Impact of Charging Infrastructure Surroundings on Temporal Characteristics of Electric Vehicle Charging Sessions

Milan Straka[1(✉)], Ľuboš Buzna[1,2], and Gijs van der Poel[3]

[1] Department of Mathematical Methods and Operations Research, Faculty of Management Science and Informatics, University of Žilina, Univerzitná 8215/1, Žilina, Slovakia
{milan.straka,lubos.buzna}@fri.uniza.sk

[2] Department of International Research Projects - ERAdiate+, University of Žilina, Univerzitná 8215/1, Žilina, Slovakia

[3] ElaadNL, Utrechtseweg 310 (bld. 42B), 6812 AR Arnhem (GL), The Netherlands

Abstract. In this paper, we apply a data-driven approach to analyse the temporal characteristics of charging sessions performed at a slow charging infrastructure. By using the variable selection ability of the Lasso method, combined with the bootstrap driven post-selection inference, we evaluate measures quantifying the potential impacts of charging infrastructure surroundings. We derive the description of the surroundings of the charging infrastructure from several publicly available datasets, representing social, demographic, business and physical environments. From the temporal characteristics, we focus on the average and standard deviation of the connection and charging time. We uncover a non-linear relationship between the connection time and the charging time. The main driving factors behind the connection time are linked with the employment-related predictors and certain types of traffic influencing the variation of the connection time. The charging time is mainly affected by the economic wealth of residents. This study extends the knowledge about the electric vehicle driver charging behaviour and can be used to inform charging infrastructure deployment strategies.

Keywords: Electric vehicles · Smart charging · Data analysis · Temporal characteristics

This document is a result of the Operational Program Integrated Infrastructure 2014–2020 "Innovative solutions for propulsion, power and safety components of transport vehicles" code ITMS313011V334, co-financed by the European Regional Development Fund. It was partly supported by the research project VEGA 1/0089/19 "Data analysis methods and decisions support tools for service systems supporting electric vehicles", APVV-19-0441 "Allocation of limited resources to public service systems with conflicting quality criteria" and by Slovak Research and Development Agency under the contract no. SK-IL-RD-18-005.

A. L. Martins et al. (Eds.): INTSYS 2020, LNICST 364, pp. 160–174, 2021.
https://doi.org/10.1007/978-3-030-71454-3_10

1 Introduction

It is expected, that the electrification could contribute to the decarbonisation of transport and energy systems. With potentially zero emissions, electric vehicles (EVs) are an alternative to internal combustion engine (ICE) vehicles, additionally decreasing the noise pollution [1].

Charging behaviour of EV drivers is not aligned with the refuelling patterns of ICE vehicles. The fast charging stations, primarily used for long trips, are not as well utilised as slow charging stations [2]. According to [3], 95% of public charging infrastructure will be slow charging stations in 2030, compared to the current 70%. Hence, more research is required to better understand the charging patterns observed at a slow charging infrastructure.

1.1 Previous Relevant Work

Slow charging stations, compared to fast charging stations, offer higher flexibility to the electrical grid, as often just a fraction of the time that an EV is plugged into a station, is really used for charging. The length of the period when an EV is plugged into a charging station is here referred to as connection time. The part of this period during which is the EV receiving energy is denoted as the charging time. The difference between the connection and the charging time, called an idle time, may take longer than the charging time, e.g. due to the use of free parking at charging stations, forcing other EV drives to search for charging opportunities elsewhere, while decreasing the utilisation of charging infrastructure.

Although there have already been studies utilising GIS data, typically they do not use a broad range of predictors and are limited to datasets with a small number of charging events. Often, data-driven studies use GPS data from vehicles with a combustion engine to build the EV charging infrastructure [16], but this can lead to inaccurate results as the behaviour of EV users may differ. In recent years, also data-driven studies utilising larger EV charging datasets have emerged [31], proposing how to expand the existing charging infrastructure while aiming at its efficient use. A review paper [9], explores the preferences of EV users in relation to charging infrastructure while reporting on papers that include socio-economic and GIS data which are utilised to improve the location of public charging infrastructure. Among the first publications utilising socio-economic data for charging station placement is Ref. [8], where authors aim at covering the night- and day-charging demand. A GIS based charging station location approach, considering a few GIS criteria mostly related to risk, is provided in [7]. Another GIS data-based study proposing a methodology on how to locate EV charging stations at both, low-and high-level spatial scales, was provided in [5].

In [17] authors addressed the problem of estimating idle times while applying the Random Forest, Gradient Boosting and XGBoost methods using characteristics of charging sessions as predictors. The most influential predictors turned out to be start time of charging sessions, ID of EV drivers identification card, and total charged energy. The only predictor used in the study that is describing the surroundings of charging infrastructure was the type of the closest road segment. A similar analysis including more factors characterising the vicinity of charging infrastructure was conducted by Pevec et al. [25]. Authors use historical charging data, GIS data as points of interests and driving distances between charging stations to predict the utilisation of charging infrastructure and provide recommendations where to deploy charging stations.

Prediction of the EV charging infrastructure popularity with GIS data and charging infrastructure parameters by logistic regression with l_1 regularisation, Gradient Boosted Regression Trees and Random Forests methods was performed in [28]. Together with the predictive ability of the methods, authors also evaluated the influence and significance of individual predictors using the logistic regression with l_1 regularisation, where the largest impact was attributed to the residential areas and charging infrastructure parameters such as the roll-out strategy, maximum charging power and the number of other charging stations located in the proximity. Authors in [30] performed a similar study to explore the relationship between the energy consumption of charging infrastructure using the Lasso method and the GIS data describing the charging infrastructure environment. The most influential factors identified by this study are pointing at the economic prosperity of inhabitants living and working in the proximity of the charging infrastructure. Authors also applied various stratification of the charging infrastructure, where the city size returned the most promising results.

1.2 Our Contribution

In this paper, we extend the previous studies [28,30], by analysing the temporal characteristics of charging sessions. We use the Lasso method combined with the post-selection inference to identify the predictors characterising the surroundings of the charging infrastructure that are potentially influencing the temporal properties of charging sessions, specifically the connection and the charging time and their variability. Moreover, the identified features are interpreted and compared with the previous studies.

2 Data

2.1 Terminology

A document [21], issued by the Dutch authorities, proposes a unified terminology for the electromobility field. It states that a charging station may consist of several charging points. A charging point must be equipped by at least one connector, where only one connector can be used at time to charge an EV. The term charging pool denotes a set of charging stations, which share the same address and the same operator. Since we are interested to investigate the influence of surroundings of charging infrastructure, we consider charging pools as the study objects.

2.2 EVnetNL Dataset

The EVnetNL dataset contains over one million charging transactions, performed on around 1700 charging pools, by more than 50000 EV drivers. The charging pools are spread across the whole Netherlands. The first transaction was recorded in January 2012 and the most recent one in March 2016. The dataset consists of two database tables, the first is named Charging transactions. The second, named Meterreadings, contains meter reading records collected with a frequency of 15 min. All the charging pools offer slow charging, with the maximum charging power of 11 kW. In the analysis we consider only transactions performed in 2015, as this is the first and the only complete year, contained in the EVnetNL dataset when the number of EV drivers and charging pools was relatively stable. A more detailed description of the dataset can be found in Ref. [30].

2.3 Geospatial Datasets

We gathered open GIS data from various sources to model the surroundings of charging pools. Table 1 provides a brief overview of the used datasets. More elaborated descriptions are available in Ref. [30].

Table 1. Overview of collected GIS datasets.

Dataset	Brief description	Source
Population cores	Detailed population data at the resolution of municipalities	[20]
Neighbourhoods	Population data aggregated to the level of statistical units called neighbourhoods	[19]
Land use	Land use data (organised in 25 categories) modelled by high resolution polygons	[18]
Energy	Aggregated natural gas and electricity consumption of companies and households	[6]
Liveability	General index describing the quality of living in 5 categories (housing, residential area, services, safety, and living environment) at the level of neighbourhoods	[13]
Traffic flows	Database of traffic volumes (cars, buses and trucks) on individual road segments	[26]
LandScan	Ambient population density in the raster format	[15]
OpenStreetMap	OpenStreetMap (OSM) points of interest (POI)	[23]
Charging pools 2015	Locations of charging pools in 2015	[22, 24]

3 Methods

3.1 Data Processing

The GIS data were processed using circular buffers with 350 m radius centred at the locations of charging pools. Predictors were extracted from the data in vector and raster formats using attribute values and estimating their values for the buffer areas by applying the methodology which is detailed in [29]. From the POI data, the minimum distance from the charging pool to the POI and the number of POIs inside the buffer were calculated. From the GIS data, we obtained a matrix with 195 predictors.

The following response vectors were extracted from the EVnetNL dataset: average connection time, standard deviation of connection time, average charging time and the standard deviation of charging time. To obtain these quantities we grouped the charging transactions by charging pools and calculated the elements of response vectors, one for each charging pool.

Based on [4,12,14] we formed a data processing procedure displayed in Fig. 1. First, we excluded from further analysis predictors with more than 95% zero values, i.e. those which carry a small information value [14, p. 44]. To mitigate the multicollinearity we at first removed the predictors with Pearson's correlation coefficient $\rho \geq 0.95$, and in the next step iteratively removed predictors with highest variance inflation factor (VIF) value, until all the VIF values were below 10. After exploring the basic transformations from the set $\sqrt{\boldsymbol{y}}, \log(\boldsymbol{y}), \boldsymbol{y}^2, \boldsymbol{y}^3$ to the response vector $\boldsymbol{y}$ with the ordinary least squares (OLS) model, we found out the transformation $log(\boldsymbol{y})$ returned the highest R^2. So from now on, we will consider the response variable $\boldsymbol{y}$ to be log transformed. After the processing, the data matrix has 1256 rows (observations) and 119 columns (predictors).

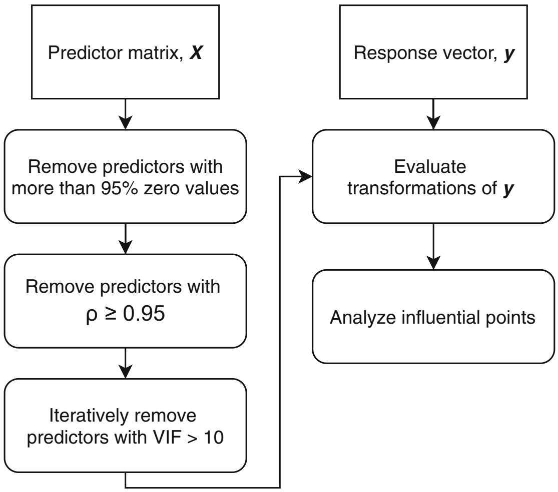

Fig. 1. Data processing procedure.

3.2 The Lasso Method and Statistical Inference

For the collection of n observations of predictor and response variables $\{(\boldsymbol{x}_i, y_i)\}_{i=1}^n$, the Lasso method [12, p. 219] solves the following optimisation problem

$$\underset{\beta_0,\beta}{\text{minimize}} \left\{ \frac{1}{2n} \sum_{i=1}^{n} (y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{p} |\beta_j| \right\}, \tag{1}$$

where $\lambda \geq 0$ is the hyperparameter, the scalar β_0 (intercept) and the vector β are optimisation variables, also called regression coefficients. The first term represents the least squares objective function and the second is the penalty function. Hence, the parameter λ determines a trade-off between the goodness of the fit and the number regression coefficients with non-zero value (variable selection).

We determine the value of the hyperparameter λ in Eq. (1) using the k-fold cross-validation [10], where we select the λ with the smallest MSE, denoted as λ^{CV} and the corresponding coefficients we denote as $\hat{\beta}_0^{CV}$ and $\hat{\beta}^{CV}$.

Due to the adaptive nature of the Lasso method, it is difficult to estimate the p-values. From the available approaches, we selected the bootstrap [10, p. 142], which was also recommended as a suitable method for assessing the stability of coefficient after the application of the Lasso method in [11]. With the bootstrap, we create a number samples of the original dataset and we apply k-fold cross-validation to each of them. This gives us an estimate of the distribution of regression coefficients $\hat{\beta}^{CV}$. For each coefficient, we evaluate the number of cases when the coefficient took a non-zero value and the consistency of its sign to assess the significance of the corresponding predictor [10, p. 153].

4 Results

4.1 Software Libraries and Settings

The data were processed in R language with packages *sf*, *raster* and *osmar* and we used the implementation of the Lasso method in the *glmnet* package. The number of folds in the cross-validation was set to 10. To determine the value of the hyperparameter λ by the cross-validation procedure combined with the Lasso method, we searched through the values 10^i, for i ranging from -4 to 0, in steps of 0.02. To sample the distribution of regression coefficients, we used 1000 bootstrap samples of the data as the input for the Lasso method.

4.2 Preliminary Analyses

We examined the distribution of all four response vectors, which we display in Fig. 2. It should be noted that the connection time is longer and varies more than the charging time. When looking at individual charging sessions, the charging time is similar to connection time mostly in cases when the connection time is short. This can be seen in Fig. 3, which shows a nonlinear relationship between the connection time and the charging time. To improve the readability of the scatter plot, we used the log-normal scale. Different charging rates of EVs and a simple fact that afterwards an EV is fully charged the connection time grows while the charging time does not, shape the nonlinear relationship.

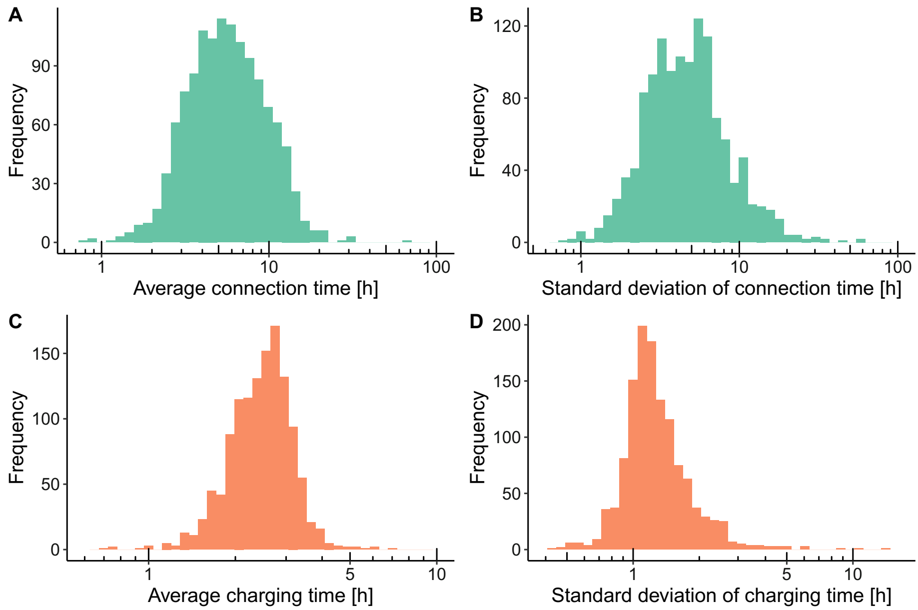

Fig. 2. Histograms of considered response variables attributed to charging pools after applying the function $log_{10}(y)$. **A** The average connection time. **B** The standard deviation of the connection time. **C** The average charging time. **D** The standard deviation of the charging time.

In Table 2 we show the coefficient of determination R^2 of the fit between the individual response vectors and the predictor matrix using the ordinary least squares method. It is notable that with the available data the connection time is better explainable than the charging time.

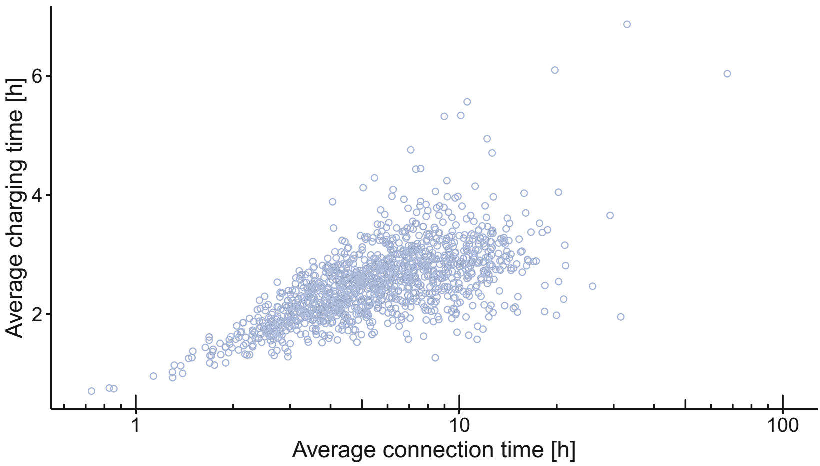

Fig. 3. Scatter plot of the average charging time vs. the average connection time at charging pools after applying the function $log_{10}(y)$. The log-normal scale improves the readability of the plot and emphasises a nonlinear relationship between these two quantities.

Table 2. The R^2 obtained by regressing the response vectors with the predictor matrix using the ordinary least squares method.

Response vector	R^2
Average connection time	0.327
Standard deviation of connection time	0.304
Average charging time	0.252
Standard deviation of charging time	0.163

4.3 Explaining the Temporal Characteristics of Charging Sessions with the Surroundings of the Charging Pools

To be able to compare multiple regression coefficients and to evaluate the influence of individual predictors, we standardised each element of $\hat{\beta}^{CV}$ by dividing it by the sample standard deviation of the corresponding predictor bootstrap sample. Standardised coefficients value shows, how much does the response changes if a predictor changes by one standard deviation. With the absolute value of a standardised coefficient increases the corresponding predictor's potential impact on the response variable [27, p. 372].

The sign of some regression coefficients tends to change across samples, which can be attributed to the low significance of the corresponding predictors and the simultaneous selection of correlated predictors [10, p. 144]. Therefore, we consider as significant those predictors, where the number of cases when the coefficient is equal to zero is less than 5% of all samples and the number of samples with the opposite sign to the sign of the median is negligible. To get a broader view on results, in Fig. 4–5 we show predictors having value zero at maximum in 10% of samples ordered descendingly by the median value of bootstrap realisations.

We can see that the number of significant coefficients is decreasing with R^2, presented in Table 2. It is worth noting that if predictors representing the minimum distance of certain objects from a charging pool have a positive coefficient, the proximity of the object has a negative impact on the response vector and the response value increases with the distance of the object from the charging pool.

For easier understanding, we divide significant predictors into three categories, namely demographics, businesses and physical environment. The minus sign in the brackets ($-$) indicates the negative impact of the feature on the response variable.

For the average connection time we find the following predictors as significant:

- *Demographics:* percentage of employed residents working in the wholesale, ICT, finance, and non-commercial sector, net number of immigrants,
- *Businesses:* number of financial and real estate businesses, density of the work-related OSM amenities, density of food OSM amenities ($-$),
- *Physical environment:* area taken by the open wet natural terrain and water, railway, terrain for retail and catering industry ($-$), area taken by roads ($-$), area taken by sports fields ($-$), density of OSM amenities related to education ($-$).

The largest positive impact on the average connection time has the area taken by the open natural terrain and water, which spreads primarily over the more developed western parts of the Netherlands, where also the four biggest Dutch cities are located. Residents employed in well-paying working sectors, and the work-related OSM amenities, which are not too close to the charging pool, have both positive influences, as well as the financial and real estate businesses. It points to the longer connection times observed at charging pools situated at locations where EV drivers tend to charge during the work. From the view of the physical environment, a positive impact has the proximity of the railway infrastructure. The negative impact has the presence of educational OSM amenities, sport fields, and terrain for retail and catering industry, i.e. locations where EV drivers tend to park for a short time and thus are not suitable for slow charging. Moreover, the negative impact is also associated with charging pools surrounded by the dense road infrastructure.

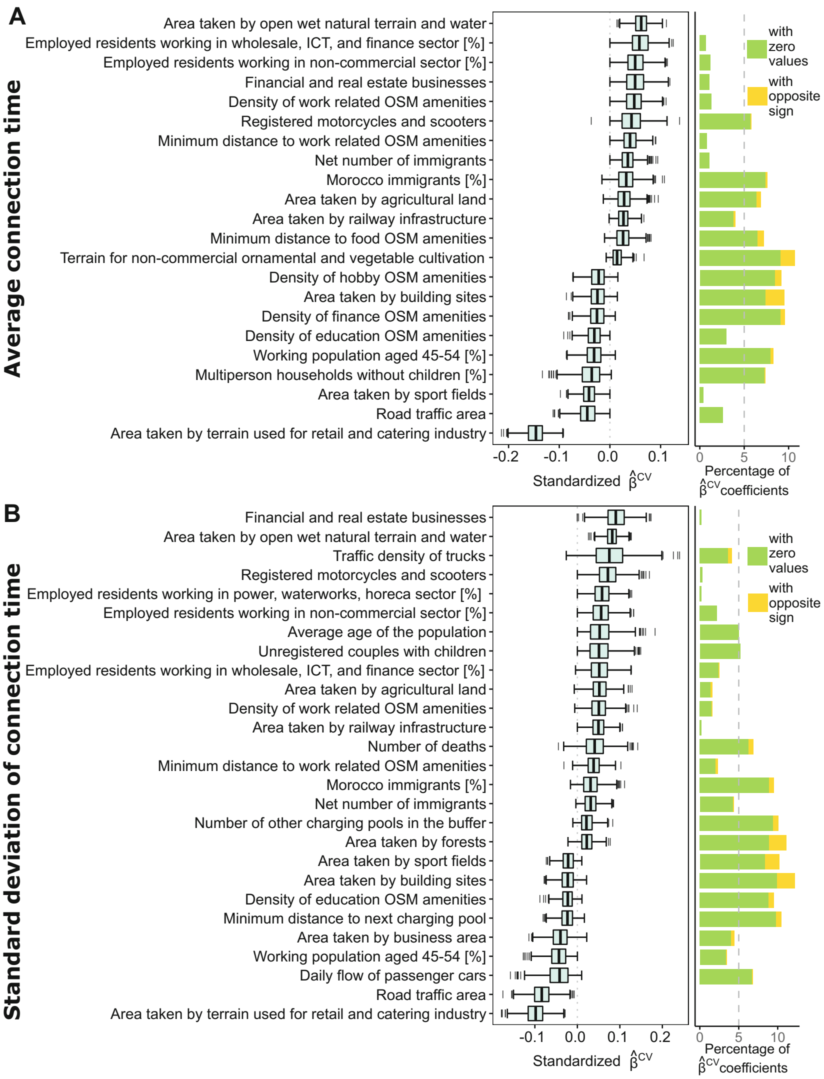

Fig. 4. (**A**) Selected regression coefficients for the average connection time and (**B**) the selected coefficients for the standard deviation of the connection time, both evaluated across the charging pools. The empirical distributions of standardised regression coefficients, obtained by the Lasso method applied to 1000 samples of the bootstrapped data, are displayed using Tukey's box plot (left panel). We show only regression coefficients with zero value in less than 10% of the samples. The coefficients are descendingly ordered, from the largest to the smallest median value of a coefficient sample. The stacked bar plot (right panel) shows the percentage of samples when the regression coefficient $\hat{\beta}^{CV}$ was set to zero and the number of samples where it reached the opposite sign as the sign of the median. We consider as significant those predictors where the percentage of coefficients with the zero value is less than 5% (indicated by the dashed line) and the number of samples with opposite sign is small.

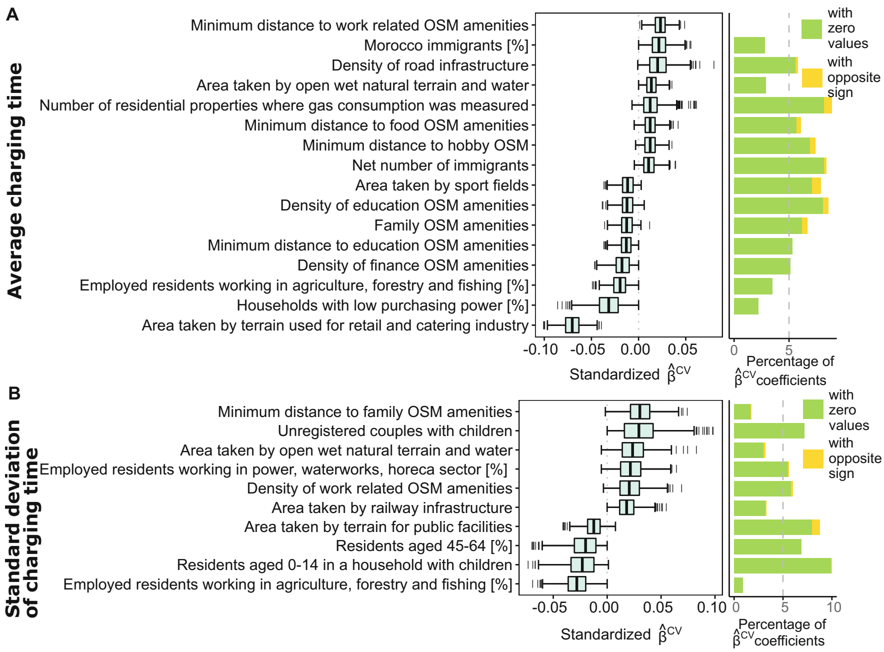

Fig. 5. (**A**) Selected coefficients for the average charging time and (**B**) the selected coefficients for the standard deviation of the charging time, both evaluated across the charging pools. The empirical distributions of standardised regression coefficients, obtained by the Lasso method applied to 1000 samples of the bootstrapped data, are displayed using Tukey's box plot (left panel). We show only regression coefficients with zero value in less than 10% of the samples. The coefficients are descendingly ordered, from the largest to the smallest median value of a coefficient sample. The stacked bar plot (right panel) shows the percentage of samples when the regression coefficient $\hat{\beta}^{CV}$ was set to zero and the number of samples where it reached the opposite sign as the sign of the median. We consider as significant those predictors where the percentage of coefficients with the zero value is less than 5% (indicated by the dashed line) and the number of samples with opposite sign is small.

For the standard deviation of the connection time, we identified the following predictors as significant:

– *Demographics:* percentage of employed residents working in the wholesale, ICT, finance, non-commercial, power, waterworks, and horeca (hotel/ restaurant/ cafe) sector, the number of registered motorcycles and scooters, average age of the population, net number of immigrants, percentage of working population aged 45–54 (−),
– *Business:* number of financial and real estate businesses, density of the work-related OSM amenities,

– *Physical environment:* area taken by agricultural land, railway, terrain for retail and catering industry (−), roads (−), businesses (−) and the minimum distance to the work-related OSM amenities.

For the standard deviation of the connection time we find similar coefficients as for the average connection time. In addition, we got the number of registered motorcycles and scooters and traffic density of trucks, indicating that higher intensity of certain type of road traffic in the proximity of charging pools, increases the variance of the connection times. The size of the area taken by businesses, terrain for retail and catering industry tend to decrease the variability in the connection times, representing locations where EV drivers typically park for a specific time, which does not vary much across different cases.

Analysis of the average charging time revealed the following predictors as significant:

– *Demographics:* percentage of Morocco immigrants, residents employed in agriculture, forestry and fishing sector (−), percentage of households with low purchasing power (−),
– *Business:* minimum distance to the work-related OSM amenities (−), density of finance OSM amenities,
– *Physical environment:* area taken by open wet natural terrain and water, area taken by terrain for retail and catering industry (−).

In difference to the connection time, the charging time has larger variance (see Fig. 2), what is most likely caused by different battery capacities and different state of charge when the charging of EVs is initiated. Surprisingly, the highest impact have the percentage of Morocco immigrants and area taken by open wet natural terrain and water. Obviously, both these predictors represent a proxy for some other hidden effects that get expressed through them. For example, higher presence of the open wet natural terrain and water is in the western part of the Netherlands, where also the largest cities are located. Considerable negative impact has area taken by terrain for retail and catering industry and proximity of the work-related OSM amenities. Here, we also see the possible impact of the wealth of residents, represented by the percentage of households with low purchasing power, and the percentage of residents with less paid occupation (agriculture, forestry and fishing).

For the standard deviation of the charging time the following predictors are significant:

– *Demographics:* residents working in agriculture, forestry and fishing industries,
– *Business:* the minimum distance to the work-related OSM amenities (−), density of finance OSM amenities,
– *Physical environment:* area taken by the open wet natural terrain and water, area taken by railway infrastructure, the minimum distance to family OSM amenities (−).

Similar predictors are linked with the standard deviation of the charging time as with the average charging time. Besides, the standard deviation gets influenced by the proximity of a railway.

4.4 Comparison with the Previous Studies

In this study, we evaluated the response vectors expressed per one charging transaction, providing a different point of view on EV charging compared to Ref. [30]. For the charging time, we found only a few significant predictors that are similar to those observed for the consumed energy and connection time. Noteworthy, the connection time is associated with a few additional predictors, e.g. the minimum distance to work or family related OSM amenities.

5 Conclusions and Discussion

In this study, we applied the Lasso method together with the post-selection inference to identify characteristics of charging infrastructure surroundings (described by predictors) that are potentially influencing the temporal characteristic of EV charging sessions.

We found out, that the connection time and charging time on charging pools exhibit a nonlinear relationship, possibly caused by the fact that battery is often fully charged long before an EV departure. The arithmetic average and the standard deviation of the connection time, are mostly affected by the working sector of residents and the work-related amenities, as well as the road density. In particular, a certain type of traffic flows present in the vicinity of charging pools appear to be influencing the variation of the connection time. In the case of the arithmetic average of the charging time at charging pools, significant predictors point to the wealth of individuals. The standard deviation of the charging times exhibits only four significant coefficients. By looking at the selected coefficients and the values of R^2 for individual variables in Table 2, we can conclude that the connection time can be better explained than the charging time from the surroundings of the charging infrastructure. This suggests that the parking behaviour, which is to large extent determining the connection time, is more influenced by the surroundings of the charging infrastructure than the state of charge at the time when an EV charging is initiated, which is strongly driving the charging time.

This study not only extends the available knowledge about the charging behaviour of EV drivers but can be also used to improve charging infrastructure deployment strategies, e.g. by utilising the information about the significant predictors in the design of predictive and prescriptive data analytic models. As future work, we recommend to identify of clusters of charging pools, e.g. to capture occasional, home and work charging and to study for them more precisely how the surroundings of charging infrastructure influence the charging behaviour of EV drivers.

References

1. Cazzola, P., Gorner, M., Schuitmaker, R., Maroney, E.: Global EV outlook 2016. Technical report, International Energy Agency, France (2016)
2. Cazzola, P., Gorner, M., Schuitmaker, R., Maroney, E.: Global EV outlook 2017: Two million and counting. Technical report, International Energy Agency (2017)
3. Cazzola, P., Gorner, M., Schuitmaker, R., Maroney, E.: Global EV outlook 2019: scaling-up the transition to electric mobility. Technical report, International Energy Agency (2019)
4. Chatterjee, S., Hadi, A.S.: Regression Analysis by Example. Wiley, Hoboken (2015)
5. Csiszár, C., Csonka, B., Földes, D., Wirth, E., Lovas, T.: Urban public charging station locating method for electric vehicles based on land use approach. J. Transp. Geogr. **74**, 173–180 (2019)
6. Nationale EnergieAtlas: Energy Atlas. https://www.pdok.nl/introductie/-/article/cbs-aardgas-en-elektriciteitsleveri-1. Accessed 16 Oct 2018
7. Erbaş, M., Kabak, M., Özceylan, E., Çetinkaya, C.: Optimal siting of electric vehicle charging stations: a GIS-based fuzzy multi-criteria decision analysis. Energy **163**, 1017–1031 (2018)
8. Frade, I., Ribeiro, A., Gonçalves, G., Antunes, A.P.: Optimal location of charging stations for electric vehicles in a neighborhood in Lisbon, Portugal. Transp. Res. Rec. **2252**(1), 91–98 (2011)
9. Hardman, S., et al.: A review of consumer preferences of and interactions with electric vehicle charging infrastructure. Transp. Res. Part D: Transp. Environ. **62**, 508–523 (2018)
10. Hastie, T., Tibshirani, R., Wainwright, M.: Statistical Learning With Sparsity: The Lasso And Generalizations. Chapman and Hall/CRC, Boca Raton (2015)
11. Hsu, D.: Identifying key variables and interactions in statistical models of building energy consumption using regularization. Energy **83**, 144–155 (2015). https://doi.org/10.1016/j.energy.2015.02.008
12. James, G., Witten, D., Hastie, T., Tibshirani, R.: An Introduction to Statistical Learning: With Applications in R. Springer, New York (2014). https://doi.org/10.1007/978-1-4614-7138-7
13. Koninkrijksrelaties, M.v.B.Z.e.: Liveability Meter. https://data.overheid.nl/dataset/leefbaarometer-2-0---meting-2016. Accessed 15 Oct 2018
14. Kuhn, M., Johnson, K.: Applied predictive modeling. Springer (2013)
15. ORN Laboratory: Landscan Datasets. https://landscan.ornl.gov/landscan-datasets. Accessed 20 May 2018
16. Liu, Q., Liu, J., Le, W., Guo, Z., He, Z.: Data-driven intelligent location of public charging stations for electric vehicles. J. Clean. Prod. **232**, 531–541 (2019)
17. Lucas, A., Barranco, R., Refa, N.: EV idle time estimation on charging infrastructure, comparing supervised machine learning regressions. Energies **12**(2), 269 (2019)
18. Netherlands CBS: CBS land cover. https://www.pdok.nl/introductie/-/article/statistics-netherlands-land-use-2015. Accessed 14 Nov 2017
19. Netherlands CBS: Neighbourhoods Dataset 2015. https://www.cbs.nl/nl-nl/dossier/nederland-regionaal/geografische-data/wijk-en-buurtkaart-2015. Accessed 20 Aug 2018
20. Netherlands CBS: Population cores in the Netherlands. https://www.cbs.nl/nl-nl/achtergrond/2014/13/bevolkingskernen-in-nederland-2011. Accessed 17 Nov 2017

21. Netherlands Enterprise Agency: Electric vehicle charging - definitions and explanation (2019). https://bit.ly/2LehwPk. Accessed 19 June 2019
22. OpenChargeMap. https://openchargemap.org. Accessed 10 Jan 2019
23. OpenStreetMap. https://www.openstreetmap.org. Accessed 13 Feb 2019
24. OplaadPalen. https://www.oplaadpalen.nl/. Accessed 20 Feb 2019
25. Pevec, D., Babic, J., Kayser, M.A., Carvalho, A., Ghiassi-Farrokhfal, Y., Podobnik, V.: A data-driven statistical approach for extending electric vehicle charging infrastructure. Int. J. Energy Res. **42**(9), 3102–3120 (2018)
26. National Institute for Public Health and the Environment: Traffic flows database. http://www.rivm.nl/. Accessed 07 Jan 2019. The database was provided for research purposes by the National Institute for Public Health and Environment
27. Siegel, A.: Practical Business Statistics. Academic Press, Cambridge (2016)
28. Straka, M., et al.: Predicting popularity of electric vehicle charging infrastructure in urban context. IEEE Access **8**, 11315–11327 (2020). https://doi.org/10.1109/ACCESS.2020.2965621
29. Straka, M., Buzna, L.: Preprocessing of GIS data for electric vehicle charging stations analysis and evaluation of the predictors significance. Transp. Res. Proc. **40**, 1583–1590 (2019)
30. Straka, M., Carvalho, R., van der Poel, G., Buzna, L.: Explaining the distribution of energy consumption at slow charging infrastructure for electric vehicles from socio-economic data. arXiv preprint arXiv:2006.01672 (2020)
31. Yang, Y., Zhang, Y., Meng, X.: A data-driven approach for optimizing the EV charging stations network. IEEE Access **8**, 118572–118592 (2020)