BIO ACCENT

International Journal of Environment & Agricultural Sciences
An Open Access Journal

Int J Environ Agric Sci
Research Article
Volume 2

Prediction of reference evapotranspiration in South Africa using an artificial intelligence approach

Akinola Ikudayisi[1]; Andre Calitz[2]; Mahmoud Nasr[3]; Samuel Abejide[1]*

[1]*Department of Civil Engineering, Walter Sisulu University, East London, 5200 South Africa.*

[2]*Department of Computing Sciences, Nelson Mandela University, Gqeberha, South Africa.*

[3]*Sanitary Engineering Department, Faculty of Engineering, Alexandria University, Egypt.*

***Corresponding Author: Samuel Abejide**

Department of Civil Engineering, Walter Sisulu University, East London, 5200 South Africa.

Email: sabejide@wsu.ac.za

Received: Aug 12, 2022

Accepted: Sep 22, 2022

Published: Sep 28, 2022

Keywords: Artificial neural networks; Feed-forward back propagation; Irrigation; Evapotranspiration; Vaalharts.

Abstract

Correct estimation of reference evapotranspiration (ETo) is vital to the sustainability of water resources management practices around the world. A key factor in irrigation scheduling and development. This study adopted a Feed-forward back-propagation Artificial Neural Network (BP-ANN) model to predict ETo from temperature, rainfall, wind speed, relative humidity and determined the effect of different numbers of neurons and hidden layers on the predictive ability of BP-ANN models. Eight different BP-ANN predictive models were developed with varied number of neurons and hidden layers. The Pearson correlation coefficient (r), coefficient of determination (R2) and root mean square error (RMSE) were used as the performance evaluation criteria. The results indicated that the model performances, the BP-ANN model with 5 neurons and 1 hidden layer was the optimal model and predicted ETo of the study area with a root-mean-square error (RMSE) of 0.76, a correlation coefficient (r) of 0.94 and coefficient of determination (R2) of 0.88. The BP-ANN predicted ETo was finally compared with 40 years ETo estimated in the study area and the performance of the BP-ANN model performed the best. The findings indicate that the BP-ANN models are efficient in predicting ETo in locations with limited datasets and can be applied in irrigation scheduling and management.

Citation: Ikudayisi A, Calitz A, Nasr M, Abejide S. Prediction of reference evapotranspiration in South Africa using an artificial intelligence approach. Int J Environ Agric Sci. 2022; 2(1): 1002.

Introduction

Evapotranspiration (ET) is a climatic parameter, which describes the evaporation power of the atmosphere. It is an important component in hydrology, agriculture, meteorology and climatology, because it is required for minerals and nutrient transport for plant growth [1]. Factors such as climate, landscape heterogeneity, topography, climate, vegetation type, soil properties, management and environmental constraints are considered while estimating ET [2,3]. ET can be either directly measured using lysimeter, water balance approaches or estimated indirectly using climatological data [4].

The evapotranspiration rate from a reference surface is called the reference ET and denoted by ET_o. Reference evapotranspiration is a combination of the process by which water is lost from soil surface through evaporation and from vegetation through transpiration with both processes occulting at the same time [5]. Reference surface is a hypothetical grass reference crop having an assumed height of 12 cm, fixed canopy resistance of 70 S per m and an albedo of 0.23, it closely resembles an extensive surface of green grass, which is actively growing with a uniform height shading the ground completely [6].

According to [7], in order to estimate ET, reference crop evapotranspiration from a standard surface must be estimated first. Paramount among the combination-type category is the Penman - Monteith (PM) equation, also called FAO-56 method and approved by the Food and Agriculture Organisation (FAO) [8]. However, in many developing countries such as South Africa, a major limitation to the successful use of this FAO-56 PM method is non-availability or limited data sets of the required parameters. Therefore, many alternative methods and techniques are adopted for the estimation of ET_o, among which is Artificial Neural Networks (ANN).

Different ANN models have been developed to estimate and predict ET_o around the world. [1] developed ANN models for the estimation of ET_0 in Burkina Faso. In the study, a Generalized Regression neural network (GRNN) was adopted to estimate ET_o using minimum and maximum temperatures from 1996 to 2006 as the only available input variables. The result of the study shows that using GRNN with minimum climatic data variables as input, performs excellently in the estimation of ET_o.

In a study conducted by [9], ET_o was predicted via an Evolutionary Artificial Neural Network (EANN). The ANN model was trained using Differential Evolution (DE), which is an evolutionary algorithm. In the study, daily climatic weather data obtained from three weather stations in the United States were used to calibrate the model. After the model simulation, it was proved that neural networks have the capacity to model ET_o effectively.

[10] estimated ET_o using an ANN for a paddy field in Indonesia. The model was calibrated using minimum, average and maximum temperature as input variables, because other needed data were not available. From the result of the prediction model, soil moisture was further estimated through another ANN model. This further shows the suitability of ANN in predicting ET_o with limited datasets.

[11] used ANN models to predict ET_o in the irrigation district of Hasanloo dam in Iran. A 21- year (1985-2005) dataset, which include wind-speed, dry and wet temperature, air humidity, percent saturation humidity, air pressure, maximum daily temperature, minimum daily temperature, and period of sunshine were used in the study. Mean Square Error (MSE) statistical analysis was used to evaluate model performance and results shows that feed-forward back propagation model was better for the prediction of ET_o.

The objectives of this study are: (1) to estimate ET_o from limited climatic data using Feed-Forward Back-Propagation Artificial Neural Network (BP-ANN) predictive models; (2) to determine the effect of different numbers of neurons and hidden layers on the predictive ability of BP-ANN models and (3) to find the correlation between the BP-ANN predicted results and the estimated long term average of ET_o at Vaalharts Irrigation Scheme (VIS) in South Africa.

Material and research method

Study area

The meteorological and weather data for this study were collected from the South African Weather Service (SAWS) and Agricultural Research Council. These datasets were extracted from the meteorological stations at Jankempdorp, Vaalharts irrigation scheme in South Africa. Vaalharts Irrigation Scheme (VIS) is the largest irrigation scheme in South Africa and the entire world [12,13]. The scheme is located on a vast land area of about 370 km^2 and majorly used for irrigation. It is located in the Northern Cape Province of South Africa, which is the driest Province in the country. The scheme is supplied with water abstracted from the Vaal River at the Vaal Harts weir about 8 km upstream of Warrenton [12]. A detailed description of the study area is given by [14]. Figure 1 shows the geographical location of the study area.

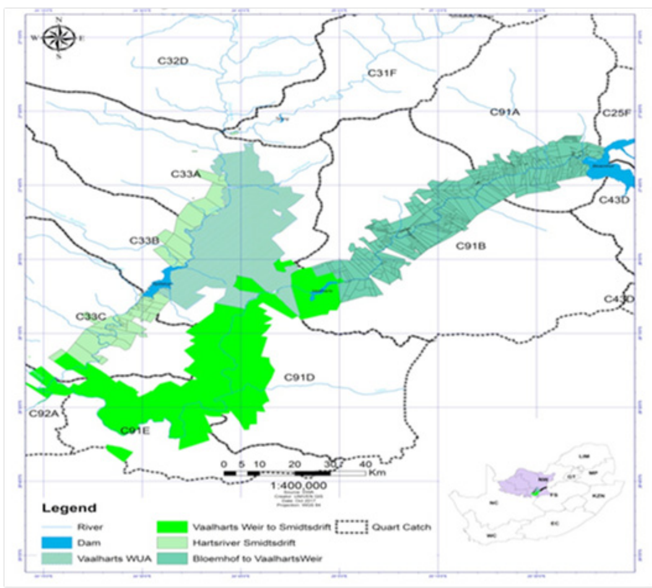

Figure 1: Location map of Vaalharts irrigation scheme.

Artificial Neural Networks (ANNs) modelling

ANNs are black box technique, non-linear data-driven networks, which are opposed to the traditional model based methods. ANNs are computational intelligence method, which was designed and inspired by the theory of neuroscience [15], hence, the name 'neural'. ANNs are mathematical models based on the capabilities of the human brain to predict and classify problem domains.

ANNs work by the input and output process so as to find the function approximation for complex non-linear relationships [6,16]. With the acquired knowledge, the model will be able to learn from examples and generalize the relationships in the data set in order to solve problems [17]. ANNs have been widely adopted for predicting and forecasting in diverse fields of research such as finance, medicine, agriculture, economics, engineering and sciences and also to solve extraordinary range of problems [18].

According to [19], ANNs are very effective in modelling non-linear systems, such as reference evapotranspiration, which is a complex and non-linear phenomenon due to its dependence on climatic parameters, such as humidity, temperature, wind speed, radiation, type and growth stage of a crop. Different types of networks are available for application in ANN and the choice of any network application is influenced by the problem to be solved and the available data. In hydrological applications, a feed-forward Back-Propagation Multilayer Perceptron Artificial Neural Network (BP-ANN) is commonly used [6].

Multilayer Perception (MLP), which was adopted in this study, is the most common and basic structure of an artificial neural network, which is usually composed of three layers of neurons, namely input layer, output layer and hidden layer [20]. During the development of an ANN architecture, the trial and error procedure is used to determine the number of hidden layers and number of neurons (nodes) in each hidden layer, while the neurons in each layer are connected to the neurons in the adjacent layer by means of weighted links, called Synapses [21,22]. Weighted input vectors (x1,x2,x2….xn) are received and processed by each neuron in a layer and then the output is transmitted to the next layer through links. The weighted summation of inputs to a node is converted to an output according to a transfer function [23]. The ANN model activation process is illustrated in Figure 2.

At the input layer of the ANN models, a linear activation function was used, at the hidden layer and output layers, sigmoid function (logistic and hyperbolic) was implemented as recommended by [24], because it makes the speed of learning faster than other functions.

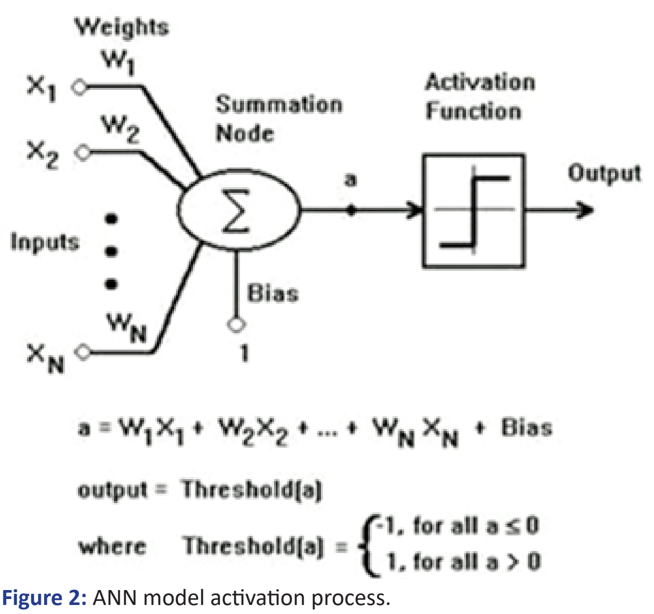

Figure 2: ANN model activation process.

In applying neural networks to predict ETo, the following steps were followed in developing the model: (i) Variable selection; (ii) Formation of training, testing and validation data sets; (iii) Neural network architecture; (iv) Evaluation criteria; (v) Neural network training [18,25-27]. The advantages of adopting neural networks models as outlined by [25] are numerous. These include: (i) they exhibit mapping capabilities; (ii) they learn by example; (iii) The ANN architecture can be trained; (iv) they have the capacity to generalize. They can predict new outcomes from an old trend (iv) They are robust systems and are fault tolerant; and (v) They can possess information in parallel, at high speed and in a distributed manner.

Design and evaluation of BP-ANN models

The average monthly historical data for six parameters were provided by the South African Weather Service (SAWS) and Agricultural Research Council (ARC). These dataset, which covers a period of 40 years (1980-2019) include: minimum and maximum temperature (°C), rainfall (mm), relative humidity (%), wind speed (m/s) and reference evapotranspiration. A major limitation is the unavailability of daily values for most of these parameters, therefore monthly average data were used for this study. This study adopted MATLAB tools in writing scripts that helps to develop the BP-ANN models. The input matrix consists of 480 column vectors of 5 – variables, and the target matrix (output) consists of the corresponding 480- relative valuations. This makes a total of 2400 data records. After data collection, the dataset was first normalized and then randomized as advised by [28].

Different structures, with different number of hidden layers, neurons in each layer, transfer function in each layer, training function, weight/bias learning function, and performance function were selected randomly for the eight models (Table 1). This was done to investigate the effects of different number of neurons and layers on the predictive performance of BP-ANN models. The feed-forward back propagation neural network type was adopted for all the models.

The Levenberg-Marquardt method (trainlm), which applies to small and medium size networks, was used to train all the models. Twenty-eight years' data between 1980and 2007 (1680 data records) were used for training the network, ten-year data between 2008 and 2017 (600 data records) were used to evaluate the models, while two-year data between 2018 and 2019 (120 data records) were used to validate the network. The performances of the developed BP-ANN models were evaluated by statistical model error parameters. The three statistical error parameters used in this study are Pearson coefficient of correlation (r), Root Mean Square Error (RMSE) and coefficient of determination (R^2).

Pearson correlation coefficient (r) indicates the strength and direction of a linear relationship between two variables (for example model output and observed values). It is obtained by dividing the covariance of the two variables by the product of their standard deviations. If we have a series *n* observations and *n* model values, then the Pearson correlation coefficient can be used to estimate the correlation between model values and observed values. The mathematical expression is given in equation (1).

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \cdot \sum_{i=1}^{n}(y_i - \bar{y})^2}} \quad (1)$$

The Root Mean Square Error (RMSE) is a frequently used measure of the difference between values predicted by a model and the values actually observed from the environment that is being modelled. The RMSE of a model prediction with respect to the estimated variable X_{model} is defined as the square root of the mean squared error. The lower the RMSE, the more accurate is the estimation capacity of the developed model. The mathematical expression for RMSE is shown in equation (2).

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(X_{obs,i} - X_{model,i})^2}{n}} \qquad (2)$$

Where X_{obs} is observed values and X_{model} is modelled values at time/place i.

The coefficient of determination (R^2) is the summary measure in a two-variable regression model that indicates the magnitude of this 'goodness of fit'. Coefficient of determination is defined by equation (3).

$$R^2 = \frac{ESS}{TSS} = \frac{\Sigma\, \hat{y}_i^2}{\Sigma\, y_i^2} \qquad (3)$$

But TSS = ESS + RSS

The Total Sum of Squares (TSS) is equal to the sum of the Explained Sum of Squares (ESS) and the Residual Sum of Squares (RSS). It is known the "decomposition of variance". The percentage of the total variation in y_i explained by the regression model". It is always between 0& 1. A larger R^2 means higher explanatory power for the explanatory variable.

Results and discussions

Seasonal variations of temperature, rainfall, relative humility, wind speed and ET_o

In South Africa, there exists four seasons (autumn, winter, spring and summer) [29], however for the purpose of this study, they were classified into two categories, based on rainfall and storm events. The rainy season is between November and March and dry season is between April and October each year. From the seasonal variations of average climatic characteristics of 40 years' data for VIS (Figure 3), it was observed that lowest monthly ET_o is from May to July and highest ET_o is from October to February. This tends to establish a relationship between rainfall and ET_o. Also, considering the effects of other parameters on ET_o, between the months of May and July, VIS experienced the lowest amount of rainfall, lowest temperature, lowest wind speed and highest relative humidity than the remaining months. Based on these observations, it can be concluded that ET_o changes with climate [30,31] and a study of these seasonal variations must be considered further in irrigation schedules and other water resources planning at VIS.

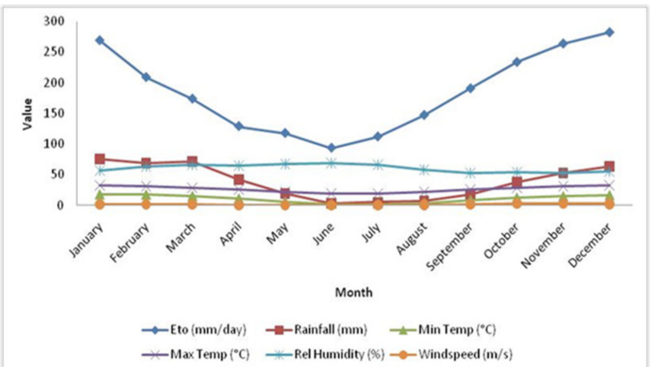

Figure 3: Variations of monthly mean climatic characteristics at VIS from 1980 to 2019.

BP-ANN model design and evaluation

For the eight BP-ANN models developed, model training stopped when the error starts to increase for the validation datasets. The Pearson correlation coefficient (r), Root-Mean-Square Error (RMSE) and coefficient of determination (R^2) for each of the trials were observed and recorded in Table 1.

Table 1: Statistical error parameters of developed BP-ANN models.

Model	Number of hidden layers	No of neurons in hidden layer	r	RMSE	R^2
1	1	5	0.92	0.79	0.78
2	**1**	**10**	**0.94**	**0.76**	**0.88**
3	1	15	0.93	0.78	0.84
4	1	20	0.92	0.79	0.80
5	1	25	0.92	0.82	0.85
6	2	5	0.91	0.91	0.81
7	3	5	0.93	0.86	0.82
8	4	5	0.86	1.19	0.83

As indicated in Table 1, it can be observed that model 2 (with bold face) has the lowest values of RMSE of 0.76 with an acceptable ***r***-value of 0.94and R^2 value of 0.88. This model comprises of ten neurons, one hidden layer and one output element. This is true according to the assertions of [32,33] that ANN models with one hidden layer outperform well multiple layers within its networks and that the model with the lowest RMSE gives the best model performance. Hence, model 2 was selected as the optimal model for predicting ET_O for Vaal harts irrigation scheme in South Africa.

The four regression plots of model 2 (Figure 4) shows the correlation between the output and target values of the network under training at different output values of *R*. The dashed lines represent the best fit for the result, that is, the output is equal to the target. The solid lines represent the best fit linear regression. The training plot gives R value of 0.9796, testing gives 0.9581, while validation plot gives R value of 0.905. It shows that all the data points in the optimal model have good fits. This indicates a good fit for all these datasets and it is similar to the result of [34].

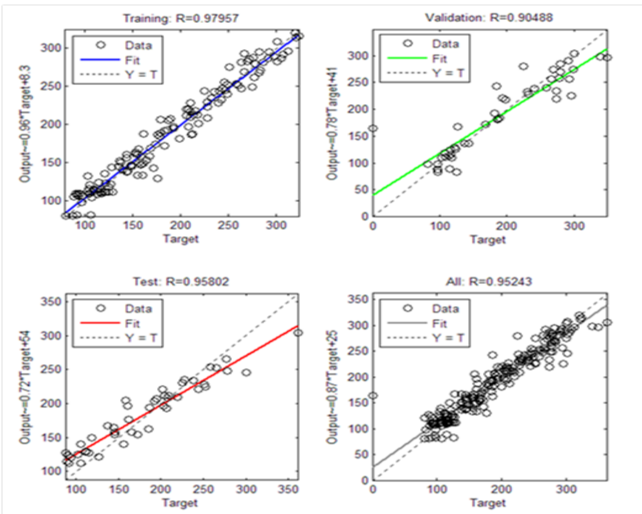

Figure 4: Regression plots of data points for training, testing and validation of model 2.

Correlation of ANN model result with estimated long term average of ET_o

The BP-ANN predicted ET_o was compared with the estimated long term average of ET_o for 40 years in the study area. The novelty is to further explore the accuracy of adopting long term average of ET_o as a solution to the challenges of limited or no datasets. Figures 5 and 6 shows the correlation between the measured ET_o value for year 2019, BP-ANN predicted result and the estimated long term average of ET_o at VIS. The statistical error parameter used to evaluate these two results is the coefficient of determination (R^2).

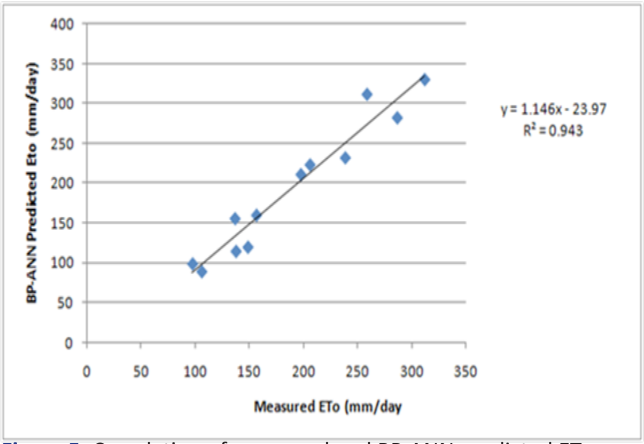

Figure 5: Correlation of measured and BP-ANN predicted ETo values at VIS.

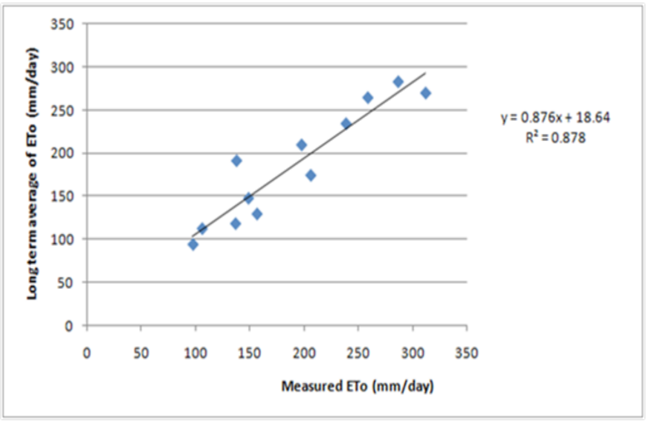

Figure 6: Correlation between measured and estimated long term average ETo values at VIS

From the results, the BP-ANN predicted ET_o produced a R^2 value of 0.943 while the estimated long term average gave R^2 value of 0.878. It can be concluded that BP-ANN models performs better than the estimated long term average ET_o in predicting ET_o in VIS.

Conclusions, limitations and future research

The performance of BP-ANN models and long term average ETo for estimating ET_o at Vaalharts Irrigation Scheme (VIS) in South Africa was evaluated in this study. The aim of the study was to find an alternative method to the standard Penman equation (FAO-56 method), which requires a high data demand. This method often experience limited or no datasets in developing countries such as South Africa. In this study, an alternative method known as Artificial Neural Networks was adopted to estimate ETo at the study area. Eight different BP-ANN models were developed with varying number of neurons and hidden layers but from the study, it was observed that BP-ANN models with a single hidden layer performs better than models with multiple layers in prediction and forecast problems. This is in consonance with the assertion of [33].

Also, from the seasonal variations of average climatic characteristics of 40 years, it was observed that lowest monthly ET_o is from May to July (dry season) and highest ET_o is from October to February (rainy season). This tends to establish a relationship between rainfall and ET_o. Also, considering the effects of other parameters on ET_o, between the months of May and July, VIS experienced the lowest amount of rainfall, lowest temperature, lowest wind speed and highest relative humidity than the remaining months. Based on these observations, it can be concluded that ET_o changes with climate. The BP-ANN predicted ET_o was finally compared with the estimated long term average of ET_o for 40 years in the study area and the performance of the BP-ANN was the best. Therefore, BP-ANN models are efficient in predicting ET_o in locations with limited datasets.

This research is limited to the weather and meteorological parameters obtained for the Vaalharts Irrigation scheme in South Africa. A major limitation in this study is the non-availability of daily values for these measured variables, therefore monthly average data was used for this study. Future research will compare the proposed versus the actual ANN observations for year 2022 to further evaluate the model.

References

1. Traore S, Kerh T, Gibson LA. Modeling Reference Evapotranspiration by Generalized Regression Neural Network in Semiarid Zone of Africa. Wseas Transactions on Information Science & Applications. 2008; 6: 991-1000.
2. Rao LY, Sun G, Ford CR, Vose JM. Modeling potential evapotranspiration Oftwoforestedwatersheds in the Southern Appalachians. American Society of Agricultural and Biological Engineers. 2011; 54: 2067-2078.
3. Ramoelo A, Majozi N, Mathieu R, Jovanovic N, Nickless A, et al. Validation of Global Evapotranspiration Product (MOD16) using Flux Tower Data in the African Savanna, South Africa. Remote Sensing. 2014; 6: 7406-7423.
4. Pereira LS, Allen RG, Smith M, Raes D. Crop evapotranspiration estimation with FAO56: Past and future. Agricultural Water Management. 2015; 147: 4-20.
5. Tfwala SS, Wang YM. Neural Network formulation for evapotranspiration modelling in Swaziland. OIDA International Journal of Sustainable Development. 2014; 7.
6. Kariyama ID. Temperature-Based Feed-Forward Back propagation Artificial Neural Network For Estimating Reference Crop Evapotranspiration In The Upper West Region. International Journal of Scientific & Technology Research.2014; 3.
7. Ojha S, Bhakar SR. Estimation of evapotranspiration for wheat crop using Artificial Neural Network. Journal of Indian Water Resources Society. 2012; 32.
8. Allen RG, Jensen ME, Wright JL, Burman RD. Operational Estimates of Reference Evapo-transpiration. Agronomy Journal. 1989; 81: 650-662.
9. Kisi O. Modeling Reference Evapotranspiration Using Evolutionary Neural Networks. Journal Of Irrigation And Drainage Engineering-ASCE. 2011; 137: 636-643.

10. Arif C, Mizoguchi M, Setiawan BI, Doi R. Estimation of soil moisture in paddy field using Artificial Neural Networks. International Journal of Advanced Research in Artificial Intelligence. 2012; 1: 17-21.

11. Masoud P, Yazdi AB, Araghinejad S, Nejadhashemi AP, Tabrizi MS. Optimal water allocation in irrigation networks based on real time climatic data. Agricultural Water Management. 2013; 117.

12. Ojo OI. Mapping and Modeling of Irrigation Induced Salinity of Vaal-Harts Irrigation Scheme in South Africa. D Tech, Tshwane University of Technology. 2013.

13. Vaalharts Irrigation Scheme (online). 2013. Wikimedia Foundation, Inc. 2015.

14. Ikudayisi A, Adeyemo J. Effects of Different Meteorological Variables on Reference Evapotranspiration Modeling: Application of Principal Component Analysis. International Journal of Environmental, Chemical, Ecological, Geological and Geophysical Engineering. 2016; 10: 623-627.

15. Morimoto T, Ouchi Y, Shimizu M, Baloch MS. Dynamic Optimization of watering Satsung mandarin using neural networks and genetic algorithms. Agricultural Water Management. 2007; 93: 1-10.

16. Kale MU, Nagdeve MB, Bagade SJ. Estimation of evapotranspiration with ANN technique. Journal of Indian Water Resources Society. 2013; 33.

17. Antonopoulos VZ, Antonopoulos AV. Daily reference evapotranspiration estimates by artificial neural networks technique and empirical equations using limited input climate variables. Computers and Electronics in Agriculture. 2017; 132: 86-96.

18. Maier H, Dandy G. Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications. Environmental Modelling & Software. 2000; 15: 101-124.

19. Bhagat AD, Popale PG. Estimation of Reference Evapotranspiration by Artificial Neural Network. International Journal of Innovative Research in Science, Engineering and Technology. 2017; 6.

20. Khoshhal J, Mokarram M. Model for Prediction of Evapotranspiration Using MLP Neural Network. International Journal of Environmental Sciences. 2012; 3.

21. Chauhan S, Shrivastava RK. Estimating reference evapotranspiration using neural computing technique. Journal of Indian Water Resources Society. 2012; 32 (1-2).

22. Laaboudi A, Mouhouche B, Draoui B. Neural network approach to reference evapotranspiration modeling from limited climatic data in arid regions. Int J Biometeorol. 2012; 56: 831-841.

23. Chowdhury A, Gupta D, Das DP, Bhowmick A. Estimation of reference evapotranspiration using artificial neural network for mohanpur, nadia district, west bengal: A case study. International Journal of Research in Engineering and Technology.2017; 6.

24. Oladunjoye AO, Oyewole SA, Singh S, Ijabadeniyi OA. Prediction of Listeria monocytogenes ATCC 7644 growth on fresh-cut produce treated with bacteriophage and sucrose monolaurate by using artificial neural network. LWT-Food Science and Technology. 2017; 76: 9-17.

25. Jha G. Artificial neural networks and its applications. 2007.

26. Kisi O. The potential of different ANN techniques in evapotranspiration modelling. Hydrological Processes. 2008; 22: 1449-2460.

27. Maier HR, Jain A, Dandy GC, Sudheer KP. Methods used for the development of neural networks for the prediction of water resource variables in river systems: Current status and future directions. Environmental Modelling & Software. 2010; 25: 891-909.

28. Al Shamisi MH, Assi AH, Hejase HAN. Using MATLAB to Develop Artficial Neural Network Models for Predicting Global Solar Radiation in Al Ain City - UAE. In: Assi, A. H. ed. Engineering Education and Research Using MATLAB. Shanghai, China: InTech China. 2011; 219-238.

29. Gibson LA, Jatrmain C, Su Z, Eckardt FE. Estimating evapotranspiration using remote sensing and the Surface Energy Balance System- A South African perspective. Water South Africa. 2013; 39.

30. Todorovic M, Karic B, Pereira LS. Reference evapotranspiration estimate with limited weather data across a range of Mediterranean climates. Journal of Hydrology. 2013; 481: 166-176.

31. Xing X, Liu Y, Zhao Wg, Kang Dg, Yu M, et al. Determination of dominant weather parameters on reference evapotranspiration by path analysis theory. Computers and Electronics in Agriculture. 2016; 120: 10-16.

32. Kim S, Kim H. Neural networks and genetic algorithm approach for nonlinear evaporation and evapotranspiration modelling. Journal of Hydrology. 2008; 351: 299-317.

33. Nasr M, Mahmoud A, Fawzy M, Radwan A. Artificial intelligence modeling of cadmium(II) biosorption using rice straw. Applied Water Science. 2015; 10: 295-394.

34. Jain SK, Nayak PC, Sudheer KP. Models for estimating evapotranspiration using artificial neural networks, and their physical interpretation. Hydrological Processes. 2008; 22: 2225-2234.