





Data-driven Model-free Hosting Capacity Estimation in a Low Voltage Node

Estimación de Hosting Capacity en un nodo de bajo voltaje basada en datos sin modelo de red

Luis Daniel Andagoya-Alba¹ , Johanna Jara² , Pablo Catota³ , Rommel Valencia⁴ 

¹ Instituto Tecnológico Superior Universitario Rumiñahui, luis.andagoya@ister.edu.ec, Quito, Ecuador

² Instituto Tecnológico Superior Universitario Rumiñahui, johanna.jara@ister.edu.ec, Quito, Ecuador

³ Instituto Tecnológico Superior Universitario Rumiñahui, pablo.catota@ister.edu.ec, Quito, Ecuador

⁴ Instituto Tecnológico Superior Universitario Rumiñahui, rommel.valencia@ister.edu.ec, Quito, Ecuador

Corresponding author: luis.andagoya@ister.edu.ec, andagoyaalba@gmail.com

ABSTRACT

Currently, distributed generation systems have been proposed to increase the energy's potential in electrical grids, however, the insertion of this form of energy generation can cause negative effects, like overvoltages, that can exceed the allowed levels by the local operator. For this reason, it is necessary to set a limit for the injection of distributed generation into the network. This parameter is called Hosting Capacity, which is used to restrict the power coming from distributed sources, while maintaining the normal functioning of the network. Some studies have been developed to define and calculate the Hosting Capacity, mainly based on scenario simulations, however, these need detailed and accurate network models which at low voltage levels are very difficult to have. This work focuses on determining a Hosting Capacity value in a low voltage node, using values from a smart meter. The proposed methodology implements a lineal regression model considering voltage and power consumption parameters. The results were validated through a network model obtaining voltage values lower to the limit, therefore, concluding that the proposed model is useful for calculating Hosting Capacity without employing a network model.

Keywords: Hosting Capacity, Data-driven analyses, Linear regression model, Distributed energy resource

RESUMEN

En la actualidad, los sistemas de generación distribuida han sido desarrollados para incrementar el potencial energético en las redes eléctricas, sin embargo, la introducción de esta forma de generación de energía puede ocasionar efectivos negativos, como sobrevoltajes, que exceden los niveles permitidos por el operador local. Por este motivo, es necesario establecer un límite para la inyección de generación distribuida en la red. Este parámetro, denominado Hosting Capacity, ha sido desarrollado para limitar la potencia de fuentes distribuidas manteniendo el funcionamiento apropiado de la red. Algunas investigaciones han definido y calculado el Hosting

Capacity mediante simulaciones de escenarios, esto implica definir modelos de red detallados y precisos que a niveles de bajo voltaje son difíciles obtener. El presente trabajo se enfoca en determinar el valor de Hosting Capacity en un nodo de bajo voltaje, utilizando datos obtenidos de un medidor inteligente. En la metodología propuesta se implementa un modelo de regresión lineal que considera parámetros de voltaje y potencia consumida. Los resultados se validaron mediante un modelo de red obteniendo valores inferiores al límite, lo que permite concluir que el modelo implementado es útil para el cálculo de Hosting Capacity sin utilizar modelo de red.

Palabras clave: Hosting Capacity, Análisis de datos, Modelo de regresión lineal, Recursos energéticos distribuidos

1. INTRODUCTION

In the coming years, distributed generation systems will increase the energy's potential of those countries that are taking appropriate measures to change the energy matrix and introducing renewable energies as an alternative, especially solar energy, as a global trend (Esau et al., 2023). To meet these challenges, the implementation of ways to monitor and manage distribution networks are required, which leads to the analysis of smart grids designed to take into account the increase of photovoltaic systems, as a contribution to distributed generation (Munikoti et al., 2022). Distributed Energy Resource (DER) can increase the voltage levels in distribution networks if their connection is not monitored. To provide a solution to this problem, regulating devices are used like load tap changers, helping to maintain the voltage level at the desired values (Huo et al., 2023).

Therefore, the need for a new network parameter, called Hosting Capacity (HC), has been evaluated to assess the ability of the system to manage the parameters of the distributed generation load without exceeding the permissible limits of the operational performance of the distribution networks (Castelo de Oliveira et al., 2020; Fatima et al., 2020; Qamar et al., 2023a); thus, methods for the analysis and real-time monitoring of the network parameters, which are factors to be analyzed for the optimization of the power distribution system, should be evaluated. These analyses allow the combination of DER with new systems, such as electric vehicle charging and grid-connected photovoltaic generation systems (Silva & Vieira, 2022; Ismael et al., 2019).

Network models are the most commonly employed techniques to determine the parameters of a network. HC calculations rely heavily on network models, which are often incorporated into their design using tools such as traditional power flows and optimal power flow optimization. These calculations face various challenges, particularly when it comes to application scalability. Errors or missing data make it logistically difficult to update network topologies and the demands associated with them, which intensifies the problem in distribution networks.

One possibility to solve this problem is to calculate voltages without having the corresponding electrical models. Currently in distribution networks, Advanced Metering Infrastructure (AMI) is becoming increasingly prevalent and can be used to achieve this goal. The physical

relationships of the circuit can be determined by using linear and nonlinear regressions, focusing on measuring the distribution of data in multiple domains. In literature reviews, there are references to analyzed network models and it is not common to find accurate network models having an adequate amount of existing elements and variables in their structure. Bassi et al. (2021a) as Rajabi et al. (2022a) propose calculations of voltages without network model through neural network models, with the validation of historical data performed by energy meters, and also using low voltage feeders are analyzed methodologies based on hyperparameters in the feeders. As a consequence, the obtained results can demonstrate approaches of calculations that establish limits for the photovoltaic capacity in a distribution system. The evaluation of photovoltaic (PV) systems penetration using smart meters helps to reduce the error rate by 10 % in the estimation (Cunha et al., 2020).

The work of Bassi et al. (2022b) laid the groundwork for a model-free calculation method for electrical voltage determination, proposed by the University of Melbourne. The report also analyzes data from distribution network service providers (DNSPs) in Victoria. A proposed approach is presented here, which aims to calculate data-driven loads without predefined network models. This method is based on the analysis of historical smart meter data and can be interpreted accurately without being affected by excessive noise or erroneous measurements. The report includes several analyses to understand the received data, such as the distribution of the data provided by each DNSP, the number of data points in the raw value set, and the direction of active and reactive power in each quadrant. Overall, this paper provides a detailed description of the proposed methodology and analysis of the data obtained, thus laying the groundwork for future work on model-free activity projects.

The paper written by Qamar et al. (2023b) the world has seen a great shift toward renewable energy resources from conventional ones. But the ever-increasing integration of distributed generation (DG) presents an analysis of a network-free model for controlling low-voltage solar-abundant distribution networks. The method uses a neural network (NN) trained with historical smart meter data to capture the nonlinear relationship between historical smart meter data (P, Q, V). This paper presents a practical case study of an Australian low-voltage power supply with 31 single-phase solar customers. The results show that the calculated PV settings are comparable to traditional three-phase Optimal Power Flow (OPF) analysis. The proposed approach is a solution that does not require a detailed three-phase low-voltage network model, which makes it practical for distribution utilities to manage PV-rich low-voltage networks. In the same context, Bassi et al. (2022b) propose an HC analysis method by training a NN based on historical data obtained from smart meters assigned to each consumer of the analysis network. This model offers the use of DER for network analysis without the use of a network model, which is a great advantage since it is difficult to obtain an accurate network model at the low voltage distribution level. The model offers a preprocessing to eliminate abnormal data that can cause errors in the determination of the voltage parameters in the proposed scenario and then uses this

data to train a NN whose output variable will be the voltage at the node. Based on these results and the definition of the allowed voltage limits, the HC value of the analyzed network can be determined.

A new method for voltage calculation in low-voltage networks with a large number of residential PV systems is evaluated in Bassi et al. (2021b) voltage calculations can help determining specific settings (e.g., PV curtailment. The proposed method uses smart meter data and deep neural networks (DNNs) to calculate the voltage without requiring a full electrical model. The results show that this method can solve the problem of voltage spikes caused by domestic solar power generation systems and improve the quality of power supply. The work of Wu et al. (2023) tested a new distribution network HC analysis method called spatiotemporal deep learning, which uses deep learning techniques to calculate the (HC) in distribution systems. The paper proposes a deep learning model, called Long Short-Term Memory (LSTM), that is capable of “remembering” relevant data in a sequence and retaining it for several moments. Therefore, it can have as much short-term memory (as a basic recurrent network) as long-term memory. In Procopiou et al. (2020) voltage rise issues are expected to occur on the very infrastructure they are connected to: the low-voltage (LV) a method based on data obtained from smart meters is proposed, which allow a fast estimation of HC in low voltage networks and thus avoid complex studies based on network models. Using data collected from smart meters, a simple invariant regression model is developed to estimate the average active power of each customer, in particular the values that can cause the specified voltage limits to be exceeded. This calculation is used to determine the additional PV capacity that can be supported by the low-voltage grid. The proposed method is validated using real smart meter data obtained by simulating the gradual penetration of PV on Australian high-voltage feeders supplying 79 low-voltage grids. The results show that this analytical method can accurately estimate the HC and is a faster and simpler alternative to more complex methods based on network models greatly improved.

The present work focuses on determining a methodology for estimating an HC value in a low-voltage node using data from a smart meter that records the voltage and power demand values of a common consumer. The data is processed to determine a regression model that allows estimating future scenarios based on historical data of voltages and power demand. Finally, the proposed methodology is applied in a node of grid connection of the electrical load of the Instituto Superior Tecnológico Universitario Rumiñahui. This work differs from other studies in the following points:

The methodology is developed only with the power and voltage values of a consumer, scenarios with photovoltaic systems are not considered as in some previously detailed developments.

- The data used is only from the smart meter, no scenarios based on network models are simulated to complete the data.

- The proposed methodology is developed for a specific user with the possibility of scaling to a network if there is enough data.

The proposed methodology is developed below.

2. MATERIALS AND METHODS

The proposed methodology is composed of four processes: Data Acquisition, Data Processing, Predictive Model Development, and HC Estimation. It involves an internal data extraction, transformation, and loading (ETL) process, which was developed in Python programming language.

Figure 1. Data-Driven Hosting Capacity Estimation Flow chart

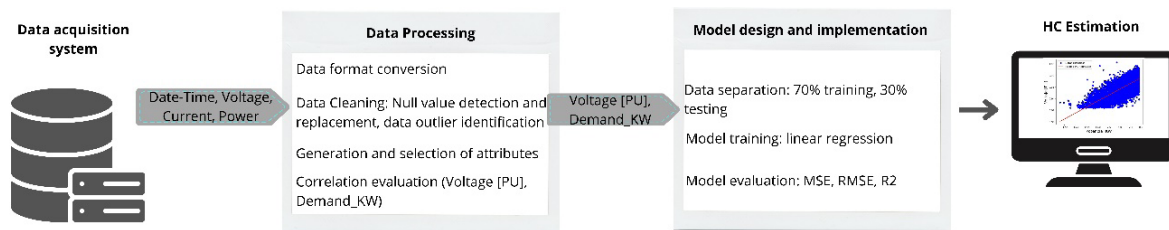


Figure 1 summarizes the proposed methodology for HC determination using PCC electrical parameter data. It is important to note that the proposed methodology involves an internal data extraction, transformation, and loading (ETL) process. Each one is detailed below.

1. *Data Acquisition.* A Data Acquisition (DAQ) located at a common connection point (PCC) of the network was used, the system stores measurements of electrical parameters (voltage, current, and consumed power) with the date and time at which it records each sample, during a period of time established as representative. The data is stored on a secure digital (SD) card in comma-separated value (CSV) format, which is referred to as the source file.
2. *Data Processing:* The source file format is verified and converted to facilitate its manipulation, and cleaning is performed to detect null and outliers that may alter the result of the project; this process culminates with the calculation or generation of the necessary attributes and evaluation of their correlation.
3. *Development of the Predictive Model.* To design the linear regression model () the selected attributes were loaded and separated into two data sets, where 70 % was used for training and 30 % for testing the model; where a and b are the coefficients that allow relating the dependent variable x , which corresponds to the power consumed in KW, with the independent variable y , which represents the voltage in per unit (p.u.) values. To evaluate the performance of the algorithm, the Mean Square Error (MSE), Root Mean Square Error (RMSE), R^2 , and Mean Absolute Error (MAE) metrics were used.
4. *HC Estimation.* With the determination of the model that describes the behavior of voltage levels at the point of analysis, it is possible to estimate values of power injected to the low

voltage distribution system for a given voltage level. In the case of HC, an active power value is determined for an overvoltage level established by the regulations in force at the analysis location.

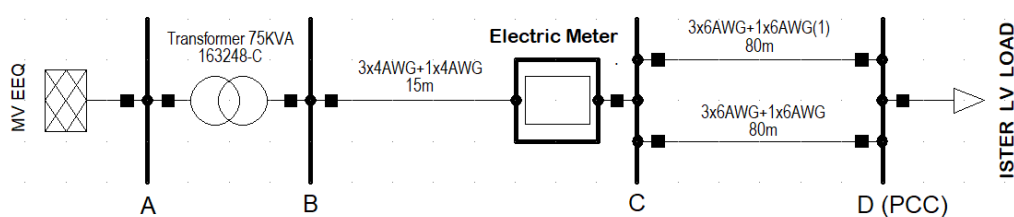
3. STUDY CASE

To validate the proposed methodology, the electrical load of the Instituto Universitario Rumiñahui (ISTER) was considered as case study. The data was acquired through a Data Acquisition System (DAQ) that records voltage, current, and power values. It was connected to the common connection point PCC, which in this case was the main distribution board of the ISTER. The system stores daily measurements in five-minute intervals, with electrical parameters (voltage, current, and power consumed), date and time at which each sample is recorded. This collection was performed for three months. The data is stored on a secure digital card (SD) in Comma Separated Value (CSV) format, which is called the source file.

MV–LV ISTER low voltage electrical system

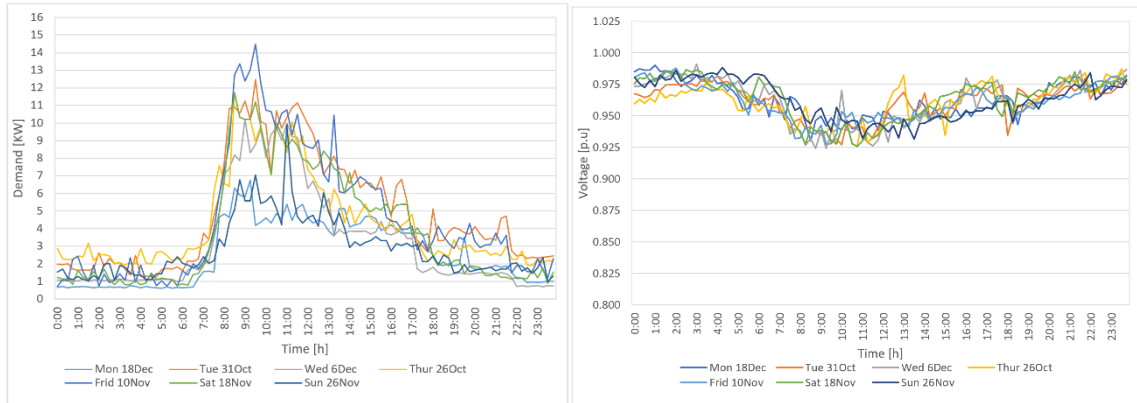
The used ISTER electrical system starts from the 22.8 KV bus that feeds the primary of the 75 KVA; 22.8 KV/220 V three-phase distribution transformer; the low voltage bus is connected to the general meter, and this in turn connects to the Main Distribution Board of the ISTER, considered the PCC for this study. The transformer tap is at step 1 with +1.5 % (line voltage of 223.4 V and single voltage of 129 V). Therefore, the base single-phase voltage considered is 129 V, which corresponds to voltage 1.00 p.u. The network is shown in Figure 2.

Figure 2. MV–LV ISTER low voltage electrical system



To validate the results obtained from the application of the methodology, a modeling of the ISTER network was carried out in the OpenDSS program, using the physical network data and the demand profiles obtained from the smart measurements. The measurements were performed from September 26 to December 23, 2023, with a period of 5 min. Figure 3 shows the demand profiles and voltages for each day of the week taken randomly within the analysis period.

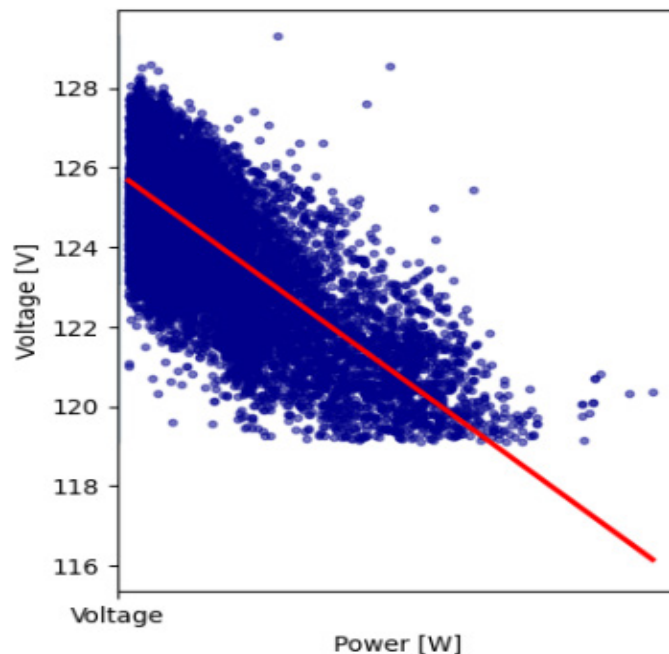
Figure 3. ISTER PCC demand and voltage profiles



Data acquisition

For the acquisition process, a DAQ was installed at a common connection point located on the Institute’s main distribution board. The data were stored in an SD memory in CSV format. The file contains measurements of electrical parameters (voltage, ampacity, and active power) in 5-minute periods carried out by the DAQ. The extraction process began by reading the CSV file with the read_csv function of the Python numpy library. Figure 4 shows the voltage (V) and power (W) data collected by the DAQ over three months and their trend.

Figure 4. Data acquired from the DAQ



Data processing

This process consists of preparing the data by converting formats, cleaning (detection of null and outlier values), and generating or calculating new attributes from existing ones. This process began by transforming the Date field of type Object into a datetime64[ns] to facilitate its manipulation. Table 1 shows the description and data type of the fields contained in the original extracted source file.

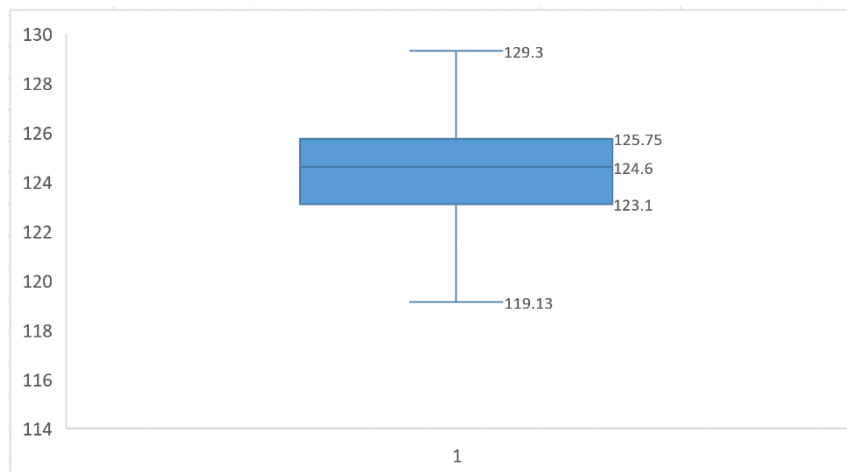
Table 1. Source file fields

Field	Description	Data Type
Date.	Indicates the date the meter records data.	Object
Time	Indicates the hour, minutes, and seconds in which the meter takes the data.	Object
Voltage	Voltage Measurement Parameter in Volts.	float64
Ampacity	Ampacity Measurement Parameter in Amperes.	float64
Power	Power Measurement Parameter in Watts.	float64

To perform data cleaning, the data was checked for null values with the `isnull()` function, however, it did not return empty fields. It was found that each day contains 288 measurements.

To identify outliers, the distribution of the data corresponding to Voltage was analyzed by comparing it with the 25th (123.1 V), 50th (124.6 V), and 75th (125.75 V) percentiles represented in a box plot. From Figure 5, it is obtained that no Voltage value is outside the limit lines (129.3 and 119.13V), therefore, there are no outliers.

Figure 5. Voltage box plot



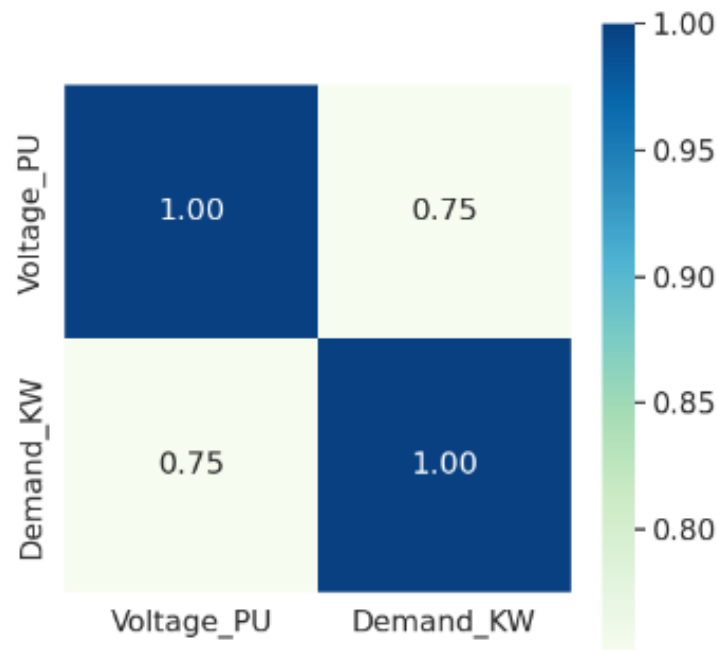
For the study, it was required to generate attributes such as Voltage in p.u. and Power consumed in KW. To change the Voltage parameter, 129 V was taken as a reference [1.00 p.u.].

In the study, to identify the power consumed, the range of negative numbers was used, and for the power that can be exported to the network, the range of positive numbers was used.

Because of the DAQ recorded power consumption values in Watts, it was converted to KW in the negative range. Power_Consumption_KW and Voltage_p.u fields were generated.

To complete the data processing, a correlation matrix was used, because it is a tool that allows identifying the variables that are related to each other. Figure 6 shows a correlation of 0.75 for the selected variables (Voltage in p.u. and power consumed in KW), which indicates that the two variables tend to increase or decrease at the same time, that is, they evolve in the same direction. Therefore, in the loading process, they will be the attributes sent for the design and implementation of the predictive model.

Figure 6. Correlation matrix

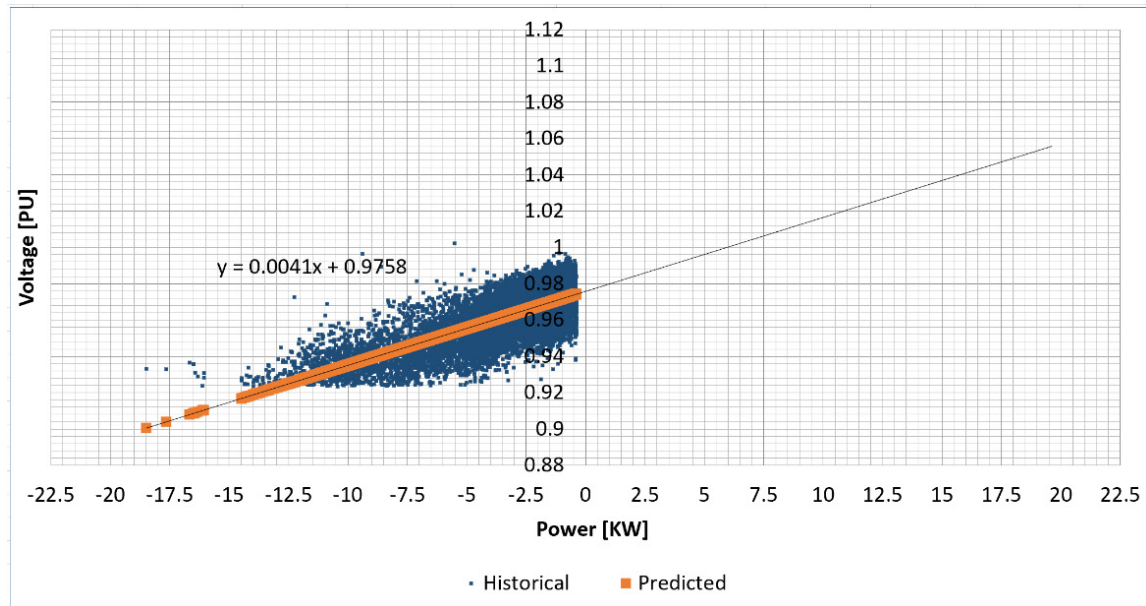


Development of the predictive model

The values obtained in the correlation matrix led to considering linear regression as an optimal model that can describe the behavior of the variables.

The design process began with the training of the linear regression model, the separation of data into training and test sets were carried out, in this case, they were divided into 70 % and 30 % correspondingly, using the LinearRegression() function of the sklearn library, that executes the process to obtain coefficients of the algorithm; being 0.00407 and 0.976. The equation that represents the regression model is , as shown in Figure 7.

Figure 7. Linear regression



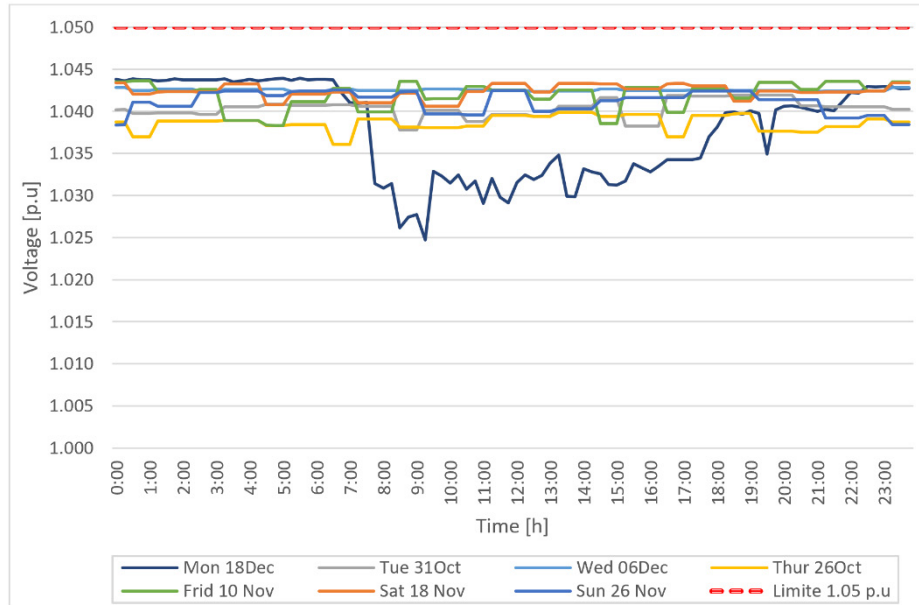
In the evaluation of the algorithm performance, the MSE values were obtained at 9.686×10^{-5} . This value is close to 0, which indicates how close the regression fit is to the observations. The RMSE had a result of 0.00984 and r^2 which indicates how well the model fits the real observations, which had a value of 0.569 and the MAE that indicates the average error is 0.00803.

By the IEEE ANSI C84.1 ($\pm 5\%$) and IEEE 1547-2018 (0.95 to 1.05) standards, it was parameterized that the overvoltage limit will be 1.05 p.u. For this reason, the HC calculation was carried out for an exact point, this being 1.05 p.u. to obtain the value of maximum power that could be injected into the network, in this case, 18,253 KW.

4. RESULTS AND DISCUSSION

The results obtained from the proposed methodology are used to model a DER system of 18,253 KW [maximum HC calculated], that is connected to the PCC to subsequently perform the analyses in time series of each day of the week, taken randomly from the considered test period. The network modeling and the simulation of power flows in time series were carried out in the OpenDSS software. Figure 8 shows the voltage profiles of the considered node, with the constant HC value determined.

Figure 8. Voltage profile in low-voltage node analyzed with 18,253KW of DER



It is shown that the HC value estimated through the proposed methodology allows the operation of the network under the determined operating limit of 1.05 p.u. in the node voltage, that is, no voltage violation occurs at any time of the day, which demonstrates the validity of the study. This demonstrates that the proposed HC estimation model was capable of providing an adequate HC estimation in the considered node associated with a low-voltage network only considering the data of the current behavior of the network.

Future Gaps

The use of other estimation models can be an interesting advance in the development of HC without network models or the use of current methodologies considering the dynamics of the network.

5. CONCLUSIONS

The linear regression model can be efficiently applied to determine a future scenario based on historical data that relates the variables that determine its behavior, in this case, the determination of a maximum value of Hosting Capacity that allows limiting the amount of generation distributed connected to a node.

The proposed methodology adequately approximated an HC value that allowed the user to remain within the overvoltage limits established by the applied regulations. Furthermore, the methodology can be scaled to several users with different characteristics as long as adequate data is available.

Network model-free analyses are considered a very useful tool for determining the behavior of an electrical system, especially those systems where it is difficult to have exact and detailed

network models, such as low-voltage distribution systems.

The proposed methodology considers the cleaning of the input data to determine the regression model as a very important process. This is important since in real systems some of the information may be lost or there may be errors in their registration, which would cause errors in HC estimation.

These methodologies are very useful for network operators in system planning for the integration of DER at the user level, considering current regulations and the network's limitations.

REFERENCES

- Bassi, V., Jaglal, D., Ochoa, L., Alpcan, T., & Leckie, C. (2022a). *Deliverable 0 «Concept, Smart Meter Data, and Initial Findings»*. <https://doi.org/10.13140/RG.2.2.29492.94089/1>
- Bassi, V., Jaglal, D., Ochoa, L., Alpcan, T., & Leckie, C. (2022b). *Deliverable 0 «Concept, Smart Meter Data, and Initial Findings»*.
- Bassi, V., Ochoa, L., & Alpcan, T. (2021a). *Model-Free Voltage Calculations for PV-Rich LV Networks: Smart Meter Data and Deep Neural Networks*. <https://doi.org/10.1109/PowerTech46648.2021.9494847>
- Bassi, V., Ochoa, L., & Alpcan, T. (2021b). Model-Free Voltage Calculations for PV-Rich LV Networks: Smart Meter Data and Deep Neural Networks. *2021 IEEE Madrid Power-Tech*, 1-6. <https://doi.org/10.1109/PowerTech46648.2021.9494847>
- Castelo de Oliveira, T. E., Ribeiro, P. F., Zobia, A. F., Abdel Aleem, S. H. E., & Ismael, S. M. (2020). An Overview of Hosting Capacity for Modern Power Grids. En A. F. Zobia, S. H. E. Abdel Aleem, S. M. Ismael, & P. F. Ribeiro (Eds.), *Hosting Capacity for Smart Power Grids* (pp. 1-9). Springer International Publishing. https://doi.org/10.1007/978-3-030-40029-3_1
- Cunha, V. C., Freitas, W., Trindade, F. C. L., & Santos, S. (2020). Automated Determination of Topology and Line Parameters in Low Voltage Systems Using Smart Meters Measurements. *IEEE Transactions on Smart Grid*, *11*(6), 5028-5038. <https://doi.org/10.1109/TSG.2020.3004096>
- de Silva, L. E. S., & Vieira, J. P. A. (2022). Combined PV-PEV Hosting Capacity Analysis in Low-Voltage Distribution Networks. *Electric Power Systems Research*, *206*, 107829. <https://doi.org/10.1016/j.epsr.2022.107829>
- Esau, Z., Ryoichi, H., & Hiroyuki, K. (2023). A flexible stochastic PV hosting capacity framework considering network over-voltage tolerance. *Energy Reports*, *9*, 529-538. <https://doi.org/10.1016/j.egy.2022.11.101>
- Fatima, S., Püvi, V., & Lehtonen, M. (2020). Review on the PV Hosting Capacity in Distribution Networks. *Energies*, *13*(18), Article 18. <https://doi.org/10.3390/en13184756>
- Guo, Y., Yuan, Y., & Wang, Z. (2022). Distribution Grid Modeling Using Smart Meter Data. *IEEE Transactions on Power Systems*, *37*(3), 1995-2004. <https://doi.org/10.1109/TPWRS.2021.3118004>
- Huo, Y., Li, P., Ji, H., Yu, H., Yan, J., Wu, J., & Wang, C. (2023). Data-Driven Coordinat-

- ed Voltage Control Method of Distribution Networks With High DG Penetration. *IEEE Transactions on Power Systems*, 38(2), 1543-1557. <https://doi.org/10.1109/TPWRS.2022.3172667>
- Ismael, S. M., Abdel Aleem, S. H. E., Abdelaziz, A. Y., & Zobaa, A. F. (2019). State-of-the-art of hosting capacity in modern power systems with distributed generation. *Renewable Energy*, 130, 1002-1020. <https://doi.org/10.1016/j.renene.2018.07.008>
- Munikoti, S., Abujubbeh, M., Jhala, K., & Natarajan, B. (2022). A novel framework for hosting capacity analysis with spatio-temporal probabilistic voltage sensitivity analysis. *International Journal of Electrical Power & Energy Systems*, 134, 107426. <https://doi.org/10.1016/j.ijepes.2021.107426>
- Procopiou, A., Liu, M., Ochoa, L., Langstaff, T., & Harding, J. (2020). *Smart meter-driven estimation of PV hosting capacity*.
- Procopiou, A. T., Liu, M. Z., Ochoa, L. F., Langstaff, T., & Harding, J. (2020). Smart meter-driven estimation of PV hosting capacity. *CIREN 2020 Berlin Workshop (CIREN 2020)*, 2020, 128-131. <https://doi.org/10.1049/oap-cired.2021.0287>
- Qamar, N., Arshad, A., Mahmoud, K., & Lehtonen, M. (2023a). Hosting Capacity in Distribution Grids: A Review of Definitions, Performance Indices, Determination Methodologies, and Enhancement Techniques. *Energy Science & Engineering*, 11. <https://doi.org/10.1002/ese3.1389>
- Qamar, N., Arshad, A., Mahmoud, K., & Lehtonen, M. (2023b). Hosting capacity in distribution grids: A review of definitions, performance indices, determination methodologies, and enhancement techniques. *Energy Science & Engineering*, 11(4), 1536-1559. <https://doi.org/10.1002/ese3.1389>
- Rajabi, A., Elphick, S., David, J., Pors, A., & Robinson, D. (2022). Innovative approaches for assessing and enhancing the hosting capacity of PV-rich distribution networks: An Australian perspective. *Renewable and Sustainable Energy Reviews*, 161, 112365. <https://doi.org/10.1016/j.rser.2022.112365>
- Taheri, S., Jalali, M., Kekatos, V., & Tong, L. (2021). Fast Probabilistic Hosting Capacity Analysis for Active Distribution Systems. *IEEE Transactions on Smart Grid*, 12(3), 2000-2012. <https://doi.org/10.1109/TSG.2020.3038651>
- Wu, J., Yuan, J., Weng, Y., & Ayyanar, R. (2023). Spatial-Temporal Deep Learning for Hosting Capacity Analysis in Distribution Grids. *IEEE Transactions on Smart Grid*, 14(1), 354-364. <https://doi.org/10.1109/TSG.2022.3196943>