This is the author's accepted version prior to journal typesetting. Please cite as :

Arsenault, R., Breton-Dufour, M., Poulin, A., Dallaire, G. and Romero-Lopez, R., 2019. Streamflow prediction in ungauged basins: analysis of regionalization methods in a hydrologically heterogeneous region of Mexico. Hydrological Sciences Journal, 64(11), pp.1297-1311. https://doi.org/10.1080/02626667.2019.1639716

Streamflow Prediction in Ungauged Basins: Analysis of regionalization methods in a hydrologically heterogeneous region of Mexico

Richard Arsenault[1]*, Mélissa Breton-Dufour[1], Annie Poulin[1], Gabrielle Dallaire[1], Rabindranarth Romero-Lopez[2]

[1] Department of construction engineering, École de technologie supérieure, 1100 Notre-Dame Ouest, Montréal (Québec), Canada, H3C 1K3.

[2] Universidad Veracruzana, Lomas del Estadio S/N, Zona Universitaria, Xalapa, 91000, Mexico

* Corresponding author: richard.arsenault@etsmtl.ca

Abstract

This paper investigates the ability of three regionalization methods to predict streamflow in ungauged catchments in Mexico, namely the Multiple Linear Regression (MLR), Spatial Proximity (SP) and Physical Similarity (PS) methods. Three hydrological models (GR4J, HMETS and MOHYSE) were calibrated on 30 diverse catchments in Mexico. A leave-one-out cross-validation implementation enabled estimating the regionalization skill at each of the 30 sites, in turn considered as being ungauged. This study allowed showing that regionalization in a hydrologically heterogeneous area such as the Mexican area under study poses problems to regionalization approaches that depend on physical catchment descriptors such as MLR and PS. The transfer of complete parameter sets from a neighboring catchment provided the most robust method to estimate streamflow in semi-arid and humid ungauged basins. The arid catchments performed worse in the context of regionalization, with GR4J being more robust than the other models due to its simpler structure.

Keywords: hydrological modeling, ungauged basin, regionalization, spatial proximity, physical similarity.

1. Introduction

Hydrological modeling often requires access to hydrometric data in order to calibrate the hydrological model. The calibrated model can then be used to simulate historic flows on the catchment, perform climate change impact studies or generate forecasts for water resources management applications. Unfortunately, in many cases, studies must be performed at sites where no gauging station exists. In these cases, it is possible to estimate the historic streamflow at the ungauged sites using so-called “regionalization” methods. These methods attempt to transfer parameter sets from hydrological models successfully calibrated at other sites. The parameter transfer function can rely on either catchment similarity, proximity or other descriptive factors.

In 2003, the International Association of the Hydrological Sciences (IAHS) launched the “Decade on prediction in ungauged basins” (Sivapalan et al. 2003), which led to the development and analysis of many regionalization approaches in a variety of climates and contexts. Multiple authors provided reviews of the developments during that period, notably Bloschl et al. (2013), Hrachowtiz et al. (2013), He et al. (2011), Parajka et al. (2013) and Razavi and Coulibaly 2013. Of the available methods, three have stood out as the most versatile and robust depending on the climate and hydrological regime of the study region. The methods are known as Physical Similarity (PS), Spatial Proximity (SP) and Multiple Linear Regression (MLR). The three regionalization methods are present in most comparative studies and are generally recognized as performing strongly under certain conditions. For example, Oudin et al. (2008) showed that a high density of hydrometric gauges favored the SP method followed by the PS method, whereas both methods become similar under a certain threshold. Parajka et al. (2013) performed a meta-analysis of regionalization studies and found that in general, MLR performed worse than the other two methods except in arid catchments (aridity index > 1.0), where MLR and PS return similar results. Another method of regionalization is to determine catchment hydrologic signatures at ungauged sites and then calibrate a hydrological model to these signatures (Bárdossy 2007, Yadav et al. 2007), however in this paper we focus on the direct determination of parameter sets at ungauged locations. The next section details the study objectives and historical streamflow prediction attempts in the region of study. . A brief overview of the regionalization methods is given in sections 1.2 to 1.5.

1.1 Regionalization in Mexico and study objectives

This study aims to determine if any of the classical regionalization methods can be used effectively to estimate streamflow in ungauged basins in Mexico. An extensive literature review has shown that only a handful of studies tackling regionalization in Mexico have been performed even though there is a need to predict streamflow in ungauged basins (Vicente-Serrano, 2006). Ouarda et al. (2008) used statistical analysis to estimate homogeneous hydrological regions for regionalization and found that Canonical Correlation Analysis (CCA) outperformed Canonical Kriging and hierarchical clustering for three basins in Mexico. Álvarez-Olguín et al. (2011) also applied homogeneous hydrological region analysis tools on 17 catchments in the Iberian Peninsula in Mexico and found that they could be grouped into three classes. Based on this work, Domínguez-Mora et al. (2016) developed a procedure to regionalize maximum flows on 37 catchments in Mexico, and Allende et al. (2009) performed regionalization on a catchment undergoing land-use change but limited the analysis to two years. Rojas-Serna et al. (2006) worked on calibrating the parameters of the daily lumped GR4J rainfall–runoff model, using a few streamflow measurements, in combination with a priori knowledge of the parameters. The approach was applied to 1111 catchments in five continents, including 260 catchments in Mexico. They found that this approach could be more efficient than classical regionalization studies, as soon as about thirty measurements could be made, at random, during a period of three to five years. Luis-Pérez et al. (2011) worked on regional flood-frequency estimation at ungauged sites in nine catchments from the Mexican Mixteca region using two different clustering approaches for the delineation of homogeneous zones. MLR was applied for the flood frequency estimation in each homogeneous zone. Of the studies found in the literature, none analyzed long-term, daily and continuous streamflow simulation in Mexico.

This study will fill a gap by performing a large-scale analysis of regionalization methods in central Mexico using three regionalization methods (MLR, SP and PS) and using three hydrological models. The breadth of climate types and variety of physical characteristics of the catchments will allow studying the regionalization methods and help answer the three following questions:

1- Which regionalization method, if any, should be preferred in the study region?
2- Do methods that perform well elsewhere in similar conditions still perform well in Mexico (to validate previous findings)?
3- Does the complexity of hydrological models influence the regionalization methods' performance?

The next sections present the regionalization methods used in this study as well as variations applied to help improve the regionalization skill in the study region.

1.2 Multiple Linear Regression (MLR)

MLR was amongst the first regionalization methods to be developed and studied. The idea behind MLR is that hydrological model parameters are, in theory, adjusted to represent certain physical processes. These processes should theoretically be driven by the catchment's physical characteristics such as slope, land-use and soil type. Therefore, MLR aims to define a relationship between catchment descriptors and hydrological model parameters calibrated on a large number of sites. These relationships can then be used to estimate the best parameter values on the ungauged site by using the ungauged catchment's descriptors in the regression equation.

Each parameter is evaluated independently and therefore the parameter set loses its cohesion, which can lead to a loss in performance (Arsenault and Brissette 2014).

Kokkenan et al. (2003) used a variant of MLR in which parameter values were first found for a global region and the regression analysis was performed on smaller regions using their own catchment descriptors in the regression estimation. They established that when similar catchments are available, it is better to use the entire parameter set from a similar catchment than to estimate the parameters one by one. MLR has been used successfully in flat and semi-arid catchments (Bloschl et al. 2013), which is not representative of our study site. MLR also requires large amounts of information to establish significant relationships between parameters and descriptors (He et al, 2011), which are not easily available in Mexico.

1.3 Physical similarity (PS)

PS uses the simple idea of calibrating a hydrological model on a similar catchment to the ungauged one, and then transferring the parameter set to the ungauged site. The idea is that the similar catchment should behave similarly hydrologically, and that the calibrated parameter set should be transferable to the similar site. Similarity is defined as the normalized distance between two points in N-dimensional space, where each dimension refers to a particular catchment descriptor, such as elevation, land-use, soil type and drainage density. The distance, also called the similarity index, can be computed using a simple formula such as that of Burn and Boorman (1993):

$$\theta = \sum_{i=1}^{k} \frac{|CD_i^G - CD_i^U|}{\Delta CD_i} \quad \text{(eq. 1)}$$

Where θ is the similarity index, CD_i represents the descriptor values vector for gauged site (G) and ungauged site (U), k is the number of catchment descriptors and ΔCD_i is the range of values that CD can take in the database. The distance is computed between the ungauged catchment and a set of calibrated catchments. The most similar catchment, i.e. the one with the shortest distance, is considered the "donor" catchment as it will donate its parameters to the ungauged basin. However, Oudin (2010) cast doubt that similar catchments are inherently hydrologically similar. Nonetheless, it is widely used in regionalization studies across the world. Reichl et al. (2009) found that having a good prior knowledge of the catchment descriptors could lead to better performance. When seven descriptors or more were used, the PS method could outperform the SP approach.

Wagener et al. (2007) found that an adequate catchment classification scheme was an important element for PS to perform satisfactorily. The advantage of PS over MLR is that the parameter set is transferred completely and remains intact, whereas in MLR the parameter set is reconstructed one parameter at a time by regression (McIntyre et al 2005, Parajka et al 2005, Oudin et al 2010). Nonetheless, Arsenault and Brissette (2014) proposed a method by which a parameter set that is transferred from a donor basin can be modified by applying the MLR approach only to the parameters whose linear regression R^2 skill is over a certain threshold. For example, if 2 out of 10 parameters have high R^2 scores (e.g. 0.8 or higher), then 8 parameters are preserved as-is from the donor set and the other two parameters are estimated using the regression function with the ungauged basin's catchment descriptors. This method, referred to as "regression-augmented regionalization", outperformed the classic approaches on 268 catchments in Canada. Catchment elevation has been considered an important descriptor in PS as it is correlated to multiple hydrological and climatic variables (Parajka et al 2013). The elevation range in the study site is quite variable; therefore, this parameter will be included as a catchment descriptor.

1.4 Spatial Proximity (SP)

SP is the simplest method to put into practice. Requiring no information regarding catchment attributes, it is based on the hypothesis that neighboring (or nearby) catchments must share physical attributes such as land cover, soil type, elevation, slope, climate data and so on. Therefore, under this hypothesis, there is no need to look for the most similar catchment as the nearby ones could be “similar enough” simply based on their location. The implementation of SP requires a few nearby catchments that have been successfully modelled and calibrated by the hydrological model. The closest catchment becomes the de facto donor and the hydrological model parameters are transferred to the ungauged site. The distance between the ungauged basin and the donor candidates is computed using the simple Euclidian distance between the catchments centroids as shown in equation 2.

$$d = \sqrt{(X_G - X_U)^2 + (Y_G - Y_U)^2} \qquad \text{(eq. 2)}$$

Where d is the distance between the centroids and (X_G, Y_G) and (X_U, Y_U) are respectively the centroid coordinates for the gauged and ungauged catchments.

Many studies used the SP method due to its simplicity and relatively good performance. Oudin et al. (2008) found that for 913 basins in France, the SP method performed better than the PS and

MLR methods, mainly due to the high density of the hydrological network. Merz and Bloschl (2004) and Paparakja et al. (2005) also came to the same conclusions, whereas Arsenault and Brissette (2014) found that even with a relatively sparse hydrological network the SP method slightly outperformed the PS method and was far more skillful than the MLR method. As is the case with the PS approach, a regression-augmented variant is also available for SP and was shown to be a good choice in some specific circumstances.

1.5 SP-PS integration

Oudin et al. (2008) showed that it was possible to improve upon the PS method by adding the latitude and longitude as catchment descriptors, which can be likened to integrating the SP method into the PS method. This method, sometimes called integrated similarity, combines the known and measurable catchment descriptors (i.e. land use, elevation, etc.) to the unknown or implicit descriptors such as soil type. Therefore, the advantages of having similar catchments for the known variables is paired with the advantage of having a nearby catchment that should be similar for the unknown or unmeasurable attributes. This method has been shown to outperform the PS, SP and MLR in many cases (Parajka et al. 2013, Samuel et al. 2011, Zhang and Chiew 2009). For this reason, PS often implicitly includes latitude and longitude as catchment descriptors.

1.6 Multiple donor averaging

Many authors have shown that for the PS and SP methods, transferring the parameter sets of a few donors and averaging the generated hydrographs could lead to significantly improved results over using a single donor (Oudin et al., 2008, Reichl et al., 2009, Samuel et al. 2011, Zhang and Chiew 2009). There exist numerous implementations of multiple donor averaging, but the two most common are the arithmetic mean and the inverse-distance weighted approaches. In the first case, the hydrographs generated by running the hydrological model on the ungauged site with a few donated parameter sets are averaged using equal weights (Mc Intyre et al. 2005, Oudin et al. 2008). In the second case, the averaging is performed using the inverse-distance weighting scheme, in which the distance is either the Euclidian distance for SP or the similarity index for PS (Samuel et al. 2011, Zhang and Chiew 2009). It has been shown that between 5 and 10 donors usually maximizes the performance, and that inverse-distance weighting slightly outperforms the arithmetic mean, although the differences might not be statistically significant (Arsenault and Brissette 2014).

The next section will detail the study area and section 3 will explain the methodology. Section 4 presents the results and their analysis, followed by concluding remarks.

2. Study area and data

2.1 Study area

The study area is composed of 30 small to medium sized catchments in central Mexico. The catchment locations are presented in Figure 1. The catchments are heterogeneous in multiple ways, as described in section 2.3. In Figure 1, the 30 catchments are overlaid on an average annual runoff map provided by the National Institute of Statistics and Geography (INEGI; INEGI 2016). In Figure 2, it can be seen that the catchments are located in arid, semi-arid and humid regions of Mexico.

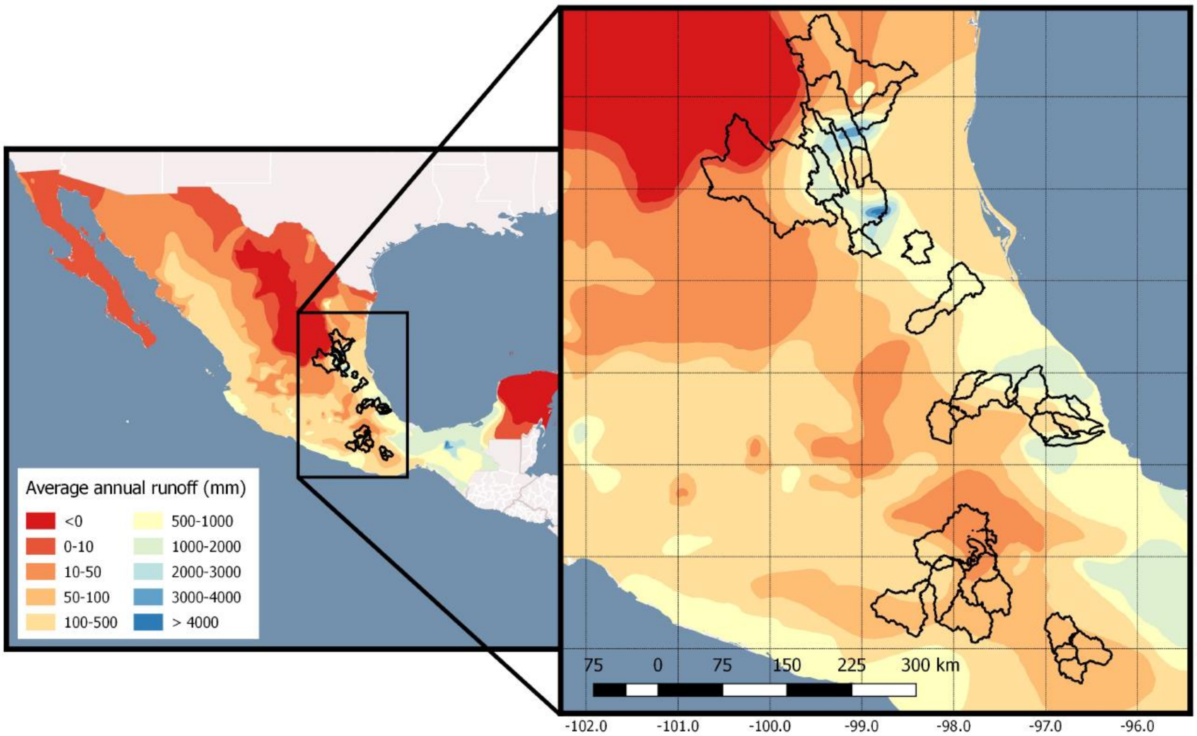

Figure 1: Map of average annual runoff throughout Mexico (left panel), with a zoom on the region and watersheds (black contours) under study (right panel).

The region's topology ranges from coastal plains to continental mountains and vegetation depends on the altitude and precipitation regime, which also vary considerably in space. Furthermore, the region is influenced by wet and dry seasons, with the wet season generally lasting from May/June to October, and the dry season lasting from November to April. The precipitation and runoff are thus strongly correlated to this pattern.

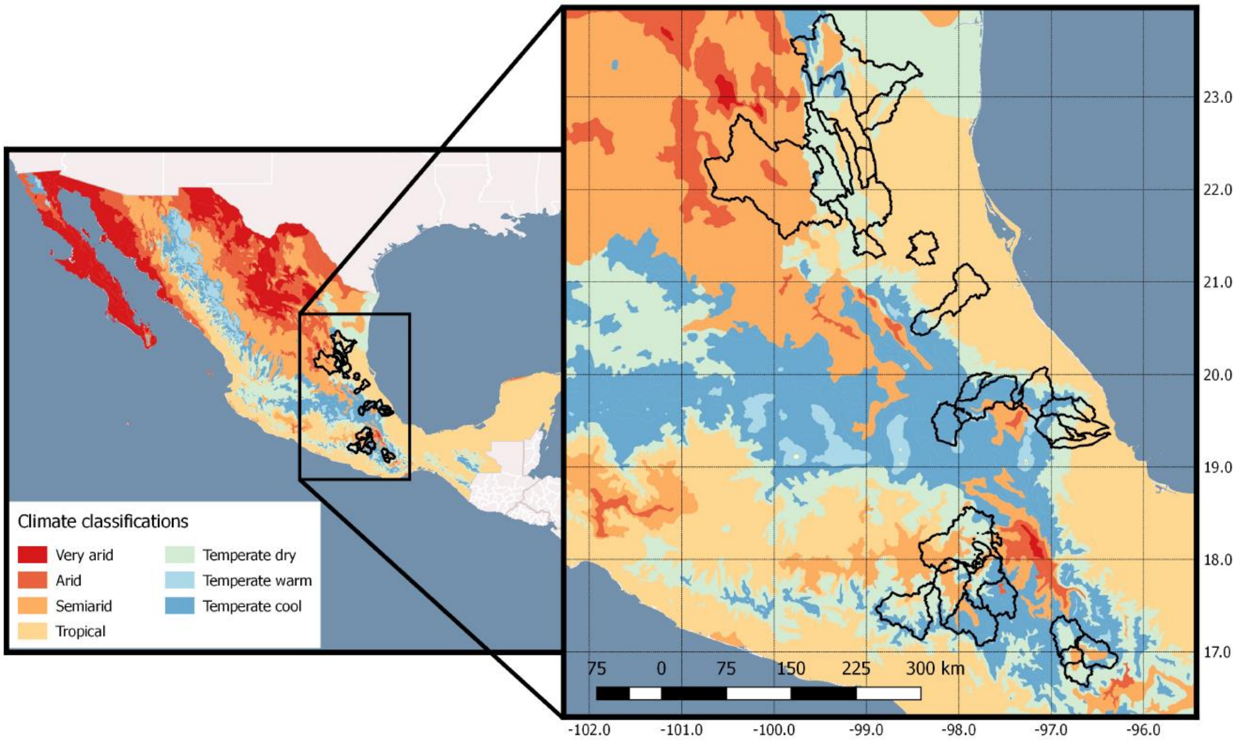

Figure 2: Climate classifications in Mexico overlaid with the boundaries of the 30 catchments in this study.

2.2 Hydrometeorological data

Hydrometric and meteorological data over a large number of Mexican catchments were required to perform the regionalization analyses. The hydrometric data were provided by the National Water Commission of Mexico (CONAGUA) and the meteorological data was sourced from the Livneh et al. (2015)'s comprehensive hydrometeorological database. The database is a gridded climate data product and includes daily minimum and maximum temperature along with daily precipitation from 1950-2013 at a 1/16° resolution over Mexico, the United-States and Southern Canada. Livneh et al. (2015) preformed a tedious data validation and correction work over the entire domain and it is expected that this dataset is of better quality than if simple observation stations were used to drive the hydrological models. The precipitation and temperatures were spatially averaged over the catchment domains to produce a single vector of inputs per variable for each catchment.

2.3 Geophysical data

The geophysical data were provided by the INEGI. The catchment descriptors required to implement the MLR and PS regionalization methods were selected from this database, and include the attributes presented in table 1.

Table 1. Statistics of the catchment descriptors from the 30 basins in this study.

Catchment Descriptor	Min	25th	50th (median)	75th	Max
Area (km^2)	397	1106	1943	3264	10968
Mean Elevation (m)	88	845	1484	1971	2579
Mean Annual Precipitation (P) (mm)	590	780	848	1596	3198
Mean Annual PET (mm)	959	1154	1250	1343	1512
Average discharge (Q) (m3/s) *	2.1	7.2	25.7	46.2	185.2
Runoff Ratio – Q/P (mm/mm)	0.07	0.11	0.30	0.52	0.92
Aridity index – P/PET (mm/mm)	0.44	0.61	0.64	1.17	2.77
Land Use: Agriculture (%)	35	71	87	97	100
Land Use: Forest (%)	0	2	12	28	64
Soil Type: Cambisol (%)	0	0	0.5	1	37.7
Soil Type: Chernozem (%)	0	3.5	5.9	20.5	45.6
Soil Type: Leptosol (%)	0.1	8	30.8	61.8	93.7
Soil Type: Luvisol (%)	0	0.3	2	7	86
Soil Type: Regosol (%)	0	0.6	10.5	15.3	56
Soil Type: Vertisol (%)	0	2.5	7.6	16.6	39.1

* The average discharge is presented here for information purposes but was not used in regionalization because the streamflow is by definition not available in ungauged catchments.

Soil types are edaphological classifications based on maps produced by the Food and Agriculture Organization of the United Nations (FAO, 2016). Land use data was provided by CONAGUA for the most part, however for some catchments satellite imagery was used to estimate land use percentages. It can be noted that Table 1 omits some information regarding certain descriptors. For example, there is no category for urban land use, and some soil types that can be found in Mexico are not present in the table. This is because categories that were deemed too insignificant to be used as predictors of hydrological regime were not included in order to maximize the regression method's ability to discriminate important from unimportant descriptors. Urban areas never exceeded 2% of the total catchment area, and the remaining soil types were only found in two or three catchments. Furthermore, indices based on streamflow such as the runoff ratio (Q/P) are not available as descriptors because the streamflow is simply not available on the ungauged basins.

3. Methodology

This section describes the hydrological models used in this study along with the calibration objectives and the implementation of the regionalization assessment framework.

3.1 Hydrological models

Three lumped hydrological models were used in this study. They are briefly described here along with any modifications to their structure to prepare them for the particularities of the hydrology in Mexico. The hydrological models were selected based on their relatively simple but highly flexible structure, allowing them to be used on a variety of hydrological conditions such as the ones in this study. Lumped models were used instead of distributed models due to the difficulty in obtaining enough high quality data to drive more data-intensive models.

3.1.1 Hydrological model GR4J

The GR4J hydrological model is a simple yet efficient hydrological model based on empirical equations and is conceptually a reservoir-based model (Perrin et al. 2003). Unit hydrographs rout the flow volumes to the catchment outlet. Having only 4 calibration parameters, it is more robust to equifinality and is easier than the other models to calibrate. It requires daily precipitation and daily potential evapotranspiration (PET) as inputs. For this study, the PET formulation of Oudin et al. (2005) was implemented as described in equation 3:

$$PET = \frac{R_e T_a + 5}{\lambda_p * 100} \ if\ T_a + 5 > 0;\ else\ PET = 0 \qquad \text{(eq. 3)}$$

Where PET is the daily potential evaporation, R_e is the extra-terrestrial solar radiation, T_a is the daily average air temperature and λ_p is the latent heat flux. The computation of R_e requires the latitude and day of year as inputs.

GR4J was considered in this paper because it was previously shown to simulate flows adequately on Mexican catchments (Velázquez et al., 2015). It has also been used in regionalization studies due to its small parameter space which makes it easier to find relationships between model parameters and catchment descriptors (Poissant et al. 2017, Arsenault et al. 2015, Oudin et al., 2008, Rojas-Serna et al., 2006). Conceptually, GR4J is the simplest model of the ones tested, with water infiltration percolating to a production store and then routed by two parameterized unit hydrographs, one leading 90% of the water to a routing store and the other routing the remaining 10% directly to the outlet. The model parameters define the production store depth (X1), a groundwater exchange coefficient defining water transfers between the aquifer and the routing store (X2), the routing store depth (X3) and the duration of the unit hydrographs (X4).

One of the key limitations of GR4J relates to its treatment of PET. Indeed, the model does not scale the PET by a mass-balance parameter. Therefore, the choice of a proper PET formulation is critical. This model is more rigid due to the limited parameter space, and this could be seen as both an advantage (fewer parameters to link to catchment descriptors) and a drawback (less flexibility to adapt to a variety of basins). Thus, in this study, the model's capacity to adapt to the local hydrology and its use in regionalization will be evaluated.

3.1.2 Hydrological model MOHYSE

The MOHYSE hydrological model is a simple, 10-parameter lumped hydrological model which was initially developed for teaching hydrological modelling in an academic setting (Fortin and Turcotte, 2007). Its relatively good performance and unparalleled versatility and ease of use quickly made MOHYSE a model of choice in model intercomparison studies and multi-model approaches (Troin et al. 2017, 2015, Velazquez et al. 2011). The MOHYSE model has the ability to simulate snowpack accumulation and melt; however, these processes are not required in the study area. The related module and associated parameters (2 parameters) were therefore removed from the model to reduce equifinality and increase the odds of finding good relationships between model parameters and catchment descriptors for the MLR and PS approaches. The version used in this study therefore has 8 calibration parameters. Castaneda-Gonzales (2014) applied the MOHYSE model in a tropical setting for the first time and found that it performed satisfactorily on a catchment in the central part of the Veracruz state between 2000 and 2010. MOHYSE only requires mean daily temperatures and daily precipitation as inputs. It computes PET internally using a temperature and extra-terrestrial radiation-based formulation. However, as opposed to GR4J, MOHYSE has a calibration parameter that can scale the PET values to improve mass-balance fitting. Furthermore, the routing process is similar to GR4J (i.e. reservoir based and routed with a unit hydrograph), however MOHYSE also has two more reservoirs: One vadose-zone routing reservoir and a more reactive surface runoff module to treat large storm events. Water depths from all three reservoirs are combined and routed to the outlet through a unit hydrograph. Due to the study site properties, it is expected that this component will bring more flexibility to the hydrological modelling process and perform better on the peak flows caused by rainfall events.

3.1.3 Hydrological model HMETS

The HMETS model is a lumped 21-parameter model that was designed to be robust on various conditions, to which flexible snow accumulation and melt processes for Nordic catchments were added (Martel et al. 2017). It is freely available on the Mathworks File Exchange. The complex snow module requires 10 of the 21 parameters. As is the case with the MOHYSE model, the snow model was decoupled and removed from HMETS, which means that the model version in this study has 11 parameters. It is considered here because of its use and performance in a few studies in arid and snow-less catchments in the United-States (Arsenault et al. 2015, Chen et al. 2018) and due to its more complex routing scheme. Indeed, HMETS has one more routing source, for a total of four (surface runoff, delayed runoff from infiltration, hypodermic flow from the vadose zone reservoir and groundwater flow from the phreatic zone reservoir). The model requires daily precipitation, daily average air temperature and daily PET as inputs. The PET formulation used to generate the PET time-series is the same as for the GR4J model, namely the Oudin formulation (Oudin et al. 2005), but the PET timeseries can be scaled as in the case of MOHYSE. This extra flexibility (more routing options and scalable PET) confer to HMETS the most potential to adapt to the wide array of hydrological conditions in the study site. It also means that the regression-based methods will have difficulty finding links between the catchment descriptors and HMETS' 11 parameters in regionalization.

3.2 Hydrological model calibration

All three hydrological models were calibrated on the 30 catchments in this study. This allowed determining the parameter set related to each model in the case that it is selected as a donor catchment. Furthermore, in some regionalization variants, it is possible to filter donor candidates according to their skill attained in calibration. Therefore, the calibration step is a prerequisite for the project.

The objective function used to calibrate the models is the Nash-Sutcliffe Efficiency metric (NSE; Nash-Sutcliffe 1970) as described in equation 4:

$$NSE = 1 - \frac{\sum_{i=1}^{n}(Q_{io} - Q_{iS})^2}{\sum_{i=1}^{n}(Q_{io} - \overline{Q_o})^2} \qquad \text{(eq. 4)}$$

Where i is the simulation day, Q_{io} is the observed streamflow for day i, Q_{is}, is the simulated streamflow for day i and $\overline{Qo}$ represents the mean of the observed streamflows. An NSE score of 1 means perfect simulation, 0 indicates that the simulation is as good as using the mean observed streamflow as a predictor and negative values indicate that using the mean is a better indication than using the hydrological model. The NSE criterion was used due to its widespread use, making it possible to compare results across studies in meta-analyses such as in Parajka et al. (2013) and Blöschl et al. (2013). Since one of the objectives of this study is to evaluate the performance of regionalization studies in Mexico and compare results with other regions, the use of NSE makes the comparison easier with the other studies. One drawback of using NSE is that the peak flows are weighted more than the low-flows, making the model slightly biased in the peak-flow range. However, this study aims to determine the skill in regionalization, therefore any objective function that considers bias and correlation could be used. Results might differ by changing the objective function to more specific criteria, which would require analyzing the aim of the regionalization in the first place (i.e. which metrics and hydrological indicators are to be modelled at the ungauged site).

The calibration itself was performed using the Shuffled Complex Evolution – University of Arizona (SCE-UA) optimization algorithm of Duan et al. (1993). It has been shown to perform well on these models and for the type of optimization problem at hand (Arsenault et al. 2014).. Even though there could be equifinality in the calibrated parameter sets, Arsenault and Brissette (2014) showed that equifinality did not play an important role in the total uncertainty of regionalization methods over 268 catchments in Canada, which included snow processes. Therefore, a single set of parameters (the best out of 5 trials) for each model-catchment pair was used in this study. Finally, to prevent the possibility of model overfitting, the models were calibrated on the first half of their available years and then validated on the second half. If performance was deemed acceptable in validation, then the calibration was performed once more on the entire time series to maximize the information content in the parameter set, following the findings of Arsenault et al (2018) which demonstrated that the optimal calibration strategy is to use all available years of data if satisfied with the calibration and validation skill on a subset of available years.

3.3 Regionalization framework

The three regionalization approaches were implemented on the 30 catchments using a leave-one-out cross-validation technique. Here, one catchment at a time is considered as pseudo-ungauged and the regionalization methods are applied to this target catchment. Because the streamflow is actually available at the site, it is possible to evaluate the skillfulness of the regionalization methods. The process is repeated for all catchments in the database, and then repeated again independently for each hydrological model.

The regionalization approaches and their variants that were implemented are MLR (classic approach), SP (classic, classic with IDW, and regression-augmented with IDW) and PS (integrated similarity, integrated similarity with IDW and regression-augmented with IDW) for a total of 6 methods. The classic PS was not implemented due to previous studies all finding that the integrated similarity outperforms the classic PS. In all cases where parameter sets are transferred (donor-based; SP and PS), multi-donor averaging was performed using 1 to 10 donors. In addition, following the methodology developed in Oudin et al. (2008), donor catchments were filtered to ensure that they scored an NSE value of at least 0.50 in calibration. Oudin et al. (2008) uses 0.70 as the threshold, but this constraint was relaxed in the present study due to the relatively poor calibration skill of the entire set of model-catchment pairs. This is to exclude parameter sets for which the confidence level from the candidate donor set is low (or very low). However, those catchments are still considered as targets in the cross-validation phase of the work.

The regionalization methods were then analyzed according to the hydrological model and catchment characteristics (climatological and geophysical). Statistical tests were used to classify the regionalization approaches and determine which catchment-model pairs should be considered in future prediction at ungauged sites in Mexico.

Finally, it is important to note that the regionalization performance was measured according to a normalized NSE (NNSE) value from Nossent and Bauwens (2012) which rescales the values between 0 and 1, as shown in equation 5:

$$NNSE = \frac{1}{2 - NSE} \qquad \text{(eq. 5)}$$

Where NNSE is the Normalized NSE value. The NNSE allows easier display of NSE values when there is a large spread between results obtained with different models and methods. When NSE is perfect (NSE=1), the NNSE also takes a value of 1. When NSE is equal to 0, thus as good as a predictor as the mean of observations, NNSE evaluates to 0.5. Finally, the lowest possible NSE (negative infinity) resolves to a NNSE of 0. This linearizes the scale for negative values and allows displaying all information rather than “cutting” bad values from the figures.

4. Results and discussion

4.1 Calibration

The results for the hydrological model calibration are presented in Figure 3. The NNSE values are those obtained when calibrating on the length of the entire available time series, which varies from site to site. The calibration objective function was the NSE value, but the NNSE is shown here to allow comparing with the regionalization results.

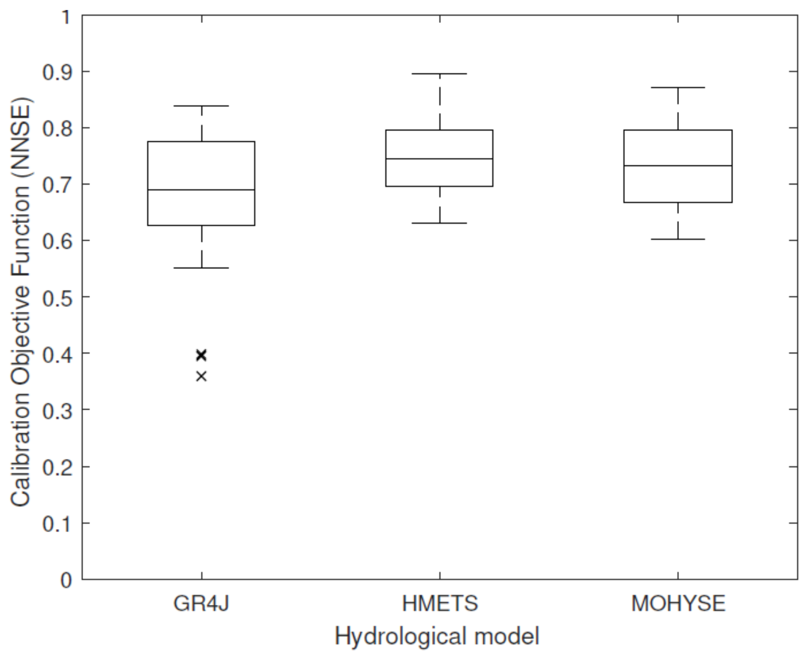

Figure 3: Boxplots showing the distributions of NNSE values obtained for the calibration of the GR4J, HMETS and MOHYSE hydrological models, on the 30 river basins under study.

The three catchments with an NNSE skill below 0.5 were still kept as HMETS and MOHYSE displayed much better performance on them. Clearly, the calibration skill is better for HMETS, followed by MOHYSE and then GR4J. This could be explained by the number of parameters, which follows the same descending order, or simply because the model structure allows more flexibility for HMETS and MOHYSE than GR4J, due to the mass-balance scaling of PET during calibration. Their more complex routing stores can also explain the more reactive and thus more efficient simulations on the smaller catchments.

The parameter sets found during calibration were generally good, but some catchments offered mediocre calibration skill. These were still preserved for the regionalization step because they could still be used as pseudo-ungauged targets. Of course, it is theoretically impossible for a regionalized parameter set to outperform a locally calibrated parameter set for the same model if the parameter optimization algorithm converged properly. However, the fact that multiple donors are added and averaged could lead to better performance than the calibration skill thanks to the increased number of degrees of freedom, although this was not observed in this study. The results must then be analyzed along with the calibration skill on the catchments.

One possible source of error is that of non-stationarity. In this study, model parameter sets are determined based on the calibration data, which span different periods from one site to the next. It is quite possible that a donor catchment could transfer parameters calibrated on the 1960-1980 period to a pseudo-ungauged site, which has verification data available for the years 1990-2015. If a structural shift in temperature or precipitation were to have occurred between those periods, one could imagine that the regionalized streamflow would contain a certain bias. However, we do not anticipate that this plays a major role in this study because all data cover the same time horizon. The residual effects of different trend amplitudes in non-stationarity conditions (e.g. rainfall could be increasing on one catchment and diminishing on another) should be much smaller than the error on the observed streamflow, weather and geophysical data.

4.2 Multiple Linear Regression

The MLR approach was the first method to be implemented in this study and serves as a comparison baseline for the other methods. Figure 4 shows the average NNSE value obtained for each hydrological model on the 30 catchments, each in turn considered as pseudo-ungauged.

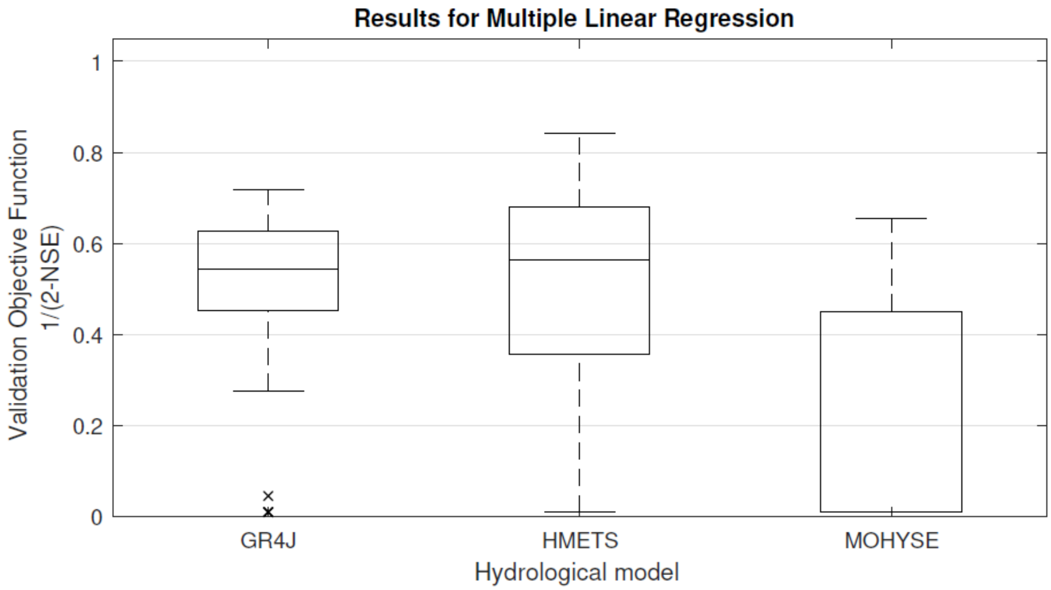

Figure 4: Boxplots showing the distributions of NNSE values when the MLR approach is applied to the 30 pseudo-ungauged river basins with GR4J, HMETS and MOHYSE hydrological models.

It can be seen in figure 4 that the GR4J model is more robust than the HMETS and MOHYSE models when using the MLR approach. MOHYSE is particularly poor as more than 75% of the returned NNSE values are below 0.5, which translates to an NSE value of 0. HMETS, on the other hand, has a higher median and maximum NNSE, but the spread is also larger. This could be explained by the fact that HMETS has a more complex structure, which means that it has more potential to score very well or very badly depending on the transferability of the model parameters. GR4J, having a more rigid structure, is constrained to less extreme variations than HMETS. The variance is thus larger with HMETS than with GR4J. An analysis of the parameter-descriptor regression strength is presented in Table 2. This analysis shows a good correlation level between the catchments' descriptors and a majority of the models' parameters.

Table 2: Regression results for the three hydrological models using the catchment descriptors from the 30 catchments in this case study as predictors of model parameter values.

Hydrological model	Number of parameters	Number of parameters with regression R^2>0.6	Percentage of parameters that are replaced by linear regression in SP and PS
GR4J	4	3	75%
HMETS	11	8	73%
MOHYSE	8	5	63%

4.3 Donor-based methods

The donor-based methods depend on the transfer of entire parameter sets rather than reconstructing a parameter set from statistical regressions. Figures 5, 6 and 7 present the results of the 6 donor-based regionalization algorithms for the hydrological models GR4J, HMETS and MOHYSE, respectively. Note that up to 10 donors were used in this study, but the conclusions are the same as using a maximum of 5 donors. For clarity, the number of donors is thus limited to 5 in these figures.

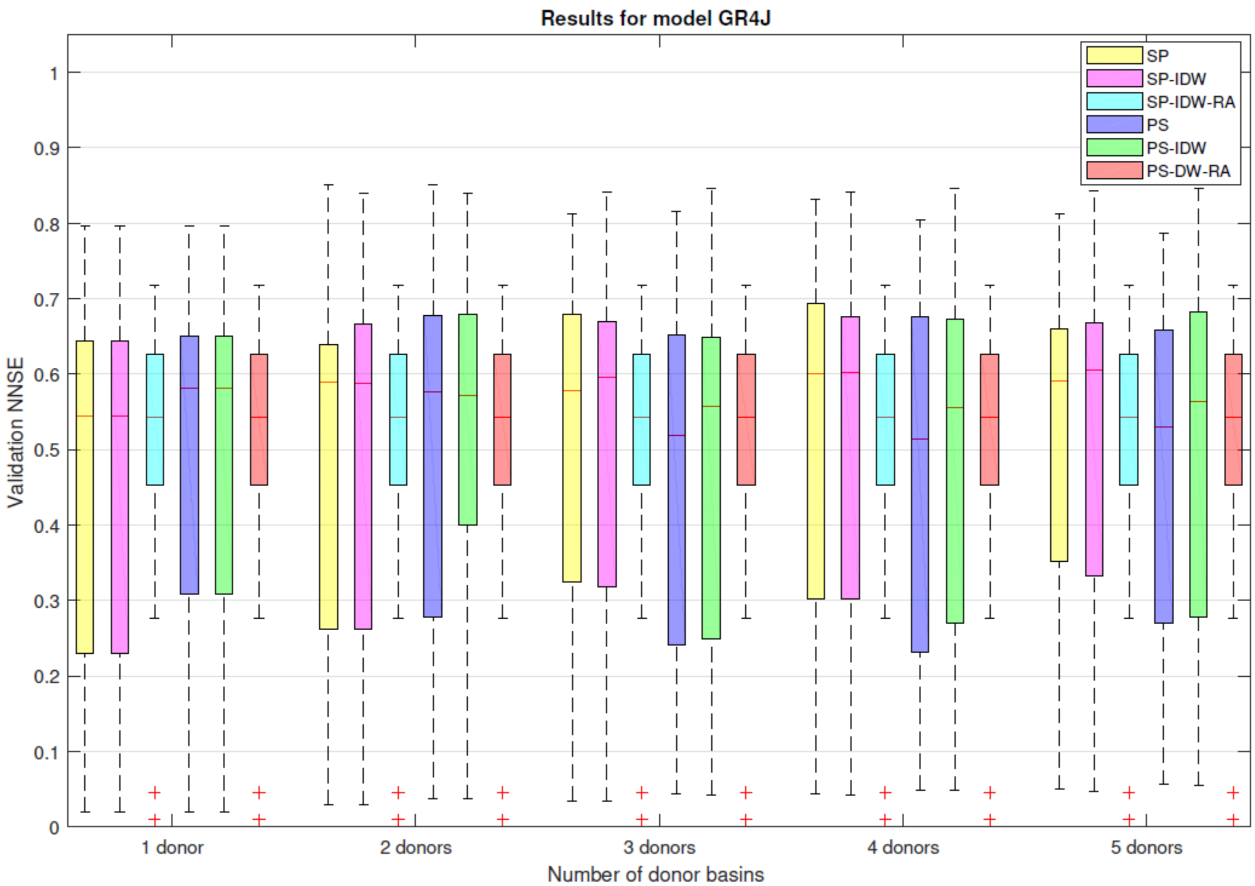

Figure 5: Boxplots showing the distributions of NNSE values when donor-based regionalization methods are applied to the 30 pseudo-ungauged river basins with the GR4J hydrological model.

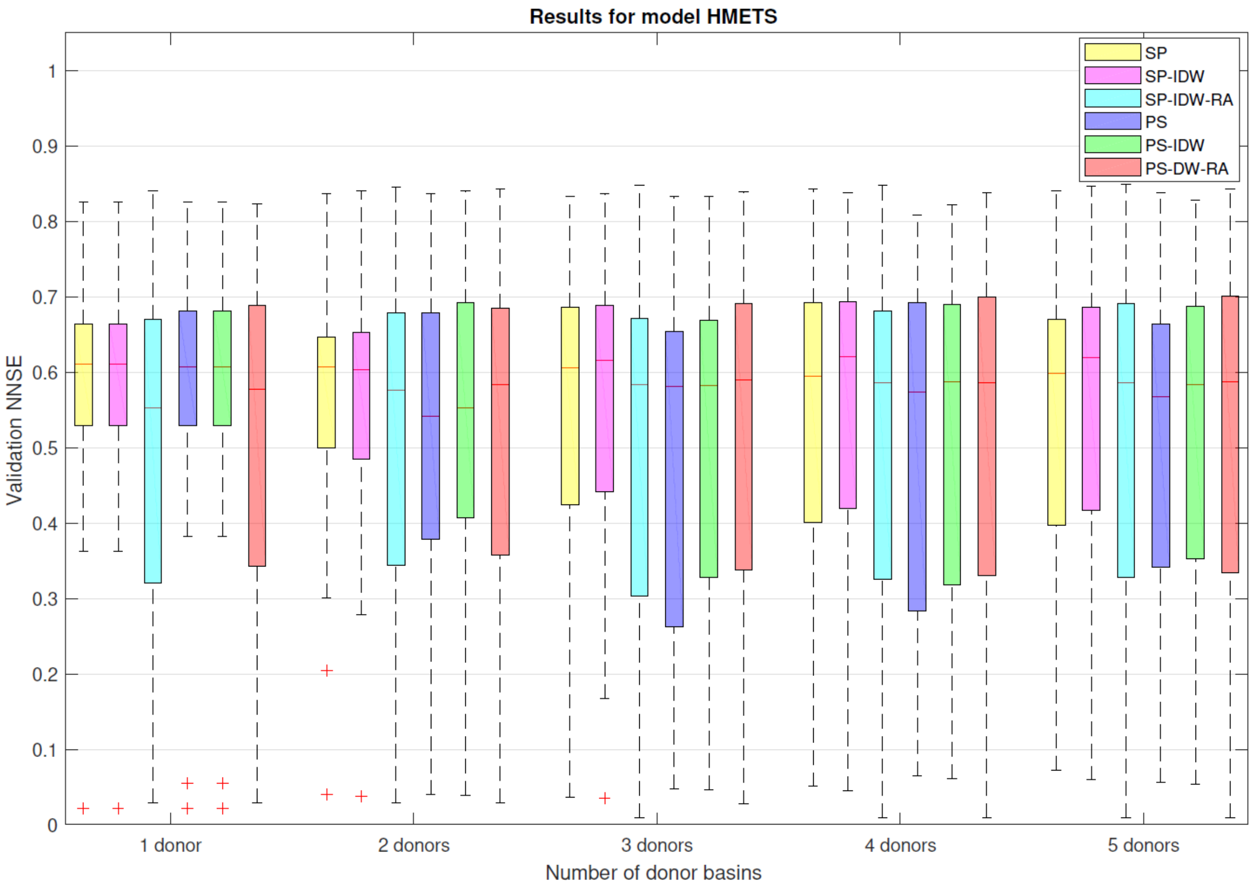

Figure 6: Boxplots showing the distributions of NNSE values when the donor-based regionalization methods are applied to the 30 pseudo-ungauged river basins with the HMETS hydrological model.

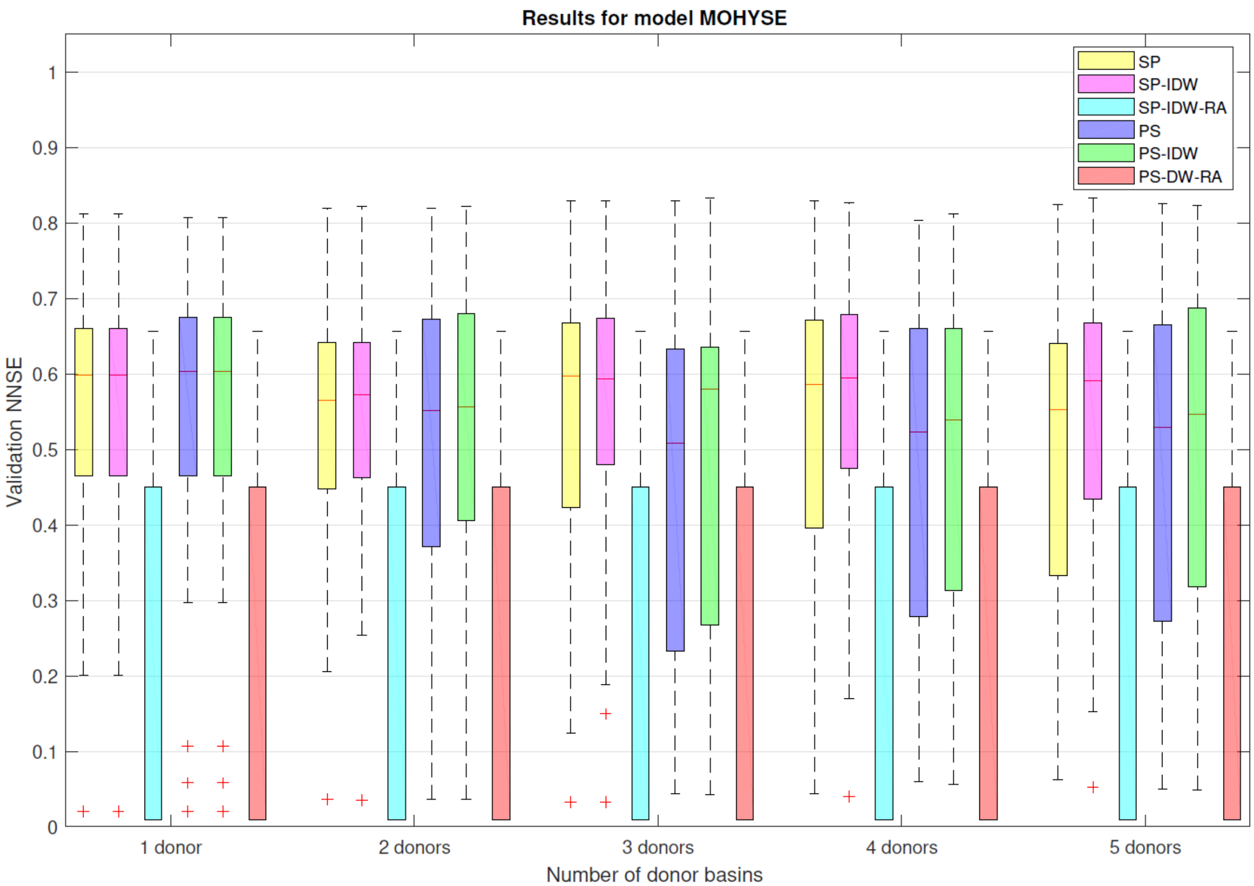

Figure 7: Boxplots showing the distributions of NNSE values when the donor-based regionalization methods are applied to the 30 pseudo-ungauged river basins with the MOHYSE hydrological model.

A few notable points emerge from the analysis of figures 5-7. First, it can be seen that using multiple donors does not always improve upon the single-donor approach and that the result depends on the regionalization method. When looking at the median values and inter-quartile ranges, multi-donor averaging seems to return positive outcomes for the SP method, but negative ones for the PS method. This is true for all models, but MOHYSE is to benefit the least from the procedure. However, in cases where multiple donors are used, the inverse distance weighting is more robust than the equal weights averaging. The hydrologically heterogeneous study area contributes to these results, in part by the fact that adding simulations generated with parameters coming from progressively farther catchments will likely add increasingly different catchments, thus using inverse distance weighting mitigates some of that risk. It is worth noting that no statistical analysis was performed on these data, namely due to the small sample size, the fact that some catchments have the same “spatial proximity donor” and “physical similarity donor” (thus reducing effective differential sample size), and that comparing results in the negative NSE range (NNSE<0.5) is generally not useful for quantitative analyses. Therefore, the results should be viewed as qualitative more than quantitative.

Second, the choice of hydrological models has a strong influence on the regionalization method performance. GR4J, for example, scores lower than the other two models for SP and PS methods, but is more robust to the regression-augmented approaches. This may be due to its simpler structure, making it more robust but less adaptive to varying conditions. While MOHYSE is strongly penalized with the regression-augmented approaches, GR4J yields the smallest distributions of NNSE values but with a decrease in performance with respect to the other SP approaches in particular, and HMETS is able to maintain a similar performance when multiple donors are used. MOHYSE performs quite badly in all these cases and is clearly not suited to make use of the regression-augmented approaches.

Overall, using all catchments (including those with NNSEs below 0.5), HMETS displays a slightly better performance than the others do. The more complex structure seems to allow it to adapt to different catchments even in regionalization mode. For donor-based methods, HMETS is followed closely in terms of performance and robustness by GR4J, with MOHYSE lagging behind. For the MLR approach, HMETS performs better than GR4J on average, as the median value is higher for HMETS than GR4J. Again, this could be due to the more flexible model structure, with the 2 extra routing components and the scalable PET, giving HMETS a better chance to fit to the catchment characteristics through parameter-descriptor correlations. As for the lower quartiles, both methods have approximately 50% of their validation NNSE values below 0.50, which translates to an NSE value of 0. Therefore, the lower quartiles can be considered similar in terms of bad performance. MOHYSE has over 75% of its NNSE values below 0.50, which ranks it last between the three hydrological models. While MOHYSE lies between GR4J and HMETS in terms of complexity and in calibration performance, the results obtained throughout this study show that it performs worse than the others in the MLR approach. This also explains the poor results in the regression-augmented PS and SP variants.

It is clear that the SP methods performed generally better than the PS methods, which might be an indication of the difficulty of finding hydrologically similar catchments in such a heterogeneous region, of the quality of the descriptor and hydrological data or of the difficulty in establishing reliable relationships between the descriptors and hydrological parameters. Indeed, the 30 catchments cover a wide array of climatological and geophysical attributes, making it somewhat more difficult to find adequate donor catchments. The results seem to be in-line with the model structure complexity, with increased performance being linked to increasing model complexity. Future work could include even more complex models to determine if it could be a way forward in regionalization in hydrologically heterogeneous regions.

A detailed example from the Atoyac watershed at the Zimatlan gauging station is presented in figure 8 in order to illustrate the regionalization methods' effects on the simulated hydrographs. The catchment was selected as it represents an average catchment on many levels amongst the ones available in this study. Namely, the 2647 km^2 catchment has a Runoff Ratio (defined as the ratio of mean runoff to mean aerial precipitation) of 0.1, an Aridity Index (defined as the precipitation to PET ratio) of 0.63, is mainly agricultural with 94% of the catchment area being allocated to agriculture and receives an average of 800mm in total precipitation every year. These values fall between the 25th and 75th percentile of all basins (see Table 1).

Most of the previously discussed points can be seen in Figure 8. First, it is possible to see the effect of using multiple donors as the top panels were generated using a single donor and the lower panels use three donors. It can be seen that in this case, the closest donor (in terms of physical distance) is the same as the most similar donor, resulting in overlapping curves for SP, PS, SP-IDW and PS-IDW. The same happens for SP-IDW-RA and PS-IDW-RA. This scenario was the same for 15 out of the 30 catchments, meaning that for half of the database, the closest and most similar donor were the same. This is not too surprising under the hypothesis that closer catchments should be more similar, and is even more likely in very heterogeneous regions such as the one under study. From the lower panels in figure 8, it is clear that weighting the multi-donor hydrographs by the inverse of the distance plays a crucial role in mitigating the effects of selecting progressively more different catchments. Indeed, the IDW variants are kept closer to the observed flow than the unweighted PS and SP methods, which produce significantly larger hydrographs.

As for the Regression-augmented approaches, their skill is directly correlated to the MLR performance as more than half of the parameters are replaced by the MLR technique for each model. For example, in panels 8a) and 8c), it can be seen that the shape of the PS-IDW-RA hydrograph is shaped much like the MLR hydrograph. For GR4J, this is because all parameters except for the groundwater exchange parameter are replaced by the MLR value. For MOHYSE, the mass-balance parameters are preserved but the routing parameters are replaced by the MLR-derived ones, giving the odd hydrograph seen in figures 8c) and 8f). HMETS, on the other hand, preserves the routing store depths as well as some of the unit hydrograph parameters. This explains why the hydrograph shape is relatively well modelled. In all cases, the regression-based methods modify the PET-scaling parameters, thus possibly removing some of the important relationships between mass-balance and other production and routing parameters.

In terms of choice of hydrological model, in this test-case the MOHYSE and HMETS models seem to perform relatively well. However, in some cases (not shown) involving MLR, the parameter sets that are transferred are not physically coherent anymore and generate errors (such as returning infinite streamflow) during the hydrological modelling for these two models, whereas GR4J was always able to generate a hydrograph. Therefore the complexity of the hydrological model can become a drawback when attempting to regionalize parameters in this context.

The fact that the catchment is very dry except for a 2-3 month period during rain season is also important to consider. Basically, other catchments that receive water year-round are parameterized as such and take this into account, for example in the size of the routing stores and groundwater exchange coefficients. When transferring these parameters from a humid catchment to a dry one for regionalization purposes, problems can arise in terms of mass-balance and event timing, as can be seen in figures 8d)-9f), where a humid catchment is considered as the 3rd donor and distorts the hydrographs for the unweighted PS and SP methods. Overall, weighted multi-donors seem to improve the overall performance and the more complex models also seem to make use of their flexibility to generate acceptable hydrographs. Unfortunately, as mentioned earlier, the more complex models are not guaranteed to find feasible results if MLR is involved.

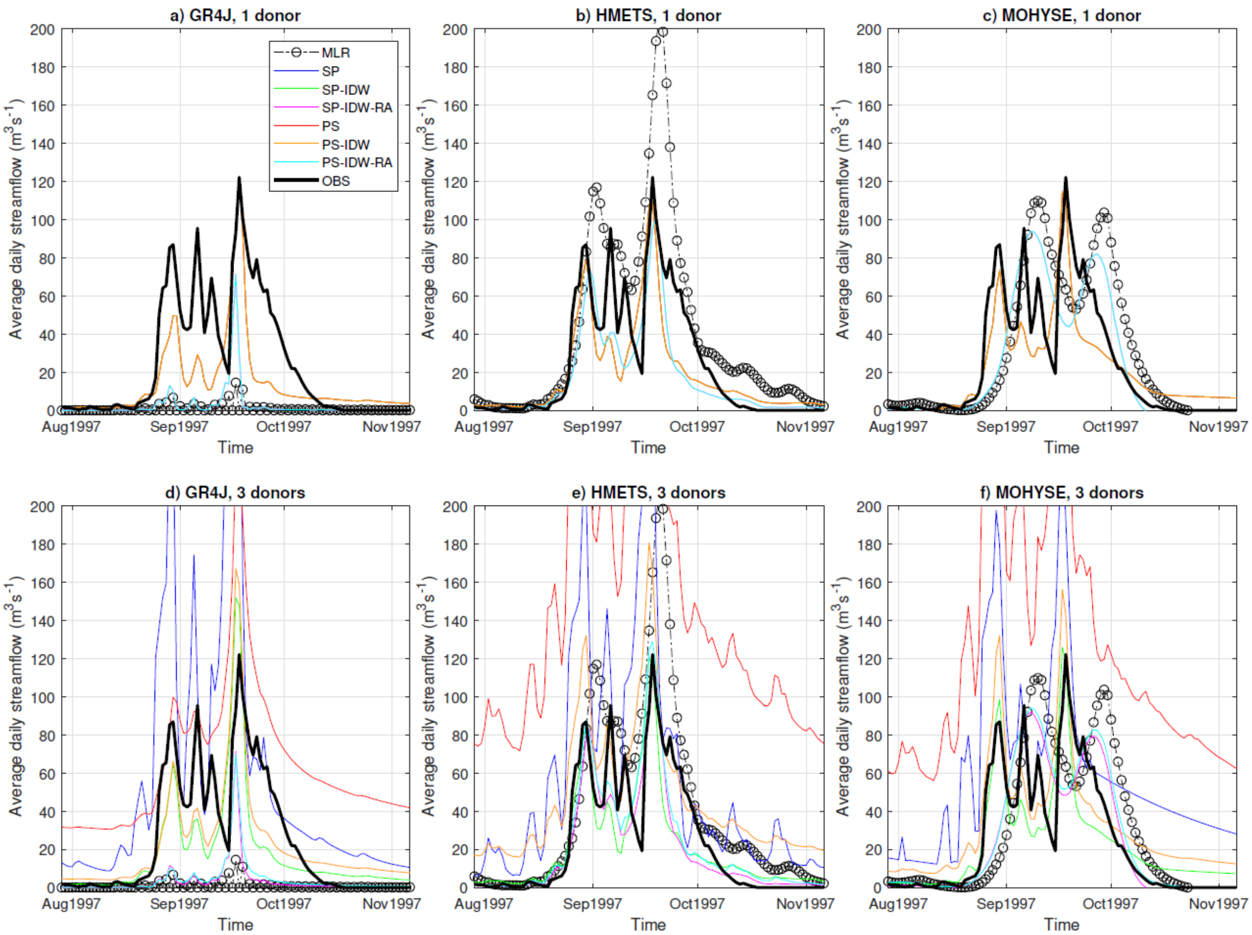

Figure 8: Regionalized hydrographs on the Atoyac catchment from the seven methods used in this study with hydrological models GR4J (leftmost panels), HMETS (center panels) and MOHYSE (rightmost panels) with one donor (top panels) and three donors (bottom panels). In the top panels, SP, PS, SP-IDW and PS-IDW are overlapping. Likewise, PS-IDW-RA and SP-IDW-RA are also overlapping.

4.4 Regionalization methods performance as a function of the basins' Runoff Ratio and Aridity index

In this paper, we excluded streamflow-based indices for the regionalization step due to the unavailability of data, by definition, in ungauged sites. However, to better understand the limitations of predicting streamflow in ungauged basins, an analysis of the model performance relative to the runoff ratio and aridity index for each of the studied catchments is presented in figures 9-12.

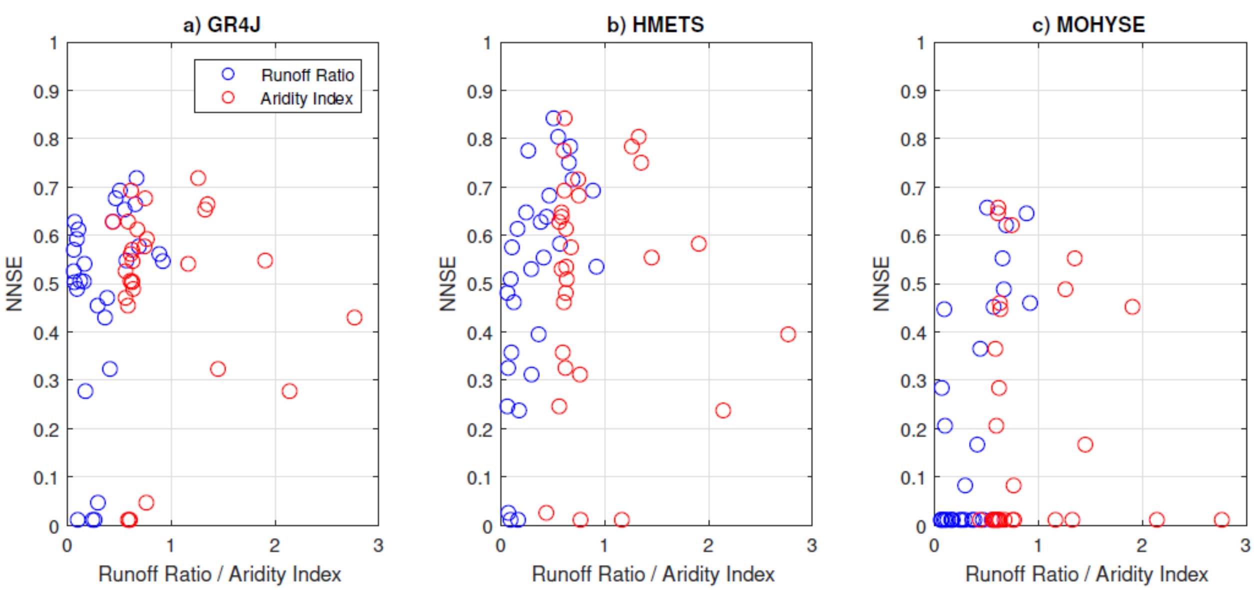

Figure 9: Ranges of runoff ratios for the 30 basins under study, and of the NNSE values when the MLR method is used with the GR4J, HMETS and MOHYSE hydrological models.

From Figure 9, it is possible to see that with the MLR method, GR4J and HMETS perform oppositely on the catchments with low runoff ratio, where GR4J seems better suited than the other models on most arid basins and HMETS performs better than GR4J on the others. MOHYSE performs generally much worse than the others for all Runoff Ratios, often providing incoherent and physically impossible parameter sets, resulting in error-state or extremely poor simulations. Furthermore, there seems to be a trend in which the humid and semi-arid catchments display higher NNSE values than the arid basins, adding weight to the multiple studies on this subject in the literature as described in section 1.

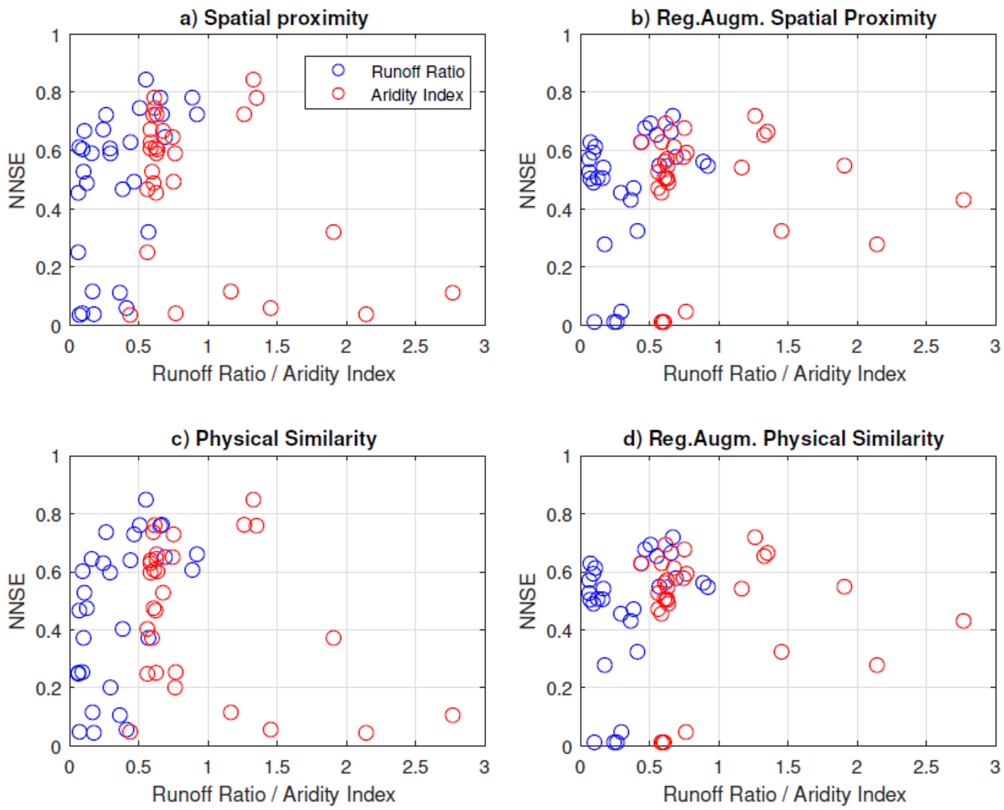

Figure 10: Ranges of runoff ratio and aridity index for the 30 basins under study, and of the NNSE values when donor-based regionalization methods are used with the GR4J hydrological model.

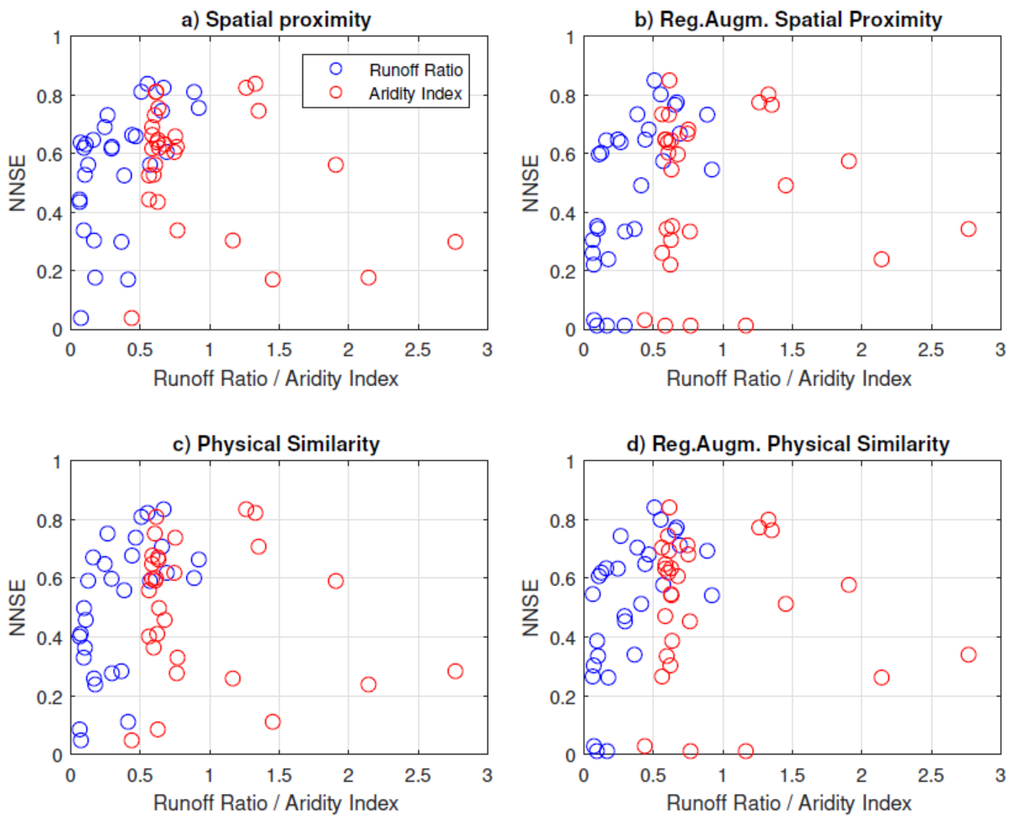

Figure 11: Ranges of runoff ratio and aridity index for the 30 basins under study, and of the NNSE values when donor-based regionalization methods are used with the HMETS hydrological model.

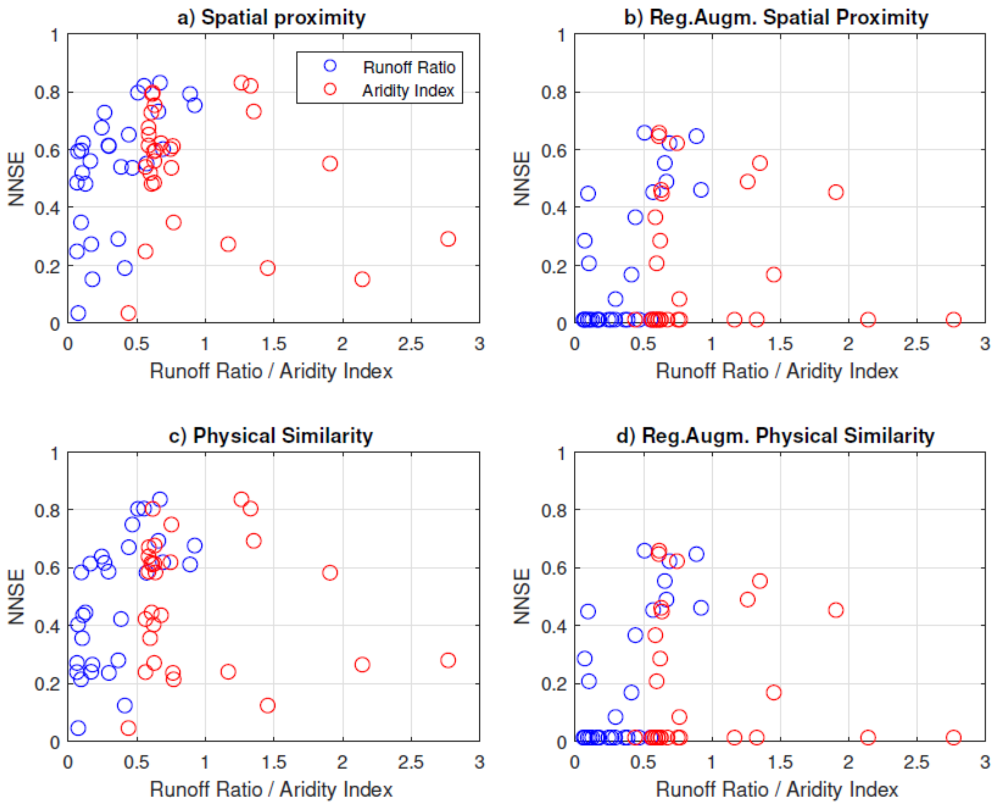

Figure 12: Ranges of runoff ratio and aridity index for the 30 basins under study, and of the NNSE values when donor-based regionalization methods are used with the MOHYSE hydrological model.

For GR4J and HMETS (figures 10 and 11), the behaviour is similar between the Regression-augmented approaches and the MLR results from figure 9 due to the fact that 75% of the parameters were replaced by the regression-derived parameters (see Table 2). This indicates that even though there were high correlations between the model parameters and the catchment descriptors, they did not carry predictive power in the regionalization step. Two possible reasons for this are that the data might contain quality issues and that the hydrological model-parameter identifiability hypothesis is wrong. Parameter identifiability problems seem to be the underlying cause because the PS method, which is based on the parameter transposition based on physical catchment descriptors, performs relatively well, leading to believe that the data quality is at least sufficient to perform this type of analysis. However, the spatial proximity method was generally found to perform the best in all situations, meaning that the data quality (namely the catchment descriptors) could be one of the pitfalls of the regionalization framework in this project. In any case, evidence suggests that performing regionalization on heterogeneous catchments is a risky proposition and depends on data quality, data availability and the number of catchments that can be used as donors in the region.

Figures 10-12 also show that for all models, the catchments with the lowest runoff ratios offer the lowest regionalization skills. This might be due to the initial calibration skill, as a poor parameter set in calibration would mean that the parameters do not represent the underlying processes very well, leading to a propagation of those errors during regionalization. For example, the aridity index is inversely correlated to the calibration skill (R^2=0.49) for GR4J, which could generate less reliable parameter sets, but the two other models see no correlation between the aridity index and calibration skill (R^2=0.03 in both cases). A second verification revealed that the calibration skill is not strongly correlated to the runoff ratio either (R^2 between 0.07 and 0.19), indicating that the results are most likely only due to the particularities of regionalizing on arid and semi-arid catchments from a hydrological point of view. Furthermore, although there is a difference in calibration NSE between the “arid” group of catchments (median NSE of 0.61) and the “non-arid” group (median NSE of 0.70), this difference cannot explain the results in Figures 10-12, which show that the performance in regionalization skill in the NNSE transformed space is significantly lower. Therefore, the calibration skill cannot explain this difference by itself. Figures 10-12 also show that the Regression-Augmented versions of SP and PS depend on the hydrological model, being generally positive for GR4J on humid catchments but overwhelmingly negative for MOHYSE.

Finally, the regionalization methods were applied once again but this time without the donor catchment quality filter, i.e. by completely removing the constraint requiring that a donor basin has a calibration NSE value greater than 0.50. The results (not shown here) were slightly worse than when the filter is applied, therefore confirming that this practice is beneficial when regionalizing in a hydrologically heterogeneous region.

4.5 Impact of the hydrological model

The choice of hydrological model was shown to be particularly important. The MOHYSE model was the worst one in most approaches and on almost all catchments. It performed particularly poorly when regionalizing on the arid catchments and in methods involving MLR. HMETS offered the best overall performance, performing slightly better than GR4J in the MLR and donor-based methods. As for GR4J, its simplicity and robustness to MLR make it a good choice for regionalization in heterogeneous regions, especially if the catchment is suspected to be located in an arid climate region. As could be expected, no model was better everywhere but the HMETS model paired with a donor-based regionalization method offered the best chance at obtaining a good skill at the cost of deteriorating the NNSE of the de facto poor regionalization candidate catchments.

5. Conclusion

This study investigated the performance of common regionalization approaches on a set of diverse catchments using three hydrological models. It was found that the linear-regression based methods were incapable of competing with the other methods and that the donor-based spatial-proximity regionalization method was to be preferred. Its simplicity and relative robustness make it a prime candidate, especially if there are a few neighboring catchments than can be used to generate a multiple-donor average. In this way, there is less risk related to the quality of catchment descriptors. The HMETS model, with 11 parameters, was overall slightly better than the 4-parameter GR4J model, and the 8-parameter MOHYSE model was the least reliable model in these circumstances. Therefore, the choice of a hydrological model plays an important role, but the sheer number of parameters and model complexity do not seem to be predictors of model behaviour in regionalization. It was also found that the regionalization methods generally performed as expected based on their hydroclimatic characteristics. Arid catchments were more difficult to calibrate and displayed lower skill in regionalization, whereas humid catchments were generally more reliably to regionalize. This study highlights the need to (1) improve the quality of measured data in the region, (2) ensure that a sufficient number of catchments are available to donate parameter sets and (3) evaluate the hydrological model's robustness to regionalization. Future work should investigate the impacts of data quality on regionalization, include more complex hydrological models and validate this study's conclusions on other hydrologically heterogeneous regions of the world.

Acknowledgements

The authors would like to thank the École de technologie supérieure and the Natural Sciences and Engineering Research Council of Canada for partial funding of this study. The authors would also like to thank the INEGI and CONAGUA for providing the hydrometeorological and geographical data used in the study. Finally, we would like to thank the two anonymous reviewers whose comments on a previous version of this work were insightful and helped shape the paper in its current form.

References

Allende, T. C., Mendoza, M. E., Granados, E. M. L., & Manilla, L. M. M. (2009). Hydrogeographical regionalisation: an approach for evaluating the effects of land cover change in watersheds. A case study in the Cuitzeo Lake watershed, Central Mexico. *Water Resources Management*, *23*(12), 2587-2603.

Álvarez-Olguín, G., Hotait-Salas, N., & Sustaita-Rivera, F. (2011). Identificación de regiones hidrológicas homogéneas mediante análisis multivariado. *Ingeniería, investigación y tecnología*, *12*(3), 277-284.

Arsenault, R., Poulin, A., Côté, P., & Brissette, F. (2013). Comparison of stochastic optimization algorithms in hydrological model calibration. *Journal of Hydrologic Engineering*, *19*(7), 1374-1384.

Arsenault, R., & Brissette, F. P. (2014). Continuous streamflow prediction in ungauged basins: The effects of equifinality and parameter set selection on uncertainty in regionalization approaches. *Water Resources Research*, *50*(7), 6135-6153.

Arsenault, R., Poissant, D., & Brissette, F. (2015). Parameter dimensionality reduction of a conceptual model for streamflow prediction in Canadian, snowmelt dominated ungauged basins. *Advances in water resources*, *85*, 27-44.

Arsenault, R., Gatien, P., Renaud, B., Brissette, F., & Martel, J. L. (2015). A comparative analysis of 9 multi-model averaging approaches in hydrological continuous streamflow simulation. *Journal of Hydrology*, *529*, 754-767.

Arsenault, R., Brissette, F., Martel, J-L. (2018).) The hazards of split-sample validation in hydrological model calibration. *Journal of Hydrology, 566,* 346-362.

Bárdossy, A. (2007). Calibration of hydrological model parameters for ungauged catchments. *Hydrology and Earth System Sciences*, *11*, 703-710.

Blöschl, G. (Ed.) et al. (2013). *Runoff prediction in ungauged basins: synthesis across processes, places and scales*. Cambridge University Press.

Burn, D. H., & Boorman, D. B. (1993). Estimation of hydrological parameters at ungauged catchments. *Journal of Hydrology*, *143*(3-4), 429-454.

Castaneda Gonzales, Mariana. (2014). *Análisis y simulación hidrológica de cuencas tropicales en México aplicando el modelo hidrológico global simplificado al extremo (MOHYSE).* Diploma thesis. Veracruzana University, Xalapa, Mexico.

Chen, J., Li, C., Brissette, F. P., Chen, H., Wang, M., & Essou, G. R. (2018). Impacts of correcting the inter-variable correlation of climate model outputs on hydrological modeling. *Journal of Hydrology*, *560*, 326-341.

Conagua. (2008). Banco Nacional de Datos de Aguas Superficiales (BANDAS). Retrieved from http://www.conagua.gob.mx/CONAGUA07/Contenido/Documentos/Portada%20BANDS.htm

Domínguez-Mora, R., Arganis-Juárez, M. L., Guzmán-García, H., Carrizosa-Elizondo, E., Esquivel-Garduño, G., & Fuentes-Mariles, G. E. (2016). Modelos regionales de escurrimientos máximos instantáneos en la república mexicana. *Tecnología y ciencias del agua*, *7*(5), 15-32.

Duan, Q. Y., Gupta, V. K., & Sorooshian, S. (1993). Shuffled complex evolution approach for effective and efficient global minimization. *Journal of optimization theory and applications*, *76*(3), 501-521.

FAO. (2016). Base de référence mondiale. Retrieved from http://www.fao.org/soils-portal/soil-survey/classification-des-sols/base-de-reference-mondiale/fr/

Fortin, V. & R. Turcotte. (2007). SCA7420 : *Le modèle hydrologique MOHYSE (bases théoriques et manuel de l'usager)*. Université du Québec à Montréal.

He, Y., Bárdossy, A., & Zehe, E. (2011). A review of regionalisation for continuous streamflow simulation. *Hydrology and Earth System Sciences*, *15*(11), 3539.

Hrachowitz, M., Savenije, H. H. G., Blöschl, G., McDonnell, J. J., Sivapalan, M., Pomeroy, J. W., ... & Fenicia, F. (2013). A decade of Predictions in Ungauged Basins (PUB)—a review. *Hydrological sciences journal*, *58*(6), 1198-1255.

Inegi. (2016). Geografia. Retrieved from http://www.fao.org/soils-portal/soil/survey/classification -des-sols/base-de-reference-mondiale/fr/

Kokkonen, T. S., Jakeman, A. J., Young, P. C., & Koivusalo, H. J. (2003). Predicting daily flows in ungauged catchments: model regionalization from catchment descriptors at the Coweeta Hydrologic Laboratory, North Carolina. *Hydrological Processes*, *17*(11), 2219-2238.

Livneh, B., Bohn, T. J., Pierce, D. W., Munoz-Arriola, F., Nijssen, B., Vose, R., ... & Brekke, L. (2015). A spatially comprehensive, hydrometeorological data set for Mexico, the US, and Southern Canada 1950–2013. *Scientific data*, *2*, 150042.

Luis-Pérez, F. E., Cruz-Barbosa, R., & Álvarez-Olguin, G. (2011). Regional flood frequency estimation for the Mexican Mixteca region by clustering techniques. In Mexican International Conference on Artificial Intelligence, Springer, Berlin, Heidelberg. 249-260.

Martel, J. L., Demeester, K., Brissette, F., Poulin, A., & Arsenault, R. (2017). HMETS-A simple and efficient hydrology model for teaching hydrological modelling, flow forecasting and climate change impacts. *International Journal of Engineering Education*, *33*(4), 1307-1316.

McIntyre, N., Lee, H., Wheater, H., Young, A., & Wagener, T. (2005). Ensemble predictions of runoff in ungauged catchments. *Water Resources Research*, *41*(12), W12434.

Merz, R., & Blöschl, G. (2004). Regionalisation of catchment model parameters. *Journal of hydrology*, *287*(1-4), 95-123.

Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I—A discussion of principles. *Journal of hydrology*, *10*(3), 282-290.

Nossent, J., & Bauwens, W. (2012). Optimising the convergence of a Sobol'sensitivity analysis for an environmental model: application of an appropriate estimate for the square of the expectation value and the total variance. International congress on environmental modelling and software, Managing Resources of a Limited Planet: Pathways and Visions under Uncertainty, Sixth Biennial Meeting, Leipzig, Germany, edited by: Seppelt, R., Voinov, A., Lange, S., and Bankamp, D., 1080–1087.

Ouarda, T. B. M. J., Bâ, K. M., Diaz-Delgado, C., Cârsteanu, A., Chokmani, K., Gingras, H., ... & Bobée, B. (2008). Intercomparison of regional flood frequency estimation methods at ungauged sites for a Mexican case study. *Journal of Hydrology*, *348*(1-2), 40-58.

Oudin, L., Hervieu, F., Michel, C., Perrin, C., Andréassian, V., Anctil, F., & Loumagne, C. (2005). Which potential evapotranspiration input for a lumped rainfall–runoff model?: Part 2—Towards a simple and efficient potential evapotranspiration model for rainfall–runoff modelling. *Journal of hydrology*, *303*(1-4), 290-306.

Oudin, L., Andréassian, V., Perrin, C., Michel, C., & Le Moine, N. (2008). Spatial proximity, physical similarity, regression and ungaged catchments: A comparison of regionalization approaches based on 913 French catchments. *Water Resources Research*, *44*(3), W03413.

Oudin, L., Kay, A., Andréassian, V., & Perrin, C. (2010). Are seemingly physically similar catchments truly hydrologically similar?. *Water Resources Research*, *46*(11), W11558.

Parajka, J., Merz, R., & Blöschl, G. (2005). A comparison of regionalisation methods for catchment model parameters. *Hydrology and earth system sciences*, *9*(3), 157-171.

Parajka, J., Viglione, A., Rogger, M., Salinas, J. L., Sivapalan, M., & Blöschl, G. (2013). Comparative assessment of predictions in ungauged basins–Part 1: Runoff-hydrograph studies. *Hydrology and Earth System Sciences*, *17*(5), 1783-1795.

Perrin, C., Michel, C., & Andréassian, V. (2003). Improvement of a parsimonious model for streamflow simulation. *Journal of hydrology*, *279*(1-4), 275-289.

Poissant, D., Arsenault, R., & Brissette, F. (2017). Impact of parameter set dimensionality and calibration procedures on streamflow prediction at ungauged catchments. *Journal of Hydrology: Regional Studies*, *12*, 220-237.

Rojas-Serna, C., Michel, C., Perrin, C., Andreassian, V., Hall, A., Chahinian, N., & Schaake, J. (2006). Ungauged catchments: how to make the most of a few streamflow measurements?. *IAHS publication*, *307*, 230.

Razavi, T., & Coulibaly, P. (2012). Streamflow prediction in ungauged basins: review of regionalization methods. *Journal of hydrologic engineering*, *18*(8), 958-975.

Reichl, J. P. C., Western, A. W., McIntyre, N. R., & Chiew, F. H. S. (2009). Optimization of a similarity measure for estimating ungauged streamflow. *Water Resources Research*, *45*(10), W10423.

Samuel, J., Coulibaly, P., & Metcalfe, R. A. (2011). Estimation of continuous streamflow in Ontario ungauged basins: comparison of regionalization methods. *Journal of Hydrologic Engineering*, *16*(5), 447-459.

Sivapalan, M., Takeuchi, K., Franks, S. W., Gupta, V. K., Karambiri, H., Lakshmi, V., ... & Oki, T. (2003). IAHS Decade on Predictions in Ungauged Basins (PUB), 2003–2012: Shaping an exciting future for the hydrological sciences. *Hydrological sciences journal*, *48*(6), 857-880.

Troin, M., Arsenault, R., & Brissette, F. (2015). Performance and uncertainty evaluation of snow models on snowmelt flow simulations over a Nordic catchment (Mistassibi, Canada). *Hydrology*, *2*(4), 289-317.

Troin, M., Arsenault, R., Martel, J. L., & Brissette, F. (2017). Uncertainty of hydrological model components in climate change studies over two Nordic Quebec catchments. *Journal of Hydrometeorology*, 19(1), 27-46.

Velázquez, J. A., Anctil, F., Ramos, M. H., & Perrin, C. (2011). Can a multi-model approach improve hydrological ensemble forecasting? A study on 29 French catchments using 16 hydrological model structures. *Advances in Geosciences*, *29*, 33-42.

Velázquez, J. A., Troin, M., & Caya, D. (2015). Hydrological modeling of the Tampaon River in the context of climate change. *Tecnología y Ciencias del Agua*, *6*(5), 17-30.

Vicente-Serrano, S. M. (2006). Spatial and temporal analysis of droughts in the Iberian Peninsula (1910–2000). *Hydrological Sciences Journal*, *51*(1), 83-97.

Wagener, T., Sivapalan, M., Troch, P., & Woods, R. (2007). Catchment classification and hydrologic similarity. *Geography compass*, *1*(4), 901-931.

Yadav, M., Wagener, T. & Gupta, H., (2007). Regionalization of constraints on expected watershed response behavior for improved predictions in ungauged basins. *Advances in Water Resources*, 30(8), 1756-1774.

Zhang, Y., & Chiew, F. H. (2009). Relative merits of different methods for runoff predictions in ungauged catchments. *Water Resources Research*, *45*(7), W07412.

APPENDIX A – ACRONYMS AND INITIALISMS

CD	Catchment Descriptor
GR4J	Génie Rural à 4 paramètres Journalier (hydrological model)
HMETS	Hydrological Model – École de technologie Supérieure (hydrological model)
IDW	Inverse Distance Weighting
MLR	Multiple Linear Regression
MOHYSE	Modèle Hydrologique Simplifié à l’Extrême (hydrological model)
NSE	Nash-Sutcliffe Efficiency
NNSE	Normalized Nash-Sutcliffe Efficiency (NSE with values normalized to [0, 1])
PET	Potential Evapotranspiration
PS	Physical Similarity
RA	Regression-Augmented (added to SP-IDW/PS-IDW to indicate some parameters in the donor set are modified by Multiple Linear Regression)
SCE-UA	Shuffled Complex Evolution – University of Arizona (Global optimization algorithm)
SP	Spatial Proximity