VU Research Portal

Imputing parking usage on sparsely monitored areas within Amsterdam through the application of machine learning

Schmidt, Jeroen; Dugundji, Elenna; Schotten, Bas

published in

Transportation Research Record
2021

DOI (link to publisher)

[10.1177/03611981211017141](https://doi.org/10.1177/03611981211017141)

document version

Publisher's PDF, also known as Version of record

document license

Article 25fa Dutch Copyright Act

[Link to publication in VU Research Portal](#)

citation for published version (APA)

Schmidt, J., Dugundji, E., & Schotten, B. (2021). Imputing parking usage on sparsely monitored areas within Amsterdam through the application of machine learning. *Transportation Research Record*, 2675(11), 320-333. <https://doi.org/10.1177/03611981211017141>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

E-mail address:

vuresearchportal.ub@vu.nl

Imputing Parking Usage on Sparsely Monitored Areas Within Amsterdam Through the Application of Machine Learning

Transportation Research Record
2021, Vol. 2675(11) 320–333
© National Academy of Sciences:
Transportation Research Board 2021
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/03611981211017141
journals.sagepub.com/home/trr

Jeroen Schmidt¹ , Elenna Dugundji², and Bas Schotten³

Abstract

Effective parking policy is essential for cities to reduce the demand their road networks experience and to combat their carbon footprints. Existing research in the application of machine learning to understand parking behavior assumes that cities have prohibitively expensive stationary parking sensors installed, while no research has yet attempted to use machine learning to impute for parking behavior using mobile probe data of sparsely monitored areas. To this end, this paper shows that it is indeed feasible to impute parking pressure (occupation as a percentage). Gradient boosted trees were found to perform the best with an R^2 score of 0.20 and root mean squared error (RMSE) score of 0.087. This paper also found that three unique parking occupancy patterns exist in Amsterdam and that this information, in combination with neighborhood characteristics, has an impact on imputation under certain conditions.

Effective parking policy is essential for city planning as city governments strive to become more efficient with their use of resources. Recent studies show that parking policy has a direct effect on the traffic patterns that cities experience; a study by Shoup found that roughly 30% of all urban traffic is a result of cruising congestion caused by individuals looking for on-street parking (1). Parking policy, as a result, also has an impact on the carbon footprint that many cities are aiming to reduce.

Parking behavior is a challenging phenomenon to track, making the development of good parking policy particularly hard. Existing methods are either not practical or expensive if a city wishes to have a complete spatio-temporal view of how parking is used; for example, the manual counting of parking usage or the use of stationary parking sensors. Imputing (the process of determining missing observations) parking pressure (the percentage of parking occupation in a specific area) is thus an important area of research because it aids city governments to better understand parking and vehicle usage, using existing monitoring infrastructure.

The City of Amsterdam “wants to improve the quality of life and accessibility of the city by making space for pedestrians, cyclists, children playing and green spaces. In that context, the Commission (city government) is focusing on a car-free city. Several concrete measures,

including an increase in parking rates and the cancellation of 7,000 to 10,000 parking spaces, have already been announced or are in progress. However, more is needed to realize the big ambitions and broad support is essential” (2).

The city government is furthering this process with the aid of quantitative insights to better understand parking usage throughout the city. Knowing the behavior of parking usage aids policymakers and government officials to implement policies that balance the needs of individual citizens with the goals of the city.

The city government’s ability to develop quantitative-based policies is hampered because they do not have a complete view of the parking behavior. The majority of parking usage is not continuously trackable because they involve long term (housing, work) parking permits that do not self-report. Temporary parking passes purchases (which are valid for a few hours) account for roughly $\pm 15\%$ of daily parking behavior and provide a

¹University of Amsterdam, NL

²Vrije Universiteit Amsterdam, NL

³Amsterdam City Government, NL

Corresponding Author:

Jeroen Schmidt, research@schmidt.ai

continuous time-series view for the whole city. This behavior is easy to track because it is possible to link the purchased passes to tariff zones, associated pay stations, and time of purchase, plus the duration of the pass being valid.

It so happens that the City of Amsterdam has mobile sensors attached to cars that patrol the city (also referred to as “probe data” in other literature) to detect parking infringements; these mobile sensors provide the city with a sparse temporal-spatial view of parking usage. Thus the City of Amsterdam wishes to use the mobile sensor data to help impute parking pressure.

A survey on *smart parking solutions* conducted by Lin et al. showed that the area of research concerning imputation of parking usage had been sparsely explored in the context of when no continuous view of parking behavior within a single area is available (3). Three papers were identified that are closely related to the problem at hand: (1) Simulated data to resemble that of probe vehicle observations to forecast parking usage (not to be confused for predicting for unseen areas); (2) Imputed parking usage for unmonitored parts of a city by taking the predictions of a machine learning model that was trained on complete data from sensors monitoring another part of city that matched the characteristics of the unmonitored parts; (3) The use of inverse distance weighting (IDW) to impute parking pressure for unmonitored roads that were adjacent to roads that were monitored continuously with sensors (4–6). Of these three papers, only the work by Bock et al. dealt with using sparse spatio-temporal data while the other two papers used continuous data to impute data from completely unmonitored areas (4).

This paper investigates the feasibility of imputing parking pressure for sparsely monitored areas using sparse spatio-temporal data. Toward that goal, the paper is structured around the following research question:

Is it possible to impute parking usage of sparsely monitored city areas using sparse spatio-temporal data and machine learning methods? This is divided into the following sub-questions:

1. Given the structure of the data, what supervised machine learning methods produce the best results?
2. Can neighborhood characteristics of over-represented neighborhoods be used by supervised machine learning methods to better predict for under-represented neighborhoods?
3. Do neighborhoods show distinct parking behavior and does the classification of these unique behaviors help impute overall parking pressure?

Related Literature

No research was identified that tackles the problem of imputing parking usage for sparsely observed areas

within cities, but extensive related work exists on the topic of smart parking. The related research is broken down into two subsections; first, there is a general overview of the field of smart parking solutions in the context of imputing parking usage, followed by a discussion of research that has directly or very closely looked at imputing parking usage for unobserved areas.

Overall Field

An academic survey on smart parking solutions was conducted by Lin et al. over the past 15 years and looked at over 100 papers (3). A categorisation framework was proposed for the different types of research conducted within the domain of smart parking solutions. This framework offered clarity into the current research, but it demonstrated, in its omission of the topic, the lack of research conducted on the imputation of parking usage in sparsely monitored areas through predictive machine learning methods.

The work by Lin et al. provided guidance on relevant papers based on the types of data sources they used, the two notable ones being non-stationary sensors and crowd sourcing (3). Papers that looked at non-stationary sensors used cars equipped with LIDAR, Magnetron, and cameras to provide a variety of smart parking solutions and were mainly concerned with the problem of *detecting* parking occupancy; while none were related to *predicting* occupancy, they did offer insight into the methodologies used to detect occupancy. The papers that looked at using crowd-sourcing methods to determine parking usage primarily investigated the simulated situation if many individuals volunteered their parking behavior through a mobile application. These papers overlap somewhat with this paper but differ in that they assume complete city scan coverage within ± 25 min—which is not the case in the Amsterdam context (4, 7).

Predicting Parking for Unseen Areas

Three papers were identified that directly or very closely dealt with the topic of predicting parking usage within sparsely monitored areas, a majority of which comes from Fabian Bock who has been focusing on the topic of unknown parking behavior since 2016.

The work by Bock et al. looked at the performance of random forests (RFs) to predict future parking usage using data that was simulated to resemble mobile app crowd-sourcing and probe vehicle parking detection (analogous to the scanner cars used in Amsterdam) (4). The research used the San Francisco fixed point parking data to simulate the two cases to predict into a 30 min time horizon (8). This research assumed that there were sufficient probe vehicles (300 cars) in the city to allow for

a 25 min time window for a complete view of the city. This research lends credence to the idea that sensor data in the form of scanner cars can be used to predict for parking behavior if there are enough mobile sensors.

The work by Bock and Sester investigated the performance of using IDW to impute parking usage in adjacent roads from nearby road segments that had continuous sensor data (6). They found strong correlations between the adjacent parking usage the observed road segment of up to 50 m. They found that IDW was 5% better than taking the aggregated occupation values for the road segments. This research showed that there is a possibility for information gain when considering spatial relations between observations.

The paper by Ionita et al. used Gaussian clustering on the average amenity duration visits to identify pairs of areas that had monitored and unmonitored parking spaces (5). The paper showed that it is possible to train a model on one area of a city and then use that trained model to predict for another part of the city (that was completely unmonitored) on the condition that they belonged to the same cluster class. This paper also experimented with several machine learning models and determined that gradient boosting methods performed the best. This paper relied on data from San Francisco OpenParking data for the occupancy values.

Auxiliary research of interest includes the work by Bock et al. and Richter et al. (9, 10). Bock et al. looked at using clustering to aggregate parking meters according to temporal behavior to simplify the analysis of those devices (9). The research found that there were distinct parking ticket purchasing behaviors within the city of Hannover, Germany, over a one year time period. Hierarchical clustering was used on the hourly aggregated occupancies of the sold passes. This work is of interest because it can help guide the understanding of different neighborhoods when building predictive models to impute the parking usage, and it lends itself well to the work that was done by Ionita et al. (5).

Richter et al. investigated how best to forecast parking occupancy with the limited resources available to on-board car navigation systems (10). They built time-series models that only used parking occupancy figures, and they employed spatial and temporal clustering through the use of complete linkage and dynamic time warping (DTW).

Methodology

The *data science process* as defined by Blitzstein et al. was employed to conduct this study (11). This methodology proved ideal in dealing with the complexities posed by the data, particularly in refining the mobile sensor data.

Several data sources were combined and used in to impute parking usage: (1) *Scan data* containing the observation of parking occupation from the mobile sensor data; (2) *City road network data* contained the spatial geometry road networks of the city's road network and was used to add contextual information to the scans; (3) *Parking pass data* which contains the transnational log of all on-street purchases tied to a paypoint; (4) *Neighborhood data* which is an amalgamation of various statistics on individual neighborhoods; specifically, a catalogue of buildings (types) that exist in each neighborhood, the population, housing densities, and counts within each neighborhood. The data was then processed to make it suitable for the various machine learning methods.

Experiments were then conducted to find the best clustering method to identify similar neighborhoods based on parking pass behavior. The clustering results were then combined with the other data sources and used with the supervised learning methods to predict the missing parking pressure values.

Experiments also investigated the performance of the supervised learning methods depending on the level (city, cluster, neighborhood) at which they were trained. This was done to determine if the methods could learn generalizable properties from the over-represented neighborhood and thus help better predict for neighborhoods with fewer observations.

Mobile Parking Sensor Data

The mobile sensor data was obtained internally from the Amsterdam city government. This data provided the sparse observations of parking usage throughout the city and was used to determine the parking pressure for each neighborhood for sparse intervals in time.

Scan Volumes. The City of Amsterdam has between 15 to 20 vehicles (the number of vehicles used varies depending on the year) equipped with panorama license plate scanners which are used to drive throughout the city to detect parking offenders who do not have valid on-street parking passes. This data holds over 160 million unique scans and spans from January 2016 to December 2018.

The scans were categorised into five overarching categories. The first three parking pass types are held for long periods, often between 1 to 12 months; these include residential parking passes (BEWONERP), work parking passes (BEDRIJFP) and citywide parking passes (STADSBREED). Next, there is on-street temporary parking passes (BETAALDP), for short stays. The three remaining pass types were SPORTP (sport) BEZOEKP (visiting), and ELAADP (charging locations for electric cars), and were grouped under a general category called "Other."

Figure 1. Parking locations scanned by a single mobile sensor for the day of September 24, 2017.

Note: Each dot represents a scan and the colour corresponds to the time of day that scan occurred. Blue lines represent neighborhood boundaries.

Business Rules and Bias. The operation of the scanning devices is outsourced to a third-party private company that is contractually committed to satisfy two performance indicators: (1) the revenue generated from identifying non-valid parking, and (2) the chance that a vehicle without a valid parking right will be detected per hour. A softer rule expects each area to be visited at least once every day.

These business rules result in the mobile scanning devices scanning neighborhoods a certain minimum number of times to meet the performance indicators, but these rules also do not stop them from prioritizing other areas of the city to generate as much revenue as possible. Notably, the centre of Amsterdam is heavily scanned compared with the peripheries of the city. It should be noted that the north of Amsterdam (across the IJ river) was only added to the routes of the mobile scanners at the end of 2018 and thus are under-represented.

These business rules also result in a bias of when a neighborhood is scanned because it will prioritize times when the neighborhoods are likely to have high offender rates. Investigation of the data revealed distinct scanning bias patterns throughout the day: uniform, early morning and/or afternoon, and increasing/decreasing visitation trends as the hours progress.

Large parts of the city are also not scanned during weekends because they become free parking during those periods and thus there is no need for the third-party company to scan those areas. A consequence of this is that the data available for weekends is biased toward neighborhoods that are still scanned. The weekends were thus removed for the cases when models were built on the city and cluster training levels.

The low number of scanning devices results in a spatio-temporal sparse view of the city. Figure 1 is a snapshot example of single mobile probe's patrol behavior for a day and, by extension, illustrates how a sparse view of the city is obtained when a low number of mobile probes are used.

Determining Routes. Observations of parking pressure for neighborhoods can only be reliably determined if it is known what neighborhoods were scanned completely (or close to). The data poses a problem in this regard because it only provides the locations of parking spots that have been scanned; consequently, it is not possible to know the route that the probes take, what empty spots have been scanned, or what percentage of a neighborhood was scanned.

This problem obscures the ground truth of what the neighborhood parking pressure looks like because it is not possible to know when a mobile scanner traveled through a neighborhood or when it scanned the whole neighborhood. Figure 1 illustrates a snapshot example of single mobile scanner's behavior within a day and shows how a mobile sensor will partially travel through neighborhoods, that is, the dark red line traveling diagonally north-west across various neighborhoods.

The routes of the mobile scanners within a neighborhood thus had to be determined. A naive routing algorithm was employed (but future work on the topic that employs offline map matching is encouraged):

1. Match (historical) parking spots to road segments (road segment in the simplified Open Street Map [OSM] topology is defined as an edge between nodes, where nodes represent intersections) defined by simplified OSM road network (the topology defined in June-2019 was retrospectively used) (12).
2. Match scan locations to OSM road segments.
3. Calculate window differences between scan observations grouped by scan devices and neighborhood.
4. Determine when long breaks occur between scans in-order to demarcated new trips.
5. For each trip in a neighborhood, determine what unique road segments were driven over.

Routes that drove past more than 80% of the parking spots within a neighborhood were then assumed to have traveled through the whole neighborhood; routes that had less than 80% were discarded. A total of 248,000 neighborhood pressure observations were produced after the naive route matching was applied with the 80% coverage filtering.

The following features were thus created from the mobile sensor data: Parking Pressure (as a percentage),

Hour of Day, Neighborhood Code, Year, Month, Week of Year, Quarter of Year.

Parking Pass Purchases Data

Parking pass purchases were internally obtained from the *Nationaal Parkeer Register* (NPR) through relations provided by the Amsterdam City Government. The data is a transnational log of various types of parking passes purchases from January 2016 to December 2018.

The NPR data indicated if parking passes were of type: BEWONERP, BEDRIJFP, STADSBREED, or BETAALDP. The data did not provide information on how the first three (long-term) passes were used throughout the city, while the the BETAALDP are for short-term stays and thus offer a good estimate of where and how they are being used. The BETAALDP passes account for $\pm 15\%$ of daily parking usage.

The BETAALDP passes were aggregated for hourly intervals per neighborhood to align with the data generated from the scan data. The aggregation of the BETAALDP parking passes were performed by spatially joining the parking pass purchase locations to the neighborhood geometries and then aggregating statistics about the parking pass purchases on hourly intervals. Four types of aggregations were built: (1) number of passes bought; (2) parking occupation as a function of how much parking time was bought in that hour (i.e., a pass bought at 13:30 will count for 0.5 of a pass at hour 13:00); (3) Parking pressure as expressed in relation to point 1; 4) Parking pressure as expressed in relation to point 2.

The average peak parking pressure resulting from BETAALDP passes was also determined for each neighborhood. These values are shown in Figure 2. It can be seen that a handful of neighborhoods experience over 20% of their peak parking pressure from BETAALDP passes.

A total of 621,000 observations were created from the NPR data, but only 204,000 remained after joining it with the scan data. Table 1 shows a summary of the data that was created to be used with the imputation machine learning methods.

Neighborhood Characteristics Data

Descriptive characteristics of neighborhoods were obtained from the City of Amsterdam public data portal (see <https://data.amsterdam.nl/datasets/>). These datasets were only available for 2018 and were retrospectively used for 2017 and 2016.

Statistics relating to population and housing figures and densities were obtained from the Key Neighborhood

Figure 2. Average peak parking pressure resulting from parking from on-street BETAALDP pass occupation.

Figures dataset (13) and contained the following columns (13):

- area in Ha
- area in Ha (land)
- population
- population density per km² (land)
- houses
- housing density per km² (land)
- average occupancy

A dataset of all non-residential buildings in Amsterdam (mostly businesses) was obtained from the Buildings Functions Map dataset (14). It classified each building according to the following 13 classifications;

- Enterprises
- Retail
- Offices
- Care
- Activities and meeting
- Unclear
- Parking
- Hotels bars restaurants
- Education
- Going out and tourism
- Religion
- Sports
- Public transport

Clustering Experiments

Three experimenters were performed to determine which clustering methods identified the most informative clusters. The clustering was performed on the normalised

Table 1. *Nationaal Parkeer Register (NPR) Data Used for Modeling $i \in [0, 3]$ Where i is an Hour*

Feature name	Description
Date	Data of observation
Hour (t)	Hour t in day when NPR window starts
Neighborhood code	Neighborhood of observation
NPR purchases % @ $t-i$	Percentage of temporary parking passes bought with relation to the total parking spaces in a neighborhood at time $t-i$
NPR purchases # @ $t-i$	Number of temporary parking passes bought at time $t-i$
NPR occupation % @ $t-i$	Percentage of parking spots occupied with relation to the total parking spaces in a neighborhood at time $t-i$
NPR occupation # @ $t-i$	Number of parking spots occupied at time $t-i$
Average peak parking pressure from NPR	The peak parking pressure from the hourly aggregated average from all the data of that neighborhood

hourly occupations per neighborhood resulting from the BETAALDP passes.

Each experiment looked into what the optimal number of clusters should be. Clustering evaluation metrics were used to determine the optimal number of clusters and then they were used to compare the methods with respect to one another. The best clustering method was then used to create features based on the cluster classifications of the neighborhoods and then combined with the constructed features. This was inspired by the works of Bock et al. and Ionita et al. who each respectively showed that clustering methods could help impute for completely unobserved areas by identifying unique areas within a city based on on-street pass sales (5, 9).

Clustering Methods. Hierarchical and K-means clustering were used to determine similar neighborhoods based on similar on-street parking behavior from the NPR data. Hierarchical clustering with a DTW similarity matrix and Euclidean K-means was used. These methods were assessed to determine which of the methods used in Bock et al. and Richter et al. was better (9, 10).

Evaluation of Number of Clusters Used. The Silhouette score and Davies-Bouldin index were chosen to determine the optimal number of clusters that the Hierarchical and K-mean clustering methods should use (15). By choosing these two metrics we were able to assess the intra and inter cluster performance of both clustering methods.

Imputation Experiments

Three different machine learning methods were investigated: gradient boosted method (GBM), RF, and least absolute shrinkage and selection operator (LASSO). These methods were chosen because they are easy to interpret compared with black box methods, and the cur-

rent research showed a history of good performance in the domain of parking usage (4, 5).

The three methods were trained at three levels of data slicing: 1) one model is trained for the whole city; 2) one model is trained per cluster (identified from the clustering experiments); and 3) one model is trained per neighborhood. Each one of these sliced levels is respectively referred to as the *city level*, *cluster level*, and *neighborhood level*.

These models were then evaluated using R^2 and root mean squared error (RMSE) with 5-fold cross validation (CV) to determine their ability to impute the parking pressure that a neighborhood experiences. Nine experiments were conducted: three data levels tests for each of the three models.

Supervised Learning Models. Supervised machine learning models were used to predict the missing parking pressure values for hourly windows per neighborhood. Gradient boosted tree methods, GBM, RF, and LASSO (linear regression) methods were trained and tested against the known parking pressure obtained from the scan data.

GBM and RF methods were chosen because they are robust against outliers in the training data and can handle categorical and continuous data very well because of their use of ensemble decision trees (discrete decision structures) (16). LASSO was selected as it generalizes well because of its use of a penalty term, and is thus less prone to overfit the data when compared with other linear methods (16).

Three baseline models were also created to have a benchmark for the regression methods to be compared against. The baseline model used the aggregated average on an hourly basis for the whole city/cluster/neighborhood (depending on the training level) and used the corresponding result to impute the missing values for a specific hour and neighborhood.

Table 2. Best Results for Different Clustering Methods

Model	No. clusters	Silhouette index (SI) score	Davies-Bouldin index (DBI) score
K-means (Euclidean)	3	0.374	1.048
Hierarchical (dynamic time warping [DTW])	3	0.369	1.385
Hierarchical (Euclidean)	5	0.379	0.785
Hierarchical (Euclidean)	3	0.375	1.086

Evaluation of Regression Model Results. The data was split via 5-fold CV and stratified across neighborhood codes and the hour of the day to fairly assess each model's performance for an individual neighborhood. The special case of time-series CV train-test splitting was not necessary for the context of this problem because forecasting is not being done, and thus the phenomena of temporal forecasting information leakage does not need to be accounted for. RMSE (Equation 1) and R^2 (Equation 2) errors are calculated by comparing the true values (y) versus the predicted values ($\hat{y}$).

$$RMSE(y, \hat{y}) = \sqrt{\frac{\sum_{i=1}^{n_{\text{samples}}} (\hat{y}_i - y_i)^2}{n_{\text{samples}}}} \quad (1)$$

$$R^2(y, \hat{y}) = 1 - \frac{\sum_{i=1}^{n_{\text{samples}}} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n_{\text{samples}}} (y_i - \bar{y})^2} \quad (2)$$

The RMSE score was selected because it gives an average estimate of how "off" the predictions are in the units of the dependent variable. The R^2 score was chosen as it shows how much of the variance the model can capture with respect to the dependent variable.

The mean and standard deviation scores of each neighborhood were then calculated across the k-folds to understand the performance of the methods at the neighborhood level. The mean, standard deviation, and median were then calculated on top of the aggregated neighborhood metrics to get a single metric value per training level. This single value was obtained to allow for the various methods at different training levels to be fairly and easily assessed against one another.

Concretely, the aggregated mean metrics across the k-folds are: the mean of means (Equation 3a), the standard deviation of means (Equation 3b), and the median of means (Equation 3c). The aggregated standard deviation metrics across the k-folds are: mean of standard deviation (Equation 3d), standard deviation of standard deviation (Equation 3e), and median of standard deviation (Equation 3f).

Mean of Metrics:

$$\text{mean-mean}_g = \text{mean}(\text{mean}(L(\hat{y}_{gcn}, \tilde{y}_{gcn}), c), n) \quad (3a)$$

$$\text{std-mean}_g = \text{std}(\text{mean}(L(\hat{y}_{gcn}, \tilde{y}_{gcn}), c), n) \quad (3b)$$

$$\text{med-mean}_g = \text{med}(\text{mean}(L(\hat{y}_{gcn}, \tilde{y}_{gcn}), c), n) \quad (3c)$$

Standard Deviation of Metrics:

$$\text{mean-std}_g = \text{mean}(\text{std}(L(\hat{y}_{gcn}, \tilde{y}_{gcn}), c), n) \quad (3d)$$

$$\text{std-std}_g = \text{std}(\text{std}(L(\hat{y}_{gcn}, \tilde{y}_{gcn}), c), n) \quad (3e)$$

$$\text{med-std}_g = \text{med}(\text{std}(L(\hat{y}_{gcn}, \tilde{y}_{gcn}), c), n) \quad (3f)$$

where

$L(\hat{y}, \tilde{y})$ = Error Metric; e.g. R^2 or RMSE

$n \in N$ = [neighbourhoods]

$c \in C$ = [K-Folds]

$g \in G$ = [training level group]

X_c = c-Fold X subset for training

$\tilde{X}_c$ = c-Fold X subset for testing

$\hat{y}_c$ = c-Fold Y subset for testing

$X_{gc} = \{k \in X_c | \text{training level}(k) = g\}$

$\tilde{X}_{gc} = \{k \in \tilde{X}_c | \text{training level}(k) = g\}$

$\hat{y}_{gc} = \{k \in \hat{y}_c | \text{training level}(k) = g\}$

$f_{gc}(x)$ = Supervised Machine Learning Method

trained on X_{gc}

$\tilde{y}_{gc} = f_{gc}(\tilde{X}_{gc})$

$\tilde{y}_{gcn} = \{k \in \tilde{y}_{gc} | \text{neighbourhood}(k) = n\}$

$\text{mean}(v, j)$ = mean of tensor v , applied along axis j

$\text{std}(v, j)$ = standard deviation of

tensor v , applied along axis j

$\text{med}(v, j)$ = median of tensor v , applied along axis j

Results

Clustering of Parking Pass Occupation Behavior

The best-performing configurations for each of the clustering methods that were tested are shown in Table 2.

Hierarchical clustering with Euclidean distance using five clusters appeared to be the best-performing model but was outperformed by K-means (Euclidean) clustering. The SI score for the hierarchical (Euclidean) clustering method plateaued after three clusters were used (Figure 3); this means that the first three clusters explain the most variance within the data and that using more than three clusters complicates the model with little gain observed. Inspection of clusters 4 and 5 also revealed that they labelled a total of three neighborhoods. These neighborhoods had above-normal peaks compared with

Figure 3. Scores for different numbers of clusters for hierarchical (Euclidean) clustering.

the other groups, but otherwise showed similar behaviors (this would explain why the DBI score was dramatically lower at five clusters). Accordingly, hierarchical (Euclidean) clustering with three clusters was the optimal configuration for that method, but a consequence of this is that K-Means (Euclidean) is the real best-performing model—with a better SI score of 0.374 and DBI score of 1.048.

The normalised on-street parking occupancy for each cluster is plotted in Figure 4. The first three clusters capture the majority of the neighborhoods, with respective neighborhood counts of 208, 72, and 38. Each cluster appears to capture unique time-series behaviors. Clusters 0, 1, and 2 all increase and peak in similar ways during the morning and mid-day hours but differ in how they decrease near the end of the day: cluster 0 only starts decreasing 3 h after working hours stop at around 8 p.m., while cluster 1 starts decreasing at the end of working hours but never increases again, and cluster 2 starts decreasing at the end of work hours (± 5 p.m.) and then increases around 8 p.m. Visual inspection of the normalised occupancy for K-means clustering versus hierarchical clustering (DTW and Euclidean) reveals that the similar time-series behaviors arise regardless of the method that is used, which indicates robustness to the findings. The neighborhoods and their corresponding cluster numbers are projected onto the Amsterdam city map in Figure 5.

Performance of Imputation Models

The aggregated R^2 and RMSE 5-fold CV mean scores for the city, cluster, and neighborhood training levels are shown in Table 3. The aggregated R^2 and RMSE 5-fold CV standard deviation scores for various training levels are shown in Table 4. In other words Table 3 communicates the performance that can be expected from the model on unseen data while while Table 4 gives an

indication of how stable the the performance is across the CV folds.

The RMSE scores performed very similarly for the GBM, RF, and LASSO methods across all the training levels and metric aggregations. This is interesting, in that the R^2 metrics vary wildly from method to method for each metric aggregation. The negative R^2 are of particular concern because it indicates that those models perform worse than using the mean of the data. It should be noted that it is possible to get negative R^2 values when the predicted values perform worse than the line defined by $y_i - \bar{y}$ ($\bar{y}$ being the mean), that is, when the fraction is larger than 1 (17).

The RF method significantly outperforms the other modeling methods at the city and cluster training levels when the median of mean and mean of standard deviation scores are looked at, but it performs similarly (worse) when the mean of mean metrics are looked at. This indicates that the method is performing very poorly for a small subset of neighborhoods which must have extremely bad R^2 scores. The RF method performs similarly to the gradient boosting machine at the neighborhood training level when the median of means scores are looked at, the respective R^2 differences being 0.012 and 0.027.

The best training and method combination came from using GBM. It had a mean of mean R^2 score of 0.2, a median of mean R^2 score of 0.24, and a standard deviation of mean R^2 score of 0.24. Figure 6a shows the behavior of these metrics. GBM also obtained a mean standard deviation R^2 score of 0.095 and a median standard deviation of 0.064. Figure 6b illustrates the behaviors of these last two metrics. The metrics of each distinct model trained per unique neighborhood is represented by the points on Figures 6, a and b; the colour of each point indicates how many data points were used to train that particular model. The distribution plots of the RMSE and R^2 scores are illustrated on each respective axis.

The difference in the median and mean R^2 scores can be explained by the observed long tail distribution in Figure 6a. Furthermore, it can be seen that the individual models built per neighborhood are packed around the median R^2 score of 0.24, and a large majority are packed around the interval (0.0, 0.5). The individual models achieve very good RMSE scores, tightly packed around 0.08, that is to say, that most of the models are wrong by 8% parking pressure units for their respective neighborhoods.

A weak inverse relationship between RMSE and R^2 can also be seen in Figure 6a, that is to say, as the models can explain more variance in the data (R^2), the lower the RMSE error will be (relative error), but the weakness of this relationship is of concern.

Figure 4. Average occupation as a result of BETAALDP pass parking in hourly intervals from 2016 to 2018 for each cluster group identified by K-means (Euclidean) clustering.

Figure 5. Neighborhood clusters identified by K-means (Euclidean) clustering mapped onto Amsterdam.

The standard deviation of the RMSE and R^2 scores for the models that were trained and tested across the 5-folds are shown in Figure 6b. The models achieve very consistent RMSE scores across the 5-folds and deviate by only 0.005. The R^2 scores, however, are centred around 0.061 and have a more aggressive standard deviation of 0.102. The implication of this is that many individual models for specific neighborhoods will have a different R^2 score by around 10% parking pressure units.

Discussion

Given the Structure of the Data, What Supervised Machine Learning Methods Produce the Best Results?

The results indicate that it is possible to impute for parking usage using sparse spatio-temporal parking data from

mobile devices with mild confidence by using machine learning methods. In all cases, the GBM and RF models outperformed the baseline models and LASSO models. The LASSO and baseline models were unable to learn the variance of the dependent variable but maintained acceptable RMSE scores.

The best-performing models were the gradient boosting machines trained at the neighborhood level; they achieved mean of mean R^2 and RMSE scores across all the neighborhoods with 0.203 and 0.087, respectively. Thus, each model per neighborhood is, on average, incorrect by 8% parking pressure units. Likewise, the model is only able to capture roughly 20% of the variance present in the data. What is curious is that the GBM method performs poorly when trained at the city and cluster level, while the RF method radically outperforms it; a possible explanation is because GBM models use an ensemble of weak learners. Thus the GBM method is unable to discriminate the different neighborhoods to make the appropriate prediction, while the RF models can discriminate better what neighborhood to predict for by using the properties of the neighborhood because they are more prone to overfitting.

The low RMSE scores are encouraging in that they indicate that it is indeed possible to use machine learning to impute parking pressure. The R^2 score indicates that the models are able to explain a small amount of variance, in most cases, around 20%, but this is still several magnitudes better than the baseline model.

Possible explanations for the low variance are that it is a result of the obfuscated ground truth in the data and the bias present in how observations are obtained. The routes that the mobile scanners took were naively constructed and thus did not account for edge cases of false road scans occurring.

Table 3. Aggregated Scores of the Means of the Root Mean Squared Error (RMSE) and R^2 Scores Calculated across the 5-Fold Cross Validation (CV) Folds for Various Modeling Methods Trained at Different Levels

Aggregated statistic	Training level	Model type	RMSE score	R^2 score
Mean	City	Baseline	0.184	-148.5
		GBM	0.133	-7.89
		RF	0.086	0.081
		LASSO	0.171	-6.42
	Cluster	Baseline	0.173	-114.8
		GBM	0.119	-4.69
		RF	0.086	-0.24
		LASSO	0.166	-97.9
	Neighborhood	Baseline	0.090	-6.89
		GBM	0.087	0.203
		RF	0.088	0.191
		LASSO	0.099	0.003
Standard deviation	City	Baseline	0.00	275.76
		GBM	0.056	39.31
		RF	0.042	1.039
		LASSO	0.171	-6.42
	Cluster	Baseline	0.00	206.5
		GBM	0.053	25.18
		RF	0.042	4.244
		LASSO	0.083	1172.
	Neighborhood	Baseline	0.00	12.87
		GBM	0.041	0.261
		RF	0.042	0.240
		LASSO	0.044	0.089
Median	City	Baseline	0.184	-25.36
		GBM	0.121	-0.34
		RF	0.075	0.201
		LASSO	0.171	-6.42
	Cluster	Baseline	0.173	-24.30
		GBM	0.106	-0.11
		RF	0.076	0.211
		LASSO	0.144	-0.81
	Neighborhood	Baseline	0.090	-1.879
		GBM	0.077	0.226
		RF	0.080	0.199
		LASSO	0.089	-0.00

Note: GBM = gradient boosted method; LASSO = least absolute shrinkage and selection operator; RF = random forest.

There is also the issue of observation bias skewing the ground truth because of the business rules, that is, mobile scanners will drive through areas when they think there might be a high offender rate. This bias also breaks the normally distributed random route assumption made in Bock et al. and might be the leading cause of why the variance is so low (4). This possible explanation could be further explored by simulating the mobile probe patterns observed in Amsterdam with that of the San Francisco data where there is a better ground truth to validate against (8).

This paper also did not consider spatial relationships between entities, such as adjacent neighborhood parking pressures and the inter-intra traffic loads. The research by Bock et al. indicates that these features might also better explain the parking pressure, and future work

should consider incorporating more spatial relationships between entities (4).

Can Neighborhood Characteristics of Over-Represented Neighborhoods be Used by Supervised Machine Learning Methods to Better Predict for Under-Represented Neighborhoods?

The results in Table 3 show that the methods trained at the neighborhood level outperform the methods trained with neighborhood properties (the city and cluster levels). This indicates that the methods that were tested are unable to use neighborhood properties to better predict for under-represented neighborhoods. This could be because the methods themselves have their loss functions

Table 4. Aggregated Scores of the Standard Deviations of the Root Mean Squared Error (RMSE) and R^2 Scores Calculated across the 5-Fold Cross Validation (CV) Folds for Various Modeling Methods Trained at Different Levels

Aggregated statistic	Training level	Model type	RMSE score	R^2 score
Mean	City	Baseline	1598.	0.097
		GBM	0.008	6.343
		RF	0.007	0.303
		LASSO	0.008	31.090
	Cluster	Baseline	1269.	0.091
		GBM	0.008	5.012
		RF	0.007	0.755
		LASSO	0.008	195.4
	Neighborhood	Baseline	90.88	0.045
		GBM	0.007	0.095
		RF	0.006	0.108
		LASSO	0.007	0.025
Standard deviation	City	Baseline	3210.	0.00
		GBM	0.006	38.60
		RF	0.006	1.674
		LASSO	0.006	334.6
	Cluster	Baseline	2382.	0.00
		GBM	0.007	33.78
		RF	0.007	6.064
		LASSO	0.008	2603.
	Neighborhood	Baseline	58.8	0.00
		GBM	0.005	0.102
		RF	0.004	0.115
		LASSO	0.005	0.048
Median	City	Baseline	136.8	0.097
		GBM	0.007	0.132
		RF	0.005	0.076
		LASSO	0.006	0.177
	Cluster	Baseline	241.34	0.091
		GBM	0.007	0.121
		RF	0.005	0.071
		LASSO	0.006	0.180
	Neighborhood	Baseline	31.43	0.045
		GBM	0.005	0.061
		RF	0.005	0.074
		LASSO	0.005	0.009

Note: GBM = gradient boosted method; LASSO = least absolute shrinkage and selection operator; RF = random forest.

defined according to the loss of the whole training vector and do not calculate a stratified weighted error. This might create a system that optimises for neighborhood accuracy regardless of the imbalance of training points between neighborhoods.

Do Neighborhoods Show Distinct Parking Behavior and Does the Clustering of these Behaviors Help?

The clustering results showed that three distinct time-series patterns exist and correspond to distinct zones within the city of Amsterdam.

Cluster 0 captures time-series behavior that extends beyond normal work hours, until 7 p.m. This behavior is puzzling because it corresponds to areas that are exclusively high-density apartments: neighborhoods within

IJburg, neighborhoods around the inbound highway to the north of Amsterdam, and around Sloterpas. A possible explanation for this is that residents perform errands near their homes after work and thus they use on-street parking on the way home. This could be validated in future work by looking at commuter patterns within these areas.

Cluster 1 captures time-series behavior that corresponds exclusively to business hours. Visual validation of this shows that cluster 1 neighborhoods corresponding to areas that are predominantly business-focused, such as the south and south-west of Amsterdam. Cluster 1 was also able to identify the surrounding areas around the “knowledge highway” (centre east of Amsterdam); this area has a mix of municipal, higher education, and business buildings. Toward the north of Amsterdam, it

Figure 6. Scatter plot of the R^2 versus root mean squared error (RMSE) test set metrics from gradient boosted method (GBM) trained at the neighborhood level: (a) Mean of RMSE and R^2 across 5-fold cross-validation (CV) per neighborhood; (b) Standard deviation of RMSE and R^2 across 5-fold CV per neighborhood.

Note: Each point represents a unique neighborhood. A kernel density estimate plot is attached to each metric on their respective axis.

can also be seen that neighborhoods were classified as cluster 1. These areas also correspond to dense business zones, such as the dockyards, shell corporate offices, and businesses around Sloterdijk Station.

Cluster 2 captures time-series behavior corresponding to behavior that would be expected from mixed-use areas: occupation throughout work hours followed by a decrease as people leave for home, but then a slight increase as people come for the nightlife. Plotting these zones on the Amsterdam geography further validates this assumption because it shows how the neighborhoods belonging to cluster 1 inhabit mixed-use areas of the city with both housing and venues for nightlife, which is a key characteristic of neighborhoods that reside between the highway encircling the city and Old Amsterdam.

The feature importance analysis of the RF method trained at the city level revealed that the identification of cluster 2 was an important feature (top-10 most important) for predicting parking pressure. This could be because cluster 2 overlaps with areas that have low parking pressure peaks due to BETAALDP passes (Figure 2) and thus the method uses it to discriminate zones where knowing the BETAALDP passes information is of little influence to the total pressure of that neighborhood.

Conclusion

This paper has proposed a working solution to impute parking pressure in sparsely monitored areas within the City of Amsterdam and provided an outline for future work in an understudied sub-field of smart parking.

Policy Application and Data Recommendations

Policymakers traditionally would have to base their decisions on a handful of costly manual counts per neighborhood per year. The imputation of parking offers policymakers the ability to follow time-dynamic trends and to move away from relying on static views of parking usage.

A continuous dynamic parking of pressure view (obtained from the imputed techniques) could allow policymakers to do the following:

- monitor the outcome/effect of measures taken (removal/adding of parking)
- identify problematic areas and time periods
- allocate public transportation resources by using parking pressure as a proxy for transportation demand
- democratize city planning by keeping citizens informed about the state/development of parking pressure in their areas
- identify parking areas in the city that could be replaced by pedestrian/green spaces

The imputation methods are heavily reliant on a good source of historical information on the location and availability of individual parking spots, as this is a prime component of the ground truth. It would be very problematic if this kind of data is not kept up to date, since that will result in matching scans/roads with the wrong locations, and thus wrong parking pressure values.

The ground truth of the scan vehicle data could be enhanced by having a complete GPS trace of a vehicle's route, as well as a flag indicating whether or not it is scanning for parked cars. Inaccuracies of the GPS coordinates could pose an issue in cities with many high-rises, narrow streets, or overhead trees, and should be considered in future work.

Machine Learning Results

It was found that gradient boosted tree and RF methods outperformed both the average baseline method and the LASSO method. Gradient boosted trees performed the best when models were built for each neighborhood, with an average R^2 score of 0.20 and RMSE score of 0.087 across all the neighborhoods. Higher R^2 scores may be achieved if the ground truth of the data is further refined through the use of map matching and historical road networks. This will likely result in improved estimates of the observed parking pressure neighborhood experiences. This analysis may be refined by investigating spatial components through explicit features of adjacent neighborhoods or the use of more advanced deep learning methods as outlined by Yang et al. (18).

RFs performed the best when compared with alternative methods within the city- and cluster-training-level experiments. The models built using the RF method performed marginally worse (± 0.02) than GBM when it was trained at the neighborhood level. The most significant predictors of parking pressure from the best-performing method trained at the city level (RF) were the following; neighborhood characteristics associated with density, buildings associated with leisure, temporal features, and features built from the NPR data.

This paper showed that neighborhood characteristics do not improve the predictive power for under-observed neighborhoods when methods are trained at higher levels of abstraction (cluster and city level). Research in this area may be refined by defining the loss function with respect to a stratification criteria of neighborhoods.

K-means clustering with three clusters was shown to be the best-performing clustering configuration out of the experiments that were conducted. This method identified three unique time-series NPR occupation patterns that corresponded with distinct areas categorised by mixed spaces, work offices, and residential areas. Also, the clustering results help perform imputation when they were used with the RF method trained at the city level. Visual validation indicated that the unique time-series behaviors were independent of the clustering methods used. Further validation could be performed by inspecting properties of the neighborhoods, such as the percentage of building types that exist in each cluster grouping.

Author Contributions

The authors confirm contribution to the paper as follows: study conception and design: J. Schmidt, E. Dugundji, B. Schotten; data collection: J. Schmidt, B. Schotten; analysis and interpretation of results: J. Schmidt, E. Dugundji, B. Schotten; draft manuscript preparation: J. Schmidt, E. Dugundji. All authors reviewed the results and approved the final version of the manuscript.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iDs

Jeroen Schmidt <https://orcid.org/0000-0002-3106-9970>

Bas Schotten <https://orcid.org/0000-0001-9626-2088>

References

1. Shoup, DC. Cruising for Parking. *Transport Policy*, Vol. 13, No. 6, 2006, pp. 479–486.
2. *Dialoog met de stad over nieuwe stappen richting autoluw Amsterdam*. Gemeente Amsterdam, 2019. <https://web.archive.org/web/20200830124537/https://www.amsterdam.nl/nieuwsarchief/persberichten/2019/persberichten-sharon-dijksma/dialoog-stad-nieuwe-stappen-richting/>.
3. Lin, T., H. Rivano, and F. Le Mouel. A Survey of Smart Parking Solutions. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 18, No. 12, 2017, pp. 3229–3253.
4. Bock, F., S. D. Martino, and M. Sester. What are the Potentialities of Crowdsourcing for Dynamic Maps of On-Street Parking Spaces? *Proc., 9th ACM SIGSPATIAL International Workshop on Computational Transportation Science - IWCTS '16*. ACM Press, Burlingame, CA, 2016, pp. 19–24.
5. Ionita, A., A. Pomp, M. Cochez, T. Meisen, and S. Decker. Where to Park? Predicting Free Parking Spots in Unmonitored City Areas. *Proc., 8th International Conference on Web Intelligence, Mining and Semantics - WIMS '18*. ACM Press, Novi Sad, Serbia, 2018, pp. 1–12.
6. Bock, F., and M. Sester. Improving Parking Availability Maps using Information from Nearby Roads. *Transportation Research Procedia*, Vol. 19, 2016, pp. 207–214.
7. Mathur, S., T. Jin, N. Kasturirangan, J. Chandrasekaran, W. Xue, M. Gruteser, and W. Trappe. ParkNet: Drive-By Sensing of Road-Side Parking Statistics. *Proc., 8th International Conference on Mobile Systems, Applications, and Services - MobiSys '10*. ACM Press, San Francisco, CA, 2010, p. 123.
8. *Parking Sensor Data Guide*. San Francisco Municipal Transportation Agency, CA, 2013. <http://sfpark.org/>

- wp-content/uploads/2014/06/docs_sensordata.pdf. Accessed June 20, 2019.
9. Bock, F., K. Xia, and M. Sester. Mapping Similarities in Temporal Parking Occupancy Behavior Based on Citywide Parking Meter Data. *Proceedings of the ICA*, Vol. 1, 2018, pp. 1–5.
 10. Richter, F., S. Di Martino, and D. C. Mattfeld. Temporal and Spatial Clustering for a Parking Prediction Service. In *2014 IEEE 26th International Conference on Tools with Artificial Intelligence*. IEEE, Limassol, 2014, pp. 278–282.
 11. Blitzstein, J., P. Hanspeter, and K. F. Verena. *CS109 Data Science*. Harvard. <https://cs109.github.io/2015/>.
 12. Boeing, G.. OSMnx: New Methods for Acquiring, Constructing, Analyzing, and Visualizing Complex Street Networks. *Computers, Environment and Urban Systems*, Vol. 65, 2017, pp. 126–139.
 13. Government AC. *Key Neighborhoods Figures*. <https://data.amsterdam.nl/datasets/LnsAX8TtqdtJ2A/>. Accessed June 20, 2019.
 14. Government AC. *Amsterdam Building Functions*. <https://data.amsterdam.nl/datasets/CWujX-uXU9R8Sg/>. Accessed June 20, 2019.
 15. Halkidi, M., Y. Batistakis, and M. Vazirgiannis. On Clustering Validation Techniques. *Journal of Intelligent Information Systems*, Vol. 17, No. 2/3, 2001, pp. 107–145.
 16. James, G., D. Witten, T. Hastie, and R. Tibshirani. *An Introduction to Statistical Learning*. Vol. 103, Springer Texts in Statistics. Springer New York, New York, NY, 2013.
 17. Bock, F., J. Liu, and M. Sester. Learning On-Street Parking Maps from Position Information of Parked Vehicles. In *Geospatial Data in a Changing World* (T. Sarjakoski, M. Y. Santos, and L. T. Sarjakoski, eds.), Springer International Publishing, Cham, 2016, pp. 297–314.
 18. Yang, S., W. Ma, X. Pi, and S. Qian. A Deep Learning Approach to Real-Time Parking Occupancy Prediction in Spatiotemporal Networks Incorporating Multiple Spatio-Temporal Data Sources. arXiv:190106758 [cs, stat]. 2019 Jan;ArXiv: 1901.06758. <http://arxiv.org/abs/1901.06758>.