Agricultural Water Management 198 (2018) 28–38

Contents lists available at ScienceDirect

Agricultural Water Management

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

Using gene expression programming in monthly reference evapotranspiration modeling: A case study in Egypt

Mohamed A. Mattar

Agricultural Engineering Department, College of Food and Agriculture Sciences, King Saud University, P.O. Box 2460, Riyadh 11451, Saudi Arabia

ARTICLE INFO

Article history:
Received 29 June 2017
Received in revised form
19 November 2017
Accepted 18 December 2017

Keywords:
Reference evapotranspiration
Gene expression programming
Penman-Monteith
Empirical models

ABSTRACT

The Penman-Monteith FAO-56 equation requires the complete climatic records for estimating reference evapotranspiration (ET_o). The present study is aimed at developing and evaluating a gene expression programming (GEP) model for estimating mean monthly ET_o by using minimal amount of climatic data. The data used in the analysis are collected from 32 weather stations in Egypt through the CLIMWAT database. The results showed that the accuracy of the GEP model significantly improved when either mean relative humidity (RH) or wind speed at 2-m height (u_2) was used as additional input variables. The GEP model with the inputs as maximum and minimum air temperature, RH, and u_2 showed the lowest root mean square error (0.426 mm d^{-1} and 0.430 mm d^{-1}) and, the highest coefficient of determination, (0.963 and 0.962) overall index of model performance (0.960 and 0.960), and index of agreement (0.991 and 0.990) for training and testing sets, respectively. Comparing the results of GEP models with other empirical models showed that the GEP technique are more accurate and can be employed successfully in modelling ET_o.

1. Introduction

Over the last few decades, one of the critical problems encountered in water management was the decrease in water availability in some parts of the world and obtaining accurate information on the agricultural demand. Overcoming these problems and improving the efficiency of water use can be achieved through an accurate irrigation schedule that identifies the required water and the right time for irrigation. Irrigation scheduling targets are meeting crop water requirements as specified in the amounts of evapotranspiration under certain climatic conditions (Hunsakar and Pinter, 2003).

Reference evapotranspiration (ET_o) is a demonstration of the environmental demand for evaporation that is independent of crop type and development, and is both spatially and temporally variable. ET_o is function of weather variables, i.e., wind speed immediately above the surface, air temperature, solar radiation, and relative humidity. These variables depend primarily on the latitude and geographical site of the concerned area. Consequently, ET_o values can be calculated or estimated by using hydro-meteorological methods, which are either physically based equations or empirical relationships between meteorological variables. One of these methods is the Penman-Monteith method (physically based), which estimates the monthly and daily ET_o . The Food and Agriculture Organization of the United Nations (FAO) recommended this method as the standard equation (Allen et al., 1998; Naoum and Tsanis, 2003; Saghravani et al., 2009); it will be referred hereafter as FAO-56. The Penman-Monteith FAO-56 (PMF-56) method is widely used to estimate the ET_o in environmental and agricultural research. PMF-56 coincides well with field observations and calibrates other models under various climates all over the world (Allen et al., 1998; Kashyap and Panda, 2001; Garcia et al., 2004; Popova et al., 2006).

A survey of the literature clearly indicated that the PMF-56 method is superior compared to all other commonly used empirical methods such as Hargreaves-Semani (HS), Blaney-Criddle, Priestley-Taylor (PT), Jensen-Haise, Irmak (IR), and Turc (TR) (Allen

Abbreviations: C_i, GEP estimated value; C_{sx}, skewness coefficient of the applied data; $\bar{C}$, average GEP estimated value; e_a, actual vapor pressure; e_s, saturation vapor pressure; E_i, PMF-56 estimated value; E_n, minimum PMF-56 estimated value; ET_c, crop evapotranspiration; ET_o, reference evapotranspiration; E_x, mamimum PMF-56 estimated value; $\bar{E}$, average PMF-56 estimated value; G, soil heat flux; K_x, kurtosis coefficient of the applied data; n, number of data points; R_a, extraterrestrial radiation; R_H, mean relative humidity; R_n, net radiation; R_s, solar radiation; S_x, standard deviation of the applied data; T, mean monthly air temperature at 2-m height; $T_{\max}$, maximum air temperatures; $T_{\min}$, minimum air temperatures; u_2, wind speed at 2-m height; x_{max}, maximum of the applied data; x_{mean}, mean of applied data; x_{min}, Minimum of the applied data; Δ, slope of the saturation vapor pressure-temperature curve at mean air temperature; λ, latent heat of vaporization; α_1, intercept of fit line equation; α_o, slope of fit line equation; γ, psychometric constant.

E-mail address: dr.mohamedmattar@gmail.com

https://doi.org/10.1016/j.agwat.2017.12.017
0378-3774/

Table 1
Statistical parameters for the data sets used in the study.

	Variable	x_{mean}	$x_{\max}$	$x_{\min}$	S_x	C_{sx}	k_x
Training set	$T_{\max}$	28.04	41.00	17.60	6.07	0.07	−1.04
	$T_{\min}$	14.95	26.30	2.50	5.65	−0.16	−1.01
	RH	58.70	98.20	6.80	16.02	−0.57	0.21
	R_s	19.94	27.90	10.21	5.17	−0.28	−1.22
	u_2	3.14	6.90	0.20	1.43	0.38	−0.46
	ET_o	4.75	1.07	11.45	2.22	0.70	0.10
Testing set	$T_{\max}$	27.56	41.00	16.60	6.09	0.20	−0.96
	$T_{\min}$	14.61	26.20	5.00	5.58	0.06	−1.11
	RH	60.05	87.60	12.60	15.07	−0.66	0.55
	R_s	19.60	27.76	10.46	5.33	−0.27	−1.35
	u_2	3.15	6.70	0.20	1.49	0.28	−0.75
	ET_o	4.57	1.23	12.05	2.20	0.98	1.23
Validation set	$T_{\max}$	29.16	41.00	17.70	6.64	0.03	−1.18
	$T_{\min}$	15.00	23.20	5.20	5.51	−0.15	−1.43
	RH	50.88	98.40	16.70	22.70	0.60	−0.77
	R_s	19.95	26.86	9.71	5.10	−0.40	−1.04
	u_2	2.73	5.30	1.20	1.15	0.77	−0.16
	ET_o	4.64	1.13	8.38	2.05	0.11	−1.25

et al., 1998; Kashyap and Panda, 2001; Yoder et al., 2005; Berengena and Gavilán, 2005; Berti et al., 2014; Feng et al., 2017; Liu et al., 2017). Unfortunately, the PMF-56 method requires complete climatic data that may be unavailable in certain locations, particularly within developing countries, and therefore is difficult to apply. In these cases, alternative methods that rely on a few climatic data are used.

Recently, artificial intelligence techniques like gene expression programming (GEP) are proposed as alternative approach. Ferreira (2001a) had developed GEP. GEP is a computational technique which allows automatically generating algorithms and expressions to solve the problems. This algorithm is used to implement symbolic regression to get a mathematical function that fits a data set (Sakthivel et al., 2012). GEP is the natural development of genetic algorithms (GAs) (Goldberg, 1989) and genetic programming (GP) (Koza, 1992). GEP involves encoded computer programs (nonlinear entities) of different shapes and sizes (expression trees) in linear strings of fixed lengths (chromosomes) (Ferreira, 2001a, 2001b).

Many studies examined the application of GEP in hydraulic and hydrological modeling. Whigham and Crapper (2001) applied GEP to model rainfall-runoff in Australia. Shiri and Kişi (2011) used recorded climatic parameters to compare GEP, adaptive neuro-fuzzy inference system (ANFIS), and artifical neural network estimating the daily pan evaporation. Shiri et al. (2012) used GEP to estimate the daily ET_o for four climatic stations in Northern Spain, where the GEP model's performance was better than the ANFIS, HS, and PT models. Traore and Guven (2013) successfully used GEP to model ET_o using climatic data from tropical dry regions in Burkina Faso. Terzi (2013) compared ANFIS wth GEP to estimate daily pan evaporation in Turkey. Shiri et al. (2014a, 2014b) evaluated a GEP model to estimate evaporation and ET_o by spatial and temporal data scanning techniques. Marti et al. (2015) compared lysimetric vs. Penman–Monteith ET_o targets in GEP models. Alazba et al. (2016) and Yassin et al. (2016a, 2016b) used GEP to estimate the daily ET_o under arid environment in Saudi Arabia. Shiri (2017) compared GEP with empirical models to estimate daily ET_o in hyper-arid regions of Iran. Their results indicated that GEP technique is the best methodology for estimating ET_o.

More water savings will definitely bring about the opportunity to expand the irrigated areas, improve water resources management, and increase food production; this will be great in ensuring food security in countries. Effective and serious action programs are required to increase crop water productivity and reduce water losses. Accordingly, ET_o needs to be calculated accurately. The problem of incomplete or missing climatic data has a significant impact on the estimation of ET_o. Therefore, ET_o must be simulated by using the available minimal number of climatic variables. Therefore, the objectives of this study are to: (1) study the feasibility of GEP models with limited climatic variables to predict the mean monthly ET_o, (2) evaluate the performance of GEP models developed with PMF-56 set as the true reference values using statistical criteria, and (3) compare the accuracy of the results obtained from these models with the results of other empirical equations.

2. Material and methods

2.1. Study area and data

Egypt, with area of about 1,002,450 km^2, is located between 22° and 31° 36′ 15″ N latitude and between 24° 41′ 49″ and 36° 53′ 42″ E longitude. The climate in Egypt is characterized as being extremely dry all over the country, except on the northern Mediterranean coast that receives rainfall in winter. Extremely hot weather during summer months is a general climatic feature of Egypt even though the daytime temperatures are obviously moderate along the northern coast. In northern coast region, the average minimum temperature vary from 9.5 °C in winter to 23 °C in summer and the average maximum temperature vary from 17 °C in winter to 32 °C in summer. In the central and the southern regions, the daytime temperature is higher, particularly in summers; the average maximum temperature exceeds 40 °C. Therefore, different regions were considered in the study to cover all the areas in the country.

The climatic data used in this study were collected from 32 weather stations, obtained from the United Nations Food and Agriculture Organization (UN-FAO) database known as CLIMWAT (Smith, 1993) from 2013 to 2015. The spatial distribution of the selected stations is shown in Fig. 1. The data include the long-term mean monthly for the maximum and minimum air temperatures ($T_{\max}$ and $T_{\min}$) [°C], mean relative humidity (RH), solar radiation (R_s) [Mj m^{-2} d^{-1}], and wind speed at 2-m height (u_2) [m s^{-1}], in addition to ET_o [mm d^{-1}] computed with the PMF-56 equation. The PMF-56 equation is highly rated across a wide range of climate (Allen et al., 1998), and is used to evaluate the results of mathematical ET_o models as a reference standerd (Irmak et al., 2002; Zanetti et al., 2007; Landeras et al., 2008; Jain et al., 2008; Dai et al., 2009; Traore et al., 2010). The PMF-56 equation is defined by Allen et al. (1998) as:

$$ET_o = \frac{0.408\,\Delta\,(R_n - G) + \gamma\,\frac{900}{T+273}\,u_2\,(e_s - e_a)}{\Delta + \gamma\,(1 + 0.34u_2)} \quad (1)$$