IET Intelligent Transport Systems

Research Article

IET Journals
The Institution of Engineering and Technology

Comparative analysis of implicit models for real-time short-term traffic predictions

ISSN 1751-956X
Received on 17th July 2015
Revised on 7th January 2016
Accepted on 15th February 2016
doi: 10.1049/iet-its.2015.0136
www.ietdl.org

Gaetano Fusco, Chiara Colombaroni, Natalia Isaenko ✉

Department of Civil, Environmental and Constructional Engineering, Sapienza University of Rome, via Eudossiana 18, Rome, Italy
✉ *E-mail: natalia.isaenko@uniroma1.it*

Abstract: Predicting future traffic conditions in real-time is a crucial issue for applications of intelligent transportation systems devoted to traffic management and traveller information. The increasing number of connected vehicles equipped with locating technologies provides a ubiquitous updated source of information on the whole network. This offers great opportunities for developing data-driven models that extrapolate short-term future trend directly from data without modelling traffic phenomenon explicitly. Among several different approaches to implicit modelling, machine-learning models based on a network structure are expected to be more suitable to catch traffic phenomenon because of their capability to account for spatial correlations existing between traffic measures taken on different elements of the road network. The study analyses and applies different implicit models for short-term prediction on a large road network: namely, time-dependent artificial neural networks and Bayesian networks. These models are validated and compared by exploiting a large database of link speeds recorded on the metropolitan area of Rome during seven months.

1 Introduction

The huge increase of traffic data availability is producing an overwhelming profusion of experimental studies and a continuous development of theoretical models for traffic state estimation and prediction. In addition to existing traditional detectors that provide traffic counts and aggregate speed estimates at fixed monitoring stations, individual detection is supplied by identification technologies such as video sensors for plate recognition, radio-frequency identification or Bluetooth. Algorithms for optimal location of monitoring stations and simulation models were developed to derive some estimate on the traffic on the whole network from point measures [1]. However, such estimates are usually affected by a high degree of uncertainty. Personal mobile devices embedding locating technology provide an additional source of ubiquitous information, which make possible to monitor different road segments on the network by periodic sampling of probe vehicles. Unlike traditional traffic monitoring technologies, probe vehicles provide sample estimates of both segment speeds and individual travel times. So novel and different sources of information offer additional opportunities for new traffic models and prediction methods. However, they pose new problems concerning the effect of statistical significance of measures and need introducing new indicators to assess the reliability of estimates. A second open issue is the capability of implicit models, which derive future values from observed trend of traffic variables, to provide accurate predictions of non-recurrent congestion by exploiting spatial correlation existing among measures collected on different segments on the network, other than their observed time series.

2 Related work

Short-term traffic predictions are object of a huge literature, whose comprehensive up-to-date reviews have been recently published [2, 3]. To provide a general overview of the state of the art, we mention here the main different approaches in the literature but we limit to quote just the earliest references for each of them, to the best of our knowledge. Instead, we reserve more space to discuss the few papers that have recently introduced the Bayesian framework, which is the main focus of this paper.

Data-driven methods were applied since ‘70s for short-term traffic predictions on single segments and later on road networks [4]. Thereafter, many different techniques with increasing levels of complexity were designed to exploit time and spatial correlation among sequential traffic variables observations: time-series approaches such as AutoRegressive Integrated Moving Average (ARIMA) [5] and seasonal ARIMA [6]; state identification methods such as: simple cluster analysis [7], *k*-neighbouring [8, 9], spectral analysis [10], neural network (NN) classifiers [11, 12], support vector regression [13, 14]; direct applications of NNs to short-term traffic predictions: namely, feed-forward (FF) [15, 16], time-recurrent [17], state-space NNs [18]; and Bayesian networks (BNs) [19]. Several authors proposed combinations of different methods. Bayes paradigm, which merges a priori information with experimental results, provides a rational framework for the combination purpose and is often used for short-term traffic prediction. Zheng *et al.* [20] combined linearly two single predictors – back propagation and radial basis function NNs – into a Bayesian combined NN model. Van Hinsbergen *et al.* [21] introduced the concept of a Bayesian committee, which combines the predictions of multiple NNs with different structures and different weight distributions. Khan [22] proposed a Bayesian approach to combine estimates obtained by naïve estimation from real-time detected data and estimates provided by even different models. Wang *et al.* [23] modified the Bayesian combination method by introducing a dynamic computation of credits and incorporated three single predictors. Antoniou *et al.* [24] proposed a two-step methodology, consisting of a Bayesian clustering technique for traffic state identification and a subsequent application of a state-specific function to estimate/predict the corresponding speed. In a recent work, Tselentis *et al.* [25] compared combinatorial approaches with traditional single time-series modelling.

BNs model correlation among measures by graph theory. Sun *et al.* [19] noted that most of state-of-the art data-driven methods often neglect information from adjacent roads to analyse the trends of the object road. Thus, they proposed the use of a BN to take into account the causal relationship between random variables statistically and model traffic flows among upstream and downstream road segments. Since BNs allow inferring the value of an object node by its neighbour nodes, the message passing mechanism of BNs can be exploited to make forecasting from incomplete data. Yu and Cho [26] included neighbour upstream and downstream segments into the BN. They assumed a Gaussian mixture model to approximate the joint probability distribution in the BN and applied the expected maximisation (EM) algorithm to train the model. Queen and Albers [27] introduced external intervention in the context of BNs to identify causal relationships between variables and in dynamic BNs to identify lagged causal relationships between time series. Pascale and Nicoli [28] proposed an adaptive BN method that selects the graph structure based on the local phase detected through a mutual information learning procedure. This method reduces complexity of the model but reduces also prediction accuracy, as authors experienced when applied their method to predict traffic flows on a freeway stretch. Hofleitner *et al.* [29] developed a graphical model connecting travel times with congestion state of each road segment and a traffic theoretical model that reproduces the distribution of delay within a road segment. The two models were combined into a dynamic BN, which demonstrated to highly exceed performances provided by a time-series model for travel time estimation. The main advantage of BNs is to combine Bayesian approach to posterior probability with network structure, which makes possible to apply graph algorithms to estimate existing correlations among variables. Traffic prediction is a particularly favourable case, since the road network topology provides direct information on the structure of BN that reflects correlations among traffic variables.

Our contribution focuses on opportunities and issues given by big data collected from floating cars. We study the most suitable structure of BNs and NNs to exploit such data. With respect to previous works, the BN architecture we introduce is closer to Bayes approach. It assumes the a priori estimate be independent of current measures and derives it from patterns obtained from analysis of historical data. Since current measures from floating cars are often unreliable or are lacking at all, a multi-level structure is envisaged to get information from farther segments when data from adjacent segments are defective. Moreover, a multi-period structure is introduced to provide predictions on several time intervals in the future in a compact form.

Finally, we present a test application on a large portion of the road network of Rome, Italy, where a big dataset from floating car data allows a systematic investigation of the most suitable BN structure for segment speed estimates. This BN is then compared with two different state-of-the-art architectures of artificial NNs: FF NN, as previously implemented in [15, 16] and non-linear auto-regressive (NARX) model with exogenous inputs [17], other than a simple naïve method and statistical estimate.

3 Methodology

3.1 FF neural network

FF is a static non-linear vector multivariate function that relates future values of speed on an output segment $\tilde{\boldsymbol{v}}\ (t+1, t+2, \ldots, t+h) = \{v_1(t+1), v_1(t+2), \ldots, v_1(t+h)\}$ within a time horizon h to the observed speed $\boldsymbol{u}(t, t-1, \ldots, t-k) = \{v_1(t), v_1(t-1), \ldots, v_1(t-k), \ldots, v_m(t), v_m(t-1), \ldots v_m(t-k)\}$ detected in k previous time intervals on segments 1, 2, …, m, including the output one

$$\tilde{\boldsymbol{v}}\ (t+1, t+2, \ldots, t+h) = \boldsymbol{f}_c(\boldsymbol{Cz} + \boldsymbol{\vartheta}_C)$$
$$\boldsymbol{z} = \boldsymbol{f}_c(\boldsymbol{Bu}(t, t-1, \ldots, t-k) + \boldsymbol{\vartheta}_D)$$

where t is the current time interval, k is the number of previous time intervals taken into consideration, $\boldsymbol{u}$ is the input vector, $\boldsymbol{z}$ is a vector representing the output of the hidden layer, $\boldsymbol{f}_c$ is a non-linear activation function and $\boldsymbol{\vartheta}_C$ and $\boldsymbol{\vartheta}_D$ are threshold values associated to the output and hidden layers, respectively. The coefficient matrices, $\boldsymbol{B}$ and $\boldsymbol{C}$ must be estimated during the training phase.

3.2 NARX NN with eXogenous inputs

NARX is a recurrent NN that relates the future values of speed on the output segment $v_1(t+1)$ to previous values of traffic variables on the same segment $\{v_1(t), v_1(t-1), \ldots, v_1(t-k)\}$ and on other segments $\boldsymbol{x}(t) = \{v_2(t), v_3(t), \ldots, v_m(t)\}$, $\boldsymbol{x}(t-1) = \{v_2(t-1), v_3(t-1), \ldots, v_m(t-1)\}$, …, $\boldsymbol{x}(t-k) = \{v_2(t-k), v_3(t-k), \ldots, v_m(t-k)\}$, which represent the exogenous inputs, through tapped delay line connections that provide delayed values of speed to be used in short-term predictions

$$\tilde{v}(t+1) = f(v_1(t), v_1(t-1), \ldots, v_1(t-k), \boldsymbol{x}(t), \boldsymbol{x}(t-1), \ldots, \boldsymbol{x}(t-k))$$

with the same meaning of the symbols and f denoting a non-linear transfer function.

3.3 Bayesian networks

BNs are probabilistic graphical models. This definition outlines the two components of a BN: a graphical component, represented by a directed acyclic graph and a probabilistic component, expressed by probability distributions. Each node of the graph represents a random variable, whereas each link represents a probabilistic dependency between the random variables corresponding to the nodes connected by the link. BN provides the joint probability density function p of future values of speed on the output road segment $v_1(t+1)$, $v_1\ (t+2)$, …, $v_1(t+h)$ conditioned by their a priori estimate p^0, the previous traffic states $\boldsymbol{x}(t, t-1, \ldots, t-k)$ on the estimated segment and on the previous traffic states on the conditioning segments, whose nodes of the BN are connected with the output nodes. The set B of connected nodes, called parents, can be seen as time dependent, because when measures on a parent road segment are defective, the joint probability is estimated by using its parent values. By assuming that the probability density functions of parents are normally and independently distributed and applying the chain rule, the density function of the predicted speed is

$$p_{V_1}\left(\tilde{v}_1(t+1), \ \ldots, \ \tilde{v}_1(t+h)\right) = \prod_{i=0}^{h} p^0_{V_1}\left(v_1(t+i)\right) \times \prod_{l \in B(t)} \prod_{j=0}^{k} p_{v_l}\left(v_l(t-j)\right)$$

3.4 Remarks

The expected advantage of prediction methods based on network architecture such as NNs and BNs is that their graph structure should have the capability to catch the time-dependent spatial correlation of traffic states on the road network. Several architectural specifications shall be preliminary studied and tested to represent the time–space correlation among different segments on the road network. Both upstream and downstream segments of the prediction segment should be included to take into account both forward flow progression that occur in light traffic and spillback progression that arises in congested conditions.

4 Application to the case study

A case study is presented that exploits a large database consisting of speed estimates from individual floating car data supplied by TomTom within a research project. Speed estimates were aggregated every 5 min. A confidence factor was provided that expresses an informal degree of belief in the reliability of the estimated values of the average segment speed.

For both NN and BN models, the training phase was carried out on about the 70% of the dataset. The remaining data was used for the validation.

The following measures of errors are applied for evaluating models accuracy

mean absolute error (MAE): $\sum_{i=1}^{n}\left(|\tilde{v}_i - v_i|/n\right)$,
mean absolute percentage error (MAPE): $\left(100\%/n\right)\sum_{i=1}^{n}\left(|\tilde{v}_i - v_i|/v_i\right)$,
root mean square error (RMSE): $\sqrt{\sum_{i=1}^{n}\left((\tilde{v}_i - v_i)^2/n\right)}$ and
root mean square percentage error (RMSEP)

Fig. 1 *Segments selected for NN architecture selection*

$$100\%\sqrt{\frac{\sum_{i=1}^{n}\left(\left(\tilde{v}_i - v_i/v_i\right)\right)^2}{n}},$$

where $\tilde{v}_i$ is the forecast, v_i is the observed value at time i and n is the size of observation set.

4.1 Training and model selection

Model selection consists of choosing the most appropriate architecture and its correspondence to variable specification: that is, for NNs, defining the number of layers, nodes, connections and the length of past and prediction horizon. Different tests were performed by exploring the neighbourhood of the target road segment and including different numbers of upstream and downstream segments, on the basis of correlations observed between target segment and surrounding segments. The best trade-off between complexity of the architecture and accuracy of predictions was provided by a simple structure composed by the segment itself, the backward star and the forward star. This structure has also two great practical advantages: it is modular and can be easily implemented on large networks through simple automatic routines that explore the road graph and select the forward star of end node and the backward star of initial node for each prediction segment. Architecture with one hidden layer was chosen for both FF and NARX, since it is demonstrated that such architecture can approximate any non-linear function.

4.2 NARX neural networks

Four alternative architectures were tested on one target segment, labelled 55,932 in Fig. 1, by considering different inputs corresponding to different correlated segments, other than the speed profile of the same segment and a number representing the weekday:

Architecture 0: average speed on segment 55,932, speed profile on segment 55,932, number of weekday;
Architecture 1: as Architecture 0 and average speed on upstream segment 54,295;
Architecture 2: as Architecture 0 and average speed backward star segments 54,295, 239,947, 220,688; and
Architecture 3: as Architecture 0 and average speed on backward star segments 54,295, 239,947, 220,688 and on forward star segment 30,915.

Each model was trained on the set of data collected in March, April and May 2014. Since NN training is an NP-complete process, the result of training depends on the specific algorithm used. Classical Levenberg–Marquardt and Levenberg–Marquardt with Bayesian regularisation algorithms were applied and compared. Performances of the training process of NARX with different numbers of hidden neurons, carried out by the two training algorithms, are shown in Fig. 2.

It is evident from this figure that Bayesian regularisation realises an approximately linear trend of the mean square error function, which decreases as the number of hidden layers increases and outperforms Levenberg–Marquardt. This figure underlines also a significant distinction between different network architectures. Specifically, including the forward and backward star improves the mean square error up to 4 km^2/h^2 (~20%) with respect to past information on the target segment only. Additional tests on positive signs of outputs, convergence of training algorithm, no cross-correlation between inputs and errors and autocorrelation of inputs revealed that architectures with more than 22 hidden neurons do not fulfil all tests.

A second step of model selection is the definition of time intervals in the past horizon. Different NARX NNs, structured according to Architecture 3 defined above, were trained assuming different combinations of the number of hidden neurons and time intervals. Tests results show (Fig. 3) that there is not a strict decrease of errors when the number of previous time intervals increases, for a given number of hidden neurons. Since NARX with 16 hidden neurons did not comply additional tests on correlation and convergence, the selected architecture has 12 hidden neurons and input nodes corresponding to speeds observed for six previous time intervals on road segments selected in Architecture 3 (target segment, forward and backward star).

4.3 FF neural networks

While NARX are time-dependent recursive networks, FF has a static structure. Predictions on different future time periods can be

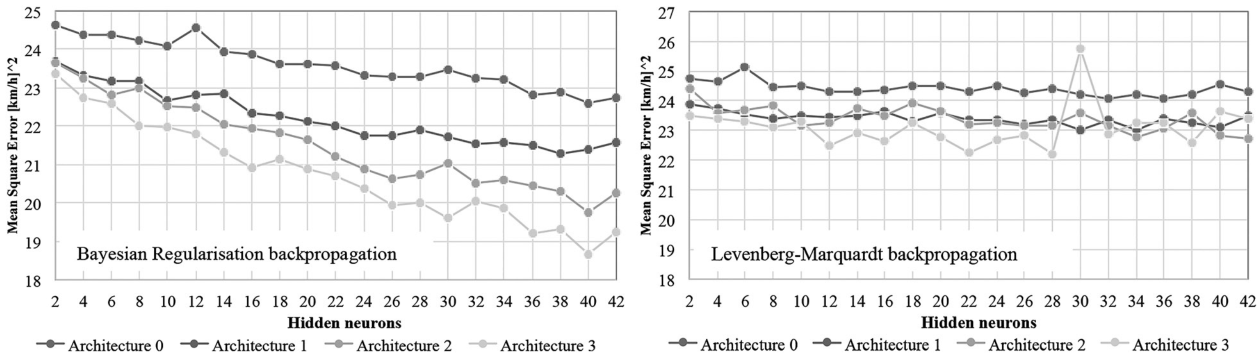


Fig. 2 *Mean square errors resulting after the training phase of four different NN architectures versus the number of hidden neurons, for two different training methods*

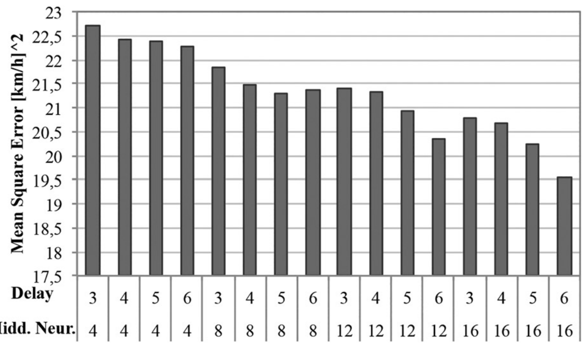
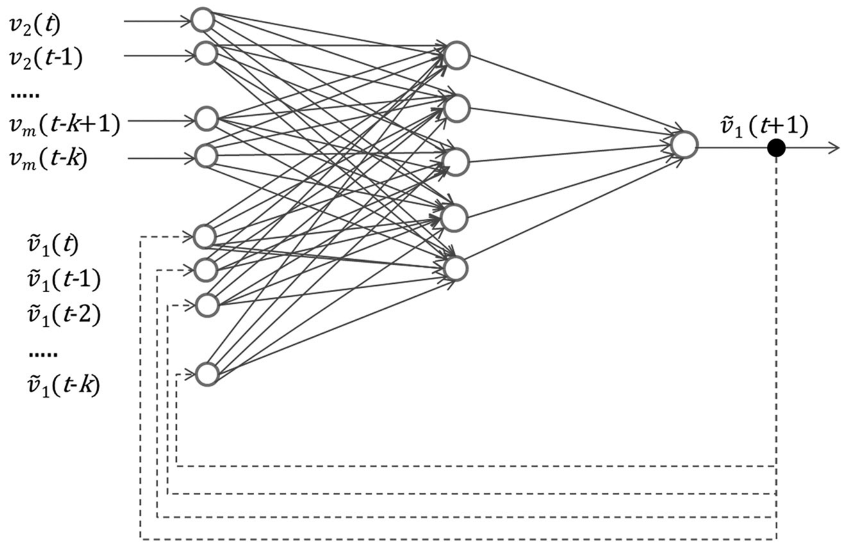


Fig. 3 *Mean square errors resulting after training of NARX NNs for different numbers of hidden neurons and different delay time intervals*

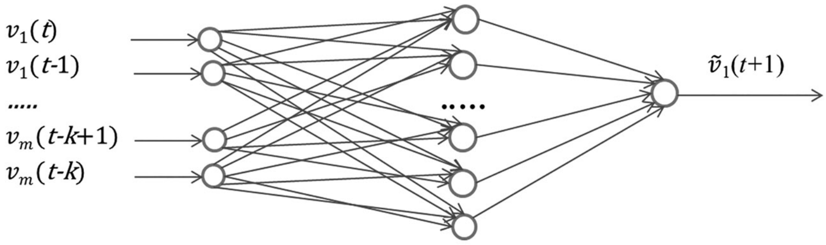
performed by applying two different methods. In the first (denoted in the following simply as FF), a different FFNN is trained for each future time interval which the forecast is related to; in the second (denoted in the following as FF'), a unique FFNN is trained to provide forecasts on all time intervals composing the prediction horizon simultaneously. The two alternative models FF and FF' have been trained and tested against the same time horizon, assuming the same architecture of input and hidden neurons found to be optimal for the NARX network. Specifically, five FFNNs were trained to provide single predictions for each of the time intervals 5; 15; 30; 45 and 60 min; one FF' was trained to provide multiple predictions for each period of the prediction interval simultaneously. Levenberg–Marquardt algorithm was used for training in both cases. Selected architectures are shown in Fig. 4. Results of the comparative training highlighted that the two methods have very similar performances. Thus, they were both selected for the validation phase.

4.4 Bayesian network

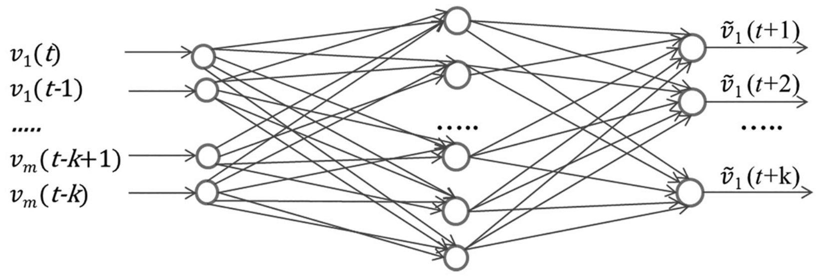
The selection of the best BN architecture was performed by conducting an empirical analysis of performances of different networks, whose architectures describe existing correlations between



a



b



c

Fig. 4 *Selected architectures*

a NNs architectures: NARX
b One output FF FF
c Multiple output FF FF′

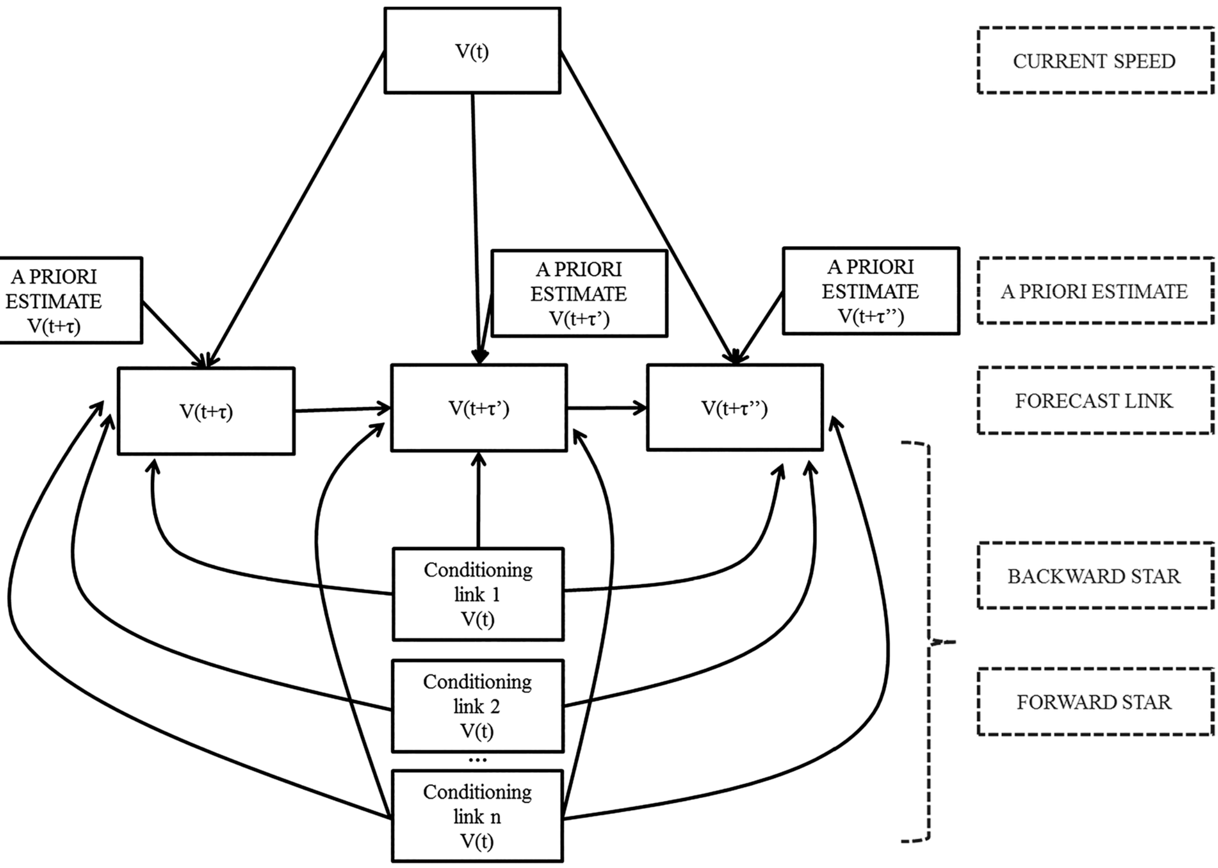


Fig. 5 *Multi-period BN structure with a priori estimate*

the speed on target segment and the speed previously observed on the same segment and on different surrounding segments.

A preliminary statistical analysis was performed to compute the average speed for every weekday in the training period, which was taken as a priori prediction.

Since the EM training algorithm of the BN can deal with incomplete data by exploiting existing correlations between other observed random variables, the database could be cleaned by removing measures having a low confidence factor, in order to train and then validate the model only against reliable data. Moreover, many model structures were investigated. Several upstream or downstream road segments were included in the BN structure and were connected among them consistently with the road network topology.

The analysis conducted by training such architectures did not point out significant differences among them. Architectures formed by many upstream or downstream segments provided only slight improvements of speed estimates with respect to the architecture formed by forward star and backward star segments.

On the other hand, with a view to the implementation of an automated procedure that builds one prediction model for each segment of the road network, architectures formed by many segments seem to be less suitable than that identified by the forward star and backward star, since they would require a more burdensome exploration of the graph.

Thus, the selected multi-period architecture whose nodes represent random variables V is shown in Fig. 5. It relates the future speed on the target segment with the current speed on road segments forming the forward star and backward star, other than the link itself and a priori estimates from historical data. Forecasts are provided on three future intervals simultaneously.

Table 1 Measures of errors for different NN models (complete dataset)

		RMSE	RMSEP	MAE	MAPE
5 min	FF	8.56	29.35	6.48	19.28
	FF′	8.60	29.55	6.57	19.52
	NARX	8.43	28.85	6.41	19.08
15 min	FF	10.54	39.71	9.18	28.00
	FF′	10.54	39.74	9.21	28.11
	NARX	10.29	38.87	8.74	26.96
30 min	FF	11.12	44.38	9.91	31.01
	FF′	11.13	44.47	9.93	31.10
	NARX	11.09	45.30	9.53	30.82
45 min	FF	11.51	48.11	10.32	33.06
	FF′	11.49	47.72	10.29	32.88
	NARX	11.53	49.74	9.95	33.03
60 min	FF	11.80	51.07	10.56	34.40
	FF′	11.81	51.08	10.57	34.45
	NARX	12.26	58.61	10.51	37.00

5 Validation and comparison of different models

Comparison of different models selected in the training and test phases was performed on a different set of data, collected in the period August–December 2014 on the segment having the opposite direction of that used for model selection. All selected models were trained on the dataset formed by observations collected in August, October, November and December and validated against the data gathered during the month of September.

The comparison was conducted in two phases, against two different databases: the complete database and the cleaned database, formed by only reliable speed measures.

The need for introducing two distinct phases comes from the different structure of the models. In fact, the NARX network is a recursive model that needs working on uninterrupted time series of data; however, BN trained by the EM algorithm is designed to deal with even incomplete data. Finally, FFNN is a static model that applies an input–output relation between a set of data and does not require outputs to be in a regular time sequence.

Table 2 Measures of errors for different prediction network models (clean dataset)

		RMSE	RMSEP, %	MAE	MAPE, %
5 min	historical Avg.	7.33	34.65	5.83	23.81
	naïve	10.12	39.47	6.91	24.86
	BN	6.13	25.69	4.69	18.03
	FF	5.6	23.83	4.22	16.16
15 min	historical Avg.	7.33	34.65	5.83	23.81
	naïve	13.76	55.75	10.64	39.54
	BN	6.85	31.50	5.44	21.83
	FF	6.55	30.30	5.17	20.71
30 min	historical Avg.	7.33	34.65	5.83	23.81
	naïve	14.95	63.65	11.93	24.86
	BN	7.07	33.56	5.62	22.93
	FF	7.01	34.33	5.59	23.03
45 min	historical Avg.	7.33	34.65	5.83	23.81
	naïve	15.36	67.58	12.36	47.82
	BN	7.16	34.36	5.72	23.52
	FF	7.38	37.55	5.88	24.82
60 min	historical Avg.	7.33	34.65	5.83	23.81
	naïve	15.91	71.01	12.93	50.81
	BN	7.24	34.89	5.77	23.80
	FF	7.66	40.10	6.12	26.23

5.1 Overall prediction performances on the complete dataset

Thus, a first comparison phase concerning the FF and NARX NNs was conducted on the complete database, both for training and validation phases. Results reported in Table 1 highlight that the NARX provides more accurate estimates of the speed on the target segment than FF for shorter time intervals, when a higher autocorrelation exists among data. However, FFNNs outperform NARX for predictions on time intervals farther than 30 min. Differences between the two models are anyway small; for instance, the difference of RMSE between NARX and FF ranges from −2 to 4%. The two FF and FF' models exhibit almost equivalent performances, with a slight outperformance of FF for shorter-term predictions, up to 15 min. The simpler FF model, formed by independent predictions for each time horizon, should be preferred.

5.2 Overall prediction performances on the clean dataset

The second comparison phase was carried out on the clean database formed only by speed measures with a high confidence factor. BN and FF architecture were trained and validated on only the reliable data collected during the month of September.

The two models were compared, other than with each other, with the statistical average speed profile computed for every weekday in training period and a naïve prediction method consisting on assuming the last observed value as forecast for the prediction time interval.

The following remarks can be drawn from Table 2:

(a) Naïve forecasts are always the worst method.
(b) BN always improves the statistic average, which represents the a priori estimate, though improvements reduce as the prediction time

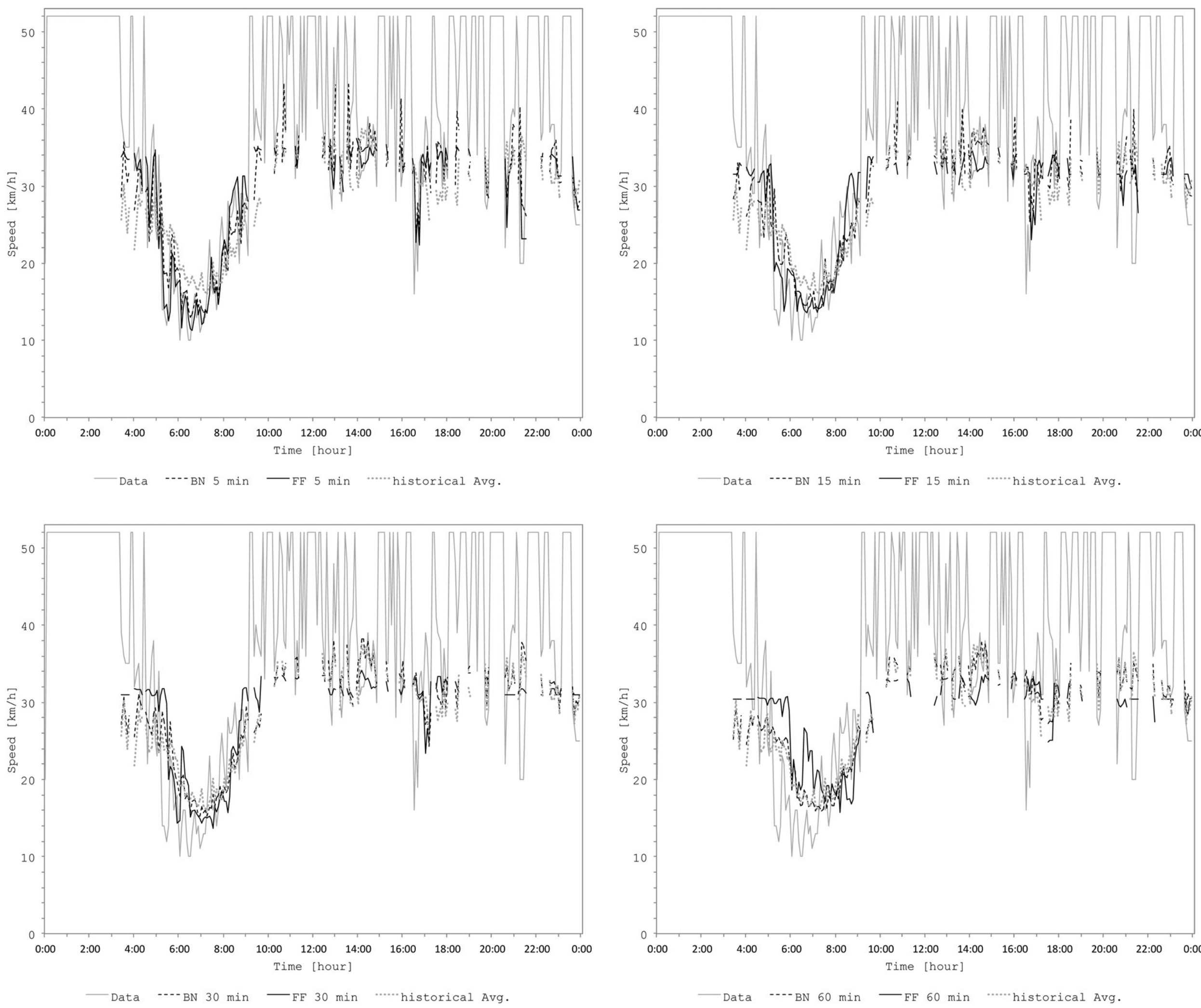


Fig. 6 *Observed values and speed forecasts for 5, 15, 30, 60 min intervals provided by FFNN, BN and historical average (historical Avg.) on via Cristoforo Colombo*

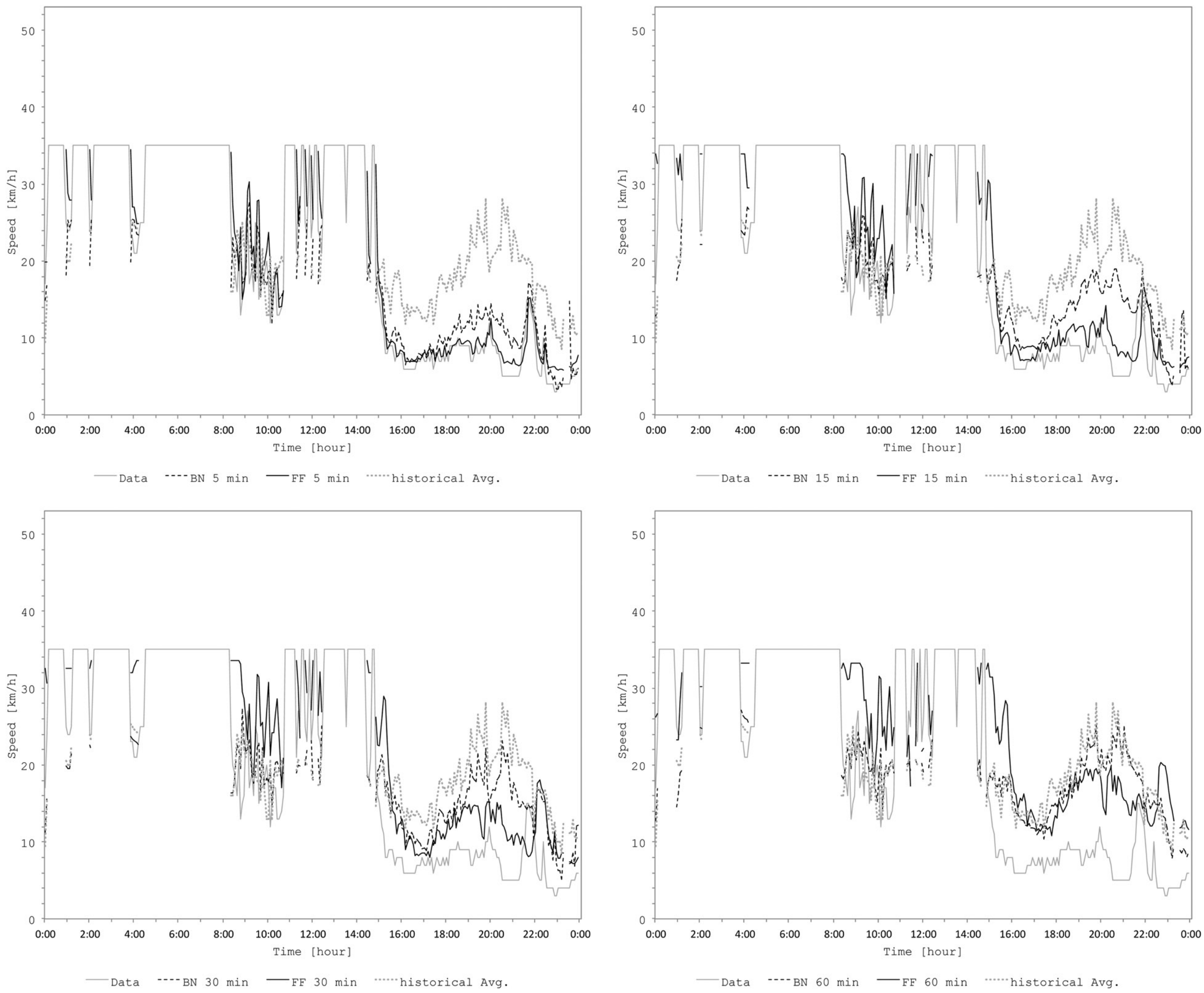


Fig. 7 *Observed values and speed forecasts for 5, 15, 30, 60 min intervals provided by FFNN, BN and historical average (historical Avg.) on Lungotevere Tor Di Nona*

interval becomes farther (for instance, RMSE for predictions based on the historical average is 7.33 km/h; RMSE for BN is 6.13 km/h for 5 min forecasts; and 7.24 km/h for 60 min forecasts).

FF performances are better than those of BNs for predictions up to 15 min (RMSE for 5 min forecasts is 5.60 km/h) but deteriorate for longer time intervals and become worst than historical average for 60 min predictions (RMSE for FFNN is 7.66 km/h).

It is worth noting that the errors experienced on the clean dataset are significantly lower than those measured on the complete dataset, which were biased by unreliable data.

5.3 Analysis of prediction performances under different traffic conditions

In addition to average error measures, it is worth investigating the performances of the various models in different traffic conditions. Some examples are provided in the following figures, which illustrate the observed values of speed on the target segment in two representative days and the corresponding forecast values provided by historical average, Bayesian network BN and feed-forward NN FF. Only forecasts corresponding to time intervals whose measures have a high confidence factor are shown. Fig. 6 refers to different prediction intervals (5; 15; 30 and 60 min) on the target segment on a typical workday.

Measured values have two very different patterns: in the peak hours (from about 4 a.m. to 9 a.m. Greenwich Mean Time, equivalent to 6 a.m. to 11 a.m., legal time in Central European Summer Time) measured speed exhibits an evident U-shaped trend, with frequent small oscillations; in the off-peak hours, it is characterised by frequent high oscillations from values close to the historical average and free-flow speed, whose values were inputted in the database when too few measures were detected in real-time.

Short-term forecasts by both models are very close to the observed values in the peak hours. As for 5 min predictions, FF provides an excellent approximation to data, which revealed a slightly higher congestion than the average for that workday. The BN forecasts provide an as good approximation as the FF at the beginning and at the end of peak period, when congestion is lighter; they overestimate the segment speed in the middle of the peak period and are closer than FF to the historical average, which was used as a priori estimate.

A priori knowledge is clearly advantageous for 60 min forecasts, where BN replays historical average, while FF is affected by an evident delay with respect to the true values.

Fig. 7 exemplifies an interesting case of non-recurrent congestion, occurred on Saturday afternoon on Lungotevere Tor di Nona, a segment belonging to a main road corridor in the historic Centre of Rome, which runs aside the embankment along the Tiber River. It is evident from this figure that the measured speeds from 3 p.m. to about 12 p.m. were much lower than the average values detected during the same time period on that road segment. Also, in this case FF provides good short-term forecasts for the next 5 min; BN forecasts exhibit a very similar shape to observed values but overestimate them slightly in the shorter-term. For 15 and 30 min forecasts, FF is still able to capture the anomalous conditions occurring in the afternoon while forecasts of the BN for 30 and 60 min approximate the historical average value: the influence of a

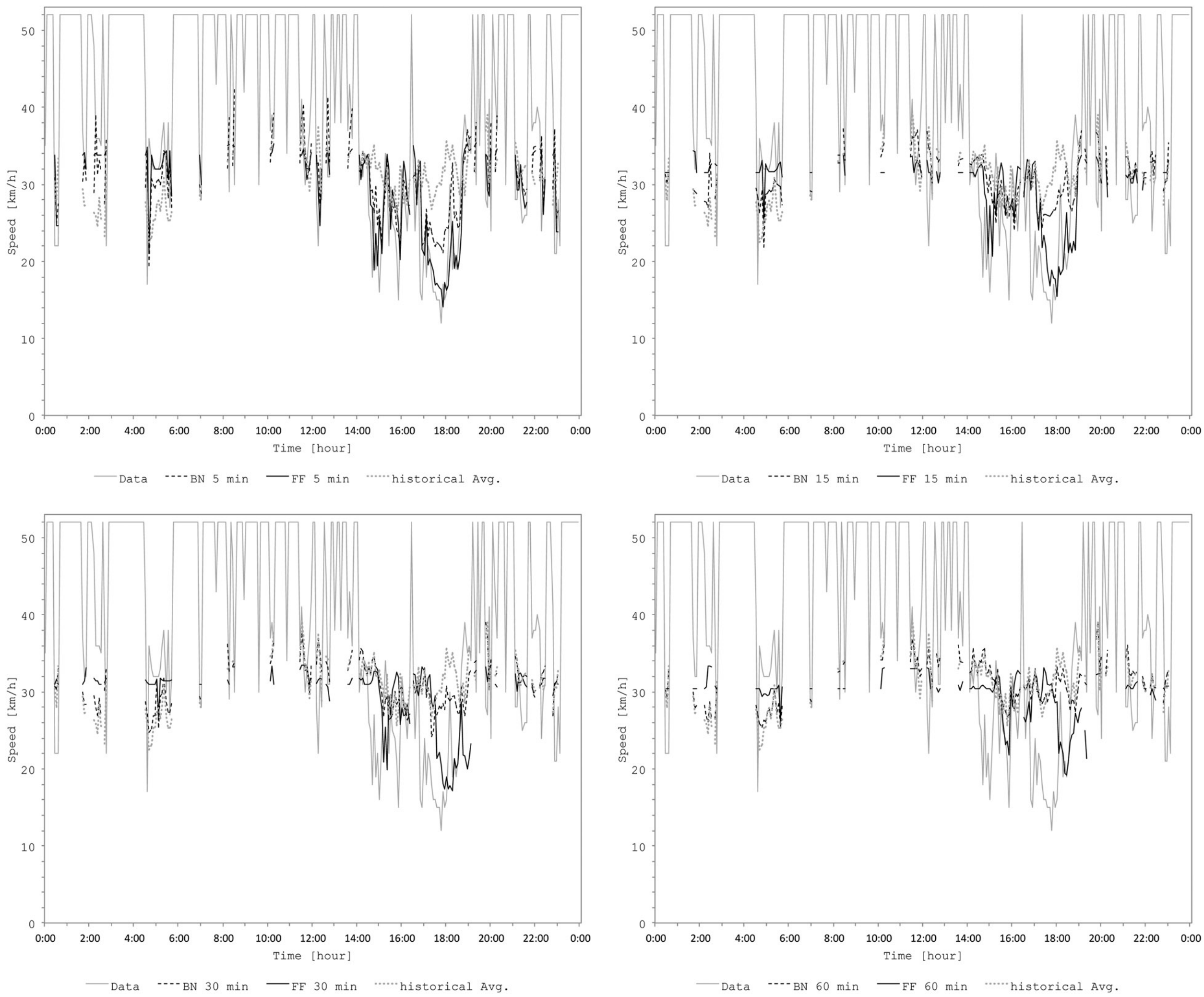


Fig. 8 *Observed values and speed forecasts for 5, 15, 30, 60 min intervals provided by FFNN, BN and historical average (historical Avg.) on via Cristoforo Colombo*

priori estimate becomes stronger for longer-term predictions as well as for anomalous conditions. It is worth noting the recurrent but unstable traffic conditions with marked oscillations occurring from 9 a.m. to 10 a.m. It can be seen that FF forecasts for this period are very noisy and overestimated for 15 and 30 min. This result can be explained by the fact that the FF forecasts are based only on the information concerning the previous traffic state; in case of significant fluctuations of input data the forecasts exhibit even increased fluctuations that reduce model robustness.

Another case of non-recurrent congestion is shown in Fig. 8. The observation refers to a portion of via Cristoforo Colombo, a main artery connecting the historic Centre of Rome to the seaboard. Non-recurrent congestion conditions occurring from 4:30 p.m to 6:30 p.m are well forecasted by both models for 5 min forecasting interval; however, for 15 and 30 min intervals only FF is able to forecast the actual trend of speed, while BN approaches the historical average profile (resulting MAE of about 10 km/h).

6 Conclusions

This paper presented a comparative analysis of different implicit short-term traffic prediction models, designed for online applications. A large dataset of average speeds detected through floating car data for ten months was used for tests and comparisons of the models.

NARX model with exogenous inputs NN and two models of FFNNs were introduced and compared. These models exhibited similar performances. The model consisting in one independent FFNN for each prediction time interval, which is the simplest among the three tested, can be seen as the most convenient for this case study. A BN that exploits a priori knowledge consisting in historical average outperforms the NN in medium-term forecasts (that is, from 30 to 60 min later), on average.

On the other hand, NNs provide more accurate forecasts in the short-term (that is, from 5 to 15 min later) and in case of non-recurrent conditions. So, NNs appear to be a convenient model for online applications of intelligent transportation system, which require high promptness to face sudden congestion occurrences and reliable predictions of anomalous conditions. However, BNs have to be seen anyway as a very powerful tool, which can be better exploited in database having a highly detailed level of information. The proposed FFNN model provides speed forecasts based only on the current traffic state condition which can reduce the model robustness in case on unstable traffic conditions. A contemporary study performed by the same authors (Fusco *et al.* [30]) on a different database composed by individual floating car data made it possible to include inter-vehicle variance of speeds into the model and to provide the BN with estimates of data reliability and uninterrupted time series. The qualitative difference in the performance of the two models suggests a further enhancement, which is object of the ongoing research, consisting in combining two models that exhibit complementary characteristics and introducing an additional model that acts as a supervisor. The supervisor should select the most appropriate forecast according to specific detected conditions or should merge different forecasts with suitable weights when they both correspond to reasonable conditions.

7 Acknowledgments

The research has been supported by Duel Company through a co-research project co-funded by Lazio Innova, the financing agency of the Italian Latium Region. Authors are grateful to Eng. Aurelio Arenella who performed NNs computations.

8 References

1 Cipriani, E., Fusco, G., Gori, S., *et al.*: 'Heuristic methods for the optimal location of road traffic monitoring stations'. IEEE Conf. on Intelligent Transportation Systems, pp. 1072–1077, art. no. 1707364
2 Vlahogianni, E.I., Karlaftis, M.G., Golias, J.C.: 'Short-term traffic forecasting: where we are and where we're going', *Transp. Res. C, Emerg. Technol.*, 2014, **43**, (1), pp. 3–19
3 Oh, S., Byon, Y.J., Jang, K., *et al.*: 'Short-term travel-time prediction on highway: a review of the data-driven approach', *Transp. Rev.*, 2015, **35**, (1), pp. 4–32
4 Stephanedes, Y., Michalopoulos, P., Plum, R.: 'Improved estimation of traffic flow for real time control', *Transp. Res. Rec.*, 1981, **795**, pp. 28–39
5 Ahmed, M.S., Cook, A.R.: 'Analysis of freeway traffic time-series data by using Box–Jenkins techniques', *Transp. Res. Rec.*, 1979, **722**, pp. 1–9
6 Williams, B.M., Hoel, L.A.: 'Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: theoretical basis and empirical results', *J. Trans. Eng.*, 2003, **129**, (6), pp. 664–672
7 Bellamy, P.H.: 'Seasonal variation in traffic flows', TRRL Report SR437, TRRL, Crowthorne, UK, 1978, pp. 1–17
8 Davis, G.A., Nihan, N.L.: 'Nonparametric regression and short-term freeway traffic forecasting', *J. Trans. Eng.*, 1991, **117**, (2), pp. 178–188
9 Smith, B., Demetsky, M.: 'Traffic flow forecasting: comparison of modeling approaches', *J. Trans. Eng.*, 1997, **123**, (4), pp. 261–266
10 Nicholson, H., Swann, C.D.: 'The prediction of traffic flow volumes based on spectral analysis', *Transp. Res.*, 1974, **8**, (6), pp. 533–538
11 Kirby, H.R., Parker, G.B.: 'The development of traffic and transport applications of artificial intelligence: an overview', in Bielli, M., Ambrosino, G., Boero, M. (Eds.): 'Artificial intelligence applications to traffic engineering' (VSP, Utrecht, 1994), pp. 3–27
12 Florio, L., Mussone, L.: 'Neural-network models for classification and forecasting of freeway traffic flow stability', *Control Eng. Pract.*, 1996, **4**, (2), pp. 153–164
13 Wu, C.H., Ho, J.M., Lee, D.T.: 'Travel-time prediction with support vector regression', *IEEE Trans. Intell. Transp. Syst.*, 2004, **5**, (4), pp. 276–281
14 Vlahogianni, E.I.: 'Optimization of traffic forecasting: intelligent surrogate modeling', *Transp. Res. C, Emerg. Technol.*, 2015, **55**, pp. 14–23
15 Fusco, G., Gori, S.: 'The use of artificial neural networks in advanced traveler information and traffic management systems'. Proc. Int. Conf. Applications of Advanced Technologies in Transportation Engineering, Capri, Italy, 27 June–30 June 1995, pp. 341–345
16 Dougherty Mark, S., Kirby, H.R., Boyle, R.D.: 'The use of neural networks to recognise and predict traffic congestion', *Traffic Eng. Control*, 1993, **34**, (6), pp. 311–314
17 Yun, S.Y., Namkoong, S., Rho, J.H., *et al.*: 'A performance evaluation of neural network models in traffic volume forecasting', *Math. Comput. Model.*, 1998, **27**, (9), pp. 293–310
18 Van Lint, J.W.C., Hoogendoorn, S.P., van Zuylen, H.J.: 'Accurate freeway travel time prediction with state-space neural networks under missing data', *Transp. Res. C, Emerg. Technol.*, 2005, **13**, (5), pp. 347–369
19 Sun, S., Zhang, C., Yu, G.: 'A Bayesian network approach to traffic flow forecasting', *IEEE Trans. Intell. Transp. Syst.*, 2006, **7**, (1), pp. 124–132
20 Zheng, W., Lee, D.H., Shi, Q.: 'Short-term freeway traffic flow prediction: Bayesian combined neural network approach', *J. Transp. Eng.*, 2006, **132**, (2), pp. 114–121
21 Van Hinsbergen, C.I., Van Lint, J.W.C., Van Zuylen, H.J.: 'Bayesian committee of neural networks to predict travel times with confidence intervals', *Transp. Res. C, Emerg. Technol.*, 2009, **17**, (5), pp. 498–509
22 Khan, A.M.: 'Bayesian predictive travel time methodology for advanced traveller information system', *J. Adv. Transp.*, 2012, **46**, (1), pp. 67–79
23 Wang, J., Deng, W., Guo, Y.: 'New Bayesian combination method for short-term traffic flow forecasting', *Transp. Res. C, Emerg. Technol.*, 2014, **43**, pp. 79–94
24 Antoniou, C., Koutsopoulos, H.N., Yannis, G.: 'Dynamic data-driven local traffic state estimation and prediction', *Transp. Res. C, Emerg. Technol.*, 2013, **34**, pp. 89–107
25 Tselentis, D.I., Vlahogianni, E.I., Karlaftis, M.G.: 'Improving short-term traffic forecasts: to combine models or not to combine?', *IET Intell. Transp. Syst.*, 2015, **9**, (2), pp. 193–201
26 Yu, Y.J., Cho, M.G.: 'A short-term prediction model for forecasting traffic information using Bayesian network'. Third Int. Conf. on Convergence and Hybrid Information Technology, 2008. ICCIT'08, 2008, pp. 242–247
27 Queen, C.M., Albers, C.J.: 'Intervention and causality: forecasting traffic flows using a dynamic Bayesian network', *J. Am. Stat. Assoc.*, 2009, **104**, (486), pp. 669–681
28 Pascale, A., Nicoli, M.: 'Adaptive Bayesian network for traffic flow prediction'. Statistical Signal Processing Workshop, 2011, pp. 177–180
29 Hofleitner, A., Herring, R., Abbeel, P., *et al.*: 'Learning the dynamics of arterial traffic from probe data using a dynamic Bayesian network', *IEEE Trans. Intell. Transp. Syst.*, 2012, **13**, (4), pp. 1679–1693
30 Fusco, G., Colombaroni, C., Comelli, L., *et al.*: 'Short-term traffic predictions on large urban traffic networks: applications of network-based machine learning models and dynamic traffic assignment models'. Proc. Int. Conf. Models and Technologies for Intelligent Transportation Systems (MT-ITS), Budapest, Hungary, 3–5 June 2015, doi: 10.1109/MTITS.2015. 7223242