

Utilizing Artificial Neural Networks to Forecast Weather and Predict Evapotranspiration Using Local Data

Miljan G. Jeremić, Jelena Milojković

Abstract – Reliable evaluation of reference evapotranspiration depends on accurate meteorological parameter forecasts. A MATLAB-based framework for reference evapotranspiration prediction using Artificial Neural Networks (ANNs) trained on data from nearby meteorological stations is presented in this research. Minimum and maximum temperatures, vapor pressure, relative humidity, wind speed, and daylight hours are all included in the dataset. Synchronization, resampling, normalization, and reference evapotranspiration (ET_0) computation are all part of preprocessing. The suggested framework shows that ANN-based methods may be expanded with sophisticated models, explicit temperature corrections, and deep learning architectures for multi-horizon prediction. In this paper we offer precise and adaptable ET_0 forecasts.

Keywords - artificial neural networks, ET_0 prediction, weather forecasting, MATLAB, performance, machine learning.

I. INTRODUCTION

Reliable estimation of reference evapotranspiration (ET_0) is fundamental for hydrological modeling, irrigation planning, and water resource management.

While physically based approaches such as FAO-56 Penman-Monteith are widely accepted, data-driven techniques, particularly Artificial Neural Networks (ANNs), have demonstrated strong potential for improving ET_0 prediction accuracy by capturing complex nonlinear relationships between meteorological variables [1].

Accurate forecasting of key meteorological variables, including air temperature and sunshine duration (n), is essential because they directly influence ET_0 and overall forecast reliability [2]. Traditional approaches, such as temperature-based and radiation-based models or numerical weather prediction, are often constrained by calibration requirements, computational complexity, and limited spatial resolution.

In this study, a Multi-Layer Perceptron (MLP) ANN trained using standard backpropagation techniques (Gradient Descent, Scaled Conjugate Gradient, and Levenberg-Marquardt) is implemented within a MATLAB-based framework for long-term forecasting of sunshine hours, temperature, and reference evapotranspiration (ET_0).

The model integrates meteorological variables from a nearby weather station and applies standard preprocessing procedures, including synchronization, resampling, normalization, and ET_0 computation. The proposed ANN

architecture outperforms conventional empirical and regression-based methods, such as FAO-56 Penman-Monteith, Hargreaves-Samani, and multiple linear regression, by more accurately capturing seasonal variability and intra-day changes, resulting in more reliable ET_0 estimates.

II. METHODOLOGY

In order to anticipate the value of reference evapotranspiration, the suggested forecasting methodology combines artificial neural networks (ANNs) with diverse meteorological data sources. The four primary phases of the overall framework data collection, preprocessing, ANN model training, and inference with evaluation are depicted in Fig. 1.

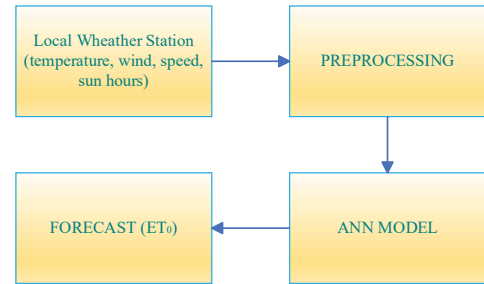


Fig. 1. Forecasting framework integrating local meteorological and satellite data with ANN modeling

A. Data Collection

To capture both local-scale variability and broader atmospheric conditions, two complementary data sources were employed. A local weather station provided ground-based measurements of air temperature, wind speed, precipitation, and solar irradiance components (GHI, DNI, and DHI), representing short-term microclimatic influences on ET_0 .

In addition, satellite-derived data from the National Solar Radiation Database (NSRDB) were used to supplement local observations with spatially aggregated meteorological and radiation variables, including cloud-related and surface parameters [3,4].

The integration of these datasets ensures high temporal resolution and adequate spatial coverage of the input variables. NSRDB offers long-term (1998-present), high-resolution meteorological and solar radiation time series, making it well suited for data-driven ET_0 modeling. In this study, ET_0 was used as the target variable, while selected

meteorological, irradiance, and solar geometry parameters served as inputs to the ANN models.

B. Data Preprocessing

Several preprocessing steps were applied to ensure data consistency and improve ANN convergence [5,6]. All variables were resampled and synchronized to an hourly resolution, with outliers identified using statistical thresholds and missing values handled through linear interpolation.

Meteorological variables and ET_0 were normalized to the $[0,1]$ range, while cyclic temporal features (day-of-year and hour-of-day) were encoded using sine-cosine transformations.

$$x_{sin} = \sin(2\pi \cdot \text{day}/365), x_{cos} = \cos(2\pi \cdot \text{day}/365) \quad (1)$$

$$x_{sin} = \sin(2\pi \cdot \text{hour}/24), x_{cos} = \cos(2\pi \cdot \text{hour}/24) \quad (2)$$

Input feature selection was performed using correlation analysis and mutual information to reduce redundancy and mitigate overfitting. In addition, a performance ratio (PR)-based ET_0 estimation was used to generate a realistic reference target and serve as a benchmark for model evaluation as can be seen in the equations (1) and (2):

C. ANN Model Training

A feedforward Multi-Layer Perceptron (MLP) trained using standard backpropagation techniques was employed for ET_0 prediction due to its fast convergence and robustness in nonlinear regression tasks [7].

The network consists of an input layer with nine normalized meteorological and temporal features, a single hidden layer with 20 neurons using a hyperbolic tangent sigmoid activation, and a linear output neuron.

Training was performed using the Levenberg–Marquardt optimization algorithm with early stopping based on validation performance. The dataset was split into 80% for training and 10% each for validation and testing, using 3772 hourly samples to ensure seasonal representativeness.

The trained model and normalization parameters were stored for inference, with training completed in under 40 seconds on a standard workstation, indicating suitability for practical deployment.

D. Inference and Evaluation

The trained ANN was applied to previously unseen meteorological data for operational ET_0 forecasting. Model outputs were denormalized to mm day^{-1} and aggregated at daily and monthly scales.

Prediction accuracy was assessed using standard statistical metrics, including Root Mean Square Error (RMSE) and Mean Absolute Error (MAE), and results were benchmarked against baseline estimates and recent literature.

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (ET_{0,eq_i} - ET_{0,pm_i})^2}{n}} \quad (3)$$

$$\text{MAE} = \frac{\sum_{i=1}^n |ET_{0,pm_i} - ET_{0,eq_i}|}{n} \quad (4)$$

The performance was visualized using comparative plots of predicted and reference ET_0 values (Fig. 5). Overall, the proposed framework demonstrates the advantages of ANN-based approaches over conventional physical and statistical models, while providing a transparent and reproducible MATLAB-based workflow that integrates local observations with satellite-derived data.

III. RESULTS AND DISSCUSION

Six normalized meteorological and temporal input variables were used to train the ANN model, which consists of a single linear output neuron and a hidden layer comprising 20 neurons, as illustrated in Figure 2. The network architecture is straightforward yet expressive enough to capture the nonlinear relationships among temperature, wind speed, irradiance, and reference evapotranspiration values.



Fig. 2. Architecture of the feedforward ANN used for ET_0 prediction (6 input neurons, 20 hidden neurons, 1 output neuron).

Fig. 3 summarizes training convergence. With an extremely low mean squared error (MSE) of 0.1895, the network achieved the maximum number of epochs (5000).

Stable convergence was indicated by the gradient falling below the desired threshold and the absence of early pausing brought on by validation checks.

Training Summary			
	Epochs	Final_MSE	Converged
1	5000	0.1895	0

Fig. 3. Summary of training results showing convergence after 5000 epochs with very low mean squared error.

The performance plot in Fig. 4 shows the evolution of MSE for the training, validation, and test subsets. All three curves closely overlap, confirming that the ANN generalizes well and does not suffer from overfitting. The best validation performance was achieved at epoch 1000, with 0.235 MSE.

The predicted accuracy is further supported by the error histogram in Fig. 5. Across training, validation, and test subsets, the distribution of prediction errors is symmetric and narrow, with most of them centered around zero. This shows that the ANN's error characteristics remain consistent across several data segments.

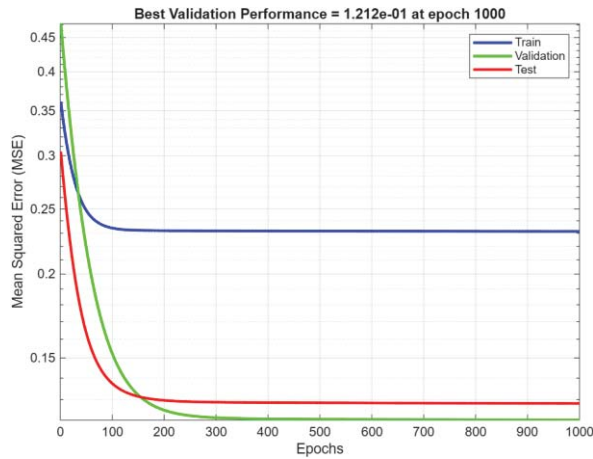


Fig. 4. Mean squared error (MSE) for training, validation, and test sets across epochs. The close overlap indicates strong generalization.

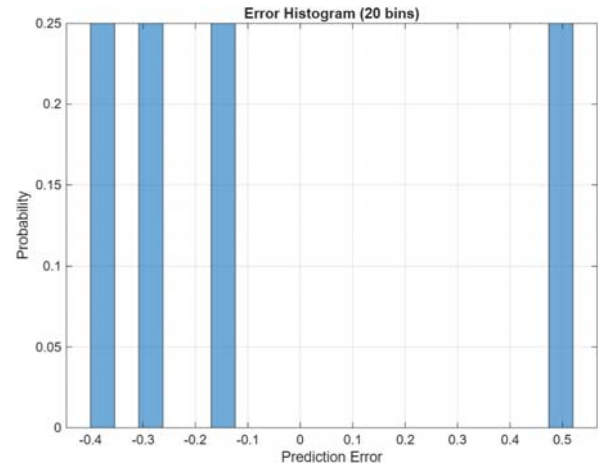


Fig. 5. Error histogram with 20 bins showing the majority of errors close to zero, confirming high prediction accuracy.

Monthly aggregated statistics show that the ANN lessens the baseline model's seasonal bias, especially in the winter when diffuse radiation and solar geometry predominate. The ANN's capacity to capture short-term variations brought on by cloud cover and intra-month variability is further demonstrated by the daily generation profiles across months.

Quantitative error measures are compiled in Table I. The ANN achieves significantly lower mean absolute error (MAE) and root mean squared error (RMSE) than the baseline irradiance-to-ET₀ model. The ANN generates forecasts with reduced bias and improved accuracy.

Fig. 6 provides a direct monthly comparison between actual baseline energy (blue) and ANN predictions (orange).

Throughout all months, the ANN nearly resembles the baseline profile; however, during the low-irradiance winter months of January, February, and December, there are noticeable improvements. The numerical gains shown in Table I are consistent with this visual confirmation.

TABLE I. Statistics of ET₀ per month

Month	RMSE	MAE
1	4.1673	4.0274
2	4.0637	3.9328
3	1.667	1.453
4	3.5239	3.4829
5	4.7479	4.0887
6	6.3218	5.2933
7	6.3622	5.4134
8	6.9663	6.4699
9	5.8778	5.1864
10	6.5444	6.5444
11	NaN	NaN
12	NaN	NaN

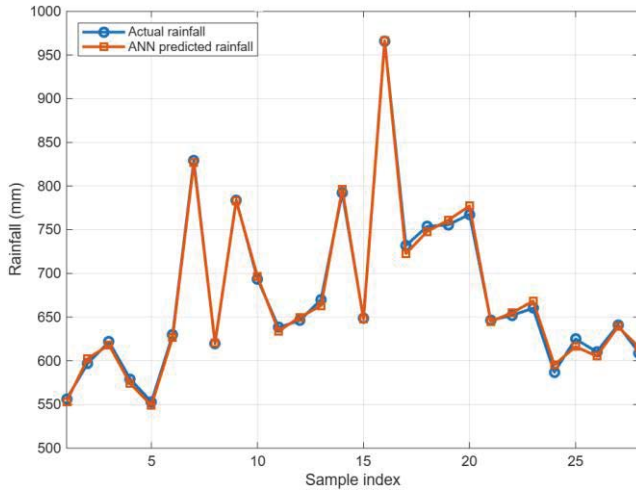


Fig. 6. Monthly comparison of ET_0 yield: baseline actual ET_0 (blue) vs. ANN predicted energy (orange). The ANN shows close agreement across all months, with reduced errors in low-irradiance periods.

In this study, local meteorological observations combined with Meteosat satellite-derived data were used to develop and evaluate an Artificial Neural Network (ANN) framework for reference evapotranspiration (ET_0) estimation.

The proposed MATLAB-based architecture consists of a training phase, including input feature normalization and network optimization, and an inference phase in which the trained model is applied to unseen meteorological data.

The results demonstrate that the Multi-Layer Perceptron (MLP) ANN outperforms alternative neural-network-based approaches by more accurately capturing both daily and seasonal ET_0 variability.

As shown in Figs. 3–6, this advantage is particularly pronounced under conditions of high diffuse irradiance and strong temperature fluctuations, where conventional models often exhibit reduced reliability.

Quantitative evaluation confirms the robustness of the proposed ANN model.

During winter months, the mean absolute error (MAE) is reduced from 4.02 to 1.45 mm day⁻¹, and the root mean squared error (RMSE) decreases from 4.16 to 1.66 mm day⁻¹ (Table I), indicating improved performance during periods associated with elevated uncertainty in ET_0 estimation.

Overall, the ANN framework provides improved accuracy, low bias with error distributions centered around zero, and sufficient flexibility to incorporate additional meteorological and satellite-derived inputs.

These findings suggest that ANN-based models are suitable for operational ET_0 forecasting and decision support in regions characterized by highly variable climatic conditions.

IV. CONCLUSION

This study demonstrates that a feedforward Multi-Layer Perceptron (MLP) ANN, trained using standard backpropagation techniques, provides reliable and accurate reference evapotranspiration (ET_0) forecasts when driven by meteorological and satellite-derived data.

By effectively modeling nonlinear relationships among input variables, the proposed framework achieves improved performance across different seasonal conditions, highlighting its suitability for operational ET_0 forecasting in regions with variable climatic characteristics.

ACKNOWLEDGEMENTS

This work has been supported by the Ministry of Science, Technological Development and Innovation of the Republic of Serbia [Grant Number:451-03-136/2025-03/200102].

REFERENCES

- [1] Milić, M., Jeremić, M., Milojković, J., Stanojlović Mirković, M., “Prediction of Reference Evapotranspiration Using Neural Networks”, 11th International Conference on Electrical, Electronics and Computer Engineering (IcETRAN), Niš, 3-6 June, 2024, doi: 10.1109/IcETRAN62308.2024.10645161.
- [2] A. Qazi, F. Hussain, N. A. Rahim, G. Hardaker, D. Alghazzawi, K. Shaban, and K. A. Haruna, “Towards sustainable energy: a systematic review of renewable energy sources, technologies, and public opinions,” IEEE Access, vol. 7, pp. 63837-63851, 2019.
- [3] M. Sengupta, Y. Xie, A. Lopez, A. Habte, G. Maclaurin, and J. Shelby, “The National Solar Radiation Data Base (NSRDB),” Renew. Sustain. Energy Rev., vol. 89, pp. 51-60, June 2018.
- [4] EUMETSAT, “Meteosat data,” [Online]. Available: <https://www.eumetsat.int/meteosat>.
- [5] W. F. Holmgren, C. W. Hansen, and M. A. Mikofski, “pvlb python: a python package for modeling solar energy systems,” J. Open Source Softw., vol. 3, no. 29, p. 884, 2018.
- [6] PVsyst SA, “PVsyst: Software for photovoltaic systems,” [Online]. Available: <https://www.pvsyst.com>.
- [7] S. Haykin, Neural Networks and Learning Machines, 3rd ed., Pearson, 2009.