CrossMark

ORIGINAL PAPER

Wavelet-multivariate relevance vector machine hybrid model for forecasting daily evapotranspiration

Roula Bachour · Inga Maslova · Andres M. Ticlavilca · Wynn R. Walker · Mac McKee

Abstract Evapotranspiration (ET) is one of the main components of the hydrological cycle. It is a complex process driven mainly by weather parameters, and as such, is characterized by high non-linearity and non-stationarity. This paper introduces a methodology combining wavelet multiresolution analysis with a machine learning algorithm, the multivariate relevance vector machine (MVRVM), in order to predict 16 days of future daily reference evapotranspiration (ETo). This methodology lays the ground for forecasting the spatial distribution of ET using Landsat satellite imagery, hence the choice of 16 days, which corresponds with the Landsat overpass cycle. An accurate prediction of daily ETo is needed to improve the management of irrigation schedules as well as the operations of water supply facilities like canals and reservoirs. In this paper, various wavelet decompositions were performed and combined with MVRVM to develop hybrid models to predict ETo over a 16-days period. These models were compared to a MVRVM model, and models accuracy and robustness were evaluated. The addition of 10 days of forecasted air temperature as additional inputs to the forecasting models was also investigated. The results of the wavelet-MVRVM hybrid modeling methodology showed that a reliable forecast of ETo up to 16 days ahead is possible.

Keywords Evapotranspiration · Wavelet · Multivariate relevance vector machine · Forecasting

R. Bachour (✉) · W. R. Walker
Civil and Environmental Engineering Department, College of Engineering, Utah State University, Logan, UT 84322-4100, USA
e-mail: bachour.roula@gmail.com

W. R. Walker
e-mail: wynn.walker@usu.edu

I. Maslova
Department of Mathematics and Statistics, American University, Washington, DC, USA
e-mail: maslova@american.edu

A. M. Ticlavilca · M. McKee
Civil and Environmental Engineering Department, Utah Water Research Laboratory, Utah State Univ, 1600 Canyon Rd., Logan, UT 84321, USA
e-mail: andres.t@aggiemail.usu.edu

M. McKee
e-mail: mac.mckee@usu.edu

1 Introduction

Evapotranspiration (ET) is a complex process affected by several environmental factors and driven mainly by weather parameters. Among numerous analytical methods that have been proposed to estimate ET, the Food and Agricultural Organization of the United Nations (FAO) Penman–Monteith model has become a generally accepted standard for calculating reference evapotranspiration (ETo) (Allen et al. 1998, 2006). The relationship between ETo and its driving factors is complicated and not easily modeled (Gao et al. 2012; Partal 2009; Torres et al. 2011; Park and Choi 2014). The non-stationarity nature of ETo time series leads to difficulties in forecasting future values (Pandey et al. 2009; Hernandez et al. 2011; Darshana et al. 2013).

Time series analysis techniques such as the seasonal autoregressive integrated moving average (SARIMA) have been widely used for modeling and predicting different hydrological parameters including ETo (Mariño et al. 1993; Trajkovic 1998; Cigizoglu 2003; Landeras et al. 2009; Gorantiwar et al. 2011). However, difficulties related to these techniques has motivated researchers to look for other modeling approaches including the use of data-driven tools

🖄 Springer

or statistical learning machines, such as artificial neural networks (ANN), multiple regression methods, support vector machines (SVM), and relevance vector machines (RVM). For instance, Landeras et al. (2009) showed that ANN's have good performance for weekly ETo forecasts. Kisi (2007) estimated daily ETo using ANN methods and compared ANN test results to those of the Penman, Hargreaves and Turc empirical models. Although, ANN's have been used extensively as a useful tool for prediction, they have difficulty dealing with non-stationary data (Cannas et al. 2006; Partal 2009; Pulido-Calvo and Gutiérrez-Estrada 2009; El-Shafie et al. 2013). On the other hand, RVM's were used extensively by many researchers for modeling and forecasting hydrological parameters. Unfortunately, these models are limited to one step ahead of forecasting, i.e. single output. When predictions are to be made for multiple steps in time, multiple outputs, the multivariate relevance vector machine (MVRVM) reported by Thayananthan et al. (2008) has proven more effective (Ticlavilca and McKee 2011; Torres et al. 2011; Ticlavilca et al. 2013).

In addition to its ability to predict multiple outputs, the MVRVM is a Bayesian regression tool that has high prediction accuracy, robustness, and estimation of uncertainty in the predictions. The MVRVM is an extension of the RVM algorithm developed by Tipping and Faul (2003) to produce multivariate outputs when given a set of inputs. Therefore, developing a model with all these properties provides a good forecasting tool to produce multiple predictions that are difficult or not practical to obtain from traditional modeling approaches. Thus, the MVRVM algorithm was used in this study to forecast daily ETo values for multiple future time steps.

In the last decade, wavelet transformation has become a useful technique for analyzing variations, periodicities and trends in time series (Chou and Wang 2002; Labat et al. 2005; Küçük et al. 2009). Wavelet transforms, which can produce a good local representation of a signal in both time and frequency domains, provide considerable information about the structure of the physical process to be modeled (Li et al. 1999). Recently, there has been an increased interest in the use of wavelet analyses in a wide range of fields related to water resources (Labat 2005; Mishra et al. 2011). Most of these studies demonstrated that the application of wavelets leads to several improvements in the analysis of global hydrological signal fluctuations and of their mutual time varying relationships. Wavelets provide a formal method to break down a complex time series into simpler units to facilitate accurate prediction (Ahmad et al. 2005; Cobaner 2013) proving to be a more effective technique than the Fourier and windowed Fourier transform for studying non-stationary time series (Labat et al. 2005; Partal and Cigizoglu 2008). However, there have been very few applications of wavelet transform techniques to ET modeling (Kisi 2011; Mishra et al. 2011; Cobaner 2013).

Researchers have developed hybrid models that combine wavelet transforms with time series models and artificial intelligence algorithms. In such approaches, a wavelet transform is used first to decompose the time series into various scales of temporal resolution. Then, a time series model or a regression model is applied. The results of such hybrid models show significant advantages over traditional time series analysis and prediction (Kisi 2011). Partal (2009) modeled ETo using wavelet decomposition and neural networks showing improvement of ETo modeling with the hybrid models. Wang and Luo (2007) combined the wavelet transformation and neural network techniques and were able to develop a wavelet-neural network hybrid model to forecast ETo 1 day ahead. Since the evapotranspiration process is characterized by high non-linearity and non-stationarity (Hernandez et al. 2011), hybrid wavelet models provide a new alternative to the evapotranspiration estimation and forecasting problem.

None of the aforementioned studies have attempted to forecast ETo multiple days ahead using the wavelet hybrid models. This type of forecasts are very important for irrigation management, as it helps the operation of the canal and reservoirs by predicting ahead of time the required amounts of water to be delivered. It can also a beneficial tool for managing the distribution systems especially when ETo is extrapolated spatially. The spatial distribution of ET is usually done using remote sensing algorithms and, for example, Landsat satellite imagery. The latter data set is only available every 16 days over a specific area for a specific Landsat satellite. Hence, forecasting ETo up to 16 days ahead is very crucial for irrigation management and operation of canal delivery and distribution systems. Therefore, the objective of this study was to develop a wavelet-MVRVM model to forecast daily ETo simultaneously, up to 16 days ahead. Wavelet-based decompositions were performed and combined with the MVRVM. The performance and accuracy of these hybrid models were then compared to the performance of a MVRVM model.

The remainder of the paper describes the data used in this study, the wavelet multiresolution analysis (MRA), and the MVRVM learning model in the Methodology Section. The Results and Discussion Section summarizes the selected wavelet-MVRVM hybrid models, and their forecasting results along with a discussion of these results. Finally the conclusions are drawn and future work is discussed.

2 Methodology

2.1 Data collection and description

The weather data for this study were taken from the meteorological station located in Delta, Utah. This station is a

Table 1 Average weather data for delta from 2002 till 2012

Month	T_{max} (°C)	T_{min} (°C)	RH_{min} (%)	RH_{max} (%)	R_s (W m^{-2})	U (m s^{-1})
January	3.5	−7.5	89.5	52.4	58.4	1.03
February	7.0	−4.8	88.9	42.4	93.2	1.24
March	13.2	−0.8	80.0	26.7	140.5	1.64
April	17.3	2.4	76.5	22.9	180.4	1.74
May	23.0	7.2	67.4	19.1	218.3	1.35
June	29.2	12.4	56.2	15.1	242.8	1.13
July	34.3	17.6	53.1	14.8	233.3	0.90
August	32.0	15.7	57.5	16.1	214.4	0.94
September	26.5	9.7	65.5	18.7	170.9	0.80
October	18.2	3.5	77.5	28.9	109.6	0.86
November	10.3	−2.8	82.9	36.7	65.4	1.01
December	3.9	−7.1	89.0	51.5	47.9	1.07
Average	18.2	3.8	73.7	28.8	147.9	1.1

part of the Community Environmental Monitoring Program (CEMP) network of 29 monitoring stations located in the Western states. It is operated and monitored by the Desert Research Institute (DRI) of the Nevada System of Higher Education. The station is located at 39°21′11″N latitude, 112°34′42″W longitude and is 1,415 m above sea level, it is surrounded with a turf grass area of about 1,000 m². It records daily solar radiation, minimum and maximum temperatures, wind speed, precipitation and relative humidity. Daily records over the full period of January 2002 until June 2012 were used in this study. These data were available on the CEMP website (CEMP 2012).

Delta is characterized by a semiarid to arid climate, with an average annual precipitation of 200 mm. The average monthly weather data from 2002 until 2012 (Table 1) showed the coldest month to be January, with a minimum temperature of −7.5 °C, and the hottest month—July, with 34.3 °C as average maximum temperature. The average wind speed throughout the year was 1.14 m s^{-1}, with 148 W m^{-2} average solar radiation per day.

The daily values of reference crop evapotranspiration (ETo) used in this study were calculated from the Delta weather data using FAO 56 Penman–Monteith (PM) equation as suggested by Allen et al. (1998, 2006):

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

where ETo = reference evapotranspiration (mm days^{-1}); Δ = slope vapor pressure curve (kPa °C^{-1}); R_n is the net radiation at the crop surface (MJ m^{-2} days^{-1}); G = soil heat flux density (MJ m^{-2} days^{-1}); γ is the psychrometric constant (kPa °C^{-1}); T = mean daily air temperature at 2 m height (°C); U_2 = wind speed at 2 m height (m s^{-1}); e_s = saturation vapor pressure (kPa); e_a = actual vapor pressure (kPa); $(e_s - e_a)$ = saturation vapor pressure deficit (kPa). It is assumed that Eq. (1) is applied to a

hypothetical crop with a height of 0.12 m, having a surface resistance of 70 s m^{-1} and an albedo $\alpha = 0.23$. Soil heat flux was ignored as suggested by Allen et al. (1998, 2006) for daily time steps. The computations of all required data for calculating ETo were done using the method given in Chapter 3 of FAO paper 56 (Allen et al. 1998, 2006).

The data sample consisted of 11 years (2002–2012) of daily ETo records. The first 7 years (2002–2008) were used for training, 2 years (2009–2010) for calibration, and the remaining data (2011–2012) have been used for unseen test of the models. Only the growing season data (April 1 till October 31) were considered in this study.

It is important to mention that this is a case study of one location. The main purpose of which was to present the potential of hybrid model and the methodology of combining wavelets decomposition with multivariate relevance vector machine in order to improve to forecast several days ahead of evapotranspiration which is crucial of better irrigation management. Further work is being done to validate the results of this methodology to other stations and explore the effects of the use of additional weather parameters as inputs.

2.2 Wavelet multiresolution analysis

Wavelet multiresolution analysis was used in this paper to study the ETo characteristics in time and frequency domains. One of the advantages of wavelet-based techniques is the ability to deal with non-stationary data. It is an alternative to a windowed Fourier transform that requires selection of a window where data are stationary and assumes that time series variability pattern stays the same over time, which is not the case with ETo series analyzed in this paper.

There are two main types of discrete wavelet transforms: orthogonal, usually referred to as discrete wavelet

transform (DWT) and non-orthogonal, also known as maximal overlap discrete wavelet transform (MODWT) (Torrence and Compo 1998).

DWT decomposes the original signal into components at dyadic frequencies. It uses the so-called mother and father wavelets to capture the detailed and smooth parts of a signal. Mother wavelets are used to extract the high-frequency components of the signal, father wavelets—the low-frequency components. The signal is then represented by its features, which are referred to as wavelet and scaling coefficients (Mallat 1989; Daubechies 1992). However, the DWT suffers from a lack of translation invariance. This means that circularly shifting a time series will not necessarily shift its DWT coefficients in a similar manner (Daubechies 1992; Lau and Weng 1995). So, the results of such decomposition depend on the starting point of the series. This transform is also limited to the series of a 2^j length, where $j = 1, 2, 3 \ldots$. This problem is solved by means of a highly redundant non-orthogonal transform called maximal overlap discrete wavelet transform (MODWT) also known as non-decimated DWT. For this transform, an input time series of any length N results in the same number of wavelet and scaling coefficients at each resolution level. Therefore, the features of wavelet coefficients in the wavelet multiresolution analysis (MRA) are aligned with the original time series (Percival and Walden 2000).

In this study, a MODWT-based MRA was used to pre-process ET series and produce the approximately independent components. The latter were used for ET modeling. For MRA decomposition based on MODWT the wavelet filter selection is not critical (see Percival and Walden 2000). Therefore, the filter that produces least artifacts at the beginning and the end of the series was used, i.e. Haar filter. Let $ETo(t)$ denote the daily evapotranspiration computed using the PM equation. We can write it as:

$$ET_o(t) = \sum_{j=1}^{J} D_j(t) + S_J(t) \qquad (2)$$

where $D_j(t) = D_j(1),\ldots, D_j(N)$ are vectors of length N, also known as jth level detail, and $S_J(t) = S_J(1),\ldots, S_J(N)$ is the smooth (or approximation) at level J. The decomposition in Eq. (2) is known as wavelet MRA. The details, $D_j(t)$, correspond to the component at time of approximately 2^j days, and capture a part of the record that corresponds to the frequencies in the range from $2^{-(j+1)}$ to 2^{-j} cycles per day. This range corresponds to physical scales between 2^j and 2^{j+1} days. The smooth, $S_J(t)$, captures the low frequency variations of the time series that correspond to time of approximately 2^J days, or to averages over intervals of 2^J to 2^{J+1} days. For further details see Percival and Walden (2000).

2.3 Multivariate relevance vector machine

A MVRVM model was used in this study to model and forecast ET. The MVRVM is an extension of the RVM developed by Thayananthan et al. (2008) for multiple outputs of the machine learning model, which in this case are the ET forecasts for multiple days ahead. Tipping (2001) introduced the RVM, as a general sparse Bayesian modeling approach for classification and regression. In RVM regression model, the weight of each input is governed by a set of hyperparameters that describe the posterior distribution of these weights. They are estimated iteratively during the machine learning training step. The value of most of the hyperparameters approaches infinity, and the corresponding weights become zero. The remaining non-zero weights are called the relevance vectors. RVMs have good generalization performance and produce a sparse representation of the nonlinear processes involved (Thayananthan et al. 2008).

For developing a MVRVM, a training data set, as input-target vector pairs $\{x_n, y_n\}_{n=1}^N$ is needed, where N is the number of observations, x is an input vector, and $y \in R^M$ is the multiple output vector. The model "learns" the dependence between input and output target with the purpose of making accurate predictions of the target vector y for previously unseen values of x:

$$y = W\Phi(x) + \varepsilon \qquad (3)$$

Here, $\Phi(x) = [1, f(x, x_1), \ldots, f(x, x_N)]^T$ is a set of $N + 1$ vectors of basis functions f, W is the $M \times P$ matrix of weights of these basis functions ($P = N + 1$), and ε is the noise vector assumed to be Gaussian with zero-mean and diagonal covariance matrix $B = diag(\sigma_1^2, \ldots, \sigma_M^2)$. The kernel basis functions, f, considered in this paper were the Gaussian kernel function, $f(x, x_n) = \exp(-r^{-2}||x - x_n||^2)$, Laplace kernel function, $f(x, x_n) = \exp\left(-\left(r^{-2}||x - x_n||^2\right)^{1/2}\right)$ and Cauchy kernel $f(x, x_n) = \exp 1/\left(1 + r^{-2}||x - x_n||^2\right)$, where r is the kernel width parameter. These types of kernels have been used by many authors in hydrology applications (Kaheil et al. 2008; Ticlavilca and McKee 2011; Torres et al. 2011).

Let $y = [\tau_1, \ldots, \tau_m, \ldots, \tau_M]^T$ and $W = [w_1, \ldots, w_m, \ldots, w_M]^T$. The multivariate Gaussian likelihood distribution for the target vector y can be written as:

$$p\left(\{y_n\}_{n=1}^N | W, B\right) = \prod_{n=1}^{N} N(y_n | W\Phi(x_n), B)$$
$$= \prod_{m=1}^{M} N(\tau_m | w_m\Phi, \sigma_m^2) \qquad (4)$$

where $\Phi = [1, \Phi(x_1), \ldots, \Phi(x_n)]$ is the design matrix, and H is the diagonal covariance matrix.

In order to avoid over-fitting in the maximum likelihood estimation of W and σ^2, Tipping (2001) proposed adding a Gaussian prior term for the weights of each basis function. The prior distribution over the weights is shown in Eq. (5):

$$p(W|V) = \prod_{m=1}^{M} \prod_{k=1}^{P} \mathrm{N}(w_{mk}|0, \alpha_k^{-2}) = \prod_{m=1}^{M} \mathrm{N}(w_m|0, V) \quad (5)$$

where $V = diag(\alpha_1^{-2}, \ldots, \alpha_p^{-2})$ with each α_k being an independent hyperparameter that determines the relevance of the associated basis function. This provides the sparsity of the model (Tipping and Faul 2003). w_{mk} is the element in mth row and kth column of the weight matrix W.

The posterior distribution of the model parameters is then given by the combination of the likelihood and prior distributions within Bayes' rule:

$$p(W|\{y\}_{n=1}^{N}, B, V) \propto p(\{y\}_{n=1}^{N}|W, B)p(W|V)$$
$$\propto \prod_{m=1}^{M} \mathrm{N}(w_m|\mu_m, \sum_m) \quad (6)$$

The posterior distribution of the weights is Gaussian $N(\mu_m, \sum_m)$ with mean $\mu_m = \sigma_m^{-2} \sum_m \Phi^T \tau_m$ and covariance $\sum_m = (V + \sigma_m^{-2} \Phi^T \Phi)^{-1}$.

Given the posterior of the weights, an optimal set of hyperparameters V^{opt} can be obtained by maximizing the data likelihood in Eq. (6) (Tipping and Faul 2003; Thayananthan et al. 2008). The data likelihood is marginalized as:

$$p(\{y_n\}_{n=1}^{N}|V, B) = \int p(\{y_n\}_{n=1}^{N}|W, B)p(W|V)dW$$
$$= \prod_{m=1}^{M} \int \mathrm{N}(\tau_m|w_m\Phi, \sigma_m^2)\mathrm{N}(w_m|0, V)$$
$$= \prod_{m=1}^{M} |C_m|^{-1/2} \exp\left(-\frac{1}{2}\tau_m^T C_m^{-1} \tau_m\right)$$
$$(7)$$

where $C_m = \sigma_m^2 I + \Phi V^{-1} \Phi^T$. And the optimal set of hyperparameters $\alpha^{opt} = \{\alpha_k^{opt}\}_{k=1}^{P}$ and noise $(\alpha^{opt})^2 = \{\alpha_m^{opt}\}_{m=1}^{M}$ are obtained using a bottom-up basis function selection approach described by Tipping and Faul (2003). During this optimization process, many elements of α go to infinity setting the corresponding posterior probability of the weight to zero. The few non-zero weights correspond to the so-called "relevance vectors" (RV) that are the sparse core of the RVM model (Tipping and Faul 2003). The optimal parameters are then used to obtain the optimal weight matrix with optimal covariance $\sum^{opt}$ and mean μ^{opt}. The mathematical formulation and optimization procedure of the RVM are discussed in detail in Tipping (2001), and

Tipping and Faul (2003). Also, details of the MVRVM model can be found in Thayananthan et al. (2008).

Given a new input x^*, the predictive distribution for the corresponding target y^* can be computed as (Tipping 2001):

$$p\left(y^*|y, \alpha^{opt}, (\sigma^{opt})^2\right) = \int p\left(y^*|W, (\sigma^{opt})^2\right) \\ p\left(W|y, \alpha^{opt}, (\sigma^{opt})^2\right)dW \quad (8)$$

$$p\left(y^*|y, \alpha^{opt}, (\sigma^{opt})^2\right) = N\left(y^*|y_{mean}^*, (\sigma^*)^2\right) \quad (9)$$

where y_{mean}^* is the predictive mean, and $(\sigma^*)^2 = \left[(\sigma_1^*)^2, \ldots, (\sigma_M^*)^2\right]^T$ is the predictive variance. This predictive variance is the sum of variances of two terms: the noise in the data and the uncertainty in the prediction of the weight parameters (Tipping 2004). The standard deviation σ^* of the predictive distribution was used to estimate the 95 % Bayesian confidence interval as: $y_{mean}^* \pm 1.96\sigma^*$.

2.4 Performance estimation criteria

In this paper, the accuracy of the models has been estimated and model robustness is evaluated using a bootstrap approach. The accuracy was estimated using the root mean square error, RMSE, the coefficient of determination, R^2, and the Nash–Sutcliffe efficiency coefficient, E. The RMSE was calculated as shown in Eq. (10):

$$RMSE = \sqrt{\frac{\sum_{t=1}^{N}(\hat{y}_t - y_t)^2}{N}} \quad (10)$$

The coefficient of determination, R^2, is computed as:

$$R^2 = 1 - \frac{\left[\sum_{t=1}^{N}(y_t - \bar{y})(\hat{y}_t - \bar{\hat{y}})\right]^2}{\sum_{t=1}^{N}(y_t - \bar{y})^2 \sum_{t=1}^{N}(\hat{y}_t - \bar{\hat{y}})^2} \quad (11)$$

where $\hat{y}_t$ = predicted ETo for day t (mm days^{-1}); y_t = calculated ETo using PM equation for day t (mm days^{-1}); $\bar{y}$ = mean of the observed ETo; $\hat{y}$ = mean of the estimated ETo; and N = total number of observations.

The Nash–Sutcliffe efficiency, E, is a normalized statistic that determines a relative magnitude of the residual variance, "noise", compared to the measured data variance, "information" (Nash and Sutcliffe 1970). It is recommended by the ASCE (1993) and Legates and McCabe (1999) as a measure of model performance. It indicates how well the plot of observed versus simulated data fits the 1:1 line and was computed as shown in Eq. (12):

$$E = 1 - \frac{\sum_{t=1}^{N} (y_t - \hat{y})^2}{\sum_{t=1}^{N} (y_t - \bar{y})^2} \tag{12}$$

It must be noted that an efficiency value of 1 ($E = 1$) corresponds to a perfect match between modeled ETo and observed data. An efficiency of 0 ($E = 0$) indicates that the model predictions are as accurate as the mean of the observed data. An efficiency less than zero ($E < 0$) occurs when the observed mean is a better predictor than the model, which indicates unacceptable performance.

In order to select the forecasting model, the main goal was to select the adequate number of inputs, or days to the past, the kernel width and the kernel type. The optimal values of these parameters were selected by trial and error procedure to obtain the best RMSE and E values.

The bootstrap method (Efron and Tibshirani 1998) was used in this study to guarantee good generalization ability and robustness of the machine learning model. The bootstrap data set was created by randomly sampling with replacement from the training data set. In the bootstrap estimation, the selection process was repeated 1,000 times to yield 1,000 bootstrap training data sets, which were treated as independent sets. For each of the bootstrap sets, the MVRVM was retrained and its performance was evaluated by calculating E for the 2 years of unseen test data set. It is important to note that the kernel type, kernel width, and numbers of days of past time series data used as input to the MVRVM stayed the same, but the hyperparameters changed with each bootstrap sample. The bootstrap method provides implicit information on the uncertainty of the estimator (E) evaluated in the model. A more detailed description of the results in detailed in Sect. 3.

3 Results and discussion

A classical time series modeling approach, the seasonal autoregressive integrated moving average (SARIMA) model was first used to forecast ETo. But the results were converging the forecast to the mean value of the ET time series. The Nash–Sutcliffe coefficient of efficiency values for SARIMA models were negative indicating that the observed mean is a better predictor than the model, which is considered as unacceptable performance in hydrological modeling (Legates and McCabe 1999). Using high resolution daily data was also another limitation for the SARIMA model, since the season is 365 days which did not allow the model to converge making it not applicable for real time modeling and forecasting. The results of the SARIMA models are not shown in this paper. In this Sect. 3 we only present the results of the wavelet-based MRA and the MVRVM models.

3.1 Wavelet decomposition selection

The properties of the ETo time series, both physical and statistical, were examined in time and frequency domains. Table 2 shows the power spectrum results of the maximum overlap discrete wavelet transform (MODWT). From this table, it was obvious that the major changes in ETo were the seasonal and annual components (Levels d7 through s8), while the short-term variations had the lowest percentage of ETo changes. Based on the power spectrum in the wavelet domain, a wavelet MRA was performed on the ETo time series. Two wavelet decomposition designs were considered to aid ETo forecasting. In Design 1, three levels of MRA decomposition were used to focus on the short-term changes of ETo. This design, given in (13), decomposed the records into four high frequency changes of ETo and the smooth part that contains the low-frequency approximation. This way the high-frequency part (up to 16 days) of the series gets more attention during the modeling than the low-frequency:

$$ETo_{(1)}(t) = D_1(t) + D_2(t) + D_3(t) + S_3(t) \tag{13}$$

where the MRA decomposition was performed for $J = 3$ levels, using MODWT, Haar filter, and the reflection boundary rule (see Fig. 1). In this model, the highest level of the detail captured the ETo changes on scale of 8–16 days.

The second design of the wavelet-MRA analysis was based on the basic energy balance equation that defines ETo:

$$ETo(t) = R_n(t) + H(t) + G(t) \tag{14}$$

where $G(t)$ = daily heat transfer from/to the ground; $H(t)$ = daily heat transfer to the atmosphere; and $R_n(t)$ = net daily solar radiation. The energy balance concept was used to extract meaningful components for building the forecasting model. Here, $G(t)$ corresponds to a long-

Table 2 Wavelet MODWT-based power distribution by level

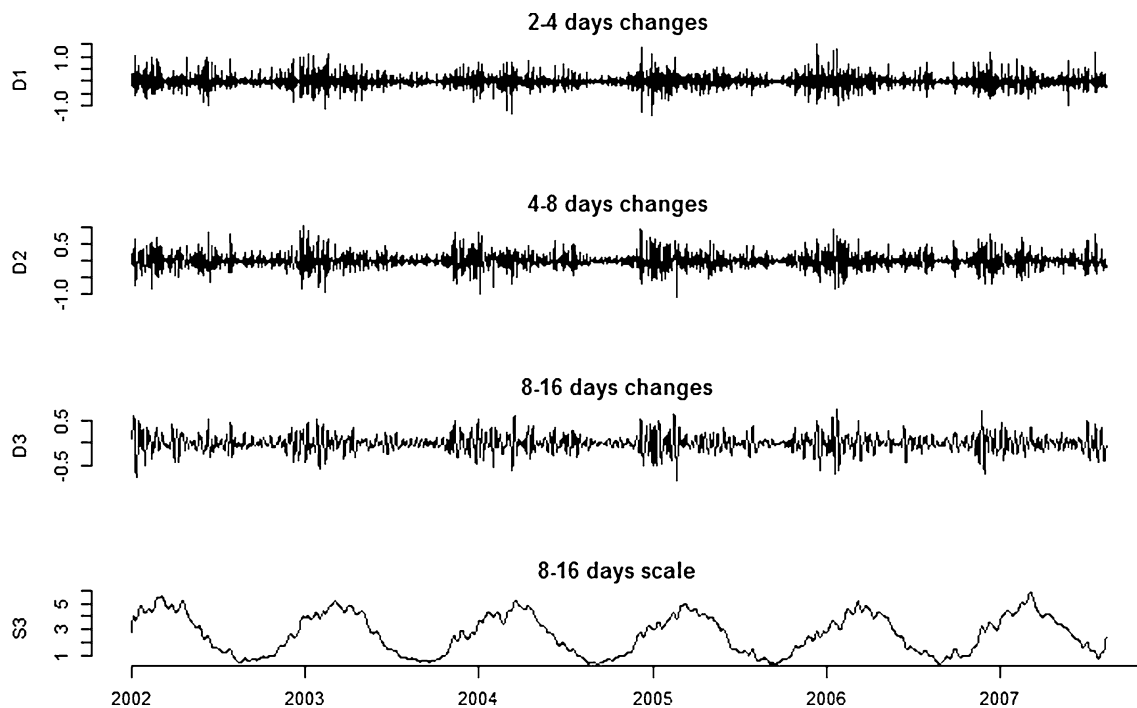| Level	ETo (%)	Physical scale in days (months, years)
d_1	4.70	2–4
d_2	4.41	4–8
d_3	3.28	8–16
d_4	2.44	16–32
d_5	2.92	32–64
d_6	6.89	64–128
d_7	21.06	128–256 (4.3–8.5 m)
d_8	42.38	256–512 (8.5–17.06 m)
s_8	11.84	256–512 (8.5–17.06 m)

Fig. 1 Design 1 of Wavelet-MRA decomposition

term component that capture the variations of the low frequency scales, $H(t)$ contains the short-term variations, which capture the high frequency changes in ETo, and $R_n(t)$ corresponds to the mid-frequency variations. Based on this, Design 2 used a wavelet-based MRA performed for $J = 8$ levels to breakdown the ETo time series into three meaningful components, and was defined as:

$$ETo_{(2)}(t) = D(t) + S(t) + A(t) \qquad (15)$$

where $D = D_1 + D_2 + D_3$, $S = D_4 + D_5 + D_6 + D_7$, $A = D_8 + S_8$. Here, D, S, and A are referred to as the daily (short-term variations), seasonal, and annual (long-term variations) components as shown in Fig. 2. The daily component, D, captured ETo variability from two to 16 days, which accounted for 12.39 % of total variability. The seasonal component, S, captured changes in ETo on the scale from 16 days to 8.5 months, which amounted about 33.31 % of total variability. The annual component, A, captured the changes of ETo on a scale of 8.5–17.06 months which contributed 54.22 % of the total variability.

3.2 Forecasting model selection

Two hybrid models, each consisting of a multivariate relevance vector machine (MVRVM) and a wavelet model developed from one of the wavelet decomposition designs were developed. A total of five different forecasting models

were evaluated (the inputs for each model are specified below):

- *Model 1 (M1)* ETo time series were used as an input for the MVRVM, which served as control a model for comparison,
- *Model 2 (M2)* Each wavelet decomposition component of Design 1 defined in Eq. (13) was used *separately* as input. The model layout produced individual forecasting models for the short-term variations, i.e. D_1, D_2, and D_3, which allowed to better model the high-frequency part of the signal. The rest of the time series was captured by the smooth approximation, S_3. Four MVRVM models were built for each decomposition. For each model the estimated output values, i.e. predictions, were found independently. These outputs were then added to get the estimated ETo,
- *Model 3 (M3)* All the levels of wavelet decompositions of Design 1: D_1, D_2, D_3, and S_3 were used *simultaneously* as multivariate inputs for the MVRVM. On contrary to Model 2, only one model was built to forecast ETo,
- *Model 4 (M4) Separate* MVRVM models were built for daily, seasonal, and annual components of Design 2 as defined in Eq. (15). Three independent models were constructed. The resulting outputs were then added to get forecasted ETo,
- *Model 5 (M5)* All the components of Design 2 were used *simultaneously* as multivariate inputs for the MVRVM to get ETo as output. One multivariate model was built to predict ETo.
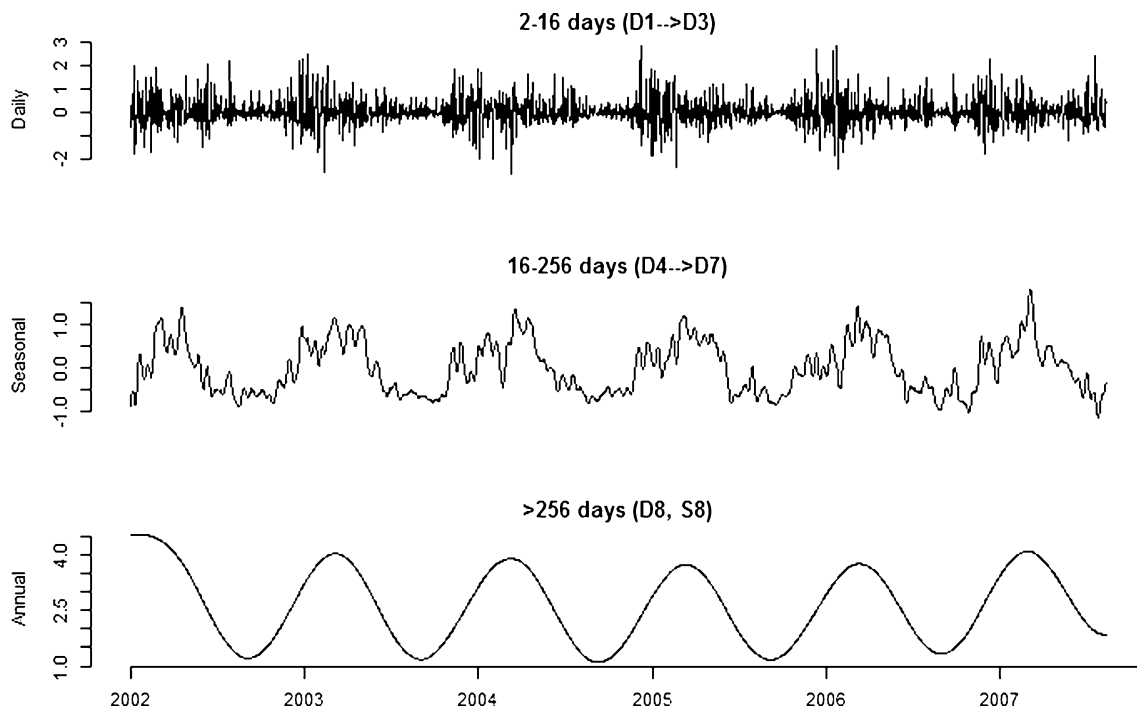
Fig. 2 Design 2 of Wavelet MRA decomposition

Table 3 Models inputs and average statistics of the unseen test data set for the 16 days of forecasted ETo

Model	Input variables	Maximum days to past	Kernel type	Kernel width	Statistic		
					E	R^2	$RMSE$ (mm days^{-1})
M1	ETo	50	Gauss	10	0.561	0.580	0.805
M2	D_1, D_2, D_3, S_3	56	Laplace/Cauchy	14	0.604	0.619	0.766
M3	All Ds together	60	Cauchy	20	0.595	0.604	0.774
M4	D, S, A	70	Gauss/Laplace	34	0.586	0.608	0.872
M5	All DSA together	9	Laplace	17	0.602	0.616	0.767

For the MVRVM application, 7 years of daily ETo were used for training the machine. In this phase, several combinations of kernel type, kernel width, and number of inputs (or days in the past of the time series) were analyzed to get the optimal MVRVM parameters. Although the inputs for these models consist of only one variable (ETo), but the outputs are the forecasts for first, second,…, 16th day ahead. Therefore, the multivariate RVM was used since it allows predicting multiple steps ahead simultaneously. The performance of the model, E and $RMSE$, for the different combinations was evaluated based on a 2 years of calibration phase data. The model that gave the highest E and lowest $RMSE$ for this calibration phase was selected, and then applied to two-year unseen test data.

The optimal kernel width, kernel type, and number of inputs (days to the past) of each model were presented in Table 3. This table also showed the performance (E, R^2, and $RMSE$) of each model for the two-year unseen test data set. These statistics were the average values for each of the forecasted 16 days of ETo for this testing phase. From the statistics presented in Table 3, it is obvious that all wavelet-MVRVM hybrid models outperformed the MVRVM model M1. M2 showed the best performance measured by E, R^2 and $RMSE$ ($E = 0.604$, $R^2 = 0.619$ and $RMSE = 0.766$ mm days^{-1}). The second best model was M5 with $E = 0.602$, $R^2 = 0.616$ and $RMSE = 0.767$ mm days^{-1} with no significant difference from M2. However, it should be noted that M5 required only 9 days of the past records in order to forecast ETo, while this number increased to 56 for M2. Requiring fewer number of inputs may be an advantage of M5 over M2 when dealing with data retrieval and incomplete data. Hence, less computational time is needed which is beneficial for real-

Table 4 Statistics of the models for all the forecasted days

Days ahead	Model														
	M1			M2			M3			M4			M5		
	E	R^2	RMSE (mm)	E	R^2	RMSE (mm)	E	R^2	RMSE (mm)	E	R^2	RMSE (mm)	E	R^2	RMSE (mm)
1	0.595	0.604	0.750	0.651	0.615	0.738	0.637	0.642	0.710	0.599	0.565	0.896	0.652	0.654	0.697
2	0.542	0.553	0.798	0.592	0.599	0.759	0.587	0.594	0.758	0.561	0.572	0.887	0.594	0.598	0.752
3	0.527	0.542	0.814	0.582	0.599	0.767	0.585	0.593	0.764	0.574	0.587	0.877	0.602	0.608	0.748
4	0.542	0.554	0.807	0.578	0.603	0.772	0.575	0.585	0.777	0.574	0.591	0.878	0.601	0.610	0.753
5	0.552	0.565	0.804	0.584	0.615	0.771	0.586	0.600	0.773	0.580	0.596	0.879	0.601	0.609	0.758
6	0.548	0.561	0.811	0.581	0.610	0.779	0.581	0.591	0.781	0.587	0.603	0.873	0.593	0.603	0.769
7	0.546	0.561	0.815	0.581	0.606	0.782	0.577	0.588	0.787	0.586	0.601	0.874	0.584	0.596	0.780
8	0.536	0.557	0.822	0.582	0.606	0.777	0.584	0.593	0.778	0.577	0.599	0.876	0.581	0.595	0.785
9	0.537	0.563	0.818	0.588	0.610	0.767	0.588	0.595	0.772	0.574	0.602	0.869	0.580	0.597	0.784
10	0.554	0.578	0.810	0.596	0.615	0.767	0.580	0.588	0.786	0.580	0.610	0.864	0.592	0.609	0.774
11	0.554	0.582	0.815	0.603	0.619	0.767	0.579	0.588	0.792	0.580	0.614	0.864	0.591	0.610	0.780
12	0.558	0.587	0.819	0.610	0.624	0.768	0.588	0.597	0.791	0.586	0.622	0.864	0.593	0.612	0.785
13	0.577	0.602	0.809	0.625	0.636	0.762	0.606	0.614	0.781	0.596	0.630	0.864	0.609	0.627	0.777
14	0.594	0.615	0.799	0.635	0.644	0.760	0.617	0.625	0.776	0.604	0.639	0.862	0.616	0.635	0.776
15	0.605	0.625	0.793	0.639	0.648	0.759	0.619	0.628	0.778	0.609	0.644	0.861	0.620	0.640	0.777
16	0.613	0.633	0.791	0.644	0.653	0.760	0.628	0.636	0.775	0.615	0.651	0.860	0.625	0.646	0.777
Average	0.561	0.58	0.805	0.604	0.619	0.766	0.595	0.604	0.774	0.586	0.608	0.872	0.602	0.616	0.767

Fig. 3 *Left* Predicted ETo versus observed ETo time series for two-years of unseen test data for Model 2 for selected days. *Right* Plots of predicted ETo versus observed ETo for the same time period
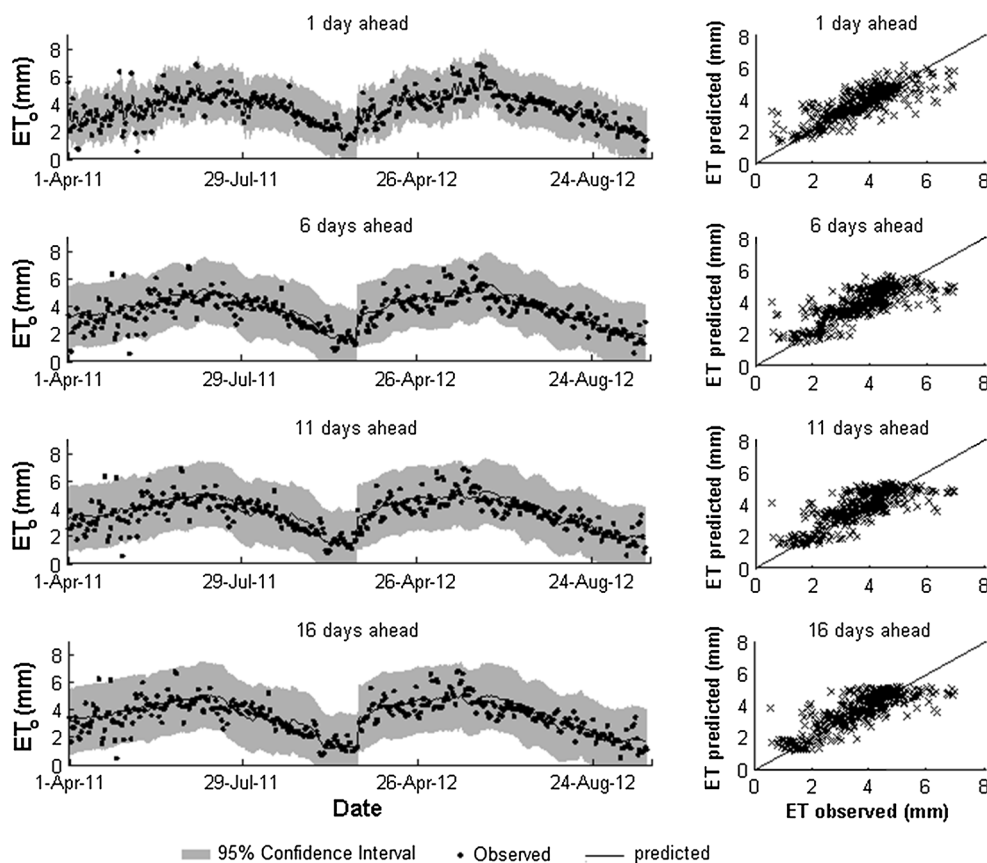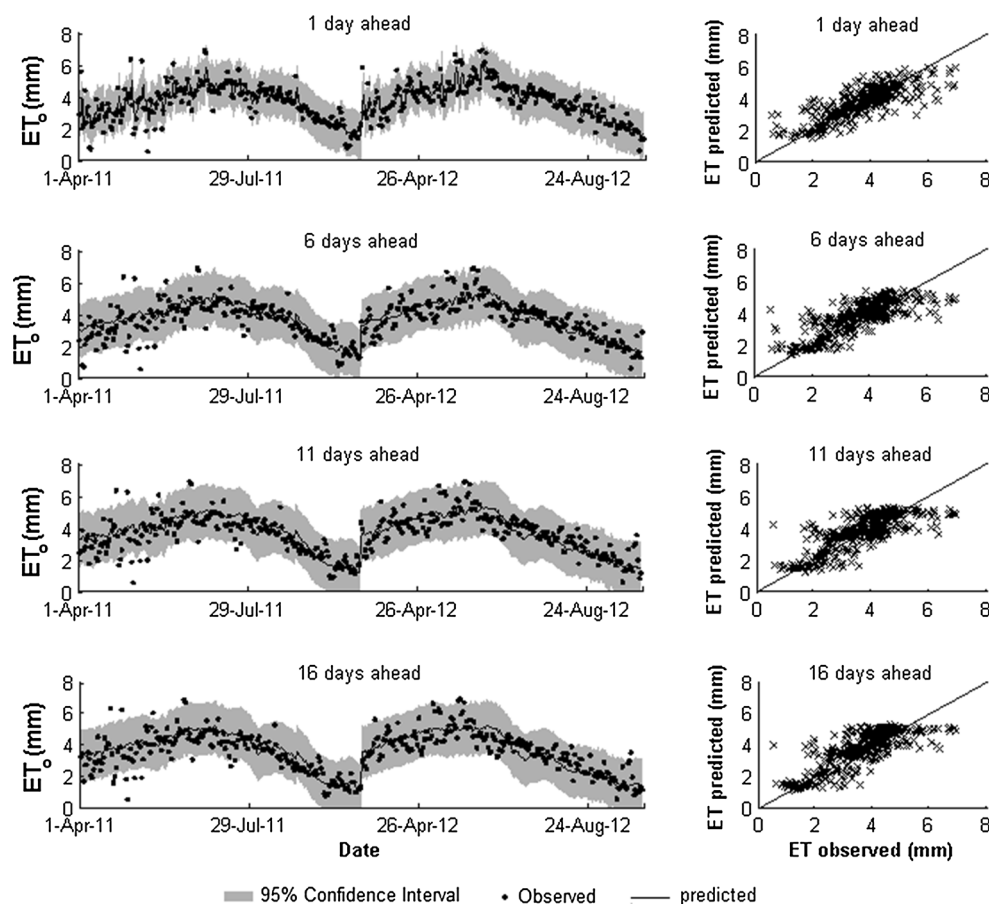
Fig. 4 *Left* Predicted ETo versus observed ETo time series for two-years of unseen test data for Model 5 for selected days. *Right* Plots of predicted ETo versus observed ETo for the same time period

time applications of the model. The overall performance of these hybrid models showed their ability to forecast daily ETo for 16 days simultaneously. This is crucial for reliable irrigation systems management.

Table 4 included the statistics for 16 forecasted days of all the models. It indicated that in some cases, e.g. M2, *E* was higher when forecasting the16th day ahead (0.644) as compared to forecasting the 7th day ahead (0.581). One would expect the efficiency to decline as we increased the forecasting horizon. This was not the case here due to the selection of the parameters by the model that minimizes the average of *E* over all the outputs. Another reason is that the MVRVM forecasting is not iterative as it is in the time series regression models. The prediction errors do not accumulate.

Figures 3 and 4 showed the plots of the results of the best two hybrid models, M2 and M5, respectively. Four days (1st, 6th, 11th and 16th days ahead) were selected to display the results. The left panels in these figures provided a graph of the observed ETo time series calculated by PM equation (points) and predicted ETo by the hybrid models (line), and the 95 % confidence bounds for the 2 years of unseen test data set (2011–2012). The shaded areas in the left panels of these figures represented a 95 % confidence

interval of the forecasting models. M2 gave wider confidence interval bounds, hence the error over the prediction in M5 was smaller. This was due to the design of the wavelet MRA decompositions. M2 was decomposed using Design 1 (described in previous section) which focused on the short-time variations. On the other hand, M5 was decomposed using Design 2 in which the annual and seasonal components (D and S) presented cyclic trends. Therefore, M5 seemed to capture the variation in the ETo time series better than M2. In the right panels of these figures, predicted ETo was plotted against observed ETo. These plots showed that the models underestimated ETo values when observed ETo was greater than about 6 mm days^{-1}. The underestimation was more pronounced when forecasting for more than 1 day ahead. This tendency has also been reported by Cobaner (2013) who used wavelet regression techniques for estimating ETo. Since the MVRVM models used are regression models, they tend to underestimate larger values and overestimate low values. This was clear in those graphs, as the model was trying to regress towards the mean.

Another comparison was done to compare M2 and M5 forecasts for the unseen data with the historical average of the ETo time series in the area of study (Fig. 5). The historical

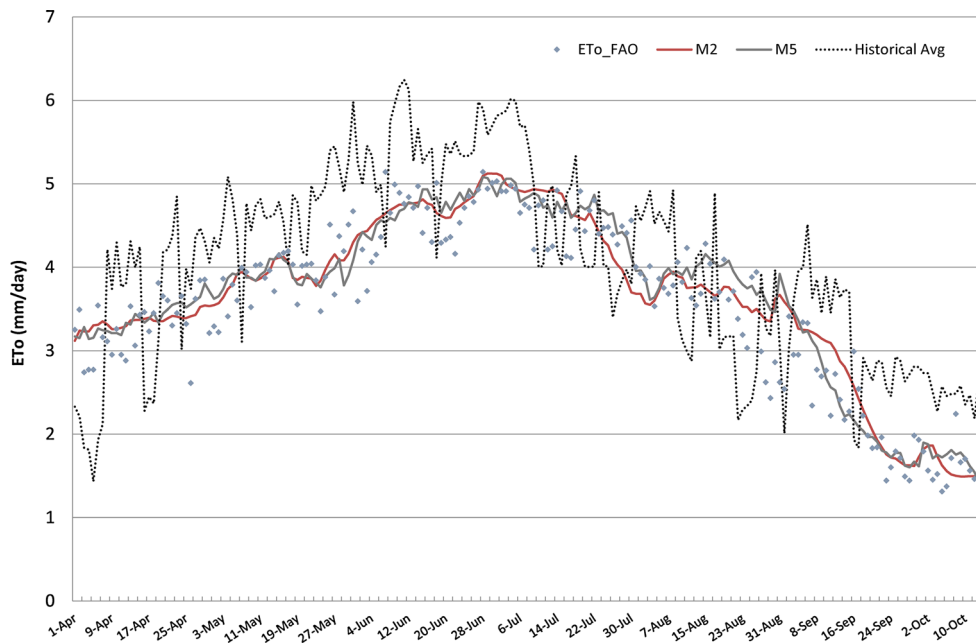**Fig. 5** Comparison of the selected models (M2 and M5) to the historical average

Table 5 Statistical results of M2, M5 and 9-years average as compared to the observed data for the 2 years of unseen test data set

Model	E	RMSE (mm/day)
M2	0.604	0.766
M5	0.602	0.767
9-years Avg	0.408	1.162

average was calculated for 9 years of the available daily data, and had an RMSE of 1.16 mm/day compared to 0.76 mm/day for the wavelet hybrid models M2 and M5 compared to the observed ETo calculated using FAO-PM method (Table 5) presenting an advantage of those hybrid models over traditional application for forecasting daily ETo.

3.3 Model robustness

Bootstrapping was performed to check for over-fitting and model generalization capability of the best two wavelet-MVRVM hybrid models, M2 and M5, and for M1 for comparison. Figure 6 showed a boxplot of the results for 1,000 bootstraps samples for which the Nash–Sutcliffe coefficient of efficiency, E, was computed. No assumptions were made about the distribution of the data. Repeated samples were drawn from the population with replacement. Since this type of samples are good approximation of the population, the bootstrap method provides a good approximation of the sampling distribution of the statistic E. For displaying purposes, only the results for 4 days (1st, 6th, 11th and 16th days ahead) were presented. On each

box, the central mark was the median, the edges of the box were the 25th and 75th percentiles, the whiskers extended to the most extreme data points not considered outliers, and outliers were plotted individually. M5 showed a higher E for all the selected days, with no significant difference from M1 and M2. From these boxplots, it was noticed that the wavelet-MVRVM hybrid models M2 and M5 did not lose the robustness while improving the accuracy of the forecasts. The boxplots confirmed that the models were robust and can be used as ETo forecasting models for real-time applications. The real-time application of ETo forecasting models allows farmers to estimate the water demand for their fields and place the water orders. It will also allow the canal and reservoir operators to release the required amount of water needed for the command area.

3.4 Forecasting ETo using 10-days of forecasted temperature

In an attempt to improve the forecasting potential of the ETo models, including additional model inputs was considered. The National Oceanic and Atmospheric Administration (NOAA) provides daily forecast of minimum and maximum temperatures (T_{min} and T_{max}) up to 10 days ahead. In order to determine whether or not adding these data to the hybrid models that performed best (i.e. M2 and M5) would improve the forecasts, the M2 and M5 analyses were repeated using additional temperature inputs.

Table 6 showed the results for the modified models M1*, M2* and M5*. It is clear from this table that adding the forecasted T_{min} and T_{max} improved the performance of

Fig. 6 Boxplots for the results of bootstrapping analysis (1,000 times) for models M1, M2 and M5 for selected days
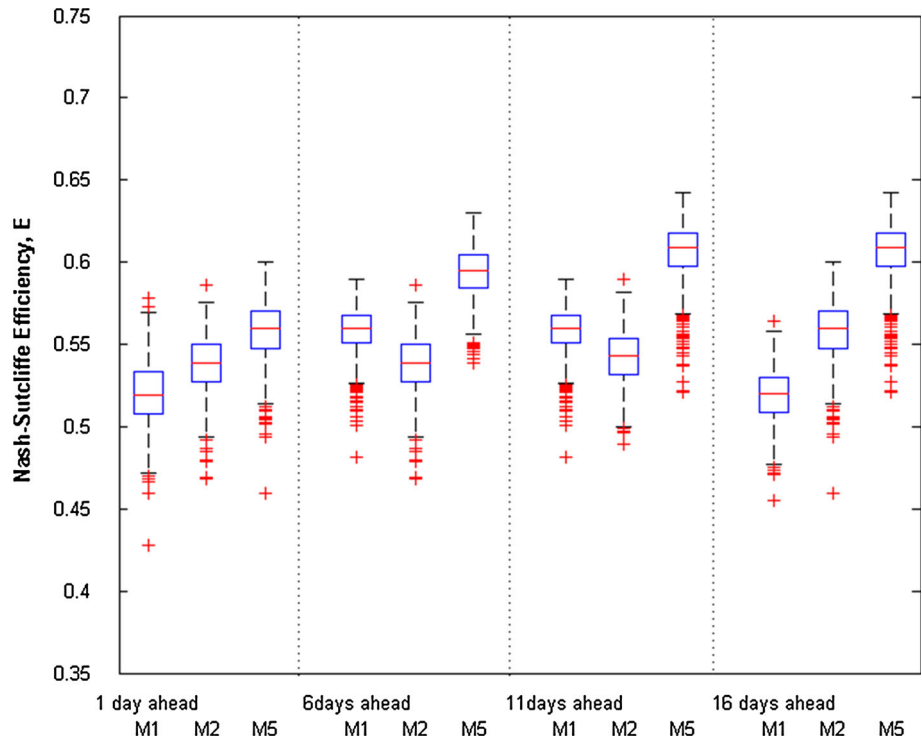

Table 6 Statistics of the models for the 16 days with the additional 10-days forecasted T_{min} and T_{max} as input

Days ahead	Model								
	M1*			M2*			M5*		
	E	R^2	RMSE (mm)	E	R^2	RMSE (mm)	E	R^2	RMSE (mm)
1	0.694	0.703	0.652	0.729	0.771	0.759	0.752	0.772	0.588
2	0.658	0.666	0.690	0.722	0.737	0.716	0.722	0.752	0.622
3	0.630	0.644	0.721	0.729	0.732	0.705	0.711	0.751	0.637
4	0.615	0.637	0.740	0.731	0.733	0.720	0.717	0.756	0.634
5	0.613	0.636	0.747	0.728	0.733	0.720	0.715	0.757	0.641
6	0.606	0.619	0.757	0.727	0.736	0.725	0.705	0.751	0.654
7	0.594	0.606	0.771	0.726	0.739	0.729	0.683	0.738	0.681
8	0.589	0.606	0.774	0.722	0.739	0.731	0.683	0.745	0.683
9	0.581	0.603	0.779	0.724	0.740	0.717	0.653	0.724	0.712
10	0.611	0.628	0.756	0.715	0.731	0.697	0.643	0.706	0.724
11	0.580	0.587	0.791	0.685	0.700	0.718	0.619	0.672	0.753
12	0.527	0.533	0.848	0.640	0.655	0.745	0.574	0.629	0.803
13	0.537	0.541	0.846	0.628	0.641	0.756	0.568	0.628	0.817
14	0.543	0.548	0.847	0.632	0.644	0.757	0.572	0.635	0.819
15	0.547	0.553	0.848	0.643	0.655	0.751	0.583	0.643	0.814
16	0.547	0.552	0.856	0.642	0.656	0.759	0.577	0.638	0.825
Average	0.592	0.604	0.776	0.695	0.709	0.732	0.655	0.706	0.713

the models, especially for forecasting the first 10 days for which the predicted temperature data were available. It is important to note that since the historical records for forecasted minimum and maximum temperature of NOAA were not available, the recorded temperatures of the 10

'forecasted' days were used instead. On average, the modified wavelet-MVRVM hybrid models M2* and M5*, respectively, outperformed again the MVRVM (M1*) model by 17.4 % and 10.6 % for E. Figure 7 showed the scatter plots of the three modified models: M1*, M2* and

Fig. 7 Scatter plots of forecasted ETo using modified models: M1* (**a**), M2* (**b**) and M5*(**c**) compared to observed ETo
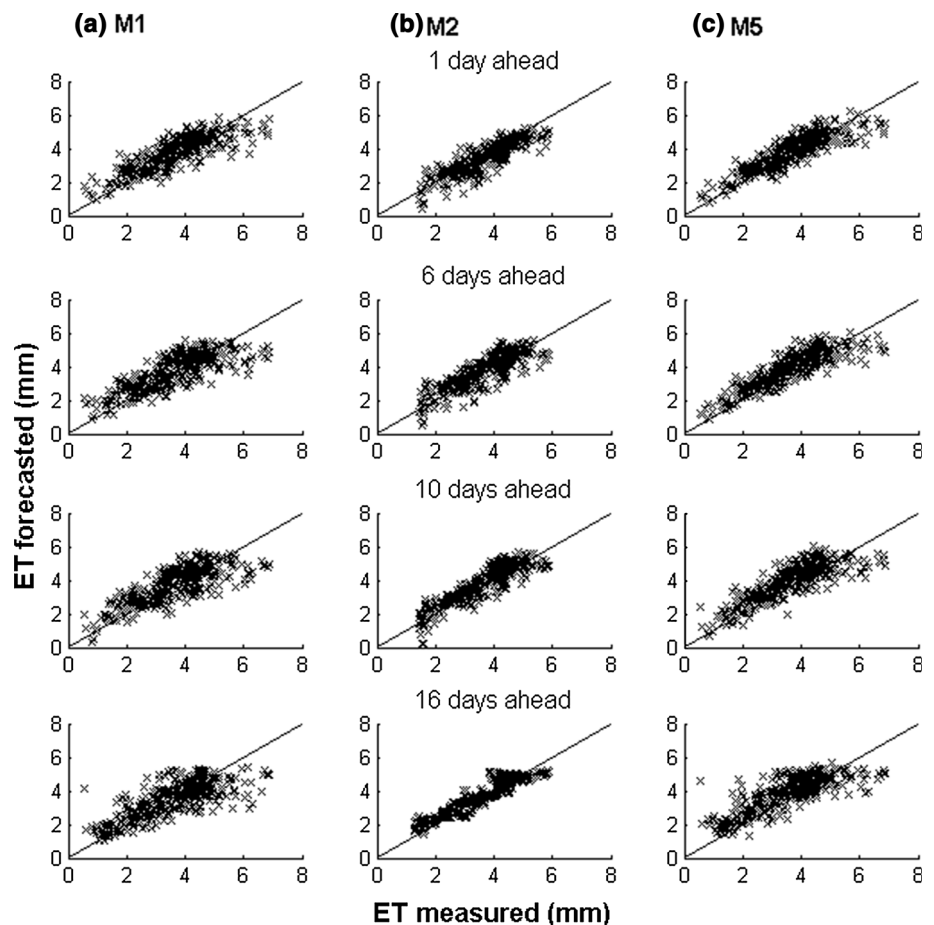
M5*. M2* showed improvement with underestimation of high ETo values. The MVRVM uses Gaussian distribution which is not designed for capturing extremes. Hence, the tendency to underestimate higher values of ETo and overestimate lower values. The introduction of wavelet decomposition helps overcoming this problem. M2* also showed that the forecasts were more clustered around the 1:1 line. This highlights the possible improvement of such hybrid models by adding information about the forecasted future temperatures.

4 Conclusions and future work

In practice, predicting daily ETo is difficult because it is characterized by high non-linearity and non-stationarity. Therefore, models that use components at different temporal resolutions provide a new alternative to ETo forecasting problems. The potential of wavelet-MVRVM hybrid modeling for forecasting daily ETo up to 16 days in advance was investigated in this paper. The study introduced the methodology of decomposing an ETo time series

using wavelet multiresolution analysis methods, and combining the wavelet decompositions with a MVRVM to develop an ETo forecasting model. The resulting models predicted daily ETo up to 16 days ahead with good accuracy. Model that combined a wavelet daily, seasonal and annual multiresolution analysis components with MVRVM, performed the best in this study. It is recommended as a forecasting model for ETo in the study area. The results showed that the wavelet-MVRVM hybrid models performed better than the MVRVM. The bootstrapping analysis showed robustness in forecasting of the wavelet-MVRVM hybrid models.

Inclusion of 10-days of forecasted minimum and maximum air temperatures as additional inputs to the models improved the performance for the first 10 days for which the weather forecast is generally available. Further research might be of interest to explore the potential of adding other weather variables, and their forecasts to the input vector.

The methodology presented in this study lays the grounds for further investigations and studies that could lead to forecast ETo at the spatial level using remote sensing algorithms and the Landsat imagery.

References

Ahmad S, Popoola A, Ahmad K (2005) Wavelet-based multiresolution forecasting. UniS Technical Report, June 2005

Allen RG, Pereira LS, Raes D, Smith M (1998) Crop evapotranspiration: guidelines for computing crop water requirements. FAO Irrigation and Drainage Paper No. 56, FAO, Rome, 300

Allen RG, Pruitt WO, Wright JL, Howell TA, Ventura F, Snyder R, Itenfisu D, Stedudo P, Berengena J, Yrisarry JB, Smith M, Raes D, Perrier A, Alves I, Walter I, Elliot R (2006) A recommendation on standardized surface resistance for hourly calculation of reference ETo by the FAO 56 Penman-Monteith method. Agric Water Manag 81:1–22

ASCE Task Committee on Definition of Criteria for Evaluation of Watershed Models of the Watershed Management, Irrigation, Drainage Division (ASCE) (1993) Criteria for evaluation of watershed models. J Irr Drain Eng 119(3):429–442

Cannas B, Fanni A, See L, Sias G (2006) Data preprocessing for river flow forecasting using neural networks: wavelet transforms and data partitioning. Phy Chem Earth 31(18):1164–1171

CEMP- Community Environmental Monitoring Program website, http://www.cemp.dri.edu. Accessed Dec 2012

Chou CM, Wang RY (2002) On-line estimation of unit hydrographs using the wavelet- based LMS algorithm. Hydrol Sci J 47(5): 721–738

Cigizoglu HK (2003) Incorporation of ARMA models into flow forecasting by artificial neural networks. Environmetrics 14(4): 417–427

Cobaner M (2013) Reference evapotranspiration based on Class A pan evaporation via wavelet regression technique. Irrig Sci 31(2):119–134

Darshana, Pandey A, Pandey RP (2013) Analysing trends in reference evapotranspiration and weather variables in the Tons River Basin in Central India. Stoch Environ Res Risk Assess 27(6): 1407–1421

Daubechies I (1992) Ten Lectures on Wavelets. SIAM, Philadelphia

Efron B, Tibshirani R (1998) An introduction of the bootstrap, monographs on statistics and applied probability, vol 57. CRC Press LLC, Boca Raton

El-Shafie A, Alsulami AM, Jahanbani H, Najah A (2013) Multi-lead ahead prediction model of reference evapotranspiration utilizing ANN with ensemble procedure. Stoch Environ Res Risk Assess 27:1423–1440. doi:10.1007/s00477-012-0678-6

Gao G, Chong-Yu X, Chen D, Singh VP (2012) Spatial and temporal characteristics of actual evapotranspiration over Haihe River basin in China. Stoch Environ Res Risk Assess 26:655–669

Gorantiwar SD, Meshram DT, Mittal HK (2011) Seasonal ARIMA model for generation and forecasting evapotranspiration of Solapur district of Maharashtra. J Agrometeorol 13(2):119–122

Hernandez S, Morales L, Sallis P (2011) Estimation of reference evapotranspiration using limited climatic data and Bayesian model averaging. In: 5th European symposium on computer modeling and simulation, 2011, pp 59–63

Kaheil YH, Rosero E, Gill MK, McKee M, Bastidas LA (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

Kisi O (2007) Evapotranspiration modelling from climatic data using a neural computing technique. Hydrol Process 21:1925–1934

Kisi O (2011) Evapotranspiration modeling using a wavelet regression model. Irrig Sci 29:241–252

Küçük M, Kahya E, Cengiz TM, Karaca M (2009) North Atlantic oscillation influences on Turkish lake levels. Hydrol Process 23(6):893–906

Labat D (2005) Recent advances in wavelet analyses: part 1. A review of concepts. J Hydrol 314(1–4):275–288

Labat D, Ronchail J, Guyot JL (2005) Recent advances in wavelet analyses: part 2-Amazon, Parana, Orinoco and Congo discharges time scale variability. J Hydrol 314(1–4):289–311

Landeras G, Ortiz-Barredo A, Lopez JJ (2009) Forecasting weekly evapotranspiration with ARIMA and artificial neural network models. J Irrig Drain Eng ASCE 135(3):323–334

Lau KM, Weng H (1995) Climate signal detection using wavelet transform: how to make a time series sing. Bull Am Meteorolo Soc 76:2391–2402

Legates DR, McCabe GJ (1999) Evaluating the use of "goodness-of-fit" measures in hydrological and hydroclimatic model validation. Water Resour Res 35(1):233–241

Li X, Ding J, Li H (1999) Combing neural network models based on wavelet transform. J Hydraul 2:1–4

Mallat SG (1989) A theory for multi resolution signal decomposition: the wavelet representation. IEEE Trans Pattern Anal Mach Intell 11(7):674–693

Mariño MA, Tracy JC, Taghavi SA (1993) Forecasting of reference crop evapotranspiration. Agric Water Manag 24:163–187

Mishra AK, Özger M, Singh VP (2011) Wet and dry spell analysis of Global Climate Model-generated precipitation using power laws and wavelet transforms. Stoch Environ Res Risk Assess 25(4):517–535

Nash JE, Sutcliffe JV (1970) River flow forecasting through conceptual models. I. J Hydrol 10:282–290

Pandey PK, Pandey V, Singh R, Bhakar SR (2009) Stochastic modelling of actual Black Gram evapotranspiration. J Water Resour Prot 1:448–455

Park J, Choi M (2014) Estimation of evapotranspiration from ground-based meteorological data and global land data assimilation system (GLDAS). Stoch Environ Res Risk Assess. doi:10.1007/s00477-014-1004-2

Partal T (2009) Modelling evapotranspiration using discrete wavelet transform and neural networks. Hydrol Process 23(25):3545–3555

Partal T, Cigizoglu HK (2008) Estimation and forecasting of daily suspended sediment data using wavelet-neural networks. J Hydrol 358:317–331

Percival DB, Walden AT (2000) Wavelet methods for time series analysis. Cambridge University Press, Cambridge

Pulido-Calvo I, Gutiérrez-Estrada JC (2009) Improved irrigation water demand forecasting using a soft-computing hybrid model. Biosyst Eng 102:202–218

Thayananthan A, Navaratnam R, Stenger B, Torr PH, Cipolla R (2008) Pose estimation and tracking using multivariate regression. Pattern Recognit Lett 29(9):1302–1310

Ticlavilca AM, McKee M (2011) Multivariate bayesian regression approach to forecast releases from a system of multiple reservoirs. Water Resour Manag 25:523–543

Ticlavilca AM, McKee M, Walker WR (2013) Real-time forecasting of short-term irrigation canal demands using a robust multivariate Bayesian learning model. Irrig Sci 31(2):151–167

Tipping ME (2001) Sparse Bayesian learning and the relevance vector machine. J Mach Learn 1:211–244

Tipping ME (2004) Bayesian inference: an introduction to principles and practice in machine learning. In: Bousquet O, von Luxburg U, Ratsch G (eds) Advanced lectures on machine learning. Springer, Berlin, pp 41–62

Tipping ME, Faul AC (2003) Fast marginal likelihood maximization for sparse Bayesian models. In Proceedings of the ninth international workshop on artificial intelligence and statistics, vol 1(3)

Torrence C, Compo GP (1998) A practical guide to wavelet analysis. Bull Am Meteorol Soc 79:61–78

Torres AF, Walker WR, McKee M (2011) Forecasting daily potential evapotranspiration using machine learning and limited climatic data. Agric Water Manag 98(4):553–562

Trajkovic S (1998) Comparison of prediction models of reference crop evapotranspiration. Facta Universitatis Ser Archit Civil Eng 1(5):617–626

Wang WG, Luo YF (2007) Wavelet network model for reference crop evapotranspiration forecasting. In: Wavelet analysis and pattern recognition (ICWAPR '07), vol 2, pp 751–755