

USING INTELLIGENT TRANSPORTATION SYSTEMS DATA ARCHIVES FOR TRAFFIC SIMULATION APPLICATIONS

Patricio Alvarez, Universidad del Bío Bío, palvarez@ubiobio.cl
Mohammed Hadi, Florida International University, hadim@fiu.edu
Chengjun Zhan, Florida International University, zhanc@fiu.edu

ABSTRACT

Collecting data for traffic simulation modeling applications is expensive. Data collected using traditional methods may not represent the variations in traffic demands and conditions throughout the year and may require additional efforts to compensate for missing and erroneous data. This paper discusses a series of data manipulation procedures for the utilization of ITS data archives to support simulation modeling. These procedures allow the extraction of collected volume data from ITS data archives, automatic identification of temporal patterns in the data, automatic segmentation of daily demands into dynamically captured sub-periods, resolving possible spatial inconsistencies in the data, and estimating missing volumes.

Key words: microscopic simulation, traffic parameters, traffic data

1. INTRODUCTION

Traditionally, traffic simulation applications have been developed using volume data collected using tube and/or manual turning volume counts. These applications are calibrated using data from travel time studies combined with volume data and field observations of queues and other traffic conditions. The collection of the required data, however, is expensive, particularly for large simulated systems. In addition, the data collected using traditional methods are normally for one or few days that may not represent the traffic demands and conditions throughout the year.

Intelligent Transportation Systems (ITS) agencies have used devices as traffic detectors, closed circuit television cameras (CCTV), electronic toll readers, and license plate readers to collect traffic parameter measurements for operational purposes. In recent years, these agencies have started archiving the data collected by these devices (FHWA, 2004). Because ITS detectors and communicators are already in place to collect data for operational purposes, the extra cost to archive and manage the data is relatively low. As ITS data archives become more widely available, the utilization of such archives for the development and calibration of simulation applications will be an attractive option. This utilization will provide a significantly lower cost and a more efficient data collection method compared to traditional methods and will increase safety by reducing the need for personnel to go out to the field for data collection purposes.

The additional details provided by the ITS data, both in time and space resolutions, will allow better representations of real-world environments in simulation applications. For example, the use of archived ITS data will allow the simulation of seasonal variations in traffic, special events, accidents, work zones, weather events, other types of incidents, and incident management strategies. This paper discusses the development of procedures and tools for the utilization of data from the ITS data archives to support the use of simulation models.

2. PREVIOUS EFFORTS

Few studies have investigated the use of archived ITS data for simulation modeling applications. Gomez (Gomez *et al*, 2004) presented a procedure for constructing and calibrating a detailed model of a freeway, based on detector data using VISSIM. Field data used as input for the model was compiled from two separate sources: loop-detectors on the on-ramps and mainline, stored in a central database referred to as the Performance Measurement System (PeMS) and a manual survey of on-ramps and off-ramps. Gaps in both sources made it necessary to use both traffic detector data and manually collected data sets. A data processing algorithm was implemented to filter, aggregate, and correct the PeMS data.

Barcelo (Barcelo *et al*, 2002 and Barcelo *et al*, 2003) described an implementation of a microscopic simulation tool (AIMSUN) to support traffic management strategies. The project integrated an ITS data warehouse with the AIMSUN modeling environment. This integration allowed the analysis and fine-tuning of traffic management strategies.

Xin (Xin *et al*, 2006) proposed a methodology for checking and correcting temporal errors integrated with an optimization-based algorithm for reconciling spatial inconsistencies in traffic counts collected using traffic detectors. First the data is filtered using a time-series model to

detect outliers. The time model is fitted based on a number of stations randomly selected from a freeway network. In a second step, volumes are corrected when the difference between adjacent stations exceed a given threshold. The volume correction is achieved by minimizing the difference between the observed and corrected volumes for a set of spatially related stations.

3. PROVIDED FUNCTIONALITIES

The use of ITS data that covers a long period of time provides the opportunity to classify the days throughout the year into different patterns. For example, on certain corridors, it may be important to differentiate between different seasons or to simulate days with special events. In addition, it is necessary to exclude days with unusual demands or congestion when simulating typical day patterns. Thus, a procedure was developed to categorize the demand data for different days into patterns based on the similarity of travel demands measured by the traffic detectors.

ITS data can include inconsistent, non-balanced, and missing measurements. Thus, a procedure was developed to produce consistent and balanced traffic demands and to estimate missing traffic demands based on measured demands. Another provided functionality was the automatic segmentation of the time period for each identified patterns into sub-periods of similar demands. The details of the procedures developed in this study to implement the required functionalities are discussed in the following section.

4. DEVELOPED TOOL

The section presents a discussion of the developed modules that deliver the functionalities identified in the previous section.

4.1. Identification and Selection of Simulated Patterns

A module was developed to categorize the demand data for different days into patterns based on the similarity of the time series of volume counts of different days. The k -means clustering algorithm (Alpaydin, 2004) was used for the categorization. The analyst can specify all or a subset of the detector measurement to be used in the categorization. This is an iterative partitioning algorithm that minimizes the sum of time series distances to cluster centroids, summed overall clusters. In this study, the times series distance is measured by the Euclidian distance defined as follows:

$$dist((v_j, c_k) = \sum_i (v_j(t_i) - c_k(t_i))^2 \quad \forall j \in k \quad (1)$$

$$c_k(t_i) = \frac{1}{n_k} \sum_j v_j(t_i)^2, \forall j \in k \quad (2)$$

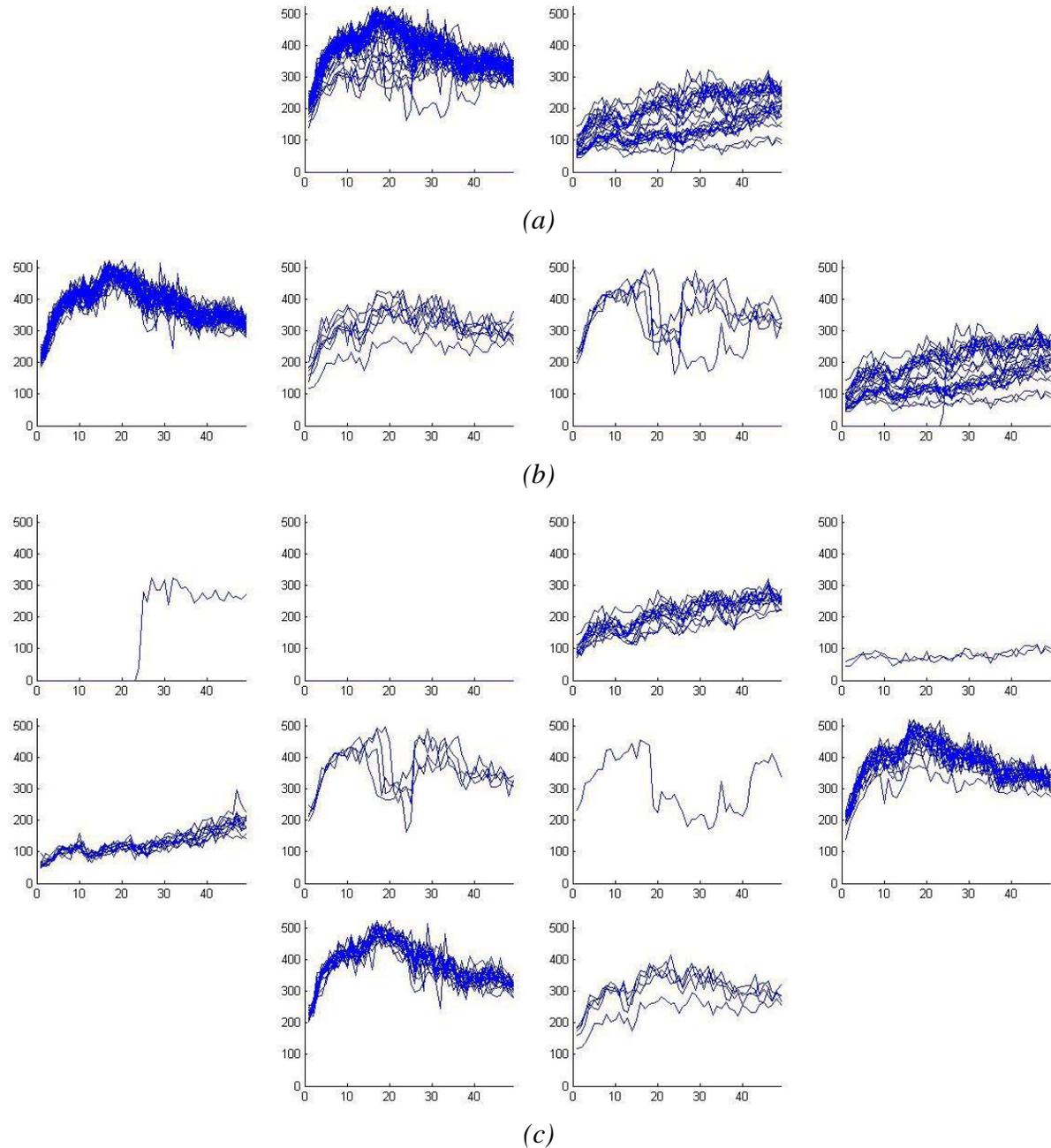
where

- $v_j(t_i)$ = time series measurement j at time interval i from STEWARD,
 $c_k(t_i)$ = centroid of cluster k at time interval i , and
 n_k = total number of time series in cluster k .

The optimization routine used in the clustering algorithm achieves a local optimal that can be different each time the algorithm is run depending on the starting point of the optimization. The starting point is a set of centroids that will serve as initial centroids in the first iteration of the algorithm. This set is chosen randomly from the data available. The number of centroids in the set is equal to the number of clusters requested by the analyst. Thus, the analyst should run the algorithm for a number of replications to associate the measured daily demands with the clusters. The results presented in this paper are based on 10 replications of the algorithm.

With the developed module, the analyst has the option of specifying the number of clusters that result from the analysis. Figure 1 shows the results of applying the data selection procedure to a set of 40 days using different number of clusters. Figure 1 contains 4 hours of data reported every 5 min. The initial dataset contains weekdays; weekends; and days with incidents, bad weather, special events, and detector malfunctions. Of course, the more clusters are used the more homogeneous each cluster will be. However, too many clusters will not be useful since the analyst's aim in most cases is to identify major differences in the patterns to be able to simulate a limited number of patterns. Figure 1 shows the results of the clustering when specifying two, four, and ten as the number of patterns resulting from the clustering procedure. As can be seen from Figure 1-a, specifying two patterns is not sufficient, since the algorithm basically classifies the days into a weekday and a weekend pattern. Figure 1-b shows the results of requesting four patterns to be produced. The procedure was able to classify the patterns in two different weekday clusters. The first cluster from the left in Figure 1-b represents higher demand weekdays compared to those days represented by the second pattern from the left in the figure. The third pattern from the left represents incident days and the fourth pattern represents weekends. Figure 1-c shows the results of the analysis when ten patterns are specified. A visualization routine was also included in the developed tool to allow the analyst to associate each pattern with specific days. This allowed the determination of the reasons for the difference in the patterns such as different seasons, different weather, special events, different incident attributes, and so on. By examining the resulting patterns and the associated information, the analyst can determine what cluster to use in the analyses, which days should be excluded as outliers, and which clusters should be divided further into sub-clusters. For example, based on the data included in Figure 1, the analyst may decide to simulate two weekday patterns and one heavy weekend day pattern. In addition, the analyst may want to classify incident days further into different incident categories and use these days in calibrating simulation models for incident conditions. It is interesting to note that the second pattern from the left in Figure 1-c does not have any detector measurements. This pattern represents days with malfunction of the detection station at this location.

Figure 1: Results of clustering using different number of clusters. (a) Two clusters. (b) Four clusters. (c) Ten clusters. Vertical axis is traffic volume per 5 min and horizontal axis is time in minutes



4.2. Time Period Segmentation

Microscopic simulation requires segmenting the day into discrete time intervals. Traditionally, analysts have divided the day into intervals that represent different peak periods during the day (e.g., AM, PM, and midday). These periods are then simulated separately. The analysts can also subdivide the peak period into subintervals to account for the variation in demands within the

peak period. Most microscopic simulation tools allow coding sub-intervals to be ran in the same run. With more detailed data available from the ITS archives, it is useful to automate this segmentation of the time periods.

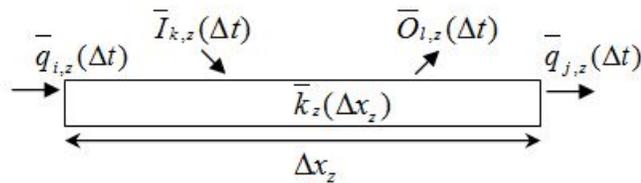
A procedure was developed in this study to segment the 24 hour or peak period volumes based on the measurements from all or a subset of the detection stations. The segmentation was done using an algorithm referred to as the Bottom-Up algorithm that has been used in data mining for linear piece wise segmentation (Keogh *et al*, 2001). First, the Bottom-Up algorithm creates the finest possible approximation of the time series, therefore n segments are used to approximate the n -length time series. Next, the cost of merging each pair of adjacent segments is calculated, and the algorithm begins to iteratively merge the lowest cost pair until a stopping criteria is met. The analyst should decide on the stopping criteria based on the number of segments appropriate for the purpose of the analysis, and the quality of the data.

The number of segments to represent the time series can be selected by the user. There is a trade-off between the number of segments and the complexity of the developed simulation application. So, it is desirable to select the lowest number of segments that capture the main temporal variations in demands.

4.3. Spatial Conciliation and Estimation of Missing Demands

Although most ITS databases implement data filtering and imputation methods, it was found that inconsistencies between adjacent detector measurements still exist. In addition, in many cases, detectors are not placed on the ramps. Thus a procedure was developed to resolve inconsistencies and non-balanced traffic between upstream and downstream detectors and to estimate missing link measurements (on the ramps with no detectors) based on other link measurements.

Let's consider the following segment:



where

- Δt = period of time,
- Δx_z = length of the section z ,
- $\bar{q}_{i,z}$ = average flow at location i in section z during Δt ,
- $\bar{q}_{j,z}$ = average flow at j in section z during Δt ,
- \bar{k}_z = average density in section z during Δt ,

$\bar{I}_{k,z}$ = average in-flow at ramp k in section z during Δt , and
 $\bar{O}_{l,z}$ = average off-flow at ramp l in section z during Δt .

The conservation equation results in the following equations:

$$\bar{k}_z \Delta x_z = (\bar{q}_{i,z} - \bar{q}_{j,z} + \bar{I}_{k,z} - \bar{O}_{l,z}) \Delta t \quad (3)$$

$$(\bar{k}_z(t + \Delta t) - \bar{k}_z(t)) \Delta x_z = (\bar{q}_{i,z} - \bar{q}_{j,z} + \bar{I}_{k,z} - \bar{O}_{l,z}) \Delta t \quad (4)$$

With all variables are as defined above. This equation is applied between every two consecutive detection stations. Further, we introduce in the formulation error terms to account for errors in detector measurements of volume and occupancy resulting in the following formulation:

$$(\bar{k}_z(t + \Delta t) + \bar{\delta}_z(t + \Delta t) - \bar{k}_z(t) - \bar{\delta}_z(t)) \Delta x_z = ((\bar{q}_{i,z} + \bar{\varepsilon}_{i,z}) - (\bar{q}_{j,z} + \bar{\varepsilon}_{j,z}) + (\bar{I}_{k,z} + \bar{\varepsilon}_{k,z}) - (\bar{O}_{l,z} + \bar{\varepsilon}_{l,z})) \Delta t \quad (5)$$

where

$\bar{\varepsilon}_{x,z}$ = flow correction at location x in section z , and
 $\bar{\delta}_z(X)$ = density (occupancy) correction in section z during period X .

For steady state conditions (where no queue occurs), the problem can be simplified assuming that the density in section z does not vary significantly during Δt , resulting in the following:

$$(\bar{\delta}_z(t + \Delta t) - \bar{\delta}_z(t)) = \frac{\alpha}{2(L + C)} ((Occ_i(t + \Delta t) + Occ_j(t + \Delta t)) - (Occ_i(t) + Occ_j(t))) = 0 \quad (6)$$

Thus the conservation equation becomes:

$$\bar{\varepsilon}_{i,z} - \bar{\varepsilon}_{j,z} + \bar{\varepsilon}_{k,z} - \bar{\varepsilon}_{l,z} = -\bar{q}_{i,z} + \bar{q}_{j,z} - \bar{I}_{k,z} + \bar{O}_{l,z}, \forall z \quad (7)$$

It is possible to formulate several optimization criteria to minimize the error values $\bar{\varepsilon}_{i,z}, \bar{\varepsilon}_{j,z}, \bar{\varepsilon}_{k,z}, \bar{\varepsilon}_{l,z}, \bar{\delta}_z(t + \Delta t)$ and $\bar{\delta}_z(t)$ (Hillier *et al*, 2004). Three different formulations of quadratic error summation minimization and linear programming optimization were investigated. The first summation minimizes the squares of all error corrections subject to complying with all conservation equations of the system and constraining all corrections to a reasonable maximum and minimum pre-defined values. The second formulation is similar to the first formulation but the error corrections are weighted by the original volumes. The third is a linear programming problem that minimizes the maximum correction. Testing revealed that the results from the first formulation results were as good or better than the other two formulations, thus it was used for the rest of this study. This formulation is as below:

Minimize

$$\sum_{i,j,k,l} \varepsilon^2 + \sum_{z,t} \delta^2(t) \quad \forall z \in \text{System} \quad (\Delta t \text{ given}) \quad (8)$$

Subject to:

$$\begin{aligned} & (\bar{\varepsilon}_{i,z} - \bar{\varepsilon}_{j,z} + \bar{\varepsilon}_{k,z} - \bar{\varepsilon}_{l,z}) \Delta t - (\bar{\delta}_z(t + \Delta t) - \bar{\delta}_z(t)) \Delta x_z = \\ & = -(\bar{q}_{i,z} - \bar{q}_{j,z} + \bar{l}_{k,z} - \bar{O}_{l,z}) \Delta t + \frac{\alpha}{2(L+C)} ((Occ_i(t + \Delta t) + Occ_j(t + \Delta t)) - (Occ_i(t) + Occ_j(t))) \Delta x_z \quad \forall z \end{aligned}$$

$$\varepsilon_{i,lower} \leq \bar{\varepsilon}_{i,z} \leq \varepsilon_{i,upper} \quad \forall i, z$$

$$\varepsilon_{j,lower} \leq \bar{\varepsilon}_{j,z} \leq \varepsilon_{j,upper} \quad \forall j, z$$

$$\varepsilon_{k,lower} \leq \bar{\varepsilon}_{k,z} \leq \varepsilon_{k,upper} \quad \forall k, z$$

$$\varepsilon_{l,lower} \leq \bar{\varepsilon}_{l,z} \leq \varepsilon_{l,upper} \quad \forall l, z$$

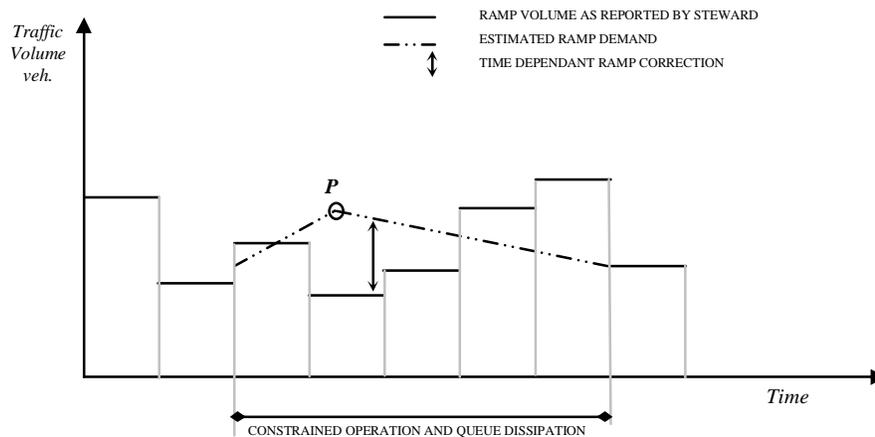
$$\delta_{z,lower} \leq \bar{\delta}_z(t) \leq \delta_{z,upper} \quad \forall z, t$$

The above formulation was also extended to cases where additional information is available from other sources, which possibly have different levels of reliability than the ITS data. Such sources may include short term counts, previous corridor studies, or old tube counts. In these cases, the analysts will have the option to assign weights to different information to account for their different levels of reliabilities.

During testing of the above model, it was determined that the results from the optimization should be examined to determine if there are large volume corrections due to measurements at one or two locations that are clearly not consistent with other measurements in the system. In this case, it is advised to take the measurements at these locations out of the optimization model. This will be illustrated using the case study presented later in this document.

The correction of the inconsistencies and non-balanced volumes must account for recurrent bottleneck locations that prevent a portion of the demand from being served during a given time period. In these cases, the counts from the data archives may not actually represent the actual demands but volumes constrained by downstream bottleneck throughputs. A procedure was developed to approximate the demand for the ramps and mainline locations affected by the bottlenecks during the constrained demand periods. The procedure first detects the presence of bottleneck and the affected locations. In addition, it identifies the time period during which the demand is constrained (traffic is queuing) and the time period during which the queue is dissipating. The sum of the volumes during these periods represents the total demands. The challenge is to distribute these demands among the sub-periods (e.g., 15 minute intervals) during the queuing and queue dissipation periods. One of three options are given to the analysts: the assumption of linear increase and decrease in traffic flow (triangular pattern) during the period, as shown in Figure 2; inputting the distribution of demands as the proportions of demand for each sub-period during the queuing and queue dissipation periods; or allowing the model to automatically specify these proportions based on measurements at detector locations specified by the user.

Figure 2: Ramp volume correction



5. CASE STUDY

The feasibility of using ITS data to develop microscopic simulation applications was tested using a freeway corridor equipped with ITS devices including traffic detectors and CCTV cameras. The traffic corridor is the eastbound section of State Road (SR) 826, also known as the Palmetto Expressway, located in Miami, Florida. This corridor includes six interchanges and begins a quarter mile west of the NW 67th Avenue interchange and ends a quarter mile east of the NW 12th Avenue interchange with a total length of 6.5 miles. This study focus is on the AM period between 5:00 and 10:00 AM and the PM period between 4:00 and 6:00 PM.

5.1. Data Collection

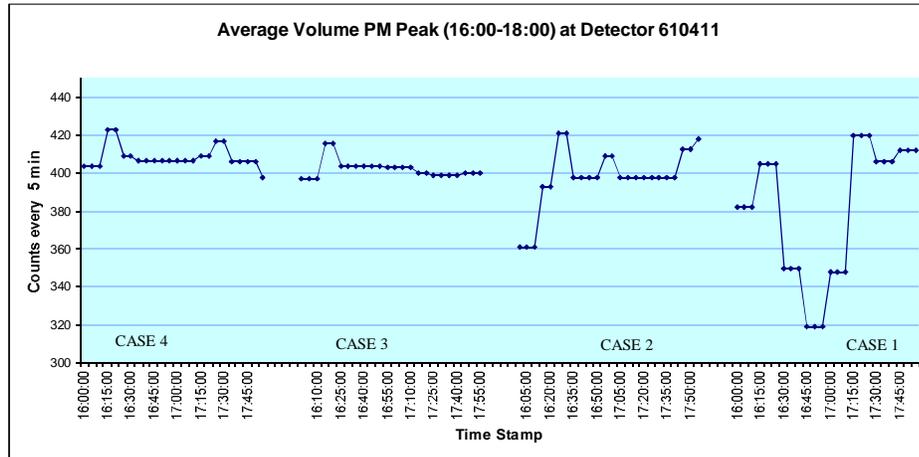
Volume, speed, and occupancy data were collected from the STEWARD ITS data warehouse. These parameters were measured by true presence microwave detectors located at 0.3-0.5 mile intervals on the test section. In total, there are 21 detectors on the eastbound direction of SR 826. The data were downloaded at the 5 minute aggregation level.

5.2. Pattern Selection

This section presents a comparison between four options to demonstrate the use of the pattern selection procedure. Figure 3 shows a comparison between four cases two consecutive hours during the PM peaks (4:00 PM to 6:00 PM). The first case uses the average of three consecutive day volume measurements collected in random from STEWARD. In the second case, the average of only two of these three days was used to exclude one day that seems to involve an incident conditions based on manual inspection of the data. The other two options utilize the pattern selection procedure with 22 days and 44 days, respectively. For these two options, the pattern selection procedure identified 16 days and 33 days belonging to a typical recurrent traffic pattern cluster. This indicates that for this corridor, the volumes on 30% of the days are non typical and should not be included when estimating the average typical demands on the corridor.

It can be seen that using a larger sample provide a more stable estimation of the average volumes. This stability helps in obtaining better data for the other procedures of this study.

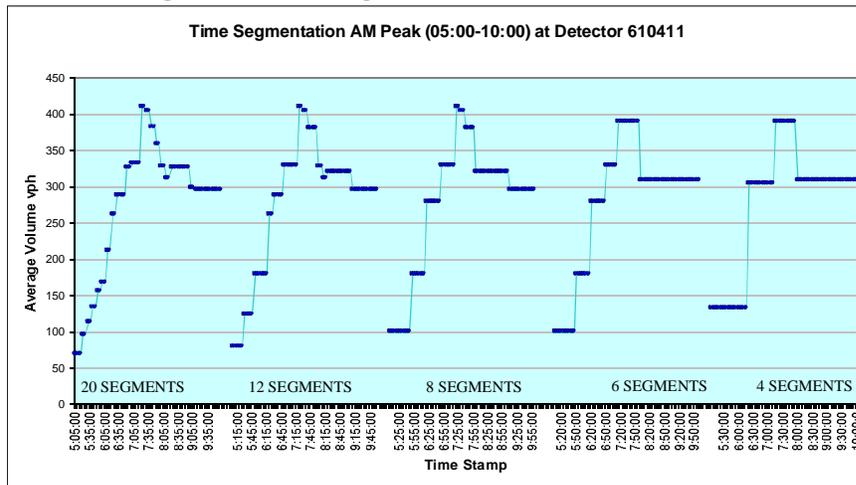
Figure 3: Average volumes based on different cases



5.3. Period Segmentation

The five consecutive hours during the AM peaks (5:00 AM to 10:00 AM) were segmented using the period segmentation procedure mentioned earlier in this paper. Figure 4 shows a comparison of using four, six, eight, twelve, and twenty segments in the segmentation procedure. Based on the results, the analyst can select the appropriate period segmentation based on the scope of the analysis.

Figure 4: Time segmentation based on volumes



It is interesting to quantify the difference in performance between the segmentation of a time series into different number of segments using the segmentation algorithm developed in this study versus a baseline segmentation, in which the length of each subinterval is fixed at 15 minutes without using the segmentation procedure. This performance was assessed using data for the two PM peak hours from one detector station on SR 826. Eight, six, and four segments

were requested for this data using the segmentation algorithm. The algorithm produced variable periods ranging from 10 min to 40 min.

As a measure of the quality of the segmentation, the sum square error between the segmented series and the original STEWARD data is calculated as follows:

$$error(v(t_i), r(t_i)) = \sum_i (v(t_i) - r(t_i))^2 \tag{9}$$

where

$v(t_i)$ = time series volume value at time interval i from STEWARD, and

$r(t_i)$ = average volume value for the resulting time segment that represent the volume of the segment that covers interval i .

Figure 5 shows a comparison between different approximation strategies, particularly it is shown that using a higher number of periods improve the quality of the segmentation. Requesting eight segments in the segmentation algorithm produced significantly lower error compared to the other numbers of segments. It is interesting to see that using the four segments produced by the developed segmentation procedure was able to achieve the same error as that obtained using eight consecutive 15 minute period without using the segmentation procedure. The algorithm with eight segments produced a 34% reduction in the error compared to using eight segments with consecutive 15 minute period. Table 1 provides the Mean Square Error and the Correlation Factor for each approximation.

Figure 5: Square error comparisons for different approximation strategies

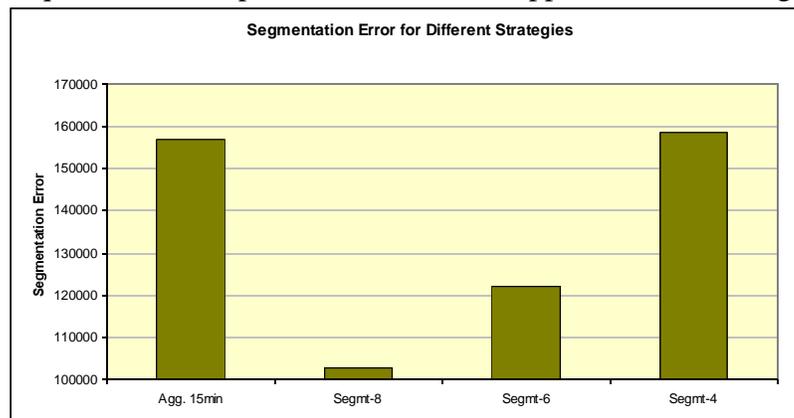


Table 1: Mean Square Error and Correlation Factor

	Agg. 15 min	Segmt-8	Segmt-6	Segmt-4
Mean Square Error	6534	4287	5093	6613
Correlation Factor	0.93	0.94	0.92	0.90

5.4. Spatial Conciliation and Missing Demand Estimation

The spatial conciliation procedure was used to correct inconsistencies between traffic detector measurements and to estimate missing volumes. On SR 826, traffic detectors are available at 0.3 to 0.5 mile intervals on the mainline. However, there are a number of ramps that do not have detectors. At other ramps, the locations of the detectors do not allow accurate measurements of the volumes. Fortunately, the detection stations on the mainline were located such that each ramp volume can be calculated as the difference between the volumes of the upstream and downstream link volumes. The following cases were compared:

- In Case 1, volume measurements from both mainline and ramp locations (where available) were used.
- In Case 2, ramp count measurements were not used. Rather, these measurements were calculated based on upstream and downstream locations.
- Case 3 is an extension of Case 2, where main line detector station 610011 was removed from the optimization model, as discussed later in this section.

Sensitivity analysis showed that data collected for only few days (e.g. three days) exhibits significant inconsistencies between detector locations and required significant corrections. Using longer periods of time (data from 30 and 60 days) reduced the inconsistencies significantly and produced better results (see Figure 3). As can be seen from Figure 3, in cases 1 and 2 where two to three days were used, the data showed unrealistic peaks. These peaks were not observed when more days were used as cases 3 and 4. Thus, 60 day data was used in this study.

Figure 6 shows the results of using only mainline detectors results, namely Case 2. From the results, it appears that the correction of the volumes on the mainline is less than 12% on the mainline in most cases. Also it can be seen that most corrections occurred in the second half of the corridor. Most of the corrections for the ramps occur for the last four on-ramps. The volumes for these ramps were reduced significantly during the correction process. Further examination indicates that this is due to the lower volume measurements than expected at the last mainline detector location. For this reason, the spatial conciliation algorithm decreased the volumes on the on-ramps to reduce the total upstream arrivals at this location. It also increased the volume significantly at this last detector location.

Figure 6: Mainline and ramp volume corrections for case 2

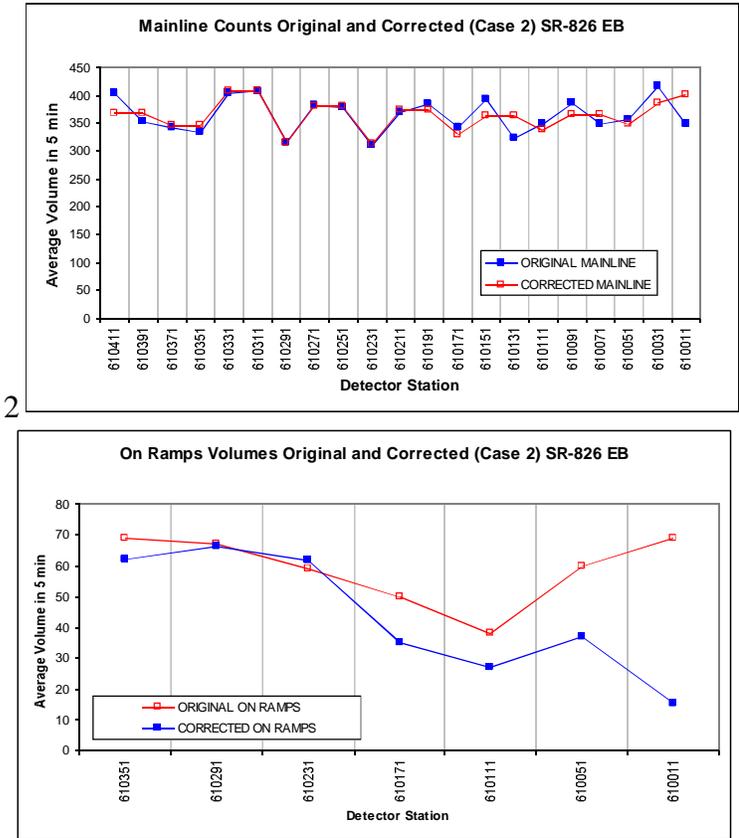
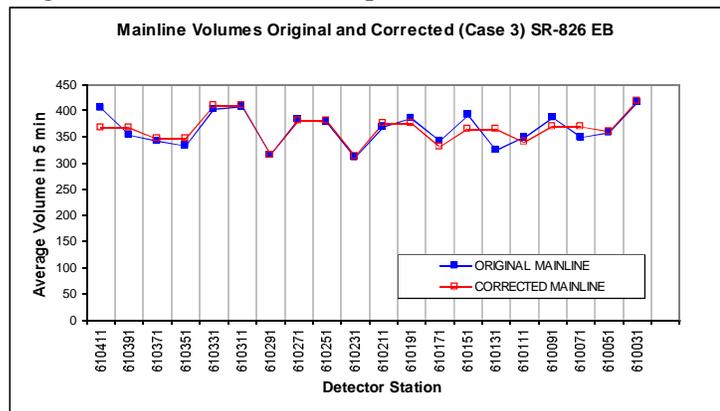
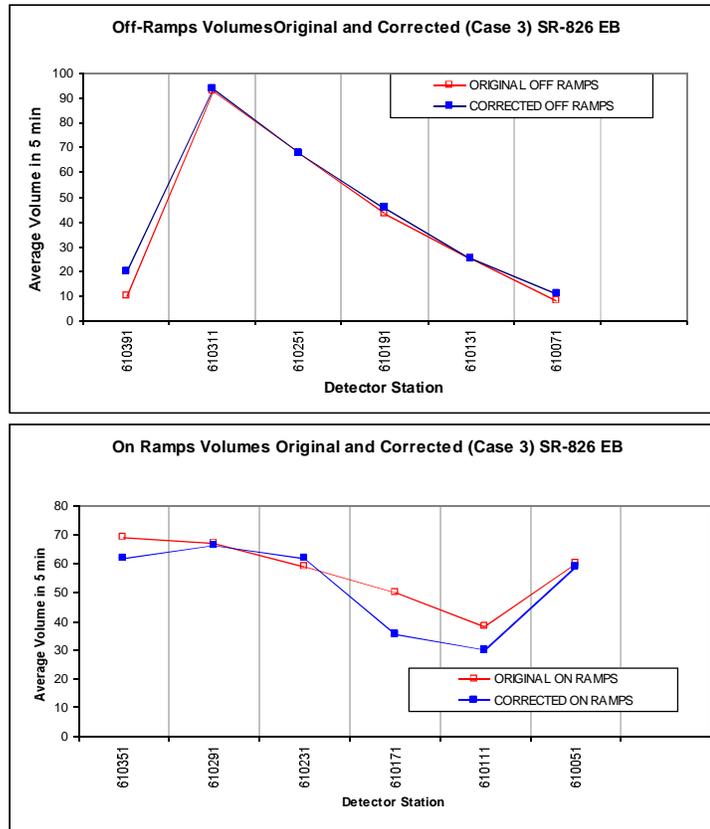


Figure 7 shows the results of Case 3, in which this last detector station which has significant inconsistency with the rest of the system was removed from the optimization. Notice how removing this detector station reduces the correction needed in adjacent stations. Manual counts for a short period of time confirmed that the last detector had data quality problems. The above results indicate that it is useful for the analyst to examine the results and revise the inputs to the optimization process, if the results from the optimization show significant corrections due to one or two suspicious detection station measurements.

Figure 7: Mainline and ramp volume corrections for case 3





The results presented above also provide additional insights regarding the impacts of data quality reported by the data warehouse. If significant errors remain in the data, the quality of the results will be affected. Further research is needed to determine the minimum data quality requirement to produce acceptable results.

6. CONCLUSIONS

This paper has illustrated the development and application of a series of data manipulation procedures for the utilization of ITS data archives to support simulation modeling. The procedures allow the extraction of collected volume data from ITS data archives, automatic identification of temporal patterns in the data, automatic segmentation of daily demands into dynamically captured sub-periods to best fit the variations in the demands, resolving possible spatial inconsistencies in the data, and estimating missing volumes. The developed procedures have been implemented as an automated tool for simulation model generation. The tool provides a graphic interface for end users to download data from the STEWARD data warehouse, identify and select ideal traffic patterns, perform segmentation on traffic demands, conduct spatial conciliation to reconcile data inconsistency and estimate missing volumes, and generate new simulation model files based on the purified data. The procedures and the developed tool can easily be adapted by other traffic agencies to interface with their ITS data archives. Although the main objective of the developed procedures and the tool is to produce data for microscopic simulation applications, they can also be used to support other applications such as macroscopic, mesoscopic, and demand forecasting modeling applications.

Acknowledgement

The work reported in this paper was conducted as part of a project funded by the Florida Department of Transportation Research Center. The authors gratefully acknowledge the FDOT support of this effort.

References

- Alpaydin, E. (2004) **Introduction to Machine Learning**. The MIT Press, Massachusetts.
- Barceló, J. García, D. and Kirschfink, H. (2002) Scenario Analysis a Simulation Based Tool for Regional Strategic Traffic Management. **9th World Conference on Intelligent Transport Systems**, Chicago.
- Barceló, J. Kirschfink, H. and Torday, A. (2003) An Integrated Software Platform to Assist Advanced Traffic Management Decisions. **10th World Conference on ITS**, Madrid.
- Federal Highway Administration (2004) **Archived Data Management Systems - A Cross-Cutting Study**. Publication FHWA-JPO-05-044. FHWA, U.S. Department of Transportation.
- Gomez, G., Horowitz, R. and May A. (2004) **Calibration of VISSIM for a Congested Freeway**. Report UCB-ITS-PRR-2004-4, University of California, Berkeley.
- Hillier, F., Frederick S., Gerald J. and Lieberman K. (2004) **Introduction to Operations Research**. McGraw-Hill, New York.
- Chu, S. Hart, D. Keogh, E. and Pazzani, M. (2001) An Online Algorithm for Segmenting Time Series, **International Conference on Data Mining**.
- Hourdos, J. Michalopoulos, P. and Xin, W. (2006) Preprocessing Volume Input Data for Improved Traffic Simulation. **Transportation Research Record: Journal of the Transportation Research Board**, No. 1965, Transportation Research Board of the National Academies, Washington D.C., pp. 192-200.