ISSN 1807-1929
Revista Brasileira de Engenharia Agrícola e Ambiental
Brazilian Journal of Agricultural and Environmental Engineering
v.28, n.9, e276836, 2024
Campina Grande, PB – http://www.agriambi.com.br – http://www.scielo.br/rbeaa

DOI: http://dx.doi.org/10.1590/1807-1929/agriambi.v28n9e276836 LITERATURE REVIEW

Streamflow regionalization in Brazil: Traditional methods and state of the art[1]

Regionalização de vazão no Brasil: Métodos tradicionais e estado da arte

Sérgio N. Duarte[2*], Wagner Wolff[2], Jéssica G. Nascimento[3], Tárcio R. Lopes[4], Thaís da S. Charles[2], Patrícia A. A. Marques[2], Adriano B. Pacheco[5] & Hugo C. Ricardo[2]

[1] Research developed at Universidade de São Paulo, Escola Superior de Agricultura "Luiz de Queiroz", Piracicaba, SP, Brazil
[2] Universidade de São Paulo/Escola Superior de Agricultura "Luiz de Queiroz", Piracicaba, SP, Brazil
[3] University of Nebraska/Nebraska Innovation Campus/Robert B. Daugherty Institute, Linlcon, USA
[4] Universidade Estadual de Maringá/Campus Umuarama, Umuarama, PR, Brazil
[5] Universidade Federal Rural da Amazônia/Campus Tomé-Açú, Tomé-Açú, PR, Brazil

HIGHLIGHTS:

Nowadays, no one can think of regionalization without GIS tools.
Traditional methods remain the most widely used in practice.
Machine learning methods have great potential in this area.

ABSTRACT: Water resources management aims to solve problems arising from intensive use of water. The proper management of this resource is based on understanding water availability, often using information from hydrometric stations; flow data is the most important information. The availability of information on river flows is often insufficient for all regions of interest. A technique called hydrological regionalization can be an alternative for obtaining information on streamflow. The objective of this study was to review the main regionalization techniques used, their advantages and limitations, as well as perspectives for the future. Traditional and widely used methods for forecasting hydrological variable, such as spatial proximity and multiple linear regression, were addressed, as well as new technologies, such as the geostatistical approach, techniques using volume balance in watersheds based on remote sensing products, and machine learning techniques. These techniques allow working with several physical characteristics of basins, generally ensuring better performances than the multiple linear regression. Further advancements in this area of knowledge are expected shortly, as the great potential of machine learning has been explored only to a small extent for hydrological regionalization purposes.

Key words: reference flows, watersheds, hydrological data

RESUMO: A gestão dos recursos hídricos visa resolver problemas decorrentes do uso intensivo da água. A gestão adequada deste recurso baseia-se no conhecimento sobre a disponibilidade hídrica dos cursos de água, e para isso utiliza-se informação de estações hidrométricas. Dentre essas informações, a vazão é a mais importante. A disponibilidade de informações sobre as vazões dos rios muitas vezes não é suficiente para todas as regiões de interesse. Para solucionar esse problema, existe uma técnica chamada regionalização hidrológica, que é uma alternativa para obtenção de informações sobre as vazões. Este estudo busca revisar as principais técnicas de regionalização utilizadas, suas vantagens e limitações, bem como perspectivas para o futuro. Abordam-se métodos tradicionais como o da proximidade espacial e o da regressão linear múltipla que são métodos mais antigos e amplamente utilizados, e novas tecnologias como abordagem geoestatística, técnicas que utilizam balanço de volume em bacias hidrográficas baseadas em produtos de sensoriamento remoto e técnicas de aprendizado de máquina. Estas técnicas permitem que se trabalhe com um maior número de características físicas das bacias, geralmente garantindo melhores performances que a regressão linear multivariada. Novos desenvolvimentos nesta última área do conhecimento são esperados em um futuro próximo, uma vez que o grande potencial do aprendizado de máquina foi explorado apenas em pequeno grau para propósitos de regionalização hidrológica.

Palavras-chave: vazões de referência, bacias hidrográficas, dados hidrológicos

• Ref. 276836 – Received 21 Jul, 2023
* Corresponding author - E-mail: snduarte@usp.br
• Accepted 21 Apr, 2024 • Published 10 Jun, 2024
Editors: Ítalo Herbet Lucena Cavalcante & Carlos Alberto Vieira de Azevedo

Introduction

Climate change and constant increase in population growth in recent decades have raised concerns about the availability of water for its multiple uses (Florêncio et al., 2019; Carvalho et al., 2020). Water management must efficiently minimize potential conflicts, balancing the demands of different human activities and avoiding environmental degradation (Dinh & Dang, 2022).

Water availability in a watershed is determined by information on river flows (Cecílio et al., 2018; Lelis et al., 2020). Frequently used reference flows are Q_{90} and Q_{95}, which represent flows exceeded or equaled in 90 and 95% of the time, respectively, as well as the minimum 7-day average flow with a 10-year return period ($Q_{7,10}$). Additionally, the long-term mean flow (Q_m) is another essential hydrological parameter representing the highest flow that can be regulated (Wolff & Duarte, 2021).

Information on flow in watersheds can be considered a challenge for managing water resources in many countries (Novaes et al., 2007; Panthi et al., 2021; Lopes et al., 2022; Ribeiro at al., 2022). However, hydrological models allow the optimization of information obtained at gauged locations for predicting flow in ungauged watersheds. Streamflow predictions can be obtained by transferring historical data between watersheds using spatial proximity (SP) and multiple linear regression (MLR), which are pioneering methods; however, new technologies also can be used for this purpose, such as the geostatistical approach (GEA), techniques using volume balance in watersheds (VBW) from remote sensing products, and machine learning techniques (MLT).

In this context, the objective of this study was to review the main regionalization techniques used, their advantages and limitations, as well as perspectives for the future.

Spatial Proximity Method

Methods based on spatial proximity (SP) and multiple regression (MR) stand out among the models found in the literature (Arsenault et al., 2019; Yang et al., 2020; Guo et al., 2021). SP is based on the idea that nearby basins have similar flow pattern with similar hydrological characteristics and responses. According to this method, there is a direct transfer of specific flow from geographically close basins to ungauged basins (Eq. 1):

$$Y = q \cdot A + \varepsilon \quad (1)$$

where:

Y - dependent variable (flow of the target basin);

q - specific flow of the donor basin;

A - target basin area; and,

ε - model residuals.

Usually, the target basin is inserted within the donor basin in this method. Chaves et al. (2002) implemented four different situations for Eq. 1. Althoff et al. (2021) recommended that the target basin should not be smaller than one-third of the donor basin. The simplicity of the SP method allows it to be used by designers without further studies (Pruski et al., 2012; 2016).

Multiple Linear Regression

Multiple linear regression (MLR) is one of the oldest and most widely used methods for predicting hydrological variables at ungauged locations, using environmental physical descriptors associated with watersheds (Ribeiro et al., 2022). According to this method, dependent variables (usually streamflow) are calibrated to correlate to independent variables of watersheds (climatic and landscape physical attributes) to build empirical relationships that can be used to predict dependent variables in ungauged watersheds (Panthi et al., 2021).

A general MLR model for estimating a hydrological parameter is represented by Eq. 2:

$$Y = a_0 + a_1 \cdot x_1 + a_2 \cdot x_2 + a_k \cdot x_k + \varepsilon \quad (2)$$

where:

Y - dependent variable (streamflow);

x with 1, 2, …, k = physical descriptors (independent variables);

a with 0, 1, 2, …, k = regional coefficients; and,

ε - model residuals.

Wolff & Duarte (2021) used 24 independent variables in a study for the state of Santa Catarina, Brazil. The most used variables are area, annual rainfall, drainage density, and talweg length and declivity. Streamflow prediction by the MLR method and its variations (Pruski et al., 2015) generally show good performance, mainly when applied to hydrologically homogeneous regions with similar climatological and topographic characteristics and similar hydrological response trends, even when these regions are not geographically adjacent. Considering hydrologically homogeneous regions, linear functions usually provide a good approximation to regional models, but it is not the case in a larger area (Li et al., 2022); alternative methods may better predict flows in heterogeneous regions. Segundo Nascimento et al. (2021), the MLR method outperformed the Random Forest (RF) machine learning technique for reference minimum flow predictions in a subtropical Brazilian state, but this outcome is not very common (Saadi et al., 2019).

The identification of different groups of basins with similar runoff generation and hydrological response mechanisms is an important factor for regional analysis of runoff frequencies (Nascimento et al., 2021). Theoretically, regional homogeneity is required for the adjustment of regional distribution frequencies and the estimation of flow quantiles using the MLR technique. However, in practice, regionalization conducted with more appropriate homogeneity conditions has resulted in lower accuracy than regionalization conducted under conditions of less homogeneity.

Ahani et al. (2022) proposed a method to rank several different regionalization approaches and identify the most appropriate regionalization for the analysis and estimation

of flow frequencies based on homogeneity, accuracy of quantile estimation, and region size. The results showed an acceptable performance of the MLR method in identifying the regionalization that provides the most appropriate homogeneity, accurate quantile estimation, and a satisfactory average size of homogeneous regions. Furthermore, the study results indicated a relative compatibility of the homogeneity ratio-based classification method with the classification based on quantile estimation accuracy. According to these results, the most appropriate regionalization of 41 basins classified by the K-nearest neighbor method in the study area included three homogeneous regions obtained by clustering algorithms.

Panthi et al. (2021) developed an MLR-based model to estimate the streamflow at several segments of a flow duration curve, incorporating basin physical characteristics and different climate data for each segment. They analyzed the sensitivity of proximity and characteristics between the donor (basins with gauging stations) and the receptor (ungauged basins) for historical streamflow data series in Nepal. The results show that regionalization techniques perform better for medium to low flows (Q_{90} and Q_{95}) than for high flows (Q5 and Q10).

Flow duration curve models are essential components in water resources study, particularly in poorly gauged or ungauged basins (Wolff & Duarte, 2021). Identifying these models in arid regions, where intermittent or ephemeral watercourses are common, constitutes an additional challenge for regionalization techniques (Jahanshahi et al., 2022; Nogueira Filho et al., 2022). Costa et al. (2020) proposed an alternative approach for modeling duration curves under zero flow conditions. Initially, they evaluated a set of frequently used flexible statistical distributions in hydrology for modeling the entire range of observed flow duration curves, selecting the eBXII distribution. Subsequently, regional models for the parameters of this distribution were identified using the evolutionary polynomial regression technique, which provided structurally complex equations but could identify the relationships between those parameters and the basins, and more importantly, could predict them in a cross-validation process, outperforming other simpler MLR techniques.

Minimum reference flows are important tools for assessing water availability in rural communities, especially those facing issues of over-extraction of water. However, the lack of flow data and gauging points has leading to the use of regionalization methods to predict the minimum reference flows required to maintain water uses. Basso et al. (2022) conducted a survey covering 92 river basins and 46 selected rural communities in a state in Brazil. Twenty-one basins were selected for having flow gauging stations and allowing for the estimation of Q_{95} using three different MLR-based methodologies. The results showed significant variation between the measured values and those estimated by the three methodologies; however, statistical analyses showed that regression equations from the methodology used by the state's official management agency (based on specific flow estimation) were more suitable for application in rural river basins in the state, mainly in larger ones. Similarly, Lelis et al. (2020) found discrepancies among MLR-based regionalization models applied in studies in other Brazilian states.

Despite advances in data collection and modeling, Althoff et al. (2021) reported a significant lack of streamflow forecasting in poorly gauged basins. They presented an approach using gridded data to regionalize flow rates (Q_m and Q_{95}) along rivers covered by the grid. The methodology was based on using the Terrain Analysis Using Digital Elevation Model (TauDEM) tool to obtain the necessary input variables for regionalization regression for each pixel of the study area, including average slope, total annual rainfall from remote sensing, and total annual evapotranspiration corresponding to each basin in the grid. These variables were suitable for the resolution of the Multi-Error-Removed Improved-Terrain (MERIT) Digital Elevation Model (DEM) (90 × 90 m). The result was a 90 × 90 m river grid with corresponding flow rates (mm per day). Thus, the gridded products improved the grid's ability to capture missing spatial patterns when using poorly instrumented basins.

Geostatistical Methods

Streamflow regionalization is a technique used in areas with scarce or nonexistent hydrological data. Although numerous studies have been developed to improve this technique, unsatisfactory results are still frequent. In addition to streamflow regionalization methodologies based on SP or MLR techniques applied to hydrologically homogeneous regions (Costa et al., 2020; Vieira et al., 2022), other alternatives include methodologies based on automatic interpolation and extrapolation techniques within a geographic information system environment (Wolff & Duarte, 2021).

The geostatistical approach (GEA) has the following advantages: (i) it does not depend on the input of the mean annual rainfall to perform calculations, as it can be spatially distributed; (ii) it is not linked to the determination of hydrologically homogeneous regions; and (iii) it can hypothetically be applied to basins of all sizes, although further studies should be conducted to corroborate this hypothesis (Wolff et al., 2014).

The general GEA model for estimating a hydrological parameter is represented by Eq. 3:

$$Y = a_0 + a_1 \cdot x_1 + a_2 \cdot x_2 + a_k \cdot x_k + S + \varepsilon \quad (3)$$

where:

Y - dependent variable (flow);

x with 1, 2, …, k = physical descriptors (independent variables);

a with 0, 1, 2, …, k = regional coefficients;

S - random effects accounted for spatial correlation; and,

ε - model residuals.

S represents the distance effect for the data; its determination is beyond the scope of the present review research. Wolff et al. (2014) used a GEA approach to develop comprehensive hydrological regionalization models for a state in Brazil and found better results than the traditional regionalization through MLR-based models applied to homogeneous regions by the state's official management agency in the 1980s.

Flow duration curves provide quick and direct information on the pattern of water resources in a basin. Thus, estimating the flow duration curve is important for basins with limited or no monitoring data for seasonal or annual periods. Using GEA to predict flow duration curves for ungauged locations represents a significant advancement in this research field. However, few results have been found, generally overestimates (positive trends), particularly for low flows (Jahanshahi et al., 2022).

Wolff & Duarte (2021) estimated flow duration curves for full and seasonal periods using a GEA-based model and streamflow data from 81 stations to obtain unbiased predictions. These stations had a high spatial density and were well distributed throughout the study area. Twenty-four independent variables were used to describe the basins: compactness coefficient; total stream length; main channel length; drainage density; highest distance difference; shape factor; first-order stream frequency; mean elevation; mean slope; drainage area; number of streams; topological diameter; perimeter; elongation ratio; area ratio; bifurcation ratio; circularity ratio; length ratio; slope ratio; annual precipitation; and spring, summer, autumn, and winter precipitations.

Wolff & Duarte (2021) applied geostatistical modeling to map flow duration curves (FDC) and, consequently, all quantiles and shape and scale parameters. Initially, they analyzed some basic assumptions of FDC parameters, including data normality and spatial stationarity. Subsequently, a maximum likelihood inference was performed to fit the geostatistical models and estimate the best shapes, comparing these GEA models with traditional models considered standards (MLR models). Finally, spatial interpolation was carried out and the performance was tested by using leave-one-out cross-validation. The GEA models fit and performed better than the MLR models. The medians of the relative residuals for the full and seasonal periods were unbiased for the entire duration. Possibly, the GEA fixed effect associated with the external deviations led to this better and unbiased result (Wolff & Duarte, 2021).

However, the position and proximity of basins and a high station density alone are not sufficient for good GEA modeling. According Jahanshahi et al. (2022), the effectiveness of model transferability depends on the proper selection of information from pairs of donor and receptor basins. Thus, they used rainfall-runoff models to evaluate two types of hydrological similarity: (I) apparent similarity, which was assessed using a similarity distance based on basin descriptive variables and the Euclidean distance based on a physical similarity method; and (II) behavioral similarity, which was determined by the best performance of parameter transfer models between the gauged donor basin and the ungauged receiver basin (best donation alternative). They proposed verifying the validation of the hypothesis that apparently similar basins in terms of descriptive variables have a hydrologically similar pattern. Spatial proximity was also implemented to evaluate its use as an alternative to physical similarity between basins. The HBV rainfall-runoff conceptual model was used in 576 basins located in four climatically distinct regions in Iran to test this hypothesis.

Finally, the results indicated that: (1) the best donor basin had the best performance, as expected, and more than 75% of similar basins exhibited hydrological similarity; (2) physical similarity outperformed proximity similarity, showing that descriptive physical variables more significantly affected transferability within each climatic region than geographical distance; however, the spatial proximity increased as the distance between the donor and target basins decreased (less than or equal to 20 km); (3) considering the spatial proximity method and consistency with basin physical characteristics, the geographical distance has a variable effect on model transferability depending on the region's climate, with spatial proximity resulting in better performance in humid than in dry regions; (4) overall, the prevailing transfer model varies from region to region in Iran; thus, climatic (aridity or potential evapotranspiration to precipitation ratio), topographic (mean elevation), and physiographic (basin area) properties have a greater effect on the transferability to for ungauged basins than other variables; and (5) the runoff ratio (Q_m / precipitation) confirms the superiority of humid regions over arid regions in terms of controlling transfer parameters (Jahanshahi et al., 2022).

Methods for Volumetric Water Balance Using Satellite Products

Recently, a remote sensing application in a recent hydrology study resulted in the development of well-performing models for monitoring and estimating variables of interest (Nascimento et al., 2021). Additionally, these tools allow for the spatialization of important variables, such as precipitation (PPT) and evapotranspiration (ET), which are essential for improving hydrological models to estimate essential information for the management of water resources (Charles et al., 2022; Moura et al., 2022).

Conventional methods for predicting river discharges require a large amount of hydrological and meteorological data. Measuring these data is costly and time-consuming, making it a challenging process (Singh et al., 2018). Thus, the hydrology research community has applied several methodologies to estimate river discharges, using available data from stream gauges to develop hydrological models based on PPT and ET of watersheds (Junges et al., 2022; Manke et al., 2022).

A simple model for estimating flow (Q_m) is through a water balance equation applied to the watershed, in which the water budget is calculated by subtracting ET, discharge, and positive soil water storage from PPT (the water inflow), thus representing the water outflow of the watershed. Thus, the flow can be obtained when the other components of water balance are known. The methodology involves estimating the annual flow (Q_m) based on the value obtained by subtracting ET from the annual PPT, using a simplified water balance equation.

A general water balance model (WBM) to estimate the hydrological parameter Q_m is represented by Eq. 4:

$$Q_m = PPT - ET + \varepsilon \quad (4)$$

where:

Q_m - mean flow;

PPT - annual precipitation;
ET - actual annual evapotranspiration; and,
ε - model residuals.

Data from Integrated Multi-satellitE Retrievals for Global Precipitation Measurement - GPM (IMERG) and Atmosphere-Land Exchange Inverse for ET mapping (ALEXI) can be used as remote sensing products in WBM, while Q_m is estimated as the residual.

The Q_m estimated by Nascimento et al. (2021) using WBM was successfully compared with observed data from stream gauges in each watershed. The main advantages of this methodology are its simplicity and good performance, as well as the free availability of PPT and ET data, especially in regions without gauging stations. They analyzed the effect of area, slope, and vegetation cover type on the model's performance in estimating Q_m. Similar performance was observed when considering the effect of different areas and slope percentages; however, the vegetation cover affected the model's performance. WBM using remote sensing products performed better in watersheds with a higher percentage of forest and pasture areas (>25 and 15%, respectively), and a smaller percentage of soybean areas (≤15%). WBM overestimated Q_m in the watersheds in the study region, which is reasonable, as changes in soil water storage were not considered. The estimation of Q_m by a simple WBM using remote sensing products is an important hydrological tool for water resources management, with the potential to use the same approach in other watersheds with different climatological and topographic characteristics. The uncertainty of IMERG and ALEXI products may result in uncertainty in the water balance estimate. However, the good performance of the model in estimating Q_m using only remote sensing products supports the recommendation of this method for hydrological forecasting in watersheds (Pereira at al., 2016).

Ribeiro et al. (2022) proposed and evaluated the performance of new exploratory regionalization variables, which represent the river flow formation process based on ET obtained from remote sensing products. They used the regional regression method to estimate Q_m and Q_{90}. The explanatory variables were: precipitation volume (Peq); the value obtained by subtracting an empirical value of 750 mm from Peq; the volumetric water balance for each stream segment; and the volumetric water balance for each hydrologically homogeneous region. These variables were obtained by combining drainage area, PPT, and ET data. ET was obtained using two remote sensing products: MOD 16 and Global Land Evaporation Amsterdam Model. The streamflow regionalization models were evaluated by statistical, physical, and risk analyses. The study was applied to the Rio Grande River Basin in the Southeast Region of Brazil. All variables, expect Peq, showed good relevance and representativeness in streamflow regionalization for considering variations in edaphoclimatic and vegetative conditions along the basin area. This study showed that step of selecting independent variables is significantly important, as a large number of independent variables is often not necessary to obtain a well-performing model.

Machine Learning Methods

Machine learning techniques (MLT) have interested hydrologists in recent years due to their ability to work with big data and solve several problems. Random Forest (RF) is an MLT developed by Breiman (2001) that can be used for prediction and classification purposes. RF regression can work with nonlinear relationship between variables by combining many regression trees, extracting multiple bootstrap samples from the original training data, and by analyzing decision trees (Breiman, 2001; Xu et al., 2019).

A decision tree is a hierarchical analysis diagram where each internal node represents an independent variable, the branch represents the result of the test, and each terminal node (leaf) represents a decision (Xu at al., 2019). The decision rules for node splitting are adjusted aiming to optimize the homogeneity of the dependent variable. Further details can be found in Tyralis et al. (2019) (Figure 1).

The use of RF for water resources management is recent (Tyralis et al., 2019); RF has been used for predicting water prices (Xu et al., 2019), regionalizing parameters of hourly hydrological models (Saadi et al., 2019), simulating large-scale flood discharges, and for several hydrological parameters and signatures (Schoppa et al., 2020). However, few studies focused on the applicability of RF to predict specific quantiles along the flow duration curve. According to Tyralis et al. (2019), RF-based models allowed the interpretation of obtained results and can complement other approaches. Additionally, most RF variants have been implemented in R programming language and are freely available.

Considering the importance of specific exceedance frequencies in flow duration curves (Q_{90} and Q_{95}) and long-term mean flow (Q_m) for water resources management, Nascimento et al. (2021) analyzed the performance of MLR and RF models in predicting these flows using a large-scale sample composed of 81 watersheds in a Brazilian state. They focused on reference flows because they are used as a tool, specifically, in the study area, thus effectively contributing to water resources management by providing models and identifying relevant descriptors for flow prediction in the region. Additionally, the study advanced the state of the art in hydrology by addressing the following questions: (i) Does RF overperform MLR in predicting reference flows in large-scale watershed? (ii) Does the low cost of RF, in terms of execution time and free software implementation, support its application for predicting reference flows? (iii) Can a subset of landscape and climatic descriptors be used for predicting long-term flow

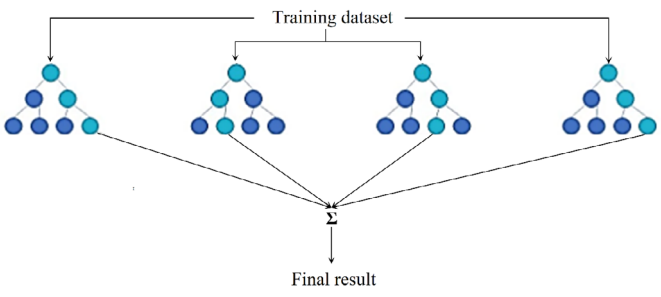

Figure 1. Hierarchical analysis diagram of the Random Forest algorithm

and minimum flows by RF models? They concluded that the MLR method outperformed RF, but their performances were similar in terms of error metrics. Thus, RF can be proposed as a new method for predicting reference flow in subtropical regions.

Golian et al. (2021) compared regionalization methods in 41 catchments in Ireland and concluded that MLR better estimated Q_m, whereas RF performed better in estimating high and low flows.

Ferreira et al. (2021) evaluated the advantages and limitations of different MLT approaches to regionalize flows in tropical watersheds. The models used were based on the RF, EARTH, MLR algorithms. The response variables were three types of minimum flows ($Q_{7,10}$; $Q_{95;}$ and Q_{90}) and Q_m. The database involved 76 covariates (independent variables) related to morphometry, topography, climate, land use and cover, and surface conditions. Two processes were used for elimination of covariates: Pearson correlation coefficient and recursive feature elimination (RFE). Several performance indices were used to validate the models. The results showed poor performance of the MLR model. The most predictive independent variable was the flow equivalent to precipitation volume, considering an abstraction factor of 750 mm. Overall, the RF and EARTH models showed similar performance and great ability to predict minimum flows and long-term mean flows.

Recently, deep learning (DL) models have shown state-of-the-art regionalization performance in gauged basin scenarios through the construction of a global hydrological model (Nogueira Filho et al., 2022). These models predict streamflow based on certain basin physical characterization variables and climate data. However, these descriptive variables are inherently uncertain and often incomplete, making them impractical in some cases, limiting their applicability. Li et al. (2022) showed that Random Vector (RV) models emerge as viable alternatives in the absence of physical descriptors of catchments. The results showed that RV models achieved predictive performance comparable to that of models using descriptive physical characteristics. The RV approach yields robust performance under different data scarcity conditions and DL model types. Additionally, the use of RV improves streamflow regionalization performance in gauged basins when physical descriptive variables are uncertain or insufficient.

According to Nogueira Filho et al. (2022) and Wang et al. (2022), modeling rainfall in ungauged basins remains a major challenge for hydrological research. A new approach to this issue is the Long-Short-Term-Memory neural network from the DL toolkit, with which few works on rainfall flow regionalization have been developed. Nogueira Filho et al. (2022) discussed the application of this new procedure on powerful computers compared to a traditional neural network and a MLR model in a practical framework under adverse conditions: limited data availability, shallow soil basins in semiarid regions, a high variability in rainfall and monthly time stages. The selected watersheds were in a state of the Northeast Region of Brazil. Regionalization by both neural networks had better performance when compared to the MLR procedure; however, the traditional neural network had slightly better relative performance. The neural network methods also showed the ability to aggregate understanding processes for different watersheds, as neural networks trained with MLR regionalization data (hybrid model) performed better when compared to networks trained for individual watersheds.

Conclusions

1. The spatial proximity method is simple and does not require previous studies of the region, but it is generally limited to target basins inserted in the donor basin.

2. The traditional method of multiple linear regression generally provides good solutions, but often requires multivariate cluster analysis.

3. The geostatistical method typically replaces multiple linear regression under conditions of dense basin grids, with relatively uniform distribution in the study area.

4. Models that perform volumetric water balance in watersheds based on satellite products generally provide good preliminary estimates of long-term mean flow (Q_m) for large regions with a small number of gauging stations.

5. The current state of the art lies in machine learning models, such as Random Forest, which allow working with a large number of physical characteristics.

Contribution of authors: Sergio N. Duarte: Research conception and administration. Wagner Wolff: Research conception. Jéssica G. Nascimento: Literature review. Tárcio R. Lopes: Data analisys. Thaís S. Charles: Work supervision. Patrícia A. A. Marques: Article preparation. Adriano B. Pacheco: Data analisys. Hugo C. Ricardo: Data colection..

Supplementary documents: There are no supplementary sources.

Conflict of interest: The authors declare no conflict of interest.

Financing statement: There was no funding for this research.

Literature Cited

Ahani, A.; Nadoushani, S. S.; Moridi, A. A ranking method for regionalization streamflow. Journal of Hydrology, v.609, p.1-10, 2022. https://doi.org/10.1016/j.jhydrol.2022.127740

Althoff, D.; Ribeiro, R. B.; Rodrigues, L. N. Gauging and ungauged: Regionalization flow indices at grid level. Journal of Hydrologic Engineering, v.26, p.1-11, 2021. https://doi.org/10.1061/(ASCE)HE.1943-5584.0002067

Arsenault, R.; Breton-Dufor, M.; Poulin, A.; Dallaire, G.; Romero-Lopez, R. Streamflow prediction in ungauged basins: analysis of regionalization methods in a hydrologically heterogeneous region of Mexico. Hydrological Sciences Journal, v.64, p.1297-1311, 2019. http://doi.org/10.1080/02626667.2019.1639716

Basso, R.; Honório, M.; Costa, I.; Bezerra, N.; Baumann, L.; Silva, F.; Albuquerque, A.; Scalize, P. Comparison between regionalized minimum reference flow and one-site measurements in hydrographic basins of rural communities in the state of Goiás, Brazil. Water, v.14, p.1-15, 2022. https://doi.org/10.3390/w14071016

Breiman, L. Random forests. Machine Learning, v.45, p.5-32, 2001. Available on: https://link.springer.com/content/pdf/10.1023/A:1010933404324.pdf. Acessed on: August 12 2022.

Carvalho, A. A. de; Montenegro, A. A. de A.; Silva, H. P. da; Lopes, I.; Morais, J. E. F. de; Silva, T. G. F da. Trends of rainfall and temperature in Northeast Brazil. Revista Brasileira de Engenharia Agrícola e Ambiental, v.24, p.15-23, 2020. http://dx.doi.org/10.1590/1807-1929/agriambi.v24n1p15-23

Cecílio, R. A.; Zanetti, S. S.; Gasparini, K. A. C.; Catrinck, C. N. Avaliação de métodos para regionalização das vazões mínimas e médias na bacia do rio Itapemirim. Scientia Agraria, v.19, p.122-132, 2018. http://dx.doi.org/10.5380/rsa.v19i2.52726

Charles, T. da S.; Lopes, T. R.; Duarte, S. N.; Nascimento, J. G.; Ricardo, H. de C.; Pacheco, A. B. Estimating average annual rainfall by ordinary kriging and TRMM precipitation products in midwestern Brazil. Journal of South American Earth Sciences, v.118, p.1-12, 2022. https://doi.org/10.1016/j.jsames.2022.103937

Chaves, H. M. L.; Rosa, J. W. C.; Vadas, R. G.; Oliveira, R. V. T. Regionalização de vazões mínimas em bacias através de interpolação em sistemas de informações geográficas. Revista Brasileira de Recursos Hídricos, v.7, p.43-51, 2002. http://dx.doi.org/10.21168/rbrh.v7n3.p43-51

Costa, V.; Fernandes, W.; Starick, A. Identifying regional models for flow duration curves with evolutionary polynomial regression: Application for intermittent streams. Journal of Hydrologic Engineering, v.25, p.1-12, 2020. https://doi.org/10.1061/(ASCE)HE.1943-5584.0001873

Dinh, T. K. H.; Dang, T. A. Potential risks of climate variability on rice cultivation regions in the Mekong Delta, Vietnam. Revista Brasileira de Engenharia Agrícola e Ambiental, v.26, p.348-355, 2022. http://dx.doi.org/10.1590/1807-1929/agriambi.v26n5p348-355

Ferreira, R. G.; Silva, D. D. da; Eleshon, A. A. A.; Fernandes Filho, E. I.; Veloso, G. V.; Fraga, M. de S; Ferreira, L. B. Machine learning models for streamflow regionalization in a tropical watershed. Journal of Environmental Management, v.280, 111713, 2021. https://doi.org/10.1016/J.JENVMAN.2020.111713

Florêncio, G. W. L.; Martins, F. B.; Ferreira, M. C.; Pereira, R. A. A. Impacts of climatic changes on the vegetative development of olive cultivars. Revista Brasileira de Engenharia Agrícola e Ambiental, v.23, p.641-647, 2019. http://dx.doi.org/10.1590/1807-1929/agriambi.v23n9p641-647

Golian, S.; Murphy, C.; Meresa, H. Regionalization of hydrological models for flow estimation in ungauged catchments in Ireland. Journal of Hydrology: Regional Studies, v.36, p.1-17, 2021. https://doi.org/10.1016/J.EJRH.2021.100859

Guo, Y.; Zhang, Y.; Zhang, L.; Wang, Z. Regionalization of hydrological modeling for predicting streamflow in ungauged catchments: A comprehensive review. Wiley Interdisciplinary Reviews-Water, v.8, p.1-32, 2021. http://doi.org/10.1002/wat2.1487

Jahanshahi, A.; Patil, S. D.; Goharian, E. Identifying most relevant controls on catchment hydrological similarity using model transferability - A comprehensive study in Iran. Journal of Hydrology, v.612, p.1-13, 2022. https://doi.org/10.1016/J.JHYDROL.2022.128193

Junges, A. H.; Bremm, C.; Fontana, D. C. Rainfall climatology variability, and trends in Veranopolis, Rio Grande do Sul, Brazil. Revista Brasileira de Engenharia Agrícola e Ambiental, v.23, p.160-166, 2022. http://dx.doi.org/10.1590/1807-1929/agriambi.v23n3p160-166

Lelis, L. C. da S.; Nascimento, J. G.; Duarte, S. N.; Pacheco, A. B.; Bosquilia, R. W.; Wolff, W. Assessment of hydrological regionalization methodologies for the upper Jaguari River basin. Journal of South American Earth Sciences, v.97, p.1-10, 2020. https://doi.org/10.1016/J.JSAMES.2019.102402

Li, X.; Khandelwal, A.; Jia, X.; Cutler, K.; Glosh, R.; Renganathan, A.; Xu, S.; Tayal, K.; Nieber, J.; Duffy, C.; Steinbach, M.; Kumar, V. Regionalization in a global deep learning model: from physical descriptors to random vectors. Water Resources Research, v.58, p. 1-19, 2022. https://doi.org/10.1029/2021WR031794

Lopes, T. R.; Nascimento, J. G.; Pacheco, A. B.; Duarte, S. N.; Neale, C. M. U.; Folegatti, M. V. Estimation of sediment production and soil loss in a water supply basin for the metropolitan region of São Paulo - Brazil. Journal of South American Earth Sciences, v.118, p.1-14, 2022. https://doi.org/10.1016/j.jsames.2022.103929

Manke, E. B.; Teixeira-Gandra, C. F. A.; Damé, R. de C. F.; Nunes, A. B.; Chagas-Neta, M. C. C.; Karsburg, R. M. Seasonal intensity-duration-frequency relationships for Pelotas, Rio Grande do Sul, Brazil. Revista Brasileira de Engenharia Agrícola e Ambiental, v.26, p.85-90, 2022. http://dx.doi.org/10.1590/1807-1929/agriambi.v26n2p85-90

Moura, M. de P.; Ribeiro Neto, A.; Costa, F. A. da. Application of satellite imagery to update depth-area-volume relationships in reservoirs in the semiarid region of northeast Brazil. Revista Brasileira de Engenharia Agrícola e Ambiental, v.26, p.44-50, 2022. http://dx.doi.org/10.1590/1807-1929/agriambi.v26n1p44-50

Nascimento, J. G.; Althoff, D.; Bazame, H. C.; Neale, C. M. U.; Duarte, S. N.; Ruhoff, A. L.; Gonçalves, I. Z. Evaluating the latest IMERG products in a subtropical climate: the case of Paraná State, Brazil. Remote Sensing, v.13, p.1-18, 2021. https://doi.org/10.3390/rs13050906

Nogueira Filho, F. J. M.; Souza Filho, F. de A.; Porto, V. C.; Rocha, R. V.; Estácio, A. B. S.; Martins, E. S. P. R. Deep learning for streamflow regionalization for ungauged basins: Application of long-short-term-memory cells in semiarid regions. Water, v.14, p.1-21, 2022. https://doi.org/10.3390/w14091318

Novaes, L. F. de; Pruski, F. F.; Queiroz, D. Q. de; Rodriguez, R. del G.; Silva, D. D.; Ramos, M. M. Avaliação do desempenho de cinco metodologias de regionalização de vazões. Revista Brasileira de Recursos Hídricos, v.12, p.51-61, 2007. https://doi.org/10.21168/rbrh.v12n2.p51-61

Panthi, J.; Talchabhadel, R.; Ghimire, G. R.; Sharma, S.; Dahal, P.; Baniya, R.; Boving, T.; Pradhanang, S. M.; Parajuli, B. Hydrologic regionalization under data scarcity: Implications for streamflow prediction. Journal of Hydrologic Engineering, v.26, p.1-11, 2021. https://doi.org/10.1061/(ASCE)HE.1943-5584.0002121

Pereira, D. dos R.; Martinez, M. A.; Silva, D. D. da; Pruski, F. F. Hydrological simulation in a basin of typical tropical climate and soil using the SWAT Model Part II: Simulation of hydrological variables and soil use scenarios. Journal of Hydrology: Regional Studies, v.5, p.149-163, 2016. https://doi.org/10.1016/j.ejrh.2015.11.008

Pruski, F. F.; Nunes, A. de A.; Rego, F. S.; Souza, M. F. Extrapolação de equações de vazões mínimas: Alternativas para atenuar os riscos. Water Resources and Irrigation Management, v.1, p.1-10, 2012. Available on: https://www3.ufrb.edu.br/seer/index.php/wrim/article/view/1587. Accessed on: August 12 2022.

Pruski, F. F.; Rodriguez, R. D. G.; Nunes, A. A.; Pruski, P. L.; Singh, V. P. Low-flow estimates in regions of extrapolation of the regionalization equations: A new concept. Engenharia Agrícola, v.35, p.808-816, 2015. http://dx.doi.org/10.1590/1809-4430-Eng.Agric.v35n5p808-816/2015

Pruski, F. F.; Rodriguez, R. del G.; Pruski, P. L.; Nunes, A. de A.; Rego, F. S. Extrapolation of regionalization equations for long-term average flow. Engenharia Agrícola, v.36, p.830-838, 2016. https://doi.org/10.1590/1809-4430-Eng.Agric.v36n5p830-838/2016

Ribeiro, R. B.; Pruski, F. F.; Oliveira, J. S.de; Filgueiras, R.; Althoff, D.; Pinto, E. J. de A. Stream flow regionalization considering water balance with actual evapotranspiration estimated from remote sensing. Journal of Hydrologic Engineering, v.27, p.1-14, 2022. https://doi.org/10.1061/(ASCE)HE.1943-5584.0002183

Saadi, M.; Oudin, L.; Ribstein, P. Random Forest ability in regionalizing hourly hydrological model parameters. Water, v.11, p.1-22, 2019. https://doi.org/10.3390/w11081540

Schoppa, L.; Disse, M.; Bachmair, S. Evaluating the performance of Random Forest for large-scale flood discharge simulation forest for large-scale flood discharge simulation. Journal of Hydrology, v.590, p.1-13, 2020. https://doi.org/10.1016/j.jhydrol.2020.125531

Singh, V. P.; Yadav, S.; Yadava, R. N. Hydrologic modeling: Select proceedings of ICWEES-2016. v.81, 2018. Springer. Available on: https://www.springer.com/gp/book/9789811058004. Accessed on: August 12 2022.

Tyralis, H.; Papacharalampous, G.; Langousis, A. A brief review of Random Forests for water scientists and practitioners and their recent history in water resources. Water, v.11, p.1-37, 2019. https://doi.org/10.3390/w11050910

Vieira, P. R.; Pruski, F. F.; Souza, J. R. C. Dimensioning of reservoirs for semiarid regions using synthetic series. Revista Brasileira de Engenharia Agrícola e Ambiental, v.24, p.581-589, 2022. http://dx.doi.org/10.1590/1807-1929/agriambi.v24n9p581-589

Xu, Z.; Lian, J.; Bin, L.; Hua, K.; Xu, K.; Chan, H. Y. Water price prediction for increasing market efficiency using Random Forest regression: A case study in the western United States. Water, v.11, p.1-10, 2019. https://doi.org/10.3390/w11020228

Yang, Y.; Magnusson, J.; Huang, S.; Beldring, S.; Xu, C.-Y. Dependence of regionalization methods on the complexity of hydrological models in multiple climatic regions. Journal of Hydrology, v.582, p.124357, 2020. http://doi.org/10.1016/j.jhydrol.2019.124357

Wang, S.; Peng, H.; Hu, Q.; Jiang, M. Analysis of runoff generation driving factors based on hydrological model and interpretable machine learning methods. Journal of Hydrology: Regional Studies, v.42, p.1-16, 2022. https://doi.org/10.1016/J.EJRH.2022.101139

Wolff, W.; Duarte, S. N.; Mingoti, R. Nova metodologia de regionalização de vazões, estudo de caso para o Estado de São Paulo. Revista Brasileira de Recursos Hídricos, v.19, p.21-33, 2014. Avaiable on: https://abrh.s3.sa-east-1.amazonaws.com/Revistas/173. Acessed on: August 12 2023.

Wolff, W.; Duarte, S. N. Toward geostatistical unbiased predictions of flow duration curves at ungauged basins. Advances in Water Resources, v.152, p.1-13, 2021. https://doi.org/10.1016/J.ADVWATRES.2021.103915