COMPARISON OF VARIOUS MODELING APPROACHES FOR THE IMPUTATION OF INTELLIGENT TRANSPORTATION SYSTEM VDS DATA

Authors:

Felipe Castrillon *
Ph.D. Student
Phone: 954-651-7531
E-Mail: fcastrillon3@gatech.edu

Angshuman Guin, Ph.D.
Senior Research Engineer
angshuman.guin@ce.gatech.edu

Randall Guensler, Ph.D.
Professor
randall.guensler@ce.gatech.edu

Jorge Laval, Ph.D.
Assistant Professor
jorge.laval@ce.gatech.edu

School of Civil and Environmental Engineering
Georgia Institute of Technology
790 Atlantic Drive
Atlanta, GA 30332

*corresponding author

Word Count: xxxx (text: xxxx; tables: 2 x 250: 500; figures: 6 x 250: 1500)

Submission Date: xxxxxx

ABSTRACT

Gaps in real-time and archived traffic data are very common and can be attributable to several different factors, such as sensor failures, data communications interruptions, etc. Regardless of the cause of the missing data, these gaps often have to be filled with reliable and accurate estimates before the data can be used for planning, operations, or congestion mitigation purposes. This research compares different methods for imputing missing values on Video Detection System (VDS) data, including historical averages, simple-linear regression, multiple-linear regression, spatial average and Newell's simplified kinematic wave model. The study uses the fundamental relationship between speed and flow in filtering the data for quality control. A sensitivity analysis is performed to test the response of different methods to factors such as the size of training dataset, time-of-day adjustments to the algorithms and others. The results indicate that the time of day and volume adjustment factors have a non-trivial impact on the accuracy of the outputs. It was found that in spite of the significant errors in the base dataset, the Newell algorithm performs on par with the other methods in terms of the bias and the mean absolute percentage, but that the more simple factoring methods also provide comparable results and are easier to implement.

INTRODUCTION

Replacing or imputing missing values in Intelligent Transportation System (ITS) data is important for many applications. This is especially true for real-time applications, such as information displayed on variable message signs (VMS) to provide updates to drivers about travel times, congestion etc. as part of an Advanced Traveler Information System (ATIS). In addition, a coherent and comprehensive dataset is important from the planning perspective as well, where aggregated volume counts are used since these are typically more sensitive to the impact of error compounding by accumulation.

Missing traffic data are typical of most traffic detection systems. A study in San Antonio found that missing data ranged from 5-25%, even though only 5-15% of this missing data comes from loop detector failure (*1*). A study in Georgia found an average missing rate of 4-14% of data for GA 400 (*2*). Also, 25-30% of single-loop detectors at intersections were found to be offline at any given time in a study done in Virginia (*3*). According to ongoing research in Purdue University, detector loops not installed properly can have count errors greater than +/- 15 percent (*4*).

Data outages can be of different types and extents. Data can be missing at the individual lane detector level or at the detection station (multiple lane) level. Missing lane detector data can occur when one or more loop detectors are broken due to construction, maintenance operations, breakage of loop wires and other reasons. Missing detector station level data can occur when a controller, to which multiple detectors are connected, loses power or the communication link to the data server. Video detection systems (VDS) that use virtual loops drawn on a single camera view can yield missing detector station data with the outage or error of a single camera.

Missing data can be classified as random or block-level. Missing data that are random in nature can result from short-term software or hardware malfunctions that can be difficult to identify. Missing block-level data can occur for a period of half a day, a few weeks or months, or even for a whole year. Block outages can be caused by construction or long periods of malfunctioning hardware or software (*5*).

Different imputation techniques have their own strengths and weaknesses and can have different performance levels under different data availability scenarios. For example, certain imputation techniques require larger data sets to train the algorithm. Also, some algorithms might be more robust to longer outages, or to spatial outages on a whole corridor. Plus, some methods may be more sensitive to input data errors than other methods.

The purpose of this paper is to compare different infill algorithms on station-level VDS volume data based on imputed error measures and applicability. Novel algorithms are developed in order to reduce the cross-station volume differences observed from the VDS volume counts. Also, methods from past research efforts are included to provide some error comparisons across a variety of algorithms.

LITERATURE REVIEW

The body of literature examines various methods that are computationally simple and heuristic. These methods are most commonly employed by transportation agencies to impute missing data values and include historical methods, spatial linear interpolations using nearby detectors, temporal interpolations using past and future data from the same detector, and more complicated methods such as weighted moving averages. Zhong, et al., suggest that the more simplistic methods have very low performance and more complex methods generate much better results and are more robust to existing traffic data (*6*).

Advanced statistical methods have been developed to reduce the errors of heuristic methods. Zhong, et al., employed the Box-Jenkins forecasting procedure (ARIMA models), and a pattern-matching algorithm, that compares various candidate curves with the curve under study, along with some simple imputation methods. The pattern matching algorithm can outperform the ARIMA model and the more simplistic methods in most hours during the day based on AAPE measures (*6*). In one effort, the researchers applied two types of regression (polynomial and kernel) in combination with an Expectation Maximization (EM). The EM algorithm provided a narrow improvement over the Marginal Absolute Percent Errors (MAPE) and Root-Mean Square Error (RMSE) measures, but might not be feasible due to the extra time and effort needed to employ (*7*). Smith, et al. compared simple approaches to two advanced methods, Expectation Maximization (EM) and Data Augmentation (DA). These advanced methods iteratively calculate regression equations and imputation values until the mean and covariance converge. The more advanced methods perform better based on MAPE values, but can be unreasonable for some applications due to their complexity (*8*). Kwon combined linear regression and Normal Bayesian imputation to come up with a Nonormal Bayesian Linear Regression (NBLR) algorithm for randomly missing data and block-level missing data. The algorithm is not compared to other methods in the study (*4*). Ni and Leonard develop a stochastic method based on Bayesian networks that is able to produce unbiased estimates and preserve the natural characteristics of the raw data (i.e. variability within variables and relationships among variables) (*2*).

Other types of methods use genetic algorithms to learn from the differences in traffic patterns. Zhong, et al. developed two genetic algorithms, locally weighted regression and time delay neural networks, to estimate imputation accuracy for one-hour. This study found that the locally weighted regression models provided the most accurate results of all of the models tested and the time-delay neural networks yielded very large errors, which might be caused by the small amount of training and test data available (*9*). Chen, et al. compared the imputation accuracy of an advanced neural network model, a hybrid algorithm that integrates the historical and average and time series analysis, and various simple methods. The authors suggest that the time-delay neural network is less applicable than other methods because it has large data requirements and did not provide much improvement over other methods based on RMSE and MAPE values. The study also found that for different missing percentage rates, the relative performance of the algorithms was the same, even though the MAPE and RMSE measures differed. Furthermore, the day-of-the-week factor did not impact significantly the accuracy of the replaced values (*10*).

Other types of imputation methods include regression models based upon spatial relationships with other detector stations. Al-Deek and Chandra compared a pairwise linear regression model, and two pairwise quadratic regression models for data imputation accuracy. The quadratic models outperformed the linear models since they take into account the nonlinear relationships between speed, volume and occupancy at different regions of volume (*11*). Another study implemented the simple linear regression and multiple regression models developed by Al-Deek and Chandra, as well as the non-normal Bayesian linear regression (NBLR) technique previously mentioned. Other algorithms are included such as local and global regression, historical imputation, and fuzzy imputation. The study results indicated that the regression based models are the most accurate, with the multiple regression model and the local regression model having the lowest RMSE compared to ground truth data. The methods using historical data, such as the historical mean, and the fuzzy imputation, did not yield results that were as accurate as the regression models (*12*).

An additional method developed by Haj-Salem and LeBacque uses traffic theory methods by applying the Lighthill-Whitham-Ricards (LWR) first-order model that uses an exponential functional form for the fundamental diagram. The method is shown to have better performance than other algorithms based on a chosen set of data which reflects real data measurements (*13*). In this study, we use a triangular fundamental diagram, in accordance with empirical observations and use an exact numerical solution method, i.e., Newell's simplified kinematic wave model (*14,15,16,17*).

QA/QC methods have been applied during past research efforts, with most methods flagging impossible values for volume, speed and occupancy (*4,17*). Other efforts have employed traffic volume relationships between volume speed and occupancy to establish thresholds for highly improbable values (*11,18*). The QA/QC filters developed in this paper are based on the methods from Turner, et al. (*18*). Another study by Turner gives a breakdown of the various QA/QC methods that have been employed in past research efforts and recommends data filtering thresholds (*17*).

METHODOLOGY

Data set

A set of three consecutive detection stations from the state road GA400 corridor in metro Atlanta is chosen as the case study for the imputation methodology comparisons. The corridor is congested during both peak periods, but experiences more congestion during the inbound morning peak period. To limit the number of confounding factors, the section chosen includes no exits or entrances between the detection stations. The study corridor consists of three four-lane detection stations, at mile posts 11.47 10.77and 10.11 respectively.

The data used in this study was obtained from the Georgia NaviGAtor system which uses VDS as the primary detection technology. Spots speeds and counts are reported as 20 second aggregates on a per-lane basis at detection stations that are located approximately 1/3 miles apart.

Station-level detector data (aggregate over all lanes at the station) are used in this study, based upon the observation that in a VDS system, such as Georgia NaviGAtor, an overwhelming majority of missing data occurs at the station-level. This is understandable because in most systems, all lane detectors are connected to a single video processing unit fed by a single camera view. Failure of either equipment component causes the complete failure of the detector station. In addition, it can be contended that a complete outage at a station is typically more severe and is also a more difficult problem to resolve.

Light-volume periods such as weekends, holidays and late-night and early-morning data points are not included in this study because those data are either not critical to most planning and operations studies or are treated as exogenous factors in such studies. Data points between 5 AM and 10 PM on weekdays are used in this study.

Quality Assurance/Quality Control

Highly improbable data values are removed from the data set and replaced with null values using filtering scripts applied to the reported 20-second bin station-level data. The conditional logic and the thresholds used in data filtering are provided in TABLE 1. For example, data for which average vehicle speeds for 20-second bins exceed 110 mph are removed from the data stream, as are 20-second traffic volumes that exceed 20 vehicles (3600 vehicles/hour).

Before applying QA/QC, 99.3% of the data points are available for training data, and

98.3% are available for testing data. After applying data filtering methods, 96.8% of the data points are available for training, and 98.03% are available for testing.

(Insert TABLE 1)

Experimental Setup

Independent data sets are used to train and then to test the algorithms. Testing data constitutes two weeks in May 2009. Two sets of training data are chosen from April and May 2009 previous to the testing data. The first training set covers four weeks and the second training set takes out two of the weeks from the first training data set. In the experiments we compare the results for models developed from four weeks of training data vs. two weeks of training data to test for sensitivity of the models to the size of the training dataset.

To test the performance of the imputation algorithms, artificial gaps are created in the testing data from the center station by removing data points from the original testing dataset. Algorithms are calibrated using training data for all detectors. Once calibrated, the algorithms are used to predict the missing values for the testing data. The replacement values are then compared to the original data using statistical measures such as root-mean square error (RMSE), mean absolute percent error (MAPE), mean absolute error (MAE) and statistical bias (BIAS).

Testing is performed for three periods during the day. The morning peak consists of 6:30AM to 9:30AM, the mid-day off-peak is from 9:30AM to 4:30PM, and afternoon peak is from 4:30 PM to 7:00 PM. A full blackout is reproduced for the center station to test the algorithms. Because the algorithms do not depend on temporal data for the center station, comparable results should be achieved in simulating randomly missing data or block-level missing data.

Imputation is performed at the 20-second level, considering that any non-imputed aggregated data will incorporate missing values and pre-imputed values. To avoid using data that have already been manipulated, imputation is performed at the raw data level. Because most analysis and applications use data at a minimum of 5-minute aggregation interval, the imputation performance is examined at 5 minute and 1-hour aggregations in addition to the 20-second level. However, the imputation is still performed at the 20-second level for the 5-minute and 1-hour aggregation comparisons.

Algorithms

Previous research efforts have indicated that historical methods are not very accurate compared to other methods. The historical model is included in our analysis because it serves as a benchmark for comparison, has been used extensively in past research efforts, and is still employed by many agencies as their primary data imputation method because of its inherent simplicity of application. Regression models developed by Al-Deek, et al. (*11*) are included because they have shown in the past to have better performance than other more complicated historical and regression methods (*12*) without an extensive calibration effort. Also, the Al-Deek, et al. models overcome some of the large data requirements required for other regression models by being more robust to failures from surrounding stations.

Some of the algorithms developed in this study are designed to reduce the cross-station differences in volumes and counts. Figure 1 displays an example of a single day cumulative vehicle count curve for the three stations. This figure is displayed as an oblique plot to illustrate the differences more clearly, where the curves are subtracted by the average slope line of the center station count. The counts have a time-delay factor such that the upstream counts begin at a free volume region at initial time *t*, while the middle station starts counting at $t + \delta_u / u$ and the downstream station begins at $t+(\delta_{u+} \delta_d)/ u$. With the time translation of the starting time of counts, it is expected that cumulative counts are roughly equal during free volume regions, and any upstream station counts are slightly higher than any downstream station counts at congested periods.

The upstream station exhibits lower cumulative counts than the other stations, while the downstream station exhibits higher cumulative counts. This happens pretty consistently for every day in the data set. Given that there are no entrances and exits, one expects conservation of vehicles through the section, which would yield close to overlapping curves for the three stations. The large difference between the station cumulative traffic counts (0-8% between upstream and middle, and 8-18% between downstream and middle for a single day) suggests that there are substantial detector errors which are not consistent across stations and may vary by time period. The differences are most likely due to video occlusion or video system splashover errors where larger vehicles are detected by detectors in their own lane as well as in adjacent lanes, as a function of different camera viewing angles and detector configurations. Hence, as large truck volumes and traffic densities vary throughout the day, departures in adjacent lane detector values can also vary.

(Insert FIGURE 1)

The volume differences across stations were confirmed to differ greatly by time of day. Figure 2 shows the cross-station differences for volumes throughout the day for one particular day. All days in the testing period exhibit a similar behavior. From this observation, two methods to adjust the volumes before using the imputation algorithm are implemented to lessen the cross-stations errors observed in the VDS data:

- Period-of-day factors (P)- Considering that the volume difference is higher during peak periods, this method uses three correction factors calibrated from the training data; one each for the morning-peak, mid-day off-peak, and the afternoon-peak. These factors are obtained from the cumulative counts of the training data and are calibrated so that the final counts during each period are equal. These factors are then used to scale the upstream/downstream volumes in the testing data to exhibit roughly similar vehicle counts at all stations.
- Regression factors (R) - In this method a regression analysis is performed to develop a relationship between the speed and the volume of the upstream/downstream station to the volumes of the middle station. After the regression weights are calibrated, these are applied to the volumes and speeds of the upstream/downstream station in order to estimate a volume for the center station.

Note that the adjustments affect only the inputs to the imputation algorithm and not the test data. The outputs from the imputation algorithm are still compared to the original data from the test dataset to quantify the algorithm performance.

(Insert FIGURE 2)

Two algorithms are applied to the two sets of volume adjusted data. One will take the average of the two neighboring upstream and downstream station detector volumes. The other method will be developed from Newell's simplified kinematic wave model, which is a simplified version of the LWR first-order model using "N-curves" or cumulative vehicle count curves. Finally, two additional regression algorithms will be developed by extending the Al-Deek and Chandra models to include dummy variables for peak periods to adjust for the noted changes in cross-station volume differences during different periods. The detailed structure of each of the nine imputation approaches compared in this study is outlined in the following subsections:

Historical (HS) Model
This algorithm replaces values using historical average values from the same weekday and time period. This method is sometimes the only method available when there is widespread spatial outage on a corridor.

Simple Linear Regression (LR) Model
This method uses the predicted average result from various simple linear regression models. Past research has employed this method at the individual lane detector level, but for our application it is employed at the detection station level. The method takes the average result from two regression models (one upstream and one downstream) or one result if data from one of the adjacent stations are missing.

LR1 model The original linear regression model is represented as:

$$v_d^i = B_0^i + B_1^i v^i \qquad (1)$$
$$v_d = \frac{1}{k}\sum_i^k v_d^i \quad (2)$$

Where:

i = upstream or downstream station level detector,
d = testing detector station (center station level detector),
k = number of stations excluding the testing station,
v = volume
B = regression coefficients

LR2 model The modified linear regression model is an extension of the original linear regression model, but includes dummy variables for peak-periods. The model is presented below:

$$v_d^i = B_0^i + B_1^i v^i + B_2^i m_1 + B_3^i m_2 + B_4^i(m_1 * v^i) + B_5^i(m_2 * v^i) \qquad (3)$$
$$v_d = \frac{1}{k}\sum_i^k v_d^i \quad (4)$$

Where:

m_1 = dummy variable for morning peak
m_2 = dummy variable for afternoon peak

Multiple Linear Regression (MR) Model
This method uses the predicted average result from multiple regression models. Just like the previous method, there are two regression models employed, and the predicted result provides the value that is used to replace missing data. Unlike the simple linear regression model, this method is able to capture some of the quadratic, or cross-term, relationships between speed, density and volume.

MR1 Model The original multiple linear regression model is represented as:

$$v_d^i = B_0^i + B_{\mathbf{1}}^i * v^i + B_2^i\left(v^i\right)^2 + B_3^i s^i + B_4^i (s^i)^2 + B_5^i * \left(v^i * s^i\right) \qquad (5)$$

$$v_d = \frac{1}{k}\sum_i^k v_d^i \qquad (6)$$

Where:

s = speed,

MR2 model The modified multiple linear regression model extends the original multiple regression model by including dummy variables for peak-periods. The model is presented below:

$$v_d^i = B_0^i + B_{\mathbf{1}}^i * v^i + B_2^i\left(v^i\right)^2 + B_3^i s^i + B_4^i (s^i)^2 + B_5^i * \left(v^i * s^i\right) + B_6^i m_1 + B_7^i m_2 + B_8^i (m_1 * v^i) + B_9^i (m_2 * v^i) + B_{10}^i (m_1 * s^i) + B_{11}^i (m_2 * s^i)) \qquad (7)$$

$$v_d = \frac{1}{k}\sum_i^k v_d^i \qquad (8)$$

Newell's simplified kinematic wave model with Adjusted Volumes

Newell's method is a special case of the kinematic wave model *(14).* The model assumes a triangular fundamental diagram between volume and density and requires that boundary at an upstream and a downstream location are known. A solution to the problem is found in *(15)* and later demonstrated using variational theory in (*16,17*). The middle-count solution to the "three-detector method" for a homogenous roadway with boundary data known for two locations x_u *and* x_d, is given by:

$$N_{cs}(t) = min\left\{N_{us}\left(\tau_{us}(t)\right), N_{ds}\left(\tau_{ds}(t)\right) + \kappa\delta_{ds}\right\} \quad (9)$$

$$\tau_{us}(t) = t - \delta_{us}/u \quad (9a)$$

$$\tau_{ds}(t) = t - \delta_{ds}/w \quad (9b)$$

$$\delta_{us} = x_{ms} - x_{us} \quad (9c)$$

$$\delta_{ds} = x_{ds} - x_{us} \quad (9d)$$

Where:

us = upstream station,
ds = downstream station,
cs = center station
N = vehicle daily counts,
w = congestion wave speed
u = free-volume speed
δ = absolute value of distance between two stations
x = location of station
κ = density
$\tau(t)$ = time translation of wave speed

For the purpose of this paper, to estimate the volumes for a given time period, the difference between counts is obtained:

$$v_{cs}(t) = N_{cs}(t) - N_{cs}(t-1) \qquad (10)$$

Newell's method requires that cumulative count curves or "N-curves" are consistent, such that the conservation of vehicles is followed and that curves do not diverge. This condition is not met in the study data (Figure 1). To reduce some of these errors, the volume adjusted curves are used in the analysis.
As previously discussed, a study using first-order LWR models has found the model to provide accurate results on real data measurements (*13*) when the underlying data are very accurate. Here, the model is evaluated using VDS data with known accuracy issues by using a volume adjustment strategy that mitigates the impact of these errors.

NW-P model This model will apply the Newell's method to volumes adjusted with the period-of-day factors.

NW-R model This model will apply the Newell's method to volumes adjusted with the regression factors.

Spatial Average Model with Adjusted Volumes
The method takes the average of the upstream and downstream adjusted volumes to infill a value for the center station. It is developed to study the impact of the volume factors on the Newell method since an application of the Newell method on the unadjusted volumes is not expected to produce meaningful results.
AVG-P This model will apply a simple average from two neighboring stations to the volumes adjusted with the period-of-day factors
AVG-R This model will apply a simple average from two neighboring stations to volumes adjusted with regression factors.

Summary of Algorithms
Overall there are a total of nine algorithms analyzed in this study. Three of them come from previous research methods, while five of them are developed by combining some previous methods to adjustments to cross-station volume differences. The nine algorithms are highlighted below:

1. HS-historical average, based upon raw detector training data and applied to raw testing data
2. LR1- simple linear regression model, calibrated using raw detector training data and applied to raw testing data
3. LR2- simple linear regression model with peak period dummy variables, calibrated using raw detector training data and applied to raw testing data
4. MR1- multiple regression model, calibrated using raw detector training data and applied to raw testing data
5. MR2- multiple regression model with peak period dummy variables, calibrated using raw detector training data and applied to raw testing data
6. NW-P - Newell's method with "period-of –day factor" adjusted curves, calibrated using Method A adjusted detector training data, applied to adjusted testing data
7. NW-R - Newell's method with Method B adjusted curves, calibrated using "regression factor" adjusted detector training data, applied to adjusted testing data

8. AVG-P- Upstream and downstream factored station averages with “period-of –day factor” adjusted curves, calibrated using adjusted detector training data, applied to adjusted testing data from the downstream and upstream stations.
9. AVG-R- Upstream and downstream factored station averages with “regression factor” adjusted curves, calibrated using adjusted detector training data, applied to adjusted testing data from the downstream and upstream stations.

RESULTS

The results below are obtained from the 2-week testing data, and the 4-week training dataset. Table 2 shows the results for morning peak, mid-day off-peak and afternoon peak for errors at 20 second, 5 minute and 1 hour aggregation levels. At the 20 second level, all models have comparable results based on MAPE and RMSE except for the historical method during peak periods. This is especially true for the regression models. The LR2 and MR2 show the lowest comparable RMSE and MAPE in the morning and some of the lowest in the mid-day off-peak and afternoon off-peak. The AVG-P method performs much better than the NW-P, and the AVG-R performs better than the NW-R for all three periods of the day.

For the 5-minute data, the error values reduce as expected, since aggregation decreases some of the jitter in the 20 second measurements. It is interesting to note that the relative performance between algorithms changes for the aggregated data. For instance, the algorithms that use the “period-of-day” volume adjustment factors have lower MAPE and RMSE than most of the regression methods for peak periods, especially the AVG-P method. Generally, the methods that exhibit lower bias for 20 second data improve their relative performance for aggregated data. This is expected because bias errors get added to the aggregated estimates. The LR2 and MR2 methods show very similar performance and perform better than the LR1 and MR1 models during the morning period. Surprisingly, they perform worse during the mid-day and afternoon peak. The historical method has the highest MAPE and RMSE throughout the day.

For the 1-hour aggregation, the relative performances of the algorithms do not change much to the 5-minute performances. The NW-P and AVG-P models have the lowest MAPE and RMSE for all time periods, except for LR1 during the mid-day off-peak.

(Insert TABLE 2)

Figure 3(a) shows the 1 hour averages for 20-second MAPE for the historical models and regression algorithms, while Figure 3(b) shows the same results for the NW and the AVG methods, as well as the LR2 model as comparison reference. These figures show that the LR2 and MR2 models in general perform better than the LR1 and MR1 models. It is also observed that the MAPE values of the LR2 and the MR2 models are very similar. The AVG models can be observed to perform better than the NW models for all time periods. The NW-R and AVG-R models have performance numbers very similar to the LR2 (and MR2) models.

Figure 4(a) shows the 1-hour averages for 5-minute MAPE for the historical models and regression algorithms, while Figure 4(b) shows the same results for the NW and the AVG methods, as well as the LR1 model as a comparison reference. Here, we can clearly see the higher MAPE for the MR2 and LR2 models during the morning peak, but higher MAPE during the other time periods compared to the LR1 and MR1 models. The AVG-P model has some of the lowest MAPE throughout the day, especially during the peak periods.

(Insert FIGURE 3)

(Insert FIGURE 4)

Figure 5 shows the differences in 5 minute MAPE and MAE between 4 week training data and 2 week training data. A positive value on the charts represents an improvement in performance by using 4 weeks of training data instead of 2 weeks. The only method that shows a noticeable improvement in performance by increasing the training dataset from 2 weeks to 4 weeks is the historical model during the morning peak period. In fact, most models show a small decrease in performance. The differences might be caused by over-training of data where the additional two weeks have some differences in traffic patterns to the other weeks of data. Further in-depth studies with a wider range of lengths of datasets need to be performed in order to obtain a better estimate of the impact of the amount of training data on the performance of the methods.

(Insert FIGURE 5)

Figure 6 shows the frequency distribution of the percentage errors for each method. The errors do not appear to be normally distributed but the distributions follow a similar pattern. The research team will be undertaking more detailed analysis of the residuals in the near future.

(Insert FIGURE 6)

CONCLUSIONS

All of the methods compared in this paper are simple to apply and need only one run each to calibrate the parameters. In terms of robustness, all methods can impute values if there is a complete outage (temporal) in one station. However, if there is an area-wide outage (spatial) over a whole corridor, all of the models fail, except for the historical method. If the models are calibrated to include more stations, then the models will be more robust to failures, but their performance might suffer. It might be worthwhile to study if the alternative approach where the imputations are based on imputes at adjacent stations when there are no inputs available, performs better than models with increased number of stations.

The volume adjusted methods' performance is found to be comparable to the regression models when comparing the 20-seconds errors. However, the scaling factor model shows better performance of the "period-of-day" methods with the 5-minute and 1-hour aggregations, especially for AVG-P method. It can be noted that in general the methods with low bias values have better and more consistent performance throughout the different aggregation levels. The Newell methods are more complex and computationally intensive, but do not appear to show an improvement over the scaling and regression factor methods, which are very simple to calibrate and execute for this application. Additional research is also required to study the effectiveness of the imputation strategies when the models are extended to incorporate ramps.

The regression models with dummy variables were observed to perform better on an average than the regression models for the 20-second errors and the 5-minute errors during the morning peak. The LR2 and MR2 models show promise for future development.

Imputation at the station-level is important for VDS data for the reasons previously mentioned. For loop detector data, the regression models are still applicable at the lane detector level.

Recommendations

The research team recommends that practitioners test some of the best performing algorithms in their own data, due to differences in types of data. If the practitioner is more interested in raw data applications, regression models can be examined since they show good relative performance in this paper and previous research efforts. If aggregated data applications are desired, the practitioner must choose a model with low bias at the raw-data level (e.g. AVG-P). Finally, we recommend choosing a back-up method that is robust in case of any failures, and is applicable within given resources.

REFERENCES

1. Turner, S., L. Albert, B. Gajewski and W. Eisele. Archived Intelligent Transportation System Data Quality. In *Transportation Research Record: Journal of the Transportation Research Board*, No 1719, Transportation Research Board of the National Academies, Washington, D.C., 2000, pp. 77-84.

2. Ni, D., and J.D. Leonard II. Markov Chain Monte Carlo Multiple Imputation using Bayesian Networks for Incomplete Intelligent Transportation Systems Data. In *Transportation Research Record: Journal of the Transportation Research Board*, No 1935, Transportation Research Board of the National Academies, Washington, D.C., 2005, pp. 57-67.

3. Nguyen, L., and W. T. Scherer. *Imputation Techniques to Account for Missing Data in Support of Intelligent Transportation Systems*. Center for Transportation Studies, University of Virginia Report No. UVACTS-13-0-78. Office of University Programs, Research and Special Programs Administration, U.S. Department of Transportation, Washington, D.C., 2003.

4. *Using Signalized Loops at Signalized Intersections for Traffic Counts*. ITE Traffic Engineering Council, Publication No. IR-122, 2007.

5. Kwon, T. M. *TMC Traffic Data Automation for Mn/DOT's Traffic Monitoring Program*. University of Minnesota Duluth Report No. 2004-29. Minnesota Department of Transportation, St. Paul, Missnesota, 2004.

6. Zhong, M., S. Sharma, and Z. Liu. Assessing Robustness of Imputation Models Based on Data from Different Jurisdictions: Examples of Alberta and Saskatchewan, Canada. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1917, Transportation Research Board of the National Academies, Washington, D.C., 2003, pp. 116-126.

7. Zhong, M., and S. Sharma. Matching Hourly, Daily and Monthly Traffic Patterns to Estimate Missing Volume Data. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1957, Transportation Research Board of the National Academies, Washington, D.C., 2006, pp. 32-42.

8. Gold , D., S. Turner, B. Gajewski and C, Spiegelman. Imputing Missing Values in ITS Data Archives for Intervals Under 5 Minutes. Presented at the 80th Annual Meeting of the Transportation Research Board, Washington, D.C., 2001.

9. Smith, B., W. Scherer, and J. Conklin. Exploring Imputation Techniques for Missing Data in Transportation Management Systems. In *Transportation Research Record: Journal of Transportation Research Board*, No. 1836, Transportation Research Board of the National Academies, Washington, D.C., 2003, pp. 132-142.

10. Zhong, M., S. Sharma, and P. Lingras. Genetically Designed Models for Accurate Imputation of Missing Traffic Counts. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1879, Transportation Research Board of the National Academies, Washington, D.C., 2004, pp. 71-79.

11. Chen, M., J. Xia, R. Liu. Developing a Strategy for Imputing Missing Traffic Volume Data. *Journal of the Transportation Research Forum.* Vol. 45, No. 3, 2006, pp. 59-77.

12. Al-Deek, H. M., and C. Chandra. New Algorithms for Filtering and Imputation of Real-Time and Archived Dual-Loop Detector Data in I-4 Data Warehouse. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1867, Transportation Research Board of the National Academies, Washington, D.C., 2004, pp. 116-126.

13. Boyles, S. Comparison of Interpolation Methods for Missing Traffic Volume Data. Presented at the 90th Annual Meeting of the Transportation Research Board, Washington, D.C., 2011.

14. Haj-Salem, H., J.P. Lebacque. Reconstruction of false and missing data with first-order traffic volume model. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1802, Transportation Research Board of the National Academies, Washington, D.C., 2002, pp. 155-165.

15. Lighthill, M. J., Whitham, G. On Kinematic Waves. I Volume Movement in Long Rivers. II A Theory of Traffic Volume on Long Crowded Roads. *Proceedings of the Royal Society of London,* vol. 229 (A), 1955, pp. 281–345.

16. Newell, G. F. A Simplified Theory of Kinematic Waves in Highway Traffic, I General Theory, II Queuing at Freeway Bottlenecks, III Multi-destination Volumes. *Transportation Research Part B*, Vol. 27, No. 4, 1993, 281–313.

17. Daganzo, C. F. A Variational Formulation of Kinematic Wave Theory: Basic Theory and Complex Boundary Conditions. *Transportation Research Part B*, Vol. 39, No. 2, 2005, pp. 187–196.

18. Daganzo, C. F. A Variational Formulation of Kinematic Waves: Solution Methods. *Transportation Research Part B: Methodological* Vol. 39, No. 10, 2005, pp. 934 – 950.

19. Turner, S., Archived Intelligent Transportation System Data Quality: Preliminary Analysis of San Antonio TransGuida Data. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1719, Transportation Research Board of the National Academies, Washington, D.C., 2000, pp. 77-84.

List of Tables and Figures

TABLE 1 QA/QC values and thresholds

Thresholds		
Volume (veh/20sec)	**Speed (mph)**	**Density (veh/mi)**
(Two conditions must be true to be declared invalid)		
Zero (=0)	Zero (=0)	Zero (=0)
(All conditions must be true in order to be declared invalid)		
Any	Low (<20)	Medium (<120, >= 40)
High (>= 15)	Any	Low(< 40) or High(>200),
Nearly Zero (<= 2)	Nearly Zero (<=10)	Not Nearly Zero (>=8)
Too High (>= 20)	Any	Any
Any	Any	Too High (>290)
Any	Too High (>110)	Any

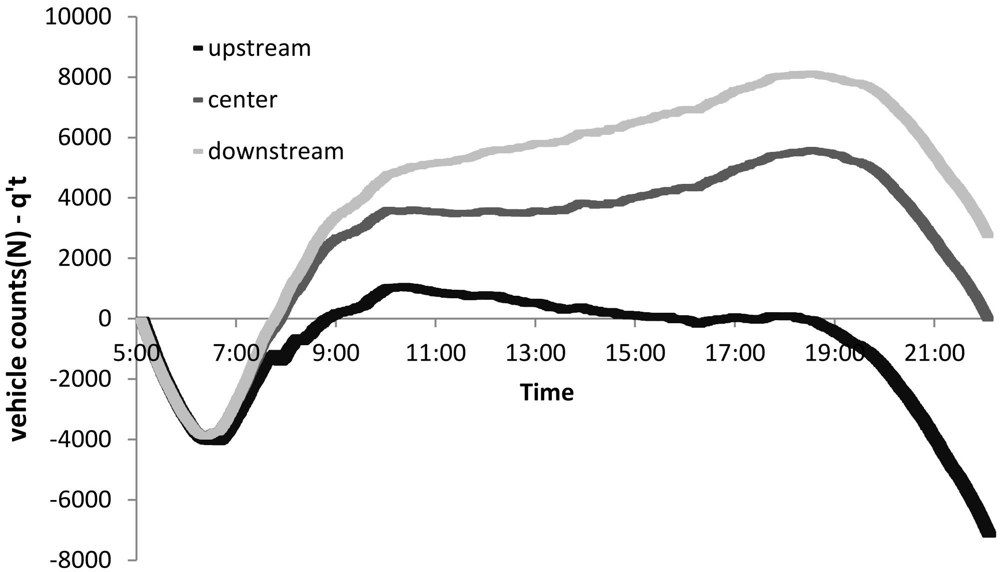


FIGURE 1 Cumulative vehicle counts for May 8th, 2009 (Oblique plots)

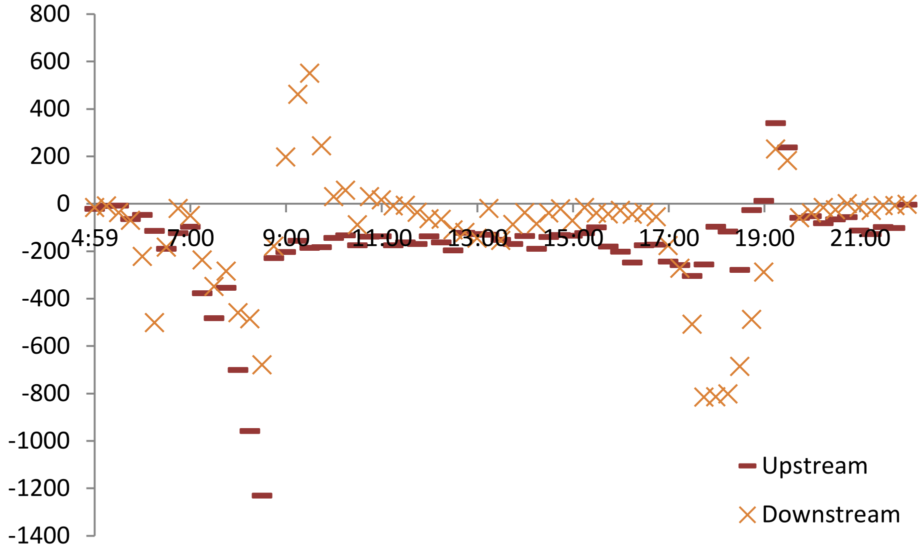
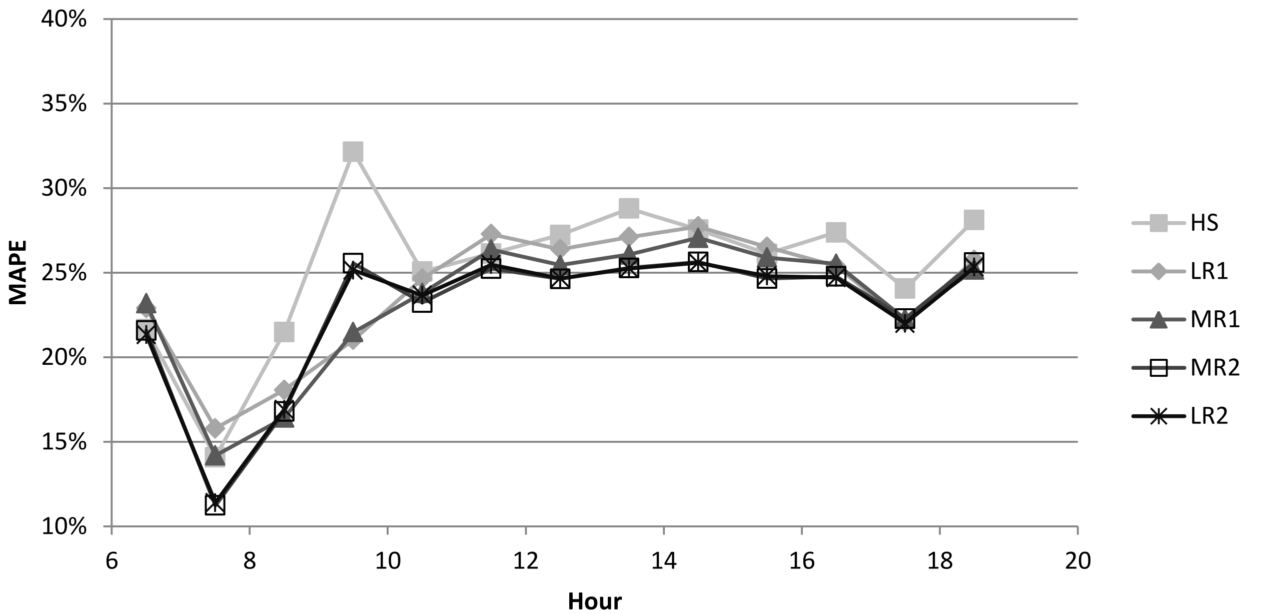


FIGURE 2- Differences in 15 minute volumes between upstream and center, and down-stream and center stations on one day from hours 5:00 to 22:00

TABLE 2 Results for 4 week training data

	20 second				*5 minutes*		*1 hour*	
	Morning Peak				**Morning Peak**		**Morning Peak**	
	RMSE	MAPE	MAE	BIAS	RMSE	MAPE	RMSE	MAPE
HS	9.7	19.2%	7.5	-0.3	103.3	12.0%	888.6	9.3%
LR1	9.8	17.4%	7.8	-5.4	104.1	11.8%	1087.5	11.7%
LR2	8.2	16.4%	6.3	0.6	76.0	8.8%	629.6	6.3%
MR1	9.4	16.5%	7.2	-4.2	96.0	10.8%	953.4	9.8%
MR2	8.3	16.5%	6.3	0.7	76.9	8.9%	641.6	6.4%
NW-P	9.7	18.3%	7.3	0.6	72.2	7.5%	499.7	4.5%
NW-R	9.9	17.5%	7.6	-4.3	101.1	11.1%	985.1	9.7%
AVG-P	8.8	16.8%	6.7	1.2	64.2	6.4%	505.4	4.9%
AVG-R	9.1	16.1%	7.0	-3.9	90.6	10.1%	884.0	9.1%
	Midday Off-peak				**Midday Off-peak**		**Midday Off-Peak**	
	RMSE	MAPE	MAE	BIAS	RMSE	MAPE	RMSE	MAPE
HS	9.1	26.8%	7.3	-1.3	56.4	10.4%	526.2	8.5%
LR1	8.7	26.6%	7.0	0.0	33.8	6.1%	337.3	5.1%
LR2	8.8	24.9%	7.0	-2.2	49.0	9.2%	533.0	8.8%
MR1	8.7	25.8%	6.9	-0.9	38.0	7.0%	391.1	6.1%
MR2	8.7	24.8%	6.9	-2.0	48.0	8.9%	514.3	8.5%
NW-P	9.9	28.9%	7.8	-0.7	41.4	7.5%	346.5	5.7%
NW-R	8.6	25.2%	6.8	-1.0	42.2	7.7%	400.5	6.3%
AVG-P	9.4	28.1%	7.5	-0.5	32.0	6.0%	322.7	5.2%
AVG-R	8.7	25.9%	6.9	-0.9	37.9	7.0%	392.7	6.1%
	Afternoon Peak				**Afternoon Peak**		**Afternoon Peak**	
	RMSE	MAPE	MAE	BIAS	RMSE	MAPE	RMSE	MAPE
HS	10.2	26.2%	8.2	-3.2	86.9	13.9%	921.7	11.6%
LR1	9.3	24.2%	7.5	-2.4	59.4	8.5%	690.7	7.8%
LR2	9.1	23.9%	7.3	-2.6	66.5	10.1%	739.8	8.5%
MR1	9.3	24.2%	7.5	-2.7	61.2	9.9%	768.4	9.5%
MR2	9.1	24.2%	7.4	-2.5	68.7	10.4%	756.1	8.8%
NW-P	10.8	28.1%	8.7	-2.2	71.2	9.7%	768.1	8.2%
NW-R	9.5	24.8%	7.7	-2.8	70.2	11.2%	846.2	10.3%
AVG-P	9.9	25.8%	8.0	-2.2	61.9	8.1%	694.9	7.8%
AVG-R	9.2	24.3%	7.4	-2.5	59.3	9.6%	731.2	8.8%

(a)



(b)

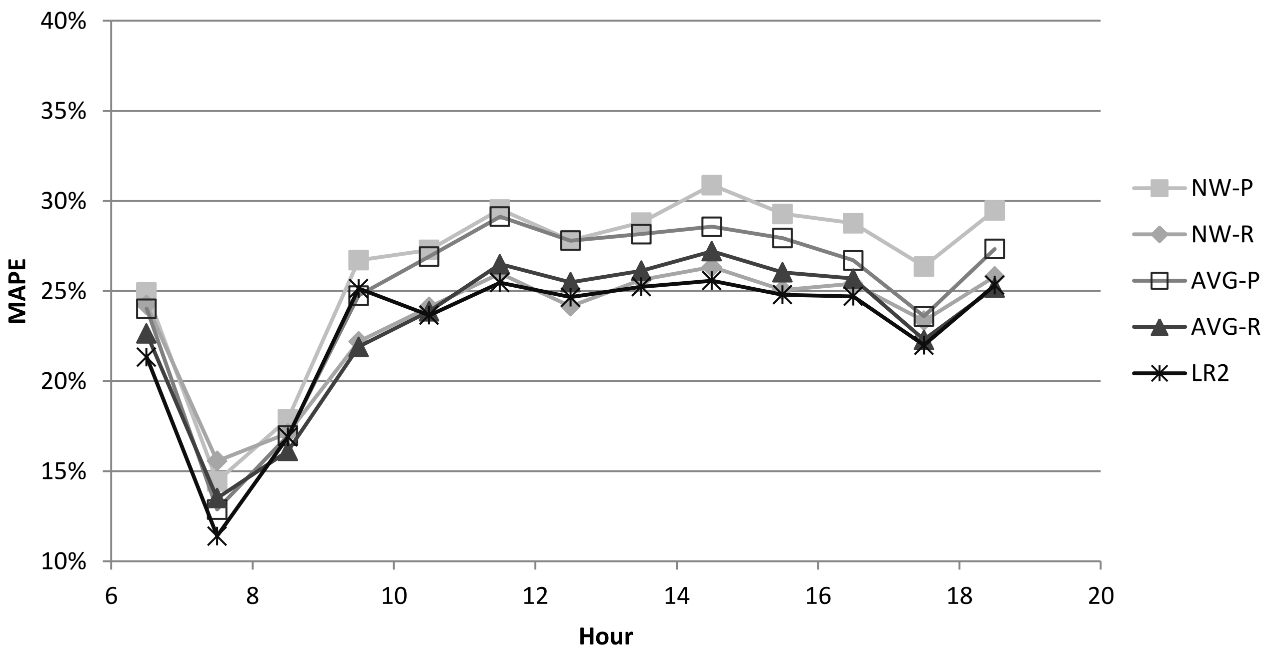
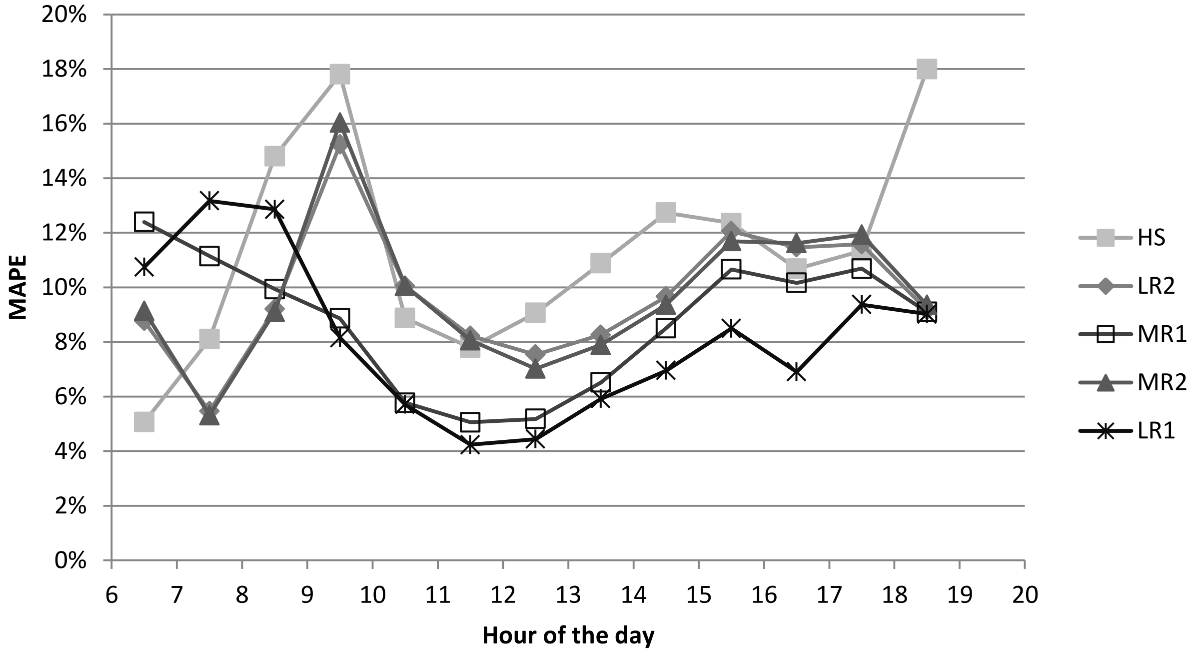


FIGURE 3 20 second MAPE averaged over one hour (a) regression models and historical model, (b) Newell's method and factored models (and LR2 as reference for comparison)

(a)



(b)

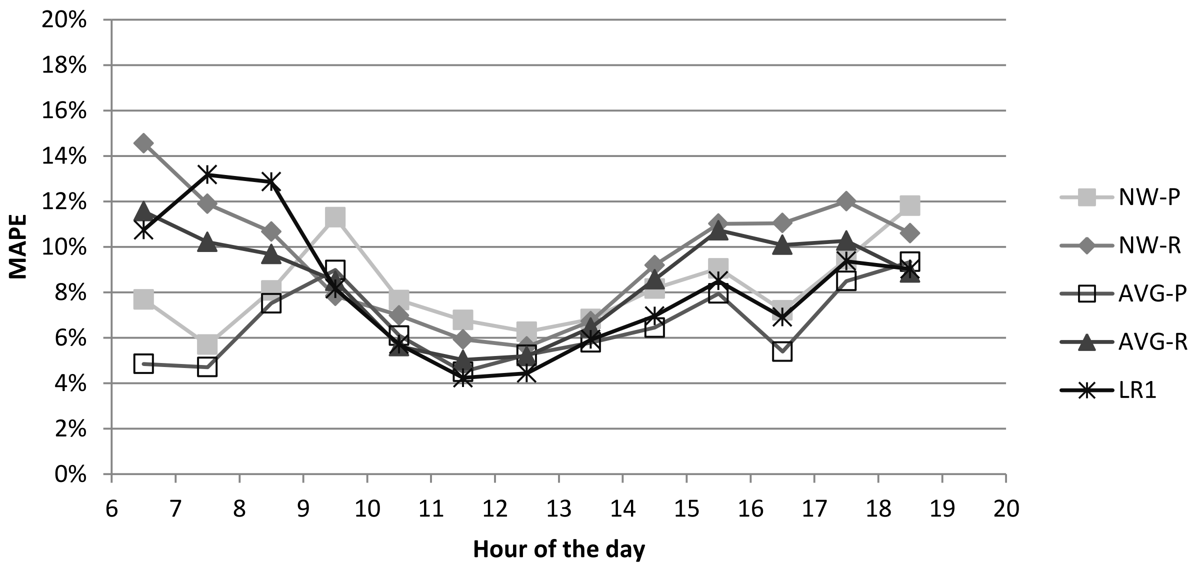
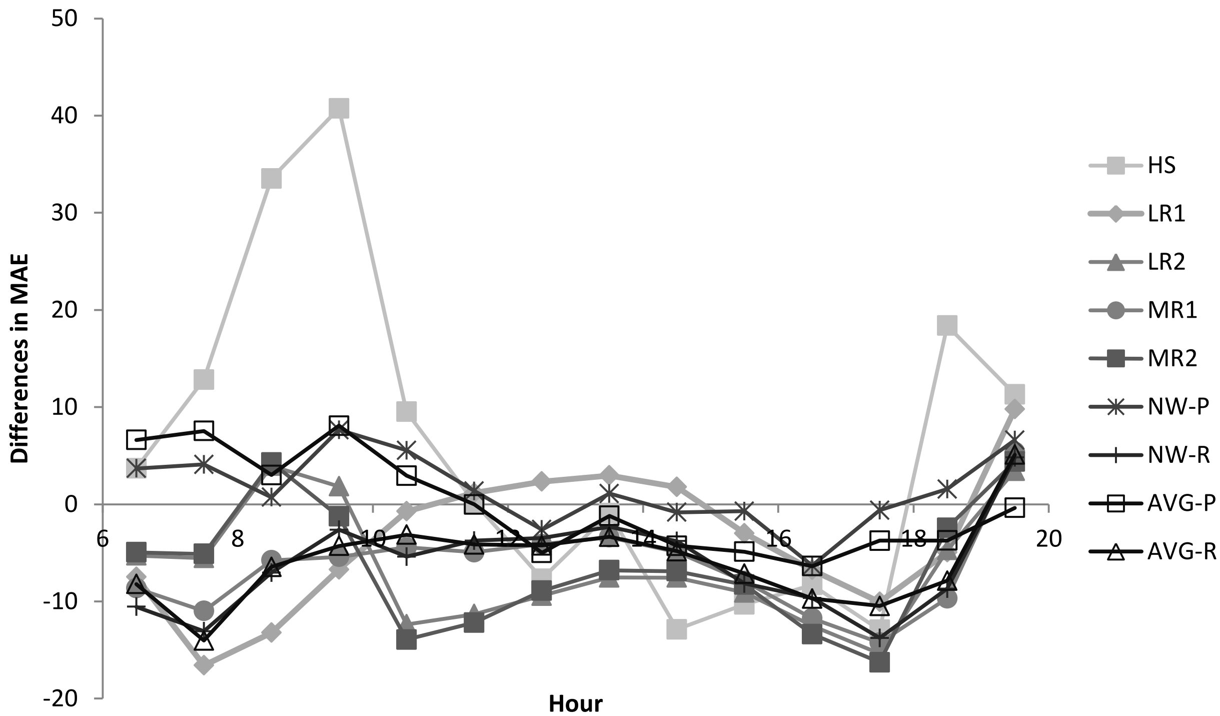


FIGURE 4 5 minute MAPE averaged over one hour (a) regression models and historical model, (b) Newell's models and factored models (and LR1 as reference for comparison)

(a)



(b)

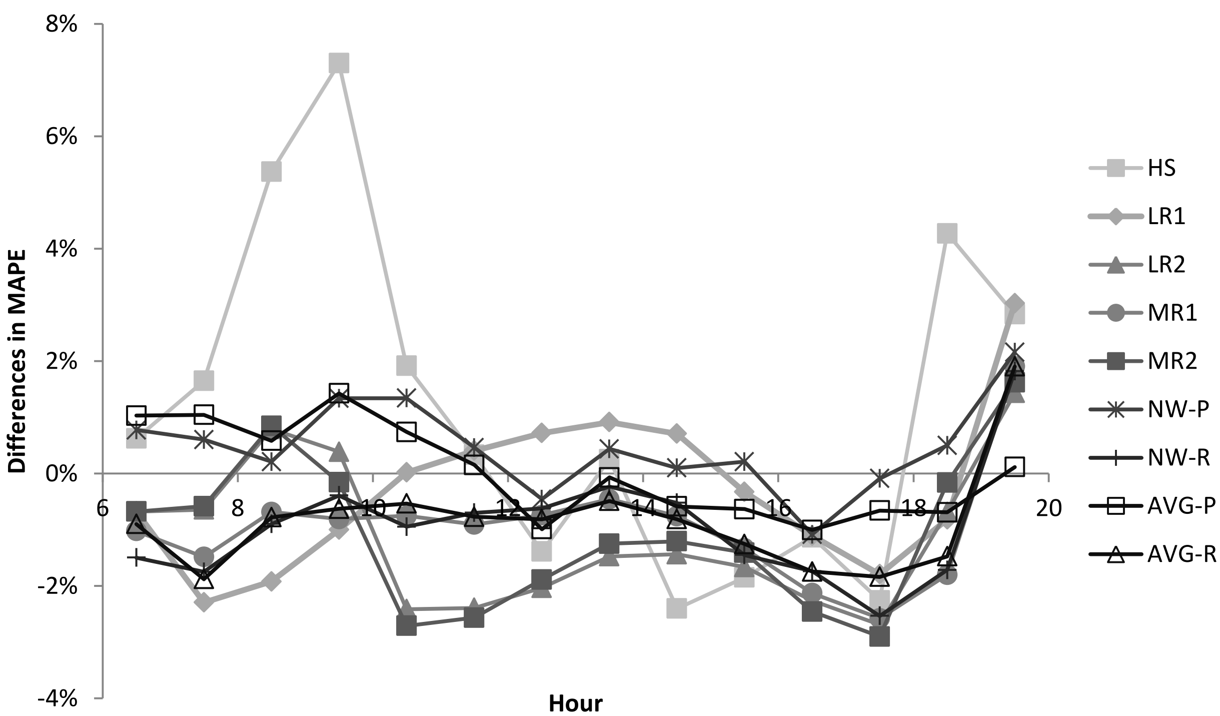


FIGURE 5 Differences in 5-minute (a) MAE and (b)MAPE between 4 week training data and 2 week training data (positive value means an improvement in the measure)

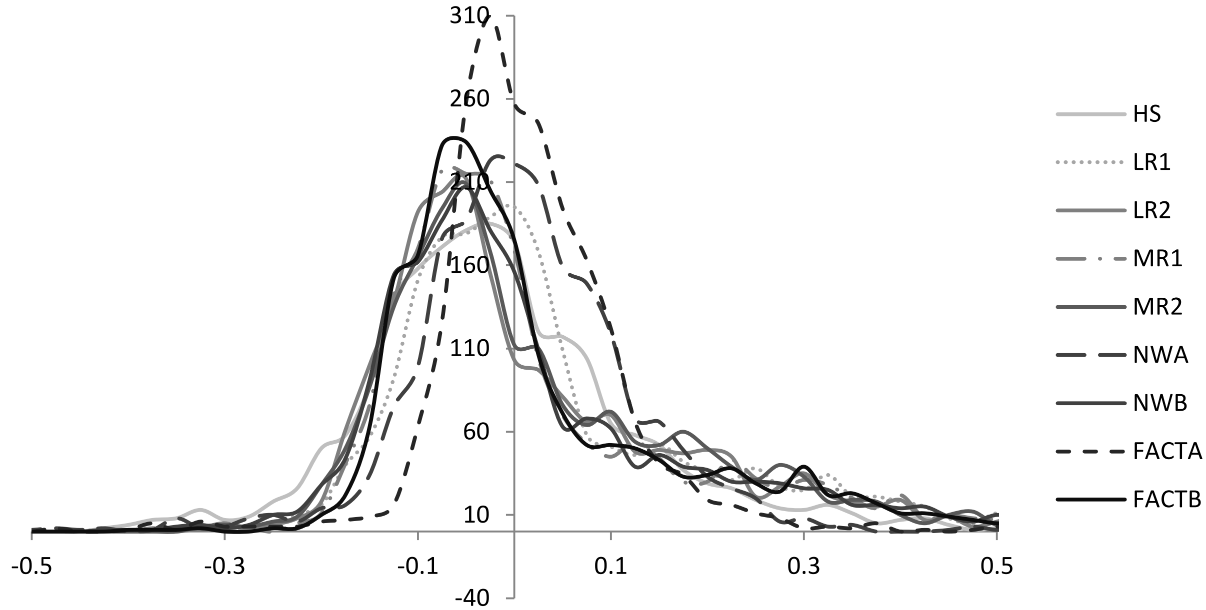


FIGURE 6 Distribution of the percentage errors for 5 minute aggregated results