Received June 30, 2018, accepted July 27, 2018, date of publication August 9, 2018, date of current version August 28, 2018.

*Digital Object Identifier 10.1109/ACCESS.2018.2864157*

# Predicting Available Parking Slots on Critical and Regular Services by Exploiting a Range of Open Data

**CLAUDIO BADII, PAOLO NESI, AND IRENE PAOLI**
Distributed Systems and Internet Technologies Lab, Department of Information Engineering, University of Florence , 50139 Florence, Italy
Corresponding author: Paolo Nesi (paolo.nesi@unifi.it)

This work was supported in part by MIUR, in part by the University of Florence, and in part by companies through the Sii-Mobility National Project on Smart City Mobility and Transport under Grant SCN112.

**ABSTRACT** Looking for available parking slots has become a serious issue in contemporary urban mobility. The selection of suitable car parks could be influenced by multiple factors—e.g., the walking distance to destination, driving and waiting time, parking prices, availability, and accessibility—while the availability of unused parking slots might depend on parking location, events in the area, traffic flow, and weather conditions. This paper presents a set of metrics and techniques to predict the number of available parking slots in city garages with gates. With this aim, we have considered three different predictive techniques, while comparing different approaches. The comparison has been performed according to the data collected in a dozen of garages in the area of Florence by using Sii-Mobility National Research Project and Km4City infrastructure. The resulting solution has demonstrated that a Bayesian regularized neural network exploiting historical data, weather condition, and traffic flow data can offer a robust approach for the implementation of reliable and fast predictions of available slots in terms of flexibility and robustness to critical cases. The solution adopted in a Smart City Apps in the Florence area for sustainable mobility has been welcomed with broad appreciation or has been praised as successful.

**INDEX TERMS** Smart city, available parking lots, prediction model, parking garage, machine learning.

## I. INTRODUCTION

In our cities, the number of vehicles is getting higher and higher if compared to the development of the surrounding urban spaces; thus, the services providing available parking slots are becoming even more relevant for urban mobility management. Drivers are wasting a considerable amount of time while trying to find a vacant parking lot, especially during peak hours and in specific urban areas (e.g., hospitals, stations, parks, sport stadium). Car drivers, in dense city districts, usually spend from 3.5 to 14 minutes to look for a slot [1]; this means spending money and producing pollution, thus affecting the general society costs. Consequently, looking around for available parking spaces may depend on a peculiar number of different reasons: different travel motivations, garage proximity to final destination, price differences among garages, the lack of familiarity with the selected urban area, etc. Looking for parking slots does not only cause annoyance and frustration to drivers, but it is expected to have a significant negative impact on the efficiency of the transportation system within the urban tissue, and sustainability. To look for an available parking brings forth unnecessary traffic workload and may affect the environment negatively due to an increase of vehicle emissions. These issues are true for parking silos with gates, as well: they can be full in certain areas and time windows; while in other areas, they may become full unexpectedly and/or due to apparently unknown conditions to drivers. Since a long time now, it is possible to collect real-time parking information – i.e., capacity, garage prices, number of empty parking slots in the silos or in the area, thus being able to realize statistics predictive models. Recently, researches have discovered that big data and artificial intelligence may exploit the relevance of other data sources, such as the garage proximity, traffic flow information, and any information related to weather conditions, to calculate precise predictions more reliably.

In this paper, a solution to predict the number of available parking slots (not taken) has been analyzed as to parking garages with gates (e.g., silos, or on flat, or under station) belonging to two different types: they carry out a regular easily predictable service or they deal with strongly randomized cases (e.g., from suburbs hospital parking to parking locations accomplishing multiple services: stations, theaters, fairs). The approach has the advantage to be robust with respect to critical cases such as when the number of free slots reaches zero, or when some data are missing in the stream. The proposed prediction model has been created in the context of the national smart city Sii-Mobility research project of the Italian Ministry of Research for terrestrial mobility and transport [48]. It exploits open data and real-time data of the Km4City infrastructure located in the Florence/Tuscany areas [49] and corresponding to the current Smart City solution. The Sii-mobility project aims at defining solutions for sustainable mobility, with suggestions s to the driver of virtuous behavior, providing info mobility, etc.

The next section is dedicated to the related works to allow for a better contextualization of the research topic. After presenting the state of the art, the paper's structure and the proposed solution's description are presented.

### A. RELATED WORKS

The car parking activity by a driver is influenced by multiple factors – i.e., the walking distance to destination, driving and waiting time, parking fees, service level, parking size, safety [2], [3], parking price, availability and accessibility [4]. In particular, two important aspects in the parking decision-making process by any driver are: the number of available parking spaces (if known), and past experience in finding available lots. In fact, drivers who are aware about parking availabilities are 45% more successful in their decisions than the ones without such knowledge, when arriving to their parking facilities [5]. Parking facilities can be indoor/outdoor and public/private. In this context, *pareto-optimal routes are selected for drivers when planning trips* [6].

In more details, parking slots can be located on the **street or in parking garages with gates**. In terms of prediction models, there is a substantial difference between parking garages and street parking. In fact, in parking garages, it is very easy to count the total number of available slots by considering the tickets released at the entrance gate, and the outputs from the exits. On the other hand, as to street-parking, occupancy could be detected by means of some distributed sensor systems. For such reasons in literature there are two distinct research lines, focused on both street-parking prediction and free/available parking slots inside garages [7]. Tiexin *et al.* [7] have also proposed some integrated theoretical models for street parking predictions, taking into account the effectiveness of both solutions in a central commercial district. Moreover, in [8], identifying where people actually park on the basis of a trajectory analysis has been proposed as a solution. On the other hand, those data have to be accessible.

The **street-parking** problem in San Francisco has been tackled in [9], predicting the occupancy rate (defined as the number of occupied parking spots over the total availability) of parking lots in a given geo-located zone in a future time [9]. The solution works with aggregated parking lots, aiming at reducing errors in parking prediction according to different travel behavior along different regions. On the other hand, Chen [9] discretized the day into 24 intervals, and performed the principal component analysis, PCA, on time series to model the trend of occupancy. Thus, four different predictive approaches (Auto-Regressive Integrated Moving Average approach, Linear Regression, Support Vector Regression, and Feed Forward Neural Network) have been used to investigate the prediction errors. Comparison has shown that Feed Forward Neural Networks produced the best predictive model, presenting a Mean Absolute Percentage Error (MAPE), 1 hour ahead, of about 3.57%. *In this case, only well-defined and stationary cases have been addressed using historical data without taking into account contextual data.* Therefore, the solution works well only on regular days, which have easy predictable conditions for regular parking clients. On the same path, in [11], an unsupervised clustering approach (Neural-Gas Network [10], [11]) has been adopted on the data to identify the similar street-parking behavior over 24 hours, using a small data sample, and a temporal resolution of 15 minutes. In reality, in [10] a strong variability among the behavior of different street-parking spaces has been presented but what is clearly missing is an effective prediction model. In [12] the solution proposed in [13] (which was a method based on Wavelet Neural Network) has been improved with the aim to predict the availability of a parking lot every minute, in an interval time of 15 hours (from 6:00 AM to 10:00 PM), using a three-days training set and one day as test set. Also, in this case, the predicting precision has been in the range of 3-10% in term of Mean Square Error (MSE). The authors have declared that in critical cases (where available slots are close to zero) the prediction error rapidly increases, and the only way to reduce it is to modify the training set. *On the other hand, we would like to stress that it is precisely in critical cases when free slots are getting fewer and fewer, that precision has to be higher, so as to provide a good service for final users; thus, predictive models and services for prediction are much more needed and relevant.*

As to street parking, in [14] a two-step methodology for occupancy prediction based on sensor data has been proposed: the first step consisted in a real-time prediction scheme based on recurrent artificial neural networks; the second module estimated the probability of finding available parking space in relation to traffic volume, day type and time slot along the day. The resulting MAPE for the prediction at 30 minutes was in the range of 1-4%. Moreover, in [15] a mathematical model has been proposed and it is based on queueing theory and Markov chain to predict parking slots occupancy based on the information exchanged among vehicles, which are connected to an ad-hoc network. The obtained predictive error at 30 minutes (in terms of average deviation of the predicted occupancy) has been in the range of 8%. In the same thematic area, in [40] a distinction among different sources of data – i.e., parking data, user data, open data, has been presented, thus emphasizing the relevance of open data to provide an independent and sustainable system to search for parking spaces on street. Thus, Pflügler *et al.* [40] proposed a prediction model based on using neural network presenting an MSE of 16% *without addressing the critical situations of parking spaces with non-stationary attitudes.* In [46] the use of car parking data, pedestrian data and car traffic data has been investigated to predict available on-street car parking in 15-minute intervals in the city of Melbourne.

On the other hand, addressing the prediction of **free slots in parking garages/silos** is a completely different problem with respect to the street parking prediction. In parking lots, the number of offered slots is typically high, and clearly reported at the entrance gate of the garage, and therefore they have a strong appeal to drivers who may arrive all together. They are typically located closer to center of attraction such as commercial centers, hospitals, railway stations, theaters, and multiservice areas, where large events may rapidly overstock the structure. Therefore, the prediction of free/available parking slots in garages is not an easy task. Some of them may have a stable stationary behavior over time (due to the served facilities, such as suburbs hospitals), thus making the job prediction easier. Others may be affected by several factors, which makes the prediction of free/available slots over time much more difficult, especially during critical situation when the parking becomes crammed. Furthermore, the data related to garages are not easily available or they can be available only against payment or when one arrives at the entrance.

Therefore, according to [16], the number of available lots of a garage may depend on road traffic flow, weather events, road condition, etc., and the best prediction of available spaces is a combination of short time data and historical information. Thus, a neural network model can be used to predict the available spaces of fourteen garages in Beijing and yet the *prediction results were not reported*. The authors keep on saying that their prediction results showed problems of performance. Teodorovic and Lucic [17] and Yan *et al.* [18] have developed an ''intelligent'' parking system *without predicting the real number of free/available parking spaces*. While, in the discrete choice a model for combining online and historical data for real-time to predict the parking availability of a single garage (665 available lots on four levels) has been used in [15]. Thus, achieving an average error in prediction of 1 hour lower than 3% without addressing the critical condition, when available parking lots are close to zero. A more traditional solution was also proposed in [19], predicting availability in one Pittsburg parking garages (totaling 691 parking spaces) using historical and real-time data, by using multilinear regression model. In [48] the authors proposed a queuing model (well-established continuous-time Markov queue) to describe and predict the stochastic occupancy change of parking facility for a single garage in San Francisco, involving historical data occupancy only.

### B. ARTICLE OVERVIEW

This article is focused on presenting the research results regarding a solution to predict the number of available parking slots for each garage in the city of Florence and some areas of Tuscany for the next day (24 hours in advance), every 15 minutes. Prediction of available parking spaces is a complex non-linear process whose dynamic changes involve multiple kinds of factors. Parking facilities provide several different working conditions. Some of them are dedicated to a specific facility (football stadium, hospital), others on multipurpose (station, expo, etc.), and others on outskirts of town. Variability and performance are one of the problems to be addressed, together with the precision in critical time slots, which is when the parking is getting full, running out of available slots. The prediction model proposed has been created in the context of the Sii-Mobility national smart city research project of Italian Ministry of Research for terrestrial mobility and transport [48] and by exploiting open data and real-time data of Km4City [49] infrastructure in the Florence area, Italy for Smart City. Sii-Mobility project aimed at defining solutions for sustainable mobility, suggesting parking status to drivers 30 minutes/1 hour in advance to allow them to take a conscious decision, and maybe change their own plan (by selecting a difference parking or exploiting public transportation).

The article is organized as follows. In Section II, Sii-Mobility general architecture for data collection and analytics is described. Section III provides a description of the forecasting methods adopted to identify and validate the predictive models and framework. In Section IV, a description of the typical trends for parking services is presented. In Section V, the identified relevant metrics for the prediction are reported and described. They are related to: historical data, the traffic flow; and the weather. Section VI focusses on the comparison of the predictive models exploiting the data collected within Florence area garages, to achieve the identification of the best resulting approach in terms of prediction error and processing time. Conclusions are drawn in Section VII.

## II. OVERVIEW OF Sii-MOBILITY ARCHITECTURE

The prediction algorithm has been developed exploiting Sii-Mobility infrastructure and architecture which present a semantic data aggregation layer and service towards control rooms and web and mobile Apps via Smart City API [20]. The architecture is depicted in **Figure 1.** City Operators and Data Brokers provide accessible data. In turn such data are collected by using Extract Transform and Load (ETL) processes scheduled on the Big Data processing back office by using a Distributed Smart City Engine Scheduler (DISCES). Among the data, the Open Data has been provided by the municipality, Tuscany region (mobility monitoring center), LAMMA weather agency, ARPAT environmental agency, etc., while private data are provided by City Operators such as mobility centers. All data are cleaned, reconciled and converted in triples for the RDF store of the Knowledge Base [21], [22]. The collection of the data into an RDF graph database enables the semantic search by relationship, spatial, temporal, etc., with specific efficient reasoners using

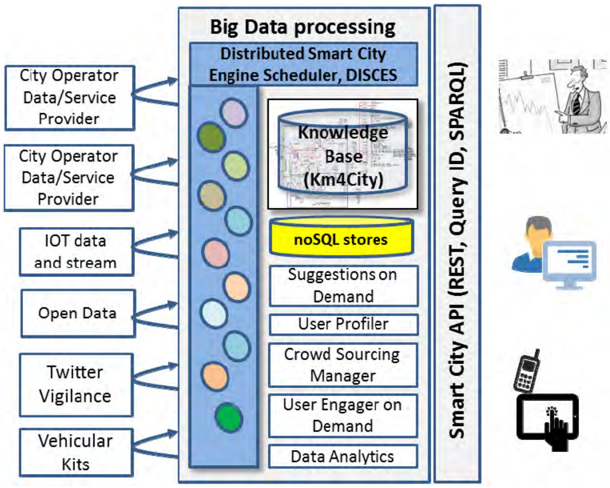



**FIGURE 1.** Sii-mobility architecture.

SPARQL queries. It can be very useful to discover which kind of entities and sensors data are available in a certain area.

The collected data include historical and real time data that can be exploited by Data Analytics algorithms to estimate recommendations, suggestions, personal assistance hints, and also origin destination matrices, prediction on traffic flow and recently prediction on free parking based on the solution presented in this paper. Most of the data modeled on Km4City store and most of the computed predictions are made accessible to mobile and web Apps via the Smart City APIs, other on Control Room Dashboard. A Smart City Dashboard must represent on the city maps and structure, real time data, and key indexes. Among them, the status and prediction of major services such as parking, Wi-Fi, traffic flow, etc.

In the context of monitoring and predicting the parking garage status, on Sii-Mobility more than 200 garages are monitored in whole Tuscany (an area of 3.5 inhabitants), and among them, about 12 are in Florence city. The status of each one of those garages is updated every 15 minutes, while the goal consists in providing in advance, namely 30 minutes and 1 hour before, guesses about the parking status for each parking garage. In this manner, car drivers will be given enough time to decide to park in different parking areas and/or to drop the idea of using a private car and decide, instead, to reach the same destination with a more sustainable solution like public transportation. The same App provides several kinds of information about parking and public transportation services in general.

## III. FORECASTING TECHNIQUES COMPARED

This section provides an overview of the techniques we have considered and compared with the aim of creating a solution to predict the number of available/free slots in parking garages. During our research different techniques have been discharged since they did not produce satisfactory results. Among possible techniques, our choice has been focused on the comparison of the most effective solutions, which are: Bayesian Regularized Artificial Neural Networks, the Support Vector Regression the Recurrent Neural Network, and the more traditional statistical approach such as Auto-Regressive Integrated Moving Average approach (e.g., ARIMA).

### A. ARTIFICIAL NEURAL NETWORKS WITH BAYESIAN REGULARIZATION

The Artificial Neural Network (ANN) is a very popular technique which relies on supervised learning. Beginning with the very first proponents, they have used ANN as powerful nonlinear regression techniques inspired by theories on how human brain works [23]–[25]. The primary application of neural networks involves the development of predictive models to forecast future values of a particular response variable from a given set of independent variables; resulting particularly useful in coping with problems showing a complex relationship between input and output variables. The outcome is modeled by an intermediary set of unobserved variables (hidden neurons), that are typically linear combinations of the original predictors. The connection among neurons in each layer is called "a link". A link is stored as a weighted value, which provides a measure of the connection between two nodes, as shown in [26] and [27]. The supervised learning step changes these weights in order to reduce the chosen error function, generally mean squared error, in order to optimize the network for use on unknown samples. ANNs tend to overfit, which means to have trained the NN to fit the noise trend, but without producing a good generalization, as expected by the ANN.

However, Bayesian Regularized ANNs (BRANNs) tries to overcome the overfitting problem by incorporating Bayes' modeling into the regularization scheme [28]. In general, the overfitting risk increases when a neural network grows in size through additional hidden layer neurons. BRANN approach avoids the overfitting because the regularization pushes unnecessary weights towards zero. The BRANN method is more robust, parsimonious, and efficient than classical ANNs, and the network weights are typically more significant in modeling the phenomena [28]. The BRANN model fits a three-layer neural network as described in [29] and [31]. The layer weights the network, which is initialized by the Nguyen-Widrow initialization method [30], and thus, the model is given by:

$$y_i = g(x_i) + e_i$$
$$y_i = \sum\nolimits_{k=1}^{s} w_k g_k \left(b_k + \sum\nolimits_{j=1}^{p} x_{ij}\beta_j^{[k]}\right) + e_i, \quad i = 1, \ldots, n \tag{1}$$

where:

- $e_i \sim N(0, \sigma_e^2)$;
- $s$ is the number of neurons;
- $w_k$ is the weight of the $k$-th neuron, $k = 1, \ldots, s$;
- $b_k$ is a bias for the $k$-th neuron, $k = 1, \ldots, s$;
- $\beta_j^{[k]}$ is the weight of the $j$-th input to the net, $j = 1, \ldots, p$;

- $g_k(\cdot)$ is the activation function: in this case

$$g_k(x) = \frac{e^{2x} - 1}{e^{2x} + 1}$$

The objective function consists of minimizing $F = \alpha E_W + \beta E_D$, where $E_W$ is the sum of squares of network parameters (weight and bias), and $E_D$ is the error (sum of squares), $\alpha$ and $\beta$ are the objective function parameters.

### B. SUPPORT VECTOR REGRESSION

The SV (Support Vector) algorithm is a nonlinear generalization of the generalized portrait algorithm developed in Russia in the sixties and further developed for decades [32]. This theory characterizes properties of learning machines which allow the generalization of unseen data, thus obtaining excellent performances in regression and time series prediction applications [42]. In the following, the Support Vector Regression model (SVR) with linear kernel has been adopted as a predictive method. The idea of SVR is based on the computation of a linear regression function $f(x) = \boldsymbol{w}^{\boldsymbol{T}}\boldsymbol{x} + b$ to a given data set $\{(\boldsymbol{x_i}, y_i)\}_{i=1}^{N}$ in a high dimensional feature space where the input data are mapped via a nonlinear function. Instead of minimizing the observed training error, SVR attempts to minimize the generalization error bound; so as to achieve generalized performance. The generalization error bound is the combination of the training error and the regularization term controlling the complexity of the hypothesis space [43].

### C. ARIMA MODELS

We have used the Auto Regressive Integrated Moving Average (ARIMA) model as an alternative forecasting method with respect to the above-mentioned techniques. The predictive model has been developed by using Box-Jenkings methodology for ARIMA modeling [33]. ARIMA model is composed of two parts: Auto-Regressive and Moving Average. The Auto-Regressive part (AR) creates the basis of the prediction and can be improved by a Moving Average (MA) modeling for errors made in previous time instants of prediction. The order of ARIMA models is defined by the parameters $(p, d, q)$: $p$ is the order of AR model; $d$ is the degree of differencing, and $q$ is the order of the MA part, respectively; and by the corresponding seasonal counterparts $(P, D, Q)$.

### D. RECURRENT NEURAL NETWORK

Neural Networks have arisen great interest for many decades, due to the desire to understand the brain, and to build learning machines. Recurrent Neural Networks (RNNs) are basically a Feedforward Neural Network with a recurrent loop [34]. They are considered a powerful model for sequential data, and they are applied to a wide variety of problems involving time sequences of events and ordered data. RNN are neural networks consisting of a hidden state $\boldsymbol{h}$ and an output $\boldsymbol{y}$ operating on a sequence of variables $\boldsymbol{x} = (x_1, \ldots, x_T)$. At each time step $t$, the hidden state of the RNN is updated by $h_{(t)} = f(h_{(t-1)}, x_t)$, where $f$ is a non-linear activation function. While in principle the recurrent network is a simple and powerful model, in practice, it is hard to train it properly [35].

## IV. DATA DESCRIPTION

As mentioned in the introduction, the main goal was to find a solution to predict the number of available parking slots (not occupied) within parking garages controlled by a gate.

The Sii-Mobility Km4City infrastructure, presented in Section II, collected the data used for the prediction during the period from January 5, 2017, to March 26, 2017. For each car park, the number of available slots has been checked and registered every 15 minutes. Therefore, our study, refers to 12 garages located in the municipality of Florence as depicted in **Figure 2**.

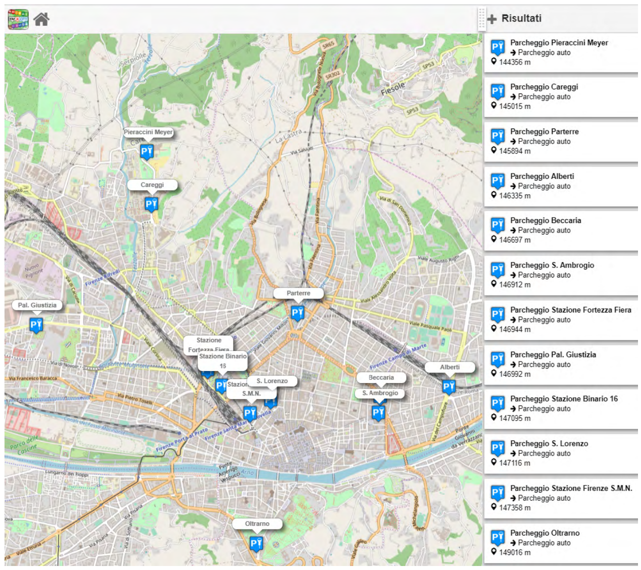



**FIGURE 2.** Map of the main 12 car parks in Florence. As depicted by "Toscana dove, cosa... Km4city" App https://www.km4city.org/webapp/

These garages are located in three main different areas of Florence: close to hospitals, downtown (near to the main touristic area) and in the outskirts. The latter are also called park and ride systems, specifically created to stimulate the usage of public transportation. They are meant to provide parking space for commuters deciding to drop their cars out of the city and switch to public transportation.

**Figure 3** reports a number of typical daily trends of available lots for some of these parking garages. According to **Figures 3**(a) and **3**(b), workdays are readily recognizable with respect to weekends (numerically higher since we have 5 working days per week). This strong difference is visible in cases (a) and (b), which are located close to hospitals: *Careggi* and *Pieraccini Meyer*, respectively. For example, in parking 2(b) close to the hospital, the available lots go close to zero from the 9:15 in the morning. While, the trend is more relaxed during week-ends, where the maximum usage of the parking (minimum number of available lots) is focused on the time slots related to visits at: lunch and dinner time.

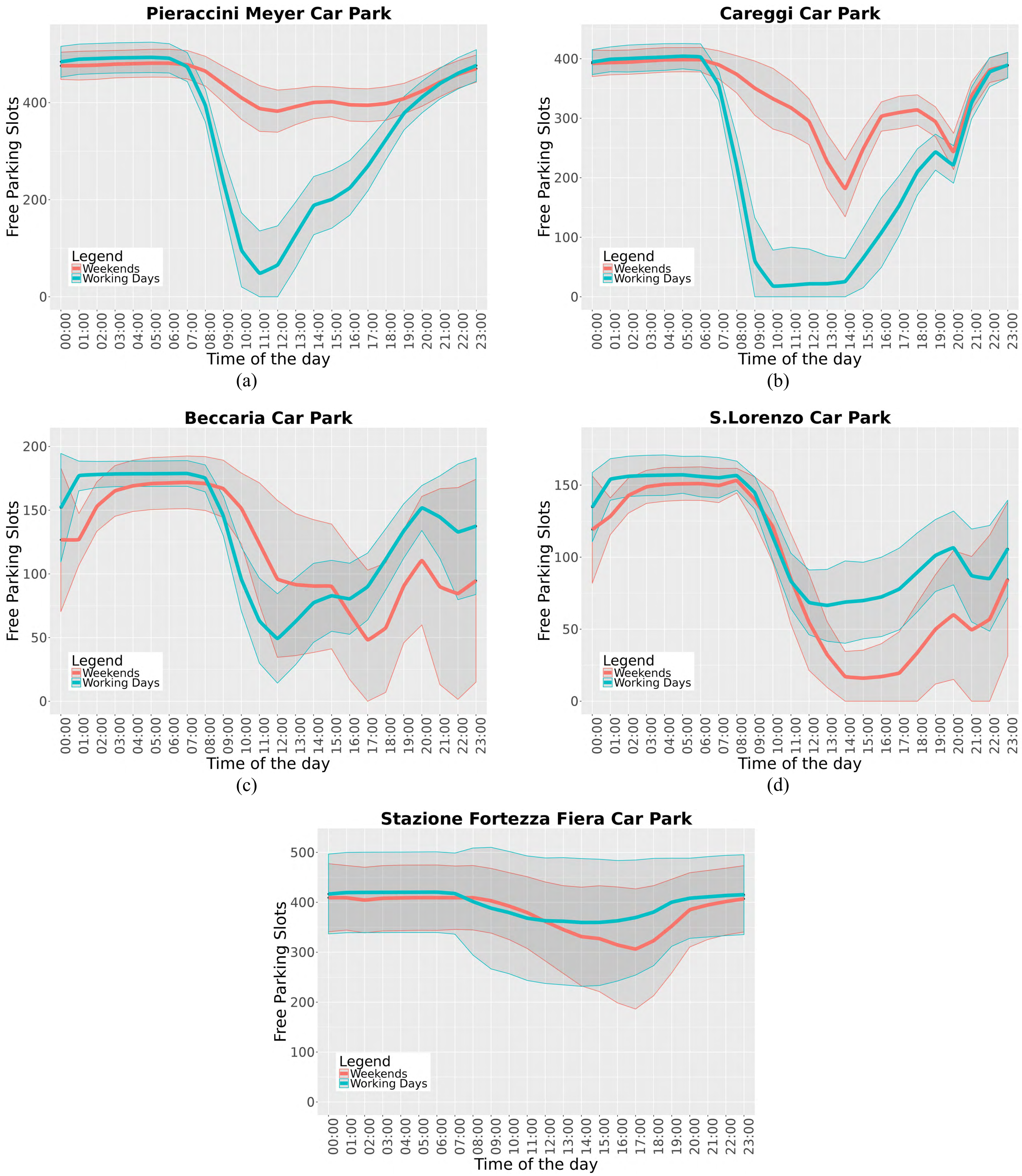



**FIGURE 3.** **Typical daily trends of free slots every 15 minutes (from 05 January 2017 to 10 April 2017) for car parks of (a) *Pieraccini Meyer*, (b) *Careggi*, (c) *Beccaria*, (d) *S. Lorenzo*, and *(e) Stazione Fortezza Fiera*.**

In the cases of the parking areas of *S. Lorenzo* and *Beccaria* (see **Figures 3**(c) and **3**(d), respectively) the number of available lots seems to be independent from workdays, and the daily trend is not repeated regularly. Observing **Figure 3**(d), it represents an area of Florence's nightlife and restaurants. It is self-evident that the parking garage is often full at lunch/dinner time and after dinner time.

A more complex situation is reported in **Figure 3**(e). This garage is located between the main railway station and the expo area of Florence. As depicted by data, the presence of a large exposition (*Pitti Fashion Expo* held from the 9th to 13th January 2017) has changed the situation. The data trend is mixed presenting a number of days with moderated usage. While, the arrival of the major event has changed radically the parking usage, since it has been used during nights all along the fair duration. And, in the day a strong saturation of the garage has been reached during the event/fair.

The above considerations become clearer by analyzing the weekly curves for *Careggi* and *S.Lorenzo* car parks (case (a) and (d) of **Figure 3**, respectively). In fact, in **Figure 4**(a), the different trends registered for working day and weekend and for the hospital parking is clear. Similarly, the same difference is registered in **Figure4**(b). Please note that, n both cases, it is possible to observe that Epiphany vacation on 6th January, created a trend similar to that related to weekends.

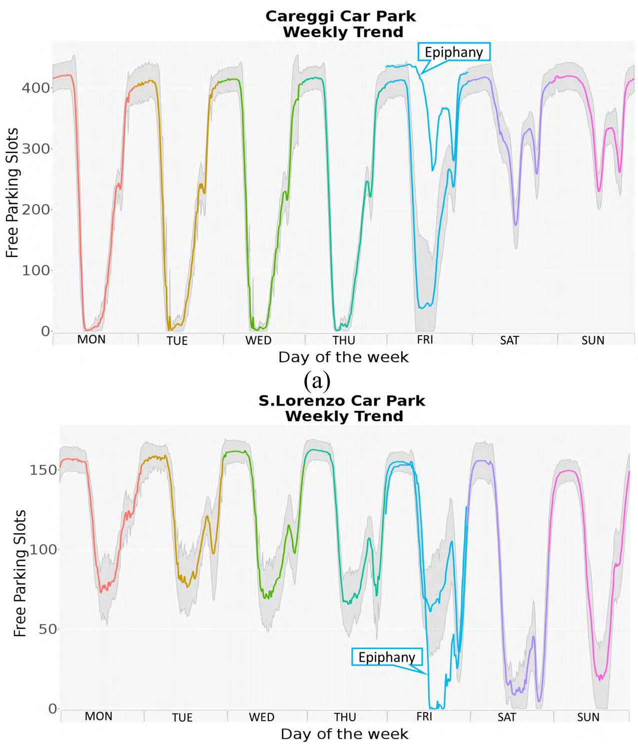



**FIGURE 4. Weekly curves of free parking slots every 15 minutes (from 05 January 2017 to 26 March 2017) for (a) *Careggi* car park and (b) *S.Lorenzo* car park.**

According to [12], it is reasonable to think that changes in patterns between workdays and weekends can be due to different travel purposes: people usually traveling for work reasons on workdays, and for entertainment on weekends [12]. However, when considering that *Careggi* car park is close to the hospital, noteworthy is that car parks have a different trend in relation with the time windows hospitals set up to allow patient visits.

In both daily and weekly curves (see Figures 3 and 4), it is possible to better understand the critical conditions of a garage, i.e., when the available parking slots become close to zero. That is the situation when drivers have to be alarmed in advance giving as more precise prediction as possible. The ability of the proposed algorithm to predict when the garage is becoming complete with a significant precision (a small prediction error), while handling missing data, can be defined as robustness.

## V. FEATURES/METRICS DEFINITION

According to the above presented state of the art, there is a substantial difference between a parking garage and a street-parking in terms of distribution of free spaces in the parking area. In the context of street-parking, it may be necessary to make a clustering to understand the free space distribution of an area; thus, aggregating the street-parking areas with the same behavior. On the opposite end, taking into account garages, the trend of the available slots is very peculiar of the specific contextual conditions of each garage (as depicted in **Figure 3**). For this reason, the adoption of clustering approach is not successfully. Moreover, one of the difficulties experienced in identifying a common predictive and precise model for all parking garages, was due to the fact that different parking garages have different behaviors in different days of the week, and period of the day. Some of them may experience critical condition when the available parking slots are close to zero and this is the moment when drivers have to be alerted in advance.

According to the above considerations, before evaluating the predictive capabilities of forecasting techniques mentioned in Section III, three groups of features have been identified as possible predictive metrics and are briefly discussed. They have been reported in **Table 1**.

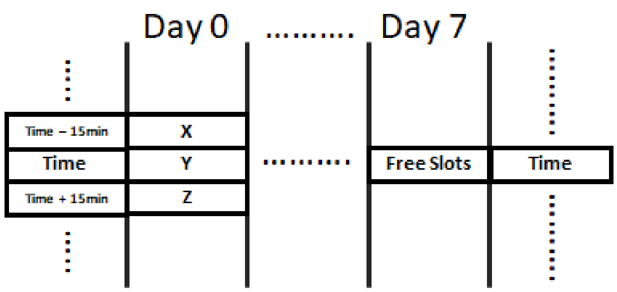
The potential metrics at the basis of the predictive models are discussed in the following beginning with the category they belong to.

Features belonging to the **Baseline category** refer to measures related to the direct statistical observation of garage data and derived information. To this category belong the date and time when measures are taken, working day or not, number of available slots, etc. All the values are recorded every 15 minutes. These variables are used to consider the seasonality of the data which may have different trends – i.e., working days with respect to weekends, etc. When the car parks have the same trend during the same day and time between different weeks, two other features have been included in the model:

- POD: the difference between the actual and previous number of available space at the same time, recorded one week before;
- SOD: the difference between the actual number of parking spaces and the next one at the same time, recorded one week before.

**TABLE 1.** Overview of Features that can be used to describe the context of parking usage with their: category, features and description.

| Category | Features | Description of features variable |
|---|---|---|
| Baseline features of free slot data | Free parking slots | Real number of available slots recorded every 15 minutes |
| | Time | Hours and minutes |
| | Month | Month of the year (1-12) |
| | Day | Day of the month (1-31) |
| | Day week | Day of the week (0-6) |
| | Weekend | 0 for working days, 1 else |
| | Previous observation's difference (POD) | Difference between the number of free spaces at time $i$ and number of free spaces at time ($i - 15$ minutes) recorded in the previous week |
| | Subsequent observation's difference (SOD) | Difference between the number of free spaces at time $i$, and the number of free spaces at time ($i + 15$ minutes) recorded in the previous week |
| Weather features | Temperature | City temperature measured one hour earlier than Time (°C) |
| | Humidity | City humidity measured one hour earlier than Time (%) |
| | Rainfall | City rainfall measured one hour earlier than Time (mm) |
| Traffic Sensors features | Average Vehicle Speed | Average speed of vehicles on the road being closest to the parking, over one-hour period (km/h) |
| | Vehicle Flow | Number of vehicles passing by closest to the parking, over one-hour period |
| | Average Vehicle Time | Average of distance between vehicles, over one-hour period |
| | Vehicle Concentration | Number of vehicles per kilometer, over one-hour period |




$$POD_{Day\,7,Time} = X_{Day\,0} - Y_{Day\,0}$$
$$SOD_{Day\,7,Time} = Z_{Day\,0} - Y_{Day\,0}$$

**FIGURE 5.** Construction of *Previous observation's difference (POD)* and *Subsequent observation's difference (SOD)* features described in Table 1.

Please see **Figure 5**, at a specific observation of a specific date and time corresponds the *POD* and *SOD* of the previous week.

Features belonging to the **Weather** are also collected every 15 minutes (i.e., temperature, humidity and rainfall). According to our analysis, the significant values are those related to the hour before any parking time. Therefore, in order to predict the number of available spaces in a garage at 3 pm, the weather features at 2 pm are relevant. In fact, the weather conditions typically influence the decisions on using the car or the public transportation. For example, the expected behavior of citizens when it rains, is to drop the motorcycle and drive a car. By doing so, more parking lots will be taken. On this line, you would suppose to exploit long term weather forecast (6 hours or days in advance) since they could also influence decisions (weather forecasts are accessible on the Km4City smart city). On the contrary, according to our experiments, the weather forecast features are less significant with respect to the real weather features, and thus have been not reported in the table of relevant features.

In **Table 1**, the features and data belonging to **Traffic Sensors** refer to the values of traffic recorded by the sensors which are located nearby the garage, and mainly on the streets leading to the garage (the distance of influence depends on the density of the city; in Florence case, over 400 meter they are marginally influencing the prediction). These traffic sensors' values are relevant if available for the previous hour with respect to the time of prediction. As described in Table 1, typical values are related to vehicle flow, concentration, average time and average speed. They are estimated every 15 minutes. The metrics adopted for traffic flow estimation are typically the ones accessible from city traffic flow sensors. In this context, the value of traffic flow is used for assessing the traffic conditions, and thus the average values are satisfactory. On the other hand, as to other applications, such as routing path finding, more precise data and predictions should be used [36]–[38].

The traffic sensors which are relevant for each garage may be one or more and they should be chosen taking into account the direction of travel and the most likely route leading to the garage. Traffic sensors are also used as detectors to identify the occurrence of relevant events such as those of Figure 3(e), even if unexpected. An option could be to perform a specific solution able to take into account any planned event. On Sii-Mobility, also the list of the city major events and their GPS coordinates is available. This approach fails to address precise predictions in the event of unplanned occurrences.

## VI. RESULTS AND DISCUSSION

According to the data and considerations reported, the identified challenge was to create a model and tools to predict the number of available parking slots in the garages with a resolution of 15 minutes for the next 24 hours.

As a training data set we have selected a sample of three months, from January 5, 2017, to the day before the one when observations were carried out in the analyzed test set. The test set is made of 96 daily observations (every 15 minutes) recorded during the weeks from March 27th (Monday) to April 2nd (Sunday), i.e., seven test sets were considered to calculate the error on the one-day prediction avoiding noise.

Note that, in the real-time application the model has been trained once a day, providing predictive models with 96 values. Predictions are displayed on the Km4City applications in Florence and whole Tuscany region, for one hour in advance every 15 minutes.

### A. ERROR MEASUREMENT DEFINITION

In the literature, most researchers have adopted the MAPE or MSE in order to calculate the prediction error. The identification of the model for measuring the error is very relevant, since it has to work well, even when close to zero. This is related to the particular issue of street-parking predictions where critical cases occur when the available parking slots are close to zero. Measures based on percentage errors (e.g., MAPE) have the disadvantage of becoming infinite or undefined when the observed value is equal to zero. However, as to garage parking prediction, the possibility of reaching zero lots is part of the problem as depicted in **Figure 3.** They are full every day for several hours (that is, the feature recording the number of available parking lots assumes values very often close to zero). For this reason, we have chosen the Mean Absolute Scaled Error (MASE) by Hyndman and Koehler [45]. The Mean Absolute Scaled Error is calculated as follows

$$MASE = mean\left(|q_t|\right), \quad t = 1, \ldots, n \tag{2}$$

and

$$q_t = \frac{obs_t - pred_t}{\frac{1}{n-1}\sum_{i=2}^{n}|obs_i - obs_{i-1}|} \tag{3}$$

where:

- $obs_t$ = observation at time $t$
- $pred_t$ = prediction at time $t$
- $n$ is the number of the values predicted over all test sets (96 daily observations per 7 days).

Note that, MASE is clearly independent on the scale of the data. When MASE is used to compare predictive models, the best model is the one presenting the smaller MASE.

MASE can be used as measure to define the robustness of the proposed approach. In this case, robustness means the ability of an algorithm to produce quite reliable results in the event of critical cases (e.g., when the number of free parking lots is zero, and/or in the event of missing data in the stream of observations. For this reason, apart from MASE daily prediction, the MASE related to night, morning, afternoon and evening have been calculated.

### B. KALMAN FILTER IMPUTATION OF MISSING DATA

One of the main problems related to the robustness of a possible approach to predict the number of free slots, lies in its capability of producing good results in critical conditions – – e.g., when slots are close to zero and/or when the data stream of observations is not providing every data continuously. In most predictive algorithms the lack of some observations could become a problem to produce good results in terms of MASE: for example, if the data related to the traffic volume within the selected park area are missing, the prediction error could become higher, as it is based only on weather data and historic data. To overcome this problem, a Kalman filter has been used for the imputation of missing data in real time. This solution, together with the capability of the model to be precise in terms of Mean Absolute Scaled Error (MASE), has turned out to be robust especially when slots are close to zero.

The data from traffic sensors are the most prone to missing data: this can be due to sporadic and discontinuous malfunctions of sensors or network connection. To avoid any consequent prediction error increase, data have been imputed through the Kalman filter approach. The system can identify not only the sporadic faults of data, but also the faults which need to be recovered only with extraordinary maintenance. In this case, the algorithm is able to provide the guess by using the historical data, weather data and the remaining real time data. In a regular situation, missing data are about 5% of the entire training set.

### C. PREDICTION MODELS RESULT

In the general framework, four different approaches were tested – i.e., BRANN, SVR, RNN and ARIMA model – applied on the features presented above. In detail, the number of input neurons in BRANN model corresponds to the number of the features reported in **Table 1**. Note that, all features are considered as an individual neuron, except *Time* which has 96 neurons, one for each slot of 15m (“00:00”, “00:15”,. . ., “23:45”), while a single output neuron represents the predicted value. The model fits a three-layer neural network with three intermediate neurons – i.e., the number of neurons corresponding to the lowest error rate [44].

The processing time comparison, among the models considered above, is also relevant and it is reported in **Table 3** for each parking garage. **Table 2 shows** that all the approaches can produce predictions every hour for the next hour in a quite small average estimation time. On one hand, in order to produce satisfactory predictions, the ARIMA approach needs to re-compute the training every hour. This is a quite expensive cost of about 9s for each car park. On the other hand, BRANN, SVR and RNN allow their being “trained” once a day, providing predictive models with 96 values in advance with quite precise results. For this reason, the ARIMA solution has been discharged as performed by other researchers in the literature, as reported in Section 1. Note that, the identified ARIMA was $(5, 1, 2) \times (1, 0, 1)$ and allowed to perform short-term predictions with a MASE of about 1.2.

Our aim was, not only to find a satisfactory solution to make predictions computationally viable and able to suit for several cases, but also to produce satisfactory results in terms of precision in the context of the critical cases discussed before.

As a further step, the comparison has been focused by considering BRANN, SVR and RNN on the whole set of car parks in Florence. As a result, **Table 3** reports the predictive capabilities obtained for reference cases of **Figure 2**. **Table 3** reports the comparison in terms of MASE over the predicted week, and a specific MASE estimated for morning, afternoon, evening and night, for each of the predicted numbers of free parking lots. The comparison of the predictive models has been estimated on a training period of 3 months, considering only the features belonging to the baseline category. MASE

**TABLE 2.** **Comparison among model processing time in training and estimation for a single garage.**

| Training | Forecasting Techniques | | | |
|---|---|---|---|---|
| | BRANN | SVR | RNN | ARIMA |
| Average Training processing time (sec) | 76.3 | 9.1 | 598.7 | 9.2 |
| Re-Training frequency | Daily | Daily | Daily | Hourly |
| Training period | 3 months | 3 months | 3 months | 3 months |
| **Estimation** | **BRANN** | **SVR** | **RNN** | **ARIMA** |
| Average Estimation time (sec) | 0.0031 | 0.0052 | 0.034 | 0.0015 |
| Estimation frequency | Hourly | Hourly | Hourly | Hourly |
| Estimation predicted period | 1 hour | 1 hour | 1 hour | 1 hour |

**TABLE 3.** **Comparison among predictive models using the features belonging to the baseline category. Darker cells are those showing better values.**

| Comparison Error | Forecasting Techniques | | |
|---|---|---|---|
| | BRANN | SVR | RNN |
| ***Careggi car park*** | | | |
| MASE Night | 34.85 | 16.29 | 20.01 |
| MASE Morning | 0.76 | 1.42 | 2.82 |
| MASE Afternoon | 1.89 | 4.34 | 3.66 |
| MASE Evening | 1.99 | 1.51 | 2.33 |
| MASE | 1.87 | 2.34 | 3.16 |
| ***Pieraccini Meyer car park*** | | | |
| MASE Night | 6.08 | 12.83 | 10.03 |
| MASE Morning | 0.86 | 1.27 | 4.90 |
| MASE Afternoon | 1.87 | 2.91 | 6.75 |
| MASE Evening | 1.36 | 1.57 | 10.23 |
| MASE | 1.37 | 2.06 | 6.67 |
| ***S. Lorenzo car park*** | | | |
| MASE Night | 10.33 | 11.81 | 18.34 |
| MASE Morning | 2.13 | 1.91 | 3.93 |
| MASE Afternoon | 2.70 | 3.15 | 2.37 |
| MASE Evening | 2.15 | 3.09 | 3.82 |
| MASE | 2.72 | 3.21 | 4.19 |
| ***Beccaria car park*** | | | |
| MASE Night | 9.32 | 7.80 | 12.47 |
| MASE Morning | 0.95 | 1.25 | 4.87 |
| MASE Afternoon | 2.49 | 2.14 | 2.45 |
| MASE Evening | 2.96 | 4.75 | 5.91 |
| MASE | 2.13 | 2.67 | 4.85 |

has been estimated on a testing period of 1 week after the 27th of March for (a) *Careggi*, (b) *Pieraccini Meyer*, (c) *S.Lorenzo*, and (d) *Beccaria* car parks. This comparison has highlighted that BRANN approach achieved the most reliable results, especially in critical time slots, where the car parking garages risk being full. This fact is also highlighted by the best MASE for BRANN in all reference cases.

An additional analysis has been performed in order to identify the set of combination of feature categories expected to produce the best predictions (see **Table 4**). The combinations of features have considered: baseline features; baseline and weather features; baseline and traffic sensors; baseline, weather, and traffic sensors features together. The comparison has been performed by both using the BRANN model which turned out to be the one better ranked and estimating R-squared, RMSE, and MASE. As it can be observed from **Table 4)**, the differences among cases are not very relevant. Results suggest that the best choice in terms of precision is still to use a model exploiting the baseline only. However, extending the assessment to all parking garages results are substantially different as discussed in the sequel.

**TABLE 4.** **The results of BRNN model training in terms of R-squared, RMSE and the estimated prediction error MASE for (a) *Careggi*, (b) *Beccaria* car parks.**

| Careggi car park – Table 4(a) | | | |
|---|---|---|---|
| **Model features** | **BRANN model results** | | |
| | **R-squared** | **RMSE** | **MASE** |
| Baseline | 0.974 | 24 | 1.87 |
| Baseline + Weather | 0.975 | 24 | 1.75 |
| Baseline + Traffic sensors | 0.975 | 24 | 2.04 |
| Baseline + Weather + Traffic sensors | 0.975 | 24 | 1.87 |

| Beccaria car park– Table 4(b) | | | |
|---|---|---|---|
| **Model features** | **BRANN model results** | | |
| | **R-squared** | **RMSE** | **MASE** |
| Baseline | 0.888 | 16 | 2.13 |
| Baseline + Weather | 0.890 | 15 | 2.15 |
| Baseline + Traffic sensors | 0.892 | 16 | 2.24 |
| Baseline + Weather + Traffic sensors | 0.895 | 16 | 2.33 |

In fact, while extending the assessment to all parking garages, **Figure 6** reports the comparison of the 4 models as

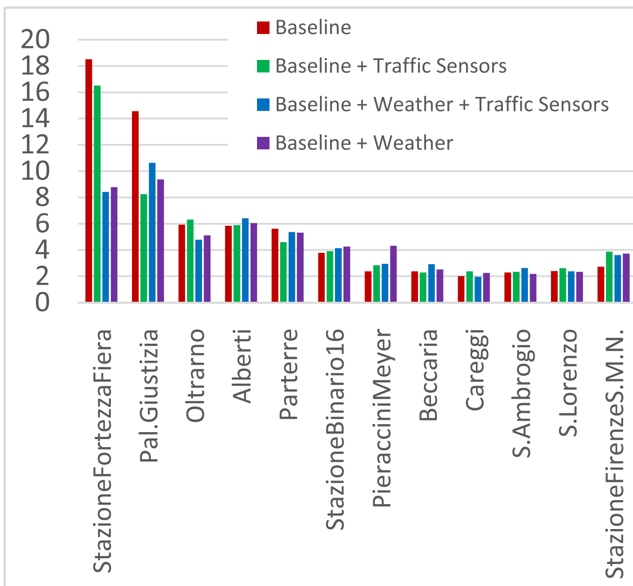
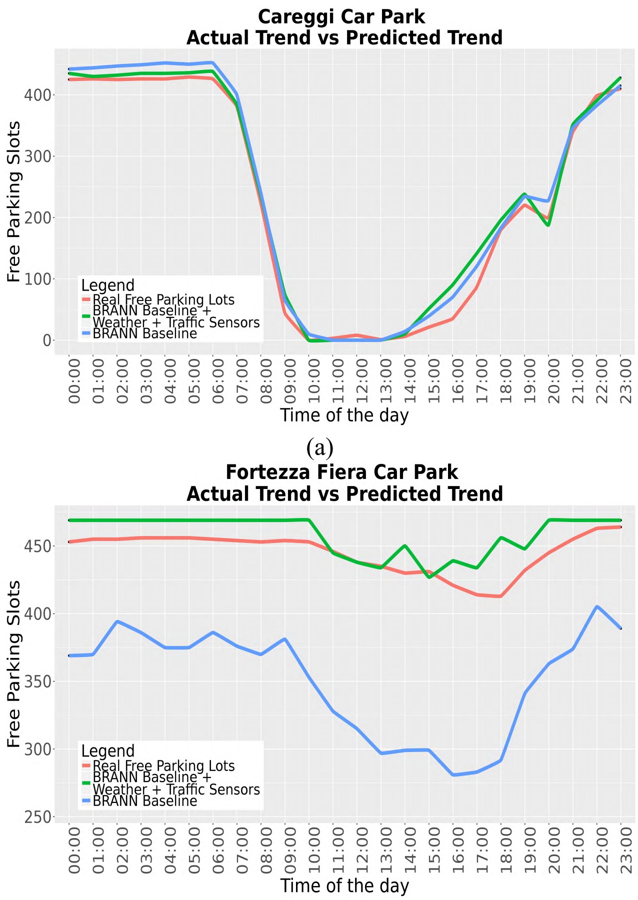



**FIGURE 6.** **Comparison of the predictive models applied to the 12 garages in Florence in terms of MASE assessed in the last week of predictions.**

compared in **Table 4** but related to all garages according to the estimation of MASE for the last week. The comparison stresses that in the cases when the daily trend of available slots:

- is regular (such as cases (a) and (b) of **Figure 3**, *Careggi* or *Pieraccini Meyer*), the 4 models of Table 4 are not so much different in terms of result quality.
- presents non-stationary critical conditions (such as Case (e) of **Figure 3**, *Stazione Fortezza Fiera, Palazzo di Giustizia, and other as Parterre*), the best model turned out to be the one considering both weather and traffic sensor features together with baseline.




(a)

(b)

**FIGURE 7.** **Comparison between the actual trend of free parking lots and the predicted trend according to BRANN using baseline features and all features for (a) *Careggi*, (b) *Stazione Fortezza Fiera* car parks along 24 hours.**

For example, **Figure 7** presents the typical comparison of the real daily trend with respect to the prediction using: (i) baseline features only, (iii) the combination of baseline, weather and traffic sensors features, for *Careggi* and *Stazione Fortezza Fiera* car parks. Noteworthy is that the addition of weather and traffic sensor features decreases the mean difference between the real values and the predictions in the *Stazione Fortezza Fiera* car park.

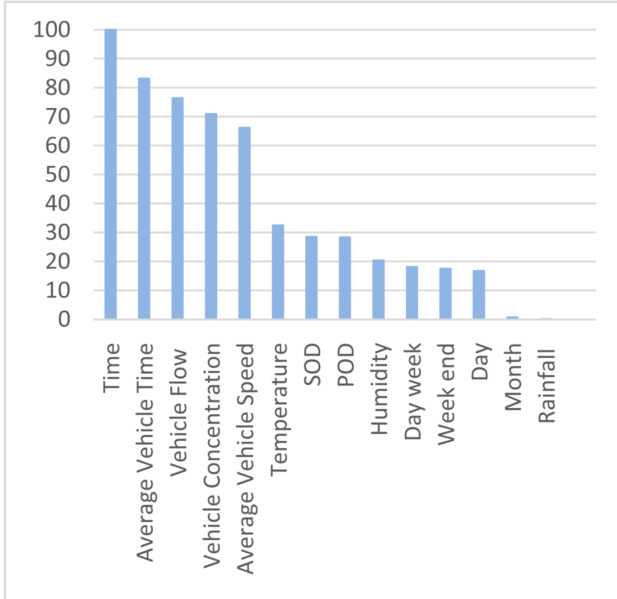



**FIGURE 8.** **Variables Importance of the BRANN full model.**

To highlight the above presented results, **Figure 8** reports the analysis of importance for the features listed in **Table 1.** They are listed in order of relevance for the BRANN full model prediction – i.e., the model with all the categories of covariates (the relevance assesses the relationship between each predictor and the outcome is evaluated). In particular, the importance of each predictor is evaluated individually: during the BRNN model training, a LOESS [39] smoother, (i.e., a nonparametric method for regression estimation) is fitted between the outcome and the predictor. To obtain a relative measure of variable importance, the $R^2$ statistic is calculated for the model containing the considered variables against the null model (intercept only). The resulting histogram depicts that variable *Time* (of the baseline) is the most relevant to predict the number of free slots for all garages. The second in terms of relevance turned out to be *Average Vehicle Time* of the traffic sensors features. According to these results, traffic variables are of primary importance, as already mentioned in [16]. These statements seem to be quite coherent with the finding of [40] for street parking. On the other hand, in making predictions for garages (as in our case) it is easier to choose traffic sensors related to the car park under investigation – i.e., only the sensors on the streets leading from the path to the garage. Whereas, in street-parking prediction, only the general traffic situation may be of interest. The selection of suitable sensors can be performed only in the cases where data are publicly available, as emphasized in [40].

As a general consideration, solutions to predict available slots in garages in current state of the art, are based only on baseline and stationary conditions [18], [41]. In our case, we have demonstrated that exploiting classic historical garage parking data together with traffic and weather features has produced better predictions.

Finally, our results cannot be directly comparable in terms of prediction errors, because we have analyzed the precision in the event of critical cases. These conditions have not yet been addressed in literature before. Please note that when the parking slots are close to zero, measures based on MAPE and MSE have the disadvantage of being infinite or undefined. For this reason, MASE was the best choice, resulting to be 1.75 in the best case.

## VII. CONCLUSIONS

Looking for available parking slots is a serious issue in today urban sustainable mobility. The solution can be to provide suggestions to drivers about the parking availability. Suggestions should reach drivers 30 minutes and 1 hour in advance (thus *producing a precise time stamp of which time they refer to*) to allow their conscious decision-making process. To this end, reliable prediction models are needed.

Prediction of available parking spaces is a complex non-linear process involving multiple kinds of factors, as the variety of parking area (downtown, nearby hospital and others on the outskirts, close to theaters, airports, etc.). In fact, a critical factor is the different trend of each garage: provided the aim is to cover a higher number of garages, the precision of the prediction is relevant, especially in critical cases (full garage)). To this aim, we have considered many techniques and provided satisfactory precision on the resulting solution. The Bayesian Regularized Neural Network has proved to be the best solution in terms of precision. In the model we have considered several metrics and features, such as the historical data, the weather conditions and the traffic flow data. In almost all predictive models, the historical data, traffic flow sensors and weather data have demonstrated high predictive capabilities in explaining the number of free parking slots. In parking garages without a recurrent daily trend of available slots, traffic sensors and weather covariates have improved the precision in predicting. The entire approach can be considered flexible, robust to critical cases and robust to sporadic lack of data. The research documented in this paper has demonstrated that a Bayesian Regularized Neural Network exploiting historical data, weather condition and traffic flow can be a robust approach for reliable and fast estimation of available slots predictions. The predictive model can produce predictions 24 hours in advance, while they are provided on mobile applications, 30 minutes, 1 hour in advance directly, and if requested also a day in advance as possible general trend. The proposed prediction model has been created by exploiting open data in the context of Sii-Mobility (national smart city research project of Italian Ministry of Research for terrestrial mobility and transport [48]), exploiting Km4City infrastructure [49] in the Florence area, Italy. The solution has been deployed as an additional feature on Smart City Apps in the Tuscany and Florence area to encourage sustainable mobility. Most computations were conducted in R Statistical Environment [50] by using different R libraries. The code to facilitate the replication of the experiment is accessible from https://www.km4city.org/parkpred/CarParkPredictions.zip

## ACKNOWLEDGEMENTS

The authors would like to thank the MIUR, the University of Florence and companies. Km4City is an open technology and research of DISIT Lab. Sii-Mobility is built on and has contributed to Km4City open solution and infrastructure.

## REFERENCES

[1] D. C. Shoup, "Cruising for parking," *Transp. Policy*, vol. 13, no. 6, pp. 479–486, 2006.
[2] S. An, B. Han, and J. Wang, "Study of the mode of real-time and dynamic parking guidance and information systems based on fuzzy clustering analysis," in *Proc. Int. Conf. Mach. Learn. Cybern.*, Aug. 2004, pp. 2790–2794.
[3] W. H. K. Lam, Z.-C. Li, H.-J. Huang, and S. C. Wong, "Modeling time-dependent travel choice problems in road networks with multiple user classes and multiple parking facilities," *Transp. Res. B, Methodol.*, vol. 40, no. 5, pp. 368–395, Jun. 2006.
[4] Z. Qian and R. Rajagopal, "Optimal parking pricing in general networks with provision of occupancy information," *Procedia-Social Behav. Sci.*, vol. 80, pp. 779–805, Jun. 2013.
[5] F. Caicedo, "The use of space availability information in 'PARC' systems to reduce search times in parking facilities," *Transp. Res. C, Emerg. Technol.*, vol. 17, no. 1, pp. 56–68, Feb. 2009.
[6] B. Yang, C. Guo, C. S. Jensen, M. Kaul, and S. Shang, "Stochastic skyline route planning under time-varying uncertainty," in *Proc. IEEE 30th Int. Conf. Data Eng.*, Mar./Apr. 2014, pp. 136–147.
[7] C. Tiexin, T. Miaomiao, and M. Ze, "The model of parking demand forecast for the urban CCD," *Energy Procedia*, vol. 16, pp. 1393–1400, 2012.
[8] B. Yang, N. Fantini, and C. S. Jensen, "iPark," in *Proc. 16th Int. Conf. Extending Database Technol. (EDBT)*, Mar. 2013, pp. 705–708.
[9] X. Chen, "Parking occupancy prediction and pattern analysis," Dept. Comput. Sci., Stanford Univ., Stanford, CA, USA, Tech. Rep. CS229-2014, 2014.
[10] T. M. Martinetz, S. G. Berkovich, and K. J. Schulten, "'Neural-gas' network for vector quantization and its application to time-series prediction," *IEEE Trans. Neural Netw.*, vol. 4, no. 4, pp. 558–569, Jul. 1993.
[11] T. Tiedemann, T. Voegele, M. M. Krell, J. H. Metzen, and F. Kirchner, "Concept of a data thread based parking space occupancy prediction in a Berlin pilot region," in *Proc. AAAI Workshop, AI Transp.*, 2015, pp. 1–6.
[12] Y. Ji, D. Tang, P. Blythe, W. Guo, and W. Wang, "Short-term forecasting of available parking space using wavelet neural network model," *IET Intell. Transp. Syst.*, vol. 9, no. 2, pp. 202–209, 2015.
[13] W.-X. Zhu and E.-X. Chi, "Analysis of generalized optimal current lattice model for traffic flow," *Int. J. Modern Phys. C*, vol. 19, no. 5, pp. 727–739, 2008.
[14] E. Vlahogianni, K. Kepaptsoglou, V. Tsetsos, and M. Karlaftis, "A real-time parking prediction system for smart cities," *J. Intell. Transp. Syst.*, vol. 20, no. 2, pp. 192–204, 2016.
[15] M. Caliskan, A. Barthels, B. Scheuermann, and M. Mauve, "Predicting parking lot occupancy in vehicular ad hoc networks," in *Proc. IEEE 65th Veh. Technol. Conf. (VTC7-Spring)*, Apr. 2007, pp. 277–281.
[16] Z. Yang, H. Liu, and X. Wang, "The research on the key technologies for improving efficiency of parking guidance system," in *Proc. IEEE Int. Conf. Intell. Transp. Syst.*, Oct. 2003, pp. 1177–1182.
[17] D. Teodorović and P. Lučić, "Intelligent parking systems," *Eur. J. Oper. Res.*, vol. 175, no. 3, pp. 1666–1681, 2006.
[18] G. Yan, W. Yang, D. B. Rawat, and S. Olariu, "SmartParking: A secure and intelligent parking system," *IEEE Intell. Transp. Syst. Mag.*, vol. 3, no. 1, pp. 18–30, Jan. 2011.
[19] T. Fabusuyi, R. C. Hampshire, V. Hill, and K. Sasanuma, "A predictive model and evaluation framework for smart parking: The case of ParkPGH," in *Proc. 18th ITS World Congr.*, Orlando, FL, USA, Oct. 2011, pp. 16–20.
[20] P. Nesi, C. Badii, P. Bellini, D. Cenni, G. Martelli, and M. Paolucci, "Km4City smart city API: An integrated support for mobility services," in *Proc. IEEE Int. Conf. Smart Comput. (SMARTCOMP)*, May 2016, pp. 1–8.
[21] C. Badii, P. Bellini, D. Cenni, A. Difino, P. Nesi, and M. Paolucci, "Analysis and assessment of a knowledge based smart city architecture providing service APIs," *Future Gener. Comput. Syst.*, vol. 75, pp. 14–29, Oct. 2017.
[22] P. Bellini, M. Benigni, R. Billero, P. Nesi, and N. Rauch, "Km4City ontology building vs data harvesting and cleaning for smart-city services," *J. Vis. Lang. Comput.*, vol. 25, no. 6, pp. 827–839, Dec. 2014.

[23] C. M. Bishop, *Neural Networks for Pattern Recognition*. London, U.K.: Oxford Univ. Press, 1995.
[24] B. D. Ripley, "Statistical ideas for selecting network architectures," in *Neural Networks: Artificial Intelligence and Industrial Applications*. London, U.K.: Springer, 1995, pp. 183–190.
[25] M. Titterington, "Neural networks," *Wiley Interdiscipl. Rev., Comput. Statist.*, vol. 2, no. 1, pp. 1–8, 2010.
[26] G. D. Garson, "Interpreting neural-network connection weights," *AI Expert*, vol. 6, no. 4, pp. 46–51, Apr. 1991.
[27] A. T. C. Goh, "Back-propagation neural networks for modeling complex systems," *Artif. Intell. Eng.*, vol. 9, no. 3, pp. 143–151, 1995.
[28] F. R. Burden and D. A. Winkler, "Robust QSAR models using Bayesian regularized neural networks," *J. Med. Chem.*, vol. 42, no. 16, pp. 3183–3187, 1999.
[29] D. J. C. MacKay, "A practical Bayesian framework for backpropagation networks," *Neural Comput.*, vol. 4, no. 3, pp. 448–472, May 1992.
[30] D. Nguyen and B. Widrow, "Improving the learning speed of 2-layer neural networks by choosing initial values of the adaptive weights," in *Proc. IJCNN Int. Joint Conf. Neural Netw.*, Jun. 1990, pp. 21–26.
[31] F. D. Foresee and M. T. Hagan, "Gauss-Newton approximation to Bayesian learning," in *Proc. Int. Conf. Neural Netw. (ICNN)*, 1997, pp. 1930–1935.
[32] V. Vapnik, "The support vector method of function estimation," in *Nonlinear Modeling*. Boston, MA, USA: Springer, 1998, pp. 55–85.
[33] G. E. Box, G. M. Jenkins, and G. C. Reinsel, "Forecasting," *Time Ser. Anal.*, vol. 12, pp. 137–191, Jun. 2008.
[34] J. Connor and L. Atlas, "Recurrent neural networks and time series prediction," in *Proc. IJCNN-Seattle Int. Joint Conf. Neural Netw.*, Jul. 1991, pp. 301–306.
[35] R. Pascanu, T. Mikolov, and Y. Bengio, "On the difficulty of training recurrent neural networks," in *Proc. ICML*, vol. 28, 2013, pp. 1310–1318.
[36] B. Yang, J. Dai, C. Guo, C. S. Jensen, and J. Hu, "*PACE*: A *PAth-CE*ntric paradigm for stochastic path finding," *VLDB J.*, vol. 27, no. 2, pp. 153–178, Apr. 2017.
[37] J. Hu, B. Yang, C. Guo, and C. S. Jensen, "Risk-aware path selection with time-varying, uncertain travel costs: A time series approach," *VLDB J.*, vol. 27, no. 2, pp. 179–200, Apr. 2018.
[38] J. Dai, B. Yang, C. Guo, C. S. Jensen, and J. Hu, "Path cost distribution estimation using trajectory data," *VLDB Endowment*, vol. 10, no. 3, pp. 85–96, Jan. 2016.
[39] W. S. Cleveland, S. J. Devlin, and E. Grosse, "Regression by local fitting: Methods, properties, and computational algorithms," *J. Econometrics*, vol. 37, no. 1, pp. 87–114, 1988.
[40] C. Pflügler, T. Köhn, M. Schreieck, M. Wiesche, and H. Krcmar, "Predicting the availability of parking spaces with publicly available data," in *Proc. GI-Jahrestagung*, 2016, pp. 361–374.
[41] F. Caicedo, C. Blazquez, and P. Miranda, "Prediction of parking space availability in real time," *Expert Syst. Appl.*, vol. 39, no. 8, pp. 7281–7290, 2012.
[42] H. Drucker, C. J. Burges, L. Kaufman, A. Smola, and V. Vapnik, "Support vector regression machines," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 9, 1997, pp. 155–161.
[43] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Statist. Comput.*, vol. 14, no. 3, pp. 199–222, Aug. 2004.
[44] G. Zhang, B. E. Patuwo, and M. Y. Hu, "Forecasting with artificial neural networks: The state of the art," *Int. J. Forecasting*, vol. 14, no. 1, pp. 35–62, Mar. 1998.
[45] R. J. Hyndman and A. B. Koehler, "Another look at measures of forecast accuracy," *Int. J. Forecasting*, vol. 22, no. 4, pp. 679–688, 2006.
[46] W. Alajali, S. Wen, and W. Zhou, "On-street car parking prediction in smart city: A multi-source data analysis in sensor-cloud environment," in *Security, Privacy, and Anonymity in Computation, Communication, and Storage* (Lecture Notes in Computer Science). Cham, Switzerland: Springer, 2017, pp. 641–652.
[47] J. Xiao, Y. Lou, and J. Frisby, "How likely am I to find parking?—A practical model-based framework for predicting parking availability," *Transp. Res. B, Methodol.*, vol. 112, pp. 19–39, Jun. 2018.
[48] *Sii-Mobility Portal*. Accessed: Aug. 8, 2018. [Online]. Available: https://www.sii-mobility.org
[49] *Km4City Portal*. Accessed: Aug. 8, 2018. [Online]. Available: https://www.km4city.org
[50] *R Project Portal*. Accessed: Aug. 8, 2018. [Online]. Available: https://www.R-project.org/

**CLAUDIO BADII** received the Degree in computer engineering from the University of Florence, and the Ph.D. degree with the DISIT Lab. He is currently a Research Fellow with the DISIT Lab, Department of Information Engineering, University of Florence. He focused on cloud simulator, Internet of Things, NodeJS, user behavior analysis, and machine learning. He currently focuses on the development of mobile applications in the smart city and big data.

**PAOLO NESI** is Full Professor with the Department of Information Engineering, University of Florence, where he is also the Chief of the DISIT Lab and Research Group (https://www.disit.org). His research interests include massive parallel and distributed systems, physical models, semantic computing, big data, artificial intelligence, smart city, cloud, and Internet of Things. He is the General Chair of the IEEE SCI, the IEEE ICSM, the IEEE ICECCS, DMS, WEDELMUSIC, and AXMEDIS international conferences, and the program chair of several others. He is a coordinator of several research and development multipartner international research and development projects of the European Commission such as Snap4City, RESOLUTE, ECLAP, AXMEDIS, WEDELMUSIC, MUSICNETWORK, and MOODS. He is involved in many other projects. He is the Sii-Mobility, Km4City, and Snap4City Project Coordinator.

**IRENE PAOLI** received the Degree in statistics from the University of Florence, where she is currently pursuing the Ph.D. degree in information and communication technology with the DISIT Lab. Her main competences are on statistical analysis, predictive models, and machine learning algorithms. At the DISIT Lab, her research interests are in the statistical analysis of social media and smart city data.