Report for 2004UT46B: Data Fusion for Improved Management of Large Western Water Systems

- Articles in Refereed Scientific Journals:
  - Khalil, A., M. McKee, M. Kemblowski, and T. Asefa. 2005. Basin-scale water management and forecasting using multi-sensor data and artificial neural networks. J. American Water Resources Association.
  - Khalil, A., M. McKee, M. Kemblowski, and T. Asefa. 2005. Sparse Bayesian learning machine for real-time management of reservoir release. Water Resources Research.
  - Asefa, T., M. Kemblowski, M. McKee, and A. Khalil. 2005. Multi-time scale stream flow prediction: The support vector machines approach. J. Hydrology.
- Conference Proceedings:
  - Khalil, A., and M. McKee. 2004. Hierarchical Bayesian analysis and statistical learning theory, I: Theoretical concepts. In: Proceedings of the 2004 Water Management Conference, Water Rights and Related Water Supply Issues, U.S. Committee on Irrigation and Drainage. Salt Lake City, October, 2004. pp. 433-444.
  - Khalil, A., and M. McKee. 2004. Hierarchical Bayesian analysis and statistical learning theory, II: Water management application. In: Proceedings of the 2004 Water Management Conference, Water Rights and Related Water Supply Issues, U.S. Committee on Irrigation and Drainage. Salt Lake City, October, 2004. pp. 445-455.
  - Khalil A. and M. McKee, 2004. An adaptive model paradigm for water resources management. In: AWRA Annual Conference, Orlando, Florida. November, 2004.

Report Follows

Data Fusion for Improved Management of Large Western Water Systems

Problem

The relative scarcity of water in the western US is growing due to population and economic growth, pollution, and diversification of the types of demands that are being placed on water use (e.g., traditional consumptive uses such as irrigation and municipal supply, as well as emerging uses for such concerns as water quality maintenance and endangered species protection). This increasing relative scarcity brings: (1) a greater need to more intensively manage the resource, and, as a result, (2) a requirement for better information about the current and potential future states of our water resources systems so that management decisions can be better informed.

In spite of these increasing needs for better water resources management information, investments in traditional water resources data collection programs (e.g., point stream flows, snow pack, soil moisture, etc.) are declining at the federal and state levels. For example, USGS support for maintenance of several stream gages in Utah has been withdrawn in recent years due to a lack of state cost-sharing commitments. In contrast, investments on the part of other Federal agencies (that have not traditionally played a significant role in support of water resources management) in new data collection methods are increasing (e.g., satellite imagery of land cover, snow cover, ocean surface temperatures, etc.; radar estimation of precipitation; aircraft and satellite imagery for estimation of evapotranspiration). These new data streams will have to be used to back-fill the decline in availability of traditional data. Moreover, analytic methods will need to be developed to apply to these data in order to improve the quality of the information base available to managers of large water systems.

Today's managers have not been schooled in new ways of collecting data or in the analytic approaches required to understand the data. Before new methods of gaining information and making decisions can be practical, investments must be made to place the resulting capabilities into the hands of the water managers who need them. These must be practical and effective, and the water managers must themselves see the value of the information that results.

These are information problems facing managers of large water systems toady, especially large irrigation systems, in several places in Utah as well as in many other arid river basins in the western US.

Research Objectives

Research is needed to develop the data now becoming available from emerging telemetric and remote sensing sources into useful information for all temporal and geographical scales of water resources management. This must be done in such a way as to maximize the total value of the information coming from both these new, emerging data sources and from the traditional water resources monitoring approaches. Further, the products of such research must be of practical use to the water resources managers who (1) are now losing access to traditional data sources and (2) have not been trained in how to access and use the information flowing from new remote sensing capabilities. In addition, the research products must also be of use to a growing range of stakeholders who have heterogeneous technical backgrounds and skill levels.

The purpose of this project is to develop a significantly enhanced capability within the state of Utah--that will also be appropriate for application in the arid West--to more efficiently manage the state's scarce water resources by exploiting emerging technologies in data collection and analysis.

The objectives of the research are to:

1. Develop and test methodologies from statistical learning theory for combining meteorological and hydrological data from traditional and new remote sensing sources to produce information valuable to managers of large water resources systems. These methodologies will be directed at supplying information for improving water management decisions on problems having a wide range of time and spatial scales (e.g., from daily reservoir releases and canal diversions to long-term commitments for reservoir operations and water exchanges).

2. Develop and implement inexpensive and effective web-based methodologies to disseminate the resulting decision-relevant information to all potential stakeholders.

3. Evaluate and report on the results of the application of the methodologies.

Methodology

The Sevier River Basin, managed by the Sevier River Water Users Association (SRWUA) and the area served by the Emery Water Conservancy District (EWCD) were used as case study areas for the project. Project research personnel have excellent working relationships with the SRWUA and EWCD, as well as with the Provo, Utah, office of the US Bureau of Reclamation, who assists both the SRWUA and EWCD in the maintenance, operation, and extension of the monitoring programs that collect, archive, and display real-time data on the state of both of these large water systems.

Research Tasks

Development of Models from Statistical Learning Theory

The major objective of this research effort is to develop and implement a set of tools that can be used to reconcile data and measurements that come from various sources and are characterized by different temporal and spatial scales of resolution, such as NASA remote sensing information and local, on-ground measurements such as stream flow data collected by the USGS. The tools will be used to fuse various pieces of information into one coherent and seamless representation of river basin states, both current and predicted. The resulting information will be provided to stakeholders, decision-makers, and water managers using modern information technologies.

This effort takes advantage of recent research experience to formulate a general framework for hydrologic data assimilation and forecasting, utilizing the following components:

- Bayesian Belief Networks (BBNs). This is a graph-based methodology that was developed by the AI community over the last decade. It is recognized as the most consistent and powerful tool for building statistical expert systems, and is used, for example, by the army to fuse intelligence information coming from multiple sources. Project researchers have published the results of research using this tool for various water and environmental problems (for example, see Ghabayen and McKee, 2003; and Ghabayen et al., 2003, 2004), and a graduate student on the project will examine the use of BBNs in assessing uncertainties in salt loading estimates in the EWCD.

- Statistical Learning Theory (SLT). This approach constitutes a general framework for machine learning, and is currently a subject of intensive research in academia and industrial research laboratories. It provides a sound and very efficient environment for building statistical models that include physical insights into the modeled phenomena. Researchers on this project have used this approach to build models for subsurface quality monitoring network design, evaluating information value for groundwater measurements, and spring runoff predictions. Examples of development of models from STL applications that show excellent promise for application to the problems posed for this research include Hassan and McKee (2003), Pande and McKee (2003), and Khalil and McKee (2003).

Specific models to be developed by the project consist of:

- a model for providing long-term forecasts of spring runoff in the Upper Sevier River Basin
- a models for providing short-term operations guidance for reservoir operations for Piute Reservoir in the Upper Sevier Basin
- a model for providing short-term operations guidance for canal diversions on the Piute-Sevier Valley Canal in the Upper Sevier Basin
- a model for estimating monthly and annual salt loading from the San Rafael River into the Colorado River, and quantification of uncertainty in those estimates

Implementation of Models

The models are to be integrated with the SRWUA or EWCD databases and implemented on the websites of those organizations. Implementation will be accomplished through collaborative work among USU project staff, representatives of the managing water districts, and contractors who work for the districts to maintain the water resources monitoring systems, databases, and websites.

Evaluation and Dissemination of Results

In addition to the use of standard statistical methods for assessing the quality of model predictions, the project will conduct an analysis of the real-time performance of the operation of the short-term forecasting models as they are being applied during the 2004 irrigation season. As necessary, project personnel will work directly with end-users in the SRWUA and EWCD to identify and correct problems with the models, and to educate the users in proper model operation and interpretation. The lessons learned from these interactions will be documented and recommendations for implementation and management of such models within the institutional framework of these types of water resources organizations will be documented. In addition to the use of web-based dissemination of research products, the results of the project will be published in the normal academic venues.

Principal Findings

Water Management in the Sevier River Basin

Real-time operations models were developed using methods from statistical learning theory. These include artificial neural network (ANN) models, support vector machine (SVM) models, and "lazy learner" (LL) models. These models have been constructed to help provide real-time management information for determining releases from Piute Reservoir and diversions into the Sevier Valley/Piute Canal. The models have been made operational on the SRWU web site for use by reservoir and canal operators. Comparisons of model predictions versus actual canal operations are given in Figure 1.

Short-term predictive models were also built using artificial neural network approaches to forecast diurnal flows from Clear Creek into the Sevier River. An example of these forecasts, made hourly for a period 24 hours in advance, is given in Figure 2.

Long-term predictive models were constructed to forecast stream flows at the Hatch gage in the Upper Sevier River Basin. These predictions come from an artificial neural network model that uses historical stream flow data, Snotel data, and sea surface temperature anomaly data from the Pacific and Atlantic Oceans. A comparison of the forecasts obtained from the ANN model versus historically measured flows is shown in Figure 3. The work shown in Figures 1-3 has been recently published (Khalil et al., 2005).

In FY 2004, work focused on development and statistical verification of an hourly operational model for predicting required releases from Piute Reservoir. The modeling process utilizes a combination of support vector machines and relevance vector machines (RVMs) to screen incoming data to recognize outliers and/or "drift" in the underlying probability distribution of the input data, develop a revised predictor model if drift in the underlying distribution is detected, and then make a prediction for required reservoir releases for the next hour. Adoption of the RVM approach for developing the predictor model has provided the capability of estimating confidence intervals on the prediction made by the model. This capability, which has not been previously possible, gives the reservoir operator valuable information about the uncertainty in the prediction made by the model. The suite of models is designed to run in real time and to provide the reservoir operator with an hour-by-hour recommendation for releases needed from the reservoir in order to meet downstream demands for nine irrigation canals. It does this in order to meet water orders that arrive 24 to 48 hours in advance of deliveries, even though travel times from the reservoir to the end of the furthest canal is on the order of five or six days. The suite of models was developed using data from the 2001 and 2002 irrigation seasons, and then tested against the 2003 and 2004 irrigation seasons. Figure 4 provides a comparison of actual reservoir releases and model-generated recommendations for release quantities, as well as confidence intervals, for the 2003 season. At the time of preparation of this report, the SRWUA was in the process of operationalizing the software that will connect the output of this model to the gate controllers that manage releases from Piute Reservoir. In the coming irrigation season, the model will be running the reservoir and its performance will be monitored by the reservoir operators and project researchers.

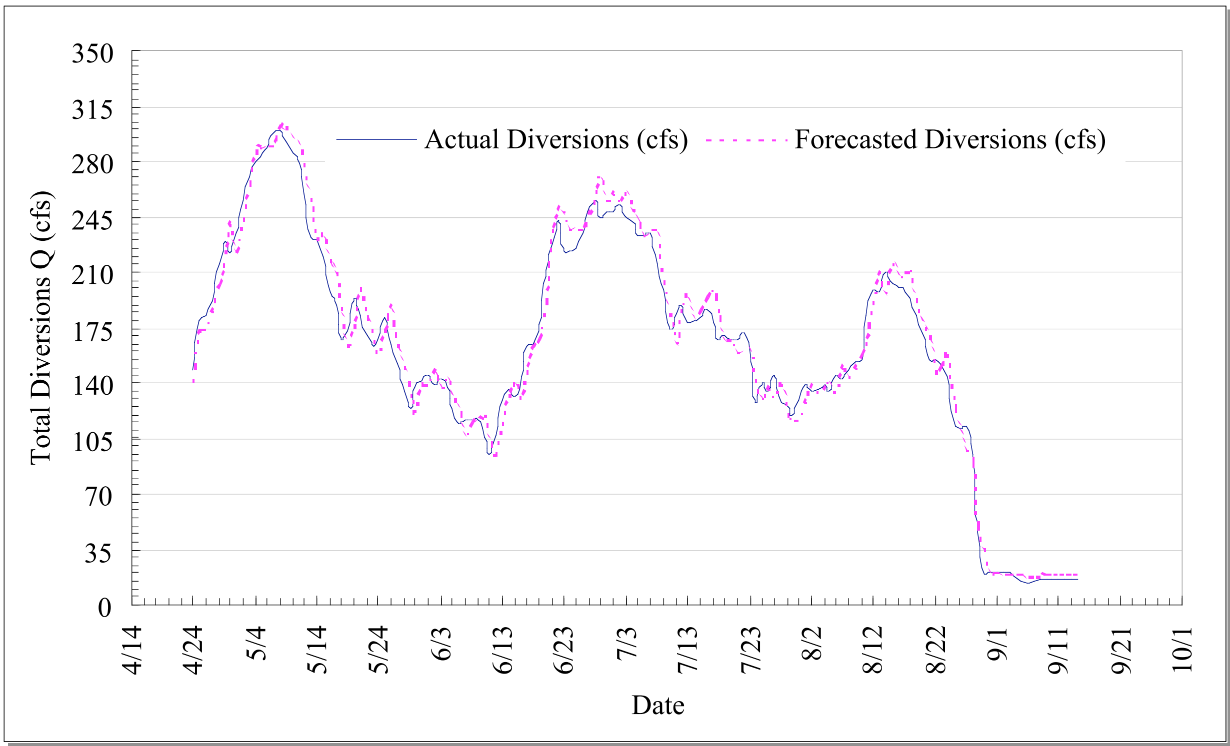

Figure 1. Comparison of Actual Sevier Valley/Piute Canal Diversions in 2002 with ANN Model Forecasted Diversions

Salt Loading in the San Rafael River Basin

Work in FY 2004 focused on statistical analysis of data available from historic stream flow and salt concentration measurements, including the real-time data provided in the on-line database operated by the Emery Water Conservancy District (EWCD) (see http://www.ewcd.org/). Work has focused thus far on analysis of salt loading from the San Rafael into the Green River. In particular, statistical relationships between stage and discharge, and between conductivity and salt concentration have been developed for the major tributaries of the San Rafael. This has been done so as to provide the basis for a Bayesian belief network (BBN) model that will be used to quantify the uncertainty in the estimate of salt loading from the basin. Additional probabilistic relationships between conductivity and salt concentration have been developed from the fundamental principals of salt chemistry. A preliminary BBN model has been constructed that estimates total daily, weekly, monthly, and annual salt loading from the San Rafael drainage into the Green River. The model is being verified and prepared for implementation on the EWCD web site.

Significance

The Sevier River Basin, managed by the Sevier River Water Users Association (SRWUA), and the Emery Water Conservancy District (EWCD) in Utah have been used as case studies and experimental sites. They were chosen because of their significant size, their importance in the agricultural sector of the state, their highly developed on-line databases, and the willingness of local water resources managers to cooperate with the research and make use of the outputs of the project. The project has focused on development of data sensor fusion approaches to reduce the uncertainty that accompanies significant water management decisions through the implementation of real-time management and long-term forecasting models. In the case of the Sevier River Basin, these models are being useful for real-time reservoir release and canal diversion decisions and for long-term forecasting of water availability. For the EWCD, the modeling is resulting in improved estimates of salt loading into the Colorado River from the region, as well as improved quantification of the uncertainty in these salt loading estimates. The output of these models is being utilized for development and deployment of decision-support systems that are being made available to water managers from these organizations. Project staff are continuing to work with these water districts to help integrate these models with their local database systems so that the resulting information will be available to water managers and stakeholders via the internet.

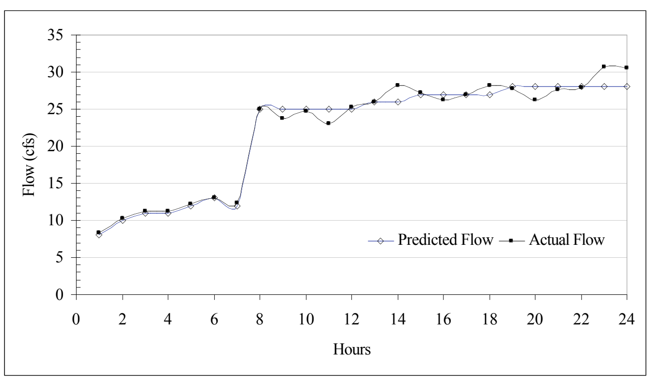

Figure 2. Predicted Versus Actual Diurnal Fluctuation of Flows in Clear Creek on 4/4/2001

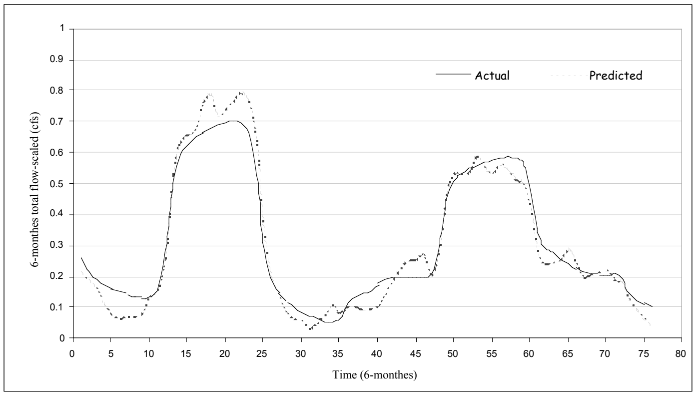

Figure 3. Time-Series Performance of the ANN Model in Predicting Seasonal Flows at Hatch

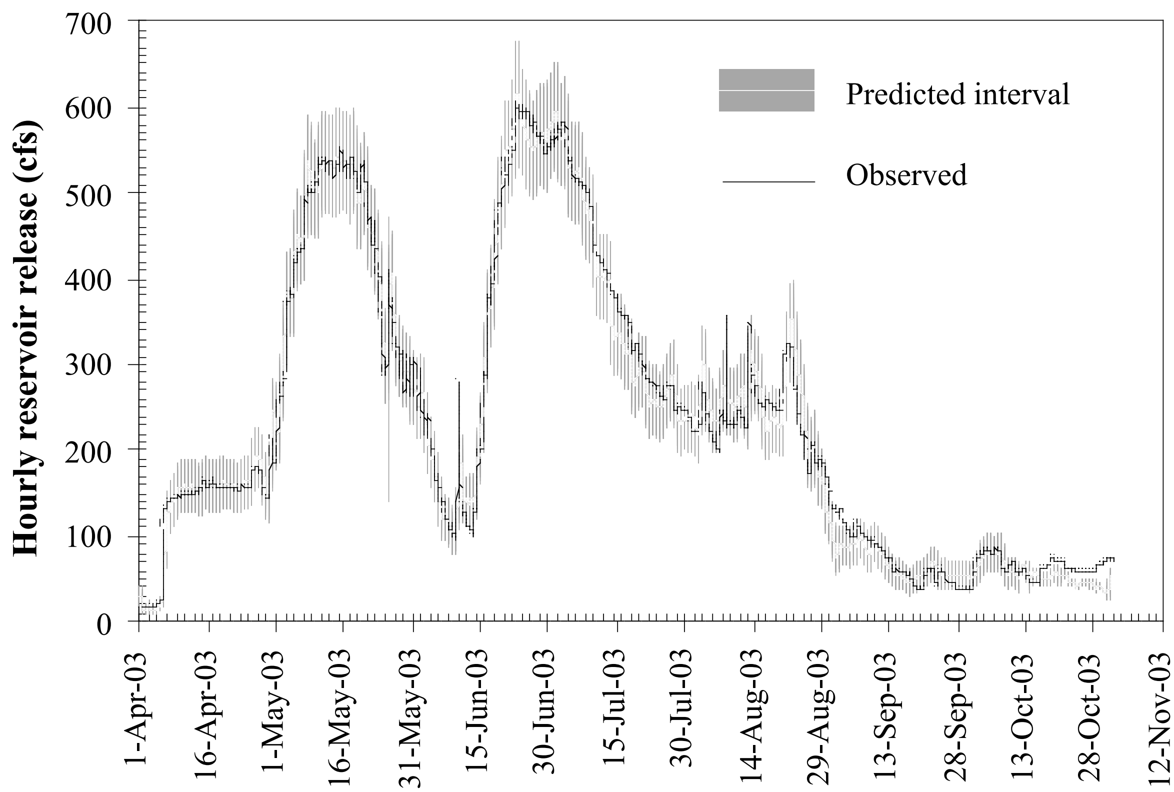

Figure 4. Comparison of Actual Piute Reservoir Releases in 2004 with the SVM/RVM Model Forecast Releases

References

Almasri, M.N., and J.J. Kaluarachchi. 2002. Application of artificial neural networks in predicting stream-aquifer interaction. *ACTA Universitatis Carolina-Geologica*, 46(2/3):54-57.

Bowman, C.L., and A.N. Steinberg. 2001. A systems engineering approach for implementing data fusion systems. In: D. L. Hall and J. Llinas (eds.). *Handbood of Multisensor Data Fusion*, Chapter 16. CRC Press, London.

Carroll, T.R. 1990. Operational airborne and satellite snow cover products of the National Operational Hydrologic Remote Sensing Center. *Proceedings of the 47th Eastern Snow Conference*, Bangor, Maine, CRREL Special Report 90-44, June 7-8.

Grody, N.C., and A.N. Basist. 1996. Global identification of snow cover using SSM/I measurements. *IEEE Transactions on Geoscience and Remote Sensing*, 34(1):237-249.

Gautam, M.R., K. Watanabe, and H. Saegusa. 2000. Runoff analysis in humid Forest catchment with artificial neural network. *Journal of Hydrology*, 235:117-136.

Ghabayen, S., and M. McKee. 2003. A Bayesian belief network model for multi-objective optimization of the Gaza Water Resources System. Presented at the *Annual Conference of the American Water Resources Association*, San Diego, CA, November 3-6.

Ghabayen, S., M. McKee, and M. Kemblowski. 2003. Using Bayesian belief networks, ionic molar ratios, and isotopes for identification of salinity origin and monitoring requirements in the Gaza Coastal Aquifer. Presented at the *Annual Conference of the American Water Resources Association*, San Diego, CA, November 3-6.

Ghabayen, S., M. McKee, and M. Kemblowski. 2004. Characterization of uncertainties in the operation and economics of the proposed seawater desalination plant in the Gaza Strip. Accepted for publication in *Desalination*, January.

Govindaraju, R.S. (2000). Artificial neural networks in hydrology. II: Hydrologic Applications. *Journal of Hydrologic Engineering*, 5(2):124-137.

Hall, D.K., A.B. Tait, J.L. Foster, A.T.C. Chang, and M. Allen. 2000. Intercomparison of satellite-derived snow-cover maps. *Annals of Glaciology*, 31:369-376.

Hassan, R., and M. McKee. 2003. Canal flow regulation using support vector machines. Presented at the *Annual Conference of the American Water Resources Association*, San Diego, CA, November 3-6.

Hiramatsu, K., S. Shikasho, and K. Mori. 1999. Nonlinear prediction of river water-stages by Feedback Artificial Neural Network. *J. Fac. Agr., Kyushu Univ.*, 44(1.2):137-147.

Kaplan, A., Y. Kushnir, M. Cane, and M. Blumenthal. 1997. Reduced space optimal analysis for historical datasets: 136 years of Atlantic sea surface temperatures. *Journal of Geophysical Research*, 102:27,835-27,860.

Kaplan, A., M. Cane, Y. Kushnir, A. Clement, M. Blumenthal, and B. Rajagopalan. 1998. Analyses of global sea surface temperature 1856-1991. *Journal of Geophysical Research*, 103☹18)567-18,589.

Khalil, A., and M. McKee. 2003. Basin-scale water management and forecasting using multisensor data and neural networks. Presented at the *Annual Conference of the American Water Resources Association*, San Diego, CA, November.

Khalil, A., M. McKee, M. Kemblowski, and T. Asefa. 2005. Basin scale water management and forecasting using artificial neural networks. *J. American Water Resources Association*, 41(1):195-208.

Kim, G., and A. P. Barros. 2001. Quantitative flood forecasting using multisensor data and neural networks. *Journal of Hydrology*, 246:45-62.

Liong, S. Y., and C. Sivapragasam. 2002. Flood stage forecasting with support vector machines. *Journal of the American Water Resources Association*, 38(1):173-186.

Nghiem, S. V., and W. Tsai. 2001. Global Snow Cover Monitoring with Spaceborne Ku-band Scatterometer. *IEEE Transactions of Geoscience and Remote Sensing*, 39:2118-2134.

Pande, S., and M. McKee. 2003. Probably approximately optimal (PAO) feature subset selection for incremental local learning algorithms. Presented at the *Annual Conference of the American Water Resources Association*, San Diego, CA, November 3-6.

Pulliainen, J., and M. Hallikainen. 2001. Retrieval of regional snow water equivalent from space-borne passive microwave observations. *Remote Sensing of Environment*, 75:76-85.

Sivapragasam, C., S.Y. Liong, and M.F.K. Pasha. 2001. Rainfall and runoff forecasting with SSA-SVM approach. *Journal of Hydroinformatics*, 3:141-152.

Vapnik, V. N. 1995. The nature of statistical leering theory. Springer-Verlag.

Wilson, L. L., L. Tsang, J. N. Hwang, and C. Chen. 1999. Mapping snow water equivalent by combining a spatially distributed snow hydrology model with passive microwave remote-sensing data. *IEEE Transactions on Geoscience and Remote Sensing*, 37:690-704.

Yu, X., S. Liong, and V. Babovic. 2002. An approach combining chaos-theoretic approach and support vector machines: case study in hydrologic forecasting. Volume 2: 690-695. In Johan Junke guo (Ed.). *Advance in Hydraulics Water Engineering*, Proceedings of the 13th IAHR-APD Congress. Singapore, 6-8 August.