Adaptive Fuzzy Level Set Streamflow Modeling and Forecasting*

Leandro Maciel[1][0000−0002−1900−7179], Rosangela Ballini[2,3][0000−0001−6683−4380], and Fernando Gomide[3][0000−0001−5716−4282]

[1] School of Economics, Business, Accounting and Actuary, University of São Paulo
leandromaciel@usp.br
[2] Institute of Economics, Universidade Estadual de Campinas, Brazil
ballini@unicamp.br
[3] School of Electrical and Computer Engineering, Universidade Estadual de Campinas, Brazil
gomide@unicamp.br

Abstract. The paper addresses the use of an adaptive, recursive fuzzy modeling based on the notion of level set to forecast monthly streamflows of a major hydroelectric power plant reservoir at the northeast of Brazil. Streamflows are highly complex nonstationary time series with high variability during the year, a feature that turns modeling and forecasting very hard and challenging. The adaptive level set method is evaluated against periodic autoregressive moving average models, currently adopted by many power industries, and against granular, neural, neural fuzzy, recurrent neural, and data driven level set models. The results show that adaptive level set modeling achieves the best root mean square error performance, surpassing all the models considered herein.

Keywords: Adaptive fuzzy systems · Data driven level set modeling · Time series forecasting.

1 Introduction

Planning and operation of hydro power and water resource systems involve many complex production relationships of several hydrological and operational assets. The natural streamflow is the key asset in hydro power and water resources planning. Streamflow data covering the entire planning period are required to effectively plan production and operate hydroelectric power plants. Streamflow forecast is of utmost importance in hydro energy system analysis, simulation, optimization, decision-making, and control as well.

Because hydroelectric systems are geographically spread out, hydrometric data are collected using sparse, distributed data acquisition systems, which often result in imprecise data and geophysical information. The highly nonlinear

* Supported by the Brazilian National Council for Scientific and Technological Development (CNPq) grants 304456/2020-9, 311260/2022-5, and 302467/2019-0, respectively, and by the Research Foundation of the State of Sao Paulo (FAPESP), grant 2020/09838-0.

relationship between the input and output flow significantly complicates streamflow modeling and forecasting. Another difficulty is the nonstationary nature of streamflows [1]. Often, wet periods show high stream variability, while in dry periods the variability is usually low. This scenario makes streamflow modeling and forecasting very difficult.

Several streamflow modeling and forecasting approaches have been reported in the literature. Many of them use hydrological models considering the meteorological forcing inputs, particularly the precipitation, measured *in situ*, or estimated from satellite and radar observation. Although physical hydrological models can be adopted, their use is difficult in practice because of the need for observation data, their complex structure, and the computational effort needed to calibrate [2]. Alternatively, machine learning methods can encode relationships in data and are an interesting alternative to forecast streamflows without detailed knowledge of the underlying physical phenomena [3]. A combination of machine learning with hydrometeorological approaches has been addressed for hourly streamflow forecasting [4], for instance. Granular computation is another alternative that has successfully addressed streamflow modeling and forecasting and has shown to be competitive with neural and neural fuzzy methods [5].

This paper addresses the use of an adaptive, recursive, data driven fuzzy modeling technique using the concept of level set recently introduced in [6] to forecast monthly streamflows of a major hydroelectric power plant reservoir in the northeast of Brazil. The level set method is compared with periodic autoregressive moving average, the model currently adopted by the power industry, with models developed by a granular pattern recognition method, a multilayer feedforward neural network, a multilayer fuzzy neural network, and a long short-term memory recurrent neural network. The results suggest that the data driven level set model outperforms the remaining ones.

The paper is organized as follows. The next section briefly reviews the level set [8], the data driven level set [7], and the recursive level set method [6]. Section 3 addresses the monthly streamflow modeling and forecast problem and evaluates the adaptive, recursive data driven level set method against classic and machine learning-based forecasters. Section 4 concludes the paper summarizing its contributions and suggesting issues for future work.

2 Level Set fuzzy Modeling

2.1 Fuzzy Level Set Models

Consider a fuzzy rule-based model with N fuzzy rules of the form:

$$\mathcal{R}_i: \text{ if } x \text{ is } \mathcal{A}_i \text{ then } y \; is \; \mathcal{B}_i \tag{1}$$

where $i = 1, 2, \ldots, N$ and $\mathcal{A}_i$ and $\mathcal{B}_i$ are known convex fuzzy sets with membership functions are $\mathcal{A}_i(x) : \mathcal{X} \rightarrow [0, 1]$ and $\mathcal{B}_i(y) : \mathcal{Y} \rightarrow [0, 1]$. Given an input data $x \in \mathcal{X}$, the output of the fuzzy model produced by the level set method is as follows [8].

1. Compute the activation level of each rule $\mathcal{R}_i$

$$\tau_i = \mathcal{A}_i(x) \tag{2}$$

2. Find the level set $\mathcal{B}_{\tau_i}$ for each $\tau_i > 0$

$$\mathcal{B}_{\tau_i} = \{y|\tau_i \leq \mathcal{B}_i(y)\} = [y_{il}, y_{iu}] \tag{3}$$

3. Compute the midpoints of the level sets

$$m_i(\tau_i) = \frac{y_{il} + y_{iu}}{2} \tag{4}$$

4. Compute the output of the model $\hat{y}$ using

$$\hat{y} = \frac{\sum_{i=1}^{N} \tau_i m_i(\tau_i)}{\sum_{i=1}^{N} \tau_i} \tag{5}$$

If the fuzzy set $\mathcal{B}_i$ is discrete, then m_i is the average of the elements of $\mathcal{B}_{\tau_i}$.

Expression (5) is an instance of a mapping $\mathcal{F} : [0, 1] \to \mathcal{Y}$:

$$\hat{y} = \frac{\sum_{i=1}^{N} \tau_i \mathcal{F}_i(\tau_i)}{\sum_{i=1}^{N} \tau_i} = \mathcal{F}(\tau) \tag{6}$$

in which $\mathcal{F}_i(\tau) = m_i(\tau_i)$ is one of the simplest. Other choices are possible and shall be addressed in a future paper.

It should be pointed out that $\mathcal{F}$ is a mapping from membership degrees to the elements of the output domain $\mathcal{Y}$, which significantly differs from the usual functional and linguistic fuzzy model rules processing mechanisms [11].

2.2 Data Driven Fuzzy Level Models

Suppose that the fuzzy sets $\mathcal{A}_i(x) : \mathcal{X} \to [0, 1]$ and $\mathcal{B}_i(y) : \mathcal{Y} \to [0, 1]$ are not known in advance. Let $\mathcal{D} = \{(x^k, y^k)\}, x^k \in R^p, y^k \in R$ be a data set such that $y^k = f(x^k), k = 1, 2, \ldots, K$. The task is to develop a fuzzy model $\mathcal{F}$ to approximate the function f using $\mathcal{D}$ [7].

Development of fuzzy linguistic rule-based models from data requires the specification of the number of rules N, and the membership functions of the antecedent and the consequent of each rule [11]. The number of rules can be identified using domain knowledge, clustering, or any structural identification method. When clustering is used, it is common to assign a cluster to a rule on a one-to-one basis, one cluster - one rule. Typically, the fuzzy c-means algorithm, or its variations, is chosen to cluster the data space. Determination of the number of clusters may need cluster validity measures or self-organizing methods [9].

In the data driven fuzzy modeling framework, once the antecedents membership functions $\mathcal{A}_i$ are determined, an estimate of the output functions $\mathcal{F}_i$, $i = 1, 2, \ldots, N$, can be derived directly from the data instead of identifying $\mathcal{B}_i$. An easy and effective way to do this is to assume $\mathcal{F}_i$ affine, namely

$$\mathcal{F}_i(\tau_i) = v_i\tau_i + w_i \tag{7}$$

The values of the coefficients v_i and w_i can be estimated using e.g. a least squares-based estimation procedure. Originally, [7] developed a pseudo-inverse-based solution whose steps are the following.

1. For each data par $(x^k, y^k) \in \mathcal{D}$
2. Compute activation levels

$$\tau_i^k = \mathcal{A}_i(x^k), \quad i = 1, \ldots, N$$

3. Let

$$\begin{aligned} z^k &= \frac{\tau_1^k(v_1\tau_1^k + w_1)}{s^k} + \ldots + \frac{\tau_N^k(v_N\tau_N^k + w_N)}{s^k}, \quad s^k = \sum_{i=1}^{N} \tau_i^k \\ \mathbf{d}^k &= [(\tau_1^k)^2/s^k, \tau_1^k/s^k, \ldots, (\tau_N^k)^2/s^k, \tau_N^k/s^k] \\ \mathbf{u} &= [v_1, w_1, \ldots, v_N, w_N]^T \end{aligned}$$

4. Assemble the matrices

$$\mathbf{z} = [z^1, \ldots, z^K]^T, \quad \mathbf{D} = [(\mathbf{d}^1)^T, \ldots, (\mathbf{d}^K)^T]^T, \quad \mathbf{y} = [y^1, \ldots, y^K]^T$$

5. Compute the vector of coefficients $\mathbf{u}$

$$\mathbf{u} = \mathbf{D}^+\mathbf{z} \tag{8}$$

6. Compute the model output $\hat{y}$

$$\hat{y} = \mathbf{du} \tag{9}$$

which corresponds to the output equation (5). $\mathbf{D}^+$ is the Moore-Penrose pseudo inverse of $\mathbf{D}$ [10]. The vector $\mathbf{u}$ is the solution of $min_u\|\mathbf{y} - \mathbf{z}\|^2$.

The processing steps of the data driven level set method, using clustering to granulate the input-output space, proceed as follows:

1. Cluster the data set $\mathcal{D}$ into N clusters.
2. Assign membership function $\mathcal{A}_i$ to each cluster i.
3. Find the vector of coefficients $\mathbf{u}$ using (8).
4. Compute the model output $\hat{y}$ using (9).

2.3 Adaptive Fuzzy Level Models

The adaptive fuzzy level set modeling proceeds similarly to the data driven level set modeling. Both use clustering to granulate the input-output space to determine the antecedent membership functions $\mathcal{A}_i$. They differ in how the vector of coefficients of the output functions are estimated. The adaptive variant of the data driven level set processes the data pairs (x^k, y^k) sequentially. In this case, the least squares estimates of the vector of coefficients are computed for each value of $k = 1, \ldots, K$. Therefore, it is desirable to systematize the parameter estimation computations recursively. Like in the previous section, the output functions are affine, but the estimates of coefficients v_i and w_i are found recursively. Recursive parameter estimation is well addressed in the literature, e.g. [16], [15]. In the context of recursive level set fuzzy modeling, the detailed steps are the following [6].

For each data par (x^k, y^k), $k = 1, \ldots$

1. Compute activation levels $\tau_i^k = \mathcal{A}_i(x^k)$, $i = 1, \ldots, N$
2. Let

$$\mathbf{d}^k = [(\tau_1^k)^2/s^k, \tau_1^k/s^k, \ldots, (\tau_N^k)^2/s^k, \tau_N^k/s^k]$$
$$s^k = \sum_{i=1}^{N} \tau_i^k$$

3. Compute gain matrix

$$\mathbf{P}^k = \frac{1}{\lambda}\left(\mathbf{P}^{k-1} - \frac{\mathbf{P}^{k-1}(\mathbf{d}^k)^T \mathbf{d}^k \mathbf{P}^{k-1}}{\lambda + \mathbf{d}^k \mathbf{P}^{\mathbf{k-1}}(\mathbf{d}^k)^T}\right)$$

4. Update output functions parameters

$$\mathbf{u}^k = \mathbf{u}^{k-1} + \mathbf{P}^k(\mathbf{d}^k)^T(y^k - \mathbf{d}^k\mathbf{u}^{k-1}) \tag{10}$$

5. Compute the model output $\hat{y}^k$

$$\hat{y}^k = \mathbf{d}^k\mathbf{u}^k \tag{11}$$

The adaptive, recursive level set fuzzy modeling needs initial estimates $\mathbf{u}^0$ of the parameters and the initial gain matrix $\mathbf{P}^0$. Usually this is done setting $\mathbf{u}^0 = 0$ and $\mathbf{P}^0 = \alpha\mathbf{I}$, where α is a large enough number [17]. As it is well known, λ is a forgetting factor used to weight the relevance of a sample in the data sequence. When λ is small, recent data are heavily weighted, and the estimation algorithm tracks time-varying systems parameters more efficiently [16].

The adaptive data driven level set method can be summarized as follows:

1. Cluster the data set $\mathcal{D}$ into N clusters.
2. Assign membership function $\mathcal{A}_i$ to cluster i.
3. For each step k do:
   (a) Find the vector of coefficients $\mathbf{u}$ using (10).
   (b) Compute the model output $\hat{y}$ using (11).

3 Streamflow Modeling and Forecast

This section uses the adaptive, recursive, data driven level set to model and forecast the average monthly inflows for a major hydroelectric plant called Sobradinho, located in Northeast Brazil. The hydrologic data covering the period from 1931 to 1990 (720 samples) are used to develop the models, and the data from 1991 to 1998 (96 samples) are used to test and to compare the performance of the different models. This is the same scenario considered in [5], and that will be used in this section to evaluate the forecasters.

The inflows oscillate between minimum and maximum values following the seasonal variation during the 12 months period. The seasonality of the flows justifies the use of 12 different models, one for each month of the year, as currently adopted by many hydrological systems worldwide. In this vein, twelve forecasting models were developed in [5] to forecast each of the 12 monthly inflow averages from January to December, from 1991 to 1998.

The forecasting methods adopted to evaluate and compare with the adaptive, recursive data driven level set method (RLSM) are the periodic autoregressive moving average (PARMA) [12], the granular functional model (GFM), the granular relational model with median recognition procedure (GRM-MRP), and the granular relational model with pattern recognition procedure (GRM-PRP) [5], as well as a multilayer neural network (MLP), and a fuzzy neural network (FNN) [13]. All these methods use the same forecasting modeling scheme considered in [5], namely, an individual model is developed to forecast each of the 12 monthly inflow averages. The RLSM, on the contrary, process data sequentially and develops only one forecasting model with two inputs of the form

$$\hat{y}^k = f(y^{k-1}, y^{k-2}) \tag{12}$$

encoded by fuzzy rules with a Gaussian membership function for each input variable in their antecedents. The rules were designed using the fuzzy c-means clustering algorithm.

The testing period is from January 1991 to December 1998. RLSM is also compared with the data driven level set method (DLSM) [6], a batch mode modeling method that uses the generalized inverse form of the least squares learning, and with a long short-term memory (LSTM) recurrent neural network with one dense layer and two recurrent layers. RLSM and LSTM process the 720 training samples sequentially to capture the temporal relationships in the data. Likewise, they process the 96 testing data samples sequentially.

Evaluation uses the root mean square error RMSE (m^3/s) computed from the forecasts produced by the models using test data. The result is summarized in Table 1 whose values, except for RLSM and DLSM, were taken from [5] and [18]. The RLSM achieves a RMSE = 997.87 in 0.15 seconds when processing training and testing data sequentially. If the consequent parameters of the fuzzy rules are kept frozen after training, then it reaches RMSE = 1029,44. In this case RLSM ran in 0.094 seconds.

Table 1. Performance evaluation of streamflow forecasting

Forecasting Method	Rules or Neurons	Number of Parameters	Testing Data	
			Samples	RMSE
PARMA	-	2[1]	96	1079.30
GFM	-	6[1]	96	1471.20
MLP	21[1]	84	96	1462.80
FNN	46[1]	184	96	1330.40
GRM-MRP	-	8[1]	96	1191.60
GRM-PRP	-	12[1]	96	1005.00
LSTM	1+2[2]	43	96	1162.90
DLSM	4	24	96	1029.28
RLSM	4	24	96	**997.87**

[1] Average/month [2] 1 Dense and 2 LSTM layers

Fig. 1 shows the forecast produced by the adaptive level set model, and Fig. 2 and Fig. 3 show the forecast of the data driven level set and the long short-term memory, respectively. Interestingly, notice that the RLSM is particularly superior in predicting higher levels of inflows, a very difficult and challenging task for streamflow forecasters.

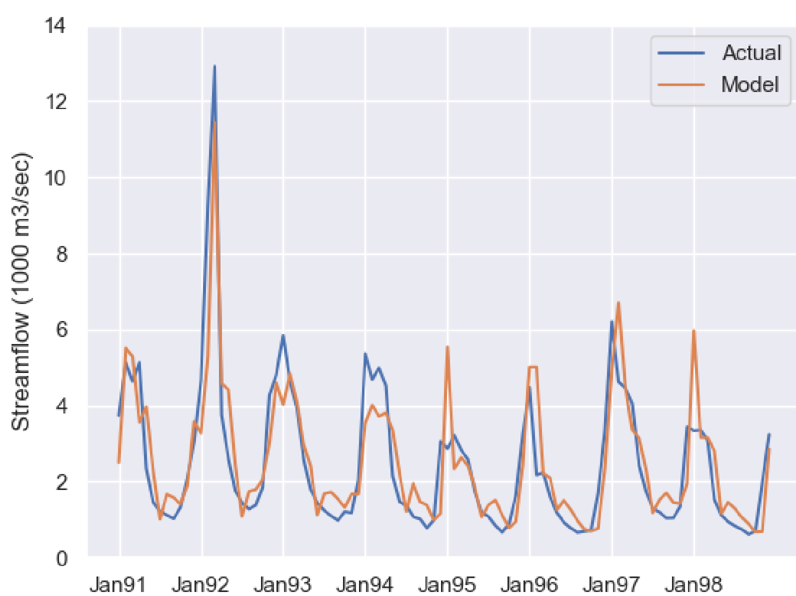

Fig. 1. RLSM forecast for the testing period.

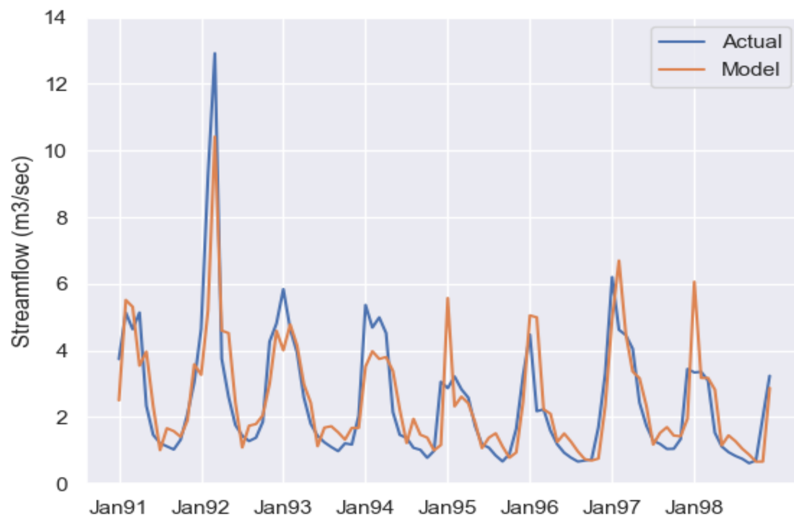

Fig. 2. DLSM forecast for the testing period.

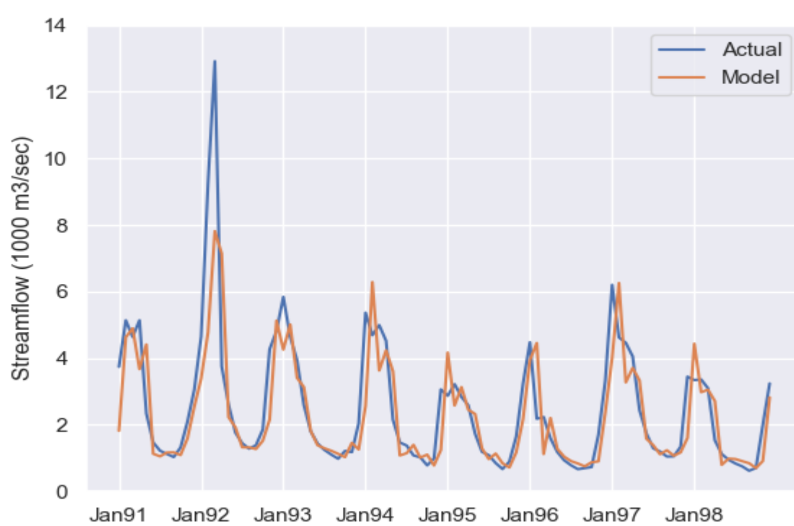

Fig. 3. LSTM forecast for the testing period.

4 Conclusion

This paper addressed adaptive data driven fuzzy modeling based on the of level set concept to forecast average monthly streamflows of a major hydroelectric power plant reservoir at the northeast of Brazil. The adaptive level set model was compared with the current power industry standard, the periodic autoregressive moving average, with granular pattern recognition-based model, a multilayer feedforward neural network, a fuzzy neural network, a long short term memories, and the data driven level set model. The results suggest that the adaptive data driven level set model performs best among the remaining models. Future work will consider extensions of the level set fuzzy modeling of granular time series, and applications in system identification and control.

References

1. Maier, H. and Dandy, G.: Neural networks for the prediction and forecasting of water resource variables: A review of modelling issues and applications. Env. Modeling and Software **15**, 101–124 (2000)
2. Yaseen, Z., Sulaiman,S., Deo, R., Chau, K.: An enhanced extreme learning machine model for river flow forecasting: State-of-the-art, practical applications in water resource engineering area and future research direction. J. Hydrol. **569**, 387–408 (2018)
3. Adnan, R., Liang, Z., Trajokovic, S., Kermani, M., Li, B., Kisi, O.: Daily streamflow prediction using optimally pruned extreme learning machine. J. Hydrol. **577**, 123981 (2019)
4. Liu, J., Yuan, X., Zeng, J., Jiao, Y., Li, Y., Zhong, L., Kisi, O.: Ensemble streamflow forecasting over a cascade reservoir catchment with integrated hydrometeorological modeling and machine learning. Hydrol. and Earth Sys. Sciences **26**(2), 265–278 (2022)
5. Hirota, M., Ballini, R., Gomide, F.: Granular models for time-series forecasting. in Handbook of Granular Computing, Pedrycz, W., Skowron, A., Kreinovich, V. (eds.) 949–967. Wiley Online Library, https://doi.org/10.1002/9780470724163.ch45 (2007).
6. Maciel, L., Gomide, F.: Recursive Level Set Fuzzy Modeling. IEEE Int. Conf. on Evolving and Adaptive Intelligent Systems 2024, IEEE Press, Madrid (2024).
7. Leite, D., Gomide, F., Yager, R.: Data driven fuzzy modeling using level sets. 2022 IEEE Int. Conf. on Fuzzy Systems, 1–5, IEEE Press, https://doi:10.1109/FUZZ-IEEE55066.2022.998, Padova (2022).
8. Yager, R.: An alternative procedure for the calculation of fuzzy logic controller values. J. of the Japanese Soc. for Fuzzy Theory and Systems **3**(4), 736–746 (1991)
9. Lughofer, E.: Evolving Fuzzy Systems: Methodologies, Advanced Concepts and Applications, Springer, Heidelberg (2011)
10. Serre, D.: Matrices: Theory and Applications, Springer, New York (2010)
11. Pedrycz W., Gomide, F.: Fuzzy Systems Engineering: Toward Human-Centric Computing, John Wiley & Sons, Hoboken (2007)
12. Vecchia, A.: Maximum likelihood estimation for periodic autoregressive moving average models. Technometrics **27**(4), 375–384 (1985)
13. Figueiredo M., Gomide, F.: Design of fuzzy systems using neurofuzzy networks. IEEE Trans. on Neural Networks **10**(4), 815–827 (1999)

14. Greff, K., Srivastava, R., Koutnik, J., Steunebrinkand, B., Schmidhuber, J.: LSTM: A search odyssey. IEEE Trans. on Neural Networks and Learning Systems **28**(10), 2222–2232 (2017)
15. Astrom K., Wittenmark, B.: Computer Controlled Systems: Theory and Design, Dover, Mineola (2011)
16. Jang J., Sun, C., Mizutani, E.: Neuro-Fuzzy and Soft Computing, Prentice Hall, Upper Saddle River (1997)
17. Jung L.: System Identification: Theory for the User, Pearson, Upper Saddle River (1999)
18. Maciel, L., Ballini, R., Gomide, F.: Streamflow forecasting using data driven level set modeling. 2023 IEEE Int. Conf. on Fuzzy Systems, 1–5, IEEE Press, ISBN: 978-8-3503-3228-5, Seoul (2023).