Multivariable AR Data Assimilation for Water Level, Flow, and Precipitation Data

Jackson B. Renteria-Mena, Eduardo Giraldo

***Abstract*—A novel method for data assimilation of a multivariable system that describes the behavior of water level, flow, and precipitation variables is presented. The proposed multivariable auto-regressive model considers correlations between the water level, flow, and precipitation and is directly estimated using measurements. In order to obtain the system parameters, a regularized estimation of the model is applied. This estimation is achieved by using the Tikhonov regularization method with generalized cross-validation for parameter selection. The proposed approach is evaluated using data from a Colombian river located in the Chocó department. Therefore, the resulting multivariable autoregressive regularized model is compared with three simultaneous univariable models. An additional comparison is performed by considering a least squares solution for parameter estimation. In addition, the proposed approach is also evaluated for data from the meteorological information center of Argentina. As a result, the proposed regularized method for data assimilation adequately tracks the data dynamics even for rank-deficient scenarios.**

***Index Terms*—Identification, regularization, multivariable, auto-regressive, data assimilation.**

I. Introduction

Chocó department in Colombia is one of the places with the highest average annual precipitation rates. Three main rivers flow through the Chocó department: the Atrato, the Baudó, and the San Juan [1]. Due to their relevance, a monitoring system is required to preserve the security of their nearby inhabitants. The Institute of Hydrology, Meteorology and Environmental Studies of Colombia (IDEAM) is the authority in charge of monitoring and predicting the possible risk to the communities of any variation in level, flow, and precipitation around all the rivers in Colombia. Several stations of monitoring variables, such as level, flow, and precipitation, are installed to evaluate the risk of the towns located around the rivers [2]. However, these monitoring systems do not provide an early warning system. In [3] several prediction methods for emergency management are presented based on statistic analysis, artificial intelligence, and simulation method. In [4] an early warning system based on fuzzy logic model is proposed in order to determine the status of flood disaster. In [5] a analysis of real-time modelling methods for flood forecasting is presented where the system identification and forecasting is preferred due to the system dynamics based on data measurements. Another example is presented in [6], where a decision support system is proposed for early prediction of water rise levels, based on a neural network. However, these methods require an expert knowledge of the system dynamics or large amount of data to obtain an adequate performance.

Several methods for system identification can be used to estimate multivariable data [7], [8]. The methods are based on a polynomial linear representation, including AR models with exogenous inputs [9]. Data assimilation is an alternative to updating the model parameters and improving the estimation of a system. In [10] and [11] are proposed estimation methods based on a Piecewise Auto-Regressive eXogeneous (PWARX) in order to model precipitation, level, and flow data based on. However, the methods do not include correlation and time variability which is inherent to the system dynamics. For example, in [12], a neural network is combined with an ensemble Kalman filter to emulate a dynamic model. In [13], a large fraction of data are used for operational weather forecasts based on the ensemble methodology. In addition, in [14] are evaluated the prediction performances of flood models of a Multiple-Input Single-Output (MISO) Auto regressive with Exogenous Input (ARX) and MISO Auto regressive Moving Average with Exogenous Input (ARMAX) where the ARMAX structure shows a better performance than the ARX structure in terms of the mean squared error. In [15] a prediction model based on multi-layer perceptron networks is presented with and optimized algorithm that improves the peformance of an hydrologycal model. However, these approaches require a large amount of data for reliable estimations (in some cases, as in [15], more than 20 years of data measurements). On another hand, in [16] are proposed optimal combinations for ARX-based forecast models, where the nonlinear models estimate more adequately the system dynamics.

A requirement in some scenarios is to design an estimation method to estimate model dynamics where a reduced amount of data is available [17], which results in a rank-deficient inverse problem [18], [19]. An ill-posed, rank-deficient, and ill-conditioned inverse problem can be solved using regularization approaches like Tikhonov regularization. For example, in [20], a multivariable AR model is proposed to describe the dynamic model of a time series and improve the solution of an inverse problem for state estimation. In [21], an alternative to estimate a model based on a regularized approach is proposed where the estimator successfully suppresses the adverse effects of the output noise. The regularization parameter selection is computed using the generalized cross-validation method, as proposed in [22], [23]. Another approach to obtain the estimation is presented in [24], where a novel stochastic gradient algorithm based on minimum Shannon entropy is proposed to estimate the parameters of an ARX model with random impulse noise by using a reduce amount of data.

This work presents a novel method for data assimilation of a multivariable system that describes the behavior of

Manuscript received May 17, 2022; revised January 26, 2023.

Jackson B. Renteria-Mena is an Adjunct Professor at Universidad Tecnológica del Choco, Quibdó, Colombia. Research group in Automatic Control. E-mail: jackson.renteria@utp.edu.co.

Eduardo Giraldo is a Full Professor at the Department of Electrical Engineering, Universidad Tecnológica de Pereira, Pereira, Colombia. Research group in Automatic Control. E-mail: egiraldos@utp.edu.co.

water level, flow, and precipitation variables. The proposed multivariable AR model considers correlations between the water level, flow, and precipitation and is directly estimated using real measurements. In order to estimate system parameters, a regularized estimation of the model is performed using Tikhonov regularization method with generalized cross-validation for regularization parameter selection. The main contributions of the proposed approach are: first, only a reduced amount of data is required to train the system, second, a linear multivariable model with correlations among inputs and outputs is proposed to model the system dynamics, and third, the proposed approach can be generalized to several data-sets by including single output or multiple outputs. The proposed approach is evaluated by using data measured from a Colombian river located in the Choco department in Colombia and data from the meteorological information center of Argentina. The proposed multivariable autoregressive regularized estimated model is compared with three simultaneous univariable models. An additional comparison is performed by considering a least squares solution for parameter estimation. This paper is organized as follows: In section II are presented the multivariable AR model and the multivariable AR regularized solution. In section III are introduced the results and discussions of data estimation for two databases and several order validations. And finally, in section IV are presented the conclusions and future works.

II. Theoretical framework

A. AR multivariable model

Consider an AR multivariable model described as follows:

$$y[k] + A_1 y[k-1] + A_2 y[k-2] + \cdots + A_p y[k-p] = e[k] \tag{1}$$

where $A_i \in \mathbb{R}^{m\times m}$ are the model matrix parameters, with $i = 1, \ldots, p$ being p the order of the system, $e[k] \in \mathbb{R}^{m\times 1}$ the noise with m the number of outputs, and $y[k] \in \mathbb{R}^{m\times 1}$ the measurement defined as:

$$y[k] = \begin{bmatrix} y_1[k] \\ y_2[k] \\ \vdots \\ y_m[k] \end{bmatrix} \tag{2}$$

Equation (3) can be rewritten as follows:

$$y^T[k] + y^T[k-1]A_1^T + \cdots + y^T[k-p]A_p^T = e^T[k] \tag{3}$$

and then

$$y^T[k] = \begin{bmatrix} -y^T[k-1] & \cdots & y^T[k-p] \end{bmatrix} \begin{bmatrix} A_1^T \\ \vdots \\ A_p^T \end{bmatrix} + e^T[k] \tag{4}$$

By considering the values of $k = 0, \ldots, K$, being K the total number of samples, the following matrix relation can be obtained:

$$\underbrace{\begin{bmatrix} y^T[1] \\ \vdots \\ y^T[k] \\ \vdots \\ y^T[K] \end{bmatrix}}_{Y} = \underbrace{\begin{bmatrix} -y^T[0] & \cdots & 0 \\ \vdots & & \\ -y^T[k-1] & \cdots & -y^T[k-p] \\ \vdots & & \\ -y^T[K-1] & \cdots & -y^T[K-p] \end{bmatrix}}_{M} \underbrace{\begin{bmatrix} A_1^T \\ \vdots \\ A_p^T \end{bmatrix}}_{\Theta} + \underbrace{\begin{bmatrix} e^T[1] \\ \vdots \\ e^T[k] \\ \vdots \\ e^T[K] \end{bmatrix}}_{\epsilon} \tag{5}$$

resulting in a discrete time measurement equation, as proposed in [18], as follows:

$$Y = M\Theta + \epsilon \tag{6}$$

where matrix $Y \in \mathbb{R}^{K\times m}$ holds the measurements, $M \in \mathbb{R}^{K\times(m\times p)}$ is the Hankel matrix that holds the past measurements, and $\Theta \in \mathbb{R}^{(m\times p)\times m}$ is the matrix that include the AR model parameters, and $\epsilon \in \mathbb{R}^{K\times m}$ represents the non-modeled features of the system, i.e. observation noise, and is assumed to be additive, white and Gaussian with zero mean and with covariance matrix defined by C_ϵ.

B. Multivarible AR Regularized Solution

The naive solution of an inverse problem associated to (6) can be achieved by the least squares solution. This can be performed by defining a functional given by

$$J_{LS} = \|Y - M\Theta\|^2_{2,C_\epsilon} \tag{7}$$

or

$$J_{LS} = (Y - M\Theta)^T C_\epsilon^{-1} (Y - M\Theta) \tag{8}$$

and

$$\frac{\partial J_{LS}}{\partial \Theta} = M^T C_\epsilon^{-1} M\Theta - M^T C_\epsilon^{-1} Y \tag{9}$$

by equaling (9) to zero, the following equation is obtained:

$$\widehat{\Theta}_{LS} = (M^T C_\epsilon^{-1} M)^{-1} M^T C_\epsilon^{-1} Y \tag{10}$$

being Θ_{LS} the least-squares solution of the AR model of (6). If $C_\epsilon = I$ the solution for Θ_{LS} proposed in (10) can be simplified to

$$\widehat{\Theta}_{LS} = (M^T M)^{-1} M^T Y \tag{11}$$

However, when the problem is rank-deficient, ill-posed or ill conditioned [18], the application of the Tikhonov Regularization Method can be performed, by defining a functional as follows:

$$J_{Tikh} = \|Y - M\Theta\|^2_{2,C_\epsilon} + \lambda^2 \|\Theta\|_2^2 \tag{12}$$

or

$$J_{Tikh} = (Y - M\Theta)^T C_\epsilon^{-1} (Y - M\Theta) + \lambda^2 \Theta^T \Theta \tag{13}$$

and

$$\frac{\partial J_{Tikh}}{\partial \Theta} = M^T C_\epsilon^{-1} M\Theta - M^T C_\epsilon^{-1} Y + \lambda^2 \Theta \tag{14}$$

by equaling (14) to zero, the following equation is obtained:

$$\widehat{\Theta}_{Tikh} = (M^T C_\epsilon^{-1} M + \lambda^2 I)^{-1} M^T C_\epsilon^{-1} Y \tag{15}$$

being λ the regularization parameter, being $\widehat{\Theta}_{Tikh}$ the regularized AR solution of (6). Equation (15) can be simplified by defining $C_\epsilon = I$, resulting in

$$\widehat{\Theta}_{Tikh} = (M^T M + \lambda^2 I)^{-1} M^T Y \tag{16}$$

It is worth mentioning that the regularization parameter λ is computed by using the generalized cross validation (GCV) method [25]. This method chooses a regularization parameter λ that minimizes the following functional:

$$\Gamma(\lambda) = \frac{\|Y - M\Theta_\lambda\|_2^2}{(\text{trace}(I - M_\lambda))^2} \tag{17}$$

being the influence matrix M_λ defined by

$$M_\lambda = M(M^T M + \lambda^2 I)^{-1} M^T \tag{18}$$

and Θ_λ defined as

$$\Theta_\lambda = (M^T M + \lambda^2 I)^{-1} M^T Y \tag{19}$$

It is worth noting that for the selection of the regularization parameter λ, the function (17) must be evaluated several times for several λ values.

C. AR Hydrological model

In order to consider the correlation of water level, water flow and precipitation, a multivariable AR model structure is selected as presented in (3), being $y[k]$ defined as follows:

$$y[k] = \begin{bmatrix} y_L[k] \\ y_F[k] \\ y_P[k] \end{bmatrix} \tag{20}$$

where $y_L[k]$ is the water level at sample k, $y_F[k]$ is the water flow, and $y_P[k]$ is the precipitation. Therefore, the multivariable AR model can be defined as

$$y[k] = -\sum_{j=1}^{p} A_j y[k-j] + e[k] \tag{21}$$

being $A_j \in \mathbb{R}^{3\times 3}$ the model parameters.

The model parameters are estimated by using (15), resulting in a regularized AR multivariable estimated model represented by Θ_{Tikh}, as follows:

$$\Theta_{Tikh} = \begin{bmatrix} A_1^T \\ \vdots \\ A_p^T \end{bmatrix} \tag{22}$$

The model (22) is updated for each new measurement, by performing the data assimilation task.

An univariable AR model can also be defined for each variable as follows:

$$y_L[k] = -\sum_{j=1}^{p} a_j^L y[k-j] + e_L[k] \tag{23}$$

$$y_F[k] = -\sum_{j=1}^{p} a_j^F y[k-j] + e_F[k] \tag{24}$$

$$y_P[k] = -\sum_{j=1}^{p} a_j^P y[k-j] + e_P[k] \tag{25}$$

being $a_j^L \in \mathbb{R}$ the model parameters for the water level variable, $a_j^F \in \mathbb{R}$ the model parameters for the water flow variable, and $a_j^P \in \mathbb{R}$ the model parameters for the precipitation variable. The model parameters for each variable are estimated by using (15), resulting in a regularized AR univariable estimated model represented by Θ_L, Θ_F and Θ_P, as follows:

$$\Theta_{Tikh}^L = \begin{bmatrix} a_1^L \\ \vdots \\ a_p^L \end{bmatrix}, \Theta_{Tikh}^F = \begin{bmatrix} a_1^F \\ \vdots \\ a_p^F \end{bmatrix}, \Theta_{Tikh}^P = \begin{bmatrix} a_1^P \\ \vdots \\ a_p^P \end{bmatrix} \tag{26}$$

The model parameters for each variable presented in (26) are also updated for each new measurement, by performing the data assimilation task.

It is worth mentioning that the parameters can also be estimated by using the least squares method as described in (10). In that case, the resulting parameters for the multivariable AR model described in (22) are defined as Θ_{LS}, and the parameters (26) for level, flow and precipitation variables are defined as Θ_{LS}^L, Θ_{LS}^F an Θ_{LS}^P respectively.

III. Results

A. Experimental setup

In order to validate the multivariable AR regularized estimation method for data assimilation, a data set of hydrological variables of Level, Flow and Precipitation are analyzed. The data set is measured at a hydrological station of the Institute of Hydrology, Meteorology and Environmental Studies (IDEAM). The IDEAM hydrological station number 11047010 is located at the Colombia country, in the Chocó Department, Municipality of Quibdo, at the Atrato river. The sample time is 12 hours, and a total amount of 1478 samples are considered.

In Table III-A is presented the geographical location of the hydrological station where the data-set is measured.

TABLE I
Location of the hydrological station

	Station Coordinates
Longitude	76° 39’ 44.13” W
Latitude	5° 41’ 52. 77” N N
Altitude	20.83 MASL

The performance evaluation of the Multivariable AR regularized estimation is compared with univariable models estimated by using the same regularized approach. In addition, the proposed approach is also compared with the least squares solution by considering a sufficient amount of data. The performance is analyzed in terms of the least squares error. Additional analysis is performed using a reduced amount of data for parameter estimation.

B. *Regularized AR Univariable estimation results*

The estimation results for the AR model parameters are computed for each of the variables: Level, Flow, and Precipitation. The regularized AR univariable solution by using Tikhonov is compared with the real data and the least squares AR estimation. The regularization parameter λ is selected independently for each data set using the GCV method. A system of order 10 is selected to exemplify the proposed approach's behavior.

In Fig. 1 is presented the selection of the regularization parameter by using GCV method for the Level variable.

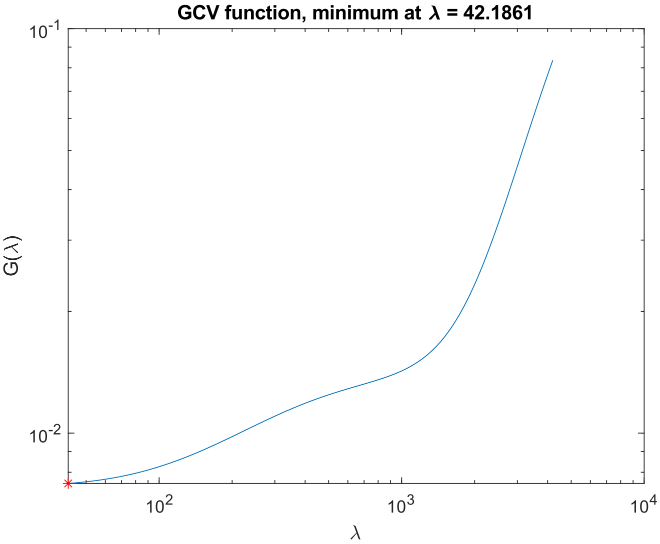

Fig. 1. Selection of the regularization parameter λ by using the GCV method for Level variable.

The selected value for λ regularization parameter is $\lambda = 42.1861$. By using this value, the vector of estimated parameters Θ^L for the Level variable by using the regularized AR estimation method can be computed. It is worth noting that the regularization parameter implies a smoothing effect in the estimated signal, where an increase in the regularization parameter can be viewed as a smoother estimated signal.

In (27) are shown the vectors for estimated parameters by using the regularized AR method Θ^L_{Tikh} and the least squares method Θ^L_{LS} for a system of order 10.

$$\Theta^L_{LS} = \begin{bmatrix} 0.8174 \\ -0.1804 \\ 0.1300 \\ -0.0007 \\ -0.007 \\ 0.1326 \\ -0.0918 \\ 0.1929 \\ -0.1506 \\ 0.1525 \end{bmatrix}, \Theta^L_{Tikh} = \begin{bmatrix} 0.6396 \\ -0.0095 \\ 0.0715 \\ 0.0298 \\ 0.0159 \\ 0.0878 \\ -0.0094 \\ 0.1083 \\ -0.0463 \\ 0.1053 \end{bmatrix} \quad (27)$$

By considering the estimated parameters of (27) for the regularized AR model Θ^L_{Tikh}, and the least squares AR model Θ^L_{LS} for a system of order 10, a comparison with the real data can be performed. In Fig. 2 is presented the comparison of the estimated signals by using the real data, Θ^L_{LS} and the Θ^L_{Tikh} is presented. An additional zoom of the first 200 samples is also shown to clarify the results.

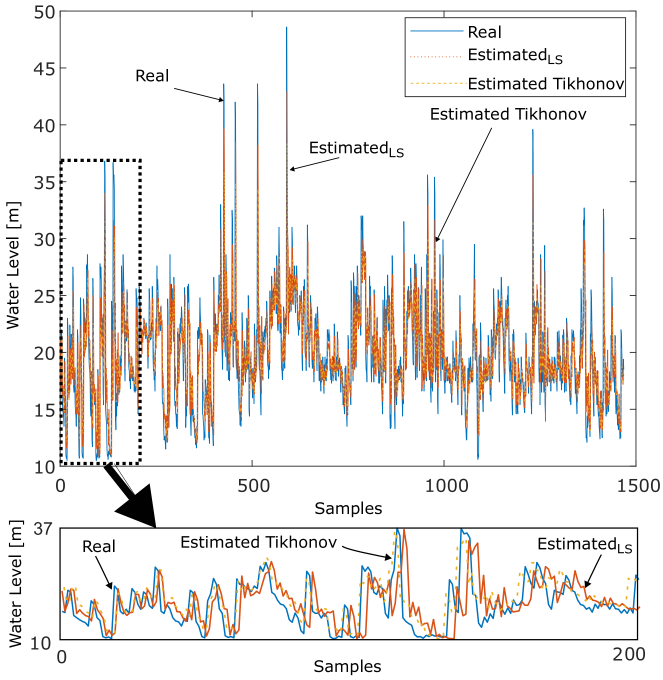

Fig. 2. Comparison of the estimated signals by using the real level data, Θ^L_{LS} and the Θ^L_{Tikh}, order 10

In Fig. 3 is presented the selection of the regularization parameter by using GCV method for the Flow variable.

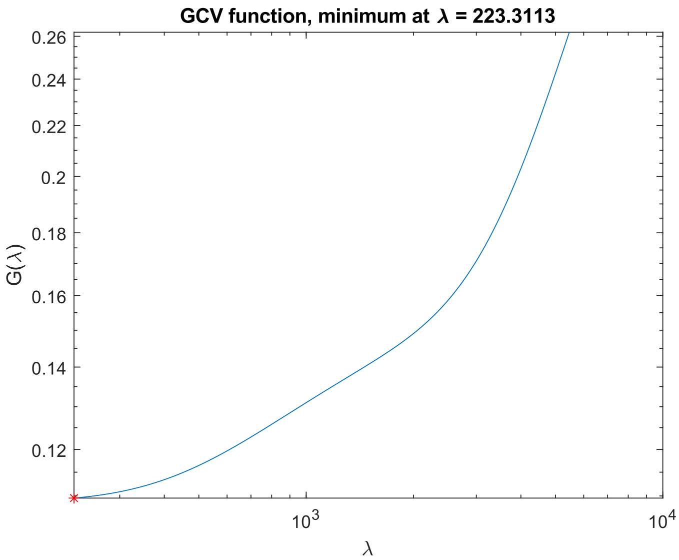

Fig. 3. Selection of the regularization parameter λ by using the GCV method for Flow variable.

The selected value for λ regularization parameter is $\lambda = 223.3113$. By using this value, the vector of estimated parameters Θ^F for Flow variable by using the regularized AR estimation method can be computed.

In (28) are shown the vectors for estimated parameters by using the regularized AR method Θ^F_{Tikh} and the least squares method Θ^F_{LS} for a system of order 10.

$$\Theta^F_{LS} = \begin{bmatrix} 0.6132 \\ 0.0008 \\ 0.0325 \\ 0.0221 \\ 0.0022 \\ 0.0811 \\ 0.0259 \\ 0.0676 \\ -0.0169 \\ 0.1388 \end{bmatrix}, \Theta^F_{Tikh} = \begin{bmatrix} 0.4613 \\ 0.0827 \\ 0.0417 \\ 0.0314 \\ 0.0218 \\ 0.0660 \\ 0.0474 \\ 0.0605 \\ 0.0264 \\ 0.1169 \end{bmatrix} \quad (28)$$

By considering the estimated parameters of (28) for the regularized AR model Θ^F_{Tikh}, and the least squares AR model Θ^F_{LS} for a system of order 10, a comparison with the real Flow data can be performed. In Fig. 4 is presented the comparison of the estimated signals by using the real data, Θ^F_{LS} and the Θ^F_{Tikh} is presented. An additional zoom of the first 200 samples is also shown.

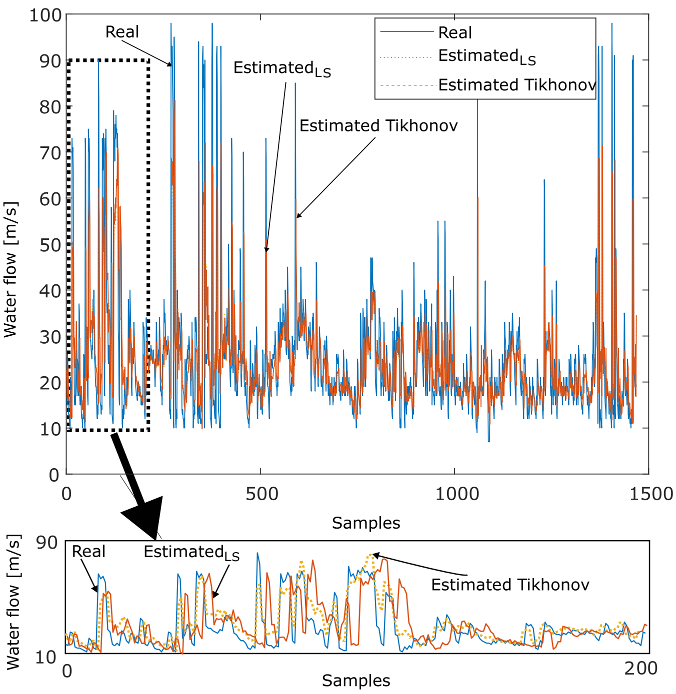

Fig. 4. Comparison of the estimated signals by using the real Flow data, Θ^F_{LS} and the Θ^F_{Tikh}, order 10

In Fig. 5 is presented the selection of the regularization parameter by using GCV method for the Precipitation variable.

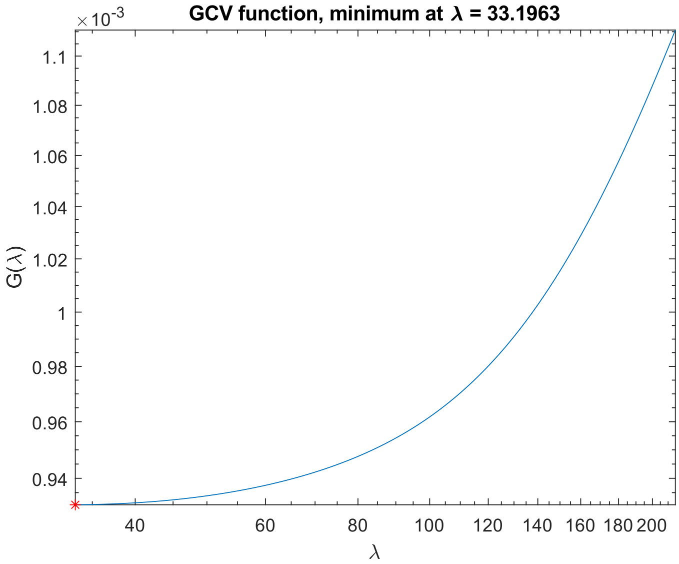

Fig. 5. Selection of the regularization parameter λ by using the GCV method for Precipitation variable

The selected value for λ regularization parameter is $\lambda = 33.1963$. By using this value, the vector of estimated parameters Θ^F for Precipitation variable by using the regularized AR estimation method can be computed.

In (29) are shown the vectors for estimated parameters by using the regularized AR method Θ^P_{Tikh} and the least squares method Θ^P_{LS} for a system of order 10.

$$\Theta^P_{LS} = \begin{bmatrix} 0.1548 \\ 0.1467 \\ 0.0332 \\ 0.1137 \\ 0.1068 \\ 0.0442 \\ 0.0940 \\ 0.0983 \\ 0.0513 \\ 0.0292 \end{bmatrix}, \Theta^P_{Tikh} = \begin{bmatrix} 0.11852 \\ 0.1138 \\ 0.0537 \\ 0.0947 \\ 0.0899 \\ 0.0589 \\ 0.0833 \\ 0.0843 \\ 0.0611 \\ 0.0461 \end{bmatrix} \quad (29)$$

By considering the estimated parameters of (29) for the regularized AR model Θ^P_{Tikh}, and the least squares AR model Θ^P_{LS} for a system of order 10, a comparison with the real Precipitation data can be performed. In Fig. 6 is presented the comparison of the estimated signals by using the real data, Θ^P_{LS} and the Θ^P_{Tikh} is presented. An additional zoom of the first 200 samples is shown in order to clarify the results.

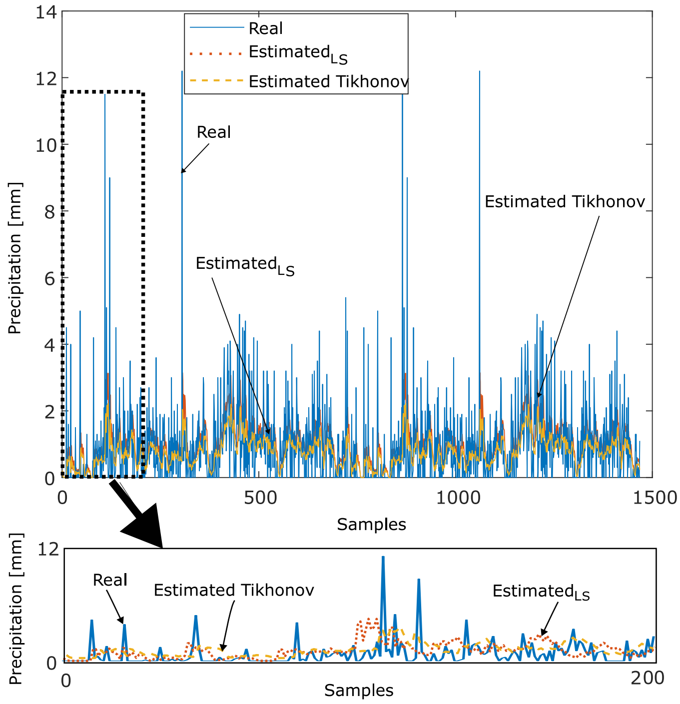

Fig. 6. Comparison of the estimated signals by using the real Precipitation data, Θ_{LS}^{P} and the Θ_{Tikh}^{P}, order 10

An additional comparison of the estimated results for the regularized AR method and the least squares AR method for a system of order 30 are presented in Fig. 7, Fig. 8, and Fig. 9. It can be seen that the obtained results are similar to the ones presented in Fig. 2, Fig. 4 and Fig. 6.

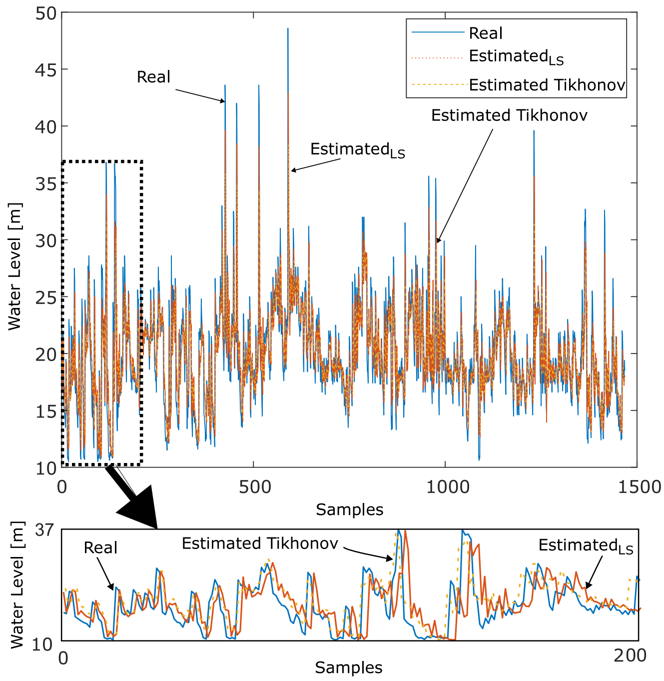

Fig. 7. Comparison of the estimated signals by using the real Level data, Θ_{LS}^{P} and the Θ_{Tikh}^{P}, order 30.

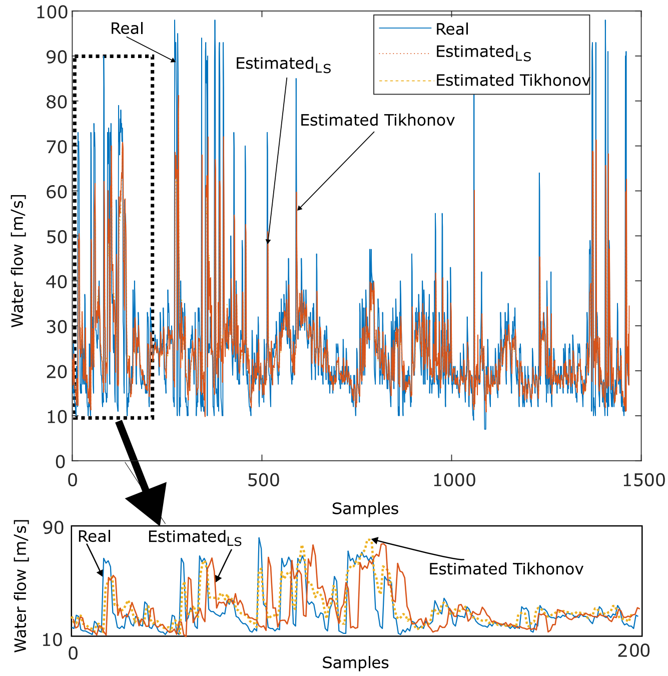

Fig. 8. Comparison of the estimated signals by using the real Flow data, Θ_{LS}^{P} and the Θ_{Tikh}^{P}, order 30.

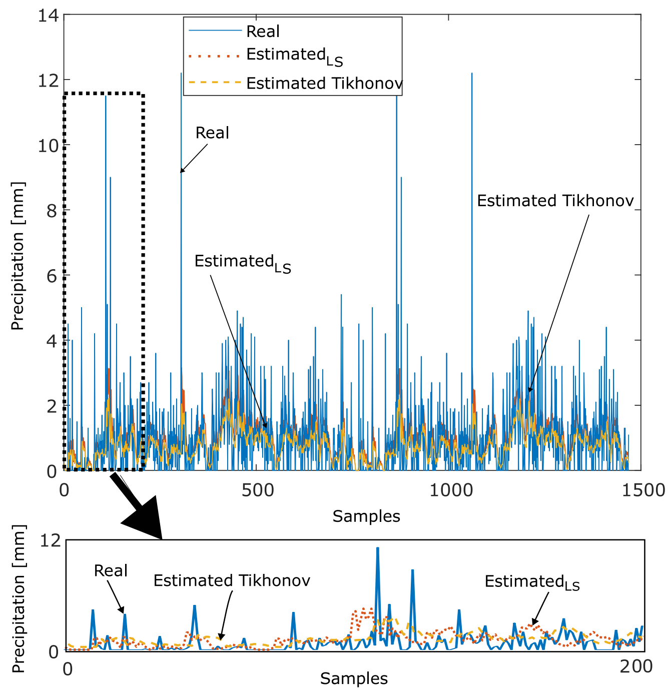

Fig. 9. Comparison of the estimated signals by using the real Precipitation data, Θ_{LS}^{P} and the Θ_{Tikh}^{P}, order 30.

It is worth mentioning that the selection of the regularization parameter λ is directly related to how much we want to penalize or adjust the flexibility of our model.

C. Regularized AR multivariable estimation results

The regularized AR multivariable solution by using Tikhonov is also compared with the real data and the least squares AR estimation. The selection of the regularization parameter λ is performed for each data-set by using the GCV method. A system of order 10 is selected in order to exemplify the behavior of proposed approach.

Three λ values are obtained by using the GCV method related to each of the variables analyzed: Level, Flow and Precipitation. In Fig. 10 is presented the selection of the regularization parameter by using GCV method for the Level variable.

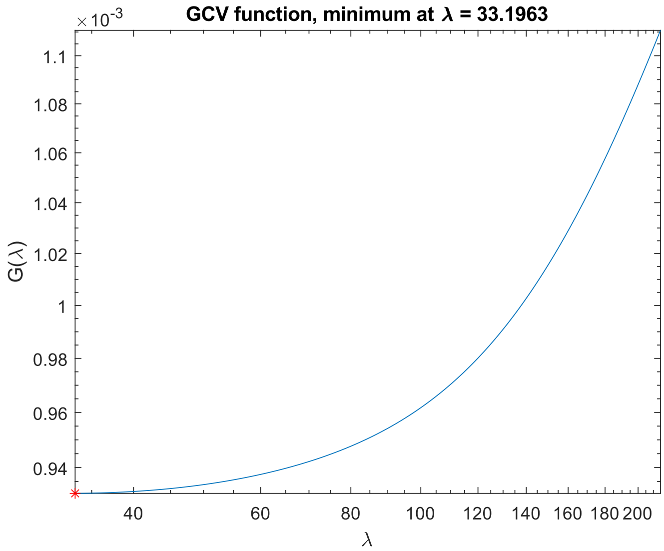

Fig. 10. Selection of the regularization parameter λ by using the GCV method for Level variable

In Fig. 11 is presented the selection of the regularization parameter by using GCV method for the Flow variable.

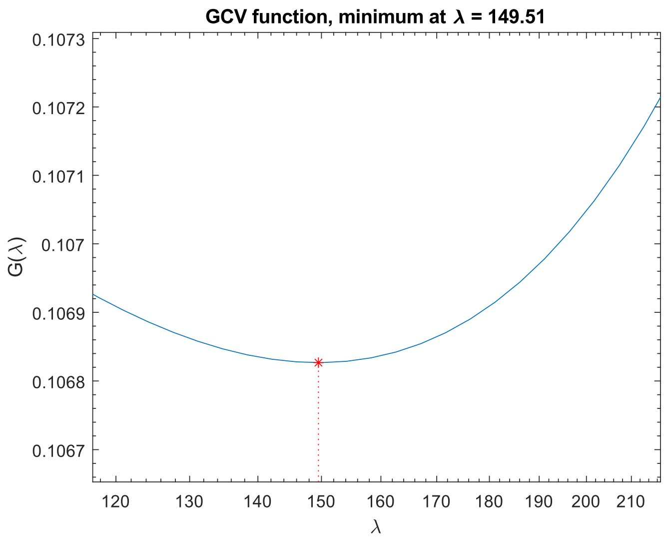

Fig. 11. Selection of the regularization parameter λ by using the GCV method for Flow variable

In Fig. 11 is presented the selection of the regularization parameter by using GCV method for the Precipitation variable.

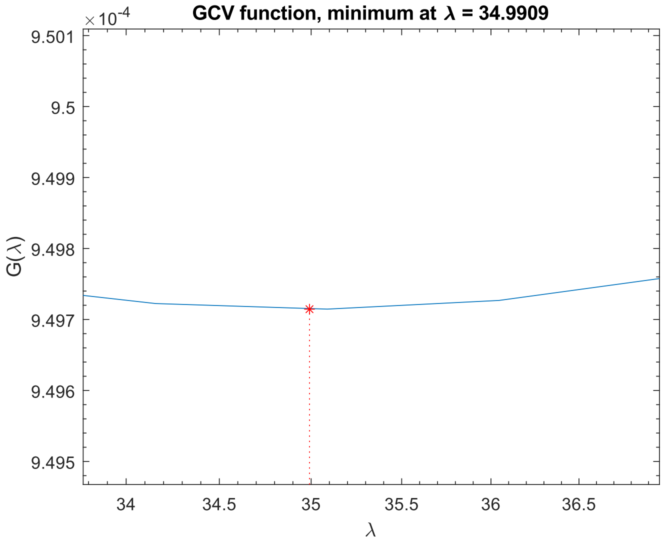

Fig. 12. Selection of the regularization parameter λ by using the GCV method for Precipitation variable

The selected values of λ for each variable are $\lambda_L = 32.3677$, $\lambda_F = 149.51$, $\lambda_P = 34.9909$. By considering these values the mean of λ_L, λ_F and λ_P is selected as the regularization parameter for the regularized AR multivariable estimated solution, as $\lambda = 72.28$.

In (30) and (31) are shown the matrices of estimated parameters by using the regularized AR method Θ_{Tikh} and the least squares method Θ_{LS} for a system of order 10.

$$\Theta_{LS} = \begin{bmatrix} 0.5898 & 0.0098 & -0.0033 \\ 0.1733 & 0.8223 & 0.0076 \\ 0.5969 & 0.1230 & 0.1225 \\ 0.0281 & -0.0012 & 0.0041 \\ -0.3180 & -0.1342 & 0.0183 \\ 0.3011 & 0.1065 & 0.1182 \\ 0.0160 & 0.01959 & -0.0044 \\ 0.2094 & 0.1111 & -0.0021 \\ 0.0956 & 0.02219 & 0.0028 \\ 0.0127 & -0.0087 & 0.0021 \\ -0.0091 & 0.02068 & -0.0073 \\ 0.2143 & 0.0307 & 0.0832 \\ 0.0168 & -0.0002 & 0.0007 \\ 0.3114 & 0.1271 & 0.0118 \\ -0.3450 & 0.1539 & 0.0809 \end{bmatrix} \tag{30}$$

$$\Theta_{Tikh} = \begin{bmatrix} 0.5883 & 0.0199 & -0.0023 \\ 0.1445 & 0.7242 & 0.0103 \\ 0.2978 & 0.0747 & 0.0664 \\ 0.0245 & -0.0069 & 0.0040 \\ -0.2332 & -0.0351 & 0.01676 \\ 0.1756 & 0.0635 & 0.0649 \\ 0.0202 & 0.0203 & -0.0041 \\ 0.1527 & 0.0854 & 0.0010 \\ 0.0748 & 0.0293 & 0.0148 \\ 0.0093 & -0.0074 & 0.0018 \\ 0.0403 & 0.0450 & -0.0040 \\ 0.1136 & 0.0299 & 0.0483 \\ 0.0198 & -0.0009 & 0.0006 \\ 0.2794 & 0.1266 & 0.0117 \\ -0.1349 & 0.0839 & 0.0464 \end{bmatrix} \tag{31}$$

By considering the estimated parameters of (30) and (31) for the regularized multivariable AR model Θ_{Tikh}, and the least squares AR model Θ_{LS} for a system of order 10, a comparison with the real Level, Flow and Precipitation data can be performed. In Fig. 13, Fig. 14 and Fig. 15 are presented the comparison of the estimated signals by using the real data, an the estimated model parameters Θ_{LS} and Θ_{Tikh}, for level, flow and precipitation respectively.

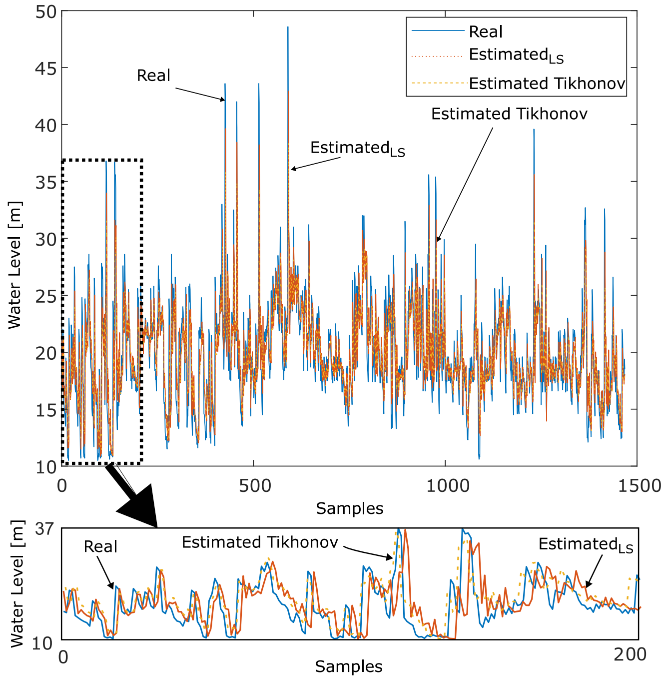

Fig. 13. Comparison of the estimated signals by using the real level data, and estimated data by using Θ_{LS} and the Θ_{Tikh} for a system of order 10

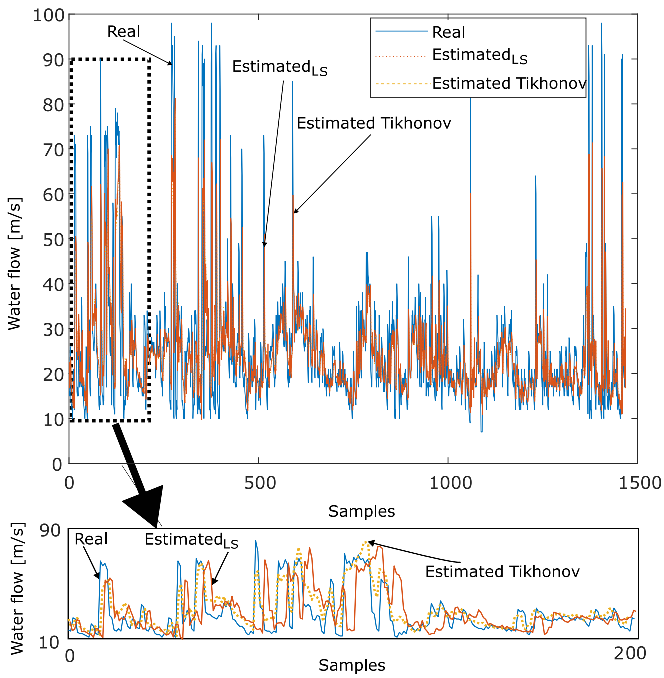

Fig. 14. Comparison of the estimated signals by using the real flow data, and estimated data by using Θ_{LS} and the Θ_{Tikh} for a system of order 10

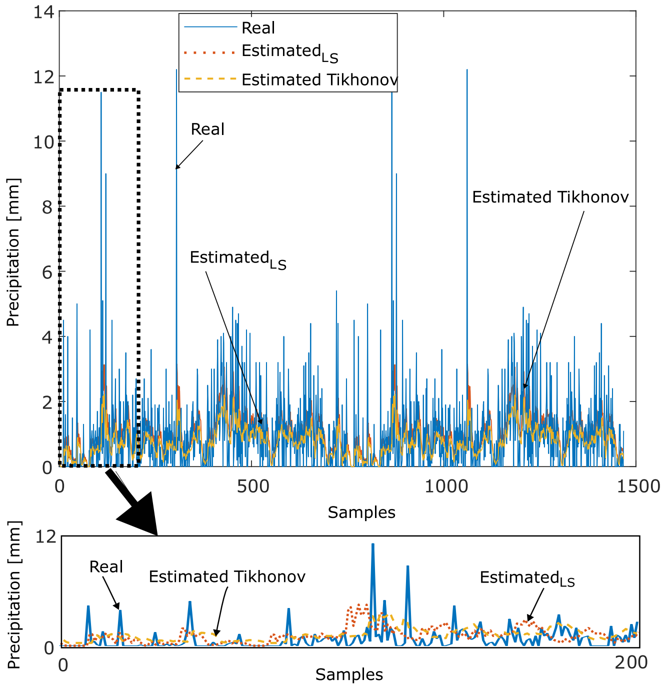

Fig. 15. Comparison of the estimated signals by using the real precipitation data, and estimated data by using Θ_{LS} and the Θ_{Tikh} for a system of order 10

A comparison of the estimated results for the univariable and multivariable AR model estimated by Tikhonov regularization and least squares is also presented. The mean squared error is used for this comparison by considering models of orders 2 to 30. Fig. 16 shows the error comparison analysis for the Level variable, estimated for the univariable AR model with the least squares method (ELS) and the Tikhonov method (ELST), and the multivariable AR model with the least squares method (ELM), and the Tikhonov method (ELMT).

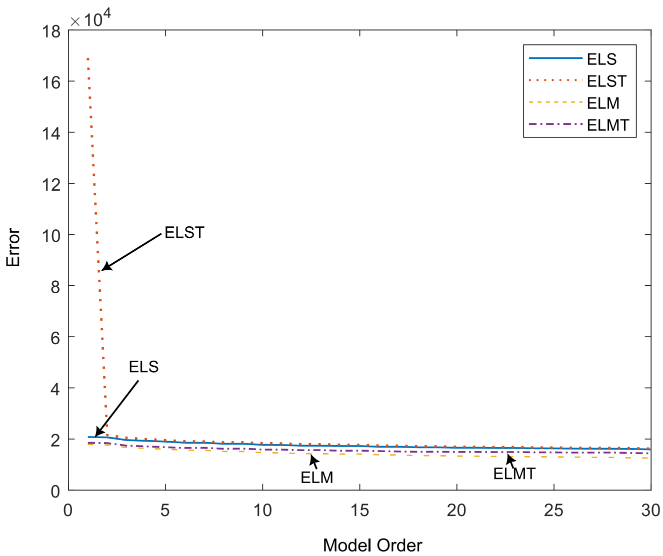

Fig. 16. Level variable estimation error comparison for the univariable AR model with the least squares method (ELS) and the Tikhonov method (ELST), and for the multivariable AR model with the least squares method (ELM), and the Tikhonov method (ELMT)

A similar comparison is presented in Fig. 17 for the flow variable.

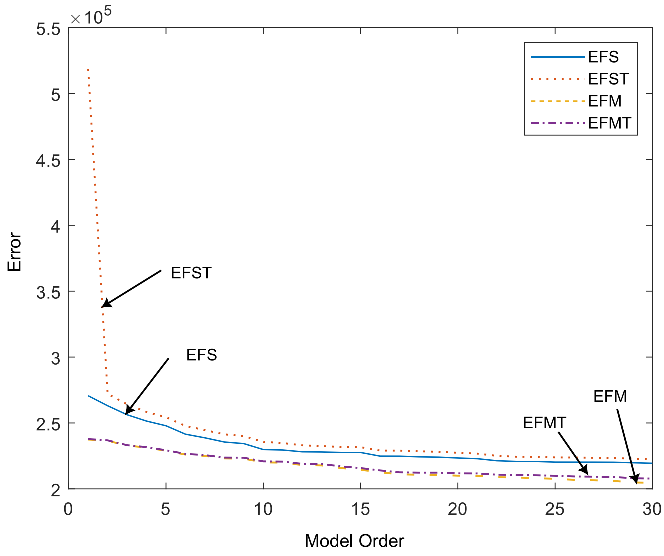

Fig. 17. Flow variable estimation error comparison for the univariable AR model with the least squares method (EFS) and the Tikhonov method (EFST), and for the multivariable AR model with the least squares method (EFM), and the Tikhonov method (EFMT)

A similar comparison is presented in Fig. 18 for the precipitation variable.

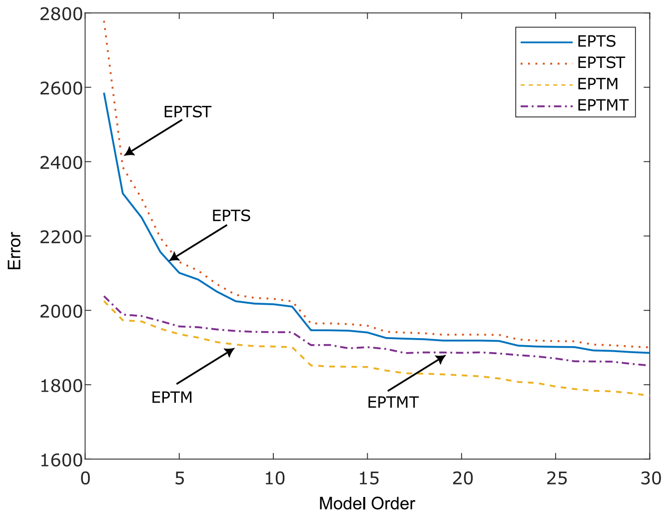

Fig. 18. Precipitation variable estimation error comparison for the univariable AR model with the least squares method (EFS) and the Tikhonov method (EFST), and for the multivariable AR model with the least squares method (EFM), and the Tikhonov method (EFMT)

From Fig. 16, Fig. 17 and Fig. 18 it can be seen that the lower error is achieved by the multivariable AR models, being the regularized AR model the least error.

D. *Estimation under a Rank-deficient scenario*

An additional evaluation is performed by using a reduced amount of data. This evaluation allows to verify the performance of the proposed approach under near rank-deficient conditions. In Fig. 19, Fig. 20 and Fig. 21 are presented the estimation results for an univariable AR system for level, flow and precipitation data respectively. The estimation is performed for 40 data samples with a system of order 30.

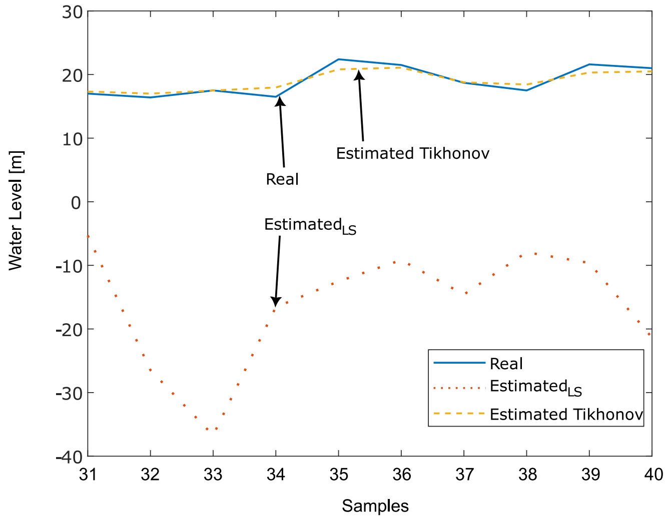

Fig. 19. Estimation results for an univariable AR system for level order 30 with a sample of 40 data.

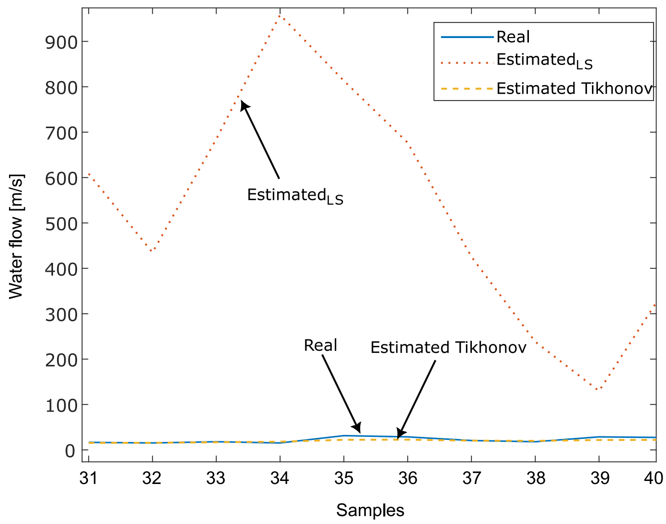

Fig. 20. Estimation results for an univariable AR system for flow order 30 with a sample of 40 data.

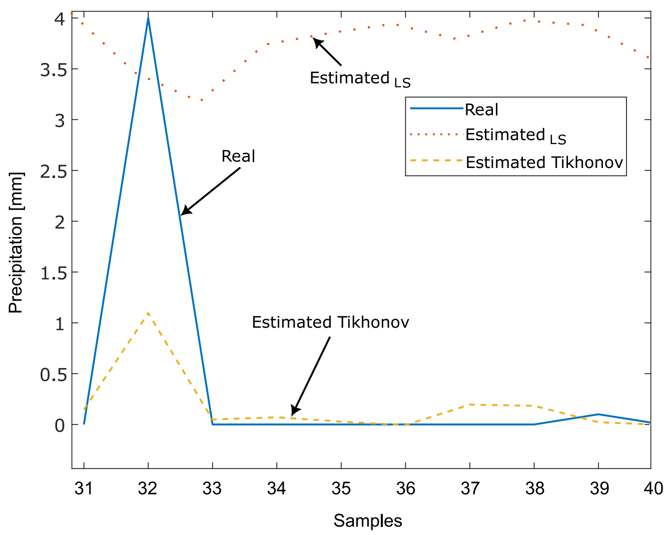

Fig. 21. Estimation results for an univariable AR system for precipitation order 30 with a sample of 40 data.

It can be seen that the regularized AR model estimated adequately the real data with lower estimation error than the least squares approach.

E. Validation of univariable AR model for a Parana river level data sample

In order to validate the proposed regularized AR model, a data sample from the province of Formosa in Argentina is considered. This station shows the river level in the Paraná basin from April 4, 2021, to April 4, 2022. The data is measured by Argentina's Meteorological Information Center (CIM) [10]. A comparison analysis is performed for the estimation of model parameters by using the Tikhonov and the least squares estimation methods. The resulting model parameters are presented in (32).

$$\Theta_{LS}^{L} = \begin{bmatrix} 0.5974 \\ 0.3094 \\ 0.1146 \\ 0.0192 \\ 0.0288 \\ 0.0246 \\ -0.0024 \\ 0.0019 \\ -0.0558 \\ -0.0408 \end{bmatrix}, \Theta_{Tikh}^{L} = \begin{bmatrix} 0.5974 \\ 0.3094 \\ 0.1146 \\ 0.0191 \\ 0.0288 \\ 0.0246 \\ -0.0024 \\ 0.0019 \\ -0.0558 \\ -0.0408 \end{bmatrix} \quad (32)$$

In Fig. 22 are presented the estimation results for the level of a model of order 10.

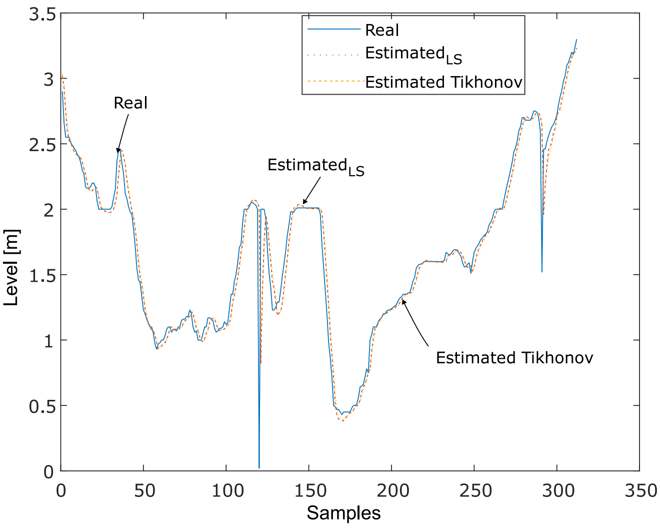

Fig. 22. Estimation results by using the least squares and the Tikhonov estimation methods by a model of order 10

Additional results are obtained by considering an estimated model of order 30. In Fig. 23 are presented the estimation results for the level of a model of order 30.

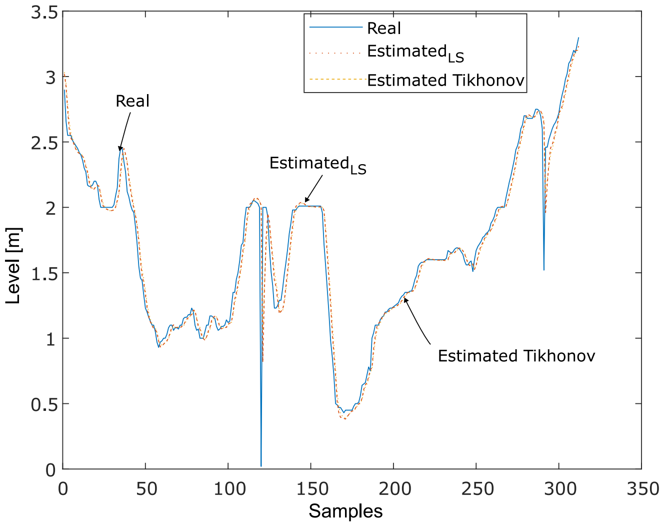

Fig. 23. Estimation results by using the least squares and the Tikhonov estimation methods by a model of order 30

In order to evaluate the performance of the proposed approach, a comparison analysis in terms of the mean squared error is computed. In Fig. 24 are shown the estimation error by using least squares (ELS) and Tikhonov regularization (ELST).

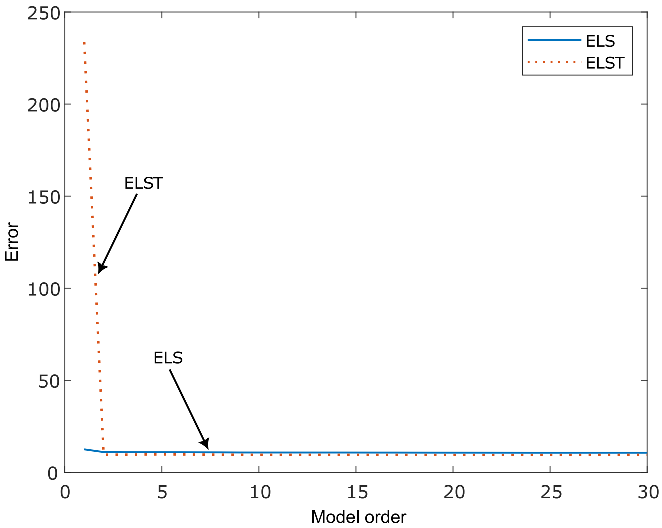

Fig. 24. Estimation error for a system of order 30 by using the estimated model by least squares (ELS) and Tikhonov regularization (ELST)

IV. Conclusions

This work evaluates a multivariable regularized AR model for hydrological variables. These results are critical in understanding the required model for predicting a real-time risk evaluation of variables. It can be seen that a large order model is required to adequately describe the data behavior, which is validated for several orders (from 1 to 30). The proposed approach adequately models Colombian river data and can be generalized for other systems. In addition, when a reduced amount of data is required, the regularized AR model still tracks the data adequately. It is worth mentioning that the procedure to update model parameters is performed according to the data assimilation techniques, typically a sequential time-stepping process, in which estimation is compared with the new measurements, and then the model is updated to reflect all the observations. In addition, the regularized multivariable AR model can describe the data behavior's correlation. Instead, the univariate model can not model this correlation. In future works, a model for data assimilation that considers an AR with exogenous inputs and a dynamical neural network for model identification will be considered. An additional evaluation by considering the terrain's topography can also be computed to design an effective Flood Early Warning System.

References

[1] D. Ortega-Alonso, "Pintando con acuarelas en el lugar más lluvioso del mundo." *Revista Afluir*, no. 5, 2021.

[2] O. Jiménez, *El Chocó, Un Paraíso Del Demonio: Nóvita, Citará Y El Baudó, Siglo Xviii*. Editorial Universidad de Antioquia, Medellín, Colombia, 2004.

[3] D. Huang, S. Wang, and Z. Liu, "A systematic review of prediction methods for emergency management," *International Journal of Disaster Risk Reduction*, vol. 62, p. 102412, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2212420921003733

[4] T. Yuwono, Ruzardi, R. Nordin, R. Mohamed, and M. Ismail, "The model of fuzzy logic in iot system as decision support system for determining flood disaster status," *Journal of Physics: Conference Series*, vol. 1529, no. 5, p. 052029, may 2020. [Online]. Available: https://dx.doi.org/10.1088/1742-6596/1529/5/052029

[5] F. Piadeh, K. Behzadian, and A. M. Alani, "A critical review of real-time modelling of flood forecasting in urban drainage systems," *Journal of Hydrology*, vol. 607, p. 127476, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0022169422000518

[6] E. Palchevsky, V. Antonov, R. Enikeev, and T. Breikin, "A system based on an artificial neural network of the second generation for decision support in especially significant situations," *Journal of Hydrology*, vol. 616, p. 128844, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0022169422014147

[7] G. C. Goodwin and K. S. Sin, *Adaptive Filtering Prediction and Control*. USA: Dover Publications, Inc., 2009.

[8] J. S. Velez-Ramirez, M. Bueno-Lopez, and E. Giraldo, "Adaptive control approach of microgrids," *IAENG International Journal of Applied Mathematics*, vol. 50, no. 4, pp. 802–810, 2020.

[9] F. Osorio-Arteaga, J. J. Marulanda-Durango, and E. Giraldo, "Multivariable adaptive control of time-varying systems," *IAENG International Journal of Computer Science*, vol. 47, no. 4, pp. 605–612, 2020.

[10] B. Hadid, E. Duviella, and S. Lecoeuche, "Data-driven modeling for river flood forecasting based on a piecewise linear arx system identification," *Journal of Process Control*, vol. 86, pp. 44–56, 2020.

[11] Y. Yang and Y. Morozov, "Minimizing multistep-ahead prediction error for piecewise arx model identification," *Computers & Chemical Engineering*, vol. 160, p. 107734, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0098135422000758

[12] M. Fan, Y. Bai, L. Wang, and L. Ding, "Combining a fully connected neural network with an ensemble kalman filter to emulate a dynamic model in data assimilation," *IEEE Access*, vol. 9, pp. 144 952–144 964, 2021.

[13] E. V. Hólm, S. T. K. Lang, M. Fisher, T. Kral, and M. Bonavita, "Distributed observations in meteorological ensemble data assimilation and forecasting," in *2018 21st International Conference on Information Fusion (FUSION)*, 2018, pp. 92–99.

[14] F. A. Ruslan, K. Haron, A. M. Samad, and R. Adnan, "Multiple input single output (miso) arx and armax model of flood prediction system: Case study pahang," in *2017 IEEE 13th International Colloquium on Signal Processing & its Applications (CSPA)*, 2017, pp. 179–184.

[15] Z. LV, J. Zuo, and D. Rodriguez, "Predicting of runoff using an optimized swat-ann: A case study," *Journal of Hydrology: Regional Studies*, vol. 29, p. 100688, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2214581820300823

[16] H.-T. Ouyang, S.-S. Shih, and C.-S. Wu, "Optimal combinations of non-sequential regressors for arx-based typhoon inundation forecast models considering multiple objectives," *Water*, vol. 9, no. 7, 2017. [Online]. Available: https://www.mdpi.com/2073-4441/9/7/519

[17] G.-Y. Chen, M. Gan, C. P. Chen, and H.-X. Li, "A regularized variable projection algorithm for separable nonlinear least-squares problems," *IEEE Transactions on Automatic Control*, vol. 64, no. 2, pp. 526–537, 2018.

[18] P. C. Hansen, *Rank-deficient and discrete ill-posed problems: numerical aspects of linear inversion*. SIAM, 1998.

[19] A. Alqahtani, S. Gazzola, L. Reichel, and G. Rodriguez, "On the block lanczos and block golub–kahan reduction methods applied to discrete ill-posed problems," *Numerical Linear Algebra with Applications*, vol. 28, no. 5, p. e2376, 2021.

[20] J. I. Padilla-Buriticá, E. Giraldo, and G. Castellanos-Domínguez, "Eeg source localization based on multivariate autoregressive models using kalman filtering," in *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2011, pp. 7151–7154.

[21] L. Yin and H. Gao, "Optimal estimation for armax processes with noisy output," in *2020 Chinese Automation Congress (CAC)*, 2020, pp. 5048–5051.

[22] C. Fenu, L. Reichel, G. Rodriguez, and H. Sadok, "Gcv for tikhonov regularization by partial svd," *BIT Numerical Mathematics*, vol. 57, no. 4, pp. 1019–1039, 2017.

[23] G. Pillonetto, T. Chen, A. Chiuso, G. De Nicolao, and L. Ljung, *Regularization of Linear Regression Models*. Cham: Springer International Publishing, 2022, pp. 33–93. [Online]. Available: https://doi.org/10.1007/978-3-030-95860-2

[24] S. Jing, "Identification of the arx model with random impulse noise based on forgetting factor multi-error information entropy," *Circuits, Systems, and Signal Processing*, vol. 41, no. 2, pp. 915–932, 2022.

[25] P. C. Hansen, J. G. Nagy, and D. P. O'leary, *Deblurring images: matrices, spectra, and filtering*. SIAM, 2006.