Data in Brief 61 (2025) 111782

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Electricity demand from national interconnected system of paraguay and meteorological dataset

Félix Morales-Mareco[a,*], Matías Verdún[a], Diego H. Stalder[a], Carlos Sauer[a], Sebastián Grillo[b]

[a] *Facultad de Ingeniería, Universidad Nacional de Asunción, San Lorenzo 111421, Paraguay*
[b] *Facultad de Ciencia y Tecnología, Universidad Autónoma de Asunción, Asunción 001013, Paraguay*

ARTICLE INFO

Article history:
Received 21 March 2025
Revised 26 May 2025
Accepted 6 June 2025

Dataset link: Electricity consumption and meteorological data from Paraguay (Original data)

Keywords:
Electric power consumption
Weather
Forecasting
Machine learning

ABSTRACT

This article presents a comprehensive dataset integrating hourly electricity demand data from Paraguay's National Interconnected System (SIN) with meteorological records from 2008 to 2022. Electricity demand data were obtained from the National Electricity Administration (ANDE), while meteorological data were sourced from the Global Data Assimilation System (GDAS).
The dataset provides a long-term perspective on trends and patterns, with high-frequency data enabling detailed analyses. Electricity demand data comprise hourly readings measured in megawatts (MW), while the meteorological dataset includes temperature and humidity readings recorded at three-hour intervals. GDAS synthesizes observations from diverse sources, such as weather stations, satellites, and radar systems, and integrates them into a numerical weather forecasting model to deliver a consistent and detailed representation of atmospheric conditions.
This integrated supports robust analyses of the relationship between electricity demand and meteorological variables. Additionally, it offers valuable resources for testing and developing forecasting models, while contributing to the optimization of energy distribution, infrastructure planning,

* Corresponding author.
E-mail address: felixmorales@fiuna.edu.py (F. Morales-Mareco).
Social media: @FelixM69 (F. Morales-Mareco), @diego_sta (D.H. Stalder), @carlossauer (C. Sauer)

https://doi.org/10.1016/j.dib.2025.111782

and efficiency improvements in Paraguay's electricity network.

Specifications Table

Subject	Engineering & Materials science
Specific subject area	Energy. Electricity consumption and meteorological data.
Type of data	Table. Raw and processed.
Data collection	Raw electricity consumption data were obtained from a set of electricity metering spreadsheets provided by the National Electricity Administration (ANDE), and meteorological data were extracted from the Global Data Assimilation System (GDAS). Weather data contains the following variables: air temperature at 2 meters above the ground (Celsius), relative humidity at 2 meters above the ground (%), surface pressure (Pa), total precipitation over six hours (mm), eastward wind speed at 10 meters above ground (m/s), and northward wind speed at 10 meters above ground (m/s). The data corresponding to holidays in Paraguay are also attached.
Data source location	Electricity consumption data: Countrywide electric consumption. Meteorological data: Luque, Central Department, Paraguay (25°14′30.0"S, 57°31′08.0"W). This locality has been chosen as the point to obtain meteorological data because it is located in a strategic area, where the largest population and industrialization of the country is concentrated. At the same time, since Paraguay has a relatively small territory, the temperature variability in the Eastern Region, which comprises 97% of the total population, is small. This explanation is developed further in the Data Description section.
Data accessibility	Repository name: Electricity consumption and meteorological data from Paraguay. Data identification number: 10.17632/dm78h95ghk.3 Direct URL to data: https://data.mendeley.com/datasets/dm78h95ghk/3
Related research article	[1]

1. Value of the Data

- The data are invaluable for analyzing and understanding electricity consumption patterns, including trends, seasonal variations, and peak demand periods.
- The dataset enables the development of robust forecasting models that integrate historical power consumption with external factors such as weather conditions, economic activity, and calendar variables.
- Because the dataset is made up of global electricity consumption data and not individual consumption data for each department or region, it can serve as a parameter in the development of more specific models for a single location.
- The dataset serves as a benchmark for testing and validating predictive models in the field of energy demand forecasting.

2. Background

Accurate energy demand forecasting is essential for ensuring the stability and resilience of power grids. By predicting electricity requirements, utilities can proactively manage resources to guarantee uninterrupted power supply [1,2]. Historical data on energy consumption, combined with insights into influencing factors, serves as an invaluable resource for this process [3].

The analysis of historical consumption data reveals trends, seasonal variations, and peak demand periods, which are critical for creating reliable forecasts. These forecasts mitigate the risk of power shortages and blackouts. Additionally, understanding influencing factors such as weather patterns, economic activity, and population demographics provides deeper insights into the drivers of energy demand. This understanding enables us to develop forecasting models that account for these external influences, leading to more robust predictions.

Reliable demand forecasts are also vital to support informed policy and infrastructure planning. Policymakers and energy companies can use them to guide planning for future infrastructure investments, such as building new power plants or transmission lines, and developing targeted energy efficiency for demand response programs for managing peak demand periods. Moreover, utilities can optimize resource allocation by strategically deploying a mix of power sources to meet the expected demand while balancing cost-effectiveness and environmental sustainability.

3. Data Description

The data presented in this paper were compiled from prior work addressing the challenge of short-term electricity demand forecasting [1]. The data coverage period is from 2008 to 2022, and were sourced from documents provided by ANDE, the institution responsible for electricity demand management and monitoring in Paraguay, as well as from the GDAS, which integrates mathematical weather forecasting models based on historical observations. The published files include the following:

1. consumption-raw.csv
2. weather-raw.csv
3. consumption.csv
4. weather.csv
5. holidays.csv
6. complete-dataset.csv

First, we have the consumption-raw.csv file, which contains historical electricity consumption records at the national level. On the other hand, the weather-raw.csv file includes several weather indicators that can be correlated to energy consumption. Furthermore, consumption.csv and weather.csv are the output files resulting from a rigorous processing stage applied to consumption-raw.csv and weather-raw.csv, respectively. The holidays.csv file provides information about holidays throughout the study period, which can significantly impact energy demand patterns.

The provided complete-dataset.csv file integrates all acquired data after rigorous processing to eliminate data corruption. Due to the implementation of daylight-saving time in Paraguay, the dataset reflects seasonal UTC changes between summer and winter. Specifically, on the day of the time change, the 00:00 timestamp is omitted-data jumps directly from 23:00 of the preceding day to 01:00 of the following day. This adjustment, results in intentional gaps in the time series that users should consider when analyzing temporal trends.

Therefore, the details of the fields included in the complete processed data file are presented in Table 1 below.

The information contained in these files is intended to support other researchers and scientific groups in the development of competent knowledge for the efficient management of electricity in Paraguay. To emphasize the approach of this work, a more detailed explanation of the complete dataset is explained in the following sections.

3.1. Electricity consumption data

The following subsection presents an exploratory analysis of the electricity consumption variable from the dataset (SIN). The goal is to summarize the characteristics of the electricity

Table 1
Detailed summary of the fields included in the complete-dataset.csv file and the source files from which these fields were extracted. The abbreviations shown correspond to the following, CR: consumption-raw.csv file, WR: weather-raw.csv file, H: holidays.csv file.

Field	Description	Unit of measurement	Data type	File Source
DATETIME	Timestamp of data recorded	Date and time, including offset according to ISO 8601 standard	Datetime	-
SIN	SIN electricity demand	MW	Float	CR
T02M	Air temperature at 2 meters above the ground	°C	Float	WR
RH2M	Relative humidity at 2 meters above the ground	%	Float	WR
PRSS	Surface pressure	Pa	Float	WR
TPP6	Total precipitation over the past 6 hours	mm	Float	WR
U10M	Eastward wind speed at 10 meters above ground	m/s	Float	WR
V10M	Northward wind speed at 10 meters above ground	m/s	Float	WR
ISHOLIDAY	Holiday indicator corresponding to the date and time recorded	Bit	Binary	H

consumption behavior over time. To begin, we analyzed and compared variations in electricity consumption curves based on the days of the week from 2008 to 2022.

The average consumption behavior for each day of the week is shown in Fig. 1. First, in Fig. 1(a), the line graph shows the average hourly profile of electricity consumption (in megawatts, MW) throughout the day for each day of the week, including holidays.

On the other hand, Fig. 1(b) shows the box plot summarizing the distribution of daily electricity consumption by day of the week and holidays. For each group, the median, quartiles and dispersion of the data are displayed, including possible outliers.

In addition, electricity consumption has a similar behavior for all days of the week, varying in amplitude in some cases. It is significantly lower on weekends and holidays than on weekdays, as shown in Fig. 1.

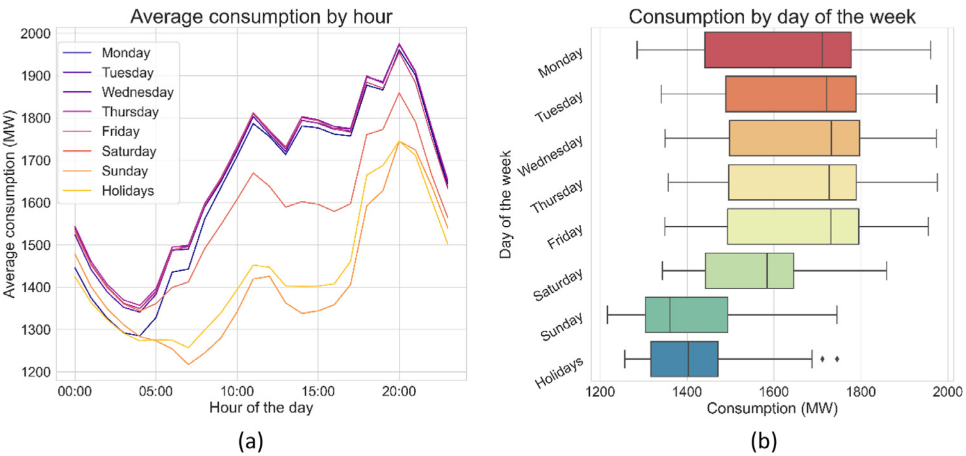

Fig. 1. Weekday electricity demand behaviour and holidays. **(a)** Average consumption curves for each hour of the day. **(b)** Electricity consumption boxplot.

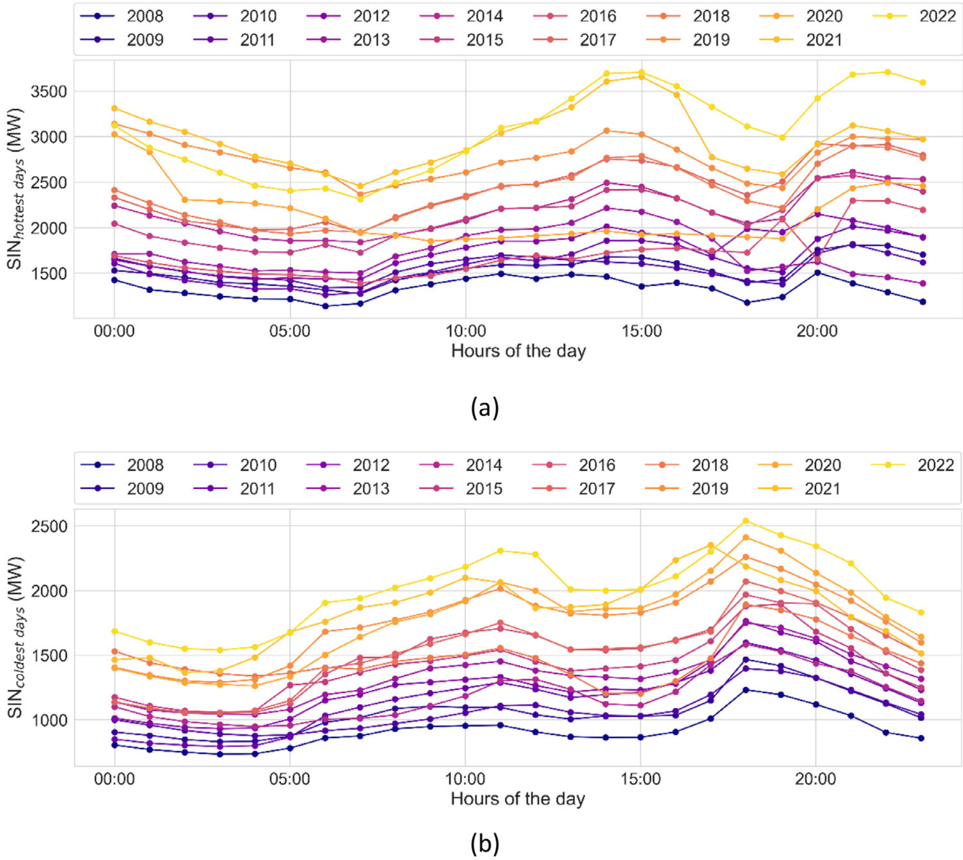

Fig. 2. Hourly variations in electricity consumption for the **(a)** hottest and **(b)** coldest days of each year.

On the other hand, Fig. 2 depicts two-line graphs representing the power consumption during the hottest and coldest days from 2008 to 2022. This leads to a constant increase in the electricity consumption over time.

3.2. Meteorological data

This dataset incorporates meteorological data extracted from the GDAS [4]. The GDAS system merges observations from a vast network of weather stations, satellites, radars, and other meteorological instruments. By integrating these diverse datasets into a numerical weather prediction model, GDAS offers a comprehensive and consistent portrayal of atmospheric conditions throughout the world [5]. This approach ensures a high degree of accuracy and eliminates potential biases that might arise from relying on a single data source.

The reference location for the meteorological data was the Silvio Pettirossi Airport (25.2419° S, 57.5188° W). This geographical point is located in the department of Central, which is the most populated in Paraguay and the one that has the greatest influence on internal electricity consumption. In turn, the Central department is located in the eastern region of Paraguay, which covers 39% of the national territory and concentrates approximately 97% of the population [6]. The relatively homogeneous climate of Paraguay, especially in the eastern region, minimizes the

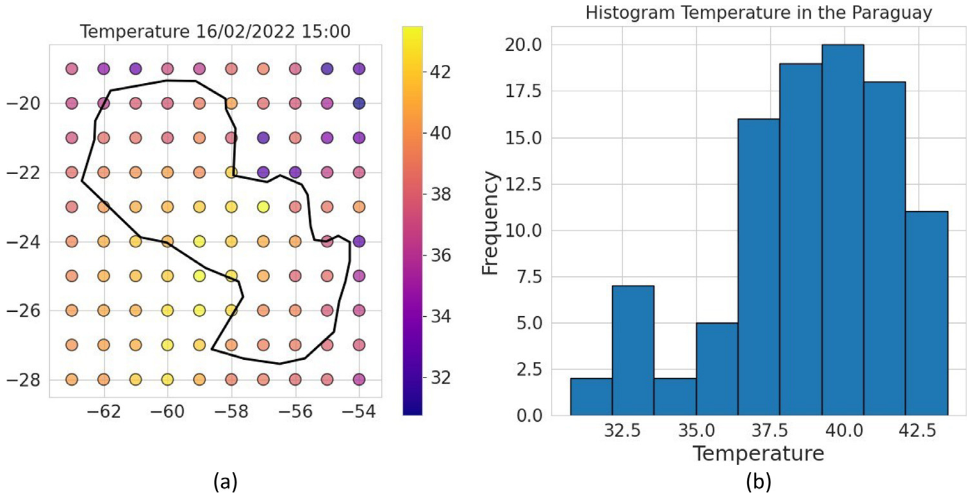

Fig. 3. Spatial distribution of surface temperature across Paraguay on February 16th, 2022 recorded at 15:00 PYST (Paraguay Standard Time) according to the GDAS model. **(a)** Temperature at different coordinate points along Paraguay. **(b)** Histogram of temperature variation.

likelihood of substantial regional variations in weather patterns. Therefore, the weather conditions at the Silvio Pettirossi Airport are likely representative of broader national trends, allowing us to capture key weather dynamics that impact electricity consumption throughout Paraguay. In relation to the above, Fig. 3 shows the temperature distribution on a day recorded in 2022. The histogram provides an insight into the distribution of these temperatures across the country. The relatively uniform distribution, with a single peak, suggests similar temperatures across most regions.

Fig. 4 illustrates the hourly temperature variations for the coldest and hottest days of each year from 2008 to 2022. In relation to the hottest days, the temperature has increased over the years, always maintaining the same profile; however, this increase does not turn out to be abrupt or to a great extent. In contrast, for the coldest days of each year, as the years progress, the coldest temperature has increased by a couple of degrees and generally remains constant throughout the day.

The specific meteorological variables included in this dataset are temperature, relative humidity, surface pressure, total precipitation over six hours, eastward wind speed at 10 meters above ground, and northward wind speed at 10 meters above ground. These measures are detailed in Table 1.

Based on studies [7,8], meteorological variables have a strong influence on electricity consumption because extreme heat often leads to increased reliance on air conditioning systems.

4. Experimental Design, Materials and Methods

Accurate historical data is essential for building robust energy demand forecasting models. However, real-world data often contains missing values and outliers that can significantly impact the model's performance. This subsection details the multi-step data imputation scheme employed in this study to address these shortcomings in the electricity demand and weather data. Fig. 5, shows the procedure to acquire, process and store the data that are part of this work.

By imputing missing values and identifying and handling outliers, we ensure the data accurately reflect the underlying patterns and relationships that are crucial for effective demand

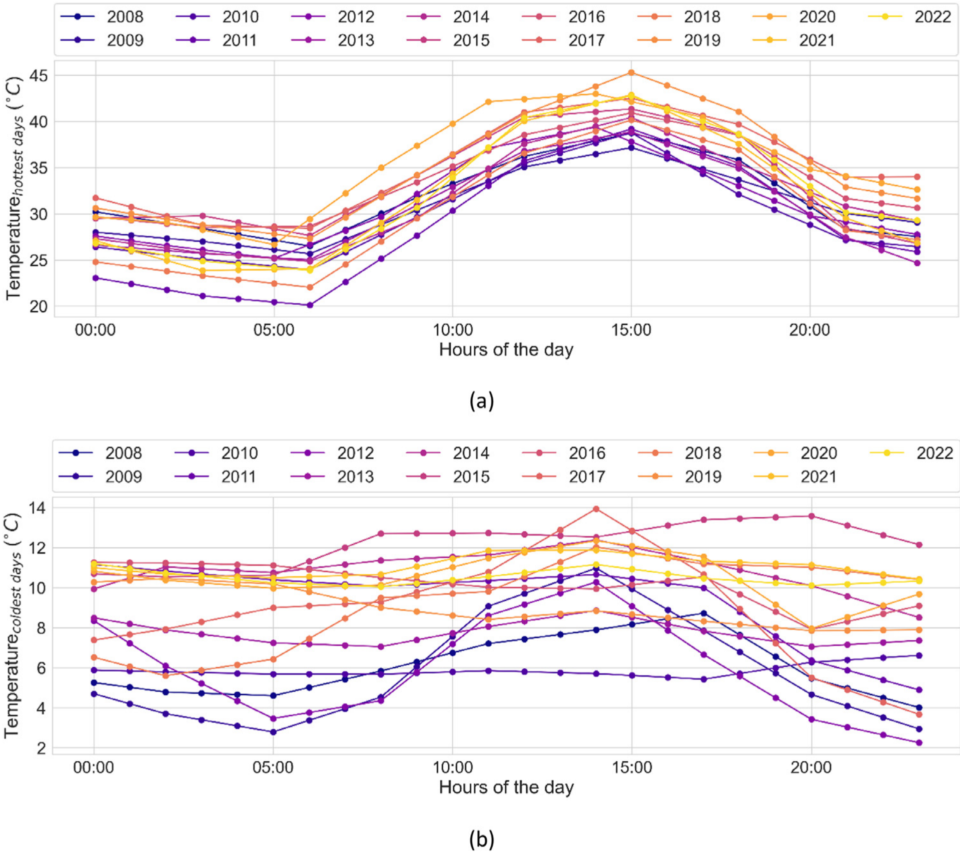

Fig. 4. Hourly variations in temperature for the **(a)** hottest and (b) coldest days of each year.

forecasting. To address missing values and outliers in the electricity demand, we employed a multi-step imputation scheme detailed as follows:

- **Missing Value Imputation:** We first addressed negative and missing values in the electricity demand time series. Standard practice dictates non-negative energy consumption values. These missing and negative values were replaced by a minimum electricity demand value corresponding to 0.1 (MW). This is because a complete time series is required to apply the following techniques, and these substituted values will necessarily be identified and replaced when applying the outlier detection and imputation techniques.
- **Normalization and Outlier Detection:** Let Y be the original data and λ a power parameter, the Box-Cox transformation

$$Y^* = \begin{cases} \frac{Y^\lambda - 1}{\lambda}, & if\ \lambda \neq 0 \\ \log Y, & if\ \lambda = 0 \end{cases}$$

was applied to obtain the normalized electricity demand data Y^*. This transformation improves the normality of the data and allows for the application of outlier detection techniques like the three-sigma rule (with a centered sliding window). We experimented with different window sizes, from 24 to 360, and found that a 168-hour (one week) window was the most effective in capturing global outliers without overly affecting local outliers that can become false positives.

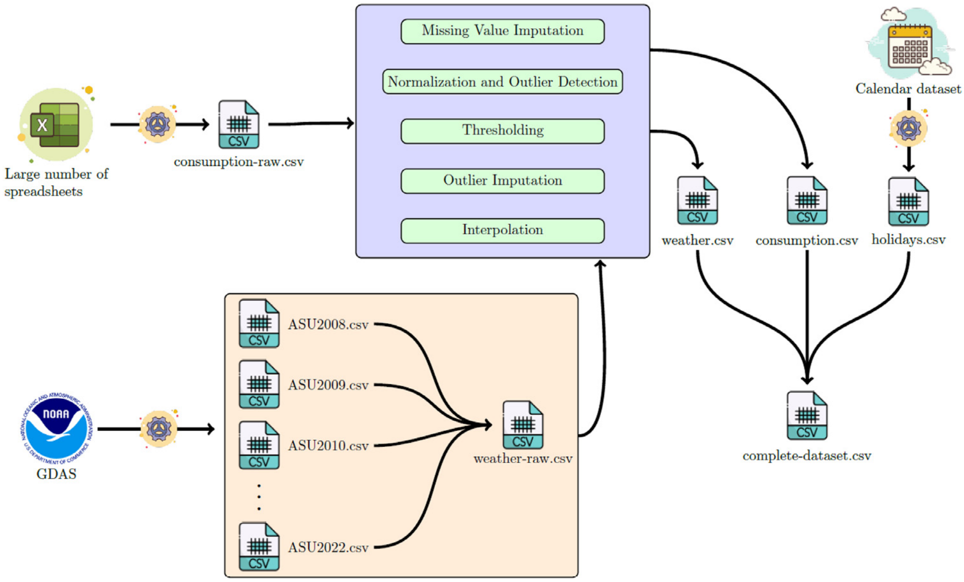

Fig. 5. Data processing and storage pipeline.

This window value has been the most conservative, knowing the typical behavior and fluctuation of electricity demand.

- **Outlier Imputation:** Detected outliers were excluded from the data. To estimate and replace these missing values, a historical mean imputation approach was employed. This method estimates the missing data by calculating the average of observations recorded at the same hour during the two preceding and two subsequent days**.** This technique helps maintain the internal consistency and temporal dynamics of the dataset, ensuring that the imputed values align with typical patterns of electricity consumption.
- **Inverse Transformation:** Finally, the inverse Box-Cox transformation was applied to the imputed data to recover the original scale of the electricity demand time series.

For the meteorological data, we also implemented a data cleaning process as follow:

- **Humidity Thresholding:** The relative humidity time series was adjusted to address values exceeding 100%. Any value exceeding this threshold was set to 100%, the maximum possible relative humidity.
- **Temperature Outlier Detection:** Although no significant outliers were detected in the temperature data, similar outlier detection techniques can be applied if required.
- **Interpolation:** Linear interpolation was used to ensure hourly data granularity for all meteorological variables, as samples were obtained at a frequency of 1 sample every 3 hours.

Limitations

Not applicable.

Ethics Statement

Authors declare no ethical conflicts with raw data collected and presented in this article.

Credit Author Statement

F. Morales-Mareco: Investigation, Visualization, Writing - Original Draft. **M. Verdún:** Investigation, Visualization, Writing - Original Draft. **D. H. Stalder:** Conceptualization, Validation, Writing - Review & Editing. **C. Sauer Ayala:** Validation, Writing - Review & Editing. **S. Grillo:** Visualization, Writing - Original Draft, Writing - Review & Editing.

Data Availability

Electricity consumption and meteorological data from Paraguay (Original data) (Mendeley Data).

Acknowledgements

Authors acknowledge the financial support of the Facultad de Ingeniería of the Universidad Nacional de Asunción (FIUNA) and the Consejo Nacional de Ciencia y Tecnología of Paraguay (CONACYT) in the project INIC01-6.

Declaration of Competing Interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] F. Morales-Mareco, C.S. Ayala, D.H. Stalder, L. Salgueiro, S.A. Grillo, Short-term electricity demand forecasting: Evaluating the effectiveness of statistical, machine learning, and deep learning models, in: 2023 XLIX Latin American Computer Conference (CLEI), 2023, pp. 1–10.

[2] J.A. Gallardo, M. García-Torres, F. Gómez-Vela, F. Morales, F. Divina, D. Becerra-Alonso, G. Velázquez, F. Daumas-Ladouce, J.L.V. Noguera, C.S. Ayala, D.P. Pinto-Roa, P.E. Gardel-Sotomayor, J.C.M. Román, Forecasting electricity consumption data from paraguay using a machine learning approach, in: H. Sanjurjo González, I. Pastor López, P. García Bringas, H. Quintián, E. Corchado (Eds.), 16th International Conference on Soft Computing Models in Industrial and Environmental Applications (SOCO 2021), Springer International Publishing, Cham, 2022, pp. 685–694.

[3] G. Velázquez, F. Morales, M. García-Torres, F. Gómez-Vela, F. Divina, J.L. Vázquez Noguera, F. Daumas-Ladouce, C. Sauer Ayala, D.P. Pinto-Roa, P.E. Gardel-Sotomayor, J.C. Mello Román, Distribution level electric current consumption and meteorological data set of the east region of paraguay, Data Brief 40 (2022) 107699 URL: https://www.sciencedirect.com/science/article/pii/S2352340921009744.

[4] For Environmental Information (NCEI), N. C. [n.d.], 'Global data assimilation system (gdas)'. URL: https://www.ncei.noaa.gov/products/weather-climate-models/global-data-assimilation.

[5] R. Buizza, P.L. Houtekamer, G. Pellerin, Z. Toth, Y. Zhu, M. Wei, A comparison of the ecmwf, msc, and ncep global ensemble prediction systems, Monthly Weather Rev. 133 (5) (2005) 1076–1097 URL: https://journals.ametsoc.org/view/journals/mwre/133/5/mwr2905.1.xml.

[6] Organización de las Naciones Unidas para la Alimentación y la Agricultura (FAO). (s.f.). *Diagnóstico Territorial: Paraguay*. URL: https://www.fao.org/4/j3292s/j3292s05.htm.

[7] M. Ali, M.J. Iqbal, M. Sharif, Relationship between extreme temperature and electricity demand in pakistan, Int. J. Energy Environ. Eng. 4 (1) (2013) 36.

[8] J.A. Sathaye, L.L. Dale, P.H. Larsen, G.A. Fitts, K. Koy, S.M. Lewis, A.F.P. de Lucena, Estimating impacts of warming temperatures on california's electricity system, Global Environ. Change 23 (2) (2013) 499–511 URL https://www.sciencedirect.com/science/article/pii/S0959378012001458.