A Generic Predictive Model for On-Street Parking Availability

Eren Unlu[(✉)], Jean-Baptiste Delfau, Bich Nguyen, Eric Chau,
and Mehdi Chouiten

Datategy, 1 Parvis de la Defense, 92800 Puteaux, France
`eren.unlu@datategy.net`

Abstract. Despite the previously demonstrated considerable negative effects of on-street parking availability on a city's traffic flux, the developed literature on this issue is far from being voluminous. It is shown that, the duration for finding a vacant parking space consume a sizeable portion of a driver's time. Especially, for huge megacities, even small, local traffic disturbances can generate chaotic results due to their complex, inter-connected nature. Hence, being able to predict the probability of finding a vacant on-street parking place on a spot at a given time up to a reasonable degree shall be at paramount of interest for future smart-city oriented conurbations. In this paperwork, we present a generic framework supported by a machine learning model, which predicts the spatio-temporal on-street parking availability, where spots are characterized according to amenities in their vicinity.

Keywords: On-street parking · Machine learning · Smart city

1 Introduction

High price index and lack of available vacant areas make it difficult to build new off-street parking sites in contemporary metropolitan cities. Hence, most of the time, drivers have to rely on limited on-street parking. The impact of drivers searching for a parking space on urban traffic congestion ranges from around 14% [6] to 30% [19, 23], highlighting the importance of this issue. A recent study shows that the average parking space search per driver per day is approximately 9 min and that one third of all drivers gave up on searching for a vacant on-street parking place at least once during the past year [11]; the average driver looses roughly 17, 41 and 44 h per year while searching for parking spot in the US, UK and Germany respectively; finally, 72.7 billion US dollars, 23.3 billion UK pounds and 40.4 billion euros are wasted every year due to traffic disturbances [11]. Taking into account the hundreds of millions of drivers in large cities around the globe, even few minutes of blockage may result in sizeable environmental hazard due to extra carbon emission and economic loss due to wasted time and fuel.

The academic literature and industrial attempts on optimizing vehicle parking are relatively scarce compared to its potential impact. Existing literature generally focuses on parking price optimization [15,16,25] or calculates the required parking space of an area based on the maximum demand [5,13], targeting decision makers. On the other hand, studies on real-time parking availability for drivers are quite rare. The tracking of vacant on-street parking spots is generally coined with the term *smart parking* under the smart city concept. For instance, a private entrepreneurship [1] offers its clients real time vacancy predictions of parking spots in various cities using the cellular and GPS data acquired from their partners. Due to industrial copyright, they do not provide any details of their algorithm.

On the academic side, there were various attempts to solve the parking place occupancy prediction problem with different approaches. Generally, researchers apply time-series regression methods, where they estimate the future states of parking slots based on the historical data. [22] uses a Long-Short Term Memory (LSTM) Recurrent Neural Network (RNN) to predict the future occupancy rates of a parking location. In [7], authors propose a Bayesian regularized neural network using parking spot's historical data, current weather and traffic flow conditions. On the other hand, [21] develops a spatio-temporal auto-regressive model to predict the occupancy rates of parking slots in San Francisco and Los Angeles. In addition to these, there are various other similar approaches in the literature such as [10,24,26].

As it can be noticed, all of these approaches propose local models that rely on a single parking spot's historical data to predict its future state. In the best case, a spatial or spatio-temporal auto-correlation model is employed, which means the locality of information is constrained by the city. This approach thus requires previous data of an existing parking location in order to make predictions. This is a very strong limitation as parking occupancy data is not available in most of the cities in the world. Moreover even if it exists, the data are generally dispersed and confidential. Therefore, a unified, generic framework capable of making predictions for parking occupancy in various cities in the world is essential.

Considering this, we propose a predictive model for on-street parking occupancy rates, based on the hypothesis that cities, especially large metropolitan areas in our current global world shall have similar spatio-temporal characteristics. Our idea is to represent a location in a city by its surrounding amenities, where amenity is synonymous to any social and/or commercial point of interest, such as a restaurant or grocery shop. At the end of the day, it would be a reasonable assumption to consider that on-street parking demand is mostly generated by the nearby amenities. The information about the amenities of a city is taken from the well known, open source geolocation initiative OpenStreetMap [14]. *Volunteered Geographic Information (VGI)* has huge potential for any kind of smart city oriented application and is not only limited to parking. However it is quite rare to find studies which exploit these rich data sources for urbanism.

We present a global on-street parking occupancy level prediction framework, which is trained by the historical parking information of a few cities with the aim of projection on other cities without data. To the best of our knowledge, this is the first such global attempt in the literature for on-street parking. As mentioned previously, there exists various non-global approaches for on and off-street parking vacancy regression. Due to locality, it can be expected that these models may give more precise prediction compared to a global one. Nevertheless, considering the limited accessibility of parking data on most of the cities in the world, study of a global predictive framework is highly influential using the amenity content of urban locations. Therefore, we desire to construct a pioneering data scientific scheme addressing the issue and evaluate the future potential with extended datasets.

2   Training Dataset

The principal aim of the study is to be able to predict the parking occupancy ratio in various areas of any kind of metropolitan city in the world at a given time, as mentioned previously. For this purpose, one needs to choose certain common spatial characteristics of cities, which are universally correlated to temporal parking demands. As mentioned previously, we propose to use numerous types of social and/or commercial points of interest, which are tagged as *amenities* in OpenStreetMap.

OpenDataParis initiative provides all on-street parking transactions for the city center of Paris in 2014, across its 7800 parking meters [2]. It consists of a list of parking transactions records. Each transaction has a parking meter id along with its latitude and longitude (where the purchase is validated). Note that, we wish to create a global predictive structure, where we are interested in the ratio of occupancy of parking spots. Compared to off-street parking, the ratio of occupancy can only be defined loosely for on-street parking. Indeed, the maximum capacity is usually not well defined for parking meters. A driver parks to an available curbside spot in a permitted area and validates its purchase at the closest parking meter. In addition to this, there can be multiple parking meters in the same area so that one can validate his/her purchase at several locations, which makes the notion of capacity arbitrary. First, we convert the transactions to temporal statistics per parking meter indicating the instantaneous number of cars registered at a given time. For this study, we have chosen to follow hourly statistics. The capacity of a parking meter will be estimated by looking at the distribution of transactions for a given parking meter.

In Paris, on-street parking is charged only during weekdays and saturday, between 8 a.m. and 7 p.m.. Even if there are variations across cities in the world, the on-street parking is generally free of charge during night-time and weekends. For the sake of generality, we have excluded the transactions of Saturday. And in a global sense, without loss of generality, it is more convenient to assume that there shall be no significant variations for on-street parking demand within weekdays. Thus, the only temporal feature we have considered in this work is the hour of the day, which is treated as a categorical variable.

2.1   On-Street Parking Occupancy Indicator

When it comes to on-street parking, finding a universal indicator for the parking place availability on a global scale is non-trivial. First of all, on-street parking regulations are highly diverse from city to city, but also within a city. On-street parking can be prohibited, free of charge during different hours within different parts of a city. While certain streets may be available for on-street parking on both sides, others may only allow it on one side of the road.

In addition to regulations, there exists also the issue of on-street parking capacity due to street geometry. Defining a capacity as for off-street parking is not evident. Even, we may able to predict the number of parking demands, defining the occupancy level from this, is problematic. At this scale of diversity and obscurity, it is challenging to reach a common indicator for parking space availability. In order to develop a global scale prediction framework as accurate as possible, a normalized indicator for on-street parking load shall be calculated. After evaluating the number of cars assigned to each parking meter for every hour of the dataset, we have calculated the means (μ) and standard deviations (σ) of each parking meter. Then, we have defined a virtual parking meter capacity as $\mu + 1.5\sigma$ for each parking meter separately. This value is attained by empirical analysis of the parking transactions record. Next, we calculate hourly occupancy ratio of each parking meter by dividing the instantaneous number cars to the virtual capacity of the parking meter. In the case where the number of cars is larger than this value, the ratio is set to 1.0. At the end of the day, the idea is to represent a normalized universal spatio-temporal on-street parking availability metric.

3   Amenities

Each parking meter is characterized by the number of major amenity types contained in a rectangle of 150, 200, and 300 m centered on it. We assume that the on-street parking demand will depend on the points of interest within these range limits. Indeed, the distribution of amenities within these ranges shall represent the type of neighbourhood as a residential, business, touristic, leisure, dense or sparse sector; which at the end, is related to the temporal on-street parking demand. Rather than using a circular periphery, we have defined range limits in squares as in Fig. 1, where we believe it is more convenient with the rigid street geometry of most of the major cities.

In OpenStreetMap, there are hundreds of amenity types which are mostly tagged by voluntary contributors, including rare definitions which are specific to certain countries (e.g. *biergarten* in germany) or no definition at all (empty amenity type). Hence, in order to construct a universal framework, we should focus on amenities which are common to all cities, such as pharmacies or grocery shops. In addition to this, we should group amenities together into major amenity types which shall have similar temporal parking occupancy characteristics. For instance, it would be logical to claim that restaurants and cafes attract customers in similar days and hours. The categorization of amenities into four main types
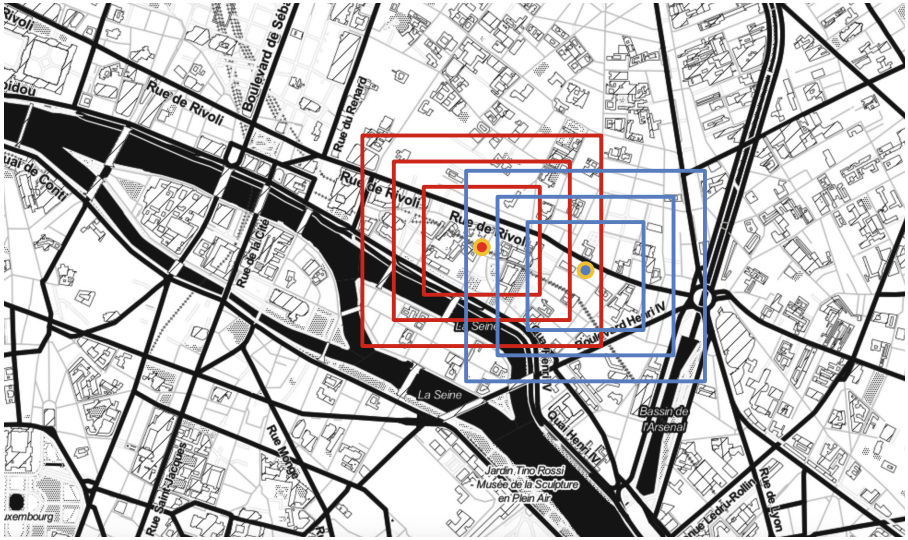
Fig. 1. Each parking meter or point in a city is represented by the number of 4 major amenity types in 150, 200 and 300 m.

is shown in Table 1. Note that it is important to consider amenity types which are expected to show high similarities in all cities. For instance, a university or an administrative amenity can be highly variant in terms of size and impact, hence also for the parking demand. However, ATM machines, banks or cafes are much more similar across the world with respect to these criteria.

Let us describe the four amenity types we chose: first, a *financial amenity* is defined, which is composed of ATM machines, banks and money transfer offices. In addition to generating parking demands directly, the density of these financial of points interest is highly correlated with the human activity around them. For example, a location with a high number of ATM machines is expected to be a more central node compared to others. Second, a *social food amenity* type is considered, including restaurants, cafes, bars etc. These are points of interest expected to have similar correlations with on-street parking occupancy within their peripheries. Another considered amenity type is the *commercial amenities* such as grocery shops, supermarkets, bakeries etc. And finally, all the rest of tagged amenities are grouped in *other amenities* general type, for which central locations tend to be more densely tagged in OpenStreetMap.

The correlation coefficients of counts of these 4 major amenity types in 3 radii of interest of 150, 200, 300 m with the normalized on-street parking occupancy for the training dataset parking meters are given in Table 2. Note that, we have also considered correlation coefficients while choosing the major amenity types and interested radii. As it can be seen, for all ranges and amenity type combinations, there exists a positive correlation with occupancy ratio up to a degree.

Table 1. 4 major amenity types defined to reflect on-street parking demand corresponding to amenities in OpenStreetMap.

Major amenity type	
Financial	Bank, ATM, Money Transfer Office
Commercial	Grocery, Supermarket, Florist, Bakery, Butcher, Rotisserie, Chocolate, Cosmetics, Pet, Dry Cleaning, Dairy, Health Food, Internet Cafe, Copyshop, Tabacco, E-Cigarette, Laundry, News Agent, Real Estate, Hair Dresser, Electronics, Clothes, Pharmacy
Social food	Cafe, Restaurant, Bar, Pub, Fast-Food
Other	–

Table 2. Correlation Coefficients of 4 major amenity types and 3 ranges for training dataset parking meters' hourly occupancy rates.

Major amenity type/range (meters)	150	200	300
Financial	0.032	0.026	0.054
Commercial	0.065	0.069	0.073
Social food	0.020	0.028	0.041
Other	0.046	0.047	0.049

4 Predictive Machine Learning Model

4.1 CatBoost

For each parking meter, we have 12 static physical features due to the number of 4 major amenity types in 150, 200 and 300 m periphery. As a temporal feature, we use the hour of the day, which we treat as a categorical variable. Before feeding these features to a machine learning algorithm, we shall convert hour category to a numerical feature. For this purpose, we have used *categorical boosting (CatBoost) encoding* algorithm [12]. Even though, this encoding scheme has been introduced recently, it has gained a significant reputation in the research community, thanks to its reported performances [4,27]. Each hour category is converted to a single numerical values between 0 and 1 after encoding, and combined with 12 physical features of each parking meter, thus producing a final feature vector with length of 13.

4.2 Random Forest Regression

We have chosen random forest regression [18], which is known for producing plausible results for voluminous datasets while avoiding overfitting. After

shuffling our OpenDataParis dataset, we divided it into 80% training and 20% test datasets for evaluation. Note that category encoding and numerical scalings are only used on the training dataset. Following detailed experimentation, the optimal number of estimators for Random Forest algorithm is found to be 150, with a maximum depth of 30. Mean absolute error for the test dataset is found to be approximately 0.19, which can be considered as an acceptable deviation for our amenity related model.

Another main advantage of the Random Forest algorithm is its high level of interpretability, similarly to other tree based approaches [8,17]. It is indeed crucial to be able to evaluate the relative importance of the features used in our model for understanding the impact of major amenity types on on-street parking demand. A well established metric of features relevance is the Mean Decrease Gini or more generally the Mean Decrease of Impurity (MDI) [20]. The MDI can be computed as follows: a decision tree is built by splitting the data in a way which minimizes a measure of impurity i (such as the Gini, Shannon entropy or Renyi entropy for example). For each node t of tree T, we thus want to find the split s_t that maximizes the impurity decrease given by,

$$\Delta i(s,t) = i(t) - p_L i(t_L) - p_R i(t_R) \tag{1}$$

where, $i(t_L)$ and $i(t_R)$ refer to impurity measures of the left and right portions of the dataset split by this node. And p_L, p_R are the proportions of samples in left and right parts of the node respectively, so that $p_L = N_{tL}/N_t$ and $p_R = N_{tR}/N_t$.

The MDI of a feature X_m is given by: [9,20]:

$$Imp(X_m) = \frac{1}{N_t} \sum_T \sum_{t \in T; v(s_t) = X_m} p(t) i(s_t, t) \tag{2}$$

where, N_T is the total number of trees of the forest, p_t is the proportion of samples reaching node t so that $p_t = N_T/N$, N being total number of samples on tree T and $v(s_t)$ is the feature used in split s_t [9,20]. Intuitively speaking, MDI thus measure how many times a feature was used for a split, highlighting its importance.

After training our model, we have reached the weighted normalized importance of 13 features as in Table 3. As expected, hour feature has by far the most significant effect, constituting more than half of the total impact. We observe that up to 300 m range, the effect of major amenity types do not vary significantly, whereas all contribute to the prediction process.

5 Hourly Predictions for the Streets of Various Cities

Unfortunately, there are only a very limited number of tagged on-street parking meters in OpenStreetMap. For most of the cities, tagged parking meters do not even exist. Due to this fact, in order to have a more universal model, we

Table 3. Mean impurity decrease based normalized feature importance of the trained random forest model for 13 features.

Mean impurity decrease			
Normalized feature importance	*150*	*200*	*300*
Category/Range (meters)			
Financial	0.013	0.017	0.034
Commercial	0.036	0.037	0.050
Social food	0.040	0.043	0.062
Other	0.045	0.043	0.051
Hour	0.51		

estimate the hourly occupancy levels of streets over world. Without loss of generality, we only consider roads which are tagged as *residential* or *living street* in OpenStreetMap. Note that, major avenues and roads may not be eligible for parking with greater probability. We do not make any assumptions about the parking regulations of streets which is obscure and we make predictions for all the considered streets. The geometrical center of the street is considered as the location for our predictive model.
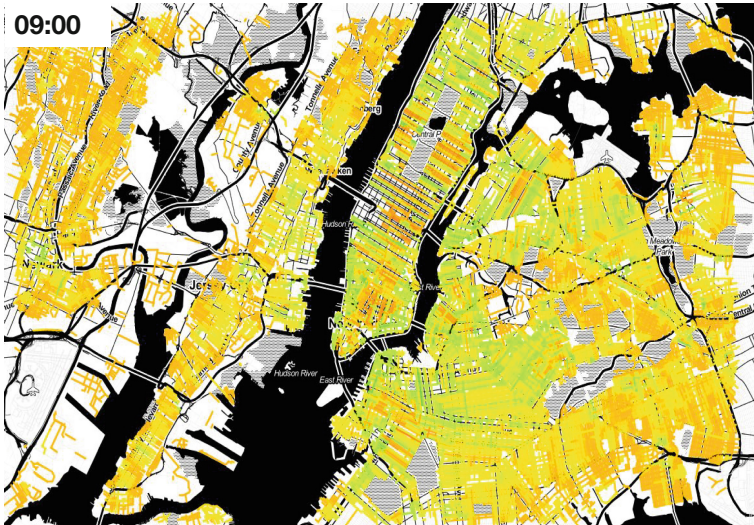
Fig. 2. Predicted on-street parking occupancies in New York, USA for 9 a.m. in weekdays. Higher occupancy levels are represented with redder hue and lower occupancy levels are represented with greener hue. (Color figure online)
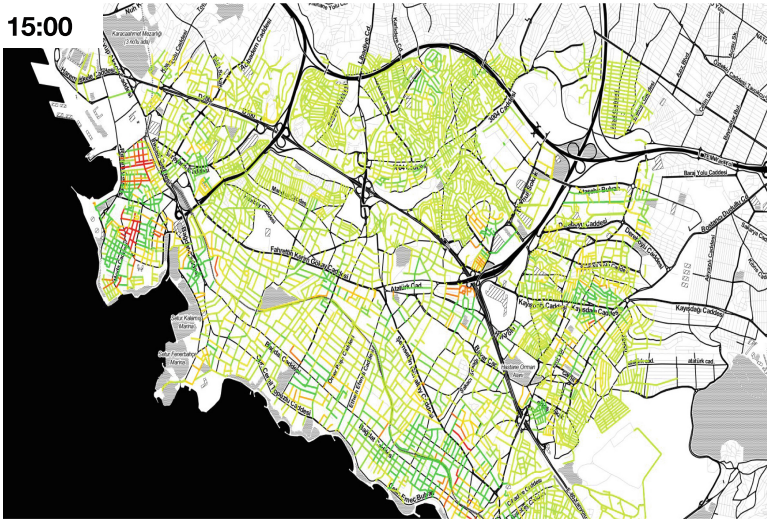
Fig. 3. Predicted on-street parking occupancies in Istanbul, Turkey for 3 p.m. in weekdays. Higher occupancy levels are represented with redder hue and lower occupancy levels are represented with greener hue. (Color figure online)

Due to limited space, we only demonstrate the results for 6 cities around globe as in Figs. 2, 3, 4, 5, 6 and 7 for three different periods of a weekday. One can observe the higher parking demands are generally centered around important hot spots as expected, especially in early morning and afternoon. Also, we can notice the medium to high occupancy levels around residential suburban areas in the morning and evening times.

In order to present a better validation, we have compared our predictions with the municipality of Seattle's parking data [3]. The dataset contains the instantaneous number of cars and capacity of each parkmeter in the city, for 2017. As we are developing a model which considers the hours between 8 a.m. and 7 p.m., we have taken the overall mean of each parkmeter for each weekday in the dataset for each hour. Then, we have calculated the amenity based static features of each parkmeter and performed our predictions. Figure 8 shows the means of real data and our predictions in a weekday at 6 p.m.

As it can be observed from Fig. 8, our model is capable of capturing the parking hot-spot regions highly accurately. For all hours considered, the mean difference (error) between real data and our predicitions is approximately 6%.

Fig. 4. Predicted on-street parking occupancies in Rennes, France for 7 p.m. in weekdays. Higher occupancy levels are represented with redder hue and lower occupancy levels are represented with greener hue. (Color figure online)
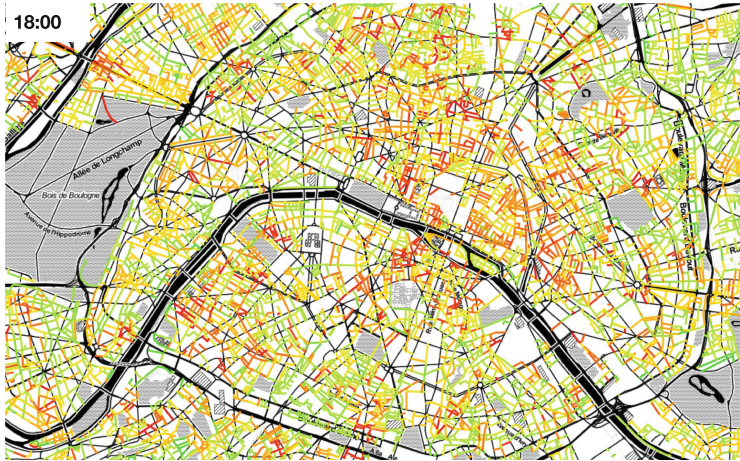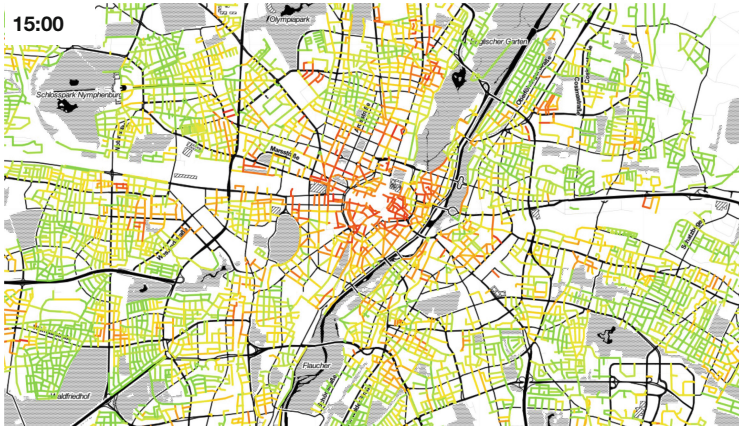

Fig. 5. Predicted on-street parking occupancies in Paris, France for 6 p.m. in weekdays. Higher occupancy levels are represented with redder hue and lower occupancy levels are represented with greener hue. (Color figure online)

Fig. 6. Predicted on-street parking occupancies in Munich, Germany for 12 a.m. in weekdays. Higher occupancy levels are represented with redder hue and lower occupancy levels are represented with greener hue. (Color figure online)
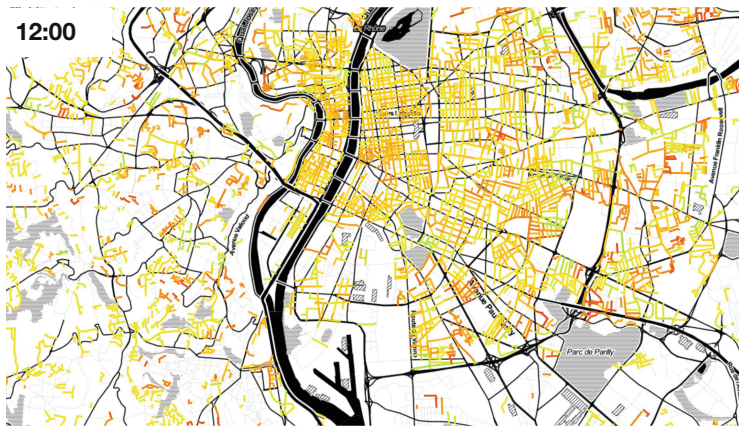

Fig. 7. Predicted on-street parking occupancies in Lyon, France for 12 a.m. in weekdays. Higher occupancy levels are represented with redder hue and lower occupancy levels are represented with greener hue. (Color figure online)

Fig. 8. Predicted on-street parking occupancies in Rennes, France for 7 p.m. in week-days. Higher occupancy levels are represented with redder hue and lower occupancy levels are represented with greener hue. (Color figure online)
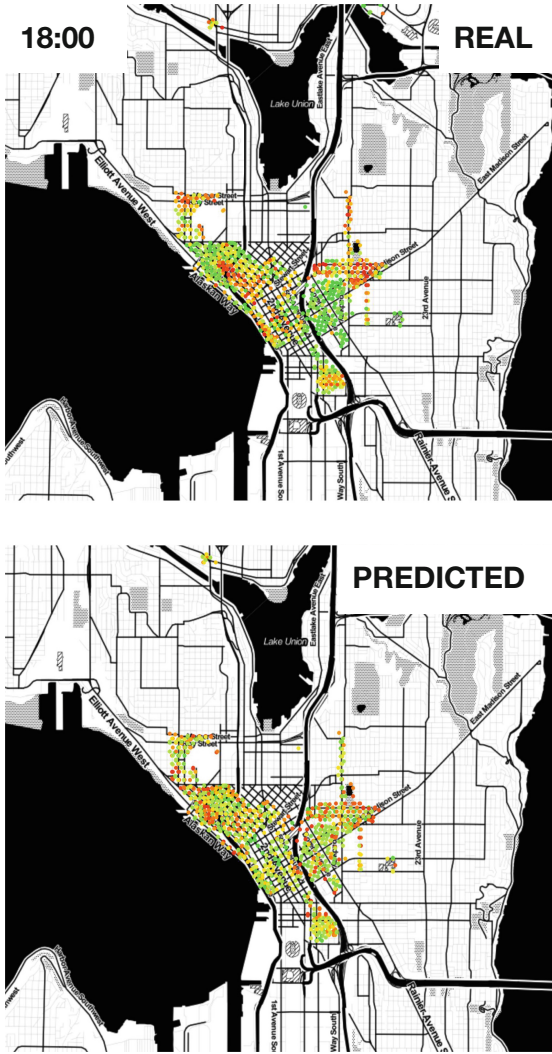
6   Conclusion and Future Work

Expecting similarities in terms of parking dynamics in contemporary global cities is a reasonable approach. Especially, if proper common points of interests (i.e. amenities) are chosen as a reference, one can estimate the on-street parking slot vacancy probabilities up to a certain extent. Even though on-street parking is a highly important subject considering the proven negative economic impact, a

universal predictive model for occupancy levels had not been proposed to the best of our knowledge. In this study, we have presented a generic framework for this purpose, where locations in cities are characterized by the number of various types of points of interests within three different radii. Constructing a unified, accurate model is quite complex due to highly variant dynamics, geometries and regulations in different cities. However, we believe this issue of a global regressive model is required to be investigated in detail due to aforementioned motivations and reasons. Therefore, we have introduced a pioneering study addressing this challenge by employing state-of-the-art machine learning algorithms. As it can be observed from the presented results in this paper, we can attain justifiable results for different cities. Unfortunately, we believe the most important bottleneck is the scarcity of open datasets about on-street parking. One can expect more and more accurate predictions with more available data sources. As a next step, we also would like to consider additional urban features such as bus stops, traffic lights, individual buildings and street geometry.

References

1. eParkomat, city smart parking s.r.o. http://eparkomat.com. Accessed 12 May 2019
2. opendataParis, 2014 on-street parking transactions. https://opendata.paris.fr/explore/dataset/horodateurs-transactions-de-paiement/information/. Accessed 31 May 2019
3. AlAwadhi, S., Scholl, H.J.: Aspirations and realizations: the smart city of Seattle. In: 2013 46th Hawaii International Conference on System Sciences, pp. 1695–1703. IEEE (2013)
4. Anghel, A., Papandreou, N., Parnell, T., de Palma, A., Pozidis, H.: Benchmarking and optimization of gradient boosting decision tree algorithms
5. Arnott, R., Inci, E., Rowse, J.: Downtown curbside parking capacity. J. Urban Econ. **86**, 83–97 (2015)
6. Arnott, R., Rowse, J.: Downtown parking in auto city. Reg. Sci. Urban Econ. **39**(1), 1–14 (2009)
7. Badii, C., Nesi, P., Paoli, I.: Predicting available parking slots on critical and regular services by exploiting a range of open data. IEEE Access **6**, 44059–44071 (2018)
8. Boulesteix, A.L., Janitza, S., Kruppa, J., König, I.R.: Overview of random forest methodology and practical guidance with emphasis on computational biology and bioinformatics. Wiley Interdisc. Rev.: Data Min. Knowl. Discov. **2**(6), 493–507 (2012)
9. Breiman, L.: Manual on setting up, using, and understanding random forests v3. 1. Statistics Department University of California Berkeley, CA, USA 1 (2002)
10. Chen, X.: Parking occupancy prediction and pattern analysis. Department Computer Science, Stanford University, Stanford, CA, USA, Technical Report CS229-2014 (2014)
11. Cookson, G., Pishue, B.: The impact of parking pain in the us, UK and Germany. Hg. v. INRIX Research (2017). Online verfügbar unter http://inrix.com/research/parking-pain/. zuletzt geprüft am 21, 2018
12. Dorogush, A.V., Ershov, V., Gulin, A.: CatBoost: gradient boosting with categorical features support. arXiv preprint arXiv:1810.11363 (2018)

13. Franco, S.F.: Downtown parking supply, work-trip mode choice and urban spatial structure. Transp. Res. Part B: Methodol. **101**, 107–122 (2017)
14. Haklay, M., Weber, P.: OpenStreetMap: user-generated street maps. IEEE Pervasive Comput. **7**(4), 12–18 (2008)
15. Inci, E.: A review of the economics of parking. Econ. Transp. **4**(1–2), 50–63 (2015)
16. Kobus, M.B., Gutiérrez-i Puigarnau, E., Rietveld, P., Van Ommeren, J.N.: The on-street parking premium and car drivers' choice between street and garage parking. Reg. Sci. Urban Econ. **43**(2), 395–403 (2013)
17. Kursa, M.B., Rudnicki, W.R., et al.: Feature selection with the Boruta package. J. Stat. Softw. **36**(11), 1–13 (2010)
18. Liaw, A., Wiener, M., et al.: Classification and regression by randomforest. R News **2**(3), 18–22 (2002)
19. Lin, T.S.: Smart parking: network, infrastructure and urban service. Ph.D. thesis, Lyon, INSA (2015)
20. Louppe, G., Wehenkel, L., Sutera, A., Geurts, P.: Understanding variable importances in forests of randomized trees. In: Advances in Neural Information Processing Systems, pp. 431–439 (2013)
21. Rajabioun, T., Ioannou, P.A.: On-street and off-street parking availability prediction using multivariate spatiotemporal models. IEEE Trans. Intell. Transp. Syst. **16**(5), 2913–2924 (2015)
22. Shao, W., Zhang, Y., Guo, B., Qin, K., Chan, J., Salim, F.D.: Parking availability prediction with long short term memory model. In: Li, S. (ed.) GPC 2018. LNCS, vol. 11204, pp. 124–137. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-15093-8_9
23. Shoup, D.C.: Cruising for parking. Transp. Policy **13**(6), 479–486 (2006)
24. Tiedemann, T., Vögele, T., Krell, M.M., Metzen, J.H., Kirchner, F.: Concept of a data thread based parking space occupancy prediction in a Berlin pilot region. In: Workshops at the Twenty-Ninth AAAI Conference on Artificial Intelligence (2015)
25. Van Ommeren, J., Russo, G.: Time-varying parking prices. Econ. Transp. **3**(2), 166–174 (2014)
26. Vlahogianni, E.I., Kepaptsoglou, K., Tsetsos, V., Karlaftis, M.G.: A real-time parking prediction system for smart cities. J. Intell. Transp. Syst. **20**(2), 192–204 (2016)
27. Zhang, F., Fleyeh, H.: Short term electricity price forecasting using CatBoost and bidirectional long short term memory neural network (2018)