POLITECNICO
MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

Executive Summary of the Thesis

A Two-Step Integrated Framework for Evaluating the Impact of Electric Mobility

Laurea Magistrale in Electrical Engineering - Ingegneria Elettrica

Author: Davide Fratelli

Advisor: Prof. Marco Merlo

Co-advisor: Corrado Maria Caminiti

Academic year: 2023-24

---

1. Introduction

The increasing penetration of electric vehicles (EVs) demands a thorough analysis of their impact on the distribution grids. This issue is addresses in this work with a two-pronged approach in a geographically referenced framework: first the flow of EVs is estimated and located and then integrated in a model of the distribution network. The results of these two elaborations provide the data necessary to investigate qualitatively the impact onto the power system.

A comprehensive review on the literature dealing with the two issues is necessary to understand the current state of the art in the research. Jia et al. [1] train a Machine Learning (ML) model for the estimation of the regional penetration of EVs in Texas based on a Federal Unites States public dataset, while Li et al. [2] provide an extensive review of the factors affecting the adoption of an EV.

Regarding synthetic grids (SGs), the most relevant and thorough papers (Pisano et al. [3] and Oneto et al. [4]) present two end-to-end frameworks which, starting from public input data, detail procedures to generate georeferenced grids: these two works provide the reference backbone for the techniques adopted and the procedure developed in this thesis.

The development of a procedure for the analysis and elaboration of traffic data was instead discussed in a published work [5] which provides the foundation for part of this thesis.

This work details a novel framework for assessing the impact of e-mobility primarily based on public data. The two stand-alone procedures de-

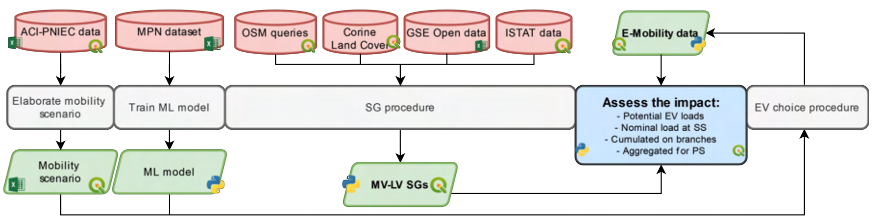


Figure 1: Flowchart of the integrated procedure

tailed allow for many future adoptions in the fields of e-mobility and power system research.

2. Methodology

The framework is structured in two stand-alone procedures: the first, called *vehicle choice procedure* (or EV choice), returns a model of EV flows. The second, *synthetic grid procedure*, returns a geographical model of the distribution networks serving a particular area. A SG is a model representing a proxy the distribution network obtained from public data: the aim is to have a dataset which is open, replicable and thus not subject to the confidentiality imposed by grid operator models. The objective is thus to best replicate the behaviour and location of the real grid, with no particular interest in having a perfect double of it. Additionally, SGs could also provide a model of the optimal network design to serve a particular area.

The two outputs are then integrated to allow for a qualitative evaluation of the impact of e-mobility, as reported in Figure 1.

2.1. Input data and preliminary elaborations

2.1.1 Vehicle choice

A dataset of car trips, geographically and temporally referenced, constitutes the foundation for the framework. For future elaborations, it is necessary that the dataset provides information on the trips itself such as time of the departure, distance travelled and duration. Furthermore, some other features such as day of the trip (mainly to distinct between weekdays and weekends) allow for some optional considerations. Socio-economic geo-referenced data must also be gathered, from now on referred to as census data, necessary to characterize the drivers. The granularities of the geographical references should be as detailed as possible while suited to allow matching the trips with the correspondent census data.

A mobility scenario indicating the distribution of EVs on the territory must be adopted. In case of analysis on future data, since no such detailed data is likely to be available, a procedure for the elaboration is detailed. The necessary data for this is indicatively the current municipal EV penetration data and future country-wide projected overall number of EVs.

Lastly, it is convenient to *rescale* the test dataset (following the same algorithm adopted in the training steps [6]) to address likely structural differences with respect to the target trips dataset, mainly regarding trips distances.

2.1.2 Routing of a Synthetic Grid

The datasets representing the buildings, the roads and the PS location in the relevant areas are necessary, along with the census data already introduced above. Moreover, a layer providing information on land use is necessary for a better characterization of the grid and the correct implementation of the framework. Ultimately the electrical data, such as installed generating units and conventional areas, while not strictly essential, allow to design a model closely resembling the real grid.

After the input data is gathered, several manipulations are performed to make sure the datasets are consistent and ready for the procedure.

2.2. Machine Learning model

The learning dataset is published by the Netherlands Mobility Panel (MPN) ([7] MPN survey, 2020, wave 8), chosen for the large number of trips and the suitability of its variables [8]. The ML model is tailored to the variables which are identified in the relevant literature as being the most correlated to the adoption of an EV.

Particularly, the model classifies the trips based on variables for both the driver's features and those of the trip itself: these were investigated in

Variable	**Implementation source**
Gender	ISTAT
Status of employment	ISTAT
Highest completed education	ISTAT
Age category	ISTAT
Trip purpose	Traffic elaboration
Length of the trip	Traffic elaboration
Main transport mode	Traffic elaboration
Departure time frame	Traffic elaboration
Travel duration	Traffic elaboration
Weekday or weekend	Traffic elaboration
Gross aggregate household income	ISTAT
Household owns EV	*Procedure output*

Table 1: ML selected variables

the literature review and are reported in Table 1. Several tests performed to achieve the best accuracy suggest to train two separated *Random Forest* models, on outward and return-home trips: a further characterization of the models based on the purpose for outward trips slightly improves the accuracy.

For each section in the relevant area, a distribution for each of the identified socio-economic variables in the model is created (with data provided by ISTAT [9]), reflecting the likely hood that for a user, each of its feature belongs to a category. Then, each trip is matched to its departure section and a category is assigned for each of the driver features, based on a random extraction from the distributions. This allows to have the target dataset which shares an identical structure to the learning trips, thus allowing the model to recognise the coding of each regressor.

2.3. EV choice procedure

The ML model is implemented on the geo-referenced target trips dataset, which are thus characterized by a likely hood of being electric. This probability is the key to the vehicle selection step. With reference to the departure section, the algorithm cycles through all the cities and assigns the EVs to the trips, starting from the most likely to be electric: this stops if the city does not have any more EV to assign (the pool of vehicles, with reference to the elaborated scenario, is exhausted) or there are no more trips to characterize.

The output of this implementation is the dataset of trips performed with an EV or alternatively the entire trips dataset characterized by their adoption of an electric or conventional vehicle.

2.4. Synthetic grids procedure

The procedure is mainly inferred from open data, with the exception of the load estimation task, which uses data from a Distribution System Operator (DSO) to build a linear regression.

As a first step, the census sections are divided based on a simplified density-based classification aimed at having three homogeneous layers with similar *urbanization patterns*. The SG procedure is run in parallel for the three layers up to the routing, which instead joins all the elements back into one layer. For each census section, a linear *regression* trained on both numerical and categorical regressors estimates the power and demand and splits it among the aggregated load points within, roughly corresponding to the buildings selected.

A case study provided also a chance to design a way to adjust the procedure to estimate the entire LV load, since the regression specifically addresses residential load. This was done by finding the ratios between total and residential yearly consumptions, provided by TERNA [10] at a province level (most granular data available).

A contemporaneity factor is assigned to each load, to replicate the connection to a transformer of load larger than its nominal size. These points are subject to a *multi-step clustering* procedure, which simulates the aggregation of loads connected to a same Secondary Substation (SS). Several techniques are adopted, namely *knapsack model*, *DBSCAN* and *K-means* clustering.

Two instances of *routing* take place, separately on MV and LV: the algorithms aim at building a quite realistic grid by taking into account the road network and the urbanization which constrain its development (underground or overhead routing). For every set of points to route, road and straight line paths are computed: a parameter selected by the user is introduced in a weighted comparison of the paths lengths to penalize a particular itinerary for each of the three levels of urbanisation.

A further review on the literature relevant to the validation of synthetic grids is conducted to select and design a set of metrics for the validation of the framework.

Input	**Source**
Training dataset	MPN survey
EV distribution	ACI data
Future scenarios	PNIEC
Census sections	ISTAT (National Statistic Agency)
PS location	OSM
Conventional areas	GSE
Buildings	OSM
Conventional generators and DGs	AtlaImpianti [11]
CLC	Copernicus project[12]
Roads	OSM

Table 2: Framework inputs

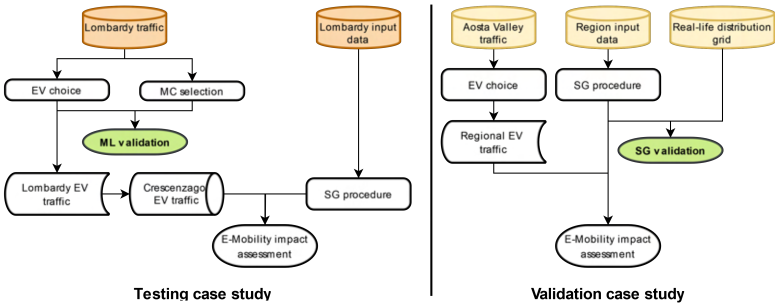


Figure 2: Flowchart of the validation and testing procedures

3. Implementation datasets

The proposed framework is implemented in two case studies, summarized in the flowcharts in Figure 2. Firstly it is validated for the Valle d'Aosta region, allowing for a comparison between the real grid (RG) and the SG. Moreover, the EV choice procedure is run for the entire Lombardy region on a projected mobility scenario in 2030: then, the results are extracted for the conventional area of Crescenzago, in the north of Milan, and used as input to assess the impact of e-mobility after the SG procedure is run for this smaller area.

3.0.1 Vehicle choice

The dataset of trips in Lombardy, published in the form of an Origin Destination (OD) matrix [13], is elaborated into a dataset of geographically and temporally referenced entries, implementing the model of the traffic flow in the region proposed in [5]. This dataset constitutes an input for this procedure.

As a preliminary step, a mobility scenario is elaborated by combining current municipal EV penetration data from ACI [14] and future country-wide adoption mandated by the National Plan for Energy and Climate (PNIEC) [15]. This is done by computing the share of EVs in each city with respect to the national total, assuming as a necessary simplification that this share will be quite constant in the coming years: the future scenario is defined by applying this distribution to the total EVs projected.

3.0.2 Synthetic grids

The datasets representing the buildings, the roads and the PS location in the relevant areas are obtained with a query on OpenStreetMaps (OSM); the census sections are downloadable from the ISTAT's publications [9], as for the EV choice procedure; the Corine Land Cover (CLC), a layer providing information on Land Use, can be found on the website of the Copernicus project [12]; finally, the electrical data, such as generators and conventional areas, can be found on GSE (Gestore Servizi Energetici). After the input data is gathered, several manipulations are performed to make sure the datasets are consistent and ready for the procedure. Mainly, the accuracy of the layers queried from OSM is subject to notoriety and population of the area, thus a manual inspection of the datasets is strongly advised.

The necessary input data is summarised in Table 2.

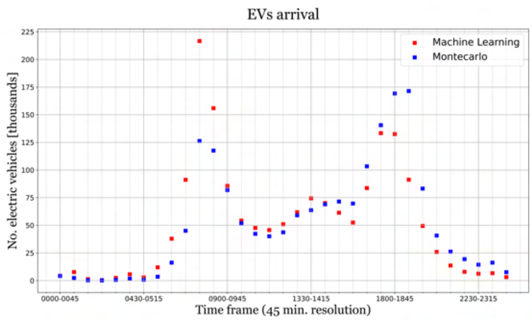
4. Implementation results

The results hereby presented are a combination of the most significant takeaways for each of the two study cases.

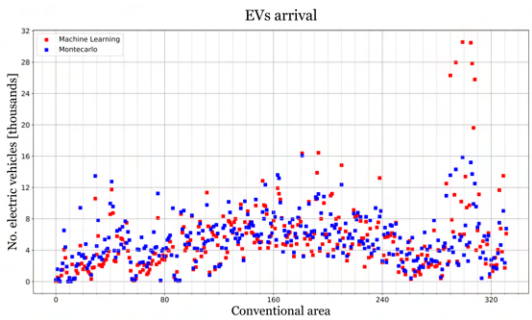
4.1. E-mobility

The plots represented in Figure 3 show a comparison of the spatial and geographical distribution of the same numbers of electric trips selected by this procedure with those drawn as 2.5% of the trips with a Monte Carlo (MC) approach: the ML model provides a further level of independence from the traffic data, selecting more often as electric trips those in rush-hours and in some specific areas, awarding particularly those which are recognised as work trips.

Focusing on the Crescenzago conventional area, the analysis on the mobility is carried out by extracting the data from the traffic elaborations run for Lombardy region, particularly only those trips arriving in one of the sections inside of the



(a) Total arrivals for the same time frames



(b) Arrivals for the same conventional area

Figure 3: Temporal and geographical comparison of regional EV arrivals

conventional area. Since the EV choice models were applied on a regional scale, despite setting an equal number of EVs to be extracted, the number of electric trips selected in the relevant sections are significantly different. The ML model in fact selected more trips in this area.

Figure 4 shows the metric indicating the difference in daily arrivals comparing ML and MC methods (particularly, ML arrivals - MC arrivals): it can be appreciated that the former are comparatively more concentrated in areas of particular significance, such as large factories, where the metric is positive.

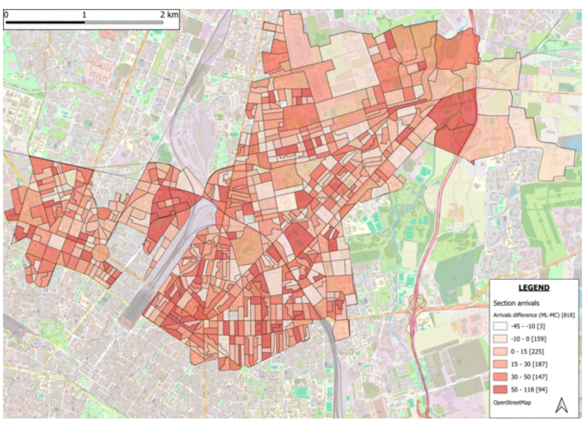


Figure 4: Differential of arrivals in the Crescenzago area according to the two models

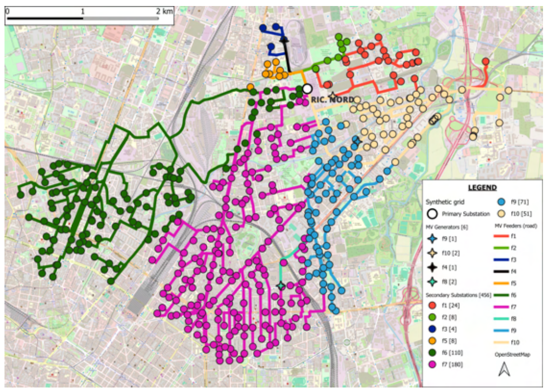
4.2. Synthetic grid

The MV and LV SGs resulting from the procedure are shown in the Figures in 5.

4.3. Assessing the impact of E-Mobility

While the temporal dimension is not quite addressed in this work (trips are aggregated on a daily basis), the geo-referenced nature of the data allows to superimpose the mobility layer onto the SGs generated by the framework, thus obtaining the number of EVs which could potentially request a connection to a PS (conventional area granularity), a single SS (SS granularity) or eventually to a same branch, after the *potential loads* are cumulated upstream each SS. This methodology allows to obtain a rich multi-faceted qualitative assessment of the impact of e-mobility on the distribution grid.

Figure 6 shows an example of the possible elabo-



(a) Overview of the MV distribution network

(b) Combined view Voltage distribution network generated in Crescenzago

Figure 5: Distribution network in Crescenzago

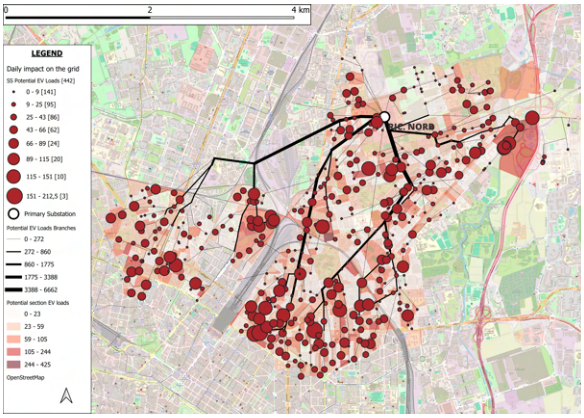


Figure 6: Combined view of most impacted branches and SSs (dimension criteria

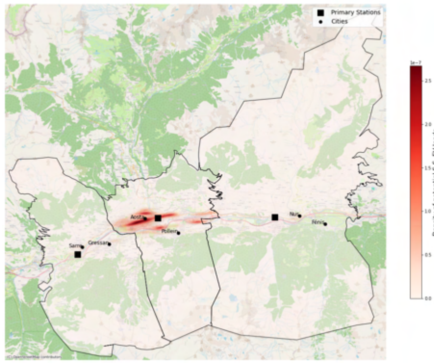
ration and presentation of these results, particularly the combined impact on branches and SSs. Examples of heatmaps of potential EV loads are provided in the upcoming section.

4.4. Validation of the synthetic grid

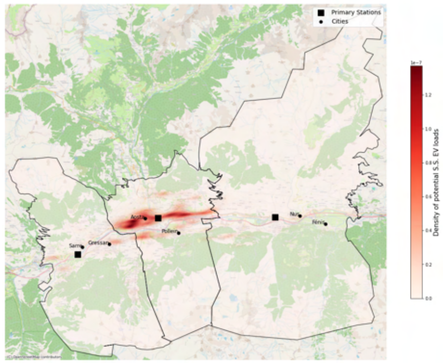
The outputs of the procedure implemented for the region of Valle d'Aosta are compared to the model of the grid operated by the main DSO of the region. The availability of a real-life model allows to validate the SG procedure by comparing some significant results for the output SG and the RG. Table 3 shows the comparison of the grid metrics analysed in the literature for the MV and LV grids. The rationale for the comparison is not to evaluate the absolute errors in the metrics, since the real life grid is the result of a long development process characterized by historical and contingent features which make it non-ideal. Conversely, the aim of the comparison is to grasp the plausibility of the synthetic grid with respect to the real one.
Finally, pictures in Figure 7 provide a visual comparison of the impact of e-mobility on the three most impacted conventional areas in the region for the real and synthetic grids.

Metric	RG	SG
Size [km]	4 214	4 582
Aerial conductor [km]	2 268	2 744
Underground conductor [km]	1 946	1 838
Cable ratio	0.86	0.67

Table 3: Comparison of routing metrics



(a) Impact on the real grid



(b) Impact on the synthetic grid

Figure 7: Impact heatmap of potential EV loads

5. Conclusions

This thesis proposes a novel integrated framework based primarily on open-data which is able to first locate EVs on the territory by employing a ML model capable of grasping the geo-temporal characterization of the phenomena. Secondly, an end-to-end procedure for the creation of a geo-referenced synthetic grid based mainly on inference from public data is proposed. The generated SG is used as a proxy for the real distribution grid, thus allowing to assess the impact of electric mobility on the power system, particularly on feeders and SSs.

6. Acknowledgements

Many thanks to Deval for providing the data at the foundation of much of this work.
Moreover, this publication makes use of data from the Netherlands Mobility Panel, which is administered by KiM Netherlands Institute for Transport Policy Analysis.

List of Acronyms

Acronym	Meaning
EV	Electric Vehicle
SS	Secondary Substation
PS	Primary Substation
DSO	Distribution System Operator
PNIEC	Piano Nazionale Integrato per l'Energia e il Clima
OD	Origin-Destination
OSM	OpenStreetMaps
ML	Machine Learning
MPN	Netherlands Mobility Panel
ACI	Automobile Club d'Italia
MC	Monte Carlo
LV	Low Voltage
MV	Medium Voltage
GSE	Gestore Servizi Energetici
CLC	Corine Land Cover
SG	Synthetic grid
RG	Real grid

References

[1] Jianmin Jia, Baiying Shi, Fa Che, and Hui Zhang. Predicting the Regional Adoption of Electric Vehicle (EV) With Comprehensive Models. *IEEE Access*, 8:147275–147285, 2020.

[2] Wenbo Li, Ruyin Long, Hong Chen, and Jichao Geng. A review of factors influencing consumer intentions to adopt battery electric vehicles. 78:318–328, October 2017.

[3] Giuditta Pisano, Nayeem Chowdhury, Massimiliano Coppo, Nicola Natale, Giacomo Petretto, Gian Giuseppe Soma, Roberto Turri, and Fabrizio Pilo. Synthetic models of distribution networks based on open data and georeferenced information. *Energies*, 12(2323):4500, January 2019.

[4] Alfredo Ernesto Oneto, Blazhe Gjorgiev, Filippo Tettamanti, and Giovanni Sansavini. Large-scale inference of geo-referenced power distribution grids using open data. *TechRxiv*, December 2023.

[5] Ghaffar Yousefi, Aleksandar Dimovski, Lucio Radaelli, and Marco Merlo. Estimating the impact of electric mobility on distribution networks through GIS techniques. *Sustainable Energy, Grids and Networks*, 38:101379, April 2024.

[6] Xi Hang Cao, Ivan Stojkovic, and Zoran Obradovic. A robust data scaling algorithm to improve classification accuracies in biomedical data. *BMC Bioinformatics*, 17(1):359, September 2016.

[7] MPN Data. https://www.mpndata.nl/. (Accessed:10-06-2024).

[8] Sascha Hoogendoorn-Lanser, Nina T. W. Schaap, and Marie-José OldeKalter. The Netherlands Mobility Panel: An Innovative Design Approach for Web-based Longitudinal Travel Data Collection. *Transportation Research Procedia*, 11:311–329, January 2015.

[9] Istat. https://www.istat.it/. (Accessed:23-08-2024).

[10] Terna SpA. Pubblicazioni Statistiche. https://www.terna.it/it/sistema-elettrico/statistiche/pubblicazioni-statistiche. (Accessed: 11-10-2024).

[11] GSE. AtlaImpianti. https://atla.gse.it/atlaimpianti/project/Atlaimpianti_Internet.html. (Accessed:07-08-2024).

[12] Copernicus Project. Corine Land Cover. https://land.copernicus.eu/en/products/corine-land-cover. (Accessed:05-08-2024).

[13] Matrice Origine/Destinazione: i dati sulle abitudini di spostamento in Lombardia. https://www.regione.lombardia.it/wps/portal/istituzionale/HP/DettaglioServizio/servizi-e-informazioni/Imprese/Imprese-di-trasporto-e-logistica/ser-matrice-od-infr/matrice-od. (Accessed:27-06-2024).

[14] Automobile Club d'Italia. Aci - parco veicoli. https://www.aci.it/. (Accessed:15-07-24).

[15] National Energy and Climate Plans. https://commission.europa.eu/energy-climate-change-environment/implementation-eu-countries/energy-and-climate-governance-and-reporting/national-energy-and-climate-plans_en. (Accessed:15-07-2024).