ARX Models as a Useful Tool to Generate Design Hydrograph with Rainfall

Dra. **Maritza ARGANIS[1,*],** M.Eng. **Margarita PRECIADO[2]**

[1] National Autonomous University of Mexico, MArganisJiingen.unam.mx

[2] Mexican Institute of Water Technology, preciado@tlaloc.imta.mx

* Corresponding author's e-mail address: MArganisJ@iingen.unam.mx

Abstract: *ARX-type parametric autoregressive models were used to identify the best fit to precipitation effective data and direct statistical runoff corresponding to a 100 years return period, under assumption that a runoff rain process can be treated as a linear system; this model was validated with data from another rain runoff event corresponding to a 10,000 years return period; the best model obtained was the ARX 1 1 1 type. It consists of two correspondence rules, one hydrograph for rise branch and another for recession, since in this way the highest model adjustment percentages (76%) were achieved with both the measured data and those of the validation.*

Keywords: *Return period, Rainfall-Runoff, ARX models, Canseco Dam*

1. Introduction

The problem to estimating direct runoff hydrograph generated by an effective precipitation event associated with a return period has a lot importance in hydraulic and hydrological studies for hydraulic design structures [1]. Recently, many studies are focused to estimating runoff for forecasting purposes and considering real-time systems; in these cases, the problem of non-linearity is argued [2, 3]. Selecting variables, one of the most influence a rain-runoff process has led to the implementation of sensitivity analysis procedures [4]. In [5] three types of rainfall runoff models are identified: physically based ones that take into account rainfall, runoff, soil moisture, soil type and use, and basin physiography; somewhat more simplified conceptual models that group the concepts of infiltration and retention losses in shrubs and theoretical or data-driven models that only find the correspondence rule between the rain process and runoff. In this study we will highlight models of this nature.

Rainfall runoff process can be analyzed as a system with an input signal to a system (excess precipitation) and an output from the system (direct runoff hydrograph). Runoff commonly occurs with a certain time delay with respect to the occurrence of precipitation and its behavior comes to depend both on the rain in the time analyzed and on the rain and the runoff itself in previous times. Under this conception, an analogy can be made with a parametric identification of systems and applied to the case of statistical rains, that is, data from a mean hyetogram of the basin and direct runoff that it would produce, for a given return period (previously obtained with help for example, a distributed model physically based or conceptual type to which a theoretical data-drive model is applied to map rainfall in runoff and this model in turn applied to estimation of direct runoff hydrographs for storms. corresponding to another return period for the same analyzed basin.

This paper describes the Matlab software toolbox tools use [6, 7], to obtain with an optimization process, the best fit ARX model for the transformation of effective rain into direct runoff. Input data correspond to the hyetogram of mean precipitation corresponding to a 10 years return period and the inflow to Canseco reservoir dam located in Veracruz, Mexico [8, 9]; which is identified as a closed basin. Validation data are from another statistical event corresponding to an effective rainfall hyetogram corresponding to a 10,000 years return period. In a first stage, 50 percent of the data were considered to identify the best AX model and 50% data from same event for validation; but in that case, ARX model gave high adjustment errors, barely a 42% fit, so it was decided to separate problem considering a function defined with two correspondence rules, one for ascent branch of the hydrograph up to time peak and another correspondence rule for descent leg to base time. With this consideration applied to the event with a 10 years return period, an ARX model was obtained that gave a 76% correspondence in adjustment in ascending branch with used data to identify (Tr = 10 years) and a similar value when applying data model for validation (Tr = 10,000 years), while in the descending branch the percentage of adjustment was 98%.

2. Methodology

2.1 Study site

In this paper, we used as a starting mean hyetogram point data for basin and runoff, corresponding to 100 and 10,000 years return periods (Table 1), they were obtained with a statistical procedure and a runoff rain model for the Canseco Dam. located in Laguna de Catemaco, that is, it is a closed basin (see Fig. 1), located in Los Tuxtlas Volcanic Massif, in the southeast of the state of Veracruz, Mexico. It is located at the coordinates 95 ° 04 '6.98"W, 18 ° 24' 6.71"N, limited by the extreme geographic coordinates 18º 21 'and 18º 27' of northern latitude and 95º 01 'and 95º 07' of western longitude, 332 m above sea level. It is part of the Rio Papaloapan basin [8, 9].

Table 1: Mean basin hyetogram (effective rainfall hpe) and direct runoff (Qd) for return period Tr = 100 and 10,000 years

	Tr=100 years		Tr=10000 years	
Tr	**hpe**	**Qd**	**hpe**	**Qd**
h	**(mm)**	**(m^3/s)**	**(mm)**	**(m^3/s)**
1	0.55	0	0.9	0
2	0.86	1.76	1.4	2.88
3	1.19	9.57	1.94	15.61
4	1.53	23.11	2.5	37.68
5	1.98	39.36	3.21	64.16
6	2.99	57.97	4.86	94.47
7	3.5	79.86	5.68	130.12
8	3.96	106.89	6.43	174.02
9	5.07	136.89	8.22	222.64
10	7.15	168.64	11.6	274.05
11	14.73	208.49	23.93	338.64
12	156.85	276.04	255.07	448.3
13	37.04	639.11	60.29	1039.15
14	15.24	1641.23	24.81	2669.33
15	9.27	2585.45	15.1	4204.42
16	8.37	2828.08	13.63	4598.63
17	6.13	2554	9.99	4153.31
18	3.96	2062.49	6.44	3354.46
19	3.8	1634.43	6.19	2658.55
20	2.64	1287.27	4.3	2094.06
21	1.48	1011	2.42	1644.79
22	0.34	791.91	0.55	1288.45
23	0	614.5	0	999.86
h	**(mm)**	**(m^3/s)**	**(mm)**	**(m^3/s)**
24	0	467.12	0	760.09
25	0	345.48	0	562.17
26	0	250.13	0	407.02

	Tr=100 years		Tr=10000 years	
Tr	**hpe**	**Qd**	**hpe**	**Qd**
27	0	179.13	0	291.48
28	0	127.95	0	208.2
29	0	91.39	0	148.72
30	0	65.28	0	106.23
31	0	46.63	0	75.88

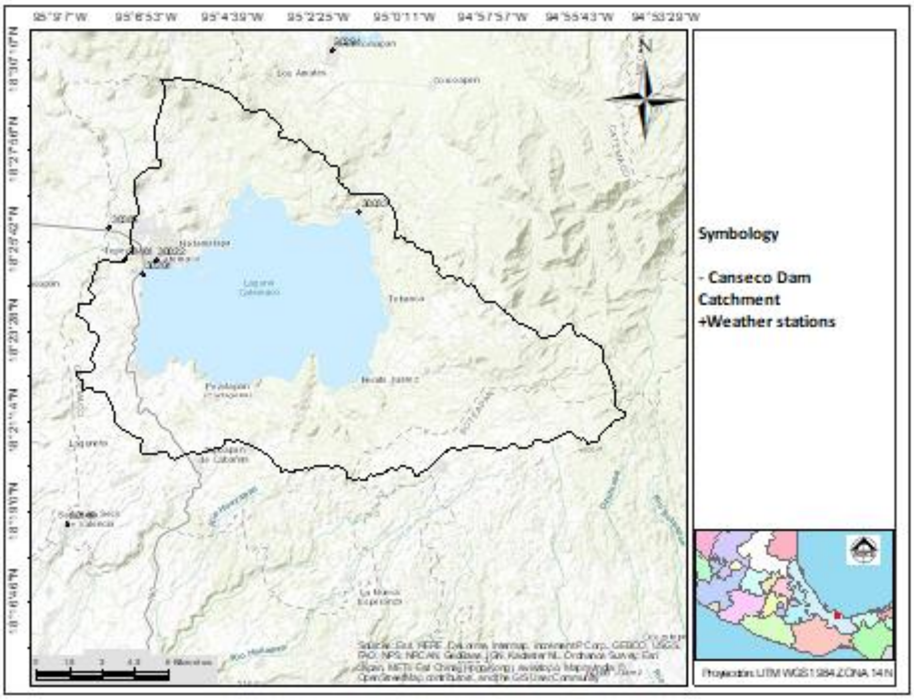

Fig. 1. Canseco Dam basin and climatologic weather stations. Source [8]

2.1 Parametric models for linear systems

For simplicity, parameter models general form for linear systems identification is given by equation 1 [6], which is known as the prediction error model (PEM)

$$A(q^{-1})y(t) = \frac{B(q^{-1})}{F(q^{-1})}u(t) + \frac{C(q^{-1})}{D(q^{-1})}e(t) \tag{1}$$

Where u (t), y (t) and e (t) are input, output and system noise respectively; A, B, D and F are polynomials as a shift operator function (q-1).
By choosing a structure from previous general model, simplified models are obtained, of which following ARX (exogenous autoregressive model), ARMAX (Autoregressive Moving Average Exogenous), OE (Output Error) and Box-Jenkins structures are for practical use:

2.1.1 ARX Model

$$A(q^{-1})y(t) = B(q^{-1})u(t) + e(t) \tag{2}$$

2.1.2 ARMAX Model

$$A(q^{-1})y(t) = B(q^{-1})u(t) + C(q^{-1})e(t) \tag{3}$$

2.1.3 OE Structure

$$y(t) = \frac{B(q^{-1})}{F(q^{-1})}u(t) + e(t) \tag{4}$$

2.1.4 Box-Jenkins Structure

$$y(t) = \frac{B(q^{-1})}{F(q^{-1})}u(t) + \frac{C(q^{-1})}{D(q^{-1})}e(t) \qquad (5)$$

In this paper, only ARX models use is highlighted.

2.2 Parametric models identification in Matlab software Functions

Within Matlab editor, you can call functions that allow linear models generation; this is a general way to obtain vector model is:

$$th = \text{model}([output \quad imput], ths) \qquad (6)$$

Where: model: it will be ARX if you are looking for ARX model shape, ARMAX, if you want to identify an ARMAX model; column vector output with system output data, column vector input with system input data and *ths* row vector containing parameters number of the model being tested and in its the most general form it and has the elements:

$$ths = \begin{bmatrix} na & nb & nc & nd & nf & nk \end{bmatrix} \qquad (7)$$

Where na, nb, nc, nd, nf are polynomials coefficients number of A, B, C, D and F of ec 1 or chosen structure (2 to 5) and nk is input and output delay times number.

2.3 Optimal structure selection

To select the model order to choose, tests can be made with different coefficients and delays in *ths* vector to make a comparison with a criterion help, for example, mean square error or by checking model parameters number to choose the best. Matlab software basin with functions designed to perform a several models automated test, obtaining their loss functions (which corresponds to the case in which mean square error between observed data and calculated with model gives minimum value) and between them indicates one that provides the smallest adjustment error; For this, it must consider data to make identification or obtain from model and data to make model validation; necessary instructions for this are highlighted in Fig. 2

```
nn=struc([1:10],[1:10],[1:10]);

v=arxstruc(datos_ident, datos_val,nn);



nn=selstruc(v)
```

Fig. 2. Instructions for obtaining optimal structure for ARX models.

In Fig. 2 struc is a matrix that contains structures result from combinations na, nb and nk in case of ARX models, arxstruc returns loss functions given vector that contains t input and output data to identify, vector that contains data for model validation and matrix struc and instruction selstruc returns the model that has the lowest loss function value.

Once the optimal model is identified, vector th of equation 6 is formed to generate it.

For model validation, instruction comparison is used (vector to identify or vector to validate, th) to presents the figure of to fit between data used to estimate model and data calculated by model or fit between data for validation and data that model calculates for those validation points.

We can also obtain residuals or errors that model commits graphically with instruction resid (vector to identify or to validate, th).

The following section highlights an example of the application of these procedures.

3. Application and results

3.1 Test 1 with ARX model and 50% storm 1 data for identification and 50% for validation

Statistical storm data were used for the return period Tr = 100 years and data first half was used for identification vector and next half for validation vector, when using instructions to optimize it was obtained that the best model is:

$$Q(t) = 1.654Q(t-1) + 1.741Pe(t-1) \tag{8}$$

When we applied validation data model to compare instruction, a 42% adjustment percentage was observed, but when we comparing against data entered for identification, an 88% adjustment percentage was observed. (Fig. 3).

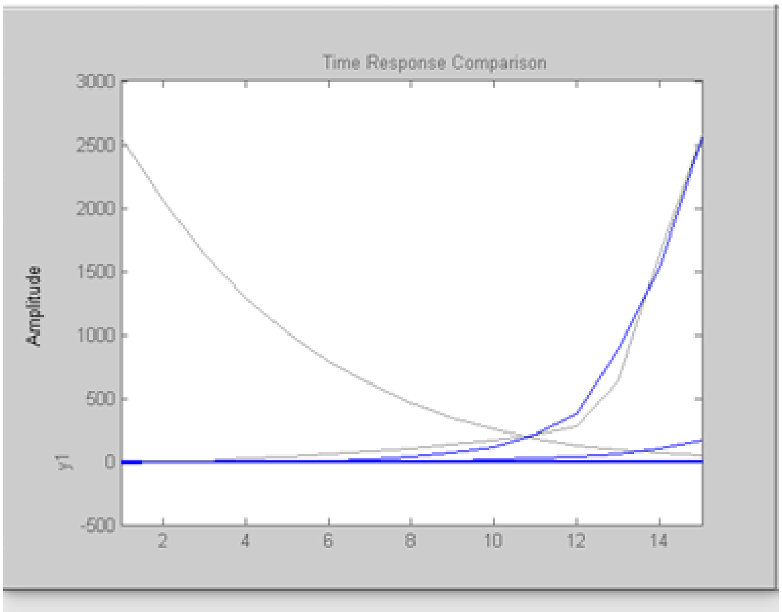

Fig. 3. Data Comparison used to identify and validate vs ARX model calculations

Given that direct runoff hydrograph upward and downward branches pass through an extreme point, identifying two behaviors; it problem can be analyzed with a function defined by two correspondence rules.

3.2 Test 2 with ARX model

Data in table 1 were divided into two groups considering storm 1 hydrograph of ascending branch (Tr = 100 years) as data to identify and storm 2 ascending branch (Tr = 10000 years as data to validate first model correspondence rule to estimate runoff; and the other hand, hydrograph of storm 1 descending branch (Tr = 100 years) was considered to identify and storm 2hydrograph of descending branch (Tr = 10000 years) to identify validate second model correspondence rule to estimate runoff.

With these considerations and when did optimization process for ascending and descending branch, the following function was obtained with two correspondence rules:

$$Q(t) = \begin{cases} 1.258Qd(t-1) + 3.041Pe(t-1); 0 \le t \le tp \\ 0.7508Q(t-1) + 21.17Pe(t-1); tp < t \le tb \end{cases} \tag{9}$$

Ascending and descending branches comparison with data for identification and with data for validation we can observed in Figs. 4 to 7.

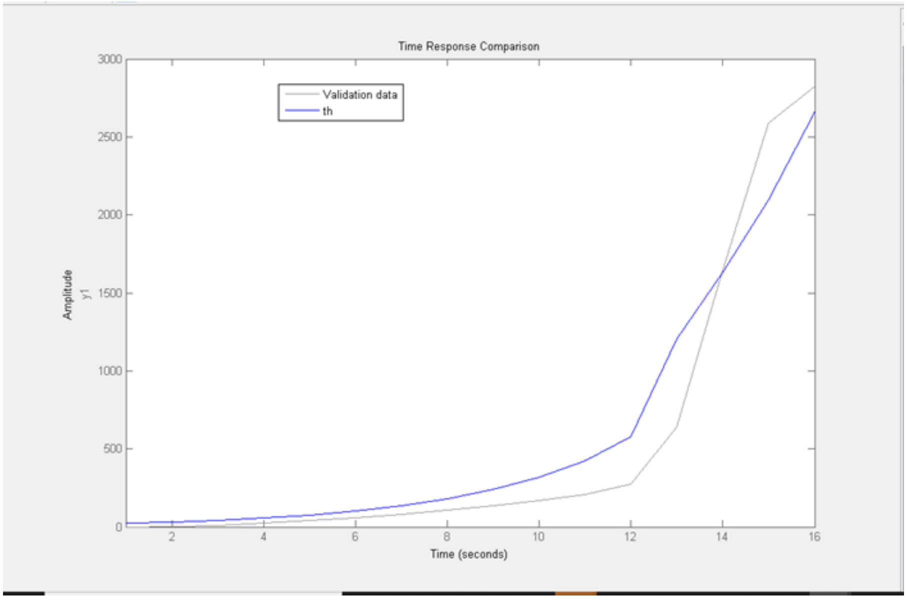

Fig. 4. Ascending branch comparison with identification data

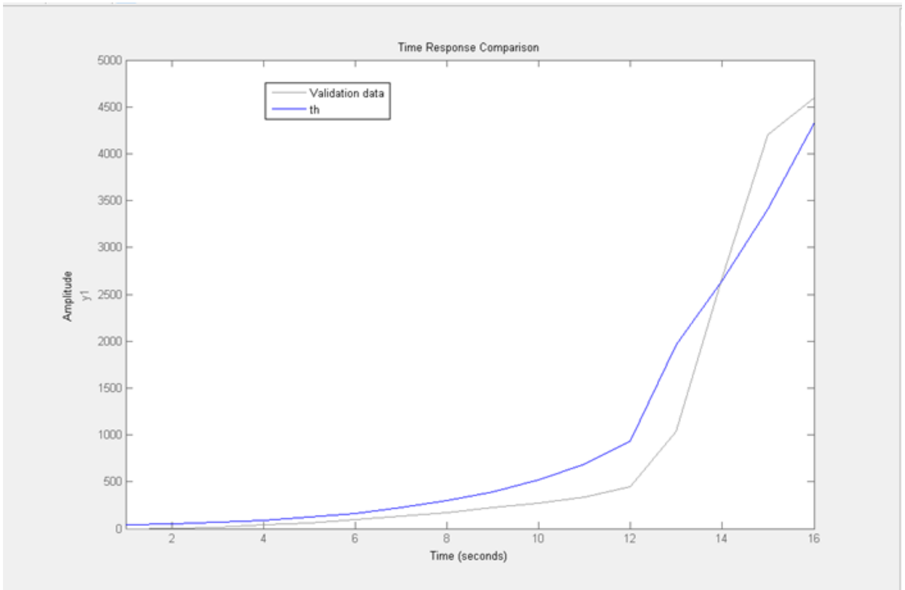

Fig. 5. Ascending branch validation with validation data

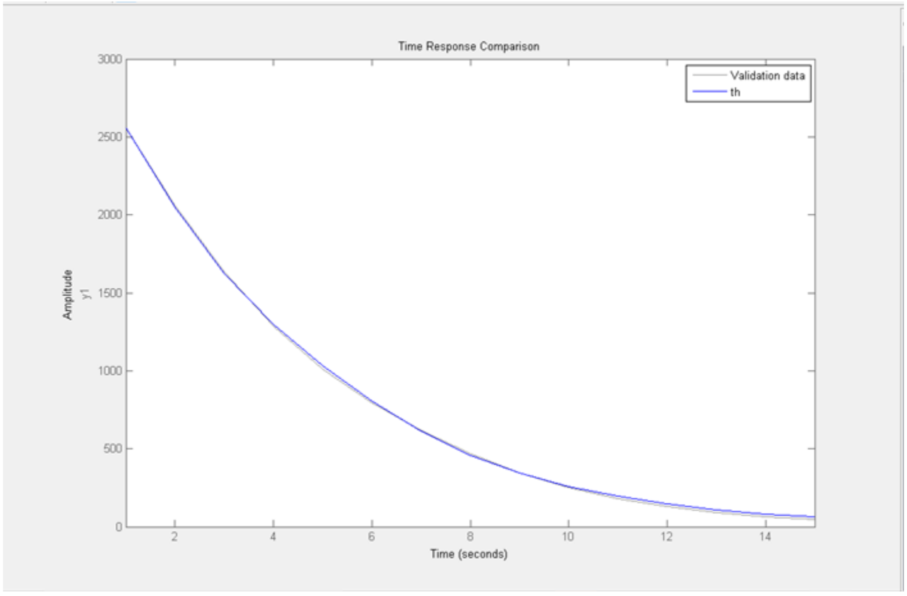

Fig. 6. Descent branch Comparison with identification data

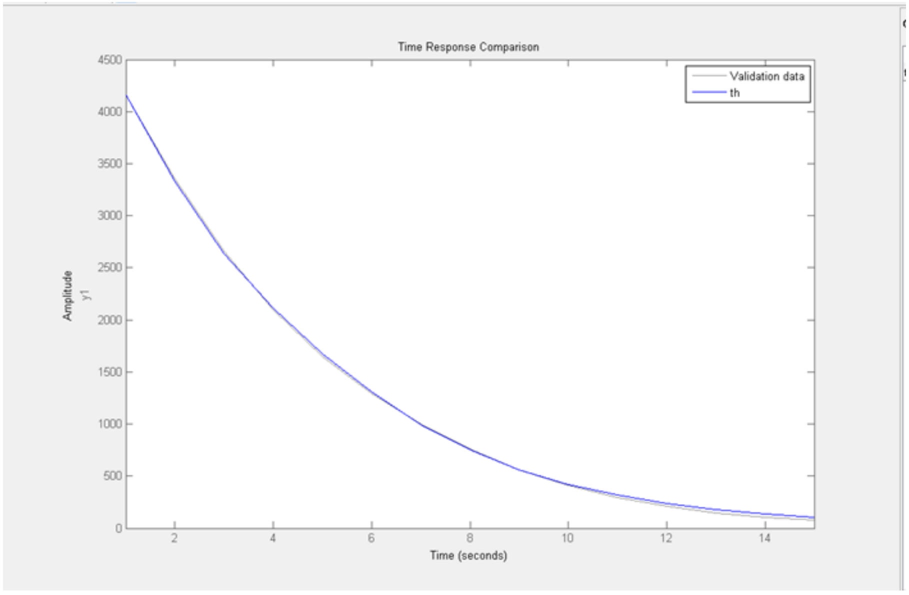

Fig. 7. Descent Branch Comparison with Validation Data

Figures 4 to 7 show a model better fit for descending hydrograph branches.
Residuals of each correspondence rule with validation data appear in Figs. 8 and 9.

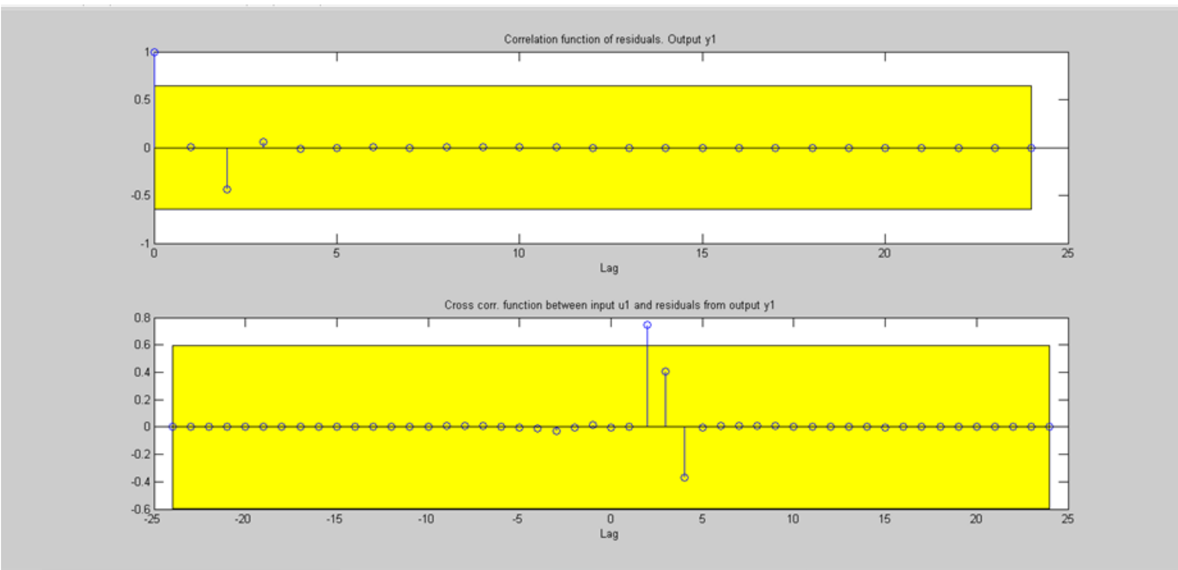

Fig. 8. Validation data residuals and those calculated by ascent branch model

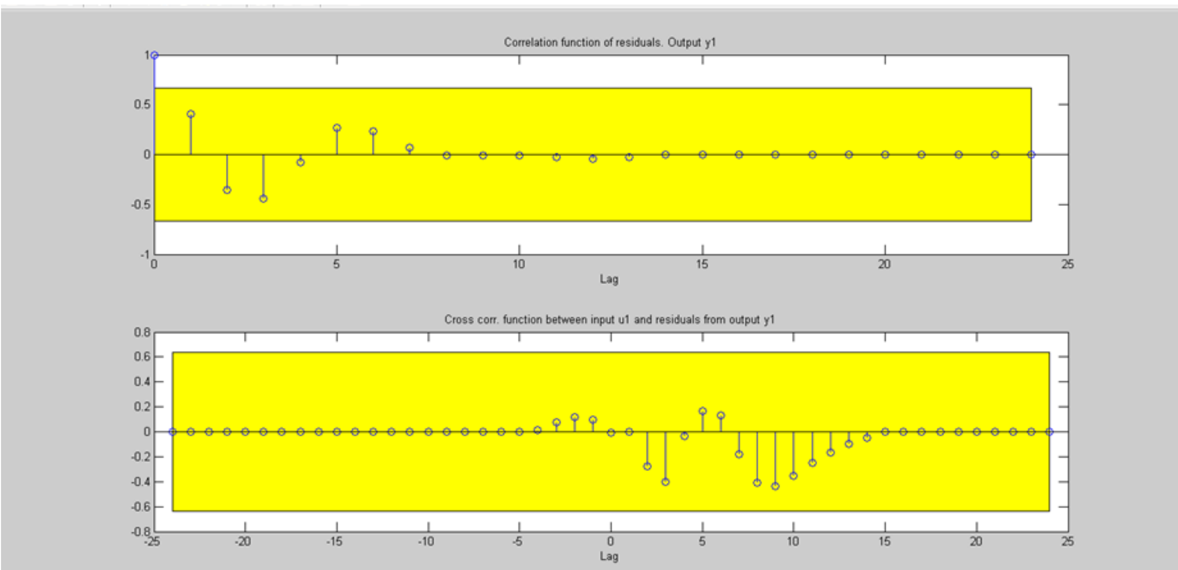

Fig. 9. Validation data residuals and those calculated by ascending branch model

From Figs. 8 and 9 above, it is observed that in ascent branch model validation presents less difficulties in reproducing data, while for descending, greater differences are observed between is estimated by model and validation data.
Hydrograph calculated complete drawing for Tr = 100 years using original input data (rainfall and runoff in previous step) is in Fig. 10 and calculated hydrograph for Tr = 100 years using runoff data that is calculating model appears in Fig. 11.

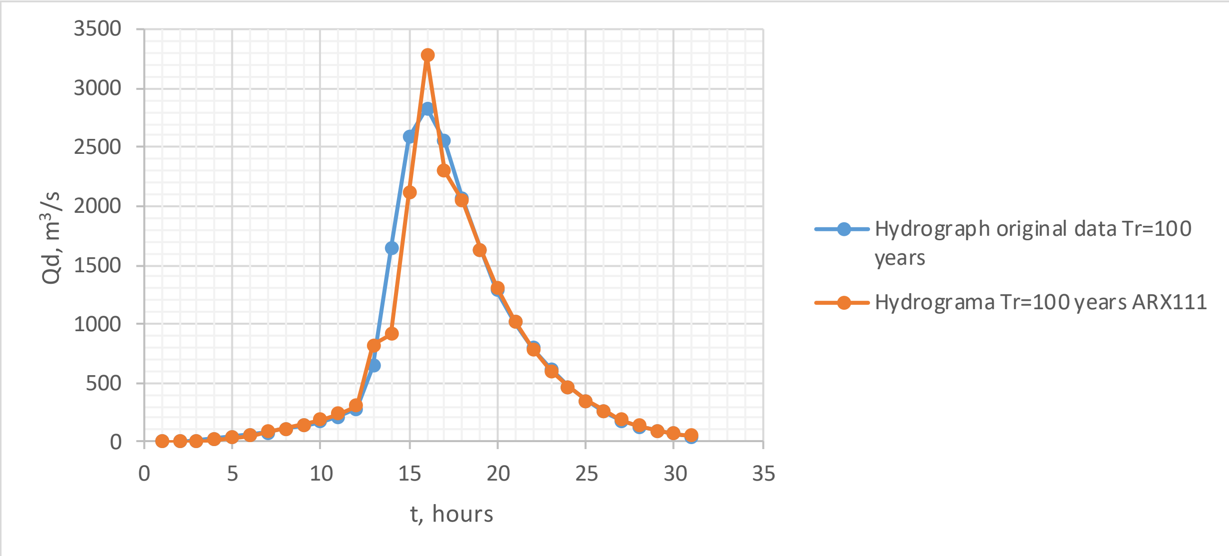

Fig. 10. Identification hydrograph comparison and the calculated hydrograph model with two correspondence rules and identification runoff data. Event Tr = 100 years

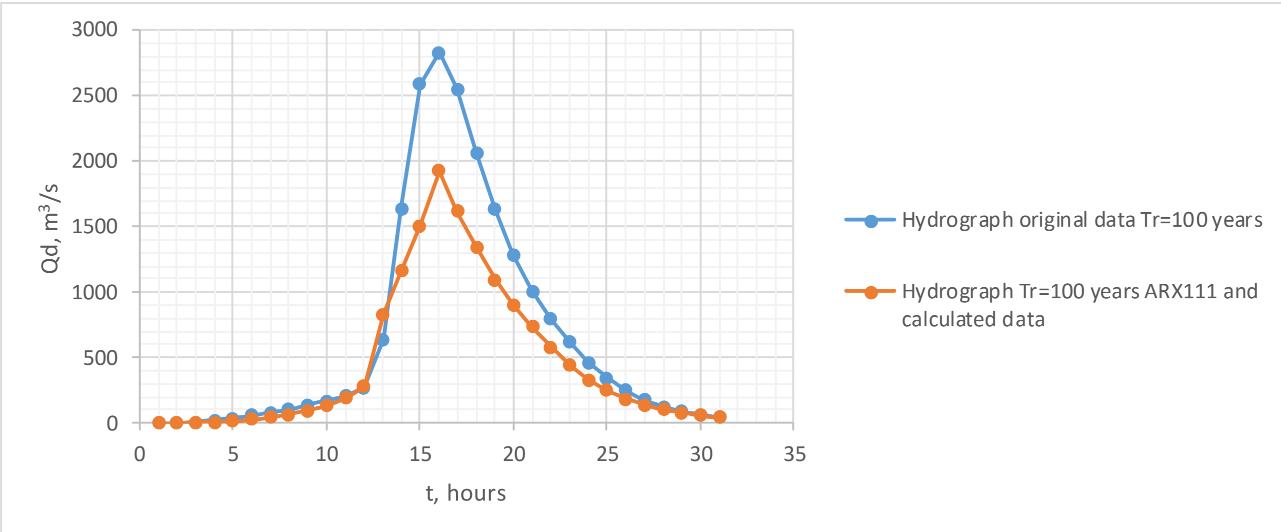

Fig. 11. Identification hydrograph Comparison and hydrograph model calculated with two correspondence rules and runoff data calculated with the model. Event Tr = 100 years

Estimated hydrograph with model for Tr = 10000 years using original rainfall and runoff data appears in Fig. 12 and the hydrograph calculated for Tr = 10000 years with given rainfall and runoff generated by the model appears in Fig. 13.

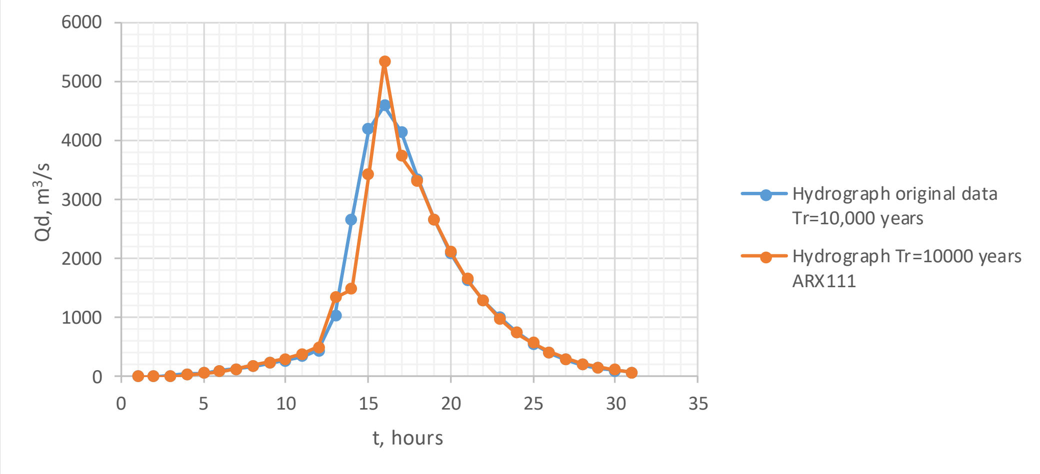

Fig. 12. Validation hydrograph comparison and calculated model hydrograph with two correspondence rules and validation runoff data. Event Tr = 10000 years

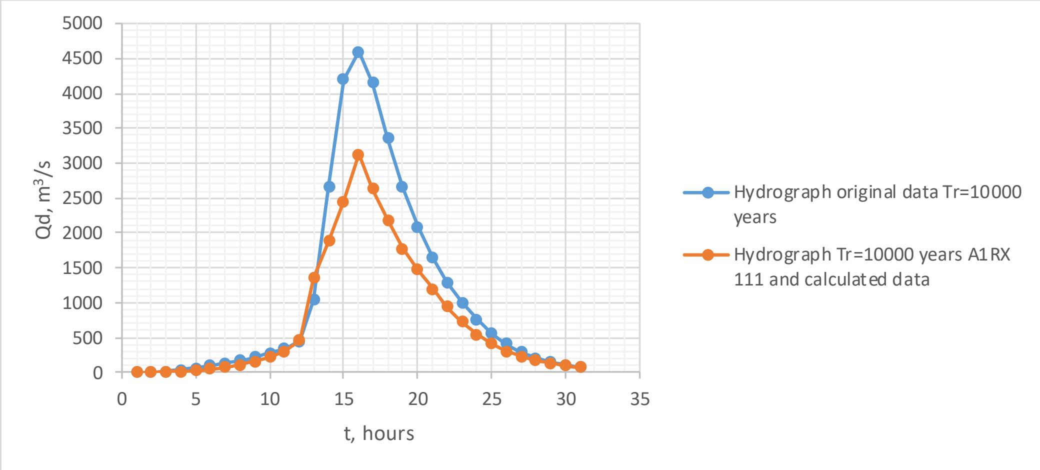

Fig. 13. Validation hydrograph comparison and calculated hydrograph model with two correspondence rules and validation runoff data. Event Tr = 10000 years

From Figures 10 to 13 an underestimation in hydrograph is observed, peak flow and therefore in runoff volume in two calculated storm cases.

Table 2 indicates peak flow and volumes obtained in each cases from Figures 10 to 13 and difference percentages found in said variables.

Table 2: Differences percentages in peak flow and hydrographs volume with original data and model calculated with two correspondence rules of equation 8

Tr	**Original Data**	**ARX**	**$Qp_{calcdcalc}$**	**% difference**	**$Qp_{calcdcalc}$**
	Qp	**$Qp_{calc\ dorig}$**		**Qpcalc dorig**	
years	**m^3/s**	**m^3/s**	**m^3/s**		
100	2828.08	3280.69	1922.70	16	32
10000	4598.63	5335.08	3126.87	16	32
	Original Data	**ARX**		**% difference**	
Tr	**V**	**V_{Qdorig}**	**V_{Qdcalc}**	**V_{Qdorig}**	**V_{Qdcalc}**
years	**hm^3**	**hm^3**	**hm^3**		

Tr	Original Data	ARX	$Qp_{calcdcalc}$	% difference	$Qp_{calcdcalc}$
	Qp	$Qp_{calc\ dorig}$		Qpcalc dorig	
100	73.11	70.55	51.10	3	30
10000	118.91	114.75	83.12	3	30

4. Conclusions

ARX Linear parametric models type was obtained to estimate direct runoff hydrograph from design precipitation data and a previously known runoff value (in this case, they were hourly rainfall and runoff data).
ARX optimal model when considering 50% of data for identification and 50% of data for validation and the storm itself presented difficulty in fully hydrograph reproducing, so it was decided to assume hydrograph shape with function behavior a definiteness by two correspondence rules, one for its ascending branch and other for its descendent branch.
Optimal models determined for each correspondence rule were of the same order na = 1, nb = 1 and nx = 1, and they were able to ascending branches reproduce with adjustments of 76% and 98%, respectively, when original data using, although in general, there is an underestimation in calculated hydrographs peak flow.
When hydrographs are calculated from known precipitation data and calculated runoff data, it is observed that model manages to reproduce ascending branch with a better fit than descending branch.
Model shape suggests that dependence that naturally occurs in basin rain runoff process, since a runoff time dependence is observed, on rainfall that occurred in a previous instant of time but also on existing runoff at instant previous.

References

[1] Aparicio Mijares, Francisco J. *Fundamentals of Surface Hydrology / Fundamentos de Hidrología de Superficie*. Ciudad de México, Limusa Publishing House, 2011.
[2] Hadid, Baya, Eric Duviella, and Stéphane Lecoeuche. "Data-driven modeling for river flood forecasting based on a piecewise linear ARX system identification." *Journal of Process Control* 1, no. 8 (February 2020): 44-56. https://doi.org/10.1016/j.jprocont.2019.12.007.
[3] Gautam, D. K., and K. P. Holz. "Rainfall-runoff modelling using adaptive neuro-fuzzy systems." *Journal of Hydroinformatics* 3, no. 1 (January 2001): 3–10. doi: https://doi.org/10.2166/hydro.2001.0002.
[4] Tang, Y., P. Reed, K. Van Werkhoven, and T. Wagener. "Advancing the identification and evaluation of distributed rainfall-runoff models using global sensitivity analysis." *Water Resour. Res.* 43, no. W06415 (June 2007): 1-14. doi:10.1029/2006WR005813.
[5] Chang, Tak K., Amin Talei, Lloyd H.C. Chua, and Sina Alaghmand. "The Impact of Training Data Sequence on the Performance of Neuro-Fuzzy Rainfall-Runoff Models with Online Learning." *Water* 11, no. 1 (December 2019): 52. https://doi.org/10.3390/w11010052.
[6] López Guillén, María Elena. "Systems Identification. Application to the modeling of a DC motor / Identificación de Sistemas. Aplicación al modelado de un motor de continua." July 28, 2021. Accessed October 22, 2021.
http://www.ie.tec.ac.cr/einteriano/control/Laboratorio/3.7Identificacion%20de%20sistemas.PDF.
[7] Mathworks. (2018). Available at: https://la.mathworks.com/help/ident/ug/what-are-polynomial-models.html.
[8] Domínguez Mora, Ramón, Maritza Liliana Arganis Juárez, E.E. Carrizosa, C.A. Jiménez, R. Mendoza, G.G. Esquivel, A.J.L. Herrera, G.J.C. Ramírez, and Edgar Antonio Sotomayor Suárez. *Hydrological study and determination of the operation policy of the Catemaco lagoon / Estudio hidrológico y determinación de la política de operación de la laguna de Catemaco.* C.H. Chilapan, for Federal Electricity Commission (CFE). Final report, 2016.
[9] Sotomayor Suárez, Edgar Antonio. *Hydrological study of the Laguna de Catemaco basin and review of the operating policy at C. H. Chilapan / Estudio hidrológico de la cuenca de la Laguna de Catemaco y revision de la politica de operación en la C. H. Chilapan*. Postgraduate Master's Thesis in Engineering, UNAM, México, 2018.