Computers and Electronics in Agriculture 113 (2015) 164–173

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

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

Soft computing approaches for forecasting reference evapotranspiration

Milan Gocić [a], Shervin Motamedi [b,c], Shahaboddin Shamshirband [d,*], Dalibor Petković [e], Sudheer Ch [f], Roslan Hashim [b,c], Muhammad Arif [d]

[a] University of Niš, Faculty of Civil Engineering and Architecture, Aleksandra Medvedeva 14, 18000 Niš, Serbia
[b] Department of Civil Engineering, Faculty of Engineering, University of Malaya, 50603 Kuala Lumpur, Malaysia
[c] Institute of Ocean and Earth Sciences (IOES), University of Malaya, 50603 Kuala Lumpur, Malaysia
[d] Department of Computer System and Technology, Faculty of Computer Science and Information Technology, University of Malaya, 50603 Kuala Lumpur, Malaysia
[e] University of Niš, Faculty of Mechanical Engineering, Department for Mechatronics and Control, Aleksandra Medvedeva 14, 18000 Niš, Serbia
[f] Department of Civil and Environmental Engineering, ITM University, Gurugaon, Haryana 122017, India

ARTICLE INFO

Article history:
Received 10 December 2014
Received in revised form 16 February 2015
Accepted 17 February 2015
Available online 10 March 2015

Keywords:
Soft computing
Forecasting
Firefly algorithm
Support vector machine
Wavelet
Serbia

ABSTRACT

Accurate estimation of reference evapotranspiration (ET_0) is needed for planning and managing water resources and agricultural production. The FAO-56 Penman–Monteith equation is used to determinate ET_0 based on the data collected during the period 1980–2010 in Serbia. In order to forecast ET_0, four soft computing methods were analyzed: genetic programming (GP), support vector machine-firefly algorithm (SVM-FFA), artificial neural network (ANN), and support vector machine–wavelet (SVM–Wavelet). The reliability of these computational models was analyzed based on simulation results and using five statistical tests including Pearson correlation coefficient, coefficient of determination, root-mean-square error, absolute percentage error, and mean absolute error. The end-point result indicates that SVM–Wavelet is the best methodology for ET_0 prediction, whereas SVM–Wavelet and SVM-FFA models have higher correlation coefficient as compared to ANN and GP computational methods.

1. Introduction

Evapotranspiration (ET) is considered as one of the most vital elements in hydrologic cycle and has an important role in the field of water resource management. It can be measured directly or estimated by calculating reference evapotranspiration (ET_0). There have been a number of equations used to estimate ET_0. However, the FAO-56 Penman–Monteith (FAO-56 PM) equation (Allen et al., 1998) has been confirmed to be superior in comparison with other techniques that are often used as the reference equation (Gavilan et al., 2007; Lopez-Urrea et al., 2006; Tabari et al., 2013; Ventura et al., 1999). Pereira et al. (2015) discuss the past and future of the FAO-56 PM as well as the accuracy and consistency of operational computation of ET_0 for agricultural uses.

Soft computing methods can be used as alternative techniques because they offer benefits such as no required knowledge of internal system variables, simpler solutions for multi-variable problems and factual calculation (Chaturvedi, 2008; Huang et al., 2010; Zadeh, 1992). Soft computing is an innovative approach in constructing computationally intelligent systems. According to Zadeh (1992), soft computing is an emerging method towards computing which parallels the significant ability of the human intelligence to comprehend in an environment of imprecision and uncertainty.

Nowadays, soft computing methods have shown to be superior and reliable in estimating and forecasting ET_0. Kumar et al. (2011) discussed the artificial neural network (ANN) in creating a model of the evapotranspiration process and selecting training algorithm in estimating and forecasting ET_0. The ANN models for ET_0 are improvised via various learning algorithms like radial basis function (RBF) that was applied by Kim and Kim (2008) and Kisi (2006) during the implementation of the generalized regression neural networks (GRNN), back-propagation (Kumar et al., 2002; Kisi, 2006), Levenberg–Marquardt algorithm (LMA) also known as the damped least-squares (DLS) method (Khoob, 2008; Kisi, 2007; Zanetti et al., 2007) and conjugate gradient decent (Landeras et al., 2008). During the development of the ANN models, different categories of ET_0 estimation methods have been considered (Trajkovic et al., 2000; Kisi, 2007; Khoob, 2007; Zanetti et al., 2007; Kumar et al., 2008, 2009; Landeras et al., 2008; Trajkovic, 2010; Citakoglu et al., 2014).

Genetic programming (GP) is used alongside with ANNs and other soft computing to create simulated hydrological processes like ET without knowing the direct relation between separate components involved (Aghajanloo et al., 2013; Kim and Kim, 2008; Guven et al., 2008; Irmak and Kamble, 2009; Kisi and Guven, 2010; Kisi and Cengiz, 2013; Parasuraman et al., 2007). Recently,

* Corresponding author. Tel.: +60146266763.
E-mail address: shamshirband@um.edu.my (S. Shamshirband).

http://dx.doi.org/10.1016/j.compag.2015.02.010
0168-1699/

explicit mathematical formulation of ET_0 can be done by using gene-expression programming (GEP) (Traore and Guven, 2012, 2013; Shiri et al., 2012, 2014a).

Support vector machine (SVM) (Vapnik, 1995, 1998) is one of the novel soft learning algorithms that has been recently realized for a wide range of applications in the field of soft computing, hydrology and environmental studies (Lee and Verri, 2003; Lu and Wang, 2005; Asefa et al., 2006; Ji and Sun, 2013; Sun, 2013). It is used in pattern recognition, classification, forecasting and regression analysis which has proven to have superior performance over previously developed methodologies; such as neural network and other conventional statistical models (Vapnik et al., 1997; Joachims, 1998; Collobert and Bengio, 2000; Huang et al., 2002; Mukkamala et al., 2002; Sung and Mukkamala, 2003). SVM has gained importance in forecasting problems of ET_0 (Kisi and Cimen, 2009; Kisi, 2013; Shiri et al., 2014b).

SVM was created in accordance with the standard statistical learning processes and structural risk minimization. This reduces the upper bound generalization error rather than the local training error. The latter is the common route used in conventional machine learning methods (Vapnik, 1998). This shows the superiority of SVM compared with other soft learning algorithms.

The following criteria for SVM modeling can be defined: first, the inclusion of the novel solution because of the convex nature of the optimal problem and second, utilization of high dimensional-space sets of kernel functions that discreetly hold non-linear transformation. In essence, there are no presumptions in functional transformation that makes data linearly unique. A linearly separable problem is one in which a decision hyper plane can be found in the input space separating the input patterns with desired output. Application of kernel functions enables mapping of input data points that are not linearly separable into higher dimensional space, which makes them linearly separable.

In the past decades, wavelet transform (WT) has been applied in the field of ET_0 modeling. WT captures both frequency and location information (location in time) (DeVore et al., 1992) and has some desirable properties compared to the Fourier transform (Strang, 1993). The transform is based on a wavelet matrix, which can be computed more quickly than the analogous Fourier matrix (Strang, 1993). Kisi (2011) combined linear regression model and discrete wavelet transform (DWT) to model ET_0 and realized that the wavelet regression (WR) approach performs better than conventional empirical models in daily ET_0 modeling. Cobaner (2013) applied WR method in estimating ET_0 based on Class A pan evaporation, while Partal (2009) and Falamarzi et al. (2014) used a form of wavelet neural network to estimate ET_0.

Our study uses forecasting model to anticipate ET_0 using SVM and firefly algorithm (FFA) (Yang, 2010a,b). FFA is a nature-inspired metaheuristic optimization algorithm that works by the flashing behavior of fireflies (Yang, 2009). The primary purpose for a firefly's flash is to act as a signal system to attract other fireflies. The FFA is used to define the optimal SVM parameters so SVM and FFA are merged in this study to enable more superior performance of SVM model (SVM-FFA). SVM is also merged with WT algorithm (SVM–Wavelet) to capture frequency and location information from the ET_0 data. The objectives of the current study are to construct, develop and evaluate the results of SVM–Wavelet, SVM-FFA, GP and ANN for ET_0 prediction.

2. Material and methods

2.1. Study site and data collection

The study area was Serbia, which is located in the central part of the Balkan Peninsula. The geographic characteristics that affect the Serbian climate are the Mediterranean Sea, the Alps, the Morava valley, the Pannonian Plains, as well as the Rhodope and Carpathian Mountains.

A set of monthly meteorological data, including minimum (T_{min}) and maximum (T_{max}) air temperatures, sunshine hours (n), actual vapor pressure (e_a), and wind speed at 2 m height (U_2) were obtained from 12 meteorological stations within Serbia throughout the period of 1980–2010. Fig. 1 illustrates the topography map of Serbia with the regional distribution of the meteorological stations. Table 1 summarizes the geographic descriptions of the stations. The mean annual T_{max} and T_{min} for most locations varied between 12.3 and 17.9 °C and between 3.8 and 8.4 °C, respectively. The mean monthly n varied for all locations between 153.6 and 182.5 h. The mean annual e_a is ranged from 0.9 to 1.4 kPa, while the mean annual U_2 varied for all locations between 0.9 and 1.9 m s^{-1}. The detailed analysis of the data collection can be found in Gocic and Trajkovic (2013, 2014).

2.2. FAO–56 Penman–Monteith equation

The FAO–56 Penman–Monteith (PM) equation (Allen et al., 1998) is used to estimate reference evapotranspiration:

$$ET_0 = \frac{0.408\Delta(R_n - G) + \gamma\frac{900}{T+273}U_2(e_s - e_a)}{\Delta + \gamma(1 + 0.34U_2)} \quad (1)$$

where ET_0 is reference evapotranspiration (mm day^{-1}), Δ is slope of the saturation vapor pressure curve (kPa °C^{-1}), R_n is net radiation (MJ m^{-2} day^{-1}), G is soil heat flux density (MJ m^{-2} day^{-1}), γ is psychrometric constant (kPa °C^{-1}), T is mean air temperature (°C); U_2 = average 24-h wind speed at 2 m height (m s^{-1}), e_s is saturation vapor pressure (kPa), e_a is actual vapor pressure (kPa), and $e_s - e_a$ is vapor pressure deficit (kPa). As the magnitude of the day soil heat flux beneath the grass reference surface is relatively small, it may be ignored and thus G was set here as 0.

The slope of the saturation vapor pressure curve is calculated as follows:

$$\Delta = \frac{4098\left[0.6108\exp\left(\frac{17.27T}{T+237.3}\right)\right]}{(T+237.3)^2} \quad (2)$$

where T is mean air temperature (°C).

R_n is the difference the net shortwave radiation (R_{ns}) and the net longwave radiation (R_{nl}):

$$R_n = R_{ns} - R_{nl} \quad (3)$$

with R_{ns} and R_{nl} being expressed as (Allen et al., 1998):

$$R_{ns} = (1-\alpha)\left(a_s + b_s\frac{n}{N}\right)R_a \quad (4)$$

$$R_{nl} = \sigma\frac{T^4_{max} + T^4_{min}}{2}(0.34 - 0.14\sqrt{e_a})\left(\frac{1.35(a_s + b_s n/N)}{a_s + b_s} - 0.35\right) \quad (5)$$

where e_a is actual vapor pressure (kPa), σ is the Stefan–Boltzmann constant (4.903×10^{-9} MJ K^{-4} m^{-2} day^{-1}), T_{max} and T_{min} are the maximum and minimum air temperatures, respectively, n is sunshine hours (h day^{-1}), N is daylight hours (h day^{-1}), R_a is extra-terrestrial radiation (MJ m^{-2} day^{-1}), a_s and b_s are regression constants. R_a and N are computed as a function of the local latitude and Julian data. According to Allen et al. (1998), a_s and b_s were set to 0.25 and 0.5, respectively. Albedo (α) was set here as 0.23.