Amsterdam University of Applied Sciences

Predicting charging infrastructure availability based on a space-time series model

Author(s)
De Clerck, Quentin ; van den Hoed, Robert; Vanhaverbeke, Lieselot

Publication date
2016

Link to publication

Citation for published version (APA):
De Clerck, Q., van den Hoed, R., & Vanhaverbeke, L. (2016). *Predicting charging infrastructure availability based on a space-time series model*. Paper presented at 30th International Battery, Hybrid and Fuel Cell Electric Vehicle Symposium, Stuttgart, Germany.

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please contact the library: https://www.amsterdamuas.com/library/contact, or send a letter to: University Library (Library of the University of Amsterdam and Amsterdam University of Applied Sciences), Secretariat, P.O. Box 19185, 1000 GD Amsterdam, The Netherlands. You will be contacted as soon as possible.

Download date:08 Oct 2026

Creating Tomorrow

EVS30 Symposium
Stuttgart, Germany, October 9 - 11, 2016

Predicting charging infrastructure availability based on a space-time series model

Quentin De Clerck[1], Robert van den Hoed[2], Lieselot Vanhaverbeke[1]
[1]*Quentin De Clerck (corresponding author), Research Group MOBI, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium, quentin.de.clerck@vub.be*
[2]*University of Applied Sciences Amsterdam, Weesperzijde 190, 1000 BA Amsterdam, Netherlands*

Abstract

Since the European Union wants to reduce the oil dependence of the transportation system, the uptake of alternative vehicle technologies are stimulated in the member states. In the Netherlands, stimulation is already largely implemented in the form of a comprehensive charging infrastructure. This infrastructure is widely used by the electric vehicle drivers and thus there may occur a form of competition for the charging points. In this paper we address this problem by predicting the short-term availability of charging points at a given location and time by using the historical charging data in a space-time series model. The model shows better accuracy with respect to a naive method for short term predictions up to one day. This will allow charging point operators to provide customers with the service of looking up estimated charging point availability in the nearby future.

Keywords: prediction, charging, infrastructure

1 Introduction

In Europe, the transportation sector depends heavily on oil (94%) and oil importation (84%) and has thus an huge negative impact on the environment (i.e. in terms of CO_2 emission) [1]. Alternative fueled vehicles, such as electric vehicles, could help diminish the oil dependence and negative environmental impacts. The European Union is aware of these alternatives and therefore developed a plan, the Clean Power for Transport plan [2], to help facilitate the uptake of EVs. Since the roll out of charging infrastructure is a part of this plan, each member state of the European Union has to come up with its own country-specific plan for the implementation of public charging infrastructure.

The uptake of electric vehicles is not as fast in all member states. One of the front-runners in term of charging infrastructure is The Netherlands. The 4 largest cities in The Netherlands have placed more than 5.600 charging points by 2015 and are aiming at a total of 8.500 charging points by the end of 2018 [3]. This infrastructure is also extensively used by EV drivers, as demonstrate the more than 1.6 million charging sessions that took place between 2013 and 2015 [3]. The capacity utilization of the infrastructure in cities like Amsterdam can reach up to 40% (average for whole day) and much higher during evening times, so that EV users are frequently confronted with competition [3]. A possible solution to this problem would be to predict when a specific charging station (i.e. the destination of the user) will be available. In this paper we apply a space-time series model for predicting the availability of charging stations. The charging data generated by the charging infrastructure in Amsterdam is used as space-time series data.

2 Literature review

Two approaches can be used for predicting space-time series data, namely, statistical methods and machine learning methods [4]. Statistical methods that have been used in the past for space-time series predictions are for example a space-time autoregressive with moving average model (STARIMA) [5, 6] or geographically (and temporally) weighted regression or G(T)WR in short[7, 8, 9]. These methodologies have both been applied in a transportation context, but never in the context of EV charging infrastructure. An example for autoregressive models is [10], where the authors developed a multivariate space-time autoregressive model for short-term traffic predictions for 5 minute intervals. Examples for the GWR methodology are [11, 12]. The authors in [11] apply GWR for forecasting metro transit in Madrid's metro-network, while the study [12] does predict the parking demand in restricted parking zones in Santander (Spain).

When it comes to predict space-time series with machine learning methods, Artificial Neural Networks (ANN) [13] is a popular technique. In contradiction to STARIMA and G(T)WR, ANNs have been used in past studies in the context of EV charing infrastructure [14, 15]. Both studies attempt to forecast the energy demand for charging infrastructure.

The statistical and machine learning methods have both their pros and cons. Statistical methods are easier to interpret, while machine learning methods allow to approximate non-linear problems, which space-time series mostly describe [4]. Therefore, in the past years, multiple studies started to combine both techniques in one framework which resulted in prediction accuracy improvements compared to the standalone techniques [16, 4]. [4] decompose the time series in two components, namely, the space-time smooth and space-time rough components. These two components describe different levels of spatial scale. Space-time smooth describes global and large-scaled characteristics and is affected by system processes. Space-time rough, on the contrary, describes local small-scaled characteristics of space-time data and is affected by random processes. The methodology applied in this paper is based on the divide and conquer methodology proposed by [4] for improving the prediction accuracy by combining space-time smooth and space-time rough predictions.

3 Data

In this paper we work with the case-study of the central district in Amsterdam, namely Amsterdam Center. This district comprises 152 distinct charging points. The data of all the charging sessions from October 2016 are used to fit and test the model. The first three weeks are considered as training data and the rest of the month as test data.
The data has been preprocessed such to obtain a continuous time series for every hour for every charging point. For example, when a charging session occurred from Sunday 12 o'clock until 15:30, 4 entries are created where the first 3 entries (i.e. 12:00, 13:00 and 14:00) show an availability of 0 (equal to 0% availability) and the last entry (15:00) shows an availability of 0.5 (or 50%). If there is no charging session at the charging point for a given hour, then its activity is equal to 1 (or 100%).
The descriptive statistics of the resulting dataset are shown in Table 1 and an overview of the mean availability of the charging infrastructure during October 2016 is pictured in Figure 1.

Table 1: Descriptives of the charging points' availability in Amsterdam Center during October 2016

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	Stand. dev.
0	0	0.5	0.446	0.744	1	0.377

4 Methodology

In this section we cover the methodology used in this study. The explanation will cover the divide and conquer technique proposed in [4]. We refer to this paper for a more in depth explanation.

4.1 Space-time decomposition

We define space-time series data as follows [4]:

$$ST^s = \{\{loc_1^s, S_1^s\}, \{loc_2^s, S_2^s\}, ..., \{loc_n^s, S_n^s\}\} \quad (1)$$

where:

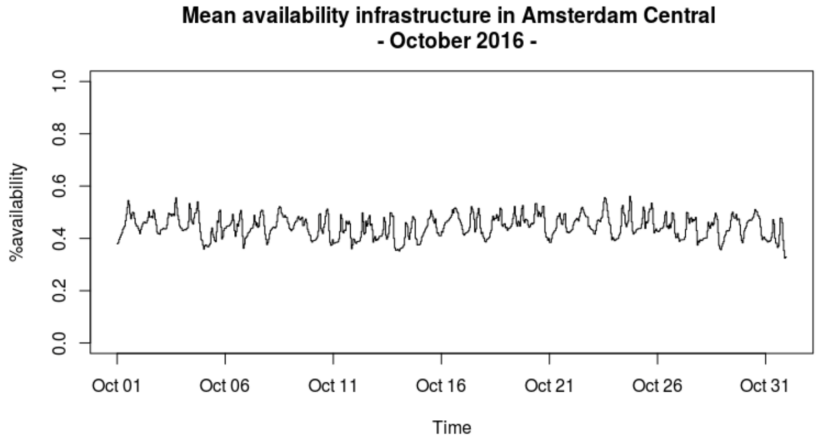


Figure 1: Mean availability of the 152 charging points in Amsterdam Center during October 2016

- s denotes the space-time series is small scaled.
- n is the total number of charging stations.
- loc_i^s is the location of the small-scaled space-time series for charging stations i $(i = 1, 2, ..., n)$.
- S_i^s is the small-scaled space-time series at charging station i $(i = 1, 2, ..., n)$.

As described in the literature review, a time series consists of two components, namely, a space-time smooth and a space-time rough component. Upscaling the observed small-scaled data allows to determine the space-time smooth characteristic of the space-time series. Upscaling the charging space-time series data is performed by first clustering the charging points and afterwards recompute the large-scaled space-time smooth component. This component is defined in an analogous way as for a small-scaled space-time series [4]:

$$ST^L = \{\{loc_1^L, S_1^L\}, \{loc_2^L, S_2^L\}, ..., \{loc_m^L, S_m^L\}\} \tag{2}$$

where:

- L denotes the space-time series is large-scaled.
- m is the total number of charging infrastructure clusters.
- loc_j^L is the location of the larger-scaled space-time series for cluster j $(j = 1, 2, ..., m)$.
- S_j^L is the larger-scaled space-time series at cluster j $(j = 1, 2, ..., m)$.

The larger-scaled space-time series S_j^L is obtained by weighing the individual time series of the charging points that are member of the same cluster j. The weight is determined based on a Voronoi diagram approach. A Voronoi is computed for the network of charging infrastructure. The total area of a charging point is equal to its surface in the Voronoi representation. The space-time smooth component is computed as a weighted sum of time series with respect to the total area of the cluster [4]:

$$S_j^L = \sum_i^{n_j} \frac{area(Voronoi(i))S_i^s}{area(Cluster(j))} \tag{3}$$

where:

- n_j is the total number of charging points belonging to cluster j $(j = 1, 2, ..., m)$.
- S_i^s is the observed time series at charging point i $(i = 1, 2, ..., n_j)$.
- $area(Voronoi(i))$ is the Voronoi area for charging point i $(i = 1, 2, ..., n_j)$.
- $area(Cluster(j))$ is the sum of the Voronoi areas of the charging points that are member of cluster j $(j = 1, 2, ..., m)$.

The decomposition of a space-time series at location i and cluster j is described by following equation [4]:

$$S_i^s = S_j^L + \varepsilon_i^s \tag{4}$$

where S_i^s is the observed space-time series, S_i^L is the larger-scaled space-time smooth component of the space-time series and ε_i^s is the smaller-scaled space-time rough component of the space-time series.

4.2 Clustering of the charging infrastructure

We use the Self-Organizing Maps algorithm [18] as clustering algorithm and the Silhouette index [19](or Sil index in short) for assessing the best number of clusters in the data. The clustering is solely based on the geographical location of the charging points. The result of the clustering procedure is shown in Figure 2. The highest Sil index value indicates the best number of clusters. Therefore, we clustered the time series in 33 clusters, resulting in 33 smooth space-time series for the data of Amsterdam Center.

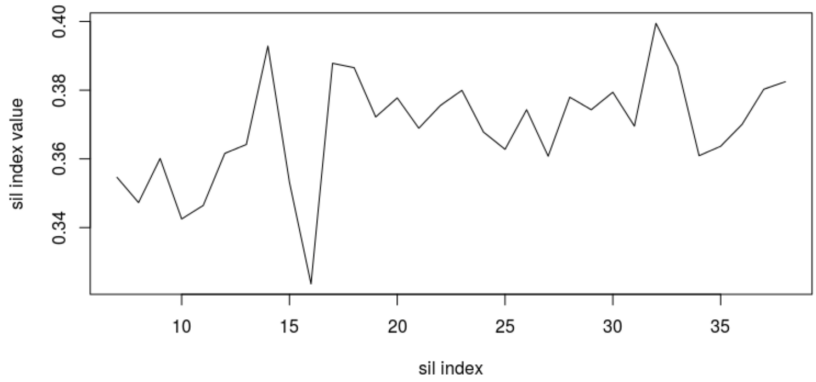


Figure 2: Sil index values for different numbers of clusters

4.3 Space-time smooth prediction

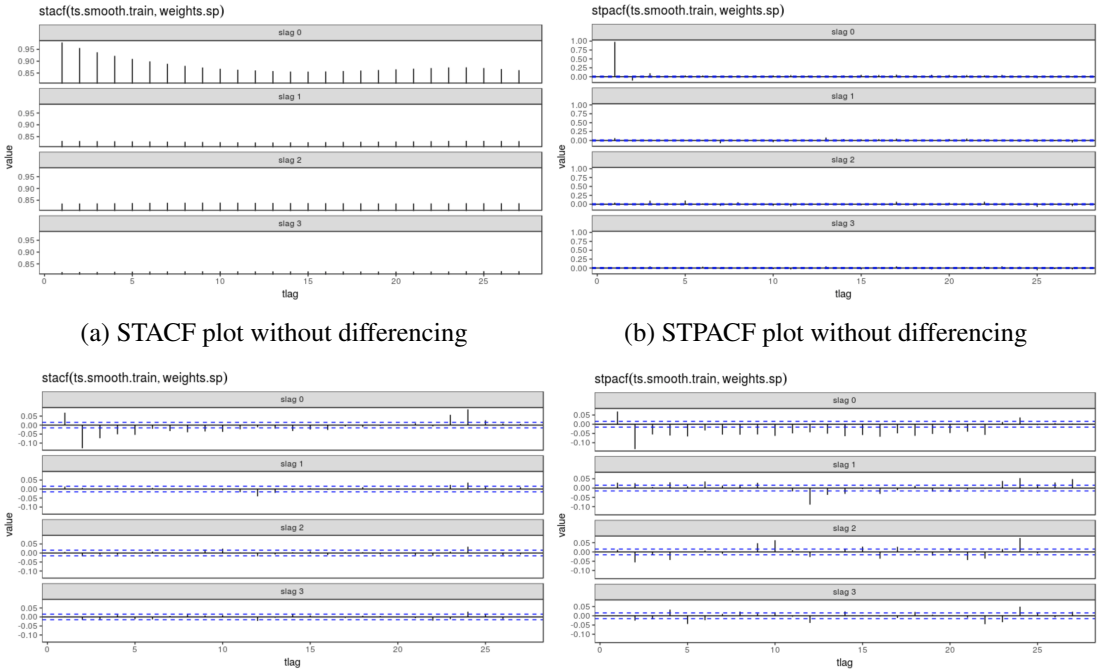
Usually, the Box-Jenkins methodology [17] is used for estimating a time series using ARIMA-based models (such as STARIMA). This methodology consists of three steps. The first step is the identification of the autoregressive and/or moving average components. These parameters can be identified by performing plots of the autocorrelation function (ACF) and partial autocorrelation function (PACF). These plots give indications which parametrization would be more appropriate given the data. The second step is the estimation step. In this step, the selected time series model is estimated for the training data. This step results in a model that should describe the temporal dynamics of the time series and that enables short-term predictions. The last step of the Box-Jenkins methodology is the diagnosis of the model. Multiple tests can be applied such as tests testing the normality or autocorrelation in the residuals. Plotting the ACF and PACF for the residuals can also give a visual clue to where the model could be further improved.

4.3.1 Test stationarity of time series

Before starting with the identification of the time series' parameters, it is important to test whether the time series is stationary or not, since ARIMA based methods assume a stationary time series in order to produce statistical significant results. A stationary time series is a time series where the mean and variance is constant over time. There are two methods to make non stationary time series stationary, namely differencing and transformation of the time series data. Differencing stabilizes the mean of the time series whereas transformations, such as logarithms, are used to stabilize the variance.

To test non stationarity, two popular tests are applied: Augmented Dickey Fuller (ADF) test and the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test. The ADF test did not indicate any non stationarity whereas KPSS did indicate non stationarity for 17 of the 33 space-time smooth series. Therefore we plotted the space-time ACF (STACF) and space-time PACF (STPACF) to visually confirm the possible presence of non stationarity. Figure 3a depicts the STACF plot of the space-time series. This figure

suggests that the space-time series is non stationarity since the spikes diminish rather slowly. Once differenced, the STACF plot in Figure 3c shows a much steeper descent which is typical for stationary space-time series.



(a) STACF plot without differencing

(b) STPACF plot without differencing

(c) STACF plot with differencing

(d) STPACF plot with differencing

Figure 3: STACF and STPACF plots

4.3.2 Identification of STARIMA parameters

STACF and STPACF plots are plotted in Figures 3c and 3d for the identification of the STARMA parameters. These plots indicate the AR (PACF plot) and MA (ACF plot) parameters required for fitting the model well. The STACF and STPACF are plotted for the neighborhood orders or spatial lags(slag) considered in this paper (slag 0, 1, 2, 3), where slag0 is a spatial lag of order 0 (meaning the charging point itself), slag3 is a spatial lag of order 3 (meaning the third order neighborhood). We refer to [6] for more details concerning neighborhood orders.

4.3.3 Estimation of the STARIMA model

The variables included in a tentative, first estimation of the model were all the variables that show significant spikes in the STACF (for the MA variables) and the STPACF (for the AR variables). Significant spikes are the spikes crossing the blue lines above and below 0 in the different slags.

This model included 84 different variables. Once this model estimated, we eliminated the variables that showed a significance (p-value) above 0.20. We repeated this procedure two times, such that most insignificant variables (with p-values superior to 0.05) were eliminated. We did not repeat this procedure until all remaining variables are significant such to ensure the model residuals would not suffer from autocorrelation. The resulting model contains 38 different variables, which are reported in Table 2 with their coefficients and significance levels.

Variables with larger (positive or negative) coefficient values have a greater impact on the space-time rough predictions. Positie This means the variables with the most impact on the predictions are for the AR variables: tlag 2 and slag 0, tlag 7 and slag 0 and tlag 24 and slag 0; and for the MA variables: tlag 24 and slag 0.

Table 2: Estimated model after filtering most insignificant variables

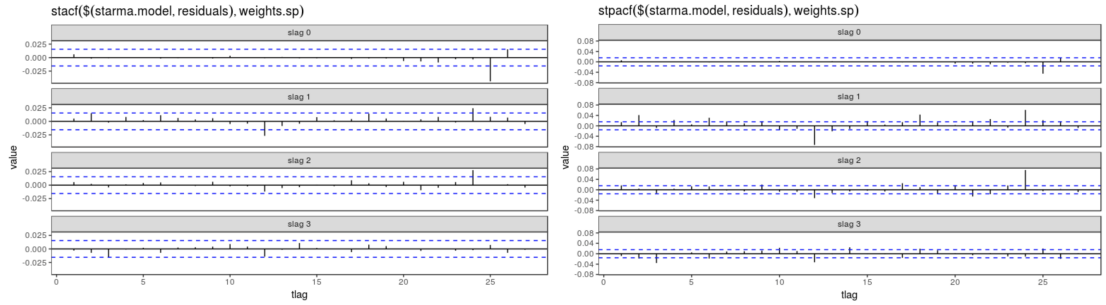
Component	tlag	slag	Coefficient	p-value
	2	0	-0.1895955	2.2e-16 '***'
	2	2	-0.0401034	0.089692 '.'
	3	0	-0.1208886	2.2e-16 '***'
	4	0	-0.1223177	2.2e-16 '***'
	4	2	-0.0413582	0.080958 '.'
	4	3	0.0357768	0.058965 '.'
	5	0	-0.1309294	2.2e-16 '***'
	5	3	-0.0397421	0.035704 '*'
	6	0	-0.0962015	2.2e-16 '***'
	7	0	-0.1896243	2.017e-05 '***'
	8	0	-0.1167275	2.2e-16 '***'
	8	3	0.0229125	0.228510
	9	0	-0.1248682	2.2e-16 '***'
	10	0	-0.1308069	2.2e-16 '***'
	10	2	0.0499932	0.035672 '*'
AR	11	0	-0.1131262	2.2e-16 '***'
	12	0	-0.1078057	2.2e-16 '***'
	13	0	-0.1088246	2.2e-16 '***'
	14	0	-0.1203823	2.2e-16 '***'
	15	0	-0.1074503	2.2e-16 '***'
	15	2	0.0248959	0.296942
	16	0	-0.1155406	2.2e-16 '***'
	16	2	-0.0357168	0.132909
	17	0	-0.0925406	2.2e-16 '***'
	18	0	-0.1010450	2.2e-16 '***'
	19	0	-0.0854966	2.2e-16 '***'
	20	0	-0.0782614	2.2e-16 '***'
	21	0	-0.0638221	9.895e-12 '***'
	22	0	-0.0790542	2.2e-16 '***'
	23	3	-0.0260733	0.169836
	24	0	0.2747809	2.2e-16 '***'
	27	1	0.0363098	0.089855 '.'
	7	0	0.0727438	0.108928
	22	3	-0.0282990	0.166847
MA	23	0	-0.0279205	0.003268 '**'
	23	1	0.0565931	0.013404 '*'
	24	0	-0.2775341	2.2e-16 '***'
	24	3	0.0599093	0.003433 '**'
Signif. codes:	'***' ≤ 0.001	'**' ≤ 0.01	'*' ≤ 0.05	'.' ≤ 0.1

4.3.4 Validation of the STARIMA model

The validation is the last step of the Box-Jenkins methodology. The model is validated both statistically and visually. We applied a multivariate Box-Pierce non correlation test. The null-hypothesis, that states that there is no correlation in the models residuals, was not rejected (p-value: 0.214). Visually we confirmed the model fits the data well since most spikes are within the blue lines in Figures 4. The spikes crossing the blue lines indicate that extra temporal orders could be tested for an alternative model.

4.4 Space-time rough prediction

The space-time rough predictions are obtained using ANNs [13]. Each time series is fitted by an individual ANN that is parameterized specifically for this time series. The ANN parameters are the number of tlags to consider and the size of its hidden layer. The parameterization is carried out by an heuristic built in the *nnetar* function from the *forecast* R package [21, 20]. This heuristic determines the number of tlags as being the same number of tlags for an optimal autoregressive model given the Akaike Infor-



(a) STACF plot (b) STPACF plot

Figure 4: STACF and STPACF plots for the model's residuals

mation Criterion and determines the size of the hidden layer as being the half of the input nodes (number of observations) + 1.

4.5 Space-time recomposition

Once both space-time smooth and space-time rough data has been predicted by respectively STARIMA and ANNs, the two predictions are recomposed as formulated in:

$$\hat{S}_i^s = \hat{S}_j^L + \hat{\varepsilon}_i^s \tag{5}$$

where $\hat{S}_j^L$ is the predicted space-time smooth series, $\hat{\varepsilon}_i^s$ is the predicted space-time rough series and $\hat{S}_i^s$ are the recomposed predictions for the original time series.

5 Results

Figure 5 depicts 4 random selected time series (in black) pictured with predictions for one week starting from the third week of October (in red) and the mean availability of the first three weeks of October (the green line).

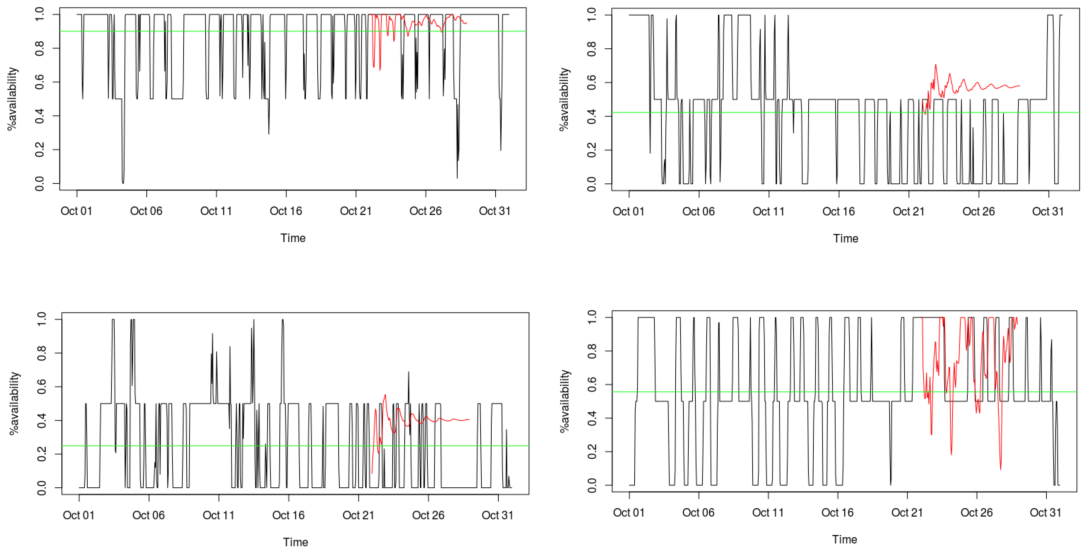


Figure 5: Example predictions of randomly selected time series

The prediction accuracy is benchmarked against the mean of the first three weeks of October. The mean is a naive, but robust indication of the expected future availability.

We applied a correction to predictions that are outside the boundaries [0,1], such that the resulting predictions keep their original meaning (i.e. 0 for unavailable and 1 for totally available during a specific hour). The benchmark compares the accuracy within 6 hours, 12 hours, 1 day, 2 days and 1 week between the average availability, the model's predictions and the corrected predictions. We use the Mean Absolute Error (MAE) as measure of performance. This measure indicates in our case the average error margin in percentage. The results of the benchmark are reported in Table 3.

Table 3: Accuracy results using the MAE metric

Method	6 hours	12 hours	1 day	2 days	1 week
Mean	0.2736	0.2684	0.2714	0.2741	0.2728
Model	0.1588	0.2099	0.2557	0.2797	0.2998
Corrected	0.1449	0.1990	0.2417	0.2671	0.2902

The benchmark shows that the model performs better than the naive solution for predictions up to one day. The model accuracy is equal or worse beyond 2 days predictions. These results indicate the model is more fit for short term predictions rather than for predictions over a longer period.

6 Conclusion

This paper aims at predicting the short-term availability of charging infrastructure based on space-time series data (i.e. charging infrastructure data from the city of Amsterdam, The Netherlands). Since the space-time domain has different scales, namely space-time smooth and space-time rough, a space-time series model combining these two scales is considered. The model shows better accuracy with respect to a naive method for short term predictions up to one day.

The resulting model could be used by charging point operators as service for their users, enabling them to check in advance the predicted availability of the charging facilities at their destination and estimated time of arrival in the nearby future.

Future research will integrate this short-term availability prediction in an optimization algorithm for route planning for EV drivers in which minimization of the risk that the selected charging point is unavailable (due to charging demand competition) is combined with the shortest path towards the actual final destination.

Acknowledgments

The authors would like to thank the Municipality of Amsterdam for providing access to the charging data.

References

[1] *European Commission*, http://ec.europa.eu/transport/themes/urban/cpt_en, accessed on 2017-01-18.

[2] *European Parliament and Council*, Directive 2014/94/EU on the deployment of alternative fuels infrastructure, European Union, Brussels, 2014.

[3] R. Wolbertus, R. van den Hoed, S. Maase, *Benchmarking Charging Infrastructure Utilization*, in 29th International Electric Vehicle Symposium 2016 (EVS29), Montréal, Québec, Canada, June 19-22, 2016, 1-16.

[4] M. Deng, W. Yang, Q. Liu, Y. Zhang, *A divide-and-conquer method for space–time series prediction*, Journal of Geographical Systems, *ISSN 1435-5930*, 19(2017), 1-19.

[5] A. D. Cliff, J. K. Ord, *Space-Time Modelling with an Application to Regional Forecasting*, Transactions of the Institute of British Geographers, *ISSN 0020-2754*, 64(1975), 119-128.

[6] P. E. Pfeifer, S. J. Deutsch, *A Three-Stage Iterative Procedure for Space-Time Modeling*, Technometrics, *ISSN 1537-2723*, 22(1980), 35-47.

[7] C. Brunsdon, A. S. Fotheringham, M. E. Charlton, *Geographically Weighted Regression: A Method for Exploring Spatial Nonstationarity*, Geographical Analysis, *ISSN 1538-4632*, 28(1996), 281-298.

[8] B. Huang, B. Wu, M. Barry, *Geographically and temporally weighted regression for modeling spatio-temporal variation in house prices*, International Journal of Geographical Information Science, *ISSN 1365-8816*, 24(2010), 383-401.

[9] A. S. Fotheringham, R. Crespo, J. Yao, *Geographical and Temporal Weighted Regression (GTWR)*, Geographical Analysis, *ISSN 1538-4632*, 47(2015), 431-452.

[10] W. Min, L. Wynter, *Real-time road traffic prediction with spatio-temporal correlations*, Transportation Research Part C: Emerging Technologies, *0968-090X*, 19(2011), 606-616.

[11] D.O. Cardozo, J.C. García-palomares, J. Gutiérrez, *Application of geographically weighted regression to the direct forecasting of transit ridership at station-level*, Applied Geography, *0143-6228*, 34(2012), 548-558.

[12] A. Ibeas, R. Cordera, L. Olio, J.L. Moura, *Modelling demand in restricted parking zones*, Transportation Research Part A: Policy and Practice, *0965-8564*, 45(2011),485-498.

[13] T. Kohonen, *An Introduction to Neural Computing*, Neural Networks, *ISSN 0893-6080*, 1(1988), 3–16.

[14] H. Le Cadre, C. Auliac, *Energy Demand Prediction in a Charge Station : A Comparison of Statistical Learning Approaches*, in European Electric Vehicle Congress 2012, Brussels, Belgium, November 19-22, 2012, 1-8.

[15] , M. Majidpour, C. Qiu, C. Chung, P. Chu, R. Gadh, *Fast Demand Forecast of Electric Vehicle Charging Stations for Cell Phone Application*, in PES General Meeting, Conference & Exposition, National Harbor, MD, USA, July 27-31, 2014.

[16] T. Cheng, J. Wang, X. Li, *A Hybrid Framework for Space–Time Modeling of Environmental Data*, Geographical Analysis, *ISSN 1538-4632*, 43(2011), 188–210.

[17] G.E. Box, G.M. Jenkins. *Time series analysis: forecasting and control*, ISBN 0-8162-1104-3, San Francisco, Holden-Day, 1976.

[18] T. Kohonen, *Self-organized formation of topologically correct feature maps*, Biological Cybernetics, *ISSN 1432-0770*, 43(1982), 59–69.

[19] P.J. Rousseeuw, *Silhouettes: A graphical aid to the interpretation and validation of cluster analysis*, Journal of Computational and Applied Mathematics, *ISSN 0377-0427*, 43(2011), 53–65.

[20] R.J. Hyndman, *forecast: Forecasting functions for time series and linear models*, R package version 8.1, 2017.

[21] R.J. Hyndman, Y. Khandakar, *Automatic time series forecasting: the forecast package for R*, Journal of Statistical Software, *ISSN 1548-7660*, 26(2008), 1-22.

Authors

Quentin De Clerck graduated in 2014 of a Master of Science in Applied Sciences and Engineering: Computer Science at the Vrije Universiteit Brussel and is since 2014 a Research Associate, under the supervision of Lieselot Vanhaverbeke, for the research group MOBI (Mobility, Logistics and Automotive Technology). His fields of expertise are electric vehicles, location analysis for charging infrastructure and Total Cost of Ownership.

Robert van den Hoed is Applied Professor Energy and Innovation at the Amsterdam University of Applied Sciences (AUAS), and is coordinator of the CleanTech research program. Research topics include electric mobility, analysis and development of charging infrastructures and smart grids.

Lieselot Vanhaverbeke is Assistant Professor at the Vrije Universiteit Brussel, at the department of Business Technology and Operations (BUTO) and the research group Mobility, Logistics and Automotive Technology (MOBI). She teaches Operations Research and Research Methods. Her research on location analysis, consumer mobility and economic aspects of electric vehicles is published in academic ISI journals. She is member of the Belgian Operational Research Society (ORBEL) and The Institute for Operations Research and the Management Sciences (INFORMS).