IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received April 12, 2019, accepted May 9, 2019, date of publication May 16, 2019, date of current version June 6, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2917228

A Taxonomy of Traffic Forecasting Regression Problems From a Supervised Learning Perspective

JUAN S. ANGARITA-ZAPATA[1], ANTONIO D. MASEGOSA[1,2], AND ISAAC TRIGUERO[3]

[1]DeustoTech, Faculty of Engineering, University of Deusto, 48007 Bilbao, Spain
[2]IKERBASQUE, Basque Foundation for Science, 48011 Bilbao, Spain
[3]Automated Scheduling, Optimization, and Planning (ASAP), School of Computer Science, University of Nottingham, Nottingham NG8 1BB, U.K.

Corresponding author: Juan S. Angarita-Zapata (js.angarita@deusto.es)

This work was supported by the European Union Horizon 2020 Research and Innovation Programme under Agreement 769142, and in part by the Marie Sklodoska-Curie under Agreement 665959.

ABSTRACT One contemporary policy to deal with traffic congestion is the design and implementation of forecasting methods that allow users to plan ahead of time and decision makers to improve traffic management. Current data availability and growing computational capacities have increased the use of machine learning (ML) to address traffic prediction, which is mostly modeled as a supervised regression problem. Although some studies have presented taxonomies to sort the literature in this field, they are mostly oriented to classify the ML methods applied and a little effort has been directed to categorize the traffic forecasting problems approached by them. As far as we know, there is no comprehensive taxonomy that classifies these problems from the point of view of both traffic and ML. In this paper, we propose a taxonomy to categorize the aforementioned problems from both traffic and a supervised regression learning perspective. The taxonomy aims at unifying and consolidating categorization criteria related to traffic and it introduces new criteria to classify the problems in terms of how they are modeled from a supervised regression approach. The traffic forecasting literature, from 2000 to 2019, is categorized using this taxonomy to illustrate its descriptive power. From this categorization, different remarks are discussed regarding the current gaps and trends in the addressed traffic forecasting area.

INDEX TERMS Traffic forecasting, supervised learning, machine learning, deep learning, intelligent transportation systems.

I. INTRODUCTION

Traffic congestion causes social, economic, and environmental issues. A contemporary strategy to deal with this phenomenon is the design and implementation of Traffic Forecasting (TF) schemes integrated into Intelligent Transportation Systems [86]. TF can be defined as the prediction of near future traffic measures for single locations, road segments, or entire networks [116]; allowing users to plan their movements along the roads ahead of time and enabling decision makers to improve the management of traffic flows.

With the advent of data and growing computational resources, the prediction of traffic using data-driven methods has become a preponderant approach, in which, TF can be tackled from different modelling perspectives: as a supervised regression problem [43], as a supervised classification problem [3], [70], or as a clustering-pattern recognition problem [33], [114], [121]. However, the supervised regression approach is, by far, the most widely used modelling perspective in the TF literature. Over the last few years, the number of academic publications about TF approached as a supervised regression problem has increased extensively. From a Machine Learning (ML) perspective [13], a supervised TF regression problem consists in building a predictive model using historical data to make predictions of typically continuous traffic measures based on unseen data.

The literature of TF reports the application of a wide range of ML methods, such as Neural Networks (NNs), Support Vector Machines, k-Nearest Neighbors, and Random Forest, among others (see [57], [116]). This variety has led to

The associate editor coordinating the review of this manuscript and approving it for publication was Xinyu Du.

different taxonomy proposals, published in survey and review papers between 2004 and 2015 [78], [81], [109], [110], [115], which categorise the methods based on the mathematical assumptions from which they operate w.r.t. traffic theory models and statistical methods. In spite of this, it has been hard to find clear knowledge to differentiate the latter methods with the ones that come from the ML area.

Contrarily, a few efforts have been directed to classify the TF problems approached by the aforementioned methods, and to the best of our knowledge, no taxonomy has been proposed in this regard. Instead, we do find research articles, mainly survey and review papers [30], [57], [78], [81], [109], [115], [116], which propose a set of criteria that allows for the categorization of these problems based on aspects such as, the type of data source (e.g. loop detectors or GPS sensors), the context of predictions (e.g. freeway or urban environments), the scale of predictions (e.g. single point, road segment, or network level), and the type of input data (e.g. temporal traffic data, spatial traffic data, or non-traffic data). Nevertheless, in our opinion, the categorizations proposed by these papers have the following drawbacks: 1) there is not a unified set of criteria: they vary from one paper to another; 2) the criteria shared by some of the previously mentioned papers do not have a common definition; and 3) most of the criteria proposed are related to traffic characteristics and they do not take aspects related to how the problems are modeled from a supervised regression perspective into account.

With these ideas in mind, in this paper, we propose taxonomy to categorize TF regression problems, in terms of both ML modelling and traffic specifications. The proposed taxonomy does not aim to highlight all details associated with the problems in order to maintain its comprehensibility and its size. It is instead designed according to core characteristics that may alter the complexity and the modelling of TF regression problems. The main contributions of this work are:

- A taxonomy that provides a panoramic view of the different TF regression problems with the aim of providing a common framework that helps to display similarities and differences among the problems. This allows us to identify gaps in the literature and establish the basis of a more systematic and rigorous comparison among TF methods based on ML.
- Unify and consolidate traffic related criteria available in the literature to characterize TF regression problems.
- Introduce new criteria to categorize TF problems with respect to how they can be modeled from a supervised regression learning perspective.
- Identify well-established approaches, gaps and current trends in the research area of TF regression problems using the proposed taxonomy.

The rest of this paper is structured as follows. Section II covers background and related work about TF and previous taxonomies published in this field. Section III presents the proposed taxonomy that is built according to traffic and ML modelling specifications. Then, Section IV categorizes relevant TF literature, dated from 2000 to 2019, using the proposed taxonomy to check its robustness and its ability to discriminate TF regression problems. Finally, Section V draws the main conclusions of the paper.

II. BACKGROUND

This section provides some context and background information about TF as well as reviewing literature related to other taxonomies proposed in this field. We start by presenting a brief history of how traffic prediction has evolved (Section II-A). Then, Section II-B summarizes data-driven modelling perspectives to approach TF, with special emphasis on ML. Finally, Section II-C reviews studies related to taxonomies of TF methods previously published and papers that have proposed different criteria to categorize TF problems.

A. BRIEF HISTORY OF TRAFFIC FORECASTING

For the last three decades, TF has been a relevant research topic due to its active role in Intelligent Transportation Systems as a strategy to deal with traffic congestion. The main objective of TF is the prediction of near future traffic measures based on current and past traffic data [116]. At the beginning of TF history, most research was focused on predicting traffic at a single location using traffic theory models [55] and classical statistical methods [53].

The emergence of sensing and telecommunications technologies integrated to transportation infrastructure started to generate vast volumes of traffic data, which in turn caused a switch in the modelling paradigm towards a data-driven approach [81]. Since then, a variety of methods have been proposed placing special emphasis on computational intelligence-based approaches, such as NNs [67], [132], Fuzzy logic [16], [17], and Bio-inspired algorithms [70], [87], among others [116].

Although some TF literature still relies on statistical methods, in recent years, ML methods have attracted the interest of this community and they are present in a wide proportion of current literature (see the review published by Ergamun and Levinson [30]). As computational capacities and massive data processing techniques have increased, complex scenarios with different road settings can be tackled with ML (e.g. network-wide predictions) leaving behind traditional approaches to address traffic prediction [57]. In this context, ML actively contributes in the design and development of current Advanced Traffic Management Systems and Advanced Traveler Information Systems [93].

In the following section, more details are provided regarding the modelling process of data-driven methods, with a special emphasis on ML as it is the main topic of the paper.

B. DATA-DRIVEN MODELLING APPROACHES OF TRAFFIC FORECASTING

Traffic forecasting can be tackled from different modelling perspectives. The approaches most commonly reported in transportation literature are statistical time-series [20], [30], [52], [60], [84], supervised regression [43], supervised classification [3], [21], [70], [74], [95], or clustering-pattern recognition [2], [33], [54], [107], [121], [123].

When the objective is to predict a continuous traffic variable, the possible modelling approaches can be time-series, supervised regression, or clustering-pattern recognition. In the first case, the modelling process is based on identifying the statistical distribution followed by the input data, defining the functional form, and then developing models that fit observations made previously, which are used to predict future traffic states [10].

In supervised regression, the focus is on using ML algorithms to learn a functional form based on the input data, without prior models or data distribution assumptions [30]. In this context, the goal is to approximate the learned mapping function in such a way that when the model faces new and unseen traffic data, it is able to make accurate predictions.

The third possible modelling approach to predict a continuous traffic variable is clustering-pattern recognition. In this case, the focus consists of finding the relationships of different locations by characterizing similar traffic measure values from one road to another, and grouping them in clusters that divide the network into correlated groups. Once the clusters have been identified, the next step is to use a supervised regression perspective to predict the traffic conditions, cluster by cluster, based on historical traffic data belonging to each group.

Finally, when the objective is forecasting a discrete traffic measure, the modelling approach should be supervised classification, which also learns a mapping function based on historical data. For instance, ML methods can forecast the Level of Service of a specific road. The latter is a categorical variable that measures the quality of the traffic through letters from A to E in a gradual way, category A being moderate traffic and category E extended delays [96]. It is important to clarify that the forecasting of discrete variables could also be addressed as a supervised regression problem on some occasions, predicting either speed or density (continuous values), and then discretising these predictions to obtain the categorical outputs.

C. TAXONOMIES IN THE TRAFFIC FORECASTING FIELD

Although literature in TF is abundant, there are not many research articles that address either the classification of the methods or the categorization of TF problems. The most relevant studies are presented below with a brief description of their main contributions.

In [115], Vlahogianni *et al.* characterized TF methods published from 1978 to 2003 in terms of the hypotheses made about the statistical distribution followed by the data, the quantity and quality of the data needed for making predictions, and the accuracy of the methods. Similarly to the ML field [38], the authors categorized two types of methods: parametric and non-parametric ones. The parametric category assumes the relationship between the explanatory and response variables as known; meanwhile, the non-parametric ones are able to model non-linear relationships without requiring the mentioned assumptions.

The aforementioned work also categorizes traffic attributes into three clusters: the scope, the conceptual process of specifying the output, and the type of input and output data. The first cluster includes the context under which predictions are done (freeway or urban); the second group is associated with the time step of data and time horizon of predictions, and the specification of what traffic variables are used; the third cluster consists in defining how many variables are considered in the input and output of the problem.

Two years later in 2007, van Hinsbergen *et al.* [109] proposed a new category of methods named "naive", in addition to the two categories proposed by Vlahogianni *et al.* [115], which makes predictions based on historic values or an average of the total historic values. In this case, the authors compared the methods under some of the traffic-related attributes defined in [115] and introduced a new attribute to describe the scope of the predictions, which can be single point, road segment, or network.

More recently in 2014, Vlahogianni *et al.* [116] presented a comprehensive review of TF literature to state a series of challenges for present and future work in this field given the development and expansion of data-driven methods. Such conclusions were drawn on the comparison of articles using the transportation attributes defined in [115]. In addition, the authors included the type of data source to the existing attributes.

In 2015, Mori *et al.* [78] proposed a taxonomy of methods for a subset of TF problems focused on travel time prediction. Within this taxonomy, the main categories of methods are naive models, traffic theory models, data-driven models, and hybrid models. The authors concluded the necessity of developing a comparative study of travel time forecasting that, rather than finding one best method, should be aimed at providing suggestions of what method is more appropriate in each forecasting situation at hand.

Complementing previous works, Ergamun and Levinson [30] introduced a new taxonomy wherein the methods fall into either statistical or ML categories. The authors stated that ML methods have attracted more attention in recent years outperforming statistical methods, such as historical average and exponential smoothing. Nevertheless, they concluded that there is no guarantee of superiority when ML are compared with advanced statistical methods, such as spatiotemporal auto-regressive integrated moving average. In spite of this, the authors stressed that both approaches differ on purposes and model development process. Statistical methods concern inference and estimation providing a model that offers insights on the data, considering both data distributions and model restrictions; whilst ML methods are focused on providing efficient and accurate predictions without prior models or data distribution specifications.

In summary, two big taxonomies of TF methods are consolidated: a traditional classification with parametric,

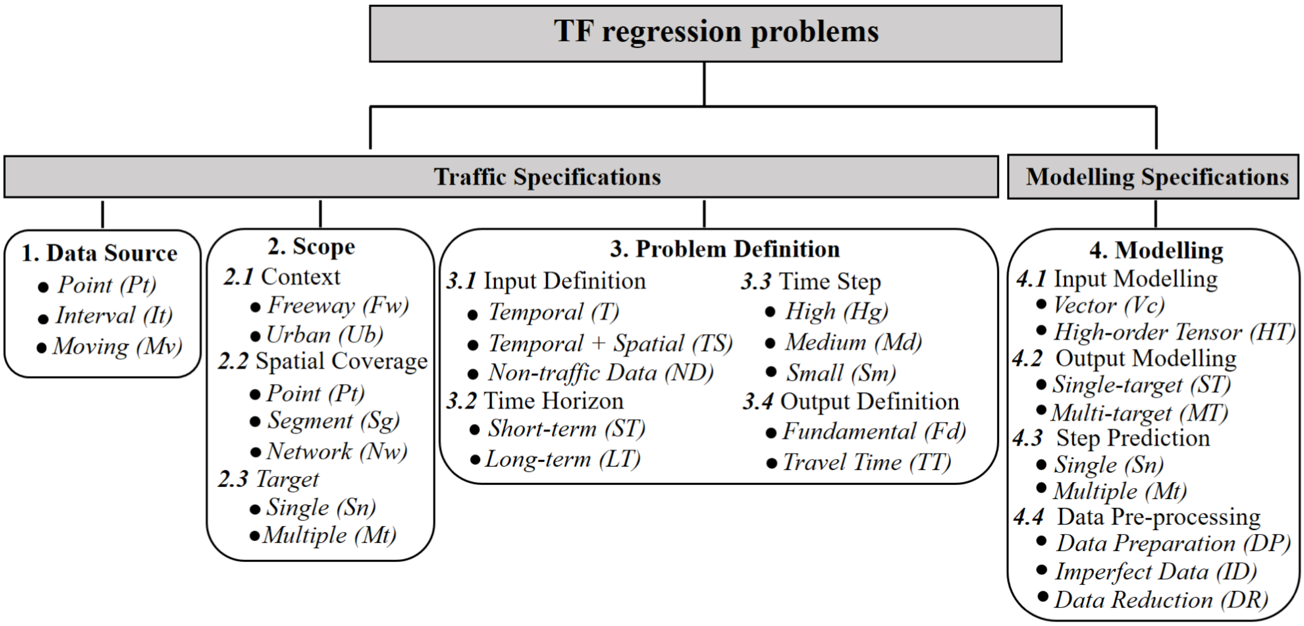


FIGURE 1. **A taxonomy of traffic forecasting regression problems.**

non-parametric and naive categories, and a taxonomy of methods in terms of ML and statistical approaches. Both taxonomies provide an overview of the available methods based on the mathematical assumptions from which they operate, and for future work suggest making predictions more representative of real traffic conditions. On the other hand, with respect to the categorization of TF problems, no taxonomy has ever been proposed before; and the described studies only present attributes that describe the transportation scenarios under which the methods were applied. Furthermore, none of the studies introduce attributes that describe the problems in terms of the modelling specifications associated with approaching TF as a supervised regression learning problem.

III. A TAXONOMY OF TRAFFIC FORECASTING SUPERVISED REGRESSION PROBLEMS

This section presents the proposed taxonomy to categorize TF supervised regression problems, which is built according to traffic and modelling specifications, as can be seen in Figure 1. These two classes of specifications have a hierarchical inner structure in which, from top to bottom, there are attributes of at most three levels. The attributes of the first and the second level correspond to general characteristics that determine the types of features that the problems may include; whilst the bullet points are the final attributes that assign, within each first or the second level class, the particular features of each TF problem.

The set of traffic specifications contain three blocks of attributes that classify the context and definition of the problems: data source, scope, and problem definition. The second set of specifications categorizes how to model the TF problem from a ML supervised regression perspective: the input and output modelling, the steps of predictions into the future, and the pre-processing approaches of the input data.

It is important to clarify that the taxonomy does not aim to describe all the details associated with the problems in order to maintain its comprehensibility and its size. Rather, it is designed according to some main attributes that may alter the complexity and the modelling of TF regression problems, keeping a moderate level of granularity. The remainder of this section is devoted to presenting and describing in more detail the attributes of the taxonomy.

A. TRAFFIC SPECIFICATIONS

Previous surveys have proposed a set of traffic attributes to describe the transportation scenarios in which TF methods can be applied [109], [115], [116]. Such attributes are the data source used; scope related attributes such as the context, spatial coverage and target; and problem definition related attributes as the input data used, the time horizon of predictions, the time step resolution of the data and the traffic measures to be predicted.

In this section, we present the aforementioned traffic-related attributes and their respective sub-attributes, including some important references that present more in depth information about them. It is important to clarify that although these transportation attributes have been defined before in the aforementioned studies, to the best of our knowledge this is the first time that: 1) their definitions are unified and consolidated in a single taxonomy, and (2) their influence over the preparation of the ML data-sets has been approached.

In the following subsections, they are formally defined in detail and graphical explanations are also used so that their concepts can be better understood.

1) DATA SOURCE

This attribute is related to the type of data source technologies and the traffic data that can be obtained from them. Recent progress in technologies for Intelligent Transportation Systems has enabled the extraction of traffic data from different sources that can be classified in several ways [69]. Nevertheless, as we are proposing a generic taxonomy able to categorize the diversity of data sources used in TF, the categorization presented here is divided in three groups: point detectors, interval detectors, and moving sensors. They differ in their spatial coverage capacity at the moment of sensing traffic. The three groups of data sources are described below.

a: POINT

This type of data source is placed at specific locations on the roads to detect the presence of nearby vehicles. Point detectors are able to output basic traffic measures as flow (number of passing vehicles per hour), occupancy (percentage of the time that the detector is occupied), and density (number of vehicles per unit length of the road). The most conventional sensors within this category are Loop detectors, Microwave radar, Laser radar sensor, Active and Passive infrared, among others (for more details see [69], [78], [79]).

Data of a point sensor can be described as an ordered sequence of measurements m_p in a given position p (Equation 1), wherein $m_{p,t}$ is the value of the traffic measurement at time t and position p. In this case, the traffic measures and their respective prediction are only valid for describing the traffic conditions where the sensor is situated.

$$m_p = \{m_{p,t}\} \qquad t = 1, 2, \ldots, T \tag{1}$$

The advantage of point sensors is that they are reliable data sources, capturing all vehicles passing near them and collecting macroscopic traffic measures, which means averages of many vehicles. Their drawback is that they are not able to observe the paths of vehicles, and as a result, it is hard to find relationships between different road segments. Such an issue can be overcome using spatial correlation analysis to find relationships between the road segments with installed sensors, and then determining what specific sensors will be included in the data-set preparation phase [50], [132]. Unfortunately, the cost of deployment and maintenance of a big sensor network can be excessively high, and this has an impact on the volume and quality of the available data [115].

b: INTERVAL

The second type of data source is interval detectors, which are capable of calculating traffic measures between two fixed points on a road. With respect to point detectors, they enable the direct sensing of travel time. The most common technologies in this category are Automatic toll collection systems, Video cameras, and License plate recognition, among others (see [69], [78]).

Even though data from interval detectors can be represented with Equation 1, their spatial coverage at the moment of predicting traffic is not the same as point sensors. In this case, data is valid for representing traffic measures between two points of a road segment, which means a broader spatial coverage beyond a single location.

In contrast to point sensors, interval detectors are not able to detect all the vehicles on the road positions where they are located [78]. In this sense, having a sample size of sensed vehicles that represent the traffic conditions on the roads in a realistic way is a challenge [68]. Therefore, forecasting traffic with this type of data source requires large volumes of data, which in real scenarios can be difficult to obtain.

c: MOVING

The appearance of Global Positioning Systems (GPSs) in smartphones and vehicles has given rise to a new type of data source that gathers more detailed traffic information. Contrary to fixed sensors, moving sensors provide individual traffic data related to vehicles' trajectories on the roads. This data allows for the identification of path patterns of vehicles in large areas with lower infrastructure costs, which means that it is more feasible to predict traffic at the network level [12], [40]. Nevertheless, aggregate traffic measures (e.g. flow or occupancy) can only be approximated to a certain point, depending on the number of available moving sensors [78], [79]. This last condition affects data preprocessing because extrapolating the GPS samples to obtain estimations of aggregated traffic measures is necessary [41].

GPS devices send location, direction and speed information every few seconds. This data can be represented as an ordered sequence of measurements, such as shown in Equation 2. Every sample p_t can be defined as spatio-temporal data $p_t = (x_t, y_t)$, wherein the spatial component contains GPS coordinates, latitude (x_t) and longitude (y_t), and the temporal part includes the time stamp (t).

$$\overline{p} = \{p_1 \rightarrow p_2 \rightarrow \ldots \rightarrow p_t \rightarrow \ldots\} \qquad t = 1, 2, \ldots, T \tag{2}$$

However, the sequence in Equation 2 sometimes contains non-exact records that impact the quality of the data [78], [130]. In order to overcome this issue and prepare the ML data-sets, the technique most commonly used is map-matching; a procedure that pins the drifting positions of data to the correct road links on which vehicles are travelling [44], [45], [105].

2) SCOPE

This attribute consists of the context wherein traffic predictions are made and their spatial extent. We identify three sublevels within the Scope attribute. The first sublevel corresponds to the transportation environments wherein traffic predictions can take place. Based on [115], the contexts can be Urban and Freeway. The second sublevel determines the spatial coverage of predictions, which is divided into

Point, Segment, and Network, as is defined in [109]. Finally, the third sublevel is related to the number of locations simultaneously considered to make predictions, which can be a single-target (one Point, Segment, or Network) or multiple locations, according to [57].

a: CONTEXT

Regarding the context of prediction, every area of a traffic network has a set of characteristics that determines the behavior of traffic in it. We identify two main types of traffic contexts: Urban and Freeway based on [57], [78], [81], [115]. Most of the ML models in the transportation literature are built using traffic data collected within freeway contexts wherein traffic is generally uninterrupted [18], [97], [103], [126]. The main reason behind this trend is the availability of fixed position sensors already installed in freeways in many cities around the world, which makes the acquisition of data easier [78].

Until recently research in urban contexts was not so common because of fixed sensor coverage issues [57], [115]. However, the advent of GPS devices has allowed more research in urban contexts that have not yet been covered with fixed position detectors [79]. Regarding the development of predictive models, traffic in urban locations is more complex and the modelling process has to take aspects such as the influence of traffic lights and intersections into account. This requires more elaborate ML models that consider such variability and provide reliable traffic predictions [57], [81].

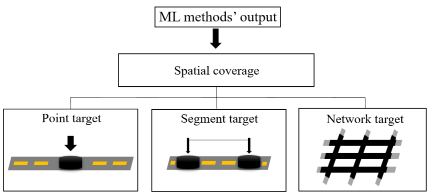


FIGURE 2. Spatial coverage of predictions provided by ML methods.

b: SPATIAL COVERAGE

This attribute represents the spatial coverage of the predictions provided by ML methods (Figure 2). As is defined in [109], the spatial covers are Point, Segment, and Network. Until the most recent and comprehensive literature reviews [81], [116], most forecasting efforts were focused on point and segment predictions [4], [18], [89], [94], [102], [103], [127], [129], [132]. However, as more moving sensor data becomes available and ML methods able to deal with temporal and spatial data appear, an increase of traffic forecasting at the Network level is reported in the transportation state-of-the-art [57], [64], [67], [88].

c: TARGET

The target attribute represents the number of locations for which predictions are carried out. Based on [57], the TF problem at hand can require traffic predictions to be made for more than one point or segment, or for a network. For one single target, the TF problem is approached as a simple regression problem [7] in which a single ML model is trained and tested to predict traffic at the target location.

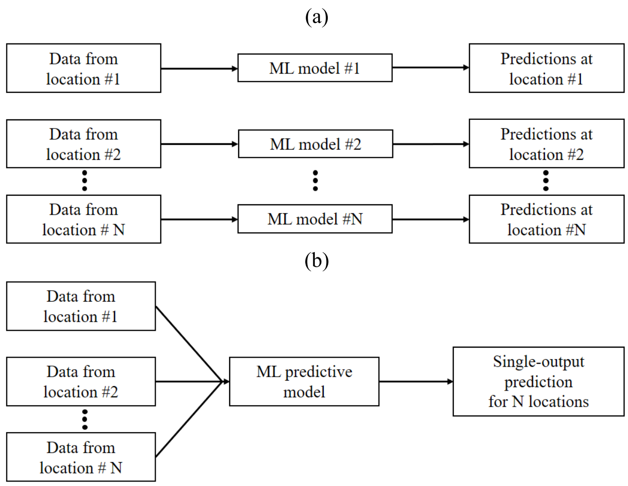


FIGURE 3. Multi-target approaches to forecast traffic in multiple locations. (a) Local approach. (b) Global approach.

On the other hand, when the forecasting problem requires predictions to be made simultaneously for multiple locations, it can be handled as a multi-target regression problem [7]. In this case there are two possible strategies to solve this problem, such as the ones shown in Figure 3. The first is to define a local approach [7] that transforms the multiple locations into independent problems, and each of them is then solved using a simple regression approach. This means developing one ML model for every location in which predictions are performed. The second strategy is to use a global approach [7] that adapts a single ML model to directly handle multiple data-sets coming from the locations under study, and then forecasting traffic for all of them using a single output. This last approach is usually more challenging because it aims not only to predict traffic at the multiple locations, but also to model the dependencies among the locations [7], which can increase the computational cost.

3) PROBLEM DEFINITION

Within this attribute, we have identified four sublevels. The first sublevel is related to the type of data used as an input. Based on [116], there are three types of inputs that can be used at the moment of feeding a ML model: only using temporal traffic data, temporal and spatial traffic data, and non-traffic data (e.g. calendar data) to enrich any of the two aforementioned inputs. The second sublevel consists of the time horizons for which predictions are made. In this taxonomy we categorize them into two groups: short- and long-term time horizons. The third sublevel determines the time step defined for the input data, which according to [81], [115] can be grouped in three ranges: high, medium, and small resolutions. Finally, the fourth sublevel is the output definition in terms of the kind of traffic variable to be predicted. Based on [115], the output variables can be travel time or a variable within the group of fundamental macroscopic variables (flow, density, occupancy, speed). The four aforementioned sublevels are presented below.

a: INPUT DEFINITION

A proper definition of input data is of great importance to ensure the good performance of ML methods in general [92], which also applies to TF. The main idea behind TF is to make predictions from a few seconds to possibly a few hours into the future using current and past traffic data, which is known as the temporal domain of traffic data [116]. Nevertheless, in recent research, including spatial traffic data has been an important consideration in TF [57]. Several research articles have supported the improvement of predictions due to the incorporation of upstream or downstream traffic data [18], [19], [50], [64], [65], [67], [97], [127], [129], [131], [132].

On the contrary, there are factors that affect traffic but are not part of its pattern behavior, such as weather or calendar data. Feeding this type of non-traffic data into ML methods can enhance their predictions [6]. In this context, TF problems can contain up to three categories of input data, which are described below: temporal, temporal and spatial, and non-traffic data.

TEMPORAL

The traffic predictions can be made using only temporal traffic data from the scale of prediction under study. In the transportation literature, we can find studies that use either one time-series associated with a single traffic measure or several time-series of different measurements taken by the same point detector or a pair of interval sensors [116]. These two approaches are compared in Figure 4.

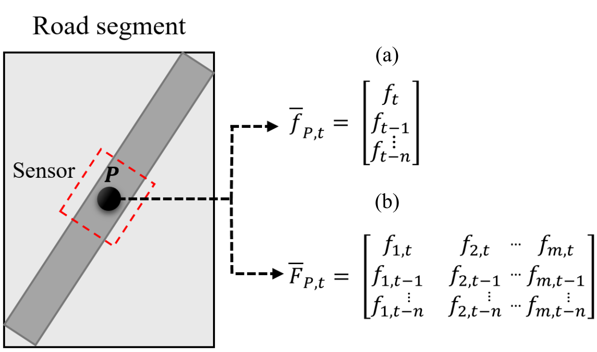


FIGURE 4. Temporal traffic data. Adapted from [79]. (a) One time-series. (b) More than one time-series.

Feeding ML models with only temporal data can be considered the most simple TF problem [46], [66], [73], [76], [101], [140]. In the case of only using one time-series, the definition of the input data ends with a vector representation, which includes the current state f of the traffic variable of interest at time t and its previous n states f_{t-n}. On the other hand, when using more than one time-series belonging to m traffic variables, the input definition is a matrix representation that includes the current traffic states $f_{m,t}$ of the variables and their previous n values $f_{m,t-n}$. It is important to clarify that in both cases, the preparation of the data-set results in a structure that contains features and samples. This is explained in more detail in the Modelling specifications (see Section III-3).

TEMPORAL AND SPATIAL

Although there is extensive literature demonstrating that reasonable accuracy can be achieved using only temporal traffic data [1], [61], [76], [77], [82], [133], [139], there is consistent evidence that shows how incorporating the spatial component of traffic can improve the accuracy of predictions [30], [57]. Representing the predictions of traffic as a function of time and space is theoretically valid taking into account that temporal and spatial data from other physical locations allows the dynamics of traffic to be captured [115].

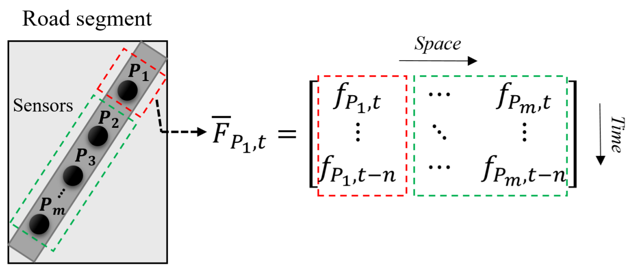


FIGURE 5. Temporal and spatial traffic data. Adapted from [79].

The input definition of this data leads to a matrix representation that incorporates time and space domains using the temporal data from both the target position $P_{1,t-n}$ and other m positions $P_{m,t-n}$, as is shown in Figure 5. This process requires identifying the spatial correlations of the locations under study to determine what positions are included and which can increase the computational complexity of the problem at hand. To reveal this dependency, it is necessary to examine the target location with its upstream and downstream locations and surrounding areas, which can be tackled by means of different methods, such as correlation analysis or by including a set of measurement positions within a pre-defined r radius [64], [67], [88]. Once this has been done, the next step is to transform the matrix representation to a data-set format with features and samples (see Section III-3).

NON-TRAFFIC DATA

Factors that are not part of the seasonal dynamism of traffic can play an important role in the accuracy of predictions. Weather and time of year, among others, are elements that matter in the forecasting process. Although some of them can be difficult to predict, their inclusion in the preparation of data-sets enhances the performance of ML models [6], [57], [79], and leads to developing responsive forecasting schemes that improve the decision making process of traffic management [57], [116].

Transportation literature reports that the inclusion of these types of data with temporal and spatial traffic data is an open issue [57], wherein calendar and weather data are the most frequently used exogenous variables for TF [11], [42], [45], [57], [80], [101], [119], [122], [124]. From a data preprocessing perspective, the incorporation of non-traffic data increases the model complexity and dimensionality of the data-sets. Besides, it is necessary to take standard procedures into account to integrate exogenous factors from different data sources into the traffic data provided by point, interval, and moving sensors (read further about approaches used for data fusion in traffic forecasting in [31]).

b: TIME HORIZON

This attribute represents the extension of time into the future over which a traffic variable is predicted. In it we have identified two sublevels: a Short-term attribute that categorizes the most common TF problems, all of which fall into the time horizon interval at less than 60 minutes ([4], [18], [19], [50], [64], [65], [67], [88], [88], [89], [94], [97], [102], [103], [106], [129], [131], [132]), and a Long-term attribute that enables the categorization of TF problems focused on long-term time horizons at more than 60 minutes [1], [42], [51], [82], [101], [122], [126], [127], [133].

According to transportation literature [27], [81], [115], it has been observed that longer time horizons generally lead to greater inaccuracy, and therefore most recent literature reviews show that research commonly predicts traffic up to 60 minutes into the future [57], [116]. More specifically, research suggests the appropriate prognosis horizon is between the range of 5 to 30 minutes into the future [28], [115]; however, Laña *et al.* [57] recommend that long-term time horizons, beyond 60 minutes, are needed to improve the management of traffic flows at the network level. In this context, enhancing the forecasting capacity of ML methods is still an open issue approached by very few authors (see Section III and [57]).

c: TIME STEP

In ML data-sets preparation, defining the appropriate time step is an important issue because it affects the quality of information about traffic lying in the data. This attribute categorizes the time interval upon which predictions are made and the frequency of time used to reach the time horizon defined in every TF problem. Three sublevels are identified: High, Medium, and Small, and are described below.

In general, high resolution time steps (e.g. traffic data sensed every 30 seconds) incorporate noise to the input data [115], [116], and the resultant ML models are more prone to overfitting. The repercussion in data pre-processing is the necessity of using techniques to reduce noise in the input data [9], [49]. In contrast, there is the case of Small resolution time steps. They influence the elimination of important data variations and traffic information within the data [113], needed during the training of ML models to keep a balance between bias and variance.

According to [116], there is no solid approach to select the appropriate time step. However, the category of Medium resolution, which encompasses TF problems characterized for having time steps between 5 and 15 minutes, contributes to having an equilibrium between noise and the lose of valuable information within the data [115].

d: OUTPUT DEFINITION

Within every TF problem, there is a clear relation between the traffic measure to be predicted and the data source used for such a task [78]. This attribute defines what traffic measure is considered as the output of ML methods to describe the anticipated traffic conditions on the roads, depending on the type of data available. Here, we have identified two sublevels described below. First, there is the category of fundamental macroscopic traffic variables [115] that includes flow, density, occupancy, and speed.

On the other hand, the second sublevel is related to forecasting travel time which, as it is explained in the Travel Time Data Collection Handbook [108], provides a common ground for communication between transportation engineers, planners, administrators and non-expert travelers [78], [109], [115].

FUNDAMENTAL

This category includes flow, occupancy, and speed measures. Flow and occupancy are traffic measures directly taken by point and interval detectors at the locations where the sensors are situated [78]. Sometimes, the sensors are placed in every lane of a freeway or urban road, and it is up to the modeler to decide if the prediction would be performed for each individual lane or an aggregation of some or all of the lanes. In any of the two cases, the predictions obtained by ML methods describe the traffic state only at the place where the detectors are installed [78]. Current transportation literature shows that research endeavors are well distributed between flow and occupancy forecasting without any special considerations at the moment of choosing any of the two traffic measures [4], [18], [50], [67], [89], [94], [102], [103], [126].

The prediction of speed is strongly connected to whether the available data source senses this traffic measure or not. If the available data comes from point detectors, which do not have the capacity of sensing speed [78], this is calculated using flow and occupancy features to generate a third feature of velocity, giving rise to a speed estimation rather than a prediction. If the available data source is interval detectors, they can collect the speed of each vehicle by using its travel time between the two sensors [59]. This measure is called point speed, and it is only valid when describing the speed at the points where the sensors are located [98]. To prepare ML data-sets with this measure, the point speed of all vehicles passing the sensors must be aggregated, generally, by means of the time mean speed technique [108] to generate the velocity feature.

Finally, if the available data comes from moving sensors, speed can be incorporated in the GPS traces or not. In the event of not being included, after performing the map-matching task, it is possible to calculate the average speed for each road segment, travelled by every moving sensor, which has at least two GPS data points. The resultant measure is the traffic speed in that specific road segment for each vehicle sensed (for more details about how to pre-process this data see [35], [91]). Then to forecast speed, a time-series of traffic speed must be built by aggregating the calculated speeds of every moving sensor available. Such predictions have segment and network spatial coverage that offers more details, about the real traffic conditions on the roads, with respect to speed predictions based on either point or interval detectors data [19], [64], [129], [132].

TRAVEL TIME

This attribute is associated with another major direction of TF problems focusing on forecasting travel time, which is defined as the time needed to cross two fixed points along a freeway or urban road [115]. As in the case of speed prediction, travel time forecasting is connected to the availability of the appropriate data for such a task [78], [116]. If the available data source technology supports the direct sensing of travel time (e.g. GPS, AVI systems), it is predicted using the measured data taken by detectors [115] (see [78] for more details regarding sensing technologies of travel time).

In the case of using point sensors [5], [65], [90], [117], travel time forecasting is based on their capability of sensing point speed, and then, making use of trajectory methods [78] during the preparation of ML data-sets to calculate travel time. The main idea of these methods is to take a whole road and divide it into smaller segments, where each of them is defined as the length between two detectors. The detector at the beginning of one segment is called the upstream detector, while the one at the end of the segment is the downstream detector. With this configuration, the most simple way to extend the point speed measurements to the segment is by using piece-wise constant methods, where the speed taken in the initial sensor, which delimits the segment, represents the entire segment [8], [99], [111]. Thus, a travel time feature is generated based on the point speed measures belonging to the areas of interest.

MODELLING SPECIFICATIONS

In supervised ML, a regression problem is the process of learning a mapping function ($f : X \rightarrow Y$) between Y, the dependent variable/s, and X, the independent variable/s [13]. The focus is on modelling and predicting how the dependent variable/s change/s when the independent variable/s vary/ies over time, as is the case for TF.

Most of the transportation literature describes characteristics of TF problems centered on traffic-related attributes [30], [57], [78], [81], [109], [115], [116]. Nevertheless, none of them gives an account of criteria that categorize how to model: 1) the input and output data, 2) the steps of predictions, and 3) the pre-processing process of the input data.

In this section, we define the attributes Input Modelling, Output Modelling, Step Prediction, and Data Pre-processing. Each of them, together with their sub-attributes, are presented and describe below.

4) MODELLING

a: INPUT MODELLING

This attribute is related to how the samples of the input data are modeled to generate a data-set. We have identified two sub-levels within it; in the first case the samples are modeled as vectors, whilst the second sub-level categorizes TF problems where the input samples are modeled as high-order tensors. These two input modelling categories are described below.

VECTOR

This attribute presents the TF problems in which every input data sample is modeled with a vector representation. Two illustrate this, we present the following two examples.

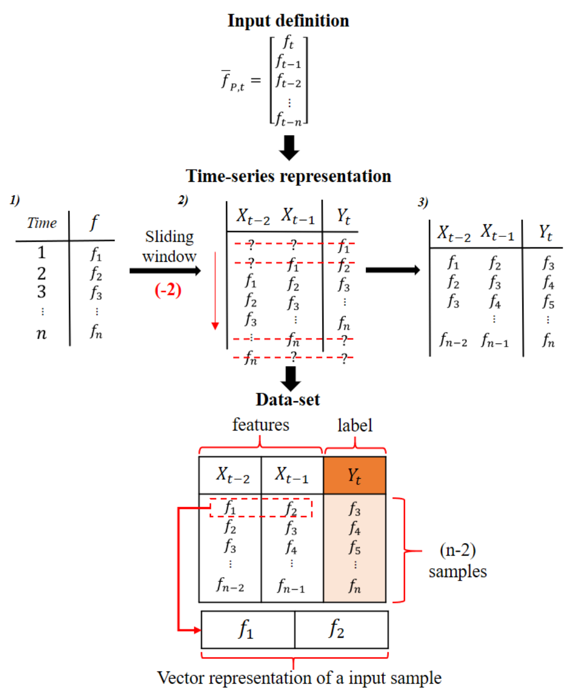


FIGURE 6. Vector to modelling temporal traffic data.

The first example corresponds to a basic TF problem whose input contains only temporal traffic data, as is depicted in Figure 6. More specifically, given a sequence of historical data of (n) measures, which belongs to a traffic parameter (f) sensed during a time interval (t) at a position (P), the sequence can be modeled to look like a supervised regression learning problem by means of the sliding window method [10].

According to Figure 6, after the input definition phase, the next step is to generate a time-series representation from the original sequence of data in which the order of the samples is preserved. Then, two copies of the time-series are used to generate two lagged-input variables named $X_{(t-1)}$ and $X_{(t-2)}$. Having done this, the sliding window method consists of using the values of the previous time steps, within $X_{(t-1)}$ and $X_{(t-2)}$, to predict the values at the next time steps in $Y_{(t)}$. For the illustrative purposes of this example, we use a window size value of two. Nevertheless, careful thought and experimentation are needed within every TF problem to find a window width that results in acceptable model performance.

After applying the sliding window method, the first two rows of $X_{(t-1)}$ and $X_{(t-2)}$ can be seen to have insufficient data to predict the $f_{(1)}$ and $f_{(2)}$ values in $Y_{(t)}$. Besides, there are no known next values in $Y_{(t)}$ to be predicted using the $f_{(n)}$ measures of $X_{(t-1)}$ and $X_{(t-2)}$. These four rows of missing values are deleted to obtain the final data-set, which has two features ($X_{(t-1)}$ and $X_{(t-2)}$) and a column of target labels of real values ($Y_{(t)}$). In this resultant data structure, each pair ($X_{(i-1)}$, $X_{(i-2)}$) of samples is modeled as a two-dimension vector that can be fed into any of the standard linear and nonlinear ML methods.

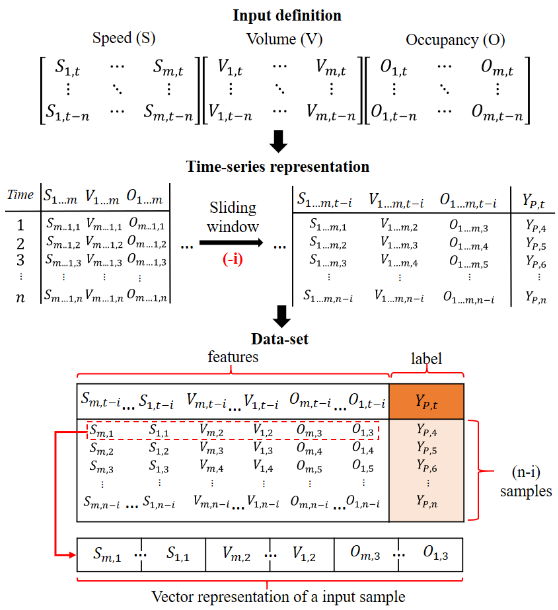


FIGURE 7. **Vector to modelling temporal and spatial traffic data.**

The second example is presented in Figure 7, which shows a more complex TF problem. In the input definition phase, there are three different traffic measures: speed (S), volume V, and occupancy O. Each of them has (m) sequences of historical data sensed by (m) detectors spatially ordered along a road segment, during a time interval (t).

Following the process described in Figure 6, the sliding window method re-frames the input data as a supervised regression learning problem with a generic window value of (i). Having done this, the resultant process leads to up to $S_{(m,t-i)}$, $V_{(m,t-i)}$, and $O_{(m,t-i)}$ features being obtained. Their values at the ($n-i$) previous time steps are used to forecast the real values within $Y_{(P,t)}$, P being the position where the target detector is located.

HIGH-ORDER TENSOR

This attribute categorizes the TF problems wherein the input data is modeled using a tensor of an order greater than two. Figure 8 illustrates how the three traffic measures (*speed*, *volume*, *occupancy*), used in the example of Figure 7 are put together to generate a three-dimensional data matrix of size (m, n, 3). This 3-D data structure contains ($3*m$) rows whose historical sequences are sensed during (n) time steps over (m) detectors spatially ordered on a road segment.

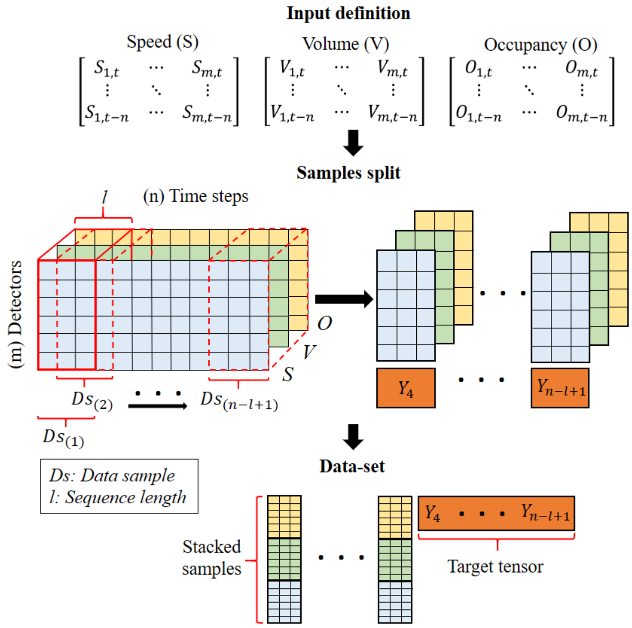


FIGURE 8. **High-order tensor to modelling temporal and spatial traffic data.**

The next step is to define a sequence length (for this example, with a value of 3) to simultaneously sample the 3 dimensions of the matrix and extract ($n - l + 1$) data samples of size (m, 3, 3), as is shown in Figure 8. These data samples are vertically stacked to produce the data-set that is commonly fed into Deep Learning methods, which are able to capture the spatial dependencies of traffic data with respect to traditional ML methods in a more realistic way [50], [65], [132]. As with the sliding window method, the sequence length needs to be carefully defined by means of experimentation to achieve appropriate model performance.

b: OUTPUT MODELLING

This attribute categorizes TF problems depending on the number of traffic variables to be predicted. Within it, we have identified two sub-levels; on the one hand, if there is only one variable to forecast, the modelling process leads to a single-target regression problem. Contrarily, if the TF problem at hand is focused on making predictions for more than one variable, the modelling process of the output is a multi-target

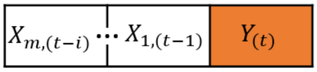


FIGURE 9. Single-target output.

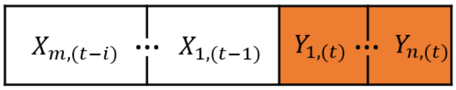


FIGURE 10. Multi-target output.

regression problem. These two input modelling processes are described below.

SINGLE-TARGET

This attribute categorizes TF problems in which there is only one traffic measure to be predicted (Figure 9). From a ML perspective, this is known as a standard regression problem [13] that consists in funding via learning a function able to predict, for either a given vector or high-order tensor sample, a real value among a continuous range.

As shown in Figure 9, this problem has the form $Y_{(t)} = X_{1,(t-i)}$ wherein $Y_{(t)}$ is the output vector that contains the value of the following steps values to forecast with the previous time steps of the input features $X_{m,(t-i)}$.

Currently, most research on the transportation literature is focused on predicting a single traffic variable using either only temporal traffic data or both temporal and spatial traffic data [4], [18], [19], [50], [64], [65], [67], [88], [88], [89], [94], [97], [103], [106], [126], [127], [129], [131], [132]. More specifically, the most common target is any of the fundamental macroscopic traffic variables [1], [11], [28], [46], [61], [66], [73], [76], [77], [80], [82], [101], [124], [139], [140]; whilst, the prediction of travel time has been less explored [90], [117], [127], [137].

MULTI-TARGET

This attribute presents those TF problems that have more than one output variable to be predicted (Figure 10). From a ML approach, this type of problem has an output space of more than one dimension [13]. As can be seen in Figure 10, a multi-target problem has the form $(Y_{(1)}, \ldots, Y_{(n)}) = (X_{1,(t-1)} \ldots X_{m,(t-i)})$ in which there are $Y_{(n)}$ traffic measures to be predicted based on the previous time steps of $X_{(m)}$ input features.

This multi-target regression approach has recently been assumed by the transportation community in order to model, in a comprehensive way, the exact form of traffic dynamics. It does generate, however, issues related to the selection of the proper data-driven method. The overall experience in multi-target modelling points out the use of non-parametric techniques, such as NNs [56], [115], to predict fundamental macroscopic traffic variables together with travel time [5], [102], [128].

c: STEP PREDICTION

This attribute is related to modelling the number of time horizons, in which traffic predictions are made in the future, within a TF problem. We have identified two sub-levels within this attribute: single, in which predictions are performed for a single time step in the future, and multiple, wherein it is necessary to forecast traffic over more than one time horizon. These two sub-levels are presented below.

SINGLE

This attribute categorizes TF problems in which there is only one time horizon. It corresponds to a single-target regression problem wherein there is a single traffic measure to be predicted, in the form $Y_{(t)} = X_{m,(t-i)}$, (t) being the target step.

In the transportation literature, TF problems with a single step of predictions are approached by means of parametric and non-parametric ML methods [18], [19], [64], [67], [89], [94], [102], [103], [106]. Nevertheless, from a practical perspective, single prediction intervals cannot support neither the short-term operational decision-making nor the medium- and long-term transportation planning, as they cannot provide a description of how traffic will evolve over time beyond few minutes into the future [57], [115].

MULTIPLE

This attribute classifies the TF problems wherein there is more than one future time step to be predicted. Specifically, there can be multiple time steps for either one target traffic measure or a set of them; which leads to a multi-target regression problem in the form $Y_{n,(t+i)} = X_{m,(t-i)}$. Let (n) be the output traffic measures and $(t+i)$ the number of time horizons to forecast.

Similar to the case of single step literature, TF problems with more than one time horizon in the future are quite well-documented in the literature [4], [50], [65], [88], [88], [97], [126], [127], [131], [132]. However, in contrast to single time step problems, multiple steps provide a stronger ground for decision makers during traffic flow management [56], [115].

d: DATA PRE-PROCESSING

This attribute categorizes TF problems in terms of the pre-processing tasks required to model their input data in a way that allows them to be processed by ML methods. According to Data Mining literature [34], Data Pre-processing (DPP) can be divided into Data Preparation and Data Reduction (more details of these two categories, with their respective techniques, can be consulted in [34]). However, for the purpose of the proposed taxonomy aimed at categorizing TF regression problems, we have split them into three categories: Data Preparation, Imperfect Data, and Data Reduction, all of which are presented below.

DATA PREPARATION

This attribute categorizes TF problems with respect to the approaches followed to properly prepare and transform the raw traffic data into the minimum format accepted by ML methods. This Data Preparation guarantees that the methods operate correctly without reporting errors during their run-time due to a no valid data format. Within this attribute, there are three approaches that are described below.

The first approach is data cleansing that is focused on detecting and discarding corrupt records within the raw data; this means, eliminating incorrect data that does not make sense in the context of traffic measures, for instance, negative values. The second is the data transformation process whose objective is to improve the input data to become more efficient in the forecasting process of ML methods. This task involves generating new features and normalizing or rescaling the input data to establish the same measurement unit or scale.

Finally, the third approach is data integration, which is related to the merging process of data that comes from different sources (for more details see [31]). In this case, such integration corresponds to the fusion of traffic data sensed by multiple sensors, or even, the inclusion of non-traffic data into the traffic data-sets.

IMPERFECT DATA

This attribute classifies TF problems in regard to the processes needed to fill in missing values and to identify noise in the raw data. In the first case, faulty reading, malfunctioning hardware or transmission errors of the traffic detectors can cause empty records. Filling in these gaps is important to guarantee accurate predictions, and can performed using different strategies ranging from a simple null value imputation to complex spatio-temporal context imputation models (see more details of imputation of missing data for road traffic forecasting in [58]). In the latter, the objective is dealing with noisy input data to detect random errors and to reduce data variability using smoothing techniques. This can improve the quality of the training data and avoid overfitting issues.

DATA REDUCTION

This attribute categorises TF problems in terms of Data Reduction techniques applied to obtain a reduced representation of the input data, which enables ML methods to reduce the computational cost of their training process. This representation is smaller; at the same time this maintains the integrity and variability of the original data [34], [37]. The three approaches that categorises the Data Reduction of TF problems are explained below.

First, the feature reduction approach, which aims to remove irrelevant-redundant features to find a minimum set of traffic attributes that increases the speed of the learning process (for more details see the review in [85]). The second approach is instance reduction, which is related to choosing a subset of samples from the whole traffic data-set to achieve the desired forecasting performance as if the complete data was used by the ML methods [34].

Lastly, the third approach is discretisation, which transforms continuous data values into categorical values within a finite number of intervals. This is useful for the integration of non-traffic data such as weather conditions, wherein raining continuous values are mapped into discrete attributes that classify the climate conditions on the roads based on nominal values.

IV. TAXONOMY ANALYSIS

This section presents the categorization of transportation literature by means of the proposed taxonomy. Specifically, the first part of the section is devoted to classifying the TF problems approached by the transportation literature reviewed between 2000 and 2019. Then, the second part introduces a hierarchical clustering analysis of this categorization of the literature to extract families of TF problems and to map the ML methods and DPP approaches used to tackle the extracted families of problems.

A. CATEGORISATION OF TRANSPORTATION LITERATURE

In this section, we analyses the taxonomy to check its robustness and its ability to discriminate papers that have approached different TF regression problems. Table 1 is devoted to categorizing a sample of relevant articles published between 2017 and 2019, whilst Table 2 includes references dated from 2014 to 2016 that are taken from the most updated survey in TF under a data-driven perspective [57]. Finally, Table 3 exhibits a sample of the main references from 2000 to 2013.

Every paper is categorized according to the sub-attributes within the four main blocks of the taxonomy whose names (1. Data Source, 2. Scope, 3. Problem Definition, 4. Modelling) are the shaded columns in the Tables 1, 2 and 3. When a particular sub-attribute is present in the TF problem approached by an article, its acronym (see Figure 1) is added to that specific cell. All described sub-attributes are present in at least one paper, which shows that no unnecessary attributes have been introduced in the taxonomy.

Figure 11 presents a summary of the findings extracted from the categorization of the literature published between 2000 and 2019. Concretely, on the x-axis of the figure all the attributes within the four blocks of the taxonomy can be seen. For each attribute, there is a stacked bar that depicts the percentage of the transportation literature (y-axis) in terms of the number of articles, which has addressed its sub-attributes represented with bullet points in the taxonomy (Figure 1).

As can be seen in Figure 11, regarding attribute **1. Data Source**, point detectors are the most common sensing technology used. The majority of the transportation literature categorized has approached TF problems that contain exclusively this type of data source. Specifically, only three papers [19], [119], [128] highlighted in Tables 1 and 2 include more than one kind of data source. In this sense, the integration of multiple data sources is an opportunity from the transportation perspective and a challenge in the ML area. Using multiple data technologies can contribute to representing the traffic conditions on the roads in a more realistic way;

TABLE 1. Transportation literature, published between 2017 and 2019, categorized by the taxonomy.

References	1. Data source	2. Scope	2.1 Context	2.2 Spatial coverage	2.3 Target	3. Problem definition	3.1 Input definition	3.2 Time horizon	3.3 Time step	3.4 Output definition	4. Modelling	4.1 Input modelling	4.2 Output modelling	4.3 Step prediction	4.4 Data pre-processing
[88]	Pt		Ub	Sg	Mt		T	ST	Md	Fd		Vc	ST	Mt	ID + DR
[131]	Pt		**Fw + Ub**	Sg	Mt		TS	ST	Md	Fd		**HT**	ST	Mt	DP + ID
[102]	Pt		Ub	Pt	Sn		T	ST	Md	Fd		Vc	**MT**	Sn	-*
[4]	Pt		Ub	Pt	Mt		T	ST	Md	Fd		Vc	ST	Mt	-*
[126]	Pt		Fw	Sg	Mt		T	LT	Hg + Md	Fd		Vc	ST	Mt	-*
[97]	Pt		Fw	Sg	Sn		TS	ST	Hg	Fd		Vc	ST	Mt	DP
[18]	Pt		Fw	Pt	Mt		TS	ST	Md	Fd		Vc	ST	Sn	DP
[103]	Pt		Fw	Pt	Sn		T	ST	Hg	Fd		Vc	ST	Sn	-*
[106]	Pt		Fw	Pt	Mt		T	ST	Md	Fd		Vc	ST	Sn	-*
[127]	It		Fw	Sg	Sn		TS	LT	Md	TT		Vc	ST	Mt	DP
[65]	Pt		Fw	Pt	Mt		TS	ST	Md	Fd		**HT**	ST	Mt	DP
[132]	Mv		Ub	Nw	Sn		TS	ST	Md	Fd		**HT**	ST	Mt	-*
[50]	Mv		Ub	Nw	Sn		TS + ND	ST	Md + Sm	Fd		**HT**	ST	Mt	DP
[67]	Pt		Fw	Pt	Mt		TS	ST	Md	Fd		Vc	ST	Sn	DP
[19]	**Pt + Mv**		**Fw + Ub**	Pt + Sg	Mt		TS	ST	Md	Fd		**HT**	ST	Sn	DP + ID
[64]	Mv		Ub	Nw	Sn		TS	ST	Md	Fd		Vc	ST	Sn	DR
[94]	It		Fw	Sg	Sn		T	ST	Md	Fd		Vc	ST	Sn	-*
[89]	Pt		Ub	Sg	Sn		T	ST	Md	Fd		Vc	ST	Sn	ID + DR
[129]	Mv		Ub	Sg	Mt		TS	ST	Md	Fd		Vc	ST	Sn	ID + DR

TABLE 2. Transportation literature, published between 2014 and 2016, categorized by the taxonomy.

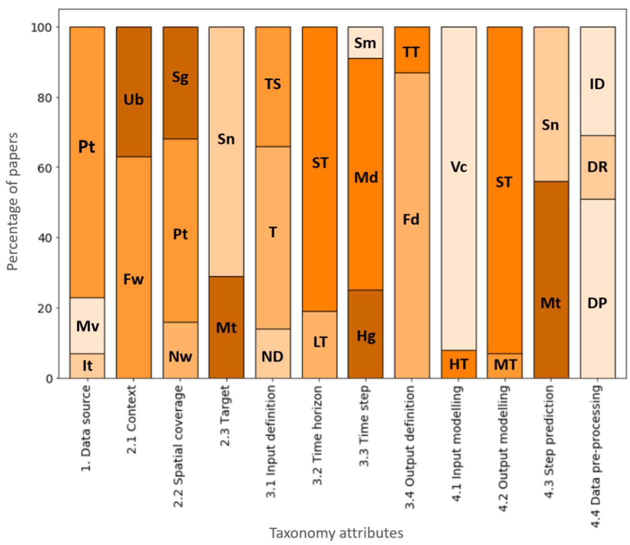
References	1. Data source	2. Scope	2.1 Context	2.2 Spatial coverage	2.3 Target	3. Problem definition	3.1 Input definition	3.2 Time horizon	3.3 Time step	3.4 Output definition	4. Modelling	4.1 Input modelling	4.2 Otput modelling	4.3 Step prediction	4.4 Data pre-processing
[73]	Pt		Ub	Sg	Sn		T	ST	Hg	Fd		Vc	ST	Sn	DP + ID
[128]	**Pt + It + Mv**		Fw	Sg	Sn		T	ST	Md	Fd + TT		Vc	**MT**	Mt	ID + DR
[11]	Mv		Ub	Nw	Sn		TS + ND	ST	Md	Fd		Vc	ST	Mt	DP + ID
[101]	Pt		Fw	Sg	Sn		T + ND	LT	Md	Fd		Vc	ST	Mt	DP
[82]	Pt		Ub	Pt	Mt		T	LT	Sm	Fd		Vc	ST	Mt	DP + ID
[124]	Pt		Ub	Nw	Sn		TS + ND	ST	Hg	Fd		Vc	ST	Mt	DP
[61]	Pt		Ub	Pt	Mt		T	ST	Md	Fd		Vc	ST	Sn	ID
[77]	Pt		Ub	Pt	Mt		T	ST	Sm	Fd		Vc	ST	Mt	-*
[80]	It		Ub	Nw	Sn		TS + ND	ST	Hg	Fd		Vc	ST	Mt	DP + DR
[76]	Pt		Fw	Pt	Sn		T	ST	Hg	Fd		Vc	ST	Mt	ID
[1]	Pt		Fw	Sg	Sn		T	LT	Hg	Fd		Vc	ST	Mt	-*
[46]	Pt		Fw	Pt	Sn		T + ND	ST	Md	Fd		Vc	ST	Sn	DP
[140]	Pt		Fw	Sg	Sn		TS	ST	-	Fd		Vc	ST	-	DP
[28]	Pt		Fw	Nw	Sn		TS	ST	Hg	Fd		Vc	ST	Sn	DP
[139]	Pt		Ub	Pt	Mt		T	ST	Md	Fd		Vc	ST	Sn	-*
[66]	Pt		Fw	Pt	Sn		T + ND	ST	Hg	Fd		Vc	ST	Mt	DP + DR
[5]	Pt		Fw	Pt	Mt		TS	ST	Hg	Fd + TT		Vc	**MT**	Sn	DP + ID
[133]	Pt		Fw	Pt	Mt		T	LT	Md	Fd		Vc	ST	Mt	DR
[90]	Pt		Fw	Nw	Sn		TS	ST	Hg	TT		Vc	ST	Mt	DP
[42]	Pt		Fw	Sg	Sn		T + ND	LT	Hg	Fd		Vc	ST	Mt	DP
[122]	Pt		Fw	Pt	Sn		T + ND	LT	Md	Fd		Vc	ST	Mt	DP
[138]	Pt		Fw	Sg	Sn		T	ST	Md	Fd		Vc	ST	Sn	ID
[62]	Pt		Fw	Pt	Sn		T	ST	Md	Fd		Vc	ST	Sn	-*
[71]	Pt		Fw	Pt	Sn		TS	ST	Md	Fd		Vc	ST	Sn	DP
[24]	Pt		Fw	Sg	Sn		TS + ND	ST	Md	Fd		Vc	ST	Mt	DP
[72]	Pt		Fw	Pt	Sn		T	ST	Md	Fd		Vc	ST	Sn	DP
[23]	Pt		Fw	Pt	Sn		T	ST	Md	Fd		Vc	ST	Mt	DP
[32]	Mv		Ub	Nw	Sn		T	ST	Md	Fd		Vc	ST	Mt	DR
[120]	Mv		Fw	Sg	Sn		TS	ST	Hg	Fd		Vc	ST	Sn	-*
[22]	Mv		Ub	Pt	Sn		T	ST	Md	Fd		Vc	ST	Sn	DP
[117]	It		Ub	Nw	Sn		TS	ST	Md	TT		Vc	ST	Mt	ID
[137]	Mv		Fw	Sg	Sn		TS + ND	ST	Md	TT		Vc	ST	Mt	ID
[51]	Pt		Ub	Pt	Sn		T	LT	Md	Fd		Vc	ST	Mt	DP
[119]	**Pt + It + Mv**		Fw	Nw	Sn		T + ND	ST	Md	Fd		Vc	ST	Sn	DP
[56]	Pt		Ub	Pt	Mt		T	LT	Md	Fd		Vc	ST	Mt	DR
[75]	Pt		Ub	Pt	Mt		T	ST	Md	Fd		**HT**	ST	St	-*
[125]	Pt		Ub	Nw	Sn		TS	ST	Md	Fd		Vc	ST	Mt	DP + ID

however, such a data fusion is a demanding modelling task during the pre-processing the input data.

The aforementioned data source availability also affects the spatial coverage of predictions. The stacked bar of attribute **2.2 Spatial Coverage** shows that few papers have approached TF problems at the network level, which is in concordance with the low usage of moving data due to privacy and availability issues [78]. On the other hand, the most

TABLE 3. Transportation literature, published between 2000 and 2013, categorized by the taxonomy.

References	1. Data source	2. Scope	2.1 Context	2.2 Spatial coverage	2.3 Target	3. Problem definition	3.1 Input definition	3.2 Time horizon	3.3 Time step	3.4 Output definition	4. Modelling	4.1 Input modelling	4.2 Output modelling	4.3 Step prediction	4.4 Data pre-processing
[112]	Pt		Fw	Sg	Sn		TS	ST	Hg	TT		Vc	ST	Mt	DP + ID
[26]	Pt		Fw	Pt	Sn		T	ST	Hg	Fd		Vc	ST	Mt	-*
[63]	Pt		Fw	Sg	Sn		T	ST	Md	Fd		Vc	ST	Sn	DP + DR
[15]	Pt		Fw	Pt	Sn		T	ST	Md	Fd		Vc	ST	Sn	DR
[48]	Pt		Fw	Pt	Sn		TS	ST	Hg	Fd		Vc	ST	Mt	DP + ID
[47]	Pt		Fw	Sg	Sn		T	ST	Md	Fd		Vc	ST	Mt	ID
[135]	Pt		Fw	Pt	Sn		T	LT	Sm	TT		Vc	ST	Mt	-*
[130]	Mv		Ub	Sg	Sn		TS	ST	Hg	Fd		Vc	ST	Sn	-*
[83]	Pt		Fw	Pt	Sn		T	ST	Md	Fd		Vc	ST	Sn	-*
[29]	Pt		Fw	Pt	Mt		T	ST + LT	Hg + Md + Sm	Fd		Vc	MT	Mt	-*
[141]	Pt		Fw	Pt	Mt		T	ST + LT	Sm	Fd		Vc	ST	Mt	-*
[100]	Pt		Ub	Pt	Sn		T	ST	Md	Fd		Vc	ST	Sn	-*
[134]	Pt		Fw	Pt	Mt		T	ST	Md	Fd		Vc	ST	Sn	DP + ID
[36]	Pt		Ub	Pt	Sn		T	ST	Md	Fd		Vc	ST	Mt	-*
[39]	Pt		Fw	Pt	Sn		TS	ST	Md	Fd		Vc	ST	Sn	DP
[118]	Pt		Ub	Pt	Mt		T	ST	Hg + Md	Fd		Vc	ST	Mt	-*
[14]	Pt		Fw	Pt	Sn		TS	ST	Md	Fd		Vc	MT	Sn	DP + ID
[136]	Pt		Fw	Pt	Sn		T	LT	Sm	TT		Vc	ST	Mt	-*
[25]	Pt		Fw	Sg	Sn		T	LT	Md	TT		Vc	ST	Mt	-*
[104]	Pt		Fw	Sg	Sn		TS	ST	Md	Fd		Vc	ST	Sn	DP + ID + DR



1. Data Source: Point (Pt), Interval (It), Moving (Mv); ***2.1 Context:*** Freeway (Fw), Urban (Ub); ***2.2 Spatial Coverage:*** Point (Pt), Segment (Sg), Network (Nw); ***2.3 Target:*** Single (Sn), Multiple (Mt); ***3.1 Input Definition:*** Temporal (T), Temporal + Spatial (TS), Non-traffic Data (ND); ***3.2 Time Horizon:*** Short-term (ST), Long-term (LT); ***3.3 Time step:*** High (Hg), Medium (Md), Small (Sm); ***3.4 Output Definition:*** Fundamental (Fd), Travel Time (TT); ***4.1 Input Modelling:*** Vector (Vc), High-order Tensor (HT); ***4.2 Output Modelling:*** Single-target (ST), Multi-target (MT); ***4.3 Step Prediction:*** Single (Sn), Multiple (Mt); ***4.4 Data Pre-processing:*** Data Preparation (DP), Imperfect Data (ID), Data Reduction (DR).

FIGURE 11. Summary of 2000-2019 TF literature categorized by the taxonomy.

common TF problems are the ones in which the prediction of traffic is focused on either point or segment levels, which can be done using point or interval detectors, the last data source being the least used in the transportation literature.

In the case of attribute **2.1 Context**, there is an apparent balance in the literature between the TF problems handled in either urban or freeway contexts. Only two studies in Table 1 approach the prediction of traffic in both contexts at the same time. In this case, the computational challenge is focused on developing methods that learn traffic patterns of both contexts or developing models that separately predict traffic in the two environments.

With respect to the attribute **2.3 Target**, Figure 11 shows that most of the literature approaches TF problems that consider a single spatial target for predictions. Multiple targets are considered when traffic is only forecasted for more than one point or segment. This because approaching multiple networks does not have any practical application from a transportation perspective. With respect to multiple targets, research is still needed to find out whether adopting a local or a global strategy to handle the prediction along multiple locations would more suitable in terms of computational cost and accuracy.

In the case of attribute **3.1 Input Definition**, the majority of papers have addressed the prediction of traffic using only temporal traffic data. Nevertheless, the inclusion of spatial traffic data and non-traffic data to enhance TF has been taken into account, as can be seen in Figure 11. Nevertheless, in most of the cases such input enrichment is modeled using a vector representation (stacked bar **4.1 Input Modelling**), which can lead to losses in the spatial dependencies of traffic data. As is highlighted in Tables 1 and 2, few studies [19], [50], [65], [75], [131], [132] have used a high-order tensor to generate multi-dimensional data input structures that maintain the spatial relationships of traffic data whilst they are being feed into ML methods.

Regarding attributes **3.2 Time Horizon** and **3.3 Time Step**, for the most part of TF problems, the forecasting of traffic is focused on short-term predictions using medium resolution data (between 5 and 15 minutes). In the case of both small and high time steps, the literature avoids these data resolutions because of the loss of valuable information and the inclusion of noise in the input data, as is stated in [57], [115]. In this context, the data science challenge lies in determining how the structure of the input data, in terms of attributes **3.1 Input Definition** and **4.1 Input Modelling**, and the time step resolution (attribute **3.3 Time Step**) can contribute to obtain long-term predictions, without losing accuracy and increasing the computational cost of training ML models.

In the area of what traffic measure is predicted (attribute **3.4 Output Definition**), choosing any of the fundamental macroscopic variables is the predominant approach, despite the relevance of forecasting time travel being previously stated [78], [115], [116]. From the perspective of how many traffic parameters are predicted, the taxonomy demonstrates that most TF problems are focused on single-target output predictions (stacked bar **4.2 Output Modelling**). The latter opens the possibility to explore the benefits and challenges of multiple output targets, which according to Tables 1 and 2 has been approached by very few studies [5], [102], [128].

Finally, the DPP in TF problems (attribute **4.4 Data Pre-processing**) is mostly focused on data preparation tasks, particularly, data cleansing, data integration and data transformation. Notwithstanding, Tables 1, 2 and 3 show that many papers reviewed do not include any DPP. These cases are highlighted with the symbol (-*) in the column **4.4 Data Pre-processing** of the aforementioned tables, and although the authors performed the aggregation of input data into small, medium, or high time steps, this is not consider to be DPP because it is an implicit task for all TF problems. The absence of DPP could be due to the fact that many articles use data sources sensed by third parties, which pre-process the traffic data before making it available, such as the Caltrans Performance Measurement System case.

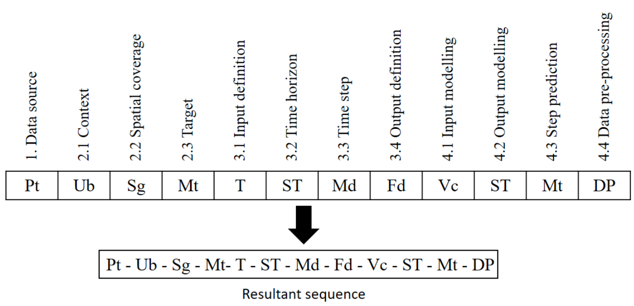


FIGURE 12. Example of sequence structure for a given TF problem.

B. FAMILIES OF TRAFFIC FORECASTING PROBLEMS

Based on the literature categorization shown above, this section presents families of TF problems, or, groups of problems that share common attributes of the taxonomy. To accomplish this purpose, first we generated a unique sequence for each reference in Tables 1, 2 and 3, which identifies the TF problem approached within every article. As can be seen in the example of Figure 12, the structure of each sequence is determined by both the traffic- and the modelling-related attributes assigned with the proposed taxonomy. We obtained 76 sequences that correspond to the total of papers reviewed between 2000 and 2019.

In the next step, we carried out a hierarchical clustering analysis considering only the traffic-related attributes. The clustering process was done comparing the 76 generated sequences, and those with up to 4 different transportation attributes were grouped in the same family. The modelling attributes were excluded in the generation of the families because we wanted to highlight common groups of TF problems, from a transportation perspective, regardless of how they have been modeled in the literature.

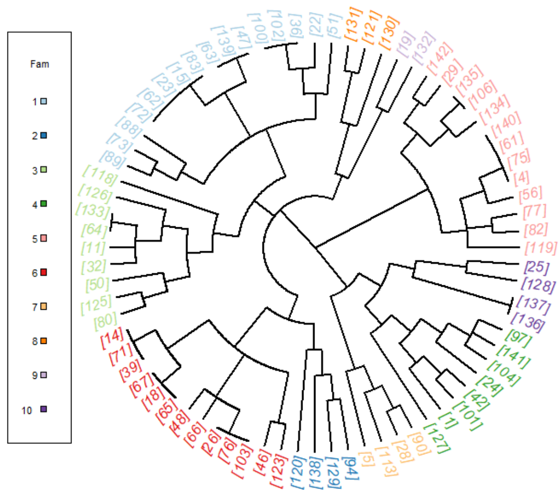


FIGURE 13. Hierarchical clustering analysis to extract families of TF problems.

We identified 10 families of TF problems that are shown in the dendrogram of Figure 13. The branches represent how close one paper is to other studies in terms of familiarity determined by the transportation attributes, and the colors depict the families. The leaf nodes of the tree structure contain the papers categorized in Tables 1, 2 and 3.

The 10 families of problems are presented with more details in Table 4, which shows what traffic attributes characterize them and how many of the papers reviewed have been approached within each family. The missing cells along a single family indicate that those transportation attributes, arranged in the columns of Table 4, do not belong to the family under consideration.

According to the information consolidated in Table 4, the families that have more complex characteristics, in terms of the integration of multiple data sources and the prediction of traffic at the network level, are those that represent the traffic conditions on the roads in a more realistic way (for instance, families 2, 3 and 9). However, these families have been approached by very few papers, which means that they are still open issues within the transportation literature, especially because of the computational challenges that they bring with them. In contrast, the simplest families (numbers 1,5 and 6) are those that contain the highest number of papers together with the more traditional transportation scenarios (point data sources and prediction both at the segment level and at single

TABLE 4. Families of TF problems with their main traffic attributes together with the number of works classified within each family. Guide of acronyms: ***1. Data Source***: Point (Pt), Interval (It), Moving (Mv); ***2.1 Context***: Freeway (Fw), Urban (Ub); ***2.2 Spatial Coverage***: Point (Pt), Segment (Sg), Network (Nw); ***2.3 Target***: Single (Sn), Multiple (Mt); ***3.1 Input Definition***: Temporal (T), Temporal + Spatial (TS), Non-traffic Data (ND); ***3.2 Time Horizon***: Short-term (ST), Long-term (LT); ***3.3 Time step***: High (Hg), Medium (Md), Small (Sm); ***3.4 Output Definition***: Fundamental (Fd), Travel Time (TT).

Family	Family's main traffic attributes								Total of works
	1.	2.1	2.2	2.3	3.1	3.2	3.3	3.4	
Family 1	Pt	-	-	Sn	T	ST	Md	Fd	16
Family 2	Pt + It + Mv	Fw	Sg	Sn	-	ST	Md	-	4
Family 3	Mv, It	Ub	Nw	Sn	TS + ND	ST	-	Fd	9
Family 4	Pt	Fw	Sg	Sn	-	ST, LT	-	Fd	8
Family 5	Pt	-	Pt	Mt	T	ST, LT	-	Fd	13
Family 6	Pt	Fw	Pt	-	T + ND, TS	ST	-	Fd	13
Family 7	Pt	Fw	-	Sn	TS	ST	Hg	-	4
Family 8	Mv	Ub	Sg	-	TS	ST-	Md	Fd	3
Family 9	Pt + Mv	Fw + Ub	Pt + Sg	Mt	TS	ST	Md	Fd	2
Family 10	Pt	Fw	-	Sn	T	LT	-	TT	4

points on the roads). How ML methods can successfully deal with these families is well documented in the literature, but, at the same time, it is still necessary to determine the most suitable ML methods, in terms of computational cost and efficiency, to approach these families of problems.

Having extracted the families of TF problems, Figure 14 maps how different classes of ML methods have been used to approach them over the last few years. The families of methods were identified grouping the prediction models used by the literature in Tables 1, 2 and 3. As can be seen in Figure 14, NNs were the dominant methods in those studies dated from 2000 to 2010, to approach families of TF problems that include point detectors as their main data source (families 1, 4, 5, 6, 7 and 10). From then to on, Instance-based methods, such as k-Nearest Neighbors and Support Vector Machines, have been used to tackle the same problems.

More complex families of TF problems that incorporate moving and interval data sources (families 2,3 and 9) only appeared in recent years, according to Figure 14. These problems include temporal and spatial dependencies of traffic within their data, which traditional ML methods have difficulties dealing with. In this context, Deep NNs have been introduced to the transportation literature to approach such complex problems and to enable making predictions at the network level. Nevertheless, many of the Deep NNs reported in the literature are still using vector representations to model the input data. In this sense, the benefits of these methods can be further exploited by using high-order tensors.

The remaining families of ML methods (Decision Tree-based, Linear Regression, Probabilistic), have been broadly applied to handle problems where the main data source is point detectors, the context of forecasting is freeway environments, and their spatial coverage of prediction are segments or single points. In such transportation contexts, Deep ANN could also handle these families of problems; however, the implementation of traditional ML methods guarantees low computational costs, in terms of time associated with the training process of the methods and the configuration of their hyper-parameters, while obtaining fair accuracy in predictions.

Finally, the influence of data DPP tasks on the families of TF problems is also discussed. Figure 15 shows how many

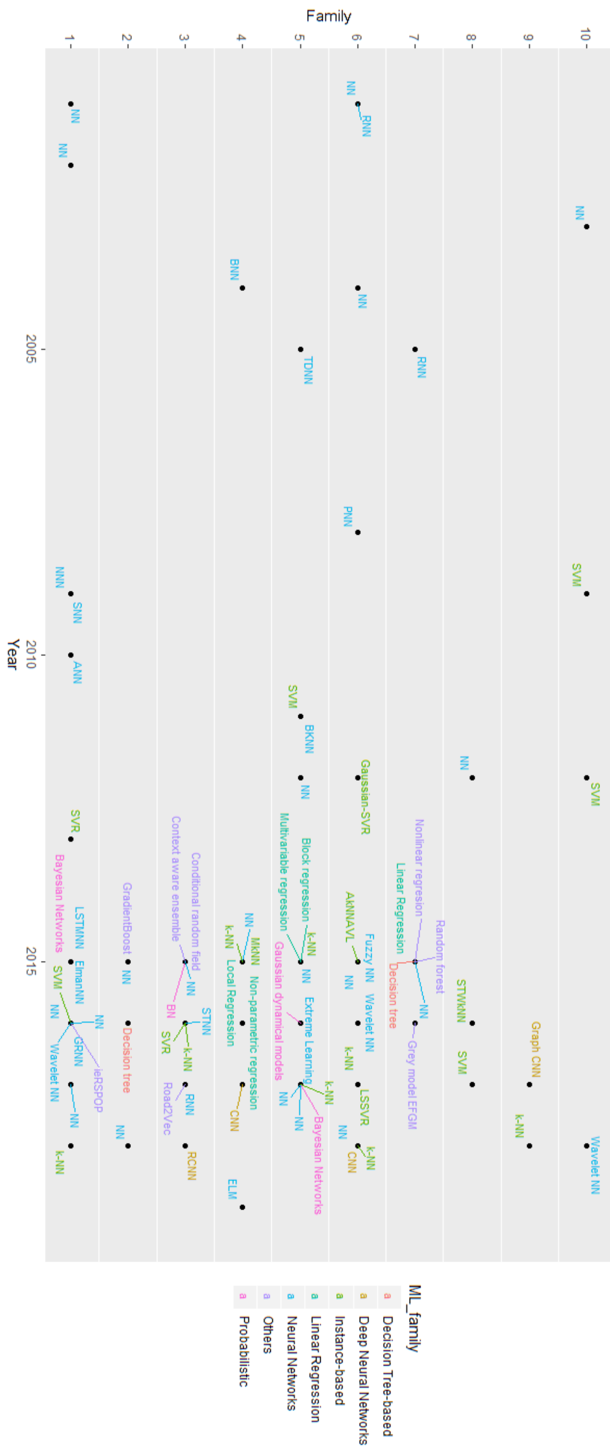


FIGURE 14. ML methods used to approach the 10 families of TF problems.

papers within every family of problems have included DPP. As shown below, there are 9 families in which there is, at least, one paper within them that does not incorporate DPP.

More concretely, Figure 16 presents how many times the DPP attributes of the proposed taxonomy (Data Preparation - DP, Imperfect Data - ID, Data Reduction - DR)

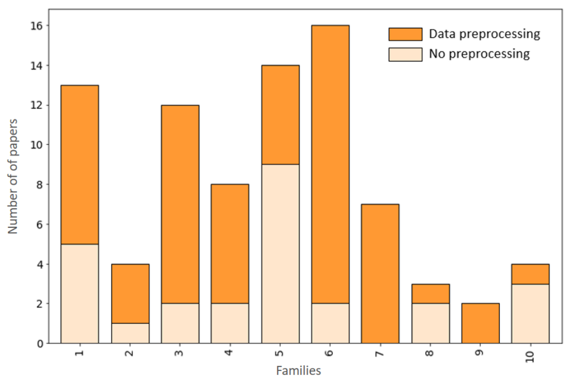


FIGURE 15. Number of papers that include DPP within families of TF problems.

FIGURE 16. DPP approaches within each family of TF problems.

have been used in the studies that include DPP in the 10 families. As can be seen, DP is the most common approach, followed by ID and DR, respectively. In this sense, although the pre-processing of input data is a fundamental stage in the modelling process of ML methods, in the transportation literature it has been scarcely studied, and therefore their influence in the model selection problem, given the characteristics of a particular TF family under study, is still an open issue.

V. CONCLUSIONS, FINDINGS, AND FUTURE TRENDS

In this paper, we have proposed a taxonomy to categorize families of TF problems from a supervised regression learning perspective. Concretely, the taxonomy is built based on two types of attributes to categorize the problems: traffic and modelling specifications. The first one includes transportation-related attributes, which are the type of data source, the context and spatial coverage of predictions, the input and output variables considered, the time horizon of predictions, and the time step of data. Diversely, the second set of attributes introduces specification about how the input and output data and the steps of predictions can be modeled, together with DPP approaches.

To check the robustness of the taxonomy, we categorized research studies published in the transportation literature from 2000 to 2019. As a result, the taxonomy analysis allowed the extraction of 10 families of TF problems, whose complexity change mainly depending on the number and type of data sources used and on the scale of predictions.

The families of TF problems most commonly approached are those that incorporate point detectors as the main data source. In these cases, traditional ML methods such as NNs, regression and instance-based methods are the predominant strategies to address the above-mentioned problems. Conversely, the families of less frequently addressed and more complex problems are those that include moving and interval data sensors to make predictions beyond single points on the roads. For this last type of problem, the common approach is to use Deep NNs to exploit the temporal and spatial characteristics of traffic data.

From the aforementioned analysis, it can be said that NNs were the predominant traffic prediction methods from 2000 to 2010. Since then, Instance-based methods started becoming more commonly used to tackle TF, especially in the cases where the input data came from point and interval detectors. However, nowadays, Deep NNs have progressively substituted these approaches as data obtained by moving sensors has become available. The latter methods are the most suitable for handling complex transportation scenarios which require predictions at the network level; nevertheless, their potential is far from being totally exploited because of most them are still using vector representations to model their input data. In this sense, the benefits of Deep NNs methods can be further exploited by using high-order tensors.

In addition to map out and discuss what ML methods have been used to approach the families of TF problems, it is important to highlight the absence of DPP techniques in the modelling process of TF problems. The current trend is focused on the integration of traffic data coming from different data sources and the improvements of imperfect data (missing values and noisy data). Nevertheless, in spite of this, the influence of data reduction approaches, such as instance and feature reduction, over the complexity of TF problems are still an open issue.

In summary, the taxonomy analysis suggests that there is no best ML algorithm that suits all forecasting situations. However, authors state the necessity of a comparison framework that allows the description and analysis of the performance of different ML methods in diverse TF problems. In this context, the taxonomy presented sets up the basis for a common framework that, in further research, will facilitate experimentation to determine which ML algorithms are more appropriate for each family of TF problems. It may also facilitate determining what would be a suitable baseline of algorithms to make fair comparisons, depending on the chosen algorithm for the TF problem at hand.

REFERENCES

[1] J. Abdi and B. Moshiri, "Application of temporal difference learning rules in short-term traffic flow prediction," *Expert Syst.*, vol. 32, no. 1, pp. 49–64, Feb. 2015.

[2] T. H. H. Aldhyani and M. R. Joshi, "Clustering to enhance network traffic forecasting," in *Information and Communication Technology for Sustainable Development*, D. K. Mishra, M. K. Nayak, and A. Joshi, Eds. Singapore: Springer, 2018, pp. 357–364.

[3] J. S. Angarita-Zapata, I. Triguero, and A. D. Masegosa, "A preliminary study on automatic algorithm selection for short-term traffic forecasting," in *Intelligent Distributed Computing XII*, J. Del Ser, E. Osaba, M. Bilbao, J. Sanchez-Medina, M. Vecchio, and X. S. Yang, Eds. Cham, Switzerland: Springer, 2018, pp. 204–214.

[4] A. Attanasi, L. Meschini, M. Pezzulla, G. Fusco, G. Gentile, and N. Isaenko, "A hybrid method for real-time short-term predictions of traffic flows in urban areas," in *Proc. 5th IEEE Int. Conf. Models Technol. Intell. Transp. Syst. (MT-ITS)*, Jun. 2017, pp. 878–883.

[5] A. Bezuglov and G. Comert, "Short-term freeway traffic parameter prediction: Application of grey system theory models," *Expert Syst. Appl.*, vol. 62, pp. 284–292, Nov. 2016.

[6] E. Bolshinsky and R. Freidman, "Traffic flow forecast survey," Dept. Comput. Sci., Technion–Israel Inst. Technol., Haifa, Israel, Tech. Rep. CS-2012-06, 2012.

[7] H. Borchani, G. Varando, C. Bielza, and P. Larrañaga, "A survey on multi-output regression," *Data Mining Knowl. Discovery*, vol. 5, no. 5, pp. 216–233, 2015.

[8] H. Botma and P. Bovy, *The free speed functions of drivers*. Delft, The Netherlands: Delft Univ. Press, 2000, pp. 53–74.

[9] D. Boto-Giralda *et al.*, "Wavelet-based denoising for traffic volume time series forecasting with self-organizing neural networks," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 25, no. 7, pp. 530–545, 2010.

[10] J. Brownlee, *Introduction to Time Series Forecasting With Python*, vol. 1, 1st ed. Melbourne, VIC, Australia: Jason Brownlee, 2018.

[11] P. Cai, Y. Wang, G. Lu, P. Chen, C. Ding, and J. Sun, "A spatiotemporal correlative k-nearest neighbor model for short-term traffic multistep forecasting," *Transp. Res. C, Emerg. Technol.*, vol. 62, pp. 21–34, 2016.

[12] P. S. Castro, D. Zhang, and S. Li, "Urban traffic modelling and prediction using large scale taxi GPS traces," in *Pervasive Computing*, J. Kay, P. Lukowicz, H. Tokuda, P. Olivier, and A. Krüger, Eds. Berlin, Germany: Springer, 2012, pp. 57–72.

[13] D. Charte, F. Charte, S. García, and F. Herrera, "A snapshot on nonstandard supervised learning problems: Taxonomy, relationships, problem transformations and algorithm adaptations," *Prog. Artif. Intell.*, vol. 8, no. 1, pp. 1–14, Apr. 2019.

[14] H. Chen, M. S. Dougherty, and H. R. Kirby, "The effects of detector spacing on traffic forecasting performance using neural networks," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 16, no. 6, pp. 422–430, 2001.

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

[16] J.-F. Chen, S.-K. Lo, and Q. H. Do, "Forecasting short-term traffic flow by fuzzy wavelet neural network with parameters optimized by biogeography-based optimization algorithm," *Comput. Intell. Neurosci.*, vol. 2018, Oct. 2018, Art. no. 5469428.

[17] W. Chen *et al.*, "A novel fuzzy deep-learning approach to traffic flow prediction with uncertain spatial–temporal data features," *Future Gener. Comput. Syst.*, vol. 89, pp. 78–88, Dec. 2018.

[18] X. Chen, Z. Wei, X. Liu, Y. Cai, Z. Li, and F. Zhao, "Spatiotemporal variable and parameter selection using sparse hybrid genetic algorithm for traffic flow forecasting," *Int. J. Distrib. Sens. Netw.*, vol. 13, no. 6, pp. 1–14, 2017.

[19] S. Cheng, F. Lu, P. Peng, and S. Wu, "A spatiotemporal multi-view-based learning method for short-term traffic forecasting," *Int. Soc. Photogram. Remote Sens.–J. Geo-Inf.*, vol. 7, no. 6, pp. 1–21, 2018.

[20] T. Cheng, J. Haworth, and J. Wang, "Spatio-temporal autocorrelation of road network data," *J. Geograph. Syst.*, vol. 14, no. 4, pp. 389–413, Oct. 2012.

[21] N. K. Chowdhury and C. K.-S. Leung, "Improved travel time prediction algorithms for intelligent transportation systems," in *Knowlege-Based and Intelligent Information and Engineering Systems*, A. König, A. Dengel, K. Hinkelmann, K. Kise, R. J. Howlett, and L. C. Jain, Eds. 2011, pp. 355–365.

[22] Y. Cong, J. Wang, and X. Li, "Traffic flow forecasting by a least squares support vector machine with a fruit fly optimization algorithm," *Procedia Eng.*, vol. 137, pp. 59–68, Dec. 2016.

[23] R. T. Das, K. K. Ang, and C. Quek, "ieRSPOP: A novel incremental rough set-based pseudo outer-product with ensemble learning," *Appl. Soft Comput.*, vol. 46, pp. 170–186, Sep. 2016.

[24] P. Dell'Acqua, F. Bellotti, R. Berta, and A. De Gloria, "Time-aware multivariate nearest neighbor regression methods for traffic flow prediction," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 6, pp. 3393–3402, Dec. 2015.

[25] A. Dharia and H. Adeli, "Neural network model for rapid forecasting of freeway link travel time," *Eng. Appl. Artif. Intell.*, vol. 16, nos. 7–8, pp. 607–613, 2003.

[26] H. Dia, "An object-oriented neural network approach to short-term traffic forecasting," *Eur. J. Oper. Res.*, vol. 131, no. 2, pp. 253–261, Jun. 2001.

[27] L. Do, N. Taherifar, and H. L. Vu, "Survey of neural network-based models for short-term traffic state prediction," *Wiley Interdiscipl. Rev., Data Mining Knowl. Discovery*, vol. 9, Sep. 2018, Art. no. e1285.

[28] C. Dong, Z. Xiong, C. Shao, and H. Zhang, "A spatial–temporal-based state space approach for freeway network traffic flow modelling and prediction," *Transportmetrica A, Transp. Sci.*, vol. 11, no. 7, pp. 547–560, 2015.

[29] S. Dunne and B. Ghosh, "Regime-based short-term multivariate traffic condition forecasting algorithm," *J. Transp. Eng.*, vol. 138, no. 4, pp. 455–466, 2012.

[30] A. Ermagun and D. Levinson, "Spatiotemporal traffic forecasting: Review and proposed directions," *Transp. Rev.*, vol. 38, no. 6, pp. 786–814, Nov. 2018.

[31] N.-E. El Faouzi, H. Leung, and A. Kurian, "Data fusion in intelligent transportation systems: Progress and challenges—A survey," *Inf. Fusion*, vol. 12, no. 1, pp. 4–10, 2011.

[32] G. Fusco, C. Colombaroni, L. Comelli, and N. Isaenko, "Short-term traffic predictions on large urban traffic networks: Applications of network-based machine learning models and dynamic traffic assignment models," in *Proc. Int. Conf. Models Technol. Intell. Transp. Syst. (MT-ITS)*, Jun. 2015, pp. 93–101.

[33] Y. Gao, S. Sun, and D. Shi, "Network-scale traffic modeling and forecasting with graphical lasso," in *Advances in Neural Networks*, D. Liu, H. Zhang, M. Polycarpou, C. Alippi, and H. He, Eds. Berlin, Germany: Springer, 2011, pp. 151–158.

[34] S. Garcia, J. Luengo, and F. Herrera, *Data Preprocessing in Data Mining*. Cham, Switzerland: Springer, 2014.

[35] J. Grengs, X. Wang, and L. Kostyniuk, "Using GPS data to understand driving behavior," *J. Urban Technol.*, vol. 15, no. 2, pp. 33–53, 2008.

[36] B. Gültekin, M. Sari, and O. Borat, "A neural network based traffic-flow prediction model," *Math. Comput. Appl.*, vol. 15, no. 2, pp. 269–278, 2010.

[37] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, 3rd ed. Waltham, MA, USA: Morgan Kaufmann, 2012.

[38] T. Hastie, R. Tibshirani, and J. H. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (Springer Series in Statistics), 2nd ed. New York, NY, USA: Springer-Verlag, 2009.

[39] S. He, C. Hu, G.-J. Song, K.-Q. Xie, and Y.-Z. Sun, "Real-time short-term traffic flow forecasting based on process neural network," in *Advances in Neural Networks*, F. Sun, J. Zhang, Y. Tan, J. Cao, and W. Yu, Eds. Berlin, Germany: Springer, 2008, pp. 560–569.

[40] J. C. Herrera, D. B. Work, R. Herring, X. Ban, Q. Jacobsond, and A. M. Bayen, "Evaluation of traffic data obtained via GPS-enabled mobile phones: The mobile century field experiment," *Transp. Res. C, Emerg. Technol.*, vol. 18, no. 4, pp. 568–583, Aug. 2010.

[41] R. Herring, A. Hofleitner, P. Abbeel, and A. Bayen, "Estimating arterial traffic conditions using sparse probe data," in *Proc. Int. IEEE Conf. Intell. Transp. Syst.*, Sep. 2010, pp. 929–936.

[42] Y. Hou, P. Edara, and C. Sun, "Traffic flow forecasting for urban work zones," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 4, pp. 1761–1770, Aug. 2015.

[43] S. Howell, "Meta-analysis of machine learning approaches to short-term urban traffic prediction," in *Proc. Scottish Transp. Appl. Res. Conf. (STAR)*, 2018, pp. 1–15.

[44] Y. Hu, M. Li, H. Liu, X. Guo, X. Wang, and T. Li, "City traffic forecasting using taxi GPS data: A coarse-grained cellular automata model," *Comput. Res. Repository*, pp. 1–30, Dec. 2016.

[45] J. Huang, C. Liu, and J. Qie, "Developing map matching algorithm for transportation data center," in *Proc. Int. Conf. P2P, Parallel, Grid, Cloud Internet Comput.*, Nov. 2014, pp. 167–170.

[46] M.-L. Huang, "Intersection traffic flow forecasting based on ν-GSVR with a new hybrid evolutionary algorithm," *Neurocomputing*, vol. 147, no. 1, pp. 343–349, 2015.

[47] S. Huang and A. W. Sadek, "A novel forecasting approach inspired by human memory: The example of short-term traffic volume forecasting, " *Transp. Res. C, Emerg. Technol.*, vol. 17, no. 5, pp. 510–525, Oct. 2009.

[48] S. Ishak and C. Alecsandru, "Optimizing traffic prediction performance of neural networks under various topological, input, and traffic condition settings," *J. Transp. Eng.*, vol. 130, no. 4, pp. 452–465, 2004.

[49] X. Jiang and H. Adeli, "Wavelet packet-autocorrelation function method for traffic flow pattern analysis," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 19, no. 5, pp. 324–337, 2004.

[50] W. Jin, Y. Lin, Z. Wu, and H. Wan, "Spatio-temporal recurrent convolutional networks for citywide short-term crowd flows prediction," in *Proc. Int. Conf. Comput. Data Anal.*, 2018, pp. 28–35.

[51] H.-J. Yang and X. Hu, "Wavelet neural network with improved genetic algorithm for traffic flow time series prediction," *Optik-Int. J. Light Electron Opt.*, vol. 127, no. 19, pp. 8103–8110, 2016.

[52] M. Karimpour, A. Karimpour, K. Kompany, and A. Karimpour, *Online Traffic Prediction Using Time Series: A Case study*. Cham, Switzerland: Springer, 2017, pp. 147–156.

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

[54] D. Kehagias, A. Salamanis, and D. Tzovaras, "Speed pattern recognition technique for short-term traffic forecasting based on traffic dynamics," *IET Intell. Transp. Syst.*, vol. 9, no. 6, pp. 646–653, 2015.

[55] B. S. Kerner, "The physics of traffic," *Phys. World*, vol. 12, no. 8, pp. 25–30, 1999.

[56] I. Laña, J. Del Ser, M. Vélez, and I. Oregi, "Joint feature selection and parameter tuning for short-term traffic flow forecasting based on heuristically optimized multi-layer neural networks," in *Harmony Search Algorithm*, J. Del Ser, Ed. Singapore: Springer, 2017, pp. 91–100.

[57] I. Lana, J. Del Ser, M. Velez, and E. I. Vlahogianni, "Road traffic forecasting: Recent advances and new challenges," *IEEE Intell. Transp. Syst. Mag.*, vol. 10, no. 2, pp. 93–109, Apr. 2018.

[58] I. Laña, I. Olabarrieta, M. Velez, and J. Del Ser, "On the imputation of missing data for road traffic forecasting: New insights and novel techniques," *Transp. Res. C, Emerg. Technol.*, vol. 90, pp. 18–33, May 2018.

[59] G. Leduc, "Road traffic data: Collection methods and applications," Eur. Comission, Joint Res. Centre, Inst. Prospective Technol. Studies, Seville, Spain, Tech. Rep. JRC 47967, 2008, pp. 1–51.

[60] L. Li, X. Su, Y. Zhang, Y. Lin, and Z. Li, "Trend modeling for traffic time series analysis: An integrated study," *Intell. Transp. Syst.*, vol. 16, no. 6, pp. 3430–3439, Dec. 2015.

[61] L. Li, X. Su, Y. Wang, Y. Lin, Z. Li, and Y. Li, "Robust causal dependence mining in big data network and its application to traffic flow predictions," *Transp. Res. C, Emerg. Technol.*, vol. 58, pp. 292–307, Sep. 2015.

[62] X. Li and W. Gao, "Prediction of traffic flow combination model based on data mining," *Int. J. Database Theory Appl.*, vol. 8, no. 6, pp. 303–312, Dec. 2015.

[63] M. Lippi, M. Bertini, and P. Frasconi, "Short-term traffic flow forecasting: An experimental comparison of time-series analysis and supervised learning," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 2, pp. 871–882, Jun. 2013.

[64] K. Liu, S. Gao, P. Qiu, X. Liu, B. Yan, and F. Lu, "Road2Vec: Measuring traffic interactions in urban road system from massive travel routes," *Int. J. Geo-Inf.*, vol. 6, no. 11, p. 321, 2017.

[65] Q. Liu, B. Wang, and Y. Zhu, "Short-term traffic speed forecasting based on attention convolutional neural network for arterials," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 33, no. 11, pp. 999–1016, 2018.

[66] S.-Y. Liu, D.-W. Li, Y.-G. Xi, and Q.-F. Tang, "A short-term traffic flow forecasting method and its applications," *J. Shanghai Jiaotong Univ.*, vol. 20, no. 2, pp. 156–163, 2015.

[67] Z. Liu, J. Guo, J. Cao, Y. Wei, and W. Huang, "A hybrid short-term traffic flow forecasting method based on neural networks combined with K-nearest neighbor," *J. Traffic Transp. Technol.*, vol. 30, no. 4, pp. 445–456, 2018.

[68] J. Longfoot, "An automatic network travel time system—ANTTS," in *Proc. Vehicle Navigat. Inf. Syst. Conf.*, vol. 2, Oct. 1991, pp. 1053–1061.

[69] J. Lopes, J. Bento, E. Huang, C. Antoniou, and M. Ben-Akiva, "Traffic and mobility data collection for real-time applications," in *Proc. IEEE Conf. Intell. Transp. Syst.*, Sep. 2010, pp. 216–223.

[70] P. Lopez-Garcia, E. Onieva, E. Osaba, A. D. Masegosa, and A. Perallos, "A hybrid method for short-term traffic congestion forecasting using genetic algorithms and cross entropy," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 2, pp. 557–569, Feb. 2016.

[71] H.-P. Lu, Z.-Y. Sun, W.-C. Qu, and L. Wang, "Real-time corrected traffic correlation model for traffic flow forecasting," *Math. Problems Eng.*, vol. 2015, Feb. 2015, Art. no. 348036.

[72] W. Ma and R. Wang, "Traffic flow forecasting research based on Bayesian normalized Elman neural network," in *Proc. IEEE Signal Process. Signal Process. Edu. Workshop (SP/SPE)*, Aug. 2015, pp. 426–430.

[73] X. Ma, Z. Tao, Y. Wang, H. Yu, and Y. Wang, "Long short-term memory neural network for traffic speed prediction using remote microwave sensor data," *Transp. Res. C, Emerg. Technol.*, vol. 54, pp. 187–197, May 2015.

[74] X. Ma, H. Yu, Y. Wang, and Y. Wang, "Large-scale transportation network congestion evolution prediction using deep learning theory," *PLoS ONE*, vol. 10, no. 3, 2015, Art. no. e0119044.

[75] Z. Ma, G. Luo, and D. Huang, "Short term traffic flow prediction based on on-line sequential extreme learning machine," in *Proc. Int. Conf. Adv. Comput. Intell. (ICACI)*, Feb. 2016, pp. 143–149.

[76] M. Meng, C.-F. Shao, Y.-D. Wong, B.-B. Wang, and H.-X. Li, "A two-stage short-term traffic flow prediction method based on AVL and AKNN techniques," *J. Central South Univ.*, vol. 22, no. 2, pp. 779–786, 2015.

[77] F. Moretti, S. Pizzuti, S. Panzieri, and M. Annunziato, "Urban traffic flow forecasting through statistical and neural network bagging ensemble hybrid modeling," *Neurocomputing*, vol. 167, pp. 3–7, Nov. 2015.

[78] U. Mori, A. Mendiburu, M. Álvarez, and J. A. Lozano, "A review of travel time estimation and forecasting for advanced traveller information systems," *Transportmetrica A, Transp. Sci.*, vol. 11, no. 2, pp. 119–157, 2015.

[79] A. M. Nagy and V. Simon, "Survey on traffic prediction in smart cities," *Pervas. Mobile Comput.*, vol. 50, pp. 148–163, Oct. 2018.

[80] X. Niu, Y. Zhu, Q. Cao, X. Zhang, W. Xie, and K. Zheng, "An online-traffic-prediction based route finding mechanism for smart city," *Int. J. Distrib. Sensor Netw.*, vol. 11, no. 8, pp. 1–16, 2015.

[81] S. Oh, Y.-J. Byon, K. Jang, and H. Yeo, "Short-term travel-time prediction on highway: A review of the data-driven approach," *Transp. Rev.*, vol. 35, no. 1, pp. 4–32, 2015.

[82] H. Pan, J. Liu, S. Zhou, and Z. Niu, "A block regression model for short-term mobile traffic forecasting," in *Proc. Int. Conf. Commun. China (ICCC)*, Nov. 2015, pp. 1–5.

[83] B. Park, "Hybrid neuro-fuzzy application in short-term freeway traffic volume forecasting, " *Transp. Res. Rec.*, vol. 1802, no. 1, pp. 190–196, 2002.

[84] D. Pavlyuk, "Short-term traffic forecasting using multivariate autoregressive models," *Procedia Eng.*, vol. 178, pp. 57–66, Oct. 2017.

[85] D. Pavlyuk, "Feature selection and extraction in spatiotemporal traffic forecasting: A systematic literature review," *Eur. Transp. Res. Rev.*, vol. 11, no. 1, p. 6, 2019.

[86] A. Perallos, U. Hernandez-Jayo, E. Onieva, and I. J. García-Zuazola, *Intelligent Transport Systems: Technologies and Applications*, 1st ed. Hoboken, NJ, USA: Wiley, 2015.

[87] S. Rahimipour, R. Moeinfar, and S. M. Hashemi, "Traffic prediction using a self-adjusted evolutionary neural network," *J. Modern Transp.*, pp. 1–11, Dec. 2018.

[88] A. Raza and M. Zhong, "Hybrid lane-based short-term urban traffic speed forecasting: A genetic approach," in *Proc. Int. Conf. Transp. Inf. Safety (ICTIS)*, Aug. 2017, pp. 271–279.

[89] A. Raza and M. Zhong, "Lane-based short-term urban traffic forecasting with ga designed ANN and LWR models," *Transp. Res. Procedia*, vol. 25, pp. 1430–1443, 2017.

[90] J. Rupnik, J. Davies, B. Fortuna, A. Duke, and S. S. Clarke, "Travel time prediction on highways," in *Proc. Int. Conf. Comput. Inf. Technol.*, Oct. 2015, pp. 1435–1442.

[91] D. Satrinia and G. A. P. Saptawati, "Traffic speed prediction from GPS data of taxi trip using support vector regression," in *Proc. Int. Conf. Data Softw. Eng. (ICoDSE)*, Nov. 2017, pp. 1–6.

[92] V. Sessions and M. G. Valtorta, "The effects of data quality on machine learning algorithms," in *Proc. Int. Conf. Inf. Qual.*, 2006, pp. 1–14.

[93] M. Shahgholian and D. Gharavian. (2018). "Advanced traffic management systems: An overview and a development strategy." [Online]. Available: https://arxiv.org/abs/1810.02530

[94] B. Sharma, S. Kumar, P. Tiwari, P. Yadav, and M. I. Nezhurina, "ANN based short-term traffic flow forecasting in undivided two lane highway," *J. Big Data*, vol. 5, no. 1, p. 48, Dec. 2018.

[95] Q. Shi and M. Abdel-Aty, "Big data applications in real-time traffic operation and safety monitoring and improvement on urban expressways," *Transp. Res. C, Emerg. Technol.*, vol. 58, pp. 380–394, Sep. 2015.

[96] I. B. M. Skycomp, *Major High-Way Performance Ratings and Bottleneck Inventory*. Baltimore, MD, USA: Maryland State Highway Administration, the Baltimore Metropolitan Council and Maryland Transportation Authority, 2009.

[97] C. Song, H. Lee, C. Kang, W. Lee, Y. B. Kim, and S. W. Cha, "Traffic speed prediction under weekday using convolutional neural networks concepts," in *Proc. IEEE Intell. Vehicles Symp.*, Jun. 2017, pp. 1293–1298.

[98] F. Soriguera and F. Robusté, "Estimation of traffic stream space mean speed from time aggregations of double loop detector data," *Transp. Res. C, Emerg. Technol.*, vol. 19, no. 1, pp. 115–129, Feb. 2011.

[99] F. Soriguera and F. Robusté, "Requiem for freeway travel time estimation methods based on blind speed interpolations between point measurements," *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 1, pp. 291–297, Mar. 2011.

[100] D. Srinivasan, C. W. Chan, and P. G. Balaji, "Computational intelligence-based congestion prediction for a dynamic urban street network," *Neurocomputing*, vol. 72, nos. 10–12, pp. 2710–2716, Jun. 2009.

[101] F. Su, H. Dong, L. Jia, Y. Qin, and Z. Tian, "Long-term forecasting oriented to urban expressway traffic situation," *Adv. Mech. Eng.*, vol. 8, no. 1, pp. 1–16, 2016.

[102] F. Su, H. Dong, L. Jia, Y. Qin, and Z. Tian, "Short-term traffic forecasting using self-adjusting k-nearest neighbours," *IET Intell. Transp. Syst.*, vol. 12, no. 1, pp. 41–48, Feb. 2018.

[103] B. Sun, W. Cheng, P. Goswami, and G. Bai, "Flow-aware WPT k-nearest neighbours regression for short-term traffic prediction," in *Proc. IEEE Symp. Comput. Commun. (ISCC)*, Jul. 2017, pp. 48–53.

[104] S. Sun, C. Zhang, G. Yu, N. Lu, and F. Xiao, "Bayesian network methods for traffic flow forecasting with incomplete data," in *Proc. Eur. Conf. Mach. Learn.*, J. F. Boulicaut, F. Esposito, F. Giannotti, and D. Pedreschi, Eds. Berlin, Germany: Springer, 2004, pp. 419–428.

[105] S. Taguchi, S. Koide, and T. Yoshimura, "Online map matching with route prediction," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 1, pp. 338–347, Jan. 2019.

[106] T.-I. Theodorou, A. Salamanis, D. D. Kehagias, D. Tzovaras, and C. Tjortjis, "Short-term traffic prediction under both typical and atypical traffic conditions using a pattern transition model," in *Proc. Int. Conf. Vehicle Technol. Intell. Transp. Syst.*, 2017, pp. 1–11.

[107] I. Triguero, G. P. Figueredo, M. Mesgarpour, J. M. Garibaldi, and R. I. John, "Vehicle incident hot spots identification: An approach for big data," in *Proc. IEEE Trustcom/BigDataSE/ICESS*, Aug. 2017, pp. 901–908.

[108] S. Turner, W. Eisele, R. J. Benz, and D. J Holdener, *Travel Time Data Collection Handbook*. Washington, DC, USA: Federal Highway Administration, 1998.

[109] C. P. van Hinsbergen, J. W. van Lint, and F. M. Sanders, "Short term traffic prediction models," in *Proc. World Congr. Intell. Transp. Syst.*, 2007, pp. 1–20.

[110] H. van Lint and C. van Hinsbergen, *Transportation Research Circular*. Washington, DC, USA: Transportation Research Board-National Research Council, 2012.

[111] J. van Lint and N. van der Zijpp, "Improving a travel-time estimation algorithm by using dual loop detectors," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 1855, no. 1, pp. 41–48, 2003.

[112] J. Van Lint, S. P. Hoogendoorn, and H. J. van Zuylen, "Accurate freeway travel time prediction with state-space neural networks under missing data," *Transp. Res. C, Emerg. Technol.*, vol. 13, nos. 5–6, pp. 347–369, Oct./Dec. 2005.

[113] E. Vlahogianni and M. Karlaftis, "Temporal aggregation in traffic data: Implications for statistical characteristics and model choice," *Transp. Lett.*, vol. 3, pp. 37–49, Sep. 2013.

[114] E. I. Vlahogianni, "Enhancing predictions in signalized arterials with information on short-term traffic flow dynamics," *J. Intell. Transp. Syst.*, vol. 13, no. 2, pp. 73–84, 2009.

[115] E. I. Vlahogianni, J. C. Golias, and M. G. Karlaftis, "Short-term traffic forecasting: Overview of objectives and methods," *Transp. Rev.*, vol. 24, no. 5, pp. 533–557, 2004.

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

[117] J. Wang, I. Tsapakis, and C. Zhong, "A space–time delay neural network model for travel time prediction," *Eng. Appl. Artif. Intell.*, vol. 52, pp. 145–160, Jun. 2016.

[118] J. Wang, P. Shang, and X. Zhao, "A new traffic speed forecasting method based on bi-pattern recognition," *Fluctuation Noise Lett.*, vol. 10, no. 1, pp. 59–75, 2011.

[119] A. Wibisono, W. Jatmiko, H. A. Wisesa, B. Hardjono, and P. Mursanto, "Traffic big data prediction and visualization using fast incremental model trees-drift detection (FIMT-DD)," *Knowl.-Based Syst.*, vol. 93, pp. 33–46, Feb. 2016.

[120] D. Xia, B. Wang, H. Li, Y. Li, and Z. Zhang, "A distributed spatial-temporal weighted model on MapReduce for short-term traffic flow forecasting," *Neurocomputing*, vol. 179, pp. 246–263, Feb. 2016.

[121] J. Xia, W. Huang, and J. Guo, "A clustering approach to online freeway traffic state identification using ITS data," *KSCE J. Civil Eng.*, vol. 16, no. 3, pp. 426–432, 2012.

[122] Y. Chun-Xia, F. Rui, and F. Yi-Qin, "Prediction of short-term traffic flow based on similarity," *J. Highway Transp. Res. Develop.*, vol. 10, no. 1, pp. 92–97, 2016.

[123] B. Zhang, K. Xing, X. Cheng, L. Huang, and R. Bie, "Traffic clustering and online traffic prediction in vehicle networks: A social influence perspective," in *Proc. IEEE Int. Conf. Comput. Commun. (IEEE INFOCOM)*, Mar. 2012, pp. 495–503.

[124] J. Xu, D. Deng, U. Demiryurek, C. Shahabi, and M. van der Schaar, "Mining the situation: Spatiotemporal traffic prediction with big data," *IEEE J. Sel. Topics Signal Process.*, vol. 9, no. 4, pp. 702–715, Jun. 2015.

[125] Y. Xu, H. Chen, Q.-J. Kong, X. Zhai, and Y. Liu, "Urban traffic flow prediction: A spatio-temporal variable selection-based approach," *J. Adv. Transp.*, vol. 50, no. 4, pp. 489–506, 2016.

[126] H.-F. Yang, T. S. Dillon, E. Chang, and Y.-P. P. Chen, "Optimized configuration of exponential smoothing and extreme learning machine for traffic flow forecasting," *IEEE Trans. Ind. Informat.*, vol. 15, no. 1, pp. 23–34, Jan. 2019.

[127] H. Yang, Y. Zou, Z. Wang, and B. Wu, "A hybrid method for short-term freeway travel time prediction based on wavelet neural network and Markov chain," *Can. J. Civil Eng.*, vol. 45, no. 2, pp. 77–86, 2018.

[128] H. Yang, T. S. Dillon, and Y.-P. P. Chen, "Evaluation of recent computational approaches in short-term traffic forecasting," in *Artificial Intelligence in Theory and Practice IV*, T. Dillon, Ed. Cham, Switzerland: Springer, 2015, pp. 108–116.

[129] B. Z. Yao *et al.*, "Short-term traffic speed prediction for an urban corridor," *Comput.-Aided Civil Inf.*, vol. 32, no. 2, pp. 154–169, 2017.

[130] Q. Ye, W. Y. Szeto, and S. C. Wong, "Short-term traffic speed forecasting based on data recorded at irregular intervals," *IEEE Trans. Intell. Transp. Syst.*, vol. 13, no. 4, pp. 1727–1737, Dec. 2012.

[131] B. Yu, H. Yin, and Z. Zhu, "Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting," *CoRR*, pp. 1–7, Jul. 2017.

[132] H. Yu, Z. Wu, S. Wang, Y. Wang, and X. Ma, "Spatiotemporal recurrent convolutional networks for traffic prediction in transportation networks," *Sensors*, vol. 17, no. 7, p. 1501, 2017.

[133] Y. Wu, H. Tan, J. Peter, and B. Shen, "Short-term traffic flow prediction based on multilinear analysis and k-nearest neighbor regression," in *Proc. CICTP*, 2015, pp. 556–569.

[134] N. Zhang, Y. Zhang, and H. Lu, "Seasonal autoregressive integrated moving average and support vector machine models: Prediction of short-term traffic flow on freeways," *Transp. Res. Rec.*, vol. 2215, no. 1, pp. 85–92, 2011.

[135] Y. Zhang and Y. Liu, "Traffic forecasting using least squares support vector machines," *Transportmetrica*, vol. 5, no. 3, pp. 193–213, 2009.

[136] Y. Zhang and M.-L. Wang, "Peak traffic forecasting using nonparametric approaches," *J. Shanghai Jiaotong Univ.*, vol. 17, no. 1, pp. 76–81, 2012.

[137] Y. Zhang and A. Haghani, "A gradient boosting method to improve travel time prediction," *Transp. Res. C, Emerg. Technol.*, vol. 58, pp. 308–324, Sep. 2015.

[138] Y. Zhang and Y. Zhang, "A comparative study of three multivariate short-term freeway traffic flow forecasting methods with missing data," *J. Intell. Transp. Syst.*, vol. 20, no. 3, pp. 205–218, 2016.

[139] J. Zhao and S. Sun, "High-order Gaussian process dynamical models for traffic flow prediction," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 7, pp. 2014–2019, Jul. 2016.

[140] J.-T. Zhong and S. Ling, "Key factors of K-nearest neighbours nonparametric regression in short-time traffic flow forecasting," in *Proc. Int. Conf. Ind. Eng. Eng. Manage.*, E. Qi, J. Shen, and R. Dou, Eds. Paris, France: Atlantis Press, 2015, pp. 9–12.

[141] M. Zhong, S. Sharma, and P. Lingras, "Short-term traffic prediction on different types of roads with genetically designed regression and time delay neural network models," *J. Comput. Civil Eng.*, vol. 19, no. 1, pp. 94–103, 2005.

JUAN S. ANGARITA-ZAPATA received the B.Sc. and M.Sc. degrees in systems engineering and computer science from the Industrial University of Santander, Bucaramanga, Colombia, in 2014 and 2016, respectively. He is currently pursuing the Ph.D. degree with the Faculty of Engineering, University of Deusto, Bilbao, Spain, where he is a Marie Sklodowska-Curie Fellow. He has published various papers at international conferences related to automatic algorithm selection techniques, complex and dynamical systems, and transportation. His research interests include the intersection of machine learning, data Science, and intelligent transportation systems.

ANTONIO D. MASEGOSA received the M.Sc. degree in computer engineering and the Ph.D. degree in computer sciences from the University of Granada, Spain, in 2005 and 2010, respectively. From 2010 to 2014, he was a Postdoctoral Researcher with the Research Center for ICT, University of Granada. In 2014, he received an IKERBASQUE Research Fellowship to work with the Mobility Unit, Deusto Institute of Technology, Bilbao, Spain. He has published four books, 17 JCR papers, and more than 20 papers in both international and national conferences. He has participated in several research projects at regional, national, and European level. His main research interests include artificial intelligence, intelligent systems, soft computing, hybrid metaheuristics, machine learning, deep learning, intelligent transportation systems, logistics networks, travel demand estimation, and traffic forecasting.

ISAAC TRIGUERO received the M.Sc. and Ph.D. degrees in computer science from the University of Granada, Granada, Spain, in 2009 and 2014, respectively. Since 2016, he has been an Assistant Professor of data science with the Automated Scheduling, Optimization, and Planning (ASAP) Research Group, School of Computer Science, University of Nottingham. His work is mostly concerned with the research of novel methodologies for big data analytics. He has published more than 60 international publications in the fields of big data, machine learning, and optimization (H-index=21 and more than 1700 citations on Google Scholar). He has acted as a Program Co-Chair of the IEEE Conference on Smart Data, in 2016, the IEEE Conference on Big Data Science and Engineering, in 2017, and the IEEE International Congress on Big Data, in 2018. He is currently leading a KTP project funded by the Innovative U.K. and the energy provider E.ON that investigates Smart Metering data. He is a Section Editor-in-Chief of the *Machine Learning* and *Knowledge Extraction* journal, and an Associate Editor of the *Big Data and Cognitive Computing* journal, and the IEEE ACCESS journal.