




## Article

# Machine Learning-Driven Calibration of Traffic Models Based on a Real-Time Video Analysis

Ekaterina Lopukhova <sup>1,\*</sup>, Ansaf Abdunagimov <sup>2</sup>, Grigory Voronkov <sup>1</sup> and Elizaveta Grakhova <sup>1</sup>

<sup>1</sup> School of Photonics Engineering and Research Advances (SPHERA), Ufa University of Science and Technology, 32 Z. Validi Street, Ufa 450076, Russia; voronkov.gs@ugatu.su (G.V.); grakhova.ep@ugatu.su (E.G.)

<sup>2</sup> Department of Automated Control Systems, Ufa University of Science and Technology, 32 Z. Validi Street, Ufa 450076, Russia; abdunagimov.ai@ugatu.su

\* Correspondence: lopukhova.ea@ugatu.su

**Abstract:** Accurate traffic simulation models play a crucial role in developing intelligent transport systems that offer timely traffic information to users and efficient traffic management. However, calibrating these models to represent real-world traffic conditions accurately poses a significant challenge due to the dynamic nature of traffic flow and the limitations of traditional calibration methods. This article introduces a machine learning-based approach to calibrate macroscopic traffic simulation models using real-time traffic video stream data. The proposed method for creating and calibrating a traffic simulation model has significantly improved the statistical correspondence between the generated vehicle characteristics and real data about cars on the simulated road section. The correspondence has increased from 37% to 73%. Machine learning models trained on generated data and tested on real data show improved accuracy rates. Mean absolute error, mean square error, and mean absolute percentage error decreased by more than two orders of magnitude. The coefficient of determination has also increased, approaching 1. This method eliminates the need to deploy wireless sensor networks, which can reduce the cost of implementing intelligent transport systems.

**Keywords:** traffic simulation models; connected car; calibration; V2I; machine learning; intelligent analysis of the video stream



**Citation:** Lopukhova, E.; Abdunagimov, A.; Voronkov, G.; Grakhova, E. Machine Learning-Driven Calibration of Traffic Models Based on a Real-Time Video Analysis. *Appl. Sci.* **2024**, *14*, 4864. <https://doi.org/10.3390/app14114864>

Academic Editors: Eui-Nam Huh and Suchao Xie

Received: 9 April 2024

Revised: 24 May 2024

Accepted: 1 June 2024

Published: 4 June 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The current level of development in communication, information, and sensor technologies enables the automation of monitoring and managing complex systems that previously required expert intervention. This automation enhances scalability and responsiveness [1–3].

Scalability, real-time monitoring, and control are increasingly feasible and essential for transport systems, enhancing their safety and efficiency [4]. Intelligent transport systems (ITSs) can enable the interconnected operation of monitoring, control, and automation systems within the transport environment [5].

### 1.1. Implementation of Intelligent Transport Systems via Simulation Model of Traffic Flow

ITS utilizing vehicle-to-everything (V2X) technology [6] meet the demands of modern transport systems [7–9] by continuously collecting data on road traffic through various sensors [4]. The collected and systematized data, which represent the context of traffic situations, are crucial for determining the evolving requirements of ITS functionality over time [10]. Context data play a vital role in analyzing sensor data, reasoning, and modeling intelligent systems, significantly impacting the effectiveness of ITSs. Two concepts and their interaction can be used to visually represent the real context of the traffic situation and its display in the ITS: physical and virtual transport space (PTS and VTS). The interaction diagram of the PTS and VTS within the metaverse transportation system offers a comprehensive representation of the context of the entire transport system [11].

The PTS environment is the data source that the ITS sensor system is designed to collect. The VTS is created by simulating the processes occurring in the PTS by processing sensor system data. The VTS can generate data similar to the PTS but with controlled process dynamics parameters and various associated factors. The VTS supports computational experiments with feedback optimization relative to the real PTS [11], which connects the operation of many systems at a high level and controls data generation.

However, the representativeness of data generated using VTS depends on the accuracy of the correspondence between real and simulated processes. In VTS, these PTS processes are modeled using the simulation model of traffic flow (SMTF). The SMTF generates random events representing vehicle appearances on simulated routes based on the context of traffic situations embedded in the model. Predetermined parameters and time within the simulated space define this context [12], influencing movement scenarios, interactions, traffic density, and infrastructure behavior on the road.

The choice of model for implementing SMTF is primarily determined by the type of PTS processes that require mapping to the VTS. Existing SMTF models can be divided into macroscopic, microscopic, and mesoscopic models [13].

Macroscopic models study the transport flow as a whole, making it possible to use elements of the physical phenomena analysis [14,15]. These models describe traffic flow characteristics like speed, density, and intensity. Macroscopic models depict general trends and are suitable for efficient analysis of large networks, but they cannot account for the behavior and interactions of individual drivers [14]. Microscopic models can represent the behavior of vehicles and drivers. These models are computationally intensive and suitable for studying detailed situations [14,16,17]. On the other hand, mesoscopic models represent traffic as transport packets, combining the detailed information of microscopic models with the statistical approach of macroscopic models [14].

To construct a general SMTF for a limited section of the highway without the need for a detailed analysis of vehicle behavior, the most suitable option would be to use a macroscopic model [14]. This solution requires less input data than microscopic and mesoscopic models [18] while providing sufficient accuracy in modeling the main PTS processes [13,19]. Macroscopic traffic models focus on vehicle behavior and statistical characteristics. Therefore, most implementations of these models concentrate on depicting the dynamics of a specific section of the route based on available data [20].

A vital element of the macroscopic model is the demand model. This concept is taken from the queuing theory and, in the context of the theory of transport flows, is characterized as the origin–destination Matrix (ODM) [21]. The ODM determines the number of trips between each pair of zones or nodes in the network, directly impacting the mapping of the spatio-temporal characteristics of the PTS processes into the VTS. It is essential to consider that the demand for different routes changes over time, showing clear patterns during peak and off-peak periods [21–23]. The ODMs require calibration to account for deviations over a more extended period [22,24,25]. Thus, SMTF with a calibrated ODM is crucial for accurately modeling PTS processes in VTS.

Various software solutions are available to simplify the construction of a macroscopic (and mesoscopic) model (with ODM calibration capability). One notable example is the simulation of urban mobility v. 1.20.0 (SUMO) software, which is widely used in research fields, particularly in traffic management and transportation communications [26]. SUMO offers capabilities for constructing, adjusting, and modeling custom traffic flow architectures and supporting tools for importing network and ODMs of existing route sections. SUMO's versatility enables the study of traffic control strategies and vehicular communication systems [27]. The high-detail simulation in SUMO is essential for analyzing and optimizing traffic management strategies, evaluating the impact of infrastructure modifications, and examining vehicle interactions in complex urban environments [28]. Additionally, SUMO supports V2X technology modeling, facilitating the development and testing of interoperable automotive applications [29].

However, a significant drawback of such models is their challenge in adapting to dynamic changes in the PTS [30]. The accuracy of the simulated model data may differ significantly from that of real-life PTS data, since the simulation's source data rely on information from the OpenStreetMap project. Updating the OpenStreetMap database can be slow and dependent on the location and activity of volunteers in the relevant region [31].

The ODM and vehicle behavior parameters show the most significant differences from real-world traffic data in PTS. The ODM affects how traffic density is distributed over space and time, which can be monitored using data from ITS sensor systems [32]. To address discrepancies between the outdated ODM and current traffic conditions, adjustments can be made to the spatial and temporal distribution of traffic density in the SMTF through calibration of simulation models.

Various calibration techniques can help bridge the gap between the simulated model and actual traffic conditions in PTS. These include direct calibration of the ODM [33–37], calibration of the macroscopic model based on physical processes [38–40], and the use of evolutionary algorithms to optimize calibration parameters [41–43]. Studies suggest that relying on a single traffic flow model may not capture all relevant phenomena accurately, potentially impacting simulation accuracy [13]. Hybrid traffic flow models, combining microscopic, mesoscopic, and macroscopic approaches [18,44–47], offer a more comprehensive representation of traffic dynamics with varying levels of scalability.

Receiving transport system data from a PTS or VTS opens up the possibility of using machine learning (ML) algorithms for both SMTF calibration and applications that can be implemented in the VTS.

### *1.2. Machine Learning for Intelligent Transport System Technologies*

The continued development of automation technologies and ITS applications implies using ML methods to analyze information about the transport system and make decisions based on selected contextual data [48]. ML methods are used in ITS applications to dynamically analyze traffic conditions, enhance operational precision, minimize delays, and ensure the safety of road users.

The use of applications for analyzing traffic congestion and providing real-time alternative routes has become widespread in ITSs [49–52]. The results of their application show that using ML methods can reduce the waiting time during peak loads by half and detect congestion during unexpected road closures almost five times faster [49]. ML techniques have been successfully used to reduce latency, manage resources, and optimize the performance of ITS applications [53–55]. They also help reduce power consumption and improve ITSs bandwidth efficiency [54].

Improving highway safety is one of the most developed areas of ML application in ITSs. It includes security analysis of the road network and its elements [56–58], improving information security in ITSs [59,60], and detecting anomalies in sensor data of connected and autonomous vehicles [61–63]. It also involves intelligent analysis of road traffic accidents to identify critical factors in their occurrence, essential for preventing similar incidents in the future [64]. ML for big data analysis enables the effective development and calibration of SMTFs. Integrating macroscopic modeling with ML techniques can enhance the accuracy of traffic flow modeling by improving prediction accuracy and adaptability to changing traffic conditions [65–67]. This hybrid approach can provide greater flexibility to changing traffic patterns by combining the overall traffic flow dynamics captured by macroscopic models with ML's ability to handle dynamic changes effectively. By leveraging ML methods instead of traditional statistical approaches, traffic algorithm modeling can achieve higher accuracy and adaptability [68].

In [69], it is demonstrated that ML can be helpful in dynamically calibrating parameters in macroscopic traffic flow models. For instance, ML can generate a set of hyperparameters for calibration through generative adversarial simulation learning. However, a challenge arises in selecting a data source to timely calibrate the VTS to reflect changes in the PTS.

Creating ML applications requires training and testing data corresponding to the real-world processes with which the applications interact [70]. Moreover, the data collection process is a non-trivial task requiring a separate approach.

### *1.3. Capturing Intelligent Transport System Data to Develop Machine Learning Applications*

A dataset for training and testing ML models can be obtained from the ITS sensor system (PTS) and the SMTF, which is part of the VTS. One increasingly popular method for collecting traffic flow data from specific highway sections in PTS is the use of wireless sensor networks (WSNs) due to their flexibility, scalability, and cost-effectiveness, as highlighted in [71–76]. WSNs allow data collection on parameters recorded by sensors from each vehicle connected to the communication network, providing more detailed information about the analyzed route section than analyzing traffic video streams [77]. These richer data can help address a broader range of ITS problems through ML.

Despite the benefits of using WSNs, video recording systems, especially those based on autonomous aerial vehicles, are becoming more prevalent [78,79] thanks to advancements in ML methods for processing video streams [80] and the proliferation of city cameras. This trend enables the creation of extensive and affordable systems for collecting transport network data to build and calibrate simulation models [81–83]. Video systems allow data collection from all vehicles without specialized sensors or communication modules. On the other hand, integrating a large number of vehicles into a WSN can present challenges.

However, it is essential to note that a fixed dataset is required to address various challenges in constructing SMTF for specific route sections, conducting computational experiments, and pre-training ML models for integration into the VTS [84–86]. This dataset cannot always be directly obtained from video streams, and deploying WSNs may be hindered by the lack of vehicles equipped with transceiver systems and sensors.

The amount of actual data obtained from the PTS may be limited in representing various traffic flow phenomena and the amount of data. Obtaining it may also be complicated by technical limitations. Receiving data from the VTS will only be effective for developing an ML application if the generated data are comparable to the data from the PTS. Suppose the statistical characteristics of the simulated data do not match the statistical characteristics of the real data. In that case, the ML model trained on the simulated data may not perform effectively in real conditions.

When developing an ML application for a vehicle-to-infrastructure (V2I) system, we encountered a similar problem of lack of statistical proximity between VTS and PTS data. It required us to develop an approach to reducing this discrepancy in datasets, which is the topic of this paper. In our previous work [27], we showcased an application of ML in ITSs focusing on specific traffic flow data, based on which the location of the “connected car” was determined to generate a dedicated radio channel, tuning the phased antenna array radiation pattern. The approach involved generating synthetic data to train the ML algorithm within the V2I system, leveraging the SUMO simulation model and the Longley–Rice propagation model [87]. The primary objective was to reduce delays in the V2I network, including minimizing network connection time. By creating a method for generating diverse motion scenarios, the algorithm’s ability to generalize for antenna radiation tuning was enhanced.

The ML algorithm, trained on the synthetic dataset, exhibited a 94% accuracy in determining the optimal main lobe position crucial for establishing and sustaining a dependable connection between the V2I infrastructure and vehicles. However, when tested on real data extracted through intelligent processing of video streams from the simulated area, the accuracy dropped by over 40% compared to synthetic data. This discrepancy highlighted a significant deviation between the characteristics distribution generated by the simulation model and actual traffic flow data.

These limitations in the development and updating of the traffic simulation model, as well as in the collection of data for the integration of ML applications for the effective operation of VTS within the ITS, necessitate the creation of a single adaptive SMTF. This model should provide data for analysis and development of ML models and adapt to the dynamics of real traffic flow.

This paper proposes a solution to develop SMTFs that closely align with real-world PTS. The methodology involves a hybrid regression calibration of a macroscopic model using SUMO data, leveraging real-time video stream analysis with the ML algorithm to capture vehicle characteristics within the modeled area.

The paper is structured as follows: Section 2 outlines how to represent PTS dynamics within a VTS and details the process for creating and calibrating the proposed SMTF. Section 3 introduces a method for analyzing video streams and extracting relevant data. Section 4 presents simulation results demonstrating the algorithm's performance in calibrating a macroscopic traffic model with new data from the PTS. Additionally, it showcases an experiment assessing the calibrated SUMO model's alignment with real PTS based on video stream data and connected car sensors.

## 2. Analysis of the Influence of the Dynamics of the PTS on the Operation of the VTS

The processes within the transportation flow can vary depending on the specific focus of the study. When examining traffic flow on a specific section of the road, several overarching factors come into play. These include the overall road infrastructure both within and outside the section being studied [88–90], as well as temporary factors that can be cyclic, like changes in time of day and seasons [35,91,92], or random, such as variations in weather conditions [93].

These global and temporal factors also fluctuate and influence traffic flow characteristics to different extents, which can make simulation modeling quite complex [35,92]. To simplify this complexity, a limited set of traffic flow parameters is required to develop SMTF and assess the dynamics of each parameter.

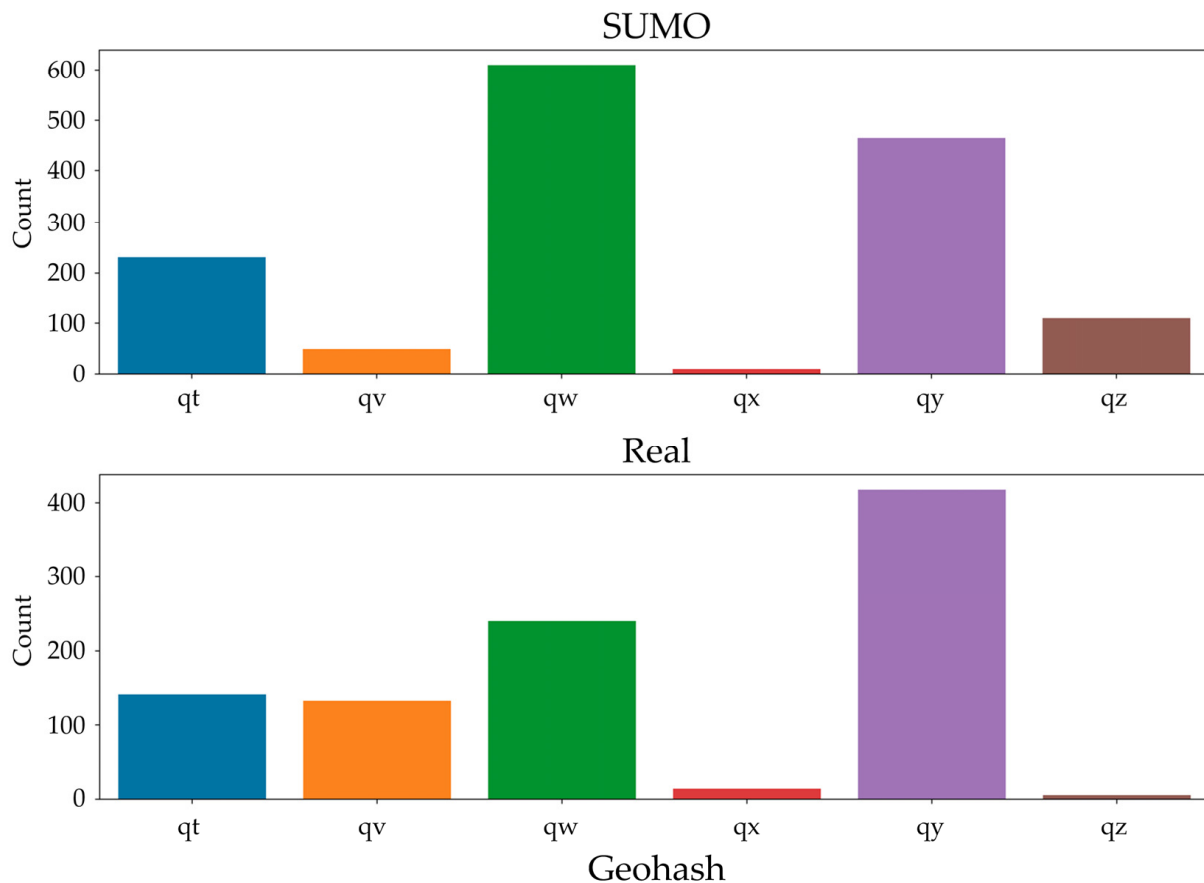
To compare the data generated in our work [29] using the SMTF we developed with real data from the PTS, we initiated data collection by analyzing the video stream of the simulated path section described in Section 3.

In the context of the previous work [27], the selected parameters for addressing the vehicle tracking issue were sensor data that reflect the internal attributes of vehicles while they are in motion. Since these vehicle sensor parameters exhibit a spatial and temporal distribution that aligns with the traffic ODM, analyzing these data offers valuable preliminary insights into the modeled area. Additionally, it supplements global navigation satellite system (GNSS) data in cases where GNSS data are inaccurate or unavailable due to interference or a lack of visible satellites [94,95].

The data used to train the developed algorithm for tracking vehicle movement included the time mark when a request was sent to connect to an intelligent infrastructure facility, the vehicle's speed, and its rotation angle relative to geographic north. This set of characteristics of vehicles participating in the transport flow demonstrates the dependence of their values on the spatial distribution in the PTS territory, which allows the clustering of many vehicles in this territory. In our application, the ML model classified these vectors, selecting the sector in the direction in which the main lobe of the radiation pattern should be directed to implement the connection or maintain a communication channel between the vehicle and the infrastructure object.

During the results analysis [27], a deviation in the ODM was specifically noted, resulting in fluctuations in the frequency of vehicle presence along different segments of the route overall and over specific time intervals. We divided the simulated area into sectors using the Geohash method [96] of GNSS coordinate space, encoding with a code sequence length of 8. Each sector measured 38.2 m by 19.1 m. This level of detail is considered optimal for V2I technology, which typically has an average range of 600 m [6]. The sector sizes were chosen to match the dimensions of route elements like intersections and turns.

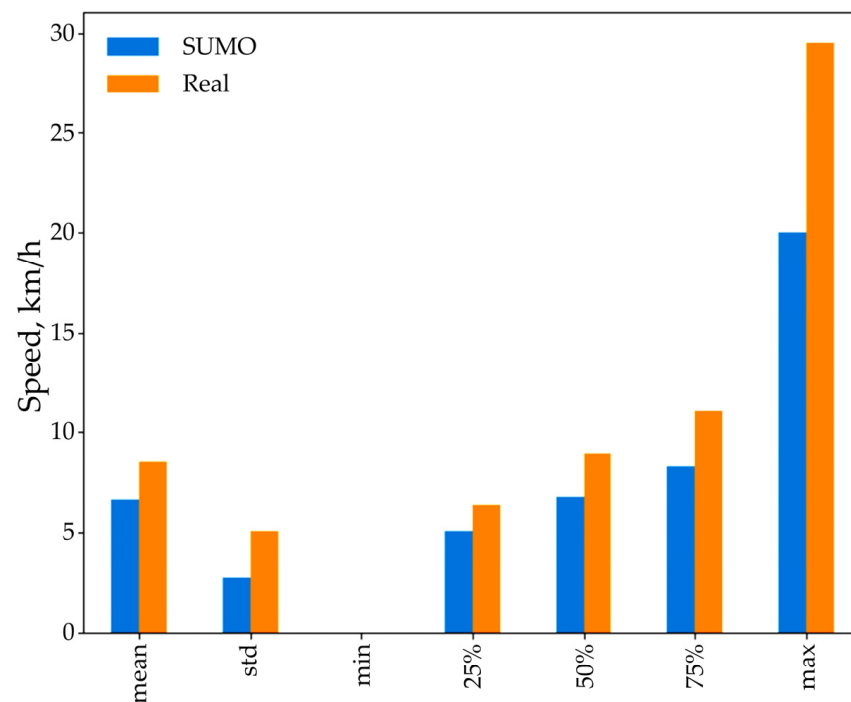
To analyze the ODM deviation, we carried out simulations in the SUMO environment, where the coordinates of vehicles were tracked when crossing sector boundaries. These coordinates were then ranked by the sectors the vehicles were leaving. The resulting ranked set was compared with real data obtained after processing the video stream, which recorded a similar section of the route. These data also went through a ranking procedure. The differences in the frequency of crossing sector boundaries between the SUMO model and the video stream processing data are presented in Figure 1, where the intersections of the main sectors for the section of the route under consideration were highlighted: v1xz2fqt, v1xz2fqv, v1xz2fqw, v1xz2fqx, v1xz2fgy, and v1xz2fqz. In Figure 1, only the last two symbols of the cluster hashes are shown on the x-axis for ease of perception.



**Figure 1.** Distribution of frequencies of cluster intersections by vehicles in the simulated area: SUMO—according to the SUMO model, Real—according to the PTS model.

Changes in the ODM led to variations in the distribution characteristics of the generated traffic flow parameters. Figure 2 illustrates the deviation in time-averaged vehicle speed values generated compared to data from the video stream.

To address deviations in the SMTF from real-world PTS, we developed a new ML-based simulation model. Initially, SUMO was used to train the model, which was then fine-tuned using data from PTS. Since the ML model is the primary building block for SMTF, assessing vehicle performance dynamics when adapting ML models to data drift was a vital development step.



**Figure 2.** Statistical indicators of deviation of speed distribution characteristics obtained from SMTF and PTS: mean, standard deviation, minimum, 25th percentile, 50th percentile, 75th percentile, and maximum.

### 2.1. Impact of Traffic Flow Dynamics on the Performance of ML Models

In the context of the VTS intelligent algorithms, changes in the probability distribution of data from the PTS can be seen as data drift [97]. This drift can be categorized into three types when PTS is viewed through a statistical model with input data  $X$  and output processing data  $Y$ . The first type (or drift of the first kind) is the drift of input data  $X$ , which reflects changes in traffic flow density due to daily, weekly, or seasonal cycles [98] and other time-related factors. The second type (the drift of the second kind) is the drift of output data  $Y$ , which could manifest as a shift in the simulation model's trace graph relative to the PTS due to external factors. The third type (the drift of the third kind) is simultaneous changes in both input and output data  $X$  and  $Y$ , indicating the combined influence of global and local factors.

Data drift from PTS dynamics directly impacts the accuracy of simulation modeling, especially when using ML models, as the distribution of input data used for analysis and training becomes outdated [11]. Therefore, calibrating the SMTF is essential to adapt to PTS data drift.

The data calibration process includes four stages [97]:

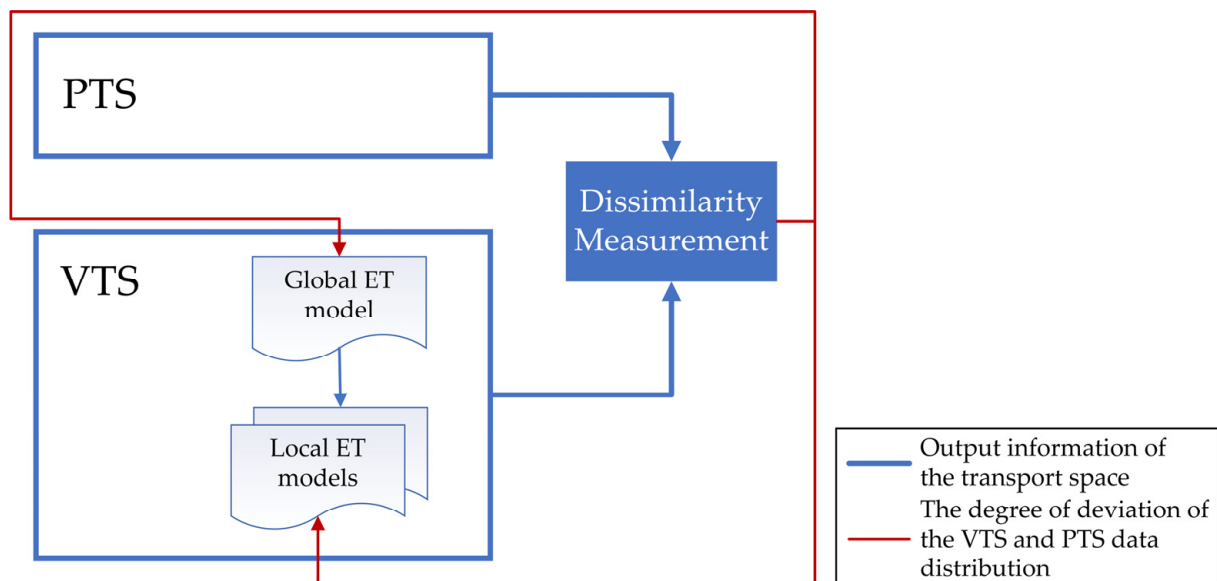
1. Data extraction: Collecting data fragments from streams to create a current sample of the studied parameter;
2. Extracting critical characteristics from the collected data;
3. Calculating test statistics to evaluate the static deviation degree of the extracted data from the original or current VTS distribution;
4. Determining the statistical significance of the deviation to initiate system adaptation to the new distribution.

These steps can be implemented as a monitoring system for the macroscopic model generated to determine when calibration is necessary.

### 2.2. Scheme of Operation and Calibration of the SMTF

As discussed in the Introduction, traffic flow variations are primarily influenced by changes in traffic density over space and time globally (across the entire study area) and locally in specific areas, such as hybrid systems that combine macroscopic and mesoscopic models. This leads to the evolution of the SMTF by transitioning from analyzing a macroscopic model of a controlled route area to studying changes in various sectors across a given territory over time.

The interaction system between the PTS and the VTS, as shown in Figure 3 (Appendix A, Figure A1), includes an initial simulation model implemented in SUMO, divided into sectors with initial global and several local patterns of SMTF timing characteristics. It also incorporates a monitoring system to track the updates of these global or local patterns. Figure 3 illustrates the iterative process of calibrating global and local SMTFs.



**Figure 3.** Iterative process for calibrating global and local SMTF models.

Once the route graph is divided into sectors, an initial temporal dynamics pattern is established through three stages. The first stage involves initializing simulation modeling in SUMO, where fixed parameters are gathered in general and local databases for each sector. The accuracy of the global ET model output and SMTF SUMO data is at least 85%. The second stage focuses on extracting the global time trend of  $Y_n(t)$  for each of the  $n$  fixed parameters using a statistical method called seasonal trend decomposition using locally estimated scatterplot smoothing (LOESS) [99].

By decomposing the time series of fixed parameter dynamics into seasonal  $S_n(t)$ , trend  $T_n(t)$ , and residual  $R_n(t)$  components, we can pinpoint the trend of the parameter  $T_n$  and the seasonal component  $S_n(t)$ . These components set the statistical proximity boundaries within which the calibration process of SMTF does not commence.

$$Y_n(t) = T_n(t) + S_n(t) + R_n(t) \tag{1}$$

The resulting trend  $T_n$  is the foundation for training the global regression model. Additionally, the average duration of seasonal deviations  $\tau_s$  was determined, representing the period during which  $S_n(t)$  values exceed half of its maximum value  $\max(S_n(t))$ .



The regression model was chosen based on three key criteria: high accuracy for variables with varying dynamics periods, computational complexity impacting training and retraining speed, and noise resistance to random short-term processes on the route. The extra trees regressor (ET) algorithm was selected for the simulation because it met these criteria [100].

As ET is an ensemble algorithm, a method was employed to adjust trained models considering regional drift to accelerate retraining [97]. This involved updating models by incorporating new classifiers trained on new data while removing some older ones until reaching the desired accuracy threshold on the test sample selected during the calibration stage.

One advantage of ET is that it utilizes the average prediction from all trees in the ensemble, potentially reducing model variance compared to other ensemble algorithms when dealing with noisy data. Additionally, introducing an extra random process in selecting input parameter split points enhances data noise robustness [101].

The third stage involves constructing the initial template of each cluster’s dynamics through additional local training on the global ET model using the described ML model adjustment method within the local database of each sector. The implementation diagram for training local models when calibrating the global SMTF model is shown in Figure 4.

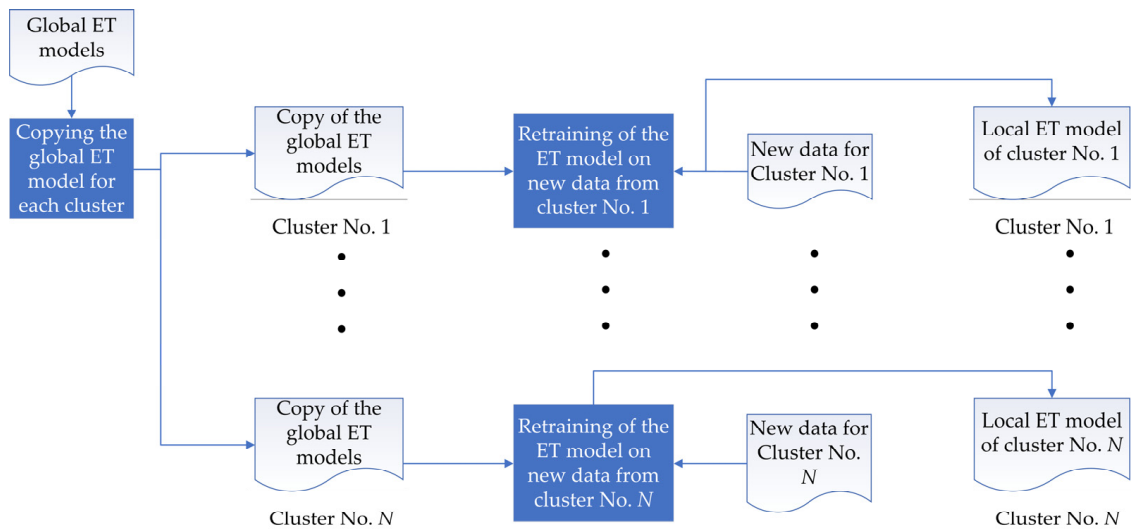


Figure 4. The process of training local models when calibrating the global SMTF model.

Subsequently, the SMTF transitions into run mode while the data drift monitoring system continuously compares fixed parameter distributions over time. The SMTF generates data from ET models, allowing the calibration process to operate simultaneously.

When new data from the PTS are received (either from the video stream analysis system or from the connected vehicle’s sensors), they are stored in the global database and then distributed to local databases based on the location of the vehicle when the data were collected. The accumulation of these new data from the PTS happens over a period of time denoted as  $\tau_S$ .

The process of calibrating the SMTF was started when we received a new dataset that included a time series of vehicle characteristics. Each sample in the dataset contained information about the time and geographic coordinates of its recording. It allowed us to distribute the data across local areas within the modeled region, providing insight into the overall trend across the modeling area.

Once the new data are collected, extracting key characteristics begins by calculating the trend of the new data  $T_n'(t)$  and relevant test statistics  $T_n(t)$ . This involves determining the maximum absolute difference between two trends for comparison, denoted as  $\max(S_n(t))$ .

The decision to initiate global or local calibration is based on specific conditions:

$$\sup_t |T'_n(t) - T_n(t)| > S_n(t) \quad (2)$$

If these conditions are met, additional training is conducted on the existing global ET model. The updated model is then incorporated into each sector by additional training on a local dataset, similar to the third stage of constructing the original SMTF. This additional model training occurs in parallel with the operation of the replaced models without causing any interruption. Additional training is only performed on a specific local model if a shift in local data is detected.

Calibrating the SMTF model involves training multiple copies of the global ET model on local data for each cluster. It happens when the acceptable deviation threshold of the statistical criterion for the distribution of the vehicle characteristics being considered is exceeded.

Evaluating the simulation's adherence to the established and calibrated SMTF involves a comprehensive process. It involves a statistically comparing the distributions of generated quantities with those obtained from the PTS data. Additionally, the performance of the ML algorithm, which has been extensively trained and integrated into the VTS, is assessed using a test dataset from the PTS. The results of these evaluations are presented in Section 4.

### 3. Methodology for Analyzing Video Stream Data

In the Introduction, it was mentioned that data from stationary objects (such as video cameras) of road safety infrastructure, or sensors from vehicles connected to the WSN, are used to analyze the ITS transport situation. To create a flexible hybrid SMTF calibration system, separate preprocessing algorithms for each data source must be developed to represent the input data uniformly [102,103].

When preprocessing data from a connected vehicle sensor system, classical steps like data cleaning, integrating data from multiple vehicles in a time-sorted manner, and normalizing parameters under consideration may suffice [104]. However, processing highway video monitoring data poses challenges in obtaining vehicle characteristics directly, which are crucial for many V2X ML applications. An additional algorithm is needed to extract vehicle characteristics from a video stream.

#### 3.1. Algorithm for Extracting Traffic Data from a Video Stream

The data extraction algorithm involves three main steps:

- Camera calibration and conversion of image coordinates to GNSS coordinates;
- Tracking and calculation of spatiotemporal characteristics of vehicles;
- Applying the Kalman filter to smooth the extracted values.

Initially, the YOLOv5 real-time video analytics system [105] was used to analyze the video stream, detecting and tracking the movements of visible vehicles. The YOLO algorithm divides the original image into a grid of  $N \times N$  cells. If the object's center falls within the coordinates of a cell, then this cell is considered responsible for determining the object's location parameters. Each object is defined by five values: center coordinates of the bounding box, its width, height, and confidence level that the box contains the object. For each object class–cell pair, the probability of the cell containing an object of that class must be determined.

The choice of YOLOv5 architecture was made for its high image processing speed and good object detection accuracy with minimal parameters and low learning rate. Training parameters included an IoU threshold value of 0.5, image compression to  $960 \times 960$  px, learning rate of 0.001, and SGD optimization algorithm. The training set comprised 3000 images, with 500 in the validation set. An example of a frame from a traffic surveillance camera that was used to collect traffic data prior to image alignment is shown in Figure 5.

To accurately detect vehicles using YOLOv5 and analyze their parameters, it was essential to address the issue of image distortions caused by various factors and establish a method to convert image coordinates to GNSS coordinates. Figure 4 shows an example of a distorted image.

To calibrate the camera images and map their coordinates to the GNSS coordinate system, the direct linear transformation method [106] was employed, along with the pose from orthographic and scaling with iterations (POSIT) method [107] to calculate the position and orientation of vehicles. The necessary input data for these algorithms included details about the dimensions of the modeled area, reference points with known coordinates in both systems, and information about the camera's optical configuration.

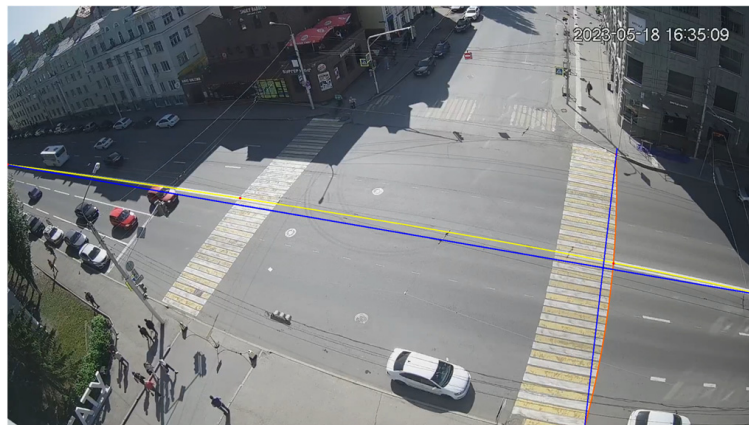


Figure 5. Original image with distortion.

The specific section of the route that was simulated represented an intersection measuring 29.4 m in width and 24.4 m in length. Reference points were established in the camera coordinates to align with points in the intersection coordinate system: (0, 244), (0, 122), (0, 0), (147, 0), (294, 0), (294, 122), (294, 244), (147, 244).

Frame marking was conducted after correcting distortions using the Nelder–Mead method implemented through the Python SciPy library in conjunction with the Python OpenCV library. Figure 6 illustrates this process.

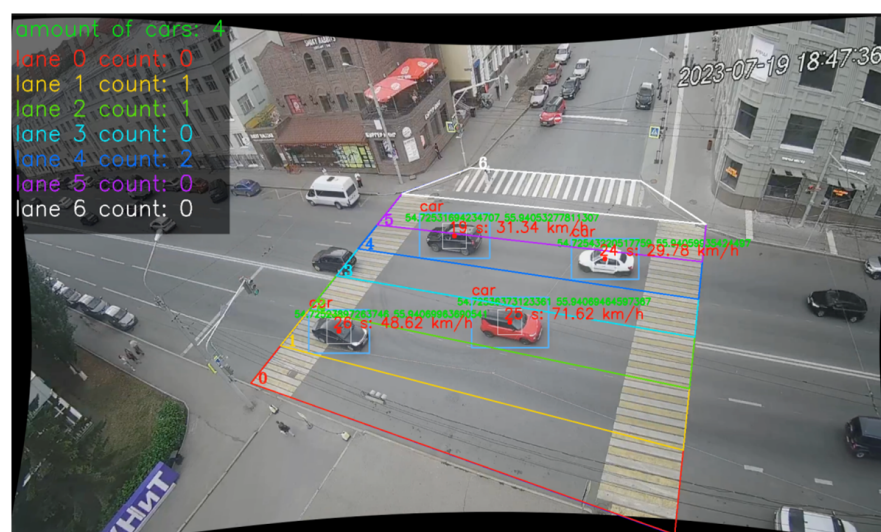


Figure 6. YOLOv5 video stream frame marking.

### 3.2. Tracking Algorithm and Indirect Calculation of Vehicle Characteristics

Four straight-line equations were formulated to define the boundaries of the visibility area along the route to improve the accuracy of tracking vehicle trajectories. The centers of objects within this area were identified in each frame, and the distances between these centers in consecutive frames were calculated using the Euclidean metric. When the distance traveled by a vehicle between frames exceeded a set threshold, its movement was recorded, and its trajectory was determined.

The distance traveled in pixels between two frames for each object was used to calculate the speed of vehicles as an example of indirect characteristics. This distance allowed for determining the speed at the current time step:

$$v_{\text{km/h}} = \frac{d(C_1, C_2)}{3600 \cdot FPS}, \quad (3)$$

where  $d(C_1, C_2)$  is the calculation of the Euclidean metric, and the distance in kilometers is then calculated from the resulting value through interpolation.  $C_i$  is the selected object center coordinates on the frame's pixel grid.  $FPS$  is the number of frames per second.

One-dimensional Kalman filtering was applied to all indirect characteristic calculations to provide a recursive estimation of the current speed value:

$$\dot{x}_k = \dot{x}_{k-1} + KG(z_k - \dot{x}_{k-1}), \quad (4)$$

where  $z_k$  is the calculated speed value,  $\dot{x}_{k-1}$  is the speed estimate at the previous step, and  $KG$  is Kalman Gain, which is calculated:

$$KG = \frac{E_{EST_i}}{E_{EST_i} + E_{EMA}}, \quad (5)$$

where  $E_{EST_i}$  is the uncertainty of the current value estimate, and  $E_{EMA}$  is the tracking algorithm's error in calculating the speed.

## 4. Simulation Results of the SMTF Calibration Algorithm and Comparison of Results with PTS

The intersection of Karl Marx and Sverdlov streets in Ufa was chosen as a simulated section for creating and calibrating the SMTF. The coordinates of the video surveillance camera at this location are latitude 54.725376 and longitude 55.940998.

Using OpenStreetMap data, a SUMO model was developed with traffic parameters set to through traffic factor = 5 and count = 12. The initial simulation involved 5000 steps, divided into six parts to establish the SMTF and subsequent calibration during the initial simulation. For example, to extract an indirect parameter of vehicles, we chose the parameters of speed, time of signal transmission from the vehicle to the infrastructure object, and the vehicle's rotation angle relative to the geographic north.

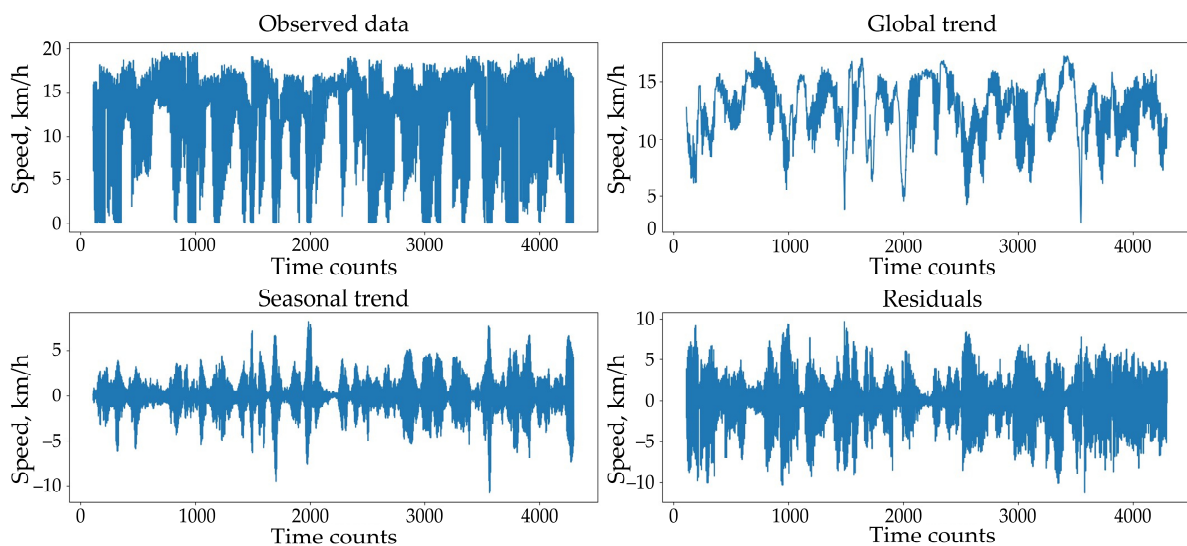
The simulated route area within a 600 m radius from the camera location was divided into 40 sectors using geohash coding (see Figure 7).

Traffic speed trends were extracted using LOESS for the first 1600 simulation steps (Figure 8). After averaging the calculated values of  $\tau_{Sn}$ , a time interval  $\tau_S$  of 60 simulation steps was chosen, approximating a minute of real time.

In the subsequent five iterations generating new data from the SUMO model, the average global and local learning times were 3500 and 110 ms, respectively. Global SMTF calibration initialization occurred in 5 out of 70 drift tracking iterations, while local calibration was observed in 26 cases.



**Figure 7.** Division of the simulated area into sectors.



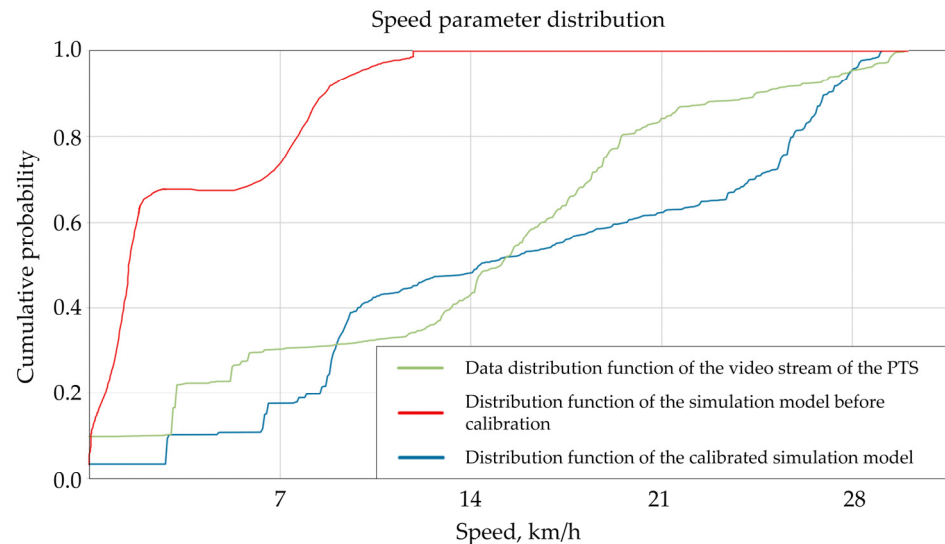
**Figure 8.** Extracting global, seasonal, and random trends from vehicle speed time series.

After simulating the SMTF calibration, the model calibration was first tested based on data from an actual video stream of the PTS intersection. Of the 40 sectors in the new dataset, information was only available for 6 sectors in the center of the intersection. When initializing a new data drift monitoring cycle, global calibration of the model was not required due to a slight divergence in the trends of all parameters under consideration. Local calibration was initiated only for four sectors for the speed parameter, two for the signal transmission time parameter, and one for the rotation angle.

Data obtained directly from a moving vehicle in the simulated area were recorded using a programmable microcontroller ELM327 from ELM Electronics, which supports OBD-II protocols [108], to monitor vehicle data in real time.

The performance of the calibration system was assessed by comparing data from the ELM327 and data generated by the SMTF before and after SMTF calibration. Two methods were used to compare the data: assessing the degree of statistical elimination of drift in data from the SMTF relative to the PTS and assessing the performance of the ML regression model trained on the SMTF data to determine the positions of the main lobe of the radiation pattern for establishing communication between the V2I infrastructure object and the vehicle.

To assess the statistical efficiency of SMTF calibration, the Kolmogorov–Smirnov (K–S) test was used, which is used to compare two samples to confirm that they belong to the same distribution [109]. The results of the K–S test show a decrease in the maximum absolute difference between the distributions of the video stream data and the calibrated SMTF compared to the original model trained on SUMO data. An example of the original velocity distribution obtained from the SUMO model, the velocity distribution from the PTS, and the distribution after model calibration are presented in Figure 9.



**Figure 9.** Cumulative distribution functions of the speed from the PTS and its generated values before and after calibration.

Thus, the results of the average level of reduction in the maximum absolute difference between the distributions of vehicle parameters are:

- For the signal transmission time from the vehicle to the infrastructure object, a reduction of 37.64% was reached;
- For vehicle speed, a reduction of 72.89% was achieved;
- For the angle of rotation of the car relative to the geographic north, a reduction of 49.27% was acquired.

To test the stability of the phased array antenna beam steering algorithm when working with data from the PTS, input data, including the specified generated parameters, were created to train the ML model.

A special algorithm was developed to generate the target variable—the number of the sector in which the main lobe of the phased array antenna should be directed—in order to minimize connection delays and maintain a stable communication channel for the car. This algorithm associates clusters with sectors based on the vehicle’s location at the time of connection to the V2I network or its movement through the simulated area. The algorithm generated random coordinates within the sector area and assigned properties to simulate a request to connect a vehicle.

Two identical ET models were trained after generating two SMTF datasets—before and after calibration. Their accuracy was assessed on ELM327 data using metrics to measure regression performance: mean absolute error (MAE), root mean square error (MSE), root mean square log error (RMSLE), coefficient of determination ( $R^2$ ), and mean absolute percentage error (MAPE). The average cross-validation results on ELM327 data are presented in Table 1.

**Table 1.** Average values of cross-validation of ML models on data from ELM327.

Model	MAE	MSE	RMSE	$R^2$	RMSLE	MAPE
Trained on SMTF data before calibration	4.1787	24.2507	4.9245	−6.9719	0.7595	1.4117
Trained on SMTF data after calibration	0.0318	0.0189	0.1376	0.9938	0.0266	0.0076

A decrease in the MAE, MSE, RMSE, RMSLE, and MAPE metrics indicates a reduction in the discrepancy between the predicted directions of the main lobe of the radiation pattern and the vehicle's location. A significant increase in  $R^2$  indicates a decrease in the variance of the dependent variable relative to the independent ones, which indicates a decrease in data drift for the ML model integrated into the VTS.

The results confirm the effectiveness of the calibration algorithm based on global and local regression models and the vehicle data synthesis algorithm using YOLOv5-processed video stream data. Tests also show that the proposed method for extracting vehicle characteristics from a video stream allows data collection without creating a WSN, which requires upgrading city infrastructure and installing sensors on vehicles. This could reduce the costs of implementing urban ITSs.

## 5. Discussion

The developed system for creating an SMTF with the possibility of global and local calibration using data from the PTS has demonstrated high adaptability to temporal and spatial deviations from the actual dynamics of traffic flow. The algorithm for extracting vehicle characteristics from video stream analysis ensured high registration accuracy, which was confirmed by tests on the correspondence of the distributions of parameters of the data from the video stream analysis and from the sensor system of the test car (the K-S test will be carried out using data from the AI and the tracker).

The results obtained were compared with the research [110], which had a very similar task. In the calibration process for the SUMO model generated previously in [110], linear regression was used to create a time trend line for the random variable, and spatial interpolation was used to adjust it based on new data. In [110], the number of vehicles in the simulated area was chosen as the generated characteristic. The metric used to assess the statistical similarity between the generated traffic flow data and the real data was  $GEH$ :

$$GEH = \sqrt{\frac{2(M - C)^2}{M + C}}, \quad (6)$$

where  $M$  is hourly traffic volume from model, and  $C$  is a real-world hourly traffic count.

For all simulated areas, the calibration method specified in [110] showed  $GEH$  values not exceeding a dimensionless value of 5, which was considered acceptable. The results of our algorithm, calculated using the  $GEH$  metric, were as follows:

- For the signal transmission time from the vehicle to the infrastructure object, the average value is  $GEH = 3.4$ ;
- For vehicle speed, the average value is  $GEH = 1.55$ ;
- For the angle of rotation of the car relative to the geographic north, the average value is  $GEH = 5.98$ .

The obtained  $GEH$  values indicate a higher statistical proximity of the data generated by the developed SMTF to real traffic flow data than the results of [110]. It refers to data on the signal transmission time from the vehicle to the infrastructure object and speed. However, the statistical proximity is lower for data on the car's turning angle relative to the geographic north. It indicates that additional research is needed on some categories of vehicle data to improve the statistical proximity of data generation to PTS data.

A comparative analysis of the developed SMTF's results with the results presented in [110] show an uneven increase in statistical proximity for different data types. It indicates the need to develop additional preprocessing algorithms for various data types.

In our future work, we plan to introduce a cluster formation algorithm to enhance the spatial accuracy of the simulation model. This algorithm will be based on the selection of elements of the simulated section of the path. Importantly, it will be designed to adapt their boundaries to the operation of specific VTS algorithms, such as the boundaries of the sectors of the beam steering algorithm of the phased array antenna, which plays a crucial role in our model.

An essential part of expanding the areas of application of SMTF will be integrating a microscopic traffic model into it, which will allow for simulation of the movement of specific vehicles within and between the cluster.

## 6. Conclusions

The proposed method for creating and calibrating a traffic simulation model showcased a significant increase in the statistical alignment between generated vehicle characteristics and actual car data on a simulated highway segment, improving from 37% to 73%. This enhancement was enabled by the approach's ability to calibrate the simulation model globally for overall trend changes and locally for specific route segments, offering flexibility and potentially saving time as local calibration could often replace global adjustments. A comparative analysis with the results of [110] demonstrated higher rates of statistical similarity of the generated vehicle characteristics for most types of data. Additionally, by reducing data drift for machine learning models integrated into the VTS system, the results indicate the algorithm's accuracy in synthesizing vehicle characteristics based on video traffic flow analysis. This capability eliminates the need to deploy a WSN, potentially reducing the costs of implementing intelligent transportation systems.

**Author Contributions:** Conceptualization, E.L., E.G. and G.V.; methodology, E.L.; software, E.L. and A.A.; validation, E.L.; formal analysis, E.L., E.G. and G.V.; investigation, E.L. and A.A.; resources, E.L., E.G. and G.V.; data curation, E.L.; writing—original draft preparation, E.L.; writing—review and editing, E.G. and G.V.; visualization, E.L. and G.V.; supervision, E.G.; project administration, E.G.; funding acquisition, E.G. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was funded under the grant of the Russian Science Foundation (Project N° 21-79-10407).

**Data Availability Statement:** A dataset for training the ML algorithm is available on request to the corresponding author's e-mail.

**Conflicts of Interest:** The authors declare no conflicts of interest.



### Appendix A

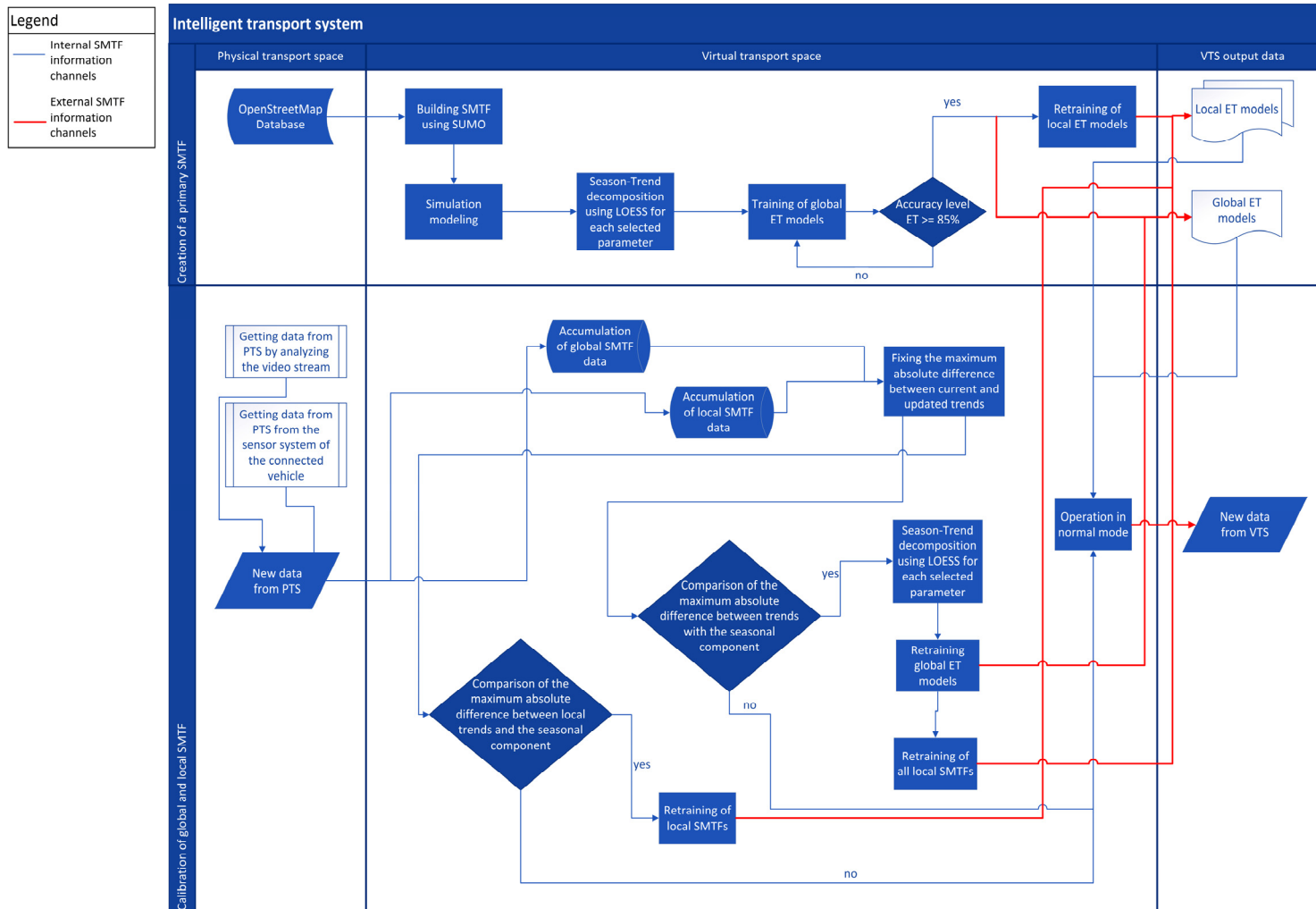


Figure A1. SMTF creation and calibration process for a single vehicle's characteristics.

## References

- Bronzino, J.D.; Peterson, D.R. (Eds.) *Medical Devices and Human Engineering*; CRC Press: Boca Raton, FL, USA, 2018.
- Statsenko, A.A.; Rogov, A.A.; Obukhov, I.V.; Smirnova, E.E. Developing Software and Hardware for Automation of Ground Urban Transport Traffic Management. In Proceedings of the 2021 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus), St. Petersburg/Moscow, Russia, 26–29 January 2021; pp. 1102–1105. [\[CrossRef\]](#)
- Momdad Salah, W.; Tie Gang, Z. Comparison of Hospital Logistics Systems. *Int. J. Sci. Res. Publ.* **2021**, *11*, 696–708. [\[CrossRef\]](#)
- Jin, J.; Rong, D.; Pang, Y.; Ye, P.; Ji, Q.; Wang, X.; Wang, G.; Wang, F.-Y. An Agent-Based Traffic Recommendation System: Revisiting and Revising Urban Traffic Management Strategies. *IEEE Trans. Syst. Man. Cybern. Syst.* **2022**, *52*, 7289–7301. [\[CrossRef\]](#)
- Miles, J.C. *Intelligent Transport Systems: Overview and Structure (History, Applications, and Architectures)*; Crolla, D., Foster, D.E., Kobayashi, T., Vaughan, N., Eds.; Wiley: Hoboken, NJ, USA, 2014; pp. 1–16.
- Vukadinovic, V.; Bakowski, K.; Marsch, P.; Garcia, I.D.; Xu, H.; Sybis, M.; Sroka, P.; Wesolowski, K.; Lister, D.; Thibault, I. 3GPP C-V2X and IEEE 802.11p for Vehicle-to-Vehicle Communications in Highway Platooning Scenarios. *Ad Hoc Netw.* **2018**, *74*, 17–29. [\[CrossRef\]](#)
- Liao, X.; Zhao, X.; Wang, Z.; Han, K.; Tiwari, P.; Barth, M.J.; Wu, G. Game Theory-Based Ramp Merging for Mixed Traffic With Unity-SUMO Co-Simulation. *IEEE Trans. Syst. Man Cybern. Syst.* **2022**, *52*, 5746–5757. [\[CrossRef\]](#)
- Wang, J.; Bi, L.; Fei, W. Multitask-Oriented Brain-Controlled Intelligent Vehicle Based on Human–Machine Intelligence Integration. *IEEE Trans. Syst. Man Cybern. Syst.* **2023**, *53*, 2510–2521. [\[CrossRef\]](#)
- Hu, Z.; Lou, S.; Xing, Y.; Wang, X.; Cao, D.; Lv, C. Review and Perspectives on Driver Digital Twin and Its Enabling Technologies for Intelligent Vehicles. *IEEE Trans. Intell. Veh.* **2022**, *7*, 417–440. [\[CrossRef\]](#)
- Huang, G.-L.; Zaslavsky, A.; Loke, S.W.; Abkenar, A.; Medvedev, A.; Hassani, A. Context-Aware Machine Learning for Intelligent Transportation Systems: A Survey. *IEEE Trans. Intell. Transp. Syst.* **2023**, *24*, 17–36. [\[CrossRef\]](#)
- Zhang, H.; Luo, G.; Li, Y.; Wang, F.-Y. Parallel Vision for Intelligent Transportation Systems in Metaverse: Challenges, Solutions, and Potential Applications. *IEEE Trans. Syst. Man Cybern. Syst.* **2023**, *53*, 3400–3413. [\[CrossRef\]](#)
- Treiber, M.; Kesting, A. *Traffic Flow Dynamics: Data, Models and Simulation*; Springer: Berlin/Heidelberg, Germany, 2013; Volume 67, ISBN 978-3-642-32460-4.
- Storani, F.; Di Pace, R.; Bruno, F.; Fiori, C. Analysis and Comparison of Traffic Flow Models: A New Hybrid Traffic Flow Model vs Benchmark Models. *Eur. Transp. Res. Rev.* **2021**, *13*, 58. [\[CrossRef\]](#)
- Khan, Z.H.; Gulliver, T.A. A Macroscopic Traffic Model for Traffic Flow Harmonization. *Eur. Transp. Res. Rev.* **2018**, *10*, 30. [\[CrossRef\]](#)
- Gipps, P.G. A Behavioural Car-Following Model for Computer Simulation. *Transp. Res. Part B Methodol.* **1981**, *15*, 105–111. [\[CrossRef\]](#)
- Gora, P.; Katrakazas, C.; Drabicki, A.; Islam, F.; Ostaszewski, P. Microscopic Traffic Simulation Models for Connected and Automated Vehicles (CAVs)—State-of-the-Art. *Procedia Comput. Sci.* **2020**, *170*, 474–481. [\[CrossRef\]](#)
- Postigo, I. *Developing Microscopic Traffic Simulation Models for the Transition towards Automated Driving*; Springer: Cham, Switzerland, 2022; ISBN 978-91-7929-438-0.
- Imai, Y.; Fujii, H.; Okano, K.; Matsudaira, M.; Suzuki, T. Development of Dynamic Micro- and Macroscopic Hybrid Model for Efficient Highway Traffic Simulation: Model Extension to Merging Sections and Validation with Probe Data. *Int. J. Intell. Transp. Syst. Res.* **2024**, *22*, 159–170. [\[CrossRef\]](#)
- Imran, W.; Khan, Z.H.; Aaron Gulliver, T.; Khattak, K.S.; Nasir, H. A Macroscopic Traffic Model for Heterogeneous Flow. *Chin. J. Phys.* **2020**, *63*, 419–435. [\[CrossRef\]](#)
- Nazarov, F.M.; Eshtemirov, B.S.; Saydullayev, Q.S. Microscopic and Macroscopic Flow Models of Traffic Management. *Tips* **2023**, *1*, 1–11. [\[CrossRef\]](#)
- Andersen, N.S.; Chiarandini, M.; Debrabant, K. A Comparison of Different Approaches to Dynamic Origin-Destination Matrix Estimation in Urban Traffic. *arXiv* **2022**, arXiv:2202.00099.
- Bochenina, K.; Taleiko, A.; Ruotsalainen, L. Simulation-Based Origin-Destination Matrix Reduction: A Case Study of Helsinki City Area. *SUMO Conf. Proc.* **2023**, *4*, 1–13. [\[CrossRef\]](#)
- Englezou, Y.; Timotheou, S.; Panayiotou, C.G. Bayesian Estimation of the Origin-Destination Matrix Using Traffic Flow Dynamics. In Proceedings of the 2019 IEEE Intelligent Transportation Systems Conference (ITSC), Auckland, New Zealand, 27–30 October 2019; pp. 2545–2550. [\[CrossRef\]](#)
- Amini, S.; Tilg, G.; Busch, F. Calibration of Mesoscopic Simulation Models for Urban Corridors Based on the Macroscopic Fundamental Diagram. 2019. Available online: [https://transp-or.epfl.ch/heart/2019/abstracts/hEART\\_2019\\_paper\\_181.pdf](https://transp-or.epfl.ch/heart/2019/abstracts/hEART_2019_paper_181.pdf) (accessed on 8 April 2024).
- Lee, J.-B.; Ozbay, K. Calibration of a Macroscopic Traffic Simulation Model Using Enhanced Simultaneous Perturbation Stochastic Approximation Methodology. 2008. Available online: <https://trid.trb.org/View/848861> (accessed on 8 April 2024).
- Lopez, P.A.; Behrisch, M.; Bieker-Walz, L.; Erdmann, J.; Flötteröd, Y.-P.; Hilbrich, R.; Lücken, L.; Rummel, J.; Wagner, P.; Wiessner, E. Microscopic Traffic Simulation Using SUMO. In Proceedings of the 2018 21st International Conference on Intelligent Transportation Systems (ITSC), Maui, HI, USA, 4–7 November 2018; pp. 2575–2582.
- Lopukhova, E.; Abdulnagimov, A.; Voronkov, G.; Kutluyarov, R.; Grakhova, E. Universal Learning Approach of an Intelligent Algorithm for Non-GNSS Assisted Beamsteering in V2I Systems. *Information* **2023**, *14*, 86. [\[CrossRef\]](#)

28. Krajzewicz, D.; Erdmann, J.; Behrisch, M.; Bieker, L. Recent Development and Applications of SUMO—Simulation of Urban Mobility. *Int. J. Adv. Syst. Meas.* **2012**, *5*, 128–138. Available online: <https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=d1dbc56fb58e437806505f8e865d69555d868af#page=48> (accessed on 8 April 2024).
29. Szendrei, Z.; Varga, N.; Bokor, L. *A SUMO-Based Hardware-in-the-Loop V2X Simulation Framework for Testing and Rapid Prototyping of Cooperative Vehicular Applications*; Jármay, K., Bolló, B., Eds.; Springer International Publishing: Cham, Switzerland, 2018; pp. 426–440.
30. Antoniou, C.; Ben-Akiva, M.; Koutsopoulos, H.N. Online Calibration of Traffic Prediction Models. *Transp. Res. Rec.* **2005**, *1934*, 235–245. [[CrossRef](#)]
31. Salles, D.; Kaufmann, S.; Reuss, H.-C. Extending the Intelligent Driver Model in SUMO and Verifying the Drive Off Trajectories with Aerial Measurements. *SUMO Conf. Proc.* **2020**, *1*, 1–25. [[CrossRef](#)]
32. Ross, M.; Du, L.; Seibold, B. Spatial-Temporal EV Charging Demand Model Considering Generic Second-Order Traffic Flows. In Proceedings of the 2021 IEEE Transportation Electrification Conference & Expo (ITEC), Chicago, IL, USA, 21–25 June 2021; pp. 789–794.
33. Hashemi, H.; Abdelghany, K. Integrated Method for Online Calibration of Real-Time Traffic Network Management Systems. *Transp. Res. Rec.* **2015**, *2528*, 106–115. [[CrossRef](#)]
34. Keler, A.; Kunz, A.; Amini, S.; Bogenberger, K. Calibration of a Microscopic Traffic Simulation in an Urban Scenario Using Loop Detector Data: A Case Study within the Digital Twin Munich. In *SUMO Conference Proceedings*; TIB Open Publishing Technische Informationsbibliothek: Hannover, Germany, 2023; Volume 4, p. 153. [[CrossRef](#)]
35. Bamdad Mehrabani, B.; Sgambi, L.; Maerivoet, S.; Snelder, M. Development, Calibration, and Validation of a Large-Scale Traffic Simulation Model: Belgium Road Network. In *SUMO Conference Proceedings*; TIB Open Publishing Technische Informationsbibliothek: Hannover, Germany, 2023; Volume 4, pp. 15–27. [[CrossRef](#)]
36. Gonzalez-Delicado, J.J.; Gozalvez, J.; Mena-Oreja, J.; Sepulcre, M.; Coll-Perales, B. Alicante-Murcia Freeway Scenario: A High-Accuracy and Large-Scale Traffic Simulation Scenario Generated Using a Novel Traffic Demand Calibration Method in SUMO. *IEEE Access* **2021**, *9*, 154423–154434. [[CrossRef](#)]
37. Balakrishna, R.; Antoniou, C.; Ben-Akiva, M.; Koutsopoulos, H.N.; Wen, Y. Calibration of Microscopic Traffic Simulation Models: Methods and Application. *Transp. Res. Rec.* **2007**, *1999*, 198–207. [[CrossRef](#)]
38. Lighthill, M.J.; Whitham, G.B. On Kinematic Waves II. A Theory of Traffic Flow on Long Crowded Roads. *Proc. R. Soc. Lond. Ser. A Math. Phys. Sci.* **1997**, *229*, 317–345. [[CrossRef](#)]
39. Huang, L.; Zhang, S.-N.; Li, S.-B.; Cui, F.-Y.; Zhang, J.; Wang, T. Phase Transition of Traffic Congestion in Lattice Hydrodynamic Model: Modeling, Calibration and Validation. *Mod. Phys. Lett. B* **2024**, *38*, 2450012. [[CrossRef](#)]
40. Rompis, S.Y.R.; Habtemichael, F.G. Calibration of Traffic Incident Simulation Models Using Field Data. *Int. J. Sustain. Transp. Technol.* **2019**, *2*, 19–26. [[CrossRef](#)]
41. Dadashzadeh, N.; Ergun, M.; Kesten, A.S.; Zura, M. Improving the Calibration Time of Traffic Simulation Models Using Parallel Computing Technique. In Proceedings of the 2019 6th International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS), Cracow, Poland, 5–7 June 2019; pp. 1–7. [[CrossRef](#)]
42. Cobos, C.; Erazo, C.; Luna, J.; Mendoza, M.; Gaviria, C.; Arteaga, C.; Paz, A. *Multi-Objective Memetic Algorithm Based on NSGA-II and Simulated Annealing for Calibrating CORSIM Micro-Simulation Models of Vehicular Traffic Flow*; Luaces, O., Gámez, J.A., Barrenechea, E., Troncoso, A., Galar, M., Quintián, H., Corchado, E., Eds.; Springer International Publishing: Cham, Switzerland, 2016; Volume 9868, pp. 468–476.
43. Yu, M.; Fan, W. Calibration of Microscopic Traffic Simulation Models Using Metaheuristic Algorithms. *Int. J. Transp. Sci. Technol.* **2017**, *6*, 63–77. [[CrossRef](#)]
44. Burghout, W.; Koutsopoulos, H.N.; Andréasson, I. Hybrid Mesoscopic–Microscopic Traffic Simulation. *Transp. Res. Rec.* **2005**, *1934*, 218–225. [[CrossRef](#)]
45. Bourrel, E.; Lesort, J.-B. Mixing Microscopic and Macroscopic Representations of Traffic Flow: Hybrid Model Based on Lighthill–Whitham–Richards Theory. *Transp. Res. Rec.* **2003**, *1852*, 193–200. [[CrossRef](#)]
46. Bourrel, E.; Lesort, J. Mixing Micro and Macro Representations of Traffic Flow: A Hybrid Model Based on the LWR Theory. 2003. Available online: <https://www.semanticscholar.org/paper/Mixing-Micro-and-Macro-Representations-of-Traffic-a-Bourrel-Lesort/9f5d08d4f35180a9b3d2dd2385a036910fba626d> (accessed on 8 April 2024).
47. Magne, L.; Rabut, S.; Gabard, J.; Toulouse, O. Towards an Hybrid Macro-Micro Traffic Flow Simulation Model. 2000. Available online: <https://www.semanticscholar.org/paper/Towards-an-hybrid-macro-micro-traffic-flow-model-Magne-Rabut/7b073b5a3ced03297e383cea558559a920623d0> (accessed on 8 April 2024).
48. Khan, I.; Ahmed, Z. ML and DL Classifications of Route Conditions Using Accelerometers and Gyroscope Sensors. In Proceedings of the 2023 3rd International Conference on Artificial Intelligence (ICAI), Islamabad, Pakistan, 22–23 February 2023; pp. 242–249. [[CrossRef](#)]
49. Sujatha, A. Traffic Congestion Detection and Alternative Route Provision Using Machine Learning and IoT-Based Surveillance. *J. Mach. Comput.* **2023**, *3*, 475–485. [[CrossRef](#)]
50. Liu, Y.; Wu, H. Prediction of Road Traffic Congestion Based on Random Forest. In Proceedings of the 2017 10th International Symposium on Computational Intelligence and Design (ISCID), Hangzhou, China, 9–10 December 2017; Volume 2, pp. 361–364.

51. Pamula, T. Road Traffic Conditions Classification Based on Multilevel Filtering of Image Content Using Convolutional Neural Networks. *IEEE Intell. Transp. Syst. Mag.* **2018**, *10*, 11–21. [[CrossRef](#)]
52. Mandhare, P.A.; Kharat, V.; Patil, C.Y. Intelligent Road Traffic Control System for Traffic Congestion A Perspective. *Int. J. Comput. Sci. Eng.* **2018**, *6*, 908–915. [[CrossRef](#)]
53. Tan, K.; Bremner, D.; Kerneć, J.L.; Imran, M. Federated Machine Learning in Vehicular Networks: A Summary of Recent Applications. In Proceedings of the 2020 International Conference on UK-China Emerging Technologies (UCET), Glasgow, UK, 20–21 August 2020; pp. 1–4. [[CrossRef](#)]
54. Fabian, P.; Rachedi, A.; Guéguen, C. Selection of Relays Based on the Classification of Mobility-type and Localized Network Metrics in the Internet of Vehicles. *Trans. Emerg. Telecommun. Technol.* **2021**, *32*, e4246. [[CrossRef](#)]
55. Javed, M.A.; Zeadally, S.; Usman, M.; Sidhu, G.A.S. FASPM: Fuzzy Logic-based Adaptive Security Protocol for Multihop Data Dissemination in Intelligent Transport Systems. *Trans. Emerg. Telecommun. Technol.* **2017**, *28*, e3190. [[CrossRef](#)]
56. Yun, K.; Yun, H.; Lee, S.; Oh, J.; Kim, M.; Lim, M.; Lee, J.; Kim, C.; Seo, J.; Choi, J. A Study on Machine Learning-Enhanced Roadside Unit-Based Detection of Abnormal Driving in Autonomous Vehicles. *Electronics* **2024**, *13*, 288. [[CrossRef](#)]
57. Taha, A.-E.M. An IoT Architecture for Assessing Road Safety in Smart Cities. *Wirel. Commun. Mobile Comput.* **2018**, *2018*, 8214989. [[CrossRef](#)]
58. Amorim, B.d.S.P.; Firmino, A.A.; Baptista, C.d.S.; Júnior, G.B.; Paiva, A.C.d.; Júnior, F.E.d.A. A Machine Learning Approach for Classifying Road Accident Hotspots. *ISPRS Int. J. Geo-Inf.* **2023**, *12*, 227. [[CrossRef](#)]
59. Jyothi, N.; Patil, R. Enhanced Machine Learning Based Techniques for Security in Vehicular Ad-Hoc Networks. In Proceedings of the 2023 International Conference on Advancement in Computation & Computer Technologies (InCACCT), Gharuan, India, 5–6 May 2023; pp. 386–393. [[CrossRef](#)]
60. Marouane, H.; Dandoush, A.; Amour, L.; Erbad, A. A Review and a Tutorial of ML-Based MDS Technology within a VANET Context: From Data Collection to Trained Model Deployment. *Authorea Preprints* **2023**. [[CrossRef](#)]
61. Baccari, S.; Hadded, M.; Ghazzai, H.; Touati, H.; Elhadeif, M. Anomaly Detection in Connected and Autonomous Vehicles: A Survey, Analysis, and Research Challenges. *IEEE Access* **2024**, *12*, 19250–19276. [[CrossRef](#)]
62. Han, X.; Zhou, Y.; Chen, K.; Qiu, H.; Qiu, M.; Liu, Y.; Zhang, T. ADS-Lead: Lifelong Anomaly Detection in Autonomous Driving Systems. *IEEE Trans. Intell. Transp. Syst.* **2023**, *24*, 1039–1051. [[CrossRef](#)]
63. Zekry, A.; Sayed, A.; Moussa, M.; Elhabiby, M. Anomaly Detection Using IoT Sensor-Assisted ConvLSTM Models for Connected Vehicles. In Proceedings of the 2021 IEEE 93rd Vehicular Technology Conference (VTC2021-Spring), Helsinki, Finland, 25–28 April 2021; pp. 1–6.
64. Bokaba, T.; Doorsamy, W.; Paul, B. Comparative Study of Machine Learning Classifiers for Modelling Road Traffic Accidents. *Appl. Sci.* **2022**, *12*, 828. [[CrossRef](#)]
65. Würth, A.; Binois, M.; Goatin, P. Traffic Prediction by Combining Macroscopic Models and Gaussian Processes. 2023. Available online: <https://hal.science/hal-04345140> (accessed on 8 April 2024).
66. Sroczynski, A.; Czyżewski, A. Road Traffic Can Be Predicted by Machine Learning Equally Effectively as by Complex Microscopic Model. *Sci. Rep.* **2023**, *13*, 14523. [[CrossRef](#)] [[PubMed](#)]
67. Son, S.; Qiao, Y.-L.; Sewall, J.; Lin, M.C. Differentiable Hybrid Traffic Simulation. *ACM Trans. Graph.* **2022**, *41*, 258:1–258:10. [[CrossRef](#)]
68. Das, S.; Tsapakis, I. Interpretable Machine Learning Approach in Estimating Traffic Volume on Low-Volume Roadways. *Int. J. Transp. Sci. Technol.* **2020**, *9*, 76–88. [[CrossRef](#)]
69. Xu, W.; Wei, H. Learning to Calibrate Hybrid Hyperparameters: A Study on Traffic Simulation. In Proceedings of the ACM SIGSIM Conference on Principles of Advanced Discrete Simulation, Orlando, FL, USA, 21–23 June 2023; pp. 144–147. [[CrossRef](#)]
70. Niel, O. A Novel Algorithm Can Generate Data to Train Machine Learning Models in Conditions of Extreme Scarcity of Real World Data. *arXiv* **2023**, arXiv:2305.00987.
71. Shubhi; Singh, A.K. Wireless Sensor Network: A Survey. 2015. Available online: [https://www.researchgate.net/publication/320385994\\_Wireless\\_Sensor\\_Networks\\_A\\_Survey](https://www.researchgate.net/publication/320385994_Wireless_Sensor_Networks_A_Survey) (accessed on 8 April 2024).
72. Das, A.; Desai, M.; Mugatkar, N.; Ponraj, A.S. *Emergency and Traffic Congestion Avoidance Using Vehicle-to-Vehicle Communication*; Thalmann, D., Subhashini, N., Mohanaprasad, K., Murugan, M.S.B., Eds.; Springer: Singapore, 2018; Volume 492, pp. 147–153.
73. Yu, H.; Guo, M. An Efficient Freeway Traffic Information Monitoring Systems Based on Wireless Sensor Networks and Floating Vehicles. In Proceedings of the 2010 First International Conference on Pervasive Computing, Signal Processing and Applications, Harbin, China, 17–19 September 2010; pp. 1065–1068. [[CrossRef](#)]
74. Qua, W. Compressed Sensing for Data Collection in Wireless Sensor Network. *J. Transduct. Technol.* **2014**, *27*, 1562–1567.
75. Wang, H.; Ouyang, M.; Meng, Q.; Kong, Q. A Traffic Data Collection and Analysis Method Based on Wireless Sensor Network. *J. Wirel. Com. Netw.* **2020**, *2020*, 2. [[CrossRef](#)]
76. Mednis, A.; Strazdins, G.; Liepins, M.; Gordjusins, A.; Selavo, L. *RoadMic: Road Surface Monitoring Using Vehicular Sensor Networks with Microphones*; Zavoral, F., Yaghob, J., Pichappan, P., El-Qawasmeh, E., Eds.; Springer: Berlin/Heidelberg, Germany, 2010; Volume 88, pp. 417–429.
77. Goliya, H.S. Data Collection and Modeling of Heterogeneous Traffic—A Review. *Int. J. Res. Appl. Sci. Eng. Technol.* **2018**, *6*, 1765–1767. [[CrossRef](#)]

78. Hourdakis, J.; Michalopoulos, P.G.; Morris, T. Deployment of Wireless Mobile Detection and Surveillance for Data-Intensive Applications. *Transp. Res. Rec.* **2004**, *1900*, 140–148. [CrossRef]
79. Manjusha, M.; Sunitha, V. A Review of Advanced Pavement Distress Evaluation Techniques Using Unmanned Aerial Vehicles. *Int. J. Pavement Eng.* **2023**, *24*, 2268796. [CrossRef]
80. Park, S.-S.; Tran, V.-T.; Lee, D.-E. Application of Various YOLO Models for Computer Vision-Based Real-Time Pothole Detection. *Appl. Sci.* **2021**, *11*, 11229. [CrossRef]
81. Hidas, P.; Wagner, P. Review of Data Collection Methods for Microscopic Traffic Simulation. In Proceedings of the WCTR, Istanbul, Turkey, 4–8 July 2004; Institute of Transport Research: Berlin, Germany, 2004.
82. Apeltauer, J.; Babinec, A.; Herman, D.; Apeltauer, T. Automatic vehicle trajectory extraction for traffic analysis from aerial Video data. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2015**, *40*, 9–15. [CrossRef]
83. Qasim, M.A.; Ali, Q.A.; Sahab, N.M.; Jaleel, R.A.; Zahra, M.M.A. Multimedia Imaging System of Data Collection and Antenna Alignment for Unmanned Aerial Vehicles Based Internet of Things. *Fusion Pract. Appl.* **2023**, *12*, 19–27. [CrossRef]
84. Kornyei, L.; Horvath, Z.; Ruopp, A.; Kovacs, A.; Liszcai, B. Multi-Scale Modelling of Urban Air Pollution with Coupled Weather Forecast and Traffic Simulation on HPC Architecture. In Proceedings of the International Conference on High Performance Computing in Asia-Pacific Region Companion, Virtual Event, 20–22 January 2021; pp. 9–10. [CrossRef]
85. Basnayake, C.; MacIver, A.; Lachapelle, G. A Gps-Based Calibration Tool for Microscopic Traffic Simulation Models. 2003. Available online: <https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=9bd08e6f4e8579b7693e5b66e828b5bd90ae9a8f> (accessed on 8 April 2024).
86. Barsi, M.; Barsi, A. Building opendrive model from mobile mapping data. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2021**, *43*, 9–14. [CrossRef]
87. Oughton, E.J.; Russell, T.; Johnson, J.; Yardim, C.; Kusuma, J. Itmlogic: The Irregular Terrain Model by Longley and Rice. *J. Open Source Softw.* **2020**, *5*, 2266. [CrossRef]
88. Savitha, B.G.; Murthy, R.S.; Jagadeesh, H.S.; Sathish, H.S.; Sundararajan, T. Study on Geometric Factors Influencing Saturation Flow Rate at Signalized Intersections under Heterogeneous Traffic Conditions. *J. Transp. Technol.* **2017**, *07*, 83–94. [CrossRef]
89. Zhang, F.; Qian, Y.; Zeng, J. Characterizing Heterogeneous Traffic Flow at a Slope Bottleneck via Cellular Automaton Model. *IEEE Trans. Intell. Transport. Syst.* **2023**, *24*, 6507–6516. [CrossRef]
90. Ai, W.; Hu, J.; Liu, D. Bifurcation Analysis of Improved Traffic Flow Model on Curved Road. *J. Comput. Nonlinear Dyn.* **2023**, *18*, 071005. [CrossRef]
91. Guido, G.; Shaffiee Haghshenas, S.; Shaffiee Haghshenas, S.; Vitale, A.; Astarita, V.; Park, Y.; Geem, Z.W. Evaluation of Contributing Factors Affecting Number of Vehicles Involved in Crashes Using Machine Learning Techniques in Rural Roads of Cosenza, Italy. *Safety* **2022**, *8*, 28. [CrossRef]
92. Al-Bayati, A.H.; Shakoree, A.S.; Ramadan, Z.A. Factors Affecting Traffic Accidents Density on Selected Multilane Rural Highways. *Civ. Eng. J.* **2021**, *7*, 1183–1202. [CrossRef]
93. Zakharov, D.; Magaril, E.; Rada, E. Sustainability of the Urban Transport System under Changes in Weather and Road Conditions Affecting Vehicle Operation. *Sustainability* **2018**, *10*, 2052. [CrossRef]
94. Cui, Y.; Ge, S.S. Autonomous Vehicle Positioning with GPS in Urban Canyon Environments. *IEEE Trans. Robot. Autom.* **2003**, *19*, 15–25. [CrossRef]
95. Meguro, J.; Murata, T.; Takiguchi, J.; Amano, Y.; Hashizume, T. GPS Multipath Mitigation for Urban Area Using Omnidirectional Infrared Camera. *IEEE Trans. Intell. Transp. Syst.* **2009**, *10*, 22–30. [CrossRef]
96. Ji, Z.; Xie, Q. Bidirectional Mapping Trajectory Similarity Measurement Algorithm Based on Geohash Grid. In Proceedings of the 2021 6th International Symposium on Computer and Information Processing Technology (ISCIPT), Changsha, China, 11–13 June 2021; pp. 281–285. [CrossRef]
97. Lu, J.; Liu, A.; Dong, F.; Gu, F.; Gama, J.; Zhang, G. Learning under Concept Drift: A Review. *IEEE Trans. Knowl. Data Eng.* **2019**, *31*, 2346–2363. [CrossRef]
98. Ostrowski, K.; Budzynski, M. Measures of Functional Reliability of Two-Lane Highways. *Energies* **2021**, *14*, 4577. [CrossRef]
99. Cleveland, R. STL: A Seasonal-Trend Decomposition Procedure Based on Loess. 1990. Available online: <https://www.proquest.com/openview/cc5001e8a0978a6c029ae9a41af00f21/1?pq-origsite=gscholar&cbl=105444> (accessed on 8 April 2024).
100. Arjasakusuma, S.; Swahyu Kusuma, S.; Phinn, S. Evaluating Variable Selection and Machine Learning Algorithms for Estimating Forest Heights by Combining Lidar and Hyperspectral Data. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 507. [CrossRef]
101. Martiello Mastelini, S.; Nakano, F.K.; Vens, C.; de Leon Ferreira de Carvalho, A.C.P. Online Extra Trees Regressor. *IEEE Trans. Neural Netw. Learn. Syst.* **2023**, *34*, 6755–6767. [CrossRef]
102. Kim, J.; Hwangbo, H.; Kim, S. An Empirical Study on Real-Time Data Analytics for Connected Cars: Sensor-Based Applications for Smart Cars. *Int. J. Distrib. Sens. Netw.* **2018**, *14*, 155014771875529. [CrossRef]
103. Al-Ahmadi, H.; Hassan, H. Real-Time Simulation of Traffic Monitoring System in Smart City. 2020. Available online: <https://www.semanticscholar.org/paper/Real-Time-Simulation-of-Traffic-Monitoring-System-Al-Ahmadi-Hassan/fbd7a10eda11b03614e3a319a867c2e2f8d7ebe7> (accessed on 8 April 2024).
104. Mallikharjuna Rao, K.; Saikrishna, G.; Supriya, K. Data Preprocessing Techniques: Emergence and Selection towards Machine Learning Models—A Practical Review Using HPA Dataset. *Multimed. Tools Appl.* **2023**, *82*, 37177–37196. [CrossRef]

105. Algiriyage, N.; Prasanna, R.; H Doyle, E.; Stock, K.; Johnston, D.; Punchihewa, M.; Jayawardhana, S. Towards Real-Time Traffic Flow. Estimation Using. YOLO and SORT from Surveillance Video Footage. 2021. Available online: [https://www.researchgate.net/publication/353327177\\_Towards\\_Real-time\\_Traffic\\_Flow\\_Estimation\\_using\\_YOLO\\_and\\_SORT\\_from\\_Surveillance\\_Video\\_Footage](https://www.researchgate.net/publication/353327177_Towards_Real-time_Traffic_Flow_Estimation_using_YOLO_and_SORT_from_Surveillance_Video_Footage) (accessed on 8 April 2024).
106. Barone, F.; Marrazzo, M.; Oton, C.J. Camera Calibration with Weighted Direct Linear Transformation and Anisotropic Uncertainties of Image Control Points. *Sensors* **2020**, *20*, 1175. [[CrossRef](#)] [[PubMed](#)]
107. Dementhon, D.F.; Davis, L.S. Model-Based Object Pose in 25 Lines of Code. *Int. J. Comput. Vis.* **1995**, *15*, 123–141. [[CrossRef](#)]
108. *ISO 15031-5:2015; Road Vehicles—Communication between Vehicle and External Equipment for Emissions-Related Diagnostics*. ISO: Geneva, Switzerland, 2015. Available online: <https://www.iso.org/standard/66368.html> (accessed on 3 April 2024).
109. Riffenburgh, R.H. Chapter 14—Tests on Variability and Distributions. In *Statistics in Medicine*, 3rd ed.; Riffenburgh, R.H., Ed.; Academic Press: San Diego, CA, USA, 2012; pp. 299–323. ISBN 978-0-12-384864-2.
110. Patil, M.; Tulpule, P.; Midlam-Mohler, S.; Patil, M.; Tulpule, P.; Midlam-Mohler, S. *An Approach to Model a Traffic Environment by Addressing Sparsity in Vehicle Count Data*; SAE International: Warrendale, PA, USA, 2023.

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.