Prediction Intervals to Account for Uncertainties in Travel Time Prediction

Abbas Khosravi, *Member, IEEE*, Ehsan Mazloumi, Saeid Nahavandi, *Senior Member, IEEE*, Doug Creighton, *Member, IEEE*, and J. W. C. (Hans) van Lint, *Member, IEEE*

***Abstract*—The accurate prediction of travel times is desirable but frequently prone to error. This is mainly attributable to both the underlying traffic processes and the data that are used to infer travel time. A more meaningful and pragmatic approach is to view travel time prediction as a probabilistic inference and to construct prediction intervals (PIs), which cover the range of probable travel times travelers may encounter. This paper introduces the delta and Bayesian techniques for the construction of PIs. Quantitative measures are developed and applied for a comprehensive assessment of the constructed PIs. These measures simultaneously address two important aspects of PIs: 1) coverage probability and 2) length. The Bayesian and delta methods are used to construct PIs for the neural network (NN) point forecasts of bus and freeway travel time data sets. The obtained results indicate that the delta technique outperforms the Bayesian technique in terms of narrowness of PIs with satisfactory coverage probability. In contrast, PIs constructed using the Bayesian technique are more robust against the NN structure and exhibit excellent coverage probability.**

***Index Terms*—Bayesian inference, delta method, neural networks (NNs), prediction intervals (PIs).**

I. Introduction

THERE HAS BEEN a significant amount of work identifying the importance of accurate travel time predictions in a transportation system. From the travelers' perspective, accurate travel time predictions reduce the uncertainty in decision making about departure time and route choice, which in turn reduce the travelers' stress and anxiety [1]. From the operators' point of view, travel time prediction models may be used to determine the reliability of a transportation system [2]. They can also be dynamically monitored to provide information about the current traffic state in a network and to find the problematic locations [3]. Consequently, travel time prediction methods are central to advanced traveler information systems and advanced traffic management systems [4]–[7].

Manuscript received March 5, 2010; revised June 30, 2010 and October 25, 2010; accepted January 6, 2011. Date of publication February 4, 2011; date of current version June 6, 2011. This work was supported by the Centre for Intelligent Systems Research, Deakin University. The Associate Editor for this paper was W. Fan.

A. Khosravi, S. Nahavandi, and D. Creighton are with the Centre for Intelligent Systems Research, Deakin University, Geelong, Vic. 3217, Australia (e-mail: akhos@deakin.edu.au; saeid.nahavandi@deakin.edu.au; douglas.creighton@deakin.edu.au).

E. Mazloumi is with the Institute of Transport Studies, Department of Civil Engineering, Monash University, Melbourne, Vic. 3800, Australia (e-mail: ehsan.mazloumi@eng.monash.edu.au).

J. W. C. van Lint is with the Department of Transport and Planning, Faculty of Civil Engineering and Geosciences, Delft University of Technology, Delft 2600, The Netherlands (e-mail: j.w.c.vanlint@tudelft.nl).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TITS.2011.2106209

Aside from those general benefits, in the public transport community, information about the future travel times of buses may support proactive real-time operational management strategies, such as holding and expressing [8]. Planners also use future travel times in their offline planning, including fleet size planning and schedule design [9].

Among the many types of existing methodologies, data-driven models have been frequently used for travel time prediction. Many efforts have been reported on the application of methods based on Kalman filtering [10], [11], generalized linear regression [12], nearest-neighborhood methods [13], and neural networks (NNs) [14], to name a few. Among these methods, NNs have shown their promising performance in predicting urban travel times [15] and bus travel times [16]. Since NNs have proven accurate and reliable traffic predictors [17], in this paper, we will develop a travel time prediction approach on the basis of NNs, which allows for the construction of prediction intervals (PIs).

Bus travel times are the result of nonlinear and complex interactions of many different constituent factors influencing either demand (e.g., passenger's demand or traffic flow) or capacity (e.g., accidents, weather condition, route characteristics) [18]. The probabilistic nature of some determining factors, such as passenger demand, bus drivers' behavior, traffic accidents, and, more importantly, signal delay experienced by different buses, leads to stochastic behaviors in this system. For the case of freeway travel times, there are also similar uncertainties, including inflow and capacity constraints at the boundaries of the considered freeway stretch, random acceleration and deceleration, unpredictable behaviors from drivers, and accidents.

Despite all the attempts made so far, the inherent stochasticity of variables and their complexity and nonlinear interrelationships make the application of NNs, and, in general, any predictive model, very challenging. In essence, the presence of uncertainty in determinants of travel time significantly degrades the performance of the exploited prediction models. As pointed out in [19], the performance of any deterministic model like NNs highly deteriorates under these conditions. Therefore, a more meaningful (and at the same time more pragmatic) approach is to consider travel time prediction as an example of probabilistic inference, which naturally leads to predicting the most probable distribution of travel time, rather than one crisp (mean) value.

To effectively ameliorate these defects, part of the research in the past has materialized into a series of studies with the purpose of developing PIs. PIs take into account both the uncertainty in model structure and the noise in input data. They allow the users of the produced point predictions to perceive the accuracy of the predictions and thus to make informed decisions to keep or reject the result. Among the most common techniques for constructing PIs for NN forecasts are the bootstrap technique [20], [21], the Bayesian technique [22], [23], the delta technique [24], [25], and the mean-variance estimation method [26], [27]. Despite their ease of implementation, Bootstrap-based methods for PI construction are computationally expensive [25], [28]. Mean-estimation variance often underestimates the variance of data [29], leading to low coverage probability [30]. The Bayesian and delta techniques are powerful methods on the basis of strong mathematical foundations [22], [24]. The application of these methods for the construction of PIs for NN predictions has proliferated in recent years, as demonstrated in many different fields [31]–[34].

To the best of our knowledge, there are a few reports on the successful construction and exploitation of PIs in the transportation society. Sun *et al.* [35] developed and compared asymptotic and bootstrap PIs for linear models. The application of the Bayesian technique has recently been reported for the construction of confidence intervals [36]. The mean-variance estimation method was also used in [27] to predict the bus travel time variability.

While in the literature many measures like mean square error (MSE), mean absolute percentage error (MAPE), and coefficient of determination (R^2) have been proposed for the assessment of the performance of point prediction models, there is no such measure for PIs. Often, the coverage probability of PIs is considered as the main index for judging their goodness. Although this measure is theoretically very important, assessing PIs only based on this may ban their further applications. We will argue on this issue in Section III when proposing some new measures for quantitative PI evaluation.

Motivated by these gaps in the literature, this paper first develops a practically useful measure for the quantitative evaluation of PIs in terms of their length and coverage probability. This measure is mainly composed of indices about how well PIs cover the underlying targets and how wide they are compared with the range of targets. Then, these measures are used for comparing PIs constructed using the Bayesian and delta techniques. All the experiments will be conducted using data sets taken from a bus route in Melbourne, Australia, and a freeway in Delft, The Netherlands.

The rest of this paper is organized as follows: Section II provides a brief review of the fundamental theories of the Bayesian and delta technique. The new PI assessment measures are described in Section III. Experimental results are demonstrated in Section IV. Finally, Section V concludes this paper with some remarks for further study in this domain.

II. Techniques for Prediction Interval Construction

By definition, a PI with a confidence level of $(1-\alpha)\%$ is a random interval developed based on past observations $x = (x_1, x_2, \ldots, x_n)$ for future observations

$$\mathrm{PI} = [L(x), U(x)] \tag{1}$$

such that

$$\Pr\left(L(x) \leq x_{n+1} \leq U(x)\right) = 1 - \alpha. \tag{2}$$

$L(x)$ and $U(x)$ correspond to the lower and upper bounds of PIs. The confidence level $((1-\alpha)\%)$ of a PI refers to the expected probability that the real value is within the predicted interval.

A. Bayesian Technique

The Bayesian technique interprets the NN parameter uncertainty in terms of probability distributions and integrates them to obtain the probability distribution of the target conditional on the observed training set [22]. In fact, it is based on the NN Bayesian training algorithm, which aims at improving the generalization capabilities of the conventional backpropagation algorithm. From a Bayesian inference perspective, the NN parameters are considered as random variables with unknown distributions. The posterior probability of an NN model with a set of parameter and a training data set can be determined using Bayes' rule as

$$P(\theta|D,\rho,\beta,M) = \frac{P(D|\theta,\beta,M)P(\theta|\rho,M)}{P(D|\rho,\beta,M)} \tag{3}$$

where $P(D|\theta,\beta,M)$ and $P(\theta|\rho,M)$ are the likelihood function of data occurrence and the prior density of parameters, respectively. $P(D|\rho,\beta,M)$ is also a normalization factor enforcing the fact that the total probability is 1. β and ρ are hyperparameters of the Bayesian technique for training parameters of NN (θ). M is also the particular NN model considered for the training. It is possible to consider different distributions for $P(D|\theta,\beta,M)$ and $P(\theta|\rho,M)$. To simplify the further analysis, it is often assumed that they are Gaussian, i.e.,

$$P(D|\theta,\beta,M) = \frac{1}{Z_D(\beta)} e^{-\beta E_D} \tag{4}$$

$$P(\theta|\rho,M) = \frac{1}{Z_\theta(\rho)} e^{-\rho E_\theta} \tag{5}$$

where $Z_D(\beta) = (\pi/\beta)^{n/2}$, and $Z_\theta(\rho) = (\pi/\rho)^{p/2}$. n and p are number of training samples and NN parameters, respectively. Substituting (4) and (5) into (3) results in

$$P(\theta|D,\rho,\beta,M) = \frac{1}{Z_F(\beta,\rho)} e^{-(\rho E_\theta + \beta E_D)}. \tag{6}$$

When training NNs using the Bayesian method, the purpose is to maximize $P(\theta|D,\rho,\beta,M)$. Such maximization can be achieved through minimization of the following cost function:

$$E(\theta) = \rho E_\theta + \beta E_D \tag{7}$$

where E_D is the sum of squared errors, and

$$\mathrm{SSE} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{8}$$

where y_i and $\hat{y}_i$ are the ith target and the predicted value, respectively. In addition, E_θ in (6) is the sum of squares of the network weights ($\theta^T\theta$). The optimal values for β and ρ maximizing the posterior probability function in (6) are as follows:

$$\beta^{MP} = \frac{\gamma}{E_D(\theta^{MP})} \tag{9}$$

$$\rho^{MP} = \frac{n-\gamma}{E_\theta(\theta^{MP})} \tag{10}$$

where γ is the effective number of parameters, and

$$\gamma = p - 2\rho^{MP} tr(H^{MP})^{-1} \tag{11}$$

where $H^{MP} = \beta\nabla^2 E_D + \rho\nabla^2 E_\theta$ is the Hessian matrix. Application of this technique for training NNs results in NNs where the variance of their prediction is

$$\sigma_i^2 = \sigma_D^2 + \sigma_{\theta^{MP}}^2 = \frac{1}{\beta} + \nabla_{\theta^{MP}}^T \hat{y}_i (H^{MP})^{-1} \nabla_{\theta^{MP}} \hat{y}_i \tag{12}$$

where $\nabla_{\theta^{MP}} \hat{y}_i$ is the gradient of the NN output with respect to its parameters (θ^{MP}). Therefore, a $(1-\alpha)\%$ PI for future samples can be constructed as follows:

$$\hat{y}_i \pm z_{1-\frac{\alpha}{2}} \left(\frac{1}{\beta} + \nabla_{\theta^{MP}}^T \hat{y}_i (H^{MP})^{-1} \nabla_{\theta^{MP}} \hat{y}_i\right)^{\frac{1}{2}} \tag{13}$$

where $z_{1-(\alpha/2)}$ is the $1-(\alpha/2)$ quantile of a normal distribution function. Further information about this technique and its detailed mathematical discussion can be found in [22] and [23].

B. Delta Technique

The delta technique is based on the interpretation of NNs as nonlinear regressors containing many parameters to be adjusted. Such an interpretation suggests calculating PIs by applying standard asymptotic theory to them. For a given set of n pairs of data $(x_i, y_i)_{i=1}^n$, where x_i is m-dimensional, suppose that the NN model is represented as follows:

$$y_i = f(x_i, \theta) + \epsilon_i, \qquad i = 1, 2, \ldots, n \tag{14}$$

where θ represents the true set of NN parameters (weights and biases). ϵ_i's are assumed to be identically and independently distributed with mean zero and variance σ^2 ($N(0, \sigma^2)$). Finding the true sets of NN parameters in reality is problematic. Therefore, often, NN parameters are adjusted based on minimization of (8), i.e.,

$$\hat{y}_i = f(x_i, \hat{\theta}) \tag{15}$$

where $\hat{\theta}$ represents the obtained set of NN parameters through the minimization of (8). Linearization of the NN model [see (15)] around the true values of model parameters (θ) is as follows:

$$\hat{y}_i = f(x_i, \theta) + g_i^T(\hat{\theta} - \theta) \tag{16}$$

where g_i^T is the gradient of NN models with respect to its parameters. This linearization allows us to construct PIs for the NN model. Based on the assumptions made regarding the distribution of ϵ_i in (14), the $(1-\alpha)\%$ PI for $\hat{y}_i$ is

$$\hat{y}_i \pm t_{1-\frac{\alpha}{2}, df}\, s\sqrt{1 + g_i^T (F^T F)^{-1} g_i} \tag{17}$$

where α is the significance level, and $(1-\alpha)\%$ is the level of confidence to the constructed PIs. $t_{df}^{1-(\alpha/2)}$ is the $1-(\alpha/2)$ quantile of a cumulative t-distribution function with df degrees of freedom. df here is the difference between the number of training samples (n) and the number of NN parameters (p). F is also the Jacobian matrix of the NN model with respect to its parameters. Veaux *et al.* [25] extended this PI construction method to the case where NN parameters are adjusted based on the weight decay cost function

$$\mathrm{SSE} + \lambda\hat{\theta}^T\hat{\theta} \tag{18}$$

where λ is the regularizing factor [22]. Compared with the cost function in (8), the latter highly penalizes NNs with higher complexity and thus improves the generalization power of networks. PIs constructed using this variation of the delta technique are as follows:

$$\hat{y}_i \pm t_{1-\frac{\alpha}{2}, df}\, s\sqrt{1 + g_i^T \Delta g_i} \tag{19}$$

where $\Delta = (F^T F + \lambda I)^{-1}(F^T F)(F^T F + \lambda I)^{-1}$. s is the estimate of the standard deviation obtained for the training samples, i.e.,

$$s = \frac{\sqrt{\mathrm{SSE}}}{n - trac(2\Gamma - \Gamma^2)} \tag{20}$$

where $\Gamma = J^T(J^T J + \lambda I)^{-1} J$.

A more detailed discussion on the basis and implications of the delta technique can be read in [24] and [25].

III. Prediction Interval Assessment Indices

While the literature offers a variety of methods for the evaluation of the performance of point prediction methods (e.g., error-based measures), there is no well-established index for PI assessment. This is why the majority of studies done on the area of PI fail to assess PIs quantitatively. As evaluation is often only made based on PI empirical coverage probability, and without inclusion of their length, the obtained results sound more subjective than objective. Often, the discussion about the length of PIs is either ignored [24], [37] or vague and companied with fuzzy terms, such as "small" [25] and "short" [38]. In this section, an examination measure is proposed, which covers both important aspects of PIs: length and coverage probability.

The spontaneous measure for the assessment of PI quality is the empirical coverage probability. PI coverage probability (PICP) is the key property of PIs and is interpreted as the probability that a set of PIs will result in a high coverage of the targets. In mathematical terms, PICP is defined as follows:

$$\mathrm{PICP} = \frac{1}{n}\sum_{i=1}^{n} c_i \tag{21}$$

where $c_i = 1$ if $y_i \in [L(x_i), U(x_i)]$; otherwise, $c_i = 0$.

Theoretically, PICP should be as close as possible to its nominal value $(1-\alpha)\%$, which is the confidence level for which PIs have been constructed. The imperfectness of PICP in real-world applications is mainly attributable to the presence of noise in samples and the severe effects of uncertainty. Other issues, such as underfitting and overfitting (which are a direct result of the inappropriate size of NNs), also contribute to the unsatisfactory smallness of PICP.

The PICP described in (21) is the most frequently used measure in the literature for the assessment of the performance of a method for construction of PIs. A high PICP has often been interpreted as a strong indication of good performance of PI construction methods. A very important point that has been frequently ignored is that a high PICP can easily be achieved through the selection of upper and lower bounds of PIs to be a large fraction of extreme values of the target (maximum and minimum values). In the case of using bounds to be exactly like the extreme values, 100% PICP is achievable. The argument here is that such wide PIs are useless as they carry no information about targets. Therefore, it is essential to read and interpret PICP in conjunction with a width-based index. Following this, a measure called mean PI length (MPIL) can be obtained as follows:

$$\text{MPIL} = \frac{1}{n}\sum_{i=1}^{n}\left(U(x_i) - L(x_i)\right). \tag{22}$$

MPIL is the mean of length of the constructed PIs. The PI lengths can also be assessed through normalizing each PI length with regard to the target range. If the range of the underlying target is *a priori* known (R), then the normalized MPIL (NMPIL) is calculated as follows:

$$\text{NMPIL} = \frac{MPIL}{R}. \tag{23}$$

Normalization of the PI length by the range of targets makes the objective comparison of PIs possible, regardless of techniques used for their construction or the magnitudes of the underlying targets. From a practical standpoint, NMPIL is an indication of how PIs are narrow. In the case of using extreme values as PIs, NMPIL will be equal to 1.

Generally, PI lengths and PICP have a direct relationship. The wider the PIs, the higher the corresponding PICP. Therefore, a measure is required to simultaneously cover and address both aspects for the comprehensive assessment of PIs. Focus on each side of the PIs may lead to misleading results. For instance, very narrow PIs with a low coverage probability are not very reliable. On the other hand, very wide PIs with a high coverage probability are not practically very useful. With regard to this argument, the following coverage-length-based criterion (CLC) is proposed for the comprehensive evaluation of PIs in terms of their coverage probability and lengths:

$$\text{CLC} = \text{NMPIL}\left(1 + e^{-\eta(\text{PICP}-\mu)}\right) \tag{24}$$

where η and μ are two controlling parameters. The level of confidence associated with PIs can appropriately be used as a guide for selecting the hyperparameters of CLC. One reasonable principle is that we highly penalize PIs if their PICP is less than $(1-\alpha)\%$. This is based on the theory that the coverage probability of PIs in an infinite number of replicates will approach toward $(1-\alpha)\%$.

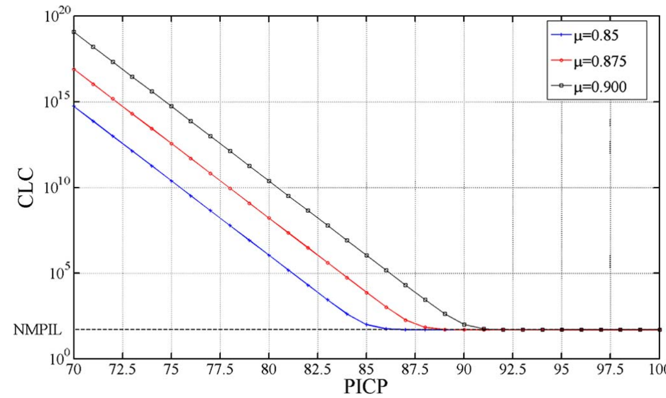

Fig. 1. CLC for different values of its parameters μ while $\eta = 200$ and $NMPIL = 50\%$.

Fig. 1 demonstrates the CLC for different values of μ and PICP. It has been assumed that NMPIL is 50% and η has been set to 200. The three curves correspond to three values of μ. As PICP decreases, CLC starts to jump very quickly. The point of quick rise is controlled by μ, whereas the jump intensity is determined by η. Setting η and μ is done based on the confidence level of PIs, i.e., $(1-\alpha)\%$. For instance, if the confidence level is 90%, then the values of η and μ can easily be adjusted to guarantee a great increase of CLC for $PICP \leq 90\%$. Based on this, the CLC will highly increase, no matter what the length of the PI is. This way, PIs with unsatisfactory coverage probability are heavily penalized. Generally, the smallness of CLC is an indication of the goodness of the constructed PIs (simultaneously achieving small NMPIL and high PICP). Whether CLC has a large or small value is totally case dependant. However, if PICP is sufficiently high, CLC and NMPIL will be almost the same. This has been demonstrated in Fig. 1 using the dashed line. As PICP takes its nominal value, CLC approaches NMPIL.

IV. Experiments and Results

A. Data

Data for the first transportation problem were supplied from GPS-equipped buses operating on an 8-km segment of bus route 246 in inner Melbourne, Australia. This segment comprises four sections (see Fig. 2). Each section is defined as the distance between two consecutive timing point stops, where the arrival and departure times of buses are recorded by GPS to monitor service consistency and schedule adherence. The four sections are similar in length. The buses in this section operate in mixed traffic, and there is no separate lane allocated to them. In this paper, a data set including weekday travel times in one direction of this segment is used. The data set provides travel times of almost 1800 trips corresponding to each route section, which were collected over a period of six months in 2007. The travel times over each route section vary considerably across different days. The range of bus travel times is 945, 713, 616, and 615

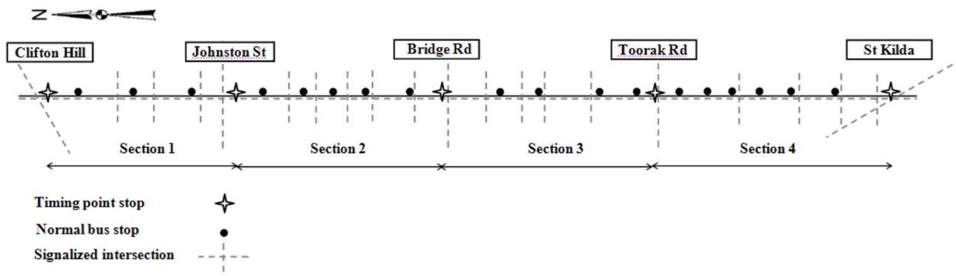

Fig. 2. Schematic representation of the study testbed.

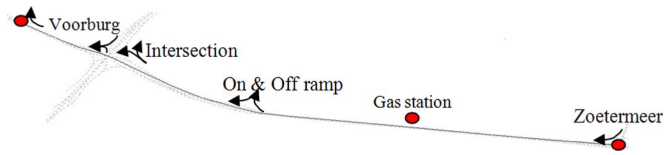

Fig. 3. A12 freeway from Zoetermeer to The Hague in the Netherlands.

in sections 1–4, respectively. This range might be related to factors including variations in passenger demand and traffic flow over different days, various signal delays experienced by different buses, and variation in driving style of bus drivers on different days.

The study needed explanatory variables to construct PIs. Provided for the research were traffic counts and traffic saturation degree values collected by the Sydney Coordinated Adaptive Traffic System (SCATS) loop detectors from each intermediate signalized intersection. In addition, provided was a measure of schedule adherence. The latter was quantified for each bus by subtracting the observed arrival time from the scheduled arrival time at each timing point stop. For each route section, this paper aims to relate the travel times to the traffic state at intermediate signalized intersections before each bus departed from the upstream timing point. To this end, and consistent with [27] and [39], which worked on the same data sets, this paper considers the schedule adherence at the upstream timing point along with the traffic saturation degree values aggregated over the 15-min period before the departure of each bus.

As the second case study, the PI construction methods are also applied to an 8.5-km-long route of the A12 freeway in the Netherlands, from an on ramp (Zoetermeer) to an off ramp (Voorburg) (see Fig. 3). At both on and off ramps, license plate cameras record the license plate of each vehicle. Individual travel times, based on matches of license plates, were available for 95 days in the winter and spring of 2007. The data were filtered for outliers, which were a considerable number, mainly due to the fact that only four characters out of six are recorded due to privacy legislations. After filtering the data and inspecting them visually, the travel times of the vehicles leaving in the same 5-min time period were averaged. A total of 39 peak periods of about 3.5 h each were selected from the data set. As inputs to the NNs, 19 double-loop detectors are available, reporting speeds every minute. The speed data are available in 1-min arithmetic mean speeds of all vehicles that are recorded (i.e., time mean speeds). The travel times of the vehicles leaving in the same 5-min time period are averaged. Time and weekday are also considered as inputs to NN models (in total, 21 inputs). No additional information was available on the occurrence of, for example, incidents or accidents. The travel times are between 273 and 1280 s, which means that the congestion has occurred many times, leading to long travel times. In addition, the coefficient of variation for the travel times is 43.02%, indicating the high variability of travel times around the mean.

TABLE I
DATA SETS USED IN THE EXPERIMENTS

Target	Samples	Inputs	Case	$R^2(\%)$
Bus travel time				
Section one of the bus route	1643	5	#1	46.29
Section two of the bus route	1741	6	#2	30.06
Section three of the bus route	1750	6	#3	38.80
Section four of the bus route	1613	6	#4	25.42
Freeway travel time				
The last five minute data	1643	22	#5	83.73
The last ten minute data	1642	22	#6	77.34
The last fifteen minute data	1641	22	#7	73.44

To predict the freeway travel times, three cases are considered. We use the data in the last 5- to 15-min intervals to predict the travel times. This means using the data of 8:55 (case study #5), 8:50 (case study #6), and 8:45 (case study #7) to predict the travel times at 9:00, and so forth.

Table I summarizes the characteristics of the used data sets in this paper. As previously discussed, cases #1–#4 relate to the bus travel time prediction, and cases #5–#7 relate to the freeway travel time prediction. All experiments described in the following section are applied to these data sets.

B. Results

The data mentioned in Table I are randomly split into two sets: 1) training set (80%) and 2) test set (20%). It is first shown that the NN performance for forecasting travel time is unsatisfactory, regardless of its size, structure, and parameters. In the experiments, the NNs have two layers with variable number of neurons. A grid of different structures is developed through changing the number of neurons between 1 and 10 in each layer ($n_1 \in [1, 10]$, and $n_2 \in [1, 10]$). To avoid any subjective argument about the performance of NNs, networks are initialized and trained five times. Then, the coefficient of determination (R^2) is calculated and averaged for test samples to measure the goodness of fit.

Table I summarizes the best prediction results obtained using the developed NNs on the grids (in total, 500 NNs were trained and tested for each case study). The small values of R^2 indicate the inability of NNs to explain stochasticity in the bus and freeway travel times. This inability is much more severe for

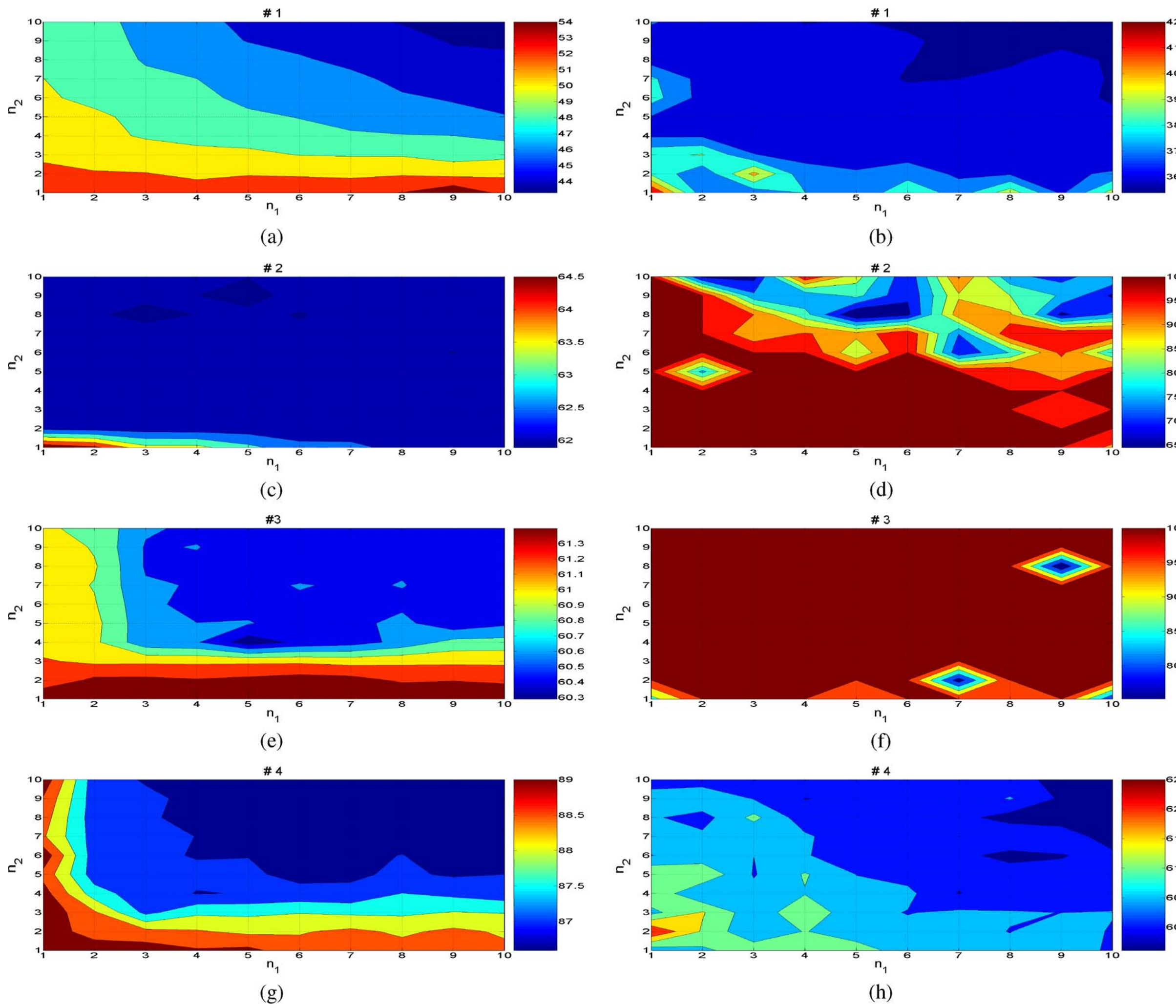

Fig. 4. Averaged CLCs on a grid composed of n_1 and n_2 for PIs constructed using the Bayesian (left side) and delta (right side) techniques for cases #1–#4. (a) Bayesian method for case #1. (b) Delta method for case #1. (c) Bayesian method for case #2. (d) Delta method for case #2. (e) Bayesian method for case #3. (f) Delta method for case #3. (g) Bayesian method for case #4. (h) Delta method for case #4.

the case of bus travel times. The best R^2 for cases #1–#4 is 46.29% (section one of the bus route). For the case of freeway travel times, the prediction accuracy is higher (almost two times better than the other case) but still far from the acceptable range (at least 90%). The smallness of R^2 is not attributable to the NN structure or its training process, because we have examined different structures (through varying number of hidden neurons) and repeated the training process five times.

The prediction results in Table I clear that the level of uncertainty in the bus travel time data set is much higher than the freeway travel times. This can be justified based on a number of known and unknown factors significantly affecting the bus movement in its routes. Examples are traffic lights, passengers' demand, and a high rate of congestion in rush hours. As many of these factors/variables are not measurable, the level of uncertainty associated with the travel time predictions generated by NNs is very high. This well justifies the construction of PIs and their application instead of point predictions.

In the next experiments, Bayesian and delta techniques are applied to bus and freeway data sets for the construction of PIs. The performance of both methods intimately hinges on the NN structure and the training process. Similar to the previous experiments, a grid of NNs with different structures is used ($n_1 \in [1, 10]$, and $n_2 \in [1, 10]$). In addition, experiments on each node are repeated five times and averaged to minimize the effect of random initialization. All PIs are constructed with 90% confidence level. The measures described in Section III are applied for the evaluation and comparison of the quality of constructed PIs. The purpose of comparison is to find which method leads to narrower PIs with a higher coverage probability.

In the beginning, some experiments were conducted to compare the performance of PIs constructed based on (17) and (19). It was found that in the case of using the former, many PIs are not reliable due to the singularity of the Jacobian matrix. In contrast, using (19) resulted in more reliable PIs. This explicitly points out the superiority of the weight decay cost function method over the traditional MSE cost function method for training NNs and development of PIs. Therefore, all PIs described and discussed later have been built using (19) instead of (17). The regularizer factor λ is fixed at 0.9.

Contour plots of the averaged CLCs of PIs constructed using the Bayesian and delta techniques are shown in Fig. 4

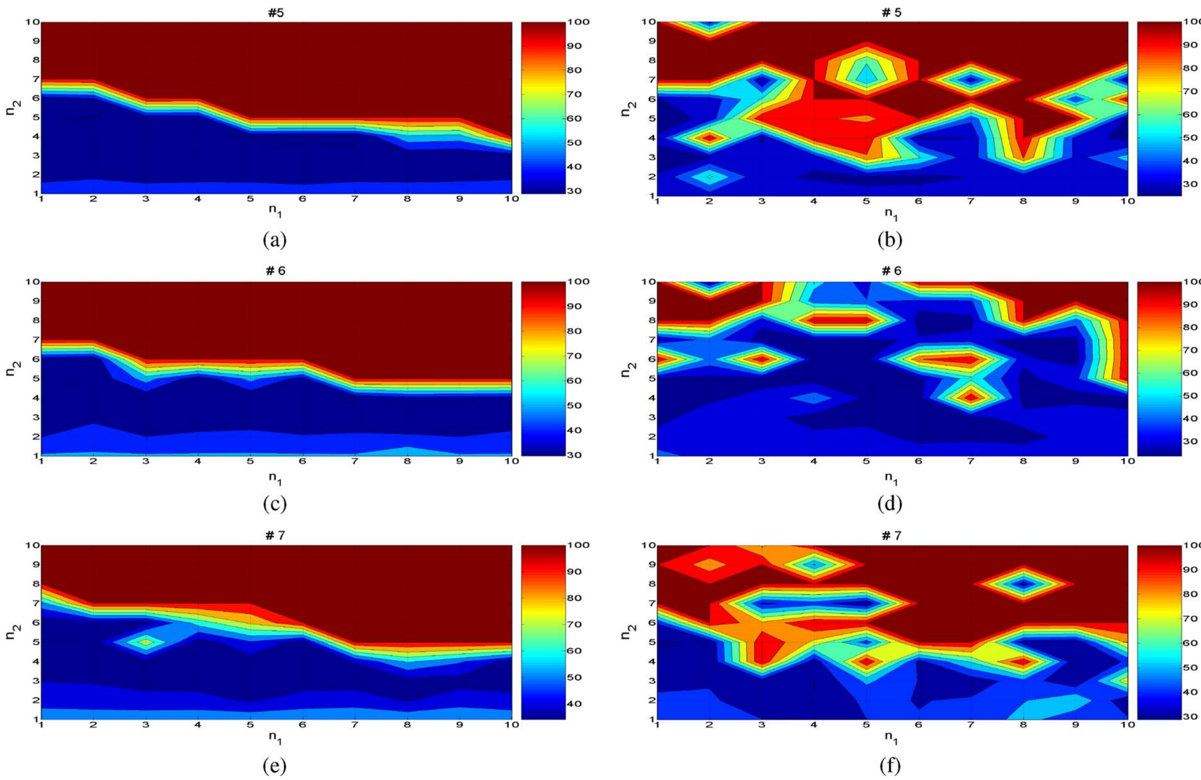

Fig. 5. Averaged CLCs on a grid composed of n_1 and n_2 for PIs constructed using the Bayesian (left side) and delta (right side) techniques for cases #5–#7. (a) Bayesian method for case #5. (b) Delta method for case #5. (c) Bayesian method for case #6. (d) Delta method for case #6. (e) Bayesian method for case #7. (f) Delta method for case #7.

TABLE II
QUANTITATIVE MEASURES FOR PI_B AND PI_D

Case	Bayesian Technique			Delta Technique		
	PICP(%)	NMPIL(%)	CLC	PICP(%)	NMPIL(%)	CLC
#1	91.4	43.1	43.3	90.5	35.2	35.4
#2	97.8	61.9	61.9	87.9	42.3	64.5
#3	98.7	60.3	60.3	87.7	41.8	75.1
#4	100	86.6	86.6	90.9	59.5	59.6
#5	89.6	27.7	28.9	89.2	24.1	24.8
#6	90.7	29.6	29.7	90.5	23.5	23.6
#7	90.5	33.8	34.0	89.2	27.9	28.8

(cases #1–#4) and Fig. 5 (cases #5–#7).[1] To simplify the referencing, we refer to the constructed PIs using the Bayesian and delta techniques as PI_B and PI_D, respectively. In addition, indices/measures related to them will be referred in the same manner (e.g., CLC_B and CLC_D). In addition, the best results (in terms of smallness of CLC) obtained using each method are summarized in Table II.

The obtained results can be analyzed and discussed from the following aspects.

1) Quality of PIs: As reported in Table II, PI_D's have a better quality than PI_B in five out of the seven cases investigated here (#1 and #4–#7). The judgment has been made based on the values of CLCs. For these cases, CLC_D is smaller than CLC_B. A very interesting point is that $PICP_B$ is always greater than $PICP_D$. In cases #2–#4, it is far beyond the nominal confidence level (90%). Such an excellent coverage probability has come with the huge cost of wide PIs. $NMPIL_D$ is always smaller than $NMPIL_B$, indicating the more narrowness of PI_D compared with PI_B. The maximum difference is for case #4, where PI_D is 27.08% narrower than PI_B. In terms of sensible time, this is a 166.54-s difference in length of the PIs. In average, PI_D's are 12.67% narrower than PI_B.

The reported results indicate that the delta method often suffers from low coverage probability. In four out of seven cases (#2, #3, #5, and #7), $PICP_D$ is (slightly) below the nominal confidence level (90%). In the other cases, the method just meets the PI construction requirements, and $PICP_D$ is equal to or greater than 90%.

A clear conclusion from the foregoing discussions is that the Bayesian method is more oriented toward producing PIs with a high coverage probability. This means that the variance term obtained in (12) is an overestimate of the actual variance of NN outcomes, leading to very wide PIs. In contrast, the NN outcome variance obtained using the delta technique appears to underestimate the actual variance. This results in PIs that are satisfactorily narrow, but the risk of their low coverage probability is high.

2) Sensitivity to the NN Structure and Training Process: The filled contour plots of CLCs shown in Figs. 4 and 5 provide insights into the behavior of PIs and their dependence on the NN structure. Regions with (dark) red color correspond to PI with unsatisfactorily low coverage probability. Contour plots for the experiments conducted using the Bayesian method include large continuous regions that CLCs remain almost stable within them. These regions can be found in different plots and for different colors. The uniform color of a region means

[1] With the purpose of better graphical representation, CLCs have been postprocessed, and outliers have been set to 100 (the maximum value of CLC in the case of using the extreme values of targets as upper and lower bounds of PIs).

that CLC remains unchanged within that region, regardless of the NN structure and complexity. This is an indication of the robustness of the quality of constructed PIs against the network complexity.

For cases #1–#4 (bus travel time data sets), the quality of PI_B has a direct relationship with the NN complexity. Such a relationship can be well observed for case #1. There are continuous regions stretched alongside each axis (number of neurons in the two hidden layers). As NNs become more complex, the quality of the constructed PI_B improves. The best results for this case are obtained for NNs that have at least seven neurons in each hidden layer. The same results are obtained with networks with similar complexity (e.g., five neurons in the first layer and ten neurons in the second layer). The same scenario happens in other cases as well.

PI_B shows a special quality pattern on the grid of NN structures for freeway travel times (cases #5–#7). When the network complexity goes beyond some specific values, the quality of PIs significantly drops. This drop includes all networks with excessive complexity. For this data set, the less the NN complexity, the better the PI quality. Therefore, ample care should be exercised when selecting the network structure to achieve the best possible results.

The quality of PI_D is highly sensitive to the network structure and its complexity. For these PIs, there are isolated regions in each plot that the CLCs within them are better than the neighborhood area. The CLC highly fluctuates on the grid of NN structures between deep valleys (appropriate structures) and sharp peaks (inappropriate structures). These valleys and peaks are more frequent for the freeway travel time data set (cases #5–#7). This is probably attributable to the number of NN inputs in two data sets. The more number of attributes results in more sophisticated relationships and nonlinear dependencies between NN outcome and its inputs. As the regions covering the appropriate structures are discontinuous and restricted, it is reasonable to conclude that CLC_D is highly sensitive to the network structure. These dependencies also have a direct relationship with the number of NN inputs.

3) Effects of NN Inputs: For the bus travel time data sets (#1–#4), the set of inputs is the same. Therefore, the comparison is not possible. For cases #5–#7, the set of inputs relates to the last 5-, 10-, and 15-min recorded speeds. According to the results shown in Table II, using the more recent speeds results in more quality PIs. PI_B's for case #5 are narrower than cases #6 and #7. In addition, PI_B's for case #7 are wider than case #6. The same scenario also happens for PI_D. The quality of PIs for cases #5 and #6 is almost the same and better than case #7.

4) PI Quality Relationship With Point Prediction Accuracy: The relationship can be sought through the comparison of results in Tables I and II. The obtained results indicate that there is a linear relationship between point prediction accuracy (measured by R^2) and the quality of PIs (measured by CLC). Fig. 6 displays scatter plots of CLC_B and CLC_D versus R^2. It is obvious from the plot that the bigger the R^2, the smaller the CLC_B and CLC_D, and the better the quality of the PIs. The correlation coefficients between CLC_B and R^2 and between CLC_D and R^2 are -0.92 and -0.87, respectively. Justification for this strong dependence can be made based on (20) for the delta technique and (12) for the Bayesian method. Both of these terms depend on the sum of squared errors defined in (8).

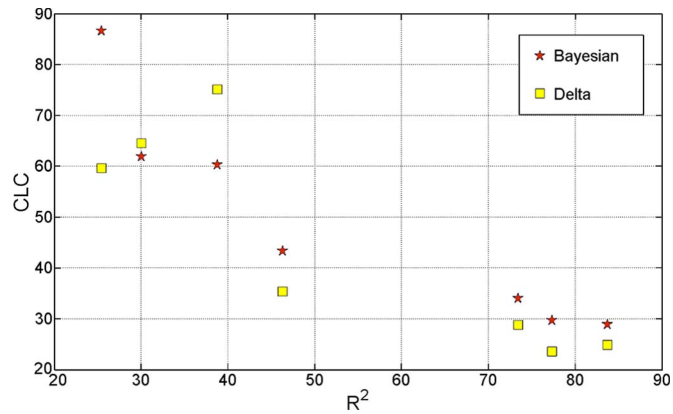

Fig. 6. Scatter plots of CLC_B and CLC_D versus R^2.

Such a strong dependence implicitly means that investigation of the error-based criteria, such as SSE or MAPE, should be part of the NN structure selection process. Although calculation of SSE is part of the process for the construction of PIs, its direct examination and assessment will positively contribute to the improvement of the PI quality.

For the first and fifth case study, some PI_B's with a confidence level of 90% have been demonstrated in Fig. 7. These PIs have been obtained from an NN with ten neurons in the first layer and five neurons in the second layer. The NN predicted values are in the center of each PI. As there is always a mismatch between predictions and targets, the reliability of the predictions is low. In contrast, the constructed PIs well cover the majority of the targets. The difference in the length of PIs is clear in the figures. As discussed before, such a difference is mainly due to the high level of uncertainty in the data sets.

C. Overall Comparison

The following can be concluded based on the demonstrated results and discussions in the previous subsections:

1) The delta method often leads to PIs whose lengths are significantly shorter than PIs constructed using the Bayesian method. However, the paid cost is a lower coverage probability that just meets the associated confidence level.
2) The PIs constructed using the Bayesian method have excellent coverage probability. In the majority of the cases investigated in this paper, $PICP_B$ is amply greater than the nominal confidence level. A wider PI_B is the cost we pay to achieve this goal.
3) The Bayesian method is more robust against the NN structure. There are many NN structures that the quality of PI_B associated with them are very similar. In contrast, the NN structures leading to the high-quality PI_D are in the format of isolated islands in the grid of NN structures.
4) In general, PIs constructed for the bus travel time data sets are significantly wider than PIs constructed for the freeway travel time data sets. This is mainly due to the presence of more uncertainties affecting the travel times of buses compared with the travel times of cars in the underlying freeway.

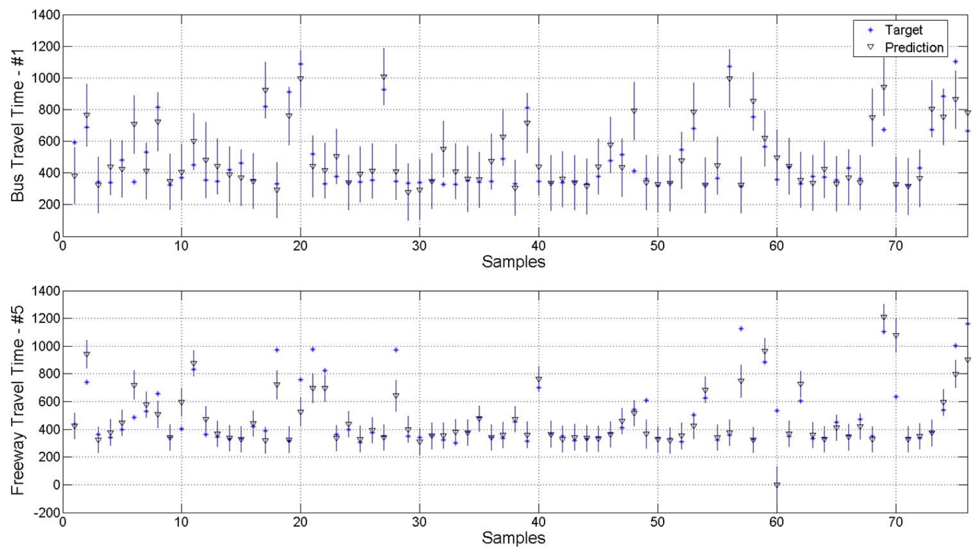

Fig. 7. Constructed PIs using the Bayesian techniques for case study #1 (top) and case study #5 (bottom) using an NN with ten neurons in the first layer and five neurons in the second layer.

V. Conclusion

Our research addressed the problem of bus and freeway travel time prediction using NN models. To appropriately accommodate uncertainties inherent to the prediction of bus and freeway travel times, PIs were developed using delta and Bayesian techniques. One of the contributions of this paper was a set of new performance measures to assess the quality of PIs. These new measures, i.e., PICP, NMPIL, and CLC, allow the analyst to balance between coverage probability (which favors wide but uninformative) prediction bounds and minimum interval width. These quantitative measures were applied for assessing the performance of PIs constructed using the Bayesian and delta techniques. It was observed that the constructed PIs using the delta technique are narrower than those developed using the Bayesian technique, while still providing a good coverage probability. In contrast, the constructed PIs using the Bayesian technique are more robust against the NN structure and training process. In both cases, the coverage probability of PIs was above the nominal confidence level (90%), which indicated the reliability of both methods for PI constructions.

The obtained results in this paper can be used in both practical and scientific studies in the transportation domain. From a practical point of view, it is possible to develop NN models (and in general any type of regression model) for the construction of PIs. Model selection can be done based on the proposed evaluation measures. These models can then be used in real time for operation planning and scheduling or implementing advanced transportation systems. From the scientific standpoint, research may go in the direction of how NNs can be developed and selected to result in PIs with the shortest length and highest coverage probability. Different methods can be developed and applied in many different areas, including the transportation field. Some preliminary results have been proposed in [34] and [40].

Acknowledgment

The authors would like to thank Ventura Bus Company and VicRoads for supplying the GPS and SCATS data, respectively. The data from the Dutch freeways were obtained from the TU Delft Regiolab data server (www.regiolab-delft.nl), whereas the travel times were kindly provided by Vialis NL.

References

[1] J. Bates, J. Polak, P. Jones, and A. Cook, "The valuation of reliability for personal travel," *Transp. Res. Part E: Logist. Transp. Rev.*, vol. 37, no. 2/3, pp. 191–229, Apr. 2001.

[2] R. Turochy and B. Smith, "Measuring variability in traffic conditions by using archived traffic data," *Transp. Res. Rec.*, vol. 1804, pp. 168–172, 2002.

[3] J. W. C. van Lint, "Online learning solutions for freeway travel time prediction," *IEEE Trans. Intell. Transp. Syst.*, vol. 9, no. 1, pp. 38–47, Mar. 2008.

[4] K. Zografos, K. Androutsopoulos, and V. Spitadakis, "Design and assessment of an online passenger information system for integrated multimodal trip planning," *IEEE Trans. Intell. Transp. Syst.*, vol. 10, no. 2, pp. 311–323, Jun. 2009.

[5] Y. Ma, M. Chowdhury, A. Sadek, and M. Jeihani, "Real-time highway traffic condition assessment framework using vehicle–infrastructure integration (VII) with artificial intelligence (AI)," *IEEE Trans. Intell. Transp. Syst.*, vol. 10, no. 4, pp. 615–627, Dec. 2009.

[6] M. Yang, Y. Liu, and Z. You, "The reliability of travel time forecasting," *IEEE Trans. Intell. Transp. Syst.*, vol. 11, no. 1, pp. 162–171, Mar. 2010.

[7] B. Bartin and K. Ozbay, "Determining the optimal configuration of highway routes for real-time traffic information: A case study," *IEEE Trans. Intell. Transp. Syst.*, vol. 11, no. 1, pp. 225–231, Mar. 2010.

[8] E. Osuna and G. Newell, "Control strategies for an idealized public transportation system," *Transp. Sci.*, vol. 6, no. 1, pp. 52–71, Feb. 1972.

[9] A. Ceder, *Public Transit Planning and Operation: Theory, Modelling and Practice*. Oxford, U.K.: Butterworth-Heinemann, 2007.

[10] C. Antoniou, M. Ben-Akiva, and H. Koutsopoulos, "Nonlinear Kalman filtering algorithms for on-line calibration of dynamic traffic assignment models," *IEEE Trans. Intell. Transp. Syst.*, vol. 8, no. 4, pp. 661–670, Dec. 2007.

[11] H. Jula, M. Dessouky, and P. Ioannou, "Real-time estimation of travel times along the arcs and arrival times at the nodes of dynamic stochastic networks," *IEEE Trans. Intell. Transp. Syst.*, vol. 9, no. 1, pp. 97–110, Mar. 2008.

[12] X. Zhang and J. A. Rice, "Short-term travel time prediction," *Transp. Res. Part C: Emerging Technol.*, vol. 11, no. 3/4, pp. 187–210, Jun. 2003.

[13] S. Robinson and J. Polak, "Modelling urban link travel time with inductive loop detector data by using the k-NN method," *Transp. Res. Rec.*, vol. 1935, pp. 47–56, 2005.

[14] J. W. C. van Lint, S. P. Hoogendoorn, and H. J. van Zuylen, "Accurate freeway travel time prediction with state-space neural networks under missing data," *Transp. Res. Part C: Emerging Technol.*, vol. 13, no. 5/6, pp. 347–369, Oct.–Dec. 2005.

[15] H. Liu, “Travel time prediction for urban networks,” Ph.D. dissertation, Delft Univ. Technol., Delft, The Netherlands, 2008.

[16] R. Jeong and R. Rilett, “Bus arrival time prediction using artificial neural network model,” in *Proc. 7th Int. IEEE Conf. Intell. Transp. Syst.*, 2004, pp. 988–993.

[17] M. Dougherty, “A review of neural networks applied to transport,” *Transp. Res. Part C: Emerging Technol.*, vol. 3, no. 4, pp. 247–260, Aug. 1995.

[18] E. Mazloumi, G. Currie, and G. Rose, “Using GPS data to gain insight into public transport travel time variability,” *J. Transp. Eng.*, vol. 136, no. 7, pp. 623–631, Jul. 2010.

[19] R. A. Kilmer, A. E. Smith, and L. J. Shuman, “Computing confidence intervals for stochastic simulation using neural network metamodels,” *Comput. Ind. Eng.*, vol. 36, no. 2, pp. 391–407, Apr. 1999.

[20] B. Efron, “Bootstrap methods: Another look at the jackknife,” *Ann. Statist.*, vol. 7, no. 1, pp. 1–26, Jan. 1979.

[21] T. Heskes, “Practical confidence and prediction intervals,” in *Proc. Neural Inf. Process. Syst.*, vol. 9. Cambridge, MA: MIT Press, 1997, pp. 176–182.

[22] C. M. Bishop, *Neural Networks for Pattern Recognition*. Oxford, U.K.: Oxford Univ. Press, 1995.

[23] D. J. C. MacKay, “The evidence framework applied to classification networks,” *Neural Comput.*, vol. 4, no. 5, pp. 720–736, Sep. 1992.

[24] J. T. G. Hwang and A. A. Ding, “Prediction intervals for artificial neural networks,” *J. Amer. Stat. Assoc.*, vol. 92, no. 438, pp. 748–757, Jun. 1997.

[25] R. D. de Veaux, J. Schumi, J. Schweinsberg, and L. H. Ungar, “Prediction intervals for neural networks via nonlinear regression,” *Technometrics*, vol. 40, no. 4, pp. 273–282, Nov. 1998.

[26] D. Nix and A. Weigend, “Estimating the mean and variance of the target probability distribution,” in *Proc. IEEE Int. Conf. Neural Netw., IEEE World Congr. Comput. Intell.*, 1994, vol. 1, pp. 55–60.

[27] E. Mazloumi, G. Rose, G. Currie, and S. Moridpour, “Prediction intervals to account for uncertainties in neural network predictions: Methodology and application in bus travel time prediction,” *Eng. Appl. Artif. Intell.*, 2010, to be published.

[28] I. Rivals and L. Personnaz, “Construction of confidence intervals for neural networks based on least squares estimation,” *Neural Netw.*, vol. 13, no. 4/5, pp. 463–484, Jun. 2000.

[29] R. Dybowski and S. Roberts, “Confidence intervals and prediction intervals for feed-forward neural networks,” in *Clinical Applications of Artificial Neural Networks*. Cambridge, MA: Cambridge Univ. Press, 2000.

[30] A. Ding and X. He, “Backpropagation of pseudo-errors: Neural networks that are adaptive to heterogeneous noise,” *IEEE Trans. Neural Netw.*, vol. 14, no. 2, pp. 253–262, Mar. 2003.

[31] A. Khosravi, S. Nahavandi, and D. Creighton, “A prediction interval-based approach to determine optimal structures of neural network metamodels,” *Expert Syst. Appl.*, vol. 37, no. 3, pp. 2377–2387, Mar. 2010.

[32] T. Lu and M. Viljanen, “Prediction of indoor temperature and relative humidity using neural network models: Model comparison,” *Neural Comput. Appl.*, vol. 18, no. 4, pp. 345–357, May 2009.

[33] Y. Jia and T. B. Culver, “Bootstrapped artificial neural networks for synthetic flow generation with a small data sample,” *J. Hydrol.*, vol. 331, no. 3/4, pp. 580–590, Dec. 2006.

[34] A. Khosravi, S. Nahavandi, and D. Creighton, “Construction of optimal prediction intervals for load forecasting problem,” *IEEE Trans. Power Syst.*, vol. 25, no. 3, pp. 1496–1503, Aug. 2010.

[35] H. Sun, C. Zhang, and B. Ran, “Interval prediction for traffic time series using local linear predictor,” in *Proc. 7th Int. IEEE Conf. Intell. Transp. Syst.*, 2004, pp. 410–415.

[36] C. P. I. van Hinsbergen, J. W. C. van Lint, and H. J. van Zuylen, “Bayesian committee of neural networks to predict travel times with confidence intervals,” *Transp. Res. Part C: Emerging Technol.*, vol. 17, no. 5, pp. 498–509, Oct. 2009.

[37] L. Zhang and S. Garcia-Munoz, “A comparison of different methods to estimate prediction uncertainty using partial least squares (PLS): A practitioner’s perspective,” *Chemometrics Intell. Lab. Syst.*, vol. 97, no. 2, pp. 152–158, Jul. 2009.

[38] J.-W. Wu, W.-C. Lee, and S.-C. Chen, “Computational comparison of the prediction intervals of future observation for three-parameter Pareto distribution with known shape parameter,” *Appl. Math. Comput.*, vol. 190, no. 1, pp. 150–178, Jul. 2007.

[39] E. Mazloumi, G. Currie, G. Rose, and M. Sarvi, “Using SCATS data to predict bus travel time,” in *Proc. 32nd ATRF*, Auckland, New Zealand, 2009.

[40] A. Khosravi, S. Nahavandi, D. Creighton, and A. F. Atiya, “Lower upper bound estimation method for construction of neural network-based prediction intervals,” *IEEE Trans. Neural Netw.*, Dec. 23, 2010, DOI: 10.1109/TNN.2010.2096824.

Abbas Khosravi (M’07) received the B.Sc. degree in electrical engineering from Sharif University of Technology, Tehran, Iran, in 2002 and the M.Sc. degree (hons.) in electrical engineering from Amirkabir University of Technology, Tehran, in 2005. His major was control with specialization on artificial intelligence (AI).

In 2006, he was with the eXiT Group, University of Girona, Girona, Spain, where he as a research academic in the area of AI applications. He is currently a Research Fellow with the Centre for Intelligent Systems Research, Deakin University, Geelong, Australia. His primary research interests include the development and application of AI techniques for the (meta)modeling, analysis, control, and optimization of operations within complex systems.

Ehsan Mazloumi received the B.Sc. degree in civil engineering and the M.Sc. degree in transportation engineering from Sharif University of Technology, Tehran, Iran, in 2001 and 2003, respectively. Since 2007, he has been pursuing the Ph.D. degree with the Institute of Transport Studies, Monash University, Melbourne, Australia.

He was with the Institute for Transportation Studies and Research, Sharif University of Technology, where his focus was on network design problems and transport network modeling. His main research interests include transport modeling, network design, application of artificial intelligence in transportation, and modeling public transport systems.

Saeid Nahavandi (SM’07) received the B.Sc. (hons.), M.Sc., and Ph.D. degrees in automation and control from Durham University, Durham, U.K.

He is the Alfred Deakin Professor, Chair of Engineering, and the Director of the Center for Intelligent Systems Research, Deakin University, Geelong, Australia. He has published over 350 peer-reviewed papers in various international journals and conferences. He designed the world’s first 3-D interactive surface/motion controller. His research interests include the modeling of complex systems, simulation-based optimization, robotics, haptics, and augmented reality.

Dr. Nahavandi is a Fellow of Engineers Australia and IET. He was the recipient of the Young Engineer of the Year Title in 1996 and has received six international awards in engineering. He is the Associate Editor of the IEEE SYSTEMS JOURNAL, an Editorial Consultant Board Member for the International Journal of Advanced Robotic Systems, and an Editor (South Pacific Region) of the *International Journal of Intelligent Automation and Soft Computing*.

Doug Creighton (M’10) received the B.Eng. (Honours) degree in systems engineering and the B.Sc. degree in physics from the Australian National University, Acton, Australia, in 1997 and the Ph.D. degree in simulation-based optimization from Deakin University, Geelong, Australia, in 2004.

He spent several years as a Software Consultant. He is currently a Research Academic and stream leader with the Centre for Intelligent Systems Research, Deakin University. He has actively been involved in modeling, discrete-event simulation, intelligent agent technologies, human-machine interfaces, and visualization- and simulation-based optimization research.

J. W. C. (Hans) van Lint (M'08) received the M.Sc. degree in civil engineering in 1997 and the Ph.D. degree in traffic modeling and prediction from Delft University of Technology (TU Delft), Delft, The Netherlands, in 2004.

He was a Software Engineer, a Consultant, and a Project Coordinator with the Rotterdam Regional Traffic Information Centre, Rotterdam, The Netherlands. Since 2004, he has been an Assistant Professor with the Transport and Planning Department, Civil Engineering, TU Delft, and was promoted and tenured as Associate Professor in 2009, teaching various subjects both in the B.Sc. and M.Sc. of civil engineering. He acquired and managed several large interdisciplinary research projects in the field of traffic monitoring, traffic estimation, and prediction ($>$ 3M€). He supervises on average six Ph.D. students and has already copromoted two Ph.Ds. He received a personal four-year (Dutch) NSF Grant in 2007 ("reversing the traffic jam") and has (co)authored 18 Institute for Scientific Information (ISI)-rated journal articles, two book parts, and over 70 peer-reviewed international conference papers.

Dr. van Lint is a member of the U.S. Transportation Research Board (TRB) committees on traffic flow theory and simulation, the TRB Committee on Artificial Intelligence, and the Intelligent Transportation Systems (ITS) Society of IEEE and is the Chair of the 2010 IEEE ITS Conference.