

# Short-term real-time traffic prediction methods: a survey

Joaquim Barros  
Faculty of Engineering  
University of Porto  
Porto, Portugal  
Email: joaquim.barros@fe.up.pt

Miguel Araujo  
Carnegie Mellon University &  
CRACS/INESC-TEC  
Porto, Portugal  
Email: maraujo@cs.cmu.edu

Rosaldo J. F. Rossetti  
LIACC-DEI, Faculty of Engineering  
University of Porto  
Porto, Portugal  
Email: rossetti@fe.up.pt

**Abstract**—Short-term traffic prediction provides tools for improved road management by allowing the reduction of delays, incidents and other unexpected events. Different real-time approaches provide traffic managers with varying but valuable information. This paper reviews the literature regarding model-driven and data-driven approaches focusing on short-term real-time traffic prediction. We start by analyzing real-time traffic data collection, referring network state acquisition and description methods which are used as input to predictive algorithms. According to the input variables available, we describe common and useful traffic prediction outputs that should contribute to understand the panorama verified on a road network. We then discuss metrics commonly used to assess prediction accuracy, in order to understand a standardized way to compare the different approaches. We list, detail and compare existing model-driven and data-driven approaches that provide short-term real-time traffic predictions. This research leads to an understanding of the many advantages, disadvantages and trade-offs of the approaches studied and provides useful insights for future development. Despite the predominance of model-driven solutions for the last years, data-driven approaches also present good results suitable for Traffic Management usage.

**Keywords**—traffic; prediction; estimation; simulation; data mining; model-driven; data-driven; machine learning

## I. INTRODUCTION

The increasing number of people visiting cities has led to a greater number of vehicles traveling through main access roads. Queues and more frequent incidents [1] require wider network monitoring, faster incident detection and shorter response times. Traffic managers usually follow a *wait and see* approach, which includes 24/7 monitoring, high technical and operational costs, difficult access to incident locations and overall complex management of the road network.

The application of preventive measures is recognized to lead to more effective management of road capacity and performance [2]. These measures should be the result of real-time short-term prediction, allowing road managers to understand traffic dynamics and to anticipate future events affecting network flow [3].

How can one predict short-term traffic dynamics using current network state? What inputs should be considered in

order to obtain accurate predictions and what metrics are the most helpful for short-term decision making? What tools and algorithms are currently available to empower active traffic management? This paper presents a literature review and related work around traffic prediction using model-driven and data-driven approaches as an attempt at clarifying such questions. The contributions of this paper are the following:

- 1) **Real-time data collection:** we analyze network state acquisition and description methods used as inputs to current methods.
- 2) **Prediction targets:** we describe common and useful prediction targets.
- 3) **Evaluating prediction metrics:** we detail metrics commonly used for evaluating prediction accuracy and their importance for active traffic management.
- 4) **Model-driven and data-driven approaches:** we compare and contrast existing model-driven and data-driven approaches for the short-term prediction task.

This survey introduces the reader to some important concepts around traffic prediction, providing important characteristics that should be considered when choosing or developing new real-time forecasting algorithms and systems.

The remainder of this paper is organized as follows. In section II, we discuss important concepts related to prediction and contextualize forecasting systems. In section III we describe important characteristics that are relevant for estimation in real-time systems. Sections IV and V present traffic simulators and data analysis tools that have been applied to this task. Section VI compares the different approaches and provides valuable remarks for future work, whereas VII summarizes our findings.

## II. PRELIMINARIES

Congestion, incidents, special events and road works affect traffic dynamics and need to be addressed by either changing supply conditions or through active management of the existing infrastructure.

While constant improvement of supply conditions is necessary (i.e. modifying public infrastructures that are built in order to satisfy users' needs [4]), the construction or modification

of the road network implies high costs and is difficult due to the existence of many constraints, such as physical space. Furthermore, changes in capacity may generate hidden traffic changes [5], so the outcome might differ from the expected.

Active Traffic Management (ATM) systems represent a cheaper and more effective way to control demand and to address anomalous incidents on road networks. Nowadays, it is possible to continuously monitor and to adjust road management strategies when traffic conditions change [6]. Despite its many advantages, the common *wait-and-see* approach is lackluster and largely denies the possibility of avoiding complex situations through preventive action.

There are two main approaches adopted for road traffic prediction: model-driven and data-driven. Model-driven approaches try to computationally model the road network and, usually through simulation, analyze the performance and behavior of the drivers [1]. In order to obtain accurate predictions, network topology often include junctions, lanes, speed limits, roundabouts and signals. This information is generally represented as a graph, where nodes represent junctions/intersections and edges represent roads. During simulation, the system moves all entities according to the road characteristics and taking interactions with other vehicles and pedestrians into account. By analyzing a vehicle movement through simulation, or by analyzing traffic conditions at a specific location, it is possible to obtain a prediction of the future network state. There are many systems developed which differ by their ability to incorporate real-time information.

Model-driven approaches are extensively used for long-term prediction, especially when planning changes to existing infrastructure. In this situation, we are modeling future traffic conditions in a distant future (months or years) and disregarding any type of real-time information. Long-term predictions are also useful to model expected events such as concerts and football matches where changes in demand are unlikely to significantly deviate from the expected. However, technology and communication improvements in the last years allowed the development and use of predictive systems [7] which, according to current traffic conditions, try to provide an accurate forecast taking into account diverse parameters. The forecasts from these simulations are typically short-term (within the same day) and they are meant to complement the active management of the road network through constant update and assessment of road conditions [8].

In order to measure and verify road traffic conditions, authorities have installed different devices on road networks, such as induction loop detectors, sensors, controllers, surveillance systems [2] and even GPS devices on buses and taxis. Also, with the spread of smartphones, it is now easier to collect location, incident information, travel times and routes typically used by drivers. Along with increased storage capacity and processing power, this additional data allowed the direct application of data-driven approaches, which are able to relate traffic conditions to external information sources such as weather, incidents, road works and other special events [9] but that, in general, disregard network topology.

### III. CHARACTERIZATION OF PREDICTIVE METHODS

Application context and motivation must be considered when choosing or developing short-term predictive methods. In particular, the data available impacts both the quality of the prediction and the possible prediction targets, which must be defined *a priori*. Finally, the metric chosen to compare the different methods influences the possible errors in our forecast. We further develop these issues in this section.

#### A. Collection of real-time data

The type and quality of the information available in real-time plays a crucial role in dictating the possible approaches to our prediction problem. As hinted before, main sources of real-time information include induction loop detectors and surveillance systems, GPS-enabled public vehicles and data collected from users in real-time, for instance through smartphones. Taxis and buses often incorporate location systems which can also be used for planning and traffic monitoring [2].

In addition to a source of real-time information, historic data is also essential regardless of the approach. While data-driven methods rely on the network's history to predict its evolution, model-driven methods require it in order to calibrate the parameters used in the traffic simulation. Network topology information is often obtained from Traffic Management Bureaus or other public entities when necessary.

The nature of the data source will impact real-time definition of the network state. While ideally the position of every vehicle in the network would be known, methods realistically need to deal with partial and often derived observations. After pre-processing, data available can be divided in three groups:

- 1) **Static observations:** real-time information obtained in fixed locations in the network, e.g. from induction loops or surveillance systems. This often includes metrics such as flow, density, volume, average velocity, occupancy or queue length at specific points in the network.
- 2) **Route observations:** usually obtained from GPS-enabled vehicles, smartphones or traffic operators. This includes metrics such as travel time, queue length and accidents that are obtained from a random point in the network.
- 3) **Global observations:** usually obtained from a secondary source, this information can be as varied as weather reports, special events (e.g. football match) or other observations which should be considered for prediction purposes, such as the day of the week.

The uncertainty in this information might also have to be analyzed when incorporated into the system, especially when using information from third parties which we cannot validate, such as smartphone data.

#### B. Prediction metrics and targets

Depending on the selected approach and desired output, a large number of variables can be forecasted. In its simplest

form, the target is the direct prediction of an input variable at some point in the future. Common targets try to define the state of the network at some specific location in terms of vehicle density, traffic flow and average velocity. Other common yet often derived targets include travel time and queue length.

Other complex outputs could be useful for road managers. It is often important to predict propagation on road networks caused by traffic condition changes [10], weather, incidents or road works. For emergency teams, making sure that travel time in key routes (taking into account lane availability and maximum speed) stays below defined thresholds can help to get to accidents quicker and in safety. Incidents prediction (in order to activate or deploy preventive measures) is another possibly desired output, yet very complex.

In terms of possible targets, model-driven methods might provide valuable information about unobservable parts of the network through simulation, while data-driven methods are necessarily limited by the data available. For instance, model-driven methods might provide alternative paths for routing applications, or predict origin-destination demand matrices to be used by public transportation authorities.

However, the key aspect of prediction output is its relation to the available inputs. Good predictions are necessarily grounded on good data and its availability influences our options in predicting a future network state.

### C. Measuring prediction accuracy

In order to assess the quality of the traffic prediction systems, it is essential to establish metrics that allow the comparison of the different methods. This evaluation must consist on a comparison between the traffic prediction result and the actual traffic conditions at that point in time. The following metrics are the most common:

- Mean Relative Error - Also called Mean Absolute Percent Error (MAPE), Percentage Absolute Error (PAE), Ratio of Absolute Error and accuracy ( $1 - MAPE$ ). This metric corresponds to the average absolute percentage change between predicted and true value, relative to the true value [11].
- Mean Absolute Error (MAE) - This metric corresponds to the average absolute difference between predicted and true values [11], also called Mean Absolute Deviation.
- Root Mean Square Error (RMSE) - This metric corresponds to the square root of the mean of the square difference between observed and predicted values.

Choosing a metric to optimize is a trade-off that will direct the predictions and impact the type of errors that will occur. As some roads are more difficult to predict than others (e.g. due to frequent but unpredictable traffic jams), some roads are more prone to produce big errors than others. In one extreme, minimizing the absolute error might lead the algorithm to overfit to easy to predict values, as a big improvement in

the prediction of a single road does not compensate many small deviations. In the other extreme, minimization of squared errors is significantly influenced by outliers in the prediction.

Relative error was the most cited metric over data-driven papers referred on section V, probably because absolute errors are difficult to compare and influenced by, among other factors, speed limits. However, there isn't an universal procedure and metric which lets us to compare and assess directly the quality of the different algorithms/frameworks developed. This is especially evident given the lack of open datasets that can be used for comparison.

## IV. MODEL DRIVEN

In this section we will focus our study on five simulators frequently cited and used nowadays to obtain information about current and future traffic conditions.

### A. DynaMIT

The DynaMIT (Dynamic Network Assignment for Management of Information to Travelers) project consists of a Dynamic Traffic Assignment (DTA) system<sup>1</sup> for estimation of network conditions, real-time traffic prediction and generation of drivers guidance developed at MIT's Intelligent Transportation Systems Laboratory [13].

This system requires both offline and real-time information. For the first case, it is important to represent network topology using a set of links, nodes and loading elements [14]. Then, historic database should contain traveler socioeconomic data such as age, gender, income, auto ownership and trip purpose obtained from census and questionnaires. Finally, users should provide in real-time surveillance data, traffic control strategies and incident characteristics (location, starting time, expected duration and road capacity affected). Effective integration of all data provide estimation, prediction and travel information generation capabilities to this system [13], including link density, flow speed and driver characteristics (travel time, route choice and departure time) [15].

### B. DynaSMART-X

The DynaSMART-X (Dynamic Network Assignment Simulation Model for Advanced Road Telematics) project is a real-time network traffic estimation and prediction system based on the DTA methodology for effective support of Advanced Traveler Information Systems (ATIS) and Advanced Traffic Management Systems (ATMS). This system interacts with different traffic data sources such as loop detectors, roadside sensors and vehicle probes [16], and provides different statistics at link and network levels, performance plots of short-term and long-term consistency checking modules. It also allow to view paths, temporal demand patterns and attributes of the network items.

---

<sup>1</sup>Dynamic Traffic Assignment models represent time-varying network and demand interaction using a behavioral approach, which can be used to determine individual measures (travel time and costs) and network measures for planning [12].



In this project, vehicles are represented by macroparticles that move according to speed-density relations and take into account conservation law describing traffic flow theory. All interactions happen within a road network representation, similarly to DynaMIT, recurring to OD time-dependent matrix built from data collected previously. Relatively to real-time information, CHART project (Maryland) used as input data containing timestamp, detector location, traffic direction, vehicle counts, flow, speed and occupancy. Moreover, users can add data related to bus, pricing and paths for example.

At the end DynaSMART produces as output link volume and density [17]. CHART project performance was measured during peak hours using root mean square error calculated after 4 consecutive periods of 5 minutes. Output data shows that short-term results are more accurate than long-term results and with a very limited number of inductive loop detectors (in strategic locations) is possible to obtain reliable estimates and predictions [17].

### C. VISTA

VISTA (Visual Interactive System for Transport Algorithms) is a tool that allows traffic simulation of a road network, including short-term traffic prediction. It was developed in 2000 by A. Ziliaskopoulos and S. Waller [18] and is especially useful for planning, engineering and at operational level.

In order to estimate and predict future traffic data (traffic flow, volume and travel time) after some time intervals (1, 5, 10, 15 minutes), this software uses as input detailed road network geometry, Geographic Information System (GIS) data and traffic control information, including signal timing, speed limits and lane designation. After path assignment, it is possible to obtain spatio-temporal trajectories for travelers [19]. Model can also be used to obtain traffic estimation and prediction in roads with partially loop detector coverage [13].

### D. TRANSIMS

TRANSIMS (Transportation Analysis Simulation System) was made public in 2001 by Los Alamos National Laboratory (Los Alamos, New Mexico) and allows the usage of driver preference models in order to create traffic demand for different means of transportation (particular or public). It is a microscopic simulator which has been used for traffic simulation and real-time prediction in a specific network and assess its performance [20].

In order to forecast future traffic states, this software requires network representation, usually using simple car-following models and lane-changing logic based on cellular automata [21]. Each vehicle moves from cell to cell (each one can be occupied by only one car), allowing the simulation of movements for discrete time intervals and the representation of complex dynamics [22]. This characteristic has its basis on cellular automaton developed and presented by Nagel and Schreckenberg [23]. Users should also provide results from conducted surveys to determine people's activities which will

be used by the activity generator module. After that, TRANSIMS receives as input (in real-time) OD time-dependent demand which results from period counts or from a time-sliced demand model. At the end, TRANSIMS produces as output OD demand matrices, gives a "complete picture" of the road network and provides travel times [24]. Each vehicle can be followed, providing time, location and speed. All calculations result from activities assigned to each person simulated in the system, provided by the activity generator module.

### E. DYNEMO

The DYNEMO (*Dynamisches Netz Modell*) was developed by Schwerdtfeger in 1984 (University of Karlsruhe, Germany) and is included in the Payne-Cremer traffic flow model family [25]. Traffic flow unit is defined by each individual vehicle and movement is controlled by the average density of traffic on the corresponding link, not taking into account neighbor vehicle behavior (unlike microscopic models) [26]. This system was developed to help evaluating traffic management schemes on highways, especially in the enforcement of speed limits and ramp metering. Vehicles' paths are calculated via the shortest path algorithm based on an OD time-dependent matrix that is given to the model [27]. It is also possible to obtain real-time traffic predictions. Fellendorf, Nökel and Handke present applications of DYNEMO to a large part of the German motorway network [28].

In order to forecast traffic conditions, the simulator requires as input the network model and the relationship between density and average speed in each section. As output, DYNEMO produces traffic volume patterns, average speed in each point of the network, travel time, fuel consumption and travel time distribution for all calculated paths. These values are passed to the traffic control system which indicates new paths to vehicles and recalculates flows in all sections [26].

## V. DATA DRIVEN

In this section, we describe 8 recent data-driven approaches to the short-term prediction problem.

### A. An Improved K-nearest Neighbor Model for Short-term Traffic Flow Prediction

Zhang et al. focused their study on the development of a system capable of predicting traffic conditions on a road network [29]. The solution was tested using data collected from sensors installed on a highway in Shanghai. The algorithm developed was based on k-Nearest Neighbors and was integrated in a system composed of three main components: historical database, search mechanism and prediction component.

The historical database is updated in real-time using data obtained from loop sensors, including traffic flow, density and speed. Data is pre-processed following a process of standardization or deletion. Real-time traffic data is then vectorized and matched against the historical database using hysteresis (in order to avoid excessive similarity between observed data and database collection when comparing). The resulting flow

predictions, in passenger cars per 5 minute, had a relative error under 10% and absolute error between 6.2 and 7.6. A weighted algorithm based on Euclidean distance was also tried, resulting in lower relative but higher absolute error.

#### *B. A Hidden Markov Model for short term prediction of traffic conditions on freeways*

In this study, Qi and Ishak [5] focused their work in the ability of making predictions using Hidden Markov Models. Through the analysis of data collected from dual loop sensors (flow, occupancy and speed) they analyzed traffic conditions during peak periods on a highway in Orlando, California. Before prediction, their system starts by classifying data according to normal and heavy traffic states. Traffic states are then placed in a two-dimensional space using first-order (average) and second-order (contrast) statistics. Some variables for each state were represented using symbols which categorize different speed levels. Changes between traffic states are modeled by transitions with some probability associated.

Through this work they found that that traffic always tends to states that have low contrast and can evolve into any state (by transition probability matrices analysis). It was also verified that there are over 5 different speed patterns, making single models difficult to use. They have also tried to predict average speed. Using data obtained from the above mentioned dual loop sensors, they were able to obtain relative error lower than 10% which confirms its applicability to situations with heavy traffic conditions.

#### *C. Freeway Traffic Estimation in Beijing based on Particle Filter*

This paper presents a different approach, an hybrid between model- and data-driven methods; in particular, they connect the Particle Filter algorithm and a macroscopic traffic model [30]. It was considered a stochastic system with discrete and continuous states, where each state only interacts with its neighbors (using a directed graph of the road network) and sensors only exist in the first and last sections. Forecasting is based on parameters characterizing each road segment, reducing the computational cost, and it becomes possible to handle high data volumes and non-linear and non-gaussian large models.

They predicted flow, speed and density from traffic flow data collected in 2-minute's intervals. The maximum error obtained (using relative root-mean square error) was 9.691% for speed and 11.90% for density.

#### *D. Short-Term Freeway Traffic Flow Prediction: Bayesian Combined Neural Network Approach*

Zheng, Li & Chi [31] present a new approach based on the combination of neural networks. The algorithm proposed uses back propagation and radial basis function neural networks, which are designed and combined, resulting on a Bayesian-combined neural network model (BCNN). It is assigned a

credit value to each predictor, which changes and adapts its behavior and results.

All operations described depend on the classification of traffic flow state, selecting and adapting the appropriate model for forecasting. In order to achieve this behavior, real-time traffic flow data is aggregated in 15-min intervals from Singapore's Ayer Rajah Expressway. This data set was divided, using one for single neural network training and the other to compare the performance between the combined and individual models. Results show that hybrid models outperform single neural networks most of the time, obtaining good traffic flow predictions.

Authors found that the probability of having a relative error lower than 10% was 77.8% for the back propagation (BP) neural network, 85% for the radial basis function (RBF) neural network and 86.9% for BCNN.

#### *E. Real-time travel time prediction using particle filtering with a non-explicit state-transition model*

This paper presents a short to medium-term travel time prediction algorithm which is based on the real-time and historical traffic data collected [11]. Despite its focus on travel time, this algorithm is flexible for other data sources and areas.

Chen and Rakha propose a new algorithm based on the particle filter algorithm which does not require an underlying physical model in order to model the state transition function, but rather only depends on historical trends. They apply a resampling process, which identifies invalid or low weighted particles which are removed or replaced so to avoid the degeneracy problem.

In order to assess the performance of their algorithm, INRIX probe data was collected from a 95-mile freeway from Richmond to Virginia. Using this data, travel time (instantaneous and experienced) was derived and used as historic and real-time input data. They show increased performance when compared to other prediction methods, namely two Kalman filters and a k-NN algorithm. For prediction horizons between 0-60 minutes (10 minutes intervals) mean absolute error varied between 8.26-10.54 minutes while relative error changed between 7.32-9.29%.

#### *F. Adaptive Kalman filter approach for stochastic short-term traffic flow rate prediction and uncertainty quantification*

Guo, Huang and Williams [3] focus their study on SARIMA + GARCH process that has gained special relevance in recent years. SARIMA captures the dynamics of traffic levels (first conditional moment) while the GARCH component captures the dynamics of traffic flow variances (second conditional moment). Kalman filters were used for real-time processing, but they assume constant parameters which limits their adaptability to real-time. The authors then suggest an adaptive Kalman filter in order to improve their results.

15-minute test data (flow rates) was obtained from freeways in the United Kingdom and in the United States. They compared traffic flow prediction using four different approaches

and concluded that BATCH presented the best forecasting levels when analyzing absolute error, relative error and root mean squared error. They also analyzed uncertainty in the quantification, which led them to conclude that when combined to the previous results obtained, Adaptive Kalman Filter presented better results.

#### *G. An Aggregation Approach to Short-Term Traffic Flow Prediction*

This paper shows a hybrid approach, attempting to obtain traffic flow forecasts using results from three distinct models: moving average (MA), exponential smoothing (ES), autoregressive MA (ARIMA) [32]. Tan et al. developed an innovative system which combines the previous results using a Neural Network.

Taking into account the repeatable pattern of traffic flow, they built three time series:  $s_1$  (daily),  $s_2$  (weekly) and  $s_3$  (hourly). The proposed approach showed better results (in terms of absolute, relative and root mean squared error) when compared to naïve ARIMA, non-parametric regression and Artificial Neural Networks. Traffic data was collected from a detector installed on National Highway 107 in China and flow rates were aggregated in 1-hour periods.

#### *H. Real-time road traffic prediction with spatio-temporal correlations*

This paper proposes a different approach which is based on the MSTAR-MA( $p, q$ ) model [9]. Receiving traffic speed and volume from loop detectors and taxis, this framework can predict traffic speed and volume up to 1-hour intervals. The authors show lower relative errors when compared to other proposals.

The algorithms were developed in order to minimize the number of parameters needed, leading to a computationally light process. Other distinct characteristic is the possibility of taking into account weather, incidents and road works. The system adapts weights of some variables periodically, balancing results according to the pattern verified in the network.

## VI. DISCUSSION

After analyzing the characteristics of predictive methods and different approaches, this section presents a brief summary of important considerations that should be present when developing or choosing a new short-term traffic prediction system. We then list important observations and attempt to provide a standard way to compare different pieces of work.

### *A. Summary*

Through sections IV and V, we analyzed different methods which attempt to provide accurate traffic predictions. Many approaches and algorithms were used and compared, providing guidelines for future development. In the following paragraphs, we describe results obtained and our analysis of the works referred.

Model and data-driven approaches are characterized by input data collected, results obtained and by internal functioning. While model-driven approaches execute their operations on a road network virtual representation, data-driven approaches only take into account data collected from information sources such as detectors, location devices or surveillance systems. This difference determines data that can be obtained through processing. Figure 1 summarizes the decision-making process when deciding between model or data-driven approaches.

Most model-driven approaches receive real-time input using OD demand matrices. This allows vehicle generation, path definition (for each vehicle) and specification of other variables such as speed. However, user can provide flow, density or even mean speed to these systems. This information is processed in order to represent vehicles simulated that move throughout the network representation. Data-driven approaches are not as limited by the type of information received. Besides traffic conditions parameters such as flow, speed, volume or occupancy, user can integrate data available from weather stations, incident information, paths, travel time registered, and even other types of data, depending only on the devices installed. The devices usually referred are loop detectors, sensors, controllers and surveillance systems. As an additional remark, some systems such as DynaMIT allow to receive incident information [13], which is useful to predict effect on traffic conditions and determine best measures to overcome its consequences.

After receiving real-time information, results obtained depend largely on the system functionality. In model-driven approaches, the system attempts to move vehicles or sets of vehicles throughout the network. This enables users to follow the path of each vehicle represented and to understand traffic dynamics in an integrated way. Most of such systems allow the placement of "virtual sensors" all over the network, providing forecast measures such as traffic flow, speed and density at a specific point. When following individual vehicles, user can obtain travel times for each OD pair, becoming a useful resource to understand traffic conditions and route choices.

Table I summarize data-driven approaches analyzed. A ✓ indicates the variable was explicitly used, whereas a + means that authors indicate it as a possible extension. Due to the relationship between some input variables (e.g. traffic flow, density and speed), some metrics might be derived and thus omitted. Furthermore, techniques might be agnostic of the specific metric used and easily incorporate it even if not mentioned.

In addition to input and output variables, managers should also consider application context. Model-driven techniques analyze traffic as a whole, as well as queues and driver behavior. This enables forecasts which take into consideration changes to traffic dynamics after information to travelers has been broadcast. By recommending best routes, managers can avoid overreaction, oversaturation and concentration phenomena [33], enabling better travel times.

Another way to compare predictive methods is based on the quality of the results obtained. In order to assess different



TABLE I. COMPARISON OF DATA DRIVEN PREDICTION METHODS.

		A	B	C	D	E	F	G	H
Real-time input	Flow	✓	✓	✓	✓		✓	✓	✓
	Density	✓	✓						
	Speed	✓	✓						✓
	Occupancy		✓						
	Travel time					✓			
	Incidents		+						+
	Weather		+						+
Output	Flow	✓			✓		✓	✓	✓
	Density			✓					
	Speed		✓	✓					✓
	Travel time (mean)					✓			
	Travel time (distribution)					✓			

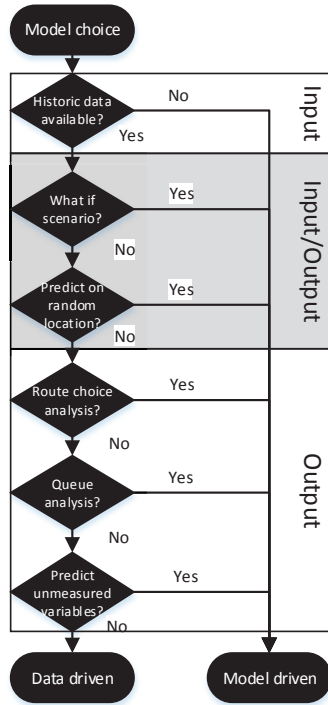


Fig. 1. Questions to consider when choosing different approaches.

systems, authors usually use mean absolute percent error (MAPE), mean absolute error (MAE) or root mean square error (RMSE) to compare deviation between predicted variable and the true value obtained. However, there is no standard which can be used to compare different approaches.

### B. Observations and future work

This paper analyzed what traffic data can be predicted and what data sources are needed for this purpose. Authorities should start building a historic database in advance and consider techniques which avoid overfitting [32] and degeneracy [11] problems. Other important consideration refers to the possibility to forecast everywhere on the road network (due to its representation) in model-driven based frameworks while data-driven tools are limited in general by the collection points.

During this literature review we found 2 papers referring to solutions that could integrate weather and incident information.

This type of frameworks takes the most from capabilities of data-driven systems, exploring patterns and relationships between traffic conditions and other unexpected factors. No paper measured the impact of weather on the accuracy of the prediction, suggesting more research in this area is required. Min and Wynter suggest that it is fundamental to obtain more knowledge about the data collected in order to understand whether the quality of results is improved [9].

Assessment and comparison between different alternatives require evaluation of deviation between predicted and observed value. As seen before, there are many different measures to evaluate result quality, making it difficult to determine which are the best ones. As referenced by Guo, Huang and Williams, the definition of an uniform measure may allow to evaluate prediction performance in a standardized way [3]. Another problem verified refers to the data sample analyzed which depends on the application context and determines results obtained. Development of a reference sample could be an important contribution to research in short-term real-time traffic prediction in order to evaluate variables such as accuracy and computing time.

As a final remark, there are alternatives that try to get the best out of multiple techniques. Researchers should consider new developments such as hybrid approaches [30] or a mix of predictors [31]. In the first case, we are associating advantages from model and data-driven approaches which become even more important when considering incidents, road works and other unexpected events which limit forecasts using past data. Fusco's and Colombaroni's approach [34] represents a good alternative, considering Artificial Neural Networks for normal situations and simulation for special conditions, depending on an incident detection algorithm. Other alternatives depend on a credit value that is given to each predictor which controls contribution according to the traffic situation verified.

## VII. CONCLUSION

In this paper, we presented a literature review and related work about traffic prediction models. We explored model-driven and data-driven approaches, comparing and understanding their capabilities. Taking into account all the information analyzed, it was possible to identify some important characteristics that should be considered when characterizing and developing new short-term traffic prediction methods.

We analyzed the main methods and inputs which allow us to describe real-time traffic states. Most traffic prediction methods receive as input traffic flow information, travel times, paths or OD matrices which are obtained from data collected through loop detectors, surveillance systems, and AVL systems (GPS-equipped vehicles). There are also weather or incident information sources which are an additional contribution to more accurate forecasts.

Taking into account real-time input data and internal functioning from these frameworks, useful prediction targets were identified including flow, speed, density, travel times, paths and best alternatives. However, while data-driven approaches

forecast traffic states on specific points (data collection points), model-driven approaches simulate the whole network allowing more complex predictions, including unpredicted situations and driver behaviors.

Following this analysis, we detailed metrics commonly used for prediction assessment. Mean Absolute Percent Error (MAPE), Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) are the most referred in the papers analyzed. However, we consider that future work should focus on the definition of a standard measure to compare different approaches.

After analyzing sections IV and V, we verify that nowadays data-driven techniques offer good solutions for real-time traffic prediction. However, some specific behaviors can only be analyzed recurring to simulators, leading to the introduction of hybrid approaches [30] and coupling of predictors [31] which we consider important references that should be analyzed in order to explore the best from both worlds. This analysis provided some insight into short-term real-time traffic prediction, which can be used for the development of a new framework for testing, assessment and development of new approaches.

## REFERENCES

- [1] S. Blatnig, "Microscopic Traffic Simulation with Intelligent Agents," Ph.D. dissertation, Alpen-Adria-Universität Klagenfurt at Fakultät für Technische Wissenschaften, 2008.
- [2] E. Bolshinsky and R. Freidman, "Traffic Flow Forecast Survey," Technion - Computer Science Department, Tech. Rep., 2012.
- [3] J. Guo, W. Huang, and B. M. Williams, "Adaptive Kalman filter approach for stochastic short-term traffic flow rate prediction and uncertainty quantification," *Transportation Research Part C: Emerging Technologies*, vol. 43, pp. 50–64, 2014.
- [4] J.-P. Rodrigue, C. Comtois, and B. Slack, *The Geography of Transport Systems*, third edit ed. New York: Routledge, 2013.
- [5] Y. Qi and S. Ishak, "A Hidden Markov Model for short term prediction of traffic conditions on freeways," *Transportation Research Part C: Emerging Technologies*, vol. 43, pp. 95–111, 2014.
- [6] V. P. Sisiopiku, "Active Traffic Management as a Tool for Addressing Traffic Congestion," *InTech, Intelligent Transportation Systems*, 2012.
- [7] W. Min, L. Wynter, Y. Amemiya, and Y. Heights, "IBM Research Report Road Traffic Prediction with Spatio-Temporal Correlations," vol. 24275, 2007.
- [8] D. A. Rozeemond, "Forecasting travel times based on actuated and historic data," *Urban Transport and the Environment for the 21st century III*, 1997.
- [9] W. Min and L. Wynter, "Real-time road traffic prediction with spatio-temporal correlations," *Transportation Research Part C: Emerging Technologies*, vol. 19, no. 4, pp. 606–616, 2011.
- [10] F. Maier, R. Braun, F. Busch, and P. Mathias, "Pattern-based short-term prediction of urban congestion propagation and automatic response," *Traffic Engineering and Control*, vol. 49, no. June 2008, pp. 227–232, 2008.
- [11] H. Chen and H. a. Rakha, "Real-time travel time prediction using particle filtering with a non-explicit state-transition model," *Transportation Research Part C: Emerging Technologies*, vol. 43, pp. 112–126, 2014.
- [12] Y.-C. Chiu, J. Bottom, M. Mahut, A. Paz, R. Balakrishna, T. Waller, and J. Hicks, "Dynamic Traffic Assignment: A Primer," no. June, 2011.
- [13] J. Lopes, "Traffic prediction for unplanned events on highways," 2011.
- [14] M. Ben-akiva, M. Bierlaire, H. Koutsopoulos, and R. Mishalani, "DynaMIT: a simulation-based system for traffic prediction," *DACCORD Short Term Forecasting Workshop*, pp. 1–12, 1998.
- [15] M. Milkovits, E. Huang, C. Antoniou, M. Ben-Akiva, and J. A. Lopes, "DynaMIT 2.0: The Next Generation Real-Time Dynamic Traffic Assignment System," *2010 Second International Conference on Advances in System Simulation*, pp. 45–51, Aug. 2010.
- [16] H. S. Mahmassani, X. Fei, S. Eisenman, X. Zhou, and X. Qin, "DYNASmart-X Evaluation for Real-Time TMC Application: Chart Test Bed," University of Maryland, Maryland, Tech. Rep. July, 2005.
- [17] X. Qin, "Traffic flow modeling with real-time data for on-line network traffic estimation and prediction," 2006.
- [18] A. K. Ziliaskopoulos and S. Waller, "An Internet-based geographic information system that integrates data, models and users for transportation applications," *Transportation Research Part C: Emerging Technologies*, vol. 8, no. 1-6, pp. 427–444, Feb. 2000.
- [19] D.-Y. Lin, N. Eluru, S. T. Waller, and C. R. Bhat, "Integration of activity-based modeling and dynamic traffic assignment," *Research Record: Journal*, no. 512, 2008.
- [20] R. Fujimoto and J. Leonard II, "Grand Challenges in Modeling and Simulating Urban Transportation Systems," 2002.
- [21] S. A. Boxill and L. Yu, "An Evaluation of Traffic Simulation Models for Supporting ITS Development," Center for Transportation Training and Research, Houston, Texas, Tech. Rep. 2, 2000.
- [22] B. B. Park and J. Kwak, "Calibration and validation of TRANSIMS microsimulator for an urban arterial network," *KSCE Journal of Civil Engineering*, vol. 15, no. 6, pp. 1091–1100, 2011.
- [23] K. Nagel and M. Schreckenberg, "A cellular automaton model for freeway traffic," *Journal de Physique I*, vol. 2, no. 12, pp. 2221–2229, 1992.
- [24] M. J. Koohbanani, "Enhancements to Transportation Analysis and Simulation System (TRANSIMS)," Ph.D. dissertation, Faculty of the Virginia Polytechnic Institute and State University, 2004.
- [25] T. Schwerdtfeger, "DYNEMO: A model for the simulation of traffic flow in motorway networks," *Ninth International Symposium on Transportation and Traffic Theory*.
- [26] K. Nökel and M. Schmidt, "Parallel DYNEMO : Meso-Scopic Traffic Flow Simulation on Large Networks," pp. 387–403, 2002.
- [27] D. Watling, "A Review Of Models Of Urban Traffic Networks (With Particular Reference To The Requirements For Modelling Dynamic Route Guidance Systems)," 1991.
- [28] M. Fellendorf, K. Nökel, and N. Handke, "VISUM-online – traffic management for the EXPO 2000 based on a traffic model," pp. 1–8, 2000.
- [29] L. Zhang, Q. Liu, W. Yang, N. Wei, and D. Dong, "An Improved K-nearest Neighbor Model for Short-term Traffic Flow Prediction," *Procedia - Social and Behavioral Sciences*, vol. 96, no. Cictp, pp. 653–662, 2013.
- [30] S. Ren, J. Bi, Y. F. Fung, X. I. Li, and T. K. Ho, "Freeway traffic estimation in Beijing based on particle filter," *Proceedings - 2010 6th International Conference on Natural Computation, ICNC 2010*, vol. 1, pp. 292–296, 2010.
- [31] W. Zheng, D.-H. Lee, and Q. Shi, "Short-Term Freeway Traffic Flow Prediction: Bayesian Combined Neural Network Approach," *Journal of Transportation Engineering*, vol. 132, no. February, pp. 114–121, 2006.
- [32] M. C. Tan, S. C. Wong, J. M. Xu, Z. R. Guan, and P. Zhang, "An aggregation approach to short-term traffic flow prediction," *IEEE Transactions on Intelligent Transportation Systems*, vol. 10, no. 1, pp. 60–69, 2009.
- [33] H.-L. Chang and P.-C. Chen, "The Impact of Traffic Information on Drivers' Route Choice-Using Competence Sets Analysis," *Journal of the Eastern Asia Society for Transportation Studies*, vol. 6, pp. 2425 – 2440, 2005.
- [34] G. Fusco and C. Colombaroni, "An Integrated Method for Short-Term Prediction of Road Traffic Conditions for Intelligent Transportation Systems Applications," *7th European Computing Conference (ECC '13)*, pp. 339–344, 2013.