

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/222419172>

# Estimating dynamic roadway travel time using automatic vehicle identification data for low sampling rates

Article in *Transportation Research Part B Methodological* · November 2006

DOI: 10.1016/j.trb.2005.10.002 · Source: RePEc

---

CITATIONS

80

---

READS

130

2 authors:



**Francois Dion**

University of California, Berkeley

37 PUBLICATIONS 1,236 CITATIONS

SEE PROFILE



**Hesham Rakha**

Virginia Polytechnic Institute and State Uni...

284 PUBLICATIONS 3,331 CITATIONS

SEE PROFILE

# Estimating Dynamic Roadway Travel Times using Automatic Vehicle Identification Data for Low Sampling Rates

By

Francois Dion<sup>1</sup> and Hesham Rakha<sup>2</sup>

## ABSTRACT

The paper describes a low-pass adaptive filtering algorithm for predicting average roadway travel times using Automatic Vehicle Identification (AVI) data. The algorithm is unique in three aspects. First, it is designed to handle both stable (constant mean) and unstable (varying mean) traffic conditions. Second, the algorithm can be successfully applied for low levels of market penetration (less than 1 percent). Third, the algorithm works for both freeway and signalized arterial roadways. The proposed algorithm utilizes a robust data-filtering procedure that identifies valid data within a dynamically varying validity window. The size of the validity window varies as a function of the number of observations within the current sampling interval, the number of observations in the previous intervals, and the number of consecutive observations outside the validity window. Applications of the algorithm to two AVI datasets from San Antonio, one from a freeway link and the other from an arterial link, demonstrate the ability of the proposed algorithm to efficiently track typical variations in average link travel times while suppressing high frequency noise signals.

---

<sup>1</sup> Dept. of Civil and Environmental Engineering, Michigan State University, East Lansing, MI 48824, USA. E-mail: fdion@vtti.vt.edu. Tel: (517) 353-8883.

<sup>2</sup> Corresponding Author. Charles Via Dept. of Civil and Environmental Engineering, Virginia Tech Transportation Institute, Blacksburg, VA 24061, USA. E-mail: hrakha@vt.edu. Tel: (540) 231-1505.

## 1. INTRODUCTION

In recent years, there has been a growing interest in utilizing Automatic Vehicle Identification (AVI) data for the provision of real-time travel time information to motorists within Advanced Traveler Information Systems (ATIS). Examples of existing AVI systems include the TranStar system in Houston (Houston TranStar, 2001), the TransGuide system in San Antonio (SwRI, 1998), and the TRANSMIT system in the New York/New Jersey metropolitan area (Mouskos et al., 1998). Both the TranStar and TRANSMIT systems estimate link travel times by monitoring the successive passage times of vehicles equipped with electronic tags at freeway toll booths. In the case of the TransGuide system, data are gathered using AVI tag readers that are installed solely for the purpose of estimating travel times.

While AVI systems may provide time stamps at which vehicles pass successive monitoring stations, the data that are gathered by these systems require some form of filtering in order to remove outlier observations. Examples of outlier data would be vehicles making a stop or taking a detour between two stations. Since these vehicles would experience a travel time that is atypical, these observations should therefore be removed from the data set of valid observations to avoid producing erroneous travel time estimates.

The paper addresses the problem of obtaining reliable travel time estimates from AVI data by presenting a robust data-filtering algorithm. The paper starts with a description of the data filtering algorithms that are currently being used in the TranGuide, TranStar and TRANSMIT systems. These algorithms are then evaluated through a theoretical discussion and an application to a dataset of observed freeway travel times from the San Antonio AVI system. Problems with the state-of-the-art algorithms are discussed and a new algorithm is presented that has been developed to filter AVI travel time data in both stable and unstable traffic conditions. A more extended evaluation of the proposed algorithms is then performed on two series of observed travel times from the San Antonio AVI system. The main conclusions of the evaluations and some recommendations for future work are finally presented.

## 2. BACKGROUND

There are several systems commercially available that are capable of estimating real-time travel times. These can be broadly classified into spot speed measurement systems, spatial travel time systems, and probe vehicle technologies. Spot speed measurement systems, specifically inductance loop detectors, have been the main source of real-time traffic information for the past two decades. Other technologies for measuring spot speeds have also evolved, such as infrared and radar technologies. Regardless of the technology, the spot measurement approaches only measure traffic stream speeds over a short roadway segment at fixed locations along a roadway. These spot speed measurements are used to compute spatial travel times over an entire trip using space-mean-speed estimates. In addition, new approaches that match vehicles based on their lengths have also been developed (Coifman, 1998; Coifman and Cassidy, 2002; Coifman and Ergueta, 2003). However, these approaches require raw loop detector data as opposed to typical 20- or 30-second aggregated data. Alternatively, spatial travel time measurement systems use fixed location equipment to identify and track a subset of vehicles in the traffic stream. By matching the unique vehicle identifications at different reader locations, spatial estimates of travel times can be computed. Typical technologies include AVI and license-plate video detection systems. Finally, probe vehicle technologies track a sample of probe vehicles on a second-by-second basis as they travel within a transportation network. These emerging technologies include cellular geo-location, Global Positioning Systems (GPS), and Automatic Vehicle Location (AVL) systems.

Kaysi et al. (1993) recommend that traffic routing strategies under recurring and non-recurring strategies be based on forecasting of future traffic conditions rather than historical and/or current conditions. In general the traffic prediction approaches can be categorized into three broad areas: (i) statistical models, (ii) macroscopic models, and (iii) route choice models based on dynamic traffic assignment (Ben Akiva et al., 1992). Time series models have been used in traffic forecasting mainly because of their strong potential for online implementation. Early examples of such approaches include Ahmed and Cook (1982) and more

recently Lee and Fambro (1999) and Ishak and Al-Deek (2003). In addition, researchers have applied Artificial Neural Network (ANN) techniques for the prediction of roadway travel times (Park and Rilett, 1998; Abdulhai et al., 1999; Rilett and Park, 1999). These studies demonstrated that prediction errors were affected by a number of variables pertinent to traffic flow prediction such as spatial coverage of surveillance instrumentation, the extent of the loop-back interval, data resolution, and data accuracy.

The algorithm that is developed as part of this research effort is a statistical approach that is suitable for short-term prediction (2- to 5-minute predictions) of roadway travel times under stable (constant mean) and non-stable (varying mean) traffic conditions.

### 3. EXISTING AVI TRAVEL TIME ALGORITHMS

#### 3.1. TransGuide Algorithm

Within the TransGuide system, link travel times between successive AVI readers are estimated using a rolling average algorithm that automatically filters out all recorded travel times that exceed a user-defined threshold link travel time. This algorithm, which was developed by the Southwest Research Institute (SwRI), is defined by Equations 1 and 2. Equation 1 defines the set of valid recorded travel times  $Stt_{ABt}$  that is used at each evaluation time to estimate the current average travel time between two AVI readers  $A$  and  $B$ , while Equation 2 defines how average travel time estimates  $tt_{ABt}$  are calculated for the corresponding set of observations (SwRI, 1998).

$$Stt_{ABt} = \{ t_{Bi} - t_{Ai} \mid t - t_w \leq t_{Bi} \leq t \text{ and } tt'_{ABt} (1 - l_{th}) \leq t_{Bi} - t_{Ai} \leq tt'_{ABt} (1 + l_{th}) \} \quad [1]$$

$$tt_{ABt} = \frac{\sum_{i=1}^{|Stt_{ABt}|} (t_{Bi} - t_{Ai})}{|Stt_{ABt}|} \quad [2]$$

In the above equations,  $t_{Ai}$  and  $t_{Bi}$  represent the detection times of a vehicle  $i$  at readers  $A$  and  $B$ , respectively;  $t$  the time at which the travel time estimation takes place; and  $tt'_{ABt}$  the previously estimated average travel time from reader  $A$  to reader  $B$ . The main operating parameters of the TransGuide algorithm are the rolling-average window  $t_w$  and the link threshold travel time  $l_{th}$ . The rolling-average window determines the period of time that should be considered when estimating the current average travel time while the link threshold parameter is used to identify and remove outlier observations. For the TransGuide system, the observation window was set at 2 min and the link threshold parameter at 0.20. The selection of a 2-minute window meant that the travel times would be re-estimated every 2 min. The selection of a link threshold of 0.20 further meant that any observed travel times between a pair of readers that differs by more than 20 percent from the average travel time associated with observations made in the previous 2 min would be deemed invalid and not be considered in the calculation of the new average travel time for the current interval.

#### 3.2. TranStar Algorithm

The TranStar algorithm is generally similar to the TransGuide algorithm. This algorithm, which was also developed by SwRI, again uses the filtered data set defined by Equation 1 and the arithmetic average of Equation 2 to calculate current average link travel times between successive AVI stations. The main difference is that travel times are updated each time new travel time information is obtained from a vehicle equipped with an electronic tag instead of being updated at fixed intervals (Vickich, 2001).

Similar to the TransGuide system, the TranStar algorithm uses a link threshold parameter of 0.20 to identify outlier data. However, the algorithm uses a much short rolling average window of only 30 s. This means

that any recorded travel time between a pair of AVI stations will be considered invalid and rejected from the statistical analyses if this travel time is greater or lower by 20 percent than the average travel time associated with observations made in the previous 30 s.

### 3.3. TRANSMIT Algorithm

Travel time estimation within the TRANSMIT system is relatively similar to the TransGuide and TranStar systems. However, instead of using a rolling average to obtain estimates of current travel times between AVI stations, this system uses fixed 15-minute observation intervals. For each interval  $k$ , the system collects a sample  $n_k$  individual link travel times, up to a maximum of 200 observations, and uses this sample to estimate a current average travel time for the interval,  $tt_{ABk}$ , on the basis of Equation 3 (Mouskos et al., 1998).

$$tt_{ABk} = \frac{\sum_{i=1}^{n_k} (t_{Bi} - t_{Ai})}{n_k} \quad [3]$$

Equation 4 is further used to smooth the estimated travel times against historical data from the same 15-minute interval in the previous week or weekend day, depending on the case, to obtain an updated historical average travel time. In this equation,  $tth_{ABk}$  represents the historical smoothed travel time for the  $k^{\text{th}}$  sampling interval while  $tth''_{ABk}$  and  $tth''_{ABk-1}$  updated historical smoothed travel time for current ( $k$ ) and previous ( $k-1$ ) sampling intervals respectively. The smoothing process is currently set using a robust exponential smoothing algorithm. The robustness of this algorithm is associated to its use of a smoothing factor  $\alpha$  of 10 percent when no incident is detected and a factor of 0 percent when an incident is reported. This form of exponential smoothing ensures that incident data are not included in the moving average, and thus, that the historical database only includes typical non-recurring congested conditions.

$$tth''_{ABk} = (\alpha) \cdot tth_{ABk} + (1 - \alpha) \cdot tth''_{ABk-1} \quad [4]$$

Within the TranStar system, incidents are reported either manually or through an automatic incident detection algorithm that was developed by PB Farradyne Inc. This algorithm is based on the observation that link travel times tend to be normally distributed under free-flowing, non-incident traffic conditions. When a number of vehicles fail to arrive at a monitoring station within expected travel times, the algorithm increases the probability of the presence of an incident on the link upstream of the monitoring station and decreases the probability of a false alarm. Once the confidence level of a possible incident is reached, the occurrence of an incident increases to its user-defined threshold and an alarm is then set off at the central computer.

## 4. EVALUATION OF EXISTING AVI FILTERING ALGORITHMS

As was discussed earlier, the TransGuide/TranStar algorithms remove outlier data prior to computing the moving average travel time. Alternatively the TRANSMIT algorithm does not remove outlier data, but instead constructs a smoothed historical dataset. In this case, the smoothing process combines 10 percent of the newly estimated average with 90 percent of the observed historical average for the corresponding evaluation interval. Another important difference between the TRANSMIT and TransGuide/TranStar algorithms relates to the ability of each algorithm to reflect short-term fluctuations in traffic conditions. In the TRANSMIT algorithm, the response to changes depends on the weights assigned with the new and historical travel time averages in the data smoothing process. Alternatively, in the TranStar/TranGuide algorithms, the ability to react to changes in traffic conditions depends on the values assigned to the rolling-average window and link threshold parameters.

While both the TransGuide and TranStar algorithms were developed to follow short-term changes in link travel times and are generally similar, there appear to be differences in their ability to track changes in travel

times. For example, Figure 1 illustrates an application of the TransGuide algorithm to a dataset of weekday freeway travel times from the San Antonio AVI system. This algorithm was applied using the same parameters used in San Antonio, i.e., a 2-minute rolling average window and a 20 percent link travel time threshold. It can be observed that the algorithm is unable to track all the changes in traffic conditions, particularly the sudden onset of congestion during the morning peak. However, no problems have been reported regarding the ability of the TranStar filtering algorithm to correctly track changes in traffic conditions within the Houston area.

The main reason for the observed differences in performance is linked to the number of AVI tag readings made within each system. In the TranStar system, the main source of travel time information is from commuters using the EZ-Tag automatic toll collection system that is operated by the Harris County Toll Road Authority. In contrast, the main source of information for the TransGuide system is from commuters that volunteered to install a tag on their vehicle. While the Harris County EZ-Tag program has now about 1 million tags in circulation with over 500,000 active accounts, only 38,000 tags were distributed to volunteer drivers in San Antonio (SAIC, 2000). The 1 million tags that were distributed in Houston resulted in a much higher level of market penetration and more specifically in an ability to collect valid travel time information from about 9 percent of all the vehicles passing an AVI antenna location. In the late 1990s, the circulation of only 150,000 tags translated into an ability to collect approximately seven valid travel times per minute during peak periods at AVI locations in Houston, and five readings per minute during off-peak periods (Turner et al., 1998; ETTM on the Web, 2001). With a million tags in circulation, the detection rate should now be significantly higher.

On the other hand, the 38,000 tags that were distributed in San Antonio resulted in a maximum level of market penetration of less than 1 percent, and resulted more specifically in the occurrence of significant temporal gaps in the data. For instance, in the data set illustrated in Figure 1, there is typically only one or no travel time observation(s) per 2-minute interval. While there are more observations during peak periods, there are still typically only two or three observations per 2-minute interval, with no more than five in any interval. With such a low sampling rate, it is therefore possible that successively recorded travel times suddenly jump by more than 20 percent, especially if several minutes elapse between successive readings.

To demonstrate the impacts of sampling rates on the ability of the TransGuide algorithm to efficiently track sudden changes in travel times, Figure 2 illustrates an application of the algorithm to a series of synthetic datasets mimicking the travel time profile of Figure 1. The synthetic travel times used in this experiment were generated using a function that produces random travel time variations around the observed travel mean of each interval. For completeness, the real-world travel times shown in Figure 1 were also included in the synthetic datasets. As indicated in Figure 2, various applications scenarios were considered by assuming sampling rates ranging from one travel time observation every 5 s to one observation every 30 s. As can clearly be observed, the algorithm is unable to effectively track changes in traffic conditions when the sampling rate drops below one observation every 20 s.

## 5. PROPOSED FILTERING ALGORITHM

A review of existing AVI filtering algorithms indicates that there is a need for the development of an algorithm that would be capable of tracking abrupt changes in observed travel times without requiring a high sampling rate. First, the TRANSMIT system does not truly report existing traffic conditions due to its use of a weighting process combining current observed travel times with historical data. Second, while reports indicate that the TranStar system can adequately track daily travel time fluctuations within the Houston freeway network, this success is attributed to the high level of market penetration of AVI-equipped vehicles. Contrary to the Houston system, the San Antonio system has a much lower level of market penetration. These low sampling rates thus create difficulties for the TransGuide AVI filtering algorithm to adequately track sudden changes in traffic conditions.

To address the shortcomings of the current state-of-the-art algorithms, an enhanced filtering algorithm is presented in this paper. This algorithm determines average travel times between successive AVI readers by first ignoring all duplicate records that might be generated by the communication equipment and then by applying a series of filters to the collected travel times to remove invalid observations. As will be explained in the following sub-sections, the algorithm considers as invalid any observed travel time that falls outside a validity range that is determined based upon the following four factors:

- a) Expected average trip time and trip time variability in future time interval,
- b) Number of consecutive intervals without any readings since the last recorded trip time,
- c) Number of consecutive data points either above or below the validity range, and
- d) Variability in travel times within an analysis interval.

Similar to the TransGuide/TranStar algorithms, the proposed algorithm is designed for real-time estimation and forecasting of roadway travel times using AVI data.

### 5.1. Expected Interval Average Travel Time and Travel Time Standard Deviation

Within the filtering algorithm, the expected smoothed average travel time  $tts$  and smoothed travel time variance  $\sigma^2_{stt}$  between a pair of readers  $A$  and  $B$  for a given sampling interval  $k$  are computed using a robust exponential smoothing low-pass filter. As shown in Equation 5, the technique estimates the expected average travel time within a given sampling interval based on a set of  $n_{vk}$  valid observations in the previous sampling interval ( $k-1$ ) and the smoothed average travel time  $tts_{AB\ k-1}$  that was estimated at the end of the previous interval using an adaptive exponential smoothing filter. A similar process is used to estimate the expected standard deviation in Equation 6. It should be noted at this point that a more detailed description of how the variance is computed is provided later in the paper.

$$tts_{AB\ k} = \begin{cases} e^{[(a) \cdot \ln(tts_{AB\ k-1}) + (1-a) \cdot \ln(tts_{AB\ k-1})]} & \text{if } n_{vk-1} > 0 \\ tts_{AB\ k-1} & \text{if } n_{vk-1} = 0 \end{cases} \quad [5]$$

$$\sigma^2_{stt_{AB\ k}} = \begin{cases} a \cdot (\sigma^2_{stt_{AB\ k-1}}) + (1-a) \cdot (\sigma^2_{stt_{AB\ k-1}}) & \text{if } n_{vk-1} > 1 \\ \sigma^2_{stt_{AB\ k-1}} & \text{if } n_{vk-1} = \{0, 1\} \end{cases} \quad [6]$$

In both equations the expected average travel time and travel time variance are calculated assuming a lognormal travel time distribution to reflect the fact that travel times are skewed towards longer travel times. For instance, in Figure 1 the recorded travel times during the non-congested portion of the day continuously fluctuate around 2.4 min (travel time at 60 mi/h posted speed limit) and do not typically decrease below 2 min. Consequently, while some vehicles may travel at speeds higher than the speed limit, these vehicles will not exceed the speed limit significantly. Alternatively, it is possible for vehicles to travel at speeds that are much lower than the speed limit as traffic congestion builds up. This, in turn, creates a travel time distribution that is skewed towards longer travel times.

The exponential smoothing factor  $\alpha$  used in both equations is a low-pass filter that accentuates lower frequencies and suppresses higher noise frequencies for better travel time forecasts. Due to the stochastic nature of traffic, significant fluctuations in estimated travel times may be observed, particularly if the sampling intervals are very short. In turn, these fluctuations make it difficult to recognize underlying trends, thus creating a need for such a filter. To do so, the proposed procedure utilizes an adaptive exponential smoothing factor  $\alpha$  that varies depending on the number of observations in the sampling interval under consideration based on the value attributed to a sensitivity parameter  $\beta$ , as demonstrated in Equation 7.

$$a = 1 - (1 - \beta)^{n_{vk}} \quad [7]$$

The introduction of an adaptive smoothing factor  $\alpha$  in the algorithm is based on the concept that the level of confidence that should be placed on a travel time estimate from a given sampling interval should be proportional to the number of observations on which this estimate is derived. In other words, it recognizes that while the availability of two or three observations over a sampling interval provides a means to estimate the expectation of the population, the resulting average travel time is not as likely to be as accurate to an estimate that is based on, for example, 15 to 20 observations.

Figure 3 illustrates the variation in the smoothing factor  $\alpha$  based on the value assigned to the sensitivity parameter  $\beta$  and the number of valid observations in the current sampling interval. As can be observed, values for the smoothing factor  $\alpha$  typically vary between 0 and 1. A value of 0 means that no confidence is put on the estimated travel time from the current interval and that no fraction of this estimate should be used to update the smoothed travel time. The algorithm considers smoothing factors of 0 when no valid observations are recorded within an analysis interval. In such a case, the algorithm assigns the previously smoothed travel time to the current interval. Alternatively, a value of 1 means that full confidence should be put on the average travel time that is estimated from the current sampling interval and that this estimate should replace, in its entirety, the moving average. Any value between 0 and 1 would finally result in the calculation of an updated moving average travel time that is a weighted combination between the previously computed moving average travel time and the average estimated travel time from the current interval.

Finally, as can be observed in Equation 7, the sensitivity parameter  $\beta$  has not been assigned a fixed value, thus allowing the user to calibrate the smoothing parameter to local conditions under consideration. Specifically, as illustrated in Figure 3, the user has the flexibility to allow the smoothing factor to respond rapidly or slowly to the number of observations in a recording interval. A sensitivity analysis of this factor is presented later in the paper.

## 5.2. Travel Time Estimation within Basic Data Validity Window

Within each sampling interval, the basic data validity window is computed based on a confidence interval that is estimated using a user-defined number of standard deviations above and below the expected interval average travel time,  $n_{\sigma}$  as defined in Equations 8, 9 and 10.

$$Stt_{ABk} = \left\{ t_{Bi} - t_{Ai} \mid t_k - t_{k-1} < t_{Bi} \leq t_k \quad \text{and} \quad tt_{AB\min k} \leq t_{Bi} - t_{Ai} \leq tt_{AB\max k} \right\} \quad [8]$$

$$tt_{AB\min k} = e^{\left[ \ln(tt_{ABk}) - n_{\sigma}(\sigma_{st_{ABk}}) \right]} \quad [9]$$

$$tt_{AB\max k} = e^{\left[ \ln(tt_{ABk}) + n_{\sigma}(\sigma_{st_{ABk}}) \right]} \quad [10]$$

In the above equations, the parameters  $tt_{AB\min k}$  and  $tt_{AB\max k}$  represent the lower and upper limits for the valid travel time observations, while  $t_k$  represents the time at the end of each interval  $k$  at which the calculation takes place. In computing the confidence limits for the next time interval, the average travel time and travel time standard deviation of all valid observations within the current sampling interval must be known as these two elements are used in Equations 5 and 6. In developing the basic data filtering process, Equation 11 is used to estimate the average travel time for the current time interval between a pair of readers,  $tt_{ABk}$ , while Equation 12 is used to estimate the travel time standard deviation for the current time interval,  $\sigma_{st_{ABk}}$ .

$$tt_{ABk} = \frac{\sum_{i=1}^{n_{vk}} (t_{Bi} - t_{Ai})}{n_{vk}} \quad [11]$$



$$\sigma^2_{t_{ABk}} = \begin{cases} 0 & \text{for } n_{vk} = 0 \\ \frac{[\ln(t_{Bi} - t_{Ai})_k - \ln(tts_{ABk})]^2}{n_{vk}} & \text{for } n_{vk} = 1 \\ \frac{\sum_{i=1}^{n_{vk}} [\ln(t_{Bi} - t_{Ai})_k - \ln(tts_{ABk})]^2}{n_{vk} - 1} & \text{for } n_{vk} \geq 2 \end{cases} \quad [12]$$

In Equation 12, the variance is not calculated against the current average travel time but against the predicted average travel time, as given by Equation 5. Since the filtering algorithm is intended to be used in real-time applications in which there may be no a priori knowledge of future traffic conditions, its success in tracking changes in traffic conditions will heavily depend on its ability to adjust to changes. As shown in Figure 4, the utilization of the observed average travel time to calculate the variance within each interval would produce a filtering algorithm that is relatively insensitive to changes in traffic conditions. On the other hand, calculating the variance based on the predicted average travel time estimates larger variances, and thus larger confidence intervals, when travel times deviate from the previously estimated moving average travel time. In Figure 5, such a calculation clearly allows for a better tracking of changing traffic conditions. Another advantage of using the expected average travel times is the ability that is then given to the algorithm to calculate travel time variances, albeit crude estimates, for intervals with only one valid observation.

While the number of standard deviations that define the size of the validity window is user-definable in Equation 9 and 10, it is envisioned that basic validity ranges encompassing two or three standard deviations be utilized. The use of a search window that is two standard deviations wide would mean that all data points within a 95 percent lognormal confidence interval are to be considered as valid and that all other points falling outside this range are to be rejected from consideration when estimating average link travel times. Similarly, the use of a validity window that is three standard deviations wide would mean that all data points within a 99 percent confidence interval are to be considered as valid. A sensitivity analysis of this parameter is presented later in the paper.

To evaluate the ability of the filtering criterion to follow travel time fluctuations, Equations 5 through 12 were applied to the dataset of Figure 1 while making the following assumptions:

- Travel time information is updated every 2 min, as is done within the San Antonio AVI system,
- A value of 0.2 is used for the sensitivity parameter  $\beta$  in Equation 7, which determines the value of weighting factor  $\alpha$  with respect to the number of valid observations in the current sampling interval, and
- A value of 2 is assigned to the parameter  $n_\sigma$  in Equations 9 and 10, which results in the definition of a basic validity range encompassing two standard deviations.

As can be observed in Figure 6, the application of the filtering criterion defined by Equations 8 through 12 does not produce good results. First, the criterion is unable to track the sudden increase in travel times that occurs during the morning peak period. Second, there are a significant number of apparently valid data points that are being rejected over the entire day.

To test the hypothesis that the poor performance of the filtering algorithm is due to a large inherent variability in travel times, the filtering criterion was reapplied using a basic search window of three standard deviations instead of only two. As can be observed in Figure 7, the consideration of a larger search window greatly reduces the number of data points that are incorrectly assumed to be invalid. However, despite this improvement, it is observed that the filtering criterion remains unable to follow the underlying travel time increase between 6:30 and 9:00 a.m. This failure is caused by the combination of the abrupt

increase in roadway travel times and the low number of observations within each sampling interval. Because travel times change rapidly at the onset of the morning peak period, a gap of a few minutes is sufficient in this case to raise the average link travel time beyond the validity bounds. Since all subsequent observations lie outside the validity window, they are considered invalid and rejected. Only the reduction of travel times to values within the validity window, at the conclusion of the peak period, allows the filtering algorithm to accept new observations as valid.

### 5.3. Expanded Data Validity Range

In order to increase the algorithm's responsiveness to abrupt changes in travel times, modifications were made to allow the algorithm to search for trends of increasing or decreasing travel times outside the basic validity window (defined by Equations 8 through 12). The modifications specifically allow the algorithm to consider as valid the third of three consecutive points outside the validity window, provided that all three observations are either above or below the validity window.

Figure 8 illustrates how the expanded search window enhances the performance of the algorithm. The figure illustrates the variation in the interval average travel time, as well as the lower and upper limits of the validity window, after consideration of trends outside the basic validity range. In this case, most of the travel times sampled between 6:00 and 6:34 a.m. are considered valid since they almost all fall within the validity window limits defined by Equations 9 and 10. The next two data points, at 6:38 and 6:49 a.m., are then rejected based on the fact that they lie outside the validity window limits. This leads to no changes in the predicted average travel time and validity range of all sampling intervals between 6:34 and 6:54. The detection of a third consecutive observation above the limits of the current validity range in the 6:52 to 6:54 interval finally indicates that a trend of increasing travel times may exist. This results in the inclusion of a 6.6-minute travel time in the set of valid measurements, and in a subsequent update of the predicted average travel time and search window limits for the 6:54 to 6:56 interval. In turn, the inclusion of this data point in the set of valid observations leads to an increase in the expected average travel time for the next intervals, and allows the filtering algorithm to correctly track the increasing travel times that are observed after time 6:58.

To allow the tracking of sudden variations in traffic conditions, changes were made to both Equations 7 and 12. Equation 7, which determines the value of the smoothing factor  $\alpha$ , is substituted by Equation 13, while Equation 12, which estimates the variance of travel times against the expected interval average, is substituted by Equation 14.

$$a = \begin{cases} 1 - (1 - \beta)^{n_{vk}} & \text{for } n_a < 3 \text{ and } n_b < 3 \\ \max(0.5, 1 - (1 - \beta)^{n_{vk}}) & \text{for } n_a \geq 3 \text{ or } n_b \geq 3 \end{cases} \quad [13]$$

$$\sigma^2_{t_{ABk}} = \begin{cases} 0 & \text{for } n_{vk} = 0 \text{ and } n_a < 3 \text{ and } n_b < 3 \\ \frac{[\ln(t_{Bi} - t_{Ai})_k - \ln(tts_{ABk})]^2}{n_{vk}} & \text{for } n_{vk} = 1 \text{ and } n_a < 3 \text{ and } n_b < 3 \\ \frac{\sum_{i=1}^{n_{vk}} [\ln(t_{Bi} - t_{Ai})_k - \ln(tts_{ABk})]^2}{n_{vk} - 1} & \text{for } n_{vk} \geq 2 \text{ and } n_a < 3 \text{ and } n_b < 3 \\ 0.01 \cdot (tts_{ABk}) & \text{for } n_a \geq 3 \text{ or } n_b \geq 3 \end{cases} \quad [14]$$

Both equations introduce the parameters  $n_a$  and  $n_b$ , which is a counter for the number of consecutive observations above or below the validity window limits. The values for both parameters are determined by processing the travel time data in the order they are received. For instance, the reception of an observation

above the validity window results in  $n_a$  being incremented by 1. If the following observation is again above the validity range,  $n_a$  is increased to 2. If, instead, the following observation is valid,  $n_a$  is then reset to 0. If the following observation is however below the validity range,  $n_a$  is again reset to 0 but  $n_b$  is incremented by 1 in this case.

The main difference between Equations 7 and 13 is the addition of a fixed smoothing factor  $\alpha$  of 0.5 that is applied to the estimation of the next interval's expected average travel time and travel time variance each time a third consecutive data point either above or below the basic validity range is identified. Because the arbitrary inclusion of such a data point in the set of valid observations constitutes a break in the normal application of the exponential smoothing process defined by Equations 8 through 10, it was determined that a constraint should be applied on the weighting factor determined by Equation 7 to ensure that the filtering algorithm quickly adjusts to the trends of increasing or decreasing travel times. The impact of this constraint is very apparent in Figure 8, where it is observed that the inclusion of the 6.6-minute observed travel time at 6:53 in the set of valid travel times provides only one valid observation for the 6:42 to 6:54 interval. If Equation 7 were used, a value of 0.2 would be assigned to the smoothing factor  $\alpha$  with a sensitivity parameter  $\beta$  of 0.2. This would have resulted in an updated expected interval travel time of 179 s for the 6:54 to 6:56 interval, instead of the 242 s expected travel time. By 7:10, the expected interval travel time would have been estimated to be only 388 s, instead of 571 s, which is already lagging behind the true average.

The main difference between Equation 14 and 12 is again linked to the inclusion of a third consecutive data point outside the basic validity range in the set of valid observations. In this case, a new criterion for calculating the variance of travel times within a sampling interval is introduced in the equation. However, the purpose of this criterion is not to increase the sensitivity of the filtering algorithm to trends of changing travel times but to decrease its sensitivity. Since the variance of travel times within an interval is calculated against the expected average travel time, large variances are typically calculated for intervals containing observations lying outside the basic validity range. If these large variances are used to determine the basic validity range of the next interval, very wide search windows would be utilized. Such wide limits may then lead to the inclusion of very long, suspicious, travel times within the set of valid observations. This situation is thus prevented by constraining the value of the variance.

To evaluate the impacts of the proposed algorithm enhancements, Equations 8, 9, 10, 11 and 14 were applied to the dataset of Figure 1. To ensure a consistent comparison with previous results, identical operating parameters as those used to produce the results of Figure 7 were considered ( $t_w = 2\text{-min}$ ,  $\beta = 0.2$ ,  $n_\sigma = 3$ ). As can be observed in Figure 9, the proposed model enhancements significantly improve the operation of the algorithm by clearly allowing it to respond to abrupt changes in average travel times during the morning peak travel period.

At this point, it may be argued that both the TranStar and TransGuide algorithms could be fixed by implementing this criterion in their algorithms to accept as valid any third consecutive data point either above or below each algorithm's validity range. Figure 10 illustrates the results of considering such a change for the same dataset of Figure 9. In both cases, the changes made to the algorithms clearly enable them to track the sudden changes in travel times during the morning peak period. Consequently, the analysis has demonstrated that the current algorithms that are implemented in the TransGuide and TranStar systems can be enhanced by incorporating this criterion. However, as indicated by the results in the third diagram, the proposed algorithm is superior to the TransGuide and TranStar algorithms because it incorporates a low-pass filter that suppresses the high frequency white noise and thus produces more stable travel time estimates.

#### 5.4. Consideration of Low Sampling Rates

To further improve the filtering algorithm, additional enhancements were made. Given the variable nature of traffic, it was first observed that predicting the average travel times during an interval while using data collected in the previous intervals does not ensure that the resulting estimates are truly representative of the

interval's real average trip time. For instance, if traffic demand is slowly increasing during a given portion of the day, it can then be expected that the average travel time that is measured in consecutive intervals should gradually increase. Second, it was observed that the assumption that the expected average travel time and standard deviation of the validity window remain constant during intervals with no observations, as defined in Equations 5 and 6, can potentially result in the algorithm using outdated average travel times to determine the validity window limits.

In this case, a period with no recorded travel times does not mean that there is no traffic passing through the pair of AVI readers, but simply means that no vehicles equipped with tags are traveling. Thus, to increase the responsiveness of the algorithm to changes in traffic conditions to situations with low sampling rates, Equation 15 was introduced to modify the search window limits by computing the number of standard deviations  $n_{\sigma k}$  that should be used in Equations 9 and 10 within each sampling interval  $k$  based on the number of intervals with zero observations.

$$n_{\sigma k} = \lambda + \lambda \left[ 1 - (1 - \beta_{\sigma})^{n_{ok}} \right] \quad [15]$$

Equation 15 provides a model that dynamically adjusts the size of the validity window based on the number of preceding sampling intervals without AVI observations,  $n_{ok}$ , a user-defined parameter  $\lambda$  representing a minimum number of standard deviations to consider, and a sensitivity parameter  $\beta_{\sigma}$ . For any interval for which at least one observation was made in the preceding interval, the equation defines a validity window that corresponds to the minimum size specified by the user,  $\lambda$ . If no observation were made in the previous interval, the size of the validity window is increased to  $\lambda + \lambda (\beta_{\sigma})$ . If no observations were made in the two previous intervals, the validity window is then further increased to  $\lambda + \lambda [1 - (1 - \beta_{\sigma})^2]$ . The size of the validity window will continue to increase with every increase in the number of consecutive preceding intervals without observations, until a maximum size of  $2\lambda$  is reached.

Figure 11 illustrates the impact of Equation 15 on the algorithm's performance. The figure compares the various algorithms considering a freeway segment with a relatively low travel time sampling rate by applying a version of the filtering algorithm that includes only Equation 8, 9, 10, 11 and 14, to a version that also includes Equation 15. As can be observed there is a noticeable difference in the size of the validity window used by both versions of the filtering algorithm. While identical validity window limits are used by both versions of the filtering algorithm for all sampling intervals for which travel time observations were made in the preceding interval, increasing differences are observed for intervals that are preceded by an increasing number of intervals without observations.

In Figure 11, the impact of the modified filtering algorithm is particularly apparent in the intervals between times 16:50 and 17:30. Within this period, only three travel times are observed: a 2.42-minute travel time at 16:51, a 3.35-minute travel time at 17:25, and a 3.18-minute travel time at 17:27. After detection of the first vehicle passage, the minimum and maximum search window limits for the subsequent sampling interval, 16:52 to 16:54, are set at 1.94 min and 3.15 min, respectively, by both versions of the algorithm. As time passes without any additional observations, the filtering algorithm based only on Equations 8, 9, 10, 11 and 14 maintains a fixed validity window, while the algorithm that includes Equation 15 gradually alters the size of the validity window to 1.66 min and 4.17 min by the time the 17:24 to 17:26 interval is reached. This results in the 3.35-minute travel time observation at 17:25 being considered as valid by the modified algorithm that includes Equation 15. A visual analysis of the time series of recorded travel times appears to indicate that the 3.35-minute observation should be considered.

## 5.5. Consideration of Successive Link Arrival and Departure Times

Testing of the algorithm on a number of additional travel time datasets further revealed that the use of an additional filter based on consecutive travel times could provide increased robustness to the filtering process. In a hypothetical situation where all vehicles travel at identical speeds, two vehicles entering a link in succession would be expected to exit the link in the same order in which they entered. Alternatively, if the vehicles travel at different speeds the order of vehicles entering and exiting the link may change as a result

of one vehicle overtaking the other. If, however, the difference in vehicle travel times is assumed to vary within 2 to 3 standard deviations, an additional criterion can be set to ensure that longer travel times do not exceed a user-defined difference relative to other vehicles that travel the same link during the same time interval. This criterion is based on the fact that if two vehicles are traveling on the same link at about the same time they should therefore be subject to similar traffic conditions. For instance, consider a vehicle B that is detected to exit a link a few seconds after a vehicle A. If the vehicle B has been detected to enter the link 5 min earlier than the vehicle A, there is then a reasonable indication that this vehicle may have stopped along the link and that its observed travel time should be eliminated from the dataset of valid link travel times.

Equation 16 is used to verify the validity of observed travel times on a given link based on the sequence of vehicle entry and exit times from the link. The equation indicates that any observed travel time from a vehicle  $i$  that is exiting a link will be considered part of the set of valid observations  $Stt_{ABk}$  provided that this vehicle does not experience a travel time that is significantly different from a similar vehicle within the same time frame. Similar to the determination of the basic data validity range, the allowed variation in link entry time is set to correspond to the estimated shorter travel time plus the confidence interval of two standard deviations.

$$Stt_{ABk} = \left\{ t_{Bi} - t_{Ai} \mid (t_{Bi} - t_{Ai}) \leq (t_{B_{i-1}} - t_{A_{i-1}}) + e^{2(\sigma_{st_{ABk}})} \text{ for } t_{Bi} \geq t_{B_{i-1}} \text{ and } t_{Ai} \leq t_{A_{i-1}} \right\} \quad [16]$$

## 6. ALGORITHM TESTING AND VALIDATION

To evaluate the ability of the proposed filtering algorithm to correctly follow fluctuations in observed travel times, the algorithm defined by Equations 8, 9, 10, 11, 14, 15 and 16 was applied to two series of travel time readings from the San Antonio AVI network. The first series consists of observations that were made on I-35 South, between Ritiman Rd. (AVI station 45) and Walzden Rd. (Station 44) over a period of 10 consecutive days in June of 1998. This freeway segment was utilized earlier to illustrate the development of the algorithm. This link is 3.955 km (2.458 mi) long and has a posted speed limit of 96 km/h (60 mi/h), yielding a nominal free-flow link travel time of 2.45 min. The second series of observations consists of travel time readings that were made along an arterial. This series consists of travel times that were observed along Fredericksburg Rd., between Datapoint Dr. (AVI station 16) and Magic Dr. (AVI station 17) over the same 10 weekdays as the first series of freeway observations. This arterial segment is 1.873 km (1.165-mi) long and has a posted speed limit of 72 km/h (45 mph), yielding a free-flow link travel time of 1.55 min.

Figure 12 and Figure 13 illustrate the application results of the proposed filtering algorithm to the travel time series described above. The results shown in each figure are those associated with the calibrated parameters producing the best results. As can be observed, the differing numbers of observation and differing nature of traffic behavior along the freeway and arterial segments created a need to use slightly different parameters. For both the freeway and arterial segments, the filtering algorithm was able to correctly track all major changes in observed travel times. While one may argue about the validity of some observations, these are relatively few in number and do not appear to significantly affect the underlying trend. In particular, it is observed in Figure 12 that the filtering algorithm is able to effectively respond to abrupt increases and decreases in observed travel times. This ability to track these changes is particularly apparent in the freeway data of June 19, which illustrates a significant change in travel times that was likely caused by the onset of congestion created by an incident on the freeway.

In addition to successfully demonstrating the ability to respond to the underlying trend, the results of Figure 12 show more specifically the ability of the proposed algorithm to follow the general travel time fluctuations on links with very low sampling rates. As indicated earlier, this is a problem with existing algorithms. As shown in the figure, the total number of AVI readings made throughout an entire day on the test arterial segment varies between 56 and 75. In contrast, between 519 and 718 readings were made each day on the freeway link of Figure 11, which are also considered small numbers. As a result of these very low sampling

rates, the diagrams of Figure 12 typically feature large intervals with no observations. In most of the diagrams for instance, there are typically no AVI readings before 7:00 a.m. or after 10:00 p.m. Between these two time points, it is also not uncommon to observe periods of more than 30 min without any observation. The presence of these large gaps thus greatly increases the task of the filtering algorithm to identify underlying low frequency trends from the high frequency white noise, or to determine whether the travel times that are observed after a large gap truly represent existing traffic conditions or should be considered as invalid.

The scenarios of Figure 12, and to some extent those of Figure 11, may be considered as extreme conditions as AVI systems are not typically intended for estimating travel times in such conditions. The application results nevertheless clearly demonstrate the robustness of the proposed filtering algorithm in such conditions. In particular, it can be expected that the availability of larger numbers of observations would result in improved reliability and accuracy for the travel time estimates that are produced by the filtering algorithm, as the availability of more information reduces the number of periods without observations and reduces the impacts of incorrectly rejecting or accepting travel time observations.

## 7. SENSITIVITY OF ALGORITHM TO MODEL PARAMETERS

Other test applications of the algorithm looked more specifically at the sensitivity of the algorithm's results to values assigned to various input parameters. Figures 14 through 17 present the results of these evaluations. All figures use the same dataset for consistency of comparisons, which was used to illustrate the model development in Figures 1 through 10. A sensitivity analysis of the impact of the number of standard deviations in setting the validity range is not examined here as it was previously discussed.

Figure 14 first illustrates the sensitivity of the algorithm to the value of the  $\beta$  parameter. As was indicated in Figure 2, this parameter is used to determine the value of the smoothing parameter  $\alpha$ . As indicated in the figure, the assignment of decreasing values to this parameter increases the suppression of the high frequency noise signal while bringing out the underlying low frequency signal. At the limit, the assignment of a value of 1.00, the maximum value allowed, results in an unsmoothed signal. It is observed from Figure 14 that high parameter values tend to create significant fluctuations in the estimated average travel time from one interval to the next. On the other end, parameter values approaching 0 tend to create a lag in the algorithm's response to changes in travel times as a result of over-smoothing the data. Specifically, a parameter value of 0 produces average travel time estimates that are insensitive to traffic changes, as illustrated in the lower right diagram. In this case, the jump that occurs near 7:00 a.m. is simply the result of the criterion allowing the algorithm to readjust when three consecutive data points are observed outside the validity window. Any effect of lower parameter values is to increase the width of the validity range. This impact is linked to the use of the parameter  $\alpha$  in Equation 6 to calculate smoothed estimates of the variance of travel times. For this example, desirable values for the parameter  $\beta$  appear to be in the 0.20 to 0.60 range given that they suppress the high frequency noise without altering the underlying low frequency trend.

Figure 15 illustrates the sensitivity of the algorithm to the parameter  $\beta_{\sigma}$  which is used in Equation 15 to determine the size of the validity window to consider based on the number of preceding intervals without observations. In this case it can be observed that increasing the value of this parameter tends to produce wider validity ranges as expected. While changes in the parameter typically do not affect the operation of the filtering algorithm for the selected example, it would be expected that increasing the parameter values would affect applications to scenarios with low sampling rates and significant travel time fluctuations by allowing the algorithm to better capture travel times that significantly differ from the previously estimated average following a period without observations.

Figure 16 illustrates the sensitivity to the number of consecutive data points outside the validity range triggering the acceptance of an outside point as valid,  $n_{skips}$ . As observed, increases this parameter primarily leads to an increased lag in the algorithm's ability to respond to the onset of the morning peak congestion.

Because of the smoothing process, any increased delay in the algorithm's response to changing in traffic conditions also leads to a delay in the recovery period during which the estimated average travel times trail the true averages. For  $n_{skips} = 10$ , the estimate average traffic time can be seen to trail until 7:30 a.m. For  $n_{skips} = 4$  or 2, the trailing only lasts until 7:15. Therefore, while the adoption of a greater number of consecutive observations triggering the acceptance of a data point outside the validity range would reduce the occurrence of sudden adjustments, it would also increase the lag from actual conditions. Experiments with the algorithm suggest that a good compromise may be to select values between 3 and 5. In particular, it can be expected that increases in the sampling rates would reduce the occurrence of sudden adjustments as there would then be a reduced likelihood of sudden changes in travel times not reflected in the averages. Higher sampling rates would thus tend to reduce the influence of the parameter considered in this paragraph.

Finally, Figure 17 illustrates the sensitivity of the algorithm to the validity window size. Based on the diagrams shown in the figure, the size of the validity window appears not to affect significantly the operation of the algorithm. An absence of impact can be expected for periods with stable traffic conditions, as adding more observations within an interval should not significantly change the average travel time based on the observations and the determination of validity range for the next intervals. Using shorter or longer intervals will, however, affect the behavior of the algorithm in periods of fluctuations. Typically, shorter intervals will allow the estimated average travel times to produce smoother transitions. Short intervals would also allow the average travel time estimated for each sampling interval to better match the individual travel times as there would then be less probability of individual travel times differing significantly from the average as a result of changing conditions. However, with the use of shorter intervals also comes the risk of obtaining intervals without observations. A rule of thumb would therefore be to allow the use of increasingly shorter intervals with increasing sample rates.

## 8. CONCLUSIONS AND RECOMMENDATIONS

The paper described a low-pass filtering algorithm for predicting average link travel times using AVI data. The proposed algorithm overcomes a number of shortcomings of existing algorithms by handling both stable and unstable traffic conditions and functioning at levels of market penetration (less than 1 percent of the traffic volume). The low-pass filtering algorithm determines the range of valid travel time observations within each sampling interval by considering the number of observations in the current and previous sampling intervals, as well as the number of consecutive observations outside the validity range. Applications of the algorithm to two datasets of observed link travel times that were collected by the San Antonio AVI system, one from a freeway link and the other from an arterial link, demonstrate the ability of the proposed algorithm to correctly track the underlying average roadway travel times, while suppressing high frequency noise. A limited sensitivity analysis demonstrated that the proposed algorithm operates best when the smoothing factor  $\beta$  ranges between 0.2 and 0.5 and the  $n_{skips}$  parameter ranges between 3 and 5.

The paper also demonstrated that the existing TranStar and TransGuide algorithms can be improved by considering as valid any observation following a trend of similar invalid observations, as is done in the proposed algorithm. The modified TranStar and TransGuide algorithms, however, do not perform any low-pass filtering of the data and thus do not suppress high-frequency noise. In addition, by assuming a fixed validity window with respect to the estimated average travel time, both algorithms would also remain insensitive to the true variability in traffic conditions.

Despite the successful application of the proposed algorithm to two distinct datasets of AVI data and the limited sensitivity analysis that was conducted, further tests are required to validate the findings of this study. The use of historical data as an additional validation criterion or a means to provide travel time information during long intervals without travel time observations from AVI readers should also be investigated.

## ACKNOWLEDGEMENTS

This research effort was supported of the Virginia Department of Transportation (VDOT) and the Intelligent Transportation System (ITS) Implementation Center.

## REFERENCES

- Abdulhai B., Porwal H. and Recker W. (1999) “Short-term Freeway Traffic Flow Prediction Using Genetically Optimized Time-Delay-Based Neural Networks.” *Transportation Research Board 78th Annual Meeting*, Washington D.C.
- Ahmed M. and A Cook. (1982), *Analysis of Freeway Traffic Time Series Data by Using Box-Jenkins Techniques. Transportation Research Record 722*, TRB, National Research Council, Washington D.C., 1-9.
- Ben Akiva M., Kroes E., De Ruiter J, Smulders S, Whittaker, J and Lindveld, K. (1992), “Real-Time Prediction of Traffic Congestion.” *Vehicle Navigation and Information Systems: Conference Record of Papers*, New York, N.Y, Institute of Electrical and Electronics Engineers, 557-562.
- Coifman, B. (1998), *Vehicle Reidentification and Travel Time Measurement in Real-Time on Freeways Using the Existing Loop Detector Infrastructure, Transportation Research Record*, No. 1643, Transportation Research Board, 181-191.
- Coifman, B. and Cassidy, M. (2002), *Vehicle Reidentification and Travel Time Measurement on Congested Freeways, Transportation Research: Part A*, Vol. 36(10), 899-917.
- Coifman, B. and Ergueta, E. (2003), *Improved Vehicle Reidentification and Travel Time Measurement on Congested Freeways, ASCE Journal of Transportation Engineering*, Vol. 129(5), 475-483.
- ETTM on the Web (2001). *United States ETC Systems In Production / Development*. <http://www.ettm.com>. Updated August 1, 2001. Accessed August 28, 2001
- Houston TranStar (2001). *TranStar Description*. <http://traffic.tamu.edu/central2.html>. Accessed June 14, 2001.
- Ishak S. and Al-Deek H. (2003) “Performance Evaluation of a Short-Term Freeway Traffic Prediction Model.” *Transportation Research Board 82nd Annual Meeting*, Washington D.C..
- Lee S. and Fambro D. (1999) “Application of the Subset ARIMA Model for Short-Term Freeway Traffic Volume Forecasting.” *Transportation Research Record* 1678, 179-188.
- Mouskos, K.C., E. Niver, L.J. Pignataro, and S. Lee (1998). *TRANSMIT System Evaluation*. Final Report, Institute for Transportation, New Jersey Institute of Technology, Newark, NJ.
- Park D. and Rilett L. (1998) “Forecasting Multiple-Period Freeway Link Travel Times using Modular Neural Networks.” *Transportation Research Record* 1617, 163-170.
- Rilett L. and Park D. (1999) “Direct Forecasting of Freeway Corridor Travel Times using Spectral Basis Neural Networks.” *Journal of Transportation Engineering*, 125(6), Nov/Dec.
- SAIC (2000). *Automated Vehicle Identification Tags in San Antonio: Lessons Learned from the Metropolitan Model Deployment Initiative*. Science Applications International Corporation, McLean, VA.
- SwRI (1998). *Automatic Vehicle Identification Model Deployment Initiative – System Design Document*. Report prepared for TransGuide, Texas Department of Transportation, Southwest Research Institute, San Antonio, TX.
- Turner, S.M., W.L. Eisele, R.J. Benz, and D.J. Holdener (1998). *Travel Time Data Collection Handbook*. Report FHWA-PL-98-035, Federal Highway Administration, Office of Highway Information Management, Washington, D.C.
- Vickich, M. (2001). Personal e-mail communications, June-August 2001.



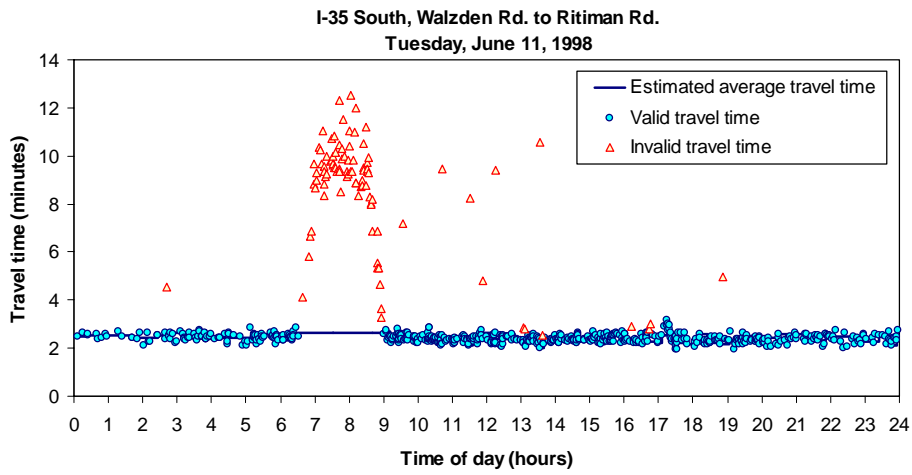


Figure 1. Application of TransGuide filtering algorithm to travel times from the San Antonio AVI system.

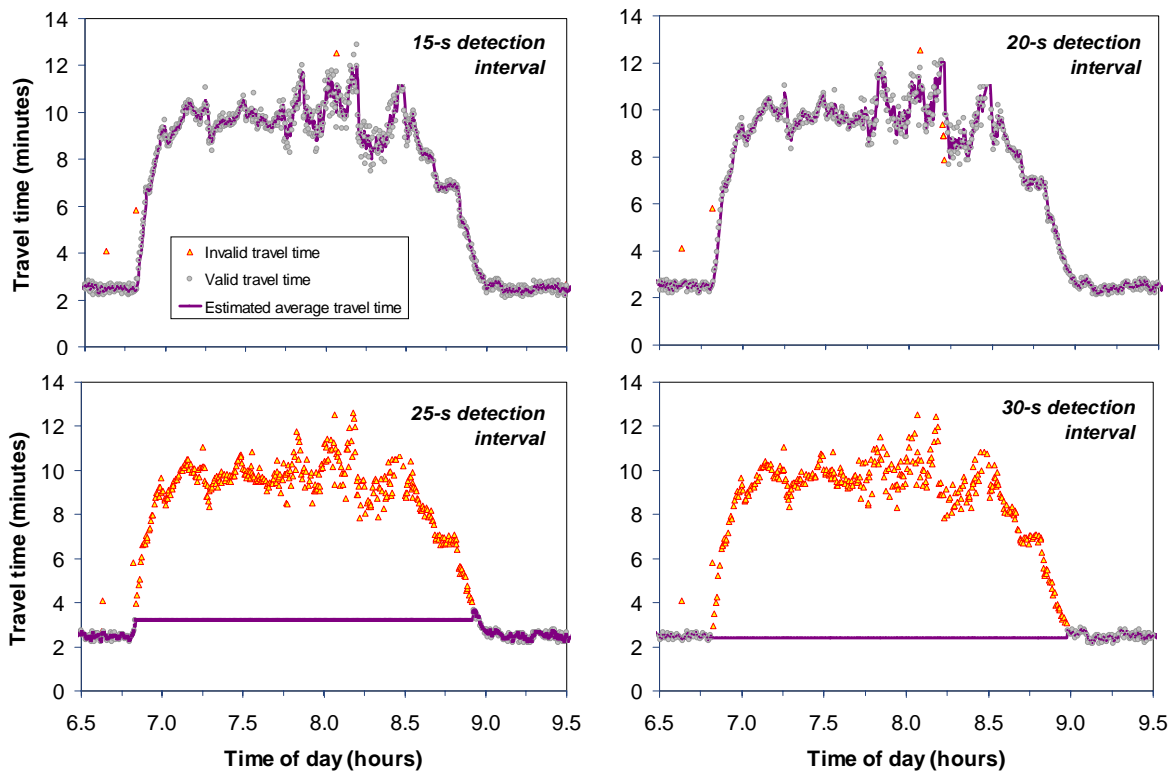


Figure 2. Impact of sampling rates on TranStar filtering algorithm.

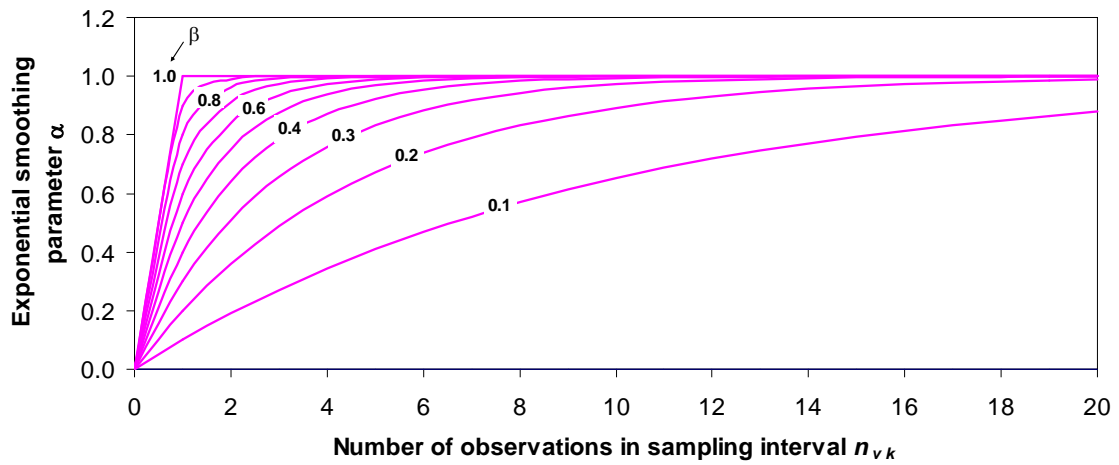


Figure 3. Value of smoothing factor  $\alpha$  as a function of number of observations in sampling interval and sensitivity parameter  $\beta$ .

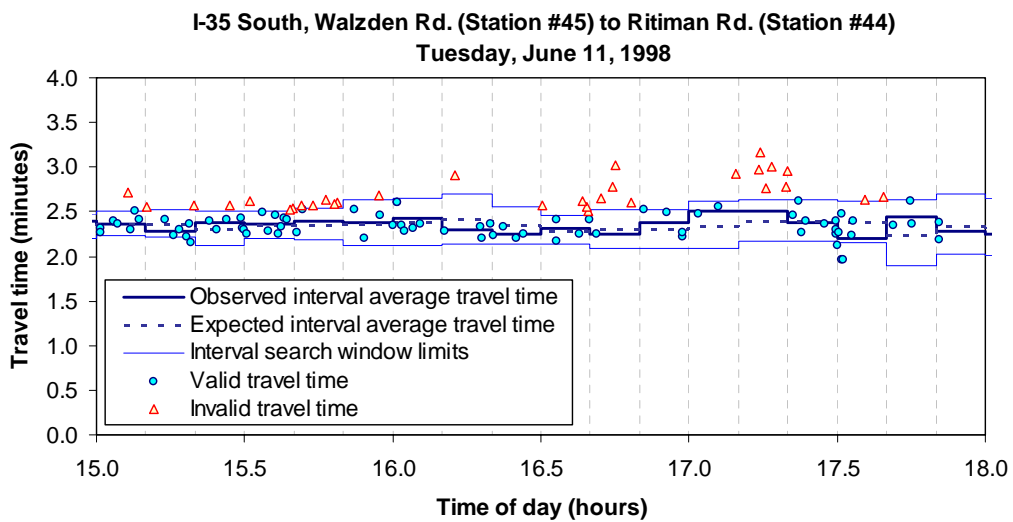


Figure 4. Application of filtering algorithm to dataset of Figure 1 using observed interval average travel times to determine the limits of the basic search widow.

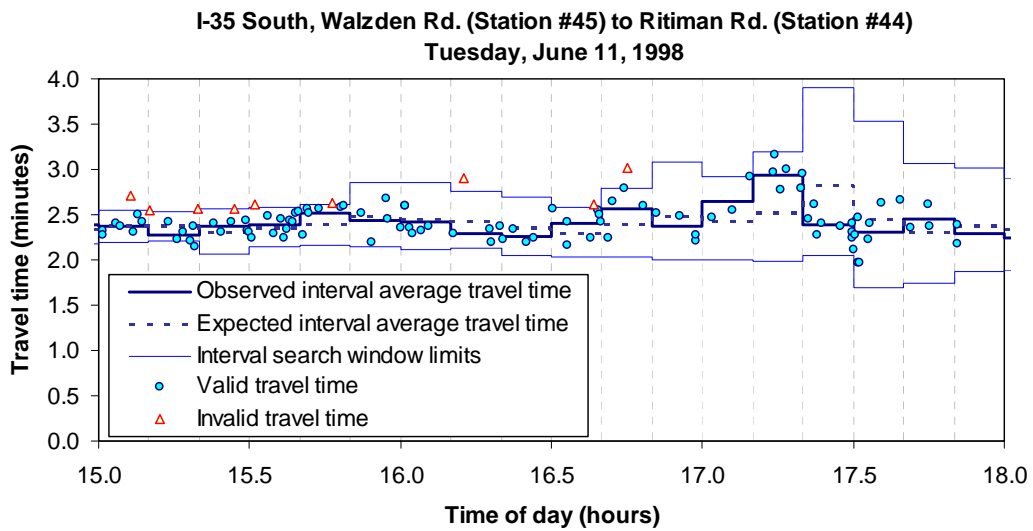


Figure 5. Application of filtering algorithm to dataset of Figure 1 using expected interval average travel times to determine the limits of the basic search widow.

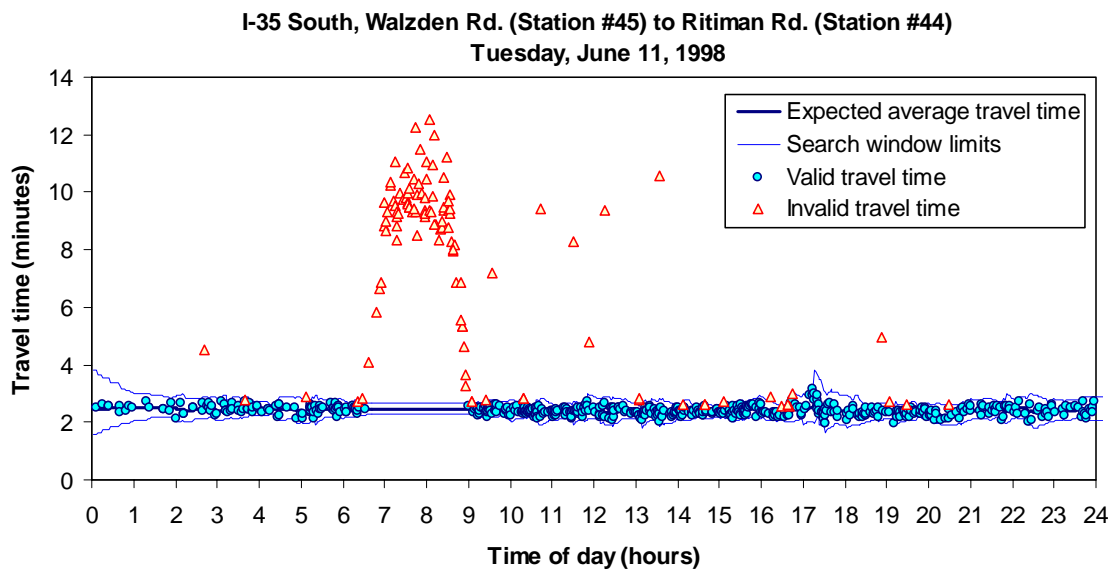


Figure 6. Application of filtering algorithm with two standard deviations as validity window on dataset of Figure 1.

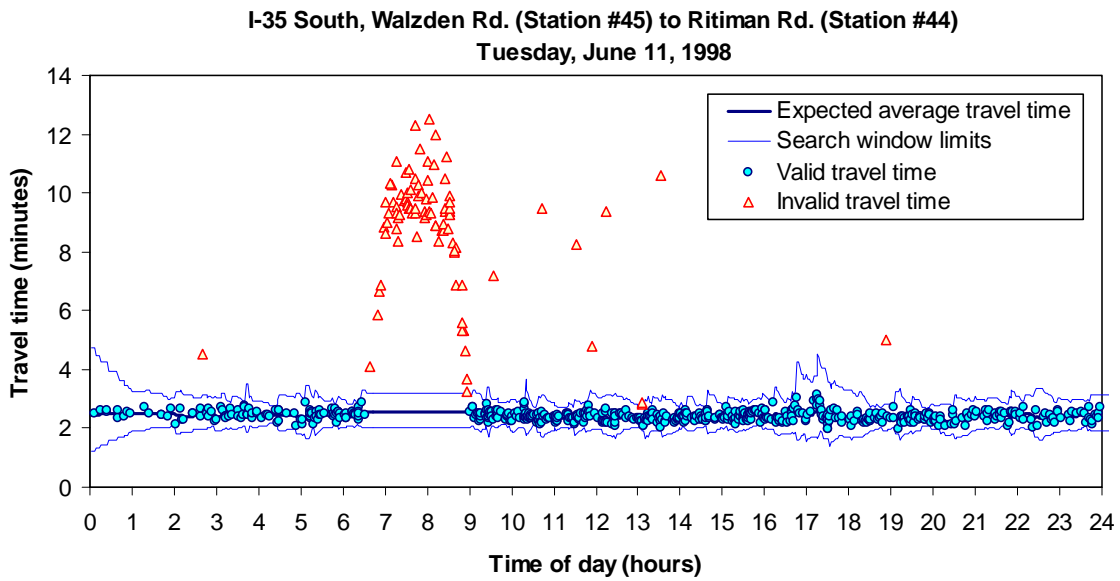


Figure 7. Application of filtering algorithm with three standard deviations as validity window on dataset of Figure 1.

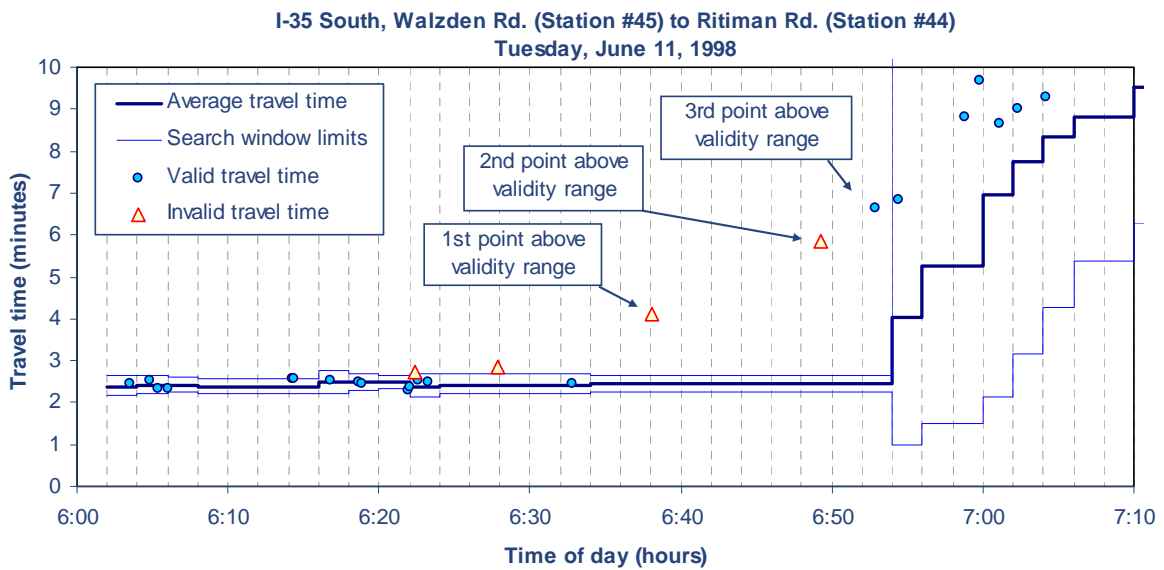


Figure 8. Example of expanded search beyond limits of validity window.

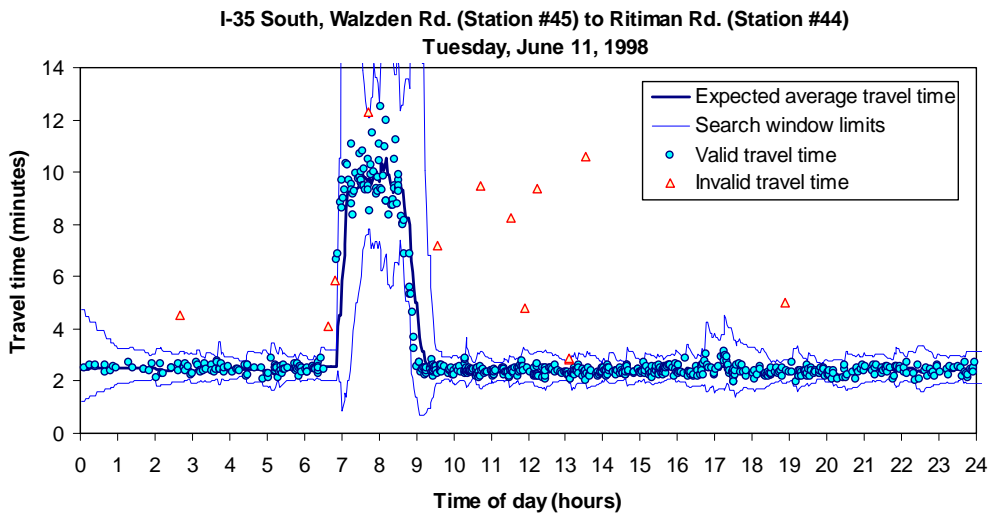


Figure 9. Application of filtering algorithm with expanded search algorithm on dataset of Figure 1.

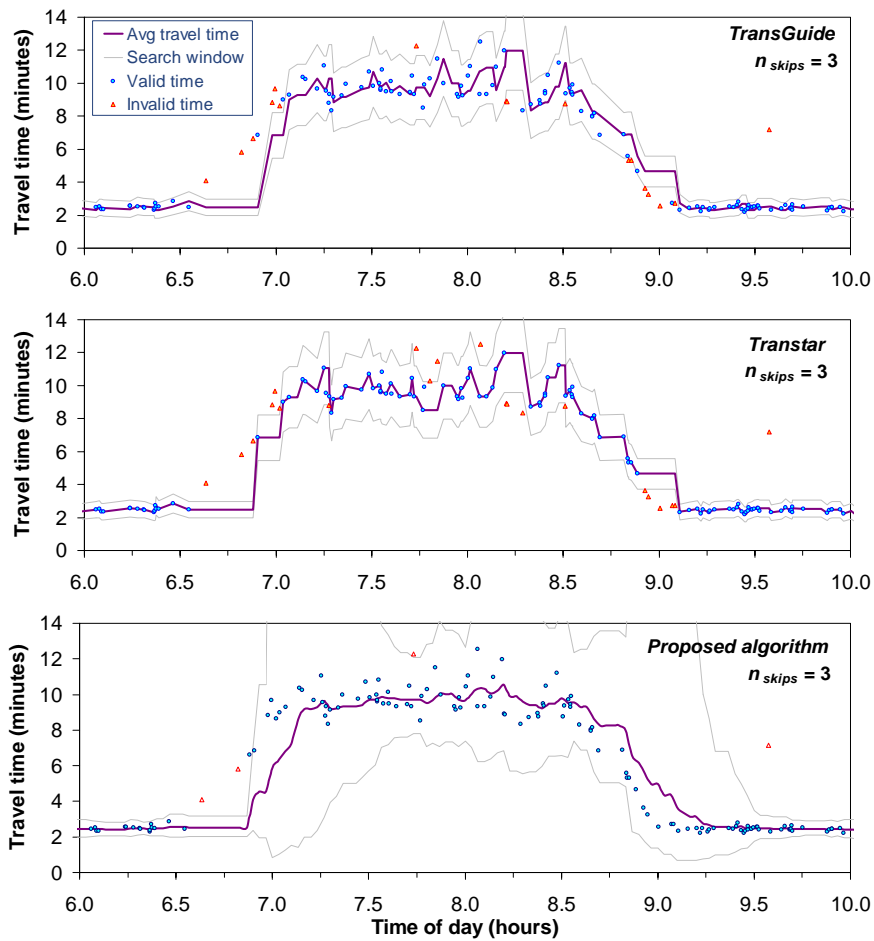
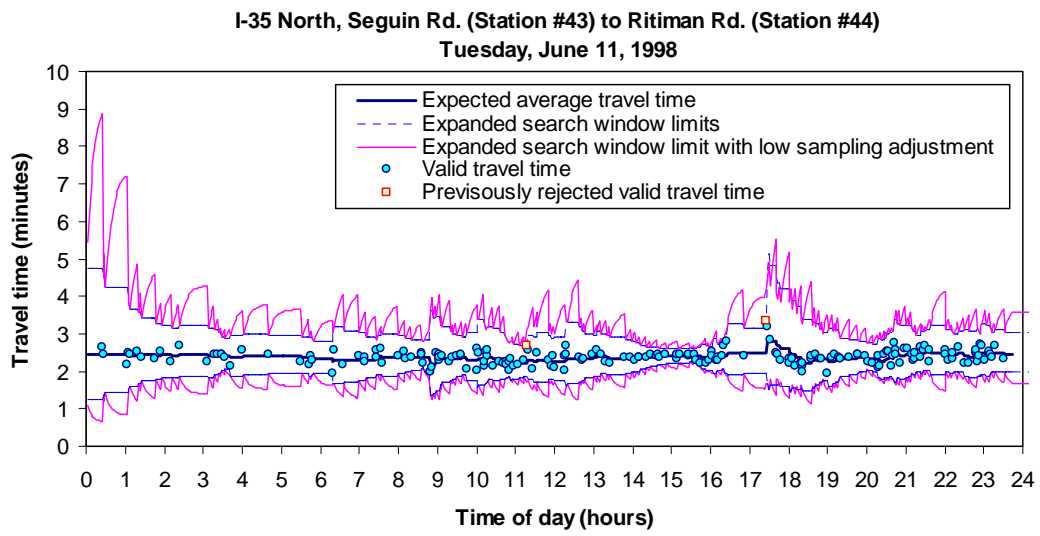


Figure 10. Application of filtering algorithm with expanded search algorithm on dataset of Figure 1.



**Figure 11. Impact of low sampling search limits on the operation of the filtering algorithm**

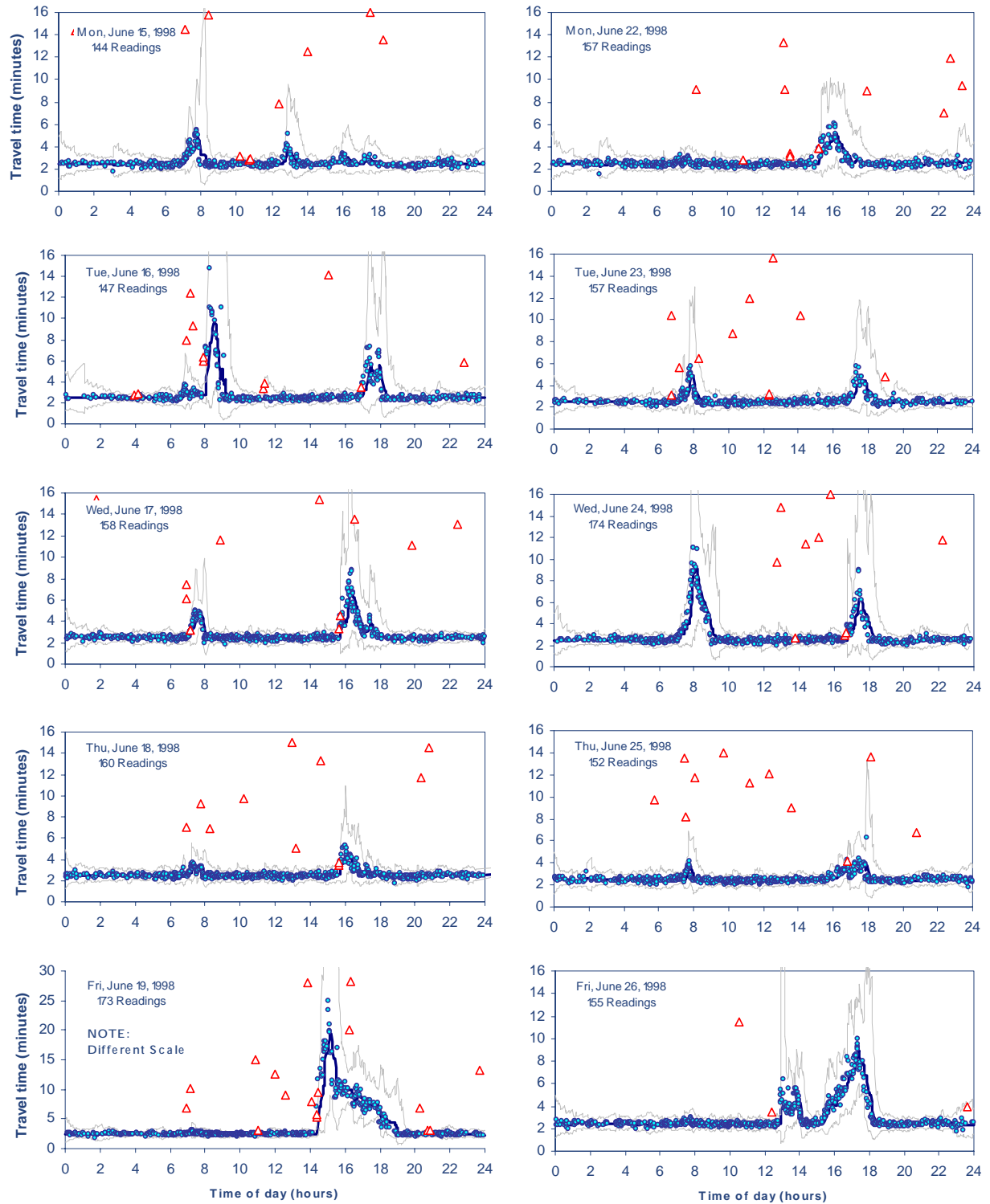


Figure 12: Sample application to a freeway roadway segment with  $t_w = 2$  min,  $\beta = 0.20$ ,  $\beta_\sigma = 0.05$ ,  $\lambda = 3$ .

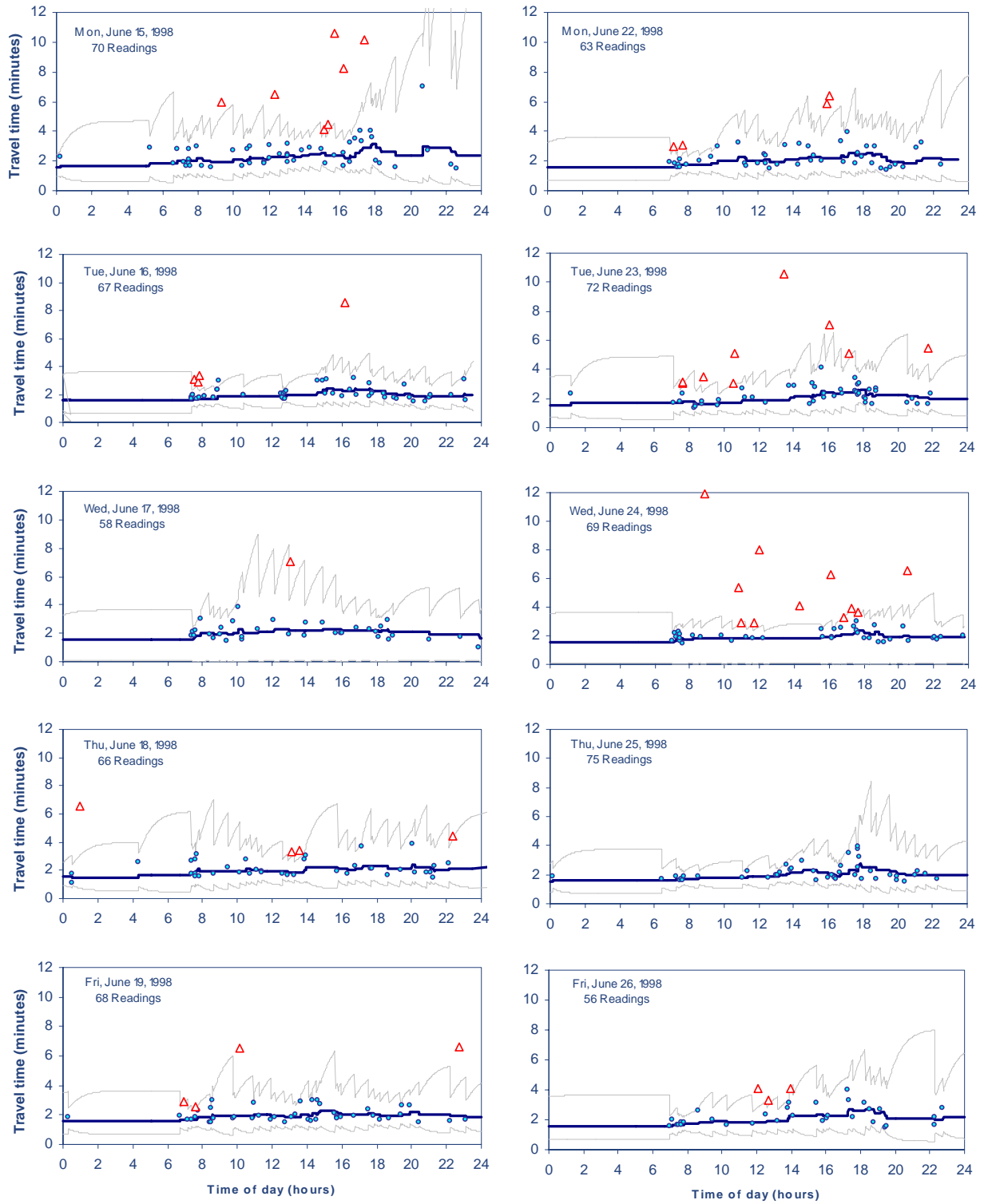


Figure 13: Sample application to an arterial roadway segment with  $t_w = 2$  min,  $\beta = 0.30$ ,  $\beta_\sigma = 0.05$ ,  $\lambda = 2$ .



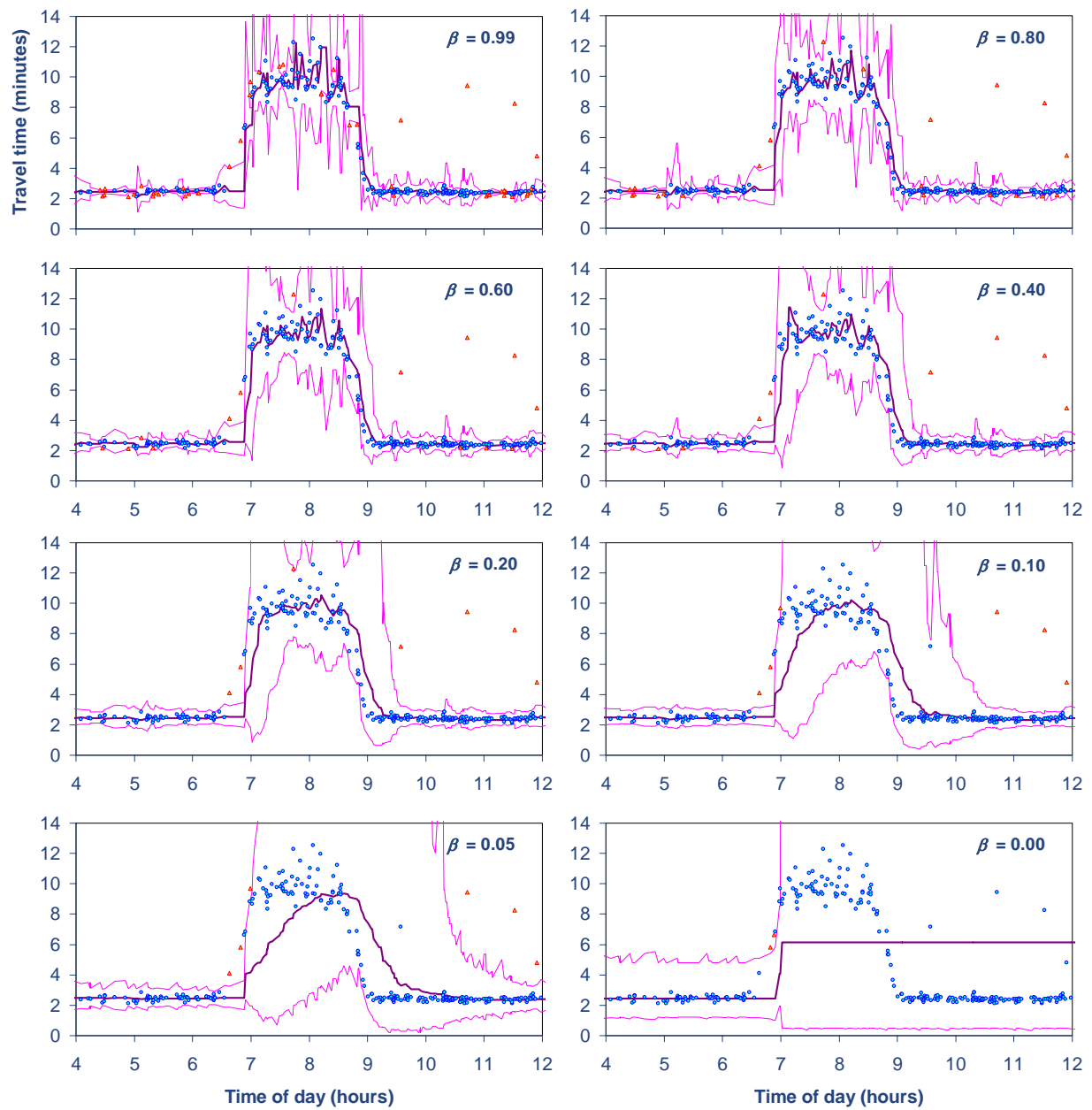


Figure 14: Impact of parameter  $\beta$  used for determining the smoothing parameter  $\alpha$  ( $t_w = 2$  min,  $\lambda = 3$ ,  $\beta_\sigma = 0.2$ ,  $n_{skips} = 3$ ).

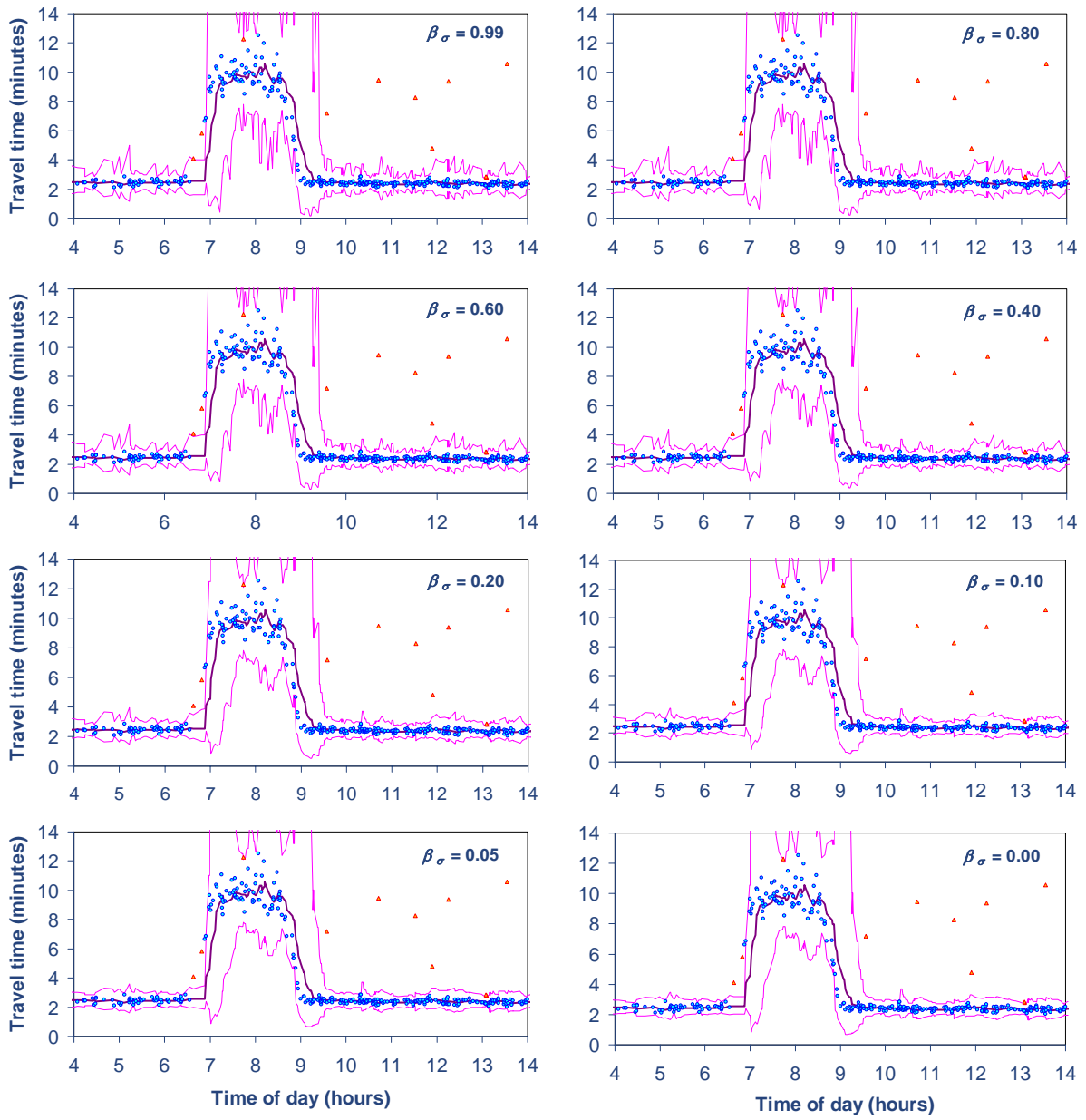


Figure 15: Impact of parameter  $\beta_\sigma$  used to determine the increase in validity range with increasing intervals without observations ( $t_w = 2$  min,  $\lambda = 3$ ,  $\beta = 0.2$ ,  $n_{skips} = 3$ ).

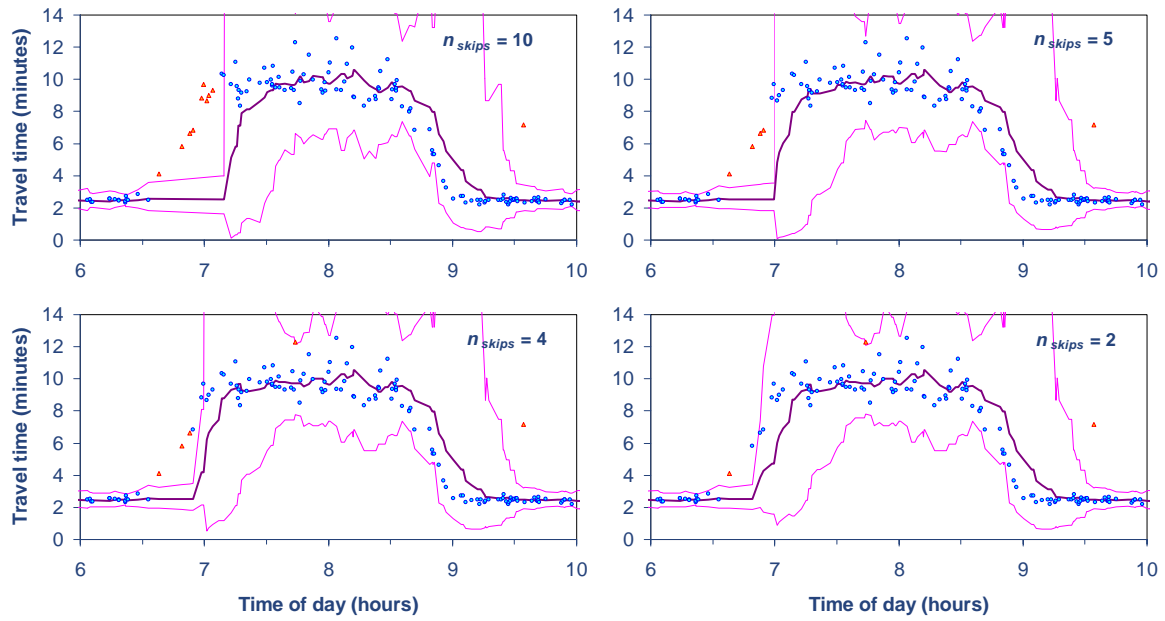


Figure 16: Impact of number of data points outside validity range triggering the acceptance of an observation ( $t_w = 2$  min,  $\lambda = 3$ ,  $\beta = 0.2$ ,  $\beta_\sigma = 0.2$ ).

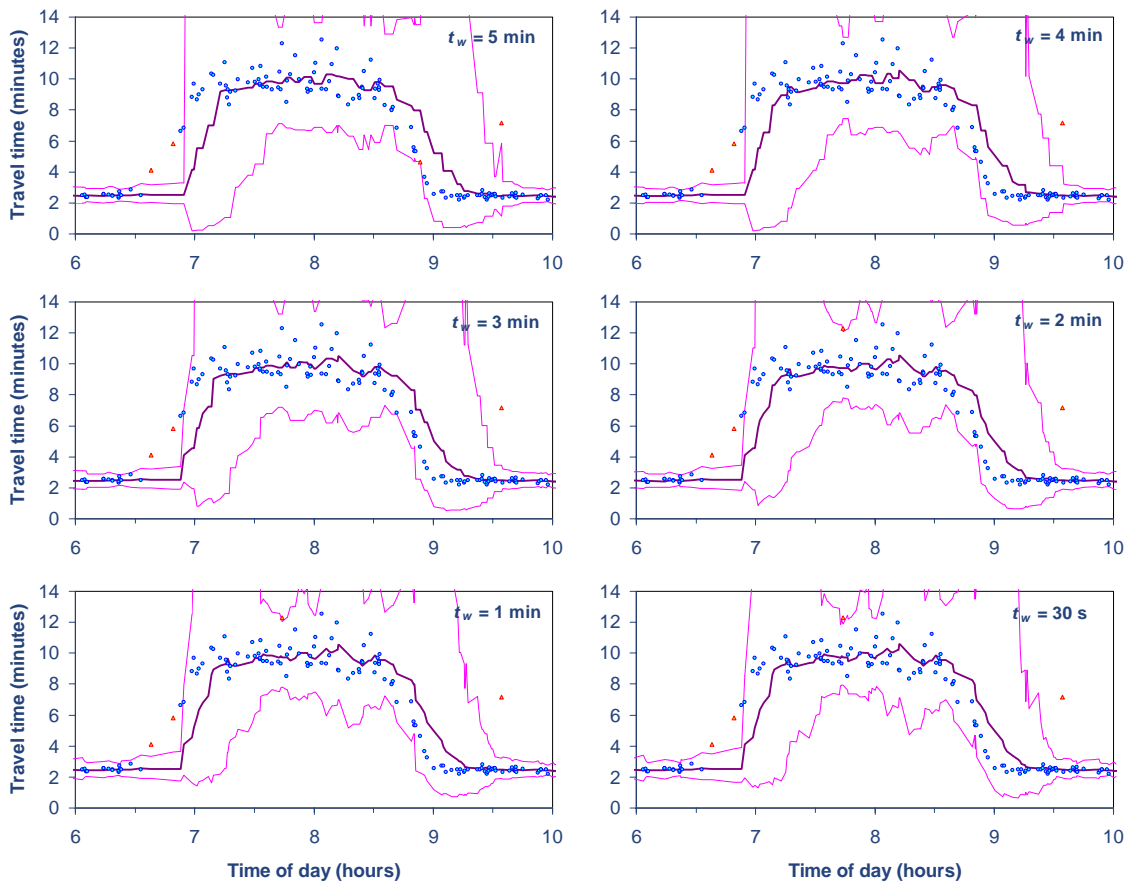


Figure 17: Impact of sampling window size ( $\lambda = 3, \beta = 0.2, \beta_\sigma = 0.2, n_{skips} = 3$ ).