

RESEARCH ARTICLE

REVISED **Estimating annual average daily traffic (AADT) data on low-volume roads with the cokriging technique and census/population data [version 2; peer review: 1 approved, 2 approved with reservations]**

Edmund Baffoe-Twum ¹, Eric Asa², Bright Awuku²

¹Department of Construction Management, West Virginia University Institute of Technology, Beckley, WV, 25801, United States

²Civil, Construction and Environmental Engineering, North Dakota State University, Fargo, ND, 58108, United States

V2 **First published:** 22 Apr 2022, 4:20
<https://doi.org/10.35241/emeraldopenres.14632.1>
Latest published: 22 May 2023, 4:20
<https://doi.org/10.35241/emeraldopenres.14632.2>

Abstract

Background: Geostatistics focuses on spatial or spatiotemporal datasets. Geostatistics was initially developed to generate probability distribution predictions of ore grade in the mining industry; however, it has been successfully applied in diverse scientific disciplines. This technique includes univariate, multivariate, and simulations. Kriging geostatistical methods, simple, ordinary, and universal Kriging, are not multivariate models in the usual statistical function.

Notwithstanding, simple, ordinary, and universal kriging techniques utilize random function models that include unlimited random variables while modeling one attribute. The coKriging technique is a multivariate estimation method that simultaneously models two or more attributes defined with the same domains as coregionalization.

Objective: This study investigates the impact of populations on traffic volumes as a variable. The additional variable determines the strength or accuracy obtained when data integration is adopted. In addition, this is to help improve the estimation of annual average daily traffic (AADT).



Methods Procedures, Process: The investigation adopts the coKriging technique with AADT data from 2009 to 2016 from Montana, Minnesota, and Washington as primary attributes and population as a controlling factor (second variable). CK is implemented for this study after reviewing the literature and work completed by comparing it with other geostatistical methods.

Results, Observations, and Conclusions: The Investigation employed two variables. The data integration methods employed in CK yield more reliable models because their strength is drawn from multiple variables. The cross-validation results of the model types explored with the CK technique successfully evaluate the interpolation technique's performance and help select optimal models for each

Open Peer Review

Approval Status ✓ ? ?

	1	2	3
version 2 (revision) 22 May 2023	✓		?
	↑		
version 1 22 Apr 2022	?	?	

1. **Watheq J. Al-Mudhafar** , Basrah Oil Company, Basrah, Iraq
2. **Balázs Varga** , Budapest University of Technology and Economics, Budapest, Hungary
3. **Mike Pereira**, Mines Paris - PSL University, Saint-Michel, France

Any reports and responses or comments on the article can be found at the end of the article.

state. The results from Montana and Minnesota models accurately represent the states' traffic and population density. The Washington model had a few exceptions. However, the secondary attribute helped yield an accurate interpretation. Consequently, the impact of tourism, shopping, recreation centers, and possible transiting patterns throughout the state is worth exploring.

Keywords

AADT, Cokriging, Population Density, Cross-Validation, Geostatistics



This article is included in the [Sustainable Cities](#) gateway.

Corresponding author: Edmund Baffoe-Twum (edmund.baffoetwum@mail.wvu.edu)

Author roles: **Baffoe-Twum E:** Writing – Original Draft Preparation, Writing – Review & Editing; **Asa E:** Supervision; **Awuku B:** Data Curation

Competing interests: No competing interests were disclosed.

Grant information: The author(s) declared that no grants were involved in supporting this work.

Copyright: © 2023 Baffoe-Twum E *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Baffoe-Twum E, Asa E and Awuku B. **Estimating annual average daily traffic (AADT) data on low-volume roads with the cokriging technique and census/population data [version 2; peer review: 1 approved, 2 approved with reservations]** Emerald Open Research 2023, 4:20 <https://doi.org/10.35241/emeraldopenres.14632.2>

First published: 22 Apr 2022, 4:20 <https://doi.org/10.35241/emeraldopenres.14632.1>

REVISED Amendments from Version 1

The new version has been completed based on concerns raised by peer reviewers. Changes and additions were made to sections. Further explanations as sought were included in this new version. I have incorporated all questions asked.

Any further responses from the reviewers can be found at the end of the article

Introduction

Varga *et al.* (2023) assert that a cost-effective elucidation for many Intelligent Transport Solutions (ITS) applications are to estimate traffic flow at not measured locations. However, accurately predicting annual average daily traffic (AADT) volumes by considering the effects of AADT and count location (recorder station) could increase understanding of spatial variability. Also, at the engineering design stage, it is necessary to have accurately predicted AADT volumes at specific locations. Besides, it is necessary to understand variables that impact the AADT values generated at a specific location. These variables include population, land use, recreation, and tourist activities. These factors can lead to high or low traffic volumes on a road segment. Researchers have explored several techniques for some of these factors in conjunction with the AADT data collected. These techniques include linear, logistic, and geographically weighted regression methods (Apronti *et al.*, 2016; Cheng, 1992; Deacon *et al.*, 1987; Lu *et al.*, 2007; Mohamad *et al.*, 1998; Raja *et al.*, 2018; Shon, 1989; Xia *et al.*, 1999; Zhao & Park, 2004). Others are artificial neural networks (Sharma *et al.*, 2001), traditional factor approach (Sharma *et al.*, 2000), smoothly clipped absolute deviation penalty (Yang *et al.*, 2011; Yang *et al.*, 2014), ubiquitous probe vehicle data (Zhang & Chen, 2020), and satellite imagery, geographical information system-based travel demand models, and spatial interpolation and geostatistical Kriging (Eom *et al.*, 2006; Shamo *et al.*, 2015; Yang *et al.*, 2014; Zhong & Hanson, 2009; Zou *et al.*, 2012). In addition, machine-learning techniques have recently been explored in estimating traffic flow patterns ((Das & Tsapakis, 2020; Varga *et al.*, 2023). These techniques differ in weighting, thus varying degrees of accuracy and acceptance based on fulfilling some conditions. However, a few of these techniques are not suitable for every location. Therefore, the authors caution that their techniques should be restricted to the areas studied in such cases (Staats, 2016). Consequently, there is a need for a more universal or generalized method that is able to generate accurate predictions and interpolation surface maps to assist with AADT data collection optimization without restrictions.

Geostatistical techniques are typically universal and have been applied and proven in several fields of study. According to Stein and Corsten (1991), some areas of applicability of geostatistical techniques include soil science, mining, hydrology, meteorology, medicine, agriculture, biology, public health, and environmental sciences (e.g., atmospheric or soil pollution). By

2009, Zhou *et al.* (2007) and Hengl *et al.* (2008) demonstrated that the top application fields of geostatistics were geosciences, water resources, environmental sciences, agriculture and/or soil sciences, mathematics and statistics, ecology, civil engineering, petroleum engineering, and meteorology.

In geostatistics, the most applied technique is the Kriging technique. However, the Kriging geostatistical methods only consider the sample values of a single variable to make predictions. In contrast, coKriging, as an alternative geostatistical method, uses more than a single piece (up to four) of the available information from several variables such as the population, land use, and so forth from the study area. Furthermore, coKriging as a geostatistical technique is a multivariate kriging method. Accordingly, it is a much better approach to predict AADT volumes and simultaneously consider the influence of variables (factors) on the dataset at data collection locations. The assumption is that data integration methods such as cokriging may yield more reliable models because their strength is drawn from multiple variables. In addition, cokriging can be extremely valuable when highly correlated covariables are thoroughly sampled.

The coKriging approach does not smooth variables and thus makes accurate predictions. However, coKriging is a good choice for determining how the various factors contribute to AADT volume changes at a given location. This makes coKriging a helpful decision-making tool.

Cokriging is used in several non-transportation-related analyses and yields more accurate and robust data than the classic or ordinary Kriging techniques (Ahmadi & Sedghamiz, 2008; Amiri *et al.*, 2017; Ersahin, 2003; Laurenceau & Sagaut, 2008; Stein & Corsten, 1991; Tziachris *et al.*, 2017; Zhang & Cai, 2015). However, Eldeiry & Garcia (2009) cautiously state otherwise.

Al-Mudhafar (2019) demonstrates the strength of geostatistics in investigating the feasibility of Bayesian Kriging to generate the most realistic spatial permeability model. Zou *et al.* in 2012 suggested the Kriging methods' usefulness when considering spatial analysis. They conclude that the results obtained by applying the method in traffic data interpolation were promising. In addition, geostatistics can effectively select a spatial resolution for image data and the support size for ground data (Atkinson & Quattrochi, 2000). Varga *et al.* (2023) use extended Kriging techniques to estimate and predict in real-time traffic volume and speed, respectively, at several unsampled locations. The results from their studies proposed that spatio-temporal prediction can accomplish a more significant extent of accurate predictions. Varga *et al.* (2023) again suggest that the deep learning technique results compared well to the Kriging technique.

Bae *et al.* (2018) confirm the need to rely on traffic dataset accuracy in literature documentation of transportation systems. The dependence on the dataset is to ensure operational

traffic conditions are monitored for performance assessment. Consequently, missing or unsampled data may result in ineffectual decision-making if not resolved appropriately. [Bae et al. \(2018\)](#) assert that most traffic data often ignore spatial correlations and consider only temporal continuity. However, they state that some studies have explored only the randomness in the missing data patterns. As a result, [Bae et al. \(2018\)](#) explore spatial and temporal characteristics of the traffic data, adopting two coKriging methods and the classic simple and ordinary kriging methods. These methods are set as standards to allow for accurate comparison. Using multiple data sources where the missing data locations are clustered or in blocks, the spatiotemporal cokriging method can effectively improve imputation accuracy. In contrast, the classic ordinary or simple kriging methods are effective if the primary data source has the missing data randomly scattered in time and location. Therefore, [Bae et al. \(2018\)](#) conclude that a Kriging-based imputation approach generates accurate and reliable predictions.

[Wackernagel \(1994\)](#) compares and confirms the advantages of coKriging over Kriging. [Wackernagel \(1994\)](#) states that between estimating a sum and the separate estimation of each of its terms, cokriging guarantees coherence. Contrariwise, with a set of auxiliary variables (autokrigeability- intrinsically correlated), coKriging and Kriging are similar. The intrinsic correlation suggests that the computed fundamental features are not a coregionalization analysis. Instead, the computed fundamental features are from classical multivariate data analysis.

[Ahmadi and Sedghamiz \(2008\)](#) evaluate groundwater depth across a plain using Kriging and coKriging methods. Their technique accurately evaluates water resource conditions in arid and semi-arid regions. They confirm the spatial relationship between groundwater depth and the prevailing climatic conditions. Based on the calculated root mean square error (RMSE), coKriging outperformed Kriging. Also, Kriging underestimated real groundwater depth for dry, wet, and normal conditions. [Ahmadi and Sedghamiz \(2008\)](#) confirmed that the coKriging estimates were unbiased.

[Laurenceau and Sagaut \(2008\)](#), in varying sampling and modeling techniques, adopt Kriging (Kriging and gradient-enhanced Kriging) and coKriging (direct and indirect). Their model constructs efficient response surfaces of aerodynamic functions. However, the authors note that coKriging did not circumvent the slow linear phase of error convergence with increased sample size. Likewise, [Stein and Corsten \(1991\)](#) use universal Kriging and coKriging as a regression procedure to confirm that Kriging is the optimum technique among all linear procedures when comparing spatial interpolation and prediction techniques. In addition, [Stein and Corsten \(1991\)](#) specify that Kriging techniques are unbiased, and the prediction error variance is nominal.

Nevertheless, coKriging has properties similar to Kriging and is more precise in its predictions. The coKriging technique

uses one or more covariable(s) in the processes. [Stein and Corsten \(1991\)](#) emphasize that there is only a slight difference between Kriging and coKriging. Yet, they confirmed that cokriging is most valuable when highly correlated covariables are thoroughly sampled. [Eldeiry and Garcia \(2009\)](#) estimated soil salinity with the best band combinations in a two-fold evaluation. They compared Kriging and coKriging regression techniques. The authors use these techniques with LANDSAT images to create accurate soil salinity maps. The regression Kriging technique outperformed the coKriging technique because the regression Kriging technique included most of the insignificant discrepancies in soil salinity.

On the other hand, [Zhang and Cai \(2015\)](#) determine when Kriging outperforms coKriging. Accordingly, they state that the outperformance occurs due to the nonexistence of theoretical results for coKriging. Furthermore, they point out that conceptually, the prediction variance of coKriging should be smaller than or equal to kriging. However, in some circumstances, it occasionally outperforms Kriging.

[Tziachris et al. \(2017\)](#) use different interpolation techniques to estimate soil iron (Fe) content at unsampled locations. They assess and compare the procedure using spatial autocorrelation and semivariograms to present the best technique. The methods are ordinary Kriging, Universal Kriging, and coKriging. The results show evidence for yearly spatial cross-correlation of soil Fe and pH. The results indicate improving the interpolated results' accuracy for the unsampled locations. Furthermore, [Tziachris et al. \(2017\)](#) confirm that coKriging takes advantage of the covariance between the two regionalized variables (pH and Fe) and achieves better yearly results than the other interpolation techniques.

Some researchers have introduced other types of coKriging to assist in the needs of research analyses. A typical example is multivariate universal cokriging (MUCK), introduced by [Clark et al. \(1989\)](#). With this example, the authors use MUCK to estimate dataset variables without having similar locations, as in the traditional multivariate coKriging technique. However, the authors quickly caution that the MUCK estimates did not vary from the traditional coKriging. Therefore, there are no restrictions on the model output and estimation processes.

Similarly, [Myers \(1991\)](#) uses data on correlated variables in coKriging to improve primary variable estimation (but this does not improve the estimation process for all variables). Cokriging makes it possible to use data collected regarding an auxiliary variable to determine data for under-sampled areas with insufficient data. From the cross-correlation structures of variables, sampled information is predicted using coKriging techniques. However, coKriging techniques inadequately represent the estimate's complexities, especially when the bivariate has non-linear and complex variables. Furthermore, caution is required since coKriging is a linear geostatistical algorithm; it has a smoothing effect on the estimated block model ([Myers, 1991](#)). Thus, it either overestimates or underestimates the original distribution of the variables

(Madani, 2019). Madani (2019) proposes a combined factor-based methodology called projection pursuit multivariate transform and coKriging to overcome shortfalls, such as the complexity of variables and the smoothing effect in traditional coKriging algorithms. The process preserves the complexity and improves the smoothing effect (Madani, 2019).

In another development, Amiri *et al.* (2017) prove that coKriging for spatial interpolation accurately predicts fish abundance. The complete model contained chlorophyll-a content to understand the ecological and anthropogenic drivers for fish population dynamics. Their research objective is to use ordinary kriging and cokriging geostatistical methods to predict the spatial density and distribution of kilka species in the southern Caspian Sea. In addition, the study determines whether the distribution of kilka relates to satellite-derived sea surface temperature, chlorophyll-a concentration, turbidity, and water depths.

Smith *et al.* (2020) show the successful application of Poisson coKriging (bivariate structure model) in predicting a Poisson outcome for the pollen counts variable when an auxiliary variable such as temperature or precipitation data is adopted. Poisson cokriging is a multiple-variable technique that assumes a covariance matrix similar to simple cokriging. Results from Poisson coKriging are minor average errors for about 95% coverage. Chen *et al.* (2018) propose an error compensation method to improve the aviation drilling robot's positioning accuracy. They verify the error compensation method's correctness and effectiveness using coKriging. A precision laser tracker is used to check the measurements. The results based on this technique result in a reduced average absolute positional error of 0.7168 mm to 0.1150mm. In addition, the maximum average absolute positional error decreases from 1.3073 mm to 0.2664 mm. Thus, Chen *et al.* (2018) confirm that the technique helps improve aviation robots' absolute position accuracy and may help meet aircraft assembly requirements.

In finding a solution to ecologists' challenges in mapping vegetation quantities over a large area, Dungan *et al.* (1994) explore point-based interpolation, such as cokriging and conditional simulations. The authors confirm that the information covering the entire area is generated with the adopted methods.

Meng *et al.* (2009) discuss the new systematic geostatistical techniques for predicting forest parameters (basal area, height, health conditions, biomass, or carbon as a response variable) or inventory. The combined methods consist of Landsat 7 enhanced thematic mapper plus (ETM+) images, a global positioning system (GPS), and geographic information systems (GIS). The GIS techniques used were univariate kriging (ordinary and universal Kriging) and multivariable kriging (cokriging and regression Kriging). Meng *et al.* (2009) confirm that geostatistical approaches can better predict parameter values for unmeasured locations. Cokriging and regression Kriging combined with the normalized difference vegetation index (NDVI) and principal components (PCs) are used to validate

200 random sampling points. From the results, the kriging techniques performed better. Regression Kriging is the best geostatistical method for spatial predictions. Furthermore, the regression Kriging results have the least errors and the highest r-squared (Meng *et al.*, 2009).

Doyen *et al.* (1996) utilized the simplified collocated coKriging technique based on a Bayesian in an interpolation where seismic impedance was a second variable to an associated primary variable, porosity. The model predicted and generated the lateral porosity variations in a reservoir layer of the Ekofisk Field, Norwegian North Sea.

This study investigates, demonstrates, and validates population distribution as a controlling factor in AADT data. Therefore, this study implements cokriging using known AADT data and the various county population data from Montana, Minnesota, and Washington as a second variable. The estimated AADT volumes were simulated using population data.

Various locations in the three selected states were compared to demonstrate the effectiveness of coKriging in spatial distribution. In addition, it illustrates the importance of considering other variables instead of a single variable to predict sampled and unsampled locations accurately. Also, the study determines the accuracy of predicted values of unsampled locations. Finally, the study is to demonstrate that the more variables there are for a predictive model, the better the model output.

Methodology

This section discusses the procedures (Figure 1) in completing the explored models to determine the applicability of the geostatistical technique to estimating AADT data.

Data description and processing

The data used for this study is acquired from the Departments of Transportation in Minnesota, Montana, and Washington states. The classification of the datasets is based on the Cornell Local Roads Program (CLRP). The categories limit AADT data to a minimum of 1 and a maximum of 400, 500, 1000, and 2000 vehicles per day. The downloaded datasets were in the Microsoft Excel spreadsheet format. Each spreadsheet contained several years of annual average daily traffic datasets. The annual average daily traffic from 2009 to 2016 is extracted from each state's data set. After extracting the needed data, the datasets are subjected to conditional clauses in Excel to generate annual average daily traffic values of less or equal to 400, 500, 1000, and 2000 vehicles per day. Each dataset generated is subjected to further screening to remove data collection stations with zero counts and missing data. Finally, the data were explored using exploratory spatial data analysis (ESDA) to visualize outliers easily, delineate global trends in the data, locate areas with high and low values, and possible transformation. The population data were drawn primarily from the U.S. Census Bureau dataset. The population is used as an AADT data control factor because it is universal in all aspects of the analysis.

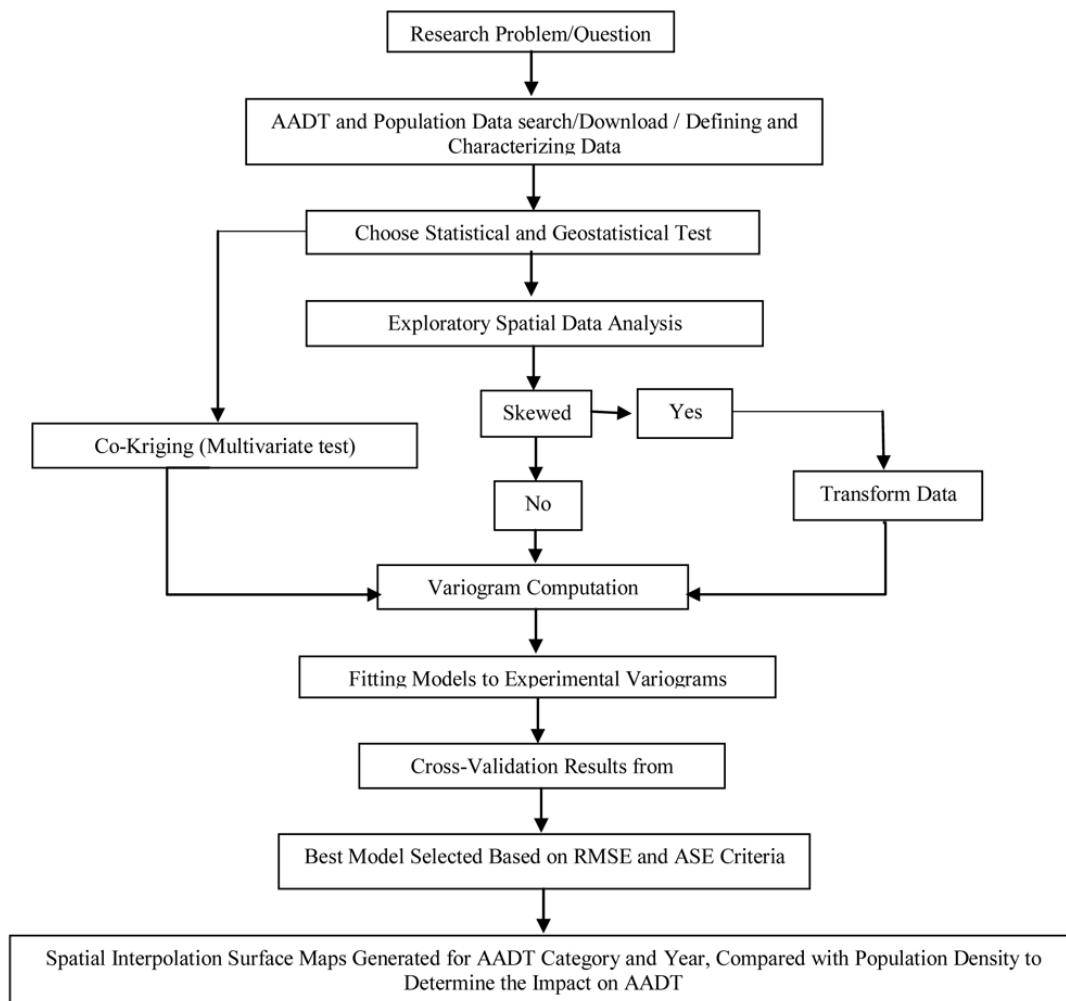


Figure 1. Processes involved in developing and interpreting the cokriging method (annual average daily traffic - AADT).

Geostatistical simple Kriging assumes that the global mean of a dataset is constant (Veeken, 2007). Kriging generates spatial interpolation and extrapolations of the control points through a multi-varied statistical approach. Therefore, Kriging prediction utilizes weighting functions dependent on the probability of distribution and spatial variations of the dataset. As a result, the process ensures that the error variance is minimized in relation to the predicted values of the least square (Veeken, 2007). Simple Kriging is similar to ordinary Kriging; however, the weighted sum equation, which equals 1, is not added in simple Kriging. Also, in simple Kriging, the mean is a known constant. Therefore, the entire dataset's average is used. In contrast, ordinary kriging uses the local averages, which correspond to the average of the subset points for points specified in the interpolation.

Sunila (2015) outlined the following as a step-by-step process for Kriging:

1. Studying the gathered data: data analysis

2. Fitting variogram models: experimental variogram and theoretical variogram models
3. Estimating values at locations that have not been sampled, e.g., ordinary Kriging, simple Kriging, indicator Kriging, etc.
4. Examining standard errors which may be used to quantify confidence levels
5. Kriging interpolation

Korn (2013), in the Handbook of Geomathematics, presents “various classical geostatistical prediction methods with a focus on interpolation methods known as Kriging. The main types of Kriging interpolation methods, such as simple, ordinary, and universal Kriging, are derived as the best linear predictors in the mean squared sense. Also discussed are the multivariate and non-linear generalizations such as cokriging or indicator Kriging and their application.”

In all geostatistical methods, only the sample values of a single variable are used to generate estimates. Cokriging uses information on several variables. It uses at least two, and a maximum of four, variables to refine the predicted values (Queiroz *et al.*, 2008). This assumes an appropriate autocorrelation exists between or among the variables explored. Cokriging is like Kriging because of similarities in premise, but interpolated surface accuracy is far better than Kriging. Multivariable use ensures accuracy and eliminates biases between actual and estimated values (Salith *et al.*, 2002). Similarly, variance among these estimates is minimized (Salith *et al.*, 2002).

Cokriging uses autocorrelation and cross-correlation algorithms to produce interpolated surfaces that envisage values at unmeasured places. According to Salith *et al.* (2002), improved accuracy for predicting the primary variable (for example, AADT) in the cokriging model is obtained if there is a stronger autocorrelation among the multivariable. Applying coKriging in this research allowed for determining factors that may be assumed as controlling AADT values at sampled locations. Thus, AADT and population data from the sampled locations are modeled to determine the population effect on the number of vehicles per day counted at each sampled location. Cokriging assumes a linear combination of primary and secondary data values with the equation given as Equation 1.

$$\hat{u}_0 = \sum_{i=1}^n a_i \cdot u_i + \sum_{j=1}^m b_j \cdot v_j \tag{Equation 1}$$

\hat{u}_0 is the estimate of U at location 0; u_1, \dots, u_n and v_1, \dots, v_m represents the primary and secondary data at nearby locations n and m, respectively. a_1, \dots, a_n and b_1, \dots, b_m represents the cokriging weights needed to be determined. However, it is worth noting that the development coKriging is like ordinary kriging, and the determined estimation error equation is as in Equation 2:

$$R = \hat{U}_0 - U_0 = \sum_{i=1}^n a_i \cdot U_i + \sum_{i=1}^m b_i \cdot V_i - U_0 \tag{Equation 2}$$

For phenomenon U with nearby locations n and V nearby locations m, U_0, \dots, U_n and V_0, \dots, V_m represent their random variables, respectively. Equation 2 may be written in matrix notation as Equation 3 and Equation 4 (Variance).

$$R = w^t Z \tag{Equation 3}$$

$$\text{Var}\{R\} = w^t C_z w \tag{Equation 4}$$

Where $w^t = (a_1, \dots, a_n, b_1, \dots, b_m, -1)$, $Z^t = (U_1, \dots, U_n, V_1, \dots, V_m, U_0)$, C_z is the covariance of the matrix Z.

The system used in coKriging can be written in semivariograms because the cross-variance matrix is symmetric, as in Equation 5.

$$\text{Cov}\{U_i, V_j\} = \text{Cov}\{V_j, U_i\} \tag{Equation 5}$$

Although the cross-variance is modeled as a symmetric function, it may be nonsymmetric. The spatial continuity is demonstrated with semivariograms converted to covariance values for the cokriging matrix with a relationship as in Equation 6.

$$C_{uv}(h) = \gamma_{UV}(\infty) - \gamma_{UV}(h) \tag{Equation 6}$$

Solutions based on the coKriging equations could be unique and exist when the auto and cross-variograms are positively fixed. Nevertheless, there are certain conditions where the coKriging model may not improve on the ordinary Kriging estimates. It occurs when the auto and cross-variograms are related to the uncomplicated model. The coKriging model may not improve the ordinary Kriging estimates when primary and secondary variables exist at all data locations. Cokriging will not provide a better model when the variogram models are relatively analogous in shape. The primary variable does not indicate conspicuous under-sampling (Goovaerts, 1997; Isaaks & Srivastava, 1989).

The correlogram (correlation statistics chart), covariance (degree at which random variables vary similarly), and semivariogram (variogram) describe the observation of spatial and temporal correlation in geostatistics. According to Deutsch and Kumara (2017), in geostatistical analysis, establishing a reliable variogram to represent each regionalized variable is a necessary step. The variogram is used in geostatistics to fit the model's spatial and temporal correlation. The theoretical variogram in spatial statistics is a function that describes the extent of reliance on a spatial random field or stochastic process. First, the variogram measures the variation between two samples taken at a location with dependence on the distance between the two. Samples nearer to each other vary less compared to samples taken far apart. Matheron, in 1963 defined a semivariogram as half the average squared difference between data values at two points separated at a distance. Since sampling all locations to obtain data values is impractical, the empirical variogram is used. Thus, it is the variance between data values at the two sampled locations (Cressie, 1993). The empirical variogram is used in geostatistics first to estimate the variogram model needed for spatial interpolation by kriging. The variogram is twice the semivariogram. The spatial autocorrelation of the measured sample points is depicted with a semivariogram. A model is fit through when each pair of locations is plotted. The specific characteristics commonly used to describe these models are range, sill, and nugget. A critical look at the semivariogram models shows that the model levels out at a certain distance. The range of the model is the distance where the model flattens out first. Therefore, samples from locations with distances separating them less than the range are considered spatially autocorrelated.

In contrast, sample locations further apart than the range are not spatially autocorrelated. The value on the y-axis, the value the semivariogram model realizes, is the sill. The partial is attained by subtracting the nugget from the sill. The nugget

is the semivariogram jump height at the origin's discontinuity (ESRI Web Support, 2023). The nugget effect may be assigned to error measurement or spatial sources of variation at distances smaller than the sampling interval. For example, innate errors in the measuring devices may result in errors in measurements obtained. However, it is worth noting that trends vary spatially over innumerable scales. Any scale of variation even smaller than the sampling distances may be introduced in the nugget effect. Consequently, understanding the scales of variation spatially is essential before data collection is implemented (ESRI Web Support, 2023).

For this study, the processes used are the following:

1. The coKriging model is completed using ArcGIS Pro., SGeMS, and QGIS are open software programs that can be used to achieve similar results. The input data were AADT data and the studied area's corresponding population data for each year studied. At the start of the model, data are input into the model. The input dataset consists of the AADT data for the year under review for all completed models. The source dataset relates to the AADT year and the number of vehicles per day. The data field is the recorded volume of vehicles at each location. Input data 2 uses the population data in the source dataset, and the data field is the population year. The next button is used to access the next stage of the model.
2. After several trials, the conditions are selected and applied to all the completed models. For dataset 1, the normal score transformation is adopted as the transformation type; true is selected for the input decluster before transformation and first for trend removal. The transformation type is set to log for the second dataset and the trend removal order as first.
3. The default for the trend's general properties and the declustering method are accepted.
4. Similarly, the default, the normal score transformation, is adopted.
5. The general properties remain the default settings at the semivariogram/covariance modeling stage; however, the Var1 and Var2 are changed to semivariogram. In addition, the default for the model nugget is accepted, whereas model 1 is varied during the processing stage to select any of the following model types; stable, circular, spherical, exponential, or Gaussian.
6. The default is accepted for the neighborhood search, except the sector type is changed to four sectors and 45 degrees.
7. The final process generates the summary statistics or cross-validation to assess the best models.

Cross-validation

The cross-validation process was completed with the full complement (100%) of the dataset attained using the *n-1* method

with an output produced from the entire data statistics and not individual points. The cross-validation output is a computerized result generated from the geostatistical optimized processes (ESRI Web Support, 2018b). The process systematically eliminates each point in the data and predicts a missing value at the surface and compares predicted and actual values (ESRI Web Support, 2018a). It produces an accurate system of measurement used to determine the accuracy and efficiency of the model. The cross-validation output is the best-evaluated accurate semivariogram and best fits the model. The measure of the best fit is based on the mean error (M.E.), the root-mean-square error (RMSE), the mean standardized error (MSE), the root-mean-square standardized error (RMSSE), and the average standard error (ASE). The closer the ME and MSE are to zero, the better the model – precisely, the more accurately the model predicts. Likewise, supposing the ASE value is comparable to RMSE, the better the result.

Consequently, a small value of ASE is preferred. Among a series of models, the model with the least difference between ASE and RMSE values is the best predictive model. The RMSSE assumes the mean of the standard error divided by the RMSE. An RMSSE approximating one (1) makes the prediction accurate and reliable. A large RMSSE is an indication of an unstable model. An RMSSE greater than one (1) indicates a model underestimating the variability of the dataset. The mathematical expression for the performance measures for the goodness-of-fit is presented in Equation 7 – Equation 10.

$$\text{Mean Error (M.E. or MAE)} = \frac{1}{n} \sum_{i=1}^n [z^*(x_i) - z(x_i)] \quad \text{Equation 7}$$

$$\text{Mean Standardized Error (MSE)} = \frac{1}{n} \sum_{i=1}^n \left[\frac{z^*(x_i) - z(x_i)}{\sigma^2(x_i)} \right] \quad \text{Equation 8}$$

$\sigma^2(x_i)$ represents the Kriging variance of the location x_i

$$\text{Average Standard Error (ASE)} = \sqrt{\frac{1}{n} \sum_{i=1}^n \sigma^2(x_i)} \quad \text{Equation 9}$$

$$\text{Root Mean Square Standardized Error (RMSSE)} = \sqrt{\frac{1}{n} \sum_{i=1}^n \left[\frac{z^*(x_i) - z(x_i)}{\sigma^2(x_i)} \right]^2} \quad \text{Equation 10}$$

Where x_i represents location, $Z^*(x_i)$ and $Z(x_i)$ are predicted, and observed parameter values at the location x_i , σ is the standard deviation, and n is the total number of observations.

Results

Table 1 summarizes the results obtained for the best model output for Montana state. The Table consists of summaries for AADT up to 400, 500, 1000, and 2000 for each year from 2009 to 2016. Model types found in the Table are Gaussian, spherical, and stable. The Table shows that the Gaussian model outperformed the stable model types. For the eight years studied, using AADT data of up to 400 vehicles per day, all but one turned out to have Gaussian as the best model. The 2011 model shows that the spherical model is the best for AADT up to 400. With AADT data up to 500 vehicles per day, the eight years studied have indicated Gaussian as the best

Table 1. Cross-validation summary for each year based on the AADTs explored, Montana (root mean square error - RMS, mean square error - MS, root mean square standardized error - RMSS, average standardized estimation error - ASE).

AADT /Year / Best Model Type	Count	Mean	RMS	MS	RMSS	ASE	ASE-RMS
400/2009 /Gaussian	1110	-3.27	106.20	-0.02	0.98	108.71	2.51
400/2010 /Gaussian	1083	-3.76	107.20	-0.03	0.99	108.71	1.51
400/2011 /Spherical	1101	-6.00	106.77	-0.05	0.99	108.14	1.37
400/2012 /Gaussian	1091	-4.76	107.79	-0.04	0.99	109.21	1.42
400/2013 /Gaussian	1107	-3.77	107.29	-0.03	0.99	108.38	1.09
400/2014 /Gaussian	1150	-4.33	106.90	-0.03	0.98	108.83	1.92
400/2015 /Gaussian	1193	-6.13	109.21	-0.05	0.98	110.90	1.68
400/2016 /Gaussian	1802	-0.78	102.86	0.00	0.96	106.29	3.43
500/2009 /Gaussian	1284	-7.75	135.15	-0.05	0.99	137.69	2.54
500/2010 /Gaussian	1261	-7.71	136.73	-0.05	0.99	139.14	2.41
500/2011 /Gaussian	1261	-7.75	132.40	-0.05	0.97	136.53	4.14
500/2012 /Gaussian	1272	-6.89	133.45	-0.04	0.99	135.62	2.17
500/2013 /Gaussian	1286	-8.85	134.47	-0.06	0.98	137.33	2.86
500/2014 /Gaussian	1313	-5.68	130.51	-0.04	0.98	133.72	3.20
500/2015 /Gaussian	1360	-8.15	133.05	-0.05	0.98	135.51	2.47
500/2016 /Gaussian	2013	-1.96	130.50	0.00	0.95	137.24	6.73
1000/2009 /Gaussian	1877	-21.16	256.94	-0.07	0.97	266.72	9.78
1000/2010 /Gaussian	1829	-19.14	256.10	-0.06	0.97	265.52	9.41
1000/2011 /Stable	1867	-9.82	250.98	-0.03	0.97	260.15	9.17
1000/2012 /Stable	1862	-9.89	249.66	-0.03	0.98	258.28	8.61
1000/2013 /Stable	1880	-7.67	248.13	-0.02	0.98	255.07	6.94
1000/2014 /Gaussian	1917	-21.88	257.04	-0.07	0.97	267.48	10.44
1000/2015 /Stable	1958	-9.00	251.10	-0.03	0.96	262.57	11.47
1000/2016 /Stable	2731	-0.54	236.48	0.02	0.87	268.18	31.70
2000/2009 /Stable	2616	-21.93	497.53	-0.03	0.95	538.72	41.19
2000/2010 /Stable	2580	-25.41	505.75	-0.04	0.95	547.50	41.75
2000/2011 /Stable	2630	-27.26	503.33	-0.04	0.95	537.74	34.41
2000/2012 /Gaussian	2634	-54.42	519.70	-0.09	0.96	546.37	26.68
2000/2013 /Stable	2626	-21.85	502.14	-0.03	0.96	535.68	33.55
2000/2014 /Gaussian	2703	-32.89	512.88	-0.06	0.97	548.13	35.25
2000/2015 /Stable	2722	-24.18	508.07	-0.04	0.98	535.84	27.76
2000/2016 /Stable	3583	3.22	469.88	0.03	0.89	526.79	56.91

model. The best models for AADT up to 1000 and 2000 were Gaussian and stable. However, the stable model dominates in the prediction accuracy. Data collection locations for Montana state remained about the same, with slight variations for the entire years reviewed.

The appraisal of the optimum models generated mainly depended on the cross-validation output. Therefore, the evaluation is based on the root means square standardized error, mean standard error, root mean square error, and average standardized estimation error. Thus, the root means square standardized error approximates 1; the mean standard error approximates zero; the difference between the root mean square error and the average standardized estimation error reaching zero authenticates confidence to the predicted model; and a small value of the average standardized estimation error.

In Figure 2 and Figure 3, the red dots represent each county's population superimposed on the surface interpolation map generated from the cokriging model. The bar chart graphs are generated using the population data of each county in the state. The blue dots are the same red dots highlighted to align with the county population on the map. There is enough evidence from these two Figures for Montana to conclude that the traffic pattern is related to population data. The highly populated areas coincide with a high traffic pattern. Thus, it indicates that population density impacts traffic volumes. However, one may be cautious about the conclusions as it may not be the only factor impacting traffic patterns in Montana. As a result, decision-makers and policymakers may generate plans and trends from these outputs for future work. For example, conclusions can be drawn on the predictions

for future occurrences regarding traffic safety and so forth. Similarly, optimization of data collection locations may also be appropriately determined and completed.

Table 2 shows the yearly best-predicted models for the state of Montana. All of the yearly best models resulted from AADT data of up to 400 vehicles per day and are associated with the Gaussian model. Out of the eight years reviewed, the spherical model successfully predicts a single year; the rest were Gaussian models. The traffic count varies from a low of 1,091 to a high of 1,802. The root mean square errors were similar and the lowest compared to the other models in Table 1. The root mean square errors ranged between 102.86 and 109.21; thus, a maximum difference of 6.35 is recorded. Likewise, the average standardized estimation error is similar. The average standardized estimation error values ranged between 106.29 and 110.90, with a maximum difference of 4.61. The closeness of the root mean square and average standardized estimation errors indicated that the models' prediction was accurate. Therefore, the result is the difference between root mean square error and average standardized estimation error. All of the root mean square standardized errors approximate one. The mean standard error approximates zero for the models (see Table 2). The mean standard errors and root mean square standardized errors further demonstrated the models' accuracy. The optimum model is the Gaussian model for AADT up to 400 vehicles per day for 2013. The least optimum model is the 2016 AADT up to 400, also a Gaussian model.

Both models had a root mean square standardized error of approximately one and a mean standardized error of approximately zero. The difference between the root mean square

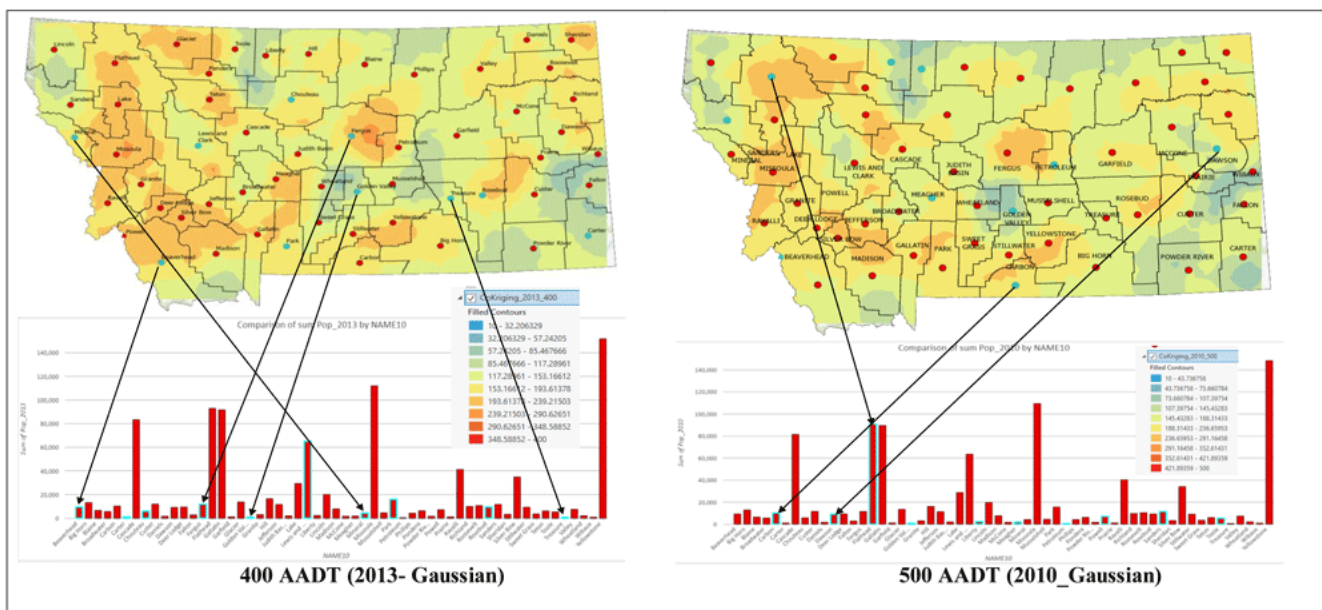


Figure 2. Generated surface maps of annual average daily traffic (AADT) data and countywise population graph for Montana. (The left side corresponds to the best predictive model for AADT up to 400 and the right for AADT up to 500, corresponding to 2013 and 2010, respectively).

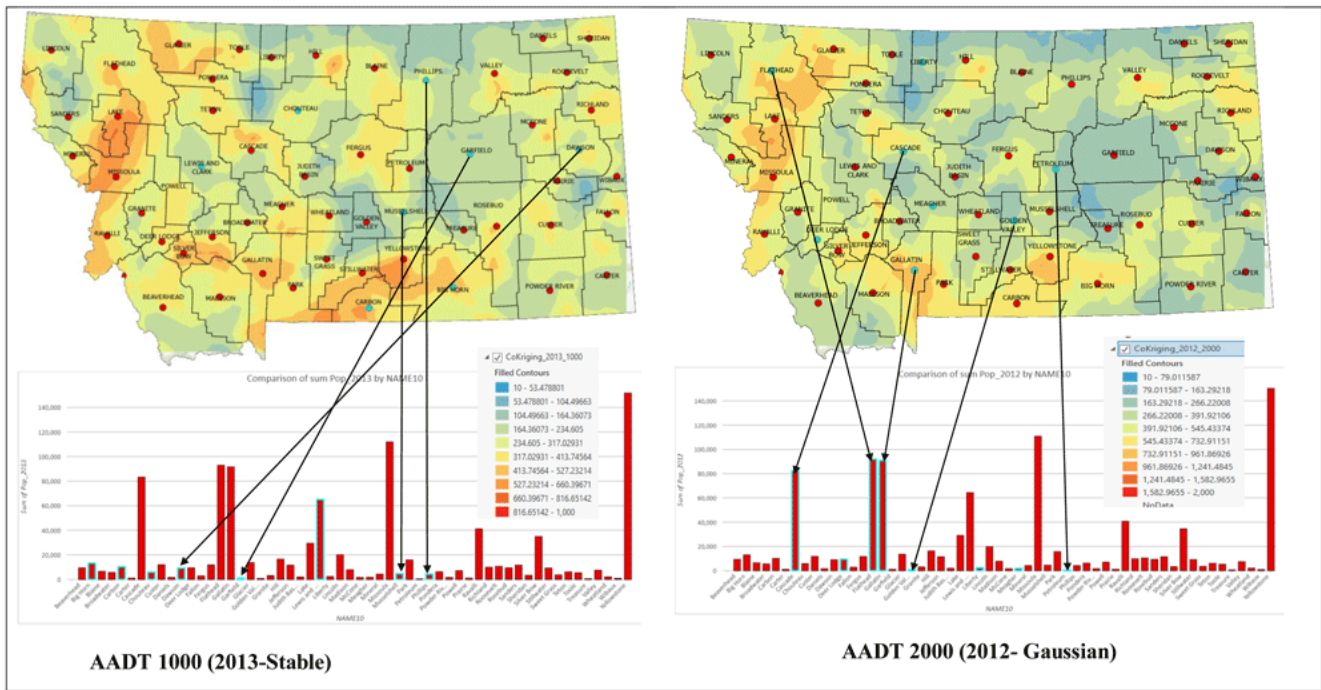


Figure 3. Generated surface maps of annual average daily traffic (AADT) data and countywise population graph for Montana. (The left side corresponds to the best predictive model for AADT up to 1000 and the right for AADT up to 2000, corresponding to 2013 and 2012, respectively).

Table 2. Montana yearly best-predicted models based on cross-validation (root mean square error - RMS, mean square error - MS, root mean square standardized error - RMSS, average standardized estimation error - ASE).

Year / AADT / Best Model Type	Count	Mean	RMS	MS	RMSS	ASE	ASE-RMS
2009 /400/Gaussian	1110	-3.27	106.20	-0.02	0.98	108.71	2.51
2010 /400/Gaussian	1083	-3.76	107.20	-0.03	0.99	108.71	1.51
2011 /400/Spherical	1101	-6.00	106.77	-0.05	0.99	108.14	1.37
2012 /400/Gaussian	1091	-4.76	107.79	-0.04	0.99	109.21	1.42
2013 /400/Gaussian	1107	-3.77	107.29	-0.03	0.99	108.38	1.09
2014 /400/Gaussian	1150	-4.33	106.90	-0.03	0.98	108.83	1.92
2015 /400/Gaussian	1193	-6.13	109.21	-0.05	0.98	110.90	1.68
2016 /400/Gaussian	1802	-0.78	102.86	0.00	0.96	106.29	3.43

and average standardized estimation errors for the optimum and the least optimum models is 1.09 and 3.43. All other models are between the differences. The graphs for the optimum predictors are in Figure 4a and Figure 4b. Figure 4a shows the difference between the root mean square errors and the average standardized estimation errors for the various years reviewed. In Figure 4b, the red lines are the predicted values,

while the blue line represents the actual data. The predicted and actual data differences are shown in the shapes of the various graphs.

Again, similar to the procedure adopted for Montana, the best models are obtained for Minnesota and Washington states from cross-validation analyses of AADT data processed.

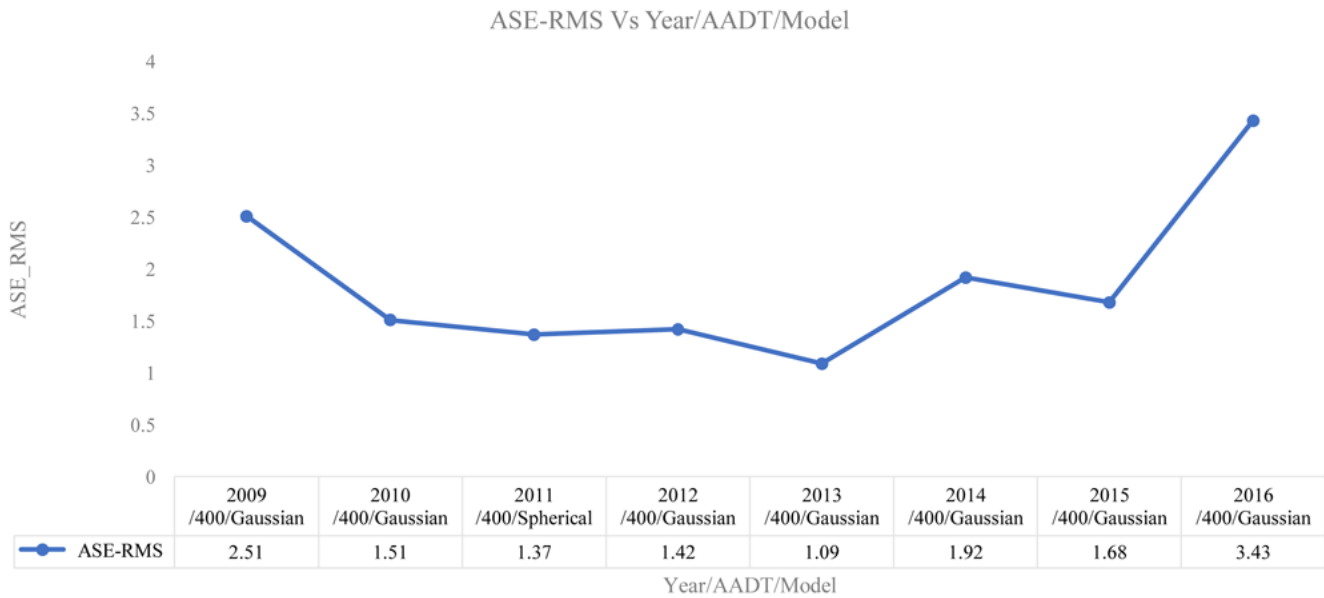


Figure 4a. A graph of the difference between root mean square error and average standardized estimation error (root mean square error - RMS, average standardized estimation error - ASE).

Based on the earlier assumptions from the cross-validation analyses, the summaries of AADT up to 400, 500, 1000, and 2000 for each year from 2009 to 2016 are generated. The model types were circular, exponential, Gaussian, spherical, Stable, and Gaussian-stable. The Gaussian-stable model shows that the Gaussian and Stable models result are similar. Therefore, any of the two models can be used for analysis. No dominant model type is said to have outperformed the other model types. However, for Washington, the stable model showed some consistency.

For the years reviewed, Minnesota state had varying locations used for each year’s data collection. Therefore, the prevailing conditions and spatial patterns or distribution impacted explored models. However, the robustness of the geostatistical technique makes it possible to have accurate models for each year. Meanwhile, for the years reviewed, Washington State maintained similar locations for data collection. Even though a critical examination of the entire dataset showed additional locations for some years, there is not much variation in locations. Data collection may have primarily been based on proximity as well as easy access to locations because of the landforms or geomorphic features. Consequently, the impact on the models explored, and the outputs generated are due to the prevailing conditions and the spatial patterns or distribution of the data. The geostatistical technique makes it possible to obtain accurate models to mimic reality.

For the study of eight years of data from Minnesota, the Gaussian model dominates the AADT data of up to 400 vehicles per day. It is also the best-fit model for the years assessed. The 2011, 2012, 2013, and 2015 models are all Gaussian,

whereas 2009 and 2010 had the spherical model as the best. The 2014 and 2016 are associated with exponential and stable models, respectively. A mix of models is selected as the best for AADT up to 500, 1000, and 2000 vehicles per day for the period studied. A few models overestimated, and others underestimated the variability of the dataset. The output for AADT up to 400 for 2012, AADT up to 500 for 2011, AADT up to 1000 for 2011, and AADT up to 2000 for 2009 and 2012 models underestimates the data variability. On the other hand, models for AADT up to 400 for 2009 and 2013, AADT up to 500 for 2009 and 2013, and AADT up to 2000 for 2013, 2014, and 2015 overestimate the variability.

In **Figure 5**, the red dots represent each county’s population superimposed on the interpolated map generated. The bar charts are graphs generated from each county’s population data. The blue dots are county populations, and the red is highlighted for comparison. The model output shows a good correlation between the traffic patterns and the population in the figure. Therefore, there is enough evidence to conclude that high-traffic areas in Minnesota are associated with population. Highly populated counties coincided with high traffic patterns. However, the population may not be the only factor significantly impacting traffic patterns. Hitherto decisions and policymakers may depend on the output, generate plans for future work, predict future traffic safety occurrences, etc. Optimizing data collection locations for better coverage may be based on and completed appropriately from this output.

Table 4 shows the yearly best-predicted models for the state of Minnesota. The yearly best models result from AADT of up to 400, 500, and 2000 vehicles daily. All the model types are

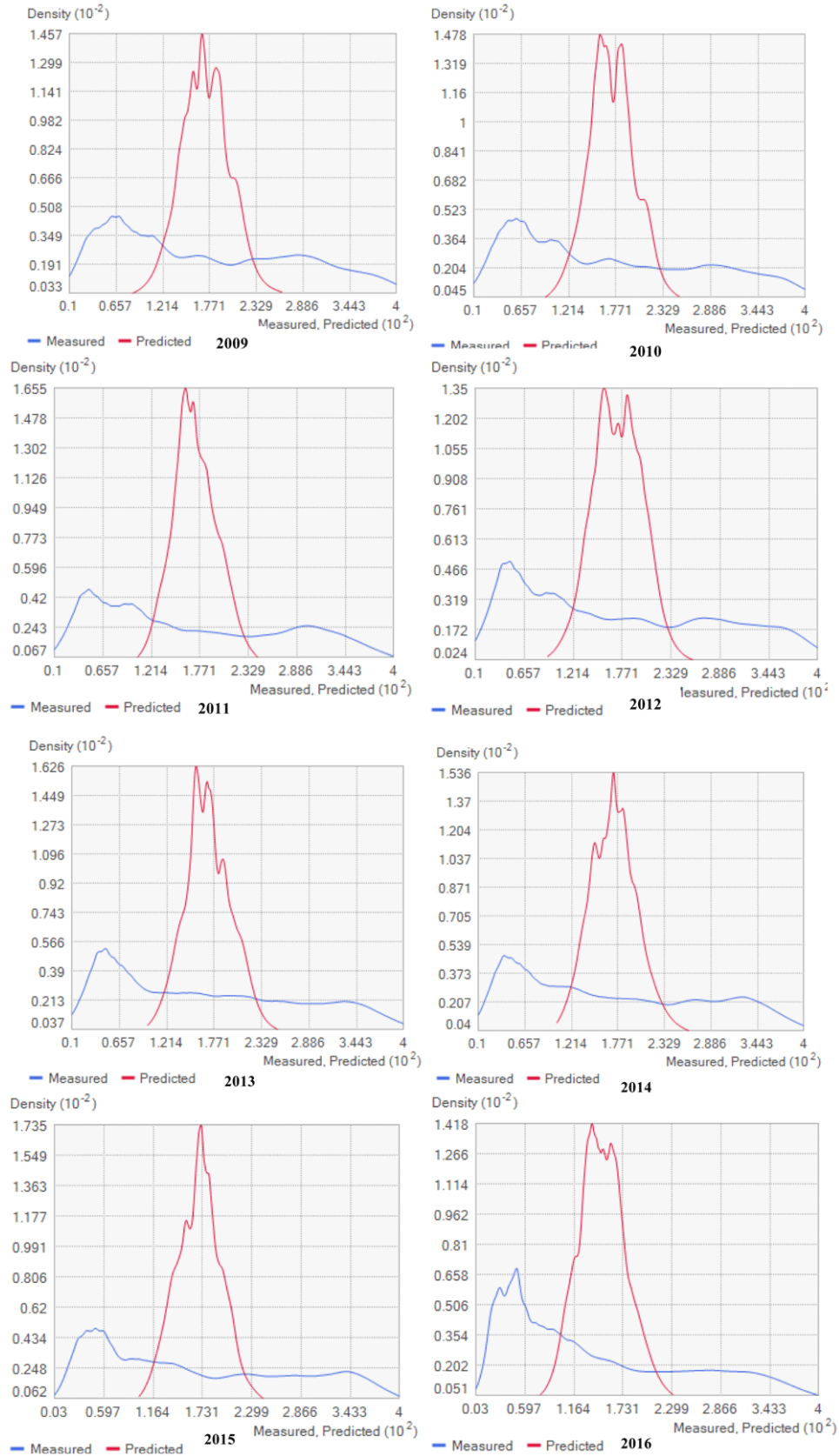


Figure 4b. Measured (blue) and predicted (red) distribution graphs.

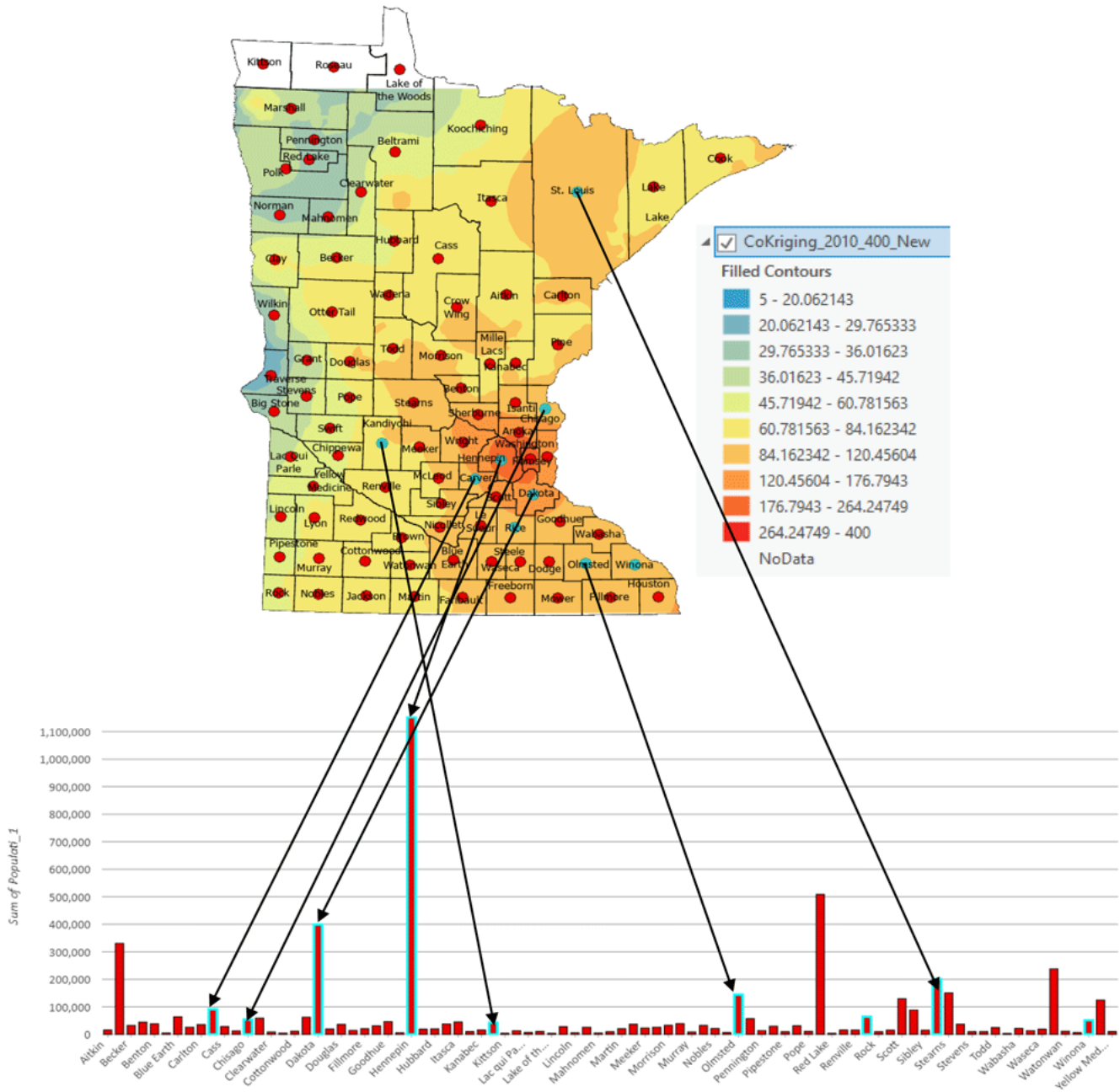


Figure 5. Annual average daily traffic (AADT) 400 (Spherical model, 2010).

represented. The results indicate that the data collection sites varied for each year explored. Only one year is best predicted with the spherical model from the eight years reviewed, the optimal model. The data counts varied from 527 to a high of 2056.

The closeness of the root means square error, and the average standardized estimation error confirms the model predictions' accuracy. However, two models overestimate AADT data values, whereas one underestimates data variability. The overestimated model corresponds with the best models for

2009 and 2013, while the underestimated corresponds to 2012. The overestimation is from AADT up to 500 and 400 from exponential and Gaussian models. The underestimation corresponds to AADT up to 2000 and the stable model. The least of the differences between the root mean square error and the average standardized estimation error and root mean square standardized error approximating one are used to determining the best predictive model.

The root mean square standardized errors and mean standard errors approximate one and zero for the models. The mean

standard error and the root mean square standardized error further demonstrate the accurateness of the models. The optimum model is the spherical model with AADT up to 400 for 2010. The least optimum model is 2013 at an AADT of up to 400 and associated with the Gaussian model. Both models have a root mean square standardized error approximating one and a mean standardized error of approximately zero. The differences between the root mean square error and the average standardized estimation error were 0.61 and 12.45, respectively. The other errors were between the stated differences. The lowest average standardized estimation errors are consistent with the optimal model.

From Washington state, for the eight years studied utilizing AADT up to 400 vehicles per day, the exponential model best fits the years assessed. The best output for 2009, 2010, and 2011 are associated with the exponential model. Whereas with 2012 and 2014, the Gaussian is the best model for AADT

up to 400. The years 2015 and 2016 are associated with the stable model. The 2013 best model for the AADT up to 400 is produced using the circular model.

For AADT of up to 2000 vehicles per day, the eight years studied generated the stable model as the best-fit model. While for AADT of up to 500 and 1000 vehicles per day, the output for five of eight years modeled are stable models, and two years are associated with Gaussian. Furthermore, 2013 for AADT up to 500 and 2010 for AADT 1000 vehicles per day are associated with spherical and circular models, respectively. The models are represented in Table 3 with their respective root mean square standardized error, mean standard error, root mean square error, and average standardized estimation error. The root means square standardized error at approximately one; the mean standardized error is approximately zero, and the comparison of root mean square error and average standardized estimation error. The output for AADT up to 400 vehicles per

Table 3. Minnesota yearly best-predicted models based on cross-validation (root mean square error - RMS, mean square error - MS, root mean square standardized error -RMSS, average standardized estimation error - ASE).

Year / AADT / Best Model Type	Count	Mean	RMS	MS	RMSS	ASE	ASE-RMS
2009 /500/Exponential	527	-0.20	71.51	0.02	0.93	74.50	3.00
2010 /400/Spherical	961	-4.00	74.62	-0.03	1.03	74.01	0.61
2011 /400/Gaussian-Stable	1320	-0.35	84.85	-0.01	1.06	84.11	0.74
2012 /2000/Stable	686	6.97	356.55	-0.03	1.20	358.36	1.81
2013 /400/Gaussian	723	7.98	98.27	0.07	0.89	110.72	12.45
2014 /400/Exponential	1551	0.87	104.40	0.01	0.99	105.86	1.45
2015 /400/Gaussian-Stable	1555	4.02	107.26	0.04	0.99	108.45	1.19
2016 /500/Circular	2056	-1.93	126.77	-0.01	0.98	128.95	2.18

Table 4. Washington’s yearly best-predicted models based on cross-validation (root mean square error - RMS, mean square error - MS, root mean square standardized error - RMSS, average standardized estimation error - ASE).

Year / AADT / Best Model Type	Count	Mean	RMS	MS	RMSS	ASE	ASE-RMS
2009 /500/Gaussian	294	-5.44	122.66	-0.04	0.99	124.13	1.47
2010 /400/Exponential	315	9.53	107.11	0.09	0.98	108.70	1.59
2011 /400/Exponential	317	9.68	104.93	0.09	0.99	105.75	0.82
2012 /400/Gaussian	322	7.16	106.22	0.07	1.00	106.59	0.37
2013 /400/Circular	324	3.14	101.16	0.03	1.00	101.23	0.07
2014 /400/Gaussian	316	-0.62	102.53	-0.01	0.99	103.55	1.01
2015 /500/Stable	353	0.97	120.84	0.01	0.99	121.54	0.69
2016 /500/Stable	334	2.07	124.06	0.02	1.00	124.63	0.57

day shows no overestimation or underestimation variability in the dataset. On the other hand, the AADT of up to 1000 vehicles per day for 2010 and 2011 is appraised to overestimate the dataset’s variability. Apart from 2009 for AADT up to 2000 vehicles per day, all models explored under AADT up to 2000 vehicles per day overestimate the data variability.

There were some correlations between the population and the traffic pattern. In such areas, the conclusion is that high traffic in Washington state relates to population density. Highly populated counties coincided with high traffic patterns as per output. However, in locations like the northwestern part of the state, the population density did not correlate well with the interpolation surface maps’ traffic intensity. The population data inversely correlated with traffic patterns. This observation may be attributed to the number of vehicles transiting to Canada or other parts of the United States. In addition, the sitting of recreation and sightseeing or tourism sites may have contributed to the highs in traffic volume when the data collection is completed.

Therefore, the county’s population may not be the only factor impacting traffic patterns. Nevertheless, the interpolation surface maps generated can be used as a baseline for determining the factors impacting traffic patterns. In addition, as a result of the findings, the decision and policymakers can develop plans for future work. For example, data collection points optimization can be selected from the interpolation surface maps produced to cover better all the sites needed for decision making.

The Washington State’s yearly best predictive models are shown in Table 4. All of the annual best models resulted from

AADT up to 400 and 500 vehicles per day. Four out of the five evaluated models are represented in Table 4. Three of the eight years reviewed were associated with the Gaussian model. Two years were each associated with the stable and exponential models, while the circular model is associated with one year. The circular model turns out to be the optimum model. The count varies from a low of 294 to a high of 353. The closeness of root means square error and average standardized estimation error revealed the models’ accuracy. None of the models overestimated or underestimated data variability.

The resulting differences between the root mean square error and the average standardized estimation error are used to determine the best predictive model and the optimum model for the years reviewed. The root means square standardized errors and mean standard errors approximating one and zero for each model are considered. The mean standard and root mean square standardized errors further reveal the models’ accuracy. The optimum model is the circular model for AADT of up to 400 vehicles per day for 2013. In contrast, the least optimum model is associated with the year 2010 at an AADT of up to 400 vehicles per day and is the exponential model. The optimum and least optimum models have a root mean square standardized error of one and 0.98, respectively, and a mean standardized error of approximately zero. The difference between the root mean square and average standardized estimation errors was 0.07 and 1.59, respectively. All of the other models are between the stated differences. The lowest average standardized estimation error is consistent with the optimal model. Figure 6a and Figure 6b show the plots of the optimum predictors with respect to the cross-validation analysis. The red lines in Figure 6b are predicted, whereas the

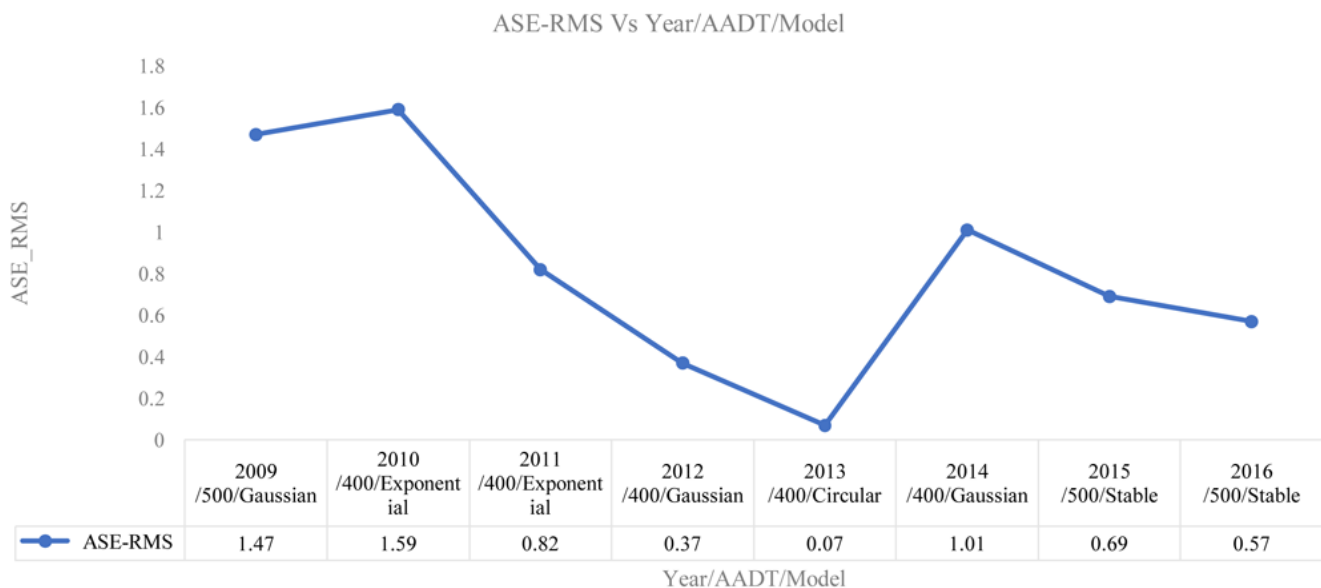


Figure 6a. A graph of the difference between the root mean square error and the average standardized estimation error (root mean square error - RMS, average standardized estimation error - ASE).

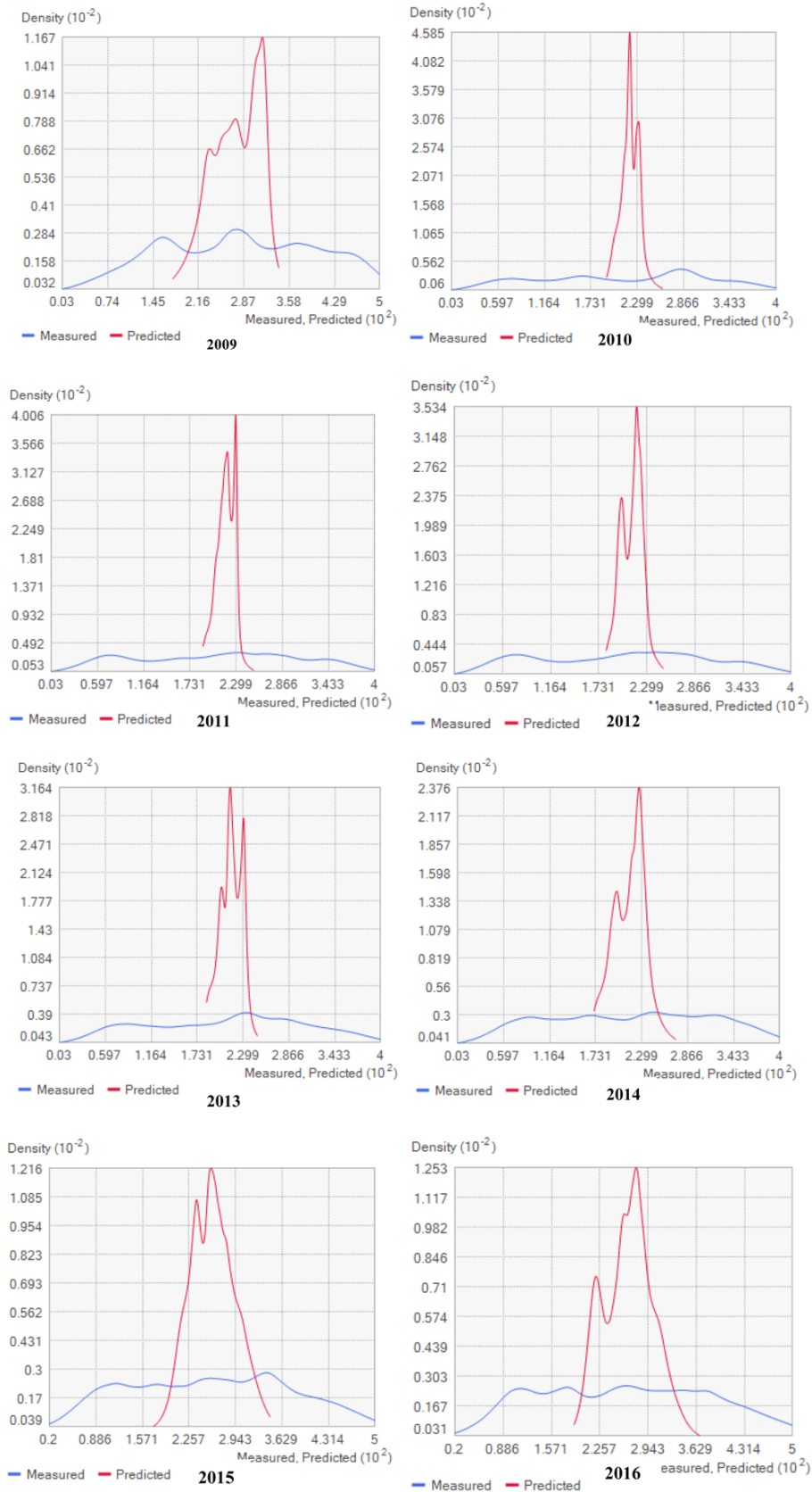


Figure 6b. Measured (blue) and predicted (red) distribution graphs.

blue lines correspond to the actual measured data. The difference is depicted in the shapes of the various graphs.

Conclusion

The presence of spatial variability in data is studied using geostatistical modeling. As a result, this study adopts the cokriging multivariate approach to determine the relationship between countywide population and the traffic density experienced on roadways classified as low volume or local roads in these three states: Montana, Minnesota, and Washington. The data used are the AADT datasets and the population data from the three states from 2009 to 2016. The geostatistical modeling technique cokriging, which has been successfully explored in other scientific studies, is used in this research. In addition, spatial interpolation surface maps of the AADT datasets are generated.

The resulting cross-validation of the various models explored under the cokriging tool is evaluated to determine the different models' performance and select the best-fit model. The evaluation is completed using the least differences between root mean square error and average standardized estimation error; the root mean square standardized error is approximately one, and the mean standardized error is approximately zero. The optimum of the best-fit models is also generated using the same process. The explored models are circular, exponential, gaussian, spherical, and stable. The optimal models are selected based on the same criteria. The models adequately predict the various AADT datasets based on these assumptions, which provides confidence to the model outputs.

The Montana interpolation models correlated well with population density per the analysis and conventions. Similarly, Minnesota's results show a good correlation between traffic patterns and population density. However, unlike Montana and Minnesota, the Washington models generating the interpolation surface maps for the traffic patterns did not directly relate to population density at every location or county. For some reason, some areas of the interpolation are inversely correlated. Nevertheless, the population densities were in this study considered a universal factor that impacts traffic patterns. In addition, other factors, such as tourism, mountains

(topographic features), recreation, and shopping center locations, may affect the density of traffic distribution on certain roads or in some sections of the state of Washington.

Also, travelers transiting through Washington state to reach neighboring states and Canada, who are not necessarily residents of the state, may have created high-volume predictions in parts of the state, especially to the northwest. The model analysis shows that model predictions are reliable enough to be used by state transportation engineers and administrators to make meaningful decisions for current and future developments. Also, the administrators may rely on these models to reduce data collection and planning costs. The models are verifiable at similar prevailing conditions; therefore, additional studies utilizing cokriging and other factors besides population may help tune and compare conclusions.

Data availability statement

All data, models, or code that support the findings of this study are publicly available from the websites of the transportation departments of the states of

Minnesota (<http://www.dot.state.mn.us/traffic/data/data-products.html#volume>),

Montana (<https://www.mdt.mt.gov/publications/datastats/traffic-maps.aspx>),

and Washington (<https://gisdata.wsdot.opendata.arcgis.com/search?q=Annual%20Average%20Daily%20Traffic>).

Population data are available at Population Clock ([census.gov](https://www.census.gov)).

Author contribution

Edmund Baffoe-Twum - Writing Original Draft, Reviewing, and Editing

Eric Asa - Supervision

Bright Awuku - Data Curation

References

Ahmadi SH, Sedghamiz A: **Application and evaluation of kriging and cokriging methods on groundwater depth mapping.** *Environ Monit Assess.* 2008; **138**(1–3): 357–368.

[PubMed Abstract](#) | [Publisher Full Text](#)

Al-Mudhafar WJ: **Bayesian kriging for reproducing reservoir heterogeneity in a tidal depositional environment of a sandstone formation.** *J Appl Geophys.* 2019; **160**: 84–102.

[Publisher Full Text](#)

Amiri K, Shabanipour N, Eagderi S: **Using kriging and co-kriging to predict distributional areas of Kilkia species (*Clupeonella* spp.) in the southern Caspian Sea.** *Int J Aquat Biol.* 2017; **5**(2): 108–113.

[Publisher Full Text](#)

Apronti D, Ksaibati K, Gerow K, *et al.*: **Estimating traffic volume on Wyoming low-volume roads using linear and logistic regression methods.** *J Traffic Transp (English Edition).* 2016; **3**(6): 493–506.

[Publisher Full Text](#)

- Atkinson PM, Quattrochi DA: **Geostatistics and geospatial techniques in remote sensing**. 2000.
[Reference Source](#)
- Bae B, Kim H, Lim H, et al.: **Missing data imputation for traffic flow speed using spatiotemporal cokriging**. *Transp Res Part C*. 2018; **88**: 124–139.
[Publisher Full Text](#)
- Chen D, Yuan P, Wang T, et al.: **A Compensation Method for Enhancing Aviation Drilling Robot Accuracy Based on Co-Kriging**. *International Journal of Precision Engineering and Manufacturing*. 2018; **19**(8): 1133–1142.
[Publisher Full Text](#)
- Cheng C: **Optimum sampling for traffic volume estimation**. Ph.D. Dissertation. University of Minnesota, Minneapolis, 1992.
- Clark I, Basinger KL, Harper WV: **MUCK - a novel approach to Co-kriging Geostatistical, Sensitivity, and Uncertainty Methods For Groundwater Flow And Radionuclide Transport Modeling**. *Proc. DOE/AECL conference*, San Francisco, 1989; 473–493.
[Reference Source](#)
- Cressie NAC: **Statistics for spatial data**. Wiley-Interscience Publication, 1993.
[Publisher Full Text](#)
- Das S, Tsapakis I: **Interpretable machine learning approach in estimating traffic volume on low-volume roadways**. *Int J Transp Sci Technol*. 2020; **9**(1): 76–88.
[Publisher Full Text](#)
- Deacon J, Pigman J, Mohenzadeh A: **Traffic volume estimates and growth trends**. Report UKTRP-8732: Kentucky Transportation Research Program, University of Kentucky, Lexington, 1987.
[Reference Source](#)
- Deutsch CV, Kumara P: **Transforming a Variogram of Normal Scores to Original Units**. In: J. L. Deutsch (Ed.), *Geostatistics Lessons*. 2017.
[Reference Source](#)
- Doyen PM, den Boer LD, Pillet WR: **Seismic Porosity Mapping in the Ekofisk Field Using a New Form of Collocated Cokriging**. Paper presented at the SPE Annual Technical Conference and Exhibition, Denver, Colorado, 1996.
[Publisher Full Text](#)
- Dungan JL, Peterson DL, Curran PJ: **Alternative Approaches for Mapping Vegetation Quantities Using Ground and Image Data**. *Environmental Information Management and Analysis: Ecosystem to Global Scales*. 1994.
[Reference Source](#)
- Eldeiry A, Garcia LA: **Comparison of Regression Kriging and Cokriging Techniques to Estimate Soil Salinity Using Landsat Images**. *Hydrology day*. 2009.
[Reference Source](#)
- Eom J, Park M, Heo T, et al.: **Improving the prediction of annual average daily traffic for nonfreeway facilities by applying a spatial statistical method**. *J Transp Res Rec*. Transportation Research Board of the National Academies, Washington, D.C.: 2006; **1968**(1): 20–29.
[Publisher Full Text](#)
- Ersahin S: **Comparing Ordinary Kriging and Cokriging to Estimate Infiltration Rate**. *Soil Sci Soc Am J*. 2003; **67**(6): 1848–1855.
[Publisher Full Text](#)
- ESRI Web Support: **Comparing models**. 2018a.
[Reference Source](#)
- ESRI Web Support: **Performing cross-validation and validation**. 2018b.
[Reference Source](#)
- ESRI Web Support: **ArcGIS Pro 3.1: Understanding a semivariogram: The range, sill, and nugget—ArcGIS Pro | Documentation**. 2023.
[Reference Source](#)
- Goovaerts P: **Geostatistics for Natural Resources Evaluation**. Oxford University Press, 1997; 483.
[Reference Source](#)
- Hengl T, van Loon EE, Shamoun-Baranes J, et al.: **Geostatistical Analysis of GPS Trajectory Data: Space-Time Densities**. *Proceedings of the 8th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences*, Shanghai, 2008; 17–24.
[Reference Source](#)
- Isaaks EH, Srivastava RM: **An Introduction to Applied Geostatistics**. Oxford University Press, 1989; 561.
[Reference Source](#)
- Korn R: **An Introduction to Prediction Methods in Geostatistics**. In: Freeden W., Nashed M., Sonar T. (eds) *Handbook of Geomathematics*. Springer, Berlin, Heidelberg/Ralf Korn. 2013; 1–19.
[Publisher Full Text](#)
- Laurenceau J, Sagaut P: **Building Efficient Response Surfaces of Aerodynamic Functions with Kriging and CoKriging**. *AIAA Journal*. 2008; **46**(2).
[Publisher Full Text](#)
- Lu J, Pan T, Liu P: **Assignment of estimated average annual daily traffic volumes on all roads in Florida**. Final Report. Florida Department of Transportation, Tallahassee, 2007.
- Madani N: **Application of projection pursuit multivariate transform to alleviate the smoothing effect in cokriging approach for spatial estimation of cross-correlated variables**. *Bollettino di Geofisica Teorica ed Applicata*. 2019; **60**(4): 583–598.
[Publisher Full Text](#)
- Matheron G: **Principles of geostatistics**. *Econ Geol*. 1963; **58**(8): 1246–1266.
[Publisher Full Text](#)
- Meng Q, Cieszewski C, Madden M: **Large area forest inventory using Landsat ETM+: A geostatistical approach**. *ISPRS Journal of Photogrammetry and Remote Sensing*. 2009; **64**(1): 27–36.
[Publisher Full Text](#)
- Mohamad D, Sinha K, Kuczek T, et al.: **Annual Average Daily Traffic prediction model for County Roads**. *Transp Res Rec*. TRB, National Research Council, Washington, D.C.: 1998; **1617**(1): 69–77.
[Publisher Full Text](#)
- Myers DE: **Pseudo-Cross Variograms, Positive-Definiteness, And Cokriging**. *Mathematical Geology*. 1991; **23**(6): 805–816.
[Reference Source](#)
- Queiroz JC, Sturaro JR, Saraiva AC, et al.: **Geochemical Characterization of Heavy metal Contaminated Area Using Multivariate Factor Kriging**. *Environmental Geology*. 2008; **55**(1): 95–105.
[Publisher Full Text](#)
- Raja P, Doustmohammadi M, Anderson M: **Estimation of Average Daily Traffic on low-volume roads in Alabama**. *Transportation Research Board 97th Annual Meeting*. Washington DC, United States. 2018.
- Salith I, Pettersson H, Sivertun A, et al.: **Spatial Correlation Between Randon (222 Rn) in Groundwater and Bedrock Uranium (238U) and Geostatistical analyse**. *Journal of Spatial Hydrology*. 2002; **2**(2): 1–10.
[Reference Source](#)
- Sharma S, Lingras P, Liu G, et al.: **Estimation of Annual Average Daily Traffic on low-volume roads**. *Transportation Research Record: Journal of the Transportation Research Board*. Paper No. 00-0125: 2000; **1719**(1): 103–111.
[Publisher Full Text](#)
- Sharma S, Lingras P, Xu F, et al.: **Application of Neural Networks to estimate AADT on low-volume roads**. *J Transp Eng*. 2001; **127**(5): 426–432.
[Publisher Full Text](#)
- Shamo B, Asa E, Membah J: **Linear spatial interpolation and analysis of Annual Average Daily Traffic Data**. *J Comput Civ Eng*. 2015; **29**(1).
[Publisher Full Text](#)
- Shon D: **Traffic Volume Forecasting for Rural Alabama state highways**. M.S. thesis. Auburn University, Auburn, Alabama, 1989.
- Smith LM, Stroup WW, Marx DB: **Poisson cokriging as a generalized linear mixed model**. *Spat Stat*. 2020; **35**: 100399.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Staats WN: **Estimation of Annual Average Daily Traffic On Local Roads In Kentucky**. *Theses And Dissertations-Civil Engineering*. 2016; 36.
[Publisher Full Text](#)
- Stein A, Corsten LCA: **Universal Kriging and Cokriging as a Regression Procedure**. *International Biometric Society*. 1991; **47**(2): 575–587.
[Publisher Full Text](#)
- Sunila R: **Geostatistics: Kriging**, Konetekniikka 1, Otakaari 4,150. 2015.
[Reference Source](#)
- Tziachris P, Metaxa E, Papadopoulos F, et al.: **Spatial Modelling and Prediction Assessment of Soil Iron Using Kriging Interpolation with pH as Auxiliary Information**. *Int J Geoinf*. 2017; **6**(9): 283.
[Publisher Full Text](#)
- Varga B, Pereira M, Kulcsár B, et al.: **Data-Driven Distance Metrics for Kriging-Short-Term Urban Traffic State Prediction**. In: *IEEE Transactions on Intelligent Transportation Systems*, 2023; 1–12.
[Publisher Full Text](#)
- Veeken PCH: **Seismic Stratigraphy, Basin Analysis And Reservoir Characterisation**. *Handbook Of Geophysical Exploration: Seismic Exploration*. 2007; **37**: 1–509.
[Reference Source](#)
- Wackernagel H: **Cokriging versus kriging in regionalized multivariate data analysis**. *Geoderma*. 1994; **62**(1–3): 83–92.
[Publisher Full Text](#)
- Xia Q, Zhao F, Shen L, et al.: **Estimation of Annual Average Daily Traffic for non-state roads in a Florida County**. Research Report. Florida Department of Transportation, Tallahassee. 1999.
[Publisher Full Text](#)
- Yang B, Wang S, Bao Y: **Efficient Local AADT estimation via SCAD variable selection based on regression models**. *Chinese Control and Decision Conference, IEEE Transactions on Intelligent Transportation System*. 2011.
[Publisher Full Text](#)
- Yang B, Wang S, Bao Y: **New efficient regression method for local AADT estimation via SCAD variable selection**. *IEEE Trans Intell Transp Syst*. 2014; **15**(6): 2726–2731.
[Publisher Full Text](#)
- Zhang H, Cai W: **When Doesn't Cokriging Outperform Kriging?** *Statistical Science*. 2015; **30**(2): 176–180.
[Publisher Full Text](#)
- Zhang X, Chen M: **Enhancing Statewide Annual Average Daily Traffic**

Estimation with Ubiquitous Probe Vehicle Data. *Transp Res Rec.* 2020; **2674**(9): 649–660.

[Publisher Full Text](#)

Zhao F, Park N: **Using geographically weighted regression models to estimate Annual Average Daily Traffic.** *Transportation Research Record: Journal of the Transportation Research Board.* TRB, National Research Council, Washington, D.C.: 2004; **1879**(1): 99–107.

[Publisher Full Text](#)

Zhong M, Hanson B: **GIS-Based travel demand modeling for estimating**

traffic on low-class roads. *Transp Plan Technol.* 2009; **32**(5): 423–439.

[Publisher Full Text](#)

Zhou F, Guo HC, Ho YS, *et al.*: **Scientometric analysis of geostatistics using multivariate methods.** *Scientometrics.* 2007; **73**(3): 265–279.

[Publisher Full Text](#)

Zou H, Yue Y, Li Q, *et al.*: **An improved distance metric for the interpolation of link-based traffic data using kriging: a case study of a large-scale urban road network.** *Int J Geogr Inf Sci.* 2012; **26**(4): 667–689.

[Publisher Full Text](#)

Open Peer Review

Current Peer Review Status:   

Version 2

Reviewer Report 31 July 2023

<https://doi.org/10.21956/emeraldopenres.16205.r28790>

© 2023 Pereira M. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Mike Pereira

Mines Paris - PSL University, Saint-Michel, Paris, France

The authors present a study on the estimation of average daily traffic data using a geostatistical approach. They propose to estimate traffic by cokriging using both traffic and census data. They apply their approach to traffic data from Montana, Minnesota, and Washington and conclude that using population data as an auxiliary variable for traffic state estimation helps to improve the results.

Here are my comments:

- I have some reservations about the work presented in this paper. Mainly, the contributions and novelty of the work are not clear to me. If it is the use of cokriging to estimate traffic, then several works have already been made in that direction (see eg. ¹). It is the fact that population numbers are correlated to traffic volumes and could be used as an auxiliary variable, one could argue that this is natural since the traffic volumes seem to be averaged at a city-level (and the more people in the city, the more vehicles in the city). Hence, the authors should make their contribution clearer.
- I do not see how the method proposed by the authors is particularly suited for low-volume roads, as stated in the article title. Similarly, as stated at the end of the introduction, I do not see how the study presented here allows us to conclude that "the more variables there are for a predictive model, the better the model output".
- The choices made in the cokriging implementation (eg. data transformation, experimental variogram parameters, nugget,..) are not discussed. Instead, the authors seem to only take the default values given by the software. However, the tuning of these parameters can sometimes greatly improve the model and its performances. A discussion on how to properly choose and tune such parameters would have been beneficial to the paper.
- The presentation of the geostatistical tools (kriging, cokriging, variogram) is muddled. Some statements made in the text are false: for instance, cokriging can be used for more than 4 variables, and cokriging does not always yield better results than kriging (it depends on

which variables are used, and how well the multivariate model fits the data). Some notions are not very well-defined (for instance, there is no explanations of what the various terms of eq. 6 mean). Hence, the presentation of the geostatistical tools should be rewritten in a more concise and accurate manner.

- The introduction could be improved: there is not enough focus on the task at hand, which is predicting AADT. Instead, too much time is spent on a lengthy review of all the possible uses of cokriging. Hence, the introduction is in need of restructuring.
- To conclude that cokriging performs good traffic predictions, the authors look at leave-one-out cross-validation errors. This is not appropriate, as such errors should rather be used for model selection, not to assess the quality of the predictions. Indeed, each time, the traffic is predicted at only one location using all the other data points. Using a training and test set would have been better.
- A comparison between cokriging and ordinary / universal kriging could have been made to really assess the supposed "superiority" of cokriging.

References

1. Bae B, Kim H, Lim H, Liu Y, et al.: Missing data imputation for traffic flow speed using spatio-temporal cokriging. *Transportation Research Part C: Emerging Technologies*. 2018; **88**: 124-139
[Publisher Full Text](#)

Is the work clearly and accurately presented and does it cite the current literature?

Partly

Is the study design appropriate and is the work technically sound?

Partly

Are sufficient details of methods and analysis provided to allow replication by others?

Partly

If applicable, is the statistical analysis and its interpretation appropriate?

No

Are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions drawn adequately supported by the results?

Partly

Is the argument information presented in such a way that it can be understood by a non-academic audience?

No

Does the piece present solutions to actual real world challenges?

Yes

Is real-world evidence provided to support any conclusions made?

Not applicable

Could any solutions being offered be effectively implemented in practice?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Geostatistics

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Reviewer Report 30 May 2023

<https://doi.org/10.21956/emeraldopenres.16205.r28887>

© 2023 Al-Mudhafar W. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Watheq J. Al-Mudhafar 

Basrah Oil Company, Basrah, Iraq

The authors have addressed all the comments and included their responses in the manuscript. Therefore, it can be accepted for passing peer review.

Is the work clearly and accurately presented and does it cite the current literature?

Yes

Is the study design appropriate and is the work technically sound?

Yes

Are sufficient details of methods and analysis provided to allow replication by others?

Yes

If applicable, is the statistical analysis and its interpretation appropriate?

Yes

Are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions drawn adequately supported by the results?

Yes

Is the argument information presented in such a way that it can be understood by a non-academic audience?

Yes

Does the piece present solutions to actual real world challenges?

Yes

Is real-world evidence provided to support any conclusions made?

Yes

Could any solutions being offered be effectively implemented in practice?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Geostatistics. Machine Learning.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Version 1

Reviewer Report 02 May 2023

<https://doi.org/10.21956/emeraldopenres.15734.r28787>

© 2023 Varga B. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Balázs Varga

Budapest University of Technology and Economics, Budapest, Hungary

The paper uses co-kriging to estimate annual average daily traffic (AADT) using co-kriging with population data. The dataset used is large, and focuses on state-wide estimation. The reviewer thinks some points need further analysis and/or clarification.

- Kriging is a well-established methodology and does need a detailed introduction in the abstract.
- In the introduction, co-kriging is introduced earlier than kriging. Additionally, some paragraphs are repetitive. Consider reorganizing the intro.
- In the literature review, the most recent reference is from 2020.

- Several contributions have recently been made to spatial traffic estimation, mainly from the machine learning field. The more, the authors need to mention machine learning as an alternative tool for multivariate traffic estimation^{1,2}).
- The literature review focuses much on how kriging was used before and less on how traffic estimation was tackled.
- The authors use separate kriging estimators for each year's traffic data. Do the authors see an opportunity to connect the temporal data to identify socio-economic trends?
- Data description does not discuss the types of roads used. Are the traffic flows in both directions? Do they consider every road from the dataset? Please elaborate on the data collection and pre-processing.
- The spatial distribution of the roads needs to be adequately discussed. However, the selection of measurement sites is a critical component in kriging.
- AADT is defined for one road section. Do the authors use detector locations as measurement sites?
- State highways (or main traffic arterials) are visible from the 2000 AADT heatmaps. I.e., there are higher traffic volumes along these highways. The authors should elaborate on this.
- Eq. 4.: "Variance as" should be outside the equation environment.
- Have the authors considered using ordinary kriging only on the traffic data as a benchmark? How much does introducing population census improve the prediction accuracy?
- Geographical distance of the sites might not accurately reflect the traffic on low-traffic roads. For example, the traffic of a busy highway correlates weakly with a nearby minor street. Using different distance metrics^{3,4}) might yield better estimates.

References

1. Das S, Tsapakis I: Interpretable machine learning approach in estimating traffic volume on low-volume roadways. *International Journal of Transportation Science and Technology*. 2020; **9** (1): 76-88 [Publisher Full Text](#)
2. Yeboah A, Codjoe J, Thapa R: Estimating Average Daily Traffic on Low-Volume Roadways in Louisiana. *Transportation Research Record: Journal of the Transportation Research Board*. 2023; **2677** (1): 1732-1740 [Publisher Full Text](#)
3. Varga B, Pereira M, Kulcsar B, Pariota L, et al.: Data-Driven Distance Metrics for Kriging-Short-Term Urban Traffic State Prediction. *IEEE Transactions on Intelligent Transportation Systems*. 2023. 1-12 [Publisher Full Text](#)
4. Zou H, Yue Y, Li Q, Yeh A: An improved distance metric for the interpolation of link-based traffic data using kriging: a case study of a large-scale urban road network. *International Journal of Geographical Information Science*. 2012; **26** (4): 667-689 [Publisher Full Text](#)

Is the work clearly and accurately presented and does it cite the current literature?

Partly

Is the study design appropriate and is the work technically sound?

Partly

Are sufficient details of methods and analysis provided to allow replication by others?

Partly

If applicable, is the statistical analysis and its interpretation appropriate?

Yes

Are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions drawn adequately supported by the results?

Yes

Is the argument information presented in such a way that it can be understood by a non-academic audience?

Yes

Does the piece present solutions to actual real world challenges?

Yes

Is real-world evidence provided to support any conclusions made?

Yes

Could any solutions being offered be effectively implemented in practice?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Intelligent Transportation Systems

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Reviewer Report 16 November 2022

<https://doi.org/10.21956/emeraldopenres.15734.r28226>

© 2022 Al-Mudhafar W. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Watheq J. Al-Mudhafar 

Basrah Oil Company, Basrah, Iraq

Moderate revisions are required:

1. The Abstract must be improved to reflect the following structure, especially the methods and procedure part:
 - **Objectives/Scope:** Please list the objectives and/or scope of your paper.
 - **Methods, Procedures, Process:** Briefly explain your overall approach, including your methods, procedures and process.
 - **Results, Observations, Conclusions:** Please describe the results, observations and conclusions of the proposed paper.
 - **Novel/Additive Information:** Please explain how your paper will present novel (new) or additive information to the existing body of literature that can be of benefit to and/or add to the state of knowledge in the petroleum industry.
 - The literature review section should be included in the Introduction Section.
 - There is a lack of literature review about other conventional geostatistical algorithms for petrophysical property modeling, such as Collocated Cokriging, universal kriging, and bayesian kriging. Therefore, the literature review should be improved by adding one more paragraph to review these methods. The following references are necessary to cover the aforementioned kriging approaches:

Al-Mudhafar, W. J. (2018). Bayesian Kriging for Reproducing Reservoir Heterogeneity in a Tidal Depositional Environment of a Sandstone Formation: A Case Study. *Journal of Applied Geophysics*. <https://doi.org/10.1016/j.jappgeo.2018.11.007>¹

Doyen, P.M., L.D. Den Boer, and W.R. Pillet. (1996). Seismic Porosity Mapping in the Ekofisk Field Using a New Form of Collocated Cokriging. SPE-36498-MS paper presented at the SPE Annual Technical Conference and Exhibition, Denver, Colorado².

Journel, A. G. and F. G. Alabert. (1990). New method for reservoir mapping. *Journal of Petroleum technology*, 42(02), 212-218³.

Al-Mudhafar, W. J. (2021). Geostatistical Simulation of Facies and Petrophysical Properties for Heterogeneity Modeling in A Tidal Depositional Environment: A Case Study From Upper Shale Member in A Southern Iraqi Oil Field. URTEc: 5551, the Unconventional Resources Technology Conference, Houston, TX⁴.

Journel, A.G., 1990. Geostatistics for Reservoir Characterization. SPE Annual Technical Conference and Exhibition, New Orleans, Louisiana <https://doi.org/10.2118/20750>⁵

Xu, W., T.T. Tran, R.M. Srivastava, and A.G. Journal. (1992). Integrating Seismic Data in Reservoir Modeling: The Collocated Cokriging Alternative. SPE-24742-MS presented at the SPE Annual Technical Conference and Exhibition, Washington, DC⁶.

Alabert FG, Massonnat GJ (1990) Heterogeneity in a complex turbiditic reservoir: stochastic modelling of facies and petrophysical variability. SPE-20604-MS paper presented at the SPE Annual Technical Conference and Exhibition, New Orleans, Louisiana⁷.

- The authors should provide in the last paragraph of the Introduction Section the conducted workflow, its strengths, advantages, and how it is different from previous approaches.
- A full description of the Variogram Analysis should be provided in the Methodology section. The following references may be useful:

Gringarten, E. and C. V. Deutsch. (1999). Methodology for variogram interpretation and modeling for improved reservoir characterization. In SPE annual technical conference (pp. 355-367)⁸.

- Also, the cross-validation description should be added in the Methodology section. You may refer to the following paper that describes the types of cross-validation techniques:

Wang, G., Ju, Y., Carr, T. R., Li, C., & Cheng, G. (2014). Application of Artificial Intelligence on Black Shale Lithofacies Prediction in Marcellus Shale, Appalachian Basin. Unconventional Resources Technology Conference. doi:10.15530/URTEC-2014-1935021⁹.

Al-Mudhafar, W. J. (2016). Incorporation of Bootstrapping and Cross-Validation for Efficient Multivariate Facies and Petrophysical Modeling. Society of Petroleum Engineers. doi:10.2118/180277-MS¹⁰.

Pirrone, M., Battigelli, A., & Ruvo, L. (2014). Lithofacies Classification of Thin Layered Turbidite Reservoirs Through the Integration of Core Data and Dielectric Dispersion Log Measurements. Society of Petroleum Engineers. doi:10.2118/170748-MS¹¹.

- Again, the variograms should be constructed and fitted prior to conducting the kriging interpolation. You provided the spatial maps then you showed the variogram fitting. The disorder should be addressed.
- The conclusions should be revised considering the raised points.
- References should be improved including the suggested references above.

References

1. Al-Mudhafar W: Bayesian kriging for reproducing reservoir heterogeneity in a tidal depositional environment of a sandstone formation. *Journal of Applied Geophysics*. 2019; **160**: 84-102 [Publisher Full Text](#)
2. Doyen PM, Den Boer LD, Pillet, WR: Seismic Porosity Mapping in the Ekofisk Field Using a New Form of Collocated Cokriging. 1996. [Publisher Full Text](#)
3. Journal A, Alabert F: New Method for Reservoir Mapping. *Journal of Petroleum Technology*. 1990; **42** (02): 212-218 [Publisher Full Text](#)
4. AlMudhafar W: Geostatistical Simulation of Facies and Petrophysical Properties for Heterogeneity Modeling in A Tidal Depositional Environment: A Case Study From Upper Shale Member in A Southern Iraqi Oil Field. 2021. [Publisher Full Text](#)

5. Journel AG: Geostatistics for Reservoir Characterization. 1990. [Publisher Full Text](#)
6. Wenlong X, Tran TT, Srivastava RM, Journel AG: Integrating Seismic Data in Reservoir Modeling: The Collocated Cokriging Alternative. 1992. [Publisher Full Text](#)
7. Alabert FG, Massonnat GJ: Heterogeneity in a complex turbiditic reservoir: stochastic modelling of facies and petrophysical variability. 1990. [Publisher Full Text](#)
8. Gringarten E, Deutsch CV: Methodology for variogram interpretation and modeling for improved reservoir characterization. 1999. 355-367 [Publisher Full Text](#)
9. Wang G, Ju Y, Li C, Carr TR, et al.: Application of Artificial Intelligence on Black Shale Lithofacies Prediction in Marcellus Shale, Appalachian Basin. 2014. [Publisher Full Text](#)
10. Al-Mudhafar WJ: Incorporation of Bootstrapping and Cross-Validation for Efficient Multivariate Facies and Petrophysical Modeling. 2016. [Publisher Full Text](#)
11. Pirrone M, Battigelli A, Ruvo L: Lithofacies Classification of Thin Layered Turbidite Reservoirs Through the Integration of Core Data and Dielectric Dispersion Log Measurements. 2016. [Publisher Full Text](#)

Is the work clearly and accurately presented and does it cite the current literature?

Yes

Is the study design appropriate and is the work technically sound?

Yes

Are sufficient details of methods and analysis provided to allow replication by others?

Partly

If applicable, is the statistical analysis and its interpretation appropriate?

Partly

Are all the source data underlying the results available to ensure full reproducibility?

Partly

Are the conclusions drawn adequately supported by the results?

Yes

Is the argument information presented in such a way that it can be understood by a non-academic audience?

Yes

Does the piece present solutions to actual real world challenges?

Yes

Is real-world evidence provided to support any conclusions made?

Yes

Could any solutions being offered be effectively implemented in practice?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Geostatistics. Machine Learning.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.
