Multivariate Forecasting of Energy Demand using Recurrent Neural Networks

Mariela N. Uhrig[a], Leandro D. Vignolo[b], Omar V. Müller[c]

[a] *Centro de Investigación Científica y de Transferencia Tecnológica a la Producción, CICYTTP-CONICET-Gobierno de Entre Ríos-UADER, Dr. Matteri y España s/n, Diamante, 3105, Entre Ríos, Argentina*

[b] *Research Institute for Signals, Systems and Computational Intelligence, sinc(i), FICH-UNL/CONICET, Ciudad Universitaria, Ruta Nac. N°168, Km. 472, Santa Fe, 3000, Santa Fe, Argentina*

[c] *Centro de Estudios de Variabilidad y Cambio Climático, FICH-UNL/CONICET, Ciudad Universitaria, Ruta Nac. N°168, Km. 472, Santa Fe, 3000, Santa Fe, Argentina*

Abstract

Accurate short-term electricity demand forecasting remains challenging due to complex relationships between demand and meteorological drivers, lagged effects acting at different temporal scales, and increased uncertainty during peak-demand periods. This paper addresses the design of recurrent neural networks for predicting electricity demand based on multiple input variables. Given the pressing need to enhance energy efficiency and reduce environmental impact, deep learning methods offer a promising approach to address these challenges.

The analysis of daily electricity demand data from the Province of Entre Ríos (Argentina), including a rich set of 21 meteorological, temporal, and energy-related variables reveals significantly higher demand during extreme temperature events, weekdays, and a sustained long-term growth trend. Capturing these patterns is essential for robust operational forecasting. This

study proposes a multivariate LSTM-based forecasting model capable of learning complex nonlinear temporal dependencies from heterogeneous inputs. LSTM networks are able to jointly model short- and long-term dynamics without manual feature engineering, while maintaining a rasonable architecture suitable for multi-horizon forecasting.

Results show that the proposed LSTM model clearly outperforms a baseline statistical approach and a state-of-the-art model based on gated recurrent neural networks, providing more accurate and stable forecasts across different prediction horizons, including during peak-demand periods. These findings highlight the effectiveness of parsimonious deep learning architectures for operational electricity demand forecasting and support their adoption in real-world power system planning and decision-making.
Keywords: electricity demand forecasting, long short-term memory networks, time series prediction

1. Introduction

Electricity constitutes a fundamental service sustaining a broad spectrum of human activities, including industrial production, residential lighting, the use of household appliances, and energy use in commercial and public infrastructure. At the global level, electricity demand has steadily increased in recent decades, driven by a combination of interrelated factors such as demographic expansion, urbanization, industrial growth, economic and technological development, and global warming [12, 19]. These forces have collectively intensified the demand placed on energy systems to improve efficiency, affordability, and sustainability, which has led to a growing body of research

exploring AI-driven approaches for improving load forecasting in energy systems [14, 1].

At a regional scale, electricity demand exhibits strong variability and is conditioned not only by socio-economic conditions but also by short-term dynamics such as seasonal weather patterns, temperature extremes, and human activity cycles [12, 18]. In this context, developing a comprehensive understanding of electricity consumption patterns and designing effective demand management strategies are essential to ensuring energy security and promoting sustainability. These strategies play a key role in optimizing resource allocation, reducing operational costs, and fostering informed energy use, especially considering that electricity expenditure constitutes a significant portion of the overall societal and industrial budget [18].

To support such decision-making processes, numerous forecasting models have been developed that rely on historical and contextual data. Traditional approaches, including Autoregressive Integrated Moving Average (ARIMA), Seasonal ARIMA with Exogenous Variables (SARIMAX), and exponential smoothing methods, have been widely applied due to their simplicity and interpretability. However, recent reviews have emphasized that these models are often inadequate for capturing complex dynamics in modern energy systems, particularly in multivariate contexts that involve nonlinear interactions and weather-driven variability [13, 24, 23]. Statistical approaches are inherently limited to modeling predominantly linear relationships and frequently fall short in scenarios where temporal dependencies and meteorological signals strongly influence electricity consumption [10].

This limitation has motivated the adoption of machine learning and deep

learning techniques, which are better suited for modeling intricate temporal dependencies including exogenous climate variables [8, 4]. Among these techniques, Recurrent Neural Networks (RNNs), and particularly Long Short-Term Memory (LSTM) networks, a specialized architecture designed to overcome the limitations of standard RNNs, have demonstrated notable success in time series forecasting, especially in energy applications [19, 10].

LSTMs are explicitly designed to capture both short- and long-term dependencies in sequential data through memory-cell mechanisms, allowing for more accurate modeling of energy consumption patterns over time. Hybrid architectures such as CNN–LSTM and attention-based models have also exhibited strong performance in multi-step forecasting tasks [4, 9]. Recent studies confirm that LSTM-based models can significantly improve electricity demand forecasting when they integrate not only historical load data but also relevant exogenous features such as temperature, humidity, wind speed, and solar radiation [19, 18, 10]. Moreover, LSTM-based models have been increasingly adopted in operational energy systems due to their adaptability, scalability, and ability to integrate multiple data sources in real time, making them suitable for grid-aware forecasting in dynamic environments [14, 13, 2].

As evidenced by the studies discussed above, forecasting electricity demand poses several challenges. The relationship between demand and meteorological variables is highly nonlinear, often governed by lagged or interacting effects that cannot be adequately modeled using linear statistical models. In addition, the integration of a large number of exogenous features introduces multicollinearity and temporal dependencies operating at different time scales. These difficulties become even more pronounced when forecast-

ing across multiple horizons or during peak-demand events, where abrupt weather changes drive rapid and sharp consumption fluctuations. Addressing these challenges requires a modeling approach capable of learning complex temporal patterns and integrating heterogeneous sources of information. LSTM architectures are advantageous because they can learn complex nonlinear relationships and capture both short- and long-term dependencies in electricity demand without extensive manual feature engineering.

This study develops and evaluates a multivariate LSTM-based model for short-term electricity demand forecasting in the province of Entre Ríos, Argentina, incorporating 21 meteorological, temporal, and energy-related variables. The model is built using real operational data from a regional electricity distribution system and is evaluated across multiple forecasting horizons (1, 7, and 14 days) and it is optimized through grid search with 5-fold cross-validation to ensure robustness and reproducibility of the results. Model performance is systematically benchmarked against SARIMAX [16] and DA-SKIP [11], a recent deep learning architecture, including a detailed assessment during peak-demand periods. Unlike many existing works that rely on limited feature sets or synthetic benchmark datasets, it focuses on an understudied South American context [8] and leverages operational data from a real electricity distribution system. The combined evaluation across multiple forecasting horizons, the explicit analysis of peak-demand events, and the use of a grid-searched and cross-validated optimization strategy distinguish this work from prior studies. By addressing both methodological and operational challenges, this study extends existing knowledge on multivariate forecasting for electricity demand and provides insights for regional utilities seeking to

modernize their grid management practices.

2. Methodology

2.1. Dataset Description

This study addresses a time series forecasting problem using a multivariate dataset composed of electricity consumption records and meteorological data. The electricity demand data for the province of Entre Ríos was provided by ENERSA, the local electricity provider, covering the period from 2012-01-01 to 2023-05-31. The dataset represents total provincial consumption at daily resolution. Meteorological variables were obtained from ERA5, the fifth-generation atmospheric reanalysis dataset developed by the European Centre for Medium-Range Weather Forecasts (ECMWF). To align with the time-step of the energy data, all weather variables were aggregated on a daily basis.

The final dataset includes 21 features, categorized into temporal, energetic, and meteorological variables, each contributing to the modeling of daily electricity consumption patterns. The output variable corresponds to the total daily energy demand measured in megawatt-hours (MWh). All variables were integrated to construct a coherent multivariate time series that enables the model to capture complex dependencies between electricity consumption and external conditions. Table 1 summarizes the input and output variables, grouped by type and accompanied by a brief description.

In Figure 1, the top panel illustrates the electricity demand dynamics over the entire period, revealing high variability, characterized by higher consumption during the austral summer and winter months. This pattern reflects the

Table 1: Summary of input and output variables used for electricity demand forecasting

Category	Type	Variable	Description
Input	Temporal	Year	Year of observation (2012–2023)
		Month	Month of the year (1–12)
		Day	Day of the month (1–31)
		Encoded Day	Encoded weekday (1–8; 8 = Holiday)
	Energetic	Users	Number of electricity service users
	Meteorological	Minimum temperature (tmin)	Minimum 2-m air temperature
		Mean temperature (tmean)	Mean 2-m air temperature
		Maximum temperature (tmax)	Maximum 2-m air temperature
		Skin temperature (skt)	Mean skin surface temperature
		Surface solar radiation downwards (ssrd)	Incoming solar radiation at surface
		Surface thermal radiation downwards (strd)	Downward thermal radiation
		Surface net solar radiation (ssr)	Net shortwave radiation balance
		Surface net thermal radiation (str)	Net longwave radiation balance
		Surface latent heat flux (slhf)	Latent heat flux at surface
		Surface sensible heat flux (sshf)	Sensible heat flux at surface
		Total precipitation (tp)	Daily accumulated precipitation
		Surface runoff (sro)	Water runoff at surface
		Surface pressure (sp)	Surface atmospheric pressure
		Total cloud cover (tcc)	Fraction of sky covered by clouds
		Wind speed (ws)	10-m wind speed
		Relative humidity (rh)	Atmospheric moisture content
Output	Target	Energy Demand	Total daily electricity consumption

climatic influence on electricity usage. The bottom panel zooms in on the year 2019, showing weekly demand dynamics with higher consumption on weekdays that sharply decreases during weekends. Given the temporal dynamics of the electricity demand data, we split the dataset chronologically: 80% for training and 20% for testing (see the top panel of Figure 1). The training set spans from 2012-01 to 2022-04, and the testing set covers 2022-05 to 2023-05. This chronological split preserves the temporal structure of the data across both subsets, including the strong seasonality of electricity demand. Main-

taining this temporal coherence is essential for accurately capturing seasonal trends and other time-dependent patterns that affect predictive performance.

Before training the models, all input features were scaled to the [0,1] range using Min–Max normalization based on the training set, and the same transformation was applied to the test data, with predictions later rescaled to original units. This step is critical for training artificial neural networks, as gradient-based optimization is sensitive to feature scale. Without normalization, large-magnitude variables can lead to unstable gradients and degraded performance. Min–Max normalization ensures balanced feature contributions and more stable convergence and is widely used in neural network–based energy forecasting studies [7].

2.2. Methods

Three methods are applied in this study: SARIMAX, DA-SKIP and LSTM. SARIMAX serves as a baseline for benchmarking, while the LSTM deep learning architecture is designed as a superior alternative due to its ability to capture nonlinear temporal dependencies, which are key for accurate electricity demand forecasting. DA-SKIP is a composite deep learning model designed specifically for the challenge of periodic time series forecasting, with a more complex architecture. The three approaches account for the influence of meteorological conditions on electricity demand. The LSTM model is further optimized with particular focus on temporal window size and hyperparameter configurations.

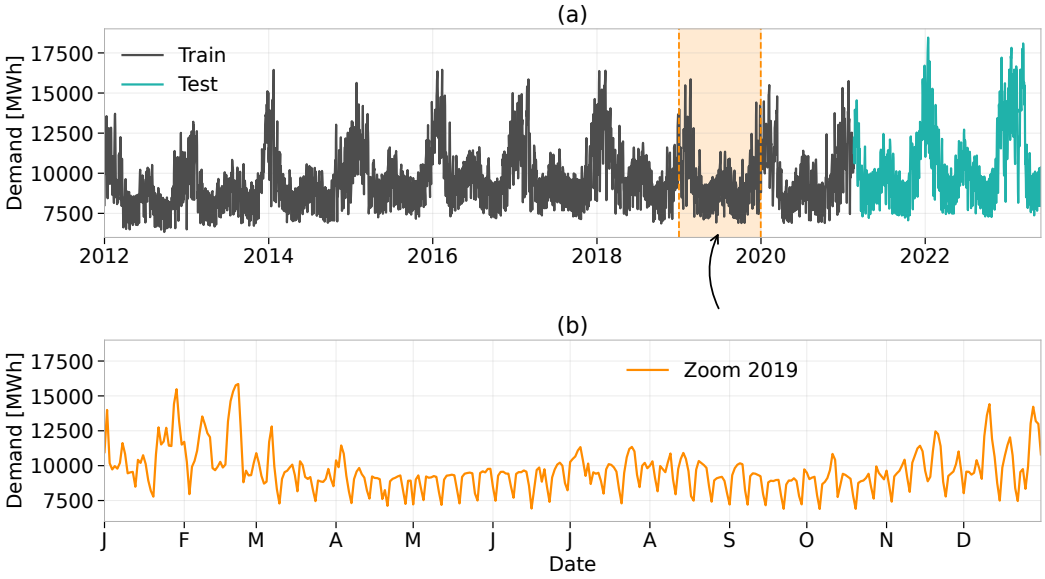

Figure 1: (a) Data split visualization: 80% of the daily electricity demand time series was used for training and 20% for testing, preserving the original chronological structure. (b) A zoom for the year 2019 to highlight the weekly and seasonal variability of the time series.

2.2.1. SARIMAX

The SARIMAX model extends the classical ARIMA model to incorporate exogenous predictors that influence the target time series. An ARIMA(p, d, q) model captures autocorrelations and moving average structures within a differenced series. It is expressed as:

$$y_t = \mu + \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} + \varepsilon_t \tag{1}$$

where y_t denotes the observed value of the target time series at time t, μ is a constant term, ϕ_i and θ_j are the coefficients of the autoregressive (AR) and moving average (MA) components, respectively, ε_t is a white noise error

term, p and q indicate the order of the AR and MA terms, and d is the degree of differencing applied to the original series to achieve stationarity. This formulation is mathematically equivalent to the more compact polynomial representation commonly used in time series literature, where AR and MA components are expressed as functions of the backshift operator.

The SARIMAX model introduces an additional regression component to include exogenous variables $X_{i,t}$, modifying the model as:

$$y_t = \mu + \sum_{i=1}^{k} \beta_i X_{i,t} + \text{ARIMA terms} + \varepsilon_t \tag{2}$$

where μ is a constant term, β_i are coefficients for the exogenous predictors, and k is the number of external variables. This formulation enables the SARIMAX model to simultaneously account for internal serial dependencies and the influence of external factors. In the context of this study, these external factors correspond to meteorological variables presented in Table 1.

In this formulation, the autoregressive and moving average components model linear temporal dependencies in electricity demand, while the exogenous regression term captures the direct and instantaneous influence of meteorological variables. This structure allows SARIMAX to represent climate-driven demand variability under the assumption of linear relationships and stationarity after differencing.

2.2.2. LSTM Networks

Artificial Neural Networks (ANNs) are powerful computational models inspired by the way the human brain processes information [17]. They consist of interconnected layers of processing units (neurons) capable of learning

complex mappings from input to output data. While feedforward neural networks can be applied to time series prediction, their inability to capture sequential dependencies over long periods limits their effectiveness in forecasting tasks. RNNs address this limitation by introducing recurrent connections, allowing information to persist across time steps. However, traditional RNNs suffer from vanishing and exploding gradient problems, making them ineffective at learning long-term dependencies. LSTM networks, a special type of RNN, were designed to overcome these issues by introducing a memory cell structure and gating mechanisms [13, 15]. LSTMs can selectively retain or forget information through input, output, and forget gates, thus enabling the modeling of complex temporal patterns across long sequences. The architecture of an LSTM cell with a forget gate is defined below.

The forget gate, which determines the unnecessary component from the previous cell state, is defined as:

$$f_t = W_f X_t + U_f h_{t-1} + b_f \tag{3}$$

where W_f is the weight matrix applied to the input X_t, U_f is the recurrent weight matrix applied to the previous hidden state h_{t-1}, and b_f is the bias term for the forget gate.

The input gate, which decides how much new information to store in the cell state, is defined as:

$$i_t = W_i X_t + U_i h_{t-1} + b_i, \tag{4}$$

where W_i is the weight matrix applied to the input X_t, U_i is the recurrent weight matrix applied to the previous hidden state h_{t-1}, and b_i is the bias term for the input gate.

sinc(i) Research Institute for Signals, Systems and Computational Intelligence (sinc.unl.edu.ar)
M. Uhrig, L. D. Vignolo & O. Müller; "Multivariate forecasting of energy demand using recurrent neural networks"
Advanced Engineering Informatics, Vol. 71, 2026.

The candidate cell state at time t is a new candidate memory created from the input and the previous hidden state, and is calculated as:

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \tag{5}$$

where the hyperbolic tangent activation function ensures that the output values are between -1 and 1, W_C is the weight matrix applied to the input X_t, U_C the recurrent weight matrix applied to the previous hidden state h_{t-1}, and b_C the bias term for the candidate cell state.

The cell state is a combination of the previous cell state, modulated by the forget gate, and the new candidate memory, modulated by the input gate. At time t, the cell state is updated as:

$$C_t = f_t C_{t-1} + i_t \tilde{C}_t \tag{6}$$

where $f_t C_{t-1}$ is the element-wise (Hadamard) product between the forget gate activation and the previous cell state C_{t-1}, representing the part of the previous cell state that is retained, and $i_t \tilde{C}_t$ is the element-wise (Hadamard) product between the input gate activation and the candidate cell state, representing the new information being stored.

The output from the LSTM cell to the next cell is given by the output gate, which determines how much of the cell state will be used for the output:

$$o_t = W_o X_t + U_o h_{t-1} + b_o \tag{7}$$

where o_t is the output gate activation at time t, W_o is the weight matrix applied to the input X_t, U_o the recurrent weight matrix applied to the previous hidden state h_{t-1}, and b_o the bias term for the output gate.

$$h_t = o_t \tanh(C_t) \tag{8}$$

where h_t is the hidden state at time t. This is the output of the LSTM cell, based on the output gate activation o_t and the current cell state C_t. This modulates the influence of the cell state on the output, while the application of the hyperbolic tangent function to the cell state C_t ensures that the output values are between -1 and 1.

These equations describe the internal gating mechanisms of the LSTM cell. The forget gate f_t determines which information from the previous cell state should be retained, while the the input gate i_t and candidate cell state $\tilde{C}_t$ regulate how much new information enters the memory. The updated cell state C_t integrates both retained and incoming information, enabling the model to preserve long-term dependencies. Finally, the output gate o_t controls how much of the internal state is exposed to produce the hidden state h_t. This structure allows the LSTM to capture nonlinear and temporally lagged relationships in sequential energy demand data, which is essential for multivariate forecasting tasks.

These mechanisms allow LSTM networks to capture both short-term and long-term temporal dependencies effectively, making them particularly suited for time series forecasting tasks involving complex, nonlinear relationships, such as electricity demand influenced by weather conditions.

By dynamically regulating information flow through these gates, the LSTM cell effectively adapts to delayed and nonlinear demand responses driven by weather variability and temporal patterns.

2.2.3. DA-SKIP

This state-of-the-art deep learning architecture was recently proposed by Huang et al. [11]. DA-SKIP is a composite model that integrates a dual-

stage attention-based encoder–decoder (DA-RNN), a gated recurrent unit with skip connections component (GRU-SKIP) explicitly designed to capture periodic patterns with a predefined period length. Finally, a simple autoregressive component handles the linear aspect of the forecasting. By integrating these three mechanisms—DA-RNN, GRU-SKIP, and the autoregressive part—the DA-SKIP is able to capture both long-term dependencies and strong periodic characteristics. While this design allows specialized modeling of nonlinear, periodic, and linear dynamics, it also results in a substantially more complex architecture with multiple recurrent modules and attention mechanisms.

2.3. Experimental Setup

All experiments were executed on an Intel Core i7 CPU with 32 GB of RAM. The hyperparameter grid search for the LSTM model required substantial computational effort, with training times ranging from approximately 4 to 9 days per configuration. Once the optimal model configuration was selected, training and evaluation for a single experiment required between 1 and 3 minutes, depending on the forecasting horizon.

2.3.1. SARIMAX Setup

The SARIMAX model, considered a baseline for electricity demand forecasting in our experiments, leverages daily meteorological and consumption data using a seven-day historical window (HTW=7) to forecast one day ahead (FTW=1). A grid search was carried out over various combinations of autoregressive (p), differencing (d), and moving average (q) orders, as summarized in Table 2. After identifying the optimal SARIMAX configuration based on

Table 2: Grid search configuration for SARIMAX hyperparameters.

Parameter	**Values Considered**
Autoregressive order (p)	{7}
Differencing (d)	{0, 1, 2}
Moving average order (q)	{1, 2, 3, 7, 14}

FTW=1, the model was also applied to multi-step forecasts (FTW=7 and FTW=14) using the same parameters.

2.3.2. LSTM Setup

The proposed deep learning model is based on a LSTM architecture, specifically designed for sequential pattern recognition in time series forecasting. To deepen our understanding of model behavior under different temporal configurations, a first series of experiments was conducted using LSTM networks trained with a fixed seven-day historical window to predict the following day (HTW7+FTW1) [22, 21, 20]. A manual grid search was used to adjust key hyperparameters, such as the number of LSTM layers, hidden units, learning rate, batch size, and number of epochs. This exploratory phase was instrumental in identifying general performance trends and narrowing the search space for more refined experiments. Notably, configurations using excessively large batch sizes (e.g., 32) consistently underperformed, guiding subsequent stages toward smaller batch sizes, which yielded more stable and accurate results.

Building on this preliminary phase, a systematic and automated evaluation model was implemented. All LSTM models were trained using a maxi-

mum of 300 epochs, with early stopping applied to prevent overfitting. The Mean Squared Error (MSE) was selected as the loss function, and optimization was performed with the Adam algorithm due to its adaptive learning capabilities, which improve convergence in deep learning models. A 5-fold cross-validation scheme was adopted to enhance generalization and mitigate overfitting.

An exhaustive grid search was conducted to explore combinations of hyperparameters, including the number of layers, hidden units, learning rates, batch sizes, and early stopping patience, using the value ranges detailed in Table 3. The historical temporal window (HTW) was fixed at seven days across all experiments, while forecasting performance was assessed for multiple prediction horizons: 1-day, 7-day, and 14-day ahead (FTW = 1, 7, and 14, respectively). This progressive setup enabled the evaluation of model robustness under increasing temporal uncertainty while maintaining consistent input conditions.

Before presenting the forecasting results, it is important to describe how the experimental setup guided the design and evaluation of the LSTM models, as well as the implementation of a classical SARIMAX model used as a comparative benchmark. The use of daily-level data spanning more than a decade (2012–2023) enabled the models to capture both seasonal patterns and short-term fluctuations, which are critical for reliable electricity demand forecasting in operational contexts.

In addition, particular emphasis was placed on evaluating the model's behavior not only under average conditions but also during peak demand scenarios, which are operationally critical for electricity providers. By in-

Table 3: Grid search configuration for LSTM hyperparameters.

Parameter	**Values Considered**
LSTM Layers	{1, 2, 3}
Hidden Neurons	{5, 10, 15, 20, 25, 30, 35, 40, 45, 50}
Learning Rate	{0.0001, 0.001}
Patience (Early Stopping)	{10, 50, 100, 150}
Batch Size	{4, 8, 10, 16, 32}

cluding complementary metrics such as P90, R2 and RMSE on high-demand days, the analysis provides a nuanced view of model reliability in extreme situations. This dual approach, quantitative optimization of model architecture and targeted evaluation of forecast performance, ensures that the conclusions are grounded in both statistical rigor and practical relevance.

2.4. Performance Evaluation Metrics

We evaluate two key aspects of model performance. The first is the overall predictive ability, which assesses how well the model reproduces the entire electricity demand time series, considering all days in the test period. The second focuses specifically on the model's ability to capture extreme demand conditions, which are particularly relevant for preventing power outages or unnecessary system recharges. In this case, performance is evaluated only on days when electricity demand exceeds the 90th percentile threshold.

Both aspects are assessed using the same performance metrics to ensure consistency: Root Mean Squared Error (RMSE) and the coefficient of determination (R^2). RMSE ranges [0, inf) and measures the average magnitude

of the errors between observed and predicted values and is expressed in the same units as the target variable (MWh). Lower RMSE values indicate better performance, with a perfect score of 0. R^2 ranges [0, 1] and quantifies the proportion of variance in the demand explained by the model, where values closer to 1 represent stronger predictive ability. These metrics provide complementary insights into both overall accuracy and performance under critical peak-demand conditions.

In addition to these primary metrics, two complementary indicators were incorporated to provide a more comprehensive evaluation of forecast quality: Willmott's Index of Agreement (WI) and the percentage bias (%Bias) [6]. The WI measures the degree of agreement between observed and predicted values by jointly accounting for both systematic and unsystematic errors, with values ranging from 0 to 1, where higher values indicate stronger agreement [3, 24]. Unlike R^2, WI is sensitive to differences in magnitude and is particularly useful for assessing overall model consistency. The %Bias metric quantifies the average tendency of the model to overestimate or underestimate electricity demand, with values close to zero indicating an unbiased forecast, and positive or negative values reflecting systematic overestimation or underestimation, respectively [4, 5, 9].

3. Exploratory Data Analysis

A preliminary analysis of the dataset was conducted to characterize the main features of the electricity demand time series and its relationship with meteorological variables. As shown in Figure 2, the monthly average energy demand reveals a clear seasonal pattern and a general upward trend

sinc(i) Research Institute for Signals, Systems and Computational Intelligence (sinc.unl.edu.ar)
M. Uhrig, L. D. Vignolo & O. Müller; "Multivariate forecasting of energy demand using recurrent neural networks"
Advanced Engineering Informatics, Vol. 71, 2026.

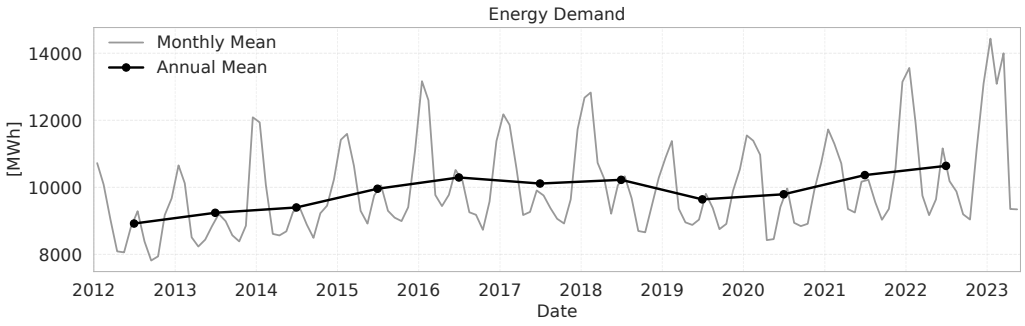

Figure 2: Energy demand [MWh] time series at monthly and annual time-scale.

throughout the period 2012–2023. However, this growth is not uniform: a slight decline or stagnation is observed between 2018 and 2020, which contrasts with the overall increasing tendency. This deviation aligns with significant increases in electricity tariffs, which typically lead to more conservative consumption behavior among users. These changes in consumer behavior highlight the relevance of socio-economic factors in demand forecasting and justify the potential inclusion of other exogenous variables in predictive modeling. Furthermore, the long-term dynamics of electricity demand underscore the influence of broader structural drivers such as population growth, economic activity, and regulatory policies on consumption patterns.

Figure 3 reveals the presence of a well-defined seasonal pattern in electricity consumption, displaying a "W-shaped" curve throughout the year (also evident in Figures 1 and 2). The interquartile range highlights the variability in demand from year to year, reinforcing the importance of accounting for seasonal effects in forecasting models. This pattern is marked by increased demand during the austral summer and winter months, with reduced consumption during the transitional seasons (spring and autumn). Such behavior is likely driven by temperature-related factors: higher demand arises from

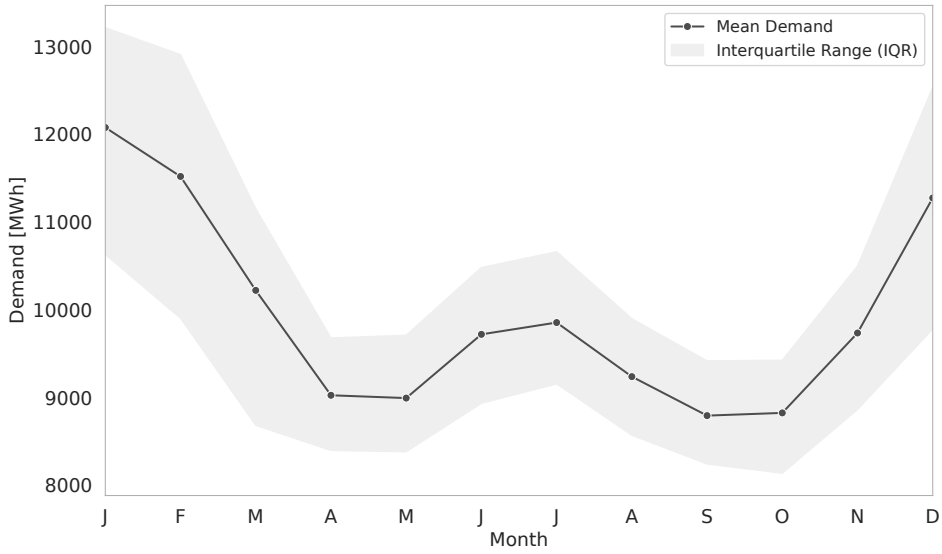

Figure 3: Mean electricity demand annual cycle.

cooling needs in summer and heating needs in winter. While both summer and winter show higher electricity demand due to temperature-related needs, the summer peaks tend to be more pronounced. This can be attributed to the fact that cooling during hot months relies almost exclusively on electricity, whereas winter heating is often provided by natural gas, depending on household infrastructure. Additionally, in Entre Ríos, heatwaves in summer are typically more frequent and persistent than cold spells in winter, leading to more sustained electricity use during the warmer season. This particular behavior underlines the importance of seasonal adjustments in electricity demand forecasting models.

Figure 4 presents the distribution of daily electricity demand across the days of the week. Panel (a) shows that electricity consumption is highest on weekdays (Monday to Friday), while weekends, particularly Sundays, exhibit the lowest levels. Panel (b) analyzes holidays as a separate category, revealing

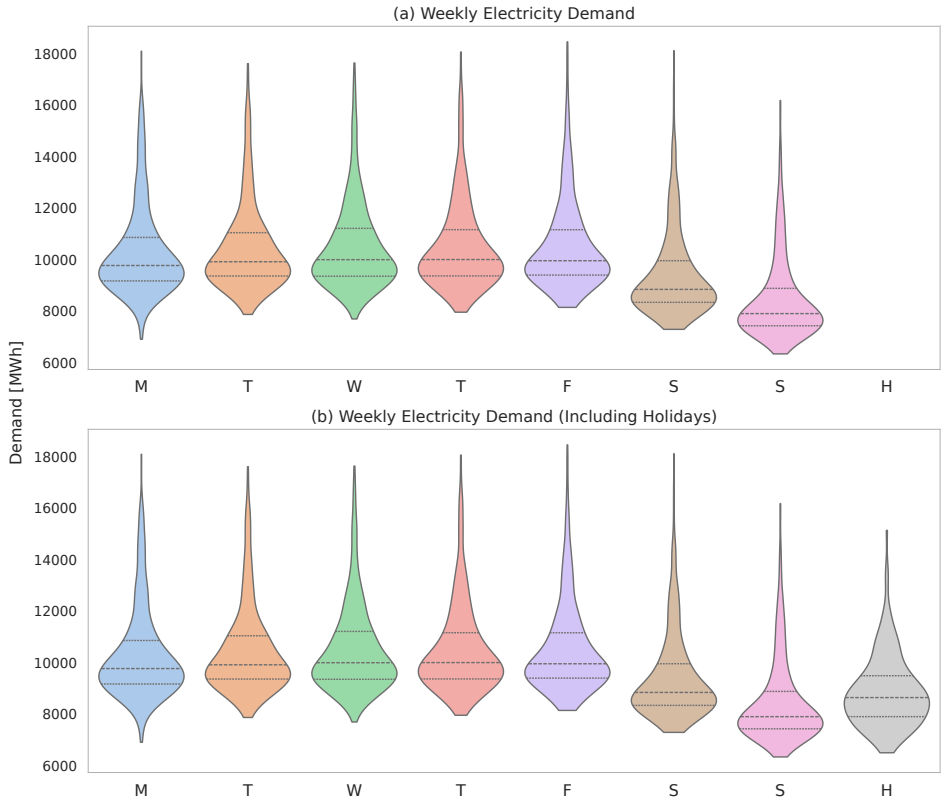

Figure 4: Daily distribution of electricity demand represented by violin plots. (a) Consumption distribution from Monday to Sunday. (b) Consumption distribution including holidays as an additional category.

that their consumption pattern more closely resembles that of weekends than that of weekdays. This observation supports the need to model holidays as a distinct temporal feature in forecasting models, rather than grouping them with standard days of the week, to improve predictive accuracy.

Understanding the relationship between meteorological conditions and electricity demand is essential for improving forecast accuracy, particularly in regions where climate variability strongly influences consumption patterns. To investigate these associations, Figure 5 presents scatter plots

sinc(i) Research Institute for Signals, Systems and Computational Intelligence (sinc.unl.edu.ar)
M. Uhrig, L. D. Vignolo & O. Müller; "Multivariate forecasting of energy demand using recurrent neural networks"
Advanced Engineering Informatics, Vol. 71, 2026.

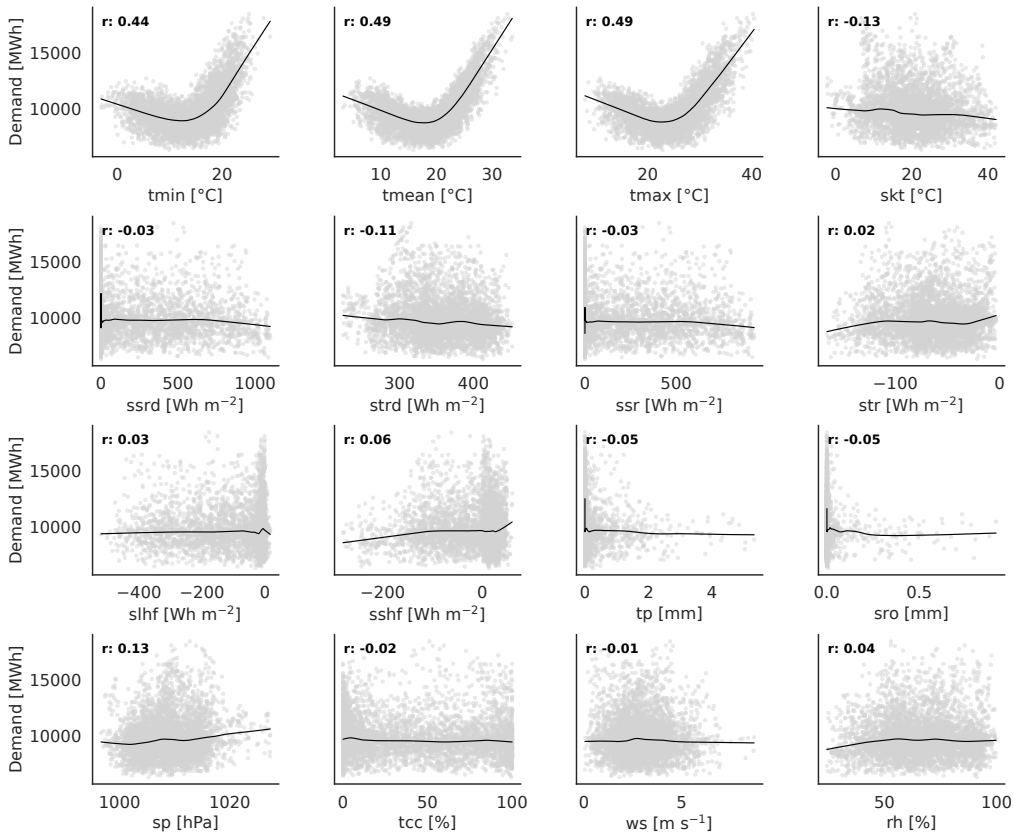

Figure 5: Scatter plots of input meteorological variables vs energy demand.

between individual weather variables and electricity demand, overlaid with locally weighted smoothing (LOWESS) curves that capture potential nonlinear trends. Among the variables analyzed, temperature indicators, minimum (tmin), mean (tmean), and maximum (tmax), show the strongest correlations with electricity demand. The observed relationship follows a distinct asymmetric U-shape, with electricity demand increasing more steeply at higher temperatures than at lower ones, reflecting a right-skewed pattern driven by cooling requirements. This behavior is consistent with the consumption pattern identified in the time-series.

sinc(*i*) Research Institute for Signals, Systems and Computational Intelligence (sinc.unl.edu.ar)
M. Uhrig, L. D. Vignolo & O. Müller; "Multivariate forecasting of energy demand using recurrent neural networks"
Advanced Engineering Informatics, Vol. 71, 2026.

In contrast, other meteorological variables such as solar radiation, wind speed, and cloud cover display weaker or more complex associations with demand. These effects may be indirect, interacting with other climatic or socio-behavioral factors not captured by univariate analysis. The use of LOWESS smoothing helps reveal these subtleties, underscoring the importance of flexible, nonlinear modeling techniques, such as artificial neural networks, for effectively capturing these complex dependencies.

4. Modeling Results

4.1. Hyperparameters Grid Search

As a starting point, a grid search was performed to evaluate different SARIMAX configurations by testing several values for the autoregressive (p), differencing (d), and moving average (q) parameters, as detailed in Table 2. While some settings yielded acceptable levels of accuracy, the overall performance was limited, especially under multistep forecasting scenarios, underscoring the challenges of relying solely on linear statistical models for capturing the complex dynamics of electricity demand.

For the LSTM models, Table 4 summarizes the best-performing configurations obtained for each of the seven experiments, highlighting the optimal combinations of temporal windows and hyperparameters. Across all cases, batch sizes of 4, 8, and 10 consistently delivered the highest predictive performance, with R^2 values ranging from 0.84 to 0.94 and RMSE values between 727.2 and 1454.4 MWh. The best results were achieved with models using one or two LSTM layers and a moderate number of hidden units (5 to 15), indicating that increased architectural complexity was not necessarily

sinc(i) Research Institute for Signals, Systems and Computational Intelligence (sinc.unl.edu.ar)
M. Uhrig, L. D. Vignolo & O. Müller; "Multivariate forecasting of energy demand using recurrent neural networks"
Advanced Engineering Informatics, Vol. 71, 2026.

advantageous in this setting. Experiment C (HTW7+FTW1) achieved the highest overall performance ($R^2 = 0.94$, RMSE = 727.2 MWh), while forecast accuracy declined gradually as the prediction horizon increased, with Experiment F (HTW7+FTW14) reaching an R^2 of 0.66 and an RMSE of 1696.8 MWh. Nevertheless, even under extended forecasting windows, the models maintained acceptable levels of accuracy, reinforcing the applicability of the proposed approach for short-term operational planning in real-world electricity distribution systems.

Furthermore, early experimental stages provided valuable insights into the influence of batch size on model performance. Experiments C and D, reconfirmed the poor performance of batch size 32, which was consequently excluded from all subsequent configurations. Similarly, in Experiment E, the use of batch size 16 led to suboptimal results and was also discarded. As a result, the remaining experiments, F, G, A, and B, were conducted exclusively with batch sizes of 4, 8, and 10, which consistently achieved superior predictive accuracy and stability. The results confirm that batch size is a key hyperparameter in LSTM training, with smaller batches improving generalization and training stability in multivariate electricity demand forecasting. This is consistent with LSTM optimization dynamics, where smaller batches introduce beneficial gradient stochasticity, while larger batches yield smoother but less adaptive updates that can hinder convergence in highly nonlinear time series. This structured and iterative approach, combining automated hyperparameter tuning, cross-validation, and incremental refinement of architectural choices, ensured a comprehensive and reproducible evaluation of LSTM performance across diverse forecasting scenarios.

Table 4: Grid search results for all experiments. Best configuration for each experiment is highlighted in bold.

Exp	HTW	FTW	Batch	Layers	Neurons	LR	Patience	R^2	RMSE [MWh]
			4	2	15	0.0001	150	0.85	1090.8
A	0	1	**8**	**3**	**5**	**0.001**	**150**	**0.90**	**848.4**
			10	1	25	0.001	50	0.82	1212.0
			4	1	35	0.0001	100	0.84	1212.0
B	0	7	**8**	**2**	**10**	**0.001**	**10**	**0.91**	**848.4**
			10	3	20	0.001	50	0.74	1454.4
			4	3	30	0.0001	10	0.85	1090.8
			8	**1**	**15**	**0.001**	**10**	**0.94**	**727.2**
C	7	1	10	2	45	0.001	100	0.78	1333.2
			16	1	15	0.001	50	0.77	1454.4
			32	3	5	0.001	100	0.71	1575.6
			4	1	30	0.0001	100	0.88	969.6
			8	**1**	**5**	**0.001**	**50**	**0.91**	**848.4**
D	7	3	10	1	10	0.001	150	0.81	1212.0
			16	3	5	0.001	150	0.76	1454.4
			32	2	40	0.001	100	0.62	1818.0
			4	1	30	0.0001	100	0.81	1212.0
E	7	7	**8**	**3**	**10**	**0.001**	**50**	**0.91**	**848.4**
			10	3	10	0.001	150	0.78	1333.2
			16	2	45	0.0001	50	0.59	1818.0
			4	3	20	0.001	150	0.57	1939.2
F	7	14	**8**	**1**	**10**	**0.001**	**100**	**0.66**	**1696.8**
			10	1	35	0.001	150	0.63	1818.0
			4	2	15	0.0001	150	0.82	1212.0
G	14	7	**8**	**2**	**5**	**0.001**	**50**	**0.90**	**969.6**
			10	3	40	0.0001	50	0.58	1939.2

Table 5: Comparison of LSTM, SARIMAX, and DA-SKIP models for different forecasting horizons.

HTW	**FTW**	**Model**	$\mathbf{R^2}$	**RMSE [MWh]**	**WI**	$\%Bias$
		SARIMAX	0.40	1878.2	0.55	-11.6
7	1	DA-SKIP	0.83	942.3	0.93	-3.0
		LSTM	**0.94**	**727.2**	**0.98**	**-1.0**
		SARIMAX	0.41	1866.1	0.53	−12.1
7	7	DA-SKIP	0.34	1832.5	0.55	−8.4
		LSTM	**0.91**	**848.4**	**0.97**	**-3.0**
		SARIMAX	0.45	1803.1	0.57	-12.0
7	14	DA-SKIP	0.31	1883.1	0.54	**-9.2**
		LSTM	**0.66**	**1696.8**	**0.90**	-10.7

Across all forecast horizons, the LSTM model demonstrated strong predictive capabilities for short-term electricity demand forecasting, particularly in the 1-day (FD1) and 7-day (FTW7) setups. As shown in Figures 1 to 3, the predicted values closely track the observed demand profiles, achieving high determination coefficients ($\mathbf{R^2}$ up to 0.94) and low root mean square errors (RMSE as low as 727.2 MWh). This strong agreement highlights the model's ability to capture both daily and weekly consumption dynamics with high fidelity. When extending the forecast to 14 days (FTW14), model performance naturally declines, with $\mathbf{R^2}$ decreasing to 0.61 and RMSE increasing to 1696.8. This reduction is expected in time series forecasting, as longer horizons introduce greater uncertainty and temporal complexity.

4.2. Models Comparison

To perform a comprehensive comparison of the proposed approach, the

LSTM model was compared with both a classical statistical baseline, SARIMAX, and a recent composite deep learning model, DA-SKIP [11]. Table 5 summarizes the comparative performance of the three models across different forecasting horizons.

As shown in Table 5, the LSTM model consistently outperforms both SARIMAX and DA-SKIP across all considered forecasting horizons. SARIMAX exhibits limited predictive capability, with relatively low R^2 and persistently high RMSE values, reflecting its difficulty in capturing nonlinear dynamics and multi-day temporal dependencies in the presence of exogenous meteorological variables. DA-SKIP attains acceptable performance for the 1-day forecasting horizon; however, its accuracy remains inferior to that of the LSTM model, with higher prediction errors. Moreover, as the forecasting horizon extends to 7 and 14 days, DA-SKIP performance degrades markedly, showing a pronounced reduction in explained variance and a substantial increase in RMSE. In contrast, the LSTM model maintains more stable accuracy across increasing horizons, demonstrating greater robustness to temporal uncertainty while relying on a significantly simpler and more parsimonious architecture.

Similar findings can be observed when comparing the models by means of Willmott's Index (WI) and %Bias, following the methodology of recent studies in multivariate energy and environmental forecasting [6, 4]. As shown in Table 5, the consistently high WI values obtained by the LSTM model indicate a strong agreement between observed and predicted electricity demand across all forecasting horizons. Moreover, the negative values of %Bias reflect a systematic but controlled underestimation behavior, which becomes more

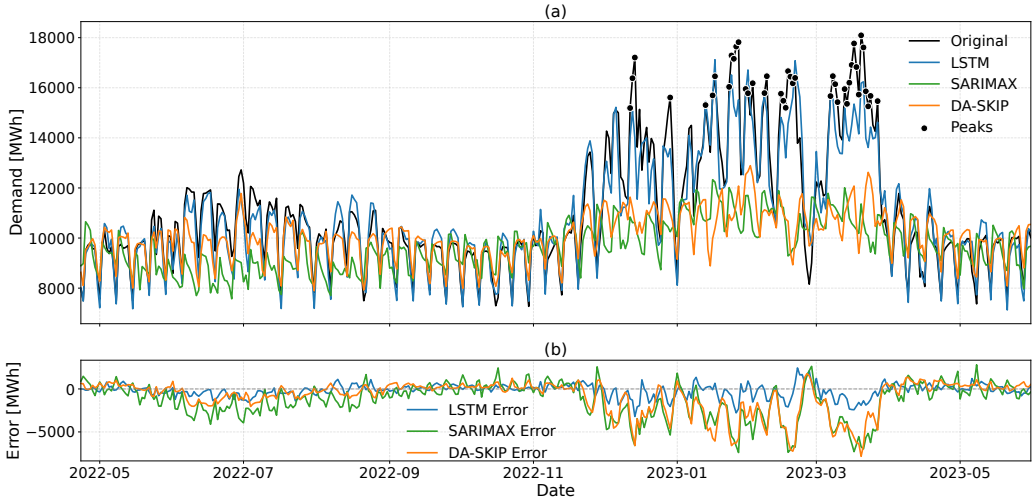

Figure 6: Comparison of LSTM, SARIMAX and DA-SKIP models for electricity demand forecasting. The panel (a) shows predicted vs. observed values, highlighting peak demand periods. The panel (b) displays the error trajectories over time, computed as the difference between predicted and observed value

pronounced for longer forecasting horizons. This tendency is commonly reported in meteorology-driven load forecasting studies and is often considered operationally preferable to overestimation during peak-demand conditions [6].

Figure 6 provides a comprehensive comparison of the LSTM, SARIMAX, and DA-SKIP models under operational conditions. The panel (a) contrasts their predictions against actual electricity demand, with peak demand days highlighted using black dots. While all models approximate the general trend of consumption, the LSTM network demonstrates a markedly superior ability to capture the demand variability. In contrast, SARIMAX and DA-SKIP systematically underestimates consumption demand. Focusing on the 41 days in the test period when electricity demand surpassed the 90th percentile thresh-

old, LSTM predictions maintain temporal coherence and amplitude under stress, whereas both SARIMAX and DA-SKIP exhibit larger deviations and fail to consistently replicate the actual peak patterns. These peak-demand days represent a small but operationally crucial subset of scenarios where accurate forecasting is essential.

Figure 6b shows the evolution of prediction errors over time, further confirming the stability of the LSTM model compared to the more erratic and dispersed error trajectories of SARIMAX and DA-SKIP. Quantitatively, LSTM errors during peak-demand days remain mostly below 2,000 MWh, while SARIMAX and DA-SKIP frequently exceeds 4,000 MWh, explaining the notably higher RMSE and severely negative R^2 observed for these models in peak-demand intervals. Rather than being anomalies, these poor metrics reflect their limited ability to adapt to abrupt demand surges. Overall, the figure reinforces the robustness and reliability of the LSTM model under stress conditions, validating its suitability for real-world applications.

4.3. Electricity Forecasting with LSTM

Figure 7a-b illustrates the predictive alignment of the LSTM model with actual demand for FTW1 and FTW7, where R^2 values reached 0.94 and 0.91, respectively, and RMSE values were 727.2 and 848.4 MWh. Similarly, 7c-d show the corresponding relationships for longer forecasting horizons, including FTW14, where a wider dispersion around the identity line and a lower R^2 (0.66) indicate the expected degradation in predictive accuracy as the forecast horizon increases. Nevertheless, the overall linear trend is preserved, confirming that the LSTM model maintains a coherent representation of demand dynamics even under extended prediction windows.

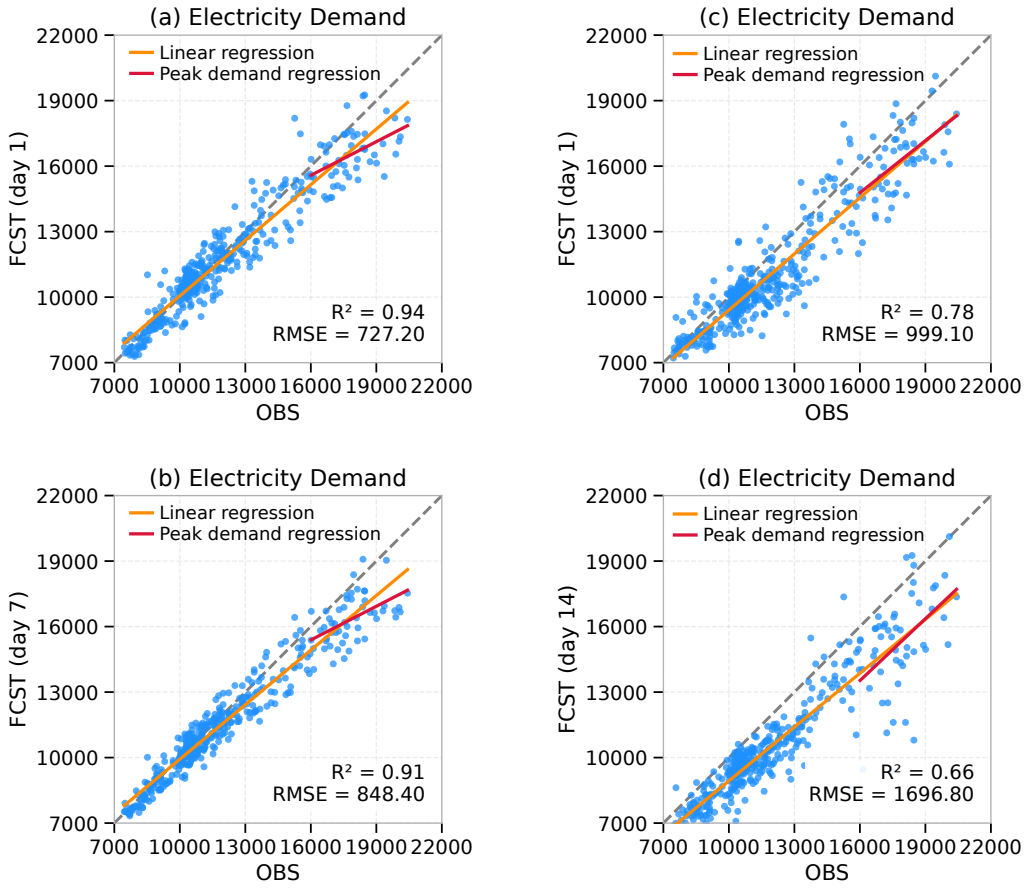

Figure 7: Predicted vs. observed values (Table 5)

Figure 8 presents a summary of RMSE and R^2 values across forecast horizons for the LSTM model using its best-performing configuration (batch size = 8). The panel (a) reports short- and mid-term forecasts (1 to 7 days), while the panel (b) includes the extended 14-day horizon. For short-term predictions (FTW1 and FTW7), both metrics remain stable and favorable, confirming the model's suitability for operational use. In contrast, FTW14 shows a noticeable drop in R^2 and a rise in RMSE, reflecting the expected trade-off

sinc(i) Research Institute for Signals, Systems and Computational Intelligence (sinc.unl.edu.ar)
M. Uhrig, L. D. Vignolo & O. Müller; "Multivariate forecasting of energy demand using recurrent neural networks"
Advanced Engineering Informatics, Vol. 71, 2026.

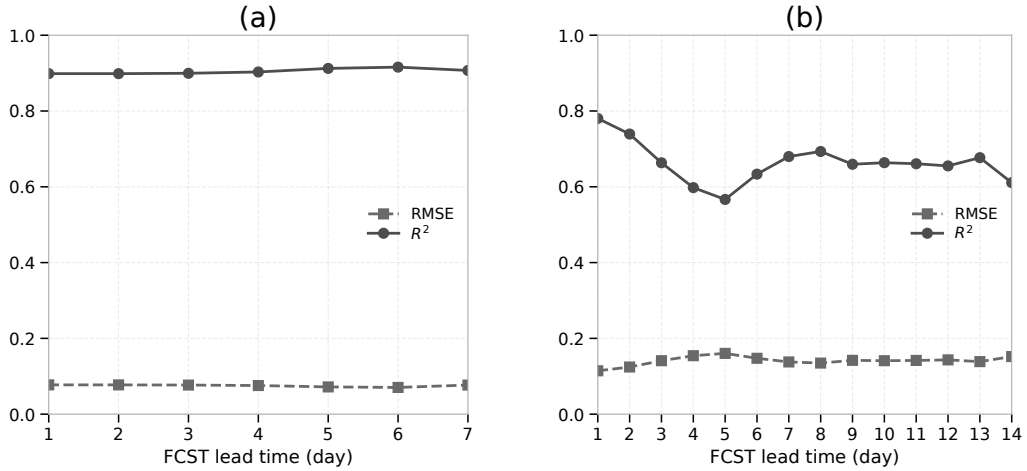

Figure 8: RMSE and R^2 for the LSTM model across forecast horizons using the best-performing configuration. (a): Forecasts up to 7 days and (b) extended to 14 days.

between forecast length and accuracy. These results highlight the model's current limitations for long-range forecasts and point to future work involving additional input enrichment or hybrid architectures to improve long-term predictive performance.

Unlike SARIMAX, the LSTM model consistently delivered superior performance across all forecast horizons, both under average and peak demand conditions. Its ability to capture nonlinear temporal dependencies and respond effectively to abrupt consumption shifts positions it as a robust and reliable tool for short-term electricity demand forecasting, especially in operational contexts where accurate peak prediction is critical for grid stability and energy planning.

These results emphasize the main contribution of this work: the development of a multivariate LSTM-based forecasting model that consistently outperforms SARIMAX and DA-SKIP across multiple horizons under real-

istic operational conditions.

5. Conclusions

This study developed and evaluated a multivariate deep-learning model based on LSTM networks for short-term electricity demand forecasting, explicitly comparing its performance against representative statistical and machine-learning benchmarks under operational conditions. The main results and contributions of this study can be summarized in four key findings.

First, the study demonstrates that incorporating exogenous meteorological variables into LSTM-based architectures is key for short-term electricity demand forecasting. Weather conditions, together with workable and non-workable days, strongly modulate electricity demand. Second, the proposed LSTM model achieves its highest accuracy at short-term horizons of 1, 7, and 14 days, with WI values reaching up to 0.90 and a bias of 11% even at a 14-day horizon. Although performance decreases for longer horizons due to increasing uncertainty, the LSTM model consistently outperforms both SARIMAX and DA-SKIP across all scenarios, demonstrating superior robustness and predictive skill. Third, the main advantage of the LSTM over the alternative methods lies in its superior ability to capture and predict high electricity demand peaks, which is a fundamental requirement for energy companies given the potential risk of service interruptions.

Lastly, the forecasting model shows clear potential for real-world deployment in regional electricity distribution systems, particularly for supporting operational decision-making during critical peak demand events. Beyond predictive accuracy, this work provides a robust and transferable framework

supported by a rigorous model design and optimization strategy, ensuring stable performance, strong generalization, and reliability under operational conditions.

Despite the promising results, some limitations of this study should be acknowledged. First, the forecasting model was developed and evaluated for a specific province of Argentina, modeling the electricity demand of Entre Ríos as a single aggregated system, which may mask localized consumption patterns; moreover, its application to other regions would require retraining and recalibration to account for different climatic, socio-economic, and consumption characteristics. Second, the analysis was limited to forecasting horizons of up to 14 days. While this temporal window is well suited for operational decision-making, forecasting accuracy degrades for longer horizons due to accumulated uncertainty and changing system dynamics; extending the model to longer-term predictions remains a topic for future research. Finally, the model relies mainly on historical demand and meteorological variables, excluding other potentially relevant factors (e.g. socioeconomic variables related to prices and income levels) due to the lack of consistent and reliable public data. Their inclusion could further improve forecasting performance and interpretability.

Future work will focus on the development of hybrid deep learning models aimed at further improving forecasting accuracy and interpretability. In addition, the deployment of the optimized model within transformer substations is envisioned to support localized energy management, incorporating sector-specific performance evaluations to adapt forecasts to residential, industrial, and touristic consumption patterns.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of Competing Interest

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.

Declaration of Generative AI and AI-assisted Technologies in the Writing Process

During the preparation of this work the authors used generative AI and AI-assisted technologies in order to improve the readability and language of the manuscript. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

References

[1] Mohamed Abdallah, Manar Abu Talib, Mariam Hosny, Omnia Abu Waraga, Qassim Nasir, and Muhammad Arbab Arshad. Forecasting highly fluctuating electricity load using machine learning models based on multimillion observations. *Advanced Engineering Informatics*, 523:101707, 2022.

[2] Mobarak Abumohsen, Amani Yousef Owda, and Majdi Owda. Electrical load forecasting using lstm, gru, and rnn algorithms. *Energies*, 16(5), 2023.

[3] Kadhem Al-Daffaiea, Mohanad S AL-Musaylh, Qasim Rhaif Mrebit Al-Faisal, and Qahtan Makki Shallal. Monthly exchange rate prediction based on artificial intelligence models and iraqi dinar against united states dollar. In *AIP Conference Proceedings*, volume 3232, page 030001. AIP Publishing LLC, 2024.

[4] Mohanad S. Al-Musaylh, Kadhem Al-Daffaie, Nathan Downs, Sujan Ghimire, Mumtaz Ali, Zaher M. Yaseen, Damien P. Igoe, Ravinesh C. Deo, Alfio V. Parisi, and Mustapha A. A. Jebar. Multi-step solar ultraviolet index prediction: integrating convolutional neural networks with long short-term memory for a representative case study in queensland, australia. *Modeling Earth Systems and Environment*, 11(1):77–90, 2025.

[5] Mohanad S. AL-Musaylh, Kadhem Al-Daffaie, and Ramendra Prasad. Gas consumption demand forecasting with empirical wavelet transform based machine learning model: A case study. *International Journal of Energy Research*, 45(10):15124–15138, 2021.

[6] Mohanad S. AL-Musaylh, Zahra Gharineiat, Kadhem Al-Daffaie, Khalid Fadhil Jasim, Ekta Sharma, and Abdullah A. Nahi. Predicting near-real-time total water level with an artificial intelligence model based on australia's tidal wave energy belt dataset. *Journal of Ocean Engineering and Marine Energy*, 11:679–699, 2025.

[7] Kelsy Cabello-Solorzano, Isabela Ortigosa de Araujo, Marco Peña, Luís Correia, and Antonio J. Tallón-Ballesteros. The impact of data normalization on the accuracy of machine learning algorithms: A comparative analysis. In *International conference on soft computing models in industrial and environmental applications*, pages 344–353. Springer, 2023.

[8] Sujan Ghimire, Mohanad S. Al-Musaylh, Thong Nguyen-Huy, Ravinesh C. Deo, Rajendra Acharya, Zaher Mundher Yaseen Casillas-Pérez, and Sancho Salcedo-Sanz. Explainable deeply-fused nets electricity demand prediction model: Factoring climate predictors for accuracy and deeper insights with probabilistic confidence interval and point-based forecasts. *Applied Energy*, 378:124763, 2025.

[9] Sujan Ghimire, Thong Nguyen-Huy, Mohanad S. AL-Musaylh, Ravinesh C. Deo, David Casillas-Pérez, and Sancho Salcedo-Sanz. Integrated multi-head self-attention transformer model for electricity demand prediction incorporating local climate variables. *Energy and AI*, 14:100302, 2023.

[10] Alok Raj Hans, Abhijeet Kumar Sharma, and Vandana Devi Wangkheirakpam. Electricity demand and price forecasting using lstm for smart grid operations. In S. Sachdeva, Y. Watanobe, and S. Bhalla, editors, *Big Data Analytics in Astronomy, Science, and Engineering*, volume 15546, page 236–250, Cham, 2025. Springer Nature Switzerland.

[11] Bingqing Huang, Haonan Zheng, Xinbo Guo, Yi Yang, and Ximing Liu. A novel model based on da-rnn network and skip gated recurrent neural network for periodic time series forecasting. *Sustainability*, 13(21), 2021.

[12] International Energy Agency. Electricity 2025: Entering the age of electricity. https://www.iea.org/reports/electricity-2025, 2025. Licence: CC BY 4.0.

[13] Vaia I Kontopoulou, Athanasios D. Panagopoulos, Ioannis Kakkos, and George K. Matsopoulos. A review of arima vs. machine learning approaches for time series forecasting in data driven networks. *Future Internet*, 15, 2023.

[14] X. J. Luo and Lukumon O. Oyedele. Forecasting building energy consumption: Adaptive long-short term memory neural networks driven by genetic algorithm. *Advanced Engineering Informatics*, 50:101357, 2021.

[15] Ivan Malashin, Vadim Tynchenko, Andrei Gantimurov, Vladimir Nelyub, and Aleksei Borodulin. Applications of long short-term memory (lstm) networks in polymeric sciences: A review. *Polymers*, 16(18), 2024.

[16] Shahenaz Mulla, Chaitanya B Pande, and Sudhir K Singh. Times series forecasting of monthly rainfall using seasonal auto regressive integrated moving average with exogenous variables (sarimax) model. *Water Resources Management*, 38(6):1825–1846, 2024.

[17] Keshab K. Parhi and Nanda K. Unnikrishnan. Brain-inspired computing: Models and architectures. *IEEE Open Journal of Circuits and Systems*, 1:185–204, 2020.

[18] Jason Runge and Radu Zmeureanu. A review of deep learning techniques for forecasting energy use in buildings. *Energies*, 14(3), 2021.

[19] Joaquín F. Torres, Francisco Martínez-Álvarez, and Alicia Troncoso. A deep lstm network for the spanish electricity consumption forecasting. *Neural Computing and Applications*, 34:10533–10545, 2022.

[20] Mariela Noelia Uhrig, Miguel Ángel Lovino, Leandro Daniel Vignolo, and Omar Vicente Müller. A recurrent neural network approach to energy demand forecasting using meteorological and energy data. In *CPAM-SIMCLEA*, page 622, Brazil, 2024.

[21] Mariela Noelia Uhrig, Leandro Daniel Vignolo, and Omar Vicente Müller. Electricity demand forecast model based on meteorological and historical demand data using artificial neural networks. In *JAIIO, Jornadas Argentinas De Informática*, page 106–118, Bahía Blanca, Argentina, 2024.

[22] Mariela Noelia Uhrig, Leandro Daniel Vignolo, and Omar Vicente Müller. Modelling and forecasting electricity demand: comparing statistical and machine learning approaches. *SADIO Electronic Journal of Informatics and Operations Research*, 24(1), 2025.

[23] Dong Yongquan, Zhang Zichen, and Hong Wei-Chiang. A hybrid seasonal mechanism with a chaotic cuckoo search algorithm with a support vector regression model for electric load forecasting. *Energies*, 11(4):1009, 2018.

[24] Zhang Zichen, Hong Wei-Chiang, and Li Junchi. Electric load forecasting by hybrid self-recurrent support vector regression model with

variational mode decomposition and improved cuckoo search algorithm. *IEEE Access*, 8:14642–14658, 2020.