

Received August 31, 2021, accepted September 8, 2021, date of publication September 24, 2021, date of current version October 5, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3115709

On the Impact of Floating Car Data and Data Fusion on the Prediction of the Traffic Density, Flow and Speed Using an Error Recurrent Convolutional Neural Network

JESÚS MENA-OREJA¹ AND JAVIER GOZALVEZ², (Senior Member, IEEE)

¹Uwicare Laboratory, Universidad Miguel Hernández de Elche (UMH), 03202 Elche, Alicante, Spain

Corresponding author: Jesús Mena-Oreja (jesus.mena@goumh.umh.es)

This work was supported in part by the Local Government of Valencia and co-supported by the European Social Fund under Grant ACIF/2017/160, in part by the Spanish Dirección General de Tráfico under Grant SPIP2017-02204, and in part by the Spanish Ministry of Science and Innovation (MCI), the State Research Agency (AEI), and the European Regional Development Fund (FEDER), under Project TEC2017-88612-R and Project PID2020-115576RB-I00.

ABSTRACT Traffic prediction helps mitigate the impact of traffic congestion. The accuracy of traffic predictions depends on the availability of the data used for the prediction as well as the prediction model. Data from fixed traffic detectors is only available at certain locations. On the other hand, connected vehicles can provide Floating Car Data (FCD) at any location and time. However, FCD may not be available at all vehicles, and this can impact predictions since the FCD may not reflect the state of all traffic. This impact is larger when predicting traffic density or flow, and existing studies generally use FCD to predict traffic speed or travel time only. This study proposes a traffic prediction model that can accurately predict the three fundamental traffic variables (traffic density, flow, and speed) using FCD and an error recurrent convolutional neural network that takes as input the three variables. These are estimated using FCD and data from induction loops. These estimates depend on the penetration rate of FCD, so we propose a method to locally and dynamically estimate this penetration rate. This method improves the estimation of the traffic variables, and hence their prediction. The proposed model is used to analyze the impact of the FCD penetration rate on the prediction of the traffic variables. We show how our proposal reduces the amount of FCD needed to improve the prediction obtained with data from traffic detectors. We show that our proposal only requires FCD from 4% of the vehicles to improve the prediction accuracy achieved with traffic detectors. Augmenting this percentage increases the accuracy of our model for the three traffic variables. However, we also show that our prediction model reduces the FCD sample size (or FCD penetration rate) needed to achieve prediction accuracy levels close to that obtained if all vehicles provided FCD.

INDEX TERMS Traffic prediction, floating car data, connected vehicle, induction loops, data fusion, neural networks, deep learning, convolutional neural networks, error recurrent.

I. INTRODUCTION

Traffic prediction can help anticipate and mitigate traffic congestions, and hence reduce their negative economic, environmental and comfort impact. An accurate traffic prediction requires reliable road traffic data, and this data can be obtained from different sources. This includes traditional fixed traffic detectors such as induction loops or traffic cameras, and also floating car data (FCD) from

connected vehicles. Fixed traffic detectors have the advantage of sensing all vehicles driving across a road section, and hence provide information about the full traffic. However, fixed traffic detectors can only provide traffic data of the location where they are deployed. This is not the case for FCD as connectivity transforms vehicles into moving sensors that can provide traffic data at any location and time. In addition, FCD devices are deployed and maintained directly by the drivers and not the traffic managers or authorities. However, traffic managers and road authorities need to acquire the FCD from third party providers like GPS providers [1], [2] or insurance

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

companies [3], and the price can vary based on the amount of FCD while it is yet unclear how much FCD they really need for their traffic estimations and predictions. In addition, FCD devices (e.g. GPS, smartphones, V2X onboard units) may not be installed on all the vehicles or may not always actively transmit traffic data. In this case, and unlike fixed traffic detectors, FCD may not provide information of the full traffic. This might slightly reduce the accuracy in the estimation or prediction of traffic variables such as the speed or travel time that do not depend directly on the number of vehicles in the scenario. Averaging the speed information of the FCD has the uncertainty of the speed of all the non-connected vehicles. However, this average is still a good estimate of the mean speed of the traffic since the speed of the rest of the vehicles is usually similar. The impact of a low penetration rate of connectivity or FCD devices can be significantly higher when considering the estimation or prediction of traffic variables that depend on the number of vehicles driving a road. This is for example the case of the traffic density or the traffic flow that provide information about the number of vehicles per unit length and per unit time respectively. The number of vehicles in a road can only be estimated using FCD if we have an estimate of the rate of connected vehicles in the traffic. However, it is important to note that such rate can vary in space and time and hence using a fixed rate across a road can ultimately negatively impact the traffic estimations and predictions. These limitations explain why most traffic prediction studies using FCD focus on predicting variables such as the speed or the travel time. Some studies propose fusing different data sources to improve the prediction accuracy and mitigate the disadvantages of each data source when processed individually [4]. However, it is yet unclear how much FCD is needed to match with FCD (alone or in combination with other data sources) the accuracy of the traffic predictions achieved using only data from existing fixed traffic detectors. To this aim, it is first important to understand the impact of the FCD penetration rate on the accuracy of traffic predictions.

This study analyzes for the first time the impact of the FCD penetration rate on the accuracy of the short-term prediction or forecast of the three fundamental traffic variables, i.e. the traffic density, the traffic flow and the space mean speed, as well as the impact of fusing FCD and data from traffic detectors (in our case, induction loops) on the prediction accuracy for the three traffic variables. To this aim, this study progresses the state of the art by presenting a short-term traffic prediction or forecast model (we will refer to as traffic prediction model in the rest of the paper) that can predict with high accuracy the three fundamental traffic variables using FCD. Predicting the three fundamental traffic variables is important, since the traffic state can only be determined when the three traffic variables are known. To this aim, we propose a model based on an error recurrent convolutional neural network (eRCNN) that takes as input estimates of the three traffic variables represented in the form of traffic images. The variables can be estimated using different data sources.

To estimate the traffic variables using FCD, we propose a method to estimate the FCD penetration rate and hence the number of vehicles in a road. The method estimates the rate at each road section and time instant so it can better adapt to the spatiotemporal evolution of the traffic. The proposed method improves the estimation of the three traffic variables, and hence their short-term prediction or forecast (prediction from now on), since the estimates of the variables are used as input to our traffic prediction model. The proposed traffic prediction model is used to analyze the impact of the FCD penetration rate and of fusing FCD and data from traffic detectors on the prediction accuracy. Our study shows that it is possible to accurately predict the three traffic variables using FCD, in particular if this data is combined with data from traffic detectors. This combination can significantly reduce the amount of FCD needed for accurate traffic predictions when utilized with an error feedback deep learning-based traffic prediction model. In addition, the combination of FCD and data from induction loops significantly improves the prediction accuracy compared to when using only data from traffic detectors, and reduces the amount of FCD needed to achieve such improvements. The conducted analysis provides valuable information to understand how much FCD is really needed to achieve accurate traffic predictions.

The remainder of this paper is organized as follows. Section II reviews previous studies that focused on traffic prediction using FCD. Section III describes our proposed traffic prediction model, including the proposed method to locally and dynamically estimate the FCD penetration rate, the process to estimate the three fundamental traffic variables using FCD and data from induction loops, as well as the error feedback eRCNN prediction module. Section IV describes the evaluation scenario and the generation of datasets for our prediction model, and Section V evaluates its performance and analyzes in detail the impact of the FCD penetration rate and of fusing FCD and traffic detectors data on the prediction accuracy. Finally, Section VI summarizes the main contributions and findings of this study.

II. RELATED WORK

Most short-term traffic prediction proposals to date have been designed to utilize data from fixed traffic detectors. For example, [5] proposes a traffic prediction technique based on Long Short-Term Memory (LSTM) recurrent neural networks for predicting the traffic speed. A similar technique is proposed in [6] for predicting the traffic flow. The proposal in [7] combines the k-nearest neighbors algorithm and neural networks to predict the evolution of the traffic density. The study in [8] combines different statistical methods and machine learning methods with a trigonometric regression function to predict the evolution of the traffic speed. The previous study is extended in [9] by combining different statistical and machine learning methods with different periodic functions. Traffic predictions using data from fixed traffic detectors are restricted to the location of the detectors. However, it is possible to predict any traffic variable using

data from fixed traffic detectors since they provide information about the full traffic at the location of the detectors. This is not the case of the FCD since the penetration rate of FCD devices is still not significant and the availability of FCD is usually limited and, in many cases, restricted to specific fleets of vehicles, e.g. taxis like in [10]. These limitations result in that most studies using FCD for traffic prediction (e.g. [1]–[3], [10]–[19]) generally focus on predicting the traffic speed or travel times since these two traffic variables can be computed without knowing the penetration rate of FCD devices and hence what fraction of the full traffic the FCD data represents.¹ This is for example the case of the study presented in [11] where authors predict the travel time using the k-nearest neighbors algorithm and GPS data from a private fleet of vehicles in the city of Munich. The proposal divides Munich's road network in different road sections, and aggregates the GPS data at each road section by averaging the speed of all the GPS samples located at the section. The proposal computes the travel time loss for each road section as the product of the length of the road section and the difference between its average speed and its maximum speed. The travel time loss is the input for the k-nearest neighbor algorithm, and the travel time prediction is the average of the future travel time associated to the k nearest travel time losses for each road section. [20] also uses the k-nearest neighbors' algorithm to predict travel times using the speed of GPS data of intercity buses. However, the proposal uses a more elaborated approach to estimate the input of the algorithm. The authors fit a linear regression model that returns the speed of a road section (between fixed traffic detectors) as a function of the speed of all the GPS samples of the road section. The model is fitted so that the speed varies linearly between the boundaries of the road section. The model is used to compute the travel time of each road section, and this travel time is then used as the input to the k-nearest neighbors algorithm. The algorithm returns then the prediction of the travel time for each road section. Other proposals also divide the network into road sections and average traffic variables over each section but differ on the prediction techniques. For example, [13] uses a Kalman filter to predict the travel time of each road section. In [1], authors aggregate the traffic speed and travel time per road section, and they propose different grey systems that model the time series of the two traffic variables per road section and that are used to predict their future values. The authors of [10] use GPS data from vehicles of the city of Berlin and GPS data from taxis of the city of Thessaloniki to obtain the instantaneous speed of a sample of vehicles in the traffic. Authors average these speeds per road section and use a STARIMA model to predict the travel time of each road section. The proposals in [2] and [14] use the gradient boosting regression tree to predict the travel time. In [2], authors train a gradient boosting regression tree to

¹In this case, it is only necessary to average the values of the speed or the travel time of each sample in the FCD in order to have an estimate of these traffic variables.

predict the travel time of each road section using as input the travel time of each road section in previous time steps. The same approach is used in [14] to predict the speed per road section.

More recently, traffic prediction proposals have shifted towards the use of machine learning and deep learning-based neural networks that are the state of the art. Many of these studies have been reviewed in [15]. For example, the study in [16] compares the accuracy of a multilayer perceptron, a non-linear autoregressive neural network and a Bayesian network for predicting the traffic speed using FCD. The input of all these models is the average speed of different road sections, the number of FCD samples, and the standard deviation of the speed. Including the number of FCD samples as input may give the model some intuition about the traffic volume driving each road section. However, the number of FCD samples can provide information about the real traffic volume only if the fraction of the total number of vehicles equipped with FCD devices is known. The study in [17] proposes the use of a convolutional neural network to predict the traffic speed in a freeway using FCD. The freeway is divided into road sections of equal length, and the speed of the FCD is averaged for each road section. The authors represent then the spatiotemporal evolution of the traffic speed as an image where each pixel of the image corresponds to the average speed in a road section at a certain time step. [18] proposes an error recurrent convolutional neural network model to predict the traffic speed. The architecture of the model is similar to that of [17], and the proposal also takes as input an image representing the spatiotemporal evolution of the traffic. The difference is that the model proposed in [18] also takes as input a vector composed of the prediction error in the last time steps. [19] proposes combining a convolutional neural network with an LSTM recurrent neural network to predict the traffic speed. The input of this model at each time step is also an image of the traffic but the image only represents the spatial evolution of the traffic. In this case, a city map is divided into a grid, and the traffic speed is averaged per cell of the grid. The image represents then the city map and each pixel of the image is the average speed of a cell in the grid.

Most of the studies using FCD focus on predicting the traffic speed or travel time. However, the contributions in [21]–[25] study the prediction of other traffic variables. The proposal in [21] uses support vector regression and FCD from taxis in Beijing to predict the traffic speed and the traffic flow. The support vector regression model takes as inputs the traffic flow and the traffic speed in previous time steps. The traffic speed is aggregated by averaging the speed of the FCD. The traffic flow is estimated by dividing the number of vehicles detected with FCD by the penetration rate of the FCD devices. The penetration rate used in [21] is a statistic published in an annual report on the traffic published by the city of Beijing. It should be noted that assuming that the penetration rate is constant may lead to an incorrect estimation of the traffic flow due to fluctuations of the real penetration rate over space and time. [22] also predicts the traffic flow

using FCD from taxis and the k-nearest neighbors' algorithm. The study differs from [21] in that authors use the FCD to compute the flow of taxis, and then use the flow of taxis as an estimate of the total traffic flow. The authors use the evolution of this estimate in a road section and its neighbor road sections as input to the k-nearest neighbors' algorithm. The output of the algorithm is the prediction of the traffic flow on the road section under evaluation. The proposal in [22] has the advantage that the prediction does not depend on the penetration rate of vehicles with FCD devices. However, the flow of taxis can fluctuate over time and space and may not correctly reflect the evolution of the real total traffic flow. The same approach to estimate the traffic flow is followed in [23] where authors use a convolutional neural network to predict the traffic flow using FCD from taxis of the city of Beijing. A similar approach is followed in [24] where authors propose a probabilistic method to predict the traffic density using FCD from the GPS of taxis. In this case, the authors assume that the density of taxis is an estimate of the real traffic density. The study in [25] uses FCD from taxis and environmental data like the wind speed or the rainfall to predict the traffic flow condition using Boltzmann and support vector machines and conditional random fields. The study defines the traffic flow condition as a discretization of the range of values that the traffic flow can take, so the prediction results into a classification problem. The study uses the FCD to estimate the flow of taxis, and this estimate is converted to a traffic flow condition for the total of the traffic flow. Working with traffic flow conditions instead of working with the real value of the traffic flow allows abstracting the results from the penetration rate of FCD devices. However, we lose the versatility of a numerical representation of the traffic flow.

We should note that existing FCD-based traffic estimation or prediction studies are generally limited by low penetration rates of the technology and by the scarce availability of quality FCD datasets. It is then challenging to analyze the effect that the penetration rate of FCD devices or the quality of the FCD have on the accuracy of traffic predictions without resorting to traffic simulation. This is for example the case of the study presented in [26] where authors use FCD datasets obtained using the microscopic traffic simulator PARAMICS in order to predict the travel time. Using PARAMICS, the authors create a simulated FCD dataset for each penetration rate under study, and they train their prediction models with each dataset. The authors analyze then the prediction error as a function of the penetration rate. The study concludes that the prediction accuracy does not vary greatly with the penetration rate. However, the study does not compare the accuracy of FCD-based predictions with that obtained using data from fixed traffic detectors, and it is then not possible to estimate the amount of FCD needed to outperform predictions with fixed traffic detectors. The study in [27] analyzes the effect of different quality indicators of the FCD on different tasks, including the prediction of the travel time. However, the analysis does not cover the impact of the penetration rate of FCD devices on the traffic prediction. The study in [4] analyzes the

effect of the penetration rate of FCD devices on the accuracy of the travel time estimation. The study fuses FCD with data from induction loops using an ensemble Kalman filter that computes the estimation of the travel time. The authors conclude that fusing both data sources leads to a more accurate estimation of the travel time, and they demonstrate that fusing both data sources is preferable over doubling the number of induction loops. Nonetheless, [4] only covers travel time estimation and not prediction, so the conclusions may not be extrapolated to the prediction of all traffic variables.

The conducted literature review shows the potential of using FCD for traffic prediction but also highlights the challenges in achieving a reliable and realistic prediction of the traffic variables that depend on the penetration rate of FCD, e.g. the traffic flow or the traffic density. Some studies propose techniques to predict the traffic flow or traffic density with FCD, but unrealistic assumptions are made or the studies lose the versatility of the predicted variable. In addition, these studies cannot analyze the effect of the penetration rate on the accuracy of the prediction since they do not have FCD from all the vehicles. Some other studies try to analyze this effect, but they do not compare the accuracy of their techniques with that achieved with fixed traffic detectors data or they just focus on traffic estimation rather than traffic prediction. A better understanding of the impact of the FCD penetration rate on the accuracy of traffic predictions is hence necessary as well as a better understanding of how much such rate affects the accuracy of FCD-based predictions compared to predictions using data from existing traffic detectors. This is relevant to estimate how much FCD is really necessary for reliable traffic predictions in scenarios where the FCD is used alone or in combination with the data from fixed traffic detectors.

III. PROPOSAL

This study proposes a method for predicting or forecasting at the short-term the three fundamental traffic variables (traffic density, traffic flow, and the space mean speed) using FCD. The proposed method is also utilized for predicting the three traffic variables when FCD is used together with data provided by induction loops. The space mean speed can be directly estimated from the FCD independently of the FCD penetration rate since this variable does not depend on the number of vehicles in the road. However, the traffic density and the traffic flow do depend on the number of vehicles in the road, and it is necessary to compute this number in order to estimate these variables using FCD. The number of vehicles on the road can be computed from the FCD if we know the penetration rate of FCD devices. This section presents then first a proposal to estimate the penetration rate of FCD. This estimate is then used to estimate the three fundamental traffic variables using FCD. The section also describes how the three fundamental traffic variables can be estimated using measurements provided by induction loops. The estimates of the three fundamental traffic variables are then used as input to our traffic prediction model that is presented last.

We propose the use of an error recurrent convolutional neural network (eRCNN) for traffic prediction. The model introduces an error feedback mechanism that improves traffic predictions. The proposed model is utilized to predict the three fundamental traffic variables using FCD and induction loops data.

A. ESTIMATION OF THE FCD PENETRATION RATE

The penetration rate of FCD is equal to the ratio of the number of vehicles equipped (and transmitting) with a FCD device to the total number of vehicles. This rate can be obtained from market statistics as done in [21]. However, this approach has two drawbacks. First, the market statistic may not be the exact penetration rate at the time it is used to estimate the traffic variables. Second, the penetration rate may fluctuate over space and time, and using the same rate in all the road network can lead to inaccurate estimates of the traffic variables. In this context, this study proposes an approach to locally and dynamically estimate the FCD penetration rate. The objective is to provide a reliable estimate of the penetration rate at any time and location. Having a reliable estimate of the rate will provide a more precise computation of the number of vehicles on the road, and ultimately a more accurate estimation of the traffic density and the traffic flow that are used as input to the eRCNN prediction model.

We define the local penetration rate as the penetration rate at a certain location and time. We define the penetration rate as the penetration rate on a complete scenario (e.g. a city or freeway); the market statistic used in [21] is a penetration rate. We propose to estimate the local penetration rate by means of data fusion and interpolation. We first utilize FCD and data from fixed traffic detectors to compute the local penetration rate in the location of fixed traffic detectors. We count the number of vehicles equipped with FCD devices that traverse a fixed traffic detector using the FCD, and we compute at the same time the total number of vehicles that traverse the fixed traffic detector. We then compute the local penetration rate at the location of a fixed traffic detector as:

$$\alpha = \frac{N_T^\alpha}{N_T} \tag{1}$$

where α is the local penetration rate, N_T is the total number of vehicles that traverse the fixed traffic detector during a time interval T , and N_T^α is the number of vehicles equipped with FCD devices that traverse the fixed traffic detector during the same time interval T . To compute N_T^α , we use Virtual Trip Lines (VTLs) following [28]. Figure 1 illustrates the concept of a VTL that is defined as an imaginary line that crosses a road from side to side and provides a geographic marker to determine if a vehicle has crossed a certain road section. We place VTLs in the location of the induction loops and using the FCD we determine if a vehicle has traversed a VTL in order to compute N_T^α .

We compute the local penetration rate at locations that do not have a fixed traffic detector by interpolating the local penetration rate between fixed traffic detectors. However, instead

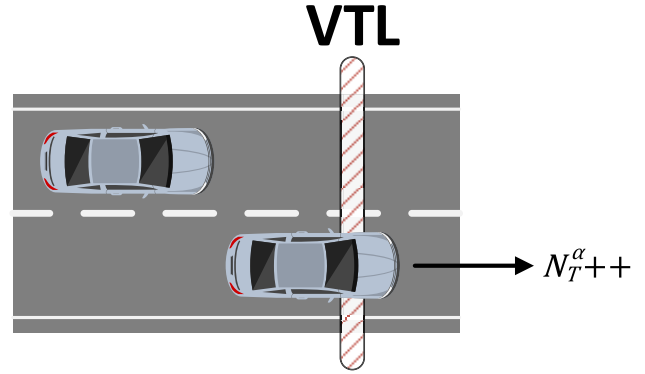


FIGURE 1. Virtual trip lines.

of computing the local penetration rate at any location, we compute it per road section. To this aim, we divide the road network into k road sections of equal length following [17] and [18], and we compute the local penetration rate using (1) for those road sections with a fixed traffic detector. For the rest of road sections, the local penetration rate is interpolated as a function of the distance between the target road section and the road sections with fixed traffic detectors. To this aim, we apply a nearest neighbor interpolation or a linear interpolation.

When we apply the nearest neighbor interpolation to estimate the local penetration rate, we assign a road section the local penetration rate of the nearest road section with a fixed traffic detector. We identify each road section by an index i and define Ω as the set of road sections so that

$$\Omega = \{i \in \mathbb{Z}/0 \leq i < k\} \tag{2}$$

Let $\Omega_D \subseteq \Omega$ be the subset of road sections that have a fixed traffic detector, let $d(i, j)$ be the distance between road sections i and j that have all equal length, and let l_r be the length of the road sections. The distance $d(i, j)$ is equal to:

$$d(i, j) = |i - j| \cdot l_r \tag{3}$$

Then, we can estimate the local penetration rate at any road section α_i using the nearest neighbor interpolation as follows:

$$\alpha_i = \alpha_j / j = \arg \min_{j \in \Omega_D} d(i, j) \tag{4}$$

If the target road section is at equal distance to two road sections with fixed traffic detectors, the local penetration rate of the target road section is set equal to the average of the local penetration rate at the two road sections.

The linear interpolation assumes a linear evolution of the local penetration rate between road sections with fixed traffic detectors. For a road section with index i that is located between two road sections (with indexes m and n) with fixed traffic detectors, we estimate the local penetration rate α_i as:

$$\alpha_i = \frac{\alpha_m - \alpha_n}{m - n} \tag{5}$$

If a road section is not located between two road sections with fixed traffic detectors (e.g., at the beginning of a road

before traversing any traffic detector), the local penetration rate is extrapolated from the two nearest road sections with fixed traffic detectors. In this case, we must ensure that this local penetration rate has a value between 0 and 1. Let m and n be the indexes of the two nearest road sections to the target road section with index i . The local penetration rate of the target road section i is then computed as:

$$\alpha_i = \max\left(0, \min\left(1, \frac{\alpha_m - \alpha_n}{m - n}\right)\right) \quad (6)$$

Using the estimated local penetration rate, we can estimate the total number of vehicles in the road solving (1) for N_T :

$$N_T = \frac{N_L^\alpha}{\alpha_i} \quad (7)$$

B. ESTIMATION OF THE TRAFFIC VARIABLES USING FCD

We are interested in the three fundamental traffic variables, i.e. the traffic flow, the traffic density, and the space mean speed. The traffic flow is defined as the number of vehicles that traverse a reference point in a road per unit time. It is usually measured in veh/h/lane. The traffic density is defined as the number of vehicles that are in a road section with a fixed length at a certain instant. It is usually measured in veh/km/lane. The space mean speed is defined as the average speed of the vehicles that are in a road section with a fixed length at a certain instant. It is usually measured in m/s. Let Q , ρ , and \bar{v}_s be the traffic flow, traffic density, and space mean speed, respectively. These variables are defined as follows:

$$Q = \frac{N_T}{cT} \quad (8)$$

$$\rho = \frac{N_L}{cL} \quad (9)$$

$$\bar{v}_s = \frac{1}{N_L} \sum_{i=1}^{N_L} v_i \quad (10)$$

where N_T is the number of vehicles that traverse a reference point in the road during a time interval of duration T , N_L is the number of vehicles that are in a road section with a fixed length L , c is the number of lanes of the road section, and v_i is the instantaneous speed of the vehicle i at the time of computing \bar{v}_s .

We consider scenarios with FCD penetration rates below 100%. In this case, the traffic variables must be estimated using the local FCD penetration rate we previously computed, and we do so over a time interval of duration T . The estimates of the traffic density and the space mean speed are then estimates of their average during the time interval of duration T .² The estimate of the traffic density $\hat{\rho}$ can be computed as:

$$\hat{\rho} = \frac{\sum_{t=1}^T N_{L,t}^\alpha}{\alpha_i c L T}, \quad N_{L,t}^\alpha \leq N_L, \quad 0 < \alpha_i \leq 1 \quad (11)$$

where α_i is the local penetration rate and $N_{L,t}^\alpha$ is the number of connected vehicles that provide FCD and are in the road

section in the time step t . This number is lower or equal to the total number of vehicles that are in the road section, N_L . The estimate of the space mean speed \hat{v}_s does not depend on the local FCD penetration rate and can be computed as:

$$\hat{v}_s = \frac{1}{\sum_{t=1}^T N_{L,t}^\alpha} \sum_{t=1}^T \sum_{i=1}^{N_{L,t}^\alpha} v_{ii}, \quad N_{L,t}^\alpha \leq N_L \quad (12)$$

where v_{ii} is the speed of vehicle i in the time step t .

We can then estimate the traffic flow \hat{Q} as:

$$\hat{Q} = \hat{\rho} \hat{v}_s \quad (13)$$

The three traffic variables can be estimated using equations (11), (12) and (13) except when $\alpha_i = 0$ or $N_L^\alpha = 0$. If $\alpha_i = 0$ then $N_L^\alpha = 0$ and the estimates of the traffic variables are indeterminate. $N_L^\alpha = 0$ does not necessarily imply that $\alpha_i = 0$. N_L^α might be null, for example, when $N_L = 0$. In this case, we are unable to determine α_i and estimate the traffic variables. In both cases, we are not getting FCD from any vehicle because there are no connected vehicles in the studied road section. It is very unlikely that there are no connected vehicles in case of traffic congestion, so when $\alpha_i = 0$ or $N_L^\alpha = 0$ we assume that the traffic conditions correspond to free flow, and $\hat{\rho} = 0$, $\hat{Q} = 0$, and $\hat{v}_s = v_{FF}$, where v_{FF} is the speed at free flow conditions in the studied road section. Following [29], we compute v_{FF} for each road section as the average of the speed at the section during all time intervals with a Level Of Service (LOS) A and with $N_L^\alpha \neq 0$. [29] defines LOS A as the level of service experienced when the traffic density is lower or equal to 6.84veh/km/lane (11veh/mi/lane).

C. ESTIMATION OF THE TRAFFIC VARIABLES USING DATA FROM INDUCTION LOOPS

We consider induction loops (like those used in the evaluation scenario) that provide measurements of the traffic flow, the time mean speed, the occupancy, and the mean length of the vehicles. The time mean speed \bar{v}_t is here defined as the average speed of the vehicles that traverse the induction loop during a time interval, it is measured in m/s and it is used as an estimate of the space mean speed. The occupancy occ is here defined as the percentage of time that the induction loop is covered by a vehicle. It is used as an estimate of the occupancy of the road, and hence provides information about the traffic density. The mean length of the vehicles \bar{l} is defined as the average of the lengths of the vehicles that traverse the induction loop, and it is measured in m. These three variables can be computed as follows:

$$\bar{v}_t = \frac{1}{N_T} \sum_{i=1}^{N_T} v_i \quad (14)$$

$$occ = 100 \frac{t_{occ}}{T} = 100 \frac{\sum_{i=1}^{N_T} t_i}{T} = 100 \frac{\sum_{i=1}^{N_T} \frac{l_{loop} + l_i}{v_i}}{T} \quad (15)$$

$$\bar{l} = \frac{1}{N_T} \sum_{i=1}^{N_T} l_i \quad (16)$$

²The traffic flow is already defined for a time interval.

where N_T is the number of vehicles that traverse the induction loop during a time interval of length T . In this study, we set the time interval equal to the time between measurements provided by the induction loop (60 seconds). v_i is the speed of the vehicle i when traversing the induction loop. t_{occ} is the amount of time that the induction loop is covered by a vehicle, i.e. it is equal to the sum of the times t_i that each vehicle i covers the induction loop. t_i is the time a vehicle i takes to travel a distance equal to the sum of the length of the induction loop l_{loop} and the length of the vehicle l_i .

We can then estimate the traffic density $\hat{\rho}$ as:

$$\hat{\rho} = \frac{\frac{occ}{100}}{\frac{\bar{l}}{1000}c} = \frac{10occ}{\bar{l}c} \quad (17)$$

where c is the number of lanes in the road section where the induction loop is located. Note that in (17) the mean length is converted to km in order to express the traffic density in veh/km/lane. The traffic flow Q is directly measured by the induction loops.

D. TRAFFIC PREDICTION

Deep learning-based neural networks are currently the state of the art in short-term traffic prediction or forecast [30]. Neural networks predict the future values of the traffic variables using as input their past evolution. For traffic prediction or forecasting, neural networks are trained using supervised learning. Supervised learning uses datasets that include inputs to the neural network and the correct output for that input. These datasets are used to train the neural networks in order to find the set of parameters of the network that minimize a loss function, for example, the prediction error (usually the quadratic error of the prediction).

In this study, we predict the traffic variables using an error recurrent convolutional neural network (eRCNN) [18], since we previously demonstrated that this network achieves the best traffic predictions using data from fixed traffic detectors under general traffic conditions and under traffic congestion [31]. The eRCNN model takes as input the estimates of the three fundamental traffic variables. The input is organized in the format of an (traffic) image with different channels (like an RGB image). The number of channels of the input image depends on the prediction approach. When the traffic variables are predicted with FCD or induction loop data only, the input will have three channels, one for each traffic variable. When the traffic variables are predicted using both FCD and induction loop data, the input image has six channels, three for the traffic variables estimated with FCD and three for the traffic variables estimated with induction loops. Figure 2 illustrates the format of the input of the eRCNN for an image with three channels. Each channel is a matrix with one dimension corresponding to the spatial evolution of a traffic variable (the vertical dimension in our implementation) and the other dimension corresponding to the temporal evolution of the variable (the horizontal dimension in our implementation). The columns of the matrix represent the state of the traffic variable for the complete road at a certain

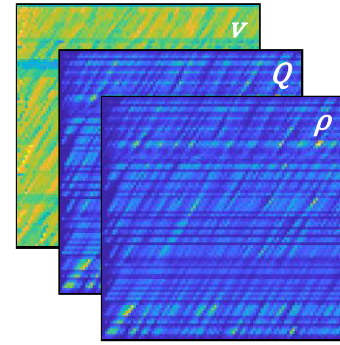


FIGURE 2. Illustration of the format of the input image to an eRCNN where the image of the traffic has three channels representing each the spatio-temporal evolution of a traffic variable.

time step. The rows of the matrix represent the temporal evolution of the traffic variable at a specific location or road section. Each pixel of an image contains then the value of the represented traffic variables at a specific location and time step.

The input image is fed to the eRCNN, which is composed by a convolutional layer, an average pooling layer and three fully connected layers. The convolutional layer applies 32 convolutional filters of size 3×3 to the input of the eRCNN. This convolutional layer outputs 32 feature maps, which are fed to an average pooling layer of size 2×2 that reduces the height and width of the feature maps. The output of the average pooling layer is converted into a one-dimensional vector that is processed by a fully connected layer composed by 256 neurons. In parallel, another fully connected layer composed by 32 neurons process a vector of the prediction error in the last six time steps. The output of these two fully connected layers is concatenated and fed to the output layer. The output layer is a fully connected layer composed by a single neuron, which is in charge of computing the prediction. The convolutional layer and the parallel fully connected layers use the ReLU activation function [32], while the output layer uses the identity activation function. The architecture of the eRCNN is depicted in Figure 3. We demonstrated in [31] that this architecture and the design of the eRCNN is the optimum one for predicting the traffic under normal traffic conditions and under traffic congestion. All the eRCNNs used in this work are trained using the back-propagation through time algorithm (BPTT) and the ADAM algorithm [33], which is a variation of the stochastic gradient descent (SGD) algorithm. For this purpose, we use batches of 20 training examples, each one consisting in a sequence of 20 time steps. All eRCNNs have been trained to predict the value of one of the three fundamental traffic variables in the next 15 minutes. The loss function minimized during the training is the squared L2 norm of the prediction error:

$$\mathcal{L} = \frac{1}{2} \|\hat{y} - y\|_2^2 \quad (18)$$

where \mathcal{L} is the value of the loss function, y is the ground truth, \hat{y} is the prediction of the eRCNN, and $\|\cdot\|$ is the L2 norm.

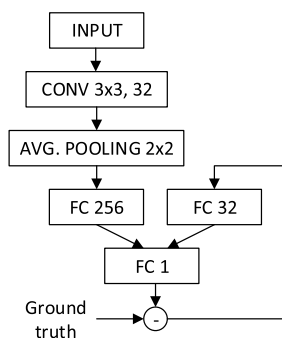


FIGURE 3. Architecture of the eRCNN.

By minimizing the L2 norm, we train the eRCNN to predict traffic conditions with minimum squared error. To facilitate the training process, we employ learning rate exponential decay. This technique allows using larger learning rates at the beginning of the process so that the training converges faster, and smaller learning rates at the end so that the final training error is small. We start with a learning rate with a value of 0.1, and this value is multiplied by 0.5 every epoch. We also use gradient clipping so that the training process is more stable at the beginning. Gradient clipping limits the maximum value of the norm of the gradient (to a value of 40 in this study) in order to avoid oscillations in the training process. All eRCNNs are trained for 10 epochs with early stopping. Early stopping is a technique that consists in stopping the neural network training process when the validation error stops improving and starts increasing. The use of early stopping helps avoiding overfitting. The training process and the eRCNN model have been implemented using the TensorFlow framework [34].

IV. SCENARIO AND DATASETS

A. SCENARIO

This study is conducted using the road traffic on the Spanish A-7 freeway section that connects the cities of Alicante and Murcia and that is shown in Figure 4. This section is 97 km long and serves three mid-sized cities (Alicante, Murcia and Elche) and an important industrial and touristic area with a total population of around 2 million people. The traffic on the road section can vary from free flow to traffic congestion, but is usually quite busy with certain areas near Murcia experiencing an Average Daily Traffic (ADT) higher than 100000 vehicles per day (88.8% light vehicles, 11.2% heavy vehicles). The ADT between Alicante and Elche can reach values as high as 83000 vehicles per day (93.7% light vehicles, 6.3% heavy vehicles).

This study uses simulated road traffic for the selected freeway section using SUMO³ [35]. The traffic is generated using the digital simulation scenario presented in [36] and that is openly available for download in a public repository.⁴ The scenario realistically simulates the road traffic over the

³SUMO (Simulation of Urban MObility) is a popular open-source microscopic traffic simulator.

⁴<https://github.com/jjgonde/Alicante-Murcia-SUMO-Scenario>

selected section for 9 full days of traffic. The simulation scenario has been calibrated using real traffic flow, speed and occupancy measurements provided for the 9 full days of traffic by induction loops managed by the Spanish road authority DGT. The measurements were collected using 99 induction loops deployed along the scenario (on the mainline and on the on- and off-ramps). Figure 4 shows the location of some of these induction loops. The calibration process ensures that the simulated traffic generated with the scenario matches very closely the real measurements provided by the induction loops on the mainline and on the on-ramps and off-ramps of the freeway.⁵ The digital scenario hence accurately models the traffic flow, speed and road's occupancy over the complete 97 km long freeway for 9 full days considering mixed traffic with light and heavy vehicles. To the authors' knowledge, the selected traffic scenario is one of the largest (both in space -97km long- and time- 9 full days of traffic) and more accurate traffic simulation scenarios openly available in SUMO.

Using SUMO, we reproduce the full nine days of traffic and collect the FCD for the vehicles in the scenario as well as the measurements at the induction loops. The data is collected at each time step of the simulation that is set equal to 1 second. The induction loops are placed in SUMO at the same locations as the induction loops deployed along the 97 km by the Spanish road authority and that provided the measurements for the calibration of the scenario.⁶ Like the induction loops deployed along the selected 97 km of freeway, the induction loops in SUMO measure the traffic flow, the time mean speed, the occupancy, and the mean length of the vehicles. We collect these measurements for the 20 induction loops in the mainline of the scenario.⁷ We collect FCD for different FCD penetration rates, in particular, penetration rates of 1%, 2%, 3%, 4%, 5%, 10%, 20%, 30%, 50%, 80% and 100%.

B. NEURAL NETWORK FOR TRAFFIC PREDICTION

We use the collected FCD and induction loop data to estimate the three fundamental variables following the process described in Sections III.B and III.C. In this study, we compute the estimates of the variables for intervals of $T = 60$ seconds. We divide the 97 km freeway segment into 1 km sections of equal length like in [3] and [18], and we only consider 95 sections to estimate the traffic variables. The first and last road sections are not considered since the simulated vehicles appear and disappear from the simulation in these sections.

The estimates of the traffic variables for the first seven days of traffic are used to train the neural network model for short-term traffic prediction or forecasting. The datasets for the last

⁵The induction loops on the mainline measure the flow of light and heavy-duty vehicles.

⁶The measurements collected at the induction loops in SUMO closely match the measurements obtained by the real induction loops deployed along the 97 km of freeway for the nine days of traffic.

⁷These measurements are used to compute the variables used with the eRCNN.



FIGURE 4. Simulated scenario and location of some of the induction loops.

two days are used for validation and testing (one day each). The training is done by means of supervised learning, so we create (input, output) pairs where the input is the estimates of the traffic variables in the form of images and the output is the ground truth, the real value of the traffic variables at the time instant that they are predicted (i.e. 15 min after the time instant when the input image is computed). The training and validation sets are used to optimize the parameters of the neural network in order to minimize the prediction error. The ground truth for these two datasets correspond to the future values of the traffic variables estimated using the local FCD penetration rate corresponding to each scenario.⁸ For the test set, the ground truth corresponds to the future values of the traffic variables computed considering an FCD penetration rate equal to 100%.⁸ This is the case because the test set is used to check the real accuracy of the traffic prediction so we need to compare the prediction of the neural network with the real state of the traffic. We compared the evolution of the training and validation errors and none of the trained eRCNNs experienced overfitting, which demonstrates that the datasets utilized are sufficient to validate the performance of the prediction models.

The input of the neural network for traffic prediction is organized into images. The images have three channels when predicting the traffic with FCD or with induction loop data. Each channel corresponds to the spatiotemporal evolution of a traffic variable. The images have six channels when the traffic is predicted using the FCD and induction loop data. Three channels correspond to the spatiotemporal evolution of the three fundamental traffic variables estimated using FCD, and the three other channels to the spatiotemporal evolution of the three fundamental traffic variables estimated using data

from the induction loops. When the traffic is predicted using FCD, the images have three channels, a width of 72 pixels and a height of 95 pixels since the freeway is divided into 95 sections.⁹ The 72 pixels for the width represent the temporal evolution of a traffic variable for the previous 3.6 hours. Each pixel corresponds to traffic data aggregated for three minutes. When the traffic is predicted using data from induction loops only, the images also have three channels and a width of 72 pixels, but the height is equal to 20 pixels since there are 20 induction loops on the mainline of the 97 km freeway segment. When the traffic is predicted using FCD and induction loops data, the images have six channels, a width of 72 pixels and a height of 95 pixels. In this case, the channels corresponding to the traffic variables estimated using data from the induction loops need to be resized to match the same height as those corresponding to the variables estimated with FCD. To this aim, all pixels for the channels estimated with the data from the induction loops have a value equal to zero, except those corresponding to the freeway sections where an induction loop is located. These pixels contain the values of the traffic variables estimated with the data from the induction loops. Figure 5 illustrates a channel representing a traffic variable estimated with the data from the induction loops that has been resized prior to being introduced as input to the neural network together with the images obtained from the FCD. We should note that setting all the pixels for freeway sections without an induction loop equal to zero may represent a challenge for the eRCNN if it cannot distinguish these preset values from real measurements of a traffic variable that are equal to zero. [37] proposes to address this challenge by adding an occupancy layer or mask to the input that indicates the neural network which

⁸This applies to traffic predictions using FCD only or FCD and induction loop data.

⁹As explained, it is originally divided into 97 1 km-long sections but we discard the first and last ones since they are used to introduce and eliminate vehicles at the beginning and end of the simulations.

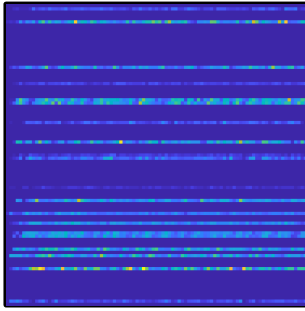


FIGURE 5. Format of a channel of a traffic variable estimated with induction loops.

pixels correspond to real measurements equal to zero and which ones correspond to a preset value indicating the lack of measurements. We evaluated the impact of adding this occupancy layer to our traffic prediction model but it did not improve the prediction accuracy. This is the case because our input to the eRCNN is not a single traffic variable but the three traffic variables (both for FCD and induction loop data). If a pixel corresponds to a location where there is not an induction loop, then all three traffic variables will be equal to zero. This can only happen when the pixel corresponds to a location without an induction loop since an induction loop that detects no traffic (traffic flow and density equal to 0) measures a speed equal to the free flow speed. Our eRCNN was able to use this information to distinguish between preset values equal to zero for pixels representing freeway sections without induction loops and real traffic measurements. This is why the proposed occupancy layer did not improve our prediction accuracy. However, the occupancy layer proposed in [37] can improve the accuracy of prediction models when the input to these models is a single traffic variable like in [37].

V. EVALUATION

This section evaluates the short-term traffic prediction or forecasting accuracy achieved with our proposal when predicting the three fundamental traffic variables, i.e. the traffic density, the traffic flow, and the space mean speed. The prediction accuracy is evaluated when using FCD only, induction loop data only, or using both FCD and induction loop data. The evaluation allows us analyzing the impact of the FCD penetration rate on the accuracy of the prediction of the three traffic variables, as well as the impact of fusing FCD and data from traffic detectors on the prediction accuracy.

A. ESTIMATION OF THE TRAFFIC VARIABLES

The traffic prediction model uses as input the estimates of the traffic variables. When these estimates are obtained from FCD, we need information about the FCD penetration rate to estimate two of the traffic variables. This section analyzes the estimation error achieved with our proposals to compute the local FCD penetration rate using data from FCD and induction loops, and compare it to that achieved when considering a FCD penetration rate that is maintained constant

along the road and time like in [21]. For the comparison, we compute three error metrics, the Mean Average Percentage Error (*MAPE*), the Mean Average Error (*MAE*) and the Root Mean Squared Error (*RMSE*), that are defined as follows:

$$MAE = \frac{1}{N} \sum_{i=1}^N |\hat{y}_i - y_i| \quad (19)$$

$$MAPE = \frac{100}{N} \sum_{i=1}^N \left| \frac{\hat{y}_i - y_i}{y_i} \right| \quad (20)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i)^2} \quad (21)$$

where \hat{y}_i is the estimation or prediction of a traffic variable, y_i is the ground truth, and N is the number of samples used to compute the *MAPE*, *MAE*, and *RMSE*. The ground truth corresponds to the real value of the traffic variable, i.e. to the value of the variable computed in the scenario where all vehicles are connected (i.e. $\alpha = 1$).

Figure 6 depicts the *MAPE* of the estimation of the traffic density and the traffic flow using FCD only. Results are shown using a constant FCD penetration rate in the scenario (Fixed), and using the local estimate of the FCD penetration rate obtained with the nearest neighbor interpolation (Nearest) and with the linear interpolation (Linear). The *MAPE* is represented as a function of the FCD penetration rate in the scenario. We do not depict the *MAPE* for the estimate of the space mean speed since this estimate does not depend on the accuracy of the estimate of the FCD penetration rate.¹⁰ Figure 6 clearly shows that our proposals to locally estimate the FCD penetration rate achieve a more accurate estimation of the traffic variables than assuming that the FCD penetration rate remains constant. This is important because the estimates of the traffic variables are used as input to the traffic prediction model, so an inaccurate estimation of the traffic variables can impact the traffic prediction accuracy. Both interpolation approaches achieve similar performance, being the linear interpolation approach slightly better than the nearest neighbor approach. The same trends have been observed for the *MAE* and the *RMSE*.

The results obtained show that we can achieve a more accurate estimate of the traffic variables using our local estimate of the FCD penetration rate than assuming a constant FCD penetration rate. Figure 6 shows that the benefit of using a local FCD penetration rate are particularly relevant at low (or below 30%) FCD penetration rates.¹¹ We can then conclude that our proposals are capable of accurately estimating the local penetration rate at any location and

¹⁰However, the estimate of the space mean speed is influenced by the sample of FCD data used for the computation.

¹¹When the FCD penetration rate increases, the local FCD penetration rate on the freeway sections becomes more homogeneous and closer to the global rate. In this case, the differences observed in Figure 6 due to the assumption that the FCD penetration rate is constant (in time and space) decrease, and a more accurate estimation of the traffic variables can be achieved using the Fixed approach.

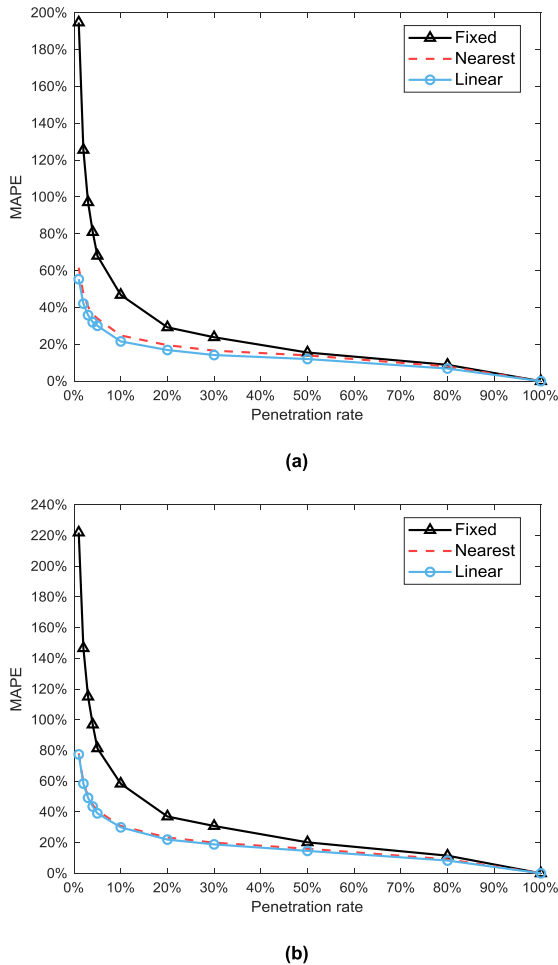


FIGURE 6. MAPE of the estimation of (a) the traffic density, and (b) the traffic flow as a function of the penetration rate of FCD devices. The results are shown for the three approaches to estimate the penetration rate.

time instant. As expected, Figure 6 also shows that the error of the estimation of the traffic variables decreases with the FCD penetration rate for the three approaches. The estimate of the traffic density depends on the FCD penetration rate, so when more FCD is available we can better estimate the local FCD penetration rate and the uncertainty on the estimate of the traffic density decreases. The same occurs with the estimate of the traffic flow since this estimate depends on the estimate of the traffic density, and hence on the estimate of the local FCD penetration rate.

Figure 7 compares the MAPE of the estimation of the traffic density that is achieved with the linear interpolation approach and with the fixed approach. Results are represented for the linear interpolation since it achieved lower estimation errors than the nearest neighbor interpolation (Figure 6). The results are represented for penetration rates of FCD in the scenario equal to 1% ($\alpha = 0.01$), 10% ($\alpha = 0.1$), and 50% ($\alpha = 0.5$). Figure 7 clearly shows that the interpolation approach outperforms the fixed approach as the estimation of the traffic density is closer to the ground truth value for

all penetration rates or values of α . The differences are particularly visible for the lower penetration rates. For example, Figure 7 shows that the interpolation approach achieves an estimation of the traffic density for $\alpha = 0.1$ close to that obtained with the fixed approach for $\alpha = 0.5$, i.e. for a significantly higher FCD penetration rate. This trend is also observed in Figure 6, and similar results and comparisons are obtained with the traffic flow as it can be directly derived from the estimate of the traffic density and equation in (13). These results demonstrate again the importance and impact of the proposed techniques to accurately and locally estimate the FCD penetration rate. The proposed techniques result in a more accurate estimation of the traffic variables that are then introduced as input to our traffic prediction model.

B. SHORT-TERM TRAFFIC PREDICTION ACCURACY

The estimated traffic variables are formatted into traffic images as described in Section IV.B and introduced to the eRCNN short-term prediction model. Figure 8 compares the accuracy of the prediction of the three fundamental traffic variables as a function of the FCD penetration rate in the scenario. Results are shown when the traffic prediction uses: 1) data from inductions loop only (Loops),¹² 2) FCD only and assumes a constant FCD penetration rate (FCD-Constant), 3) FCD only and estimates the local penetration rate with linear interpolation (FCD-Linear), 4) FCD and induction loops data and assumes a constant FCD penetration rate (Fusion-Constant), and 5) FCD and induction loops data and estimates the local penetration rate with linear interpolation (Fusion-Linear). We chose to represent results only with the linear interpolation as it outperformed the nearest neighbor interpolation in the results reported in the previous section.

The comparison of Figure 8 and Figure 6 shows that the prediction error is lower than the estimation error for all traffic variables and all configurations tested.¹³ This is the case because the eRCNN neural network model used in this study is capable to adapt its predictions based on its prediction errors in previous time steps. Thanks to this error feedback, the eRCNN learns how to react after a bad prediction. This improves the prediction error even when the training data is not fully reliable, for example, when the sample of FCD is small due to a low FCD penetration rate. Despite the error feedback in the eRCNN model, the prediction accuracy is of course affected by the FCD penetration rate and the FCD sample size for training the prediction model. This is visible in Figure 8 that shows how the MAPE of the prediction decreases with the FCD penetration rate. The same trends have been observed for the MAE and RMSE.

¹²The prediction using data from induction loops only is independent of the FCD penetration rate and hence results in a constant prediction error in Figure 8.

¹³Except when the penetration rate is 100%. In this case, a lower error is observed for the estimation compared to the prediction because the estimation of the traffic variables is the ground truth, hence a null estimation error is achieved.

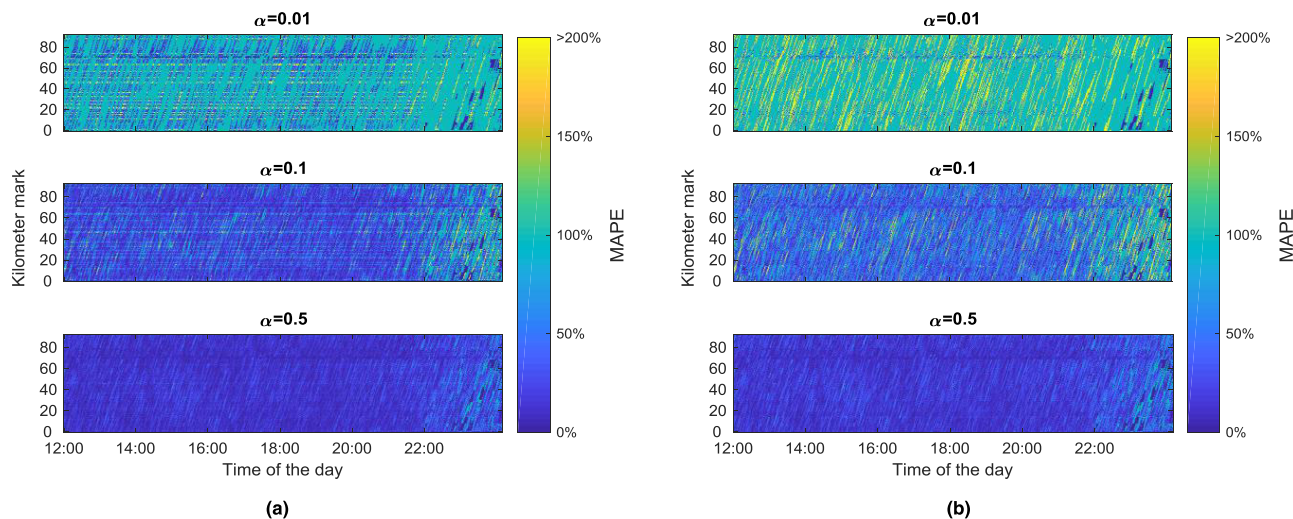


FIGURE 7. MAPE of the Estimation of the traffic density using the (a) linear interpolation approach and the (b) fixed approach in scenarios with FCD penetration rates of 1%, 10%, and 50%.

Figure 8 shows that the prediction with data from inductions loop only (Loops) achieves the lowest prediction error only when the FCD penetration rate is low. This is the case because the size of the FCD sample is too low to achieve accurate traffic predictions using FCD. For higher FCD penetration rates, the best prediction accuracy is achieved when using both FCD and induction loops data and we estimate the local FCD penetration rate with linear interpolation (Fusion-Linear). On the other hand, the lowest prediction accuracy is obtained when using FCD only and assuming a constant FCD penetration rate (FCD-Constant). The differences among configurations are maintained for all FCD penetration rates and the three fundamental traffic variables, except when estimating the traffic density and traffic flow with a FCD penetration rate equal to 100%. In this case, all vehicles are connected and provide FCD, and the value of the local FCD penetration rates and the data from induction loops decreases.

Figure 8 demonstrates that fusing FCD and data from induction loops improves the prediction accuracy in general. In this study, we have used both sources of data at two different stages of the traffic prediction process: when computing the local FCD penetration rate to estimate the traffic variables, and when generating the input image to the eRCNN neural network model. The former proves to improve the prediction as the FCD-Linear configuration reduces the prediction error compared to the FCD-Constant configuration. Using both data sources for creating the input image to the eRCNN model also improves the prediction since we provide more complete information to the prediction model. This results in that the two configurations that achieve the best prediction accuracy are Fusion-Linear and Fusion-Constant.

Figure 8 allows determining the minimum FCD penetration rate that is necessary to improve the prediction accuracy compared to that obtained when using only induction loop data (Loops). This penetration rate is equivalent to the minimum fraction of the vehicles we need to sample in order

to have a traffic prediction as accurate as that using only induction loop data. This is important as FCD is usually provided by third party providers, and traffic managers and/or authorities need to acquire FCD for their traffic estimations and predictions and the price may vary depending on the FCD sample size (and hence on the FCD penetration rate). Figure 8 shows that the prediction with data from inductions loop only (Loops) achieves the lowest prediction error only when the FCD penetration rate is low. The minimum FCD penetration rate to achieve better prediction accuracy compared to the Loops configuration depends on the data used for the prediction and the estimation of the FCD penetration rate. For example, when using the FCD-Constant configuration, we need a 5% FCD penetration rate for improving the prediction accuracy for the traffic density and the space mean speed. This value increases to 10% in the case of the traffic flow. This result shows that traffic managers and authorities need a sample of FCD corresponding to a 10% of the vehicles to be able to start improving the traffic prediction accuracy compared to what they can achieve nowadays with the induction loops deployed in the freeway scenario under evaluation. This is an important information to guide them into how much data they need to acquire to be able to improve the predictions achieved with their current traffic detectors. Introducing the local FCD penetration rate proposal (FCD-Linear) helps reducing a bit the FCD penetration rate required to improve the prediction accuracy of the three traffic variables compared to when using only data from induction loops. In particular, the FCD penetration rate required reduces from 10% for FCD-Constant to 8.6% for FCD-Linear. Higher reductions in the FCD penetration rate (or sample of FCD needed) that is necessary to improve the prediction accuracy of the Loops configuration are obtained with our proposal to combine the FCD and induction loops data for the estimation of the local FCD penetration rate and the input image to the eRCNN prediction model. In particular, the Fusion-Linear

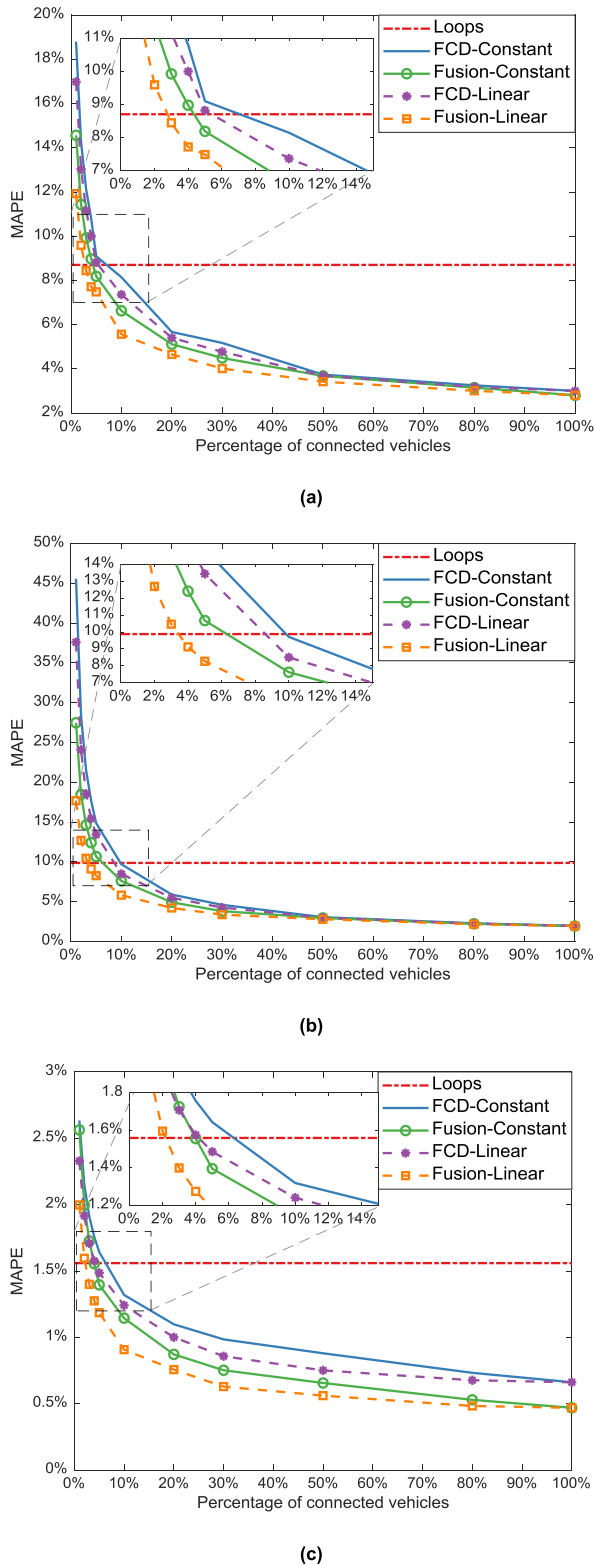


FIGURE 8. MAPE of the prediction of (a) the traffic density, (b) the traffic flow, and (c) the space mean speed as a function of the FCD penetration rate.

configuration requires approximately three times less FCD than the FCD-Constant configuration to improve the prediction accuracy compared to when using only data from

induction loops (Loops in Figure 8).¹⁴ Figure 8 shows that the Fusion-Linear configuration improves the prediction of the three traffic variables compared to the Loops configuration with just a sample of FCD corresponding to a 4% FCD penetration rate. This is again an important information to traffic managers and authorities since it shows that using the induction loops data and an advanced traffic prediction model like the one presented in this study can provide predictions of the three fundamental traffic variables with a relatively small sample of FCD.

Figure 8 shows that the prediction error decreases for all configurations (except Loops) as the FCD penetration rate increases. It is though interesting to observe that increasing the FCD penetration rates beyond certain values does not result in a significant gain in prediction accuracy. For example, doubling the FCD penetration rate from 5% to 10% using the Fusion-Linear configuration decreases the prediction error for the traffic density from 7.49% to 5.57%. On the other hand, increasing the FCD penetration rate from 20% to 50% only decreases the prediction error for the traffic density from 4.66% to 3.43%. The gains are even smaller when we increase the FCD penetration rate above 50%. Similar trends are also observed when using only the FCD for the traffic prediction (FCD-Linear and FCD-Constant). These results are important since they provide indications to the traffic managers and authorities on the impact of the amount of FCD on the traffic prediction accuracy. This will help them limit the purchase of FCD to only the data they need to reduce the prediction errors to their target values.

Finally, it is important to highlight that the Fusion-Linear configuration of our prediction model (i.e. using FCD and induction loops data as input to the eRCNN and for estimating the FCD penetration rate) significantly reduces the amount of FCD needed to achieve high accuracy levels compared to when using only FCD for the prediction (FCD-Constant configuration). For example, the Fusion-Linear configuration achieves a MAPE of 5% in the prediction of the traffic density and the traffic flow when the FCD penetration rate is 16.24% and 15.19% respectively. The FCD-Constant configuration increases these values to 32.46% and 27.01%. The FCD penetration rate required by the FCD-Constant configuration to achieve a 1% MAPE in the prediction of the space mean speed is 28.66% while it is only 8.34% with the Fusion-Linear configurations. These results highlight how an adequate combination of FCD and induction loops data together with an error recurrent eRCNN prediction model significantly reduces the amount of FCD necessary to achieve high prediction accuracy levels.

VI. CONCLUSION

This study has analyzed for the first time the impact of the FCD penetration rate and the impact of fusing FCD and

¹⁴The Fusion-Constant configuration requires approximately half of the FCD needed by the FCD-Constant configuration to start improving the accuracy of the traffic prediction compared to when using only induction loop data.

data from traffic detectors (in our case, induction loops) on the accuracy of the short-term prediction or forecasting of the three fundamental traffic variables (the traffic density, the traffic flow, and the space mean speed). To do so, this study presents a short-term traffic prediction model that can accurately predict the three fundamental traffic variables using FCD. The model includes an error recurrent convolutional neural network that takes as input estimates of the three traffic variables. The variables are estimated using FCD and data from induction loops as well as a method to locally and dynamically estimate the FCD penetration rate. This method improves the estimation of the three traffic variables that are used as input to the eRCNN model, and hence their short-term traffic prediction. The conducted evaluation has demonstrated that our proposed model can achieve high prediction accuracy levels for the three traffic variables using a small FCD sample or FCD penetration rate if the FCD is combined with data from traffic detectors. In this case, the study shows that our proposal only requires FCD from 4% of the vehicles in the scenario to improve the prediction accuracy achieved with traffic detectors. This is compared to needing FCD from 10% of the vehicles in the scenario when using only FCD for the prediction. The study also shows that combining FCD and traffic detectors data with our prediction model helps achieving high prediction accuracy levels with significantly less FCD than when only using the FCD for the prediction. Increasing the FCD sample size (or FCD penetration rate) significantly improves the prediction accuracy achieved with our prediction model compared to that obtained with data from traffic detectors. However, our study shows that increasing the FCD penetration rates beyond certain values does not result in a significant gain in prediction accuracy. This result provides valuable insights into the amount of FCD needed to exploit the potential of connected vehicles to achieve high accuracy prediction levels.

REFERENCES

- [1] 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.
- [2] 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.
- [3] C. de Fabritiis, R. Ragona, and G. Valenti, "Traffic estimation and prediction based on real time floating car data," in *Proc. IEEE Conf. Intell. Transp. Syst.*, Beijing, China, Oct. 2008, pp. 197–203.
- [4] A. D. Patire, M. Wright, B. Prodhomme, and A. M. Bayen, "How much GPS data do we need?" *Transp. Res. C, Emerg. Technol.*, vol. 58, pp. 325–342, Sep. 2015.
- [5] 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.
- [6] Y. Tian and L. Pan, "Predicting short-term traffic flow by long short-term memory recurrent neural network," in *Proc. IEEE SmartCity*, Chengdu, China, Dec. 2015, pp. 153–158.
- [7] J. Raj, H. Bahuleyan, and L. D. Vanajakshi, "Application of data mining techniques for traffic density estimation and prediction," in *Proc. TPMDC*, Mumbai, India, 2016, pp. 321–330.
- [8] X. Yang, Y. Zou, J. Tang, J. Liang, and M. Ijaz, "Evaluation of short-term freeway speed prediction based on periodic analysis using statistical models and machine learning models," *J. Adv. Transp.*, vol. 2020, Jan. 2020, Art. no. 9628957.
- [9] X. Miao, B. Wu, Y. Zou, and L. Wu, "Examining the impact of different periodic functions on short-term freeway travel time prediction approaches," *J. Adv. Transp.*, vol. 2020, Aug. 2020, Art. no. 3463287.
- [10] A. Salamanis, D. D. Kehagias, C. K. Filelis-Papadopoulos, D. Tzovaras, and G. A. Gravvanis, "Managing spatial graph dependencies in large volumes of traffic data for travel-time prediction," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 6, pp. 1678–1687, Jun. 2016.
- [11] F. Rempe, G. Huber, and K. Bogenberger, "Travel time prediction in partitioned road networks based in floating car data," in *Proc. ITSC*, Rio de Janeiro, Brazil, Dec. 2016, pp. 1982–1987.
- [12] 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, Jan. 2016.
- [13] T. Zhu, X. Kong, and W. Lv, "Large-scale travel time prediction for urban arterial roads based on Kalman filter," in *Proc. Int. Conf. Comput. Intell. Softw. Eng.*, Wuhan, China, Dec. 2009, pp. 1–5.
- [14] C. Guo, "Traffic time prediction based on imputation algorithms for missing values," in *Proc. IC-NIDC*, Guiyang, China, Nov. 2018, pp. 223–228.
- [15] A. Boukerche and J. Wang, "Machine learning-based traffic prediction models for intelligent transportation systems," *Comput. Netw.*, vol. 181, Nov. 2020, Art. no. 107530.
- [16] 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. MT-ITS*, Budapest, Hungary, Aug. 2015, pp. 93–101.
- [17] X. Ma, Z. Dai, Z. He, J. Ma, Y. Wang, and Y. Wang, "Learning traffic as images: A deep convolutional neural network for large-scale transportation network speed prediction," *Sensors*, vol. 17, no. 4, pp. 1–16, 2017.
- [18] J. Wang, Q. Gu, J. Wu, G. Liu, and Z. Xiong, "Traffic speed prediction and congestion source exploration: A deep learning method," in *Proc. IEEE 16th Int. Conf. Data Mining*, Barcelona, Spain, Dec. 2016, pp. 499–508.
- [19] 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, pp. 1–16, 2017.
- [20] J. Y. Wang, K. I. Wong, and Y. Y. Chen, "Short-term travel time estimation and prediction for long freeway corridor using NN and regression," in *Proc. ITSC*, Anchorage, AK, USA, Oct. 2012, pp. 582–587.
- [21] Q. Yang, J. Wang, X. Song, X. Kong, and Z. Xu, "Urban traffic congestion prediction using floating car trajectory data," in *Proc. ICAPP*, Zhangjiajie, China, Dec. 2015, pp. 18–30.
- [22] 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. 246–265, Feb. 2016.
- [23] J. Zhang, Y. Zheng, and D. Qi, "Deep spatio-temporal residual networks for citywide crowd flows prediction," in *Proc. 31st AAAI Conf. Artif. Intell.*, San Francisco, CA, USA, Feb. 2017, pp. 1655–1661.
- [24] P. S. Castro, D. Zhang, and S. Li, "Urban traffic modeling and prediction using large scale taxi GPS traces," in *Proc. ICPC*, Newcastle, U.K., Jun. 2012, pp. 57–72.
- [25] X. Niu, Y. Zhu, Q. Cao, and X. Zhang, "An online-traffic-prediction based route finding mechanism for smart city," *Int. J. Distrib. Sensor Netw.*, vol. 11, no. 8, 2015, Art. no. 970256.
- [26] R. Herring, A. Hofleitner, and S. Amin, "Using mobile phones to forecast arterial traffic through statistical learning," in *Proc. Conf., Transp. Res. Board*, Washington, DC, USA, Aug. 2009, pp. 1–22.
- [27] V. Cerqueira, L. Moreira-Matias, J. Khiari, and H. van Lint, "On evaluating floating car data quality for knowledge discovery," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 11, pp. 3749–3760, Nov. 2018.
- [28] B. Hoh, M. Gruteser, R. Herring, J. Ban, and D. Work, "Virtual trip lines for distributed privacy-preserving traffic monitoring," in *Proc. 6th Int. Conf. Mobile Syst., Appl., Services*, Breckenridge, CO, USA, Jun. 2008, pp. 15–28.
- [29] *Highway Capacity Manual*, Transp. Res. Board, Nat. Res. Council, Washington, DC, USA, 2010.
- [30] 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.
- [31] J. Mena-Oreja and J. Gozalvez, "A comprehensive evaluation of deep learning-based techniques for traffic prediction," *IEEE Access*, vol. 8, pp. 91188–91212, 2020.
- [32] V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proc. IEEE Int. Conf. Mach. Learn. (ICML)*, Haifa, Israel, Feb. 2010, pp. 807–814.

- [33] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," in *Proc. ICLR*, San Diego, CA, USA, May 2015, pp. 1–15. [Online]. Available: <https://arxiv.org/abs/1412.6980>.
- [34] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, and G. S. Corrado, "TensorFlow: Large-scale machine learning on heterogeneous distributed systems," 2016, *arXiv:1603.04467*. [Online]. Available: <http://arxiv.org/abs/1603.04467>
- [35] D. Krajzewicz, J. Erdmann, M. Behrisch, and L. Bieker, "Recent development and applications of SUMO-simulation of urban mobility," *Int. J. Adv. Syst. Meas.*, vol. 5, nos. 3–4, pp. 128–138, 2012.
- [36] J. J. González-Delgado, "Alicante-Murcia freeway scenario: A high accuracy and large-scale traffic simulation scenario generated using a novel traffic demand calibration method in SUMO," *IEEE Access*, to be published.
- [37] F. Rempe, P. Franek, and K. Bogenberger, "Estimating traffic speeds using probe data: A deep neural network approach," 2021, *arXiv:2104.09686*. [Online]. Available: <http://arxiv.org/abs/2104.09686>



JESÚS MENA-OREJA received the degree in industrial engineering from the Universidad Miguel Hernández de Elche (UMH), Spain, in 2015, where he is currently pursuing the Ph.D. degree, working on the use of machine learning and big data for traffic characterization and prediction under the presence of connected and automated vehicles. He was recognized by the Professional Organization of Industrial Engineers with the Best Student Award.



JAVIER GOZALVEZ (Senior Member, IEEE) received the degree in electronics engineering from the Engineering School, ENSEIRB, Bordeaux, France, and the Ph.D. degree in mobile communications from the University of Strathclyde, Glasgow, U.K. Since October 2002, he has been with the Universidad Miguel Hernández de Elche (UMH), Spain, where he is currently a Full Professor and the Director of the Uwicore Laboratory. He is an Elected Member on the Board of Governors of the IEEE Vehicular Technology Society (IEEE VTS). He was previously an IEEE Distinguished Lecturer of the IEEE VTS. From 2016 to 2017, he was the President of the IEEE VTS. He is currently the Editor-in-Chief of the *IEEE Vehicular Technology Magazine* and an IEEE Distinguished Speaker.

• • •