Agricultural Water Management 163 (2016) 110–124

Contents lists available at ScienceDirect

Agricultural Water Management

journal homepage: www.elsevier.com/locate/agwat

Artificial neural networks versus gene expression programming for estimating reference evapotranspiration in arid climate

Mohamed A. Yassin[a], A.A. Alazba[a,b], Mohamed A. Mattar[b,c,*]

[a] Alamoudi Water Chair, King Saud University, P.O. Box 2460, Riyadh 11451, Saudi Arabia
[b] Agricultural Engineering Department, King Saud University, P.O. Box 2460, Riyadh 11451, Saudi Arabia
[c] Agricultural Engineering Research Institute (AEnRI), Agricultural Research Center, P.O. Box 256, Giza, Egypt

ARTICLE INFO

Article history:
Received 7 March 2015
Received in revised form 2 August 2015
Accepted 11 September 2015
Available online 29 September 2015

Keywords:
Reference evapotranspiration
Penman–Monteith
Artificial intelligence
Arid environments

ABSTRACT

Artificial neural networks (ANNs) and gene expression programming (GEP) were compared to estimate daily reference evapotranspiration (ET_{ref}) under arid conditions. The daily climatic variables were collected by 13 meteorological stations from 1980 to 2010. The ANN and GEP models were trained on 65% of the climatic data and tested using the remaining 35%. The generalised Penman–Monteith (PMG) model was used as a reference target for evapotranspiration values, with h_c varies from 5 to 105 cm with increment of a centimetre. The developed models were spatially validated using climatic data from 1980 to 2010 taken from another six meteorological stations. The results showed that the eight ET_{ref} models developed using the ANN technique were slightly more accurate than those developed using the GEP technique. The ANN models' determination coefficients (R^2) ranged from 67.6% to 99.8% and root mean square error (RMSE) values ranged from 0.20 to 2.95 mm d^{-1}. The GEP models' R^2 values ranged from 64.4% to 95.5% and RMSE values ranged from 1.13 to 3.1 mm d^{-1}. Although the GEP models performed slightly worse than the ANN models, the GEP models used explicit equations.

1. Introduction

Evapotranspiration is the principal variable of the hydrological cycle affecting irrigation water requirements and the future planning and management of water resources. It can be determined either experimentally (directly) or mathematically (indirectly). It can be measured directly by using either a lysimeter or a water balance in a controlled crop area (Gavilan et al., 2007). However, this approach is difficult, time-consuming and expensive.

As the ET_{ref} depends on several interacting climatological factors, such as temperature, humidity, wind speed and radiation, it is difficult and complex to estimate it. Over the last 50 years, experts have developed many methods for estimating the ET_{ref}. Method selection essentially depends on the availability of measured climatic variables. The generalised Penman–Monteith (PMG) method is widely used in agricultural and environmental research to estimate the ET_{ref} and it coincides well with field observations. Many researchers acknowledge that the PMG model is the most promising standardised method for estimating the ET_{ref}. However, it requires a significant amount of climatic data, which may be unavailable or not be reliable in certain locations, especially when dealing with developing countries. In these cases, alternative methods that rely on fewer weather inputs are necessary.

Over the past decade, intelligent computational models have been developed as alternative methods for estimating the ET_{ref}, such as the artificial neural network (ANN) technique (Gorka et al., 2008). With the development of computer technology, ANNs have become increasingly important because of their wide application to different scientific areas. ANNs are defined as massive, parallel-distributed processors made of simple processing units, which have a natural propensity for storing experimental knowledge and making it available for use. ANNs are effective tools for modelling nonlinear processes, as they require few inputs and are able to map input-output relationships without any understanding of the physical process involved (Haykin, 1999).

Several studies have used ANN to estimate the ET_{ref} as a function of climatic variables. Bruton et al. (2000) first developed ANN models to estimate daily pan evaporation using weather data from Rome, Plains and Watkinsville, Georgia. Their ANN models estimated pan evaporation slightly better than multiple linear regression models and the Priestley–Taylor equation.

Kumar et al. (2002) developed an ANN model to estimate the ET_{ref} and evaluated appropriate combinations of various measured

* Corresponding author.
E-mail addresses: mmattar@ksu.edu.sa, dr.mohamedmattar@gmail.com (M.A. Mattar).

http://dx.doi.org/10.1016/j.agwat.2015.09.009
0378-3774/

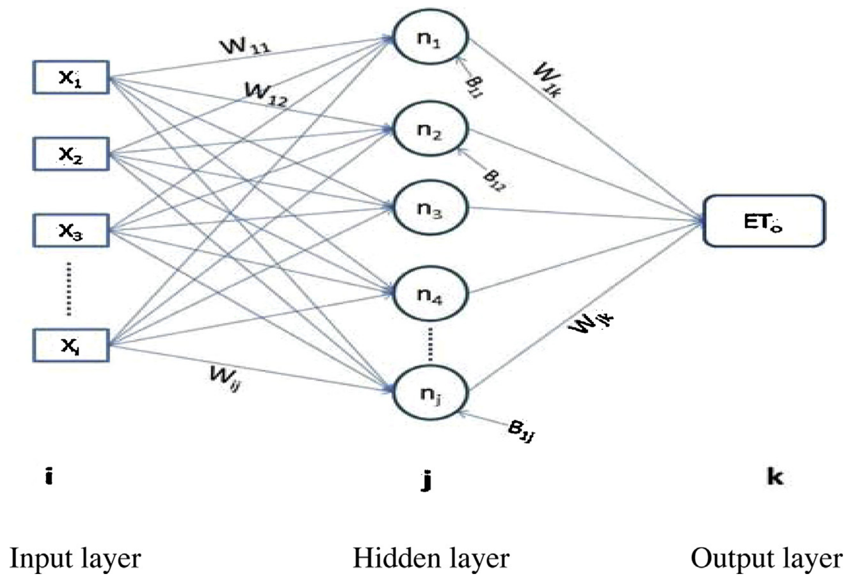

Fig. 1. Architecture of the ANN used to model the ET_{ref}.

weather data. The results indicated that their ANN model predicted the ET_{ref} better than the FAO-56 Penman–Monteith (PMFAO) method. Sudheer et al. (2003) and Trajkovic et al. (2003) reported the performance of radial basis function ANN models in ET_{ref} estimation. Arca et al. (2004) tested 11 combinations for estimating the ET_{ref} using ANN models. Under the most simplified combination, ET_{ref} was estimated as a function of two variables, the extra-terrestrial solar radiation and air temperature.

Landeras et al. (2008) used weather data collected from four weather stations of the Basque Meteorological Service from 1992 to 1996. They compared seven ANN models with different input combinations with ten locally calibrated empirical and semi-empirical ET_{ref} models, using PMFAO daily ET_{ref} values as a reference. The results showed the ANN models obtained better results than the locally calibrated ET_{ref} equations. Huo et al. (2012) trained and tested ANN models to forecast the ET_{ref} using 50 years of meteorological data from three stations in north-west China. They compared the ANN models' performances to multiple linear regressions, the Penman equation and two empirical equations. The results showed that the ANN models exhibited high precision compared to the other models and that ANN models with five inputs were more accurate than those with four or three inputs.

Gene expression programming (GEP) was invented by Ferreira (2001b) and is the natural development of genetic algorithms and genetic programming (GP). GEP has been applied in fields as diverse as artificial intelligence, artificial life, engineering and science, financial markets, industrial, chemical and biological processes and mechanical models. It has been used to solve problems such as symbolic regression, multi-agent strategies, time series prediction, circuit design and evolutionary neural networks (Samadianfard, 2012).

GEP has been used in a number of hydrological and hydraulic modelling problems. Guven and Aytek (2009) used a GEP approach to model the stage–discharge relationship and compared the results with conventional methods. They found that the explicit algebraic formulations resulting from the GEP approach gave the best results. In a similar study, Azamathulla et al. (2011) developed mathematical models to estimate the stage–discharge relationship for the Pahang River based on GP and GEP techniques.

Ghani and Azamathulla (2011) used GEP to model the functional relationships of sediment transport in sewer pipe systems. More recent, Azamathulla and Ahmad (2012) used GEP to predict the transverse mixing coefficient in open channel flows. Zahiri and Eghbali (2012) used GEP to predict the flow discharge in compound channels.

Of the many published studies on the application of GEP in hydrological modelling. However, the use of GEP for modelling evapotranspiration has been recorded by only a few studies. Aytek and Kişi (2008) presented GP as a new tool for estimating the ET_{ref} using daily climatic variables obtained from the California Irrigation Management Information System database. The results obtained were compared to seven conventional ET_{ref} models. They found that the new model produced satisfactorily results and could be used as an alternative to the conventional models. However, Kişi and Guven (2010) investigated the accuracy of linear genetic programming, which is an extension of the GP technique, in modelling the daily ET_{ref} using the PMFAO equation. The linear genetic programming model was found to perform more accurately than the support vector regression model, artificial neural network and four empirical models. Terzi (2013) compared GEP, ANFIS as an alternative approach to estimate daily pan evaporation in Turkey. Traore and Guven (2013) used GEP for modelling the ET_0 using routing weather data from tropical seasonally dry regions of West Africa in Burkina Faso. This study investigates the application of the GEP and ANN for modelling daily ET_{ref}. Moreover, the performance of the GEP models is statistically compared with the ANN models developed.

2. Materials and methods

2.1. Artificial neural network

An artificial neural network (ANN) consists of a large number of interconnecting processing elements and is similar in structure to a biological neural network (Eslamian et al., 2012). ANN usually consists of layers of neurons, weights representing the connection strengths and a transfer or activation function.

In this study, an ANN model of multilayer perception with a universal function approximator is used. Fig. 1 depicts the model layers. The input layer (i) is connected to the hidden layer (j), which is in turn connected to the output layer (k) by means of the connection weights (W) and biases (B). The weights are used to change the throughput parameters and vary the connections to the neurons (n). The biases are used as additional elements inside the hidden