Identification of Open-Channel Characteristics

A. HOF, L.G. SILVIS AND R. BROUWER
Delft University of Technology, Civil Engineering, Land and Water Management, PO Box 5048 2600 GA Delft, the Netherlands
P.M.J. VAN DEN HOF
Delft University of Technology, Mechanical Engineering Systems and Control Group, Delft, the Netherlands
A.J. CLEMMENS
US Water Conservation Laboratory, United States Department of Agriculture, USDA-ARS, Phoenix, U.S.A.

Summary: Efficient design methods for automated control systems have been developed in control engineering. If these methods are to be used to design a linear control system for irrigation canals this will require a linearized model of the water movement in a given irrigation canal. The simplest model is probably the Integrator-Delay model that contains two parameters, a delay time and the surface area of the backwater curve. Both these parameters can be deduced from the geometry and roughness of the canal. Very often, however, the necessary physical parameters such as canal geometry and roughness are not known. In control engineering a technique known as system identification (building a model based on measurement data of the system) has proven to be useful for overcoming such problems. System identification was applied successfully as a test case at the Upper Arizona Canal near Phoenix. This paper describes the principle of system identification and its application at the Upper Arizona Canal.

1 INTRODUCTION

The water movement in irrigation canals can be modeled quite accurately on the Saint Venant equations. This is a set of non-linear partial differential equations that is too complex for control systems design. To overcome this problem the Saint Venant equations can be linearized and simplified to the following Integrator-Delay model (Schuurmans et al. 1995):

$$h(x,s) = \frac{1}{As}\left(q_{(0,s)}e^{-\tau s} - q_{(L,s)}\right) \qquad (1)$$

Where A = surface area of the backwater curve (m^2); τ = delay time (s); s = Laplace variable; q = flow rate (m^3/s); h = water level in the backwater part (m); x = distance (m).

The approximation of the water movement by the ID-model divides a canal reach into two parts (Figure 1).

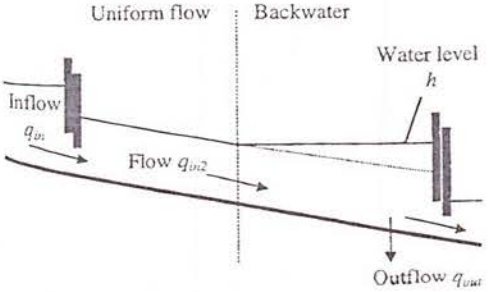

Figure 1 The ID-model divides canal into two parts

One part has a uniform flow that is approximated by a pure delay time. In this section it is assumed that waves will travel only in a downstream direction and that the deformation of waves can be ignored. In the backwater part the dynamics become complicated. Resonance waves move up and down through the canal and reflect against the boundaries. However, at low frequencies, the water level just 'integrates' flow variations in the backwater part. A detailed derivation of the ID-model is described by (Schuurmans et al., 1995.)

The ID-model of equation (1) is expressed in the continuous Laplace domain. In real existing systems, measured data and control signals are never continuous. Due to sampling these signals are discrete. In that context it is therefore better to use the discrete counterpart of the Laplace transform, the z-transform (e.g. Ljung and Glad, 1994). The z-variable can be expressed as:

$$z^{-1}h(t) = h(t - T_S) \qquad (2)$$

Here T_s = sample time. Assuming that all signals are measured at a sample time of T_s and that τ/T_s is an integer, the backwater coefficient b (sm^{-2}) and the delay nk (-) can be defined as:

$$b = \frac{T_s}{A} \quad \text{and} \quad nk = \frac{\tau}{T_s} \qquad (3)$$

With equations (2) and (3) it is possible to rewrite equation (1) to:

$$h(t) = \frac{b}{1-z^{-1}}\left(z^{-nk} q_{in}(t) - q_{out}(t)\right) \quad or$$
$$h(t) = z^{-nk} G(z) q_{in}(t) - G(z) q_{out}(t) \qquad (4)$$

Here $G(z)$ is the transfer function that models the part of the canal that is affected by the backwater. The discrete ID-model has only two parameters: the backwater coefficient b and the delay nk, both of which are in fact rewritten expressions of the backwater curve surface area A and of the delay time τ. Both these parameters can be computed quite accurately from the geometry of the canal and from the steady state values of the flow rates and water depths. Very often however, the geometrical data is lacking or is hard to obtain. Applying system identification can be one way of overcoming the problem of lack of information.

The system identification technique was applied at the Arizona Canal. The canal is 30.5 kms long and consists of five reaches with a capacity of 53 m^3/s. The two upstream reaches are long and steep in relation to the three lower reaches. The canal is lined with reinforced concrete sprayed over the existing surface. Water transported by SRP is monitored and controlled by one or two watermasters stationed in the water control room. A Supervisory Control and Data Acquisition (SCADA) system is used for that purpose. The results of the experiments on the second and third reach of the Upper Arizona Canal are worked out in this paper, which will lead to a good understanding of the possibilities and the limitations of system identification.

2 PRINCIPLES OF SYSTEM IDENTIFICATION.

System identification uses sets of measured data to determine a model. System identification is based on the fact that a system can be seen to exhibit itself through its externally measurable signals. For a canal system the measurable signals are the flow rates on entering and leaving the canal and the water level in the canal. Measurements taken from the dynamic system and used for parameter estimation will always be contaminated by disturbance signals such as measurement noise. This can be accounted for in the system equations by adding a disturbance term $v(t)$, which is supposed to reflect the deviations that occur in measurement data around the condition in which the assumption of linearity is valid.

$$h(t) = \frac{b}{1-z^{-1}}\left(z^{-nk} q_{in}(t) - q_{out}(t)\right) + v(t) \qquad (5)$$

When performing system identification using prediction error methods in the way that will be done in this paper, the disturbance $v(t)$ is modeled as a sequence of zero-mean identically distributed, independent random variables (Ljung and Glad, 1994). This means that we can write:

$$v(t) = H_0(z) w(t) \qquad (6)$$

in which $H_0(z)$ is a proper rational transfer function that is stable, and $w(t)$ is a sequence of zero mean, identically distributed, independent random variables (white noise). The basic description that is going to be used to identify the open-channel system is:

$$h(t) = z^{-nk} G(z) q_{in}(t) - G(z) q_{out}(t) + H_0(z) w(t) \qquad (7)$$

Ljung (1997) describes the following basic steps of system identification:

- Design the experimental set up and collect the relevant data.
- Examine the data
- Select and define a model structure
- Estimate parameters based on measurement data
- Examine/validate the model.

3 SET UP OF EXPERIMENTS

First it is necessary to design the experiment and collect input and output data on the process that is to be identified.

In order to capture the dynamics of a system the input variables of that system should vary sufficiently. In the case of a channel reach the two input signals that can be distinguished are the flow rates on entering and leaving the canal, which are q_{in} and q_{out}. (see also equation 1). In this particular case with an open-channel flow it would be sufficient to just vary the inflow q_{in}. In the block diagram of Figure 2 it can be seen that if q_{in} is varied all the dynamics will be excited.

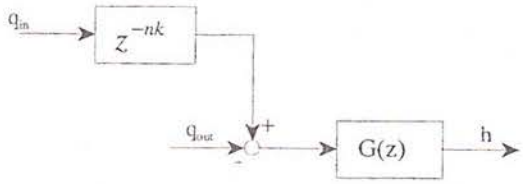

Figure 2 Block diagram of ID-model

If on the other hand, only the outflow q_{out} is varied then it will not be possible, by means of system identification, to find the part of the system that has a delay time.

For the experiments done in the Upper Arizona Canal it was decided that the flow into the canal reach should be varied. The openings of the gates at the downstream end of the canal reach were fixed.

4 INPUT SIGNAL

The discharge into the reach, which constitutes the input of the experiment, should vary sufficiently to capture the dynamics of the canal system in experimental conditions. It is important that the input covers a wide frequency range (Ljung and Glad, 1994). Besides discrete-time white noise, that has a flat frequency spectrum, there is an alternative signal, the so-called Pseudo Random Binary Signal (PRBS). This signal shifts randomly between two levels at discrete points in time. Examples of these signals shown in Figure 3.

A PRBS is a popular test input signal for system identification, because of its ability to prescribe the amplitude of excitation, which is something that can not be done for a white noise signal (Ljung and Glad, 1994). Furthermore,

canal operators are very familiar with step-wise discharge changes, which is what makes the PRBS 'user friendly' for experiments on irrigation canals.

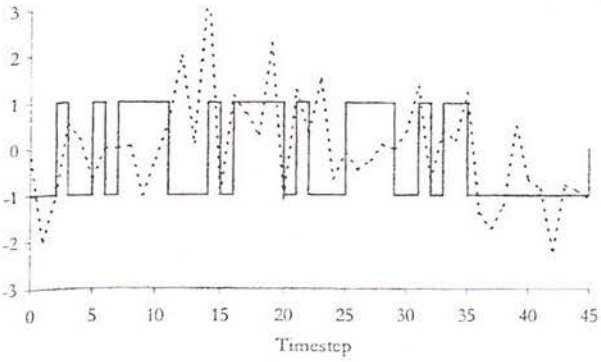

Figure 3 Examples of white noise signal (dotted) and PRBS (solid)

The PRBS switches between two levels around the nominal nflow. The amplitude chosen should be such that the flow ate variation results in a measurable and safe water level ariation at the downstream end of the reach. In the Upper Arizona Canal, an amplitude of between 5 and 10% of the ominal flow was used. Finding the proper amplitude might nvolve some trial and error, but very often a good first stimate can be made on the basis of the knowledge of the anal operators.

When the basic interval choice between two consecutive step hanges is made the following has to be considered. The ffects of friction gradually reduce the height of a wave in a anal: the wave flattens out propagating downstream (Chow, 959). So, if the basic interval is chosen too small, two onsecutive waves can merge into each other, and cannot be cognized as the consequence of different step changes. nother important aspect when designing the basic interval is at of the sample time that is going to be used for entification. The sample time should be smaller than the asic interval, otherwise the sampled input signal will not present the actual input. Taking these two aspects into onsideration, the minimum time between two consecutive ep changes can be roughly estimated to be half the estimated ne delay. This value depends on the characteristics of the nal, which are unknown at the experiment design stage. In se of the Upper Arizona Canal the estimates of the canal erators were used for this purpose. For reach two the delay ne was estimated at 90 minutes and for reach three at 30 nutes.

practice, identification experiments have to be of limited ration, because normal operation of the canal will have ority. A general rule of thumb for the lower limit of the al duration of the experiment is given by 5-10 times the gest relevant time constant in the process. (Ljung and Glad, 94). In this case of open-channel flow it is possible to use time delay as estimated by the canal operators for that pose. For the second reach the total duration of the eriment came to 540 minutes and for reach three to 240 nutes.

5 EXAMINATION OF DATA

Proper examination of the data, prior to parameter estimation, is a necessity if consistent results are to be obtained. The raw process data consists of measured values of the discharge into the reach and of the water level at the downstream end of the reach. The first step when preprocessing data is to choose a data part of which is known that the data acquisition process was completed in a proper way.

Visual inspection of the data record will lead to the detection of drastic outliers. If outliers are accepted uncritically, they will have a devastating effect on the estimated model. The quadratic criterion that is presented later in this paper gives disproportionately large weight to data points that give large prediction errors. To protect the identification process from this, outliers were removed by means of linear interpolation between good measurements that border on outliers.

Any disturbances acting on the process may contribute to process output as well. These contributions are often slowly varying trends, which are not correlated to the input. For example, the flow through the downstream gate can slowly increase due to a fall in water level downstream of that gate. This may result in a water level that drifts away. Because the data set used for the system identification experiment is of a limited length, these trends may significantly affect the identification results. In this study, linear trends from the signal were removed after a least-squares fit.

To get proper identification, it is necessary to make the average input and the signal zero (Ljung, 1987). Subtracting the mean of the data from the original data can do this.

6 MODEL STRUCTURE AND EVALUATION

The system identification procedure includes selecting an appropriate model set. Going on sets of measurements of the inflow, the outflow and the water level at the downstream end, it is the model within this set that minimizes a certain criterion that is selected. Consider the situation as depicted in the diagram in Figure 4.

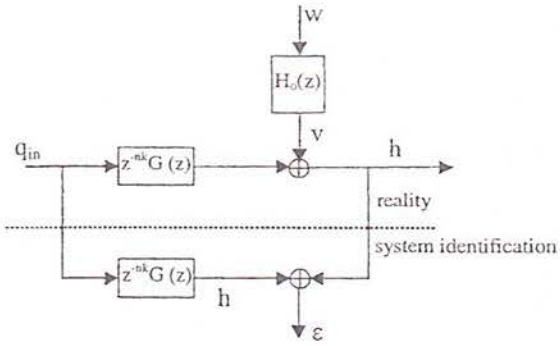

Figure 4 Scheme of System Identification

A simple model structure is the ARX parameterization. (Ljung and Glad, 1994) The ARX parameterization leads to the following transfer functions for G (canal model) and for H (noise model):

$$G(z,\theta)=\frac{B(z)}{A(z)}=\frac{b_1+b_2z^{-1}+\ldots+b_{nb}z^{-nb+1}}{1+a_1z^{-1}+\ldots+a_{na}z^{-na}} \qquad (8)$$

$$H(z,\theta)=\frac{1}{A(z)}=\frac{1}{1+a_1z^{-1}+\ldots+a_{na}z^{-na}} \qquad (9)$$

$$\theta^T=[a_1\ a_2\ \ldots\ a_{na}\quad b_1\ b_2\ \ldots\ b_{nb}] \qquad (10)$$

Here, the integers na and nb are the orders of the polynomials. The number nk represents the number of delays from input to output.

The capability of the model to predict future values of signals is an important property when evaluating models. For each value of the parameter vector θ the model provides us with an estimation, a prediction of $\hat{h}(t,\theta)$. The prediction can be written in the following form:

$$\hat{h}(t)=\varphi^T(t)\cdot\theta \qquad (11)$$

Here $\hat{h}(t)$ = the predicted water level; $\varphi(t)$ = the regression vector and θ = the parameter vector, which in the ARX case for open-channel flow satisfy:

$$\varphi^T=\left[-h(t-1)\ldots-h(t-na)\ q_{in}(t)\ldots q_{in}(t-nb)\right] \qquad (12)$$

$$\theta^T=[a_1\ a_2\ \ldots\ a_{na}\quad b_1\ b_2\ \ldots\ b_{nb}] \qquad (10)$$

What is evaluated is how good this prediction is at time t. This is done by calculating the prediction error ε.

$$\varepsilon(t,\theta)=h(t)-\hat{h}(t)=h(t)-\varphi^T(t)\cdot\theta \qquad (13)$$

Here $h(t)$ is the measured value from the experiments.

If data sequences of the inflow $q_{in}(t)$ and the water level $h(t)$ are collected over a period t=1,2,...,N, it is possible to evaluate how well the model with parameter value θ describes the performance of the open-channel system. An accurate model generates 'small' prediction errors.

The most simple and most frequently used criterion function in system identification is the quadratic function:

$$V_N(\theta)=\frac{1}{N}\sum_{t=1}^{N}\left(h(t)-\varphi^T\cdot\theta\right)^2 \qquad (14)$$

Here $V_N(\theta)$ is the cost criterion that is to be minimized. The solution can be found by equating the first derivative of the quadratic function to zero and by calculating the corresponding parameter values. This strategy leads to the optimal parameter set $\bar{\theta}$:

$$\bar{\theta}=\left[\frac{1}{N}\sum_{t=1}^{N}\varphi(t)\varphi^T(t)\right]^{-1}\cdot\left[\frac{1}{N}\sum_{t=1}^{N}\varphi(t)h(t)\right] \qquad (15)$$

Equation (15) defines the model generating the smallest prediction error.

7 FILTERING THE DATA

As explained in (Ljung and Glad, 1994) the ARX-model structure has a natural tendency to stress high frequency behavior of the model. As mentioned in the introduction of this paper, the strength of the ID-model is its low frequency behavior. Therefore, the results of the identification using an ARX structure can be improved by prefiltering the input and output through a low pass filter:

$$h_f(t)=L(z)\cdot h(t)$$

$$q_f(t)=L(z)\cdot q(t)$$

Here L(z)=the low-pass filter. A simple and useful procedure to design such low pass filter was described by (Ljung and Glad, 1994). First an ARX-model is fit to the unfiltered data, which will give an estimate of the transfer function $G(z)$. Afterwards the estimate of the denominator is used to filter the data

$$L(z)=\frac{1}{1-a_1\cdot z^{-1}}$$

In this way the focus of the model fit can be moved to the lower frequencies.

8 IDENTIFICATION

While the experiments were executed the gate openings at the downstream end of the canal reach were fixed. Consequently, the water level at the downstream end is not only affected by the inflowing discharge but by the varying outflow as well. The flow rate through the downstream structure depends on the water level upstream of the structure and on the position of the structure. Hence the water levels at the downstream end of the canal and the flow rate on leaving the canal depend on each other. By fixing the gate opening at the downstream end of the canal a feedback mechanism is actually created (see Figure 5).

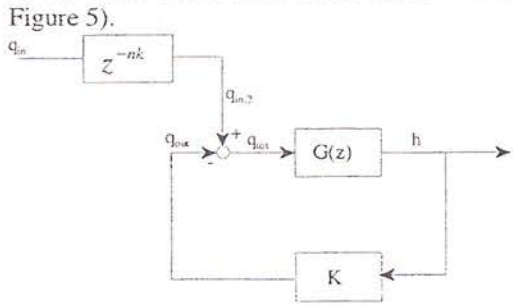

Figure 5 Block diagram of system during experiment

The feedback mechanism is a proportional one (K=constant). This can be seen quite easily from a linearization of the gate-discharge equation. (Silvis et al., 1998)

In the Upper Arizona Canal a three-step identification procedure was applied. (e.g. van den Hof and Schrama, 1993) The first step involved estimating the time delay. The data series of the flow rate q_{in} and the water level h were used to identify the model that describes the relation between q_{in} and h. While doing this one should realize that it is not the model of the water movement that is estimated. Instead a model that describes the closed loop system, that is created by the feedback mechanism, is estimated.

$$h(t) = G(z)F(z)q_{in} + F(z)v(t) \tag{16}$$

$$F(z) = \frac{1}{1+KG(z)} \tag{17}$$

But by identifying the model $G(z)F(z)$ the delay nk can be obtained. All reasonable delays were tested in a fourth order ARX. The delay minimizing the identification criterion of equation (14) was selected. This resulted in a delay nk of 40 minutes for reach two and 16 minutes for reach three of the Upper Arizona Canal.

The inflow q_{in} can be shifted nk time steps which will lead to the delayed inflow q_{in2}. Next, the total flow rate into the backwater area q_{tot} can be computed by subtracting the measured outflow q_{out} from the delayed inflow q_{in2}.

$$q_{tot} = q_{in2} - q_{out} \tag{18}$$

The relation between the total discharge and the water level h is the model $G(z)$, that describes the backwater part of the canal, can be extended by the following relation which describes the influence of the feedback loop on the input signal:

$$q_{tot}(t) = F(z)q_{in2}(t) - Kh(t) \tag{19}$$

Due to the presence of the "proportional controller" K, the input $q_{tot}(t)$ is not uncorrelated to the noise $v(t)$, which makes consistent open-loop identification of $G(z)$ difficult. When using equation (17) and (18), equation (19) can be rewritten as:

$$q_{tot} = F(z)q_{in2}(t) + F(z)Kv(t) \tag{20}$$

The following step of the three-step method identifies the transfer function $F(z)$ by parameterizing a high order ARX-model (e.g. fourth order) using the delayed inflow $q_{in2}(t)$ and the signal $q_{tot}(t)$. This is an open-loop type of problem as $q_{in2}(t)$ and $v(t)$ are uncorrelated. Next, the estimate of the sensitivity function

$$(z) = \frac{1}{1+KG(z)} \tag{17}$$

used to reconstruct a noise-free signal $\hat{q}_{tot}(t)$:

$$q_{tot} = \hat{F}(z)q_{in2}(t) \tag{21}$$

Next, the physical knowledge present in the ID-model is used. The total discharge q_{tot} is integrated.

$$q_{tot2}(t) = \frac{1}{1-z^{-1}}\hat{q}_{tot}(t) \tag{22}$$

The final step of the procedure consists of the application of the standard identification method. The backwater coefficient b is estimated by minimizing the identification criterion as explained before. Here the water level h is the output signal and the integrated total discharge q_{tot2} the input.

For this identification an ARX-model with $na = 0$ and $nb = 1$ was applied. For reaches two and three of the Arizona Canal the following values b=0.0058 and b=0.0026 were found respectively. With the aid of equation (3) these two values for b can be transferred to the following backwater curve surface areas: reach two 10.3 ha and reach three 9.2 ha.

9 RESULTS

The system identification lead to parameters of the ID-model as presented in table 1.

Table 1 Results of system identification

reach number	Delay time [min.]	Back water Surface area [ha.]
2	40	10.3
3	16	9.2

These identified models were used to simulate the water movement in the reaches using measurement data of the inflow and the outflow as input signal. The output of this simulation should be in accordance with the measured water level at the downstream end of the canal reach. Both data sequences are plotted in Figure 6 for the second reach of the canal. It can be seen that the identified model overestimates the water level variation in the beginning, but is very well capable of describing the peaks after three, six and eight hours.

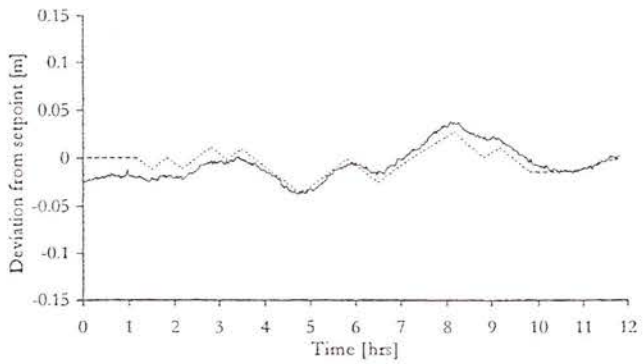

Figure 6 Reach two, simulation results (dotted) and measurements (solid)

Figure 7 shows the results of reach three. It can be seen that the identified model gives an adequate description of the water levels.

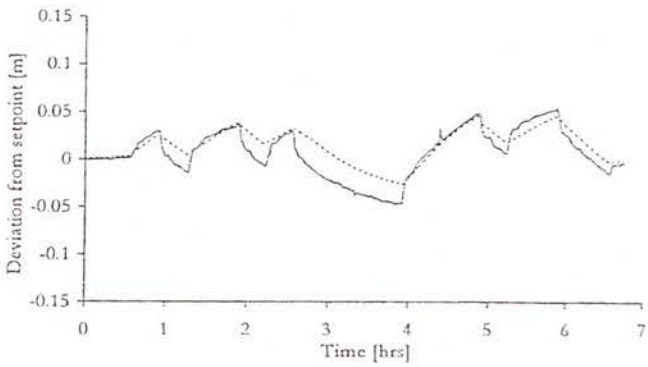

Figure 7 Reach three, simulation results (dotted) and measurements (solid)

The experiments on the Upper Arizona Canal showed that system identification can lead to an accurate linear model of the water movement that can be used for the design of control systems. Further, it was found that for situations with long delays the deduced models were less accurate. This was not caused by the system identification procedure but caused by the assumptions that were made deriving the linear model. It was assumed that in a part of the canal with uniform flow, the water movement could be modeled by a pure delay. In practice however this is not true and a dispersion of waves will occur. The assumption has no consequences for small delay times but causes larger errors as the delay time increases. The ID-model is less accurate for long delays.

The time delays estimated using system identification differ from the values used by SRP's watermasters (table 2). The latter are much higher. This can be explained by the fact that the time delay of the model and the time delay used by the operator have different backgrounds. The identified delay reflects the time between the upstream discharge change and the moment the first part of the wave arrives at the downstream end. The delay used by the watermasters is based on their operational point of view. For them it is more important to know when they can schedule the full discharge change to the downstream part of the channel system.

Table 2. Comparison of time delays

Reach number	Time Delay [min.]	
	Identification	Watermasters
2	40	90
3	16	30

The experiments as described in this paper were carried out with fixed gate positions at the downstream end of the canal. This introduced a feedback mechanism into the system. It was therefore necessary to apply a three-step identification. If on the other hand a flow controller had been used at the downstream end, which kept the flow rate constant despite water level changes, the feedback mechanism would be avoided. The delay time *nk* and the transfer function *G(z)* could have been found in one step.

A linear approximation was made for the water movement of the canal. In reality the water movement is not linear and the assumption of linearity in valid only for small variations around the nominal inflow. Simulations on a non-linear computer model showed that the identification procedure would lead to more accurate models if the step changes of the input signal were smaller (Silvis et al., 1998). But in practice the minimum size of the step changes are limited by the accuracy of the measurement equipment. If the step changes are too small the measurement equipment will not notice any variations. The interests of linearity and the capacity of the equipment have to be balanced.

10 CONCLUSIONS

The system identification procedure led to acceptable estimations of both parameters, the delay time and the surface area of the backwater curve, for all reaches of the Upper Arizona Canal. The approximate model thus obtained gave a good description of the canal characteristics at the nominal flow rate. Further, it was found that in situations with relative long delay times the obtained model was less accurate. This was caused by the simplifications that were made to derive the approximate model.

It can be concluded that system identification is a valuable technique that enables the control systems engineer to find a model in situations for which this was too difficult or too laborious to make such a model by schematization of the physical parameters.

REFERENCES

Ljung, L., (1987). "System Identification, theory for the user", New Jersey, PTR Prentice Hall. ISBN: 0-13-881640-9

Ljung, L. and T. Glad (1994). "Modeling of Dynamic Systems" New Jersey, PTR Prentice Hall. ISBN: 0-13-597097-0

Ljung, L., (1997). "The System Identification Toolbox, User's guide" The Mathworks, inc.

Schuurmans, J., O.H. Bosgra and R. Brouwer (1995). "Open Channel flow Model Approximation for Controller Design", App. Math. Modelling 1995, vol. 19, pp.525-530.

Silvis, L.G. and A. Hof (1998). "System Identification at the Upper Arizona Canal", Communications of the Department of Water Management, Environmental and Sanitary Engineering, Delft University of Technology, Faculty of Civil Engineering

Van den Hof, P.M.J. and RJP Schrama, (1993), "An indirect method for transfer function estimation from closed loop data", Automatica, vol. 29, pp. 1523-1527.