Framework for real-time traffic forecasting methodology under exogenous parameters

Sofia Samoili, EPFL/LAVOC
Prof. André-Gilles Dumont, EPFL/LAVOC

Conference paper STRC 2012

STRC | 12th Swiss Transport Research Conference
Monte Verità / Ascona, May 02-04,2012

Framework for real-time traffic forecasting methodology under exogenous parameters

Sofia Samoili
Laboratory of Traffic Facilities (LAVOC)
École Polytechnique Fédérale de Lausanne (EPFL)
GC C1 383, Station 18, 1015, Lausanne

Phone: +41 21 693 06 02
Fax: +41 21 693 63 49
sofia.samoili@epfl.ch

André-Gilles Dumont
Laboratory of Traffic Facilities (LAVOC)
École Polytechnique Fédérale de Lausanne (EPFL)
GC C1 389, Station 18, 1015, Lausanne

Phone: +41 21 693 23 89
Fax: +41 21 693 63 49
andre-gilles.dumont@epfl.ch

April 2012

Abstract

Ensuing from the deteriorating conditions of road networks, traffic forecasting techniques with mathematical and computer-theory methods have been employed to address the real-time prediction of traffic conditions and their dynamic control, with the optimum use of the current infrastructure. The prediction of traffic conditions, acknowledging multiple regimes, its transitions and driver's behaviour parameters, is a highly desired attribute in intelligent transportation systems (ITS), as it could increase operational performance.

The study in question aims to provide a framework to define the reasons that might preserve the conditions of recurrent and non-recurrent congestion occurrence in highways, so as to develop in long term an efficient dynamic traffic prediction model that would mitigate congestion emergence, before being triggered by interactive exogenous parameters, such as weather conditions, traffic composition, incident occurrence, traffic direction and seasonality. Contrary to currently existent prediction models, the challenge of the model to be developed is to capture traffic dynamics and enhance predictability upon multiregime (congestion, near-congestion, free flow) and transitional traffic behaviour, combining exogenous multi-dimensional determinants with real-time or near-real-time data, resulting to dynamic prediction models for highway traffic. Apart from the dynamic aspect, the transferability issue will be attempted to be addressed. Furthermore, active traffic management (ATM) highway management schemes will be formed in microscopic scale, with principal goal to control traffic by optimising the operation of an emergency lane as additional traffic lane.

Keywords

Traffic forecasting – traffic prediction – weather impact – incident occurrence – traffic composition

1. Introduction

Traffic congestion is an escalating phenomenon with multidimensional impact. A potential increase in the network capacity with an expansion of the network would only mask and transpose the problem ulteriorly, even more reinforced, without addressing the causality. An aspiring research could stand in the hypothesis that a traffic forecasting method, in order to conduce to restrain or prevent network's performance aggravation, should acknowledge several exogenous parameters for a more plausible representation of traffic evolution, as they retain mutual connections that affect congestion occurrence.

A number of interdisciplinary studies evoked computer-theory and mathematical methods to estimate and/or predict traffic conditions, without acknowledging though the dialectical interrelationships of prevailing exogenous parameters as a system. Furthermore, the overwhelming majority of the developed traffic prediction models have not integrated the weather determinant, even though it is undeniably affecting network operations' performance and traffic safety. In USA, weather's socio-economic impact has been recorded (Table 1), but for the rest of the world only the deaths and injuries from traffic accidents are enumerated, disdaining any relationship to weather and its power to accentuate the deterioration of effectiveness of existing infrastructure (WHO, 2004; FHWA, 2011).

Table 1 Traffic and socio-economic impact per year in freeways from adverse weather conditions in USA, based on FHWA Road Weather Management Overview (2011).

	Speed Reduction (% out of total observations)		**Weather-related Vehicles crashes** (% out of total observations)		**Deceased** (people)		**Injured** (people)		**Winter Road Maintenance**
	Light Weather Condi-tions	**Heavy Weather Condi-tions**	**Wet/ Snowy/ Icy pavement**	**During Rainfall/ Snow**	**Wet/ Snowy/ Icy pavement**	**During Rainfall/ Snow**	**Wet/ Snowy/ Icy pavement**	**During Rainfall/ Snow**	
Rain	2-13%	3-17%	75%	47%	5’700	3’400	544’700	357’300	n/a
Snow	3-13%	5-40%	24%	15%	1’300	900	116’800	76’000	2.3 billion $ 20% of state DOT maintenance budgets

Until recently weather was regarded as a fact and none research has been yet published that attempts to challenge or operate with the factor, in order to mitigate its impact in traffic.

Responding to this omission a mathematical model will be developed that will accede to quantify the control power of weather conditions, in order to proceed beyond the empirical qualitative studies of the weather impact. This research perspectives are among others to initiate a parameterised worldwide applicable model that detects traffic patterns and predicts the congestion duration in highways, thus the potential deterioration of level of service in a network, under exogenous variables, namely weather parameters, traffic composition (presence of vehicles' classes), incident occurrence, traffic direction, seasonality, and that leads not only to the quantification of the weather impact on road networks in form of traffic units, but also to the potential linked alteration of the network's performance through time, that is of the configuration of traffic evolution.

The prediction of traffic conditions, acknowledging multiple regimes, its transitions and driver's behaviour parameters, is a highly desired attribute in ITS, as it could increase operational performance. The model to be developed will attempt to capture traffic dynamics and enhance predictability upon multiregime (congestion, near-congestion, free flow) and transitional traffic behaviour in real-time or near-real-time, resulting to dynamic prediction models for highway traffic. Apart from the dynamic aspect, the traffic forecasting will be available regardless the site, ensuring the transferability issue.

1.1 Objectives statement and motivation

A framework for a real-time multi-step ahead traffic evolution prediction model (upon recurrent and non-recurrent congestion) under exogenous parameters for highways, will be described in the current study, aiming to quantify the impact of exogenous parameters (adverse weather conditions, traffic composition, incident occurrence, traffic direction, seasonality) on traffic congestion and predict the occurrence and duration of congestion. The weather impact quantification is regarded to be a novel approach for congestion prediction, as currently is only empirically confirmed that deteriorates traffic operations. The real-time prediction model that will be structured in a research following the framework described in the study in question, is expected to be for congestion prediction in highways, transferable to several sites and applicable in ITS and ATIS (Advanced Traveller Information System) environments for traffic control management by capturing traffic evolution congestion and mitigating its impact. Furthermore, an improvement of forecasting accuracy is expected, because of the exogenous parameters that have not been applied in their entirety in existing traffic forecasting models.

2. State of Research

Short-term traffic prediction algorithms were set to the center of research interest since the ‘90s, in order to improve the efficiency of network management in both urban arterials and highways via ITS strategies. The developed forecasting models estimate traffic parameters, such as speed, flow, density, occupancy, with short time forecasting horizons (a few minutes to a few hours) into the future and various prediction intervals (a few seconds to a few minutes), based on the behaviour of traffic dynamics that has to be detected and predicted.

According to empirical observations (Florio and Mussone, 1996; Kirby et al. 1997; Sheu, 1999; Ishak and Al-Deek, 2002), it is concluded that a large *forecasting horizon* leads to a degradation of prediction accuracy and a short forecasting step dictates a complex prediction. Nevertheless, it is suggested that an interval should not be very short (eg. 30s), as the potentially predicted traffic parameters, such as flow and speed, are strongly fluctuating causing noise in short time periods and thus decreasing the forecasting accuracy. For this reason data of short periods are usually aggregated in 5-minutes steps or greater (Florio and Mussone, 1996; Park and Rilett, 1998; Sheu, 1999), but the final prediction interval and horizon are defined based on the type of network (urban or highway) and on the type of the ITS system (traffic management system, real-time adaptive control system etc.) that the algorithm is called to operate. Implementations of traffic forecasting models in highways are implying that the optimum prediction step is in a range of 10-min and 15-min, as with a corresponding data aggregation the data variability declines and the accuracy of prediction horizon is improved (Head, 1995; Smith and Demetsky, 1997; Abdulhai et al., 1999). In European highways, where the average time headway is most commonly of 30 minutes, a 15-min or less prediction horizon is adequately envisaging the traffic performance (Smith and Demetsky, 1997).

Regarding to the *suitable parameters* that have been examined for capturing and forecasting traffic conditions, four dominate the literature: flow, occupancy, speed and travel time. The selection of a variable is intrinsically linked to type of available data (loop detectors, ITS/surveillance-derived stream data, real-time or not etc), the category of the network (urban, highway etc.) and the type of system that is aimed to be developed (control system, informative system-ATIS etc.).

In studies at the University of Leeds (Dougherty et al., 1994; Dougherty and Cobbett, 1997; Kirby et al., 1997) traffic flow, occupancy and speed have been found convenient for forecasting traffic dynamics. In particular, Dougherty and Cobbett developed three models to predict individually the aforementioned parameters in highways. The results for 5-min, 15-min and 30-min forecast horizons, indicated that the flow and occupancy prediction accuracy was greater than the speed’s, probably due to slow-moving vehicles in very low flow conditions. However, there are various contradicting results, which imply either that predictions based on traffic flow are more reliable (Levin and Tsao, 1980), or that occupancy is more stable and thus representative, as it is affected by traffic composition and vehicles' length (Lin et al. 2009). Of paramount's importance is the ascertainment that the forecasting accuracy is impetuously declining when an model is formed to predict multiple traffic variables, rather than multiple models for multiple traffic parameters, mainly as a result of their strong correlation (Florio and Mussone, 1996; Innamaa, 2000).

With the emergence of ITS technologies, from which traffic stream data could be derived in real-time and high accuracy, accurate travel time forecasting became possible, along with delays and queue lengths for both urban arterials and highways (Chen et al., 2001; Zhang and Rice, 2003; vanLint et al., 2005; Skabardonis and Geroliminis, 2008). In studies that these data were inaccessible, double-loop or single-loop detectors data were employed and travel time prediction was approached via space mean speed prediction from flow, density, occupancy and vehicle length estimation (Dailey, 1997; Kwon et al., 2000; Wang and Nihan, 2000; Dia, 2001). Regardless the aspiring approaches, the results are poorer and the accuracy levels redirect to an estimation rather than prediction.

When aiming to provide prediction for Advance Traveller Information Systems (ATIS), travel time or speed is favoured as predictive variables over flow and occupancy, because they are more efficient and meaningful for the users that are non-familiar to transportation theory users. Nevertheless, this possibility is usually given for highways that are monitored and real-time data can be provided, and less commonly for urban areas. Hence, for traffic control applications in several types of networks, namely highways, urban networks etc., and for ATIS applications that monitoring is not supported, traffic forecasting is suggested to be based on traffic flow and occupancy, either as final predictive outcome or for computing indirectly the travel time or speed (Williams, 2001; Stathopoulos and Karlaftis, 2003; Liu et al., 2008; Vlahogianni et al., 2008).

In parallel to the delineation of an optimum prediction horizon and interval and of the aim-corresponding suitable parameters, studies are focused on the optimum *prediction method* to be used. A variety of mathematical and artificial intelligent methods were employed in the past two decades that venture to capture efficiently the traffic dynamics, namely the abrupt fluctuations and peaks during the state transitions. These methods can be categorised statistically in parametric and non-parametric approaches.

The parametric techniques that were initially used for forecasting models, encompassed prediction approaches that were mainly developed based on historical average algorithms, which as the self-explanatory title implies are using an average of past traffic variables to forecast future values of the variables (Stephanedes et al, 1981; Kaysi et al., 1993; Smith and

Demetsky, 1997), and time-series models, such the Autoregressive Integrated Moving Average (ARIMA) models (Stephanedes et al., 1981; Kaysi, 1993 ; Jeffrey et al., 1987). As Smith and Demetsky discuss, ARIMA models are mathematical models that based on the stochastic processes of traffic conditions, explain the past behaviour of uninterrupted data series and then apply it to the forecast future behaviour. However, the predictions are neither stable nor representative of the rapid variations and the unexpected boundaries or edges of traffic (Davis et al, 1991; Hamed et al., 1995) that indicate transitional conditions (Addison, 2002) between free-flow, congested or near-congested states. To overcome this deficiency, researchers turned to another time series model category but of multivariate nature, the state-space models using Kalman filter algorithm that is based on successive updates of parameters. The multivariate character of the approach allows data from multiple detectors to be jointly considered and results in outperforming the ARIMA models in prediction accuracy, especially when the traffic data are classified into different time periods during a day (Stathopoulos and Karlaftis, 2003).

In order to mitigate the distributional constraints, in form of residuals, or assumptions for input or output variables upon transition of traffic states (Clark, 2003), recent studies utilized real-time datasets derived from ITS systems to forecast traffic conditions with non-parametric approaches. Non-parametric regression models and artificial neural network-based models (ANN) dominated over the last decade. Specifically, as non parametric regression is based on pattern recognition, states from similar traffic dynamic sites can be predicted, establishing a more stable and transferable method comparing to the ARIMA models category and to the historical average algorithms, with higher accuracy in the prediction of flow and occupancy in motorways than speed (Smith and Demetsky, 1997; Smith et al., 2002; Clark, 2003). Howbeit, ANN's models have been extensively used in numerous combinations with other algorithms and methods, demonstrating enhanced accuracy results, through their induction conception that permits likely generalisations and pattern recognition for multiple steps into the future with limited computational costs. Even though the functional relationship within neurons is non-linear, the more sophisticated approach leads to model highly non-linear relationships in multivariate setting and so to outperform classic linear statistical models (Clark et al., 1993; Kirby et al., 1997; Zhang et al., 1998). The models are data-driven, since the neurons (interconnected nodes) are connected in an input/output manner, namely the output data of one neuron (antecedent) are the input data of another (descendant) (Rokach, 2010). Numerous researches employed the most widely studied neural network: the multilayer feedforward network (Clark et al., 1993; Kwon and Stephanedes, 1994; Smith and Demetsky, 1994; Zhang, 2000; Vlahogianni et al., 2005). It is consisted of neurons, which are organised in input, hidden and output layers, and is based on multilayer units that are called perceptrons (MLP) and compute linearly the combination of its inputs, while in its simplest structure (single layer feedforward network) the perceptrons are invoking an activation function to transform the weighted sum into a binary output (Rokach, 2010).

The theory of neural networks was merged in hybrid approaches with rule-based fuzzy logic and genetic algorithms (GA), so that the optima network parameters emerge and the optima matches between input and hidden layers are detected. GA maintain a population of classifiers for the training part of a model and serve as search engines through multivariate complex spaces aiming to find the optimum classifier and hence the maximum statistical correlations via local minima that would indicate the optima connections between input an hidden layers (Lingras and Mountford, 2001; Adeli and Jiang, 2003). Empirical observations indicated that with the use of GA as an optimisation technique of the ANN, the high values of traffic parameters are better identified, the prediction model performs better and that computational effort is decreased in comparison to traditional ANN, since the input space is reduced by the effectuated generalisations (Yin et al., 2002; Ishak and Alecsandru, 2003). However, attention has to be drawn to the fact that GA may indicate erroneous traffic patterns as representative of the input dataset, because of over-heuristics enforcement (over-training) that produce poor classifiers, which chanced to perform well on the training dataset (Rokach, 2010), and thus decrease remarkably the prediction accuracy. In view to that and to the greater deficiency of efficient constructive methods for the implementation of ANN, in the context of determining the input parameters of the neurons and of selecting the network structure, wavelet networks surfaced to fill in this void of ANN models with integration of enhance the detection of traffic patterns even for non-recurrent conditions (Zhang, 1997; Zhang and Benveniste, 1992; Adeli and Karim, 2000; Karim and Adeli, 2002a; Karim and Adeli, 2002b; Jiang and Adeli, 2005).

In conclusion, the most studied non-parametric techniques, namely neural networks and non-parametric regression, outperform traditional statistical methods, such as historical average algorithms, as detect and predict remarkably better the extreme traffic conditions, the rapid fluctuations and the transitions between states (Smith and Demetsky, 1997). As various research results suggest, between ANN and ARIMA models, ANN are the more aspiring for predicting traffic variables particularly for multiple steps ahead, because of the adaptive nature of the non-parametric ANNs', even with the addition of new input data, that predict efficiently the rapid and high variations of traffic patterns and are more suitable for real and near-real predictions, mainly in highways (Kirby et al., 1997; Williams et al. 1998; Smith et al., 2000; Lint et al., 2005; Jiang and Adeli, 2005).

Nevertheless, the majority of the deployed forecasting models have not incorporated exogenous variables that would ensure the consistency of the model on predicting traffic conditions and evolution upon diverse traffic composition, incidents and weather conditions. Even though few empirical studies confirm and assess the weather impact (Kyte et al., 2000; Pisano and Goodwin, 2004; Agarwal et al, 2005; Chung et al., 2006; Billot et al., 2009), there are even fewer the models that predict adequately certain traffic parameters (speed or travel time) (Huang and Ran, 2003; Wiley, 2006; Butler et al., 2007; Faouzi et al., 2010) and none that encompasses the entirety of the aforementioned exogenous variables. Addressing to this under-researched part of traffic forecasting domain, the methodology that will be described in the next section will attempt to capture the traffic dynamics and predict traffic evolution under diverse traffic composition, incidents and weather conditions.

A comparative table of the most prominent methods presented in relevant literature is following in the Appendix (Table A.1).

3. Methodological Framework

Envisaging the adaptability and transferability of the proposed framework on networks with similar geometric characteristics and level of demand, methodologies that demand the heuristic definition of several coefficients were not preferred.

The approach will be centred on the development of a traffic congestion prediction methodology that will be able to provide real-time traffic forecasting upon multiple regimes and its transitions (free flow, near-congestion, congestion, recovery) incorporating also parameters for traffic prediction under various weather conditions (rain, no rain, snow, no snow), incidents, traffic compositions (categories based on the percentage of heavy vehicles) and day types (weekday, weekend/holidays). The predicted values of the traffic parameters will be utilised, aiming to mitigate the congestion incidence in highways, by developing a dynamic decision-making algorithm that allots the emergency lane to the traffic, whenever the network performance deteriorates. For the same reason, the emergence of an optimum control strategy and of predominant factors that have not been incorporated, will be pursued. Finally, the robustness of the proposed methodology will be assessed.

The following proposed framework is subject to further adaptations and will be finalized to a proposed methodology following the completion of the analysis. It is an approach to forecast traffic evolution with considering as exogenous factors the traffic composition, incidents and weather conditions. The proposed short-term traffic forecasting algorithm will be structured as multivariate, predicting recurrent and non-recurrent congestion multiple steps ahead, and will have a long prediction horizon with prediction interval as close to the traffic parameters evolution as possible. The transferability of the model will be pursued via a non-parametric methodology and will serve for traffic control and traveller information systems. To ensure the transferability and the variability adaptation of the traffic parameters a framework with an ANN prediction model is proposed. The proposed prediction model is planned to be applied in highways in real-time environments and given the disposability of highway traffic data, the traffic variables of flow, speed, and then occupancy and congestion duration is aimed to be predicted. The following table (Table 2) summarizes the processes, inputs, outputs and the proposed methodological tools that will be employed for the subsections of the research part in question.

Table 2. Summarizing table of processes, inputs, outputs and methodological tools for the proposed traffic prediction framework.

Process	Methodological Tool	Inputs	Outputs
1. Pre-process. Adressing missing data & outliers issue	Statistical indicators (MRE, SRE, RMSEP)	Flow and speed (Q & V) time series	Flow and speed (Q & V) time series without missing data
2.Traffic patterns recognition	Discrete Wavelet Transform with Multi-Resolution Analysis (DWT MRA)	Flow and speed (Q & V) time series without missing data	Flow and speed (Q & V) time series with singularities
3.Dominant traffic regimes detection	Fuzzy C-Mean Algorithm (FCM)	Flow and speed (Q & V) time series with singularities, exogenous variables,	Regimes of traffic flow
4. Traffic Conditions Forecasting	Artificial Neural Networks (ANN) with various: - structures - algorithms - no. of layers - prediction horizon - time step	Flow and speed (Q & V) time series with various regimes of flow, exogenous variables	Congestion prediction vector

3.1 Data Pre-process, Outliers and Missing Data Approach

Before detecting the various traffic states and its transitions in highways, which will lead to a better traffic evolution prediction, the issue of outliers and missing or corrupted data will be addressed. Missing or corrupted data occur following a malfunction of the equipment by producing either randomly missing data, when i.e. a very short-time and temporary cease of recording data is emerged due to communication disruption, or “blocks” of missing data, when the device ceases recording for a longer period of time for re-initialisation, maintenance reasons or structural damage. Additionally, as a result of noise measurement, unreliable data (outliers) may be recorded that may be witnessed throughout the dataset without a specific pattern. Each of these cases of missing data and all the combinations will be tested with interpolation and moving average and will be compared to the null replacement scenario via traditional statistical indicators (i.e. mean relative error-MRE (eq.1), standard deviation of relative error-SRE (eq.2), root mean square error proportional-RMSEP (eq.3) that as Van Lint et al. discussed, quantify efficiently the performance of each scenario (van Lint et al., 2005).

$$MRE = 100 \cdot \frac{1}{N} \sum \frac{y_n - o_n}{o_n} \tag{1}$$

$$SRE = 100 \cdot \sqrt{\frac{1}{N-1} \sum \left(\frac{y_n - o_n}{o_n} - \frac{MRE}{100} \right)^2} \tag{2}$$

$$RMSEP = \frac{100}{o_{mean}} \cdot \sqrt{\frac{1}{N} \sum \left(y_n - o_n \right)^2} \tag{3}$$

where N is the total number of observations, y_n is the n^{th} predicted value of the n^{th} input, o_n is the n^{th} predicted value of the n^{th} output, o_{mean} is the predicted mean and y_n-o_n is the prediction error for data pattern n (van Lint et al., 2005).

It is noted that wavelets are not strongly envisaged as potential denoising method, as the proposed prediction model aims to be used in a real-time system environment (Adeli and Karim, 2000) and wavelet filter coefficients need heuristic and subjective criteria to be used in order to provide a simplified linear combination of elementary functions, rendering the data-driven model intransferable to other sites.

3.2 Traffic Patterns Recognition with Wavelet Analysis

The detection of traffic patterns will be firstly addressed, by identifying the important parts of the dataset, where the traffic states and potentially its transitions may subsist, in order that in the subsequent phase of the study the search will be correctly oriented and with reduced size of data. Via the analysis of abrupt fluctuation in the measurements, produced by traffic variables (flow, speed), weather conditions (rain, no rain, snow, no-snow), incident occurrence (incident, no-incident), traffic composition (% of heavy vehicles), traffic direction and day type (weekday, weekend/holidays), the regions of the dataset where the dominant states may occur will be unveiled.

Regarding the properties and evolution of traffic demand, hence the properties of times series scaled measurements of traffic flow in a fixed spatial point, research led to the conclusion that are characterised by periodicity or seasonality depending on the day or the time period. The transitions between traffic states are depicted as irregular structures (*singularities*) with noticeable fluctuations along a dataset and as they are phenomena with scaling behaviour, a dyadic sequence of scales can be applied using Discrete Wavelet Transform (DWT) Wavelet Multi-Resolution Analysis (WMRA), in order to localise characteristics in space/time and frequency/scale for flow and speed through wavelet decomposition (Mallat and Hwang, 1992; Addison, 2002; Jiang and Adeli, 2004; Vlahogianni, 2008).

Wavelet analysis was selected amongst other artificial intelligence techniques, and in particular the Fourier analysis that was vastly used until recently, because of the following aspiring points. First, it is more efficient in identifying patterns and sharp transitions in time and scale in non-linear input signals, as the traffic datasets. Furthermore, wavelets are more adept in representing discontinuities in chaotic time series, which could be the case for the kind of datasets in question, as the signals can be decomposed hierarchically and so analysed in various levels of details and in various components of characteristics. Lastly, even though wavelets are not limited to stationary inputs, the required storage for representing time series is small and consequently the computational time is sufficiently brief for real-time applications. Wavelet transforms of a signal/dataset evolving in time depend on two variables: the time and the local scale/frequency as a measure of similarity, thus provide time and frequency localisation for continuous time signals with discrete values (Adeli and Karim, 2000; Jiang and Adeli, 2004; Daubechies, 2006). They will be used to subset the input data into blocks of frequency components and to examine each one with a resolution corresponding to its scale in different levels of detail, so as to detect singularities, namely specific traffic patterns with an important rapid variation at each of them, herein corresponding to flow or speed, throughout a wide scale range.

The *Discrete Wavelet Transform (DWT) of traffic flow* can be written as f^t of time variable *t*, in terms of a series of orthonormal wavelet basis functions as follows in equation 4:

$$\begin{gathered} W_{f_{i,j}} = 2^{-j/2} \int_{-\infty}^{\infty} f^t \psi\left((2^{-j})^t - k\right) dt, \\ j,k \in \mathbb{Z}, \psi \in L^2(\mathbb{R}) \end{gathered} \tag{4}$$

The coefficient ψ is the scaled and translated version of the elementary *wavelet function* called *mother* or *generating wavelet*, which is presented in equations 5 and 6 in Continuous Wavelet Transform (CWT) (Daubechies, 1992), and describes the original signal at time $b_{j,k} = 2^{-j} k$ at the frequency band $a_j = 2^j \neq 0$, where parameters k denotes the time index, j the frequency index, both belonging to the $\mathbb{Z}$ set of integers, a the frequency (or scale) and b the temporal content (or space or dilation) location, both belonging to the $\mathbb{R}$ set of real numbers.

$$\psi_{a,b}(t) = \frac{1}{\sqrt{|a|}}\psi\left(\frac{t-b}{a}\right), \quad a,b \in \mathbb{R}, \psi \in L^2(\mathbb{R}) \tag{5}$$

that by replacing the a and b the following wavelet expansion function in DWT:

$$\psi_{j,k}{}^{t} = 2^{-j/2}\psi(2^{-j}t-k), \quad j,k \in \mathbb{Z}, \psi \in L^2(\mathbb{R}) \tag{6}$$

The notation $L^2(\mathbb{R})$ corresponds to the square summable time series space of traffic flow f, where the superscript 2 denotes the square of the modulus of the function. When for a certain temporal content b the frequency a in a subset decreases, then $\psi\left(\frac{t-b}{a}\right)$ narrows in width while maintaining the same space, as it is examined for a certain b, emphasising a sharper area and thus indicating a potential singularity. The DWT representation of traffic flow in eq. 4, can be also regarded as a subsampling of the CWT coefficients with dyadic scales, e.g. 2^{j-1} for $j = 1, 2, ..., N$, where N is the maximum number of decomposition level and thus less accurate than the CWT (Jiang and Adeli, 2004). Nevertheless, the reason that DWT was selected for the current methodology addressing a real-time application, is that the required number of coefficient vectors is smaller and thus the integration of every version of values for the scale and dilation, a and b respectively, demands less computational effort, is less time-consuming and reduced in size.

Regarding the properties of wavelets, it has to be noted that they have to be *non-orthonormal*, in case of missing data (Daubechies, 2006), and *non-orthogonal* so as to ensure the non-independency of each wavelet, in order that their integration in the ANN to be rendered possible. The redundant wavelet coefficients (*bases*) that emerge from DWT, or else the sets of independent vectors, will be further referred to as wavelet frames. Each frame contains supernumerary wavelets for a signal reconstruction and consequently renders the traffic forecasting more accurate, as encompasses more possible subsets that allow the emergence of a more representative vector. A multiresolution analysis will be evoked for various levels of detail/resolution (Mallat, 1989), which demands the implication of a sequence of nested closed approximation scaling function subspaces V_i with $i \in \mathbb{Z}$ (space of integers) that as result of the dilations and translations of the scaling functions, have the following properties (Zhang and Benveniste, 1992; Adeli and Karim, 2000; Daubechies, 1988, 1992):

$$.... \subset V_{-2} \subset V_{-1} \subset V_0 \subset V_1 \subset V_2 \subset ... \subset L^2(\mathbb{R})$$

$$\overline{\bigcup_{i \in \mathbb{Z}} V_i} = V_{j_0} \oplus W_{j_0} \oplus W_{j_0+1} \oplus ... = L^2(\mathbb{R}), \qquad \bigcap_{i \in \mathbb{Z}} V_i = \{0\} \tag{7}$$

That is, $V_i (i \in \mathbb{Z})$ is a subspace of V_{i+1} and V_i is defined by spanning $L^2(\mathbb{R})$ by the *scaling functions* of *father wavelet* (eq. 8):

$$\varphi_k^t (k \in \mathbb{Z}) \text{ as } V_i = \overline{\underset{k}{Span}\{\varphi_k^t\}}, \tag{8}$$

with the overbar indicating the closure of the subspace. The wavelet subspace is W_i the orthogonal complement of V_i in V_{i+1}, so the entirety of the time series space is represented as the summation of all subspaces, with $\oplus$ indicating the direct space sum and the integer subscript j_0 any starting scaling parameter index from negative infinity to positive infinity, including zero. Furthermore, the scaling function subspace V_i satisfies the natural scaling condition (eq. 9):

$$f^t \in V_i \Leftrightarrow f(2^t) \in V_{i+1} \tag{9}$$

where traffic flow time series have a membership to the subspace V_i and are scaled by a dyadic factor in the next subspace, that is a downsampling by a factor of 2 is effectuated.

Consequently, based on the aforementioned properties, the shifted scaling father function $\varphi(2^t)$ and the mother wavelet ψ^t (eq. 10) for the solution of the DWT of the traffic flow expressed in MRA (eq. 11), are as follows:

$$\psi^t = \sum_n h^n \sqrt{2\varphi}\left(2^t - n\right), \quad n \in \mathbb{Z} \tag{10}$$

$$f^t = \sum_k s_{j_0}^k \varphi_{j_0}^t + \sum_k \sum_{j=j_0} w_j^k \psi_{j,k}^t \tag{11}$$

where n is the order of the wavelet function and h^n is a sequence of n real or complex numbers, called the *scaling function coefficients,* or else the *scaling filters*, with values resulted from any given type of wavelet function satisfying the two fundamental wavelet properties: *a)* the integral of ψ^t is zero and *b)* the integral of the square of ψ^t is unity (Daubechies, 1988; WUTAM, 1998; Jiang and Adeli, 2004). The first term of eq. 11 is the coarse resolution at scale j_0 and the second term is the frequency and time breakdown of the signal (Jiang and Adeli, 2004).

After the data analysis, the optimum wavelet function and depth of wavelet analysis will be selected, so as the singularities of the inputs – time series of flow and speed – and hence only the noteworthy parts of the dataset emerge more accurately, allowing to proceed to the following part of the research, the detection of the dominant traffic states.

3.3 Dominant Traffic States/Regimes and Factors Detection with Fuzzy Approach

From the reduced in dimensions dataset that resulted by the Discrete Wavelet Transforms Wavelet Multi-Resolution Analysis (DWT WMRA), certain singularities in the traffic evolution have been unveiled, indicating the parts that traffic regimes may occur. An efficient classification is proposed to be invoked through a rule-based fuzzy data clustering, by which the dominant states will be emerged without losing important characteristics. Even though several methodological techniques have been used in literature, which vary in compact support and smoothness, a hybrid method of wavelet and fuzzy analysis was selected, as the computational effort will be reduced and a more robust model will be established. The emergence of this approach was evoked, following the aforementioned literature review and attempts of dominant regimes identification with the use of genetic algorithms, which suggested erroneous traffic patterns as representative of the input dataset, because of premature convergence and over-training that led to the definition of poor classifiers, which chanced to perform well on the training set, and thus would have decreased remarkably the prediction accuracy.

The dataset can be classified into homogeneous clusters, which their *objects* are traffic flows, speeds, the aforementioned exogenous parameters and the singularities, that correspond to certain traffic states. These parameters form a set of vectors $X = \{x_1, x_2, x_3, ..., x_n\}$. A representative set $Z = \{z_1, z_2, z_3, ..., z_k\}$ of z classes in X is produced by a data clustering technique, where x, z $\in \mathbb{R}^p$, thus the most representative traffic states. Among the most efficient fuzzy algorithms, namely the Fuzzy C-Mean algorithm (FCM) (Dunn, 1973; Bezdek, 1981), applied by Adeli and Karim for freeway incident detection within a neural network (Adeli and Karim, 2000), and the Fuzzy K-Mean algorithm (FKM) (MacQueen, 1967), applied by Vlahogianni et al. for traffic flow regimes detection in signalised arterials (Vlahogianni et al., 2008), the FKM will be employed for this classification, as the FCM demands a heuristic definition of coefficients that renders the prediction model unsuitable for real-time application.

Consequently with fuzzy partitioning, one traffic pattern is allowed to belong to more than one cluster with a different membership degree from 0 to 1 with 1 denoting the maximum membership to the cluster, rather than a crisp assignment of one traffic pattern exclusively to one cluster. A number of *k* clusters are a priori defined with *k* centroids, which must be placed considering that they represent the traffic states, which will form the initial-seed groups centroids. Each object/set of parameters/initial representative of a traffic state is assigned to the nearest centroid, until all objects are associated. To the new classes that have been created, k new centroids are located and each object is re-assigned to a new centroid. The loop is terminated when the location of centroids is fixed. This algorithm aims to minimize the following squared error objective function (eq. 12):

$$J_{\beta}(z) = \sum_{i=1}^{n}\sum_{j=1}^{k} A_{ij}^{\beta} \left\| \mathrm{x}_i - \mathrm{z}_j \right\|^2 \tag{12}$$

subject to:

$$\sum_{j=1}^{k} A_{ij} = 1, \quad 1 \le i \le n \tag{13}$$

$$A_{ij} \ge 0, \quad 1 \le i \le n, \quad 1 \le i \le k, \; 2 \ge \beta > 1 \tag{14}$$

where J_β is the objective function for given value of data fuzziness degree β, with 2 indicating the greatest fuzziness in the dataset, A_{ij} the membership degree of vector i in class j, k the number of classes, $\left\| \mathrm{x}_i - \mathrm{z}_j \right\|^2$ is a chosen distance measure between data point x_i and the clusters' centers z_j , and z_j are the clusters' centers that indicate the distance between the n data points from their assigned cluster centers, computed by the equation 15 (MacQueen, 1967; Adeli and Karim, 2000):

$$\mathrm{z}_j^t = \frac{\sum_{i=1}^{n} A_{ij}^{\beta} x_i}{\sum_{i=1}^{n} A_{ij}^{\beta}} \tag{15}$$

The algorithm is composed as follows:

1. Set k points that represent the initial group of traffic states (objects) and the initial seed group of centroids, in respect to the limit of equation 13. Compute the cluster center z_j for each class (eq. 15)
2. Assign each object to the group that has the closest centroids.
3. After the assignment of the last object, recompute the positions of the k centroids.
4. Repeat steps 2 and 3 until the centroids location remains fixed.

The procedure always terminates, and regarding its sensitivity to the initial random selection of seed-points as cluster centers, it can be reduced after multiple iterations. The algorithm results to a classification of the aforementioned objects into groups of one traffic regimes each, and consequently to the dominant traffic regimes of the dataset, reducing also the dimensionality of the dataset.

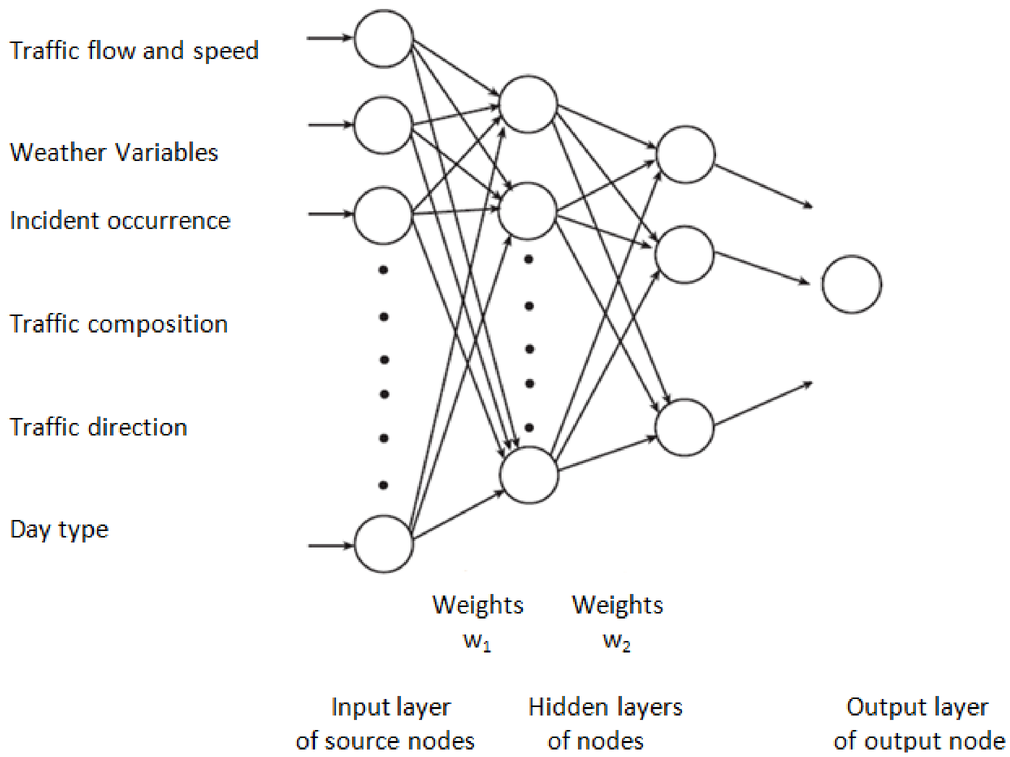
3.4 Traffic Conditions Prediction with Wavelet Neural Network

The development of an efficient and transferable model for traffic prediction on highways will be addressed in this part, as the majority of traffic prediction models in existing literature are oriented to forecast in urban areas that the traffic evolution is expressed more limitedly, and thus they are not representative for the rapid fluctuations that are observed in highways. The data-driven approaches of neural networks have been proven accurate traffic predictors (Chen et al., 2001; Ishak & Alecsandru, 2003), so for modeling and forecasting traffic conditions' evolution in highways, and in particular traffic congestion, the Artificial Neural Network (ANN) theory will be explored, in order to unveil a potential approach for employment. The reason for not searching an approach among the common time series and several other non-parametric and parametric methodologies and heuristic methods, as the genetic algorithm or the trial-and-error, lies in the resulted overpredicted parameters and the unstable and not representative predictions of the rapid and high variations of traffic parameters, which indicate transitional traffic conditions (Davis et al., 1992; Smith and Demetsky, 1997; Hamed et al., 1995; Addison, 2002). Regarding the genetic algorithms, empirical observations indicated erroneous traffic patterns as representative of the input dataset, because of over-heuristics enforcement that produced non-representative classifiers and subsequently decreased remarkably the prediction accuracy. Therefore, the non-linearity of traffic conditions is inducing the development of a neural *architecture*, as it has the ability to solve complex problems with non-linear functions, and of an automatic adaptive *algorithm* that will have forecast efficiently real-time the traffic congestion upon multiple regimes in highways. The proposed methodology is based on a hybrid of wavelet and neural network models (WNN), aiming to utilise the advantages of both fields on several structures and algorithms and on prediction horizons and time steps, as neural networks can be trained incrementally.

A neural network is described by Haykin as a parallel distributed processor that temporarily stores, via synaptic weights, experiential knowledge acquired through a learning process. During this process the free parameters of the ANN, such as the synaptic weights and bias levels, are formed as a result of the constant incitation by the environment in which the ANN is embedded. Namely, the type of learning that will be selected for the research, depends on the behaviour of the parameter in question (Haykin, 1999). Architecturally, as Rokach cites, the development of a model by ANN is attained with a network of interconnected units – in the sense of inputs-outputs – hereinafter called *neurons* or *nodes* (Anderson and Rosenfeld, 2000; Rokach, 2010). The nodes are organized in three layers (input, hidden, output) and are connected via the hidden layers of neurons that serve as extractors of the features of the input data by using a nonlinear function without loops, also referred to as direct cycle. The neurons in the hidden layer are connected to both the input and the output neurons and they are the key to the activation of the classifier. In order to compute the output of a single neuron, the weighted sum of inputs to the neuron is calculated, then the bias is added to the sum, which is finally given as input to the activation function of the neuron. In the simplest structure of feedforward NNs, the single-layer *perceptron* network – a classifier that calculates a linear combination of its inputs and invokes an activation function that transforms the weighted sum into binary output – with one-direction nodes and a single layer of outputs, has the same analytical function as the logistic regression model, and is also known as sigmoid function. Subsequently this is not an adequate ANN structure for the current research. Even though the input space was reduced as subject to the previously presented parts of the research, since the information is spread across many attributes (traffic parameters time series and exogenous parameters in various traffic regimes, for both upstream and downstream sections, so as to predict incoming traffic and back propagating shockwaves), it remains complex and it is required respectively a more sophisticated structure that ANNs can provide and for which were selected for the prediction methodology.

Firstly, the structure of the multiple-layer perceptron network with different algorithms will be explored. In figure 1 is illustrated the topology of a representative multi-layer feedforward (MLF) neural network adjusted to the current research. This ANN architecture consists of interconnected and usually feed-forward way nodes, and although the dominant supervised learning algorithm in the literature was the back propagation algorithm (BP), a gradient descending method for training the network, herein it will only be addressed as a part of the comparative study of ANN architectures and algorithms for the appointment of a more efficient algorithm, since it cannot be applied in real-time ATM systems.

Figure 1. Graphic representation of a multi-layer feedforward (MLF) neural network with one input layer, one output layer, two hidden layers and their weights, adjusted in the research in question.



Source (of the neurons part of the image): Rokach, 2010.

Recent studies have presented a more efficient structure for the multilayer feedforward NN, especially for real-time applications, with several aspiring algorithms and results more accurate in comparison to the BP. Even though the feedforward networks have static structure by definition, by incorporating the time as dimension into the design of an ANN implicitly (Tapped Focus Lagged Feedforward Network – TLFN) or explicitly (recurrent networks), i.e. as a short-term memory into their input layer, they transformed into non-linear dynamic models with the advantage of enhanced stability (Haykin, 1999). These approximating networks have two-stage structure, a linear preprocessing phase and a memoryless nonlinear network of basis transfer functions, followed by the incorporation of the time dimension. For the nonlinear part, simple nonlinear structures can be used, such as the radial basis function (RBF), the sigmoidal etc., which learn the input space and transform its vector to the output space. Haykin refers to a better approximation of large classes, based on the observation that biological neurons are more receptive to activation when the input is closer to the center of the field (Adeli and Karim, 2000), feature that is promising for more efficient forecasting results, given the large datasets of the research in question.

Another structure that by empirical observations it is promising to outperform the MLF NNs, as forecasts with greater accuracy the time-dependent traffic flow along with the entity of the exogenous parameters, is the Time Delay Neural Network (TDNN). TDNN uses BP techniques for the weights of the neurons, but does not employ only single connections between nodes, as the majority of ANNs. Instead it establishes multiple connections that affect the hidden and output layer and resulting in saving in memory previously developed spatio-temporal patterns and not only inputs. Furthermore, the Adaptive Time Neural Network (ATNN) is an even more aspiring NN, as during the learning process it seeks phase relationships that produce higher correlation over history and automatically adapts its intervals (Abdulhai et al., 1999).

Regardless the structure and the applied algorithm, in order to estimate the parameters of the forecasting neural network, a training phase is included that calculates the connection weights, which optimize a given evaluation function of the training data. Various search methods can be used to train these networks, of which the most widely applied one is back propagation (BP) (Rumelhart, 1986). This method propagates values of the output evaluation function backward to the input, allowing the network weights to be adapted so as to obtain a better evaluation score. Radial basis function (RBF) networks employ Gaussian nonlinearity in the neurons (Moody and Darken, 1989), and can be seen as a generalization of nearest neighbour methods with an exponential distance function (Poggio and Girosi, 1990).

For the final emergence of the most suitable type of ANN for traffic forecasting in highways, a comparison of all the aforementioned supervised and unsupervised learning algorithms will be effectuated in terms of accuracy. The TDNN and ATNN is expected to outperform static MLF-type NNs as they contain in the forecasting process spatial patterns that are time-evolving, as the traffic parameters behaviour that is to be predicted.

Finally, the proposed methodology will be compared to simple deterministic models (non-parametric regression etc.), to assess if the computational cost will be considerably equivalent to the accuracy improvement. According to Lu (Lu et al., 1996), there are deficiencies in the application of neural networks, namely the difficulty in interpreting the model, in incorporating prior knowledge about the application domain, and also long training time, both in terms of CPU time, and of manually finding parameter settings that will enable successful learning i.e. optimize the evaluation function. Nevertheless, due to the spatial coefficient that is not incorporated in the time series approach and to the time coefficient that is not comprised in the static MLF-type NNs, the time delay-type NNs are envisaged to be more accurate in traffic forecasting.

Regarding the proposed prediction horizon and the time step that is considered to be selected for the prediction framework in question, two are two rules of thumb that tend to be efficient when followed:

a) the larger the prediction horizon, the less the accuracy prediction
b) the shorter the time step/interval of the forecasting model, the more difficult is the prediction effectuated, because of the rapid fluctuations of traffic parameters in short time periods

However, based on implementations of traffic forecasting models (Florio and Mussone, 1996; Smith and Demetsky, 1997) and the range of traffic parameters in European highways, the time interval is suggested to not be too short (i.e. less than 30’’), as the traffic parameters have variability and the fluctuation in short time interval would result in decreasing the accuracy prediction. The optimum prediction horizon and time interval will be defined following the completion of the analysis from a range of 30-seconds, 1-minute, 2-minutes, 4-minutes, 5-minutes, 10-minutes, 15-minutes for the horizon prediction and 30-seconds, 2-minutes and 5-minutes for the time interval.

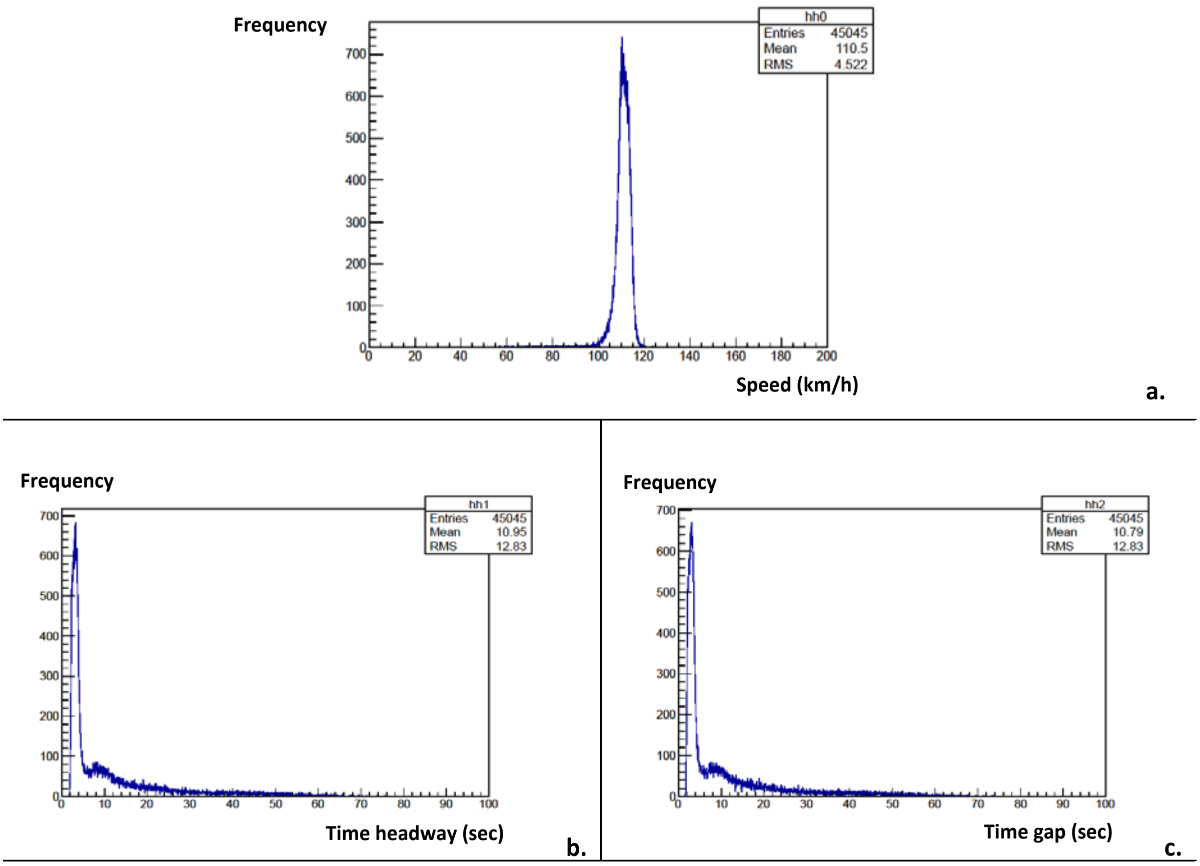
4. Initial Investigation of Naïve Forecasting Models – Case Study Results

In an initial attempt to comprehend the behaviour of forecasting models, simple deterministic models were evoked to be analysed with a dataset of disaggregated speed, time headway and time gap collected from a loop detector on A1 Swiss highway between Geneva and Lausanne, for 365 days of year 2010.

The data were aggregated per 10-minutes, since the daily average time headway during peak hours was most commonly of 30-minutes and a lower than 15-minutes prediction horizon was considered at that primal stage as adequate. To find the best matches in the observations in terms of traffic flow patterns and time period classification (weekday, weekend), the Nearest Neighbour method was applied. In addition, two simple deterministic models were developed as extrapolations of historical data, to set a measure to assess the degree of forecast accuracy of this research proposed model. In the first naïve model, the current traffic speed and headway is used as the predicted value and in the second, the average of the current and the four previous speeds and headways as its prediction.

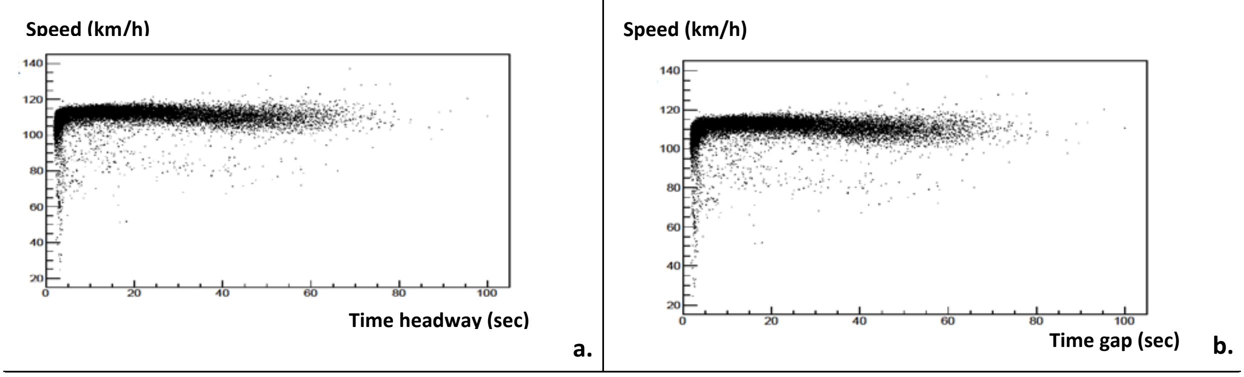
In figure 2, are presented the abovementioned data aggregated by 10-min, in order to define via the Root Mean Square (RMS) the weight of each variable in the calculation of the mahalanobis distance, that would be used in the Nearest Neighbour method. The use of weights was aiming to normalise the different magnitudes of each variable, resulted from the RMS of speed and of headway distribution, namely $w_s=4.5\ km/h$ $w_h=12.8 sec.$

Figure 2. a. Speed, b. time headway and c. time gap distributions for 10-min aggregated traffic data of 2010 from detector 24 on A1 Swiss motorway (Data courtesy of OFROU).



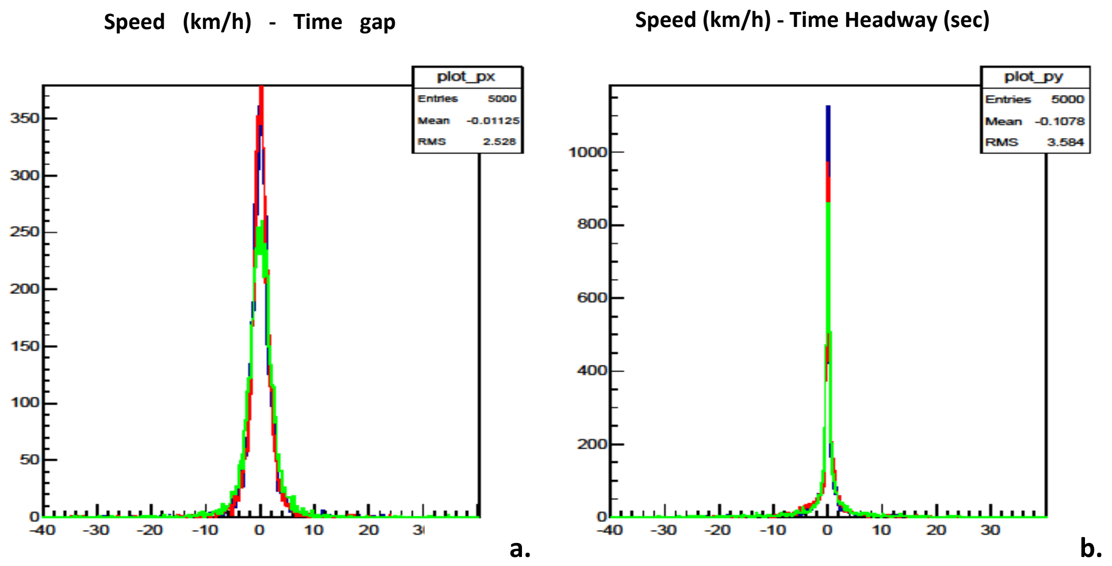
In figure 3 is demonstrated that the speed and the time gap/or headway variables are independent, as for various values of time gap and headway the speed does not alter its behaviour.

Figure 3. a. Speed - headway and b. speed - time gap relationship scattergrams for 10 min aggregated traffic data of 2010 from detector 24 on A1 Swiss motorway. (Data courtesy of OFROU)



As it is indicated from the comparison of the three tested methods in the following figure (Figure 4), without any conditions imposed to refine the input and examine the interactions upon multiple regimes and exogenous parameters, the predictions are characterized by high inaccuracy. Even though the steepness of the headway graph (Fig. 4.b.), depicts a better predictability for the headway variable, an all-day forecast misleads the assessment of the methods. The case of restricting the input data, using a subset around the period of peak hours (before, after and during), is expected to improve the prediction accuracy.

Figure 4. Error of prediction for a. speed - time gap and b. speed - headway relationship scattergrams for three traffic prediction methods: time series using one-step (red line) and five-step (green line) and Nearest Neighbour (blue line) for 10 min aggregated historic data of 2010 dataset from detector 24 on A1 Swiss motorway. (Data courtesy of OFROU)



5. Conclusions and Perspectives

Amongst the various prediction methods that are presented in current literature, the most adaptable regarding the main objective that was targeted for the scope of traffic forecasting of this study and the available data, were selected to be implemented. To provide a baseline for the traffic forecasting potentials of these methods, deterministic forecasting methods were employed to be compared with, in order to assess whether the computational effort reciprocates in accuracy. The implementation of the simple deterministic models of time series and nearest neighbour affirmed the expected results. With the application of an unconditional approach, the headway and speed predictions derived from both methods, which were not remarkably dissimilar, were characterised by sizable errors of 10 seconds per 10 minutes aggregated data for daily observations during 2010. Following, stochastic methodological approaches will be examined, contemplating to unveil a representative, consistent and transferable method, able to anticipate network's behavioural alterations triggered by weather conditions, traffic composition and incidents, that will be part of the proposed methodology, in order to optimally form it and provide real-time traffic evolution dynamic forecast in highways, that will increase operational performance and enhance predictability, acknowledging multiregime (congestion, near-congestion, free flow), transitional traffic behaviour and the entirety of exogenous parameters that are remised from current ATM control strategies. This valuable model is envisaging mitigating congestion emergence and moderating its impacts while maintaining traffic safety, without any additional costs, since it is using optimally the current infrastructure. Furthermore, its transferability is ensured in view of its structure, permitting to be implemented in ITS (Intelligent Transportation Systems) and ATIS (Advanced Traveller Information System) environments for highways with negligible modifications.

Following the validation of the model, the modification of existent traffic behavioural models will be pursued, by incorporating the emerged one into existing micro-simulations tools that will be able hereinafter to evaluate weather-responsive strategies. Furthermore, ATM highway control management schemes will be formed in microscopic scale, in order to moderate congestion emergence where a hard shoulder/emergency lane is non-dynamically operating. A dynamic algorithm will be implemented that incorporates exogenous parameters to the current procedure, avoiding unstable or forced flow and maintaining correspondingly traffic safety. In a future part of the research, and in parallel to the sensitivity analysis of the proposed methodology, a hybrid ATM approach will be suggested to confront the persistent recurrent and non-recurrent congestion, as a combination of applied and tested individual strategies in highway systems of Switzerland and abroad.

Among the deliverables of the research will be the development of a decision support system and hybrid traffic control strategies for the dynamic operation of hard shoulder in highways, for further network performance amelioration and depression of traffic congestion occurrence in the same safety level. In addition, an Application Programming Interface (API) that re-structures the existing traffic behavioural models of micro-simulation analysis, encompassing a set of predominant exogenous parameters dialectical to the traffic evolution, will be derived at the end of the research. Therefore, the models will be rendered more relevant to the adversities that a network encounters, degrading its performance and impeding its operations, and will represent more accurately traffic evolution for future exogenous-responsive traffic management strategies.

6. References

Abdulhai, B., Porwal, H. and Recker, W. (1999) *Short-term Freeway Traffic Flow Prediction Using Genetically-optimized Time-delay-based Neural Networks.* UCB, UCB-ITS-PWP–99–1, Berkeley, CA: Institute of Transportation Studies, University of California, Berkeley.

Adeli J. and Karim A. (2000). Fuzzy-wavelet RBFNN Model for Freeway Incident Detection. *Journal of Transportation Engineering*, 126(6), pp. 464-471.

Adeli H. and Jiang X. (2003). Neuro-Fuzzy Logic Model for Freeway Work Zone Capacity Estimation. *Journal of Transportation Engineering, ASCE,* 129(5), pp. 484–493.

Addison, P. S. (2002). *Illustrated wavelet transform handbook: Introductory theory and applications in science, engineering, medicine and financ*e. Institute of physics publishing, Bristol, U.K., chapter 5.

Agarwal M., Maze T. and Souleyrette R. (2005). Impacts of Weather On Urban Freeway Traffic Flow Characteristics And Facility Capacity. *Proceedings of the 2005 Mid-Continent Transportation Research Symposium*, Ames, Iowa, August 2005.

Anderson J.A. (1995). *Introduction to Neural Networks*. Cambridge, MA: MIT Press.

Arsham H. (2011). *Deterministic Modeling, Linear Optimization with Applications.* http://home.ubalt.edu/ntsbarsh/opre640a/partviii.htm. Accessed 11.10.2011.

Bezdek J. C. (1981). *Pattern Recognition with Fuzzy Objective Function Algorithm*s. Plenum Press, New York.

Billot R., Faouzi N.-E.El, Sau J. and deVuyst F. (2009). Integrating the Impact of Rain Into Traffic Management: Online Traffic State Estimation Using Sequential Monte Carlo Techniques. *Transportation Research Record*, 2169, pp. 141-149.

Butler S., Ringwood J. and Fay D. (2007). Use of Weather Inputs in Traffic Volume Forecasting. Proceeding of Irish Signals and Systems Conference.

Clark S. D., Dougherty M. S. and Kirby H. R. (1993). The Use of Neural Networks and Time Series Models for Short Term Traffic Forecasting: A Comparative Study. *Transportation Planning Methods*: Proceedings, *PTRC 21st Summer Annual Meeting*.

Clark S. (2003). Traffic Prediction Using Multivariate Nonparametric Regression. *Journal of TransportationEngineering*, 129(2), pp. 161–168.

Chen H. and Grant-Muller S. (2001). Use of sequential learning for short-term traffic flow forecasting. *Transportation Research Part C: Emerging technologies*, 9(5), pp. 319–336.

Chen H., Grant-Muller S., Mussone L., Montgomery F. (2001). A Study Of Hybrid Neural Network Approaches And The Effects Of Missing Data On Traffic Forecasting. *Neural Computing and Applications,* 10(3), pp. 277-286.

Chung E., Ohtani O., Warita H., Kuwahara M. and Morita H. (2006). Does Weather Affect Highway Capacity? *Proceedings of 5th International Symposium on Highway Capacity and Quality of Service - Country Reports and Special Session Papers,* 1, pp. 139-146.

Dailey D. J. (1997). Travel Time Estimates Using A Series Of Single Loop Volume And Occupancy Measurements. *Proceedings of the Transportation Research Board 76th annual meeting*, Washington DC.

Daubechies I (1988). Orthonormal Bases of Compactly Supported Wavelets, *Communication on Pure and Applied Mathematics*, 41, pp. 909-996.

Daubechies I. (2006). *Ten Lectures on Wavelets.* Philadelphia: SIAM, 1st ed. 1992, 2nd ed. 2006.

Davis G. A., Niham N. L., Hamed M. M. and Jacobson L. N. (1991). Adaptive Forecasting of Freeway Traffic Congestion. *Transportation Research Record*, 1287, pp. 29–33.

Dia H. (2001). An Object-Oriented Neutral Network Approach to Short-Term Traffic Forecasting. *European Journal of Operational Research*, 131, pp. 253–261.

Dougherty M.S., Kirby H.R. and Boyle R.D. (1994). Using Neural Networks To Recognise, Predict And Model Traffic. *Artificial Intelligence Applications to Traffic Engineering,* VSP, Utrecht, The Netherlands, pp. 233-250.

Dougherty M. S. and Cobbet M. R. (1997). Short-Term Inter-Urban Traffic Forecasts Using Neural Networks. *International Journal of Forecasting*, 13, pp. 21–31.

Dunn J. C. (1973). A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters. *Journal of Cybernetics*, 3, pp.32-57.

Durbin J. (2000). The Foreman Lecture: The State Space Approach to Time Series Analysis and its Potential for Official Statistics. *Australian and New Zealand Journal of Statistics*, 42(1), pp. 1–23.

Faouzi N.-E.El, Billot R. and Souzebda S. (2010). Motorway Travel Time Prediction based on Toll Data and Weather Effect Integration. *IET Intelligent Transportation Systems*, 4(4), pp. 338–345.

Federal Highway Administration – FWHA (2011). *Road Weather Management Overview*, http://ops.fhwa.dot.gov/Weather/overview.htm, http://ops.fhwa.dot.gov/weather/weather_events/rain_flooding.htm, Accessed 31.08.2011.

Florio L. and Mussone L. (1996). Neural Network Models for Classification and Forecasting of Freeway Traffic Flow Stability. *Control Engineering Practice*, 4(2), pp. 153–164.

Geroliminis N., Shrivastava A., Michalopoulos P. (2010). A Dynamic Zonebased Coordinated Ramp Metering Algorithm with Queue Constraints for Minnesota's Freeways. *13th International IEEE Annual Conferenceon Intelligent Transportation Systems*, Madeira Island, Portugal, September 19-22, 2010, pp. 1456–1461.

Hamed M. M., Al-Masaeid H. R. and Bani Said Z. M. (1995). Short-Term Prediction of Traffic Volume in Urban Arterials. *ASCE Journal of Transportation Engineering*, 121(3), pp. 249–254.

Harvey A.C. (1984). A Unified View of Statistical Forecasting Procedures. *Journal of Forecasting*, 3(3), pp.245–283.

Haykin S. (1999). *Neural Networks: A Comprehensive Foundation*. Englewood Cliffs, NJ: Prentice-Hall, 2nd ed.

Henson M.A. (1998). Nonlinear Model Predictive Control: Current Status and Future Directions. *Computers & Chemical Engineering*, 23(2), pp. 187-202.

Hoogendoorn S. & Bovy P. (1998). New Estimation Technique for Vehicle-Type-Specific Headway Distributions. *Transportation Research Record: Journal of the Transportation Research Board* (TRB), 1646, 18-28 p.

Huang S. and Ran B. (2003). An Application of Neural Network on Traffic Speed Prediction Under Adverse Weather Condition. *TRB 2003 Annual Meeting CD-ROM.*

Huang S., Sadek A.W. (2009). A novel forecasting approach inspired by human memory: The example of short-term traffic volume forecasting. *Transportation Research Part C: Emerging technologies*, 17(5), pp. 510-525.

Innamaa S. (2000). Short-Term Prediction of Traffic Situation Using MLP-Neural Networks. *Proceedings of the 7th World Congress on Intelligent Transportation Systems*, Turin, Italy.

Ishak S. and Al-Deek H. (2002). Performance Evaluation Of Short-Term Time-Series Traffic Prediction Model. *Journal of Transportation Engineering*, 128 (6), pp. 490–498.

Ishak S. and Alecsandru C. (2003). Optimizing Traffic Prediction Performance of Neural Networks Under Various Topological, Input, And Traffic Condition Settings. *Transportation Research Board Annual Meeting CD-ROM*, Washington, DC, USA.

Jeffrey D. J., Russam K. and Robertson D. I. (1987). Electronic Route Guidance by Autoguide: The Research Background. *Traffic Engineering and Control*, 28(10), pp.525-529.

Jiang X. and Adeli H. (2004). Wavelet Packet-Autocorrelation Function Method for Traffic Flow Pattern Analysis. *Computer-Aided Civil and Infrastructure Engineering*, 19, pp. 324–337.

Jiang X. and Adeli H. (2005). Dynamic Wavelet Network for Traffic Flow Forecasting. *Journal of Transportation Engineerin*g, ASCE, October 2005, pp. 771-779.

Karim A. and Adeli H. 2002a. Comparison of the Fuzzy—Wavelet RBFNN Freeway Incident Detection Model with the California Algorithm. *Journal of Transportation Engineering*, 128(1), pp. 21–30.

Karim A. and Adeli H. 2002b. Incident detection algorithm using wavelet energy representation of traffic patterns. *Journal of Transportation Engineering*, 128(3), pp.232–242.

Kaysi I., Ben-Akiva M. and Koutsopoulos H. (1993). An integrated approach to vehicle routing and congestion prediction for real-time driver guidance. *Transportation Research Record Transportation Research Board*, Washington D.C., 1408, pp. 66-74.

Kirby H., Dougherty M. and Watson S. (1997). Should we use neural networks or statistical models for short term motorway forecasting. *International Journal of Forecasting*, 13, pp. 45–50.

Kwon J., Coifman B. and Bickel P. (2000). Day-to-Day Travel Time Trends and travel Time Prediction from Loop Detector Data. *Transportation Research Record*, 1717, Transportation Research Board, pp. 120-129.

Kwon E. and Stephanedes Y. J. (1994). Comparative Evaluation of Adaptiveand Neural-Network Exit Demand Prediction for Freeway Control. *Transportation Research Record*, 1446, pp. 66–76.

Kyte M, Khatib Z., Shannon P. and Kitchener P. (2001). Effect of Environmental Factors on Free-Flow Speed. *Transportation Research Circular E-C018: 4th International Symposium on Highway Capacity*, pp. 108-119.

Levin M. and Tsao Y.-D. (1980). On Forecasting Freeway Occupancies and Volumes. *Transportation Research Record*, 773, pp. 47–49.

van Lint J.W.C., Hoogendoorn S.P., van Zuylen H.J. (2005). Accurate Freeway Travel Time Prediction with State-Space Neural Networks Under Missing Data. *Transportation Research Part C: Emerging technologies*, 13(5-6), pp. 347-369.

Lin W.-H., Lu Q. and Dahlgren J. (2002). A Dynamic Procedure for Short-Term Prediction of Traffic Congestion. *Proceedings of the 81st Transportation Research Record Annual Meeting*, Washington, DC.

Lingras P. and Mountford P. (2001). Time Delay Neural Networks Designed Using Genetic Algorithms for Short-Term Inter-City Traffic Forecasting. *IEA/AIE 2001, LNAI*, 2070, pp. 290–299.

Liu Y., Dai Y., Cao J.-H. (2008). Traffic Flow Chaotic Time Series Prediction Based on Wavelet Neural Network. *Computer Engineering*, 16.

Lu H., Setiono R. and Liu H. (1996). Effective Data Mining Using Neural Networks. *Knowledge and Data Engineering, IEEE Transaction*, 8(6), pp. 957-961.

MacQueen J. B. (1967). Some Methods for Classification and Analysis of Multivariate Observations. *Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley, University of California Press, 1, pp.281-297.

Mallat S. G. (1989). A theory for Multiresolution Signal Decomposition: The Wavelet Representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 11(7), pp. 674-693.

Mallat S. G. and Hwang W.L. (1992). Singularity Detection and Processing with Wavelets. *IEEE Transactions on Information Theory*, 38(2), pp. 617-643.

Min W. and Wynter L. (2011). Real-time Road Traffic Prediction with Spatio-Temporal Correlations. *Tranportation Research Part C:Emerging technologies*, 19(4), pp. 606-616.

Moody J. and Darken C. (1989). Fast learning in networks of locally-tuned processing unites. *Neural Computation*, 1, pp. 281–294.

Okutani I. and Stephanedes, Y. J. (1984). Dynamic Prediction of Traffic Volume through Kalman Filtering Theory, *Transportation Research Part B*, 18(1), pp. 1-11.

Park D. and Rilett L. R. (1998). Forecasting Multiple-Period Freeway Link Travel Times Using Modular Neural Networks, *Transportation Research Record*, 1617, pp. 163–170.

Park D., Messer C. J. and Urbanik T. ii (1998). Short-term freeway traffic volume forecasting using radial basis function neural networks, *Transportation Research Record*, 1651, pp. 39–47.

Pisano P.A. and Goodwin L. (2004). Arterial Operations in Adverse Weather. *Publications of FHWA Office of Operations.*

Poggio T. and Girosi F. (1990). Networks for approximation and learning. *Proceedings IEEE,* 78, pp. 1481–1497.

Rokash L. (2010). *Pattern Classification Using Ensemble Methods.* Series in Machine Perception and Artificial Intelligence, 75, pp. 1-18.

Rumelhart D. E., Hinton G. E. and Williams R. J. (1986). Learning internal representations by error propagation. *Paralled Distributed Processing*, *Explorations in the Microstructure of Cognition*, Cambridge, MA: The MIT Press, 1, chapt.8, pp. 318-362.

Sheu, J.-B. (1999). A stochastic modeling approach to dynamic prediction of section-wide inter-lane and intra-lane traffic variables using point detector data, *Transportation Research Part A*, 33(2), pp.79–100.

Skabardonis A., Geroliminis N. (2008). Real-time Monitoring and Control on Signalized Arterials. *Journal of Intelligent Transportation Systems Technology, Planning, and Operations*, 12 (2), pp. 64-74.

Smaragdis E. and Papageorgiou M. (2004). Series of New Local Ramp Metering Strategies. Transportation Research Record: *Journal of the Transportation Research Board*, 1856, Washington D.C.: TRB, National Research Council, pp. 74–86.

Smith B. L. and Demetsky M. J. (1994). Traffic Flow Forecasting: Comparison of Modelling Approaches. *Journal of Transportation Engineering*, 123(4), pp. 261–266.

Smith B. L. and Demetsky M. J. (1997). Traffic Flow Forecasting: Comparison of Modelling Approaches, *Journal of Transportation Engineering*, 123(4), pp. 261–266.

Smith B. L., Williams B. M. and Oswald K. R. (2000). Parametric and Nonparametric Traffic Volume Forecasting. *Proceedings of the 80th TRB Annual Meeting*, Washington, DC.

Smith B. L., Williams B. M. and Oswald K. R. (2002.) Comparison Of Parametric And Non-Parametric Models For Traffic Flow Forecasting. *Transportation Research Part C: Emerging technologies*, 10(4), pp.303–321.

Stathopoulos A and Karlaftis M. (2003). A Multivariate State-Space Approach for Urban Traffic Flow Modelling and Prediction. *Transportation Research Part C: Emerging technologies,* 11(2), pp. 121-135.

Stephanedes Y. J., Michalopoulos P. G., and Plum R. A. (1981). Improved estimation of traffic flow for real-time control. *Transportation Research Record Transportation Research Board,* Washington D.C., 795, pp. 28-39.

TSS-Transport Simulation Systems S.L. (2010). *Microsimulator and Mesosimulator Aimsun User's Manua*l. 25-29, 273-278 p.

Vlahogianni E., Geroliminis N., Skabardonis A. (2008). Empirical and Analytical Investigation of Traffic Flow Regimes and Transitions in Signalized Arterials. *Journal of Transportation Engineering*, 134(12), pp. 512-522.

Wang Y. and Nihan N. L. (2000). Freeway Traffic Speed Estimation With Single-Loop Outputs. *Transportation Research Record*, 1727, pp. 120–126.

Williams B. M., Durvasula P. K. and Brown D. E. (1998). Urban Traffic Flow Prediction: Application of Seasonal Autoregressive Integrated Moving Average and Exponential Smoothing Models. *TransportationResearch Record*, 1644, pp. 132–144.

Williams B.M. (2001). Multivariate Traffic Flow Prediction: An Evaluation Of Arimax Modelling. *Transportation Research Board 80th Annual Meeting, CD-Rom,* Washington, DC, USA.

World Health Organization - WHO (2004). *World Report on Road Traffic Injury Prevention*, pp. 157, 76-78, 119-127. http://whqlibdoc.who.int/publications/2004/9241562609.pdf, Accessed 31.08.2011.

WUTAM, The Wutam Consortium (1998). Basic Properties of Wavelets. *Journal of Fourier Analysis and Applications*, 4, pp. 575-594.

Yin H., Wong S. C. and Xu J. (2002). Urban Traffic Flow Prediction Using Fuzzy-Neural Approach. *Transportation Research Part C*, 10, pp. 85–98.

Zhang Q. and Benveniste A. (1992). Wavelet Networks. *IEEE Transactions on Neural Networks*, 3(6), pp. 889-898.

Zhang Q. (1997). Using Wavelet Network in Nonparametric Estimation. *IEEE Transactions On Neural Networks*, 8(2). pp. 227-236.

Zhang G., Patuwo B. E. and Hu M. Y. (1998). Forecasting With Artificial Neural Networks: The State of Art, *International Journal of Forecasting*, 14, pp. 35–62.

Zhang H. M. (2000). Recursive Prediction of Traffic Conditions With Neural Networks. *Journal of Transportation Engineering*, 126(6), pp. 472–481.

Zhang X. and Rice J.A. (2003). Short-term Travel Time Prediction. *Transportation Research Part C: Emerging technologies*, 11(3-4), pp.187-210.

7. Appendix

Table A.1. Comparative presentation of recent traffic prediction methods.

Methodological Approach	**Characteristics**	**Advantages**	**Deficiencies**
Historical Average (Stephanades et al., 1981 ; Jeffrey et al., 1987; Kaysi et al., 1993)	- deterministic - relies on cyclical nature of traffic flow - avg. volume of each time interval at each site	- simple structure	-impossible to predict the dynamic behaviour of traffic (incidents, transitions etc.)
Autoregressive Integrated Moving Average (ARIMA) (Okutani and Stephanades, 1984; Davis et al., 1991; Kim and Hobeika, 1993; Hamed et al., 1995)	- stochastic - parametric - linear or non-linear - periodic predictions	- well-established theoretical background	- not stable and not representative of rapid variations and unexpected edges of traffic - difficult multivariate modeling - weak transferability - need of uninterrupted series of data - not well-suited for freeway traffic flow forecasting
State-space model/ Kalman filter state estimators (Harvey, 1984; Henson, 1998; Durbin, 2000; Stathopoulos and Karlaftis, 2003)	- successive updates of parameters from different time periods during a daily observation - linear or non-linear	- possible multivariate modeling - non stationarity of variables - flexible in changes of structure - better accuracy than univariate time series models (i.e.	- demands full state of the system - the systems must be controllable - strong background of Hilbert space theory, multivariate

		ARIMA)	statitistics etc.
Non-parametric regression (i.e Nearest Neighbour, Kernel etc.) (Smith and Demetsky, 1997; Smith et al. 2002; Clark, 2003)	- deterministic - non-parametric - non-linear - identifies groups of past or neighbourhood cases with similar features (states, input value etc.) around current and not prior input state	- dynamic clustering - identifies past cases of the current prediction and of its prior - high accuracy, ameliorated via training simple structure	- data-driven accuracy - demands extensive dataset - difficult multivariate modeling
Artificial Neural Networks (ANN) (Zhang and Benveniste, 1992; Clark et al., 1993; Kwon and Stephanedes, 1994 ; Smith and Demetsky, 1994; Zhang, 1997; Zhang et al., 1998; Zhang, 2000; Kirby et al., 1997; Adeli and Karim, 2000; Karim and Adeli, 2002a,2002b; Yin et al., 2002; Ishak and Alecsandru, 2003; Vlahogianni et al., 2005; Jiang and Adeli, 2005)	- non-parametric - non-linear	- transferability - high accuracy - permits generalisations	- demands extensive dataset - complex internal structure and often heuristic
Part of Hybrid Models			
Spatio-Temporal Correlations 0 (real-time)	- For transient behaviour: Multi-variate Spatial-Temporal autoregressive (MSTAR) model - Prediction: Vector-Auto-Regressive Integrated Moving Average (VARIMA) (p,d,q)(autoregressive terms, nonseasonal differences, lagged forecast errors in prediction equation) - Decompose/cut data: Time periods into intervals of peak/off-peak of day/week Space periods into speed-based links	- accurate even for 12time periods of 15-min into future	- weather, incidents, traffic composition not included as parameters - only two regime segregation (free-flow, congested)
Wavelet, Fuzzy, Bayesian , Analysis of multi-regimes and transitions (Vlahogianni et al., 2008)	- Empirical identification of traffic flow regimes and transitions - Transitional conditions detection via flow singularity	- four-regime consideration & inter-regime& intra-regime - real-time applicable - site-transferable	- not established causality between traffic flow phenomena - identification of traffic flow regimes under incidents and adverse weather conditions

	detection - Flow regime identification - Intra-&Inter-regime Transitions identification		not included
State-Space Neural Networks (SSNN) (Lint et al., 2005)	- Predict: Discrete State Space Model (DSSM) or State-Space Neural Networks (SSNN) : depends on its previous state and the previous states of other sections (free-flowing congested) - Training: Self-learning alg. with cost function & regularization terms from Levenberg–Marquardt and Bayesian Regularization (LM–BR) - Missing data replacement: Null replacement, Simple imputation: interpolation, Exponentially moving average	- transferability and non-heuristic preparation/training method, - prediction capability under missing data - Simple imputation: interpolation and Exponential moving average: good results, insensitive to missing data - low performance time cost (sequential prediction of 15min: 0.5s)	- reciprocity between predictive accuracy & robustness. - large input dataset demanded (8191 combinations tested and failed) - weather, incidents, etc. exogenous parameters not included - only two regime segregation (free-flow, congested) - not transitional states detected
Spinning network (SPN) with BP algorithm (ANN) and Nearest Neighbour0 (Huang and Sadek, 2009)	- Historical volume vectors stored in the “rings”, the SPNetwork - Functions: Merge and Compare Merge: calculates the average of the two items to be merged (eg. traffic volume vectors) However, considering that an item may have been merged with many items before, which should give it more “stability”, the function keeps a record of how many other items were merged into a given item before the current “merge” process, and uses that count as a weight when calculating the average. Compare: calculates the Euclidean distance of two vectors to assess how similar the two vectors are. - 5min prediction for now	- low computational requirements - transferability - no training procedure	- extremely complex structure