

<https://doi.org/10.1590/2318-0331.241920180114>

## Regionalization of hydrological model parameters for the semi-arid region of the northeast Brazil

### *Regionalização de parâmetros de modelo hidrológico para a região semiárida do nordeste brasileiro*

Múcio Valença Virães<sup>1</sup>  and José Almir Cirilo<sup>2</sup> 

<sup>1</sup>Companhia de Pesquisa de Recursos Minerais, Serviço Geológico do Brasil, Recife, PE, Brasil

<sup>2</sup>Universidade Federal de Pernambuco, Caruaru, PE, Brasil

E-mails: [mucio.viraes@cprm.gov.br](mailto:mucio.viraes@cprm.gov.br) (MVV), [almir.cirilo@gmail.com](mailto:almir.cirilo@gmail.com) (JAC)

Received: July 07, 2018 - Revised: August 08, 2019 - Accept: September 10, 2019

## RESUMO

Este estudo analisa a regionalização de parâmetros de modelos hidrológicos em regiões semiáridas com escassez de dados, tendo como foco parte do semiárido brasileiro que envolve os Estados do Ceará, Rio Grande do Norte, Paraíba e Pernambuco. O modelo hidrológico adotado foi o MODHAC - Modelo Hidrológico Auto-Calibrável. Foram selecionadas 45 estações fluviométricas com um mínimo de oito anos consecutivos de dados consistidos a partir do ano 2000, além dos registros de anos anteriores. O principal critério para regionalização foi a proximidade entre as estações fluviométricas disponíveis, sendo também avaliada a escolha com base em propriedades físicas e climáticas para algumas estações. Foi feita análise de fatores que afetam a qualidade das séries de vazões, bem como da rede de monitoramento pluviométrico. A influência de reservatórios sobre esses dados e a desativação de muitas estações de monitoramento das precipitações foram os problemas mais frequentes. Considerando diferentes indicadores de desempenho, como o Coeficiente de Nash para a raiz quadrada das vazões satisfatório em 63% dos casos, de acordo com valores de referência apresentados na literatura, o procedimento teve razoável sucesso para quantificar as vazões mais elevadas e médias, porém não apresentou em geral bom ajuste das vazões mais baixas.

**Palavras-chave:** Modelagem hidrológica; MODHAC; Regionalização de vazões; Hidrologia de regiões semiáridas.

## ABSTRACT

This work deals with the regionalization of parameters of hydrological models in semi-arid regions with data scarcity, focusing on part of the Brazilian semi-arid region that involves the States of Ceará, Rio Grande do Norte, Paraíba and Pernambuco. The hydrological model adopted was the MODHAC - Self-Calibrating Hydrological Model. A total of 45 fluviometric gauge stations were selected with a minimum of eight consecutive years after 2000, in addition to the records from previous years. The main regionalization criterion was the proximity between the available flow gauge stations, also evaluating the choice based on physical and climatic properties for some stations. An analysis was made of factors that affect the quality of the flow series, as well as the rainfall gauge station network. The influence of reservoirs on these data and the deactivation of many rainfall stations were the most frequent problems. Considering different performance indicators, such as the Nash Efficiency Coefficient for the flow square root that was satisfactory in 63% of the cases, according to reference values presented in the literature the procedure was reasonably successful in quantifying the higher and medium flow, but it did not present in general a sound adjustment of the lower flows.

**Keywords:** Hydrological modeling; MODHAC; Flow regionalization; Semi-arid regions hydrology.



## INTRODUCTION

Hydrological models have been used to simulate flow series from actual or predicted rainfall. However, such models require calibration of their parameters with pluviometric and fluvimetric series of adequate duration and quality. But it is not always possible to have series with available data under these conditions, and thereby it is imprecise to establish the calibration parameters of these models to simulate water availability in basins or sub-basins without flow data. Thus, seeking to regionalize the parameters of the hydrological model obtained in one basin to simulate the water availability of another similar basin, in which there is no available hydrological series, is a strategy that can bring outstanding results for the water resources planning. This paper aims to analyze the possibility of regionalizing parameters of a hydrological model with several known applications in studies and projects in the Brazilian semi-arid region and discuss the challenges regarding them.

Pilgrim et al. (1988) explain that arid and especially semi-arid regions are often in delicate hydrological equilibrium. The whole hydrological nature (and thus the values of the model parameter) can be altered by prolonged sequences of wet or dry weather. In these regions, rainfall, in turn, tends to be more variable both in space and time than in wetter areas.

According to Huang et al. (2016), most of the hydrological models described in the literature on this theme can well represent runoff in humid regions, but this does not happen when dealing with watersheds in semi-arid territories. Al-Qurashi et al. (2008), referencing several authors, attribute this difficulty to the general scarcity of data on precipitation, flow, soil properties and initial humidity conditions; the influence of seasonal and interannual vegetation variability; the complexity of the watercourse morphology; the difficulty of quantifying overflow losses; and inaccuracy in estimating potential evaporation.

The term regionalization is used to denote the transfer of information from one river basin to another. According to Samuel et al. (2011), most regionalization methods of hydrological models parameter are based on three sets of criteria: physical similarity of the basin, spatial proximity and regression-based approaches. A detailed analysis of the regionalization experiences of hydrological models parameters is presented by Beck et al. (2016). The authors describe experiments in different parts of the world involving models with 1 to 28 parameters to be calibrated. In their work, they apply the HBV model (BERGSTRÖM, 1992; SEIBERT; VIS, 2012), with 14 parameters to be calibrated for 1787 watersheds in different regions of the world with areas from 10 to 10,000 km<sup>2</sup>, considering the physio-climatic factors in the regionalization analysis. Based on performance criteria the authors considered the results satisfactory, particularly for basins that are less than 5,000 km away from those for which the parameters were calibrated. Beck et al. (2016) discuss various regionalization approaches that seek to identify the flow signatures (such as the slope of the permanence curve and the base flow rate). The literature on the topic records an extensive discussion about the results obtained from different regionalization criteria. Many authors highlight sound results with the transfer of parameter sets according to a particular measure of climatic and / or physiographic similarity (KOKKONEN et al., 2003; MCINTYRE et al., 2005;

PARAJKA et al., 2005; OUDIN et al., 2008; LI et al., 2009; REICHL et al., 2009; WALLNER et al., 2013; SINGH et al., 2014; SELLAMI et al., 2014; GARAMBOIS et al., 2015).

Even under similar conditions, the scale of the studied watershed often becomes a limiting factor for the transfer of information. This implies that a smaller sub-basin that is in a region homogeneous for several larger basins may not have hydrological behavior similar to the larger basins in which it is inserted, as shown by the scale effect studies in the research developed by Girardi et al. (2011).

Other authors, on the other hand, report many analyzes in which the transfer of parameter values, based only on the physical proximity of the watersheds, led to good results in the regionalization process. Patil and Stieglitz (2014) discuss regionalization criteria based on physical-climatic similarity and proximity between the basins, recording that in many applications there is no clear evidence of advantage for one criterion or another. Oudin et al. (2008), Zhang et al. (2015), for river basins in Austria, France, and Australia, reported better results in the transfer of hydrological models parameters when the proximity between the basin was taken into account. Petheram et al. (2012) reached the same conclusion in a study about the flow regionalization for watersheds with low-density of hydrological monitoring in Northern Australia.

Several authors proposed the regionalization of flow permanence curves as the hydrological response or one of the “signatures” of the watershed (LIU; HAN 2010; PINHEIRO; NAGHETTINI, 2010; WESTERBERG et al., 2011; PRAFULLA; YILMAZ; GUPTA, 2012; COSTA et al., 2014). In this last article, the authors use the regionalization methodology of permanence curves for watersheds in two Brazilian regions. One of them is formed by intermittent rivers of the Ceará state, but the results were inferior for those obtained for basins with perennial rivers.

## Runoff modeling in semi-arid regions

The semi-arid region has been a major challenge for the application of hydrological models due to data scarcity, especially of flows. According to McIntyre and Al-Qurashi (2009), the selection of a rainfall-flow model for application in an arid or semi-arid region should consider the spatial characteristics of precipitation, the variability, and nonlinearity of losses, and the availability and quality of data.

There are not many hydrological models developed especially for arid or semi-arid regions. Then, to overcome this difficulty, several studies use models developed for general application, with no specificity for semi-arid regions. Al-Qurashi et al. (2008) seeking to represent the soil and rainfall spatial variability during 27 hydrological events applied the Kineros 2 distributed model to an arid watershed in Oman. The authors conclude based on performance indicators that the model validation was not satisfactory for all events and all calibration strategies that were tested to evaluate the highest flow rates. They state that their results are consistent with the experience of other hydrology modelers in arid and semi-arid climate regions and that further scientific research is needed, especially regarding the rainfall observation and spatial modeling.

Another experiment with hydrological modeling in the semi-arid is cited by Adam et al. (2017) in their study of the El Hawad watershed in the semi-arid region of Sudan, where the average annual rainfall is about 250 mm. In this work was used the SCS Curve Number model. Then, according to the authors, this method enabled the estimation of runoff depths consistent with the rainfall for different soil moisture conditions. Another case is presented in the study of Kan et al. (2017) for nine Chinese river basins. Three of them are in humid regions, three in semi-arid semi-humid regions and three in arid regions. Kan et al. (2017) investigated the application of three hydrological models in flood forecasting: XAJ (Xinjiang Model, based on the runoff generation mechanism due to excess of saturation, which is suitable for wet and semi-humid regions), NS (Northern Shaanxi Model, based on the flow generation due to excess of infiltration supposedly common in arid areas) and MIX (its mechanism is based on vertical mixing, combining excess saturation and excess infiltration in the flow generation).

The results confirmed the complexity of the drier basins for flood forecasting. All models tested worked satisfactorily in wet watersheds, and only the NS model was applicable in arid watersheds. The XAJ and MIX models, which consider excess of saturation in the flow generation, showed a better performance than the NS model, which is based only on the excess of infiltration for the case of the addressed semi-arid and semi-humid basins.

A study using the HEC-HMS hydrological model is presented by Wang et al. (2016), in the Hailiutu watershed, semi-arid northwestern China, calibrated from 1978 to 1992. In this case, the model systematically underestimated the flows in winter and spring and some flows in summer. This is due to the discrepancy between the nonlinear rainfall-flow response in the basin and the linear structure of the Soil Moisture Accounting (SMA) model. Another example is dealt in the work of Traore et al. (2014) in the Koulountou River basin, a tributary of the Gambia River, located in the Republic of Guinea-Conakry. This study involved two hydrological models: GR4J and GR2M, which were developed by the French Research Institute for Agriculture and Environmental Engineering (CEMAGREF) today IRSTEA - National Research Institute for Science and Technology for the Environment and Agriculture. According to the authors, the modeling obtained satisfactory results from the established criteria of the Nash Coefficient and its variations, with square root variations and decimal logarithms greater than 0.7 in both models, when compared observed and calculated flows. These results led to the establishment of parameters of these models that would allow restoring missing flows from rainfall.

McIntyre and Al-Qurashi (2009) present a discussion of the IHACRES model, which was defined by the authors as semi-empirical and was developed jointly by the Institute of Hydrology (IH) in the United Kingdom and the Center for Resource and Environmental Studies (CRES) of the Australian National University. The justification of the model is based on the calculation of a unit hydrograph for the total runoff after deducting losses in nonlinear relationship with precipitation. The model was initially designed for watersheds in a temperate climate and was progressively adapted to represent runoff in temporary rivers. It is currently presented in a concentrated and semi-distributed version, with daily or

sub-daily time steps. The authors report the application in several semi-arid basins of Australia and Oman, pointing out that the best results are obtained using the semi-distributed model. However, the authors emphasize the limitation of this type of model when the quantity and quality of hydrological data are not adequate.

Regarding modeling experiences for the Brazilian semi-arid, Cabral et al. (2017), for example, have satisfactorily calibrated the hydrological model HEC-HMS (Hydrologic Engineering Center – Hydrologic Modeling System) in a region between Zona da Mata and Agreste, in the São Miguel river basin in the State of Alagoas. This method was used to simulate flow rates by both observed and estimated radar precipitation to compare them with the observed flow. However, their results showed that the flow estimated by precipitation, obtained by radar, underestimated the magnitude of the peak flow and volume, while adequately representing the time of the peak flow with sound Nash-Sutcliffe coefficient values (0.75 - 0.79). An experience in the semi-arid in the State of Paraíba is described in the study of Felix and Paz (2016) in the Piancó basin. They used the drainage area of the river station of same name as a research sub-basin, simulating the flows from the model distributed by MGB-IPH (Large Basin Model) modules. This study reveals the difficulties of the model in representing the lower flows of the Piancó River. The permanence curve is observed at 5% of the time, when the flows exceed the value of 50m<sup>3</sup>/s, happening a very sound adherence between simulated and measured flows according to the authors. On the other hand, in approximately 40% of the time there is no runoff, but the model maintains a residual flow. This behavior was also identified by other authors discussed in this article, due to the application of a type of model that does not present mathematical formulation for rivers with characteristics of semi-arid regions.

An experiment on the evaluation of general models with suitability for semi-arid regions was developed by Huang et al. (2016). Four classic hydrological models were selected and applied to three semi-arid basins in northern China. They were: TOPMODEL, Xinjiang (XAJ), SAC-SMA and Tank. Based on the analysis and comparison of the results of these classic models, the authors developed four new models designated by them as flexible and seeking to improve modeling responses. The application of flexible models aims to identify the dominant flow processes.

Considering the classic models, the authors believe that the high flexibility of the nonlinear components in the XAJ and SAC-SMA models may be the reason why these models had a better performed than the TOPMODEL in the semi-arid basins.

### *MODHAC simulation model*

The MODHAC (Modelo Hidrológico Auto-Calibrável - Self-Calibrating Hydrological Model) (Lanna, 1997), although it is commonly used as the vast majority of rainfall-flow models, was also developed to meet semi-arid characteristics (Lanna, 1997). The model was used for hydrological studies of the Ceará State Water Resources Plan and Master Plans of several basins in Bahia. Martins et al. (2006) used the model in an operational study of reservoirs of the Jaguaribe-Metropolitan system in Ceará. Due to the various applications, this was the model chosen to evaluate

the regionalization possibility in the semi-arid area of the present study, although it is a model of general use.

The selection of the model, study areas, and data used in this research had a significant influence on the Northeast Atlas project (ANA, 2006). This project is under the coordination of ANA (Agência Nacional de Águas) - National Water Agency and involved numerous actors, such as water resources state bodies and institutions. For the Atlas hydrological studies, regarding the regions of the states of Ceará, Rio Grande do Norte, Paraíba and Pernambuco, the MODHAC hydrological model was used to generate estimations of water availability provided by several reservoirs of the Northeast region.

In the case of Atlas, in turn, the model was chosen based on the studies of the São Francisco water transfer project (Brasil, 2001; ENGEORPS; HARZA, 2000). In the MODHAC were used parameters adjusted in the Atlas studies, but complementing the calibrations for areas not contemplated or with adjustments considered unsatisfactory (initially based on the  $R^2$  and  $Pbias$  performance indicators), as well as analyzing the application of these parameters to most recent data series.

The MODHAC is a lumped hydrological model, i.e., it does not explicitly consider the spatial variability of the hydrological process and the physiographic characteristics of the basin, but it can be used as a semi-distributed model in a basin composed of interconnected sub-basins. This model simulates the terrestrial phase of the hydrological cycle, that is, the transformation process that leads to the calculation of river flows. It was opted

for the heuristic calibration of the parameters to ensure a better sensitivity and refinement, but mainly to enjoy the experiences of the mentioned applications, without using the automatic option for such purpose.

The water storage process in the basin is simulated through three fictitious reservoirs: surface storage, subsurface storage and underground storage. Precipitation goes through an initial corrective filter, which is needed due to the rises from deficiencies in obtaining data because of the low rainfall or systematic reading errors. Afterward, this precipitation feeds the potential evapotranspiration (PET), which may be or not supply. If water remains in this process, then the surface reservoir (vegetation and topographic depressions) will be fed, where the unmet PET will be reevaluated. This situation will occur even without rain if the reservoir is not empty. If this reservoir overflows, will happen a runoff, and therefore, also a percolation to the subsurface and underground tank. The water from the subsurface reservoir will be able to meet the remaining PET, besides being responsible for the base runoff, and in case of overflow, it will attend the hypodermic runoff. The water from the underground reservoir will also form the base runoff and, in the event of an overflow, the water will flow into the subsurface reservoir that may serve its runoff. Figure 1 schematically demonstrates the processes of the hydrological cycle in the MODHAC.

The model requires the calibration of 14 parameters, described in Table 1. Despite the large number, some of these can be considered constant or even disregarded in the calibration

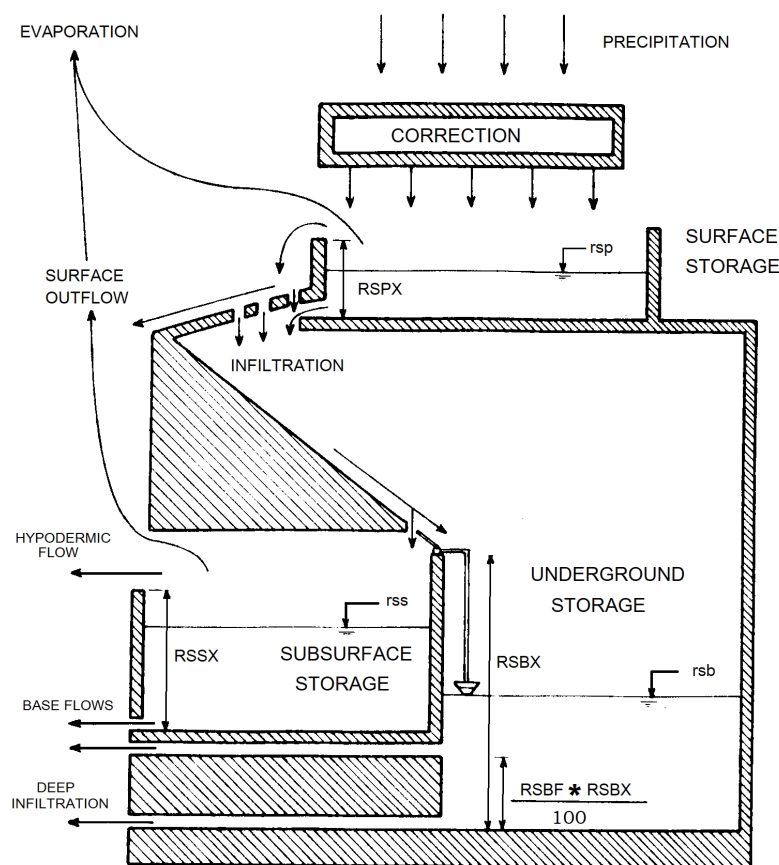


Figure 1. Schematic representation of the processes of the hydrological cycle in MODHAC. Source: (Lanna, 1997).

**Table 1.** Description of the MODHAC parameters.

Parameters	Designation
RSPX	Maximum capacity of the surface reservoir [mm].
RSSX	Maximum capacity of the subsurface reservoir [mm].
RSBX	Maximum capacity of the underground reservoir [mm].
RSBY	Minimum storage to occur contribution from the underground reservoir to base flow [mm].
IMAX	Soil permeability [mm].
IMIN	Minimum infiltration [mm].
IDEC	Empirical infiltration coefficient (0-1) [dimensionless].
ASP	Exponent of the surface reservoir emptying law [dimensionless].
ASS	Exponent of the subsurface reservoir emptying law [dimensionless].
ASBX	Exponent of the underground reservoir emptying law [dimensionless].
ASBY	Adjustment of the hydrographic recession curve [dimensionless].
PRED	Precipitation correction [dimensionless].
CEVA	Soil evapotranspiration [dimensionless].
CHET	ETP fraction supplied directly from the rain [dimensionless].

due to the characteristics of the basin that will be modeled, thus reducing the adjustment work.

The underground reservoir simulates water storage in the lower soil layers. It encompasses the underground aquifer and gives rise to a more delayed underground or basal runoff that supports the runoff in the drought. This reservoir has two capacities that should be to consider. These are RSBX and RSBY, which are parameters of the model. RSBX is the full capacity and RSBY controls the leaks, or contributions to the runoff promoted by this reservoir. When storage is less than RSBY the water leakage will be controlled by the ASBX parameter and when it is higher, is controlled by ASBY.

## Study area and hydrological data

### *Physical characteristics of the Brazilian semi-arid region and limitations of hydrological data*

The semi-arid regions, as most of the Brazilian Northeast, are areas with large volumetric, temporal and spatial rainfall irregularities. For this reason, several strategies are adopted to meet the regularization of water flow throughout the dry period. This is the case of thousands of dams built for the accumulation and redistribution of water for various purposes during the drought (CIRILO et al., 2017). Historical series of hydrological data do not always exist due to the lack of monitoring or are not long enough, especially the level observations and the monitoring campaigns for the establishment of rule curves and then the determination of flow rates. Besides, the existence of reservoirs alters the natural conditions of runoff, being necessary the identification of this effect when simulations are performed.

The northeastern semi-arid region is characterized by intermittent flow in most of its rivers, having the flow interrupted during most of the year and only becoming perennial in areas where they reach wetter regions, that is, near the river mouth on the Atlantic ocean. Moreover, the crystalline basement that occupies more than 80% of the region has characteristics not contemplated by most hydrological models, in the aspect of river-aquifer accumulation and water exchange. It is observed, especially

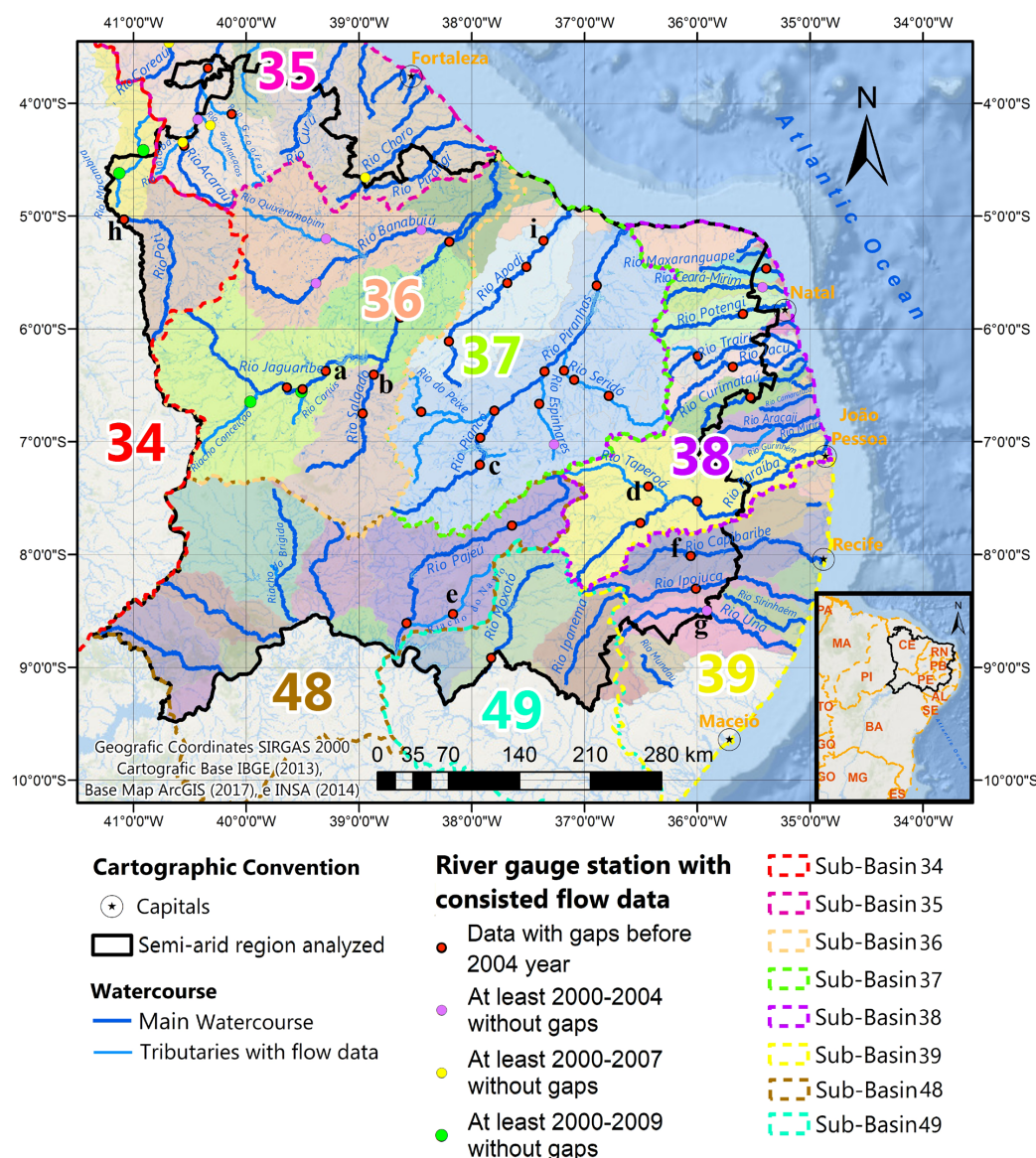
when the simulations are developed with a time step of one day or less, that the flow ceases soon after the precipitation occurs, characterizing the almost absence of storage and underground flow. In addition, runoff losses can be considerable (recharging the cracking of the crystalline stratum, for example). Therefore, based on the described characteristics, the modeling is restricted to simulate the surface runoff, because the groundwater component is impaired by the physical environment characteristics.

Regarding the hydrological monitoring networks, the low density of fluviometric stations with continuous data series of adequate duration for the elaboration of studies and projects should be recorded. The rainfall network has already been constituted by a significant number of measuring stations at the time that the SUDENE - Northeast Development Superintendence had a fruitful performance in the sector, until the 80's of the last century. However, most of these stations have been deactivated. There is currently an effort by ANA - National Water Agency and related state agencies to rebuild the hydrometric network, but there is not yet significant historical series.

Another relevant aspect is the presence of dams along most of the rivers, built from the sevenths on last century. The flow series of the rivers are therefore affected by the operation of the reservoirs. Thus, hydrological analyzes should consider both physical interventions in the basins and the reduction of rainfall monitoring stations.

The studied region is part of the Brazilian northeastern semi-arid, belonging to the states of Ceará, Rio Grande do Norte, Paraíba and Pernambuco, located between the parallels 03° 30' and 09° 30' S and meridians 41° 30' and 35° 13' W, within the UTM Zone 24 South for most of its part. This territory has an area of 281,965 km<sup>2</sup> that occupies 108,956 km<sup>2</sup> or 73.20% of Ceará, 48,752 km<sup>2</sup> or 92.22% of Rio Grande do Norte, 45,788 km<sup>2</sup> or 81.01% of Paraíba and 78,469 km<sup>2</sup> or 79.91% of Pernambuco, involving 79.11% of the total area of these states, 17.75% of the Northeast area and 3.31% of Brazil's land area, as shown in Figure 2.

The wettest quarter occurs in February, March and April in most of the states of Ceará, Rio Grande do Norte (except in its eastern portion) and in the western and central regions of Paraíba. In Pernambuco the wettest quarter is divided into regions: in the



**Figure 2.** Study area in northeastern semi-arid boundaries, according to cartographic data of IBGE (2013) and INSA (2014). Fluviometric reference stations: Iguatu (a), Icó (b), Piancó (c), Poço de Pedras (d), Ilha Grande (e), Toritama (f), Capivara (g), Oiticica (h) and Mossoró (i).

west, it rains most in January, February and March and on the east coast the most significant rainfall occurs in May, June and July. Between these territorial extremes of the state the rainy season is not well defined, depending on the atmospheric phenomena that generate the precipitation.

According to the Pluviometric Atlas of Brazil project, developed by CPRM (2011), the study area has a range of average annual rainfall isoietes between 400 mm and 1,300 mm. In general, the state of Ceará has average annual rainfall in the semi-arid region around 400 mm to 1,200 mm, with an average around 750 mm, for Rio Grande do Norte this value ranges between 500 mm and 1,300 mm, with an average of 632 mm. For Paraíba, the annual rainfall ranges from 400 mm to 1,000 mm, with an average of 636 mm, and Pernambuco ranging from 500 mm to 1,000 mm, with an average of 581 mm.

#### Fluviometric stations

From the ANA inventory and the database built with data from CPRM and ANA were selected 170 stations with flow data for the states of Ceará, Rio Grande do Norte, Paraíba, and Pernambuco. Stations located on the São Francisco River have a drainage area outside the study area and are not part of this research.

From the 170 stations were selected flows series with at least eight consecutive years of consisted data to represent the flows observed in MODHAC and that had drainage areas partially or entirely within the semi-arid polygon. This restriction led to the selection of 45 stations, which are represented in Figure 2 and with physical and climatic characteristics indicated in Table 2. However, of these 45 stations only four have consisted flow

**Table 2.** Fluviometric stations selected for analysis, where A is the station's drainage area, Qm is the average annual flow, PMA is the average annual rainfall,  $C_{esc}$  is the surface runoff coefficient, and  $C_{ris}$  is the percentage of the station's drainage area over the karstic and/or fractured aquifer.

Code	A (km <sup>2</sup> )	Qm (mm/ano)	PMA (mm)	$C_{esc}$	$C_{ris}$ (%)
34730000	1,031	80.37	654.9	0.123	0.0
34740000	2,216	57.03	791.1	0.072	0.0
34741000	10,730	32.86	682.3	0.048	92.2
35125000	1,498	193.00	989.9	0.195	96.4
35210000	1,563	102.80	758.3	0.136	100.0
35223000	698	214.45	1,011.0	0.212	82.6
35235000	3,653	113.72	844.6	0.135	95.7
35240000	1,530	96.81	747.2	0.130	100.0
35258000	2,375	29.69	736.9	0.040	100.0
35275000	11,273	155.22	811.4	0.191	95.9
35830000	1,599	89.89	809.2	0.111	100.0
36045000	3,454	24.38	684.0	0.036	93.3
36070000	13,671	27.11	649.9	0.042	95.8
36125000	3,632	44.73	739.3	0.061	80.4
36130000	6,183	59.84	826.9	0.072	77.8
36160000	20,824	40.70	717.0	0.057	89.3
36270000	8,910	81.58	885.3	0.092	47.2
36290000	12,203	66.41	894.2	0.074	59.1
36320000	39,689	43.32	787.1	0.055	79.0
36390000	48,283	62.74	780.5	0.080	80.7
36470000	4,525	63.26	780.5	0.081	100.0
36520000	7,038	49.47	656.2	0.075	100.0
36580000	17,237	193.00	705.3	0.274	99.5
37030000	2,197	77.72	709.1	0.110	98.8
37080000	6,588	61.91	728.1	0.085	85.9
37084000	7,690	16.31	718.6	0.023	85.2
37090000	9,786	26.70	698.3	0.038	86.6
37260000	1,451	55.99	826.3	0.068	63.5
37340000	4,555	83.93	704.7	0.119	96.4
37410000	15,212	78.60	725.1	0.108	89.1
37470000	21,639	75.74	686.1	0.110	91.6
37515000	1,868	33.12	497.8	0.067	99.7
37559000	6,316	6.95	504.0	0.014	96.1
37570000	9,714	41.02	517.8	0.079	97.3
37710150	38,061	50.35	616.6	0.082	93.3
38830000	5,033	31.60	597.2	0.053	100.0
38850000	3,181	38.71	537.1	0.072	100.0
38860000	13,648	80.25	520.7	0.154	99.6
39130000	2,443	31.67	529.4	0.060	100.0
39340000	1,973	45.66	582.6	0.078	100.0
39540000	2,628	44.45	592.1	0.075	100.0
48820000	3,537	88.42	617.4	0.143	99.8
48860000	12,276	68.74	622.8	0.110	87.9
48880000	2,263	26.21	583.3	0.045	89.7
49160000	8,241	16.40	589.1	0.028	58.9

data, without gaps, between 2000 and 2009 (10 years), four other stations with data between 2000 and 2007 (8 years), and six other fluviometric stations serving the period between 2000 and 2004 (5 years) without gaps with consistent flow rates. Thus, only eight stations attend a continuous monthly period of 8 years of consistent flows, without missing data, i.e., 96 consecutive monthly flow data, from the year 2000.

At the end, nine fluviometric stations were considered as reference for the regionalization: seven with data used in the Northeast Atlas: Iguatu (a) and Icó (b) in Ceará State, Piancó (c) and Poço de Pedras (d) in Paraíba State and Ilha Grande (e), Toritama (f), and Capivara (g) in the State of Pernambuco. Two other stations that were included in this study are Oiticica (h) in Ceará and Mossoró (i) in Rio Grande do Norte.

### Pluviometric stations

Two rainfall databases were used for the calculation of the average rainfall series weighted by the Thiessen polygon, which was the method used in this study. One from the Northeast Atlas project from 1933 to 2000, originally from 1933 to 2001, and another from available data from ANA, including more recent records for the compatibility of the studied period with the data of the fluviometric stations. The first database was generated with consistency analysis and series extension to standardize the data from 830 rainfall stations with the same duration. Inconsistencies were detected in the data from 2001 in 56 stations and thereby the series were reduced until 2000.

The second database with rainfall data from the year 2000 was obtained from ANA and CPRM. These data sources present discontinuity in some stations' rainfall series, and there was no treatment extending extinct station series or new facilities to maintain a single period to all. Therefore, it was decided to select rainfall stations that had at least one continuous data series of eight years for the construction of this second database.

Thus, the corrected Northeast Atlas project database has 830 rainfall stations from 1933 to 2000, and the second database, dating from a period after 2000, has data from 228 rainfall stations from ANA with at least the period between 2000 and 2007 without missing data.

### Evaporimetric data

The Potential evapotranspiration (PET) data used in the Northeast Atlas project were estimated from evaporation data, a methodology that was also adopted in this study. The calculations were performed according to the data available from the climatological station nearest to the drainage area of the fluviometric station to be modeled. Thus, most of the calibrations/simulations performed in the present study maintained the evapotranspiration data used in the Northeast Atlas project. But, for new simulations were estimated evapotranspiration data from the INMET (Brazilian National Institute of Meteorology) database.

## RESULTS

From the data selection for the specified time periods, the chosen hydrological model was calibrated, also taking into account the results of previous studies and then, through the cross-validation process seeking to identify its suitability or not for other basins.

### Calibration of the model parameters and performance indicators

In the calibration process, results from the Northeast Atlas project were used, when the selected basins coincided, and with new adjustments of parameter values for non-contemplated areas. The evaluation was done at a monthly level comparing the flow calculated by MODHAC with the flow observed from the corresponding fluviometric stations.

The results were evaluated by comparing the observed and the calculated flows by MODHAC, considering the following performance indicators:

NSE: Nash-Sutcliffe efficiency;

$r^2$ : determination coefficient;

RMSE: root mean square error;

Pbias%: percent bias;

RSR: ratio of the root mean square error to the standard deviation of measured data;

AAPE: Average absolute percentage error.

The Nash-Sutcliffe Efficiency Coefficient (NSE) ranges from  $-\infty$  to 1 and can be obtained from Equation (1).

$$NSE = 1 - \frac{\sum_{i=1}^n (Q_{i,obs} - Q_{i,cal})^2}{\sum_{i=1}^n (Q_{i,obs} - \overline{Q_{obs}})^2} \quad (1)$$

Where  $n$  is the total number of event data,  $Q_{i,obs}$  is the observed flow and  $Q_{i,cal}$  is the calculated flow, both at time  $i$ , while  $\overline{Q_{obs}}$  is the average flow observed in the period.

Other variations of this indicator have been used, such as the Nash-Sutcliffe Efficiency Ratio of the square root flow ( $NSE_{sqrtQ}$ ) and its logarithms ( $NSE_{logQ}$ ):

$$NSE_{sqrtQ} = 1 - \frac{\sum_{i=1}^n (\sqrt{Q_{i,obs}} - \sqrt{Q_{i,cal}})^2}{\sum_{i=1}^n (\sqrt{Q_{i,obs}} - \sqrt{\overline{Q_{obs}}})^2} \quad (2)$$

$$NSE_{logQ} = 1 - \frac{\sum_{i=1}^n (\log Q_{i,obs} - \log Q_{i,cal})^2}{\sum_{i=1}^n (\log Q_{i,obs} - \log \overline{Q_{obs}})^2} \quad (3)$$

RMSE, Pbias, AAPE, and RSR are obtained through the following Equations (4), (5), (6) and (7).

$$RMSE = \frac{1}{n} \left[ \sum_{i=1}^n (Q_{i,obs} - Q_{i,cal})^2 \right]^{1/2} \quad (4)$$

$$Pbias = \left[ \frac{\sum_{i=1}^n (Q_{cal}(i) - Q_{obs}(i))}{\sum_{i=1}^n (Q_{obs}(i))} \right] 100 \quad (5)$$

$$AAPE = \frac{1}{n} \cdot \sum_{i=1}^n \left| \frac{Q_{i,obs} - Q_{i,cal}}{Q_{i,obs}} \right| \quad (6)$$

$$RSR = \frac{\left[ \sum_{i=1}^n (Q_{i,obs} - Q_{i,cal})^2 \right]^{1/2}}{\left[ \sum_{i=1}^n (Q_{i,obs} - \overline{Q_{obs}})^2 \right]^{1/2}} \quad (7)$$

The vast majority of authors use the Nash-Sutcliffe efficiency coefficient as one of the main performance indicators, some of them are Cabral et al. (2017), Felix and Paz (2016) and Traore et al. (2014). Many articles address the values of the performance indicators individually, and thereby, few aggregate them together. Reference values for NSE, for example, have been referenced in the literature. According to Gotschalk and Motovilov (2000, apud Van Liew et al. (2007)), NSE values above 0.75 (M - Very Good) are considered very good, if between 0.75 and 0.36 (A - Acceptable) they are acceptable, and unacceptable if below 0.36 (I - Insufficient),



for both daily and monthly time steps. However, authors such as Zappa (2002) propose values above 0.5 for NSE.

On the other hand, it is recommended to associate the Nash-Sutcliffe efficiency coefficient values with other performance indicators. Table 3 refers to the minimum and maximum values of NSE, *Pbias*, and RSR for calibration and validation from various works compiled by Moriasi et al. (2007). The authors gathered works that aggregated several performance indicators and selected the three mentioned to represent the accuracy of the modeling jointly. A similar analysis developed by Van Liew et al. (2007) reaches the same performance reference ranges for the NSE and *Pbias* indicators.

In addition to the indicators cited, several authors use NSE derivations, like square root ( $NSE_{sqrtQ}$ ) and decimal logarithm of flows ( $NSE_{logQ}$ ) instead of the natural flows in the Nash-Sutcliffe formula, indicated in Equations (2) and

(3). According to Traore et al. (2014), NSE closer to the unit indicate the adjustment of the highest flows,  $NSE_{sqrtQ}$  indicate the average flows and  $NSE_{logQ}$  the lowest flows. Pushpalatha et al. (2012) analyze researches developed by several authors with other transformations on the NSE indicator and conclude that  $NSE_{sqrtQ}$  provides more adequate information about the errors made in the simulation of high and low flow rates, which agree with results of Oudin et al. (2006).

The calibration analysis is presented below, highlighting the external factors that caused distortions in the adjustments. The performance indicator values for the nine reference stations that were calculated for the model calibration are concentrated in Table 4. The analysis of each station is relevant to understand the peculiarities that must be found in the regionalization process. Table 5 presents the parameter values calibrated for the 9 reference gauge stations.

**Table 3.** Recommended performance values for monthly simulations.

Recommended Performance Values for Monthly Simulations	NSE	<i>Pbias</i> (%)	RSR
Very good	0.75 to 1.00	< ±10	0 to 0.5
Good	0.65 to 0.75	±10 to ±15	0.5 to 0.6
Satisfactory	0.50 to 0.65	±15 to ±25	0.6 to 0.7
Insufficient	< 0.50	> ±25	> 0.7

Source: Adapted from Moriasi et al. (2007).

**Table 4.** Best performance indicators for full series data of the reference gauge stations.

Code	Name	Parameters "donor basins"	NSE	R <sup>2</sup>	RMSE	AAPE	RSR	PBias	NSE $sqrtQ$	NSE $logQ$
34741000	Oiticica	Oiticica	0.55	0.64	7.75	1.15	0.67	-5.94	0.54	-5.52
36160000	Iguatu	Icó	0.81	0.81	4.97	19.57	0.44	0.69	0.81	0.53
36290000	Icó	Icó	0.72	0.76	7.26	6.38	0.53	4.49	0.75	0.59
37090000	Mossoró	Mossoró	0.59	0.78	4.50	4.03	0.64	-1.01	0.65	-0.31
37470000	Piancó	Icó	0.66	0.67	11.32	6.09	0.58	-10.95	0.76	0.67
38850000	Poço de Pedras	Piancó	0.60	0.62	7.70	1.46	0.63	-39.33	0.62	-2.42
39130000	Toritama	Capivara	0.39	0.59	7.85	18.02	0.78	25.39	0.41	-0.05
39540000	Capivara	Capivara	-0.66	0.49	11.36	2.84	1.29	37.11	0.34	0.11
48880000	Ilha Grande	Piancó	0.52	0.53	6.15	1.69	0.69	-40.82	0.31	-1.97

**Table 5.** Calibrated parameters of Iguatu (a), Icó (b) and Oiticica (h) stations in the state of Ceará, Piancó (c) and Poço de Pedras (d) in the state of Paraíba, Ilha Grande (e), Toritama (f), and Capivara (g) in the state of Pernambuco, and Mossoró (i) in the state of Rio Grande do Norte, which were selected for regionalization.

Parameters	Stations								
	a	b	c	d	e	f	g	h	i
RSPX	97	121	49.82	40	6.5	8.0	17.3	300	300
RSSX	376.6	519.3	530	300	288	239.2	103	200	300
RSBX	30.05	19.67	0.0	20	0.0	0.0	0.0	0	30
RSBY	145.8	145.8	0.0	20	0.0	0.0	0.0	0	1.0
IMAX	50	78.91	181	24.18	228.5	267.7	100	80	30
IMIN	0.4969	0.4969	2.0	7.0	31.65	8.813	0	0.5	0.5
IDEC	0.3025	0.2275	0.29	0.7964	0.534	0.4546	0.19	0.1	0.3
ASP	3.902	1.668	0.01	0.2525	0.4478	0.1776	0.9	0.5	100
ASS	0.1017	0.1017	0.08	0.9555	0.152	0.1	0.18	0.1	0.035
ASBX	0.4585	0.4585	0.0	0.5	0.0	0.0	0.0	0.0	0.0001
ASBY (*10 <sup>-3</sup> )	66,192	37.9	0.0	500	0.0	0.0	0.0	0.0	0.005
PRED	999	999	999	999	999	999.9	999	999	999
CEVA	0.99	1.0	0.7459	0.1	0.3	0.3	0.3	1.0	0.99
CHET	1.0	1.0	0.9	0.8	0.8	0.75	0.9	1.0	1.0

### Iguatu Station – 36160000

This station is located on the Jaguaribe river, in the Ceará state. The modeled periods were from 1962 to 1980, 1980 to 2000 and 2001 to 2009. Some factors should be taken into account for the period from 2001 to 2009:

1. Indications of rating curve change;
2. Missing data in periods of heavy rainfall;
3. Reduction in the number of rainfall stations from 55 to 14.

These issues greatly influenced the performance indicators of the adjustments. Registering NSE as a reference, the values were 0.89 for the period from 1962 to 1980; 0.57 for 1981 to 2000; -0.78 for 2001 to 2009; finally,  $NSE = 0.49$  for the complete series using the calibrated parameters for the station. In the end, as shown in Table 4, the parameters calibrated for the data of the Icó station led to the best result, meeting the criteria established by Moriasi et al. (2007) and also satisfactorily representing the minimum flow rates according to  $NSE_{logQ} > 0.5$ .

### Influence of evaporation data

To identify how much the estimated evapotranspiration data impact on modeling, the sensitivity of the model was evaluated according to the nature of these data. For this, the flows in the Iguatu station were simulated in two situations: using the INMET data from the Iguatu climatological station and using the evaporimetric averages of the four climatological stations adopted in the Northeast Atlas project. The INMET data are the

climatological normals obtained with Piché evaporimeter. Table 6 shows the considerable differences in evaporation values and Figure 3 their reflection on simulated flow rates. Hence the need to maintain the same potential evapotranspiration dataset when comparing results of parameter regionalization.

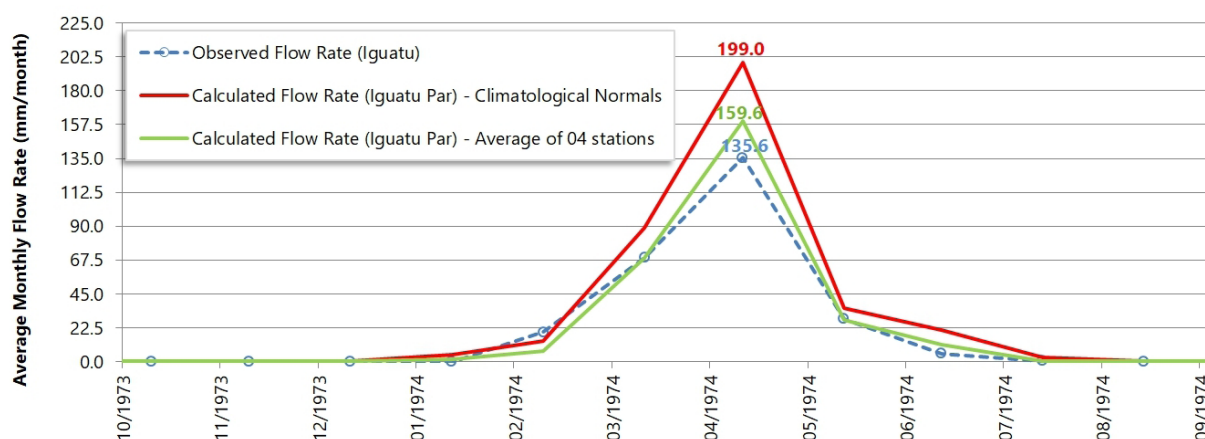
### Icó station - 36290000

The Icó station, located on the Salgado river, such as Iguatu, is also situated in the semi-arid region of Ceará. The periods studied here are subdivided from 1959 to 1987, 1988 to 1999 and 2000 to 2007.

There are some issues to be considered when analyzing flow data at this station, that is, the period from 1988 to 1999 presents loss in modeling quality. It was also found that from 1988 to 1993 the historical series showed zero flow between July and December, a situation that is not observed in the previous 20 years. On the other hand, analyzing the period from 1990 to 1996 (Figure 4) were observed flood peaks, but it was not identified the presence of rainfall compatible with this behavior. It happens that the affluent São João stream has its confluence upstream from where Icó station is. Then, the Lima Campos dam is located in this affluent and it is suspected that due to the occurrence of six years drought, this dam may have released water seasonally through the affluent to the Salgado River for regularization, which would explain the decrease in modeling efficiency for the period from 1959 to 1987. Nevertheless, the Icó station presented the best performance among all analysis with its calibrated values for the parameters approaching the optimal classification proposed by Moriasi et al. (2007), and having an adequate adjustment for the minimum flow rates.

**Table 6.** Evaporimetric data used to estimate the evapotranspiration for modeling the Iguatu station.

Jan	Feb	Mar	Apr	Mai	Jun	Jul	Aug	Set	Oct	Nov	Dec
Evaporimetric Data (Average of 04 basin stations - Northeast Atlas project)											
190.8	138.55	104.15	105.67	136.07	159.90	216.37	265.65	281.47	282.97	262.25	247.47
Evaporimetric Data (INMET Climatic Standard) - 82686 - Iguatu											
158.8	103.7	90.2	80.2	95.7	127.9	172.6	205.3	217.5	230.1	215.5	203.1



**Figure 3.** Flow rates calculated in the Iguatu station with MODHAC for evapotranspiration estimated by the INMET Climatological Normals of the station and by the average of 04 stations in the region (Northeast Atlas). Period 1973-1974.

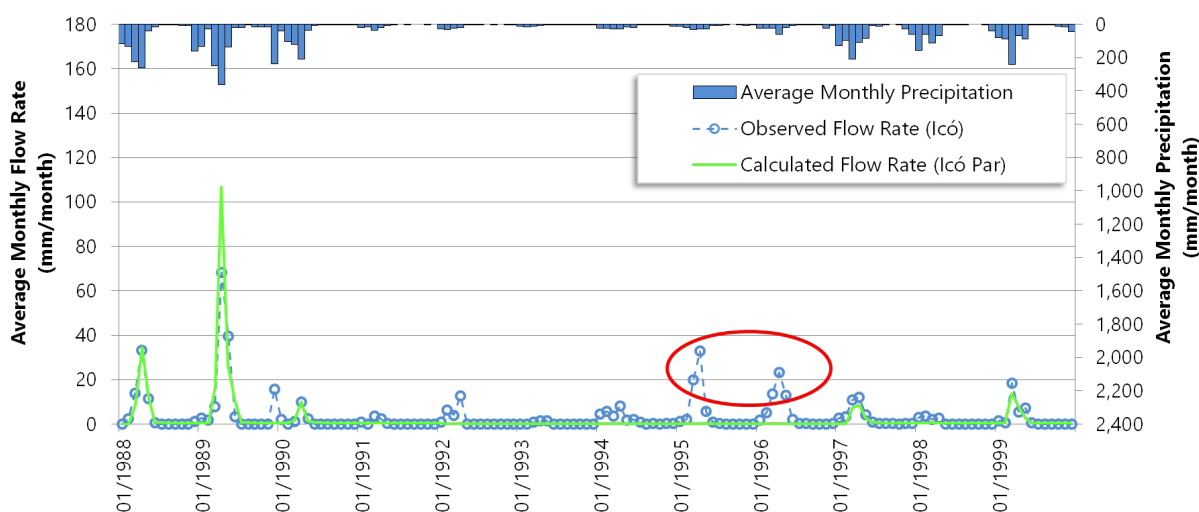


Figure 4. Modeling of the Icó station, period 1988-1999.

#### Oiticica station - 34741000

This station is far west in Ceará, in the Parnaíba river basin (Sub-Basin 34). Its installation occurred only in 2004. Then, for the study was selected the period of available consistent flow, which was between 2005 and 2011. The performance indicator values obtained for this station, presented in Table 4, are classified as satisfactory according to Moriasi criteria (2007). The  $NSE_{logQ}$  indicates that the adjustment of the low flows was not sound, a trend also reflected in the AAPE indicator, which is very restrictive, because it weights the flows equally, regardless of their magnitude.

#### Piancó station - 37340000

The Piancó fluviometric station is located in the Piranhas-Açu river basin. The periods from 1964 to 1981, 1982 to 1999 and 1999 to 2006 were considered. The change in the observation period did not affect the adjustment performance indicators, with satisfactory to outstanding results in all analyzed time intervals. According to Table 4 the adjustment is considered sound even for low flows.

#### Mossoró station - 37090000

This station is located in the Apodi basin, in Rio Grande do Norte semi-arid region. The most sensitive parameter in Mossoró modeling was RSPX, which deals with the maximum capacity of the superficial reservoir. As in other cases that were analyzed, this parameter seeks to compensate for the storage effect in upstream reservoirs.

Referring only to the Nash-Sutcliffe Coefficient, the value fell from 0.79 (period 1987-1998) to  $NSE = 0.20$  in the period from 1998 to 2005. The fall on performance is attributed to the reduction of 38 pluviometric stations with data consisting and only 11 stations in the subsequent period. Nevertheless, according to Table 4, the indicators suggested by Moriasi et al. (2007) for

the entire period were satisfactory. However, an unsatisfactory adjustment for low flow rates expressed by  $NSE_{logQ}$  was observed.

#### Poço de Pedras station - 38850000

The MODHAC simulations for this station data in the state of Paraíba, located in the Taperoá River, affluent of the Paraíba River, presented different behavior over the studied period. From 1986 on the quality of the adjustment worsened. This must have been due to the interference of the Taperoá II (1983) reservoir construction, which is upstream of the station. This fact is confirmed by observing the fall from  $8.14 \text{ m}^3/\text{s}$  to  $1.50 \text{ m}^3/\text{s}$  of the monthly average flow of the first period, even with the average monthly rainfall remaining at 47 mm for both periods in the station drainage area.

Even so, only the adjustment indicators  $Pbias$  and  $NSE_{logQ}$  were not satisfactory.

#### Toritama station - 39130000

The Toritama river gauge station in the State of Pernambuco is closest to the river head of the Capibaribe River, being located upstream of the Jucazinho Dam and downstream of the Poço Fundo reservoir. Between 1973 and 1986, the simulations results were sound ( $NSE = 0.76$ ), but they have worsened in the following years, probably due to Poço Fundo, which started operations in 1987.

The indicators proposed by Moriasi et al. (2007) presented in Table 4 for the whole period are slightly lower than the acceptable minimums. The negative  $NSE_{logQ}$  indicates that the adjustment of the low flow was not proper either.

#### Ilha Grande station - 48880000

This is a fluviometric station located in the Navio Creek, in the Pajeú river basin, downstream of the Barra do Juá reservoir, which started operating in 1982. The values obtained for the  $NSE_{logQ}$ ,  $NSE_{sqrtQ}$ , and  $Pbias$  performance indices were not satisfactory.

## Capivara station – 3954000

This fluviometric station is in the upper course of the Una river, in Central Agreste region of Pernambuco. The period selected for the studies were from 1978 to 1993 and 1998 to 2006. For the first adjustment was obtained a  $NSE = 0.82$ . In the transition periods, the availability of the rainfall stations decreased from 17 to only 3 stations, which may explain the reduction in the adjustment quality.

Considering the entire period, the adjustment was not satisfactory, as indicated in Table 4.

## Evaluation of parameter value transfer

In order to compare the characteristics of the basins under analysis, the following indicators are presented in Table 1: Area of the basin ( $\text{km}^2$ ); mean annual flow  $Q_m$  ( $\text{mm} / \text{year}$ ); Mean annual rainfall PMA ( $\text{mm}$ ); Cesc (runoff coefficient:  $Q_m / \text{PMA}$ ); and percentage of area located on fractured aquifer Cris (%). These calculations were made for the area controlled by each of the 45 fluviometric stations. Crystalline soils are predominant, which is typical of the semi-arid region of northeastern Brazil.

Among the problems mentioned, the ones that mainly stand out are the influence of reservoirs on the observed flow and the deactivation of a considerable portion of pluviometric stations in the most recent years. Therefore, given the problems regarding the flow and precipitation data series, it should be expected that climate similarity indicators will be influenced. Thus, it was decided to test the application validity of the calibrated parameters between the different regions based on the greater proximity between them. The process consists of using the set of parameter values calibrated for one station by simulating the flow series for other neighboring stations and evaluating the adjustment indicators.

A commonly raised issue concerns the block transfer of parameter values. Bárdossy (2007) and Oudin et al. (2008)

emphasize the fact that parameter values are interdependent and thus should be considered as a set that must be transferred entirely to maintain the correlations between the adjusted values. This interdependence also exists with evaporation data, especially when it comes to monthly level simulations, as occurs in this paper.

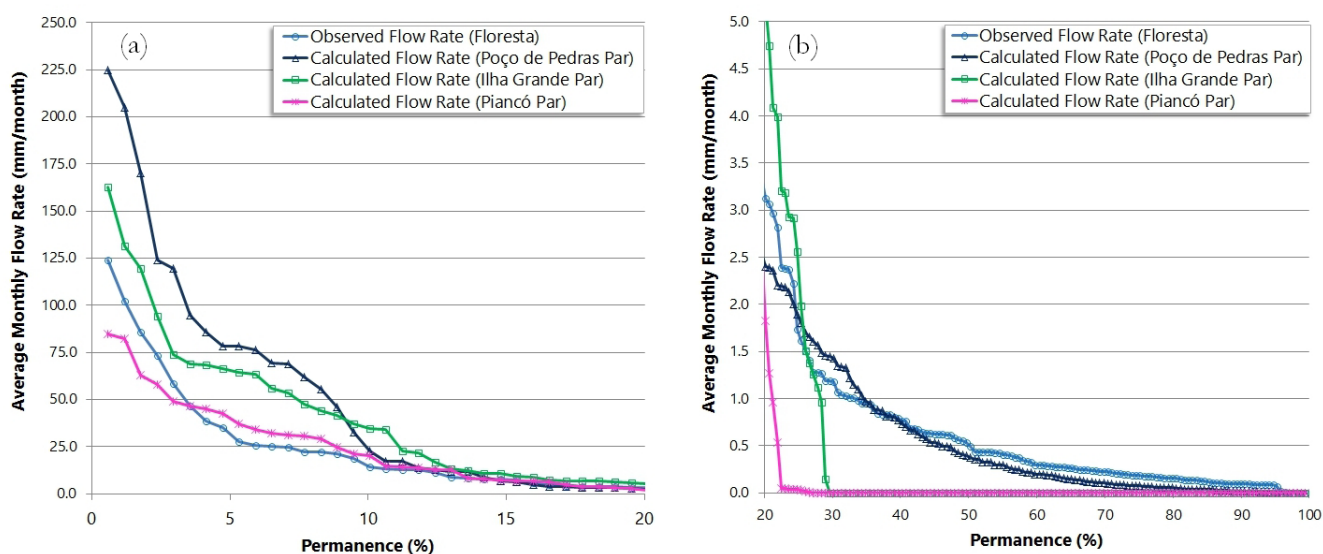
Patil and Stieglitz (2014), in turn, highlight that, as some parameters are more sensitive in the calibration process than others, the convenience of block parameter transfer should be evaluated according to the characteristics of the parameters.

## Regionalization result

The simulation results for several crossings parameter were analyzed. From the flow calibration at the selected stations, indicated in Table 2, the transfers of the parameter sets were tested to simulate the runoff in closer basins.

Figures 5a, 5b, 6a, and 6b represent the permanence curves of the flows observed at the Iguatu and Floresta stations (representing “receiving basins”) and the permanence curve generated with the calibrated parameters for the “donor basins”. In the case of Iguatu, the parameters come from calibration with data from the Icó station. To simulate the flow rates corresponding to the Floresta station, the parameters used came from the calibration of Poço de Pedras, Ilha Grande and Piancó stations. The permanence curves in both cases were separated into two parts to highlight very low flow rates. For these it is verified that the models did not properly reproduce the recorded flows. This behavior was identified by Costa et al. (2014) in the application of the regionalization method by synthetic permanence curve for Ceará basins with daily data. Probably this is due to the inadequacy of general models used in both cases (MODHAC and Rio Grande). Two examples of parameter transfer are presented as permanence curves for the stations of Iguatu, in Ceará, and Floresta, in Pernambuco.

In the general context, 113 analyzes were performed from all 45 stations using the parameters of the “donor” stations

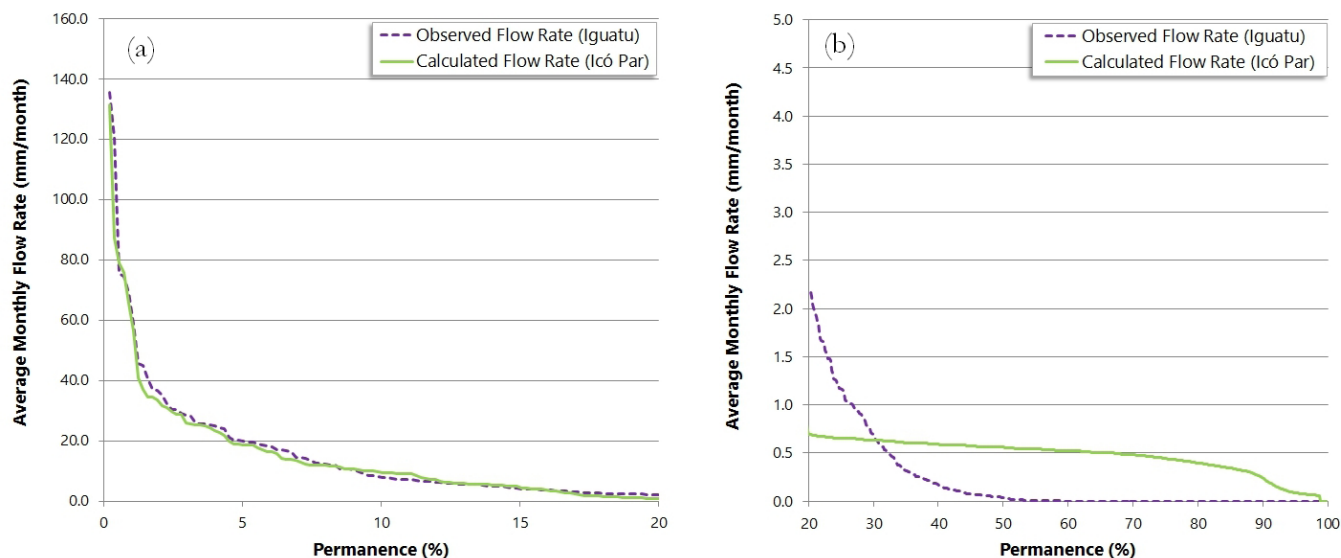


**Figure 5.** Permanence curves of the observed and simulated flow for the Floresta station from the parameters calibrated at the Poço de Pedras, Ilha Grande, and Piancó stations: a) larger flows; b) smaller flows.

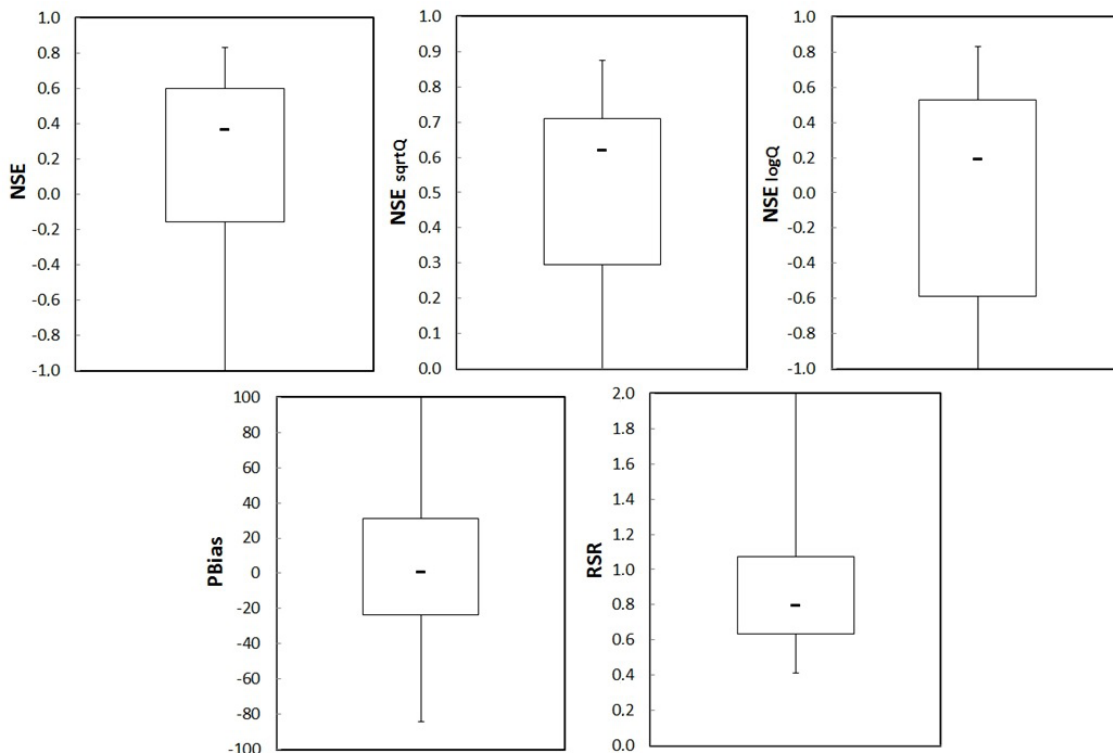
applied to the flow simulation for the periods in which they presented data. Figure 7 shows box plot graphs of all analyzes performed considering the NSE,  $NSE_{sqrtQ}$ ,  $NSE_{logQ}$ , PBias and RSR indicators. NSE values greater than 0.50 were obtained in 40% of the analyzes of all performed simulations and 51% presented NSE above 0.36.  $NSE_{sqrtQ}$  and  $NSE_{logQ}$  above 0.5 were obtained in 63% and 26% of the simulations, respectively. RSR values less than 0.7 and PBias less than 25 were obtained in 40%

and 72% of the results, respectively. However, in only 39% of the analyzed results, the set of NSE, RSR, and PBias parameters were classified as satisfactory, according to Moriasi et al. (2007) proposal.

Table 7 shows the regionalization performance indicators from the permanence curves. As can be seen, the best results were obtained from data of the “donor stations” Icó, for Iguatu, and Piancó, for Floresta.



**Figure 6.** Permanence curves of the observed and simulated flow for the Iguatu station from the parameters calibrated at the Icó station: a) higher flows; b) smaller flows.



**Figure 7.** Box plot graph of all results obtained between observed flow and simulated flow considering the performance indicators NSE,  $NSE_{sqrtQ}$ ,  $NSE_{logQ}$ , PBias, and RSR (first and third quartiles, median and extreme values).

**Table 7.** Performance indicators for regionalization of parameters concerning the permanence curves: Iguatu and Floresta stations.

Code	Name	Parameter	NSE	R <sup>2</sup>	RMSE	AAPE	RSR	PBias	NSE <sub>sqrtQ</sub>	NSE <sub>logQ</sub>
36160000	Iguatu	Icó	0.98	0.98	1.63	20.53	0.14	0.69	0.93	0.62
48860000	Floresta	Poço de Pedras	-0.13	0.98	17.82	0.47	1.06	88.41	0.78	0.89
48860000	Floresta	Ilha Grande	0.70	0.96	9.19	0.88	0.55	52.41	0.84	0.82
48860000	Floresta	Piancó	0.93	0.94	4.52	0.82	0.27	-10.11	0.90	-1.16

### Evaluation of selected “donor” fluviometric stations

The reference stations indicated in Table 4 were chosen based on their spatial distribution so that they could represent a closer set of stations, and on the best quality of their historical series when compared to the others. To evaluate the representativeness of this selection, an analysis was performed with all 45 stations indicated in Table 2, presenting then the physio-climatic properties of the basins under study (area, flow, and annual average precipitation, runoff coefficient, percentage of fractured soil). Afterward, the aim is to evaluate the regionalization criteria from these variables. For this, it was adopted the methodology that was proposed to measure the physical similarity between watersheds described by Kay et al. (2007), and adapted by Oudin et al. (2010). The similarity is defined by the Euclidean distance in the space of each physical or climatic property, with values normalized by their standard deviation throughout the complete set. This is done using Equation (8).

$$dist_{a,b} = \sqrt{\sum_{j=1}^J w_j \left( \frac{x_{a,j} - x_{b,j}}{\sigma_{x,j}} \right)^2} \quad (8)$$

Where  $j = 1, J$  are the basin descriptors,  $x_{a,j}$  is the value of each descriptor,  $\sigma_{x,j}$  is the standard deviation of each property and  $w_j$  is the weight assigned to each descriptor. In this case, the following values were estimated for the weights: 0.15 for the area; 0.225 for mean flow and precipitation; 0.25 for the flow coefficient and 0.15 for the crystalline basement fraction. The estimation of these values was performed to give prime importance to the variables that depend on hydrological data. The Euclidean distance described in Equation (8) was calculated for all 45 stations indicated in Table 2.

To compare the results of the selected donor stations, the results obtained for the data of stations Iguatu, in Ceará, and Floresta, in Pernambuco were chosen as examples, which are presented in Table 7.

In the case of regionalization for Floresta station, the shortest Euclidean distances were 0.435 for Piancó station; 0.598 for Capivara station; 0.701 for Poço das Pedras station; 0,856 for Ilha Grande station. Except for Capivara station, the selection of the other stations coincided with the choice made for Piancó station, with inversion in the ranking of Ilha Grande and Poço das Pedras stations.

Regarding the analysis of data from the Iguatu station, the choice as “donor basin” based on the defined metric fell on the Oiticica station. The criterion of shortest Euclidean distance pointed the Icó station as the 6th priority among the 9 stations chosen as “donors”, and this was the station that at the end presented the best fit. For the stations which regionalization could not be established with any of the others that were used as reference (Croatá, Saudoso, Fazenda Paraná, Caicó, Bodocongó,

Caraúbas and Inajá), the Euclidean distances criterion indicated as stations able to act as “donors” some of them that did not meet the minimum available flow data requirements. Therefore, these stations were not selected.

## CONCLUSIONS

This study involved the survey of an extensive hydrological and spatial database on the Brazilian Northeastern semi-arid, analyzing data from precipitation, flow and evaporation series in four states.

The fragility of the observed data was evident, being necessary to develop a detailed analysis of the flows and precipitations series, as well as external effects that affect the quality of the flow data. It is also required to verify the presence of dams and the reduction of the amount of pluviometric stations that remained in operation. These changes are reflected in climate indicators and thereby, being more challenging to establish similarity relationships. This gathering of data and information sought to explain the reduction in the simulation’s quality for the most recent periods. The regionalization of the model parameters presented better adjustments for watersheds of states that have older and more consistent hydrological series, such as Ceará. In the case of Pernambuco, the results were impaired due to the absence of fluviometric stations with long, consistent and available series.

A portion of non-regionalized areas is due to reservoir interference. These situation does not mean that the parameters used cannot lead to sound results, but it would be necessary to expand the study to consider only the incremental areas upon the reservoirs and to complement their inflows.

In regionalization analyzes the scale effect was not noticed. Where happened a rejection of the parameter values of the “donor basins” it is not noticed influence of the basin area. The rainfall data deficiencies and the influence of reservoirs were evaluated as responsible for reducing the quality of the simulation of the most recent periods. Therefore, it cannot be safely deduced that the calibrated parameter values are not appropriate for the data series from 2001 onwards. The short duration or presence of long gap periods compromised the use of other stations that could be “donors” and thus broaden the regionalization universe.

Another essential issue that should be analyzed concerns the hydrological model used, although Petheram et al. (2012), for example, consider that modeling results are more dependent on the quality of input data than on the model. However, certain behaviors of intermittent river basins are often not reproduced by the equations that represent the simulated processes.

The results obtained in the regionalization are average, as can be observed in the performance indicators. Although in 63% of the results the values obtained from NSE<sub>sqrtQ</sub> were at least

satisfactory, representing an adequate adjustment for the average flows. On the other hand, only 26% of the values obtained for  $NSE_{logQ}$ , the minimum flow adjustment indicator, and 39% of the set of three indicators (NSE, RSR, and Pbias) were simultaneously met, according to the Moriasi et al. (2007) criteria.

The challenge of improving regionalization for semi-arid regions with scarce data is related to the parameterization with a higher physical basis. In this way, a significant portion of the parameters can be calculated directly from the basin characteristics, which will facilitate the transfer of parameter values.

## ACKNOWLEDGEMENTS

The authors acknowledge the CPRM for encouraging the development of the project that led to the presented work, and CNPq, Rede Clima and INCT Clima for their support to the UFPE research group. We also acknowledge the editors and reviewers of RBRH for the many contributions that have enhanced this article.

## REFERENCES

- ADAM, E. O.; ELBASIT, M. A. M. A.; TESHAMICHAEL, S.; AHMED, F. Integration of satellite rainfall data and curve number method for runoff estimation under semi-arid wadi system. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, v. XLII-3, n. W2, p. 1-7, 2017.
- AL-QURASHI, A.; MCINTYRE, N.; WHEATER, H. S.; UNKRICH, C. Application of the Kineros2 rainfall-runoff model to an arid catchment in Oman. *Journal of Hydrology*, v. 355, n. 1-4, p. 91-105, 2008. <http://dx.doi.org/10.1016/j.jhydrol.2008.03.022>.
- ANA. Atlas nordeste: abastecimento urbano de água: alternativas de oferta de água para as sedes municipais da região Nordeste do Brasil e do Norte de Minas Gerais. Brasília: Agência Nacional de Águas, 2006. 80 p. Available from: <<http://www.integracao.gov.br/documents/10157/3678963/Atlas+Nordeste+-+Abastecimento+Urbano+de+%C3%81gua.pdf/81e3a0c7-2b28-4d3e-9dd7-cd9a9ba3e50e>>. Access on: 22 mar. 2017.
- BÁRDOSSY, A. Calibration of hydrological model parameters for ungauged catchments. *Hydrology and Earth System Sciences*, v. 11, n. 2, p. 703-710, 2007. <http://dx.doi.org/10.5194/hess-11-703-2007>.
- BECK, H. E.; VAN DIJK, A. I. J. M.; DE ROO, A.; MIRALLES, D. G.; MCVICAR, T. R.; SCHELLEKENS, J.; BRUIJNZEEL, L. A. Global-scale regionalization of hydrologic model parameters. *Water Resources Research*, v. 52, n. 5, p. 3599-3622, 2016. <http://dx.doi.org/10.1002/2015WR018247>.
- BERGSTRÖM, S. *The HBV model - its structure and applications, SMHI Rep. RH 4*. Norrköping: Swedish Meteorological and Hydrological Institute, 1992
- BRASIL. Ministério da Integração. *Projeto de integração do Rio São Francisco com bacias hidrográficas do nordeste setentrional*. São Paulo: FUNCATE – Fundação de Ciência, Aplicações e Tecnologia
- Espaciais, 2001. Available from: <<http://www.setechidrobrasileira.com.br/port/Projetos/recursos-hidricos/transposicao-das-aguas-do-rio-sao-francisco-f>>. Access on: 06 nov. 2018.
- CABRAL, S. L.; SAKURAGI, J.; SILVEIRA, C. S. Incertezas e erros na estimativa de vazões usando modelagem hidrológica e precipitação por RADAR. *Revista Ambiente & Água*, v. 12, n. 1, p. 57, 2017. <http://dx.doi.org/10.4136/ambi-agua.1924>.
- CIRILO, J. A.; MONTENEGRO, S. M. G. L.; CAMPOS, J. N. B. The issue of water in the Brazilian semi-arid region. In: BICUDO, C. M.; TUNDISI, J. G., SCHEUENSTUHL, M. C. B. (Eds.), *Waters of Brazil*. 1 ed. Cham: Springer International Publishing, 2017. p. 59-71. [http://dx.doi.org/10.1007/978-3-319-41372-3\\_5](http://dx.doi.org/10.1007/978-3-319-41372-3_5).
- COSTA, V.; FERNANDES, W.; NAGHETTINI, M. Regional models of flow-duration curves of perennial and intermittent streams and their use for calibrating the parameters of a rainfall-runoff model. *Hydrological Sciences Journal*, v. 59, n. 2, p. 262-277, 2014. <http://dx.doi.org/10.1080/02626667.2013.802093>.
- CPRM – Companhia de Pesquisa de Recursos Minerais. *Atlas pluviométrico do Brasil*. Brasília, 2011. Escala 1:5.000.000. Available from: <<http://www.cprm.gov.br/publique/Hidrologia/Mapas-e-Publicacoes/Atlas-Pluviometrico-do-Brasil-1351.html>>. Access on: 01 dec. 2017.
- ENGEORPS; HARZA. Projeto: transposição de águas do Rio São Francisco para o Nordeste Setentrional. São Paulo: ENGEORPS/HARZA, 2000. Relatório R32 - Relatório síntese de viabilidade técnico-econômica e ambiental. Available from: <https://www.bndes.gov.br/arquivos/pisf/r32-sintese.pdf>. Access on: 04 jul. 2019.
- FELIX, V. S.; PAZ, A. R. Representação dos processos hidrológicos em bacia hidrográfica do semiárido paraibano com modelagem hidrológica distribuída. *Revista Brasileira de Recursos Hídricos*, v. 21, n. 3, p. 556-569, 2016. <http://dx.doi.org/10.1590/2318-0331.011616009>.
- GARAMBOIS, P. A.; ROUX, H.; LARNIER, K.; LABAT, D.; DARTUS, D. Parameter regionalization for a process-oriented distributed model dedicated to flash floods. *Journal of Hydrology*, v. 525, p. 383-399, 2015. <http://dx.doi.org/10.1016/j.jhydrol.2015.03.052>.
- GIRARDI, R. V.; CASTRO, N.; GOLDENFUND, J. A.; SILVEIRA, A. L. L. Avaliação do efeito de escala em características de chuva e vazão em sub-bacias embutidas da Bacia do Potiribu-RS. *Revista Brasileira de Recursos Hídricos*, v. 16, n. 2, p. 49-64, 2011. <http://dx.doi.org/10.21168/rbrh.v16n2.p49-64>.
- HUANG, P.; LI, Z.; CHEN, J.; LI, Q.; YAO, C.; LI, Q.; YAO, C. Event-based hydrological modeling for detecting dominant hydrological process and suitable model strategy for semi-arid catchments. *Journal of Hydrology*, v. 542, p. 292-303, 2016. <http://dx.doi.org/10.1016/j.jhydrol.2016.09.001>.
- IBGE – Instituto Brasileiro de Geografia e Estatística. *Base cartográfica*. Rio de Janeiro, 2013. Escala 1:250.000. Available from: <<http://>>

- mapas.ibge.gov.br/bases-e-referenciais/bases-cartograficas/cartas>. Access on: 02 apr. 2013.
- INSA – Instituto Nacional do Semiárido. *Acervo digital*. Campina Grande, 2014. Available from: <www.insa.gov.br/sigsab/acervoDigital>. Access on: 26 sept. 2017.
- KAN, G.; HE, X.; DING, L.; LI, J.; LIANG, K.; HONG, Y. Study on Applicability of Conceptual Hydrological Models for Flood Forecasting in Humid, Semi-Humid Semi-Arid and Arid Basins in China. *Water (Basel)*, v. 9, n. 10, p. 719, 2017. <http://dx.doi.org/10.3390/w9100719>.
- KAY, A. L.; JONES, D. A.; CROOKS, S. M.; KJELDSEN, T. R.; FUNG, C. F. An investigation of site-similarity approaches to generalisation of a rainfall–runoff model. *Hydrology and Earth System Sciences*, v. 11, n. 1, p. 500-515, 2007. <http://dx.doi.org/10.5194/hess-11-500-2007>.
- KOKKONEN, T. S.; JAKEMAN, A. J.; YOUNG, P. C.; KOIVUSALO, H. J. Predicting daily flows in ungauged catchments: model regionalization from catchment descriptors at the Coweeta Hydrologic Laboratory, North Carolina. *Hydrological Processes*, v. 17, n. 11, p. 2219-2238, 2003. <http://dx.doi.org/10.1002/hyp.1329>.
- LANNA, A. E. *MAG1 – MODHAC97: modelo hidrológico auto-calibrável*. Porto Alegre: Instituto de Pesquisas Hidráulicas da UFRGS, 1997.
- LI, H.; ZHANG, Y.; CHIEW, F. H. S.; XU, S. Predicting runoff in ungauged catchments by using Xinanjiang model with MODIS leaf area index. *Journal of Hydrology*, v. 370, n. 1-4, p. 155-162, 2009. <http://dx.doi.org/10.1016/j.jhydrol.2009.03.003>.
- LIU, J.; HAN, D. Indices for calibration data selection of the rainfall-runoff model. *Water Resources Research*, v. 46, n. 4, p. W04512, 2010. <http://dx.doi.org/10.1029/2009WR008668>.
- MARTINS, E. S. P. R.; SOUZA FILHO, F. A.; CARDOSO, G. B. B. Uso de otimização/simulação e previsão de afluências na operação tática dos reservatórios dos sistema Jaguaribe-Metropolitano-CE. *Revista Brasileira de Recursos Hídricos*, v. 11, n. 4, p. 175-186, 2006. <http://dx.doi.org/10.21168/rbrh.v11n4.p175-186>.
- MCINTYRE, N. R.; LEE, H.; WHEATER, H. S.; YOUNG, A.; WAGENER, T. Ensemble predictions of runoff in ungauged catchments. *Water Resources Research*, v. 41, n. 12, p. W12434, 2005. <http://dx.doi.org/10.1029/2005WR004289>.
- MCINTYRE, N.; AL-QURASHI, A. Performance of ten rainfall-runoff models applied to an arid catchment in Oman. *Environmental Modelling & Software*, v. 24, n. 6, p. 726-738, 2009. <http://dx.doi.org/10.1016/j.envsoft.2008.11.001>.
- MORIASI, D. N.; ARNOLD, J. G.; VAN LIEW, M. W.; BINGNER, R. L.; HARMEL, R. D.; VEITH, T. L. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *American Society of Agricultural and Biological Engineers*, v. 50, n. 3, p. 885-900, 2007.
- LOUDIN, L.; ANDRÉASSIAN, V.; MATHEVET, T.; PERRIN, C.; MICHEL, C. Dynamic averaging of rainfall–runoff model simulations from complementary model parameterizations. *Water Resources Research*, v. 42, n. 7, p. W07410, 2006. <http://dx.doi.org/10.1029/2005WR004636>.
- LOUDIN, L.; ANDRÉASSIAN, V.; PERRIN, C.; MICHEL, C.; LE MOINE, N. Spatial proximity, physical similarity, regression and ungauged catchments: A comparison of regionalization approaches based on 913 French catchments. *Water Resources Research*, v. 44, n. 3, p. W03413, 2008. <http://dx.doi.org/10.1029/2007WR006240>.
- LOUDIN, L.; KAY, A.; ANDRÉASSIAN, V.; PERRIN, C. Are seemingly physically similar catchments truly hydrologically similar? *Water Resources Research*, v. 46, n. 11, p. W11558, 2010. <http://dx.doi.org/10.1029/2009WR008887>.
- PARAJKA, J.; MERZ, R.; BLÖSCHL, G. A comparison of regionalisation methods for catchment model parameters. *Hydrology and Earth System Sciences*, v. 9, n. 3, p. 157-171, 2005. <http://dx.doi.org/10.5194/hess-9-157-2005>.
- PATIL, S. D.; STIEGLITZ, M. Modeling daily streamflow at ungauged catchments: What information is necessary? *Hydrological Processes*, v. 28, n. 3, p. 1159-1169, 2014. <http://dx.doi.org/10.1002/hyp.9660>.
- PETHERAM, C.; RUSTOMJI, P.; CHIEW, F. H. S.; VLEESHOUWER, J. Rainfall-runoff modelling in northern Australia: A guide to modelling strategies in the tropics. *Journal of Hydrology*, v. 462-463, p. 28-41, 2012. <http://dx.doi.org/10.1016/j.jhydrol.2011.12.046>.
- PILGRIM, D. H.; CHAPMAN, T. G.; DORAN, D. G. Problems of rainfall-runoff modelling in arid and semiarid regions. *Hydrological Sciences Journal*, v. 33, n. 4, p. 379-400, 1988. <http://dx.doi.org/10.1080/02626668809491261>.
- PINHEIRO, V. B.; NAGHETTINI, M. Calibração de um modelo chuva-vazão em bacias sem monitoramento fluviométrico a partir de curvas de permanência sintéticas. *Revista Brasileira de Recursos Hídricos*, v. 15, n. 2, p. 143-156, 2010. <http://dx.doi.org/10.21168/rbrh.v15n2.p143-156>.
- PRAFULLA, P.; YILMAZ, K. K.; GUPTA, H. V. Multiple-criteria calibration of a distributed watershed model using spatial regularization and response signatures. *Journal of Hydrology*, v. 418, p. 49-60, 2012.
- PUSHPALATHA, R.; PERRIN, C.; LE MOINE, N.; ANDRÉASSIAN, V. A review of efficiency criteria suitable for evaluating low-flow simulations. *Journal of Hydrology*, v. 420-421, p. 171-182, 2012. <http://dx.doi.org/10.1016/j.jhydrol.2011.11.055>.
- REICHL, J. P. C.; WESTERN, A. W.; MCINTYRE, N. R.; CHIEW, F. H. S. Optimization of a similarity measure for estimating ungauged



- streamflow. *Water Resources Research*, v. 45, n. 10, p. W10423, 2009. <http://dx.doi.org/10.1029/2008WR007248>.
- SAMUEL, J.; COULIBALY, P.; METCALFE, R. A. Estimation of Continuous Streamflow in Ontario Ungauged Basins: Comparison of Regionalization Methods. *Journal of Hydrologic Engineering*, v. 16, n. 5, p. 447-459, 2011. [http://dx.doi.org/10.1061/\(ASCE\)HE.1943-5584.0000338](http://dx.doi.org/10.1061/(ASCE)HE.1943-5584.0000338).
- SEIBERT, J.; VIS, M. J. P. Teaching hydrological modeling with a user-friendly catchment-runoff-model software package. *Hydrology and Earth System Sciences*, v. 16, n. 9, p. 3315-3325, 2012. <http://dx.doi.org/10.5194/hess-16-3315-2012>.
- SELLAMI, H.; LA JEUNESSE, I.; BENABDALLAH, S.; BAGHDADI, N.; VANCLOOSTER, M. Uncertainty analysis in model parameters regionalization: A case study involving the SWAT model in Mediterranean catchments (Southern France). *Hydrology and Earth System Sciences*, v. 18, n. 6, p. 2393-2413, 2014. <http://dx.doi.org/10.5194/hess-18-2393-2014>.
- SINGH, R.; ARCHFIELD, S. A.; WAGENER, T. Identifying dominant controls on hydrologic parameter transfer from gauged to ungauged catchments: A comparative hydrology approach. *Journal of Hydrology (Amsterdam)*, v. 517, p. 985-996, 2014. <http://dx.doi.org/10.1016/j.jhydrol.2014.06.030>.
- TRAORE, V. B.; SAMBOU, S.; TAMBA, S.; FALL, S.; DIAW, A. T.; CISSE, M. T. Calibrating the Rainfall-Runoff Model GR4J and GR2M on the Koulountou River Basin, a tributary of the Gambia River. *American Journal of Environmental Protection*, v. 3, n. 20, p. 36-44, 2014. <http://dx.doi.org/10.11648/j.ajep.20140301.15>.
- VAN LIEW, M. W.; VEITH, T. L.; BOSCH, D. D.; ARNOLD, J. G. Suitability of SWAT for the conservation effects assessment project: comparison on USDA agricultural research service watersheds. *Journal of Hydrologic Engineering*, v. 12, n. 2, p. 173-189, 2007. [http://dx.doi.org/10.1061/\(ASCE\)1084-0699\(2007\)12:2\(173\)](http://dx.doi.org/10.1061/(ASCE)1084-0699(2007)12:2(173)).
- WALLNER, M.; HABERLANDT, U.; DIETRICH, J. A one-step similarity approach for the regionalization of hydrological model parameters based on self-organizing maps. *Journal of Hydrology*, v. 494, p. 59-71, 2013. <http://dx.doi.org/10.1016/j.jhydrol.2013.04.022>.
- WANG, M.; ZHANG, L.; BADDOO, T. D. Hydrological modeling in a semi-arid region using HEC-HMS. *Journal of Water Resource and Hydraulic Engineering*, v. 5, n. 3, p. 105-115, 2016. <http://dx.doi.org/10.5963/JWRHE0503004>.
- WESTERBERG, I. K.; GUERRERO, J.-L.; YOUNGER, P. M.; BEVEN, K. J.; SEIBERT, J.; HALLDIN, S.; FREER, J. E.; XU, C.-Y. Calibration of hydrological models using flow-duration curves. *Hydrology and Earth System Sciences*, v. 15, n. 7, p. 2205-2227, 2011. <http://dx.doi.org/10.5194/hess-15-2205-2011>.
- ZAPPA, M. *Multiple-response verification of a distributed hydrological model at different spatial scales*. 2002. 167 f. Thesis (PhD.) - Swiss Federal Institute of Technology, Zurich, 2002.
- ZHANG, Y.; VAZE, J.; CHIEW, F. H. S.; LI, M. Comparing flow duration curve and rainfall-runoff modelling for predicting daily runoff in ungauged catchments. *Journal of Hydrology*, v. 525, p. 72-86, 2015. <http://dx.doi.org/10.1016/j.jhydrol.2015.03.043>.

### Authors contributions

Múcio Valença Virões: Paper conception, research, literature review, data analysis, results, discussion and paper writing.

José Almir Cirilo: Paper conception, research, literature review, discussion, paper revision and writing.