Estimation of Annual Average Daily Traffic in a Florida County Using GIS and Regression

Fang Zhao, Lehman Center for Transportation Research, Department of Civil & Environmental Engineering, Florida International University, 10555 W. Flagler Street, Miami, FL 33174. Phone: (305) 348-3821, Fax: (305) 348-2802, Email: fang@eng.fiu.edu.

Soon Chung, Lehman Center for Transportation Research, Department of Civil & Environmental Engineering, Florida International University, 10555 W. Flagler Street, Miami, FL 33174. Phone: (305) 348-4103, Fax: (305) 348-2802, Email: schung01@fiu.edu.

ABSTRACT

Accurate Annual Average Daily Traffic (AADT) information is important to many applications including roadway design, air quality compliance, travel model validation, etc. Yet complete or even extensive coverage of a network to collect traffic count data is impractical due to the cost involved. There have been studies to estimate AADT for interstate highways, expressways, urban roads, and rural roads. Such efforts are often hampered by the lack of relevant detailed information that adequately explains variations in the traffic counts. In this paper, a multiple linear regression model of AADT on local roadways in a large urban area is presented. By employing geographic information system (GIS) technology, a variety of land use and accessibility measurements are developed and tested. While most are statistically significant, few add enough explanatory power to be practical and useful. Four models are presented, which achieved R^2 of 0.66 to 0.82. The predictive power of the models are also examined.

INTRODUCTION

Estimation of Annual Average Daily Traffic (AADT) is extremely important in traffic planning and operations for the state departments of transportation, because AADT provides information for the planning of new road construction, determination of roadway geometry, congestion management, pavement design, safety considerations, etc. Accurate AADT data are crucial to the calibration and validation of travel demand models. AADT is also used to estimate state-wide vehicle miles traveled on all roads and is used by local governments and the environmental protection agencies to determine compliance with the 1990 Clean Air Act Amendment. Additionally, AADT is reported annually by the state departments of transportation to the Federal Highway Administration.

While all local and state governments collect traffic counts, due to budgetary constraints, the data coverage is often limited, especially for local roads or in rural areas. Typically, when count data are unavailable, estimates are made based on comparisons to roads that are considered to be similar. Such comparisons are inherently subject to large errors, and also may not be repeated often enough to remain current.

Literature on estimating AADT for roads that do not have traffic counts is limited. Most of the relevant literature is on either estimating traffic volume for freeways or extrapolation of 24-hour or 48-hour traffic count data to obtain AADT. One attempt to estimate AADT for county roads was made by a research group at Purdue University, Indiana, which developed a multiple regression method utilizing aggregated data at the county level (*1*). Data from 89 count stations were used for predicting AADT in 40 counties in Indiana. Four predictors are chosen to estimate AADT. They include county population, location type (urban/rural), access to other roads, and total arterial mileage in a county. A R^2 of 0.75 is achieved.

Several other studies have been reported on estimating AADT for state roads using statistical methods (*2, 3, 4, 5, 6*). The predictors include county population size, total number of through lanes, state/non-state code indicating jurisdictions, location type (rural/urban), personal income level, vehicle registrations, etc. A common problem reported was the lack of data for some potentially important predictors. For example, in a Minnesota study, it was proposed to use the population within a certain distance of the roadway as a predictor, but the attempt was abandoned due to unavailability of the data (*2*).

Xia et al. (7) developed a regression model for estimating AADT for non-state-owned roads in Broward County, Florida. The predictors used in the model include function classification, number of lanes, area type, auto ownership, presence of non-state roads nearby, and service employment. The model has an adjusted R^2 of 0.5961. The prediction errors range from 1.31% to 57%. 50% of the test points has an error smaller than 20%, while 85% of the test points has an error smaller than 40%. The model also underestimates the AADT for about 5% for the entire test data set.

The AADT model for Broward County was later modified (8). The model does not include the service employment. The adjusted R^2 is slightly higher at 0.6069. The prediction errors range from 0.57% to 61.99%. 47.2% of the predictions has errors smaller than 20%, while 83.3% has errors smaller than 40%.

The study presented in this paper is a continued effort following the aforementioned two studies. The main differences between this and the previous studies are the use of a larger data set that also includes all the AADTs for state roads, replacing the old state roadway function classification system with the new federal function classification system, and more extensive analysis of land use and accessibility variables. Several multiple regression models are developed and compared. In the remainder of this paper, an overview of the characteristics of the study area is provided first. The data used in the model, spatial analyses performed to derive land use and accessibility measurements, and model development are then described. Finally, the results are discussed and conclusions provided.

CHARACTERISTICS OF THE STUDY AREA

The study area is comprised of the entire Broward County in South Florida (see Figure 1). The county is situated between Miami-Dade County and Palm Beach County. The population of the county is approximately 1.4 million, while the population of the tri-county area is about 4.5 million. The development pattern in the county is similar to those in many coastal areas: the central business district (CBD) is located near the coast; intense economic developments occur in the I-95 corridor (which also include the CBD) near the coastal line across the entire county length; and a gradual decrease in economic activity intensity and an increase in typical suburban residential developments west of the I-95 corridor. The county borders the Everglades National Park, its development limit, in the west. The areas near the Everglades have not been fully developed. The maps showing employment and population distribution are given in Figure 2 (areas uninhabited are not shown).

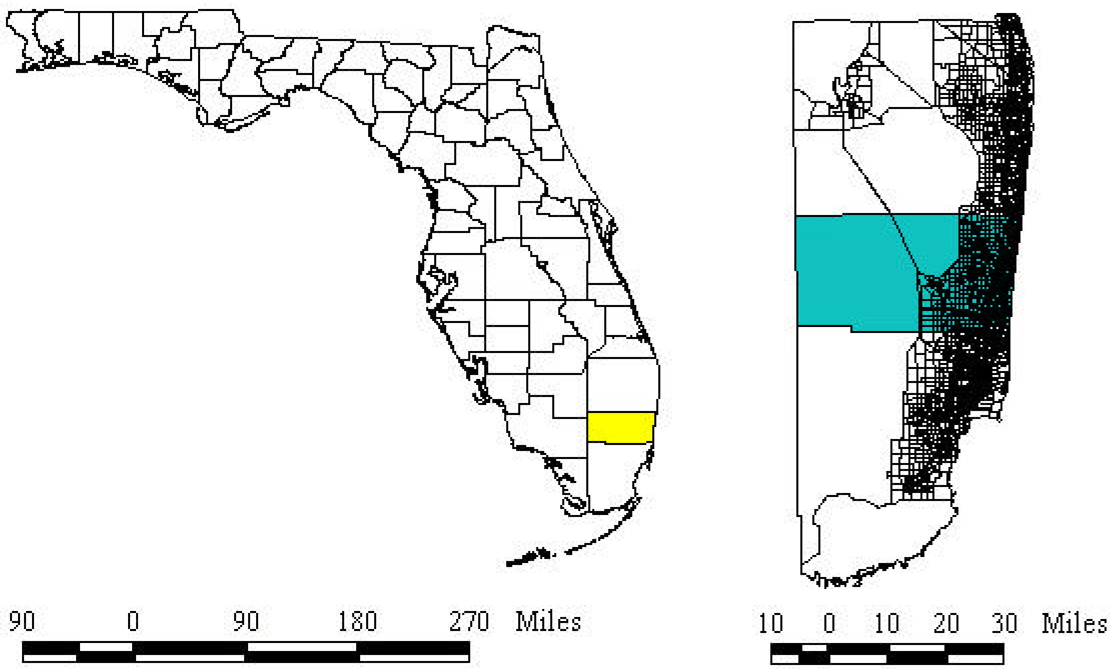

FIGURE 1 Florida State and Broward County

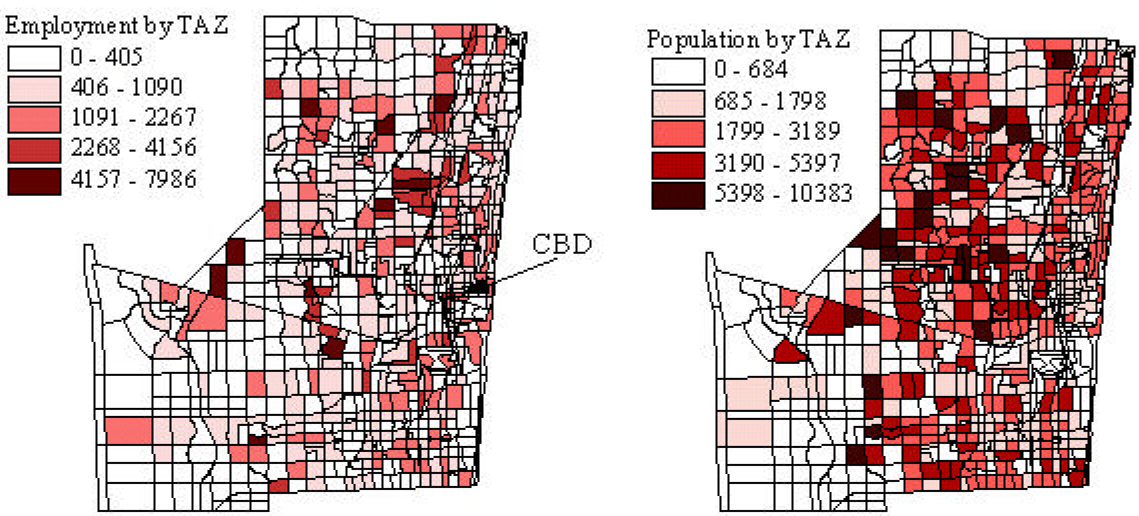

FIGURE 2 Broward County Employment and Population Distribution

DATA COLLECTION AND SPATIAL ANALYSIS

AADT Data

The response variable in the regression models is AADT in thousands. AADT data are obtained from average quarterly traffic counts in 1998 from the Broward County Metropolitan Planning Organization. The counts are adjusted by seasonal factors developed based on traffic data obtained from a number of permanent count stations on state roads. There are four seasonal factors, each applied uniformly to one of the four areas (east, I-95, central, and west). After excluding expressway and zero AADTs, there are 898 data points, 816 (90.9% of the database) of which are used for model development, and 82 (9.1% of the database) for model validation and testing. Of the 816 count stations, 184 are on principal arterials, about 278 on minor arterials, 305 on collectors, and approximately 49 on local roads. These two sets of data are selected randomly from the entire AADT database. The roadway function classes and AADT at various count stations are shown in Figure 3.

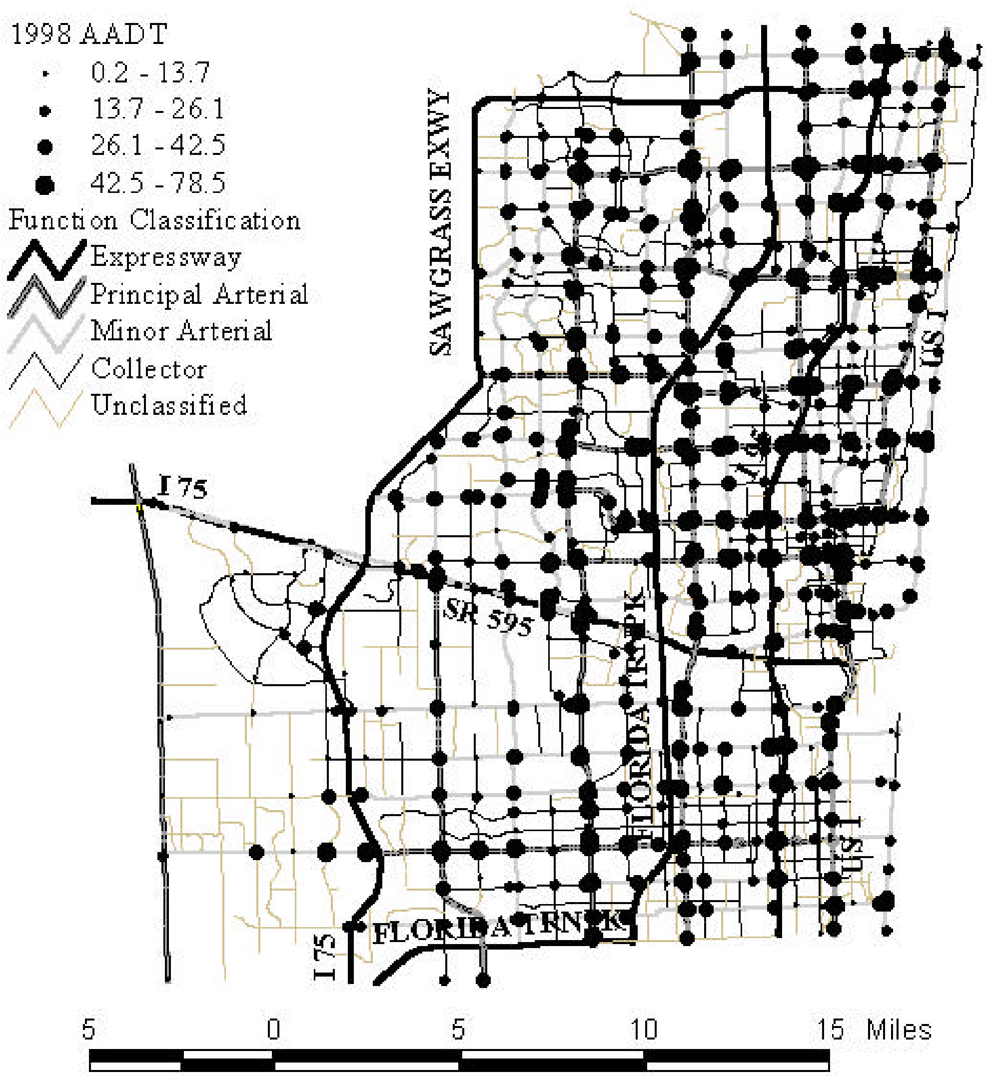

FIGURE 3 Roadway Function Classes and AADT

Initial Predictors

Four groups of predictors, or independent variables, are examined for potential inclusion in the models: roadway characteristics, socioeconomic characteristics, expressway accessibility, and accessibility to regional employment centers. These variables are described below.

Roadway Characteristics Data

These data are related to the characteristics of the roadway sections and include:

1. Number of lanes (*LANE*): the number of lanes on a roadway in 1998;

2. Area type (*AREA*): land use type that is one of the following:
   Central Business District (CBD) = 5.17
   CBD Fringe = 3.16
   Residential = 3.24
   Outlying Business District = 5.63
   Rural Area = 1.65
   Undefined = 1.0
   The original nominal values are replaced by the above numeric values determined based on the mean values of AADT in each area group.

3. Function classification (*FCLASS*): four function classes are used, which are based on the federal function classification system. Their numeric values are determined based on the mean values of AADT in each function class group:
   Urban principal arterial = 3.4
   Urban minor arterial = 2.2
   Urban collector = 1.0
   Unclassified = 0.6

The data for the above predictors are provided by Broward County Metropolitan Planning Organization (MPO) as a GIS layer.

The function classification has the highest correlation with AADT (0.84), followed by number of lanes (0.77). Area type also has a reasonably high correlation with AADT (0.49). Preliminary regression analysis shows that its partial R^2 is only 0.002. Also considering that its definition is rather subjective, this variable is not selected for inclusion in the final regression models.

Socioeconomic Characteristics

These variables reflect the socioeconomic characteristics surrounding a count station. From previous studies (7, 8), it has been determined that population and employment (total and by categories) aggregated within a buffer of a fixed size around a count station are not significant variables. For this study, several new socioeconomic indicators are developed and tested, which are described below.

1. Employment size in a corridor near a group of count stations. To capture the effect of commercial developments in a corridor, count stations that are near each other are grouped

together and a common point on the network (usually an intersection) is created to represent their location. Next, network paths of one-mile (1.61-km) length are created originating from such common points. A 100-foot (30.5-meter) buffer is then generated around the network paths and used to aggregate the employment in the corridor. The employment data used are for the year of 1999 and are purchased from InfoUSA for a separate project. The database contains the number of employees at each business location and the SIC code. Employers are geocoded when possible. The data provide more accurate locational information on the employment sites than aggregate data based on TAZs. This indicator, however, turns out to have a low correlation with AADT and is not used.

2. Employment, population, and total dwelling units around a count station. In another effort to account for the impact of local land use developments on AADT, employment, population, and dwelling units are aggregated using a variable-sized buffer around each count station. The employment data are again the 1999 InfoUSA data. The data on population and total dwelling units are aggregate at TAZ level. The 1998 data are obtained by interpolating 1995 and 2000 population and dwelling units forecast by the Broward County Department of Strategic Planning and Growth Management.

   The size of the buffer is based on count station location and the function class of the road. The study area is divided into four sub-areas based on the original zones used for seasonal factors and based on roadway density. The buffer sizes reflect the average spacing of roads in a given class of roadways in a particular sub-area. They are considered to be the "service" area of a road. The buffer sizes based on sub-areas and roadway function classes are given in Table 1.

TABLE 1. Buffer Sizes Based on Count Station Location and Function Class

Location	Buffer Size (miles)			
	Principal Arterial	Minor Arterial	Collector	Unclassified
East	1.0	0.5	0.25	0.12
Central	1.5	0.5	0.25	0.12
Central West	2.0	1.0	0.50	0.25
West	3.0	2.0	0.50	0.25

   The three indicators (employment, population, and dwelling units) have a correlation factor of 0.325, 0.240, and 0.252 with AADT, respectively. However, they are highly correlated to each other. Regression results show that the employment indicator contributes to the increase of R^2 the most. Therefore, the employment indicator (*BUFFEMP*) is used and the other two are not included in the final models.

3. Employment and population around a count station aggregated using buffer sizes determined based on functional classification.

Accessibility to Expressways

This group of variables measures whether a road has easy access to expressways. The measurements developed include the following:

1. Minimum distance to expressway from a count station to an expressway access point, which is measured as network distance in miles.

2. Minimum travel time in minutes from a count station to an expressway access point. The travel speed is assumed to be 55 mph (88.5 kmph) for expressways, 35 mph (56.3 kmph) for principal arterials, 30 mph (48.3 kmph) for minor arterials, 25 mph (40.2 kmph) for collectors, and 20 mph (32.2 kmph) for unclassified roads.

3. Number of expressway access points within a four-mile (6.44 km) radius from a count station.

4. Direct Access (*DIRECTAC*), which is a binary variable that assumes a value of 1 and 0. 1 means a count station is on a road that connects to an expressway, and 0 otherwise.

Correlation and regression analyses show that the number of expressway access points close to a count station is not an important factor. The minimum distance between a count station and an expressway access point has a correlation coefficient of 0.20 but again is not an important factor. Direct access (*DIRECTAC*) has a high correlation (0.48) with AADT and is further considered.

Regional Accessibility to Employment

Several measurements are developed in an attempt to model influence of regional economic activities on the traffic of a road. There are five general kinds of measurements:

1. Network distance to the regional mean centers of employment (*DECNTR*). The regional mean center of employment is determined by finding the spatial mean center of all the TAZs weighted by the TAZ total employment. This measure is developed in an attempt to reflect the gradual decrease in development intensity as one moves away from the urban core.

2. Network distance to the regional mean centers of population (*DPCNTR*). The regional mean center of population is determined by finding the spatial mean center of all the TAZs weighted by the TAZ total population. The purpose for this measure is the same as for *DECNTR*.

3. Regional accessibility to population centers defined as follows:

$$RPAccess_k = \sum_{i=1}^{N_P} P_i e^{-0.0954 t_{ki}}$$

Where $RPAccess_k$ is the accessibility of count station k to regional population, P_i is the population of the ith population center, t_{ki} is the network travel time from count station k to the ith population center, and N_P is the number of population centers. To calculate

$RPAccess_k$, 17 population centers are created in Broward County by examining the population data with the aid of GIS visualization tools such as population distribution and density maps, three-dimensional surface maps created from population data by TAZs with population as the z coordinate, and contour maps generated from the surface map.

4. Regional accessibility to employment centers defined as follows:

$$REAccess_k = \sum_{j=1}^{N_E} E_j e^{-0.0954 t_{kj}}$$

Where $REAccess_k$ is the accessibility of count station k to regional employment, E_j is the employment of the jth employment center, t_{kj} is the network travel time from count station k to the jth employment center, and N_E is the number of employment centers. Using the same approach for generating population centers, 16 employment centers in Broward County are created.

5. Regional accessibility to population and employment defined as the product of the previous two accessibility measures:

$$RPEAccess_k = \left(\sum_{i=1}^{N_P} P_i e^{-0.0954 t_{ki}} \right) \left(\sum_{j=1}^{N_E} E_j e^{-0.0954 t_{kj}} \right)$$

The correlation between AADT and *RPACCESS*, *REACCESS*, and *RPEACCESS* are 0.03, 0.27, and 0.20, respectively. Regression analysis further shows that *REACCESS* is a better indicator than *RPEACCESS*. Therefore, only *REACCESS* is chosen for inclusion in the models.

Because Broward County is part of the urbanized tri-county area, there is a significant amount of trip interchanges among the counties. To account for the internal-external (I-E) trips, a separate set of regional accessibility is created that includes the population and employment centers within 10 miles (16.1 km) from the county border in Palm Beach County, and those within 12 miles (19.3 km) in Miami-Dade County, respectively. The 10-mile and 12-mile limits are chosen based on the fact that the majority (over 75 percent) of the I-E trips are either originated or destined within these limits.

Correlation and regression analyses show, however, that the benefit of including the two adjacent counties in the three accessibility measurements is negligible. The correlation coefficients are also low, typically below 0.11. The data processing cost is not offset by the improvement to the model with the other two counties included. Therefore, a set of *REACCESS* is created only using Broward County data and included in the models.

REGRESSION MODELS

Based on preliminary analyses, the following variables are included in the final regression models:

1. Function classification (*FCLASS*)
2. Number of lanes (*LANE*)
3. Direct access from a count station to expressway access points (*DIRECTAC*)

4. Accessibility to regional employment in Broward County (*REACCESS*)
5. Employment in a variable-sized buffer around a count station (*EMPBUF*)
6. Population in a variable-sized buffer around a count station (*POPBUF*)

Using the above six variables, four regression models are generated. Variables are selected using the stepwise procedure. The models are checked for multicollinearity, and outliers are diagnosed. Consequently**,** outliers (about *5.0%*) are detected from the data set for model development and the regressions are redone. All the statistical analyses are conducted using the Statistical Analysis System (SAS) software. The four models have the following forms:

Model 1: $AADT = -9.520386 + 8.480001\ FCLASS + 3.428939\ LANE$
$+ 0.596752\ REACCESS + 2.991573\ DIRECTAC + 0.069086\ BUFFEMP$

Model 2: $ADDT = -6.15742 + 6.55471\ LANE + 0.61433\ REACCESS$
$+ 7.88344\ DIRECTAC - 0.34494\ DISTPOPMCNTR$

Model 3: $ADDT = -4.66034 + 4.95341\ LANE + 0.51119\ REACCESS$
$+ 4.52713\ DIRECTAC - 0.10689\ DPOPCNTR + 0.00112\ POPBUFF$

Model 4: $AADT = -4.26565 + 4.86271\ LANE + 0.47286\ REACCESS + 4.34780$
$DIRECTAC - 0.10197\ DPOPCNTR + 0.00104\ POPBUFF$
$+ 0.00022820\ EMPBUFF$

For all the four models, there is a strong relationship between *AADT* and the independent variables. There is no multicollinearity among them. The coefficient of determination, R^2, ranges between 0.66 and 0.82 for the four models. In other words, *at least 66%* of the sum of squares in *AADT* can be associated with the variation in these independent variables. The overall *F*-test is very significant (*Prob. > F = 0.0001 or smaller*).

Among the four models, Model 1 includes the functional class (*FCLASS*) as a variable, which is the most significant predictor with a partial R^2 of 0.7103. *LANE* has a partial R^2 of 0.0836. The partial R^2's for *REACCESS*, *DIRECTAC*, and *BUFFEMP* (listed in order of their importance based on their partial R^2) are all small at less than 0.02.

A concern about the use of functional class is that functional classes are to a large degree determined based on the traffic volume. Therefore, while a strong correlation may exist between AADT and functional classes, the use of functional classes in a model fails to capture the underlying causes of varying traffic volumes. Additionally, since other criteria are also used in determining the functional classes such as anticipated future land developments or linkage of facilities of significance, this correlation may be weak for certain roads.

In an attempt of directly capturing the underlying causes of traffic volume variations, three additional models are tested. Model 2 uses no information of function classification, but it also has the lowest R^2 (0.66). Models 3 and 4 include two variables *POPBUF* and *EMPBUF* that depict the land use intensity in the “service” areas surrounding different types of facilities. The size of these service areas is determined based on the functional class of the road. The functional classes here are used as a way to define the hierarchy of roads in a transportation network thus the general service areas of roads. The inclusion of variables *POPBUF* and *EMPBUF* resulted in improvements of models, with a R^2 of 0.76 for both.

In all the models, the signs of the coefficients are as expected. For instance, the positive sign for *FCLASS* means that a road of a function class higher in the network hierarchy will carry more daily traffic. The positive sign for *LANE* indicates that more lanes usually mean more traffic. The positive sign for *REACCESS* indicates that a road easily accessible to regional employment centers as measured by employment weighted by network travel time tends to have a higher traffic volume. The positive sign for *DIRECTAC* is consistent with the expectation that a road with direct access to expressways tends to have more traffic. The positive sign for *POPBUF* and *EMPBUF* explain the positive relationship between *AADT* with population and employment concentrations in the "service" area of a road. The negative sign of *DPOPCNTR* means the farther away a road is from the center of population of the urbanized area, the less traffic it tends to carry. The SAS output for the models are given in Figures 4, 5, 6, and 7, respectively.

```
Model: MODEL1
Dependent Variable: AADT

                              Analysis of Variance

                               Sum of         Mean
        Source         DF      Squares        Square      F Value       Prob>F
        Model           5 148822.82461   29764.56492      691.281       0.0001
        Error         769  33110.90976      43.05710
        C Total       774 181933.73437

            Root MSE       6.56179     R-square       0.8180
            Dep Mean      24.33303     Adj R-sq       0.8168
            C.V.          26.96660

                               Parameter Estimates

                    Parameter      Standard     T for H0:                  Variance
Variable  DF         Estimate         Error    Parameter=0    Prob > |T|     Inflation
INTERCEP   1        -9.520386    0.69724437       -13.654        0.0001    0.00000000
LANE       1         3.428939    0.20361749        16.840        0.0001    1.83129936
FCLASS     1         8.480001    0.34578021        24.524        0.0001    1.98849722
REACCESS   1         0.596752    0.07343288         8.126        0.0001    1.04482819
BUFFEMP    1         0.069086    0.01757626         3.931        0.0001    1.09992301
DIRECTAC   1         2.991573    0.60794499         4.921        0.0001    1.29422806
```

FIGURE 4 SAS Output for Model 1

```
Model: MODEL2
Dependent Variable: AADT
                                     Analysis of Variance

                                 Sum of           Mean
        Source          DF      Squares         Square     F Value     Pr > F
        Model            4       120204          30051      374.85     <.0001
        Error          770        61730       80.16819
        C Total        774       181934

            Root MSE              8.95367     R-Square      0.6607
            Dependent Mean       24.33303     Adj R-Sq      0.6589
            Coeff Var            36.79636

                                     Parameter Estimates
                             Parameter       Standard                                Variance
      Variable      DF        Estimate          Error     t Value     Pr > |t|       Inflation
      Intercept      1        -6.15742        1.75116       -3.52       0.0005               0
      lane           1         6.55471        0.22540       29.08       <.0001         1.20521
      racs           1         0.61433        0.12017        5.11       <.0001         1.16673
      dirac          1         7.88344        0.80287        9.82       <.0001         1.21231
      pop99          1        -0.34494        0.09402       -3.67       0.0003         1.20404
```

FIGURE 5 SAS Output for Model 2

Model: MODEL3
Dependent Variable: AADT

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	138715	27743	493.64	<.0001
Error	769	43219	56.20103		
Corrected Total	774	181934			

Root MSE	7.49673	R-Square	0.7624
Dependent Mean	24.33303	Adj R-Sq	0.7609
Coeff Var	30.80888		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	1	-4.66034	1.46853	-3.17	0.0016	0
lane	1	4.95341	0.20833	23.78	<.0001	1.46865
racs	1	0.51119	0.10077	5.07	<.0001	1.17045
dirac	1	4.52713	0.69720	6.49	<.0001	1.30406
pop99	1	-0.10689	0.07980	-1.34	0.1808	1.23747
bfpop	1	0.00112	0.00006179	18.15	<.0001	1.52898

FIGURE 6 SAS Output for Model 3

Model: MODEL4
Dependent Variable: AADT

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	139147	23191	416.27	<.0001
Error	768	42787	55.71229		
Corrected Total	774	181934			

Root MSE	7.46407	R-Square	0.7648
Dependent Mean	24.33303	Adj R-Sq	0.7630
Coeff Var	30.67463		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|	Variance Inflation
Intercept	1	-4.26565	1.46899	-2.90	0.0038	0
lane	1	4.86271	0.20996	23.16	<.0001	1.50490
racs	1	0.47286	0.10128	4.67	<.0001	1.19250
dirac	1	4.34780	0.69715	6.24	<.0001	1.31530
pop99	1	-0.10197	0.07947	-1.28	0.1999	1.23808
bfpop	1	0.00104	0.00006782	15.37	<.0001	1.85787
bfemp	1	0.00022820	0.00008199	2.78	0.0055	1.60900

FIGURE 7 SAS Output for Model 4

MODEL VALIDATION

A rigorous test of the models' predictive capability involves using the models to estimate AADTs for the roads other than those used for the calibration and then comparing the estimation with the actual AADTs observed. At the beginning of the analyses, the data set is divided into two groups, one used for model calibration and the other for model validation. The 82 testing data are used to examine the models' predictive capability. Table 2 summarizes the testing results.

MSE is the mean square of errors, and total error is the error in percentage for the entire testing data set, which is under 3% for all models.

TABLE 2 Model Testing Result Summary and Comparison

Model	Model 1	Model 2	Model 3	Model 4
Variables Used	*LANE* *FCLASS* *REACESS* *BUFFEMP* *DIRECTAC*	*LANE* *REACESS* *DIRECTAC* *DPOPCNTR*	*LANE* *REACESS* *DIRECTAC* *DPOPCNTR* *POPBUF*	*LANE* *REACESS* *DIRECTAC* *DPOPCNTR* *POPBUF* *EMPBUF*
R-square	0.8180	0.6607	0.7624	0.7648
Adj R-square	0.8168	0.6589	0.7609	0.7630
MSE	50.00	80.17	56.20	55.71
Total Error[a] (%)	+0.25	-2.37	+1.95	+1.72

a. Total Error = (SumPredt. – SumADT)/SumADT*100

Table 3 gives the cumulative percent errors for the four models. For instance, for Model 1, 58.54% of the testing points have an error less than 20%. The maximum error for each model is also given in the last row. Figure 8 illustrates the error distributions and Figure 9 the cumulative percentages of testing points within certain error ranges for the four models.

Comparison of the models shows that Model 1 has the best performance in terms of prediction errors, with about 37% of testing points having an error smaller than 10% and about 73% of the testing points having an error smaller than 30%, respectively. Model 2 has the worst performance. Models 3 and 4 have better than that of Model 2. They also have close performances. This indicates that Model 3 may be adequate since it has one fewer variable thus reduce the data collection and processing costs.

Data for all the variables are relatively easy to obtain. The function classification and number of lanes are always readily available. Determination of accessibility to regional employment centers (*REACCESS*) requires moderate effort, which involves identification and creation of employment centers, building of an appropriate network including a turn table, and use of network analysis tools such as finding the shortest paths. Direct access (*DIRECTAC*) to expressways and employment in a buffer zone (*BUFFEMP*) are both easy to determine.

TABLE 3 Cumulative Percentage of Testing Points in Different Error Ranges

Error Range (%)	Cumulative Percentage of Testing Points			
	Model 1	Model 2	Model 3	Model 4
10	36.59	17.07	31.71	31.71
20	58.54	32.93	53.66	52.44
30	73.17	59.76	69.51	68.29
40	79.27	73.17	78.05	75.61
50	89.02	81.71	82.93	82.93
60	93.90	86.59	90.24	90.24
70	95.12	90.24	93.90	95.12
80	95.12	92.68	95.12	95.12
90	95.12	93.90	96.34	96.34
100	95.12	95.12	96.34	96.34
200	100.00	97.56	98.78	98.78
300	100.00	100.00	100.00	100.00
Max Error (%)	155.67	162.94	173.12	185.40
# with Error > 100%	4	4	3	3
# outliers	3	3	3	3

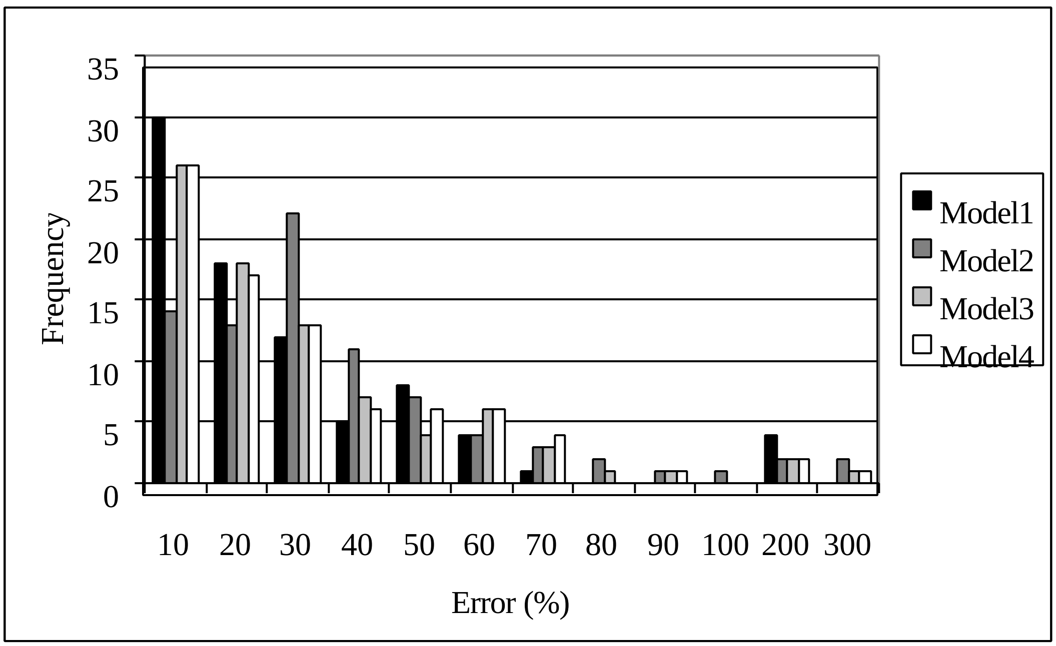

FIGURE 8 Error Distributions of Four Models

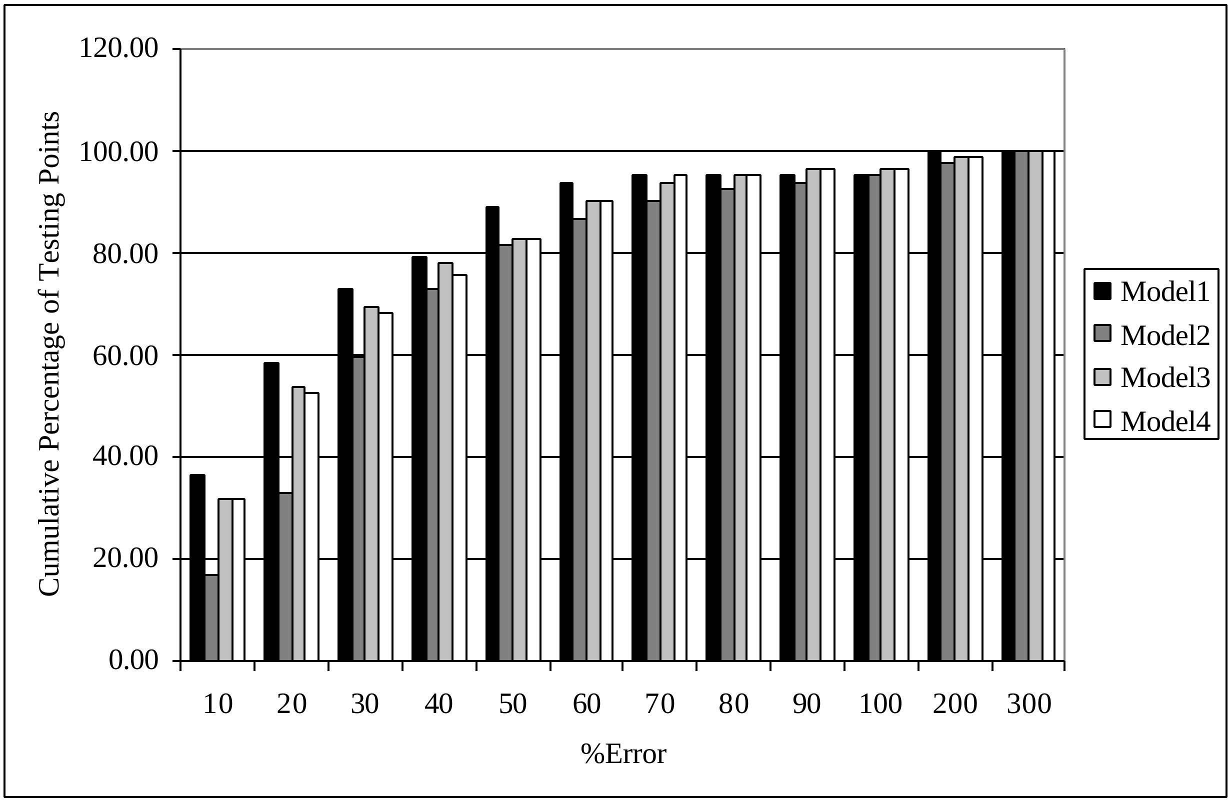

FIGURE 9 Cumulative Percentage of Testing Points in Different Error Ranges

The percent difference between the predicted and actual AADT values ranges from 0.3% to 155.6% for Model 1 and 0% to 288.1% for Model 3. Figure 10 shows the locations of testing points and the percent errors for Model 1. The largest errors (> 100%) all occur on low volume roads. Three of the four largest errors happen on US 1. Even though US 1 is a minor arterial, the traffic volumes at these locations are only 9,400, 7,100, and 4,200, respectively. Most of the north-south traffic is carried by I-95. Additionally, near the coast, employment concentration is relatively low, also contributing to the low volumes on these roads. In fact, all these three test points are outliers.

The fourth large error occurs west and south of I-75. It is a minor arterial, but its AADT is only 9,200 and the surrounding area is mostly rural. This examination of the conditions at locations of the largest errors reveals special cases where function class is high but high volume is not likely to occur due to land use patterns and the network configurations. This also suggests the need for caution when applying a model to determine if a particular location might represent an outlier.

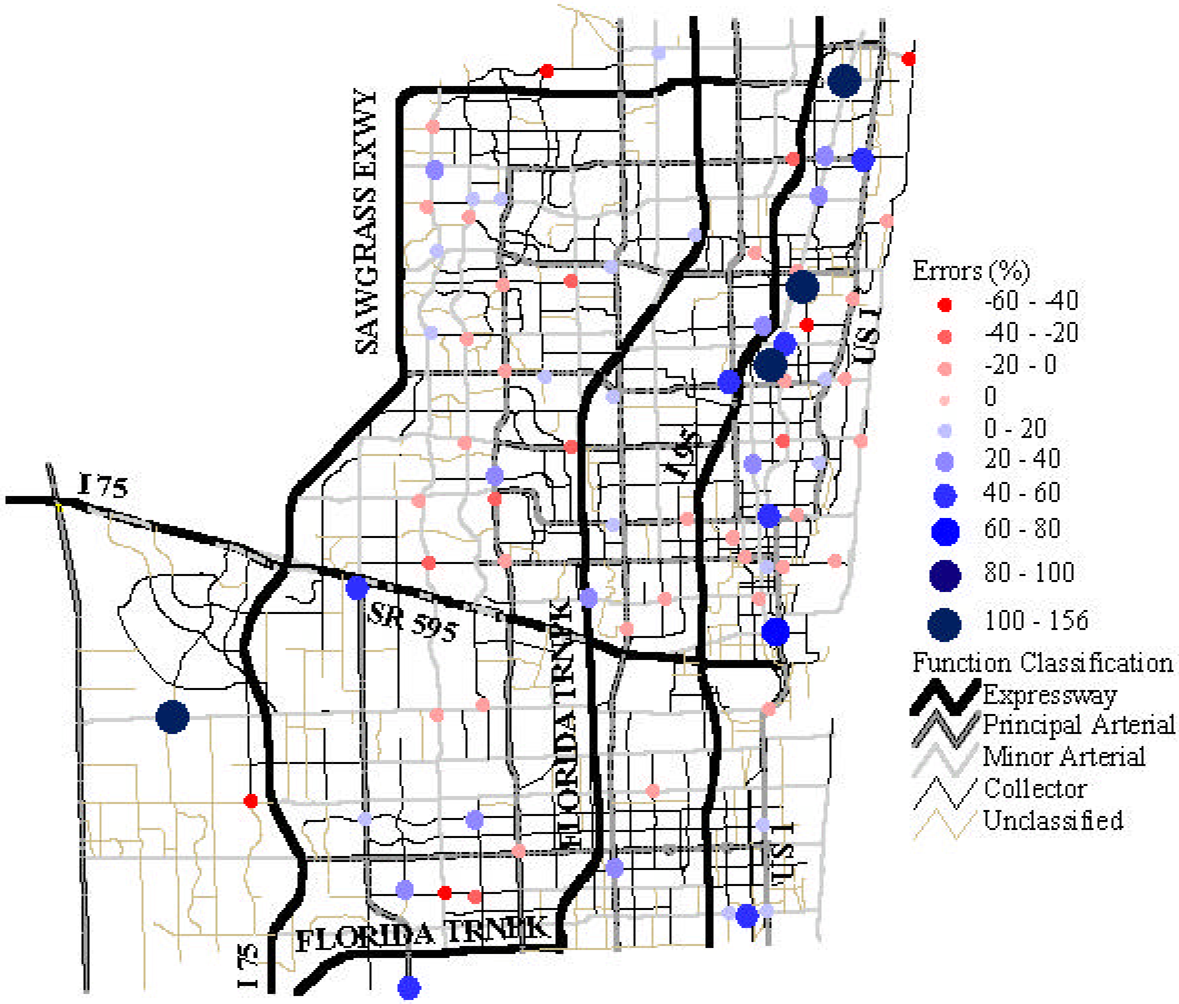

FIGURE 10 Locations of Testing Points and Associated Error in Percentage for Model 1

Figure 11 shows the error distribution for Model 3. The largest errors occur east of I-95. The error that was large in Model 1 and occur southwest of I-75 now is small in Model 3, which may indicate that local land use variables POPBUF and the distance to population center have help better reflect the actual land use intensity and the through traffic carried. There is also an interesting pattern of locations where underestimation and overestimation occur, which suggests that there are certain land use factors not being reflected.

CONCLUSIONS

Four multiple regression models for estimating AADT on non-expressway roads have been developed for Broward County. A large set of predictors are developed and investigated. Up to six predictors have been used in the final models. They include function classification, number of lanes, accessibility to regional employment centers, directness of expressway access, and population and employment around a count station. The models are able to explain 66% to 83% of the total variability depending on the variables used. Generally, the more variables used, the better a model performance is. The choice of a model therefore is likely to be based on data processing cost. On the other hand, the model data are generally available and relatively easy to process.

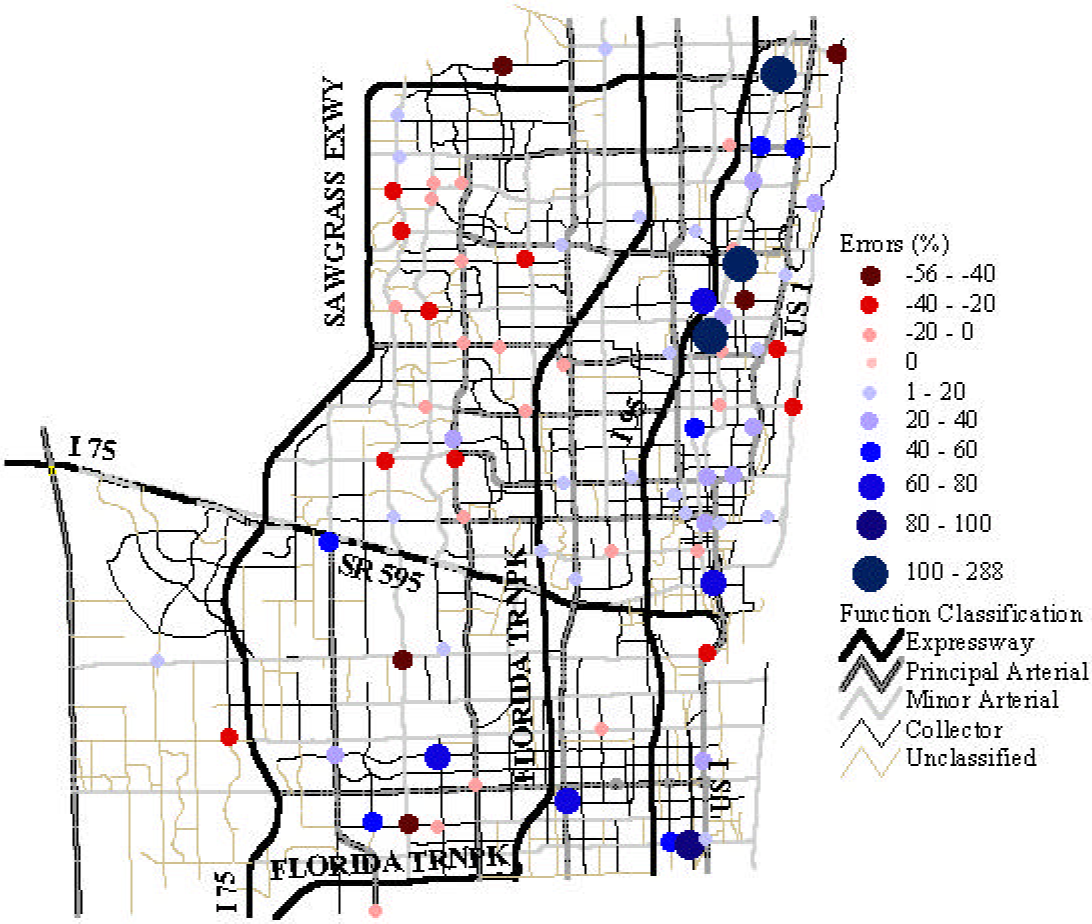

FIGURE 11 Locations of Testing Points and Associated Error in Percentage for Model 3

Function class and number of lanes prove to be the most significant predictors. Although function classes as a predictor outperform other variables, they do not explain the causes that determine AADT since they themselves are determined by other factors including traffic volume. Model testing has already shown that function classes are not consistently correlated with AADT. However, the model that does not use any information related to function classes has the worst performance. Additional land use variables that are determined based on function classes significantly improve the models although they are not strong predictors as function class. Such indirect use of function classes may acceptable since they reflect the hierarchy of the network system, which may be difficult to capture otherwise.

The models in their current forms may not be adequate to meet the need of engineering design or the calibration of travel demand models, but their performances have improved. They may be used for tasks that do not require high level of accuracy at individual sites such as estimating system-wide vehicle miles traveled.

Future work includes careful examination of the spatial patterns of errors resulted from different models and the causes for errors. More effective land use indicators are still needed as well as the study of temporal stability of the regression models.

ACKNOWLEDGEMENTS

The authors would like to thank Ms. Lina Koulekowski in Broward County MPO for providing GIS data on AADT and function classification system, and Ms. Lee-Fang Chow for assistance in GIS analysis.

REFERENCES

1. Mohamad, D., Sinha, K. C., and Kuczek, T. (1998), "An Annual Average Daily Traffic Prediction Model for County Roads," presented at Transportation Research Board, 77th Annual Meeting, January 11-15, 1998, Washington D. C.

2. Cheng, C., *Optimum Sampling for Traffic Volume Estimation*, Ph.D. Dissertation, University of Minnesota, Carlson School of Management, Minnesota, 1992.

3. Deacon, J. A., Pigman, J. G. and Mohenzadeh, A., *Traffic Volume Estimates and Growth Trends*, UKTRP-87-32, Kentucky Transportation Research Program, University of Kentucky, Kentucky, 1987.

4. Fricker, J. D. and Saha, S. K. *Traffic Volume Forecasting Methods for Rural State Highways – Final Report*, FHWA/IN/JHRP-86/20, Joint Highway Research Project, School of Civil Engineering, Purdue University, Indiana, 1987.

5. Aldrin, M., "A Statistical Approach to the Modeling of Daily Car Traffic", *Traffic Engineering and Control*, Vol. 36, Issue 9, 1995, pp. 489-493.

6. Saha, S.K. "The Development of a Procedure to Forecast Traffic Volumes on Urban Segments of the State and Interstate Highway Systems," Ph.D. Dissertation, School of Civil Engineering, Purdue University, 1990.

7. Xia, Q., Zhao, F., Chen, Z., Shen, L. D., and Ospina, D. "Development of a Regression Model for Estimating AADT in a Florida County," Transportation Research Record, No.1660, Transportation Research Board, National Research Council, Washington, D.C., 1999, pp. 32-40.

8. Shen, L.D., Zhao, F., and Ospina, D. "Estimation of Annual Average Daily Traffic for Off-System Roads in Florida," Research Report, Florida Department of Transportation, Tallahassee, FL, June 1999.