Electricity demand forecast model based on meteorological and historical demand data using artificial neural networks

Mariela N. Uhrig[1], Leandro D. Vignolo[2], and Omar V. Müller[3]

[1] Centro de Investigación Científica y de Transferencia Tecnológica a la Producción, CICYTTP-CONICET-PROV. E. R.-UADER
[2] Research Institute for Signals, Systems and Computational Intelligence, sınc(i), FICH-UNL/CONICET
[3] Centro de Estudios de Variabilidad y Cambio Climático,FICH-UNL/CONICET
mnuhrig@conicet.gov.ar, ldvignolo@sinc.unl.edu.ar, ovmuller@gmail.com

Abstract. Accurate forecasting of electricity demand is crucial for improving transmission system operation through optimized use of resources, operation planning, and minimized outages. The dynamic of electricity demand depends on exogenous factors (e.g., meteorological conditions), but the relationships between demand and factors are complex and nonlinear, posing a challenge for accurate prediction.
With the aim of predicting electricity demand, this work explores the relationship with meteorological conditions for the province of Entre Ríos (Argentina). We propose a recurrent neural network model based on long short-term memories, which receives the raw input data without feature extraction. We evaluate its performance and compare it with a state-of-the-art method.
The exploratory analysis of the data shows that temperature extremes present a strong influence on consumption patterns. The proposed models achieve a performance of 0.77 in determination coefficient when comparing predicted electricity demand with observations. This indicates the potential as a powerful tool for optimizing the system operation in Entre Ríos.

Keywords: electricity demand forecast · weather conditions · deep learning · artificial neural network.

1 Introduction

Electricity demand forecasting is a key input for operational and strategic decision-making [1]. Each country tries to use as little energy as possible in different areas from buildings to farms, from industrial processes to vehicles [2]. Having prior information on the electricity needs for the coming days is crucial for companies providing this service, as it enables them to plan their daily operations effectively and determine necessary contingency measures. Thus, they can minimize operational issues, prevent service disruptions or equipment damage due to overloading, and avoid blackouts and financial losses. The electricity consumption is influenced by different external factors such as water, wind, temperature, which make its prediction complex [3].

Numerous studies have explored methodologies to forecast energy demand, ranging from traditional statistical approaches to advanced techniques such as artificial neural networks (ANNs). The main challenge is the difficulty in capturing non-linear relationships and the limited long-term forecasting capacity. In [4] they tested different ANNs to forecast the daily electricity demand in Greece, including ambient temperature, relative humidity, among others, as input variables. In [5] they chose to apply the backpropagation neural network model and trend extrapolation method to forecast energy demand. The information used as input was the primary, secondary, and tertiary value of the industry, energy consumption, the level of urbanization, among others, without taking into account weather data. They found that the precision of the neural network was much higher than trend extrapolation. Furthermore, the ANN backpropagation network model is used to forecast Turkey's electricity demand based on different socio-economic indicators [6]. The applications of traditional techniques such as econometric and time series models along with soft computing methods such as neural networks, fuzzy logic, and other models are reviewed in [7] and verified the remarkable performance of the neural network models. Particularly, an important advantage of deep neural networks is that they can automatically extract features from raw data to support predictions [12]. While deep neural network techniques have shown promising results across various domains in recent years [8,9,10,11], their application to energy demand forecasting based on meteorological variables remains relatively unexplored.

Focusing on ENERSA, the company that provides the service in the province of Entre Ríos, relies on rudimentary tools to make decisions based on energy demand. For instance, they adjust the service cost according to an analysis of the seasonal variations in the demand from historical records. Other aspects they take into account for decision-making are the GPD (from English, Gross Domestic Product) forecasts or available economic forecasts, the weather seasonal forecast by the SMN (from Spanish, Servicio Meteorológico Nacional), and the behavior of the activity of large customers and the residential sector [O. Bustamante, 2019, personal communication]. In this study, we contribute to the needs of ENERSA by developing an energy demand forecast model based on machine learning methods, aiming to improve decision-making processes and optimize electricity distribution in the province.

The proposed forecast is based on historical daily demand and meteorological data. We initially focused on developing a method to forecast the electricity demand in Entre Ríos for one day using information from the previous 7 days. In the future, the developed model could be forced with 7-day weather forecasts produced by climate models, and thus, identify critical days of the coming week that may lead to failures in the regional system.

2 Methods

In practice, it is often challenging to determine whether a time series is generated from a linear or nonlinear underlying process, as thereby, which method is more effective in out-of-sample forecasting. Thus, it is difficult for forecasters to choose the right technique for particular applications. Typically, a number of different models are tried and the one with the most accurate result is selected. We started with ARIMA given that most studies in the literature use it as a benchmark. [13,14,15].

2.1 ARIMA

In an autoregressive integrated moving average model, the future value of a variable is assumed to be a linear function of several past observations and random errors [16]. That is, the underlying Hprocess that generates the time series has the following form:

$$(1-B)^d y_t = \theta_0 + \phi_1 y_{t-1} + ... + \phi_p y_{t-p} + \epsilon_t - \theta_1 \epsilon_{t-1} - ... - \theta_q \epsilon_{t-q} \quad (1)$$

where $(1-B)^d y_t$ indicates that the series has been differenced d times; y_t and t are the actual value and random error at time period t, respectively; ϕ $(i = 1, 2, ..., p)$ and θ $(j = 0, 1, 2, ..., q)$ are model parameters. d, p, and q are integers and are often referred to as orders of the model. Random errors, ϵ_t, are assumed to be independently and identically distributed with a mean of zero and a constant variance of 2. Eq. 1 entails several important special cases of the ARIMA family of models. This means p are the last values of the time series used as regressors, q is the size of the moving average window, which means that an average term based on the error of the previous time step is included, and d indicates the number of times the data is differenced to remove trends and make the series more stationary.

2.2 Artificial Neural Networks (ANNs)

ANNs are powerful functions that simulate how the human brain processes information. An ANN is composed of a network of processing nodes (or neurons), which perform numerical manipulations and are interconnected in a specific order. The historical data can be used by ANNs to predict the future values of the noisy multivariate times series [17]. Feed-forward ANNs can be applied to sequential or time series data, however, there are several issues that render them unsuitable for these types of problems.

Long Short-Term Memory (LSTM) Recurrent neural networks (RNN) are able to capture the dynamics of sequences via recurrent connections, overcoming the limitations of feed-forward ANNs. LSTM is a special type of RNN, which

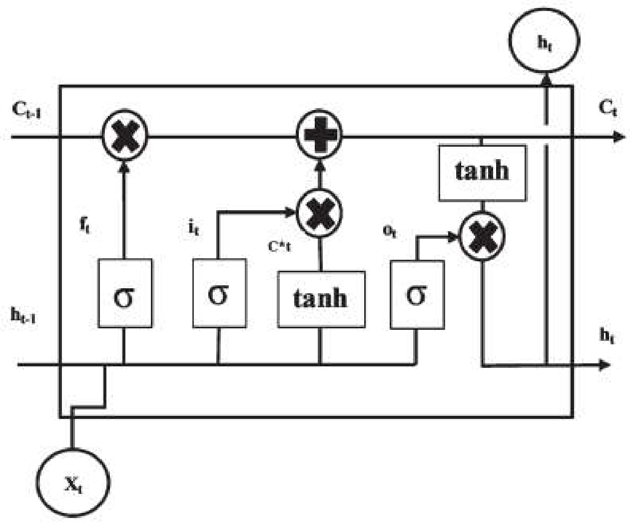

Fig. 1. LSTM unit architecture

has the capability to learn longer dependencies in the dataset [17,18]. LSTMs have been effectively used in several time series forecasting applications, e.g. in [19].

The simple LSTM architecture with forget gate [20] is depicted in Figure 1. The forget gate determines the unnecessary component from the previous cell state which can be computed as follows:

$$f_t = W_f X_t + U_f h_{t-1} + b_f \tag{2}$$

where f_t is the forget gate activation at time t. This vector determines which part of the past information to forget, W_f is the weight matrix applied to the input X_t at time t, U_f is the recurrent weight matrix applied to the previous hidden state h_{t-1} (the output of the LSTM cell in the previous time step) and b_f the ias term for the forget gate. The state update of the cell is determined by the input gate and *tanh* layer, which is calculated as

$$i_t = W_i X_t + U_i h_{t-1} + b_i, \tag{3}$$

where i_t is the input gate activation at time t. This vector decides how much new information to store in the cell state, W_i is the weight matrix applied to the input X_t, U_i recurrent weight matrix applied to the previous hidden state h_{t-1}, b_i the bias term for the input gate.

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

$$C_t^* = \tanh(W_c X_t + U_c h_{t-1} + b_c) \tag{4}$$

where tanh hyperbolic tangent activation function, ensuring the output values are between *-1* and *1*, W_c 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 at time t is a combination of the previous cell state, modulated by the forget gate, and the new candidate memory, modulated by the input gate which is calculated as

$$C_t = f_t C_{t-1} + i_t \tilde{C}_t \tag{5}$$

where $f_t C_{t-1}$ 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 $i_t \tilde{C}_t$ 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 cell to the next cell is calculated by the output gate as

$$o_t = W_o X_t + U_o h_{t-1} + b_o \tag{6}$$

where o_t the output gate activation at time t. This determines which part of the cell state will be used for the output. W_o 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 bias term for the output gate.

$$h_t = o_t \tanh(C_t) \tag{7}$$

where h_t hidden state at time t. This is the output of the LSTM cell, based on the output gate activation and the current cell state, o_t output gate activation, modulating the influence of the cell state on the output and $\tanh(C_t)$ application of the hyperbolic tangent function to the cell state C_t), ensuring the output values are between *-1* and *1*.

2.3 Metrics for performance evaluation

For evaluating training and prediction performances we considered the Root Mean Squared Error (RMSE). It provides a measure of the average size of the errors between predicted and actual values, with lower values indicating better model performance. In addition to the RMSE, we also use the coefficient of determination (R^2). It represents the proportion of the variance in the dependent variable that can be predicted from the independent variables. A higher R^2 value indicates that the model explains a greater proportion of the variance in the target variable and thus has better predictive performance.

3 Data description

The electricity and meteorological data are daily time series for the period 2012 to 2023 obtained from two sources: Energía de Entre Ríos S.A. (ENERSA)

Table 1. Dataset variables

Variable	Description
Year	The year of the observation.
Month	The month of the observation.
Encoded Day	Encoded representation.
Minimum Temperature	The lowest temperature recorded.
Mean Temperature	The average temperature recorded.
Maximum Temperature	The highest temperature recorded.
Skin Temperature	The temperature of the skin surface.
Surface Runoff	The amount of water flowing over the surface.
Total Precipitation	The total amount of precipitation.
Surface Latent Heat Flux	The flux of latent heat at the surface.
Surface Pressure	The pressure exerted by the atmosphere at the surface.
Surface Sensible Heat Flux	The flux of sensible heat at the surface.
Surface Short-Wave Solar Radiation Downwards	The amount of solar radiation reaching the surface.
Long-Wave Thermal Radiation Downwards	The amount of thermal radiation emitted by the surface.
Surface Net Solar Radiation	The net balance of solar radiation at the surface.
Surface Net Thermal Radiation	The net balance of thermal radiation at the surface.
Total Cloud Cover	The percentage of sky covered by clouds.
Wind Speed	The speed of the wind.
Relative humidity	Measurement of water vapor content in air.
Number of Service Users	The number of users of an electricity service.

and the fifth-generation ECMWF atmospheric reanalysis of the global climate (ERA5). After preprocessing, the final dataset includes a total of twenty variables and it can be seen in Table 1:

The input of the model includes historical values of variables over 7 days. The output is the demand for electricity on the eighth day. To create the data set, we define a time window of 7 days, indicating the number of historical data points the neural network will consider to predict the next day's demand. Observed electricity demand is the target variable. This data was split as follows: 80% for training, 10% for validation, and 10% for testing. The split is performed using consecutive periods of daily data, for instance, January 2012 to February 2021 for training, March 2021 to April 2022 for validation, and May 2022 to May 2023 for testing. This approach ensures that each subset of data (training, validation, and testing) preserves the inherent temporal structure which includes a marked annual cycle. This is crucial for capturing seasonal fluctuations and other temporal patterns that could impact the predictive performance of the model.

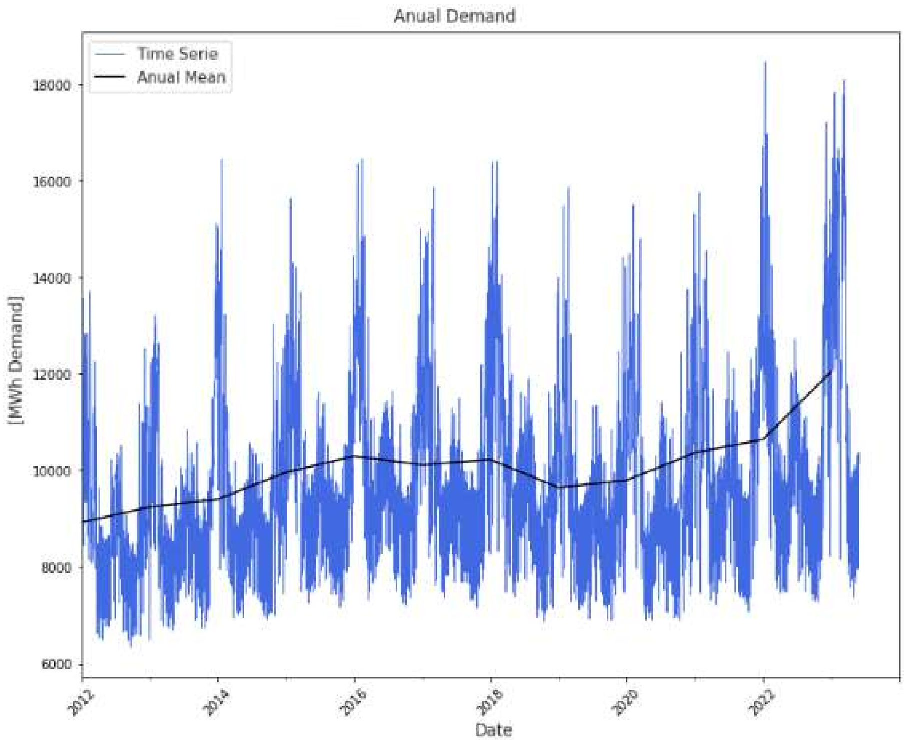

Fig. 2. Annual demand time series with trendline [MWh]

4 Experiments and results

4.1 General features of the demand series

Exploratory analysis was conducted to explore the relationship between energy and meteorological variables. The electricity demand over the entire study period shows an upward trend in demand over the years, as can be seen in Figure 2. Notably, there were discernible spikes in demand in 2016 and 2019, corresponding to the increase of tariffs suffered in Argentina during those years. These fluctuations highlight the influence of external factors such as economic conditions and population growth on energy consumption patterns.

From the annual cycle analysis represented in Figure 3, we observe that energy consumption follows a “W” pattern, with high consumption during summer months, but also during winter, and less consumption during intermediate seasons.

The left panel of Figure 4 illustrates the dynamic of energy consumption across different days of the week. The results indicate that energy consumption peaks from Tuesday to Thursday, while Sunday records the lowest levels of con-

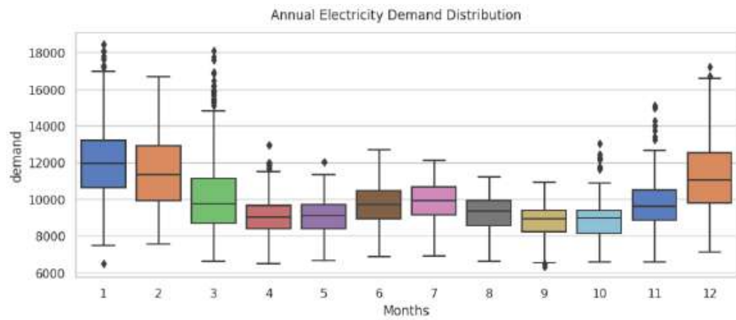

Fig. 3. Annual demand cycle [MWh]

sumption. When holidays are isolated in the analysis, as shown in the right panel of Figure 3, the boxplot suggests that energy consumption during holidays is comparable to that on weekends. Conversely, the consumption on working days shows a slight increase compared to the boxplots presented in the left panel, which include holidays. These findings align with previous studies [21,22,23].

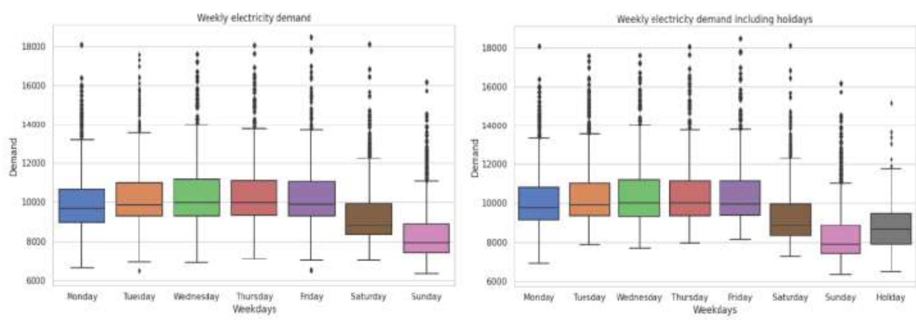

Fig. 4. Left: Energy consumption across different days of the week. Right: Energy consumption across different days of the week including holidays.

Lastly, we evaluate the relationship between input variables and energy demand with scatterplots (Figure 5). The results highlight the strong influence of temperature on energy consumption, while the direct impact of other variables is not as clear.

4.2 Models evaluation

For the ARIMA model, several parameter configurations were tested, with p=*7*, d=*2*, and q=*7* achieving the best performance. This indicates that order 7 was

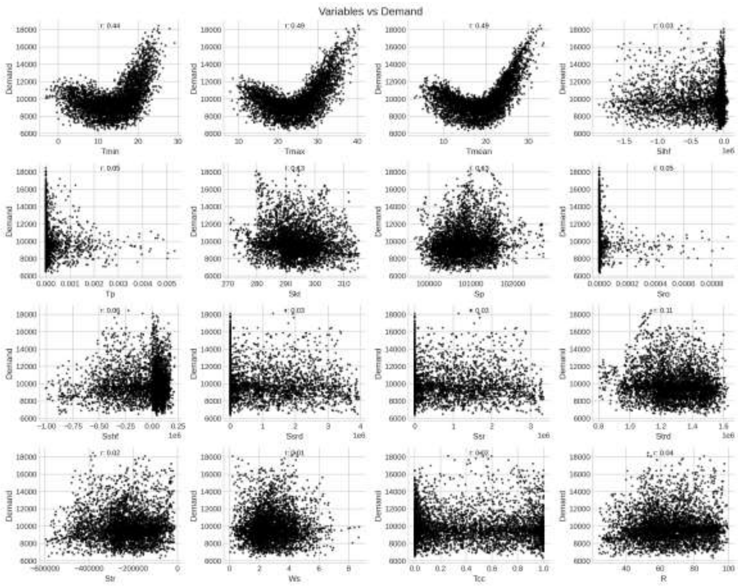

Fig. 5. Relationship between input variables and energy demand

used for the seasonal autoregressive component, 2 indicates that a second-order differentiation was necessary, and order 7 was used for the moving average component. The approximation of ARIMA models to complex non-linear problems may not be adequate, obtaining R^2=0,05 and RMSE=0,22. This can be seen in the Figure 6.

For LSTM models, a total of thirty-one different settings were explored using the options described in section II. 1-layered, 2-layered, and 3-layered LSTM structures are used for modeling with different amounts of neurons in each hidden layer. After the LSTM layers, a fully connected layer is incorporated to further process the learned features and capture complex patterns in the data. The number of layers and units in the LSTM was selected via trial-and-error, as well as the training parameters learning rate, and batch size. The performance of each architecture considered was evaluated using different configurations regarding learning rate, batch size, and number of epochs.

In Table 2, different configurations used in the experiments along with their corresponding metric values for each dataset can be observed. For the experiments in the table, Adam optimizer [24] was used and a patience of 150 epochs for early stopping. It was observed in the experiments that larger batch sizes and increasing the number of neurons in the layers led to a loss of generalization capacity and performance and required a greater number of epochs during training.

For each batch size explored, the hyperparameters and their respective values for the two performance metrics achieved good prediction performance. The

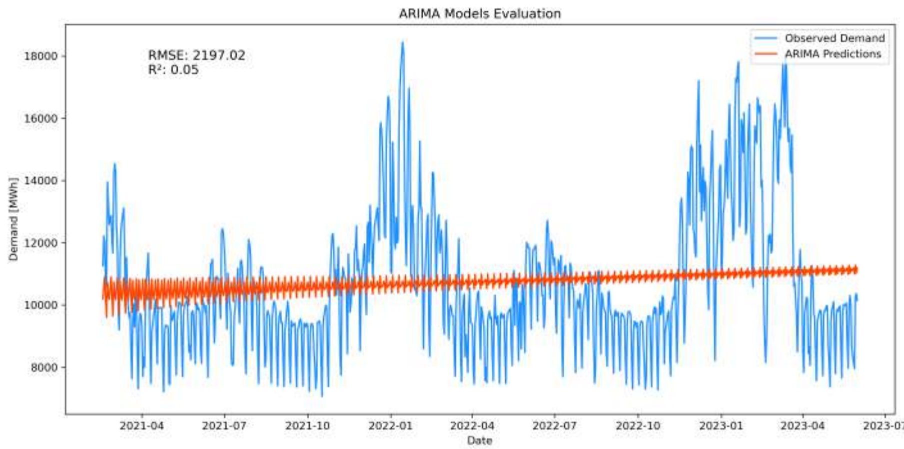

Fig. 6. ARIMA model evaluation in data test

best configurations are those with, 2 layers and 5 neurons, and, 2 layers and 10 neurons, all yielding a test R^2 above 0.70, as highlighted in Table 2. Furthermore, as shown in Figure 7, the model has successfully captured the underlying patterns and dynamics of the time series data, significantly outperforming the ARIMA model, which did not perform well.

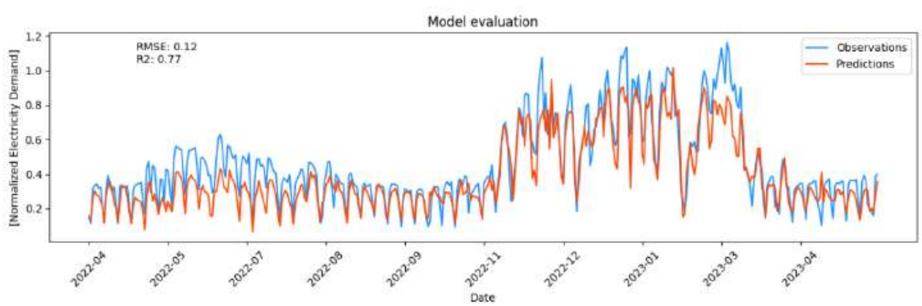

Fig. 7. LSTM model evaluation in data test

For the optimal model, which achieves an R^2 value of 0.77 on the test set, the corresponding R^2 value on the training set is 0.64. This discrepancy indicates an absence of overfitting.

Table 2. Configurations evaluated for the LSTM model and their metrics.

					Train		**Validation**		**Test**	
Layers	**Units**	**Lr**	**Epochs**	**Batch**	**RMSE**	R^2	**RMSE**	R^2	**RMSE**	R^2
1	10	0,0001	200	4	0,09	0,73	0,16	0,38	0,20	0,29
1	20	0,0001	200	4	0,08	0,78	0,11	0,68	0,17	0,51
1	20	0,0001	200	8	0.09	0.69	0,14	0,53	0,20	0,35
1	40	0,0001	200	4	0,07	0,84	0,12	0,62	0,19	0,41
1	40	0,0001	200	8	0,08	0,79	0,13	0,59	0,20	0,37
1	40	0,0001	200	10	0,08	0,78	0,13	0,60	0,18	0,43
1	40	0,0001	400	8	0,06	0,87	0,13	0,57	0,20	0,34
1	60	0,0001	200	4	0,08	0,84	0,15	0,46	0,22	0,14
1	40	0,001	200	4	0,05	0,92	0,12	0,63	0,15	0,59
1	40	0,001	200	8	0,05	0,90	0,13	0,56	0,17	0,50
1	40	0,001	200	10	0,05	0,90	0,13	0,56	0,19	0,41
2	5	0,001	150	10	0,10	0,67	0,12	0,64	0,17	0,50
2	5	0,001	150	16	0,09	0,73	0,13	0,58	0,18	0,42
2	**5**	**0,001**	**100**	**4**	**0,08**	**0,78**	**0,10**	**0,76**	**0,12**	**0,75**
2	5	0,001	100	8	0,08	0,77	0,11	0,69	0,17	0,48
2	**5**	**0,001**	**100**	**10**	**0,10**	**0,64**	**0,10**	**0,73**	**0,12**	**0,77**
2	5	0,001	100	16	0,10	0,62	0,11	0,68	0,14	0,68
2	5	0,001	100	32	0,10	0,64	0,12	0,65	0,18	0,48
2	5	0,001	50	8	0,10	0,63	0,13	0,59	0,17	0,53
2	5	0,001	50	10	0,11	0,55	0,12	0,67	0,14	0,66
2	5	0,001	50	16	0,12	0,45	0,14	0,49	0,19	0,40
2	5	0,001	50	32	0,12	0,51	0,14	0,52	0,18	0,47
2	**10**	**0,001**	**50**	**4**	**0,08**	**0,77**	**0,10**	**0,75**	**0,12**	**0,75**
2	10	0,001	50	8	0,08	0,76	0,10	0,73	0,14	0,67
2	**10**	**0,001**	**50**	**10**	**0,08**	**0,75**	**0,10**	**0,75**	**0,13**	**0,72**
2	10	0,001	50	16	0,10	0,67	0,12	0,63	0,17	0,52
2	20	0,001	50	4	0,08	0,78	0,12	0,64	0,17	0,52
2	20	0,001	50	8	0,08	0,75	0,11	0,07	0,15	0,62
2	20	0,001	50	10	0,09	0,68	0,13	0,60	0,17	0,49
2	20	0,001	50	16	0,09	0,71	0,12	0,64	0,17	0,50
2	20	0,001	50	32	0,09	0,72	0,13	0,60	0,19	0,37

5 Conclusions and future work

This paper presents a recurrent artificial neural network architecture for electricity demand forecasting using a case study of the province of Entre Ríos, Argentina. In particular, this study innovates by incorporating both historical energy consumption variables and meteorological variables as inputs to the forecasting model, providing a comprehensive approach to energy demand forecasting. Preliminary results show a consistent level of competence in accurate demand forecasting. The proposed models achieve a performance of 0.77 in coefficient of determination with 0.12 RMSE when comparing predicted electricity demand with observations. To potentially improve the performance of the models, it could be beneficial to extend the exploration of the hyperparameters or modify the early stopping criteria. Finally, future work could include testing other deep learning models, such as combining convolutional neural networks with LSTM, to further improve the accuracy and robustness of electricity demand forecasting.

References

1. Pai, Hong, Forecasting regional electricity load based on recurrent support vector machines with genetic algorithms, Electric Power Systems Research, Volume 74, 3, 417-425, ISSN 0378-7796 (2005)
2. Barak, S. and S.S. Sadegh, Forecasting energy consumption using ensemble ARIMA– ANFIS hybrid algorithm. International Journal of Electrical Power & Energy Systems (2016). 82: p. 92-104.
3. Zhao, H.-x. and F. Magoulès, A review on the prediction of building energy consumption. Renewable and Sustainable Energy Reviews (2012). 16(6): p. 3586-3592.
4. Ekonomou, L., Oikonomou, D.S.: Application and comparison of several artificial neural networks for forecasting the Hellenic daily electricity demand load. In: Proceedings of the 7th WSEAS International Conference on Artificial Intelligence, Knowledge Engineering and data bases, pp. 67–71. World Scientific and Engineering Academy and Society (WSEAS) (2008).
5. Wang, J.M., Liang, X.H.: The forecast of energy demand on artificial neural network. In: Artificial Intelligence and Computational Intelligence. AICI'09. International Conference. vol. 3, pp. 31—35. IEEE Computer Society, Washington, DC, United States (2009)
6. Kavaklioglu, K.: Modeling and prediction of Turkey's electricity consumption using Support Vector Regression. Appl. Energy (2011). 88(1), 368–375
7. Ghalehkhondabi, I., Ardjmand, E., Weckman, G.R. et al. An overview of energy demand forecasting methods published in 2005–2015. Energy Syst (2017). 8, 411–447. https://doi.org/doi:10.1007/s12667-016-0203-y
8. Rivenson, Y., Zhang, Y., Günaydın, H. et al. Phase recovery and holographic image reconstruction using deep learning in neural networks. Light Sci Appl 7, 17141. (2018) https://doi.org/doi.org/10.1038/lsa.2017.141
9. Belthangady, C., Royer, L.A. Applications, promises, and pitfalls of deep learning for fluorescence image reconstruction. Nat Methods (2019). 16, 1215–1225. https://doi.org/doi.org/10.1038/s41592-019-0458-z
10. Pereira, T.D., Aldarondo, D.E., Willmore, L. et al. Fast animal pose estimation using deep neural networks. Nat Methods (2019). 16, 117–125. https://doi.org/doi.org/10.1038/s41592-018-0234-5
11. M. M. Sachin and Melvin Paily Baby and Abraham Sudharson Ponraj. "Analysis of energy consumption using RNN-LSTM and ARIMA Model", Journal of Physics: Conference Series. (Dec, 2020) vol.1716, n°1, sp. 012048. https://doi.org/dx.doi.org/10.1088/1742-6596/1716/1/012048
12. P. Vincent, H. Larochelle, Y. Bengio, and P. Manzagol, Extracting and composing robust features with denoising autoencoders, in Proc. 25th Int. Conf. Mach. Learn. (ICML'08) , pp. 1096—1103. Association for Computing Machinery, New York (2008)
13. R. Sharda, R.B. Patil, Neural networks as forecasting experts: an empirical test, in: Proceedings of the International Joint Conference on Neural Networks, Washington, D.C., Vol. 2 (1990), pp. 491– 494.

14. Z. Tang, C. Almeida, P.A. Fishwick, Time series forecasting using neural networks vs Box–Jenkins methodology, Simulation 57 (1991) 303–310.
15. Z. Tang, P.A. Fishwick, Feedforward neural nets as models for time series forecasting, ORSA J. Comput. 5 (1993) 374–385.
16. Brockwell, P. J., & Davis, R. A. Introduction to time series and forecasting (3a ed.). Springer, New York (2016) https://doi.org/10.1007/978-1-4757-2526-1
17. Adamowski, J., Fung Chan, H., Prasher, S.O., Ozga-Zielinski, B., Sliusarieva, A.: Comparison of multiple linear and nonlinear regression, autoregressive integrated moving average, artificial neural network, and wavelet artificial neural network methods for urban water demand forecasting in Montreal, Canada. Water Resources Research 48(1) (2012)
18. Q. Gao, Y. Liu, J. Yang and Y. Hong, Çomparative Research on Electricity Consumption Forecast Based on Deep Learning" 2nd International Conference on Artificial Intelligence and Education (ICAIE), Dali, China. (2021) pp. 213-217.https://doi.org/doi:10.1109/ICAIE53562.2021.00052
19. F. A. Gers. Learning to forget: continual prediction with LSTM, in: 9th Int. Conf. Artif. Neural Networks ICANN '99, IEE pp. 850–855. (1999) https://doi.org/doi:10.1049/cp:19991218
20. R. Jozefowicz, W. Zaremba and I. Sutskever, An empirical exploration of recurrent network architectures, in: Int. Conf. Mach. Learn, pp. 2342-–2350. Association for Computing Machinery, New York (2015) https://doi.org/doi:10.1109/CVPR.2015.729876
21. Gutiérrez, E. La demanda residencial de energía eléctrica en la Comunidad Autónoma de Andalucía; un análisis cuantitativo. Tesis de Doctorado. Facultad de Ciencias Enconómicas y Empresariales, Universidad de Sevilla (2003).
22. Medina, S.; García, J. Predicción de demanda de energía en Colombia mediante un sistema de inferencia difuso neuronal. Revista Energética (2005) 33,15-24.
23. Murillo, J; Trejos, A. Carvajal, P. Estudio del pronóstico de la demanda de energía eléctrica utilizando modelos de series de tiempo. Scientia et Technica (2003) 23, 37-42.
24. Kingma, D. P., Ba, J. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980. (2014) Retrieved from https://arxiv.org/abs/1412.6980