



ASTES

Advances in Science, Technology & Engineering Systems Journal

VOLUME 6-ISSUE 6 | NOV-DEC 2021

www.astesj.com

ISSN: 2415-6698

EDITORIAL BOARD

Editor-in-Chief

Prof. Passerini Kazmerski
University of Chicago, USA

Editorial Board Members

Dr. Jiantao Shi
Nanjing Research Institute of
Electronic Technology, China

Dr. Lu Xiong
Middle Tennessee State
University, USA

Dr. Hongbo Du
Prairie View A&M University, USA

Dr. Nguyen Tung Linh
Electric Power University, Vietnam

Dr. Tariq Kamal
University of Nottingham, UK

Sakarya University, Turkey

**Dr. Mohmaed Abdel Fattah
Ashabrawy**
Prince Sattam bin Abdulaziz
University, Saudi Arabia

Mohamed Mohamed Abdel-Daim
Suez Canal University, Egypt

**Prof. Majida Ali Abed
Meshari**
Tikrit University Campus,
Iraq

Mr. Muhammad Tanveer Riaz
School of Electrical Engineering,
Chongqing University, P.R. China

Dr. Heba Afify
MTI university, Cairo, Egypt

Dr. Omeje Maxwell
Covenant University, Nigeria

Dr. Daniele Mestriner
University of Genoa, Italy

Mr. Randhir Kumar
National Institute of Technology Raipur, India

Regional Editors

Dr. Hung-Wei Wu
Kun Shan University, Taiwan

Dr. Maryam Asghari
Shahid Ashrafi Esfahani,
Iran

Dr. Shakir Ali
Aligarh Muslim University, India

Dr. Ahmet Kayabasi
Karamanoglu Mehmetbey
University, Turkey

Dr. Ebubekir Altuntas
Gaziosmanpasa University,
Turkey

Dr. Sabry Ali Abdallah El-Naggar
Tanta University, Egypt

Mr. Aamir Nawaz
Gomal University, Pakistan

Dr. Gomathi Periasamy
Mekelle University, Ethiopia

Dr. Walid Wafik Mohamed Badawy
National Organization for Drug Control
and Research, Egypt

Dr. Abhishek Shukla
R.D. Engineering College, India

Mr. Abdullah El-Bayoumi
Cairo University, Egypt

Dr. Ayham Hassan Abazid
Jordan University of Science and
Technology, Jordan

Mr. Manu Mitra
University of Bridgeport, USA

Dr. Qichun Zhang
University of Bradford, United Kingdom

Editorial

Advances in Science, Technology and Engineering Systems Journal (ASTESJ) is an online-only journal dedicated to publishing significant advances covering all aspects of technology relevant to the physical science and engineering communities. The journal regularly publishes articles covering specific topics of interest.

Current Issue features key papers related to multidisciplinary domains involving complex system stemming from numerous disciplines; this is exactly how this journal differs from other interdisciplinary and multidisciplinary engineering journals. This issue contains 19 accepted papers in Electrical and Information Technology domains.

Editor-in-chief

Prof. Passerini Kazmersk

ADVANCES IN SCIENCE, TECHNOLOGY AND ENGINEERING SYSTEMS JOURNAL

Volume 6 Issue 6

November-December 2021

CONTENTS

<i>Building Health Monitoring using Visualization of Sound Features Based on Sound Localization</i>	01
Mitsuru Kawamoto, Takuji Hamamoto	
<i>Light Modulation Enhancement by using an Impedance Matching Scheme for a Subcarrier Multiplexed Light Transmitter</i>	07
Seiji Fukushima, Satoshi Yanagihara, Toshio Watanabe, Tsutomu Nagayama	
<i>Lean Six Sigma Implementation in the Food Sector: Nexus between Readiness-Critical Success Factors</i>	12
Sarina Abdul Halim-Lim, Nurul Najihah Azalanzazllay, Anjar Priyono, Guven Gurkan Inan, Muhammad Iqbal Hussain	
<i>Service Lifetime Loss Evaluation Method for Transformers Exclusively Serving Solar Power Plants</i>	22
Bonginkosi Allen Thango, Jacobus Andries Jordaan, Agha Francis Nnachi	
<i>Assessment of Transformer Cellulose Insulation Life Expectancy Based on Oil Furan Analysis (Case Study: South African Transformers)</i>	29
Bonginkosi Allen Thango, Jacobus Andries Jordaan , Agha Francis Nnachi	
<i>Service Robot Management System for Business Improvement and Service Extension</i>	34
Hideya Yoshiuchi, Tomohiro Matsuda	
<i>Antecedents to Learners' Satisfaction with Serious Games: An Investigation Using Partial Least Square</i>	43
Ruben Chambilla, Daniel Tomiuk, Cataldo Zuccaro, Michel Plaisent, Prosper Bernard	
<i>Hiragana and Katakana Minutiae based Recognition System</i>	54
Piotr Szymkowski, Khalid Saeed, Nobuyuki Nishiuchi	
<i>Modelling and Testing Services with Continuous Time SRML</i>	60
Ning Yu, Martin Wirsing	
<i>Intermediation in Technology Transfer Processes in Agro-Industrial Innovation Systems: State of Art</i>	66
Leidy Dayhana Guarin Manrique, Hugo Ernesto Martínez Ardila, Luis Eduardo Becerra Ardila	

<i>The Influence of Online Learning on Learning Interest and Motivation and Their Impact on Student Achievement at Educational Technology Study Program – Ibn Khaldun University Bogor</i>	82
Mohammad Givi Efgivia, Al-Bahra, Abdul Karim Halim, R Andi Ahmad Gunadi	
<i>Innovations in Recruitment—Social Media</i>	88
Lucie Böhmová, Antonín Pavlíček	
<i>Improving of Heat Spreading in a SiC Propulsion Inverter using Graphene Assembled Films</i>	98
Sepideh Amirpour, Torbjörn Thiringer, Yasin Sharifi, Marco Majid Kabiri Samani	
<i>Control and Monitoring Systems in Medium Voltage Distribution Networks in Poland – Current Status and Directions of Development</i>	112
Janusz Gurzynski, Lukasz Kajda, Marcin Tarasiuk, Tomasz Samotyjak, Zbigniew Stachowicz, Slawomir Kownacki	
<i>Associated Risk Factors for the Development of Extensively Resistant Pulmonary Tuberculosis in the First Level of Health Care: From a Public Health Perspective</i>	119
Mauricio Mamani, Mario Chauca, Edward Huamani, Richard Gonzales	
<i>Study and Implementation of LEDs Drivers with Dimming Capability</i>	130
Brahim Mrabet, Abdeljelil Chammam	
<i>Analysis of Reading Time and the Number of Characters within One Gazing Point</i>	137
Yusuke Nosaka, Miho Shinohara, Hidemi Ishikawa, Yuko Hoshino, Mitsuho Yamada	
<i>Video Risk Detection and Localization using Bidirectional LSTM Autoencoder and Faster R-CNN</i>	145
Idir Boulfrifi, Mohamed Lahraichi, Khalid Housni	
<i>Spatio-Temporal Analysis of Drought over the Guinean-Sudano Ecological Zone, Northern Nigeria</i>	151
Gloria Langyit Dasat, Dickson Dare Ajayi, Ademola Aremu	

Building Health Monitoring using Visualization of Sound Features Based on Sound Localization[†]

Mitsuru Kawamoto^{*1}, Takuji Hamamoto²

¹Human Augmentation Research Center, National Institute of Advanced industrial Science and Technology (AIST), Kashiwa, 277-0882, Japan

²Tokyo City University, Tokyo, 158-8557, Japan

ARTICLE INFO

Article history:

Received: 21 July, 2021

Accepted: 11 October, 2021

Online: 06 November, 2021

Keywords:

Understanding sound environment

Sound visualization technique

Sound features

Sound map

Monitoring the collapse of buildings

ABSTRACT

This paper describes what can be accomplished by understanding sound environments. Understanding sound environments is achieved by extracting the features of the sound and visualizing the features. The visualization is realized by converting the three features, namely, loudness, continuity, and pitch, into RGB values and expressing the sound with color, where the color is painted in the estimated direction of the sound. The three features can distinguish falling objects within a building and roughly estimate the direction of the generated sounds. The effectiveness of the proposed sound visualization was confirmed using the sounds of cans and stones falling in a building; hence, it is shown that the proposed visualization method will be useful for monitoring the collapse of buildings by sound.

1. Introduction

Techniques for understanding environments are used to obtain sensor-related information in environmental measurements. For example, in terms of monitoring buildings, the analysis of camera and accelerometer sensor data is used to understand the current state of the building and the differences from its past state (e.g., [1]). A camera can capture images of an entire building, and an accelerometer can detect building vibrations. However, the camera only detects events within its field of view and the accelerometer only detects vibrations in its surroundings. Further, these sensors can only be installed in the monitoring area, thereby providing limited coverage, since it is difficult and risky to install sensors under collapse risk. Alternatively, sounds can reach sensors such as microphones for detection. Moreover, if a microphone array is used as the sound sensor, the sound source direction can be estimated. Therefore, sound sensors can allow more flexible monitoring of buildings than cameras and accelerometers.

Considering the benefits of sound sensing, we propose a method for analyzing environmental sounds aiming to evaluate the

difference from the past state of buildings, which is called building health monitoring, by extracting their features and estimating the direction of sound sources. As precursory sounds often occur before building collapse, such sounds may be detected and characterized by extracting sound features.

Aiming to perform building health monitoring, a visualization method for sound features based on sound localization to facilitate analysis has been introduced. In most cases, sound monitoring is simply performed by recognizing measured environmental sounds for applications, such as elderly people (e.g., [2], and references therein). On the other hand, the proposed method can provide awareness of changes in buildings by providing a visual representation of environmental sounds. The sound features of loudness, continuity, and pitch are considered. These features are quantified using spectrograms and chromagrams. Then, a visual representation of the sound features along with their estimated source position are obtained. The proposed visualization technique was evaluated using experimental sound signal data generated at a building in Gunkanjima (Hashima), Japan. Our method is a novel paradigm for sound monitoring and a novel contribution to building health monitoring.

^{*}Corresponding Author: Mitsuru Kawamoto, Kashiwa II Campus, University of Tokyo, 6-2-3 Kashiwanoha, Kashiwa, Chiba 277-0882, +81 3 3599 8543 & m.kawamoto@aist.go.jp.

[†] A preliminary version of this paper was presented at DCOSS2020.

2. Materials and Methods

2.1. Data Acquisition

We evaluate the proposed visualization technique which can be used for building health monitoring, using measurements obtained from a sound sensor (microphone array). Figure 1 shows Gunkanjima (Hashima), a World Heritage site in Nagasaki, Japan. The microphone array is installed as shown in Figure 2 on the second floor of building No. 30 (enclosed in black circle of Figure 1(a)) of Gunkanjima. The method is applied to verify the feasibility of determining the building status by analyzing environmental sounds. Building No. 30 is the oldest reinforced concrete building in Japan and is at risk of collapsing. The microphone array is comprised of 16 microphones (each arrow in the righthand side of Figure 2 shows a cluster of four microphones).



(a)



(b)

Figure 1: (a) Bird's eye view and (b) cityscape of Gunkanjima, Japan

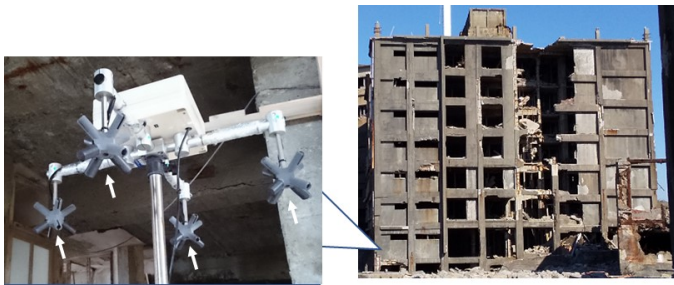


Figure 2: Microphone array installed in Building No. 30 of Gunkanjima (Hashima Island), Japan.

2.2. Sound Model of an Array with M Microphones

The sounds measured with the microphone array (Figure 2) are analyzed in the frequency domain. Then, taking the short-time Fourier transform of each microphone input at time t , the following model is obtained [3]:

$$\mathbf{y}(t, \omega) = \mathbf{A}(t, \omega)\mathbf{s}(t, \omega) + \mathbf{n}(t, \omega), \quad (1)$$

where $\mathbf{y}(t, \omega) = [Y_1(t, \omega), \dots, Y_M(t, \omega)]^T$ is the input vector of the microphone array, with m -th element $Y_m(t, \omega)$ of the vector, at time t and frequency ω ; M ($= 16$) denotes the number of microphones in the array; and the superscript T denotes the transpose. Additionally, $\mathbf{A}(t, \omega)$ is a matrix of transfer function vectors defined as

$$\mathbf{A}(\omega) = [\mathbf{a}_1(\omega), \dots, \mathbf{a}_L(\omega)], \quad (2)$$

where $\mathbf{s}(t, \omega) = [S_1(t, \omega), \dots, S_L(t, \omega)]^T$ is the source spectrum vector of the L sound sources in the measurement environment and $\mathbf{n}(t, \omega) = [N_1(t, \omega), \dots, N_M(t, \omega)]^T$, a background noise spectrum vector, carries the assumption of following a zero-mean Gaussian distribution. We assume $\mathbf{a}_l(\omega) = [A_{l1}(\omega), \dots, A_{lM}(\omega)]^T$ to be a transfer function that can be premeasured using time-stretched pulses [4]. We design time-stretched pulses to measure the impulse response, and the energy of the impulse signal is dispersed over time using a filter [4]. This filter is used to advance (or delay) the phase in proportion to the square of the frequency. Therefore, $\mathbf{a}_l(\omega)$ can be obtained by applying the inverse of the filter to the time-stretched pulse response.

2.3. Sound Source Features

In building health monitoring, it is necessary to detect the location, magnitude, and duration of an eventual collapse. Hence, the loudness, continuity, and pitch features are extracted from the sounds observed by the microphone array. To visualize the sound features, a color map is drawn at the estimated sound locations to analyze the types of generated sounds. The proposed method for the building health monitoring displays information in the color map, which we call a sound map.

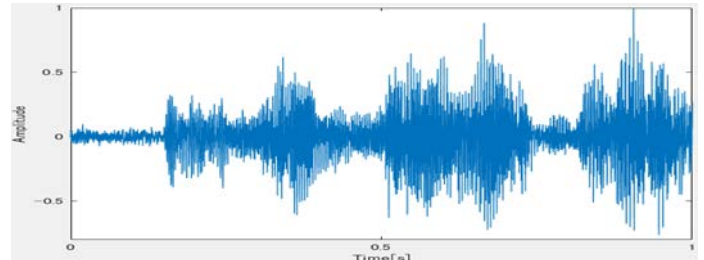


Figure 3: Sample environmental sound.

The observed environmental sounds are divided into 1-second segments, as illustrated in Figure 3. The short-time Fourier transform is applied to each segment to determine the loudness, continuity, and pitch, as detailed below.

1) Loudness Feature

Loudness is important to characterize environmental sounds during building collapse. This information is represented by a sonogram, which we obtain via the MATLAB Music Analysis toolbox [5]. The estimation of the loudness per frequency band is performed using auditory models and the function `ma_sone` in the MATLAB toolbox, where the specific loudness sensation (in sones) per critical band (in Bark scale) is calculated in six steps. 1) The fast Fourier transform is used to calculate the power spectrum of the audio signal. 2) According to the Bark scale [6], the frequencies are bundled into 20 critical bands. 3) The spectral masking effects are calculated as in [7]. 4) The loudness is

calculated in decibels relative to the threshold of hearing (decibels with respect to sound pressure level—dB-SPL). 5) From the dB-SPL values, equal loudness levels in unit phones are calculated. 6) The loudness is calculated in sones based on [8]. (Regarding detail of the six steps, see [9]). Figure 4 shows the sonogram in the Bark scale of the sound depicted in Figure 3.

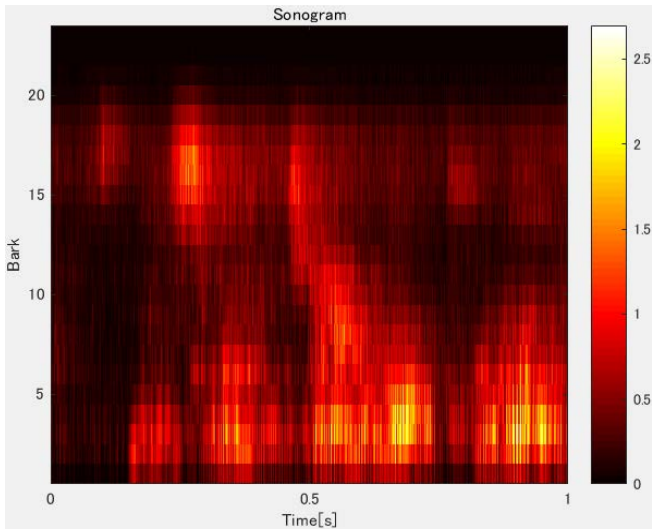


Figure 4: Sonogram in Bark scale of the sound in Figure 3.

The frequency histogram of the sonogram is also computed [5], and the histogram is resampled to 8 bits using the magnitude relation for the median of the histogram. The score expressing the information can then be obtained by converting each 8-bit value into the corresponding decimal value. An environmental sound with several variations in loudness sensation per frequency band is indicated by a higher score.

2) Continuity Feature

We use pitch variations of environmental sounds with respect to time t to measure the continuity of sounds. Therefore, continuity is expressed as a score obtained from a feature representing the pitch variation of the environmental sound.

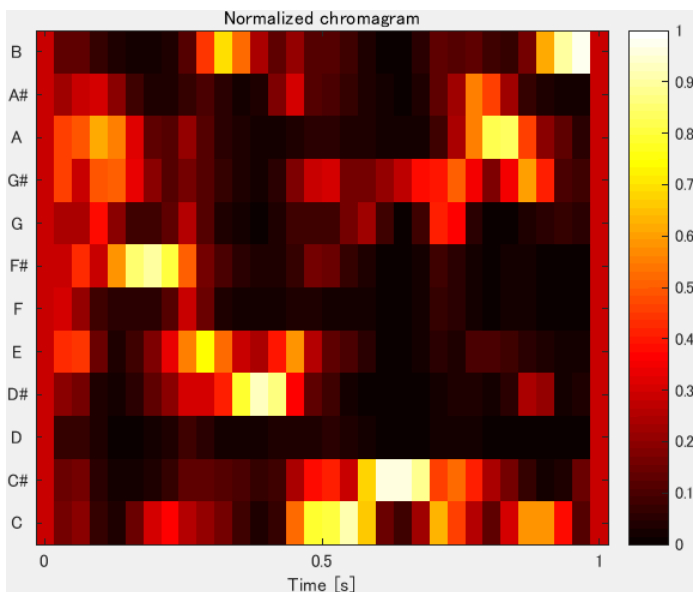


Figure 5: Chromagram of the sound in Figure 3

The corresponding chromagram, like the one shown in Figure 5, can be calculated using pitch features as those illustrated in Figure 3 by applying the method in [10]. Namely, we use the MATLAB Chroma toolbox to calculate the chromagram [11]. Subsequently, a differential chromagram is determined from the initial chromagram (Figure 5) using command diff in MATLAB, and the discrete cosine transform is applied to the differential chromagram (Figure 6). Based on this result, a time-domain histogram is then calculated. Continuity information is thus represented by a score obtained from the time-domain histogram in a method analogous to that used to obtain the loudness.

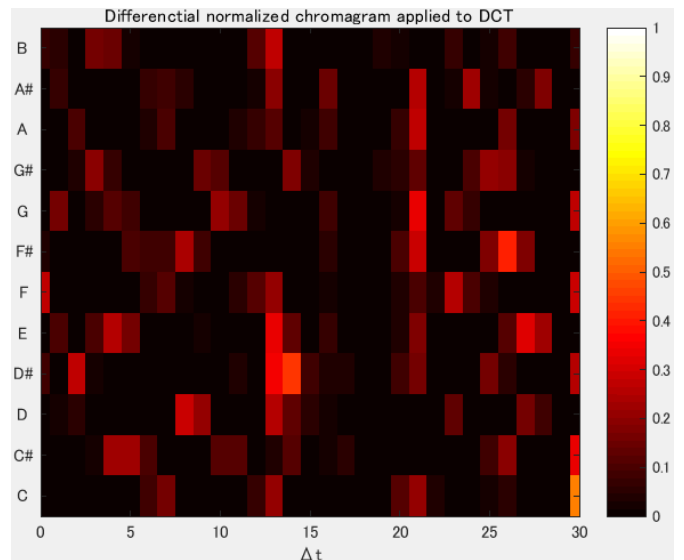


Figure 6: Differential chromagram after applying discrete cosine transform for the sound in Figure 3.

The histogram represents the variation of keys over time t . Therefore, a low score suggests that key variations of the environmental sound do not occur frequently, that is, the sound has low continuity. Conversely, a high score indicates frequent key variations of the environmental sound, indicating high continuity. As environmental sounds contain various keys, we use this score to represent continuity.

3) Pitch Feature

Pitch information is represented by the corresponding score, which is used to determine the type of environmental sound.

The spectrogram of the sound depicted in Figure 3 is shown in Figure 7. Edge extraction is applied to the spectrogram allowing the calculation of the number of pixels in its frequency feature areas and centroid frequencies. The improved affinity propagation method [12] is used to categorize the detected frequency characteristic areas of the spectrogram. More details on affinity propagation can be found in [13] and [14].

Each centroid frequency obtained by improved affinity propagation is classified into low-, medium-, or high-frequency groups. Then, a frequency group histogram can be established. The pitch is obtained from the histogram as a score, for which the calculation details can be found in [15]. A low score indicates a low dominant frequency of the environmental sound, whereas a high score indicates the presence of various frequency components.

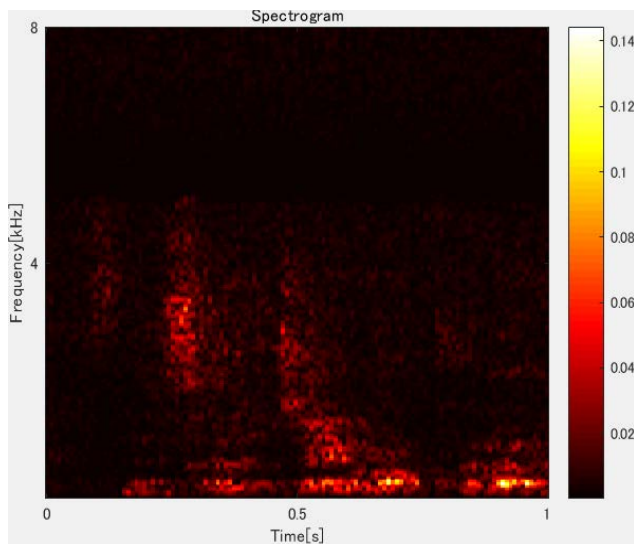


Figure 7: Spectrogram of the sound in Figure 3.

2.4. Sound Localization

Various methods for estimating sound source directions has been proposed (e.g., [16], [17], [18]). In this paper, using $A(\omega)$ obtained from (2), the MUSIC [19] is used to estimate the locations of sound sources. Details about the estimation of virtual 3D sound source positions can be found in [10].

2.5. Sound Map from Features

The loudness, continuity, and pitch scores are used to construct a sound map by representing the three scores in a red–green–blue color model. The obtained colors are then overlaid on the estimated sound source positions. Therefore, unlike conventional sound visualization methods such as power spectrum, the proposed sound map reflects not only the pitch but also the loudness and continuity of the sound in a color model, thus establishing a novel representation.

3. Experimental Results

3.1. Experimental Setup

The effectiveness of the proposed method is evaluated by using the sounds of falling cans and stones measured by the microphone array installed as shown in Figure 2.



Figure 8: Experimental scenario using falling cans and stones that hit the second floor of the building as sound sources

The three hitting points for evaluation are marked with an X in Figure 8 and are located on the 2nd floor of the building. The cans

are dropped from the 4th and 6th floors of the building to investigate the difference in the sound features caused by the drop height. Figure 8 also shows the placement of the microphone array and the person in charge of dropping the cans and stones.

3.2. Sound Maps

Figure 9 shows the sound maps obtained from the sound sources generated by cans (Figure 9(a)) and stones (Figure 9(b)) hitting on the second floor. The color differences mainly correspond to loudness and continuity, as listed in Table 1, where each feature score is the average across nine trials. There is also a difference in pitch. Hence, the sounds of hitting stones have more frequency components than those of hitting cans. The three extracted features allow to distinguish differences in objects that cause sudden sounds, and Figure 9 also demonstrates the correct localization of the sound sources.



(a)



(b)

Figure 9: Sound maps of sound sources generated by (a) cans and (b) stones hitting the 2nd floor of the building. The red circle indicates the position of the microphone array.

Table 1: The averages of loudness, continuity, and pitch scores for the sounds generated by falling cans and stones hitting on second floor of the building.

	Loudness score	Continuity score	Pitch score
Can	158.2	107.3	19.2
Stone	115.3	179.8.8	32.2



(a)



(b)

Figure 10: Sound maps of the sound sources of dropping cans at the (a) 4th and (b) 6th floors of the building.

Therefore, if there are frequency differences in sounds generated before and after a building collapses, the three features, especially the pitch, can be used to detect them.

Figs. 10 shows the sound maps of the dropping cans from the fourth (Figure 10(a)) and sixth (Figure 10(b)) floors of the building, respectively, averaged across three trials. The X marks in Figure 11 show the corresponding dropping points for the maps in Figure 10. As expected, the colors on the estimated directions of the sound sources for both floors are similar.



(a)



(b)

Figure 11: Dropping points of cans (X marks) from the (a) 4th and (b) 6th floors of the building.

Table 2: Averages of loudness, continuity, and pitch scores for the generated sounds of falling cans at the 4th and 6th floors of the building.

Floor	Loudness score	Continuity score	Pitch score
2	240.0	113.7	25.0
4	240.0	84.3	16.3
6	245.3	28.3	26.0

The colors of the sound maps in Figs. 10(a) and (b) reflect the difference in floors and are mainly related to variations in the continuity score, as listed in Table 2.

The experimental results indicate the feasibility of using the proposed method for visualizing sound features in a building health monitoring system that can roughly determine the locations and types of environmental sounds in buildings.

4. Discussion and Conclusions

Figure 12 shows snapshots of building 30 before (Figure 12(a)) and after (Figure 12(b)) a floor collapse (red area). These snapshots illustrate the difficulty in determining the collapse through the use of images from cameras. Thus, the use of sounds may allow effective monitoring of buildings such as those in Gunkanjima, which is an uninhabited island.



(a)



(b)

Figure 12: Building No. 30 (a) before and (b) after the collapse of the floor (red area).

We propose a method of visualizing sounds which can be used for implementing building health monitoring by calculating the direction and features of environmental sounds. The proposed visualization technique of sound features considering sound localization relies on sound maps that reflect the loudness, continuity, and pitch of multiple sounds.

Experiments considering falling cans and stones were considered to simulate sounds of a collapsing building. The experimental results suggest that, using the proposed method, collapsing floors and walls attributable to the damage and deterioration of buildings can be localized. Moreover, the mechanism of building collapse may be analyzed and clarified by using the proposed method.

The proposed method for building health monitoring using the sound measurement will be used to continuously monitor building No. 30 of Gunkanjima, Japan, and the building monitoring data will be collected. In addition, we will further improve the accuracy of sound localization.

Conflict of Interest

The authors declare no conflict of interest.

www.astesj.com

Acknowledgment

This work was partly funded by Japan Society for the Promotion of Science (grant number 16H02911 and 17HT0042) and JST CREST under grant JPMJCR18A4.

References

- [1] H. Hamamoto, N. Kurata, S. Saruwaari, M. Kawamoto, A. Tomioka, T. Daigo, "Field test of change detection system of building group in preparation for unexpected events in Gunkanjima," *AIJ J. Technol. Des.*, **24**(57), 553-558, 2018, doi:10.3130/aijt.24.553 (article in Japanese with an abstract in English).
- [2] A. Sasou, N. Odontselgel, S. Matsuoka, "An acoustic-based tracking system for monitoring elderly people living alone," in 4th International Conference on Information and Communication Technologies for Ageing Well and e-Health, **2018-March**, 89-95, 2018, doi:10.5220/0006664800890095.
- [3] A. Quinlan, M. Kawamoto, Y. Matsusaka, H. Asoh, F. Asano, "Tracking Intermittently Speaking Multiple Speakers Using a Particle Filter," *EURASIP Journal on Audio, Speech, and Music Processing*, **2009**(1), 1-11, 2009, doi:10.1155/2009/673202.
- [4] Y. Suzuki, F. Asano, H. Kim, T. Sone, "An optimum computer-generated pulse signal suitable for the measurement of very long impulse responses," *The Journal of the Acoustical Society of America*, **97**(2), 1119-1123, 1995, doi: 10.1121/1.412224.
- [5] E. Pampalk, *Music analysis toolbox* [Online]. Available: <http://www.pampalk.at/ma/documentation.html>.
- [6] H. Fastl, E. Zwicker, *Psychoacoustics Facts and Models*, Springer-Verlag Berlin Heidelberg, 1999.
- [7] M. R. Schroeder, D.S. Atal, J.L. Hall, "Optimizing digital speech coders by exploiting masking properties of the human ear," *The Journal of the Acoustical Society of America*, **66**(6), 1647-1652, 1979, doi:10.1121/1.383662.
- [8] R. A. W. Bladon, B. Lindblom, "Modeling the judgment of vowel quality differences," *The Journal of the Acoustical Society of America*, **69**(5), 1414-1422, 1981, doi: 10.1121/1.385824.
- [9] E. Pampalk, A. Rauber, D. Dueck, "Content-based organization and visualization of music archives," in the 10th ACM International Conference on Multimedia 2002, 570-579, 2002, doi:10.1145/641007.641121.
- [10] M. Kawamoto, T. Hamamoto, "Building Health Monitoring Using Computational Auditory Scene Analysis," in 16th Annual International Conference on Distributed Computing in Sensor Systems, DCOSS 2020, 144-146, 2020, doi:10.1109/DCOSS49796.2020.00033.
- [11] M. Müller, S. Ewert, "Chroma toolbox: Matlab implementations for extracting variants of chroma-based audio features," in the 12th International Society for Music Information Retrieval Conference, ISMIR 2011, 215-220, 2011.
- [12] M. Kawamoto, "Sound-Environment Monitoring Method Based on Computational Auditory Scene Analysis," *Journal of Signal and Information Processing*, **8**(2), 65-77, 2017, doi:10.4236/JSIP.2017.82005.
- [13] B. J. Frey, D. Dueck, "Clustering by passing messages between data points," *Science*, **315**(5814), 972-976, 2007, doi:10.1126/SCIENCE.1136800.
- [14] K. Wang, J. Zhang, D. Li, X. Zhang, T. Guo, "Adaptive Affinity Propagation Clustering," *Acta Automatica Sinica*, **33**(12), 1242-1246, 2007.
- [15] M. Kawamoto, "Sound-environment monitoring technique based on computational auditory scene analysis," in 25th European Signal Processing Conference, EUSIPCO 2017, 2017-January, 2516-2524, 2017, doi:10.23919/EUSIPCO.2017.8081664.
- [16] T. Nishiura, T. Yamada, S. Nakamura, K. Shikano, "Localization of multiple sound sources based on a CSP analysis with a microphone array," in 2000 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), **2**, 1053-1056, 2000, doi:10.1109/ICASSP.2000.859144.
- [17] H. L. Van Trees, *Optimum Array Processing*, Wiley Interscience: New York, 2002, doi:10.1002/0471221104.
- [18] D. H. Johnson, D. Dudgeon, *Array Signal Processing: Concepts and Techniques*, New Jersey: Englewood Cliffs: PTR, Prentice Hall, 1993.
- [19] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Transactions on Antennas and Propagation*, **AP-34**(3), 276-280, 1986, doi:10.1109/TAP.1986.1143830.

Light Modulation Enhancement by using an Impedance Matching Scheme for a Subcarrier Multiplexed Light Transmitter

Seiji Fukushima*, Satoshi Yanagihara, Toshio Watanabe, Tsutomu Nagayama

Graduate School of Science and Engineering, Kagoshima University, Korimoto 1, Kagoshima-shi, 890-0065, Japan

ARTICLE INFO

Article history:

Received: 30 July, 2021

Accepted: 15 October, 2021

Online: 06 November, 2021

Keywords:

Radio-on-Fiber

Laser Diode

Subcarrier Multiplexing

Impedance Matching

ABSTRACT

We propose and analyze new impedance matching schemes to enhance applied voltage to an optical modulator and light modulation for a subcarrier multiplexed light transmitter or a radio-on-fiber transmitter that carries radio-frequency signal through an optical fiber. Our proposal includes two methods using a quarter-wavelength impedance transformer and a tapered microstrip line for impedance matching between a driver circuit and an electro-absorption modulator. Simulation results for both two schemes show that large enhancement is observed for 20 GHz and lower frequency and that some resonant boosts are observed for higher frequency as well. Discussions are described to design a circuit with improved performances. It is shown that our schemes can drive the electro-absorption modulator at a voltage higher than 1 V that is sufficient to drive the modulator.

1. Introduction

The use of optical fiber network has become more popular to send both broadband digital data [1] and modulated radio frequency (RF) signal [2]. The latter is referred to as radio-on-fiber (RoF) technology, where the signal is transmitted as subcarrier-multiplexed (SCM) laser light. The digital network requires optical devices such as a laser diode (LD), a light modulator, and a photodiode to have broadband response. In contrast, the RoF network does not require such broadband response but requires only a few spot-frequency or segmented-frequency responses. Such RoF systems are employed in a network between a data center and an access point for cellular services, where radio carrier frequency becomes higher, especially at the fifth generation cellular services using a Ka-band [3]. In addition to these services, some satellites employ a Ku-band so that the use of optical fiber link is practical between an antenna and a control site in an earth station. For example, some small cube satellites, developed by universities and consortiums, sent the Ku-band signals to earth stations in its downlink system [4]. A RoF link was employed in the earth station to reduce RF signal loss in conventional coaxial cables [5-7]. In addition to the low link loss, the optical fiber will benefit us in many aspects such as weight and cost.

To achieve a higher frequency operation and to avoid chromatic dispersion in an optical fiber, an electro-absorption modulator integrated laser diode or an externally modulated laser (EML) would be one of the strong candidates to develop a SCM

transmitter. The EML is monolithic integration of an electro-absorption modulator (EAM) and a distributed feedback laser diode (DFB-LD), as shown in Figure 1. A current source is connected to the DFB-LD to obtain continuous laser light and a digital or a RF source to the EAM to obtain modulated light. The EML emits SCM light when the RF signal is applied to the EAM. Both the optical devices can concentrate on either high-speed modulation or single-mode emission of laser light [8, 9]. Most EMLs are designed for 1.5- μm wavelength networks and are widely used in commercial optical networks due to their performances, compactness, and reliability. In addition to the commercial optical networks, they are applied to the shorter-length, high-throughput Ethernet services [10-12], which makes an optical transmitter smaller in size and compatible with optical transmitter standards. Challenges have been done to develop a 1.3- μm wavelength device [11].

A link length limit due to chromatic dispersion has been analyzed for the RoF link in the satellite earth station that employs the EML [13], which is different from the digital network systems. Effort has been made to achieve a more-functional device such as radio frequency conversion in the SCM light [14].

One of the interesting approaches to enhance the response of EMLs is the use of via connection in three-dimensional substrate to mount the EML [12]; however, this technique needs the state-of-the-art manufacturing process. On the other hand, this paper focuses on a passive transmission line between a RF source and the EAM, as shown in Figure 2. The EAM is a reversely biased

*Corresponding Author: Seiji Fukushima, fukushima@eee.kagoshima-u.ac.jp

www.astesj.com

<https://dx.doi.org/10.25046/aj060602>

diode with an external resistive load of 50 Ω in general. The conventional 50-Ω matching may be a bottle neck for driving the device. Our idea to overcome this issue is a simple matching scheme such as a quarter-wavelength impedance transformer (QIT) and a tapered microstrip line (TML) that can match the driver and the modulator with different impedance [15-17]. Hence, we can increase the external load resistance to increase the applied voltage to the modulator and its modulation index.

Our proposed scheme is described in Section 2 and some results in Section 3. We discuss characteristics and features of the QIT and TML in Section 4. We will conclude this paper in Section 5.

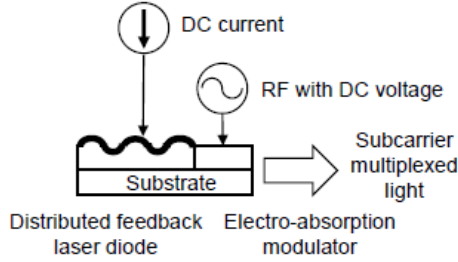


Figure 1: Cross Section of an Electro-Absorption Modulator Integrated Laser Diode (EML) and Power Sources

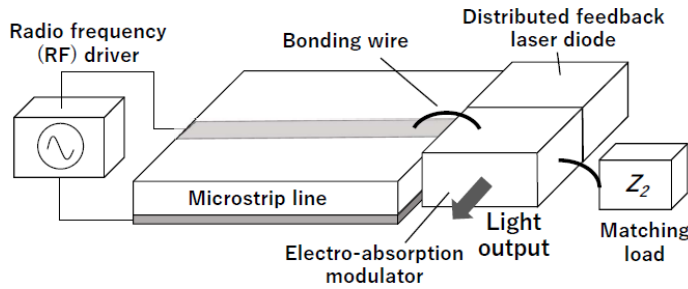


Figure 2: Schematic Model Including a RF Source, a Transmission Line, and an EML

2. Proposed Structure and Circuit Model

A circuit around the EML is shown in Figure 2. A RF signal is transferred from a RF driver to the EML in the EML through a microstrip line. Impedance of the driver, the EML, and the microstrip line is 50 Ω in general. We should notice the reversely biased EML has very high impedance but it can be impedance-matched by using an external parallel resistance of approximately 50 Ω. We could choose favorite impedance if the entire circuit can be matched by other methods. We propose the use of QIT and TML that are shown in Figure 3 (a) and (b), respectively. The QIT must follow equation below

$$Z_0^2 = Z_1 Z_2 \quad (1)$$

where Z_0 is microstrip line impedance, and Z_1 and Z_2 impedance of the RF driver and the parallel load to the EAM, respectively, and the microstrip line length must be a quarter of a wavelength ($\lambda/4$) or multiplied length by odd numbers ($3\lambda/4, 5\lambda/4, \dots$). If the TML is long enough and if the impedance at the both end matches with adjacent circuits, no electric reflection occurs at the boundaries.

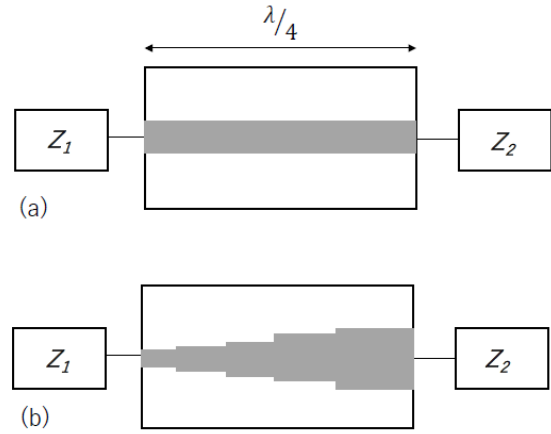


Figure 3: Structure of the Transmitter with (a) a Quarter-Wavelength Impedance Transformer (QIT) and (b) a Tapered Microstrip Line (TML).

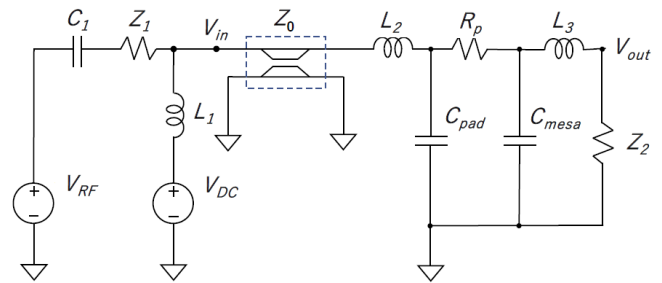


Figure 4: Equivalent Circuit Consisting of a RF Source, a Transmission Line, and an EAM.

Figure 4 shows the entire equivalent circuit of the driver, the RF transmission line and, the EAM. A symbol, V_{RF} , on the left hand with its load impedance Z_1 is the RF driver and the next the reverse voltage source, where the both sources are connected to a RF line Z_0 through a bias-tee of C_1 and L_1 . The RF line Z_0 is placed on the center to connect the RF source and the EAM. All the electric signals, dc and ac, pass through Z_0 to the EAM section consisting of two inductors, one resistor, and two capacitors; L_2 and L_3 : inductances of bonding wires to the microstrip and to the external load, respectively, R_p : a resistance of electrode, C_{mesa} : a capacitance of the modulator diode, and C_{pad} : a capacitance of an electrode pad. An external matching resistive load Z_2 is connected to the EAM in parallel. In our simulation, we used the values as follows: $V_{RF}=1$ V, $R_p=5$ Ω, $L_1=79.4$ μH, $L_2=0.11$ nH, $L_3=0.04$ nH, $C_1=51$ nF, $C_{pad}=0.029$ pF, $C_{mesa}=0.45$ pF. These values have been obtained from measurements or theoretical calculations. Other parameters will be described when they are needed. We assumed the central operation frequency to be 28 GHz, one of the fifth generation cellular service bands. We analyzed the proposed circuits by using a circuit simulator LTspice. A tapered microstrip line circuit model was not available so that cascaded, stepped microstrip line (Tline) models, shown, in Figure 3(b), were employed instead. The number of steps was five. As a result, we were able to use the LTspice circuit models including Tline for all the components. Any commercially available substrates were not assumed except for MEGTRON7 to have general discussions in this paper.

If we assume that the DFB-LD in the EML is operated with an automatic-power-control current supply, the DFB-LD emits

continuous single-mode laser beam. The EAM modulates the laser beam into SCM laser light that carries information of RF signal.

3. Simulation Results

3.1. Results for Quarter-wavelength Impedance Transformer

Assuming a dielectric constant of the substrate as 3.4 and the output impedance of EML, Z_2 , as 50 (Ω , gray), 70 (Ω , blue), or 100 (Ω , orange), we calculated terminal voltage spectra for circuits with QIT, as shown in Figure 5. Here, the impedance of the RF driver, Z_1 , can be calculated as 50, 35.7, or 25 Ω by using Eq. (1). The terminal voltage is defined by the voltage across the EAM. The voltage at the lower frequency or at the left edge in Figure 5 is higher for the higher impedance and the enhancement ratio is nearly proportional to Z_2 . This tendency can be found at most of the frequency ranges between 1 and 20 GHz. The enhancement is obtained simply due to higher voltage division in Z_2 . The voltage curves decrease as the frequency becomes higher in a case, $Z_2 = 50 \Omega$, which is the same with the conventional structure, while some peaks are found for Z_2 of 70 and 100 Ω . The peaks are obtained at 6.1, 30, 60, and 92 GHz for Z_2 of 100 Ω . In addition to the simple boost of the voltage in the lower frequency range, we obtained response peaking in the higher frequency range, which is due to resonance by some inductive and capacitive elements.

We obtained powerful voltage boost by using the QIT microstrip line between the driver and the EML with the load impedance higher than 50 Ω , as expected. The resonant peaks enable the applied voltage boost at three or four frequencies.

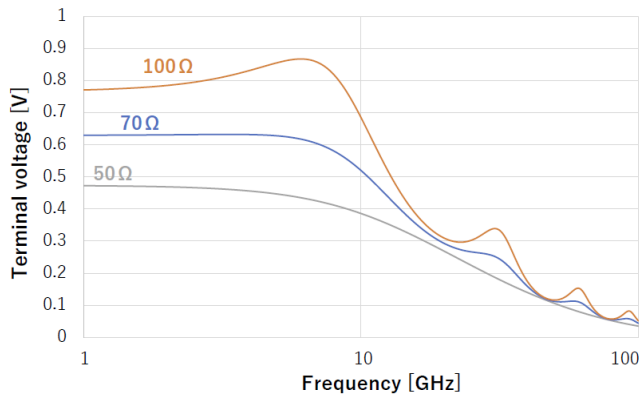


Figure 5: Spectra of the Terminal Voltage across the EAM when the QIT is Employed.

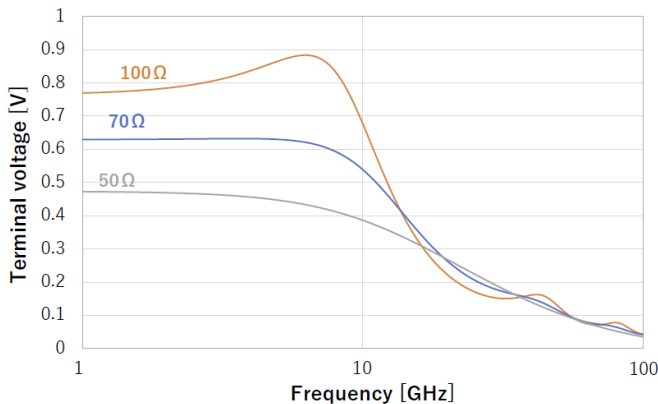


Figure 6: Spectra of the Terminal Voltage across the EAM when the TML is Employed.

3.2. Results for Tapered Microstrip Line

Figure 6 shows the calculated terminal voltage spectra for the circuit including the TML. The results are the same with the results of QIT circuit qualitatively. In the lower frequency range, the terminal voltage is amplified with magnification of Z_2/Z_1 approximately, which means that we can obtain higher voltage from the higher impedance of Z_2 . In the higher frequency range, fewer number of resonant peaks are found and their amplitudes are smaller even if they exist. The peak at 6.4 GHz is observed as very similar to the results of QIT while two lower peaks are observed at 41.7 and 79.8 GHz for Z_2 of 100 Ω .

When we have to determine which circuit we should adopt, we have to note that both will function at the same performance for frequencies lower than 20 GHz but that the details differ between the QIT and the TML. The QIT exhibits a higher voltage or a high light modulation index when we manufacture all the devices and a circuit, as specified and as designed. If we are not able to manufacture the devices as specified nor as designed, the TML would be a good choice since it allows some margin in the resonant frequencies to provide lower and broader peaks.

For clearer comparison, we show terminal voltage dependence on the load impedance Z_2 for a frequency at 1 GHz (blue), the first resonant frequency near 6 GHz (orange), and the second resonant frequency near 30 GHz (gray) in Figure 7. The plots and lines are overlapped for the two models of QIT and TML at 1 GHz and the first resonant frequency while they differ at the second resonant frequency.

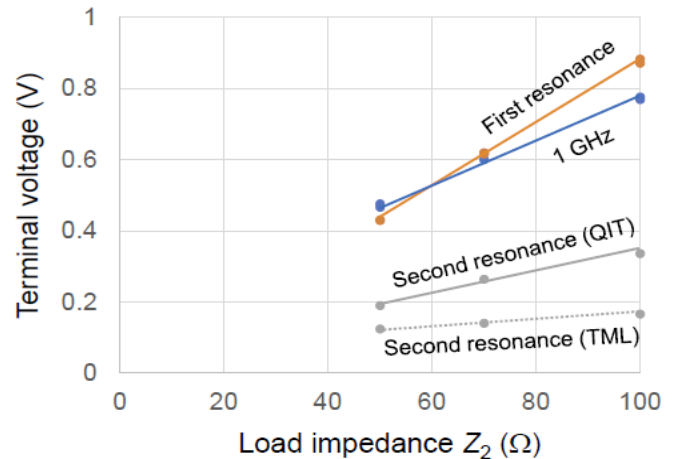


Figure 7: Terminal Voltage Dependence on the load impedance Z_2 .

The terminal voltages always increase as the load impedance becomes higher. All the lines are linear in a Z_2 range between 50 and 100 Ω ; however, the number of calculated points of the load impedance is limited to only three. Thus, we can double the voltage at 1GHz and the first resonant frequency when we employ a load impedance of 100 Ω other than the original 50 Ω . When we pay our attention to the second resonance, we should choose the QIT model (solid line) since it provides much higher voltages than the TML model (broken line). In both cases, the terminal voltages are even higher with the 100- Ω load than the voltages with the original 50- Ω load. Our ideas are worth applying to the optical transmitter especially when we consider that the second resonant

frequency, 30 GHz, is within the fifth-generation cellular service band.

4. Discussion

Difference of the improvement between the two proposed structures, QIT and TML, is small at a 20-GHz frequency and lower. However, we should analyze the results in the previous section for better performance at higher frequencies. Resonant frequency dependences on the order of the resonance are shown in Figures 8 and 9 for the QIT and TML, respectively. We plotted these figures, assuming dielectric constants to be 3.4, 10, and 15, where some parameters of MEGTRON7 were adopted for the constant of 3.4. We took account of dielectric losses due to the substrates. The load impedance to the EML, Z_2 , was assumed to be 100Ω . The number of the orders is larger within the plotted frequency range as the dielectric constant becomes larger. The number of plotted points are three, four, and, five for dielectric constants of 3.4, 10, and, 15, respectively, as shown in Figure 8. The relationship between the resonant frequency and the order is close to a linear function. The number of the points for the QIT is larger than that for the TML. We obtained the dependency on the dielectric constant, the order of resonant peak, and the type of the matching structure, which provides us more flexibility in designing and manufacturing the matching circuits.

The results described above focus only on the resonant frequency; however, the terminal voltage at a resonant frequency is of importance as well. Figures 10 and 11 show a voltage across the EAM versus a resonant frequency for the QIT and TML, respectively. The employed parameters are the same with those for Figures 8 and 9. The figures include results for the conventional $50\text{-}\Omega$ -matching circuit. The curves are overlapped for all the three dielectric constants and the voltages are inversely proportional to the logarithmic plot of the frequencies, as shown in Figure 10. At a 30-GHz frequency near the fifth generation cellular band, the terminal voltage of 0.34 V is achieved for all the dielectric constants. We obtained only 0.2 V from the conventional structure in contrast. Similar results are obtained for the TML, as shown in Figure 11. Here, we have to note the terminal voltages are dependent on a dielectric constant at the TML model. The lower dielectric constant substrate provides us higher voltage.

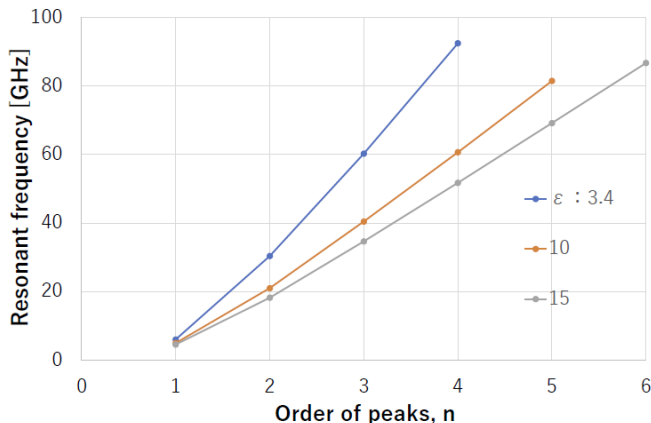


Figure 8: Resonant Frequency Dependence on the Order of Peak for the QIT Model.

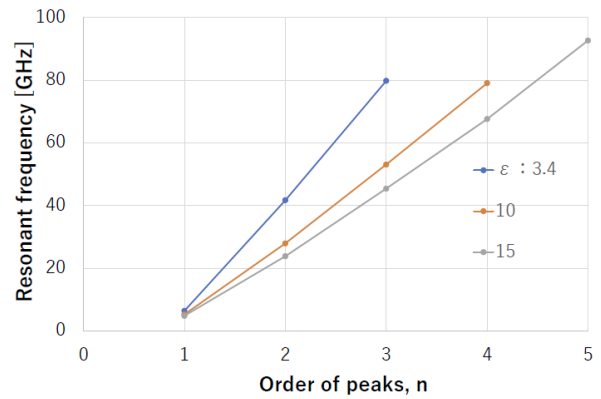


Figure 9: Resonant Frequency Dependence on the Order of Peak for the TML Model.

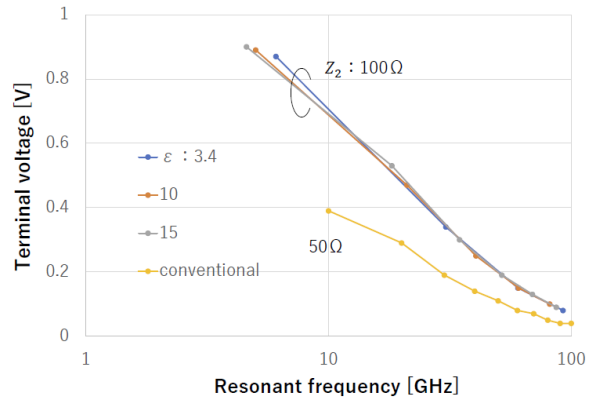


Figure 10: Terminal Voltage across the EAM versus Resonant Frequency for the QIT Model.

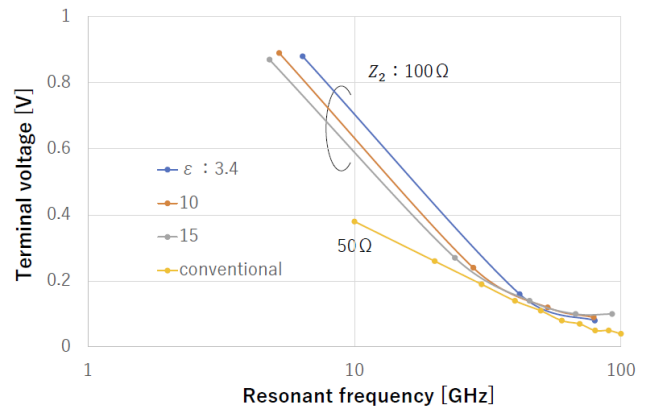


Figure 11: Terminal Voltage across the EAM versus Resonant Frequency for the TML Model.

Numeric analysis has been successfully done for the proposed structure that employs a higher impedance load in parallel to the EAM. In addition, we analyzed the behavior at higher impedance, Z_2 , for the QIT. Figure 12 shows a terminal voltage dependence on a resonant frequency for the QIT circuit with the impedance Z_2 of 500Ω . The tendency is similar qualitatively with the previous results. It should be noted that the terminal voltage becomes approximately 5.5 times those obtained from the conventional circuit. The voltages at resonant frequencies 30 and 60 GHz are 1.05 and 0.52 V, respectively, which are sufficient to drive the EML and to transmit the SCM laser light. This trial may not be applied to the TML model unfortunately since such a high-

impedance microstrip line cannot be fabricated on a typical substrate.

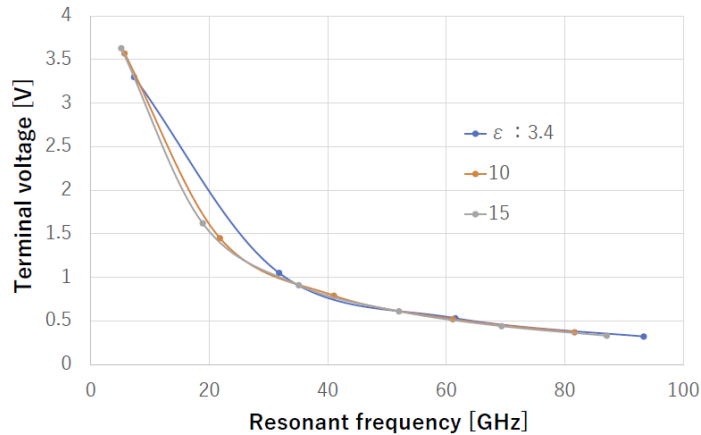


Figure 12: Terminal Voltage across the EAM versus Resonant Frequency for the QIT Model, Assuming the Load Impedance, Z_2 , is 500 Ω .

5. Conclusions

We have proposed new structures to boost applied voltage to the EAM and its modulation index in a SCM or RoF transmitter that carries radio-frequency signal through an optical fiber. The two proposed models employ a QIT and a TML, both are capable of impedance matching for two adjacent devices with different impedance. Our analysis has been done numerically by using an electronic circuit simulator. Comparing with the conventional circuit, the voltage applied to the modulator increases approximately at the ratio of the load impedance to the transmission line impedance at 20 GHz and lower frequency in both the structures. Due to the capacitive and inductive elements in a circuit, some resonant peaks were found at higher frequency in both the models. The applied voltage to the EAM is enhanced at the resonant frequencies.

To further improve performances, we have analyzed characteristics of our models with parameters of dielectric constant of substrates and load impedance. The results are similar between the two models. The impedance of the QIT microstrip line is constant and a standard one of 50 Ω . On the other hand, the impedance must be varied in a wide range from Z_1 to Z_2 on a TML substrate, which might be difficult to fabricate in some cases. We should note the matter described above in designing a circuit.

It is finally concluded our proposal does not require any active device but that it can enhance the voltage across the EAM and modulation index of the SCM light. This technique would be easily applied to most of the light transmitters such as a direct modulation laser and a Mach-Zender light modulator. In addition, this kind of RoF link would be applied to higher-frequency applications [18].

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

This work was supported by JSPS KAKENHI Grant Number 17K06432.

References

- [1] E. B. Basch, R. Egorov, S. Gringeri, S. Elby, "Architectural tradeoffs for reconfigurable dense wavelength-division multiplexing systems," *IEEE Journal of Selected Topics in Quantum Electronics*, **12**(4), 615-626, 2006. DOI: 10.1109/JSTQE.2006.876167
- [2] S. Torihata, H. Tsuchiya, K. Haeiwa, T. Yamashita, S. Tanizawa, Y. Ozaki, "The optical remote antenna for the digital TV broadcasting relay networking," *International Topical Meeting on Microwave Photonics*, Awaji, Japan, 375-378, 2002.
- [3] M. Sung, J. Kim, E.-S. Kim, S.-H. Cho, Y.-J. Won, B.-C. Lim, S.-Y. Pyun, H. Lee, J. K. Lee, J. H. Lee, "RoF-based radio access network for 5G mobile communication systems in 28 GHz millimeter-wave," *Journal of Lightwave Technology*, **38**(2), 409-420, 2020/doi 10.1109/JLT.2019.2942636.
- [4] Y. Sakamoto, N. Nishio, "Determination using radio interferometer of small-diameter antennas for LEO satellites," *IEEE Transaction on Aerospace and Electronic Systems*, **47**(3), 2111-2118, 2011.
- [5] S. Fukushima, N. Miura, T. Shimaki, T. Funasako, K. Yamashita, T. Hachino, Y. Igarashi, "Electro-absorption modulator integrated laser application to a cube satellite earth station," *2013 Conference on Lasers and Electro-Optics Pacific Rim (CLEO-PR 2013)*, paper TuPO-11, 2013..
- [6] S. Fukushima, T. Shimaki, T. Funasako, N. Miura, K. Yamashita, T. Hachino, Y. Igarashi, "Cube satellite earth station based on bidirectional WDM radio-on-fiber link," *2013 Asia-Pacific Radio Science Conference*, paper DCA-6, 2013.
- [7] S. Fukushima, T. Shimaki, K. Yamashita, T. Funasako, T. Hachino, "Frequency division multiplexed radio-on-fiber link employing an electro-absorption modulator integrated laser diode for a cube satellite earth station," *IEICE Transaction on Electronics*, **E99-C**(2), 212-218, 2016.
- [8] H. Kamioka, M. Fukuda, T. Tohmori, Y. Itaya, "Reliability of an electro-absorption modulator integrated with a distributed feedback laser," *Pacific Rim. Conference on Lasers and Electro-Optics (CLEO-PR '99)*, **4**, 1202-1203, 1999.
- [9] K. Wakita, K. Sato, K. Kotaka, M. Yamamoto, T. Ktaoka, "20-Gbit/s, 1.55-mm strained-InGaAsP MQW modulator integrated DFB laser module," *Electronics Letters*, **30**(4), 302-303, 1993.
- [10] H. Takahashi, T. Shimamura, T. Sugiyama, M. Kubota, K. Nakamura, "High-power 25-Gb/s electroabsorption modulator integrated with a laser diode," *IEEE Photonic Technology Letters*, **21**(10), 633-635, 2009.
- [11] T. Fujisawa, K. Takahata, W. Kobayashi, T. tadokoro, N. Fujisawa, S. Kanazawa, F. Kano, "1.3-mm, 50-Gbit/s EADFB lasers for 400GbE," *2011 Optical Fiber Communication Conference and Exposition*, paper OWD4, 2011.
- [12] X. Zhang, X. Liu, A. Rainier van Dommele, M. K. Matters-Kammerer, "Dual-channel 56 Gb/s PAM-4 electro-absorption modulator driver for 3D wafer scale packaging," *2018 Asia-Pacific Microwave Conference (APMC 2018)*, paper WE4-K-03, 2018.
- [13] T. Uezono, S. Fukushima, T. Watanabe, T. Nagayama, "Link length analysis of a radio-on-fiber system for a cube satellite earth station," *Annual Conference on Engineering and Applied Science (ACEAT)*, 18-23, 2017.
- [14] S. Fukushima, Y. Doi, Y. Matsuoka, H. Takeuchi, "EA-modulator/LD integrated device application to mm-wave optical link," *IEICE Society Conference*, paper B-5-219, 1998.
- [15] D. M. Pozer, *Microwave Engineering*, 4th Ed., Wiley, 2011.
- [16] A. Ueda, T. Noguchi, H. Iwashita, Y. Sekimoto, M. Ishiguro, S. Takano, T. Nagatsuma, T. Ito, A. Hirata, T. Ishibashi, "W-band waveguide photomixer using a uni-traveling-carrier photodiode with 2-mW output," *IEEE Transaction on Microwave Theory and Techniques*, **51**(5), 1455-1459, 2003. DOI: 10.1109/TMTT.2003.810154
- [17] S. Yanagihira, S. Fukushima, T. Watanabe, T. Nagayama, "Impedance-matched feedline from a semiconductor laser emitting subcarrier multiplexed light," *25th Optoelectronics and Communications Conference*, paper T4-4.5, 2020.
- [18] M. Nishio, Q. Liu, T. Miyazaki, M. Hirata, Y. Kuroki, M. Kusuhata, N. Iwashita, C. Minamitake, S. Yasuda, N. Iino, T. Omodaka, O. Kameya, N. Kawano, T. Suzuyama, Y. Shibuya, N. Kurihara, "Observation site atmospheric phase fluctuations observed by three-element VLBI," *IEEE Transaction on Antennas and Propagation*, **55**(7), 2056-2063, 2007.

Lean Six Sigma Implementation in the Food Sector: Nexus between Readiness-Critical Success Factors

Sarina Abdul Halim-Lim^{*1}, Nurul Najihah Azalanzazllay¹, Anjar Priyono², Guven Gurkan Inan³, Muhammad Iqbal Hussain⁴

¹Department of Food Technology, Faculty of Food Science and Technology, Universiti Putra Malaysia, 43400, Malaysia

²Department of Management, Universitas Islam Indonesia, Yogyakarta, 55000, Indonesia

³Industrial Engineering Department Samsun University, Samsun, 55020, Turkey

⁴Faculty of Mechanical Engineering Technology, Universiti Malaysia, Perlis, 02400, Malaysia

ARTICLE INFO

Article history:

Received: 15 July, 2021

Accepted: 24 October, 2021

Online: 09 November, 2021

Keywords:

Lean Six Sigma

LSS Readiness

Critical Success Factors

Change Management

Food sector

ABSTRACT

Lean Six Sigma (LSS) is a renowned approach for boosting operational excellence and competitive advantage through integrated core objectives of value creation and variation reduction. Despite its proven benefits in many leading companies, LSS implementation in the food sector is still behind compared with other sectors. LSS implementation is costly, and most businesses have failed due to a lack of preparation and an unsupportive organizational culture. Therefore, there is a need to identify LSS readiness factors that suit the food sector to minimize the risk of implementation failure in the industry. The current study concentrates on the LSS pre-implementation phase to determine the competency criteria to adopt LSS customized for the food business. This study will explore the LSS readiness criteria during the pre-implementation stage and critical success factors (CSFs) during the implementation stage in the food sector through Lewin's Change Theory. Twelve food sector employees who were associated with quality management activities were interviewed using a semi-structured approach. The interview was recorded, transcribed and the transcription was analyzed using content analysis. The results showed six readiness themes in the food manufacturing sector with twenty-nine LSS readiness attributes, while seventeen factors out of thirty-one CSFs for the LSS at the implementation stage. The identified readiness factors are management commitment and leadership (ten attributes), organizational culture (nine attributes), employee involvement (six attributes), process management (four attributes), project management (four attributes) and external factors (three attributes). Through Pareto analysis, the most prioritized CSFs are from top management and leadership and employee involvement themes, with the training program being identified as the most important LSS CSFs (85%). This study will serve as a foundation for a benchmarking tool for managers to improve the effectiveness of an LSS implementation in the food sector.

1. Introduction

A strategic business plan is required in the food sector to minimize the process errors that would impact the product quality. This study which, is an extension of work initially presented in the 2020 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM), addresses Lean Six Sigma (LSS) application as the approach. Recently, there was an increase in quality and lower-cost products demand. The food business experienced challenges due to high customer

expectations, strict food standards, and a competitive market [1]. Furthermore, the food businesses have been combating waste that can occur at any point in the supply chain [2]. But waste elimination requires a structured way of thinking, where new manufacturing initiatives should be considered a valuable innovation for the food businesses. Since 2000, the model for integrating Lean and Six Sigma has been sought by the organization globally to improve process performance, maximize profit and minimize operating costs [3]. Lean has a role to reduce non-value-added activities, which is also considered waste in the

*Corresponding Author: Sarina Abdul Halim-Lim, sarinalim@upm.edu.my

customer's eyes. Whilst Six Sigma is viewed as a problem-solving methodology focusing on making a process effective by reducing process variations [4]. The hybrid of LSS depicted the mixture of best values of these two effective approaches, enabling both variations and waste reduction across the process [2, 3]. In [4], the author confirmed the prominent advantage of LSS projects in this sector mostly is cost saving. This is what drives and motivates quality executives in the food sector to continue their CI efforts, however, there is another motivation such as the need to abide by the laws and regulations. By reducing overfilling confectionary cream that can be expensive [2], LSS enabled food companies not to distribute underweight food products which can cause them to be penalized. LSS activities also encourage circular economy activities by decreasing waste of overfilling cream [5].

Studies have highlighted that the food business has been slower to adopt the LSS mindset and has remained passive compared to other industries [4] and [6]. The reason is that the cost of products is relatively low compared to the cost structure of the business [7], prioritization on the strict regulations that characterize the food sector [8], the nature of the food sector being conservative towards change.

The perceived management difficulties that the businesses in this sector encounters, for instance resistance to change, may be mitigated through LSS preparedness evaluation outlined by organizational change theory [9]. Resistance to the new effort such as LSS, occurred at various levels within an organization. It considered as a significant issue that could result in the failure of LSS implementation [1], [10] and [11]. Thus, the importance of conducting an assessment on the state of LSS readiness prior the implementation need to be highlighted, because preparedness is one of the LSS implementation success factor by [11]. It also will facilitate the effort to integrate LSS with management innovation such as digital technologies [12]. The pre-implementation stage receives insufficient attention compared to the LSS implementation phase elements, such as crucial success factors (CSFs) and barriers.

Thus, this paper aims to investigate the factors that contribute to LSS preparedness customized to the food context. Variables contributing to the preparedness of LSS and their impact on food sector readiness for LSS implementation must be determined to achieve this research aim.

2. Literature Review

2.1. Lewin's change theory

Readiness study was pioneered by [13] to technological adoption, in which he established that the term is semantically contrary to the resistance to change. [14] associated readiness for a practice that promotes organizational transformation by minimizing inhibitors of success or uncovering the important points to improve the success of the change. Based on Lewin's theory of change model, it is appropriate to explore the factors contributing to unfreezing and freezing phases to drive any kind of change [15]. Unfreezing refers to accepting the need to make

the change and preparing the people ready for the change [15]. By adopting the theory, this study intends to understand two phases of the theory which are the unfreezing, followed by the change phase. in the context of LSS as depicted in the research framework in Figure 1. Several studies have identified the importance of readiness outside of the implementation process and encouraged the longevity of the LSS system [10]. Due to the large venture capital and continuous commitments involved with the adoption of LSS, the company should tactically plan to avoid misspending on CI activities [6] and [16]. Furthermore, the understudied LSS issue especially reluctance towards change, high turnover and complacency of current process performance, are highly involved in the two phases of Lewin's change theory.

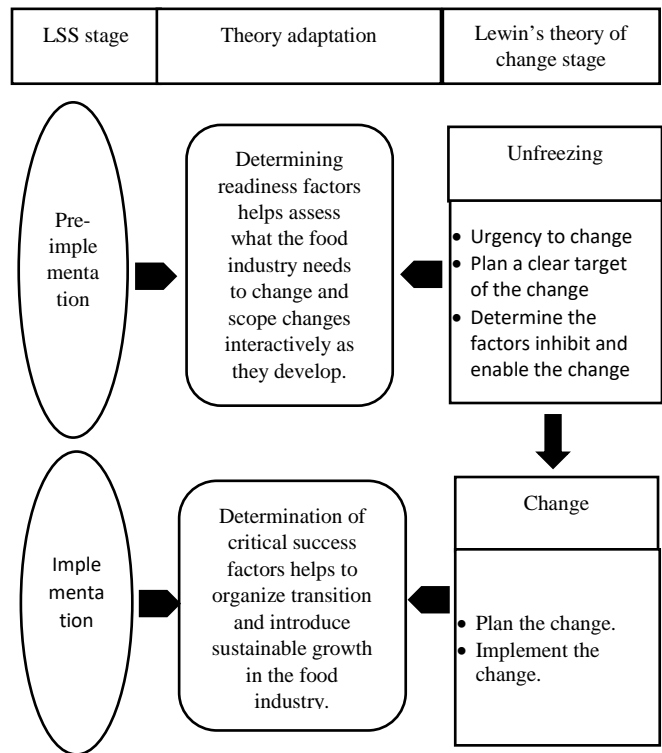


Figure 1: Research framework

3. Methodology

Based on the current body of knowledge, there is no research addressing the preparedness of LSS in the context of the food business. This study is primarily an inductive qualitative investigation based on semi-structured interviews comprising practitioners from twelve different food firms familiar with LSS. The inductive approach used in this study appreciates the insights and views of the interviewees as data [17]. This study is filling the gap of knowledge on LSS preparedness in the food business, necessitating a qualitative semi-structured interview rather than a quantitative technique commonly employed in readiness studies in other industries [3]. The research process was guided by the iterative qualitative data analysis model by [18], which comprises data collection, data reduction, data coding and analysis and conclusion, contribution, limitation and future research agenda.

3.1. Sampling

This study applies purposive sampling to choose a suitable sample using the criterion technique. Respondents involving food practitioners that are familiar with LSS and involved in CI projects [19] as depicted in Table 1. The size of a food business can have a causal relationship to LSS implementation effectiveness [6] and [20]. Hence, one of this study's sampling criteria is the experts working in a big company [6]. The interview guideline was created to not only to minimize bias but direct the interview sessions. Then, a pilot interview was carried with three respondents, which was shown by [21]. The interview is inductive in nature, maintaining an adaptive approach to give the chance for probing the questions based on the answers given.

Table 1: Details of the experts

Respondents	Current industry	Position	Experience in the food sector (year)	Role in LSS
E1	Cooking oil	Operational Excellence Manager	9	LSS Black Belt
E2	Dairy	Production Manager	14	SS Green Belt Lean Black Belt
E3	Dairy	Process Engineer	7	LSS Green Belt
E4	Fats	Assistant Manager	9	LSS Green Belt
E5	Beverages	Production Engineer	7	LSS Green Belt
E6	Butter	Continuous Improvement Manager	1	LSS Black Belt
E7	Dairy	Operation Manager	15	LSS Black Belt
E8	Dairy	Operation Manager	15	LSS Black Belts
E9	Beverages	Quality Assurance Executive	2	Certified Quality Engineer
E10	Confectionary	Value Stream Manager	13	LSS Green Belt
E11	Oils	Supply Chain Manager	9	LSS Green Belt
E12	Meat and poultry	Quality Assurance Manager	10 years	Six Sigma Green Belt

3.2. Data collection

There are two steps of data collection involved in this study. The first step is three pilot interview sessions were conducted as outlined by [21] in the literature, followed by the twelve semi-structured interviews using revised interview protocol. The interviews lasted ranges 60 to 180 minutes and were conducted face-to-face in the English language. Interviews were performed using a checklist, with notes and commentaries inserted throughout the session. With the participants' permission, the researcher recorded the interview sessions and comprehensive notes taken during the sessions. In grounded theory, the sample

size cannot be established in advance because it is dependent on the emerging theoretical categories [22]. The theory becomes evident when gathering new data no longer sparks new theoretical insights of the theoretical categories [8]. The twelve interviews with practitioners resulted in data saturation due to data overlap and no new insight materials being offered after the ninth interview, exhibiting the aspects of data saturation [21] and [22]. Despite reaching data saturation with the tenth respondent, the researcher conducted two additional interview sessions to confirm that no new information emerged, as shown in Figure 2.

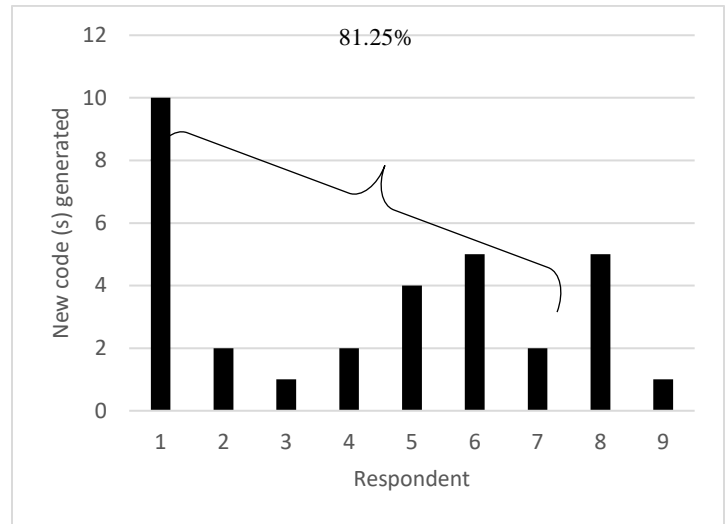


Figure 2: New LSS readiness codes generated at each interview

3.3. Data coding and analysis

Shortly after the interview, the interview sessions were familiarized and verbatim transcribed before the process of data reduction. The data was analyzed following the abductive approach, which combined interpretative content analysis and deductive themes identification [18]. The data was transcribed using software QSR Nvivo 12, and a field note was used to complement the transcription. The transcripts of the interviews were familiarized during data reduction, and the coding procedure was carried out to identify appropriate data pertinent that appropriate with the topic, as described by [18].

There are three fundamental coding modes in grounded theory: open, axial, and selective [22]. In open coding, the transcribed interviews are accurately conceptualized into appropriate readiness themes. The researchers were able to infer new meanings from the data collected using inductive coding, while the research objectives guided the emergent coding. During the axial coding phase, the content from the open coding themes was then assessed to develop linkages between the themes. The themes that resulted from the axial coding describe the emergent pattern in the LSS implementation and organizational readiness to change theory. Selective coding was conducted as the final step, where the purpose is to identify the core themes of the LSS readiness framework.

Comprehensive data analysis and explanation were performed to verify and clarify previously developed logical findings. The four components of [23] were applied to this investigation to guarantee the data presented was reliable. The causal map depicting the causal relation of readiness factors is developed based on inductive data analysis [24], [25] and [26]. The map depicts connected links by arrows, that showed the reviewers' inferred causality [25]. Descriptive analysis through Pareto analysis was conducted to identify the most important LSS CSFs.

4. Results

The determination of the key factors of the LSS pre-implementation stage will be conducted with consideration at both the 'unfreezing' and 'freezing' stages outlined by several readiness studies [27]. The findings revealed CSFs that could help coordinate LSS implementation in the context of the food sector during transitional periods [4].

4.1. Readiness factors in the food sector

Through the qualitative evaluation of employees in the food sector that has been practicing LSS, this study gathered a lot of data about the influence of LSS readiness criteria. When 29 readiness factors were identified through the interview process, the results of the interviews were described in terms of the highest number of mentioned readiness factors by those who took part in interviews.

The result shows that management support has topped all the identified LSS readiness factors. The respondents stated leadership and management are both have a critical role in LSS deployment. Top management is highly suggested to communicate their vision on LSS adoption and set realistic targets for the LSS program in the company. In the food sector, effective communication by the top management is one of the critical ingredients in encouraging the employees' involvement in the LSS related activities under their strict quality assurance programme.

The result shows that, although the regulations and industry standards outline the most quality programmes in the food sector, the top management has a critical role in developing the urgency towards the LSS adoption at the organizational level. According to respondent E7, senior management should keep a consistent cadence when encouraging the industry to adopt LSS. Top management support can be viewed by their overall knowledge on the "why" and "how" LSS implementation can contribute to their operational performance and company's bottom line and subsequently oversee towards business excellence.

The second critical readiness factor is financial support allocated for the program's implementation and its continuation in the company. Financial investment is the most important type of support from upper management, as it has a cascading effect on other readiness factors. For example, as reflected in Figure 3.0, each dimension may correlate with each other, and these findings may suggest there is a correlation between the financial resources, training, reward and recognition, and investment in essential

resources (trainers, machines, and software) under top management support theme. The respondents highly suggested the top management should be eager to spend in LSS activities including appreciation programs, soft and hard technologies, training and contracting advisors to facilitate LSS implementation.

Two emerging readiness factors in the context of the food manufacturing industry. Respondents E3, E5, E7, E8, and E9 pointed out the appointing a change agent to lead the LSS readiness phase is particularly critical in the food business. The purpose is to ensure the LSS initiation is guided and have a smooth transition between the pre-implementation and implementation stages. The assignment of the LSS champion as the change agent was implemented at the implementation phase reflected the initiation of the LSS journey of the business [28].

Another newly identified readiness factor in the study is top-level management awareness of the importance of LSS implementation. Top management should understand the demand for LSS implementation in their business and fully comprehend the reasons for embarking on the LSS. The top management was well-informed, primarily about the importance of a compulsory food safety programme in the food businesses depending on the type of food commodity. However, a brief LSS awareness session for the top management is suggested in the food sector to obtain the buy-in, commitment, and engagement in the LSS programme.

According to the vast majority of respondents, the ability to identify and maintain skillful human resources in LSS is essential for the successful implementation of LSS initiatives. Companies that initiate LSS with poor understanding and skills are deemed to fail to sustain the programme. The need for external experts is critical in educating theoretically and practically the team members because the training programme in class would be insufficient given the LSS project duration of six months, which would necessitate project consultation with the coach. According to E1 and E2, it is common for the food companies to have the same staff becoming the internal expert for both, food safety and continuous improvement activities.

According to E3 and E8, the food industry has been characterized as a complex industry, that could be a factor in the effectiveness of LSS. The industry's complexity included industrial norms and standards, raw material types, procedures involved, and critical-to-quality performance monitoring. As respondents advocate updating machinery and software to examine the capability of measuring variations and performances of critical-to-quality characteristics, common concerns in production (e.g. long waiting duration when devices break down) must be considered.

Respondent E2 advocated that the LSS should be implemented within a realistic, feasible, and achievable time frame, with awards and recognition included in the continuous improvement programme. As a result, respondents recommended that relevant guidelines be developed as a handbook to direct the organization's implementation of the LSS.

The involvement of qualified personnel in the quality management project is a significant indicator of preparedness at the LSS level. Identifying and training the right applicant would result in the addition of extrinsic value to the organization. However, several respondents emphasized that senior management has an ultimate say to choose the change agent/champion and let the decision of the project team's appropriate members to the champion. Typically, team members appointment happened at the implementation stage of the LSS. However, the respondents suggested that the existence of nurturing the top talent in the company should already have existed.

"It is critical to choose the suitable employees for the LSS belt training to see the return of investment of the training" (E6).

In the LSS implementation, having top talent is believed to have a better result, attracting more talent where it motivates the other employees to get involved [28].

Respondents E6, E7, E9, and E11 asserted that increasing LSS awareness would improve employees' willingness to adapt in favour of LSS implementation. It is agreed by the respondents that the employees will embrace better on the program when their understanding of the principal and advantage of LSS towards the business and themselves. LSS should be introduced as early as in tertiary education, where E5 stated that LSS can be embedded under related subjects in food bachelor degree program such as Operation Management, Quality Management and Supply Chain Management.

Additionally, several respondents indicated that specific tactics should focus on different projects. Implementing LSS should not be viewed as the answer for all production process issues. The importance of selecting the correct process and project contribute to the subsequent LSS adoption direction in the company. According to [3], project prioritization should be done in considering the best products in the business, contribution of the project to bottom-line saving and profit, contribution of the project to growth and impact of the project on the sustainability enhancement.

E9 warned that lack of LSS awareness would risk the success of the LSS adoption. In the food sector, interest in LSS can be nurtured through the other quality-related activities in the businesses, for instance, quality assurance and food safety activities. Another significant issue identified is the process for forming teams for CI activities, as they are considered as the main personnel accountable for quality improvement projects. Typically, food companies commonly will have a food safety team that adhered to both food safety and quality improvement issues.

Process management's fundamental notion is to understand quality criteria from the eyes of the customer. The guidelines and suggestions from LSS projects should be aligned with customer demands and the characteristics of each commodity in the food sector. E6 indicated that government assistance, such as

establishing LSS awareness campaigns for the business, can be crucial in its readiness. The government can assist by engaging industry players through experience-sharing workshops and learning visits to LSS projects that have been successful as benchmarks and case studies.

4.2. LSS CSFs in the food sector

The findings indicated that 31 identified factor nodes contributed to the success of LSS implementation. The number of the codes was analyzed using Pareto analysis to determine their frequency and criticality under 80/20 rules. In [1], the author suggested that Pareto analysis helps classify "vital few" success factors from "trivial many," which focus on the factors to accentuate the desired goals.

Figure 3.0 shows that 17 CSFs were identified as the prioritized factors for the success of LSS implementation through 80/20 rules. The food sector context identified training programmes as the most critical factors, followed by top management support, financial resources, communication structure employee involvement and as the top five main success factors for LSS implementation. Some of the CSFs are correlated and linked from the "unfreezing" stage to the "freezing stage" of LSS implementation. All of the LSS readiness factors in Figure 3.0, such as external factors (EF), employees involvement (EI), organizational culture (OC), management support and leadership (MSL) and project management (PM) are the common factor identified at the readiness and implementation phase of LSS.

Based on the prioritized CSFs, a causal mapping in Figure 4 developed based on the interview sessions showed that the potential of some of the factors have an impact on the emerging of other factors as the critical factor for the LSS program to be successful.

The respondents mentioned the training program has the most causal impact on successful LSS adoption, especially in this industry. Three distinct factors arose from the training: coaching and mentoring, LSS awareness and understanding level of LSS.

"Our LSS training programmes will include coaching and mentoring session as our company has hired in-house trainers to train the Green Belt/Black Belt programme for our employees."(E1)

"LSS basic training is crucial to create awareness among employees and their critical roles in LSS implementation." (E5)

"I would say continuous training for the employees is critical to increasing their understanding of the function and benefits of each LSS tool." (E12)

One of the most critical factors following the training programme was the support and commitment of senior management on LSS projects. Respondent E12 mentioned that the requirement of LSS project approval highly depends on the top management vision and buy-in of the project.

"If there is new management, there must be massive support from the top to ensure they acknowledged and understand the benefits of current initiatives. When management is willing to invest in LSS projects, I would say they are supportive." (E7)

"I think that management support is influenced by their awareness on the advantages of LSS projects." (E10)

Another important factor to consider is the availability of financial resources and the budget allocation communicated by the top management to implement LSS initiatives. Another emerging factor that is relatable to the allocation of funding is the rewards and recognition. Respondents stated that their company allocates a budget for rewarding the employees and investment in cost-saving projects, which serve as a driving factor in making it a successful project. Another respondent responded on this issue:

"Having a strong financial capability to invest in systems, training, the hiring of trainers, and software is critical for any organization, in my opinion." (E2)

"LSS projects necessitate investment on the part of the management, particularly in terms of staff training and programme maintenance." (E6)

Another major component that contributed to the success of the LSS initiative in the food business was the participation of employees. In response to respondent R1, he stated that his company engages their employees by empowering them through hands-on training to provide the best exposure to them in the workplace. E7 stated that employee involvement led to awareness and an increased understanding of the LSS concept and benefits among those who worked for the company. Respondent E11 revealed that his company allows shop-floor employees to participate in decision-making, motivating and building trust among employees and management.

Medium of communication was identified as one of the most important CSFs that can be implemented to achieve a good result in the food business for successful LSS implementation. An effective communication program is essential, especially in convincing top management. He stated that top management roles in communicating the LSS strategic thinking to all departments are critical. Besides, it was to ensure the awareness of employees on LSS initiatives is under the same vision of the management. Other respondents stated that:

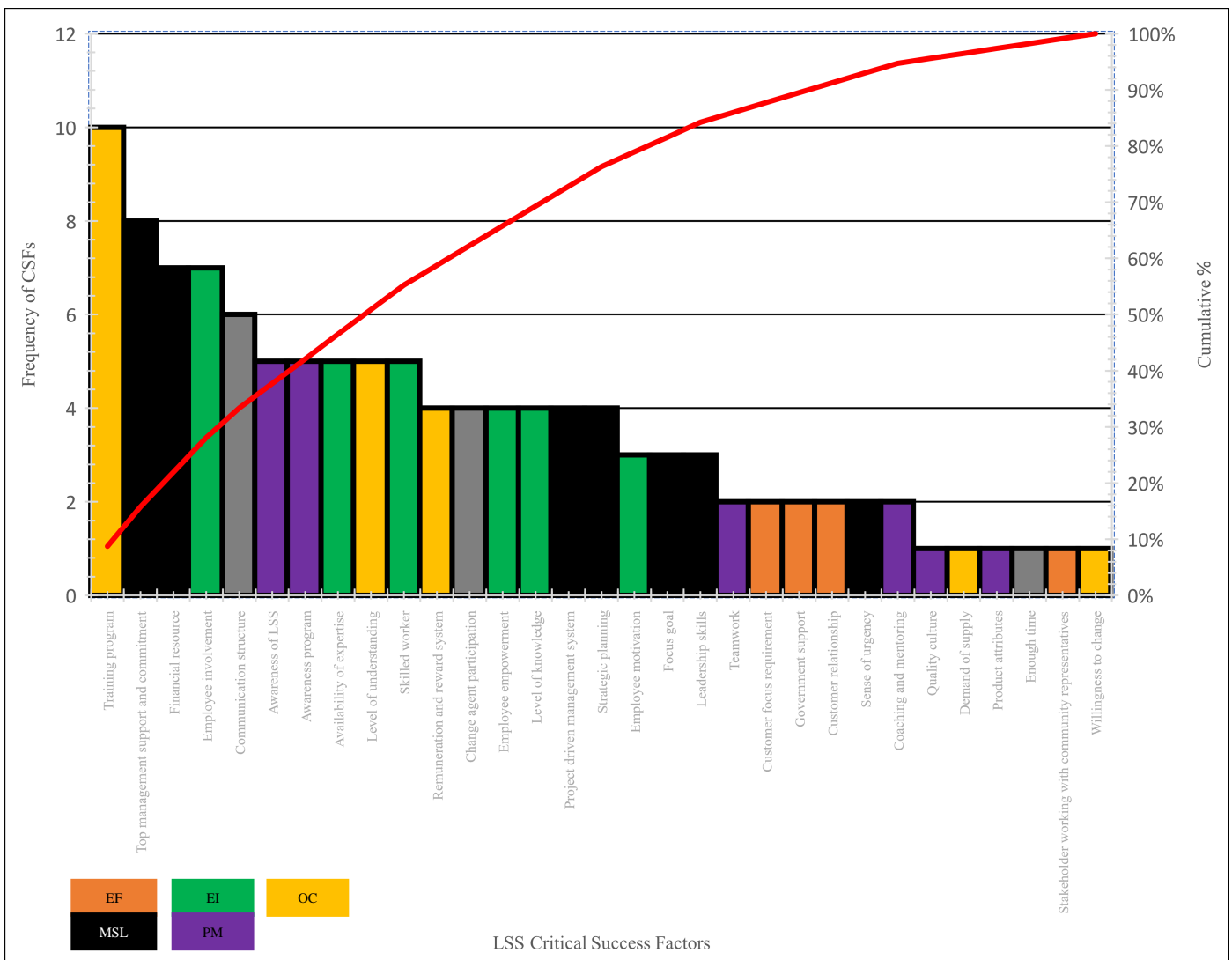


Figure 3: Pareto chart of CSFs of LSS implementation in the food sector

"The acquisition is having a simple briefing to the shop floor to deliver the standard procedure of the LSS projects and ensuring that they are aware of how these projects will impact them is critical" (E8).

The Pareto chart in Figure 3 also revealed that the LSS enablers for the food sector mainly comprised of soft factors. The change process has an impact on any form of factors that could lead to the LSS implementation success in the food sector, including the "support from the government" and customer focus business. Some of the quotes being mentioned by the respondents:

"The government should provide help and support by rewarding those who complete successful LSS projects in a competition or by providing financial assistance to those who attend LSS training. Sources of coaches from other successful industries should be able to be an eye-opener" (E10)

"The LSS concept is driven by the ability to accrue the trust of the customer through value creation and variation reduction. The project should be tailored to the specific customer to ensure that it meets all of the customer requirements." (E1, E5).

5. Discussions

5.1. Readiness factors and their causal relationship

It was claimed that there is a necessity to explore critical and readiness factors for each sector because the nature and operations of various businesses may differ in terms of their factor, applicability, and value to the business [29]. The food sector's distinguishing traits, including the obligation to comply with strict

food legislation rules, necessary cleaning of machinery and production sites, and the rapid changeover of various products, can impede the success and readiness of adopting LSS [2] and [30].

Venturing into LSS implementation necessitates high budget and continuous efforts, mandating strategic planning focused on the essential criteria during the pre-implementation stages to avoid overspending or underspending on CI operations [6, 11].

The findings revealed that management support and leadership, organizational culture readiness, employee involvement, process management, project management, and external interactions all impacted the preparedness of the LSS program. Figure 5 depicts the identified preparedness criteria to demonstrate the relationship between them.

5.1.1. Management support and leadership (MSL)

Management support and leadership was identified as the most critical factor at the pre-implementation phase. If any attempts at continuous improvement are to be made in a long time and with unwavering funding, this indicates a need for management support [11]. The organization's top management is capable of incorporating LSS as a business strategy for improving performance and business excellence. Implementing LSS demands significant capital expenditures and high commitments, demanding strategic planning to avoid cutting CI operations budget [6] and [16]. Management support implies that top management is taking action and developing plans to facilitate and steer the business's implementation of the LSS [1].

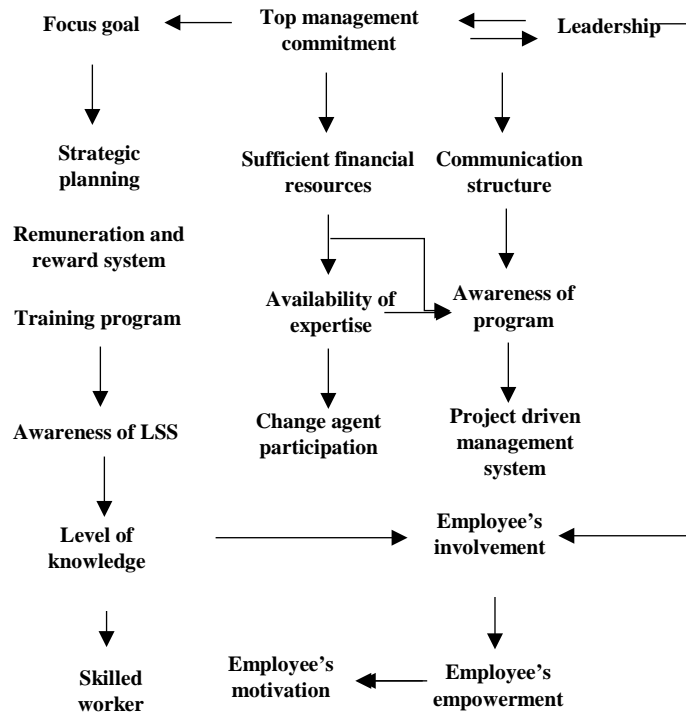


Figure 4: Causal diagram of the prioritized LSS CSFs in the food sector

5.1.2. Organizational culture readiness

Following the interview, it was determined that in this industry, an organization's preparedness to adopt LSS relies on its organizational culture's willingness to adapt. According to a previous study, quality-driven culture should be promoted among employees in the organization. The foundation of success can be achieved with the corporate culture change [31] and [32] and [33]. Regular training enables the company to produce talented personnel with sufficient tools and techniques to be used by staff [32].

5.1.3. Employee involvement

Employee engagement is referring to the level of understanding of their roles within the company indicated the readiness of LSS activities. Employees who have adequate skills to apply the tools of LSS and an awareness of the principle of LSS are thought to be less reluctant to change. Operating as a team has been demonstrated to decrease the unwillingness towards the change, which validates the view articulated in the literature that management may appoint members and educate the team to tackle issues in process improvement [34]. By assigning employees to the problem-solving activity, the organization can help them feel more confident in their abilities. This would foster a sense of trust between top management and the employees.

5.1.4. Project management

The result shows that project management is a prominent LSS readiness factor as LSS is a project-based program. When

administering the LSS project, prioritization should be made to deliver the most financial benefits to the firm [27], [32] and [35]. LSS project management entails determining the companywide level of understanding about LSS, which can inform the management on the company preparedness for LSS implementation [32]. Participation in the LSS awareness activities will increase the familiarity of the employees and minimize the unwillingness towards the change in the program's direction.

5.1.5. Process management

Considering the nature of the business, food businesses should consider integrating the LSS with customer relation activities. This will enable the voice of consumers will always be considered in the process and critical to quality attributes. The integrated components should be made as to the business's major priority for the future. Consumer requirements must be considered for the LSS projects to consider and increase the "consumer voice" while taking into account the particular features of the food sector. Gaining operational excellence is stated to be critical in the survival of a business [3]. Efficiency can be improved by identifying root causes, developing improvement solutions, and preventing them from occurring again [27], [34], [35], [36] and [37]. Aside from that, when a reliable system for measurement is accessible in the process, organizations are better prepared for LSS.

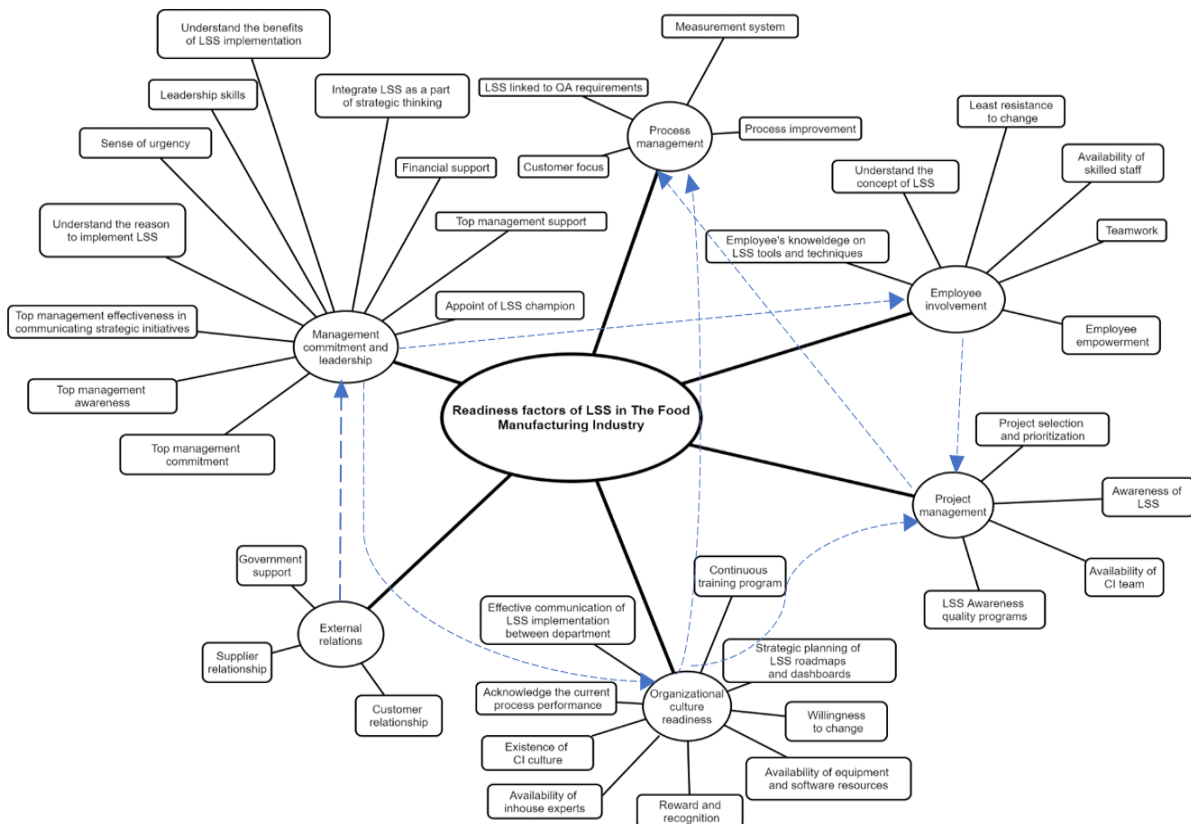


Figure 5: The conceptual LSS readiness factors in the food sector and its constructs.

5.1.6. External factors

Most of the respondents discussed the importance of government support in initiating LSS in the food businesses. In the previous studies on LSS readiness, this factor is not commonly found. However, this component was identified as affecting the leanest implementation in the red meat business. The authorities play the policymaker's role and are responsible for strategizing effective adoption strategies [37]. However, cooperation between the external factors is defined as their events with their suppliers and customers [38]. Upon examination of the factors, it was discovered that management support has become the most vital determinant of LSS readiness. The management support entails the highest number of tasks in ensuring the preparedness of a food business to uptake LSS.

6. Conclusions

The implementation of a successful LSS can result in significant financial advantages for a company. However, most companies fail to implement LSS because they are pressured into doing so as part of a fad of business excellence approaches without taking the time to assess whether the organization is prepared to execute it successfully. The key research objective is discovering relevant factors that need to be assessed before investing in the LSS adoption customized for the food business, as well as the CSFs that must be considered throughout this process, as it is the linking phase following the readiness phase as according to Lewin's change theory.

According to the findings of this study, the emerging LSS readiness factors determined are management support and leadership, the readiness of organizational culture, employee involvement, process and project management, and finally external factors. The results show there is a proposition that the discovered factors for LSS implementation in this sector are causally related. The findings show 17 CSFs, where some of these factors are common factors under the readiness stage. In the food business, the most prevalent CSFs of the LSS are training, top management commitment and leadership, financial capacity, employee engagement, and communication.

There are two ways to look at the significance of this study. First, the determined LSS readiness factors provided a simple and easy approach as a decision-making tool to assist food sector practitioners in understanding their current practices, evaluating if their organizational environment supports LSS and finally deciding whether to invest in LSS Implementation. The managers also will be informed on the factors that are important in the pre-implementation stage. Since some of the readiness themes were found will be an enabling factor in the implementation phase, the managers will be able to plan the implementation strategically. Whilst, for the food business that has implemented LSS, the result of this study opens the window to identify factors that require more attention to reach the targeted level of LSS readiness. Second, this paper extends to the limited body of knowledge about LSS implementation in the food sector and organizational

readiness to change theory. Additionally, the findings of this study can be evaluated by comparing to those of similar studies conducted in various industries or countries, determining the impact of LSS readiness in different countries. Additionally, the findings of this study can be evaluated by comparing to the other similar studies conducted in various industries or countries, evaluating the impact of LSS readiness in different countries.

There were several limitations concerning the generalizability of this work, which provide the opportunity for future research. This study produced results that are described by the features of the interpretive type of study. While analyzing the data sets, it was found that hypothesis development must be built on a solid foundation, allowing for an empirical study to be conducted. Because the study used a small and focused sample, it was not possible to determine whether the findings had any external validity. Future research should consider that the result from grounded theory can be developed to hypothesis to be tested through quantitative study to establish a generalized LSS readiness factor with a larger sample size for the food sector. Future studies can also focus on the size of the company in the food production business such as big enterprise, medium or small. The identified factors can be used as the readiness measurement instrument in other industrial settings. Finally, the result of this study can be used as a vital theory for the construction of an LSS readiness diagnostic tool for the food sector.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgement

This work is supported by the Universiti Putra Malaysia under research grant Putra Young Initiative (IPM/2017/9552400) and Malaysia Ministry of Education (MOE) for providing research grant under Fundamental Research Grant Scheme (FRGS/1/2018/TK03/UPM/02/7).

References

- [1] S.A.H. Lim, J. Antony, "Statistical process control readiness in the food sector: Development of a self- assessment tool," *Trends in Food Science and Technology*, **58**, 133-139, 2016, doi:10.1016/j.tifs.2016.10.025.
- [2] M. Dora, X. Gellynck. "Lean six sigma implementations in a food processing SME: A case study," *Quality and Reliability Engineering International*, **31**, 1151–1159, 2015, doi:10.1002/qre.1852
- [3] V. R. Sreedharan, R. Raju, M. V. Sunder, J. Antony, "Assessment of lean six sigma readiness (LESIRE) for manufacturing industries using fuzzy logic," *International Journal of Quality and Reliability Management*, **36**(2), 137–161, 2019, doi:1108/IJQRM-09-2017-0181.
- [4] L. B. M. Costa, M. G. Filho, L. D. Fredendall, F. J. Gómez Paredes, "Lean, six sigma and lean six sigma in the food sector: A systematic literature review," *Trends in Food Science and Technology*, **82**, 122– 133, 2018, doi:10.1016/j.tifs.2018.10.002.
- [5] L. Marrucci, M. Marchi, T. Daddi, "Improving the carbon footprint of food and packaging waste management in a supermarket of the Italian retail sector," *Waste Management*, **105**, 594-603, 2020, doi:10.1016/j.wasman.2020.03.002.
- [6] L.B.M. Costa, M. G. Filho, L. D. Fredendall , G.M.D Ganga, "The effect of lean six sigma practices on food sector performance: Implications of the Sector's experience and typical characteristics," *Food Control*, **112**, 2020, doi:10.1016/j.foodcont.2020.107110
- [7] J. A. Garza-reyes, I. E. Betsis, V. Kumar, M.A.R. Al-Shboul, "Lean readiness-the case of European pharmaceutical manufacturing industry,"

- International Journal of Productivity and Performance Management, **67**(1), 20-44, 2018, doi: 10.1108/IJPPm-04-2016-0083
- [8] B.V. Chowdary, D. George, "Improvement of manufacturing operations at a pharmaceutical company: a lean manufacturing approach," *Journal of Manufacturing Technology Management*, **23** (1), 56-75, 2012, doi: 10.1108/17410381211196285.
- [9] A. A. Armenakis, S. G. Harris, K. Mossholder, "Creating readiness for organizational change," *Human Relations*, **46**, 681–703, 1993, doi:10.1177/001872679304600601.
- [10] Z. Radnor, 2011. "Implementing lean in health care: making the link between the approach, readiness and sustainability," *International Journal of Industrial Engineering and Management*, **2**(1), 1-12, 2011, doi: 10.1016/j.socscimed.2011.01.011.
- [11] J. Antony, "Readiness factors for the lean six sigma journey in the higher education sector," *International Journal of Productivity and Performance Management*, **63**(2), 257–264, 2014, doi:10.1108/IJPPM-04-2013-0077
- [12] C. Andrea, and M. Kumar. "Lean Six Sigma and Industry 4.0 integration for Operational Excellence: evidence from Italian manufacturing companies." *Production Planning & Control*, **32**(13), 1084-1101, 2020, doi:10.1080/09537287.2020.1784485
- [13] E.H. Jacobson, "The effect of changing industrial methods and automation on personnel," Paper presented at the Symposium on Preventive and Social Psychology, Washington, DC, 1957.
- [14] S. Al-Balushi, P.J. Sohal, P.J. Singh, A. Al Hajri, Y.M. Al Farsi, R, Al Abri, "Readiness factors for lean implementation in healthcare settings – a literature review," *Journal of Health Organization and Management*, **28**(2), 135-153, 2014, doi: 10.1108/JHOM-04-2013-0083.
- [15] B.J. Weiner, "A theory of organizational readiness to change," *Implementation Science*, **4**(1), 67, 2009, doi: \10.1186/1748-5908-4-67.
- [16] E. Drohomerski, S. E. G. da Costa, E. P. de Lima, P. A. Garbuio, "Lean, six sigma and lean six sigma: An analysis based on operations strategy," *International Journal of Production Research*, **52**(3), 804–824, 2014, doi:10.1080/00207543.2013.842015.
- [17] C. Robson, K. McCartan, *Real world research* (4th ed.), John Wiley & Sons, 2006.
- [18] M. B. Miles, A. M. Huberman, *Qualitative data analysis: An expanded sourcebook.*, Sage Publications, 1994
- [19] M. Q. Patton, *Qualitative research and evaluation methods*, Thousand Oaks, Sage, 2002.
- [20] MG Filho, GMD. Ganga, A. Gunasekaran, "Lean manufacturing in Brazilian small and medium enterprises: Implementation and effect on performance," *International Journal of Production Research*, **7543**, 1–23, 2016, doi: 10.1080/00207543.2016.1201606.
- [21] M. Dikko, "Establishing construct validity and reliability: Pilot testing of a qualitative interview for research in takaful (islamic insurance)," *The Qualitative Report*, **21**, (3), 521- 528, 2016, doi: 10.46743/2160-3715/2016.2243.
- [22] R. Bogdan, S.K. Biklen, *Qualitative research for education: An introduction to theories and methods*, Pearson A & B, 2007.
- [23] E. Guba, "Criteria for assessing the trustworthiness of naturalistic inquiries," *Educational Technology Research and Development*, **29**(2), 75-91, 1981, doi:.
- [24] M.A. Abernethy, M. Horne, A.M. Lillis, M.A. Mallina, F.H. Selto, "A multi-method approach to building causal performance maps from expert knowledge," *Management Accounting Research*, **16**(2), 135-155, 2005, doi: 10.1016/j.mar.2005.03.003.
- [25] I. Pyrko, V. Dörfler, "Using causal mapping in the analysis of semi-structured interviews," In *Academy of Management Proceedings* , 2018(1), 2018, doi: 10.5465/ambpp.2018.14348abstract.
- [26] A. J. Scavarda, T. Bouzdine-Chameeva, S. M. Goldstein, J.M. Hays, A. V. Hill, "A methodology for constructing collective causal maps," *Decision Sciences*, **37**(2), 263-283, 2006, doi: 10.1111/j.1540-5915.2006.00124.x.
- [27] K. Jayaraman, T. Leam Kee, K. Lin Soh, "The perceptions and perspectives of lean six sigma (LSS) practitioners," *The TQM Journal*, **24**(5), 433–446, 2012, doi: 10.1108/17542731211261584.
- [28] V. F., Vallejo, J. Antony, Douglas, J. A., Alexander, P., & Sony, M. (2020). Development of a roadmap for Lean Six Sigma implementation and sustainability in a Scottish packing company. *The TQM Journal*.
- [29] J. Rockart, "Chief executives define their own data needs," *Harvard Business Review*, **57**(2), 1979.
- [30] A. Panwar, B.P. Nepal, R. Jain, A. P. S. Rathore, "On the adoption of lean manufacturing principles in process industries," *Production Planning & Control*, **26**(7), 564–587, 2015, doi: 10.1180/09537287.2014.936532.
- [31] A. Gurumurthy, P. Mazumdar, S. Muthusubramanian, "Graph theoretic approach for analyzing the readiness of an organization for adapting lean thinking: a case study," *International Journal of Organizational Analysis*, **21**(3), 96-427, 2013, doi:10.1108/IJOA-04-2013-0652.
- [32] Y. Lagrosen, R. Chebl, M. Rios Tuesta, "Organizational learning and six sigma deployment readiness evaluation: A case study," *International Journal of Lean Six Sigma*, **2**(1), 23–40, 2011, doi: 10.1108/2040146111119431.
- [33] SC. Cross, "Lean cuisine." *Industrial Engineer*, **41**(1), 46-47, 2009
- [34] P.B. Keliji, B.S.D Abadi, M. Abedini, "Investigating readiness in the Iranian steel industry through six sigma combined with fuzzy delphi and fuzzy DANP," *Decision Science Letters*, **7**, 465–480, 2018, doi: 10.5267/j.dsl.2018.1.001.
- [35] M. N. Mishra, "Identify critical success factors to implement integrated green and lean six sigma," *International Journal of Lean Six Sigma*, 2018, doi:10.1108/IJLSS-07-2017-0076.
- [36] T. Y. Lee, W. K. Wong, K. W. Yeung, "Developing a readiness self-assessment model (RSM) for six sigma for China enterprises," *International Journal of Quality and Reliability Management*, **28**(2), 169–194, 2011, doi: 10.1108/02656711111101746.
- [37] A. Cox, D. Chicksand, M. Palmer, "Stairways to heaven or treadmills to oblivion?: Creating sustainable strategies in red meat supply chains," *British Food Journal*, **109**, 2007, doi:10.1108/00070700710780689
- [38] C. A. Moya, D. Galvez, L. Muller, M. Camargo, "A new framework to support lean six sigma deployment in SMEs," *International Journal of Lean Six Sigma*, **10**(1), 58–80, 2019, doi:10.1108/IJLSS-01-2018-0001.

Service Lifetime Loss Evaluation Method for Transformers Exclusively Serving Solar Power Plants

Bonginkosi Allen Thango*, Jacobus Andries Jordaan, Agha Francis Nnachi

Department of Electrical Engineering, Tshwane University of Technology, Emalahleni, 1034, South Africa

ARTICLE INFO

Article history:

Received: 29 May, 2021

Accepted: 07 September, 2021

Online: 13 November, 2021

Keywords:

Renewable Energy

Transformers

Service lifetime

Total Ownership Cost

ABSTRACT

In the last decade, South Africa has attracted and attained more investment by means of the Renewable Energy Independent Power Producer Procurement Programme (REI4P), which is, a structured invitation to Independent Power Producers (IPPs) to submit a bid to generate and supply power to the electrical grid. In spite of REI4P's undeniable triumph, a much-discussed drawback has been the related service lifetime cost of equipment that facilitate the renewable energy technologies. The description of the service lifetime loss method (SLLM) gets more complex in the new dawn of decarbonized electricity market. The SLLM must be modified for determining the total ownership cost of transformers facilitating Sustainable Energy Systems (SES's) in the decentralized energy market. The main focus of this work is to indicatively formulate a fundamental advancement upon the conventional service lifetime losses evaluation formula to contemplate the service lifetime loss evaluation method for transformers exclusively of service to solar power plants in South Africa. The distinct operational features of a solar plant have been embedded in the formulated service lifetime loss evaluation formula by way of the plants' Generation (GM) and Non-Generation Mode (NGM). Further, a levelized cost of energy supplied per unit of time by the solar plant is employed to determine the energy cost of the no-load and load losses that will be consumed by the studied transformers during their service life. Ultimately, the premier findings of this study indicate that the annual solar potential has an effect on the transformer service lifetime loss factors and the conventional method is not suitable thereof for this application. This is a characteristic that should be precisely considered, as it may influence the tender adjudication process to purchase a transformer based on the total ownership cost offers of various transformer manufacturers.

1. Introduction

The rapidly escalating cost of energy from coal power generation in South Africa has aroused a substantial demand for feasible recourse energy sources. In spite of the fact that at present it contributes to only a small proportion of South Africa's power generation, deployment of solar power plants (SPPs) is rapidly increasing for both utility-scale and distribute generation power generation systems (DPGS). Reduction of the energy tariffs are impelled by the technological advances, proportionate saving in costs gained by an increased level of production in manufacturing, and novelties in finance has introduced solar power at arm's length of grid parity in an increasing renewable energy markets. Ongoing advances and additional cost savings will broaden these window of opportunities especially in South Africa where opportune solar conditions prevail. The energy tariffs for renewable energy

technologies has been incessantly plummeting since the inception of the Integrated Resource Plan (IRP) [1] and [2]. Figure 1 shows the annualized cost of energy (ACOE) for wind and solar [3]. Wind and solar photovoltaic (PV) are observed to cost about 0.91 ZAR/kW and 0.755 ZAR/kW while coal-based energy cost about 1.03 ZAR/kW [3].

Meanwhile the rapid increase of the South African energy mix has been attended by an observed uneven proficiency revealed the new dawn of renewable energy market. For instance, the energy mix in South Africa by the vertically-integrated utility and Independent Power Producers (IPP's) exist together at the same time but have distinctive methodologies of evaluating their capital expenses, network expenditures and load curves. Therefore, the processes for the transformer loss capitalization must be distinctive. The methods available in the literature straggle the proficiency to appropriately unbundle the loading and energy components of the cost transformer losses in the concerned

*Corresponding Author: Bonginkosi Thango, thangotech@gmail.com

entities. This will significantly guarantee that each loss component is apportioned to suitable stakeholders in decentralized energy markets with regards to who is accountable to insure the losses of the transformer.

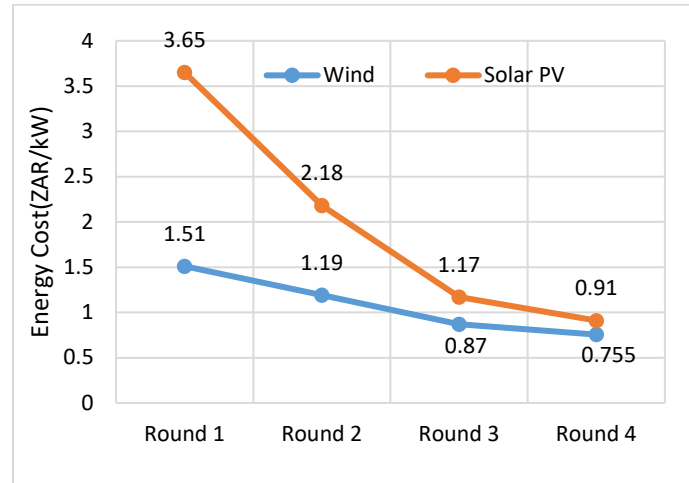


Figure 1: Energy tariffs for Wind and Solar in South Africa

Transformers are often anticipated to be in service between 25-30 years [4]. The jurisdiction procedure by utility owners for purchasing a transformer should therefore not only be based on the initial purchase cost but also the operating and maintenance costs. The total ownership cost (TOC) method [5] take into consideration the purchase cost, operational cost and maintenance costs over the transformer designed service lifetime. It has become possible for utility owners to compute the most technical and economic choice between competing transformer designs over its designed service lifetime through the TOC method. The TOC method also incorporates other cost related parameters such as annualized cost of electricity (ACOE), discount rates, tax rates et cetera. Considering that the TOC method comprises of the cost of service losses, in which will materialize in the future, there is a need to discount these losses to equalize them to current day South African rand (ZAR) value. The capitalization values cost of no-load loss (CNLL) and cost of load losses (CLL) in ZAR/kW studied by authors in [6 - 10] derive the cost of service losses. Transformer manufacturers submit a technical and commercial offer in a competitive bid and if the capitalization values are provided, then the manufactures should be grounded on these values in all cases. The service no-load losses are reliant on the core steel grade and are unvarying for a transformer during service [11] and [12]. The service load losses on the other side are dependent on the loading profile of the transformer [13] and [14]. Significant service losses of a transformer during service can potentially culminate in excessive heating and consequently degradation of insulation of materials due to overloading and presence of harmonics especially in renewable energy technologies suchlike wind and solar [15] - [20]. As a result, a low-loss transformer would practically be in service for a longer duration with minimum potential for premature failure. Additionally, a low-loss transformer discounts the amount of electrical energy necessitated to supply the service losses.

Firstly, this work presents classical SLLM for transformers that are intended to be in service for 8760 hours annually and from a single utility energy source. This method is based on transformer

loss data supplied by the manufacturers and the utility characteristics.

Nevertheless, the understanding of the SLLM becomes knotty within the bounds of transformers intended to be of service to renewable energy application. A SLLM for solar power plants is then proposed in this work. This method corrected the classical method intended for transformers intended to be in service for 8760 hours annually and from a single utility energy source. In the case of solar power plants, during the non-generating period, the solar plant will be feeding-off the electric grid with coal power generation

2. Transformer cost analysis

In order to bestow the surged energy demand, the excessive loading of transformers during service and related remaining service lifetime on account of the hotspot temperature rise is structural components must be cautiously estimated in opposition to the replacement of transformers or supplementing existing unit with new transformer. IPP's generally ascertain definitely the size of suitable units for service based on current loading cycle, forecasted load growth and other sound engineering acumen. Additionally, IPP's determine the economic effect of service losses of the transformer over its planned service lifetime and the initial purchase price prior to procuring new transformers. The classical Total Ownership Cost (TOC) procedure provides a comprehensive analysis for the economic planning aspects of investing in purchasing and operating a transformer

2.1. Service lifetime loss method: Classical Method

The purchasing of transformers by SLLM is predominantly employed based on the premise of a lowest TOC amongst competing offers of various transformer manufacturers. This method provide details on the cost of transformer service losses and the initial purchasing price of the unit. The service lifetime of the unit is presumed to be even irrespective of the MVA rating. Various transformer loading guides suggest a normal operating service lifetime of 30 years [4]. The aspect of investing in purchasing a transformer is determined to reflect an annual cost of the service no-load and load losses inclusive of the demand rate. The total revenue needed is treated by the ratio of the annual cost of these losses to the fixed charge rate (FCR). The FCR illustrates all the incurred charges in the utility independent of energy (kWh) traded, suchlike depreciation, tax, insurance and service maintenance.

For each unit of kilowatt (kW) of the service no-load and load losses, there is an inclusion of the demand cost (DC) portion derived from capital cost of the transformer. Consequently, there is an inclusion of the component for the energy cost (EC) in kilowatt-hour (kWh) by the utility. The cost of service losses (CSL) are hence computed as follows in (1).

$$CSL(ZAR/kW) = DC + EC \quad (1)$$

here:

$$DC = \text{System Investment}(ZAR/kW)$$

$$EC = \frac{8760 \times ACOE}{FCR}$$

System investment (SI) illustrates the acquisition in the bulk movement of electrical energy from a generating site to distribution facilities essential to deliver supplementary energy demand arising from transformer service losses at the system peak. There are essentially two approaches to assessing the SI rate in ZAR/kW. The first one is classical and is based on the established costs of constructing a generating site and related distribution facilities. The second approach is based on a scenario where the power utility does not self-generate power and instead purchase power. The SI rate can be ascertained by ratio of the DC (ZAR/kW-yr.) and the FCR.

Consequently, the ACOE is evaluated by forecasting and weighting the annual inflated cost of energy by means of capital recovery factor (CRF). Analytically, the latter can be expressed as shown in (2).

$$ACOE = \text{Current Energy Cost} \times CRF \quad (2)$$

here,

$$CRF = \left(i \times \left[\frac{(1+i)^n}{(1+i)^n - 1} \right] \right) \left(X \times \left[\frac{1-X^n}{1-X} \right] \right)$$

where,

i – Discount rate

n – Designed transformer service lifetime

e – Energy escalation rate

$$X = \frac{1+e}{1+i}$$

It follows that the cost of losses can be expressed as follows in (3). This equation is handily feasible for power utilities with their own electricity generation.

$$CSL(ZAR/kW) = \frac{DC(ZAR/kW-Year) + 8760 \times ACOE}{CRF} \quad (3)$$

Additionally, (3) segregate apiece class of service losses in ZAR/kW as shown in (4), (5) and (6). The cost of no-load losses (CNLL) are expressed as shown in (3).

$$CNLL(ZAR/kW) = CGTD + \frac{8760 \times ACOE}{CRF} \quad (4)$$

Here, $CGTD$ is the utility costs for the generation, transmission and generation. The cost of load losses (CLL) are expressed as shown in (4).

$$CLL = CGTD \times PRF^2 \times EPR^2 + \frac{8760 \times ACOE \times TLF \times EPR^2}{CRF} \quad (5)$$

Here, PRF and EPR is the peak responsibility factor and the equivalent peak ratio. The PRF is the ratio of the load's demand to peak power demand at the course of the system peak as expressed in (7).

$$PRF = \frac{\text{Transformer loading during system peak}}{\text{Transformer peak load}} \quad (7)$$

The PRF can be implemented to unique customer classes. Needless to say that a loading with a PRF of one peaks concurrently with the overall system. The electrical loading of

particular customer classes have a propensity to fluctuate at a similar fashion. For instance in South Africa, commercial electrical load are at peak from about 08:00AM in the morning to about 18:00PM in the evenings while residential electrical loads peak from around 18:00PM in the evenings. Inevitably, the weather has a direct impact on the loading extent. In the winter for instance, electrical heaters increase the demand and lower the variety among loads. At the distribution transformer level, the loading factors are largely ranging from about 0.4 to 0.9 with 0.75 being common [6 - 8] and [10]. The standard values of PRF can vary from 1.0 to about 0.35. Given that PRF is a ratio based on the loading (kVA), the services losses are a function of PRF^2 based on transformer manufacturers empirical survey. The most common values of the PRF for various transformers is tabulated in Table 1.

Table 1: PRF values for various transformer classes

Power System Level	PRF
Generator Step-Up	1
Transmission Substation	0.9
Distribution Substation	0.8
Distribution	0.35

The interrelation between the service losses and the rated transformer loading is described by the peak ratio (PR). The latter can be expressed analytically as follows in (8).

$$PR = \frac{\text{Annualized peak transformer loading}}{\text{Maximum rated transformer load}} \quad (8)$$

Supposing the load increase by a specific percentage annually, then from (5), the EPR may be employed as expressed as follows in (9).

$$EPR = PR \times \sqrt{\frac{(1+g)^{2t} - 1}{\ln(1+g)^{2t}}} \quad (9)$$

where,

g – Rate of annual load growth

t – Time in years

Lastly from (5), the annualized TLF can be described by the ratio of the average service load loss to the maximum load losses. The latter can be expressed as follows in (10).

$$TLF = \frac{\text{Average transformer load loss}}{\text{Maximum transformer service load loss}} \quad (10)$$

If the power utility provides the annual loading cycle data, then the services load losses can be evaluated from this data. Nevertheless, the utility owners endeavour to link up the transformer service loss factor with a more easily accessible information viz. the load factor (LF). The determination of the TLF is dependable on some empirical data and can then be evaluated as expressed in (11) [11].

$$TLF = a \times LF^2 + b \times LF \quad (11)$$

here,

a – varies between 0.8-0.89

b – Varies between 0.2-0.11

In practice, the recommended a and b constants are 0.84 and 0.16 respectively. The total ownership cost is then expressed as shown in (12). The CSL are them multiplied by the corresponding no-load (NLL) and load losses (LL) in kW .

$$TOC(ZAR) = TP + CNLL \times NLL + CLL \times LL + CAL \times AL \quad (12)$$

Here, TP is the transformer price offered by the most competitive transformer manufacturer and CL is the cost of auxiliary losses. Prevalently, (4) and (5) are referred to as the A and B factors. Subsequently, (13) can be introduced as new factor C to the classical TOC method, in which ignores this factor.

$$TOC(ZAR) = PP + A \times NLL + B \times LL + C \times AL \quad (13)$$

2.2. Service lifetime loss method: Proposed Method

A majority of SPP's in South Africa are owned by the Independent Power Producers (IPP's). The plant consists of an array of solar photovoltaics modules connected in series and thereupon connected to an inverter that carries out the direct current (DC) to alternating current (AC) conversion. Furthermore, a step-up transformer is connected to the inverter output to increase the voltage to the desired transmission voltage level. Transformers in SPP's are connected on a permanent basis to the electric grid to guarantee the plant is fed with power to fulfil its auxiliary losses during non-generating mode (NGM). The energy cost of the coal power generation (EC_{CPG}) will be affected in this regard. Conversely, during SPPs generation mode (GM) the solar irradiation is available and the solar modules can harvest power. Only, the energy cost of the solar PV (EC_{PV}) in kWh will be affected.

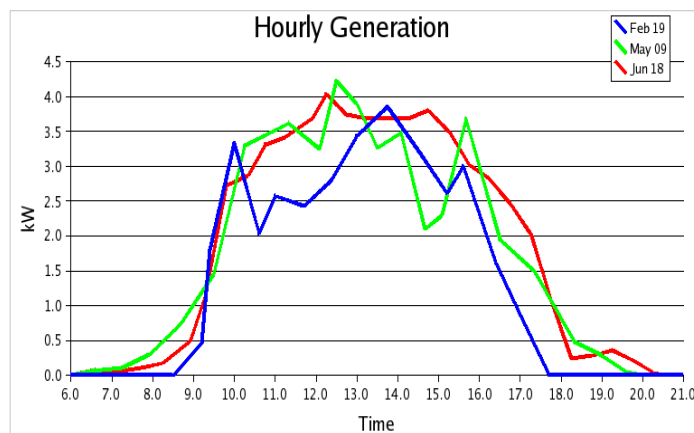


Figure 2: SPP generation profile in South Africa

Figure 2 demonstrate a solar PV generation profile of a SPP as acquired by field measurements and the data is collected over the period of five months. It should be noted that the SPP profile is reliant on the effective average daily solar radiation profile on the

surface area of the solar PV modules and corresponding solar PV technology. As demonstrated in Figure 2, the GM of the SPP can be described by the affective area under the curve between the period 08:30AM to about 18:30PM

The generation modes of the SPP generation profile is then tabulated as shown in Table 2. A 24-hours generation state is considered in the analysis and about 14 hours the solar radiation will be available for the solar modules to harvest power (GM) while about 10 hours of that time the sun will be down (NGM). The ratios of GM and NGM against the 24-hours daily generation state result in the generating (FGM) and non-generation factors (FNGM)

It is reaffirmed that the proposed SLLM proposed in this work is relevant to Independent Photovoltaic Power Producers

Table 2: Generation modes: solar power plant

Generating State	24 hours
GM (Hours)	14
NGM (Hours)	10
FGM	0,583
FNGM	0,417

(13P) supplying electricity to the electric grid by means of step-up transformer. The critical element in capitalizing the service losses in these transformers is the correct definition of the energy component of the cost of service no-load (ZAR/kW) and load losses (ZAR/kW). The energy component is the current energy tariff that will be utilised by a unit kilowatt (kW) of each service loss over the course of designed service lifetime of the SPP. In (14), the proposed method to evaluate the $CNLL$ for transformer intended to operate in SPP. Over the course of the day, the SPP has a greater likelihood to operate GM and NGM. Whenever, the SPP is operating in GM, then it is solely accountable to generate, solely supply of energy and service loss demands as well as transmission of energy to the electric grid. Conversely, during the course of NGM, the SPP purchase energy in a South African context from a coal power generation utility in order to supply its auxiliary requirements and losses. The latter is expressed by (14).

$$CNLL(ZAR/kW) = CGTD + \frac{8760 \times \text{NGM} \times EC_{CPG}}{FCR} \quad (14)$$

In this case, the energy tariff of the coal power generation utility (EC_{CPG}) applies. Moreover, the annual operating hours are weighted in accordance with the availability of the solar radiation. In (15), the case of the evaluating for the cost of service load losses for transformers intended for SPP is also demonstrated by improving (5) from the previous section. The energy tariff of the SPP (EC_{PV}) and corresponding generating factor is considered when the plant is generating its power.

$$\frac{CLL(ZAR/kW)}{FCR} = CGTD \times PRF^2 \times EPR^2 + \frac{8760 \times \text{GM} \times EC_{PV}}{FCR} + \frac{8760 \times \text{NGM} \times EC_{CPG}}{FCR} \quad (15)$$

Additionally, NGM component is added into the formulation to take into account for the when the transformer is operated in reverse as a step-down transformer to supply the auxiliary and loss requirements of the SSP from the grid supply.

3. Case Study: 1500KVA oil-immersed transformer

In this section, the data in Table 3 is considered to conduct the service lifetime loss evaluation. The data is intended to make a substantiated decision to purchase a 1500kVA oil-immersed transformer with 65 °C mean winding temperature. Transformer designs from three distinct manufactures will be considered in this case study. Each design attempt to attain coveted specifications with cost-effective materials, minimum dimensions and lowest overall cost of transformer ownership. In South Africa, increasingly in South Africa, utility owners now purchasing transformers on the basis of the transformer TOC over its intended service lifetime, than just the purchasing price.

Table 3: Service lifetime parameters

Parameter	Value
Designed Service Lifetime	30 years
No. of service hours per year	8760
Annualized Energy Cost	0,61 ZAR/kW
Fixed Charge Rate (FCR)	0,192
System Investment (SI)	23000 ZAR/kW
Discount Rate (<i>i</i>)	10,00%
Energy Escalation rate (<i>i</i>)	2,00%
Load Factor (LF)	0,6
TLF	0,3984
EPR	0,97
Peak Responsibility Factor (PRF)	0,6
Peak Ratio (PR)	0,7
Load Growth Rate (LGR)	2,00%
Coal	1,03 ZAR/kW

In Table 4, a design comparison (A, B and C) of the service losses, selling prices and respective cost of service losses and TOC is presented for three different competing transformer manufacturers. This information is useful in the jurisdiction of the more technically and economically suitable transformer. Design A is the highest loss design of the three offers. Design B and C are lower loss designs. Manufacturer B and C on that account have used better grades of core steel, optimized conductor dimensions and improved cooling medium and procedures

Table 4: Service lifetime loss analysis: classical method

Mn	NLL	LL	SP	CLL	TOC
A	6	25	R460,839	R970 966	R1,431,806
B	5	20	R510,215	R789 677	R1,299,892
C	5	16	R559,591	R696 260	R1,255,851

In Table 4, the classical TOC method has been employed to evaluate the SLLM. This method makes the assumption that the transformer will be in service for 8760 days annually over its designed service lifetime. This premise does not however hold true for renewable energy applications considering the spasmodic nature of renewable energy sources. In Table 5, the proposed solar power plant procedure for evaluating the SLLM is presented. This method takes into consideration the GM and NGM of the plant over the designed transformer service lifetime as described in the previous section. Design C is observed to have lower service losses in comparison to manufacturer B and C even though the

selling price of the manufacturer is higher. It follows that the TOC of the unit is also lower over the normal operating service lifetime

Table 5: Service lifetime loss analysis: solar PV method

Mn	NLL	LL	SP	CLL	TOC
A	6	25	R460,839	R737,986	R1,198,825
B	5	20	R510,215	R597,308	R1,107,523
C	5	16	R559,591	R512,443	R1,072,033

The case studies in Table 4 and Table 5 illustrate that cost of the transformer service losses over the transformer designed lifetime are critical to evaluate over just the purchase price.

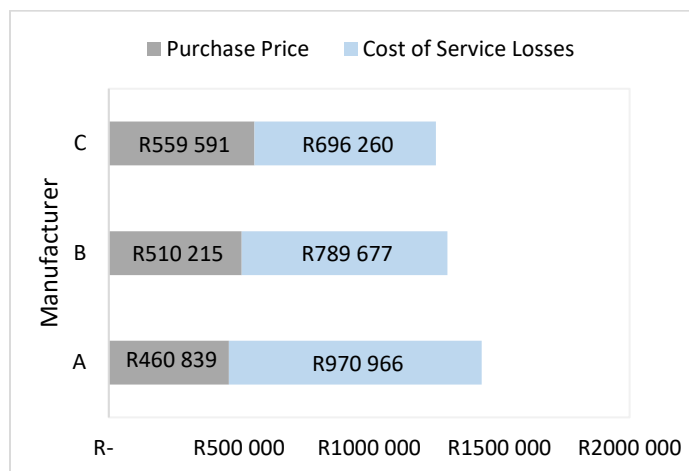


Figure 3: TOC of the transformer manufacturers: Classical method

The results in Figure 3 indicate that manufacturer C has the highest purchase price in comparison to manufacturer B and C by 9% and 18 % respectively. Additionally, the CSL for manufacturer C is lower by 13% and 39% respectively over the designed service lifetime. By employing the classical method, this results indicate that over the designed service lifetime of the transformer under study the IPP can purchase the most expensive transformer while keeping a relatively lower cost of service losses over the intended transformer lifetime. Alternatively, the IPP can choose the cheapest transformer offer with significantly higher service losses over the designed lifetime. However, from a maintenance perspective of the IPPP the later will not be economical.

Results of the proposed SLLM for the three transformer offers are presented in Figure 4. As indicated earlier, throughout the day the SPP will undoubtedly operate in two modes. When the plant is operating in GM, the SSP is accountable to supply its own energy and losses in addition to supplying power to the electric grid. When the SSP is operating on the NGM, then auxiliary requirements of the SSP will be supplied by the electric grid, that is to say that the SPP will purchase power from the coal power generation utility and their energy tariffs will apply.

Figure 4, the results indicate that manufacturer A has the cheapest purchase price and highest CSL. Manufacturer B is the second cheapest offer while manufacturer C is more expensive. The CSL is cheaper by 44% and 17% in comparison to manufacturer A and B respectively. In this regard, the most techno-economical purchase decision for the IPP will be that of manufacturer C.

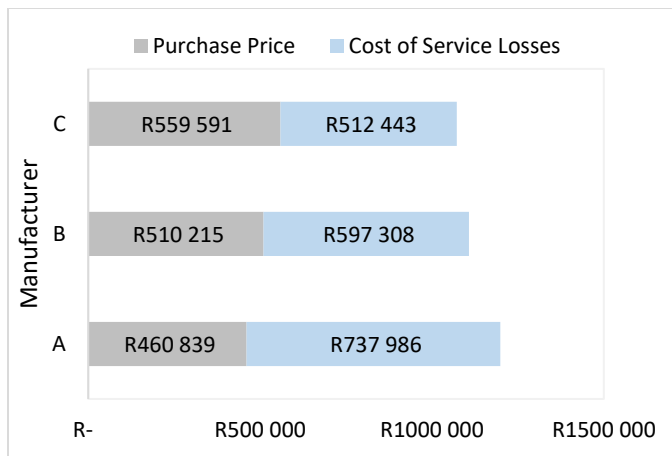


Figure 4: TOC of the transformer manufacturers: Solar power plants

In order to demonstrate the differences between the classical and proposed SPP SLLMs, Figure 5 is presented. It can be observed from Figure 3 and Figure 4 that the transformer offers from each respective method that manufacturer C has a significantly higher selling price but is more economical over the 30 years designed service lifetime.

Manufacturer B and C have managed to optimise their no-load losses by 17% in comparison to manufacturer A. In practice, this loss is minimised by four methods:

- Enhancement of the core steel construction
- Enhanced cutting of the core laminations and employing step-lapped joints as they interleave the core laminations that connect the joint gap.
- Enhancement of the core laminations stacking; and
- Computational modelling of the no-load losses.

The use of amorphous steel in the transformer design also plays a critical role and can minimise the no-load loss by 70-80% as opposed to silicon graded core steel.

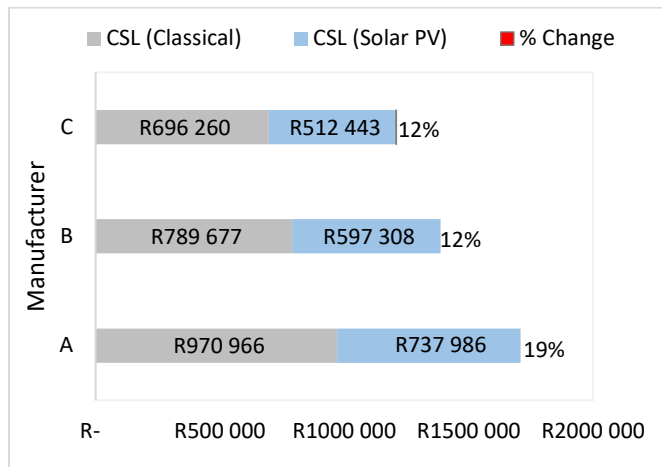


Figure 5: TOC of the transformer manufacturers: Comparison

Overall, manufacturer C has managed to optimise the service load losses by 56% and 25% in comparison to manufacturer A and B respectively. In practice, the service load losses is instrumental to the efficiency of the transformer during service. It therefore critical to reduce these losses for the purpose of enhancing and achieving the most optimal service performance. Reduction of the load loss is treated by the selection of large conductor dimensions

in the interest of minimizing the resistance per unit length of winding conductors. A comparison of the CSL between manufacturer A and B is 12% while between manufacturer A and manufacturer C it is 19%.

The results presented in Figure 5, show some benefits of using the proposed method over the classical method. Taking into consideration the generation modes of the SPP and the ACOE has some techno-economic benefits for the IPP.

4. Conclusion

This work proposes a holistic SLLM to compute the TOC of transformers designed to be of service to solar power plants. These transformers are predominantly owned by IPP’s in South Africa. More notably, the proposed method acknowledge precisely how the cost of the service no-load and load losses should be calculated, taking into consideration the TOC of the transformer in relation to the operational requirements of solar power plants they operate in. The distinct operational features of a SPP have been incorporated in the classical SLLM by means of the two generating modes (GM and NGM). Additionally, the proposed SLLM for the transformer under study is influenced by the reality that this transformer will exclusively facilitate the SSP, as distinguished from another generation utility. As a result the ACOE for SPP in South Africa is employed to evaluate the CSL that will be consumed by the transformer under study during service. At large, it is shown that the solar irradiation plays a pivotal role on the calculation of the CNLL and CLL. This is a property of the SPP that should be cautiously covered, as it may influence the tender jurisdiction criterion to choose the lowest TOC of a transformer over its designed service lifetime.

The insights gained in this work indicate that most economical transformer purchase choice for SPP owners is a high purchase price and low-cost transformer over low-purchase price and high-loss transformer.

Conflict of Interest

The authors declare no conflict of interest.

References

- [1] Department of Energy, “Executive Summary of the Draft Integrated Electricity Resource Plan for South Africa - 2010 to 2030 IRP 2010,” in Department of Energy, 2010, doi: executive-summary-draft-integrated-electricity-resource-plan-south-africa-2010-2030-irp
- [2] Department of Energy, “Integrated Resource Plan 2019, in Department of Mineral Resources and Energy, 2019, doi: IRP/2019/IRP-2019.pdf.
- [3] J. Calitz, C. Mushwana and T. Bischof-Niemz, “Statistics of utility-scale solar PV, wind and CSP in South Africa in 2016,” in CSIR Energy Centre, 2017.
- [4] K.S. Takayuki, K.T. Kobayashi, “How Transformers Age,” in T&D World, 2019, doi: substations/article/20972255/how-transformers-age.
- [5] A. Indarto, I. Garniwa, R. Setiabudy and C. Hudaya, “Total cost of ownership analysis of 60 MVA 150/120 kV power transformer,” 2017 15th International Conference on Quality in Research (QIR) : International Symposium on Electrical and Computer Engineering, Nusa Dua, 291-295, 2017, doi: 10.1109/QIR.2017.8168499.
- [6] A. L. Lazari and C. A. Charalambous, "Life-cycle loss evaluation of power transformers serving large photovoltaic plants in vertically integrated and decentralised systems," in IET Generation, Transmission & Distribution, **9** (8), 759-766, 2015, doi: 10.1049/iet-gtd.2014.0465.
- [7] T. Knutson, "Conducting Distribution Transformer Evaluations Using the Total Ownership Cost Method," 2015 IEEE Rural Electric Power Conference, Asheville, NC, 97-101, 2015, doi: 10.1109/REPC.2015.14.

- [8] C. A. Charalambous, A. Milidonis, S. Hirodiontis and A. Lazari, "Loss Evaluation and Total Ownership Cost of Power Transformers—Part II: Application of Method and Numerical Results," in *IEEE Transactions on Power Delivery*, **28** (3), 1881-1889, 2013, doi: 10.1109/TPWRD.2013.2262507.
- [9] A. Eliasson, H. Elvfing and V. R. Ramanan, "Amorphous Metal core material shows economic and environmental benefits when pre-existing transformers are to be replaced within Vattenfall Group's distribution network," 2010 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT Europe), Gothenberg, 1-7, 2010, doi: 10.1109/ISGTEUROPE.2010.5638963.
- [10] A. L. Lazari and C. A. Charalambous, "Probabilistic Total Ownership Cost of Power Transformers Serving Large-Scale Wind Plants in Liberalized Electricity Markets," in *IEEE Transactions on Power Delivery*, **30** (4), 1923-1930, 2015, doi: 10.1109/TPWRD.2014.2365832.
- [11] IEEE Guide for Transformer Loss Measurement - Redline," in *IEEE Std. C57.123-2019 (Revision of IEEE Std C57.123-2010) - Redline*, 1-86, 2020, doi: 10.1109/IEEESTD.2010.5954117.
- [12] J. Wojtkun, B. Bródka and D. Stachowiak, "The influence of core geometry on no-load losses of medium power transformers," 2018 International Interdisciplinary PhD Workshop (IIPHDW), Swinoujście, 123-127, 2018, doi: 10.1109/IIPHDW.2018.8388339.
- [13] P. Tandeaw, C. Manop and N. Jirasuwankul, "Design and Testing of Low Loss Distribution Transformer for Non-Linear Loading Current," 2018 International Electrical Engineering Congress (iEECON), Krabi, Thailand, 1-4, 2018, doi: 10.1109/IEECON.2018.8712270.
- [14] S. Taheri, H. Taheri, I. Fofana, H. Hemmatjou and A. Gholami, "Effect of power system harmonics on transformer loading capability and hot spot temperature," 2012 25th IEEE Canadian Conference on Electrical and Computer Engineering (CCECE), Montreal, QC, 1-4, 2012, doi: 10.1109/CCECE.2012.6334834.
- [15] B.A Thango, D.B Nyandeni, P.M Molepo J.A Jordaan, A.F Nnachi, "Solar Power Plant Transformer Loss Computation under Harmonic Currents using Finite Element Method", 9th CIGRE Southern Africa Regional Conference, 2019.
- [16] DB Nyandeni, M Phoshoko (Pr. Eng.), R Murray, BA Thango, "Transformer Oil Degradation on PV Plants – A Case Study", 8th South Africa regional conference, 14-17, 2017,
- [17] Bonginkosi A. Thango, Jacobus A. Jordaan, Agha F. Nnachi, "Effects of Current Harmonics on Maximum Loading Capability for Solar Power Plant Transformers", 2020 International SAUPEC/RobMech/PRASA Conference, 1-5, 2020, doi: 10.1109/SAUPEC/RobMech/PRASA48453.2020.9041101.
- [18] A. Elmoudi, M. Lehtonen, H. Nordman, "Effect of harmonic on transformers loss of life", in: *Conference Record of IEEE International Symposium on Electrical Insulation*, **2**, 408-411, 2006.
- [19] R. Singh and A. Singh, "Aging of distribution transformers due to harmonics," *Proceedings of 14th International Conference on Harmonics and Quality of Power - ICHQP 2010*, Bergamo, 1-8, 2010.
- [20] B. A. Thango, J. A. Jordaan and A. F. Nnachi, "Selection and Rating of the Step-up Transformer for Renewable Energy Application," in *SAIEE Africa Research Journal*, **111**(2), 50-55, 2020, doi: 10.23919/SAIEE.2020.9099492.

Assessment of Transformer Cellulose Insulation Life Expectancy Based on Oil Furan Analysis (Case Study: South African Transformers)

Bonginkosi Allen Thango*, Jacobus Andries Jordaan, Agha Francis Nnachi

Department of Electrical Engineering, Tshwane University of Technology, Emalahleni, 1034, South Africa

ARTICLE INFO

Article history:

Received: 04 June, 2021

Accepted: 21 August, 2021

Online: 13 November, 2021

Keywords:

Ageing

Oil

Transformer

Dissolved Gas Analysis (DGA)

Degree of polymerization (DP)

ABSTRACT

The ageing of oil-immersed power transformers triggers several defects and damages in the insulating materials, particularly in the cellulose insulation. The decomposition of the cellulose paper produces dissolved gases into the insulating oil, in which the Dissolved Gas Analysis (DGA) of the oil samples can provide insights to incipient faults sustained by the transformer as well as the status of the insulation. The condition of the cellulose insulation can be established by the measurement of the Degree of Polymerization (DP). In some instances, when the DP is intricate to measure, the estimation can be realized by collating information with respect to oil Furan analysis (2-furfuraldehyde (2FAL)). In this work, regression analysis based on 120 transformer oil samples is developed to establish new formulae by generating an analogy between the DP and 2FAL to indicate the status of the cellulose insulation and percentage of remnant life expectancy. In present study, five different transformers are tested using the proposed formulae to compute the DP based on the 2FAL concentrations. The results, are compared with measurement results and four other existing models. The results indicate an error of estimate of less than 2% and 1% against the measured DP.

1. Introduction

Oil-immersed power transformers form an integral part in the continuous operation of the South African power grid. Retaining the continuous operation of power transformer is thereby a critical issue for the power utility for the reason that any operational malfunction leads to loss of revenue. On that account, scheduled maintenance contributes to preventing transformer failures and power outages. A large percentage of transformer failures is contributed by the damage of the insulation materials, which is composed of insulating oil and cellulose paper [1-5].

The decomposition process of transformer insulating materials amplify at higher hotspot temperature, presence of oxygen and production of moisture. The examination of the transformer insulation condition can be established by diagnostic methods among which includes DGA [5, 6], DP by measuring furan compounds [7, 8], or Gel permeation chromatography [9, 10].

Furfural analysis has gained immense popularity in the South African transformer manufacturing industry in the evaluation of the strength of insulating cellulose paper and corresponding remnant service lifetime. Furans are the derived products of the

decomposition of the cellulose paper that disseminate into the transformer oil. In measuring the furan concentration on the transformer oil, then the cellulose paper conditions can be ascertained. A rigorous method to assess the tensile strength in the cellulose paper that is intricately connected to cellulose insulation strength of a transformer is unearthing the Degree of Polymerization (DP) of the insulation. However, it is infeasible to obtain the paper samples from a transformer that is in-service, particularly paper samples in winding hotspot areas that are typically localized in the centre of the transformer steel tank. A number of researchers [11-14] have concluded that there is a remarkable connection between the DP and furan concentration, specifically 2-Furaldehyde (2FAL). The benefit of analysing the 2FAL concentration correlated to the DP is that the transformer oil samples can be appropriately extracted from the unit in-service without engendering any impairment to the transformer.

The most common procedure in measuring Furanic compounds in the transformer oil is recommended by the standard published in [15]. In view of the fact that this procedure necessitates particular expertise and high-end equipment, utility owners and manufacturers subcontract the Furan test to certified service enterprises. The utilisation optical spectroscopy techniques is also

*Corresponding Author: Bonginkosi Allen Thango, thangotech@gmail.com

more and more notable in the Southern African transformer condition monitoring.

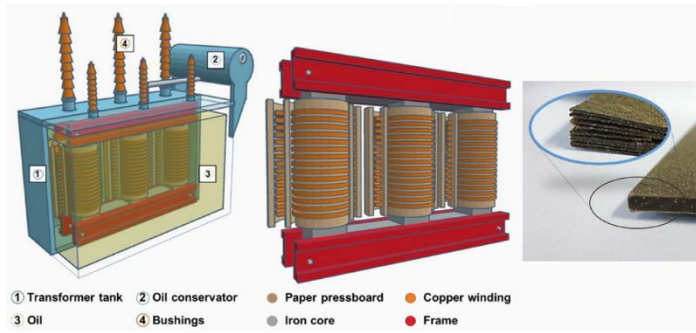


Figure 1:Transformer cellulose paper

The current work proposes an expedient alternative to evaluate the DP in the transformer cellulose paper using a statistical survey. A set of tests were carried out in the laboratory based upon oil and cellulose paper samples of transformers that have aged and also units that were collected from site as a results of various faults. Regression models have been developed to estimate the DP in the cellulose paper using the measured 2FAL concentrations. The results indicate an error of estimate of less than 2% and 1% of the proposed logarithmic and polynomial fit DP models against the measured DP respectively.

2. Furan compounds mechanism

The Furan compounds present in the transformer oil that can be measured by the High-Performance Liquid Chromatography (HPLC) are namely the 2FAL, 2-Furfural (2FOL), 5-Hydroxy methyl-2-furfural (5HMF), 5Methyl-2-Furfural (5MEF), and 2-acetylfuran (2ACF) as demonstrated in Figure 2 [16].

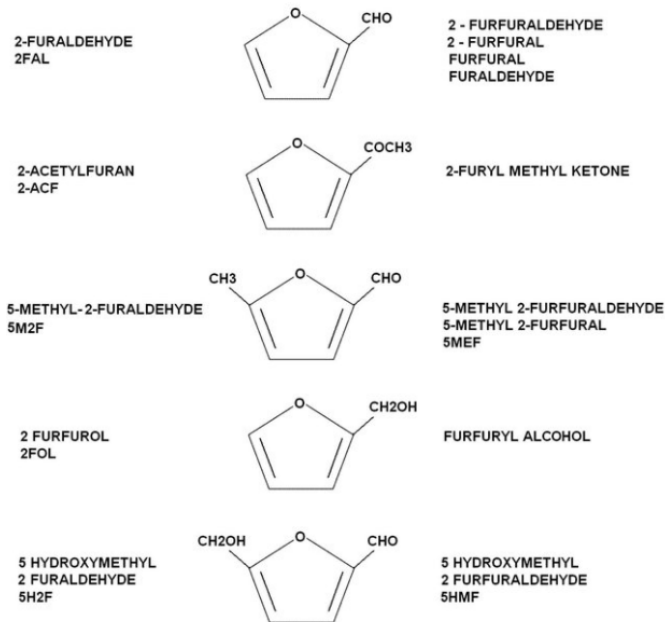


Figure 2: Furan compounds present in the transformer oil.

It is widely adopted the Furans are triggered from the decomposition of cellulose insulating material during the course of the transformer operational lifetime.

In several researches from the literature, case studies of transformer diagnostic status have been presented where the measured concentration of the furan in oil is employed as a key performance indicator of the chemical decomposition of the insulating cellulose paper. Evaluation of the results of furan analysis may be implemented with other assessments such as the monitoring of in-service transformer insulating oil and winding temperatures. As a result of intricacy of the cellulose decomposition mechanics and plethora of factors that can influence the process, the correlation between the 2FAL concentration and the service life of the transformer is not fully unravelled. Based on field experience, the factors that have a material impact on the composition of 2FAL in the transformer insulation oil are summarized below:

- The commencing 2FAL concentration in the transformer oil is characterised by the ratio of the generation/decomposition rate to a particular temperature;
- The occupancy of moisture considerably affect the dissemination of furan compounds between insulating cellulose and oil. The foregoing espouse the 2FAL composition;
- The winding temperature allures the formation / decomposition of 2FAL. The variability of the operating temperature as a result of miscellaneous loading conditions substantially affect the formation of furans;
- Types of transformer oil are also critical on the formulation of 2FAL. Oxidation inhibitors have a particular safeguarding effect on cellulose thermoset and triggers the deceleration of furan compounds in comparison with non-additive oils;
- Thermally upgraded paper generate minimal quantity of furan compounds than standard Kraft paper in oil;
- Types of transformer (cellulose/oil ratio) and cooling state is of paramount importance. Accruing the quantity of cellulose add up to the composition of furans;
- Oil reclamation bring about a lower furan content throughout a lapse of time. However, as a chemical equilibrium, the initial 2FAL concentration will be attained once more.

Several studies have been published in modelling the interrelation between the oil 2FAL concentration and the DP of cellulose material. A universal correlation that is endorsed by the transformer manufacturing industry is presently not available. Nevertheless, there is a common characteristic on the models proposed by authors in [17–21], which concludes that there is a correlation between the 2FAL concentration and the DP of cellulose. Other furan compounds such as 2FOL, 5HMF, 5MEF, and 2ACF are more erratic than 2FAL and tend to decompose rapidly.

Table 1 depicts the DP models that are employed in the comparison approach to substantiate the proposed models. The proposed Eq. (6) and Eq. (7) above are formulated based on logarithmic and polynomial fit regression models using furan analysis data collected from a fleet of mineral oil-filled transformers, particularly in the South African network. These models show excellent correlation coefficient (R^2) between the 2FAL concentration and the DP of cellulose with the polynomial fit model having an improved R^2 . A graphical of these models and existing models is then demonstrated as shown in Figure 3.

Table 1: Various DP models

Eq.	Author	DP exponential formula	Ref.
1	Chendong	$DP = \frac{\log_{10} 2FAL - 1,51}{-0,0035}$	[17]
2	Dong	$DP = 402,47 - 220,87 \times \log_{10} 2FAL$	[18]
3	Vaurchex	$DP = \frac{2,6 - \log_{10} 2FAL}{0,0049}$	[19]
4	Heisler	$DP = 325 \times \left(\left(\frac{19}{13} \right) - \log_{10} 2FAL \right)$	[20]
5	De Pablo	$DP = \frac{1850}{2FAL + 2,3}$	[21]
6	Thango (Eq. 6)	$DP = -123,6 \times \ln 2FAL + 456,38$ $R^2 = 0,9809$	
7	Thango (Eq. 7)	$DP = 0,2358 \times 2FAL^6 - 5,4819 \times 2FAL^5 + 49,471 \times 2FAL^4 - 220,64 \times 2FAL^3 + 520,55 \times 2FAL^2 - 692,51 \times 2FAL + 811,15$ $R^2 = 0,9913$	

The percentage of the decomposed cellulose life for a unit in-service can be ascertained the Eq. (9) and Eq. (10) below.

Table 2: percentage of the decomposed cellulose

Eq.	Author	%RSL exponential formula	Ref.
9	Kanumuri	$\% RSL = \frac{LOG10(DP) - 2,903}{-0,006021}$	[23]
10	Doble	$\% RSL = 166,1 \times \log_{10}(2FAL) - 382,2$	[24]

Eq. (9) gives the decomposed cellulose life as a percentage, but to measure transformer paper life in years, Eq. (5) is proposed in [22] may be employed. The parameters used assume initial DP (DP₀) as 1100 and a 35-year transformer lifespan.

3. Case Study: South African oil-immersed transformers

In this section, the sample results of furan analysis of five transformers are presented. The DP has been measured for each of the case studies. At glance, transformer #5 has the lowest 2FAL concentration with a slightly higher measured DP, which suggest this unit is older than the other units. Additionally, transformer #2 has a higher 2FAL value and lowest measured DP.

Table 3: Service lifetime parameters

#	2FAL	Measured DP
1	1,563	408
2	0,231	648
3	1,601	405
4	1,778	393
5	2,377	351

In Table 4, the proposed models in Eq. (6) and Eq. (7) to evaluate the amount of DP in the transformer insulating cellulose paper are compared with measured data and other models in the literature. The comparison will be carried out based upon the error of estimated of the estimated and measured samples.

Table 5 presents the error of estimate of some case studies using Eq. (6) and Eq. (7) and other DP models from the literature. The 5 cases in Table 3 are employed to examine the accuracy of Eq. (6) and Eq. (7). Only one case (Case 2) shows a slightly higher error of estimate against the measured data, but all the cases for Eq. (6) and Eq. (7) are less than 2% and 1 % respectively. The error of estimates for Chendong are less than 8%, Dong less than 12% except for higher case (case 2), Vaurchex has less than 29% and De Pablo less than 17%.

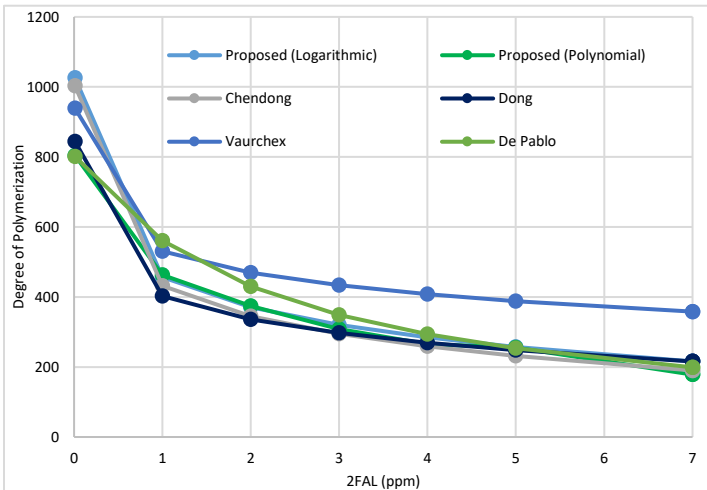


Figure 3: Comparison of DP models

2.1. Transformer remaining service lifetime

Consequently, the remaining life of the transformer can determined using Eq. (8), where the DP values range from 1200 ppm to 1000 ppm, with 200 ppm for degraded paper insulation [22].

$$RSL = 20,5 \times \ln \left(\frac{1100}{2FAL} \right) \quad (8)$$

Table 4: Degree of Polymerization

#	2FAL	Measured DP	Thango (Eq. 6)	Thango (Eq. 7)	Chendong	Dong	Vaurchex	De Pablo
1	1,563	408	401	406	376	360	491	479
2	0,231	648	637	676	613	543	660	731
3	1,601	405	398	403	373	357	489	474
4	1,778	393	385	390	360	347	480	454
5	2,377	351	349	349	324	319	454	396

Table 5: Error of estimate (Degree of Polymerization)

#	Thango (Eq. 6)	Thango (Eq. 7)	Chendong	Dong	Vaurchex	De Pablo
1	2%	1%	8%	12%	20%	17%
2	2%	4%	5%	16%	2%	13%
3	2%	1%	8%	12%	21%	17%
4	2%	1%	8%	12%	22%	15%
5	0%	1%	8%	9%	29%	13%

Table 6: Remnant Service Lifetime

#	2FAL	Measured DP	Thango (Eq. 6)	Thango (Eq. 7)	Chendong	Dong	Vaurchex	De Pablo
1	1,563	20,33	20,68	20,46	22,01	22,92	16,53	17,05
2	0,231	10,85	11,18	9,97	11,98	14,47	10,46	8,38
3	1,601	20,48	20,83	20,60	22,17	23,05	16,62	17,25
4	1,778	21,10	21,51	21,27	22,90	23,64	17,02	18,16
5	2,377	23,42	23,51	23,54	25,06	25,35	18,15	20,97

Table 7: Percentage Remnant Service Lifetime (Eq. 9)

#	2FAL	Measured DP	Thango (Eq. 6)	Thango (Eq. 7)	Chendong	Dong	Vaurchex	De Pablo
1	1,563	48,55	49,77	48,99	54,44	57,66	35,19	37,00
2	0,231	15,18	16,36	12,09	19,16	27,93	13,81	6,50
3	1,601	49,09	50,31	49,51	55,02	58,12	35,51	37,70
4	1,778	51,26	52,69	51,85	57,58	60,18	36,89	40,90
5	2,377	59,41	59,74	59,86	65,18	66,21	40,87	50,79

Table 8: Percentage Remnant Service Lifetime (Eq. 10)

#	2FAL	Measured DP	Thango (Eq. 6)	Thango (Eq. 7)	Chendong	Dong	Vaurchex	De Pablo
1	1,563	51,43	50,21	50,99	45,54	42,33	64,79	62,99
2	0,231	84,80	83,62	87,89	80,83	72,05	86,18	93,49
3	1,601	50,90	49,68	50,47	44,97	41,86	64,48	62,28
4	1,778	48,73	47,29	48,13	42,41	39,80	63,09	59,08
5	2,377	40,58	40,24	40,12	34,80	33,77	59,12	49,20

After examining the veracity of Eq. (6) and Eq. (7) to evaluate the DO, the remnant service lifetime of the insulating cellulose paper was established, at which point, the studied transformers was determined by applying Eq. (8). Table 6 shows the corresponding remnant service lifetime of the transformer.

Table 7 and Table 8 demonstrate the percentage remnant service lifetime of the 5 cases in Table 6. The results indicate that there is an inversely proportionate relationship between the percentage remnant service lifetime and the 2FAL concentration while a directly proportional relationship is observed against the degree of polymerization. The results of the percentage remnant service lifetime by applying Eq. (9) is tabulated below.

An increase in the 2FAL concentration results in the decrease of the degree of polymerization and consequently the transformer's remnant in-service lifetime. The degree of polymerization of insulating cellulose paper is high for newly manufactured transformers ranging above 800 to 1200 and the cellulose paper with decomposition activity due to the operational conditions will decrease radically until it reaches a DP of about 2000 where the unit must be removed from service as it indicates

end of lifespan. Subsequently, the results of the percentage remnant service lifetime by applying Eq. (10) is tabulated.

Based on the results presented in Table 7 and Table 8, it can be observed and concluded that the proposed Eq. (6) and Eq. (7) may be adopted by local manufactures to reliably evaluate the remnant in-service lifetime of transformers connected to the South African grid. Results of Chendong, Dong, Vaurchex and De Pablo are erratic in evaluating the transformer service life.

4. Conclusion

Examination of cellulose paper in transformers is highly complex when the unit is in-service, consequently chemical tests were conducted to survey the condition of the insulating cellulose paper. The degree of polymerization was wielded as a key benchmark for evaluating the condition of the cellulose paper by measuring the Furan (2FAL) concentration. A statistical survey through logarithmic and polynomial fit regression models was realized to develop the prognosis equations allying the degree of polymerization and the measured 2FAL concentration. The statistical survey results demonstrated that there is a high correlation between the cellulose DP and the 2FAL concentration,

yielding correlation coefficients of 0.9809 and 0.9913 for the proposed models.

The results of the studied cases inveterate that the 2FAL concentration is a suitable index for predicting the degree of polymerization. By comparing the proposed models in Eq. (6) and Eq. (7) with the DP models in the literature, the results revealed that the proposed DP models based on South African transformers to be an excellent measure of the degree of polymerization at which point the remnant and percentage service lifetime of the units can be evaluated. The error of estimate of the proposed Eq. (6) and measured data is less than 2 % while Eq. (7) is less than 1% except for an anomaly observed in case number 2. Chendong, Dong, Vaurchex and De Pablo yielded and error of estimate of no more than 8%, 16%, 29% and 17% in comparison to the measured data. Similar trends in the estimation of the remnant and percentage service lifetime are observed.

The continuation of this work in the future will comprise the application of least squares support vector machine for classifying the condition of the insulating cellulose paper and then evaluate the remnant service lifetime of a transformer.

Conflict of Interest

The authors declare no conflict of interest.

References

- [1] M. K. Pradhan and T. S. Ramu, "Diagnostic testing of oil-impregnated paper insulation in prorated power transformers under accelerated stress," Conference Record of the 2004 IEEE International Symposium on Electrical Insulation, 66-69, , 2004, doi: 10.1109/ELINSL.2004.1380452.
- [2] C. Chakradhar Reddy, E. Johnson and T. S. Ramu, "On the life estimation of oil impregnated paper insulated transformer bushings," 2007 Electrical Insulation Conference and Electrical Manufacturing Expo, 13-16, 2007, doi: 10.1109/EEIC.2007.4562578.
- [3] Y. Tu, W. Sun, C. Yue, G. Chen and W. Wang, "Research of insulation properties of polymer materials using in oil-filled transformers under high temperature," 2011 Annual Report Conference on Electrical Insulation and Dielectric Phenomena, 583-586, 2011, doi: 10.1109/CEIDP.2011.6232724.
- [4] S. Missas, M. G. Danikas and I. Liapis, "Factors affecting the ageing of transformer oil in 150 / 20 kV transformers," 2011 IEEE International Conference on Dielectric Liquids, 1-4, 2011, doi: 10.1109/ICDL.2011.6015409.
- [5] T. K. Saha, M. Darveniza, Z. T. Yao, D. J. T. Hill and G. Yeung, "Investigating the effects of oxidation and thermal degradation on electrical and chemical properties of power transformers insulation," in IEEE Transactions on Power Delivery, **14**(4), 1359-1367, 1999, doi: 10.1109/61.796229.
- [6] D. M. Mehta, P. Kundu, A. Chowdhury and V. K. Lakhiani, "DGA diagnostics save transformers - case studies," 2015 International Conference on Condition Assessment Techniques in Electrical Systems (CATCON), 116-120, 2015, doi: 10.1109/CATCON.2015.7449519.
- [7] S. Apte, R. Somalwar and A. Wajirabadkar, ""Incipient Fault Diagnosis of Transformer by DGA Using Fuzzy Logic"," 2018 IEEE International Conference on Power Electronics, Drives and Energy Systems (PEDES), 1-5, 2018, doi: 10.1109/PEDES.2018.8707928.
- [8] K. Baburao, N. M. Bhangre, A. M. Wagle, A. Venkatasami and S. E. Chaudhari, "The experience of DP and furan in remnant life assessment of power transformer," 2008 International Conference on Condition Monitoring and Diagnosis, 555-558, 2008, doi: 10.1109/CMD.2008.4580348.
- [9] N. Das, A. Abu-Siada and S. Islam, "New approach to estimate furan contents in transformer oil using spectroscopic analysis," 2012 22nd Australasian Universities Power Engineering Conference (AUPEC), pp. 1-4, 2012.
- [10] S. P. Lai, A. Abu-Siada, S. M. Islam and G. Lenco, "Correlation between UV-Vis spectral response and furan measurement of transformer oil," 2008 International Conference on Condition Monitoring and Diagnosis, 659-662, 2008, doi: 10.1109/CMD.2008.4580372.
- [11] D. Feng, Z. Wang and P. Jarman, "Transmission power transformer assessment using furan measurement with the aid of thermal model," 2012 IEEE International Conference on Condition Monitoring and Diagnosis, 521-524, 2012, doi: 10.1109/CMD.2012.6416194.
- [12] M. Meissner, M. Darmann, S. Schober, M. Mittelbach and C. Sumereder, "Reliability Study of Furan Level Analysis for Transformer Health Prediction," 2019 IEEE 20th International Conference on Dielectric Liquids (ICDL), 1-4, 2019, doi: 10.1109/ICDL.2019.8796785.
- [13] K. B. Shaban, A. H. El-Hag and K. Benhmed, "Prediction of Transformer Furan Levels," in IEEE Transactions on Power Delivery, **31**(4), 1778-1779, 2016, doi: 10.1109/TPWRD.2016.2521320.
- [14] M. A. Talib, M. A. A. Aziz, Y. Balasubramaniam, Y. Z. Y. Ghazali, M. R. Z. Abidin and M. F. M. Yousof, "Transformer Ageing Investigation and Correlation of Furanic Compound and Degree of Polymerization for Power Transformer Life Assessment," 2020 IEEE International Conference on Power and Energy (PECon), 240-244, 2020, doi: 10.1109/PECon48942.2020.9314543.
- [15] "IEEE Guide for Evaluation and reconditioning of Liquid Immersed Power Transformers," in IEEE Std. C57.140-2006, 1-73, 2007, doi: 10.1109/IEEESTD.2007.353650.
- [16] I. Hohlein and A. J. Kachler, "Aging of cellulose at transformer service temperatures. Part 2. Influence of moisture and temperature on degree of polymerization and formation of furanic compounds in free-breathing systems," in IEEE Electrical Insulation Magazine, **21**(5), 20-24, 2005, doi: 10.1109/MEI.2005.1513426.
- [17] X. Chendong, F. Qiming, X. Shiheng, "To Estimate the Ageing Status of Transformers by Furfural Concentration in the Oil", In Proceedings of the CIGRE Study Committee 33 Colloquium, Leningrad, Moscow; 1991.
- [18] M. Dong, Z. Yan and G. J. Zhang, "Comprehensive diagnostic and aging assessment method of solid insulation in transformer," 2003 Annual Report Conference on Electrical Insulation and Dielectric Phenomena, 137-140, 2003, doi: 10.1109/CEIDP.2003.1254813.
- [19] H. Vaurchex, I. H hlein and A. J. Kachler, "Transformer aging research on furanic compounds dissolved in insulating oil", CIGRE, 2002, doi: publication/15-302_2002-transformer-ageing-research-on-furanic-compounds-dissolved-in-insulating-oil
- [20] H. Andreas et al. "Zustandsbeurteilung von Transformatoren mit Furfuroil-Bestimmung." 2003.
- [21] A. de Pablo, "Furfural and ageing: how are they related," IEE Colloquium on Insulating Liquids (Ref. No. 1999/119), 5/1-5/4, 1999, doi: 10.1049/ic:19990667.
- [22] M. K. Pradhan and T. S. Ramu, "On the estimation of elapsed life of oil-immersed power transformers," in IEEE Transactions on Power Delivery, **20** (3), 1962-1969, 2005, doi: 10.1109/TPWRD.2005.848663.
- [23] D. Kanumuri, V. Sharma, O.P. Rahi, O P, "Analysis Using Various Approaches for Residual Life Estimation of Power Transformers", International Journal on Electrical Engineering and Informatics, Bandung **11**(2), 389-407, 2019, doi:10.15676/ijeei.2019.11.2.11.
- [24] Doble," Testing & Sampling", doi: services/laboratory-services/testing-and-sampling/.

Service Robot Management System for Business Improvement and Service Extension

Hideya Yoshiuchi*, Tomohiro Matsuda

Hitachi, Ltd., Research & Development Group, Center for Technology Innovation – Digital Platform, 1-280, Higashi-koigakubo, Tokyo, 185-8601, Japan

ARTICLE INFO

Article history:

Received: 14 July, 2021

Accepted: 26 October, 2021

Online: 13 November, 2021

Keywords:

Service Robot System

Data Analysis

External Equipment Association

ABSTRACT

Service robots are applied in many kinds of business fields and its various kinds of data can be collected with service robot in business scenes. Therefore, it becomes more important to utilize such data for business improvement. Additionally, due to limitation of design such as body size and battery life, a service robot cannot prepare much peripheral equipment on its body. In this paper, we will show two research results as for service robot. One is data analysis technology for business improvement as a function of operation and management system for service robot. Another is service extension of service robot by association of external equipment with proper design of utilization condition of external equipment. Through evaluation experiment, we confirmed potential effect of business improvement is 8.1 % by modifying service scenario of robot.

1. Introduction

This paper is an extension of work originally presented in 5th International Conference on Robotics and Artificial Intelligence [1]. Due to the several social trends such as labor power and human resource issue, service robots technologies are being developed for realizing not only simple operation automation but also business improvement [1]. Recently, more people highly expect service robot system [2], [3] can solve social issues such as lack of labor power [4], improvement of productivity, human care [5] and so on. In manufacturing field, industrial robots were deployed and applied in 1970s for factory automation, and Internet of Things (IoT) technologies are utilized for productivity improvement and reduction of operation cost [6] with safe environment [7]. In service business field, service robots are expected to realize business and service quality improvement [8], [9] in the future aging society with fewer children. Based on these social and business background, we develop service robot and its application with intellectual media processing technology such as voice recognition and image analysis.

As for the robot deployment, people have higher attention to how to utilize various kinds of robots in actual fields for the purpose of collaboration with people. In industry field, Industry 4.0 [10] aims effective manufacturing with various technologies. Industrial robots are one of the most important technology in this fields. Industrial robots are designed to realize both production efficiency and safety for work with human in the same fields [11].

On the other hand, service robots are assumed to be applicable to many kinds of service and business scene through verbal and multimodal communication [12] not only for replacement of human labor load by automation but also for giving helpful function to people, therefore, applications for service robot have many variations for their purposes. For example, in a large facility such as an airport or a shopping mall, customers sometimes have difficulties for finding the way to their destination due to the scale and complexity of the facility. In such case, a service robot can give customers facility guidance information and support customers. In another case, a service robot can provide detail explanation of sold goods in a large shopping mall.

The value provided by service robots is considered in Business-to-Business-to-Consumer (B to B to C) scheme, and there are differences between the value of service provider (business side) expected and the value of consumer expected. Service providers apply service robots to their facilities for not only labor power but also some special skills such as multi-language guidance that normal employees cannot learn in short time. Consumers think service robots give them dedicated guidance. For example, when customers visit a large shopping mall for the first time and does not have enough knowledge for the mall, they ask a service robot where the shop that they are looking for. The service robot can give the location information by facility map, or the service robot can bring the customers to their destination by physical movement in the mall.

*Corresponding Author: Hideya Yoshiuchi, hideya.yoshiuchi.fq@hitachi.com

People consider that service robots have much potential for social innovation. On the other hand, especially service providers, people would like to know how effective service robots are. Therefore, it is important to evaluate and visualize the effect of service robot for service providers and consumers. To evaluate service robots' effect by a kind of quantum key performance indicator (KPI), we have to clarify what service provider expect to service robots for their business improvement in each service field. After specifying KPI for business improvement measurement, we consider service scenario for robots and verify the effect of service robots operated by the service scenario with monitoring the specified KPI.

In addition to business improvement, it is important to consider how to realize applications by service robots with limited capability. Because service robot has several limitations such as body size and battery duration, it is difficult to prepare enough devices for applications to be realized by service robot. In such case, we need to consider how to extend capability of service robot for extended service.

In this paper, we explain two technologies. One is the service robot system and data analysis technology for business improvement. The other is external device association technology for service extension with service robot system. For both proposed technology, we will show evaluation results. The structure of this paper is as follows: In chapter 2, we will show basic architecture of service robot system. In chapter 3, we will explain data analysis technology of service robot system for business improvement and its evaluation result. In chapter 4, We show external device association technology for service extension. Chapter 5 conclude this research.

2. Service Robot System

2.1. Summary of Service Robot System Architecture

Figure 1 shows general system architecture of service robot system. The service robot system consists of service robots and application servers that provides media processing function such as voice recognition, image processing and language processing. Additionally, the service robot system includes operation and management server that gives us management function for system operation such as machine condition monitoring and remote operation for emergency case. Because the main function of a service robot is service application through interaction with people, service robots are deployed in facilities of service providers. The methods of interaction between a service robot and people are sound and image. Therefore, a service robot has microphone and video camera for voice and image processing. A service robot system gathers voice and video data through microphone and video camera on a service robot, processes such media data, and provides basic interaction function to end users such as replying questions and giving useful information. Media processing function like voice recognition and video analysis consume much computational resources such as Central Processing Unit (CPU), sometimes media processing function is realized by application servers located out of a service robot.

Implementing media processing function by external application servers makes the size of a service robot smaller one. However, the data size of voice, image and video is relatively

large and sending such media data from a service robot to application servers requires much larger network resources. Therefore, it is better to put application servers in the same service facilities as service robot's location. However, if it is not possible to put application servers in service facilities for some reasons such as facility size or no server room, application servers can be set in a management center.

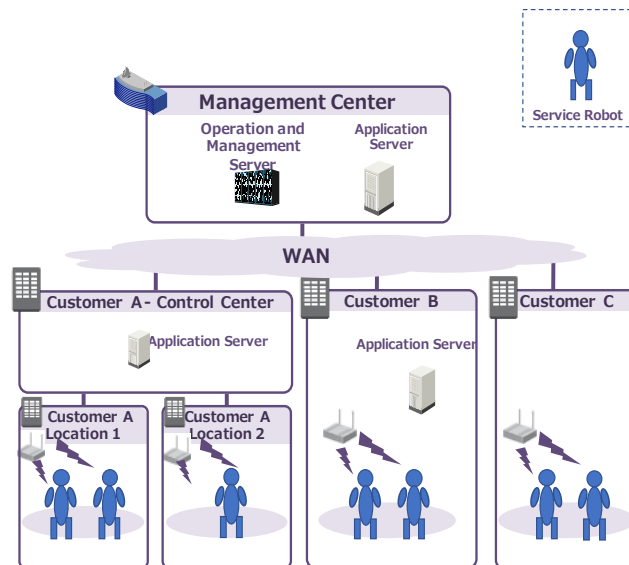


Figure 1: General architecture of service robot system

2.2. Research Task of Service Robot System

In service robot system, media processing is primitive function for interaction with end users. The combination of media processing realizes voice communication, and questions and answers. The one of the main purposes of developing service robot system is to organize service scenario by the set of communication and Q&A list for various kinds of business scene such as guidance in shopping mall or large transportation facilities (ex. airports and railway stations). Service scenario consists of timing information of voice communication, contents of communication, gestures and movement of a service robot. Service scenarios are designed for satisfying business and service requirement of service provider. It is difficult to evaluate the effect of service robot system before service robot system start its operation. Therefore, one of the research challenges of service robot system is to establish the evaluation method of the application of service robot system for assumed KPI suitable for business and service requirement, and proposal scheme of service scenario modification to realize business improvement. Besides KPI issues, we have to consider how to realize service application with limited resource of service robot. If we set many devices on service robot, its size is getting bigger, and weight is getting heavier. Because service robots work among peoples, sometimes in busy crowded area, it needs to keep human safety [13], [14]. If robots are bigger and heavier, service robots easily violate human safety. Therefore, robot developer has to consider not only functionality realized by various devices, but also risk assessment for safer robot operation [15], [16].

3. Data Analysis Technology for Service Robot System

In this section, data analysis technology for service robot system that realizes business improve is explained.

3.1. System Architecture

Figure 2 shows the architecture of service robot system with data analysis technology for business improvement. This system consists of the following components.

- Service robot: Communicate with end users by interaction with voice and video, gesture and movement and provides various kinds of service to the end users
- External equipment and system: Realize additional function that the service robot cannot provide such as GUI by digital signage.
- Application server: Providing media processing primitive function such as voice, video and language processing.
- Integrated service system: Store and manage all the data that can be collected in service robot system. For the detail of collected later, we will explain in later section
- Business analysis server: Provide data analysis function for business improvement with all the data stored in the integrated service system.
- Customer business system: Various external system for customers' business purpose such as marketing, accounting and so on.

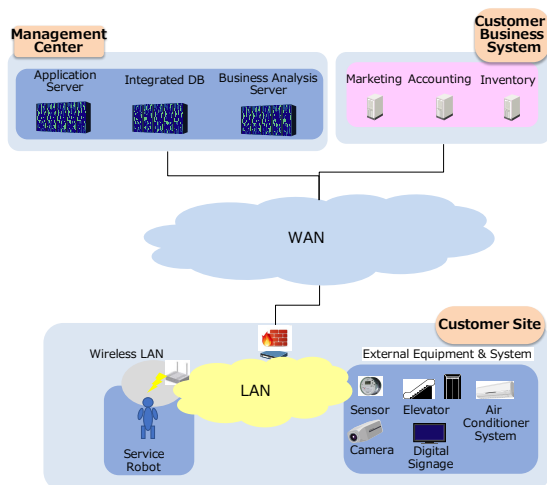


Figure 2 Architecture of Service Robot System with Data Analysis Technology

3.2. Data Analysis for Business Improvement

In service robot system, service robot provides various functions to end user by service scenario. The possible actions of a service robot consist of communication method by voice, image and language processing, Service scenario combines these actions. From the viewpoint of service provider, they expect service robots to do some business operations instead of their staff. The typical business operation realized by service scenario is guidance activity in commercial facilities or transportation facilities.

To realize business improvement, KPI related to business operation is measured and monitored. KPIs are indicators that

evaluates business activities. For example, ISO 22400 defines important KPIs for manufacturing field mainly focusing on Manufacturing Execution System (MES) [17]. Commonly, KPIs focus on the following items:

- Cost: Cost KPI allows you to evaluate the activities that utilize economic resources in different processes. This can include energy consumption, cost of raw materials, cost of human resources and overhead costs.
- Time: Time KPI measures the time spent on an activity or a process. This can be the time from the beginning to the end of a process or a step of a process. Ways to measure time can be by the percentage of products or services delivered on time or the percentage of late deliveries.
- Productivity: Productivity KPI allows your organization to measure the efficiency of the operation. This can include measuring the percentage of downtime your equipment or machine experience. It also can be an output of operations such as product items or the number of served customers.
- Quality: Quality KPI help calculates the effectiveness of processes. These numbers may be based on the number of errors occurring and the number of completed processes. Other way to measure is the percentage of products that need corrective action and the number of satisfied customers.

Service robot applications have different tendency from industrial robot use case. Service robots are mainly used for support of end user's activities in service facilities such as commercial mall, railway station, airport, hospital and so on. Service robots mainly provide intellectual works to those customers; therefore, it is not always suitable to measure KPIs related to manufacturing productivity or automation activities. Suitable KPI candidates are service-related value such as service quality, the number of users that communicate with service robots and so on. After we decided to utilize service robot into our service environment, we have to choose KPIs that have relationship with service robot application. To analyze how the behavior of a service robot affects the KPIs, we need to monitor and store the data which describe service robots' activities. As for KPI evaluation, one or more service scenario affects the KPI value change. For example, daily sales information of commercial complex facilities is a sum of daily sales of each tenant. Some shop items are very popular and have strong effect to the daily sales of the tenant. In another viewpoint, daily sales information has relationship with likability of a facility.

To realize business improvement measured by KPIs, it is effective to figure out the service scenario that has strong relationship with business KPI and try to apply the service scenario as much as possible to service robots. For this purpose, we need to gather the enough data from service robot system for analysis of relationship between business KPI and service scenario. Then, we consider how to realize KPI improvement by modifying service scenario that much affects the KPI. The modification ways of the service scenarios are just candidate method and there is no assured result. Therefore, we must apply

the modified service scenario to service robots and evaluate its efficiency by checking KPI increase or decrease. If the KPI is improved, we decide to apply scenario modification permanently. If the KPI decreases, we discard scenario modification and try another modification. The data type to be collected in service robot system for business improvement is as follows:

- KPIs to be improved as time-series data
- Service scenarios that are applied to service robots
- The type of service scenario
- Relationship between the KPI and service scenarios
- Execution result of service scenario
- Action history of service robots as time-series data
- Motion history of service robots: raising a hand, nod, showing bye-bye, etc.
- Image processing results
- Voice recognition results
- Statements of service robots corresponding to the end users

Here we show the detail of data analysis process of service robot system for business improvement.

- Select one KPI for business improvement that a service provider applying service robot system would like to improve.
- Extract service scenario for service robot that has strong relationship with the KPI to be improved. The service scenario that gives very good effect to the KPI to be solved is selected the scenario that has possibilities for further business improvement.
- Analyze the data that relates to the extracted service scenario and decide whether the execution result of service scenario gives good or bad effect. By analyzing the result of service scenario execution and KPI change, we can understand tendency of service scenario and KPI behavior.
- Consider how to modify the extracted service scenario for business improvement: One of the ways to realize business improvement is to increase service scenario execution for good result.
- Modify the service scenario according to the way of step 4 and deploy the modified service scenario. After that, do evaluation test for a pre-determined term.
- Evaluating the KPI to be improved. If the KPI is improved, the service provider permanently applies the modification of the service scenario. If KPI does not change or show bad result, the service provider discards the modification and try to consider another modification way.

3.3. Evaluation

We applied service robot system to the exhibition. The task of a service robot is guidance of our company booth in the exhibition hall.

Figure 3 shows the layout of exhibition hall. The exhibition was held in 2 days and around 40,000 visitors came. In the exhibition, there were 171 topics are shown within 7 business categories. The service robot was set only in one category booth.

It has microphone, speaker and camera. By utilizing these devices, the service robot can communicate with people through verbal way. The service robot also has wheel and it can move around the exhibition hall, however, due to safety issue, we do not use moving function of the service robot. The contents of guidance are concierge service. The service robot waits for the visitors at the entrance of the booth and give the information related to the exhibition contents or facilities when visitors get close to the service robot. Additionally, stage show is periodically done and the service robot make announcement when the stage show begins soon. The basic behavior of the service robot for the exhibition is as follows:

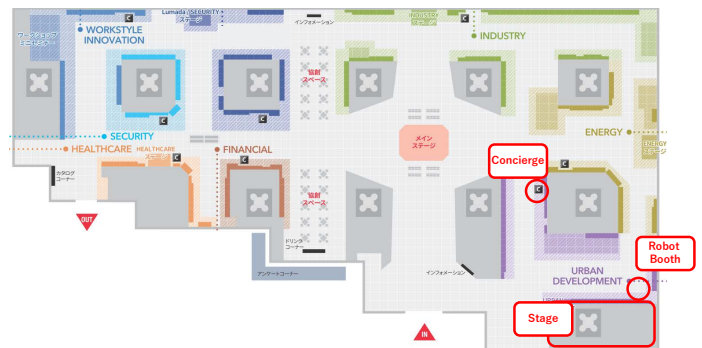


Figure 3: Architecture of Service Robot System with Data Analysis Technology

- The service robot waits for visitors' arrival at the entrance of the booth
- If there is no visitor, the service robot make announcement for its own service and try to get attention of visitors: "Hello, here we show our company's technology and service. Please come and visit our booth!"
- If the stage demo begins soon, the service robot makes announcement for the stage demo: "Our stage show starts soon. Please come and see the show!"
- When a visitor comes in front of the service robot, the service robot starts to talk and ask the visitor whether they have some questions or not.
- When the visitor asks a question, the service robot try to understand the question and reply answer. The assumed questions are questions for the exhibition contents, facility guidance and general question such as daily conversation.
- After answering a question, the service robot asks the visitor whether the visitor still have question or not.
- If the visitor has more question, the service robot continues to answer the question. If the visitor has no more question, the service robot finishes the service for the current visitor and wait for another visitor's arrival.

The service scerio of the service robot for the exhibition includes the following items:

- Announcement for the visitors
- Physical Motions that attract visitors' attention and promote natural communication
- Understanding the question by voice recognition and language processing
- Answering the question

Service robot has two kind of tasks. The important task is guidance of exhibitions and it has higher priority than other task. When the service robot detects somebody's talking, it start guidance service to the person in front of the service robot. Service robots are recently deployed in actual fields. It has basic communication function through microphone and voice recognition technology. However, if people has no experience to communicate with service robot, they might have difficulty to let them understand what people say. The communication of service robots are controlled by computer program and it has detection function of human utterance section. Because not all the people understand this detection mechanism, sometimes communication gap happens between service robot and human. Semiotics[18] is an approach for smooth mutual communication. It sometimes applies sign or non-verbal communication method. In our system design, we applied sign of service robot to let the people understand the service robot is ready for communicating with the customers. For example, a service robot aim his face to the customer or raise its hand to start communication in natural way. The data related to the previous listed items are recorded as time-series data and we used the data for business improvement analysis. For the business improvement analysis, there are two important tasks.

- Clarifying the KPI to be improved and related service scenario
- Making data formatting for data analysis of service scenario related data

Several KPI is suitable for evaluating the effect of exhibition. In this case, we focused on the understanding contents of exhibition by visitors. We set up one service robot in exhibition booth. It is difficult to measure how much the visitor understands exhibition contents, we chose the number of question and contents of question as a KPI that reflects visitors' interest to the exhibition. We classified the contents of question by judging how much the question relates to the exhibitions and monitored how the service robot behaves according to the service scenario. we recorded the data of service robot action for each visitor.



Figure 4: Structure of Concierge Service Data for Exhibition

Figure 4 shows the architecture of service robot action data for exhibitions. The data is classified into two categories. One is behavior with visitors, and the other is behavior without visitors. To separate time-series data of the service robot for each visitor, we used voice recognition result with specific words to indicate service start and end for a visitor. For example, when the service robot starts to ask question, the service robot says, "Are there any question?". In the exhibition, the service robot serves to visitors for two hours and 25 minutes for 37 visitors. After we prepare data for each visitor, we extract the feature values that represents the data. In this evaluation, we chose the following items:

- The number of speaking of the service robot
- The number of voice recognition of the service robot (that means the number of speaking of a visitor)
- The number of motion that the service robot executes.
- The number of repeated questions

- Service time for each visitor

When the service robot cannot understand the meaning of the visitor's question, the service robot asks the visitor to repeat the same question once more. We calculated the feature values shown above, then we classified the result of service scenario for each visitor. We considered the best result of service scenario is directly related to the exhibition content because the purpose of service robot system for the exhibition is to let visitors understand the exhibition contents (that is, our company's technology and service). However, service robots are just now in the beginning stage and start to be deployed in actual world. Therefore, some visitors do not know what to ask to the service robot. In such case, visitors make the question for facilities such as "Where is a rest room?" or start daily life conversation. We inspected all the data for 37 visitors and classified the result of service scenario into the following four categories, then we give the label for the data for each visitor to represent the category.

- Answer the question for the exhibition contents
- Answer the question for facility information (such as restroom or exit)
- Start daily conversation
- Cannot understand the visitor's question and does not answer anything

Smaller number means visitors have higher interest for the exhibition contents. After data classification, we did cluster analysis with K-means method. We analyzed the clustering result with changing the number of clusters. When we set the number of clusters three, we confirmed the following tendencies.

- Cluster 1 (data of 22 visitors belongs to): In this class, the service time is relatively short, few visitors repeated their questions. The service robot answers a few questions.
- Cluster 2 (data of three visitors belongs to): Service time is quite long. The visitor repeated question and asked so many questions, and the service robot answered all the related questions.
- Cluster 3 (data of 12 visitors belongs to): The service time is middle. In this class, visitors keep silent and the service robot try to make the visitor do some questions. However, visitors seem to have no idea and do not know what to be asked. Finally, the service robot regarded as the visitor has no question and ends the service for this visitor.

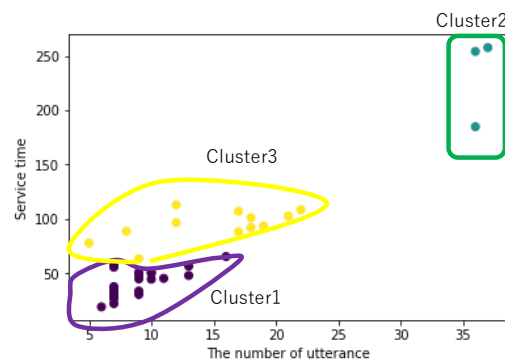


Figure 5: Clustering Result of Exhibition Data

Figure 5 shows the result of clustering. In Figure 4, service time and the number of utterance of the service robot are chosen as the

evaluation axis that most reflects the clustering result. In cluster 1, visitors and the service robot make smooth conversation. In some case, a visitor repeated the question, however, in most case, a visitor asked the exhibition contents and the service robot properly answered the question. In cluster 2, visitor seems to have some knowledge for service robots, and try to figure out what kind of questions. In this case, the service robot recognizes the visitors' question with high accuracy and the visitor and the service robot can talk long time. In cluster 3, the service robot cannot understand the questions, or the visitor does not know what to say.

We made detail analysis of the data classified into cluster 3 and figure out two major cases.

- Case1: The service robot failed to understand what the visitor says. The visitor quitted to communicate with the service robot.
- Case2: The visitor does not know what to ask to the service robot and there exists long silence time. Even though the service robot talks to the visitors and try to make the visitor ask some questions, the visitor has no idea and conversation is terminated.

As for the number of cluster, we have tested several cases. When we set the number of cluster as 4, we confirmed the samples included in cluster 3 and cluster 4 show the similar features. Therefore, we chose 3 as the number of the cluster. In case 1, main reason is the accuracy of voice recognition and technology improvement is needed. In case 2, the issue is scenario design problem. In the evaluation experiment, the service robot just asked visitors whether the visitor has question or not. However, if we designed the service robot gives more detail information such as what kind of question is assumed, we can expect the service robots behaves to visitors more effective way. For case 2, three visitors are classified in this case. We defined service quality as the ratio of number of the visitors that the service robot properly gives service for exhibition demo. The number of whole visitor for service robot is 37, that is, the sum of the data in all the clusters. If we can modify the service scenario and give better guidance to these three visitors, we can expect 8.1% (3/37) business improvement.

This is our first trial of service scenario analysis. Through this field test, we found the following issue to be solved in the future.

- In some cases, the service robot cannot correctly understand human speaking contents. In such case, we should provide alternative communication method by remote operation by human.
- Information on voice is sometimes not correctly understood by listener. In such case, visualizing method is useful for proper information providing. Since our service robot used in this field test does not have monitor, we should consider how to associate external device and the service robot in this application.

For further service scenario improvement, it is important to set target value of selected KPI, and making evaluation with Plan-Do-Check-Action cycle. Through continuous improvement, we can figure out root causes of the problem by utilizing analytic tools such as fishbone diagrams [19].

4. Service Extension Technology with External Equipment Association

In this chapter, we explain external device association technology for service robot system to realize service extension.

4.1. Functions for External Equipment Association

Figure 6 shows function blocks of external device association technology for the service robot system. Application server provides the following functions to realize external device association for service robots.

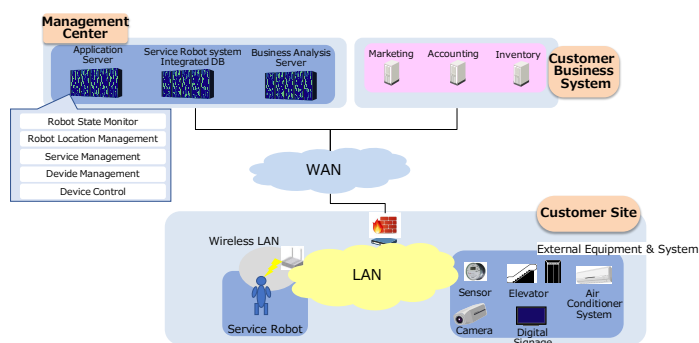


Figure 6: Functions for External Equipment Association

- **Robot State Monitor:** Collecting information on the status of various parts (motors, cameras, etc.) of service robot, then monitoring and visualizing the operating status of service robots. When some error occurs, this function notifies an alert.
- **Robot Location Management:** Managing location information of service robots based on the map information of a facility where service robots are working.
- **Service Management:** Managing services provided in facilities where service robots are installed. It includes the list of external equipment required for each service.
- **Equipment Management:** Managing equipment information about the installation location of external equipment in a facility and the conditions of use of external devices.
- **Device Control:** Provides an interface to control external equipment to realize various operation.

4.2. Usage of External Equipment with Association Condition

Service robot has several limitations on its design. Typical one is body size. If we set many peripheral equipment on a service robot, its size gets bigger and bigger. Even though we satisfy size limitation with great effort, many peripherals consume much electric power, and it requires larger battery. Therefore, service robot cannot prepare all the equipment for service realization and some equipment should be prepared as external devices. To solve this issue, we propose service management function that manages a list of devices required for services. The contents of service information are shown in Table 1.

For example, if a service robot does not have a display, an external display must be used to display visual information. The

facility guidance service using a map is a typical example, and the facility guidance service can be provided only when the external display is available for a service robot.

Table 1: Service Information

Item name	Description
Service Identifier	Unique identifier of a service
Device List	Device list that is essential for a service providing (ex. camera, display, etc.)

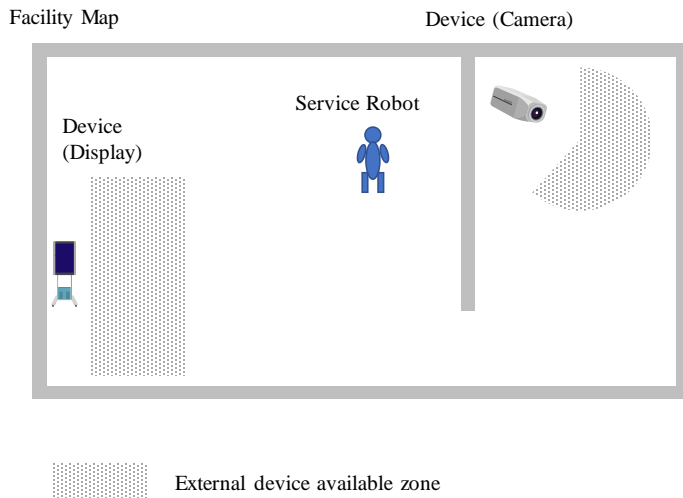


Figure 7: Example of external device availability

When a service robot starts to provide service, the service robot must check whether all the external devices necessary for the service are available or not in advance. The availability of external device depends on the position of external devices and a service robot. Also, we should consider whether the aimed devices are free and not used by other service robot. Figure 7 illustrates the positional relationship between a service robot and external devices. In this example, two external devices, a display and a surveillance camera, are installed in a facility. Taking a display as an example, for a service robot to use the display for facility information guidance service to end users such as facility users, the service robot must be close to the display. In addition, since the display has a display surface, the service robot needs to take care so that the display surface of the display can be seen by the end user. Based on this assumption, the available zone of the display is shown as the shaded area in Figure 7. The available zone of the camera differs depending on the service used, but if the recording area of the camera needs to include EMIEW, the available zone of the camera is the fan-shaped shaded area on the upper right of Figure 7. In either case, it is not enough to simply consider the physical distance between the service robot and the external devices.

Table 2 is a list of information managed by equipment management function. The external device has X and Y

coordinates on the map created by using a laser range finder or the like as the position information of the device in the facility. In addition, information of available zone of external devices is stored as a list of parameters that express the range of area. Using these parameters, a service robot realizes extended service with external devices by the following procedure shown in Figure 8.

Table 2: Item List of Device Management

Item name	Description
Device Identifier	Unique identifier of a device
Device Location	Device location in a facility: X and Y coordination in a map
Device Availability Parameter List	Device list that is essential for a service providing (ex. camera, display, etc.)

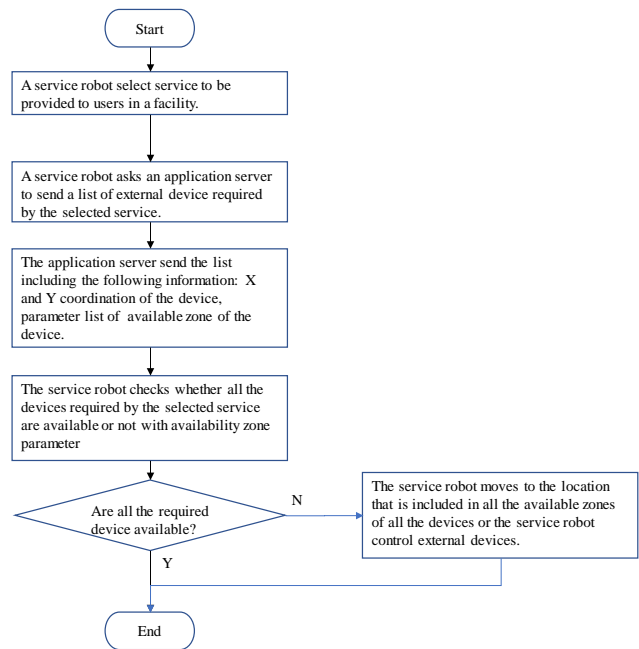


Figure 8: Flow chart of device availability check

4.3. Evaluation

In order to measure the effect of the proposed method, consider a concrete application example in a business field. As one of the typical application examples, let us consider a product and facility guidance service in a commercial facility. A typical commercial facility in Japan opens at 10 am and closes at 9:00 pm. A service robot provides shoppers with on-site facility guidance and product explanation services for 11 hours when commercial facilities are open. Figure 9 shows the timetable for commercial facilities.

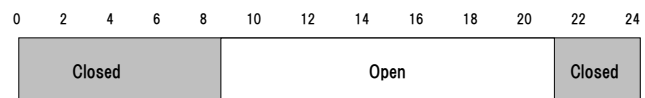


Figure 9: Timetable of the Shopping Mall

Figure 10 shows the layout of the shopping mall. A shopping mall consists of an information center that guides visitors to the facility and a shopping area that sells products. A service robot shall provide facility guidance using digital signage at the information center and product explanations at the shopping area.

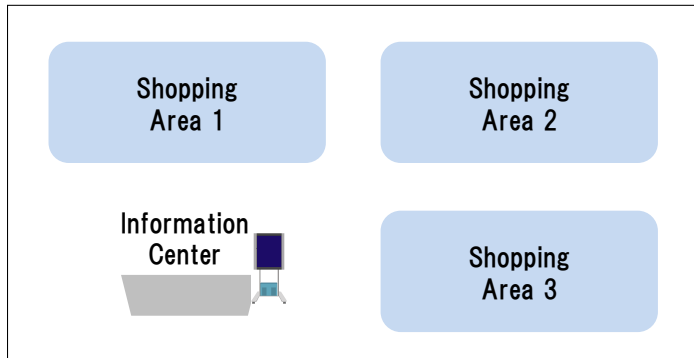


Figure 10: Layout of the shopping mall

Without the proposed method, the service robot cannot determine available services by comparing its location with the location of external devices. For this reason, shopping mall owner must set service robots at each location that provides different services. In this case, the arrangement of service robots is shown in Figure 11. In this case, 4 service robots are required, one in the information center and one in each shopping area.

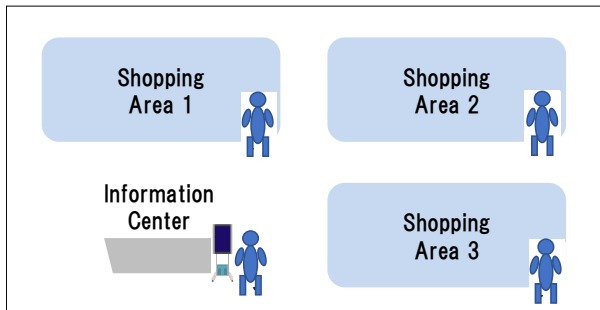


Figure 11: Arrangement of Service Robots with Conventional Technology

When applying the external device association technology, service robots can check whether its position is the information center and near the signage, or the location is the shopping center and away from the signage. Therefore, after providing facility information at the information center, it is possible to move to the shopping area and provide product information if necessary. We show the arrangement of service robots in this case in Figure 12. As for the association among service robots and signage, Simultaneous Localization And Mapping (SLAM) [20] technology is applied. The utilization area of signage is indicated in robot map managed by application server. The robot map has both signage utilization area and robot current location and application server can check whether signage is available by robot or not based on the robot location and signage condition. If robot is not at near signage, robot cannot use signage. Also, if the signage is used by another robot, signage is not available, too.

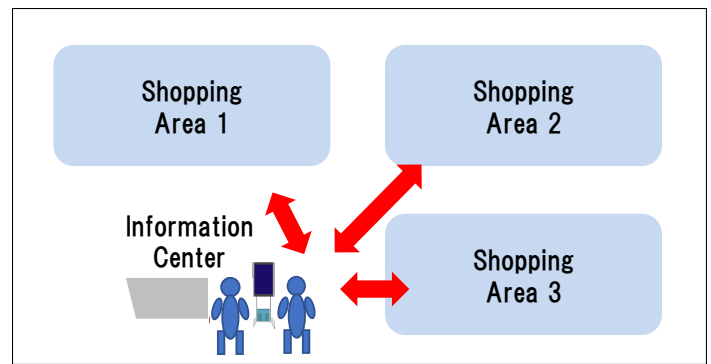


Figure 12: Arrangement of EMIEWs with Proposed Technology

The following conditions are assumed in the shopping mall.

- Time required for commercial facility guidance: 3 minutes
- Travel time between shopping area and information center: 1 minute
- Time required for product description: 5 minutes
- Visitor frequency: Once every 5 minutes

It takes 10 minutes for a service robot to move to the shopping area, explain the product, and return to the original position after guiding the commercial facility at the information center. If the customer's visit frequency is once every 5 minutes, it is possible to respond to customers without waiting for the customer with two service robots. In this case, the number of service robots installed in one commercial facility can be reduced by 50% compared to the conventional technology.

5. Conclusions

We developed data analysis technology of service robot system for business improvement and service extension with external device association with service robots. Through the evaluation by exhibition case, we found out the reason that the service robot failed to serve to the visitors, and how to modify the service scenario. 8.1% business improvement can be expected for exhibition case. In another case, we consider facility and product guidance service in a commercial facility and confirmed our proposed external device association technology can reduce the number of service robots set in a facility 50% compared with conventional technology.

Conflict of Interest

The authors declare no conflict of interest.

References

- [1] H. Yoshiuchi, T. Matsuda, and J. Dai, "Data Analysis Technology of Service Robot System for Business Improvement," *Proceedings of the 2019 5th International Conference on Robotics and Artificial Intelligence (ICRAI)*, 7-11, 2019, doi:10.1145/3373724.3373733
- [2] Z. Zhaohui, M. Xuesong, B. Xu, C. Hanghang and T. Jian, "Development of an intelligent interaction service robot using ROS," *Proceedings of 2016 IEEE Advanced Information Management, Communicates, Electronic and Automation Control Conference, IMCEC 2016, 2017*, doi:10.1109/IMCEC.2016.7867516
- [3] C. Park, S. Kang, J. Kim and J. Oh, "A study on service robot system for elder care," *2012 9th International Conference on Ubiquitous Robots and*

- Ambient Intelligence (URAI), IEEE, 2012, doi:10.1109/URAI.2012.6463070
- [4] T. Jyh-Hwa and L. Su Kuo, "The Development of the Restaurant Service Mobile Robot with a Laser Positioning System," Proceedings of the 27th Chinese Control Conference, IEEE, 2008, doi:10.1109/CHICC.2008.4605839
- [5] H. S. Ahn, J. Choi, H. Moon, M. Jang, S. S. Kwak and Y. Lim, "Social Human-Robot Interaction of Human-Care Service Robots," 2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI), 2019, doi:10.1109/HRI.2019.8673282
- [6] D. J. Ahn, J. Jeong and S. Lee, "A Novel Cloud-based Fog Computing Network Architecture for Smart Factory Big data Applications," 2018 South-Eastern European Design Automation, Computer Engineering, Computer Networks and Society Media Conference (SEEDA_CECNSM), 2018, doi:10.23919/SEEDA-CECNSM.2018.8544946
- [7] M. Dohi, K. Okada, I. Maeda, S. Fujitani and T. Fujita, "Proposal of Collaboration Safety in a Coexistence Environment of Human and Robots," 2018 IEEE International Conference on Robotics and Automation (ICRA), 2018, doi:10.1109/ICRA.2018.8460869
- [8] T. Savickas and O. Vasilecas, "Business Process Event Log Use for Activity Sequence Analysis," 2015 Open Conference of Electrical, Electronic and Information Sciences (eStream), 2015, doi:10.1109/eStream.2015.7119500
- [9] G. Park, L. Chung, L. Khan and S. Park, "A Modeling Framework for Business Process Reengineering Using Big Data Analytics and a Goal-orientation," 2017 11th International Conference on Research Challenges in Information Science (RCIS), 2017, doi: 10.1109/RCIS.2017.7956514
- [10] K. Zhou, T. Liu and L. Zhou, "Industry 4.0: Towards future industrial opportunities and challenges," 2015 12th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD), 2147-2152, 2015, doi: 10.1109/FSKD.2015.7382284.
- [11] J. A. Marvel, J. Falco and I. Marstio, "Characterizing Task-Based Human-Robot Collaboration Safety in Manufacturing," IEEE Transactions on Systems, Man, and Cybernetics: Systems, **45**(2), 260-275, 2015, doi: 10.1109/TSMC.2014.2337275.
- [12] Y. Kim and W. C. Yoon, "Generating Task-Oriented Interactions of Service Robots," IEEE Transactions on Systems, Man, and Cybernetics: Systems, **44**(8), 2014, doi: 10.1109/TSMC.2014.2298214
- [13] G. Yu, S. Hong, K. Sung and J. Seo, "A study on the risk investigation and safety of personal care robots," 2017 17th International Conference on Control, Automation and Systems (ICCAS), 904-908, 2017, doi:10.23919/ICCAS.2017.8204353
- [14] A. J. Kim, B. B. Kang, C. J. Shin, D. H. Kim and E. S. Jang, "A Study of Safety Requirements of Service Robot Appliances for Household and Similar Use," 2019 19th International Conference on Control, Automation and Systems (ICCAS), 1410-1411, 2019, doi:10.23919/ICCAS47443.2019.8971666
- [15] W. Wang, Z. Wang, F. Gao and Y. Ma, "Design of Wireless Remote Control Service Robot," 2020 IEEE International Conference on Information Technology, Big Data and Artificial Intelligence (ICIBA), 651-654, 2020, doi: 10.1109/ICIBA50161.2020.9276882
- [16] S. B. Lee and S. Hun Yoo, "Design of the companion robot interaction for supporting major tasks of the elderly," 2017 14th International Conference on Ubiquitous Robots and Ambient Intelligence (URAI), 655-659, 2017, doi: 10.1109/URAI.2017.7992695
- [17] K. Kumagai, M. Fujishima, H. Yoneda, S. Chino, S. Ueda, A. Ito, T. Ono, H. Yoshida and H. Machida, "KPI element information model (KEI Model) for ISO22400 using OPC UA, FDT, PLCopen and AutomationML," 2017 56th Annual Conference of the Society of Instrument and Control Engineers of Japan (SICE), 602-604, 2017, doi: 10.23919/SICE.2017.8105550.
- [18] W. Takano and Y. Nakamura, "Statistically integrated semiotics that enables mutual inference between linguistic and behavioral symbols for humanoid robots," 2009 IEEE International Conference on Robotics and Automation, 646-652, 2009, doi: 10.1109/ROBOT.2009.5152574.
- [19] O. A. Makinde, T. Munyai and B. I. Ramatsetse, "Establishing suitable process improvement methodologies for optimizing servicing operations in the banking industries," 2017 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM), 860-864, 2017, doi: 10.1109/IEEM.2017.8290014.
- [20] F. Hashikawa and K. Morioka, "Mobile robot navigation based on interactive SLAM with an intelligent space," 2011 8th International Conference on Ubiquitous Robots and Ambient Intelligence (URAI), 788-789, 2011, doi: 10.1109/URAI.2011.6146017.

Antecedents to Learners' Satisfaction with Serious Games: An Investigation Using Partial Least Square

Ruben Chambilla¹, Daniel Tomiuk^{2,*}, Cataldo Zuccaro², Michel Plaisent², Prosper Bernard²

¹Faculté des Sciences, Université du Québec à Montréal, Montréal, H2X 3Y7, Canada

²Faculté des Sciences de la Gestion (ESG), Université du Québec à Montréal, Montréal, H3C 3P8, Canada

ARTICLE INFO

Article history:

Received: 16 July, 2021

Accepted: 11 October, 2021

Online: 13 November, 2021

Keywords:

Serious Games

Learner's Satisfaction

Partial Least Square

ABSTRACT

Serious games are at the forefront of discussions about the future of learning. Research shows that they can help improve student motivation and knowledge transfer by making learning experiences more pleasurable. Teaching practices are increasingly enhanced or complemented by serious games; skills and knowledge are developed by recreating realistic situations allowing learners to enhance their procedural knowledge, all while having fun. Literature is scarce when it comes to identifying factors that influence learners' satisfaction when using serious games. In this paper, we present empirical results from an ongoing research project. We developed a questionnaire using factors identified in the literature and collected the responses of $n = 110$ business students from classes using serious games as a tool. Analyses were performed using partial least squares structural equation modeling (PLS-SEM). We used the following predictor variables for our model: performance and status feedback and tracking, ease of use, reliability, perceived control, instructor support, aesthetics, realism entertainment, goal clarity, immersion, and progressive challenge. Results show that sense of control, entertainment and effectiveness have a direct positive influence on learners' satisfaction while other factors influence satisfaction by mediation or as components of theoretically justified higher-order constructs. The sample size and its composition limit the generalization of results. Further studies are needed.

1. Introduction

With today's many new applications and technological innovations, learning using serious games appears as an increasingly viable alternative to more traditional teaching environments. Serious games combine both playful aspects (e.g., rules of control, scoring, challenge, and artificial characters with utilitarian contents (e.g., learning, teaching, information). [1]. Serious games are often thought of as educational games and these terms are often used synonymously [2]. Although they are games, they do not pursue entertainment, enjoyment, or fun as their primary purpose. Instead, the gaming component is used as a motivator to engage and captivate players to encourage and facilitate learning. Serious games, particularly those that simulate reality follow a learning model often qualified as *experimental* [3]. This category has two characteristics: (1) it is linked to a learning cycle based on experimentation and reflection and (2) it is based on anchoring learning in a simulated reality, building on practical experiences [3] while learning from one's mistakes without the inherent risks associated with failure in the real-world.

Virtually recreating real-world situations allows learners to develop procedural knowledge and skills by performing complex tasks without any of the real risks typically associated with bad decisions [4]. Learning objectives are all achieved while having fun. The ludic aspect is, in fact, what separates serious games from other learning environments that may also make use of information technology. Industry predictions attest to the popularity of game-based learning. According to a recent Global Game-Based Learning Market report, serious games will probably reach \$17 billion by the year 2023, surpassing 37% in terms of growth rate. This is being driven by improvements and innovation in virtual and augmented reality and artificial intelligence [5].

Nevertheless, the literature on serious games also acknowledges limits and difficulties. Primarily, good games are not easy to design. Games developed by the entertainment industry, set up exclusively for amusement, often fail. Arguably, designing successful serious games is even more difficult. Serious games must not only be entertaining; they must also help

*Corresponding Author: Daniel Tomiuk, tomiuk.daniel@uqam.ca

students achieve learning goals. Presented by this dual challenge, many serious game projects continue to fail [6].

When a learning environment is well-designed, it generates a more positive impact on the player's attitude and learning experience [7]. However, relatively few studies have tried to identify the antecedents of serious game success in the eyes of learners. Although their popularity continues to steadily increase, few studies have empirically examined the factors that affect learners' satisfaction with serious games. The purpose of our research was to identify factors affecting learner satisfaction in a serious gaming environment and to assess the relative importance of these factors.

We collected data collection from 110 university business students enrolled in courses using serious games to simulate real life managerial situations. We used SmartPLS3 software to conduct statistical analysis. Results from our analysis show that among the twelve antecedents of satisfaction identified through our literature review, sense of control, fidelity, effectiveness, and entertainment influence directly students' satisfaction with serious games. However, further analyses revealed the presence of multicollinearity among the independent variables of the model making some factors appear statistically unimportant. Treatment of our initial model was attempted to remedy the situation by considering the existence of mediating variables and theoretically justified superordinate constructs depicted as second-order factors. Our results show that while some factors affect learners' satisfaction with serious games directly, others do so indirectly through mediating factors or as part of higher-order factors and, as such, should not be neglected when designing successful games.

2. Literature Review

2.1. Benefits of Serious Games

Serious games are increasingly used in academia [8]; [9]. Several reasons explain the growing interest in serious games. First, an increasing number of educational institutions are transitioning from traditional pedagogical and didactic to more learner-centered approaches. The latter emphasizes the more active and participative role of learners [10] and serious games can create authentic (ecologically valid) and immersive environments [11]. Their interactivity helps learners in better understanding complex tasks, knowledge acquisition, and retention. Serious games also promote better concentration more than passive learning environments [12]. This represents a major transition from a traditional listening/observing only learning model to one founded on live experiences and practice. Compared to traditional classrooms where knowledge transmission is typically passive, serious games are more effective because of the interactivity they allow [13]. Moreover, the playfulness of games promotes greater interest in learning activities, engagement, focus, and enthusiasm. Serious games provide several benefits to learners [14]. Well-designed games are fun, enjoyable and increase student motivation. Linked to motivation is the concept of satisfaction. Studies have evidenced the importance of this concept in education [15], and yet, few studies have examined the antecedents of learners' satisfaction while interacting with serious games [16]; [17].

2.2. Satisfaction with Serious Games

In business, the concept of customer satisfaction is a vital component of a company's success. Research shows that employee satisfaction enhances productivity and organizational performance [18]. Customer satisfaction is positively linked to perceived service quality. Satisfaction also impacts positively on customer loyalty [19]. In the area of education, studies have demonstrated a positive relationship between learner satisfaction and both motivation and achievement [20]. Gameplay can generate strong, positive feelings in learners. These, in turn, motivate learners to continue playing (i.e., continue using the technology) and learning. Learners' satisfaction with serious games is more pronounced when positive emotions are high [21].

Satisfaction is one of the most researched concepts in the area of Information Systems (IS) [22]. When compared to other measures, user satisfaction is among the strongest of indicators employed by researchers to measure IS success [23]. According to [24], user satisfaction is "*the affective attitude towards a specific computer application by someone who interacts with the application directly*". Through a literature review, our paper first seeks to identify antecedents of serious game satisfaction from the perspective of learners. We then propose and empirically test a model to investigate the degree to which these antecedents relate to satisfaction.

2.3. Antecedents of Satisfaction with Serious Games

Twelve main factors were identified from an intensive review of the literature [25] as potential predictors of satisfaction, defined below :

Ease of Use is borrowed from the Technology Acceptance Model [26], it is defined as the degree to which a particular technology is perceived to be intuitive and simple to use.

Reliability is the degree to which the serious game worked smoothly and without problems. In Marketing, reliability has been identified as a crucial dimension of perceived service quality (i.e., the literature on SERVQUAL), while in Information Systems (IS) /Information Technology (IT) research, it refers to whether a system functions properly and can be defined as the degree to which a user possesses inherent confidence in the system [27].

Aesthetics reflects a subjective experience that is evoked by players during the process of interaction with the game interface [28]; [29]; [30], namely using sound, sight, and motion [31].

Fidelity/Realism describes the psychological and physical similarities that exist between a particular game and the environment it seeks to represent [32], the realism of a simulation [33].

Instructor assistance and support. Several studies show that the teacher plays an essential role in the success of IT integration in classrooms [34]. We define instructor assistance as the degree to which learners perceive that the instructor provided encouragement, help, and information on the use of the serious game. Instructors provide assistance at two levels, theoretical and technical [35].

Perceived control is defined as the degree to which learners feel capable of influencing and bringing about change to the world of play [36]. This can be achieved via multiple customization options such as, the pace of the game, the choice of the type of feedback the learner believes would be useful to progress more successfully through the game and even the capability of manipulating the appearance of one’s playable character [37]. Research shows that when the learner has more control over their learning, they become more involved in the game and even try more complex strategies [38] and a resulting increase in positive attitudes [39].

Progressive challenge is the degree to which the game challenges the learner with increasingly difficult problems and increasingly complex challenges. Successful game progression is built on previously acquired knowledge, abilities, and skills developed and learned while playing [40]. Successful learning ensues when players progressively adapt to and succeed in performing tasks that are increasingly challenging and complex.

Effectiveness is defined as the capacity/power of the serious game to improve the knowledge of learners [41]. Whereas conventional games seek to boost the playfulness aspect of the game, serious games seek to create learning environments that support improvements in skills training (e.g., professionals). These skills are other than those associated purely with play. Indeed, research shows that serious games can be very effective in improving professional and academic skills [42].

Status/Performance tracking and feedback is the degree to which the simulation’s progress information and feedback is perceived as constructive and helpful. Assessment and feedback are pillars of teaching effectiveness [43]. They help learners achieve the game’s objectives [44]. Performance feedback must be specific and presented at the right time [45]. Feedback includes information on the state of the player, instructor help/assistance during play, tips and clues [41]. It can take the form of visual, auditory, tactile prompts [46].

Entertainment is the degree to which the learner finds the game play pleasurable. Research defines *pleasure* as an attitude or an emotion that can satisfy intrinsic needs and produce a positive reaction [47]. The entertainment value derived from gameplay is largely determined by the quality of design of the “educational ludo” environment [48]. Entertainment value impacts positively on learners’ satisfaction.

Immersion is the ability of a game to create an immersive experience in the learner [49], the feeling of "being there" [50]. It occurs when the learner temporarily accepts the reality of a game and focuses on this reality through his involvement in the game. It reflects the degree of player engagement in the game environment [49].

Goal clarity refers to the need to clearly define what objectives must be achieved during game play. It is a fundamental principle of instructional design [51]; [52]. Studies (e.g., [53]) demonstrate that serious game environments can fail to promote concentration or rouse curiosity in learners mainly because their goals are ill-defined, ambiguous, or misunderstood. Clarity of objectives is crucial in effective game design because it serves to increase commitment and concentration [54].

3. Hypotheses

In this section, we present the hypothesized relationship between each factor and the learner’s satisfaction with the serious game (see Table 1). The questionnaire items for each measure/factor are presented in the Appendix. For the sake of parsimony, only questionnaire items that successfully passed the card-sorting exercise (item validation) are presented.

Table 1: Hypotheses of Original Model

H1	A positive and significant relationship exists between ease of use and learners’ satisfaction
H2	A positive and significant relationship exists between reliability and learners’ satisfaction
H3	A positive and significant relationship exists between aesthetics and learners’ satisfaction
H4	A positive and significant relationship exists between fidelity and learners’ satisfaction
H5	A positive and significant relationship exists between instructor assistance and support and learners’ satisfaction
H6	A positive and significant relationship exists between feeling of control and learners’ satisfaction
H7	A positive and significant relationship exists between progressive challenge and learners’ satisfaction
H8	A positive and significant relationship exists between and learners’ satisfaction
H9	A positive and significant relationship exists between performance/status tracking and feedback and learners’ satisfaction
H10	A positive and significant relationship exists between entertainment and learners’ satisfaction
H11	A positive and significant relationship exists between immersion and learners’ satisfaction
H12	A positive and significant relationship exists between goal clarity and learners’ satisfaction

4. Methodology

PLS-SEM (Partial Least Square - Structural Equation Modeling) is a multivariate statistical modelling technique increasingly being used by researchers in IS/IT, education and in other areas because of several advantages. It is often favoured when analyzing models using smaller sample sizes or data with non-normal distributions. Specifically, we used the statistical software package called SmartPLS.

4.1. Item Creation and Refinement

We generated an initial set of items and then refined and validated them before including them in a questionnaire. Steps were as follows.

- Creation or adaptation of items based on the definition of concepts [55].
- Validation of the questionnaire items using the card-sorting method [56].

The card sorting method seeks to verify, whether the items initially chosen by the researcher to be included in the questionnaire relate accurately to the concepts that (s)he is trying to measure. For this, four (4) judges were recruited: two professors from the IS/IT area and two students enrolled in a master’s program in Information Systems/Information Technology. The validation process required each judge to sort the initial questionnaire items by properly combining each one with the actual construct that item was intended to measure. When at least 3 out of 4 judges (or 75% of judges) would successfully associate a specific item to its expected construct, the item in question was deemed to accurately reflect the concept it was intended to measure and was therefore included in the questionnaire.

4.2. Data Collection and Sample Size

After validating the initial set of items using the pretest and the sort-by-card method, the items were included in an online questionnaire created using the Google Forms application. Three (3) lecturers from the business school of our university were contacted to help us recruit students. These lecturers were chosen because, in their classes, they used a serious gaming platform that simulated the typical challenges and problems encountered by managers of small- and medium-sized businesses. Over a period of approximately one (1) week, we visited the classrooms of these lecturers to explain to students the purpose of our study and invite them to participate. We printed information sheets containing our contact details and the URL for the online questionnaire. During our visits to classrooms, these sheets were provided to students. All questionnaire items were of the Likert type ranging from 1 (Strongly disagree) to 5 (Strongly agree).

Data collection was done over a period of three (3) weeks. Two emails were sent to the lecturers to remind their students to participate in the study. Participation was completely voluntary. All in all, there were 345 students enrolled in all the classes we had visited. A total response of n = 110 was received across all students enrolled in the courses, a response rate of about 33%.

Demographic information was as follows: As for age, 23- to 27-year-old students represented the largest group of participants (45% of respondents). As for gender, there was just about the same number of women as there were of men: 55 and 49 respectively while six (6) respondents preferred not to specify their gender or chose “other”. In terms of education, 37% of participants hadn’t finished their undergraduate studies. The remaining students specified that they had, at least, obtained a bachelor (or equivalent degree).

4.3. Original Model

The Partial Least Square - Structural Equation Modeling (PLS-SEM) approach allows the modelling and testing of complex cause-and-effect relationships. Two of the strengths of PLS-SEM is that the technique allows the modeling of latent variables and works well with smaller samples [57]. We used SmartPLS version 3.3.0.

Measurement Model - In SmartPLS, the measurement model was solved using the *factor* weighing scheme. Evaluating the measurement model includes an analysis of (1) reliability, (2) convergent validity, and (3) divergent validity [58]. We used www.astesj.com

Cronbach's alpha to analyze reliability and refine our scales. The cut-off value was 0.7 as recommended by [59]. Convergent validity was achieved by retaining items with loadings above or equal to 0.7. We also tentatively retained items with loadings between 0.4 and 0.7. These were later scrutinized further by examining their *composite reliability* (CR) and *average variance extracted* values (AVE). If CR was greater or equal to 0.7 and AVE was greater or equal to 0.50, the item was considered acceptable and was retained in the model [60]. All items with loadings that did not attain a minimum threshold of 0.4 were removed, one by one, starting with the item with the smallest loading. After each item removal, the model was recalculated and re-evaluated. This process continued until only acceptable items remained. The final loadings of indicators from the measurement model are presented in the Appendix.

Next, we turned towards establishing discriminant validity. For this, the *Heterotrait-Monotrait Ratios* (HTMT) were examined. A value of 0.85 or less for HTMT indicate good discriminant validity [61]. The two highest HTMT values in our model were between *effectiveness* and *entertainment* and between *entertainment* and *immersion*, with values of 0.724 and 0.703, respectively. Acceptable Cronbach alpha, CR, and AVE values presented in Table 2 all point to a good measurement model.

Table 2 - Construct Validity and Reliability

Construct	Cronbach alpha	Composite Reliability	Average Variance Extracted
Assistance and Support	0.922	0.944	0.809
Fidelity	0.922	0.945	0.810
Effectiveness	0.902	0.928	0.720
Aesthetics	0.874	0.911	0.641
Entertainment	0.859	0.914	0.780
Goal Clarity	0.839	0.886	0.611
Immersion	0.834	0.881	0.599
Satisfaction	0.832	0.900	0.749
Reliability	0.829	0.883	0.654
Progressive Challenge	0.804	0.872	0.633
Performance and Feedback	0.775	0.869	0.689
Ease of Use	0.767	0.852	0.660
Control	0.700	0.869	0.768

Structural Model - The structural model was generated using a path weighing scheme. A bootstrap of 5000 subsamples was used to generate the p-values. Path coefficients and t-values using SmartPLS3 are presented in Table 3 below which also shows whether the associated hypothesis is supported or not.

Out of all the paths specified in our original model, only four (4) were significant (at the 0.05 level). The highest path value is between entertainment and satisfaction (0.258), followed by effectiveness and satisfaction (0.205), control and satisfaction (0.187), and fidelity and satisfaction (0.164), in decreasing order. As such, only hypotheses H4, H6, H8, and H10 were supported. The paths not showing up as significant were perplexing. For instance, in our model, ease of use has no significant effect on

learners’ satisfaction. This result contradicts numerous studies which evidence the importance of perceived ease of use on users’ technology acceptance. We suspected the presence of multicollinearity in our model.

Table 3: Path Coefficients and p-values (Original Model)

Supported Hypotheses	Path Coeff.	p-value
H4 (Fidelity → Satisfaction)	0.164	0.049*
H6 (Control → Satisfaction)	0.187	0.049*
H8 (Effectiveness → Satisfaction)	0.205	0.038*
H10 (Entertainment → Satisfaction)	0.258	0.014*
Note: * = p < 0.05		
Unsupported Hypotheses	Path Coeff.	p-value
H1 (Ease of Use → Satisfaction)	0.006	0.937
H2 (Reliability → Satisfaction)	0.088	0.250
H3 (Aesthetics → Satisfaction)	-0.021	0.788
H5 (Assistance and Support → Satisfaction)	0.035	0.656
H7 (Progressive Challenge → Satisfaction)	0.098	0.250
H9 (Performance and Feedback → Satisfaction)	-0.027	0.702
H11 (Immersion → Satisfaction)	0.096	0.267
H12 (Goal Clarity → Satisfaction)	0.082	0.334

4.4. Multicollinearity, Higher-Order Constructs, and Mediation

Statistical models that contain several independent variables that are highly correlated to one another and not just with the dependent variable can cause a serious problem called multicollinearity. It makes accurate interpretation of statistical results difficult because it increases the standard errors of the coefficients. By overinflating the standard errors, multicollinearity may cause coefficients of some independent variables to appear to be non-significant. In other words, multicollinearity makes it look like the relationships between some independent variables and the dependent variables are statistically insignificant.

To test for multicollinearity, we first ran a SmartPLS model which contained only two (2) constructs: ease of use and satisfaction. As expected, and supported by literature, the path was now highly significant (path=0.408; p=0.000) which suggested the presence of multicollinearity [62] and/or mediating variables [63] in the original model (i.e., misspecified paths).

According to [62], the presence of multicollinearity among the independent factors of a model may indicate that higher-order structures are present and should be considered. Respecifying a model with these higher-order structures can effectively reduce the effects of multicollinearity but should be theoretically argued and empirically tested. In other words, higher-order constructs should not be extracted through a purely data-driven approach. They should be grounded in theoretical logic extracted from the existing literature.

In a model, a lack of statistically significant paths when these are supported by theory (or strong conjectural evidence), can suggest the presence of mediating variables that are not being considered or are incorrectly specified in that model. When mediation occurs (figure 1), an independent variable (X) has little or no statistically discernable direct effect on a dependent variable (Y). Instead, X acts on Y via an intervening variable (M). As such, the independent variable (X) affects the mediating variable (M) which, in turn, affects the dependent variable (Y). In other words, an independent variable (X) influences the dependent variable (Y), but this influence is not direct. Instead, variable X influences variable Y through some other variable (M). Mediation can be complete or partial.

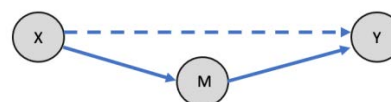


Figure 1: Mediation

4.5. Alternate Model

Following established guidelines to assess your measurement model and testing that discriminant validity falls within acceptable boundaries are very important steps in Structural Equation Modelling (SEM) but it does not guarantee that your structural model will yield significant paths [61]. Indeed, many of our initial hypotheses were not supported in our original model. Given that our analyses suggest the presence of higher-order constructs and mediation, we reworked our original model by respecifying alternative paths, paired with theoretical justification. These modifications to our original model are currently specified as propositions given that, in the coming year, we intend to explore in more detail existing literature to better substantiate these proposed alternate paths and the addition of higher-order structures.

Higher-Order Construct 1 → Platform Transparency

During our initial analyses, a significant correlation was noted between reliability and ease of use (0.434). Also, among all constructs considered in our study, these two were the only factors that were “technological” in nature.

In a now-famous article entitled “*The Computer for the 21st Century*” published in *Scientific American* in 1991 (pp 66-75), the author Mark Weiser writes that “(t)he most profound technologies are those that disappear. They weave themselves into the fabric of everyday life until they are indistinguishable from it”. The article argues that the best technologies are made or eventually become transparent to users. Technology should never be cumbersome to use. Learners’ attention should be drawn to the contents of the communication rather than to the equipment used to communicate it, and players/learners should not be concerned by technological issues. Arguably, difficult to use, unintuitive interfaces, and unreliable IS/IT platforms riddled with recurrent bugs and glitches, are likely to take away from the learner’s overall serious game experience and diminish one’s enjoyment of it, and thus, should reflect negatively on the user’s satisfaction with the serious game.

To represent this in our model, we re-specified *reliability* and *ease of use* as dimensions of a second-order construct which we named “*Platform transparency*”. We define this second-order construct as *the perception that the game’s underlying technological platform does not distract from the user’s overall gaming experience and learning objectives because of inadequate design and technical problems*.

When technological platforms supporting serious gaming are inherently unstable, unintuitive, and/or difficult to use, it is likely to impede on the overall learning experience, reducing gaming enjoyment and negatively impacting one’s sense of control over the entire experience. Previous empirical studies in IS support our assertions. There is a negative relationship between *ease of use* and one’s perceptions of control and enjoyment while using information technologies (e.g., [64]).

- P1: The more the game’s technological platform is perceived to be transparent during gaming (easy to use and reliable), the greater the learner’s entertainment while playing the serious game.
- P2: The more the game’s technological platform is perceived to be transparent during gaming (easy to use and reliable), the greater the learner’s sense of control while playing the serious game.

Higher-Order Construct 2 → Guidance

In the original model, our analysis showed non-significant relationships between satisfaction and three (3) independent variables, namely *goal clarity*, *performance and feedback*, and *assistance and support*. Rather than modeling three (3) direct paths to satisfaction, we considered the possibility that mediation was present. It could be argued that, rather than impacting on satisfaction directly, these three (3) antecedents may be factors influencing learners’ perceived effectiveness of the serious game, which would then impact satisfaction. In other words, perceived effectiveness could be a moderating variable. However, the path coefficients from *assistance and support* (path=0.132; p=0.102), from *performance and feedback* (path=0.016; p=0.843), and from *goal clarity* (path=0.130; p=0.112) to perceived *effectiveness* of the serious game were all non-significant.

Next, we reviewed the definitions of these three constructs collectively while re-examined their individual items to see whether an over-arching concept could connect them together. This led us to re-specify all three (3) concepts as dimensions of a common, second-order construct. We labeled this higher-order structure as “*Perceived guidance*”. We define this as *the degree to which the user perceives that the serious gaming experience provides a clear understanding of goals and sufficient guidance and orientation on how to achieve them by adequately informing learners of their progress and by providing information about what they are doing right/wrong*. We propose that the more effective serious games are those characterized by clear, achievable goals where learners’ progress is guided by regular, useful feedback reinforcing successful actions and results all while encouraging users to take corrective measures or try alternate strategies following unsuccessful ones. As such, we reason that *guidance* will impact satisfaction, but indirectly. The

relationship will likely be mediated by perceived *effectiveness*. The following proposition is then stipulated.

- P3: The greater the *perceived guidance* received during the serious gaming experience, the more the learner will perceive the serious game as effective.

Mediation

Based on our literature review, our original model hypothesized a significant path between the serious game’s *progressive challenge* and learner’s *satisfaction*. Statistically, this path was not significant. We then reasoned that the learner’s perceived *effectiveness* of the serious game could be mediating this relationship, making it look like *progressive challenge* has no effect on learners’ satisfaction. Arguably, games designed to be progressively challenging are also likely to be perceived as being more effective by learners given that such games would allow learners to progress in their knowledge and skill acquisition gradually. We postulate the following proposition:

- P4: The greater the perceived progressive challenge of the serious game, the more the learner will perceive the game as being effective for learning.

In our original model, the path between *immersion* and *satisfaction* was statistically non-significant. This path was dropped and, instead, we specified a path between the concepts of *immersion* and *entertainment*. Studies have recognized that games’ ability to create immersive experiences is among the most pleasurable aspects of video gameplay [65]. This brings us to the following proposition:

- P5: The greater the perceived immersion during serious gameplay, the greater the entertainment.

A path was also drawn linking *aesthetics* to *immersion*. Aesthetically superior games help in generating an immersive experience while playing suggesting that both concepts are linked. This relationship is expressed in the following proposition:

- P6: The better the aesthetics of a serious game, the greater the immersive experience perceived by the learner.

Research has also identified a link between the *entertainment* value of games and their perceived *effectiveness* [66]. Well-designed, fun games can put learners in a better mood which then positively impacts their perceived effectiveness of the serious game in supporting learning [67]. As such, we propose that a positive relationship exists between the entertainment value of a serious game and the perceived effectiveness of the game as a learning platform.

- P7: The greater the perceived entertainment value of a serious game, the greater the perceived effectiveness of that serious game.

In our original model, the path between *fidelity* and *satisfaction* was found to be positive and significant. However, the literature that pertains to the realism of games also suggests that there may be mediating effects between these two factors. It is likely that positive relationships exist between fidelity (how closely the game simulates reality) and the training effectiveness

of the serious game and immersion as well [68]. This brings us to the following propositions:

- P8: The greater the perceived fidelity of the serious game, the more the learner will perceive it as effective.
- P9: The greater the perceived fidelity of the serious game, the more the learner will perceive it as immersive.

With these propositions, we specified a new, alternative model which included the higher-order constructs described previously and added paths for mediation (Figure 2). We ran it using bootstrapping with 5000 subsamples.

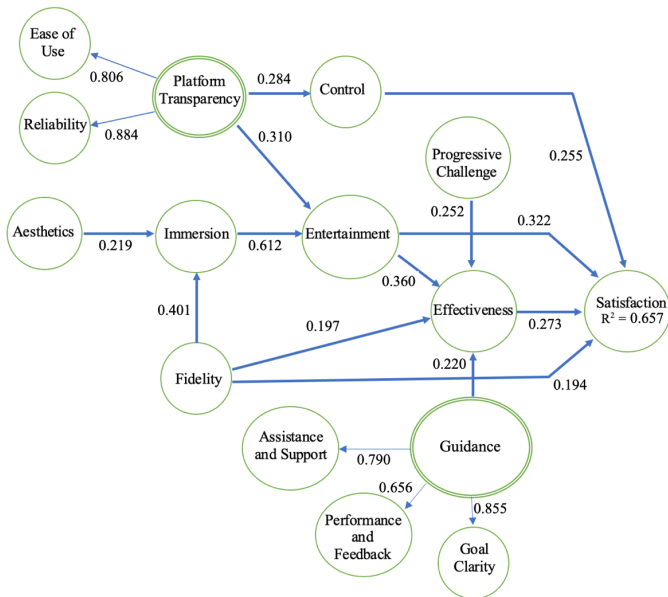


Figure 2: Structural Model (alternate model)

Tables 4 (hypotheses) and 5 (propositions) present the results of the structural model. In this respecified/alternate model, all paths are significant. This alternate model explains $R^2 = 65.7\%$ of the variance in our data on satisfaction.

Table 4: Hypotheses Path Coefficients and p-values (Alternate Model)

Hypotheses	Path Coeff.	p-value
H4 (Fidelity → Satisfaction)	0.194	0.009**
H6 (Control → Satisfaction)	0.255	0.001**
H8 (Effectiveness → Satisfaction)	0.273	0.002**
H10 (Entertainment → Satisfaction)	0.322	0.000***
Note: ** = $p < 0.01$; *** = $p < 0.001$		

Table 5: Proposition Path Coefficients and p-values (Alternate Model)

Propositions	Path Coeff	p-value
P1 (Platform Transparency → Entertainment)	0.310	0.000***
P2 (Platform Transparency → Control)	0.284	0.017*
P3 (Guidance → Effectiveness)	0.220	0.001**
P4 (Progressive Challenge → Effectiveness)	0.252	0.002**
P5 (Immersion → Entertainment)	0.612	0.000***

P6 (Aesthetics → Immersion)	0.219	0.011*
P7 (Entertainment → Effectiveness)	0.360	0.000***
P8 (Fidelity → Effectiveness)	0.197	0.017*
P9 (Fidelity → Immersion)	0.401	0.000***
Note: * = $p < 0.05$; ** = $p < 0.01$; *** = $p < 0.001$		

5. Conclusion

Through a literature review on serious games, we identified several factors which seemingly impact learners' satisfaction with serious games. Using these factors, we tested an original model using PLS-SEM. This original model specified paths going directly from each factor directly into satisfaction suggesting direct effects.

However, several paths from our original model were statistically non-significant. We respecified an alternate model supported by theoretical foundations and previous research. It included both mediation and second-order factors which helped provide a richer, more complete picture of how factors identified through our literature review relate with one another to positively impact learners' satisfaction with serious games. Every factor considered in this study was found to have a positive effect on learners' satisfaction, although not necessarily a direct effect. Several relationships were mediated through other factors while some factors are best combined into higher-order structures. Results suggest that none should be neglected when designing serious games.

Overall, the ludic component plays a central role in serious games. Entertainment value derived from playing serious games has a strong, positive impact on learner satisfaction. Although they are primarily learning tools, people perceive them as games, and as such, expect them to be designed to be fun to play. The importance of entertainment value is followed by whether the learner perceives the serious game as effective, the learners' sense of control during play, and last, by the fidelity/realism of the game. These findings suggest that to design successful serious games that satisfy learners, it is important to create a ludic environment that is fun, simulates reality well, all while making learners feel like they are in control of this environment. Perceived effectiveness of the serious game holds an important role in generating satisfaction in learners. Perceived effectiveness is a direct antecedent to satisfaction and acts as a mediating variable for several of the other variables identified in our literature review. Although other variables gained mediator status in our alternate model, none mediated a greater number of variables than did effectiveness. Our results also show that some variables do not impact satisfaction directly. Their effects on satisfaction are completely mediated through other factors. This includes aesthetics, immersion, and progressive challenge.

An important limitation of this study is the sample size. Although PLS-SEM handles small sample sizes well, small sample sizes raise questions about the generalizability of results. Another limitation of this study is that it is cross-sectional. This means that results do not take into account how time influences the relationship among variables. A factor can be more influential on satisfaction from the onset of play. Others may become more important in explaining satisfaction later on, as the learner amasses more experience. For instance, the importance of providing increasingly challenging and more varied learning

situations may become more and more important in explaining whether players continue to derive satisfaction from a serious game as their experiences and skill set grow. As such, the temporal stability of factors as antecedents of satisfaction should be investigated in future studies. Moreover, our surveys use self-reporting. These can cause respondents to inflate their opinions and responses making it hard to capture the true relationships among constructs. Interviews could be conducted in the future to validate our results. Since the respondents are all business students a generalization can be challenged for other disciplines. Finally, research is further needed to identify additional factors that may impact satisfaction with serious games and to explore the nature of the relationships that may exist between these factors.

Conflict of Interest

The authors declare no conflict of interest.

References

- [1] J. Alvarez, D. Djaouti, O. Ramouroux, *Apprendre avec les serious games?* Editions Canopé, Poitiers, France, 2016.
- [2] B. Cowan, B. Kapralos, "An overview of serious game engines and frameworks," *Intelligent Systems Reference Library*, **119**, 15-38, 2017, doi:10.1007/978-3-319-49879-9_2.
- [3] P. Cohard, "L'apprentissage dans les serious games : proposition d'une typologie," @GRH, **16**(3), 11-40, 2015, doi : 10.3917/grh.153.0011.
- [4] M.S. Calabora, A. Moraa, S. Moyab, "The future of serious games in accounting education: a delphi study," *Journal of Accounting Education*, **46**, 43-52, 2019, doi:10.1016/j.jaccedu.2018.12.004.
- [5] R. Valentine, "Metaari: game-based learning market will reach \$17 billion by 2023," *Gamesindustry.biz*, 8th August 2018, <https://www.gamesindustry.biz/articles/2018-08-08-metaari-game-based-learning-market-will-reach-usd17-billion-by-2023>.
- [6] A. Usoltsev, "Serious games: 5 gamification pitfalls to avoid," August 8 2018, <https://zenodo.org/record/2847262>
- [7] A. Messaoud, "Conception d'un jeu sérieux actif pour l'apprentissage intergénérationnel du français," Québec, Canada: Université de Laval, 2016.
- [8] S. De Freitas, "Are games effective learning tools? a review of educational games," *Educational Technology & Society*, **21**(2), 74-84, 2017, <http://www.jstor.org/stable/26388380>.
- [9] C.J. Stettina, T. Offerman, I. Sidhu, B. De Mooij, "Gaming for agility: using serious games to enable agile project & portfolio management capabilities in practice," *Engineering, Technology and Innovation (ICE) & IEEE International Technology Management Conference*, Stuttgart, 2018, doi.org/10.1109/ice.2018.8436384
- [10] R. Rodela, A. Ligtenberg, R. Bosma, "Conceptualizing serious games as a learning-based intervention in the context of natural resources and environmental governance," *Water*, **11**(245), 1-15, 2019, doi.org/10.3390/w11020245.
- [11] J. Lumsden, A. Skinner, A.T. Woods, N.S. Lawrence, M. Munafò, "The effects of gamelike features and test location on cognitive test performance and participant enjoyment," *PeerJ.*, **4**, 2-19, 2016, doi.org/10.7717/peerj.2184.
- [12] I. Bououd, R. Fafi. "Using serious games to correct french dictations: proposal for a new unity3d/nooj connector," *Association for Computational Linguistics*, 49-52, Septembre 4 2017, doi.org/10.18653/v1/W17-3808.
- [13] A. Duguet, S. Morlaix, "Le numérique à l'université : facteur explicatif des méthodes pédagogiques?," *Revue Internationale de Pédagogie de l'Enseignement Supérieur*, **34**(3), 1-18, 2018, doi.org/10.4000/ripes.1682.
- [14] J. Stewart, L. Bleumers, J. Van Looy, I. Mariën, A. All, D. Schurmans, K. Willaert, F. De Grove, A. Jacobs, G.J. Misuraca, L. Stewart, V.-L.J. Bleumers, M. Ilse, A. All, D. Schurman, K. Willaert, "The potential of digital games for empowerment and social inclusion of groups at risk of social and economic exclusion: evidence and opportunity for policy," *Institute for Prospective Technological Studies - European Commission*, **11**, 172, 2013, doi.org/10.2791/88148.
- [15] W.S. Chen, A.Y. Yao, "An empirical evaluation of critical factors influencing learner satisfaction in blended learning: a pilot study," *Universal Journal of Educational Research*, **4**(7), 1667-1671, 2016, doi:10.13189/ujer.2016.040719.
- [16] K.C.M. Bul, I.H.A. Franken, S. Van der Oord, P.M. Kato, M. Danckaerts, L.J. Vreeke, A.Willems, H.J.J. van Oers, R. van den Heuvel, R. van Slagmaat, and A. Maras. "Development and user satisfaction of plan-it commander a serious game for children with ADHD," *Games for Health Journal*, **4**(6), 502-512, 2015, doi.org/10.1089/g4h.2015.0021.
- [17] M.A. Khenissi, F. Essalmi, M. Jemni, "Comparison between serious games and learning version of existing games," *Procedia - Social and Behavioral Sciences*, **191**, 487-494, 2015, doi.org/10.1016/j.sbspro.2015.04.380.
- [18] B. Budie, R. Appel-Meulenbroek, A. Kemperman, M. Weijjs-Perree, "Employee satisfaction with the physical work environment: the importance of a need based approach," *International Journal of Strategic Property Management*, **23**(1), 36-49, 2019, doi.org/10.3846/ijspm.2019.6372.
- [19] A. Akhmedova, A.Manresa, D. Escobar Rivera, A. Bikfalvi, "Service quality in the sharing economy: a review and research agenda," *International Journal of Consumer Studies*, **45**(4), 889-910, 2021, doi.org/10.1111/ijcs.12680.
- [20] N.M. Martirosyan, D.P. Saxon, R. Wanjohi, "Student satisfaction and academic performance in armenian higher education," *American International Journal of Contemporary Research*, **4**(2), 1-5, 2014, http://www.aijcrmet.com/journals/Vol_4_No_2_February_2014/1.pdf.
- [21] E. Alsadoon, "The impact of social presence on learners' satisfaction in mobile learning," *TOJET: The Turkish Online Journal of Educational Technology*, **17**(1), 226-233, 2018, <https://files.eric.ed.gov/fulltext/EJ1165749.pdf>.
- [22] R. Vaezi, A. Mills, W.W. Chin, H. Zafar, "User satisfaction research in information systems: historical roots and approaches," *Communications of the Association of Information Systems*, **38**, 501-532, 2016, doi.org/10.17705/1CAIS.03827
- [23] W.H. DeLone, E.R. McLean, "Information systems success: the quest for the dependent variable," *Information Systems Research*, **3**(1), 60-95, 1992, doi.org/10.1287/isre.3.1.60.
- [24] W.J. Doll, G. Torkzadeh, "The measurement of end-user computing satisfaction," *MIS Quarterly*, **12**(2), 259-274, 1988, doi.org/10.2307/248851.
- [25] G. Petri, C. Gresse von Wangenheim, "How to evaluate educational games: a systematic literature review," *Journal of Universal Computer Science*, **22**(7), 992-1021, 2016, doi.org/10.1016/j.compedu.2017.01.004.
- [26] F.D. Davis, "Perceived usefulness, perceived ease of use and user acceptance of information technology," *MIS Quarterly*, **13**(3), 319-340, 1989, doi.org/10.2307/249008.
- [27] Y.H. Al-Mamary, A. Shamsuddin, A.H. Nor Aziati, "Key factors enhancing acceptance of management information systems in yemeni companies," *Journal of Business and Management Research*, **5**, 108-111, 2014, <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.1054.2814&rep=rep1&type=pdf>.
- [28] S. Dhar, V. Ordonez, T.L. Berg, "High level describable attributes for predicting aesthetics and interestingness," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, **1**, 1657-1664, 2011, doi:10.1109/CVPR.2011.5995467.
- [29] M. Pajusalu, The evaluation of user interface aesthetics, Master Thesis, Rennes/Tallinn: Tallinn University Institute of Informatics, 2012, http://www.cs.tlu.ee/teemad/get_file.php?id=202.
- [30] J.A. Abubakar, A.S. Bahrin, M.K. Ahmad, A.N. Zulkifli, "Conceptual model of game aesthetics for perceived learning in narrative games," *International Journal on Advanced Science Engineering Information Technology*, **7**(3), 993-999, 2016, doi: 0.1109/IUSER.2016.7857944.
- [31] Z. Herbert, Sight, Sound, Motion: Applied Media Aesthetics (6th ed.), Wadsworth Publishing Co. Inc., 2010.
- [32] C. Crawford, *The Art of Computer Design*, Berkeley, CA: Osborne/McGraw-Hill, 1984.
- [33] J. Pellegrino, A. Scott, "The Transition from Simulation to Game-Based Learning," *Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC)*, Dec. 6-9, 2004, <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.454.5486&rep=rep1&type=pdf>
- [34] K. McKnight, K. O'Malley, R. Ruzic, M.K. Horsley, J.J. Franey, K. Bassett, "Teaching in a digital age: how educators use technology to improve student learning," *Journal of Research on Technology in Education*, **48**(3), 194-211, 2016, doi.org/10.1080/15391523.2016.1175856.
- [35] D. Djaouti, *Serious Game Design: Considération Théoriques et*

- Techniques sur la Création de Jeux Vidéo à Vocation Utilitaire, PhD Thesis, Toulouse University, 2011.
- [36] M. Busch et al., "Personalization in Serious and Persuasive Games and Gamified Interactions," in Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play (CHI PLAY '15) ACM, 811-816, 2015, <https://uwspace.uwaterloo.ca/bitstream/handle/10012/12804/p811-busch.pdf?sequence=3>
- [37] N. Whitton, Learning and Teaching with Computer Games in Higher Education, Games-Based Learning Advancements for Multi-sensory Human Computer Interfaces: Techniques and Effective Practices, Chapter II, 18-33, 2009, doi.org/10.4018/978-1-60566-360-9.ch002
- [38] D.I. Cordova, M.R. Lepper, "Intrinsic motivation and the process of learning: beneficial effects of contextualization, personalization and choice," Journal of Educational Psychology, **88**, 715-730, 1996, doi.org/10.1037/0022-0663.88.4.715.
- [39] J.J. Vogel, D.S. Vogel, J. Cannon-Bowers, C.A. Bowers, K. Muse, M. Wright, "Computer gaming and interactive simulations for learning: a meta-analysis," Journal of Educational Computing Research, **34**(3), 229–243, 2006, doi.org/10.2190/FLHV-K4WA-WPVQ-H0YM.
- [40] N. Shin, L.M. Sutherland, C.A. Norris, E. Soloway, "Effects of game technology on elementary student learning in mathematics," British Journal of Educational Technology, **43**(4), 540-560, 2011, doi.org/10.1111/j.1467-8535.2011.01197.x.
- [41] B. Ingadottir, K. Blondal, D. Thue, S. Zoega, I. Thylen, T. Jaarsma, "Development, usability, and efficacy of a serious game to help patients learn about pain management after surgery: an evaluation study," JMIR Serious Games, **5**(2), 1-15, 2017, doi: 10.2196/games.6894.
- [42] S. Tobias, J.D. Fletcher, Computer Games and Instruction, Charlotte: IAP Information Age Publishing, 2011.
- [43] D. Drummond, A. Hadchouel, A. Tesnière, "Serious games for health: three steps forwards," Advances in Simulation, **2**(3), 1-8, 2017, doi.org/10.1186/s41077-017-0036-3.
- [44] J. Loisier, Étude sur l'Apport des Jeux Sérieux pour la Formation à Distance au Canada Francophone, Réseau d'Enseignement Francophone à Distance (REFAD), 2015, https://www.refad.ca/documents/Etude_Jeux_serieux_en_FAD.pdf.
- [45] M.D. Kickmeier-Rust, D. Albert, "Micro-adaptivity: protecting immersion in didactically adaptive digital educational games," Journal of Computer Assisted Learning, **26**(2), 95-105, 2010, doi.org/10.1111/j.1365-2729.2009.00332.x.
- [46] N. Iten, D. Petko, "Learning with serious games: is fun playing the game a predictor of learning success?," British Journal of Educational Technology, **47**(1), 151–163, 2016, doi.org/10.1111/bjet.12226.
- [47] R. Tamborini, N.D. Bowman, A. Eden, M. Grizzard, A. Organ, "Defining media enjoyment as the satisfaction of intrinsic needs," Journal of Communication, **60**, 758–777, 2010, doi.org/10.1111/j.1460-2466.2010.01513.x.
- [48] I. Marfisi-Schottman, S. George, F. Tarpin-Bernard, P. Prevot, "Comment Évaluer la Qualité d'un Learning Game Pendant sa Conception?," in Technologies de l'Information et de la Communication pour l'Enseignement, TICE2012, Lyon: France, 80-90, 2012, <https://hal.archives-ouvertes.fr/hal-00963678/document>.
- [49] G. Calleja, "Revising Immersion: A Conceptual Model for The Analysis of Digital Game Involvement," in Proceedings of the 2007 Digital Games Research Association (DiGRA) International Conference, 82-90, 2007, <http://www.digra.org/wp-content/uploads/digital-library/07312.10496.pdf>.
- [50] A. McMahan, "Immersion, Engagement and Presence: A New Method for Analyzing 3-D Video Games," in M. J. Wolf, & B. Perron (Eds), The Video Game, Theory Reader, New York: Routledge: Taylor & Francis Group, 67-87, 2003.
- [51] M. Ebrahimzadeh, S. Alavi, "Digital video games: e-learning enjoyment as a predictor of vocabulary learning," Electronic Journal of Foreign Language Teaching, **14**(2), 145–158, 2017, <http://e-flt.nus.edu.sg/v14n22017/ebrahimzadeh.pdf>.
- [52] J.G. Peery, C. Pasalar, "Designing the learning experiences in serious games: the overt and the subtle-the virtual clinic learning environment," Informatics, **5**(30), 1-14, 2018, doi.org/10.3390/informatics5030030.
- [53] M. Ma, A. Oikonomou, L.C. Jain, Serious Games and Edutainment Applications (ed. London Limited). New York: Springer-Verlag, 2011.
- [54] Y. Wang, R. Pramod, C. Sankar, P.K. Raju, "Relationships between Goal Clarity, Concentration and Learning Effectiveness When Playing Serious Games," in Proceeding of Twentieth Americas Conference on Information Systems, 2014 https://www.researchgate.net/publication/263426149_Relationships_betweeen_Goal_Clarity_Concentration_and_Learning_Effectiveness_when_Pl
- aying_Serious_Games
- [55] G.A. Churchill, "A paradigm for developing better measures of marketing constructs," Journal of Marketing Research, **16**, 64-73, 1979, doi.org/10.1177/002224377901600110.
- [56] G.C. Moore, I. Benbasat, "Development of an instrument to measure the perceptions of adopting an information technology innovation," Information Systems Research, **2**(3), 192-222, 1991, doi.org/10.1287/isre.2.3.192.
- [57] F. Ali, S.M. Rasoolimanesh, M. Sarstedt, C.M. Ringle, K. Ryu, "An Assessment of the use of partial least squares structural equation modeling (pls-sem) in hospitality research," International Journal of Contemporary Hospitality Management, **30**(1), 514-538, 2018, doi.org/10.1108/IJCHM-10-2016-0568.
- [58] Chin, W. W., Esposito-Vinzi, V., Henseler, J., & Wang, H. (2009). How to Write Up and Report PLS Analyses. Handbook of Partial Least Squares, pp 655-690
- [59] J.C. Nunnally, I.H. Bernstein, Psychometric Theory (ed. 3rd). New York: McGraw-Hill, 1994.
- [60] J.F. Hair, C.M. Ringle, M. Sarstedt, "PLS-SEM: indeed a silver bullet," Journal of Marketing Theory and Practice, **19**(2), 139-152, 2011, doi.org/10.2753/MTP1069-6679190202.
- [61] J. Henseler, C.M. Ringle, M. Sarstedt, "A new criterion for assessing discriminant validity in variance-based structural equation modeling," Journal of the Academy of Marketing Science, **43**(1), 115-135, 2015, doi.org/10.1007/s11747-014-0403-8.
- [62] C. Schwarz, A. Schwarz, W.C. Black, "Examining the impact of multicollinearity in discovering higher-order factor models," Communications of the Association for Information Systems, **34**(62), 1191-1208, 2014, doi:10.17705/ICAIS.03463.
- [63] R. Agler, P. De Boeck, "On the interpretation and use of mediation: multiple perspectives on mediation analysis," Frontiers in Psychology, **15**, 1-11, 2017, doi:10.3389/fpsyg.2017.01984.
- [64] V. Venkatesh, "Determinants of perceived ease of use: integrating control, intrinsic motivation, and emotion into the technology acceptance model," Information Systems Research, **11**(4), 342–365, 2000, doi.org/10.1287/isre.11.4.342.11872.
- [65] L. Michailidis, E. Balaguer-Ballester, X. He, "Flow and Immersion in Video Games: The Aftermath of a Conceptual Challenge, Frontiers in Psychology, 05 September 2018, doi: 10.3389/fpsyg.2018.01682.
- [66] N. Nazry, M. Nazrina, D.M. Romano, "Mood and learning in navigation-based serious games," Computers in Human Behavior, **73**, 596–604, 2017, doi.org/10.1016/j.chb.2017.03.040.
- [67] Y. Zhonggen, "A meta-analysis of use of serious games in education over a decade," International Journal of Computer Games Technology, **1**, 1-8, 2019, doi.org/10.1155/2019/4797032.
- [68] D. Rojas, B. Kapralos, S. Cristancho, K. Collins, A. Hogue, C. Conati, A. Dubrowski, "Developing effective serious games: the effect of background sound on visual fidelity perception with varying texture resolution," Studies in Health Technology and Informatics, **173**, 386-392, 2012, doi:10.3233/978-1-61499-022-2-386.

Appendix

<i>Item code</i>	<i>Aesthetics</i>	<i>Loading</i>
AES1	The simulation is visually appealing	0.827
AES2	The visual design of the simulation is attractive	0.881
AES3	I appreciated the aesthetic side of the simulation	0.841
AES4	The color contrast makes it easier to understand windows	0.758
	<i>Control</i>	

CON1	I was able to complete the simulation at my own pace	0.883
CON2	The simulation gives the user control to better manage his learning	0.870
	Ease of Use	
EOU1	The interface is intuitive	0.815
EOU2	This simulation is easy to use	0.872
EOU4	The simulation controls are logical and consistent	0.791
	Effectiveness	
EFF1	This game has helped me improve the knowledge I acquired in class	0.867
EFF2	The game asks me to put into practice what I've learned in class	0.855
EFF3	The game covers the material and concepts that are related seen in class	0.885
EFF4	I feel like I have acquired new knowledge using this simulation	0.773
EFF5	The simulation allowed me to put into practice all the skills I acquired in class	0.859
	Entertainment	
ENT1	The simulation was fun	0.884
ENT2	I would like to play this game again	0.886
ENT3	The time I spent playing the game was pleasant	0.880
	Fidelity	
FID1	The simulation does reflect real situations	0.935
FID2	There was a very good physical and psychological similarity between the simulation and the environment it sought to represent	0.885
FID3	The simulation encourages learners to use the same cognitive processes they will have to use in the real-world	0.911
FID4	The simulation is faithful to the types of problems encountered in the real-world	0.867
	Goal Clarity	

CLA1	The game's instructions were precise and obvious	0.831
CLA2	Objectives were clearly defined	0.858
CLA3	I understood the learning objectives throughout the simulation	0.733
CLA4	From start to finish, I understood the task to be accomplished in the simulation	0.694
CLA5	The purpose of the game was always clear while I was playing	0.782
	Immersion	
IMM1	While playing this game, I did not feel the time pass	0.835
IMM2	I was very involved in this simulation when I was playing it	0.746
IMM3	I lost track of time while playing the game	0.822
IMM4	When I was playing the game, I became very focused on what I was doing	0.732
IMM5	The simulation had me focused on the task at hand	0.724
	Instructor assistance and support	
INS1	The teacher's mastery with simulation was useful for learners	0.900
INS2	The competence of the teacher with this technology was appreciated by the students	0.891
INS3	The teacher was able to help learners to use technology	0.909
INS4	The teacher's knowledge of the simulation was useful to us	0.898
	Performance / status and feedback	
PERF3	The feedback offered during the simulation was constructive	0.875
PERF4	The information offered by the simulation helped me to make better decisions for the rest of the game	0.768
PERF6	The simulation kept me informed of my status (e.g., score, feedback)	0.837
	Progressive challenge	

CHA4	The increasingly difficult situations presented by the game stimulated my desire to learn	0.699
CHA6	This simulation offers an adequate amount of difficulty to stimulate the interest of the learner	0.795
CHA7	The simulation was able to sharpen my skills allowing me to meet increasingly complex challenges	0.880
CHA8	This game allows different levels of challenges, well suited for different learners	0.797
Reliability		
REL1	The simulation worked well	0.871
REL2	This technology was reliable	0.820
REL3	I found that the technology worked as expected	0.852
REL4	I had technical problems (reverse item)	0.704
Satisfaction		
SAT1	In general, I consider that this simulation was well designed	0.809
SAT2	I enjoyed using this game for learning	0.887
SAT3	Overall, I am satisfied with this simulation	0.899

Hiragana and Katakana minutiae based recognition system

Piotr Szymkowski^{*1}, Khalid Saeed¹, Nobuyuki Nishiuchi²

¹Bialystok University of Technology, Department of Computer Science, Bialystok, 15-351, Poland

²Tokyo Metropolitan University, Department of Computer Science, Tokyo, 191-0065, Japan

ARTICLE INFO

Article history:

Received: 12 July, 2021

Accepted: 12 October, 2021

Online: 20 November, 2021

Keywords:

Handwriting recognition

Biometric recognition

Minutiae

ABSTRACT

The Japanese writing system is unique due to the number of characters employed and the methods used to write words. It consists of three different 'alphabets', which may result in the methods used to process Latin script not being sufficient to obtain satisfactory results when attempting to apply them to a recognition of the Japanese script. The authors present an algorithm based on minutiae, i.e., feature points, to recognise the hiragana and katakana characters. A method using image processing algorithms is compared with a method using a neural network for the purpose of automating this process. Based on the distribution and type of minutiae, vectors of features have been created to recognise 96 different characters. The authors conducted a study showing the effect of the chosen segmentation method on the accuracy of the character recognition. The proposed solution has achieved a maximum accuracy at the level of 65.2%.

1 Introduction

This paper is an extension of work originally presented in The International Conference on Graphics and Signal Processing [1]. The Japanese writing system consists of three types of characters: kanji, hiragana and katakana. The hiragana and katakana characters are found only in Japan. They are syllabaries, which means that a single written character stands for a single spoken syllable. Hiragana is mainly used to write prefixes and suffixes of the Japanese words, particles and words of Japanese origin which for some reason are not written in the kanji characters. Hiragana is also used to write furigana, a reading aid for the difficult or unusual kanji characters. It is often found in books for children and teenagers.

Katakana is used to write first names, surnames and words of foreign origin. It is also applied for onomatopoeia (mimetic words), which frequently occur in Japanese. The two previously discussed syllabaries each contain 46 characters each.

The last group of characters are the kanji characters. These characters have been borrowed from the Chinese language. The number of the kanji characters in Japan is estimated to be approximately 50,000, but the average Japanese needs circa 3,000 characters for the use on an everyday basis. The kanji characters are used to write the subject of words. Unlike the hiragana and katakana, the kanji characters have different manners of reading. The way they are read depends on whether they stand alone or are accompanied by

other characters. One kanji character can have even more than 10 different ways of reading.

The quantity and complexity of the writing system in relation to the 26 characters of the Latin alphabet results in the fact that recognition methods and algorithms for the purpose of writing should be subject to a change.

2 State of the Art

For the purpose of the recognition of the Japanese writing system, publications on the on-line recognition are easier to find than the off-line ones. The key publications presenting a cross-section of ongoing studies on this very subject matter will be introduced below.

2.1 Deep Convolutional Recurrent Network

The first solution presented uses deep convolutional recurrent neural networks in order to recognise single lines of text written in Japanese [2]. The network used consists of three layers: a convolutional feature extractor, recurrent layers and transcription layers [3]. Prior to the recognition itself, a preprocessing of the images is required. This consists of normalising, i.e., homogenising the samples, and binarizing them with the use of the Otsu's method. The authors use the TUAT Kondate database. Another aspect presented in this paper constitute the artificially generated samples, developed on the basis

*Corresponding Author: Piotr Szymkowski, prszymkowski@gmail.com

of the previously mentioned database. The model presented here shows a character recognition rate at the level of 96.35% when using only the data from the database, and at the level of 98.05% when using also the artificially generated data. The generation of these data is based on the use of the local and global elastic distortions.

2.2 Extended Peripheral Direction Contributivity

The recognition method of the Japanese writing system presented by the authors [4] adopts the Extended Peripheral Direction Contributivity method. It is one of the more commonly used directional features. The feature vector is developed in the following manner. First of all, the image is scanned from their peripheral to inner part in the eight directions. Then, on each black pixel of the contour of the first, second and third line encountered in each direction, the run-length l_i in each of the eight directions is calculated and the direction contributivity d_i ($i = 1, 2, \dots, 8$) is calculated for the pixel using (1) [5].

$$d_i = \frac{l_i}{\sqrt{\sum_{j=1}^8 l_j^2}} \quad (1)$$

Thirdly, the image is divided into several zones in each direction of the peripheral scan. Finally, a feature vector is obtained by recording the directional components for each zone, in each direction and in each order encountered. The information that can be obtained by the aforementioned method are as follows [6]:

- A quantity of lines and their orientation,
- A position of the lines,
- The relations between the lines.

2.3 Address Recognition

The algorithm presented, represents the use of a system that reads information from the letters and postal shipments concerning the information on a destination address [7]. For the purpose of reducing the information to be extracted, a priori knowledge about the structure of the Japanese address has been used. The introduced system is based on five processes: segmentation, segment combination, character recognition, character combination and path search. The system is supported by a glossary of the possible addresses.

3 Image Processing Algorithms

The algorithm presented by the authors uses many image processing methods. Some of the currently used solutions use a priori knowledge, which also reduces a probability of misclassification of the characters. The principles of operation of the most important of them will be specified below.

3.1 Morphological Closing

One of the essential operations performed on the images are the morphological ones. As their name suggests, their application results in a change in the appearance of the result image. In case of

the algorithm presented by the authors, the image on which these operations are employed is black and white. The two most basic morphological operations are the image erosion and image dilation. The result of applying the filter is shown in Figure 1. Their application is equivalent to the minimum and maximum of the statistical filters [8].

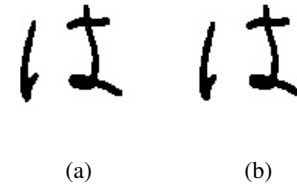


Figure 1: Original image (a) and image after morphological closing applied (b).

The closing filter consists in the performance of the dilation and erosion operations on the image using the same structural element. As a result of this operation, objects lying close to each other are merged, their edges are filled and smoothed.

3.2 Bilateral Filter

The bilateral filter is one of the blurring filters used to remove the initial noise present in the images. In contrast to the Gauss filter, it is not linear, which means that it blurs the image while retaining transparency of the edges [9].

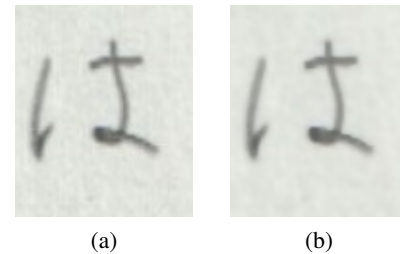


Figure 2: Original image (a) and image after bilateral filter application (b).

This is achieved by varying the blur rate, which depends on the size of the structural object and the value of the local gradient. It results in stronger blurring of homogenous areas and lesser blurring in the regions of sharp local changes, i.e. edges, what can be seen at Figure 2. The bilateral filter is presented in (2).

$$I_{filter} = \frac{1}{W_p} \sum_{x_i \in \Omega} I(x_i) f_r(\|I(x_i) - I(x)\|) g_s(\|x_i - x\|) \quad (2)$$

The normalising formula is presented in (3).

$$W_p = \sum_{x_i \in \Omega} f_r(\|I(x_i) - I(x)\|) g_s(\|x_i - x\|) \quad (3)$$

where:

- I_{filter} – result image,
- I – input image,

- x - coordinates of pixel under analysis,
- Ω - structural object,
- f_r - smoothing factor (intensity),
- g_s - smoothing factor (distance).

3.3 Modified K3M Algorithm

Thinning is one of the operations used to simplify the description of the shapes of objects in the image. The operation is performed on a binarized image. As shown in Figure 3, all lines in the image are narrowed until they reach a width of 1 pixel. The modified K3M algorithm used in this publication has been created by one of the authors [10].

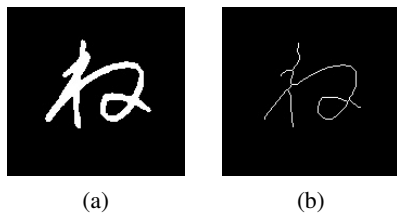


Figure 3: Original image (a) and sign thinned using Modified K3M algorithm (b).

3.4 Otsu's Method

The Otsu's method is one of the binarization techniques, which uses the thresholding what means that the algorithm uses a predetermined threshold to segment the entire input image. An example of how the method works is shown in Figure 4. It is a technique based on the discriminant analysis [11]. The optimal threshold is chosen based on one of the presented in (4) equivalent criterion functions.

$$\eta = \frac{\delta_B^2}{\delta_T^2}, \lambda = \frac{\delta_B^2}{\delta_W^2}, \kappa = \frac{\delta_T^2}{\delta_W^2} \quad (4)$$

Where:

- $\delta_B^2 = P_0(\mu_0 - \mu_T)^2 + P_1(\mu_1 - \mu_T)^2 = P_0P_1(\mu_0 - \mu_1)^2$ - between class variation,
- $\delta_W^2 = P_0\delta_0^2 + P_1\delta_1^2$ - within class variation,
- $\delta_T^2 = \sum_{i=0}^{L-1} (i - \mu_T)^2 P_i$ - global variation,
- $\delta_T^2 = \delta_W^2 + \delta_B^2$.

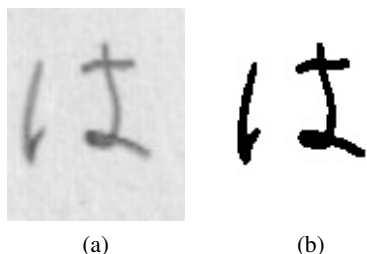


Figure 4: Original image (a) and image after Otsu method application (b).

4 Algorithm Description

To achieve the results, a Python application was prepared by the authors using keras library to apply artificial neural networks and sklearn to apply the SVM algorithm. The presented proprietary solution uses two artificial intelligence algorithms for the purpose of automating the recognition process of Japanese handwriting, therefore two types of feature vectors are created. It is similar to another used in the biometric algorithms, more specifically while recognising a user by their fingerprints. In order to create feature vectors for the character recognition, characteristic points - minutiae - are used. Figure 5 shows a diagram demonstrating how the presented solution works.

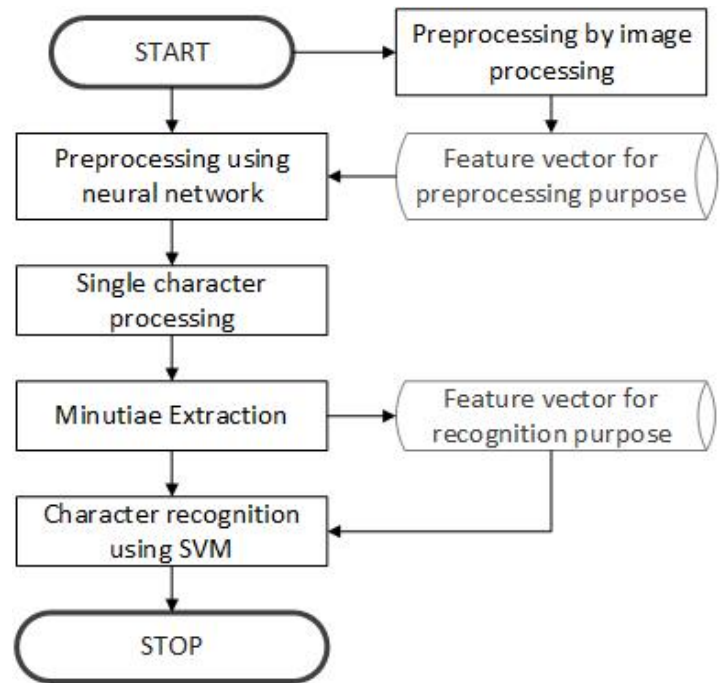


Figure 5: Algorithm diagram

4.1 Preprocessing Using Image Processing

This step is followed by the creation of feature vectors further used during the learning of the neural network in order to perform an automatic segmentation. The steps of this process are shown at Figure 6. The first stage consists in the use a bilateral filter [12]. As a consequence of the aforementioned, some of the noise that appeared in the image is eliminated. However, this has failed to affect the edge condition.

The next step has been to apply the background subtraction algorithm. It is usually applied when segmenting a series of photographs or films. After subtracting two consecutive photographs or films frames, the result image shows changes in the image, movement of animals, objects. In case of the presented solution, an image on which an averaging filter has been applied using a structural object with a diameter of 5 pixels is subtracted from the original image upon removing noise. The result image shows the differences in the image frequency.

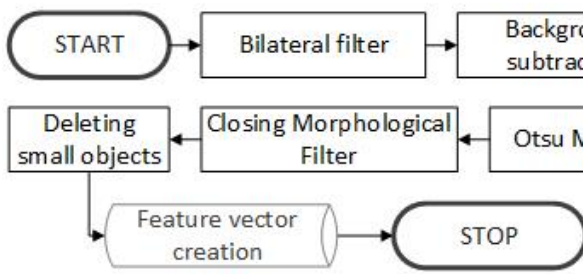


Figure 6: Image processing approach to preprocessing.

Binarization is the third preprocessing step. For this purpose, the algorithm according to the Otsu’s method has been used. As in the event of the biometric algorithms, the green channel has been used for this purpose. Upon completing this stage, the image shows a partially or fully completed character. Then a closing morphological filter is applied. After this operation, the edges in the image are smoothed and some of the noise is deleted. Finally, the noise is removed by applying the last algorithm.

An algorithm has been developed that calculates the area of each object appearing in the image. If it is less than a threshold one, the object is deleted (Figure 12).

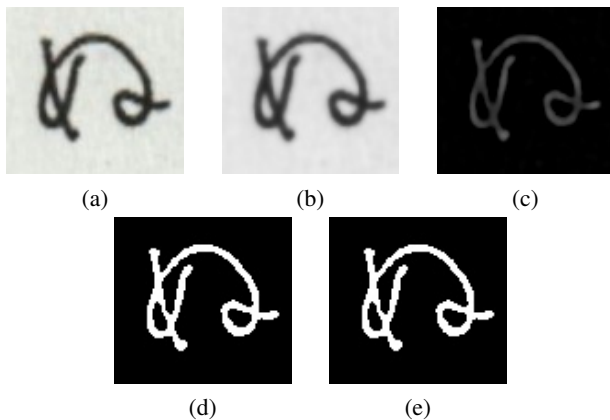


Figure 7: Character after each operation of preprocessing: (a) original image, (b) image after bilateral filter application, (c) image after background subtraction, (d) binarized image and (e) image after small object deletion and closing filter application.

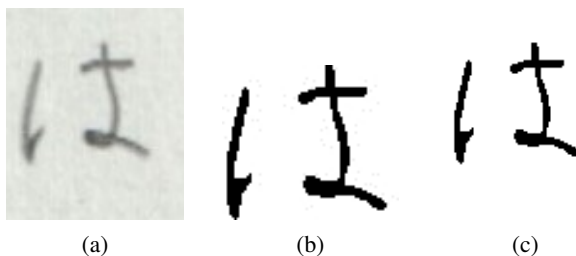


Figure 8: Sample (a), sample preprocessed by hand (b) and sample preprocessed using image processing approach.

The final step is to create a feature vector based on the output image. A square structural object with a length and a width of 11 pixels is used. Each vector possesses information about the central pixel, its 120 neighbouring ones, as well as information about its

classification upon preprocessing (1 - background, 0 - object). The result of these operations can be seen in Figure 8.

4.2 Segmentation Using a Neural Network

A neural network has been used to automate the preprocessing process. The large illustrations (2900 x 1200 pixels) representing Japanese writing system have been used to develop the sufficient feature vectors. The principle of operation of this part of the algorithm is presented in Figure 9.

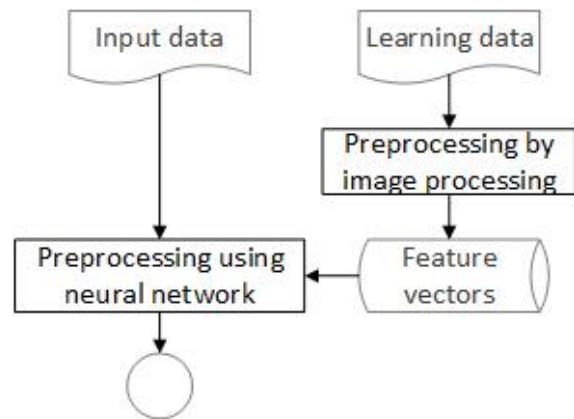


Figure 9: Neural network approach to preprocessing.

A sequential neural network model with the ReLU activation function has been employed for the research purposes. The neural network consists of four layers: an input layer, two hidden layers, and an output layer. The input layer has 363 nodes, the first hidden layer possesses 120 nodes, the second hidden layer has 30 nodes, and the output layer possesses 1 node. As the output, information is received as to whether the pixel under study (located in the centre of the structural object) should be classified as a background or as an object [13]. A comparison of manually processed samples and those processed using neural networks is presented in Figure 10.

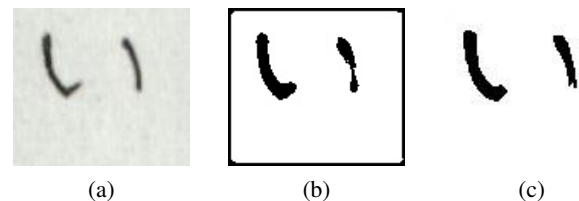


Figure 10: Original sample (a), sample after image processing based preprocessing (b) and sample preprocessed using neural network (c).

4.3 Processing of Single Characters

Upon preprocessing the image containing the object of interest, you obtain the image that requires the further adjustments to extract the features of interest. The next steps of single character processing are shown in Figure 11. The first step is to remove the part of the image that does not carry any information. This is done with the help of the vertical and horizontal histograms. Commencing from the ends of the image, you search for the maximum value of a distance from the edge, where the number of pixels of the object is 0.

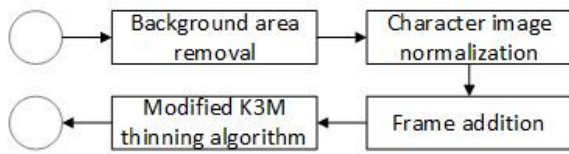


Figure 11: Single character processing

4.4 Development of Feature Vectors and Character Recognition

The next step is the normalisation, which consists in standardising the size of the images [14]. The operations are performed on the images to adjust their size to 62 by 62 pixels. A one-pixel frame is then added to enable to perform further operations on the images (Figure 12). The last operation performed at this stage of the image processing is skeletonization using a modified K3M algorithm. The choice of the skeletonization algorithm has been made using an experimental method.

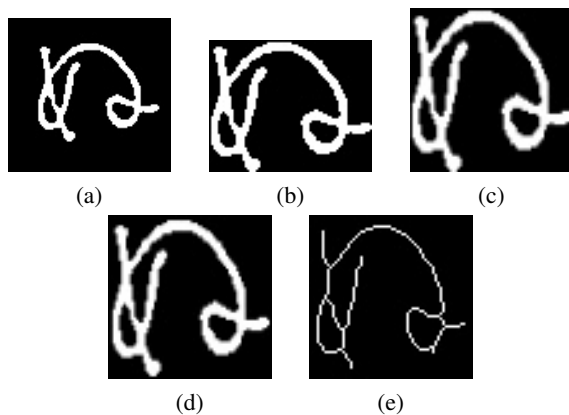


Figure 12: Character after each operation of single character processing step: (a) original image, (b) image after background area removal, (c) image resized, (d) one pixel-wide frame added and (e) image after thinning algorithm application.

While determining the feature vectors, the authors have used the methods applied in the biometric recognition using fingerprints. They use minutiae - characteristic points present in the fingerprints (Figure 13). Among all available types, three types have been selected:

- Ends and beginnings of lines,
- Bifurcations,
- Intersections or Trifurcations.

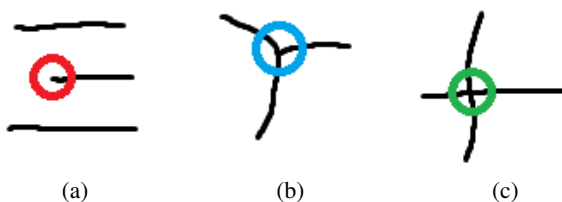


Figure 13: Minutiae types: (a) beginning or ending, (b) bifurcation and (c) trifurcation.

For the purpose of developing a feature vector, the image has been divided into 16 equal areas (16 by 16 pixels) and the sums of each type of minutiae are stored in the appropriate places. Each feature vector contains the feature values and the decision class.

On the basis of an in-house database containing approximately 50 samples for each of the 96 hiragana and katakana characters, experiments have been conducted to test the ability of the SVM (Support Vector Machine) to recognise the characters [15].

5 Results

During preprocessing, a prediction match of 86.20% with the intended result has been obtained. It is a relatively high value, but even the smallest deviation from an exemplar one can result in, and has resulted in, the creation of the additional minutiae or a loss of some of them in the image. In the course of the study, the division of samples into a learning set and a test set has been carried out in the ratio 3:2.

Two experiments have been conducted during the character recognition based on the feature vectors of minutiae occurrences. The first experiment has constituted in testing the effect of the choice of preprocessing method on the results of the recognition of the Japanese characters. Three trials have been carried out: data upon a manual segmentation has been used within the first trial, data preprocessed with the use of an algorithm based on the image processing algorithms has been applied in the second trial, and data preprocessed using artificial intelligence has been used in the third one. Table 1 shows the results of the first experiment. Depending on the technique used, the results vary. It can be noted that the more complicated the method processing of a single character is complicated, the lower recognition results were obtained. The highest precision was achieved when using the authors' algorithm is 46.6% when applying image processing approach to preprocessing.

Table 1: First experiment results.

Circumstances	Precision of recognition
Hand preprocessed samples	50,3%
Samples preprocessed using image processing	46,6%
Samples preprocessed using neural network	42,6%

Table 2: Second experiment results.

Circumstances	Precision
Text preprocessed using neural network with added artificial samples	63,3%
Text preprocessed using image processing approach with added artificial samples	65,2%

In the second experiment, a 50 percent chance of development of a descendant from it has occurred. The newly created sample has been subject to twisting and stretching by random parameters. It has been further adapted to the requirements of the algorithm and a new feature vector has been developed on the basis of it. In both

cases the results have been significantly improved by more than 15%, reaching 63.3% when using a neural network and 65.2% when using image processing for preprocessing, as presented in the Table 2.

Table 3: Solution comparison.

Authors	Approach	Database	Declared precision (Best)
Author's Solution	Minutia based	Own database	65,2%
N.T. Ly, C.T. Nguyen, M.Nakagawa [2]	Deep Convolutional Recurrent Network	TUAT Kondate	98,05%
M.Mori, T.Wakahara, K.Ogura [4]	Extended Peripheral Direction Contrubutivity	ETL9B	95%
X. Xia, X. Yu, W. Liu, et al. [7]	Over segmentation strategy	Own database	83.4%
S. Tsuruoka, M. Hattori, M. F. b. A. Kadir, et al. [16]	Modified quadratic discriminate function and users dictionaries	ETL9B	82.5%

6 Conclusions

The presented results enable to conclude that the introduced method is not sufficient to allow the unaided character recognition at a high enough level. One of the reasons may be the use of a non-uniform database. The obtained result (62.5%) is the lowest among those compared 3. However, those results allow to suppose that combining this method with other will allow to obtain better solution. The future papers of the authors will focus on the development of the presented solution in order to obtain a system allowing for the off-line text recognition with the use of the hybrid methods.

Conflict of Interest The authors declare no conflict of interest.

Acknowledgment This work was supported by grant PROM from the Polish National Agency for Academic Exchange NAWA, supported by the Bialystok University of Technology and partially funded with resources for research by the Ministry of Science and Higher Education in Poland, under Grant WZ/WI-IIT/4/2020.

References

- [1] P. Szymkowski, K. Saeed, N. Nishiuchi, "SVM Based Hiragana and Katakana Recognition Algorithm with Neural Network Based Segmentation," in Proceedings of the 2020 The 4th International Conference on Graphics and Signal Processing, ICGSP 2020, 51–55, Association for Computing Machinery, New York, NY, USA, 2020, doi:10.1145/3406971.3406978.
- [2] N. T. Ly, C. T. Nguyen, M. Nakagawa, "Training an End-to-End Model for Off-line Handwritten Japanese Text Recognition by Generated Synthetic Patterns," in 2018 16th International Conference on Frontiers in Handwriting Recognition (ICFHR), 74–79, 2018, doi:10.1109/ICFHR-2018.2018.00022.
- [3] S. D. Budiwati, J. Haryatno, E. M. Dharma, "Japanese character (Kana) pattern recognition application using neural network," in Proceedings of the 2011 International Conference on Electrical Engineering and Informatics, 1–6, 2011, doi:10.1109/ICEEL.2011.6021648.
- [4] T. Wakahara, Y. Kimura, M. Sano, "Handwritten Japanese character recognition using adaptive normalization by global affine transformation," in Proceedings of Sixth International Conference on Document Analysis and Recognition, 424–428, 2001, doi:10.1109/ICDAR.2001.953825.
- [5] M. Mori, T. Wakahara, K. Ogura, "Measures for structural and global shape description in handwritten Kanji character recognition," 1998, doi:10.1117/12.304621.
- [6] B. Lyu, R. Akama, H. Tomiyama, L. Meng, "The Early Japanese Books Text Line Segmentation base on Image Processing and Deep Learning," in 2019 International Conference on Advanced Mechatronic Systems (ICAMechS), 299–304, 2019, doi:10.1109/ICAMechS.2019.8861597.
- [7] X. Xia, X. Yu, W. Liu, C. Zhang, J. Sun, S. Naoi, "An Efficient off-Line Handwritten Japanese Address Recognition System," in 2019 International Conference on Document Analysis and Recognition (ICDAR), 714–719, 2019, doi:10.1109/ICDAR.2019.00119.
- [8] M. Nixon, A. S. Aguado, Feature Extraction and Image Processing for Computer Vision, Third Edition, Academic Press, Inc., USA, 3rd edition, 2012.
- [9] M. Buczkowski, P. Szymkowski, K. Saeed, "Segmentation of Microscope Erythrocyte Images by CNN-Enhanced Algorithms," Sensors, **21**(5), 2021, doi:10.3390/s21051720.
- [10] M. Tabedzki, K. Saeed, A. Szczepański, "A modified K3M thinning algorithm," International Journal of Applied Mathematics and Computer Science, **26**(2), 439–450, 2016, doi:10.1515/amcs-2016-0031.
- [11] C. Solomon, T. Breckon, Fundamentals of Digital Image Processing: A Practical Approach with Examples in Matlab, Wiley Publishing, 1st edition, 2011.
- [12] Y. Seki, "Collection of Online and Offline Handwritten Japanese Characters and Handwriting Classification Using the Data," in 2020 17th International Conference on Frontiers in Handwriting Recognition (ICFHR), 264–269, 2020, doi:10.1109/ICFHR2020.2020.00056.
- [13] N.-T. Ly, C.-T. Nguyen, K.-C. Nguyen, M. Nakagawa, "Deep Convolutional Recurrent Network for Segmentation-Free Offline Handwritten Japanese Text Recognition," in 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR), volume 07, 5–9, 2017, doi:10.1109/ICDAR.2017.357.
- [14] I. Uddin, D. A. Ramli, A. Khan, J. I. Bangash, N. Fayyaz, A. Khan, M. Kundi, "Benchmark Pashto Handwritten Character Dataset and Pashto Object Character Recognition (OCR) Using Deep Neural Network with Rule Activation Function," Complexity, **2021**, 6669672, 2021, doi:10.1155/2021/6669672.
- [15] K. Nguyen, C. Nguyen, M. Nakagawa, "A Segmentation Method of Single- and Multiple-Touching Characters in Offline Handwritten Japanese Text Recognition," IEICE Transactions on Information and Systems, **E100.D**, 2962–2972, 2017, doi:10.1587/transinf.2017EDP7225.
- [16] S. Tsuruoka, M. Hattori, M. F. b. A. Kadir, T. Takano, H. Kawanaka, H. Takase, Y. Miyake, "Personal Dictionaries for Handwritten Character Recognition Using Characters Written by a Similar Writer," in 2010 12th International Conference on Frontiers in Handwriting Recognition, 599–604, 2010, doi:10.1109/ICFHR.2010.98.

Modelling and Testing Services with Continuous Time SRML

Ning Yu, Martin Wirsing*

Institute for Informatics, LMU Munich, 80538 Munich, Germany

ARTICLE INFO

Article history:

Received: 31 July, 2021

Accepted: 26 October, 2021

Online: 20 November, 2021

Keywords:

continuous time SRML

Service-Oriented Hybrid
Doubly Labelled Transition
System

differential equations

traffic control system

ABSTRACT

The *SENSORIA* Reference Modeling Language (SRML) aims at modelling composite services at a high level. Continuous time SRML extends SRML so that it can model services whose components can perform both discrete processes and continuous processes. In order to show how continuous time SRML is applied, in this paper, we systematically introduce our study on continuous time SRML in the following approach: First we introduce the theoretical foundation of continuous time SRML, the Service-Oriented Hybrid Doubly Labeled Transition Systems. This is the semantic domain over which continuous time SRML is defined and interpreted. Then we design a case study of a traffic control system. In the case study, we illustrate the scenario of the system, explain the continuous time SRML model of the system, and show how to transform the model to a kind of Deterministic Finite Automata that can be used for testing and verification. Finally, we show our idea of testing our model with the IBM WebSphere Process Server. With this approach, we come to a conclusion that continuous time SRML can be used to model certain systems in the real-world, and can be tested with proper tools.

1 Introduction

Service-Oriented Computing (SOC) [1] is a paradigm for distributed computing, in which computation units are abstracted as services. The *SENSORIA* Reference Modelling Language (SRML) [2, 3] has been developed in the IST-FET integrated project *SENSORIA* [4], and it is a prototype domain-specific language for abstracting service-oriented systems at a high level abstraction. Hybrid systems arise in embedded control when discrete components are coupled with continuous components. Continuous time SRML [5] is an extension of SRML, and it aims at abstracting service-oriented systems in which hybrid systems are embedded. That is: the discrete components and continuous components are abstracted as components of services. Being different from SRML, continuous time SRML can be used to formally model services, whose components can perform both discrete processes and continuous processes that can be describe by differential equations.

In [6], we introduced the application of continuous time SRML through the case study of a Traffic Control System. In this paper, we give a novel systematic presentation of our approach from the semantic domain of continuous time SRML, to the case study, and finally to the test environment. Continuous time SRML is interpreted over Service-Oriented Hybrid Doubly Labeled Transition Systems [5] (SO-HL²TS). This kind of transition systems extends Service-Oriented Doubly Labeled Transition Systems [7] (SO-L²TS), so

that each state of a SO-L²TS is lifted to a function mapping a time interval to an infinite trace of state, over which the differential equations in continuous time SRML can be interpreted. Our approach includes two parts: specifying the system with continuous time SRML modules, and transforming the modules to a hybrid automaton-like Deterministic Finite Automaton (DFA). In order to test the system module, we combine IBM Websphere Integration Designer [8] (WID) and Matlab to form the test environment. WID is used to implement the continuous time SRML module, and Matlab [9] is used to compute the variables controlled by differential equations. Thus, the main contribution of this paper is to provide a way to test or verify the complex systems (like the Traffic Control System) that is specified with continuous time SRML.

This paper is arranged as follows: In Section 2, we introduce the related work such as hybrid automata, model checking tools and the early research of this topic. In Section 3, we introduce HL²TS, the semantic domain of continuous time SRML. In Section 4, we show the case study of the traffic control system. In Section 5, we introduce the testing environment and show the idea of implementing the traffic control system. In Section 6, we make a conclusion of this paper and introduce our future work.

*Corresponding Author: Martin Wirsing, Oettingenstraße 67,80538 Munich (Germany), +49 89 2180 9154 & wirsing@pst.ifi.lmu.de

2 Related work

SRML is brought forward in [7], and it is used to model important properties of services and the business conversations within and among services in SOC. A service specified with SRML can only perform discrete activities. Continuous time SRML extends the semantic domain and syntax of SRML, such that it could express differential equations which represents the continuous process performed by services. This essential property of continuous time SRML enables it be able to model the combination of service-oriented systems and hybrid systems. We will introduce this in Section 3.

Hybrid automata [10] are formal models for hybrid systems. They abstract hybrid systems with graphs and workable representations. A hybrid automaton consists of a set of variables, a control graph, some essential conditions and a set of events. These compositions match the compositions of a service component that is specified with continuous time SRML. Thus, a service specified by continuous time SRML can be transformed to a hybrid automata-like DFA, this will be introduced in Section 4. Hybrid automata are interpreted over timed transition systems, while continuous time SRML is interpreted over SO- HL^2TS . Timed transition systems abstract the duration of a continuous process as a parameter of a transition. In this way, it is simple and clear for abstracting continuous processes. SO- HL^2TS specify continuous processes as traces of infinite states with definite initial states and final states. This enables us to know the values of all the variable of an arbitrary state.

In [11], the author provides another way for modelling hybrid systems. In that paper, continuous processes are also interpreted over traces of states. This is the original idea from which the traces of states in SO- HL^2TS come. But in [11], discrete processes are not modelled between traces of states. In [12], the author also provides rule schemata for verifying the hybrid systems. Moreover, based on this work, a model checking tools for hybrid systems, KeYmaera X, is brought forward in [13]. In [14], provides an approach for simulating embedded systems with Real-Time Maude. Compared with our approach for testing, this can be another way to test the traffic control model, but the interactions of a service would not be implemented as easy and explicate as that in WID.

3 Continuous time SRML and its semantic domain

Continuous time SRML models composite services, whose elementary components may involve in continuous time executions. The services in continuous time SRML are represented with service modules. A service module is composed of service components, external service interfaces and internal wires linking between them. In Figure 1, we see the compositions of a service module. In the service module, SC is a service component, EX-P is an external service interface (provides-interface), every EX-R is an external service interface (requires-interface), and every IW is an internal wire.

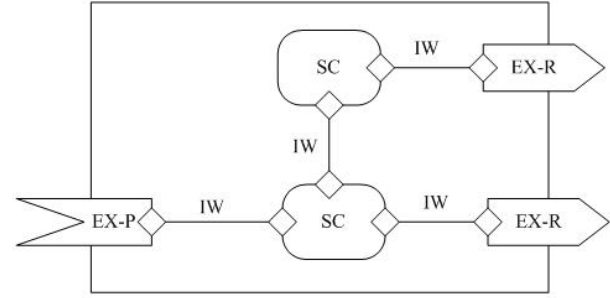


Figure 1: A service module of continuous time SRML

In a service module of continuous time SRML, a service component is specified with a business role, an external service interface is specified with a business protocol, and an internal wire is specified with a connector. The full syntax of continuous time SRML can be found in [5], and they are interpreted over SO- HL^2TS s.

SO- HL^2TS s are extensions of HL^2TS s [5, 15] such that they enrich HL^2TS s with service-oriented features. HL^2TS s extends the L^2TS [16] in that it defines a set of functions Σ . These functions map from the real number domain to the state domain of the HL^2TS and can be used to interpret the continuous processes described by differential equations. The following definition of HL^2TS is from [5, 15].

Definition 1 (Hybrid Doubly Labeled Transition System) A Hybrid Doubly Labeled Transition System (HL^2TS) is a tuple

$$\langle S, s_0, \Sigma, Act, R, AP, L \rangle$$

where:

- S is a set of states;
- $s_0 \in S$ is the initial state;
- Σ is a set of functions and for every function $\sigma \in \Sigma$ there is $r_\sigma \in \mathbb{R}$ and $r_\sigma \geq 0$, such that $\sigma : [0, r_\sigma] \rightarrow S$;
- Act is a finite set of observable actions;
- $R \subseteq \{\sigma(r_\sigma) : \sigma \in \Sigma\} \times 2^{Act} \times \{\sigma(0) : \sigma \in \Sigma\}$ is the transition relation. A transition $\sigma(r_\sigma) \xrightarrow{\alpha} \sigma'(0)$ with $\alpha \subset Act$, is denoted by $(\sigma(r_\sigma), \alpha, \sigma'(0)) \in R$;
- AP is a set of atomic propositions;
- $L : S \rightarrow 2^{AP}$ is a labeling function such that $L(s)$ is the subset of all atomic propositions that are true in state s .

In Definition 1, the set S of state can be finite or infinite; for every $\sigma \in \Sigma$, the real number r_σ is unique, and function σ on the interval $[0, r_\sigma]$ represents the prolongation of states in the continuous process with the duration r_σ ; function L is a state labeling function that defines which propositions are true in each state, and these propositions are from the fixed set of atomic propositions AP .

In order to add service-oriented features to HL^2TS s, we refine HL^2TS s with Service-Oriented Transition Systems (SO-TSs)

[2]. SO-TSs are used that models the communications in service-oriented systems during the computing processes. In a SO-TS, states and transitions are labelled with certain sub-sets of events that appear in the computing processes. With such refinements, we can obtain SO-HL²TSs. The following definition of SO-HL²TSs is from [5, 15].

Definition 2 (Service-Oriented Hybrid L²TS) A Service-Oriented Hybrid L²TS (SO-HL²TS) that refines a SO-TS $\langle S, \rightarrow, s_0, G \rangle$ is a tuple

$$\langle S, s_0, \Sigma, Act, R, AP, L, TIME \rangle$$

where:

- $Act = \{e! : e \in E\} \cup \{e_j : e \in E\} \cup \{e? : e \in E\} \cup \{e_i : e \in E\}$;
- $R \subseteq \{\sigma(r_\sigma) : \sigma \in \Sigma\} \times 2^{Act} \times \{\sigma(0) : \sigma \in \Sigma\}$ is such that,
 $\sigma(r_\sigma) \rightarrow \sigma'(0)$ iff $(\sigma(r_\sigma), \alpha, \sigma'(0)) \in R$ for some $\alpha \in 2^{Act}$
- $AP = Act \cup PLG$;
- For every $\sigma \in \Sigma$ and every $\zeta \in [0, r_\sigma]$, $L : S \rightarrow 2^{AP}$ is such that:

$$L(\sigma(\zeta)) = \{e!, e_j, e?, e_i : e \in HST^{\sigma(\zeta)}\} \cup PLG^{\sigma(\zeta)}$$

In Definition 2, the set *Act* is refined with the sets of events that are published (denoted by $e!$), delivered (denoted by e_j), executed (denoted by $e?$) and discarded (denoted by e_i) respectively, E is the set of all the events. The set *AP* of atomic propositions is labelled with the set of actions and the set of pledges that associate with every interaction (the propositions that hold during the time when the interaction is valid). Every state is labelled with the actions whose associating events are from the history of the events of that state, and the pledges that hold in the state. We also specify the time in each state with the function *TIME*. The full definition can be seen in [5]

When applying a function $\sigma \in \Sigma$ to its interval $[0, r_\sigma]$, we obtain a infinite trace of states $\sigma_0, \dots, \sigma_{r_\sigma}$. By connecting traces of states with possible actions, paths of states are formed. Because a trace of states represent the evolution of a continuous process, actions which represent the discrete processes can only take place at the initial and final states of each trace. Thus, a transition in a path only exists between the last state of one trace and the first state of another trace. This can be seen in the definition of transition relation *R* in Definition 2. The following definition of traces and paths of states is from [5, 15].

Definition 3 (Paths of SO-HL²TSs) Given a SO-HL²TS $m = \langle S, s_0, \Sigma, Act, R, AP, L, TIME, \Pi \rangle$, paths of m are defined as follows:

- For every $\sigma \in \Sigma$, $\sigma(0) \dots \sigma(r_\sigma)$ denotes the trace of states of σ where $\sigma(0)$ is the first state of the trace and $\sigma(r_\sigma)$ is the last state of the trace. $\sigma(0) \dots \sigma(r_\sigma)$ includes all the states in the set $\{\sigma(\xi) | 0 \leq \xi \leq r_\sigma\}$, which can be finite or infinite.
- $\rho = (\sigma_1(0) \dots \sigma_1(r_{\sigma_1}), \sigma_2(0) \dots \sigma_2(r_{\sigma_2}), \dots)$ is a path of m if there exists an $\alpha \in 2^{Act}$ such that for every $\sigma_i(r_{\sigma_i})$ and $\sigma_{i+1}(0)$ with $i \in \mathbb{N}$, there exists an $\alpha \in 2^{Act}$ such that $(\sigma_i(r_{\sigma_i}), \alpha, \sigma_{i+1}(0)) \in R$;

- The states in a path are ordered lexicographically such that, for every $i, j = 1, 2, \dots$ and $\zeta \in [0, r_{\sigma_i}], \xi \in [0, r_{\sigma_j}]$, there is $\sigma_i(\zeta) < \sigma_j(\xi)$ iff either $i < j$, or $i = j$ and $\zeta < \xi$;
- For the states in a path, there is:

- for every $i, j = 1, 2, \dots$ and $\zeta \in [0, r_{\sigma_i}], \xi \in [0, r_{\sigma_j}]$, there is $TIME^{\sigma_i(\zeta)} \leq TIME^{\sigma_j(\xi)}$ if $i < j$, and $TIME^{\sigma_i(\zeta)} < TIME^{\sigma_j(\xi)}$ if $i = j$ and $\zeta < \xi$;

In particular, if $TIME^{\sigma_i(r_{\sigma_i})} < TIME^{\sigma_{i+1}(0)}$ then the transition $(\sigma_i(r_{\sigma_i}), \alpha, \sigma_{i+1}(0))$ takes time, otherwise if $TIME^{\sigma_i(r_{\sigma_i})} = TIME^{\sigma_{i+1}(0)}$ the transition is executed in zero time;

- A path ρ terminates if it is a finite sequence $\sigma_1(0) \dots \sigma_1(r_{\sigma_1}), \dots, \sigma_n(0) \dots \sigma_n(r_{\sigma_n})$. In such case the first state of the trace $\sigma_1(0)$ is denoted by *firstp* and the last state $\sigma_n(r_{\sigma_n})$ is denoted by *lastp*;
- The concatenation of traces $\rho_1 = (\sigma_1(0) \dots \sigma_1(r_{\sigma_1}), \sigma_2(0) \dots \sigma_2(r_{\sigma_2}), \dots)$ and $\rho_2 = (\zeta_1(0) \dots \zeta_1(r_{\zeta_1}), \zeta_2(0) \dots \zeta_2(r_{\zeta_2}), \dots)$, denoted by $\rho_1 \circ \rho_2$, is defined as follows:
 - $\rho_1 \circ \rho_2 = (\sigma_1(0) \dots \sigma_1(r_{\sigma_1}), \dots, \sigma_n(0) \dots \sigma_n(r_{\sigma_n}), \zeta_1(0) \dots \zeta_1(r_{\zeta_1}), \dots)$ iff ρ_1 terminates at $\sigma_n(r_{\sigma_n})$ and $(\sigma_n(r_{\sigma_n}), \alpha, \zeta_0(0)) \in R$;
 - $\rho_1 \circ \rho_2 = \rho_1$ iff ρ_1 does not terminate;
 - $\rho_1 \circ \rho_2$ is not defined in other cases;
- λ is an empty hybrid trace such that for any arbitrary hybrid trace ρ , $\rho \circ \lambda = \lambda \circ \rho = \rho$.

The full semantics of continuous time SRML is defined based on Definition 2 and Definition 3, and can be found in [5].

4 Traffic control system and the transformation process

In order to show the application of continuous time SRML, we design a case study of a traffic control system to control the self-driving trains. A traffic control system includes a local traffic control center, a built-in intelligent driving system on a self-driving train, and a monitoring center. Suppose depending on the data obtained from the sensors on the train and the built-in intelligent driving system, the self-driving train can move on their own in most of the driving time without communicating to the outside. But when approaching train stations, they need to get additional information from the local traffic control center, e.g. to know if the station is free, in order to guarantee safe driving. In addition, the monitoring center needs to know the status of the train in some condition, so that they can react in time if the train is not driving safely. In our case study, we show in a traffic control system, how the intelligent driving system on a self-driving train communicates with the local traffic control center and the monitoring center when the train is approaching a train station. The control scheme of the intelligent driving system is adapted from the model of a train system in [17], which extends ETCS [18] level 3 for rail-road crossings.

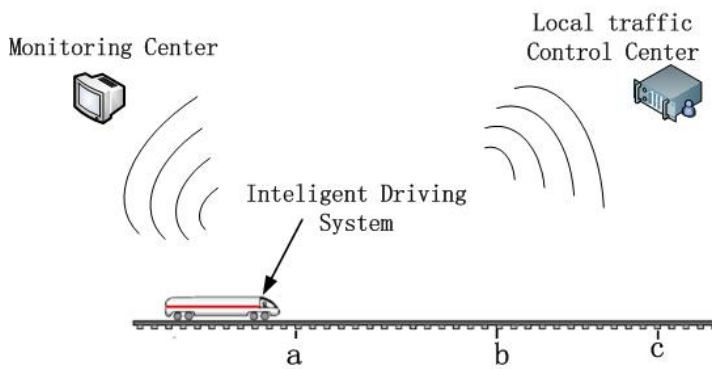


Figure 2: A brief scenario of the traffic control system

```

BUSINESS ROLE IDS is
INTERACTIONS
s&r VControl
_receive newSpeed: speed
rcv VMoveOn
_receive newSpeed: speed
s&r Brake
_receive dec: acc
r&s BrakeInfo
_send currPos: position
currTime: time
ORCHESTRATION
var C: position, B: acc, dC: acc, dC: speed, v1: speed, v1: speed, t: time
init C=0^v1=100^v1=0^B=0^t=0 ^dC=v1^dC=B
transition Nego
trigger VControl_receive
guard dC=v1
effect dC=v1
transition Corr
trigger Brake_receive
guard dC=v1
effect dC=B
transition Cont
trigger VMoveOn_receive
guard dC=v1
effect dC=v1
transition Stop
trigger BrakeInfo_receive
guard dC=B
effect dC=B
    
```

Figure 3: Business Role *IDS*

Figure 2 shows a brief view of the scenario. In this figure, we assume that the train station is at point *c*, and on reaching point *a*, the train moves with a constant speed v_0 . When the train reaches point *a*, as it needs to move at a lower speed, the intelligent driving system sends a request to the local traffic control center, to ask with which speed the train can move on. Based on the condition of the current traffic condition, the local traffic control center replies with the value of the proper speed. On receiving this message, the intelligent driving system sets a constant speed v_1 of this value to the train. If the local traffic control center finds the traffic is lighter and the train can move with a higher speed, it sends a message to the intelligent driving system with the value of the higher speed. On receiving this message, if the train hasn't passed point *b*, the intelligent driving system sends another request to the local traffic control center, to ask with which lower speed it can move on in the future. Before the train reaches point *b*, this process can repeat for several times. When the train reaches point *b*, the intelligent driving

system sends a request to the local traffic control center, to ask with which deceleration the train can brake. Based on the condition of the current traffic condition, the local traffic control center replies with the value of the proper deceleration. On receiving this message, the intelligent driving system sets the deceleration of the train to the required deceleration. After the train brakes, if the intelligent driving system receives the request from the monitoring center, it sends the current condition (position, speed, deceleration, and so on) of the train to the monitoring center.

We abstract the whole system with continuous time SRML as a traffic control service module, and abstract the intelligent Driving System with business role *IDS*, the Local Traffic Control Center with business protocol *TCC*, and the Monitoring Center with business protocol *MOC*. In this paper, we concentrate on the computational reality of a service module, and ignore the interactions caused by communications with the outside parties at the service interfaces, thus, we only make study to the business role *IDS*.

A business role includes a set of interactions and an orchestration in which these interactions are organized. An interaction consists of a type (e.g. *s&r*) and a name (e.g. *VControl*) and attributes of event types (e.g. *_receive*) and parameters (e.g. *newSpeed: speed*). With this structure, we can specify events that associate with an interaction (e.g. *VControl_receive*), and parameters that associate with an event (e.g. *VControl_receive.newSpeed*). An orchestration of a business role includes a set of variables that is local to the service component abstracted by the business role, a set of initial values of the variables that identifies the initial state of the service component, and a set of transitions that model the activities performed by the service component. A transition consists of a *trigger*, a *guard* and an *effect*. A *trigger* is an event or a state condition that specifies the condition for the transition to occur. A *guard* is a condition that identifies the states in which the transition can occur. An *effect* is the proposition that is true in the state that follows the state specified by the guard condition, and represents the effect of the transition. In the declaration of a transition, we don't model the actions caused by the transition (e.g. the assignments of local variables) because they are not considered in the verification procedure.

Figure 3 shows the specification of business role *IDS*. *IDS* contains four interactions and four transition rules. Interaction *VControl* denotes the communication between *IDS* and *TCC*: when the train reaches point *a*, the *IDS* sends a request to *TCC* to ask for a lower speed, and it can receive the reply from *TCC* with parameter *newSpeed* when the train is between point *a* and point *b*, the value of *newSpeed* is then assigned to local variable v_1 ; interaction *VMoveOn* denotes the communication between *IDS* and *TCC*: when the train is between point *a* and point *b* and the traffic is lighter, *IDS* can receive a message from *TCC* with parameter *newSpeed*, indicating that the train can move with a higher speed, the value of *newSpeed* is then assigned to local variable v_0 ; interaction *Brake* denotes the communication between *IDS* and *TCC*: when the train reaches point *b*, the *IDS* sends a request to *TCC* to ask for a braking deceleration, and it can receive the reply from *TCC* with parameter *dec* when the train is after point *b*, the value of *dec* is then assigned to local variable *B*; interaction *BrakeInfo* denotes the communication between *IDS* and *MOC*: when the train is between point *b* and point *c*, and it brakes, *IDS* can receive a request from *MOC* asking the current position of the train and the current time, and it replies to

MOC with parameter $currPos$ and $currTime$, which take the values of local variable C and t respectively. Transition $Nego$ describes that the train changes its speed from v_0 to v_1 when it receives the slow down message from the local traffic control center; transition $Cont$ describes that the train changes its speed from v_1 to v_0 when it receives the speed up message from the local traffic control center; transition $Corr$ describes that the train brakes with the deceleration B when it receives the braking message from the local traffic control center; transition $Stop$ describes that the train receives the request message from the monitoring center, and replies with its current position and the current time. In business roles, the continuous time executions are specified in the form of differential equations. (e.g. $dC = v_0$ in Figure 3). In general, in continuous time SRML, the n th derivative of variable a to time variable t , $\frac{\partial^n a}{\partial t^n}$, is denoted by $d^n a$. The formal semantics of business roles can be found in [5].

Next we take transition $Nego$ as an example to illustrate how a transition in IDS can be interpreted over a SO-HL²TSs: Suppose a SO-HL²TSs $h = \langle S, s_0, \Sigma, Act, R, AP, L, TIME, \Pi \rangle$ satisfies IDS (the satisfaction of a business role over a SO-HL²TS can be found in [5]), $\sigma(r_\sigma) \xrightarrow{\alpha} \sigma'(0)$ is a transition in h ($\sigma, \sigma' \in \Sigma$ and $\alpha \in Act$), if $VControl_receive \in \alpha$ and $\sigma(r_\sigma) \models dC = v_0$, then $\sigma'(0) \dots \sigma'(r_{\sigma'}) \models dC = v_1$. Particularly, formula $dC = v_1$ represents the differential equation $\frac{\partial C}{\partial t} = v_1$, where t denotes the continuous time variable. Such a formula expresses that variable C changes according to the differential equation $\frac{\partial C}{\partial t} = v_1$ along the trace $\sigma'(0) \dots \sigma'(r_{\sigma'})$ of h .

In order to be able to test or verify the Traffic-Control module that is explained in this section, we show how to transform the business role IDS to a hybrid automata-like DFA. A Deterministic Finite Automata (DFA) [19] consumes a string of input symbols over a set of states, and it operates as follows: when inputting a symbol, the DFA transits from one state to another state, and it does the transitions until all input symbols have been consumed. In a business role, if we take the trigger and guard of every transition as an input symbol, and label the effect of every transition to a state, then we can construct a DFA based on the business role. According to the scenario of the traffic control system and the specification of business role IDS , we can construct the DFA shown in Figure 4.

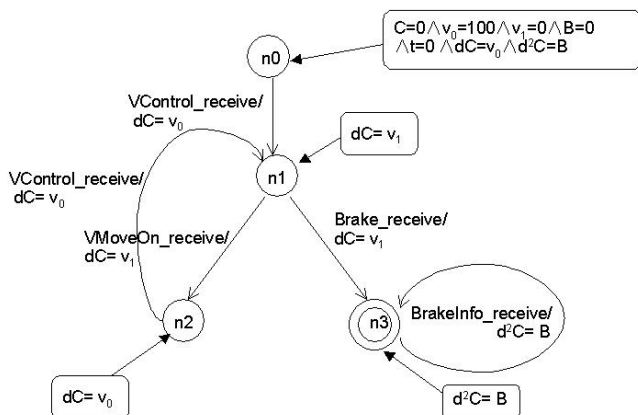


Figure 4: The DFA for business role IDS

5 The testing environment

IBM WebSphere Process Server (WPS) is an integration platform for Service Oriented Architecture (SOA), and supports the Service Component Architecture (SCA) programming model. As a process engine, WPS provides a hosting environment for business processes and several Web-based applications. These features of WPS match the structure and the requirements of continuous time SRML modules, so we choose WPS as the basic testing environment.

As the implementing tools, we combine WID with Matlab. In WID, an assembly editor is the central panel where service modules are built by creating and linking services components and service interfaces. A service component is specified with an implementation, some interfaces defining the inputs, outputs and faults of the service component, and zero or several references which denote the interface of other service components connecting to this service component. Service components can be implemented in various types, such as Java objects, business processes, and business state machines. We choose business state machines to implement our Traffic-Control module, because from Figure 5, we find that the hybrid automaton based DFA can be mapped to WID business state machines almost directly. Figure 5 shows the mapping relations between a continuous time SRML module and a WPS service module.

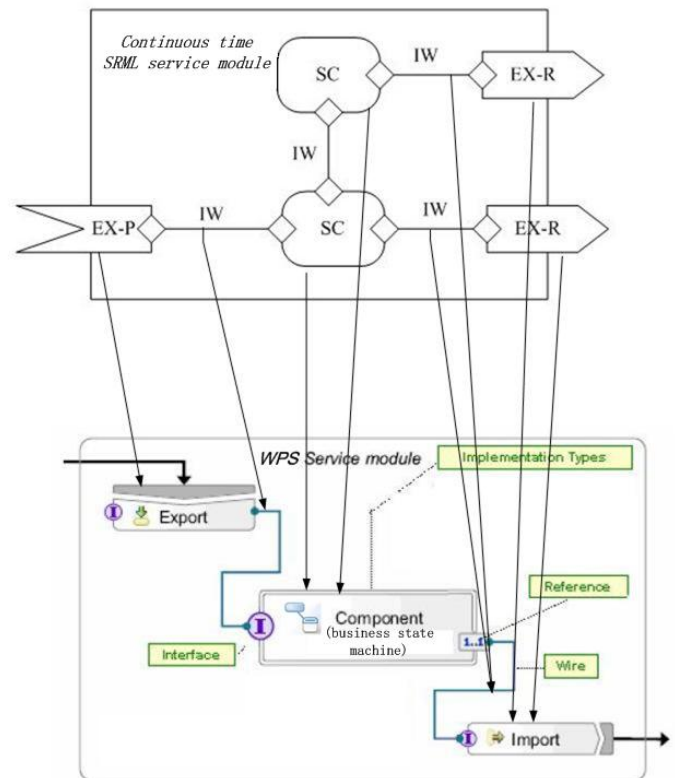


Figure 5: Mappings from continuous time SRML module to WPS service module

In a service component, the continuous processes described with differential equations can be directly implemented with Java code when there are exact solutions of the differential equations. Otherwise, we need to solve the differential equations numerically.

In such case, we combine WID with Matlab to achieve our goal. Matlab [9] is a multi-paradigm numerical computing environment, and it allows interfacing with programs written in languages such as C, C++ and Java. In the latest versions of Matlab, there is a function to wrap programs to Java packages. This function enables us to write Matlab programs to solve the differential equations in the service components in WID, and pass the solutions to the service components by importing the programs packed into Java packages to WID. We'll continue to do the implementation and testing in our future work.

6 Conclusion and future work

In this paper, we give a complete introduction to continuous time SRML, including the semantic domain of continuous time SRML, the application of continuous time SRML, and the testing environment. Continuous time SRML abstracts hybrid system-embedded service-oriented systems at a high level. The semantic foundation of continuous time SRML is SO-HL²TS, over which this modelling language is interpreted. The transitions of SO-HL²TSs enable continuous time SRML to model the communications between services or within a service; and the traces of states of SO-HL²TSs enable continuous time SRML to model the continuous processes performed by service components. Systems modelled with continuous time SRML can be transformed to hybrid automata-like DFA for testing and verification. This can be demonstrated by the case study of the Traffic Control System. By combining IBM WID and Matlab, we are able to implement the hybrid automata-like DFA.

In our future work, we will show the implementation of the hybrid automata-like DFA for the Traffic Control model, and provide the result of testing. We will also try to verify the Traffic Control system with KeYmaera X, which is a newly developed model checking tool for hybrid systems. For the transformation presented in Section 4, we are also working on a systematic approach so that this can be done automatically by programs. Further, we will try to enrich continuous time SRML with the expressions of actions, so that models specified with continuous time SRML match the business state machines of IBM WID better.

References

- [1] D. Georgakopoulos, M. Papazoglou, *Service-Oriented Computing*, The MIT Press, Cambridge, Massachusetts, 2009.
- [2] J. Fiadeiro, A. Lopes, J. Abreu, "A formal model for service-oriented interactions," *Science of Computer Programming*, **77**, 577–608, 2012, doi: 10.1016/j.scico.2011.12.003.
- [3] J. Fiadeiro, A. Lopes, L. Bocchi, J. Abreu, "The Sensoria Reference Modelling Language," *Lecture Notes in Computer Science*, **6582**, 61–114, 2011, doi:10.1007/978-3-642-20401-2_5.
- [4] M. Wirsing, M. Hölzl, *Rigorous Software Engineering for Service-Oriented Systems*, Springer, 2011.
- [5] N. Yu, *Injecting Continuous Time Execution into Service-Oriented Computing*, Munich University, Ph.D. thesis, 2016.
- [6] N. Yu, M. Wirsing, "The Application of Continuous Time SRML," in *Proceedings of 2020 ICICSE&ICACTE*, 99–103, 2020, doi:10.1145/3424311.3424316.
- [7] J. Abreu, *Modelling Business Conversations in Service Component Architectures*, University of Leicester, Ph.D. thesis, 2009.
- [8] IBM, *IBM Integration Designer V8.5.5 documentation*, https://www.ibm.com/support/knowledgecenter/SSTLXK_8.5.5/com.ibm.wbpm.wid.main.doc/kc-welcome.html, 2014.
- [9] MathWorks, <https://www.mathworks.com/>, 2014-2021.
- [10] T. A. Henzinger, "The theory of hybrid automata," in *Proceedings of 11th Annual IEEE Symposium on Logic in Computer Science*, 1996, doi:10.1109/LICS.1996.561342.
- [11] K. Cordwell, A. Platzer, "Towards Physical Hybrid Systems," in *Proceedings of International Conference on Automated Deduction, CADE'19*, 216–232, 2019, doi:10.1007/978-3-030-29436-6_13.
- [12] A. Platzer, *Logical Foundation of Cyber-Physical Systems*, Springer, Cham, 2018.
- [13] A. Platzer, *KeYmaeraX: An aXiomatic Tactical Theorem Prover for Hybrid Systems*, <http://www.ls.cs.cmu.edu/KeYmaeraX/index.html>, 2021.
- [14] M. Fadlisyah, E. Ábrahám, P. C. Lepri, D. Ölveczky, "A Rewriting-Logic-Based technique for modelling Thermal Systems," in *Proceedings of First International Workshop on Rewriting Techniques for Real-Time Systems, PTRTS 2010*, 82–100, 2010, doi:10.4204/EPTCS.36.5.
- [15] N. Yu, M. Wirsing, "A SOC-Based Formal Specification and Verification of Hybrid Systems," in *Proceedings of the 22nd International Workshop on Algebraic Development Techniques (WADT2014)*, 151–169, 2015, doi: 10.1007/978-3-319-28114-8_9.
- [16] R. De Nicola, F. Vaandrager, "Three Logics for Branching Bisimulation," *Journal of the ACM*, **42(2)**, 458–487, 1995, doi:10.1145/201019.201032.
- [17] A. Platzer, J. Ousel, "European Train Control System: A Case Study in Formal Verification," in *Proceedings of 11th International Conference on Formal Engineering Methods (ICFEM 2009)*, 246–265, 2009, doi:10.1007/978-3-642-10373-5_13.
- [18] The Worldwide Railway Organization (UIC), *The European Train Control System*, <https://uic.org/etcs#Standards-and-Specifications>, 2015.
- [19] J. E. Hopcroft, R. Motwani, J. D. Ullman, *Introduction to Automata Theory, Languages, and Computation (3rd Edition)*, Addison-Wesley Longman publishing Co., Inc., USA., 2006.

Intermediation in Technology Transfer Processes in Agro-Industrial Innovation Systems: State of Art

Leidy Dayhana Guarin Manrique^{1,*}, Hugo Ernesto Martínez Ardila², Luis Eduardo Becerra Ardila²

¹School of Electrical, Electronic and Telecommunications Engineering, Industrial University of Santander, Bucaramanga, 680002, Colombia

²INNOTECH Research Group, School of Industrial and Business Studies, Industrial University of Santander, Bucaramanga, 680002, Colombia

ARTICLE INFO

Article history:

Received: 16 July, 2021

Accepted: 21 October, 2021

Online: 23 November, 2021

Keywords:

Intermediation

Technology transfer

Innovation

Innovation systems

Agro-industrial innovation systems

ABSTRACT

Considering the importance of addressing innovation issues that impact the development of economic sectors, this document presents the research work aimed at establishing a state of the art related to technology transfer and intermediation issues, which can be adopted in the innovation systems. In this case, the agro-industrial innovation system is taken as a reference from the perspective of a country, Colombia. Likewise, it is proposed to consider the synergy between actors such as the university and the industry, from a holistic vision of the systems. In this sense, and making use of the Scopus, Web of Science and Google Scholar databases, through the implementation of a methodology in which three main phases of search, selection and reading of scientific publications were generated, a set of documents was obtained, and through these it was possible to identify: concepts on innovation systems, aspects that intervene in transfer processes of technology especially related to the articulation of the actors that are part of the innovation systems, as well as the way in which these issues can be adopted, taking the agro-industrial sector as a reference. Thus, through this research, the existence of structural gaps in the networks of actors is highlighted as a key factor, which, when trying to be moderated through the intervention of government actors, also require the participation of intermediary innovation agents, that facilitate the articulation and flow of data, information and communication between those who develop technology and those who require these developments to mitigate a need in the productive sectors.

1. Introduction

This paper is an extension of the work "Innovation systems supported in technology transfer processes" that is published in 2020 *International Congress of Innovation and Trends in Engineering* (CONITI) [1].

In Colombia, agribusiness is identified as a key sector for the sustainable development of the country [2], which is why the Law 1876 of 2017 has been approved, to support the creation of the National Agricultural Innovation System of Colombia where articulation, market orientation, sustainable development, associativity, intellectual property, development and innovation, and producers as research agents are highlighted as principles [3]. Additionally, this study seeks to contribute from the approach of technology transfer processes associated to the promotion of

innovation, initiatives that seek the consolidation of actions related to the fulfillment of the Sustainable Development Goals [4], and the adoption of challenges that try to contribute to the food assurance of the population by 2050, which is contemplated to be close to 9,100 million people [5].

In this sense, it is proposed that it is from the academic and industrial actors, as well as the identification of intermediary agents, mechanisms, or tools, that actions aimed at the articulation of actors in innovation systems are adopted; and the generation of research results, as well as their dissemination and appropriation, contribute both, to the sustainable development of the territory and the strengthening of innovation activities in the agro-industrial sector. Dissemination of research results is understood as the activity that makes known to the public or interested third parties, the results of the research carried out by a knowledge-generating actors [6].

*Corresponding Author: Leidy Dayhana Guarin Manrique, leidy2127825@correo.uis.edu.co

Therefore, and based on the importance that technology transfer processes have in the development of innovation systems of different geographical or spatial location, and sectoral scope, this work presents a set of concepts highlighting as a key aspect identified in the research carried out, the need to take into account the incorporation of innovation intermediary agents who through their contacts, networks, capacities, and experience are seen as those facilitators that can mitigate the gap of social distancing between the actors that can be part of the technology transfer processes in different sectors. Likewise, the novelty of this document is also associated with the analysis of the case of an agroindustrial innovation system that is visualized from government initiatives as a potential system to generate opportunities for articulation around innovation initiatives, which may finally arrive to have an impact in the markets.

2. Analysis Scenario

In the understanding of the complex nature of innovation systems in the agribusiness sector, within the framework of technology transfer processes, the intermediary agents are identified as actors of special interest for the analysis of synergies, because they assume a key role in the articulation of the various actors that make up these systems [7]. In this regard, universities should disseminate their research results through their technology transfer offices, which can contribute or be aligned with the needs and requirements of the industry. However, there are barriers related to the connections or articulation between actors of an innovation system [8], which hinder the development of dissemination activities.

Likewise, and based on the interaction between the university-industry-government and civil society actors that are part of the innovation systems [9], there are other barriers identified in the literature, which at a technological level hinder the implementation of activities of innovation-oriented to generate products and services with added value in the markets. One of the barriers is the asymmetry of information, which refers to the fact that both, the receiver and the sender in a technology transfer process do not have complete information on the technology developed and its exploitation possibilities to be able to reach a transfer agreement that satisfies the parties involved [10].

Based on the above, it is possible to identify problems associated with the knowledge and information necessary to establish a common language among the actors that participate in the technology transfer processes [9], which can lead, for example to lack of financing for technological development by potential investors, and the low opportunity for commercialization of technologies from universities.

Consequently, it is proposed that the lack of articulation of actors such as the university and industry in an innovation system, can negatively influence the generation of information flows between these two actors and, consequently, raise the levels of information asymmetry in a technology transfer process.

3. Methodology

The results obtained in this work have been framed in the execution of the phases presented in Figure 1.

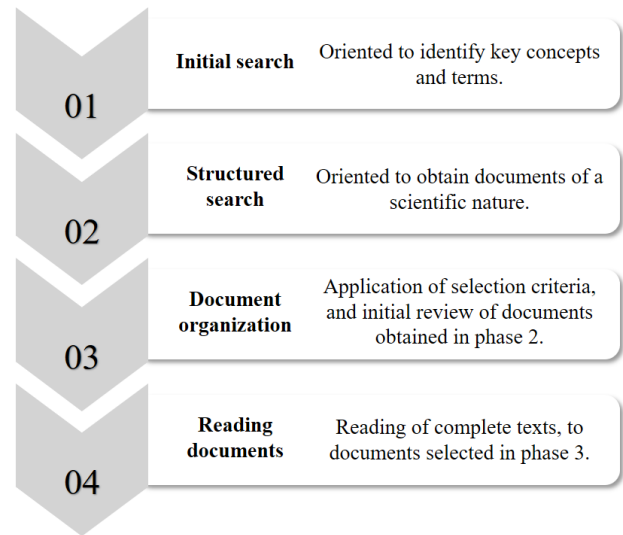


Figure 1: Research methodology..

In the first phase, the search for information was carried out with the purpose of identify the first set of documents related to technology transfer and intermediation in innovation systems. Consequently, from this information, key terms were identified for structuring the search equation and relevant authors on the aforementioned topics. In this search, the Google Scholar¹ tool was used, understanding that it is a global search engine that is constantly being updated, and in which recognized authors and leading institutions in the publication of various topics can be identified.

In the second phase, the selection of the specialized databases Scopus and Web of Science was carried out². These databases are considered robust in the global scope of publications they handle, as well as the multidisciplinary of the areas they cover, which allows addressing the issue of intermediation in technology transfer processes in a broad way. Likewise, it is highlighted that both Scopus and Web of Science handle quality documents, that is, the documents that are published are reviewed and validated by experts.

The search was executed on October 26, in 2019, and the used general equation was: TITLE-ABS-KEY (("technolog* transfer" OR "know* transfer" OR "technolog* exchang*" OR "know* exchang*") AND ("structur* mediat*" OR intermediar* OR gatekeep* OR broker* OR liaison* OR betweenness OR bridg* OR mediat* OR coordinat*)) AND (universit* OR academic)). The selection criteria applied to the results of this review were: 1) Spanish and English languages; 2) Include indexed articles in peer-reviewed scientific journals; 3) Articles belong to the citation databases: Social Sciences Citation Index (SSCI); 4) Select documents that are classified as articles or reviews, foreseeing that

¹ For more information: https://guiasbuh.uhu.es/Google_Scholar_Citations/Ventajas_y_desventajas_de_GSC

² For more information: <https://www.recursoscientificos.fecyt.es/licencias/productos-contratados/wos>

they must necessarily be reviewed by peer reviewers, representing an aspect of quality.

In the third phase, the results were grouped into two sets, which were: (i) Documents in the period from 2017 to 2019, which were read title and abstract, to select those associated with the subject under study, giving a total of 15 documents to be read completely; and, (ii) Documents obtained by adding (rev * OR meta *) to the equation, to which the title and summary were read to select those associated with the subject under study, giving a total of 52 documents that would be completely read.

In the last phase of the methodology, the complete documents were read, and information related to the topics of innovation systems, technology transfer, agro-industrial innovation systems, and intermediation was obtained. Finally, government-type sources associated with science, technology, and innovation issues and the agro-industrial sector in Colombia were reviewed.

In short, the protocol of the review to identify works associated with technology transfer and intermediation processes, included the following steps:

- Preliminary search in the Google Scholar database in order to identify relevant studies and keywords.
- Definition of keywords to search, establish and refine the search equation.
- Search for information in Web of Science and Scopus.
- Categorization and tabulation of results, including the number of publications according to countries, years, areas, in addition to the analysis of the most relevant journals, authors and keywords, in order to identify the guidelines followed by research on the subject of interest.
- Identification of articles to be analyzed applying the selection criteria, by reading the abstracts, titles and keywords.
- Reading and analysis of the documents.

4. State of the Art

Based on the results obtained from the methodology, two approaches have been found; the first, associated with the behavior of scientific knowledge as well as those publication sources that are relevant when reviewing works associated to the subject, and secondly, the most relevant issues associated with the action of the intermediary agents in technology transfer processes.

4.1. Scientific publications

On the one hand, regarding scientific publications, as reflected in Figure 2, technology transfer issues and intermediation actions related to the aforementioned processes have gained momentum during the last decade, highlighting the years 2017 and 2018 (until the moment of the search) as those spaces of time where the researchers found the topic more relevant. An example is found in the university context, where through intermediary agents such as their technology transfer offices, they seek to bring the inventions developed in the academy to the markets [9].

On the other hand, with respect to the most prominent journal sources obtained from the Scopus and Web of Science search, are shown in Table 1.

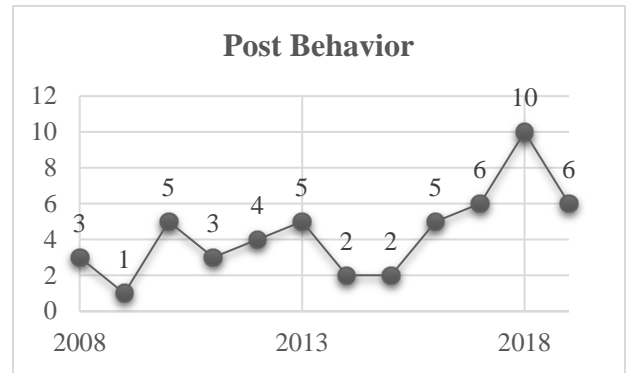


Figure 2: Publication in the time. Own elaboration based in Scopus and Web of Science.

Table 1: Most Outstanding Publication Sources. Own elaboration based in Scopus and Web of Science.

Source title	Number of Publications
Journal of Technology Transfer	5
Journal of Knowledge Management	4
Administrative Sciences	2
Technology Analysis & Strategic Management	2
Technological Forecasting and Social Change	2

Regarding the Journal of Technology Transfer³, it is identified that its focus is given by issues related to the study of technology transfer, from multiple dimensions of analysis, which can range from the identification of cases to issues of regulations or policies associated to technology transfer processes.

Another journal that has been highlighted is the Journal of Knowledge Management⁴ which relates mainly to knowledge management and innovation derived from different analysis contexts.

the following section is presented based on the results obtained from reading the documents found in the search for information, including those mentioned in the previous journals and during the period from 2008 to 2021.

4.2. Key information identified

For this second part, the following information compendium has been proposed, related to the topics of technology transfer, intermediation and agro-industrial innovation systems.

4.2.1. Technology Transfer

Technology transfer is understood as the process in which knowledge and technologies are disseminated between a holder and an interested party to be appropriated, used, or exploited [11],

³ For more information: <https://www.springer.com/journal/10961>
www.astesj.com

⁴ For more information: <https://www.emeraldgroupublishing.com/journal/jkm>

[12], [13], generally seeking to be brought to the market [14], in exchange for an incentive or monetary value [15].

Within the framework of these technology transfer processes, the following factors are taken into account: the generation of innovation in organizations [16], social capital [17] involved in the social networks of the actors who participate in the technology transfer [18], their interactions [19] and their previous relationships [20]; the university-industry relationships [21], the formality of the process (which can be formal or informal) according to knowledge management and the interactions between actors [17], [22], the limits of organizations and incentives [23], economic benefits e.g. profits on patents [24], technological cooperation between organizations [24], patents seen as a transfer mechanism [18], [25], the business culture that universities can adopt [21], the option of allowing students to participate in industry [17], the culture of the parties [17], legal frameworks [24], the uncertainties of innovations involved in technology transfer processes [26], and the importance of considering the intellectual property in the processes [13], [27]. This intellectual property can be seen materialized in the adoption of technology transfer mechanisms such as the creation of spin-off, patent licensing, collaboration actions between academia and industry [28], and the training of professionals with knowledge that contribute to the solution of industry needs [29].

Besides, when seeking to identify whether the technology transfer has been successful or not, it is found that depends of the accumulation or assimilation of knowledge in organizations that adopt the technologies that have been transferred [17]. In turn, at the macroeconomic level, technology transfer can be measured through business creation and job creation [30]. Additionally, other ways of identifying the success of technology transfer have been through the licensing of inventions [20], [24], incentives for personnel who participate in technology transfer processes [23], the indicators of patent families [24], and knowledge intensive entrepreneurship at universities [30].

However, in addition to factors in the technology transfer process, barriers are also identified that should be mitigated, some of these are: the lack of incentives from universities [9] and industry [16], cognitive and organizational distances between the university and industry [20], the time required to implement technological development activities, absorption capacity, the characteristics of knowledge, context and actors, and the culture of the parties participating in the process [17], the common language between the actors [9], the level of trust [17], [31], and the asymmetries of information between the parties [32], [33].

Finally, based on both the factors and the barriers that must be managed, it is necessary to understand that among the actors and medium involved in the process, it is possible to distinguish the sender that seeks to transfer the technology to a receptor, through a medium or mechanism, known as a channel [12]. In some scenarios, the channel may not exist, generating in the aforementioned process, a structural void that indicates the need for an agent to mitigate the gap in the flow of information, knowledge, and technologies between those who participate in the technology transfer processes. This agent is known as an intermediary [34], [35].

4.2.2. Intermediaries in the technology transfer

Intermediaries are agents that connect the components that structure an innovation system [36], also known as facilitators [37], who within their objectives seek to optimize interaction [38] to generate trust among the system's actors [39], strengthen interrelation processes, and at the same time mitigate the structural gaps that are identified in the social network of its actors [34].

At a general level, intermediaries are recognized as actors of special interest in innovation systems, for their intervention in actions aimed at mitigating barriers in technology transfer processes. They have key elements that help to the technology transfer process such as contacts, previous relationships with inventors, and experiences in technology commercialization [20]. In the case of the academic context that seeks to transfer technology to the industrial sector, the commercialization processes of the research results of a university, are more probable to achieve if there are links with industry [40]. In some cases, these links can be facilitated through the intervention of intermediary agents.

Thus, with regard to intermediation, the following variables have been analyzed in studies associated to the articulation between the university and industry actors: the experience of the technology transfer offices, which influences research and development agreements in the university [41], the previous relationships between intermediary agents and actors that participate in technology transfer processes [20], the performance of knowledge transfer offices measured through a number of supported patent applications, monetary income obtained from university-industry collaborations, and the creation of spin-offs [42], the efficiency of the technology transfer offices of the universities based on their age [43], [44], and the performance of technology transfer offices measured through support in licensing, patenting and Spin-off creation activities [45].

An example of intermediation is the case of technology and science parks. They support the creation of networks, especially those where university plays an important role due to both its link to the park and the legitimacy they provide to companies through their image of reliability and reputation [46]. Another example are the technology transfer offices, which within their functions provide advice on legal protection issues [45] of the technologies that are waiting to be brought to the market, in addition to providing support in the creation of technology-based companies, technological marketing [45], and assessment and due diligence in commercialization tasks of technological developments.

However, some studies have focus on the idea that an intermediary agent could only have a particular interest in obtaining economic benefits from technological results [37], beyond considering its action as positive in aspects of knowledge transfer, articulation [7] and interaction [47]. Consequently, the purpose of the intermediary agent could be an important barrier that has been associated to their nature, whether they really seek to generate linking actions, or simply act with individual commercial or political interests [48].

Therefore, it is from the nature of the intermediary agents and the objective they pursue, that they act in the configuration of the different innovation systems, especially when taking into account

the different barriers that limit the successful result in terms of technology transfer [20].

4.2.3. Agro-industrial innovation systems

The concept of agro-industrial innovation systems considers the definition of a system that refers to the set of components that are interrelated with a defined objective in a given context [49], in accordance with the synergy that it experiences [50], [51]. Furthermore, when systems are complemented with the term innovation, they open the space to identify the possible achievement of competitive advantages [52]. When it comes to competitive development issues in the countries, innovation should be aimed at promoting the transfer of knowledge to reality considering the results of research, observation, discoveries, or intuitions that are generated in the nation [53].

Therefore, the conceptualization of innovation systems is open and unfinished [54]. However, some characteristics are critical such as it incorporates actors of different nature with the purpose of exchange of technologies [55], have different scope at the geographical or sectoral level [51], [56], [57] it is associated with social networks [58], and other aspects such as synergy management [58], knowledge management processes [59], [60], and the procurement of economic advantages [61].

In the case of sectoral innovation systems, which are those of interest for this research associated with the agro-industrial sector, the following stand out:

- Actors are related based on the economic sector to which their activities are associated [62].
- A differential element of these systems is the knowledge associated with innovation processes [63] in the sector.
- The key factor for the development of these innovation systems is the group of companies involved in accordance to the conditions of the sector [64].

- A set of elements of interest that sustains the development of innovation systems at the sectoral level are related to social networks, capacities of industry actors, the policies that direct their development [65], assimilation and knowledge dissemination [66], and finally, the supply chain [67].

According to the analysis level different study dimensions can be proposed as possible approaches to understand the innovation processes and its possible obstacles [53].

A representation of the set of dimensions and elements that can be considered for study purposes of innovation systems in the agro-industrial sector, and that have been previously adopted in the research activities [68], [69] in Colombia, can be seen in Figure 2.

Figure 2 shows that the dynamics of the innovation system is supported by the integration of different elements, such as a specific context, dimensions that make it possible to study the dynamics of the system from different perspectives, agents that externally intervene in the system, system behavior, and a set of interrelated actors represented under the concept of the quadruple helix [70] within the framework of a knowledge economy [71], [72]. These interrelationships are shown as a key attribute of the system, since from them, it is possible to identify realities of the different sectors [53] and generate networks or strengthen those that already exist, thus enabling information flows, knowledge and technology transfers, the realization of agreements, the establishment of collaborations, and articulation actions organized around a purpose. However, these interrelationships may require the intervention of facilitating agents, which generate links between the actors of the system [39], [73], [74]. These agents identified as intermediaries, facilitate the creation of networks of actors promoting collaboration and the flow of knowledge and information [20], [36], [37], [39], [42], [75]-[82].

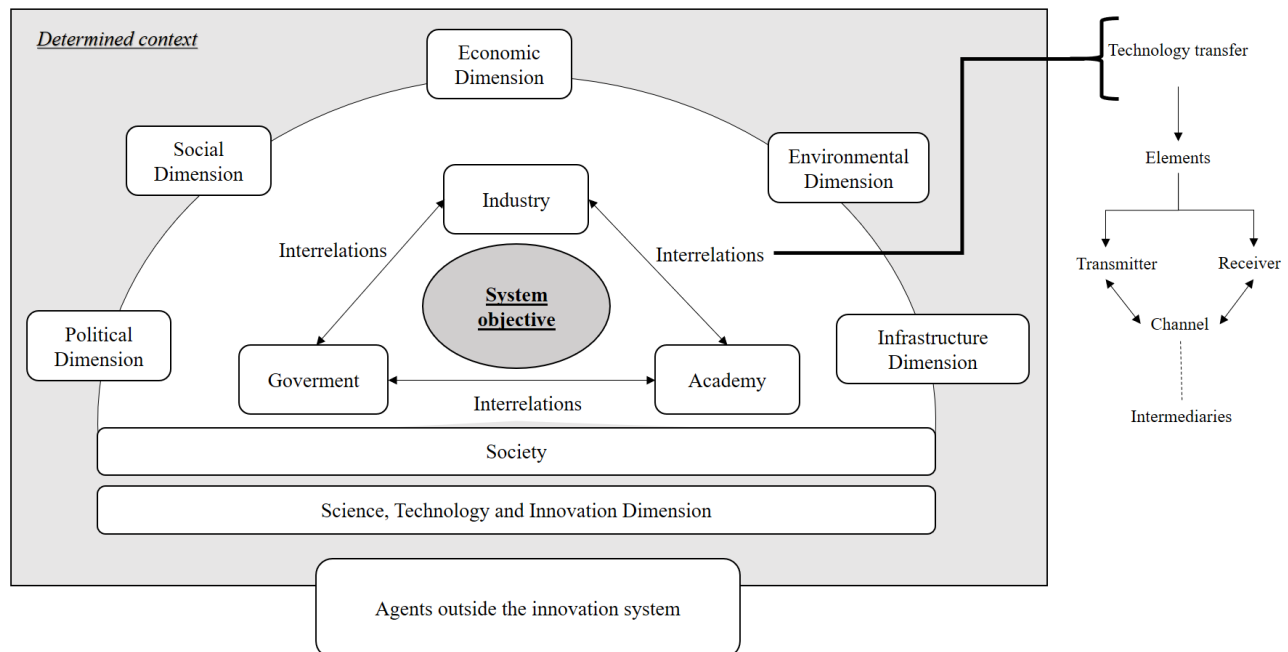


Figure 3: Dimensions and elements in innovation systems. Based on [53], [69].

Based on what is illustrated in Figure 2, it is found that there are similarities with other models of innovation systems, (e.g. the Tropical Agriculture Platform) [83]. The following are highlighted as common key elements:

- The proposal of a specific context, made up of institutional actors and innovation policies, which interact to achieve innovation objectives applied to the agricultural sector. In this regard, public and private research and technological development organizations stand out as key actors [22].
- The existence of institutions that act as intermediaries between the actors that carry out research, and the industry that commercialize the research output. An example is the technology transfer offices, whose mission is to promote technology-based entrepreneurship [74], [77], [84], [85], facilitate the commercialization of knowledge, manage intellectual property [9], [86], provide advice on inventions, manage academic human capital, promote contract research and promote articulation between the different actors of the university, industry, and government [86].
- Interaction with actors from other sectors, which according to their own purposes come to influence the synergy of innovation activities that take place in the sector system.
- The proposal of a political dimension, which guides both the planning and the execution of the guidelines that regulate the synergy of the system.

From the articulation of actors in the innovation systems, the sectoral innovation system like the agro-industrial, can achieve results in processes of diffusion, adoption, and appropriation of technologies among the university, industry, government, and society in general.

In addition, the value of including intermediary agents in technology transfer processes is also recognized, anticipating the possible existence of gaps in terms of social networks that emerge from the dynamics of these systems, understanding that within their role as articulatory agents, they allow generating a higher level of cohesion in the relationship of the various actors in the system.

5. Discussion

Faced with the development of innovation processes and technological management, in 2019 Colombia ranks 67th out of 129 countries in the global innovation index, from which aspects such as infrastructure, market, business sophistication, institutions, knowledge and technology production, human capital, researching, and creative production, were analyzed [87], [88]. In addition, as a result of this evaluation, in terms of collaboration activities between the university and the industry to develop research, the country ranks third in the Latin American region [88]. In the same orientation towards technology management processes, at the national level, it finds that most of the innovations in Colombia are developed from the research actions of the universities [89].

However, despite identifying that innovation represents an opportunity for economic development, the country requires an optimal infrastructure and necessary conditions to carry out technology transfer processes. In addition, it was identified in the

science, technology, and innovation policies, that there is a low level of intermediary agents such as technology transfer offices, incubators, and science and technology parks, which by their nature support the linkage, interaction, and flow of information and knowledge between the actors of the quadruple helix [90]. The existence of a low capacity of the industrial sector to articulate and transfer knowledge and technologies among the actors of the innovation system in Colombia is pointed out in this policies. Thus, in the national development plan of Colombia, it is identified that only 3.8% of companies work collaboratively with universities, a figure that has influenced the proposal to reach a goal of 59 technology transfer agreements between the university and the industry in the country, in the period 2018-2021 [91].

From the approach of sectoral innovation systems associated with agribusiness in Colombia, the Strategic Plan for Agricultural Science, Technology and Innovation has been established [92], which takes into account the need to generate strategies aimed at promoting programs of transfer and adaptation of technologies associated with the different productive sectors, in which the proposal of mechanisms oriented to generate articulation between the different actors of the agricultural innovation system should be included. In this sense, from the Colombian Agricultural Research Corporation (Agrosavia), the Ministry of Agriculture and Rural Development of Colombia, and the Ministry of Science, Technology, and Innovation of Colombia, in the formulation of the Strategic Plan for Agricultural Science, Technology and Innovation [92], formulate two actions focusing on science, technology and innovation in the area of technology transfer, technical assistance and innovation [92], which are: (i) “the development, adjustment, and validation of extension tools and specialized comprehensive technical assistance”, and (ii) “the generation of collaborative workspaces through territorial innovation systems”.

Based on the previous approaches, and understanding the value chain made up of the primary sector associated with agricultural activities, and the secondary sector related to agribusiness activities, the Santander agro-industrial innovation system is proposed as an example, in which numerous technological developments and research results carried out by universities have been identified, which tend to be unknown by the sector and have not had the possibility of being commercialized, particularly due to the lack of articulation between the actors that are part of the innovation system [93]. In addition, there are shortcomings in the information flows required from both the receiver and the sender side, to develop technology transfer processes from the university to the agro-industrial sector.

Another case of the agricultural sector in which similar problems have been studied in technology transfer processes, and in which the intermediary actions of science parks, technology transfer offices, consultants, and in general, intermediary innovation agents have favored its mitigation in the productive and market contexts, is reflected in the Dutch case [37] where intermediaries act as facilitators or forgers of networks of actors that interact under the synergy of transfer processes.

Based on the above, it is necessary to understand that when technology transfer processes are developed, elements associated with information and knowledge are involved. Therefore, it is

necessary to take into account the existence of cognitive differences between the actors who seek to interrelate [80] around these processes. These cognitive differences can be related to the information asymmetry barrier, since it represents the gaps in terms of the information that a technology owner requires to determine the interest of a second party in its appropriation, as well as the information that the recipient needs to adopt it [10] and therefore use, exploit or commercialize it. In this sense, this barrier can lead to inefficiencies in technology commercialization activities between the university and industry [94].

Overcoming the asymmetry of information can facilitate the investment activities of industry players in the technological developments of the university, which can be considered as university ventures [95], thus opening the possibility of promoting commercial management actions among inventors or holders of technological developments and representatives of the industry, where intellectual property rights play a key role [20].

Additionally, it is important to bear in mind how the information asymmetry barrier is mitigated in the technology transfer process between the university and the industry, and how it will depend on the inventors' ability to disseminate [74], the absorption capacity of industry representatives [96], and the participation of intermediary agents that moderate information flows.

Therefore, it is proposed that the experience in commercialization and the social networks of intermediaries, as well as their previous relationships with inventors, can help reduce the information asymmetry between the senders and recipients of a technology transfer process [20]. In the case of universities among the factors that can be taken into account in the process of reducing the level of asymmetry of information is its prestige, since it generates more trust among the actors of the different economic sectors [97].

6. Conclusions

Based on the information presented above, it is necessary to highlight the importance of encouraging the different actors of the agribusiness innovation systems, to develop articulation that in addition of being focused on technology transfer processes, also generate alliances between the university and the industry, both in research processes and in the dissemination and exploitation of their results. Likewise, and anticipating the possible existence of structural gaps in the social networks of these systems, the option of identifying and interacting with intermediary actors is proposed as an alternative, understanding that their intervention in the technology transfer processes between the university and the industry can be assumed as a strategic factor that would contribute to the economic development of a region.

In front of this scenario, it is also necessary for government sector actors to contribute to the strengthening of the synergy of innovation systems, through regulatory mechanisms focused on expanding articulation actions and alliances between the actors of the aforementioned systems, to promote territorial development from the different economic sectors of the regions.

When it comes to contributing to economic sectors, including the agro-industrial sector, innovation is evidenced as a key factor

in strengthening the markets. In addition to generate competitive advantages in the market, it can be associated to the processes of technology transfer between the university and the industry. A whole development takes place that starts from the scientific research and matures until reaching a viable product or service to be offered to the industry. Therefore, in this scenario, it is proposed that the university should have capacities to disseminate its technologies, and the industry must have capacities to adopt the technologies that are transferred to it, the foregoing to achieve, as a result, the use, exploitation, or commercialization of inventions and technological developments generated by universities.

In addition, the level of connectivity that the actors representing the university and industry can achieve also affects technology transfer processes, which is influenced by the asymmetry of information generated by the parties involved in the process of technology transfer, which is why an agent that acts as an intermediary between the university and the industry is identified as a third actor, that manages to moderate the asymmetry of information and at the same time make the transfer process effective by facilitating the flow of information between the parties involved in the process.

Based on the above, and on what was stated during the document, it is necessary to take into account that technology transfer processes can use different mechanisms that range from the creation of spin-offs to the generation of agreements between universities and companies, where the human capital of the universities brings their knowledge to the industry and reinforces their capacities through the presumption of real market scenarios. However, it is also relevant to understand that technology transfer processes include a set of factors and barriers that must be viewed from a multidisciplinary approach, for which hybrid capabilities between the university and industry require an effort to be developed by the actors that participate in these processes, thus allowing the intervention of intermediary agents who, due to their nature, can support the management of actions against the factors and barriers mentioned.

7. Future work

It is recommended to generate research actions around topics such as:

The evaluation of the effect generated by the different types of intermediary actors in the technology transfer processes in the agro-industrial sector, and also in other sectors of interest to the various actors that make up the innovation systems at the national, regional, or sectoral level.

Generate alliance options around various research and technological development projects, executed between the university and the industry, in such a way that their effectiveness can be evaluated in terms of technology transfer processes.

The dissemination of the results that the different types of intermediary agents have achieved in technology transfer processes, in such a way that in the second phase of this activity, an initial set of indicators can be proposed that allow measuring the impact of the intervention of these actors in the development of the systems of innovation.

Regarding future lines of research, it is important to focus towards the study of the nature of the actors that are part of the technology transfer processes, in such a way that it can be evidenced the types of academic, industrial and hybrid capacities that allow facilitating the transfer of inventions that are developed in the university to the market.

Another alternative study is identified in the need to generate university-industry articulation methodologies, in which the objective is to identify market challenges that can be developed by the academy.

References

- [1] L. Guarín, H. Martínez, "Innovation systems supported in technology transfer processes," in 2020 Congreso Internacional de Innovación y Tendencias en Ingeniería (CONIITI), 1–9, 2020, doi:10.1109/CONIITI51147.2020.9240319.
- [2] DANE, "Boletín técnico - Producto Interno Bruto PIB -Cuarto trimestre de 2017," 28, 2018.
- [3] MinAgricultura, Ley 1876 por medio de la cual se crea el Sistema Nacional de Innovación Agropecuaria y se dictan otras disposiciones. Congreso de Colombia, 29 de diciembre de 2017, 1–28, 2017.
- [4] PNUD, Proyecto desarrollo económico incluyente, 2011.
- [5] FAO, Cómo alimentar al mundo en 2050, 2009.
- [6] J.J.S. Gil, "La difusión del conocimiento como un elemento clave de la innovación y la competitividad en el entorno universitario bajo el sistema Open Journal System (OJS)," International Journal of Information Systems and Software Engineering for Big Companies (IJSEBC), 4(1), 63–71, 2017.
- [7] M. Ramírez, I. Clarke, L. Klerkx, "No Analysing intermediary organisations and their influence on upgrading in emerging agricultural clusters," Environment and Planning A: Economy and Space, 50(6), 1314–1335, 2018, doi:10.1177/0308518X17741316.
- [8] AGROSAVIA, Misión para la transformación del campo. Diagnóstico en CTeI, 2015.
- [9] E.S. Olaya, J. Berbegal-Mirabent, O.G. Duarte, "Performance of university transfer offices as intermediary for knowledge market development [Desempeño de las oficinas de transferencia universitarias como intermediarias para la potencialización del mercado de conocimiento]," Intang. Cap, 2014, doi:10.3926/ic.497.
- [10] OMPI, Comité permanente sobre el derecho de patentes. Decimocuarta sesión., 2010.
- [11] B. Bozeman, H. Rimes, J. Youtie, "The evolving state-of-the-art in technology transfer research: Revisiting the contingent effectiveness model," Research Policy, 44(1), 34–49, 2015, doi:10.1016/j.respol.2014.06.008.
- [12] C. Battistella, A.F. De Toni, R. Pillon, "Inter-organisational technology/knowledge transfer: a framework from critical literature review," The Journal of Technology Transfer, 41(5), 1195–1234, 2016, doi:10.1007/s10961-015-9418-7.
- [13] C.S. Hayter, E. Rasmussen, J.H. Rooksby, "Beyond formal university technology transfer: innovative pathways for knowledge exchange," The Journal of Technology Transfer, 45(1), 1–8, 2020, doi:10.1007/s10961-018-9677-1.
- [14] F. Donneys González, B. Blanco Campins, "La transferencia de tecnología en universidades colombianas," Economía y Desarrollo, 157(2), 182–198, 2016.
- [15] F.C. Sercovich, "Dependencia tecnológica en la industria Argentina," Desarrollo Económico, 33–67, 1974, doi:10.2307/3466047.
- [16] T.D.H. Gusberti, A.C. Bretas, "Diagnosis of the market for ideas and the role of industrial associations as intermediaries in the Brazilian context," Industry and Higher Education, 32(1), 9–22, 2018, doi:10.1177/0950422217733087.
- [17] C. Van Horne, D. Vincent, "Challenges in technology transfer: An actor perspective in a quadruple helix environment," The Journal of Technology Transfer, 42(2), 285–301, 2017, doi:10.1007/s10961-016-9503-6.
- [18] Y. Ye, K. De Moortel, T. Crispeels, "Network dynamics of Chinese university knowledge transfer," The Journal of Technology Transfer, 45(4), 1228–1254, 2020, doi:10.1007/s10961-019-09748-7.
- [19] P. D'Este, O. Llopis, F. Rentocchini, A. Yegros, "The relationship between interdisciplinarity and distinct modes of university-industry interaction," Research Policy, 48(9), 1–57, 2019, doi:10.1016/j.respol.2019.05.008.
- [20] Y.-C. Kim, M. Rhee, R. Kotha, "Many hands: The effect of the prior inventor-intermediaries relationship on academic licensing," Research Policy, 48(3), 813–829, 2019, doi:10.1016/j.respol.2018.11.007.
- [21] P. Centobelli, R. Cerchione, E. Esposito, "Exploration and exploitation in the development of more entrepreneurial universities: A twisting learning path model of ambidexterity," Technological Forecasting and Social Change, 141, 172–194, 2019, doi:10.1016/j.techfore.2018.10.014.
- [22] S. Zaichenko, "The human resource dimension of science-based technology transfer: lessons from Russian RTOs and innovative enterprises," The Journal of Technology Transfer, 43(2), 368–388, 2018, doi:10.1007/s10961-017-9567-y.
- [23] S. Horner, D. Jayawarna, B. Giordano, O. Jones, "Strategic choice in universities: Managerial agency and effective technology transfer," Research Policy, 48(5), 1297–1309, 2019, doi:10.1016/j.respol.2019.01.015.
- [24] C. Martínez, L. Bares, "The link between technology transfer and international extension of university patents: Evidence from Spain," Science and Public Policy, 45(6), 827–842, 2018, doi:10.1093/scipol/scy008.
- [25] D. Chatterjee, L. Leisyte, S. Dasappa, B. Sankaran, "University research commercialization in emerging economies: A glimpse into the 'black box,'" Science and Public Policy, 45(3), 361–372, 2018, doi:10.1093/scipol/scx066.
- [26] N. Baines, H.L. Smith, "Key driving factors for product and service innovations in UK university spin-offs," Industry and Higher Education, 33(3), 161–171, 2019, doi:10.1177/0950422219825579.
- [27] G.A. Van Norman, R. Eisenkot, "Technology transfer: From the research bench to commercialization. Part 1: Intellectual property rights —basics of patents and copyrights," JACC: Basic to Translational Science, 2(1), 85–97, 2017, doi:10.1016/j.jacbts.2017.01.003.
- [28] J. González, Manual de transferencia de tecnología y conocimiento, The Transfer Institute, 2011.
- [29] A. Anduaga, "The engineer as a 'linking agent' in international technology transfer: The case of Basque engineers trained in Liège," Engineering Studies, 3(1), 45–70, 2011, doi:10.1080/19378629.2010.550922.
- [30] F. Iqbal, P.C. Hung, F. Wahid, S.M.Q.A. Mohammed, "A glance at research-driven university's technology transfer office in the UAE," International Journal of Technology Management, 78(1–2), 70–87, 2018, doi:10.1504/IJTM.2018.093939.
- [31] E. De Wit-de Vries, W.A. Dolfisma, H.J. Van der Windt, M.P. Gerkema, "Knowledge transfer in university–industry research partnerships: a review," The Journal of Technology Transfer, 44(4), 1236–1255, 2019, doi:10.1007/s10961-018-9660-x.
- [32] N.T. Gallini, B.D. Wright, "Technology transfer under asymmetric information," The RAND Journal of Economics, 147–160, 1990, doi:10.2307/2555500.
- [33] J.-W. Min, N.S. Vonortas, Y. Kim, "Commercialization of transferred public technologies," Technological Forecasting and Social Change, 138, 10–20, 2019, doi:10.1016/j.techfore.2018.10.003.
- [34] M. Fritsch, M. Kauffeld-Monz, "The impact of network structure on knowledge transfer: An application of social network analysis in the context of regional innovation networks," The Annals of Regional Science, 44(1), 21, 2010, doi:10.1007/s00168-008-0245-8.
- [35] S. Parjanen, H. Melkas, T. Uotila, "Distances, knowledge brokerage and absorptive capacity in enhancing regional innovativeness: A qualitative case study of Lahti region, Finland," European Planning Studies, 19(6), 921–948, 2011, doi:10.1080/09654313.2011.568804.
- [36] J. Howells, "Intermediation and the role of intermediaries in innovation," Research Policy, 35(5), 715–728, 2006, doi:10.1016/j.respol.2006.03.005.
- [37] L. Klerkx, C. Leeuwis, "Establishment and embedding of innovation brokers at different innovation system levels: Insights from the Dutch agricultural sector," Technological Forecasting and Social Change, 76(6), 849–860, 2009, doi:10.1016/j.techfore.2008.10.001.
- [38] L. Klerkx, C. Leeuwis, "Matching demand and supply in the agricultural knowledge infrastructure: Experiences with innovation intermediaries," Food Policy, 33(3), 260–276, 2008, doi:10.1016/j.foodpol.2007.10.001.
- [39] W.L. Ruíz Castañeda, S. Quintero Ramírez, J. Robledo Velásquez, "Impacto de los intermediarios en los sistemas de innovación," Journal of Technology Management & Innovation, 11(2), 130–138, 2016, doi:10.4067/S0718-27242016000200013.
- [40] R. Welsh, L. Glenna, W. Lacy, D. Biscotti, "Close enough but not too far: Assessing the effects of university-industry research relationships and the rise of academic capitalism," Research Policy, 37(10), 1854–1864, 2008, doi:10.1016/j.respol.2008.07.010.
- [41] A. Caldera, O. Debande, "Performance of Spanish universities in technology transfer: An empirical analysis," Research Policy, 39(9), 1160–1173, 2010, doi:10.1016/j.respol.2010.05.016.
- [42] J. Berbegal-Mirabent, F. Sabaté, A. Cañabate, "Brokering knowledge from universities to the marketplace: The role of knowledge transfer offices," Management Decision, 50(7), 1285–1307, 2012, doi:10.1108/00251741211247012.

- [43] W. Chapple, A. Lockett, D. Siegel, M. Wright, "Assessing the relative performance of U.K. university technology transfer offices: Parametric and non-parametric evidence," *Research Policy*, **34**(3), 369–384, 2005, doi:10.1016/j.respol.2005.01.007.
- [44] D.S. Siegel, D.A. Waldman, A.N. Link, "Assessing the impact of organizational practices on the productivity of university technology transfer offices: An exploratory study," 1999, doi:10.1016/S0048-7333(01)00196-2.
- [45] D.S. Siegel, R. Veugelers, M. Wright, "Technology transfer offices and commercialization of university intellectual property: performance and policy implications," *Oxford Review of Economic Policy*, **23**(4), 640–660, 2007, doi:10.1093/oxrep/grm036.
- [46] L. Lecluyse, M. Knockaert, A. Spithoven, "The contribution of science parks: a literature review and future research agenda," *The Journal of Technology Transfer*, **44**(2), 559–595, 2019, doi:10.1007/s10961-018-09712-x.
- [47] J.M. Ekboir, A.O. Vera-Cruz, "Intermediary organisations to foster the agricultural system of innovation: The Mexican Produce Foundation," *International Journal of Technological Learning, Innovation and Development*, **5**(1–2), 111–125, 2012, doi:10.1504/IJTLID.2012.044879.
- [48] H. Yang, L. Klerkx, C. Leeuwis, "Functions and limitations of farmer cooperatives as innovation intermediaries: Findings from China," *Agricultural Systems*, **127**, 115–125, 2014, doi:10.1016/j.agsy.2014.02.005.
- [49] B. Wilson, *Sistemas: Conceptos, Metodología y Aplicaciones*, Editorial Limusa S.A. De C.V., 1993.
- [50] Y.A. Cáceres, L.C. Flórez, "Una reflexión sistémica sobre los fundamentos conceptuales para sistemas de información," *Revista Colombiana de Computación*, **8**(1), 1–19, 2007.
- [51] J. Heijts, *Sistemas nacionales y regionales de innovación y política tecnológica: Una aproximación teórica*, 2001.
- [52] L. Mathison, J. Gándara, C. Primera, L. García, "Innovación: Factor clave para lograr ventajas competitivas," *Negotium*, **3**(7), 65–83, 2007.
- [53] J. Sábato, N. Botana, *La ciencia y la tecnología en el desarrollo futuro de América Latina*, Lima, 1970.
- [54] C. Freeman, *Technology, policy, and economic performance: lessons from Japan*, Frances Printer Publishers, 1987.
- [55] E.M. Balmaseda, I.Z. Elguezal, G.I. Clemente, "Evolución de los modelos sobre el proceso de innovación: desde el modelo lineal hasta los sistemas de innovación," in *Decisiones basadas en el conocimiento y en el papel social de la empresa: XX Congreso anual de AEDEM, Asociación Española de Dirección y Economía de la Empresa*, 2007.
- [56] *Minciencias, Actores del Sistema Nacional de Ciencia, Tecnología e Innovación*, 2016.
- [57] C. Freeman, "The 'national system of innovation' in historical perspective," *Cambridge Journal of Economics*, **19**, 15–24, 1995.
- [58] L.J. Quintero-Campos, "Aportes teóricos para el estudio de un sistema de innovación," *INNOVAR. Revista de Ciencias Administrativas y Sociales*, **20**(38), 57–76, 2010.
- [59] D.J. Spielman, *Innovation systems perspectives on developing-country agriculture: A critical review*, 2005, doi:10.22004/ag.econ.59692.
- [60] R. Rajalahti, W. Janssen, E. Pehu, *Agricultural innovation systems: From diagnostics toward operational practices*, 2008.
- [61] B.-Å. Lundvall, *The learning economy and the economics of hope*, Anthem Press, 2016.
- [62] P. Adams, F. Malerba, *Sectoral Systems of Innovation*, Oxford University Press, 2014.
- [63] J. Jasso Villazul, *Los sistemas de la innovación como espacios regionales, sectoriales y empresariales: Características y taxonomía*, 2000.
- [64] M.N. Arancegui, "Los sistemas regionales de innovación: una revisión crítica," *EKONOMIAZ. Revista Vasca de Economía.*, **70**(1), 25–59, 2009.
- [65] F. Malerba, R. Nelson, "Sistemas sectoriales, alcance y desarrollo económico," *Economía: Teoría y Práctica*, **SPE1**, 41–62, 2009.
- [66] M. Amaro Rosales, M. Morales Sánchez, "La Biotecnología en México, una aproximación desde los sistemas sectoriales de innovación," *Ide@ s CONCYTEG*, **5**(64), 1224–1246, 2010.
- [67] T. Mets, "Developing the Sectorial Innovation System of Estonian Biotechnology," *Engineering Economics.*, **5**(50), 73–79, 2006.
- [68] UIS, UNAB, UPB, INAL, *Diseño de un modelo de trabajo colaborativo entre actores del sector agropecuario para el desarrollo de la Agrópolis de Santander Magdalena Medio*, 2016.
- [69] UIS, Findeter, BID, *Agrópolis de Santander Magdalena Medio - Estudios de profundización escala diamante*, 2016.
- [70] O. Afonso, S. Monteiro, M. Thompson, "A growth model for the quadruple helix," *Journal of Business Economics and Management*, **13**(5), 849–865, 2012, doi:10.3846/16111699.2011.626438.
- [71] H. Etzkowitz, *La triple hélice: universidad, industria y gobierno. Implicaciones para las políticas y la evaluación*, 2002.
- [72] H.G.C. Castillo, "El modelo de la triple hélice como un medio para la vinculación entre la universidad y empresa," *Revista Nacional de Administración*, **1**(1), 85–94, 2010, doi:10.22458/ma.v1i1.286.
- [73] W.L. Ruíz Castañeda, *Análisis del impacto de los intermediarios en los sistemas de innovación: Una propuesta desde el modelado basado en agentes*, Universidad Nacional de Colombia, 2016.
- [74] W.L. Castañeda Ruíz, J. Robledo Velásquez, "Evaluación del Impacto de los Intermediarios en los Sistemas de Innovación: Marco de Análisis," in *ALTEC*, 1–17, 2013.
- [75] H. Van Lente, M. Hekkert, R. Smits, B.A.S. Van Waveren, "Roles of systemic intermediaries in transition processes," *International Journal of Innovation Management*, **7**(03), 247–279, 2003, doi:10.1142/S1363919603000817.
- [76] J. Vereijssen, T.M. Williams, N.M. Park, M. Nielsen, N.M. Agnew, "Evaluating co-innovation principles in a fundamental bioprotection research programme addressing challenges to potato production," *New Zealand Plant Protection*, **70**, 16–24, 2017, doi:10.30843/nzpp.2017.70.20.
- [77] E. Villani, E. Rasmussen, R. Grimaldi, "How intermediary organizations facilitate university–industry technology transfer: A proximity approach," *Technological Forecasting and Social Change*, **114**, 86–102, 2017, doi:10.1016/j.techfore.2016.06.004.
- [78] N. Aguilar-Gallegos, E.G. Martínez-González, J. Aguilar-Ávila, H. Santoyo-Cortés, M. Muñoz-Rodríguez, E.I. García-Sánchez, "Análisis de redes sociales para catalizar la innovación agrícola: de los vínculos directos a la integración y radialidad," *Estudios Gerenciales*, **32**(140), 197–207, 2016, doi:10.1016/j.estger.2016.06.006.
- [79] M. De Silva, J. Howells, M. Meyer, "Innovation intermediaries and collaboration: Knowledge-based practices and internal value creation," *Research Policy*, **47**(1), 70–87, 2018, doi:10.1016/j.respol.2017.09.011.
- [80] C.W. Kilelu, L. Klerkx, C. Leeuwis, A. Hall, "Beyond knowledge brokering: an exploratory study on innovation intermediaries in an evolving smallholder agricultural system in Kenya," *Knowledge Management for Development Journal*, **7**(1), 84–108, 2011, doi:10.1080/19474199.2011.593859.
- [81] P. Kivimaa, W. Boon, S. Hyysalo, L. Klerkx, "Towards a typology of intermediaries in sustainability transitions: A systematic review and a research agenda," *Research Policy*, **48**(4), 1062–1075, 2019, doi:10.1016/j.respol.2018.10.006.
- [82] L. Klerkx, M. Schut, C. Leeuwis, C. Kilelu, "Advances in Knowledge Brokering in the Agricultural Sector: Towards Innovation System Facilitation," *IDS Bulletin*, **43**(5), 53–60, 2012, doi:10.1111/j.1759-5436.2012.00363.x.
- [83] P. de A.T. TAP, *Marco Común sobre el Desarrollo de Capacidades para los Sistemas de Innovación Agrícola: Documento de Síntesis*, CAB International, Wallingford, Reino Unido., 2016.
- [84] J.M.B. Garmendia, A.R. Castellanos, "Estructuras de intermediación para la transferencia de conocimiento universitario: las oficinas de transferencia tecnológica," *Propiedad Intelectual*, **9**(13), 152–176, 2010.
- [85] J. Maira, J. Etxabe, P.A. Serena, "Strategies on Technology Transfer and Patents Commercialization for Nanotechnology at the Spanish National Research Council," *Recent Patents on Nanotechnology*, **12**(1), 45–58, 2018, doi:10.2174/1872210511666170601115516.
- [86] F. Alvarado-Moreno, "El Papel de las Oficinas de Transferencia Tecnológica (OTT) en las Universidades: Una Perspectiva de la Última Década," *Journal of Technology Management & Innovation*, **13**(3), 104–112, 2018, doi:10.4067/S0718-27242018000300104.
- [87] *Portafolio*, "Colombia cae en clasificación global de los más innovadores," *Portafolio*, 2019.
- [88] J. García, C. Gómez, C. Jiménez, J.C. Castellanos, *La eficiencia de la Innovación en Colombia frente al mundo: Un análisis desde el Global Innovation Index, 2016 - 2019*, 2019.
- [89] *Minciencias, Transferencia de conocimiento, una estrategia para potencializar la investigación*, Minciencias, 2016.
- [90] *CONPES, Política nacional de ciencia, tecnología e innovación 2015 - 2025*, 2015.
- [91] *DNP, Plan Nacional de Desarrollo 2018 - 2022*, <https://www.dnp.gov.co/DNPN/Paginas/Plan-Nacional-de-Desarrollo.aspx>, 2018.
- [92] *Agrosavia, MADR, Minciencias, PECTIA: Plan Estratégico de Ciencia, Tecnología e Innovación del sector Agropecuario colombiano (2017-2027)*, 2016.
- [93] L. Guarín, E. Romero, L.E. Becerra, "Agrópolis de Santander Magdalena - Medio en Colombia: Alternativa para el desarrollo rural en las regiones," in *Congreso Internacional de Investigación e Innovación - Multidisciplinario*, 7382–7393, 2018.
- [94] G. Abramo, C.A. D'Angelo, F. Di Costa, M. Solazzi, "The role of

information asymmetry in the market for university–industry research collaboration,” *The Journal of Technology Transfer*, **36**(1), 2011, doi:10.1007/s10961-009-9131-5.

- [95] P. Gubitta, A. Tognazzo, F. Destro, “Signaling in academic ventures: the role of technology transfer offices and university funds,” *The Journal of Technology Transfer*, **41**(2), 2016, doi:10.1007/s10961-015-9398-7.
- [96] A.C. Schulze-Krogh, “Firms’ Absorptive Capacity for Research-Based Collaboration—an Analysis of a Norwegian R&D Brokering Policy Program,” *Science and Public Policy*, **45**(4), 2018, doi:10.1093/scipol/scx081.
- [97] M. Cattaneo, M. Meoli, S. Vismara, “Cross-border M&As of biotech firms affiliated with internationalized universities,” *The Journal of Technology Transfer*, **40**(3), 2015, doi:10.1007/s10961-014-9349-8.

The Influence of Online Learning on Learning Interest and Motivation and Their Impact on Student Achievement at Educational Technology Study Program – Ibn Khaldun University Bogor

Mohammad Givi Efgivia^{*1}, Al-Bahra², Abdul Karim Halim³, R Andi Ahmad Gunadi⁴

¹Postgraduate Educational Technology, University of Ibn Khaldun, Bogor, Indonesia

²Master of Informatics Engineering Study Program, Raharja University, Tangerang Indonesia

³Faculty of Teacher Training and Education, Ibn Khaldun University, Bogor Indonesia

⁴Master of Educational Technology University of Muhammadiyah Jakarta

ARTICLE INFO

Article history:

Received: 01 August, 2021

Accepted: 26 October, 2021

Online: 30 November, 2021

Keywords:

Online Learning

Interest

Motivation

Student Achievement

ABSTRACT

The aim of this research was to prove the effect of online learning on interest and motivation to learn and its effect on student achievement in the Educational Technology Study Program - Ibnu Khaldun University, Bogor. This research method uses descriptive analysis and verification with a quantitative approach. The source of this research uses primary data. Collecting data using a questionnaire. The sample of this study was 40 students at the Education Technology Study Program IK University Bogor by performing both descriptive and path analysis for the random sampling technique. of this study indicate that online learning has a positive and significant effect on interest and motivation to learn. It is known that the total effect of online learning on students' interest in learning is 79.03%, while the total effect on student learning motivation is 79.74%. Interest in learning has Both interest in learning and learning motivation have contributed on student achievement which are measured by 29.6% and 51.07% respectively, with a total effect of 29.61%. While learning motivation has a positive and significant effect on student achievement, with a total effect of 51.07%.

1. Introduction

Coronavirus is an infectious disease, which spreads very quickly in Indonesia. Corona Virus 19 is a new disease that has been transmitted from Wuhan, China, in December 2019. This disease spreads in animals, however the virus can also be transmitted to humans [1]. Indonesia's Ministry of Education and Culture has instructed, that one way to implement the physical distancing program is to apply online learning to suppress the sudden rise of COVID disease. Both students and education practitioners are asked to switch the face-to-face learning process to online [2]

In light of the ongoing covid health crisis, educators and students are in a state of uncertainty. In the conditions of uncertainty online learning, of course lots of questions arise regarding lots of question regarding to the deliverable and measurable impacts for both students and education practitioners. One of the fundamental questions for us in this research is student motivation in a new virtual platform that applies to all stakeholders. Understanding the variety that will impact student perceptions can also help administrators to make informed decisions about coursework and organizational design as well as hold professional development workshops for instructors.

^{*}Corresponding Author: Mohammad Givi Efgivia, Email: mgivi@uika-bogor.ac.id

This online learning policy applies to all educational institutions. This is done by changing the studying process which initially was in a personal approach to a remote online study. This significant adjustment has become a debate about quality. Several research in the past mentioned the significant cost of the online learning practice. Although other studies stated that it is disadvantageous, online learning is found to have many positive impacts which are shown by the interaction within the learning process and the flexibility of time and location, while other studies stated that it was a loss. Moreover, this is seen as an opportunity to adapt to an online learning experience [3]

In addition, high interactions also occur in In addition, high interaction also occurs in the digital space between educators and participants, and among participants which gives rise to the ability to share information and bread experiences [4]. The online learning process uses computer technology, this increases the interest of both educators and participants which also implies to the increase of computer literacy competence.

The digital study focuses on learning materials and deliverables. First, the learning materials affect student learning outcomes. This should involve distance learning This should involve distance learning according to the information technology by the cloud service with various medium of

deliverables such as audio and visuals. Content capacity can be adjusted to the needs of students in bringing out their competence to understand the learning material and be ready to move on to new topics [5]. Second, instruction is done online where students can explain their assignments directly. Similarly, the deliverable is aligned to the needs of pedagogy, both hardware equipment and the communication rules will be aligned to the policy of the education provider [6].

Instructions are given by teachers online as well as video conferencing using popular video conference applications and other messaging services. These platforms must be accessible from any time and anywhere.

Several studies on student learning motivation have been carried out, including by examining the influence of learning methods and learning motivation, student learning motivation on the learning process, the influence of the internet on student learning outcomes, factors that influence student learning motivation, the contribution of learning media in increasing student learning motivation. , information technology-based learning media and much more. However, research on the effect of online learning on student learning motivation is rarely done [7].

Learning motivation is an individual learner's opinion about the desire, then the learner will present different knowledge acquisition needs because of different opinions. Learning motivation is a belief that is embedded in the heart in directing individual learning goals, encouraging learning behavior to make continuous efforts, strengthening the history of cognition, and strengthening and improving learning outcomes [8].

Learning motivation is an individual learner's opinion about the desire, then the learner will offer a various processes of extracting, structuring, and organizing knowledge from one source due to a different personal need. A drive to learn is widely known coming from the heart in directing individual learning goals, encouraging learning behavior to make continuous efforts to learn, this is a student's choice of certain learning activities and efforts in these activities. Thus, this study defines learning motivation as guiding students to continue learning and efforts at learning goals set by the teacher in the learning process [9].

For online-learning environments [10] proves that students' positive emotions in various online learning activities are higher than negative emotions, especially during synchronous activities with teachers and with peers. They also proved that positive emotions experienced during exam preparation are closely related to student motivation which supports the learning process and student learning outcomes.

The emotion of student achievement is related to the domain to a large extent. However achievement emotions were analyzed in different content domains (such as school subjects, mathematics) [11], [12], research in teacher education broadly did not emerge. For example, there have been several studies on pre-service primary school teacher-specific math anxiety and student emotions in teaching internships [13].

In this study, the authors discuss the internal psychological factors, namely interest and motivation. Interest is a powerful motivation that energizes learning, guides academic and career trajectories, and is critical to academic success. Interest is a psychological state of attention and influence on a particular

object, and tends to last a long time to re-engage from time to time [14].

Recent research about brain functions demonstrates the value of interest and motivation in the development of academic skills, specifically for secondary students. FMRI have allowed researchers to study the activation of different areas of the brain during specific tasks. This research suggests that students need to be interested and motivated to learn in order for true learning, or activation of necessary brain regions, to take place [15]. On the other hand, students under stress or faced with a perceived threat are known to show decreased interest and motivation. The human brain is wired for survival, and, when presented with a perceived threat, the brain will transform into a reflex survival response.

This research aims to show the implication of the online education activity on learning enthusiasm and motivation and its impact on student achievement at Ibn Khaldun University Bogor majoring in Educational Technology

2. Method

The method in this research is a descriptive and verification method with a quantitative approach, because of the variables that will be examined in relation to the purpose of providing a structured, reliable, and carefully the description of the facts and the relationship between the variables studied online learning on interest and learning motivation and its impact on student achievement at Educational Technology Study Program – Ibn Khaldun University Bogor.

The type of data in this study used primary data, the data processed and used in this study were sourced from questionnaires/questionnaires. The sample for the study was 40 students at School X using a simple random sampling technique. The verification analysis used in this study is path analysis. Path analysis has a close relationship with Multiple Regression.

3. Result

3.1. Descriptive Analysis

Based on the descriptive analysis carried out, the findings of the percentage score of each online learning variable, learning interest, learning motivation and student achievement are shown in Table 1.

Based on the descriptive analysis above, it shows that online learning, learning interest, learning motivation and student achievement are categorized as sufficient. The indicator which has the highest percentage score at online learning variable, about using of self-learning materials with a percentage of 68.5%, but regarding the utilization of electronic technology services has the lowest percentage, which is 64.8% this is because there are still students who are constrained by internet signals, internet quotas or inadequate gadgets.

In the interest in learning variable, the indicator that has the highest percentage score is about having curiosity with a percentage of 68.5%, but regarding perseverance and enthusiasm in learning has the lowest percentage, which is 66%. While on the motivation variable, the indicator that has the highest percentage score regarding the motivation to learn and the hopes in the future with a percentage of 67.5%, but there are still students who lack motivation or desire to succeed.

Table 1: Descriptive Analysis

No.	Indicator	Actual Score	Ideal Score	%	Information
Online Learning (X)					
1	Utilization of electronic technology services	259	400	64.8%	Enough
	Utilization of computer media, such as computer networks or digital media.	263	400	65.8%	Enough
	Using learning materials to be studied independently (self learning materials).	274	400	68.5%	Good
	Take advantage of the learning schedule, curriculum, learning progress results and matters relating to education administration can be viewed at any time on the computer.	268	400	67.0%	Enough
	Total Online Learning	1064	1600	66.5%	Enough
Interest in Learning (Y1)					
2	Perseverance in learning.	267	400	66.8%	Medium
	Awareness in learning.	270	400	67.5%	Medium
	Enthusiasm.	267	400	66.8%	Medium
	Enjoy the lessons taught.	276	400	69.0%	High
	Have curiosity.	277	400	69.3%	High
	Total	1357	2000	67.9%	Medium
Interest to learn					
3	Learning Motivation (Y2)	264	400	66.0%	Medium
	There is a desire and desire to succeed	270	400	67.5%	Medium

In the Student Achievement variable, the parameter with the highest percentage score is in the Affective domain with a percentage of 67.5%, but the psychomotor domain has a low percentage score of 64% this is due to the lack of student activity due to distance/online learning during the pandemic. This shows that the perceptions of students in Educational Technology Study Program regarding Online Learning, Learning Interest, Learning Motivation and Student Achievement are considered less than optimal.

The next analysis is verification analysis, The results of the analysis can determine the impact of online learning on interest and learning motivation and its implications for Student Achievement in Educational Technology Study Program A series of data analysis processes using path analysis include testing the normality assumption, analysis of path coefficients and coefficients of determination, analysis direct and indirect effects and hypothesis testing.

3.2. Normality Assumption Test

Based on the table of Kolmogorov Smirnov test results, it can be seen that the significance value of the unstandardized residual in the substructure model 1, 2 and 3 has a significance greater than 0.05, thus it can be proved that the data is normally distributed.

Table 2: Kolmogorov Smirnov Test

One-Sample Kolmogorov-Smirnov Test				
		Unstandardized Residual	Unstandardized Residual	Unstandardized Residual
N		35	35	35
Normal Parameters ^{a,b}	Mean	.0000000	.0000000	.0000000
	Std. Deviation	6.13382878	6.40788798	5.97092413
Most Extreme Differences	Absolute	.141	.126	.081
	Positive	.096	.099	.078
	Negative	-.141	-.126	-.081
Test Statistic		.141	.126	.081
Asymp. Sig. (2-tailed)		.078 ^c	.176 ^c	.200 ^{c,d}

3.3. Path Coefficient Analysis and Coefficient of Determination

Sub Structure 1: Online Learning on Learning Interests

Path coefficient analysis will analyze the causal relationship between online learning and interest in learning.

Start from results of the questionnaires which have been distributed assisted with SPSS, the following results are obtained:

Table 3: Testing the Coefficient of Sub-Structural Paths 1

Model	Standardized Coefficients	Beta	t	Sig.
	Pembelajaran Online	.889	11.949	.000

The structural model based on the standardized coefficients of Online Learning on Learning Interest is as follows:

$$Y_1 = 0.889(X) + \epsilon_1 \tag{1}$$

The magnitude of the residual coefficient (ϵ_1) and the magnitude of the effect can be known through the calculation results as follows:

Table 4: Coefficient of Determination of Substructure 2

Model Summary ^b				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.889 ^a	.790	.784	3.52822

a. Predictors: (Constant), Pembelajaran Online

b. Dependent Variable: Minat Belajar

Based on the table above, it can be seen that the influence of Online Learning on Interest in Learning has a coefficient of determination of 0.790. These results indicate that the online learning variable has an influence of 79% on learning interest, while the remaining 21% is influenced by other variables outside the known model.

Sub Structure 2: Online Learning on Learning Motivation

Path coefficient analysis would like to analyze the causal connection between online learning and learning motivation. After Seeing the results of the questionnaires which have been distributed assisted with SPSS, the following results are obtained:

Table 5: Testing the Path Coefficient of Sub Structure 2

Model	Standardized Coefficients Beta	t	Sig.
Pembelajaran Online	.893	12.202	.000

The structural model based on the standardized coefficients of Online Learning on Learning Motivation is as follows:

$$Y_2 = 0.893(X) + \epsilon_2 \tag{2}$$

The magnitude of the residual coefficient (ϵ_2) and the magnitude of the effect can be known through the calculation results as follows

Table 6: Coefficient of Determination of Substructure 2

Model Summary ^b				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.893 ^a	.797	.791	2.68370

a. Predictors: (Constant), Pembelajaran Online
 b. Dependent Variable: Motivasi Belajar

From the table above, seems like the influence of Online Learning on Learning Motivation has a coefficient of determination of 0.797. These results have proved that the online learning variable has an effect of 79.7% on learning motivation, while the remaining 20.3% is influenced by other variables outside the known model.

Sub Structure 3: Interest and Learning Motivation on Student Achievement

Path coefficient analysis would mind analyzing the causal connection between Interest and Learning Motivation on Student Achievement. After seeing the results of the questionnaires which have distributed assisted with SPSS, the following results are obtained:

Table 7: Testing the Path Coefficient of Sub Structure 3

Model	Standardized Coefficients Beta	t	Sig.
Minat Belajar	.344	2.118	.041
Motivasi Belajar	.577	3.550	.001

The structural model based on standardized coefficients of Interest and Learning Motivation on Student Achievement is as follows:

$$Z = 0,344 (Y_1) + 0.577 (Y_2) + \epsilon_3 \tag{3}$$

The magnitude of the residual coefficient (ϵ_3) and the magnitude of the effect can be known through the calculation results as follows:

From table 8, it can be proved that the influence of Interest and Learning Motivation on Student Achievement has a

Table 9: Direct and Indirect Effect

Variable	Direct Effect	Indirect Effect	Total Effect	R-Squared
X → Y1	(0.889) ² = 79,03%	-	79,03%	79,03%
X → Y2	(0.893) ² x 100% = 79,74%	-	79,74%	79,74%
Y1 → Z	(0.344) ² x 100% = 11,83%	0.344 x 0.869 x 0.577 x 100% = 17,78%	29,61%	80,8%
Y2 → Z	(0.577) ² x 100% = 33,29%	0.577 x 0.952 x 0.344 x 100% = 17,78%	51,07%	

coefficient of determination of 0.808. These results indicate that the variables of Interest and Learning Motivation have an influence of 80.8% on student achievement, while the remaining 19.2% is influenced by other variables outside the known model.

Table 8. Coefficient of Determination of Substructure 2

Model Summary ^b				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.899 ^a	.808	.797	2.04500

a. Predictors: (Constant), Motivasi Belajar, Minat Belajar
 b. Dependent Variable: Prestasi Belajar

3.4. Direct and Indirect Effect

Based on the three equations above, the path of each variable is used to find the magnitude of the direct and indirect influence on each independent variable on the dependent variable, it could be seeing in below:

From table 9, we find that the online learning variable has a direct consequence on influencing learning motivation of 79.74%, which is greater than its direct effect on learning interest, which is 79.03%. Interest in learning has a direct effect of 11.83% smaller than the indirect effect of 17.78% on student achievement. Meanwhile, learning motivation has a direct effect of 33.39% greater than the indirect effect of 17.78% on student achievement. Therefore, it shows that in influencing student achievement, learning motivation has a higher influence than learning interest. Finally, we are able to say that if online learning is carried out properly supported by high interest and motivation in students, it will have an effect on high student achievement.

4. Discussion

4.1. The Effect of Online Learning on Interest in Learning

The Impact of online learning on interest in learning give the results of testing the first hypothesis, namely online learning on Interest in Learning, obtained a significant value (0.000) < 0.05 so that H₀ is denied, so we are able to say that online learning has a positive and important effect on Interest in Learning, with a positive direction indicating that it will be better learning done online, When students' interest in learning is getting higher, and conversely the worse the learning done online it is going to have an impact on the low interest in learning.

Research that has been carried out is in accordance with previous research which states that Interest plays a very important role in learning. Interest will lead to motivation to learn and improve learning outcomes. The strategy used by teachers in increasing students' interest in learning is to provide students with an understanding of the importance of learning, make learning materials simple, clear, and interesting, use interesting media, then carry out regular and continuous learning[1].

It was found that there were differences in learning interest between male and female students where girls had a higher interest in learning than boys [16].

4.2. *The Effect of Online Learning on Learning Motivation*

The results of testing the second hypothesis, namely Online Learning on Learning Motivation, obtained a significant value. $(0.000) < 0.05$ so H_0 is denied, it can be say that online learning has a positive and important effect on Learning Motivation, with a positive direction indicating that it will be better online learning, the higher student's motivation, and vice versa. Bad learning is done online, it is going to have an impact on the low learning motivation.

The findings of this study are reinforced in 2019. The use of specific learning strategies did not strongly relate to student performance in their course. Given the increasing popularity of online courses and programs, more attention should be paid to the way students approach and experience these courses, how student characteristics (such as different types of motivation) and activities (the way they learn and master the material) are closely related to their success [17].

The results of this study are not in accordance with the research conducted in 2017. The study proved that participants' extrinsic and intrinsic motivation were lower in online education when compared to face-to-face education and the researchers gave the reason students were less motivated in online courses was the lack of interaction with peers and teachers. Students stated that the higher level of motivation was due to the satisfying role of the teacher and adequate communication with the teacher [18]. Yantrapracorn et al. (2018) also stated that the negative impact of online learning courses in Thailand was discouraging teacher feedback [19].

The results of this study are not in accordance with the research conducted in 2021. Students' views are mostly negative in terms of online education based on the results of semi-structured interviews and students' creative writing assignments. The participants expressed various reasons for their lack of motivation in the online learning process including being dissatisfied with course content and materials, lack of discipline in attending courses, lack of communication with teachers and other students, in addition they stated that there was a lack of private space to take courses [20].

Motivation has proven to be an important, though not the most essential, element in supporting academic success. As many and as good as the tools and methods given to students to learn, there will be no progress in them, if they are not motivated to apply them [21].

4.3. *The Influence of Learning Interest on Student Achievement*

The results of testing the third hypothesis, namely Learning Interest on Student Achievement, obtained a significant value $(0.000) < 0.05$ so that H_0 is denied, can be say that Learning Interest has a positive and significant effect on Student Achievement, with a positive direction indicating that the higher students' interest in learning it will have an impact on high student achievement, and conversely the lower student interest in learning are going to have an impact on low student achievement.

The result of research in 2017 is a close relationship between student achievement and the interest of learning and

student learning motivation, high learning achievement is resulted from the interest and motivation to learn that is owned by a student [22]. There is a mutual influence on learning motivation, learning interest, family environment and learning models together on student achievement in the Department of Economic Education of UIN Suska Riau [23].

This finding is in accordance with the research conducted in 2019. There is a significant effect of interest in learning and perceptions of teacher competency on social studies learning achievement. Thus, the interest in learning and teachers' pedagogic competency plays a very important role to determine and improve the learning achievement of Social Sciences [24].

4.4. *The Effect of Learning Motivation on Student Achievement*

The results of testing the fourth hypothesis, namely Learning Motivation on Student Achievement, obtained a significant value $(0.000) < 0.05$ so that H_0 is denied, wa could say that Learning Motivation has a positive and significant effect on Student Achievement, with a positive direction indicating that the higher students' motivation to learn it will have an impact on high student achievement, and conversely the lower student motivation to learn will have an impact on low student achievement.

The findings are in line with the study conducted in 2017. Motivation has a slightly positive effect on student achievement. The moderator variables, year of publication, type of publication, country, culture of research place, school subjects and sample groups will act as moderators in terms of the influence of motivation on student achievement [25].

The findings are accordance with 2020 study. Learning motivation has a positive and significant effect on the learning achievement of state high school students in the city of Padang. Stimulating learning and providing a more collaborative learning experience is generated through the integration of technology with face-to-face teaching. It is time for higher education to adapt to these changes to keep up with the increasing demands of both students and the workplace. The involvement, motivation, and interaction of students are the determining factors in order to achieve the success of the learning process. Students become more intrinsically motivated, when they can relate what they are learning to real life and personalize it [26].

5. Conclusion

Online learning has a positive and significant effect on learning interest. and Learning Motivation. Learning motivation also has a positive and significant effect on student achievement at the Educational Technology Study Program – Ibn Khaldun University Bogor.

Lecturers are expected to be able to motivate students and create interesting learning, and always pay attention to students in the online learning process that takes place. In addition, lecturers also need to improve professional competence through training activities, seminars, and trainings so that they can use the media professionally in learning activities. Parents of students need to provide support for their children to learn and always pay attention to the condition of the child, accompanying the child in the learning process.

References

- [1] A. Sariol, S. Perlman, "Lessons for COVID-19 immunity from other

- coronavirus infections,” *Immunity*, 2020.
- [2] H. Harisman, M.S.Y. Lubis, M.R.S. Lubis, “The Implementation of Right-Fulfillment to the Education of Elementary School Students in the COVID-19 Pandemic Situation,” *Randwick International of Social Science Journal*, **2**(4), 476–484, 2021. DOI: 10.47175/rissj.v2i4.328
- [3] F. Firman, S. Rahayu, “Pembelajaran online di tengah pandemi covid-19,” *Indonesian Journal of Educational Science (IJES)*, **2**(2), 81–89, 2020.
- [4] R. Pakpahan, Y. Fitriani, “Analisa pemanfaatan teknologi informasi dalam pembelajaran jarak jauh di tengah pandemi virus corona covid-19,” *Journal of Information System, Applied, Management, Accounting and Research*, **4**(2), 30–36, 2020.
- [5] D. Alabbasi, “Exploring graduate students’ perspectives towards using gamification techniques in online learning,” *Turkish Online Journal of Distance Education*, **18**(3), 180–196, 2017.
- [6] Y. Shi, “Handbook on Facilitating Flexible Learning During Educational Disruption: The Chinese Experience in Maintaining Undisrupted Learning in COVID-19 Outbreak,” *Bulletin of the Technical Committee on Learning Technology (ISSN: 2306-0212)*, **20**(2), 36–38, 2020.
- [7] R.S.I. Kawet, “Pengaruh metode pembelajaran dan minat belajar mahasiswa terhadap hasil belajar manajemen konstruksi,” *JTP-Jurnal Teknologi Pendidikan*, **19**(3), 224–239, 2017.
- [8] K. Shabani, “Dynamic assessment of L2 learners’ reading comprehension processes: A Vygotskian perspective,” *Procedia-Social and Behavioral Sciences*, **32**, 321–328, 2012.
- [9] L. Koff, R. Mullis, “Nutrition education and technology: can delivering messages via new media technology effectively modify nutrition behaviors of preschoolers and their families?,” *Journal of Nutrition Education and Behavior*, **43**(4), S40, 2011. DOI: 10.1016/j.jneb.2011.03.120
- [10] F. D’Errico, M. Paciello, L. Cerniglia, “When emotions enhance students’ engagement in e-learning processes,” *Journal of E-Learning and Knowledge Society*, **12**(4), 2016.
- [11] K. Musnah “Zero Waste] dalam Mekanisme Bersih Pengelolaan Sampah Kota Tanpa TPA, 2015.
- [12] M. Bieg, T. Goetz, F. Sticca, E. Brunner, E. Becker, V. Morger, K. Hubbard, “Teaching methods and their impact on students’ emotions in mathematics: an experience-sampling approach,” *ZDM*, **49**(3), 411–422, 2017. DOI: 10.1007/s11858-017-0840-1
- [13] R. Yuan, I. Lee, “I need to be strong and competent’: A narrative inquiry of a student-teacher’s emotions and identities in teaching practicum,” *Teachers and Teaching*, **22**(7), 819–841, 2016. DOI: 10.1080/13540602.2016.1185819
- [14] J.M. Harackiewicz, J.L. Smith, S.J. Priniski, “The Importance of Promoting Interest in Education. Policy Insights Behav,” *Brain Sci*, **3**, 220–227, 2016.
- [15] S. Sutarto, D.P. Sari, I. Fathurrochman, “Teacher strategies in online learning to increase students’ interest in learning during COVID-19 pandemic,” *Jurnal Konseling Dan Pendidikan*, **8**(3), 129–137, 2020. DOI: 10.47175/jkdp.v2i4.328
- [16] S. Rodiah, W. Sopandi, “The Effect of Online Learning on Learning Interest and Digital Literacy of Students in terms of Gender (Gender-Based Comparative Causal Study of Students’ Learning Interest and Digital Literacy in Distance Learning Due to the Covid-19 Primary School Pande,” in *International Conference on Elementary Education*, 233–237, 2021.
- [17] E. Stark, “Examining the Role of Motivation and Learning Strategies in Student Success in Online versus Face-to-Face Courses.,” *Online Learning*, **23**(3), 234–251, 2019.
- [18] C.-H. Lin, Y. Zhang, B. Zheng, “The roles of learning strategies and motivation in online language learning: A structural equation modeling analysis,” *Computers & Education*, **113**, 75–85, 2017.
- [19] P. Yantraprakorn, P. Darasawang, P. Wiriyakarun, “Self-efficacy and online language learning: causes of failure,” *Journal of Language Teaching and Research*, **9**(6), 1319–1329, 2018.
- [20] M. Esra, Ç. Sevilen, “Factors influencing EFL students’ motivation in online learning: A qualitative case study,” *Journal of Educational Technology and Online Learning*, **4**(1), 11–22, 2021.
- [21] F. Koca, “Motivation to Learn and Teacher-Student Relationship.,” *Journal of International Education and Leadership*, **6**(2), n2, 2016.
- [22] D. Hude, M. Si, I.F. Rohmah, “Analysis of Student Learning Interest and Student Learning Motivation in Enhancement Student Learning Achievement at School,” *International Journal of Current Research*, **9**(10), 58981–58985, 2017.
- [23] L. Lastri, S. Kartikowati, S. Sumarno, “Analysis of Factors that Influence Student Learning Achievement,” *Journal of Educational Sciences*, **4**(3), 679–693, 2016. DOI: 10.1080/13542.2016.11819
- [24] S.O. Özen, *The effect of motivation on student achievement*, Springer: 35–56, 2017.
- [25] A. Susanto, S. Imawati, L. Qurrotaini, “The Effect of Learning Interest and Student’s Perception of Teacher’s Pedagogic Competency on Learning Achievement of Social Sciences: Survey at Labschool Elementary School, Faculty of Education, University of Muhammadiyah Jakarta,” in *6th International Conference on Community Development (ICCD 2019)*, Atlantis Press: 33–35, 2019.
- [26] R. Rafiola, P. Setyosari, C. Radjah, M. Ramli, “The Effect of Learning Motivation, Self-Efficacy, and Blended Learning on Students’ Achievement in The Industrial Revolution 4.0,” *International Journal of Emerging Technologies in Learning (IJET)*, **15**(8), 71–82, 2020.

Innovations in Recruitment—Social Media

Lucie Böhmová*, Antonín Pavlíček

Department of Systems Analysis, Prague University of Economics and Business, Prague, 130 67, Czech Republic

ARTICLE INFO

Article history:

Received: 16 July, 2021

Accepted: 02 October, 2021

Online: 30 November, 2021

Keywords:

Human Resources Management

Innovations in Recruitment

Recruitment

Social Networks

Data Analysis

Social Network Analysis

Model

ABSTRACT

The main objective and contribution of the paper is to describe the creation of a model to support recruitment using social media information and its deployment in practice. The model includes the design of an automated solution for downloading social media data and a proposal for the subsequent analysis and creation of a predictive model based on the evaluation of user behavior on a social network. A final assessment of the effectiveness of the proposed model was made through formal validation and a case study. For the case study validation, the proposed model was implemented through use of the recruitment application called *Prace Na Miru (PM; Tailored Work)* for Facebook data extraction. A Myers Briggs Type Indicator (MBTI) personality test was used to determine the predictors of user social network behavior. Using cluster analysis and machine learning (or decision trees), a stochastic predictive model was developed to determine the personality type of candidates—the accuracy of MBTI personality category prediction ranges between 68% and 84% for individual cases, with a confirmation rate ranging between 43% and 81%. The case study confirmed the model's usefulness for supporting recruitment in a real-life deployment of the PM model.

1. Introduction

At the beginning of the twenty-first century, communication platforms that allow users to interact with each other have proliferated globally. These platforms gradually began to replicate the social structures of the real world. Within barely five years (2001–2005), social networks as we understand them today took on an established form.

These networks are no longer mere communication channels but are permeating more and more areas and blurring the boundaries between private and professional life. Social networks have also shown their potential in human resources management (HRM), especially in recruitment. Research [1] has found that 43% of organizations use social media sites to research candidates, and approximately 50% of employers report rejecting job applicants because of social media content. [2]

The current state of the labor market is not very favorable for recruitment from the point of view of organizations struggling to find suitable employees. Low unemployment and high demand for employees are forcing organizations to look for new and more

effective recruitment solutions. In this respect, the use of social networks brings an advantage in two ways: they are economically advantageous, and they are an innovative solution [3]-[5].

Social media such as Facebook, Twitter, and LinkedIn are used by recruiters to attract talent by sharing the company culture and information related to career opportunities. For example, Microsoft has developed a specific website for recruitment, using job blogs and life at work videos to provide relevant information to both active and passive job seekers. Similarly, Deloitte places employee testimonials on social media (e.g., YouTube videos) to allow potential job applicants to learn more directly from the insiders. A shortcoming is the question of whether these job messages can reach the target audiences [6].

The results of research [7] confirmed that recruiters intended to use social media as it provides pre-hire benefits. Similarly, results showed social media recruitment is more compatible and less complex, and it ensures better trialability and observability. The study proved that social media recruitment was cost-effective while attracting and retaining better talent [8].

The combination of HRM and social networks in the context of recruitment issues expects an interdisciplinary approach based on applied informatics (as it is primarily based on the use of ICT),

*Corresponding Author: Lucie Böhmová, lucie.bohmova@vse.cz

but its implications extend to the fields of management and human resources (HR) [4].

In the field of applied informatics, our research primarily uses data analysis of social networks through our own designed application called **Prace Na Míru** (abbreviation **PM**; literal translation into English—Tailored Work), which allows for automated data acquisition. It is further related to applied informatics by the selected research design, called “design science research” [9], and the resulting artifact, which is a model verified not only by meeting the theoretical assumptions but, largely, by its application in practice [10].

1.1. Research problem

The generally defined research problem is the support of recruitment in modern HRM through the use of social networks.

The authors propose a solution in the form of a recruitment support model—the PM model.

1.2. Objectives and research questions

The main objective was to create and verify the effectiveness of a model for supporting recruitment through social networks which would be generally valid and applicable to any type of organization and job position.

In order to achieve the main objective, sub-objectives (SOs) and sub-research questions (SRQs) were identified as follows:

- **SO 1:** Create and deploy the PM model to support recruitment using a specific social network within a selected problem context and validate its effectiveness.
- **SRQ 1:** What is the design of the PM model and the process of its deployment in practice?
 - **SO 1.1:** Create a custom application to extract data from a specific social network.
 - **SRQ 1.1:** How can we create an application to mine data from social networks?
 - **SO 1.2:** Identify appropriate parameters to evaluate user behavior on social networks.
 - **SRQ 1.2:** What are the appropriate parameters to evaluate user behavior on social networks?
 - **SO 1.3:** Identify predictors of user behavior on social networks using appropriate analytical methods.
 - **SRQ 1.3:** What are the predictors of user behavior on social networks? And what are the appropriate analytical methods to determine them?

1.3. Methods used

In this paper, the authors used methods according to [10] and [11] in the publication *Design Science Research*, which consists of five phases:

- **Phase 1—problem awareness:** This is based on an analysis of the current state of knowledge on the issue, especially best practices and literature research. The output is the basis for the design of the artefact.

- **Phase 2—design/theme:** the result is a general tentative design of the artefact in Figure 1.
- **Phase 3—development:** the development of the indicative design of the artefact from phase 2 into the final form of the artefact, the result of this phase.

In terms of conceptual development, the development and design/idea phases are intertwined—the case here.

- **Phase 4—evaluation:** the artifact is evaluated in order to obtain information indicative of the artifact’s behavior and quality and useful to the artifact’s future development. Scientific methods are used in the evaluation of the artefact—formal verification—by creating an instance in the form of the PM model. These scientific methods are of three types:

- data analysis;
- cluster analysis;
- predictive modelling using decision trees (or machine learning).

Validation is done through quantitative research in the form of a questionnaire survey among experts in practice.

Evaluation is done through qualitative research in the form of a case study and subsequent practical deployment.

- **Phase 5—Conclusion:** the last phase summarizes the research’s conclusions.

This paper is primarily concerned with description of the fourth phase.

Phase 1—Problem Awareness (Background)

The main starting point for the development of the model is a recruitment application for extracting user data from a social network (or custom work).

Another important starting point for model construction is the parameters for evaluating user behavior on the social network. In the context of recruitment, the most appropriate parameters are those based on a personality test.

Therefore, it was necessary to specify requirements for selecting an appropriate test according to the purpose of the PM model, which are [2, 12] as follows:

- assessment of personality characteristics;
- assessment of interpersonal characteristics;
- evaluation of work characteristics;
- relevance for recruitment;
- speed;
- transparency;
- the ability to complete the online test from anywhere;
- immediate evaluation without the need for multiple specialist (e.g., psychologist) costs.

There are several psychological models and personality tests for assessing jobseekers that can serve as parameters to evaluate

behavior [13]-[16]. Classic tests include the Eysenck’s Personality Inventory (EPI), which is primarily aimed at identifying temperament based on measures of introversion (or extroversion) and stability (or lability); the Achievement Motivation Inventory (AMI), used to measure various dimensions of performance motivation in a professional context and useful in the selection of employees, team creation, or sport psychology; the Personal Competence Profile, which takes into account findings on emotional intelligence and social competence; and the Myers-Briggs Type Indicator (MBTI) model of personality types, which is widespread among organizations and is mainly used by HRM in recruitment, by health professions, and by educational programs to address the diversity of personalities that exist [2], [17].

Based on the analysis of the abovementioned characteristics, the MBTI personality test was chosen as the most suitable and was therefore used in the PM model [17]-[19]. The MBTI personality test meets all the above requirements—this is explored in more detail in section 1.6 “Predictors for user behavior evaluation on social media.”

This implies the need to identify predictors using appropriate analytical methods to assess user behavior on the social network. The most suitable predictors in this work were found to be areas of interest, such as favorite book, TV series, film, or sports personality.

The PM model development procedure is based on a model abstraction technique which allows the projection of alternatives through a model experiment. This, in effect, helps to track changes in individual parameters and their relationships at the desired level of simplification. The model is then used to explore, in particular, relationships, causes, and analogies. The methodological framework for the development of the recruitment support model is based on the CRISP-DM methodology [20].

The fundamental constraints in the creation of the model are mainly based on legislation. Organizations must comply with the General Data Protection Legislation (GDPR) and other employment legislation when identifying and verifying candidate information on social media. At the same time, it must not engage in discriminatory behavior. Therefore, such sensitive information about candidates is omitted from the model, even though it can be obtained from social networks relatively easily.

1.4. Validity and reliability of MBTI

We know from discussions of psychological and personality tests that critics have pointed out these tests may not be universally valid, and the results may not be reliable. For example, the popular-science journal *Scientific American* has published articles in recent years (e.g., “How Accurate Are Personality Tests?” [21] and “Personality Tests with Deep-Sounding Questions Provide Shallow Answers about the True You” [22]) that question the validity of personality tests.

However, peer-reviewed scientific journals, most recently the *Journal of Best Practices in Health Professions Diversity: Education, Research and Policy*, have published systematic (meta-analysis) reviews of the literature concerning its validity and reliability. The article “Validity and Reliability of the Myers-Briggs Personality Type Indicator: A Systematic Review and

Meta-analysis,” on the contrary, confirms the validity of MBTI tests on a large sample of 221 studies [23], [24].

Authors agree that the MBTI instrument has reasonable construct validity. Their results indicate that the extravert-introvert, sensing-intuition, and judging-perceiving subscales have satisfactory reliabilities of .75 or higher and that the thinking-feeling subscale has a reliability of .61 [23].

The only mentioned problem is that the majority of studies are usually conducted on college-age students; thus, the evidence to support the tool’s utility applies more to this group, and careful thought should be given when applying it to other individuals.

This limitation is not a problem in our research since the respondents in our case were also primarily students and young alumni.

Phases 2 and 3: Design and Development

The authors’ application Prace Na Miru is the essential component in developing the PM model. Running on the website prace-na-miru.eu/en/, the application harvests user data.

The collected data is saved in JSON format within the MongoDB database. Heroku is used to host both the page and the database. JavaScript and Node.js frameworks were used to write the code. Candidates use the Facebook login button to access the website and to approve the download/export of his data into our database.

The sample selection is derived through self-selection among students and graduates of the Prague University of Economics and Business. Information about the application Prace Na Miru was distributed through an email newsletter to the target student and alumni groups. The application was also published on social media in suitable groups, during a university job fair called *Šance*, and on the web page of the college career center. There were 960 unique candidates who registered with and logged in to the application from October 2016 to January 2017. 624 of them responded to the subsequent email invitation to complete the MBTI test and 484 completed the questionnaire in full. This data has been used for further analysis.

1.5. Data analysis

The data collected by the Prace Na Miru application was processed in the SW called Knime [25]. The spectrum of publicly available data was organized into categories. Gender, log-in device, list of friends, date and year of birth, and other details were provided by users. Some of this data, such as favorites, are important for recruitment (or, more accurately, for the construction of the model), for example, favorite TV series, favorite athletes, and so forth. Some basic demographic data about users, such as gender, age, and education, was also available.

Table 1: Correlation of Publicly Accessible User Information to Categories

Categories of publicly available data	Friend lists	TV series	Athletes	Books	Events	Games	Music	Likes	Films
Friend lists	1	0.09	0.11	0.09	0.31	-0.06	0.16	0.29	0.04
TV series	0.09	1	0.17	0.32	0.11	0.23	0.57	0.09	0.57
Athletes	0.11	0.17	1	0.15	0.07	0.25	0.16	0.03	0.12

Books	0.09	0.32	0.15	1	-0.09	0.32	0.27	0.07	0.52
Events	0.31	0.11	0.07	-0.09	1	-0.01	0.14	0.47	0.03
Games	-0.06	0.23	0.25	0.32	-0.01	1	0.11	0.04	0.31
Music	0.16	0.57	0.16	0.27	0.14	0.11	1	0.22	0.49
Likes	0.29	0.09	0.03	0.07	0.47	0.04	0.22	1	0.12
Films	0.04	0.57	0.12	0.52	0.03	0.31	0.49	0.12	1

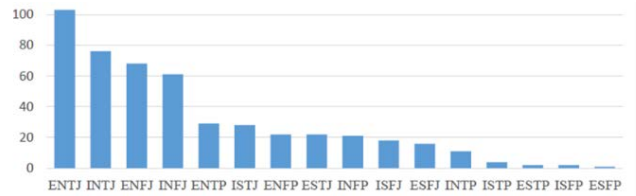


Figure 2: Number of persons represented by MBTI type

Spearman’s correlation coefficient was used to detect correlations within the collected data (Table 1). The existence of low bond tightness for the categories “favorite book” and “series” (0.32) and “favorite game” and “book” (0.32) was proven. Medium-bond tightness was found in the categories “likes” and “events” (0.47) and “favorite movie” and “book” (0.52). Significant-bond tightness was in the categories “favorite music” and “series” (0.57) and “favorite music” and “movie” (0.57). It turns out that users sharing information about their favorite series also share information about their favorite music and movies.

1.7. MBTI test and its evaluation

All users logged into Prace Na Miru were tested by the MBTI test. The response rate at this point was 50.4%. This fact helped to determine personality types and characteristic features of the questioned users and compared them with their online approach to social media. The sample size was 484 unique respondents.

1.8. PM model creation

For the creation of the PM model, the authors used training data (N=960) upon which the model learned how to decide the personality types of the candidates. The result of the modeling phase—Model PM—is shown in Figure 3.

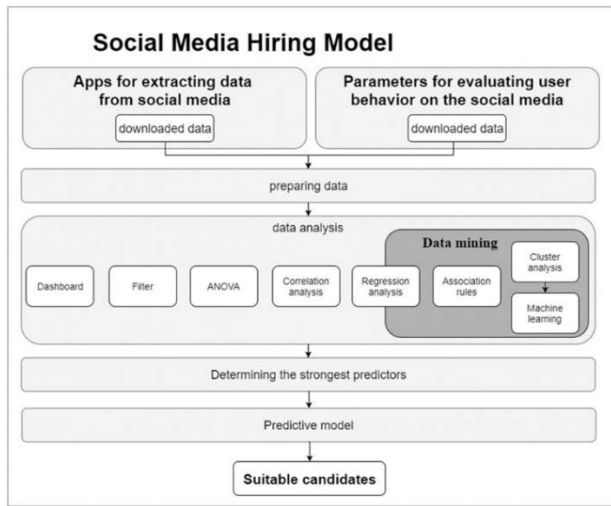


Figure 1: Social media hiring model

1.6. Predictors for user behavior evaluation on social media

An important part of the specification for the construction of a functional PM model is identifying predictors to evaluate user behavior within the social media environment. The MBTI personality test [12], [26], [27] was used in the PM model. The MBTI personality test combines four criteria:

- approach to the surrounding environment—introversion (I) or extroversion (E);
- way of obtaining information—intuition (N) or sense (S);
- way of evaluating information—feeling (F) or thinking (T);
- lifestyle—perception (P) or judging (J).

Merging the four criteria mentioned above results in 16 personality types. Figure 2 shows the number of persons represented by MBTI type. The most represented type is ENTJ, for whom employment in leadership positions such as leader or manager is suitable. This is followed by INTJ (scientist, builder), ENFJ (teacher), and INFJ (writer, creator).

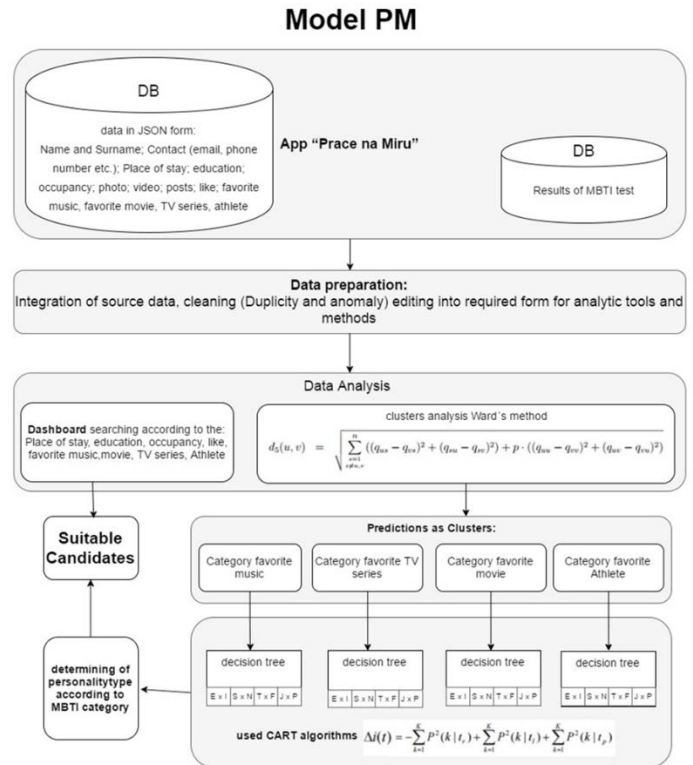


Figure 3: PM model

Phase 4: Model PM Evaluation

The model validation is realized by its implementation in a specific problem context: the recruitment of suitable candidates for positions at the University Development and Counselling Centre and internship programs at the University xPORT Business Accelerator. The target user group upon which the model is validated are students and recent university graduates.

The evaluation is performed by obtaining valuable and reliable information proving the quality of the PM model.

1.9. Formal verification

The formal verification is based on the construction of the PM model. Scientific methods included are as follows:

- data analysis and cluster analysis;
- predictive modelling using decision trees (or machine learning).

In developing the PM model, the authors used the MBTI personality test to determine predictors of user behavior on the social network, which helps to diagnose candidate characteristics such as perception of the environment or the way they acquire and process information. In addition, the PM model includes instructions for an automated solution for downloading user data from Facebook using a customized recruitment application called Jobs. Also included is a proposal for data analytics. Specifically, the PM model describes cluster analysis, decision trees, and predictive model(s) for determining personality type (which is based on the MBTI test).

After preprocessing the data, the authors segmented users into clusters, which are used in the PM model as predictors of behavior according to MBTI categories. Due to the large amount of data, it was necessary to select only some clusters that have predictive power. Out of 28 possible attributes, the author identified four with the highest predictive power as predictors:

- favorite music;
- favorite TV series;
- favorite movie; and
- favorite athlete.

For each attribute, the author created decision trees using the BigML [28] software tool, which allowed for their categorization according to the MBTI personality categories.

For each decision tree, the author includes a graphic illustration of the main tree branch that is the strongest and a ray graph that shows the overall decision tree with the relative representation of number of persons (according to the circle plot) and the confidences for the given personality category. The individual circles represent the path of the tree through the number of occurrences in the clusters. The color represents the amount of confidence of the decision tree. The darker, more saturated the color in the ray graph, the more confident is the decision tree. If confidence is 0%, the result is equivalent to random selection [29], [30].

PM model for the Attribute “Favorite Music”

For the creation of decision trees, clusters were created for individual attributes using the software tool Pajek [31]. As an example, the author presents in Table 2 the first five most frequently occurring items in each cluster.

Table 2: Cluster Overview for the Attribute “Favorite Music”

Cluster A	Cluster B	Cluster C	Cluster D	Cluster E	Cluster F	Cluster G	Cluster H	Cluster I
AC/DC	Bruno Mars	Snoop Dogg	Ed Sheeran	Electro Swing	Walk Off The Earth	Linkin Park	The Beatles	Ben Cristovao
Calvin Harris	Wiz Khalifa	Britney Spears	Fedde Le Grand	OneRepublic	Universal Music CZ	Tomáš Klus	Radiohead	Taylor Swift

Yellow card	Monika Bagárová	Fergie	INNA	Philip Glass	George Ezra	Nightwish	Kings Of Leon	Rihanna
The Prodigy	Adam Levine	T-Pain	Johny Machette	Coldplay	Lindsey Stirling	Apocalyptica	Queens of the Stone Age	Beyoncé
Christina Perri	Tyrese Gibson	Sean Paul	Ellie Goulding	River Hollywoood Undead	Thirty Seconds to Mars	Avril Lavigne	The Black Keys	Kryštof

1.10. Decision tree for category “extroverted (E) or introverted (I) perception of the environment”

Below is the text notation of the predictive model for the category “extroverted (E) or introverted (I) perception of the environment”:

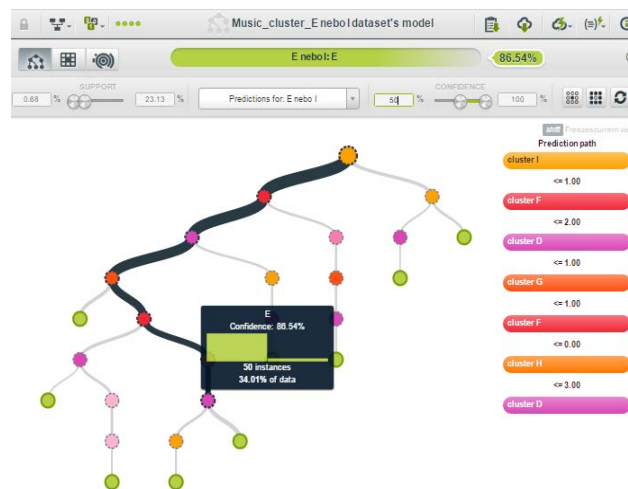


Figure 4: PM model for the attribute “favorite music”—decision tree for category “extroverted (E) or introverted (I) perception of the environment”



Figure 5: PM model for the attribute “favorite music”—decision tree bar graph for category “Perception of the environment extroverted (E) or introverted (I)”

- **Data distribution**
 - E: 70.75% (104 instances)
 - I: 29.25% (43 instances)
- **Predicted distribution**
 - E: 73.47% (108 instances)
 - I: 26.53% (39 instances)
- **Field importance**

1. cluster I: 28.47%
2. cluster F: 23.07%
3. cluster D: 19.41%
4. cluster G: 11.36%
5. cluster C: 5.59%
6. cluster E: 5.16%
7. cluster H: 4.48%
8. cluster B: 2.45%

1.11. Decision tree for category “sensory (S) or intuition (N) information acquisition”

Below is the text notation of the predictive model for the category “sensory (S) or intuition (N) information acquisition”:

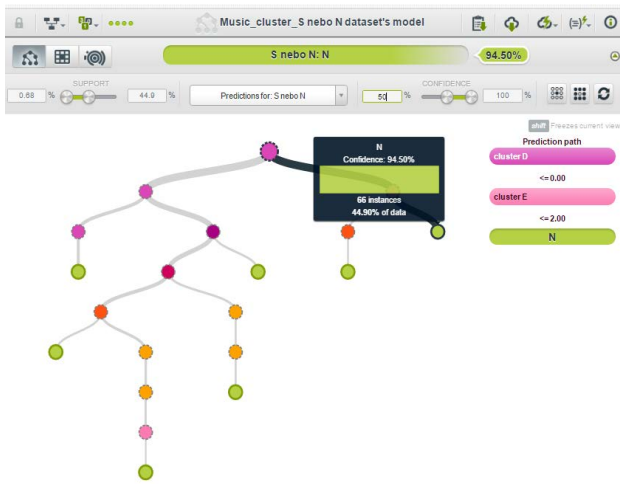


Figure 6: PM model for the attribute “favorite music”—decision tree for category “sensory (S) or intuition (N) information acquisition”

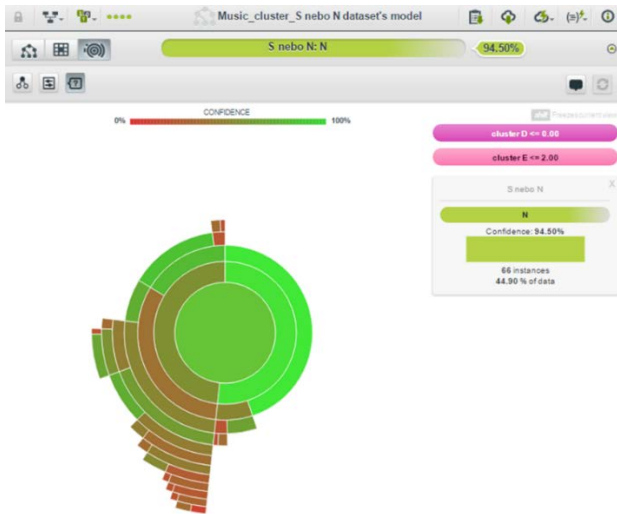


Figure 7: PM model for the attribute “favorite music”—decision tree bar chart for category “Sensory (S) or intuition (N) information acquisition”

- **Data distribution**
N: 82.31% (121 instances)

S: 17.69% (26 instances)

- **Predicted distribution**
N: 82.99% (122 instances)
S: 17.01% (25 instances)
- **Field importance**
 1. cluster D: 47.21%
 2. cluster A: 12.80%
 3. cluster I: 11.08%
 4. cluster B: 9.43%
 5. cluster E: 7.05%
 6. cluster G: 6.97%
 7. cluster C: 2.82%
 8. cluster F: 1.88%
 9. cluster H: 0.75%

1.12. Decision tree for category “thinking (T) or feeling (F) information processing”

Below is the text notation of the predictive model for the category “thinking (T) or feeling (F) information processing”:

- **Data distribution**
F: 49.66% (73 instances)
T: 50.34% (74 instances)
- **Predicted distribution**
F: 48.30% (71 instances)
T: 51.70% (76 instances)
- **Field importance**
 1. cluster F: 18.69%
 2. cluster E: 16.51%
 3. cluster C: 15.07%
 4. cluster G: 14.95%
 5. cluster H: 8.88%
 6. cluster D: 8.50%
 7. cluster I: 6.85%
 8. cluster B: 6.21%
 9. cluster A: 4.32%

1.13. Decision tree for category “inference (J) or perception (P) lifestyle”

What follows is the textual notation of the predictive model for the category “inference (J) or perception (P) lifestyle”

• **Data distribution**

J: 76.87% (113 instances)

P: 23.13% (34 instances)

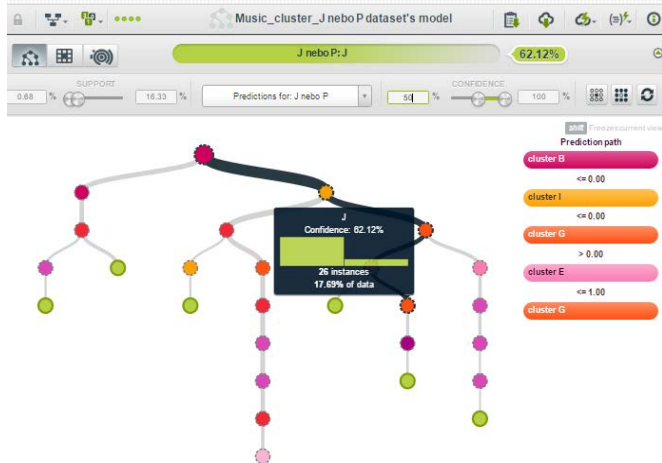


Figure 10: PM model for the attribute “favorite music”—decision tree for category “inference (J) or perception (P) lifestyle”



Figure 11: PM model for the attribute “favorite music”—decision tree bar chart for category “inference (J) or perception (P) lifestyle”

• **Predicted distribution**

J: 76.87% (113 instances)

P: 23.13% (34 instances)

• **Field importance**

1. Cluster D: 17.31%

2. Cluster F: 16.94%

3. Cluster I: 16.50%

4. Cluster B: 14.91%

5. Cluster G: 13.03%

6. Cluster E: 9.45%

7. Cluster C: 7.37%

8. Cluster H: 3.43%

9. Cluster A: 1.06%

2. PM model Validation

The validation of the PM model was carried out on 198 people from the target group.

The results of the individual predictive models are specified below in terms of the percentage of cases confirming the accuracy of the PM model and the average confidence of each predictive model.

2.1. Predictive PM model for attribute “favorite music”:

- Category “extroverted (E) or introverted (I) perception of the environment”

The model confirmed the correct result 83% of the time. The average confidence of the model is 66%.

- Category “sense (S) or intuition (N) information acquisition”

The model confirmed the correct result in 82% of cases. The average confidence of the model is 81%.

- Category “thinking (T) or feeling (F) information processing”

The model confirmed the correct result in 77% of cases. The average confidence of the model is 63%.

- Category “inference (J) or perception (P) lifestyle”

The model confirmed the correct result in 80% of cases. The average confidence of the model is 51%.

2.2. Predictive model PM for attribute “favorite TV series”:

- Category “extroverted (E) or introverted (I) perception of the environment”

The model confirmed the correct result in 80% of cases. The average confidence of the model is 56%.

- Category “sense (S) or intuition (N) information acquisition”

The model confirmed the correct result in 70% of cases. The average confidence of the model is 57%.

- Category “thinking (T) or feeling (F) information processing”

The model confirmed the correct result 74% of the time. The average model confidence is 43%.

- Category “inference (J) or perception (P) lifestyle”

The model confirmed the correct result in 81% of cases. The average confidence of the model is 49%.

2.3. Predictive model PM for attribute “favorite movie”:

- Category “extroverted (E) or introverted (I) perception of the environment”

The model confirmed the correct result in 79% of cases. The average confidence of the model is 56%.

- Category “sense (S) or intuition (N) information acquisition”

The model confirmed the correct result in 84% of cases. The average model confidence is 49%.

- Category “thinking (T) or feeling (F) information processing”

The model confirmed the correct result in 78% of cases. The average confidence of the model is 51%.

- Category “inference (J) or perception (P) lifestyle”

The model confirmed the correct result in 75% of cases. The average confidence of the model is 59%.

2.4. Predictive model PM for attribute “favorite athlete”:

- Category “extroverted (E) or introverted (I) perception of the environment”

The model confirmed the correct result in 71% of cases. The average confidence of the model is 53%.

- Category “sense (S) or intuition (N) information acquisition”

The model confirmed the correct result in 80% of cases. The average confidence of the model is 44%.

- Category “thinking (T) or feeling (F) information processing”

The model confirmed the correct result in 79% of cases. The average confidence of the model is 53%.

- Category “inference (J) or perception (p) lifestyle”

The model confirmed the correct result in 68% of cases. The average confidence of the model is 44%.

Using cluster analysis and machine learning (or decision trees), a stochastic predictive model (see Figure 12) was developed to determine the personality type of the candidates—the accuracy of the MBTI personality category prediction ranges from 68% to 84% for individual cases, with a confirmation rate ranging from 43% to 81%. The case study confirmed the model’s usefulness for supporting recruitment, and it has been put into practice.

Predictor	Node Output	Node Instan	Node Support	Node Confide	Prediction	Prediction Confidence
# 'E' is missing	N	282.00	1.00	0.75	N	75.48%
# 'E' > 3	S	1.00	0.01	0.34	N	76.07%
# 'E' <= 3	N	280.00	0.99	0.76	N	76.07%
# 'A' is missing	N	280.00	0.99	0.76	N	76.07%
# 'A' > 3	S	1.00	0.00	0.21	N	76.38%
# 'A' <= 3	N	279.00	0.99	0.76	N	76.38%
# 'P' is missing	N	279.00	0.99	0.76	N	76.38%
# 'P' > 0	N	183.00	0.65	0.67	N	78.17%
# 'D' is missing	N	96.00	0.34	0.67	N	68.61%
# 'D' > 3	N	3.00	0.01	0.68	N	63.82%
# 'D' <= 3	N	88.00	0.31	0.64	N	63.82%
# 'E' is missing	N	88.00	0.31	0.64	N	63.82%
# 'E' > 0	N	36.00	0.13	0.68	N	47.58%
# 'E' is missing	N	52.00	0.18	0.68	N	68.10%
# 'E' > 1	N	23.00	0.09	0.52	N	75.03%
# 'C' is missing	N	27.00	0.10	0.52	N	51.52%
# 'C' > 0	N	6.00	0.02	0.41	N	60.97%
# 'B' is missing	N	21.00	0.07	0.41	N	40.88%
# 'B' > 0	N	3.00	0.01	0.44	N	33.72%

Figure 12: Decision tree bar chart for category “inference (J) or perception (P) lifestyle”

2.5. Validation and evaluation

Validation and evaluation are an important part of the PM model deployment for actual use at the University Counselling Centre (CC UEP) and the University xPORT Business Accelerator (xPORT) as described in the case study below. In so doing, we used the processing method described in [32] to create the case study. The proposal and case study preparation process were separated into four steps:

1. At the March 2017 employment fair *Šance*, we reached out to employers on behalf of CC UEP with an offer to use the PM model to assist in the recruitment of interns.
2. Companies sent out job descriptions and completed surveys for each vacancy. There was a link in the survey to a dashboard that searched the app (Prace Na Miru) data for certain qualities that are available on Facebook.
3. The PM model was tested and verified on an agreed-upon employment position from *Šance* and the internship program. According to the survey’s specifications, the authors contacted chosen users from the application Prace Na Miru and offered them a job. Users interested in the opportunity submitted their CV to the perspective employers.
4. Feedback from employers about the quality of the proposed candidates was collected via an online survey containing 11 questions. The feedback response rate was 54%.

Implementation of steps one through four occurred between March and May 2017. In total, 95 companies were involved.

The survey provided feedback from professional HR recruiters on the PM model. 88% of respondents were positive/liked the PM model.

For recruiters the primarily relevant demographic information from social media was, for example, birth date and place, education, and social media network behavior, such as wall posts, photos, and videos.

Other information (favorite series, movie, music, athlete) which we use in the PM model as predictors were perceived by respondents as less important. Therefore, the PM model has an additional value in the user's behavior evaluation.

3. Discussion

The article has shown how the PM model can be successfully deployed in practice. The case study has proven that the PM model is generally valid and feasible for various companies and job positions.

It has been shown that the use of PM model requires extensive technical and analytical knowledge, especially in the data analysis phase. Therefore recruiters (specifically HR departments) should receive a functional version of the PM model, matching the needs of their particular segment, job position, and candidate target group.

Theoretically, the recruiting app Prace na Miru (Tailored Work) can gather data from any social media network with an open API. Different personality tests or other typological evaluation tests can be chosen as well.

Practically any reasonable test can be used to set parameters to evaluate predictors, providing the authors select the correct analytical method as well.

This article has shown the same results as [33].

Benefits of the Model

The PM model predicts the inclusion of a candidate within a personality type according to the MBTI categorization based on the candidate's behavior on Facebook. The PM model is assembled upon the sophisticated creation of clusters that are based on data categorization automatically obtained via a custom-created Facebook app, Prace Na Miru; its data analysis; and MBTI test evaluation.

The establishment of predictors is a major benefit of this research. Employers will be able to choose and target more accurate applicants on social media. Another advantage of the research is the model's usage in practice. The authors have verified this within the case study and also deployed it in the recruiting solution for CC UEP and the xPORT internship program.

Finally, this proposed solution has the benefit of being adaptable to multiple target groups depending on the needs of the organization.

Research limitations:

- Because social media is a rapidly changing and continually evolving medium, the major problem is the application's long-term viability.
- Only applicants with social media profiles can be evaluated by the PM model.
- The focus is limited to the Czech/Slovak labor market and its legal framework [34].
- Despite testing the model on real data, the authors are aware of possible model distortions [33].

Possible further research and activity:

- Involve students and alumni from additional Czech universities, as well as from other countries.
- Include other social media such as Instagram or Twitter.

4. Conclusion

The main output of the thesis is an artefact in the form of a model for recruitment support, its validation, and a practical application (the PM model).

The article primarily describes the fourth phase of verification, which consists of two stages: formal verification and case study. In the formal verification, a concrete instance of the artifact (in the form of a recruitment support model) is created based on the artifact—the PM model.

The stochastic predictive model confirmed the accuracy of the MBTI personality category prediction as ranging between 68% and 84% for individual cases, with confounding ranging between 43% and 81%. The case study confirmed the effectiveness of the model at supporting recruitment and the real-life deployment of the model.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgment

This paper was processed with a contribution from the Prague University of Economics and Business, IG Agency, OP VVV www.astesj.com

IGA/A, CZ.02.2.69/0.0/0.0/19_073/0016936, grant number 05/2021 and grant IGS 13/2021.

References

- [1] J. B. Becton, H. J. Walker, J. B. Gilstrap, P. H. Schwager, Social media snooping on job applicants: The effects of unprofessional social media information on recruiter perceptions. *Personnel Review*, **48**(5), 1261-1280, 2019, doi:<http://dx.doi.org/10.1108/PR-09-2017-0278>.
- [2] N. Ahmed, S. Khan, K. Latif. Job description ontology, Paper presented at the Proceedings - 14th International Conference on Frontiers of Information Technology, FIT 2016, 217-222.
- [3] M. Armstrong. Řízení lidských zdrojů: nejnovější trendy a postupy. Praha: Grada Publishing, 2015, ISBN 9788024752587.
- [4] L. Gelinas, R. Pierce, S. Winkler, I. G. Cohen, H. F. Lynch, B. E. Bierer, Using social media as a research recruitment tool: Ethical issues and recommendations. *American Journal of Bioethics*, **17**(3), 3-14, 2017.
- [5] LinkedIn, TALENT TRENDS 2014. LinkedIn Talent Solutions, 2014, available: https://business.linkedin.com/content/dam/business/talent-solutions/global/en_US/c/pdfs/linkedin-talent-trends-2014-en-us.pdf.
- [6] D. Smith, Research shows role of pre-hire and post-hire outcomes in effectiveness of social media recruitment, *Human Resource Management International Digest*, **29**(5), 17-19, 2021, doi:<http://dx.doi.org/10.1108/HRMID-02-2021-0033>.
- [7] A. Muduli, J. J. Trivedi, Social media recruitment: The role of credibility and satisfaction. *Evidence - Based HRM*, **8**(2), 237-251, 2020, doi:<http://dx.doi.org/10.1108/EBHRM-08-2019-0069>.
- [8] J. L. Bender, A. B. Cyr, L. Arbuckle, L. E. Ferris, Ethics and privacy implications of using the internet and social media to recruit participants for health research: A privacy-by-design framework for online recruitment. *Journal of Medical Internet Research*, **19**(4), 2017.
- [9] A. Dresch, D. P. Lacerda, Jr. Antunes, *Design Science Research: A Method for Science and Technology Advancement*, New York: Springer, 2015, ISBN 978-3-319-07373-6.
- [10] V. K. Vaishnavi, W. Kuechler, *Design Science Research Methods and Patterns*, CRC Press, Boca Raton, FL, 2015.
- [11] Z. Molnár, S. Mildeová, H. Řezanková, R. Brixí, J. Kalina, Pokročilé metody vědecké práce, Zeleneč: Profess Consulting, Věda pro praxi (Profess Consulting), 2012, ISBN 978-80-7259-064-3.
- [12] Y. Melanthiou, F. Pavlou, E. Constantinou, The use of social network sites as an E-recruitment tool, *Journal of Transnational Management*, **20**(1), 31-49, 2015.
- [13] C. Khatri, J. S. Chapman, J. Glasbey, M. Kelly, D. Nepogodiev, A. Bhangu, J. E. Fitzgerald, Social media and internet driven study recruitment: Evaluating a new model for promoting collaborator engagement and participation. *PLoS ONE*, **10**(3), 2015.
- [14] D. Muntinga, M. Moorman, E. Smit, 'Introducing COBRAs', *International Journal of Advertising*, **30**(1), 13-46, Business Source Ultimate, EBSCOhost, 2011, ISSN: 0265-0487.
- [15] Oracle, *Social Recruiting Guide: How to Effectively Use Social Networks*, 2012, available: <http://www.oracle.com/us/media/effectively-use-social-networks-1720586.pdf>.
- [16] H. Yin, B. Cui, L. Chen, Z. Hu, X. Zhou, Dynamic user modeling in social media systems, *ACM Transactions on Information Systems*, **33**(3), 2015.
- [17] M. Miller, Face-to-face with the digital folk: The ethics of fieldwork on facebook, *Folk culture in the digital age: The emergent dynamics of human interaction* (pp. 212-232), 2012.
- [18] A. Ladkin, D. Buhalis, Online and social media recruitment: Hospitality employer and prospective employee considerations. *International Journal of Contemporary Hospitality Management*, **28**(2), 327-345, 2016.
- [19] M. Mattare, Revisiting Understanding Entrepreneurs Using the Myers-Briggs Type Indicator®, *Journal of Marketing Development and Competitiveness*, **9**(2), 114-119, 2015, ISSN: 21552843.
- [20] P. Chapman, J. C. Clinton, R. Kerber, T. Khabaza, T. Reinartz, C. Shearer, R. Wirth, CRISP-DM 1.0, Step-by-step data mining guide, 2020, Available: <ftp://ftp.software.ibm.com/software/analytics/spss/support/Modeler/Documentation/14/UserManual/CRISP-DM.pdf>.
- [21] A. Chen, How Accurate Are Personality Tests?, *Scientific American*, Oct. 10, 2018, accessed: Sep. 14, 2021, available: <https://www.scientificamerican.com/article/how-accurate-are-personality-tests/>.
- [22] R. Stein, Personality Tests with Deep-Sounding Questions Provide Shallow Answers about the True You, *Scientific American*, 2018, accessed: Sep. 14, 2021, available: <https://www.scientificamerican.com/article/personality-tests-with-deep-sounding-questions-provide-shallow-answers-about-the-true-you/>.

- [23] K. Randall , M. Isaacson, C. Ciro, Validity and Reliability of the Myers-Briggs Personality Type Indicator: A Systematic Review and Meta-analysis, *Journal of Best Practices in Health Professions Diversity: Education, Research & Policy*, vol. 10, no. 1, pp. 1–27, Spring 2017.
- [24] Jobvite, Social Recruiting Survey, 2014, available: https://www.jobvite.com/wp-content/uploads/2014/10/Jobvite_SocialRecruiting_Survey2014.pdf.
- [25] Knime, 2017, available: <https://www.knime.com/>.
- [26] I. B. Myers, M. H. Mccauley, R. Most, *Manual: A Guide to the Development and Use of the Myers-Briggs Type Indicator*, Palo Alto, Ca.: Consulting Psychologists Press, 1985, ISBN 08-910-6027-8.
- [27] C.E. Fretwell, C. C. Lewis, M. Hannay, Myers-Briggs Type Indicator, A/B Personality Types, and Locus of Control: Where Do They Intersect?, *American Journal of Management*, **13**(3), 57-66, 2013, ISSN: 21657998.
- [28] BigML, 2017, available: <https://bigml.com/>.
- [29] M. N. Manu, K. R. Anandakumar, A current trends in big data landscape. Paper presented at the 2015 IEEE International Conference on Computational Intelligence and Computing Research, ICCIC 2015.
- [30] Q. Tran, Using ANOVA to analyze modified gini index decision tree classification, Paper presented at the Proceedings of the 2008 International Conference on Data Mining, DMIN 2008, 164-170, 2008.
- [31] Pajek, Analysis and visualization of large networks, 2017, available: <http://mrvar.fdv.uni-lj.si/pajek/>.
- [32] R.K. Yin, *Case study research: design and methods*, 5. Edice, Los Angeles, Calif. [u.a.]: Sage Publ, 2014, ISBN 14-522-4256-9.
- [33] L. Böhmová, A. Pavlíček, 'How the Modern Human Resources Management Can Take Advantage of Information From Social Media while Recruiting', in 2020 International Conference on Engineering Management of Communication and Technology (EMCTECH), Oct. 2020, 1–6. doi: 10.1109/EMCTECH49634.2020.9261522.
- [34] Official Journal of the European Union, Ec.europa, 2016, available: http://ec.europa.eu/justice/data-protection/reform/files/regulation_oj_en.pdf

Improving of Heat Spreading in a SiC Propulsion Inverter using Graphene Assembled Films

Sepideh Amirpour^{*1}, Torbjörn Thiringer², Yasin Sharifi³, Marco Majid Kabiri Samani⁴

¹China Euro Vehicle Technology AB, Department of Motion and Energy, Gothenburg, SE-41755, Sweden

²Chalmers University of Technology, Department of Electrical Engineering, Gothenburg, SE-41296, Sweden

³Fengco Real Time Control AB, Gothenburg, SE-41109, Sweden

⁴INFOTIV AB, Gothenburg, SE-41117, Sweden

ARTICLE INFO

Article history:

Received: 01 August, 2021

Accepted: 14 November, 2021

Online: 17 December, 2021

Keywords:

Silicon Carbide (SiC)

Voltage Source Inverters (VSI)

Thermal Feedback

Graphene Assembled Films

Thermal Simulations

Electrified Vehicles

ABSTRACT

The focus of this work is first to establish the effect of the chip temperature and thermal feedback on the determination of the power loss in a three-phase propulsion inverter, then to demonstrate the possibility of achieving an improved heat spreading through the different layers inside a SiC power module by using graphene assembled films in the packaging of the power module. The power loss analysis has been carried out for two Silicon Carbide (SiC) modules in a vehicle inverter, incorporating the MOSFET's reverse conduction as well as including the impact of blanking time on the inverter on-state losses. This data for calculating the losses is determined at an operating situation below the field weakening speed with a high torque for a permanent magnet synchronous machine (PMSM). The operating point is found to be the worst operating condition point when looking at the power loss point. First, it can be noted that not accounting for the thermal feedback, the power loss is considerably underrated, i.e., 11-15% on the on-state converter. Following, the analysis of utilizing the graphene layer in the SiC module reveals a reduction of 10°C per SiC chips in the junction temperature of the SiC MOSFET is achievable. The reduction is calculated based on an applied power loss per SiC chips in steady-state simulation. Furthermore, up to 15°C decrease in the transient computation over the Worldwide Harmonized Light Vehicles Test Cycle (WLTC) per SiC chip is noticed. Moreover, a reduction up to 50% for the junction to case thermal resistance ($R_{th,JC}$) is observed by adding the graphene layer in the power module.

1. Introduction

Increasing the efficiency of vehicle inverters for electrified powertrains is becoming a significant demand [1]. Improved Wide Band Gap silicon carbide (SiC) MOSFETs are promising replacements of the standard silicon insulated gate bipolar transistors (Si IGBTs) in EV applications [2–4]. This is a result of lower switching losses due to quicker switching transitions, as well as good thermal properties in the SiC MOSFETs, superior to those of Si IGBTs. Moreover, lower the on-state losses can be reduced through the utilization of the MOSFET's ability to also conduct current in the opposite direction [5, 6]. Therefore, improving the efficiency of the propulsion inverter utilizing SiC MOSFETs, can lead to reduced losses in the powertrain, which results in a higher power density, to some extent also through the reduction of cooling

circuitry. However, the development of switching devices towards miniaturization, causes severe temperature stresses that significantly threaten the performance and lifetime of the semiconductor chips and dies in high power applications. A possibility to tackle this issue is enhanced thermal conduction in various directions inside the power module with effective cooling of the semiconductor chips. Integrating heat spreading materials with excellent mechanical and thermal properties for the power electronics packaging can efficiently move the heat away from the power devices, resulting in a lower operating temperature of the system [7].

Compared to metal materials, nano-scaled carbon materials, in especial, graphene, have a thermal conductivity up to 5300 W/m.K [8], ten times higher than that of the metals [7, 9] as well as an amazing lightweight property, 2.2 g/cm³ and good stability [7], [10, 11]. Particularly, compared with the single layer graphene,

*Corresponding Author: Sepideh Amirpour, Gothenburg, SE-41755, Sweden, +46721843949, sepideh.amirpour@cevt.se

there is the incentive in fabricating graphene assembled films (GFs) as new heat spreading materials [12–15] due to their promising electrical and thermal properties. Consequently, research community as well as manufacturers have put intensive efforts into the research on new high-thermally conducting materials like graphene as an effective heat spreader and these efforts have been reported in literature [16–21]. In [22–24] research on the usage of graphene as a spreader of heat in power modules packaging have been conducted. Further interesting research are in [25–27] and a procedure for studying graphene transistors using standard TCAD tools is brought forward in [28]. However, these applications have their own limited scope regarding the design and do not satisfy the comprehensive aspects of the combined electrical and thermal analysis of the SiC semiconductor power module from the chip level towards the different material layers.

To fill this knowledge gap, a research effort conducted is presented in this paper. The specific novel contribution of this article (which is an addition of the research originally provided at the IEEE Industrial Electronics Society (IECON) 2020 conference [29]) is: How can the temperature distribution through the SiC MOSFET’s chips as well as the Direct Bonded Copper (DBC) layers in the power module be efficiently spread, through the utilization of graphene with its excellent thermal properties. The finite element method (FEM) in the COMSOL Multiphysics software is very useful when it comes to determining the thermal aspects and has accordingly been utilized. First, a comprehensive numerical analysis quantifies to what extent incorporating and omitting the thermal feedback will affect the losses in a SiC-based propulsion inverter when including the blanking time and the MOSFET’s conduction in its opposite direction [29]. Then the graphene solution is proposed to boost the heat dissipation in the SiC power module.

The analysis is carried out for two SiC half-bridge modules, CAB450M12XM3, a recently introduced 3rd generation [30], and the previous module, CAS300M12BM2 also with 1200V rating [31] from Wolfspeed /Cree. Figure 1 shows the procedure used for the power loss determination.

2. Reverse Conduction and Blanking Time

In Si IGBTs the flow of the current in opposite direction is through an anti-parallel diode, while MOSFETs can conduct this current through their ‘conduction channel’. In the inverter, when the current and voltage in one leg does not have the same signs, one of the upper or the lower diodes in the phase leg is conducting. When V_{ds} of the corresponding MOSFET exceeds the diode’s threshold voltage, both will conduct in parallel. The diode can either be the intrinsic inbuilt diode or a separate one. This ability affects the distribution of on-state losses in a SiC MOSFET module, with the consequence of lower losses.

In order to prevent a short-circuit of the dc-link, a time duration when the lower as well as upper transistor is off at the same time, is needed in a PWM-controlled inverter, a so-called blanking time. The result is that, the risk of a short circuit of the dc-link can be kept to a minimum. The total on-state loss of the modules is affected by the conduction of the diode as a consequence of the blanking time, increasing the losses slightly.

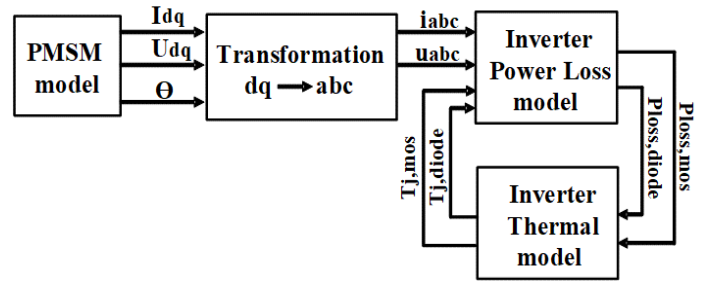


Figure 1: Overview of the system simulation model.

3. Conduction Losses including the MOSFET’s Reverse Conduction and Blanking Time

3.1. Conduction Losses in the SiC Power Module

The on-state losses for the investigated inverter are determined numerically using MATLAB. The methodology used in this work is based on the procedure presented in [6] and [32] for calculating the SiC MOSFETs’ conduction losses. In this procedure, the MOSFET on-state loss over a machine line period is determined according to

$$P_{cond,MOS} = \frac{1}{2\pi} \int_0^{2\pi} R_{on} I_M^2(\alpha) \tau(\alpha) d\alpha \quad (1)$$

here R_{on} is the MOSFET on-state resistance, I_M the MOSFET current, $\alpha = 2\pi ft$ where f is the machine fundamental frequency and τ is the duty cycle found as

$$\tau(\alpha) = \frac{1}{2} (1 + m \sin \alpha) \quad (2)$$

here m is the modulation index [33]. Similarly, the on-state losses of the diode can be formulated as

$$P_{cond,D} = \frac{1}{2\pi} \int_0^{2\pi} (R_d I_D^2(\alpha) + V_d I_D(\alpha)) \tau(\alpha) d\alpha \quad (3)$$

From the data sheet, V_d , the voltage drop and, R_d , the on-state resistance can be found, I_D is the current through the diode.

The resulting expressions for the current through the MOSFET and the diode thus becomes

$$I_M = \frac{R_d I_p \sin(\alpha - \varphi) - V_d}{R_d + R_{on}} \quad (4)$$

$$I_D = \frac{R_{on} I_p \sin(\alpha - \varphi) + V_d}{R_d + R_{on}} \quad (5)$$

where φ is the angle of fundamental power factor and I_p is the peak value of the phase current [32].

The losses when incorporating and not incorporating the MOSFET reverse conduction are displayed in Figure 2 for the upper diode /MOSFET combination in a phase leg of a CAS300, SiC inverter. A substantial decrement of up to 83% in the diodes’ total on-state losses originating from the parallel conduction was found. In Table 1 the operating condition used for calculating the

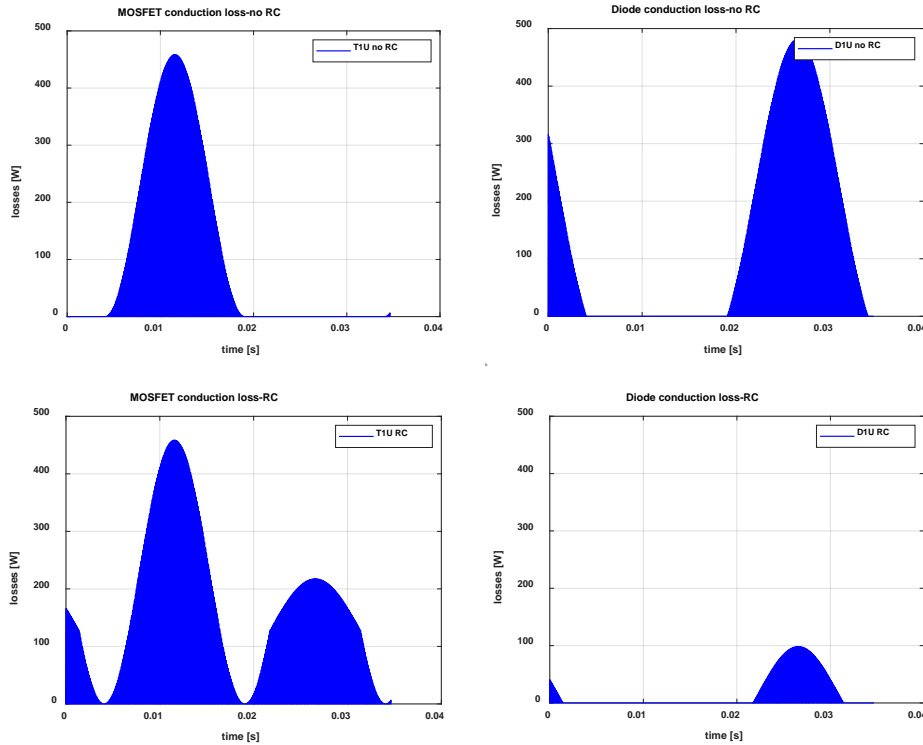


Figure 2: MOSFET and diode conduction losses with (RC), and without MOSFET reverse conduction (no RC), in CAS300 SiC inverter for upper MOSFET and diode in a phase leg.

Table 1: Chosen Operating Point of PMSM for Power Loss Calculations

Variable	Value	Unit
Current magnitude	565	[A]
DC voltage	300	[V]
Blanking time	0.5	[μ s]
Switching Frequency	10	[kHz]
Modulation index	0.084	[-]
Torque	160	[Nm]
Mechanical speed	500	[rpm]

losses is given. Likewise, a reduction up to 97% is noticed in the total diodes' conduction losses of the CAB450 SiC inverter.

3.2. Blanking Time

The influence of blanking time is incorporated in the calculation of the on-state losses by formulating a representative duty cycle. The resulting formulation becomes,

$$\tau_{eq}(\alpha) = \tau(\alpha) - t_{blanking}f_{sw} = \frac{1}{2}(1 - 2t_{blanking}f_{sw} + m\sin\alpha) \quad (6)$$

where f_{sw} is the switching frequency [32].

The calculated currents of the diodes and MOSFETs in a phase leg of the CAS300 inverter are presented in Figure 3, without (top one) and with (bottom one) utilizing blanking time, also incorporating the MOSFET reverse conduction. As can be seen in the bottom figure, there is current only in the diode for the blanking time intervals. This leads to an increase of 20 W (12%) in the diodes' conduction losses for the CAS300 inverter which is

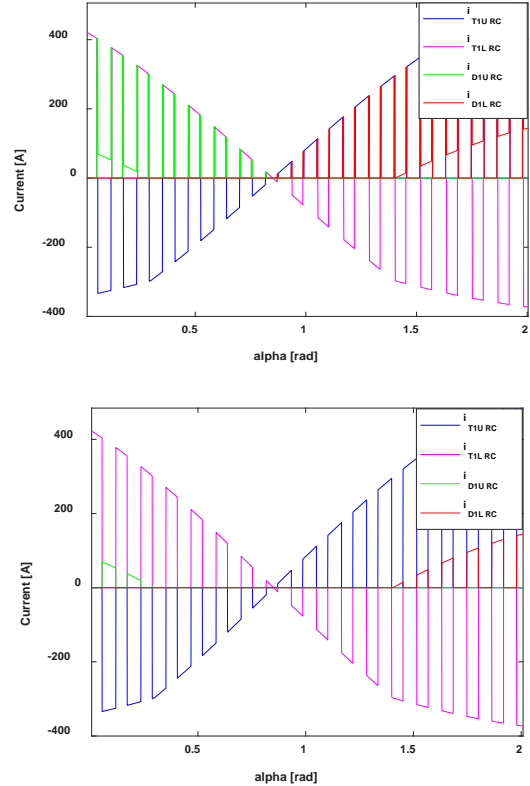


Figure 3: Influence of blanking time on the diode and MOSFET currents as a function of $\alpha = 2\pi ft$ in one phase leg of a CAS300 equipped inverter. Both when incorporating the MOSFET's reverse conduction and not. Upper figure, no blanking time, lower figure with blanking time.

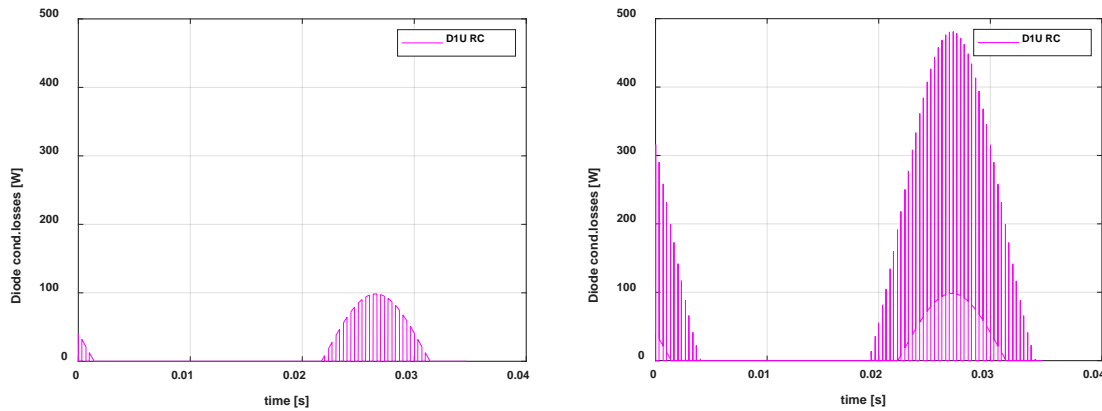


Figure 4: Diode conduction loss, without blanking time (left figure), with blanking time (right figure) in a phase leg of CAS300 SiC inverter.

illustrated in Figure 4 (right one). The impact on the MOSFETs' conduction losses is very low. Likewise, for the CAB450, a rise of up to 70 W, of the diodes' conduction losses is noted. Figure 4 presents the upper diode conduction loss in a phase leg of the CAS300, with and without blanking time, when incorporating the MOSFET reverse conduction. In Table 1 the operating condition used for the determination of the conduction losses are provided.

4. Switching Losses in SiC MOSFET

In the MOSFETs and its anti-parallel/body-diode, a resulting loss is created from each turn off and turn on. The switching losses in a MOSFET and diode can be obtained analytically by the expression as

$$P_{sw.MOSFET,Diode} = f_{sw} \cdot E_{sw(@I_{nom},V_{nom})} \cdot \left(\frac{1}{\pi} \frac{I_p}{I_{nom}}\right)^{k_i} \cdot \left(\frac{V_{dc}}{V_{nom}}\right)^{k_v} \quad (7)$$

In the expression E_{sw} represents the switching energy loss, I_p the peak phase current, I_{nom} and V_{nom} the nominal current and voltage values and k_i , and k_v the current and voltage exponents [34]. In this study, in the numerical implementation, the switching

loss is established at each switch-off and switch-on occasion of the module as

$$P_{sw} = \frac{\sum E_{sw}}{t} \quad (8)$$

with t being the simulation time.

5. Electro-Thermal Calculation Network

Due to that several parameters in the modules are temperature-dependent, for instance the on-state resistances, forward voltage drops, switching and reverse recovery energies, the inverters' power losses are determined through the usage of a thermal model, that is given in Figure 5. T_j , are the junction temperatures of the MOSFET and the diodes, while T_c is the case temperature, T_s and T_f are the heatsink and fluid temperatures. In order to have a representative comparison, the same heatsink is utilized for both inverters. Noteworthy is that, for the inverters' design, three SiC half-bridge modules were needed. All set-ups are normalized to a

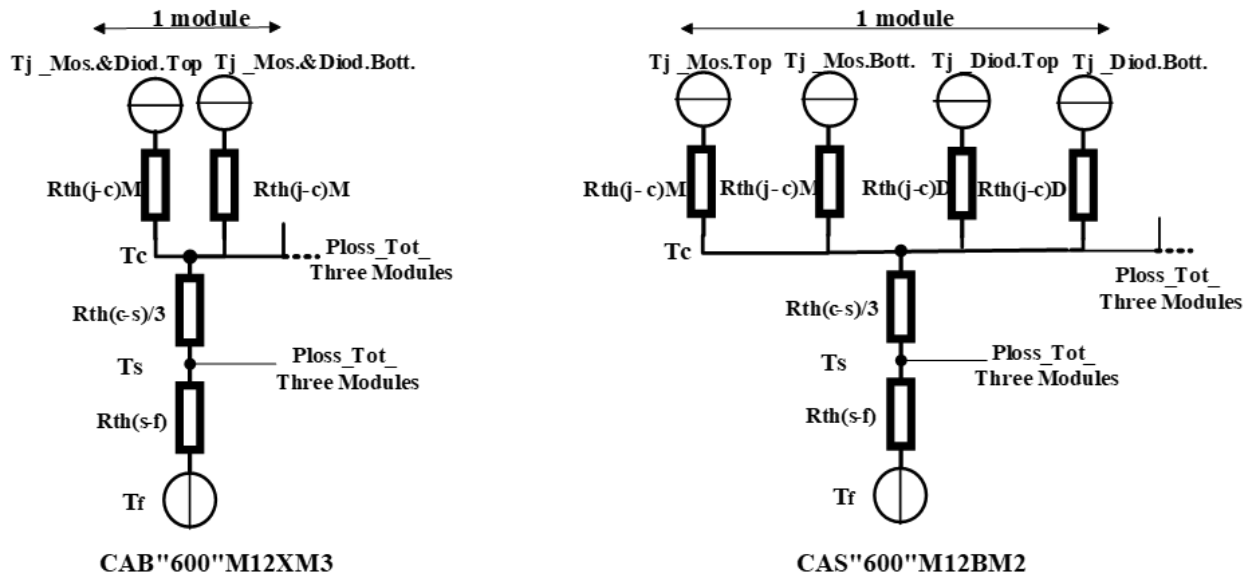


Figure 5: Thermal calculation model to find the power losses and temperatures.

Table 2: PMSM Rating Parameters

Variable	Value	Unit
DC-Link voltage	300	[V]
Rated current	400 RMS	[A]
Pole Pairs	4	[-]
Max. speed	12000	[rpm]
Max. torque	160	[Nm]

rated current of 600 A following the specification of the inverters. The temperature and flow rate of the coolant is 65 °C and 10 L/min. Worth mentioning is that the latest 3rd GEN MOSFET dies have been used in the CAB450 module which have an intrinsic body diode, and accordingly the module does not need an antiparallel diode, as can be observed in Figure 5. In comparison, the CAS300 module is equipped with an anti-parallel Schottky diode.

6. Power Losses Analysis in SiC Inverters with and without Accounting for Thermal Feedback

As mentioned before, an operating point of a PMSM at high torque, low speed and high current magnitude is used to analyze the power losses of the two investigated SiC power modules. The chosen operating point can be noticed as the toughest operating condition in the urban driving cycles. MPTA/MTPV field-oriented control is utilized to generate the inverter currents. Based on this control strategy, the machine operates in the constant torque speed range and in the partial and full field weakening speed ranges [35,36]. Table 1 gives the chosen operating point and key machine parameters are given in Table 2. As depicted in Figure 1, an iterative approach is used to find the losses based on the temperature in which, the devices' steady-state and transient parameters (V_T , V_d , R_f , R_d , R_{on} , E_{sw} and E_{rr}) are interpolated using the feedback of the junction temperature. The losses and the temperatures are looped until they have converged. Worth mentioning is that the switching energies of the devices are

interpolated as a function of junction temperature and current to fulfill both the current and temperature dependency of the switching energies. The results of the investigations are given in Table 3. It is worth to note that, for a fair comparison, a power loss scaling factor is utilized on the SiC modules in order to form them into 600 A modules.

As shown in Table 3, a significant reduction in diode conduction losses is observed for the two SiC modules. As was explained in sections 2 and 3, this is due to the MOSFET's reverse conduction ability to utilize the diode and the MOSFET to share the current. In the SiC modules, the increase of on-state losses for the built-in diodes as a function of an increase in junction temperature is substantially lower compared to that of the MOSFETs' channel. This is due to a more moderate rise in the diode dynamic resistance with respect to the temperature increase. Moreover, for the two investigated modules, in SiC-MOSFETs, a temperature increase has almost zero effect on the switching losses.

For the comparison of the two SiC modules, a substantial decrease is noticed in the on-state losses in the CAB450's diodes. This is due to the fact that, the CAB450 module is using the latest 3rd GEN MOSFET dies with a very robust intrinsic body diode and reliable operation in the 3rd quadrant. This feature is weaker in the CAS 300 Module.

All in all, the analysis shows that the thermal feedback has a considerable impact on the losses and leads to an increase of up to 11% for CAS300 and 15.6% for CAB450 on the inverters' total conduction losses at the chosen operating condition. This leads to that the demand for effective cooling strategies of the semiconductor power modules is increasing continuously, especially for traction applications which can typically have a high ambient temperature, particularly in SiC-MOSFETs where the temperature variations have larger magnitudes compared to those of Si-IGBTs [37]. Hence, in the following sections, utilizing graphene, a novel thermal conductive, cost-effective, and eco-friendly material is proposed in the power module packaging and

Table 3: Calculated Average Values of Conduction and Switching Losses of the SiC Inverters utilizing and not utilizing Thermal Feedback, Considering Blanking Time and using the MOSFET's Reverse Conduction at 160 Nm, 500 rpm Mechanical Speed and $T_f = 65^\circ\text{C}$.

CAB "600" M12XM3 (CAB450)									
	Cond. [W]	Cond. Thermal Feedback [W]	Diff. [W]	Diff.%	Sw. [W]	Sw. Thermal Feedback [W]	Diff. [W]	Diff.%	Tjunc. Thermal Feedback [°C]
MOSFETs	977.8	1140.5	162.7	16.6	206.6	206.8	0.2	0.09	99.8
Diodes	74.4	75.4	1	1.35	3.51	3.52	0.01	0.28	99.8
Tot. Inverter	1052.2	1215.9	163.7	15.6	210.1	210.3	0.22	0.1	
CAS "600" M12BM2 (CAS300)									
	Cond. [W]	Cond. Thermal Feedback [W]	Diff. [W]	Diff.%	Sw. [W]	Sw. Thermal Feedback [W]	Diff. [W]	Diff.%	Tjunc. Thermal Feedback [°C]
MOSFETs	1072	1164	92	8.6	196.3	196.5	0.2	0.1	95.6
Diodes	140.1	182.1	42	29.9	0	0	0	0	88.2
Tot. Inverter	1212.1	1346.1	134	11	196.3	196.5	0.2	0.1	

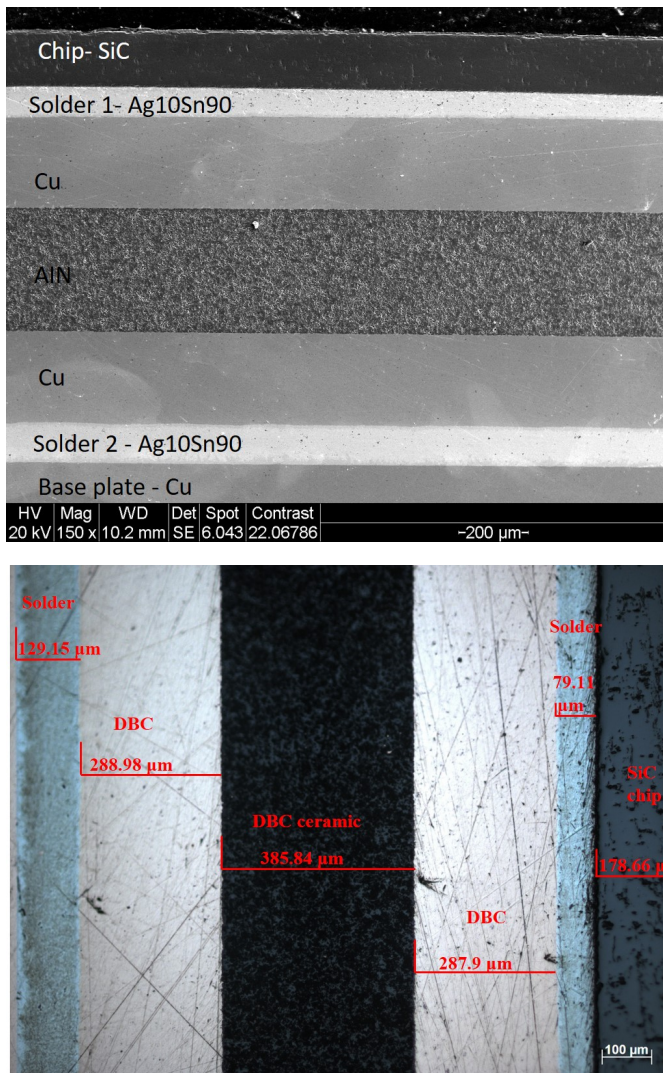


Figure 6: The Layer materials of the reference SiC module CAS300 from SEM and ESD analyses (top) and the materials respective thickness of CAS300 from SEM and ESD analyses (bottom) (given in Table 4).

analyzed by FEM in the COMSOL Multiphysics software intended to improve the heat spreading inside the module.

7. 3D-CAD Model and Thermal Modeling Set-up of the Power Module using SEM and EDS Analyses

As was mentioned before, to enhance the heat spreading and cooling of the semiconductor chips in the chosen half-bridge SiC power module, CAB450M12XM3, a thermal model is built up as close as possible to the real module starting from CAD and then

FEM simulation in COMSOL Multiphysics. Even though the thermal packaging aspects of the module was mostly unknown due to confidentiality. However, the key features were known from the reverse engineering effort, which showed that the base plate and insulator have been made of copper and silicon nitride, respectively. Therefore, a Scanning Electron Microscopy (SEM) together with an Energy Dispersive X-Ray Spectroscopy (EDS) investigation were conducted on the available similar voltage class SiC power module, CAS300M12BM2, as a reference to approximate further material information. EDS is a chemical micro-analysis technique, which identifies the x-rays released from the sample through a bombardment process with electrons, so that the element arrangement of the analyzed volume can be found. The above-mentioned analyses have been conducted on a cross-sectional cut off piece of the module. As illustrated in Figure 6 (top), the chips are all silicon-carbide, the base plate is made of copper and an aluminum nitride (AlN) insulator has been used as one of the DBC layers. The other two DBC layers are made of copper. Moreover, the solder $Sn_{90}Ag_{10}$ is used for both solder layers. The thickness of the layers which is shown in Figure 6 (bottom), is given in Table 4.

In the next step, the CAD geometry of the CAB450 SiC power module including the thermal aspects is meshed in COMSOL Multiphysics based on the SEM and EDS analyses data from the reference CAS300 module as well as the CAB450 datasheet information. Instead of AlN, silicon nitride (Si_3N_4) is used as the ceramic material in the DBC substrate based on the CAB450 datasheet. Figure 7 (left) presents a 3D-geometry of the CAB450 half-bridge power module including the chips in which each group of five chips is representing one SiC switch place. The dimensions are implemented according to the datasheet information.

8. Automated FEM Computation in COMSOL for Steady- state/Transient Heat Dissipation

As mentioned before, to be able to capture all the heat dissipation surface interactions, a 3D FEM simulation in COMSOL Multiphysics is performed. Each of the power module materials and their thicknesses were accounted for in the meshing procedure as separate computational domains including the solders, copper, DBC, etc. In order to take the thermal coupling and heat spreading effects into consideration, the contact resistances among the material layers were also modelled using the provided realistic values by material datasets. As the focus of

Table 4: Power Module Layers' Thicknesses and Thermal Properties

Layer	Material	Thickness (μm)	Thermal conductivity, K (W/m.K)	Density, ρ (kg/m ³)	Heat capacity, C (J/kg.K)
Chip	SiC	178	490	3216	690
DBC	Cu	288	400	8940	385
DBC ceramic	AlN/Si ₃ N ₄	386	20	3100	700
Solder (below the chip)/ (below the DBC layer)	Sn ₉₀ Ag ₁₀	79/129	50	9000	150
Baseplate	Cu	3	400	8940	385

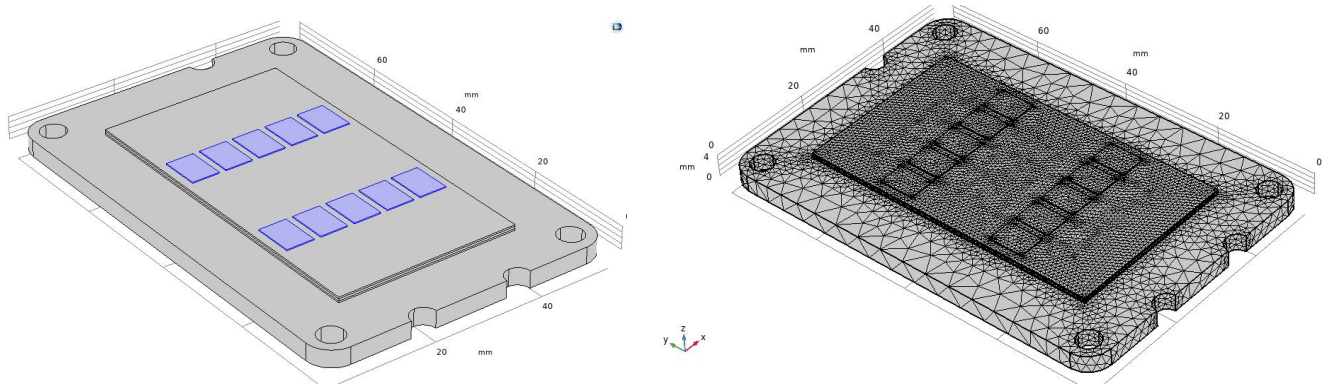


Figure 7: 3D-geometry of CAB450 SiC module built for simulation (left) and the mesh visualization of the module (right).

the analysis is on the thermal distribution through the chips and DBC layers, a manually refined mesh is applied in these domains. Overall, a uniform mesh distribution is built up as depicted in Figure 7 (right). The boundary conditions are carried out by considering the chips as heat sources and a selected power losses of 1262 W, i.e., 126.2 W per chip from the previous power loss calculations of the investigated module is applied to these heat sources.

To mimic a liquid cooling solution, the bottom surface of the baseplate is given a heat transfer coefficient of 3000 W/(m².K) to represent the convective heat flux as a boundary condition with a liquid flow of water with a rate of 1 m/s (one meter per second) and a temperature of 65°C. All other boundaries are set to have a heat transfer coefficient of 10 W/(m².K), which is the convection of low-speed air flow above a surface [38].

In general, the transportation of heat can be described as the transportation of the thermal energy as a result of a gradient in temperature. Then the heat flux can be expressed by Fourier's law as

$$q = -k \nabla T \quad (9)$$

which defines the theory behind the heat conduction. The equation shows that the thermal conductivity, k in (W/m.K), is proportional to the magnitude of the temperature gradient. q is the heat flux measured in (W/m²) and ∇T is the temperature gradient.

The heat convection, which can either be natural or forced, is added to the boundaries of the system, and is dependent on the geometry and its length. When natural convection is implemented, the system is cooled naturally through air. The steady-state heat flux density is calculated as the following equation

$$q_s = h (T_s - T_m) \quad (10)$$

when the convective cooling/heating is involved. h is the heat transfer coefficient in W/(m².K), T_s represents the surface temperature and T_m is the media temperature.

Since there are different materials involved in this study, then the heat conduction or heat diffusion must be taken into account by the expression as

$$\rho C \frac{\delta T}{\delta t} = \nabla K(T) \nabla T + q_v \quad (11)$$

where ρ is the density, C is the heat capacity, $\frac{\delta T}{\delta t}$ represents the difference of temperature over time, and ∇T is the temperature gradient.

Finally, the model is ready for analyzing the time variations of the temperature inside the different layers of the power module and the temperature profiles across it, within the boundary layers. The first results will illustrate the thermal performance of the investigated power module which is built up as close as possible to the real model and in the next phase an assembled graphene film layer is placed under the first copper layer as close as possible to the chips as heat sources to demonstrate the improvement of the heat distribution inside the module from the chip level towards the different layers.

9. Results of Steady-state Thermal Computation of the Power Module

The stationary solver in COMSOL is used to evaluate the steady-state temperature distribution across the power module from the chips towards the bottom layers. The heat distribution is measured by applying a 3D-cutline from the middle chip to the baseplate as depicted in Figure 8. The reason for choosing the middle chip is since this one will experience the highest thermal stress. As it can be seen in Figure 9 (top), the result shows an average maximum temperature of around 99°C for the chips for the applied power loss of 126.2 W per chip which is approximately close to the junction temperature calculated by the electro-thermal model presented in section 6, Table 3. The lowest temperature of 70°C is also noticed around the edges of the module. The temperature profile across all the layers is illustrated in Figure 9 (bottom).

In addition, the thermal resistance of the junction to case (R_{thJC}) of 0.2 K/W has been obtained from the simulation.

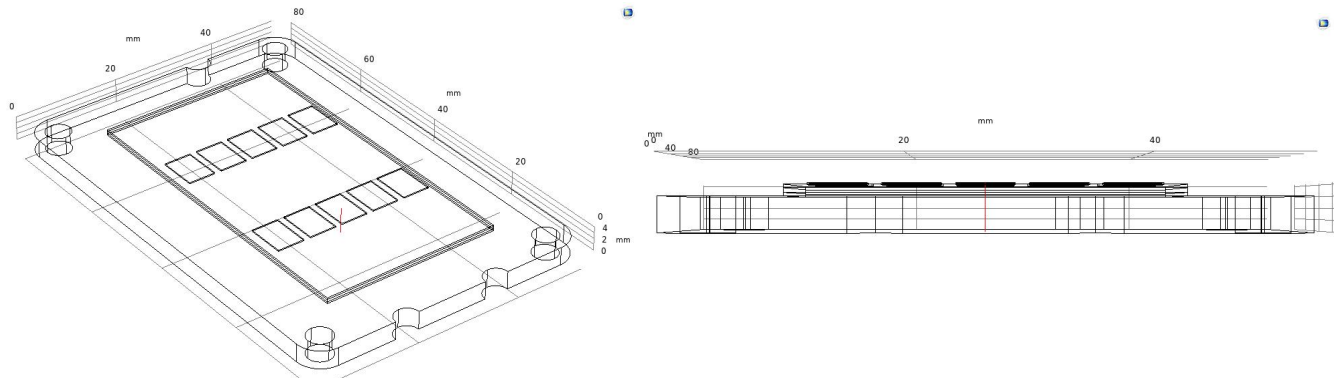


Figure 8: 3D-cutline across the module for temperature measurements, top view (left) and side view (right).

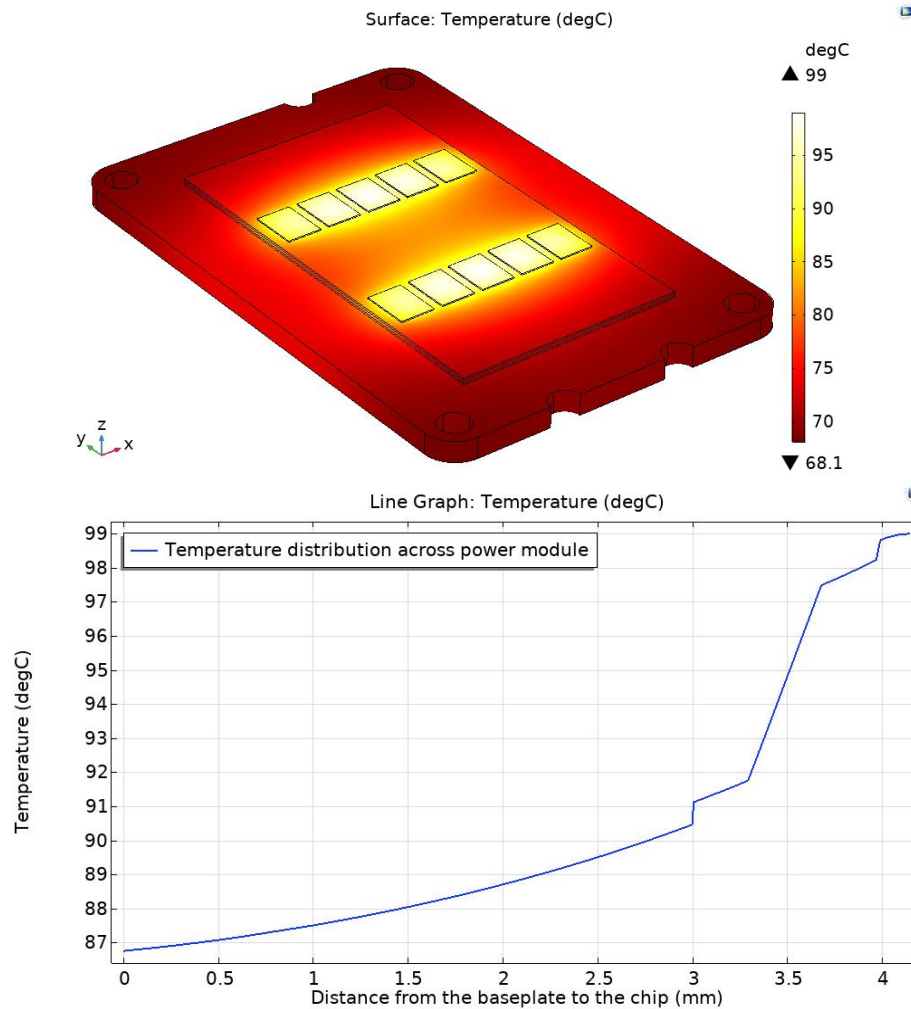


Figure 9: Temperature distribution across the power module (top) and temperature profile across all the layers (bottom).

10. Using the Graphene Assembled Film in the module Packaging

As mentioned before, a novel solution to efficiently transport the heat away from the power devices, especially in high power applications can be achieved by integrating ultra-high thermal conducting as well as flexible and robust materials into the cooling path [7]. Therefore, graphene which is one of the most promising materials that can provide the above-mentioned properties, has been used in this study to optimize the thermal impedance of the

power module as well as to make a uniform and fast heat spreading over the module.

To prove this goal, the graphene layer has been placed as close as possible to the chips which are the heat sources, with an in-plane thermal conductivity of 2900 W/m.K in x and y directions and a cross-plane thermal conductivity of 14 W/m.K in the z direction. The directions as the design parameters should be chosen in such a way that they conduct and spread the heat flow as fast and

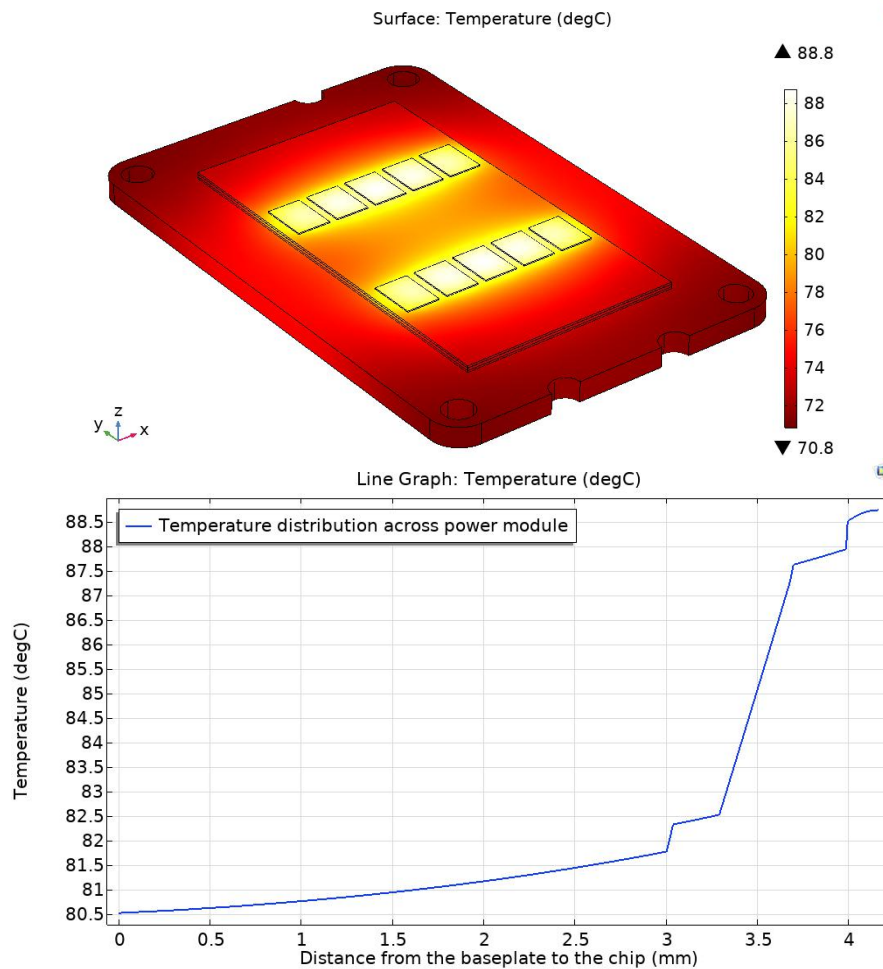


Figure 10: Temperature distribution across the power module (top) and temperature profile across all the layers (bottom) with graphene.

Table 5: Graphene Thermal Properties

Material	Thermal conductivity, K (W/m.K)	Density ρ (kg/m ³)	Heat capacity, C (J/kg.K)
Graphene	(x,y,z) = (2900,2900,14)	2267	720

efficient as possible on the surface of the module and then towards the bottom layers.

Worth mentioning is that, to reach high performing thermal and mechanical properties, the fabricating freestanding graphene films (GFs) have shown a promising potential compared to using single/few layer graphene for the thermal management of high-power electronics [7]. Hence, in this work, 1 μm thick graphene films are utilized to make an appropriate layer thickness in respect to the thermal conductivity which is challenging since the thermal conductivity levels are very much dependent on the thickness of the graphene layer [7]. The layer is placed under the first copper layer as a realistic placement in order to not weaken the electrical conductivity through the module which otherwise might happen as a result of a poor insulating ability of the graphene material.

In addition, in the fabrication processes, in order to minimize the thermal contact resistance between the graphene layer and www.astesj.com

other different substrates, a molecular functionalization method is used in which the small molecules are utilized to functionalize the surface of the graphene film which is in contact with the device surfaces. The small molecule, as the functional agent, gives as result a formation of molecular bridges between the device substrates and the grapheme-based surface and in this way the thermal resistance between these layers is strongly reduced [39].

Therefore, based on this method, the contact resistance is set to $1.1 \times 10^{-8} \text{ K.m}^2/\text{W}$ in this work to achieve the appropriate thermal coupling between the layers [39]. Table 5 gives the material properties of graphene.

11. Results of Steady-state Thermal Computation of the Power Module with Graphene Layer

To investigate the thermal behavior in the presence of the graphene layer in the power module, a 3D-cutline from the middle chip to the baseplate, similar as in section 9 is applied to the module.

The result reveals a reduction of 10°C in the average temperature per chip for the applied power loss of 126.2 W per chip, as depicted in Figure 10, and Figure 10 (bottom) which illustrates the temperature profile across all the layers. Moreover, as can be seen in Figure 11, a comparison between the two cases,

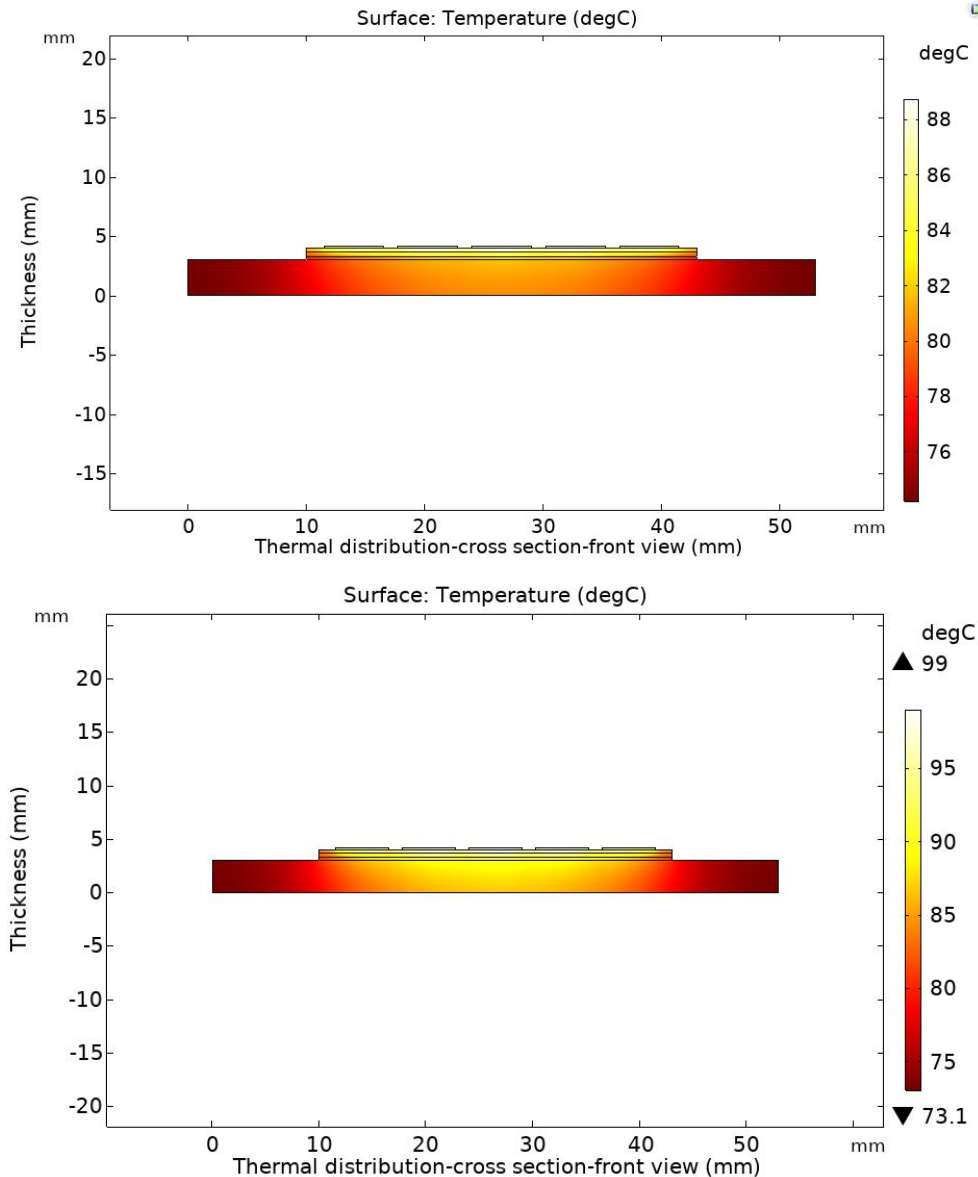


Figure 11: Temperature distribution with graphene (top) and without graphene (bottom).

with and without using graphene, confirms the expected results, i.e., a noticeable difference in the temperature distribution in the module. With graphene (top figure), the temperature is spreading flatter and more homogeneously in each layer, which means that each layer is absorbing the heat uniformly. Consequently, the heat is more distributed across the baseplate compared to that of the case without graphene (bottom figure) in which the heat is mostly concentrated in the middle where the chips are located. Therefore, this temperature uniformity leads to a significant improvement, 54.3%. The junction to case thermal resistance (R_{thJC}) equals to 0.0913 K/W, using the graphene layer compared to that of 0.2 K/W without graphene.

Moreover, Figure 12 depicts the temperature distribution across each layer which clearly shows the temperature difference of more than 10°C between the two cases for the applied power loss per chips.

12. Transient Thermal Computation of the Power Module, with/without Graphene layer using WLTC Driving Cycle

Since the load of the inverter is constantly varying, hence the real-time thermal computations are necessary to more accurately quantify the thermal behavior and reliability of the inverters.

Therefore, the time-dependent solver in COMSOL is used to evaluate the transient temperature distribution across the power module from the chips towards the bottom layers over the WLTC driving schedule consisting of both the urban and highway phases. Currently in many countries including the EU, WLTC is the main drive test procedure for light-duty vehicles. First based on the vehicle model used in this study, the required torque and speed values for the PMSM model are determined. The values are then utilized for the power loss calculation over the WLTC drive test

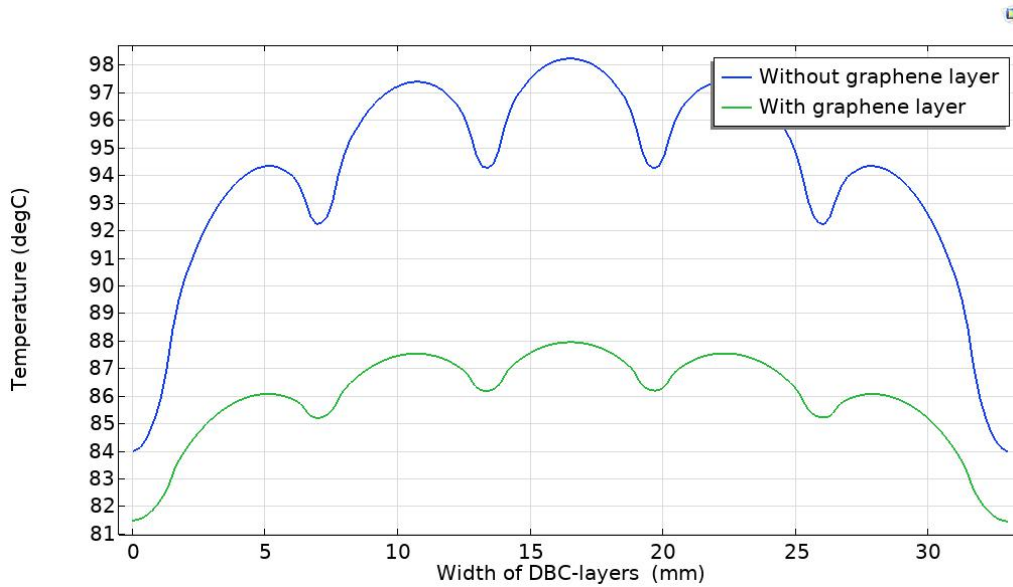


Figure 12: Temperature distribution across each layer with/without graphene.

Table 6: Vehicle Parameters

Vehicle Parameter	Value	Unit
Total mass of the vehicle	1700	[kg]
Aerodynamic drag coefficient	0.35	[-]
Vehicle cross section area	2	[m ²]
Rolling friction coefficient	0.007	[-]
Wheel radius	0.3	[m]

and thereafter for the thermal loading in COMSOL. Table 6 gives the vehicle model parameters.

In order to measure the transient heat distribution, a 3D-cutline from the middle chip to the baseplate is used similarly as done in the steady-state simulation and then the calculated power losses over WLTC are applied to the chips as the heat sources.

Figure 13 illustrates the calculated power losses in the WLTC, and Figure 14 shows the temperature distribution across the power module for the two cases (with/without) graphene. The thermal profile comparison per chip between the case of the power module with the assembled graphene film and without the graphene film is depicted in Figure 15. As can be seen in Figure 13, the highest calculated power losses are observed around 1000 second for the related speeds and torques of the PMSM based on the used vehicle model parameters. Therefore, the temperature distributions for the two cases, with/without graphene are considered at 1033 s.

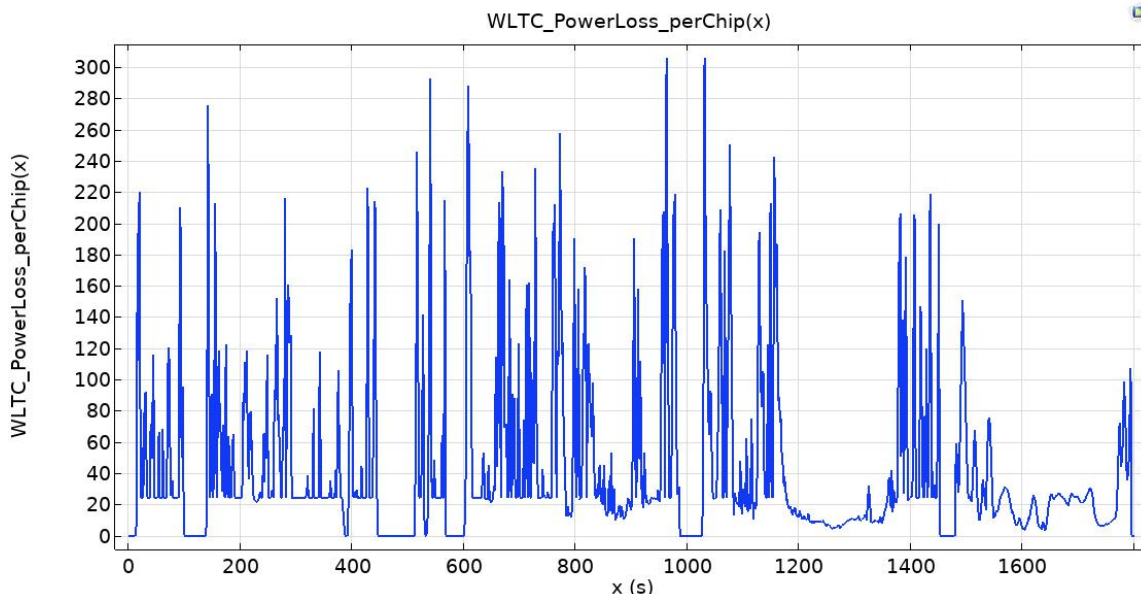


Figure 13: Calculated power losses over WLTC

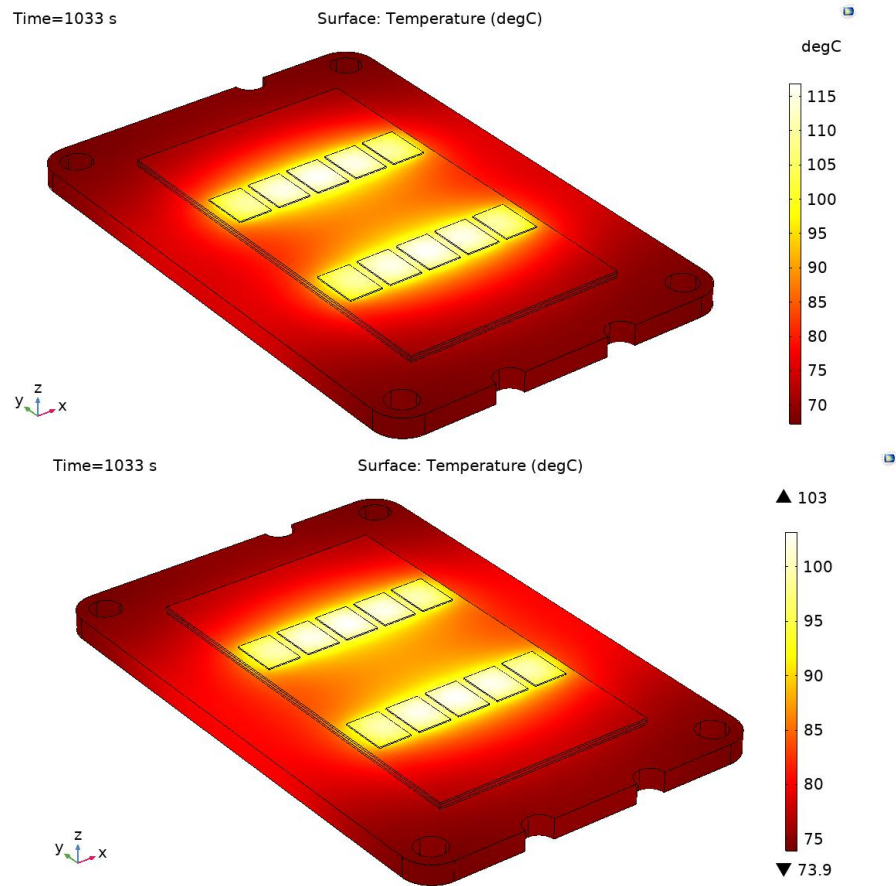


Figure 14: Temperature distribution across the power module without the graphene layer (top) and with the graphene layer (bottom) at 1033 s.

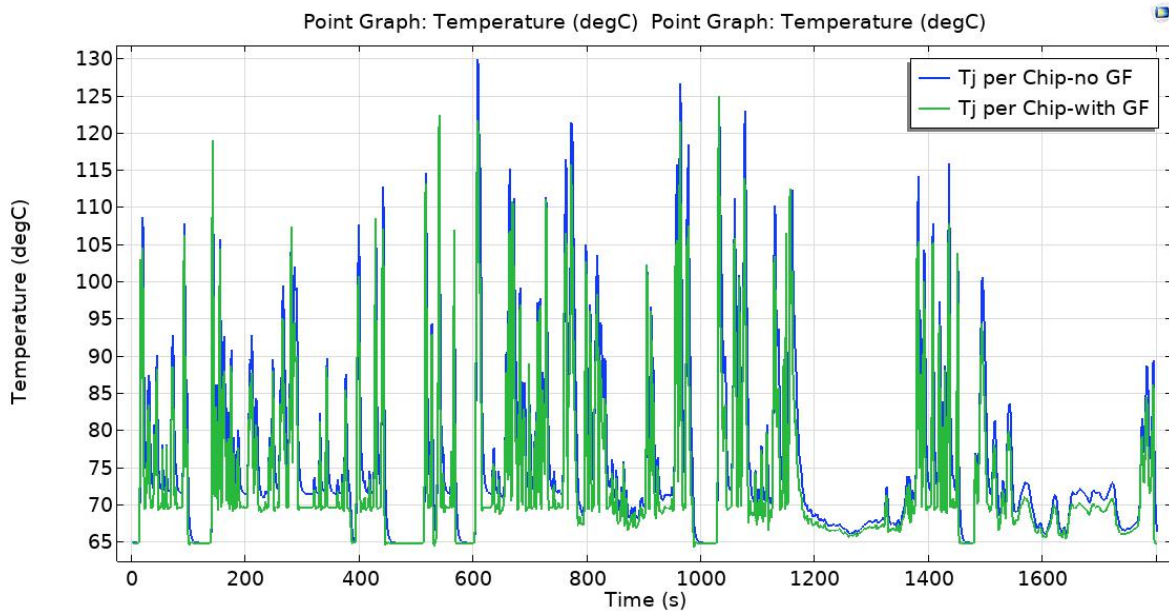


Figure 15: Comparison of the average junction temperature profile of each chip over the WLTC drive cycle.

According to Figure 14, a temperature reduction of about 15°C per chip at 1033 s is observed in case of using the graphene layer in the power module. Moreover, a close look into the heat distribution over the module shows a flatter distribution in case of using graphene. Also, with graphene the heat is more distributed across the baseplate compared to the case without graphene in www.astesj.com

which the heat is more concentrated in the middle, where the chips are located. Quite clearly proven, graphene as a heat conductor is improving the heat spreading across the module and this leads to an effectiveness of the cooling of the semiconductor power module.

Finally, Figure 15 illustrates the thermal profile comparison over the WLTC driving schedule per chip between the two cases, with and without the graphene layer. As depicted, the graphene layer gives a lower average junction temperature per chip and this can lead to a lower operating temperature of the semiconductor power module in general. Worth mentioning is that this effect can be more apparent in high-power applications where due to the high voltage and current levels used, a high number of chips are needed. Also, as observed before in the steady-state simulation, boosting the heat dissipation performance in the chips and power module by using graphene, leads to a noticeable reduction in the junction to case thermal resistance of the power module and decreasing the operating temperature accordingly.

13. Conclusion

In this paper, the impact of neglecting the thermal feedback on the conduction losses for two SiC inverters were determined. The effect is demonstrated for a chosen heavy operating point of a PMSM, high torque and low speed, which is a difficult operating point from a power loss perspective. Since accounting for the thermal feedback shows a considerable increase in the inverter total conduction losses, therefore, graphene as a novel thermal management material is utilized as a layer in the power module packaging to improve the thermal dissipation and cooling of the SiC chips in the power module. A steady-state as well as a transient thermal computation in COMSOL Multiphysics are carried out on the investigated SiC power module which is created in a way to be as close as possible to a real module in terms of layers material, electrical and thermal aspects.

The thermal computation results imply a noticeable improvement in thermal spreading over the module which leads to a lower thermal resistance from junction to case of the module and a lower operating temperature in both the steady-state and transient computations. The transient computations have been applied over a WLTC drive cycle to investigate the thermal variations of the SiC chips when the load varies in a hybrid vehicle powertrain.

Worth mentioning is that, inserting the graphene assembled films in the SiC power module and functionalizing the films to reduce the contact thermal resistances in a standard electronics lab is very challenging due to a very thin and flexible texture of the graphene material. In addition, different variables including the inner temperatures, the heat distribution, and the streamlines, etc., within the multiple layers from the SiC chips down to the whole bottom layers cannot be easily measured on the hardware implementation even with many sensors. Therefore, using FEM in COMSOL Multiphysics is found to be a reliable method to simulate the thermal aspects precisely and prove the concepts of this paper.

To sum up, with this study, the application of using graphene in the building of the power module demonstrates a promising potential to boost the heat spreading in the semiconductor power modules especially in the high-power applications where also lightweight is required as priority in the light-duty vehicles.

As most important future work, measurement verification is the next step.

Conflict of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this article.

References

- [1] J. Reimers, L. Dorn-Gomba, C. Mak, A. Emadi, "Automotive traction inverters: Current status and future trends," *IEEE Transactions on Vehicular Technology*, **68**(4), 3337–3350, 2019.
- [2] K. Kumar, M. Bertoluzzo, G. Buja, "Impact of SiC MOSFET traction inverters on compact-class electric car range," in 2014 IEEE International Conference on Power Electronics, Drives and Energy Systems (PEDES), 1–6, 2014, doi:10.1109/PEDES.2014.7042156.
- [3] J. Biela, M. Schweizer, S. Waffler, J.W. Kolar, "SiC versus Si—Evaluation of Potentials for Performance Improvement of Inverter and DC–DC Converter Systems by SiC Power Semiconductors," *IEEE Transactions on Industrial Electronics*, **58**(7), 2872–2882, 2011, doi:10.1109/TIE.2010.2072896.
- [4] K. Olejniczak, T. McNutt, D. Simco, A. Wijenayake, T. Flint, B. Passmore, R. Shaw, D. Martin, A. Curbow, J. Casady, B. Hull, "A 200 kVA electric vehicle traction drive inverter having enhanced performance over its entire operating region," in 2017 IEEE 5th Workshop on Wide Bandgap Power Devices and Applications (WiPDA), 335–341, 2017, doi:10.1109/WiPDA.2017.8170569.
- [5] S.E. Schulz, "Exploring the High-Power Inverter: Reviewing critical design elements for electric vehicle applications," *IEEE Electrification Magazine*, **5**(1), 28–35, 2017, doi:10.1109/MELE.2016.2644281.
- [6] J. Rąbkowski, T. Plątek, "Comparison of the power losses in 1700V Si IGBT and SiC MOSFET modules including reverse conduction," in 2015 17th European Conference on Power Electronics and Applications (EPE'15 ECCE-Europe), 1–10, 2015, doi:10.1109/EPE.2015.7309444.
- [7] J.L. Nan Wang, Majid Kabiri Samani, Hu Li, Lan Dong, Zhongwei Zhang, Peng Su, Shujing Chen, Jie Chen, Shirong Huang, Guangjie Yuan, Xiangfan Xu, Baowen Li, Klaus Leifer, Lilei Ye, Tailoring the Thermal and Mechanical Properties of Graphene Film by Structural Engineering, *Small*, **14**(29), 1801346, doi:10.1002/sml.201801346.
- [8] T.M. Tritt, *Thermal conductivity: theory, properties, and applications*, Springer, 2010.
- [9] L. Peng, Z. Xu, Z. Liu, Y. Guo, P. Li, C. Gao, "Ultrahigh Thermal Conductive yet Superflexible Graphene Films," *Advanced Materials*, **29**(27), 2017, doi:10.1002/ADMA.201700589.
- [10] A.A. Balandin, S. Ghosh, W. Bao, I. Calizo, D. Teweldebrhan, F. Miao, C.N. Lau, "Superior thermal conductivity of single-layer graphene," *Nano Letters*, **8**(3), 902–907, 2008, doi:10.1021/NL0731872.
- [11] Y. Liu, S. Chen, Y. Fu, N. Wang, D. Mencarelli, L. Pierantoni, H. Lu, J. Liu, "A lightweight and high thermal performance graphene heat pipe," *Nano Select*, **2**, 2020, doi:10.1002/nano.202000195.
- [12] N.J. Song, C.M. Chen, C. Lu, Z. Liu, Q.Q. Kong, R. Cai, "Thermally reduced graphene oxide films as flexible lateral heat spreaders," *Journal of Materials Chemistry A*, **2**(39), 16563–16568, 2014, doi:10.1039/C4TA02693D.
- [13] G. Xin, H. Sun, T. Hu, H.R. Fard, X. Sun, N. Koratkar, T. Borca-Tasciuc, J. Lian, "Large-area freestanding graphene paper for superior thermal management," *Advanced Materials*, **26**(26), 4521–4526, 2014, doi:10.1002/ADMA.201400951.
- [14] Z. Liu, Z. Li, Z. Xu, Z. Xia, X. Hu, L. Kou, L. Peng, Y. Wei, C. Gao, "Wet-spun continuous graphene films," *Chemistry of Materials*, **26**(23), 6786–6795, 2014, doi:10.1021/CM5033089.
- [15] J. Li, F. Ye, S. Vaziri, M. Muhammed, M.C. Lemme, M. Östling, "Efficient inkjet printing of graphene," *Advanced Materials*, **25**(29), 3985–3992, 2013, doi:10.1002/ADMA.201300361.
- [16] Y. Fu, J. Hansson, Y. Liu, S. Chen, A. Zehri, M.K. Samani, N. Wang, Y. Ni, Y. Zhang, Z. Bin Zhang, Q. Wang, M. Li, H. Lu, M. Sledzinska, C.M.S. Torres, S. Volz, A.A. Balandin, X. Xu, J. Liu, "Graphene related materials for thermal management," *2D Materials*, **7**(1), 2020, doi:10.1088/2053-1583/AB48D9.
- [17] N. Wang, Y. Liu, S. Chen, L. Ye, J. Liu, "Highly Thermal Conductive and Electrically Insulated Graphene Based Thermal Interface Material with Long-Term Reliability," in 2019 IEEE 69th Electronic Components and

- Technology Conference (ECTC), 1564–1568, 2019, doi:10.1109/ECTC.2019.00240.
- [18] M.F. Abdullah, S.A. Mohamad Badaruddin, M.R. Mat Hussin, M.I. Syono, "EVALUATION ON THE REDUCED GRAPHENE OXIDE THERMAL INTERFACE MATERIAL AND HEAT SPREADER FOR THERMAL MANAGEMENT IN HIGH-TEMPERATURE POWER DEVICE," *Jurnal Teknologi*, **83**(3 SE-), 53–59, 2021, doi:10.11113/jurnalteknologi.v83.15102.
- [19] J.S. Lewis, Reduction of Device Operating Temperatures with Graphene-Filled Thermal Interface Materials, *C*, **7**(3), 2021, doi:10.3390/c7030053.
- [20] Y. Liu, T. Thiringer, N. Wang, Y. Fu, H. Lu, J. Liu, "Graphene based thermal management system for battery cooling in electric vehicles," in 2020 IEEE 8th Electronics System-Integration Technology Conference (ESTC), 1–4, 2020, doi:10.1109/ESTC48849.2020.9229856.
- [21] N. Wang, Y. Liu, L. Ye, J. Li, "Highly Thermal Conductive and Light-weight Graphene-based Heatsink," in 2019 22nd European Microelectronics and Packaging Conference Exhibition (EMPC), 1–4, 2019, doi:10.23919/EMPC44848.2019.8951839.
- [22] S. Huang, Y. Zhang, S. Sun, X. Fan, L. Wang, Y. Fu, Y. Zhang, J. Liu, "Graphene based heat spreader for high power chip cooling using flip-chip technology," in 2013 IEEE 15th Electronics Packaging Technology Conference (EPTC 2013), 347–352, 2013, doi:10.1109/EPTC.2013.6745740.
- [23] S. Huang, Y. Zhang, N. Wang, N. Wang, Y. Fu, L. Ye, J. Liu, "Reliability of graphene-based films used for high power electronics packaging," in 2015 16th International Conference on Electronic Packaging Technology (ICEPT), 852–855, 2015, doi:10.1109/ICEPT.2015.7236714.
- [24] S. Huang, J. Bao, H. Ye, N. Wang, G. Yuan, W. Ke, D. Zhang, W. Yue, Y. Fu, L. Ye, K. Jeppson, J. Liu, "The effects of graphene-based films as heat spreaders for thermal management in electronic packaging," in 2016 17th International Conference on Electronic Packaging Technology (ICEPT), 889–892, 2016, doi:10.1109/ICEPT.2016.7583272.
- [25] P. Zhang, N. Wang, C. Zandén, L. Ye, Y. Fu, J. Liu, "Thermal characterization of power devices using graphene-based film," in 2014 IEEE 64th Electronic Components and Technology Conference (ECTC), 459–463, 2014, doi:10.1109/ECTC.2014.6897324.
- [26] M.R.M. Hussin, M.M. Ramli, S.K.W. Sabli, I.M. Nasir, M.I. Syono, H.Y. Wong, M. Zama, "Fabrication and characterization of graphene-on-silicon schottky diode for advanced power electronic design," *Sains Malaysiana*, **46**(7), 1147–1154, 2017, doi:10.17576/JSM-2017-4607-18.
- [27] J. Wang, L. Shi, Y. Li, L. Jin, Y. Xu, H. Zhang, Y. Zou, Y. Lan, X. Ma, "Thermal management of graphene-induced high-power semiconductor laser package with bidirectional conduction structure," *Optics and Laser Technology*, **139**, 2021, doi:10.1016/J.OPTLASTEC.2021.106927.
- [28] K. Hemanjaneyulu, M. Khaneja, A. Meersha, H.B. Variar, M. Shrivastava, "Comprehensive Computational Modelling Approach for Graphene FETs," in 2018 4th IEEE International Conference on Emerging Electronics (ICEE), 1–5, 2018, doi:10.1109/ICEE44586.2018.8937909.
- [29] S. Amirpour, T. Thiringer, D. Hagstedt, "Power Loss Analysis in a SiC/IGBT Propulsion Inverter Including Blanking Time, MOSFET's Reverse Conduction and the Effect of Thermal Feedback Using a PMSM Model," in IECON 2020 The 46th Annual Conference of the IEEE Industrial Electronics Society, 1424–1430, 2020, doi:10.1109/IECON43393.2020.9254297.
- [30] Wolfspeed, CAB450M12XM3, Nov. 2021.
- [31] Wolfspeed, CAS300M12BM2, Nov. 2021.
- [32] A. Acquaviva, T. Thiringer, "Energy efficiency of a SiC MOSFET propulsion inverter accounting for the MOSFET's reverse conduction and the blanking time," in 2017 19th European Conference on Power Electronics and Applications (EPE'17 ECCE Europe), P.1-P.9, 2017, doi:10.23919/EPE17ECCEurope.2017.8099052.
- [33] J.W. Kolar, H. Ertl, F.C. Zach, "Influence of the modulation method on the conduction and switching losses of a PWM converter system," *IEEE Transactions on Industry Applications*, **27**(6), 1063–1075, 1991, doi:10.1109/28.108456.
- [34] SEMIKRON, Application Manual Power Semiconductors, Ilmenau, 2015.
- [35] M. Meyer, J. Bocker, "Optimum Control for Interior Permanent Magnet Synchronous Motors (IPMSM) in Constant Torque and Flux Weakening Range," in 2006 12th International Power Electronics and Motion Control Conference, 282–286, 2006, doi:10.1109/EPEPEMC.2006.4778413.
- [36] M. Meyer, T. Grote, J. Bocker, "Direct torque control for interior permanent magnet synchronous motors with respect to optimal efficiency," in 2007 European Conference on Power Electronics and Applications, 1–9, 2007, doi:10.1109/EPE.2007.4417370.
- [37] L. Ceccarelli, R.M. Kotecha, A.S. Bahman, F. Iannuzzo, H.A. Mantooth, "Mission-Profile-Based Lifetime Prediction for a SiC mosfet Power Module Using a Multi-Step Condition-Mapping Simulation Strategy," *IEEE Transactions on Power Electronics*, **34**(10), 9698–9708, 2019, doi:10.1109/TPEL.2019.2893636.
- [38] E. EDGE, Convective heat transfer coefficients table chart, Nov. 2021.
- [39] H. Han, Y. Zhang, N. Wang, M.K. Samani, Y. Ni, Z. Y. Mijbil, M. Edwards, S. Xiong, K. Sääskilähti, M. Murugesan, Y. Fu, L. Ye, H. Sadeghi, S. Bailey, Y.A. Kosevich, C.J. Lambert, J. Liu, S. Volz, "Functionalization mediates heat transport in graphene nanoflakes," *Nature Communications*, **7**(1), 11281, 2016, doi:10.1038/ncomms11281.

Control and Monitoring Systems in Medium Voltage Distribution Networks in Poland - Current Status and Directions of Development

Janusz Gurzyński*, Łukasz Kajda, Marcin Tarasiuk, Tomasz Samotyjak, Zbigniew Stachowicz, Sławomir Kownacki

Institut Energetyki – Instytut Badawczy, 01-330 Warszawa, Mory 8, Oddział Gdańsk, Poland

ARTICLE INFO

Article history:

Received: 04 October, 2021

Accepted: 26 November, 2021

Online: 17 December, 2021

Keywords:

Communication

Detection of short circuits

ICT security

Medium voltage network

Telemetric control

ABSTRACT

The paper describes the control and monitoring systems installed in medium voltage networks by Polish distribution network operators. It also outlines the expected directions of development of these systems, specifies the functions of the individual system components and describes the requirements applicable to them. In particular, attention is paid to the implementation of functions to detect short-circuits and automatic voltage control on the low voltage side of the transformer using the on-load tap changer. It also describes the communication functions implemented by the control and monitoring systems and the methods of ensuring a guaranteed power supply for the above systems.

1. Introduction

Similarly, to the EU and other countries, the power network in Poland is split into a transmission network and a distribution network. In Poland, the transmission network consists of extra high voltage (EHV) lines with a rated voltage of 400 kV and 220 kV. The distribution network is made up of:

- high voltage (HV) lines with a rated voltage of 110 kV,
- medium voltage (MV) lines with rated voltage of 15-30 kV,
- low voltage (LV) lines with a rated voltage of 0.4 kV.

In Poland, there is one transmission system operator, whereas distribution networks are controlled and managed by five large distribution system operators (DSO) and several smaller local DSOs. Correct operation of the DSOs is supervised by the Energy Regulatory Office (ERO). This is a central governmental administration authority to regulate the fuel and energy market in Poland. The electricity market in Poland comprises more than 17.8 million consumers that consume approx. 175 TWh of electricity every year [1].

The medium voltage distribution networks in Poland consist of overhead lines with a length of more than 225 thousand kilometres and cable lines with a total length of more than 85 thousand kilometres. In 2018, the number of medium voltage substations exceeded 265 thousand.[1].

The condition of automation of MV networks in Poland has significantly improved in recent years due to the increase in the number of installed remote control and monitoring systems and as the extension and retrofit of SCADA systems. This is caused by the requirements to ensure the reliability of supply and keeping of the appropriate quality of electricity supplied. Currently, virtually all newly built substations installed within medium voltage distribution networks are equipped with telemechanics systems. A number of the existing substations are overhauled and equipped with telemechanics systems as well. These actions result from the DSO's seeking the reduction of the time of restitution of consumer supply in case of failure, and thus the improvement of the area CTP (duration of the outage) and CP (frequency of the outage) indicators defined and required by the ERO. CTP is the area indicator for the interruption of electricity supply to consumers (the equivalent of SAIDI), whereas CP is the indicator for the frequency of interruptions of electricity supply to consumers in each area (the equivalent of SAIFI).

The requirements of the ERO towards the DSOs in the area of improvement of CTP and CP indicators in the coming years are included in the document [2].

The duration of interruptions in electricity supply to consumers is reduced by installing automation systems in MV networks within cable network substations and at load switches in overhead lines. These systems allow for remote control of switches and reconfiguration of MV networks so that at the moment of failure,

*Corresponding Author Janusz Gurzyński, j.gurzynski@ien.gda.pl

the largest possible number of consumers can be supplied with electricity while shutting down the smallest possible fragment of the MV network.

2. Automation systems installed in the MV network

Telemechanics systems in medium voltage substations to which MV/LV transformers are connected are installed in parallel with devices that are elements of the Advanced Metering Infrastructure (AMI), such as balancing meters to measure energy on the low voltage side of MV/LV transformers. At certain DSOs, data concentrators are installed at the stations to enable communication with electricity meters installed at consumers and prosumers on the low voltage end, using the Power Line Communication (PLC) technique. Together with the telemetric control and monitoring systems, these devices use the communication system shared with the monitoring centres.

Along with the telemechanics systems, in medium voltage substation equipped with MV / LV transformers devices that are AMI (Advanced Metering Infrastructure) elements, such as balancing meters and data concentrators are installed.

Balancing meters measure energy on the low side of MV / LV transformers while data concentrators gather information from electricity meters installed at consumers and prosumers on low voltage, using the PLC (Power Line Communication) technique. For communication with the control centre, a common communication channel is used both by AMI and the telemechanics system.

The automation systems used for the control of MV overhead line load switches (as shown in Figure 1) and in MV cable network distribution substations (as shown in Figure 2) are made up of the below components:

- A telemetric control and monitoring system (telemechanics controller (RTU)) - allows for remote control and monitoring of load switches (LS) and MV circuit breakers (CB) equipped with an autonomous protection device (A.RELAY). The system collects and provides to the monitoring system (SCADA) binary signals (BI/BO) and additional information, e.g. concerning the status of fuses in the low voltage substation supplied from the MV/LV transformer (TR), and the operation of the guaranteed supply system and the condition of the overall facility (OVERALL) where it is installed.
- Fault detection systems (FD) - detect and indicate single-phase short circuits (ground faults) and phase-to-phase short circuits in the MV network. The system also allows for the measurement of voltage and current in the medium-voltage network. The fault detection systems may be independent devices or the fault detection function may be implemented in the telemechanics controller (as shown in the Figure 2). The current and voltage transformers or sensors (VT) (CT) are the source of measurement signals.
- Primary communication system (TETRA) based on the dedicated modem, connected with a controller using a serial link (COM) - provides communication with the dispatch centre.
- The backup communication system (GSM) ensures communication with the dispatching centre in case of failure of the primary communication system. It may be installed as

a controller module (as shown in Figure 1) or as a separate GSM router (as shown in Figure 2) connected to the controller using the Ethernet link (ETH).

- Guaranteed power supply system (PS) with a 24 VDC battery (ACCU) - allows for the operation of the entire system in case of 230 V AC power supply voltage loss.
- At least one of the communication systems (primary, backup system) also functions as an engineering monitoring channel, which enables access to the configuration of some station devices (i.a. the telemechanics controller) through a WWW server or dedicated software.
- Except for the above, in a more extended version, the automation systems may contain the below additional systems (as shown in Figure 3):
- Advanced field terminals and protection devices (RELAY) that communicate with the controller using a serial link or Ethernet link and send statuses and measurements to the dispatcher system.
- Metering system for power quality parameters on the medium and low voltage sides.
- The balancing meter or AMI data concentrator with a balancing meter (AMI DC), that reads the indications of the subscriber energy meters and sends measurements to the settlement system (AMI).

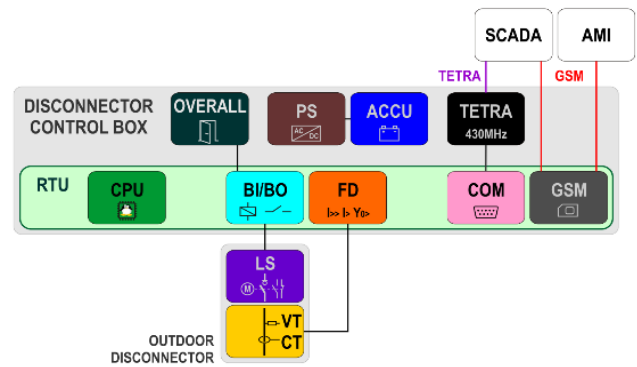


Figure 1: MV overhead switch control system

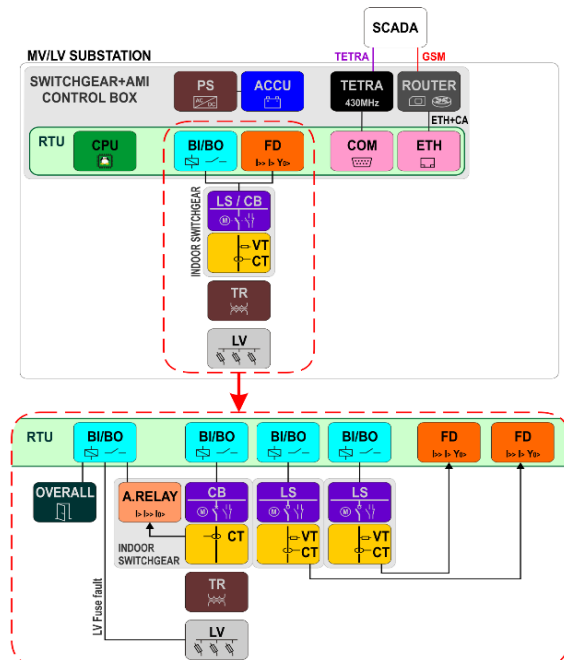


Figure 2: MV/LV substation control system

The automatic voltage control system (OTLC REG) - allowing for low voltage control through control of the on-load tap changer of the transformer (TR OTLC). This is one of the methods for limiting the voltage problems indicated above [3].

Some consumers are prosumers, e.g. have photovoltaic panels connected to the LV network through inverters (PV INV). The number of PV systems in Poland are increasing dynamically, that results in voltage problems, i.e. exceeded threshold voltage values in the LV network permitted under the regulations.

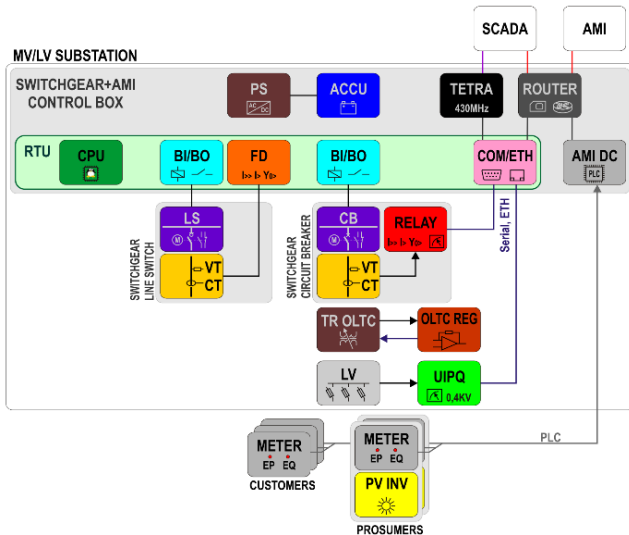


Figure 3: MV/LV station control system - extended option

3. Telemetric control and monitoring systems and systems for fault detection in the MV network

Failures in the medium voltage distribution network are the main reasons for the lack of energy supply to consumers [4], [5]. On average, 75% of the duration interruptions are associated with failures in the MV network, in particular with single-phase (earth) faults. The correct location of the fault within the network is of primary significance for the process of isolation of the damaged network section and restoration of the power supply in the remaining area.

The detection of earth faults based on current measurement only is sometimes ambiguous. The emergence of the low power voltage transformers (LPVT) and low power current transformers (LPCT) on the market allowed the application of conductance criteria (Y_0 , G_0 , B_0) which required the voltage measurement additionally [6]. LPVTs installed directly in the separable connectors of MV switchgear act as the resistant or capacity voltage dividers. As for the LPCT the Rogowski coils are used most frequently. Due to the size and safety of use (low voltage on the secondary side), these devices are particularly suitable for use in MV switchgears installed within the network and in overhead line load switches.

Implementing within the MV network monitoring of current and voltage and reliable detection of short circuit current flow for earth faults and phase-to-phase faults enables quick detection of a damaged network section. Remote control of switches makes it possible to isolate the damaged network element and supply power, as far as possible, to other network sections within a timeframe of no more than 3 minutes. This duration was imposed by the Energy Regulatory Office in the requirements presented in

[2], according to which power supply interruptions of up to 3 minutes are not included in the interruption duration indicator. It is not always possible to adhere to this time interval because the network operator has to have time to identify the existing situation, develop the switching strategy, assess the consequences of actions it wants to undertake, and carry out the sequence of switching operations.

The time needed for the performance of these actions may be reduced by implementing restitution automatics known as the Fault Detection, Isolation and Restitution (FDIR) [4], also referred to as the Self-Healing Grid (SHG) or Fault Location, Isolation and Service Restoration (FLISR). The task of the FDIR systems is to automatically analyze the network configuration [7] and perform the necessary switching operations within a timeframe of no more than 3 minutes.

The purposes of the FDIR are, i.e.:

- Reduction of the faulty network area.
- Reduction of the time of failure to supply electricity to consumers.
- Support for the organization of work of the fault-clearing services teams.
- Reduction of the interruption duration indicator and interruption occurrence frequency.

The FDIR is most often implemented in one of the three methods described below:

- Central automation system - the switches within the network are equipped with controllers with fault indicators. Information on the current network status (i.e. on the status of switches) and occurrence of faults is sent to a central system on an ongoing basis, where the sequence of switching operations to be carried out is developed to isolate the fault location and restore the power supply to the maximum number of consumers.
- Local automation system - within the MV network, circuit breakers (CB) are installed equipped with protection devices with different time delay settings, allowing for CB operation's time selectivity in case of a fault circuit. Only the circuit breakers closest to the fault location are opened and the tie switch downstream the fault location is closed. More advanced mode of load switches operation is based on the exchange of the information related to the switch status between the controllers to develop and perform the switching sequence.
- Mixed automation system - uses both central and local automation system features. Its main advantage is the possibility to change the protection devices settings fitting them to the new network scheme, which was created after the switching operation. The new settings are calculated by the central system.

In Poland, the FDIR automation system implemented centrally, locally or in a mixed manner has been tested by the largest DSOs in pilot systems and is currently at the production deployment stage.

4. Automatic voltage control systems and measurement of electricity quality indicator

The installation of a large number of Distributed Energy Resources (DER), mainly PV, in the LV network, results in the

DSO facing new challenges relating to compliance with the parameters of voltage supplying the consumers [8]. Even small power generating units connected directly to the low voltage network may result, in the case of low network load, in the increase in voltage above the permitted range. In turn, the decrease in LV at the MV/LV transformer output will fail to comply with the minimum voltage level when the DER sources do not operate.

One of the planned solutions implemented on a pilot basis to facilitate compliance with the permissible voltage levels in the LV network is the use of the transformer with an on-load tap changer [3].

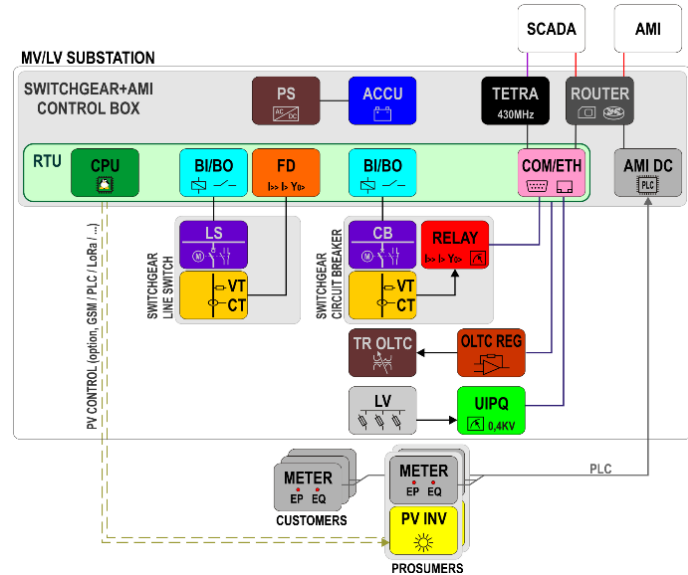


Figure 4: MV/LV station control system - voltage control in the LV network

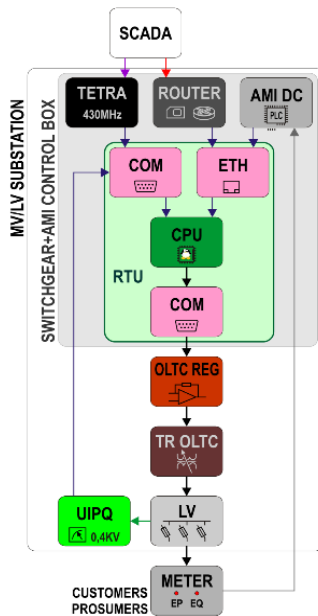


Figure 5: Algorithm for voltage control in the LV network

The algorithms for voltage maintenance in the permitted range are implemented in the telemechanics controller, which may source metering data from voltage measurements of the LV switchgear (as shown in Figure 4), measurements of electricity

quality installed in LV outlet circuits and measurements from within the network obtained from the AMI system devices.

The telemechanics controller may also set the tap changer according to the command from the dispatcher system (as shown in Figure 5).

Other methods to maintain the voltage in the permitted range may comprise, e.g. the use of a symmetrizing transformer within the LV network to balance voltage in the particular phases [8] or remote control systems for DER sources, allowing for the limitation of the energy generated thereby, as described in more detail in [9]. Currently, the exceeded values of permitted voltage levels in the LV network result in the deactivation of PV microgeneration units as a result of activation of overvoltage protection devices in PV inverters, causing the reduction of prosumer profit and voltage oscillations in the network due to subsequent microgeneration activation and deactivation cycles. The extension of the telemechanics controller to include PV microgeneration unit control functions (control of reactive power consumption and reduction or disconnection of active power generation) will allow for the elimination of uncoordinated deactivation of PV microgeneration units and the related voltage oscillations.

In the locations with a large share of PV microgeneration units, the planned use of the metering systems for electricity quality indicators in the LV network, monitored by the telemechanics controller with communication with the SCADA system in the dispatcher centre, will allow for increasing the observability of the LV network and will provide data necessary for technical and economic analyses associated with the extension of LV networks.

5. Communication with dispatcher systems

The most common communication method used by the DSO involves the connection with the telemechanics controller through access using the cellular network purchased from an external provider. The communication may be affected through the modem installed in the telemechanics controller or through an external router with an appropriate modem. The solution including the external router is most often used when the same communication channel is also used by other systems installed in a substation. Another communication system often used by the DSO is the Terrestrial Trunked Radio (TETRA) system. This system is used for voice communication between the dispatchers and the personnel working in the field, but also enables the transmission of data, hence it is used in the control and monitoring systems of medium voltage networks of several DSOs in Poland. TETRA is not a generally available system. It is intended for public services and entities performing strategic tasks, such as the power sector. In particular, it is to provide reliable and safe communication when public networks are not available or overloaded (large-scale catastrophes or failures resulting in long-term loss of power supply for cellular network base stations).

In most of the substations located in cities or their vicinity, it is not necessary to install external GSM antennas. Usually, internal antennas are sufficient to establish a stable and efficient connection. The only exception is substations located in cellars of buildings several meters below the ground level. In such places, it is necessary to use antennas outside the building. External GSM antennas are also used in telemechanics systems of over-head line load switches which are often located outside urbanized areas.

The situation is different in the case of the TETRA infrastructure which uses lower frequency - approx. 450 MHz, which allows for the installation of base stations at greater distances. Due to the smaller number of base stations, in all facilities equipped with the TETRA infrastructure, antennas are installed outside buildings and, if possible, above the roofline (to ensure the “visibility” of the base station).

Irrespective of the communication system applied the communication with the dispatching systems always uses one of the protocols used in Poland: DNP 3.0 or IEC 60870-5-104.

It is expected that like the primary MV substations, in the future solutions for the secondary substations, the integrated telemechanics controller will be replaced with devices installed in the individual MV substation bays (as shown in Figure 6).

In each MV substation bay, autonomous devices will be installed for the implementation of all functions in this bay: remote and local control of the load switch or circuit breaker, handling the set of sensors for the measurement of voltage and current, detection of faults and faults and events recording.

The division of the telemechanics controller into several independent devices requires the use of an efficient and easy to implement communication system. The use of the standard described in IEC 61850 for the communication between the central controller and the field controllers is one of the solutions which attracts great interest. This standard uses the Ethernet interfaces that ensure high transmission speed and low delays in data transmission. It has been used for several years at high voltage and extra high voltage stations. At present, the second edition of the IEC 61850 standard is most often implemented in the devices, as numerous errors occurring in the first edition have been corrected and the number of modelled devices has been significantly extended.

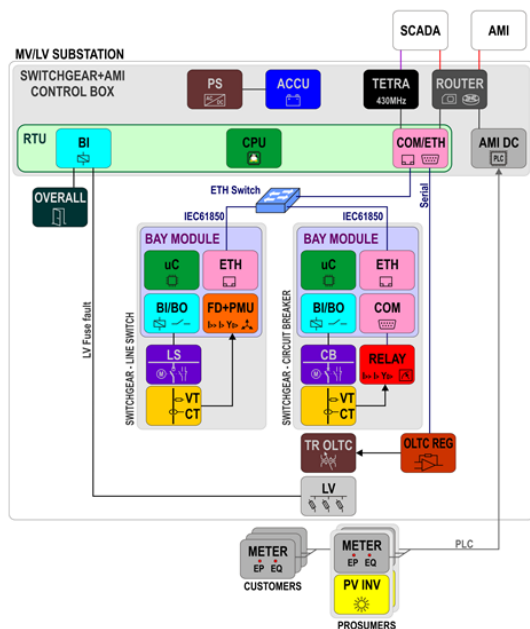


Figure 6: Distributed MV/LV station control system under IEC61850

The central controller located in the substation will be responsible for collecting general information (not associated with the individual medium voltage substation bays), e.g. information

on the power supply system operation and door opening, and will function as a concentrator of data from field controllers and translator of protocols from IEC 61850 to DNP 3.0 or IEC60870-5-104. Currently, pilot projects for this type of system are carried out, one of them being a demonstrative system implemented in Poland within the EUniversal project, in which the authors of this paper are participating.

6. Security in communication systems

The power system is a critical infrastructure system, hence its protection against external threats is particularly important. One of the protection elements is the provision of ICT security to process (OT) systems, which comprise the dispatcher system along with the used remote transmission system and smart field devices that communicate with the dispatcher system [10].

The modern telemechanics systems implemented in Poland in the MV networks, as the OT system components, are equipped with mechanisms protecting against possible cyber-attacks. The security mechanisms implemented in the telemechanics controllers enable the implementation of the OT system security policy adopted at a given DSO and ensure, i.e.:

- Mutual authentication and authorization of the communicating parties.
- Confidentiality and integrity of transmitted data.

Most of the Polish DSOs require the implementation of the above mechanisms to be compliant with the IEC 62351 standard commonly applied in the power sector. For the authentication and authorization, it is required to implement the support for the Public Key Infrastructure (certificates) (IEC 62351-9) in the telemechanics controllers, and to ensure data confidentiality and integrity, the support for the Transport Layer Security (TLS) protocol implementation is required in the presentation and transport layer (IEC 62351-3) of the communication architecture.

The authors of the paper envisage that in addition to the above-mentioned mechanisms, in the telemechanics controllers, additional functions relating to ICT security will be implemented to enable adjustment of controller configuration to the changing requirements of a given DSO, such as:

- Automation of certificate exchange, e.g. the use of the Simple Certification Enrollment Protocol (SCEP).
- Support for the Internet Protocol Security (IPSec) using certificates and for the Virtual Private Network (VPN).
- Support for the IEEE 802.1X standard – the function of a supplicant that initiates the request for authentication using certificates.
- Possibility of disabling not supported communication ports.
- Possibility of changing configuration concerning: addressing, TCP numbers, admissible master server addresses, etc.
- Update of internal software with source authorization and confirmation of new software integrity.
- Implementation of SNMP version 3.

At present, the DSOs, when preparing technical requirements for telemechanics controllers installed in the MV networks,

demand the implementation of command authentication in the application layer following the IEC 62351-5 standard for network protocols such as DNP 3.0 over TCP and IEC 60870-5-104. For authentication and authorization purposes, the Public Key Infrastructure (PKI) certificate mechanism and TLS protocol are used. The protocol ensures a sufficient protection level for the communication that uses the TCP, due to which the requirement for additional authentication in the application layer seems redundant.

7. Guaranteed power supply systems

Concerning issues relating to the guaranteed power supply systems for the telemechanics systems installed in the MV network nodes, particular attention should be paid to the method of supply of secondary circuits of the MV switchgear and switch drives. This is usually a system comprising of a power supply unit with a battery charger and a battery bank. Most DSOs require that the battery capacity be adjusted to ensure the system operation for 24 hours after the power supply voltage loss. In the telemechanics systems installed in MV substations and load switches in overhead lines, batteries with a rated voltage of 24 V and capacity of several to 26 Ah are used. In the case of the majority of larger facilities with a greater number of feeders or equipped with additional devices, e.g. analyzers of network parameters or failsafe relays, larger capacity batteries are used.

In Poland, two solutions for guaranteed power supply systems that ensure sufficient inrush current and current of switch drive operation in the MV network are popular.

In the first solution, the telemechanics system is equipped with a power supply unit with a rated current lower than that required for the switch drive motors, and the backup power supply battery is at the same time a buffer for the power supply of the drive motors also when the primary 230 VAC voltage is present.

In the second solution, the telemechanics system is equipped with a power supply unit that enables current consumption sufficient for the supply of drives, and the battery is only a source of emergency power supply.

The first solution is cheaper and more popular and is suitable for the supply of drives and switchgears that do not consume energy when drives are at rest (switchgears without additional field instrumentation or overhead load switches). In this system, the telemechanics controller is connected to the main power supply unit (with battery charger) output and the drive is connected directly to the battery. When the primary power supply is in place, a temporarily increased current consumption from the battery by the drive motor does not significantly disturb the operation of the charging system of the power supply unit with a battery charger. In the case of primary power supply loss after the battery discharge, the power supply unit will disconnect the main output supplying the telemechanics controller, and the MV switchgear connected directly to the battery terminals will not cause its deep discharge, since it does not consume energy when the drives are not operating. The first solution becomes no longer functional when the MV switchgears are equipped with devices consuming current on a continued basis. These devices include field controllers, MV field protection devices, as well as power supply and control systems for drive operation. In such a situation, the

second solution is used. At present, due to better availability and lower prices of power supply units with battery charger with a high input current, this solution is becoming increasingly popular.

A solution with supercapacitors is also implemented, as supercapacitors buffer energy to supply the drive, which according to the developers of this solution allows for maneuvering the switch when the batteries are worn out. The use of such a system is proposed mainly for overhead load switches.

8. Conclusions

The above described control and monitoring systems for switches in MV networks implemented in the form of centralized telemechanics controllers are typical systems used in recent years by Polish distribution network operators. However, in the nearest future, it is expected that the approach will change, and instead of centralized controllers, distributed telemechanics systems will be applied, which consist of field controllers placed in the particular MV switchgear feeders and a central controller. One of the possible communication solutions that may be used in MV substations is the connection of field controllers and the central controller using an Ethernet communication bus and the IEC 61850 communication standard, and GOOSE communication to send events.

Due to the current and planned dynamic development of PV micro-systems connected to the LV network in Poland and problems with the occurrence of exceeded voltage levels in the LV network, it is assumed that in section of the MV/LV transformer stations, it will be necessary to install automatic low voltage control systems, the various options of which are currently implemented and tested in pilot systems. The DSOs have not specified their requirements for such systems yet. The solution indicated in the paper, based on the use of the on-load tap changer, is one of the possible methods that may improve the parameters of the supply voltage in the LV network.

References

- [1] H. Mikołajuk, G. Parciński, J. Brasse, E. Żarek, D. Zaborska, I. Wrońska, M. Zatorska, H. Bojanowska, E. Stępiak, *Statystyka elektroenergetyki polskiej 2018*, Agencja Rynku Energii S.A., Warszawa, 2019.
- [2] Urząd regulacji Energetyki, *Regulacja jakościowa w latach 2018-2025 dla operatorów systemów dystrybucyjnych*(którzy dokonali , z dniem 1 lipca 2007 r., rozdzielenia działalności), Urząd regulacji Energetyki, Warszawa, 2018.
- [3] C. Long, L.F. Ochoa, "Voltage control of PV-rich LV networks: OLTC-fitted transformer and capacitor banks," *IEEE Transactions on Power Systems*, **31**(5), 4016–4025, 2016, doi:10.1109/TPWRS.2015.2494627.
- [4] A. Babś, Ł. Kajda, "Automatyka restytucyjna w sieciach średnich napięć sposobem na poprawę jakości dostaw energii," *Wiadomości Elektrotechniczne*, **1**(3), 36–41, 2017, doi:10.15199/74.2017.3.6.
- [5] CEER, "CEER benchmarking report 5.2 on the continuity of electricity supply data update," Report on the Continuity of Electricity Supply, (February), 54, 2015.
- [6] J. Linčiks, D. Baranovskis, "Single phase earth fault location in the medium voltage distribution networks," *Scientific Journal of Riga Technical University. Power and Electrical Engineering*, **25**(25), 13–18, 2009, doi:10.2478/V10144-009-0002-6.
- [7] A. Estebasari, E. Patti, L. Barbierato, "Fault detection ,isolation and restoration test platform based on smart grid architecture model using internet-of-things approaches," in *Proceedings - 2018 IEEE International Conference on Environment and Electrical Engineering and 2018 IEEE Industrial and Commercial Power Systems Europe, IEEEIC/I and CPS Europe 2018*, Institute of Electrical and Electronics Engineers Inc., 2018, doi:10.1109/IEEEIC.2018.8494449.
- [8] M. Ozorowski, R. Jąłoza, "Poprawa parametrów jakościowych sieci zasilającej nn z dużym nasyceniem mikroinstalacji PV i stacji ładowania

EVC,” *Wiadomości Elektrotechniczne*, **R. 88**, nr(5), 17–23, 2020, doi:10.15199/74.2020.5.3.

- [9] M.M. Viyathukattuva Mohamed Ali, M. Babar, P.H. Nguyen, J.F.G. Cobben, “Overlaying control mechanism for solar PV inverters in the LV distribution network,” *Electric Power Systems Research*, **145**, 264–274, 2017, doi:10.1016/j.epsr.2016.12.002.
- [10] J. Świdorski, P. Dopierała, M. Świniarski, “Bezpieczeństwo cybernetyczne transmisji danych pomiędzy systemami nadrzędnymi a telemetrycznymi sterownikami obiektowymi na potrzeby energetyki w świetle wymagań normy IEC 62351,” *Wiadomości Elektrotechniczne*, **1(6)**, 6–11, 2019, doi:10.15199/74.2019.6.1.

Associated Risk Factors for the Development of Extensively Resistant Pulmonary Tuberculosis in the First Level of Health Care: From a Public Health Perspective

Mauricio Mamani^{*1}, Mario Chauca², Edward Huamani³, Richard Gonzales³

¹Universidad Privada San Juan Bautista, Lima, 15067, Perú

²Universidad Ricardo Palma, Lima, 15039, Perú

³Federico Villarreal National University, Lima, 15088, Perú

ARTICLE INFO

Article history:

Received: 09 July, 2021

Accepted: 11 November, 2021

Online: 17 December, 2021

Keywords:

Risk factor

Extensively drug-resistant
Tuberculosis

ABSTRACT

To determine the risk factors associated with the development of extensively resistant pulmonary tuberculosis in the DIRIS Lima Sur, 2017. The type of research was observational, quantitative, analytical, retrospective case-control approach. The population consisted of 158 patients with MDR-TB treatment schedule between 2016 and 2017. The sample consisted of 24 cases diagnosed with extensively drug-resistant pulmonary tuberculosis (XDR-TB) and 48 controls with MDR-TB. The survey was used for both groups (cases - controls) as a data collection technique and a questionnaire as an instrument. 5 risk factors associated with XDR-TB were identified. Among the internal factors associated at the bivariate level were: drug use, previous history of MDR-TB/TB treatment, having taken MDR-TB treatment for less than one year, history of failure to primary and individualized scheme for MDR-TB/TB ($p < 0.05$); and the multivariate analysis confirmed the influence of the factor "history of failure to primary and individualized scheme for MDR-TB/TB". Among the external factors most associated at the multivariate level were: having a history of family contact deceased by XDR-TB/MDR-TB and extra domiciliary contact in the neighborhood deceased by TB ($p < 0.05$) It was concluded that the factors associated with the development of extensively resistant tuberculosis are "failure to follow the primary and individualized MDR-TB regimen" and "history of deceased family contact with XDR-TB-MDR-TB".

1. Introduction

Tuberculosis (TB) is a disease caused by bacteria corresponding to the Mycobacterium tuberculosis complex, which usually attacks the lungs, but are not exclusive to them, is considered infectious to have the ability to be transmitted from a sick person to a healthy one, when the first expectorated drops in the form of aerosols [1].

TB, according to the World Health Organization (WHO), is among the ten leading causes of death worldwide, being in 2016 an approximate of 1.7 million deaths, of which 0.4 million had association with HIV, likewise, its statistics reveal that about 10.4 million individuals became ill with TB; as for its resistant forms, in the same year, an estimated 480,000 (360,000 to 600,000) incident cases of MDR-TB, where approximately 190,000 (120,000 to 260,000) would have died from this cause.

Only 123,000 cases were diagnosed, of which 12,240 developed XDR-TB (9% of MDR-TB cases), where countries such as India, the People's Republic of China and the Russian Federation account for 50% of all extremely resistant cases; likewise, in Latin America, Peru, is among eight countries reporting presence of XDR-TB in a total of 117 cases, followed by Brazil, Mexico, Haiti, Argentina, Ecuador, Colombia and Bolivia [2], [3].

Likewise, in 2017 our country detected 31,518 incident cases of TB in all its forms, 27 578 incident cases of pulmonary TB, 16 877 cases of smear-positive TB (FP), 1335 of MDR TB and 83 of XDR-TB, representing TB morbidity statistics of 99.0 per 100 thousand inhabitants, an incidence of 86.7 per 100 thousand inhabitants and a smear-positive incidence of 53.0 per 100 thousand inhabitants.

*Corresponding Author: Mauricio Mamani, maurim130@gmail.com

Peru accounts for 14% of the estimated cases of TB in the Region of the Americas; Metropolitan Lima and Callao report the highest percentage of TB patients in all its forms in the country with 64%, in relation to multidrug-resistant TB (MDR-TB) it reaches 79% and extensively drug-resistant TB (XDR-TB) 70%.

The Metropolitan Lima region is divided into four Integrated Health Network Directorates (DIRIS), which are responsible for the overall health of the population. The Lima City DIRIS is one of the main TB hotspots, as it has the highest concentration of TB cases in all its forms and resistances [4], [5].

The growing incidence of primary resistance and acquired resistance to XDR-TB in the DIRIS Lima Sur, which categorizes it as an epidemiological scenario of high risk of TB transmission, leads us to the particular analysis of the following factors: Prolonged contact with bacilliferous patients and substance abuse (tobacco, alcohol, drugs), whose development is carried out throughout our study and lead us to the decision to execute it and achieve the identification of those risk factors that are associated with the emergence of XDR-TB in MDR-TB patients, and that this useful and accurate information will support the approach of appropriate health policies in order to decrease the morbidity and mortality of the population vulnerable to MDR-TB and XDR-TB and contribute to the control and eradication of TB.

The study of risk factors is of great importance because they will allow to predict or prevent the development of TB disease, which are present in each individual and their variability is according to the socioeconomic development of their nation and

the social status they have, so they can be classified into factors attributable to the individual, environmental factors or environment and socioeconomic factors [6].

Within the typology of risk factors they found two: the risk factors of the external environment or associated with the disease and the risk factors of the internal environment or predictors of a disease [7].

The consumption of harmful substances such as tobacco and alcohol are considered health risks and influence the therapeutic regimen of TB, the former being a common risk factor and of great importance in TB, with negative repercussions on public health [8].

The previous treatment of TB in its resistant forms as an incurable disease, together with the inadequate management of the follow-up of the cases, and an untimely and ineffective treatment, generated the mutation and appearance of new strains with greater resistance, thus the emergence of XDR-TB [9].

The deficiencies in the therapeutic regimen and its non-compliance, bring as a consequence the appearance of more resistant forms of TB (XDR-TB), which circulate freely in the community, increasing the transmission of these strains and their potential increase in morbidity and mortality [10]. XDR-TB statistics do not provide accurate official figures, but each country reports high and increasing figures for the lethality of this type of TB [11].

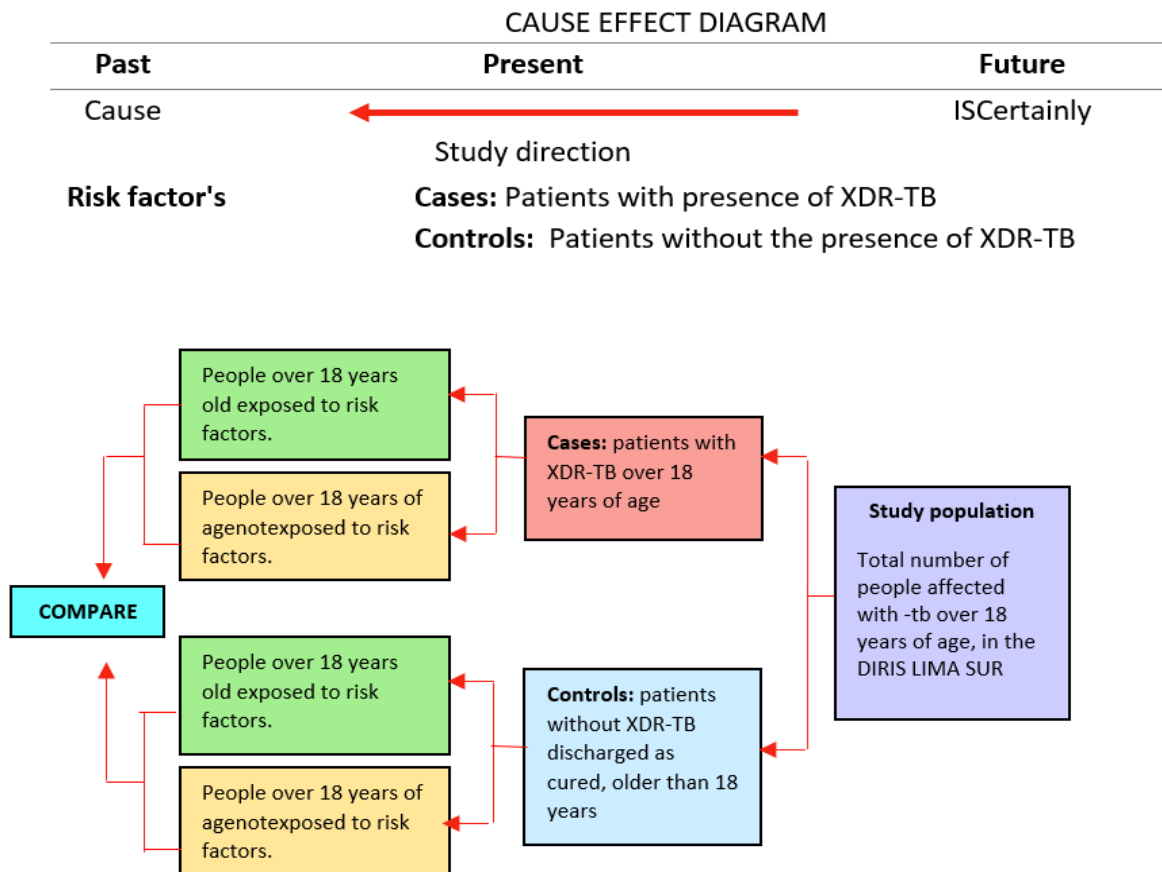


Figure 1: Cause Effect Diagram

The progressive emergence of XDR-TB is of great concern because of its contagiousness and the difficulty to combat and control, which classifies it as potentially irreversible, where the vulnerable population is the most affected, followed by the economically active, in addition to conditions such as previous history of having received treatment and epidemiological background that allow a mechanism of transmission to the person, family and community, exacerbating the deterioration of the health of the entire population; this situation leads to great repercussions not only at the individual level but economic and social for the country, considering it as a public health problem.

To date, XDR-TB remains in the new millennium, the most dangerous human infectious disease, generating high numbers of deaths in the world, however, the production of resistant strains, is not yet defined to what extent would constitute a health problem today, so this article represents the extension of the work originally presented at the 4th International Conference on Medical and Health Informatics [12].

2. Analysis y Method

2.1. Type of research

A retrospective case-control study was developed, with quantitative approach, observational and analytical level, where we identified the risk factors of extensively drug-resistant TB (XDR-TB), applying the epidemiological method [13], [14].

2.2. Variables

The factors are events or phenomena from different sources that predispose people, belonging to a certain region, whose effects could generate the initiation of a disease, that is, that individual who is exposed to a greater number of risk factors will have a greater chance of developing a disease in relation to another less exposed, therefore, will have a high health risk in contrast to one who is away from those factors. [7] A risk factor is an agent that interacts with others; it is a series of detectable particularities in an individual or group way in people, related to the possibility that they are potentiated in influencing the state of health of people. These factors are usually accompanied by three elements and an exposure that predisposes to the development of a certain result (a disease) [15].

Risk factors associated with XDR-TB: These are those characteristics discernible in individuals or groups that are associated with an increased likelihood of developing an unfavorable effect on their health. That is, the exposed individual contracts and develops extensively resistant tuberculosis (XDRTB), this has extended the pattern of resistance to at least two of the most effective drugs of each second line group: the fluoroquinolone group and the second-line injectable group; having among them ciprofloxacin, levofloxacin and moxifloxacin for the first group and amikacin, kanamycin or capreomycin for the second group respectively. [16]

Presence of XDR TB: patients diagnosed with XDR TB, with positive smear or culture results and with a susceptibility panel showing resistance to the most potent first-line antituberculous drugs and to a fluoroquinolone and aminoglycoside as second-line drugs. TB in this form of presentation has virtually no cure, reflects a high mortality rate and keeps the same transmission mechanism

as in its simple form, being its potential affected also their family and close environment [17].

2.3. Population, sample and sampling

The population confirmation for this study was 158 MDR-TB patients diagnosed between 2016 and 2017; and 24 cases who were diagnosed as XDR-TB under treatment. The study area covered health facilities belonging to the Directorate of Integrated Health Networks Lima South (DIRIS). This total was subdivided into two groups of cases and controls. Using Open Epi software, a sample of 24 cases and 48 controls was obtained for unpaired case-controls [18], [19]. For the control group, simple random probability sampling was used, because as this is a subgroup of the population, each constituent element has the same possibility of choice.

2.4. Instruments

The survey was used as a technique and the questionnaire as an instrument, which was applied to both groups: control and cases, in some cases support was sought from medical records, follow-up records, treatment cards and home visits for verification of the information. The validity of the instrument was obtained through expert judgment made up of thematic, methodological and statistical specialists with academic degrees of Doctorate, likewise, it was validated by the technical standard and in other research studies carried out in Peru [20] (The instrument has a reliability of 0.730 (acceptable higher than 0.600) for all items.

2.5. Data collection and analysis procedure

The study began by reviewing the MDR-TB and XDR-TB case follow-up books of thirty-five (35) health facilities of the DIRIS Lima Sur that have a National Health Strategy for the Prevention and Control of Tuberculosis (ESN-PCT), and then selected those patients who met the inclusion criteria and through home visits proceeded to the application of the instruments after informed consent of the participants, having knowledge of the objectives and aims of the study. The data collected were processed by means of the statistical software STATA version 15, and presented in tables, parameters or measures for a better description and analysis of each study variable. In the hypothesis testing, the Odds Ratio (OR) was used in both simple logistic regression and multivariate regression, with a confidence interval (CI) of 95%.

3. Result

3.1. Descriptive analysis

Table 1: Distribution of sociodemographic factors of patients with XDR-TB and MDR-TB in DIRIS Lima Sur, 2017

		Cases		Controls		Total	
		n ₁	%	n ₂	%	n	%
Age	10 – 19	1	4.2	4	8.3	5	6.9
	20 – 25	5	20.8	13	27.1	18	25.0
	26 – 60	17	70.8	25	52.1	42	58.3
	61 to more	1	4.2	6	12.5	7	9.7

Genre	Male	14	58.3	23	47.9	37	51.4
	Female	10	41.7	25	52.1	35	48.6
Level of education	Primary	5	20.8	17	35.4	22	30.6
	Secondary	15	62.5	18	37.5	33	45.8
	Technical superior Univ.	4	16.7	8	16.7	12	16.7
Occupation	Does not work or study	0	0.0	5	10.4	5	6.9
	Work	14	58.3	19	39.6	33	45.8
	Study	7	29.2	16	33.3	23	31.9
	Other	2	8.3	8	16.7	10	13.9
		1	4.2	5	10.4	6	8.3

Source: DIRIS Lima Sur, 2017

Table 1 describes the sociodemographic factors of the patients enrolled in the study. It can be seen that the age from 26 to 60 years is found in 70.8% of the cases and 52.1% of the controls, 58.3% of the cases were males and 52.1% of the controls were females, the secondary education level is found in 62.5% of the cases and 37.5% of the controls and the condition of not working or studying is found in 58.3% of the cases and 39.6% of the controls.

Table 2: Distribution of clinical factors of patients with XDR-TB and MDR-TB in DIRIS Lima Sur, 2017.

		Cases		Controls		Total	
		n1	%	n2	%	n	%
Body Mass Index	Normal	17	70,8	32	66,7	49	68,1
	Overweight	7	29,2	12	25,0	19	26,4
	Obesity	0	0,0	4	8,3	4	5,5
Smear test result (BK)	+	9	37,5	25	52,1	34	47,2
	++	7	29,2	13	27,1	20	27,8
	+++	8	33,3	10	20,8	18	25,0
Cultivation result	+	8	33,3	23	47,9	31	43,1
	++	13	54,2	13	27,1	26	36,1
	+++	3	12,5	12	25,0	15	20,8
Condition of entry	Previously	14	58,3	20	41,7	34	47,2

Hematocrit result	Never treated	10	41.7	28	58.3	38	52.8
	Under	7	29,2	8	16,7	15	20,8
	Normal	17	70,8	40	83,3	57	79,2

Source: DIRIS Lima Sur, 2017.

Table 2 shows that normal BMI was present in 70.8% of the cases and 66.7% of the controls, smear microscopy (+) was present in 37.5% of the cases and 52.1% of the controls, on the other hand, the culture result for 54.2% of the cases was (++) and 47.9% of the controls, (+). As for the admission condition, 58.3% of the cases were previously treated and 41.7% of the controls were never treated. Normal hematocrit was present in 70.8% of cases and 83.3% of controls.

Table 3: Distribution of internal risk factors associated with MDR-TB in DIRIS Lima Sur, 2017.

		Cases		Controls		Total		
		n1	%	n2	%	n	%	
Harmful Habits	Alcohol consumption	Yes	5	20.8	6	12.5	11	15.3
		No	19	79.2	42	87.5	61	84.7
	Drug Use	Yes	7	29.2	4	8.3	11	15.3
		No	17	70.8	44	91.7	61	84.7
	Tabaco use	Yes	4	16.7	4	8.3	8	11.1
		No	20	83.3	44	91.7	64	88.9
History of comorbidities	Positive HIV serology	Yes	2	8.3	1	2.1	3	4.2
		No	22	91.7	47	97.9	69	95.8
	Diagnosis of Diabetes Mellitus	Yes	5	20.8	4	8.3	9	12.5
		No	19	79.2	44	91.7	63	87.5
	Cortico prior therapy	Yes	2	8.3	0	0.0	2	2.8
		No	22	91.7	48	100.0	70	97.2

Source: DIRIS Lima Sur, 2017.

Table 3 shows the distribution of internal risk factors for the appearance of XDR-TB. Regarding harmful habits, 79.2%, 70.8% and 83.3% of the cases and 87.5%, 91.7% and 91.7% of the controls did not consume alcohol, drugs or tobacco, respectively.

The history of comorbidity such as a history of positive HIV serology, diagnosis of diabetes and use of corticotherapy were not present in the majority of cases or controls.

Table 4: Distribution of risk factors related to treatment history associated with MDR-TB in DIRIS Lima Sur, 2017.

Treatment History		Yes	Cases		Controls		Total	
			n ₁	%	n ₂	%	n	%
Previous TB/MDR-TB treatment	Yes	17	70.8	19	39.6	36	50.0	
	No	7	29.2	29	60.4	36	50.0	
Poor adherence to TB/MDR-TB treatment	Yes	9	37.5	10	20.8	19	26.4	
	No	15	62.5	38	79.2	53	73.6	
TB/MDR-TB treatment less than one year	Yes	11	45.8	6	12.5	17	23.6	
	No	13	54.2	42	87.5	55	76.4	
Failure to adhere to primary and individualized TB/TB-MDR-TB regimen	Yes	12	50.0	4	8.3	16	22.2	
	No	12	50.0	44	91.7	56	77.8	
TB/MDR-TB treatment dropout	Yes	4	16.7	7	14.6	11	15.3	
	No	20	83.3	41	85.4	61	84.7	
Early relapse in less than 6 months at discharge as cured of TB/MDR-TB	Yes	2	8.3	4	8.3	6	8.3	
	No	22	91.7	44	91.7	66	91.7	
Previous Fluoroquinolone treatment	Yes	1	4.2	0	0.0	1	1.4	
	No	23	95.8	48	100.0	71	98.6	
Self-administered home treatment for TB/TB-MDR.	Yes	2	8.3	0	.0	2	2.8	
	No	22	91.7	48	100.0	70	97.2	
Adverse reactions to antituberculosis treatment	Yes	1	4.2	1	2.1	2	2.8	
	No	23	95.8	47	97.9	70	97.2	

Source: DIRIS Lima Sur, 2017.

Table 4 shows that 70.8% of the cases were previously treated for TB/MDR-TB, 79.2% of the controls did not have poor adherence to TB/MDR-TB treatment. In 54.2% of the cases and 87.5% of the controls, no less than one year of previous treatment was required; 83.3% of the cases and 85.4% of the controls did not have a history of abandonment of MDR-TB/MDR-TB treatment;

no early relapse within 6 months of discharge as cured was present in 91.7% of cases and controls; self-administration of treatment was not present in 91.7% of cases and 100% of controls; and the presence of RAFAS was not present in the majority of cases and controls (>95%).

Table 5: Distribution of external risk factors associated with MDR-TB in DIRIS Lima Sur, 2017.

Contact history		Yes	Cases		Controls		Total	
			n ₁	%	n ₂	%	n	%
Family member at home with XDR-TB/MDR-TB/MDR-TB	Yes	14	58.3	27	56.3	41	56.9	
	No	10	41.7	21	43.8	31	43.1	
Had family members who died of XDR-TB/MDR-TB	Yes	10	41.7	7	14.6	17	23.6	
	No	14	58.3	41	85.4	55	76.4	
Have you had neighbors or friends in the neighborhood with tuberculosis?	Yes	11	45.8	25	52.1	36	50.0	
	No	13	54.2	23	47.9	36	50.0	
Have you had neighbors or friends in the neighborhood who died from TB?	Yes	5	20.8	2	4.2	7	9.7	
	No	19	79.2	46	95.8	65	90.3	
Prison history for the past two years.	Yes	5	20.8	4	8.3	9	12.5	
	No	19	79.2	44	91.7	63	87.5	
Previously worked in hospital or was a health science student	Yes	0	0.0	3	6.3	3	4.2	
	No	24	100.0	45	93.8	69	95.8	
Hospitalized in the last two years before becoming ill with TB, for more than 15 days.	Yes	1	4.2	1	2.1	2	2.8	
	No	23	95.8	47	97.9	70	97.2	

Source: DIRIS Lima Sur, 2017.

Table 5 presents the external factors associated with XDR-TB. Intradomestic contact with XDR-TB/MDR-TB was present in 58.3% of the cases and 56.3% of the controls; history of a deceased family member with XDR-TB/MDR-TB was not present in 58.3% of the cases and 85.4% of the controls. The majority of cases and controls did not have out-of-home contact with TB, were not former prisoners did not work in health facilities, were not students of health sciences, and had not been hospitalized in the last two years for TB.

3.2. Hypothesis testing

General hypothesis

H0: Risk factors are not significantly associated with extensively resistant tuberculosis in multidrug-resistant patients. in the Health Directorate Lima Sur, 2017.

H1: Risk factors are significantly associated with extensively resistant tuberculosis in multidrug-resistant patients. in Health Directorate Lima Sur, 2017.

The association of risk factors with the development of XDR-TB was contrasted from the sub-hypotheses, to which each factor separately was applied simple logistic regression (bivariate) and multivariate, for subsequent determination of the existence or non-existence of significant association; using the Odds Ratio (OR) with a confidence interval (CI) of 95% and considered as a statistically significant result at a level $\alpha=0.05$

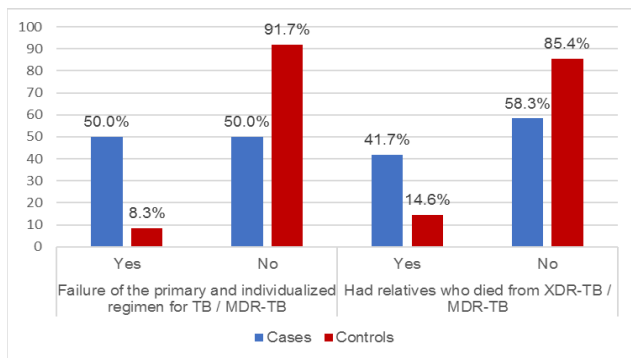


Figure 2: Factores de riesgo internos y externos asociados significativamente a la TB-XDR

According to the figure, 50% of the cases if they failed the primary and individualized scheme for TB / MDR-TB and 41.7% of the cases if they had relatives who died from XDR-TB / MDR-TB; in controls, 91.7% did not fail the primary and individualized regimen for MDR-TB / TB and 85.4% had no relatives who died from XDR-TB / MDR-TB.

Specific Hypothesis 1

H0: Internal risk factors are not significantly associated with the presence of extensively drug-resistant tuberculosis in multidrug-resistant patients.

H1: Internal risk factors are significantly associated with the presence of extensively drug-resistant tuberculosis in multidrug-resistant patients

Table 6 shows that the factors associated with XDR-TB in the DIRIS Lima Sur were: drug use, previous history of having used MDR-TB/TB treatment, having received MDR-TB treatment for less than one year and history of treatment failure for MDR-TB/TB, since their p-value is < 0.05, rejecting the H0 and concluding that these factors are significant. The multivariate analysis reaffirmed the influence of the variable: history of failure to primary and individualized treatment for MDR-TB/TB. In this sense: The odds of presenting XDR-TB in the group of cases with failure to the primary and individualized scheme for MDR-TB/TB is 6.81 times the odds of presenting XDR-TB. In the controls with primary and individualized TB/TB-MDR-TB scheme failure.

Table 6: Statistical test of internal risk factors associated with XDR-TB in DIRIS Lima Sur, 2017.

		p value	Bivariate		Multivariate	
			OR	IC95%	OR	IC95%
Harmful Habits	Alcohol consumption	0,354	1,84	0,49-6,79		
	Drug use	0,021	4.53	1,17-17,47		
	Tabaco use	0,289	2,2	0,49-9,69		
History of comorbidities	Positive serology for HIV	0,211	4,27	0,37-49,68		
	Diagnosis de Diabetes Mellitus	0,131	2,89	0,69-11,98		
	Corticosteroid therapy	0,108*	-	-		
Treatment History	Previous TB/MDR-TB treatment	0,012	3,71	1,29-10,63		
	Poor adherence to TB/MDR-TB treatment	0,130	2,28	0,77-6,72		
	TB/MDR-TB treatment less than one year	0,002	5,92	1,83-19,15		
	Failure to adhere to primary and individualized TB/TB-MDR-TB regimen	0,000	11	3,00-40,33	6,81	1,39-33,27
	TB/MDR-TB treatment dropout	0,817	1,17	0,31-4,47		
	Early relapse in less than 6 months at discharge as cured of TB/MDR-TB.	0,108*	1	0,17-5,89		
	Previous Fluoroquinolone treatment	0,333*	-	-		
Self-administered home treatment for TB/TB-MDR.	0,108*	-	-			
Adverse reactions to antituberculosis treatment	0,559*	2,04	0,12-34,16			

Source: DIRIS Lima Sur, 2017

Specific Hypothesis 2

H0: External risk factors are not significantly associated with the presence of extensively drug-resistant tuberculosis in multidrug-resistant patients.

H1: External risk factors are significantly associated with the presence of tuberculosis. extensively resistant in multidrug-resistant patients.

Table 6: Statistical test of external risk factors associated with XDR-TB in DIRIS Lima Sur, 2017.

	p-value	Bivariate		Multivariate	
		OR	IC95%	OR	IC95%
Contact history					
Family member at home with XDR-TB/MDR-TB	0,866	1,09	0,40-2,94		
Had family members who died of XDR-TB/MDR-TB	0,011	4,18	1,34-13,09	9,73	2,00-47,26
Have you had neighbors or friends in the neighborhood with tuberculosis?	0,617	0,78	0,29-2,08		
Have you had neighbors or friends in the neighborhood who died from TB?	0,024	6,05	1,08-33,96		
Prison history for the past two years.	0,131	2,89	0,69-11,98		
Previously worked in hospital or was a health science student	0,546*	-	-		
Hospitalized in the last two years before becoming ill with TB, for more than 15 days.	0,559*	2,04	0,12-34,16		

Source: DIRIS Lima Sur, 2017.

Table 7 shows that the external factors associated with XDR-TB in the DIRIS Lima Sur were: history of a family member who died of XDR-TB/MDR-TB and having friends or neighbors in the neighborhood who died of TB, due to their p-value < 0.05, rejecting the H0 and concluding that these factors are significant. The multivariate analysis reaffirmed the influence of the variable: history of a family member who died from XDR-TB/MDR-TB. In this sense: The odds of presenting XDR-TB in the group of cases who had relatives who had died of XDR-TB/MDR-TB is 9.73 times the odds of presenting XDR-TB in the controls who had relatives who had died of XDR-TB/MDR-TB.

4. Discussion

When analyzing our results, we found that according to sociodemographic factors of the 72 patients with XDR-TB and MDR-TB in the DIRIS Lima Sur; taking into account the majority

of cases, 70.8% are between 26 to 60 years old, 58.3% are male, 62.5% have secondary school education and 58.3% neither work nor study. As for the majority of controls, 52.1% were between 26 and 60 years of age, 52.1% were women, 45.8% had secondary education and 45.8% did not work or study. The results are similar in some respects to those of other studies in which the variables age, age, sex, sex, and age related variables were crossed. [21] Where the variables age, sex and occupation were crossed. Likewise with the study of [22] where the variables age, sex and occupation were considered as predictors of the appearance of XDR-TB and MDR-TB. It is also worth mentioning the study of [23] where they show age greater than or equal to 40 years as a significant protective factor. Another study by [24] Another study of the study of the study of the age of the cases shows an average age of 37 years for the case group and 38 years for the controls, as well as 56.2% and 66.7% for the male gender for cases and controls respectively.

The socioeconomic level of a person has an impact on morbidity and mortality in vulnerable population groups and therefore in their country. Studies have shown that a lower socioeconomic status has an impact on the mother's education and that a poor educational level is associated with increasing morbidity rates [25].

One of the great changes in the epidemiology of TB worldwide are the resistant forms of this disease, from mono-resistant TB to XDR-TB, constituting from a great challenge for public health to an incurable disease with very high lethality, especially if it is associated with HIV/AIDS.

The existence of factors such as those found in our study, for the group of cases with a history of a family member who died of XDR-TB, is 9.73 times more likely to develop XDR-TB and the history of having received previous anti-tuberculosis treatment is 6.81 times more likely to develop XDR-TB than in the control group. These epidemiological factors predispose morbidity and mortality in the economically active population.

The internal factors significantly associated with the development of XDR-TB were drug use, previous history of having received MDR-TB/TB treatment, having received previous MDR-TB treatment for less than one year, and multivariate analysis confirmed the significant influence of the variable failure to primary and individualized MDR-TB/TB treatment. The study of [20] study highlights that the failure of an antituberculosis treatment scheme consisting of drugs such as fluoroquinolone and a second-line injectable has a strong association with XDR-TB. Another study by [22] concludes that one of the factors associated with multidrug resistant TB is a history of previous TB treatment. Also in the research [26] also stated that problems or failure to adhere to treatment is more strongly associated with XDR-TB as well as inadequate screening, among others. Another study [21] highlights a history of anti-TB treatment as a risk factor for the development of MDR TB and XDR TB, and that failure to adhere to treatment regimens is strongly associated with XDR TB [9] is strongly associated with XDR-TB.

Regarding harmful habits, patients with a history of drug use have a higher risk of developing XDR-TB, results that are related to the study of [24] where he reaffirms drug addiction along with deterioration in adherence to treatment as risk factors for the

presence of MDR-TB. The success of a TB control program and the health of the affected person will depend on the degree of adherence to the therapeutic regimen, which is one of the most important aspects in the management of resistant TB, for which the interdisciplinary team of health institutions should formulate patient-centered and oriented strategies in order to seek adherence to supervised treatment and avoid irregularities, in partnership and involvement of grassroots organizations, social actors in the community, and community health workers [17].

The external factors with significant association to the development of XDR-TB found were extradomiciliary contact with someone who died of TB; and through multivariate analysis, the influence of the variable: family member who died of XDR-TB/MDR-TB was reaffirmed, results that are coherent with the study of [23] study in relation to the occurrence of primary MDR-TB in patients from Callao, ratifying as a risk factor having had contact with someone with MDR/XDR-TB and as a protective factor being younger than 40 years old. Other studies, such as [20] highlights as factors with a strong association to the development of XDR-TB the failure of a treatment scheme with second-line drugs and having a close relationship with someone affected by documented XDR-TB. [21] also mentions contact with MDR TB as a risk factor, and in the case-control study, the study also [27] also considers previous contact with MDRTB cases in both groups as a risk factor.

The contact is the person who establishes a close relationship with the affected person; therefore, an extra-domiciliary contact is comprised of partners, friends, neighbors and family members who have a regular relationship with the person with TB and who frequently visit the home. The primary practices for environmental control of TB and its resistant forms such as XDR is to ensure and maximize natural and cross ventilation in the environment, free of obstacles in the entrances, to ensure an exchange of air flow with a minimum of twelve air changes per hour and to ensure natural lighting, these strategies comprise primary prevention [17].

The next level of prevention is secondary, through early detection and timely treatment will ensure the cure and reduction of serious consequences of the disease, so in the case of TB, this is not aimed at reducing the incident cases but to reduce its severity (duration and complications) and lethality. In [28], Likewise, tertiary prevention aimed at rehabilitation seeks to reduce the complications of a disease through therapy and rehabilitative medicine, guaranteeing the reduction of sequelae, disabilities, and suffering, and promoting the adaptation of affected persons and their families to their environment. All these levels of prevention also include good quality medical care, which together will allow the individual to reintegrate into productive activities without being re-exposed to the same risk factors that contributed to getting sick [28].

Our findings showed that, despite the existence of a health strategy for the prevention and control of tuberculosis with defined objectives and treatment schemes approved by technical standards, there are risk factors that are associated with the emergence of extensively resistant TB, which demonstrates deficiencies in the approach to people affected by tuberculosis, which goes beyond an exclusively health approach, but also demographic and social. This situation does not contribute to an adequate isolation of the index

case, generating chains of multiplication and transmission of TB and MDR/XDR TB bacilli in the family nucleus and community. In addition, overcrowded public transportation puts at risk the other passengers who share the same reduced and unventilated space, under the same mechanism.

There are not enough human resources for the comprehensive care of TB patients, and the health personnel at the first level of care are responsible for the approach and management of TB, MDR-TB and XDR-TB cases, making early diagnosis and timely treatment of this complex disease difficult. These patients with MDR-TB and XDR-TB receive treatment divided into two doses, both in the morning and in the afternoon; however, most first-level health care institutions only provide care in the mornings, a situation that contributes to the large-scale increase in the morbidity of tuberculosis in the country.

5. Conclusions

There are risk factors that are significantly associated with the development of extensively resistant tuberculosis in multidrug-resistant patients in the Directorate of Integrated Health Networks of Lima Sur, demonstrating that these factors were determinant for its appearance.

Multivariate analysis showed that the association of the internal risk factor: having failed the primary and individualized regimen for MDR-TB/TB is 6.81 times higher risk of developing XDR-TB/MDRTB than in those who did not fail.

The external risk factor: having a deceased family member with XDR-TB/MDR-TB is 9.73 times more likely to develop XDR-TB/MDR-TB than those without a deceased family history; this association was confirmed by multivariate analysis.

5.1. Recommendations

It is suggested to implement and train health professionals in the approach and comprehensive care of the person affected by tuberculosis, as well as in the sustainable work of adherence to supervised treatment.

Implement the family and community health program to reduce the epidemiological profile of tuberculosis as a key strategy in the Directorate of Integrated Health Networks in southern Lima. Also, promote research to develop new diagnostic and susceptibility panel tests and prioritize research with an epidemiological approach

References

- [1] M. Raviglione, *Tuberculosis: The Essentials*, Fourth Edition. Boca Raton: CRC Press 2016.
- [2] WHO. *Implementation of the end of TB strategy*. Geneva: WHO 2016.
- [3] OPS. *Tuberculosis in the Region of the Americas*. Retrieved February 28, 2015.
- [4] OPS. *Tuberculosis*. Retrieved from <https://bit.ly/2rf0x1G>.
- [5] MINSA. *Tuberculosis intervention and control plan 2018 - 2020*. Lima. Retrieved on May 10, 2018.
- [6] J. Garcia, *Tuberculosis and risk factors*. *Emerging diseases*, 153, 2016.
- [7] K. Colimon, *Foundations of Epidemiology*. Madrid: Diaz de Santos S.A 2008.
- [8] WHO. *World report on tuberculosis*. Retrieved June 12, 2014, from <http://www.who.int/mediacentre/factsheets/fs104/es/.N>. Cao, M. Štěpnička, M. Burda, "Fuzzy Quantifiers and Compositions of Partial Fuzzy Relations Employing Dragonfly Algebras," in 2019 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), 1–6, 2019, doi:10.1109/FUZZ-

IEEE.2019.8858832.

[9] Y. Llamas, Extensively antibiotic-resistant tuberculosis (XDR-TB): therapies used successfully clinically to cure the disease. *Journal of Clinical Research*, **65**(3), 255-262, 2013. DOI: 10.1016/j.ijantimicag.2009.12.019

[10] J. Robledo, Control of multidrug-resistant tuberculosis: a possible goal. *Rev Biomedica*, **39**(3), 431-433, 2019. DOI: 10.2165/11595340-000000000-00000

[11] M. Contrera, Drug-resistant tuberculosis: diagnosis and treatment. *National Guide for e; tuberculosis management*, 147, 2017.

[12] M. Mamani et al., Characterization of risk factors associated with tuberculosis. *ICMHI 2020: Proceedings of the 4th International Conference on Health and Medical Informatics (230-234)*. Ibaraki-Japan: Association for Computing Machinery, 2014

[13] R. Bonita, *Basic Epidemiology*. Washington, D.C: PAHO, 2017.

[14] R. Hernandez et al., *Research methodology (Sixth ed.)*. Mexico: McGraw Hill, 2018.

[15] CDC. Risk factors for tuberculosis. Obtained from the Center for the Control and Prevention of Nurses: <https://www.cdc.gov/tb/esp/topic/basics/risk.htm>.

[16] C. Daley, *An Issue of Clinics in Chest Medicine*. Pennsylvania: Elsevier Health Sciences, 2019.

[17] MINSA. *Technical Health Standard for the comprehensive care of people affected by tuberculosis*. Lima: MINSA, 2013.

[18] J. Kelsey, *Methods in Observational Epidemiology*. New York: Oxford University Press, 1986.

[19] J. Fleiss, *J. Statistical Methods for Relationships and Proportions*. New York: John Wiley and sons, 1981.

[20] M. Ticona, Extremely resistant tuberculosis (XDR-TB), history and current situation. *Acta Med Per*, 4, 25, 2008.

[21] Y. Libón, Risk factors for the development of multidrug-resistant tuberculosis in patients of the Hospital Nacional "Dos de Mayo. Ricardo Palma University, Lima, Lima, 2017.

[22] D. Asenjo, Factors associated with multidrug resistance in. patients diagnosed with tuberculosis Guillermo Almenara Irigoyen National Hospital. San Martin de Porres University, Lima, 2016.

[23] A. Avalos, Factors associated with primary multidrug-resistant tuberculosis in. *An Fac med*, 75, 233-246, 2016.

[24] Rodriguez. Risk Factors for Multidrug-resistant Pulmonary Tuberculosis in the La Libertad Region, Peru, 2013.

[25] J. Gofin, *Global Community Health, principles and programs in the world*. Madrid, 2012.

[26] Y. Gonzales, Extensively antibiotic-resistant tuberculosis (XDR-TB): therapies used successfully clinically to cure the disease. *Clinical Research Rev.*, **60**(3), 2011. DOI: 10.1016/j.ijantimicag.2009.12.019

[27] T. Chang et al., Risk factors in the appearance of multidrug resistance in patients with pulmonary tuberculosis. *Rev medic panacea*, **4**(2), 31-35, 2014.

[28] OPS. *Module of principles of epidemiology for the control of diseases (MOPECE) (Second ed.)*. Washington D.C.: OPC, 2011.

Annexes

QUESTIONNAIRE TO DETECT RISK FACTORS ASSOCIATED WITH THE PRESENCE OF TB-XDR IN A FIRST LEVEL HEALTH CARE INSTITUTION - DIRIS LIMA SUR, 2017

I. Socio-demographic characteristics (Intervening variable)

Cases	Controls
-------	----------

A. Age (in years)

1) 10 – 19	1
2) 20 – 25	2
3) 26 – 60	3
4) 61 plus	4

B. Genero

1) Male	1
2) Female	2

C. Level of education

1) Primary	1
2) Secondary	2

3) Technical superior	3
4) Superior Univ.	4

D. Occupation

1) Not working, not studying	1
2) Works	2

3) Study	3
4) Other	4

E. BMI category.

a) Weight _____

b) Size _____

c) BMI result _____

1) Low weight <18.5	1
---------------------	---

2) Normal [18.5 – 25)	2
3) Overweight ≥ 25	3

F. Results of smear microscopy

a) BK (+)	1	b) BK (++)	2	c) BK (+++)	3
-----------	---	------------	---	-------------	---

G. Cultivation result

a) Cultivation (+)	1	b) Cultivation (++)	2	c) Cultivation (+++)	3
--------------------	---	---------------------	---	----------------------	---

H. CONDITION OF ENTRY:

<input type="checkbox"/> TREATED	<input type="checkbox"/> a) BEFORE TREATED	<input type="checkbox"/> b) NEVER
-------------------------------------	---	--------------------------------------

I. Result of number of drugs with resistance IN VITRIO (PS)

1) Resistance (R)	R	H	Km	Am	Cm	Lfx	Mfx	Eto	Cs	PAS	Cfz	Amx/Clv	Cpx
2) Sensitive (S) (Mark with X)													

J. Haematocrit result

a) Result	_____
b) Men: Normal: <41% - 53%>	1
Low: < 41%	2
c) Women: Normal: <36% - 46%>	1
Low: < 36%	2

II. Risk factor associated with TB- XDR-TB-MDR cases and controls

2.1.	Internal case factors and controls		
	A. Harmful Habits	Yes(1)	No (0)
1	Consumed alcohol more than 3 times per week		
2	Used drugs in the past year		
3	Consumed more than 5 cigarettes per week		
	B. History of comorbidities		
4	Showed positive serology for HIV		
5	Diabetes Mellitus		
6	Prior use of corticosteroid therapy		
	C. History of Tuberculosis Treatment		
7	You were previously treated for tuberculosis / MDR-TB		
8	Showed poor adherence to TB/MDR-TB treatment		
9	You were on TB/ MDR-TB treatment for less than 1 year		
10	You failed the primary, individualized TB/TB-MDR-TB regimen		
11	You have previously stopped treatment for TB/MDR-TB.		
12	Had early relapse within 6 months of being discharged as cured of TB/MDR-TB		
13	Previous use of Fluoroquinolones		
14	You received self-administered home treatment for TB- /TB- MDR.		
15	Had adverse reactions to anti-tuberculous treatment (RAFA).		
2.2.	External case-control factors		
	D. History of intra- and extra-home contact	Yes(1)	No (0)
16	Had a family member at home with TB -XDR/TB - MDR		
17	Had family members who died of XDR-TB/MDR-TB		
18	Have you had neighbors or friends in the neighborhood with tuberculosis?		
19	Have you had neighbors or friends in the neighborhood who died from TB?		
20	History of incarceration in a correctional facility within the last two years		
21	Previously worked in a hospital or was a health science student		
22	He was hospitalized in the last two years before getting sick with TB, for more than 15 days.		

Study and Implementation of LEDs Drivers with Dimming Capability

Brahim Mrabet*, Abdeljelil Chamman

Electrical Engineering Department, Prince Sattam bin Abdulaziz, Alkharj, 11942, Saudi Arabia

ARTICLE INFO

Article history:

Received: 26 August, 2021

Accepted: 29 November, 2021

Online: 17 December, 2021

Keywords:

LED

LED driver

DC-DC buck converter

DC-DC boost converter

Constant Voltage driver

Constant Current driver

LED dimming

ABSTRACT

Nowadays, LED lights take an important place in our daily lives and they have known a great growth in indoor as well as in outdoor lighting applications. LED (Light Emitting Diode) light sources including their own drivers have excluded many systems fitted with both inefficient light sources. In this paper we present, a LED driver handling a DC input power supply and piloting a series of powered LEDs. Some topologies of LED driver supplying a constant voltage or a constant current toward a LED load and their operations are detailed and presented. In this regard, two types of DC-DC converter in occurrence buck and boost converter are analyzed, designed, and simulated. Furthermore, a design of laboratory prototype of constant-current LED drivers based on DC-DC buck converter totally dimmed is achieved, and tested. Calculated and experimental results are in good correspondence with very small deviation of LED current in both cases for different levels of dimming.

1. Introduction

Light Emitting Diodes (LEDs) usage has known an extraordinary growth in indoor and outdoor lighting applications. Consequently, nowadays LED lights *take an important place* in our daily lives and have become a viable alternative to conventional low efficiency light sources like incandescent, CFL (compact fluorescent lamp) and HID (high intensity discharge) lamps. LED lighting systems have better characteristics than conventional lighting systems such as energy saving, long life spans and color rendering [1]-[8] and they play an important role in future indoor as well as outdoor lighting applications. To operate properly with respect of energy saving, high efficiency and minimum electric constraints, these light sources demand a specific driver. These LED drivers play an important role on the type of lighting system performances, such as dimming capability, power quality, Light quality and product lifetime.

The overall lighting system efficiency is affected by the choice of driver's topology and the associated electronic components. For outdoor applications some other constraints must be added to those known before such as humidity and ambient temperature. In indoor applications, the drivers are integrated in the luminaries in order to make them compact and facilitate their integration in any decoration [9].

Two principal electronic circuits are used to drive LEDs under specific electrical conditions: constant current (CC) or constant voltage (CV) [10]-[15]. Such drivers, should also satisfy intensity and color requirements of LEDs.

In the second section of this paper, after a short presentation on the fundamentals and characteristics of LEDs and their driver circuits, we present the operating principles and the design of DC-DC buck converter and boost converter used to drive LEDs with current and voltage regulation. by varying voltage source respectively. The DC-DC converter can supply LEDs up to 1A current and insure full dimming of rated current. The third section of this paper, describes the design and discussion of these converters and their control. Finally, in sections 4 and 5, some simulations and experimental results are given for a DC-DC buck converter four LED light-spot connected in series and mounted on a board drawing 300mA rated current.

2. LEDs and LEDs drivers

2.1. LEDs

LEDs are made of doped gallium nitride (GaN) semiconductor material. Their electronic structure is a diodes or tandem structures, [16]. Power LEDs are by their superior life time, cost-effective, and high luminous intensity compared to classical incandescent lamp and low pressure discharge lamp. Nevertheless, they are known by low environment impact and they are used

*Corresponding Author: Brahim Mrabet, Prince Sattam bin Abdulaziz University, Alkharj, +966 115 88 8253, b.mrabet@psau.edu.sa

solid-state lighting source in many lighting applications. Power LEDs light sources are presented in figure 1, they require 2 to 4V of dc forward voltage and several hundred mA of dc forward current for their optimal operating point [17].



Figure 1: Samples of power LEDs

A powerful light spot, is made by combining small LED array devices according different matrix architectures. In addition, it's required to take account of some LEDs characteristics when using them, such as:

- Their electric characteristics, light output and the color of the emitted light are sensitive to the temperature.
- The forward voltage of high-Power LED changes with temperature which will cause drawing more current. The relation between VF (Forward Voltage) and temperature is reverse as shown in Figure 2. When the temperature increases, forward voltage decreases, this is known as Thermal Runaway, which causes a semiconductor failure. [16].

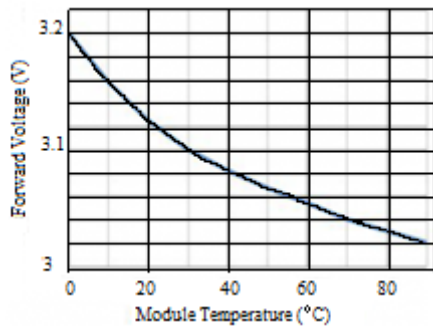


Figure 2: Forward voltage versus Temperature for LED

Stable temperature and dc current as much as possible increases the reliability and guaranties the longest life of LEDs.

3. LEDs Drivers

LEDs need equipment called a “driver” for their supply and control. Many topologies of LED drivers are deployed and many special ICs are available and used to support these different topologies. Two kinds of drivers are commonly employed to drive LEDs:

- Constant-Voltage regulated drivers able to maintain a DC output voltage. regardless variation of input voltage or output current within specified limits-
- Constant-Current regulated drivers able to maintain a DC output current regardless variation of input voltage or output voltage within specified limits [18].

Converter topologies such as buck, boost, or buck-boost converters are widely employed for LED drivers [19]-[23].

3.1. DC-DC converter

Llinear DC-DC converters and switched-mode DC-DC converters are among the converters used for DC-DC conversion function. They can deliver different dc voltage level from dc voltage source and they often provide a regulated output.

- *Linear DC-DC converter*

Figure 3 shows a linear DC-DC converter.

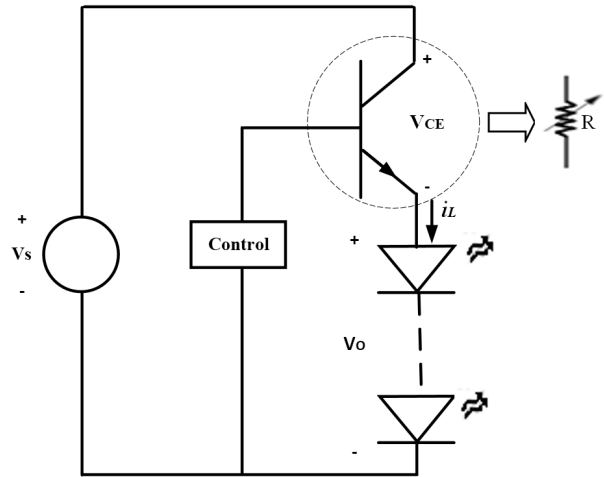


Figure 3: A basic linear regulator

The load current i_L is controlled by adjusting the transistor dc biasing, the output voltage may be controlled over a range of 0 to roughly V_s . The transistor in effect operates as a variable resistance. The low efficiency of this circuit is a serious drawback for power applications caused by BJT(Bipolar Junction Transistor) losses.

- *Switched-mode DC-DC converters*

In a switching mode, the MOSFET operates as an electronic switch by being totally ON or totally OFF [24]. In the chopper circuit of figure 4 (a), the output is the same as the input when the switch is closed during DT , and it's zero when the switch is open. Figure 4 (b) shows the output voltage across the load.

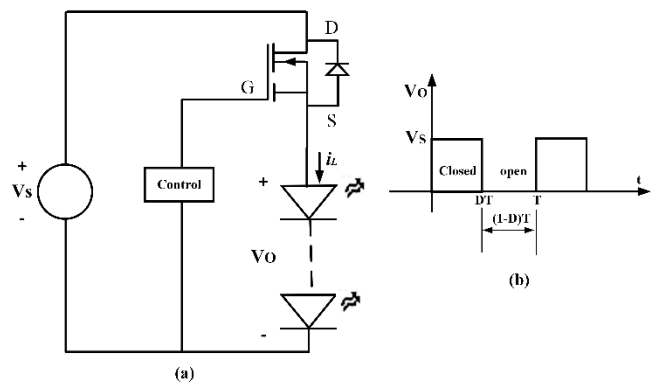


Figure 4: (a) Switching converter (b) Output voltage

The output average voltage is: $V_o = DV_s$ and thus it's less than the input voltage and its dc component can be controlled by adjusting D. Such converter is nominated buck converter. For boost converter, the voltage V_o is always higher than the supply

voltage V_s . The switch mode solution is more efficient than the linear solution. As shown in Figure 5 and in Figure 6, a filter stage should be inserted in the converter to obtain pure continuous output voltage instead of square wave output [25]-[28]. In Figure 5, during the period where the MOSFET is on and the diode D1 is off, the energy is stored in passive components L and C_{out} . Conversely, when the MOSFET is off and the diode D1 is on, the power supply V_s is separated from the LEDs load which will be supplied only from the energy stored in inductor L and in the capacitor C_{out} during the previous period.

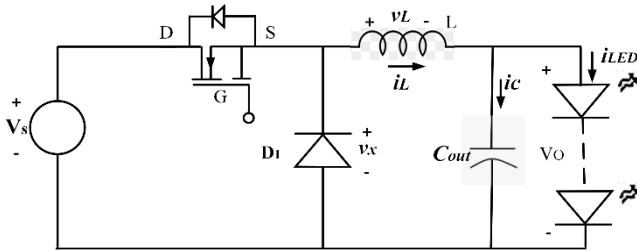


Figure 5: Buck converter

For boost converter in the circuit of Figure 6, during the first interval when the transistor and the diode are switched on and off respectively, the voltage source is supplying the inductor L and the capacitor is providing power to the load formed by a series connection of LEDs. During the second interval when the transistor and the diode are switched off and on respectively, energy is stored in L and in C_{out} , and provide power to LEDs.

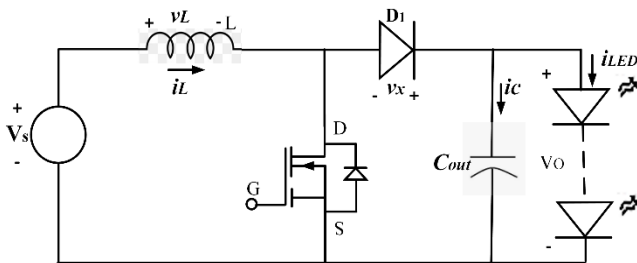


Figure 6: Boost converter

3.2. LED Driver based on Buck converter

A block scheme of a constant-current regulated driver for LED is shown in Figure 7. The output current is set out to a specific value-by inserting a resistive current sensor, R_s , in the load branch.

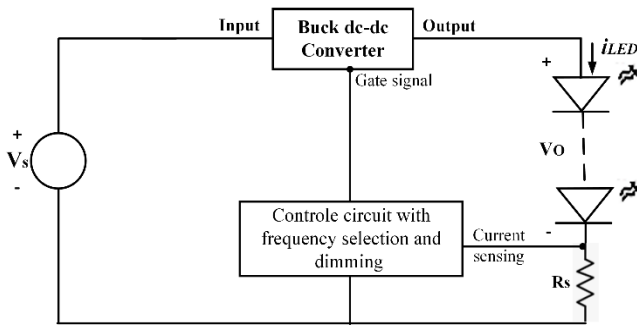


Figure 7: Current controlled LED driver circuit

A block scheme of a constant-voltage boost converter used to drive LED is shown in Figure 8. The output voltage is programmed

by placing an appropriate voltage divider, R_1 - R_2 , in parallel with the LEDs.

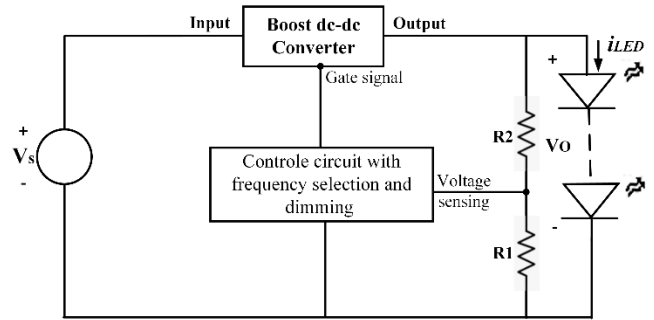


Figure 8: Voltage controlled LED driver circuit

The choice of the switching frequency is a compromise between performance and external component size.

3.3. Constant current versus constant voltage LED driver circuit

LEDs are always driven by a constant current driver. Indeed, based on the LED V-I characteristic in terms of temperature, the forward voltage of LED is decreased when the temperature is increased and vice versa. For a long time, operation of LED under constant voltage, the temperature will raise and consequently the current will raise too which in its turn will cause new increase of temperature. As a result, if the heat created in the LED is not released in an appropriate time, the current will continue raising until the failure of LED. However, in case of constant current source the raise of heat in the LED only results in the drop of forward voltage across it. For applications where the lighting system does specify the voltage, constant voltage driver is used.

4. The design of buck and boost converter

As for any type of converter or driver system, the design aims to satisfy the desired application needs including the load requirements and to determine the system requirements and its parameters.

4.1. Load requirements

The load requirements define the type of light emitting diodes to be used, the rated current, the rated forward voltage, the rated power, the number of LEDs, and the type of connection.

4.2. System requirements and components values

Based on the load requirements, the system requirements is determined by selecting the inductor, the MOSFET, the diode, and the capacitors. The output voltage ripple is reduced by inserting an output ceramic capacitor. Table 1 shows the equations used for calculating the different components [24].

With f : frequency, D : duty cycle, V_s : voltage source, V_o : output voltage, V_{DS} : Drain Source voltage, V_r : reverse voltage, I_f : forward current, I_o : output current.

The inductor is selected with a 25% larger current the calculated one in order to avoid its saturation. In addition, the optimal ripple current $I_{L-ripple}$ is 40% of the output current. The input ceramic capacitor is calculated using the following equation

$$C_{IN} = \frac{I_o \times D(1 - D)}{f \times \Delta V_{pp}} = \frac{I_o \times L \times I_{L-ripple}}{V_s \times \Delta V_{pp}}$$

The requirements for this design are: $V_S=16V$, four 300mA high power LEDs, $f=500\text{ kHz}$, $V_{O\text{-ripple}}=3\%$. Based on this, the design parameters required for the converter are:

$$L=10\mu\text{H}, C_{\text{IN}}=2.2\mu\text{F}, C_{\text{out}}=2.2\mu\text{F}, \text{Diode: 1N5819}$$

Table 1: Component selection

	DC-DC buck converter	DC-DC boost converter
Inductor L	$V_S \frac{(1-D) \times D}{I_{L\text{-ripple}} \times f}$	$\frac{D \times V_S}{I_{L\text{-ripple}} \times f}$
Capacitor C_{out}	$\frac{(1-D)}{8 \times L \times \frac{V_{O\text{-ripple}}}{V_O} \times f^2}$	$\frac{D \times V_O}{V_{O\text{-ripple}} \times f \times I_O}$
Diode D1	$V_r > V_{IN}$ $I_F > I_O(1-D)$	$V_r > V_{out}$ $I_F > I_O(1-D)$
MOSFET	$V_{DS} > V_{IN}$ $I_D > I_O(1-D)$	$V_{DS} > V_{out}$ $I_D > I_O \frac{D}{1-D}$

5. Simulation and experimental results

5.1. Simulation results

The buck dc-dc- driver circuit of Figure7 is simulated using LTspice. The output voltage V_{LED} , output current I_{LED} , and inductor current I_L are shown in Figure. 9, Figure. 10, and Figure. 11. As seen in these figures, the LED current is controlled and it remains constant at 300 mA regardless the variation of the output voltage across the load formed by a series connection of LEDs. By observing the waveform of inductor current it's clear that its duty cycle is varying according the output voltage variation and its form is a saw-tooth signal as expected.

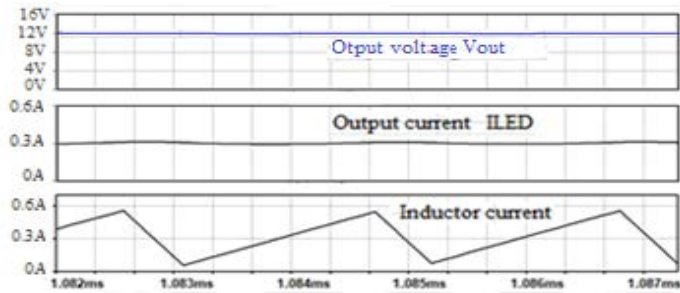


Figure 9: For 4 LEDs: waveforms of V_{LED} , I_{LED} , and I_L , $f = 500\text{ kHz}$, and $I_{LED}=300\text{ mA}$

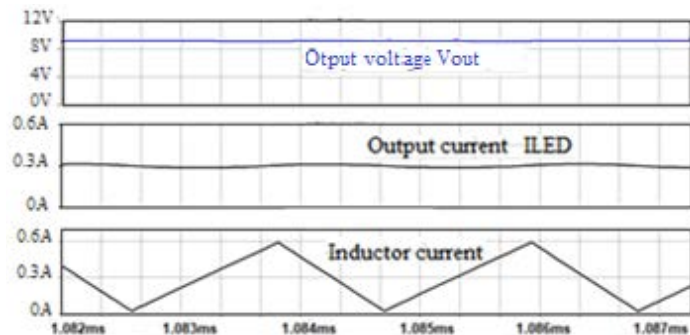


Figure 10: For 3 LEDs: waveforms of V_{LED} , I_{LED} , and I_L , $f = 500\text{ kHz}$, and $I_{LED} = 300\text{ mA}$.

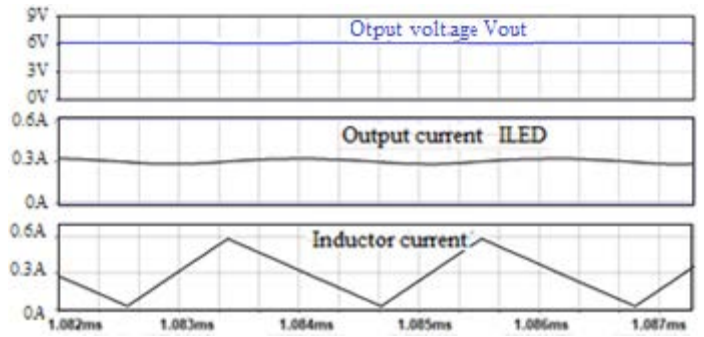


Figure 11: For 2 LEDs: waveforms of V_{LED} , I_{LED} , and I_L , $f = 500\text{ kHz}$, and $I_{LED}=300\text{ mA}$.

Table 2 gives numerical values of duty cycle as a response to the change of output voltage for a constant output current as shown in Figures 9 through 11.

Table 2: Duty cycle versus output voltage change for buck converter

Output voltage (V)	Output current (mA)	Duty cycle
12.6	300	77.6%
9.4	300	57.56%
6.2	300	39.5%

Regarding dimming capability of the converter, a low frequency pulse width modulation (PWM) signal is used to control the amount of the output current which is closely related to the amount of the emitted light. Simulations conducted for various duty cycles of PWM signal show that the dimming is obtained as illustrated in Figure. 12, Figure. 13, and Figure. 14.

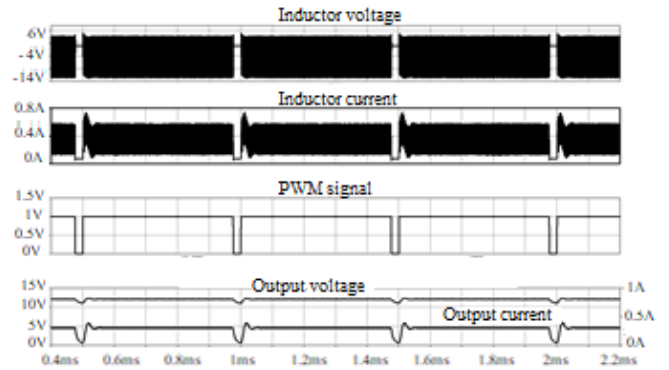


Figure.12: Waveforms of V_L , I_L , PWM signal, V_{LED} , I_{LED} for dimming level of 95% and $I_{LED\text{-rms}}=287.2\text{mA}$.

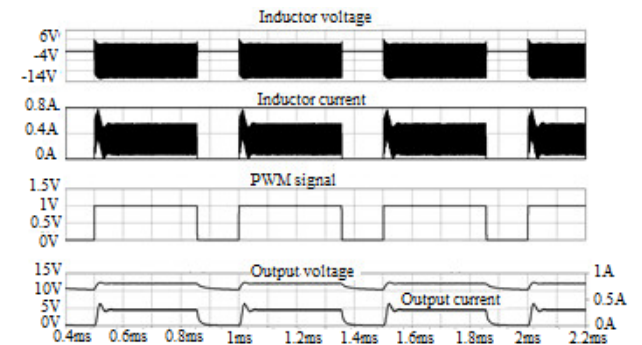


Figure 13: Waveforms of V_L , I_L , PWM signal, V_{LED} , I_{LED} for dimming level of 70% and $I_{LED, \text{rms}} = 248.22\text{mA}$

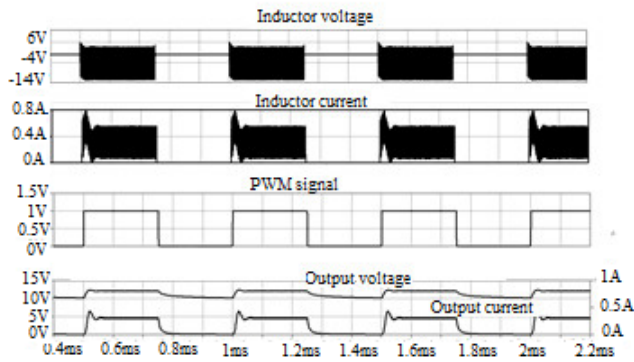


Figure 14: Waveforms of V_L , I_L , PWM signal, V_{LED} , I_{LED} for dimming level of 50% and $I_{LED, rms} = 206.69\text{mA}$

Simulation results for the boost dc-dc- driver circuit of Figure 8 show that the voltage across LEDs (V_{LED}) is regulated at 12.8 V regardless the variation of the load current or the input voltage. Figures 15 and 16 depict the output voltage when the load current is changed by connecting in parallel other LEDs. It's clear that the voltage regulation is achieved by adjusting the duty cycle as a response to the feedback signal issued from the voltage divider connected in parallel with the load.

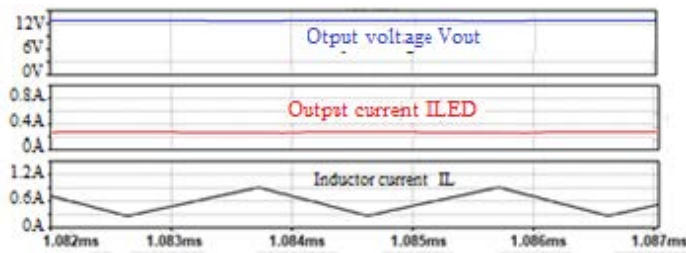


Figure 15: For 4 LEDs: output voltage V_{LED} , output current I_{LED} , and inductor current I_L , $f = 500\text{ kHz}$, and $V_{LED} = 12.8\text{V}$

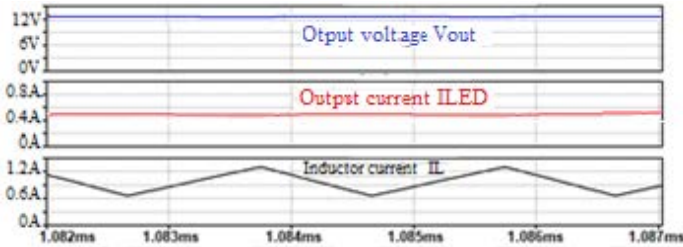


Figure 16: For 8 LEDs (two sets of 4 LEDs connected in parallel): output voltage V_{LED} , output current I_{LED} , and inductor current I_L , $f = 500\text{ kHz}$, and $V_{LED} = 12.8\text{V}$

Waveforms displayed in Figure 17, Figure 18, and in Figure 19 show that voltage V_{LED} is also maintained constant when the 6V input voltage is increased or decreased by 1 V in DC-DC boost converter.

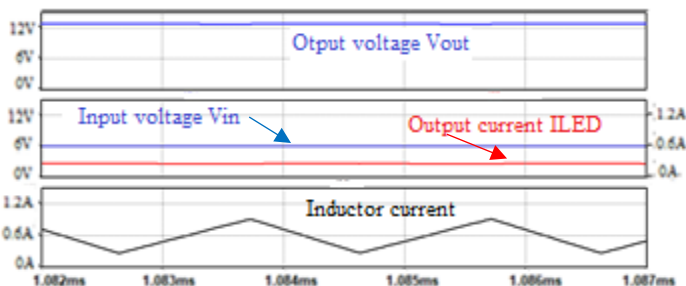


Figure 17: Waveforms of V_{out} , I_{LED} , and I_L for $V_{in} = 6\text{V}$, $f = 500\text{ kHz}$, and $V_{LED} = 12.8\text{V}$.

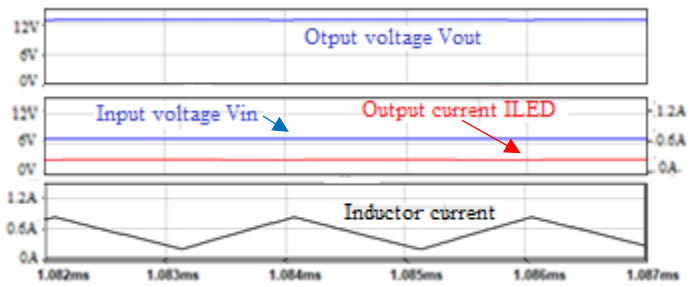


Figure 18: Waveforms of V_{out} , I_{LED} , and I_L for $V_{in} = 7\text{V}$, $f = 500\text{ kHz}$, and $V_{LED} = 12.8\text{V}$.

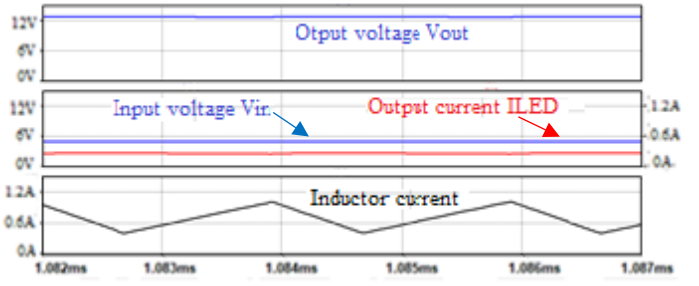


Figure 19: Waveforms of V_{out} , I_{LED} , and I_L for $V_{in} = 5\text{V}$, $f = 500\text{ kHz}$, and $V_{LED} = 12.8\text{V}$.

Table 3 gives numerical values of duty cycle as a response to the change of input voltage for a constant output voltage as shown in Figures 17 through 19.

Table 3: Duty cycle versus input voltage change for boost converter

Input voltage (V)	Output voltage (mA)	Duty cycle
5	12.8	60.1%
6	12.8	54.4%
7	12.8	43%

5.2. Experimental results

In previous sections, analysis and design have been performed for buck and boost converter. In this section, experimental results for a prototype of constant-current LED buck converter are presented. The load is formed by four LED light-spot connected in series and mounted on a board as shown in Figure.20 and Figure 21. The reasons for focusing on constant current DC-DC buck converter in experimental study are to ensure that the current doesn't exceed the rated current, to obtain consistent brightness and color for each LED. In addition, at equal load, the same LC filter has to work much harder in a boost mode compared to buck mode. In the boost case, the capacitor is switching high current and the inductor is carrying high average current. This means that the passive components in the boost converter case must be oversized compared to the buck converter, leading to a higher cost and size. And also means higher ohmic losses and lower efficiency due to the capacitor ESR and inductor ESL. The constant-current LED driver prototype has been built and four LED light-spot have been connected as shown in Figure.20 and Figure 21.

The dimming for LEDs is obtained by application of low frequency pulse width modulation signal with the possibility of duty cycle variation. Various oscilloscope captions are recorded. Figure. 22 depicts the waveform of the output voltage across LEDs when there is no dimming signal.

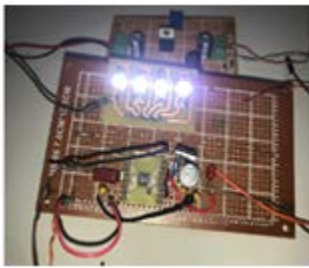


Figure 20: Driver for 4 LED light-spot, $f = 500$ kHz, and $I_{LED} = 300$ mA.



Figure 21: Implementation of the prototype for easy measurements.



Figure 22: Waveform of V_{LED} for for full duty cycle and $I_{LED} = 300$ mA

Figure 23 shows the voltage across LEDs and the control signal for pulse width modulation duty cycle set at 0.95

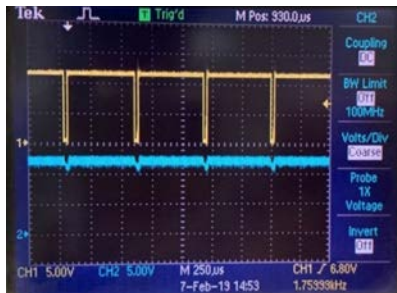


Figure 23: V_{LED} (Bottom) and pulse width modulation signal (Top) with Dcyle = 0.95 and rms current $I_{LED} = 284$ mA

Figure 24, shows the voltage cross LEDs and the control signal for duty cycle equal 0.7

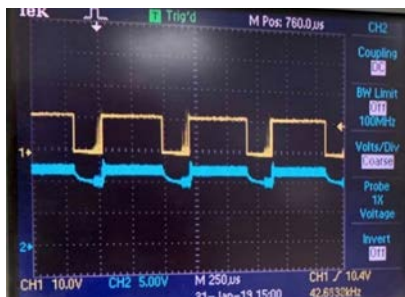


Figure 24: V_{LED} (Bottom) and pulse width modulation signal (Top) with Dcyle = 0.7 and rms current $I_{LED} = 246$ mA

Figure 25, shows the voltage cross LEDs and the control signal with pulse width modulation duty cycle set at 0.5.

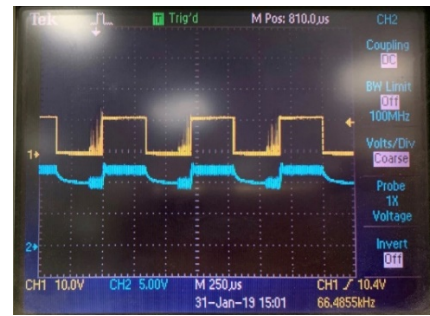


Figure 25: V_{LED} (Bottom) and pulse width modulation signal (Top) with Dcyle = 0.5 and rms current $I_{LED} = 203$ mA

Simulation results depicted in Figure. 12, Figure.13, and Figure.14 for 95%, 70%, and 50% dimming levels respectively show good agreement with experimental results conducted under the same conditions of dimming. and presented in Figure. 23, Figure.24, and Figure.25. Indeed, for the three levels of duty cycle 95%, 70%, and 50%, the deviations of LED current in simulation and experimental results are 3.2mA, 2.2mA, and 3.7 mA respectively as shown in Table 2.

Table 2: Dimming of output current: deviation between simulation and experimental result

Dcyle		95%	70%	50%
$I_{LED, rms}$ (mA)	Simulation	287.2	248.2	206.7
	Experimental	284	246	203
Deviation (mA)		3.2	2.2	3.7

6. Conclusion

In this work, design, simulation, implementation, and test of a constant current LED driver with dimming capability have been done. Firstly, during the design stage we started by calculating and selecting the passive components and the semiconductor switches satisfying the system requirements and needs. Then, some simulations and measurements on a prototype of LED driver have been conducted. Comparison of simulation results to those obtained by experiment shows good agreement between them. In addition, regulation of the output current drawn by the LEDs and dimming of the brightness of these LEDs have been achieved. Indeed, the rated output current of the LED is maintained constant at 300 mA regardless the increase or the decrease of the output voltage. Such variation of this voltage is obtained by varying the number of LEDs forming the load. Furthermore, a full dimming is obtained by adjusting the pulse width modulation signal. Each dimming level of the light emitted by LEDs is achieved by controlling their rms current. In addition, a constant voltage LED using boost topology is designed and simulated. The obtained results demonstrated a good driver performance in terms of voltage regulation. This dc-dc boost driver is able to maintain a constant output voltage across the LEDs at 12.8V when the input voltage is adjusted or when the output current is changed by connecting LEDs in parallel.

In both cases, the output voltage is sensed by a voltage divider connected in parallel with the load then applied to a controller to adjust the suitable duty cycle of the signal controlling the power switch.

As a future work, these drivers will be combined in a single buck-boost driver operating as a constant-voltage LED driver (CV)

or as a constant-current LED driver (CC) under certain conditions of operation.

Conflict of Interest

The authors declare no conflict of interest.

References

- [1] H. Yamanaka, H. Yamada, "Dual active bridge DC-DC converter based wide dimming range LED driver with high-speed turn-off for high-brightness LED floodlight," *IEEE Journal of Industry Applications*, **8**(3), 556–557, 2019, doi:10.1541/ieejia.8.556.
- [2] B. Hanna, H. Markus, "Comparison of LED circuits," *OSRAM Opto Semiconductors*, 1–10.
- [3] P.S. Almeida, D. Camponogara, M.A.D. Costa, J.M. Alonso, "and Driver Life Spans," (June 2015), 36–47.
- [4] A.S.O. Ogunjuyigbe, T.R. Ayodele, V.E. Idika, O. Ojo, "Effect of lamp technologies on the power quality of electrical distribution network," *Proceedings - 2017 IEEE PES-IAS PowerAfrica Conference: Harnessing Energy, Information and Communications Technology (ICT) for Affordable Electrification of Africa, PowerAfrica 2017*, 159–163, 2017, doi:10.1109/PowerAfrica.2017.7991216.
- [5] Y.K. Cheng, K.W.E. Cheng, "General Study for using LED to replace traditional lighting devices," 173–177, 2006.
- [6] Y. Jeff, "Solid-state lighting: Lamps, chips, and materials for tomorrow," *IEEE Circuits and Devices Magazine*, **20**(June), 28–37, 2004, doi:10.1109/MCD.2004.1304539.
- [7] P.A. Meshram, "Investigation and Performance Analysis of Dc-Dc Converter for High Efficiency Led Driver," **2**(12), 444–452, 2016.
- [8] M. Al-Absi, Z. Khalifa, A. Hussein, "A New Capacitor-Less Buck DC-DC Converter for LED Applications," *Active and Passive Electronic Components*, **2017**, 0–5, 2017, doi:10.1155/2017/2365848.
- [9] LED drivers, *LEDs Magazine*, (September), 2019.
- [10] R. Xu, H. Li, Y. Li, X. Hou, "Research on a High-Efficiency LED Driving Circuit Based on Buck Topology," *Circuits and Systems*, **02**(04), 352–357, 2011, doi:10.4236/cs.2011.24048.
- [11] A. Report, "AN-1656 Design Challenges of Switching LED Drivers," (May), 1–7, 2013.
- [12] S. Rajesh, S.S.C. Bharathi, "A Universal Multistage High Brightness LED Driver for Domestic and Industrial Lighting Application," **5**(4), 107–111, 2014.
- [13] Y. Wang, J.M. Alonso, X. Ruan, "A Review of LED Drivers and Related Technologies," *IEEE Transactions on Industrial Electronics*, **64**(7), 5754–5765, 2017, doi:10.1109/TIE.2017.2677335.
- [14] T.I. Incorporated, "How to Design a Simple Constant Current/Constant Voltage Buck Converter," *Texas Instruments Incorporated, Application Report, SNVA829*, (June), 1–9, 2018.
- [15] D. Agrawal, R.K. Karn, D. Verma, R. Agrawal, "Dc-dc converter topologies for led driver circuit: A review," *International Journal of Circuits, Systems and Signal Processing*, **14**, 542–547, 2020, doi:10.46300/9106.2020.14.70.
- [16] I.M. Mrabet, "Buck converter Design Supervisor :"
- [17] H.J. Chiu, Y.K. Lo, J.T. Chen, S.J. Cheng, C.Y. Lin, S.C. Mou, "A high-efficiency dimmable LED driver for low-power lighting applications," *IEEE Transactions on Industrial Electronics*, **57**(2), 735–743, 2010, doi:10.1109/TIE.2009.2027251.
- [18] M. Nishikawat, Y. Ishizukat, H. Matsuot, K. Shigematsutt, "Circuit with Constant-Output- Current Control and Constant-Luminance Control," (2), 6–11, 2006.
- [19] P.S. Chavan, M.F.A.R. Satarkar, M.T. Asst, "Overview of control technique for DC-DC Converters," (May), 51–55, 2015.
- [20] M. Truntic, M. Milanovic, "Voltage and current-mode control for a buck-converter based on measured integral values of voltage and current implemented in FPGA," *IEEE Transactions on Power Electronics*, **29**(12), 6686–6699, 2014, doi:10.1109/TPEL.2014.2301935.
- [21] U. Nasir, Z. Iqbal, M.T. Rasheed, M.K. Bodla, "Voltage mode controlled buck converter under input voltage variations," *2015 IEEE 15th International Conference on Environment and Electrical Engineering, IEEEIC 2015 - Conference Proceedings*, 986–991, 2015, doi:10.1109/IEEEIC.2015.7165298.
- [22] J.N. Hemalatha, S.A. Hariprasad, G.S. Anitha, "Control strategy to generate pwm signals with stability analysis for dual input power converter system," *Indonesian Journal of Electrical Engineering and Informatics*, **7**(4), 609–619, 2019, doi:10.11591/ijeii.v7i4.976.
- [23] Reshma B R, Arun S Mathew, "Gain Scheduling Implementation in DC/DC Buck Converter using PID Controller," *International Journal of Engineering Research And*, **V4**(07), 1063–1067, 2015, doi:10.17577/ijertv4is070656.
- [24] D.W. Hart, *Power electronics*, 2011.
- [25] J. Byun, I. Hong, B. Lee, S. Park, "Intelligent Household LED Lighting System Considering Energy Efficiency and User Satisfaction," **59**(1), 70–76, 2013.
- [26] A. Raciti, S.A. Rizzo, G. Susinni, "Steady-State Electrical Modeling of LED and CF Bulb Lamps Under Variable Voltage on the Main Steady-state electrical modeling of LED and CF bulb lamps under variable voltage on the main," (June), 2018, doi:10.1109/EEEIC.2018.8494611.
- [27] L. Kukacka, P. Dupuis, R. Simanjuntak, G. Zissis, W. Silver, "Simplified models of LED ballasts for spice," 1–5, 2014.
- [28] R.V. Monteiro, B.C. Carvalho, G.C. Guimarães, M.A. Tamashiro, "Computational & Math. Modeling of a Buck Driver Type of a Tubular Led Lamp," *IEEE Latin America Trans.*, **14**(5), 2016, doi:10.1109/TLA.2016.7530414.

Analysis of Reading Time and the Number of Characters within One Gazing Point

Yusuke Nosaka*, Miho Shinohara, Hidemi Ishikawa, Yuko Hoshino, Mitsuho Yamada

Tokai University, Tokyo, 108-8619, Japan

ARTICLE INFO

Article history:

Received: 18 October, 2021

Accepted: 05 December, 2021

Online: 23 December, 2021

Keywords:

Gaze point

Reading time

Display Method

ABSTRACT

Concomitant with the ubiquity of lightweight electronic terminals, such as smartphones and tablets, e-books have increased in popularity. However, studies have shown that their display method and information-presentation efficiency need improvement. We investigated visual information processing during reading to develop a more efficient display method that presents information in an easier to see fashion than conventional methods. Specifically, we focused on the information-processing ability of the brain within the gazing point during reading to clarify various factors that make a book easier to read such as the minimum display time and line length that can be read accurately and the difference in reading time depending on the font. An experimental method using a “window” was utilized that facilitates changing of the display information in the window by moving the window in conjunction with the eye movement. Using this method, it was found that if the number of characters displayed in the window is more than six and the window display time is 56 ms or more, reading can be performed in the same manner as when the window is not displayed. Thus, here we focused on letter size to investigate the relationship between reading time and font size. Experimental results indicate that among sentences in three font sizes 12, 18, and 24 pt normal reading without any restriction of view is possible when the font size is 18 pt or more. These results suggest that normal reading can be performed even with a very short display time of 56 ms by using characters with a font size of 18 pt or more in Japanese, which is useful for proposing a new display method for e-books. In addition, this experimental method can be applied to languages other than Japanese, and can contribute to the development of an optimum display method for each language.

1. Introduction

In recent years, small and lightweight electronic terminals, such as smartphones and tablets, have become widespread. Their importance has been increasing because information is digitized and can be accessed from various terminals. Simultaneously, e-books are becoming more widespread, and dedicated terminals and smartphone applications for e-books have appeared. An environment that facilitates the reading of e-books rather than conventional paper books has been established, and the advantages of e-books such as term search function, character size change, and percentage display of the current reading position further support their spread. Studies exist on eye movements related to reading: analyses have been performed on the impact of negative and positive words on reading [1], and the conjecture of the concentration degree from eye movements during reading [2]. In recent years, studies have been conducted on the effectiveness of

reading and education using e-books, for example, learning via reading e-books and paper books has been compared and analyzed [3, 4], and effective learning methods using e-books have been developed and evaluated for foreign languages [5, 6].

There are differences in eye movements between reading e-books and conventional paper books. In particular, eye movements when turning pages differ between paper books and e-books, depending on the manner the pages are turned. In the case of e-books, the eye often moves linearly from the end of the page to the beginning of the screen where the next page is displayed, whereas in the case of paper books, it may follow the movement of the paper on the page to be turned [7]. We hypothesized that, by analyzing eye movements during reading, it would be possible to develop new display methods reflecting the acceptance characteristics of human visual information not only in e-books but also in other applications that present visual information on displays.

*Corresponding Author: Yusuke Nosaka, 23-3- 2 Minato Tokyo Japan, 03-3441-1171 & yusuke_n@star.tokai-u.jp

www.astesj.com

<https://dx.doi.org/10.25046/aj060618>

A previous study clarified the minimum display time and the minimum number of characters required for display within one gazing point required for normal reading. Consequently, in this current study, we investigated the differences in reading time by changing the font size displayed in one gaze point. By clarifying the minimum font size required for normal reading through this study, we hope to propose a display method that enables efficient and fast reading.

The eye movement during reading is a repetition of saccade, which is a high-speed and intermittent eye movement, from one gaze point to the next gaze point [8–11]. During saccade, the visual-information-processing ability is reduced owing to the perceptual phenomenon of saccadic suppression [12–14]. More specifically, the information processing during reading is mainly performed within the gaze period, and it is necessary to examine the information-processing ability within the gaze point to clarify the visual-information-processing ability during reading. However, the only way to analyze the information-processing capacity within the gazing point is to use a “window” that can control the visual information displayed within the gazing point. To this end, we used a device developed in previous research for controlling the display image in the gaze point [15]. By so doing, it is possible to obtain the knowledge necessary for studying a display method that adaptively processes the size and display time of display characters based on the characteristics of human visual information processing.

Previous studies in this area are extremely important for understanding this paper. Thus, Section 2 summarizes the research on the perceptual range (effective field of view) that had a major influence on previous research. In Section 3, we discuss the analyzer [15] and the experimental results on the reading time when the gaze time is limited to 98 ms. Furthermore, the experimental results that clarified the lower limit of the gaze time required for reading are described [16]. Section 4 describes our analysis method and the experimental results obtained for different font sizes. Section 5 concludes the paper and discusses future developments.

2. Effective Field of View during Reading

Humans do not recognize all the information in their field of view, only the part on which they are focused. The size of the effective visual field changes depending on the type of visual work, but the range in which such effective information can be obtained is called the effective field of view [17]. Many experiments have been conducted to investigate the effective field of view, but all of them used estimation methods. Subsequently, by proposing an experimental method using windows [18], the research advanced dramatically by progressing from estimation.

The visual task in this study is reading. The experimental method using a window is a method of changing the display information in the window by moving the window in conjunction with the eye movement. The effective field of view is investigated by moving the window within a period called saccadic suppression in which the function of visual information processing deteriorates in the brain; the window refers to the display range of the image on the display.

2.1. Study Conducted by Saida and Ikeda

In [19], the author acquired the eye movements of their subjects using the limbus tracking method. Limbus tracking is a method of irradiating infrared light near the boundary between the sclera and cornea and measuring eye movements by utilizing the difference in reflectance. The researchers used two TV cameras. They input the acquired X and Y positions of the eyeball into the X-Y oscilloscope and displayed a horizontally long rectangle at that position. The image of the X-Y oscilloscope was recorded with the first camera, and the board on which the text was written with the other camera. The area corresponding to the X and Y positions of the eyeball was switched with a switcher similar to a wipe process using the rectangular image generated by shooting the X-Y oscilloscope from the video of the TV camera that recorded the text. In this way, the window for displaying characters was moved in conjunction with the eye movements. Based on this experiment, Saida and Ikeda indicated that sentences can be read normally, as in the condition without a window, if the characters displayed on the rectangle (window) exceeded 13 characters. This indicates that the processing range of one gazing point is approximately 13 characters.

However, three limitations of this experiment should be noted. (1) The rectangle is displayed directly on the oscilloscope by the output of the eye movement, the gaze point is not detected, and the gaze part is not switched. To clarify the visual information processing at the gazing point, it is necessary to first detect the gazing point and change the image of this area. (2) Because of the camera and cathode-ray tube (CRT) display used, it is expected that afterglow and afterimage would occur, but no description was provided. Therefore, it is not possible to obtain an accurate display time for presenting a window image. (3) The display time of the window changes depending on the gaze time because the window moves in conjunction with the eye movement. Controlling the display time of information at the gazing point is indispensable when considering information processing at the gazing point.

2.2. Studies Conducted by McConkie and Rayner

In [20–23], the author input the eye movements acquired by the limbus tracking method into a PDP-6 minicomputer and controlled the DEC Model 340 display with built-in uppercase and lowercase characters. With this display, they conducted experiments by preparing six types of replacement patterns, such as characters with similar glyphs, in addition to the original text characters displayed at the gazing point. In their studies, they showed that the effective field of view was 13 characters, which is similar to the findings of Saida and Ikeda.

The abovementioned experiment has two issues: (1) The time the previous character remained on the screen at the time of switching was not measured, only a description of short afterglow was provided. However, the experiment was conducted with a CRT using phosphor with the least afterglow, and the authors were aware of the effects of the afterglow. They stated that the letters were not perceived before being covered by the window, and therefore afterglow did not interfere with the experiment. (2) The saccade was not detected, the window image was switched when the eye movement exceeded a certain distance from the left end of the sentence, and the display time of the window image was not

controlled. Therefore, it cannot be replaced only for a certain period within the gazing point.

In addition, recent studies using windows introduced research on the perceptual range when learning a language other than the mother tongue [24] and on the perceptual characteristics when the resolution is changed in conjunction with eye movement. Leung investigated the range of perception of Japanese university students learning English as a foreign language while reading English sentences; however, they did not investigate the temporal characteristics of human visual information processing during reading. Loschky and McConkie examined the time to perceive blur or flicker when displayed at different resolutions from the standpoint of user performance, but they did not examine the time required to acquire visual information.

A device has subsequently been developed that can detect eye movements during reading and control the display information around the gazing point. This enables clarification of the amount of information that can be processed within one gazing point and the required processing time [7, 15, 16].

3. Previous Research on Information Processing within the Gazing Point

This study clarifies how the reading time changes depending on the size of the font displayed in the gazing point that occurs during reading. For this experiment, it is necessary to solve the problems described in Section 2, that is, afterimage, afterglow, and real-time detection of gaze point. To this end, the features of the device developed by Kushima for that purpose are explained in Section 3.1. Section 3.2 describes the basic experiments performed to detect the gaze point that occurs during reading in real time, and describes the saccade velocity as the threshold value for detecting saccade in this study using the device described in Section 3.2. Section 3.3 describes the experimental results of using this device to extract the gaze point with the saccade velocity as the threshold, display the window, and change the number of characters displayed in the window. Through this experiment, Kushima clarified the number of characters displayed in a window that allows for reading equivalent to that without a window. Section 3.4 further describes the research results obtained by Nomura regarding the window display time, that is, the minimum gaze time, which enables reading equivalent to that without a window.

Based on these results, Section 4 describes the experimental results of the current study, focusing on the number of characters displayed in the gazing point, the display time, and the font displayed in the window, especially the font size.

3.1. Development of Display Image Control Device

In [15], the author developed an experimental device to solve the problems described in the previous section. Table 1 shows the equipment used, Figures 1 and 2 show the configuration diagram and flowchart, respectively. Figure 3 shows the display for the subjects, and Figure 4 shows the eye movement measuring device, EMR-9.

Table1: Equipment used in the experimental device [15]

Eye Movement Measuring Device	EMR-9 (NAC Imaging Technology)
Data Processing PC	OS: Windows 7 Professional
	CPU: Core i7 3770 @ 3.40 GHz
	Memory: 32GB
Display	CRT Display (iiyama HM204D A)

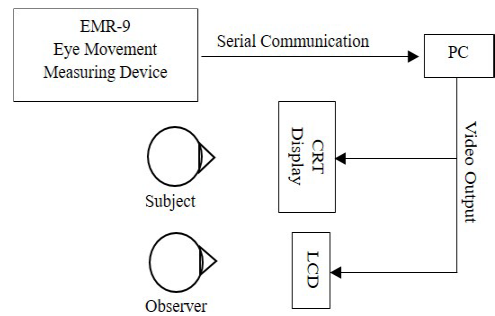


Figure 1: Configuration of the experimental device developed in [15]

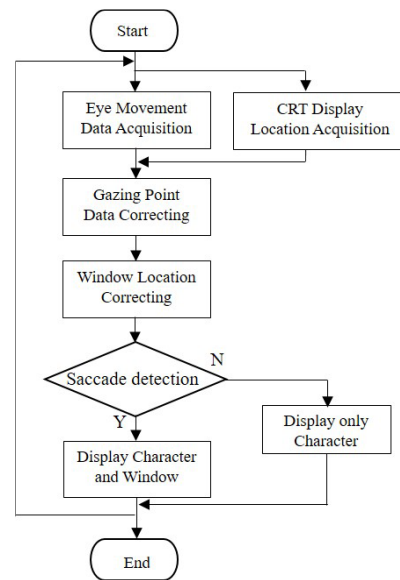


Figure 2: Flowchart of the experimental system developed in [15]

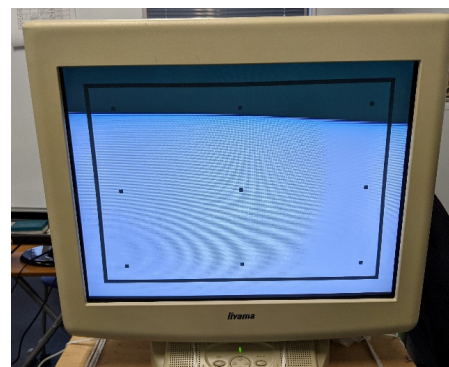


Figure 3: CRT display for the subject: a rectangle is displayed to estimate the coordinates of the field of view camera.



Figure 4: Eye movement measuring device EMR-9

It is a compact and lightweight eye tracking device based on the concept of mobility. It is possible to more naturally measure daily behavior without burdening the subject.

Eye movement is input from the eye movement measuring device to the information processing PC, and the saccade is detected by the threshold value described in Section 3.3. The displays consisted of a liquid crystal display for the experimenter and a CRT display for the subject (Figure 3). For the subject display, a short afterglow CRT display (< 1 ms) was used instead of the hold-type liquid crystal display. Since CRT displays are no longer commercially available, they have been maintained and dynamically preserved for such experiments. The text is displayed on the CRT display for the subject, and the image of the field camera of the eye movement measuring device EMR-9 is output on the liquid crystal display for the experimenter to monitor the movement of the line of sight when the subject is reading the text. The software on the information processing PC executes two threads in parallel. One thread acquires the position of the rectangle to be displayed on the CRT display from the image of the field camera of EMR-9 (Figure 3). The other thread acquires eye movement data from EMR-9 via serial communication and calculates the position and movement speed of the gazing point, the angle of rotation of the eye, and the angular acceleration. If a saccade is detected, a window image is presented at the position of the gaze point on the CRT display (lower right of Figure 5); if a saccade is not detected, only the first three characters of each line are displayed as a reference to start reading.

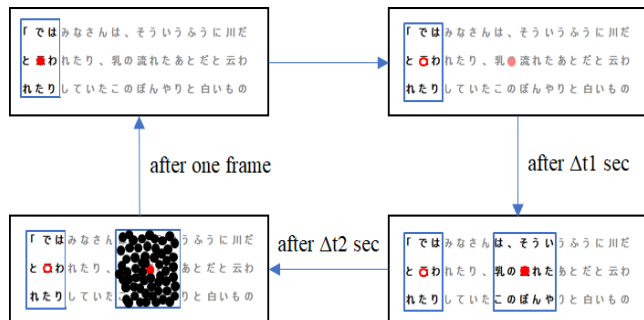


Figure 5: Examples of switching the display image in the window during the experiment

Figure 5 shows an example of switching the display image in the window during the experiment. Figure 6 shows the screen of the corresponding CRT display. The red dot in Figure 5 is the gazing point (the point with only the outline is the previous gazing point), and the blue frame is the range displayed to the subject.

When the gazing point moves and a saccade is detected, a window image is displayed after Δt_1 seconds of a saccade detection, as shown in the lower right of Figure 5. The displayed window image was cut out from the displayed text based on the size of the window. After displaying the window image for Δt_2 seconds, random dots (masking image) are displayed for one frame in the range where the window image is displayed. This process was repeated until the subject finished reading. In the experiment, the experimenter arbitrarily changed the values of Δt_1 and Δt_2 .

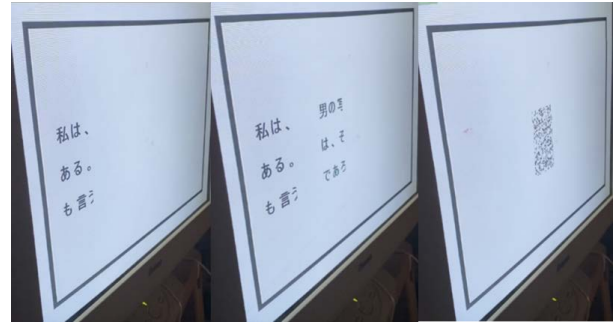


Figure 6: Changes on the CRT screen corresponding to Figure 5

Figure 7 shows the time flow of the process of displaying the window image at the gazing point. The switching of the displayed image in the window must be performed within saccadic suppression, which is the period during which the ability of visual information processing in the brain is reduced. As shown in Figure 7, it takes 52 ms to capture the eye movement data from EMR-9 into the information processing computer. Overall, it takes 85 ms to determine the gazing point. This is within the time range of a saccadic suppression, but considering that saccadic suppression lasts for approximately 200 ms, it is necessary to control the image replacement time of the gaze part in millisecond units. Because the refresh rate of a general display is 60 Hz and drawing is performed approximately every 16.7 ms, it is considered that a delay of 16.7 ms at the maximum will occur. In addition, if the image switching of the gaze part occurs immediately after the drawing of the display is performed, image switching of the gaze part may not occur on the display. To prevent this phenomenon, the author implemented a vertical synchronization function that synchronized the drawing of the display with the drawing of the program.

Delay (52 ms) until the measuring device detects the eye movement, and the information is received by the information processing computer via serial communication.

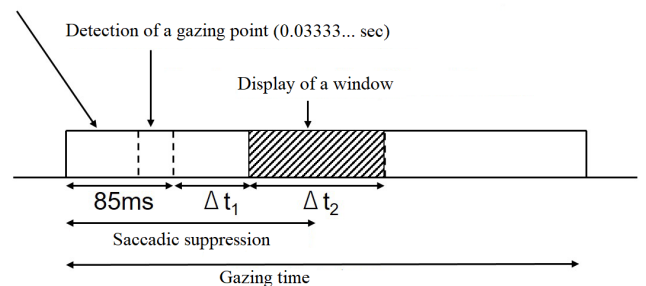


Figure 7: Time flow of the process of displaying the window image in the gazing point

Δt_1 : time from the detection of saccade until the window image is displayed
 Δt_2 : Time to display the window

3.2. Definition of a Gazing Point

Eye movement during reading is not smooth, but is instead a repetition of gaze and saccade which is an intermittent and high-speed eye movement. In this study, the text was displayed by wiping only the gaze part of the window while reading. For that purpose, it was necessary to extract only the gazing points from the eye movements. Various studies have been conducted to define the gazing point. Even in the state of gaze, the eyeballs are not completely stationary, and small involuntary movements called miniature eye movements occur. Therefore, the author defined a gazing point as including the velocity component of miniature eye movements and smooth pursuit eye movement below a certain velocity [25]. In contrast, they studied the definition of the gazing point when gazing at a video image that is constantly moving [26]. Consequently, it has been reported that the definition of the gazing point differs depending on the viewing target, and it is necessary to consider the definition of the gazing point according to the experimental conditions and environment. Thus, we conducted an experiment to define the gazing point in the same experimental environment and conditions as the experiment conducted this time.

In [26, 27], the cases where the eye movement velocities of 5°/s and 10°/s or less last for 150 ms or more are extracted as the gazing point. However, with this definition, it takes at least 150 ms to extract the gazing point, and the window cannot be displayed at the same time as the gaze. The viewing angle per character of the text used in this experiment is about 1.2°, and the average interval of saccades is three to five characters in Japanese kana/kanji mixed text [27]. Thus, viewing targets were displayed at three different intervals of 1, 2, and 3° expected from the subject who repeated the gaze to the left and right from the center of the screen every 5 s 10 times, as shown in Figure 8.

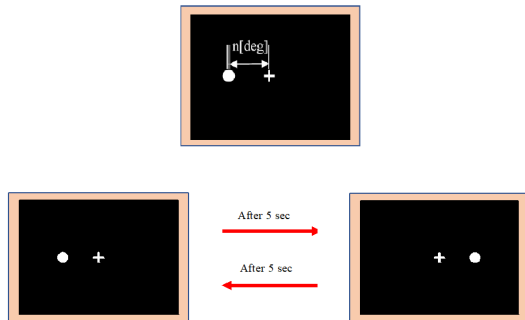


Figure 8: Gaze experiment of moving optotype conducted to determine the speed at which the gaze point is extracted

Based on the measurement results obtained in this manner, the eye movement speed of the threshold value for extracting the start and end of the saccade was changed from 30°/s to 90°/s, and the gazing point was extracted. The gazing point was defined as the period from the end of the saccade to the start of the next one. The detection accuracy of the gazing point was verified by comparing the extracted gazing point with the real gazing point.

Figure 9 shows an example in which the gazing point is detected from the eye movement when gazing at a moving target with a viewing angle of 1°. If the threshold speed is appropriate, a total of 20 gazing points, 10 on each side, should be detected. Comparing the eye movement speed thresholds of 42°/s and 48°/s,

the detection of the gazing point in the red frame failed at 48°/s, but it was detected accurately at 42°/s. Similar results were obtained with viewing angles of 2° and 3°. In the studies [15, 16], and the study reported herein, it was judged that saccade started when the eye movement speed exceeded 42°/s and ended when it was less than 42°/s. The gazing point was defined as the duration from the end of one saccade to the start of the next saccade.

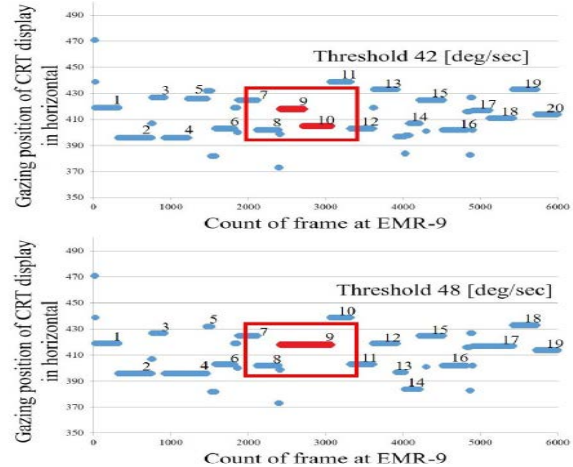


Figure 9: Detecting the gazing point by changing the threshold value for determining the start and end of a saccade (Movement amount of viewing target: 1°, detected gaze point is blue line)

3.3. Number of Characters in a Window

Kushima compared the reading time by increasing or decreasing the number of characters displayed in the window from 3 to 17 characters by two characters using the abovementioned experimental device and definition of gazing point. The image of the text read was an excerpt from a Japanese textbook for third graders. The purpose was to confirm the validity of the definition of the experimental device and the gazing point, and the experiment was conducted by setting the values of Δt_1 to 14 ms and Δt_2 to 98 ms.

They found that when the window width was greater than nine characters, there was no substantial change in the reading time; that is, when the window display time was 98 ms and nine or more characters were displayed in the window, the reading was the same as without a window [15].

3.4. Minimum Gaze Time

Kushima conducted the experiment with a window display time of only 98 ms. Nomura considered that reading was possible even with a shorter display time, and conducted a similar experiment with a display time of 98 ms or less to investigate the lower limit of the gaze time required for reading. Using the device developed in [15], Δt_1 was set to 14 ms, and Δt_2 was set to one of six values: 84, 70, 56, 42, 28 and 14 ms. The displayed image contained excerpts from three novels by Aozora Bunko: “Galaxy Railroad Night,” “I am a Cat,” and “Bochan.”

The results showed that when the window display time (Δt_2) was short, the reading time was long, and conversely, when the display time was long, the reading time was short. It was also shown that the reading time did not change significantly when the window display time was 56 ms or more and the window width

was seven characters or more. Furthermore, the reading time was long and unstable regardless of the window width when the window display time was 42 ms or less [16].

4. Font Size Experiment

Previous studies have shown that when the window display time is 56 ms or more and the window width is seven characters or more, there is no significant difference in reading time with and without a window. The purpose of this experiment is to clarify the font size that enables reading in the same manner as without a window and the relationship between reading time and the font size displayed at one gazing point.

4.1. Experimental Method

Experiments were performed using the same procedure as in [15]. The subjects were seven male students ages 21-22 years from our university. Eight text images were presented to the subjects in order, and the window width was increased by two characters from 3 to 17 characters (up). Then, the window width was presented as reducing by 2 characters from 17 to 3 characters (down). In addition, we prepared three font sizes: 12, 18, and 24 pt. The displayed images were excerpts from the beginnings of novels, "Brothers," "Restaurant with many orders," and "Human disqualification," by Aozora Bunko. We used the early part of each novel so that the difficulty level of reading did not change depending on the content of the text. Figure 10 shows an example of the text from "Brothers."

父がなくなったときは、長兄は大学を出
たばかりの二十五歳、次兄は二十三歳、
三男は二十歳、私が十四歳でありました

12 pt

父がなくなったときは、長兄は大学を出
たばかりの二十五歳、次兄は二十三歳、
三男は二十歳、私が十四歳でありました

18 pt

父がなくなったときは、長兄は大学を出
たばかりの二十五歳、次兄は二十三歳、
三男は二十歳、私が十四歳でありました

24 pt

Figure 10: Display examples for each font size (Brothers)

The reading times of the up and down experimental sequences were averaged and compared for each font size. Δt_1 was set to 14 ms, which is the same as in the previous study [15], and Δt_2 was set to 98 ms, which enables reading without visual field restriction. In addition, to suppress the influence of the presentation order on the experimental results, we adjusted the order in which the images and font sizes were presented to the subjects. To motivate the subjects to fully understand and read the content of the text, the subjects were informed that a confirmation test of comprehension

was to be performed after the measurement; this test was conducted after the measurement. This experiment was conducted in accordance with the Tokai University ethical code for human experiments.

4.2. Experimental Results

Figure 11 shows the average reading time for each font size. When the font size was 24 pt, the reading time decreased when the window width expanded from three characters to nine characters, and the fluctuation in reading time was small after nine characters. Similar results were observed at 12 and 18pt. These results are consistent with the results of previous studies [15, 25], and show that if the window width is nine characters or more, normal reading without visual field restriction is possible. From the measured reading time data, analysis of variance was performed based on two factors: font size and window width. The results are shown in Table 2. There was a significant difference at the significance level of 1% in the results of only one factor for each of the font sizes and the window widths. There was no significant difference in the interaction between font and window width sizes.

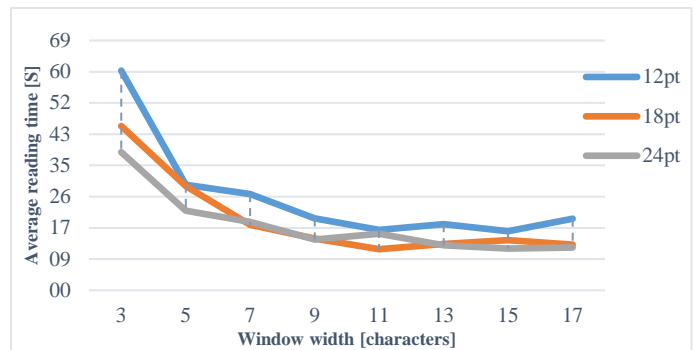


Figure 11: Average reading time for each font size

Table 2: Results of ANOVA for font size and window width

	Results
A: Font size	F (2, 144) = 8.577, p < .0001
B: Window width	F (7, 144) = 19.649, p < .0001
A×B	F (14, 144) = 0.940, p = .5183

As our focus is on font size, we performed multiple comparisons for the three font sizes 12, 18, and 24 pt; the results are shown in Table 3. There were significant differences between the 12 pt group and both the 18 and 24 pt groups, but there was no significant difference between the 18 and 24 pt groups. These results indicate that when the three font sizes, 12, 18, and 24 pt, are presented, there is no significant change in reading time when the font size is 18 pt or more, and normal reading without visual field limitation is possible.

Table 3: Results of multiple comparisons for font sizes

Pair	p-value
12pt × 24pt	0.0003071
12pt × 18pt	0.0007033
18pt × 24pt	0.8138201

4.3. Discussion

Effectively displaying text information on electronic media such as e-books can help to reduce reading time and improve

readability. Eye movements during reading are repeated saccades and gaze points. Clarifying the visual information-processing ability during gaze can be useful for proposing a new display method more suitable for information processing in the human brain.

To clarify the visual-information-processing ability in the gazing point, the number of characters displayed in the gazing point and the display time within the gazing point that enable reading similar to normal reading have been investigated by using a window that moves in conjunction with the gazing point.

The experimental results thus have helped to clarify based on the reading time that it is possible to read the same as normal reading without a window by displaying more than six characters for 56 ms or more.

On the other hand, regarding readability, not only in e-books but also in conventional paper books, information of another dimension such as the type and size of the font to be displayed is also involved.

Consequently, in this study, we focused on the size of the font. The smaller the font is, the more information that can be displayed on a limited page, and the more characters that can be displayed in one gazing point. However, it also has the problem of being difficult to read. Therefore, the reading time was evaluated by changing the font size. Our result indicate that there is a significant difference between 12 pt font and 18 and 24 pt font. Furthermore, we were able to obtain knowledge about the display time, number of display characters, and size of the display font.

There are two styles of reading, one is to read and take the time to full absorb all the content, and the other is to read it to get information as quickly as possible. We believe that the results of this study will be useful in proposing a display method for faster and more accurate reading in the latter reading style. Nevertheless, we also believe that this research will be useful for realizing a display environment that facilitates the intake of information and allows more concentration on content even in the former reading style.

The difficulty of studying reading is that it depends on the language. Correct conclusions may not be obtained unless a native language researcher analyzes the native language subject. Therefore, related papers tend to be grouped by language. In the future, it will be important to compare the experimental results between different languages, and we hope that similar research will proceed in other languages with reference to this research result.

5. Conclusion

In recent years, small and lightweight electronic terminals, such as smartphones and tablets, have become widespread. In addition, the spread of e-books is supported by their specific advantages and many associated studies have been conducted. We investigated visual information processing during reading to study a display method that is easier to see and more efficient in presenting information on a display than conventional display methods.

In a previous study, Kushima reported that they developed a measuring device that solves the problems that occurred in www.astesj.com

experiments to investigate the effective visual field in the past. They also reported that by displaying the image for 98 ms at the time of gazing, regular reading speed was the same as that without visual field limitation. Furthermore, Nomura showed that reading is not possible with the same device at a display time of 42 ms or less and that normal reading is possible by displaying sentences in a window of seven characters at 56 ms.

In this study, to investigate the relationship between reading time and the font size, an experiment was conducted in which sentences were presented in three font sizes, 12, 18, and 24 pt, and the image was displayed for 98 ms at the time of gazing. The results showed that normal reading without any restriction of view is possible when the font size is 18 pt or more.

From the results obtained this time and the results so far, it was shown that by displaying seven or more characters with a size of 18pt or more and for a period of 56ms or more in one gazing point, reading can be performed in the same manner as without a window.

In this study, the result for seven adult male subjects are reported. More participants actually volunteered for the experiment, but the number of subjects for which eye movements could be accurately detected on a character-by-character basis was limited to seven. To increase the number of subjects, it is necessary to improve the eye movement measuring device so that the eye movement can be acquired more easily and accurately.

Furthermore, research is needed that takes into account not only differences in gender and age, but also differences in reading habits. Results may differ between those who read a lot and those who do not. In addition, reading studies may not be evaluated correctly unless native language researchers conduct studies using native language subjects. Thus, the references are biased by language and it is difficult to compare the results between languages. However, the visual information-processing ability within the gazing point may be common to humans even if the language is different. In the future, we hope that this research will be useful for fostering a new culture related to languages, such as displaying the translation results of translation software that takes into account the readability of each language by making comparisons between various languages.

References

- [1] G. G. Scott, P. J. O'Donnell, S. C. Sereno, "Emotion words affect eye fixations during reading," *Journal of Experimental Psychology, Learning, Memory, and Cognition*, **38**(3), 783–792, 2012, doi:10.1037/a0027209.
- [2] L. Steindorf, J. Rummel, "Do your eyes give you away? A validation study of eye-movement measures used as indicators for mindless reading," *Behavior Research Methods*, **52**, 162–176, 2020, doi:10.3758/s13428-019-01214-4.
- [3] M. Pinto, C. Pouliot, J. A. Cordón-García, "E-book reading among Spanish university students," *The Electronic Library*, **32**(4), 473–492, 2014, doi:10.1108/EL-05-2012-0048.
- [4] J. Lim, G. E. K. Whitehead, YunDeok Choi, "Interactive e-book reading vs. paper-based reading: Comparing the effects of different mediums on middle school students' reading comprehension," *System*, **97**, 102434, 2021, doi:10.1016/j.system.2020.102434.
- [5] N. F. Almunawaroh, "The effectiveness of using an e-book in elt: worldwide cases," *journal of teaching & learning english in multicultural contexts (tl emc)*, E-ISSN: 2541–6383, **4**(2), 2020.
- [6] L. Lecailliez, B. Flanagan, M. A. Chen, H. Ogata, "Smart dictionary for e-book reading analytics," *LAK '20: Proceedings of the Tenth International Co*

- nference on Learning Analytics & Knowledge, 89–93, 2020, doi:10.1145/3375462.3375499.
- [7] H. Takahira, R. Ishikawa, K. Kikuchi, T. Shinkawa, M. Yamada, “Analysis of gaze movement while reading e-books,” *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, E97-A (2), 530–533, 2014, doi:10.1587/transfun.E97.A.530.
- [8] E.B. Huey, “Preliminary experiments in the physiology and psychology of reading,” *American Journal of Psychology*, **9**, 575–586, 1898.
- [9] E.B. Huey, “On the Psychology and physiology of reading 2,” *American Journal of Psychology*, **12**, 292–312, 1900.
- [10] E.R. Hilgard, Pain as a puzzle for psychology and physiology. *American Psychologist*, **24**(2), 103, 1969.
- [11] E.B. Huey, “The psychology and pedagogy of reading,” Macmillan Company (Currently resold at MIT Press), 1908.
- [12] L.N. Thibos, F.E. Cheney, D.J. Walsh, “Retinal limits to the detection and resolution of gratings,” *Journal of the Optical Society of America*, 1524–1529, 1987.
- [13] R.J. Jacobs, “Visual resolution and contour interaction in the fovea and periphery,” *Vision Research*, **19**, 1187–1195, 1979.
- [14] F.C. Volkman, A.M. Shick, L.A. Riggs, “Time course of visual inhibition during voluntary saccades,” *Journal of the Optical Society of America*, **58**, 4, 562–569, 1968.
- [15] T. Kushima, M. Suganuma, S. Mochiduki, M. Yamada, “Development of a novel accurate analysis system regarding information processing within the gazing point,” *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, E102-A (9), 1205–1216, 2019, doi:10.1587/transfun.E102.A.1205.
- [16] K. Nomura, Y. Furumi, T. Kushima, H. Otomo, S. Mochiduki, M. Yamada, “Analysis of gaze time required for reading,” *The Institute of Image Information and Television Engineers (ITE) Annual Convention*, 13E-2, 2019.
- [17] T. Hatada, H. Sakata, H. Kusaka, “Psychophysical analysis of the “Sensation of Reality” induced by a visual wide-field display,” *SMPTE Journal*, **89** (8), 560 – 569, 1980, doi:10.5594/J01582.
- [18] M. Saida, “Reading and eye movement,” in *Experimental Psychology of Eye Movement*, R. Imosaka, Y. Nakamizo, K. Koga eds., 167-197, The University of Nagoya Press, 1993.
- [19] M. Saida, M. Ikeda, “Sentence interpretation with limited field of view,” *Clinical Ophthalmology*, **29**(8), 923–925, 1975.
- [20] G.W. McConkie, K. Rayner, “The span of the effective stimulus during a fixation in reading,” *Perception and Psychophysics*, **17**(6), 578–586, 1975, doi:10.3758/BF03203972.
- [21] K. Rayner, “The perceptual span and peripheral cues in reading,” *Cognitive Psychology*, **7**(1), 65–81, 1975, doi:10.1016/0010-0285(75)90005-5.
- [22] K. Rayner, “Parafoveal identification during a fixation in reading,” *Acta Psychologica*, **39**(4), 271–282, 1975, doi:10.1016/0001-6918(75)90011-6.
- [23] C.Y. Leung, M. Sugiura, D. Abe, L. Yoshikawa, “The perceptual span in second language reading: An eye-tracking study using a gaze-contingent moving window paradigm,” *Open Journal of Modern Linguistics*, **4**(5), 585–594, 2014, doi:10.4236/ojml.2014.45051.
- [24] L.C. Loschky, G.A. McConkie, “User performance with gaze contingent multi-resolution displays,” in *Proceeding of the 2000 Symposium on Eye Tracking Research & Applications (ETRA)*, 97–103, 2000, doi:10.1145/355017.355032.
- [25] M. Yamada, T. Fukuda, “Definition of gazing point for picture analysis and its applications,” *IEICE Transactions*, J69-D (9), 1335–1342, 1986.
- [26] R. Fukuda, M. Sakuma, E. Nakamura, T. Fukuda, “An experimental consideration on the definition of a fixation point,” *The Japanese Journal of Ergonomics*, **32**(4), 197–204, 1996, doi:10.5100/jje.32.197.
- [27] N. Osaka, “Effect of peripheral visual field size upon eye movements during Japanese text processing,” J. K. O’Regan, A. Levy-Schoen, eds., *North-Holland*, 1987, doi:10.1016/B978-0-444-70113-8.50063-1.

Video Risk Detection and Localization using Bidirectional LSTM Autoencoder and Faster R-CNN

Idir Boulfri^{*}, Mohamed Lahraichi, Khalid Housni

Faculty of Science, Ibn Tofail University, Kenitra, 14000, Morocco

ARTICLE INFO

Article history:

Received: 09 November, 2021

Accepted: 05 December, 2021

Online: 30 December, 2021

Keywords:

Convolution neural network

Fast R-CNN

LSTM

Auto-encoder

ABSTRACT

This work proposes a new unsupervised learning approach to detect and locate the risks "abnormal event" in video scenes using Faster R-CNN and Bidirectional LSTM autoencoder. The approach proposed in this work is carried out in two steps: In the first step, we used a bidirectional LSTM autoencoder to detect the frames containing risks. In the second step, for each frame containing risks, we first used Faster R-CNN to extract all the objects containing in the scene and then for each object detected we check whether it represents a risk or not. In other words, in testing phase, the frames with events deviated from normal features learned in training phase are detected as risk. To locate objects representing risk, only the objects detected by Fast R-CNN deviated from normal feature are classified as risk. Experimental results demonstrate that the proposed method can reliably detect and locate the object representing risk in video sequences.

1. Introduction

The security of public spaces has become a very important area in recent years, hence the need to develop an automated surveillance system capable of analysing video scenes, exactly detect and locate anomalies. In order to respond to this, demand several approaches have been proposed based on sparse coding [1]-[3] or deep learning techniques [4]-[6]. The last one is divided on two types of learning techniques [1], the supervised setting which requires both normal and abnormal labelled training samples, but it is difficult to obtain a training labelled data set, moreover, unsupervised methods avoid excessive manual labelling, it can only be applied on the specific scenes because they are using underlying data and prior knowledge to design limited distributions.

In recent years, deep learning has become an important domain and have been applied for a diverse set of tasks, the anomaly detection is one, most research on this area was based on RNN and CNN network architectures, such as deep generative models such as variational autoencoder (VAE) [7], generative adversarial networks (GANs) [8], Long Short-Term memory networks (LSTMs) [9], deep learning have resolved the weakness of traditional method.

In real life there is some challenging problems to detect anomalous video, like the lack of clear definition of anomalies, difficulty of scene segmentation, high density object with random motions, occlusions, and the fact that a risk appears rarely in short time. This paper focused on unsupervised risk detection and localization approaches based on convolutional neural network algorithms.

Our approach present a model based on the convolutional neural network, Bidirectional LSTM auto-encoder and Fast C-RNN. The convolution neural network auto-encoder captures the local structure and the LSTM auto-encoder captures temporal information, it learns the normal patterns from the normal training videos then the risks are detected as events deviated from the normal patterns learned. To detect risk in frame we compute de difference between the original frame and the reconstructed frame, and performing a threshold error will detect the frame contains risk. In the second step the Fast C-RNN extract objects from the frame with risk, to localize risk in this frame the error is only computed between objects detected in normal frame and those objects in the reconstructed frame. Fig. 1 illustrates an overview of the proposed method.

The reminder of this paper is organized as follows: The second section presents the methodology, the third section shows experimental results, and finally a conclusion.

^{*}Corresponding Author: Idir Boulfri, iboulfri@gmail.com

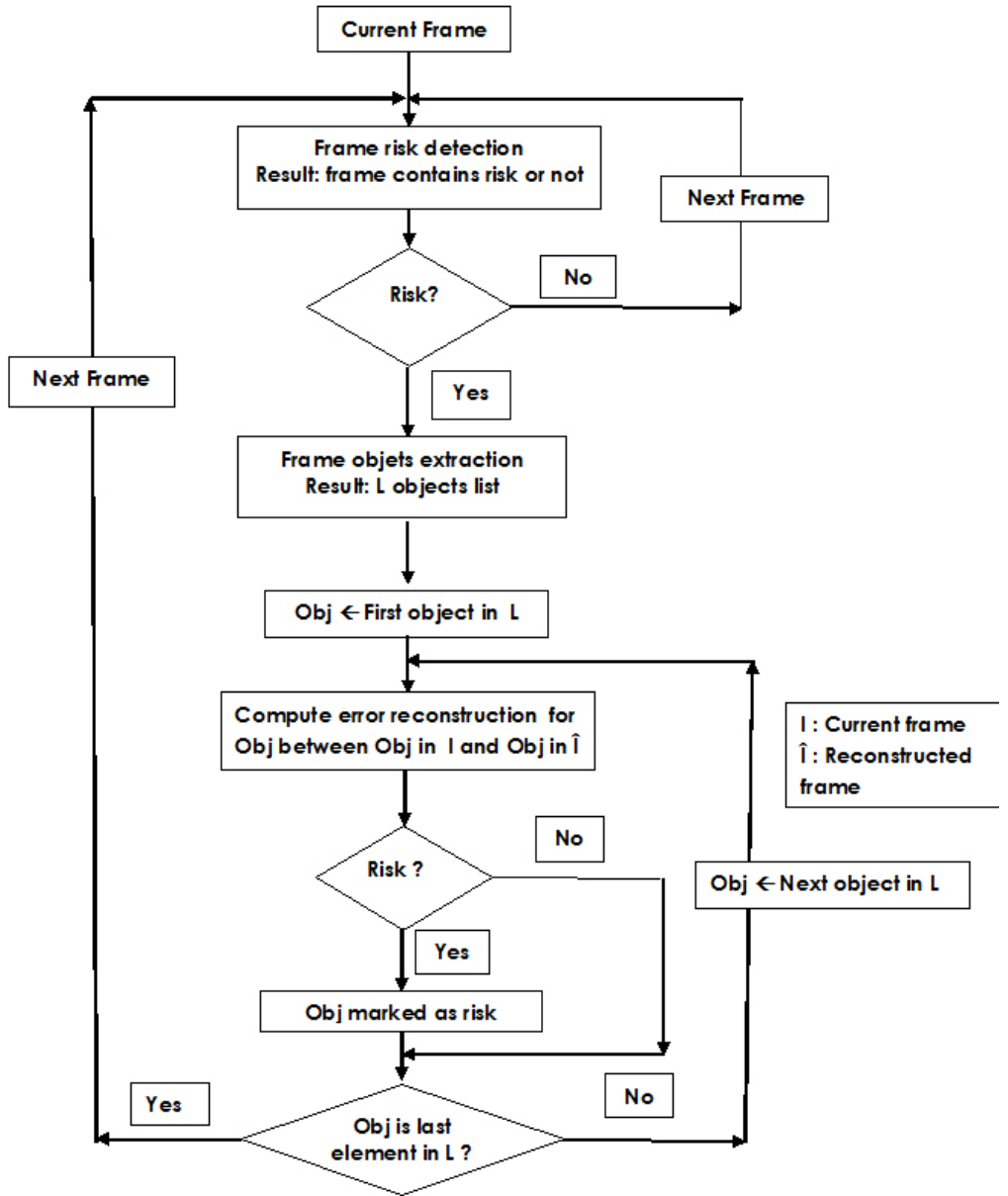


Figure 1: Overview of the proposed approach

2. Related Work

In this section, we present some of the previous works on anomaly detection based on deep learning. In [4], the authors have been introduced a cascade of auto-encoders, based on two novel cubic-patch-based anomaly detectors, the first one based on power of an auto-encoder on reconstructing an input video patch and the second is based on the power of sparse representation of an input video patch.

To analyze spatial and temporal information, the spatial temporal Convolutional Neural Networks have been used to capture spatial and temporal features encoded in frames video. The convolution is only performed in spatial temporal volumes of moving pixels to resolve the problem of local noise, and increase detection accuracy [10]. In [11], the authors have integrated the HOG and HOF motion features as input to the autoencoder to learn the reconstruction of regular motion in video frames, in the

reconstruction step, the higher error is classified as abnormal events. Hence in [5], the authors have proposed an unsupervised deep learning framework for anomalous event detection in complex video scenes based on a three-stream architecture (spatial, temporal and their joint representation) by employing the auto-encoder to learn the features, furthermore in [6], the authors built a novel model called spatio-temporal autoencoder (STAE), which learn video representation automatically using deep neural networks and extract features from both spatial and temporal dimensions by using 3-dimensional convolutions.

Whereas Generative Adversarial Nets (GANs) [12] takes as the first input the normal frames and produces corresponding optical-flow images. As the second input, GANs takes the real optical-flow of normal frames and outputs an appearance reconstruction. The abnormal areas are detected by computing local differences between the reconstructed appearance and motion and the normal frames and real optical-flow in order.

In [13], the authors have been presented a composite Conv-LSTM network able to reconstruct input frames, and predict future frames. The anomalous video segments are detected using a regularity evaluation algorithm at the model's output, video sequences containing normal events have a higher regularity score since they are similar to the data used to train the model, while sequences containing abnormal events have a lower regularity score.

3. Methodology

3.1. The LSTM Model

To resolve the problem of vanishing gradient and exploding in recurrent neural network (RNN), Hochreiter and al in [14], have been presented a neural network called Long Short-Term Memory (LSTM) for modelling long dependencies over time, and learn more semantic information and complex features. The LSTM unit is formulated with the equations as follow:

$$f_t = \sigma(W_f \otimes [h_{t-1}, x_t] + b_f) \quad (1)$$

$$i_t = \sigma(W_i \otimes [h_{t-1}, x_t] + b_i) \quad (2)$$

$$\hat{c}_t = \tanh(W_c \otimes [h_{t-1}, x_t] + b_c) \quad (3)$$

$$C_t = f_t \otimes C_{t-1} + i_t \otimes \hat{c}_t \quad (4)$$

$$o_t = \sigma(W_o \otimes [h_{t-1}, x_t] + b_o) \quad (5)$$

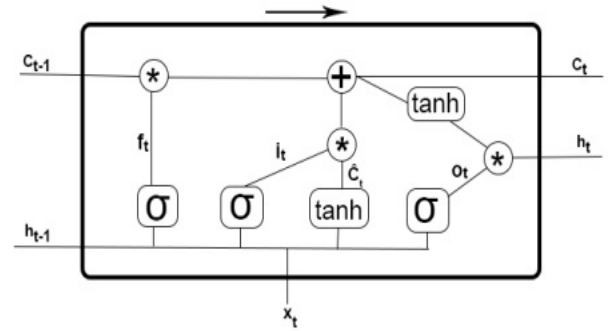
$$h_t = o_t \otimes \tanh(C_t) \quad (6)$$

Eq. (1) define the forget gate to reset the memory cell, Eq. (2) and Eq.(3) denotes the input and output gates, and essentially control the input and output of the memory cell. Eq. (4) represents the memory cell that prevents the problem of vanishing gradient and exploding problem in RNN.

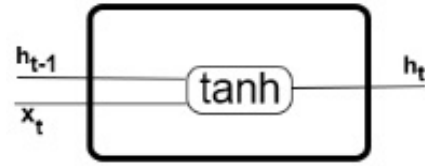
x_t denote the input at time t .

Figure 2 illustrates the difference between LSTM unit and RNN neural network.

To capture the most important semantic information and height level features in sequences video, we use Bidirectional Long Short-Term Memory (BLSTM) Networks architecture. The basic idea of BLSTM that the output of the forward and backward layer is combined at each time step to form one output, it's learned the past and future feature very fast and more accurate, the Fig. 3 illustrate the BLSTM architecture:



(a)



(b)

Figure 2: LSTM unit on the top and RNN unit on the bottom

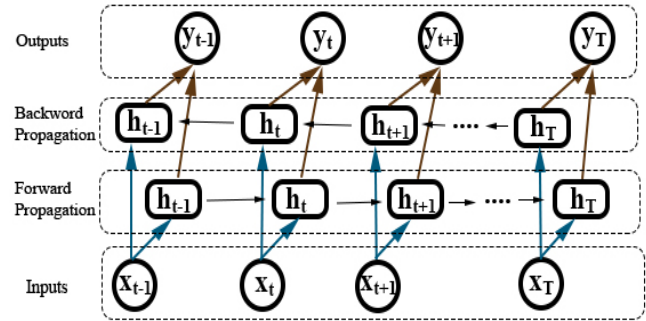


Figure 3: The architecture of BLSTM

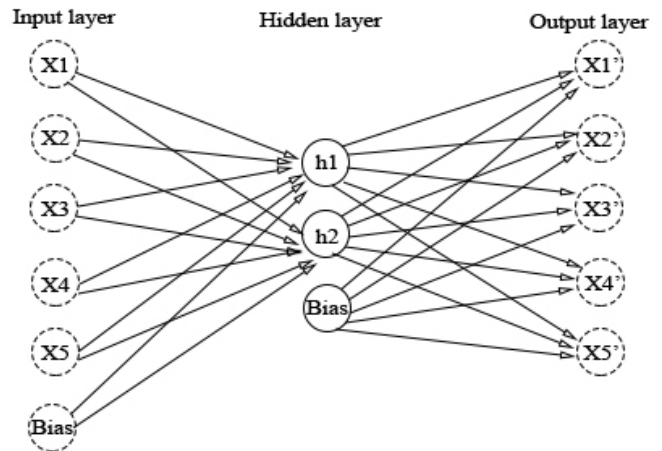


Figure 4: Autoencoder architecture

3.2. The Autoencoder

An autoencoder is an artificial neural network which performed in an unsupervised learning context, it can be seen as

the set of two components, an encoder and a decoder, the encoder that consists to reduce dimensionality of the input data, in order to represent them in a new space, the number of their input units layer is less than the output. Whereas the decoder, reconstruct data from the encoding by minimizing the reconstruction error between the encoder result (hidden layer) and the original inputs. The Fig.4 shows the autoencoder architecture.

In comparison with the PCA, the autoencoder is more efficient in case of non-linear transformation.

3.3. Faster C-RNN

Object detection is the process of finding objects in frames video, there is many approaches treating this area of research, on this paragraph we will focused on the neural network Faster R-CNN approach [14].

Faster R-CNN is a region based neural network, the first step use RPN fully convolutional network that take an image as input and outputs proposal regions with an objectness score for each one the second step integrate the Fast R-CNN network to classify those regions.

The RPN network is a complete convolutional network which slides on the feature map to indicate for each position whether there is an object or not, without taking into account the class of the object.

In order to have a system that is robust to translation and to scale, RPN uses an anchor-based algorithm. For each position of the sliding window on the feature map, 9 anchors are placed. The anchors are all centered on the sliding window, only their scale and ratio change (there are three scales and three ratios (1: 1, 2: 1 and 1: 2), which makes the 9 anchors. Each anchor is processed

through the convolutional layers of the RPN and the networks produce the probability that this anchor represents an object and potentially an offset to correct the dimensions of the anchor. Faster R-CNN generate region proposals directly in the network instead of using an external algorithm, that's make it faster, accurate and useful in real time detection.

4. The Proposed Method

For risk detection and localization in space public, we have presented an approach, which consists of recurrent neural network and convolution neural network. The proposed method focused on two pre-trained models. The Bidirectional LSTM Autoencoder [15], this model learns the normal behavior from normal training video frames. The risks on testing data are detected as behavior deviated from the normal characteristic learned, by performing a threshold on reconstruction error. The second model is the Fast RCNN [14], a pre-trained object detection model which used to detect and extract objects in video frames, the model was trained on the MS-COCO dataset, and the output of this model is the bounding box of objects detected in frames video.

The first model consists of two convolution layers followed by bidirectional LSTM layers in the Fig. 5. Inspired by [17], the reconstruction error of the frame t is defined as follow:

$$e(t) = \|I(t) - f_w(I(t))\|_2 \tag{7}$$

where f_w the learned model and the reconstruction error score is defined as follow:

$$RES = \frac{e(t) - e(t)_{min}}{e(t)_{max}} \tag{8}$$

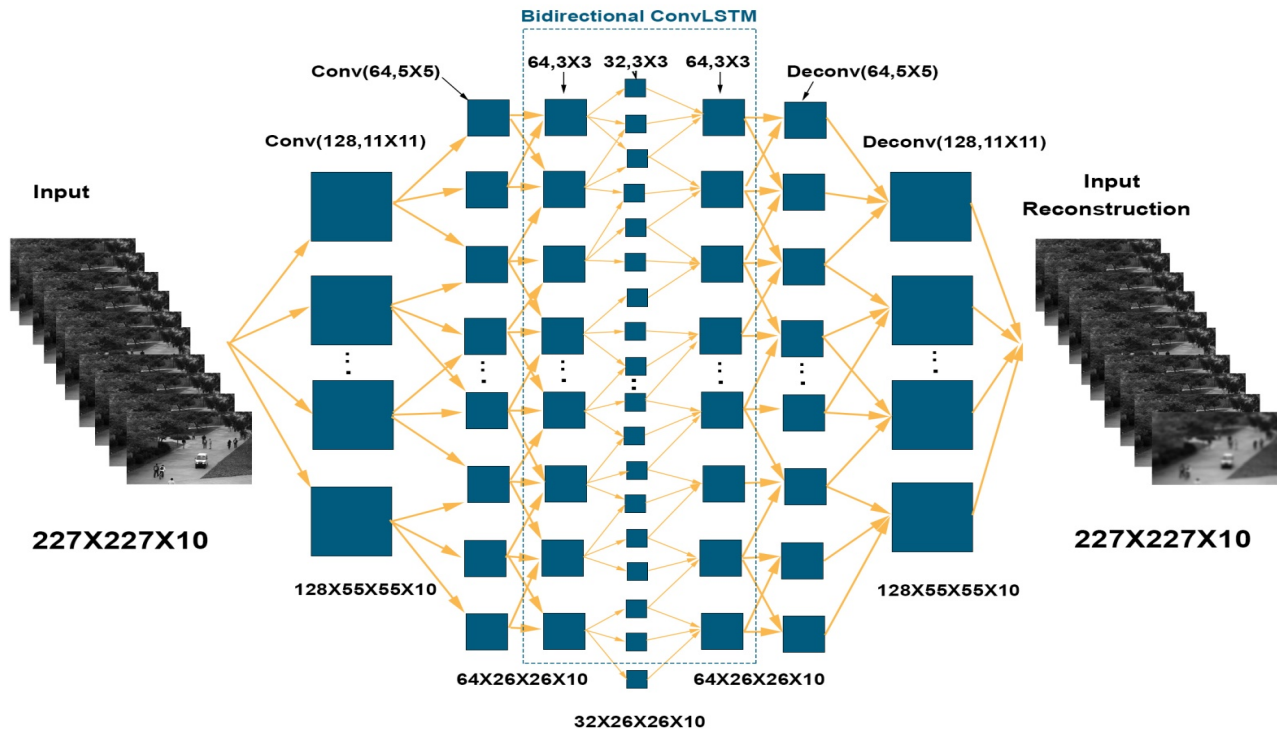


Figure 5: Architecture of the first model

To locate objects representing risk, we apply equation (8) to compute error construction of objects detected in frames with risk. In other words, the reconstruction error scores between bounding box of object in the original frame and the same bounding box in reconstructed frame must be greater than a threshold α to classify the object as abnormal, the experimental value of α is 0.005.

5. Experimental Results

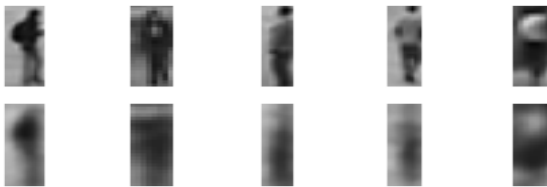
The model was trained and tested on the UCSD Ped1 datasets, the training video frames are divided into temporal cuboid of 10 frames, the resolution of each frame is 227x227, and the pixels value are normalized to take value between 0 and 1. The same processing data was performed to the testing video frames and the reconstructed error was computed for every cuboid of testing data.

The training videos contain videos without risk, and the testing videos contain both sequences video without risk and sequences video with risk. The computer used in this works has NVIDIA K80 with 12GB Memory tensorflow 2 python Library. To trains the model 40 epochs was used with a batch size of 4, a dropout of 20 percent, Adam optimizer with lr=1e-4, decay=1e-5 and epsilon=1e-6.

The Fig 6, 7 and 8 illustrates the result of our approach. (a) Represents the image difference between the original and the reconstructed frame, (b) shows objects of original frame, and the reconstructed objects, the object which has a reconstruction error greater than a threshold (0.005) is classified as abnormal object.



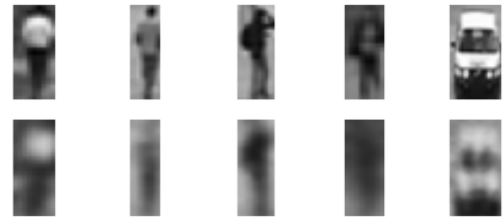
(a) Input frame on the left, reconstructed frame on the middle, and the difference between input and reconstructed frame on the right



(b) Extracted objects on the top, reconstructed frame on the bottom
Figure 6: The folder test 19 in UCSDped1, frame 150



(a) Input frame on the left, reconstructed frame on the middle, and the difference between input and reconstructed frame on the right



(b) extracted objects on the top, reconstructed frame on the bottom
Figure 7: The folder test 36 in UCSDped1, frame 120



(a) Input frame on the left, reconstructed frame on the middle, and the difference between input and reconstructed frame on the right



(b) Extracted objects on the top, reconstructed frame on the bottom
Figure 8: The folder test 19 in UCSDped1, frame 120

The Tab. 1 demonstrates the reconstructed error of detected objects. The object 5(car), where error greater than the threshold (0.005) in frame 2 was located as risk.

Table 1: Reconstruction Error for Detected Objects

	Obj1	Obj2	Obj3	Obj4	Obj5
Frame1	0.0037	0.004	0.0028	0.0032	0.0033
Frame 2	0.0043	0.003	0.0038	0.0041	0.0054
Frame 3	0.0035	0.0029	0.0028		

6. Conclusion

This work proposes an automated deep learning-based approach to detect and locate risks in public space; it exploits both convolution neural network and recurrent neural network in order to learn the spatial and temporal semantic features from frames video. Firstly, Bidirectional LSTM Autoencoder is used to detect abnormal frames, and then, a Fast R-CNN is applied to locate exactly objects representing risks in frames containing risks. Experiments were tested on UCSD datasets, and prove that the proposed approach detect and locate risks accurately. This approach performs automatically without any involvement, which makes it useful for real time video surveillance.

Finally, our future work consists to treat this idea in one convolutional neural network model.

References

- [1] M. Sabokrou, M. Fathy, M. Hoseini, "Video anomaly detection and localization based on the sparsity and reconstruction error of auto-encoder," *Electronic Letter*, **52**, 1122 – 1124, 2016, doi: 10.1049/el.2016.0440.
- [2] X. Mo, V. Monga, R. Bala, and Z. Fan, "Adaptive Sparse Representations for Video Anomaly Detection," *Circuits Syst. Video Technol. IEEE Trans*, **1**, 2013, doi: 10.1109/TCSVT.2013.2280061.
- [3] C. Li, Z. Han, Q. Ye and J. Jiao, "Abnormal behavior detection via sparse reconstruction analysis of trajectory," *Proceedings-6th International Conference on Image and Graphics, ICIG*, 807–810, 2011, doi: 10.1109/ICIG.2011.104
- [4] Hu, Jingtao, E. Zhu, S. Wang, X. Liu, X. Guo, J. Yin, "An Efficient and Robust Unsupervised Anomaly Detection Method Using Ensemble Random Projection in Surveillance Videos," *Sensors* **19**, 4145, 2019, doi:10.3390/s19194145.
- [5] D. Xu, E. Ricci, Y. Yan, J. Song, N. Sebe. "Learning deep representations of appearance and motion for anomalous event detection," *BMVC*, 1-12, 2015, doi: 10.5244/c.29.8.
- [6] Y. Zhao, Y. Deng, B. Shen, C. Liu, Y. Lu, H. Hua, X.S. "Spatio-temporal autoencoder for video anomaly detection," *Proceedings of the 25th ACM International Conference on Multimedia, Silicon Valley, CA, USA*, 23–27, 1933–1941, 2017, doi: 10.1145/3123266.3123451.
- [7] M. Welling, D. Kingma, "Stochastic gradient vb and the variational auto-encoder," *Proceedings of the 2nd International Conference on Learning Representations (ICLR)*, 2014.
- [8] J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," *Advances in neural information processing system*, 2672–2680, 2014, DOI:10.1145/3422622.
- [9] S. Hochreiter, S., Schmidhuber, J. "Long short-term memory," *Neural computation*, **9**(8), 1735-1780, 1997, doi:10.1162/neco.1997.9.8.1735.
- [10] M. Sabokrou, M. Fayyaz, M. Fathy, Z. Moayed, R. Klette, "Spatial temporal convolutional neural networks for anomaly detection and localization in crowded scenes," *Signal Processing: Image Communication*, **47**, 358–368, September 2016, doi:10.1016/j.image.2016.06.007.
- [11] M. Hasan, M., Choi, J., Neumann, J., Roy-Chowdhury, A.K., Davis, L. S. "Learning temporal regularity in video sequences," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 733–742, (June 2016), doi:10.1109/CVPR.2016.86.
- [12] M. Ravanbakhsh, E. Sangineto, M. Nabi, N. Sebe, "Abnormal Event Detection in Videos using Generative Adversarial Nets," *In Proceedings of the IEEE International Conference on Image Processing (ICIP)*, 17–20, September 2017, doi:10.1109/APSIPAASC47483.2019.9023261.
- [13] J. Medel, A. Savakis, "Anomaly detection in video using predictive convolutional long short-term memory networks," 2016.
- [14] S. Ren, K. He, R. Girshick, and J. Sun. "Faster R-CNN: Towards realtime object detection with region proposal networks," *Advances in Neural Information Processing Systems*, 91–99, 2015, doi:10.1109/TPAMI.2016.2577031.
- [15] K. Boulfrifi, K. Housni, "Bidirectional Convolutional LSTM Autoencoder for Risk Detection," *International Journal of Advanced Trends in Computer Science and Engineering*, **9**, 85-89, 2020, doi:10.30534/ijatcse/2020/241952020.
- [16] B. T. Morris and M. M. Trivedi, "A survey of vision-based trajectory learning and analysis for surveillance," *IEEE Transactions on Circuits and Systems for Video Technology*, **18**, 1114–1127, 2008, doi:10.1109/TCSVT.2008.927109.
- [17] M. Hasan, J. Choi, J. Neumann, K. Roy-Chowdhury, S. Davis "Learning temporal regularity in video sequences," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 733–742, June 2016, doi:10.1109/CVPR.2016.86.

Spatio-Temporal Analysis of Drought over the Guinean-Sudano Ecological Zone, Northern Nigeria

Gloria Langyit Dasat^{1,*}, Dickson Dare Ajayi², Ademola Aremu³

¹Pan African University, Institute of Life and Earth Sciences (PAULESI), University of Ibadan, Ibadan, 200284, Nigeria

²University of Ibadan, Department of Geography, Ibadan, 200284, Nigeria

³University of Ibadan, Department of Agricultural and Environmental Engineering, Ibadan, 200284, Nigeria

ARTICLE INFO

Article history:

Received: 07 October, 2021

Accepted: 05 December, 2021

Online: 30 December, 2021

Keywords:

Drought

SPI

NDVI

Precipitation

Climatic variations

ABSTRACT

Drought has become more common and severe in recent decades due to increased global changes in climatic conditions and anthropogenic activities thereby analyzing drought at both regional and local levels is crucial. And so, this study aims at assessing drought risk in Northern Nigeria using NDVI and SPI. Data used included satellite images (Landsat OLI-TIRS, ETM+, TM, and MSS) meteorological data, and gridded precipitation datasets (1975 – 2016) from the Climate Research Unit (CRU). The drought index was calculated using the Standardized Precipitation Index (SPI) for both short-term and long-term timescales with emphasis on drought magnitude, frequency, duration. The Mann-Kendall trend analysis was utilized on precipitation and SPI to investigate the trend changes. Results indicated that SPI dynamics over the region at varying timescales suggest the unpredictability of drought with longest drought duration were the severe and extreme drought ($SPI \leq -1.5$) was peaked at -3.12 severity that lasted 15 months between 2007 and 2008. The worst drought magnitudes were more eminent in the longer timescale (SPI12 months) with its peak of -70.8 that lasted 60 months (5 years) between 1984 and 1986. Similarly, drought intensity was observed between the mildly dry with SPI -0.47 (1986 - 1991) and severely dry with SPI -1.79 (1983 - 1984) this further revealed that spatially drought risk is increasing southward. NDVI values obtained demonstrated a consistent decrease with a better state in 1975 which ranged between -0.31 and 0.63. 2016 recorded lower values with a range between -0.21 and 0.47. NDVI minimum and maximum values were between -0.48 and 0.63 between 1975 and 2016, while in 2016 highest of 0.47 was observed as the least maximum value with 1975 recording the highest value at 0.63. The results demonstrated that drought conditions within the Guinean-Sudano ecological zone of northern Nigeria are evident due to rapid unpredictable climatic changes and anthropogenic activities.

1. Introduction

Drought is among the most destructive natural disasters in the world. It often causes far-reaching damages to natural ecosystems, results in water shortages, and increased risks of surface temperatures and desertification amongst others. Contrary to other extreme phenomena (such as floods and tornadoes), drought results from below-normal precipitation combined with warm temperatures over months to years [1] with variations in drought frequency and severity often of enormous concern. The primary cause of most droughts can be associated with fluctuating

atmospheric patterns such as low rainfall, cloud cover, and greater evaporation rates [1] which are exacerbated by anthropogenic activities.

Droughts are universally divided into four types: meteorological drought (abnormally low precipitation), hydrological drought (precipitation shortages that impact the surface and groundwater levels), agricultural drought (decreased soil moisture that impacts crops), and socio-economic drought (decreased water supply that affects people and supply of goods) [2,3]. However, the meteorological drought influences both agricultural and hydrological droughts to occur [4,5] Droughts can also be

*Corresponding Author: Gloria Langyit DASAT, glowj480@gmail.com

classified by length, intensity, regional distribution, frequency, and onset rate [6].

Droughts occur worldwide but their frequency, duration, and severity differ by location. According to Trenberth, the size, intensity, frequency, and spatial coverage of the dry area has increased since the 1970s across the globe particularly in Africa, East, and South Asia, eastern Australia, Southern Europe, Northern and Southern America, most of Alaska, and Western Canada [7]. There were about 669 incidences of drought reported worldwide between 1960 and 2016, with approximately 2.2 million deaths that affected over 2.6 billion people leading to an estimated economic impact of 146 billion US Dollars [8]. Unlike developed countries, underdeveloped and developing countries are the worst hit by the impact of climate change due to physical, social, and economic factors [9].

The extreme weather and climate events including droughts and floods have significant impacts on the economy, natural resources, and human health in Africa [9]. In this regard, the negative impacts of drought are more severe in sub-Saharan Africa, where rain-fed farming encompasses 95% of all agriculture [10,11]. Also, crop production and food security are mainly threatened by drought where the effects are inter-temporal and long-lasting [12,13]. Africa stands top with a total of 289 reported drought events that had about 700 thousand deaths and above 414 million people unsettled between 1960 and 2015 [14]. This threat is expected to further exacerbate the existing widespread poverty and food insecurity with the situation similar within other regions of sub-Saharan Africa [14–16]. In West Africa, [17] revealed there is about a 40% decline in annual precipitation total from the year 1968–1990 as compared with the 30 years between 1931 and 1960. So, frequent drought occurrences within the West African region have caused famine and are threatening the African savanna regions while making the populace highly vulnerable to drought [18].

Likewise, the situation in Nigeria is not any different, drought occurrences and reoccurrences have been reported in Nigeria for decades [19] that have grown and gradually migrated southward particularly within Nigeria's Guinean-Sudano ecological zone [20,21]. This indicates that drought risks will continue to expose residents in this area with its impacts comprising mass starvation, famine, destruction of ecosystems, socio-economic instability and poverty [22] particularly in areas where agriculture is the mainstay of the economy as the case of the Guinean-Sudano zone in northern Nigeria. Therefore, it is essential that regular drought monitoring is conducted to provide preventive policy aid for long and medium-term planning.

From the 20th century, scientific studies had adopted climatic (temperature and precipitation) and hydrological (soil moisture and streamflow) indicators as main input towards the generation of indices for quantitative modeling of drought severity [23] as depicted in the Standardized Precipitation Index (SPI) proposed by [24]. However, further advances in the study of drought between the 20th and 21st centuries led to the acceptance of over 150 indices for drought studies [25] several regions with different climatic conditions. Thus, many drought studies have depended on the availability of data from several remote sensing systems applied in a variety of methodologies. These include NDVI-based

conceptualizations like soil Adjusted Vegetation Index (SAVI) [26] and more advanced methods now include indices such as Drought Severity Index (DSI) proposed by [27] applied in [28]. In Nigeria, not until recently, statistical techniques such as SPI and SPEI were the most used indices in drought assessment. Nonetheless, the introduction of Geo-information Systems (GIS) and remote sensing that has championed the adoption of integrated methods in the monitoring of drought has proven to provide better and more robust options to mitigating drought. Lately, [21,29,30] amongst others have employed combined techniques through satellite-based applications such as NDVI and statistical techniques like SPI/ SPEI in drought assessment over the Sudano-Sahelian region and North Central of Nigeria. The integrated techniques were utilized in examining vegetation health and response to drought and in predicting the occurrence duration and severity of drought events.

Nevertheless, there have been little or no reports on drought vulnerability and risk even though it is virtually common knowledge to land users such as farmers within the Guinean-Sudano region which comprises 5 ecological zones namely: Derived, Guinea, Sudan, and Sahel Savannas, and Jos Plateau. Therefore, this study aims to analyze drought in northern Nigeria using NDVI and SPI indices for effective prevention and mitigation towards sustainable development.

2. Methodology

2.1. Study area

From south to north, the ecological zones of Nigeria are Mangrove Swamp and Coastal Vegetation, Freshwater Swamp Forest, Lowland Rain Forest, Derived Savanna, Guinea Savanna, Sudan Savanna, and Sahel Savanna [31]. A few mountainous areas are found on the Jos Plateau, Adamawa, Taraba, and northern part of Cross-River State (Figure 1), but for this study, 4 States (Bauchi, Gombe, Plateau, and Nassarawa), as well as the Federal Capital Territory (FCT) were used to illustrate the climatic variation across major ecological zones in northern Nigeria.

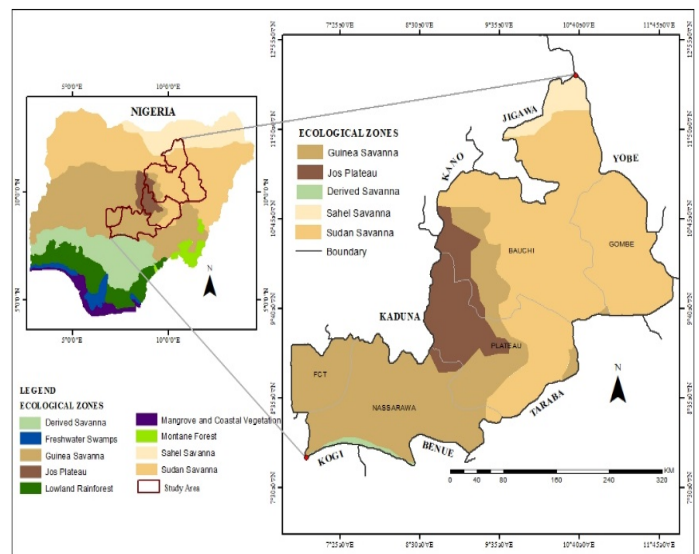


Figure 1: Ecological Zones in Nigeria

The study area lies between latitude 9°15' and 12° 30' North and longitude 6° 45' and 12° 00' East. It falls within the Sahel, Sudan, Guinea, and the Jos Plateau ecological zones with a little portion of derived Savanna southward area covering an area of 133,232 square kilometers (Km²) accounting for 14.42 percent of Nigeria's total landmass. The study area is bordered by Yobe, Jigawa, and Kano States to the north; Borno and Adamawa states to the east; Kaduna and Niger States to the west; and Kogi, Benue, and Taraba States to the south (Figure 2). Northern Nigeria constitutes about 80% of the vegetative zones, records lower rainfall, shorter wet season and they make up the Savanna land [23].

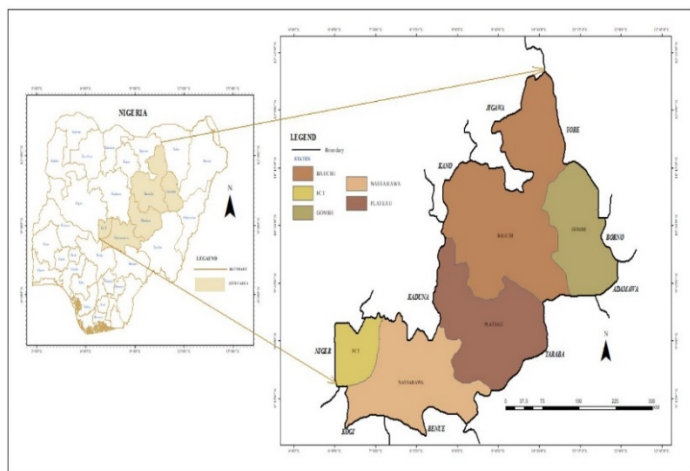


Figure 2: The study area, Bauchi, Gombe, Plateau, and Nassarawa States including FCT

Rainfall is mostly seasonal and extremely unstable through the years with two distinct seasons (dry and wet). The dry season covers up to a period of about 7-8 months, from October to April/May, and the wet season about 4-5 months, from May to September with annual rainfalls between 600 – 1500 mm [21,31]. The study area consists of predominantly 2 drainage systems: the River Niger drainage system consists of the River Niger and its tributaries, prominent among which are the: Benue, Sokoto-Rima, Kaduna, Gongola; and Lake Chad inland drainage system that draws the following in-flowing rivers from Nigeria: the Komadougou-Yobe (with headstreams including Hadejia, Jama'are, and Misau) and the Yedseram [32].

Northern Nigeria is rich in agricultural produce nevertheless the enormous inter-annual inconsistency of rainfall undermines it to intermittent dry spells which occasionally result in the widespread drought that can inflict grievous socio-economic constraints [23]. Irrigation is generally practiced and crops include grains (rice, wheat, soybeans, beans, maize, millet, amongst others), tuber crops (yam and cassava) mainly from the Guinea-savanna zone [22]. Cotton and groundnut, which are economic crops are found in this region mainly in the midland zone and Sahel zones [21]. The Savanna landforms are the tremendous natural habitat for a large number of grazing livestock such as cattle, goats, horses, sheep, camels, and donkeys [15]. The Geology to which the region belongs mostly in the basement complex rocks formation made of older and younger granites formed during the geologic period of the pre-cambrea era [15]. The region had a total population of 13,499,215 in 2006 with about 32.8% growth to 20,082,239 in

2016 [33] with a majority of the population living below the poverty line with income below 361 USD per annum as poverty data reported an average of 55% poverty rate across the region [34]. Hence the dire need to assess drought conditions in northern Nigeria and particularly across the ecological zones.

2.2. Data and Methods

This paper investigates the spatiotemporal variability of drought incidence in the Guinean-Sudano zone of northern Nigeria between 1975 and 2016. The study is based on an analysis of a calculated space-time field of the Standardized Precipitation Index (SPI), as proposed by [24] combined with NDVI to provide a robust assessment of drought within the Guinean-Sudano zone of northern Nigeria.

SPI offers several advantages which were highlighted by [35] as a starting point for meteorological drought monitoring. First, SPI is probability-based, and therefore it is easy to interpret and allows for comparison across space and time [4]. This led to its endorsement through the Lincoln Declaration on Drought Indices as the standard for evaluating the occurrence of meteorological drought [36]. However, the time series should have a minimum of 30 years of data, even when missing data are accounted for [15,37] which has influenced our 45-year temporal assessment. For these reasons, SPI has been used in drought assessment and monitoring in various climates across the globe and particularly in Africa where data availability is an issue.

The impact of drought on society and agriculture is a dire situation that requires reliable indices that identify the geospatial and temporal aspects of drought occurrences and their intensity to determine the impact and also for decision-making and research [38]. Therefore, development and advancements in space expertise on topical issues such as drought detection, monitoring, and assessment have been achieved with the formulation of plans to deal with this slow-onset disaster [4]. With the help of environmental satellites, drought can be recognized 4-6 weeks earlier than before and defined precisely, and its impact on agriculture can be diagnosed far in advance of harvest, which is the most vital for global food security and trade [39].

Furthermore, drought indices calculated at a single area are only valid at that location. As a result, one of the key disadvantages of climate-based drought analysis is their lack of spatial precision, as well as the fact that they rely on data collected at weather stations, which can be sparsely dispersed, lowering the dependability of the drought [4,39,40]. Hence, satellite-derived drought indicators based on satellite-derived surface data have become frequently used in drought research [28]. NDVI is one of the most used vegetation indicators.

The Normalized Difference Vegetation Index (NDVI) is a measure of vegetation health and density [41]. The formula for NDVI is: $NDVI = (NIR - RED) / (NIR + RED)$. The NIR and RED are the reflectance values of near-infrared and red bands which vary from +1 to -1. NDVI is a good indicator of green biomass, leaf area index, and patterns of production [41]. Remotely sensed data have been used to evaluate climatic and environmental changes at regional and global scales [42]. It can be used for the description of land cover, vegetation classification, and health but is also effective for monitoring rainfall and drought [43]. The sensitivity

of NDVI values to rainfall fluctuations varies regionally, owing to the effects of variation in features such as vegetation type and soil background [44].

2.3. Datasets and their characteristics

Secondary data were utilized for this study with primary sources from field observations and GPS measurements. Landsat satellite images of Bauchi, Gombe, Plateau, and Nassarawa States including FCT were acquired for 1975, 1990, 2005, and 2016 obtained from the United States Geological Survey (USGS) Earth Explorer platform.

In addition, rainfall data were acquired from Nigerian Meteorological Agency (NIMET) for 10 weather stations using the monthly rainfall data for the period of forty-one years (1975 - 2016). Though gauge measurements provide high-quality estimates of rainfall and other weather parameters, their availability is a major constraint especially in sparsely populated areas such as drylands [45]. Therefore, gridded data sets were utilized to supplement for no data area where climatic data for 5 States over the study area to allow interpolation were obtained from data developed by the University of East Anglia at a spatial resolution of 0.5x0.5 degrees covering a temporal range of 1975 to 2016 [45]. These datasets are developed using models and algorithms; integrating ground observations and satellite recorded imageries in multiple wavelengths [45] have provided an alternative to gauge data at a broad scale with varying Spatio-temporal resolutions thereby helping overcome the foregoing constraints by providing more homogenous spatial and temporal coverage for most locations across the world including un-gauged regions [45].

2.4. Determination of SPI and NDVI

The Standardized Precipitation Index (SPI) developed by [24] is a popular index that is used to characterize drought at varying time scales. SPI is computed by finding the probability of precipitation data of a given frequency over an area [11,25]. Positive SPI values connote wet conditions; the higher the SPI, the more unusually wet a period is [25]. Negative SPI values signify dry conditions; lower SPI implies the more abnormally dry a period is (Table 1). This was computed using SPI Generator created by [46].

Table 1: Standard Precipitation Index (SPI)

SPI VALUES	DESCRIPTION
2>	Extremely wet
1.5 - 1.99	Very wet
1.0 - 1.49	Moderately wet
0 - 1.0	Mildly wet
-1 to 0	Mildly drought
-1.5 to -1.0	Moderately drought
-2 to -1.5	Severe drought
-2 <	Extreme drought

The drought magnitude was obtained as the aggregate SPI over the drought months taken as a positive value [39]. The intensity (drought severity) was computed as the magnitude divided by drought duration [11,39].

Since the input parameter for SPI computation is precipitation, trend analysis of precipitation was done to study the local changes in climate. The Mann-Kendall non-parametric test was adopted in this work to assess the trends in precipitation and SPI aimed at testing the statistical distribution of the data records. There are several approaches to trend analyses available but the Mann-Kendall was the most preferred and used test for assessing the trends in hydro-climatic studies [38,47,48].

SPI values for 15 locations under the periods between 1975 and 2016 were interpolated using the Inverse Distance Weighted Technique (IDW) and according to [49] interpolation provided optimal areal estimates in any given situation and is applicable both for drought and flood. The interpolation maps are thus been reclassified into drought severity classes. The SPI map was further integrated using the cell statistic tool in ArcGIS to generate a drought risk map.

In calculation of SPI, we adopted methods

$$SPI = \frac{x_i - \bar{x}}{\sigma} \tag{1}$$

where: x_i = observed precipitation value of the selected period during a particular year;
 \bar{x} and σ = the mean and the standard deviation respectively over the selected period.

The criteria employed by [4,24] were utilized in the quantification and classification of the SPI drought severity status of the study. Drought classification with monthly $SPI \leq -1.00$ were considered to be as drought commencement and cessation since mild drought (SPI between - 0.99 and 0.99) showed only a slight variation from the normal rainfall distribution [50–52]. Therefore, using the SPI_SL_6.exe algorithm, the severity of the drought was generated and plotted as graphs and maps. Temporal drought event characteristics (frequency, magnitude, intensity, and severity status) were computed using one of the SPI analysis parameters.

Drought frequency is the return period between drought events that have severity threshold values of $SPI \leq -1$ [52]. The magnitude of drought event tallies with the collective water dearth during the drought period below a certain threshold (SPI -values ≥ -1) and drought intensity (DI) is the ratio between the drought magnitude and the duration of the event [50,52] computed as:

Drought magnitude

$$(DM) = - \sum_{i=1}^n SPI \tag{2}$$

$$\text{Drought intensity (Di)} = \frac{DM}{D_D} \tag{3}$$

where: D_M = drought magnitude,
 n = number of months with drought event at j timescale; and
 D_D = drought duration.

Drought severity (DS) must not be muddled with intensity, which refers to the drought event's lowest SPI number [53]. The number of droughts per 100 years for several timescales of the rainy months and seasons (SPI: 3-, 6-, and 12-months) in the study as assessed by [51,52,54] were:

$$N_{i_{100}} = \frac{N_i}{ix_n} \times 100 \quad (4)$$

where: $N_{i_{100}}$ = the number of droughts for a timescale i in 100 years;

N_i = the number of years with droughts for a timescale i in the n -year set;

i = the timescales (3-, 6- and 12- months), and n = the number of years in the data set (45 years).

The MK (Mann Kendall) trend test [55,56] has been computed to substantiate the statistical significance (increasing and decreasing trends) for wet months and seasons of SPI: 3-, 6- and 12- months timescales. The relative strength of the MK trend test in time series analysis has been quantified using Sen's Slope Estimator (SSE) [57]. The analysis was carried out by the Addinsoft XLSTAT as applied by [58].

The Mann-Kendall trend test (S) is calculated as:

$$S = \sum_{i=1}^{N-1} \sum_{j=i+1}^N \text{sgn}(x_j - x_i) \quad (5)$$

where x_i is the time series ranked from $i=1, 2, \dots, n-1$ and x_j from $j = i+1, 2, \dots, n$.

All the data values are taken as reference point to which comparison is done with the rest of the data values x_j such;

$$\text{sgn}(x_j - x_i) = \begin{cases} +1 & > (x_j - x_i) \\ 0 & = (x_j - x_i) \\ -1 & < (x_j - x_i) \end{cases} \quad (6)$$

The statistics of variance is given as

$$\text{Var}(S) = \frac{n(n-1)(2n+5) - \sum_{i=1}^m t_i(t_i-1)(2i+5)}{18} \quad (7)$$

where t_i represents number of ties up to sample value i .

Sen's Slope Trend Estimator (SSE)

The Sen's trend estimator test was described by [57] and the magnitude of the trend is given by

$$T_i = \frac{x_j - x_k}{j - k} \quad (8)$$

where x_j and x_k are considered as data points j and k ($j > k$) compatibility. The median of these N values of T_i is represented as Sen's estimator of slope which is given as

$$Q_i = \begin{cases} T_{\frac{N+1}{2}} & N \text{ is odd} \\ \frac{1}{2} \left(T_{\frac{N}{2}} + T_{\frac{N+2}{2}} \right) & N \text{ is even} \end{cases} \quad (9)$$

Positive and negative figures of Q_i represent upward (increasing) and downward (decreasing) trends, respectively. To measure the Spatio-temporal occurrence of drought over the area under study, the 3-, 6- and 12-month SPI were used. These time lags are enough for drought frequency and intensity assessment [39,40]. The SPI was computed on a monthly scale so that the consistency of drought duration and intensity can be determined according to table 2.

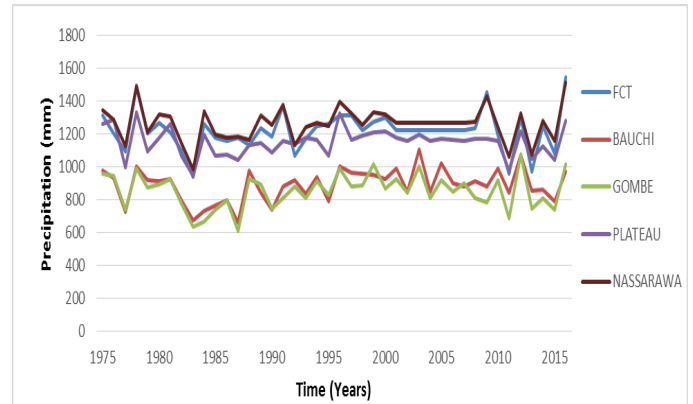
In addition, Normalized Difference Vegetation Index (NDVI) was computed. The NDVI datasets were obtained from the Landsat OLI- TIRS, ETM+, TM, and MSS with NDVI universally defined as:

$$\text{NDVI} = \frac{\text{NIR} - \text{RED}}{\text{NIR} + \text{RED}} \quad (10)$$

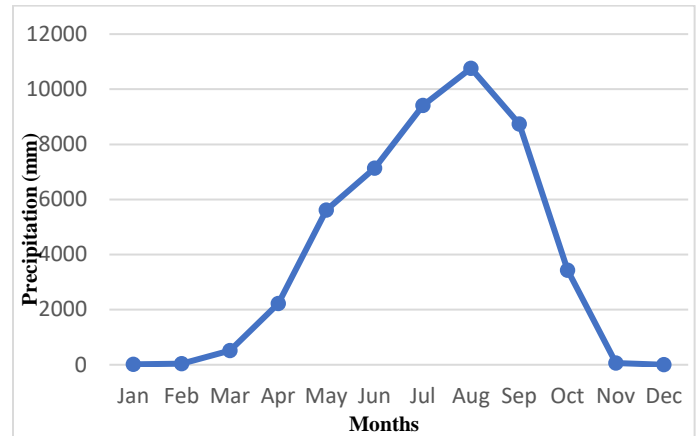
where NIR is the Near Infra-red band and Red is the red band in the electromagnetic spectrum. Furthermore, a correlation and regression analysis were carried out to observe the relationship between SPI and Vegetation cover over the study area.

3. Results

The Standardized Precipitation Index (SPI) developed [24] is a popular index that is used to characterize drought at different time scales.



(a) Annual precipitation over the ecological zones



(b) Cumulative seasonal precipitation

Figure 3: Unpredictability of precipitation over the ecological zones

3.1. The pattern of precipitation and drought events;

The characteristics of precipitation over the ecological zones Figure 3a and b revealed that long precipitation occurs between June and September with its peak in August while the short precipitation occurs between November and March. Rainfall variability was eminent year to year across the zone with states within similar ecological zones exhibiting similar precipitation patterns. Bauchi and Gombe states within the Sahel and Sudan exhibited similar rainfall dynamics with the least record of 661mm/yr and 614mm/yr in 1987 respectively while peak annual precipitation of 1547mm/year was recorded in FCT, 2016. Correspondingly, Plateau and Nassarawa states including FCT showed similar rainfall patterns within the Guinean savanna and Jos Plateau ecological zones.

Table 2: Drought Characteristics over the ecological zones

LOCATIONS	TIME SCALE	YEAR	DURATION (months)	MAGNITUDE	INTENSITY	PEAK
ABUJA	SPI 3	2013-2014	11	-11.83	-1.0755	-2.95
	SPI 6	2013-2014	11	-12.99	-1.1809	-2.85
	SPI 12	1986- 1991	59	-27.59	-0.4676	-2.27
BAUCHI	SPI 3	1977- 1978	9	-8.78	-0.9756	-2.96
	SPI 6	1977- 1978	12	-13.14	-1.0950	-2.85
	SPI 12	1983- 1988	60	-67.79	-1.1298	-2.15
GOMBE	SPI 3	1977- 1978	9	-7.40	-0.8222	-2.89
	SPI 6	1984- 1986	23	-19.46	-0.8461	-2.85
	SPI 12	1983- 1984	60	-70.80	-1.1800	-1.97
PLATEAU	SPI 3	1977-1978	11	-6.84	-0.6218	-3.06
		2013- 2014	11	-8.83	-0.8027	
	SPI 6	2007- 2008	15	-16.38	-1.0920	-3.12
	SPI 12	1986- 1988	25	-19.80	-0.7920	-2.35
		1994- 1996	25	-17.26	-0.6904	
NASSARAWA	SPI 3	1982- 1984	17	-22.03	-1.2959	-3.02
	SPI 6	1983- 1984	15	-26.80	-1.7867	-2.99
	SPI 12	2010- 2012	24	-28.65	-1.1938	-2.37

Hence, seasonal variation over the ecological zones (Figure 3b.) further depicted impacts these fluctuations could cause to food production as the region is majorly agrarian and dependent on rain-fed agriculture.

Characteristics of drought events (duration, magnitude, and intensity) Table 2 shows the observed drought duration, magnitude, and intensity of the slight, large, and long-term rainy epochs and spells as computed using SPI. SPI at various timescales of 3, 6, and 12 months were calculated with shorter timescales (3 and 6 months) exhibiting high variability in both wet and dry spells however, the longer timescale at 12 months witnessed lesser variability (Figure 4) which is an extraction of SPI over the region. SPI dynamics over the region at varying timescales suggest the unpredictability of drought. Periods with the longest drought duration were within the severe and extreme drought ($SPI \leq -1.5$) which was peaked at -3.12 severity that lasted 15 months between 2007 and 2008. The worst drought magnitudes were more eminent in the longer timescale (12 months) with its peak of -70.8 that lasted 60 months (5 years) between 1984 and 1986. Similarly, drought intensity was observed between the mildly dry with SPI -0.47 (1986 - 1991) and severely dry with SPI -1.79 (1983 - 1984). Seasonally, a visibly slight difference was observed in the drought magnitude among the SPI3/6 months (short-term) and SPI12-months (long-term)

seasons. The drought magnitude, duration, and intensity were diverse from mildly dry to severe drought irrespective of the timescales or ecological zones with the worst intensity condition during SPI6 (short-term) - 1.79 and SPI12 (long-term) - 1.19 respectively.

Plateau and Nassarawa states experienced the worst drought conditions despite being situated within the Jos Plateau and Guinea savanna ecological zones where conditions are assumed to be within the sub-humid region compared to drylands. This further reaffirms that drought incidences have grown and gradually migrated southward with very unstable patterns.

3.2. Spatial and Temporal representation of SPI over the Guinean-Sudan ecological zone

Figure 5 shows the spatial illustration of SPI from 1975 to 2016 over the Guinean-sudano ecological zone of Nigeria. Results revealed that 1975 recorded the highest precipitation than the 1990, 2005, and 2016 hydrological years. It is crucial to note that most ecological zones that recorded high precipitation in some years also recorded the least in other years thus, reaffirming the fact that precipitation over the region fluctuates and drought is not peculiar to one or a particular region. Similarly, the unstable pattern of precipitation observed indicates its randomness which imposes a grievous burden on agriculture, water resources, and livelihood within the region.

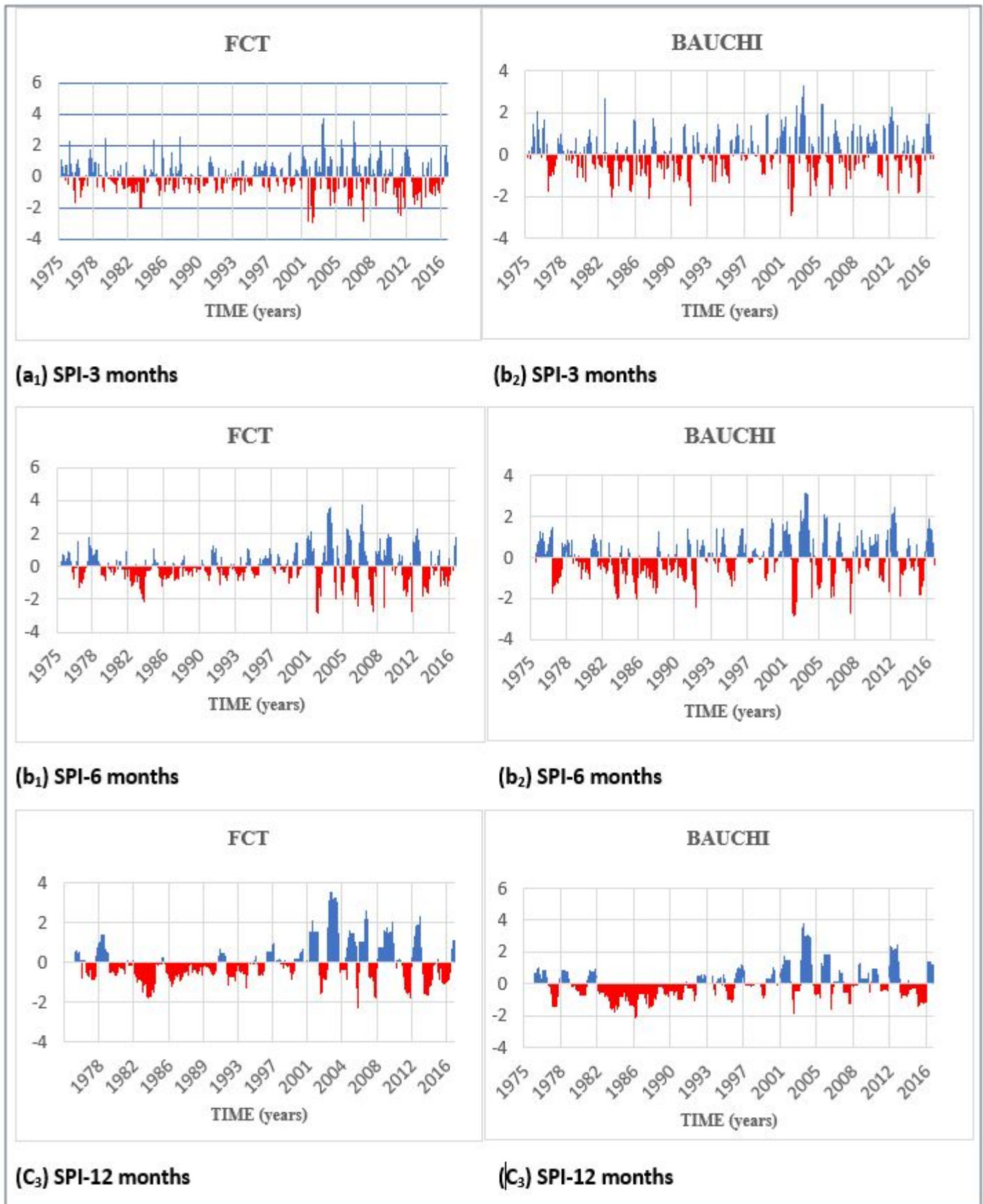


Figure 4: Extraction of SPI at different timescales

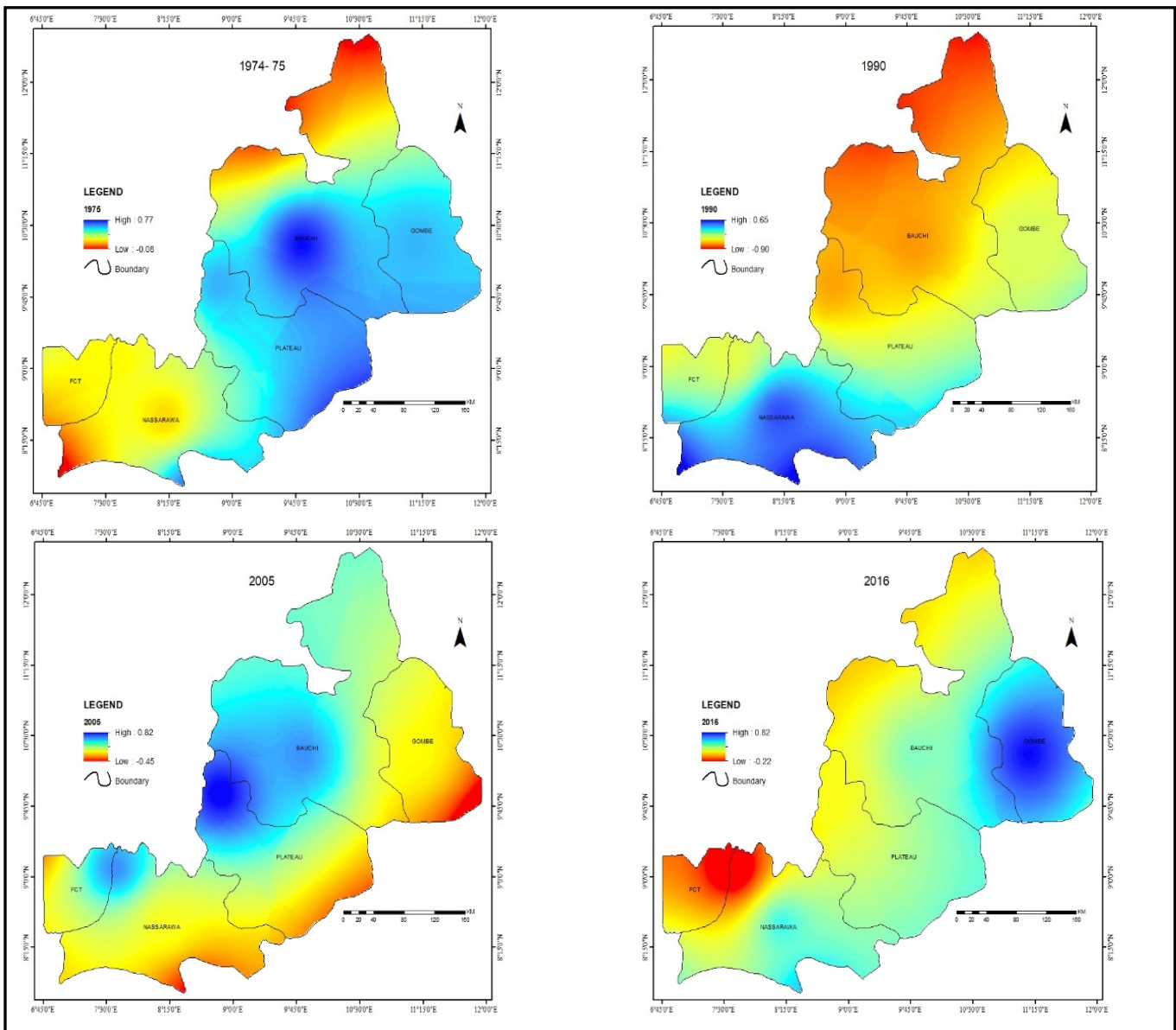


Figure 5: Spatial representation of SPI for different hydrological years over the region

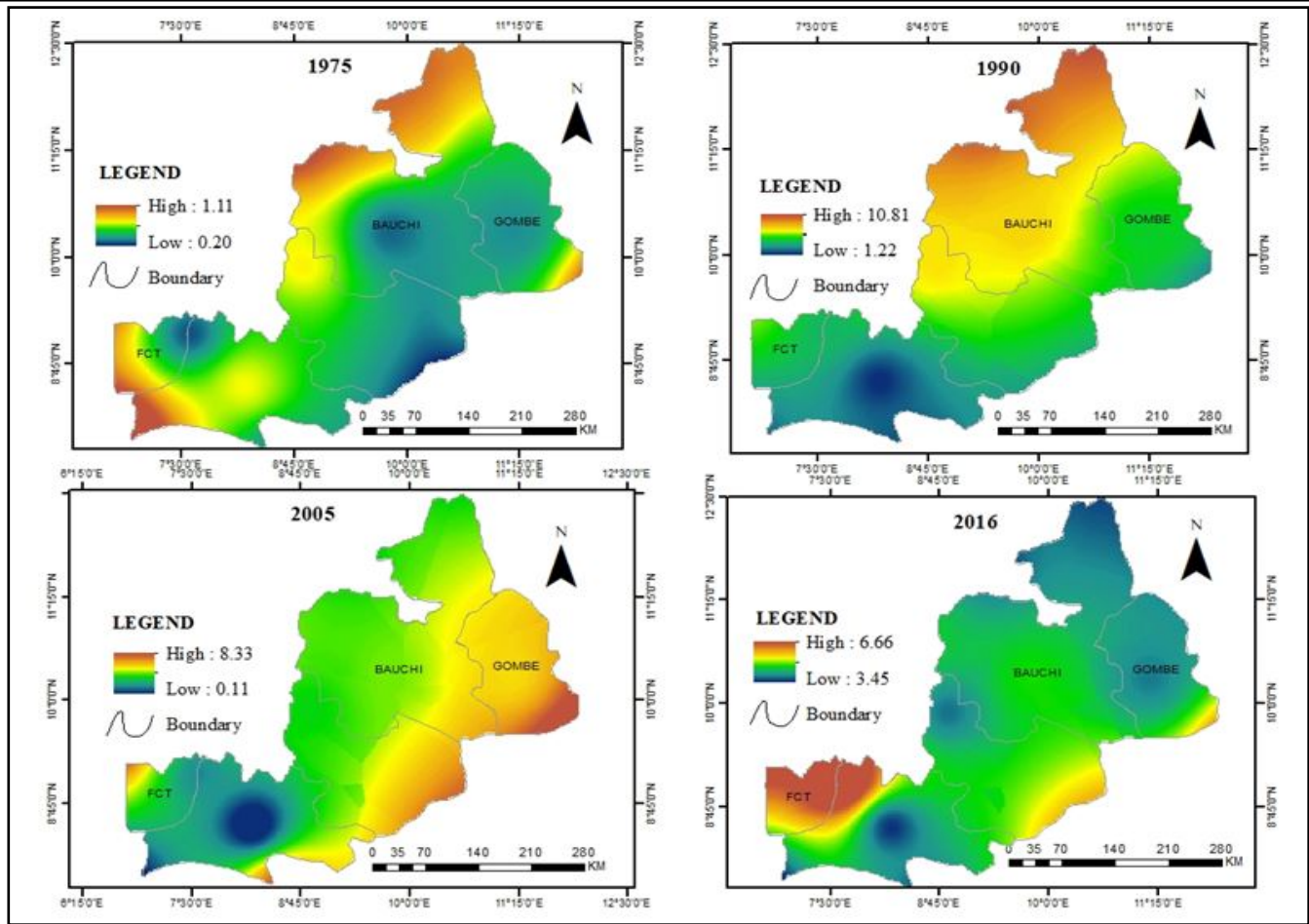
Figure 6a above shows the mean spatial drought magnitude which was depicted as positive values and corresponding duration (Figure 6b) across the ecological zones for different hydrological epochs. Results showed that drought magnitude peaked in 1983-1984, 1988 with the longest duration of 5years. Mean drought magnitude was highest in 1990 with 10.81 and closely followed by 2005 with a mean magnitude of 8.33 (Figure 6). Both mean magnitudes were observed within the Sahel, Sudan, and Jos Plateau ecological zones as represented in the states under review.

A southward shift in the magnitude reflected the inconsistency of drought in the region. 1975 recorded the least mean drought magnitude with worst values at 1.109 while the longest duration was 12 months in all 4 epochs. It appeared drought severity and magnitude incidences varied from a location to another through the years under review. However, 2002 recorded the worst-hit year as across the region Bauchi, Gombe, Plateau, and FCT experienced various degrees of drought for a duration as high as

8months with its peak at -2.85 experienced in Bauchi, Gombe states, and FCT.

3.3. Drought risk mapping over the Guinean- Sudano zone of Nigeria

Figure 7 above revealed spatial drought severity which in turn illustrated the drought risk of the region. The drought characteristics of the region were classified into three and base on the mean drought intensity (severity) values. The results revealed that drought could be classified as mildly wet, mild drought, and moderate drought at a macro level however at local levels drought mechanisms are more complex and the drivers are highly dependable on the local environmental conditions prevalent in the states [26,39]. Drought events were linked to low precipitation values that were closely reflected through the SPI values. However, it reaffirms that environmental problems are best and in-depth managed at the local levels when knowledge of the micro climatic condition is considered.



(a) Average drought magnitude across the ecological zones



Figure 6: Spatial illustration of average drought magnitude and duration for different hydrological years

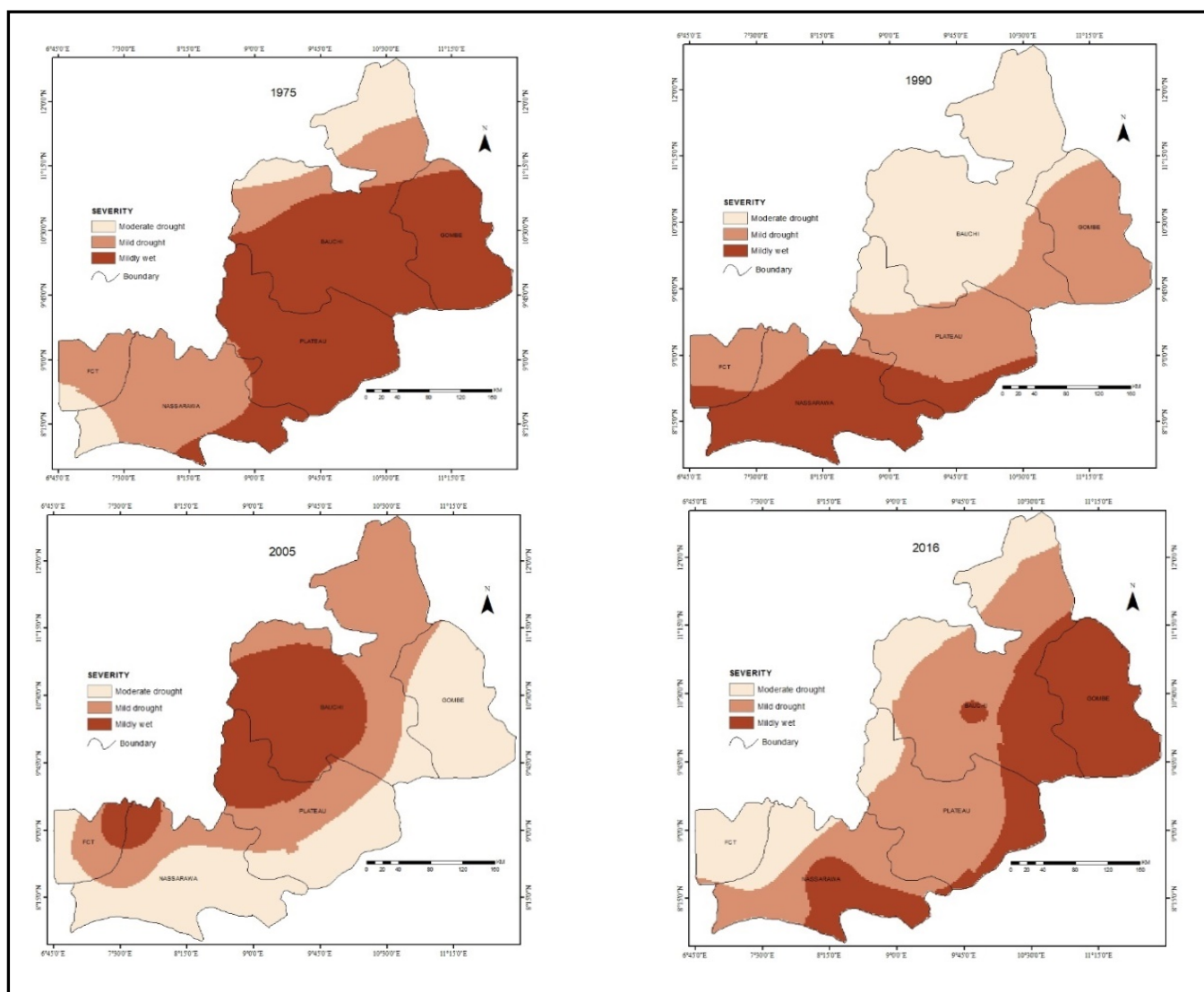


Figure 7: Drought severity over the ecological zones

Table 3: Mann-Kendall and significance level of SPI-3, -6, -12 over the Guinean-Sudano ecological zone

DURATION	PARAMETER	FCT	BAUCHI	GOMBE	PLATEAU	NASSARAWA	REGIONAL
SPI 3	Kendall tau	-0.008	0.073	0.023	0.006	-0.009	0.017
	(Significance)	0.85	0.081	0.579	0.892	0.852	0.6508
	Sen's slope	0.0002	0.008	0.005	0.001	-0.001	0.00264
	Trend	No	No	No	No	No	No
SPI 6	Kendall tau	0.055	0.092	0.06	0.045	0.024	0.0552
	(Significance)	0.254	0.074	0.271	0.376	0.641	0.3232
	Sen's slope	0.008	0.011	0.008	0.003	0.003	0.0066
	Trend	No	No	No	No	No	No
SPI 12	Kendall tau	0.111	0.148	0.077	0.06	0.043	0.0878
	(Significance)	0.041	0.009	0.016	0.269	0.454	0.1872
	Sen's slope	0.012	0.019	0.01	0.007	0.03	0.0156
	Trend	Yes	Yes	Yes	No	No	No

Table 4: Mann-Kendall and significance level of precipitation over the Guinean-Sudano ecological zones

STATES	KENDALL	SIGNIFICANCE LEVEL	SEN'S SLOPE	TREND
FCT	0.083	0.448	0.649	No
BAUCHI	0.136	0.209	1.904	No
GOMBE	0.027	0.812	0.235	No
PLATEAU	0.001	0.97	0.007	No
NASSARAWA	0.041	0.713	0.609	No
REGIONAL	0.058	0.630	0.681	No

3.4. Spatial trends in drought across the Guinean- Sudano zone of Nigeria

The spatial and temporal variability in drought trends were observed in the study area and shown in Table 3 as the Negative and Positive SPI trends at multiple time scales. Of all the SPI representations tested, only SPI-12 indicated significant trend values in Bauchi, Gombe, and FCT with Sen's slope (Kendal tau) values of 0.019 (0.148), 0.01 (0.077), and 0.012 (0.111) respectively.

However, no trend existed at the regional level with Sen's slope (Kendal tau) values of 0.0156 (0.088). The Mann-Kendall trend test of SPI-values revealed statistically significant (insignificant) decreasing trends for short-term months in FCT and Nassarawa state which are within the Guinea savanna with Kendall tau values of -0.008 and -0.009 respectively. SPI6- month timescales showed

statistically non-significant increasing trends across the ecological zones. So, the Null hypothesis (H_0) was accepted across the ecological zones for timescales -3 and 6-months as the recorded values were greater than the significance level (p-value 0.05) while the alternative hypothesis for FCT (Guinea savanna); Bauchi and Gombe states (within the Sahel-Sudan savanna) respectively were accepted principally for SPI-12. These denoted that SPI 12- shows the status of a year-round water shortage caused by drought while SPI-6 and SPI-3 are appropriate indicators of the status of a seasonal water shortage caused by drought [39].

Table 4 shows positive annual precipitation trends over the ecological zones which exhibited a significantly increasing trend occurred both locally in the states and regionally for the periods under review. Thus, in verifying the status of drought over the ecological zone, some water reservoirs were visited (Figure 8).

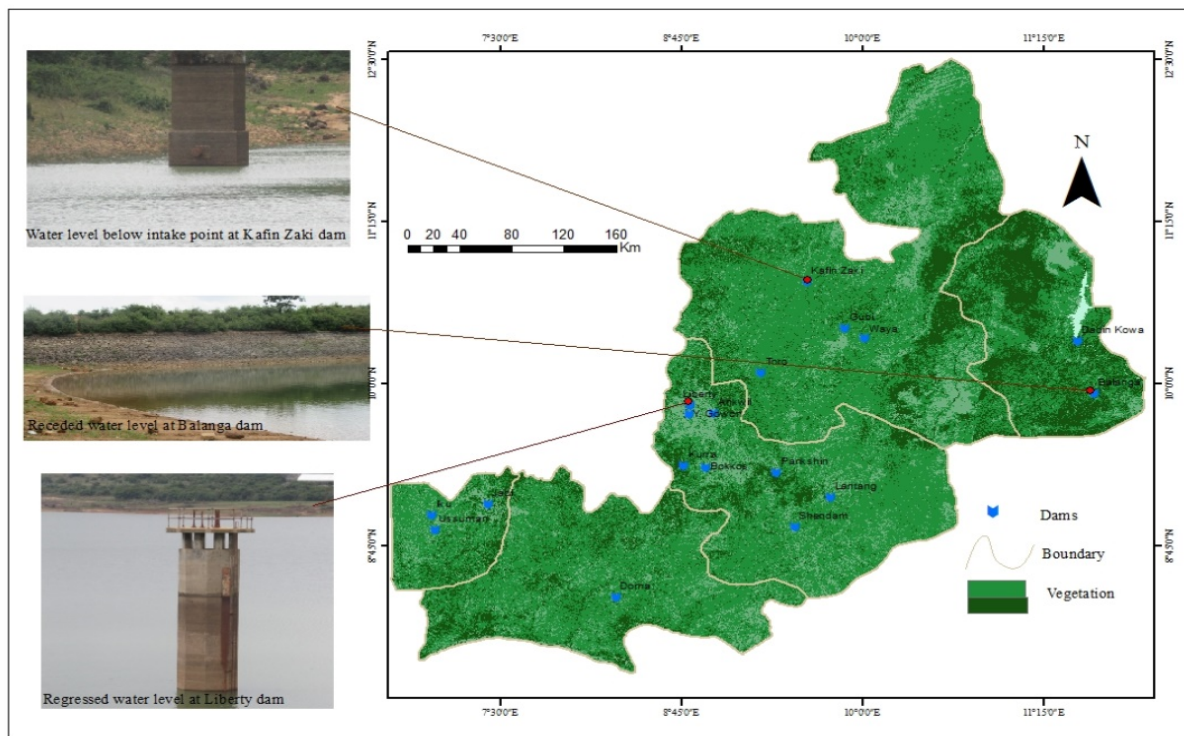


Figure 8: Validation of water death due to drought at some reservoirs visited

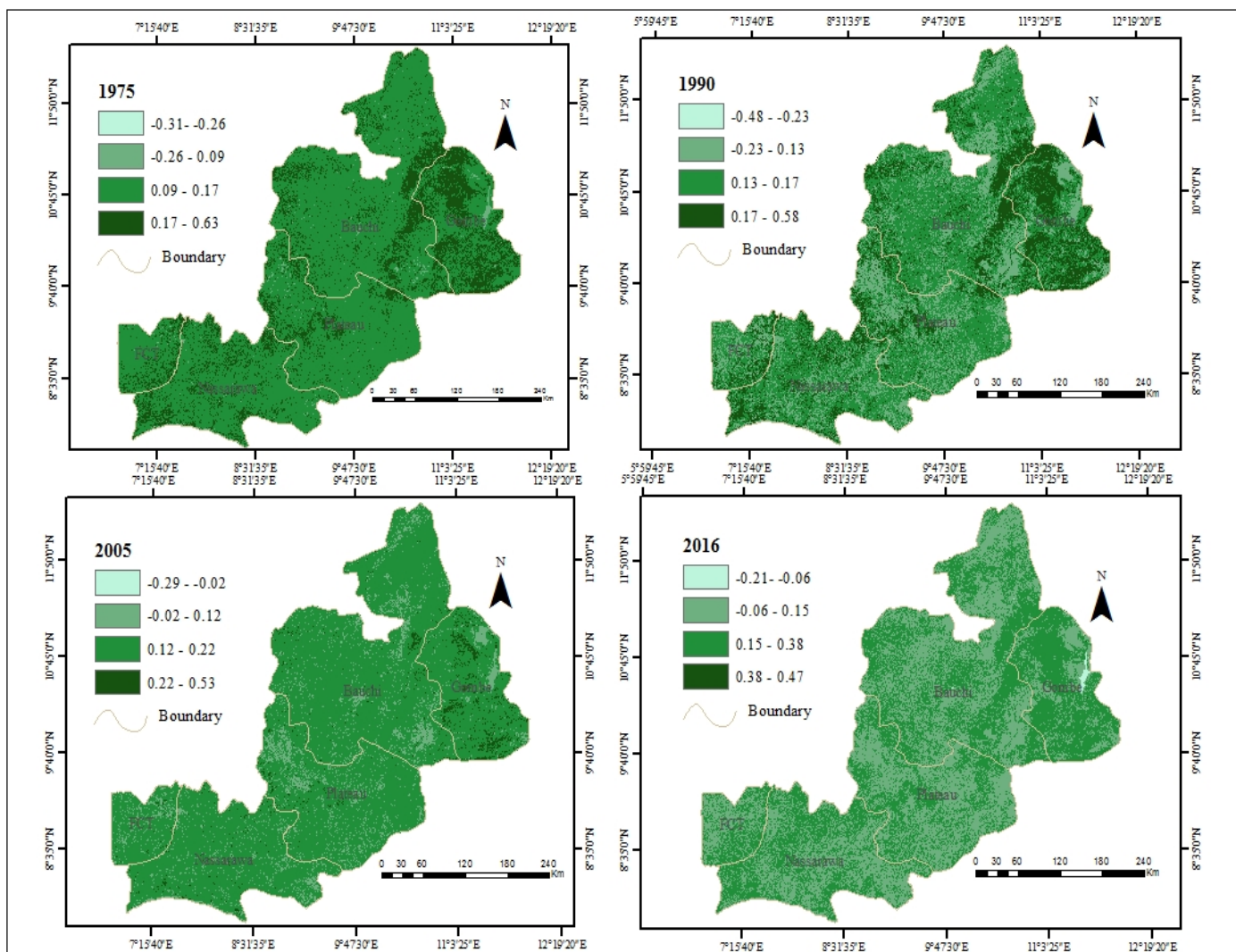


Figure 9: NDVI distribution over the ecological zones as of January and March

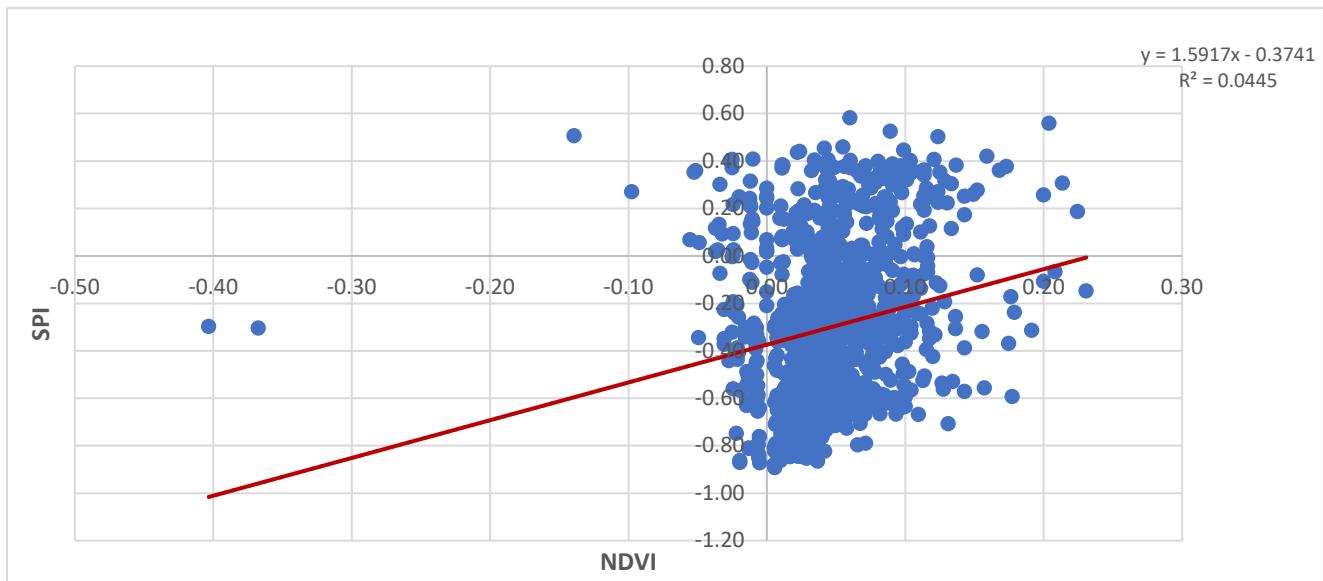
Pieces of evidence of drought on water reservoirs were lucid as receded water levels at peak of rainy season observed between July and September 2020 revealed about 10metres to 16metres depth of these dams have been lost with dry spillways visible across the visited sites (Figure 8) which were contrary to the usual water levels at the peak of rainy seasons decades ago. Observation revealed water levels were below the intake points of the tower thereby causing major water shortages particularly for irrigation farmers and water consumption in the urban centers.

3.5. Drought relationship to NDVI

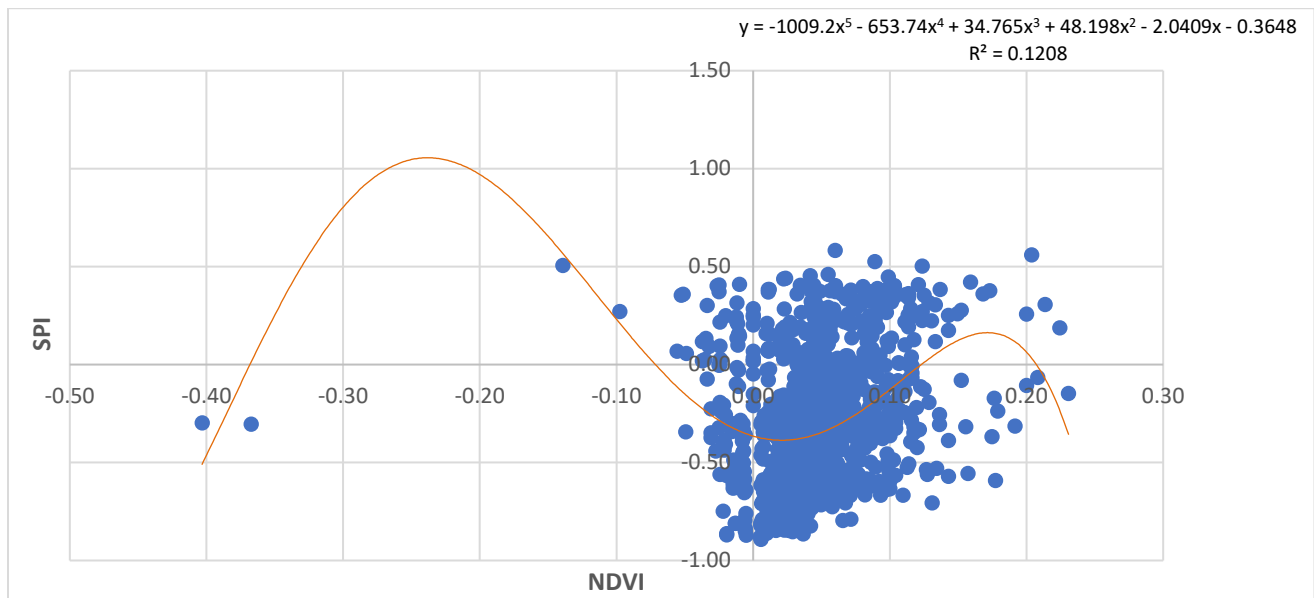
Considering Guinean- Sudano the tremendous impact of drought on vegetation cover, it is pertinent that this biological indicator is given the required attention. Thus, in this study, the Normalized Difference Vegetation Index (NDVI) was carried out alongside a reconnaissance survey to verify the values obtained. It was observed that NDVI values were higher in 1975 with a range between - 0.31 and 0.63. Areas along river channels and other water bodies were noted to have higher NDVI values than other LULC types. 2016 recorded lower values with a range between -0.21 and 0.47 (Figure 8). NDVI minimum and

maximum values were between -0.48 and 0.63 between 1975 and 2016, while in 2016 highest of 0.47 was observed the least maximum value with 1975 recording the highest value at 0.63. Negative index values were recorded over water bodies while values close to 0 over the land without vegetation.

In [59], the author suggested that values less than 0.1 will be considered as desert as they often represent bare rocks, sand, or bare surfaces [60]; between 0.1 and 0.2 as semi-desert while areas with 0.2 – 0.4 as shrub and grassland (which are similar characteristics of the savannas), and finally pixels with more than 0.5 will be assumed having forest land cover characterized by dense, green vegetation. Likewise, [61,62], classified regions that had a long-term mean monthly NDVI value between 0.1 and 0.5 as subject to drought and desertification. Therefore, the relationship between vegetation density and soil moisture with the role these parameters play in drought assessment have shown that they are strongly related and can explain the dynamics of drought and desertification [59]. Nonetheless, other parameters such as rainfall, soil quality, and socio-economic factors have strategic roles.



(a) Linear regression model



(b) Non-linear regression model

Figure 10: Relationship between NDVI and SPI using both linear and non-linear regression models

The relationship between NDVI which represents the amount of vegetation cover and SPI that denotes variability of precipitation was verified using regression analysis in this study. It was observed that precipitation and vegetation had a positive linear relationship with correlation coefficient ‘R’ for the 4 epochs under review: 1975 recorded 0.0625; 1990 at 0.045; 2005 at 0.126 and 2016 at 0.1334 respectively (Figure 10a). However, the non-linear regression appeared more suitable representation of the relationship between NDVI and SPI as a higher correlation coefficient ‘R’ 0.12 was recorded as against 0.045 observed in 1990 (Figure 10b) this affirms the suggestion of [63,64] who opined that the non-linear regression model provides relatively suitable fit for representation of relationships between variables. This further suggests that either way there is a strong relationship between precipitation and abundance of vegetation as the trend

revealed that a decrease in precipitation reflected a decrease in vegetation cover over the ecological zones. Therefore, the NDVI related to SPI are significant indicators for vegetation status and its response to precipitation variability [65].

4. Discussion

Drought is a recurrent climatic feature that occurs in virtually every agro-climatic zone of Nigeria. Characteristics such as severity, duration, and geospatial extent, deaths, socio-economic effect, and long-term impacts are used to categorize drought impacts. It adversely affects different types of systems and places various demands on the available water within the Nigerian society such as agriculture, hydropower generation, rural and urban water supply schemes, and various industries [15].

However, the prevalence of drought disasters related to the shortfall of rainfall remains very important. In this study, the temporal and spatial spread of drought frequency, magnitude, intensity, and severity were examined. The studied results indicated 2002; 2007- 2008; and 2013-2014 were the most severe drought periods with extreme values below -3.0.

The SPI values of 3-month as well as 6- and 12-months timescales had an eminent impact on the rainfall pattern of the study periods that echoes the prompt occurrence of agricultural drought. In this regard, [25,58] confirmed that the 1- to 6-months timescales of SPI values lucidly reflects the fast development of agricultural drought, the exacerbating of the actual state of water conditions, and the negative effects of rainfall deficit in agriculture (soil moisture, groundwater table depth, crop yield). The significance of recurrent incidence of droughts for 3-, 6-, and 12-months and ensuing implications on the socio-economy of rain-fed agricultural practices and productions for smallholder farmers cannot be shoved aside.

The most significant drought risk incidence of months in the dry and wet seasons under moderate and (extreme) drought intensity classes highly impacted the agricultural practices as reaffirmed by [66] that “from the 1970s onwards, the decline of rainfall amount and the increasing frequency of drought directly influenced water shortages; limited crop growth and development; and reduced yield”.

It's been cogently revealed that the most immediate impact of drought is a decrease in crop production, due to inadequate or poorly distributed rainfall. During droughts drastically reduced rainfall and high temperatures across the country lead to high exposure rates to drought impacts [25]. In Nigeria, agriculture occupies about 40 percent of the country's GDP, and employs about 60 percent of the active labor force, with women representing about 65 percent [34]. More than 70 percent of the Nation's livestock are raised in the drought-prone drylands of Nigeria [15,67]. The poor crop yields or total crop failure due to drought result in mass poverty and starvation as agriculture is the mainstay of Nigeria's rural economy with its main source of water for agriculture being rainfall and so drought has a tremendous effect on agriculture and natural resources [15]. Malnutrition due to famines which usually follow droughts occurs due to reduced food production from the spread of infectious disease and food and water-borne illness, and increased air pollution [68]. Thus, if not monitored, the drought would lead to devastations with dire repercussions.

In addition, water availability is the paramount limitation to economic growth in northern Nigeria as reduced annual average rainfall and its run-off could exacerbate the increasing drought and desertification in Nigeria. The occurrence of droughts leads to the drop in the water levels of rivers and streams and the lowering of the water table which was evident in some water reservoirs visited in the course of study July/ September 2020 (Figure 8). This means that the region truly does not have abundant water assets to sustain the current level of per capita food production while meeting rational water needs for domestic, industrial, and ecological purposes [69]. Moreover, pertinent to note is the growing rate of rural-urban migration across the region as revealed in the report [34] which depicted a shift from 20% to

52% in rural-urban population in the period under review. The resilience of the rural productive systems is based on short-term circulation from the rural areas to the urban environments [69] commonly termed (Hausa *cin rani*).

Subsequently, Nigerian forests are already under great pressure with a deteriorating state of deforestation arising from increasing populations and growing economic needs. These have led to greater demand for forest resources exacerbating drought and other forms of land degradation including desertification [70] particularly with increasing changing climate and anthropogenic activities on forest growth and development and through greater demands on forests by populations adjusting to drought. Likewise, large populations within the drought-affected region regularly migrate southwards towards the coastal zone especially for grazing and this has led to resource conflicts between farmers and herders. The effects of drought on agriculture, water resources, and other sectors have stripped people of their means of livelihood, making the populace susceptible to inter-communal conflicts; increase in recruitments into criminal gangs and armed militia; armed robbery, insurgency, increase in possession of small arms[15] and other forms of social vices.

5. Conclusion

In conclusion, this study was able to affirm that drought is in dire need of deliberate attention and the choices we make today will determine the chance our communities and nation at large have in winning the fight against this menace. Results indicated that SPI dynamics over the region at varying timescales suggest the unpredictability of drought with longest drought duration were the severe and extreme drought ($SPI \leq -1.5$) was peaked at -3.12 severity that lasted 15months between 2007 and 2008. The worst drought magnitudes were more eminent in the longer timescale (SPI-12months) with its peak of -70.8 that lasted 60months (5 years) between 1984 and 1986. Similarly, drought intensity was observed between the mildly dry with SPI -0.47(1986 - 1991) and severely dry with SPI - 1.79 (1983 - 1984) this further revealed that spatially drought risk is increasing southward. SPI values demonstrated high variability in both wet and dry spells; however, the longer timescale at 12 months witnessed lesser variability. However, due to security challenges in the northern parts of Nigeria, the study demonstrated the resourcefulness of incorporating remote sensing and GIS as a means of analysing drought risk over the region other than the use of basic statistical measures. Over the years, there have been significant variations in vegetation cover as NDVI applied in analyzing vegetation quality revealed that areas along river channels had higher NDVI values which corresponded with thicker vegetation (riparian) observed during field surveys. NDVI values obtained demonstrated a consistent decrease with a better state in 1975 which ranged between - 0.31 and 0.63. 2016 recorded lower values with a range between -0.21 and 0.47. NDVI minimum and maximum values were between -0.48 and 0.63 between 1975 and 2016, while in 2016 highest of 0.47 was observed the least maximum value with 1975 recording the highest value at 0.63. NDVI dynamics were related to rainfall, dryness, and moisture content of the area to have a better understanding of drought within the climate change context. With the recent focus on Land Degradation Neutrality (LDN), a need for integrating assessment

of key ecological problems such as desertification and other forms of land degradation will improve preventive and mitigating actions thereby better chances for combatting drought vulnerability and risk within northern Nigeria.

Acknowledgement

Our profound appreciation to African Union Commission (AUC) for funding this study by a grant provided to the first Author.

Conflict of Interest

The authors have read and understood the policy on declaration of interests and declare that there is no conflict of interest related to this article or legal bond or copyrights.

References

- [1] A. Dai, "Drought under global warming: a review," *WIREs Climate Change*, **2**(1), 45–65, 2011, doi:10.1002/wcc.81.
- [2] D.A. Wilhite, M.H. Glantz, "Understanding the Drought Phenomenon: The Role of Definitions," *Water International*, **10**(3), 111–120, 1985, doi:10.1080/02508068508686328.
- [3] I. Masih, S. Maskey, F.E.F. Mussá, P. Trambauer, "A review of droughts on the African continent: a geospatial and long-term perspective," *Hydrology and Earth System Sciences*, **18**(9), 3635–3649, 2014, doi:10.5194/hess-18-3635-2014.
- [4] World Meteorological Organization (WMO) and Global Water Partnership (GWP), *Handbook of Drought Indicators and Indices*, Geneva, 2016.
- [5] W. Zhan, K. Guan, J. Sheffield, E.F. Wood, "Depiction of drought over sub-Saharan Africa using reanalyses precipitation data sets," *Journal of Geophysical Research: Atmospheres*, **121**(18), 2016, doi:10.1002/2016JD024858.
- [6] J. Sheffield, E.F. Wood, "Projected changes in drought occurrence under future global warming from multi-model, multi-scenario, IPCC AR4 simulations," *Climate Dynamics*, **31**(1), 79–105, 2008, doi:10.1007/s00382-007-0340-z.
- [7] Trenberth Kevin, Meehl Jerry, Masters Jeff, Somerville Richard, *Current Extreme Weather and Climate Change*, 2011.
- [8] Centre for Research on the Epidemiology of Disasters (CREDE), *Annual Disaster Statistical Review: The numbers and trends*, 2016.
- [9] Intergovernmental Panel on Climate Change, *Summary for Policymakers. In Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, 2013.
- [10] M.M.Q. Mirza, "Climate change and extreme weather events: can developing countries adapt?," *Climate Policy*, **3**(3), 233–248, 2003, doi:10.3763/cpol.2003.0330.
- [11] K.V. Suryabhagavan, "GIS-based climate variability and drought characterization in Ethiopia over three decades," *Weather and Climate Extremes*, **15**, 11–23, 2017, doi:10.1016/j.wace.2016.11.005.
- [12] M.A. Miyan, "Droughts in Asian Least Developed Countries: Vulnerability and sustainability," *Weather and Climate Extremes*, **7**, 8–23, 2015, doi:10.1016/j.wace.2014.06.003.
- [13] B. Shiferaw, K. Tesfaye, M. Kassie, T. Abate, B.M. Prasanna, A. Menkir, "Managing vulnerability to drought and enhancing livelihood resilience in sub-Saharan Africa: Technological, institutional and policy options," *Weather and Climate Extremes*, **3**, 67–79, 2014, doi:10.1016/j.wace.2014.04.004.
- [14] C. Funk, P. Peterson, M. Landsfeld, D. Pedreros, J. Verdin, S. Shukla, G. Husak, J. Rowland, L. Harrison, A. Hoell, J. Michaelsen, "The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes," *Scientific Data*, **2**(1), 150066, 2015, doi:10.1038/sdata.2015.66.
- [15] N. Federal Ministry of Environment, *National Drought Plan under Federal Ministry of Environment*, Nigeria., 2018.
- [16] I.F.P. Research Institute (IFPRI), *2016 Global hunger index: Getting to zero hunger*, 2016, doi:10.2499/9780896292260.
- [17] A. Dai, T. Zhao, J. Chen, "Climate Change and Drought: a Precipitation and Evaporation Perspective," *Current Climate Change Reports*, **4**(3), 301–312, 2018, doi:10.1007/s40641-018-0101-6.
- [18] J.N. Eze, "Drought occurrences and its implications on the households in Yobe state, Nigeria," *Geoenvironmental Disasters*, **5**(1), 18, 2018, doi:10.1186/s40677-018-0111-7.
- [19] E.O. Oladipo, "A comprehensive approach to drought and desertification in Northern Nigeria," *Natural Hazards*, **8**(3), 235–261, 1993, doi:10.1007/BF00690910.
- [20] O.Y. Ekundayo, E.C. Okogbue, F. Akinluyi, A. Kalumba, I. Orimoloye, "Spatiotemporal drought assessment using vegetation health index and standardized precipitation index over Sudano-Sahelian region of Nigeria," *African Geographical Review*, **1**–13, 2020, doi:10.1080/19376812.2020.1841658.
- [21] O. Ideki, M.O. Nwagbara, "Geo-Spatial Assessment of Vegetation Response to Drought in North Central Nigeria," *Journal of Geoscience and Environment Protection*, **07**(03), 46–63, 2019, doi:10.4236/gep.2019.73003.
- [22] Federal Ministry of Environment (FME) Nigeria, *National Forest Reference Emission Level (FREL) for the Federal Republic of Nigeria*, 2019.
- [23] J. McQuigg, "A Simple Index of Drought Conditions," *Weatherwise*, **7**(3), 64–67, 1954, doi:10.1080/00431672.1954.9930323.
- [24] T. B. McKee, N. . Doesken, J. Kleist, "The Relationship of Drought Frequency and Duration to Time Scales.," in *Proceedings of the 8th Conference on Applied Climatology*, Boston, 1993.
- [25] A.A. Mekonen, A.B. Berlie, M.B. Ferede, "Spatial and temporal drought incidence analysis in the northeastern highlands of Ethiopia," *Geoenvironmental Disasters*, **7**(1), 2020, doi:10.1186/s40677-020-0146-4.
- [26] A. Huete, "A soil-adjusted vegetation index (SAVI)," *Remote Sensing of Environment*, **25**(3), 295–309, 1988, doi:10.1016/0034-4257(88)90106-X.
- [27] Q. Mu, M. Zhao, J.S. Kimball, N.G. McDowell, S.W. Running, "A Remotely Sensed Global Terrestrial Drought Severity Index," *Bulletin of the American Meteorological Society*, **94**(1), 83–98, 2013, doi:10.1175/BAMS-D-11-00213.1.
- [28] M.A. Haroon, J. Zhang, F. Yao, "Drought monitoring and performance evaluation of MODIS-based drought severity index (DSI) over Pakistan," *Natural Hazards*, **84**(2), 1349–1366, 2016, doi:10.1007/s11069-016-2490-y.
- [29] O. Adedeji, A. Olusola, G. James, H.A. Shaba, I.R. Orimoloye, S.K. Singh, S. Adelabu, "Early warning systems development for agricultural drought assessment in Nigeria," *Environmental Monitoring and Assessment*, **192**(12), 798, 2020, doi:10.1007/s10661-020-08730-3.
- [30] I.R. Orimoloye, O.O. Ololade, "Spatial evaluation of land-use dynamics in gold mining area using remote sensing and GIS technology," *International Journal of Environmental Science and Technology*, **17**(11), 4465–4480, 2020, doi:10.1007/s13762-020-02789-8.
- [31] R.W.J. Keay, *An outline of Nigerian vegetation*, 3rd ed., Federal Government Printer, Lagos, 1959.
- [32] N. Federal Ministry of Environment (FME), *State of the Nigerian Environment Report*, 2008.
- [33] N. National Bureau of Statistics, *National Bureau of Statistics, Nigeria*, 2017.
- [34] N. National Bureau of Statistics, *National Bureau of Statistics, Nigeria*, 2019.
- [35] World Meteorological Organization, Hayes M, Wood D, Svoboda M, *Standardized Precipitation Index User Guide*, Geneva, 2012.
- [36] M. Hayes, M. Svoboda, N. Wall, M. Widhalm, "The Lincoln Declaration on Drought Indices: Universal Meteorological Drought Index Recommended," *Bulletin of the American Meteorological Society*, **92**(4), 485–488, 2011, doi:10.1175/2010BAMS3103.1.
- [37] World Meteorological Organization (WMO), *Standardized Precipitation Index User Guide*, Geneva, 2012.
- [38] M.R.I. Baig, Shahfahad, M.W. Naikoo, A.H. Ansari, S. Ahmad, A. Rahman, "Spatio-temporal analysis of precipitation pattern and trend using standardized precipitation index and Mann–Kendall test in coastal Andhra Pradesh," *Modeling Earth Systems and Environment*, 2021, doi:10.1007/s40808-021-01262-w.
- [39] W. Kalisa, J. Zhang, T. Igbawua, F. Ujoh, O.J. Ebohon, J.N. Namugize, F. Yao, "Spatio-temporal analysis of drought and return periods over the East African region using Standardized Precipitation Index from 1920 to 2016," *Agricultural Water Management*, **237**, 2020, doi:10.1016/j.agwat.2020.106195.
- [40] H. Xie, C. Ringler, T. Zhu, A. Waqas, "Droughts in Pakistan: A spatiotemporal variability analysis using the Standardized Precipitation Index," *Water International*, **38**(5), 620–631, 2013, doi:10.1080/02508060.2013.827889.
- [41] P.S.D.N. Thenkabail, V.U. Smakhtin, "The Use of Remote Sensing Data for Drought Assessment and Monitoring in Southwest Asia," 5–13, 2004.
- [42] S.L. Lewis, P.M. Brando, O.L. Phillips, G.M.F. van der Heijden, D. Nepstad,

- “The 2010 Amazon Drought,” *Science*, **331**(6017), 554–554, 2011, doi:10.1126/science.1200807.
- [43] O. V. Wilhelmi, D.A. Wilhite, “Assessing Vulnerability to Agricultural Drought: A Nebraska Case Study,” *Natural Hazards*, **25**(1), 37–58, 2002, doi:10.1023/A:1013388814894.
- [44] H.Q. Liu, A. Huete, “A feedback based modification of the NDVI to minimize canopy background and atmospheric noise,” *Undefined*, **33**(2), 457–465, 1995, doi:10.1109/36.377946.
- [45] J.E. Joseph, O.O. Akinrotimi, K.P.C. Rao, A.P. Ramaraj, P.S. Traore, P. Sujatha, A.M. Whitbread, *The Usefulness of Gridded Climate Data Products in Characterizing Climate Variability and Assessing Crop Production*, 2020.
- [46] National Drought Mitigation Center, SPI Program, National Drought Mitigation Center, 2018.
- [47] H. Abghari, H. Tabari, P. Hosseinzadeh Talaei, “River flow trends in the west of Iran during the past 40years: Impact of precipitation variability,” *Global and Planetary Change*, **101**, 52–60, 2013, doi:10.1016/j.gloplacha.2012.12.003.
- [48] J. Hannaford, B. Lloyd-Hughes, C. Keef, S. Parry, C. Prudhomme, “Examining the large-scale spatial coherence of European drought using regional indicators of precipitation and streamflow deficit,” *Hydrological Processes*, **25**(7), 1146–1162, 2011, doi:10.1002/hyp.7725.
- [49] Kang-Tsung C, *Introduction to Geographic Information System*, 6th ed., McGraw-Hill International, New-York, 2012.
- [50] M.A. Degefu, W. Bewket, “Trends and spatial patterns of drought incidence in the omo-ghibe river basin, ethiopia,” *Geografiska Annaler: Series A, Physical Geography*, **97**(2), 395–414, 2015, doi:10.1111/geoa.12080.
- [51] L. Labedzki, “Estimation of local drought frequency in central Poland using the standardized precipitation index SPI,” *Irrigation and Drainage*, **56**(1), 67–77, 2007, doi:http://dx.doi.org/10.1002/ird.285.
- [52] T. Sternberg, D. Thomas, N. Middleton, “Drought dynamics on the Mongolian steppe, 1970-2006,” *International Journal of Climatology*, **31**(12), 1823–1830, 2011, doi:10.1002/joc.2195.
- [53] J. Spinoni, G. Naumann, H. Carrao, P. Barbosa, J. Vogt, “World drought frequency, duration, and severity for 1951-2010,” *International Journal of Climatology*, **34**(8), 2792–2804, 2014, doi:10.1002/joc.3875.
- [54] K. Gopal Ghosh, *Geo-Statistical Assessment of the Intensity, Duration, Frequency and Trend of Drought over Gangetic West Bengal, Eastern India*, IntechOpen, 2019, doi:10.5772/intechopen.80037.
- [55] M.G. Kendall, *Rank correlation methods*, 4th edn. Griffin, London, 4th ed., Griffin, London, 1975.
- [56] B. Mann Henry, “Nonparametric tests against trend,” *Econometrica: Journal of the Econometric Society*, 245–259, 1945.
- [57] P.K. Sen, “Estimates of the Regression Coefficient Based on Kendall’s Tau,” *Journal of the American Statistical Association*, **63**(324), 1379–1389, 1968, doi:10.1080/01621459.1968.10480934.
- [58] E.P. Agbo, C.M. Ekpo, “Trend Analysis of the Variations of Ambient Temperature Using Mann-Kendall Test and Sen’s Estimate in Calabar, Southern Nigeria,” in *Journal of Physics: Conference Series*, IOP Publishing Ltd, 2021, doi:10.1088/1742-6596/1734/1/012016.
- [59] A.C. Anyamba, Tucker C J, Eastman J R, “NDVI anomaly patterns over Africa during the 1997/98 ENSO warm event,” *International Journal of Remote Sensing*, **22**(10), 1847–1859, 2001, doi:10.1080/01431160010029156.
- [60] K. Ryder, *Differences to Measure Drought in the Russian River Watershed*, 2021.
- [61] U. Helldén, C. Tottrup, “Regional desertification: A global synthesis,” *Global and Planetary Change*, **64**(3–4), 169–176, 2008, doi:10.1016/j.gloplacha.2008.10.006.
- [62] E. Rimkus, E. Stonevicius, J. Kilpys, V. Maciulyte, D. Valiukas, “Drought identification in the eastern Baltic region using NDVI,” *Earth System Dynamics*, **8**(3), 627–637, 2017, doi:10.5194/esd-8-627-2017.
- [63] X. Qian, J. Xue, Y. Yang, S.W. Lee, “Thermal Properties and Combustion-Related Problems Prediction of Agricultural Crop Residues,” *Energies*, **14**(15), 4619, 2021, doi:10.3390/en14154619.
- [64] M. Assefa, A. Wossenu, S. Gabriel, *Extreme Hydrology and Climate Variability: Monitoring, Modelling, Adaptation and Mitigation*, 1st ed., Elsevier, 2019.
- [65] H. Barbosa, C. Tote, L. Kumar, Y. Bamutaze, *Harnessing Earth Observation and Satellite Information for Monitoring Desertification, Drought and Agricultural Activities in Developing Countries*, 1st ed., InTech, 2013, doi:10.5772/55499.
- [66] United Nations Environment Program/UNEP, *Food security in the horn of Africa: the Implications of a drier, hotter and more crowded future*, 2011.
- [67] I.U. Abubakar, M.A. Yamusa, “Recurrence of drought in Nigeria: causes, effects, and mitigation,” *Int. J. Agric. Food Sci. Technology*, **4**(3), 169–180, 2013.
- [68] A.G. Hassan, M.A. Fullen, D. Oloke, “Problems of drought and its management in Yobe State, Nigeria,” *Weather and Climate Extremes*, **23**, 2019, doi:10.1016/j.wace.2019.100192.
- [69] A.B. Alhassan, R.C. Carter, I. Audu, “Agriculture in the Oasis of the Manga Grasslands of Semi-Arid North-East Nigeria,” *Outlook on Agriculture*, **32**(3), 191–195, 2003, doi:10.5367/000000003101294460.
- [70] O. Joseph, A.E. Gbenga, D.G. Langyit, “Desertification risk analysis and assessment in Northern Nigeria,” *Remote Sensing Applications: Society and Environment*, **11**, 70–82, 2018, doi:10.1016/j.rsase.2018.04.012.