

# A comprehensive data analysis for aggregate capacity forecasting in Vehicle-to-Grid applications

Luca Patanè<sup>\*</sup>, Francesca Sapuppo<sup>†§</sup>, Maria Gabriella Xibilia<sup>‡</sup>  
Dipartimento di Ingegneria, Università degli Studi di Messina, Messina, Italy  
ORCID: <sup>\*</sup> 0000-0002-5488-9365, <sup>†</sup> 0000-0001-8772-2759, <sup>‡</sup> 0000-0001-7723-2051

Giuseppe Napoli

Consiglio Nazionale delle Ricerche (CNR), Istituto di Tecnologie Avanzate per l'Energia (ITAE), Messina, Italy  
ORCID: 0000-0003-0730-8641

**Abstract**—The integration of electric vehicles (EVs) into the Vehicle-to-Grid (V2G) framework necessitates accurate forecasting of the aggregated available capacity (AAC). This paper introduces a method for extracting AAC values from generic mobility data and other readily available data sources obtaining a comprehensive dataset that encompasses geographical information through GPS coordinates, meteorological data, and the consideration of national holidays and weekends. Different predictive models based on machine learning, such as Neural Networks, Long Short-Term Memory networks, Tree Ensembles, and Gaussian Process Regressors were implemented. The application of such models on different multidimensional datasets seeks to enhance the precision of capacity predictions, explaining and modeling the complex dependencies within the input data and the target available capacity, ultimately improving the efficiency and reliability of V2G systems. As the prevalence of EVs continues to rise, the proposed methodology offers a valuable tool for grid operators, energy planners, and stakeholders to optimize the integration of EVs into the broader energy landscape and facilitates the strategic planning of aggregator hub locations.

**Index Terms**—Vehicle-to-Grid, data-driven predictive model, time-series prediction, machine learning, aggregator hub planning

## I. INTRODUCTION

Vehicle-to-Grid (V2G) enables the bidirectional flow of electrical energy between electric vehicles (EVs) and the power grid, turning EVs into mobile energy storage units that can contribute to the stability and reliability of the grid [1]. Given the increasing penetration of electric vehicles EVs into the worldwide market (60% by 2030 [2]), the smooth integration of EVs into the V2G system is becoming a crucial issue. For V2G system to be reliable and economically viable for providers, an accurate prediction of aggregated available capacity (AAC) connected to specific aggregator hubs is required. Such a prediction depends on the availability of EVs to supply energy through a V2G hub when needed, which in turn can be affected by various factors: user behavior, location of the charging infrastructure [3], and vehicle characteristics. In addition, including meteorological data, alongside calendar

information (weekends and national holidays) might significantly influence the accuracy of AAC predictions. Meteorological factors, such as weather conditions and temperature fluctuations, might play a crucial role in determining the usage behavior of EVs. Unfavorable weather conditions, for example, can affect driving behavior [4] and thus affect the availability of EVs for grid services. In addition, weekends and national holidays lead to different time patterns in the use of EVs, as travel habits and energy consumption behavior often deviate from typical weekday routines [5]. In the literature, Machine Learning (ML) models, i.e. Persistence model, Generalized Linear Model, Artificial Neural Networks (ANNs) [6], Long-Short-Term-Memory network (LSTM) [7], K-Means Clustering [8], and Convolutional Neural Networks (CNN) [9], [10], have been applied to predict AAC based on different types of data sources: EVs fleets with a limited number of vehicles [11], [12], aggregated availability based on charging point data [6], historical vehicle information [13], calendar [8]. Very little literature presents the usage of mobility pattern [14], [15] and weather information [6] oriented to V2G AAC. This work aims to exploit the combination of such diverse data and use it as a basis for the implementation of ML prediction algorithms. In this way, the derived models are trained and tested on joint data sources, including real-world mobility data of vehicles from publicly available databases, meteorological information, and weekend and holiday data for the prediction of AAC in V2G applications. This work therefore includes pre-processing algorithms to obtain aggregated data on available capacity in space and time domain based on mobility and GPS information and fuzzy inputs combining holiday and weekday information. A comparison of the performances of the different models is also presented to explain the influence of the input characteristics on the prediction of the models and to gain a more comprehensive understanding of the factors that influence the available capacity of EVs. This will enable the exploration of a more efficient and reliable integration strategy of EVs into the wider energy ecosystem. The paper sections are organized as follows: data collection, followed by data pre-processing and analysis across diverse datasets, leading to model description and optimization techniques, culminating in the presentation of model prediction results.

This work was funded by the MASE - **Consiglio Nazionale delle Ricerche** within the project RICERCA DI SISTEMA 22-24 -21.2 Progetto Integrato Tecnologie di accumulo elettrochimico e termico. CUP Master: B53C22008540001, UNIME-DI-RdS22\_24: J43C23000670001

<sup>§</sup> Corresponding Author

## II. DATA COLLECTION

### A. Vehicle Dataset

The Vehicle Energy Dataset (VED) [16] is a comprehensive, open-access dataset of fuel and energy-related information collected from 383 individual vehicles in Ann Arbor, Michigan, USA. This open-access dataset includes GPS records of vehicle routes and time-series data on fuel consumption, energy consumption, speed, and auxiliary energy use. The dataset covers a wide range of vehicles

- 264 internal combustion engines (ICEs),
- 92 hybrid electric vehicles (HEVs),
- 27 plug-in hybrid electric vehicles/electric vehicles (PHEV/EVs)

operating in real-world conditions from November 2017 to November 2018.

In this work, a partial set of the features contained in the VED dataset was processed to obtain the AAC and other aggregated information for V2G applications. Specifically, the following features were selected: DayNum, VehId, Trip, Timestamp, Latitude, Longitude with a sampling time of 3s. The Daynum represents the difference in days between each record and November 1, 2017; VehID is a vehicle identifier; Trip is an identifier for the vehicle's trip; Timestamp represents the amount of time since the start of each trip; Latitude and Longitude are the GPS information expressed in degree. In addition, the HV Battery State of Charge (SoC[%]) is provided for the PHEVs and EVs with a sampling time of 60s.

### B. Meteo Dataset

The meteorological data was extracted from the MeteoStat database using the Python API [17]. The information on precipitation, temperature, and wind speed was extracted on an hourly basis for the period under investigation.

### C. National Holidays Dataset

State office closings for state holidays are regulated by the Michigan Department of Civil Service Regulation 5a.08. Public Act 124 of 1865 is the Michigan law governing official state holidays [18], [19]. Non-business days, as in Table I, were considered in conjunction with weekend information to obtain a comprehensive holiday rate to be incorporated into the model.

## III. DATA PRE-PROCESSING, AND ANALYSIS

### A. Vehicle Dataset

1) *Trips and Stops Extraction*: The couple of fields Vehicle ID and Trip identify a series of records belonging to a specific trip for a vehicle. Data are pre-processed to extract

- the covered distance for each trip
- the stops, intended as the time interval between two consequent trips

For each trip, the covered distance is determined by the Haversine formula [20]: it is the shortest distance over the earth's surface distance between two points. Therefore, the total distance for each trip is determined as the sum of

TABLE I  
OFFICE OF RETIREMENT SERVICES (ORS) NON-BUSINESS DAYS

Non-Business Day	Date
Weekends	Saturdays and Sundays
Thanksgiving Day	23 Nov 2017
Day after Thanksgiving	24 Nov 2017
Christmas (Eve and Day)	24-25 Dec 2017
New Year (Eve and Day)	31 Dec 2017 - 1 Jan 2018
Martin Luther King, Jr. Day	15 Jan 2018
President's Day	19 Feb 2018
Memorial Day	28 May 2018
Juneteenth	19 Jun 2018
Independence Day	4 Jul 2018
Labor Day	3 Sept 2018
Columbus Day	8 Oct 2018
Veterans Day	12 Nov 2018

the calculated great-circle distance between two points at sequential timestamps.

2) *Stop Maps and Hubs Selection*: An initial analysis of the mobility data was carried out by plotting the stops on the global map to identify geographical points suitable as V2G hubs. Figure 1 shows a preliminary analysis of the mobility data, showing the density of stops during different intervals of the day and integrated over the entire year of data. The color scale represents the duration of the stops in hours, assuming a minimum stop of 30 minutes. A circle area with a  $r = 1km$  radius was determined around two candidate points as is shown in Figure 2(a). Further insights into the hub areas are provided by the satellite images in Figures 2(b) and (c) to view the typology of urban structures and infrastructure within the area of interest.  $Hub_1$ , given the city center location and the vehicles' stop density, was chosen as the best candidate.

3) *State of Charge (SoC) Simulation*: Information on electrical features such as the SoC is only available for PHEVs and EVs. As a simplifying assumption and given the future projection for the EVs market, the ICEs and HEVs were assumed to be EVs contributing to the V2G logic. For simplicity, they were assumed to be EVs of the same make and model (Nissan Leaf with 40kWh battery capacity). SoC for ICEs and HEVs is calculated as an indirect measure of distance traveled and charging stop intervals according to [9], [11]. It is based on the following parameters:

- Battery charge at the beginning of the simulation:  $SoC = 100\%$
- The minimum state of charge that must be maintained to meet the requirements of vehicle users when calculating the available capacity:  $SoC_{min} = 50\%$
- The energy consumption per kilometer traveled by a vehicle:  $0.2km/kWh$
- Rapid charging hour rating of a vehicle, typically using DC power in the period from 7 a.m. to 7 p.m.:  $50kW$
- Slow-charging hour rating of a vehicle in the time interval from 7 a.m. to 7 p.m.:  $6kW$
- Efficiency of the charging process taking losses into account:  $90\%$
- Power rating of the export to the grid:  $50kW$

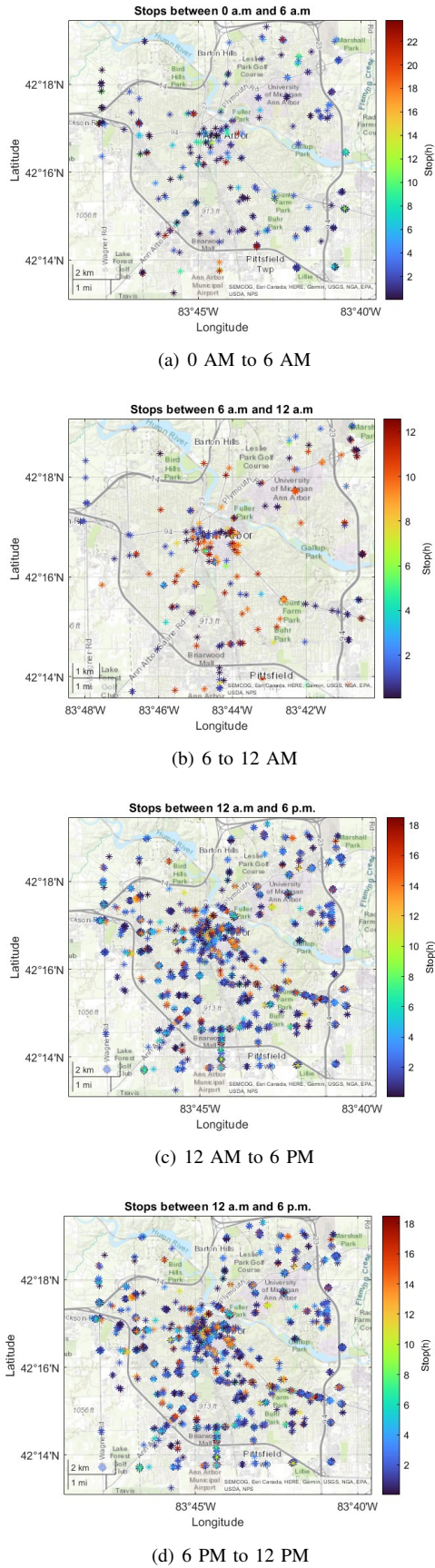


Fig. 1. Stop maps in different time intervals of the day, integrated over the entire year. The stop events have started in the time window: (a) from 0 AM to 6 AM, (b) from 6AM to 12 AM (c) from 12 AM to 6 PM, and (d) from 6 PM to 12 PM.

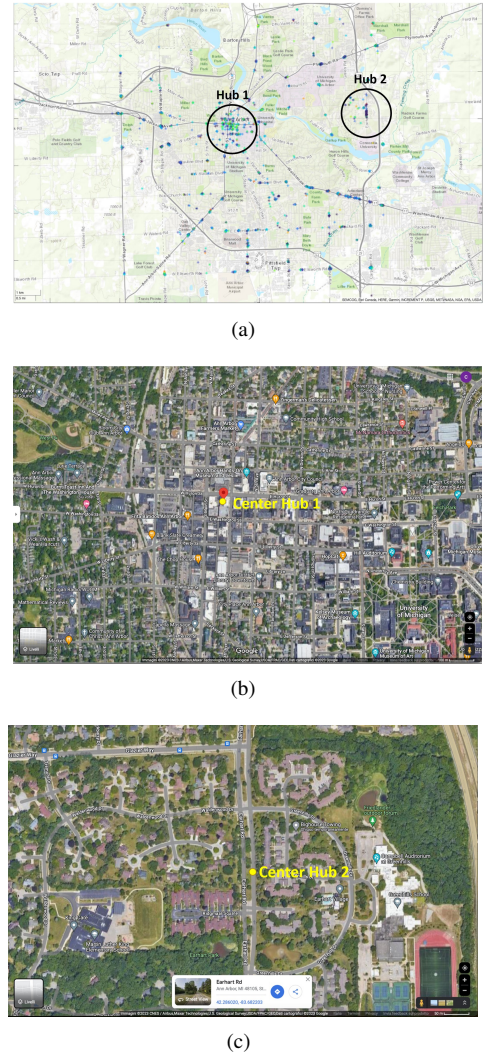


Fig. 2. Selection of the Aggregation Hubs in the Ann Arbor Area: (a) Entire zone of interest with the selection of  $hub_1$  and  $hub_2$  areas. Satellite view of (b)  $hub_1$  area in the city center and university zone, (c)  $hub_2$  area in a residential area near a school zone.

4) *Data Aggregation*: Trips and stop mapping were used to create an aggregated dataset in space and time with a 30-minute-interval time series related to the V2G  $hub_1$  to feed into the dynamic prediction model. The real or simulated  $SoC_v$  is used to determine the Available Capacity of a vehicle ( $AC_v$ ). It is defined as the capacity of each vehicle to provide energy to the grid in a half-hour ( $hh$ ) period.

$$AC_v^{hh} = \text{Max}(SoC_v^{hh-1} - SoC_{min}, 0) * BC \quad (1)$$

where the  $BC$  is the battery capacity. Equation 1 is considered for available vehicles. The aggregation in space refers to the assumption that the vehicles parked within a specific radius  $r$  from a selected V2G hub would connect to it. A vehicle parked within  $r$  from the hub in a 30-minute interval, and respecting the  $SoC_{min}$  requirement, is considered available ( $av^r$ ) to feed energy into the V2G system in that interval. The target feature to be predicted is the Aggregated Available Capacity

( $AAC_{hub^r}^{hh}$ ) in a half-hour interval and within  $r$  distance from the hub. It is defined in Equation 2.

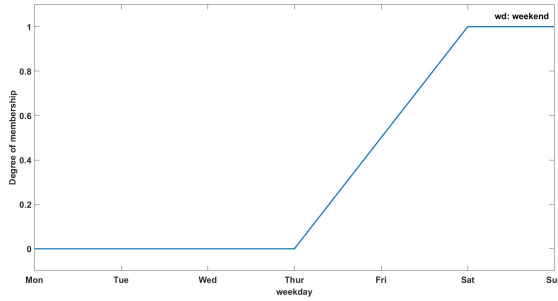
$$AAC_{hub^r}^{hh} = \sum_{av^r} AC_v^{hh} \quad (2)$$

Additional considered aggregated time series are the following:

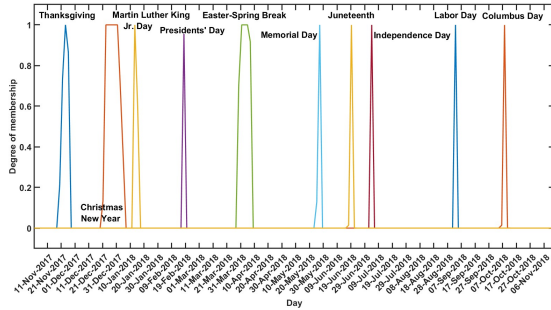
- Mean Hub Distance ( $MHD_{hub^C}^{hh}$ ): the mean distance from the hub center  $C$  of all vehicles parked in the global area of interest in the VED data
- Stop Duration Integral ( $SDI_{hub^r}^{hh}$ ): the sum of the parking stop of the available vehicles in the 30-minute time interval

### B. Meteo Dataset

The pre-processing of the meteo information involved the imputing of the missing data. The missing data were replaced with the average of the signal in the same week.



(a) Weekend



(b) National Holidays

Fig. 3. Membership function for the fuzzification of the holiday rate. (a) Weekend membership. (b) National holiday membership functions.

### C. National Holidays Fuzzy Set

The information about national holidays and weekends (Table I) is represented by a discontinuous time series unsuitable for dynamic models. The objective of the fuzzification of such inputs is to obtain a continuous time series including both weekends and national holiday information. Fuzzy membership functions were generated by considering the effect that weekends and holidays could have on drivers' habits in the previous and successive days. The membership functions for

the weekend and holidays, Figures 3(a) and (b) respectively, were both applied to the original dates, and the maximum value was taken to obtain a single continuous dynamic feature comprehensive for both information.

## IV. MODELS

Different models have been adopted and compared: Neural Networks (NNs), Long Short-Term Memory (LSTM) Networks, Ensemble of Trees, Gaussian Process Regression, and Trees. The regression models with a sequence output were employed to train the models for time series forecasting. In this setup, the  $AAC_{hub^r}$  target consists of training sequences with values shifted by one-time step. At each time step in the input sequences, the models learn to forecast the value of the subsequent time step.

### A. Architecture and Hyperparameters

Model optimization and hyperparameter selection were performed via the Bayesian optimization algorithm.

1) *NN*: NNs were implemented as fully connected hidden layers with Rectified Linear Activation Function (ReLU). The input layer depended on the size of the input set, the output layer had the size of one. The hyperparameters to be optimized were: the number of fully connected layers, the layer size, and regularization strength ( $\Lambda$ )

2) *LSTM*: An LSTM network is a type of recurrent neural network (RNN) that processes input data by iterating over time steps and updating the RNN state. The RNN state retains information from all preceding time steps. The utilization of a sequence-to-sequence LSTM neural network allows one to predict future values in a time series or sequence based on preceding time steps as input [7]. A simple LSTM architecture was applied: a sequence input layer, with size depending on the number of input data features, LSTM hidden layers with ReLU activation and dropout implementation to avoid overfitting, a fully connected layer with the size of one, for  $AAC_{hub^r}$  prediction, a regression layer. The hyperparameters to be optimized are the LSTM depth ( $LSTM_{Depth}$ ), the number of hidden units ( $NumHiddenUnits$ ), and the dropout probability ( $DropoutLayer$ ).

3) *Tree Ensemble (TE)*: Ensemble models combine results from many weak learners into one high-quality ensemble model. The hyperparameters to be optimized are method (Bag, LSBoost), number of learners, learning rate, minimum leaf size, and number of predictors to sample.

4) *Gaussian process regression (GPR)*: The hyperparameters to be optimized are: Basis function (zero, constant, linear), Kernel function (Rational Quadratic, Squared Exponential, Matern 5/2, Matern 3/2, and Exponential), Sigma.

5) *Tree*: The hyperparameter to be optimized is: Minimum Leaf Size. The Surrogate decision split was set to OFF.

## V. RESULTS

Different model identification and prediction were performed for three different input selections:

- Input Set 1 (*ISI*):  $AAC_{hub^r}$



TABLE II  
MODELS PERFORMANCE COMPARISON FOR DIFFERENT INPUT SETS (IS1, IS2, IS3)

Inputs	Model	RMSE(Valid.)	$R^2$ (Valid.)	RMSE(Test)	$R^2$ (Test)
IS1: $AAC_{hub_r}$	NN, 2 Layers (122,2)	3.5382	0.81	<b>3.6918</b>	<b>0.83</b>
	LSTM, 2 Layers (233,233)	<b>2.2161</b>	<b>0.88</b>	4.68798	0.51
IS2: $AAC_{hub_r}$ , $MHD_{hub_C}$ , $SDI_{hub_r}$	NN, 3 Layer (298,54,6)	3.2536	0.84	<b>3.3004</b>	<b>0.86</b>
	LSTM, 3 Layers (278,278,278)	<b>1.6818</b>	<b>0.94</b>	4.6262	0.58
IS3: Meteo, Holidays, $AAC_{hub_r}$ , $MHD_{hub_C}$ , $SDI_{hub_r}$	NN, 3 Layers (50,100,100)	3.3065	0.84	<b>3.2745</b>	<b>0.87</b>
	LSTM, 2 Layers (299,299)	<b>2.1989</b>	<b>0.88</b>	3.8875	0.68
	TE (LSBoost, n.learners=257)	3.1575	0.85	3.2819	0.86
	GPR (Linear,Non Isotr.Rat.Quadratic)	3.2731	0.84	3.3093	0.86
	Tree (Min Leaf Size: 15)	3.3508	0.83	3.5102	0.84

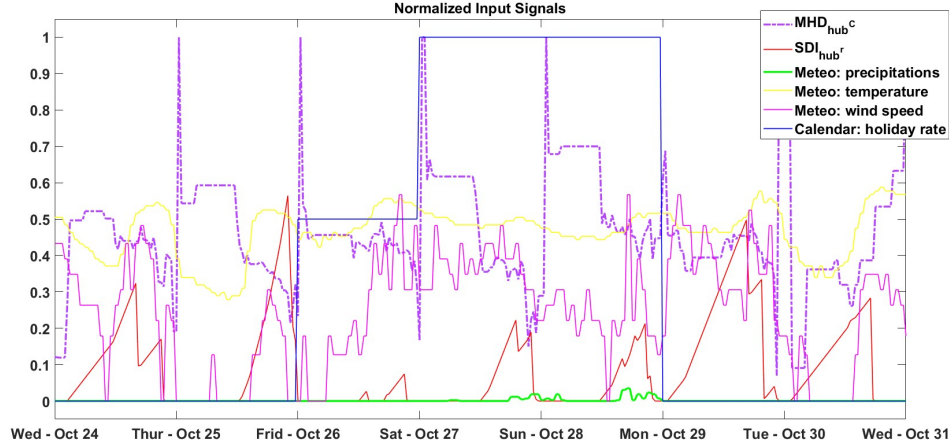


Fig. 4. Input signals normalized to the 0-1 range for the test week: Oct. 24th to Oct. 31st, 2018

- Input Set 2 (IS2):  $AAC_{hub_r}$ ,  $MHD_{hub_C}$ ,  $SDI_{hub_r}$
- Input Set 3 (IS3):  $AAC_{hub_r}$ ,  $MHD_{hub_C}$ ,  $SDI_{hub_r}$ , meteo information (precipitation, temperature, wind speed), holiday rate

Data was divided into training and validation (75%) and testing (25%) datasets. In particular, to avoid seasonal bias and unbalanced datasets, 3 weeks per month were chosen as training/validation, 1 week as test, and holiday weeks were included both in the training and test phases.

#### A. Bayesian optimization

The model optimization for hyperparameter determination was implemented via a Bayesian algorithm on the training data. The optimization metric was based on the minimization of the root mean square error (RMSE). The optimal hyperparameters and associated model performances are shown in Table II in terms of RMSE and correlation coefficient ( $R^2$ ) for both the validation and test datasets. The hyperparameters for the selected models applied to different input sets are:

1) IS1: NN (n.layers=2, layer size= 122, 2,  $\Lambda = 7.59$ ), LSTM (LSTMDepth=2, NumHiddenUnits= 233,233, DropoutLayer=0.253)

2) IS2: . NN (n. layers=3, layer size= 298, 54, 6,  $\Lambda = 0.018$ ), LSTM: LSTMDepth=3, NumHiddenUnits= 278,278,278, DropoutLayer=0.3913

3) IS3: . NN (n. layers=3, layer size= 50, 100, 100,  $\Lambda = 0.0$ ), LSTM (LSTMDepth=2, NumHiddenUnits= 299,299, DropoutLayer=0.1044), TE (method=LSBoost, n. of learners=257, learning rate=0.0578, min. leaf size=4, n. of predictors to sample=8); GPR (Basis fun.=linear, Kernel fun.=Nonisot. Rational Quadr., Sigma=78.65; Tree: Min. Leaf Size= 15, Surrogate decision split= OFF).

The NN and LSTM models were applied to all three input sets, while the TE, the GPR, and the Tree were only applied to the complete input set IS3, which resulted in being the most promising. The best-performing model in the validation dataset is the LSTM using the IS2: it has an RMSE of 1.6818 and  $R^2$  of 0.94. The best-performing model in the test dataset is the NN using the full input dataset IS3: it has an RMSE of 3.2745 and an  $R^2$  of 0.87. The autoregressive models, i.e. the ones using only the  $AAC$  at the previous step as input, are the worst-performing: it means that the dynamics of the system are affected by the exogenous inputs, thus confirming the validity of the proposed holistic methodological framework.

#### B. Aggregated Available Capacity Prediction

Figure 4 shows the input time series normalised in a range of 0 – 1 for a selected test week (Wednesday, 24 October to 31 October 2018). Figure 5 shows the 30-minute ahead prediction of AAC in the same week with (a) NNs and (b) LSTMs. It

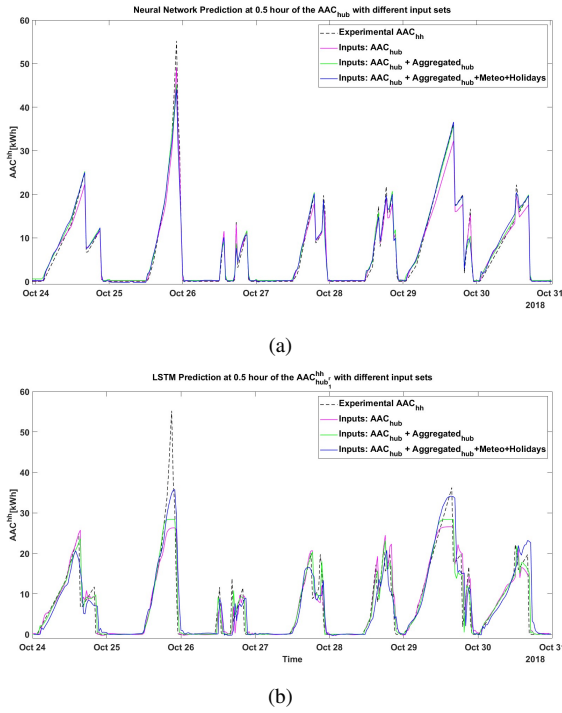


Fig. 5. 30-minute ahead prediction for  $AAC_{hub1r}$  using different input sets to the optimized models for the week in the test dataset Oct. 24th to Oct. 31st, 2018. (a) Neural Networks and (b) LSTM.

can be seen how the  $SDI$  signal correlates with the  $AAC$  forecast. The  $MHD$  shows peaks associated with hh intervals when there are no cars parked in the entire region analysed. The holiday rate shows positive values that coincide with the weekend and correlate strongly with a decrease in  $AAC$ . The comparison between the predicted time series in 5 (a) and (b) confirms that the NNs outperform the LSTMs in the test dataset. Furthermore, they both qualitatively confirm how the prediction improves when the full input set  $IS3$  is used.

## VI. CONCLUSIONS

This paper presents a novel approach to improve the accuracy of aggregate available capacity prediction in V2G systems. A distinctive contribution to the existing literature lies in the comprehensive integration of geographical, meteorological and temporal factors, in particular a fuzzified input set that considers both national holidays and weekends. By incorporating these contextual elements, our methodology provides a more holistic understanding of the complex dependencies within the V2G system. Further studies can be developed in this direction by using Explainable AI to interpret the impact of each input feature on the prediction. In addition, general mobility data was used to obtain aggregated time series, thus deviating from the traditional practice of using application-specific datasets. The application of our forecasting framework can be adapted to general mobility data to support the strategic placement of aggregator hubs and optimize their distribution to effectively manage and utilize the available capacity of geographically dispersed EVs. This

strategic planning contributes to the overall scalability and adaptability of the V2G infrastructure. These innovations not only extend the applicability of our method to generic data, but also highlight its versatility in predicting available capacity and demonstrate that it can be adapted for various EV applications beyond V2G.

## REFERENCES

- [1] S. Habib, M. Kamran, and U. Rashid, "Impact analysis of vehicle-to-grid technology and charging strategies of electric vehicles on distribution networks – a review," *J. of Power Sources*, vol. 277, pp. 205–214, 2015.
- [2] I. E. Agency, "By 2030 evs represent more than 60% of vehicles sold globally, and require an adequate surge in chargers installed in buildings," in *IEA, Paris*, 2022.
- [3] J. Dixon, W. Bukhsh, K. Bell, and C. Brand, "Vehicle to grid: driver plug-in patterns, their impact on the cost and carbon of charging, and implications for system flexibility," *eTransportation*, vol. 13, p. 100180, 2022.
- [4] V. Bakhshi, K. Aghabayk, N. Parishad, and N. Shiwakoti, "Evaluating rainy weather effects on driving behaviour dimensions of driving behaviour questionnaire," *Journal of Advanced Transportation*, vol. 2022, pp. 1–10, 3 2022.
- [5] T.-H. T. Gim, "Sem application to the household travel survey on weekends versus weekdays: the case of seoul, south korea," *European Transport Research Review*, vol. 10, p. 11, 3 2018.
- [6] J. Graham and F. Teng, "Vehicle-to-grid plug-in forecasting for participation in ancillary services markets," in *2023 IEEE Belgrade PowerTech*. IEEE, jun 2023.
- [7] F. Curreri, L. Patanè, and M. G. Xibilia, "Rnn- and lstm-based soft sensors transferability for an industrial process," *Sensors*, vol. 21, no. 3, 2021.
- [8] D. Perry, N. Wang, and S.-S. Ho, "Energy demand prediction with optimized clustering-based federated learning," in *2021 IEEE Global Communications Conference (GLOBECOM)*, 2021, pp. 1–6.
- [9] R. Shipman, R. Roberts, J. Waldron, C. Rimmer, L. Rodrigues, and M. Gillott, "Online machine learning of available capacity for vehicle-to-grid services during the coronavirus pandemic," *Energies*, vol. 14, no. 21, 2021.
- [10] S. Li, C. Gu, J. Li, H. Wang, and Q. Yang, "Boosting grid efficiency and resiliency by releasing v2g potentiality through a novel rolling prediction-decision framework and deep-lstm algorithm," *IEEE Systems Journal*, vol. 15, no. 2, pp. 2562–2570, 2021.
- [11] R. Shipman, R. Roberts, J. Waldron, S. Naylor, J. Pinchin, L. Rodrigues, and M. Gillott, "We got the power: Predicting available capacity for vehicle-to-grid services using a deep recurrent neural network," *Energy*, vol. 221, p. 119813, 2021.
- [12] H. S. Nogay, "Estimating the aggregated available capacity for vehicle to grid services using deep learning and nonlinear autoregressive neural network," *Sustainable Energy, Grids and Networks*, vol. 29, p. 100590, 2022.
- [13] S. Li, C. Gu, J. Li, H. Wang, and Q. Yang, "Boosting grid efficiency and resiliency by releasing v2g potentiality through a novel rolling prediction-decision framework and deep-lstm algorithm," *IEEE Systems Journal*, vol. 15, no. 2, pp. 2562–2570, 2021.
- [14] M. Schläpfer, H. J. Chew, S. Yean, and B.-S. Lee, "Using mobility patterns for the planning of vehicle-to-grid infrastructures that support photovoltaics in cities," *ArXiv*, vol. abs/2112.15006, 2021.
- [15] T. Zeng, S. Moura, and Z. Zhou, "Joint mobility and vehicle-to-grid coordination in rebalancing shared mobility-on-demand systems," *IFAC-PapersOnLine*, vol. 56, no. 2, pp. 6642–6647, 2023, 22nd IFAC World Congress.
- [16] G. Oh, D. J. Leblanc, and H. Peng, "Vehicle energy dataset (VED), a large-scale dataset for vehicle energy consumption research," *IEEE Trans Intel Trans Sys*, vol. 23, no. 4, pp. 3302–3312, 2022.
- [17] "Meteo stat python api," <https://dev.meteostat.net/python/>.
- [18] "Ann arbour school breaks," <https://annarborwithkids.com/articles/winter-spring-break-schedules-2022/>.
- [19] "Michigan ors non-business days," <https://www.michigan.gov/psru/ors-non-business-days>.
- [20] N. R. Chopde and M. Nichat, "Landmark based shortest path detection by using A\* and haversine formula," *Int J Adv Res Comput Comm Eng*, vol. 1, no. 2, pp. 298–302, 2013.