Proceeding Series of the Brazilian Society of Computational and Applied Mathematics, v. 9, n. 1, 2022.

Trabalho apresentado no XLI CNMAC, Unicamp - Campinas - SP, 2022.

Proceeding Series of the Brazilian Society of Computational and Applied Mathematics
Preprint

Level river forecasting using empirical hydrological modeling for Rio Negro basin Uruguay

Johan S. Duque[1] Leonardo do Santos[2] Rafael Santos[3]
Applied Computing, National Institute for Space Research, INPE, Sao Jose dos Campos, SP, BR.
Johny A. Arteaga[4] Natalie Aubet[5]
Universidad Tecnológica del Uruguay, DZ, UY.

Abstract. Climate change has influenced several of the water cycle related variables such as rainfall that contribute to increasing natural disasters. To establish new methodologies for rivers level forecasting is necessary for the implementation of early warning systems. In this work, we present results of a multilayer perceptron artificial neural network (ANN) to forecast temporal series of water levels at the outlet of Rio Negro river with 24-hour antecedence. Input data was collected by a set of hydrological monitoring stations composed of water level and rainfall measures acquired with a one-day resolution. Water-level prediction were evaluated by the Nash-Sutcliffe coefficient (NSE) and by the root mean square error (RMSE). The results show consistency between predicted and observed values, especially when combining both water level and rainfall data. In such case, values of NSE reached 0.93 to 0.54 and RMSE between 0.028 and 0.061 for antecedence of 1 to 7 days respectively with implemented topology for the empirical model.

Key-words. Empirical hydrological modeling,Water-level, Rain, Neural networks.

1 Introduction

Hydrological models are essential for monitoring water catchment sources and early warning of natural disasters such as floods and landslides, one of the most frequent natural disasters in Uruguay and Brazil. With losses in the industrial economics and human lives, these kinds of problems and disasters are becoming more common every day with extreme weather events [9, 17]. Due to meteorological consequences associated with climate change and its impacts, see [1].

The complexity in hydrological modeling has prompted a need for using emerging techniques for calibration and optimization, such as data-driven models, particularly the ANN based models have become popular in recent years for hydrological studies [4, 6, 11].

For its capacity to learn and to generalize the knowledge of pairs of sufficient data, ANN models have been successful in processes such as estimating river flows (flow, level, flow volume), making flood warnings, operating reservoirs for flood control, determining the water potential of the stream, hydroelectric production in dry periods, and planning transportation in streams, previously studied by other authors see [3, 8].

[1]johan.duque@utec.edu.uy
[2]santoslbl@gmail.com
[3]rafael.santos@inpe.br
[4]johny.arteaga@utec.edu.uy
[5]natalie.aubet@utec.edu.uy

The data-driven models can solve complex problems on a big scale such as the recognition of forms, classification, association and forecasting, it can efficiently capture the nonlinear or linear relationship between the natural processes. ANN based models can also identify the consequences of data acquisition, analysis and processing [15] [13]. Data-driven models are known as empirical or black box models because they do not include the physical conservation laws.

Mathematically, an ANN problem can be considered as following: given a dataset $S = \{(x_i, y_i = f^*(x_i), i \in [n]\}$ with $x_i \in X := [0,1]^d$ points with distribution P and $sup_{x \in X}|f^*(x)| \leq 1$, the aim is to approximate f^* accurately as possible [7]. According to Weinan E. et. al. 2020 the standard procedure in supervised learning approach is the following: 1) choose hypothesis space, the set of trial functions which will be denoted by H_m. The subscript m characterizes the size of the model, it can be the number of parameters or neurons in an ANN. 2) Choose a loss function. The primary goal is to fit the data. Therefore the most popular choice is the "empirical risk" $R_n(f) = \frac{1}{n}\sum_i (f(x_i) - f^*(x_i))^2$. 3) Choose an optimization algorithm and the hyper-parameters. The most popular choices are gradient descent (GD), stochastic gradient descent (SGD) and advanced optimizers such as Adam and RMSprop [7]. The overall objective is to minimize the "population risk", also known as the "generalization error": $R(f) = E_{x \sim P}(f(x) - f^*(x))^2$.

In practice, $R(f)$ estimated on a finite data set (which is unrelated to any data used to train the model) and is called test error, whereas the empirical risk (which is used for training purposes) is called the training error. To extend the mathematical foundations of neural networks, the reader can consult the following articles where the authors address this discussion rigorously [18].

There are many open problems and challenges to be solved for neural networks modelling in the hydrological context: can we optimize the loss function using the selected training algorithms for hydrometric input data? The solution obtained from training, generalize well enough to test data associated with the river level? What are the best ANN parameters?

In this work we present a multilayer perceptron (MLP) artificial neural network for river level prediction on the watershed of the Rio Negro basin in Uruguay. This letter aims to establish an empirical rain-level model throughout Rio Negro watershed with data taken from nineteen hydrometric stations in order to determine the level in the lowest station with antecedence of 24 hours with combination of the input data.

With the proposed network architecture and with the combination of the level and rainfall input data, our results show similarity between the observed and predicted data with good accuracy. It is relevant to mention here that the temporal dynamics of streamflow of the Rio Negro basin has not been studied by empirical methods.

2 Data and methods

2.1 Study area and data

The Rio Negro basin is a sub-basin of the Uruguay river reaching a length of 850 km, with 700 of them in Uruguayan territory, crossing entirely in an East-West direction one and comprises most of the meteorological data found in this area. Its drainage network is a trans-boundary system that occupies more than one third of the Uruguayan territory and the southern part of Rio Grande do Sul state in Brazil as shown figure 1. The natural flow of stream in this sub-basin is affected by many reservoirs and dams in the middle Rio Negro River basin and by many reservoirs and diversion for irrigation. The river basin is rich in resources such as forestry, agriculture, hydro-power that favor economic development and make the basin a region that is attractive to population and

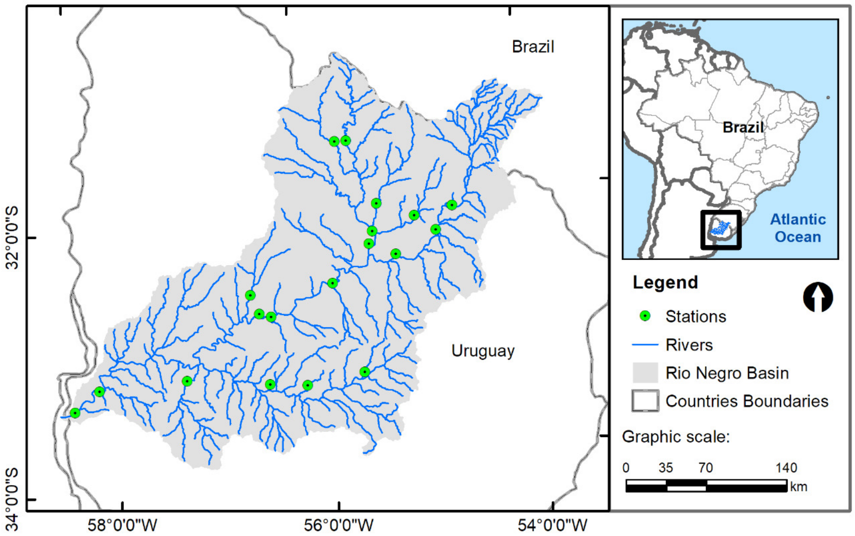


Figure 1: Study region and location of the hydrometric stations

industry [14].

The Rio Negro basin has generated flooding emergencies due to rainfall in the middle and lower basin, causing the displacement of more than 5000 people in January and June 2019, with Mercedes region being the most affected department, followed by Durazno and Río Negro [10]. The intense rains that occurred in a short period of time caused the overflow of some water courses linked to this basin causing flooding along its channel.

The daily data used in the framework of this study were obtained from the nineteen hydrometric monitoring stations from the National Administration of Power Plants and Transmission (UTE) system website, see [2]. The location of the stations in the Rio Negro basin is shown in figure 1. Rainfall and level data cover the period from January 01, 2015 to December 31, 2020, summing 2192 readings. The selected period includes some extreme precipitation events that cause floods problems in 2016 and 2019 in the study area, specifically in Mercedes city department of Villa Soriano reported by the national emergency system, see [16].

The source code and the data set used in this work are available for download on the github private repository upon request [12].

2.2 Rain-level modeling method by neural network

The setting of an ANN constitutes the following four steps: data search preprocessing, learning and evaluation of the model. In this study, a perceptron multilayer neural network was implemented with two layers: an input layer with 38 nodes (one neuron for each input parameter), 1 hidden layer with 18 neurons, and an output layer with 1 neuron. The input layer was configured to receive tuples with input characteristics: water level and rain values observed in each of the 19 hydrometric stations.

We employs ReLu, $Relu(X) = max(X, 0)$ (rectified linear unit), activation function for hidden layer and linear function for output layer. Some model parameters were calibrated through trial and error processes to fine-tune the parameters of the empirical model to get reliable estimation

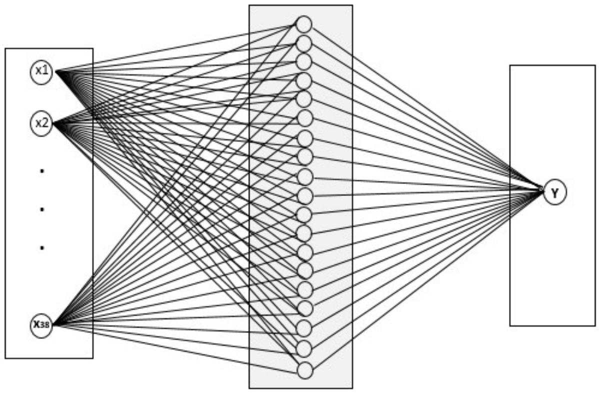


Figure 2: Architecture of the rain-level neural network model

of river level. The simulated level hydrograph model was computed in prior runs of the flood forecasting model to estimate the best parameters for the empirical model. The river rain-level model was evaluated with two metrics (Nash coefficient and RMSE) and graphic criteria between the observed and simulated levels. The Nash coefficient and RMSE are given by the following mathematical expressions,

$$NSE = \frac{\sum_{t=1}^{T}(Q_p^t - Q_0^t)^2}{\sum_{t=1}^{T}(Q_0^t - \bar{Q}_0^t)^2}, \tag{1}$$

$$RMSE = \sqrt{\frac{\sum_{t=1}^{T}(Q_p^t - Q_0^t)^2}{T}}, \tag{2}$$

where Q_0^t is the mean of the actual level in the time interval $[1; T]$ that contains T discrete values, Q_p^t is the predicted level at time t, and Q_0^t is the actual level at time t. The NSE is commonly used for hydrological models and it is equivalent to the correlation coefficient between predicted and actual data [5].

Figure 2 shows the topology of the neuronal network implemented for rain-level model, this architecture gives better performance according with the model evaluation, which must be higher than 60% for the NSE criteria equation 1 and 70% for RMSE equation 2. However, in the modeling of an ANN, the quantity of hidden layer is higher, and the model becomes complex and unstable. To reduce model complexity it is necessary to choose architectures with fewer layers [15]. The appropriate model in the framework of this study is the ANN architecture shown in figure 2 with 38 nodes for input data, 18 neurons in the hidden layer and 1 neuron in the output.

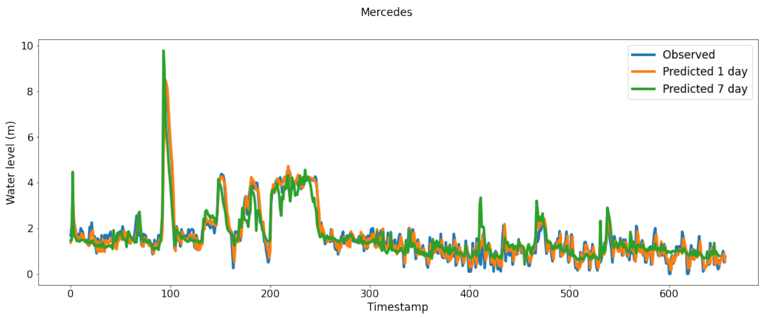
The ANN purpose is to predict water level temporal series at the outlet of the Rio Negro watershed (Mercedes station) using as input temporal series of water level and/or rainfall at different hydrological monitoring stations as described previously. The predictions were done for antecedence ranging from 1 to 7 days. Input data was split as follows: 70% for the training, 30% for the validation, and test. Data in the period from January 2015 to March 2019 with randomly sorted samples was used for model training, while data from the months of April 2019 to December 2020 was used for test, corresponding to a period of 658 days. Three set of tests were performed, based on different combinations of the available input datasets. The first test employs only water level data, the second one uses only rainfall data, while the third one inputs both water level and rainfall data following the previous methodology [5].

3 Results and Discussion

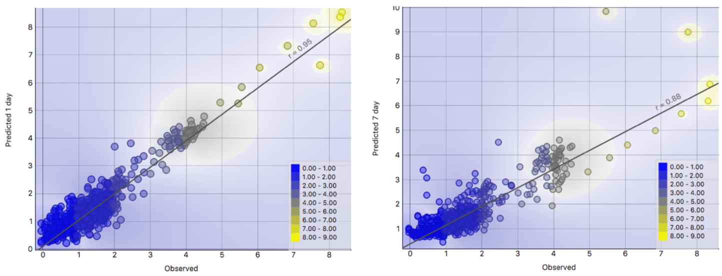
Predicted and observed temporal series of the water level at the Mercedes station with an antecedence of 1 and 7 days are shown in figure 3a for a period of 658 time intervals, corresponding to days of the months of April 2019 to December 2020. In the time series predicted it is possible to observe level increases in some instants of time that coincide with the reports of floods in the study area, for example around 100 Timestamp corresponding with days between June 18 and 24 of 2019, this results coincide with data reported by the national emergency system with a flood level of 8.4 meters causing serious disasters in the study area.

Based on network architecture for rain plus level time series input, the results show good correspondence between predicted and observed values for a prediction margin of 24 hours of antecedence, for prediction margin of 7 days of antecedence results presents an increase in the RMSE around 3.3% as presented in table 1.

In level hydrographs see figure 3a it is possible to observe levels that are well-adjusted with the ANN model, it is also possible to identify the difference between predictions with precedence of 1 and 7 days, some peaks associated with higher levels are not so good because these alterations in the level can be caused by the influence of dams located in the river basin. In this case, the model presented the highest values for the NSE evaluation metric 93% and the lowest for the RMSE values 2.8%. The complete result of the evaluation metrics of the implemented models are presented in Table 1.



(a) Observed and predicted hydrographs.



(b) Scatter plot between predicted and validation data.

Figure 3: Observed and predicted values of water level time series at the outlet for 1 and 7 days of antecedence, considering rainfall and level as input.

Figure 3b presents the relationship between the observed and the simulated river levels in the Mercedes station by the model for antecedence of 1 day and 7 days of antecedence respectively. The r^2 value associated in each case indicated in the plot was a standard fit line and significant for demonstrating model performance. The predictions time horizon made for longer periods of time

DOI: 10.5540/03.2022.009.01.0269

Table 1: Model evaluation metrics for water level and rainfall input

Day	**Rain**		**Level**		**Rain + Level**	
	NSE	**RMSE**	**NSE**	**RSME**	**NSE**	**RSME**
1	0,9252	0,0291	0,9379	0,0273	0,9326	0,0280
2	0,8209	0,0431	0,8803	0,0364	0,8663	0,0373
3	0,7225	0,0532	0,8166	0,0434	0,8059	0,0425
4	0,6188	0,0611	0,7443	0,0503	0,7323	0,0496
5	0,5175	0,0659	0,6557	0,0538	0,6256	0,0549
6	0,4578	0,0701	0,5798	0,0607	0,5682	0,0591
7	0,4213	0,0719	0,5561	0,0601	0,5432	0,0610

decreased with the increase in days, it was expected according to previously works [5] [15].

4 Conclusions

In this study, an empirical hydrological model was implemented for river level forecasting in the Rio Negro basin using a MLP neural network. The historical level and rain data of the stations were analyzed in order to estimate the future river level and evaluate the model. The model demonstrates a strong learning ability for rain-level and level time series, but it can occasionally present poor performance results due to the random selection of initialization parameters. The results show good performance with RMSE between 0.029 and 0.071 considering as input the data of the precipitation time series only, 0.027 and 0.061 considering as input the data of the level time series only and a RMSE between 0.028 and 0.061 considering the time series of rain plus level. These results indicate good accuracy considering that Nash-Sutcliff coefficients were larger than 90% for the three ANN topologies implemented.

This study can be useful for streamflow and level prediction in critical places considering the alterations in the river basin due to the presence of dams, which add a complexity factor to the model to be considered in subsequent studies. In addition, understanding the temporal dynamics of streamflow and other hydrologic processes continues to be challenging, the level and rain time series are nonlinear, and many parameters can affect these time series. This work can be reconstructed with different input parameters and prepare future hydrologic studies.

Acknowledgments

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001. This work is supported by a Grant of FAPESP 2015/50122-0, DFG-IRTG 1740/2, and CNPq 420338/2018-7. The authors are thankful to Cintia Pereira de Freitas for her help and considerations in this work.

References

[1] IPCC Fourth Assessment Report. Website Accessed 10/12/2021. In: `https://archive.ipcc.ch/publications_and_data/ar4/wg2/en/ch3s3-4-3.html` (2007).

[2] UTE official website. Accessed 06/07/2020. In: `https://portal.ute.com.uy/precipitaciones-ocurridas-y-prevision-de-niveles` ().

[3] C. W. Dawson and R. L. Wilby. "Hydrological modelling using artificial neural networks." In: **Progress in Physical Geography** 25.1 (2001), pp. 80–108.

[4] Editorial. "The rise of data-driven modelling". In: **Nat Rev Phys.** https://doi.org/10.1038/s42254-021-00336-z (2021), p. 383.

[5] C. Pereira de Freitas. "Combining Rainfall and Water Level Data for Multistep High Temporal Resolution Empirical Hydrological Forecasting". In: (2020).

[6] M Jajarmizadeh and S Harun. "A Review on Theoretical Consideration and Types of Models in Hydrology". In: **Journal of Environmental Science and Technology** (2012), pp. 249–261.

[7] Hornik K, Stinchcombe M, and H. White. "Multilayer feedforward networks are universal approximators". In: **Neural Networks.** doi:10.1016/0893-6080(89)90020-8 (1989), pp. 359–366.

[8] H.C. Daily Kilinc. "Streamflow Forecasting Based on the Hybrid Particle Swarm Optimization and Long Short-Term Memory Model in the Orontes Basin". In: **Water.** https://doi.org/10.3390/w14030490 (2022), pp. 14–490.

[9] T. Kokkonen, H. Koivusalo, and T. Karvonen. "A semi-distributed approach to rainfall- runoff modelling—a case study in a snow affected catchment." In: **Environmental Modelling and Software, 16(5).** (2001), pp. 481–493.

[10] Online site. Ministerio de Transporte y Obras Publicas. Accessed 31/08/2020. https://www.gub.uy/ministerio-transporte-obras-publicas/institucional/informacion-gestion/pedidos-informes/inundaciones-ciudad-mercedes-soriano.

[11] A. Mosavi, P. Ozturk, and K.W. Chau. "Flood prediction using machine learning models: literature review". In: **Water** 10.11 (2018).

[12] Website repository. In: https://github.com/JD39/Empirical-hidrological-forecasting.git (2021).

[13] Zhu. S et al. "An improved long short-term memory network for streamflow forecasting in the upper Yangtze River". In: **Stoch Environ Res Risk Assess,** 34 (2020), pp. 1313–1329.

[14] M. Sadres. "Desarrollo de un modelo en la cuenca alta del Río Negro (Uruguay) utilizando Hydrobid: Análisis de la Disponibilidad de Agua en Diferentes Escenarios". In: **Tesis de Maestría, Universidad de Alcalá**. 2019.

[15] L.B.L Santos et al. "An operational dynamical neuro-forecasting model for hydrological disasters". In: **Modeling Earth Systems and Environment,** (2001), p. 94.

[16] **Sistema Nacional de Emergencias. Online. Accessed 01/03/2020**. https://www.gub.uy/sistema-nacional-emergencias/comunicacion/noticias/eventos-adversos-significativos-ocurridos-uruguay-entre-anos-2015-2019.

[17] Chahine M. T. "The Hydrological Cycle and Its Influence on Climate". In: **Nature** (1992), pp. 373–380.

[18] Weinan et al. "Towards a Mathematical Understanding of Neural Network-Based Machine Learning: what we know and what we don't". In: https://arxiv.org/pdf/2009.10713.pdf (2020), pp. 1–56.

