Global Positioning System Data to Model Network-Wide Road Segment Level Number of Lanes Using Spatial Analysis and Machine Learning

Adham Badran [a*], Ahmed El-Geneidy [b], and Luis Miranda-Moreno [a]
[a] Civil Engineering Department, McGill University, 817 Sherbrooke Street West, Montreal, H3A 0C3, Canada
[b] School of Urban Planning, McGill University, 815 rue Sherbrooke West, Montreal, H3A0C2, Canada
[*] Corresponding Author

Contact: adham.badran@mail.mcgill.ca, Department of Civil Engineering, McGill University, 817 Sherbrooke Street West, Montreal H3A 0C3, Canada.

Word Count: 5363

Submitted on July 27th, 2023

For Citation Please use: Badran, A., El-Geneidy, A., Miranda-Moreno, L. (2024). *Global positioning system data to model network-wide road segment level number of lanes using machine learning.* Paper presented at the 103rd Transportation Research Board Annual Meeting, Washington DC, USA.

Abstract

One of the main features required in transport network modelling is the number of lanes used to estimate the road capacity and predict vehicular travel times based on traffic flows. Traditionally, the number of lanes information is collected manually or more recently extracted using computer vision techniques, which are two resource intensive methods. This research proposes the use of emerging crowd-sensed Global Positioning System (GPS) data to predict the number of lanes per road segment for large scale transport models through geographic operations and machine learning. The developed method consists of i) a spatial analysis to analyze the GPS trajectory data and estimate predictors and ii) a supervised machine learning model development to build a model able to predict the number of lanes per road segment.

It was found that the method predicts the number of lanes at an accuracy of 91% using two predictors: number of GPS points per road segment and a lateral distance variable containing 60% of the GPS data points, centered around the lateral distance distribution median. The best prediction model was obtained using decision trees classifier. It was also found that most of the local roads did not have sufficient data points to obtain a stable lateral distance distribution, therefore, the model was limited to a subset of road segments with sufficient observations. Given the availability of high spatiotemporal coverage GPS data, the method can be adapted and applied to large scale road network models and predict the number of lanes accurately and cost-effectively.

Keywords: Global Positioning System, Transport Model, Road Network, Number of Lanes, Road Capacity, EMME.

Introduction

Knowledge of the number of lanes on road segments within the transport network is essential for the planning and operation of the transport system. For example, conventional and autonomous vehicle navigation, transport modelling and simulation, road safety applications all require the number of lanes information as an input. In fact, lane-level digital maps are critical for advanced driver assistance systems and continuous research is being performed to improve their development (Guo et al., 2016). Moreover, autonomous vehicle navigation requires prior knowledge of the road network in addition to real-time detection of the road lanes to select the trajectory appropriately (Bounini et al., 2015). In transport modelling, the number of lanes information is essential for all modelling scales. Macroscopic models include the number of lanes information into volume-delay functions to determine the road's vehicular capacity and evaluate road segment level travel time. Meanwhile microscopic transport models consider the number of lanes through lane changing models (Treiber and Kesting, 2013). Another example is the use of the number of lanes when analyzing pedestrian-vehicle interaction at crossings and the relationship with road-user safety (Kadali and Vedagiri, 2020).

The challenge in obtaining the number of lanes information is for large-scale road networks and maps. In fact, at large scales, the required resources to develop and maintain detailed digital networks become significant which renders the manual development infeasible. With the advances in technology, new data sources and techniques are emerging and present a potential to extract transport network-related information. Global Positioning System (GPS) trajectory data is being collected by different organizations through GPS-enabled smartphones and stored on servers using cellular internet. For example, the city of Montreal has provided its residents a smartphone application that records their trajectories for a limited period to analyze the trajectory data and improve transport planning and reduce traffic delays (Montréal).

Extracting the number of lanes has been tackled in the past using different data sources. The most frequent method to extract the number of lanes for large-scale networks is based on aerial imagery and computer vision techniques. Multiple studies have been looking at extracting road network features automatically using different data sources. First, high-resolution imagery, in combination with computer vision methods have enabled the large-scale detection and extraction of road network-related attributes. One of the research groups has done extensive work using road segmentation to detect different visible features such as the road, sidewalks, vegetation, buildings, and cars to augment OpenStreetMap by adding more features (Mattyus et al., 2015). The main challenges were found to be the presence of trees, shadows, cars, as they increase heterogeneity in the images in addition to misalignment issues with respect to the road centreline file used as a priori of road segments' location. The same research group further expanded the analysis by collecting and incorporating street-level imagery in the number of lanes recognition algorithm which increased its prediction accuracy (Máttyus et al., 2016). Recognizing that collecting street-level imagery presents high collection and processing costs, they proposed a more resource-friendly version that only employs satellite imagery but takes advantage of new methodological advances in deep learning to improve the model accuracy (Máttyus et al., 2017). Another study has also extracted the number of lanes information from satellite imagery using an SVM classifier for lane identification based on brightness levels. Although they predicted the number of lanes at an accuracy of 100%, the experiment was only presented for six road segments (Tang et al., 2014). Although satellite imagery has been used to detect the number of lanes and improved by collecting street level high-resolution imagery, it is not without limitations. Data availability is limited due to the collection costs, moreover, occlusions, illumination variability and unmarked road lines reduce the capacity of such techniques (Kasmi et al., 2018). The best number of lanes prediction accuracy obtained was 83 %.

Recent research efforts have been studying the extraction of road networks from GPS data using different spatial analysis algorithms. Three main approaches were used to extract road networks: Clustering, intersection linking, and track alignment. For example, the work by Guo et al. (2021) proposes a clustering method to extract road network centreline and intersections with the accuracy of 92%. Clustering is in fact the most popular method to extract road networks from GPS trajectory data. Another study by Zhang et al. (2019) employs the intersection linking method to detect the road network and intersections at an accuracy greater than 90%. Although not very popular, studies by Leichter and Werner (2019) and Zhongyi et al. (2018) have also used the track alignment method to generate road networks. However, accuracy was either low or not compared to the ground truth. Although these road network inference methods are able in some cases to extract the road network centreline and intersections with high accuracy, they do are not designed to extract more detailed road network features such as the number of lanes.

Very few studies have examined the use of sole GPS trajectory data to extract the number of lanes. A study by Arman and Tampere (2020) proposes a method that extracts lane locations on a highway corridor. However, the number of lanes extracted is not explicitly validated by comparing to the ground truth. Therefore, no accuracy was provided. One attempt by Zhang et al. (2010) used GPS traces and a road centreline map from OpenStreetMap to improve the map quality and estimate the number of lanes. The main limitation was the assumption of normal distribution of GPS traces with respect to the road centre, which is not the case and resulted in number of lanes prediction accuracy of less than 60%.

Moreover a study by Chen and Krumm (2010) fits Gaussian mixture models to GPS trajectory data to determine the number of lanes. Although the study attempts to preserve the continuous

nature of road segments, it is limited by the sample size and the fact that this method requires prior knowledge of the number of Gaussian distributions to fit. Thus, the study resulted in relatively low accuracy predictions.

In sum, the main limitations of past studies extracting the number of lanes are the high cost of imagery data collection and the output accuracy. In fact, the high cost reduces the frequency of map updates which can result in maps not representing the continuously evolving nature of the road network. In addition, the output accuracy of past studies can potentially be improved by using large-scale GPS trajectory data.

Considering the general availability of road centreline data or algorithms to infer them from different data sources, the objective of this study is to propose a method that uses GPS trajectory data to extract the number of lanes with a relatively high accuracy. This is done through spatial analysis of GPS trajectory points to extract variables that feed into a machine learning classification algorithm that predicts the number of lanes for road segments for use in large-scale transport models.

Methodology

GPS data treatment can be divided into two main parts based on the analysis type. The first part of the analysis was the spatial analysis using Geographic Information System (GIS) software carried out in the FME software. This software was selected since it is a very efficient data integration platform capable of managing, combining, and transforming big data with advanced spatial data analysis capabilities. The second step was the number of lanes prediction model development and visualization carried out in MATLAB. The general assumption of this study is that although GPS accuracy is between 7 and 13 meters (Merry and Bettinger, 2019), GPS trajectory points will be distributed around the middle of traffic lanes when the sample size is large. Therefore, this method proposes to determine the distance distribution of GPS points for each directional link with respect to a reference line and infer the number of lanes based on the distribution properties through a machine learning classification method.

Spatial Analysis

The first step requires raw GPS trajectory data, a road network model (links and nodes), and an azimuth-direction dictionary table as input. A summary of the spatial analysis steps can be seen in Figure 1. The yellow boxes present the input data sources required to carry out the spatial analysis steps. In this study, each GPS trajectory point had two sets of longitude and latitude points; raw, and map matched coordinates, which were both used at different stages of the analysis.

The process can be divided into four main steps: 1. Determine the direction of each GPS trajectory point, 2. Remove GPS trajectory points located at intersections, 3. Associate each GPS trajectory point to a directional road segment, and 4. Calculate the lateral distance between each GPS point and the reference line. The number corresponding to each step is also presented in Figure 1.

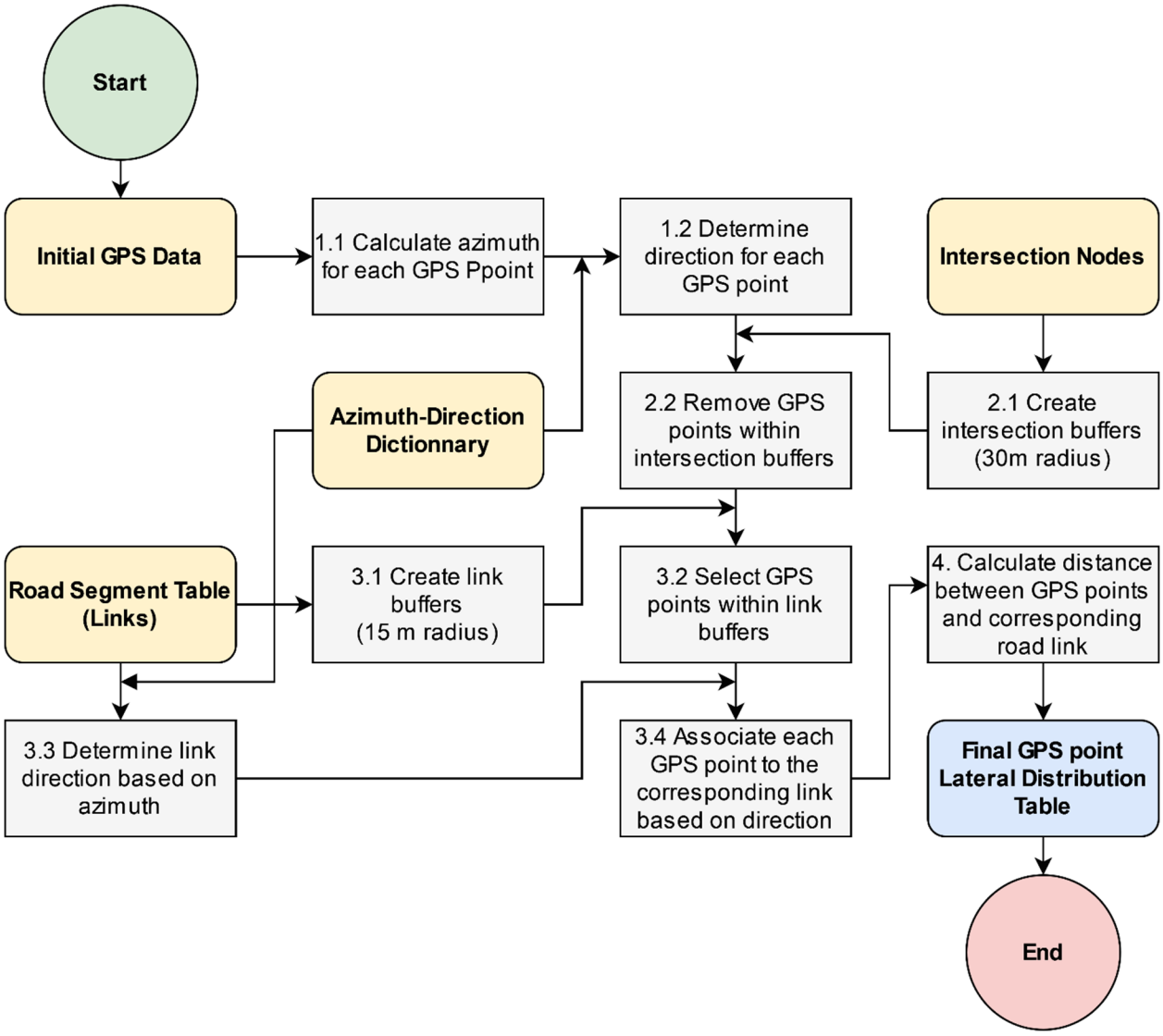
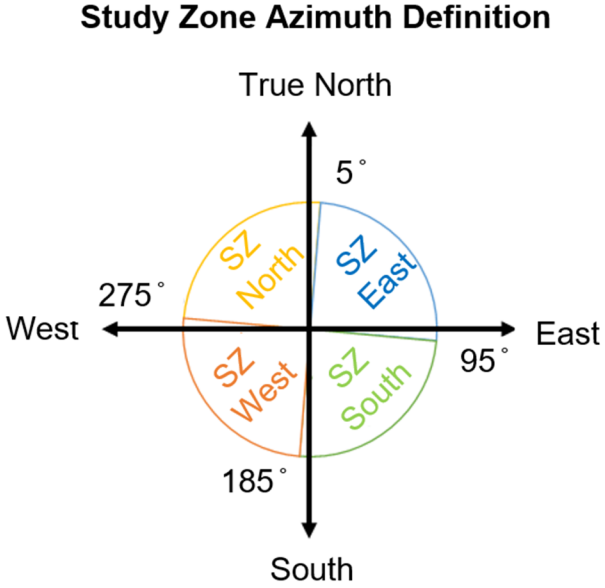


Figure 1 - GPS Trajectory Points GIS Treatment Diagram

Firstly, the azimuth of each GPS trajectory point was calculated based on its location and the location of the consecutive point within the same trip. The azimuth is defined as the orientation, in degrees, between two points as the number of degrees clockwise from the north reference. The azimuth was selected as the measure to define trip segment directions and an azimuth-direction dictionary was created for that purpose as seen in Figure 2. Map matched coordinates were used to calculate the azimuth to ensure consistent direction results and remove the fluctuations found in raw GPS point data. Following the azimuth calculation for each point, the direction was calculated using the azimuth-direction dictionary.

Secondly, intersection buffers were used to remove GPS points that fall within the vicinity of intersections. Given that this study aims to determine the mid-block road segment number of lanes, the GPS trajectory points in the vicinity of intersections were removed since the number of lanes near an intersection is sometimes different to allow for upstream dedicated turning lanes or downstream insertion lanes. Following visual inspection of the road network, a buffer size of 30-meter radius with respect to the intersection centres was used to filter GPS trajectory points within

intersection areas. This ensured that the remaining GPS trajectory points correspond to travel within the road segment.



Azimuth angle range (degrees)	Direction
275 - 4	North
5 - 94	East
95 - 184	South
185 - 274	West

SZ: Study Zone

Figure 2 - Azimuth-Direction Correspondence

Thirdly, to remove noisy GPS trajectory points, a road segment (link) buffer of 15-meter radius was created to select GPS points associated to each link through nearest neighbour analysis. This buffer size was selected to ensure that the GPS points' lateral distribution profile with the respect to the directional link is captured entirely while minimizing the number of outliers. This was validated in the following steps of the analysis by examining all lateral distance distribution histograms and kernel density estimators. The link direction was also obtained based on the azimuth to add an extra criterion when selecting the nearest neighbour and ensure that every GPS point is associated to the correct directional link.

Fourthly, the shortest distance between each GPS point and the associated directional link is calculated and serves in the following step develop a number of lanes prediction model. This distance corresponds to the length of the perpendicular line, d_i, between the GPS location point and the directional link as seen in Figure 3. It was the main variable carried to the next modelling step.

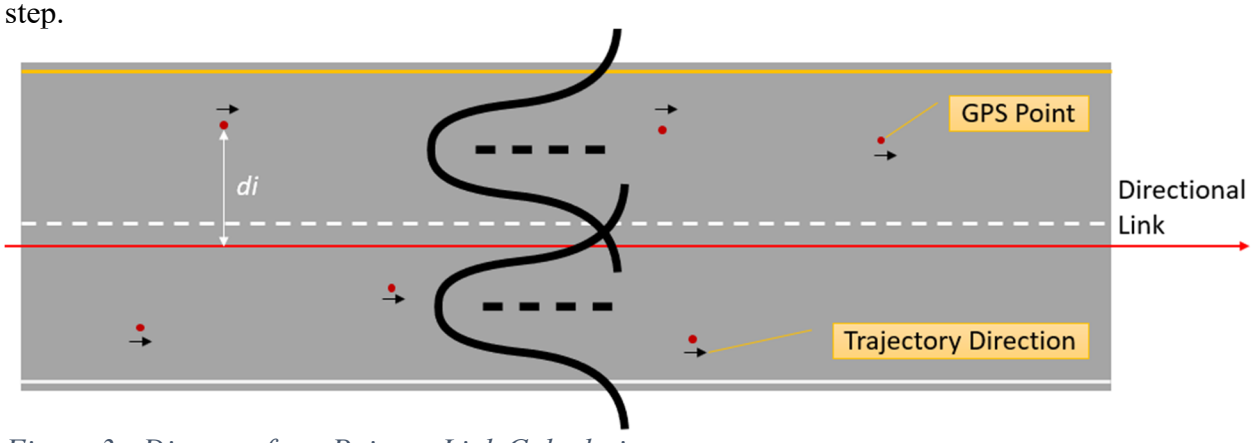


Figure 3 - Distance from Point to Link Calculation

The location of the directional link with respect to the actual road segment is approximate since it is based on a simple street centreline shapefile. Moreover, for bidirectional road segments, the links for both directions are superimposed. This was taken into account while determining road segment buffer size.

Road Segment Number of Lanes Prediction Model

Following spatial analysis, the second part of the method consisted of creating the number of lanes prediction model. Assuming that road segments with different numbers of lanes have different GPS trajectory data characteristics (such as spatial distribution pattern and number of points), roads with different numbers of lanes are seen as distinct categories and the question is formulated as a number of lanes classification problem. A classification model was calibrated using input variables derived from GPS trajectory points data to output the number of lanes for each road segment. For each road segment, input variables were compiled following the spatial analysis part and were used to train the model to predict the number of lanes as a categorical variable. The two main GPS trajectory points descriptors, used to derive input variables to the classification tree model, were the lateral distance d_i and the number of points per directional road segment. A frequency histogram and a kernel density estimator were fitted to the distance variable to visualize the distribution with respect to the reference line (directional link) and determine model parameters. First, it was observed that for some of the links, sample size was too low and resulted in unstable and unmeaningful distributions. Following inspection of the kernel density estimators and frequency histogram, the sample size was limited to a minimum of 500 GPS points per directional link to produce stable results in terms of distribution shape. Road segments with fewer GPS point observations were removed.

Based on the observed distributions and preliminary tests and aiming to create variables that reflect the lateral distribution of GPS trajectory points with respect to the link, distance percentiles, d_{ipc}, were calculated for different percentiles, i, of 5%, 10%, 15%, 20%, 80%, 85%, 90%, and 95%. To standardize these values and render them comparable across different links, new variables were created by calculating the variable D_p defined as the lateral distance containing a proportion, p, of the GPS points data. D_p is calculated using lateral distance percentiles to ensure that this new variable is centred around the median distance value. The following are the lateral distance variables that were calculated:

$$D_{90} = d_{95pc} - d_{5pc}$$
$$D_{80} = d_{90pc} - d_{10pc}$$
$$D_{70} = d_{85pc} - d_{15pc}$$
$$D_{60} = d_{80pc} - d_{20pc}$$

For example, D_{60} corresponds to the difference between the 80th percentile distance and 20th percentile distance, therefore it contains 60% of the GPS points data. A visual illustration is provided in Figure 4. The number of GPS points per link and the standard deviation of lateral distance per link were also calculated to be tested in the model specification.

Following the creation of the variables for each road segment, supervised machine learning classification methods were tested. In fact, classification tree analysis was carried out to determine if it can create an accurate model that can be used for prediction. This method is a good option when ground truth data is available for the learning step. Moreover, it is non-parametric and does not require prior knowledge of the distribution of each variable. Another advantage of this method compared to other machine learning techniques such as neural networks classification is its transparency which makes the model easy to interpret (Ian et al., 2017).

To ensure protection against overfitting, model validation was carried out using a 5-fold cross-validation. This validation method divides the dataset randomly into five groups. At each step, one of the five groups is held out to be used for validation while the other four groups are used to train the model. Once the model is specified, it is used to make predictions on the group that was held out. For a 5-fold cross validation, this process is repeated five times.

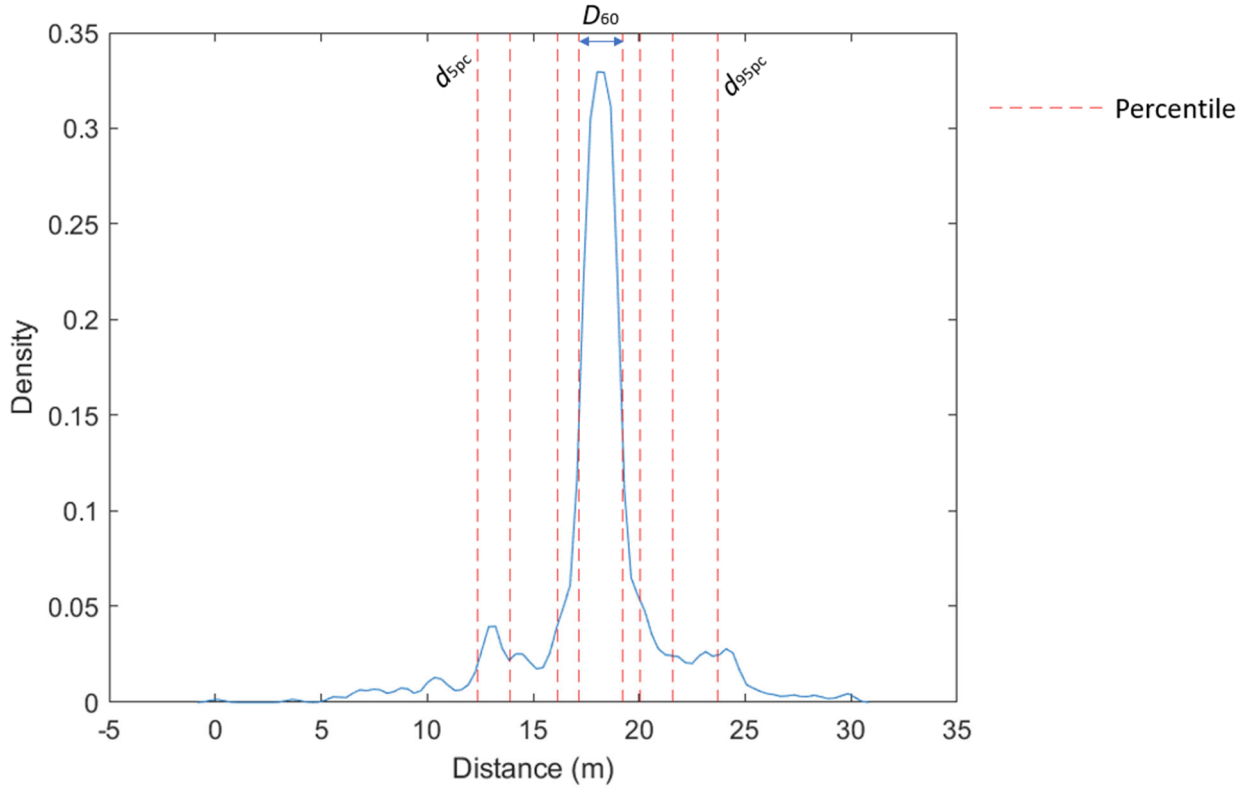


Figure 4 - Example of Percentile Visualization

Data

Three main input datasets are used: 1) GPS trajectory points, 2) Modelled directional road network (links and nodes), and 3) Google maps and Street View. GPS data was collected during the spring of 2014 in Quebec City, Canada. It was collected for 21 days by 2000 voluntary users through the Mon Trajet smartphone app, made available by the Municipality. Each point is described by the following attributes: map matched X and Y coordinates, trip ID, speed, and timestamp (Year-Month-Day-Hour-Minute-Second). Following the preprocessing steps, 245 links were selected as the experimental data to model number of lanes, which included 120 000 GPS points (excluding GPS points within the intersection buffers. This study area was selected based on its urban setting since it is in the city centre where more GPS trajectories were available. Figure 5 presents a sample of the study area where part a shows the raw GPS trajectory points, and part b shows the processed GPS trajectory points for the same road corridor following spatial analysis steps 1 to 3. In part b of the figure, GPS trajectory points are colored differently depending on the link to which they were associated.

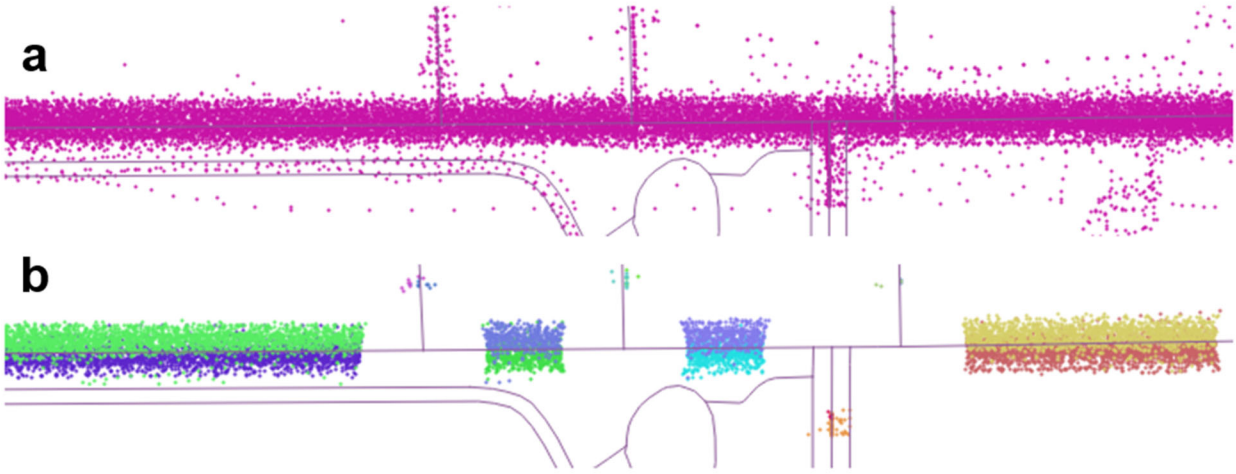


Figure 5 - Sample of GPS Points Data in Study Area - Before and After Spatial Processing

The directional road network was created using an initial road centreline shapefile which was converted in a network model compatible the EMME transport modelling software to obtain directional links and augmented using the same GPS trajectory data to ensure that road topology and connectivity are valid. Each link is defined by an origin and a destination node. The possible number of lanes per directional link was one, two, or three lanes, for which the ground truth was manually extracted using Google Maps and Street View.

For a given road segment, it is important to note that the number of lanes available for traffic can vary spatially and temporally. The presence of lanes dedicated to transit vehicles or high-occupancy vehicles at a fixed schedule on concerned road segments reduces temporally the number of lanes available to general traffic. This is also the case for lanes that are used for parking at fixed schedules. Throughout a road segment, the number of lanes can also change spatially. For example, it is common to see a higher number of lanes at the two extremities of a road segment to allow for traffic insertion and for dedicated turning lanes. The complex nature of traffic lanes can be seen in Figure 6 where a reserved bus lane (highlighted in green) is present at a fixed schedule and the number of lanes at the intersection level is different (usually greater) than the mid-block number of lanes to accommodate turning movement flows. This paper examines the mid-block number of lanes and does not consider reserved lanes.

Figure 6 - Example of a Complex Road Geometry

Results

With the proposed steps and parameters, it was possible to extract GPS points for road segments and associate each point to the correct directional link based on the trajectory direction. The sample size filter limited the number of analyzed directional links included in the analysis to 43 links. The buffer sizes were also validated based on the frequency distributions of GPS points' lateral distance with respect to the link since the entire distribution profile is captured. This can also be noted in

Figure 7 which presents the kernel density estimator fitted to the lateral distance variable distribution for six different links of varying number of lanes. The sample size, N, and the D_{60} values are also presented for each link.

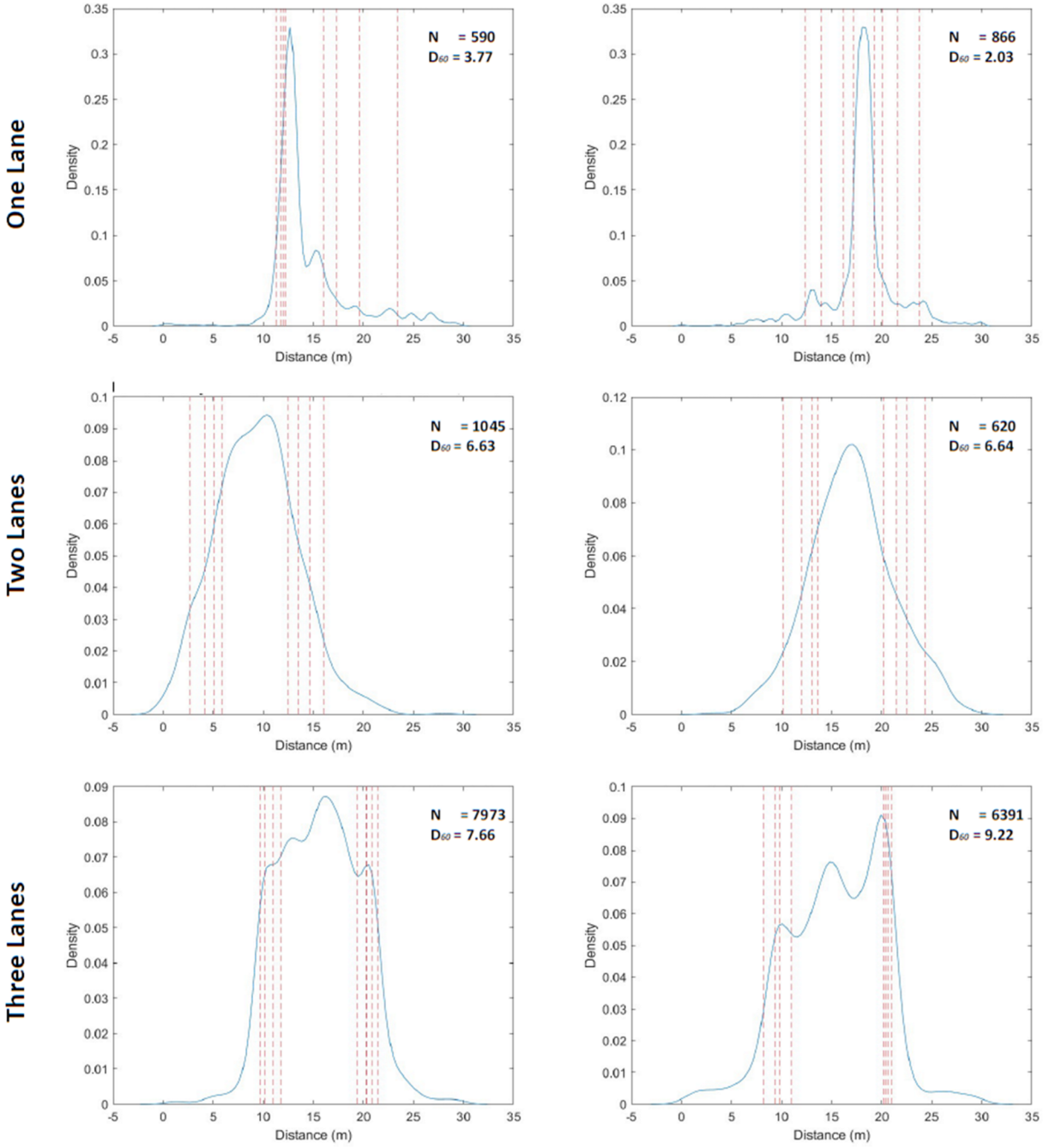


Figure 7 - Sample Kernel Density Estimator of Lateral Distance for One, Two, and Three Lanes

In addition to the distribution profile of lateral distance, the figure also demonstrates the significant difference in the distribution profile between links having one, two, and three lanes. Through observation, it was possible to identify that road segments with fewer lanes have lower values of N and smaller D_{60} values. This can be explained by the fact that roads with fewer users are designed to have fewer lanes, and GPS points are concentrated in a narrower area.

Model specification was carried out to determine the best model and best predictors for the number of lanes. Given the relatively low number of road segments, a 5-fold cross-validation method was performed to avoid overfitting the data. The highest classification prediction accuracy was found using a decision tree classifier at 91% using two predictors, the sample size N and D_{60}. The optimizable decision tree classifier tested iteratively different numbers of splits and different split criteria to reach the minimum classification parameters and error.

Figure 8Figure 9 presents a plot of the two selected predictors, showing a clear delimitation between the predictor values for roads with one, two, or three lanes.

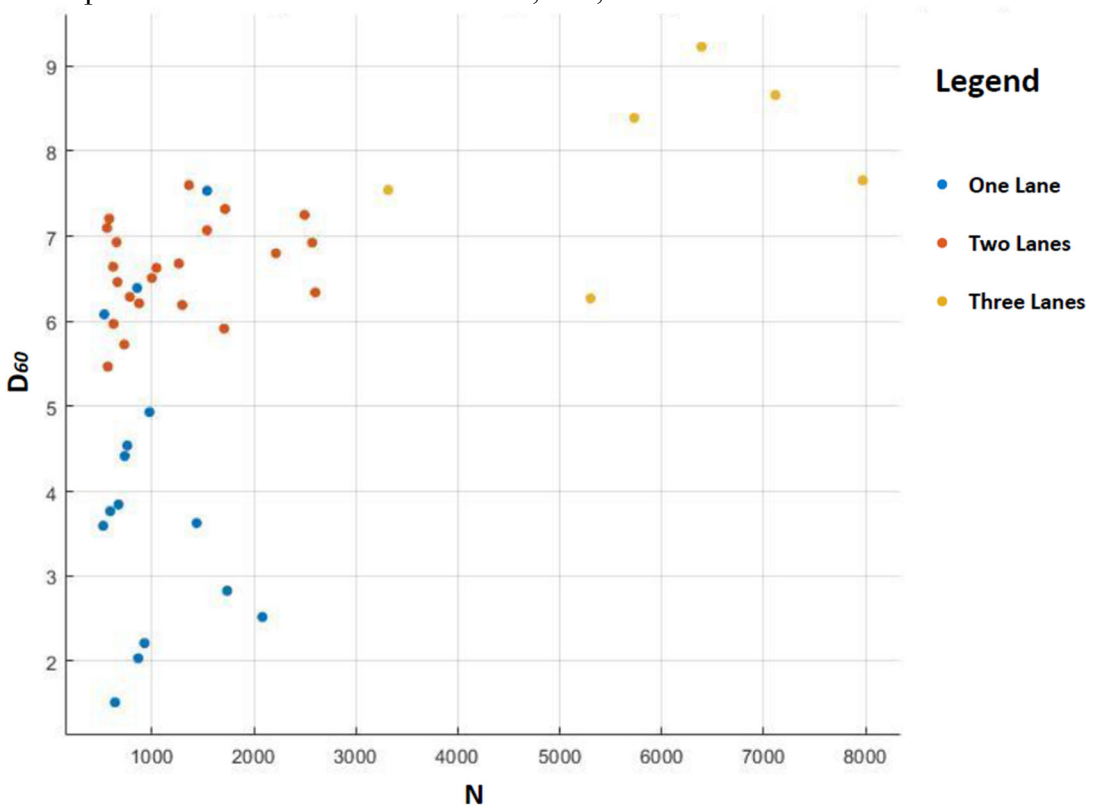


Figure 8 - D60 vs. Sample Size (N)

Moreover, the optimized decision tree is presented with the three split levels and values in Figure 9. Ensemble classifiers, such as boosted trees, bagged trees, and subspace discriminant were also tested to improve prediction accuracy and the best accuracy was using the subspace discriminant ensemble classifier at 91%. Given that the optimized decision tree was able to predict at the same accuracy level it was selected as the best model in this case since it is simpler to visualize and interpret.

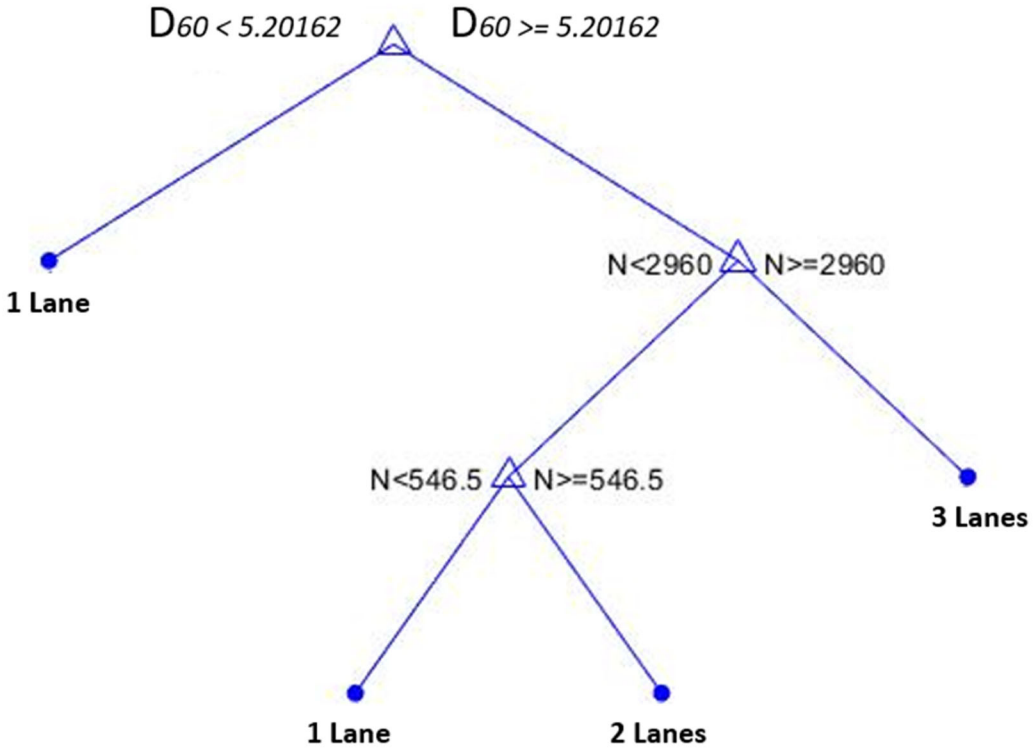


Figure 9 - Selected Classification Decision Tree

Discussion

The proposed methodology predicts the number of lanes per road segment based on the number of GPS points associated to the link and the difference between the 80th and 20th percentile distance, representing a lateral distance measure centered around the median lateral distance.

Given that the best prediction model was obtained using only two variables, an optimized decision tree classifier was sufficient to reach a good model accuracy (91%). However, adding new variables will require retesting ensemble classifier methods to verify if they are able to improve prediction accuracy. Moreover, to use this model, the sample size would need to be translated into relative terms or to be specified with respect to the sample size corresponding to a new dataset. The main hypothesis behind using the sample size as a variable is that for a given period of data collection where we assume a representative sample, it is expected to have a larger number of observations for road segments with a larger number of lanes since they generally have a higher traffic flow.

The spatial analysis and model specification steps were limited by the experimental data available. During the study, it was found that some of the GPS points were map matched in their raw form which signifies that they were snapped to a road centreline at a step prior to accessing the data. Given that this study examines the lateral distribution of raw GPS points with respect to the road link, map matching has a negative impact on data quality. It was also noted that some links had a low number of GPS points, which resulted in unstable lateral distance distribution profiles.

Ideally, larger datasets of uniquely raw GPS points need to be used to have a larger coverage to obtain more realistic distributions and potentially create more predictor variables. The objective is to have more GPS points per link, not necessarily more links as it will also become more complex

to obtain the ground truth information. An increase in the number of points per link will also increase the probability of having better coverage for different times of the day, which enables model specification for different time periods to detect the change in the number of traffic lanes temporally.

Although some studies have proposed the extraction of the number of lanes using satellite and street-level imagery with a relative high accuracy, they are not without limitations (Nieroda et al., 2022, Máttyus et al., 2016, Kasmi et al., 2018). In fact, the high cost of imagery data collection is an important limitation that is overcome in this study since GPS data is currently being crowd sensed by location-based applications through smartphones. Furthermore this study considers 43 road links for model specification which is a larger sample than the work by Tang et al. (2014) which only considers 6 road segments for the analysis.

Comparing this study to some studies using GPS data to extract the number of lanes, the prediction accuracy significantly exceeds the 60% accuracy obtained in the study by Zhang et al. (2010). In addition, the method proposed in the current study provides more accurate results and a simpler procedure than the studies by Chen and Krumm (2010) and Arman and Tampere (2020) to obtain the number of lanes for integration in large-scale transport models.

Conclusion

This study proposes a method to predict the number of lanes per road segment using crowd sensed GPS trajectory data as an input in addition to a simple geographic representation of the road network. The proposed framework is composed of two main steps: to predict the number of lanes of road segments using GPS trajectory data while aiming to keep the cost low and to obtain high prediction accuracy.

The first step is a spatial analysis process to filter and prepare the GPS trajectory data for variable creation. Due to the noise inherent to GPS trajectory, it was crucial to ensure that raw GPS data points were filtered using buffers. This is also necessary to account for the specificities in road design and for the discrepancies in the road network geographic representation. This step also served to produce variables necessary to derive the predictors for the following step. The two main variables were the number of GPS points per road segment and the lateral distance between each point and the reference line representing the road segment. The second step is the training and validation of a machine learning method using classification tree analysis and ensemble learning. Standardized predictors were derived from the lateral distance variables to ensure that the values are comparable across different road segments.

This study was able to develop a road segment number of lanes prediction model using GPS trajectory point data with an accuracy of 91% using a decision tree classifier and two predictors. This prediction accuracy is higher than prediction results obtained by previous research. This finding demonstrates that it is possible to extract the number of lanes available for general traffic by using crowd-sensed GPS trajectory data. This will facilitate road transport network model development and update. The proposed method was demonstrated using a case study in Quebec City, Canada.

However, the work is not without limitations and can be further developed by having a larger temporal sample coverage to enable the prediction of the number of lanes for different periods allowing the detection of dynamic reserved lanes or parking lanes. This study used manually collected ground truth data which limited the size of the study area, network coverage for model development and validation will be increased in future works by collecting more ground truth data or obtaining this information from another source. Moreover, it is possible to explore adding land

use variables that might be correlated with the number of lanes and help in improving the prediction model's accuracy. The potential of this method can also be maximized by automating a procedure that can use GPS trajectory points and other basic input files to create a road network containing the number of lanes per road segment.

Eventually, with the arrival of autonomous vehicles, new data sources may also be available in terms of geotagged imagery data that can be automatically collected and treated by these vehicles during their operation. These processed images may in the future be used to mine road network features at a low cost and high accuracy.

Acknowledgement

The authors would like to acknowledge the generous support of McGill University's Faculty of Engineering and the Vadasz Scholars Program.

Author Contribution

The authors confirm contribution to the paper as follows:
Study conceptualization and design: All authors.
Formal Analysis, Investigation, Writing - Original Draft: Adham Badran
Supervision: Ahmed El-Geneidy and Luis Miranda-Moreno
Interpretation of results and manuscript review and editing: All Authors

References

ARMAN, M. A. & TAMPERE, C. M. J. 2020. Road centreline and lane reconstruction from pervasive GPS tracking on motorways. *Procedia Computer Science,* 170**,** 8.

BOUNINI, F., GINGRAS, D., LAPOINTE, V. & POLLART, H. Autonomous Vehicle and Real Time Road Lanes Detection and Tracking. 2015 IEEE Vehicle Power and Propulsion Conference (VPPC), 19-22 Oct. 2015 2015. 1-6.

CHEN, Y. & KRUMM, J. Probabilistic modeling of traffic lanes from GPS traces. Proceedings of the 18th SIGSPATIAL international conference on advances in geographic information systems, 2010. 81-88.

GUO, C., KIDONO, K., MEGURO, J., KOJIMA, Y., OGAWA, M. & NAITO, T. 2016. A Low-Cost Solution for Automatic Lane-Level Map Generation Using Conventional In-Car Sensors. *IEEE Transactions on Intelligent Transportation Systems,* 17**,** 2355-2366.

GUO, Y., LI, B., LU, Z. & ZHOU, J. 2021. A novel method for road network mining from floating car data. *Geo-spatial Information Science***,** 16.

IAN, H., FRANK, E., HALL, M. & CHRISTOPHER, J. 2017. Data mining: Practical machine learning tools and techniques—Part II: More advanced machine learning schemes. Morgan Kaufmann, Burlington, MA.

KADALI, B. & VEDAGIRI, P. 2020. Role of number of traffic lanes on pedestrian gap acceptance and risk taking behaviour at uncontrolled crosswalk locations. *Journal of Transport & Health,* 19**,** 100950.

KASMI, A., DENIS, D., AUFRERE, R. & CHAPUIS, R. Map Matching and Lanes Number Estimation with Openstreetmap. 2018 21st International Conference on Intelligent Transportation Systems (ITSC), 4-7 Nov. 2018 2018. 2659-2664.

LEICHTER, A. & WERNER, M. 2019. Estimating road segments using natural point correspondences of GPS trajectories. *Applied Sciences-Basel,* 9**,** 11.

MÁTTYUS, G., LUO, W. & URTASUN, R. Deeproadmapper: Extracting road topology from aerial images. Proceedings of the IEEE international conference on computer vision, 2017. 3438-3446.

MATTYUS, G., WANG, S., FIDLER, S. & URTASUN, R. Enhancing road maps by parsing aerial images around the world. Proceedings of the IEEE international conference on computer vision, 2015. 1689-1697.

MÁTTYUS, G., WANG, S., FIDLER, S. & URTASUN, R. Hd maps: Fine-grained road segmentation by parsing ground and aerial images. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016. 3611-3619.

MERRY, K. & BETTINGER, P. 2019. Smartphone GPS accuracy study in an urban environment. *PLOS ONE,* 14**,** e0219890.

MONTRÉAL, V. D. MTL Trajet Study.

NIERODA, B., WOJAKOWSKI, T., SKRUCH, P. & SZELEST, M. A Heatmap-Based Approach for Analyzing Traffic Sign Recognition and Lane Detection Algorithms. 2022 26th International Conference on Methods and Models in Automation and Robotics (MMAR), 22-25 Aug. 2022 2022. 217-221.

TANG, L., GAN, A. & ALLURI, P. 2014. Automatic Extraction of Number of Lanes from Georectified Aerial Images. *Transportation Research Record,* 2460**,** 86-96.

TREIBER, M. & KESTING, A. 2013. Traffic flow dynamics. *Traffic Flow Dynamics: Data, Models and Simulation, Springer-Verlag Berlin Heidelberg***,** 983-1000.

ZHANG, C., XIANG, L., LI, S. & WANG, D. 2019. An intersection-first approach for road network generation from crowd-sourced vehicle trajectories. *ISPRS International Journal of Geo-Information,* 8**,** 26.

ZHANG, L., THIEMANN, F. & SESTER, M. Integration of GPS traces with road map. Proceedings of the Third International Workshop on Computational Transportation Science, 2010. 17-22.

ZHONGYI, N., LIJUN, X., TIAN, X., BINHUA, S. & YAO, Z. 2018. Incremental road network generation based on vehicle trajectories. *ISPRS International Journal of Geo-Information,* 7**,** 19.