Artificial Intelligence-Based Reference Evapotranspiration Modelling with Minimum Climatic Parameters

K. Chandrasekhar Reddy

Abstract Artificial neural networks (ANN), an artificial intelligence-based technology, have yielded numerous favourable water resource and hydrology simulation results. The investigation goal is to find the efficient ANN architecture for estimating reference evapotranspiration with the fewest climate variables. The climatic parameters, namely wind speed (W), relative humidity (RH), air temperature (T), and sunshine hours (S) were used to estimate monthly reference evapotranspiration (ET_0). Partial correlation and multiple linear correlations between the climatic parameters and FAO-56 Penman–Monteith reference evapotranspiration ($PMET_0$) were carried out in order to determine the most influential factor by eliminating one factor at a time. The influence of parameters T, S, W, and RH was observed from highest to lowest, respectively. Therefore, the best ANN models to estimate artificial neural network reference evapotranspiration ($ANNET_0$) were developed using climatic parameters as inputs and eliminating one lowest influencing parameter each time. Training of model was done with a portion of data, and the remaining was used for testing the model. Performance indices have been used to assess the ability of the model by correlating the $PMET_0$ and $ANNET_0$. The viability of the generated models was verified by using the numerical indicators (i.e. efficiency coefficient, coefficient of determination, and root mean square error). ANN (1-5-1), ANN (2-5-1), ANN (3-4-1), ANN (4-3-1) with (T), (T, S), (T, S, W), and (T, S, W, RH) as inputs, respectively, were shown to have 89.58%, 94.36%, 95.20%, and 99.44% during the testing period, respectively. Thus, in the area of study and other locations with similar climatic conditions, these ANN models can assess monthly ET_0 with adequate accuracy.

Keywords Reference evapotranspiration · Performance indices · FAO-56 Penman–Monteith method · Multiple and partial correlation coefficients · Artificial neural networks

---

K. Chandrasekhar Reddy (✉)
Civil Engineering, Siddharth Institute of Engineering & Technology, Puttur, Andhra Pradesh 517583, India
e-mail: kamasani.kcr@gmail.com

P. V. Timbadiya et al. (eds.), *Geospatial and Soft Computing Techniques*, Lecture Notes in Civil Engineering 339, https://doi.org/10.1007/978-981-99-1901-7_45

1 Introduction

The optimization of irrigation water use is the need of the hour as the water demand is increasing rapidly. Determination of amount evapotranspiration is one of the feasible methods to decrease irrigation water consumption. The Lysimeter is perhaps the only accurate, direct field measuring, and appropriate instrument for evaluating evapotranspiration. The instrument, however, has a high initial cost and requires complicated maintenance. As a result, many semi-experimental and experimental models to predict reference evapotranspiration (ET_0) using climate data have been established by many investigators. ET_0 can be estimated by Hargreaves, Jensen-Haise, Blaney-Criddle, Makkink, Pan Evaporation, Priestley-Taylor, Radiation, Christiansen, FAO-56 Penman–Monteith (PM), Modified Penman, etc. However, in all weather conditions, the PM technique can produce results comparable to Lysimeter readings, therefore considered as the standard method to estimate ET_0 [1–5]. In FAO, Penman–Monteith method for estimating ET_0 requires data of climate parameters, namely temperature (highest and lowest), relative humidity (vapour pressure) (highest and lowest), wind speed, and sunshine (radiation) along with site location. Many methods of determining ET_0 don't adequately describe the nonlinear effects intrinsic in the ET_0 function. When the correlation among the independent and dependent parameters in literary modelling is nonlinear, artificial neural networks (ANNs) represent intricate nonlinear procedures relevant to ET_0 assessment for improved interpretation. Technology based on artificial intelligence (AI) is feasible [6] in contrast to others statistical interpretation approaches; ANN performed better even the data had errors and noise.

The present investigation is interfered by establishing efficient ANN architecture for estimating monthly ET_0 with the fewest climate variables.

2 Materials and Methodology

2.1 Research Area

The researcher acquired meteorological data for Tirupati, a well-known city of Chittoor, AP, India, with altitude 161.0 m and coordinate values of 13° 37' N and 79° 25' E (research region), from the India Meteorological Department. Climate data from 1992 to 2001 was utilized to find an efficient ANN model. Part of the data (1992–1998) was utilized to train ANN model and the rest to test it.

Table 1 Particulars of Penman–Monteith (PM) technique suggested by FAO-56

Primary reference	PM equation	Input data
Allen et al. [2]	$\mathrm{ET}_0 = \frac{0.408\Delta(R_n-G)+\gamma\frac{900}{T_{\mathrm{mean}}+273}u_2(e_\mathrm{s}-e_\mathrm{a})}{\Delta+\gamma(1+0.34u_2)}$	$T_{\max}$, $T_{\min}$, $\mathrm{RH}_{\min}$, $\mathrm{RH}_{\max}$, n, u_2

2.2 ET₀ Assessment Techniques

2.2.1 Penman–Monteith (PM) Technique Suggested by FAO-56

The Penman method, which combines an intern mass transfer and energy balance approach for the reference crop, inspired the PM method. An anticipated grass crop having 0.12 m height, 70 s/m surface resistivity, and 0.23 albedo is considered as a reference crop. It is equivalent to evaporation from the thickest green grass of equal height, which thrives and is adequately irrigated. ASCE and European studies resulted in relatively consistent and accurate performance in arid and humid climates [7]. As a result, the technique is regarded as the most accurate, and its necessary particulars are listed in Table 1. The PM technique used is suggested by FAO-56.

PM equation is a typical approach to estimate ET_0 as it represents physiological and physical characteristics that impact evapotranspiration. Climate variables like relative humidity, sunshine, extreme air temperatures, and wind velocity at 2 m elevations are considered inherent data for utilizing this technique.

2.2.2 Artificial Neural Networks (ANNs) Modelling

For the present investigation, a standard multilayer feed-forward ANN with the logistic sigmoid function was used, considering the momentum factor with a constant value of 0.9 and 0.1 for the learning rate. The input data was normalized between 0.1and 0.9 to avoid saturation. During calibration, error backpropagation, an iterative nonlinear optimization technique based on the gradient descent search method [8], was used. The standardization set to reduce error and the validation set to ensure appropriate neural network training was used not to overtrain the neural networks. The performance of the model was tracked throughout every repetition to avoid over-learning and thus improve it. The hit-end-error technique was used to create a network with the lowest mean squared error by eradicating too few or too many neurons, resulting in an ideal network. When modelled in MATLAB, it has augmented training time and discovered the ideal neural network.

2.3 Multiple Linear Correlation Analysis

The relationship among independent and dependent parameters in this investigation is calculated by the following coefficients.

2.3.1 Multiple Correlation Coefficient (R)

It is a metric for the linear relation among numerous parameters, consisting of one dependent parameter, y, (ET_0), and mutually independent variables, x_i (RH, T, S, and W). It is calculated as the ratio of the standard deviation of calculated value (s_{e1}) to the standard deviation of experimental measurements (s_1). Considering S_1 as the standard deviation of residuals, the coefficient may be represented as

$$R = \frac{s_{e1}}{s_1} = \sqrt{1 - \frac{S_1^2}{s_1^2}} \tag{1}$$

2.3.2 *Coefficient of Determination (D_1)*

It is defined as squared multiple correlation coefficient (R_1), i.e. $D_1 = R_1^2$. The coefficient of multiple non-determination $(1 - D_1) = 1 - R_1^2$.

The D_1 specifies the dependent variable's portion of the variance. Meanwhile, $(1 - D_1) = 1 - R_1^2$ denotes the proportion of variance not explained by the multiple linear correlation of the variable x_1 over $x_2, x_3, \ldots, x_m$.

2.3.3 *Partial Correlation Coefficient ($r_{1\text{-}i}$)*

It measures the relationship between independent variable x_i (T, RH, S, or W) and dependent variable x_1 (ET_0) after eliminating the linear influence of other parameters on them.

R_1 is the multiple correlation coefficient between x_1 and x_i. R_{1-i} is also a multiple correlation coefficient between x_1 and x_i, after omitting the chosen independent variable x_i. It is then calculated from

$$r_{1-i} = \sqrt{\frac{\left(1 - R_{1-i}^2\right) - \left(1 - R_1^2\right)}{\left(1 - R_{1-i}^2\right)}} = \sqrt{1 - \frac{\left(1 - R_1^2\right)}{\left(1 - R_{1-i}^2\right)}} \tag{2}$$

2.4 Performance Metrics

The following metrics were utilized to assess the capability of the generated models.

2.4.1 Coefficient of Determination (D)

It is equal to R^2, i.e. $D = R^2$, where R—correlation coefficient and it can be represented as

$$R = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\left[\sum_{i=1}^{n}(x_i - \overline{x})^2 \sum_{i=1}^{n}(y_i - \overline{y})\right]^{1/2}} \times 100 \tag{3}$$

Here, x_i = $\mathrm{PMET_0}$ values and $\overline{x}$ = mean of x_i

y_i = $\mathrm{ANNET_0}$ values and $\overline{y}$ = mean of y_i.

$i = 1$ to n.

n = number of data values.

The D value indicates the extent of relationship between $\mathrm{PMET_0}$ and $\mathrm{ANNET_0}$ values.

2.4.2 Root Mean Square Error (RMSE)

It is used to find the residual error between $\mathrm{PMET_0}$ and $\mathrm{ANNET_0}$ values. It is stated as [9].

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - y_i)^2}{n}} \tag{4}$$

2.4.3 Efficiency Coefficient (EC)

Here, the efficiency coefficient (Nash and Sutcliffe 1970) is utilized to evaluate the skill of the developed ANN architectures. EC is more reliable alternative than the RMSE indicator when the training and testing data periods have different lengths [10]. EC is stated as

$$\mathrm{EC} = \left(1 - \frac{\sum_{i=1}^{n}(x_i - y_i)^2}{\sum_{i=1}^{n}(x_i - \overline{x})^2}\right) \times 100 \tag{5}$$

Table 2 Coefficient of multiple and partial correlations

Multiple correlation coefficient					Partial correlation coefficient			
Independent variable omitted								
–	T	S	W	RH	T	S	W	RH
0.9967	0.9748	0.9835	0.9922	0.9958	0.9314	0.8937	0.7589	0.4625

If EC value more than 90% shows the model performing good, value 80–90% indicates satisfactory model, and 60–80% value intimates a not acceptable model.

3 Results and Discussions

Partial and multiple correlations coefficients are used to recognize the extremely influencing weather parameters. The results of the investigation among the climate parameters (RH, S, T, and W) and $PMET_0$ in the research region are given in Table 2. The ANN algorithm is functioned with the input factors in scope and the changing number of nodes in the hidden layer to attain optimal performance and efficiency, as given in Table 3. As a result, the scatter and comparison graphs obtained are illustrated in Figs. 1 and 2.

The coefficients of multiple and partial correlations given in Table 2 indicate that the influence in the order of lowest to highest is relative humidity, wind speed, sunshine hours, and temperature on ET_0 at the study area. This is due to the area located in the semi-arid zone, which is characterized mainly by radiation and high temperature. The high EC and low RMSE values are performing adequately for the area, and Table 3 is indicating the same.

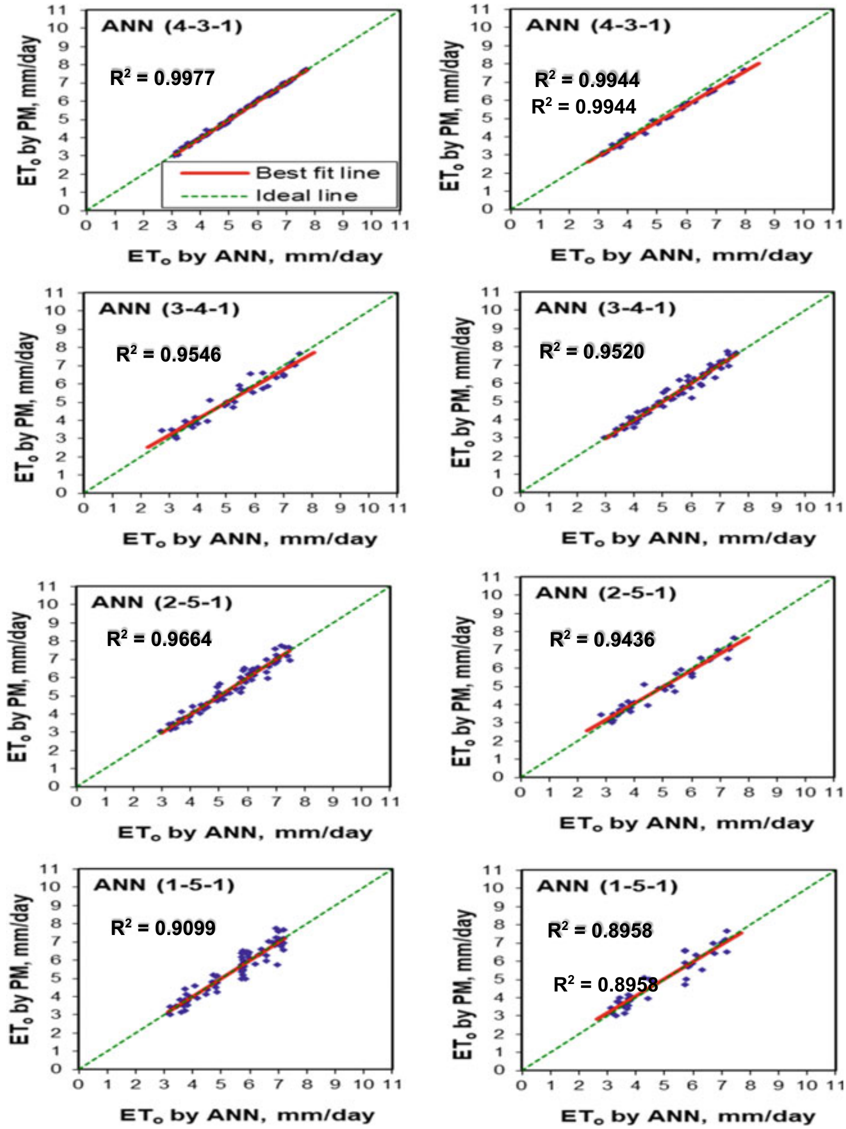
Figures 1 and 2 show scatter and comparison graphs that show the same results. The graph drawn among ET_0 obtained from the PM method as the ordinate and ET_0 attained from ANN as the abscissa yields a straight line with a unit slope and a zero intercept, highlighting the fact that ANN outcomes are fairly significant compared to PM method results. The author suggests that the developed model for predicting monthly ET_0 may be implemented for the experimental area under consideration based on the results of the ANN model data because it yields better accuracy.

4 Conclusions and Recommendations

The ANN models were established to forecast monthly ET_0 using climate variables affecting the region chosen for the current investigation as inputs. During the testing period, ANN (1-5-1), ANN (2-5-1), ANN (3-4-1), ANN (4-3-1) with (T), (T, S), (T, S, W), and, (T, S, W, RH) as inputs, respectively, were shown to have

Table 3 Performance metrics of ANN models

Optimal ANN	Input parameters	Slope (m)		Intercept		R^2		RMSE		
		Training	Testing	Training	Testing	Training	Testing	Training	Testing	Training
4-3-1	T, S, W, RH	1.0001	0.9202	−0.0004	0.2184	0.9977	0.9944	0.07	0.10	99.7
3-4-1	T, S, W	0.9990	0.9077	0.0078	0.4433	0.9546	0.9520	0.29	0.29	95.4
2-5-1	T, S	1.0019	0.8921	−0.0097	0.5184	0.9664	0.9436	0.25	0.32	96.6
1-5-1	T	1.0000	0.9421	0.0001	0.3333	0.9099	0.8958	0.41	0.43	90.9



a) During training the model b) During testing the model

Fig. 1 Scatter plots of average monthly $PMET_0$ versus $ANNET_0$

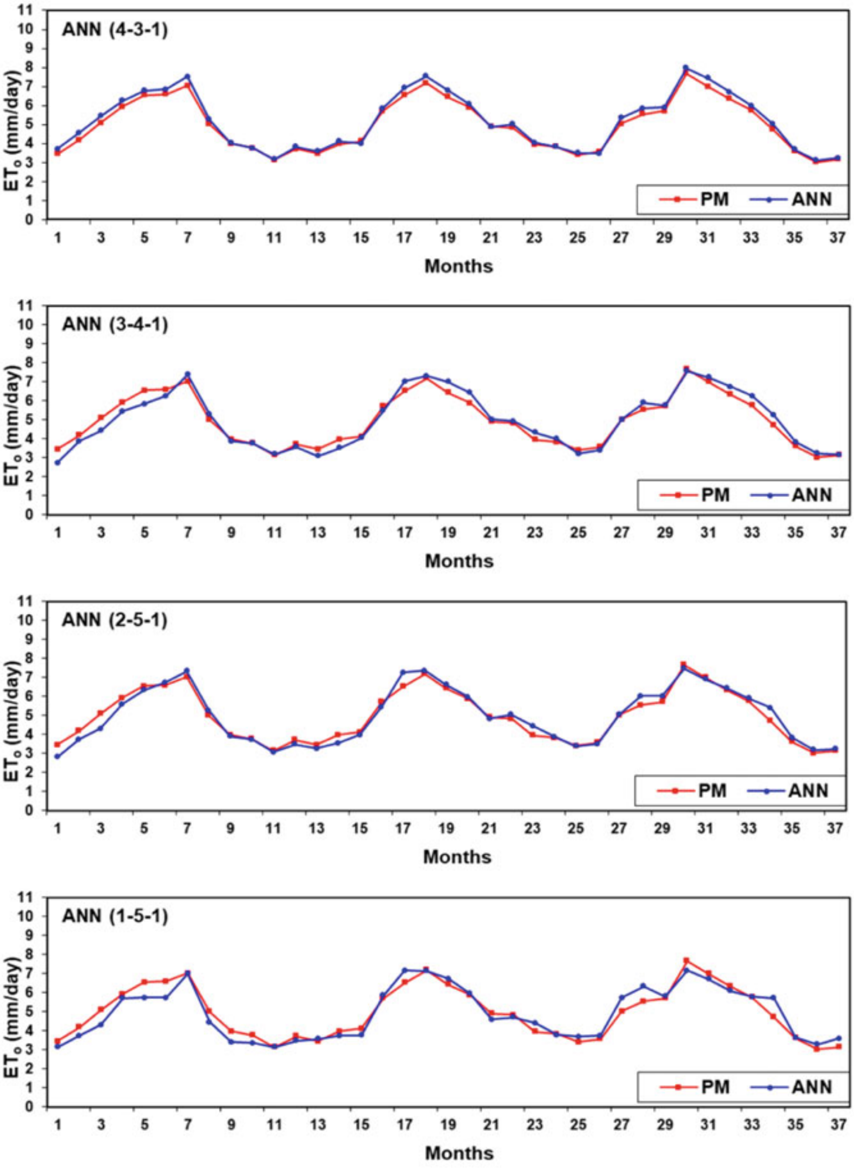


Fig. 2 Comparison graphs of mean monthly $PMET_0$ vs $ANNET_0$ during the testing period

89.58%, 94.36%, 95.20%, and 99.44%, respectively. However, the performance of ANN (4-3-1) model with inputs (T, S, W, RH) was better as compared to other.

These ANN models are suggested for forecasting monthly ET_0 with reasonable accuracy in the study location and other locations having alike climatic conditions.

Acknowledgements The corresponding author expresses gratitude to the Civil Engineering Department Staff and the college administration at SIETK, AP, India, for their suggestions and providing the necessary facilities throughout the research exploration.

References

1. Hargreaves GH, Samani ZA (1985) Reference crop evapotranspiration from temperature. Appl Eng Agric 1(2):96–99
2. Allen RG, Pereira LS, Raes D, Smith M (1998) Guidelines for computing crop water requirements-FAO irrigation and drainage paper 56, FAO—Food and Agriculture Organisation of the United Nations. Geophysics 156:178
3. Feng Y, Cui N, Zhao L, Hu X, Gong D (2016) Comparison of ELM, GANN, WNN and empirical models for estimating reference evapotranspiration in humid region of Southwest China. J Hydrol 536:376–383
4. Mallikarjuna P, Jyothy SA, Murthy DS, Reddy KC (2014) Performance of recalibrated equations for the estimation of daily reference evapotranspiration. Water Resour Manage 28(13):4513–4535
5. Nema MK, Khare D, Chandniha SK (2017) Application of artificial intelligence to estimate the reference evapotranspiration in sub-humid Doon valley. Appl Water Sci 7(7):3903–3910
6. Koch J, Berger H, Henriksen HJ, Sonnenborg TO (2019) Modelling of the shallow water table at high spatial resolution using random forests. Hydrol Earth Syst Sci 23(11):4603–4619
7. Govindaraju RS (2000) Artificial neural networks in hydrology. I: preliminary concepts. J Hydrol Eng 5(2):115–123
8. Pandey PK, Pandey V (2016) Evaluation of temperature-based Penman-Monteith (TPM) model under the humid environment. Model Earth Syst Environ 2(3):1
9. Laaboudi A (2020) Slama A (2000) Using Neuro-fuzzy and linear models to estimate reference Evapotranspiration in South region of Algeria (A comparative study). Italian J Agrometeorol 2:55–64
10. Liang GC, O'Connor KM, Kachroo RK (1994) A multiple-input single-output variable gain factor model. J Hydrol 155(1–2):185–198