Proceedings of the 2006 American Control Conference
Minneapolis, Minnesota, USA, June 14-16, 2006

Modeling and Identification of the Sevier River System

M. Maxwell and S. Warnick

Information and Decision Algorithms Laboratories
Department of Computer Science, Brigham Young University, Provo, UT, 84602
http://idealabs.byu.edu/

***Abstract*— This paper deals with the control-oriented MISO system identification process for the Piute Dam and Sevier River. The Sevier River is an open-channel river with long delays and infrequent data measurements. Based upon adaptations of two models commonly used to estimate river flow, we developed a parameterized mass balance model and compared the validation results of the three models. The parameterized mass balance model performed the best during periods of high flow. Since high flow conditions are most important for water conservation efforts, we chose this model as the basis of a robust controller to be designed and implemented on the Sevier River Basin.**

I. INTRODUCTION

The conservation of water resources has become an increasingly important task throughout the world. This is especially true in arid climates where lack of water can cause economic ruin or struggle for survival. However, from a water management perspective there are many complexities that hinder water conservation efforts. Examples of these complexities include the size of river systems, the difficulty of accurately predicting downstream flow in the presence of disturbances, and the inability of water management professionals to constantly and consistently deliver water to specified locations on demand without incurring losses.

With the modernization of the water management industry and the rise of technological advancements enabling real-time water monitoring at remote locations, there is a large potential for decreasing water losses through modeling and automatic control of river systems. There has already been much research dealing both with the modeling and the control of irrigation canals. An overview of current methods, applications, and research of this area is covered in [1].

Commonly, the modeling of open-channel hydraulics such as irrigation channels has been based on the Saint-Venant equations. These non-linear hyperbolic partial differential equations can accurately approximate the flow in such systems. Litrico, et al. have done an extensive amount of work with these equations and have developed suitable models for control design using simplifying assumptions and linearizing around a reference flow. The resulting system was able to be approximated using a second-order transfer function with delay [2]. With this model, Litrico, et al. were able to develop a number of different controllers for the system including a PI controller [3], a robust IMC controller [4], and an H_∞ controller [5].

This research was supported by the U.S. Department of Interior, Bureau of Reclamation and the Brigham Young University Office of Research and Creative Activities.

Contact M. Maxwell (maxwell.matt@gmail.com) and S. Warnick (sean@cs.byu.edu) with comments.

Additionally, Weyer has used data-driven system identification techniques to develop suitable models for irrigations systems [6]. In [7], Ooi, et al. show that data-driven models perform comparably with models derived from the physics-based Saint-Venant equations. Consequently, Weyer has created a LQ controller [8] and Li, et al. have created an H_∞ controller [9] for the data-driven model.

Although Weyer found suitable data-driven models for his system, these results do not scale well to the Sevier River Basin. This is due to the fact that the system modeled by Weyer had relatively small delays (a few minutes) and frequent data sampling (every minute). Because of the remote locations and battery-operated equipment in many areas of the Sevier River Basin only hourly data transmission is feasible. Furthermore, the lags in the Sevier River are usually over a day long.

The purpose of this system identification process is to develop a data-driven model suitable for heavily lagged systems with infrequent data. This model will be a control-oriented model and will form the basis of a robust controller to be implemented on Piute Dam in the Spring of 2006.

First, we provide a description of the system to be modeled including the physical layout and the data collection and dissemination procedures. Next, we discuss our modeling procedures including system representation, lag determination, and model classes used. We next show the results of the respective models and provide some analysis on the results. Last, we present conclusions for the system identification process and outline our future work.

II. SYSTEM DESCRIPTION

A. Physical Layout

The Sevier River Basin is located in rural south-central Utah. It covers approximately 12.5 percent of the state of Utah and is managed by the Sevier River Water Users Association (SRWUA). The basin is divided into five regions as represented in Figure 1: Upper, Central, Gunnison, Lower, and San Pitch. The Piute Reservoir is located in the Upper region. Valuable run-off collects in the Piute Reservoir each spring. Releases from this reservoir flow north into diversions located along the river in the Central region.

Excess water runs over the Vermillion Dam (located on the Central/Gunnison border) and is lost to water users in the Central and Upper regions.

Fig. 1. Sevier River Basin

Because the Sevier River Basin is located in an arid climate and generally uses all possible water resources each year, effective management of the Piute Reservoir is a high priority. To facilitate a more efficient water management system, the Sevier River Basin has been heavily instrumented and has many diversions and release structures that can be controlled remotely via radio. The Piute Dam is one of these remotely automated structures. This automated gate forms the framework from which the Piute Dam model-driven automation project is based.

Average delays from the reservoir release at Piute Dam to the lowest downstream point in the model, Sevier River at Vermillion, are between 24 and 36 hours. The reservoir release ranges from 0 m^3/s in the winter months to nearly 20 m^3/s at the height of the irrigation season. The average release during the irrigation season is around 8.5 m^3/s.

B. Data Collection

Data are collected by remote data logging stations every hour. Measurements including water height, flow, etc. are transmitted via radio, phone, and/or internet to a central datahut. Each of these measurements represent averages over the previous hour. Data collection software collects the data from each of the remote stations and stores them into a database. The locations of data collection sites along the Piute Reservoir stretch of the river are indicated by black dots in Figure 2.

As can be seen in Figure 2, there are two intermediate measuring stations located on the Sevier River: one above Clear Creek and the other near Elsinore. These intermediate measurements split the river system into three stretches as denoted by dashed lines on the graphic. Each of these stretches can be modeled independently and then combined serially to obtain a model for the entire system from Piute Reservoir to Sevier River at Vermillion.

Hourly historical data from each of the data collection stations on the river begin in January 2000 and continue to the present day. Equipment malfunctions and communication errors have caused missing data points to occur; however, over 98% of the required data are present. Fortunately, most

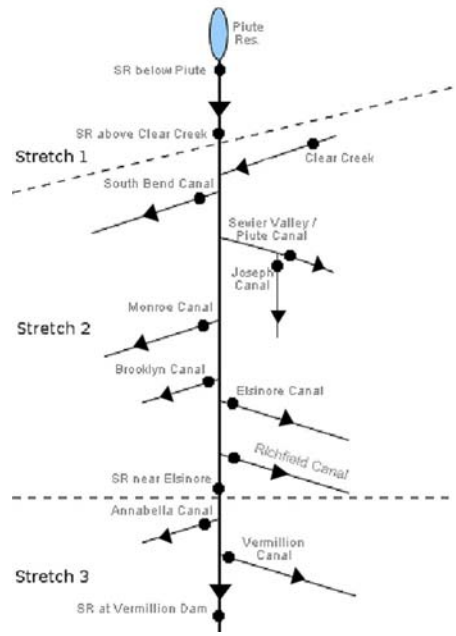

Fig. 2. Piute Reservoir to Vermillion Dam Diagram

of the missing data was infrequent and of short duration, so linear interpolation was sufficient to fill the missing data. One extended stretch of missing data occured from September 2002 to April 2003. Since this large break split the data into two nearly equal groups we used the first group of three years to fit the models and the second group of two years to validate the model.

C. Data Dissemination

To provide water users and water management officials with accurate, timely hydrologic information the SRWUA maintains the webpage http://www.sevierriver.org. This website contains detailed information on flows, reservoir heights, weather conditions, and more for the entire Sevier River Basin. As part of his previous work with the U.S. Bureau of Reclamation (USBR), the author helped design a web-based data display system entitled OpenBasin [10]. This system is the basis behind SRWUA's data dissemination process.

Website pages of particular interest for this paper include the Upper region page (http://www.sevierriver.org/upper), and the Central region page (http://www.sevierriver.org/central). The Upper region page displays data for the Piute Reservoir and the Central region page displays data for all of the downstream area that will be used in this system identification procedure. Additionally, real-time model estimates and comparisons for each of the models introduced in the next section can be found at http://www.sevierriver.org/flow/models.php.

III. MODELING PROCEDURES

This and the following sections will focus on the second stretch of the river from Sevier River above Clear Creek to

Sevier River at Elsinore (see Figure 2). This stretch of the river poses the most difficulties because it has a regulated inflow, an unregulated inflow, seven regulated outflows, and significant lag.

A. System Representation

The inflows and outflows of the river stretch are labeled u_1 through u_9 where u_1 is the flow at Sevier River above Clear Creek, u_2 is the flow at Clear Creek, and so forth downstream. There will be no forced sign changes for the different flows. Negative numbers will be treated as outflows while positive numbers will be treated as inflows. The arrows on Figure 2 indicate the direction of flow for each section of the river. An additional step in our validation process is ensuring that the correct signs correlate with correct inflows and outflows.

It should also be noted that Sevier Valley/Piute Canal is u_3 while Joseph Canal is u_4 and the final diversion, Richfield Canal, is denoted u_9. We let y denote the flow downstream at the point to be modeled. For the second stretch of the river, this location is Sevier River near Elsinore.

Both of the inflows into this stretch, $\{u_1, u_2\}$, are represented as inputs to the model. The outflows of the river, $\{u_3, \ldots, u_9\}$, are also modeled as inputs because each diversion along the river from Piute Reservoir to Sevier River at Vermillion Dam has an automated gate. These gates are adjusted to maintain a specified reference flow by independent PID controllers. Since we can specify the flow at each one of these locations and ensure that the resulting flow will be within a nominal amount of the desired flow there is no need to predict the amount of flow at each of these structures. Consequently, the resulting model type for Stretch 2 is a MISO model.

B. Lag Determination

We determined approximate time lags for each of the inflows/outflows obtained by shifting each input data set u_i by n hours where $n = 0, 1, \ldots, 9, 10$ and performing a statistical correlation test on the shifted data versus the downstream flow at Sevier River at Elsinore. The amount n of the shifted data which had the highest correlation was then taken to be the appropriate time lag of that input. The maximum of n was set at 10 for this stretch of the river because it is known that the lag from any u to the downstream y is less than 10 hours. The resulting lags are $L = \left[\begin{array}{c} 3, 2, 2, 2, 2, 1, 0, 0, 0 \end{array}\right]^T$, where l_i is the appropriate lag for input u_i.

It is important to note the monotonic order of the lags in L is actually representative of the physical layout of the system. The flow u_1 is furthest from Sevier River at Elsinore and u_9 is the closest.

The SRWUA's previous estimate of the lags, based on geographic factors, average flows, and prior experience, was $\left[\begin{array}{c} 5, 4, 3, 3, 3, 2, 1, 1, 1 \end{array}\right]^T$; however, a comparison of these two sets of lags using a generic mass balance model on the validation data showed that the lags we determined based on statistical correlation were consistently closer to the true downstream flow than those based on SRWUA's approximate lags. After this process of determining the approximate system lags L, we used this information to formulate models representative of the river basin.

C. Model Classes

We compared three different model classes in the process of developing a model for the Sevier River: a mass balance model, a parameterized second-order model with delay, and a parameterized mass balance model. We show the form of each model along with the results of identification and validation below. We used standard linear regression techniques to fit the parameters of the models.

1) Mass Balance Model: A mass balance model simply states that the amount of water flowing into a river will be the same amount that flows out of the river. This MISO mass balance model was adapted from a similar SISO model developed by Weyer in [6].

$$y(k) = \sum_{i=1}^{9} a_i u_i(k - l_i)$$

Weyer used a parameterized mass balance model to estimate flows from the height of the water running over a gate. Since the measurement devices for the Sevier River already calculate flow, we modified the model by fixing the parameters for each flow in the following form:

$$A = \left[\begin{array}{c} 1, 1, -1, -1, -1, -1, -1, -1, -1, -1 \end{array}\right]^T.$$

The values of 1 as weights indicate inflows; whereas, the values of -1 indicate outflows.

It should be noted that although a mass balance model gives a good approximation of river flows it also has weaknesses. Due to evaporation, seepage, return flows, unregulated inflows, and other disturbances, the sum of the water flowing from the inflows does not strictly equal the sum of the water flowing through the outflows. Consequently, none of these phenomenon will be captured with a strictly mass balance approach.

2) Parameterized Second-Order Model with Delay: The Saint-Venant equations are two partial differential equations that are commonly used to model open-channel dynamics. As shown by Litrico, under simplifying assumptions such as uniform width and depth these equations can be simplified to the diffuse wave equation. Linearizing around a reference discharge results in the Hayami equation. This equation can be sufficiently approximated by a second-order system plus delay [4].

Additionally, in [7], Ooi, et al. show that this model can be fitted using black box system identification methods as opposed to physics-based derivations with comparable results.

Based upon these two results, we chose a second-order parameterized model with delay to approximate the Hayami equation and model the Sevier River Basin. After evaluating ARX, ARMAX, Output-Error, and Box Jenkins models we found that for this stretch of river, the ARX and ARMAX models were most representative. Specifically, the two models were very similar, yet the ARMAX model performed slightly better than the ARX model. In an attempt to connect to the previous literature on this topic, we provide the analysis of the ARX model while noting that the same discussion hold for the ARMAX model (except for a slight decrease in the errors). The ARX model considered is

$$y(k) = b_1 y(k-1) + b_2 y(k-2) + \sum_{i=1}^{9} a_i u_i(k - l_i)$$

The parameters obtained using a linear regression on the first three years of data are

$$\begin{bmatrix} b_1 \\ b_2 \\ a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \\ a_6 \\ a_7 \\ a_8 \\ a_9 \end{bmatrix} = \begin{bmatrix} 1.2841 \\ -.3004 \\ .0115 \\ .0194 \\ -.0027 \\ -.0144 \\ -.0173 \\ -.0075 \\ -.0034 \\ -.0074 \\ -.0088 \end{bmatrix}.$$

These parameter fittings are reasonable in relation to proper identification of inflows and outflows. u_1 and u_2 are the only inflows into the river system, and a_1 and a_2 are the only parameters greater than zero. This is what we would expect of a properly fitted model.

Additionally, the average weight of the inflows $\frac{(a_1+a_2)}{2} = .015$ is similar to the average weight of the outflows $(\sum_{i=3}^{9} a_i)/7 = .018$, suggesting consistency with the mass balance model. However, upon closer analysis, the stretch of the Sevier River Basin may not fit a mass balance framework due to unmeasured inflows.

To see this effect, we calculate the average difference between inflows and outflows Q_{diff} as follows,

$$Q_{diff} = \frac{\sum_{k=1}^{n} \sum_{i=1}^{2} u_i(k - l_i) - \sum_{k=1}^{n} \sum_{i=3}^{9} u_i(k - l_i)}{n},$$

where n is the number of data points in the historical data set. For this stretch of the river we determined that there is a Q_{diff} of -0.37 m^3/s. This implies that water is consistently gained from unmeasured inflows between Sevier River at Clear Creek and Sevier River at Elsinore. With an average flow of 6.16 m^3/s this represents about 6% of the upstream flow. The Sevier River system clearly does not fit well into a strict mass balance model.

D. Parameterized Mass Balance Model

Using information concerning the Sevier River Basin and techniques from each of the two previous models, we were able to develop a more accurate model for this system.

Along the Sevier River, every device that measures diverted water from the river is located very close to the actual diversion gate. Consequently, we can assume that the measurement for every outflow is exact (i.e. the weight is 1). The only circumstance that would invalidate this assumption would be if there was some problem with the sensor; however, even in this instance we do not want to include those effects in our model since these problems are only temporary.

With the outflow weights, $a_3, \ldots, a_9$, fixed at unity, we are still able to parameterize the input weights a_1 and a_2. This is justifiable because the locations that measure the inflows to the river are physically separated from the final measurements downstream. Thus, the accuracy of the inflow measurements are disrupted by evaporation, seepage, and other unmeasured disturbances.

The result of this analysis is a partially parameterized mass balance equation.

$$y(k) = a_1 u_1(k - l_i) + a_2 u_2(k - l_2) - \sum_{i=3}^{9} u_i(k - l_i)$$

Using the three years of training data to fit the model we get

$$\begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 0.9841 \\ 1.4164 \end{bmatrix}.$$

Before continuing we need to validate that these parameters are reasonable. Earlier we discussed that the Sevier River actually gains water as it flows downstream. For this reason it makes sense that the average weight for these two inflows would be larger than one. Additionally, the weight for Clear Creek is greater than one and the weight for Sevier River above Clear Creek is less than one. This is also to be expected.

Clear Creek is an unregulated inflow. This means that Clear Creek is likely very representative of the conditions around the river, such as the weather. For example, when the snow pack begins to melt off of the mountains in the Spring, both the flow down Clear Creek and unmeasured flows will be higher than normal. However, since the Sevier River above Clear Creek is a regulated flow we would expect it to contain very little information about unmeasured inflows. Consequently, it is not unreasonable that the weight of Clear Creek be above one (to include unmeasured inflows) and the weight of Sevier River to be slightly below one (to compensate for evaporation and seepage). In general, we would expect the first parameter to be slightly lower during hot seasons and the second parameter to be slightly higher during wet seasons.

IV. RESULTS AND ANALYSIS

A. Mass Balance Model

The results of the mass balance model validation are shown in Figure 3. It should be noted that the large errors between about 4500 hours and 7500 hours are due to sensor error in the validation set and should not be considered representative of this model or any of the other models. Consequently, the calculation of error metrics for these models do not include the hours from 4500 to 7500.

The mass balance model does a fairly good job of capturing major system events like the large flows at 8000, 14000, and 16000 hours. However, the mass balance model is heavily influenced by erroneous data such as the large spikes near 11000 hours and is not very good a predicting during times of low flow such as the span from 9000 to 13000 hours.

B. Parameterized Second-Order Model with Delay

The validation results of the parameterized second-order ARX model with delay are shown in Figure 4. This model performs well at capturing general trends of the river system; however, it is not very good at estimating the true flow of the system during periods of higher flow. This model is very good at estimating the flow of the system during times of low flow (9000 to 15000) and especially when the flow is rather erratic (hours 1 to 3000 and 9000 to 11000). Finally, it should be noted that the ARMAX model performed slightly better than the ARX model for periods of higher flow, but the ARMAX model was still significantly lower than the two mass balance models for these periods.

C. Parameterized Mass Balance Model

The parameterized mass balance model is heavily influenced by erroneous data such as the large spikes near 11000 hours. Nevertheless, it does a good job of capturing the trends of the river during high flow conditions such as those at 8000, 14000, 16000, and even the flood conditions as 18000. As expected, this model behaves much like the mass balance model; however, it usually estimates a little higher than the mass balance model does–especially when there is significant flow in Clear Creek.

To accurately compare the system identification methods used we split the validation set into different ranges that are respective of various system conditions. These ranges represent the system in periods of low flow, high flow, rapidly changing flow, and overall flow. Additionally, the final condition of the river is represented in the last two and a half months of data. This condition is that of severe flood conditions during the Spring of 2005. The periods of time that are included for each range is summarized in Table I.

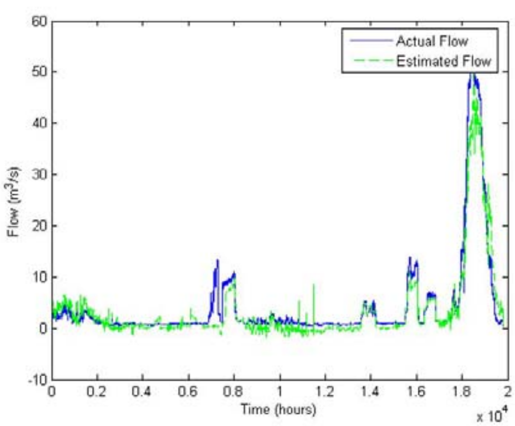

Fig. 3. Validation of Mass Balance Model

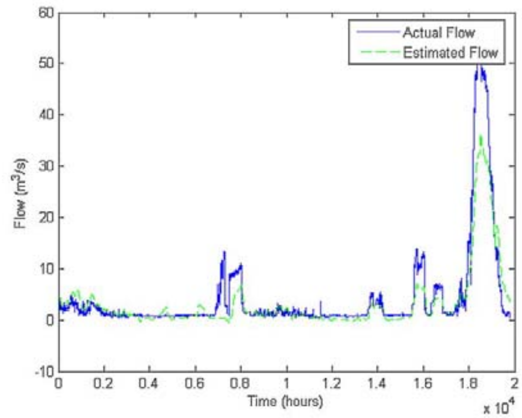

Fig. 4. Validation of Second-Order Model with Delay

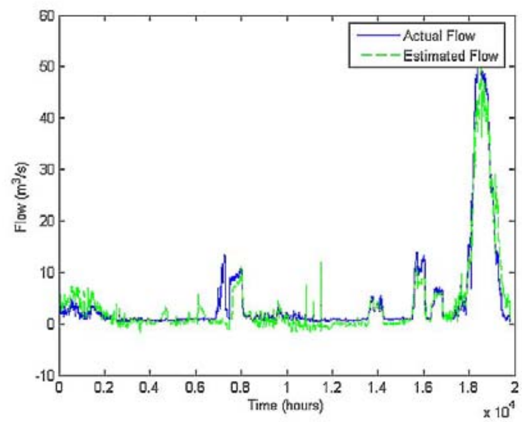

Fig. 5. Validation of Parameterized Mass Balance

The root mean square error and the mean absolute error for each model and range is shown in Table II.

D. Comparative Analysis

As can be seen from Table II, the mass balance model and the parameterized mass balance model perform comparably during most conditions on the river. The parameterized mass balance model usually has a little higher error than the usual mass balance model; however, it appears that the parameterized version is slightly better at times of high flow and much better during flood-like conditions. Thus it appears that the parameterized model is best suited for high flow conditions.

The parameterized second-order model with delay performed significantly better during times of low flow and rapidly changing flow than the other two models did. Unfortunately, this model was not very accurate at estimating during times of high flow, and was especially bad during flood conditions.

It should be noted that all three models tend to underestimate the flow during high flow levels. This discrepancy is

TABLE I
RANGES OF SYSTEM CONDITIONS

Period	Range (hours)
Low Flow	1-3000, 9000-13000
High Flow	7700-8200, 15490-16870
Changing Flow	1-2000, 9000-11000
Overall Flow	1-4000, 8000-18000
Flood	18000-19000

TABLE II
MODEL VALIDATION ERRORS: ROOT MEAN SQUARED(m^3/s), MEAN ABSOLUTE ERROR(m^3/s)

Flow Type	Mass Balance		Second Order		Param. Balance	
Low	1.17	1.05	0.95	0.74	1.30	1.09
High	1.74	1.38	3.57	2.97	1.74	1.35
Changing	1.31	1.18	1.14	0.87	1.50	1.26
Overall	1.13	0.91	1.51	0.94	1.19	0.94
Flood	7.13	6.12	11.2	9.19	5.90	4.95

most likely due to return flows coming from the farmers' excess irrigation water. Unfortunately, these return flows are not measured and are more related to operational conditions of each farmer than general seasonal, environmental, or flow conditions. Lack of data concerning these operational conditions creates a modeling limitation which will be addressed in the feedback control design.

From a water management perspective, accurately modeling the river system is most crucial during times of high flow such as the irrigation season. This is because at these times the demand for water and the potential for water conservation are the highest. Thus the most profitable model must estimate high flow conditions well. Consequently, we chose parametric mass balance model as the preferred model for the Sevier River. This model performs at least as well as the other models on high flow conditions and ever better than the other two during extreme high flow conditions.

V. CONCLUSIONS AND FUTURE WORKS

A. Conclusions

In this paper, we described the physical layout of the Sevier River Basin including data collection and dissemination related to Piute Dam. We determined the appropriate lags for the river system using statistical correlation analysis and validated those results. We also developed three different model types for this MISO system, a mass balance model, a parametric second-order model with delay, and a parametric mass balance model.

The purpose of this modeling is to create a control-oriented model capable of estimating downstream flow given upstream flow and relevant inflows/diversions. We used three years of data to fit our parametric models and two years of data to validate the models. We analyzed the results of this validation set and concluded that the parametric mass balance model works best for the Sevier River Basin during times of high flow. Since times of high flow are most important for water management purposes, this model was chosen as the preferred model for this system.

B. Future Works

The parametric mass balance model developed from this system identification procedure will be used as the basis for the design of a robust feedback controller for the Piute Dam. With the cooperation of the SRWUA and the USBR, starting in the Spring of 2006, this controller will be used to automatically adjust the Piute Reservoir release gate in real-time. The goal of this controller will be to regulate the flow at Sevier River at Vermillion to a specified amount despite disturbances such as evaporation, rain, runoff, and unregulated inflow (Clear Creek).

The technical framework for this automation process has already been developed. This framework will allow the SRWUA to dynamically specify target flow values for each of the diversions and for the downstream flow at Vermillion. A software package developed by the author will calculate the reservoir release each hour based on real-time data and autonomously adjust the gate height at Piute Reservoir and each of the downstream diversions as described in [11].

VI. ACKNOWLEDGMENTS

The authors gratefully acknowledge the contribution of Roger Hansen of the U.S. Bureau of Reclamation's Provo Area Office who provided many insights, clarifications, and resources to further this water management research.

REFERENCES

[1] P.O. Malaterre and B.P. Baume, "Modeling and regulation of irrigation canals: existing applications and ongoing researches", *in Proceedings of IEEE Conference on System, Man, and Cybernetics*", San Diego, CA, USA, 1998, pp. 3850-3855.

[2] X. Litrico, D. Georges, and J.-L. Trouvat, "Modelling and robust control of a dam-river system", *in IEEE International Conference on Systems, Man, and Cybernetics*, San Diego, CA, USA, 1998, vol. 4, pp. 3862-3867.

[3] X. Litrico, V. Fromion, F.-P. Baume, and M. Rijo, "Modelling and PI control of an irrigation channel", *in Proceedings of European Control Conference ECC03*, Cambridge, UK, 2003.

[4] X. Litrico, "Robust IMC Flow Control of SIMO Dam-River Open-Channel Systems", *in IEEE Transactions on Control Systems Technology*, Vol. 10, no. 3, pp.432-437.

[5] X. Litrico, and V. Fromion, "Advanced control politics and optimal performance for an irrigation canal", *in Proceedings of European Control Conference ECC03*, Cambridge, UK, 2003.

[6] E. Weyer, "System identification of an open water channel", *in Control Engineering Practice*, vol. 9, pp. 1289-1299, 2002.

[7] S.K. Ooi, M.P.M. Krutzen, and E. Weyer, "On physical and data driven modelling of irrigation channels", *in Control Engineering Practice*, vol. 13, issue 4, April 2005, pp. 461-471.

[8] E. Weyer, "LQ control of an irrigation channel", *in Proceedings of the 42nd IEEE CDC*, Hawaii, USA, pp. 750-755, 2003.

[9] Y. Li, M. Cantoni, and E. Weyer, "Design of a Centralized Controller for an Irrigation Channel Using H_∞ Loop-shaping", *in Control 2004*, University of Bath, U.K., September 2004.

[10] B. Berger and M. Maxwell, "OpenBasin: An Open Source Software Framework for Real-time Web-based Hydrologic Data Dissemination", *in Proceedings U.S. Committee on Irrigation and Drainage Conference on Water Management*, Salt Lake City, UT, USA, Oct. 2004, pp. 365-373.

[11] M. Maxwell, "Real-Time Model-Based Dam Automation: A Case Study of the Piute Dam", *in Proceedings U.S. Committee on Irrigation and Drainage Conference on SCADA Technologies*, Vancouver, WA, USA, Oct. 2005.