WATER RESOURCES RESEARCH, VOL. 41, W12422, doi:10.1029/2004WR003785, 2005

Support vector machines for nonlinear state space reconstruction: Application to the Great Salt Lake time series

Tirusew Asefa,[1] Mariush Kemblowski,[2] Upmanu Lall,[3] and Gilberto Urroz[2]

Received 3 November 2004; revised 8 August 2005; accepted 6 September 2005; published 20 December 2005.

[1] The reconstruction of low-order nonlinear dynamics from the time series of a state variable has been an active area of research in the last decade. The 154 year long, biweekly time series of the Great Salt Lake volume has been analyzed by many researchers from this perspective. In this study, we present the application of a powerful state space reconstruction methodology using the method of support vector machines (SVM) to this data set. SVM are machine learning systems that use a hypothesis space of linear functions in a kernel-induced higher-dimensional feature space. SVM are optimized by minimizing a bound on a generalized error (risk) measure rather than just the mean square error over a training set. Under Mercer's conditions on the kernels the corresponding optimization problems are convex; hence global optimal solutions can be readily computed. The SVM-based reconstruction is used to develop time series forecasts for multiple lead times ranging from 2 weeks to several months. Unlike previously reported methodologies, SVM are able to extract the dynamics using only a few past observed data points out of the training examples. The reliability of the algorithm in learning and forecasting the dynamics is tested using split sample sensitivity analysis, with a particular interest in forecasting extreme states. Efforts are also made to assess variations in predictability as a function of initial conditions and as a function of the degree of extrapolation from the state space used for learning the model.

Citation: Asefa, T., M. Kemblowski, U. Lall, and G. Urroz (2005), Support vector machines for nonlinear state space reconstruction: Application to the Great Salt Lake time series, *Water Resour. Res.*, *41*, W12422, doi:10.1029/2004WR003785.

1. Introduction

[2] The reconstruction of low-order nonlinear dynamics from the time series of a state variable has been an active area of research in the last decade. Dynamic reconstruction is the problem of approximating an unknown function that describes the state evolution of a chaotic system [*Abarbanel*, 1996]. It is an ill-posed inverse problem [*Tikhonov and Arsenin*, 1977]. Because of this ill-posedness the solution may not be unique, stable, or even exist. Therefore, when tackling a dynamic reconstruction problem, it is necessary to use some form of regularization in order to limit the functional space from which the reconstructed mapping function is obtained.

[3] Application of chaos theory in hydrology has increased considerably. However, most of the studies are concerned with the investigation of the existence of chaos in rainfall [e.g., *Hense*, 1987; *Rodriguez-Iturbe et al.*, 1989] and streamflows [e.g., *Jayawardena and Lai*, 1994]. Others have looked at predictions [e.g., *Lall et al.*, 1996; *Jayawardena et al.*, 2002; *Elshorbagy et al.*, 2002]. We refer the reader to *Sivakumar* [2000] for the review of chaos theory in hydrology.

[4] In what follows we present studies of chaos in the Great Salt Lake (GSL) time series and investigate the applicability of state-of-the-art learning tools called support vector machines (SVM) in learning and predicting the low-dimensional chaotic GSL volume time series. We present short-term GSL volume predictions at different time steps and highlight attractive features of these new learning tools compared with those previously used in reconstructing GSL time series.

2. Data and Method

2.1. Great Salt Lake

[5] The Great Salt Lake, located in Utah, United States of America, is the fourth largest terminal lake (with no outlet to the sea) in the world, with a drainage area of 90,000 km^2 (Figure 1a). Streams contribute ~66% of the annual inflow to the lake, precipitation is estimated to be ~31%, and 3% is groundwater contribution [*Arnow*, 1984]. The lake drains three states: Utah, Idaho, and Wyoming. The modern era record-breaking rise of GSL level between 1982 and 1986 resulted in severe economic impact. The increase in area covered by the lake from September 1982 to July 1983 was ~267 km^2 and resulted in damage to roads, railroads, recreation facilities, and industrial installation on the lake bed [*Arnow*, 1984]. According to the record level estimated by Utah Division of the State Land and Forestry the costs

[1]Department of Source Rotation and Environmental Protection, Tampa Bay Water, Clearwater, Florida, USA.
[2]Department of Civil and Environmental Engineering and Utah Water Research Laboratory, Utah State University, Logan, Utah, USA.
[3]Department of Earth and Environmental Engineering and International Research Institute for Climate Prediction, Columbia University, New York, New York, USA.

0043-1397/05/2004WR003785$09.00

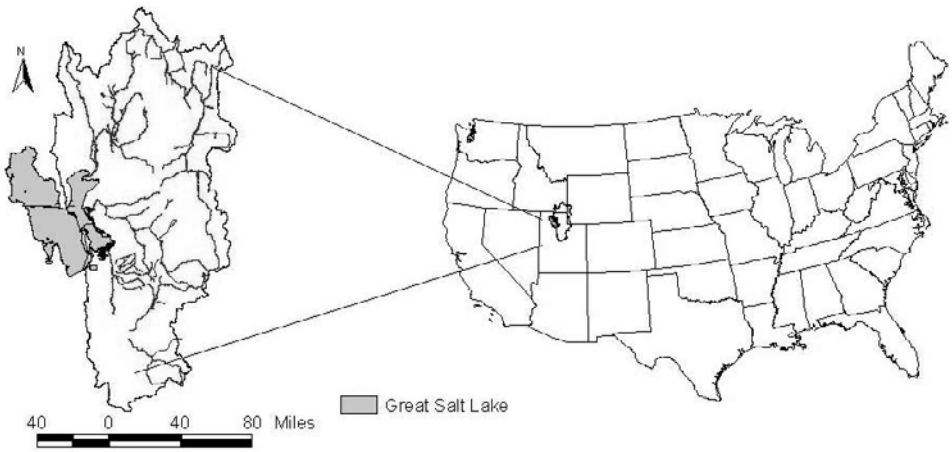

Figure 1a. Study area (41°03′35″N, 112°15′08″W). See color version of this figure in the HTML.

associated with damages and protection of these facilities as the lake rose was about $150 million. Soon GSL levels fell and pumps installed to alleviate flooding were left high and dry, seemingly solving the problem by itself.

[6] Traditional linear time series analysis models were found to be insufficient to adequately describe the dramatic rise and fall of GSL volume [*Lall et al.*, 1996]. One reason for such inadequacy may be the fact that there is not enough information prior to an event within just projected one-dimensional time series output of the system. This opens up the possibility of investigating whether there is a set of differential equations that could generate a single time series and hence the possibility of investigating chaos by unfolding (reprojecting) the dynamics through representation of the data in multidimensional state space. The observed one-dimensional 1848–2002 GSL volume time series is shown in Figure 1b. Figure 1b shows the dramatic rise and fall of GSL volume at different times throughout the measurement period.

[7] Several studies have contributed to the current understanding of the GSL volume dynamics [*Lall and Mann*, 1995; *Moon and Lall*, 1996; *Abarbanel et al.*, 1996; *Abarbanel and Lall*, 1996; *Mann et al.*, 1995; *Sangoyomi et al.*, 1996]. Dimension calculation, estimates of Lyapunov

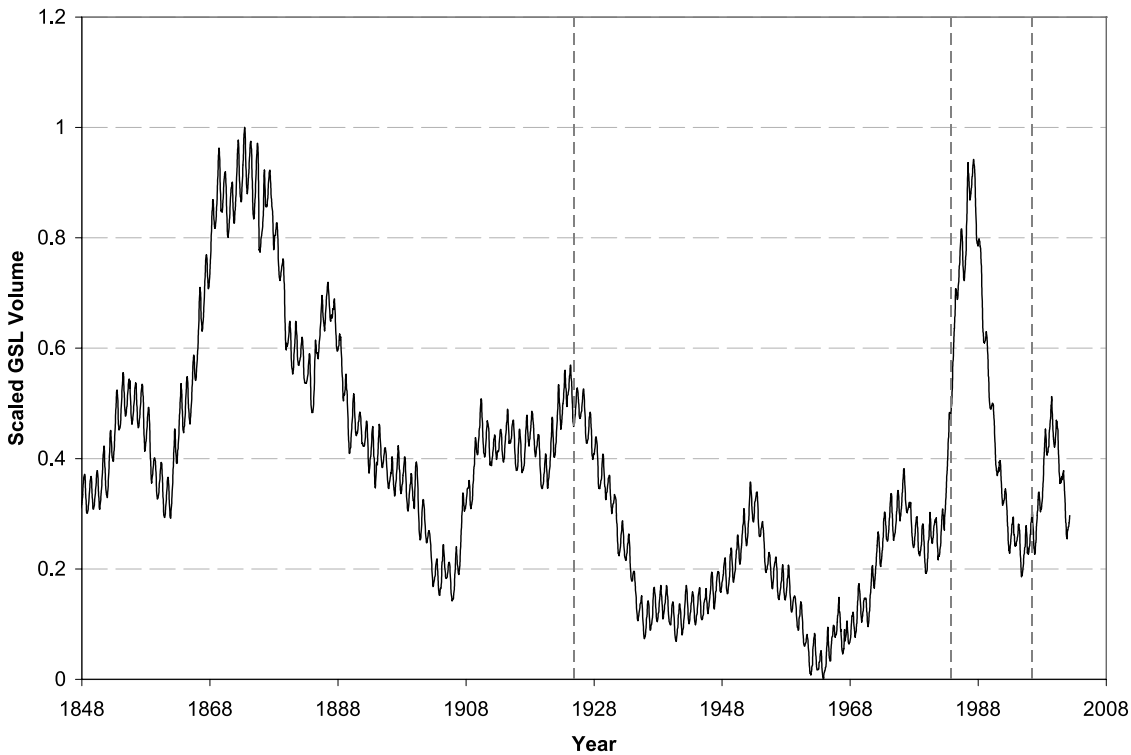

Figure 1b. Great Salt Lake (GSL) volume time series (1848–2002). Data are scaled with respect to maximum volume of 3.77×10^{10} m^3 and minimum volume of 1.05×10^{10} m^3. The dashed lines show where predictions of GSL fall and rise were conducted.

exponents for predictability assessment, tests for determinism, and tests for nonlinearity results of the above studies suggest that the GSL time series may correspond to a low-dimensional (d = 4), nonlinear, dynamical system. On the basis of the existence of positive Lyapunov exponents, different forecasting methods were utilized to predict the short-term behavior of GSL volume. The methods used were multivariate adaptive regression splines [*Friedman*, 1991; *Lall et al.*, 1996], local polynomials [*Moon*, 1995; *Abarbanel et al.*, 1996], and artificial neural networks [*Coulibaly and Baldwin*, 2005]. Even though these methodologies were successful in predicting the wet and dry conditions, the present study, unlike these previous methodologies, demonstrates a dynamical reconstruction algorithm that only uses a few samples of the training data and computes a global optimum of the objective function. This has important computational implications.

2.2. Nonlinear Dynamic System Reconstruction From Time Series of a State Variable

[8] Chaos occurs as a feature of orbits $\mathbf{v}(t)$ arising from nonlinear evolution rules which are systems of differential equations of

$$\frac{d\mathbf{v}(t)}{dt} = \mathbf{F}[\mathbf{v}(t)] \tag{1}$$

with three or more degree of freedom or invertible maps of

$$\mathbf{v}(t+1) = \mathbf{F}[\mathbf{v}(t)]. \tag{2}$$

As a class of observable signals, $\mathbf{v}(t)$, chaos lies logically between the well-studied domain of predictable, regular, or quasiperiodic signals and the totally irregular stochastic signals we call "noise" that is completely unpredictable [*Abarbanel*, 1996]. In our case, seldom do we have measurements of all the relevant dynamic variables (components of GSL volume, $\mathbf{v}$). Even if we do, we do not know the underlying physical and/or other processes responsible for the evolution of GSL volume dynamics. Instead, what we have is the system output measurements of a single variable. Hence one usually tries to construct a multivariate state space in which the dynamics unfold by creating vectors in (embedding) dimension d out of the single measurement, v(t). The purpose of embedding is to create a pseudo state space, called the reconstruction space, in which the dynamics of the original chaotic system can be reconstructed. *Takens* [1981] introduced a method of state space construction for recovering dynamics of a system from observed scalar quantities using delay coordinates

$$\mathrm{v}_t = \left(v_t, v_{t-\tau}, \ldots, v_{(d-1)\tau}\right), \tag{3}$$

where τ is an integer lag that models the delay. Then dynamic predictions will consist of estimating the state variable at some future time, T,

$$v_{t+T} = f_T\left(v_t, v_{t-\tau}, \ldots, v_{(d-1)\tau}\right). \tag{4}$$

The T step reconstructed mapping function, f_T, will be found by minimizing a regularized risk functional. Support vector machines have been proven useful in such cases because of their inherent regularization capabilities (see discussion in section 3). Future state variables at different time steps may be forecasted in two ways: (1) using a direct T step ahead prediction found by estimating the mapping f_T and (2) using iterated one step ahead forecast, f_t. We have used both approaches and will demonstrate that as T increase the quality of prediction using the first method decreases while the second method provides a better prediction.

3. Support Vector Machines

[9] In their present form, SVM for classification and regression problems were developed at AT&T Bell Laboratory by V. Vapnik and coworkers in the early 1990s [e.g., *Boser et al.*, 1992; *Vapnik*, 1998]. Since then, there have been a growing number of SVM applications in the statistics, computer science, and other fields. Recently, SVM have been introduced in surface and subsurface hydrology. *Dibike et al.* [2001] applied SVM successfully in both remotely sensed image classification and regression (rainfall/runoff modeling) problems and reported a superior performance over the traditional artificial neural networks (ANN). *Kanevski et al.* [2002] used SVM for mapping soil pollution by Chernobyl radionuclide Sr90 and concluded that the SVM were able to extract spatially structured information from the row data. *Liong and Sivapragasam* [2002] also reported a superior SVM performance compared with ANN in forecasting flood stage. *Asefa et al.* [2004] presented a support vector–based long-term groundwater head monitoring network designing approach where SVMs were used to learn a potentiometric surface. Below we will give a brief introduction. We refer interested readers to *Vapnik* [1998] for detailed treatment of the subject.

[10] Let $\hat{f}(\mathbf{v})$ be a GSL volume estimator that explains functional dependency between (input) state space values $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ taken from $\mathbf{v} \in \mathbf{R}^d$, where d is the embedding dimension; and its future scalar state variables $\{y_1, y_2, \ldots, y_n\}$ with $y \in \mathbf{R}$ drawn from a set of N-independent and identically distributed observations. One looks for $\hat{f}(\mathbf{v})$ by minimizing the following regularized risk function [*Vapnik*, 1998]:

$$\min \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{N}(\xi_i + \xi_i^*) \tag{5a}$$

subject to

$$\begin{aligned} f(\mathbf{v}) - \langle \mathbf{w}, \mathbf{v}\rangle - b &\le \varepsilon + \xi_i \\ \langle \mathbf{w}, \mathbf{v}\rangle + b - f(\mathbf{v}) &\le \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* &\ge 0 \end{aligned} \tag{5b}$$

to obtain

$$\hat{f}(\mathbf{x}) = \langle \mathbf{w}, \mathbf{v}\rangle + b. \tag{5c}$$

Here $\mathbf{w}$ are the support vector weights (basis functions), angle brackets denote the dot product in $\mathbf{v}$, and b is a bias; ξ_i and ξ_i^* are slack variables that determine the degree to which state space samples with error more than ε be penalized; and ε is the degree to which one would like to tolerate errors in constructing the predictor $\hat{f}(\mathbf{v})$ (Figure 2). Because of this the above formulation is also referred to as

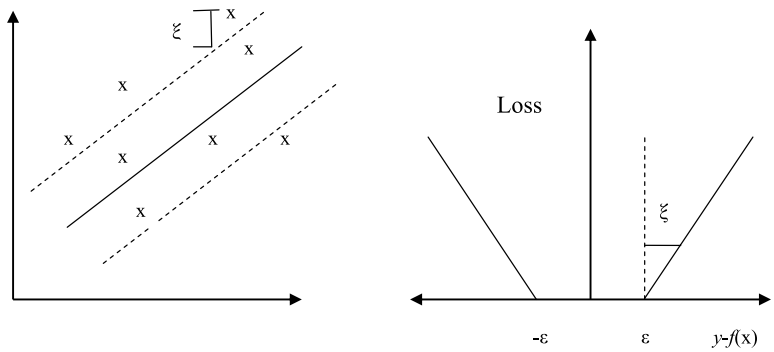

Figure 2. The ε-insensitive loss function Γ.

the ε-insensitive approach. The objective function given in equation (5) minimizes the complexity of the GSL estimator (i.e., the estimator will tend to be flat), rendering regularization of the solution, and penalizes state space data points that lie outside the ε tube (goodness of fit). In other words, for any (absolute) error smaller than ε, ξ_i, = $\xi_i^* = 0$; hence these data points do not enter the objective function. This means that not all reconstructed state space points will be used to estimate $\hat{f}(\mathbf{v})$. The constant $C > 0$ determines the trade-off between the complexity of f and the amount up to which deviations larger than ε are tolerated.

[11] Usually, the optimization problem given in equation (5) is solved in its dual form using Lagrange multipliers. Maximizing equation (5) in its dual form, differentiating with respect to primal variables ($\mathbf{w}$, b, ξ_i, ξ_i^*), and rearranging gives the following [see *Asefa et al.*, 2004, Appendix]:

$$W(\alpha^*, \alpha) = -\varepsilon \sum_{i=1}^{N} (\alpha_i + \alpha_i^*) + \sum_{i=1}^{N} Z_i (\alpha_i - \alpha_i^*) - \frac{1}{2} \sum_{i,j=1}^{N} \times (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) k(\mathbf{v}_i, \mathbf{v}_j) \quad (6a)$$

subject to constraints

$$\sum_{i=1}^{N} (\alpha_i^* - \alpha_i) = 0, \qquad 0 \le \alpha_i, \alpha_i^* \le C, \quad (6b)$$

to obtain

$$\hat{f}(\mathbf{v}) = \sum_{i=1}^{n} (\alpha_i^* - \alpha_i) k(\mathbf{v}, \mathbf{v}_i) + b, \quad (6c)$$

where α_i^* and α_i are Lagrange multipliers, $k(\mathbf{v}, \mathbf{v}_i)$ is a kernel that measures nonlinear dependence between two state space realizations, and n is the number of selected state space points that explain the underlying dynamic relationship. From the Kuhn-Tucker condition it follows that only for $|\hat{f}(\mathbf{v}) - y_i| \ge \varepsilon$ the Lagrange multipliers may be nonzero. In other words, for state space points inside the ε tube (Figure 2), α_i and α_i^* would vanish. Those data points that have no vanishing coefficients are called "support vectors," hence the name support vector machines. Intuitively, one

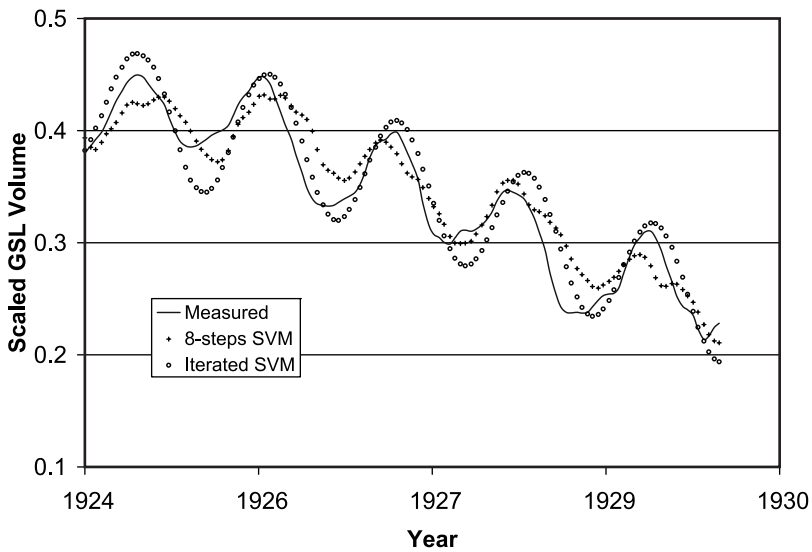

Figure 3. GSL fall forecast. Eight steps (4 months) prediction assumed that historical data were available at the time of forecast, while the iterated predictions did not use new data once prediction started. In the latter case, output from a given time step were fed back to the support vector machines (SVM).

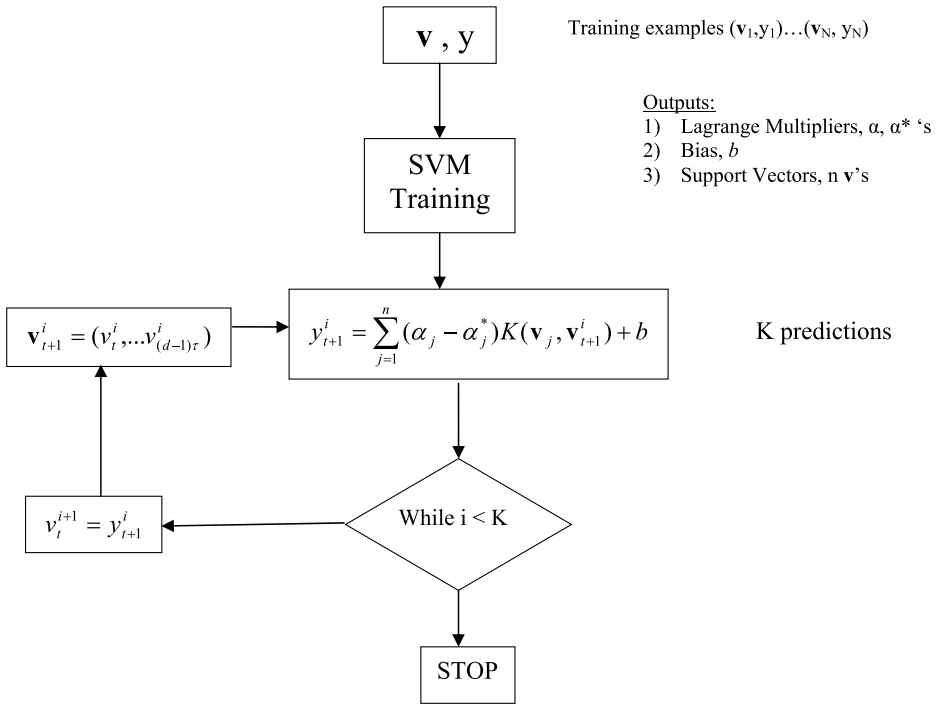

Figure 4. Architecture of (autonomous) iterated SVM. Once prediction starts, no new observation data are fed to the machine.

can imagine the support vectors as "smart" state space points that "support" the estimated GSL predictor. Note that the upper limit in summation in equation (6c) goes to the number of support vectors n (usually $n \ll N$). This is a sparse approximation to function estimation [*Girosi*, 1998] where only a small number of coefficients will be different from zero. This has important computational implications.

4. Results and Discussions

[12] Previous studies have shown that the GSL correlation and nearest-neighbor dimensions are ~4 [*Abarbanel et al.*, 1996; *Sangoyomi et al.*, 1996]. The value of the time delay parameter, τ, from average mutual information results (at which successive values of GSL measurements are somewhat independent) [*Abarbanel et al.*, 1996] suggests the use of τ between 8 and 17. An embedding dimension of 4 would therefore unfold the attractor, and no orbit crossing remains that would deteriorate predictions. Using a bigger dimension than $d = 4$, on the other hand, can improve predictions, while using a smaller dimension deteriorates them. For example, *Müller et al.* [1999] obtained a better prediction for the data set D from the Santa Fe competition [*Weigend and Gershenfeld*, 1994] by using an embedding dimension of 20, while the artificial data were generated from a nine-dimensional periodically dissipative dynamical system with slight drift on the parameters. An embedding dimension of $2d + 1$ is sufficient to unfold the trajectories where d is the actual system dimension, so there is a hypothetical upper bound for clean data. In all the following reported cases we used a radial basis kernel. Optimal SVM parameters were obtained on the basis of cross-validation results on training data. We refer readers to *Asefa et al.* [2004] for detailed exposition of SVM parameters estimation procedure.

4.1. Predicting GSL Volume Decline

[13] We will first attempt to forecast GSL volume decline from 15 December 1924 until 15 December 1929 (5 years prediction). An embedding dimension $d = 8$ and time delay $\tau = 11$ that gave the best model based on minimum predictive root mean square error was used to produce the results shown in Figure 3.

[14] Training started on 15 September 1893 (index 1) and ended 1 December 1924 (index 750). The best parameters of the SVM were $C = 0.13$, $\varepsilon = 0.02$, and $\sigma^2 = 0.03$. Two types of predictions are depicted in Figure 3: eight steps (4 months) ahead predictions that assumed that observed data were available at the time of forecast and iterated prediction that forecasts GSL volume one step ahead where computed values were, in turn, used to make the next step prediction. In the latter case, no new (observed) data were used once prediction started. Figure 4 shows the architecture of these (autonomous) iterated SVM. The SVM were able to extract only a very small percentage of state space points from the training data (less than 5%) in order to explain the underlying dynamical relationship for the next 5 years of prediction (121 iterated steps), effectively capturing transition from "average" to "low" GSL regimes. Comparable results were also found using $d = 9$ and $\tau = 8$. The use of a complete time series (e.g., 1773 for one step ahead prediction corresponds to about 73 years of GSL volume observations) did not improve model prediction much. This is especially true for longer time step predictions. For example, eight steps or 4 months ahead predictions deteriorated when complete time series were used. This runs counter to

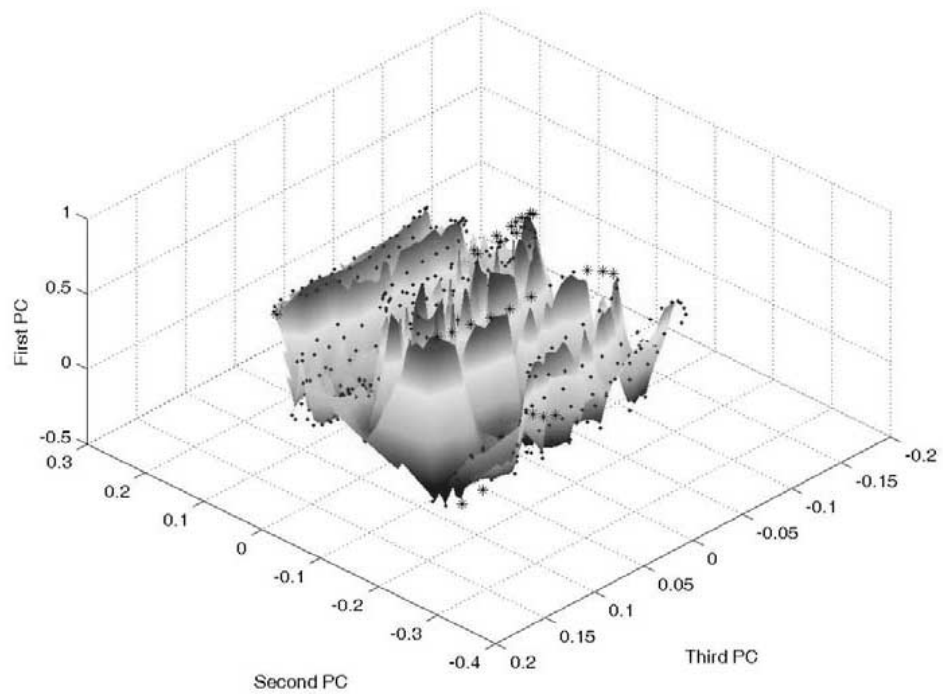

Figure 5a. Training data (dots) and support vectors (stars) for one step iterated predictions. See color version of this figure in the HTML.

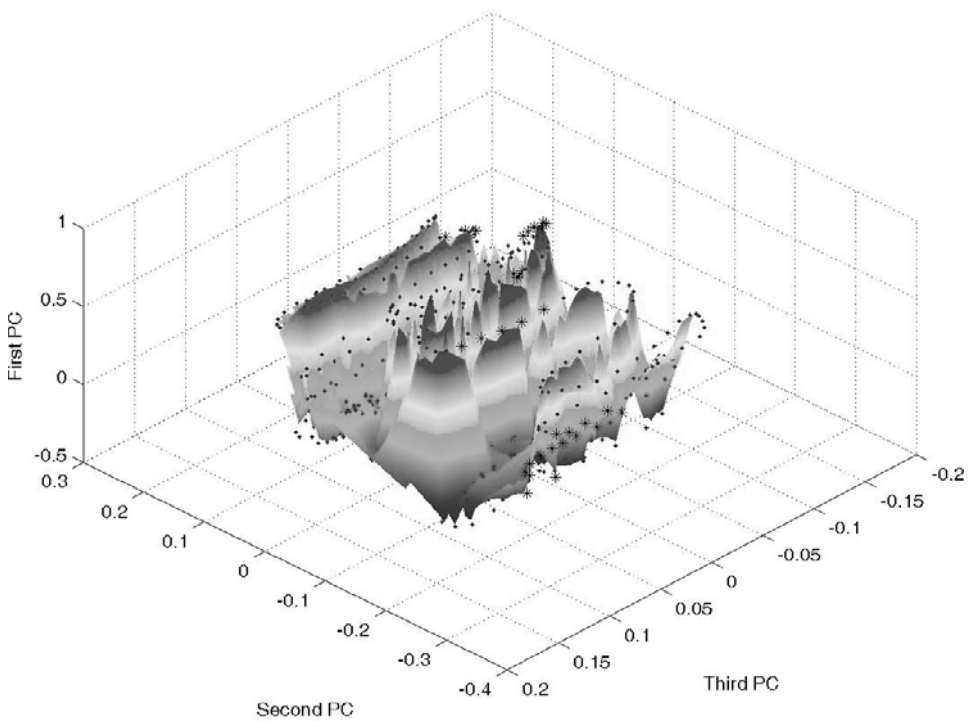

Figure 5b. Training data (dots) and support vectors (stars) for eight steps ahead (4 months) predictions. See color version of this figure in the HTML.

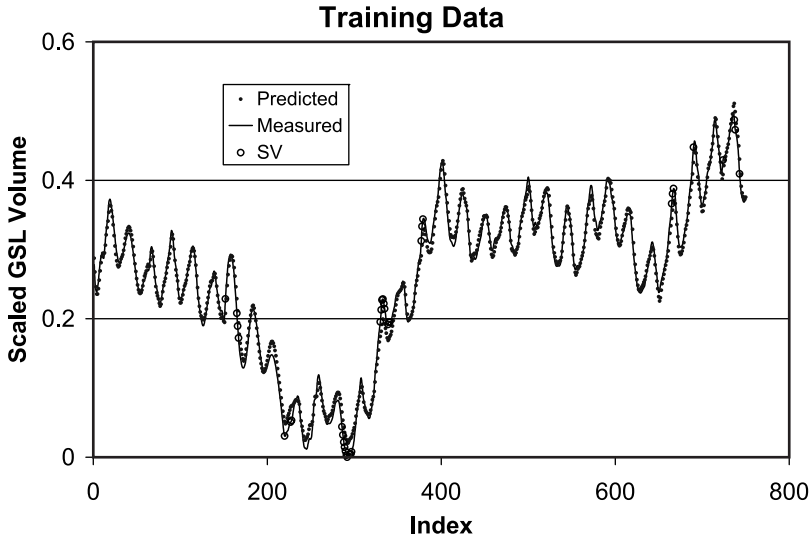

Figure 5c. Training data (dots) and support vectors (circles) for one-dimensional training data and associated support vectors.

the usual notion of asymptotic convergence where use of longer training data is expected to improve model predictions. One reason for such behavior might be the chaotic characteristics of the GSL time series and its sensitivity to initial conditions.

[15] Figures 5a–5c show the locations of extracted support vectors (SV) together with the rest of training points. For the sake of visualization the first three principal components of reconstructed state space are plotted. Figure 5a shows the locations of 37 SV corresponding to iterated prediction, whereas Figure 5b shows the locations of 90 SV associated with eight steps (4 months) prediction. Figure 5c provides the one-dimensional plot of the iterated SVM with associated support vector locations. It is interesting to observe that the SV are at "smart" locations in the convex hull of the embedding. This is one of the interesting features of SVM compared with other learning algorithms like artificial neural networks in the sense that locations of support vectors (data points that are used to learn underlying physical or other relationships) provide insight into studies of process approximation. For example, in designing groundwater head monitoring networks, *Asefa et al.* [2004] used SVM to approximate a potentiometric surface, and the support vectors were found to be located at the most uncertain areas of the groundwater watershed. State space points that are selected as support vectors are the only training data points that are used to predict future state values. The rest of the training data will not be used.

4.2. Predicting the Rise of GSL Volume

[16] The dramatic modern era GSL rise, flooding major highways, and its subsequent fall occurred during the historical period 1983–1987. Looking at pre-1983 time series, one would not expect such a rise. Traditional linear time series models (autoregressive, autoregressive moving average, and autoregressive integrated moving average) have failed to predict this event [*Lall et al.*, 1996]. Figure 6 shows SVM predictions for GSL volume rise

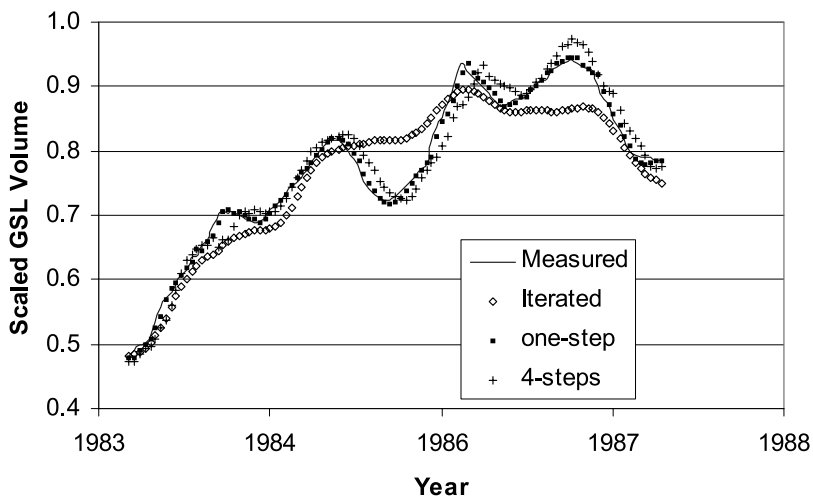

Figure 6. Historical rise of GSL volume. One step predicted 2 weeks ahead; four steps predicted monthly volumes. The iterated (autonomous) prediction did not use any new data once prediction started.

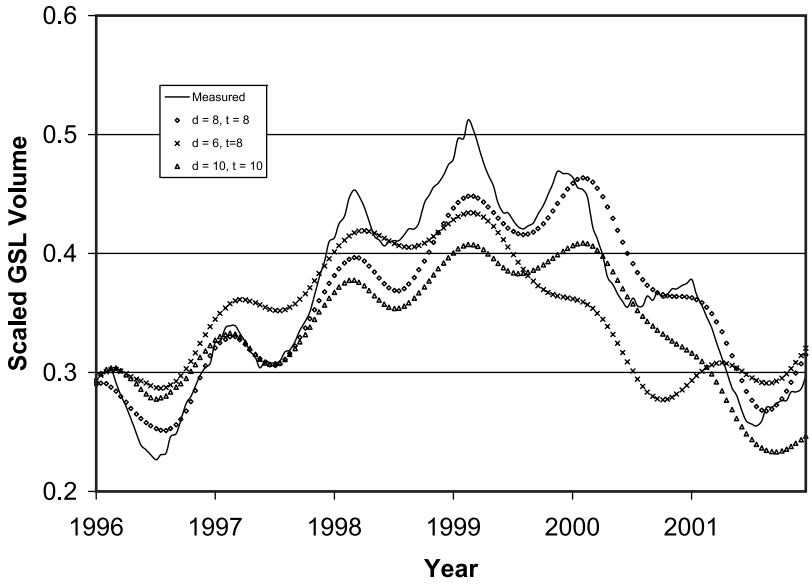

Figure 7. Continuous 6 year GSL prediction. The one step past observed values at the time of prediction, while the iterated prediction did not use new data once prediction started. A dimension of 8 and lag of 8 was used for the plot.

starting 1 October 1983. The best model based on predictive root mean square error used an embedding dimension of 10 and lag time of 8. Seven hundred fifty training examples were used. The one to four steps ahead predictions did an excellent job of capturing this dramatic rise. The autonomous SVM that did not see any new (observed) data once prediction started also gave a good performance. One has to realize that what is seen in Figure 6 is a very rarely visited piece of the attractor that is more challenging for getting accurate predictions.

4.3. Predicting Recent Events

[17] Figure 7 shows continuous prediction for the last 6 years of the observed data (from 1 May 1996 to 15 April 2002). Seven hundred fifty examples were used for training. The autonomous SVM prediction also effectively captured the variability in the observed GSL volume time series. Figure 7 shows results obtained using different embedding dimensions and time lags. On the basis of predictive mean square error the best model was the one that used a dimension and lag of 8. Again, in all the cases, only a few (7%–12%) of the training examples were used to learn the underlying dynamic relationship. These data points are support vectors and are results of the optimization algorithm. This is unlike previous methodologies that use the entire training data to make predictions.

5. Comparison With Artificial Neural Networks

[18] Artificial neural networks (ANN) are the most widely used machine learning tools in hydrology. *ASCE Task Committee on Application of Artificial Neural Networks in Hydrology* [2000] and *Govindaraju and Rao* [2000] present a wide array of ANN applications in hydrology. In order to

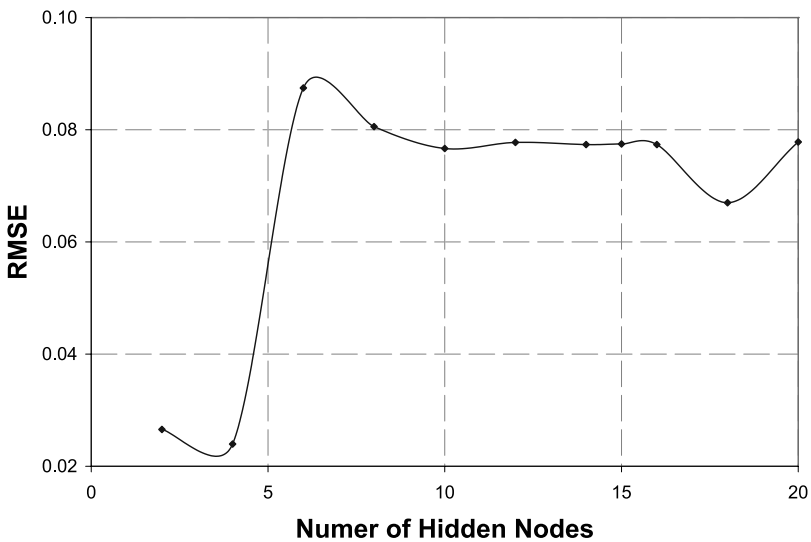

Figure 8. A three-layer artificial neural networks (ANN) performance as a function of the different number of hidden nodes.

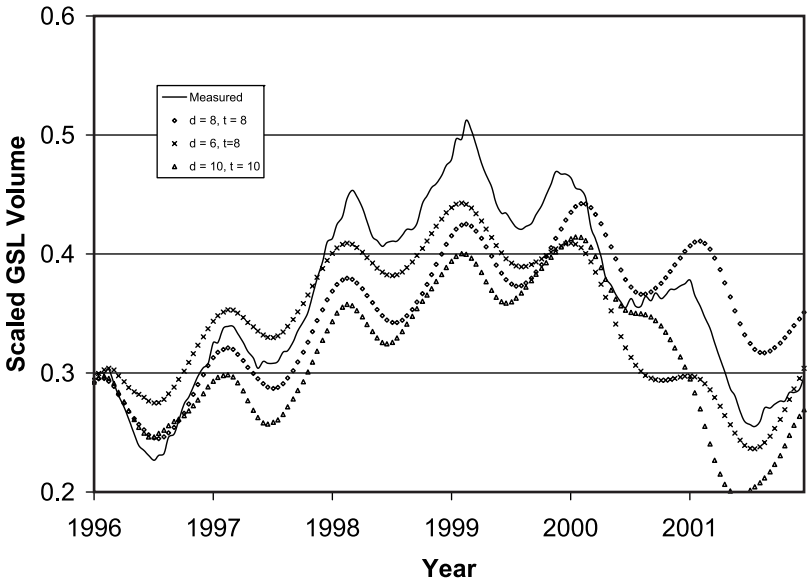

Figure 9. ANN continuous 6 year iterated GSL prediction using different embedding dimension and time lag. All plots did not use new data once prediction started.

compare ANN with SVM we used the same dynamical representation of the time series, training, and testing samples for the cases reported in section 4. In all cases, one output node, one hidden layer, and one input layer were used. Hidden layer neurons used a tan-sigmoid function, while linear transfer function was employed at the output layer. In all the cases the optimum number of hidden nodes was found to be about half the size of the input layer nodes (see, e.g., Figure 8 for GSL fall case).

[19] We used a Bayesian regularization back propagation algorithm to train the networks. This algorithm minimizes an objective function somewhat similar in principle to equation (5a) of the SVM. The difference is in the case of the ANN a least squares loss function is used as a measure of goodness of fit, and network complexity is controlled by minimizing the sum of network weights. The trade-off parameter between goodness of fit and network complexity was automatically estimated within the Bayesian framework of *MacKay* [1992]. The ANN were constructed using Matlab Neural Network Toolbox [*Demuth and Beale*, 2004]. In order to insure sufficient convergence of the training algorithm the training epoch was set to 500. We searched for the best ANN model by using different initializations (100 used here) and selected the best performing ANN based on minimum prediction errors over testing set. These results were compared with that of SVMs. Note that such an approach is different from ANN ensemble prediction where one actually averages (in some sense) different network predictions resulting from different network initialization rather than selecting the single best performing ANN model. Our method makes the ANN comparison fair with that of SVM, which unlike ANN converge to the same result irrespective of initialization points. Iterated ANN predictions were made in a similar manner to that of SVM by feeding back predictions into the model inputs.

[20] Figure 9 depicts results of iterated ANN prediction for the recent event of GSL time series. We used two types of measurement criteria: root-mean-square error (RMSE) and mean absolute error (MAE). Table 1 presents statistical performance measures of the two models. The ANN forecast captured the GSL fall event as well as that of SVM as measured by these statistics. As shown in Table 1, slightly better results are provided by SVM if one looks at the RMSE, whereas ANN gave smaller values for MAE. This might be an indication of the SVM models being robust with regard to "outliers" or extreme values. Even though both models have regularized objective functions, we attribute this difference to the form of the loss function (goodness of fit) optimized by the respective algorithms. While predicting the recent events, the best SVM model uses $\tau = 8$ and $d = 8$, whereas the best ANN model identifies $\tau = 6$ and $d = 8$. Both models gave good predictions for $\tau = 10$ and $d = 10$. Again, smaller error statistics as measured by RMSE came from SVM prediction, while MAE favored ANN.

6. Conclusions

[21] We have presented a dynamic reconstruction approach that used state-of-the-art learning methodology called support vector machines (SVM) based on statistical learning theory. The approach consisted of two parts: one related to the regularization of the solution and the other related to the goodness of fit resulting in remarkable generalization capabilities. Interested readers are encouraged to see *Vapnik* [1998].

[22] Extreme wet and dry periods of GSL have required more than traditional linear time series analysis tools. In fact, these linear models have failed to explain the dramatic rise and fall of GSL time series.

[23] Chaotic theory coupled with state-of-the-art learning tools has demonstrated the ability to reconstruct the dynamics of a low-dimensional system. Dimension and average mutual information analysis prescribed $d = 4$ and $\tau = 11$–17. In all the applications the best results of SVM have selected values of $d = 8$–10 and $\tau = 8$–11. Even though previously used methodologies (e.g., artificial neural networks) were also successful in predicting the wet and dry

Table 1. Error Statistics for SVM and ANN Models[a]

Event	SVM		ANN	
	RMSE	MAE	RMSE	MAE
GSL fall				
$\tau = 8$, $d = 11$	0.0221	0.0172	0.0240	0.0149
Recent events				
$\tau = 8$, $d = 8$	0.0265	0.0198	0.0457	0.0371
$\tau = 6$, $d = 8$	0.0483	0.0402	0.0387	0.0286
$\tau = 10$, $d = 10$	0.0445	0.0301	0.0592	0.0280

[a]SVM, support vector machines; ANN, artificial neural networks; RMSE, root-mean-square error; MAE, mean absolute error; GSL, Great Salt Lake; τ, time delay; and d, dimension.

periods of GSL volume time series, SVM provide several attractive features, including sparse approximation due to the fact that they only use a subset of the training data. While predicting GSL fall, for example, only 5% of the training data (support vectors) were able to explain the underlying relationship and iteratively predict the next 5 years. This ε-insensitive sparse solution saves CPU time during prediction because now one need not store all the training set but only the support vectors. This unique ε-insensitive feature is tolerant to outliers. Another important attribute of SVM compared with ANN is due to the nature of the optimization algorithm; the solution has global optima, avoiding the use of gradient-based search techniques that may converge to a local optima. We used 100 initialization runs to overcome this problem in ANN. One could also use ensemble "averaging" to stabilize ANN predictions. We believe the smaller RMSE error statistics given by SVM (hence robustness with regard to outliers and extreme values) compared with an ANN model applied to the same data set is the result of the nature and form of the loss function employed by the two algorithms. The SVM methodology also provides insight as to the actual data points and their locations that were used to construct the GSL predictor. However, they are not without shortcomings. Kernel as well as SVM parameter selections are still heuristic. Overall, the present application shows a promising performance of SVMs.

[24] **Acknowledgment.** We are grateful for the thoughtful review and suggested improvements by the Associate Editor and two anonymous reviewers that helped to improve this manuscript.

References

Abarbanel, H. D. I. (1996), *Analysis of Observed Chaotic Data*, Springer, New York.

Abarbanel, H. D. I., and U. Lall (1996), Nonlinear dynamics and the Great Salt Lake: System identification and prediction, *Clim. Dyn.*, *12*, 287–297.

Abarbanel, H. D. I., U. Lall, Y. Moon, M. E. Mann, and T. Sangoyomi (1996), Nonlinear dynamics and the Great Salt Lake: A predictable indicator of regional climate, *Energy*, *21*(7/8), 655–665.

Arnow, T. (1984), Water level and water quality changes in Great Salt Lake, Utah, 1847–1983, *U.S. Geol. Surv. Circ.*, *913*.

ASCE Task Committee on Application of Artificial Neural Networks in Hydrology (2000), Artificial neural networks in hydrology. II: Hydrologic applications, *J. Hydrol. Eng.*, *5*, 124–137.

Asefa, T., M. W. Kemblowski, G. Urroz, M. McKee, and A. Khalil (2004), Support vectors–based groundwater head observation networks design, *Water Resour. Res.*, *40*, W11509, doi:10.1029/2004WR003304.

Boser, B. E., I. Guyon, and V. Vapnik (1992), A training algorithm for optimal margin classifiers, paper presented at Fifth ACM Workshop on Computational Learning Theory, Assoc. for Comput. Mach., Pittsburgh, Pa., July.

Coulibaly, P., and C. K. Baldwin (2005), Nonstationary hydrological time series forecasting using nonlinear dynamic methods, *J. Hydrol.*, *307*, 164–174.

Demuth, H., and M. Beale (2004), Neural Network Toolbox user's guide, MathWorks, Inc., Natick, Mass.

Dibike, B. Y., S. Velickov, D. Solomatine, and B. M. Abbot (2001), Model induction with support vector machines: Introduction and applications, *J. Comput. Civ. Eng.*, *15*(3), 208–216.

Elshorbagy, A., S. P. Simonovic, and U. S. Panu (2002), Estimation of missing streamflow data using principles of chaos theory, *J. Hydrol.*, *255*, 123–133.

Friedman, J. (1991), Multivariate adaptive regression splines, *Ann. Stat.*, *19*(1), 1–141.

Girosi, F. (1998), An equivalence between sparse approximation and support vector machines, *Neural Comput.*, *10*(6), 1455–1480.

Govindaraju, R. S., and A. R. Rao (Eds.) (2000), *Artificial Neural Networks in Hydrology*, 348 pp., Springer, New York.

Hense, A. (1987), On the possible existence of a strange attractor for the southern oscillation, *Contrib. Atmos. Phys.*, *60*(1), 34–47.

Jayawardena, A. W., and F. Lai (1994), Analysis and prediction of chaos in rainfall and stream flow time series, *J. Hydrol.*, *153*, 23–52.

Jayawardena, A. W., W. K. Li, and P. Xu (2002), Neighborhood selection for local modeling and prediction of hydrological time series, *J. Hydrol.*, *258*, 40–57.

Kanevski, M., A. Pozdnukhov, S. Canu, and M. Maignan (2002), Advanced spatial data analysis and modeling with support vector machines, *Int. J. Fuzzy Syst.*, *4*(1), 606–615.

Lall, U., and M. E. Mann (1995), The Great Salt Lake: A barometer of low-frequency climatic variability, *Water Resour. Res.*, *31*(10), 2503–2515.

Lall, U., T. Sangoyomi, and H. D. I. Abarbanel (1996), Nonlinear dynamics of the Great Salt Lake: Nonparametric short-term forecasting, *Water Resour. Res.*, *32*(4), 975–985.

Liong, S. Y., and C. Sivapragasam (2002), Flood stage forecasting with SVM, *J. Am. Water Resour. Assoc.*, *38*(1), 173–186.

MacKay, D. J. C. (1992), Bayesian methods for adaptive models, Ph.D. dissertation, Calif. Inst. of Technol., Pasadena.

Mann, M. E., U. Lall, and B. Saltzman (1995), Decadal-to-centennial-scale climate variability: Insights into the rise and fall of the Great Salt Lake, *Geophys. Res. Lett.*, *22*(8), 937–940.

Moon, Y.-I. (1995), Climate variability and dynamics of Great Salt Lake hydrology, Ph.D. dissertation, 247 pp., Utah State Univ., Logan.

Moon, Y.-I., and U. Lall (1996), Atmospheric flow indices and interannual Great Salt Lake variability, *J. Hydrol. Eng.*, *1*, 55–62.

Müller, K. R., A. Smola, G. Rätsch, B. Schölkopf, J. Kohlmorgen, and V. Vapnik (1999), Predicting time series with support vector machines, in *Advances in Kernel Methods: Support Vector Learning*, edited by B. Schölkopf, C. J. C. Burges, and A. J. Smola, pp. 243–254, MIT Press, Cambridge, Mass.

Rodriguez-Iturbe, I., F. B. De Power, M. B. Sharifi, and K. P. Georgakakos (1989), Chaos in rainfall, *Water Resour. Res.*, *25*(7), 1667–1675.

Sangoyomi, T., U. Lall, and H. D. I. Abarbanel (1996), Nonlinear dynamics of the Great Salt Lake: Dimension estimation, *Water Resour. Res.*, *32*(1), 149–1599.

Sivakumar, B. (2000), Chaos theory in hydrology: Important issues and interpretations, *J. Hydrol.*, *227*, 1–29.

Takens, F. (1981), Detecting strange attractors, in *Dynamical Systems and Turbulence, Warwick 1980*, edited by D. A. Rand and L. S. Young, *Lect. Notes Math.*, *898*, 366–381.

Tikhonov, A., and V. Arsenin (1977), *Solution of Ill-Posed Problems*, John Wiley, Hoboken, N. J.

Vapnik, V. (1998), *Statistical Learning Theory*, John Wiley, Hoboken, N. J.

Weigend, A. S., and N. A. Gershenfeld (Eds.) (1994), *Time Series Prediction: Forecasting the Future and Understanding the Past*, Addison-Wesley, Boston, Mass.

T. Asefa, Department of Source Rotation and Environmental Protection, Tampa Bay Water, 2535 Landmark Drive, Suite 211, Clearwater, FL 33761-3930, USA. (tasefa@tampabaywater.org)

M. Kemblowski and G. Urroz, Department of Civil and Environmental Engineering and Utah Water Research Laboratory, Utah State University, 8200 Old Main Hill, Logan, UT 84321, USA. (mkem@cc.usu.edu; gurro@cc.usu.edu)

U. Lall, Department of Earth and Environmental Engineering and International Research Institute for Climate Prediction, Columbia University, New York, NY 10027, USA. (ula2@columbia.edu)