Travel Time Estimation Based on Fused Traffic State Data: Case Studies in US and South Korea

Soknath MIL\textsuperscript{a}, Mongkut PIANTANAKULCHAI\textsuperscript{b}

\textsuperscript{a,b}School of Civil Engineering and Technology, Sirindhorn International Institute of Technology, Thammasat University, P.O.Box 22, Pathum Thani 12121, Thailand
\textsuperscript{a}E-mail: soknath.siit@gmail.com
\textsuperscript{b}E-mail: mongkutp@gmail.com

Abstract: The goal of Advance Traveler Information System (ATIS) is to improve traffic flow and safety by providing up-to-date information of traffic network. In this paper, traffic information (travel time) estimation based on fused traffic state data is presented. A centralized architecture is used to fuse the traffic state data from different sensors based fusion by averaging and fusion by median and to estimate the travel time based on the simultaneous travel time estimation model accordingly. Two case studies are selected to investigate the performance of fusion models based on freeway data in USA and the arterial road data in South Korea. The results show that the fusion by median performs best. The model is able to eliminate outliers in the data with less effort of complex mathematical process. It can be used as a benchmark for comparison with other advanced fusion models.

Key Words: ATIS, Centralized architecture, Travel time estimation, Data fusion

1. INTRODUCTION

In response to congestion problem, the applications of Intelligent Transportation System (ITS) are a viable option to improve traffic flow by providing reliable traffic information to the road users. One among those application is the Advance Traveler Information System (ATIS). It aims to disseminate travel time to the road users, since the travel time is one of the most valuable information for both road users and traffic managers to deal with congestion. Road users may decide to change routes or departure time according to the information provided and their acceptance level. According to the Tokyo Metropolitan Expressways (MEX) survey, the decisions on route changes are insensitive to the trip length. The level of accuracy should stay within 5 minutes or 10 minutes (Edward et al., 2004). This is an important threshold value for evaluating the travel time estimation model. Even though provision of travel time information is a key in reducing congestion, there are adverse effects in providing the wrong information as well such as annoyance, confident destruction, and congestion. To avoid problems in estimation, one should carefully investigate the accuracy of traffic data.

Two main sources of traffic data are on-road sensor and probe vehicle sensors. On-road sensors generally utilize conventional data collection technologies such as manual count, loop detectors, passive and active infrared sensors, microwave radars, ultrasonic and passive acoustics, and video image detection. On the other hand, probe vehicle data refers to the new type of data collection called Floating Car Data (FCD\textsuperscript{1}). FCD includes both GPS equipped vehicle and cellular phone data (Leduc, 2008). The two main differences of the above two

\textsuperscript{1} FCD sometimes are called floating cellular data
sensor types are their temporal coverage and spatial coverage. Data obtained from on-road sensors are often of large temporal coverage such that all the time data are collected. However, loop detectors are generally available only on major roads and distributed over long distance because they are costly. Thus the data obtained from them are normally of low spatial coverage. In contrast, probe vehicle sensors normally send out data at regular-time interval. Hence the location is not fixed, which can be covered over the city (spatial coverage) depending on the availability of sensor-equipped vehicles (sample size). However, in general, the sample size of probe vehicles is small. In spite of that, as a complement to conventional on-road sensor traffic data, floating car data is currently becoming widely utilized.

In response to the need for the improvement in the reliability of traffic data, data fusion, which combines the data from above types of sensors, is currently becoming an interesting topic in scientific research as well as in transportation field of research. Data fusion technique was firstly developed by the U.S. Department of Defense in the late 1980s for detecting, classifying, identifying, and tracking of targets. Many different applications of data fusion were developed based on military needs (Chang et al., 2014). Later, in traffic research area, data fusion has attracted more researchers due to more availability of data, which can be accessed both off-line and on-line. There are two main fusion algorithm found in the literature, direct fusion model and indirect fusion model in Figure 1 (Mori et al., 2015). In the first approach, the data are directly fused from different sensors and the estimated model is constructed. The second approach, the data are first processed to get a unique format (i.e. travel time) and finally fused together.

**Figure 1. Fusion approaches in travel time estimation (source: Mori et al., 2015)**

For the first approach, one way of combining traffic data from different source is the feedforward neural network method present by several authors (Cheu et al., 2001; Liang and Ling-xiang, 2006; Bachmann et al., 2013). The several input nodes are taken from the loop detector data (i.e. travel time estimated from loop detector data, loop detector density) while the rest are from the probe vehicle data (i.e. travel time estimated from probe vehicle data, probe vehicle sample size). Another way is the use of the state-space models presented in Nanthawichit et al. (2003); Chu et al. (2005); Bachmann et al. (2013). The state space estimation models perform two updated equations, state equation and measurement equation. Common approaches in solving this dynamic model are the Kalman filter and others variation filter model depends on linearity of equations. Other isolated approach are based on the theory of traffic flow as found in Hellinga, (2002), and Claudel et al., (2009).

For the second approach, common methods were linear combination, Bayesian theory, evidential theory, Fuzzy theory. Linear combination or weighted average model is one of the earliest model presented in the Advance Driver and Vehicle Navigation Concept project
(ADVANCE) by Tarko and Rouphail (1993). The project utilized the method of least squares for travel time estimation, namely the simple convex combination. An improvement of simple convex combination called Bar Shalom/Campo was proposed by El Faouzi (2004). Different way of weighted calculation presented by Bachmann et al. (2013); Choi (1999); Choi and Chung (2002). Baysian theory and evidential theory are the two model based on the probability theory to assign weights on the measurements. The application of baysian theory was found in Soriguera and Robusté (2011), while the evidential theory or dempster-Shafer theory was found in Faouzi et al. (2000); El Faouzi Nour-Eddin (2006); Qing-Jie Kong (2009); Kong et al. (2007); Qing-Jie Kong (2009); Kong et al. (2009a); Kong et al. (2009b). Other methods such as fuzzy integrals and fuzzy linear regression were found in Choi and Chung (2002) and Bachmann et al. (2013).

In summary, various fusion techniques have been investigated in travel time estimation based on fusion model both direct fusion (approach 1) and indirect fusion (approach 2). For the first approach, raw traffic data are combined directly by means of neural network model, state space model and traffic theory based models. The second approach’s methods consists of weighted linear combination, Bayesian theory, evidential theory, and fuzzy theory. Techniques in this group have the same goals to seek for appropriate weights of individual data and linearly combined the data. In addition, most models in the literature finally compared their results to single source approach because there is lack of common measurement. In this study, the second approach was used. Two common techniques based on fusion by averaging and fusion by median are compared in this study. The aims of this study is to present the empirical results of the data fusion model. Two case studies along I880N freeway, US and the urban arterial road in Suwon, South Korea is presented. Because of its simplicity both in modelling and application, the model can be considered as a benchmark for comparison to other models.

2. METHODOLOGY

2.1 Modeling Framework

In general, traffic state data refers to all traffic variables such as speed, flow, occupancy, and so forth. However these traffic variables in this study are first converted into speed. So from this point on only the term traffic speed data is used instead of traffic state data to be precise. This study combined traffic speed data in a centralized data fusion architecture. The procedure was based on the combination of all available data, after converted into the traffic speed map as presented in Figure 2. First, traffic speed map derived from each type of sensor is calculated based on the selected algorithm s as presented in the section 2.2. Individual speed map are adjusted according to predefined spatial resolution (link based) and temporal resolution (30 seconds or 5 minutes). In this study, fusion by averaging and fusion by median (Equation 1) are selected to combined the data. Finally, the travel time is estimated from the resulted speed map fusion based on instantaneous travel time estimation model.

\[
\text{Fusion by averaging: } I_f(x,t) = \frac{1}{S} \sum_{s=1}^{S} I_s(x,t) \\
\text{Fusion by median: } I_f(x,t) = \text{median}\{I_1(x,t), I_2(x,t), \ldots, I_S(x,t)\} \\
\text{Where:} \\
I_f(x,t) : \text{speed value at location } x, \text{ time } t \text{ based on mean and median rule} \\
I_s(x,t) : \text{speed value at location } x, \text{ time } t \text{ of sensor } s \\
S : \text{Number of sensors.}
\]
2.2 Selected Estimation Algorithms

The goal of this research is to investigate the application of data fusion technique for travel time estimation. However, it is essential to point out the procedure and algorithms for extracting and processing data of each type. Basic estimation techniques are applied because this study does not focus on improvement of individual estimation. The next sections cover the speed matrix calculation’s procedure for each type of data.

2.2.1 Speed Estimation from Single Loop Detector

Normally single loop detectors will only provide flow and occupancy data. Thus efforts on converting those data into speed data are required. Starting with the basic equation proposed by (Edie, Port of New York Authority. and International Symposium on the Theory of Road Traffic Flow, 1963), the average speed \( v \) of the traffic in a region (A) with the space \((x)\) and time \((t)\) is given by:

\[
\bar{v} = \frac{\sum_{i} x_i}{\sum_{i} t_i},
\]

Where:
- \( x_i \): distance traveled by vehicle \( i^{th} \) in the region A.
- \( t_i \): time spent by vehicle \( i^{th} \) in the region A.

This generalized equation is neither specific to time-mean speed nor space-mean speed. The application of space-mean speed specifically can be rewritten as:

\[
\bar{v}_s = \frac{\sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} t_i} = \frac{n \Delta x}{\Delta t} = \frac{1}{\bar{v}_i},
\]

Where:
- \( n \): number of vehicle in the region A.
- \( \Delta x \): infinitesimal length along space dimension.

Occupancy \((O)\) is the percentage of time in which detector remains on all over the aggregate period \((T)\). Its formula is presented by:
One of the biggest differences between single loop detectors and dual loop detectors is that single loop detectors are unable to measure the time between the first loop and the second loop, thus no data of spacing nor vehicle length can be received. This is the reason behind the inability to obtain speed using single loop detectors. Previous research on improving speed estimation on single loop detectors includes Dailey (1999); Wang and Nihan (2000); Zhanfeng Jia (2001); Coifman (2001); Hellinga (2002); Wang and Nihan (2003); Coifman et al. (2003). These studies attempted to eliminate the bias error, which come from the assumption of constant average effective vehicle length. Nevertheless, for the sake of simplicity, the conventional method that assumes the average effective vehicle length by \( g \) is used in the first case of this study. The individual speed of the vehicle \( (i) \) across each detector can be written as:

\[
  v_i = \frac{g}{t_{on}(i)}
\]

Taking into account the equation (3), (4), and (5) the space-mean speed is:

\[
  v_s = \frac{n \cdot g \cdot \frac{F \cdot g}{O \cdot T}}{\sum_{i=1}^{n} t_{on}(i)}
\]

Where:

\( F \) : traffic volumes,

### 2.2.2 Speed Estimation from probe vehicles Data

Data from GPS are more available over space compared to those obtained from fixed location sensors. As a vehicle travels along a corridor, data are sent in non-intrusive and systematic way, such as every fixed time interval. Once the data are received, they have to be matched with the map. This continuous process is called the map-matching technique. The problem of the map-matching technique is to find the path on a network, which is closest to the location obtained from raw GPS data. Various map-matching algorithms were found in the literature. However, in the first case of this study, because the selected corridor is a simple straight highway without complicated network such as flyovers or intersections, the closest distance was used. A critical review on map-matching techniques can be found in Hashemi and Karimi (2014).

Once outliers are removed through a map-matching algorithm, simple aggregation technique is applied to map the data into an individual link. Then the link travel speed can easily be computed by the harmonic mean of speed as following formulae:

\[
  v_s = \frac{n}{\sum_{i=1}^{n} \frac{1}{v_i}}
\]
2.2.3 Travel Time Estimation algorithm

Two specific terms need to be clarified before proceeding with any travel time calculation, are “Link” and “Corridor”. Link represents as a short section of the road in which the travel time can be measured (commonly it is established by a section between a pair of detectors), while corridor is the combination of multiple links. Because link travel time is commonly related to the average travel time over a fixed section; hence space mean-speed is used in this case. Finally, the average link travel time (LT) can be written as:

\[ LT_j = \frac{\sum_{i=1}^{n} L_{ji}}{n} = \frac{\sum_{i=1}^{n} v_i}{n} = \frac{L_j}{v_s}, \]

Where:
- \( LT_j \) : travel time of link \( j \)
- \( L_j \) : length of selected link \( j \)
- \( LT_{ji} \) : travel time of individual vehicle on link \( j \)

Consequently, the corridor travel time is calculated by the summation of all links at the same time interval (the simultaneous travel time estimation model):

\[ TT(t) = \sum_j LT_j(t) \]

3. CASE STUDY ON I880 FREEWAY, CALIFORNIA, USA

3.1 Data description

![Study Corridor Map](Image)

Figure 3. Study Corridor (Source: Google Map)
To investigate the accuracy of the fusion model compared to the estimation based on individual sensor, Mobile Century Data provided at [http://traffic.berkeley.edu](http://traffic.berkeley.edu) was used. The data consists of single loop detector data, Vehicle Trip Lines (VTLs) data, GPS data, and GPS log from mobile phone. More information about the data can be found in Herrera et al. (2010). The data were collected on February 8, 2008 from 10:00 until 18:00 along I880N corridor, CA starting from Decoto Road to Winton Avenue (Figure 3). In this case study only data from the first three types are used since the fourth type data has similar characteristic to the third type data. Each data type is briefly discussed in the following sections.

### 3.2 Loop Detector Data

Based on loop detector data, around 1.2 million records were detected by 20 loop detectors placing along studied corridor. Averaging number of detected vehicles over number of loop detectors provided the total volume (8 hours) of: 11,120; 12,047; 11,020; 10,551 and 3,367 vehicles from the leftmost lane to the rightmost lane (slow vehicle lane) respectively. Detectors are placed along study corridor providing an almost equally distributed distance around 0.8 km (0.5 miles). Information available consist of loop ID, longitude/latitude, flow and occupancy of individual lanes.

### 3.3 Vehicle Trip Lines (VTLs) Data

Vehicle Trip Lines technique is basically a sampling strategy for privacy protection (Hoh et al., 2006). For Mobile Century Data, VTLs performs as spatial triggers for phone to collect measurement and sends the update (Herrera et al., 2010). In VTLs sampling technique, mobile phones send their anonymous location, speed and direction of travel whenever they cross the predefined lines. More details about VTLs data can be found in (Herrera et al., 2010). The detected lines were predefined by 41 stations; equally distributed around 250 m along the study corridor. Raw data were provided in the form of speed of on individual vehicle and longitude/latitude of each trip line. Hence space-mean speed can directly be calculated by applying equation (7).

### 3.4 GPS Data

GPS data consist of 1,388 trajectories traveling randomly on the selected corridor. The instantaneous speeds of vehicles on each link are grouped into time intervals and are used to calculate the space-mean speed using equation (8). Individual trip file is provided in a CSV file format with the record of time stamp, longitude, latitude and speed.

### 3.5 Ground Truth Data

In addition, the ground truth travel time was provided by high-resolution video cameras through the license plate re-identification process. Travel time along the corridor vary by time of the day (departure time). The ground truth travel time data are plotted with departure time (with the resolution of 30 seconds and 5 minutes) in Figure 4 and Figure 5. The ground truth travel time data are used to validate the accuracy of estimation. Three types of deviation (0.5 $\sigma$, $\sigma$, 2$\sigma$) from the mean are also plotted to show the variation of the travel time along different departure time. It is clearly seen that departure between 15:00 -17:00 experienced high variability of travel time, when significant number of trucks and motorbikes presented on traffic stream (Figure 6).
Figure 4. Mean and standard deviation of ground truth travel time for 30-second interval
(Source: calculated by authors from Mobile Century Data at http://traffic.berkeley.edu)

Figure 5. Mean and standard deviation of ground truth data for 5-minute interval
(Source: calculated by authors from Mobile Century Data at http://traffic.berkeley.edu)

Figure 6. Aggregation of travel time into 5-minutes interval
(Source: calculated by authors from Mobile Century Data at http://traffic.berkeley.edu)
3.6 Analysis and Results

3.6.1 Sensor Errors

The preliminary investigation of sensor errors can be stated based on different sensor types. First of all, loop detector’s errors commonly come from malfunctions and the assumption of effective average vehicle length \( \bar{g} \). For VTLs data and GPS data, errors come from driving dynamics of tested vehicles. In addition to these general problems, the algorithms for computing speed are different. This can lead to different levels of errors. Lastly, VTLs data and GPS data, both come from sampling. Thus if the sample size is too small, it might not be sufficient to statistically represent the whole driving population.

3.6.2 Time-Space Alignment of Data

In order to perform fusion on speed measurements, alignment for both time and space of the speed map data are required. Among three types of data, loop detector data have the lowest resolution for both spatial dimension and temporal dimension. Hence alignments are applied to VTLs data and GPS data based on loop detectors’ finest time and space dimension. Since the spatial resolutions of VTLs and GPS are smaller than the loop detector thus it is possible to replace VTLs by location of loop detectors and similar procedure are made on GPS data.

According to the time dimension, loop detector data is reported every 30-second interval while VTLs data and GPS data are reported at smaller interval. In order to combine 3 types of sensor’s data, the 30-second interval is the minimum interval that is possible. However to investigate the effects of different levels of data aggregation, 5-minute interval is considered as well in this case study. After time dimension and space dimension were aligned, speed data of each type of sensors can be fused. The algorithms fuse only available data at the time. For example if GPS is not available at specific time interval, only loop detector data and VTLs data are considered to be fused at that time.

Table 1 presents percentage of missing data of all sensor types including the one after fused. The calculation was performed by taking number of missing cells over total cells (number of time intervals multiplies with number of links). It is important to note that the data are spatially and temporally aligned according to the distance between each pair of loop detectors (link) and the time interval of 30 seconds and 5 minutes. Loop detector showed lowest percentage of missing data followed by GPS and VTLs respectively. It is noted that when combining data from all types of sensors (Fusion) the percentage of missing data is significantly reduced to 0.2% and none for 30-second interval and 5-minute interval respectively.

<table>
<thead>
<tr>
<th>Sensor Type</th>
<th>Missing per 30-second interval</th>
<th>Missing per 5-minute interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Loop</td>
<td>0.8%</td>
<td>0.1%</td>
</tr>
<tr>
<td>VTL</td>
<td>38.2%</td>
<td>10.3%</td>
</tr>
<tr>
<td>GPS</td>
<td>26.6%</td>
<td>3.2%</td>
</tr>
<tr>
<td>Fusion</td>
<td>0.24%</td>
<td>0%</td>
</tr>
</tbody>
</table>

3.6.3 Results of Speed Map Estimation

Traffic speed from individual type of sensors are represented in the form of color map of time-
space dimension (Figure 7-10) using the calculation procedure described in section 2.2. Speed maps derived from three data sources show similar patterns. It is clearly seen the missing values exist in GPS data and VTLs data as blank cells in the color map. All maps are represented in the same spatial dimension (based on location of loop detectors) and same temporal dimension of 30 seconds. The fusion speed map in Figure 10 shows complete information by applying median operation with exclusion of cells containing missing data. Similar procedure done for 5-minute interval. Given various speed maps that naturally are obtaining by different ways of calculating, which one is the most accurate and what is the level of accuracy once all are combined? Since there is no information about real speed map, the accuracy of each speed map could not directly be accessed. For this reason, the travel time from video image detection (ground truth data) presented in section 3.5 was used.

3.6.4 Validation of Travel Time Estimation

Once the speed maps are created, the travel time can be calculated by using equation (8-9). Figure 11 is plotted of the estimation of travel time based fusion by median, fusion by averaging, loop detector data alone with the ground truth travel time for 30-second interval. Similarly, Figure 12 is plotted of the estimation of travel time based fusion by median, fusion by averaging, and based on type of data (loop detector data, GPS data, VTLs data) with the ground truth travel time for 5-minute interval. Figure 13 shows the estimation errors of individual type of data compared to the ground true data. It can be seen clearly that travel time is over estimated in the morning period. This is because of an accident occurred during the day of experiment, which activated a non-recurrent congestion at that location (Herrera et al., 2010).

To access the accuracy of the estimation models, a number of common measurements of error are used in this case study including Mean absolute error (MAE), Mean absolute percentage error (MAPE), Mean absolute scale error (MASE) and Mean signed difference (MSD). MAE is the simplest measurement of model’s accuracy. It is the mean absolute value of the different between estimated value ($\hat{T}$) and the actual value ($T$). One of disadvantages of using MAE is that the size of relative error is not obvious. To deal with this problem, MAPE is used to measure...
the percentage of error, which allows us to understand the size of error obviously. MAPE, however, has disadvantage of being infinite when \( T_i = 0 \). In addition to MAPE, MASE scales the error compared to the average one-step naïve forecasting. MASE is the only available measurement of errors method, which can be used in all circumstance including when \( T_i = 0 \).

Lastly, MSD is added to the bias of estimation. Giving the error by the difference between ground truth data and estimated models (\( e_i = T_i - \hat{T}_i \)), these measurements can be computed as:

\[
\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} |e_i| 
\]

\[
\text{MAPE} = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{|e_i|}{T_i}\right) \times 100 
\]

\[
\text{MASE} = \frac{1}{n} \sum_{i=1}^{n} e_i \left(\frac{1}{n-1} \sum_{j=2}^{n} |T_j - T_{j-1}|\right) 
\]

\[
\text{MSD} = \sum_{i=1}^{n} \frac{e_i^2}{n} 
\]

There is slightly bigger error of the estimation based on median compared to the average at 30-second interval due to high-fluctuated data (Figure 11). However, in general, the estimation based on fusion by median performs better compared to fusion by averaging. The accuracy of estimation model is linearly increased according to the value of time interval from 30 seconds to 5 minutes. The reason behind is the ability to eliminate outliers (most biased data among all types) when more than two data sources available. Table 2 shows the high error of estimation based on GPS data in 5-minute intervals following by the estimation based on loop detector data and VTLs data. Improvement of accuracy is found when 5-minute interval is used in data fusion model.

![Figure 11. Travel time estimation versus ground truth travel time data (30-second interval)](image_url)

Comparing its measurement of errors between the 30-second and 5-minute intervals shows that the accuracy of travel time estimation by data fusion based on median is much improved compared to the mean. In case of using loop detector data alone, increase of aggregation interval from 30 seconds to 5 minutes does not improve the accuracy of travel time estimation at all
(Table 2). With reference to MSD, travel time derived from VTLs and loop detector data is both underestimated while it tends to overestimate in case of GPS data.

<table>
<thead>
<tr>
<th>Measurement of errors</th>
<th>GPS</th>
<th>VTL</th>
<th>Loop</th>
<th>Fusion (median)</th>
<th>Fusion (mean)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>30-second interval</strong></td>
<td>MAE(s)</td>
<td>N/A</td>
<td>N/A</td>
<td>70.91</td>
<td>60.3</td>
</tr>
<tr>
<td></td>
<td>MAPE(%)</td>
<td>N/A</td>
<td>N/A</td>
<td>8.68</td>
<td>7.7</td>
</tr>
<tr>
<td></td>
<td>MASE(%)</td>
<td>N/A</td>
<td>N/A</td>
<td>2.62</td>
<td>2.2</td>
</tr>
<tr>
<td></td>
<td>MSD(s)</td>
<td>N/A</td>
<td>N/A</td>
<td>-17.7</td>
<td>20.9</td>
</tr>
<tr>
<td><strong>5-minute interval</strong></td>
<td>MAE(s)</td>
<td>122.4</td>
<td>53.8</td>
<td>80</td>
<td>52.5</td>
</tr>
<tr>
<td></td>
<td>MAPE(%)</td>
<td>14.7</td>
<td>6.8</td>
<td>9.7</td>
<td>6.8</td>
</tr>
<tr>
<td></td>
<td>MASE(%)</td>
<td>4.4</td>
<td>1.9</td>
<td>2.9</td>
<td>1.9</td>
</tr>
<tr>
<td></td>
<td>MSD(s)</td>
<td>111.6</td>
<td>-23.6</td>
<td>-34.3</td>
<td>-4</td>
</tr>
</tbody>
</table>

Note: - Estimated travel times are done based on the interpolation of missing values of speed.
- No estimation of travel time for GPS and VTLs (30 second interval) since the percentage of missing values is too high to be interpolated.
4 CASE STUDY ON ARTERIAL ROAD IN SUWON, KOREA

4.1 Data description

The problem occurs in applying the median when there is only two types of data available, which are commonly from loop detector data and GPS data. The case study bases on the available data from the study of Choi and Chung, (2002) is investigated. In this case, only two data set are available (Loop detectors data and GPS data) along with the ground truth data. The data are collected for the period of 2 hours from 11:00 to 13:00 on November 28, 1998 including the historical data available on November 21, 1998. The resulting of travel time from individual sensor is used. However, to apply median, in this case study the historical data is considered as the third party data. Table 3-4 present the data from Choi and Chung (2002) before fuzzy regression technique was used to combine the data in their study.

<table>
<thead>
<tr>
<th>Table 3. Travel time data for link 1 to link 2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Time</strong></td>
</tr>
<tr>
<td>----------</td>
</tr>
<tr>
<td>00-04</td>
</tr>
<tr>
<td>05-09</td>
</tr>
<tr>
<td>10-14</td>
</tr>
<tr>
<td>15-19</td>
</tr>
<tr>
<td>20-24</td>
</tr>
<tr>
<td>25-29</td>
</tr>
<tr>
<td>30-34</td>
</tr>
<tr>
<td>35-39</td>
</tr>
<tr>
<td>40-44</td>
</tr>
<tr>
<td>45-49</td>
</tr>
<tr>
<td>50-54</td>
</tr>
<tr>
<td>55-59</td>
</tr>
<tr>
<td>00-04</td>
</tr>
</tbody>
</table>

Note: H. data means historical travel time data.

<table>
<thead>
<tr>
<th>Table 4. Travel time data for link 3 to link 4</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Time</strong></td>
</tr>
<tr>
<td>----------</td>
</tr>
<tr>
<td>00-04</td>
</tr>
<tr>
<td>05-09</td>
</tr>
<tr>
<td>10-14</td>
</tr>
<tr>
<td>15-19</td>
</tr>
</tbody>
</table>

2 Data are directly extract from Choi and Chung, (2002).
3 GPS data is derived from the arithmetic mean (loop detector data & GPS data) and the loop detector data from Choi and Chung, (2002).
The model combines the data based on average and mean of time series data. Similar to the first models. Step-by-step of calculating procedure is presented. It is proved by measurements of the way in which the speed map is presented allowed the fusion by averaging and fusion by case, median provides better result in term of MAPE for all four links. Different data types in order to reduce bias from outliers. This study applies simple approach by using median, and it was shown that the results are reasonable results. Finally, model might be different.

\[
\begin{array}{cccccccc}
\text{Time} & \text{Ground Truth} & \text{H. data} & \text{Loop detector} & \text{GPS} & \text{Mean} & \text{Median} & \text{Ground Truth} & \text{H. data} & \text{Loop detector} & \text{GPS} & \text{Mean} & \text{Median} \\
\hline
25-29 & 16.38 & 11.18 & 13.9 & - & 12.54 & 12.54 & 16.43 & 26.23 & 15.08 & 15.08 & 18.80 & 15.08 \\
35-39 & 23.34 & 22.64 & 17.79 & 66.07 & 35.5 & 22.64 & 26.04 & 23.44 & 14.5 & 9.84 & 15.93 & 14.5 \\
00-04 & 14.47 & 68.64 & 14.96 & - & 13.76 & 13.76 & 12.5 & 15.1 & 15.15 & - & 15.12 & 15.12 \\
10-14 & 30.18 & 54.34 & 33.23 & - & 47.95 & 47.95 & 31.23 & 22.63 & 28.91 & - & 25.77 & 25.77 \\
15-19 & 56.13 & 23.93 & 33.34 & 126.44 & 78.04 & 74.33 & 61.05 & 61.95 & 34.89 & 45.07 & 47.30 & 45.07 \\
20-24 & 35.48 & 52.26 & 27.5 & 92.16 & 53.51 & 40.88 & 42.04 & 51.54 & 32.87 & 23.71 & 36.04 & 32.87 \\
25-29 & 41.46 & 29.15 & 31.52 & 103.98 & 64.29 & 57.36 & 59.71 & 66.01 & 27.35 & 49.43 & 47.60 & 49.43 \\
30-34 & 46.51 & 57.25 & 39.06 & 127.46 & 79.81 & 72.91 & 40.36 & 30.46 & 42.85 & 100.69 & 58 & 42.85 \\
40-44 & 39.35 & 73.21 & 33.45 & 76.71 & 54.50 & 53.33 & 61.29 & 55.09 & 32.7 & 131.86 & 73.22 & 55.09 \\
45-49 & 43.71 & 26.24 & 39.45 & 211.27 & 110 & 79.51 & 61.28 & 64.18 & 34.59 & 123.81 & 74.19 & 64.18 \\
50-54 & 37.38 & 19.88 & 35.82 & 42.5 & 42.17 & 42.5 & 43.83 & 43.63 & 31.4 & 140.06 & 71.70 & 43.63 \\
55-59 & 26.32 & 37.4 & 15.93 & 84.03 & 39.96 & 19.92 & 27.79 & 25.59 & 23.98 & 25.5 & 25.02 & 25.5 \\
\end{array}
\]

4.2 Analysis Result

Since the data are in form of time series data (1D) not in form of matrix (2D), only the simple median and average are considered alone the different departure time. Table 5 show the result of fused data through average and median. It is revealed that the median performs better for all studied links. In overall, fusion by median performs around 10% better than fusion by averaging.

<table>
<thead>
<tr>
<th>Link 3</th>
<th>Link 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>Median</td>
</tr>
<tr>
<td>Mean</td>
<td>Median</td>
</tr>
</tbody>
</table>

Table 5. Level of errors compared to the ground truth data

<table>
<thead>
<tr>
<th>Link 1</th>
<th>Link 2</th>
<th>Link 3</th>
<th>Link 4</th>
<th>All links (average)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average</td>
<td>37%</td>
<td>17%</td>
<td>40%</td>
<td>26%</td>
</tr>
<tr>
<td>Median</td>
<td>30%</td>
<td>15%</td>
<td>24%</td>
<td>15%</td>
</tr>
</tbody>
</table>

5. CONCLUSIONS

This study presents case studies as an empirical results of travel time estimation using the data fusion model in the traffic research field. Two case studies are presents. In the first case study, the way in which the speed map is presented allowed the fusion by averaging and fusion by median to apply with two dimension data. Since it is difficult to collect ground truth velocity information, ground truth travel time is used to validate and measure the accuracy of the fusion models. Step-by-step of calculating procedure is presented. It is proved by measurements of accuracy (MAE, MAPE, MASE, and MSD) that fusion by median outperforms fusion by averaging and the model does provide better estimation without knowing its statistical characteristics. In the second case study, the data are used based on ready calculated travel time. The model combines the data based on average and mean of time series data. Similar to the first case, median provides better result in term of MAPE for all four links.

Many researchers have sought for a better estimation of travel time by evaluating the weight of different data types in order to reduce bias from outliers. This study applies simple approach by using median, and it was shown that the results are reasonable results. Finally, model might be
considered as a benchmark in comparison more advanced fusion approach as found in the
literature. There are some limitations in this study, which could be further improved in future
works. The ground truth data may contain some outliers (Figure 6), which might come from the
mix of vehicle type and the error of sensor detection during congestion period. Proper treatment
of these outliers would further improve the accuracy of validation process. Moreover, the travel
time estimation algorithms used in this study are based on instantaneous speed based model;
which may subject to larger errors on long corridors. Better estimation techniques should be
investigated in future work such as time slide model (Li et al., 2006), dynamic time slice model
(Cortés et al., 2002), linear model (van Lint and van der Zijpp, 2003) and trajectory
reconstruction model (Ni and Wang, 2008).

ACKNOWLEDGEMENTS
This research paper is financially supported by the Sirindhorn International Institute of
Technology, Thammasat University. However, the authors are responsible for the accuracy of
the content provided and does not, in any way, represent the ideas of the University. Moreover,
the author would like to acknowledge the data from the Mobile Century field experiment and
from the study of Choi and Chung, (2002).

REFERENCES

multi-sensor data fusion techniques for freeway traffic speed estimation using
doi:10.1016/j.trc.2012.07.003

from Heterogeneous Sources for First Response Deployment. J. Comput. Civ. Eng. 28,
04014018. doi:10.1061/(ASCE)CP.1943-5487.0000379

Proceedings, pp. 573–578. doi:10.1109/ITSC.2001.948723

Choi, K., 1999. Data fusion methodology for link travel time estimation for advanced traveler


Chu, L., Oh, S., Recker, W., 2005. Adaptive Kalman filter based freeway travel time estimation,
in: 84th TRB Annual Meeting, Washington DC.

travel times using probe and fixed data, in: 88th TRB Annual Meeting Compendium of
Papers.

Policy Pract. 35, 863–880. doi:10.1016/S0965-8564(00)00028-8

and Estimation 11, 211–222. doi:10.1016/S0968-090X(03)00025-1


Dailey, D.J., 1999. A statistical algorithm for estimating speed from single loop volume and


Zhanfeng Jia, C.C., 2001. The PeMS algorithms for accurate, real-time estimates of g-factors and speeds from single-loop detectors 536 – 541. doi:10.1109/ITSC.2001.948715