Impact of Data Loss for Prediction of Traffic Flow on an Urban Road Using Neural Networks

Teresa Pamuła

Abstract—**The deployment of intelligent transport systems requires efficient means of assessing the traffic situation. This involves gathering real traffic data from the road network and predicting the evolution of traffic parameters, in many cases based on incomplete or false data from vehicle detectors. Traffic flows in the network follow spatiotemporal patterns and this characteristic is used to suppress the impact of missing or erroneous data. The application of multilayer perceptrons and deep learning networks using autoencoders for the prediction task is evaluated. Prediction sensitivity to false data is estimated using traffic data from an urban traffic network.**

Index Terms—**Deep learning, traffic flow prediction, sensitivity to loss of data.**

I. Introduction

THE progress of deployment of complex control and management systems, commonly called intelligent transportation systems (ITS), in transport networks continues spanning over larger commuting areas not only urban. The capability to predict the evolution of traffic situations is of utmost importance for efficient functioning of ITS. The prediction methods use data provided by various traffic monitoring systems of different reliability and vulnerability to traffic incidents. Besides real time data historical accounts of traffic patterns are used for forecasting traffic. All data sources exhibit imperfections of different weight on the overall prediction capability. Traditional inductive loop vehicle detectors break down and feed no data. Video based detectors during bad weather conditions give erratic values of traffic data. Historical data contains false entries due to bad storage. In order to diminish the influence of such instances of data loss, robust methods of prediction need to be developed.

A way to enhance the prediction performance is to take into account the relations between traffic monitoring stations in the road network. The mutual spatial placement of the measuring devices along traffic corridors retains relations between road traffic characteristics. Some positions may be very important to the functioning of the transport system and play a predominant role in the work out of traffic forecasts. These spatial relations need to be coupled with traffic data and used for elaboration of predictions. Further improvements require expansion of vehicle detection resources or introduction of redundancy in the traffic monitoring systems. This approach inevitably leads to the generation of vast amounts of data which may cause unpredictable difficulties in processing.

The important query is the strategy of substituting data from malfunctioning equipment or corrupt storage with data which gives the least impact on the prediction performance [15], [19].

Fusing of traffic data, road network parameters and vehicle detector states is predominantly the domain of non parametric traffic prediction methods especially based on artificial intelligence such as variants of neural networks.

The aim of this study is to investigate the sensitivity of neural networks to loss of input data for predicting traffic flows in a road network and to propose a strategy for substituting lost data for preserving the accuracy of prediction. Representative multilayer perceptron based NN (MLP) and a deep learning network based on autoencoders (DLN) is compared The networks differ in the way the training process is done. The training process of DLN consists of two steps: first the network is trained layer-by-layer, then the network parameters are tuned. MLP is trained in one step. The DLN training adopts more efficiently to training data.

The paper sections contain: a discussion of related works highlighting the features of different prediction approaches using neural networks, section 3 presents the methodology of the study centred on the spatiotemporal characteristics of acquired traffic data, next the results of prediction experiments using real traffic data are summarized, the final section draws conclusions and some topics for future works are proposed.

II. Related Works

The task of efficient prediction of traffic parameters is approached using groups of methods based on: linear models of traffic, filtering, non parametric representations and hybrid combinations of models [20], [26]. Hybrid methods pose a processing challenge so are rarely implemented.

The linear models assume a linear form of the traffic dynamics, among which are: time series, multivariate time series, autoregressive integrated moving average (ARIMA), seasonal ARIMA models [9].

A comparison study of univariate and multivariate techniques by Kamarianakis and Prastacos [6] shows that multivariate techniques can capture the relations between measurements at different locations and time. Prediction results are comparable with results obtained by univariate techniques e.g. ARIMA models when the number of sites is small and they are dispersed on the road network.

Manuscript received October 28, 2016; revised May 8, 2017, September 3, 2017, and May 6, 2018; accepted May 9, 2018. Date of publication May 31, 2018; date of current version February 28, 2019. The Associate Editor for this paper was D. Work.

The author is with the Faculty of Transport, Silesian University of Technology, 44-100 Gliwice, Poland (e-mail: teresa.pamula@polsl.pl).

Digital Object Identifier 10.1109/TITS.2018.2836141

Min and Wynter [4] propose a multivariate spatial-temporal autoregressive (MSTAR) model based on a Vector-ARMA which includes dependencies among data from neighbouring measurement sites. Good prediction accuracy of speed and traffic volume is achieved for different road categories over 5-min intervals for up to 1 h in advance.

Seasonal ARIMA models [2], [3] incorporate seasonal differencing to achieve a stationary transformation of road traffic data. The predictive performance of these models depends on the length of the seasonal difference. A weekly difference is appropriate for urban traffic which follows patterns of commuting to work.

Filtering based methods rely on recursive least square, least mean square, generalized linear model and Kalman filters. Kalman filters with a process variances adaptation mechanism [1] demonstrate good fitting capabilities to changeable traffic characteristics.

The problem of missing data for prediction using these methods is not addressed directly; attempts are made to substitute missing values with averages or factors [11]. An approach using spatio-temporal data relations is proposed in [17] where neighbouring links traffic patterns are used for calculation of kernel regression estimator for prediction of travel times.

Whitlock and Queen [31] adapt a Multiregression Dynamic Model [30] to accommodate missing data series and use Markov chain Monte Carlo methods to provide forecasts of the missing series. Muralidharan and Horowitz [29] present an automated imputation procedure based on identifying the parameters of a ramp flow model. Link-node cell transmission model, reduced to four mode model, is used. Update equations are designed which minimise the error between simulated and measured traffic data.

Non parametric representation offers the most adaptive models [10]. Traffic dynamics can be combined with network characteristics and historical patterns of behavior [18]. Methods based on neural networks stand out in this approach [8]. These enable learning of the characteristics of the traffic without prior rigorous bounds.

van Lint *et al.* [13] and Chen *et al.* [14] investigate imputation schemes to remedy false data inputs to neural networks (MLPs) for prediction of freeway travel time. Null replacement, sample mean or exponentially moving average substitution are tested. Null replacement relies on training the NN with data series with missing values that means the NN adapts its weights to corrupt data. Although theoretically simple imputations cannot provide adequate remedy due to the nonlinear character of traffic dynamics, test results show that this approach is successful.

Introduction of temporal and spatial dependences is proposed for lowering the impact of data loss from vehicle detectors. Li *et al.* [16] introduce a kernel probabilistic principle component analysis model for combining traffic flow information from neighbouring measuring points. This model relaxes the linear mapping assumption of probabilistic principle component and thus better describes the nonlinear spatial–temporal dependence around neighbouring points. Vlahogianni *et al.* [21] discuss the idea of a modular neural network, for incorporation of spatial dependences in the prediction process. Spatially distributed, along a route, sites feed traffic flow data to ascribed NN modules which reconstruct the characteristics of the data series. The outputs of the modules compete or can be averaged to produce the final output. An extended approach is presented in [22] where the sites are placed at major intersection of an urban transport network.

Mapping of complex relations of diverse variables using multilayer perceptrons is achieved by introducing additional hidden layers in the network. This approach encounters difficulties in elaboration of weights of these multilayer structures. A successful solution to this task is brought about with the introduction of the deep learning networks. Deep learning is aimed at extracting features hierarchies in a layer-wise fashion. This is done usually using stacked autoencoders. DLNs find numerous applications in image processing, where the data reduction and abstraction of information is at the centre of attention.

As yet only a few applications of DLN in traffic prediction can be noted in literature. Huang *et al.* [23] propose a deep architecture that consists of a deep belief network (DBN) and a multitask regression layer. The DBN extracts features from the series of traffic flow values at different locations which are used by the multitask layer for prediction. Authors examine DBNs of different depth and number of nodes to achieve the best prediction results. In [24] a DLN consisting of stacked autoencoders is proposed to learn generic traffic flow features. The model successfully discovers the latent traffic flow features such as spatial and temporal correlations. A logistic regression layer on top of the network is used for supervised traffic flow prediction. Stacked autoencoders are also examined in [25] and prove to be highly effective for traffic flow prediction along a motorway.

Polson and Sokolov [32] resolve the problem of finding spatio-temporal relations in traffic data using a hierarchical linear vector autoregressive model. The parameters of the model are used to build a DLN. The optimal structure and weights are found using stochastic gradient descent. Prediction performance is improved over linear models.

A DLN composed of autoencoders complemented with denoising autoencoders is proposed in [33]. The denoising autoencoders capture statistical dependencies between inputs. This model recovers traffic data through feature extraction and statistical dependency learning. Random and continuously corrupted traffic data sets are used for testing the model. The results prove the significance of accounting for statistical dependencies between data.

III. Methodology

Data loss is modelled using deterministic or probabilistic representations. In the case of ITS the functioning of vehicle detection infrastructure can be described as deterministic. Deterministic in the sense that lengths of periods of data loss are defined. Supervisors of the system observe strict maintenance rules. A set of determined reaction periods is defined for bringing back to order the malfunctioning devices. An important question arises how to set the output of the detection device during the reaction period. Traffic flow is

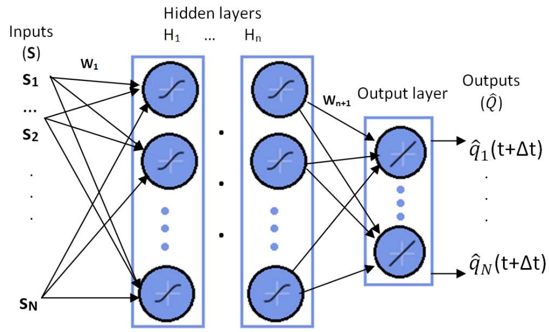

Fig. 1. Multilayer perceptron for traffic flow prediction.

chosen as the parameter for analysis, this is the basis for elaborating control and management decisions in ITS.

An urban road network comprising of N nodes is modelled. N sets $s_i(t)$ of k past values of traffic flow are used for prediction:

$$s_i(t) = (q_i(t), q_i(t-\Delta t), \ldots, q_i(t-k\Delta t)))$$

Δt is the interval of observations, i is the node number.

The flow q_i at a road network node i, usually a junction or a vehicle detection station on a traffic lane, is defined as:

$$q_i(t) = \begin{cases} f_i(t) & measured\ flow \\ B & break\ down\ value \end{cases}$$

The prediction values at each node $\hat{q}_i(t+\Delta t)$ are elaborated using data from all nodes:

$$\hat{q}_i(t+\Delta t) = P_i(s_1(t), s_2(t), \ldots, s_N(t))$$

For all nodes

$$\hat{Q} = P(S)$$

Neural networks of different structures are used for prediction of traffic flow values. Two representative neural networks are used: a multilayer perceptron MLP and a deep learning network DLN based on autoencoders.

The multilayer perceptron – fig. 1 has kN inputs and a number of hidden layers of neurons. The numbers of neurons in the layers are determined by the required complexity of mapping of the traffic flow data series [34]. Tested networks with up to two hidden layers are adequate to model the volatility of urban traffic flow changes. The output layer contains N neurons which generate prediction values of traffic flow at the nodes of the road network.

The input values are passed with weights W_1 to the first hidden layer of m neurons. The values of the outputs of the neurons of this layer $H_1(j)$ are calculated using the formula:

$$H_1(j) = \sigma(W_1(j)S + b_1(j))$$

j is the neuron number.

A chosen transfer function σ is used to evaluate the outputs. Usually a symmetric sigmoid function is used in hidden layers. Consecutive layers use these output values as inputs and elaborate subsequent outputs in the same way:

$$H_n(j) = \sigma(W_n(j)H_{n-1} + b_n(j))$$

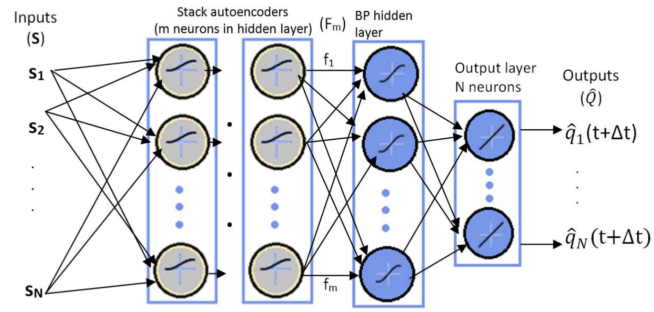

Fig. 2. Deep learning network based on autoencoders for prediction.

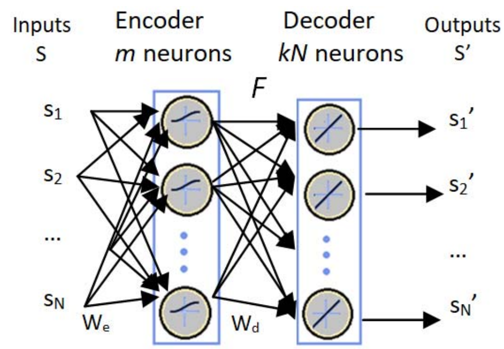

Fig. 3. Nonlinear autoencoder pretraining – the first autoencoder in the stack.

The output layer consists of N neurons with a linear transfer function. The values of the outputs of the network, that is the traffic flow prediction values for each of the nodes, are calculated using:

$$\hat{Q} = W_{n+1}H_n + b_{n+1}$$

Weights of the neurons are updated during training using the Levenberg-Marquardt algorithm.

The deep learning network consists of a set of stacked autoencoders, the outputs of the stack are passed to a MLP. The numbers of neurons in the autoencoders are a function of the number of road network nodes and the anticipated features of the traffic flow data series. The output layer contains N neurons which generate prediction values based on features elaborated by the last autoencoder of the stack. Fig. 2 presents the diagram of the DLN based on autoencoders.

An autoencoder is composed of an encoder and a decoder. The encoder and decoder can have multiple layers. The DLN autoencoder stack is pretrained layer by layer. The encoder contains m neurons, while the number of decoder neurons is equal to the number of inputs.

The number of neurons in the first autoencoder of the stack is a function of the volatility of the traffic data. Urban traffic data exhibits characteristic modes of value changes. There are plateaus, steep increases, peaks, and smooth falls of values that is changes of different linear or non linear character. This behavior can be represented by a few tens of features.

Fig. 3 introduces the process of autoencoding in the case of the first autoencoder. Consecutive encoders use the f values - generic features, as inputs and the training process is repeated.

The weights W_e of the encoder and W_d of the decoder neurons are updated until the autoencoders outputs draw level with the inputs. This is done in two steps:

- encoding step maps input data $s_1, \ldots, s_N$ to features $f_1, \ldots, f_m$,

- decoding step maps features $f_1, \ldots, f_m$ back to $s'_1 \ldots, s'_N$, which approximates the original input data.

Features $f_1, \ldots, f_m$ are calculated:

$$f_j = \sigma_e(W_e(j)S + b_e(j))$$

Then, the decoder maps the encoded representation that is features $f_1, \ldots, f_m$ back into an estimate of the original input vector S' as follows:

$$s'_i = \sigma_d(W_d\,(i)\,F + b_d(i))$$

σ_e, σ_d are transfer functions of the neurons in the autoencoder. The neurons have usually sigmoid transfer functions.

The goal of updating is to have S' to approximate S, that is the objective function G:

$$G = \sum_{i=1}^{N} (s'_{\mathbf{i}} - s_{\mathbf{i}})^2 = \sum_{i=1}^{N} (\sigma_d(W_d(i)F + b_d(i)) - s_i)^2$$
$$= \sum_{i=1}^{N} \sum_{j=1}^{kn} (\sigma_d(W_d(i)(\sigma_e(W_e(j)S + b_e(j))) + b_d(i)) - s_i)^2$$

is minimized using the scaled conjugate gradient descent method.

The output of the DLN consists of a single BP hidden layer and elaborates the prediction values using the set of features F from the last layer of the stack. The hidden layer and the output layer of the DLN are trained using the Levenberg-Marquardt algorithm.

IV. Case Study

Two representative neural networks are used: a multilayer perceptron MLP and a deep learning network DLN based on autoencoders. The aim of the study is to find the difference in sensitivity of these two types of neural networks, used for predictions, to data loss. The construction of DLN based on autoencoders is much more complex than MLP and the question is whether this brings substantial performance gains. Trained networks enable a very fast evaluation of prediction values and are suitable for modelling diverse data sets.

Least complex structures of the networks are proposed as to track the significance of differences in the way the networks map road traffic data series. In [27] and [28] results of applying a number of neuron network architectures for prediction of traffic flows are discussed. MLPs with up to three hidden layers and with a varying number of neurons (6 to over 60) in the layers are investigated. The research work shows that expanding the structures contributes to smaller error rates but does not bring decisive changes in the mapping behavior.

The proposed MLP contains one hidden layer and the DLN consists of one autoencoder. One network is used for prediction of traffic flow values at all of the measuring sites in parallel. This enables the incorporation of spatio-temporal relations between the traffic data. Such a compact solution is desirable for elaborating control and management decisions in ITS.

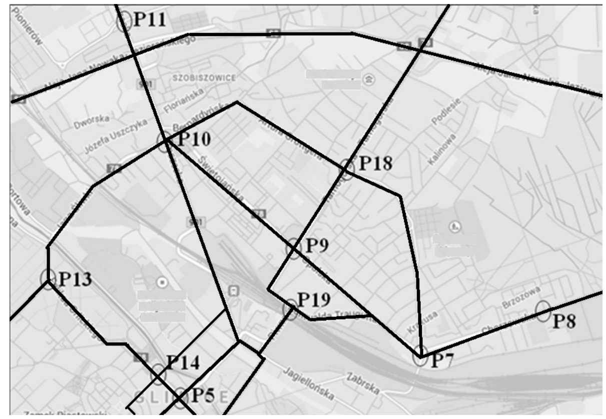

Fig. 4. The placement of the ten measuring sites.

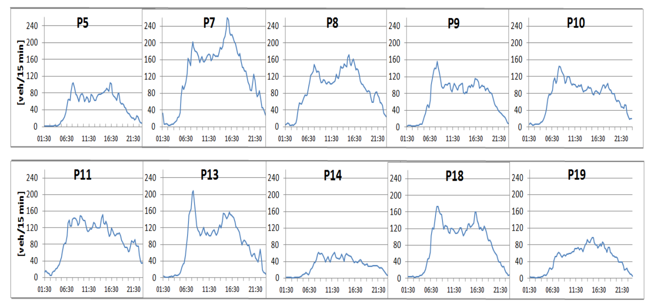

Fig. 5. Intraday traffic flow at the road network nodes.

The dataset for training and testing the networks comes from an ITS database maintained by the Gliwice Traffic Control Centre. Gliwice is a town in Upper Silesian in Poland with about 200 thousand inhabitants. Only a part of this huge database is used in this study. A set of ten traffic flow measurement sites $N = 10$ are chosen for tests.

The measurement sites are located at signalised junctions on main roads carrying in and out bound traffic to the centre of the town fig. 4. Loop detectors and video detectors are used to collect traffic data. Data is locally used for traffic control at the junctions and forwarded to the towns traffic control centre.

Morning and afternoon peak hours can be observed, no congested traffic is noted. The mean flow rates fall in the range of 40 to 160 veh/15min. Fig. 5 illustrates the diversity of road traffic data. This diversity of flow may pose a challenge for performing parallel prediction of values.

In the case of this traffic control centre, vehicles counts are registered in 5 minute intervals and converted to average flow values over 15 min periods. In all over 100 thousand flow values, that is more than 10 thousand at each site, covering over 100 working days (Monday to Friday) of traffic observations are used to train and an appropriate number to test the networks. The test values are also weekday observations but do not belong to the training set.

The traffic control centre registers additionally traffic incidents and monitors the state of the road infrastructure. This data is not used in the current study.

Intraday traffic series with gaps, larger than 3 consecutive registration periods, are omitted. Corrupted data accounts for about 10% of all of the collected vehicle counts. The problem of data loss is closely related to the maintenance capabilities of the road authorities and reliability of traffic measuring devices. Immunity to corrupted or missing data is an important feature of prediction models.

Weekday's data is used. No data standardization is carried out. The ranges of flow values at measuring sites are not highly volatile. Traffic flow values in the night (from 23.00 till 6.00 flows are very low) are not included in the calculation of errors.

Preliminary examination of the traffic flow data series and earlier studies [28] using a similar dataset permit to define the length of past observation as 6 periods of collecting flow values - k equal to 5, from time t to $t\text{-}5\Delta t$. This means that the prediction is based on previous 90 minutes of traffic observation.

This accounts for $kN = 60$ inputs of the MLP. The hidden layer consists of 12 neurons; this number comes as a result of several tests of adapting to characteristics of the data series and to the number of measurement sites. Sites with high traffic flow values require a distinct neural representation in the layer, one can conclude. Sigmoid transfer function is used for the neurons.

The autoencoder of the DLN consists of 20 neurons. Such a small number of neurons in comparison to the number of inputs kN=60 is proposed in compliance with the earlier tests done using MLP. The output layer consists of N neurons with a linear transfer function.

V. Results and Discussions

The study includes assessment of the prediction performance of the proposed networks and evaluation of the networks sensitivity to data loss.

A. Prediction Performance

Testing data for evaluating the prediction performance comes from the same database of the ITS control centre and is prepared in the same way as training sequences. Data collected on five randomly chosen workdays for each of the measuring sites are chosen for testing. MAP and RMS errors for intraday traffic flow predictions are calculated

$$MAPE = \frac{1}{M}\sum_{i=1}^{M}\frac{\left|f_i(t) - \hat{q}_i(t)\right|}{f_i(t)}$$

$$RMSE = \sqrt{\frac{1}{M}\sum_{i=1}^{M}\left(f_i(t) - \hat{q}_i(t)\right)^2}$$

M – total number of measurements at the measuring site, $f_i(t)$ – measured values, $\hat{q}_i(t)$ - predicted values corresponding to the measured $f_i(t)$.

Additionally naive prediction is used to provide some quick reference estimates of prediction errors [34]. The same test sets as for NNs are used for evaluation. The predicted traffic flow is calculated using previously registered values of flow:

$$\hat{q}(t + \Delta t) = \frac{q_{hist}(t + \Delta t)}{q_{hist}(t)} q(t)$$

Δt is the interval of observations, $q_{hist}(t)$ is the historical average volume at time point t. The average is calculated using 100 values recorded at the measuring site at time t on work days during tests. The quotient of historical averages represents a simple measure of the trend of change.

TABLE I
Prediction Errors of MLP, DLN Networks, Naïve and ARIMA Model

Measurement sites	Mean intraday traffic flow values [veh/15min]	RMSE [veh/15min]				MAPE [%]			
		MLP	DLN	Naive	ARIMA	MLP	DLN	Naive	ARIMA
P5	47.1	8.0	7.6	9.4	6.4	9.9	9.5	12.0	9.5
P7	122.6	10.5	10.7	14.9	14.6	5.9	6.0	8.0	8.3
P8	87.7	9.2	9.0	11.3	4.1	7.7	7.6	8.5	7.2
P9	68.1	7.9	7.4	10.0	10.7	7.3	7.6	8.7	8.9
P10	71.8	7.1	7.1	9.6	9.3	6.8	6.9	8.9	7.9
P11	88.0	9.3	9.1	11.7	13.8	7.3	7.1	9.2	11.3
P13	81.7	10.4	10.5	15.5	13.6	8.8	8.8	13.5	10.8
P14	30.8	5.6	5.5	6.9	7.2	11.5	10.5	13.7	13.8
P18	80.6	8.6	8.6	9.9	10.1	7.7	7.6	8.2	7.4
P19	46.5	6.6	6.7	7.0	10.0	9.1	9.5	9.7	11.1
					mean MAPE	8.2	8.1	10.0	9.6

This approach does not take into account mutual relations of traffic streams which are captured by the proposed neural networks based prediction solution. Predictions done using ARIMA based models can give better results [35].

Seasonal ARIMA(p, d, q)(P, D, Q)$_S$ model is prepared for comparison [2]. The period value S=480 is used as the construction is done for a sequence of 5 working days with 96 measurements per day. The other parameters of the model ARIMA(1, 0, 1)(0, 1, 1)$_{480}$ are selected on the basis of the Akaide information criterion (AIC) applied to the collected traffic data. The necessary outlier detection for the seasonal ARIMA models was quite time consuming.

Table 1 presents the summary of prediction errors of the networks side by side to enable the comparison of their performance. MAP errors fall in the range of 5 to 11% while RMS errors in the range of 5 to 11 veh/15min. Average MAP errors, including all nodes, for MLP amount to 8.2% while for DLN it is 8.1%, which means that the networks predict almost exactly.

The naive prediction consistently gives larger errors. At some sites the errors are 50% larger. Site P7 and P13 register large traffic flows with distinct sharp peaks. Neural networks also give large errors for these sites, but the influence of neighbouring sites P8, P9 and P10, P5 significantly reduces the prediction errors.

The seasonal ARIMA based model gives lower prediction errors than the naive prediction, at two measurement sites P8, P18, also give slightly lower errors than NN models. In all the mean MAPE is still significantly higher than for NN models.

Two extreme graphs of errors are shown in fig. 6 to emphasize the existing differences in prediction performance.

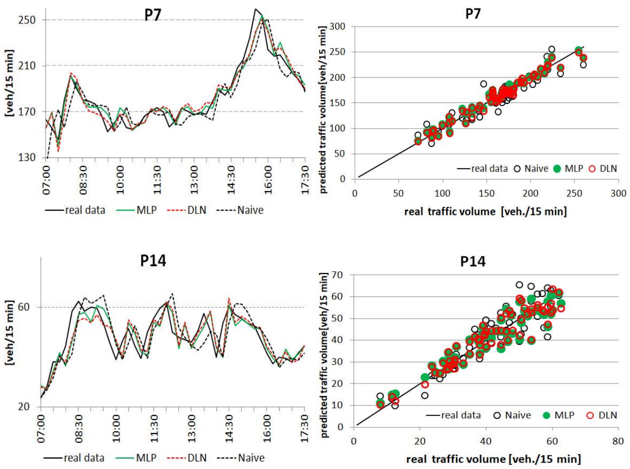

Fig. 6. Intraday (7.00-17.00) traffic flow predictions error graphs node P7 – smallest MAPE, node P14 – largest MAPE for both networks and Naïve model.

Highest differences are obtained for nodes with low traffic flows (P14) and lowest when traffic flow values are high (P7). Analysis of the space relations between sites shows that prediction errors do not spread equally in the neighbourhood but follow the routes of low traffic load.

B. Sensitivity to Data Loss

Four deterministic data imputation strategies are tested, that is the break down variable B in the prediction model takes on the values:

- zero,
- half of the maximum value of traffic flow noted at the measuring site,
- maximum value of traffic flow noted at the measuring site,
- past values (historical) of the traffic flow noted at the measuring site,

Additionally two random imputation strategies are tested where up to three randomly chosen measuring sites are shut down for two hours during morning (6.00-8.00) and afternoon (14.00-16.00) traffic peaks. The breaks down values are zero and mean of past values (historical) of the traffic flow noted at the measuring site in the 2 hour gap. This test resembles a real world case of measuring sites malfunctioning; it means that up to a third of the traffic data collection system is out of order. It is assumed that the repair time is at most two hours. Management of traffic streams at road traffic peaks is crucial for preventing congestion in road networks.

The first deterministic strategy in real world applications needs no special means of implementation as it is usually the value of data feed when there is no communication with the vehicle detector or the detector is completely broken down. Other strategies are more cumbersome to implement, as a way of registering previous values of traffic flow is required, either at the control centre or at the measuring sites. The duration of data loss is set to one day as maintenance works usually take place at night when the traffic is at its lowest level.

Tests are carried out modelling different levels of failure of the traffic measurement system. One, two or three measuring sites are shut down and their data feed is substituted with break down variables. Approximately it means disabling 10 to 30% of measuring capabilities of the system.

The networks are trained with correct data from all sites. Prediction MAPE and RMSE are evaluated with test data from working vehicle detectors and break down values from shut down detectors.

Tables II to V present MAPE for different deterministic imputation strategies and different levels of measuring system failure. Sites P5, P7, P8 at the edges of the road network are shut down to disable the measuring capabilities of the system.

Mean values of errors describe the performance of different imputation strategies. The values presented in the tables are representative also for other combinations of shut down sites and may be regarded as performance measures when 10 to 30% of the measuring capabilities are disabled.

TABLE II
PREDICTION MAPE FOR B = HISTORICAL TRAFFIC FLOW

	B = historical traffic flow					
	Malfunctioning sites					
	P5		P5 & P8		P5, P7 & P8	
Measuring sites	MLP [%]	DLN [%]	MLP [%]	DLN [%]	MLP [%]	DLN [%]
P5	12.3	12.0	12.2	12.1	12.4	12.1
P7	5.8	6.1	5.8	6.1	6.9	6.8
P8	7.3	7.5	11.2	10.0	11.1	10.1
P9	7.2	7.5	7.0	7.4	7.1	7.1
P10	6.5	7.0	6.4	6.9	6.2	6.7
P11	7.2	7.1	7.1	7.1	7.0	7.0
P13	9.4	8.8	9.4	8.8	9.3	8.6
P14	11.7	10.5	11.8	10.5	11.7	10.5
P18	8.0	7.6	7.9	7.6	8.0	7.6
P19	9.1	9.6	9.1	9.6	9.0	9.4
Mean values excluding broken detectors	8.0	8.0	8.1	8.0	8.3	8.1

TABLE III
PREDICTION MAPE FOR B = 0

	B = 0					
	Malfunctioning sites					
	P5		P5 & P8		P5, P7 &P8	
Measuring sites	MLP [%]	DLN [%]	MLP [%]	DLN [%]	MLP [%]	DLN [%]
P5	83.0	80.0	81.0	76.0	82.0	87.0
P7	6.9	6.6	7.2	7.6	79.0	85.0
P8	8.6	7.8	94.3	95.0	88.0	89.0
P9	8.4	8.2	9.5	11.4	10.7	8.9
P10	7.4	7.5	7.8	7.5	7.8	7.0
P11	7.3	8	7.5	9.3	8.7	14.2
P13	9.1	11.2	10.9	11.8	10.4	13.2
P14	11.4	11.4	11.5	11.3	10.7	12.2
P18	9.3	9.6	8.5	9.0	12.0	8.4
P19	10.5	11	10.9	13.0	9.6	9.6
Mean values excluding broken detectors	8.8	9.0	9.2	10.1	10.0	10.5

TABLE IV
PREDICTION MAPE FOR B = HALF OF THE MAXIMUM

	B = half of the maximum					
	Malfunctioning sites					
	P5		P5 & P8		P5, P7 & P8	
Measuring sites	MLP [%]	DLN [%]	MLP [%]	DLN [%]	MLP [%]	DLN [%]
P5	49.0	49.0	49.0	49.0	52.0	51.0
P7	6.4	6.6	6.5	6.9	28.3	14.3
P8	7.6	7.7	31.4	31.0	28.2	14.8
P9	8.1	8.4	6.2	85	8.6	9.3
P10	6.4	6.6	6.5	6.5	6.5	7.1
P11	6.9	7.2	7.4	7.4	8.0	8.1
P13	8.6	9.1	9.2	9.2	9.2	9.4
P14	11.7	11.5	11.7	11.5	12.4	11.5
P18	16.2	10.6	12.0	11.3	11.9	12.2
P19	9.3	9.6	9.2	9.8	9.2	10.6
Mean values excluding broken detectors	9.0	8.6	8.6	18.5	9.4	9.7

TABLE V
PREDICTION MAPE FOR B = MAXIMUM

	B = maximum value of traffic flow					
	Malfunctioning sites					
	P5		P5 & P8		P5. P7 & P8	
Measuring sites	MLP [%]	DLN [%]	MLP [%]	DLN [%]	MLP [%]	DLN [%]
P5	129.0	126.0	130.0	124.0	136.0	120.0
P7	9.2	8.4	9.2	7.8	81.3	67.5
P8	7.7	8.6	83.5	72.0	83.9	68.5
P9	9.7	11.0	13.6	11.2	13.3	18.0
P10	6.6	6.7	6.5	7.3	6.5	8.9
P11	7.5	7.2	7.8	8.6	7.6	14.6
P13	9.4	11.5	9.8	11.5	9.4	22.3
P14	14.0	16.2	12.3	17.0	14.9	15.8
P18	19.9	17.0	18.2	17.2	21.0	18.4
P19	9.1	10.4	9.2	11.0	11.0	17.3
Mean values excluding broken detectors	10.3	10.8	18.9	18.2	12.0	16.5

Detailed examination of prediction errors in the case of malfunctioning measurement sites brings about the following observations Prediction results for broken down sites reach very large errors for all investigated imputation strategies except when the break down variable takes on historical traffic flow values.

This behavior may be linked to the way the networks were trained. The training sequences did not include sequences modelling break downs.

Prediction results for functioning sites are best in the case of substituting break down values with zeros or historical flow values. Imputing zeros increases the prediction errors of MLP by 0.5 to 2% when the system loses 10 to 30% of measuring capabilities. DLN performance is slightly worse that is the error is larger reaching 0.8 to 2.4%. Imputing historical traffic

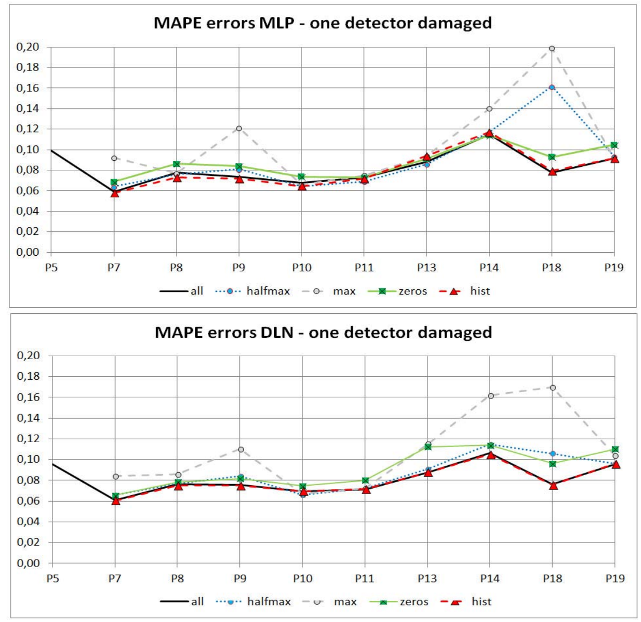

Fig. 7. Prediction errors when one measuring site is broken down (P5) for MLP and DLN networks.

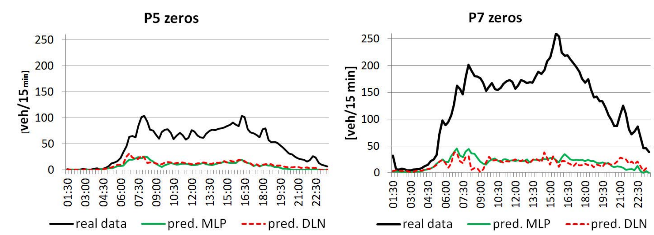

Fig. 8. Intraday traffic flows and predictions for sites P5 and P7 - zeros for all day.

flow values changes the error by +/− 0.2%, which means that there is practically no significant change in performance.

Substituting the break down traffic flow values with maximum or half of maximum traffic flow values at the sites generates the largest prediction errors especially in the case of the DLN. Spatial position of the broken site does not affect errors at sites.

Neighbouring sites are not affected by the breakdown site but relatively large errors appear at seemingly not related sites. This phenomenon is especially visible for DLN predictions. Fig. 7 illustrates this behavior. Large errors appear at sites far away from the broken one. Broken site P5 strongly changes the prediction values at P9 and P18 which lie on the other side of the road network.

Sites P9 and P18 measure the traffic stream along an artery to the town centre, part of the traffic passing P18 disperses so the values at P9 are lower but still closely related. Detailed examination of traffic patterns used for training show that the P18 measurement site provides highly volatile traffic data. This can be a source of increased sensitivity of the neural model as the number of used neurons may be inadequate to map the diversity of traffic flow changes.

Detailed graphs of predictions and test traffic flows, for *B=0* shown on fig. 8. P5 is chosen to illustrate low traffic

TABLE VI
PREDICTION MAPE FOR TWO HOUR ZERO VALUE GAPS

	B = zero in 14.00-16.00					
	Randomly chosen malfunctioning sites					
	one		two		three	
Measuring sites	MLP [%]	DLN [%]	MLP [%]	DLN [%]	MLP [%]	DLN [%]
P5	19.9	20.0	19.8	19.7	20.3	20.9
P7	6.0	6.9	6.7	6.7	19.1	19.0
P8	7.2	7.5	20.0	19.9	19.9	19.2
P9	7.5	7.2	7.7	7.5	8.0	7.3
P10	6.8	6.4	6.8	6.4	7.1	7.0
P11	7.2	6.9	7.2	6.9	8.7	7.4
P13	11.4	9.1	11.4	9.6	10.2	11.1
P14	12.6	10.9	12.6	11.2	13.2	11.7
P18	8.7	7.9	8.7	7.6	9.0	8.5
P19	9.6	9.0	9.6	9.2	9.9	9.6
Mean values excluding broken detectors	8.6	8.0	8.8	8.1	9.4	8.9

TABLE VII
PREDICTION MAPE FOR TWO HOUR MEAN VALUE GAPS

	B = mean in 14.00-16.00					
	Randomly chosen malfunctioning sites					
	one		two		three	
Measuring sites	MLP [%]	DLN [%]	MLP [%]	DLN [%]	MLP [%]	DLN [%]
P5	10.0	10.0	10.0	10.1	10.2	10.2
P7	5.7	6.0	5.7	6.0	7.4	7.5
P8	7.0	7.1	7.7	7.7	7.8	7.9
P9	7.5	6.9	7.5	6.9	7.5	6.9
P10	6.4	6.4	6.4	6.4	6.3	6.3
P11	7.2	6.8	7.2	6.8	7.2	6.8
P13	9.4	8.6	9.3	8.6	9.4	8.6
P14	11.8	10.9	11.9	10.9	11.8	10.8
P18	8.1	7.5	8.1	7.5	8.0	7.5
P19	9.2	8.9	9.2	8.9	9.2	8.9
Mean values excluding broken detectors	8.0	7.7	8.2	7.8	8.5	8.0

predictions and P7 represents heavy traffic prediction results. Although the prediction values are much smaller than the test flow values the course is preserved. MLP and DLN predictions coincide for low flow rates (P5). Naive predictions give zeros.

Tables VI and VII present MAPE for the two random imputation strategies and different levels of measuring system failure. Both traffic peaks were examined. There are no significant differences in the prediction behavior. MAPE are calculated when data gaps, for randomly chosen measuring sites, are introduced in the afternoon road traffic peaks. The tables contain results for draws that gave the same set of measuring sites as previously investigated. This enables a comparison of the performance of deterministic and random imputation strategies.

The results prove that using zero values for gap data has little influence on prediction errors of correctly functioning

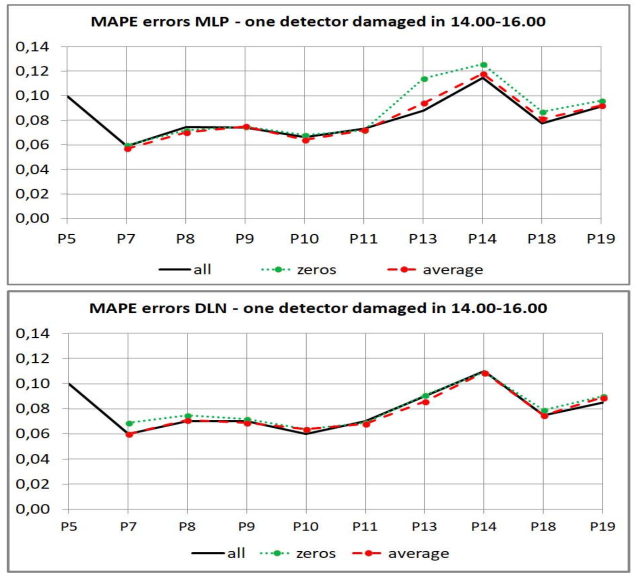

Fig. 9. Prediction errors when one measuring site is broken down (P5) in 14.00-16.00 for MLP and DLN networks.

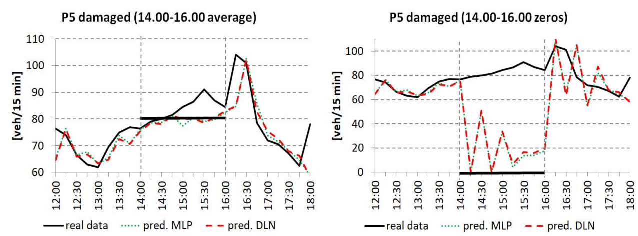

Fig. 10. Prediction results when one measuring site is broken down (P5) for MLP and DLN networks.

measuring sites. DLN prediction errors are reduced by 0.1% in the case of one broken site and are 0.8% larger when three sites are shut down. MLP prediction errors rise by 0.4% in the case of one broken site and by 1.2% when three sites are shut down.

Better prediction results are obtained for using mean values in the gaps. DLN prediction errors are reduced by up to 0.4% while MLP prediction errors at most rise by 0.3%.

In comparison with deterministic imputation the errors are lower which complies with the time the measuring sites were out of order. Deterministic imputation assumes one day shut downs while random imputation models two hour gaps in data. DLN is less sensitive to data gaps.

Fig. 9 shows the spread of errors when a data gap is introduced at one of the measuring sites. In the case of imputing the mean value in the gap there is negligible influence on other measuring sites errors for both NNs. A zeroed gap slightly and evenly increases errors on all other sites for both NNs.

Detailed prediction results for the broken measuring sites are shown on fig. 10. The predicted values in the data gaps demonstrate oscillations for both NNs.

Low impact of imputing zeros data on prediction errors can be connected to the fact that traffic data at night consists of low

TABLE VIII
PREDICTION MAPE FOR B = 0, B = HALF OF THE MAXIMUM (HMAX), B = MAXIMUM (MAX)

	MAPE [%]					
	B=0		B=HMax		B=Max	
Measuring sites	MLP [%]	DLN [%]	MLP [%]	DLN [%]	MLP [%]	DLN [%]
P5	61.3	34.3	38.5	35.5	38.7	32.9
P7	7.2	6.3	6.1	6.0	6.1	7.1
P8	7.3	8.1	7.1	7.6	7.1	8.2
P9	7.8	7.5	7.5	7.5	7.5	7.6
P10	6.8	7.0	6.3	6.3	6.4	6.7
P11	7.4	9.3	7.5	7.0	7.5	7.5
P13	9.7	9.2	8.8	8.5	9.6	10.2
P14	11.6	10.7	13.1	10.6	10.7	11.2
P18	7.7	7.6	8.3	8.2	8.2	8.0
P19	9.8	8.3	9.1	8.6	9.1	9.3
Mean values excluding broken detectors	8.4	8.2	8.2	7.8	8.0	8.4

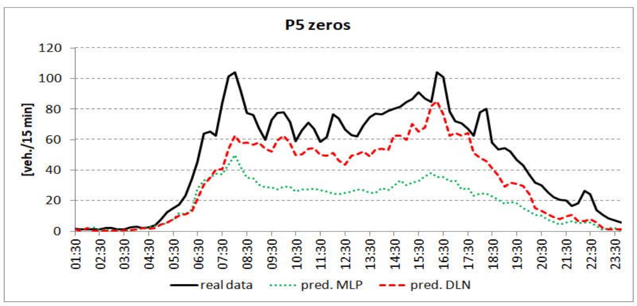

Fig. 11. Intraday traffic flows and predictions for sites P5 - zeros for all day using training sequences complemented with break down values.

values which are incorporated in the training sequences. The trained networks, especially DLN, are capable of matching this to corrupted sequences. This is further investigated. The training of the networks is changed. Training sequences are complemented with break down values. 1% of the sequences contain erroneous data. Table VIII presents the new results for the example of one malfunctioning site – P5; other sites behave in a similar way.

The mean values of MAPE are lower for all the networks. DLN networks significantly improve their performance in the case of imputing B=Max the error falls by more than 2%. The prediction performance at the malfunctioning site P5 is also highly improved as shown on fig.11.

Comparison of fig. 8 and fig. 11 distinctly indicates, that introduction of training with break down values, restores near normal traffic values instead of depicting the character of changes.

VI. CONCLUSIONS

The proposed models of road traffic flow prediction based on neural networks take into account spatiotemporal relations between traffic measuring sites. A single neural network is used for prediction of traffic at all of the investigated traffic measuring sites. The networks are trained using large sets of past historical data which is time consuming but it is done off-line. The training sets comprise summer and winter traffic data which map different weather and road conditions. Some exceptional traffic situations e.g. road works, are also included.

The trained networks predict road traffic flow, for the next measuring interval, for correctly functioning measuring sites, with an error of about 8%. The prediction is done in real time, which is advantageous for use in ITS.

The prediction performance of MLP and DLN based on autoencoders is comparable in the case of correctly functioning measuring sites. The parallel prediction of traffic flow at all of the sites takes into account mutual relations between the changes of flows at the sites.

More important is the sensitivity to data loss from the measuring sites. Random failures of measuring sites at traffic peak hours especially impair the performance. The loss of measuring capabilities in the range 10 to 30% models real word conditions of ITS functioning.

Prediction is improved by implementing an imputing strategy for broken down measuring sites. Investigations prove that substituting zeros or historical data gives the least impact on the prediction performance of examined NNs.

MLPs give better prediction results than DLNs in the case of data loss which lasts for one day. In all the differences in errors fall in the range 0.5% to 2.4%.

DLN based prediction is less sensitive to data gaps lasting for two hours. In all the differences in errors fall in the range −0.4% to 1.2%.

Incorporation of break down modelling sequences into the network training sequences improves the performance of the networks. DLN based prediction comprising data from all measuring sites and short maintenance periods can be recommended for use in ITS. The lost data is imputed with mean historical values of traffic flow.

Prediction of traffic values at broken down sites needs further investigation. The problem of devising an effective model of malfunctioning which maps different damages of the site remains to be resolved. Investigation of incorporating break down modelling sequences into the network training sequences is a prospective idea for future research.

ACKNOWLEDGMENT

The author thanks ZIR-SSR Bytom for providing video detector data from the Gliwice site.

REFERENCES

[1] J. Guo, W. Huang, and B. M. Williams, "Adaptive Kalman filter approach for stochastic short-term traffic flow rate prediction and uncertainty quantification," *Transp. Res. C, Emerg. Technol.*, vol. 43, pp. 50–64, Jun. 2014.
[2] B. M. Williams and L. A. Hoel, "Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: Theoretical basis and empirical results," *J. Transp. Eng.*, vol. 129, no. 6, pp. 664–672, Nov. 2003.
[3] M. Lippi, M. Bertini, and P. Frasconi, "Short-term traffic flow forecasting: An experimental comparison of time-series analysis and supervised learning," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 2, pp. 871–882, Jun. 2013.
[4] W. Min and L. Wynter, "Real-time road traffic prediction with spatio-temporal correlations," *Transp. Res. C, Emerg. Technol.*, vol. 19, no. 4, pp. 606–616, 2011.

[5] 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.*, vol. 24, no. 3, pp. 534–542, 2011.

[6] Y. Kamarianakis and P. Prastacos, "Forecasting traffic flow conditions in an urban network: Comparison of multivariate and univariate approaches," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 1857, no. 1, pp. 74–84, Jan. 2003.

[7] E. I. Vlahogianni, M. G. Karlaftis, and J. C. Golias, "Short-term traffic forecasting: Where we are and where we're going," *Transp. Res. C, Emerg. Technol.*, vol. 43, pp. 3–19, Jun. 2014.

[8] M. G. Karlaftis and E. I. Vlahogianni, "Statistical methods versus neural networks in transportation research: Differences, similarities and some insights," *Transp. Res. C, Emerg. Technol.*, vol. 19, no. 3, pp. 387–399, 2011.

[9] P. J. Brockwell and R. A. Davis, *Introduction to Time Series and Forecasting*. New York, NY, USA: Springer-Verlag, 2016.

[10] B. L. Smith, B. M. Williams, and R. K. Oswald, "Comparison of parametric and nonparametric models for traffic flow forecasting," *Transp. Res. C, Emerg. Technol.*, vol. 10, no. 4, pp. 303–321, Aug. 2002.

[11] M. Zhong, P. Lingras, and S. Sharma, "Estimation of missing traffic counts using factor, genetic, neural, and regression techniques," *Transp. Res. C, Emerg. Technol.*, vol. 12, no. 2, pp. 139–166, 2004.

[12] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Netw.*, vol. 61, pp. 85–117, Jan. 2015.

[13] 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. C, Emerg. Technol.*, vol. 13, nos. 5–6, pp. 347–369, Oct./Dec. 2005.

[14] H. Chen, S. Grant-Müller, L. Mussone, and F. Montgomery, "A study of hybrid neural network approaches and the effects of missing data on traffic forecasting," *Neural Comput. Appl.*, vol. 10, no. 3, pp. 277–286, 2001.

[15] Y. Li, Z. Li, and L. Li, "Missing traffic data: Comparison of imputation methods," *IET Intell. Transp. Syst.*, vol. 8, no. 1, pp. 51–57, Feb. 2014.

[16] L. Li, Y. Li, and Z. Li, "Efficient missing data imputing for traffic flow by considering temporal and spatial dependence," *Transp. Res. C, Emerg. Technol.*, vol. 34, pp. 108–120, Sep. 2013.

[17] J. Haworth and T. Cheng, "Non-parametric regression for space–time forecasting under missing data," *Comput., Environ. Urban Syst.*, vol. 36, no. 6, pp. 538–550, 2012.

[18] M. T. Asif *et al.*, "Spatiotemporal patterns in large-scale traffic speed prediction," *IEEE Transp. Intell. Transp. Syst.*, vol. 15, no. 2, pp. 794–804, Feb. 2014.

[19] B. Smith, W. Scherer, and J. Conklin, "Exploring imputation techniques for missing data in transportation management systems," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 1836, no. 1, pp. 132–142, Jan. 2003.

[20] G. Fusco, C. Colombaroni, and N. Isaenko, "Comparative analysis of implicit models for real-time short-term traffic predictions," *IET Intell. Transp. Syst.*, vol. 10, no. 4, pp. 270–278, 2016.

[21] E. I. Vlahogianni, M. G. Karlaftis, and J. C. Golias, "Spatio-temporal short-term urban traffic volume forecasting using genetically optimized modular networks," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 22, no. 5, pp. 317–325, 2007.

[22] K. Halawa, M. Bazan, P. Ciskowski, T. Janiczek, P. Kozaczewski, and P. Rusiecki, "Road traffic predictions across major city intersections using multilayer perceptrons and data from multiple intersections located in various places," *IET Intell. Transp. Syst.*, vol. 10, no. 7, pp. 469–475, Sep. 2016.

[23] W. Huang, G. Song, H. Hong, and K. Xie, "Deep architecture for traffic flow prediction: Deep belief networks with multitask learning," *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 5, pp. 2191–2201, Oct. 2014.

[24] Y. Lv, Y. Duan, W. Kang, Z. Li, and F.-Y. Wang, "Traffic flow prediction with big data: A deep learning approach," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 2, pp. 865–873, Apr. 2015.

[25] H.-F. Yang, T. S. Dillon, and Y.-P. P. Chen, "Optimized structure of the traffic flow forecasting model with a deep learning approach," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 10, pp. 2371–2381, Oct. 2017, doi: 10.1109/TNNLS.2016.2574840.

[26] T. Pamuła, "Neural networks in transportation research—Recent applications," *Transp. Problems*, vol. 11, no. 2, pp. 27–36, 2016.

[27] T. Pamuła, "Classification and prediction of traffic flow based on real data using neural networks," *Arch. Transp.*, vol. 24, no. 4, p. 519, 2012.

[28] T. Pamuła, "Short-term traffic flow forecasting method based on the data from video detectors using a neural network," in *Activities of Transport Telematics* (Communications in Computer and Information Science), vol. 395. Berlin, Germany: Springer, 2013, pp. 147–154.

[29] A. Muralidharan and R. Horowitz, "Imputation of ramp flow data for freeway traffic simulation," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2099, no. 1, pp. 58–64, Jan. 2009.

[30] M. E. Whitlock and C. M. Queen, "A Bayesian model of a traffic network," Univ. Kent, Canterbury, U.K., Tech. Rep. UKC/IMS/98/13, 1998.

[31] M. E. Whitlock and C. M. Queen, "Modelling a traffic network with missing data," *J. Forecasting*, vol. 19, no. 7, pp. 561–574, 2000.

[32] N. Polson and V. Sokolov. (2016). "Deep learning for short-term traffic flow prediction." [Online]. Available: https://arxiv.org/abs/1604.04527

[33] Y. Duan, Y. Lv, Y.-L. Liu, and F. Wang, "An efficient realization of deep learning for traffic data imputation," *Transp. Res. C, Emerg. Technol.*, vol. 72, pp. 168–181, Nov. 2016.

[34] B. L. Smith and M. J. Demetsky, "Traffic flow forecasting: Comparison of modeling approaches," *J. Transp. Eng.*, vol. 123, no. 4, pp. 261–266, Jul. 1997.

[35] M. Bernaś, B. Płaczek, P. Porwik, and T. Pamuła, "Segmentation of vehicle detector data for improved k-nearest neighbours-based traffic flow prediction," *IET Intell. Transp. Syst.*, vol. 9, no. 3, pp. 264–274, Apr. 2015.

Teresa Pamuła received the D.Phil. degree in engineering science in 2002. She is a graduate of the Faculty of Automatic Control, Electronics and Computer Science at the Silesian University of Technology (SUT), Poland.

She is currently an Assistant Professor with the Faculty of Transport, SUT. She has authored or co-authored several textbooks in programming and over 50 papers in peer-reviewed journals. Her research interests include the application of artificial intelligence methods in transport problems, especially neural networks for prediction and evaluation of road traffic parameters.