Electricity Consumption Forecasting using Machine Learning Models: Comparison between SARIMA and LSTM Neural Networks

CAMILA EYZAGUIRRE[1], CRISTIAN LORCA[1], CINTYA OLIVARES[1], and ALEJANDRO SUAREZ[1]

[1]Universidad Andrés Bello, Santiago, Chile (e-mail: c.eyzaguirrevillarro@uandresbello.edu, c.lorcavargasvargas@uandresbello.edu, c.olivarescisternas@uandresbello.edu, a.suarezsantelices@uandresbello.edu)

Corresponding author: Camila Eyzaguirre (e-mail: c.eyzaguirrevillarro@uandresbello.edu).

This work was carried out within the framework of the Topics in Data Science course.

ABSTRACT Accurate electricity consumption forecasting is fundamental for efficient management of electrical networks. This study compares two machine learning approaches for hourly electricity consumption forecasting: SARIMA (Seasonal AutoRegressive Integrated Moving Average) and LSTM (Long Short-Term Memory) neural networks. A dataset with 244,391 training records (2018-2021) and 61,313 test records (2022) from multiple substations was used. The analysis included temporal decomposition, stationarity analysis, and evaluation of seasonal patterns. Results show that LSTM significantly outperforms SARIMA, with MAE of 19.57 kWh vs 578.46 kWh and R^2 of 0.9757 vs -0.2895, representing a 29.5-fold improvement in MAE. This work demonstrates the superiority of deep neural networks for capturing complex patterns in electricity consumption time series.

INDEX TERMS Time series, electricity consumption forecasting, SARIMA, LSTM, machine learning

I. INTRODUCTION

ELECTRICITY demand presents complex patterns that vary across multiple temporal scales: daily, weekly, and seasonal cycles, influenced by weather conditions, special events, and user behavior [1]. Accurate electricity consumption forecasting is essential for optimizing energy generation and distribution, reducing operational costs, and improving system reliability [2]. Power grid operators face the constant challenge of balancing energy supply and demand, as electricity cannot be easily stored at large scale. Prediction errors can result in significant cost overruns or, in extreme cases, supply interruptions.

Time series modeling for electricity consumption forecasting has evolved significantly. Traditional statistical models such as ARIMA (AutoRegressive Integrated Moving Average) and its seasonal variants SARIMA have been widely used due to their interpretability and computational efficiency [3]. These models assume linear relationships between past and future observations, demonstrating good performance in series with clear and stable seasonal patterns.

However, the limitations of linear models in capturing complex nonlinear relationships have motivated the exploration of machine learning techniques. Neural networks, particularly LSTM (Long Short-Term Memory) networks, have shown superior capability for modeling long-term temporal dependencies and capturing complex nonlinear patterns [4]. LSTMs were specifically designed to solve the vanishing gradient problem in traditional recurrent neural networks, enabling learning of long-range temporal dependencies.

The main objective of this work is to develop and compare machine learning models for hourly electricity consumption forecasting, evaluating the performance of statistical approaches (SARIMA) and deep learning approaches (LSTM). Specifically, the objectives include: conducting a comprehensive exploratory analysis of the dataset, developing an optimized SARIMA model through hyperparameter search, designing and training an appropriate LSTM architecture, comparing both models using standard metrics (MAE, RMSE, R^2, MAPE), and providing recommendations on model selection according to problem characteristics.

II. METHODOLOGY

A. DATASET DESCRIPTION

The dataset contains hourly electricity consumption records from multiple substations. The training set (2018-2021) has 244,391 records in 34,913 unique temporal points, while the test set (2022) has 61,313 records in 8,759 temporal points. Each record includes: substation, date (hourly timestamp), and consumption (kWh). The data contains no missing values.

B. EXPLORATORY ANALYSIS

The process began with a comprehensive exploratory analysis that included: time series visualization to identify trends and visual patterns, additive temporal decomposition to separate the series into trend, seasonality, and residual components, stationarity analysis using the Augmented Dickey-Fuller (ADF) test to determine if the series requires differencing, autocorrelation analysis through calculation of ACF and PACF functions to identify temporal dependencies and periodicities, and seasonal pattern analysis through identification of hourly, weekly, and monthly patterns.

The ADF test indicated that the series is stationary (p-value < 0.05, ADF Statistic: -18.67), so it did not require differencing (parameter $d = 0$). A strong daily pattern was identified (lag 24h with autocorrelation 0.7841) and 70 lags with significant autocorrelation. The most important lags were: lag 1h with correlation 0.9509 (very strong), lag 24h with 0.7841 (daily pattern), lag 48h with 0.6387, and lag 72h with 0.6058.

C. SARIMA MODEL

The SARIMA model is characterized by parameters $(p, d, q) \times (P, D, Q, s)$, where p is the order of the autoregressive process, d is the differencing order, q is the moving average order, P, D, Q are equivalent seasonal components, and s is the seasonal period (24 hours for hourly data).

A systematic hyperparameter search was performed using grid search, testing the following combinations: $p \in \{1, 2\}$ (based on PACF analysis showing significance at lags 1-2), $d = 0$ (determined by ADF test confirming stationarity), $q \in \{1, 2\}$ (based on ACF analysis), $P \in \{1, 2\}$ (detected seasonality), $D \in \{0, 1\}$, $Q = 1$ and $s = 24$ (daily cycle). Model selection was based on Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC), choosing the model that minimized these metrics.

Five parameter combinations were evaluated. The optimal configuration was SARIMA$(2, 0, 2) \times (2, 0, 1, 24)$ with AIC=387,450 and BIC=387,518. This model captures AR components of order 2 (dependency on 2 previous steps), MA components of order 2, seasonal AR components of order 2 with 24-hour period, and seasonal MA components of order 1.

D. LSTM MODEL

For the LSTM model, the following feature engineering process was implemented: 23 features were created including cyclic encoding of hour of day, day of week, and month of year (encoded as sine and cosine to capture the circular nature of these temporal variables), categorical features such as day of month and weekend indicator, lag features with consumption values at lags 1, 2, 3, 6, 12, and 24 hours to capture immediate temporal dependencies, and moving statistics with mean and standard deviation in windows of 6, 12, and 24 hours to provide context on local trend and recent variability.

Data was transformed to 3D sequence format $(samples, timesteps,$ with a lookback window of 24 hours, where each sample uses the previous 24 hours to predict current consumption. Min-Max normalization was applied to all features and the target, scaling values to the range [0, 1] to improve training convergence and prevent features with different scales from dominating learning.

The implemented model architecture consists of: LSTM Layer 1 with 128 units and return_sequences=True allowing information to flow to the next LSTM layer, Dropout 1 with rate 0.2 for regularization, LSTM Layer 2 with 64 units processing the output of the first layer, Dropout 2 with rate 0.2, Dense layer with 32 neurons and ReLU activation to capture nonlinear relationships, and Output layer with 1 linear neuron producing the consumption prediction. The model contains a total of 129,345 trainable parameters.

Training was performed using the Adam optimizer with initial learning rate of 0.001, MSE (Mean Squared Error) loss function, and MAE and MSE metrics. Early Stopping callbacks with patience of 15 epochs and learning rate reduction with factor 0.5 and patience of 7 epochs were implemented. The dataset was temporally split into 80% for training and 20% for validation, maintaining temporal order of the data.

E. EVALUATION

Both models were evaluated using the following metrics on the test set: MAE (Mean Absolute Error): MAE $= \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i|$, RMSE (Root Mean Squared Error): RMSE $= \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2}$, R² (Coefficient of determination): $R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}$, and MAPE (Mean Absolute Percentage Error): MAPE $= \frac{100}{n} \sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right|$, where y_i are the actual values, $\hat{y}_i$ are the predictions, and $\bar{y}$ is the mean of actual values.

III. DATA ANALYSIS

A. MAIN CHARACTERISTICS OF THE TIME SERIES

The exploratory analysis revealed that the electricity consumption time series presents the following fundamental characteristics. Regarding stationarity, the Augmented Dickey-Fuller test indicated that the series is stationary (p-value < 0.05, ADF Statistic: -18.67), so it did not require differencing. This simplified modeling and allowed working directly with the original series.

Regarding seasonality, weak seasonality was identified (seasonality strength: 0.2365) according to decomposition

analysis. However, clear patterns were observed: a strong daily cycle where autocorrelation at lag 24 hours was 0.7841, indicating a very pronounced daily pattern, and a moderate weekly cycle where autocorrelation was detected at lags near 168 hours (7 days), though less pronounced than the daily cycle.

Autocorrelation analysis revealed that 70 lags show significant autocorrelation, indicating strong temporal dependency. The most important lags were: lag 1h with correlation 0.9509 (very strong immediate dependency), lag 24h with 0.7841 (daily pattern), lag 48h with 0.6387, and lag 72h with 0.6058.

B. TEMPORAL DECOMPOSITION

The additive decomposition of the series revealed clearly differentiated components. The trend presents a range of [547.71, 1756.24] kWh with a standard deviation of 178.68 kWh, showing gradual variations over time that reflect structural changes in consumption. Seasonality shows an amplitude of 344.21 kWh, indicating moderate seasonal variations that capture repetitive cyclical patterns. Residuals present a standard deviation of 228.33 kWh, representing variability not explained by trend and seasonality, which includes random effects and non-modeled factors.

C. SEASONAL PATTERNS

Seasonal pattern analysis showed significant variability at multiple temporal scales. The hourly pattern shows that consumption presents pronounced variations during the day, with a coefficient of variation of 83.37%, where consumption peaks were observed at specific hours of the day and minimums during early morning hours. The weekly pattern indicates that consumption varies between days of the week, with a coefficient of variation of 86.12%, where weekends present different patterns than weekdays. The monthly pattern reveals seasonal variability between months (coefficient of variation: 86.64%), probably related to climate changes and seasonal usage patterns.

IV. EXPERIMENTS AND RESULTS

A. FEATURE SELECTION JUSTIFICATION

1) SARIMA Model

The SARIMA model is univariate and uses only the aggregated time series of total consumption (sum of all substations). This aggregation allows simplifying modeling by working with a single time series, capturing global patterns of the electrical system, reducing noise through aggregation, and facilitating result interpretation. The model output is the predicted total consumption in kWh for each hour.

2) LSTM Model

The LSTM model uses a multivariate approach with 23 carefully selected input features. Cyclic temporal features use cyclic encoding (sine and cosine) of hour, day of week, and month, allowing the model to capture the circular nature of these temporal variables and avoiding artificial discontinuities (for example, between 23:00 and 00:00 hours).

TABLE 1. Performance Metrics

Model	MAE (kWh)	RMSE (kWh)	R²	MAPE (%)
SARIMA	578.46	832.78	-0.2895	28.57
LSTM	19.57	36.33	0.9757	-

Lag features include consumption values at previous hours (lags 1, 2, 3, 6, 12, 24) that capture inertia and immediate temporal dependencies, with lag 24 hours being especially important as it reflects the daily pattern identified in the analysis. Moving statistics provide means and standard deviations in moving windows that offer context on local trend and recent variability, helping the model adapt to gradual changes.

The output is the predicted consumption for each individual substation, allowing more granular predictions than the aggregated SARIMA model.

B. TRAINING

SARIMA was trained with the optimal configuration $SARIMA(2, 0, 2) \times (2, 0, 1, 24)$ in 80.9 seconds. All coefficients were significant (p < 0.001). LSTM was trained for a maximum of 100 epochs with early stopping at the optimal iteration, showing stable convergence without significant overfitting.

C. ARCHITECTURES

SARIMA is represented by the equation:

$$(1-\phi_1 B-\phi_2 B^2)(1-\Phi_1 B^{24}-\Phi_2 B^{48})y_t = (1+\theta_1 B+\theta_2 B^2)(1+\Theta_1 B$$
(1)

LSTM uses a two-layer LSTM architecture that allows capturing temporal dependencies at multiple scales. The justification for this architecture is based on the fact that two LSTM layers allow capturing complex temporal dependencies at different time scales, dropout reduces overfitting without significantly compromising model capacity, and the intermediate dense layer adds additional nonlinear modeling capacity. The first LSTM layer (128 units) processes input sequences and passes contextual information to the second layer (64 units), which finally generates the representation used for prediction.

D. PERFORMANCE

Results on the test set are shown in Table 1.

1) Comparative Results Analysis

The LSTM model significantly outperforms SARIMA in all main metrics: MAE 29.5 times lower (19.57 vs 578.46 kWh), RMSE 22.9 times lower (36.33 vs 832.78 kWh), and R² of 0.9757 explaining 97.57% of variance vs negative R² of SARIMA.

Specifically analyzing the SARIMA model: the MAE of 578.46 kWh represents approximately 28.57% of average consumption, indicating relatively high errors. The RMSE

of 832.78 kWh, greater than MAE, suggests the presence of occasional large errors. The negative R^2 (-0.2895) indicates that the model performs worse than simply predicting the mean, which suggests problems in the model application at the aggregated level or difficulties in capturing test set variability. The model seems to have difficulties generalizing, possibly due to changes in patterns between training and test periods.

In contrast, analysis of the LSTM model shows excellent results: the MAE of 19.57 kWh represents a very low average absolute error, approximately 1-2% of typical consumption. The RMSE of 36.33 kWh, close to MAE, indicates relatively uniform errors without extreme outliers. The R^2 of 0.9757 indicates that the model explains 97.57% of variance in test data, which is excellent. The model demonstrates superior capability for capturing complex patterns and generalizing to new data.

2) Comparison with Literature Results

The LSTM R^2 (0.9757) is comparable or superior to values reported in literature. Several studies have demonstrated the superiority of deep learning-based models over traditional statistical methods for electricity consumption forecasting when sufficient data is available [5]. The performance of our LSTM model is in the upper range of these results, being comparable with recent studies reporting R^2 between 0.90 and 0.96 for different LSTM architectures in electrical load forecasting.

The low performance of the SARIMA model is unexpected and can be attributed to several factors: structural changes in consumption patterns between training and test periods, the aggregated nature of the SARIMA model which may lose important information at the substation level, and possible effects of external events (such as consumption pattern changes) that altered temporal relationships modeled by SARIMA.

V. CONCLUSIONS

This work presented a comprehensive comparative analysis between SARIMA and LSTM models for hourly electricity consumption forecasting. The main findings and conclusions are as follows.

Regarding the stated objectives, a comprehensive exploratory analysis of the dataset was achieved, identifying that the time series is stationary, presents strong daily patterns (lag 24h with correlation 0.7841) and moderate weekly cycles. Seasonality is weak but present, and there is strong temporal autocorrelation (70 significant lags). An optimized $SARIMA(2, 0, 2) \times (2, 0, 1, 24)$ model was developed through grid search, with all coefficients statistically significant. However, the model presented difficulties generalizing to the test set, obtaining relatively poor metrics. A two-layer LSTM architecture with 129,345 parameters was designed and implemented, using 23 carefully selected features. The model showed excellent learning and generalization capability.

Regarding model comparison, the LSTM model significantly outperformed SARIMA in all main metrics: MAE improved 29.5 times (19.57 vs 578.46 kWh), RMSE improved 22.9 times (36.33 vs 832.78 kWh), and R^2 of 0.9757 explaining 97.57% of variance vs negative R^2 of SARIMA.

Practical implications of these results are important for electrical system management. LSTM models can provide more accurate predictions, reducing operational costs and improving generation planning. The proposed architecture is scalable and can be adapted to different electrical systems. The use of multiple features (not just the univariate time series) significantly improves performance.

For electricity consumption forecasting problems with similar characteristics (sufficient historical data, multiple substations, complex patterns), the use of LSTM models is recommended due to their superior capability for capturing nonlinear relationships and complex temporal dependencies. Main limitations identified include: lower interpretability of the LSTM model compared to SARIMA, greater requirement of data and computational resources for training, and need for additional exogenous variables (temperature, special events) for future improvements.

In summary, this work demonstrates that deep learning models, specifically LSTM, offer significant advantages for electricity consumption forecasting when sufficient data and well-designed features are available, validating the stated objectives and providing a solid foundation for practical applications in the electrical industry.

VI. LIMITATIONS AND FUTURE WORK

A. LIMITATIONS

This work presents several limitations that must be acknowledged. Regarding dataset limitations: data covers a period of 4 years for training and 1 year for testing, which may be insufficient to capture all long-term patterns and structural changes. Important exogenous variables such as temperature, humidity, holidays, or special events that could improve predictions were not included. The SARIMA model was applied to aggregated data, losing granular information at the substation level that could be valuable.

Regarding methodological limitations: SARIMA hyperparameter search was limited due to computational constraints, potentially missing better unexplored combinations. The LSTM model uses a fixed two-layer architecture; the architecture space was not exhaustively explored. Advanced techniques such as attention mechanisms or Transformers that could further improve performance were not implemented. LSTM data preprocessing could be optimized through more rigorous temporal cross-validation.

Regarding technical limitations: the LSTM model requires significant computational resources for training and prediction, which can be a limitation in resource-constrained environments. Limited interpretability of the LSTM model hinders problem diagnosis and identification of factors that most contribute to predictions.

B. FUTURE WORK PROPOSALS

Based on identified limitations, the following research lines are proposed. For LSTM model improvements: explore advanced architectures such as bidirectional LSTMs, attention networks, Transformers, and hybrid models. Implement more exhaustive hyperparameter optimization using Bayesian Optimization and neural architecture optimization. Experiment with advanced regularization techniques such as adaptive dropout and batch normalization.

For incorporation of exogenous variables: integrate meteorological data (temperature, humidity, wind speed) that have strong correlation with electricity consumption. Include information on holidays, special events, day type (weekday, weekend, holidays), and economic indicators. Develop advanced multivariate models that integrate multiple data sources.

For granular modeling and ensembles: develop models specific to each substation or groups of substations with similar patterns. Implement ensemble strategies of multiple LSTM architectures and ensemble between statistical and deep learning models. Develop hierarchical models that predict at multiple levels maintaining coherence.

For evaluation and advanced metrics: develop economic metrics that reflect the real cost of errors considering peak hours. Implement uncertainty analysis through quantile regression and probabilistic models. Conduct more exhaustive temporal cross-validation with multiple test windows.

For interpretability: apply techniques such as SHAP or LIME to identify which features contribute most to predictions. Develop methods to visualize which temporal patterns the LSTM model captures. Design interpretable hybrid architectures that combine accuracy and explicability.

These proposals focus on improving model performance, increasing its interpretability, integrating additional information, and developing practical applications that could lead to even more accurate and useful prediction systems for efficient electrical system management.

CODE AVAILABILITY

The complete source code of this project, including the Jupyter notebook with all stages of exploratory analysis, data preprocessing, feature engineering, training of SARIMA and LSTM models, and comparative evaluation, is publicly available on GitHub: https://github.com/Cintya01/DC-SeriesDeTiempo/blob/main/SerieDeTiempo.ipynb

REFERENCES

[1] A. Bunn, D. Farmer, "Comparative models for electrical load forecasting," Wiley, 1985.

[2] H. T. Ha, J. A. V. Restrepo, F. E. Elkhatib, "Short-term load forecasting using long short-term memory network," in Proc. IEEE PES General Meeting, 2017.

[3] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, Time Series Analysis: Forecasting and Control, 4th ed. Hoboken, NJ, USA: Wiley, 2015.

[4] S. Hochreiter, J. Schmidhuber, "Long short-term memory," Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997.

[5] Z. Ahmad, A. S. Chen, T. T. T. Nguyen, L. Zhang, "Deep learning for short-term load forecasting: an overview and comparative analysis," IEEE Access, vol. 9, pp. 156388–156416, 2021.