Input Feature Optimization for ANN Models Predicting Daylight in Buildings

C-L. Lorenz, A. B. Spaeth, C. Bleil De Souza, M. Packianather
Cardiff University, Wales, U.K.
lorenzc4@cardiff.ac.uk

Abstract. Artificial Neural Networks (ANNs) were used as prediction models to explore design solutions for the atrium design of a school building. To this end, a solution space of 165 design variants was generated via parametric modeling. This paper details the process of extracting and selecting the input features required for ANN training in order to predict the DA and sDA metric. The feature selection undertaken in this study mainly consisted of two steps: Firstly, a computationally less extensive machine learning model was used to rank the input features according to their relevance in predicting daylight levels. Secondly, ANNs were trained applying sequential forward selection. The proposed method is investigated in terms of achievable improvements to prediction accuracy, reduceable training time and the feasibility of the method.

1. Introduction

Access and exposure to daylight not only regulates the circadian rhythm but is vital to reducing fatigue (Figueiro *et al.*, 2017) and has been shown to improve well-being, work performance and students' performance in schools (Maesano and Annesi-Maesano, 2012; Veitch, Christoffersen and Galasiu, 2013). In recent years, the Daylight Factor (DF), a standard metric for assessing daylight in buildings, came under scrutiny for its unsuitability as a design driver, given that it considers neither orientation nor climate (Reinhart, Mardaljevic and Rogers, 2013). In its stead, climate-based metrics were introduced (Reinhart and Walkenhorst, 2001) and have only recently been adopted into European and British building standards (EN 17037:2018). Of particular interest in this paper was the Daylight Autonomy (DA) metric, which quantifies the percentage of occupied hours in a year, in which a threshold of 300 lux can be met with daylight. However, compared to point-in time daylight factor calculations, the annual simulation for climate-based metrics is more time consuming, making it inherently more difficult to assess design options as an integral part of the design process (Jones and Reinhart, 2019).

Research has introduced Artificial Neural Networks (ANNs) as surrogate energy simulation software in order to improve the feasibility of performance-driven design investigations (Nguyen, Reiter and Rigo, 2014). The benefits of such ANN-based models include their instantaneous response rate (once trained) and high accuracy, under the condition that ANN models are trained with sufficient performance data, an appropriate model selection and parameter settings (Zhao and Magoulès, 2012). Such models have been found to be particularly useful for automated building control, where an immediate response is required (Hu and Olbina, 2011). Their use further extends to the modeling of large design solution spaces in order to solve design problems with numerous design variables (Machairas, Tsangrassoulis and Axarli, 2014). In general, ANNs present a promising avenue in research as a mature technology that works well with noisy data (i.e. simulation data) and can operate with various types of input and output (i.e. real, discrete, Boolean), the relationship between which need not be known (Mckee, Schulz and Caruana, 2006). Such research has laid the groundwork of this paper, as we applied ANN-based modeling to efficiently predict daylight for the solution space of a building design. More specifically, this study trained ANNs to predict climate-based metrics to explore design variants for the central atrium of a school building. Within that context, this paper focused on the selection and optimization of input features.

1.1 ANN-based Daylight Predictions

In the field of daylight design, ANNs have been used as prediction models for luminous efficacy (Lopez and Gueymard, 2007), sky luminance and sky irradiance (Pattanasethanon, Lertsatitthanakorn and Atthajariyakul, 2008; Janjai and Plaon, 2011), and horizontal internal illuminance levels (Kazanasmaz, Gunaydin and Binol, 2009). The ANN models relied on historic data as training data, which constituted an obstacle in terms of the time and the effort required to generate the training data. The need of data to be collected over a longer period of time - e.g. around 3 months as in Kazanasmaz et al. (2009) - also undermines the feasibility of applying ANNs in the first place. One line of research has circumvented this problem by applying ANNs in the optimization of building design, thereby training the ANN models on data extracted from a fraction of simulations required as part of the optimization process. Magnier and Haghihat (2010) showed that simulation-based ANNs could be used to pass performance results to the fitness function of a Genetic Algorithm to considerably reduce simulation time. In further optimization studies, ANNs were used to predict electric energy consumption and visual comfort (Wong, Wan and Lam, 2010; Kim, Jeon and Kim, 2016). Concerning climate-based metrics, Zhou and Liu (2015) were able to predict the specific illuminance range of the UDI (Useful Daylight Illuminance). The studies undertaken typically predict hourly or point-in-time daylight levels. Thus, annual culminative predictions of the DA metric has been understudied. The potential for predicting DA metric was introduced in the authors' previous work (Lorenz *et al.*, 2018) and is reinvestigated for a more complex design scenario. The input features so far had been selected empirically. As a result, the need for a robust, automated and replicable method of improving the input feature selection was identified.

1.2 Input Feature Selection and Optimization

Input features, also known as predictor variables, make up the training data which is passed to ANN models in order to facilitate supervised, unsupervised and reinforcement type learning. Input feature selection refers to the reduction of the dimensionality of training data by selecting a subset of input features. An objective function is commonly used as selection criterium to minimize the predictive error and thereby identify an optimal or suboptimal subset of input features. Such input feature selection methods have been shown to improve prediction accuracy and identify the minimum number of features needed (Marcano-Cedeño *et al.*, 2010; Özşen, 2013).

As exhaustive search methods are computationally expensive, two other common methods of input feature selection are Sequential Forward Selection (SFS), and Sequential Backward Selection (SBS). SFS is a bottom-up approach that starts with an empty set of features to which features are iteratively added (Whitney, 1971). Its counterpart, SBS starts from the complete set from which features are iteratively removed (Marill and M. Green, 1963). Inevitably, both methods inhibit a nesting problem, whereby potentially important features, once removed, cannot be re-introduced. To ensure that important features would be kept alive, mixed method approaches were introduced (Pudil, Novovi and Kittler, 1994). A comparative evaluation of methods is given in (Zongker and Jain, 1996). As forward-based methods have been shown to be faster than backward-based methods, this paper works with a forward-sequential approach in combination with a machine learning algorithm, which was used to determine the sequence of input features.

2. Research Methodology

The different steps to developing, optimizing and validating ANN models are illustrated in Figure 1. The ANNs were employed to predict daylight for a design solution space. As such, performance data was collected from selected samples of the design solution space. The performance data was extracted from daylight simulations and recorded features describing the design changes and the corresponding daylight results. While a majority of data was used for training the ANN models, a part of it was retained for validation of the models.

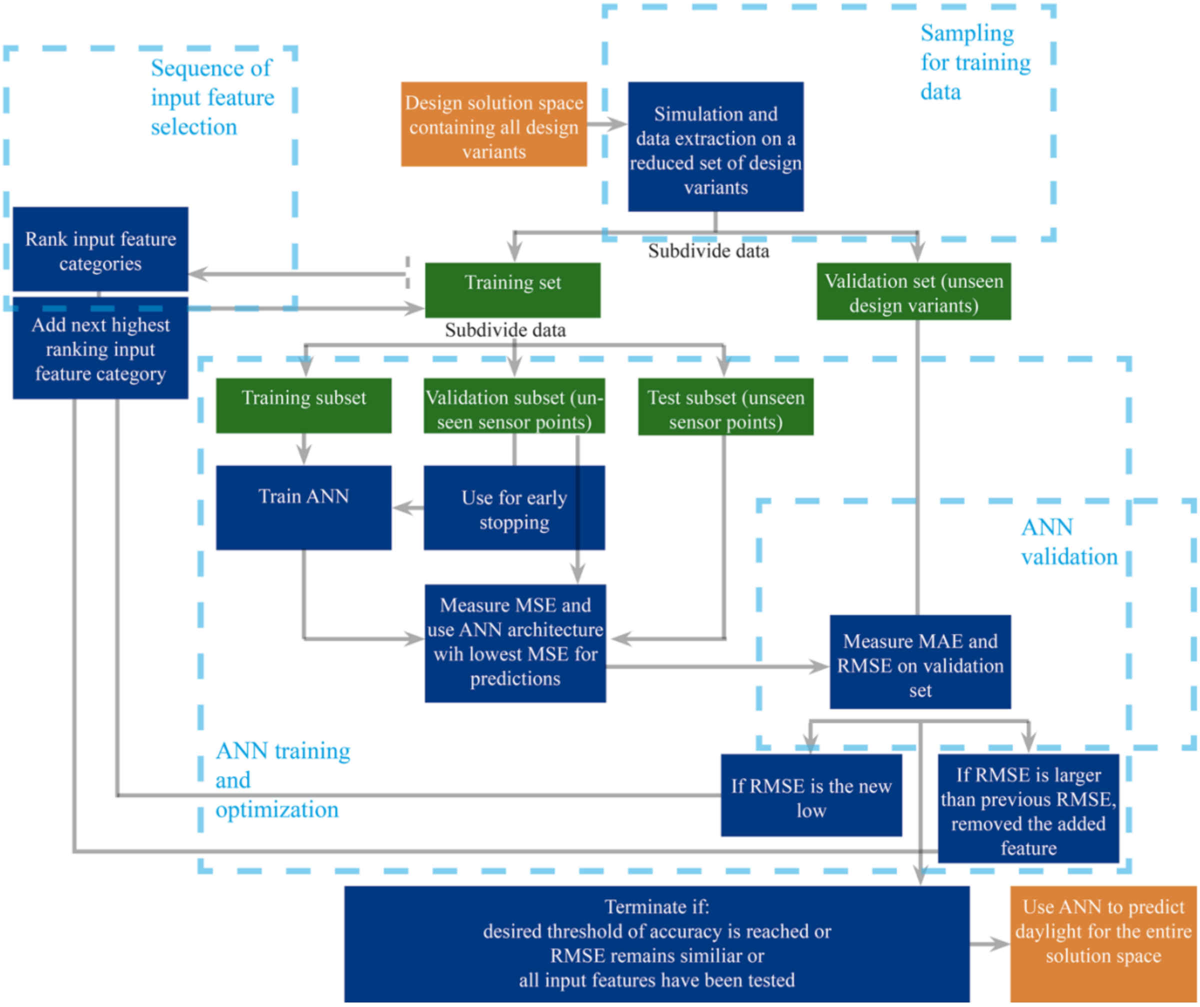

Figure 1: ANN model development and validation

Once a training set had been generated, bagged decision trees, a machine learning technique, were used to determine a sequence for the forward-sequential selection of features. Thus, the ANN models were first trained with the input features that showed the highest impact as predictors. The training data set was again subdivided into a training subset, a validation subset and a test subset at the ratio of 65:25:15. The validation subset was used for early stopping to avoid overfitting and the test subset was used to estimate prediction accuracies on new cases. Both subsets were also used to assist in the optimization of the network architecture. Each network architecture was trained ten times with the initial weight settings and distribution of samples across the subsets varying in each training run. The network architecture with the lowest error across ten training runs was used for predictions, and the output of all ten networks was averaged to improve generalization.

The predictions were validated against the simulations withheld from training. In the case of insufficient accuracy, another training feature was added to the training set and the optimization of network architecture, and validation was repeated. There are several options for terminating the training cycle: once a desired threshold of accuracy is reached, once the accuracy converges for a number of training cycles, once all input features have been tested, or a combination of the three. In this study we tested all input features to allow for a more wholistic evaluation of the method. However, if the aim is not to find an optimal subset of features but to identify the minimum number of input features required to reduce the training time, the first two options of termination better serve the purpose.

2.1 Design Solution Space and Design Variables

The exploration of design solutions was done for the central atrium of a school-building in Hamburg (Figure 2). As a first design variable, the atrium base dimension was reduced from 225m^2 to 56,25m^2, thereby splaying the atrium well walls in obtuse angles between 90 and 79°. With a second design variable, the atrium well was slanted, changing the orientation between north and south. 9 possible solutions were specified for the slant of the atrium well, modifying its splay angles between 58 and 104°. This resulted in a 9 by 6 matrix of 54 possible design variants for the atrium well geometry. As the third design variable, 3 possibilities were specified for the window-to-wall ratios (WWR) across floor levels of the 6-storey building. The WWR was reduced from the ground to higher floor levels in order to increase the reflected light and daylight levels on the ground floor (Samant, 2017). In a first option, the WWR distribution was set to 50% -6th floor, 60% - 5th floor, 70% - 4th floor, 80%- 3rd floor, 90% - 2nd floor and 100% - ground floor. In a second option, the WWR distribution was set to 20% -6th floor, 35% - 5th floor, 50% - 4th floor, 65%- 3rd floor, 80% - 2nd floor and 100% - ground floor. Lastly, in a third option, the WWR distribution was set to 20% -6th floor, 30% - 5th floor, 40% - 4th floor, 50%- 3rd floor, 60% - 2nd floor and 100% - ground floor. The entire solution space thereby contained 162 design variants for the central atrium.

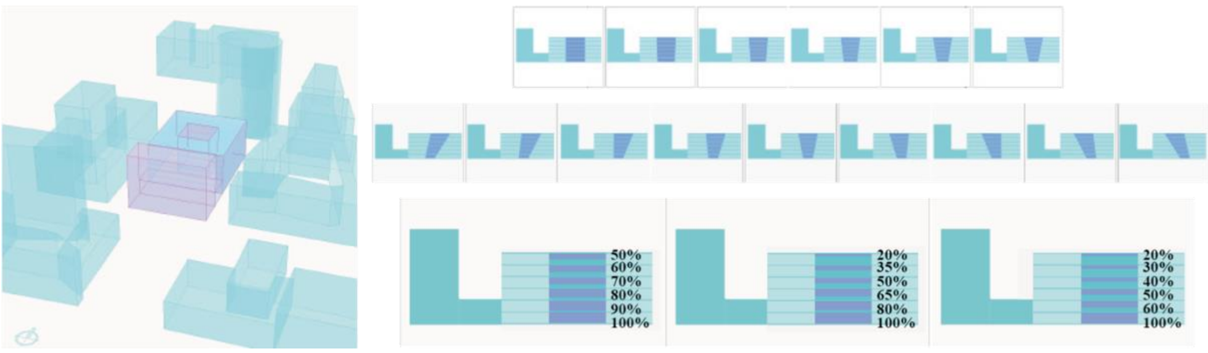

Figure 2: To the left: The school building and surrounding buildings. To the right, from top to bottom: 6 variants for a reduced atrium base dimension, 9 variants for the atrium well orientation, 3 variants for the WWR distribution across floor levels (the atrium well is highlighted in dark blue)

2.2 Daylight Simulation and Data Extraction

From the solution space of 162 variants, a reduced set of 36 variants were selected to provide training data and another 21 variants were selected for validation of ANN accuracy. The architectural models were built in Grasshopper, and daylight simulation on the selected variants were run in Diva – a radiance-based and validated software (Mohsenin and Hu, 2015). DA was calculated for sensor points at a work-plane height of .8 m above floor level and the sensor points were spaced in .6 m distance of each other. The input features were extracted for every sensor point and passed to the input layer of the network model (Figure 3). The corresponding DA levels were passed to the output layer of the network for it to undergo supervised training. As shown in Figure 3, 26 input features in total were extracted from the simulation model and grouped according to categories. This was done for the next step, in which a sequence for feature selection was determined.

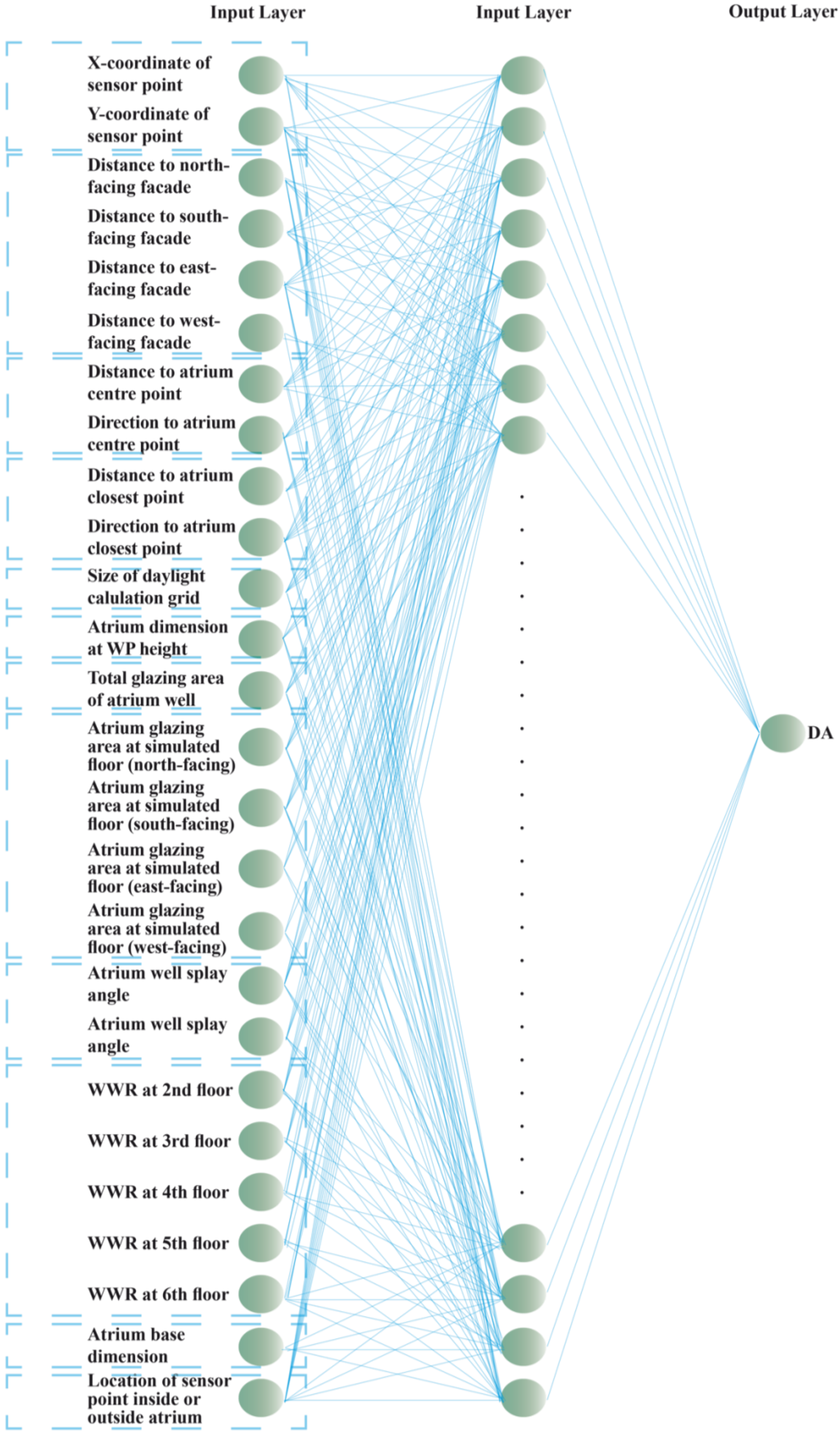

Figure 3: Representation of ANN model as a construct of neurons with an input layer, hidden layer and output layer. The extracted input features are passed to the input layer of the model and the daylight performance data (in DA) to the output layer. A suitable number of hidden neurons needs to be identified as these effect the ANN training time and prediction accuracy.

2.3 Ranking of Input Feature Categories

In order to rank the input feature categories according to their impact as predictors, the training data set was passed to machine learning (ML) models for fitness approximation. Although ANNs can be trained to identify significant features, this was not done due to their increased computation time. Among the tested ML techniques were linear regression models, fine, medium and coarse trees, boosted tree ensembles, liner, quadratic, cubic, fine Gaussian support vector machines and Gaussian Process Regression models. Bagged decision trees showed the lowest root mean squared error (RMSE) and superior performance compared to the other models. Additionally, the computation time was low, with approximation taking around four minutes. Hence, they were selected to identify a feature selection sequence.

From the data set that was passed to the bagged trees model, every input feature category was individually removed and the corresponding variance in RMSE was measured. The features were then ranked according to the variance they inflicted on the error, with those features causing the largest variance ranked highest. The RMSE during approximation on the training data set and resulting sequence of input feature categories is given in Table 1.

Table 1: Ranking of input feature categories according to variance in prediction error

Input Feature Category	**Sequence Order**	**RMSE**	**Input Feature Category**	**Sequence Order**	**RMSE**
Distance and direction to atrium closest point	1	1.19	X-, Y-coordinates of sensor points	7	.87
Distance to facade	2	.99	Glazing area at simulated floor level	8	.85
Distance and direction to atrium centre point	3	.96	Glazing area across all floors	9	.84
WWR	4	.90	Atrium base dimension	10	.84
Location of sensor point inside or outside atrium	5	.90	Atrium dimension at WP height	11	.84
Atrium well spay angles	6	.89	Daylight calculation grid size	12	.84

2.4 Sequential Feature Selection and Validation

The ANN models were trained starting from a data set with one input feature category. Once trained, the ANN model was used to predict the DA metric for the 21 retained variants and predictions were compared to the simulated DA. Two measures were used: the mean absolute error (MAE) and the root mean square error (RMSE). The MAE, which gives the absolute difference between the simulated and predicted values, was chosen for its ease of interpretation. The RMSE was selected as it weighs larger errors more heavily and is a commonly used measure of accuracy. The RMSE was observed for each added input feature and its minimization was used as the objective function to optimise the selection. As long as the RMSE reached a new low, one additional input feature category was added to the training data set. If the RMSE worsened, the last added feature was removed before adding the next feature in the sequence. In this way, all input features were added to the training data set at some point in a forward-sequential manner. The results discuss the observed errors in accuracy for every added feature.

2.5 ANN Training

Back-propagation ANN models were employed in conjunction with the Lavenberg-Marquardt algorithm. The training parameters were set to an initial mu of 1, a mu decrease factor of .8 and mu increase factor of 1.5. The training was run for 200 epochs, during which the connection strengths between neurons were adjusted to minimise the mean squared error (MSE) on the training data set. The training data set was again divided into a subset of training, validation and test data in order to ensure robustness of the trained networks through cross-validation and early stopping. The maximum number of validation failures was set to 6. Ten network architectures with 38 to 40 neurons in the hidden layer were trained and tested with said settings and the network architecture with the lowest MSE on the training, validation and test subset was used for predictions. This was done to ensure that a network optimised architecture was used to evaluate the input features.

3. Results

The MSE on the training data set (36/162 simulations with 150.706 sensor point data samples) and the RMSE and MAE on the validation set (21/162 sim with 87.850 sensor point data samples) were recorded at each training stage of the sequential search. Figure 4 shows the MSE during feature selection (for the sequence refers back to Table 1). Figure 5 shows the MAE obtained on the validation set and Figure 6 the RMSE. The minimum achieved error has been highlighted in red. The sequence at which an added input feature category was again removed from the training data has been highlighted in yellow.

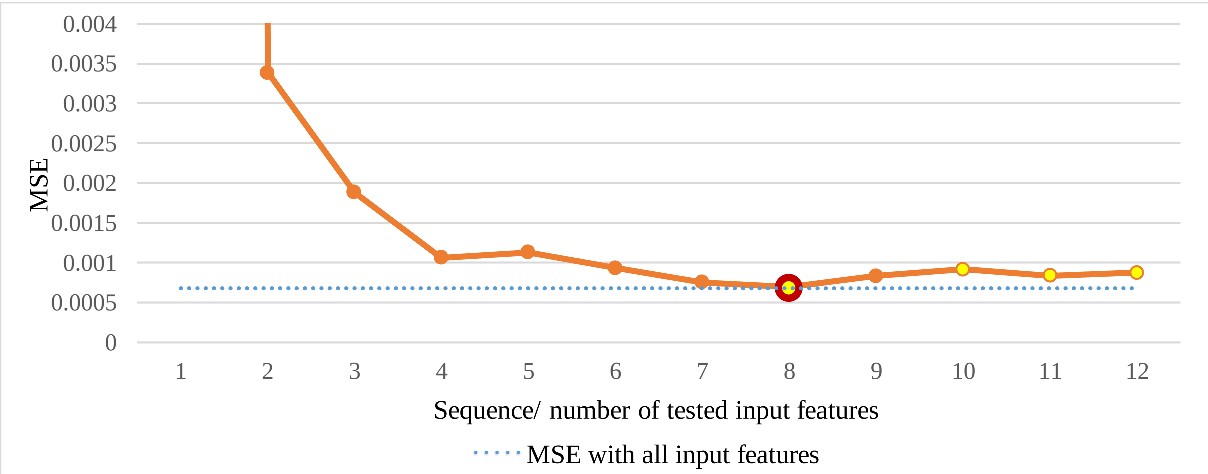

Figure 4: MSE on the training data set after every added input feature category

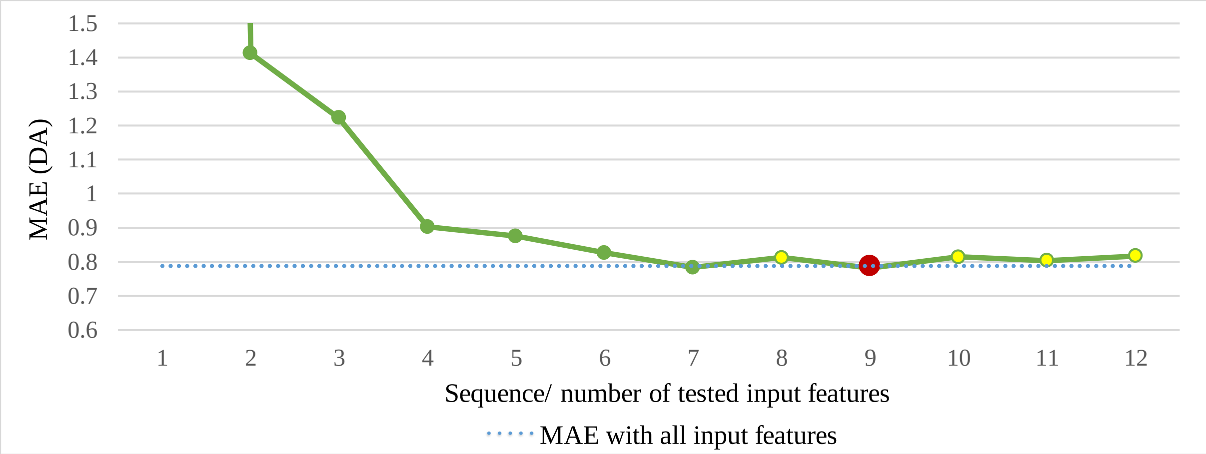

Figure 5: MAE on the training data set after every added input feature category

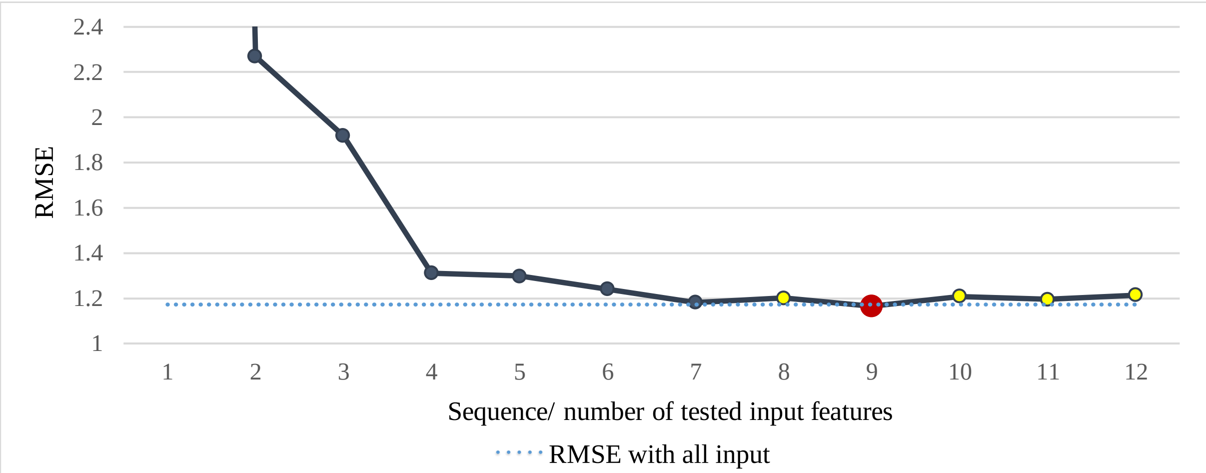

Figure 6: RMSE on the training data set after every added input feature category

The MSE dropped below 0.01 after including six input features (which include distance and direction to atrium closest point, distance to façade, distance and direction to atrium center point, WWRs, location of sensor points inside or outside atrium, and atrium well spay angles). The MAE and RMSE remained in the similar range after the 7th input feature category, the sensor point identifiers, were added as training data. The lowest MSE was reached after adding the 8th input feature describing glazing areas on the simulated floor level. The MAE and RMSE however increased, resulting in the feature being removed from the feature set. Consecutively, the MAE and RMSE reached their minimum at .78 MAE and 1.16 RMSE. In comparison, the ANNs trained with all input features showed marginally higher errors of .79 MAE and 1.17 RMSE. Although this shows that the input feature selection could not result in a significant improvement of accuracies, it highlights that not all input features are needed and overall training time can be reduced.

In fact, training time of the 'optimal' feature set with 8 added and one removed input feature category (19 individual features) including network optimization took 04:33 (hh:mm), whereas the training and optimization of the model that included all input features was 05:55. Taking a more minimal approach, the training with 7 added feature categories (18 individual features) was 03:01. Predictions on the validation set were made in less than 1 second.

Inadvertently, due to nature of the selection process, it lacks in feasibility compared to a trial and error approach validating an empirically selected feature set. The above outlined time savings can therefore only truly be achieved if: a) the selected input feature subset is extracted and trained for a larger data set, b) the feature subset holds validity for a new or similar design scenario, or c) the same input features would be selected with a smaller and thereby less computationally expensive network architecture of fewer hidden neurons. This however would still need validation. The training times as measured here were taken on a 2.6 GHz Intel Core i9 processor.

The prediction accuracy converged around .8 DA MAE, meaning that this was, on average, the absolute difference between the simulated and predicted DA. The simulated DA range from 0 to 88%, with 1% referring to 1% of occupied hours in a year. A difference of .8 DA can therefore hardly be interpreted and constitutes a negligible error. The RMSE, which converged around 1.2%, supports this finding and shows a high accuracy of ANNs in predicting the DA metric.

In order to achieve above-mentioned accuracies, data was extracted from a total of 57/162 simulations for training and validation. As daylight simulations took approximately 3 hours per design variant, a total of 315 hours of simulations were bypassed through ANN predictions. Given that the process relied on a small sample size of 13% for validation, 16,4% when including the validation subset within the training data, the results may vary. However, increasing the validation set within the ANN-integrated workflow would require a larger number of simulations to be run, thereby reducing the achievable time-savings.

4. Conclusion and Recommendations for Future Research

Overall, the following conclusions for ANN-based daylight predictions can be drawn:

a) On average, prediction accuracies of around .8 DA mean absolute error (MAE) were achieved. These accuracies could be maintained as the number of input features increased.

b) The MSE on the training data set did not directly correlate with the MAE and RMSE on the validation data set, as the prediction accuracies could improve (lower MAE and RMSE) even though the ability of the network to fit the data decreased (higher MSE).

d) Having superfluous input features did not significantly lower accuracies (e.g. calculation grid size and atrium dimension at WP height). It did however increase training time. On the other hand, too few input features (e.g. before sequence 4), though requiring less training time, compromised prediction accuracies.

e) Using the proposed method, accuracies barely improved compared to the empirically selected full set of input features. However, the training time for the ANN models could be reduced by around 50% without significantly compromising the accuracies.

The paper proposes a viable method for selecting input features useful for predicting daylight in buildings. Additionally, the study investigated in how far improvements could be achieved. Although, in terms of accuracy, those were marginal in this study, the selection method may prove more valuable in larger and more complex solution spaces with a larger number of variables. A downside of the proposed method is that it remains computationally expensive, as it requires multiple training runs with already computationally demanding ANN models. It would therefore be useful to evaluate the results of feature selection using smaller and more feasible network architectures, or completely rely on computationally less expensive ML models for feature selection. A comparison to alternative feature selection methods is recommended, as well as the application of the proposed method on more complex solution spaces.

References

Figueiro, M. G., Steverson, B., Heerwagen, J., Kampschroer, K., Hunter, C. M., Gonzales, K., Plitnick, B. and Rea, M. S. (2017) 'The impact of daytime light exposures on sleep and mood in office workers', *Sleep Health*. National Sleep Foundation., 3(3), pp. 204–215. doi: 10.1016/j.sleh.2017.03.005.

Hu, J. and Olbina, S. (2011) 'Illuminance-based slat angle selection model for automated control of split blinds', *Building and Environment*. Elsevier Ltd, 46(3), pp. 786–796. doi: 10.1016/j.buildenv.2010.10.013.

Janjai, S. and Plaon, P. (2011) 'Estimation of sky luminance in the tropics using artificial neural networks: Modeling and performance comparison with the CIE model', *Applied Energy*. Elsevier Ltd, 88(3), pp. 840–847. doi: 10.1016/j.apenergy.2010.09.004.

Jones, N. L. and Reinhart, C. F. (2019) 'Effects of real-time simulation feedback on design for visual comfort', *Journal of Building Performance Simulation*. Taylor & Francis, 12(3), pp. 343–361. doi: 10.1080/19401493.2018.1449889.

Kazanasmaz, T., Gunaydin, M. and Binol, S. (2009) 'Artificial neural networks to predict daylight illuminance in office buildings', *Building and Environment*, 44, pp. 1751–1757. doi: 10.1016/j.buildenv.2008.11.012.

Kim, W., Jeon, Y. and Kim, Y. (2016) 'Simulation-based optimization of an integrated daylighting and HVAC system using the design of experiments method', *Applied Energy*. Elsevier Ltd, 162, pp. 666–674. doi: 10.1016/j.apenergy.2015.10.153.

Lopez, G. and Gueymard, C. A. (2007) 'Clear-sky solar luminous efficacy determination using artificial neural networks', *Solar Energy*, 81(7), pp. 929–939. doi: 10.1016/j.solener.2006.11.001.

Lorenz, C.-L., Packianather, M., Spaeth, A. B. and Bleil de Souza, C. (2018) 'Artificial Neural Network-Based Modelling for Daylight Evaluations', in *Symposium on Simulation for Architecture + Urban Design*. Delft. The Netherlands.

Machairas, V., Tsangrassoulis, A. and Axarli, K. (2014) 'Algorithms for optimization of building design: A review', *Renewable and Sustainable Energy Reviews*, 31(1364), pp. 101–112. doi: 10.1016/j.rser.2013.11.036.

Maesano, C. and Annesi-Maesano, I. (2012) *Impact of Lighting on School Performance in European Classrooms*. Paris. Available at: http://velcdn.azureedge.net/~/media/com/articles/pdf/light and performance_whitepaperfinal 1.pdf (Accessed: 29 August 2017).

Magnier, L. and Haghighat, F. (2010) 'Multiobjective optimization of building design using TRNSYS simulations, genetic algorithm, and Artificial Neural Network', *Building and Environment*. Elsevier Ltd, 45(3), pp. 739–746. doi: 10.1016/j.buildenv.2009.08.016.

Marcano-Cedeño, Quintanilla-Domínguez, J., Cortina-Januchs, M. G. and Andina, D. (2010) 'Feature selection using Sequential Forward Selection and Classification applying Artificial Metaplasticity Neural Network', *IECON 36th Annual Conference on IEEE Industrial Electronics Society*, pp. 2845–2850. doi: 10.1109/IECON.2010.5675075.

Marill, T. and M. Green, D. (1963) 'On the effectiveness of receptors in recognition systems', *Information Theory, IEEE Transactions on*, 9, pp. 11–17. doi: 10.1109/TIT.1963.1057810.

Mckee, S. A., Schulz, M. and Caruana, R. (2006) 'Efficiently Exploring Architectural Design Spaces via Predictive Modeling', in *Asplos International Conference on Architectural Support for Programming Languages and Operating Systems*. CA, USA, pp. 195–206.

Mohsenin, M. and Hu, J. (2015) 'Assessing daylight performance in atrium buildings by using Climate Based Daylight Modeling', *Solar Energy*, 119, pp. 553–560. doi: 10.1016/j.solener.2015.05.011.

Nguyen, A.-T., Reiter, S. and Rigo, P. (2014) 'A review on simulation-based optimization methods applied to building performance analysis', *Applied Energy*, 113, pp. 1043–1058. doi: 10.1016/j.apenergy.2013.08.061.

Özşen, S. (2013) 'Classification of sleep stages using class-dependent sequential feature selection and artificial neural network', *Neural Computing and Applications*, 23(5), pp. 1239–1250. doi: 10.1007/s00521-012-1065-4.

Pattanasethanon, S., Lertsatitthanakorn, C. and Atthajariyakul, S. (2008) 'An accuracy assessment of an empirical sine model , a novel sine model and an artificial neural network model for forecasting illuminance / irradiance on horizontal plane of all sky types at Mahasarakham , Thailand', 49, pp. 1999–2005. doi: 10.1016/j.enconman.2008.02.014.

Pudil, P., Novovi, J. and Kittler, J. (1994) 'Floating search methods in feature selection', *Pattern Recognition*

Letters, 15(November), pp. 1119–1125.

Reinhart, F, C., Mardaljevic, J. and Rogers, Z. (2013) ‘Dynamic Daylight Performance Metrics for Sustainable Building Design’, *The journal of the illuminating Engineering Society of North America*, 3(1), pp. 7–31. doi: 10.1582/LEUKOS.2006.03.01.001.

Reinhart, C. F. and Walkenhorst, O. (2001) ‘Validation of dynamic RADIANCE-based daylight simulations for a test office with external blinds’, *Energy and Buildings*, 33(7), pp. 683–697. doi: 10.1016/S0378-7788(01)00058-5.

Samant, S. (2017) ‘Atrium and its adjoining spaces: a study of the influence of atrium façade design Swinal Samant Atrium and its adjoining spaces: a study of the influence of atrium faç ade design’, *Architectural Science Review*, 54(4). doi: 10.1080/00038628.2011.613640.

Veitch, J. A., Christoffersen, J. and Galasiu, A. D. (2013) *Daylight and View through Residential Windows : Effects on Well-being*, *Residential dayligting and well-being*.

Whitney, A. W. (1971) ‘A Direct Method of Nonparametric Measurement Selection’, *Computers, IEEE Transactions on*, 20, pp. 1100–1103. doi: 10.1109/T-C.1971.223410.

Wong, S. L., Wan, K. K. W. and Lam, T. N. T. (2010) ‘Artificial neural networks for energy analysis of office buildings with daylighting’, *Applied Energy*. Elsevier Ltd, 87(2), pp. 551–557. doi: 10.1016/j.apenergy.2009.06.028.

Zhao, H. X. and Magoulès, F. (2012) ‘A review on the prediction of building energy consumption’, *Renewable and Sustainable Energy Reviews*. Elsevier Ltd, 16(6), pp. 3586–3592. doi: 10.1016/j.rser.2012.02.049.

Zhou, S. and Liu, D. (2015) ‘Prediction of Daylighting and Energy Performance Using Artificial Neural Network and Support Vector Machine’, *American Journal of Civil Engineering and Architecture, Vol. 3, 2015, Pages 1-8*, 3(3A), pp. 1–8. doi: 10.12691/AJCEA-3-3A-1.

Zongker, D. and Jain, A. (1996) ‘Algorithms for Feature Selection: An Evaluation’, in *Proceedings of 13th International Conference on Pattern Recognition*. Vienna, Austria, pp. 18–22.