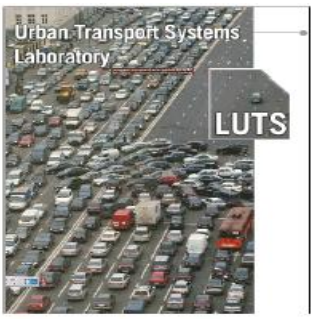
Travel Time Estimation and Prediction in Freeway Systems

Mehmet Yildirimoglu
Nikolaos Geroliminis

Conference paper STRC 2012

STRC
12th Swiss Transport Research Conference
Monte Verità / Ascona 02-04 May, 2012



ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Travel Time Estimation and Prediction in Freeway Systems

Mehmet Yildirimoglu	Nikolas Geroliminis
Swiss Federal Institute of Technology Lausanne	Swiss Federal Institute of Technology Lausanne
Lausanne	Lausanne

Phone: 0216932484
Fax: 0216932479
email:
mehmet.yildirimoglu@epfl.ch

Phone: 0216932481
Fax: 0216932479
email:
nikolas.geroliminis@epfl.ch

Abstract

Travel time is considered as an important performance measure for roadway systems, and dissemination of travel time information can help travelers to make travel decisions such as route choice or time departure. Since the traffic data collected in real time reflects the past or the current conditions on the roadway, a predictive travel time methodology should be used to obtain the information to be disseminated. However, an important part of the literature uses instantaneous travel time assumption, and sums the travel time of roadway segments at the starting time of the trip. The growing need for short-term travel time prediction also led to the development of forecasting algorithms. These methods can be broadly classified in two major categories; parametric methods (e.g. linear regression, time series models, Kalman filtering), non-parametric methods (neural network models, support vector regression, bayesian models, simulation models).

This paper presents a predictive travel time methodology based on speed data at fixed loop detectors. However, in contrast to above mentioned existing methodologies, it benefits from the available traffic flow essentials (e.g. shockwave, bottlenecks). The proposed method makes use of both historical and real time traffic information to provide travel time prediction. First, an existing bottleneck identification algorithm is used to determine the location and spatial extent of the bottlenecks. In order to use the historical dataset in a useful and efficient manner, days with similar traffic patterns (i.e. speed profiles) should be identified. Since high number of detectors and time periods in a day lead to a large number of observations, Principal Component Analysis (PCA) is used to reduce the dimensions of the dataset. Then, Gaussian Mixture Model (GMM) is used to create clusters in the historical dataset. Optimal number of clusters can be determined by the use of average silhouette width and information criteria such as Akaike Information (AIC) and Bayesian Information (BIC) criteria. In this study, optimal number of clusters is also verified by a performance measure which indicates spatial and temporal distribution of congested regions detected by the bottleneck identification algorithm. Based on this distribution, a probability map of congestion can be created for each cluster and can be used to predict travel time for a day which belongs to that cluster. However, this approach would only work under recurrent traffic conditions. In case of non-recurrent congestion, bottleneck identification algorithm is implemented in real time and the congestion propagation is estimated using the shockwave speed previously calculated for the historical dataset. The experiment results based on the loop detector data of I-80 and I-5 segments in California indicate that the proposed method provides promising travel time predictions under both recurrent and non-recurrent traffic conditions.

I. Introduction

Predictive travel time is a valuable information required by drivers and transportation managers to make better travel and control decisions. The provision of travel time information through Advanced Traveler Information Systems (ATIS) enables drivers to make decisions such as route choice and departure time. In addition, besides the fundamental traffic parameters, travel time can be used by transportation agencies to deploy efficient control measures and to prevent potential traffic congestion.

Data required to estimate travel time can be obtained through loop detectors, test vehicles, license plate matching techniques (automatic vehicle identification, AVI) and ITS probe vehicle techniques [1]. All of the detection technologies except the one based on loop detectors provide direct measurement of experienced travel time. As freeways are usually equipped with loop detectors that collect flow, speed and occupancy information, travel time estimation in freeways should rely on them. Travel time measurement can be either based on local velocity measurements, or more sophisticated models that attempt to correlate vehicle observations at multiple locations [2]. However, the essential problem with travel time information is that it always has to refer to future conditions in the roadway. On the contrary, traffic data collected in real time reflect past or current conditions in the roadway. Therefore, the provision of travel time information always requires prediction of future conditions on the roadway. The approach of instantaneous travel times might create considerable errors when traffic conditions are varying in time and space.

A speed contour plot is presented for a section in freeway I-5S in California in **Fig. 1**. A few active bottlenecks can be seen in the site that start at different times and propagate upstream. **Fig. 1** clearly shows that the difference between instantaneous and experienced travel time by plotting a few vehicle trajectories with instantaneous and experienced travel time. Note that these differences can be quite significant especially during the congestion onset and dissipation. This indicates that estimation of travel time should not be solely based on the traffic data collected in real time, but also the future recurrent traffic conditions can be integrated from historical data.

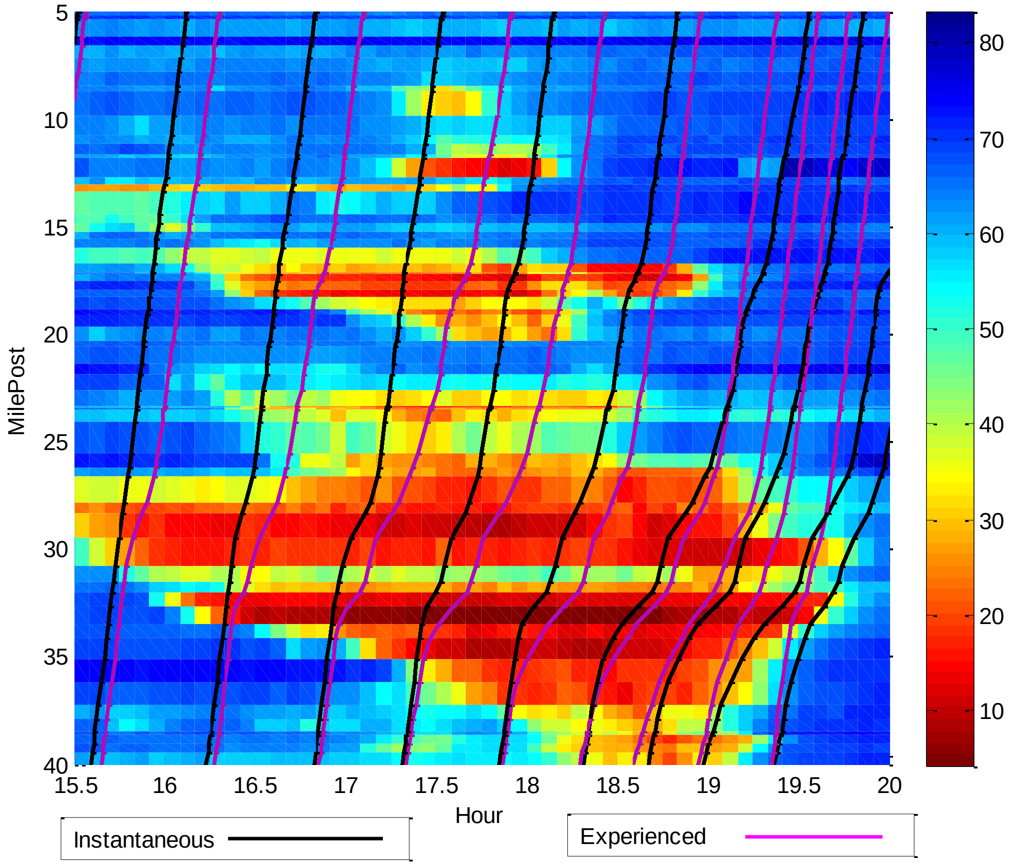


Fig. 1 Speed Contour Plot and Trajectories

The need for short-term travel time prediction led to the development of various forecasting algorithms. These methods can be broadly classified in two major categories; parametric methods (e.g. linear regression [3], time series models [4], Kalman filtering [5]) and non-parametric methods (neural network models [6], support vector regression [7], bayesian models [8], simulation models [9] etc.).

This paper presents a predictive travel time methodology based on speed data at fixed loop detectors. However, in contrast to the aforementioned existing methodologies, it benefits from the available traffic flow essentials (e.g. shockwave, bottlenecks). The proposed method makes use of both historical and real time traffic information to provide travel time prediction. Instead of identifying traffic flow patterns using statistical methods, that sometimes might not succeed to capture complex phenomena of traffic flow, we propose to integrate in the methodology, identification of traffic patterns with traffic flow theory fundamentals, for example with shockwave analysis and bottleneck estimation. First, an existing bottleneck identification algorithm is utilized to determine the location and spatial extent of the bottlenecks [10]. It uses speed readings at fixed detector locations as an indicator of bottleneck activation. Identified bottleneck locations are used in this study to restore the major traffic events likely to be observed on the roadway and to construct the link between real-time traffic information and historical dataset. Using the shockwave phenomena and identified bottleneck locations in real-time, the impact of a bottleneck can be predicted before it completely develops. Historical information can be very useful to determine the characteristics of the bottlenecks (i.e. spatial extent and duration) and so, predict their impacts. Nevertheless, as we will show later, traffic conditions significantly vary from day to day (even for similar demand conditions) and as a result the size of a bottleneck in the time-space domain and travel speed of vehicles in this domain have high fluctuations. Thus, a simple prediction based on historical average or a partitioning of traffic conditions based on days (weekdays-weekends) or times of day (AM or PM peak) might introduce significant estimation errors.

II. Methodology

A. Bottleneck Identification Algorithm

Chen et al. [10] developed an algorithm to automatically identify bottleneck locations, their activation and deactivation times, and their spatial extents using loop detector data and focusing on speed measurements. Chen method compares each pair of detectors adjacently located and determines the existence of bottleneck when

- Speed difference between upstream and downstream detectors is above the minimum speed differential, Δv_{min} threshold.
- Speed at upstream detector is below the maximum speed threshold, v_{max}.

Chen et al. [10] chose values of v_{max}=40mph and Δv_{min}=20mph with data aggregated at 5min intervals taken from California freeways. These parameters may need to be adjusted depending on the application.

Identification of bottleneck in an automated way allows us to restore the major traffic events that occur on the roadway and to keep track of traffic conditions in real time. However, since the algorithm has an offline part, it is not possible to smooth the results in real-time.

B. Clustering

Historical traffic patterns can be used for prediction of travel time. To use the historical dataset in a useful and efficient manner, days with similar traffic patterns (i.e. speed profiles) should be identified. Clustering techniques have been already used in transportation field to analyze traffic flow patterns, see for example [11]. However, since travel times are computed using local velocity measurements in this study, time-dependent speed measurements along the roadway must be used in the clustering step.

Since high number of detectors and time periods in a day lead to a large number of observations, Principal Component Analysis (PCA) is used to reduce the dimensions of the dataset, see for example [12]. PCA has been proved to be an efficient tool to reduce the dimensions of the dataset and to compress the data. PCA, using the orthogonal transformation, converts a set of observations with correlated variables into a set of observations with linearly uncorrelated variables, which are called principal components (PC).

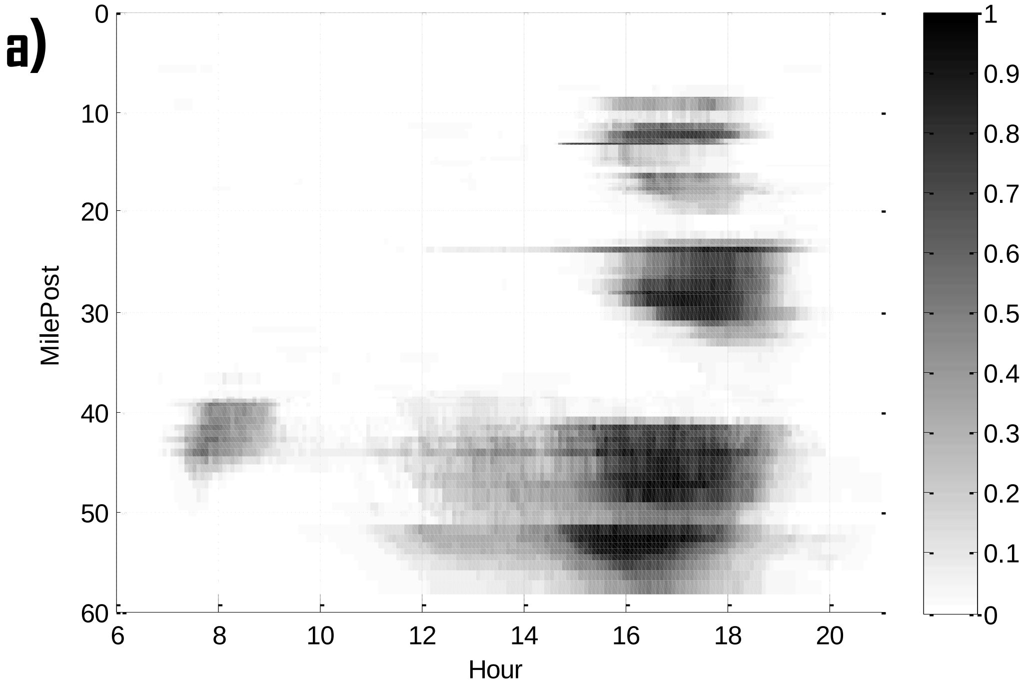
After reducing the dimensions of the dataset, Gaussian Mixture Model (GMM) is applied to create clusters in the historical dataset. GMM is the combination of multivariate normal density components, and it fits the data using expectation maximization (EM) algorithm. GMM is often used for clustering purposes, and unlike other clustering methods, it is not solely based on the distance between the observations, but it is based on the distribution of data points.

Optimal number of clusters can be determined by the use of average silhouette width and information criteria such as Akaike Information (AIC) and Bayesian Information (BIC) criteria. In addition to the optimal number of clusters, the stability of the results is crucial to clustering. GMM, whose initialization is random or based on *k*-means results, should return the same results every time it is repeated to ensure the accuracy and the robustness of the algorithm.

C. Stochastic Congestion Maps

Once clusters are created and bottleneck identification algorithm is applied, a stochastic congestion map (**Fig. 2**a), which represents the likelihood of observing bottleneck at a given space-time point, can be created for each cluster and it can be used to predict travel time for a day which belongs to the given cluster.

Each cluster is divided into subsets using certain threshold probability values (e.g. from 0.05 to 1 with spaces of 0.05), and a deterministic congestion map is created for each subset. Congestion map associated with the lowest threshold is constructed by the bottleneck points (in the time space domain) whose probability is greater than 0.05 (i.e. occurring more than 5% of the analyzed days), the map associated with second lowest threshold is constructed by the points whose probability is greater than 0.10, and so forth. **Fig. 2**b represents the subsets created within the congestion map. Note that darker colors represent higher threshold values, and the higher the value of the threshold is, the smaller the size of the bottleneck is. Also note that the difference between subsets is roughly in the shape of rings around a given core. These results are of great importance to our analysis because they show that even the location and duration of bottlenecks are roughly known a priori, a more careful look identifies strong stochastic phenomena that can vary travel times from one day to the other. For example bottleneck #5 in fig. 2b starts at location with milepost 40, but its extension in time and space varies from day to day.



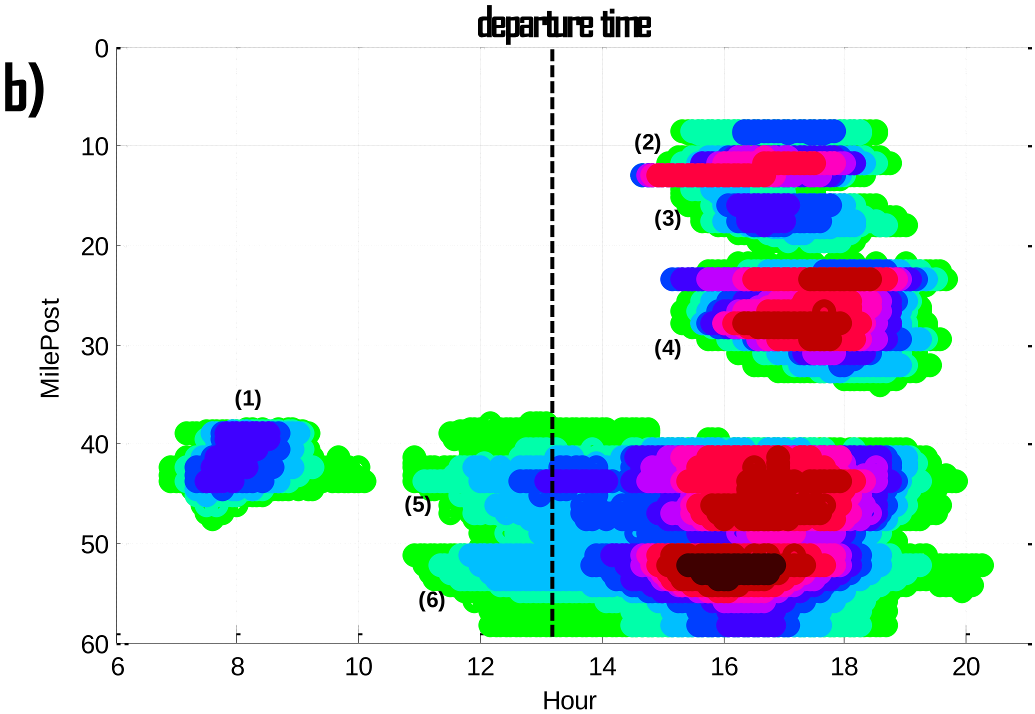


Fig. 2 a.Stochastic Congestion Map, b. Subsets in Congestion Map

Travel time for a specific departure time can vary significantly even within the cluster. To address the travel time variability and to provide more accurate travel time information, an update algorithm is developed in this study that allows us to switch between subsets and congestion maps associated with them.

D. Update Algorithm

Before delving into the update algorithm, one should notify that there are multiple cores in the congestion map around which the rings form. In addition, although there may be a correlation between the size of the blocks for a given day, it is intuitively known that they are not fully dependent. Therefore, the congestion map is divided into blocks as it is represented by the numbers in **Fig. 2**b, and the update algorithm is applied separately for each of them to select the threshold value that best represents the real-time traffic conditions on that part of space-time plot.

The idea behind the update algorithm is to provide travel times based on the expected traffic conditions at the very beginning. If there is no real-time information about the given block at the departure time, which is the case for block 2, 3 and 4 in **Fig. 2**b, the expected block size (e.g. probability of 0.5) is used to compute the predicted travel time. However, the given departure time in **Fig. 2**b allows us to compare real-time traffic information with congestion maps for blocks 5 and 6. Then, the algorithm starts switching between threshold values to find the bottleneck shape that would best represent the real-time bottleneck information obtained till the departure time. Threshold value returned by the update algorithm for each block is used to construct the congestion map, and travel time for the given departure time is computed on this time-dependent congestion map.

For a given block, selection of threshold value is done by the following similarity metric;

$$arg\ \max_i(true\ positive_i + true\ negative_i) \qquad (1)$$

where i is t threshold index, $true\ positive_i$ is the number of rightly classified bottleneck points in the congestion map associated with threshold i, and $true\ negative_i$ is the number of rightly classified non-bottleneck points in congestion map associated with threshold i. Note that the set of points that the algorithm uses to determine the threshold (Equation 1), is defined by the maximum size of the given block can get. Since real-time information is available only till departure time, the points up to the departure time are used in the update step. In addition, a moving time window of 2 hours is used to keep track of varying conditions. To summarize, the algorithm starts with the set of points defined by the lowest threshold value (the maximum size of the block) and up to the departure time, and then switches to a time window of 2 hours after the time difference between the departure time and the start of the block gets larger than 2 hours.

If the same performance value is computed for multiple threshold values, then the algorithm picks the one that is closest to the expected threshold value. This indicates that the algorithm always selects the conditions that are most likely to be observed.

E. Speed Profile

The problem in travel time prediction is twofold; prediction of major traffic events on the roadway (e.g. bottlenecks) and prediction of speed profile. By the methodology described so far, this study attempts to predict major traffic congestion events on the roadway. However, the speed profile inside and outside the bottleneck time-space domain is still unknown. There are two different approaches that can be applied for this estimation.

Instantaneous speed information (i.e. speed measurements at the departure time) can bring valuable insight into this problem. The assumption that instantaneous speed will remain constant separately for bottleneck and non-bottleneck locations can help us estimate a speed profile.

Another way to deal with this problem is to use the average speed measurements for each bottleneck block separately. Average speed for a given space-time point is calculated if the same point in historical dataset is registered as bottleneck, otherwise it is not considered in the average speed measurement. By this methodology, average speed measurement for each block and for each time period can be calculated.

The comparison between the real speed profile and the ones estimated, showed that the speed profile inside the bottleneck is generally underestimated in both instantaneous and average speed approach. Therefore, the minimum of instantaneous and average speed is used in this study to estimate the speed profile inside the bottlenecks. However, instantaneous travel time for non-bottleneck locations can be easily used to represent free flow speed.

III. Case Study

For the application, the data from California freeway performance measurement system (PeMS) is used. PeMS collects 30-sec loop detector flow and occupancy data throughout the state. Then, it processes them and fills in the missing detector data to compute 5-minute flow, occupancy and speed averages [13].

For this study, a 60 mile section of I-5S in the district of San Diego/Imperial is selected. Considering the detector quality and the effect of recurrent congestion, selected roadway section is quite suitable for this study. 5-minute loop detector data is downloaded through PeMS website for the year of 2011. The dataset is divided randomly into two parts; Training Set (~80%-294 days) and Testing Set (~20%-71 days).

A. Bottleneck Identification Algorithm

Bottleneck identification algorithm is applied for the training dataset and the results are utilized to compute the congestion maps in the next steps of the methodology.

B. Clustering

Before clustering, PCA is applied to reduce the dimensions of the dataset. 100 principal components carry 95% of the variance in the original data of 16020 variables. Therefore, the rest of the operations is carried out with the reduced dataset of 100 variables.

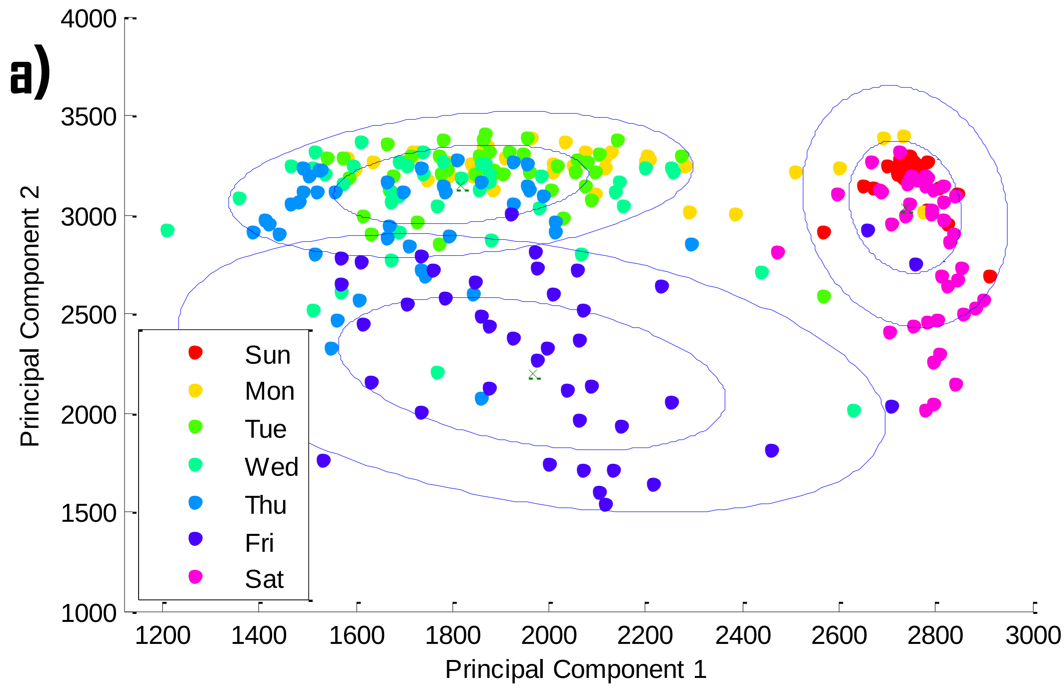
Average silhouette width, AIC and BIC criteria are used to determine the optimal number of clusters. Although AIC and BIC show a decreasing trend for increasing number of clusters, average silhouette width reaches its optimum value at three clusters. The optimal number of clusters indicated by AIC and BIC is also tested. In fact, any number of clusters except three brought unstable results. Therefore, considering both the stability and the average silhouette width results, three clusters is selected to be the most appropriate configuration in this analysis.

Fig. 3 presents the GMM results based on days of the week and based on clusters. It clearly shows that the clusters are mainly dominated by certain features of days of the week. The first cluster shown in **Fig. 3**b is dominated by weekend days, the second week days other than ‘Friday’s, and the third by ‘Friday’s.

Table 1 presents the distribution of days along the clusters. Most of the week days classified in the first cluster are holidays and they are not subject to a significant level of congestion. The second cluster is mainly composed of week days other than ‘Friday’s. These days have significant level of congestion. However, the level of congestion is not as high as it is observed on ‘Friday’s. Therefore, the clustering algorithm creates a separate cluster mainly for ‘Friday’s. Although the clusters do not totally belong to a certain day of the week, this information is very useful to identify expected traffic conditions on the roadway for a particular day. Hence, each cluster is assigned to dominating days of the week in a predetermined way, and travel time prediction for a given day is executed within the corresponding cluster and its associated congestion map.

Table 1 Clusters vs. Weekdays

				Days			
Cluster	*Mon*	*Tue*	*Wed*	*Thu*	*Fri*	*Sat*	*Sun*
1	6	1	1	1	2	40	42
2	36	41	37	35	8	0	0
3	0	0	4	6	32	2	0


a)
4000
3500
3000
2500
2000
1500
1000
Principal Component 2
Sun
Mon
Tue
Wed
Thu
Fri
Sat
1200
1400
1600
1800
2000
2200
2400
2600
2800
3000
Principal Component 1

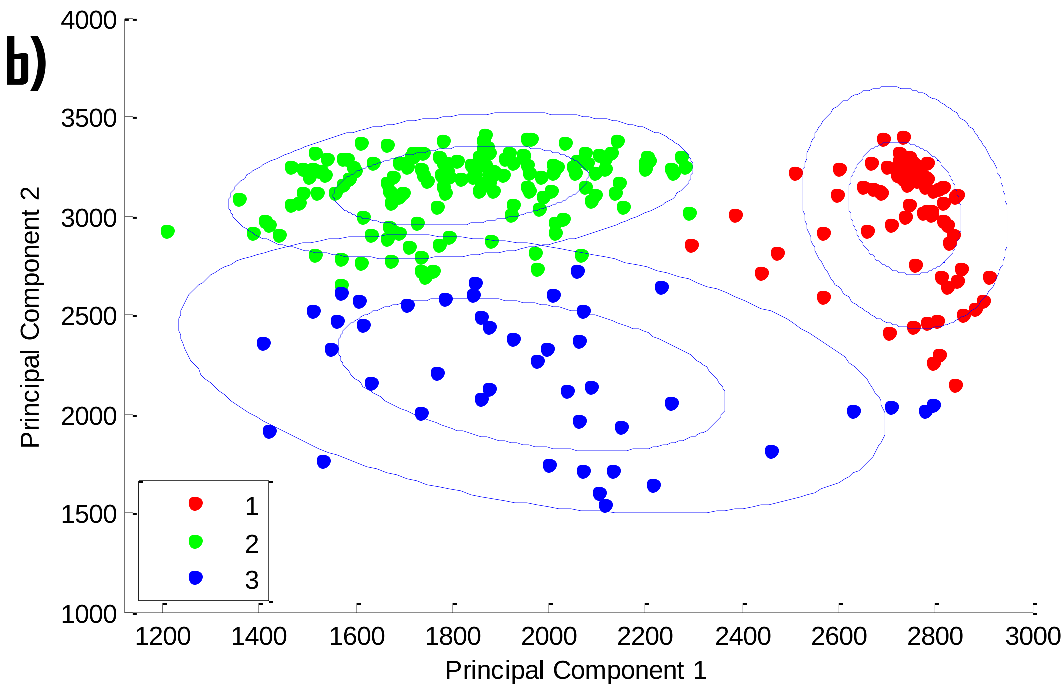

b)
4000
3500
3000
2500
2000
1500
1000
Principal Component 2
1
2
3
1200
1400
1600
1800
2000
2200
2400
2600
2800
3000
Principal Component 1

Fig. 3 GMM Results a. Based on days of the week, b. Based on clusters

C. Stochastic Congestion Map

By combining the results obtained from bottleneck identification and clustering steps, stochastic congestion maps can now be created. By simply estimating the average number of bottleneck occurrences for a space-time point, such a map can be constructed. **Fig. 2**a presents the stochastic congestion map for the third cluster described above. Then, it is divided into subsets for different threshold values.

IV. Results

Described methodology is tested on the training set (71 days). Since the weekend days are not subject to significant level of congestion, they are not considered in the evaluation step.

Travel time computation in this study is done by constant speed interpolation. For a link between two successive detectors, the speed measurement at downstream or upstream detector, or the average of two measurements can be used to represent the velocity. All constant speed interpolation methods imply instantaneous speed changes, which do not occur in real-time. However, considering the distance between the detectors (about 500m) in our study site, this phenomenon is not expected to largely affect the results. Travel time can also be computed by using linear and quadratic speed interpolation methods, which do not require instantaneous speed changes. Alternatively, one could apply a more detailed traffic flow model (of first or higher order) to estimate speed between the detectors. Nevertheless, we do not expect the accuracy of the results to improve. However, this study only uses constant speed interpolation method, with the average speed measurement of two successive detectors.

Instantaneous travel time is calculated by the use of the speed measurements at the departure time and the constant speed interpolation method, described above. On the other hand, experienced travel time is calculated by traveling a trajectory through the velocity field. The time it takes to travel each segment is calculated, and the speed measurements at the time when the trajectory reaches the next segment is used to compute its travel time. Predicted travel time is calculated in the same manner as experienced travel time. However, instead of velocity field which is unknown at the departure time, predictive trajectory walks over the congestion map and uses an estimated speed profile to compute the time it takes to traverse each segment.

First, holidays, which in general belong to weekend cluster and do not have a significant level of congestion, are excluded from dataset. Then, the congested periods in the testing set are identified with an automated algorithm. Note that a period is registered as congested if any of the instantaneous, predictive and experienced trajectory hit a bottleneck point. The implication of doing this is that a period can be registered as congested, although experienced travel times do not indicate delays. In case traffic conditions are less congested than the expected ones of the given cluster, predictive trajectory may hit some bottleneck points, although experienced travel times do not suffer from that. Therefore, our travel time prediction methodology experiences some time lag to adapt to real-time traffic conditions.

For the sake of comparison, historical travel times are also computed for each day of the week. The median value of the experienced travel times at a given departure time on a given day is taken as historical average value.

To measure the effectiveness of the methods, two statistics, namely mean absolute error (MAE) and mean absolute percentage error (MAPE) are utilized;

$$MAE = \frac{1}{n}\sum_{t=1}^{n}\left|T(t) - \hat{T}(t)\right| \tag{2}$$

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{T(t) - \hat{T}(t)}{T(t)}\right| * 100 \tag{3}$$

where n is the number of observations, $T(t)$ is the experienced travel time, and $\hat{T}(t)$ is the travel time provided by the methodology.

The weighted average of MAE (for identified 93 congested periods) with respect to congestion duration is 2.00, 2.80 and 5.30 minutes for predictive, instantaneous and historical average approaches respectively, which shows that our methodology outperforms the other two.

Fig. 4 presents MAE values of instantaneous and predictive travel time for identified 93 congested periods in the testing set. Since historical average performs clearly worse than the other methods, it is excluded from further analysis. In 72 out 93 cases, predictive travel time methodology produces better results than instantaneous travel time assumption.

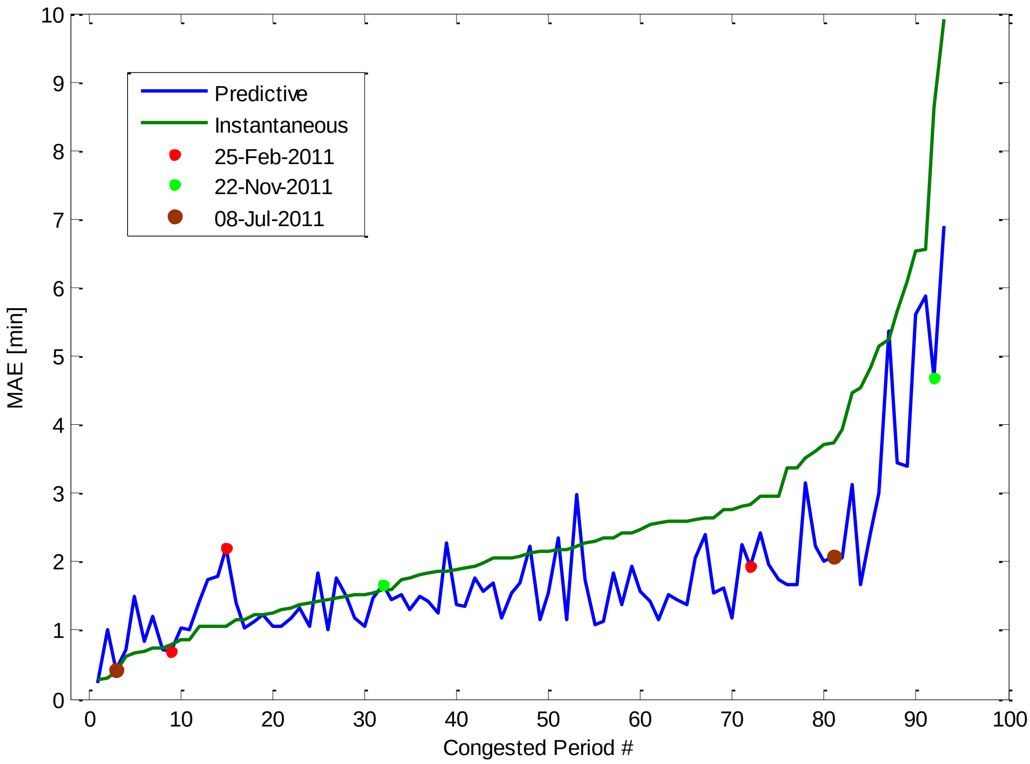


Fig. 4 Instantaneous vs. Predictive Travel time Performance for Congested Periods

Fig. 5 presents the travel times provided by predictive methodology developed in this study, instantaneous approach and historical average method for 3 different days. It clearly shows that historical average is not capable of producing accurate results under congested conditions. MAE values for the days plotted in **Fig. 5** are shown in **Fig. 4**. The morning peak in **Fig. 5**a leads to the largest difference in MAE values in the favor of instantaneous approach. This happens for a case of a short morning peak duration with light congestion. However, predictive methodology performs quite well for the rest of the day, especially in the evening peak which is highly congested. The afternoon peak in **Fig. 5**b leads to one of the largest difference in MAE values in the favor of predictive methodology. Although predictive travel time has a similar performance with instantaneous one at the congestion onset, it captures the peak of the curve and does a quite good job on the congestion dissipation. **Fig. 5**c is another example where predictive methodology mostly captures the shape of the travel time curve.

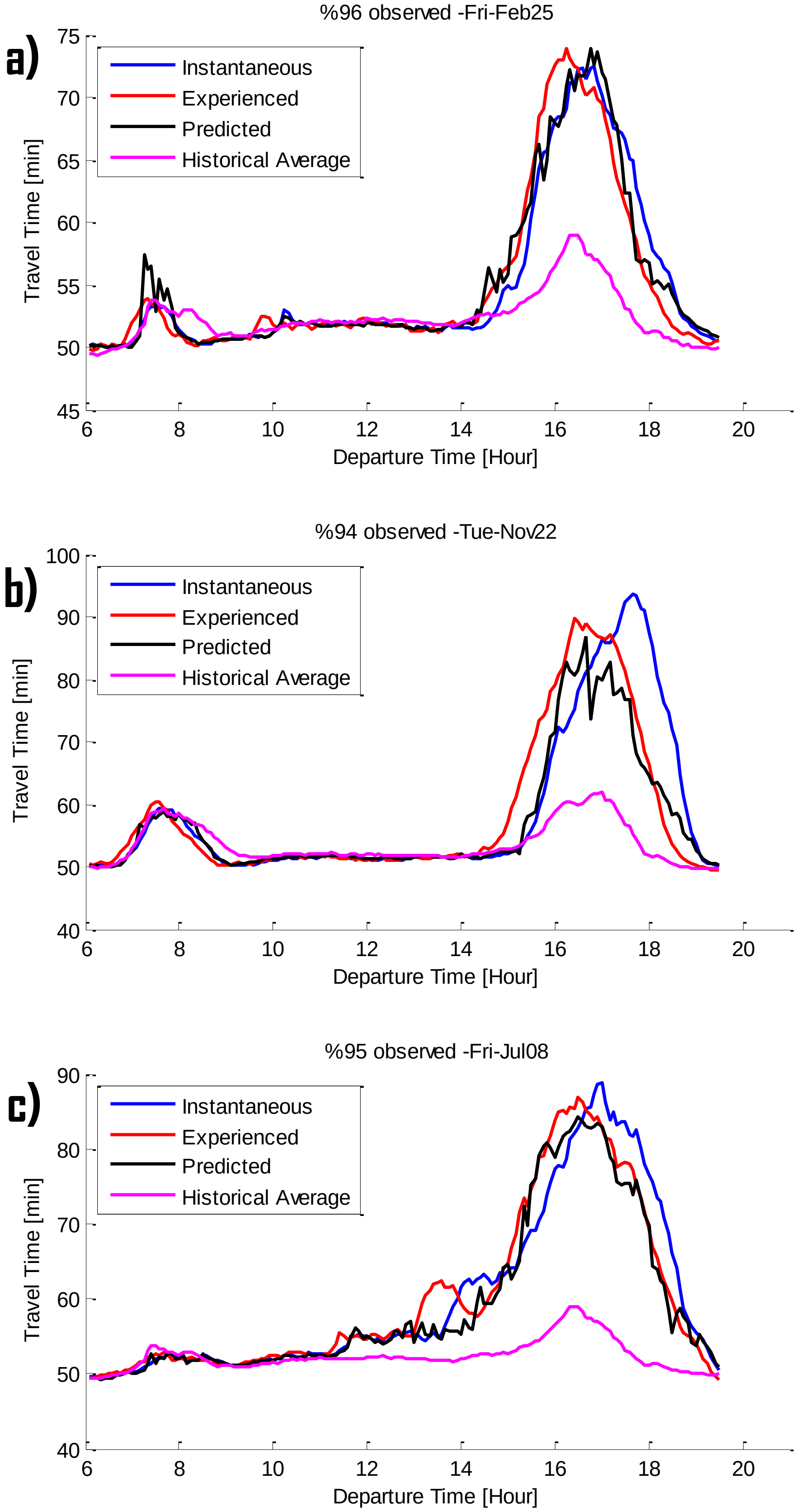


Fig. 5 Comparison of Models a. 25-Feb-2011, b. 22-Nov-2011, c. 08-Jul-2011

V. Conclusion

The disposition of travel time information through ATIS or its use as in ATMS to deploy efficient control measures always requires the prediction of traffic conditions on the freeway. The aim of this paper is to predict travel times by using traffic flow essentials, not any statistical procedure.

First, an automated bottleneck identification algorithm is used to detect the major traffic events that occur on the freeway. Then, the historical (or training) dataset is partitioned based on the clusters obtained through GMM. The results obtained from the first two parts are combined to create stochastic congestion maps for each cluster. Next, using the estimated speed profile, the congestion maps associated with threshold values and an update algorithm that compares real-time and historical traffic conditions, this study predicts the experienced travel times.

The experiment results based on the loop detector data of I-5 segment in California/San Diego indicate that the proposed method provides promising travel time predictions under varying traffic conditions.

Ongoing work analyzes the performance of the developed methodology in different study sites. Also, extensions of the algorithm include a time-dependent cluster assignment, where if conditions sharply change during a period, the algorithm will identify which cluster of days is more appropriate to capture this change. We also plan to test the performance of the algorithm in cases of some non-recurrent events in the freeways, especially accidents.

REFERENCES

1. *Travel Time Data Collection Handbook*, FHWA-PL–98–035, College Station, TX: Texas Transportation Institute, The Texas A&M University System, March 1998.
2. B. Coifman (2002) Estimating travel times and vehicle trajectories on freeways using dual loop detectors, Transportation Research Part A, vol. 36, pp. 351-364.
3. Zhang, X. and Rice, J. A. (2003) Short-term travel time prediction, Transportation Research Part C, 11(3–4), pp. 187–210.
4. Yang, J.-S.(2005) A Study of travel time modeling via time series analysis, In Proceedings of 2005 IEEE Conference on Control Applications, Toronto, Canada , pp. 855-860.
5. Van Lint, J. W. C. (2008) Online learning solutions for freeway travel time prediction, In IEEE proceedings of International Conference on Intelligent Transportation Systems, 9(1), pp. 38–47.
6. Van Lint, J. W. C. (2006) Reliable real-time framework for short-term freeway travel time prediction, Journal of Transportation Engineering, 132(12), pp. 921-932.
7. Vanajakshi, L. and Rilett, L.R. (2007) Support vector machine technique for the short term prediction of travel time, In IEEE Proceedings of Intelligent Vehicles Symposium, Istanbul, Turkey, pp. 600-605.
8. Fei, X., Lu, C-C., Liu, K., (2011) A Bayesian dynamic linear model approach for real-time short-term freeway travel time prediction, Transportation Research Part C, 19(6), pp.1306-1318.
9. Liu, Y., Lin, P.W., Lai, X.R., Chang, G.L., Marquess, A. (2006) Developments and applications of

simulation-based online travel time prediction system: traveling to Ocean City, Maryland, Transportation Research Record, 1959, pp. 92–104.

10. Chen, C., Skabardonis, A., Varaiya, P., (2004). Systematic identification of freeway bottlenecks. Transportation Research Record 1867, pp.46–52.
11. Weijermars, W., and E. van Berkum (2005), Analyzing highway flow patterns using cluster analysis, IEEE Conference on Intelligent Transportation Systems, Proceedings, Vienna, Austria, pp. 831-836.
12. Nagendra, S.M.S., Khare, M., 2003. Principal component analysis of urban traffic characteristics and meteorological data. Transportion Research Part D, vol. 8, pp. 285–297.
13. Chen, C., Petty, K., Skabardonis, A.,Varaiya. P. (2001) Freeway performance measurement: Mining loop detector data, Transportation Research Record 1748, pp. 96–102.