BILINGUAL PUBLISHING CO.
Pioneer of Global Academics Since 1984

Journal of Geographical Research
https://ojs.bilpublishing.com/index.php/jgr

ARTICLE

A Hybrid Geostatistical Method for Estimating Citywide Traffic Volumes – A Case Study of Edmonton, Canada

Mingjian Wu[*] Tae J. Kwon Karim El-Basyouny

Department of Civil and Environmental Engineering, University of Alberta, Edmonton, AB, T6G2W2, Canada

ARTICLE INFO

Article history
Received: 10 March 2022
Revised: 30 March 2022
Accepted: 05 April 2022
Published Online: 16 April 2022

Keywords:
Traffic volume
Geographical information system
Spatial modelling
Hybrid geostatistics
Network regression kriging

ABSTRACT

Traffic volume information has long played an important role in many transportation related works, such as traffic operations, roadway design, air quality control, and policy making. However, monitoring traffic volumes over a large spatial area is not an easy task due to the significant amount of time and manpower required to collect such large-scale datasets. In this study, a hybrid geostatistical approach, named Network Regression Kriging, has been developed to estimate urban traffic volumes by incorporating auxiliary variables such as road type, speed limit, and network accessibility. Since standard kriging is based on Euclidean distances, this study implements road network distances to improve traffic volumes estimations. A case study using 10-year of traffic volume data collected within the city of Edmonton was conducted to demonstrate the robustness of the model developed herein. Results suggest that the proposed hybrid model significantly outperforms the standard kriging method in terms of accuracy by 4.0% overall, especially for a large-scale network. It was also found that the necessary stationarity assumption for kriging did not hold true for a large network whereby separate estimations for each road type performed significantly better than a general estimation for the overall network by 4.12%.

1. Introduction

Traffic volume information has long played an important role in various transportation research areas, including but not limited to policy making, roadway design, safety analysis and air quality control [1-4]. However, collecting traffic count data over a large spatial area requires a significantly substantial, and potentially prohibitive, amount of manpower and financial costs [5]. For these reasons, local governments and transportation agencies are not able to place enough monitoring equipment for the whole road network, resulting in gaps and shortages of available data. Considering the irreplaceability of traffic volume data in transportation-related studies, the focus must advance from just 'measuring' to 'well-estimating' [6]. Thus, over many years, transportation researchers and

**Corresponding Author:*
Mingjian Wu,
Department of Civil and Environmental Engineering, University of Alberta, Edmonton, AB, T6G2W2, Canada;
Email: mingjian.wu@ualberta.ca

DOI: https://doi.org/10.30564/jgr.v5i2.4513

practitioners have made considerable efforts to produce reliable estimations of traffic volumes by applying various methodologies and tools.

One of the most widely adopted methods to date is Ordinary Least Squares (OLS) regression while using as many explanatory variables as possible. For example, Xia et al. estimated traffic volumes for nonstate roads [7]. In their study, 450 count-monitoring stations were involved in the investigation with up to 12 initial variables (accessibility, socioeconomic, and road characteristics) included in the multiple regression model. Using a similar method, Zhao and Chung [8] predicted the annual average daily traffic (AADT) on multiple roadways in an urban area through a variety of land-use and accessibility measures. Mohamad et al. utilized a multiple regression model to incorporate relevant demographic variables such as population, state highway mileage, per capita income, and the presence of interstate highways [9].

Various methodologies other than regression models also have been employed. Tang et al. [10] compared time series, neural network, nonparametric regression, and Gaussian maximum likelihood (GML) techniques and found that GML methods were the most promising. Lam et al.'s [11] more recent study used nonparametric models and GML methods to forecast Hong Kong's traffic volumes. Davis and Yang proposed empirical Bayesian methods that were used to compute quantiles of the predictive probability distribution of the total traffic at a highway station, given a sample of daily traffic volumes from that station [12]. These activities were based on expressions for traffic data variability and the American Association of State Highway and Transportation Officials (AASHTO) reliability concept for pavement design. Goel et al. proposed a correlation-based approach to improve AADT estimation by exploiting the inherent underlying correlations between link flows [13]. These correlations arise partially because the inflows and outflows to a node are always constrained. In addition, when the network has many origin–destination (O-D) zones, and a relatively smaller number of links, the correlation between the link flows can be large.

With the growing availability of geographic information system (GIS) datasets and the evolution of spatial analysis techniques, researchers have started to explore geostatistical methods [14,15] that exploit the spatial context of traffic and other spatial data. Studies that are most closely related to this topic include Eom et al.'s [6] use of spatial regression methods to predict AADT for nonfreeway facilities in Wake County, North Carolina, and Selby B. and Kockelman, K.M. [16] compared the AADT estimation results between universal kriging and geographically weighted regression for Texas. They conclude that the overall predictive capability of kriging methods eclipses traditional models as well as other spatial methods. Kriging presumes spatial dependence in error terms or unobserved factors, as a function of distance. Due to its capability of accounting for the variable's interaction over space and its ease of implementation for interpolating target variables at unmeasured locations, this method continues to gain popularity for application in many transportation fields in recent years [17-20]. Like all the other geostatistical methods, kriging uses Euclidean distance to quantify the spatial separation between variables during its interpolation procedures. However, when this approach is transferred to studies targeting traffic specific variables, logically the actual network distance should be a more reasonable metric. Some prior transportation-related research studies replaced Euclidean distance with network distance in their kriging models. For instance, Zhang and Wang [21] refined the standard kriging model by using subway system network to estimate the transit ridership in New York City. Although their findings indicate that kriging using subway routes outperforms the standard ones, they are based on two short subway lines whereby making their assessment less is conclusive. Likewise, a road network is much more complex than the subway network thus greater scrutiny is necessary in order to appreciate the true benefit of implementing road network distances for estimating traffic volumes. Selby B. and Kockelman, K.M. [16] evaluated universal kriging in estimating AADT using Euclidean distances and network distances for Texas but their dataset was limited to only one-year and may not be conclusive due to its limited temporal coverage. Datasets with a longer term of traffic volume observations need to be used to further validate its estimation performance.

Table 1 summarizes the reviewed literature including their corresponding study areas, road types, findings/contributions, and limitations. In summary, a variety of models have been proposed and developed to estimate traffic volume information with kriging proving to be a promising method for generating reliable estimates. Nonetheless, there exist very few studies that incorporate local auxiliary information into the kriging method, and no studies that utilize city-scale and long-term traffic volume datasets. Likewise, due to the limited datasets and significantly higher computational costs, kriging with network distances has rarely been used in transportation studies involving a large urban network. Furthermore, the second-order stationarity assumption, which assumes the spatial pattern of a target variable is the same across the entire study area is the most fundamental assumption behind kriging [15,22]. However, whether or not this holds true for traffic volumes across the urbanized network has not been investigated in past literature.

Therefore, the primary objective of this research is to develop a hybrid geostatistical interpolation framework, as well as validate its feasibility and robustness through a real-world case study. To summarize, the main novelty and contributions of this research work include the following:

• Utilized local auxiliary variables to detrend the data within the interpolation framework, and to explain variations in traffic volumes with respect to different locations,

• Replaced Euclidean distance with road network distance to improve the estimation accuracy and model explanation power,

Table 1. Summary of the Selected Relevant Studies of Traffic Volumes Estimation

Author(s)	Study Area	Road Types	Method	Findings/Contributions	Limitations
Mohamad, et al., 1998	Indiana	• County roads	• Multiple regression	• Data from multiple counties were used to validate the accuracy of the developed model	• Regression models are not able to spatially map traffic volumes across the entire road network without adequate auxiliary variables as input. • Adopted methods lacks the considerations of spatial characteristics of road networks.
Xia, et al., 1999	Broward County, Florida	• Nonstate roads	• Multiple regression	• 12 initial variables were involved in regression models	
Zhao & Chung, 2001	Broward County, Florida	• Interstate highways • Express ways • Urban roads • Rural roads	• Multiple regression	• More explanatory variables with modified data and road classifications were used to extend the previous efforts	
Tang, et al., 2003	Hong Kong	• Major road links	• Time series model • Neural network • Nonparametric regression (NPR) • Gaussian maximum likelihood (GML)	• GML model appeared to generate the most accurate predictions among the four models.	• Limited data input (road types, temporal coverage, etc.). • Traffic volumes were assumed to be normally distributed and time dependent. • Lack of spatial analysis of traffic volumes.
Lam, et al., 2006	Hong Kong	• Major road links	• Time series model • Neural network • Nonparametric regression • (NPR) • Gaussian maximum likelihood (GML)	• Extended efforts using more observations to predict the hourly traffic flows. • NPR and GML appeared to be more promising.	
Davis & Yang, 2001	NA	• Highways	• Empirical bayes method	• Proposed method could estimate the probable ranges and associated probability distribution.	• Included road type is highway only. • Adopted method is only able to provide point estimates.
Goel, et al., 2005	NA	• Highways	• Correlation-based method	• Developed method exploit the correlation imposed by O-D path flow.	• Adopted method is sensitive to the input data configuration.
Eom, et al., 2006	Wake County, North Carolina	• Urban • Suburban • Rural	• Spatial regression • Universal kriging (UK)	• Comparisons between spatial regression and ordinary regression method proved the former one could provide better predictions.	• Euclidean distance was used in developing the models; however, it is not the true representation of the distance in road network. • Lack of spatial interaction analysis of the traffic volumes.
Selby & Kockelman, 2013	Texas	• From local roads to interstate freeways	• Universal kriging (UK) • Universal network kriging • Geographically weighted regression (GWR)	• UK provided better results than GWR in the study area. • UK with network distance showed no enhanced performance.	• Input data were lack of long-term observations. • The network distance piece was only evaluated in a single network setting.
Zhang & Wang, 2014	New York City	• Subway Line	• Network universal kriging	• With using network distance in kriging model, the standard kriging method using Euclidean distance was improved in terms of the spatial variability analysis.	• A few subway lines were selected for the study, which was a relatively simple network. • Lack of long-term observed data for model performance evaluation.

• Evaluated the applicability of the proposed method using a large-scale road network (i.e., city of Edmonton),

• Investigated the second-order stationarity assumption of kriging in relation to urban network traffic volume,

• Validated the consistency of the research findings using long term traffic volume data (i.e., 10 years)

In particular, this research also aims at answering the following questions which represent the main objectives of this study.

• What local factors would contribute to enhancing the traffic volume estimations?

• To what extent and in what situations would kriging with network distances outperform standard kriging with Euclidean distances?

• How to characterize the spatial dependence (spatial variation pattern) of traffic volumes and apply them to different network configurations?

The results of this study can be an important reference source for gaining knowledge on network regression kriging and network spatial variation pattern of traffic volumes. From a practical application perspective, since kriging is capable of spatially mapping traffic volume for all unmeasured locations across an urban area, our proposed framework provides a tool for transportation practitioners and engineers to estimate traffic volume at locations abstinent of any traffic information. With complete and accurate estimations of traffic volumes, city planners and transportation agencies will be able to benefit from this in various ways [23].

2. Methodology

2.1 Regression Kriging – The Idea

As described previously, a variety of techniques have been implemented to estimate traffic volumes. Each method takes known traffic counts and/or uses auxiliary information (e.g., land use, time-steps, road geometry attributes, etc.) to make an estimation. These estimates can be separated into future-year and present-year predictions. Future-year predictions use current and past traffic count data to estimate the traffic counts at a future date. On the other hand, current-year predictions use the available data to estimate traffic counts at unmeasured locations. Kriging is typically applied in the latter situation [24,25].

Kriging is one of the most commonly used geostatistical interpolation techniques that account for the uncertainty of the estimation. It relies on the second order stationarity assumption where the mean, variance and dependence structure of the target variable do not change over space [22]. Kriging predicts the values at unsampled locations from the weighted average of nearby measured observations. The weights are determined based on their distance from the unsampled location and their closeness to each other. Commonly used variants of the kriging method include Ordinary Kriging (OK) and Regression Kriging (RK). The main difference between the two methods is that OK assumes the mean of the target variable to be unknown but constant locally, whereas RK assumes there to be a trend associated with the auxiliary variables (e.g., longitude, latitude) [26]. In this study, RK is adopted as it incorporates two conceptually different methods to model and map spatial variability to strengthen the explanation of the target variable. The estimations are made separately for the trend (OLS in this study) and residuals (via Kriging), and then added back together as shown in Equation (1).

$$\hat{Z}(x) = \hat{m}(x) + \hat{e}(x) = \sum_{i=0}^{p} \hat{\beta}_i \cdot q_i(x) + \sum_{k=1}^{m} \lambda_k \cdot e(x_k) \qquad (1)$$

where $\hat{m}(x)$ is the fitted deterministic component (i.e., the trend), $\hat{e}(x)$ is the interpolated residual, $\hat{\beta}_i$ are coefficients of the estimated trend model, $\hat{\boldsymbol{\beta}}_0$ represents constant term, p is the number of auxiliary variables, λ_k are kriging weights and $e(x_k)$ is the regression residual. Kriging weights for each sampling location is estimated based on the parameters of the semivariogram model, introduced in the following paragraphs, and the relative distance of the specific point with other sampling points and the unknown point.

2.2 Semivariogram

As previously mentioned, the semivariogram model depicts the spatial dependence of the measured sample points [14,25,27], and is used for linear interpolation via RK, thus constructing a good quality semivariogram is critical as it determines the accuracy of estimation results. A semivariogram is the plot of the expected value of the semivariance of the variable of interest. It is a statistic that shows how the level of similarity between two known points decreases as their separation distance increases [27]. The semivariance value can be calculated by taking the average of the squared difference of two measurements in a study domain separated by a specific and defined lag distance. The equation generally used for semivariogram estimation is shown below:

$$\gamma(h) = \frac{1}{2n(h)} \sum_{i=1}^{n(h)} [z(x_i + h) - z(x_i)]^2 \qquad (2)$$

Here, $\gamma(h)$ is the semivariance value. $z(x_i + h)$ and $z(x_i)$ are two measurements taken at the location x_i and $(x_i + h)$ which are separated by a lag distance, h. Figure 1 shows a typical semivariogram plot [18].

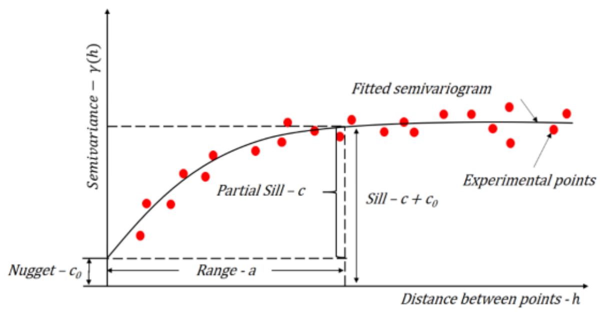


Figure 1. A Typical Semivariogram with Parameters

Three basic parameters associated with each semivariogram are the nugget, range and sill. According to theory, the semivariogram value at the origin should be zero, but due to stochastic errors such as measurement bias, the value of the semivariogram at the origin could differ significantly from zero and this is known as the nugget effect. The semivariance value at which the semivariogram levels off is known as the sill parameter. The lag distance at which the semivariogram reaches the sill value is known as the spatial range beyond which spatial dependence is considered non-existent. To assess the strength of the spatial dependence, the nugget-to-sill ratio (NSR) is typically utilized as a dimensionless measure of the proportion of total observed variation that could not be explained by the observed spatial dependence of the target variable [28,29]. In other words, a small NSR represents a strong spatial dependence while a big NSR reflects a weak spatial dependence of the variable. To fit an experimental semivariogram based on observations made, there are several common theoretical forms [30,27,31] can be selected. The selection of fitted model types and adjustment of semivariogram parameters typically have effect on kriging estimation accuracy. However, these detailed selection/comparison processes are not the focus of this presented study, so to enforce a fair comparison between the different interpolation strategies, the spherical model (Equation (3)) was adopted for all.

Spherical model:

$$\gamma(h) = \begin{cases} 0 & ,h = 0 \\ c_0 + (c - c_0) * \left(1.5\left(\frac{h}{a}\right) - 0.5\left(\frac{h}{a}\right)^3\right) & ,0 < h < a \\ c & ,h \geq a \end{cases} \quad (3)$$

Here, h = lag distance, a = spatial range of continuity and c = sill.

2.3 Hybrid Geostatistical Interpolation Framework

The standard RK method, including the construction of the semivariogram model, is all based on the Euclidian distance metric; however, this might not be a reasonable way of interpolating traffic-related variables (i.e., traffic volumes in this study) which may strongly be correlated with the road network. Therefore, when constructing the semivariogram, Euclidean distances are replaced with road network distances thus providing a better representation of the spatial dependence between the measured traffic count points. This simple change makes the standard RK wholly based on the road network distance. Figure 2 illustrates the key steps of the hybrid geostatistical interpolation framework using network regression kriging. It is worth noting that the proposed framework as a whole is transferrable to other cities. Additionally, the target variable (traffic volume in this study) can be replaced with other traffic-related variables (e.g., traffic collisions, congestions, etc.); however, the detailed parameters (e.g., auxiliary variables for detrending data) involved in this framework needs to be recalibrated, which will depend on the type and availability of the input data.

The first step is to obtain the network distances between each point. This includes the distances between measured points and also the distances between measured points and unmeasured points of interest. This can be done by the Network Analyst extension in ArcGIS [32]. The second step is to remove the trend (i.e., the deterministic component) of the measured data points, to show only the true spatial variation of the traffic volumes. The detrended traffic volumes (i.e., the residuals) would be used as the input into the semivariogram modelling where python [33] and scikit-gstat package [34] are utilized. Kriging is then used to estimate the residuals for each point based on the semivariograms constructed in the previous step, and the final traffic volume estimates can be obtained by adding the estimated residuals and the deterministic components. Olea's study regarding the steps of constructing an appropriate semivariogram model is used as a guideline during the whole process [27].

To ensure that the semivariogram model is of good quality and to compare the estimation accuracy between different estimation strategies, cross-validation is utilized to check the estimation performance. In this study, the root-mean-square-error (RMSE) was calculated using the "leave one out" method that checks the prediction error of each point individually to ensure that the overall prediction accuracy is acceptable [35,27]. RMSE is an indicator of how closely the model estimates matches the measured values. For this reason, it is widely used in cross validation to evaluate the quality of the constructed semivariogram. The equation of RMSE can be found below.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{i=n}(Z_i - \widehat{Z}_i)^2}{n}} \quad (4)$$

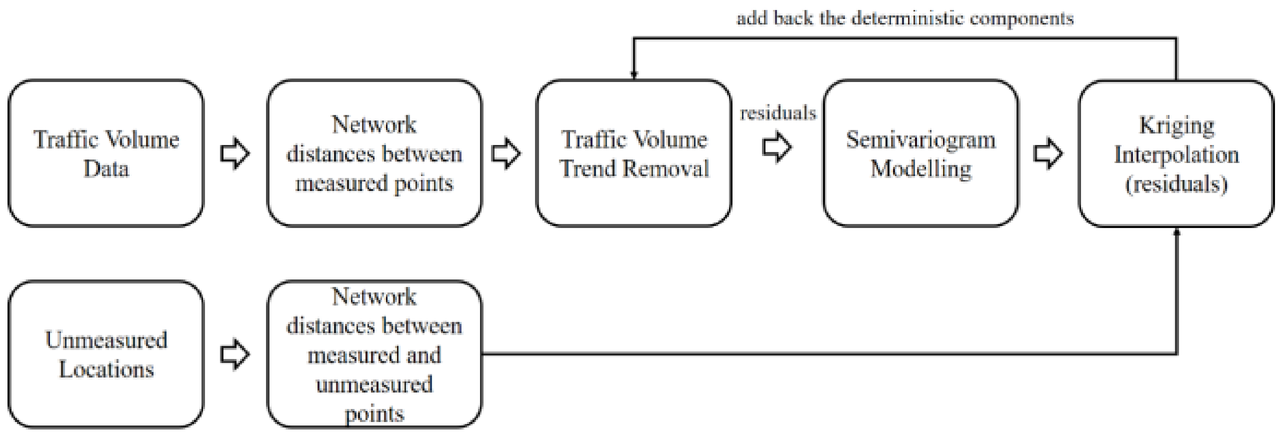


Figure 2. Proposed Hybrid Geostatistical Interpolation Framework

Here, n is the total number of measured points.

3. Case Study

3.1 Study Area: Edmonton, Alberta, Canada

As mentioned above, the road network for the city of Edmonton is used for the case study. To date, there are 233 arterials (1894 km), 699 collectors (886 km) and 2,394 local roads (2624 km) in total. Based on the function and traffic conditions of the city districts, the whole city is divided into two zones. The core zone is the dense areas in the centre of the city that have far more active transportation mode use like walking, cycling and public transportation. It mostly consists of local and collector roads and carries the main functions of residence and commerce. By contrast, the non-core zone is the rest of the city where social vehicles and trucks are the main modes of transportation. It is mainly used by commuters, intercity travelers, industrial, and agricultural industries.

3.2 Data Sources and Processing

Traffic volume data were obtained from the City of Edmonton Open Data Portal where the city road network shapefile and historical (2008-2018) traffic volume data were provided. The traffic volume data were collected by loop detectors or by manual counting on weekdays. Since not all the measurement locations had year-round monitoring, average-daily-traffic volume (ADT) for each detection site was calculated and used in this study. Table 2 summarizes the collected ADT data with descriptive statistics, while Figure 3 displays the study areas along with the different road types and the boundary of the core zone. All detection sites are also shown in the same figure. The year 2017 was not included in the following analysis due to insufficient data. The road network distance between each pair of ADT points was obtained using the Network Analyst extension in ArcGIS [32], which lets us find the shortest route between locations along a network of transportation routes.

3.3 Trend Removal of Traffic Volumes

As previously described, the first step involved in the framework is to remove the trend of the target variable (i.e., the ADT) as much as possible so that the true spatial variation pattern of ADT can be observed. In this study, the linear regression with ordinary least squares (OLS) was adopted in removing the trend or the deterministic part of the kriging model. It is worth noting that linear regression is not the only method that can be employed here, however, the focus of this study is not on the detrending techniques while the linear regression model is one of the most widely used and effective methods that allows us to gain a deeper understanding of the influence that each covariate has in explaining the variations of ADT [6,20,16,21,8]. Available local auxiliary information in Edmonton for each detection site included the road type (eight types in total which were converted into numerical variables according to road class level), speed limit (30 km/h – 110 km/h) and local accessibility factors such as the distance to the nearest school and distance to the nearest senior residence, all of which were used in the stepwise multiple linear regression analysis by fitting a first-order polynomial model. The p-value was adopted to confirm the statistical significance between ADT and the auxiliary variables in a 5% significance level. Table

Table 2. Descriptive Statistics of the Collected Average Daily Traffic (ADT) Volume Data

Year	# measured ADT points	Maximum ADT	Minimum ADT	Average ADT	Standard Deviation	Proportions in Arterial/Collector/Other roads
2008	415	26196.0	17.5	7478.8	6034.6	84.3%/ 15.0%/ 0.7%
2009	425	33989.0	28.5	5742.3	5961.9	68.7%/ 30.8%/ 0.5%
2010	494	39520.0	38.0	5796.0	5956.6	68.2%/ 31.1%/ 0.6%
2011	581	45142.0	59.0	6070.8	7337.3	62.5%/ 37.1%/ 0.4%
2012	521	52727.0	49.5	8269.6	8965.1	75.0%/ 24.8%/ 0.2%
2013	700	50818.0	16.5	9546.7	9189.5	82.8%/ 15.8%/ 1.3%
2014	905	53171.5	56.5	10016.5	9141.8	84.7%/ 14.5%/ 0.8%
2015	1174	60066.0	78.5	10646.2	9209.0	81.6%/ 17.7%/ 0.6%
2016	1035	47176.0	43.5	9961.5	7896.4	85.3%/ 13.8%/ 0.9%
2017	8	7129.0	561.5	3018.0	2406.9	62.5%/ 37.5%/ 0.0%
2018	238	49487.0	130.5	9105.0	7646.4	89.9%/ 8.0%/ 2.1%

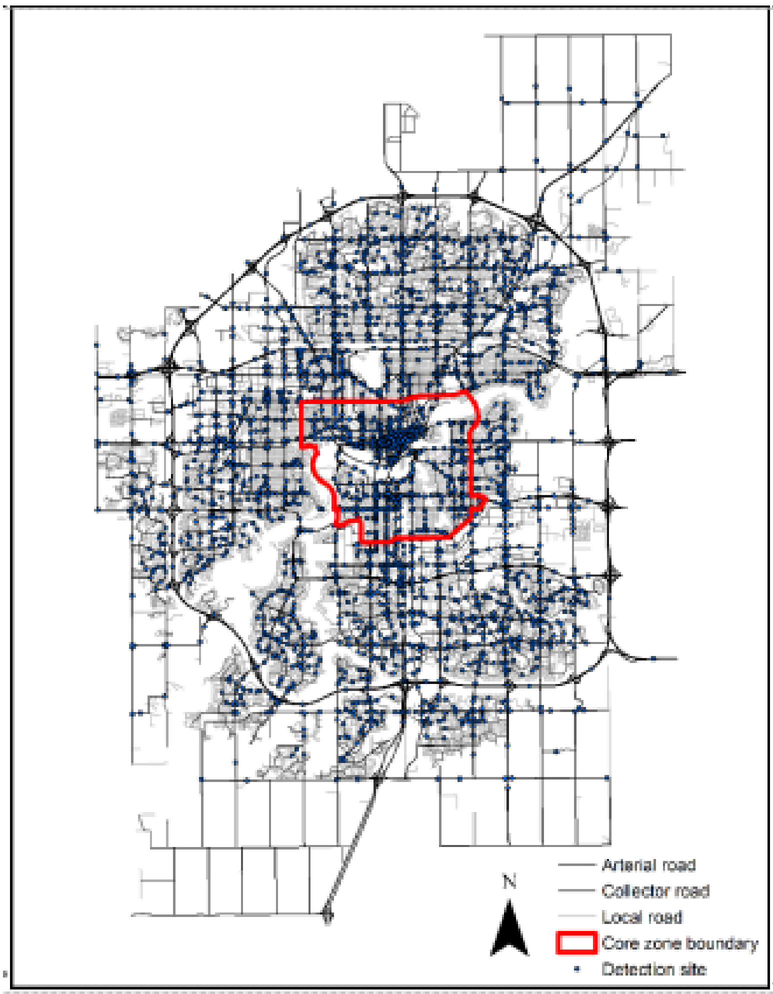


Figure 3. The Study Area – Edmonton, Alberta

Table 3. Linear Regression Results for Traffic Volume Trend Removal

Year	Road Type	Speed Limit	Distance to Nearest Senior Residence	Distance to Nearest School	R2
2008	–5185.387	1687.224	99.019	–	30.6%
2009	–10800.105	1330.000	223.670	–	36.4%
2010	–12610.012	1270.851	256.203	–	39.6%
2011	–17979.301	1308.853	375.761	–1.234	48.0%
2012	–21996.595	1369.502	447.581	–1.278	59.2%
2013	–22324.391	2185.237	389.300	–	51.2%
2014	–22547.419	1991.376	414.489	–	48.3%
2015	–24068.033	2045.341	461.151	–1.224	53.1%
2016	–17932.787	2082.277	327.148	–	42.0%
2018	–30951.479	899.345	647.308	–	52.0%

Note: eight road types are converted into numerical variables.
- Arterial-Class A (Primary Highway Truck Route) = 8;
- Arterial-Class B (Non-Primary Highway Truck Route) = 7;
- Arterial-Class C (Truck Route Low speeds) = 6;
- Arterial-Class D (Non-Truck Route Low speeds) = 5;
- Collector-Commercial (Adjoining lots zoned > 50% Commercial) = 4;
- Collector-Industrial (Adjoining lots zoned > 50% Industrial) = 3;
- Collector-Residential = 2;
- Other (e.g., local roads) = 1.

3 summarizes the calibrated parameters, and only those variables with *p*-values less than $\alpha = 0.05$ are included.

As can be seen from the linear regression results, road type, speed limit, and distance to the nearest senior residence are all consistently significant variables that can explain the variation of ADTs in each year. In 2011, 2012 and 2015, the distance to the nearest school is also a significant variable in this step. Signs of the calibrated coefficients all make sense, as higher classes of roads (e.g., primary highway truck route) hold more traffic volumes compared with lower classes (e.g., local roads); roads with higher speed limits also tend to hold more traffic volumes as they are either interstate highways or arterials that serve the most commuters. In terms of the local accessibility factors that may only be applicable to Edmonton, the shorter the distance to a senior residence is, the larger the traffic volume tends to be, as most of the senior residences in Edmonton are located near major arterial roads. This also partially applies to the schools, but it depends on the locations of detection sites in specific years, implying that changing sampling locations may have an impact on the estimation performance of the regression.

Although the regression models were able to moderately remove underlying spatial trend, the low R^2 values indicate the need for performing spatial interpolation via kriging using residual values to further improve the estimation accuracy. The next is to construct semivariogram models based on the residuals of the regression models and implement regression kriging under different scenarios and road network configurations.

3.4 Semivariogram and Regression Kriging

In this section, the comparison results including the spatial dependence and estimation accuracy between standard kriging and network kriging are presented. As shown in Table 2, ADT values can vary a lot between each year, and to understand the network model performance in different network settings, as well as assess the stationarity assumption of ADT, the study repeated the same analysis procedures for three different network configuration groupings. The first grouping (Case I) considers the whole city-wide network, the second grouping (Case II) is the district functions of core and non-core zones, and the final grouping (Case III) is a separation between arterial and collector road types. As mentioned previously, all semivariograms were fitted using the spherical model in order to enforce a consistent and fair comparison.

3.4.1 Case I: Whole Network

Following the steps of the standard kriging and network kriging (Figure 2), the semivariogram models of two distance metrics for each year were first constructed based on the same data points. Figure 4 includes all the semivariogram models that were developed by carefully following the procedures and guidelines from Olea's publication [27].

As can be seen from Figure 4, the general shapes and sill values remain similar in each pairing while the nugget values of network models are consistently smaller compared with the standard ones. This also leads to different spatial dependence patterns between the two models. Figure 5 vividly shows the comparisons between the four parameters of the two types of semivariogram models, *t*-tests were also employed to examine the significance of the differences (Note: * represents the *p*-value is significant at 5% significance level).

From the *t*-test values, it can be concluded that the nugget and the nugget-to-sill ratio (NSR) values of network-distance semivariogram models are significantly smaller as compared to their Euclidean distance counterpart. This means the stochastic effect or the variance of ADT within a very short lag distance are reduced dramatically and thus the stronger spatial dependence of ADT can be observed. The reason can be the use of network distance takes the actual road network into account which avoids the situation where one low traffic volume road segment is considered adjacent to a highway segment even if there is no direct connection between them. In addition, the nugget and NSR values change dramatically over the 10-year period, which implies that the spatial configuration of the ADT detection sites has an impact on the representativeness of the ADT spatial variation pattern. The travel behavior changes over time (i.e., temporal pattern) in the city of Edmonton is also one of the possible reasons. This will need further investigation in the future, but unfortunately, it will be out of scope for this study. Figure 5 suggests that there is no significant difference present in the sill values, and this is because sill values represent the maximum variance between ADT values within the sampling scale, points that are far away from each other in Euclidean distances are likely to be far away from each other in network distance as well, thus the use of network distance does not affect sill value much. For the range values of semivariograms, it is a parameter that is highly correlated with the ADT sampling distribution and thus makes estimating the true range complicated hampering its use in comparisons [36]. Thus, for the other two groups, only the nugget and NSR values calibrated from their corresponding semivariograms were used for comparisons.

After the semivariogram models were constructed, the cross-validation was implemented to compare the estimation accuracy between the two-distance metrics. RMSE values suggest that using network distance led to an average improvement of 4.0% for all of the years. The *t*-statistic between them is 5.473 with the *p*-value being equal to 0.000, which means network kriging is able to consistently provide significantly more accurate estimates compared to the standard one for the whole city road network.

It is worth noting that the percentages of the improvement vary for each year (i.e., 1% to 9%). The reason for the variation can be attributed to the sampling distribution of the measured ADT points, as the city keeps expanding its scale during these 10 years thus changing most of the locations of the ADT detection sites every year.

Since the computational costs of the network model are exponentially higher than the standard one, it is important to distinguish the situations under which applying network kriging is worth the increased effort. In other words, there is a need to identify the situations where the network semivariogram model and RK method can significantly outperform the standard ones.

3.4.2 Case II: Core Zone and Non-core Zone

The same comparison procedures are repeated between core zone (downtown) and non-core (suburb) zone of the city to see if the model would perform differently under varying network configurations. Calibrated nugget and NSR values are compared in Table 4. Years with insufficient data for both zones are excluded from this analysis.

Although results show that the network kriging model is able to reduce the nugget effect, better represent spatially dependence, and generate more accurate results in both core and non-core zones, the *t*-test values indicate that the differences were seen in the RMSE values between standard and network models are not statistically significant in the core zone (*t*-statistic: 0.919; *p*-value: 0.194) while they do differ significantly in the non-core zone (*t*-statistic: 2.057; *p*-value: 0.039). The reason for this can be attributed to the density and scale of the networks. For example, the core zone has a smaller spatial scale than that of the non-core zone and that varying distance metrics do not play a significant role due to having higher network density. Additionally, the shape of the road segments within the core-zone is mostly straight thus making the differences between Euclidean and network distance kriging less dramatic.

Therefore, it can be concluded that the marginal benefits gained by using network kriging on a small and dense network with straight road segments is not worth the extra effort. By contrast, a widely spread-out network, or a network with a considerable number of curvy roads, can benefit significantly more from the network model. Additionally, calibrated nugget and NSR values in Table 4 also imply the same conclusion as discussed in Case I, regardless of the zone area, ADT detections sites' configuration affects the semivariogram model. The temporal pattern of travel behavior also needs future investigation. Another point worth mentioning here is that the semivariogram models for each year were very different between the core and non-core zones, which infers that the stationarity assumption (i.e., translation invariance) of ADT may not hold true across large urban area. In other words, one single semivariogram model may not be able to represent the spatial pattern of all ADTs in the network as a whole thereby warranting a need to develop separate semivariogram models by taking into account the underlying characteristics of road networks under investigation.

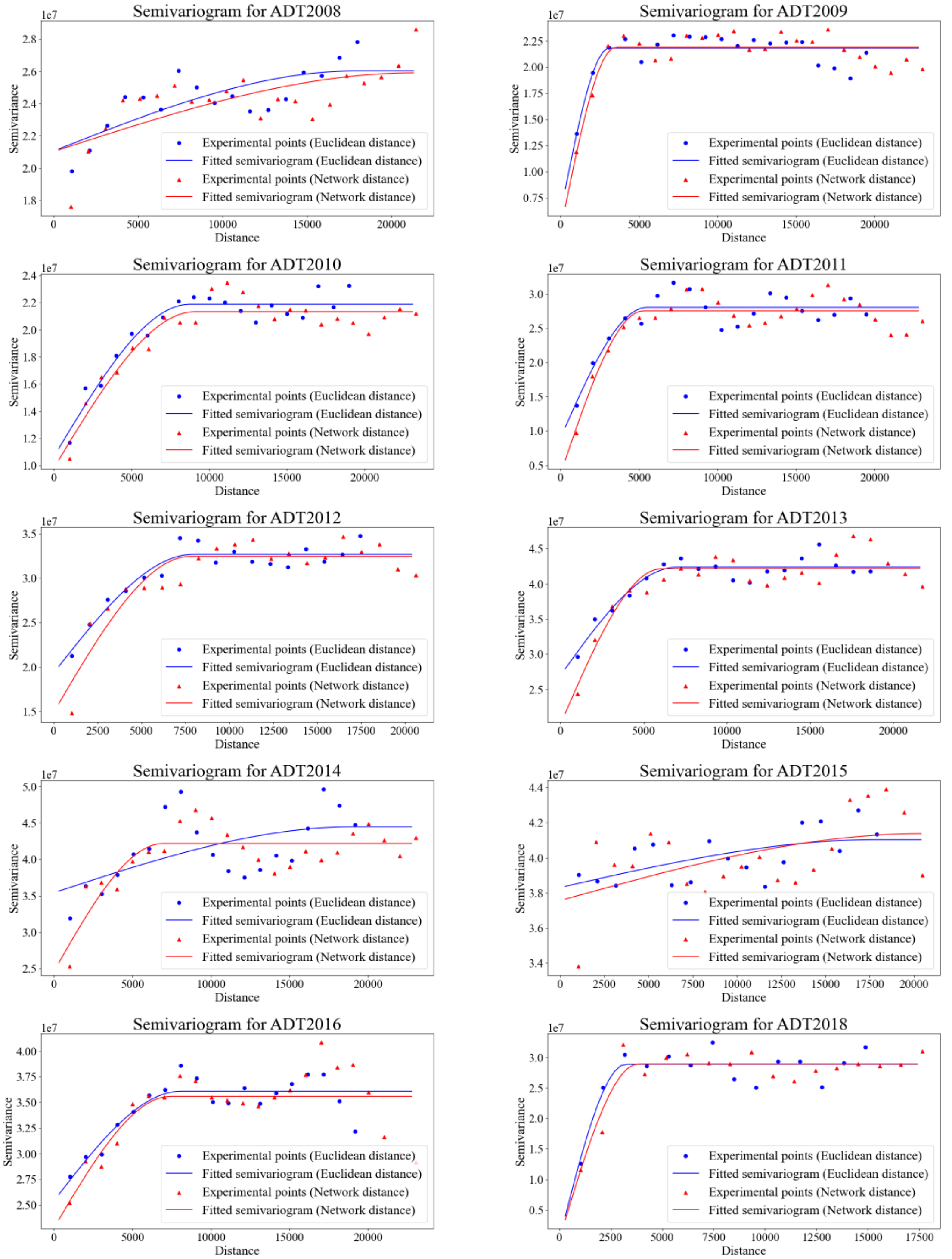
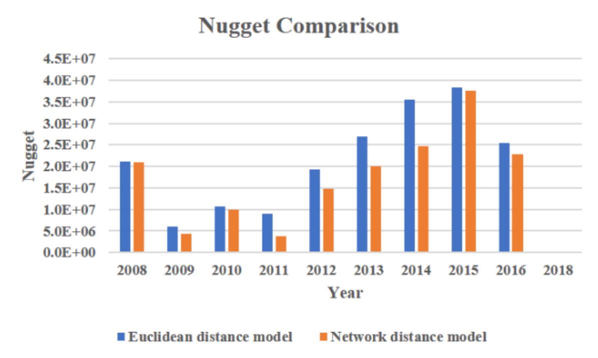
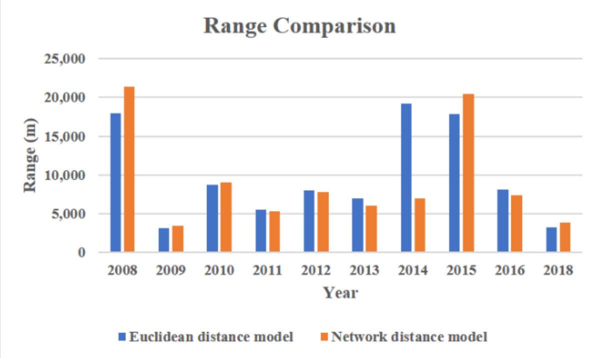
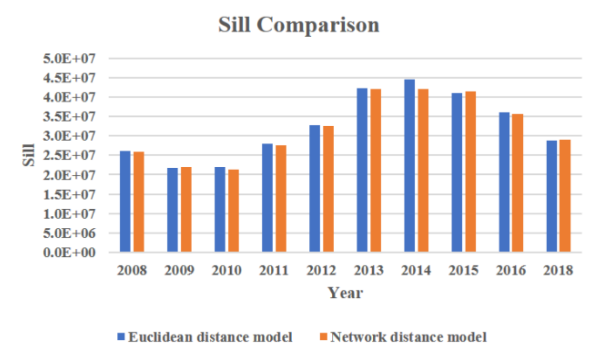


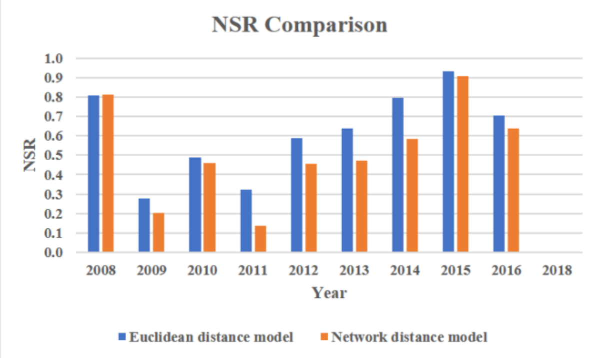
Figure 4. Constructed Semivariogram Models





t-statistic	*p*-value	*t*-statistic	*p*-value
2.991	0.001**	0.508	0.312





t-statistic	*p*-value	*t*-statistic	*p*-value
1.726	0.06	3.518	0.003**

Figure 5. Comparisons of the Semivariogram Parameters

Table 4. Calibrated Semivariogram Parameters for Core vs Non-core Zones

Year	Core Zone (Euclidean)		Core Zone (Network)		Non-core Zone (Euclidean)		Non-core Zone (Network)	
	Nugget	NSR	Nugget	NSR	Nugget	NSR	Nugget	NSR
2008	6.3E+06	3.4E-01	7.3E+06	4.0E-01	3.1E+07	7.6E-01	2.2E+07	5.7E-01
2009	-	-	-	-	1.3E+07	3.7E-01	1.1E+07	2.9E-01
2010	1.2E+07	6.0E-01	1.1E+07	5.5E-01	1.6E+07	4.1E-01	1.6E+07	3.9E-01
2011	2.3E-06	9.1E-14	5.2E+05	2.0E-02	2.5E+07	4.2E-01	1.7E+07	2.8E-01
2012	2.2E+07	9.9E-01	1.9E+07	8.4E-01	6.0E+07	6.5E-01	3.0E-10	3.3E-18
2013	2.1E+07	7.6E-01	1.7E+07	6.2E-01	2.9E+06	2.9E-02	7.9E+06	7.9E-02
2014	2.1E+07	6.1E-01	5.3E+06	1.7E-01	7.4E-09	7.6E-17	1.1E-06	1.1E-14
2015	3.8E+07	9.9E-01	3.0E+07	1.0E+00	8.1E+07	8.0E-01	7.4E+07	7.2E-01
2016	2.2E+07	6.6E-01	2.2E+07	6.6E-01	3.6E+07	5.2E-01	3.5E+07	5.1E-01
2018	-	-	-	-	1.0E+07	1.4E-01	6.0E+06	8.3E-02

3.4.3 Case III: Arterial and Collector

To further examine whether the stationarity assumption is true or not for the study area, this section investigates the differences between semivariogram models tailored for different road types, namely, arterial and collector roads. The model development procedures remain the same as described in the previous sections, and the years with insufficient data for both road types are again excluded from this analysis. All models were fitted using the same spherical form. Their calibrated semivariogram parameters are shown in Table 5.

It is implied that the spatial dependence pattern can be very different between the two road types, as the calibrated parameters differ from each other significantly. For example, the nugget values in arterials are dramatically bigger than those in collectors, this is because arterial roads tend to carry most of the traffic flows of the city especially for intercity trips, and industrial and agricultural travels, thus the magnitude of the variances can be also higher for this road type. In addition, the calibrated semivariogram parameters vary over time regardless of the road type. This implies that the impact of ADT detection sites' configuration and the temporal pattern of travel behavior in the city of Edmonton will require further investigation in future work.

The ultimate goal of this research is to provide more accurate estimates of traffic volumes, cross-validation was conducted to assess the estimation accuracy for the separate estimations for arterial and collector roads. RMSE values were used to compare and contrast the ADT estimates from the various semivariogram models developed using separate road types, keeping all roads together, and network distances. Results show that compared to using Euclidean distance in kriging without separating the road types, network distance kriging while considering semivariograms separately can, once again, improve the ADT estimation accuracy by 4.12% on average for the entire urban network while the *t*-test also validates that it is a statistically significant improvement (*t*-statistic: 5.828; *p*-value: 0.000).

3.5 Spatial Mapping of Traffic Volumes

With the semivariogram models calibrated and validated for each year, the proposed hybrid geostatistical approach was used to interpolate the missing ADTs for the entire city network. This method utilized the midpoints of each road segment that are split by intersections as shown in Figure 6.

As described previously, the major difference in implementing network kriging compared with the standard one is to replace the Euclidean distance with the network distance. Therefore, the network distance matrix based on the shortest path for all measured and unmeasured

Table 5. Calibrated Semivariogram Parameters for Arterial vs Collector

Year	Arterial (Euclidean)		Arterial (Network)		Collector (Euclidean)		Collector (Network)	
	Nugget	NSR	Nugget	NSR	Nugget	NSR	Nugget	NSR
2008	2.4E+07	8.1E-01	1.9E+07	6.7E-01	2.6E+06	6.8E-01	2.3E+06	5.7E-01
2009	3.8E+05	1.3E-02	9.7E-10	3.2E-17	1.8E-13	5.0E-20	6.5E-09	1.9E-15
2010	1.7E+07	5.4E-01	1.4E+07	4.8E-01	1.7E+06	4.3E-01	1.6E+06	4.0E-01
2011	2.6E-02	6.3E-10	6.4E-11	1.6E-18	4.1E+05	6.8E-02	2.8E-09	4.7E-16
2012	2.9E+07	6.6E-01	2.3E+07	5.6E-01	3.0E-07	5.0E-14	4.3E-10	7.1E-17
2013	3.0E+07	6.1E-01	2.4E+07	4.8E-01	-	-	-	-
2014	3.5E+07	6.8E-01	2.6E+07	5.4E-01	1.6E-11	1.0E-18	1.3E+07	1.0E+00
2015	4.3E+07	9.3E-01	4.2E+07	9.0E-01	4.3E-05	2.1E-12	6.8E-13	3.4E-20
2016	2.8E+07	6.8E-01	2.5E+07	6.2E-01	1.1E-09	1.5E-16	3.4E+05	4.6E-02
2018	8.7E-08	2.8E-15	6.7E-07	2.2E-14	-	-	-	-

locations were generated firstly via ArcGIS [32] and then input into kriging along with the corresponding best-fitted spherical semivariogram models. As a result, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015, 2016 utilized the separated-road type network semivariograms while 2013, 2018 utilized their citywide network semivariograms. These interpolated ADT values reflect the true spatial distribution pattern as the ring road and the areas within it is shown to have higher traffic volumes as compared to the areas outside the ring road (i.e., the far northern and far southern parts of the city where people rarely go). Furthermore, years with interpolation using separated semivariogram models (with * marked in title) tend to distinguish the traffic volume differences more prominently between road types.

Although more case studies using different network sizes and traffic related variables (e.g., collision frequency) are required to further attest to the conclusiveness of the results generated herein, the findings of this study provide significant contributions to areas in need of utilizing accurate traffic volume information with limited traffic detectors.

3.6 Summary of Research Findings

With the adoption of the proposed hybrid interpolation framework to estimate and spatially map ADT across the city of Edmonton, several findings were made from the three case studies (i.e., Case I, II, III). These findings are summarized below:

• Network semivariogram was able to better represent the spatial variation pattern of ADT within the road network. When compared with semivariogram models using Euclidean distance (Case I, II, III), network semivariogram has lower nugget and NSR values [37].

• Estimating ADT using network kriging was consistently more accurate than standard kriging with Euclidean distance [21]. This finding was supported by cross-validation results using the 10 years of data (Case I).

• Small and dense networks where roads are fairly straight were not worth performing network kriging due

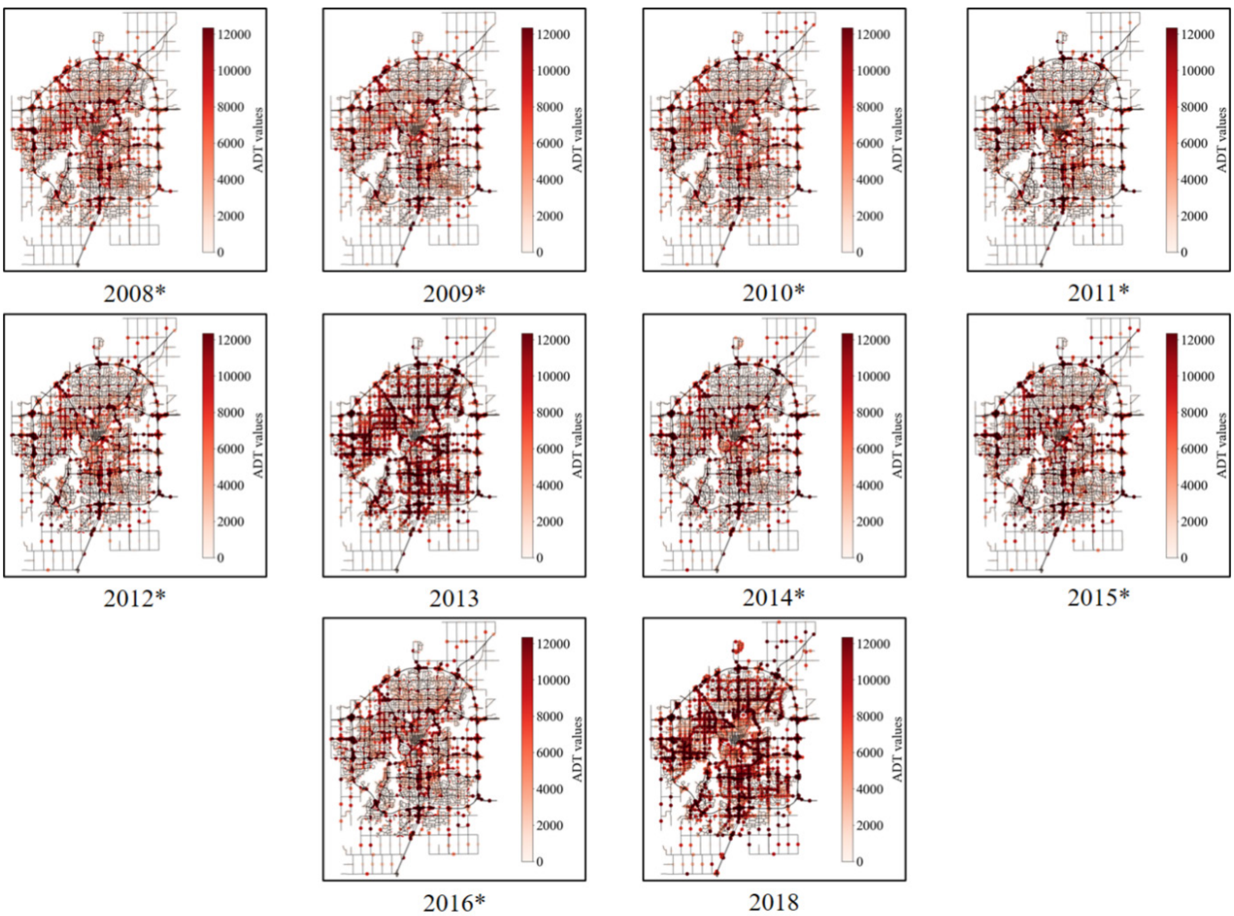


Figure 6. Interpolated ADT maps via Network Regression Kriging

to its low return and substantial computational cost. In Case II, estimated ADT values using two distance metrics have no significant difference in the non-core zone, which supports this finding.

• The imposed stationarity assumption of ADT did not hold true for the entire study area. This was evident from the different semivariogram parameters calibrated with respect to different city zones and road types (Case II, III).

• Estimating ADT using network kriging with multiple road types and semivariograms (Case III) was significantly more accurate than estimating using Euclidean distance with a single semivariogram.

• Semivariogram parameters (e.g., nugget, range, NSR) changed dramatically over time regardless of zone and road types, which implies that the configuration of ADT detection sites has an impact on the spatial variation pattern representativeness, and thus will potentially affect the overall estimation accuracy. The temporal pattern of travel behavior in the city of Edmonton also needs further investigation.

• Interpolated ADT (Figure 6) using our proposed method was able to successfully reflect the true spatial distributions of the traffic volumes in the city.

Since the data employed in this case study covers 10 years of ADT observations, and our findings regarding the network regression kriging performance are consistent with previous relevant studies [11,38,16,7,21], it can be said that our proposed hybrid geostatistical interpolation framework and research findings are transferrable and applicable to other cities as well as other traffic related variables. One thing to note is, some of the detailed parameters or settings (e.g., variables in detrending step, semivariogram type, etc.) may differ depending on the data of interest and availability.

4. Conclusions and Future Research

In this research, a hybrid geostatistical interpolation framework for estimating city-wide traffic volumes was developed by applying linear regression models to remove trends within ADT and performed kriging with network distances using semivariogram models constructed for each estimation year. A case study in Edmonton was conducted to compare the estimation accuracy between the standard and network models. Linear regression models consistently show that road types, speed limits and accessibility to senior residences are significant explanatory variables over the 10 years, and accessibility to schools may also be a significant variable depending on the distribution of sample sites. Overall, the network semivariogram model better represents the spatial interaction pattern of the traffic volumes. It significantly reduces the nugget effect while also forming a stronger spatial dependence by generating a lower NSR. For a large road network (e.g., the whole city network, non-core zone), network kriging consistently and significantly outperforms standard kriging by providing more accurate traffic volume estimates. By contrast, on a small scale and dense road network with road segments that are not very curvy (e.g., core-zone), the standard kriging method based on Euclidean distance is still able to provide reliable estimates, and the semivariogram model is also able to adequately represent the spatial interaction of traffic volumes. Furthermore, it is also found that the stationarity assumption for traffic volumes does not hold true thereby indicating that the semivariogram model constructed over a particular zone/network configuration may not be representative of the actual spatial interaction for all zones or different road types in the city. This also suggests that separate estimates for different road types using their corresponding semivariogram models can produce more accurate estimates overall.

In conclusion, within the area of traffic estimation, our research provides several contributions. First, a hybrid geostatistical interpolation framework was developed, which involves the use of local auxiliary variables and road network metrics. This interpolation framework is a powerful tool that transportation agencies can employ when their traffic information is limited. In addition, this is also the first time that long-term and large urban datasets are used to prove the feasibility and robustness of our proposed framework. And lastly, it was demonstrated that with semivariogram models, using network distances can allow us to better understand the spatial variation patterns of traffic volume. Furthermore, this research also contributes to identifying situations where the application of network kriging is not necessary considering the substantially high computational costs required of it. Furthermore, the imposed stationarity assumption placed on the whole network did not hold true since semivariogram models developed for different zones and road types led to improved estimation performance. It is expected that with more accurate ADT estimation/ mapping results (as shown in Figure 6) and a better understanding of the ADT spatial pattern, transportation authorities can make more robust and accurate decisions on transportation-related activities.

This research can be extended in several directions. First, the same approach proposed in this study can be directly implemented in any other traffic-related cases where the use of network distance is more reasonable. Given the studies done in the past using Euclidean distances, there are lots of studies that can be re-done using network distances such as traffic congestion, collisions, emissions and so on [38-40]. Secondly, the distribution of the sampling sites is suspected to affect the modeling and estimation performance as can be seen from the 10-year RMSE values and calibrated nugget values. Therefore, additional efforts into developing optimal traffic-count sampling strategies will be required to optimize the traffic monitoring capabilities and to further improve estimation performance. It would be also beneficial to use the data collected over same geographic locations but over different years in an attempt to further validate the conclusiveness and transferability of our findings. Thirdly, instead of converting road types into numerical values, they can also be hot-coded or binary-coded (e.g., major roads vs other roads) to quantify the effect of different road types have on ADT and their contributions to improving the accuracy of interpolating citywide traffic volumes. Lastly, the calibrated parameters of our linear regression models (i.e., coefficients of the auxiliary variables) and semivariograms (e.g., range) tend to vary drastically, which implies that people's travel behavior within the city of Edmonton is changing over time. Further investigation of this temporal pattern can be a potential future research topic.

Acknowledgement

The authors would like to thank the City of Edmonton for providing financial support and the data used in this study. The authors would also like to acknowledge the anonymous reviewers for their suggestions that have contributed to improving the quality and the readability of the manuscript.

Conflict of Interest

All authors declare that they have no conflicts of interest.

References

[1] Elvik, R., Høye, A., Vaa, T., et al., 2009. The Contribution of Research to Road Safety Policy-Making. The Handbook of Road Safety Measures, Emerald Group Publishing Limited. pp. 117-141.
DOI: https://doi.org/10.1108/9781848552517-006.

[2] Minderhoud, M.M., Botma, H., Bovy, P.H.L., 1997. Assessment of Roadway Capacity Estimation Methods. Transportation Research Record. 1572(1), 59-67.
DOI: https://doi.org/10.3141/1572-08

[3] Rakowska, A., Wong, K.C., Townsend, T., et al., 2014. Impact of traffic volume and composition on the air quality and pedestrian exposure in urban street canyon. Atmospheric Environment. 98, 260-270.
DOI: https://doi.org/10.1016/j.atmosenv.2014.08.073

[4] Wu, M., Kwon, T.J., El-Basyouny, K., 2020. A Citywide Location-Allocation Framework for Driver Feedback Signs: Optimizing Safety and Coverage of Vulnerable Road Users. Sustainability. 12(24), 10415.
DOI: https://doi.org/10.3390/su122410415

[5] Skszek, S.L., 2001. State-of-the-art report on non-traditional traffic counting methods. Arizona.

[6] Eom, J.K., Park, M.S., Heo, T.Y., et al., 2006. Improving the Prediction of Annual Average Daily Traffic for Nonfreeway Facilities by Applying a Spatial Statistical Method. Transportation Research Record. 1968(1), 20-29.
DOI: https://doi.org/10.1177/0361198106196800103

[7] Xia, Q., Zhao, F., Chen, Z., et al., 1999. Estimation of Annual Average Daily Traffic for Nonstate Roads in a Florida County. Transportation Research Record. 1660(1), 32-40.
DOI: https://doi.org/10.3141/1660-05

[8] Zhao, F., Chung, S., 2001. Contributing factors of annual average daily traffic in a Florida county: exploration with geographic information system and regression models. Transportation Research Record. 1769(1), 113-122.
DOI: https://doi.org/10.3141/1769-14

[9] Mohamad, D., Sinha, K.C., Kuczek, T., et al., 1998. Annual Average Daily Traffic Prediction Model for County Roads. Transportation Research Record. 1617(1), 69-77.
DOI: https://doi.org/10.3141/1617-10

[10] Tang, Y.F., Lam, W.H.K., Ng, P.L.P., et al., 2003. Comparison of four modeling techniques for short-term AADT forecasting in Hong Kong. Journal of Transportation Engineering. 129(3), 271-277.
DOI: https://doi.org/10.1061/(ASCE)0733-947X(2003)129:3(271)

[11] Lam, W.H.K., Tang, Y.F., Chan, K.S., et al., 2006. Short-term Hourly Traffic Forecasts using Hong Kong Annual Traffic Census. Transportation. 33(3), 291-310.
DOI: https://doi.org/10.1007/s11116-005-0327-8

[12] Davis, G., Yang, S., 2001. Accounting for uncertainty in estimates of total traffic volume: an empirical Bayes approach. Journal of Transportation and Statistics. 4(1), 27-38.

[13] Goel, P.K., McCord, M.R., Park, C., 2005. Exploiting Correlations between Link Flows to Improve Estimation of Average Annual Daily Traffic on Coverage Count Segments: Methodology and Numerical Study.

Transportation Research Record. 1917(1), 100-107.
DOI: https://doi.org/10.1177/0361198105191700112

[14] Matheron, G., December 1963. Principles of geostatistics. Economic geology. 58(8), 1246-1266.
DOI: https://doi.org/10.2113/gsecongeo.58.8.1246

[15] Chiles, J., Delfiner, P., 2009. Geostatistics: modeling spatial uncertainty, John Wiley & Sons.

[16] Selby, B., Kockelman, K.M., 2013. Spatial prediction of traffic levels in unmeasured locations: applications of universal kriging and geographically weighted regression. Journal of Transport Geography. 29, 24-32.
DOI: https://doi.org/10.1016/j.jtrangeo.2012.12.009

[17] Kwon, T.J., Fu, L., 2017. Spatiotemporal variability of road weather conditions and optimal RWIS density—an empirical investigation. Canadian Journal of Civil Engineering. 44(9), 691-699.
DOI: https://doi.org/10.1139/cjce-2017-0052

[18] Wu, M., 2020. Evaluating the Safety Effects of Driver Feedback Signs and Citywide Implementation Strategies.
DOI: https://doi.org/10.7939/r3-xwkc-1112

[19] Biswas, S., Wu, M., Melles, S.J., et al., 2019. Use of Topography,Weather Zones, and Semivariogram Parameters to Optimize RoadWeather Information System Station Density across Large Spatial Scales. Transportation Research Record. 2673(12), 301-311.
DOI: https://doi.org/10.1177/0361198119846467

[20] Gu, L., Wu, M., Kwon, T.J., November 2019. An Enhanced Spatial Statistical Method for Continuous Monitoring of Winter Road Surface Conditions. Canadian Journal of Civil Engineering. 47(10), 1154-1165.
DOI: https://doi.org/10.1139/cjce-2019-0296

[21] Zhang, D., Wang, X.C., 2014. Transit ridership estimation with network Kriging: a case study of Second Avenue Subway, NYC. Journal of Transport Geography. 41, 107-115.
DOI: https://doi.org/10.1016/j.jtrangeo.2014.08.021

[22] Gneiting, T., Genton, M.G., Guttorp, P., 2006. Geostatistical space-time models, stationarity, separability, and full symmetry. Monographs On Statistics and Applied Probability. 107, 151.

[23] Roll, J., 2019. Evaluating Streetlight Estimates of Annual Average Daily Traffic in Oregon.

[24] Cressie, N., 1990. The origins of kriging. Mathematical Geology. 22(3), 239-252.
DOI: https://doi.org/10.1007/BF00889887

[25] Lichtenstern, A., 2013. Kriging methods in spatial statistics.

[26] Hengl, T., Heuvelink, G., Stein, A., 2003. Comparison of kriging with external drift and regression kriging. ITC Enschede, Netherlands.

[27] Olea, R.A., 2006. A six-step practical approach to semivariogram modelling. Stochastoc Envrionment Research and Risk Assessment. 20(5), 307-318.
DOI: https://doi.org/10.1007/s00477-005-0026-1

[28] Cambardella, C.A., Moorman, T.B., Novak, J.M., et al., 1994. Field-scale variability of soil properties in central Iowa soils. Soil Science Society of America Journal. 58(5), 1501-1511.
DOI: https://doi.org/10.2136/sssaj1994.03615995005800050033x

[29] Sun, X.L., Yang, Q., Wang, H.L., et al., 2019. Can regression determination, nugget-to-sill ratio and sampling spacing? Catena. 181, 104092.
DOI: https://doi.org/10.1016/j.catena.2019.104092

[30] Bohling, G., 2005. Introduction to geostatistics and variogram analysis. Kansas Geological Survey. pp. 1-20.

[31] Solana-Gutiérrez, J., Merino-de-Miguel, S., 2011. A Variogram Model Comparison for Predicting Forest Changes. Procedia Environmental Sciences. 7, 383-388.
DOI: https://doi.org/10.1016/j.proenv.2011.07.066

[32] Desktop, ESRI ArcGIS, 2011. Release 10. Redlands, CA: Environmental Systems Research Institute. 437, 438.

[33] Van Rossum, G., Drake, F.L., 2009. Python 3 Reference Manual, Scotts Valley, CA: CreateSpace.

[34] Mälicke, M., Schneider, H.D., 2019. Scikit-GStat 0.2.6: A scipy flavoured geostatistical analysis toolbox written in Python. (Version v0.2.6). Zenodo.

[35] Oz, B., Deutsch, C.V., 2000. Cross Validation for Selection of Variogram Model and Kriging Type: Application to IP Data from West Virginia. Center for Computational Geostatistics Annual Report Papers. pp. 1-13.

[36] Brooker, P.I., 1986. A parametric study of robustness of kriging variance as a function of range and relative nugget effect for a spherical semivariogram. Mathematical Geology. 18(5), 477-488.
DOI: https://doi.org/10.1007/BF00897500

[37] Skøien, J.O., Merz, R., Blöschl, G., 2006. Top-kriging - geostatistics on stream networks. Hydrology and Earth System Sciences. 10(2), 277-287.
DOI: https://doi.org/10.5194/hess-10-277-2006

[38] Miura, H., 2010. A study of travel time prediction using universal kriging. 18(1), 257-270.
DOI: https://doi.org/10.1007/s11750-009-0103-6

[39] Thakali, L., Kwon, T.J., Fu, L., 2015. Identification of crash hotspots using kernel density estimation and kriging methods: a comparison. Journal of Modern

Transportation. 23(2), 93-106.
DOI: https://doi.org/10.1007/s40534-015-0068-0
[40] Bayraktar, H., Turalioglu, F.S., 2005. A Kriging-based approach for locating a sampling site-in the assessment of air quality. Stochastic Environmental Research and Risk Assessment. 19(4), 301-305.
DOI: https://doi.org/10.1007/s00477-005-0234-8