Irrig Sci (2013) 31:151–167
DOI 10.1007/s00271-011-0300-6

ORIGINAL PAPER

Real-time forecasting of short-term irrigation canal demands using a robust multivariate Bayesian learning model

Andres M. Ticlavilca · Mac McKee · Wynn R. Walker

Received: 5 February 2010 / Accepted: 1 August 2011 / Published online: 13 August 2011

Abstract In the lower Sevier River basin in Utah, the travel times between reservoir releases and arrival at irrigation canal diversions limit the reservoir operation in enabling delivery changes, which may not be compatible with the on demand schedule in the basin. This research presents a robust machine learning approach to forecast the short-term diversion demands for three irrigation canals. These real-time predictions can assist the operator to react promptly to short-term changes in demand and to properly release water from the reservoir. The models are developed in the form of a multivariate relevance vector machine (MVRVM) that is based on a Bayesian learning machine approach for regression. Predictive confidence intervals can also be obtained from the model with this Bayesian approach. Test results show that the MVRVM learns the input–output patterns with good accuracy. A bootstrap analysis is used to evaluate robustness of model parameter estimation. The MVRVM is compared in terms of performance and robustness with an Artificial Neural Network.

Communicated by J. Kijne.

A. M. Ticlavilca (✉) · M. McKee
Utah Water Research Laboratory,
Department of Civil and Environmental Engineering,
Utah State University, 8200 Old Main Hill,
Logan, UT 84322-8200, USA
e-mail: andres.t@aggiemail.usu.edu

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

W. R. Walker
Department of Civil and Environmental Engineering,
College of Engineering, Utah State University,
Logan, UT 84322-4100, USA
e-mail: wynn.walker@usu.edu

Introduction

Many if not most water resource systems in the Western US and worldwide have irrigated agriculture as their principal client. Efforts to modernize these systems and improve their water-use efficiency have focused on ways to optimize the supply to the farm. Where water must be released from a reservoir to a canal diversion and subsequently conveyed to a farm turnout, there is widespread use of scheduled or arranged demand water delivery. This is a bottom-up strategy that differs substantially from the top-down strategy historically used and exemplified by rotational water deliveries. Supervisory control and data acquisition systems (SCADA) are now widely used to facilitate the scheduled demand operation in real time. The SCADA systems allow hourly, or less, regulation of flow and water levels, whereas historically such regulation occurred as little as every 4–8 h. However, the lag between a flow release and its arrival at a diversion point still limits operator responses to short-term changes in demand. The purpose of this paper is to demonstrate a machine learning approach to forecast short-term irrigation demands and thus at least partially remedy the lag problem for the operator.

Many modeling techniques based on physical modeling have been developed to characterize the current and future states of irrigation and water resources systems. Most of the time, their practical applications are limited by the lack of required data and the expense of data acquisition. To overcome these limitations, researchers have used data-driven modeling based on machine learning approach as an alternative to physically based models (Pulido-Calvo and Gutierrez-Estrada 2008; Khalil et al. 2005a). Examples include artificial neural networks (ANNs), support vector machines (SVMs), and relevance vector machines (RVMs).

These are characterized by their ability to replicate the expected response of a system. They can be used to provide predictions of the system behavior using only historical data (Khalil et al. 2005b).

Operation of irrigation canals is based on water requirements for crop irrigation, physical characteristics of the irrigation system, irrigation scheduling (on-demand, rotational or continuous), travel times (from the diversion point to the field outlets), the agricultural area served (hydrologic, climatic, environmental, etc.) and human behavior, including that of the canal operator and the farmer. The canal operator must divert water from the river into a canal to fulfill farmers' requirements. The combination of all of these behaviors and factors may cause unexpected future changes in the irrigation canal system operation, making predictions difficult. Therefore, it is necessary to develop machine learning models that not only have the ability to learn the input–output patterns and make accurate predictions but also guarantee robustness toward future changes in the system behavior.

Machine learning models have often been applied in irrigation topics. Pulido-Calvo and Gutierrez-Estrada (2008) applied a hybrid model consisting of combined feed-forward computational neural networks (CNNs), fuzzy logic, and a genetic algorithm for forecasting daily irrigation water demand. Their performance results showed an accurate model. Tabari et al. (2010) compared ANN models with multivariate non-linear regression methods in order to estimate daily pan evaporation. Their results demonstrated that the ANN model presents the best estimates. Kumar et al. (2011) presented a review of several contributions on ANN modeling in evapotranspiration estimation.

RVM modeling, which is based on Bayesian machine learning theory, has been used in water resources applications. Khalil et al. (2005b) applied a RVM model to predict the real-time operation of a single reservoir. Their results demonstrated that the RVM model was able to predict future system states (generalization ability) and to estimate or characterize the uncertainty of the predictions (predictive confidence intervals). Another important advantage of utilizing RVMs for real-time application is their sparse formulation. RVMs typically utilize fewer basis functions when compared with SVMs (Tipping 2001). Inducing sparsity can be an effective method to control model complexity, avoid over-fitting, and control the computational characteristics of model performance (Tipping and Faul 2003).

The research reported here extends the capability introduced above to hourly and daily multi-step-ahead predictions with predictive confidence intervals for three canals simultaneously. Therefore, the models recognize the patterns between multivariate outputs (future irrigation diversion requirements for three canals) and multivariate inputs (recent data collected about the state of the system). In order to obtain multiple-time-ahead predictions with predictive confidence intervals, the model exploits the capability of the multivariate relevance vector machine (MVRVM) (Thayananthan 2005; Thayananthan et al. 2008).

The MVRVM is a Bayesian regression tool and represents an extension of the RVM algorithm developed by Tipping and Faul (2003). It can be used to produce multivariate outputs when given a set of multivariate inputs. The MVRVM has the same capabilities as the conventional RVM: high prediction accuracy, robustness, sparse formulation, and characterization of uncertainty in the predictions. Therefore, developing a model with all these properties can provide a practical decision support tool in real-time water resources management by providing multiple predictions that are difficult (or not practical) to obtain from traditional modeling approaches.

The remainder of the paper describes the MVRVM learning model, the time series characteristics, the area of study where the model has been applied, the way the model has been developed for an irrigation canal system, the results of the MVRVM application, a comparison with an ANN model, and conclusions that can be drawn.

Model description

Thayananthan (2005) developed the multivariate relevance vector machine (MVRVM), which is an extension of the relevance vector machine (RVM) for regression (Tipping and Faul 2003) to generate multivariate outputs. Ticlavilca and McKee (2011) applied MVRVM in water resources management research in order to develop multiple-time-ahead predictions of daily releases from multiple reservoirs. This section summarizes the description of the MVRVM model.

Given N observations consisting of input-target vector pairs $\{\mathbf{x}^{(n)}, \mathbf{t}^{(n)}\}_{n=1}^{N}$, where $\mathbf{x} \in \mathrm{R}^D$ is an input vector and $\mathbf{t} \in \mathrm{R}^M$ is an output target vector, a model learns the dependency between input and target with the purpose of making accurate predictions of $\mathbf{t}$ for previously unseen values of $\mathbf{x}$:

$$\mathbf{t} = \mathbf{\Phi}(\mathbf{x})\mathbf{W} + \varepsilon \tag{1}$$

where $\mathbf{W}$ is a $M \times P$ weight matrix and $P = N + 1$. The error or noise vector ε is conventionally assumed to have a Gaussian distribution with zero-mean (Tipping and Faul 2003) and diagonal covariance matrix $\mathbf{S} = \mathrm{diag}\left(\sigma_1^2, \ldots, \sigma_M^2\right)$ (Thayananthan 2005).

A kernel function $K(\mathbf{x}, \mathbf{x}_n)$ can be used to define a set of nonlinear basis functions of the form $\mathbf{\Phi}(\mathbf{x}) = [1, K(x, x^{(1)}, \ldots K(x, x^{(N)}))$. Then, from Eq. 1, the term $\mathbf{\Phi}(\mathbf{x})\mathbf{W} = \sum_{n=1}^{N} K(x, x_n) w_n$ is a linear-weighted sum of N basis functions. In this paper, we considered a Gaussian kernel $K(x, x_n) = \exp(-a^{-2}||x - x_n||)^2$ where "a" is the kernel width parameter. This type of kernel has been used by several authors in water resources and hydrology applications (Khalil et al. 2005b; Tripathi and Govindaraju 2007).

Let τ_r be the vector with the rth component of all the example output vectors, such that $\mathbf{t} = [\tau_1, \ldots, \tau_r, \ldots, \tau_M]^{\mathrm{T}}$. A likelihood distribution of the weight matrix $\mathbf{W} = [\mathbf{w}_1, \ldots, \mathbf{w}_r, \ldots, \mathbf{w}_M]^{\mathrm{T}}$ can be written as a product of Gaussians of the weight vectors ($\mathbf{w}_r$) corresponding to each target output (τ_r) (Thayananthan et al. 2008):

$$p\left(\left\{\mathbf{t}^{(n)}\right\}_{n=1}^{N} | \mathbf{W}, \mathbf{S}\right) = \prod_{n=1}^{N} N\left(\mathbf{t}^{(n)} | \mathbf{W}\mathbf{\Phi}(\mathbf{x}^{(n)}), \mathbf{S}\right) = \prod_{r=1}^{M} N\left(\tau_r | \mathbf{w}_r \Phi, \sigma_r^2\right) \tag{2}$$

where $\mathbf{\Phi} = [1, \Phi(\mathbf{x}_1), \Phi(\mathbf{x}_2), \ldots, \Phi(\mathbf{x}_N)]$. From Eq. 2, the model contains as many parameters as training observations. As a result, there is a danger that the optimal estimation of $\mathbf{w}_r$ and σ_r^2 will suffer from severe over-fitting (Tipping 2001). To avoid this, Tipping (2001) proposed applying a Bayesian perspective, and thereby imposing some constrain on the selection of parameters by defining an explicit zero-mean Gaussian prior probability distribution over them (Thayananthan et al. 2008):

$$p(\mathbf{W}|\mathbf{A}) = \prod_{r=1}^{M} \prod_{j=1}^{P} N(w_{rj}|0, \alpha_j^{-2}) = \prod_{r=1}^{M} N(\mathbf{w}_r|0, \mathbf{A}) \tag{3}$$

where w_{rj} is the (r, j)th element of the weight matrix $\mathbf{W}$, and $\mathbf{A} = \mathrm{diag}\left(\alpha_1^{-2}, \ldots, \alpha_P^{-2}\right)^{\mathrm{T}}$ is a diagonal matrix which contains all the hyperparameters $\alpha = \{\alpha_j\}_{j=1}^{P}$. Each individual α_j determines the relevance of the associated basis function (Thayananthan et al. 2008).

Given the likelihood and prior distribution, the posterior distribution of the model parameters can be defined within Bayes' rule:

$$p\left(\mathbf{W}|\{\mathbf{t}\}_{n=1}^{N}, \mathbf{S}, \mathbf{A}\right) \propto p\left(\{\mathbf{t}\}_{n=1}^{N}|\mathbf{W}, \mathbf{S}\right) p(\mathbf{W}|\mathbf{A}) \tag{4}$$

The posterior parameter distribution conditioned on the data is Gaussian, and it can be written as the product of Gaussians for the weight vectors of each target output dimension (Thayananthan et al. 2008):

$$p\left(\mathbf{W}|\{\mathbf{t}\}_{n=1}^{N}, \mathbf{S}, \mathbf{A}\right) \propto p\left(\{\mathbf{t}\}_{n=1}^{N}|\mathbf{W}, \mathbf{S}\right) p(\mathbf{W}|\mathbf{A}) \propto \prod_{r=1}^{M} N(\mathbf{w}_r|\mu_r, \mathbf{\Sigma}_r) \tag{5}$$

where the covariance and mean are $\mathbf{\Sigma}_r = \left(\mathbf{A} + \boldsymbol{\sigma}_r^{-2}\mathbf{\Phi}^{\mathrm{T}}\mathbf{\Phi}\right)^{-1}$ and $\boldsymbol{\mu}_r = \boldsymbol{\sigma}_r^{-2}\mathbf{\Sigma}_r\mathbf{\Phi}^{\mathrm{T}}\tau_r$, respectively.

Given the posterior, we can obtain an optimal weight matrix by getting a set of hyperparameters that maximizes the data likelihood in Eq. 5. The data likelihood is marginalized by analytically integrating the weights out (Thayananthan et al. 2008):

$$\begin{aligned} p\left(\{\mathbf{t}\}_{n=1}^{N}|\mathbf{A}, \mathbf{S}\right) &= \int p\left(\{\mathbf{t}\}_{n=1}^{N}|\mathbf{W}, \mathbf{S}\right) p(\mathbf{W}|\mathbf{A}) \mathrm{d}\mathbf{W}, \\ &= \prod_{r=1}^{M} \int N\left(\boldsymbol{\tau}_r|\mathbf{w}_r\Phi, \sigma_r^2\right) N(\mathbf{w}|\mathbf{0}, \mathbf{A}) \\ &= \prod_{r=1}^{M} |\mathbf{H}_r|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\boldsymbol{\tau}_r^{\mathrm{T}}\mathbf{H}_r^{-1}\boldsymbol{\tau}_r\right) \end{aligned} \tag{6}$$

where $\mathbf{H}_r = \sigma_r^2\mathbf{I} + \Phi\mathbf{A}^{-1}\mathbf{\Phi}^{\mathrm{T}}$. Then, the optimal values of hyperparameters $\boldsymbol{\alpha}^{\mathrm{opt}} = \left\{\boldsymbol{\alpha}_j^{\mathrm{opt}}\right\}_{j=1}^{P}$ and noise parameters $(\boldsymbol{\sigma}^{\mathrm{opt}})^2 = \left\{\sigma_r^{\mathrm{opt}}\right\}_{r=1}^{M}$ can be obtained by maximizing the marginal likelihood using the optimization algorithm proposed by Tipping and Faul (2003). Many optimal values of $\boldsymbol{\alpha}$ are infinite, for which the posterior probability of the weight becomes zero. The relatively few nonzero weights are the relevance vectors (RVs), which generate a sparse representation.

The optimal parameters are used to obtain the optimal weight matrix with optimal covariance $\mathbf{\Sigma}^{\mathrm{opt}} = \left\{\mathbf{\Sigma}_r^{\mathrm{opt}}\right\}_{r=1}^{M}$ and mean $\boldsymbol{\mu}^{\mathrm{opt}} = \left\{\mu_r^{\mathrm{opt}}\right\}_{r=1}^{M}$.

We can compute the predictive distribution for a new input $\mathbf{x}^*$, corresponding to a target $\mathbf{t}^*$, from:

$$p\left(\mathbf{t}^*|\mathbf{t}, \boldsymbol{\alpha}^{\mathrm{opt}}, (\boldsymbol{\sigma}^{\mathrm{opt}})^2\right) = \int p\left(\mathbf{t}^*|\mathbf{W}, (\boldsymbol{\sigma}^{\mathrm{opt}})^2\right) . p\left(\mathbf{W}|\mathbf{t}, \boldsymbol{\alpha}^{\mathrm{opt}}, (\boldsymbol{\sigma}^{\mathrm{opt}})^2\right) \mathrm{d}\mathbf{W} \tag{7}$$

Taking into consideration that both terms in the integrand are Gaussian, Eq. 7 is computed as:

$$p\left(\mathbf{t}^*|\mathbf{t}, \boldsymbol{\alpha}^{\mathrm{opt}}, (\boldsymbol{\sigma}^{\mathrm{opt}})^2\right) = N\left(\mathbf{t}^*|\mathbf{y}^*, (\boldsymbol{\sigma}^*)^2\right) \tag{8}$$

where $\mathbf{y}^* = \left[\mathrm{y}_1^*, \ldots, \mathrm{y}_r^*, \ldots, \mathrm{y}_M^*\right]^T$ is the predictive mean with $\mathbf{y}_r^* = \left(\mu_r^{\mathrm{opt}}\right)^{\mathrm{T}}\mathbf{\Phi}(\mathbf{x}^*)$ and $(\boldsymbol{\sigma}^*)^2 = \left[\left(\sigma_1^*\right)^2, \ldots \left(\sigma_r^*\right)^2, \ldots, \left(\sigma_M^*\right)^2\right]^{\mathrm{T}}$ is the predictive variance with $\left(\sigma_r^*\right)^2 = \left(\sigma_r^{\mathrm{opt}}\right)^2 + \mathbf{\Phi}(\mathbf{x}^*)^{\mathrm{T}}\mathbf{\Sigma}_r^{\mathrm{opt}}\mathbf{\Phi}(\mathbf{x}^*)$. The predictive variance $\left(\sigma_r^*\right)^2$ contains the sum of two variance terms: the noise on the data and the uncertainty in the prediction of the weight parameters (Tipping 2001).

The standard deviation σ_r^* of the predictive distribution is defined as a predictive error bar of y_r^* (Bishop 1995). The width of the 90% predictive confidence interval for any y_r^* can be calculated as $\pm 1.65\sigma_r^*$.

Readers interested in greater detail regarding the mathematical formulation and the optimization procedures of the MVRVM model are referred to Thayananthan (2005) and Thayananthan et al. (2008). Readers interested in researching in more depth on RVM and Bayesian topics are referred to Tipping (2001), Tipping and Faul (2003) and Tipping (2004). A MATLAB code developed by Thayananthan (2005) is available from http://mi.eng.cam.ac.uk/~at315/MVRVM.

Time series characteristics

Autocorrelation function analysis was performed to explore any dynamic behavior in the time series (on both an hourly and a daily basis). The autocorrelation functions for all three demand diversion canals have slow decays, which indicates that the time series are nonstationary (Pulido-Calvo et al. 2003). This implies that using classical statistics (such as ARMA models, which assume that the time series are stationary), would be unsuitable. ARIMA models could be used because they take into account nonstationary behaviors of the time series; however, they use a linear parametric approach, which can lead to poor prediction results on unseen data. ARIMA models are also unsuitable for long-term forecasting (e.g. forecasting up to 24 h ahead for hourly irrigation diversion demand) because the long-term forecast asymptotically approaches the mean value of the time series data (Shumway and Stoffer 2011). Moreover, studies in water resources and irrigation topics developed by Pulido-Calvo et al. (2003) and Nourani et al. (2009) have demonstrated that machine learning techniques perform better than classical ARIMA models. These studies proved the suitability of machine learning models to learn the nonlinear dynamics and nonstationary behavior of the data in order to make accurate predictions for previously unseen values.

For this paper, we apply Bayesian machine learning models that have been successfully applied in forecasting complex systems in water resources topics (Ticlavilca and McKee 2011; Khalil et al. 2005c). These models use parametric techniques that assume a functional form that can approximate a large number of complex functions by using nonlinear transformation of a large number of parameters. Moreover, the Bayesian approach of the models (explained in the previous section) allows them to find the model parameters in a framework that will avoid overfitting.

As indicated previously, the time series are characterized as nonstationary. Therefore, the proposed model will need to be retrained periodically to capture the pattern changes in the data.

Study area

The Sevier River Basin, illustrated in Fig. 1, is Utah's largest drainage area and one of its driest as well. Water shortages are particularly acute in the lower basin, where the approximately 70,000 acres (28,000 ha) are allocated an average of only 190,000 acre-feet (234 MCM). Salinity levels average about 2,400 dS/m and require careful attention to irrigation uniformity and efficiency. Irrigators have compensated for the lack of water by utilizing an extensive water measurement program, precision land leveling, and real-time reservoir management. The Sevier River basin today is fully equipped with SCADA systems on all reservoir and canal gates to provide a nearly virtual management and regulation capability. This system can be viewed at: www.sevierriver.org.

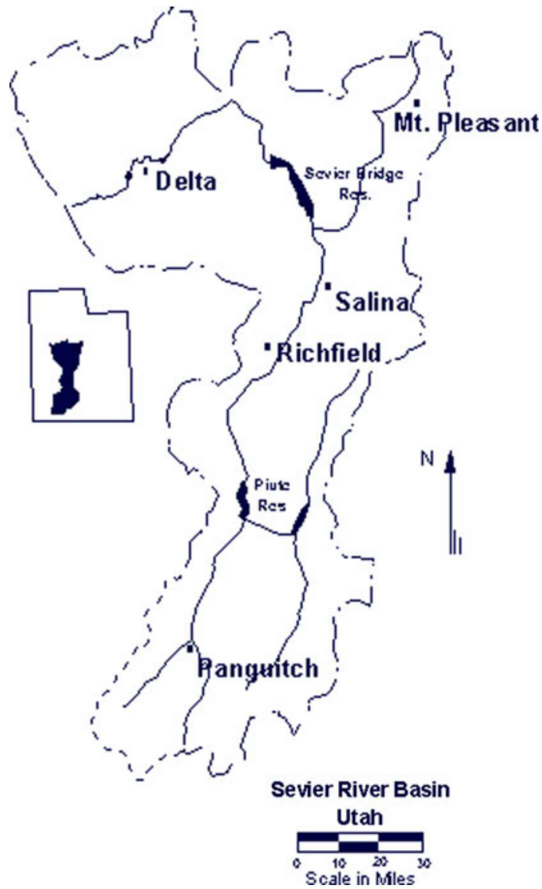


Fig. 1 Sevier River Basin

Model application

In the lower basin, illustrated in Fig. 2, irrigation water is released on a scheduled demand basis from Sevier Bridge Reservoir to three canal companies that divert directly from the river and four irrigation companies that divert from two regulating reservoirs (DMAD Reservoir and Gunnison Bend Reservoir). There is an inherent lag time of 1 day in the delivery of water from Sevier Bridge Reservoir to the

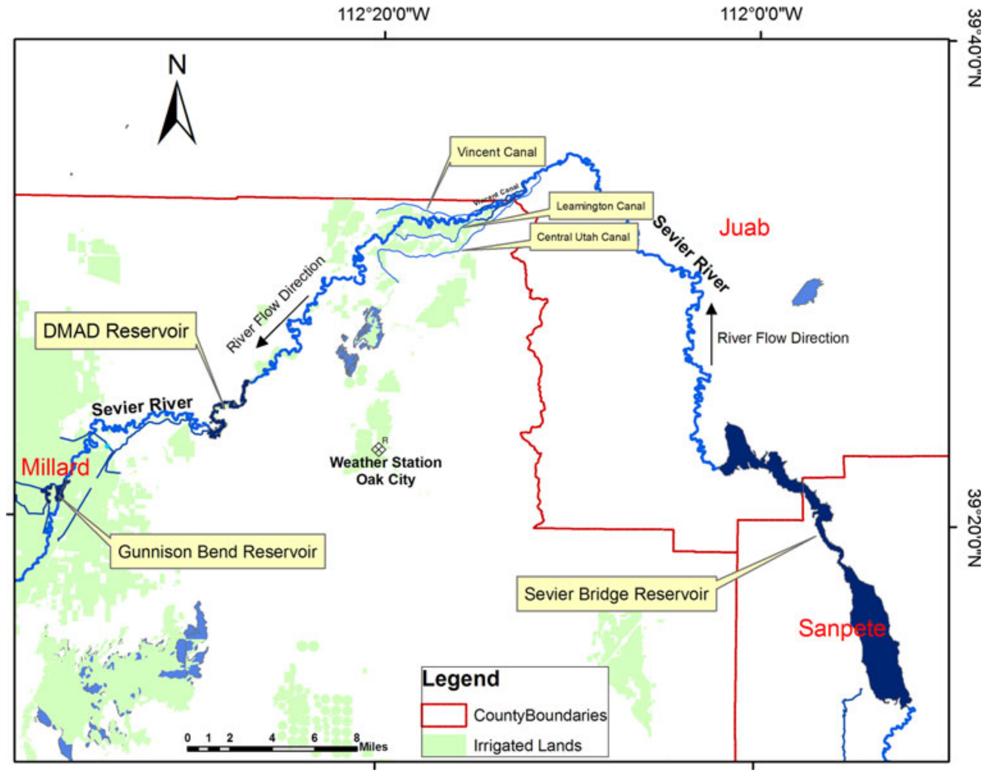


Fig. 2 Lower Sevier River Basin

Vincent, Central Utah, and Leamington Canals between Sevier Bridge and DMAD. This lag time increases to 2.5 days to DMAD and 3.5 days to Gunnison Bend. The canal and reservoir companies have not yet implemented computerized water ordering systems, and the watermasters operating the canals and the Sevier River Commissioner operating Sevier Bridge must communicate the water orders verbally. This verbal communication is inefficient, and both canal and reservoir operators must anticipate orders based on personal experience.

For the Vincent, Central Utah, and Leamington Canals (Fig. 3), the lack of coordination between the water orders and the reservoir releases can result in inadequate flows at their river intakes. The analyses reported herein are applied to the water management problem of this case to determine whether a reasonable forecast of water orders could be made, thus facilitating the proper release of water from Sevier Bridge. Two time frames are examined. A daily forecast is evaluated to program the daily release from Sevier Bridge, and an hourly forecast is evaluated to determine how the daily release should be regulated.

MVRVM forecasting hourly model

The irrigation canals are intensively monitored in real-time to enable on-demand operation of the irrigation system. The travel time from the head gate canals to the irrigated areas ranges from 1 to 24 h. A model that predicts hourly irrigation demands (1, 12 and 24 h ahead) can assist reservoir and canal operators to anticipate future hourly behavior of the system and to more intensively monitor the system.

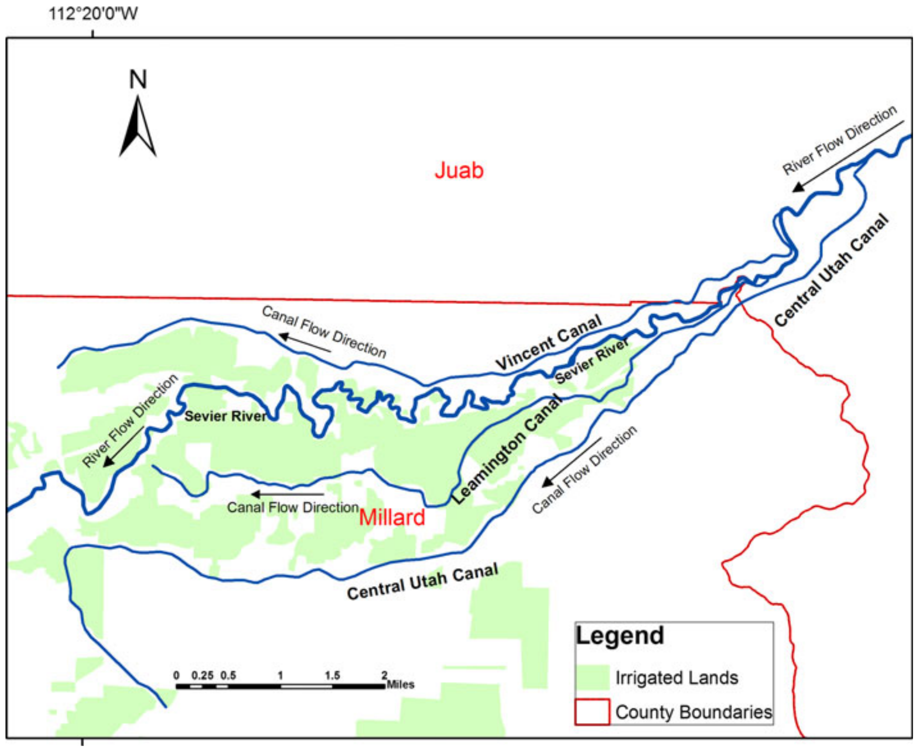
Hourly data from the 2005 and 2006 irrigation seasons (approximately 10,000 input–output pairs) were used to train the MVRVM hourly model and to find the model parameters. This large size of data set required a training time of 27 min. While this is not a considerable training time, the training of 1,000 bootstrap models to explore model robustness (one of the purposes of this research that will be explained with more detail later in this report) would have become computationally expensive. Wang et al. (2005) found that, despite its simplicity, data set reduction by uniform random sampling performs consistently well when employed before using SVM modeling for real-world problems, especially when the reduction rate is high (selection of 30% of the whole data set). Therefore, a subset of 30% of the entire training data set (3,000 observations) was selected using uniform random sampling, thus reducing the training speed to approximately 6 min.

The inputs used in the model to predict hourly canal demands are expressed as

$$\mathbf{x} = \left[\mathbf{D1}_{\mathrm{h-nh}}, \mathbf{D2}_{\mathrm{h-nh}}, \mathbf{D3}_{\mathrm{h-nh}}\right]^{\mathrm{T}} \tag{9}$$

where h = hour of prediction, nh = number of hours previous to the prediction time, $\mathbf{D1}_{\text{d-nh}}$ = Central Utah canal demand 'nh' hours previous to the prediction time,

Fig. 3 Sevier River, Central Utah Canal, Vincent Canal, and Leamington Canal



$\mathbf{D2}_{\text{d-nh}}$ = Vincent canal demand 'nh' hours previous to the prediction time, $\mathbf{D3}_{\text{d-nh}}$ = Leamington canal demand 'nh' hours previous to the prediction time

The multiple output target vector of the model is expressed as

$$\mathbf{t} = [\mathbf{D1}_{\text{h}}, \mathbf{D1}_{\text{h+12}}, \mathbf{D1}_{\text{h+24}}, \mathbf{D2}_{\text{h}}, \mathbf{D2}_{\text{h+12}}, \mathbf{D2}_{\text{h+24}}, \mathbf{D3}_{\text{h}}, \mathbf{D3}_{\text{h+12}}, \mathbf{D3}_{\text{h+24}}]^{\text{T}} \tag{10}$$

where $\mathbf{D1}_{\text{h}}$ = prediction of Central Utah canal demand 1 h ahead, $\mathbf{D1}_{\text{h+12}}$ = prediction of Central Utah canal demand 12 h ahead, $\mathbf{D1}_{\text{h+24}}$ = prediction of Central Utah canal demand 24 h ahead, $\mathbf{D2}_{\text{h}}$ = prediction of Vincent canal demand 1 h ahead, $\mathbf{D2}_{\text{h+12}}$ = prediction of Vincent canal demand 12 h ahead, $\mathbf{D2}_{\text{h+24}}$ = prediction of Vincent canal demand 24 h ahead, $\mathbf{D3}_{\text{h}}$ = prediction of Leamington canal demand 1 h ahead, $\mathbf{D3}_{\text{h+12}}$ = prediction of Leamington canal demand 12 h ahead, $\mathbf{D3}_{\text{h+24}}$ = prediction of Leamington canal demand 24 h ahead

MVRVM forecasting daily model

The reservoir operator must release water from Sevier Bridge reservoir in order to fulfill water demands for the irrigation canals and DMAD reservoir (Fig. 2). The approximate travel time between Sevier Bridge Reservoir and the canals is 1 day. A model that predicts irrigation water demands 1 and 2 days ahead can assist the operator of the Sevier Bridge Reservoir located upstream of the irrigation canals in planning and managing the available water of the reservoir efficiently.

Daily data from the 2001 through 2006 irrigation seasons (1,194 input–output pairs) were used to train the daily MVRVM model to find the model parameters. Daily data from the 2007 irrigation season were used to test the model. The inputs are the available past daily data collected by sensors on the canals and air temperature. Rainfall information was not taken into account since the average precipitation in the Lower Sevier River Basin is relatively low (ranges from 7 to 9 in. per year).

Air temperature affects the daily rate of evapotranspiration in the irrigated areas and evaporation in the river and canals. Moreover, the daily behavior of farmers and canal operators can be directly influenced by temperature information in the basin (Khalil et al. 2005b). Daily maximum and minimum temperature from Oak city were included in the model inputs. This historical data to build the model were obtained from The National Oceanic and Atmospheric Administration (NOAA), and recent online past data can be obtained from the National Weather Services web site, www.nws.noaa.gov.

The inputs used in the model to predict daily canal demands are expressed as

$$\mathbf{x} = [\mathbf{D1}_{\text{d-nd}}, \mathbf{D2}_{\text{d-nd}}, \mathbf{D3}_{\text{d-nd}}, \mathbf{Tmax}_{\text{d-nd}}, \mathbf{Tmin}_{\text{d-nd}}]^{\text{T}} \tag{11}$$

where d = day of prediction, nd = number of days previous to the prediction time, $\mathbf{D1}_{\text{d-nd}}$ = Central Utah canal demand 'nd' days previous to the prediction time, $\mathbf{D2}_{\text{d-nd}}$ = Vincent canal demand 'nd' days previous to the prediction time, $\mathbf{D3}_{\text{d-nd}}$ = Leamington canal demand 'nd' days previous to the prediction time

The multiple output target vector of the model is expressed as

$$\mathbf{t} = [\mathbf{D1}_{\mathrm{d}}, \mathbf{D1}_{\mathrm{d+1}}, \mathbf{D2}_{\mathrm{d}}, \mathbf{D2}_{\mathrm{d+1}}, \mathbf{D3}_{\mathrm{d}}, \mathbf{D3}_{\mathrm{d+1}}]^{\mathrm{T}} \tag{12}$$

where $\mathbf{D1}_{\mathrm{d}}$ = prediction of Central Utah canal demand 1 day ahead, $\mathbf{D1}_{\mathrm{d+1}}$ = prediction of Central Utah canal demand 2 days ahead, $\mathbf{D2}_{\mathrm{d}}$ = prediction of Vincent canal demand 1 day ahead, $\mathbf{D2}_{\mathrm{d+1}}$ = prediction of Vincent canal demand 2 days ahead, $\mathbf{D3}_{\mathrm{d}}$ = prediction of Leamington canal demand 1 day ahead, $\mathbf{D3}_{\mathrm{d+1}}$ = prediction of Leamington canal demand 2 days ahead.

Model selection

The statistic used for model selection is the coefficient of efficiency (E) calculated for the testing phase. It has been recommended by the ASCE (1993) and Legates and McCabe (1999), and is given by:

$$E = 1 - \frac{\sum_{n=1}^{N} (t_n - t_n^*)^2}{\sum_{n=1}^{N} (t_n - t_{\mathrm{av}})^2} \tag{13}$$

where $\mathbf{t}$ is the observed output; t^* is the predicted output; t_{av} is the observed average output and N is the number of observations. This statistic ranges from minus infinity (poor model) to 1.0 (a perfect model) (Legates and McCabe 1999).

The kernel width parameter and the number of previous time steps had to be selected in order to find the appropriated model. The term "number of previous time steps" corresponds to the term "model order" as it is defined by autoregressive modeling approaches. Since the machine learning technique applied here does not use the same approach as classical time series analysis, the model order was not calculated with traditional time series methods (e.g., partial autocorrelation and Akaike information). Instead, the model order and the kernel width parameter were estimated by a trial and error process.

Several MVRVM models were built with variation in kernel width and the number of previous time steps. For the daily model, the model order "nd" (or previous time steps as input data) was chosen from a range of 1–5 days previous to the prediction time. For the hourly model, the model order "nh" was chosen from a range of 4–48 h (incrementing by 4 h) previous to the prediction time. For each order, the kernel width was chosen from a range of 1–40 (for the daily model) and from a range of 2–40, incrementing by 2 (for the hourly model).

For both models (hourly and daily), the selected kernel width is the one with maximum E. From the list of models with selected kernel width at different "nd" and "nh" values for the daily and hourly model, respectively, we considered that the selected model is the one with the maximum E and minimum value of model order.

Bootstrap analysis

In this paper, we apply machine learning techniques for irrigation demand processes (which are a combination of human behavior and hydrological processes) to learn inverse or retrieval models (Cherkassky et al. 2006; Krasnopolsky 2009). This means we are training a model to find its parameters based on measured data. Most hydrological inverse processes are ill posed (Parker 1994), so machine learning applications deal with formulation for ill-posed problems while trying to replicate the expected response of the system based on measured data. Also, there are uncertainties in data measurements (e.g. noise in data, scarce data, and lack of relevant data) that contribute to data complexity (Cherkassky et al. 2006). This is why it is necessary to develop machine learning models to guarantee not only high accuracy but also good generalization and robustness of parameter estimation with respect to future changes in the nature of the input data. Changes in the training data used to build a model may give different test results.

The bootstrap method (Efron and Tibshirani 1998) was used to explore the implications of the change in the nature of the input data and to guarantee good generalization ability and robustness of the machine learning model (Khalil et al. 2005b). The bootstrap data set was created by randomly sampling with replacement from the whole training data set. In the bootstrap estimation, this selection process was independently repeated 1,000 times to yield 1,000 bootstrap training data sets, which are treated as independent sets (Duda et al. 2001). For each of the bootstrap training data sets, a model was built and evaluated over the original test data set.

The bootstrap method provides implicit information on the uncertainty in the statistics estimator evaluated in the model. A robust model is one that shows a narrow confidence bounds in the bootstrap histogram (Khalil et al. 2005b).

Comparison between MVRVM and ANN

ANNs have been widely applied in irrigation, water resources, and hydrologic modeling (ASCE Task Committee on the Application of ANNs in Hydrology 2000a, b; Khalil et al. 2005d; Pulido-Calvo and Gutierrez-Estrada 2008; Rahimi 2008; Adeloye 2009; Kumar et al. 2011). A

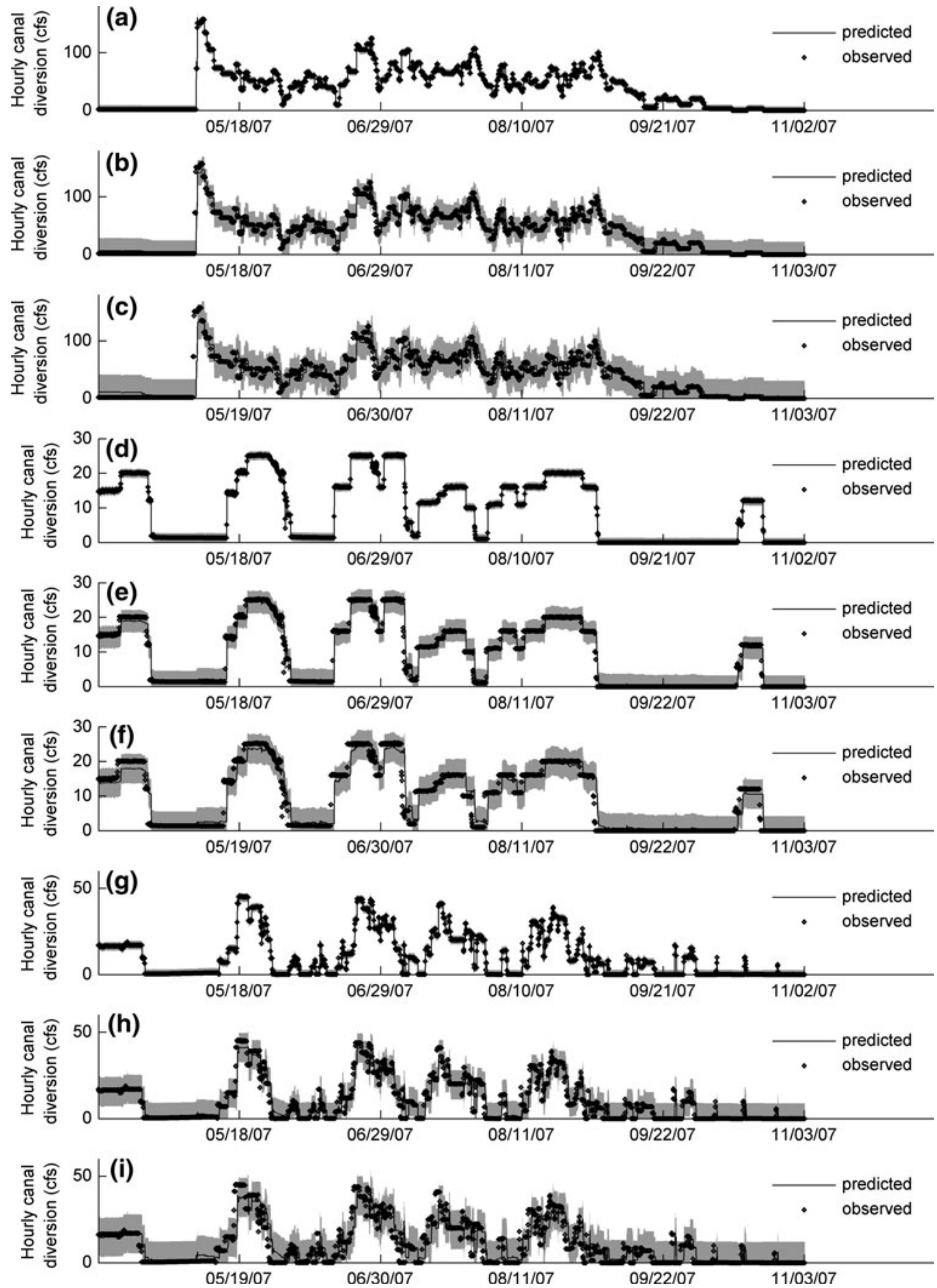


Fig. 4 Observed versus predicted diversions for the hourly MVRVM model with 0.90 confidence intervals (*shaded region*) for the testing phase (2007 irrigation season). Central Utah Canal: **a** 1 h ahead, **b** 12 h ahead, **c** 24 h ahead; Vincent Canal: **d** 1 h ahead, **e** 12 h ahead, **f** 24 h ahead; Leamington Canal: **g** 1 h ahead, **h** 12 h ahead, **i** 24 h ahead

comparative analysis between the MVRVM models, which have been developed here, and ANN models is performed in terms of performance and robustness. The ANN toolbox available in MATLAB is applied in this research. Readers interested in greater detail regarding ANNs and their training functions are referred to Demuth et al. (2009).

The Levenberg–Marquardt (LM) algorithm was used to train the ANN model (Demuth et al. 2009; Tan and Van Cauwenberghe 1999). This algorithm, which has been used by several authors in irrigation, water resources, and hydrologic research (Pulido-Calvo and Gutierrez-Estrada 2008; Rahimi 2008; Kisi 2009), is a training algorithm with second-order convergence to optimize the ANN parameters (i.e. weight and bias) (Tan and Van Cauwenberghe 1999). The LM algorithm requires a random initial guess for the ANN parameters. A preliminary ANN model with the same architecture was tested 1,000 times to explore the variation of the model efficiency. These preliminary results indicated that the random variation on the initial parameters for the diversion demand time series do not have a significant effect on the model efficiency. Therefore, the model was set to a fixed random initial parameter. This was done to

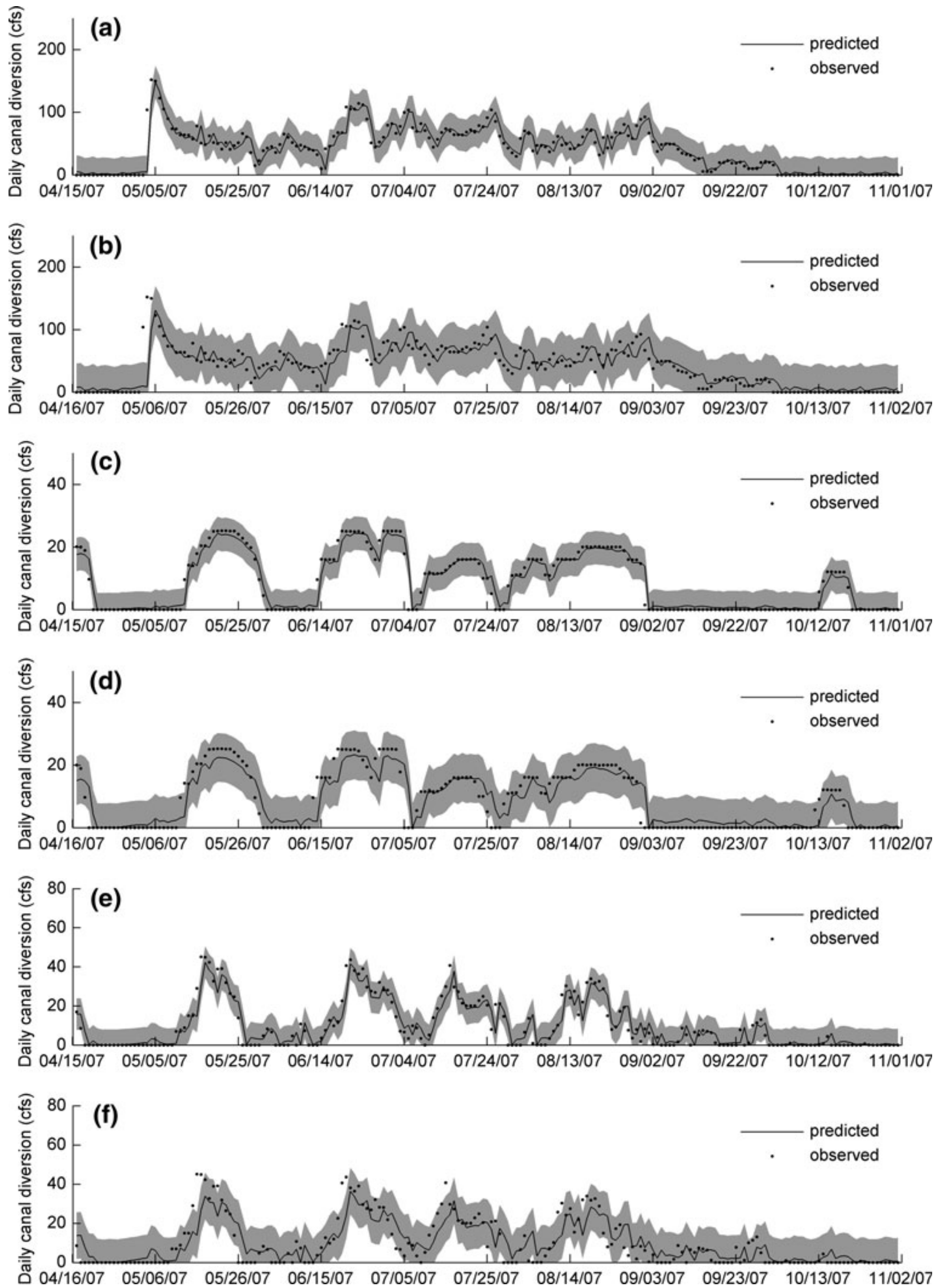


Fig. 5 Observed versus predicted diversions of the daily MVRVM model with 0.90 confidence intervals (*shaded region*) for the testing phase (2007 irrigation season). Central Utah Canal: **a** 1 day ahead, **b** 2 days ahead; Vincent Canal: **c** 1 day ahead, **d** 2 days ahead; Leamington Canal: **e** 1 day ahead, **f** 2 days ahead

allow the training computation for an ANN model with the same architecture to be repeated with the same results.

Preliminary investigations performed by the authors showed that training an ANN with two or more hidden layers did not result in significant improvement and it required significant training time. In fact, simulations performed by Tabari et al. (2010) showed that an increase in the hidden layer and the number of neurons in each hidden layer brought nearly insignificant improvement to the ANN model. Moreover, Kumar et al. (2008) pointed out that ANN architecture with multiple hidden layers can lead to overfitting. Therefore, a single hidden layer was used for the ANN model structure.

Several ANN models were built that included variation in size of the single hidden layer (i.e., number of neurons) and the number of previous time steps (or model orders). The model orders for the daily model range from 1 to 5 days; and the model orders for the hourly model ranges from 4 to 48 h. For each order, the size of the hidden layer ranges from 1 to 10 neurons. The selected hidden layer size

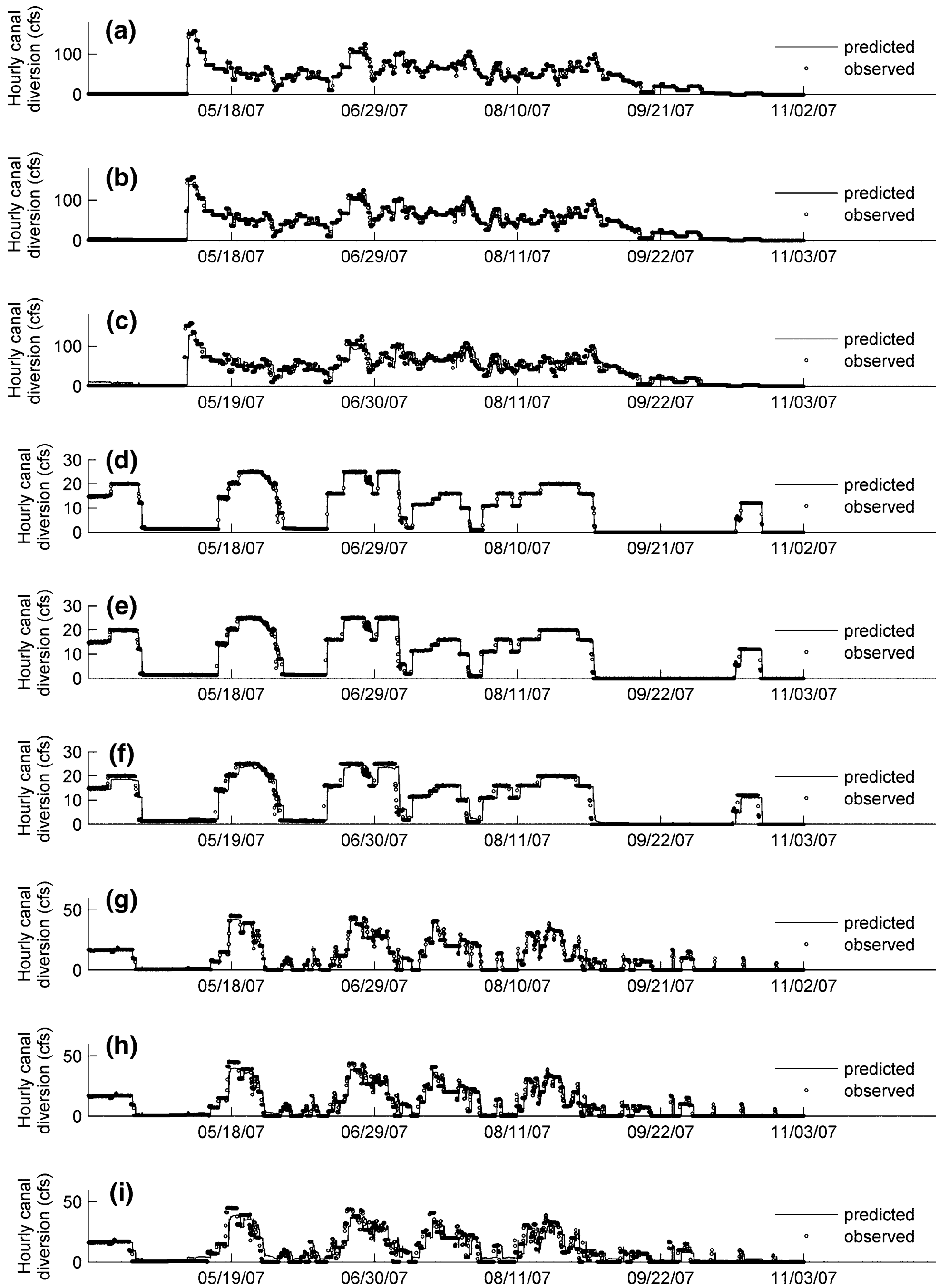


Fig. 6 Observed versus predicted diversions of the hourly ANN model for the testing phase (2007 irrigation season). Central Utah Canal: **a** 1 h ahead, **b** 12 h ahead, **c** 24 h ahead; Vincent Canal: **d** 1 h ahead, **e** 12 h ahead, **f** 24 h ahead; Leamington Canal: **g** 1 h ahead, **h** 12 h ahead, **i** 24 h ahead

is the one with maximum E. From the list of selected models at different model orders, the best model is the one with the maximum E and minimum value of model order.

Results and discussion

The predicted or anticipated water orders represented by the solid lines in Figs. 4 and 5 can serve as useful information from a water management standpoint in the basin. As previously mentioned, water-management problems are due to the one-day lag time in the delivery of water from Sevier Bridge reservoir to the irrigation diversion canals. Although the basin is fully implemented with a SCADA system, the canal and reservoir operators must still anticipate the water orders based on personal experience and communicate them verbally. The irrigation demand predictions from the proposed MVRVM model can be implemented in computerized water order systems and thus, assist the reservoir operator to respond to short-term changes in demand and to determine the amount of water that should be given to the downstream reservoirs

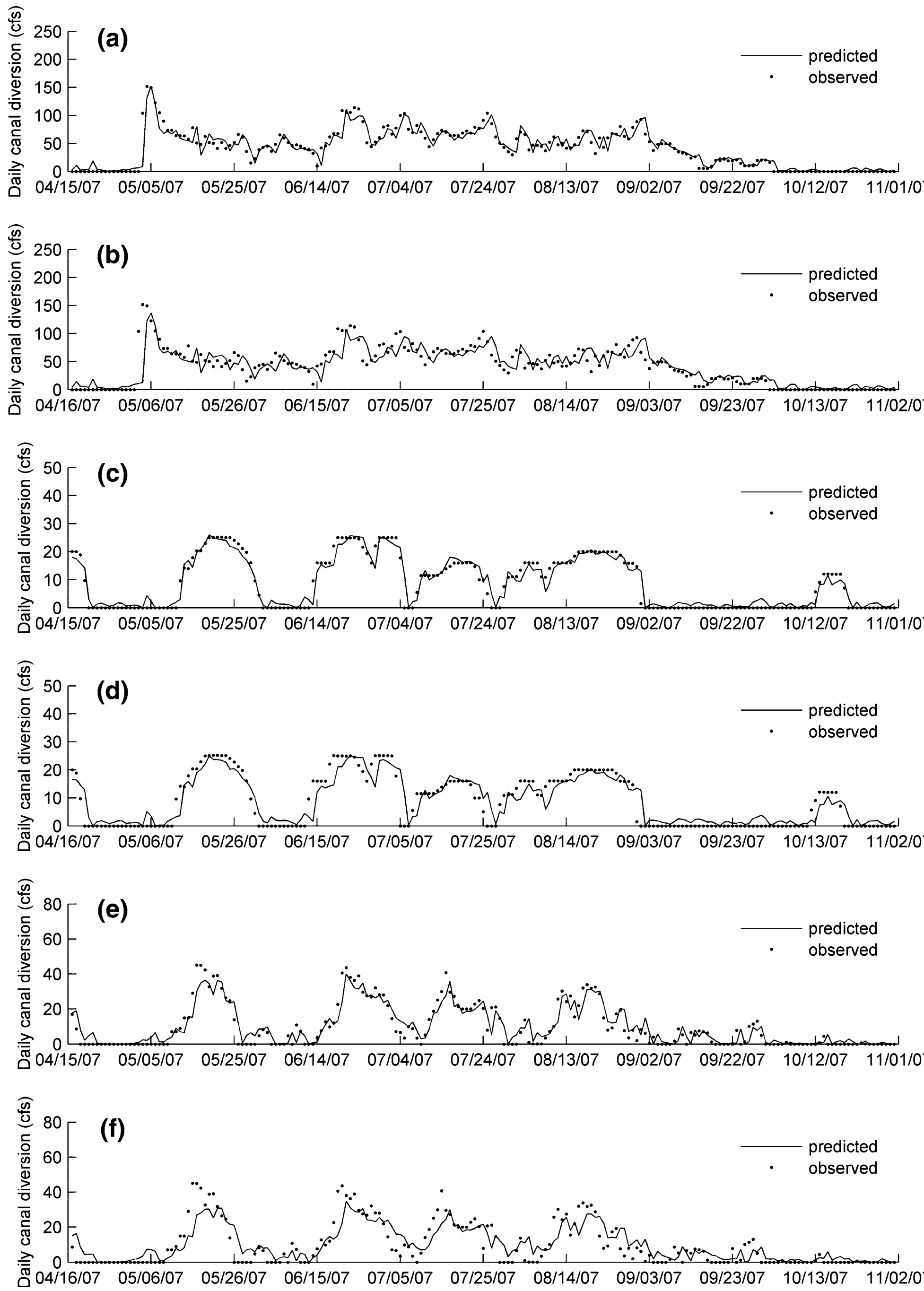


Fig. 7 Observed versus predicted diversions of the daily ANN model for the testing phase (2007 irrigation season). Central Utah Canal: **a** 1 day ahead, **b** 2 days ahead; Vincent Canal: **c** 1 day ahead, **d** 2 days ahead; Leamington Canal: **e** 1 day ahead, **f** 2 days ahead

Table 1 Hourly model performance using different statistics for the testing phase

Model	Statistics	Central Utah Canal			Vincent Canal			Leamington Canal		
		1 h	12 h	24 h	1 h	12 h	24 h	1 h	12 h	24 h
MVRVM (selected kernel width = 14)	*E*	0.993	0.898	0.785	0.996	0.942	0.881	0.986	0.848	0.736
	RMSE	2.879	10.913	15.824	0.549	2.177	3.136	1.460	4.864	6.415
ANN (selected hidden layer size = 5)	*E*	0.994	0.899	0.779	0.996	0.943	0.881	0.989	0.853	0.748
	RMSE	2.682	10.821	16.040	0.562	2.171	3.131	1.336	4.780	6.266

(DMAD and Gunnison Bend) in a proper and timely manner.

The predicted outputs of the MVRVM for the testing phase (2007 irrigation season) are shown as the solid lines in Fig. 4 for the hourly model and Fig. 5 for the daily model. The models explain well the observed irrigation demands (dots). Figures 4 and 5 also show the 0.90 confidence interval (shaded region) associated with the predictive variance of the MVRVM in Eq.8. The confidence intervals for the twenty-four-hour-ahead and the two-day-ahead predictions (Figs. 4c, f, i, 5b, d, f) become wider. We can see how the uncertainty in the predictions increases when predicting further into the future for both hourly and daily models.

The observed (dots) and predicted (solid lines) outputs of the ANN for the testing phase (2007 irrigation season) are shown in Fig. 6 for the hourly model and in Fig. 7 for the daily model. The models explain well the observed irrigation diversions (dots). We can see that the predictions for the ANN models are single target outputs (which are plotted by the solid lines), while the predictions for the MVRVM models (Figs. 4, 5) are probabilistic. As previously mentioned, the outputs are the predictive mean, and the predictive confidence intervals can then be determined (the shaded regions in Figs. 4, 5). This is an advantage when compared with the ANN model since we can determine a degree of uncertainty for the MVRVM predictions.

Renault et al. (2007) emphasized that it is better to give the uncertainty in a value rather than to treat estimates as if they are single values. Each predicted value from Figs. 4 and 5 is the mean of the predictive distribution of possible outputs indicated in Eq. 8. From each predictive distribution, we are able to estimate the 90% predictive confidence intervals, which give the degree of uncertainty associated with each prediction. Therefore, the MVRVM model can provide the operators with an interval of plausible values for predicted irrigation water demand.

Tables 1 and 2 show some statistics that measure MVRVM and ANN performances for both hourly and daily models. Again, we can see good performance of both models in the testing phase for the one-hour-ahead, twelve-hour-ahead, and one-day-ahead forecasts. The performance accuracy is reduced for both models for twenty-four-hour-ahead and two-day-ahead predictions. It is important to mention that the selected models for both MVRVM and ANN required the same number of previous time steps as inputs: 2 days as input data for the daily models, and 4 h as input data for the hourly models. We can see that the performance results for MRVM and ANN are quite similar.

Tripathi and Govindaraju (2007) pointed out that it is very important to carefully select the kernel width of a

Table 2 Daily model performance using different statistics for the testing phase

Model	Statistics	Central Utah Canal		Vincent Canal		Leamintong Canal	
		1 day	2 days	1 day	2 days	1 day	2 days
MVRVM (selected kernel width = 25)	E	0.851	0.659	0.927	0.803	0.846	0.671
	RMSE	12.998	19.648	2.515	4.116	4.890	7.158
ANN (selected hidden layer size = 5)	E	0.840	0.658	0.927	0.809	0.831	0.659
	RMSE	13.447	19.684	2.514	4.056	5.122	7.293

Table 3 Combined and individual MVRVM models (hourly model)

Model	E			Number of previous hours as inputs		
	Central Utah Canal	Vincent Canal	Leamington Canal	Central Utah Canal	Vincent Canal	Leamington Canal
MVRVM (combined)	0.892	0.940	0.857	4		
MVRVM (individual)	0.897	0.940	0.859	4	4	4

Table 4 Combined and individual MVRVM models (daily model)

Model	E			Number of previous hours as inputs		
	Central Utah Canal	Vincent Canal	Leamington Canal	Central Utah Canal	Vincent Canal	Leamington Canal
MVRVM (combined)	0.755	0.865	0.759	2		
MVRVM (individual)	0.760	0.853	0.751	3	5	3

RVM model. They proposed different techniques to find test results with good generalization performance. In their case, they used a RVM model for univariate output developed by Tipping (2001). In this research, we utilized a MVRVM model that is a RVM for multivariate outputs. As a practical application, we used a trial and error process to find an optimal kernel width parameter that satisfies all of the multiple outputs of the model. The generalization capabilities and robustness of the selected model parameter will be demonstrated later in this section. The selected kernel width parameter was 14 and 25 for the hourly and daily models, respectively (Tables 1, 2).

Figures 4 and 5 also show some differences in patterns among the diversion demand canals. It is because the numbers of offtakes relative to the service area are different. In fact, Central Utah and Leamington canal serve to approximately 50 and 30 stakeholders, respectively, while Vincent canal serves to only 1 stakeholder. Therefore, to determine whether it would make sense to combine the three canals into one prediction model, we also compared the proposed combined model with individual models in order to explore any significant differences.

The results are shown in Tables 3 and 4 for hourly and daily basis, respectively. There is no significant difference regarding the coefficient of efficiency (E) between the combined and individual models for the hourly (Table 3) and daily (Table 4) model.

In Table 4, we can see (at daily basis) that the combined MVRVM needs just 2 previous days as inputs (or model order), while each individual model needs more days (3, 5, and 3 days for the Central Utah Canal, Vincent canal and Leamington canal, respectively). Therefore, building a combined model (at daily basis) results in using fewer inputs. Using fewer inputs may be an advantage when dealing with data retrieval and incomplete data. However, more definite conclusions regarding the use of either a combined model or individual models are not provided at this time, as they may not be sufficiently well supported.

We prefer to apply the combined model to demonstrate the advantage of the MVRVM in giving multivariate outputs and forecasting the operation of three canals as a whole system. As mentioned in the introduction section, machine learning models recognize the patterns between the macro-description of the behavior of a system (in this case, a common reservoir delivers water to the three irrigation canals that belong to the same basin) and the behavior of the constituents of this system (each irrigation canal). Further research should be developed with more detailed analysis of advantage and disadvantage of building a single or combined model.

Researches have focused on real-time system for water resources and irrigation purposes (Lobbrecht and Solomatine 2002; Lobbrecht et al. 2005; De Sanctis et al. 2010). Kim et al. (2010) introduced a web-based wireless lysimeter network system and showed experimental results during an entire growing season in an agricultural field. Van Overloop et al. (2010) used a real-time predictive control model to deliver irrigation water to a canal. As mentioned in "Study area", a SCADA system has already been implemented in the lower Sevier River basin along with a web site that can be used as a delivery tool for real-time data display. This web site provides data retrieval from hourly and daily canal diversion for the previous 7 days (Berger et al. 2002, 2003).

For this paper, the models were trained and the optimal parameters were estimated in order to be used for real-time forecasting during the 2007 irrigation season (Figs. 4, 5). In the same way, the models can be used for future application in real time. The authors are currently developing standalone codes to fully implement the model in the

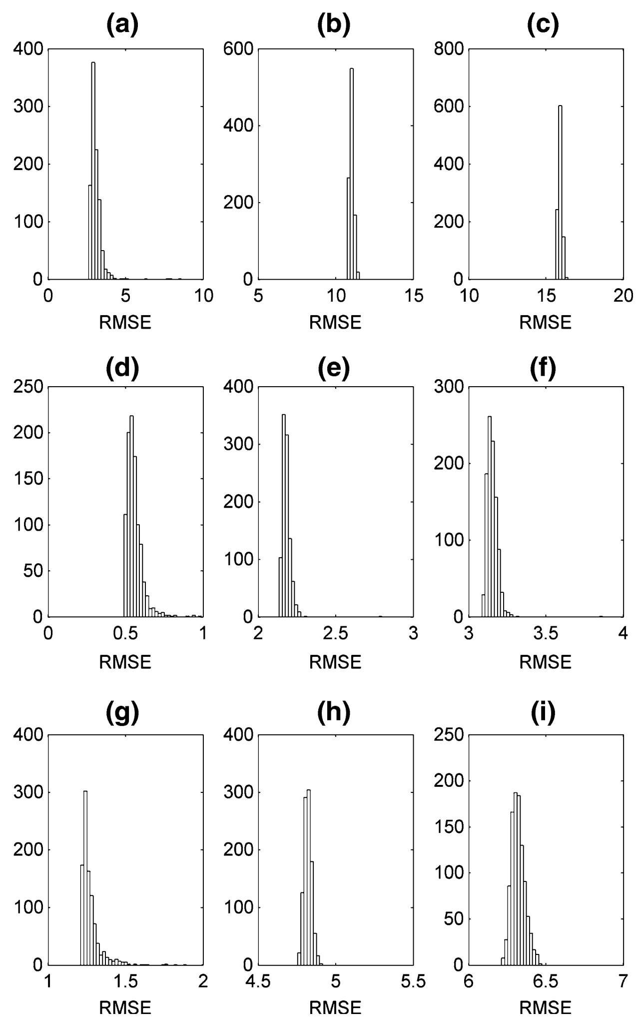


Fig. 8 Bootstrap histograms of the hourly MVRVM model for the RMSE test. Central Utah Canal: **a** 1 h ahead, **b** 12 h ahead, **c** 24 h ahead; Vincent Canal: **d** 1 h ahead, **e** 12 h ahead, **f** 24 h ahead; Leamington Canal: **g** 1 h ahead, **h** 12 h ahead, **i** 24 h ahead

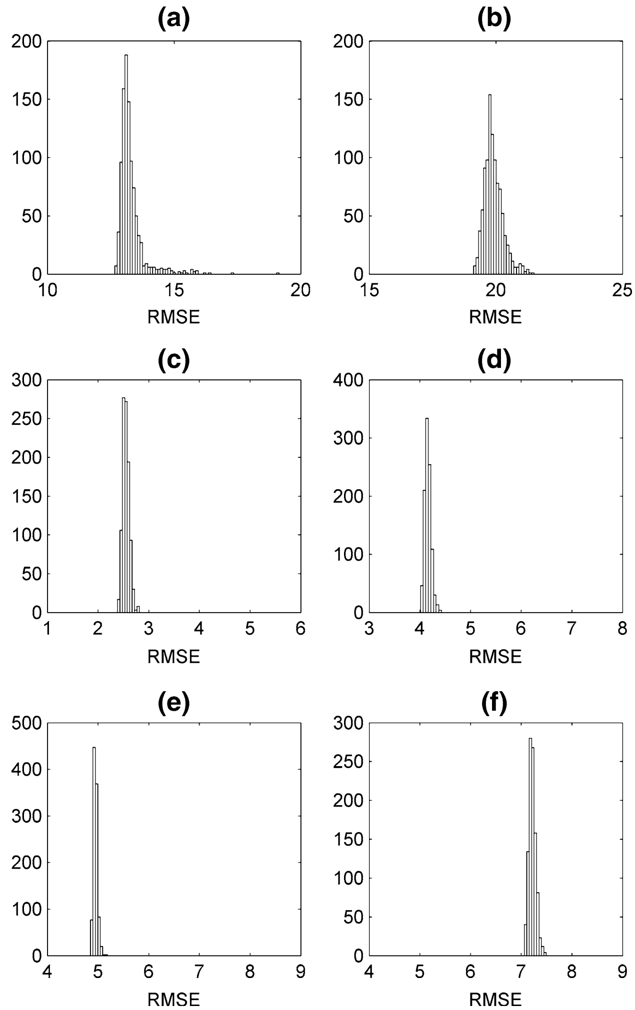


Fig. 9 Bootstrap histograms of the daily MVRVM model for the RMSE test. Central Utah Canal: **a** 1 day ahead, **b** 2 days ahead; Vincent Canal: **c** 1 day ahead, **d** 2 days ahead; Leamington Canal: **e** 1 day ahead, **f** 2 days ahead

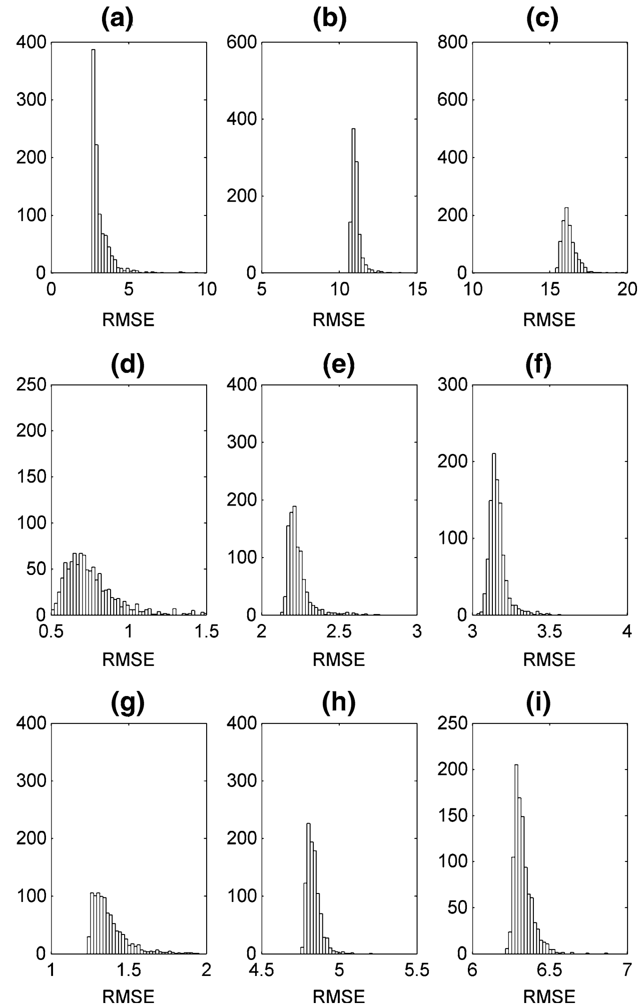


Fig. 10 Bootstrap histograms of the hourly ANN model for the RMSE test. Central Utah Canal: **a** 1 h ahead, **b** 12 h ahead, **c** 24 h ahead; Vincent Canal: **d** 1 h ahead, **e** 12 h ahead, **f** 24 h ahead; Leamington Canal: **g** 1 h ahead, **h** 12 h ahead, **i** 24 h ahead

Sevier River Water Users web site. A detailed work regarding this implementation in real time will be reported in a follow-up paper.

For the data set used in this paper, the MVRVM models resulted in relatively short training times. The training times for the models were approximately 6 and 2 min for the hourly and daily models, respectively (on an Intel Core i7 Windows Vista). It is important to recall that these training processes were based on 3,000 and 1,194 observations for the hourly and daily model, respectively. Moreover, each prediction cycle for both models (hourly and daily) is less than 0.05 s.

A key feature for real-time applications is that MVRVM models yield a solution that depends only on a very small number of relevance vectors (RVs). In the RVM approach, the RVs are subsets of the training data set that are used for prediction (Khalil et al. 2005b). As a consequence, the complexity of the model is proportional to the number of RVs. Same is true for MVRVM model. In our paper, the hourly model only utilizes 26 RVs from the full training data set (3,000 observations from the 2005 through 2006 irrigation seasons). The daily model utilizes 26 RVs from the full training data set (1,194 observations from the 2001 through 2006 irrigation seasons). This low number of relevance vectors illustrates that the Bayesian learning procedure embodied in the MVRVM is capable of producing very sparse models. This sparsity property of the MVRVM can be an effective method to control the computational characteristics of model performance (Tipping and Faul 2003), specially when dealing with large data sets. Therefore, MVRVM models are suitable for large-scale systems and real-time applications.

An important question arises regarding when the model should be retrained. Machine learning models should be retrained whenever the properties of the system change. In

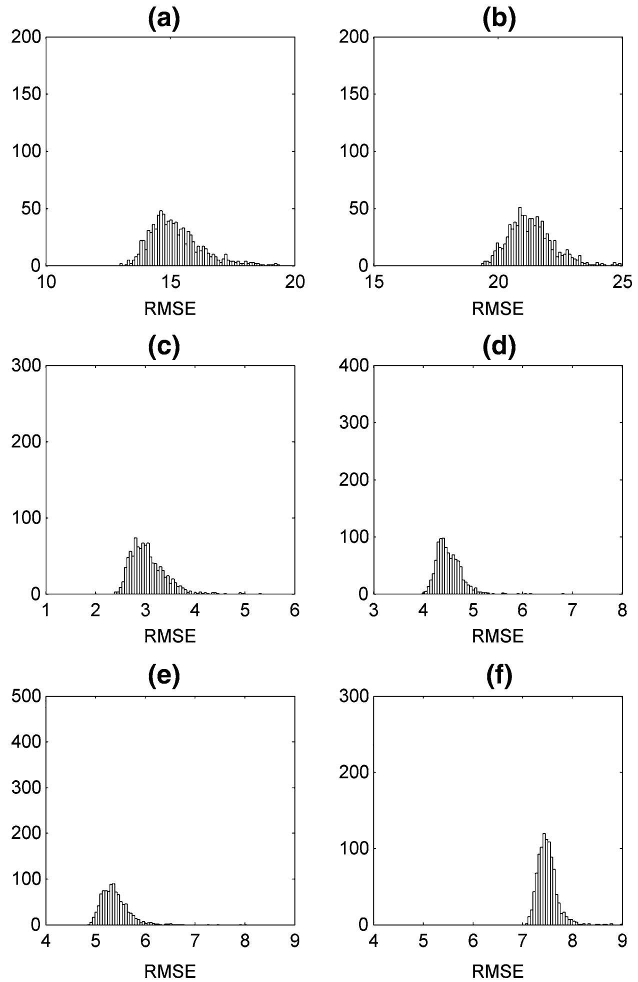


Fig. 11 Bootstrap histograms of the daily ANN model for the RMSE test. Central Utah Canal: **a** 1 day ahead, **b** 2 days ahead; Vincent Canal: **c** 1 day ahead, **d** 2 days ahead; Leamington Canal: **e** 1 day ahead, **f** 2 days ahead

section "Introduction", we mentioned that the demand at the three canal diversion depends mainly on two different behaviors that are related to a combination of natural and anthropogenic influences. These real processes can affect the irrigation diversion operation and can change in every irrigation season or even month. Therefore, a more detailed analysis of the model retraining and dynamic behavior of this special case of time series would need to be performed.

Figures 8, 9, 10 and 11 show the bootstrap histograms based on the 1,000 bootstrap training data sets of the MVRVM and ANN models for the RMSE test. The width of the bootstrapping confidence intervals provides implicit information on the uncertainty in the model parameters. A narrow confidence interval implies low variability of the statistics with future changes in the training data, which indicates that the model parameter set is robust, while a wide confidence interval indicates that future changes in the training data might be inadequate to find a robust parameter set (Khalil et al. 2005c). Most of the bootstrap histograms of the MVRVM model (Figs. 8, 9) show very narrow confidence bounds in comparison with the histograms of the ANN model (Figs. 10, 11). Therefore, the MVRVM allows more robust parameter estimation.

From the bootstrap results, the interquartile range (IQR) of the RMSE test was obtained in order to quantify and compare the robustness of the models. The IQR is a statistic that measures the range within which the middle half of the statistics (i.e., RMSE) falls. It is performed to reduce the chance of outliers affecting the results. It can be calculated by taking the difference between the 75th percentile and 25th percentile of the bootstrap results for the RMSE test. The IQR for the RMSE test for both models (MVRVM and ANN) are shown in Tables 5 and 6 for the

Table 5 Bootstrap IQR for the RMSE test for the hourly model

	MVRVM			ANN		
	Percentiles			Percentiles		
	25th	75th	IQR	25th	75th	IQR
Central Utah Canal RMSE (cfs)						
1 h	2.881	3.197	0.316	2.754	3.337	0.583
12 h	10.922	11.092	0.170	10.876	11.159	0.283
24 h	15.817	15.978	0.161	15.920	16.452	0.532
Vincent Canal RMSE (cfs)						
1 h	0.523	0.577	0.054	0.643	0.835	0.193
12 h	2.164	2.193	0.029	2.188	2.254	0.065
24 h	3.129	3.172	0.043	3.129	3.185	0.056
Leamington Canal RMSE (cfs)						
1 h	1.235	1.289	0.054	1.296	1.422	0.126
12 h	4.804	4.836	0.032	4.803	4.855	0.052
24 h	6.289	6.347	0.058	6.285	6.353	0.068

Table 6 Bootstrap IQR for the RMSE test for the daily model

	MVRVM			ANN		
	Percentiles			Percentiles		
	25th	75th	IQR	25th	75th	IQR
Central Utah Canal RMSE (cfs)						
1 day	13.027	13.372	0.344	14.549	16.086	1.537
2 day	19.666	20.103	0.437	20.773	22.142	1.369
Vincent Canal RMSE (cfs)						
1 day	2.503	2.588	0.085	2.794	3.236	0.442
2 day	4.114	4.196	0.082	4.334	4.656	0.322
Leamington Canal RMSE (cfs)						
1 day	4.923	4.971	0.048	5.198	5.534	0.336
2 day	7.173	7.261	0.089	7.355	7.585	0.230

hourly and daily models, respectively. We can see that, for all cases, the IQR for the MVRVM is much lower than the IQR for the ANN. This lower IQR implies that the MVRVM models show lower variability in the RMSE test with respect to possible future changes in the nature of the input data, which indicates that the MVRVM is more robust than the ANN model.

Conclusions

In this research, we have proposed a real-time forecasting model for proper irrigation water delivery in the lower Sevier River basin in Utah. The main water-delivery problem in the basin is the inefficient operator responses to short-term changes in demand due to the inherent lag time between a reservoir flow release and its arrival at the diversion irrigation canals. In this context, the potential benefit of accurate anticipations of short-term irrigation canal demand lies in assisting the reservoir and canal operators in making efficient real-time operation and management decisions for available water resources in the basin.

MVRVM models used in this research provide multiple-time-ahead forecasts of required hourly and daily diversions for a system of multiple irrigation canals. The MVRVM is a Bayesian regression tool extension of the RVM algorithm that produces multivariate outputs (with predictive confidence intervals) when given a set of multivariate inputs.

The test results presented in this paper have demonstrated the successful performance of MVRVM for multiple irrigation demand forecasts. The MVRVM models were compared in terms of performance and robustness with ANN models. The performance results of both the MVRVM and ANN were fairly similar; however, a bootstrap analysis showed that MVRVM models were more robust than ANN models.

Acknowledgments The authors would like to thank the Utah Water Research Laboratory (UWRL) and the Utah Center for Water Resources Research (UCWRR) for support for this research. We also thank Professor Inga Maslova, of American University, for helpful comments and discussions. The authors are grateful to Dr. Roger Hansen of the Provo, Utah, office of the US Bureau of Reclamation and Jim Walker of the Sevier River Water Users Association. We also thank two anonymous reviewers whose comments significantly contributed to the improvements in the manuscript.

References

Adeloye AJ (2009) Multiple linear regression and artificial neural networks models for generalized reservoir storage-yield-reliability function for reservoir planning. J Hydrol Eng 14(7):731–738

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

ASCE Task Committee on the Application of ANNs in Hydrology (2000a) Artificial neural networks in hydrology, I: preliminary concepts. J Hydrol Eng 5(2):115–123

ASCE Task Committee on the Application of ANNs in Hydrology (2000b) Artificial neural networks in hydrology, II: hydrologic application. J Hydrol Eng 5(2):124–137

Berger B, Hansen R, Hilton A (2002) Using the World-Wide-Web as a support system to enhance water management. Paper presented at the 18th ICID congress and 53rd IEC meeting, International Committee on Irrigation and Drainage, Montreal, Quebec, Canada

Berger B, Hansen R, Jensen R (2003) Sevier River Basin System Description. Sevier River Water Users Association, Delta

Bishop CM (1995) Neural network for pattern recognition. Oxford University Press, USA

Cherkassky V, Krasnopolsky V, Solomatine DP, Valdes J (2006) Computational intelligence in earth sciences and environmental applications: issues and challenges. Neural Netw 19:113–121

De Sanctis AE, Shang F, Uber JG (2010) Real-time identification of possible contamination sources using network backtracking methods. J Water Resour Plan Manage 136(4):444–453

Demuth H, Beale M, Hagan M (2009) Neural network toolbox user's guide. The MathWorks Inc, USA

Duda RO, Hart P, Stork D (2001) Pattern classification, 2nd edn. Wiley Interscience, USA

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

Khalil A, Almasari M, McKee M, Kemblowski MW, Kaluarachchi J (2005a) Applicability of statistical learning algorithms in groundwater quality modeling. Water Resour Res 41:W05010

Khalil A, McKee M, Kemblowski MW, Asefa T (2005b) Sparse Bayesian learning machine for real-time management of reservoir releases. Water Resour Res 41:W11401

Khalil A, McKee M, Kemblowski MW, Asefa T, Bastidas L (2005c) Multiobjective analysis of chaotic dynamic systems with sparse learning machines. Adv Water Res 29:72–88

Khalil A, McKee M, Kemblowski MW, Asefa T (2005d) Basin-scale water management and forecasting using neural networks. J Am Water Resour Res 41:195–208

Kim Y, Jabro JD, and Evans RG (2010) Wireless lysimeters for real-time online soil water monitoring. Irrig Sci. doi:10.1007/s00271-010-0249-x

Kisi O (2009) Daily pan evaporation modeling using multi-layer perceptrons and radial basis neural networks. Hydrol Process 23:213–223

Krasnopolsky VM (2009) Neural network applications to solve forward and inverse problems in atmospheric and oceanic satellite remote sensing. Artif Intell Methods Environ Sci 9:191–205

Kumar A, Bnadyopadhyay A, Raghuwanshi NS (2008) Comparative study of conventional and artificial neural network-based ETo estimation models. Irrig Sci 26:531–545

Kumar A, Raghuwanshi NS, Singh R (2011) Artificial neural networks approach in evapotranspiration modeling: a review. Irrig Sci 29:11–25

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

Lobbrecht AH, Solomatine DP (2002) Machine learning in real-time control of water systems. Urban Water 4:283–289

Lobbrecht AH, Dibike YB, Solomatine DP (2005) Neural networks and fuzzy systems in model based control of the overwaard polder. J Water Resour Plan Manage 131(2):135–145

Nourani V, Mehdi K, Akira M (2009) A multivariate ANN-wavelet approach for rainfall–runoff modeling. Water Resour Manage 23:2877–2894

Parker RL (1994) Geophysical inverse theory. Princenton University Press, Princenton

Pulido-Calvo I, Gutierrez-Estrada JC (2008) Improved irrigation water demand forecasting using a soft-computing hybrid model. Biosyst Eng 102(2):202–218

Pulido-Calvo I, Roldan J, Lopez-Luque R, Gutierrez-Estrada JC (2003) Demand forecasting for irrigation water distribution systems. Irr Drain Eng 129(6):422–431

Rahimi AK (2008) Artificial neural network estimation of reference evapotranspiration from pan evaporation in a semi-arid environment. Irrig Sci 27:35–39

Renault D, Facon T, Wahaj R (2007) Modernizing irrigation management, the MASSCOTE approach. FAO Irrigation and Drainage Paper 63, Rome

Shumway RH, Stoffer DS (2011) Time series analysis and its applications, 3rd edn. Springer, USA

Tabari H, Marofi S, Sabziparvar AA (2010) Estimation of daily pan evaporation using artificial neural network and multivariate non-linear regression. Irrig Sci 28:399–406

Tan T, Van Cauwenberghe A (1999) Neural-network-based d-step-ahead predictor for nonlinear systems with time delay. Eng Appl of Artif Intell 12:21–35

Thayananthan A (2005) Template-based pose estimation and tracking of 3D hand motion. Dissertation, University of Cambridge, UK

Thayananthan A, Navaratnam R, Stenger B, Torr PHS, 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 Manage 25:523–543

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 A (2003) Fast marginal likelihood maximization for sparse Bayesian models. Paper presented at ninth international workshop on artificial intelligence and statistics, Society for Artifical Intelligence and Statistics, Key West, Florida

Tripathi S, Govindaraju R (2007) On selection of kernel parameters in relevance vector machines for hydrologic applications. Stoch Eviron Res Risk Asses 21:747–764

Van Overloop PJ, Clemmens AJ, Strand RJ, Wagemaker RMJ (2010) Real-time implementation of model predictive control on Maricopa-Stanfield irrigation and drainage district’s WM Canal. J Irr Drain Eng 136(11):747–756

Wang J, Neskovic P, Cooper LN (2005) Training data selection for support vector machines. In: Wang L, Chen K, Ong YS (eds) Advances in natural computation. Springer, Berlin, pp 554–564