FORECASTING NETWORK DATA: SPATIAL INTERPOLATION OF TRAFFIC COUNTS USING TEXAS DATA

Xiaokun Wang
Assistant Professor
Department of Civil and Environmental Engineering
Bucknell University

Kara M. Kockelman
Associate Professor & William J. Murray Jr. Fellow
(Corresponding author)
Department of Civil, Architectural and Environmental Engineering
The University of Texas at Austin
6.9 ECJ, Austin, TX 78712-1076
kkockelm@mail.utexas.edu

The following paper is a pre-print and the final publication can be found in
Transportation Research Record No. 2105: 100-108, 2009.
Presented at the 88th Annual Meeting of the Transportation Research Board,
January 2009

ABSTRACT

Annual average daily traffic (AADT) values have long played an important role in transportation design, operations, planning, and policy making. However, AADT values are almost always rough estimates based on the closest short-period traffic counts, factored up using permanent automatic traffic recorder data. This study develops Kriging-based methods for mining network and count data, over time and space. Using Texas highway count data, the method forecasts future AADT values at locations where no traffic detectors are present. While low-volume road counts remain difficult to predict, available explanatory variables are very few, and extremely high-count outlier sites skew predictions in the data set used here, overall AADT-weighted median prediction error is 31% percent (across all Texas network sites). Here, Kriging performed far better than other options for spatial extrapolation − such as assigning AADT based on a point's nearest sampling site, which yields errors of 80%. Beyond AADT estimation, Kriging is a promising way to explore spatial relationships across a wide variety of data sets, including, for example, pavement conditions, traffic speeds, population densities, land values, household incomes, and trip generation rates. Further refinements, including spatial autocorrelation functions based on network (rather than Euclidean) distances and inclusion of far more explanatory variables exist, and will further enhance estimation.

Keywords: Annual average daily traffic (AADT), traffic counts, spatial interpolation, Kriging

INTRODUCTION

For transportation professionals, vehicle flow is perhaps the most important indicator for infrastructure management decisions, air quality estimates, and crash rate analyses. Each year, transportation agencies within the U.S. spend a significant amount of time and money collecting this information, as required by the Federal Highway Administration for the nation's Highway Performance Monitoring System (HPMS) database.

At roughly 50 to 200 sites across each U.S. state's highway network, automatic traffic recorders (ATRs) exist to continuously collect traffic count data. In essence then, just one ATR station exists for over 400 centerline miles of State-maintained roadway (or more than 1,000 miles of total public roadway), and these counts are often incomplete (as devices falter and go "off line" for days or weeks at a time). In order to estimate AADT at other network locations, agencies collect short-period traffic counts (SPTCs) at thousands of locations statewide, for one to three days every year or two – or five, as in the case of the data used here. In any given year, each short-period collection result may proxy for 5 to 20 or more centerline miles of state-maintained roadway.

The SPTC data are factored up using ATR-based day-of-week, month-of-year, axle-number and annual-growth factors, in order to obtain AADT estimates. While there is great opportunity for estimation error at any given site – estimated to be on the order of 10 to 20 percent by Gadda et al. (2007), an even bigger challenge is to forecast traffic flows at locations anywhere else in the network. Intuitively, traffic volumes between two SPTC sites should be close to counts at those two sites, particularly if they share the same road classification and number of lanes. A spatial technique worth evaluating here is the method of Kriging, which presumes spatial autocorrelation in error terms/unobserved factors, as a function of distance. This method is common to the discipline of regional science and could be put to much use within the field transportation.

Using linear extrapolation for temporal forecasting and Kriging methods for spatial interpolation, this study utilizes available data while producing AADT estimates for all sites in the Texas DOT-maintained network (of 79,500 centerline miles). The method exploits the spatial features of traffic count data while providing future-year traffic forecasts at any network location. Results indicate where traffic count uncertainty (errors in prediction) is highest, allowing agencies to more strategically sample under this not inexpensive, federally-mandated process.

The data used here are AADT estimates for 27,738 sites from 1999 to 2005, estimated using Texas DOT permanent ATR data to factor up one-day urban saturation count data. The 2006 AADT values at these 27,738 sites are predicted first , and then values at other locations throughout the network are spatially interpolated. The following sections describe existing studies on related topics, the data and method used here, and finally the analytical results.

LITERATURE REVIEW

Over many years, researchers and practitioners have sought to produce better traffic volume predictions (over time and space) based on limited data. The methods adopted most widely to date involve ordinary least squares regression with as many control variables as possible. For example, Zhao and Chung (2001) developed such models for AADT predictions in Florida, by using a variety of land-use and accessibility measures. Various other methods have also been explored. Tang et al. (2003) compared time series, neural network, nonparametric regression, and Gaussian maximum likelihood (GML) techniques and found that GML methods were most promising. Lam et al.'s (2006) later study used non-parametric models and GML methods to forecast Hong Kong's traffic volumes (using the state's Annual Traffic Census data). With the growing availability of geographic information system (GIS) datasets and evolution of spatial analysis techniques, researchers have started to explore methods that exploit the spatial context of traffic and other data. The study most closely related to this topic is Eom et al.'s (2006) use of Kriging methods to predict AADT for non-freeway facilities in Wake County, North Carolina. They conclude that the overall predictive capability of such methods eclipses traditional models.

The method of Kriging was first developed by Georges Matheron (1963), based on the Master's thesis of Krige (1951), a South African mining engineer who used a prototype of this technique to predict ore reserves. After several decades' development, Kriging has become a core geostatistics tool and is now used in many topic areas. For example, Bayraktar and Turalioglu (2005) used Kriging for air quality analysis, Emerson (2005) applied Kriging to natural resource analysis, and Zimmerman et al. (1998) used Kriging for water studies. Such methods can be used to predict count values at unmeasured locations while assessing the errors of these predictions. They rely on the notion that unobserved factors are autocorrelated over space, and the levels of autocorrelation decline with distance. Meanwhile, the values to be predicted may depend on several observable causal factors (e.g., number of lanes, posted speed limit, and facility type). These create a "trend" estimate, $\mu(s)$; so, in general, spatial variables can be defined as follows:

$$Z_i(s) = \mu_i(s) + \varepsilon_i(s) \tag{1}$$

where $Z_i(s)$ is the variable of interest (actual traffic count here) and s gives location (x,y coordinates) of site i. $Z_i(s)$ is composed of a deterministic trend $\mu_i(s)$ and a random error component $\varepsilon_i(s)$. The various $\varepsilon(s)$ values are correlated over space. Features of "trend" (often called "drift" in other studies), or the expected value of $Z(s)$, result in three types of Kriging: If $\mu(s)$ is constant across locations (or explanatory information is lacking), one can rely on Ordinary Kriging. Trends that depend on explanatory variables and unknown regression coefficients must rely on Universal Kriging. If the trend is known, one has Simple Kriging. ArcGIS's "Geostatistical Analyst" and "Spatial Analyst" (ESRI 1996) tools can be used to fit and then apply these three different methods.

Weak stationarity is assumed in all three cases, so that the correlation between $Z(s)$ and $Z(s+h)$ does not depend on actual locations, but only the distance h between the two sites. Furthermore, thanks to weak stationarity, the variance of $Z(s+h)$ - $Z(s)$ equals $2\gamma(h)$ for any s and h. Useful illustrations of such behaviors include variograms, where $2\gamma(h)$ values are plotted along the y-axis (versus distance), and semivariograms, where $\gamma(h)$ is used as the y coordinate.

In Universal Kriging, $\mu(s)$can be a deterministic function of any form. A simple assumption is to use a linear function where $\mu(s)$=Xβ (where X contains explanatory variables, like number of lanes and facility type). In contrast, $\varepsilon(s)$ reflects unobserved variation (e.g., local land use patterns and transit service levels).

For purposes of prediction, Kriging is performed on the $Z(s)$ values. The sum of interpolated random components $\varepsilon(s)$ and estimated $\mu(s)$ values provide the $Z(s)$ value estimates. The following section explains how the random component $\varepsilon(s)$ is estimated via Kriging.

As noted above, weak stationarity ensures the following:

$$\gamma(h) = 1/2 \text{ var } [Z(s+h) - Z(s)] \tag{2}$$

Where var $[Z(s+h) - Z(s)]$ is the variance (over all sites) between counts at sites s and $s+h$. The first step is to select an appropriate "curve" or semivariogram model to best fit this relationship (γ vs. h), for a given dataset. There are several commonly used models, including exponential, spherical and Gaussian models. These specifications are as follows:

1. Exponential

$$\gamma(h) = \begin{cases} c_0 + c_1\left[1 - \exp\left(-\frac{h}{a}\right)\right] & \textit{if } h > 0 \\ 0 & \textit{otherwise} \end{cases} \tag{3}$$

2. Spherical

$$\gamma(h) = \begin{cases} c_0 + c_1\left[1.5\frac{h}{a} - 0.5\left(\frac{h}{a}\right)^3\right] & \textit{if } 0 < h < a \\ c_0 + c_1 & \textit{if } h > a \\ 0 & \textit{otherwise} \end{cases} \tag{4}$$

3. Gaussian

$$\gamma(h) = \begin{cases} c_0 + c_1\left[1 - \exp\left(-\frac{h}{a}\right)^2\right] & \textit{if } h > 0 \\ 0 & \textit{otherwise} \end{cases} \tag{5}$$

These models all rely on three parameters that describe their shape while quantifying the level of spatial autocorrelation in the data. c_0 is called the "nugget effect", and it reflects discontinuity at the variogram's origin, as caused by factors such as sampling error and short scale variability. (In theory, $\gamma(h=0)$ should be zero.) Here, a is called the "range", and this scale factor determines the threshold distance at which $\gamma(h)$ stabilizes (i.e., flattens). c_0+c_1 is the maximum $\gamma(h)$ value, called the "sill", with c_1 referred to as the "partial sill" (Cressie, 1993). Figure 1 illustrates these parameters.

It is simplest to estimate the three shape parameters using Ordinary Kriging (much like non-linear least squares, when all $Z(s)$ values enjoy the same mean. In Universal Kriging, the vector of parameters β needs to be estimated simultaneously or iteratively (in synch with c_0, c_1 and a) because

$$
\begin{aligned}
&E[(Z(s_i)-Z(s_j))^2] \\
&\quad = \mathrm{var}\,[(Z(s_i)-Z(s_j)] + (\mu(s_i)- \mu(s_i))^2 \\
&\quad = 2\gamma\,(s_i-s_j) + \sum_k \beta_k^{\,2}\left(x_k\left(s_i\right)-x_k\left(s_j\right)\right)^2
\end{aligned}
\tag{6}
$$

One approach is to use a series of feasible general least square (FGLS) regressions to estimate β and $\sum$ iteratively, where $\sum$ indicates the covariance matrix of error terms (ε) and is a function of a, c_0 and c_1. Starting from an initial value of β (e.g., β=0), residuals are $e=Z-X\beta$ (where X is the matrix of explanatory information across all sites). Using these values, one can estimate the variogram's three parameters to get an estimate of $\sum$. The updated estimate of β then is $\beta = (X'\Sigma^{-1} X)^{-1} X\Sigma^{-1}Z$. This process repeats until all estimators stabilize/the system converges (which is guaranteed by the algorithm).

Another approach is to use restricted maximum likelihood estimation (REML) by assuming the errors follow a normal distribution, so that the data set's log-likelihood enjoys the following proportionality:

$$LL \propto -|\Sigma|^{-1/2}[(Z-X\beta)'\,\Sigma^{-1}(Z-X\beta)] \tag{7}$$

This likelihood can be maximized with respect to all unknown parameters using Newton-Raphson or other optimization techniques. The real challenge is the calculation of distance: currently, all available packages use Euclidean distances, which can be easily derived based on locations of the sites. The computational burden can increase dramatically if non-Euclidean distances are used and sample size is large. Nevertheless, with efficient algorithms for network-distance calculations (or patience on the part of the analyst), such extensions certainly are feasible.

DATA SET DESCRIPTION

The AADT data used here was generated for Texas using a combination of permanent (ATR) and a subset of the state's SPTCs, referred to as urban-area saturation traffic counts. There are roughly 240 ATR sites in Texas, and their locations are generally selected based on coverage need for the roadway functional classification. Other factors that can influence site selection include power and phone line availability, level of congestion, road design, and requests by local district officers. These ATRs are operated continuously, but due to construction and communication problems, some operations are often discontinued, leading to approximately 200 usable sites each year.

In addition, 75,000 to 95,000 short period traffic counts (SPTCs) are collected in rural and urban areas of Texas annually. Among these, about 28,000 urban "saturation" count locations are visited on staggered five-year cycles, and their resulting AADT estimates were provided to the research team. These 28,000 saturation-count sites are monitored for 24 hours to obtain a day's traffic count, Monday through Thursday, when school is in session. As mentioned earlier, AADT values for such sites are estimated by applying day-of-week, month-of-year and other factors, developed on the basis of values from groups of similar ATR sites. According to previous studies (e.g., Gadda et al. [2007] and Granato [1998]), this factoring process results in average estimation errors of approximately 12% to 25%, depending on data quality and site type. Thus, the data used here are measured with error. Errors introduced by the factoring algorithm may obscure errors produced by the temporal and spatial extrapolation in this analysis[1]. Moreover, not every saturation site offers a complete panel of estimated AADT values; for the period 1999 through 2005, 27,738 are complete, and thus used here. Table 1 provides some descriptive statistics of these traffic counts over different years, and these data are the basis for the temporal forecasting and spatial interpolation pursued here, as described in the following sections.

Temporal Extrapolation

Traffic counts at each site over these seven years follow a nearly linear trend. The magnitude or slope of the change, however, varies significantly across sites. In order to reasonably extrapolate future traffic counts, each site was analyzed using ordinary least square (OLS) regression based on all seven years' traffic counts. Assuming counts change linearly with time, two parameters can describe the equation for counts at each site. These are the slope and the average count (which together provide an intercept term). Figure 2(a) is a histogram of all slope estimates for the 27,738 sites. Due to the long right-side tail, the horizontal axis uses a logarithmic scale, and traffic counts at most sites increase by around 100 each year. For the 6,829 sites experiencing volume reductions, the most common reduction in daily count per year is 10 vehicles.

Figure 2(b) shows the distribution of the traffic count averages (over 7 years) also with a logarithmically-scaled horizontal axis. As shown, most sites average around 6,000 vehicles/day. To give a sense of relative change in counts, Figure 2(c) is a histogram of slope-to-mean count values. After mean count values and slopes are calculated, traffic counts at these saturation sites in future years can be extrapolated, assuming that the changes in traffic volumes follow a linear pattern. As an example, Figure 3 provides year-2006 predicted values for all 27,738 sites.

Spatial Interpolation of AADT Data

There are nearly 80,000 centerline miles of state-maintained highways Texas, and another 144,000 in the form of county roads. According to the Texas Almanac (Alvarez 2006), total street and roadway miles sum to roughly 300,000. Working with the smallest of these three values, the 27,738 urban saturation counts imply roughly one count site every 2.9 centerline miles of state-maintained highway (but sampled just once every five years). The AADT estimates between those traffic counts sites are often directly given as the AADT value of the closest site with traffic count detector. For urban areas, it is a very strong assumption that the AADT value keeps constant on the section of 2.9 miles long road. It is more reasonable to believe that the AADT values should be some function of AADT values at nearby sites. To this end, ordinary Kriging is used here, for strategic spatial interpolation of AADT values. How many total centerline miles in TX that are getting counted?

While traffic count predictions at each site can be enhanced by controlling for variables like number of lanes, speed limit, and functional class, such information was unavailable; so only

---

[1] Of course, the initial factoring process does remove much variability in the data (due to day of week, month of year, year of sample, and other effects), so it is a recommended approach when only 24-hour counts are available for a network, most acquired on different days of the year, in different years.

general location details are used here. Table 2 summarizes all variables across TxDOT's 25 districts, illustrating how different districts across Texas have experienced very different rates of traffic growth. Dallas, Bryan (home to Texas A&M and College Station) and Corpus Christi are among the fastest growing. The maximum values of traffic occur on US 59 in Houston, where the 350,000 AADT values imply that each of the 12 lanes averages 1,200 vehicles per hour, every hour of every day.

To improve location information for purposes of AADT estimation, this study obtained network details based on the TxDOT-provided network file. This was done by joining road links (in a vector layer) to the 27,738-site point layer. Each site was matched to attributes of the closest ESRI-network road section, providing a variable of functional class. Among these, the two most commonly represented classes are interstate highways (Class 1) and other principal arterials (Class 2).

Traffic counts on segments of the same class were spatially interpolated using Kriging. While using network distances interpolating spatial dependencies between locations along a network is behaviorally most reasonable, it is far more computationally intensive at the outset. (For example, the distances would have to be calculated 24,031x24,031 times for Class 1 road segments in the saturation count data set, and 3,707x3,707 times for the Class 2 segments.) Fortunately, existing studies (e.g., Hoef et al. [2006] and Kruvoruchko and Gribov [2004]) on small networks find that using Euclidean distances can yield satisfactorily results, even when the dependencies arise over a network. To ensure computational tractability, this work relies on Euclidean distances. The following two examples show how Class 1 and Class 2 road segment AADTs are estimated.

Different semi-variogram specifications were estimated and compared for each class of roadway, and the exponential form (Eqn. 6) was selected for both, thanks to its better fit. For Class 1 segments, the estimated range value (a) is 1.248, partial sill (c_1) is 1.62E7 and nugget value (c_0) is 2.33E7. For Class 2 segments, range (a) is 0.158, partial sill (c_1) is 2.82E9 and nugget (c_0) is 9.86E8. The fitted prediction lines are shown in Figure 4.

As expected, the variance (hence semivariogram) increases with distance, and becomes stable after a certain distance threshold. As compared to Class 2, Class 1 scatter is higher for a given distance, and becomes stable after a comparatively longer distance. The data also suggest that the variability increases with distance, as is typical in other studies, such as ESRI (1996). The larger sill and nugget values of the Class 1 suggests that the spatial autocorrelation of Class 1 AADT is more distance-dependent. Its wider range also suggests that Class 1 autocorrelation is more distance sensitive. One possible reason for this difference is that Class 1 roads are more physically connected than Class 2 roads. In other words, the access points along Class 1 roads may lead to significant fluctuations in AADT over space, while flow changes along Class 2 roads is likely to appear more continuous in nature.

As expected, of course, relying on a series of seven past values to predict the next year's count at any given site results in better prediction. Temporal dependence is "stronger" than spatial dependence, for a variety of reasons. Kriging methods are therefore most useful in uncounted/unsampled locations, or where past counts are getting too old to be reliable.

Assessing Goodness of Fit

In order to validate the Kriging method, this study used 80% of the Class 1 observations to spatially interpolate AADT values and compare these estimates to the actual AADT values for the remaining 20% of observations. Differences in these values were evaluated using the following error ratio indicator:

$$Error_i = \frac{AADT_{est,i} - AADT_i}{AADT_i} \tag{8}$$

Figures 5 and 6 provide the spatial distribution and histogram of these error ratios. As shown in Figure 8, the error ratio exhibits no clear spatial trends. The simple median of these errors is 0.333 (or 33%). Apparently, the spatially interpolated values tend to over-estimate AADT by about 33% – which is sizable. Figure 7 indicates that this overestimation is mainly caused by locations with very low AADT values for which Kriging methods are less reliable – in part because they do not recognize/respect a zero-count boundary. The performance is significantly better for locations with traffic counts higher than 1,000 vpd. However, when the AADT values get too high, their values tends to be under-estimated. When errors are weighted by AADT values, the average bias switches to a negative 31%. When weighted by the quarter power of AADT (i.e., $AADT^{0.25}$), the median weighted error is 0.018 (or 1.8%). Given the fact that only a limited number of influential factors are considered, such a bias indicates that this spatial interpolation method yields fairly reliable estimation results, especially for road sections with moderate to high traffic volumes.

Related to all this, Gadda et al. (2007) tested the spatial error of AADT estimation by assigning the AADT from its nearest sampling site, which is the standard technique for no-information AADT estimation. They found that the average error rises significantly with the number of lanes. Error patterns also vary by general location (using a rural/urban indicator) and functional class. As expected, such estimates of error also rose with distance between the unknown-count site and its neighbor site – rather dramatically. They averaged just 6.33% at 0.2 miles away to 79.5% at just 1.6 miles, eclipsing the 33% and 31% figures noted here (which are at an average distance of 2.9 miles away).

The comparison of their study and this one first suggests that estimation may significantly improve when other variables are controlled for. Moreover, when no other explanatory information is used, the Kriging method is capable of yielding more reliable estimations than simply relying on the count of a location's nearest count location, given the fact that the average distance between the Class 1 sites is around 2.9 miles.

CONCLUSIONS

This study describes a rather straightforward process that can be used to estimate real-time and longer-term traffic counts throughout a network based on limited information, using ArcGIS Kriging tools. Given AADT estimates at over 27,000 urban saturation traffic count stations, AADT values are first forecast for future years and then spatially interpolated for sites without detectors. Spatial interpolation using ordinary Kriging methods suggests that traffic volumes on different classes of roadways exhibit rather different patterns of spatial autocorrelation. Furthermore, model validation suggests that, while relatively uninformed Kriging tends to overestimate AADT values, the bias is still tolerable – particularly in cases where no earlier counts exist at that site or very nearby. Such methods can be applied to many other aspects of transportation studies, as long as they exhibit some sort of spatial dependence; these include variables like housing price, trip generation rates, pavement conditions, and crash rates.

Of course, the data sets used here can be dramatically enhanced by those who have direct access to such data, and the methods used here can be further refined (particularly when coding the routines directly, rather than relying on ESRI commands). Both temporal extrapolation and spatial interpolation stand to benefit from the inclusion of standard explanatory information (e.g., number of lanes and speed limits). And reliance on network (rather than Euclidean) distances between all count locations may be helpful (particularly for small data sets). New code for temporal and spatial extrapolation in the presence of heteroskedasticity while controlling for various explanatory factors will make a valuable extension of this work. Nevertheless, the predictions in the study make effective use of temporal and spatial information in existing data sets. These predicted values can be used as estimates of traffic conditions at unmonitored sites in any year, facilitating system management, data analysis and investment decisions.

ACKNOWLEDGEMENTS

This work was financially supported by Texas Department of Transportation (TxDOT) research project 0-5686, “Utilizing the Data Collected at Traffic Management Centers for Planning Purposes through Non-Traditional Sources and Improved Equipment.” The authors are grateful to those who provided kindly help for this study. These include Shannon Crum, Rhonda Christiansen and Michael Chamberlain at TxDOT, as well as Annette Perrone’s administrative and editing assistance.

REFERENCES

Alvarez, Elizabeth Cruce (ed.), 2006, *Texas Almanac 2006-2007*. College Station: Texas A&M University Press Consortium.

Bayraktar, H. and Turalioglu, F. S., 2005, “A Kriging-Based Approach for Locating A Sampling Site - In The Assessment of Air Quality.”, *Stochastic Environmental Research and Risk Assessment* 19(4): 301-305.

Cressie, N., 1993, *Statistics for Spatial Data*, Wiley Interscience: New York.

Gadda, S., Kockelman, K., and Maggon, A. 2007, “Estimates of AADT: Quantifying the Uncertainty.” Presented at the World Conference on Transportation Research, Berkeley, California. (Available at http://www.ce.utexas.edu/prof/kockelman/public_html/TRB07AADTUncertainty.pdf.)

Emery, X., 2005, “Simple and Ordinary Kriging Multigaussian Kriging for Estimating Recoverable Reserves.” *Mathematical Geology* 37(3): 295-319.

ESRI, 1996, *ArcView Spatial Analyst.* Environmental Systems Research Institute.

Eom, J.K., Park, M.S., Heo, T.Y. and Huntsinger, L.F., 2006, “Improving the Prediction of Annual Average Daily Traffic for Nonfreeway Facilities by Applying a Spatial Statistical Method.” *Artificial Intelligence and Advanced Computing Applications* 1968: 20-29.

Granato, S. (1998) The Impact of Factoring Traffic Counts for Daily and Monthly Variation in Reducing Sample Counting Error. *Proceedings of the Crossroads 2000 Conference* (Ames, Iowa), pp.122-125. Accessed July 2006 at http://www.ctre.iastate.edu/pubs/crossroads/122impact.pdf

Hoef, J., Peterson, E., Theobald, D., 2006, “Spatial Statistical Models that Use Flow and Stream Distance.” *Environmental and Ecological Statistics* 16: 449-464.

Krige, D.G, 1951, *A statistical approach to some mine valuations and allied problems at the Witwatersrand*, Master's Thesis of the University of Witwatersrand.

Kruvoruchko, K. and Gribov, A., 2004, “Geostatistical Interpolation and Simulation with Non-Euclidean Distances.” *geoENV IV*, Eds. Xavier Sánchez-Vila, Jesús Carrera, and J. Jaime Gómez-Hernandez. Kluwer Academic Publishers.

Lam, W.H.K., Tang, Y.F., Chan, K.S. and Tam, M.L., 2006, “Short-Term Hourly Traffic Forecasts using Hong Kong Annual Traffic Census.” *Transportation* 33(3): 291-310.

Matheron, G., 1963, “Principles of Geostatistics.” *Economic Geology* 58: 1246-1266, 1963.

Tang, Y.F. Lam, W.H.K. and Ng, P.L.P., 2003, “Comparison of Four Modeling Techniques for Short-Term AADT Forecasting in Hong Kong.” *Journal of Transportation Engineering-ASCE* 129(3): 271-277.

Zhao, F. and Chung, S., 2001, “Contributing Factors of Annual Average Daily Traffic in A Florida County - Exploration with Geographic Information System and Regression Models.” *Pavement Management, Monitoring, and Accelerated Testing* 1769: 113-122.

Zimmerman, D.A., Marsily, G. D., Gotway, C. A. and Marietta, M.G., 1998, “A Comparison of Seven Geostatistically Based Inverse Approaches to Estimate Transmissivies for Modelling Advective Transport by Groundwater Flow.” Water Resource Research 34(6): 1273-1413.

LIST OF TABLES AND FIGURES

Table 1. 24-HourTraffic Counts

Year	Number of Sites	Minimum Count	Maximum Count	Mean Count	Standard Deviation
1998	27,914	10	361,890	7004.8	17,748
1999	27,980	10	328,000	7437.7	19,405
2000	28,075	10	337,000	7709.4	19,961
2001	28,271	10	337,000	7928.0	20,626
2002	28,285	10	349,000	8107.8	21,309
2003	28,304	10	345,000	8225.4	21,549
2004	28,303	10	335,000	8429.1	21,859
2005	28,278	10	341,940	8702.2	22,723
Over all years	27,738	10	349,000	7943	20,647

Table 2.Patterns of Change for Urban Saturation Count Sites AADT Values by Districts

District of TxDOT	Number of SPTC Sites	Change in 24 Hour Traffic Count/Year (Slope)	Average of AADT Estimates	Ratio of Slope to Mean Value (%)
Abilene	1075	33.6	2,518	1.085%
Amarillo	1141	50.8	3,243	1.094
Atlanta	1209	82.2	4,336	1.399
Austin	1258	348.6	12,842	2.135
Beaumont	952	103.1	6,612	0.907
Brownwood	877	17.4	2,189	0.115
Bryan	1192	172.5	4,901	3.197
Childress	713	17.3	1,167	0.454
Corpus Christi	1181	201.1	5,868	3.185
Dallas	1433	535.9	18,452	3.516
El Paso	386	530.4	14,721	1.234
Fort Worth	1278	396.9	13,860	2.582
Houston	1246	729.5	20,769	2.671
Laredo	506	67.9	3,399	1.351
Lubbock	1661	39.4	2,493	1.062
Lufkin	1336	34.2	3,696	1.060
Odessa	728	51.3	3,417	0.738
Paris	1405	62.2	3,373	1.591
Pharr	1019	367.0	9,181	2.971
San Angelo	686	37.3	2,126	1.037
San Antonio	1425	375.0	11,986	2.512
Tyler	1697	82.7	4,910	1.551
Waco	1392	224.3	6,334	2.582
Wichita Falls	962	72.3	3,128	1.362
Yoakum	1511	66.9	3,630	1.910

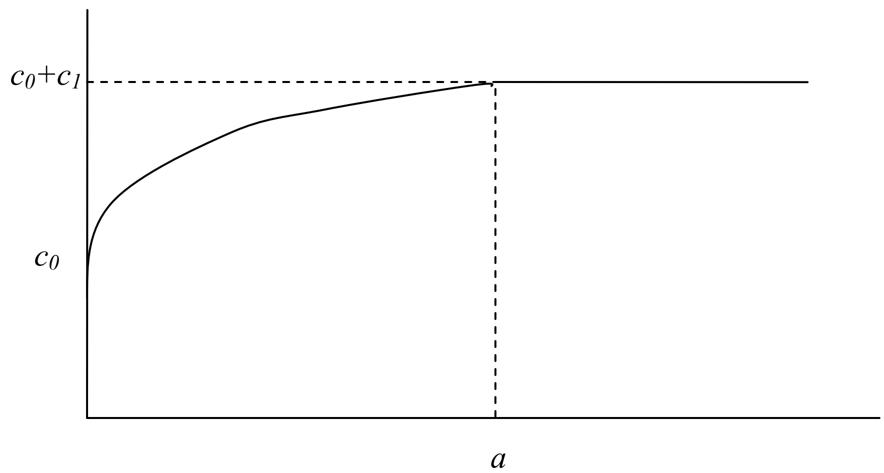
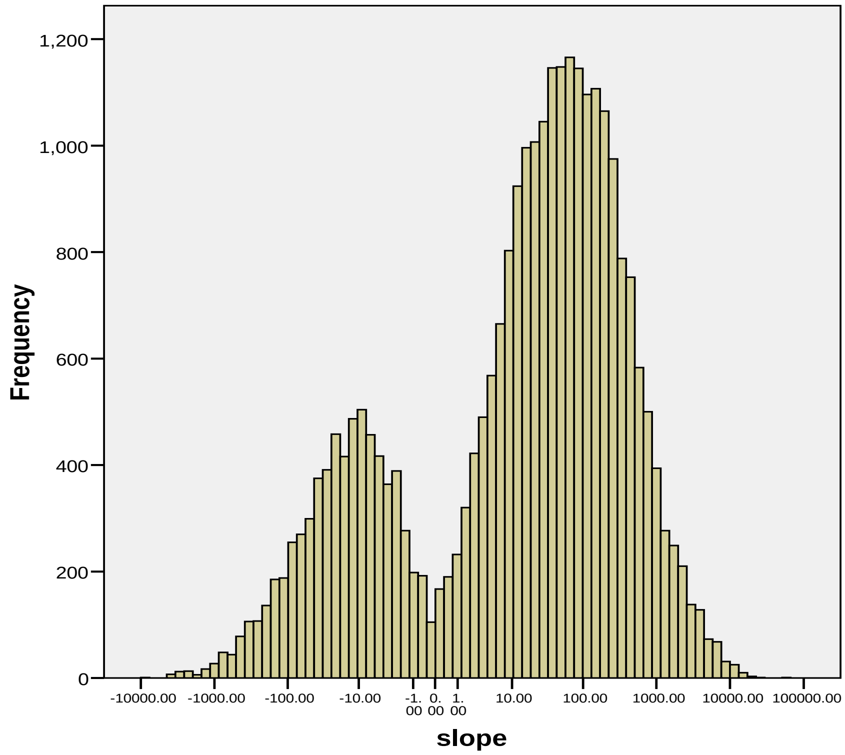
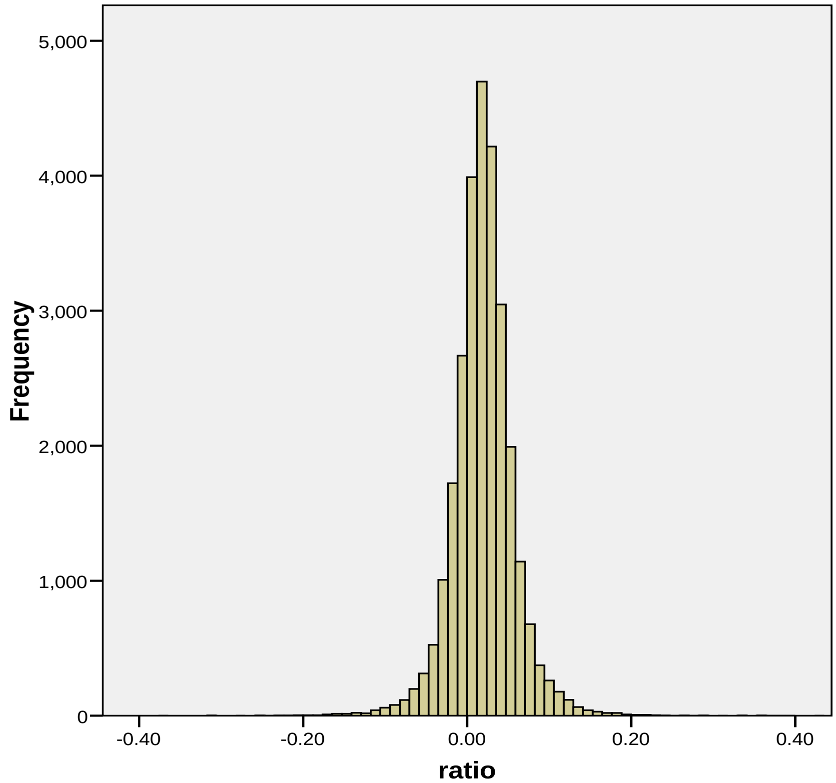
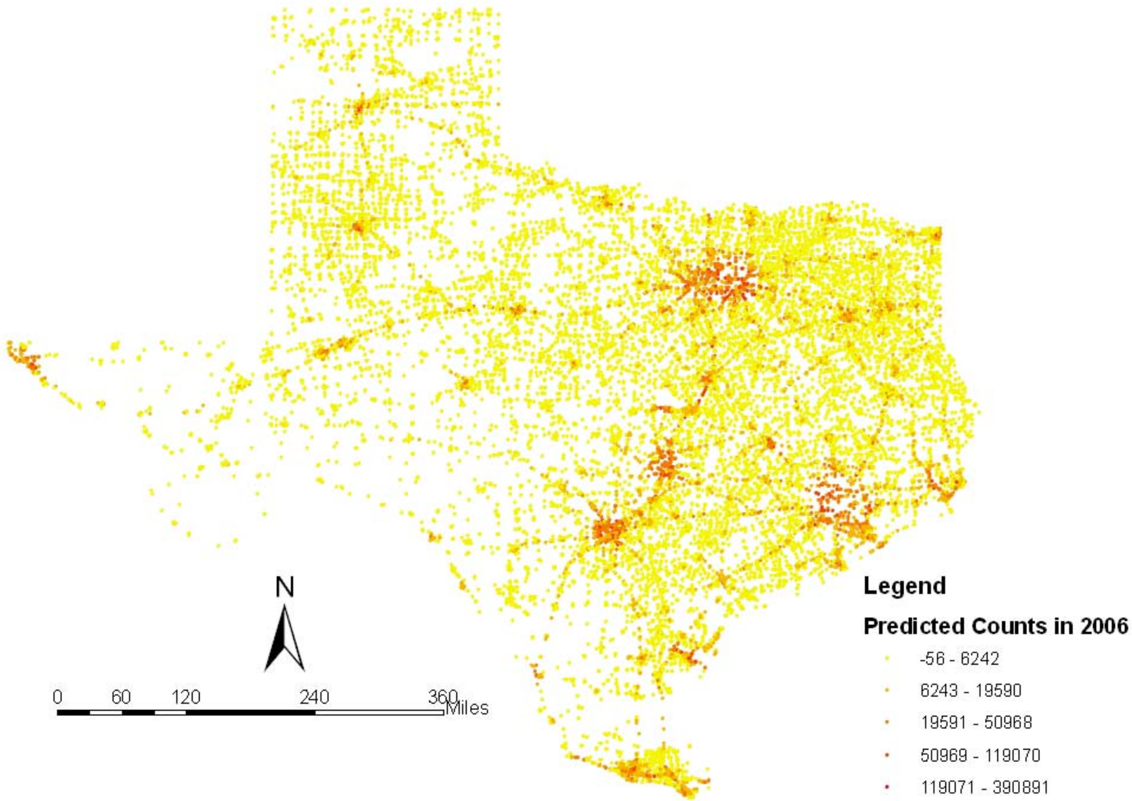


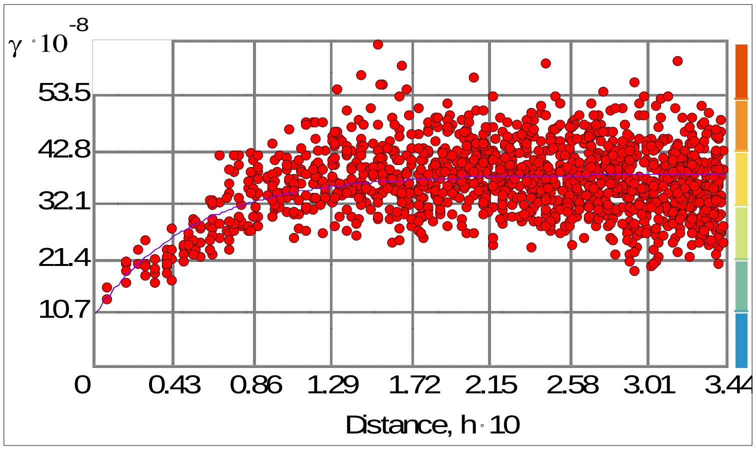
Figure 1. Illustration of Semivariogram


Frequency
1,200
1,000
800
600
400
200
0
-10000.00
-1000.00
-100.00
-10.00
-1.00
0.00
1.00
10.00
100.00
1000.00
10000.00
100000.00
slope



(c). Distribution of Slope-to-Mean Count Values (Relative Change)
Figure 2. Distribution of Parameters for 24 Hour Traffic Counts across SPTC Sites


N
0
60
120
240
360
Miles
Legend
Predicted Counts in 2006
-56 - 6242
6243 - 19590
19591 - 50968
50969 - 119070
119071 - 390891



(b). Class 2 Segments

Figure 4. Semivariogram Fitting for AADT

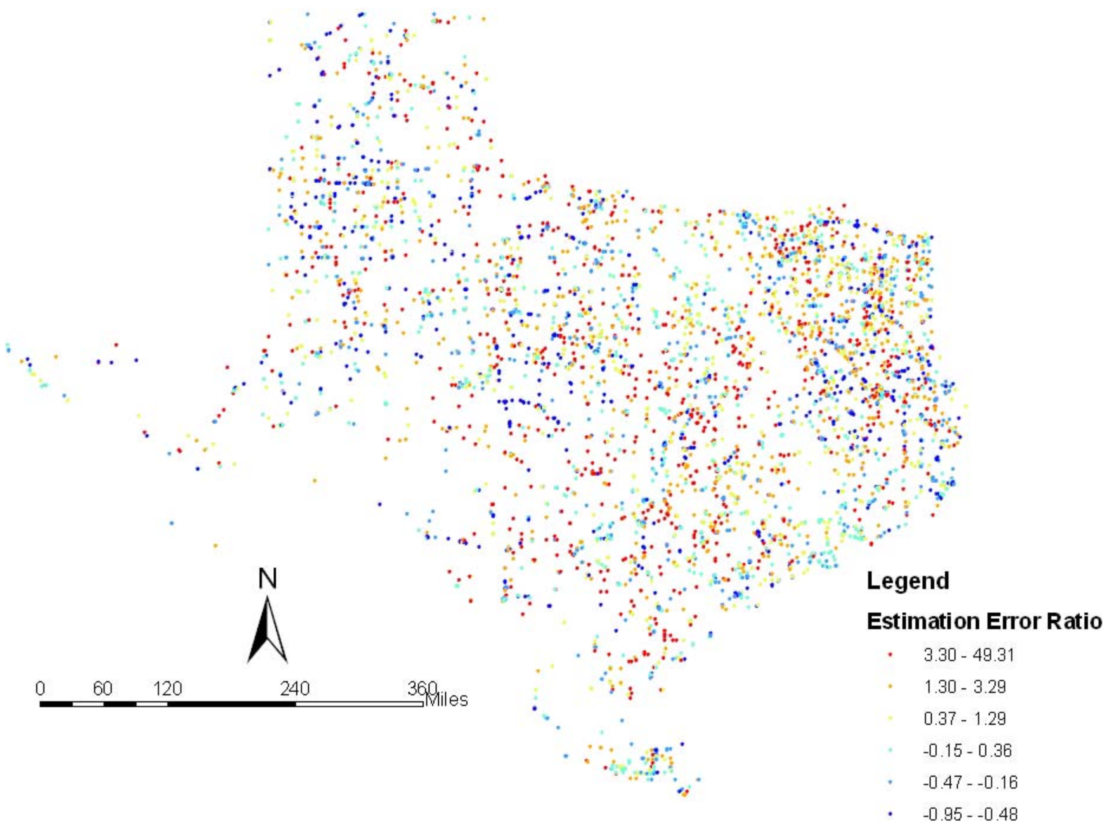
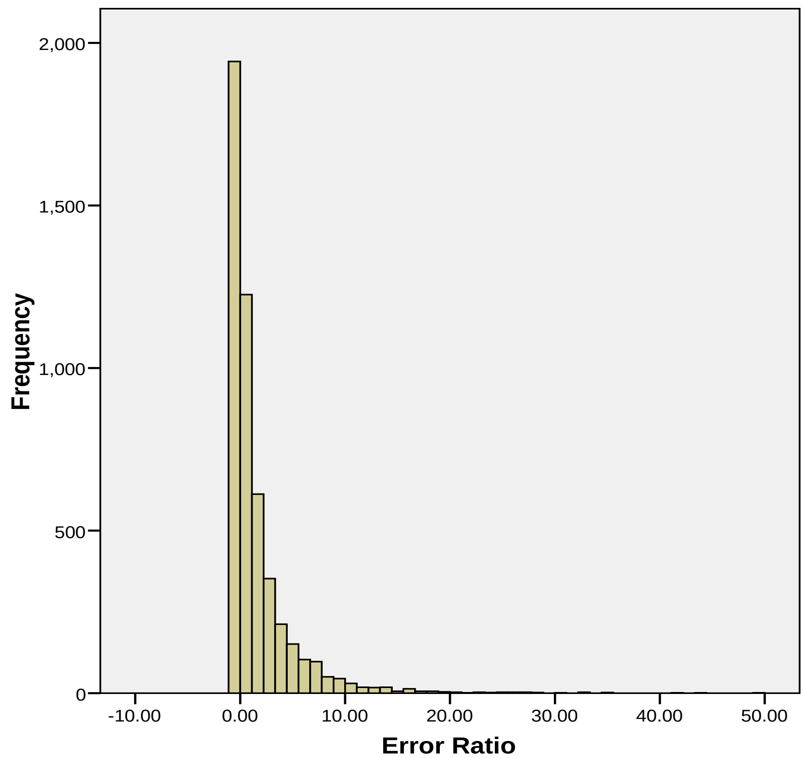


Figure 5. Differences between Kriging Estimates and Observed Traffic Counts
(Using TxDOT's road network, and the 1983 North American Global Coordinate System)



2,000
1,500
1,000
500
0
Frequency
-10.00
0.00
10.00
20.00
30.00
40.00
50.00
Error Ratio

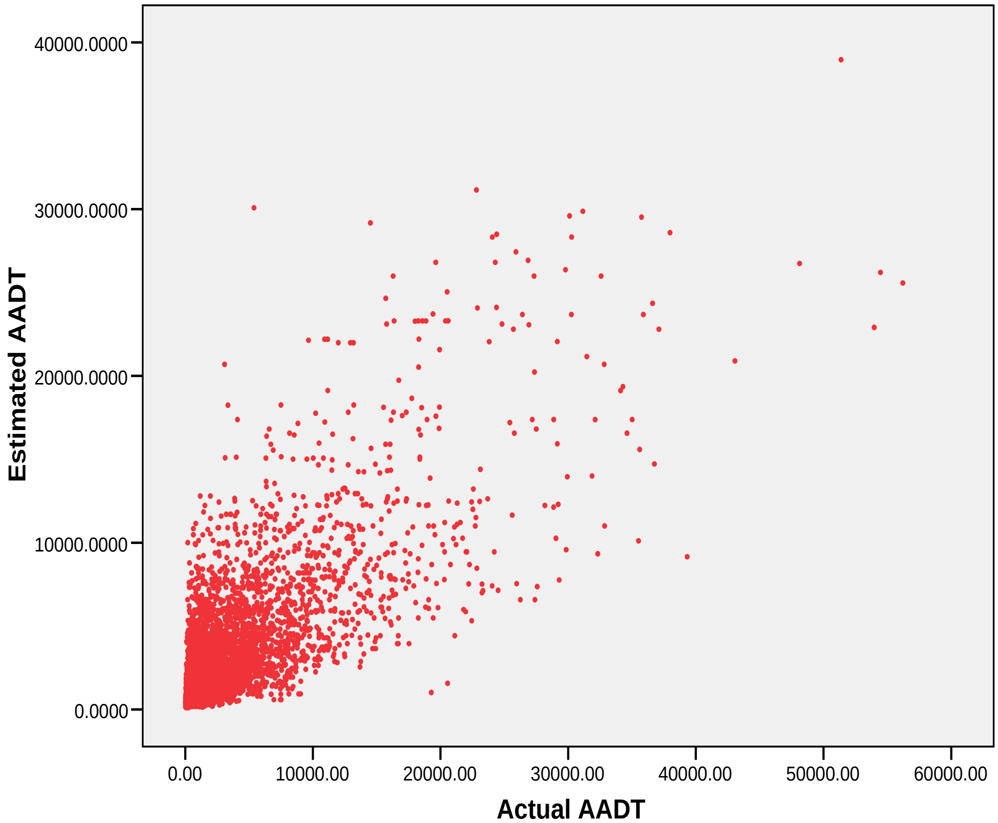


Figure 7. Relationship between Kriging Estimates and Actual Traffic Counts