
Data Driven Techniques and Wavelet Analysis for the Modeling and Analysis of Actual Evapotranspiration

Zohreh Izadifar and Amin Elshorbagy

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/52809>

1. Introduction

Evapotranspiration (ET) is one of the important components of the hydrological cycle, which its modeling and analysis is vital for better understanding of watersheds hydrology and efficient water resource designs and managements. Evapotranspiration (ET) is a combined term including the transport of water to the atmosphere in the form of evaporation from the soil surfaces and from the plant tissues as a result of transpiration. Evapotranspiration is considered as a major cause for water loss around the world (Dingman, 2002).

ET is basically a complex and not fully understood mechanism, which varies over temporal and spatial scales. ET can be conceptually expressed either in the form of potential or actual evapotranspiration. Potential evapotranspiration (PET) describes the maximum loss of water under specific climatic conditions when unlimited water is available. The actual evapotranspiration (AET) is the rate at which water is actually removed to the atmosphere from a surface due to the evapotranspiration process. The influence of soil moisture on the AET has made its physical modeling more complicated than the PET. Complexity of AET has also imposed some limitations on the previously developed estimation models. Although the AET is the preferred form of ET in the hydrological analysis, vast majority of the previous studies have investigated the modeling of PET. As a result, there is a vital need for modeling and analysis of AET mechanism. Complexity of the AET physics, limitations of the currently available AET estimation approaches, such as requirement of extensive information and reasonable estimation of models parameters has led to the investigation of some techniques/tools that can model/analyze such complicated mechanism without having a complete understanding of it.

Data driven techniques can provide a model to predict and investigate the process without having a complete understanding of it. Inductive modeling approach is also interesting because of its knowledge discovery property. Using data driven models, one can extract useful

implicit information from a large collection of data and improve the understanding of the investigated process. Machine learning (ML) techniques are modern data driven modeling methods that originated from the advances in computer technologies and mathematical algorithms. These techniques are usually employed for characterizing complicated systems, which cannot be easily understood, analyzed, and modeled. Artificial neural networks (ANNs) and genetic programming (GP) are two robust ML techniques, which apply artificial intelligence for the modeling of complex systems. ANNs are computational models that can be used for the modeling of complex relationships by simulating the functional aspects of biological neural networks. GP is an evolutionary-based technique inspired by the biological evolution to generate computer programs (e.g. models) for solving a user-defined problem.

Both ANNs and GP technique has been examined for the modeling of ET. Kumar et al. (2002) developed an ANN model for the prediction of reference evapotranspiration (ET_0) and compared its performance with that of a conventional method (Penman-Monteith equation) to examine the capabilities of ANNs in ET_0 prediction compared to the PM method. The results of the study showed that the ANN model can predict ET_0 better than the conventional method for the considered local case study. The utility of ANNs for the estimation of reference and crop evapotranspiration (ET_c) of wheat crop was examined by Bhakar et al. (2006) and it was revealed that the ANN model was suitable for the prediction of ET_0 and ET_c . Zanetti et al. (2007) found that by using ANNs, it was possible to estimate ET_0 just as a function of maximum and minimum air temperature. The results of a study conducted by Jain et al. (2008) indicated that ANNs can efficiently estimate ET_0 from the limited meteorological variables of temperature and radiation only. Landeras et al. (2008) developed seven ANNs with different input combinations and then compared ANNs to locally calibrated empirical and semi-empirical equations of ET_0 . Their proposed ANNs performed better than the locally calibrated equations particularly in situations where appropriate meteorological inputs were lacking. Dai et al. (2009) investigated the predictive ability of ANNs for the prediction of ET_0 in arid, semi-arid, and sub-humid areas of Mongolia, China, and conducted a comparison between the estimated ET_0 values from ANNs and MLR. The results showed that regional ET_0 can be satisfactorily estimated using ANN models and conventional meteorological variables.

In the majority of the conducted studies, researchers have focused on the modeling of potential and reference crop evapotranspiration but not actual evapotranspiration (AET). To the knowledge of the authors, the only publications reporting the application of ANNs for the modeling of AET include the studies conducted by Sudheer et al. (2003) and Parasuraman et al. (2006; 2007). Sudheer et al. (2003) estimated the lysimeter-measured AET of rice crop using RBF-ANNs. The results demonstrated that ANNs can successfully estimate the AET. Parasuraman et al. (2006) developed spiking modular neural networks (SMNNs) for modeling the dynamics of EC-measured hourly latent heat flux. The results demonstrated that although the SMNNs are computationally intensive, they can perform better than regular feed forward neural networks (FFNNs) in modeling evaporation flux. Parasuraman et al. (2007) developed a regular three-layered FFNN model for the estimation of EC-measured hourly AET as a function of net radiation, ground temperature, air temperature, wind speed, and relative

humidity. Their results indicated that the ANN model performed better than the currently used PM method in northern Alberta, Canada.

Among the various published studies on the application of GP in hydrological modeling, only a few studies examined the applicability and robustness of GP for modeling of the evapotranspiration process. To the best knowledge of the authors, the only publications that investigated the application of GP for modeling the evapotranspiration mechanism are the studies conducted by Parasuraman et al. (2007), Parasuraman and Elshorbagy (2008), and El-Baroudy et al. (2009). Parasuraman et al. (2007) employed equation-based GP for modeling the hourly actual evapotranspiration process as a function of net radiation, ground temperature, air temperature, wind speed, and relative humidity. The performance of the evolved GP model was compared with that of ANN model and the traditional Penman Monteith (PM) method. It was noted that GP and ANN models had comparable performances and both predicted AET values with better closeness to the measured AET than the PM method. Their analysis also indicated that the effect of net radiation and ground temperature on the AET dominated over other variables. Parasuraman and Elshorbagy (2008) investigated a GP-based modeling framework for quantifying and analyzing the model structure uncertainty on an AET case study. The results of the study demonstrated the capability of the ensemble-based GP in quantifying the uncertainty associated with the hourly AET model structure. El-Baroudy et al. (2009) did not develop a new GP model for AET, but rather developed models using a technique called evolutionary polynomial regression (EPR), and then compared its performance to the ANN and GP models developed by Parasuraman et al. (2007). With the exception of Parasuraman et al. (2007), Parasuraman and Elshorbagy (2008), and El-Baroudy et al. (2009), no other publication was observed that reports an explicit equation for the prediction of AET.

Understanding of the not fully understood mechanism of AET as well as its correlation with the interacting meteorological variables can be improved by exploiting the available time series data and some data mining tools. New digital signal processing tool, namely wavelet analysis (WA), has a robust property for providing multiresolution representation of hydrological time series. Representation of the time series data into time and scale domains makes it possible to extract useful information about temporal cyclic events existing in the underlying signal. In addition, the correlation structure of time series data, in terms of temporal cyclic variations, can be investigated using extensions of wavelet analysis such as cross wavelet analysis. In the field of hydrology, wavelet has been increasingly used for the analysis of spatial-temporal variability of hydrological processes and systems as well as their interactions with climatic variations. WA has been frequently applied for feature extraction of discharge time series data (Saco and Kumar, 2000; Kirkup et al., 2001; Cahill, 2002; Lafreniere and Sharp, 2003; Coulibaly and Burn, 2004; Labat, 2006; Schaepli et al., 2007; Labat, 2008), and characterization of temporal variability of rainfall (Gupta and Waymire, 1990; Kumar and Foufoula-Georgiou, 1993a, b; Kirkup et al., 2001; Coulibaly, 2006; Westra and Sharma, 2006; Miao et al., 2007; Chen and Liu, 2008). In the above-mentioned studies, the utility of WA was mainly employed for detecting and analyzing different periodic events existing in the time series and correlated meteorological signals. Coulibaly and Burn (2004) investigated the temporal and spatial variability of

Canadian streamflows. The results exhibited different period bands of significant activities in the streamflow time series, which were found to be correlated to the considered climatic patterns at some spatial locations. Coulibaly (2006) employed wavelet and cross wavelet analysis to investigate both spatial and temporal variability in seasonal precipitation and its relationship with climatic modes in the Northern Hemisphere. The results revealed striking climatic-related cyclic features in the precipitation time series and, in the temporal-spatial variability of the relationship between precipitation and climate throughout Canada. There are very limited case studies in the literature that investigated the application and capability of WA in analyzing the variability of the evapotranspiration process. Kaheil et al. (2008) used discrete wavelet transform (DWT) for decomposing and reconstructing processes involving the AET phenomenon at various spatial scales, and to find the relationship between the inputs and outputs using support vector machines technique. To the best knowledge of the authors, no effort has been made, in the literature, which benefited from the capability of WA in the temporal scaling of AET variations. Time-scale analysis of the AET signal seems to be an effective approach in improving the understanding of the AET process as well as the efficiency and predictive ability of AET prediction models. Temporal variations of AET and meteorological variables, as well as their correlations, can be examined using wavelet analysis. Wavelet-provided information can improve the understanding of AET temporal variations, its relationship with influential meteorological variables, and hopefully improve the modeling of the AET mechanism.

This chapter presents the ANNs and GP modeling of AET, and the WA of the AET and meteorological signals. The rest of this chapter is organized as follow. In sections 2 and 3, the three techniques of ANNs, GP, and WA are described. Section 4 presents the application of the ANNs, GP, and WA techniques for the modeling and analysis of AET in a specific case study. The hourly eddy covariance (EC)-measured AET is modeled as a function of five meteorological variables; net radiation (R_n), ground temperature (T_g), air temperature (T_a), relative humidity (RH), and wind speed (W_s), using the ANNs and GP techniques, and their performances are compared. The advantage of the investigated data driven models for revealing some information about the AET function and its most influential variables are also examined. Temporal variability of the AET and associated contribution of the meteorological variables is also examined using the wavelet analysis as an approach to modeling input determination. Conclusions of the results and analysis and possible future research are provided in section 5.

2. Data driven modelling

2.1. Artificial neural network

Artificial Neural Networks (ANNs) (Swingler, 1996) are massive networks of parallel information processing systems resembling (simulating) the human brain's analytical function, and they have an inherent ability to learn and recognize highly nonlinear and complex relationships by experience. ANNs learn from empirical examples, which make them a non-rule-based

technique, like statistical methods (Maier and Dandy, 2000). Each neuron (information-processing unit) in ANNs consists of input connection links, a central processing unit, and output connection links (Fig.1a). Input signals are received through the connection links from the outside environment or other neurons. Each connection link is assigned a synaptic weight (w) representing the strength of the connection between two nodes in characterizing input-output relationship (ASCE, 2000). Received information is processed in the central processing unit (neuron body), by adding up the weighted inputs and bias (Eq. 1), and passed through the activation function (Eq. 2). Bias (b) is the threshold value, which must be exceeded before the node (neuron) can be activated (ASCE, 2000). Activation function forms the output of the node and enables the nonlinear transformation of inputs to outputs. The log-sigmoid activation function is one of the two most commonly used activation functions in the literature because it is continuous, relatively easy to compute, its derivatives are simple (during the training process), maps the outputs away from extremes, and provide nonlinear response (ASCE, 2000).

$$t = \sum_{i=1}^n w_i x_i + b \tag{1}$$

$$f(t) = \frac{1}{1 + e^{-t}} \tag{2}$$

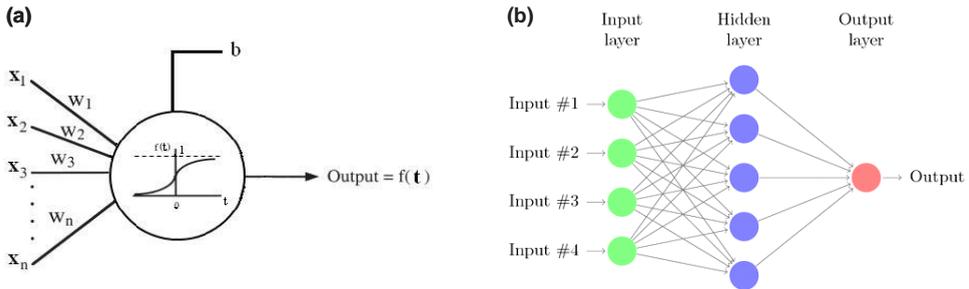


Figure 1. a) Schematic diagram of an artificial neuron, (b) Simple configuration of three-layer feed forward ANN (from Fauske, 2006).

One of the popular types of ANNs, in water resource problems, is the feed forward neural networks (FFNNs) in which the neurons are arranged in layers; input layer, one or more hidden layers, and output layer. The information in FFNNs flows and is processed in one direction from input layer, through hidden layer(s), to the output layer (Fig.1b). Each of the neurons in the hidden layer receives the input signals from the input layer. Received information is processed individually in each of the hidden layer neurons and the outputs are passed to the output layer neuron(s) to release the final response of the network. A simple configuration of three-layer feed forward ANNs is shown in Fig.1b.

It was observed in the literature that a single hidden layer has been usually sufficient for the approximation of conventional hydrological processes (Maier and Dandy, 2000), and it was

noted also, in particular, for the process of evapotranspiration (Kumar et al., 2002; Parasuraman et al., 2007). The number of hidden layers and hidden neurons is specified, based on the complexity of the problem, using different methods (usually trial-and-error procedure (ASCE, 2000)). ANNs with single hidden sigmoid layer and linear output layer are the most popular network architectures in the field of water resources (Cybenko, 1989; Hornik et al., 1989). ANNs learn the pattern of the investigated process by adjusting the connection weights and bias values using the provided examples of input-output relationship (namely, training samples). A training algorithm is employed to optimize the weight matrices and bias vectors, which minimize the value of a predetermined error function. Minimum error function results in an ANN model that can generate the most similar output vector to the target vector.

Back propagation algorithm is the most common type of training algorithm in the FFNNs in water related problems (Maier and Dandy, 2000). The network starts with random weight and bias values and generates the output of the network using the given input data; this step is called the forward step (ASCE, 2000). The network output is compared with the desired target output, and the associated error value is computed. The error is propagated backward through the network and the connection weights are adjusted accordingly. The forward and backward steps, together called an epoch, are implemented repeatedly for several times until the error function reaches its minimum value and the optimum weight and bias values are achieved.

One of the problems that threaten the learning process is over-fitting. It usually occurs when the network has memorized the training examples, but it has not learned to generalize to new situations. Various techniques can be employed to avoid over training and improve network generalization ability such as; regularization and early stopping (Neural Network Toolbox User's Guide, 2009). Regularization attempts to smooth the network response by keeping the size of the network weights adequately small (MacKay, 1992) using the modified form of the error function, which considers network weights and biases (Neural Network Toolbox User's Guide, 2009). Through early stopping approach, an independent test set, namely cross-validation, can be used to monitor the performance of the model on a set of not-yet-encountered examples at some stages of the training process. Training is stopped when error on the cross-validation dataset begins to rise to prevent the model from being over-trained (Neural Network Toolbox User's Guide, 2009). Levenberg-Marquardt (Levenberg, 1944; Marquardt, 1963) and Bayesian-regularization (MacKay, 1992) are two of the common training algorithms in ANNs. Levenberg-Marquardt is one of the high-performance algorithms that appear to be the fastest method for training moderate-sized FFNNs (Neural Network Toolbox User's Guide, 2009). Bayesian-regularization algorithm is an automated regularization algorithm. This algorithm also keeps the network size as small as possible (Neural Network Toolbox User's Guide, 2009).

2.2. Genetic programming (GP)

The origins of evolutionary computation traced back to the late 1950's (Box, 1957; Friedberg, 1958; Friedberg et al., 1959; Bremermann, 1962) when it was proposed for the first time. Genetic programming (GP) was first recognized as a different and new development in the world of evolutionary algorithms in the seminal monograph of Genetic Programming by Koza (1992).

Genetic algorithms (GA) belong to the family of evolutionary algorithms, and are generally considered as an optimization method for searching global optimum of a function using natural genetic operators. Genetic programming (GP), which was introduced by Koza (1992), is an extension of GA for inducing computer programs, as solutions for problems at hand, using an intelligent and adaptive search. This type of search uses the information gained from the performance (fitness) of individual computer programs, in the search space, for modifying and improving the current programs. Depending on the particular problem, computer programs of the GP search space may be different, e.g. Boolean-valued models and symbolic mathematical models (Koza, 1992). Symbolic regression GP evolves computer programs in the form of mathematical expressions in which both functional form and numerical coefficients of the regression symbolic model are optimized through the evolutionary process of GP. This application of GP can be adopted for obtaining explicit mathematical AET models.

In the first step of GP implementation, a population of computer programs is randomly generated. This initial population is called first generation. Symbolic regression models are represented by structured parse trees, which are composed of functional and terminal sets appropriate to the problem. A functional set can be a set of mathematical arithmetic operators such as $\{+, -, *, /\}$, mathematical functions, Boolean and conditional operators, and any other user-defined functions where the number of arguments of each function is specified. The terminal set, which is associated with the nodes that terminate a branch of a tree in tree-based GP (Banzhaf et al., 1998), is defined as independent variables; i.e. the terminal set $z=\{x,y\}$ where x and y are independent variables (Sette and Boullart, 2001).

GP begins to search in the search space of randomly generated models of initial generation. The fitness measure is used to evaluate how well each individual in the population performs. Fitness is usually measured by the errors produced by individual models. Each model in the population is run using a number of provided data instances (training dataset) to measure the performance of each individual over a variety of representative different situations (Koza, 1992). A scalar fitness value is assigned to each individual using the defined fitness evaluation function. Base on the assigned fitness values, some individuals in the population perform better than others with smaller error values, which means that they have higher chance to be selected for the next step of GP.

In the second step, genetic operators are used to create the next generation. Individuals with better performance are allowed to survive and be reproduced in the next population, called mating pool. In the mating pool, two other operations are performed on the reproduced individuals, namely crossover and mutation. Crossover acts on specific percentage of the mating pool population, *crossover probability* (P_c), and results in the creation of new individuals in the population. Crossover exploits two individuals (parents), selected based on their fitness, and splits each parent at the crossover point into two fragments (sub trees), which are swapped between the parents to create two new offspring (Fig. 2). The offspring (new models) are improved individuals, compared to their parents, which carry some genetic properties from each of them.

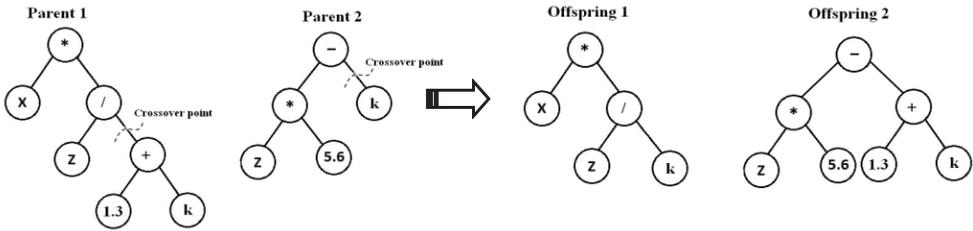


Figure 2. Crossover operation on two selected individuals.

Mutation operates on the population individuals in proportion to the *mutation probability* (P_m). A string is randomly selected from the mating pool and it undergoes some changes at the randomly selected mutation point (Fig. 3). The mutation operation also results in new individuals, which increases the genetic diversity of the population (Koza, 1992). Simply reproduced individuals from the mating pool and newly created individuals resulted from genetic operations of reproduction, crossover, and mutation form the next generation of GP search space. The described evolutionary process is performed iteratively over several generations until some *termination criterion* is satisfied. The termination criterion might be a maximum number of generations or some measure of the goodness of the generated solution and stop the algorithm once the solution is found (Koza, 1992). The result of the GP algorithm, which is a GP-evolved model for the investigated problem, based on the termination criterion, is either the best found model or the best individual of the last GP generation.

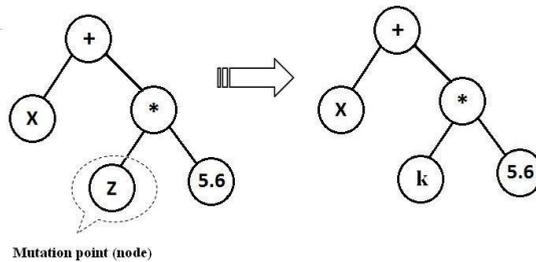


Figure 3. Mutation operation on a selected individual.

3. Wavelet analysis

Natural functions, e.g. meteorological and hydrological processes, operate over a wide range of spatial and temporal scales leading to spatial/temporal variability of interacting mechanisms. AET is a hydrometeorological signal interacting with several temporally/spatially variable meteorological signals. Evaluation of dominant cyclic variations in the AET and

correlated meteorological signals improves the understanding of the mechanism as well as its modeling. Temporal cyclic variations of natural processes are not usually stationary and contain several localized and transient frequency events. Therefore, conventional frequency domain analysis such as Fourier transform cannot reveal the localized natural cyclic events. Wavelet analysis (WA) provides a tool for decomposing the variations of a time series signal into time and scale (frequency) domains; allowing the identification and analysis of dominant temporal cyclic events. The basic component of WA is the wavelet transformation in which the studied function is represented by wave-like oscillating functions. The choice of the wavelet function is of high importance within the wavelet transformation. Wavelet functions are defined in different forms, namely mother wavelets, to have specific properties for information extraction of different types of signals. Figure 4 shows some examples of mother wavelets.

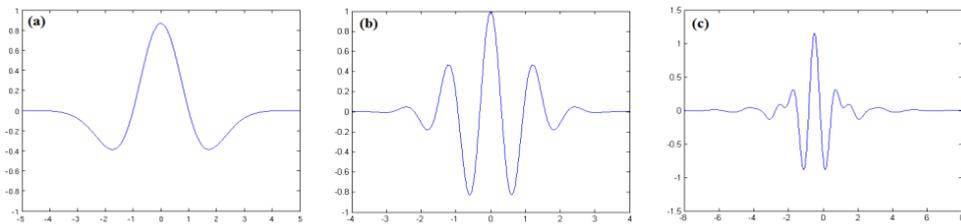


Figure 4. Examples of mother wavelet functions; (a) Mexican Hat, (b) Morlet, and (c) Meyer.

The term wavelet function generally refers to two types of wavelet functions, namely orthogonal and non-orthogonal (Torrence and Compo, 1998). Orthogonal wavelets are mainly used for decomposition of a signal into specific (preferably minimum) frequency bands (Polikar, 1996). This type of wavelet analysis is usually referred to as discrete wavelet transformation, which may not provide a physically meaningful analysis all the time (Si, 2008). Non-orthogonal wavelets are usually used for continuous wavelet transformation (CWT) of time series signals in which a continuous set of frequencies are examined. CWT results in a highly redundant time-scale resolution of the signal, which in one hand induces some uncertainties in the reconstruction of the signal and, on the other hand, provides better scale analysis of the time series (Si, 2003; He et al., 2007). Because of the wide range of possible dominant frequencies that can be obtained using CWT, Coulibaly and Burn (2004) indicated that the CWT is more appropriate for analysis of geophysical and hydrological time series.

3.1. Continuous wavelet analysis

As it was mentioned earlier, the choice of wavelet function is an important component in the wavelet transformation. Wavelet function can be a real or complex function. Complex wavelet functions make it possible to extract the information of both amplitude and phase, which is more suitable for analyzing the signal's oscillatory behavior (Torrence and Compo, 1998). Morlet, Mexican Hat, and Haar are some of the mother wavelets usually employed in the CWT. Morlet is a complex and non-orthogonal wavelet that provides sufficient resolution in time

and scale domains (Grinsted et al., 2004; Si, 2008). Morlet function, with non-dimensional frequency parameter (ω_0) equal to 6, has been shown to successfully work for the analysis of observed time series in different hydrological applications (Lafreniere and Sharp, 2003; Anctil and Tape, 2004; Coulibaly and Burn, 2004; Labat et al., 2005; Si and Zelek, 2005; Coulibaly, 2006). This Morlet wavelet is an exponential oscillatory function defined as (Torrence and Compo, 1998): $\tau_0(\eta) = \pi^{-1/4} e^{i\omega_0 \eta} e^{-\eta^2/2}$, where η and ω_0 are non-dimensional time and frequency parameters. The CWT of a discrete time series data of x_i ($i=1,2,\dots,N$) is defined as the inner product of time series signal with the scaled and translated version of mother wavelet function, $\psi_0(\eta)$, according to a specific scale (s) and time location (τ), which is given as:

$$CWT(\tau, s) = \sum_{i=1}^N x_i(t) \cdot \psi_{\tau,s}^*(t) \tag{3}$$

where $\psi_{\tau,s}(t)$ is the normalized wavelet function and (*) represents the complex conjugate. Normalized wavelet function ensures that the wavelet transform at each scale is not weighted by the magnitude of the scale, which makes a direct comparison of wavelet coefficients at different scales possible (Torrence and Compo, 1998). Normalized wavelet function is defined as:

$$\psi_{\tau,s}(t) = \frac{1}{\sqrt{s}} \psi_{0\tau,s}(t) \tag{4}$$

where τ and s are associated with the time location and scale resolution at which the wavelet transformation is performed. Localization of the time series signal into time and scale domains is implemented, first, by modulating the mother wavelet, corresponding to the current scale, and shifting the scaled wavelet through the signal to the end and performing the convolution at each discrete time location. This results in the time localization of the signal. The procedure is repeated, in the second step, for each scaled wavelet to localize the signal in the scale domain. Wavelet coefficients are computed for all time and scale steps (τ,s) to give the multiresolution representation (or CWT) of the signal. Scaled and translated wavelet at scale s and time location τ is computed by:

$$\psi_{\tau,s}(t) = \psi\left(\frac{t - \tau}{s}\right) \tag{5}$$

According to the mathematical definition of CWT, WA investigates the resemblance of the wavelet function with the in hand signal in the sense of frequency content (Polikar, 1996). In other words, "if the signal has a major component of the frequency corresponding to the current scale, then the wavelet at the current scale will be similar or close to the signal at the particular location where this frequency component occurs. Therefore, the CWT coefficient at this point in the time-scale plane will be a relatively large number" (Polikar, 1996) and will spike in the contour plot of CWT spectrum.

For implementing CWT, it is required to identify the set of analyzed scales a priori. In continuous wavelet analysis, the investigated scales must be incremented continuously to create a complete picture of the wavelet transform. These set of scales (s) can be generated using fractional powers of two (Torrence and Compo, 1998); $s_j = s_0 2^{δj}$, $j=0,1,2,\dots,J$, where s_0 is the smallest scale and J determines the maximum number of scales to be investigated. $δj$ is the scale step size whose value depends on the selected wavelet function (Torrence and Comp, 1998). Complex wavelet function, e.g. Morlet, results in complex wavelet coefficients constitute of real and imaginary parts or amplitude, $|CWT(\tau, s)|$, and phase, $\tan^{-1}[\text{Im}\{CWT(\tau, s)\} / \text{Re}\{CWT(\tau, s)\}]$, respectively. For convenient description of time series cyclic variations, it is common to use wavelet power spectrum, defined as, $|CWT(\tau, s)|^2$, instead of continuous wavelet spectrum. The obtained wavelet power spectrum is also normalized; divided by the variance of the time series (σ^2), $|CWT(\tau, s)|^2 / \sigma^2$, for easier comparison with different wavelet spectra (Torrence and Compo, 1998). Cone of Influence (COI) has been defined in the wavelet spectrum to clarify the areas that are considerably affected by the zero paddings at the ends of the time series signal. Time series data are padded by zeros at both edges to overcome the problem caused by their finite lengths. These zero values decrease the magnitude of wavelet power at the areas close to the edge from which the COI distinguishes regions that are not or negligibly influenced. Length of COI is estimated for each examined scale using a mathematical expression, which is defined as a function of scale. For Morlet wavelet, the length of COI at each scale (s) was defined as $\sqrt{2}s$.

3.2. Statistical significance test

Most of the natural processes (e.g. geophysical and hydrological) are affected with background color noise (white or red noise). The effect of noise is reflected on the signal's wavelet power spectrum. It is essential to identify the powers caused by the background noise and distinguish them from the actual wavelet power peaks. Torrence and Compo (1998) developed a statistical significance test for wavelet power spectra to establish significant levels. Following Torrence and Compo (1998), a statistical significance test is implemented by modeling the appropriate background noise (either white or red) and then testing the significance of the power spectrum peaks against the modeled background noise at certain statistical significance level. Significance test investigates if the peaks of the wavelet spectrum represented some true cyclic features or they are just caused by noise. Most of the geophysical time series are contaminated with red noise background signals (Grinsted et al., 2004). Red noise refers to the temporal fluctuations that have higher amplitude at lower frequencies and lose the magnitude as the frequency increases

According to Hasselmann (1976), lag-1 auto regressive process (AR [1]) is a suitable background noise for many climatological applications. A simple theoretical AR [1] red noise model for modeling the background time series red noise (x_n) is given by (Torrence and Compo, 1998):

$$x_n = \alpha x_{n-1} + z_n \tag{6}$$

where $x_0 = 0$, z_n is the Gaussian white noise, and α is the lag-1 autocorrelation coefficient that can be estimated from observed time series (Allen and Smith, 1996).

It was shown by Torrence and Compo (1998) that the local wavelet power spectrum of the theoretical red noise, at every randomly selected time location, is on average identical to the Fourier transform of the noise time series. In the described statistical significance test, it is assumed that the time series variables have random normal distribution. Fourier power spectrum of the theoretical noise, which is the square of the normally distributed spectrum, has chi-square (χ^2) distribution with two degrees of freedom, X_2^2 , corresponding to the real and imaginary parts. Statistical significance test can be performed at 95% confidence level. To perform the test, the 95% line is developed by multiplying the red noise spectrum by the 95th percentile value of X_2^2 . Wavelet peaks are compared with this 95% line and the peaks that are above this confidence line are identified as cyclic features that are significantly different from background red noise at 95% confidence level.

3.3. Cross wavelet analysis

Cross wavelet analysis is an extension to WA, which examines the linear correlation between two time series. Cross wavelet spectrum between two processes, X and Y , is estimated by (Torrence and Comp, 1998):

$$W^{XY}(\tau, s) = CWT^X(\tau, s)CWT^{Y*}(\tau, s) \quad (7)$$

where $CWT^X(\tau, s)$ and $CWT^Y(\tau, s)$ are the continuous wavelet transforms of the investigated time series, X and Y , and $(*)$ indicates the complex conjugate. Cross wavelet spectrum is complex and can be decomposed into amplitude and phase. Local relative phase between X and Y is estimated by the complex argument (phase), $\tan^{-1}[Im\{W^{XY}(\tau, s)\} / Re\{W^{XY}(\tau, s)\}]$ and the cross wavelet power is also defined as $|W^{XY}(\tau, s)|$. The phase information in the cross wavelet spectrum gives the phase angle difference between the components of the two-time series. Using cross wavelet spectrum, cyclic features at which the underlying time series are co-varying can be detected. The co-variations of two signals demonstrate the existence of a link, in some way, between the underlying processes and also the fact that the information of one process is capable of predicting the other process. This information is very useful when it is of interest to find out the processes that have correlation (or strong correlation) with a target time series, e.g. AET here. The signals, which are showing to have high common power with the target signal in the cross wavelet spectrum, can be used as predictors in the estimation of temporal variations of the target time series. This is important information in the modeling of complex processes, e.g. hydro-meteorological processes, in which determination of important predictors is essential and a challenging task. Statistical significance test of the cross wavelet power spectrum can be conducted using the theoretical Fourier spectra of the two underlying time series. More description on the development of cross wavelet significance test can be found in Torrence and Comp (1998).

4. Application of data driven modelling and wavelet analysis in characterizing AET in a case study

This section describes the application of the previously explained data driven modeling techniques; ANNs and GP, and wavelet analysis for the modeling and analysis of AET in a case study.

4.1. Research scope and experimental data

The experimental data, which were used in this study, were collected from the South West Sand Storage (SWSS) site, located at Mildred lake mine north of Fort McMurray, Alberta, Canada. The SWSS facility is an active tailing disposal facility (dam), which covers an area of about 23 km², holding approximately 435×10⁶ m³ of materials, with 40 m higher than the surrounding landscape and an overall side slope of 5%. The soil cover system within the SWSS consists of a 45 cm thick peat/secondary mineral soil with a clay loam texture overlying the tailing sand. Vegetation cover system varies across the SWSS site including the dominant groundcover of horsetail (*Equisetum arvense*), fireweed (*Epilobium angustifolia*), sow thistle (*Sonchus arvensis*), and white and yellow sweet clover (*Melilotus alba*, *Melilotus officinalis*), and tree and shrub species including Siberian larch (*Larix siberica*), hybrid poplar (*Populus* sp. hybrid), trembling aspen (*Populus tremuloides*), white spruce (*Picea glauca*) and willow (*Salix* sp.) (Parasuraman et al., 2007). The latent heat flux data were originally measured on a continuous basis (Baldocchi et al., 1988) using the eddy covariance technique, and the mean fluxes were recorded every 30 minutes on a data logger. In this study, the hourly Eddy Covariance latent heat (LE) flux (Wm⁻²) data from May 3 to September 21, 2005 and from May 27 to September 8, 2006 were used. The day-time data, which were used for modeling purpose, were only associated with the period of 8:00 AM to 8:00 PM. The data of net radiation (R_n ; Wm⁻²) were also recorded using net radiometer. Air temperature (T_a ; °C), ground temperature (T_g ; °C), relative humidity (RH), and wind speed (W_s ; m s⁻¹) constituted the rest of the meteorological data, which were measured by the weather station located at the site. The LE and R_n fluxes were originally recorded in the unit of Wm⁻² on half hourly basis. For convenient interpretation, the latent heat flux (Wm⁻²) was converted to the equivalent depth of water (mm m⁻²). Since the hourly data were desired to be used in the modeling procedure, conversion of the recorded half-hourly data to hourly data was also implemented in the pre-processing step.

In the first step, the data of the year 2006 were used for modeling purposes with the two proposed techniques (ANNs and GP). Disregarding the missing data, the total number of available instances for modeling in year 2006 is 1207, which were randomly divided into three datasets consisting of 604 instances (50%), 201 instances (17%), and 402 instances (33%) of the data, for training, cross-validation, and testing purposes, respectively. To obtain three statistically consistent subsets, a population of 100 groups of three sub-datasets was randomly generated by sampling from the dataset. The statistical characteristics of the data, i.e. mean and standard deviation, were determined for every subset of each group. Then, the group possessing three subsets with relatively similar statistical characteristics was selected for this study. Aside from the described modeling procedure, a rigorous test was also implemented

in the second step, using the data of 2005, to assess the generalization ability of the developed models in a more realistic way. Disregarding the missing data in 2005, 1600 instances are available. The 2005 dataset has different statistical properties from the data of 2006, which are discussed later in this study.

Multiresolution analysis of the AET and meteorological signals (wavelet analysis) was conducted using the data of the year 2006. The total number of instance that was available for wavelet analysis of the 2006 data is 2520, which constitute the hourly time series data from May 27 to September 9. All of the observed time series data were pre-treated before performing the WA to have zero mean and unite standard deviation.

4.2. ANN modeling

Three-layer FFNNs were adopted in this study for the modeling of AET process. The input layer contained five nodes providing the information of predictor variables; R_{iv} , T_{gr} , T_{ar} , RH , and W_s , to the network and the output layer consisted of a single neuron representing the model output (predicted AET values). Activation functions adopted here include the log-sigmoid and linear functions for the hidden layer and output layer neurons, respectively. The commonly used trial-and-error procedure was employed and different number of hidden neurons ranging from 1 to 14 was investigated for finding the optimum number of hidden neurons. Regularization and early stopping approaches were employed with the examined training algorithms; Levenberg-Marquardt (Levenberg, 1944; Marquardt, 1963) and Bayesian-regularization (MacKay, 1992). Neural Network Toolbox in MATLAB (MALAB® Software, 2003) was used to develop the ANN models to predict AET based on five inputs of meteorological variables, R_{iv} , T_{gr} , T_{ar} , W_s , and RH . The data pool of 2006 was randomly divided into three subsets of training, cross validation, and testing using the approach explained earlier. The training subset was used for optimizing the connection weights and bias of the network. The cross-validation subset was used for early stopping. Once the network was trained, the generalization and predictive ability of the network was evaluated using a completely unseen subset of 2006 called testing subset. The data subsets were normalized so that data fell between 0 and 1. Such scaling of data smoothness the solution space and averages out some of the noise effects (ASCE, 2000). Based on the training subset, different ANN models were trained using Levenberg-Marquardt and Bayesian-regularization training algorithms, using different number of hidden neurons ranging from 1 to 14. For each examined network architecture the training process was repeated several times, each time started with different random initial weight matrices, until satisfactory optimal network (with minimum errors) was obtained. The ANN model with the best performance measures associated with the cross-validation subset was selected as the optimal predictive network. The performance and generalization ability of the trained model was evaluated on the testing subset, which determines how well the ANN model performs on the dataset that have not been seen during the training process (Cheng and Titterton, 1994).

ANNs, as a data driven technique, have the ability to determine the critical model inputs (Maier and Dandy, 2000). In this study, the ANN modeling technique was used to identify the important meteorological variables affecting the AET process. In this approach, no prior

knowledge was assumed about the physics of AET mechanism and the relationships among variables. All possible combinations of input variables, 26 combinations, were considered to be examined as ANN model input sets. Separate optimal ANN models were developed and trained for each input combination set using the model development approach explained earlier. The developed ANN models were compared based on their prediction accuracy in order to identify the most appropriate and efficient combinations of inputs for the estimation of AET. This approach is commonly referred to as trial-and-error procedure, which is under the category of heuristic approaches. The possible combination sets of five available input variables include; five-input combination, " $R_{nv} T_{gr} T_{ar} RH, W_s$ ", four-input combinations, " $R_{nv} T_{gr} RH, W_s$ "; " $R_{nv} T_{gr} T_{ar} RH$ "; " $R_{nv} T_{gr} T_{ar} W_s$ "; " $R_{nv} T_{ar} RH, W_s$ "; " $T_{gr} T_{ar} RH, W_s$ "; three-input combinations, " $R_{nv} T_{gr} RH$ "; " $R_{nv} T_{gr} W_s$ "; " $R_{nv} T_{gr} T_{ar}$ "; " $R_{nv} RH, W_s$ "; " $R_{nv} T_{ar} RH$ "; " $R_{nv} T_{ar} W_s$ "; " $T_{gr} RH, W_s$ "; " $T_{gr} T_{ar} W_s$ "; " $T_{gr} T_{ar} RH$ "; " $T_{ar} RH, W_s$ "; and two-input combinations, " $R_{nv} T_{gr}$ "; " $R_{nv} RH$ "; " $R_{nv} W_s$ "; " $R_{nv} T_{ar}$ "; " $T_{gr} RH$ "; " $T_{gr} W_s$ "; " $T_{gr} T_{ar}$ "; " $T_{ar} RH$ "; " $T_{ar} W_s$ "; " RH, W_s ".

4.3. GP modeling

Major steps in the implementation of GP to solve a problem, e.g. evolution of AET models in the current study, include determination of functional and terminal sets, fitness measure, initializing method, selection method, levels of GP parameters over the run (crossover and mutation probabilities, population size), and the termination criterion. The functional set, which was introduced to GP, included $\{+, -, *, /\}$. The terminal set was defined as $\{R_{nv} T_{gr} T_{ar} W_s RH\}$. Root mean squared error (RMSE) was selected as the fitness function for evaluating individual performance and further fitness-based selection. Ramped-half-and-half method was adopted for initializing the first generation tree structures. Descriptions of initializing methods can be found in Koza (1992) and Banzhaf et al. (1998). The next important issue in the implementation of GP is the fitness-based selection method. Selection method determines the manner by which the individuals are selected based on the assigned fitness values for further GP operations (e.g. crossover, mutation). Roulette wheel selection method was employed here for implementing selection operation in the GP runs. Roulette wheel method is the simplest selection scheme that follows a stochastic algorithm. Several different levels of GP parameters; crossover and mutation probabilities, number of evaluated generations, and the size of population, were executed for obtaining symbolic regression AET models using the training subset. The termination criterion for each GP run was the identified maximum number of generations. The performances of the generated symbolic equations were assessed using the cross-validation subset to select the best equation (model). The selected symbolic equation was then tested using the unseen testing subset to evaluate the predictive accuracy and generalization ability of the proposed model. Data subsets that were used with the GP technique were exactly the same as those used with the ANNs. The data were normalized by dividing the values of variables by their corresponding maximum values. In this way, all variables could have dimensional consistency during the GP implementation (Parasuraman et al., 2007). In this study, GPLAB (Silva, 2005), GP toolbox for MATLAB, was used for the implementing of the GP technique and generating mathematical models based on the datasets where AET is a dependent variable as a function of the five independent variables: $R_{nv} T_{gr} T_{ar} W_s$ and RH .

The performances of the ANNs and GP models were evaluated to compare their predictive accuracies based on three statistical criteria: Pearson's correlation coefficient (R), root mean squared error (RMSE), and mean absolute relative error (MARE), which were calculated as follows:

$$R = \frac{\sum_{i=1}^N (O_i - \bar{O})(P_i - \bar{P})}{\left[\sum_{i=1}^N (O_i - \bar{O})^2 \right]^{0.5} \left[\sum_{i=1}^N (P_i - \bar{P})^2 \right]^{0.5}} \quad (8)$$

$$MARE = \frac{1}{N} \sum_{i=1}^N \frac{|O_i - P_i|}{|O_i|} \quad (9)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (O_i - P_i)^2}{N}} \quad (10)$$

where O_i , P_i , \bar{P} , and \bar{O} are observed values, simulated values, mean of simulated, and mean of observed values, respectively. N is the number of instances in the dataset.

4.4. Wavelet analysis

In this study, only temporal scaling of the variables time series was investigated whereas the spatial variability of the AET and meteorological signals was not considered. Since scale analysis of the time series data were of interest, the CWT was employed for the analysis. The Morlet wavelet with non-dimensional frequency parameter (ω_0) equal to 6 was adopted as the mother wavelet for the current wavelet transformation. For the current analysis, the scale step size of $\delta_j = 0.083$ and the maximum examined scales of $S_j = 16$ and 48 hours were selected for performing the transformation. The smallest scale (S_0) was selected as approximately equal to $2\delta t$, where δt is the time step of the measured time series data. The time step of the AET and meteorological variables is an hour ($\delta t = 1$) and subsequently $S_0 = 2$ hours. A simple theoretical AR [1] red noise model was adopted for describing the background noise. The meteorological variables, whose covariations with the AET time series were investigated in this study, include R_{nr} , T_{gr} , T_{ar} , RH , and W_s . The statistical significance test was performed at 95% confidence level.

Both continuous and cross wavelet analysis were implemented using the software package developed for MATLAB and provided on-line by Grinsted et al. (2004) (<http://www.pol.ac.uk/home/research/waveletcoherence/>). Wavelet and cross wavelet analysis were basically of interest to examine the temporal cyclic variations occurring during day-time (8:00 AM to 8:00 PM) of the AET and meteorological time series. However, wavelet transformation can only be performed on complete (continuous) time series but not non-continuous time series such as day-time data. To obtain accurate wavelet analysis, which were also associated only with the day-time variations, wavelet and cross wavelet analysis were performed using the complete time series data (day-time and night-time data). Then, the wavelet coefficients (spectrum

segments), which were associated with night-time data were cut out to give the spectrum of the day-time only time series data. Wavelet spectra provided in the next section are all associated with the day-time only time series.

4.5. Results and discussions

4.5.1. ANN model and performance analysis

Figure 5 illustrates the influence of number of hidden neurons on the performance measures for two training algorithms; Levenberg-Marquardt and Bayesian-regularization. It appears that Levenberg-Marquardt training algorithm is more sensitive to the number of hidden neurons, represented by larger fluctuations in the error measures with respect to the number of hidden neurons than the Bayesian-regularization algorithm. Figure 5a indicates that the Levenberg-Marquardt algorithm leads to lower values of correlation coefficient (R) for all numbers of hidden neurons compared to the Bayesian-regularization algorithm. Figs. 5b and 5c show that Levenberg-Marquardt results in higher values of RMSE and MARE than Bayesian-regularization for all number of hidden neurons. It indicates that the Bayesian-regularization training algorithm performs more efficiently than the Levenberg-Marquardt algorithm on the dataset under consideration. This might be attributed to some hindrance caused by the use of redundant network parameters (weights and biases) in the output estimation of the network trained by Levenberg-Marquardt algorithm, while the network trained by Bayesian-regularization training algorithm only use the effective network parameters for computing the output. Among the 28 assessed ANN models, the ANN model with eight hidden neurons trained by Bayesian-regularization algorithm resulted in relatively better statistical measures; R of 0.89, RMSE of 0.06, and MARE of 0.28 when evaluated using the cross-validation dataset. Therefore, eight hidden neurons were adopted for the ANN model for the rest of the modeling process.

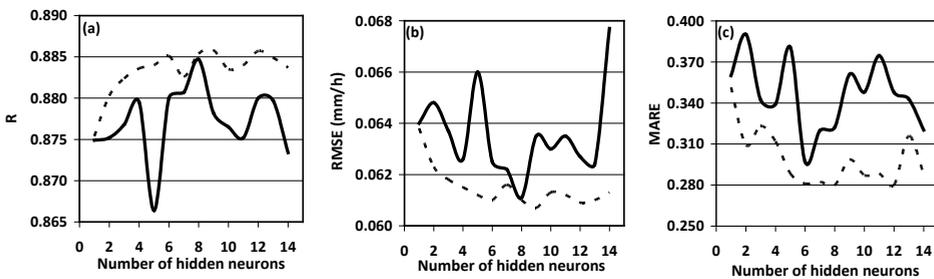


Figure 5. The influence of number of hidden neurons on the network performance for two training algorithms using the cross-validation subset: —, Levenberg-Marquardt; ---, Bayesian-regularization.

Applying the selected ANN model with eight hidden neurons to the testing dataset results in R, RMSE, and MARE values of 0.86, 0.07, and, 0.31, respectively, which indicates reasonably

low values of RMSE and MARE, and high value of R associated with the testing dataset, which imply that the ANN model has good generalization ability for predicting AET based on the unseen testing dataset. For the five available meteorological variables, R_n , T_{gr} , T_a , RH , and W_s , 26 different input combinations could be assessed, which were described earlier in this chapter. In order to examine the importance of each input combination, the associated optimum ANN model was developed. The primary results indicated that net radiation (R_n) is a crucial factor in the estimation of AET; its exclusion from the input set causes serious deterioration of the performance of the ANN models. For instance, ANN model with the predictors set of T_{gr} , T_a , RH , and W_s (excluding R_n) resulted in the performance measures of 0.11 mm/h, 0.69, and 0.54 for the RMSE, MARE, and R, respectively, when applied to the testing subset. The significant role of net radiation, as the main source of energy, in the AET mechanism is expected based on the physics of the AET process. As a result, the rest of the analysis was performed only for the input subsets, which include R_n as one of the predictors. Consequently, the total number of investigated input combinations decreased from 26 to 16. Table 1 shows the performance statistics of ANN models trained using 16 different combinations of inputs.

The best performance of ANNs was obtained when all five meteorological variables were used for the modeling of AET; however, ANN models, which employed the predictor combinations of " R_n , T_{gr} , RH , W_s "; " R_n , T_{gr} , T_a , RH "; " R_n , T_{gr} , T_a , W_s "; " R_n , T_a , RH , W_s "; " R_n , T_{gr} , RH "; and " R_n , T_{gr} , W_s ", also resulted in comparable performances. Among the input combinations of two factors only, the ANN model with predictor set of R_n and T_g performed fairly well, which shows the possibility of using fewer number of predictors for estimating AET in an efficient and parsimonious way. Obtaining acceptable prediction accuracies from different combinations of inputs demonstrates the difficulty of determining the significant input variables for modeling the AET process. Thus, the trial-and-error procedure using the ANN technique might not be the best approach for identifying the important AET predictors. This difficulty can also be associated with the complexity of the AET process itself. The interaction among multiple processes and variables involving the AET makes it possible, for ANN model, to sufficiently capture the variations of AET by using different combinations of variables. It is understood from the results that determination of a unique set of meteorological variables might not be necessary for the estimation of AET. Instead, the effort can be concentrated on the determination of the most efficient and parsimonious set of predictor variables.

4.5.2. GP modeling and performance analysis

Using GPLAB (Silva, 2005) several equation-based GP models were generated at 42 different levels of GP parameters including crossover probability, mutation probability, number of generations, and population size. Table 2 presents the values of RMSE, MARE, and R along with the associated GP parameters obtained with the best eight models generated by GP. The optimum GP models that resulted in the best statistics associated with the cross-validation subset are given below (Equations 11-18):

$$AET = -0.013 + 0.148R_n + 0.01T_g - 0.104RH \quad (11)$$

Input combination	Training			Cross-validation			Testing		
	RMSE*	MARE	R	RMSE	MARE	R	RMSE	MARE	R
R_n, T_g, T_a, RH, W_s	0.06	0.40	0.89	0.06	0.28	0.89	0.07	0.31	0.86
R_n, T_g, RH, W_s	0.06	0.43	0.88	0.06	0.28	0.88	0.07	0.33	0.86
R_n, T_g, T_a, RH	0.06	0.43	0.88	0.06	0.31	0.88	0.07	0.32	0.86
R_n, T_g, T_a, W_s	0.07	0.44	0.87	0.07	0.29	0.87	0.07	0.33	0.85
R_n, T_a, RH, W_s	0.07	0.49	0.85	0.07	0.30	0.87	0.07	0.36	0.83
T_g, T_a, RH, W_s	0.12	0.87	0.61	0.10	0.71	0.62	0.11	0.69	0.54
R_n, T_g, RH	0.06	0.44	0.88	0.06	0.30	0.88	0.07	0.34	0.86
R_n, T_g, W_s	0.07	0.42	0.87	0.07	0.29	0.86	0.07	0.32	0.85
R_n, T_g, T_a	0.07	0.48	0.85	0.07	0.31	0.87	0.07	0.35	0.84
R_n, RH, W_s	0.07	0.47	0.85	0.07	0.29	0.86	0.07	0.37	0.83
R_n, T_a, RH	0.07	0.54	0.84	0.07	0.35	0.86	0.07	0.40	0.83
R_n, T_a, W_s	0.07	0.54	0.83	0.07	0.32	0.86	0.07	0.37	0.83
R_n, T_g	0.07	0.57	0.85	0.06	0.34	0.87	0.07	0.42	0.84
R_n, RH	0.07	0.53	0.82	0.07	0.36	0.87	0.07	0.49	0.82
R_n, W_s	0.08	0.53	0.79	0.08	0.35	0.82	0.09	0.45	0.77
R_n, T_a	0.08	0.51	0.83	0.07	0.34	0.85	0.08	0.43	0.82

*RMSE in mm/h

Table 1. Performance statistics of ANN models with different combinations of inputs.

$$AET = -0.018 + 5.54 \times 10^{-3} T_g + 9.49 \times 10^{-3} R_n T_g \tag{12}$$

$$AET = 0.0784 + 9.2 \times 10^{-3} R_n T_g - 3.5 \times 10^{-3} R_n^2 T_g + 2.7 \times 10^{-4} R_n T_g T_a - 8.64 \times 10^{-7} T_a^2 \tag{13}$$

$$AET = 0.039 + 0.063 R_n + 1.88 \times 10^{-3} T_g + 7.37 \times 10^{-3} R_n T_g \tag{14}$$

$$AET = 0.0696 + 7.836 \times 10^{-3} R_n T_g + 2.569 \times 10^{-3} T_g \tag{15}$$

$$AET = 0.0633 + 3.1 \times 10^{-3} T_g + 0.011 \times R_n + 6.85 \times 10^{-3} T_g R_n \quad (16)$$

$$AET = 0.0775 + 2.23 \times 10^{-3} T_a + 6.35 \times 10^{-3} R_n T_a \quad (17)$$

$$AET = 0.129 R_n + 0.005 T_a \quad (18)$$

where, AET , R_n , T_g , T_a , and RH are the rate of actual evapotranspiration [mm h^{-1}], net radiation [MJ], ground temperature [$^{\circ}\text{C}$], air temperature [$^{\circ}\text{C}$], and relative humidity [fraction], respectively.

Model	Crossover prob.	Mutation prob.	No. of generation	Population size	RMSE (mm/h)	MARE	R
Eq. 11	0.6	0.2	50	60	0.06	0.37	0.88
Eq. 12	0.5	0.2	60	70	0.07	0.34	0.86
Eq. 13	0.6	0.3	60	60	0.07	0.37	0.86
Eq. 14	0.7	0.5	50	300	0.07	0.35	0.85
Eq. 15	0.5	0.2	200	100	0.07	0.43	0.86
Eq. 16	0.7	0.4	200	50	0.07	0.44	0.86
Eq. 17	0.8	0.3	100	40	0.07	0.46	0.85
Eq. 18	0.6	0.3	50	80	0.08	0.40	0.85

Table 2. The best generated GP-based models using various GP parameters for the cross-validation subset.

The optimum GP-evolved models are structurally simple, characterizing the variation of AET as semi-linear functions of meteorological variables, since the models are linear in parameters. Most (six out of eight) of the presented GP models contain R_n and T_g as AET predictors. The appearance of RH (one out of eight times) and T_a (three out of 8 times) was limited in the developed models. Interestingly, W_s never came up as an important predictor in the presented optimum AET models, which means that GP did not find wind speed to be an effective component in the estimation of hourly AET. The simplicity of the models seems to be interesting, especially when the error measures also indicate relatively good generalization ability of the models based on the testing subset (Table 3). It is perceived from the GP models that the AET mechanism can be characterized by structurally simple models, which are also not physically complex. This can be considered as a strong advantage of the GP technique that searches for any possible combination of predictors that can properly model the AET process.

Model	RMSE (mm/h)	MARE	R
Eq. 11	0.07	0.35	0.85
Eq. 12	0.08	0.32	0.83
Eq. 13	0.07	0.32	0.82
Eq. 14	0.08	0.32	0.82
Eq. 15	0.07	0.39	0.83
Eq. 16	0.07	0.40	0.83
Eq. 17	0.07	0.41	0.81
Eq. 18	0.09	0.36	0.79

Table 3. Performance statistics of the GP-based models using testing subset.

Based on the equation-based GP models, the contribution of each meteorological variable in the estimation of AET can also be discussed. This is only possible by using the normalized form of the equations in which all input variables receive their values from a consistent range (e.g. less than 1). Then the contribution of each input variable or factor to the AET can be assessed based on the associated coefficient's magnitude. A selective set of models, which includes only GP models with different physical structures, was identified and the input variables were normalized for further analysis. The selected models, Eq. (11), (12), (13), (14), and (17), are rewritten, in order, as follow:

$$AET = -0.013 + 0.385R'_n + 0.285T'_g - 0.095RH' \tag{19}$$

$$AET = -0.018 + 0.15T'_g + 0.53R'_nT'_g \tag{20}$$

$$AET = 0.0784 + 0.514R'_nT'_g - 0.49R'^2_nT'_g + 0.424R'_nT'_gT'_a - 0.976 \times 10^{-3}T'^2_a \tag{21}$$

$$AET = 0.039 + 0.164R'_n + 0.051T'_g + 0.412R'_nT'_g \tag{22}$$

$$AET = 0.0775 + 0.075T'_a + 0.396R'_nT'_a \tag{23}$$

In these normalized equations, input variables, which are associated with the models' linear coefficients (e.g. T'_g , $T'_gR'_nT'_a$, and $R'^2_nT'_g$), are normalized inputs by dividing each of them by its corresponding maximum values and AET is the rate of actual evapotranspiration [mm h⁻¹]. These normalized models were only developed and used for interpreting the contribution of different inputs to the estimation of AET.

Equation (19) indicates that AET can be estimated as a simple linear function of R'_n , T'_g , and RH' , which is highly dominated by the net radiation and ground temperature variables. Equation (20) also has a simple structure describing the AET process as a nonlinear function, of only net radiation and ground temperature, which is dominated by the two-factor interaction of R'_n and T'_g . The interaction factor of $R'_n T'_g$ has larger contribution to the estimation of AET than the T'_g individually, with the average contribution magnitude of 0.15 and 0.10 mm/h for the terms $0.53R'_n T'_g$ and $0.15T'_g$, respectively. This indicates that when some factors are interacting, their interactions influence the individual contribution of each variable to the AET mechanism. Consequently, the interaction term is more responsible for AET variations than the individual variables. In Eq. (21), the air temperature (T'_a) variable has been included in addition to the R'_n and T'_g . The air temperature has appeared both as an individual variable and as an interacting factor in the three-factor interaction term of $R'_n T'_g T'_a$. According to the coefficients associated with these variables in Eq. (21), T'_a can affect the rate of AET only through the influence it might have on the R'_n and T'_g (interacting coefficient of 0.424 compared to that of air temperature, 0.000976). The structure of Eq. (22) also confirms the importance of interaction effects of multiple variables rather than the individual processes. The combined component of $R'_n T'_g$ is more responsible for the variation of AET than the R'_n and T'_g individually. The average contribution magnitude of each of the terms $0.412R'_n T'_g$, $0.164R'_n$, and $0.051T'_g$ in the estimation of AET values are 0.12, 0.05, and 0.03 mm/h, respectively. Equation (23) demonstrates that the AET mechanism can even be characterized as a simple semi-linear function of R'_n and T'_a only, which are commonly available meteorological measurements. Again, the AET model is dominated by the interaction factor of the two variables. The interaction and individual terms of $0.396R'_n T'_a$ and $0.075T'_a$ are contributing to the estimation of AET by the average magnitude of 0.11 and 0.05 mm/h, respectively. Although the generated models, based on error measures, are performing well and relatively similar, they are using different combinations of inputs with different mathematical structures. This demonstrates that precise identification of the meteorological variables driving the AET process is not an easy and straightforward task, where different combinations of inputs may result in relatively good AET estimation. The results obtained from the GP-evolved models indicate that the hourly AET process can be estimated by both linear and nonlinear relationships. Using the above-listed GP models, one may choose one of them for estimating the rate of AET based on the meteorological data that are available. Thus, the proposed GP models can suit different conditions of data availability.

4.5.3. Comparison among ANN and GP models

The performances of the models from the two proposed techniques; ANNs and GP, were compared based on the testing subset. It can be seen (Table 4) that the equation evolved by GP resulted in slightly larger MARE value compared to that of ANN model. Despite discrepancies among the statistics, the differences are small, which implies that the models have comparable performances for estimating AET based on the meteorological variables. Errors produced by the ANN and GP models have similar statistical characteristics and follow the same probability distribution of LogLogistic (Figure 6).

Model	RMSE (mm/h)	MARE	R
ANN	0.07	0.31	0.86
GP (Eq. 20)	0.07	0.35	0.85

Table 4. Performance statistics of different models using testing subset..

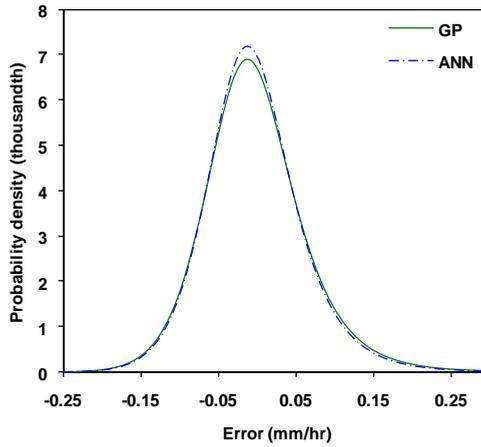


Figure 6. Probability distribution of the data driven models errors.

Figure 7 illustrates the scatter plots of the predicted AET values by ANNs and GP, respectively, versus observed data, using the testing subset. Based on the visual comparison, no substantial difference can be observed among the predictive abilities of the proposed models.

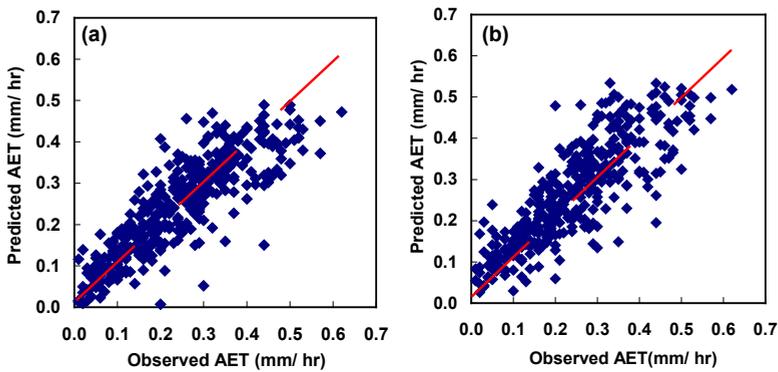


Figure 7. Scatter plots of predicted actual evapotranspiration (AET) versus observed AET by (a) ANN, (b) GP (Eq. 11) using testing subset.

In order to evaluate the generalization ability of the developed models in a more realistic and rigorous way, the optimum models obtained from the ANNs and GP techniques, using the 2006 data, were employed for the prediction of actual evapotranspiration of a different year (2005). This assessment was also conducted to identify the possible superiority of any of the proposed models for future prediction applications. The 2005 dataset, which was used for implementing the rigorous generalization test, has different statistical properties from the year 2006 dataset, which was employed for training and testing during model development (Table 5). Applying the developed models to a statistically different dataset helps to evaluate and compare the models' predictive abilities on instances from a statistically different population. The hourly meteorological variables of R_n , T_g , T_a , W_s and RH of the year 2005 were used as the inputs of the optimum models, including ANN and GP (Eq. 11 and Eq. 12) to estimate the AET.

Year	Minimum (mm/h)	Maximum (mm/h)	Average (mm/h)	Median (mm/h)	SD * (mm/h)	Coefficient of variation
2005	-0.05	0.68	0.18	0.16	0.12	0.65
2006	-0.06	0.67	0.24	0.23	0.13	0.55

*Standard deviation

Table 5. Statistical characteristics of AET data of the years 2005 and 2006.

A comparison among the performance statistics of the data driven models, used for the prediction of AET in 2005, is given in Table 6. It is apparent that GP model is performing better than the ANN model with lower error measure values and higher correlation coefficients. GP technique was able to capture the semi-linearity of the AET process and to characterize it by simple equations. However, ANN model, because of its structure, tried to fit a complex non-linear model to the AET process, which was unnecessary and resulted in its poor performance in generalization.

Model	RMSE (mm/h)	MARE	R
ANN	0.10	0.91	0.78
GP			
- Equation (11)	0.07	0.55	0.82
- Equation (12)	0.06	0.47	0.85

Table 6. Performance statistics of different models using 2005 data.

For better interpretation of models' performances, further analysis was conducted. Scatter plots of the predicted AET values by ANN and GP models versus observed values, using the 2005 dataset, are illustrated in Fig. 8. It can be seen in Fig.8a that the ANN model is overestimating most of the AET values in 2005, which was expected from the large value of MARE.

Figure 8b indicates that the GP model is performing well on the estimation of AET in 2005. This demonstrates the superiority of GP model over the ANN with regard to the generalization ability.

The two different types of comparison discussed above highlighted the importance and also the reliability of the approach one may use for comparison purposes in the modeling process. In the first approach, the unseen testing subset, which was coming from the same year (statistical population) that was used for developing (training) the models, was employed for testing the generalization ability of the models. Based on this comparison, no considerable difference was observed among the two proposed data driven models in terms of models' generalization ability. However, using the second approach; testing the developed models using data from a different year, led to a more realistic assessment of the predictive accuracy, which revealed the discrepancies among the data driven models much better. Consequently, the choice of the testing dataset on which the generalization ability of the models is evaluated is important, and in the case of inappropriate and/or insufficient testing data, incorrect conclusions might be made.

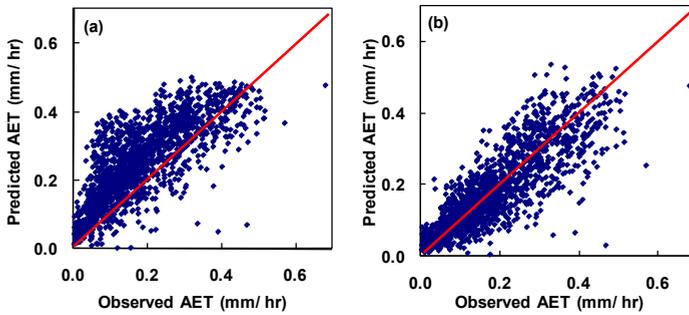


Figure 8. Scatter plots of predicted actual evapotranspiration (AET) versus observed AET by (a) ANN and (b) GP (Eq. 21) models using 2005 data.

Another point of interest, which was observed in this analysis, is the issue of representation of a set of optimum GP equations instead of representing only the best GP equation. Out of the best GP-evolved models, Eq. (20), as an example, performed better than Eq. (21) when they were applied to the testing dataset of 2006. However, Eq. (21) had better performance than that of Eq. (20) when 2005 data were tested. This indicates that no single GP equation can be adopted as the best GP model, and thus, representation of a set of GP equations as the optimum GP models is necessary (Parasuraman and Elshorbagy, 2008).

The proposed AET estimation models can also be compared in terms of their complexity and efficiency. The number and the data availability of the various inputs and the simplicity/complexity of the model usage can also be investigated for better comparison of the proposed models. Table 7 provides the sum squared error (SSE) values of the proposed data driven models. Based on the SSE values, the ANN model has better fitness to the data than the GP models although it is more complex based on the number of input variables and optimized

parameters. The GP model (Eq. 11) has also comparable SSE value, which indicates its good fitness though it is simple in terms of the number of inputs and estimated parameters.

Model	SSE*	No. of required input variables	No. of optimized parameters
ANN	2.16	5	24
GP			
- Equation (11)	2.83	3	4
- Equation (12)	3.50	2	3

* Sum squared error

Table 7. Akiak information criterion and sum squared error of the data driven models and their required inputs

The ANN technique provides an implicit model from which no explicit information about the AET process can be easily obtained. As a result, ANN models might be used when prediction of AET is the only concern of the modeling process. In other words, accurate estimation of AET is more important than the understanding of AET mechanism itself. Furthermore, the significant input variables that are employed for AET estimation in the ANN model cannot be explicitly/easily identified, since the associated information is stored in the network connection weights and cannot be easily interpreted. For the end user, application of ANN models is also not as easy as the equation-based models. The GP model is an equation-based model and is of interest for hydrologists and modellers because of its transparency and simplicity in application and usage. Explicit form of equation-based models, such as GP, makes it possible to extract some information about the physics of the process. The GP model has the advantage of using fewer input variables (Table 7) and also has simple and realistic structure. Consequently, the GP model becomes more applicable when a limited number of meteorological variables is available or can be measured. The simple structure of the GP models makes it easy for the users to understand how input variables are contributing to the AET process.

4.5.4. Wavelet analysis

As it was described in the previous chapter, the length of cone of influence (COI) is defined as a function of scale (e.g. $\sqrt{2}$ s for Morlet wavelet) and increases with the scale. Since the studied range of scales mainly constitute of small scales (less than 16 hours), for all spectra shown hereafter, edge effects (COI) are negligible near both end regions of the wavelet transformation, and consequently, cannot be seen in the spectra. The thick black contour lines, seen in the wavelet spectra hereafter, enclose areas in which the values of wavelet powers are significantly greater than the background red noise at 95% confidence level. Black contour lines might be seen only as black areas in the spectra because the small-scale wavelet is narrow in time domain (high time resolution) and the peaks appear very sharp.

Daily variations are apparently the known cyclic pattern existed in the meteorological signals. Continuous wavelet transformation (CWT) conducted at scale range of 2 to 48 hours confirmed the presence of such diurnal cyclic variations. Figure 9 shows noticeably strong wavelet powers at the scale band of 16 to 32 hours, which is most likely due to the diurnal cyclic variations in the AET and R_n time series, as example. Similar spectra for other meteorological signals of T_{gr} , T_a , RH , and W_s are provided in Appendix I. Strong wavelet powers at the band scale of 16 to 32 hours indicates that the larger-scale cyclic events are the dominant source of temporal variations in the studied time series. Consequently, small-scale cyclic events (e.g. less than 16 hours) may not play a considerable role in inducing the signals' temporal variations. In this study, the small-scale cyclic events were of interest to be investigated though they are not the main source of temporal variations. This is because the small-scale (hourly) variations and modeling of AET were the focus of this study, and the WA was examined for identifying the most important input variables in the estimation of small-scale AET values.

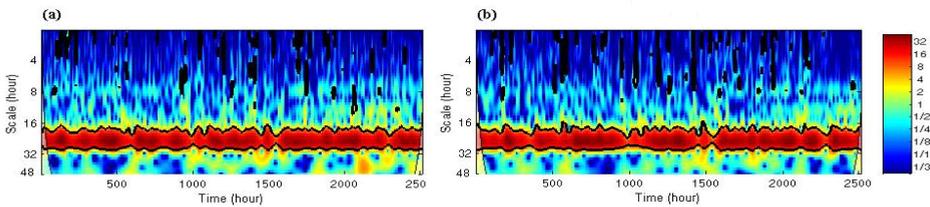


Figure 9. Continuous wavelet power spectrum of hourly time series for the scale range of 2 to 48 hours; (a) AET, (b) R_n .

Figure 10 shows continuous wavelet spectrum of the daytime hourly AET signal. Several wavelet peaks were found to be significantly different from the background red noise at scales of 2 to 8 hours, which were fairly observed along the studied period (growing season of 2006). The significant powers appeared at the scales of 2-4 hours are more frequent than those appeared at 4-8 hours showing that most of the short-time intermittent variations in the AET time series are probably produced by the 2-4 hours scale cyclic events.

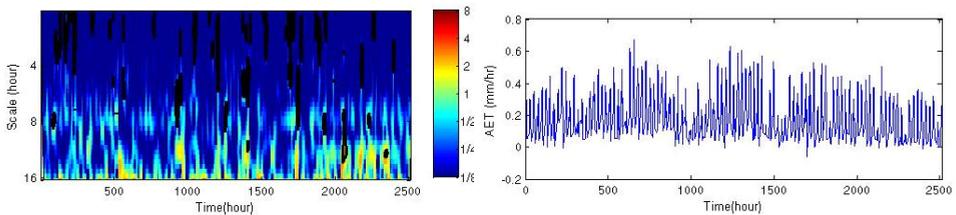


Figure 10. Continuous wavelet power spectrum (left) and time series of hourly AET (right).

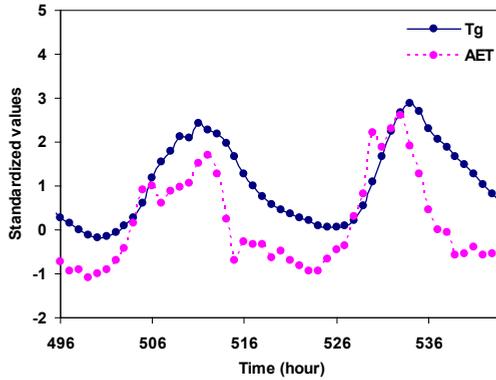


Figure 12. Time series of AET and T_g for a typical time-window of 48 hours.

Wavelet power spectrum of T_g signal is shown in Fig. 11 and exhibits no specific significant cyclic behaviour at scales less than 8 hour. The only cyclic features, which were identified to be significantly different from red noise, were at the scales of 8 to 16 hours. Detected features did not show high magnitude powers (mostly in green), which demonstrate the weak contribution of small-scale cyclic events in the temporal variations of T_g time series. This can also be observed in the zoomed time series of a typical 48-hour window of T_g signal (Figure 12). The time series of T_g does not exhibit much short-time cyclic variations, e.g., less than daily, compared to that of AET. This might be attributed to the physics of the T_g time series, which changes gradually over the short terms and is not immediately influenced by sudden fluctuations in the atmospheric condition.

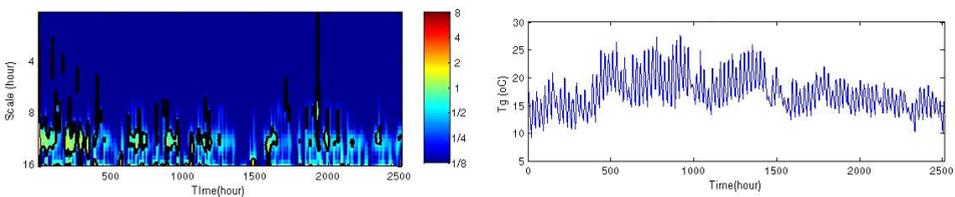


Figure 11. Continuous wavelet power spectrum (left) and time series of hourly T_g (right).

Figure 13 demonstrates limited detected cyclic features in the wavelet power spectrum of T_a , which are different from the background red noise at scales of 2 to 8 hours. The significant wavelet peaks that were identified at scales 8 to 16 hours do not contain large magnitude powers. Consequently, T_a signal might not constitute of considerable small-scale cyclic variations. Wavelet power spectrum of RH (Fig. 14) shows significant peaks at scales of 2-8 hours at several time locations along the studied period. Wavelet powers of RH spectrum at

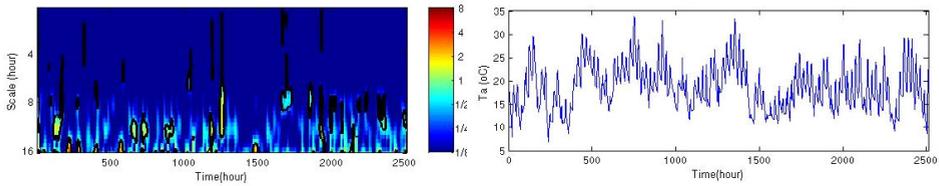


Figure 13. Continuous wavelet power spectrum (left) and time series of hourly T_a (right).

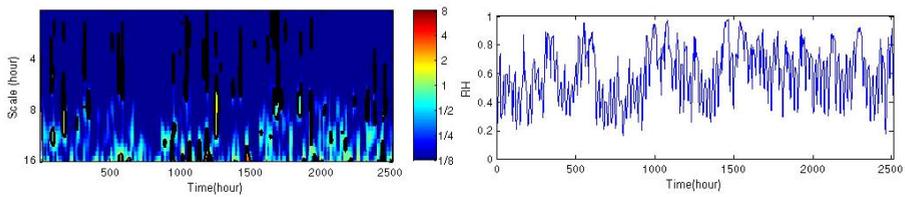


Figure 14. Continuous wavelet power spectrum (left) and time series of hourly RH (right).

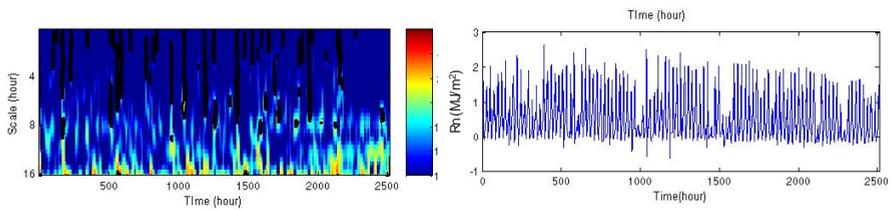


Figure 15. Continuous wavelet power spectrum (left) and time series of hourly R_n (right).

very small scales (around 2 hours) are not significantly different from the background red noise.

Cyclic temporal variations of R_n were identified to be significantly different from the red noise at small-scale band of 2 to 8 hours (Fig. 15). These significant cyclic features appeared quite frequently along the studied time duration especially at scales of 2-4 hours. Wavelet analysis of the W_s signal exhibited significant cyclic features at scales of 2 to almost 7 hours and 8 to 16 hours (Fig. 16). Small-scale cyclic features (2-4 hours) appeared more frequently than the larger-scale features (8-16 hours).

Out of the analyzed time series, the AET, R_n , RH , and W_s exhibited frequent small-scale cyclic features, which were found to be significantly different from the background red noise. No specific significant small-scale cyclic features were detected in the wavelet spectra of T_g and T_a , which could be attributed to two possible reasons. First, temporal variations of air and ground temperature signals do not involve considerable small-scale cyclic features and are mostly generated by larger-scale cyclic trends. Second, the likely existing small-scale cyclic

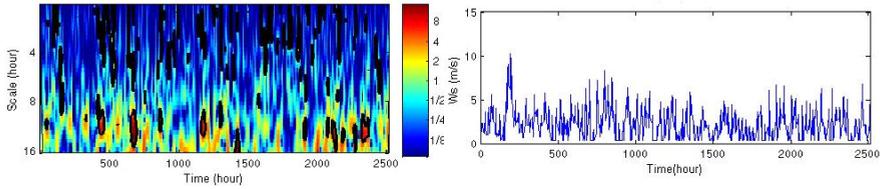


Figure 16. Continuous wavelet power spectrum (left) and time series of hourly W_s (right).

variations are not large enough, in magnitude (because of slight changes of these variables in small time scales), to be differentiated from the background red noise and consequently, cannot be detected as significant cyclic features in the wavelet power spectrum.

As it was discussed earlier in this section, although several small-scale cyclic features were found in the wavelet spectra of the time series, they might not substantially contribute to the temporal variations of the considered signals. The reason might be the existence of larger-scale features (e.g. scales of 16 to 32 hours), which induce the major temporal variations in most of the studied time series (Izadifar, 2010).

4.5.5. Cross wavelet analysis

It can be seen from the cross wavelet spectrum of AET- R_n (Fig. 17) that both time series have common significant powers at scales of 2 to 8 hours along the studied period. This demonstrates the significant linear correlation between AET and R_n signals at small 2scales at 95% confidence level. To be more specific, the significant AET- R_n correlations appeared at particular time locations but not continuously along the time axis. This means that the linear covariation of these two time series becomes significantly different from red noise only at some periods of time, which is more likely due to the low magnitude of variations at small scales compared to that of larger scales (e.g. diurnal). In other words, there might be non-noise small-scale covariances between the time series; however, since they are not major source of variations in the time series and are low in magnitude, associated cross wavelet powers cannot be distinguished from background noise. Significant powers of AET- R_n cross wavelet spectrum imply that small-scale variation of AET can be explained by R_n time series. Phase information is provided in the cross wavelet spectra using arrows. Arrows, pointing right and left, show in order in-phase and anti-phase relationship between the two time series. Fig. 17 indicates the in-phase relationship between AET and R_n at significant areas. By in-phase, it means that the two time series are positively correlated. The relationship between AET and R_n was observed to be not necessarily in-phase over all detected significant areas, since there are some cases when other involved factors affect the conventional cause and effect relationship between the two signals. Varying (not fixed) phase information was also observed in the cross wavelet spectra between AET and other considered meteorological signals, which might be attributed to the range of studied scales (small-scales). Small-scale cyclic events are not the dominant source of variation in the studied time series and consequently, might not carry solid phase

information. Larger-scale features, which have more contribution to the temporal variations, may contain less varying and more reliable phase information (Izadifar, 2010).

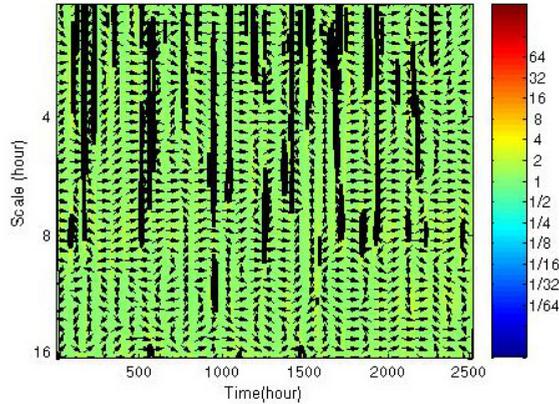


Figure 17. Cross wavelet transform of the AET- R_n time series.

Cross wavelet transform of AET and RH exhibited significant common features at the scale band of 2 to 8 hours (Fig. 18). Significant correlation between AET and RH indicates that the RH signal can describe some of the small-scale variations of AET. Although the magnitude of linear correlation between the two time series might not be identified quantitatively, it is seen that RH has a significant cause and effect relationship with the AET signal at small scales. Similar to the AET- R_n cross wavelet spectrum, phase relationship between AET and RH was not stable. However, some anti-phase relationship can be observed at specific time-scale locations. By anti-phase it means that the AET and RH are negatively correlated to each other. Unstable phase relationship, at low scales, also demonstrates the complexity that exists in the short-time variations of AET and its relationship with the involved meteorological factors, which cannot be easily explored by using the cross wavelet transformation.

W_s time series also exhibited significant covariances with AET signal at scales of 2 to 8 hours (Fig. 19), which were more frequent at scales less than 4 hours. All of the significant small-scale features found in individual wavelet transform of the W_s time series were not detected as common features between AET and W_s at 95% confidence level. It indicates that only specific numbers of short-time cyclic variations of W_s are linearly correlated to the small-scale cyclic variations of AET. Overall, the results of cross wavelet analysis of AET and W_s demonstrate the existence of linear correlation at small scales. Phase information of AET- W_s spectrum illustrates both in-phase and anti-phase relationship between the variations of the two analyzed signals.

As it was expected from the individual wavelet spectra of AET and T_g , no specific significant common power was found at small scales in the cross wavelet transform of AET- T_g (Fig. 20). This might be attributed to two possible reasons; first, the presence of non-linear correlation

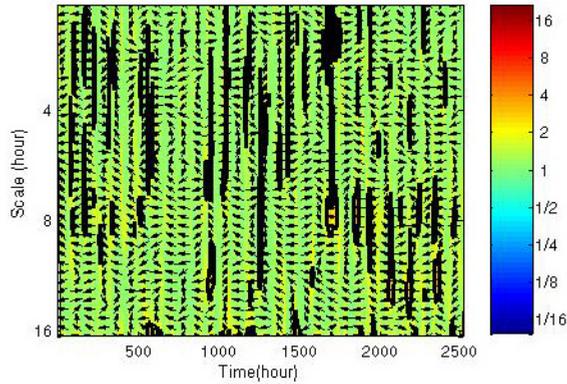


Figure 18. Cross wavelet transform of the AET-RH time series.

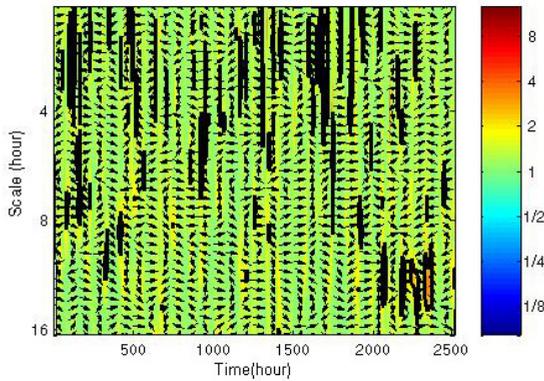


Figure 19. Cross wavelet transform of the AET- W_s time series.

between AET and T_g at small scales, which cannot be identified by the current cross wavelet analysis. Second, for a time series like T_g which is not varying much over short time intervals (e.g. hourly), small-scale cyclic features do not have high powers at scales of 2-8 hours and result in low cross wavelet powers of AET- T_g spectrum that cannot be differentiated from background red noise.

Cross wavelet spectrum of AET- T_a shows limited detected features in the time-scale domain in which the two signals were linearly correlated and the power was significantly different from red noise at 95% confidence level compared to those of AET- R_w , AET-RH, and AET- W_s (Fig. 21). Considering the rare significant peaks detected in the band scales of 2 to 8 hours of T_a univariate power spectrum (Fig. 13), the identified powers in the cross wavelet spectrum

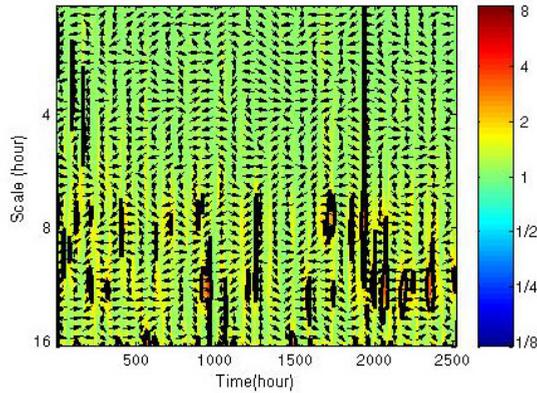


Figure 20. Cross wavelet transform of the AET- T_g time series.

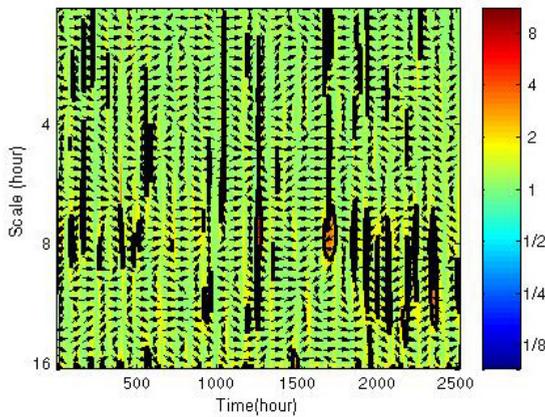


Figure 21. Cross wavelet transform of the AET- T_a time series.

might not indicate significant covariations between the two signals. The significant common powers in the cross wavelet spectrum of AET- T_a were most probably caused by the strong powers of univariate AET spectrum only, which were more frequent than those of T_a over the studied time period. Consequently, no specific and reliable cause and effect relationship can be perceived between AET and T_a time series at small scales. However, strong linear correlation, which was significantly different from background red noise, was observed between AET and T_a at about diurnal scale (scale band of 16 to 32 hours) when the range of studied scales was extended up to 48 hours, Fig. 22.

The results of the cross wavelet analysis determined, to some extent, the meteorological variables that have significant linear correlation with the AET signal at small scales at 95%

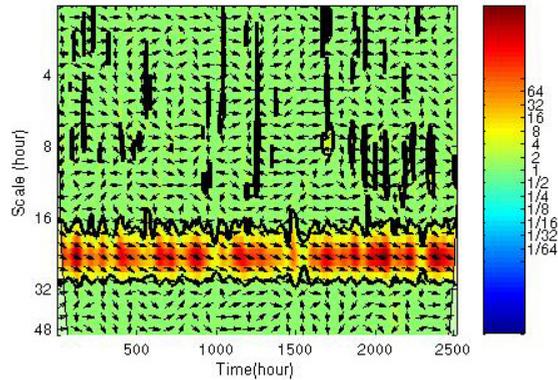


Figure 22. Cross wavelet transform of the AET- T_g time series for the scale range of 2 to 48 hours.

confidence level. Based on the cross wavelet analysis, R_w , RH , and W_s time series exhibited significant correlation with the AET signal and therefore, they are the important variables in the prediction of small-scale AET variations. Based on the values provided on the scales (color bars) of the cross wavelet spectra, R_w exhibited stronger correlation with AET than RH , which is stronger than W_s , with maximum cross wavelet power magnitudes of 256, 16, and 8, respectively. Unfortunately, the results of the cross wavelet analysis cannot be interpreted in a precise quantitative way to identify the importance of one variable over others in the prediction of AET at specific scales (or band scales). In addition, significant powers detected in the univariate wavelet spectra must always be considered when significant cross wavelet powers are interpreted. This is to avoid false common powers, which might be created by the large magnitude powers of one univariate spectrum only.

Based on the cross wavelet analysis, ground temperature has no important linear correlation with the AET time series at small scales. As a result, one might not select T_g as a predictor in the estimation of AET. However, the results of the data driven modeling demonstrated the importance of ground temperature in the prediction of AET. This inconsistency might be attributed to the previously mentioned ability of cross wavelet analysis in identifying only linear correlations between time series. As a result, any existing non-linear correlation between each pair of signals remains undiscovered using cross wavelet analysis. Another possible reason for insignificant common powers in AET- T_g cross wavelet spectrum is that the cross wavelet analysis can investigate the correlation between only two time series at a time (not multiple time series), which ignores the possible effect of other factors interacting with the considered signals. The importance of interaction effects, which exist among the variables involved in the AET process, was observed in the results of the data driven modeling. The above-mentioned limitations of cross wavelet analysis affect the accuracy of the correlation analysis of time series for determining the most important AET predictors.

The other issue, which most probably resulted in the inconsistency between the findings of wavelet analysis and data driven modeling is the time-scale at which the temporal variations

of signals were investigated using wavelet analysis or modeled by data driven techniques. The results of the cross wavelet analysis at the larger-scales might be in agreement with the results of the data driven models, regarding the most effective variables in the prediction of AET. Cross wavelet analysis exhibited strong linear correlation between AET and T_g at about diurnal scale approximately over the whole studied period. As a result, it can be perceived that the proposed data driven models mainly characterized the larger-scale variations of AET (the dominant cyclic patterns) for which R_n and T_g were found to be the most effective predictors. The cross wavelet analysis, conducted in the present study, concerned more about the small-scale variations of AET. The results obtained using the WA at small scales might be useful in the development of AET models that aim to characterize the small-scale temporal variations.

The above-mentioned discussion highlights the importance of the time-scale of temporal variations, which might be of interest to be investigated, analyzed, and modeled. Depending on the specific time-scale of variations one is interested in, the employed modeling technique and/or, for instance, the range of studied scales in the wavelet analysis may vary. As a result, it is important to have better understanding of different types of temporal variation exist in the investigated time series prior to the signal analysis and/or modeling.

5. Conclusion

In conclusion, the investigated data driven modeling techniques were promising for the estimation of the hourly AET mechanism using the observed data, without assuming or applying significant knowledge of the physics of the process. The choice of the testing dataset was found to be important for realistically assessing the generalization ability of the proposed data driven models and also for the determination of the possible superiority of any of the modeling techniques over others. The genetic programming (GP) modeling technique was found to perform similarly and better than the ANN model with regard to generalization ability. The GP-evolved models also had the advantage of being structurally simple and requiring fewer input variables, which is of interest for many hydrological practitioners.

Furthermore, the proposed equation-based GP models showed that the AET process has the potential to be estimated by structurally simple (e.g. semi-linear) models. Equation-based AET models made it possible to extract some information about the physics of the process. It was observed that the meteorological variables of R_n and T_g have larger contribution, than other variables, to the estimation of AET. In addition, the interaction effects of the meteorological variables were found to be important and effective in the estimation of AET.

The results of wavelet analysis improved the understanding of the AET mechanism by revealing the importance and contributions of different time-scale cyclic variations exist in the AET time series. This highlights the issue of time-scale and the importance of its consideration in the modeling and prior-to-modeling input selection procedure. Although several small-scale cyclic features were detected in the AET signal, larger-scale variations were found to be the major frequency events at which the predictant-predictor (AET-meteorological variables)

correlation analysis was more clear and reliable. GP models were noted to mainly model the larger-scale (dominant) temporal variations of AET, although short time-scale (hourly) data were employed for training and developing the models.

Consistency between the results of data driven modeling (especially GP) and wavelet analysis, regarding the most important predictor variables, at large time-scales (e.g. diurnal) indicated that wavelet analysis can be employed as a guide for identifying the most linearly correlated predictors for the modeling of AET. However, limitations of such signal analysis tool should be considered when it is used for input determination prior to modeling. Wavelet analysis helped to perceive the difference between the predictive abilities of various models from a new perspective, which is the time-scale of variations that the models characterize. For instance, when the performances of two prediction models are compared, it should be noted if both models are capturing the same time-scale of variations. Consideration of this point can make the models' comparative analysis more accurate and fair.

Author details

Zohreh Izadifar and Amin Elshorbagy

University of Saskatchewan, Canada

References

- [1] Akaike, H. (1974). A new look at statistical model identification. *IEEE Trans. Automat. Contr.*, AC, , 19, 716-722.
- [2] Allen, M. R, & Smith, L. A. (1996). Monte Carlo SSA: Detecting irregular oscillations in the presence of coloured noise, *J. Clim.*, , 9, 3373-3404.
- [3] Anctil, F, & Tape, D. G. (2004). An exploration of artificial neural network rainfall-runoff forecasting combined with wavelet decomposition. *J. Environ. Eng. Sci.*, 3, SS128., 121.
- [4] ASCE(2000). Artificial Neural Networks in Hydrology. I. Preliminary Concepts. *J. Hydrol. Eng.*, , 5(2), 115-123.
- [5] Baldocchi, D. D, Hicks, B. B, & Meyers, T. P. (1988). Measuring biosphere-atmosphere exchanges of biologically related gases with micrometeorological methods. *Ecology*, 69, 1331-1340.
- [6] Banzhaf, W, Nordin, P, Keller, R. E, & France, F. (1998). *Genetic Programming: An Introduction*. Morgan Kaufmann Publishers, Inc. San Francisco, California, USA.

- [7] Bhakar S.R; Oiha, S.; Singh, R.V. & Ansari, A. ((2006). Estimation of evapotranspiration for wheat crop using artificial neural network, Proceedings of 4th World Congress on Computers in Agriculture, Orlando, FL, US.
- [8] Box, G. E. P. (1957). Evolutionary operation: A method for increasing industrial productivity. *Appl. Statistics*, , 6(2), 81-101.
- [9] Bremermann, H. J. (1962). Optimization Through Evolution and Recombination, In: *Self-Organizing Systems*, M.C. Yovits et al, (Ed.), Spartan Books, Washington, DC., 93-106.
- [10] Cahill, A. T. (2002). Determination of changes in streamflow variance by means of a wavelet-based test. *Water Resour. Res.*, , 38(6), 1065-1078.
- [11] Chen, X, & Liu, D. (2008). Wavelet analysis on inter-annual variation of precipitation in Guangdong, China. *IAHS publication*, , 319, 3-9.
- [12] Cheng, B, & Titterton, D. M. (1994). Neural networks: A review from a statistical perspective. *Statistical Science*, , 9(1), 2-54.
- [13] Coulibaly, P. (2006). Spatial and temporal variability of Canadian seasonal precipitation (1900-2000). *Advances in Water Resources*, , 29, 1846-1865.
- [14] Coulibaly, P, & Burn, D. H. (2004). Wavelet analysis of variability in annual Canadian streamflows. *Water Resource Research*, 40, W03105, doi:10.1029/2003WR002667.
- [15] Cybenko, G. (1989). Approximation by superposition of a sigmoidal function. *Math. Control Signals Syst.*, , 2, 303-314.
- [16] Dai, X, Shi, H, Li, Y, Ouyang, Z, & Huo, Z. (2009). Artificial neural network models for estimating regional reference evapotranspiration based on climate factors. *Hydrological processes*, , 23(3), 442-450.
- [17] Dingman, S. L. (2002). *Physical Hydrology*, Prentice-Hall, Inc., Upper Saddle River, NJ.
- [18] El-Baroudy, I, Elshorbagy, A, Carey, S, Giustolisi, O, & Savic, D. (2009). Comparison of Three Data-driven Techniques in Modelling Evapotranspiration Process. *Journal of Hydroinformatics*, In press.
- [19] Friedberg, R. M. (1958). A learning machine: Part I. *IBM J.*, , 2(1), 2-13.
- [20] Friedberg, R. M, Dunham, B, & North, J. H. (1959). A learning machine: Part II. *IBM J.*, , 3(7), 282-287.
- [21] Grinsted, A, Moor, J. C, & Jevrejeva, S. (2004). Application of the cross wavelet transform and wavelet coherence to geophysical time series. *Nonlinear Processes in Geophysics*, , 11, 561-566.
- [22] Gupta, V. K, & Waymire, E. (1990). Multiscaling properties of spatial rainfall and river flow distribution. *Journal of Geophysical Research*, 95, 1999-2009.

- [23] Hasselmann, K. (1976). Stochastic climate models: part I. theory. *Tellus*, , 28(6), 473-85.
- [24] He, Y, Guo, R, & Si, B. C. (2007). Detecting grassland spatial variation by a wavelet approach. *Int. J. Remote Sens.*, , 28, 1527-1545.
- [25] Hornik, K, Stinchcombe, M, & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, , 2(5), 359-366.
- [26] Izadifar, Z. Modeling and analysis of actual evapotranspiration using data driven and wavelet techniques. M.Sc. thesis, University of Saskatchewan, Saskatoon, Saskatchewan, CA.
- [27] Jain, S. K, Nayak, P. C, & Sudheer, K. P. (2008). Models for estimating evapotranspiration using artificial neural networks, and their physical interpretation. *Hydrological Processes*, , 22(13), 2225-2234.
- [28] Kaheil, Y. H, Rosero, E, Gill, M. K, Mckee, M, & Bastidas, L. A. (2008). Downscaling and forecasting of evapotranspiration using a synthetic model of wavelets and support vector machines. *IEEE Trans. Geosci.Remote Sens.*, , 46(9), 2692-2707.
- [29] Kirkup, H, Pitman, A. J, Hogan, J, & Brierley, G. (2001). An initial analysis of river discharge and rainfall in Coastal New South Wales, Australia, using wavelet transforms. *Australian Geographical Studies*, , 39(3), 313-334.
- [30] Koza, J. R. (1992). *Genetic Programming: On the Programming of Computers by Means of Natural Selection*, MIT Press, Cambridge, MA.
- [31] Kumar, M, Raghuwanshi, N. S, Singh, R, Wallender, W. W, & Pruitt, W. O. (2002). Estimating evapotranspiration using artificial neural network. *Journal of Irrigation and Drainage Engineering*, , 128(4), 224-233.
- [32] Kumar, P, & Foufoula-georgiou, E. (1993a). A multicomponent decomposition of spatial rainfall fields 1. Segregation of large-and small-scale features using wavelet transforms. *Water Resources Research*, 29, 8, 2515-2532.
- [33] Kumar, P, & Foufoula-georgiou, E. (1993b). A multicomponent decomposition of spatial rainfall fields 1. Self-similarity in fluctuations. *Water Resources Research*, 29, 8, 2533-2544.
- [34] Labat, D. (2006). Oscillations in land surface hydrological cycle. *Earth and Planetary Science Letters*, 242, 1-2, 143-154.
- [35] Labat, D. (2008). Wavelet analysis of the annual discharge records of the world's largest rivers. *Adv. Water Resour.*, , 31(1), 109-117.
- [36] Labat, D, Ronchail, J, & Guyot, J. L. (2005). Recent advances in wavelet analyses: Part Amazon, Parana, Orinoco and Congo discharges time scale variability. *J Hydrol.*, 314, 1-4, 289-311., 2.

- [37] Lafreniere, M, & Sharp, M. (2003). Wavelet analysis of inter-annual variability in the runoff regimes of glacial and nival stream catchments, Bow Lake, Alberta. *Hydrological Processes*, , 17(6), 1093-1118.
- [38] Landeras, G, Ortiz-barredo, A, & Lopez, J. J. (2008). Comparison of artificial neural network models and empirical and semi-empirical equations for daily reference evapotranspiration estimation in the Basque Country (Northern Spain). *Agricultural Water Management*, , 95(5), 553-565.
- [39] Levenberg, K. (1944). A method for the solution of certain problems in least squares. *Quart. Appl. Math.* , 2, 164-168.
- [40] MacKay D. ((1992). A practical bayesian framework for backproparation networks. *Neural Computation* , 4(3), 448-472.
- [41] Maier, H. D, & Dandy, G. C. (2000). Neural network for the prediction and forecasting of water resource variables: a review of modeling issues and applications. *Environmental Modelling & Software* , , 15, 101-124.
- [42] Marquardt, D. (1963). An algorithm for least squares estimation of non-linear parameters. *J. Ind. Appl. Math.*, , 11(2), 431-441.
- [43] Marquardt, D. (1963). An algorithm for least squares estimation of non-linear parameters. *J. Ind. Appl. Math.*, , 11(2), 431-441.
- [44] MATLAB software Version [6.5.1]. Copyright © (2009). MathWorks, Inc., 3 Apple Hill Drive, Natick, MA, USA., 1994-2009.
- [45] Miao, C. Y, Wang, Y. F, & Zheng, Y. Z. (2007). Wavelet-based analysis of characteristics of summer rainfall in Nenjiang and Harbin. *Journal of Ecology and Rural Environment* , , 23(4), 29-32.
- [46] Neural Network Toolbox User's Guide(2009). MathWorks, Inc., 3 Apple Hill Drive, Natick, MA, USA.
- [47] Parasuraman, K, & Elshorbagy, A. (2008). Toward improving the reliability of hydrologic prediction: Model structure uncertainty and its quantification using ensemble-based genetic programming framework. *Water Resour. Res.*, 44, W12406, doi: 10.1029/2007WR006451.
- [48] Parasuraman, K, Elshorbagy, A, & Carey, S. K. (2006). Spiking modular neural networks: A neural network modeling approach for hydrological processes. *Water Resour. Res.*, 42, W05412, doi:10.1029/2005WR004317.
- [49] Parasuraman, K, Elshorbagy, A, & Carey, S. K. (2007). Modeling the dynamics of the evapotranspiration process using genetic programming, *Hydrological Sciences Journal*, , 52, 563-578.
- [50] Polikar, P. (1996). The wavelet tutorial, <http://www.site.uottawa.ca/~qingchen/wavelet/htm>, accessed on July 2008.

- [51] Russo, D, Bresler, E, Shani, U, & Parker, J. C. (1991). Analysis of infiltration events in relation to determining soil hydraulic properties by inverse problem, methodology. *Water Resour. Res.*, , 27, 1361-1373.
- [52] Saco, P, & Kumar, P. (2000). Coherent modes in multiscale variability of streamflow over the United States. *Water Resour. Res.*, , 36(4), 1049-1067.
- [53] SAS/STAT software Version [9.1] of the SAS System for Windows. Copyright © SAS Institute Inc. SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc., Cary, NC, USA., 2002-2003.
- [54] Schaeffli, B, Maraun, D, & Holschneider, M. (2007). What drives high flow events in the Swiss Alps? Recent developments in wavelet spectral analysis and their application to hydrology. *Advances in Water Resources*, , 30(12), 2511-2525.
- [55] Sette, S, & Boullart, L. (2001). Genetic programming: principles and applications. *Engineering Application of Artificial Intelligence*, , 14, 727-736.
- [56] Si, B. C. (2003). Spatial and scale-dependent soil hydraulic properties: A wavelet approach. In: *Scaling methods in soil physics*, Ya. Pachepsky et al., (Ed.), CRC Press, New York., 163-178.
- [57] Si, B. C. (2008). Spatial Scaling Analyses of Soil Physical Properties: A Review of Spectral and Wavelet Methods. *Vadose Zone Journal*, , 7(2), 547-562.
- [58] Si, B. C, & Zeleke, T. B. (2005). Wavelet coherency analysis to relate saturated hydraulic properties to soil physical properties. *Water Resource Res.*, 41, 11, W11424. doi:10.1029/2005WR004118.
- [59] Silva, S. (2005). GPLAB-a genetic programming toolbox for MATLAB. <http://gplab.sourceforge.net>.
- [60] Sudheer, K. P, Gosain, A. K, & Ramasastri, K. S. (2003). Estimating actual evapotranspiration from limited climatic data using neural computing technique. *Journal of Irrigation and Drainage Engineering, ASCE*, , 129, 214-218.
- [61] Swingler, K. (1996). *Applying neural networks: A practical guide*, Academic press, London.
- [62] Torrence, C, & Compo, G. P. (1998). A practical guide to wavelet analysis. *Bull. Am. Met. Soc.*, , 79, 61-78.
- [63] Westra, S, & Sharma, A. (2006). Dominant modes of interannual variability in Australia rainfall analyzed using wavelets. *Journal of Geophysical Research-part D: Atmospheres*, 111, 5, D05102, doi:10.1029/2005JD005996.
- [64] Zanetti, S. S, Sousa, E. F, Oliveira, V. P. S, Almeida, F. T, & Bernardo, S. (2007). Estimating evapotranspiration using artificial neural network and minimum climatological data. *J. Irrig. and Drain. Engrg.*, , 133(2), 83-89.