Abstract: This paper presents models to predict freeway link travel time in multiple periods. The relationships between traffic measures and average link travel time from simulation were explored using the Generalized Additive Model (GAM), a nonparametric regression technique. Parametric models to predict freeway link speed were then developed based on the relationships discovered by the nonparametric model. The models were developed on a set of simulation data and tested on another set of simulation data. The root mean square errors (RMSE) of the model developed to predict link travel time at current time-step were 88.63, 6.88, 2.38 seconds for morning peak hours, off-peak hours, and afternoon hours, respectively. The RMSE of the model to predict link travel time at 5, 10, 15 minutes forwards were 91.20, 101.86, 122.08 seconds, respectively.

Key Words: Average Link Travel Time, Intelligent Transportation Systems, Travel Time Prediction

1. INTRODUCTION

Advanced Traveler Information System (ATIS) and Advanced Traffic Management System (ATMS) are two major components of Intelligent Transportation Systems (ITS). The ATIS consists of dynamic route guidance system that advises the best route to drivers using the processed information gathered from various sources such as fixed sensors and probe vehicles. The information required for the ATIS includes traffic conditions such as link speed or link travel time. The link travel time at current time interval would benefit drivers who are entering the link. The link travel time at several minutes forwards would benefits drivers who are at several links away and traveling to the subject link. The ATMS also need to know current traffic conditions for traffic management. A successful implementation of the ITS relies on a reasonable accuracy of link speed or link travel time estimation and prediction. Hence, models to estimate and predict link travel time accurately are necessary for acquisition of reliable traffic conditions.

2. OBJECTIVES

The objectives of this research were 1) to develop models to estimate and predict freeway link travel time by using data from fixed sensors and 2) to compare the performances of the proposed models with another model such as neural network based models.
3. LITERATURE REVIEW

Sources of traffic measures can be classified into two types according to its characteristics; fixed sensors and mobile sensors. Fixed sensors such as loop detectors, microwave sensors and Autoscope, can represent temporal variation well but lack of representing spatial variation. Mobile sensors such as probe vehicles equipped with Global Positioning System (GPS), can represent spatial variation well but lack of representing temporal variation. The literatures involve in link travel time estimation using data from fixed sensors are summarized here.

Gipps (Gipps, 1977) developed a model using regression technique to estimate link travel time. Nam (Nam and Drew, 1996) developed a stochastic queuing theory based model to estimate travel time on freeways using flow measurements. Dailey (Dailey, 1993) used Cross-Correlation technique with loop data to measure the propagation time of traffic. The time delay was used to estimate the mean speed. The results showed that the estimates had good agreement with empirical data measured at 30-seconds interval. Pushkar (Pushkar, Hall et al., 1994) developed model based on catastrophe theory for estimating average speeds for single loop detectors data. The results showed that the catastrophe theory estimates were better than those made assuming a constant vehicle length. Salem (Salem and Van Grol, 1997) presented the algorithm called speed-based section level travel time (SLTT) estimator. The algorithm was based on the assumption that vehicles drive half of the section with the time-mean speed detected at the downstream end of the section and half of the section with the time-mean speed at the upstream end of the section. Another algorithm called SLTT estimator was based on mass-balance. The travel time was estimated by adding free flow travel time to delay. The delay was computed as the time needed for the extra vehicle to leave the section with an effective capacity equal to the average outflow of three time periods. Park (Park and Rilett, 1998) developed modular neural networks based models for link travel time forecasting. The performances of the model were compared with a Kalman filtering model, an exponential smoothing model, a historical profile, and a real-time profile. It was found that the modular ANN gave the best overall results. D’Angelo (D’ Angelo et al., 1999) proposed a non-linear time series model for corridor travel predictions based on travel time data collected from an 18 km freeway section in Orlando, Florida, USA. The single-variable time series model showed better performance to multivariable model. Cortes (Cortes, Lavanya et al., 2001) developed a technique to estimate travel time using loop detector data from several locations both inside and outside the link considered. The model with spot speed from loop detector as a variable provided high accuracy. Ishak (Ishak and Al-deek, 2002) developed a nonlinear time series traffic prediction model. The model was developed based on real-time data collected from the 62.5 km corridor of Interstate-4 in Orlando, Florida, USA. The results showed that travel time increased drastically under congested conditions due to high fluctuation and instability in speed observations. Wouters (Wouters et al., 2005) developed a statistical algorithm to compute the average travel time for long term travel time prediction. A database of 2 years travel time data gathered from inductive loop detector was used. The prediction travel time was compared with independent loop data and showed a good match. Emam (Emam and Al-Deek, 2006) proposed a best-fit statistical distribution (lognormal) to estimate and predict travel time reliability of freeway corridors. It showed higher sensitivity to level of congestion and bottlenecks when compared with existing Florida and buffer time methods. Van Lint (Van Lint et al., 2005, Van Lint, 2006) proposed a freeway link travel time prediction framework based on state-space neural network (SSNN). The framework showed high accuracy and reliability of travel time prediction.
4. METHODOLOGY

Generalized additive model (GAM) technique is applied to explore relationship between average link speed and traffic measures. A generalized additive model (Hastie and Tibshirani, 1990) consists of a random component, an additive component, and a link function relating the two components. The response \( y \), the random component, is assumed to have density in the exponential family,

\[
f_r(y, \theta, \phi) = \exp \left\{ \frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi) \right\}
\]

(1)

Where, \( \theta \) is the natural parameter and \( \phi \) is the scale parameter. The mean of the response variable \( \mu \) is related to the set of independent variables \( x_1, x_2, x_3, \ldots, x_p \) by a link function \( g \).

\[
g(\mu) = \eta = s_1(x_1) + \ldots + s_p(x_p)
\]

(2)

Where, \( s_1(.) \ldots s_p(.) \) are smooth functions defining the additive component, and the relationship between \( \mu \) and \( \eta \) is defined by \( g(\mu) = \eta \).

The generalized additive model is a nonparametric regression and smoothing technique that relaxes the assumptions of linearity and uncoovers the structure in the relationship between the independent variables and the dependent variables. The generalized additive models combine an additive assumption that enables relatively many parametric relationships to be explored simultaneously with the distributional flexibility of generalized linear models. Other nonparametric regressions do not perform well when the number of independent variables is large, leading to variances of the estimates to be unacceptably large or is often referred to as the ‘curse of dimensionality’. The generalized additive model overcomes these problems since each of the individual terms are estimated using a univariate smoother and the estimates of the terms explain how the response changes with corresponding independent variables.

Generalized additive model is appropriate where the relationship between the variables is expected to be of a complex form, and may not easily be modeled as a linear or non-linear model. The appropriate functional form may be suggested or the structure of relationship between independent variables and dependent variable may be explored more thoroughly using generalized additive models. While the nonparametric form for the functions \( s(.) \) makes the model more flexible, the additivity is retained and allows interpretation of the model as in generalized linear models. This modeling framework not only allows classification but also provides an understanding of the roles of the prognostic factors \( (x_i) \).

Each smoother such as a cubic spline, \( s(.) \), is estimated based on minimizing the following penalized residual sum of squares function:

\[
\sum_{i=1}^{n} \left\{ g(y_i) - s(x_i) \right\}^2 + \lambda \int_{a}^{b} \left\{ s'(t) \right\}^2 dt
\]

(3)

Where; \( \lambda \) is smoothing parameter and \( a \leq x_i \leq \ldots \leq x_n \leq b \).

For the fitted generalized additive model with cubic spline smoothers, the following equation is minimized for \( p \) smoother:

\[
\sum_{i=1}^{n} \left\{ g(y_i) - \sum_{j=1}^{p} s_j(x_i) \right\}^2 + \sum_{j=1}^{p} \lambda_j \int_{a}^{b} \left\{ s_j'(t) \right\}^2 dt
\]

(4)
The generalized additive model is estimated based on two iterative algorithms: Local Scoring Algorithm (LSA) and Backfitting Algorithm (BFA). An outer loop of the Local Scoring Algorithm and an inner loop of the Backfitting Algorithm are used until convergence. For each iteration of the Backfitting algorithm, splines are estimated. During each iteration of the Local Scoring Algorithm, an adjusted dependent variable and a set of weights are estimated to apply an iteratively reweighted least squares. The algorithms are as follows:

### 4.1 Local Scoring Algorithm

1. Initialization: \( s_0 = g(E(y)) \), \( s_0^i(.) = s_0^p(.) = ... = s_0^p(.) = 0 \), \( m = 0 \).
2. Iterations: \( m = m + 1 \)
   From the dependent variable, predictor, and mean based on the previous iteration, a new adjusted dependent variable may be estimated as:
   \[
   z_i = \eta_i^0 + \left( y_i - \mu_i^0 \right) \left( \frac{\partial \eta_i}{\partial \mu_i} \right)_0
   \]
   \[\text{(5)}\]
   The weights \( w_i \) are defined by,
   \[
   w_i^{-1} = \left( \frac{\partial \eta_i}{\partial \mu_i} \right)^2 \nu_i^0
   \]
   \[\text{(6)}\]
   Where, \( \nu_i^0 \) is the variance of response \( y \) at \( \mu_i^0 \).
3. Convergence: The iterations are continued till the deviance for the model fails to decrease or satisfies the convergence criterion. Using the Backfitting Algorithm, an additive model to \( z \) is fitted with weight \( w_i \) to obtain estimated functions \( s_j(.) \).

### 4.2 Backfitting Algorithm

The Backfitting Algorithm is applied to estimate the smoothing functions \( s_1(.), ..., s_p(.) \) in the additive model in Eq. (2) The \( j^{th} \) set of partial residuals may be estimated as,

\[
R_j = Y - s_0 - \sum_{k=1}^{p} s_k(X_k)
\]

and \( E(R_j|X_j) = s_j(X_j) \). Based on estimates \( \{s_i(.), i \neq j\} \), smoothing functions \( s_j(.) \) may be estimated. The iterative procedure is as follows:

1. Initialization : \( s_0 = E(Y); s_0^i(.) = ... = s_0^p(.) = 0 \), \( m = 0 \)
2. Iterations: \( m = m + 1 \), for \( j = 1 \) to \( p \)
   \[
   R_j = Z - s_0 - \sum_{k=1}^{i-1} s_k(x_i) - \sum_{k=i+1}^{p} s_k(x_k)
   \]
   \[\text{(8)}\]
   \[
   s_j^m = E(R_j|X_j)
   \]
   \[\text{(9)}\]
   An iteratively re-weighted least squares is obtained by smoothing weighted \( R_j \) on \( x_j \).
   \[
   s_j(x_j) = \frac{E[w;\{R_j\}|x_j]}{E(w|x_j)}
   \]
   \[\text{(10)}\]
   Where; the weights are estimated based on Eq. (6)
3. Convergence: The iterations are continued till either the residual sum of squares
\[ Y - s_p - \sum_{p} s_p(X_p) \]
fails to decrease, or satisfies the convergence criterion. \( z_i \) is re-estimated based on the Local Scoring Algorithm.

An adjusted dependent variable based on an iteratively weighted least squares. The algorithm regresses \( z_i \) on \( x \) with weight \( w_i \) to obtain revised estimates. The new \( \mu, \eta, z_i \) estimated and the process repeated till the change in deviance,
\[ D(y; \hat{\mu}) = 2(l(\mu_{max}; y)) - l(\hat{\mu}', y) \]  
(11)
is sufficiently small, where, \( \mu_{max} \) is the parameter value that maximizes likelihood \( l(\mu, y) \) over all \( \mu \), the saturated model. Here, \( l(\mu, y) \) is the log likelihood
\[ \sum_{i=1}^{n} l(\mu_i, y_i) = \sum_{i=1}^{n} \log P(y_i, \mu_i, \phi) \]  
(12)
Where; \( P = \) probability of \( Y|x \).

The degrees of freedom for a generalized additive model can be expressed as,
\[ df = \text{trace}(S(\lambda_j))-1 \]  
(13)
Where; \( S(\lambda_j) \) is a smoother operator. Thus \( df \) is the sum of the eigenvalues of \( S(\lambda_j) \).

5. DATA COLLECTION AND REDUCTION

The test link was a section of south bound direction of Srirath Freeway in Bangkok, Thailand, between Ratchadapisak on-ramp and Klong Papa on-ramp. It is a 3-lane freeway section with total distance of 7.0 kilometers and consisting of four on-ramps and an off-ramp (Figure 1). Traffic measures were collected using camcorders at four locations on mainline and all on-ramps and an off-ramp during morning peak hours from 6.30 to 9.30 AM, off-peak hours from 11.00 AM to 2.00 PM, and afternoon hours from 3.00 to 6.00 PM on November 17, 2004 and November 22, 2004. Traffic volume and spot speed at 5-minute interval were then extracted using a time recording program and the Autoscope.

For the November 17, 2004 data set, the volume varied from 3,720 to 8,724 vph, 2,040 to 5,592 vph, and 2,640 to 6,768 vph for morning peak hours, off-peak, and afternoon hours. The speed varied from 35 to 85 km/h, 81 to 110 km/h, and 62 to 109 km/h for morning peak hours, off-peak, and afternoon hours.

For the November 22, 2004 data set, the volume varied from 3,948 to 8,640 vph, 2,508 to 5,328 vph, and 2,400 to 6,132 vph for morning peak hours, off-peak, and afternoon hours. The speed varied from 31 to 86 km/h, 89 to 110 km/h, and 85 to 113 km/h for morning peak hours, off-peak, and afternoon hours. It may note that although the test link was a 3-lane freeway, a wide shoulder was utilized as a fourth lane during morning peak hours starting from 7.15 AM. Comparing ranges of data on both dates, both speed and volume data were relatively close.
6. MICROSIMULATION AND CALIBRATION

The test link was simulated using CORSIM (ITT Industries Inc., 2001), a microsimulation software. The traffic measures from simulation including 5-minute interval volume and spot speed, were then calibrated with the collected field data for all four locations that data collected.

The differences between volume from simulation and field varied from 0 to 6 percent, 0 to 10 percent, 0 to 10 percent, and 1 to 9 percent at Rachadapisek, Bangsue, Phahonyothin, and Klong Papa locations, respectively.

The differences of each location were less than 10 percent. These differences lied within the upper limit recommended in the report of Minnesota Department of Transportation (Minnesota Department of Transportation, 2004).

The differences of 5-minute interval spot speed between simulation and field data ranged from 0 to 20 percent, 0 to 19 percent, 0 to 20 percent, and 0 to 17 percent, at Rachadapisek, Bangsue, Phahonyothin, and Klong Papa locations, respectively. The differences were less than 20 percent which lied within an acceptable value suggested by Minnesota Department of Transportation (Minnesota Department of Transportation, 2004). To develop link travel time estimation and prediction models, the first set of simulation data was used. The second set of simulation data was then used for the test of model performances. The model development is presented next.
7. MODEL DEVELOPMENT

As mentioned earlier, the models were developed on the first simulation data set and tested on the second data set. Traffic measures at Ratcha da location were considered as data at the upstream location. The traffic measures at Klong Pa-Pa location were considered as data at the downstream location. The independent variables considered included volume, speed, and occupancy at upstream location and downstream location at current time interval (t), one time-step backward (t-5), two time-step backward (t-10), and three time-step backward (t-15). The dependent variable was an average link speed which was later converted to average link travel time. The relationships between dependent variable and independent variables were explored using Generalized Additive Model (GAM), a non-parametric model. The Gaussian distribution was used as a link function. Degree of freedom was set to three to avoid over fitted. The model was developed using the forward stepwise technique. The parametric models as suggested by partial prediction of non-parametric GAM were then developed. Details are described below.

**Model to estimate link travel time at current time-step for morning peak hours**

For morning peak hours, the significant independent variables for the model to estimate average speed at current time interval (t) were upstream volume at current time step (UV_t), upstream occupancy at current time interval (UOCC_t) and at a time-step backward (UOCC_t-5), downstream volume at a time-step backward (DV_t-5) (Table 1). Table 2 shows statistics values of significant variables of the model to estimate link travel time for morning peak hours.

Figure 2a shows the partial prediction of the significant independent variables by non parametric estimates. The dashed lines show 95 percent confident interval. By inspection of partial predictions, the relationship between average link speed and upstream volume at current time-step was similar to polynomial of degree three or two polynomials of degree two. The relationships between average link speed and upstream occupancy at current time-step and a time-step backward were similar to polynomial of degree two. Average link speed decreased as downstream volume at a time-step backward increased up to about 635 (vehicles per 5 minutes). Then the average increased when downstream volume increased.

From the partial prediction estimated by GAM-non parametric, the parametric model was developed. The upstream volume at current time-step, downstream occupancy at current time-step and at a time-step backward, and downstream volume at a time-step backward could be represented by polynomial functions (Figure 2b). The parametric estimates of a model to estimate average link speed at current time-step could be expressed as follows:

\[
\text{ALS}_0 = 1649.922 + 0.0012 (UV_t \leq 470)^2 - 1.0655 (UV_t \leq 470) + 0.0006(UV_t > 470)^2 - 0.7260 (UV_t > 470) - 0.0348(UOCC_t)^2 + 1.5285 (UOCC_t -0.1098 (UOCC_t-5)^2 + 3.6623 (UOCC_t-5) + 0.0035 (DV_t-5)^2 - 4.4862(DV_t-5) \]

Where;
- \(\text{ALS}_0\) = Average Link Speed at current time-step (km/hr)
- \((UV_t \leq 470)\) = Upstream volume at current time-step when it is less than or equal to 470 vehicles per 5 minutes.
- \((UV_t > 470)\) = Upstream volume at current time-step when it is greater than 470 vehicles per 5 minutes.
- \(UOCC_t\) = Upstream Occupancy at current time-step (percent)
UOCC\textsubscript{t-5} = Upstream Occupancy at a time-step backward (percent)

DV\textsubscript{t-5} = Downstream Volume at a time-step backward (Vehicle/5 minutes)

Table 1 Significant variables for models to predict link speed

<table>
<thead>
<tr>
<th>Variables</th>
<th>AM Peak Hours</th>
<th>Off-Peak Hours</th>
<th>PM Peak Hours</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>t</td>
<td>t+5</td>
<td>t+10</td>
</tr>
<tr>
<td>U/S Volume\textsubscript{t}</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>U/S Speed\textsubscript{t}</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>U/S Speed\textsubscript{t-5}</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>U/S Occupancy\textsubscript{t}</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>U/S Occupancy\textsubscript{t-5}</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>U/S Occupancy\textsubscript{t-10}</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>U/S Occupancy\textsubscript{t-15}</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>D/S Volume\textsubscript{t-5}</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

Note: \( t \) = current time interval., \( t+5, t+10, t+15 \) = 5, 10, 15 minutes forwards interval, respectively.

U/S = Upstream, D/S = Downstream

Table 2 Statistics values of model developed for morning peak hours

<table>
<thead>
<tr>
<th>Variables</th>
<th>Degree of Freedom</th>
<th>F-Value</th>
<th>P-Value</th>
<th>Null Deviance</th>
<th>Residual Deviance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UV\textsubscript{t}</td>
<td>3</td>
<td>5.35496</td>
<td>0.01572156</td>
<td>1174.489</td>
<td>168.3802</td>
</tr>
<tr>
<td>UOCC\textsubscript{t}</td>
<td>3</td>
<td>4.98832</td>
<td>0.01976648</td>
<td>1174.489</td>
<td>168.3802</td>
</tr>
<tr>
<td>UOCC\textsubscript{t-5}</td>
<td>3</td>
<td>10.94042</td>
<td>0.00088348</td>
<td>1174.489</td>
<td>168.3802</td>
</tr>
<tr>
<td>DV\textsubscript{t-5}</td>
<td>3</td>
<td>4.11770</td>
<td>0.03479037</td>
<td>1174.489</td>
<td>168.3802</td>
</tr>
</tbody>
</table>

Similar procedures were conducted for link speed (travel time) prediction models for morning peak hours at 5, 10, 15 minutes forward (t+5, t+10,t+15). The models to estimate link travel time at current time-step for off-peak hours and afternoon hours were also developed. Variations of link travel time during off-peak and afternoon hours were low due to light traffic. Therefore, models to predict link travel time for those periods were not developed, but only for morning period were developed. Consequently, a total of six different models were obtained.

Significant variables for link travel time prediction at 5, 10, and 15 minutes forward were the same as the model to estimate link travel time at current time step. Significant variables for link travel time estimation during off-peak hours were upstream speed at current time step and at a time-step backward. For afternoon hours, significant variables were upstream volume at current time-step, upstream occupancy at two and three time-step backward (Table 1). Details of the models may be found in the report by Thanasupsin (Thanasupsin, 2006). The parametric estimate of the models could be expressed as equation 15 to equation 19.
Predicting average link speed at 5 minutes forward.

\[
ALS_5 = 1711.389 + 0.0016 (UV_t \leq 470)^2 - 1.3916 (UV_t \leq 470) + 0.0008 (UV_t > 470)^2 - 0.9713(UV_{t-5})^2 + 1.8847 (UOCC_t) - 0.1188 (UOCC_{t-5})^2 + 4.0002 (DV_{t-5}) + 0.0035 (DV_{t-5})^2 - 4.4844 (DV_{t-5})
\]  

Predicting average link speed at 10 minutes forward.

\[
ALS_{10} = 1938.835 + 0.0016 (UV_t \leq 470)^2 - 1.3314 (UV_t \leq 470) + 0.0008 (UV_t > 470)^2 - 0.9713(UV_{t-5})^2 + 1.8847 (UOCC_t) - 0.1188 (UOCC_{t-5})^2 + 4.0002 (DV_{t-5}) + 0.0035 (DV_{t-5})^2 - 4.4844 (DV_{t-5})
\]
\[
470^2 - 0.9466(UV_t > 470) - 0.0650 \cdot (UOCC_t)^2 + 2.6809 \cdot (UOCC_t) - 0.1480 \]
\[
(UOCC_{t-5})^2 + 5.1385 \cdot (UOCC_{t-5}) + 0.0041(DV_{t-5})^2 - 5.2752(DV_{t-5}) \tag{16}
\]

Predicting average link speed at 15 minutes forward.

\[
ALS_{15} = 2805.143 + 0.0019(UV_t \leq 470)^2 - 1.6157(UV_t \leq 470) + 0.0010(UV_t > 470)^2 - 1.1521(UV_t > 470) - 0.0697 \cdot (UOCC_t)^2 + 2.7029(UOCC_t) - 0.1738 \]
\[
(UOCC_{t-5})^2 + 6.1238 \cdot (UOCC_{t-5}) + 0.0061(DV_{t-5})^2 - 7.8105(DV_{t-5}) \tag{17}
\]

Estimating link speed at current time-step for off-peak hours.

\[
ALS_0 = 491.1878 + 0.0278 \cdot (US_t)^2 - 5.3405(US_t) + 0.0153 \cdot (US_{t-5})^2 - 2.9197(US_{t-5}) \tag{18}
\]

Where; \(US_t\) = Upstream speed at current time-step (km/hr)
\(US_{t-5}\) = Upstream speed at a time-step backward (km/hr)

Estimating link speed at current time-step for afternoon peak hours.

\[
ALS_0 = 90.9786 - 0.0003 \cdot (UV_t)^2 - 0.1588(UV_t) + 0.1909 \cdot (UOCC_{t-10})^2 - 2.8604 \cdot (UOCC_{t-10}) + 0.0927 \cdot (UOCC_{t-15})^2 - 1.3856(UOCC_{t-15}) \tag{19}
\]

Where; \(UOCC_{t-10}\) = Upstream Occupancy at two time-step backward (percent).
\(UOCC_{t-15}\) = Upstream Occupancy at three time-step backward (percent).

For performance comparison, neural network based models were developed using a tool called Intelligent Problem Solver (IPS) in Statistica 7.0 (StatSoft, Inc., 2004). Traditionally, it is necessarily to run training algorithms a number of times with a given "neural network design," selecting the best network or a few of the best. Further, a neural network is also being selected. Therefore, a number of experiments with different designs are conducted, and the best networks selected. The Intelligent Problem Solver (IPS) follows a similar process. The search algorithms are used to determine the selection of inputs, the number of hidden units, and other key factors in the network design. These search algorithms are interleaved so that the IPS searches for optimal networks of different types (for example, Multilayer Perceptrons and Radial Basis Functions) simultaneously. However, the investigation of whether IPS actually provides the best network or enough training was not performed in the study and is recommended in future efforts. From the above procedures, the Multilayer perceptron neural network seemed to be the more suitable for both train set and test set. For all time periods and prediction time interval, the number of hidden layer was one. The numbers of nodes in hidden layer were 9, 10, and 6 for travel time estimation at current time step during morning peak, off-peak, and afternoon hours, respectively. For travel time prediction during morning peak hours at time \(t+5\), \(t+10\), and \(t+15\), the numbers of nodes in hidden layer were 11, 8, and 7, respectively. Due to the limitation of the output from software used, the synaptic weights can not be presented here. Next, a comparison of performance of the proposed models and neural network based models will be discussed.
8. MODEL PERFORMANCES

The performances of the models to predict link travel time were reported as average root mean square errors (RMSE) and mean absolute percent errors (MAPE) of 30 simulation sets. The RMSE and MPE are defined as follows;

\[
RMSE = \sqrt{\frac{\sum (TT_{sim} - TT_{estimate})^2}{N}}
\]

\[
MAPE = \frac{1}{N} \sum \left[ \frac{|TT_{sim} - TT_{estimate}|}{TT_{sim}} \times 100 \right]
\]

Where;
- \( TT_{sim} \) = Link travel time from simulation
- \( TT_{estimate} \) = Link travel time estimation or prediction using proposed model or neural network
- \( N \) = number of samples

For morning peak hours, the link travel time increased from 425 seconds at around 6.50 AM to 665 seconds at around 8.10 AM, and slightly decreased to about 620 seconds at around 8.40 AM (Figure 3). For off-peak hours, the link travel time ranged from 261 seconds to 265 seconds (Figure 4). The variation of link travel time was relatively low due to light traffic volume. The link travel time variation during afternoon hours was also moderately low. For afternoon hours, the link travel time ranged from 262 to 268 seconds (Figure 5).

Table 3 shows RMSEs and MAPEs of models to estimate link travel time at current time-step on train sets and test sets. Overall, the RMSE on the train set was higher than RMSE on the test set, as expected. The RMSEs of the proposed model on test sets were 88.63, 6.88, and 2.38 seconds for morning peak hours, off-peak hours, and afternoon hours. The RMSE of the models for off-peak hours and afternoon hours were relatively lower than that of morning peak-hour. It may be due to lighter traffic and relatively lower variation of travel time. In comparison with neural network based model, the proposed model provided lower RMSE. The MAPEs of the proposed model on the test sets were 10.89, 3.32, 2.73 percent for morning peak hours, off-peak hours, and afternoon hours. For morning peak hours, MAPE of the proposed model was lower than that of neural network, but they were slightly higher than those of neural network for off-peak hours and afternoon hours.

Table 4 shows the RMSEs and MAPEs of models to predict link travel time at one time-step forward (t+5), two time-step forward (t+10), and three time-step forward (t+15). The RMSEs were 91.20, 101.86, 122.08 seconds and the MAPEs were 11.30, 13.01, 15.52 percent, for the models to predict link travel time at t+5, t+10, and t+15, respectively. The RMSEs increase as prediction time-step proceeds. At prediction interval t+5 and t+10, the RMSEs of the proposed model were lower, though the MAPEs were higher than those of neural network.
Figure 3 Comparison of link travel time (Ratchadapisek-Klong Papa) during morning peak hours

Figure 4 Comparison of link travel time (Ratchadapisek-Klong Papa) during off-peak hours

Figure 5 Comparison of link travel time (Ratchadapisek-Klong Papa) during afternoon hours
Table 3 RMSE and MAPE of models to estimate link travel time at current time-step (t)

<table>
<thead>
<tr>
<th>Time period</th>
<th>Train Set</th>
<th></th>
<th>Test Set (average 30 sets)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RMSE</td>
<td>MAPE</td>
<td>RMSE</td>
<td>MAPE</td>
</tr>
<tr>
<td></td>
<td>(seconds)</td>
<td>(percent)</td>
<td>(seconds)</td>
<td>(percent)</td>
</tr>
<tr>
<td></td>
<td>Model</td>
<td>Neural Network</td>
<td>Model</td>
<td>Neural Network</td>
</tr>
<tr>
<td>AM Peak hours (6.30 - 9.30 AM)</td>
<td>27.60</td>
<td>37.92</td>
<td>3.86</td>
<td>5.29</td>
</tr>
<tr>
<td>Off-peak hours (11 AM - 2 PM)</td>
<td>0.40</td>
<td>0.42</td>
<td>0.51</td>
<td>0.14</td>
</tr>
<tr>
<td>PM hours (3-6 PM)</td>
<td>0.51</td>
<td>0.58</td>
<td>0.15</td>
<td>0.14</td>
</tr>
</tbody>
</table>

Table 4 RMSE and MAPE of link travel time at different prediction interval on the test set

<table>
<thead>
<tr>
<th>Time period</th>
<th>Prediction Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>t+5</td>
</tr>
<tr>
<td></td>
<td>RMSE (seconds)</td>
</tr>
<tr>
<td></td>
<td>Model NN</td>
</tr>
<tr>
<td>AM Peak hours (6.30 - 9.30 AM)</td>
<td>91.20</td>
</tr>
</tbody>
</table>

8.1 Volume Sensitivity of the Models

To test the volume sensitivity of the models, the test link during morning peak hours was re-simulated for other two cases; increasing traffic volume by 5 percent and decreasing traffic volume by 5 percent. Both cases were simulated and used to test the performance of the model developed. When traffic volume was increased by 5 percent, the average link travel time was 702 seconds and the RMSE of model developed was 97.20 seconds (Figure 6). The RMSE of neural network based model was 155.09 seconds. When traffic volume was decreased by 5 percent, the average link travel time was 609 seconds (Figure 7). The proposed model showed high fluctuation of link travel time estimates compared with that from simulation. This may be due to traffic conditions have differed from the traffic conditions that the model has been developed from, and the model relatively sensitive to those change. The recalibration of model may require when traffic conditions changed. The RMSE was 113.48 seconds and 166.75 seconds for the model developed and the neural network based model, respectively (Table 5).
Figure 6 Link travel time when volume was increased by 5 percent

Figure 7 Link travel time when volume was decreased by 5 percent

Table 5  RMSEs of the models after increased and decreased volume

<table>
<thead>
<tr>
<th>Data</th>
<th>RMSE (seconds)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Model VS. Simulation data</td>
</tr>
<tr>
<td>Volume increased by 5 percent (Re-simulated)</td>
<td>97.20 (14.82%)</td>
</tr>
<tr>
<td>Volume decreased by 5 percent (Re-simulated)</td>
<td>113.48 (20.31%)</td>
</tr>
</tbody>
</table>
9. CONCLUSIONS AND RECOMMENDATIONS

In this study, the models to estimate freeway link travel time at current time-step and to predict freeway link travel time at 5, 10, and 15 minutes forwards were developed. The parametric models were developed based on relationships suggested by non-parametric GAM. The GAM is a novel technique and was firstly applied for travel time estimation in this research. The RMSE of the models for off-peak hours and afternoon hours were relatively lower than that of morning peak-hour. It may be due to lighter traffic and relatively lower variation of travel time. For morning peak hours, MAPE of the proposed model was lower than that of neural network, but they were slightly higher than those of neural network for off-peak hours and afternoon hours. At prediction interval t+5 and t+10, the RMSEs of the proposed model were lower, though the MAPEs were higher than those of neural network.

In conclusion, the proposed parametric GAM models have strong potential for on-line implementation. Under prevailing roadway and traffic conditions of the test link, the proposed models are comparable with the neural network based models. It may be mentioned that all model performances presented in this paper are based on the limited test data set. The temporal and spatial variations of data were not investigated. It is also recommended that transferability test or test the model developed on other freeway locations and the capability of describing incident situations should be investigated in next efforts.

ACKNOWLEDGEMENTS

The research was supported by a research grant to the first author from the Thailand Research Fund (TRF). The contents of this paper reflect the views of the authors and do not necessarily reflect the official views of the Thailand Research Funds. The authors greatly appreciate the generous assistance of Mr. Korakot Kerdkaew.

REFERENCES


for Freeway Corridors, Transportation Research Record 1676, 184-191.