

A Bayesian Regression Model for Estimating Average Daily Traffic Volumes for Low Volume Roadways

Mehrnaz Doustmohammadi¹ & Michael Anderson¹

¹Department of Civil and Environmental Engineering, University of Alabama in Huntsville, Huntsville, United States

Correspondence: Mehrnaz Doustmohammadi, Department of Civil and Environmental Engineering, University of Alabama in Huntsville, Huntsville, AL, 35899, United States. Tel: 205-243-6717 E-mail: md0033@uah.edu

Received: November 25, 2018 Accepted: December 20, 2018 Online Published: December 28, 2018

doi:10.5539/ijsp.v8n1p143

URL: <https://doi.org/10.5539/ijsp.v8n1p143>

Abstract

Common Average Daily Traffic (ADT) estimation models use Linear Regression and a collection of socio-economic and roadway variables. While linear regression is widely understood, it is not always optimal for developing prediction models as the regression techniques don't have the ability to account for data distributions, or variability of the point estimates. To overcome this limitation, this paper presents a study that utilizes a Bayesian Regression model to develop a model to estimate ADT values for low volume roadways. The need for ADT estimates is critical as roadway traffic counts are the backbone of maintenance, safety and construction designs. While significant investment is made in collecting ADT values for higher functionally classified and high volume roadways, low volume roadways are often neglected in the traffic count program due to budget limitations and the misguided notion that there is limited return on investment in counting these facilities. This research developed a technique to estimate ADT for local roads in Alabama incorporating variables used in previous studies and a Bayesian Regression model. The final Bayesian Regression model relies on four independent variables: number of households in the area, employment in the area, population to job ratio and access to major roads. The model was used to generate ADT estimates on low-volume rural, local roads for 12 counties in Alabama. The paper concludes that the model can be used to predict the ADT for low-volumes roadways in Alabama for future applications.

Keywords: low-volume roads, traffic counts, estimation

1. Introduction

1.1 The Problem

Average Daily Traffic (ADT) is a vital attribute for any roadway when considering maintenance, safety and construction. While ADT data is usually collected for major roadways, low-volume roadways are generally not counted as part of a routine data collection scheme (Raja, Doustmohammadi, & Anderson, 2018). This study focuses on establishing a new model for estimating ADT for low-volume roads using a Bayesian Regression model. The advantage of the model is the ability to incorporate the fact that the independent variables tend to come from a distribution and not a single point estimate (Ma, Kockelman, & Damien, 2008). The use of the distribution allows for a potentially more accurate model than traditional linear regression, which use single point estimates of the variables (Ma, Kockelman, & Damien, 2008).

This paper presents a Bayesian Regression model for estimating low-volume ADT for a collection of roadways in Alabama. The model was developed using socio-economic factors as independent variables including: nearby population, number of households in the area, employment in the area, population to job ratio and access to major roads. The paper presents a brief literature review on ADT estimation and Bayesian Regression modeling, describes the data used for the study, model developed and accuracy of the model and makes overall conclusions about the use of Bayesian Regression modeling for ADT estimation. This paper concludes that this modeling technique performs slightly better than traditional linear regression models due to the added knowledge of variable distributions.

1.2 Background

The motivation behind developing an ADT estimation model is to produce models that can be used to supplement costly data collection efforts. While ADT estimation models are certainly not unique, the use of Bayesian Regression as an effort to improve the model versus linear regression has not been fully studied. For high volume and urban roadways, numerous studies have been attempted with a variety of independent variables including: population, employment, total number of lanes, location type (urban/rural), personal income, vehicle registrations (Doustmohammadi, Anderson, &

Doutmohammadi, 2017; Gecchele, Rossi, Gastaldi, & Caprini, 2011; Lowry & Dixon, 1996; Sharma, Gulati, & Rizak, 1996; Pan, 2008; Zhao & Park, 2004; Zhong & Liu, 2007; Doutmohammadi & Anderson, 2016; Zhao & Chung, 2001; Anderson, Sharfi, & Gholston, 2006). Attempts have been made to forecast ADT on lower volume roadways as well (Raja, Doustmohammadi, & Anderson, 2018; Mohamad, Sinha, Kuczek, & Scholer, 2013; Garber, 1984; Wang, Gan, & Alluri, 2013; Zhao & Chung, 2001; Sharma, Lingras, Xu, & Kilbum, 2001; “Estimation of Annual...”, 1999). These studies all use a regression model of some type, but do not use Bayesian Regression.

Bayesian Regression models are not new to transportation analysis as they have been used many times for crash analysis and safety studies (Ma, Kockelman, & Damien, 2008; Etz, N.d.; Spiegelhalter et al, 2002; Xie, Lord, & Zhang, 2007; Ma & Kockelman, 2006; Maher & Summersgill, 1996). Bayesian Regression analysis is similar to linear regression with some enhancements. In Bayesian methodology, each dependent and independent variable is formulated based on the distribution rather than the point estimate (Etz, N.d.). Response sampled from a normal distribution in Bayesian linear regression is $y \sim N(TX, 2I)$ (Etz, N.d.). Bayesian linear regression does not find the single best value of model parameters, but rather to estimate the posterior distribution for the model parameter. The advantage is that the posterior regression will be more accurate due to the distribution of the regression. Therefore, the Bayesian Regression model should, in theory, be superior to the linear regression model.

2. Method

2.1 Data Collection

This study used a collection of low volume traffic counts and collected socio-economic data to create the Bayesian Regression model. The traffic counts were collected for this study as additional counts that were being collected by the state to support the needs of the Highway Performance Monitoring System (HPMS). For this study, 205 low-volume counts were collected from several rural counties in Alabama. The traffic count locations are shown in Figure 1.

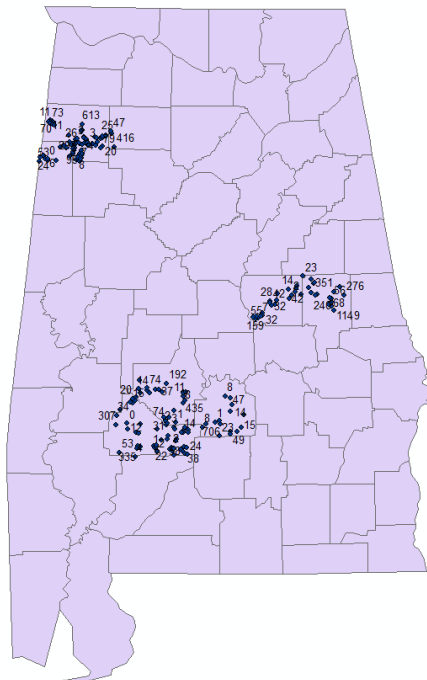


Figure 1. Traffic Count Locations

The socio-economic data included the population in the census blocks near the count location, number of households in the census blocks near the count location, employment in the census blocks near the count location and the location of state routes in Alabama. The data for the population, number of households and number of employees in the census blocks near the count locations were obtained from the Census Department. Additionally, a value of population to jobs comparison was calculated to determine if the traffic count was on a roadway that offered access to employment, as this would increase the traffic count. For the study, if the value of population to job was less than 1.0, the data was tagged with a 1, otherwise, the value was recorded as a 2. The data collection was performed using roadway counters placed according to standard traffic count collection practices and ArcGIS was used to map the data and assist in the collection of socio-economic data, as shown in Figure 2.

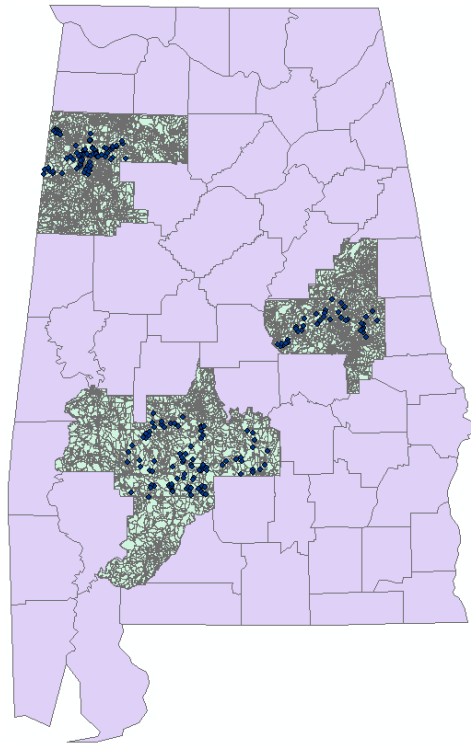


Figure 2. Count Locations and Census Demographic Data

The state routes in Alabama were obtained from the Alabama Department of Transportation and were used to identify key locations that the low-volume roadway would potentially connect. This data was collected to obtain a connection to major roadway factor. For the data used in this study, a count location near a major facility was given a value of 1 and a count location away from a major facility was given a value of 2. The routes included all Interstates, U.S. Highways and State Highways as shown in Figure 3.

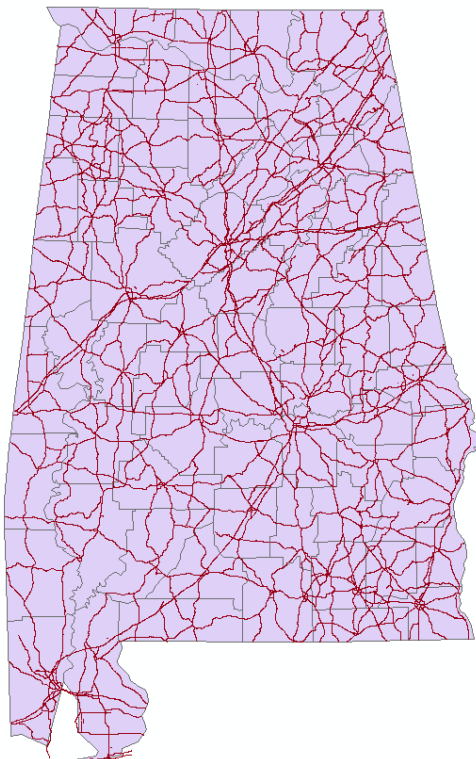


Figure 3. Alabama State Road System

The collected data were divided into two groups, data used to develop the model and data used for validation of the model. There were 150 locations that were used for model development and 55 used for validation. The traffic count data used for model development and model validation ranged from 1 vehicle in a day to 1,163 vehicles per day. The block population in the zones near the count ranged from 0 to 227, the number of households ranged from 0 to 135 and the block employment ranged from 0 to 128. Table 1 shows a summary of the data and statistical values for the data that were used for model development and model validation.

Table 1. Summary Statistics of the Data

| Model Development Data (N=150) | | | | |
|--------------------------------|------|--------------------|---------|---------|
| Parameter | Mean | Standard Deviation | Minimum | Maximum |
| Traffic Count | 129 | 180 | 1 | 1,163 |
| Block Population | 40 | 48 | 0 | 227 |
| Number of Households | 21 | 25 | 0 | 129 |
| Block Employment | 11 | 19 | 0 | 92 |
| Model Validation Data (N=55) | | | | |
| Parameter | Mean | Standard Deviation | Minimum | Maximum |
| Traffic Count | 213 | 239 | 29 | 1,149 |
| Block Population | 54 | 50 | 0 | 181 |
| Number of Households | 30 | 31 | 0 | 135 |
| Block Employment | 20 | 24 | 0 | 128 |

2.2 Model Development

The research team developed a traditional linear regression model and a Bayesian Regression model for the data. The initial testing was to determine the independence of the variables through a variance-covariance matrix expressed as (Montgomery, Peck, & Vining, 2012):

$$\text{Variance} = \begin{bmatrix} X1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & Xn \end{bmatrix} \tag{1}$$

From the matrix, it was determined that Block Population should be removed from the model as Number of Households was sufficient and reduced the necessity for both variables. Additionally, the Linear Regression model will be developed without variables that are not significant from a point estimate standpoint. Therefore, some variables from the Linear Regression models might not appear in the final model. The Bayesian Regression model, as it uses a distribution for the independent variables, all variables entered into the model will appear in the final equation.

The format of the linear regression models generated from the analysis take the same format. The regression models follow a traditional linear equation format. The Linear Regression model and Bayesian Regression model will both have the following format (Montgomery, Peck, & Vining, 2012):

$$\begin{bmatrix} y1 \\ y2 \\ \dots \\ yp \end{bmatrix} = \begin{bmatrix} 1 & x1,1 & x1,n \\ \vdots & \vdots & \vdots \\ 1 & x1,p & Xn,p \end{bmatrix} * \begin{bmatrix} \beta1 \\ \beta2 \\ \dots \\ \betan \end{bmatrix} \tag{2}$$

The models developed for estimating low volume roadway ADT using the randomly assigned 150 traffic counts are:

Traditional Linear Regression Model:

$$[\text{Traffic Count}] = \begin{bmatrix} \text{Number of Households} \\ \text{Block Employment} \\ \text{Access to Major Roadways} \end{bmatrix} * [5.209 \quad 2.564 \quad - 5.389] \tag{3}$$

Bayesian Regression Analysis Model:

$$[\text{Traffic Count}] = \begin{bmatrix} \text{Intercept} \\ \text{Number of Households} \\ \text{Block Employment} \\ \text{Population to Job Comparison} \\ \text{Access to Major Roadways} \end{bmatrix} * [-36.587 \quad 5.16 \quad 3.682 \quad 10.639 \quad 6.262] \tag{4}$$

The quality of the models was determined using statistical methods and visualizations. The statistical accuracy was

calculated using a Percent Root Mean Square Error (%RMSE), the common value for used for validating accuracy of travel demand models. (Montgomery, Peck, & Vining, 2012):

$$\% RMSE = \frac{100 * \sqrt{\frac{(Observed Value i - Expected Value i)^2}{n}}}{Average Traffic Count} \tag{5}$$

The %RMSE for the data used to develop the models was calculated to be 62.15 for the Linear Regression model and 63.53 for the Bayesian Regression model. Figure 4 shows a scatter plot of the actual traffic counts versus the predicted counts for the Linear models and Figure 5 shows a scatterplot for the Bayesian model. From the scatterplots, while the model results and plots are very similar, the Bayesian Regression model tends to predict values slightly higher than the Linear Regression model, especially for those traffic counts that would be considered the higher of the low volume counts.

To more accurately determine the differences between the models in the higher category of low volume roadways, for the 25 roadways in the model development grouping where the traffic counts were greater than 250, the Bayesian Regression model has a %RMSE of 31.09 while the Linear Regression model has a %RMSE of 32.81. The accuracy of these higher category low volume traffic counts will prevent under-prediction of these roadways, which is considered more important than being accurate in the extremely low volume category. Due to this result, the paper provides details for validation of the Bayesian Regression model only.

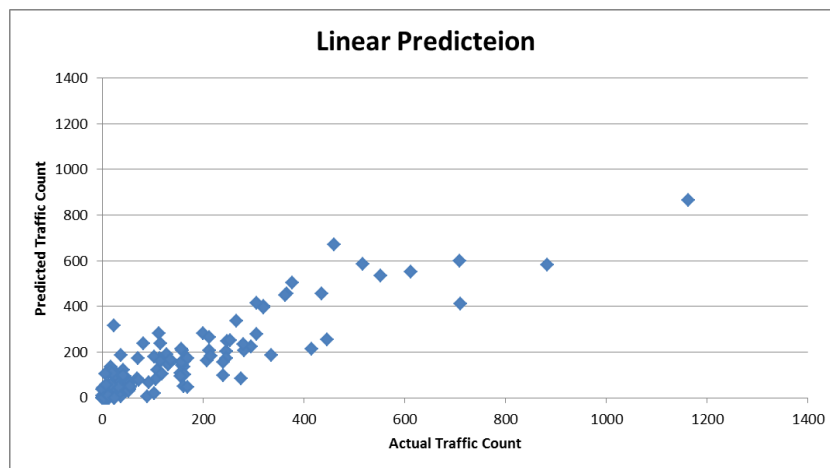


Figure 4. Linear Model Scatterplot

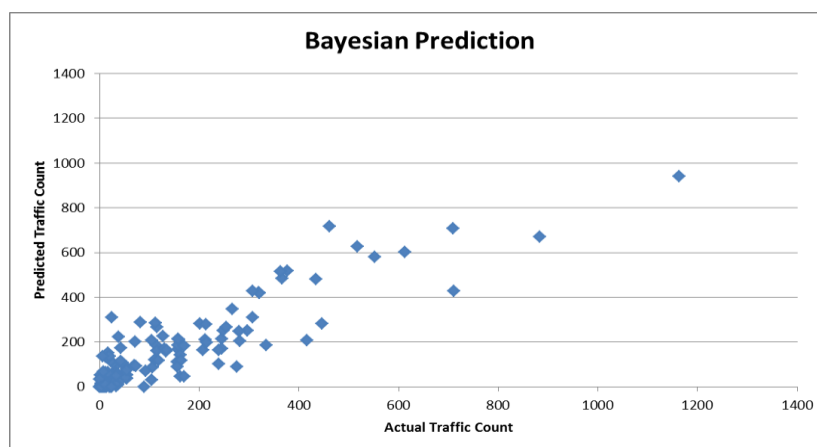


Figure 5. Bayesian Model Scatterplot

2.3 Model Validation

The validation of the models was performed using 55 traffic counts from the original data collection effort that were not used in the model development. The calculated %RMSE for the validation model set using the Bayesian Regression model was 48.30, actually better than the dataset used for model development. Figure 6 shows the scatterplot of the Bayesian Regression model to the actual traffic counts.

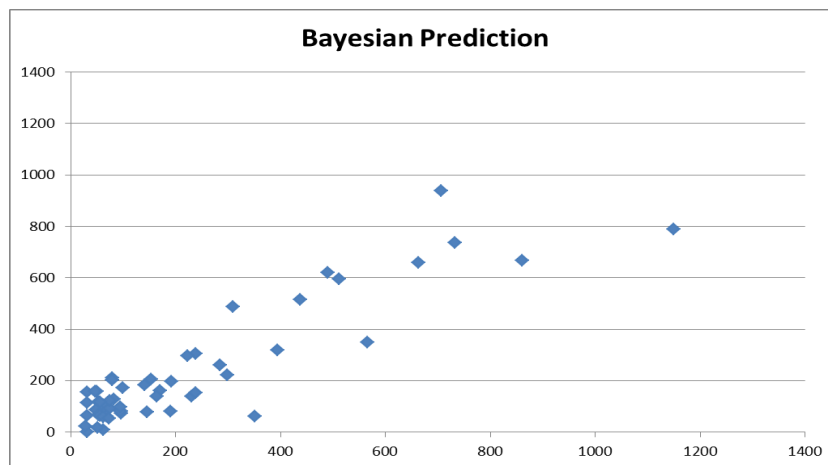


Figure 6. Validation Plot of the Bayesian Regression model

3. Conclusions

This paper examined the development of a Bayesian Regression model for predicting traffic count volumes for low-volume roadways. The model was developed using a collection of 205 traffic counts on low volume roadways and a collection of demographic variables near the count locations. The use of Bayesian Regression was performed to allow for variations in data to be used to develop the model in hopes to develop a model that was more accurate than Linear Regression. The models developed using Bayesian Regression tended to be more accurate at predicting the higher volume category of the low volume roadways.

The overall contribution to this paper is a new Bayesian Regression model that can be used to predict ADT values for low volume roadways. The volume range that the equations presented in this work are generally for roadways with an anticipated traffic volume of less than 1,000 vehicles per day. The equation presented can predict traffic counts for roadways during the current year and also have the benefit, due to the use of projected demographic values, to forecast a traffic count in the future, to continue to support maintenance, safety and construction.

Acknowledgement

This paper was made possible by funding provided by the Alabama Department of Transportation, SPR-0001(058) 2015-533.

References

- Anderson, M. D., Sharfi, K., & Gholston, S. E. (2006). Direct demand forecasting model for small urban communities using multiple linear regression. *Transportation research record*, 1981(1), 114-117. <https://doi.org/10.1177/0361198106198100117>
- Doustmohammadi, M., & Anderson, M. (2016). Developing Direct Demand AADT Forecasting Models for Small and Medium Sized Urban Communities. *International Journal of Traffic and Transportation Engineering*, 5(2), 27-31.
- Doustmohammadi, M., Anderson, M., & Doustmohammadi, E. (2017). Using Log Transformations to Improve AADT Forecasting Models in Small and Medium Sized Communities. *International Journal of Traffic and Transportation Engineering*, 6(2), 23-27.
- Estimation of annual average daily traffic for off-system roads in Florida[R]. Florida Department of Transportation, 1999.
- Etz, A. (2018). Introduction to the concept of likelihood and its applications. *Advances in Methods and Practices in Psychological Science*, 1(1), 60-69. <https://doi.org/10.1177/2515245917744314>
- Garber, N. J. (1984). *Optimizing traffic count program: a methodology for estimating AADT volumes from short-duration counts* (No. VHTRC 85-R14).
- Gecchele, G., Rossi, R., Gastaldi, M., & Caprini, A. (2011). Data mining methods for traffic monitoring data analysis: A case study. *Procedia-Social and Behavioral Sciences*, 20, 455-464. <https://doi.org/10.1016/j.sbspro.2011.08.052>
- Lowry, M., & Dixon, M. (2012). GIS tools to estimate average annual daily traffic. *National Institute for Advanced Transportation Technology. University of Idaho*, 109.

- Ma, J., & Kockelman, K. M. (2006). Bayesian multivariate Poisson regression for models of injury count, by severity. *Transportation Research Record, 1950*(1), 24-34. <https://doi.org/10.1177/0361198106195000104>
- Ma, J., Kockelman, K. M., & Damien, P. (2008). A multivariate Poisson-lognormal regression model for prediction of crash counts by severity, using Bayesian methods. *Accident Analysis & Prevention, 40*(3), 964-975. <https://doi.org/10.1016/j.aap.2007.11.002>
- Maher, M. J., & Summersgill, I. (1996). A comprehensive methodology for the fitting of predictive accident models. *Accident Analysis & Prevention, 28*(3), 281-296. [https://doi.org/10.1016/0001-4575\(95\)00059-3](https://doi.org/10.1016/0001-4575(95)00059-3)
- Mohamad, D., Sinha, K. C., Kuczek, T., & Scholer, C. F. (2013). Annual Average Daily Traffic Prediction Model for County Roads, *Transportation Research Board 2013 Annual Meeting*.
- Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). *Introduction to linear regression analysis* (Vol. 821). John Wiley & Sons.
- Pan, T. (2008). Assignment of estimated average annual daily traffic on all roads in Florida.
- Raja, P., Doustmohammadi, M., & Anderson, M. D. (2018). Estimation of Average Daily Traffic on Low Volume Roads in Alabama. *International Journal of Traffic and Transportation Engineering, 7*(1), 1-6.
- Sharma, S. C., Gulati, B. M., & Rizak, S. N. (1996). Statewide traffic volume studies and precision of AADT estimates. *Journal of transportation engineering, 122*(6), 430-439. [https://doi.org/10.1061/\(ASCE\)0733-947X\(1996\)122:6\(430\)](https://doi.org/10.1061/(ASCE)0733-947X(1996)122:6(430))
- Sharma, S., Lingras, P., Xu, F., & Kilburn, P. (2001). Application of neural networks to estimate AADT on low-volume roads. *Journal of Transportation Engineering, 127*(5), 426-432. [https://doi.org/10.1061/\(ASCE\)0733-947X\(2001\)127:5\(426\)](https://doi.org/10.1061/(ASCE)0733-947X(2001)127:5(426))
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 64*(4), 583-639. <https://doi.org/10.1111/1467-9868.00353>
- Wang, T., Gan, A., & Alluri, P. (2013). Estimating annual average daily traffic for local roads for highway safety analysis. *Transportation Research Record: Journal of the Transportation Research Board, (2398)*, 60-66. <https://doi.org/10.3141/2398-07>
- Xie, Y., Lord, D., & Zhang, Y. (2007). Predicting motor vehicle collisions using Bayesian neural network models: An empirical analysis. *Accident Analysis & Prevention, 39*(5), 922-933. <https://doi.org/10.1016/j.aap.2006.12.014>
- Zhao, F., & Chung, S. (2001). Contributing factors of annual average daily traffic in a Florida county: exploration with geographic information system and regression models. *Transportation Research Record: Journal of the Transportation Research Board, (1769)*, 113-122. <https://doi.org/10.3141/1769-14>
- Zhao, F., & Park, N. (2004). Using geographically weighted regression models to estimate annual average daily traffic. *Transportation Research Record: Journal of the Transportation Research Board, (1879)*, 99-107. <https://doi.org/10.3141/1879-12>
- ZHONG, M., & Guoxin, L. I. U. (2007). Establishing and managing jurisdiction-wide traffic monitoring systems: North American experiences. *Journal of Transportation Systems Engineering and Information Technology, 7*(6), 25-38. [https://doi.org/10.1016/S1570-6672\(08\)60002-1](https://doi.org/10.1016/S1570-6672(08)60002-1)

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).