Theme: AI based solutions in modelling methods (3)

Performance evaluation of a Transformer-based energy forecasting model for heterogeneous end-users: agriculture, port, community, and aquaculture

Mariana Akemi Ikegawa Bernabé[1], Rafael González Perea[1], Juan Antonio Rodríguez Díaz[1], Jorge García Morillo[1]

[1] University of Cordoba, 5 Medina Azahara Avenue, Cordoba 14071, Spain

Corresponding author: p82ikikm@uco.es

Abstract. Accurate energy demand (ED) forecasting is essential for optimizing the use of water and energy resources across diverse sectors. This study evaluates a Transformer-based encoder–decoder model for hourly, medium-term ED forecasting in four end-users: fisheries (Ireland), an irrigation district (Spain), a port (Spain), and a local energy community (Portugal). Input variables were selected through fuzzy logic and correlation analysis, resulting in 7 inputs for the Fisheries, Port, and Community cases and 12 inputs for the Irrigation District. The proposed model achieved consistently high predictive performance across all case studies, with R^2 values between 97.9% and 99.62% and low error metrics. The highest accuracy was obtained for the Irrigation District (R^2 = 99.62%, MSE = 0.038), while the Community and Port cases showed the lowest MAE values (0.44 kWh and 6.64 kWh, respectively). Differences in absolute error levels are mainly associated with variations in demand magnitude and operational variability. Overall, the results confirm the robustness and generalization capability of Transformer-based models under diverse input configurations and data conditions, supporting their applicability for reliable ED forecasting and operational planning in multi-sector environments.

Keywords: artificial intelligence; deep learning; energy resource management; renewable energy; time series forecasting; transformer neural network.

1 Introduction

Electricity is a fundamental energy form driving socioeconomic development [1]. With over 50% of the global population residing in urban areas, which consume two-thirds of primary energy and produce 70–80% of energy-related greenhouse gas emissions [2], there is a pressing need to implement sustainable energy practices in rural areas, where access is limited but critical for resilience [3]. Traditional fossil fuel–based energy sources, such as oil, coal, and natural gas, are increasingly unsustainable due to their environmental, economic, and health impacts. Moreover, international climate agreements, such as the Kyoto Protocol, impose strict targets for reducing greenhouse gas emissions [4]. Fossil fuel emissions are projected to reach 37.4 billion tonnes of CO_2 in 2024, an increase of 0.8% compared to 2023, with coal, oil, and gas contributing 41%, 32%, and 21% of these emissions, respectively [5].

Efficient energy management is essential in sectors such as agriculture and aquaculture, especially with the growing integration of renewable energy (RE). Intermittent RE generation requires accurate forecasting to balance supply and demand. Advances in artificial intelligence (AI) have enabled models that leverage historical and real-time data for improved energy demand (ED) prediction [6,7]. Attention mechanisms introduced by Vaswani et al. [8] led to the Transformer Neural Network (TNN), capable of capturing complex temporal dependencies and performing multi-step forecasting efficiently.

Despite their potential, the application of TNNs for hourly, medium-term ED forecasting remains limited, particularly in real-world scenarios involving heterogeneous end-users. Most existing studies focus on single-domain applications or assume homogeneous datasets [9,10], providing limited insight into how model performance is affected by differences in data quality, input variable selection, and consumption patterns.

To address these gaps, this study proposes a Transformer-based forecasting framework enhanced with a data-driven feature selection approach based on fuzzy logic (FL) and evaluates it across four distinct case studies: agriculture, aquaculture, port operations, and an energy community, representing diverse consumption scales, operational behaviours, and data conditions. This cross-sector analysis provides new insights into how variations in data quality, input dimensionality, and consumption dynamics influence forecasting performance, while also enabling the assessment of the model's generalization capability under heterogeneous real-world conditions.

While the results demonstrate strong predictive performance, this study also acknowledges the limitations associated with model generalization. In particular, the proposed approach depends on the availability and quality of relevant input variables, as well as sufficiently large datasets to effectively capture temporal dependencies. Furthermore, when applied to different regions or energy users with distinct consumption patterns, model retraining or adaptation may be required to maintain performance.

Overall, this work contributes a flexible and scalable Transformer-based framework for energy demand forecasting that integrates adaptive feature selection with cross-domain validation. By explicitly addressing both performance and generalization under heterogeneous conditions, the study provides a more realistic and robust foundation for the deployment of AI-based forecasting tools in energy management and planning.

2 Material and Methods

2.1 Case Studies: Aquaculture, Agriculture, Port and Energy Community

The Transformer-based model was implemented and validated in four locations across three countries in the Atlantic area, representing distinct end-users and consumption patterns: (1) Aquaculture: Island Seafoods LTD., Killybegs, Ireland; (2) Agriculture: Valle Inferior Irrigation District, Southern Spain; (3) Port: Port of Avilés, Northern Spain; (4) Energy community: Village of Marruge, Portugal. Each case study presents different input variables, temporal resolutions, and data quality levels, enabling a comparative analysis of model performance under diverse conditions.

In the aquaculture sector, Island Seafoods Ltd. (Northern Ireland) integrates renewable energy through a hydropower plant generating over 850,000 kWh annually, primarily used for wastewater treatment, with surplus exported to the grid. In agriculture, the Valle Inferior del Guadalquivir Irrigation District (Spain) operates a large-scale irrigation system supported by a 6 MWp photovoltaic plant, which supplies more than 50% of its energy demand, highlighting the relevance of renewable integration in energy-intensive systems.

The Port of Avilés (Spain) represents a complex industrial environment with variable energy demand linked to port logistics and cargo operations, while the Marruge energy community (Portugal) reflects small-scale, low-demand consumption patterns typical of rural areas with traditional agricultural practices. Together, these case studies provide a diverse testing framework in terms of scale, data characteristics, and operational variability, enabling a comprehensive assessment of the forecasting model under heterogeneous real-world conditions.

2.2 Problem Approach

Forecasting hourly ED over a one-week horizon constitutes a time series problem, where current demand depends on past values. A Transformer-based encoder–decoder model was developed using an Intel Core i7-1195G7 CPU @ 2.90 GHz and 16 GB of RAM computer, built using TensorFlow 2.10.0 in Python 3.9.0, leveraging attention mechanisms for multi-step forecasting and improved accuracy. The core focus of this work is the systematic evaluation of the model's forecasting performance across four heterogeneous, assessing its accuracy, robustness, and generalization capability under varying input configurations, consumption patterns, and data availability conditions.

2.3 Data Acquisition

One key advantage of comparing the TNN forecasting model across different end-users is the ability to evaluate variability in data sources and quality. For this study, ED and climatic data were collected from four distinct case studies. In Case (1) Fisheries, ED was recorded every 15 minutes via a telemetry system and stored online. Case (2) Irrigation (Spain) used hourly ED data reported to the Irrigation Community authorities, with climatic data from the *Red de Información Agroclimática de Andalucía* [11]. In Case (3) Ports, ED data were provided by port authorities and recorded separately for low- and medium-voltage zones at 15-minute intervals. Case (4) relied on hourly ED from a smart-metered agro-tourism facility (*Moleiro da Costa Má*, ~30 km from Marruge), adjusted to represent a small rural community's typical consumption patterns. Climatic data were obtained from the *Open-Meteo Historical Weather* API [12].

2.4 Data Preprocessing and Subset Generations

The raw data were cleaned by removing negative values and then standardized (inputs and targets) using *RobustScaler* to address scale differences. *RobustScaler* centres the data by subtracting the median and scales it according to the interquartile range (IQR), which is defined as the distance between the 25th and 75th percentiles. The dataset was split into 85% training and 15% validation. To ensure robust generalization, 100 Monte Carlo-based random splits were generated, and the one minimizing distributional difference between subsets was selected. The Monte Carlo approach, combined with the Kolmogorov-Smirnov statistic [13], guarantees that the distributions of key variables remain similar across the subsets. This approach enhances model generalisation and reduces the risk of overfitting.

2.5 Fuzzy Logic and Correlation Analysis

ED is influenced by multiple factors, including weather, temperature, humidity, energy prices, and past consumption. Identifying the most informative inputs is crucial to improve model performance and reduce computational complexity [14]. In this study, FL curves were used to select key input variables from potential candidates, supported by statistical analyses: Pearson correlation for numerical variables [15] Spearman correlation for discrete variables [16] and Point-Biserial correlation for categorical variables [17]. Comparing FL and statistical results is important, as FL can capture complex, nonlinear relationships that may be missed by traditional correlation analyses, while some variables may show low individual correlation but become predictive in combination with others.

2.6 Transformer Neural Network Architecture

The proposed forecasting model is an adaptation of the Transformer architecture introduced by Vaswani et al. [8]. The resulting TNN follows a recursive encoder–decoder structure and integrates multi-head attention mechanisms with positional encoding (PE). Since the Transformer architecture does not inherently model sequential order, positional information is explicitly incorporated into the input through sinusoidal positional encoding [8]. This encoding uniquely represents each position in the sequence while preserving relative temporal relationships. The formulation is given by:

$$PE_{(pos,2i)} = \sin\left(pos/1000^{2i/d_{model}}\right)$$
$$PE_{(pos,2i+1)} = \cos\left(pos/1000^{2i/d_{model}}\right) \tag{1}$$

where pos represents the position in the sequence, i is the dimension index, and d_{model} is the embedding dimension.

Following [8] original nomenclature, the attention mechanism is based on query (Q), key (K), and value (V) vectors obtained through linear projections of the input using learned weight matrices. The scaled dot-product attention computes similarity scores between Q and K, scales them to ensure numerical stability, and applies a SoftMax function to obtain normalized attention weights:

$$\text{Attention}(Q,K,V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{2}$$

where d_k is the dimension of the K.
The SoftMax function, defined in equation (13), transforms the attention scores into a probability distribution, allowing the output to be computed as a weighted sum of the values.

$$SoftMax = \frac{e^{zi}}{\sum_{j=1}^{m} e^{zj}}, \quad \text{for } i = 1, \dots, m \text{ and } z = (z_i, \dots, z_m) \in \mathbb{R}^m \tag{3}$$

To increase representational capacity, the model employs multi-head attention, equation (4), enabling simultaneous attention over different subspaces. Each attention head independently computes attention, and their outputs are concatenated and projected through a linear transformation by a linear layer with learnable weight W^O. This mechanism allows the model to capture diverse temporal dependencies within the input sequence.

$$\text{MultiHead}(Q, K, V) = Concat(head_1, \dots, head_h)W^O \tag{4}$$

where each attention head is defined as $head_h = Attention\left(QW_h{}^Q, KW_h{}^K, VW_h{}^V\right)$, where $QW_h{}^Q, KW_h{}^K$ and $VW_h{}^V$ denote distinct parameter matrices learned for each $head_h$.
The Transformer architecture is structured into stacked encoder and decoder blocks. The encoder processes the input sequence, combined with positional encoding, through N_{enc} identical layers. Each encoder layer consists of a multi-head self-attention sub-layer followed by a position-wise feed-forward network (FFN), with layer normalization applied after each sub-layer to improve training stability.
The FFN introduces non-linearity and enhance the model's representational capacity:

$$FFN(x) = ReLU(W_1 + b_1)W_2 + b_2 \tag{5}$$

where W_1, W_2, b_1 and b_2 are the weight and bias parameters associated with the two hidden layers. The ReLU activation function, defined as equation (6), enhances the model's ability to capture complex patterns in the data:

$$f(x) = x^+ = \max(0, x) \tag{6}$$

The decoder mirrors the encoder structure but includes an additional cross-attention sub-layer that attends to the encoder outputs. Furthermore, a masked multi-head attention mechanism (causal attention) is applied to ensure that predictions at time step i depend only on previous outputs, preventing access to future information during training.

2.7 Performance Evaluation

The performance of the proposed hourly forecasting model is evaluated using Mean Squared Error (MSE), Mean Absolute Error (MAE), and the coefficient of determination (R^2). Their mathematical definitions are provided in equations (7, 8, 9). MSE penalizes significant deviations, MAE provides an interpretable average error in the original scale, and R^2 quantifies the proportion of variance in the observed data explained by the model.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(ED_i - \widehat{\mathrm{ED}}_i\right)^2 \tag{7}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|ED_i - \widehat{ED}_i\right| \tag{8}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(ED_i - \widehat{ED}_i)^2}{\sum_{i=1}^{n}(ED_i - \overline{ED}_i)^2} \tag{9}$$

where ED_i represents the observed value, $\widehat{ED}_i$ denotes the predicted output value, $\overline{ED}_i$ represents the average of the observed values and n is the total number of samples.

3 Results and Discussion

3.1 Forecasting Model Input Variables for each End-user

The input variables of the model were analysed using FL and correlation analysis. The model was trained with 7 input variables in cases (1), (3), and (4), and with 12 input variables in case (2), as detailed in Table 1. The data analysis revealed differences among the four end-users in terms of temporal coverage, ED magnitude, data quality, and dataset size.

Beyond the mere application of a standard Transformer architecture, the novelty of this study lies in the integration of heterogeneous datasets from multiple sectors (fisheries, irrigation, port operations, and local communities), combined with a tailored input selection strategy based on feature importance and correlation analysis. This allows the model to adapt its input structure to each case study, enhancing both interpretability and performance across diverse operational contexts.
The study period spans from 1 January 2020 to 31 December 2023 for the Fisheries and Irrigation cases, while the Port dataset covers 28 April 2021 to 31 December 2023, and the Community dataset extends from 1 January 2020 to 31 December 2022. ED values range from zero to approximately 1,600 kWh for Fisheries, 4,300 kWh for the Irrigation District, 600 kWh for the Port, and 40 kWh for the Community case. Outliers represent a small fraction of the observations, ranging from 0.06% to 5.94%, corresponding to 2,084 of 34,890 records for Fisheries, 267 of 35,064 for Irrigation, 13 of 23,297 for the Port, and 372 of 26,304 for the Community dataset.

Table 1 Data analysis for the 4 end-users: Fisheries, Irrigation, Port and Community

Data Analysis	(1) Fisheries, Ireland	(2) Irrigation District, Spain	(3) Port, Spain	(4) Community, Portugal
Current ED (kWh)	✓	✓	✓	✓
ED at $t-1$ (kWh)	✓	✓	✓	✓
ED at $t-2$ (kWh)	✓	✓	✓	✓
ED at $t-3$ (kWh)	✓	✓	✓	✓
ED at $t-4$ (kWh)	✓	✗	✓	✗
ED at $t-5$ (kWh)	✓	✗	✗	✗
Boolean Rainfall (0-1)	✗	✓	✗	✗
Maximum temperature (°C)	✗	✓	✗	✗
Average temperature (°C)	✗	✓	✗	✓
Minimum temperature (°C)	✗	✓	✗	✗
Solar Radiation (MJ/m^2)	✗	✓	✗	✓
ET0 (mm)	✗	✓	✗	✗
Day vs. Night (0-1)	✗	✓	✓	✗
Hours of year	✓	✗	✗	✗
Days of year	✗	✓	✗	✓
Weekday vs. Weekend (0-1)	✗	✗	✓	✗
Selected input variables	7	12	7	7

These differences in data quality and scale directly influence the feature selection process and the resulting model behaviour. For instance, the Irrigation District dataset, with richer meteorological information and higher variability, requires a broader set of explanatory variables, while the Community dataset, characterised by stable and low-demand patterns, can be effectively modelled with fewer inputs. This highlights the importance of context-specific feature engineering when applying deep learning models to energy demand forecasting.

3.2 Forecasting Model Performance

The forecasting model architecture, considering computational capacity constraints, consisted of two encoder layers and two decoder layers, with an input transformation space of $d_{model} = 256$. The model incorporated four attention heads and a feed-forward dimension d_{ff} of 1024. Each encoder layer contained 1,580,032 parameters, while each decoder layer comprised 2,107,392 parameters. The final dense output layer included 257 parameters, resulting in a total of 3,687,681 trainable parameters. To mitigate overfitting, a dropout rate of 10% was applied throughout the network. The model was trained using the ADAM optimiser combined with the learning rate scheduling strategy proposed in [8], ensuring stable and efficient convergence during training.
Hyperparameter tuning included testing warm-up steps in the range of 2,000–8,000, a batch size of 96 constrained by computational resources, and early stopping with patience values between 5 and

20, ultimately selecting a patience of 10 epochs to balance convergence and overfitting prevention. The total number of epochs (E) used in the training of the model was 350.
Table 2 summarizes the forecasting accuracy achieved for the four end-users: Fisheries, Irrigation District, Port, and Community. The Transformer-based model exhibits consistently high performance across all cases, with R^2 values exceeding 97% and low error metrics (MSE and MAE), demonstrating strong predictive capability under heterogeneous consumption patterns and varying data characteristics.

Table 2 Performance of the forecasting model for the 4 end-users.

Performance Evaluation	(1) Fisheries, Ireland	(2) Irrigation District, Spain	(3) Port, Spain	(4) Community, Portugal
R^2 (%)	97.9	99.62	98.67	98.58
MSE	0.149	0.038	0.072	0.083
MAE (kWh)	23.04	32.36	6.64	0.44

Figure 1 provides a visual assessment of the model performance through scatter plots of predicted versus observed hourly ED for each end-user. In all cases, predicted values closely follow the 1:1 reference line, indicating strong agreement between model outputs and observed data. The dense concentration of points along the diagonal reflects low systematic bias and limited dispersion, in agreement with the high R^2 values reported in Table 2.

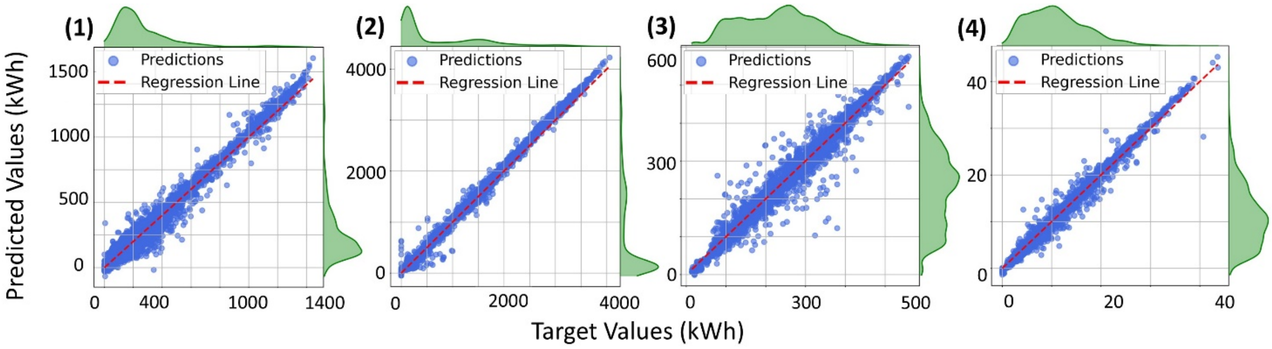

Figure 1 Comparison between predicted and observed hourly energy demand obtained from the Transformer-based model for the four end-users. Each panel corresponds to a different case study: (1) Fisheries (Ireland), (2) Irrigation District (Spain), (3) Port of Avilés (Spain), and (4) Community (Portugal).

A more detailed comparison between case studies reveals that differences in model performance are closely linked to variations in data quality, input dimensionality, and consumption dynamics. The Irrigation District case benefits from high-quality, information-rich inputs and exhibits the best performance, while the Fisheries case shows slightly higher dispersion due to larger demand variability and a higher proportion of outliers. The Port case reflects operational irregularities typical of industrial environments, and the Community case demonstrates that stable, low-variance consumption patterns facilitate highly accurate predictions even with limited input variables.
The Irrigation District case, which incorporates a larger number of input variables, shows the tightest clustering around the regression line, especially at higher demand levels, indicating that additional meteorological and temporal inputs improve performance in highly variable, energy-intensive systems. In contrast, the Fisheries, Port, and Community cases, trained with fewer input variables, still achieve robust predictive accuracy, demonstrating the Transformer model's strong generalization capability. Differences in point dispersion mainly reflect variations in ED ranges and operational characteristics among end-users.
The Port of Avilés case, characterized by moderate demand levels (up to approximately 600 kWh), shows a clear linear relationship between predicted and observed values, with slightly increased

dispersion at mid-to-high demand ranges. This behaviour reflects the operational variability of port activities, while the relatively low MAE (6.64 kWh) confirms that prediction errors remain small in absolute terms. Occasional deviations from the regression line can be attributed to isolated outliers and abrupt demand changes. The Community case, with the lowest absolute demand levels, presents a very compact distribution of points around the 1:1 line and minimal dispersion, resulting in an exceptionally low MAE (0.44 kWh). This indicates that the model is particularly effective in capturing stable consumption patterns typical of residential or small-scale community energy use.
Importantly, the consistent performance across datasets with different characteristics suggests that the model is not overfitting to a specific domain but is instead learning generalisable temporal and contextual patterns. This reinforces its applicability beyond the specific case studies analysed.
Overall, the combined quantitative metrics and visual evidence confirm the robustness and adaptability of the Transformer-based forecasting model across heterogeneous end-user profiles. The results support its suitability for multi-sector ED prediction under varying consumption scales, operational dynamics, and data availability conditions.

4 Conclusions

The results demonstrate that the proposed Transformer-based model achieves high forecasting accuracy for hourly, medium-term energy demand across all four end-users, with R^2 values ranging from 97.9% to 99.62% and consistently low error metrics. The lowest MAE values were obtained for the Community (0.44 kWh) and Port (6.64 kWh) cases, while higher absolute errors in the Fisheries and Irrigation District cases are mainly attributable to larger demand magnitudes.
Despite differences in consumption scale, data availability, and input dimensionality, the model showed robust and stable performance across all scenarios. The Irrigation District case, which included the largest number of input variables (12), achieved the highest accuracy (R^2 = 99.62%, MSE = 0.038), indicating that the inclusion of additional meteorological and temporal predictors enhances performance in highly variable, energy-intensive systems.
Furthermore, the results highlight the model's strong generalisation capability, suggesting that it can be effectively transferred to other types of energy users or geographical regions with similar temporal structures, provided that appropriate input variables are selected. This opens the possibility for broader application in energy systems with limited historical data or different operational characteristics. However, some limitations regarding generalisation capability should be acknowledged. The model performance depends on the availability and quality of input variables, particularly meteorological and temporal features in highly variable systems. In addition, the Transformer architecture requires sufficiently large and representative datasets to capture complex temporal dependencies, which may limit its applicability in data-scarce environments.
In this context, the primary contribution of the study lies in the attainment of a high predictive accuracy, in conjunction with the validation of a malleable and extensible Transformer-based framework. This framework demonstrates an aptitude for adapting to a myriad of energy consumption scenarios, while concurrently establishing the conditions under which its generalization maintains reliability. This supports its viability for energy management and operational planning across diverse sectors.

Acknowledgments

This research was supported by HY4RES - Hybrid Solutions for Renewable Energy Systems: achieving net-zero Atlantic area energy consumers & communities), grant number EAPA_0001/2022, from Interreg Atlantic Area Programme and co-funded by the European Union.

Reference

[1] M.J.B. Kabeyi, O.A. Olanrewaju, Sustainable Energy Transition for Renewable and Low Carbon Grid Electricity Generation and Supply, Front. Energy Res. 9 (2022). https://doi.org/10.3389/fenrg.2021.743114.

[2] J. Keirstead, M. Jennings, A. Sivakumar, A review of urban energy system models: Approaches, challenges and opportunities, Renewable and Sustainable Energy Reviews 16 (2012) 3847–3866. https://doi.org/10.1016/J.RSER.2012.02.047.
[3] D. Poponi, T. Bryant, K. Burnard, P. Cazzola, J. Dulac, A.F. Pales, J. Husar, P. Janoska, E.R. Masanet, L. Munuera, Energy technology perspectives 2016: towards sustainable urban energy systems, International Energy Agency, 2016.
[4] A.K. Akella, R.P. Saini, M.P. Sharma, Social, economical and environmental impacts of renewable energy systems, Renew. Energy 34 (2009) 390–396. https://doi.org/10.1016/J.RENENE.2008.05.002.
[5] G.C. Budget, Fossil fuel CO emissions increase again in 2024, (2024). https://globalcarbonbudget.org/fossil-fuel-co2-emissions-increase-again-in-2024/ (accessed September 2, 2025).
[6] R. González Perea, E. Camacho Poyato, P. Montesinos, J.A. Rodríguez Díaz, Prediction of irrigation event occurrence at farm level using optimal decision trees, Comput. Electron. Agric. 157 (2019) 173–180. https://doi.org/10.1016/J.COMPAG.2018.12.043.
[7] A. Barragán-Escandón, J. Terrados-Cepeda, E. Zalamea-León, The Role of Renewable Energy in the Promotion of Circular Urban Metabolism, Sustainability 9 (2017). https://doi.org/10.3390/su9122341.
[8] A. Vaswani, G. Brain, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, L. Kaiser, I. Polosukhin, Attention is all you need, in: Adv. Neural Inf. Process. Syst., 2017. https://doi.org/https://doi.org/10.48550/arXiv.1706.03762.
[9] M.A. Ikegawa Bernabé, R.G. Perea, J.A. Rodríguez Díaz, J.G. Morillo, Transformer-based artificial intelligence for forecasting energy demand in irrigation districts, Comput. Electron. Agric. 246 (2026). https://doi.org/10.1016/j.compag.2026.111611.
[10] R. González Perea, E. Camacho Poyato, J.A. Rodríguez Díaz, Attention is all water need: Multistep time series irrigation water demand forecasting in irrigation disctrics, Comput. Electron. Agric. 218 (2024). https://doi.org/10.1016/j.compag.2024.108723.
[11] Red de Información Agroclimática de Andalucía (RIA) | Instituto de Investigación y Formación Agraria y Pesquera (IFAPA), (n.d.). https://www.juntadeandalucia.es/agriculturaypesca/ifapa/riaweb/web/ (accessed March 6, 2025).
[12] P. Zippenfenig, [dataset] Open-Meteo.com Weather API, (2024). https://doi.org/10.5281/zenodo.14582479.
[13] G.W. Corder, D.I. Foreman, Nonparametric Statistics for Non-Statisticians: A Step-by-Step Approach, in: Nonparametric Statistics for Non-Statisticians, 2009: pp. 12–37. https://doi.org/https://doi.org/10.1002/9781118165881.ch2.
[14] Y. Lin, G.A. Cunningham, S. V. Coggeshall, Input variable identification — fuzzy curves and fuzzy surfaces, Fuzzy Sets Syst. 82 (1996) 65–71. https://doi.org/10.1016/0165-0114(95)00223-5.
[15] P. Dutilleul, J.D. Stockwell, D. Frigon, P. Legendre, The Mantel test versus Pearson's correlation analysis: Assessment of the differences for biological and environmental studies, J. Agric. Biol. Environ. Stat. 5 (2000) 131–150. https://doi.org/10.2307/1400528.
[16] C. Xiao, J. Ye, R.M. Esteves, C. Rong, Using Spearman's correlation coefficients for exploratory data analysis on big dataset, Concurr. Comput. 28 (2016) 3866–3878. https://doi.org/10.1002/cpe.3745.
[17] E.E. Cureton, Rank-biserial correlation, Psychometrika 21 (1956) 287–290. https://doi.org/10.1007/BF02289138.