Revista Brasileira de Climatologia
ISSN: 2237-8642 (Eletrônica)

A COMPARISON USING STATISTICS AND MACHINE LEARNING METHODS FOR STREAMFLOW TIME SERIES

VILLAVICENCIO, Lourdes – mmendoza@ufrn.edu.br
Universidade Federal Do Rio Grande Do Norte / UFRN

MENDES, David – david.mendes22@gmail.com
Universidade Federal Do Rio Grande Do Norte / UFRN

MONTEIRO, Felipe – felipefmonteiro@ufrn.edu.br
Universidade Federal Do Rio Grande Do Norte / UFRN

ANDRADE, Lara – Lara@ccet.ufrn.br
Universidade Federal Do Rio Grande Do Norte / UFRN

SILVA, Cassia – cassia.silva@ueg.br
Universidade de Brasília, Universidade Estadual de Goiás / UNB, UEG

ABSTRACT: This study was carried out in the Sibinacocha lake watershed in the Peruvian Andes. In this region the long-term meteorological data are scarce and there are few studies of flow forecasts. Based on this evidence, in this study we present the monthly flow simulation, using statistical models and data-oriented models, to evaluate the performance of these methodologies. The results of the comparative statistical analyses indicated that the data-oriented models, specifically the Recurrent Neural Networks, provided great improvements over the other models applied, specifically the ability to capture the minimum and maximum monthly flow, resulting in excellent statistical values (R^2=0.85, d=0.96), thus suggesting this methodology as a possible application for flow forecasts.

KEYWORDS: Time-series analysis, Streamflow Forecasting, Neural Networks.

UMA COMPARAÇÃO USANDO MÉTODOS ESTATÍSTICOS E APRENDIZADO DE MÁQUINA PARA PREVISÃO DE SÉRIES TEMPORAIS DE VAZÃO

RESUMO: Este estudo foi realizado na bacia hidrográfica do lago Sibinacocha nos Andes peruanos. Nesta região há uma escassez de dados meteorológicos de longo prazo, além de poucos estudos no que refere à previsão da vazão. Com base nesta evidência, este estudo apresenta a simulação da vazão mensal, utilizando modelos estatísticos e modelos orientados a dados, com o objetivo de avaliar o desempenho destas metodologias. Os resultados da análise estatística comparativa indicaram que, os modelos orientados a dados, especificamente o modelo de Rede Neural Recorrente, proporciona grandes melhorias em relação aos outros modelos aplicados neste estudo, conseguindo representar os valores mínimos e máximos da vazão mensal (R^2=0,85, d=0,96). Dessa forma sugere-se, a aplicação desta metodologia como possível aplicação para previsões de series temporais de vazão.

PALAVRAS-CHAVE: Análise de séries temporais, previsão de vazão, redes neurais.

UNA COMPARACIÓN USANDO MÉTODOS ESTADÍSTICOS Y DE APRENDIZAJE AUTOMÁTICO PARA SERIES DE TIEMPO DE CAUDAL

RESUMEN: Este estúdio fue realizado en la cuenca hidrográfica del lago Sibinacocha en los Andes peruanos. En esta región existe una escasa información de datos meteorológicos a largo plazo y pocos estúdios relacionados a la predicción de caudales. Con base en esta evidencia, este estúdio presenta la simulación de caudal mensual, utilizando modelos estadísticos y modelos orientados a datos, con el objetivo de evaluar

el desempeño de estas metodologias. Los resultados del análisis estadístico comparativo indicaron que, los modelos orientados a datos, especificamente el modelo de la Red Neural Recurrente, proporciona grandes mejoras en relación a los otros modelos aplicados en este estúdio, consiguiendo representar los valores mínimos y máximos del caudal mensual (R^2=0,85, d=0,96). Desta manera se sugiere esta metodologia como posible aplicación para la previsión de series temporales de caudales.

PALABRAS CLAVE: Análisis de series temporales, previsión de caudal, redes neuronales.

1. INTRODUCTION

Time series of flows are of great interest in hydrological research, as they are important for efficient river basin planning and sustainable water resources management, executing natural disaster forecasting, including droughts and floods, hydraulic infrastructure projects, electric energy generation (ZHAO & ZHAO, 2014; LEHNER et al., 2017).

Thus assisting hydrologists, water resource planners, public institutions, hydroelectric companies, and policy-makers in the effective management of water and the preservation of this natural resource. Therefore, in recent decades the development of new approaches, together with the improvement of available ones, has received much attention from hydrologists around the world to model and accurately estimate flow processes to address the problems mentioned above (MOHAMMADI et al., 2006; MEHDIZADEH et al., 2019; WU et al., 2010; NIU et al.,2020).

Generally, hydrological models can be divided into two groups: physical-based conceptual models and empirical models (FATHIAN et al., 2019). Physics-based models are sophisticated models that require a lot of information from various data sources (such as rain, evaporation, solar radiation, land use data, among others), to simulate flow behavior. However, the main disadvantage of these physics-based models is the presence of numerous region-dependent parameters that require calibration and validation, making it difficult to optimize them (FARFÁN et al., 2020). Besides, there can be enormous uncertainty in the models predictions, due to uncertainty in the input data, because inaccessibility to the various sources of information, limits the use of these models (MEHDIZADEH et al., 2019; WAGENA et al., 2020). In this regard, many researchers have studied the adoption of empirical models to predict future flow based on a long collection of historical flow records (WU et al., 2010; NIU et al.,2019; TIKHAMARINE et al., 2020). These empirical models are widely used and easier to apply to flow modeling as they do not require complex data sets or physical requirements (MEHDIZADEH et al., 2019).

Empirical models include: statistics models, the most recognized for hydrological simulations, exponential smoothing methods (CHATFIELD & YAR, 1988) e Box & Jenkins (BOX et al., 2008) as ARMA (autoregressive moving average) and ARIMA (integrated autoregressive moving average). And the data-driven models (machine learning algorithms), among the most used by hydrological scientists are the artificial neural network (ANN), for its learning ability, generalizability and satisfactory results in accurate flow forecast estimates (MONTANARI et al., 2000; SHAHWAN & ODENING, 2007; MEHDIZADEH et al., 2019).

ISSN: 2237-8642 (Eletrônica)

In this context, a comparative analysis developed to evaluate the performance of statistical modeling and data-driven modeling for forecasting average monthly flows in the Sibinacocha Lake watershed, located in the central Andes of southern Peru. Also, the empirical methodologies were not found so far for forecasting the flow in this region, which means that this article has the potential to fill this research gap and contribute to the decision-making of the managers responsible for the region.

2. DATA AND METHODS.

2.1 STUDY AREA

The Sibinacocha lake watershed is located in the heart of the Vilcanota mountain range in the Peruvian Andes, between coordinates 13° 27' 10" at 14° 29' 45" south latitude and 70° 41' 10" at 71° 20' 05" west longitude (Figure 1), with an altitude of 4860 meters above sea level (m.a.s.l). The region has a subtropical climate of rainy summers (December - March) and dry winters during the rest of the year (RUBEL & KOTTEK et al., 2011; KRONENBERG et al., 2016). The average annual rainfall in the highland region is approximately 700 mm (PERRY et al., 2013). Regarding the trends in temperature, these vary seasonally, where the minimum daily temperature increases stronger than the maximum daily temperature (SALZMANN et al., 2013).

It is important to highlight that the Sibinacocha lake watershed is of glacial origin and has the important function of supplying water to the Vilcanota-Urubamba River during the dry season. In this sense, the Machupicchu Electric Company built a dam in this watershed in 1988, to maintain the operation of the Machu-Picchu and Santa Teresa hydroelectric plants (CATACORA, 2008; DRENKHAN et al., 2019)

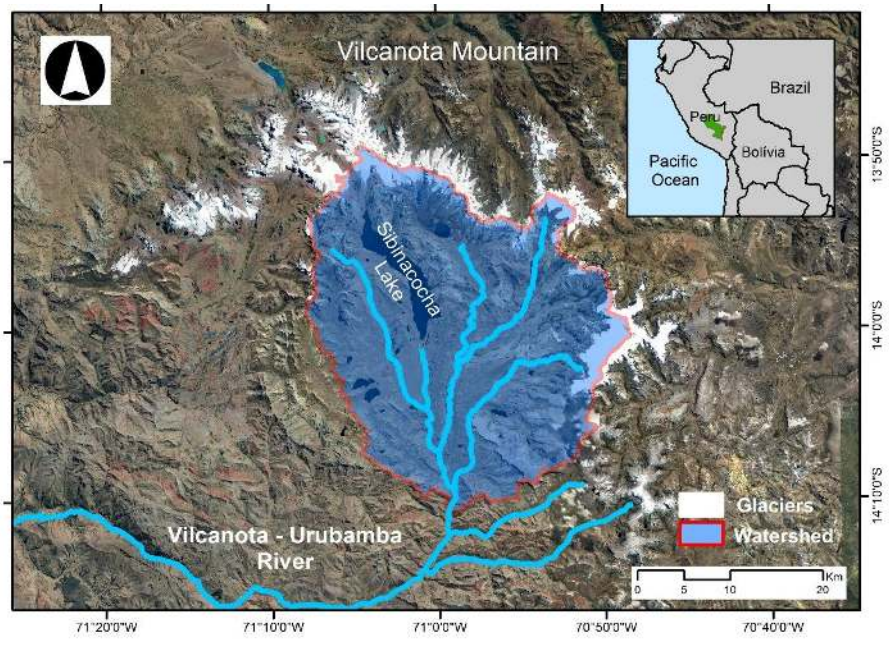

Figure 1 - In the figure, in the upper right corner, shows the location of the department of Cusco-PERU, where the Sibinacocha lake watershed is located, shown in the center of the figure.

2.2 DATASETS, ANALYSIS STATISTICS, AND FORECASTING

In this work, the data of the observed monthly flow comes from the Alto Urubamba station, this set of data is available free of charge at the National Water Authority of Peru (http://sofia.ana.gob.pe). Note that for the basin under study, flow data are available from 01-01-1965 until 31-12-2013 without missing information. Statistical methods and machine learning methods are used to calculate flow estimates. The statistical methods used are the Holt-Winters model and the Seasonal Arima model. And the machine learning methods were: Autoregressive Neural Network and Recurrent Neural Networks.

For the statistical analysis, we execute the descriptive analysis of the flow data, following, trend analysis, for the dry and wet season, using the modified Mann Kendall test, the advantage of using this technique is that it avoids problems of temporal autocorrelation in the data, besides being little influenced by abrupt changes (HAMED&RAO, 1998; ZHANG et al., 2009). After that, we applied the technique of repeated measure variance analysis (HAND & TAYLOR, 1987) for the dry and wet season, with the purpose of testing whether the mean flow before dam construction differs significantly from the mean flow after dam construction. Finally, the decomposition of the time series was executed, then we applied the Dickey-Fuller Augmented (ADF) test (DICKEY & FULLER, 1981) to verify the non-stationarity of the time series.

In machine learning methodology it is good practice to divide the available data into two subsets of data (Figure 2), generally, the division is 70% of the data for the training set and 30% of the data is used for evaluation or testing (AFAN et al., 2015). However, before executing the data division, we point out that, because it is a historical monthly flow data, it is important to consider the presence of possible monthly seasonality in the series. In this sense, the training phase consists of presenting the data to the algorithm, where the algorithm learns the characteristics or patterns useful for the objective of the knowledge discovery process. After the training, it will be evaluated or tested, using new data not seen by the algorithm. The use of this new data set will provide realistic measures of the algorithm's performance since the estimated values of the test data set will be compared with the values observed through the application of statistical metrics.

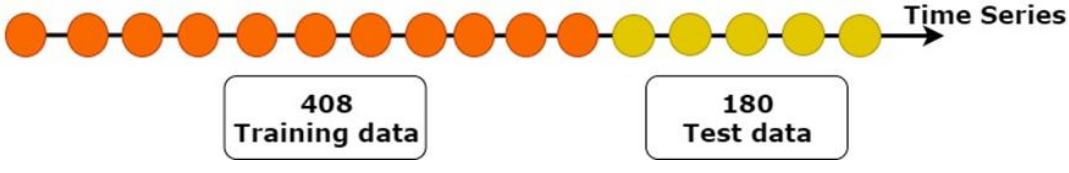

Figure 2 - Example of representation of the training and test sets.

HOLT-WINTERS

Holt-Winters (H-W) model is a statistical method used to estimate future values of time series data, being able to deal with seasonal trends and changes. The theory of the method is described in, Morettin and Toloi (2006). The objective is to capture the patterns of the time series, separating it into trend, seasonality, and error. Prediction equations are allocated in two ways: additive or multiplicative, according to the essence of the series. The additive H-W model was used in the time-series data of this work since the multiplicative prediction

produced extremely low values. Therefore, the additive model was adopted in this study applying the following equation:

$$Y_t = N_t + T_t + S_t + \varepsilon_{t,} \qquad t = 1, \ldots\ldots, n \qquad (1)$$

where: $\mathbf{N_t}$: is level of the series, shows how the expected time series evolves; $\mathbf{T_t}$: is the tendency, this relates to the fact that the predicted time series can have increased or decreasing motions in different time intervals; $\mathbf{S_t}$: is a seasonal component, which is related to the fact that the expected time series has cyclical patterns of variation that repeat at relatively constant time intervals. Thus, the additive form of the Holt-Winters algorithm is presented in Equations 2,3, and 4.

$$N_t = \alpha(y_t - S_t\text{-}s) + (1\text{-}\alpha)(N_t\text{-}1 + T_t\text{-}1), \qquad 0 \leq \alpha \leq 1 \qquad (2)$$

$$T_t = \beta\,(N_t - N_t\text{-}1) + (1\text{-}\beta)\,T_t\text{-}1, \qquad 0 \leq \beta \leq 1 \qquad (3)$$

$$S_t = \delta(Y_t - N_t) + (1\text{-}\delta)\,S_t\text{-}s, \qquad 0 \leq \delta \leq 1 \qquad (4)$$

Where: N_t, T_t, and S_t denote the time-adjusted values "t"; S is the number of times that the series is observed per year, and α, β, δ are straightening constants

SEASONAL ARIMA

In hydrological studies, it is common to find seasonality in the series. This brings us to the need to use models that incorporate this important temporal resource. The Seasonal Arima model is a seasonal autoregressive integrated moving average (SARIMA) model used to assist in seasonality (BOX et al., 2008). The SARIMA model is represented by (p, d, q) x (P, D, Q)z where "p" represents the non-seasonal autoregressive term, "q" represents the non-seasonal moving average term, "d" represents the non-seasonal differencing terms and (P, D, Q)z represents the seasonal autoregressive, seasonal moving average and seasonal difference terms respectively. SARIMA model was defined at Equation (5) (BOX et al. 2008, WIBOWO et al., 2017)

$$\phi_{NAR}(B)\,\phi_{SAR}(B^z)(1\text{-}B)^d\,(1\text{-}B^z)^D\,S_t = \theta_{NMA}(B)\,\theta_{SMA}\,(B)\varepsilon_t \qquad (5)$$

Where z = seasonality lag, ϕ_{NAR} = non-seasonal autoregressive parameter, ϕ_{SAR} = seasonal autoregressive parameter, θ_{NMA} = non-seasonal moving average parameter, θ_{SMA} = seasonal moving average parameter, D = seasonal difference, d = non-seasonal difference, S_t = Streamflow at time t, εt is white noise.

NEURAL NETWORK AUTOREGRESSIVE

The ANN is the most widely used technique in the field of machine learning, created from the inspiration of the structure of the human brain, therefore composed of a certain number of neurons. Thus the flow data that are inserted in the ANN, enter through an input layer (predictors) and are

subsequently processed through different intermediate layers, producing a result in the output layer (predictions) (HAYKIN, 2001). In recent decades, many studies have shown that ANNs are adequate to solve hydrology-related problems (ZHANG & GOVINDARAJU, 2000; JAIN & SRINIVASULU, 2004; CHANG et al. 2010; DARIANE & KARAMI, 2014). Because in their structure can have both nonlinear and linear models, presenting improved results when compared with other prediction models (KHASHEI, & BIJARI, 2010). When dealing with time-series data, lagged time series values can be used as input to a neural network by calling it a neural network autoregressive (NNAR) model. In this article we considered feedforward networks with a hidden layer using notation NNAR (p, P,k)s in which indicates that p represents late entries (y_{t-1}, y_{t-2},..., y_{t-p}, y_{t-s}, y_{t-2s}, y_{t-Ps}) and k neurons in the hidden layer (HYNDMAN & ATHANASOPOULOS, 2013).

RECURRENT NEURAL NETWORKS

Recurrent neural networks (RNN) was first developed in the 1980s. RNN is a dynamic, deep-learning neural network architecture that specializes specifically for interdependent data, such as prediction of complex hydrological and meteorological time series (COULIBALY & BALDWIN, 2005, KUMAR et al., 2004). RNNs have their architecture, an input layer, one or more hidden layers composed of recurrently connected nodes, and an output layer. Note that on all RNNs layers, they will share their parameter values over the entire sequence time interval, this feature allows scaling to longer or shorter sequences. So the main difference between RNNs and NNAR is that RNNs have connections that allow feedback on the architecture to "remember" previous information (WERBOS, 1988; ELMAN, 1990).

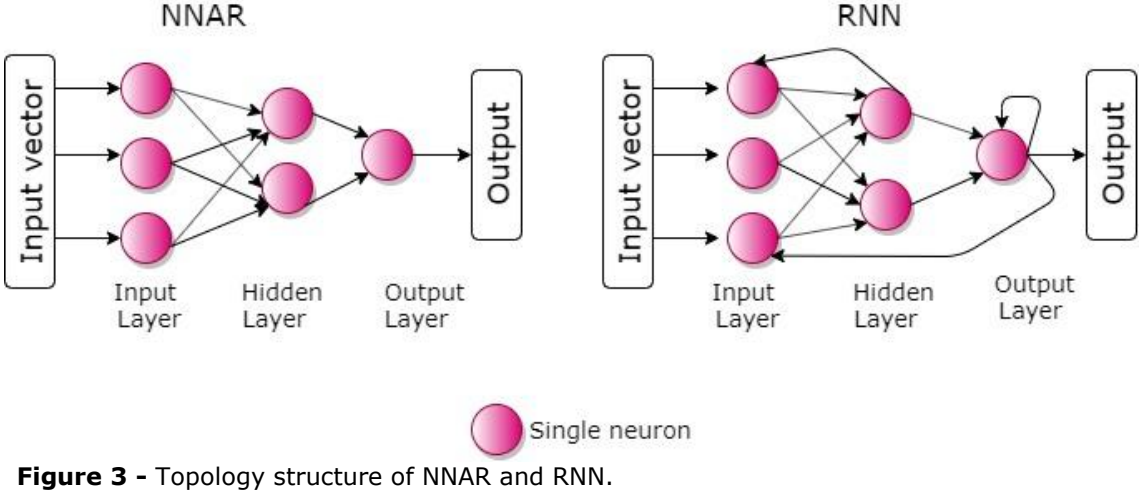

Figure 3 - Topology structure of NNAR and RNN.

2.3 PERFORMANCE METRICS FOR EVALUATION

To classify the performance of the proposed models (Table 1) were used the main techniques adopted and recommended by several authors (AHMADI et al., 2014, DAGGUPATI et al., 2015, MORIASI et al., 2007), which include the coefficients of Nash-Sutcliffe (NSE), percentage bias (PBIAS), determination

coefficients (R^2), mean absolute error (MAE), mean root square error (RMSE), and index the Willmott (d). These are calculated as follows:

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - t_i| \quad (1)$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - t_i)^2} \quad (2)$$

$$PBIAS = \left(\frac{\sum_{i=1}^{n} y_i - \sum_{i=1}^{n} t_i}{\sum_{i=1}^{n} t_i}\right) \times 100 \quad (3)$$

$$NSE = 1 - \frac{\sum_{i=1}^{n}(y_i - t_i)^2}{\sum_{i=1}^{n}(y_i - \bar{t})^2} \quad (4)$$

$$R^2 = \left(\frac{\sum_{i=1}^{n}(y_i - \bar{y})(t_i - \bar{t})}{\sum_{i=1}^{n}[(y_i - \bar{y})^2]^{0.5}\ \sum_{i=1}^{n}[(t_i - \bar{t})^2]^{0.5}}\right)^2 \quad (5)$$

$$d = 1 - \frac{\sum_{i=1}^{n}(t_i - y_i)^2}{\sum_{i=1}^{n}(|t_i - \bar{t}| + |y_i - \bar{t}|)^2} \quad (6)$$

Where, y_i is observed flow in time i, t_i is Simulated flow in time i; $\bar{y}$ is the average observed flow; $\bar{t}$ is average simulated flow and n is the number of observations.

Table 1 - Forecast quality metrics used in the present study.

Abbreviated name	Full name	Values	Optimum value	Condition (preferred values)
MAE	Mean absolute error	[0, + ∞)	0	Smaller MAE
RMSE	Root mean square error	[0, + ∞)	0	Smaller RMSE
PBIAS	Percent bias	(− ∞, + ∞)	0	Smaller |PBIAS|
NSE	Nash–Sutcliffe efficiency	(- ∞, 1]	1	Larger NSE
R²	Coefficient of determination	[0, 1]	1	Larger R²
d	Index the Willmott	[0, 1]	1	Larger d

Mean Absolute Error (MAE) provides an easily interpretable assessment, as it measures the average magnitude of errors in a set of predictions (HYNDMAN & KOEHLER 2006). The RMSE is one of the most commonly used performance measures in hydrological modeling, is a quadratic scoring rule that also measures the average magnitude of the error. It is the square root of the mean of the quadratic differences between prediction and actual observation (HYNDMAN & KOEHLER 2006). PBIAS, which measures the average tendency of

ISSN: 2237-8642 (Eletrônica)

simulated data to be higher or lower than observed values, is a potential metric for hydrology scientists (GUPTA et al., 2009).

The NSE measures the strength of the relationship between the variance of the residuals and the variance of the observed values (NASH & SUTCLIFFE, 1970; MORIASI et al., 2007). Pearson's product-moment correlation square is indicated as R^2 where it represents the proportion of the total variance of the measured data that can be explained by the simulated data, the higher the values, close to 1, represent better model performance. The index designated concordance, represented by the letter "d", related to the deviation of the estimated values from those observed (WILLMOTT et al, 1985). Its values range from 0 for no concordance to 1 for perfect concordance.

3. RESULTS AND DISCUSSION

3.1 DESCRIPTIVE ANALYSIS

The main task in dealing with time series is to define a model capable of representing the process involved (SIQUEIRA, 2013). Thus, to analyze a certain historical series of river basin fluviometric data, it is necessary for the principle to perform an exploration of the data, since the flow of a given water body has as its basic characteristic a great temporal and spatial variability. Note that Figure 4 shows the behavior for the entire period studied (1965 - 2013), and Figure 5 shows a well-defined seasonal pattern of monthly flow (May to August).

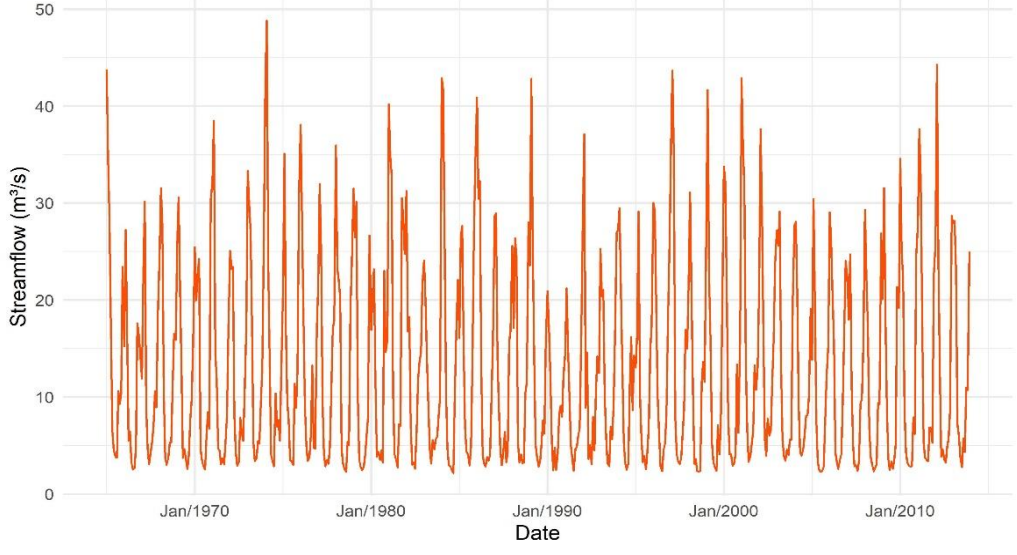

Figure 4 - Observed streamflow series (m³/s) from January 1965 to December 2013 in the Sibinacocha watershed.

The mean flow rate for the analyzed period was 12.87 m³/s with a standard deviation of 10.42 m³/s. The maximum monthly average flow rate was 48.81 m³/s in February 1974, while the minimum monthly average flow rate was 2.16 m³/s in September 1984. The high coefficient of variation (above 80%) indicates the high degree of variability of the observed series, corroborating the seasonality observed in Figures 4 and 5.

ISSN: 2237-8642 (Eletrônica)

In the Peruvian Andes, during the year there are two well-differentiated periods, the dry and wet period. Since the wet season begins in the first precipitation in September, increasing between December and March, and subsequently decreasing from April. Thus, the months from May to August are the dry season (LAGOS et al., 2008). In this sense, according to Figure 5, it can be observed that the " wet " period is marked by the months with the highest flow, and the " dry " period as the periods with the lowest flow.

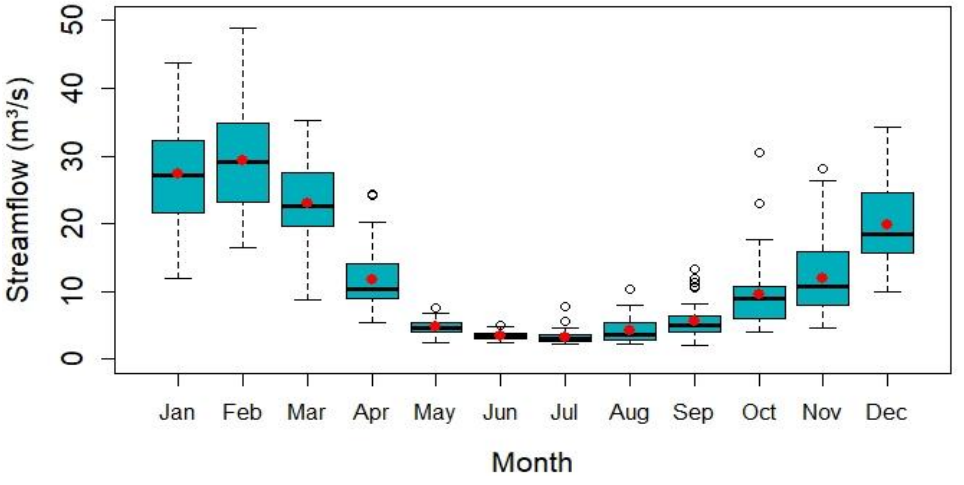

Figure 5 - Seasonal streamflow series (m³/s) on the Sibinacocha lake watershed, between the years 1965-2013.

Trend analyses were performed for the full 48-year of data, both for the dry and wet periods. As demonstrated modified Mann Kendall's test showed values of p-value = 0.38, p-value = 0.23, p-value = 0.49 respectively, which indicates that the monthly flow data do not present a significant trend for the periods under analysis.

Analysis of variance of repeated measurements was executed to evaluate the monthly means of flow before and after the dam construction. The results indicated that there was no significant statistical difference (p-value = 0.367) between the monthly flow means (Figure 6) for the dry season. On the contrary, a significant statistical difference (p=0.02) of the average monthly flow during the wet season was observed, which the average value was 17.8 m³/s before the dam construction, changing to 16.8 m³/s after the hydraulic work. (Figure 7).

ISSN: 2237-8642 (Eletrônica)

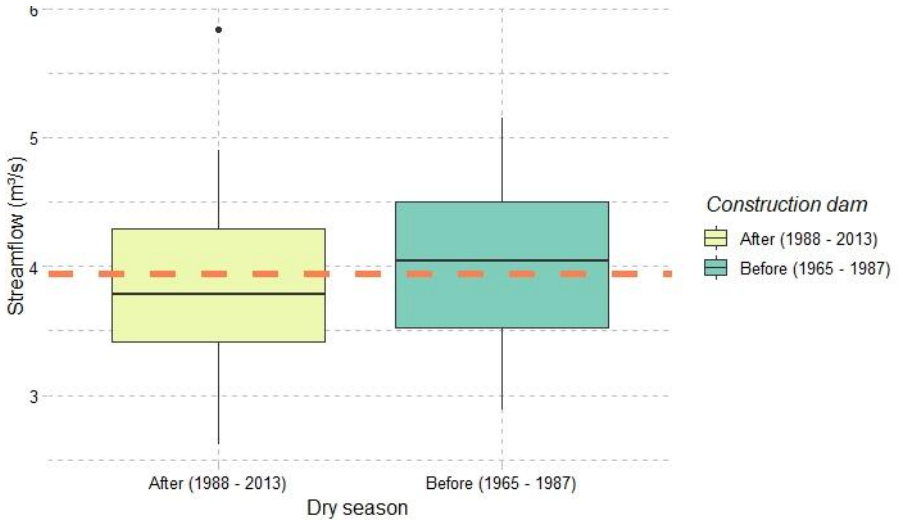

Figure 6 - Dry season flow (m³/s) in the Sibinacocha lake watershed, before and after the construction of the dam, the dashed line represents the average dry season for the period.

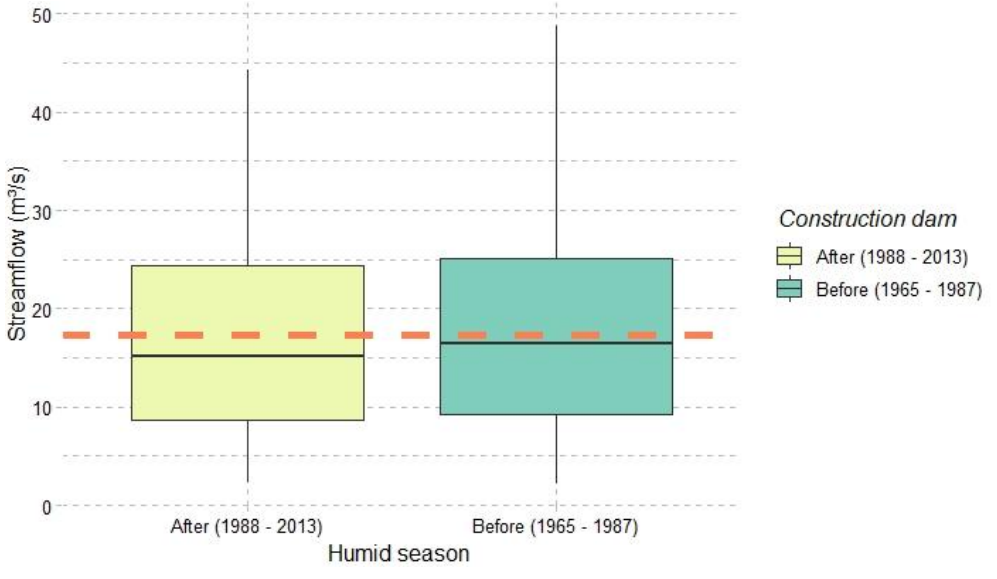

Figure 7 - Wet season flow (m³/s) in the Sibinacocha lake watershed, before and after the construction of the dam, the dashed line represents the average wet season for the period.

Figure 8 shows the classical additive decomposition of the flow series (seasonality, trend, and random components) corroborating that the series has seasonality and there is no evidence of systematic increase or decrease of the variable over time.

ISSN: 2237-8642 (Eletrônica)

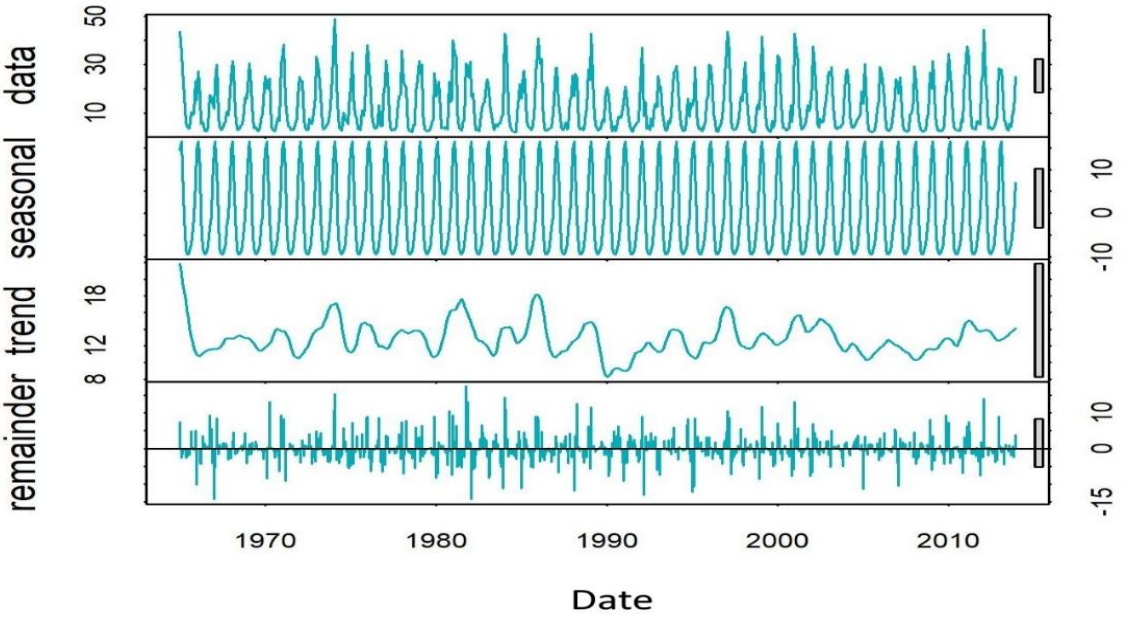

Figure 8 - Decomposition of additive time series of streamflow in seasonality, trend, and randomness components.

3.2 FORECASTING STREAMFLOW

The results of the calculation of the monthly flow estimates in the four models used represented the monthly pattern of the data observed for the Sibinacocha Lake watershed (Figure 9).

The Holt-Winters model, it presented good evaluation coefficients except for PBIAS (Table 2 and Figure 10). Also, it is observed in Figure 7, that the model presented much lower predictions of the observed minimum flow values. Regarding the Sarima model, although it obtained acceptable evaluation metrics (Table 2 and Figure 10), the model was not able to provide good predictions of monthly flows, since the Sarima model could not reach the maximum observed flow values, besides that overestimates the minimum values of monthly flow.

For the NNAR model, this type of modelling provided acceptable metrics, although a lower correlation coefficient when compared to the statistical models presented (Figure 10). It is observed that the NNAR model pointed to an improvement in the flow predictions; however, it also failed to capture the maximum values of the monthly flow. Regarding the modeling via recurrent neural networks (RNNs), it showed a better monthly flow pattern, when compared to the other models applied in this work, besides presenting good evaluation metrics, this type of neural network was able to capture the maximum and minimum values of monthly outflow over time.

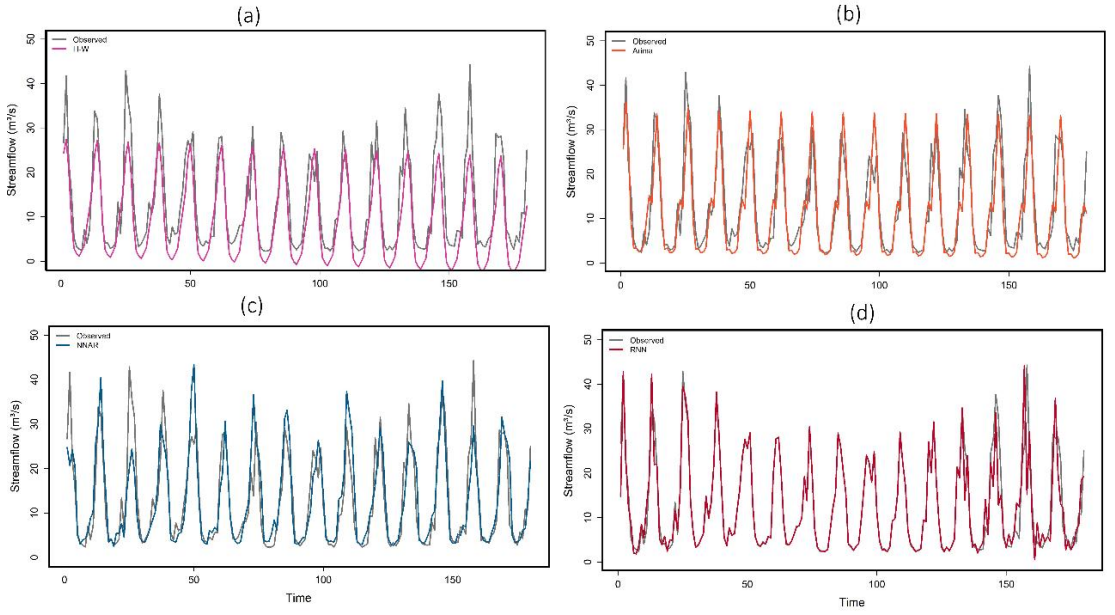

Figure 9 - Streamflow Forecasting (a) Observed and Holts-Winters, (b) Observed and Arima, (c) Observed and NNAR, (d) Observed and RNN.

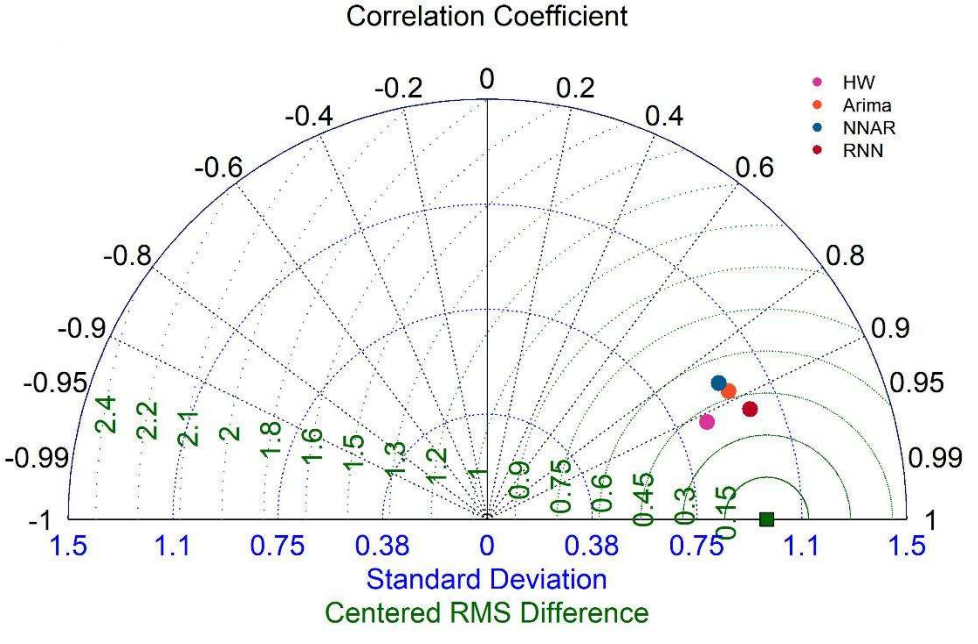

Figure 10 - Taylor Diagram summarizes model validation information (Holt-winters, Seasonal Arima, NNAR, RNN). The radial distance from the origin represents the standard deviation of the observed flow. The green lines represent the RMS error centered on the simulated field. The correlation coefficient is represented by the azimuth position of the points.

Table 2 - Results of model evaluation metrics.

Performance Metrics	Stochastic models		Machine Learning	
	Holt-Winters	Seasonal Arima	NNAR	RNN
MAE	4.48	3.71	3.49	**1.7**
RMSE	5.76	5.15	5.41	**4.05**
PBIAS	-30.2	-9.4	1.5	**-2.5**
NSE	0.7	0.76	0.73	**0.85**
R^2	0.84	0.78	0.74	**0.85**
d	0.91	0.94	0.93	**0.96**

Studies applied in other river basins have shown that data-driven models can be relatively straightforward, providing good predictive information, eliminating some shortcomings of other model types (HAN et al., 2007, MASSELOT et al., 2016, NOORI & KALIN., 2016). In this sense, based on the results obtained so far, it can be confirmed that neural networks have the potential to be a useful tool for flow prediction, being considered as alternatives to traditional approaches.

4. CONCLUSION

In this work, the monthly flow time series provided by the National Water Agency of Peru was used, where it was possible to analyze the monthly flow trends for the years analyzed. In addition, it was also observed that the construction of the dam, maintains the flow during the dry season and helps to control floods during the wet season, thus confirming the importance of the construction of the hydraulic works.Then, different methodologies were applied to evaluate their performance in calculating the monthly flow simulation. Therefore, it was demonstrated the high performance of Recurrent Neural Networks in comparison with the other models presented here, thus showing the RNN as a promising and rapid model for flow simulation with the possibility of being applied in other basins. However, it is advisable to use meteorological variables together with the observed flow data to further improve the results of the estimates. This research will continue with the application of a physics-based model for the simulation of the flow in Sibinacocha Lake Watershed, illustrating the hydrological cycle of this basin, and contributing new insights for the region.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the financial support to the first author granted by the CAPES 001 funding code and thank the reviewers for their contribution.

REFERENCES

AFAN, H. A.; EL-SHAFIE, A.; YASEEN, Z. M.; HAMEED, M. M.; MOHTAR, W. H. M. W.; HUSSAIN, A. ANN based sediment prediction model utilizing different input scenarios. Water resources management, v. 29, n.4, p.1231-1245, 2015.

AHMADI, M.; ARABI, M.; ASCOUGH II, J. C.; FONTANE, D. G.; ENGEL, B. A. Toward improved calibration of watershed models: Multisite multiobjective measures of information. Environmental Modelling & Software, v.59, p.135-145, 2014.

BAYER, F. M.; SOUZA, A. M. Wavelets e modelos tradicionais de previsão: Um estudo comparativo. Revista Brasileira de Biometria. v.28, p, 40-61, 2010.

BOX, G.; JENKINS, G. M.; REINSEL, G. Time series analysis: forecasting and control. 4.ed. Hoboken, New Jersey: John Wiley & Sons, 2008, 746p.

CATACORA, E. A. Predicciones del comportamiento de caudales de la CH Machupicchu mediante análisis arima de series temporales. Tese (Maestría en Ciencias con Mención en Energética)- Universidad Nacional de Ingeniería, Lima, 2008.

CHANG, L.C; Shen, HY; Wang, Y.F; Huang, J.Y; Lin Y.T. Clustering-based hybrid inundation model for forecasting flood inundation depths. J Hydrol 385, p, 257–268, 2010.

CHATFIELD, C.; YAR, M. Holt-winters forecasting: some practical issues. Journal of the Royal Statistical Society: Series D, v.37, p,129-140, 1988.

COULIBALY. P.; BALDWIN. C.K. Nonstationary hydrological time series forecasting using nonlinear dynamic methods Journal of Hydrology, v.307, p,1–4, 2005.

DAGGUPATI, P.; YEN, H.; WHITE, M. J.; SRINIVASAN, R.; ARNOLD, J. G.; KEITZER, C. S.; SOWA, S. P. Impact of model development, calibration and validation decisions on hydrological simulations in West Lake Erie Basin. Hydrological Processes, v. 29, n.26, p.5307-5320, 2015.

DARIANE, A.B; KARAMI, F. Deriving hedging rules of multi-reservoir system by online evolving neural networks. Water Resour Manag v.28, n.11, p, 3651–3665, 2014.

DICKEY, B. Y. DAVID A.; FULLER, 1981 Wayne A. Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root. Econometrica, v. 49, p, 1057–1072, 1981.

DRENKHAN, F.; HUGGEL, C.; GUARDAMINO, L.; HAEBERLI, W. Managing risks and future options from new lakes in the deglaciating Andes of Peru: The example of the Vilcanota-Urubamba basin. Science of the Total Environment, v. 665, n. 465-483, 2019.

ELMAN, J. L. Finding structure in time. Cognitive science, v. 14, n. 2, p. 179-211, 1990.

FARFÁN, J. F.; PALACIOS, K.; ULLOA, J.; AVILÉS, A. A hybrid neural network-based technique to improve the flow forecasting of physical and data-driven models: Methodology and case studies in Andean watersheds. Journal of Hydrology: Regional Studies, v. 27, p. 100652, 2020.

FATHIAN, F.; MEHDIZADEH, S.; SALES, A. K.; SAFARI, M. J. S. Hybrid models to improve the monthly river flow prediction: Integrating artificial intelligence and non-linear time series models. Journal of Hydrology, v. 575, p. 1200-1213, 2019.

GUPTA, H.V.; KLING, H.; YILMAZ, K.K. MARTINEZ, G.F. Decomposition of the mean squared error and NSE performance criteria: implications for improving hydrological modelling. J. Hydrol 377(1–2), p,80–91, 2009.

HAMED, K. H.; RAO, A. R A modified Mann-Kendall trend test for auto correlated data. Journal of Hydrology, v. 204, p.182-196, 1998.

HAN, T.; KWONG, S. LI Uncertainties in real-time flood forecasting with neural networks Hydrol. Process., v. 2 n.21, p, 223-228,2007.

HAND, D. J.; TAYLOR, C. C. Multivariate analysis of variance and repeated measures: a practical approach for behavioural scientists (Vol. 5). CRC press. London, 1987. 225 p.

HAYKIN, S. Redes neurais – princípios e prática. Trad. Paulo Martins Engel. Porto Alegre: Bookman, 2001. 891 p.

HYNDMAN, R.J.; ATHANASOPOULOS, G. Forecasting: principles and practice. OTexts: 2013. 377p.

HYNDMAN, R.J.; KOEHLER, A.B. Another look at measures of forecast accuracy. Int J Forecast v. 4, n. 22, p, 679–688, 2016.

JAIN. A.; SRINIVASULU, D. Development of effective and efficient rainfall–runoff models using integration of deterministic, real-coded genetic algorithms, and artificial neural network techniques Water Resour. Res., v. 40, n.4, 2004.

KHASHEI, M.; BIJARI, M. An artificial neural network (p, d, q) model for timeseries forecasting. Expert Systems with applications, v. 37, n.1, p, 479-489, 2010.

KRONENBERG, M., SCHAUWECKER, S., HUGGEL, CH., SALZMANN, N., DRENKHAN, F., GIRAÁLDEZ, C., GURGISER, W., KASER, G., SUAREZ, W., GARCÍA, J., ROHRER, M. "¿Cuál Es El Futuro Del Caudal En Las Cuencas Glaciadas de Los Andes Centrales?" In , 1364–73. Lima, Perú: XXVII Congreso Latinoamericano de Hidráulica. 2016.

KUMAR, D.N.; RAJU, K.S.T. Sathish River flow forecasting using recurrent neural networks Water resources management, v.18, n. 2, p, 143-161, 2004.

LAGOS, P.; SILVA, Y.; NICKL, E.; MOSQUERA, K. El Niño? Related Precipitation Variability in Perú. Advances in Geosciences, 2008.

LEHNER, F.; WOOD, AW.; LLEWELLYN, D.; BLATCHFORD, DB.; GOODBODY, AG.; PAPPENBERGER, F. Atenuar os impactos da não estacionariedade climática na previsibilidade sazonal de vazões no sudoeste dos EUA. Geophysical Research Letters, v. 44 , n.24, p.12-208, 2017.

MASSELOT, P.; DABO-NIANG, S.; CHEBANA, F.; Ouarda, T.B. Streamflow forecasting using functional regression J. Hydrol., v. 538, p, 754-766, 2016.

MONTANARI, A.; ROSSO, R.; TAQQU, M.S. A seasonal fractional ARIMA model applied to the Nile River monthly flows at Aswan. Water Res Res v. 36, n. 5, p,1249–1259, 2000.

MEHDIZADEH, S.; FATHIAN, F.; ADAMOWSKI, J. F. Hybrid artificial intelligence-time series models for monthly streamflow modeling. Applied Soft Computing, v.80, p. 873-887, 2019.

MORETTIN, P. A.; TOLOI, C. M. C. Análise de Séries Temporais. 2. ed. São Paulo: Edgard Blücher, 2006, 350p.

MOHAMMADI, K., ESLAMI, H. R., & KAHAWITA, R. Parameter estimation of an ARMA model for river flow forecasting using goal programming. Journal of Hydrology, v.331, n.1-2, p. 293-299, 2006.

MORIASI, D. N.; ARNOLD, J. G.; LIEW, M. W. VAN; BINGNER, R. L.; HARMEL, R. D.; VEITH, T. L. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. Transactions of American Society of Agricultural and Biological Engineers, v.50, n.3, p, 885-900, 2007.

MORTATTI, J.; BORTOLETTO JUNIOR, M. J.; MILDE, L. C. E.; PROBST, J. L. Hidrologia dos rios Tietê e Piracicaba: séries temporais de vazão e hidrogramas de cheia. Revista de Ciência & Tecnologia, v.12, n.23, p, 55-67, 2004.

NASH, J. E.; SUTCLIFFE, J. V. River flow forecasting through conceptual models part 1 – A discussion of principles. Journal of Hydrology, v.10, n.3, p, 282-290, 1970.

NIU, W. J.; FENG, Z. K.; CHEN, Y. B.; ZHANG, H. R.; CHENG, C. T. Annual streamflow time series prediction using extreme learning machine based on gravitational search algorithm and variational mode decomposition. Journal of Hydrologic Engineering, v. 25, n. 5, p. 04020008, 2020.

NOORI, N.; KALIN, L.Coupling SWAT and ANN models for enhanced daily streamflow prediction J. Hydrol., v.533, p,141-151, 2016.

PERRY, L.B.; SEIMON, A.; KELLY, G. M. Precipitation delivery in the tropical high Andes of southern Peru: new findings and paleoclimatic implications, International journal of Climatology, v. 34, p. 197-215, 2013.

RUBEL, F.; KOTTEK, M. Comments on:"The thermal zones of the Earth" by Wladimir Köppen (1884). Meteorologische Zeitschrift, v. 20, n. 3, p. 361-365, 2011.

SALZMANN, N.; HUGGEL, C.; ROHRER, M.; SILVERIO, W.; MARK, B.; BURNS, P.; PORTOCARRERO, C. Glacier Changes and Climate Trends Derived from Multiple Sources in the Data Scarce Cordillera Vilcanota Region, Southern Peruvian Andes. The Cryosphere, v. 7, p, 103-118, 2013.

SHAHWAN, T.; ODENING, M. Forecasting agricultural commodity prices using hybrid neural networks. In Computational intelligence in economics and finance, p, 63-74, 2007..

SIQUEIRA, H. V. Máquinas desorganizadas para previsão de séries de vazões. Tese (Doutorado em Engenharia Elétrica) — Universidade Estadual de Campinas, Campinas, 2013, p. 218

TIKHAMARINE, Y.; SOUAG-GAMANE, D.; AHMED, A. N.; KISI, O.; EL-SHAFIE, A. Improving artificial intelligence models accuracy for monthly streamflow forecasting using grey Wolf optimization (GWO) algorithm. Journal of Hydrology, v. 582, p. 124435, 2020.

WAGENA, M. B.; GOERING, D.; COLLICK, A. S.; BOCK, E.; FUKA, D. R.; BUDA, A.; EASTON, Z. M. Comparison of Short-Term Streamflow Forecasting using Stochastic Time Series, Neural Networks, Process-Based, and Bayesian Models. Environmental Modelling & Software, p. 104669, 2020.

WERBOS, P.J. Generalization of backpropagation with application to a recurrent gas market model. Neural Netw. v.1, p, 339–356, 1988.

WIBOWO, W.; DWIJANTARI, S.;HARTATI, A. Time Series Machine Learning: Implementing ARIMA and Hybrid ARIMA-ANN for Electricity Forecasting Modeling. In International Conference on Soft Computing in Data Science. P, 126-139. Springer, Singapore. 2017.

WILLMOTT, C. J.; ROWE, C. M.; MINTZ, Y. Climatology of the terrestrial seasonal water cycle. Journal of Climatology, v. 5, n. 6, p. 589-606, 1985.

WU, C. L.; CHAU, K. W. Data-driven models for monthly streamflow time series prediction. Engineering Applications of Artificial Intelligence, v. 23, n.8, p. 1350-1367, 2010.

ZHANG, B.; GOVINDARAJU, S. Prediction of watershed runoff using bayesian concepts and modular neural networks Water Resour. Res., v. 36, n. 3, p, 753-762, 2000.

ZHANG, W; YAN, Y; ZHENG, J; LI, L; DONG, X; CAI, H. Temporal and spatial variability of annual extreme water level in the Pearl River Delta region, China. Global and Planetary Change v. 69, p, 35-47, 2009.

ZHAO, T.;ZHAO, J. Joint and respective effects of long-and short-term forecast uncertainties on reservoir operations. Journal of hydrology, v. 517, p, 83-94, 2014