

Comparison of methodologies for flow prediction through teleconnection indices. Case study: Paute river basin

Comparación de metodologías para la predicción del caudal mediante los índices de teleconexión. Caso de estudio: Cuenca del río Paute

María Daniela González¹ <https://orcid.org/0009-0003-3436-0805>, Darío Xavier Zhiña^{1,2} <https://orcid.org/0000-0001-9556-4025>, Alexandra Guanuchi-Quito^{1,2} <https://orcid.org/0000-0002-5583-8674>, Alex Avilés-Añazco^{1,2} <https://orcid.org/0000-0001-9278-5738>

¹Facultad de Ciencias Químicas, Universidad de Cuenca, Cuenca, Ecuador
daniela.gonzalez98@ucuenca.edu.ec

²Grupo de Evaluación de riesgos ambientales en sistemas de producción y servicios (RISKEN), Universidad de Cuenca, Cuenca, Ecuador
dario.zhina0608@ucuenca.edu.ec,
alexandra.guanuchi@ucuenca.edu.ec,
alex.aviles@ucuenca.edu.ec



Esta obra está bajo una licencia internacional
Creative Commons Atribución-NoComercial 4.0

Sent: 2023/06/24

Accepted: 2023/07/07

Published: 2023/10/15

Abstract

The Paute river basin (southern Ecuador) suffers hydrological changes due to climate change and human activities. Hydrological changes cause extreme events and affect ecosystems, hydroelectric plants, and quality of life. It highlights the importance of understanding hydrological behavior to make appropriate decisions in extreme environments. This study seeks to predict discharges in the Paute river basin through global teleconnection indices. Multiple Linear Regression (MLR) was obtained using three different methodologies: multicollinearity analysis, Principal Component Analysis (PCA), and correlation with monthly delays. It was shown that the principal component analysis scenario obtained the best predictive fits, specifically by including 41 indices and 20 components. For the scenario using monthly delays, the best delay occurs within a single month for most seasons. Finally, with the multicollinearity analysis scenario, better results were obtained using 41 indices, although essentially the performance corresponds to the number and

Summary: Introduction, Materials y Methods, Results and Discussion and Conclusions.

How to cite: González, M., Zhiña, D., Guanuchi-Quito, A., & Avilés-Añazco, A. (2023). Comparison of methodologies for flow prediction through teleconnection indices. Case study: Paute river basin. *Revista Tecnológica - Espol*, 35(2), 27-49.

Recuperado a partir de <http://www.rte.espol.edu.ec/index.php/tecnologica/article/view/1028>

indices of each model. Teleconnection indices are not sufficient when used as the only input variable for download modeling and prediction, giving mostly unsatisfactory results. However, a clear trend links the behavior of flows and indices, and it is possible to improve the models based on more climatic variables or with other predictive methods.

Keywords: Discharge prediction. Teleconnection indices. Principal component analysis. Multiple regression models. Multicollinearity analysis.

Resumen

La cuenca del río Paute (sur del Ecuador) sufre cambios hidrológicos por el cambio climático y las actividades humanas. Los cambios hidrológicos causan eventos extremos y afectan a ecosistemas, centrales hidroeléctricas y la calidad de vida. Destaca la importancia de comprender el comportamiento hidrológico para tomar decisiones adecuadas en ambientes extremos. Este estudio busca predecir las descargas en la cuenca del río Paute mediante los índices de teleconexión global. Se obtuvieron modelos de Regresión Lineal Múltiple (MLR) mediante tres metodologías diferentes: análisis de multicolinealidad, Análisis de Componentes Principales (ACP) y correlación con retrasos mensuales. Se demostró que el escenario de análisis de componentes principales obtuvo los mejores ajustes predictivos, específicamente al incluir 41 índices y 20 componentes. Para el escenario que usa retrasos mensuales, el mejor retraso ocurre dentro de un solo mes, para la mayoría de las estaciones. Finalmente, en el escenario de análisis de multicolinealidad se obtuvieron mejores resultados utilizando 41 índices, aunque esencialmente el rendimiento corresponde a la cantidad y los índices de cada modelo. Los índices de teleconexión no son suficientes cuando se utilizan como la única variable de entrada para el modelado y la predicción de descargas, dando resultados en su mayoría insatisfactorios. Sin embargo, existe una clara tendencia que vincula el comportamiento de caudales e índices, y es posible mejorar los modelos en base a más variables climáticas o con otros métodos predictivos.

Palabras clave: Bases de datos, predicción de caudales, Índices de Teleconexión, Análisis de Componentes Principales, Modelos de Regresión, Análisis de Multicolinealidad.

Introduction

Within studies related to global teleconnections between ENSO (El Niño South Oscillation) and discharge, strong and regionally consistent discharge impacts were found in Central and South America, New Zealand, and Australia, while weaker signals were observed in parts of Africa and North America (Kundzewicz, Szwed, & Pińskwar, 2019). To understand more the localized effects of teleconnections, a large number of studies are currently focused on connecting flooding with climatic variability on a continental scale. Many relevant studies have been carried out in Australia, Asia, Europe, North America, and South America. These studies utilize indices such as ENSO, NAO (North Atlantic Oscillation), AMO (Atlantic Multidecadal Oscillation), TSA (Tropical Southern Atlantic), TNA (Tropical Northern Atlantic), AO (Arctic Oscillation), and PDO (Pacific Decadal Oscillation), among others (Giddings and Soto, 2006).

In a local context, the most relevant index is ENSO due to its effects. In South America, ENSO causes floods and droughts along the western coast. However, simulating ENSO in the region faces biases and uncertainties, especially when dealing with long time series. Systematic errors occur in the central equatorial Pacific, the eastern equatorial Indian Ocean, and regions with boundary current systems (such as the tropical Pacific and Atlantic) (Cai et al., 2020).

Within the confines of the study area and its regional dynamics, some studies have

been conducted in Ecuador, demonstrating that the Sea Surface Temperature (SST) variability and ENSO phenomena have an impact on discharge patterns throughout the nation. Similarly, multiple investigations have established that discharge anomalies can be found nationwide using ENSO and its modes. (Córdoba Machado et al., 2015). The research holds particular significance due to its revelation that the three most common natural catastrophes in Ecuador are floods, droughts, and landslides, with the first two having a significant impact on local living conditions (Fontaine et al., 2008).

To comprehend the impacts of teleconnections on climate, it is imperative to situate these events within a climatological context. Teleconnection indices are statistically significant correlations of recurrent atmospheric anomalies that occur in nearby and distant areas, often concurrently, at the planetary or hemispheric level (Hatzaki et al., 2007). In line with this methodology, while examining the sequence of oscillations that constitute climate variability around the mean values, these anomalies are recognized by detecting deviations. This representation exemplifies the climate's ephemeral condition in response to modifications and is demarcated by precise temporal and spatial scales (IDEAM - UNAL, 2018). Additionally, the teleconnection fluctuations encompass anomalies within their cycles, generating variability modes. These modes are determined from sub-calculations or temporal variations in the behavior of teleconnection indices (Dima and Lohmann, 2004).

Broadly speaking, focusing on Ecuador and the Andes, ENSO is the primary factor influencing SST and air pressure. El Niño (warm phase) and La Niña (cold phase) patterns constitute this phenomenon. Another factor is the PDO, which exhibits warm and cold interdecadal phases that affect the surface waters of the Pacific Ocean. Recent investigations have identified that the impact of PDO in South America has increased due to climate change (Morán et al. 2016). Regarding precipitation, indicators related to ENSO (such as Niño 3.4 and SOI), can affect rainfall in Ecuador; also, Niño 3.4 and SST exhibit a significant causal relationship. The only significant impact that NP and WP have on rainfall in South America is in Brazil. Due to spatial distance, AO and AAO have a minimal influence. Widespread droughts have been related to TNA and TSA. However, droughts or excessive rains in Ecuador or the region show no significant connection to the NAO (Giddings and Soto, 2006).

Due to its ecological and social relevance, numerous hydrological studies have focused on regional factors within the Paute river basin in southern Ecuador to comprehend patterns in sub-basin and micro-basin discharge (Celleri et al., 2007; Sotomayor et al., 2018; Ward et al., 2011). However, considering the recent understanding of these teleconnection indices, the use of these to forecast water behavior and correlate it with extreme hydrological events represents a novel exploratory strategy. These indices operate on a global scale and might exert a greater influence on hydrological dynamics than local elements do.

As previously mentioned, the significance of this study stems from the fact that substantial ecological and human losses occur in the area, particularly due to floods in the Sierra region induced by rivers within the Paute River basin. For that reason, the purpose of the project is to use teleconnection indices and variability modes to forecast the flow in the Paute river basin. Principal component analysis (PCA), correlation with monthly delay, and MLR with multicollinearity analysis are all part of the process of forecasting. The levels of prediction for each methodology will then be evaluated using statistical measures by contrasting the outcomes with the original data.

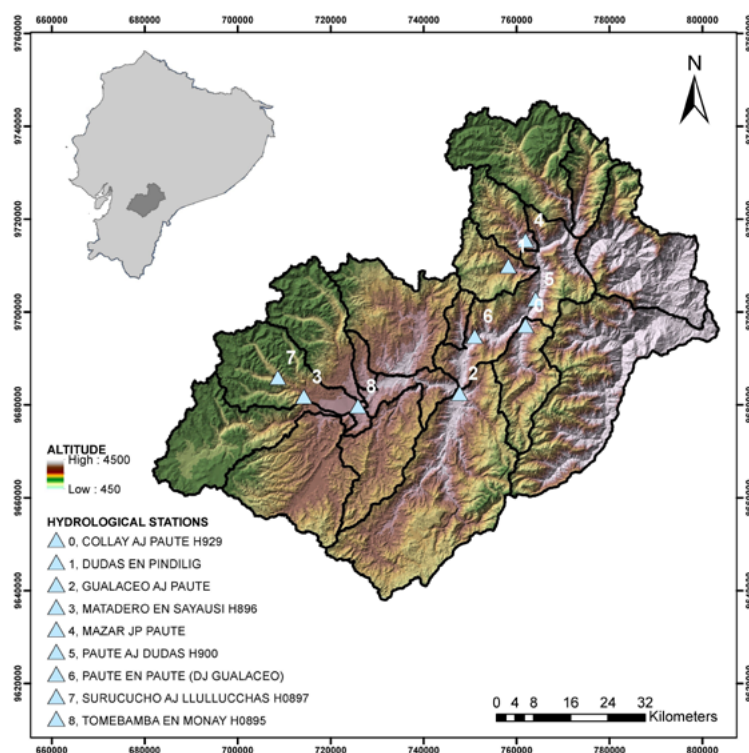
Materials y Methods

Study area

The study was conducted in the Paute river basin in southern Ecuador (Figure 1). The basin covers approximately 6437 km², with slopes ranging from 25% to 50% (CELEC EP 2013). The basin covers the Azuay, Cañar, and Morona Santiago provinces and is part of the Santiago River basin. Annual rainfall reaches its maximum average between 2500-3000 mm in the eastern region, while in the western region, it ranges between 1200-1500 mm (Institute of Regime Studies et al., 2017). The area experiences two distinct periods of frequent precipitation due to the ITCZ. The wet season for the unimodal regime occurs from June to August, while for the bimodal regime, it takes place from March to May (Campozano et al., 2016).

Figure 1

Location of the study site, as well as the location of the different sub-basins and the discharge and precipitation stations



The Paute basin consists of 18 hydrological sub-basins (Figure 1), which discharge, through a fall process, feeds the hydroelectric plants with a discharge value of Mazar (141.10 m³/s), Molino (200 m³/s), Sopladora (150 m³/s) and Cardenillo (180 m³/s), El Labrado and Chanlud (2.4 m³/s and 4.18 m³/s respectively) (Orbes & Peralta, 2017; Matute Pinos, 2014). The amount of energy produced by each of the aforementioned plants is Amaluza (1.075 MW), El Labrado and Chanlud dams feed two plants, Saymirin and Sucay (14.4 MW and 24 MW respectively); Mazar (162.6 MW) and Sopladora (500 MW). It is estimated that in total, it produces 40% of the hydroelectric production at the country level (Contreras et al., 2017).

Hydrometeorological and climatic data

Data from discharge stations of the Ecuadorian Institute for Meteorology and Hydrology (INAMHI) were collected for 20 years (1995-2015) on a monthly scale. Only stations with at least 40% data availability were considered, resulting in a total of nine stations (Table 1). Data filling was performed using the ten nearest discharge and precipitation stations for each selected station. The average monthly discharge amount was used to complete the missing data.

Table 1
Discharge and precipitation stations used for the study

STATION NAME	X	Y	ID	% MISSING VALUES
PAUTE AJ DUDAS	78.62	2.68	H900	53
GUALACEO AJ PAUTE	78.76	2.86	H0931	60
TOMEBAMBA EN MONAY	78.96	2.88	INA895	38
PAUTE EN PAUTE (DJ GUALACEO)	78.74	2.75	H894	35
SURUCUCHO AJ LLULLUCCHAS	79.12	2.83	INA897	18
MATADERO EN SAYAUSI	79.66	2.87	INA896	14
COLLAY AJ PAUTE	78.64	2.73	H929	50
DUDAS EN PINDILIG	78.68	2.62	H902_R	40
MAZAR JP PAUTE	78.64	2.57	H0892	59

In addition, data for time series of teleconnection indices and modes of variability were obtained from the database of the National Oceanic and Atmospheric Administration [41 indices (NOAA - <https://psl.noaa.gov/data/climateindices/list/>)].

Flow modeling using atmospheric and oceanic climatic indices

Teleconnection indicators, climate variability modes, and MLR were used in the flow modeling process. A variety of criteria, including PCA, correlation with monthly delays, and multicollinearity analysis, were used to generate MLRs. For calibration and validation, the study used 80% of the data. Data were utilized for calibration directly from January 1995 to September 2011, and for validation directly from September 2011 to December 2015.

Flow modeling using Principal Component Analysis (PCA)

PCA was used on teleconnection indices and climate variability modes to minimize the number of variables. MLR analyzed factors accounting for 70% and 90% of the variance (Rea and Rea 2016). R software's "prcomp" function was employed for this. PCA is essential for reducing dimensionality in data analysis while maintaining data variety and representativeness (Shabri and Shuhaida, 2014).

Modeling using Atmospheric and Oceanic Climate Indices and Monthly Delays

Three forecast scenarios were constructed to take into account the delay processes between atmospheric circulation and precipitation/flow effects. In these circumstances, the indices were delayed by the flow values by one, two, and three months. The top five teleconnection indicators for each discharge station with the highest Pearson correlation coefficients were chosen for MLR creation.

Modeling using multicollinearity analysis

A multicollinearity study was carried out to reduce the number of variables to address the problem of many variables. Multicollinearity, which indicates linear dependence across predictors and interferes with each predictor's effects on the dependent variable, was discovered using the Variance Inflation Factor (VIF) (Vega Vilca and Guzman, 2011). The appropriate VIF value was five, and variables that exceeded this limit were eliminated. For MLR building, the Leap Sequence criterion, combining forward and backward selections, was utilized (Hastie et

al., 2021). RStudio software with the MASS package and the "leapseq" function were utilized for this process.

Goodness of fit of the resulting models

To determine the accuracy of the predictions made, a series of metrics were processed in the R software V4.0.2, using the "hydroGOF" package. The metrics used in this study were KGE (Knoben, Freer and Woods, 2019), NSE (Krause, Boyle and Bäse, 2005), RMSE (Meyer, 2010), and R2. The model with the highest value in their metrics was considered the best model.

Results and Discussion

Modeling using Principal Component Analysis (PCA)

Of the 41 teleconnection indices, 10 and 20 main components were obtained, which explain 70% and 90% of the variance, respectively. It is considered that a 70% explanation of variance is a significant percentage for the study; however, the components that explained 90% of the explanation of variance were used to obtain a better prediction of flows (Rea and Rea, 2016).

The components obtained in the PCA analysis were used to perform a multiple regression model. The results obtained with the PCA analysis were plotted in time series and dispersion diagrams; the results can be seen in Figures 2 - 5.

Figures 2-5 illustrate that the results align with observed values, but predictions deviate from the actual values. Underestimation occurs at low values, while overestimation occurs at high values. Although the model performs better for lower values, it fails to accurately reflect extreme events. This limitation arises from using only MLR and station values without considering environmental behavior or hydrological processes.

Figure 2

Observed and modeled flow time series using MLR and ACP (calibration stage). a) Paute AJ Dudas, b) Gualaceo, c) Tomebamba, d) Paute DJ Gualaceo, e) Surucucho, f) Matadero, g) Collay, h) Dudas, i) Mazar

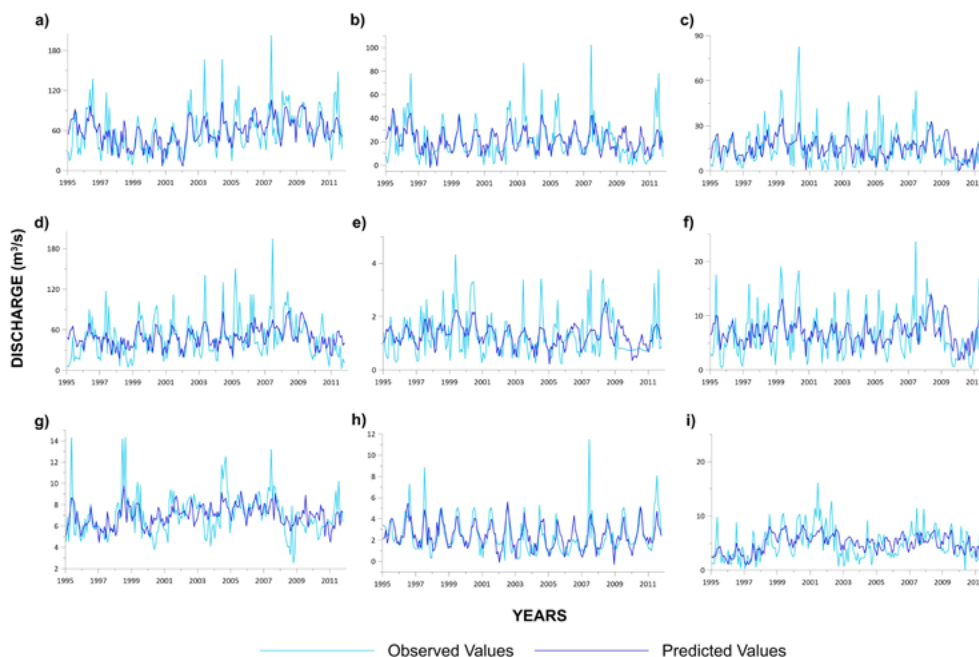


Figure 3

Observed and modeled flow time series using MLR and ACP (validation stage). a) Paute AJ Dudas, b) Gualaceo, c) Tomebamba, d) Paute DJ Gualaceo, e) Surucucho, f) Matadero, g) Collay, h) Dudas, i) Mazar

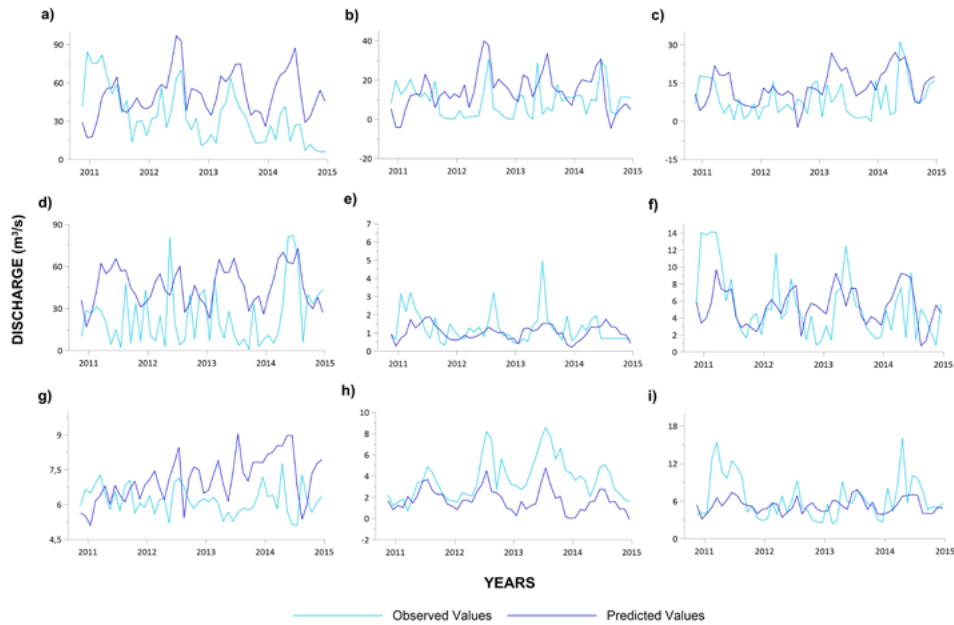


Figure 4

Dispersion plots of observed and modeled discharge using MLR and ACP (calibration stage).

a) Paute AJ Dudas, b) Gualaceo, c) Tomebamba, d) Paute DJ Gualaceo, e) Surucucho, f) Matadero, g) Collay, h) Dudas, i) Mazar

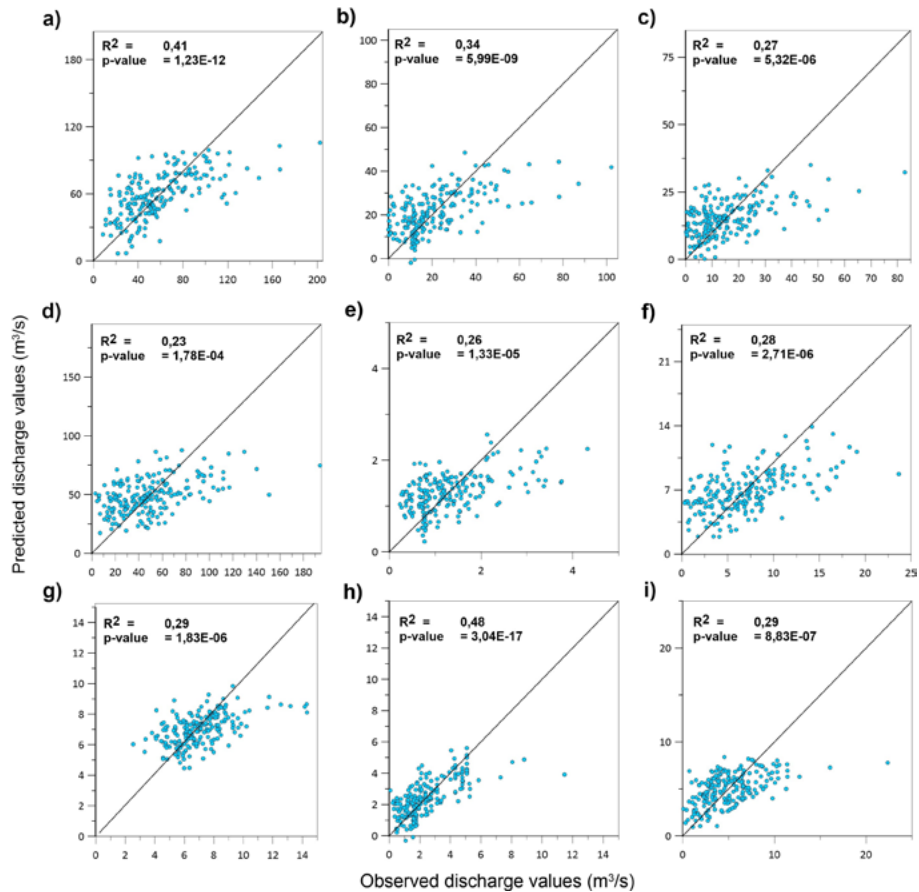
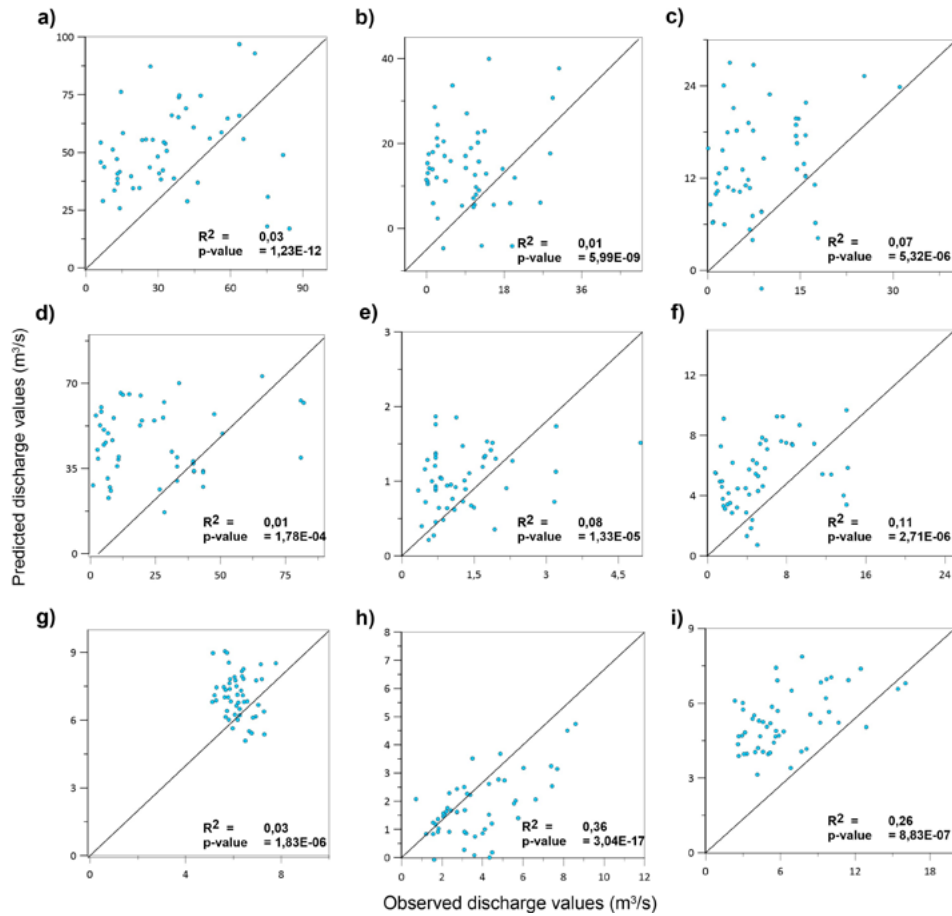


Figure 5

Dispersion plots of observed and modeled discharge using MLR and ACP (validation stage).

a) Paute AJ Dudas, b) Gualaceo, c) Tomebamba, d) Paute DJ Gualaceo, e) Surucucho, f) Matadero, g) Collay, h) Dudas, i) Mazar



During the validation phase, synchronization is seen in severe events that are within the station ranges of each participant. The precision is still less than at the calibration stage, though. The minimum p-values on the dispersion plots in both stages demonstrate statistical significance and a distinct correlation between the indices and the stations from the Paute basin. It's crucial to remember that these outcomes are dependent on the data utilized and independent of the robustness of the model.

Among all the stations, Paute AJ Dudas and Dudas (Figures 2a-2h, 3a-3h, 4a-4h, and 5a-5h) exhibit the best-fit models with correlations of 0.41 and 0.48, respectively, during the calibration stage. However, during the validation stage, none of the stations achieved satisfactory results, with the highest correlation observed in Dudas at 0.36 and Mazar at 0.26. This indicates that the models heavily rely on the initial conditions of the study and do not effectively incorporate other environmental or hydrological data.

Paute DJ Gualaceo and Surucucho demonstrate the lowest correlation results during the calibration stage, with values of 0.23 and 0.26 (Figures 2d-2e, 3d-3e, 4d-4e, and 5d-5e). Overall, the calibration results exhibit low correlations, sometimes as low as 0.01, indicating a lack of robustness in the models using this methodology.

Table 2
Metrics results at calibration and validation stage

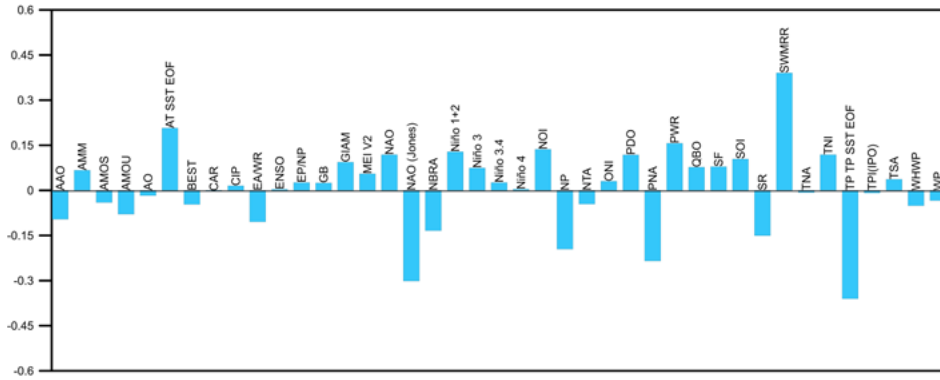
STATIONS	CALIBRATION STAGE				VALIDATION STAGE			
	RMSE	NSE	R2	KGE	RMSE	NSE	R2	KGE
PAUTE AJ DUDAS	24.35	0.41	0.41	0.49	30.25	-1.02	0.03	0.02
GUALACEO	13.27	0.34	0.34	0.41	12.93	-1.55	0.01	-0.06
TOMEBAMBA	10.74	0.27	0.27	0.33	9.5	-0.98	0.07	0.06
PAUTE DJ GUALACEO	26.05	0.23	0.23	0.27	32.07	-1.26	0.01	-0.29
SURUCUCHO	0.68	0.26	0.26	0.31	0.89	-0.05	0.08	0.08
MATADERO	3.56	0.28	0.28	0.34	3.67	0.05	0.11	0.2
COLLAY	1.6	0.29	0.29	0.34	1.52	-5.44	0.03	-0.33
DUDAS	1.19	0.48	0.48	0.57	2.5	-0.68	0.36	0.21
MAZAR	2.57	0.29	0.29	0.35	3.12	0.12	0.26	0.16

There aren't satisfactory results in these models; in Table 2 we can observe that Dudas, Collay, and Mazar were the stations with less RMSE even when only Dudas had a good model fit, which means this could be related to the low original discharge values and their predictions and not properly with the model functioning. According to Moriasi et al. (2007), the NSE does not show satisfactory results (from 0.6 or higher), classification describes a performance as very good, good, satisfactory, and unsatisfactory with values higher to 0.75, between 0.65 – 0.75, between 0.5 – 0.65 and lower to 0.5 respectively. Dudas and Paute AJ Dudas also domain this metric in the calibration stage. KGE gets better performance values than NSE, without changing the results of stations.

In the validation stage, Dudas still has one of the best performances, but Mazar and Matadero also appear here. These metrics do not have any special tendency. A study made in New Gales, Australia, using the indices ENSO, MEI, IOD, Niño 3.4, PDO, and TPI to predict the study zone discharges, got NSE values between 0.15 – 0.55, which means that indices by themselves are not enough to predict discharge (Esha, Imteaz y Nazari, 2019). A study realized in western Canada mentions that PDO (used in our study), even when related to ENSO, also has great uncertainty regarding the nature and origin of the index variability itself; which results in a waste of potential use in forecasting studies, especially because the statistic shows that it is highly related to water resources (Whitfield et al., 2010).

In comparison to other approaches, PCA had the lowest RSME, indicating that other indices may be more closely associated than the ENSO index or its variability modes (unique indices have the potential to affect only this methodology). Additionally, there may be several variables utilized for the study; however, the VIF approach challenges this idea because, in some instances, the indices used for the models soar to 13. Metrics show that incorporating 90% of variance in all models rather than just 70% results in an improvement.

Figure 6
Factorial Components of ACP using 41 indices and 20 Components



In Figure 6, we can observe the loading factor of ACP, which is the influence of every index on the 20 components. According to the study, the best models (90%) are indirectly related to ENSO indices and not to the study area, as can be identified in a comparison between the indices and their influence on each component used in the modeling (AT SST EOF, PWR, and SWMRR (South West Monsoon Region Rainfall) with positive values, NAO (Jones), NP, PNA, SR, and TP SST EOF (Tropical Pacific SST Empirical Orthogonal Function) (negatives). The highest correlations occur with SWMRR (0.39) and AT SST EOF (0.21). This correlation could occur because the rainfall season of SWMRR in Mexico and Arizona happens from June to September, matching the wet season of the Paute River Basin of the unimodal regimen from June to August (Campozano et al., 2016 & Crimmins, 2014). And AT SST EOF correlation is related to the location of this anomaly in Ecuador (Fan and Schneider, 2012).

The PNA and other ENSO patterns used in research on the Canadian Columbia River basin showed that depending on the station (certain months), ENSO variability modes could be more or less influent, even causing an anomaly in the river discharges (Gobena, Weber y Fleming, 2013). As the results of our models are not accurate, this can be considered an explanation, having indexes that are more related to anomalies in the ACP study, being necessary to realize a correlation study considering each station's anomaly values on the data time series.

In a study made in Iran, 25 variability modes of teleconnection indices were used to explain the variability of precipitation using ACP methodology (AMO, AMM, BEST, Niño 3.4, El Niño 4, NTA, SOI, and TNA). Results show eight principal components that explain 80% of the variance (Choubin et al., 2016). In Colombia, a study that analyzes the climatic variability of the Cauca River using the ACP in Index (CCC, an index of the own river; ONI; PDO; EMI, El Niño Midoki; SST; SOI; and MEI) has as results that two components that explain 80% of the variance, an only one explaining 70%. The eigenvalues (proportional to loading factors) showed that this first component with 70% of variance was more associated with ENSO and the two components with 80% of variance were more related to the index of the study zone CCC (Sedano, 2017). In our study, the models work similarly, most of the indices directly correlate with ENSO and SST. It is also visible in the studies of Choubin et al. (2016) and Sedano (2017) that the percentage of input data needed to explain 70% of the variance is from 25 – 30%, and for upper percentages, it is from 35 – 50% of them. The influence of every index involved decreases with the number of components, focusing on their correlation.

Modeling using Atmospheric and Oceanic Climate Indices and Monthly Delays

To determine the variables of each prediction model for the discharge stations, the Pearson correlation was carried out for the different stations and the teleconnection indices,

making delays of one to three months, considering the absolute value of the correlations. The first five indices with the highest values were used. The best results were obtained by performing a one-month delay between the values of the stations and the teleconnection rates. The results of the correlation using a delay month can be seen in Figure 7.

Figure 7
Graphic representation of the existing correlation between discharge stations and teleconnection indices



The absolute values of correlation are usually in a range from 0.35 to 0.61. The highest correlation is the Dudas station with a CIP index of 0.61, and the lowest belongs to the same station with an NTA index with 0.11. Once the correlation values between the discharge stations and the different teleconnection indices were obtained, the following was to identify the five indices with the highest absolute correlation values, which are presented in Table 3.

Table 3
Absolute Correlation values of the indices used for the multiple regression models of each discharge station

STATION	INDICES AND CORRELATION VALUES				
PAUTE AJ DUDAS	SF	NP	CAR	GB	WHWP
	0.35	0.29	0.21	0.19	0.18
GUALACEO	NP	CIP	GB	AMM	PWR
	0.28	0.26	0.23	0.20	0.19
TOMEBAMBA	AMM	WHWP	Niño 4	PWR	TNI
	0.23	0.22	0.22	0.22	0.21
PAUTE DJ GUALACEO	NP	CAR	PWR	GB	AMM
	0.23	0.22	0.21	0.20	0.19
SURUCUCHO	AMM	NP	Niño 4	PWR	NTA
	0.31	0.30	0.23	0.23	0.23
MATADERO	PWR	WHWP	Niño 4	AMM	TNA
	0.29	0.28	0.28	0.26	0.24
COLLAY	CAR	AMOU	TNI	SR	AT SST EOF
	0.26	0.22	0.18	0.18	0.17
DUDAS	CIP	NP	AMM	NBRA	WHWP
	0.61	0.44	0.25	0.24	0.23
MAZAR	AMM	NP	AMOS	SF	NTA
	0.30	0.28	0.28	0.28	0.23

The MLRs were generated with the indices from Table 2. The results are presented in Figures 8-11 with a one-month delay; the behavior of the modeled time series adjusts to the trend of the series of each station, especially at mean values and low values. However, it presents the same problem as the series predicted with the PCA method. All values are indifferent to the model results.

Figure 8

*Observed and modeled flow time series using MLR and one-month delay (calibration stage).
 a) Paute AJ Dudas, b) Gualaceo, c) Tomebamba, d) Paute DJ Gualaceo, e) Surucucho, f) Matadero, g) Collay, h) Dudas, i) Mazar*

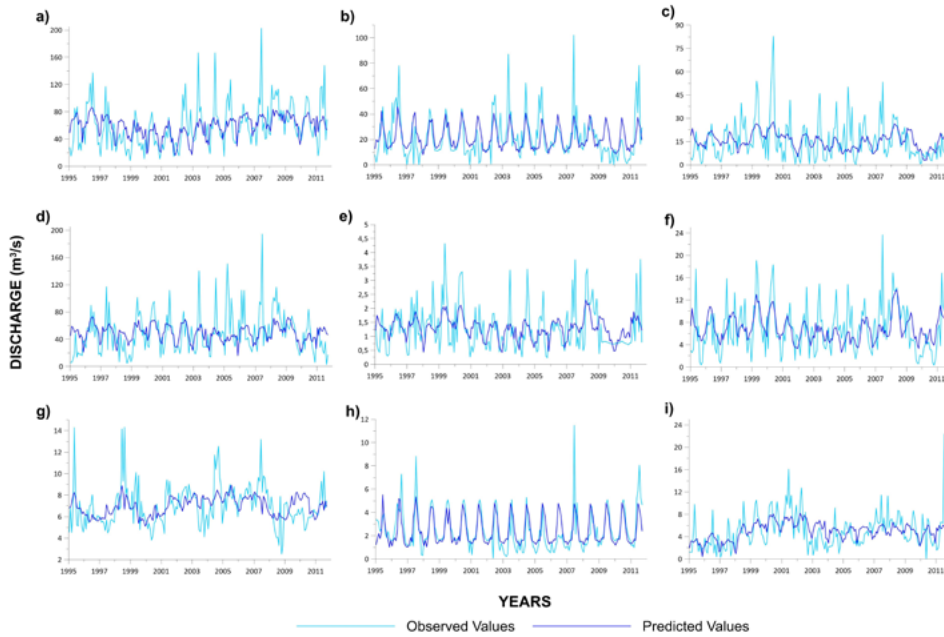
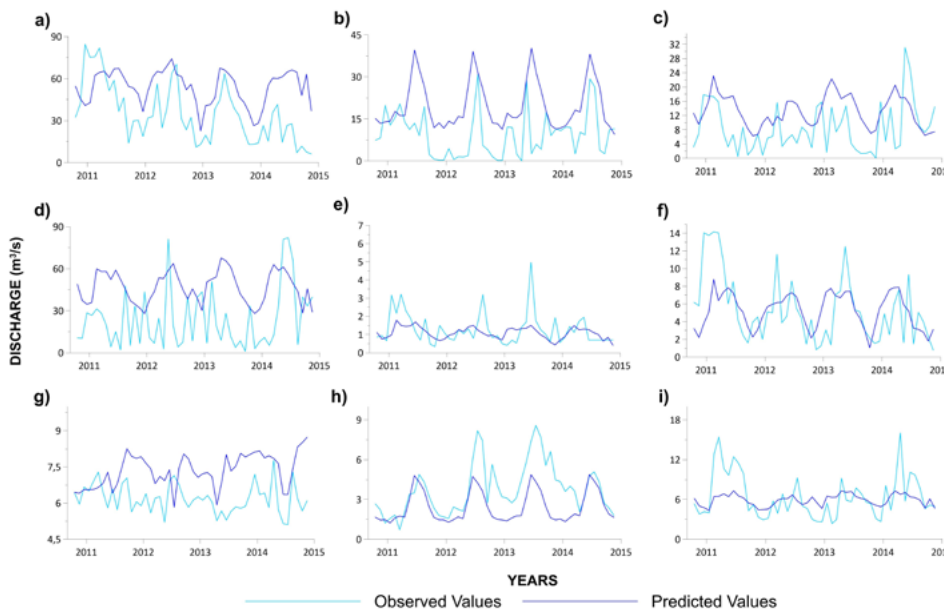


Figure 9

*Observed and modeled flow time series using MLR and one-month delay (validation stage).
 a) Paute AJ Dudas, b) Gualaceo, c) Tomebamba, d) Paute DJ Gualaceo, e) Surucucho, f) Matadero, g) Collay, h) Dudas, i) Mazar*



In 2018, an investigation into the evolution of the sediments of several lakes in the El Cajas National Park determined that, with a one-month delay for the ENSO index, the results of the correlation with the precipitations close to the local stations were solid, proving that there is a strong positive link between intense rains and therefore an increase in flow between La Niña, Niño 3.4 (which explains the variability of the Ecuadorian Andes), El Niño 4 (a more intense relationship), El Niño 1+2 (which explains the variability of the coastal plains), and a more neutral link with El Niño (Schneider et al., 2018).

The delay of one month in our study proves that there is a delay in the effects of the teleconnection rates on the climatic conditions of the Paute river basin. And that the variability of the indices fits with the level of correlation found considering the spatial configuration of the studies. In a study carried out in Brazil on the relationship between the alluvial plains and the connection with the signals (variability modes) of ENSO teleconnection (El Niño, Niño 3.4, SOI), it was determined that for the Amazon basin, there is a delay of two months. This occurs from January to March and causes lower rainfall, causing a decrease in flow (Schöngart et al., 2004). However, an investigation carried out in Peru on the relationship between monthly rainfall and SST identified that in the areas of the equatorial Pacific, ENSO has a delay of one month in the wet season and zero in the dry season (Bazo, Lorenzo and Porfirio Da Rocha, 2013). When comparing our results with the literature, it confirms our results because the indices with the highest correlation for this scenario are AMM, NP, PWR and, WHWP, which are strongly influenced by ENSO variability modes.

Figure 10

Dispersion plots of observed and modeled discharge using MLR and one-month delay (calibration stage). a) Paute AJ Dudas, b) Gualaceo, c) Tomebamba, d) Paute DJ Gualaceo, e) Surucucho, f) Matadero, g) Collay, h) Dudas, i) Mazar

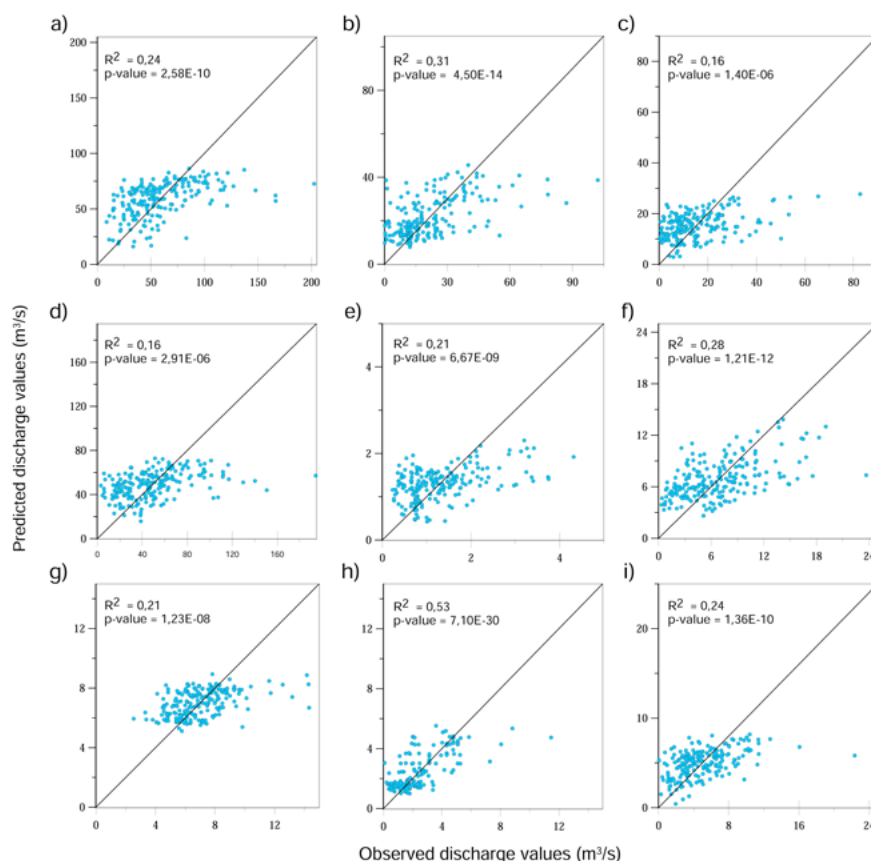
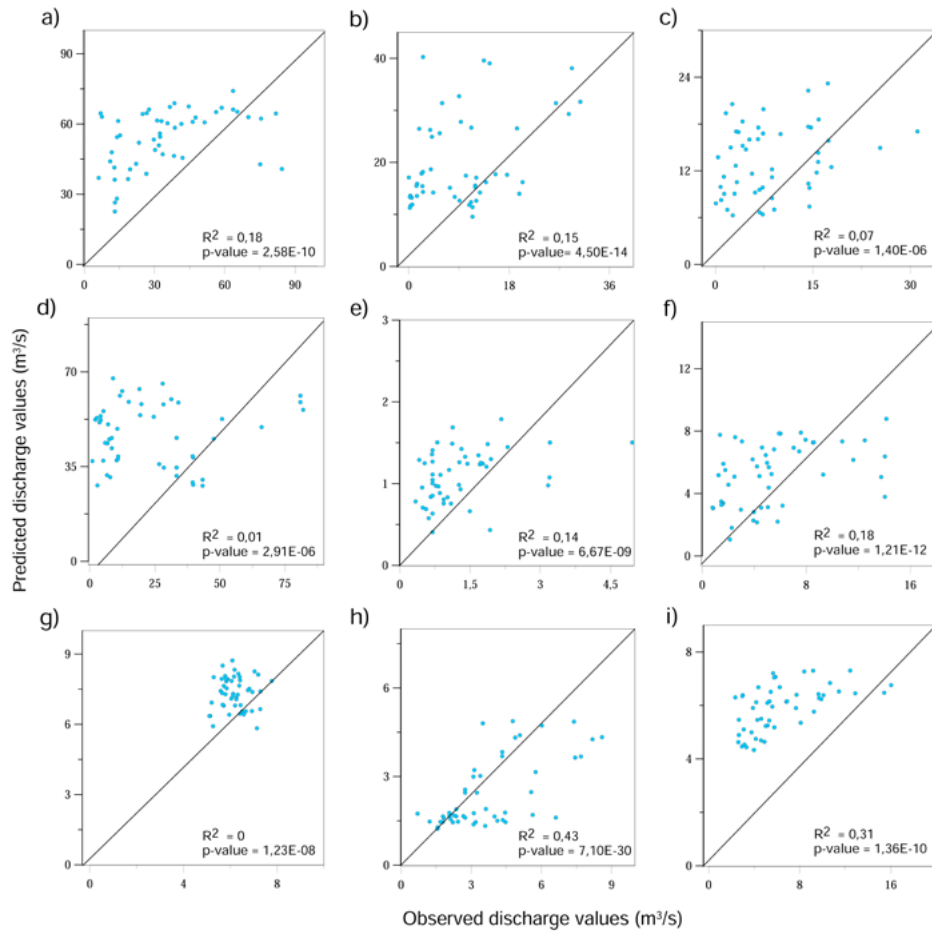


Figure 11

Dispersion plots of observed and modeled discharge using MLR and one-month delay (validation stage). a) Paute AJ Dudas, b) Gualaceo, c) Tomebamba, d) Paute DJ Gualaceo, e) Surucucho, f) Matadero, g) Collay, h) Dudas, i) Mazar



The PWR index is conditioned by El Niño (Carreric, 2020); on the other hand, the WHWP index is also directly affected by El Niño in the summer through the TNA temperature increase that occurs during the El Niño winter in the Pacific (Wang y Enfield, 2001). NP is dominated by the interannual variations of the El Niño and La Niña events (Espino Sánchez, 2014); AMM, for its part, produces anomalies in the SST north of Ecuador, but it occurs during the boreal spring (March) with greater intensity. In a case study of the rivers of Quebec in Canada was discovered that between ENSO, NAO, and PDO there is a large delay between the effect of the patterns and the answer on the discharge. It is also mentioned that the modes of variability associated with teleconnection patterns can change the periodicity and the stationary effects. This can explain the non-optimal fit models in this methodology (McGregor, 2017).

For the modeling, an acceptable R2 value is obtained in Dudas with 0.53 (Figure 10h), which confirms the ability of the models to capture the averages. The second-best performance occurs at the Gualaceo station with a coefficient of 0.31 (Figure 10b) and Matadero with 0.28 (Figure 10f). In this case, two of the stations varied in the indices used, which represented a great improvement for both cases. The models with the lowest performance are Paute DJ Gualaceo and Tomebamba, both with a value of 0.16 (Figure 10c – 10d).

When analyzing the coefficient of determination (Table 4), it can be seen that Dudas station is acceptable according to the bibliography. Since values less than 0.5 indicate in a

hydrological model that there is a large error variance integrated into the model and that it cannot be explained by it, in other words, using only the values of the indices in a multiple regression model does not give good results for the flows of the Paute river basin. There are local and regional variables that could be integrated into the model, improving its predictive capacity (Moriassi et al., 2007).

Table 4
ACP metrics results at calibration and validation stage

STATIONS	CALIBRATION STAGE				VALIDATION STAGE			
	RMSE	NSE	R2	KGE	RMSE	NSE	R2	KGE
PAUTE AJ DUDAS	27.69	0.24	0.24	0.28	26.79	-0.65	0.18	0.11
GUALACEO	13.62	0.31	0.31	0.37	13.92	-1.95	0.15	-0.27
TOMEBAMBA	11.54	0.16	0.16	0.16	8.49	-0.6	0.07	0.01
PAUTE DJ GUALACEO	27.29	0.16	0.16	0.15	32.68	-1.37	0.01	-0.41
SURURCUCHO	0.71	0.21	0.21	0.24	0.84	0.08	0.14	0.09
MATADERO	3.57	0.28	0.28	0.34	3.44	0.16	0.18	0.25
COLLAY	1.69	0.21	0.21	0.23	1.46	-4.96	0	0
DUDAS	1.15	0.53	0.53	0.61	1.96	-0.06	0.43	0.37
MAZAR	2.66	0.24	0.24	0.28	2.98	0.2	0.31	0.13

In this modeling, we get the best fit of the model. NSE reaches a satisfactory result in the calibration stage with Dudas, and KGE gets a good fit model (0.61). Paute AJ Dudas's results are reduced compared with other methodologies; this can occur when analyzing where stations are located and how the delay effects affect every station. Tomebamba and Gualaceo have the lowest results, reaching 0.15 and 0.16 on NSE and KGE, respectively. RMSE is only related to the discharge of each station and not the results of the modeling.

In the calibration stage, Dudas is still the best station with 0.37 on KGE, while Collay gets the worst results with -4.96 on NSE and zero on KGE. This means that the average monthly values are a better predictor than our model (Krause et al., 2005). The variance range is very wide, and these results prove that the models have no statistical robustness even when the p-value maintains infinity in all cases (both stages).

Modeling using multicollinearity analysis

The third case used for flow modeling was a preliminary analysis of multicollinearity, as explained in the previous section on materials and methods. When using the VIF criterion for the reduction of the variables, a total of 22 variables were obtained (reduction of almost half of the original variables); the resulting variables were PNA, WP, EA/WR, NAO, TSA, PDO, NP, AO, AAO, PWR, CAR, AMOS, QBO, SR, SF, GB, EP/NP, NAO (Jones), NOI, CIP, NBRA, SWMRR and AT SST EOF. None of the modes of variability or signals directly related to ENSO (MEI V2, Niño 1+2, Niño 3, Niño 3.4, Niño 4, SOI, ONI) are reflected in the results. Indirectly, we find NP, which indicates its high influence on the rest of the indices and its intimately connected and similar behavior. This also happens with the indices AMO UNSMOOTHED, AMM, and AMO SMOOTHED, since only one is kept after VIF analysis. With the resulting

variables, the next step was the construction of the MLR using the stepwise criterion. The best model for each of the discharge stations is presented in Table 5.

Table 5
MLR used for prediction with Multicollinearity Analysis

DISCHARGE	N° VARIABLES	INDICES
PAUTE AJ DUDAS	5	AMOS + AT SST EOF + NP + PDO + SOLAR
GUALACEO	2	NP + PDO
TOMBAMBA	5	EA/WR + NP + PWR + CIP + NBRA
PAUTE DJ GUALACEO	4	AMOS + CAR + NP + PDO
SURUCUCHO	2	NP + PWR
MATADERO	1	PWR
COLLAY	2	CAR + SR
DUDAS	1	CIP
MAZAR	5	AMOS + NP + SOLAR+TSA+NBRA

Figures 12-15 show the results of the predictions made under the VIF analysis for multicollinearity. The trends are maintained in all seasons, and visually, the results are acceptable (Dudas achieves the best modeling).

Figure 12
Observed and modeled flow time series using MLR and Multicollinearity Analysis (calibration stage). a) Paute AJ Dudas, b) Gualaceo, c) Tomebamba, d) Paute DJ Gualaceo, e) Surucucho, f) Matadero, g) Collay, h) Dudas, i) Mazar

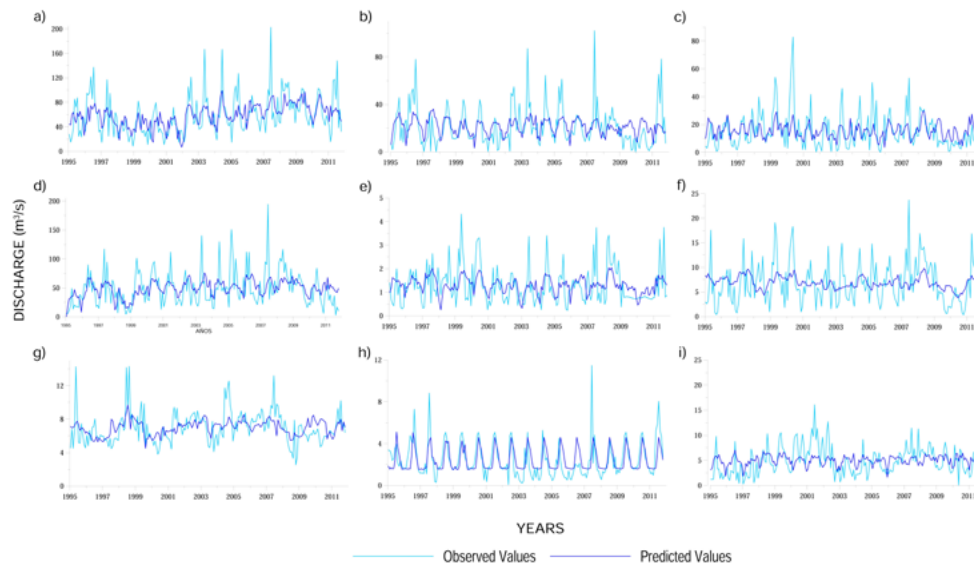
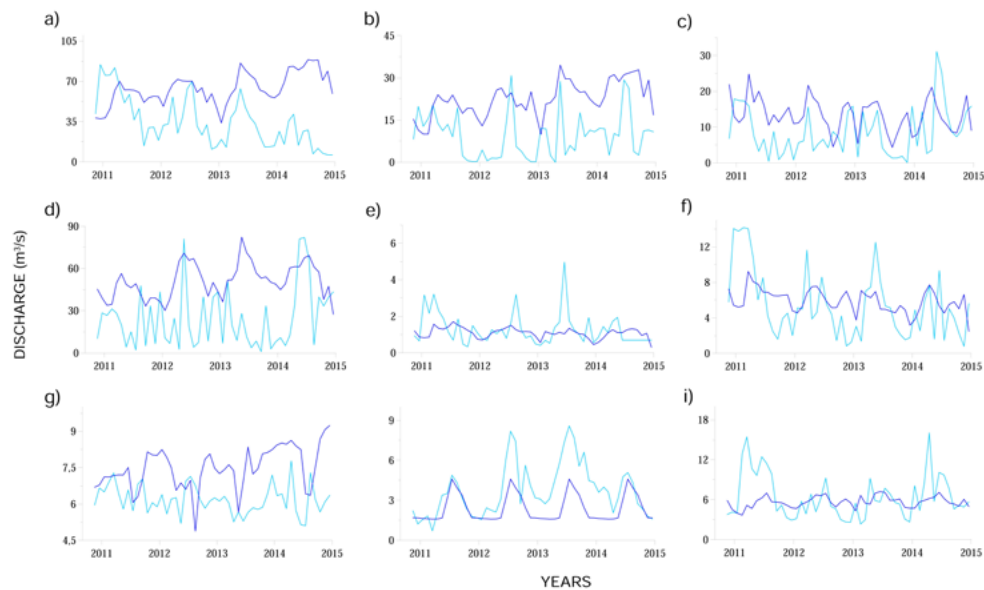


Figure 13

Observed and modeled flow time series using MLR and Multicollinearity Analysis (validation stage). a) Paute AJ Dudas, b) Gualaceo, c) Tomebamba, d) Paute DJ Gualaceo, e) Surucucho, f) Matadero, g) Collay, h) Dudas, i) Mazar

**Table 6**

Metrics results at the calibration and the validation stage

STATIONS	CALIBRATION STAGE				VALIDATION STAGE			
	RMSE	NSE	R2	KGE	RMSE	NSE	R2	KGE
PAUTE AJ DUDAS	26.2	0.32	0.32	0.38	39.73	-7.51	0.02	-0.36
GUALACEO	15.01	0.16	0.16	0.14	16.05	-5.86	0.02	-0.09
TOMBAMBA	11.46	0.17	0.17	0.18	8.76	-2.69	0.03	-0.01
PAUTE DJ GUALACEO	26.45	0.21	0.21	0.24	35	-7.26	0.02	-0.24
SURURCUCHO	0.72	0.18	0.18	0.19	0.89	-8.96	0.03	-1.25
MATADERO	4.03	0.08	0.08	-0.02	3.66	-6.92	0.07	-1.02
COLLAY	1.69	0.21	0.21	0.23	1.70	-2.83	0.01	-0.15
DUDAS	1.29	0.40	0.40	0.48	1.98	-2.77	0.45	-0.10
MAZAR	2.9	0.10	0.10	0.03	3.39	-15.73	0.02	-2.13

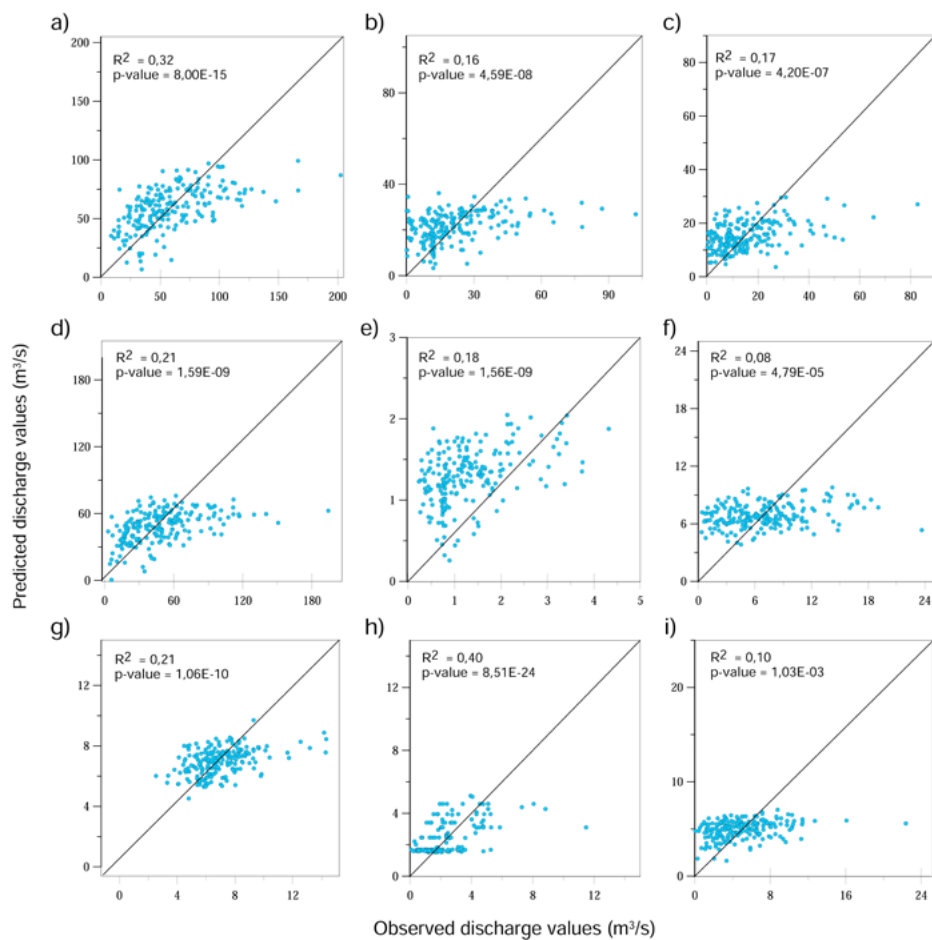
Using VIF to select the indices as variables in every model gave us different results for every station, and these results are not related to the functioning of indices but only to methodology. In the calibration stage, Dudas obtained the closest to a good result on correlation, NSE, and KGE, with 0.40 and 0.48, respectively. Paute AJ Dudas gets another close result to satisfactory with 0.38 in KGE. The worst fits occur on Matadero and Gualaceo, reaching even negative values on KGE (0.14 to -0.02). In this model, it is very clear that the results are not improved by using more variables. As a result, Dudas only uses one index to obtain better results in comparison with Mazar, which uses 5 (the maximum number of variables in each station) and gets poor results. The noise in this methodology does not influence it drastically since all RMSEs tend to maintain low values, which means the results are a show of the functioning itself. Dudas has one variable, Paute AJ Dudas 5, Matadero, and Gualaceo 1 and 2, proving

that only one index is enough to capture the variability of the discharges if it is related to the environment of the station.

In the validation stage, Tomebamba is the “best” fit model with -2.69 ; the rest have also negative values on NSE, in KGE. These results are similar, except in the case of Dudas with a 0.45 correlation, proving that is the station with the best results throughout the study. RMSE in both cases still obeys the initial values of the data to show results more than the model prediction, which is why the stations with the highest values, Paute AJ Dudas and Paute DJ Gualaceo, have high RSME, but their performance is very different, with Paute AJ Dudas the second-best station to fit the models.

Figure 14

Dispersion plots of observed and modeled discharge using MLR and Multicollinearity Analysis (calibration stage). a) Paute AJ Dudas, b) Gualaceo, c) Tomebamba, d) Paute DJ Gualaceo, e) Surucucho, f) Matadero, g) Collay, h) Dudas, i) Mazar



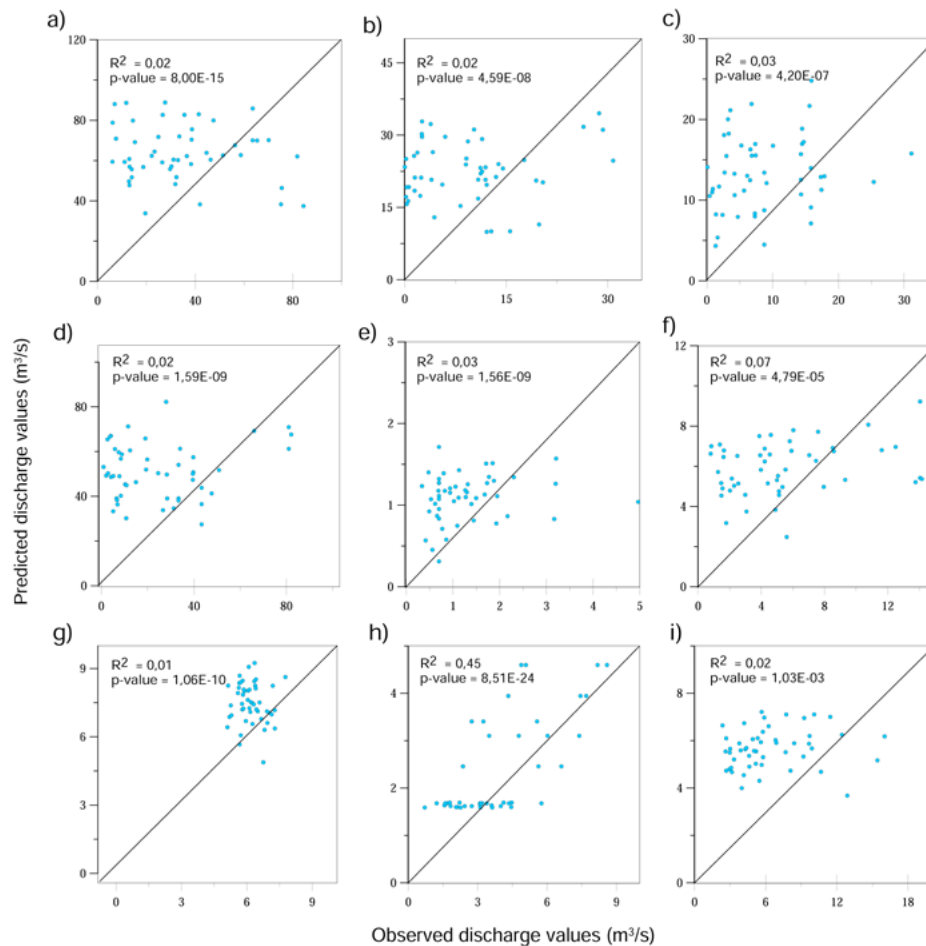
A study made in Indonesia that used SOI, Niño 3.4, and IOD indices to analyze the Java discharge regimen identified that the results on LRM with KGE metrics were best when the model included more variables (Nugroho, Tamagawa and Harada, 2022). This also happened in our study using this specific methodology, where we used a different number of variables for every station. In addition, from a general perspective, when using 20 components as variables on ACP also happened.

When we analyze Figure 14f, we can see the worst result: Matadero used only PWR to get the predictions, and we can see no relation between the behavior of the original series and

the predicted one. P- value maintains low in every station which means they are statistically representative. RSME is dominated by the same behavior as in the two other methodologies.

Figure 15

Dispersion plots of observed and modeled discharge using MLR and Multicollinearity Analysis (validation stage). a) Paute AJ Dudas, b) Gualaceo, c) Tomebamba, d) Paute DJ Gualaceo, e) Surucucho, f) Matadero, g) Collay, h) Dudas, i) Mazar



Even when the best models tend to have more variables, the variables change for every station; this is proven because stations with the same indices can have bad or good indices and are independent of each other. Also, there are stations that, using the same number of indices, can have better results than other stations. The results are directly linked with how the values of the time series couple with the indices data in an individual way. And the value of the data itself does not influence the results.

There are no similar indices used in all scenarios, but the ones that are present in the two methodologies (VIF and using monthly delays) are: NP, PWR, CAR, and CIP; also, this one has the most influence on loading factors in ACP (NP and PWR). The two better-fitting stations are Dudas and Paute AJ Dudas since this one does not have extreme events as frequently as the other stations, allowing LRM to work better. This methodology is good for approximation but it is limited to mathematical procedure because the only input data is the time series and not any environmental variable that can help to understand the hydrological behaviors (not environmental context or another variable). The stations with the worst results are Mazar, Gualaceo, and Matadero. Also, the validation range time could be amplified to improve the

models by catching more variability, especially in the extreme events that are not so frequent. In the Amazon Basin, research determined that ENSO domains the conditions of the discharges but does not fully understand the behavior of each variability mode and its effects. This is caused by the temporal SST anomalies, their magnitude, and their position in the equatorial Pacific Ocean. The study also points out that some statistical tests, such as Kendall, could not be suitable for certain regions. With a large-basin memory, it leads to autocorrelation and misleading significance (Towner et al., 2020). This is not only observed in our study but also evidenced in each of the methodologies used.

Conclusions

The best models can follow the original data tendency but cannot predict whether the values are too low or too high. The best scenarios were obtained when using PCA as a general view, but the best values in metrics were obtained when using a 1-month delay. The models except Dudas and Paute AJ Dudas mostly didn't reach a satisfactory result in all metrics (NSE, KGE, R2). This means that indices are not enough as discharge predictors. RSME is not a good metric in our study since its values were associated with the data and not with the model's performance in all cases. It shows a very light tendency on PCA, lowering their values. The validation stage determines that the models do not have robustness, and this is directly related to the initial conditions of our models not applying to different contexts. We have several indices associated with ENSO that do not represent any relevance in the study (AO, BEST, ENSO, GIAM, MEI V2, NAO, Niño 3, Niño 3.4, ONI, SOI, TPI, and IPO). It is important to recognize if this index has a direct or indirect effect on the zone and which of his variability modes better represents it. Finally, it is important to highlight that being a first approximation to the use of all teleconnection indices to measure their predictive capacity in the Paute river basin, the results are very interpretable. It is essential to mention that the teleconnection indices (signals and modes of variability), in addition to providing global information, maintain a well-studied atmospheric and oceanic circulation. This makes it possible not only to observe climate variability on a large spatial and temporal scale but also to analyze variables dependent on these phenomena without having to resort to breaking them down for a single study.

Acknowledgments

The authors would like to thank the INAMHI for the information provided. This work was founded for Corporación Ecuatoriana para el Desarrollo de la Investigación y la Academia (CEDIA) within the research project “Análisis Nexus agua-alimentos-energía-servicios ecosistémicos ante cambios del clima, uso del suelo y población. Un enfoque novedoso para el desarrollo sostenible local a escala de una cuenca hidrográfica” and Vicerrectorado de Investigaciones de la Universidad de Cuenca (VIUC).

References

- Bazo, J., Lorenzo, M. D. L. N., & Porfirio Da Rocha, R. (2013). Relationship between monthly rainfall in NW Peru and tropical sea surface temperature. *Advances in Meteorology*, 2013. <https://doi.org/10.1155/2013/152875>
- Cai, W., McPhaden, M. J., Grimm, A. M., Rodrigues, R. R., Taschetto, A. S., Garreaud, R. D., Dewitte, B., Poveda, G., Ham, Y.-G., Santoso, A., Ng, B., Anderson, W., Wang, G., Geng, T., Jo, H.-S., Marengo, J. A., Alves, L. M., Osman, M., Li, S., ... Vera, C. (2020). Climate impacts of the El Niño–Southern Oscillation on South America. *Nature Reviews Earth & Environment*, 1(4), 215–231. <https://doi.org/10.1038/s43017-020-0040-3>

- Campozano, L., Célleri, R., Trachte, K., Bendix, J., & Samaniego, E. (2016). Rainfall and Cloud Dynamics in the Andes: A Southern Ecuador Case Study. *Advances in Meteorology*, 2016, 1–15. <https://doi.org/10.1155/2016/3192765>
- Carreric, A. (2019). *Enso diversity and global warming* (Doctoral dissertation, Université Paul Sabatier-Toulouse III).
- CELEC EP. (2013). *Actualización Del Estudio De Impacto Ambiental Y Plan De Manejo Ex Post De La Central Paute Molino (Update of Environmental Impact Assessment and Post-Management Plan of Paute Molino Power Plant)*.
- Celleri, R., Willems, P., Buytaert, W., & Feyen, J. (2007). Space–time rainfall variability in the Paute basin, Ecuadorian Andes. *Hydrological Processes*, 21(24), 3316–3327. <https://doi.org/10.1002/hyp.6575>
- Choubin, B., Khalighi-Sigaroodi, S., Malekian, A., & Kişi, Ö. (2016). Multiple linear regression, multi-layer perceptron network and adaptive neuro-fuzzy inference system for forecasting precipitation based on large-scale climate signals. *Hydrological Sciences Journal*, 61(6), 1001–1009. <https://doi.org/10.1080/02626667.2014.966721>
- Córdoba Machado, S., Palomino Lemus, R., Gámiz Fortis, S. R., Castro Díez, Y., & EstebanParra, M. J. (2015). Assessing the impact of El Niño Modoki on seasonal precipitation in Colombia. *Global and Planetary Change*, 124, 41–61. <https://doi.org/10.1016/j.gloplacha.2014.11.003>
- Crimmins, M. (2014). Southwestern Monsoon. Climate Assessment for the SouthWest. <https://climas.arizona.edu/sw-climate/monsoon>
- Dima, M., & Lohmann, G. (2004). Fundamental and derived modes of climate variability: concept and application to interannual time-scales. *Tellus A: Dynamic Meteorology and Oceanography*, 56(3), 229. <https://doi.org/10.3402/tellusa.v56i3.14415>
- Esha, R. I., Imteaz, M. A., & Nazari, A. (2019). Assessing Gene Expression Programming as a technique for seasonal streamflow prediction: A case study of NSW. *IOP Conference Series: Earth and Environmental Science*, 351(1), 012004. <https://doi.org/10.1088/1755-1315/351/1/012004>
- Espino Sánchez, M. A. (2014). *Patrones de variabilidad ambiental y las pesquerías en el Pacífico Sud Este (Patterns of Environmental Variability and Fisheries in the Southeast Pacific)*. Universidad Nacional Mayor de San Marcos.
- Fan, M., & Schneider, E. K. (2012). Observed Decadal North Atlantic Tripole SST Variability. Part I: Weather Noise Forcing and Coupled Response. *Journal of the Atmospheric Sciences*, 69(1), 35–50. <https://doi.org/10.1175/JAS-D-11-018.1>
- Giddings, L., & Soto, M. (2006). *Teleconexiones y precipitación en América del Sur [Teleconnections and precipitation in South America]*. 6, 13–20.
- Gobena, A. K., Weber, F. A., & Fleming, S. W. (2013). The Role of Large-Scale Climate Modes in Regional Streamflow Variability and Implications for Water Supply Forecasting: A Case Study of the Canadian Columbia River Basin. *Atmosphere-Ocean*, 51(4), 380–391. <https://doi.org/10.1080/07055900.2012.759899>
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An Introduction to Statistical Learning* (2nd ed.). Springer US. <https://doi.org/10.1007/978-1-0716-1418-1>
- Hatzaki, M., Flocas, H. A., Asimakopoulos, D. N., & Maheras, P. (2007). The eastern Mediterranean teleconnection pattern: identification and definition. *International Journal of Climatology*, 27(6), 727–737. <https://doi.org/10.1002/joc.1429>
- IDEAM - UNAL (2018). *Variabilidad Climática y el cambio climático en Colombia (1era ed.) [Climate Variability and Climate Change in Colombia (1st ed.)]*. IDEAM

- Contreras, J., Ballari, D., & Samaniego, E. (2017). EJE 02-09 Optimización de una red de monitoreo de precipitación usando modelos Geoestadísticos: caso de estudio en la cuenca del río Paute, Ecuador [AXIS 02-09 Optimization of a precipitation monitoring network using geostatistical models: case study in the Paute river basin, Ecuador]. *Memorias Y Boletines De La Universidad Del Azuay*, 1(XVI), 115–124. <https://doi.org/10.33324/memorias.v1iXVI.55>
- Knoben, W. J. M., Freer, J. E., & Woods, R. A. (2019). Technical note: Inherent benchmark or not? Comparing Nash–Sutcliffe and Kling–Gupta efficiency scores. *Hydrology and Earth System Sciences*, 23(10), 4323–4331. <https://doi.org/10.5194/hess-23-4323-2019>
- Krause, P., Boyle, D. P., & Bäse, F. (2005). Comparison of different efficiency criteria for hydrological model assessment. *Advances in Geosciences*, 5, 89–97. <https://doi.org/10.5194/adgeo-5-89-2005>
- Kundzewicz, Szwed, & Pińskwar. (2019). Climate Variability and Floods—A global Review. *Water*, 11(7), 1399. <https://doi.org/10.3390/w11071399>
- Matute, V. (2014). *Análisis De Factibilidad De Generación Eléctrica A Pie De La Presa De Chanlud* (Doctoral dissertation, Universidad de Cuenca).
- McGregor, G. (2017). Hydroclimatology, modes of climatic variability and stream flow, lake and groundwater level variability. *Progress in Physical Geography: Earth and Environment*, 41(4), 496–512. <https://doi.org/10.1177/0309133317726537>
- Meyer, T. (2010). Technical note: Root Mean Square Error Compared to, and Contrasted with, Standard Deviation. *Surveying and Land Information Science*, 72(3), 107-108
- Morán-Tejeda, E., Bazo, J., López-Moreno, J. I., Aguilar, E., Azorín-Molina, C., Sanchez-Lorenzo, A., Martínez, R., Nieto, J. J., Mejía, R., Martín-Hernández, N., & Vicente-Serrano, S. M. (2016). Climate trends and variability in Ecuador (1966-2011). *International Journal of Climatology*, 36(11), 3839–3855. <https://doi.org/10.1002/joc.4597>
- Moriassi, D., Arnold, J., Van Liew, M., Bingner, R., Harmel, R., & Veith, T. (2007). Model Evaluation Guidelines for Systematic Quantification of Accuracy in Watershed Simulations. *Transactions of the ASABE*, 50(3), 885–900. <https://doi.org/10.13031/2013.23153>
- Nugroho, A. R., Tamagawa, I., & Harada, M. (2022). Spatiotemporal Analysis on the Teleconnection of ENSO and IOD to the Stream Flow Regimes in Java, Indonesia. *Water (Switzerland)*, 14(2). <https://doi.org/10.3390/w14020168>
- Orbes, J., & Peralta, T. (2017). *Estado del arte en Manejo de Sedimentos en cuencas Andinas en el Ecuador, caso de estudio: cuenca del Río Paute [State of the Art in Sediment Management in Andean Watersheds in Ecuador; Case Study: Paute River Basin]*. (Bachelor dissertation, Universidad de Cuenca)
- Fontaine, G., Narváez, I., and Cisneros, P. (2008). *[Geo Ecuador 2008: State of the Environment Report]*. FLACSO. Quito, Ecuador. Available online at: <https://biblio.flacsoandes.edu.ec/libros/digital/41444.pdf> (accessed April 24, 2023).
- Rea, A., & Rea, W. (2016). How Many Components should be Retained from a Multivariate Time Series PCA ?
- Schneider, T., Hampel, H., Mosquera, P. V., Tylmann, W., & Grosjean, M. (2018). Paleo-ENSO revisited: Ecuadorian Lake Pallcacocha does not reveal a conclusive El Niño signal. *Global and Planetary Change*, 168, 54–66. <https://doi.org/10.1016/j.gloplacha.2018.06.004>
- Schöngart, J., Junk, W. J., Piedade, M. T. F., Ayres, J. M., Hüttermann, A., & Worbes, M. (2004). Teleconnection between tree growth in the Amazonian floodplains and the El Niño–Southern Oscillation effect. *Global Change Biology*, 10(5), 683–692. <https://doi.org/10.1111/j.1529-8817.2003.00754.x>
- Sedano, R. (2017). *Influencia de la variabilidad climática y factores antrópicos en los extremos hidrológicos en el Valle Alto del río Cauca, Colombia [Influence of Climate Variability and Anthropogenic Factors on Hydrological Extremes in the Upper Cauca River Valley, Colombia]*. (Doctoral dissertation, Universitat Politècnica de Valencia).

- Shuhaida, I., & Shabri, A. (2014). Stream flow forecasting using principal component analysis and least square support vector machine. *Journal of Applied Science and Agriculture*, 9(11), 170–180.
- Sotomayor, G., Hampel, H., & Vázquez, R. F. (2018). Water quality assessment with emphasis in parameter optimisation using pattern recognition methods and genetic algorithm. *Water Research*, 130, 353–362. <https://doi.org/10.1016/j.watres.2017.12.010>
- Towner, J., Cloke, H. L., Lavado, W., Santini, W., Bazo, J., Coughlan de Perez, E., & Stephens, E. M. (2020). Attribution of Amazon floods to modes of climate variability: A review. *Meteorological Applications*, 27(5). <https://doi.org/10.1002/met.1949>
- Vega Vilca, J. C., & Guzman, J. (2011). Regresion PLS y PCA Como Solución al Problema de Multicolinealidad en Regresion Multiple. *Revista de Matemática: Teoría y Aplicaciones*, 18(1), 9. <https://doi.org/10.15517/rmta.v18i1.2111>
- Wang, C., & Enfield, D. B. (2001). The Tropical Western Hemisphere Warm Pool. *Geophysical Research Letters*, 28(8), 1635–1638. <https://doi.org/10.1029/2000GL011763>
- Ward, E., Buytaert, W., Peaver, L., & Wheeler, H. (2011). Evaluation of precipitation products over complex mountainous terrain: A water resources perspective. *Advances in Water Resources*, 34(10), 1222–1231. <https://doi.org/10.1016/j.advwatres.2011.05.007>
- Whitfield, P. H., Moore, R. D. (Dan), Fleming, S. W., & Zawadzki, A. (2010). Pacific Decadal Oscillation and the Hydroclimatology of Western Canada—Review and Prospects. *Canadian Water Resources Journal*, 35(1), 1–28. <https://doi.org/10.4296/cwrj3501001>