UNIVERSIDADE DE SÃO PAULO
Escola de Engenharia de São Carlos

Streamflow predictions in ungauged basins by using the Grunsky's Method

Previsão de vazões em bacias não monitoradas usando o Método de Grunsky

Aluno: Bruno Ken Marchezepe

Orientador: Paulo Tarso Sanches de Oliveira

EESC · USP

UNIVERSITY OF SÃO PAULO
SÃO CARLOS SCHOOL OF ENGINEERING
DEPARTMENT OF HYDRAULICS AND SANITARY ENGINEERING

BRUNO KEN MARCHEZEPE

Streamflow predictions in ungauged basins by using the Grunsky's Method

Previsão de vazões em bacias não monitoradas usando o Método de Grunsky

VERSÃO CORRIGIDA

São Carlos

2023

BRUNO KEN MARCHEZEPE

Streamflow predictions in ungauged basins by using the Grunsky's Method

Previsão de vazões em bacias não monitoradas usando o Método de Grunsky

Dissertation presented to the São Carlos School of Engineering, University of São Paulo, in partial fulfillment of the requirements for the degree of Master in Science: Hydraulics and Sanitary Engineering

Advisor: Prof. Dr. Paulo Tarso Sanches de Oliveira

VERSÃO CORRIGIDA

São Carlos

2023

Ficha catalográfica elaborada pela Biblioteca Prof. Dr. Sérgio Rodrigues Fontes da EESC/USP com os dados inseridos pelo(a) autor(a).

M316s Marchezepe, Bruno Ken
Streamflow predictions in ungauged basins by using the Grunsky's Method / Bruno Ken Marchezepe; orientador Paulo Tarso Sanches de Oliveira. São Carlos, 2023.

Dissertação (Mestrado) - Programa de Pós-Graduação em Engenharia Hidráulica e Saneamento e Área de Concentração em Hidráulica e Saneamento -- Escola de Engenharia de São Carlos da Universidade de São Paulo, 2023.

1. regionalization. 2. hydrological group. 3. temporal scale. 4. tropical basins. 5. mean temperature. 6. aridity index. 7. potential evapotranspiration. I. Título.

Eduardo Graziosi Silva - CRB - 8/8907

FOLHA DE JULGAMENTO

Candidato: Engenheiro **BRUNO KEN MARCHEZEPE.**

Título da dissertação: "Previsão de vazões em bacias não monitoradas usando o Método de Grunsky".

Data da defesa: 22/06/2023.

Comissão Julgadora	Resultado
Prof. Dr. **Paulo Tarso Sanches de Oliveira** **(Orientador)** (Universidade Federal de Mato Grosso do Sul/UFMS)	Aprovado
Prof. Associado **Sérgio Nascimento Duarte** (Escola Superior de Agricultura "Luiz de Queiroz"/ESALQ-USP)	Aprovado
Prof. Dr. **Adriano Rolim da Paz** (Universidade Federal da Paraíba/UFPb)	Aprovado

Coordenador do Programa de Pós-Graduação em Engenharia Hidráulica e Saneamento:
Prof. Associado **Juliano José Corbi**

Presidente da Comissão de Pós-Graduação:
Prof. Titular **Carlos De Marqui Junior**

ACKNOWLEDGEMENTS

I would like to thank my wife Rafaella for her unwavering support and encouragement throughout this journey. Her constant belief in me has been invaluable, and I am grateful for her love and patience. I am also grateful to my parents, Leila and Reginaldo, and my family and friends who have been there for me every step of the way. Their support and patience have been instrumental in helping me achieve my goals.

My immense gratitude to Prof. Dr. Paulo Tarso Sanches de Oliveira, who, since our first conversation, has been the best advisor I could wish. I certainly would not accomplish this important work without his example as a leader, researcher, and, mostly, as a human.

During my experience as a master student, I had the luck to work with Prof. Dr. Gabriel Dibbern Sacchi as an assistant at Hydrology classes for the undergraduate student. It was one of the most enriching experiences of my career, and I thank him for the partnership and for sharing teaching experiences and skills.

I am grateful to all the professionals at SHS-USP, including the professors and technical staff, for their support and guidance. Special thanks go to Sá and Rose, who provided attention and support to all PPG students. I also thank to my colleagues from São Carlos, who have been a source of joy and camaraderie, even from far away. Their willingness to share the weight of responsibilities that come with scientific research has made my journey as a master's student more fulfilling.

Last but not least, I thank to my colleagues and friends from HWater, Alexandre, Márcio, Mário and Savina, for their consistent patience and support that allowed me to study and conclude my dissertation.

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Grant Finance Code 001

“Pior que não terminar uma viagem é nunca partir.”

Amyr Klink

ABSTRACT

Marchezepe,, B. K. Streamflow predictions in ungauged basins by using the Grunsky's Method. Master Thesis, São Carlos School of Engineering, Department of Hydraulics and Sanitary Engineering, University of São Paulo, São Carlos, SP, Brazil, 2023.

Predicting runoff in ungauged basins is a significant challenge in Hydrology, and various advanced techniques have been developed to address this issue. However, returning some simplicity to the predictions might be necessary for practical uses. This master dissertation evaluates a generalized of Grunsky's approach and proposes its improvement. This approach was originally developed in the early 1900s for basins in San Francisco Peninsula, USA. First, we focus on the application of the Grunsky generalized method to 716 Brazilian catchments on interannual and monthly scales. The rainfall-runoff relation coefficient (α) is determined, and the method's performance is evaluated locally, catchment by catchment. Additionally, we regionalized and analyzed the catchments into hydrological groups. Then, an improved version of the previously tested method is proposed, incorporating a rainfall-runoff relation coefficient (α) based on mean annual temperature. Through multiple linear regression, the α values are determined using catchment attributes such as the aridity index, annual average temperature, and potential evapotranspiration. The generalized method presented a median percentage bias and Kling-Gupta Efficiency of -1% and 0.73, respectively, with favorable performances observed in certain groups. On a montly scale, more than 83% of the total studied basins had at least one month with KGE greater than 0.50. The performance of the improved method indicates the suitability of this approach for predicting runoff in Brazilian basins, with KGE = 0.899, R^2 = 0.82, and RMSE = 27.4% on the interannual scale. Here, we emphasize the practicality and reliability of the Grunsky approach as a simple and easy-to-use equation for predicting runoff. The findings suggest that this approach can serve as a viable alternative to more complex methods, especially in ungauged or poorly gauged basins. By incorporating both theoretical and empirical elements, this study contribute to the ongoing efforts to develop accessible and effective methods for runoff prediction, furthering our understanding of hydrological processes in Brazilian catchments.

Keywords: Regionalization, hydrological group, temporal scale, tropical basins, mean temperature, aridity index, potential evapotranspiration.

RESUMO

Marchezepe, B. K. Previsão de vazões em bacias não monitoradas usando o Método de Grunsky. Dissertação (Mestrado) – Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2023.

Estimar o escoamento em bacias não monitoradas é um desafio significativo em Hidrologia, e várias técnicas avançadas foram desenvolvidas para abordar esse problema. No entanto, reintroduzir certa simplicidade nas previsões pode ser necessário para usos práticos. Esta dissertação de mestrado avalia uma versão generalizada do Método de Grunsky e propõe seu aprimoramento. Essa versão foi originalmente desenvolvida no início dos anos 1900 para bacias na Península de São Francisco, EUA. Primeiramente, focou-se na aplicação do método generalizado de Grunsky a 716 bacias brasileiras em escalas interanuais e mensais. O coeficiente de relação precipitação-vazão (α) é determinado, e o desempenho do método é avaliado localmente, bacia por bacia. Além disso, as bacias foram regionalizadas e analisadas em grupos hidrológicos. Em seguida, foi proposta uma versão aprimorada do método previamente testado, incorporando um coeficiente de relação precipitação-vazão (α) baseado na temperatura média anual. Por meio de regressão linear múltipla, foram determinados os valores de α usando atributos da bacia, como o índice de aridez, a temperatura média anual e a evapotranspiração potencial. O método generalizado apresentou uma mediana de PBIAS e Eficiência de Kling-Gupta (KGE) de -1% e 0,73, respectivamente, com desempenhos favoráveis observados em certos grupos. Em escala mensal, mais de 83% do total de bacias estudadas apresentaram pelo menos um mês com KGE maior que 0,50. O desempenho do método aprimorado indica a adequação dessa abordagem para estimar o escoamento em bacias brasileiras, com KGE = 0,899, R^2 = 0,82 e RMSE = 27,4% na escala interanual. Enfatiza-se aqui a praticidade e a confiabilidade da abordagem de Grunsky como uma equação simples e fácil de usar para estimar o escoamento. Os resultados sugerem que tal abordagem pode ser uma alternativa viável a métodos mais complexos, especialmente em bacias pouco ou não monitoradas. Ao incorporar elementos teóricos e empíricos, este estudo contribui para os esforços contínuos de desenvolvimento de métodos acessíveis e eficazes para a estimativa do escoamento, aprimorando a compreensão dos processos hidrológicos em bacias brasileiras.

Palavras-chave: Regionalização, grupo hidrológico, escala temporal, bacias tropicais, temperatura média, índice de aridez, evaporação potencial.

LIST OF FIGURES

CHAPTER 1

Figure 1: Maps of the spatial distribution of a) each catchment by area and hydrological group over the territory; b) the six Brazilian Biomes (Amazon, Atlantic Forest, Cerrado, Caatinga, Pampa, and Pantanal); and c) Brazil, in yellow, on the world map. 20

Figure 2: Schematic view of the analysis method. 22

Figure 3: Scatter plot with observed and mean annual runoff versus annual precipitation for the 30 years series from the 716 CABra catchments, and Grunsky's equation plots for multiple α values (mm^{-1}). Colored according to the aridity index value from each catchment. 24

Figure 4: Cumulative Distribution Function (CDF) curves comparing KGE values for each method for determining α value on an interannual scale, for hydrological groups a) Dry, b) Extremely-dry, c) Extremely-wet, d) Non-seasonal, e) Rainforest and f) Savannah. 27

Figure 5: Cumulative Distribution Function (CDF) curves comparing KGE values for each method for determining α value on a monthly scale, for hydrological groups a) Dry, b) Extremely-dry, c) Extremely-wet, d) Non-seasonal, e) Rainforest and f) Savannah. 28

Figure 6: Scatter plot with observed annual runoff versus predicted annual runoff from MMSE α value for each catchment during the 30 years series from the 716 CABra catchments, divided by groups. Colored by the number of neighbors for each point. 30

Figure 7: (a) Map of KGE values and (b) Map of PBIAS values (mm $year^{-1}$) for catchments, comparing observed runoff with predicted runoff with α correspondent to the MMSE α for each catchment. 32

Figure 8: Maps of KGE from MMSE α value for each group for each month, from January (top left) to December (bottom right). 33

Figure 9: Scatter plot with observed annual runoff versus predicted annual runoff from median MMSE α value for each group during the 30 years series from the 716 CABra catchments, divided by groups. Colored by the number of neighbors for each point. 35

Figure 10: Boxplot with KGE from regionalized monthly runoff prediction for each month and group from the 716 CABra catchments. 36

CHAPTER 2

Figure 11: Schematic view of the analysis method. 54

Figure 12: Spearman correlation matrix between alpha and climatic parameters of CABra dataset. 56

Figure 13: Scatterplots of alpha coefficient and a) Mean Temperature (°C), b) Aridity Index, and c) Potential Annual Evapotranspiration by Penman-Monteith equation. 57

Figure 14: Scatterplot of alpha coefficient versus mean monthly temperature (°C) by group. 59

Figure 15: Observed runoff versus predicted runoff on an annual scale. 60

Figure 16: Observed runoff versus predicted runoff on a monthly scale. 61

APPENDIX

Figure S17: Grunsky's equation plot for various values of α (mm^{-1}), and lines with relations between Q (annual runoff) and P (annual precipitation). 75

Figure S18: α coefficient of all 716 evaluated catchments, divided according to their hydrological groups for both (a) interannual and (b) monthly temporal scales. 76

Figure S19: Scatter plot with observed annual runoff versus observed annual precipitation for each year of the 30 years series from the 735 CABra catchments, and Grunsky's equation plots for multiple α values (mm^{-1}), divided by cluster. 77

Figure S20: Boxplot of (a) Mean, (b) Median, and (c) MMSE α values, for the 12 months of the year, from the 716 catchments. 79

Figure S21: Map of α values correspondent to the MSE for each catchment. 80

Figure S22: Boxplots α Mean Monthly Values by hydrological group. 81

Figure S23: Boxplots α Median Monthly α values by hydrological group. 82

Figure S24: Boxplots α MMSE Monthly Values by hydrological group. 83

Figure S25: Maps of MMSE α value for each group for each month, from January (top left) do December (bottom right). 85

Figure S26: Maps of PBIAS from MMSE α value for each group for each month, from January (top left) to December (bottom right). 86

Figure S27: Boxplots with PBIAS from regionalized monthly runoff prediction for each month and group from the 716 CABra catchments. 88

LIST OF TABLES

CHAPTER 1

Table 1: Median of MMSE **α** value (mm^{-1}) of each hydrological group, for an interannual scale. 34

CHAPTER 2

Table 2: Multiple linear regression equation coefficients and performance metrics for each month.. 62

APPENDIX

Table S3: Median alpha values from the regionalization on a monthly scale. 87

TABLE OF CONTENTS

ACKNOWLEDGEMENTS .... **III**
ABSTRACT .... **V**
RESUMO .... **VI**
GENERAL INTRODUCTION **1**
GENERAL OBJECTIVES **10**
CHAPTER 1 STREAMFLOW PREDICTION IN UNGAUGED BRAZILIAN CATCHMENTS BY USING THE GRUNSKY METHOD **11**
1. ABSTRACT 12
2. KEYWORDS 13
3. INTRODUCTION 13
4. MATERIAL AND METHODS 16
4.1. Dataset *16*
4.2. Re-introducing Grunsky's method *17*
4.3. Computing α coefficient for interannual and monthly temporal scales on a local scale *18*
4.4. Regionalization *19*
5. RESULTS AND DISCUSSION 22
5.1. Relations between runoff and precipitation and Grunsky's equation *23*
5.2. Choosing the best method to define a single α for each catchment *24*
5.3. Local Performance of α values *29*
5.4. Regionalization of α values *34*
6. LIMITATIONS AND BROADER IMPLICATIONS 37
7. CONCLUSIONS 38
8. ACKNOWLEDGEMENTS 39
9. REFERENCES 40
CHAPTER 2 IMPROVED EMPIRICAL GRUNSKY METHOD FOR STREAMFLOW PREDICTION IN UNGAUGED BRAZILIAN CATCHMENTS **46**
1. ABSTRACT 47
2. KEYWORDS 47
3. INTRODUCTION 47
4. MATERIAL AND METHODS 50
4.1. Dataset *50*
4.2. Relashionship between α and catchments attributes *52*
4.3. Calculating α from a Multiple Linear Regression Method *53*
4.4. Performance Evaluation of streamflow prediction obtained by using the proposed approach *54*
5. RESULTS AND DISCUSSION 55
5.1. Relashionship between α and catchments characteristics *55*
5.2. Calculating α from a Multiple Linear Regression Method *59*
6. LIMITATIONS AND BROADER IMPLICATIONS 62
7. CONCLUSIONS 63
8. ACKNOWLEDGEMENTS 64
9. REFERENCES 64
GENERAL CONCLUSIONS **70**
APPENDIX **72**
1. CONTENTS OF THIS FILE 72
2. GRUNSKY METHOD PLOT 75
3. BOXPLOTS OF A COEFFICIENT BY GROUP 76
4. GRUNSKY CURVE FOR HYDROLOGICAL GROUPS ON ANNUAL SCALE 77
5. BOXPLOTS OF A COEFFICIENT BY MONTH 79
6. INTERANNUAL SCALE ANALYSIS 80
7. MONTHLY SCALE ANALYSIS 81
8. SPATIAL DISTRIBUTION OF LOCAL ANALYSIS ON A MONTHLY SCALE 85
9. REGIONALIZATION RESULTS FOR THE MONTHY SCALE 87

GENERAL INTRODUCTION

From the ancient philosophers to the 17th Century, Science has been taken by classical and theoretical thinking about nature and its processes. Over the last centuries, with the advance of the scientific method, research passed through a great development, using empirical and rational approaches. Hydrology was not different, going from a qualitative to a quantitative perspective, turning into a branch of natural science (Nace, 1974; Hubbart, 2011).

Three renowned pioneers in the field were the astronomer Edmund Halley (1656-1742), who made experiments with water storage and evaporation, Pierre Perrault (1608-1680), believed to be the first to make an experiment with rainfall-runoff relation, and the physicist Edme Marriotte (1620-1684), that did similar experiments as Perrault, confirming his results (Eview and Linsley, 1967; Hubbart, 2011). These first surveys led to an estimation of a value of runoff about 1/6 of the precipitation in the Seine river basin, France (Eview and Linsley, 1967).

Across the following centuries, along with the development of hydraulic structures and increase of climate measurements, rainfall-runoff relations were estimated and summarized into empirical formulas, mentioning the works of Justin (1914) and, more recently, Vogel, Wilson and Daly (1999), that used parameters such as temperature, drainage area, catchment slope, etc. (Eview and Linsley, 1967; Santos and Hawkins, 2011). One of these studies is the relation found by Grunsky (1908, 1915), which shows a simple method to directly find average annual runoff ($\bar{Q}$) from average annual precipitation ($\bar{P}$) in the São Francisco peninsula, California. The author developed this practical relation from hydrological observations during the 1900s decade. Despite of the simplicity of Grunsky's proposed framework, it achieved an excellent performance in describing the long-term hydrological fluxes.

The empirical methods entered in disused since the 1930's, with more sophisticated methods developed, isolating the water cycle components, and including other parameters to the analysis with the use of statistics, mainly from the 1950's, when computational development was put into use to help hydrologists with greater datasets and time series (Eview and Linsley, 1967; Sivapalan *et al.*, 2003).

In function of the increase of computational power over the last decades, prediction models became more accessible and widely applied by scientists across the globe. Some recent methods in the field can be mentioned, including machine learning techniques, which focus on models based solely on data to predict hydrological processes (Samantaray *et al.*, 2022), satellite-based approaches, using interpolation and clustering analysis (De Souza Fraga *et al.*, 2022), and artificial neural networks, which show promising results for hourly daily and storm events predictions (Besaw *et al.*, 2010).

Predictions are generally focused on searching the best fitting model to data, and may sometimes put aside the search for understanding the hydrological processes (Sivapalan *et al.*, 2003; Hrachowitz *et al.*, 2013). Methods of machine learning, for example, when applied to predictions in ungauged basins, result into predictions that might even outperform process-driven prediction methods (Kratzert *et al.*, 2019; Singh *et al.*, 2022), and operate by analyzing only previous data itself to predictions, not taking into account any other theoretical or physical model.

Despite the development of sophisticated methods of predictions in ungauged basins, uncertainties still remain in a great part due to heterogeneity among catchments, by vegetation, climate, human interference and non-stationary conditions, and need for understanding hydrological processes in these different environments and scales (Hrachowitz *et al.*, 2013; Blöschl *et al.*, 2019; Samantaray *et al.*, 2022). Along with all difficulties that exist in predictions in ungauged basins, even with all the cited methods, one question arises: is there any reason to use simple, empirical methods in an era of advanced techniques, with machine learning and complex climate models with dozens, or even hundreds parameters?

Returning the focus to Grunsky's empirical method, the original framework remained untouched until the development of a recent study that employed a generalized formula specifically designed for Mediterranean-type climate basins in California, Portugal, and France (Santos and Hawkins, 2011). This study produced promising outcomes, including a good performance ($r^2 = 0.89$) in long-term streamflow prediction of Portugal watersheds, and a robust correlation observed between the generalized formula and mean annual temperature. The above-mentioned method involves a straightforward approach to calculate the runoff in a catchment area.

In order to complement modern prediction methods with didactic and easy-to-use models, based in theoretical and empirical methods, in **chapter 1**, we discuss the validity of the Grunsky method for Brazilian catchments, proving its applicability even in non-Mediterranean climate, which was not experimented since this method was first proposed in the early 1900's.

In **chapter 2**, we analyzed the method's accuracy and precision by adding other parameters to the method in order to improve its performance. The main parameters for this analysis are: the aridity index, potential evapotranspiration and mean temperature. The latter was already proven to have high correspondence to precipitation and streamflow, and even proposed by Santos and Hawkins (2011).

References

Agarwal, A. *et al.* (2016) 'Hydrologic regionalization using wavelet-based multiscale entropy method', *Journal of Hydrology*, 538, pp. 22–32. doi: 10.1016/J.JHYDROL.2016.03.023.

Almagro, A. *et al.* (2020) 'CABra: a novel large-sample dataset for Brazilian catchments', *Hydrology and Earth System Sciences Discussions*, (October), pp. 1–40. doi: 10.5194/hess-2020-521.

Almagro, A. (2021) *Towards a better understanding of catchment hydrology in Brazil*. Federal University of Mato Grosso do Sul. Available at: https://repositorih.ufms.br/handle/123456789/4035.

ANA (2007) *Diagnóstico da outorga de direito de uso de recursos hídricos no Brasil. Caderno de Recursos Hídricos. Volume 4. Agência Nacional de Águas.* Brasília. Available at: www.ana.gov.br.

Arheimer, B. *et al.* (2020) 'Global catchment modelling using World-Wide HYPE (WWH), open data, and stepwise parameter estimation', *Hydrology and Earth System Sciences*, 24(2), pp. 535–559. doi: 10.5194/HESS-24-535-2020.

Arsenault, R. *et al.* (2023) 'Continuous streamflow prediction in ungauged basins: long short-term memory neural networks clearly outperform traditional hydrological models', *Hydrol. Earth Syst. Sci*, 27, pp. 139–157. doi: 10.5194/hess-27-139-2023.

Athira, P. *et al.* (2016) 'Predictions in ungauged basins: an approach for regionalization of hydrological models considering the probability distribution of model parameters', *Stochastic Environmental Research and Risk Assessment*, 30(4), pp. 1131–1149. doi: 10.1007/S00477-015-1190-6.

Ballarin, A. S. *et al.* (2022) 'The Impact of an Open Water Balance Assumption on Understanding the Factors Controlling the Long-Term Streamflow Components', *Water Resources Research*, 58(10). doi: 10.1029/2022WR032413.

Beck, H. E., de Roo, A. and van Dijk, A. I. J. M. (2015) 'Global Maps of Streamflow Characteristics Based on Observations from Several Thousand Catchments', *Journal of Hydrometeorology*, 16(4), pp. 1478–1501. doi: 10.1175/JHM-D-14-0155.1.

Besaw, L. E. *et al.* (2010) 'Advances in ungauged streamflow prediction using artificial neural networks', *Journal of Hydrology*, 386(1–4), pp. 27–37. doi: 10.1016/J.JHYDROL.2010.02.037.

Beskow, S. *et al.* (2016) 'Potential of the LASH model for water resources management in data-scarce basins: a case study of the Fragata River basin, southern Brazil', *https://doi.org/10.1080/02626667.2015.1133912*, 61(14), pp. 2567–2578. doi: 10.1080/02626667.2015.1133912.

Bloschl, G. *et al.* (eds) (2013) *Runoff Prediction in Ungauged Basins*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9781139235761.

Blöschl, G. *et al.* (2019) 'Twenty-three unsolved problems in hydrology (UPH) – a community perspective', *https://doi.org/10.1080/02626667.2019.1620507*, 64(10), pp. 1141–1158. doi: 10.1080/02626667.2019.1620507.

Bork, C. K. *et al.* (2021) 'Minimum streamflow regionalization in a Brazilian watershed under different clustering approaches', *Anais da Academia Brasileira de Ciências*, 93(suppl 4). doi: 10.1590/0001-3765202120210538.

Budyko, M. I. (1974) *Climate and Life, New York, NY: Academic Press.*

Choi, J. *et al.* (2021) 'Usefulness of Global Root Zone Soil Moisture Product for Streamflow Prediction of Ungauged Basins', *Remote Sensing 2021, Vol. 13, Page 756*, 13(4), p. 756. doi: 10.3390/RS13040756.

Dal Molin, M. *et al.* (2020) 'Understanding dominant controls on streamflow spatial variability to set up a semi-distributed hydrological model: The case study of the Thur catchment', *Hydrology and Earth System Sciences*, 24(3), pp. 1319–1345. doi: 10.5194/HESS-24-1319-2020.

Du, T. L. T. *et al.* (2020) 'Streamflow prediction in "geopolitically ungauged" basins using satellite observations and regionalization at subcontinental scale', *Journal of Hydrology*, 588, p. 125016. doi: 10.1016/J.JHYDROL.2020.125016.

Eview, R. and Linsley, A. K. (1967) 'The relation between rainfall and runoff: Review paper', *Journal of Hydrology*, 5, pp. 297–311. doi: 10.1016/S0022-1694(67)80128-8.

de Figueiredo, J. V. *et al.* (2016) 'Runoff initiation in a preserved semiarid Caatinga small watershed, Northeastern Brazil', *Hydrological Processes*, 30(13), pp. 2390–2400. doi: 10.1002/HYP.10801.

Gao, Y. *et al.* (2021) 'Diagnosis toward predicting mean annual runoff in ungauged basins', *Hydrol. Earth Syst. Sci*, 25, pp. 945–956. doi: 10.5194/hess-25-945-2021.

Grunsky, C. E. (1908a) 'Closure to Rain and Run-Off Near San Francisco', *Transactions of the American Society of Civil Engineers*, 61(2), pp. 539–543. doi: 10.1061/TACEAT.0002035.

Grunsky, C. E. (1908b) 'Rain and Run-Off Near San Francisco, California', *Transactions of the American Society of Civil Engineers*, 61(2), pp. 496–515. doi: 10.1061/TACEAT.0002036.

Grunsky, C. E. *et al.* (1915) 'Discussion: Run-Off from Rainfall and Other Data', *Transactions of the American Society of Civil Engineers*, 79(2), pp. 1156–1208. doi: 10.1061/TACEAT.0002701.

Grunsky, C. E. and Manson, M. C. E. (1908) *Reports on the water supply of San Francisco, California, 1900 to 1908, inclusive.* Transactions of the American Society of Civil Engineers. Available at: http://hdl.handle.net/2027/uc1.$b78665 (Accessed: 13 March 2022).

Gupta, H. V. *et al.* (2009) 'Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling', *Journal of Hydrology*, 377(1–2), pp. 80–91. doi: 10.1016/J.JHYDROL.2009.08.003.

Hersbach, H. *et al.* (2020a) 'The ERA5 global reanalysis', *Quarterly Journal of the Royal Meteorological Society*, 146(730), pp. 1999–2049. doi: 10.1002/qj.3803.

Hersbach, H. *et al.* (2020b) 'The ERA5 global reanalysis', *Quarterly Journal of the Royal Meteorological Society*, 146(730), pp. 1999–2049. doi: 10.1002/QJ.3803.

Honório, M. D. S. *et al.* (2020) 'Comparison of two methods for determining Q95reference flow in the mouth of the surface catchment basin of the Meia Ponte river, state of Goiás, Brazil', *Open Engineering*, 10(1), pp. 469–476. doi: 10.1515/ENG-2020-0053/MACHINEREADABLECITATION/RIS.

Hrachowitz, M. *et al.* (2013) 'A decade of Predictions in Ungauged Basins (PUB)-a review', *Hydrological Sciences Journal*. Taylor & Francis, pp. 1198–1255. doi: 10.1080/02626667.2013.803183.

Hubbart, J. A. (2011) 'Origins of Quantitative Hydrology: Pierre Perrault, Edme Mariotte, and Edmond Halley', *Journal of the American Water Resources Association*, 13(6), pp. 15–17.

Justin (1914) 'Derivation of Run-Off from Rainfall Data', *Transactions of the American Society of Civil Engineers*, 77(1). Available at: https://doi.org/10.1061/taceat.0002572.

Knoben, W. J. M., Freer, J. E. and Woods, R. A. (2019) 'Technical note: Inherent benchmark or not? Comparing Nash-Sutcliffe and Kling-Gupta efficiency scores', *Hydrology and Earth System Sciences*. doi: 10.5194/hess-2019-327.

Kratzert, F. *et al.* (2019) 'Toward Improved Predictions in Ungauged Basins: Exploiting the Power of Machine Learning', *Water Resources Research*, 55(12), pp. 11344–11354. doi: 10.1029/2019WR026065.

Liu, Y. *et al.* (2020) 'What is the hydrologically effective area of a catchment?', *Environmental Research Letters*, 15(10). doi: 10.1088/1748-9326/aba7e5.

Marchezepe, B. K. *et al.* (2023) 'Streamflow prediction in ungauged Brazilian catchments by using the Grunsky Method', *[Manuscript submitted for publication]. School of Engineering of São Carlos, University of São Paulo.*

McDonnell, J. J. (2003) 'Where does water go when it rains? Moving beyond the variable source area concept of rainfall-runoff response', *Hydrological Processes*, 17(9), pp. 1869–1875. doi: 10.1002/HYP.5132.

Meira Neto, A. A. *et al.* (2020) 'An Aridity Index-Based Formulation of Streamflow Components', *Water Resources Research*, 56(9), pp. 1–14. doi: 10.1029/2020WR027123.

Montanari, A. *et al.* (2013) '"Panta Rhei—Everything Flows": Change in hydrology and society—The IAHS Scientific Decade 2013–2022', *https://doi.org/10.1080/02626667.2013.809088*, 58(6), pp. 1256–1275. doi: 10.1080/02626667.2013.809088.

Morice, C. P. *et al.* (2012) 'Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The HadCRUT4 data set', *J. Geophys. Res*, 117, p. 8101. doi: 10.1029/2011JD017187.

Mosavi, A., Ozturk, P. and Chau, K. W. (2018) 'Flood Prediction Using Machine

Learning Models: Literature Review', *Water 2018, Vol. 10, Page 1536*, 10(11), p. 1536. doi: 10.3390/W10111536.

Nace, R. L. (1974) 'Pierre Perrault: the Man and His Contribution To Modern Hydrology', *JAWRA Journal of the American Water Resources Association*, 10(4), pp. 633–647. doi: 10.1111/j.1752-1688.1974.tb05623.x.

Parajka, J. *et al.* (2013) 'Comparative assessment of predictions in ungauged basins-Part 1: Runoff-hydrograph studies', *Hydrology and Earth System Sciences*, 17(5), pp. 1783–1795. doi: 10.5194/HESS-17-1783-2013.

Pinheiro, E. A. R. *et al.* (2016) 'Importance of soil-water to the Caatinga biome, Brazil', *Ecohydrology*. doi: 10.1002/eco.1728.

Piol, M. V. A. *et al.* (2019) 'Performance evaluation of flow duration curves regionalization methods', *Brazilian Journal of Water Resources*. doi: 10.1590/2318-0331.241920170202.

Pomeroy, J. W. *et al.* (2007) 'The cold regions hydrological model: A platform for basing process representation and model structure on physical evidence', *Hydrological Processes*, 21(19), pp. 2650–2667. doi: 10.1002/HYP.6787.

Pool, S., Vis, M. and Seibert, J. (2021) 'Regionalization for Ungauged Catchments — Lessons Learned From a Comparative Large-Sample Study', *Water Resources Research*, 57(10), p. e2021WR030437. doi: 10.1029/2021WR030437.

Popper, K. (2002) *The Logic of Scientific Discovery*. Available at: http://books.google.com/books?hl=en&lr=&id=T76Zd20IYlgC&pgis=1.

Razavi, T., Coulibaly, P. and Asce, M. (2012) 'Streamflow Prediction in Ungauged Basins: Review of Regionalization Methods', *Journal of Hydrologic Engineering*, 18(8), pp. 958–975. doi: 10.1061/(ASCE)HE.1943-5584.0000690.

Salinas, J. L. *et al.* (2013) 'Comparative assessment of predictions in ungauged basins-Part 2: Flood and low flow studies', *Hydrology and Earth System Sciences*, 17(7), pp. 2637–2652. doi: 10.5194/hess-17-2637-2013.

Samantaray, S. *et al.* (2022) 'Monthly runoff prediction at Baitarani river basin by support vector machine based on Salp swarm algorithm', *Ain Shams Engineering Journal*, 13(5), p. 101732. doi: 10.1016/J.ASEJ.2022.101732.

Santos, F. L. and Hawkins, R. H. (2011) 'Generalized Mediterranean Annual Water Yield Model: Grunsky's Equation and Long-Term Average Temperature', *Journal of Hydrologic Engineering*, 16(11), pp. 874–879. doi: 10.1061/(ASCE)HE.1943-5584.0000402.

Seo, Y., Kim, S. and Singh, V. P. (2018) 'Machine Learning Models Coupled with Variational Mode Decomposition: A New Approach for Modeling Daily Rainfall-Runoff', *Atmosphere 2018, Vol. 9, Page 251*, 9(7), p. 251. doi: 10.3390/ATMOS9070251.

Shalamu, A. (2009) 'Monthly and seasonal streamflow forecasting in the Rio Grande

Basin', *2009*, (August), p. 289.

Silva, W. M. and Simões, S. J. C. (2014) 'Spatial Intra-Annual Variability of Precipitation Based on Geostatistics. A Case Study for the Paraıba Do Sul Basin, Southeastern Brazil', *International Journal of Geosciences*, 5, p. 408. doi: 10.4236/ijg.2014.54039.

Singh, L. *et al.* (2022) 'Streamflow regionalisation of an ungauged catchment with machine learning approaches', *Hydrological Sciences Journal*. doi: 10.1080/02626667.2022.2049271.

Sivapalan, M. *et al.* (2003) 'IAHS Decade on Predictions in Ungauged Basins (PUB), 2003-2012: Shaping an exciting future for the hydrological sciences', *Hydrological Sciences Journal*, 48(6), pp. 857–880. doi: 10.1623/hysj.48.6.857.51421.

De Souza Fraga, M. *et al.* (2022) 'Assessment of streamflow regionalization using interpolated and satellite-based precipitation: a case study in a tropical watershed at Brazil', *Modeling Earth Systems and Environment 2022*, pp. 1–14. doi: 10.1007/S40808-022-01383-W.

Spinoni, J. *et al.* (2015) 'Towards identifying areas at climatological risk of desertification using the Köppen-Geiger classification and FAO aridity index', *International Journal of Climatology*, 35(9), pp. 2210–2222. doi: 10.1002/JOC.4124.

VanderKwaak, J. E. and Loague, K. (2001) 'Hydrologic-Response simulations for the R-5 catchment with a comprehensive physics-based model', *Water Resources Research*, 37(4), pp. 999–1013. Available at: https://onlinelibrary.wiley.com/doi/full/10.1029/2000WR900272.

Viglione, A. *et al.* (2013) 'Comparative assessment of predictions in ungauged basins - Part 3: Runoff signatures in Austria', *Hydrology and Earth System Sciences*, 17(6), pp. 2263–2279. doi: 10.5194/HESS-17-2263-2013.

Vogel, R. M., Wilson, I. and Daly, C. (1999) 'Regional regression models of annual streamflow for the United States', *Journal of Irrigation and Drainage Engineering*, 125(3). doi: https://doi.org/10.1061/(ASCE)0733-9437(1999)125:3(148).

Xavier, A. C., King, C. W. and Scanlon, B. R. (2016) 'Daily gridded meteorological variables in Brazil (1980–2013)', *International Journal of Climatology*, 36(6), pp. 2644–2659. doi: 10.1002/JOC.4518.

Yang, X. *et al.* (2020) 'Dependence of regionalization methods on the complexity of hydrological models in multiple climatic regions', *Journal of Hydrology*, 582, p. 124357. doi: 10.1016/J.JHYDROL.2019.124357.

Yaseen, Z. M. *et al.* (2016) 'Stream-flow forecasting using extreme learning machines: A case study in a semi-arid region in Iraq', *Journal of Hydrology*, 542, pp. 603–614. doi: 10.1016/j.jhydrol.2016.09.035.

Zhang, K. *et al.* (2023) 'Streamflow Prediction in Poorly Gauged Watersheds in the United States Through Data-Driven Sparse Sensing', *Water Resources Research*, 59(4), p. e2022WR034092. doi: 10.1029/2022WR034092.

Zhang, L. *et al.* (2008) 'Water balance modeling over variable time scales based on the Budyko framework – Model development and testing', *Journal of Hydrology*, 360(1–4), pp. 117–131. doi: 10.1016/J.JHYDROL.2008.07.021.

Zhang, R. *et al.* (2018) 'Season-based rainfall–runoff modelling using the probability-distributed model (PDM) for large basins in southeastern Brazil', *Hydrological Processes*, 32(14), pp. 2217–2230. doi: 10.1002/HYP.13154.

Zotarelli, L. *et al.* (2014) 'Step by Step Calculation of the Penman-Monteith Evapotranspiration (FAO-56 Method). AE459', *Institute of Food and Agricultural Sciences. University of Florida*, pp. 1–14. Available at: http%22//edis.ifas.ufl.edu/ae459.

GENERAL OBJECTIVES

The main objective of this project is to evaluate the performance of the Grunsky Method to characterize rainfall-runoff relationships in Brazilian catchments.

Specific Objectives

In more detail, the present study aims to:

- Assess the performance of a simple and fast way to compute catchment runoff on different time and spatial scales through its precipitation values, known as Grunsky Method;
- Apply the proposed framework for the largest number of Brazilian catchments from a reliable database, facilitating its application by decision-makers and replication for future updates;
- Evaluate the role of climatic factors on the response of runoff to precipitation in Brazilian catchments;
- Analyze methods for increasing Grunsky Method efficiency.

CHAPTER 1

Streamflow prediction in ungauged Brazilian catchments by using the Grunsky Method

Marchezepe, B. K., Almagro A., Ballarin A. S. Oliveira, P. T. S. Streamflow prediction in ungauged Brazilian catchments by using the Grunsky Method [Manuscript submitted for publication]. School of Engineering of São Carlos, University of São Paulo.

1. Abstract

Establish a reliable rainfall-runoff relation capable of predicting runoff in ungauged basins is a matter of interest across the world for a long time and has been taking importance during the past decades. Machine learning methods, regionalization, software, and hydrological models are used to estimate runoff, with reasonable precision. Although, returning some simplicity to the predictions might be necessary for practical uses. In this paper, we re-introduce C. E. Grunsky's approach, developed in the early 1900s to predict runoff from values of precipitation on a 2 equations system for San Francisco's peninsula. Here, we analyze the Grunsky generalized method applied for 716 Brazilian catchments, on an interannual and monthly scale. First, we established the best method to find the rainfall-runoff relation coefficient for each catchment. Then, we evaluate the performance of the method on a local scale, i.e., catchment by catchment. Lastly, we analyze the method of regionalization, by dividing the catchments into 6 hydrological groups. For local scale, the Kling-Gupta Efficiency (KGE) values range from 0.87 to 0.93 on an interannual scale and greater than 0.50 on a monthly scale. For the regionalized scale, KGE varies from 0.60 to 0.84 on an interannual scale. We also found suitable KGE values on a montly scale, with more than 22% of catchments with KGE higher than 0.50, being the best performances in Non-seasonal and Extremely-wet basin groups, and the worst performance in Dry group. Our findings indicate that the Grunsky's approach is suitable to predict runoff for Brazilian basins on

interannual and monthly scales. This simple and easy-to-use equation presents a reliable alternative to more complex methods to compute runoff from only using rainfall data.

2. Keywords

Regionalization, hydrological group, temporal scale, tropical basins.

3. Introduction

The understanding of the nature of temporal and spatial variability of hydrological fluxes, especially runoff, is crucial for hydrological studies, such as for the design of hydraulic structures and water resources management (Viglione *et al.*, 2013; Beck, de Roo and van Dijk, 2015; Dal Molin *et al.*, 2020). Despite its importance, the correct understanding of catchments' hydrological behavior is still a challenge, given that most basins across the world are ungauged, which precludes a precise temporal and spatial characterization. Due to this lack of data, the study of the prediction parameters in those basins has received a great deal of attention in the last decades (Razavi, Coulibaly and Asce, 2012; Bloschl *et al.*, 2013; Salinas *et al.*, 2013).

Recently, an initiative of the International Association of Hydrological Sciences called the decades of Predictions of Ungauged Basins (PUB) 2003-2012 and 2013-2022 (Sivapalan *et al.*, 2003; Hrachowitz *et al.*, 2013; Montanari *et al.*, 2013), attempted to merge the knowledge of the field in terms of the core questions, the benefits of models for the catchments study and awareness of uncertainties for more flexible approaches (Hrachowitz *et al.*, 2013). Nevertheless, these initiatives did not unify the modeling strategies for prediction in ungauged basins, and, therefore, different approaches have been proposed and applied since then (Bloschl *et al.*, 2013; Yang *et al.*, 2020; Pool, Vis and Seibert, 2021).

In general, good performance on predictions can be reached for most catchments with regionalization approaches, with the major challenge remaining in the choice of the best method and finding gauged catchments that are good characteristics and parameters “donors” for the ungauged catchments (Pool, Vis and Seibert, 2021). Approximately 10 years before the PUB Decade, during the 1990s, some of these prediction methods showed a complexity that might not be a practical solution for some applications, with the use of parameters difficult to understand and reproduce in other studies, focusing much more in the model than in the real-world processes (Hrachowitz *et al.*, 2013). One of the main challenges recognized in PUB is to find generalizable insights that would bring models as learning tools for hydrological processes (Popper, 2002). Fortunately, this mentality has changed in the last years with the search for more flexible models, which indeed may lead to better knowledge about these processes and improve prediction itself (McDonnell, 2003; Pomeroy *et al.*, 2007).

At the beginning of the 20th century, in a scenario of the early forming of Hydrology as a science and lack of computational performance, the former president of the American Society of Civil Engineers (ASCE), Carl E. Grunsky (1908, 1915), developed somewhat of a “rule-of-thumb” (Santos and Hawkins, 2011), directly relating average annual runoff ($\bar{Q}$) to average annual precipitation ($\bar{P}$) in the São Francisco peninsula, California. The author developed this practical relation from hydrological observations during the 1900s decade. Despite of the simplicity of Grunsky’s proposed framework, it achieved an excellent performance in describing the long-term hydrological fluxes.

The method remained untouched, and empirical approaches were substituted by more sophisticated methods with the increase in physical understanding of water cycle components and computers’ processing power. More than a century after Grunsky’s study with restricted information at his disposal, Santos and Hawkins (2011) generalized the method, inserting a single coefficient, called α, for studying Mediterranean catchments in California, Southern France, and Portugal, with areas ranging from 1.47 to 97,667.00 km^2. They reported that the proposed approach reached a coefficient of determination of 0.89 for observed and predicted annual runoff in Portugal watersheds, and almost 1:1 relation for the France watershed, which shows an exceptional perspective for the method, given its simplicity. Furthermore, they have shown that the α coefficient led to the possibility of application in catchments with different responses to precipitation. However, despite the possibilities, this approach has only embraced long-term data, i.e. the mean values of precipitation and runoff. In addition, the application of this approach was limited to the Mediterranean climate and has not yet been investigated using a large-sample catchment dataset. Moreover, to the best of our knowledge, no study evaluated the performance of the Grunsky method to characterize streamflow fluxes at finer temporal scales.

In this study, we investigate the performance of the Grunsky framework in computing monthly and interannual streamflow for runoff prediction in ungauged tropical basins. To this end, we applied for the first time the generalized Grunsky method for 716 Brazilian catchments available in the Catchments Attributes for Brazil (CABra) dataset (Almagro *et al.*, 2020). Then, we proposed a streamflow regionalization based on the median α computed for each of the six Brazilian hydrological groups. Our investigation shows the suitability to use the Grunsky approach to estimate streamflow in ungauged basins in different climate conditions. With this study, we seek to answer the following questions: Up to what temporal scales the method can be used? Is the Grunsky framework a suitable approach to characterize and predict hydrological fluxes at finer temporal scales than long-term scales? Can this approach be regionalized to predict streamflow in Brazilian basins? We organized this study as follows: in section 4, we describe the dataset used in the study, the methods used in the study as well as the approach to employ Grunsky's method on an interannual and monthly scale. In section 5, we present and discuss the results, respectively and, in section 7, we highlight the main conclusions of the paper.

4. Material and methods

4.1. Dataset

The study comprehended the use of the Catchments Attributes for Brazil (CABra) dataset (Almagro *et al.*, 2020), which contains data from 735 catchments spread over the Brazilian territory. The attributes are grouped into 8 classes, namely: topography, climate, streamflow, groundwater, soil, geology, land-use and land-cover, and hydrologic disturbance. In addition, for each catchment, there is a 30-year daily series of precipitation, streamflow, and actual and potential evapotranspiration, from 1980 to 2010, all of them containing less than 10% of missing data.

We used daily time series of precipitation (P) and streamflow (Q) for the climatological period between October 1st, 1980, to September 30th, 2010 as the main attributes for our analysis, comprehending 30 hydrological years. P is derived from an ensemble mean between a high-resolution ground-based reference dataset (Xavier, King and Scanlon, 2016) and the ERA5 reanalysis dataset (Hersbach *et al.*, 2020a) while Q is based on streamflow gauge observations over the Brazilian catchments. These daily values were resampled to create annualized (monthly) series. Therefore, for each catchment, 30 (360) values of annual (monthly) runoff and precipitation were computed. We excluded from our analysis years (months) with at least one day with missing value to maintain the annual and monthly values comparable within the same catchment. In other words, we only consider years (months) with complete daily records.

We also excluded 18 catchments that presented at least one year with Observed Annual Runoff (Q_{obs}) higher than Observed Annual Precipitation (P_{obs}), and 1 catchment that exhibited inconsistency between precipitation and evapotranspiration data, remaining 716 catchments for the analysis.

4.2. Re-introducing Grunsky's method

Carl E. Grunsky, as a city engineer together with Marsden Manson, published a report on water supply sources in San Francisco, California. The study occurred in 1908 and contains data on the streams in the San Francisco Peninsula and a discussion on the city's water supply situation and its recommended improvement (Grunsky and Manson, 1908). Later, in 1915, Grunsky studied the relations of rainfall-runoff in the area, which resulted in a practical empirical Equation 1 relating the mean annual runoff ($\bar{Q}$), and the mean annual precipitation ($\bar{P}$), measured in inches.

$$\begin{aligned} \bar{Q} &= \bar{P}^2/100, \quad \bar{P} \leq 50 \\ \bar{Q} &= \bar{P} - 25, \quad \bar{P} \geq 50 \end{aligned} \tag{1}$$

Through Grunsky's analysis (1908, 1915), a generalized method to correlate the annual runoff and the annual precipitation resulted in Equation 2 (Santos and Hawkins, 2011), where α is a coefficient (mm^{-1}) used for this generalization. The generalized equation was applied to California, the South of France, and Portugal, characterized by the Mediterranean climate, with dry summer and wet winter (Santos and Hawkins, 2011). The good fit for Mediterranean catchments arises the interest in the study of other types of climates. The rainfall-runoff relationships for varying values of α are presented on Supplementary Material (Figure S1). As a matter of interest, according to the generalized Grunsky method (Equation 2), a negative α value is only possible when runoff is higher than precipitation, i.e., when runoff is receives great influence from groundwater. On the contrary, when runoff is higher than precipitation, the method does not lead exclusively to a negative α value. Since catchments with $Q_{obs} > P_{obs}$ were excluded, there are no negative α coefficients in the analysis.

$$\begin{aligned} \bar{Q} &= \alpha \bar{P}^2, \quad \bar{P} < 1/(2\alpha) \\ \bar{Q} &= \bar{P} - \frac{1}{4\alpha}, \quad \bar{P} > 1/(2\alpha) \end{aligned} \tag{2}$$

4.3. Computing α coefficient for interannual and monthly temporal scales on a local scale

According to the Grunsky method, the α coefficient is calculated from $\bar{Q}$ and $\bar{P}$. As the generalized formula consists of a system of two equations, two values of α are calculated, and the chosen α would be the one that satisfies the system. Nevertheless, the work applied the proposed framework considering a long-term temporal scale, obtaining, therefore, a single α value. We, in contrast, evaluated the performance of the Grunsky method on two other temporal scales: interannual and monthly. However, in this case, we obtained an α value for each evaluated temporal unit (month or year). Therefore, to keep the consistency with the original framework, for each evaluated temporal scale (interannual and monthly), we chose a single α coefficient that best represents the streamflow in each catchment by maintaining the generalized equation but using values of runoff and precipitation different from the long-term $\bar{Q}$ and $\bar{P}$.

We evaluate three different approaches to obtain this unique α coefficient on a local scale, i.e., one α for each catchment: (i) the mean α among all 30-year series of α values; (ii) the median α among all 30-year series of α values; and the α that presented the Minimum Mean Squared Error (MMSE) between predicted annual runoff (Q_{pred}) and observed annual runoff (Q_{obs}). For the last one, to have more values to make the predictions, and since the most important is to obtain the MMSE value, we used α values from the two sides of Equation 2. To evaluate the best method to calculate α (mean, median, and MMSE), we compared their KGE's Empirical Cumulative Distribution Function (ECDF) curves for both evaluated temporal scales. The KGE was computed considering the runoff estimated through one of the three methods to define α and the observed runoff.

To evaluate the prediction performance of the Grunsky method, we applied the equations considering the selected α coefficient (using one of the 3 proposed approaches) from each catchment and the observed annual precipitation and compared estimated and observed streamflow values for both interannual and monthly temporal scales. To evaluate the model performance, we used four widely used statistical metrics: the coefficient of determination (R^2), the Kling-Gupta Efficiency (KGE) (Gupta *et al.*, 2009), the Percent Bias (PBIAS), and the Root Mean Square Error (RMSE). As a matter of interest, R^2 and KGE values close to 1 represent better model performance, positive (negative) PBIAS indicates overestimation (underestimation) in predictions, and lower RMSE values indicate reduced errors.

4.4. Regionalization

Calculating the α coefficients for each one of the 716 gauged basins would not be the final goal of this study, but a first step for the analysis since we seek to predict streamflow in ungauged basins. After the computation of α, we look for the best classification method to cluster the catchments according to their characteristics. From this cluster classification, it would be possible to define an α coefficient for each group, which, in turn, can be used to predict streamflow fluxes at ungauged basins belonging to the respective group. Prediction in Ungauged Basins passes by a reliable regionalization approach considering multi-criteria analysis (Pool, Vis and Seibert, 2021).

To investigate the performance of the regionalization of the α coefficient across Brazil, we used the six hydrological groups proposed by Almagro (2021). To define these six hydrological groups, the author used hydrological signatures and attributes of catchments from the CABra dataset and then employed a principal component analysis and a random forest algorithm. The catchments were grouped based on their hydrological similarities and the dominant processes and their driving attributes were investigated. The hydrological groups are: "Non-seasonal", "Dry", "Rainforest", "Savannah", "Extremely-dry", and "Extremely-wet" catchments (Figure 1). We chose to use these six hydrological groups for the regionalization framework because they are able to better represent the hydrological similarities than other geographical, climate, or ecosystem groups (Almagro, 2021).

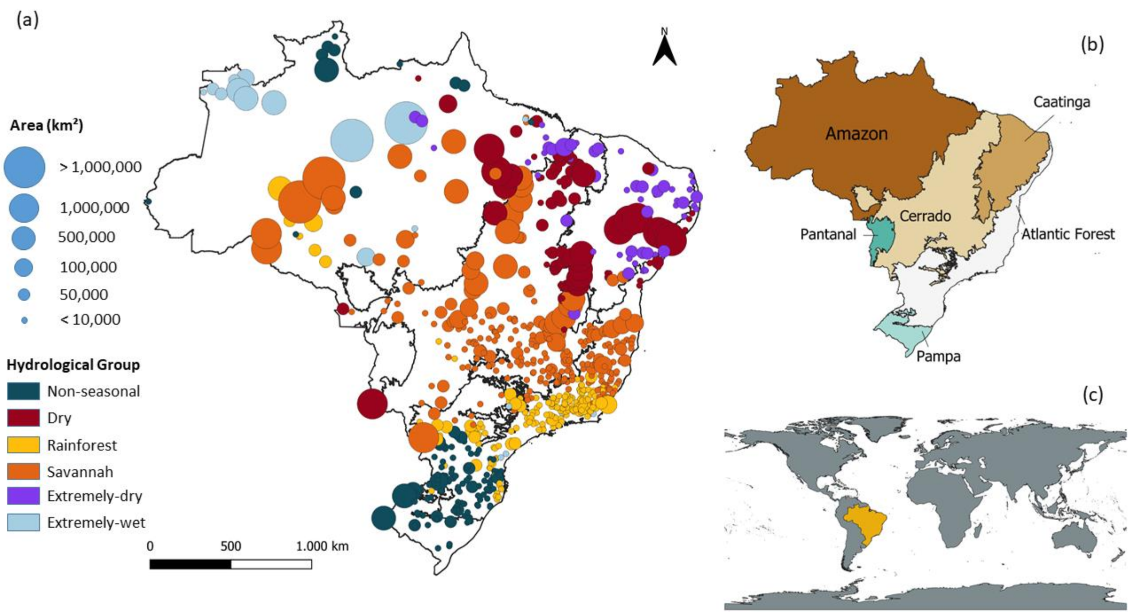

Figure 1: Maps of the spatial distribution of a) each catchment by area and hydrological group over the territory; b) the six Brazilian Biomes (Amazon, Atlantic Forest, Cerrado, Caatinga, Pampa, and Pantanal); and c) Brazil, in yellow, on the world map.

Group 1 (Non-seasonal) presents high values of mean streamflow, with no clear relation between precipitation and temperature throughout the year, with no definitive division between wet and dry seasons. The catchments are located mainly in the South region, with Pampa, Atlantic Forest biomes, and some in the Amazon biome. Group 2 ("Dry"), located mainly in Cerrado biome in the region between Amazon and Caatinga biomes, has well-defined rainy and dry seasons, with the highest temperatures of all groups (20°C to 32°C). The catchment's aridity index (the ratio between mean-annual potential evapotranspiration and mean-annual precipitation) has a great variance and low values of mean precipitation. Group 3 (Rainforest) is located mainly in the Atlantic Forest biome region, with high amounts of precipitation, and a well-defined rainy season, which coincides with the highest temperatures of summer in the region. Group 4 (Savannah) refers to catchments that present savannah characteristics and is located mainly in the Cerrado biome. The vegetation is composed basically of grass and forest, with an aridity index varying from 1 to 2. Group 5 (Extremely-dry), located in the Caatinga biome, is covered mainly by grass and shrubs, showing high values of aridity index. This group is different from Group 2 (Dry) because of its non-perennial streams. Finally, Group 6 (Extremely-wet) presents the highest values of precipitation and streamflow. The catchments are located in the Amazon and Atlantic Forest biomes and are covered mainly by forests.

To evaluate the performance of Grunsky's framework to predict runoff in ungauged basins, we conducted the following regionalization framework: (i) First, for each catchment of each hydrological group, we computed a median α value using the "leave one out" cross-validation approach. That is, the α of a specific catchment was computed considering the median α from all the other catchments of the same group, with exception of itself. This step is fundamental to preserving independence in the determination of α. Moreover, we use the MMSE value from each catchment as it obtained the best performance among the three evaluated methods (see Results section 5.2). (ii) Then, using the Grunsky method and the median α of the step (i), we estimated the runoff for the 716 catchments. (iii) Finally, we evaluated the performance of the regionalized Grunsky's method to estimate catchments' runoff through R^2, RMSE, and KGE values. We then repeated the same process for each month, on a monthly scale. Here we present a schematic view of the analysis method (Figure 2).

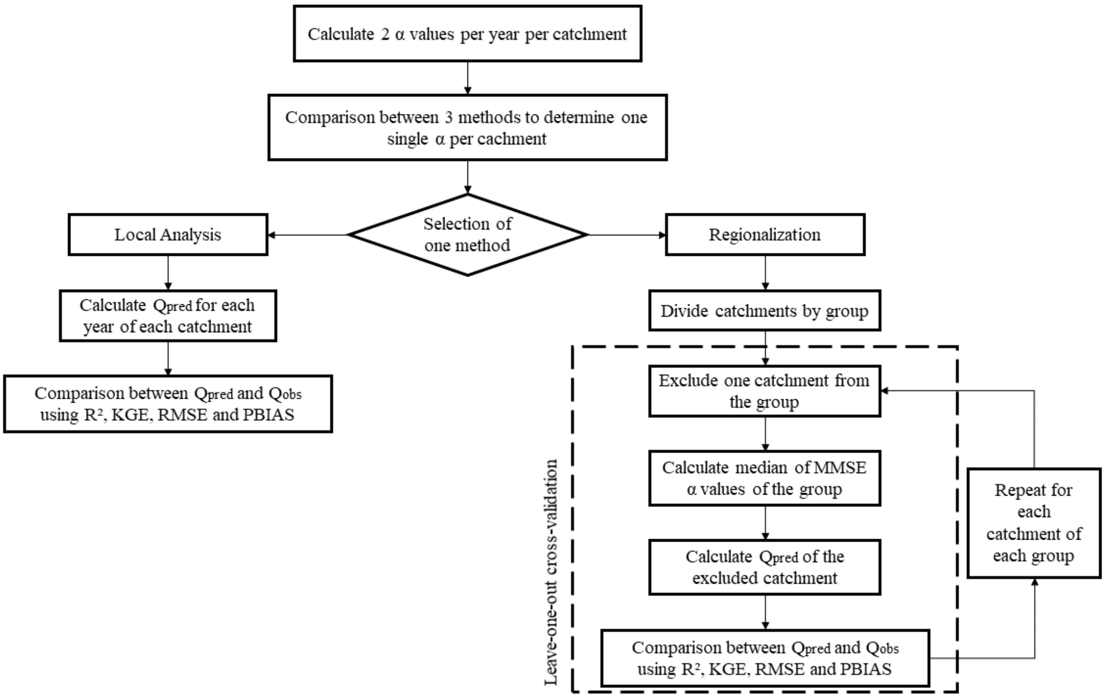

Figure 2: Schematic view of the analysis method.

5. Results and discussion

The results are presented in three steps: (i) we first assess the relationship between observed annual runoff (Q_{obs}) and precipitation (P_{obs}) with Grunsky's method and choose the best α value for each catchment in both interannual and monthly temporal scales. (ii) Then we evaluate the Grunsky's framework for each time scale on local analysis, catchment by catchment; and (iii) finally, we assess the performance of the proposed regionalization method to compute interannual and monthly runoff across Brazil.

5.1. Relations between runoff and precipitation and Grunsky's equation

Figure 3 shows a scatter plot between P_{obs} and Q_{obs} for each year of the 30-year time series of the 716 evaluated catchments, colored according to their aridity index computed annually, and Grunsky's method plots for multiple α values. The aridity index, commonly used from the Budyko framework (1974), is given as the ratio between mean potential evapotranspiration ($\overline{PET}$) and mean precipitation $(\bar{P})$, and is notably a good and worldwide used parameter to analyze climatic influence on long-term runoff (Meira Neto *et al.*, 2020). We noted that the relation between P_{obs} and Q_{obs} follows, in general, the Grunsky Generalized curve (Equation 2) and fits well for both low and high α values, indicating the Grunsky framework is able to capture the inter-catchment and annual rainfall-runoff relationship variability. However, when looking at the aridity index values, we note lower dispersion of the Grunsky's curve for aridest catchments, possibly due to the high initial abstraction during rain events in dry catchments (de Figueiredo *et al.*, 2016). The highest aridity index values occur for catchments with lower P_{obs} amounts, and they show a better fit to the Grunsky's curve than for lower values of aridity index when P_{obs} presents higher values. These results were somehow expected since the method was proposed and has already been applied to Mediterranean catchments, which are generally drier than most of climate types in Brazil (Spinoni *et al.*, 2015).

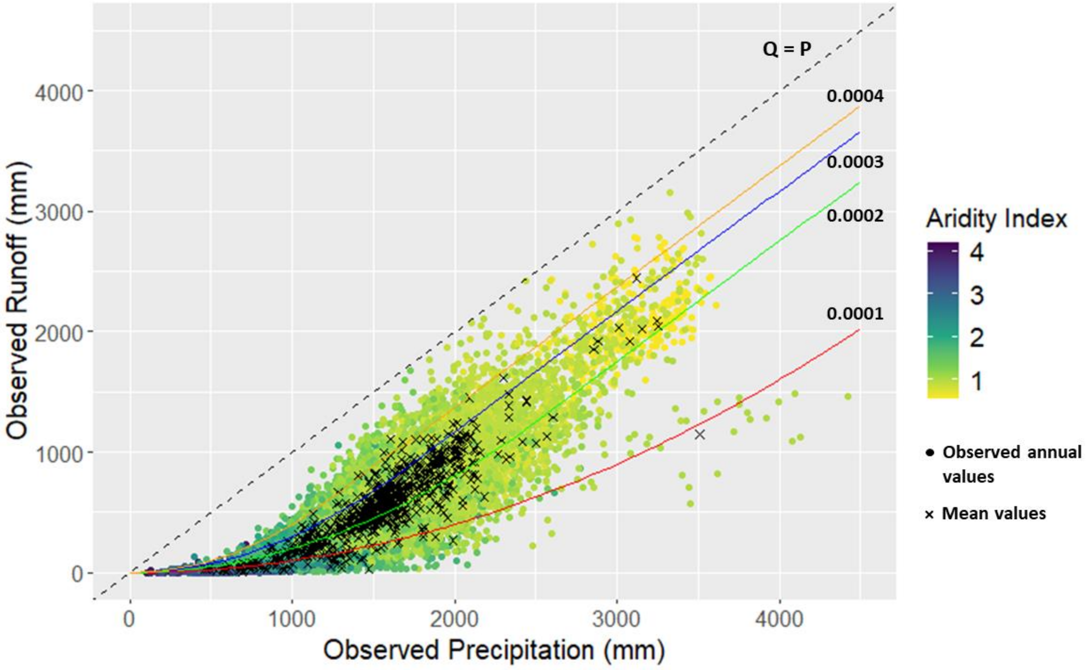

Figure 3: Scatter plot with observed and mean annual runoff versus annual precipitation for the 30 years series from the 716 CABra catchments, and Grunsky's equation plots for multiple α values (mm^{-1}). Colored according to the aridity index value from each catchment.

Figure 3 shows the same plot for the long-term (mean-annual) temporal scale. The observed scatter values agree with the behavior of Grunsky's generalized method functions, with α values ranging from 0.0001 to 0.0004. Even the catchments with higher long-term mean annual rainfall ($\bar{P}$ > 2,000 mm year^{-1}) fall inside the interval between the theoretical curves. Our results corroborate with those reported by Santos and Hawkins (2011), which presented α values from 0.00009 to 0.00043 mm^{-1}, with the highest P_{obs} = 2,770.1 mm. Values tend to be sparse for precipitations higher than 2,000 mm year^{-1}, as observed previously.

5.2. *Choosing the best method to define a single α for each catchment*

We analyzed the rainfall-runoff relations for the interannual and monthly scales, considering, however, a single α value for each catchment, calculated by the three different methods described in Section 4.3 (mean, median, and MMSE). To better visualize de variability of α values calculated from the 30-year series, we presented them on Supplementary Material (Figure S2), separated by the different hydrological groups in both interannual and monthly scales.

From the interannual analysis, it is evident that the behavior of the α coefficient changes according to each hydrological group. The difference between the boxplots of the three evaluated approaches (mean, median, and MMSE) was not so remarkable. Nevertheless, it is possible to note that the α values estimated through the mean approach exhibited a larger variability. Interestingly, although the MMSE method involves more complex calculations to determine α, using linear regressions between predicted and observed runoff to define the α that minimizes Minimum Square Error (MSE), it obtained a very similar variability to the median approach. This does not hold for the monthly temporal scale. In this case, the MMSE approach showed the lowest variations. As expected, α is higher for wetter catchments, and lower for drier catchments. Dry and Extremely-dry groups show the lowest MMSE α values for all methods. These values are coherent to scatter plots per hydrological group on Supplementary Material (Figure S3), where there's a higher density of points near the curve for α = 0.0001 for these drier groups.

The α values obtained for each month by using the three evaluated approaches were presented on Supplementary Material (Figure S4). We found that the α coefficient presents a higher range in May, June, July, and August, the driest months of the year in Brazil, which was expected since, during this period of the year, a great part of Brazilian territory is affected by climatic instability and higher variability in the rainfall-runoff relation (Silva and Simões, 2014; Zhang *et al.*, 2018). For the Caatinga biome (Dry and Extremely-dry groups), Pinheiro *et al.* (2016) found that soil water content has a stronger influence on evapotranspiration and air temperature during the June-September and December-January periods, which increases the uncertainty of the Grunsky method by adding more variance to hydrological processes. Differently from mean values, median α show lower scatter in all months, especially in May, though it still presents high variability in June, July, and August. We observed that the α MMSE value resulted in a great decrease in α dispersion, given the purpose of this method to determine an α value that minimizes errors for each catchment, with only the months June and July having a higher variability in α values. These variations were higher than that found in the same months for the median and mean values. We also noted that, on a monthly scale, α values tend to be approximately one order of magnitude higher than that obtained on an annual scale. This fact was already expected since annual precipitation is divided by the 12 months of the year, and by Equation 2, α is proportionally inverse to precipitation.

From Figure 4, where KGE from all catchments by group appears in the ECDF curves from 0 to 1.0 (normalized number of catchments for each group), we note that MMSE α values present better results for all groups, especially for Dry and Extremely-dry groups, where the difference from the other methods is more clear on an interannual scale. This result contradicts the already indicated in the scatter plot of Figure 3, which presented the best results from drier catchments, while Figure 4 shows that the best performances are from the Non-seasonal, Rainforest, and Savannah groups. Whereas KGE assumes mainly values from 0 to 1 in some groups, a part of catchments presents negative values, which does not imply in a bad simulation, since KGE values higher than -0.41 still indicate a good performance compared to a mean flow reference (Knoben, Freer and Woods, 2019). The MMSE α values spatialized within Brazilian territory, for each catchment, are found in Supplementary Material (Figure S5).

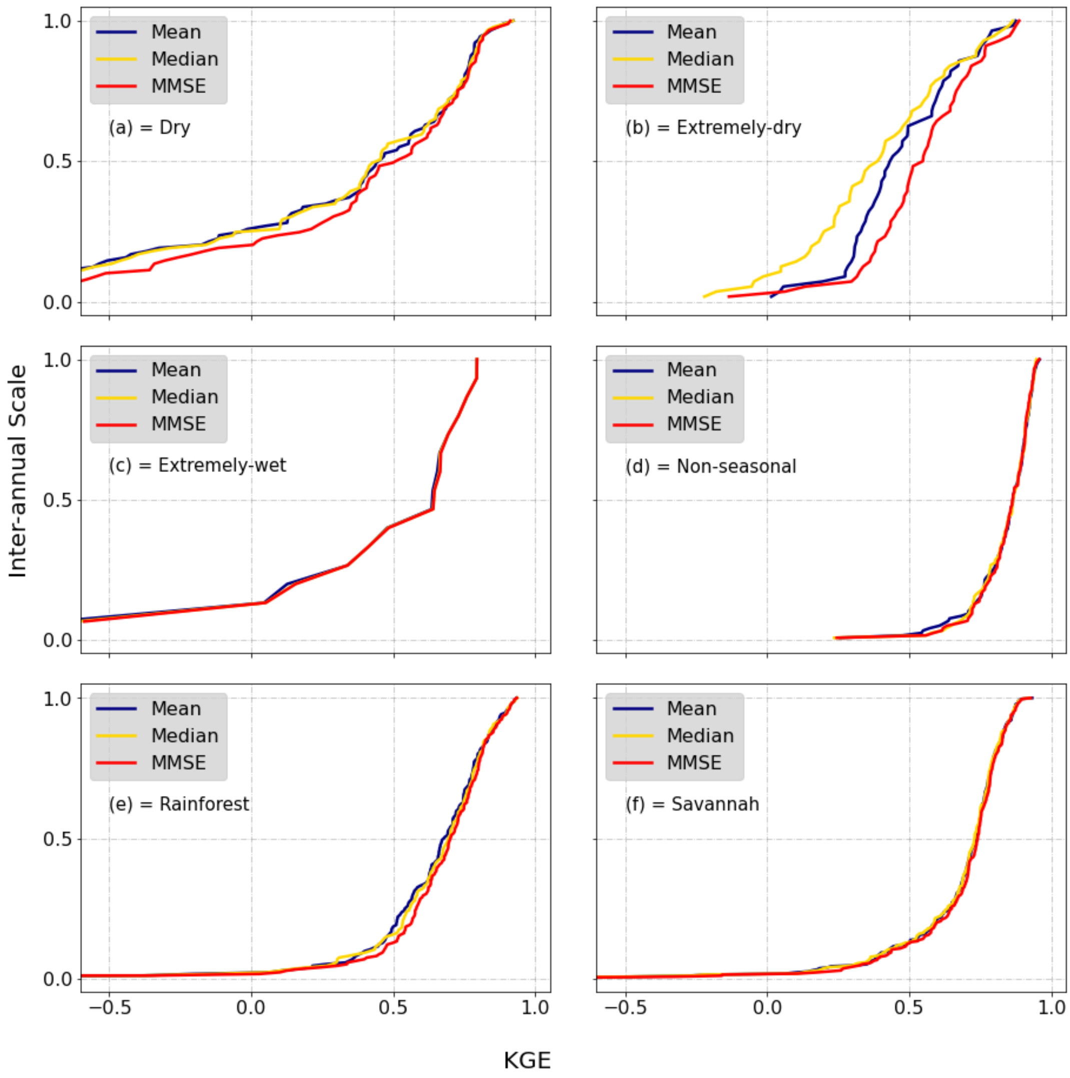

Figure 4: Cumulative Distribution Function (CDF) curves comparing KGE values for each method for determining α value on an interannual scale, for hydrological groups a) Dry, b) Extremely-dry, c) Extremely-wet, d) Non-seasonal, e) Rainforest and f) Savannah.

On a monthly scale (Figure 5), it is more evident that MMSE α values, in general, for all groups, present better results than mean and median values. Among the groups, the best results are for the Non-seasonal group, with a great percentage of KGE > 0.5. The Dry group presents more than 50% of values below 0, indicating worse predictions. As the MMSE presented a better performance in runoff prediction than the other approaches evaluated for interannual and monthly temporal scales, we used α values obtained by the MMSE in our study, for local and regionalized analysis. For details of boxplots of Monthly Observed Precipitation (mm) and Runoff (mm) for each hydrological group, see Supplementary Material (Figures S6, S7, and S8).

The difference in performance on different groups and time scales comes from the non-linearity of arid catchments behavior, especially for hydrological extremes, which might affect the intra-annual behavior more than interannual and long-term runoff. On the other hand, more humid catchments, given their more (relative) linear behavior, tend to have better performance (Parajka *et al.*, 2013; Salinas *et al.*, 2013).

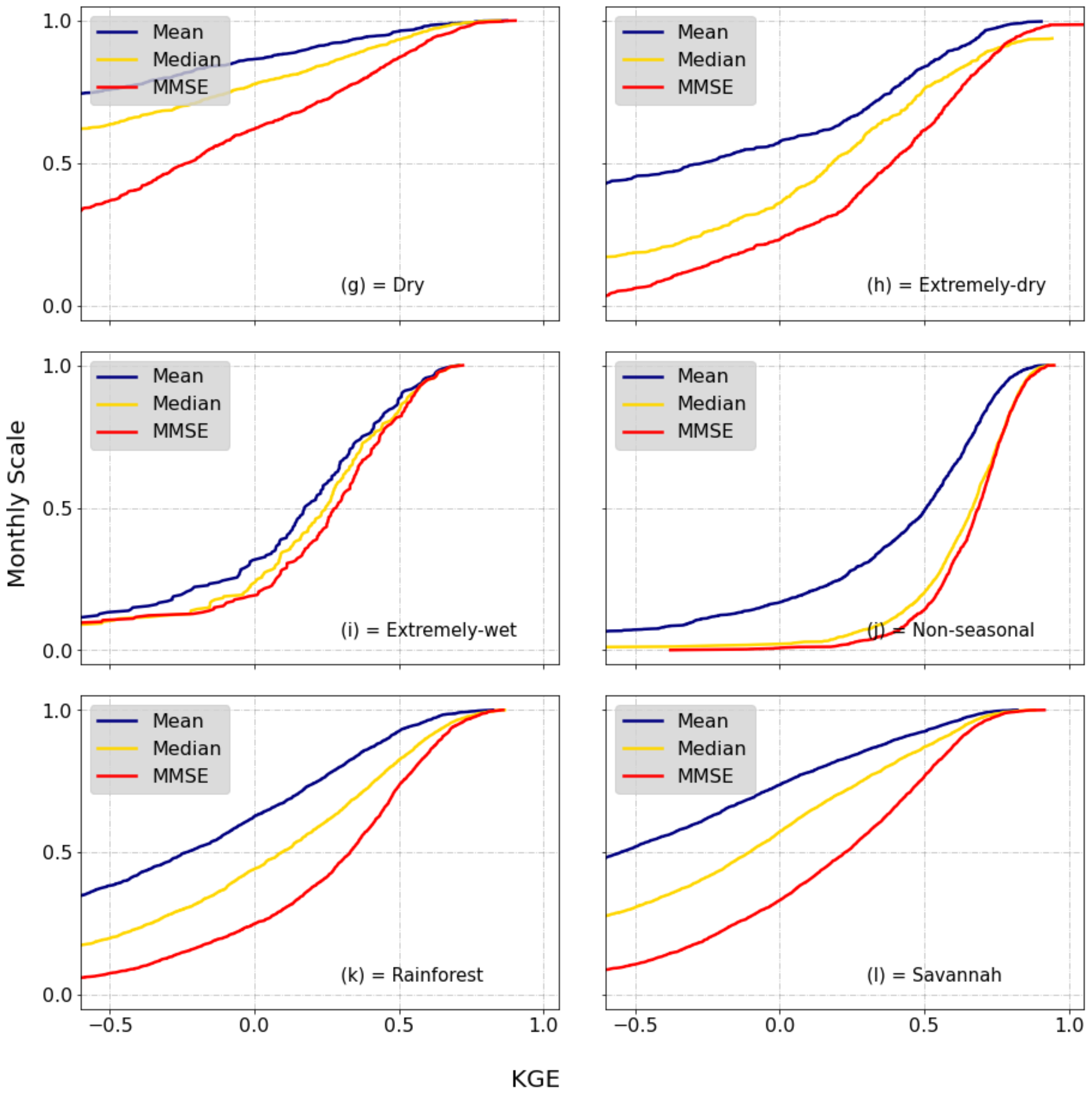

Figure 5: Cumulative Distribution Function (CDF) curves comparing KGE values for each method for determining α value on a monthly scale, for hydrological groups a) Dry, b) Extremely-dry, c) Extremely-wet, d) Non-seasonal, e) Rainforest and f) Savannah.

5.3. Local Performance of α values

The predicted runoff was calculated then by Grunsky's Equation using the MMSE α value for each catchment on an interannual and monthly scale. The correlation between Predicted Runoff and Observed Runoff was then evaluated using metrics such as R^2, KGE, relative RMSE (absolute RMSE divided by mean runoff of each group), and Percentage Bias (PBIAS).

5.3.1. Interannual scale

Figure 6 illustrates the relations between predicted and observed annual runoff for each group, colored by point density with the number of neighboring points. All groups present R^2 higher than 0.80, considered a satisfactory result, especially for the Dry group, which shows $R^2 = 0.93$. RMSE is higher for Extremely-dry (48.3%) group, and smaller for Extremely-wet (13.1%), and Non-Seasonal (18.3%) groups, also corroborating with the previous analysis from Section 5.1, as smaller RMSE indicates better correlation. KGE shows values close to 1 in almost all groups, indicating a high correlation. Dry, Savannah and Extremely-wet groups present higher KGE (0.929, 0.904, and 0.901, respectively). The predictions show higher KGE values than other studies, such as Gao *et al.* (2021), that predicted mean annual runoff in 35 catchments in the United States with the use of a water balance model calculated from Budyko-type equations and spatial distribution models for soil data, resulting in $R^2 = 0.76$.

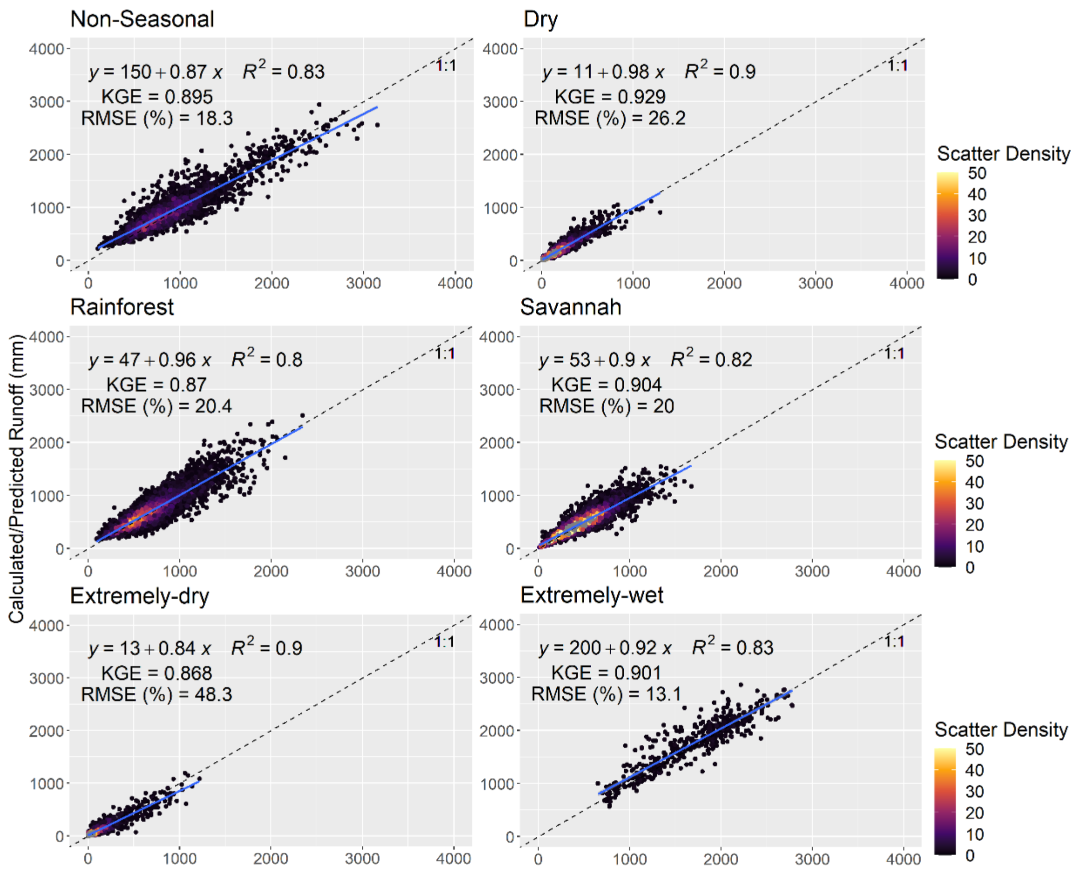

Figure 6: Scatter plot with observed annual runoff versus predicted annual runoff from MMSE α value for each catchment during the 30 years series from the 716 CABra catchments, divided by groups. Colored by the number of neighbors for each point.

The adjusted curves and respective equations indicate the general behavior of this relationship. The fact that the regression coefficient is <1 indicates that the predicted values tend to be lower than the observed values, in general. The ideal scenario would be the constant to be equal to 0 and the coefficient equal to 1, so that the predictions are the closest as possible to the observations.

The results may indicate a constant behavior of each catchment during the 30-year series, generating an α value that well represents the catchment, thus resulting in a good correlation. In addition, it can be noted that catchment characteristics according to each group may influence α for a better correlation, although it is not possible to well differentiate the performance between drier and wetter groups. It is interesting to note that even the Non-seasonal group presented a good performance even though this group does not present a very defined behavior of seasonality between water and energy availability (Almagro, 2021), with a great variety of catchments and characteristics.

To understand the correlation between predicted and observed runoff during the 30-year series, KGE and PBIAS values for each catchment were spatialized within the Brazilian territory, as shown on Figure 7. The higher values of KGE for the catchments are mainly located where there are the groups with higher values of precipitations (Non-seasonal in the south region, Rainforest in the southeast, and Extremely-wet in the northwest). Also, Dry and Extremely-dry groups present values closer to 0, indicating a relatively lower correlation according to KGE. The highest positive PBIAS values (>20%) correspond to the northeast region (Dry and Extremely-dry groups), indicating an overestimation of the model. The other regions of the territory mainly present negative PBIAS, indicating an underestimation of Q_{obs}. The values show good performance, ranging mostly from -5 to 5%. Also, as discussed for KGE, a lower value of PBIAS was found for the wettest catchments. One should remember that these values refer to one α value for each catchment, corresponding to the MMSE for each catchment. The regionalization prediction would be evaluated by using a single α for each hydrological group, discussed in Section 5.4.

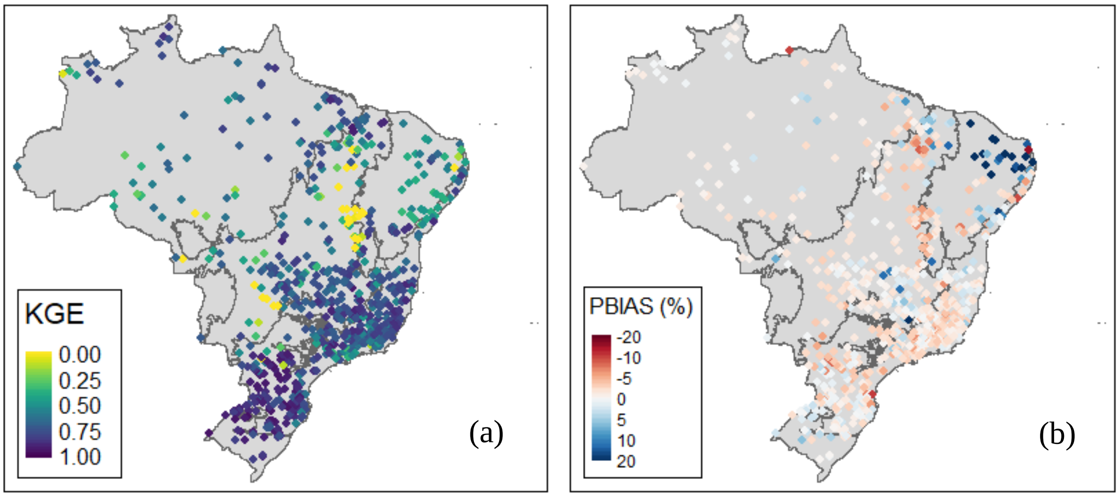

Figure 7: (a) Map of KGE values and (b) Map of PBIAS values (mm year^{-1}) for catchments, comparing observed runoff with predicted runoff with α correspondent to the MMSE α for each catchment.

5.3.2. Monthly scale

At the monthly scale, we calculated the MMSE α value for each catchment, and the spatialized values according to each month are on the Supplementary Material (Figure S10). It was observed that in the Southeast region (Non-seasonal, Savannah, and Rainforest groups), the drier months (i.e., from May to August, especially June and July) correspond to the lowest KGE values (Figure 8). During the other months, KGE values are in general higher than 0.50, also observed in Choi *et al.* (2021) when studying a prediction model for watersheds in South Korea, which indicates a good model prediction, especially for wetter months. Comparing RMSE values, the results are similar to predictions from Samantaray *et al.* (2022) using Support Vector Machine (SVM), with scenario conditions with minimum RMSE of 2.721 and 5.117 for two stations in India. The performance variability between dry and wet seasons was already observed by Zhang *et al.* (2018), which applied rainfall-runoff modeling for large basins in southeastern Brazil.

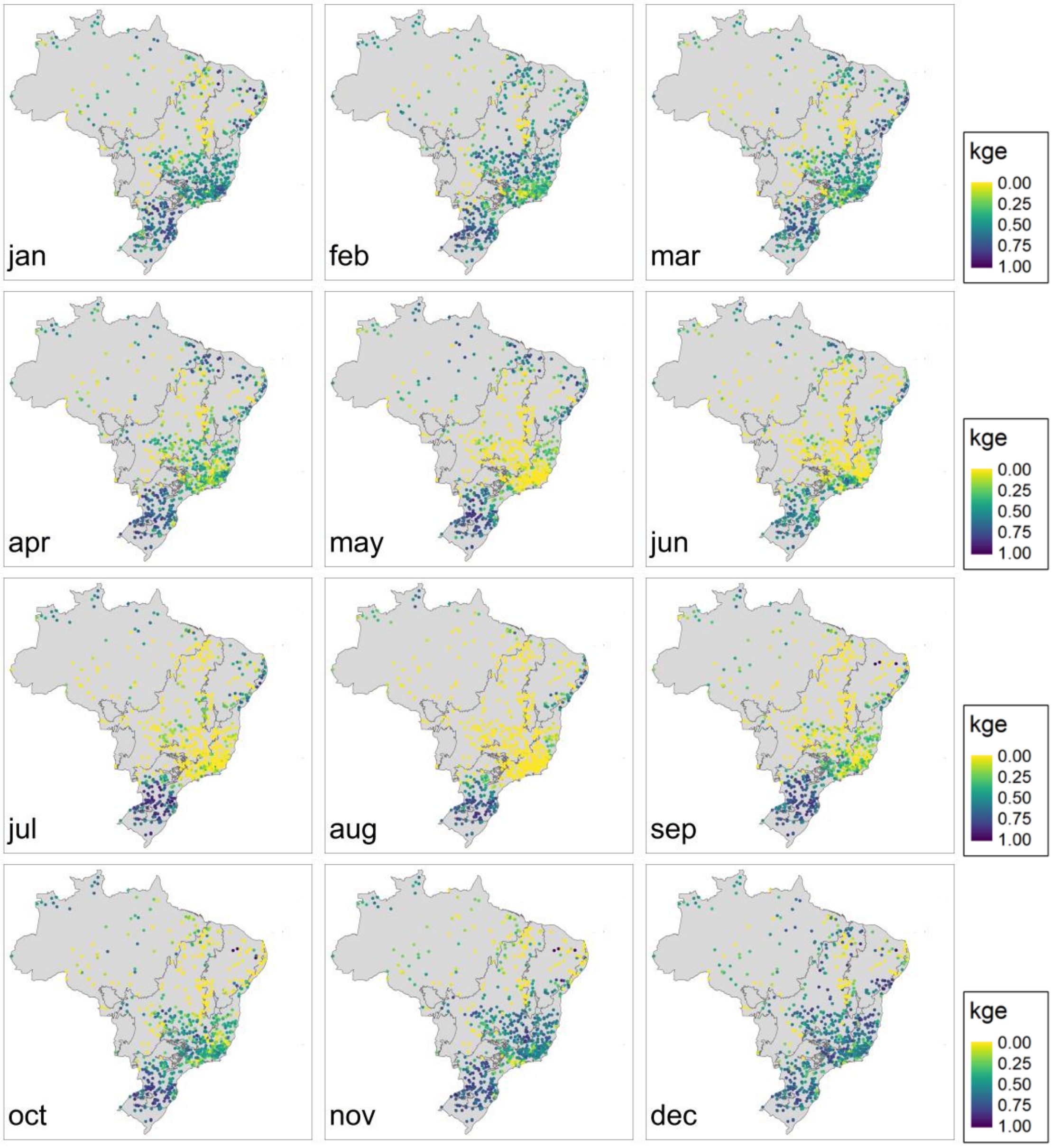

Figure 8: Maps of KGE from MMSE α value for each group for each month, from January (top left) to December (bottom right).

PBIAS between Predicted Runoff and Observed Runoff was calculated for all months, and the results are shown on Supplementary Material (Figure S11). The maps indicate a prevalence of negative PBIAS, implying an underestimation in MMSE predictions, i.e., the predicted runoff is lower than the observed one. The months with the highest bias differ between regions, being the Southeast region with the highest biases, mainly from January to March, and August to October. The lowest biases are observed in the North region (Extremely-wet and part of the Savannah group), with the highest biases from August to October.

5.4. *Regionalization of α values*

It is important to analyze the α coefficient according to catchment characteristics and to define a proper clustering proposal that can be used for predictions on ungauged basins. The clustering proposal analyzed was the hydrological group division (Almagro, 2021), on an interannual and monthly scale. Results for the regionalization approach are described below.

5.4.1. *Interannual scale*

We evaluate the use of six α values to predict runoff across Brazil, calculated from the median of the MMSE α values of the catchment belonging to the groups. The correlation between Predicted Annual Runoff and Observed Annual Runoff was then evaluated by leave-one-out cross-validation, using metrics such as R^2, KGE, RMSE, and PBIAS, as described in Section 4.4. Figure 9 illustrates the relations between predicted and observed annual runoff for each group, colored by the scatter density with the number of neighboring points.

*Table 1: Median of MMSE **α** value (mm^{-1}) of each hydrological group, for an interannual scale.*

Hydrological group	**Median of MMSE α value (mm^{-1})**
Non-seasonal	0.0002558485
Dry	0.0001249390
Rainforest	0.0002220525
Savannah	0.0002124880
Extremely-dry	0.0000923000
Extremely-wet	0.0002515930

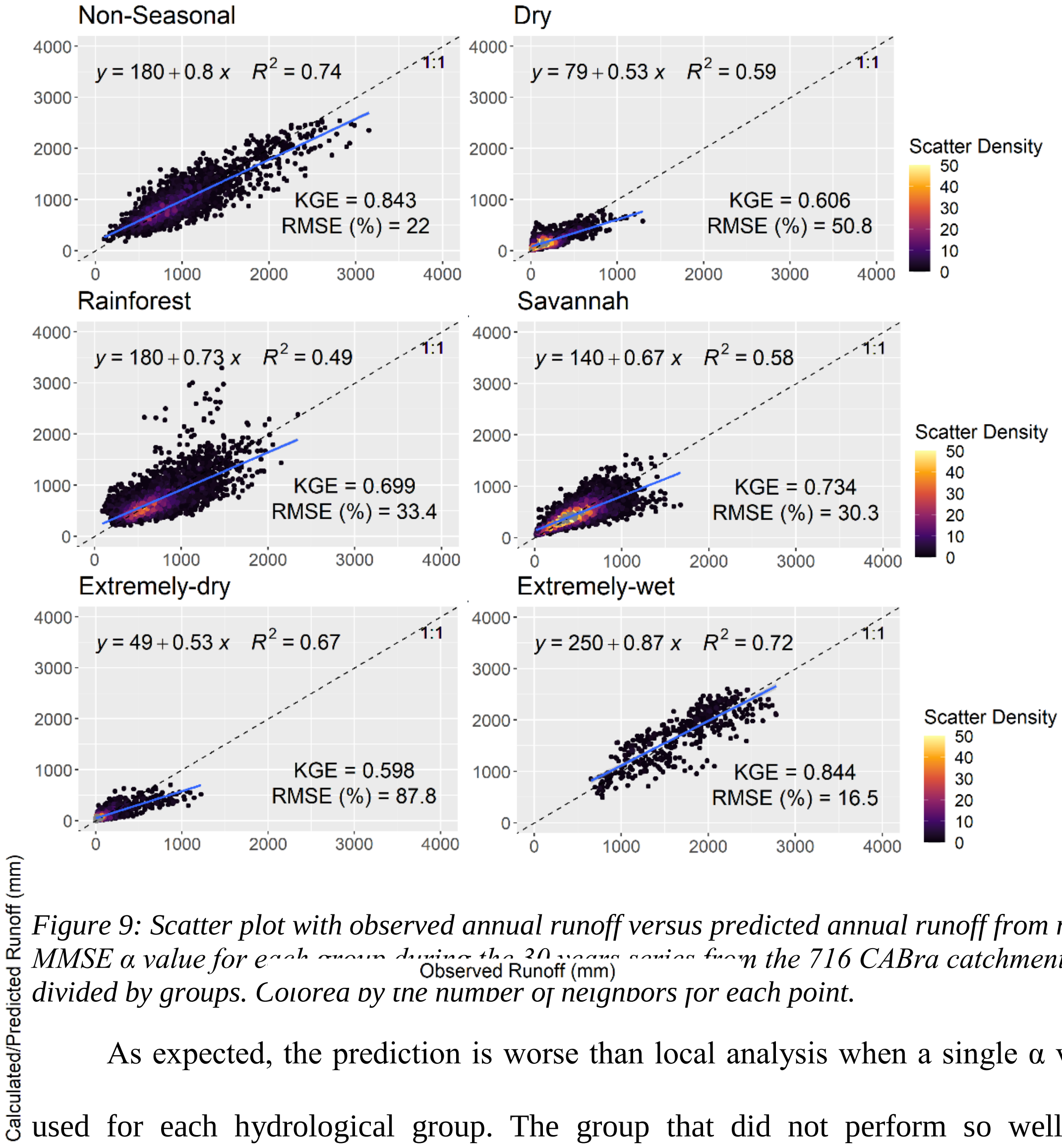

Figure 9: Scatter plot with observed annual runoff versus predicted annual runoff from median MMSE α value for each group during the 30 years series from the 716 CABra catchments, divided by groups. Colored by the number of neighbors for each point.

As expected, the prediction is worse than local analysis when a single α value is used for each hydrological group. The group that did not perform so well is the Extremely-dry group (R^2 = 0.67 and KGE = 0.598), and the best one is the Non-seasonal (R^2 = 0.74 and KGE = 0.843). Despite worse prediction indicators for each group, KGE and R^2 were higher than 0.598 and 0.49, respectively, for all groups, highlighting the good performance of the method on an interannual scale, considering all years of a series within one catchment. RMSE is higher for Extremely-dry (89.0%) and Dry (50.8%) groups, and smaller for Non-seasonal (22.1%) and Extremely-wet (16.2%). The adjusted equations follow the same behavior as the local scale analysis.

5.4.2. Monthly Scale

The regionalization on a monthly scale was done by assuming one α coefficient for each group and month, from the median values from the catchments. The boxplots of KGE values for the 12 months divided by each group are presented in Figure 10. These α values and the boxplots of PBIAS values are exhibited in the Supplementary Material (Table S1 and Figure S12, respectively).

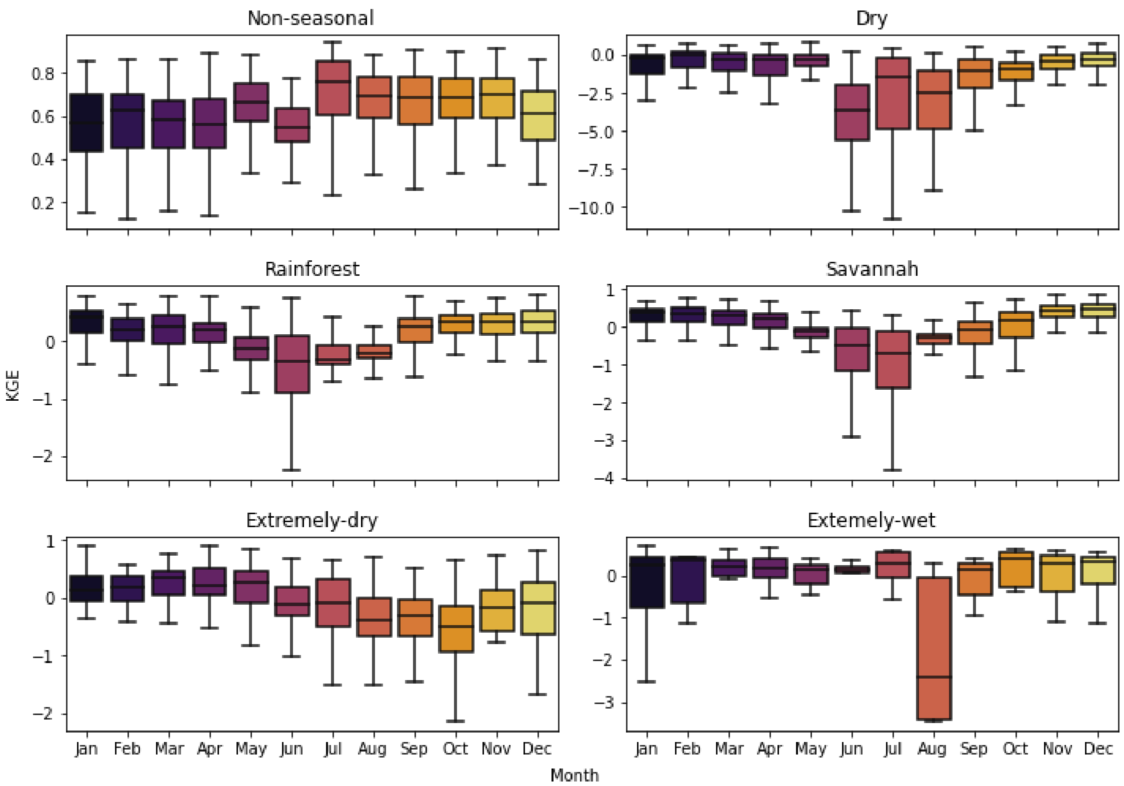

Figure 10: Boxplot with KGE from regionalized monthly runoff prediction for each month and group from the 716 CABra catchments.

The groups with the lowest values of PBIAS and highest values of KGE are Non-seasonal and Extremely-wet, given a more constant rainfall behavior on an intra-annual scale (Almagro, 2021). The worse performance was the Dry group, with median monthly KGE < 0 for all months, and PBIAS > 100% from June to August. The regionalization for a monthly scale appears to have worse results than the interannual scale for all groups, indicating the influence of interannual and intra-annual behavior of streamflow on the method's performance. The results appear to be similar to the predictions by Arheimer *et al.*, 2020, which showed relatively good results in Brazilian North region catchments when predicting monthly runoff, with median monthly KGE > 0.60, and worse performance in the North-east region, with KGE < -1. Dry and Extremely-dry groups might have higher variance on an interannual scale, and particularly on an intra-annual scale due to the influence of soil water content, and high rates of evapotranspiration (Pinheiro *et al.*, 2016).

6. Limitations and broader implications

This paper has demonstrated the usefulness of the Grunsky method on an annual and monthly scale, especially for wetter catchments. Broader studies are required for the performance improvement in drier catchments, such as consideration of the hydrological balance and inter-catchment groundwater fluxes, as proposed by Liu *et al.* (2020) and Schwamback *et al.* (2022).

The analysis methods employed, due to their relative ease of application, have great replication capacity. That is, they may be used for a large database of hydrographic basins, as is the case with CABra, and can be updated in the future, as needed. Further studies may complement and enrich the topic discussion through the adaptation of the methods already described, or further analysis of other characteristics of hydrographic basins since the database has a great variety of data. Through the CABra database, especially the values of precipitation and flow in different time scales, it is possible to analyze the influence of environmental factors on runoff response to precipitation.

Regionalization is an important tool for the estimation of water availability in poorly or ungauged catchments(Athira *et al.*, 2016), and contribute for different water uses (Agarwal *et al.*, 2016; Bork *et al.*, 2021), specially in developing countries where hydrological data is scarce (Beskow *et al.*, 2016). In Brazil, legislation on water resources allows each state to determine its values of reference for granting water use privileges (Honório *et al.*, 2020). In several Brazilian states, the reference environmental flow is Q_{95} permanence flow, which guarantees that the flow of a determined river equals or passes this value on 95% of the time (ANA, 2007; Piol *et al.*, 2019). Broader studies might be needed seeking the performance evaluation of Grunsky's Method for practical purposes such as predicting permanence flow curves for different water uses.

7. Conclusions

This study aims to predict runoff in ungauged basins from values of precipitation and the α coefficient applied on a simple and easy-to-use equation. The method could support fast decision-making processes, and hydraulic projects, from irrigation to dam constructions, and generate permanence flows for licensing projects. From this analysis, some conclusions deserve special attention and are described below.

First, results could show that the Grunsky method can be applicable for Brazilian catchments, having a good fit, especially for Non-seasonal and Extremely-wet groups on annual and monthly scales, showing less variability in both interannual and intra-annual scales. The best method to obtain α for each catchment is the MMSE method, showing good results for annual and monthly analysis, for all hydrological groups. Grunsky's Equation ranged from annual KGE = 0.60 to 0.84 for the regionalized method, by using only precipitation and the α coefficient. Considering the usage of each approach, the regionalization method, using a single α coefficient for each hydrological group, has a wider use for ungauged basins, although a local analysis might be convenient for analysis on poorly gauged catchments, since it shows a better performance.

Our study demonstrated the possibility of Grunsky's method to be applied to interannual and monthly scales. The temporal scale seems to be a paramount factor in the decision of whether the model is good for prediction or not. Non-seasonal catchments, located mainly in the South region of Brazil, show more constancy and better performance for predictions on an interannual and intra-annual scale. Considering the simplicity and practical use of the method, this work presents a reliable alternative to more complex methods to calculate runoff from very few parameters. We expanded Grunsky's method usage from Mediterranean to tropical climate basins in Brazil, which might be an important tool for future studies and also for hydraulic infrastructure projects.

8. Acknowledgements

The authors acknowledge the Graduate Program in Hydraulic Engineering and Sanitation – PPGSHS (USP-EESC) and Graduate Program in Environmental Technologies – PPGTA (UFMS-FAENG) for the scientific support. This study was supported by grants from the Coordination for the Improvement of Higher Education Personnel (CAPES) - Finance Code 001 and CAPES PrInt, the Ministry of Science, Technology, and Innovation (MCTI) and National Council for Scientific and Technological Development (CNPq) (grants numbers 441289/2017-7 and 309752/2020-5), and São Paulo Research Support Foundation (FAPESP, grants numbers 2015/03806-1, 2019/24292-7, and 2020/08140-0).

9. References

Agarwal, A. et al. (2016) ‘Hydrologic regionalization using wavelet-based multiscale entropy method’, Journal of Hydrology, 538, pp. 22–32. doi: 10.1016/J.JHYDROL.2016.03.023.

Almagro, A. et al. (2020) ‘CABra: a novel large-sample dataset for Brazilian catchments’, Hydrology and Earth System Sciences Discussions, (October), pp. 1–40. doi: 10.5194/hess-2020-521.

Almagro, A. (2021) Towards a better understanding of catchment hydrology in Brazil. Federal University of Mato Grosso do Sul. Available at: https://repositorih.ufms.br/handle/123456789/4035.

ANA (2007) Diagnóstico da outorga de direito de uso de recursos hídricos no Brasil. Caderno de Recursos Hídricos. Volume 4. Agência Nacional de Águas. Brasília. Available at: www.ana.gov.br.

Arheimer, B. et al. (2020) ‘Global catchment modelling using World-Wide HYPE (WWH), open data, and stepwise parameter estimation’, Hydrology and Earth System Sciences, 24(2), pp. 535–559. doi: 10.5194/HESS-24-535-2020.

Athira, P. et al. (2016) ‘Predictions in ungauged basins: an approach for regionalization of hydrological models considering the probability distribution of model parameters’, Stochastic Environmental Research and Risk Assessment, 30(4), pp. 1131–1149. doi: 10.1007/S00477-015-1190-6.

Ballarin, A. S. et al. (2022) ‘The Impact of an Open Water-balance Assumption on the Understanding of the Factors that Control the Long-term Streamflow Components’. doi: 10.1002/ESSOAR.10510930.1.

Beck, H. E., de Roo, A. and van Dijk, A. I. J. M. (2015) ‘Global Maps of Streamflow Characteristics Based on Observations from Several Thousand Catchments’, Journal of Hydrometeorology, 16(4), pp. 1478–1501. doi: 10.1175/JHM-D-14-0155.1.

Beskow, S. et al. (2016) ‘Potential of the LASH model for water resources management in data-scarce basins: a case study of the Fragata River basin, southern Brazil’, https://doi.org/10.1080/02626667.2015.1133912, 61(14), pp. 2567–2578. doi: 10.1080/02626667.2015.1133912.

Beven, K. J. (2000) ‘Uniqueness of place and process representations in hydrological modelling’, Hydrology and Earth System Sciences, 4(2), pp. 203–213. doi: 10.5194/HESS-4-203-2000.

Bloschl, G. et al. (eds) (2013) Runoff Prediction in Ungauged Basins. Cambridge: Cambridge University Press. doi: 10.1017/CBO9781139235761.

BORK, C. K. et al. (2021) ‘Minimum streamflow regionalization in a Brazilian watershed under different clustering approaches’, Anais da Academia Brasileira de Ciências, 93(suppl 4). doi: 10.1590/0001-3765202120210538.

Budyko, M. I. (1974) Climate and Life, New York, NY: Academic Press.

Choi, J. et al. (2021) ‘Usefulness of Global Root Zone Soil Moisture Product for Streamflow Prediction of Ungauged Basins’, Remote Sensing 2021, Vol. 13, Page 756, 13(4), p. 756. doi: 10.3390/RS13040756.

Dal Molin, M. et al. (2020) ‘Understanding dominant controls on streamflow spatial variability to set up a semi-distributed hydrological model: The case study of the Thur catchment’, Hydrology and Earth System Sciences, 24(3), pp. 1319–1345. doi: 10.5194/HESS-24-1319-2020.

de Figueiredo, J. V. et al. (2016) ‘Runoff initiation in a preserved semiarid Caatinga small watershed, Northeastern Brazil’, Hydrological Processes, 30(13), pp. 2390–2400. doi: 10.1002/HYP.10801.

Gao, Y. et al. (2021) ‘Diagnosis toward predicting mean annual runoff in ungauged basins’, Hydrol. Earth Syst. Sci, 25, pp. 945–956. doi: 10.5194/hess-25-945-2021.

Grunsky, C. E. (1908) ‘Rain and Run-Off Near San Francisco, California’, Transactions of the American Society of Civil Engineers, 61(2), pp. 496–515. doi: 10.1061/TACEAT.0002036.

Grunsky, C. E. and Manson, M. C. E. (1908) Reports on the water supply of San Francisco, California, 1900 to 1908, inclusive. Transactions of the American Society of Civil Engineers. Available at: http://hdl.handle.net/2027/uc1.$b78665 (Accessed: 13 March 2022).

Gupta, H. V. et al. (2009) ‘Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling’, Journal of Hydrology, 377(1–2), pp. 80–91. doi: 10.1016/J.JHYDROL.2009.08.003.

Hersbach, H. et al. (2020) ‘The ERA5 global reanalysis’, Quarterly Journal of the Royal Meteorological Society, 146(730), pp. 1999–2049. doi: 10.1002/qj.3803.

Honório, M. D. S. et al. (2020) ‘Comparison of two methods for determining Q95reference flow in the mouth of the surface catchment basin of the Meia Ponte river, state of Goiás, Brazil’, Open Engineering, 10(1), pp. 469–476. doi: 10.1515/ENG-2020-0053/MACHINEREADABLECITATION/RIS.

Hrachowitz, M. et al. (2013) ‘A decade of Predictions in Ungauged Basins (PUB)—a review’, https://doi.org/10.1080/02626667.2013.803183, 58(6), pp. 1198–1255. doi: 10.1080/02626667.2013.803183.

Justin, J. D. et al. (1915) ‘Discussion: Run-Off from Rainfall and Other Data’, Transactions of the American Society of Civil Engineers, 79(2), pp. 1156–1208. doi: 10.1061/TACEAT.0002701.

Knoben, W. J. M., Freer, J. E. and Woods, R. A. (2019) ‘Technical note: Inherent benchmark or not? Comparing Nash-Sutcliffe and Kling-Gupta efficiency scores’, Hydrology and Earth System Sciences. doi: 10.5194/hess-2019-327.

Liu, Y. et al. (2020) ‘What is the hydrologically effective area of a catchment?’, Environmental Research Letters, 15(10). doi: 10.1088/1748-9326/aba7e5.

McDonnell, J. J. (2003) ‘Where does water go when it rains? Moving beyond the variable source area concept of rainfall-runoff response’, Hydrological Processes, 17(9), pp. 1869–1875. doi: 10.1002/HYP.5132.

Meira Neto, A. A. et al. (2020) ‘An Aridity Index-Based Formulation of Streamflow Components’, Water Resources Research, 56(9), pp. 1–14. doi: 10.1029/2020WR027123.

Montanari, A. et al. (2013) ‘“Panta Rhei—Everything Flows”: Change in hydrology and society—The IAHS Scientific Decade 2013–2022’, https://doi.org/10.1080/02626667.2013.809088, 58(6), pp. 1256–1275. doi: 10.1080/02626667.2013.809088.

Parajka, J. et al. (2013) ‘Comparative assessment of predictions in ungauged basins-Part 1: Runoff-hydrograph studies’, Hydrology and Earth System Sciences, 17(5), pp. 1783–1795. doi: 10.5194/HESS-17-1783-2013.

Pinheiro, E. A. R. et al. (2016) 'Importance of soil-water to the Caatinga biome, Brazil', Ecohydrology. doi: 10.1002/eco.1728.

Piol, M. V. A. et al. (2019) 'Performance evaluation of flow duration curves regionalization methods', Brazilian Journal of Water Resources. doi: 10.1590/2318-0331.241920170202.

Pomeroy, J. W. et al. (2007) 'The cold regions hydrological model: A platform for basing process representation and model structure on physical evidence', Hydrological Processes, 21(19), pp. 2650–2667. doi: 10.1002/HYP.6787.

Pool, S., Vis, M. and Seibert, J. (2021) 'Regionalization for Ungauged Catchments — Lessons Learned From a Comparative Large-Sample Study', Water Resources Research, 57(10), p. e2021WR030437. doi: 10.1029/2021WR030437.

Popper, K. (2002) The Logic of Scientific Discovery. Available at: http://books.google.com/books?hl=en&lr=&id=T76Zd20IYlgC&pgis=1.

Razavi, T., Coulibaly, P. and Asce, M. (2012) 'Streamflow Prediction in Ungauged Basins: Review of Regionalization Methods', Journal of Hydrologic Engineering, 18(8), pp. 958–975. doi: 10.1061/(ASCE)HE.1943-5584.0000690.

Salinas, J. L. et al. (2013) 'Comparative assessment of predictions in ungauged basins-Part 2: Flood and low flow studies', Hydrology and Earth System Sciences, 17(7), pp. 2637–2652. doi: 10.5194/hess-17-2637-2013.

Samantaray, S. et al. (2022) 'Monthly runoff prediction at Baitarani river basin by support vector machine based on Salp swarm algorithm', Ain Shams Engineering Journal, 13(5), p. 101732. doi: 10.1016/J.ASEJ.2022.101732.

Santos, F. L. and Hawkins, R. H. (2011) 'Generalized Mediterranean Annual Water Yield Model: Grunsky's Equation and Long-Term Average Temperature', Journal of Hydrologic Engineering, 16(11), pp. 874–879. doi: 10.1061/(ASCE)HE.1943-5584.0000402.

Schwamback, D. et al. (2022) 'Are Brazilian catchments gaining or losing water? The effective area of tropical catchments', Hydrological Processes, 36(3). doi: 10.1002/HYP.14535.

Silva, W. M. and Simões, S. J. C. (2014) 'Spatial Intra-Annual Variability of Precipitation Based on Geostatistics. A Case Study for the Paraıba Do Sul Basin, Southeastern Brazil', International Journal of Geosciences, 5, p. 408. doi: 10.4236/ijg.2014.54039.

Sivapalan, M. et al. (2003) 'IAHS Decade on Predictions in Ungauged Basins (PUB), 2003-2012: Shaping an exciting future for the hydrological sciences', Hydrological Sciences Journal, 48(6), pp. 857–880. doi: 10.1623/hysj.48.6.857.51421.

Spinoni, J. et al. (2015) 'Towards identifying areas at climatological risk of desertification using the Köppen-Geiger classification and FAO aridity index', International Journal of Climatology, 35(9), pp. 2210–2222. doi: 10.1002/JOC.4124.

Viglione, A. et al. (2013) 'Comparative assessment of predictions in ungauged basins - Part 3: Runoff signatures in Austria', Hydrology and Earth System Sciences, 17(6), pp. 2263–2279. doi: 10.5194/HESS-17-2263-2013.

Xavier, A. C., King, C. W. and Scanlon, B. R. (2016) 'Daily gridded meteorological variables in Brazil (1980–2013)', International Journal of Climatology, 36(6), pp. 2644–2659. doi: 10.1002/JOC.4518.

Yang, X. et al. (2020) 'Dependence of regionalization methods on the complexity of hydrological models in multiple climatic regions', Journal of Hydrology, 582, p. 124357. doi: 10.1016/J.JHYDROL.2019.124357.

Zhang, R. et al. (2018) 'Season-based rainfall–runoff modelling using the probability-distributed model (PDM) for large basins in southeastern Brazil', Hydrological Processes, 32(14), pp. 2217–2230. doi: 10.1002/HYP.13154.

Chapter 2

Improved empirical Grunsky Method for Streamflow prediction in ungauged Brazilian catchments

1. Abstract

Despite the reported advances since the 2000s in streamflow prediction through regionalization techniques by using regression equations, hydrologic similarity, hydrological models, and machine learning, this remains a difficult issue in Hydrology, particularly for ungauged basins. Here, we proposed an improved Grunsky generalized method to streamflow prediction for ungauged and poorly gauged basins. This approach utilizes the generalization of a rainfall-runoff relation coefficient (α) based on mean annual temperature. We evaluate this approach in 716 Brazilian catchments for interannual and monthly scales. Initially, we investigate the relationship between α and various catchment attributes such as the aridity index, annual average temperature, and potential evapotranspiration. Subsequently, we calculate α for each catchment through multiple linear regression, utilizing annual average precipitation and temperature as input variables. Finally, we assess the performance of mean annual runoff prediction by employing the computed α for each catchment, employing leave-one-out cross-validation. Our findings indicate that the improved Grunsky's approach is suitable to predict runoff for Brazilian basins mainly on interannual scale, presenting KGE = 0.899, R^2 = 0.82 and RMSE = 27.4%.

2. Keywords

Aridity index, mean temperature, potential evapotranspiration, regionalization.

3. Introduction

Runoff prediction at ungauged or poorly gauged basins (PUB) has been an important matter for hydrologists and emerges as one of the most fundamental topics in the hydrology field given its fundamental role for improved water resources management and enhanced hydrological processes understanding (Agarwal *et al.*, 2016; Du *et al.*, 2020; Zhang *et al.*, 2023). For instance, the International Association of Hydrological Science (IAHS) recognized the importance of this issue and designated the decade of 2003-2012 for discussing and studying PUB (Sivapalan *et al.*, 2003; Hrachowitz *et al.*, 2013). Since then, several methods of PUB, ranging from empirical to physically-based have been developed (Arsenault *et al.*, 2023), given the advances in process understanding, data-processing capability and the greater availability of hydroclimatic time series (Sivapalan *et al.*, 2003).

Currently, one of the most used PUB methods relies on data-driven approaches, which consist in calibrating the model with an observed dataset and further applying the calibrated model to predict runoff at ungauged basins, operating as a “black box”, and not based on hydrology principles in the basin. This approach encompasses the machine learning techniques such as artificial neural networks and support vector machines, and has seen significant improvements in prediction accuracy over the past few decades (Mosavi, Ozturk and Chau, 2018; Samantaray *et al.*, 2022). Nevertheless, these computational methods might be of difficult replication and understanding for who is interested in predict runoff in a simple and fast process (Shalamu, 2009; Yaseen *et al.*, 2016; Seo, Kim and Singh, 2018). Another widely used PUB method, the physically-based models, might help the understanding of relations between the hydrological entities (VanderKwaak and Loague, 2001), but also employs complex physical equations and parametric assumptions that hampers their use in a simple way (Yaseen *et al.*, 2016).

Throughout the late 1800’s and 1900’s, several empirical equations relating rainfall and runoff were developed in order to support the design of hydraulic infrastructures (Eview and Linsley, 1967). These approaches rely on the use of catchments' parameters such as mean slope, mean annual temperature and drainage area, besides precipitation, (Justin, 1914; Vogel, Wilson and Daly, 1999) to predict runoff. The Civil Engineer Carl E. Grunsky, president of the American Society of Civil Engineers - ASCE in 1924, developed one of these empirical methods based on data collected from catchments in San Francisco Peninsula in California, United States, which consists on a 1-variable functional form that relates mean annual precipitation to runoff (Grunsky and Manson, 1908; Grunsky *et al.*, 1915).

Grunsky's original framework remained untouched until the development of a recent study that employed a generalized formula specifically designed for Mediterranean-type climate basins in California, Portugal, and France (Santos and Hawkins, 2011). This study produced promising outcomes, including a good performance ($r^2 = 0.89$) in long-term streamflow prediction of Portugal watersheds, and a robust correlation observed between the generalized formula and mean annual temperature. Marchezepe *et al.* (2023) analyzed the generalized Grunsky's method in 716 Brazilian catchments, reporting for interannual regionalized streamflow predictions, a median percentage bias and Kling-Gupta Efficiency of -1% and 0.73, respectively. Futhermore, more than 83% of the total studied basins had at least one month with KGE greater than 0.50 on a montly scale.

The results gathered so far highlighted the easiness and good performance of the Grunsky method for predicting runoff. While the Grunsky approach is highly applicable because it relies solely on precipitation data, there are currently other variables with abundant data availability (e.g. temperature) that could be incorporated into the formulation to improve its performance. Here, we seek to answer the following questions: Can Grunsky's Method be improved by adding temperature data? How much is the improvement? How can these variables be integrated into the main framework without deviating from the original objective of predicting runoff in a quick and straightforward manner?

In this study, we propose an improved version of the Grunsky's equation which encompasses/uses temperature data besides precipitation and investigate its performance in computing monthly and interannual streamflow. We applied the generalized Grunsky method for 716 Brazilian catchments available in the Catchments Attributes for Brazil (CABra) dataset (Almagro *et al.*, 2020). We organized this study as follows: in section 4, we describe the used dataset, presents Grunsky's method and the parameters used for its improvement, on an interannual and monthly scales. In section 5, we present and discuss the results, respectively and, in section 7, we highlight the main conclusions of the paper.

4. Material and methods

4.1. Dataset

We used the CABra dataset, which comprises several climatological and topographic attributes for 735 catchments in Brazil (Almagro *et al.*, 2020). The catchment attributes are divided into eight categories, which include topography, climate, streamflow, groundwater, soil, geology, land-use and land-cover, and hydrologic disturbance. In addition to these static attributes, the dataset also comprises 30-year daily records of precipitation, streamflow, and both actual and potential evapotranspiration from 1980 to 2010. The data for each catchment contains less than 10% of missing information. Here, we used the CABra's daily time series of precipitation (P) and streamflow (Q) between October 1st, 1980 and September 30th, 2010, covering 30 hydrological years. P was derived from an ensemble mean between a high-resolution ground-based reference dataset (Xavier, King and Scanlon, 2016) and the ERA5 reanalysis dataset (Hersbach *et al.*, 2020b), while Q was based on streamflow gauge observations from ANA, the National Water Agency (available at http://www.snirh.gov.br/hidroweb/). The daily values were converted to annualized (monthly) series, resulting in 30 (360) annual (monthly) runoff and precipitation values for each catchment. To maintain the consistence of annual and monthly values within the same catchment, years (months) with at least one day of missing data were excluded from the analysis. A total of 19 catchments were also excluded from the study because they exhibited inconsistent values of annual Real Evapotranspiration (E), Q and P (Q > P or E > Q). Consequently, the study used 716 catchments for analysis.

The Brazilian catchments were divided in six hydrological groups based on catchment hydrological signatures and attributes from the CABra dataset, analyzed with principal component analysis and random forest algorithms (Almagro, 2021). These hydrological groups are “Non-seasonal”, “Dry”, “Rainforest”, “Savannah”, “Extremely-dry”, and “Extremely-wet” catchments (Figure 1). We used these six groups for regionalization because they represent hydrological similarities better than other geographic, climate, or ecosystem groups. The groups are characterized as follows:

- Non-seasonal: high mean streamflow, no clear precipitation-temperature relationship, located mainly in the South region with Pampa, Atlantic Forest biomes, and some in the Amazon biome.
- Dry: well-defined rainy and dry seasons, highest temperatures, located mainly in the Cerrado biome between Amazon and Caatinga biomes, with great variance in aridity index and low values of mean precipitation.
- Rainforest: high amounts of precipitation, well-defined rainy season, located mainly in the Atlantic Forest biome region.

- Savannah: presents savannah characteristics, located mainly in the Cerrado biome with vegetation composed of grass and forest, aridity index varying from 1 to 2.
- Extremely-dry: located in the Caatinga biome, covered mainly by grass and shrubs, non-perennial streams, high aridity index.
- Extremely-wet: highest precipitation and streamflow values, located in the Amazon and Atlantic Forest biomes, covered mainly by forests.

4.2. Relashionship between α and catchments attributes

The study comprehended an exploratory analysis between the α coefficient, calculated from the generalized Grunsky's Equation described below, and some of the catchment's attributes in which may exert an important influence in the rainfall-runoff transformation. Equation 1 refers to the generalized Grunsky's quation between the mean annual runoff ($\bar{Q}$), and the mean annual precipitation ($\bar{P}$), measured in mm. Since there are only catchments with $Q_{obs} < P_{obs}$, α coefficient is positive in the annual analysis.

$$\begin{aligned} \bar{Q} &= \alpha\bar{P}^2, \quad \bar{P} < 1/(2\alpha) \\ \bar{Q} &= \bar{P} - \frac{1}{4\alpha}, \quad \bar{P} > 1/(2\alpha) \end{aligned} \tag{1}$$

The α coefficient was calculated using the Minimum Mean Square Error (MMSE) method, which is the same approach used by Marchezepe *et al.* (2023). This method selects an α value from the 30-year annual series that best fits the model for each catchment based on the KGE metrics.

To obtain a preliminary overview of the relationship between the α parameter – computed for all evaluated catchments – and catchments attributes, we computed the

Spearman's correlation between α and several catchments attributes, already available in the CABra dataset, and from this first analysis, we chose some of the higher correlations and previously recognized as the parameters that have stronger influence on Q in a catchment-scale: maximum, minimum and mean temperarature (Santos and Hawkins, 2011), aridity index (Meira Neto *et al.*, 2020; Ballarin *et al.*, 2022), and actual and potential evapotranspiration (ET and PET, respectively), calculated by Penman-Monteith (Allen et al., 1998), Priestley-Taylor (Priestley and Taylor, 1972), and Hargreaves (1975) methods.

4.3. Calculating α from a Multiple Linear Regression Method

After identifying the catchments' attributes well correlated with α, we fit a Multiple Linear Regression (MLR) considering the α as the dependent variable and mean precipitation ($\bar{P}_A$) and this specific catchment parameter as the independent variables ($\bar{X}_A$), generating an Equation 2 type equation, with A and B being the $\bar{X}_A$ and $\bar{P}_A$ coefficient, respectively, and C the constant term. This type of equation was used for mean annual and monthly scales, needing more studies for other temporal scales.

$$\alpha = \mathrm{A} \cdot \bar{X}_A + \mathrm{B} \cdot \bar{P}_A + C \tag{2}$$

We evaluate the performance of the MLR models in predicting α by using different metrics in a leave-one-out cross validation scheme. This approach involved generating a regression equation for each catchment by using data from all the other catchments except the one being analyzed. Then, we predicted the α coefficient for the excluded catchment and used Grunsky's generalized equation to obtain a predicted runoff. This process was repeated for all catchments, and the predicted runoff values were compared to the observed runoff values at two different time scales: mean annual and mean monthly values. The analysis method can be visualized on Figure 11.

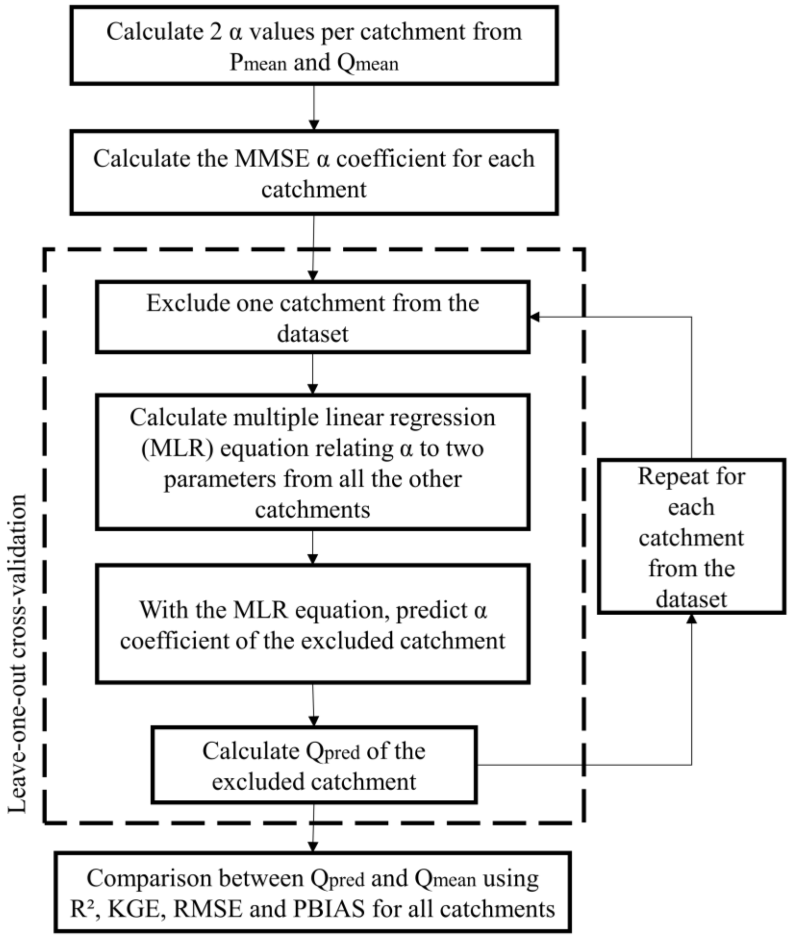

Figure 11: Schematic view of the analysis method.

4.4. Performance Evaluation of streamflow prediction obtained by using the proposed approach

Four widely used statistical metrics were employed to evaluate the model performance, namely: the coefficient of determination (R^2), the Kling-Gupta Efficiency (KGE) (Gupta *et al.*, 2009), the Percent Bias (PBIAS), and the Root Mean Square Error (RMSE). R^2 and KGE values close to 1 indicate better model performance, positive (negative) PBIAS represents overestimation (underestimation) in predictions, and lower RMSE values imply reduced errors.

5. Results and discussion

The results are presented in two steps: (i) we first assess the relationship between α coefficient and climate parameters, and choose the parameter that best correlates with α coefficient. (ii) Then we evaluate the performance of an improved Grunsky's framework, proposed in the present study, on an annual and monthly time scales for all catchments, by calculating α coefficient from a multiple linear regression equation using mean temperature.

5.1. Relashionship between α and catchments characteristics

The Spearman correlation between annual α coefficient with the selected catchments parameters is displayed in Figure 2, which also includes the inter correlation between the selected parameters. The correlation values between α and catchments attributes ranged from -0.5 to -0.57 with α and all parameters, excepting actual evapotransporation. Since potential evapotranspiration, temperature and aridity index are strictly related, these results were already expected.

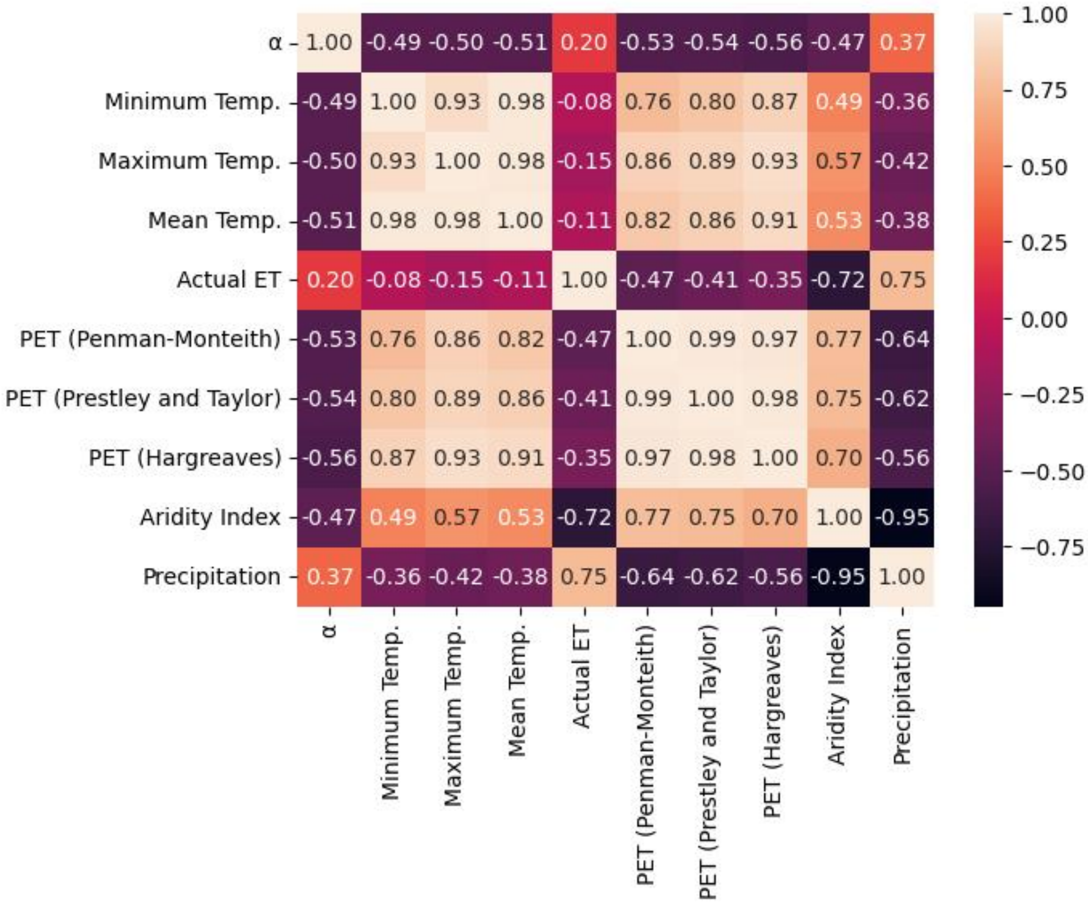

Figure 12: Spearman correlation matrix between alpha and climatic parameters of CABra dataset.

By analyzing scatterplots from Figure 13, it can be seen that temperature, PET and aridity index present negative correlation with α. From Equation 1, we observe that α presents a positive correlation with precipitation, which is negatively correlated with PET, and therefore, with mean temperature, according to the Penman-Monteith framework (Zotarelli *et al.*, 2014), and also, by definition, inversely proportional to aridity index (Budyko, 1974). Especially, mean temperature and potential annual evapotranspiration by Penman-Monteith equation present a similar scatter pattern, also demonstrated by Figure 12, where these two parameters have a positive correlation higher than 0.8. However, the aridity index plot (Figure 13b) showed a distinct distribution, due to a lower absolute correlation with α and the other parameters. One of the reasons might be that aridity index associates potential evapotranspiration and precipitation, which adds more variability to the scatter, since α is not strongly correlated with precipitation from the Spearman correlation observed on Figure 12.

Among the hydrological groups, the Dry and Extremely-dry groups exhibit the lowest values of α, suggesting that catchment attributes significantly influence the conversion of precipitation into runoff. This influence, here represented by α, is also manifested by the highest values of mean temperature, PET and Aridity Index for these groups, meaning that lower α indicates that precipitation has a stronger tendency to losses before turning into runoff, in opposition to Non-seasonal and Rainforest groups, with higher α.

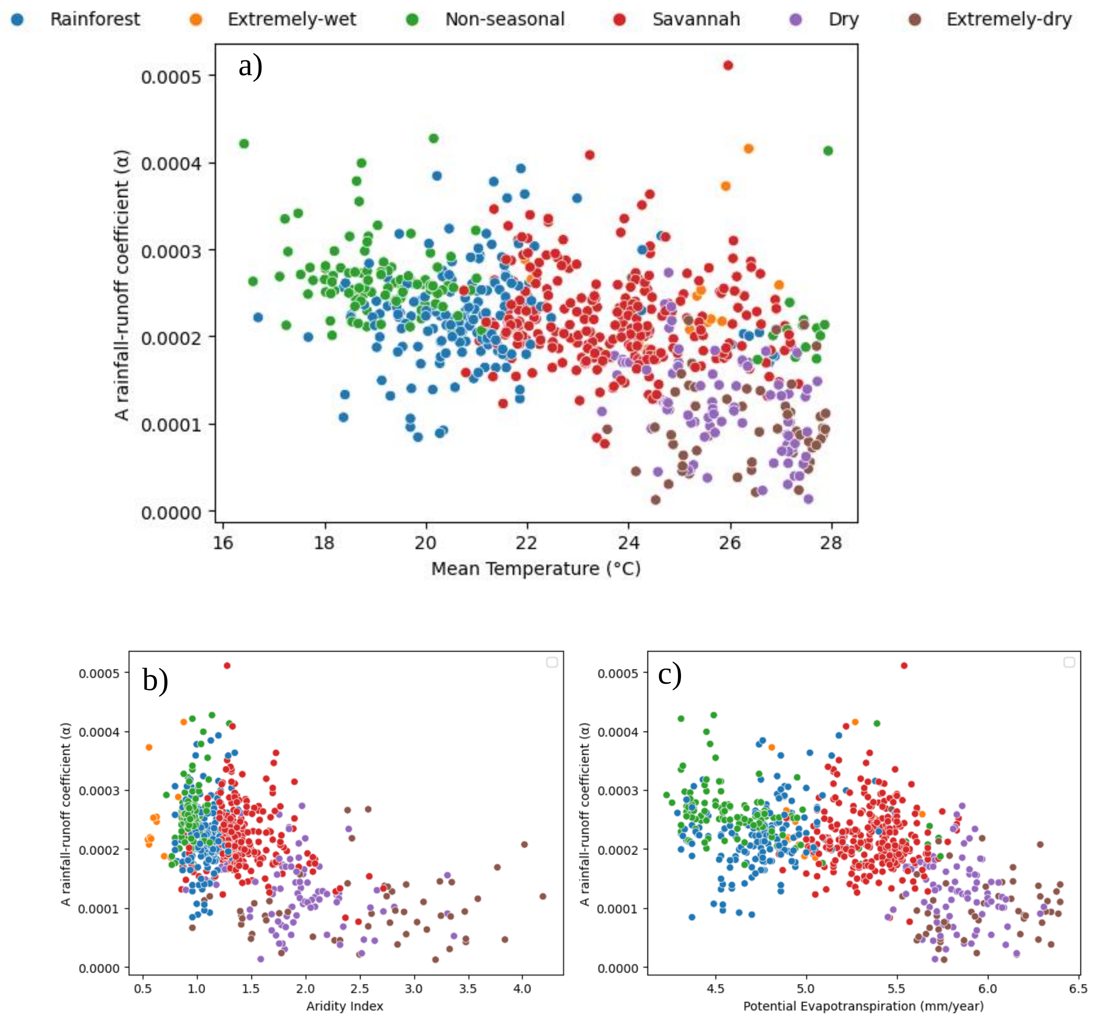

Figure 13: Scatterplots of alpha coefficient and a) Mean Temperature (°C), b) Aridity Index, and c) Potential Annual Evapotranspiration by Penman-Monteith equation.

As long as this study aims to search the most simple possible method to find runoff from an ungauged catchment's parameters, and since potential evapotranspiration and mean temperature are highly correlated, we chose the mean temperature as the center parameter of the analysis, on an annual and monthly scales, taking into account this being a direct and simpler parameter regarding quantification over the global atmosphere (Morice *et al.*, 2012; Xavier, King and Scanlon, 2016).

On a montly scale, as described in Marchezepe *et al.* (2023), the driest months of the year, i.e. from May to August, there is a higher variation of α values, and even negative results for Rainforest, Savannah and Dry groups, implying on $Q > P$ for these months through the historical series, implying that the runoff is influenced by groundwater for these months. These results put in evidence that, during these drier months, runoff receives relatively more contribution from groundwater, especially for these groups, also noted by Almagro (2021). The other groups did not present this behavior, suggested by the inexistence of dry/wet seasons for Non-seasonal group, the water scarcity on Extremely-dry catchments, leading to non-perennial streams and streamflow influenced by surface flow, and high precipitation rates throughout the year on Extremely-wet groups (Almagro, 2021). The impact of seasonality in predictions might request a proper approach for these months and groups, taking into account the groundwater effect on runoff. The wetter months presented a similar scatter behavior as the annual scale, with a negative correlation between mean temperature and α coefficient.

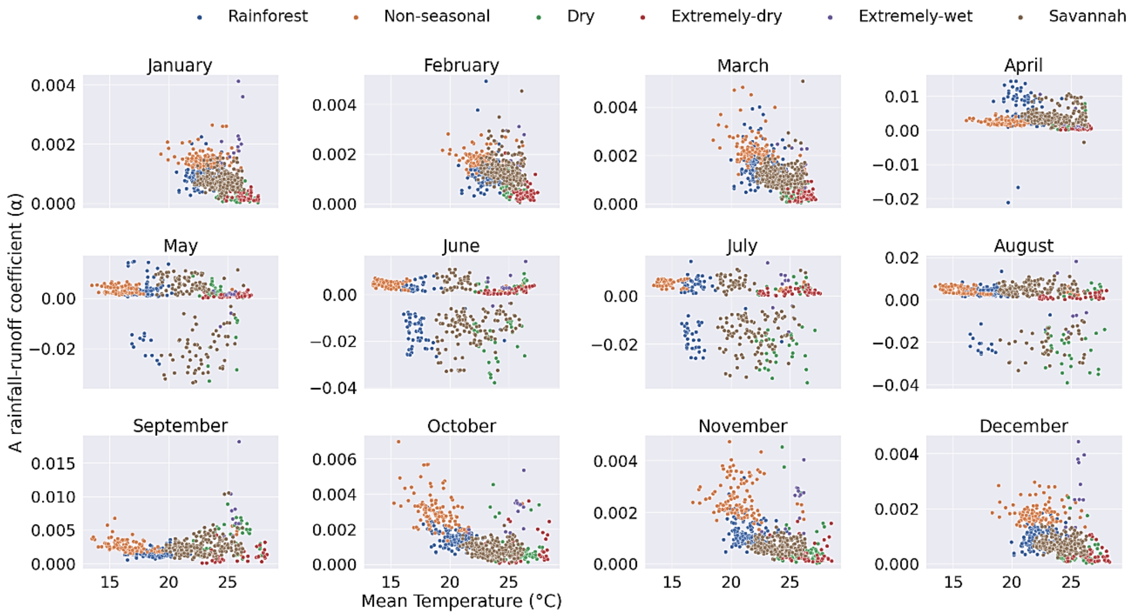

Figure 14: Scatterplot of alpha coefficient versus mean monthly temperature (°C) by group.

5.2. Calculating α from a Multiple Linear Regression Method

Since mean temperature is the chosen focus parameter of the analysis, we computed a multiple linear regression equation, relating α coefficient to the Mean Annual Temperature ($\bar{T}_A$), and Mean Annual Precipitation ($\bar{P}_A$) from all 716 catchments, generating an Equation 2 type equation. The equations were calculated by the leave-one-out cross validation.

The calculated multiple linear regression model for all catchments on an annual scale resulted on Equation 3.

$$\alpha = -1.102222\mathrm{e}-05 \cdot \bar{T}_A + 4.850557\mathrm{e}-08 \cdot \bar{P}_A + 3.819316e-04 \quad (3)$$

From the same process for all catchments, the observed and predicted annual runoff values were compared, presenting KGE = 0.9, R^2 = 0.82 and RMSE = 27.3% (Figure 15). These inputs represent an increase in performance of the original method from Marchezepe *et al.* (2023), that presented KGE = 0.6 to 0.84, R^2 = 0.49 to 0.74 and

RMSE = 16.5% to 87.8% for annual scale and each group. Results show similar performance than the water balance modeling based on the Budyko framework from Zhang *et al.* (2008), that showed a coefficient of efficiency of 0.93, and from Samantaray *et al.* (2022), that reached $R^2 = 0.97$ for monthly scale predictions using machine learning techniques for monthly scales using rainfall, temperature and humidiy. Also, results are better than studies such as from Singh *et al.* (2022) that use machine learning techniques for regionalization in ungauged basins ($R^2 = 0.7$).

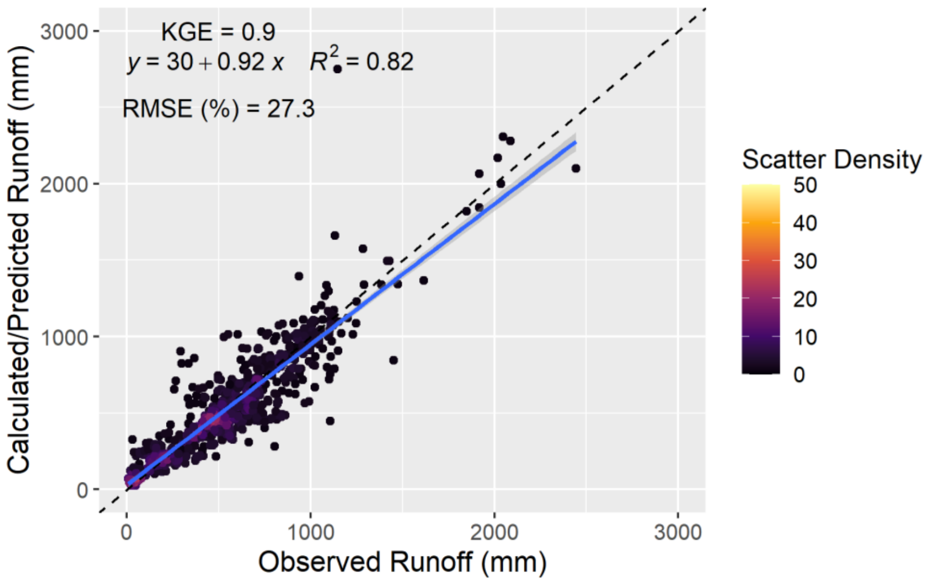

Figure 15: Observed runoff versus predicted runoff on an annual scale.

The same analysis method using leave-one-out cross validation was used for the monthly scale analysis (Figure 16). The results put in evidence the previous observations of higher variation mainly on drier months (May to August), when runoff remains supplied by groundwater, and exceed precipitation, leading to negative values of α, and decreasing prediction performance for these months. On the other hand, even though wetter months performed relatively better, results do not follow the mean annual predictions, indicating great intra-annual influence of different factors other than simply mean temperature and precipitation. This difference between wet and dry seasons predictions was already observed for Southeast Brazil (Zhang *et al.*, 2018).

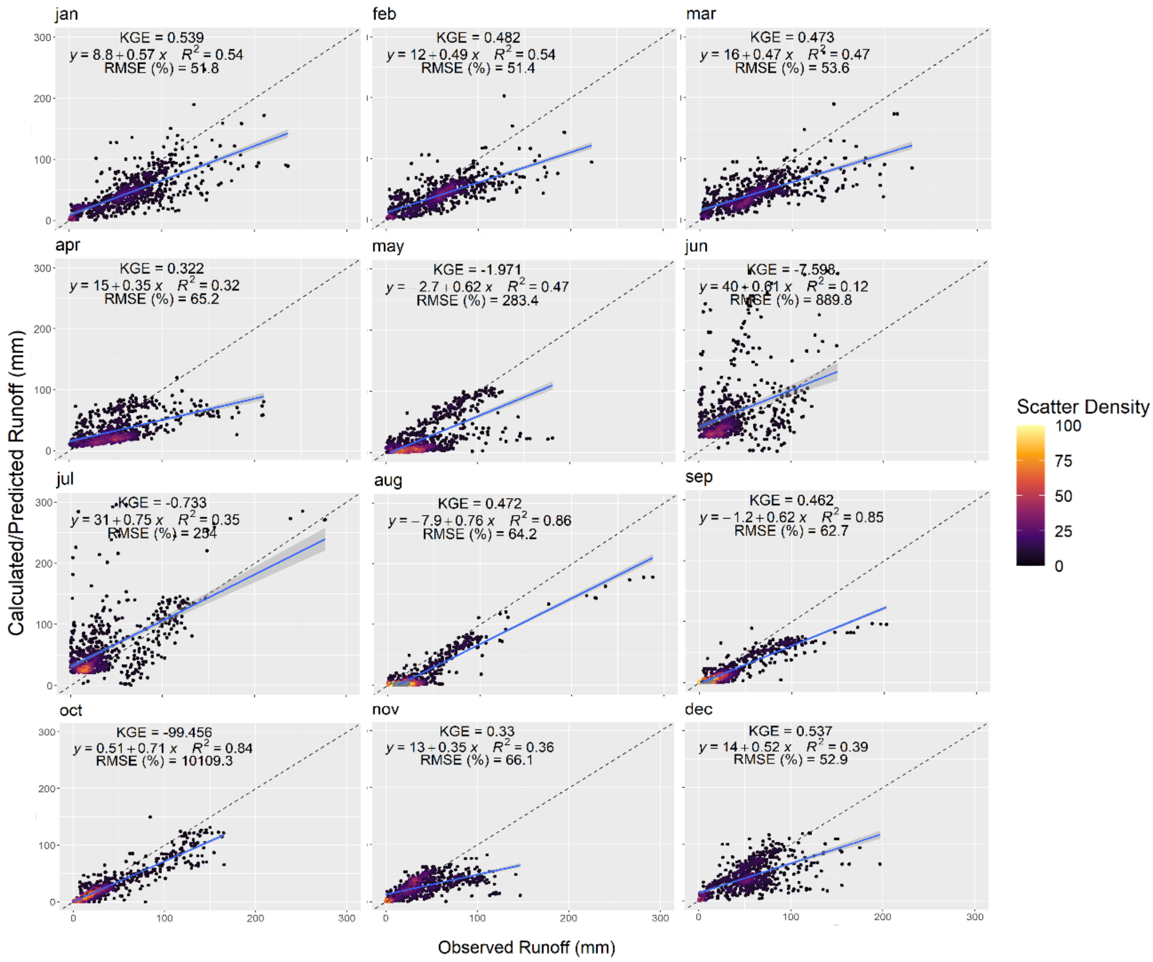

Figure 16: Observed runoff versus predicted runoff on a monthly scale.

The equations coefficients and performance indicadors for each month are described on Table 2. One can note adverse values of these indicators, mainly for October, taking into account that there was no exclusion of outliers in the analysis. There were counted data with discrepant values from the other catchments.

Table 2: Multiple linear regression equation coefficients and performance metrics for each month.

Month	A	B	C	KGE	RMSE	R^2
1	-1.550252e-04	-9.977739e-08	0.004542756	0.539	51.8	0.54
2	-1.946228e-04	-7.198937e-07	0.006048492	0.482	51.4	0.54
3	-2.548126e-04	5.026635e-07	0.007223951	0.473	53.6	0.47
4	-1.427797e-04	-6.749037e-06	0.006624364	0.322	65.2	0.32
5	-2.176534e-04	-1.334365e-05	0.009476049	-1.971	283.4	0.23
6	9.903585e-05	1.116949e-04	-0.014463027	-7.615	891.4	0.03
7	3.888679e-04	1.546483e-04	-0.021028256	-0.733	254.0	0.11
8	-9.213360e-05	-4.612289e-06	0.006441633	0.472	64.2	0.86
9	6.001941e-06	-1.639433e-07	0.001965425	0.462	62.7	0.85
10	-1.103967e-04	7.255341e-06	0.002810552	-98.812	10045.1	0.01
11	-1.780340e-04	-2.940251e-06	0.005529575	0.330	66.1	0.36
12	-1.823458e-04	-2.493800e-06	0.005695633	0.537	52.9	0.39

6. Limitations and broader implications

This paper has demonstrated the usefulness of the improved Grunsky method on an annual scale, especially. On a monthly scale, there were better results for wetter months. Broader studies are required for the performance improvement in drier catchments, such as using the direct runoff instead of total runoff, which might improve the predictability since direct runoff is closely related to precipitation events and predictions performances are influenced by dry and wet seasons (Zhang *et al.*, 2018; Meira Neto *et al.*, 2020).

The analysis methods employed, due to their relative easy application, have great replication capacity. That is, they may be used for a large database of hydrographic basins, as is the case with CABra, and can be updated in the future, as needed. Through the CABra database, it is possible to analyze the influence of other environmental factors on runoff response to precipitation, especially the values of precipitation and flow in monthly scale.

7. Conclusions

This study aims to improve the Grunsky's method, which predicts runoff in ungauged basins from values of precipitation associated with other parameter, resulting in a simple equation with better performance than the original method. From this analysis, some conclusions deserve special attention and are described below.

First, it can be said that the proposed method improves Grunsky's generalized framework by adding mean temperature to a multiple linear regression equation to predict α coefficient. Performance indicators such as KGE, R^2 and RMSE show a better performance than the original method.

Secondly, time scale do influence the method performance, as already mentioned for the original method and confirmed by the present study. On a monthly scale, results show a high variability of α and, consequently, of predicted runoff, leading to lower performance indicators than for mean annual scale. This variability might be a consequence of baseflow influence on intra-annual scales, since the highest variabilities are presented on the driest months, when total runoff exceeds total precipitation, indicating a high influence of baseflow, and leading to negative α coefficients.

At last, it is noteworthy that the improved method could support fast decision-making processes, and hydraulic projects, from irrigation to dam constructions, and generate permanence flows for licensing projects. The multiple linear regression equation is one of the fastest methods to calculate a parameter from independent variables. In addition, these variables, mean temperature and mean precipitation, are widely used and available from many sources, such as satellite images and gauge stations (Morice *et al.*, 2012; Xavier, King and Scanlon, 2016).

8. Acknowledgements

The authors acknowledge the Graduate Program in Hydraulic Engineering and Sanitation – PPGSHS (USP-EESC) and Graduate Program in Environmental Technologies – PPGTA (UFMS-FAENG) for the scientific support. This study was supported by grants from the Coordination for the Improvement of Higher Education Personnel (CAPES) - Finance Code 001 and CAPES PrInt, the Ministry of Science, Technology, and Innovation (MCTI) and National Council for Scientific and Technological Development (CNPq) (grants numbers 441289/2017-7 and 309752/2020-5), and São Paulo Research Support Foundation (FAPESP, grants numbers 2015/03806-1, 2019/24292-7, and 2020/08140-0).

9. References

Agarwal, A. *et al.* (2016) 'Hydrologic regionalization using wavelet-based multiscale entropy method', *Journal of Hydrology*, 538, pp. 22–32. doi: 10.1016/J.JHYDROL.2016.03.023.

Almagro, A. *et al.* (2020) 'CABra: a novel large-sample dataset for Brazilian catchments', *Hydrology and Earth System Sciences Discussions*, (October), pp. 1–40. doi: 10.5194/hess-2020-521.

Almagro, A. (2021) *Towards a better understanding of catchment hydrology in Brazil*. Federal University of Mato Grosso do Sul. Available at: https://repositorih.ufms.br/handle/123456789/4035.

ANA (2007) *Diagnóstico da outorga de direito de uso de recursos hídricos no Brasil. Caderno de Recursos Hídricos. Volume 4. Agência Nacional de Águas.* Brasília. Available at: www.ana.gov.br.

Arheimer, B. *et al.* (2020) 'Global catchment modelling using World-Wide HYPE (WWH), open data, and stepwise parameter estimation', *Hydrology and Earth System Sciences*, 24(2), pp. 535–559. doi: 10.5194/HESS-24-535-2020.

Arsenault, R. *et al.* (2023) 'Continuous streamflow prediction in ungauged basins: long short-term memory neural networks clearly outperform traditional hydrological models', *Hydrol. Earth Syst. Sci*, 27, pp. 139–157. doi: 10.5194/hess-27-139-2023.

Athira, P. *et al.* (2016) 'Predictions in ungauged basins: an approach for regionalization of hydrological models considering the probability distribution of model parameters', *Stochastic Environmental Research and Risk Assessment*, 30(4), pp. 1131–1149. doi: 10.1007/S00477-015-1190-6.

Ballarin, A. S. *et al.* (2022) 'The Impact of an Open Water Balance Assumption on Understanding the Factors Controlling the Long-Term Streamflow Components', *Water Resources Research*, 58(10). doi: 10.1029/2022WR032413.

Beck, H. E., de Roo, A. and van Dijk, A. I. J. M. (2015) 'Global Maps of Streamflow Characteristics Based on Observations from Several Thousand Catchments', *Journal of Hydrometeorology*, 16(4), pp. 1478–1501. doi: 10.1175/JHM-D-14-0155.1.

Besaw, L. E. *et al.* (2010) 'Advances in ungauged streamflow prediction using artificial neural networks', *Journal of Hydrology*, 386(1–4), pp. 27–37. doi: 10.1016/J.JHYDROL.2010.02.037.

Beskow, S. *et al.* (2016) 'Potential of the LASH model for water resources management in data-scarce basins: a case study of the Fragata River basin, southern Brazil', *https://doi.org/10.1080/02626667.2015.1133912*, 61(14), pp. 2567–2578. doi: 10.1080/02626667.2015.1133912.

Bloschl, G. *et al.* (eds) (2013) *Runoff Prediction in Ungauged Basins*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9781139235761.

Blöschl, G. *et al.* (2019) 'Twenty-three unsolved problems in hydrology (UPH) – a community perspective', *https://doi.org/10.1080/02626667.2019.1620507*, 64(10), pp. 1141–1158. doi: 10.1080/02626667.2019.1620507.

Bork, C. K. *et al.* (2021) 'Minimum streamflow regionalization in a Brazilian watershed under different clustering approaches', *Anais da Academia Brasileira de Ciências*, 93(suppl 4). doi: 10.1590/0001-3765202120210538.

Budyko, M. I. (1974) *Climate and Life, New York, NY: Academic Press.*

Choi, J. *et al.* (2021) 'Usefulness of Global Root Zone Soil Moisture Product for Streamflow Prediction of Ungauged Basins', *Remote Sensing 2021, Vol. 13, Page 756*, 13(4), p. 756. doi: 10.3390/RS13040756.

Dal Molin, M. *et al.* (2020) 'Understanding dominant controls on streamflow spatial variability to set up a semi-distributed hydrological model: The case study of the Thur catchment', *Hydrology and Earth System Sciences*, 24(3), pp. 1319–1345. doi: 10.5194/HESS-24-1319-2020.

Du, T. L. T. *et al.* (2020) 'Streamflow prediction in "geopolitically ungauged" basins using satellite observations and regionalization at subcontinental scale', *Journal of Hydrology*, 588, p. 125016. doi: 10.1016/J.JHYDROL.2020.125016.

Eview, R. and Linsley, A. K. (1967) 'The relation between rainfall and runoff: Review paper', *Journal of Hydrology*, 5, pp. 297–311. doi: 10.1016/S0022-1694(67)80128-8.

de Figueiredo, J. V. *et al.* (2016) 'Runoff initiation in a preserved semiarid Caatinga small watershed, Northeastern Brazil', *Hydrological Processes*, 30(13), pp. 2390–2400. doi: 10.1002/HYP.10801.

Gao, Y. *et al.* (2021) 'Diagnosis toward predicting mean annual runoff in ungauged basins', *Hydrol. Earth Syst. Sci*, 25, pp. 945–956. doi: 10.5194/hess-25-945-2021.

Grunsky, C. E. (1908a) ‘Closure to Rain and Run-Off Near San Francisco’, *Transactions of the American Society of Civil Engineers*, 61(2), pp. 539–543. doi: 10.1061/TACEAT.0002035.

Grunsky, C. E. (1908b) ‘Rain and Run-Off Near San Francisco, California’, *Transactions of the American Society of Civil Engineers*, 61(2), pp. 496–515. doi: 10.1061/TACEAT.0002036.

Grunsky, C. E. *et al.* (1915) ‘Discussion: Run-Off from Rainfall and Other Data’, *Transactions of the American Society of Civil Engineers,* 79(2), pp. 1156–1208. doi: 10.1061/TACEAT.0002701.

Grunsky, C. E. and Manson, M. C. E. (1908) *Reports on the water supply of San Francisco, California, 1900 to 1908, inclusive.* Transactions of the American Society of Civil Engineers. Available at: http://hdl.handle.net/2027/uc1.$b78665 (Accessed: 13 March 2022).

Gupta, H. V. *et al.* (2009) ‘Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling’, *Journal of Hydrology*, 377(1–2), pp. 80–91. doi: 10.1016/J.JHYDROL.2009.08.003.

Hersbach, H. *et al.* (2020a) ‘The ERA5 global reanalysis’, *Quarterly Journal of the Royal Meteorological Society*, 146(730), pp. 1999–2049. doi: 10.1002/qj.3803.

Hersbach, H. *et al.* (2020b) ‘The ERA5 global reanalysis’, *Quarterly Journal of the Royal Meteorological Society*, 146(730), pp. 1999–2049. doi: 10.1002/QJ.3803.

Honório, M. D. S. *et al.* (2020) ‘Comparison of two methods for determining Q95reference flow in the mouth of the surface catchment basin of the Meia Ponte river, state of Goiás, Brazil’, *Open Engineering*, 10(1), pp. 469–476. doi: 10.1515/ENG-2020-0053/MACHINEREADABLECITATION/RIS.

Hrachowitz, M. *et al.* (2013) ‘A decade of Predictions in Ungauged Basins (PUB)-a review’, *Hydrological Sciences Journal*. Taylor & Francis, pp. 1198–1255. doi: 10.1080/02626667.2013.803183.

Hubbart, J. A. (2011) ‘Origins of Quantitative Hydrology: Pierre Perrault, Edme Mariotte, and Edmond Halley’, *Journal of the American Water Resources Association*, 13(6), pp. 15–17.

Justin (1914) ‘Derivation of Run-Off from Rainfall Data’, *Transactions of the American Society of Civil Engineers*, 77(1). Available at: https://doi.org/10.1061/taceat.0002572.

Knoben, W. J. M., Freer, J. E. and Woods, R. A. (2019) ‘Technical note: Inherent benchmark or not? Comparing Nash-Sutcliffe and Kling-Gupta efficiency scores’, *Hydrology and Earth System Sciences*. doi: 10.5194/hess-2019-327.

Kratzert, F. *et al.* (2019) ‘Toward Improved Predictions in Ungauged Basins: Exploiting the Power of Machine Learning’, *Water Resources Research*, 55(12), pp. 11344–11354. doi: 10.1029/2019WR026065.

Liu, Y. *et al.* (2020) ‘What is the hydrologically effective area of a catchment?’, *Environmental Research Letters*, 15(10). doi: 10.1088/1748-9326/aba7e5.

Marchezepe, B. K. *et al.* (2023) ‘Streamflow prediction in ungauged Brazilian catchments by using the Grunsky Method’, *[Manuscript submitted for publication]. School of Engineering of São Carlos, University of São Paulo.*

McDonnell, J. J. (2003) ‘Where does water go when it rains? Moving beyond the variable source area concept of rainfall-runoff response', *Hydrological Processes*, 17(9), pp. 1869–1875. doi: 10.1002/HYP.5132.

Meira Neto, A. A. *et al.* (2020) 'An Aridity Index-Based Formulation of Streamflow Components', *Water Resources Research*, 56(9), pp. 1–14. doi: 10.1029/2020WR027123.

Montanari, A. *et al.* (2013) '"Panta Rhei—Everything Flows": Change in hydrology and society—The IAHS Scientific Decade 2013–2022', *https://doi.org/10.1080/02626667.2013.809088*, 58(6), pp. 1256–1275. doi: 10.1080/02626667.2013.809088.

Morice, C. P. *et al.* (2012) 'Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The HadCRUT4 data set', *J. Geophys. Res*, 117, p. 8101. doi: 10.1029/2011JD017187.

Mosavi, A., Ozturk, P. and Chau, K. W. (2018) 'Flood Prediction Using Machine Learning Models: Literature Review', *Water 2018, Vol. 10, Page 1536*, 10(11), p. 1536. doi: 10.3390/W10111536.

Nace, R. L. (1974) 'Pierre Perrault: the Man and His Contribution To Modern Hydrology', *JAWRA Journal of the American Water Resources Association*, 10(4), pp. 633–647. doi: 10.1111/j.1752-1688.1974.tb05623.x.

Parajka, J. *et al.* (2013) 'Comparative assessment of predictions in ungauged basins-Part 1: Runoff-hydrograph studies', *Hydrology and Earth System Sciences*, 17(5), pp. 1783–1795. doi: 10.5194/HESS-17-1783-2013.

Pinheiro, E. A. R. *et al.* (2016) 'Importance of soil-water to the Caatinga biome, Brazil', *Ecohydrology*. doi: 10.1002/eco.1728.

Piol, M. V. A. *et al.* (2019) 'Performance evaluation of flow duration curves regionalization methods', *Brazilian Journal of Water Resources*. doi: 10.1590/2318-0331.241920170202.

Pomeroy, J. W. *et al.* (2007) 'The cold regions hydrological model: A platform for basing process representation and model structure on physical evidence', *Hydrological Processes*, 21(19), pp. 2650–2667. doi: 10.1002/HYP.6787.

Pool, S., Vis, M. and Seibert, J. (2021) 'Regionalization for Ungauged Catchments — Lessons Learned From a Comparative Large-Sample Study', *Water Resources Research*, 57(10), p. e2021WR030437. doi: 10.1029/2021WR030437.

Popper, K. (2002) *The Logic of Scientific Discovery*. Available at: http://books.google.com/books?hl=en&lr=&id=T76Zd20IYlgC&pgis=1.

Razavi, T., Coulibaly, P. and Asce, M. (2012) 'Streamflow Prediction in Ungauged Basins: Review of Regionalization Methods', *Journal of Hydrologic Engineering*, 18(8), pp. 958–975. doi: 10.1061/(ASCE)HE.1943-5584.0000690.

Salinas, J. L. *et al.* (2013) 'Comparative assessment of predictions in ungauged basins-Part 2: Flood and low flow studies', *Hydrology and Earth System Sciences*, 17(7), pp. 2637–2652. doi: 10.5194/hess-17-2637-2013.

Samantaray, S. *et al.* (2022) 'Monthly runoff prediction at Baitarani river basin by support vector machine based on Salp swarm algorithm', *Ain Shams Engineering Journal*, 13(5), p. 101732. doi: 10.1016/J.ASEJ.2022.101732.

Santos, F. L. and Hawkins, R. H. (2011) 'Generalized Mediterranean Annual Water

Yield Model: Grunsky's Equation and Long-Term Average Temperature', *Journal of Hydrologic Engineering*, 16(11), pp. 874–879. doi: 10.1061/(ASCE)HE.1943-5584.0000402.

Seo, Y., Kim, S. and Singh, V. P. (2018) 'Machine Learning Models Coupled with Variational Mode Decomposition: A New Approach for Modeling Daily Rainfall-Runoff', *Atmosphere 2018, Vol. 9, Page 251*, 9(7), p. 251. doi: 10.3390/ATMOS9070251.

Shalamu, A. (2009) 'Monthly and seasonal streamflow forecasting in the Rio Grande Basin', *2009*, (August), p. 289.

Silva, W. M. and Simões, S. J. C. (2014) 'Spatial Intra-Annual Variability of Precipitation Based on Geostatistics. A Case Study for the Paraıba Do Sul Basin, Southeastern Brazil', *International Journal of Geosciences*, 5, p. 408. doi: 10.4236/ijg.2014.54039.

Singh, L. *et al.* (2022) 'Streamflow regionalisation of an ungauged catchment with machine learning approaches', *Hydrological Sciences Journal*. doi: 10.1080/02626667.2022.2049271.

Sivapalan, M. *et al.* (2003) 'IAHS Decade on Predictions in Ungauged Basins (PUB), 2003-2012: Shaping an exciting future for the hydrological sciences', *Hydrological Sciences Journal*, 48(6), pp. 857–880. doi: 10.1623/hysj.48.6.857.51421.

De Souza Fraga, M. *et al.* (2022) 'Assessment of streamflow regionalization using interpolated and satellite-based precipitation: a case study in a tropical watershed at Brazil', *Modeling Earth Systems and Environment 2022*, pp. 1–14. doi: 10.1007/S40808-022-01383-W.

Spinoni, J. *et al.* (2015) 'Towards identifying areas at climatological risk of desertification using the Köppen-Geiger classification and FAO aridity index', *International Journal of Climatology*, 35(9), pp. 2210–2222. doi: 10.1002/JOC.4124.

VanderKwaak, J. E. and Loague, K. (2001) 'Hydrologic-Response simulations for the R-5 catchment with a comprehensive physics-based model', *Water Resources Research*, 37(4), pp. 999–1013. Available at: https://onlinelibrary.wiley.com/doi/full/10.1029/2000WR900272.

Viglione, A. *et al.* (2013) 'Comparative assessment of predictions in ungauged basins - Part 3: Runoff signatures in Austria', *Hydrology and Earth System Sciences*, 17(6), pp. 2263–2279. doi: 10.5194/HESS-17-2263-2013.

Vogel, R. M., Wilson, I. and Daly, C. (1999) 'Regional regression models of annual streamflow for the United States', *Journal of Irrigation and Drainage Engineering*, 125(3). doi: https://doi.org/10.1061/(ASCE)0733-9437(1999)125:3(148).

Xavier, A. C., King, C. W. and Scanlon, B. R. (2016) 'Daily gridded meteorological variables in Brazil (1980–2013)', *International Journal of Climatology*, 36(6), pp. 2644–2659. doi: 10.1002/JOC.4518.

Yang, X. *et al.* (2020) 'Dependence of regionalization methods on the complexity of hydrological models in multiple climatic regions', *Journal of Hydrology*, 582, p. 124357. doi: 10.1016/J.JHYDROL.2019.124357.

Yaseen, Z. M. *et al.* (2016) 'Stream-flow forecasting using extreme learning machines: A case study in a semi-arid region in Iraq', *Journal of Hydrology*, 542, pp. 603–614. doi: 10.1016/j.jhydrol.2016.09.035.

Zhang, K. *et al.* (2023) 'Streamflow Prediction in Poorly Gauged Watersheds in the United States Through Data-Driven Sparse Sensing', *Water Resources Research*, 59(4), p. e2022WR034092. doi: 10.1029/2022WR034092.

Zhang, L. *et al.* (2008) 'Water balance modeling over variable time scales based on the Budyko framework – Model development and testing', *Journal of Hydrology*, 360(1–4), pp. 117–131. doi: 10.1016/J.JHYDROL.2008.07.021.

Zhang, R. *et al.* (2018) 'Season-based rainfall–runoff modelling using the probability-distributed model (PDM) for large basins in southeastern Brazil', *Hydrological Processes*, 32(14), pp. 2217–2230. doi: 10.1002/HYP.13154.

Zotarelli, L. *et al.* (2014) 'Step by Step Calculation of the Penman-Monteith Evapotranspiration (FAO-56 Method). AE459', *Institute of Food and Agricultural Sciences. University of Florida*, pp. 1–14. Available at: http%22//edis.ifas.ufl.edu/ae459.

GENERAL CONCLUSIONS

Grunsky's Method proved to be a relatively reliable prediction method, given its simplicity and transparency with hydrological parameters.

In Chapter 1, we discussed the validity of the method for Brazilian catchments, proving its applicability even in non-Mediterranean climate, which was not experimented since the method was first proposed in the early 1900's.

In Chapter 2, we analyzed the introduction of other parameters to the method in order to improve it. The main parameter for this analysis was mean temperature, that was already proven to have high correspondence to precipitation and streamflow, and even proposed by (Santos and Hawkins, 2011).

As it is a matter of increasing relevance within the scope of climate change, the present study will be able to offer parameters that allow a succinct, but well-founded, assessment of the predictions of ungauged Brazilian catchments. This assessment, represents another tool in favor of better strategic planning for watershed management, engineering projects and environmental studies for issuance purposes.

The analysis methods employed, due to their relative ease of application, have great replication capacity. That is, they can be used for a large database of hydrographic basins, as is the case with CABra, and can be updated in the future, as needed. If necessary, it is also possible to complement and enrich the study through the adaptation of the methods already described, or further analysis of other characteristics of hydrographic basins, since the database has a great variety of data. Through the CABra database, especially the values of precipitation and flow in different time scales, it is possible to analyze the influence of environmental factors on runoff response to precipitation. It is expected, with the large number of hydrological basins analyzed, a study of greater reliability.

APPENDIX

Supporting Information for

Streamflow prediction in ungauged Brazilian catchments by using the Grunsky Method

Bruno K. Marchezepe[1], André Almagro[2], André S. Ballarin[1], and Paulo Tarso S. Oliveira[1,2]*

[1]Department of Hydraulics and Sanitary Engineering, São Carlos School of Engineering, University of São Paulo, São Carlos, SP, Brazil

[2]Faculty of Engineering, Architecture and Urbanism and Geography, Federal University of Mato Grosso do Sul, MS, 79070–900, Brazil.

1. Contents of this file

This supporting information contains additional figures that support the findings of the paper. First, Figure S17 presents a plot of Grunsky's method equation considering different α coefficient values. In Figure S18 we dispose boxplots of α coefficient values divided by the hydrological groups. Figure S19 provides the scatter plots of Annual Observed Precipitation (mm) and Runoff (mm) for each hydrological group. Figure S20 exhibits boxplots of α coefficient values by month. In Figure S21, we show the spatialization of MMSE α values on Brazilian territory in order to observe any spatial distribution of its values. Figure S22, Figure S23 and

The regionalization on a monthly scale was done by assuming one α coefficient for each group and month, from the median values from the catchments. The boxplots of KGE values for the 12 months divided by each group are presented in Figure 10. These α values and the boxplots of PBIAS values are exhibited in the Supplementary Material (Table S1 and Figure S12, respectively).

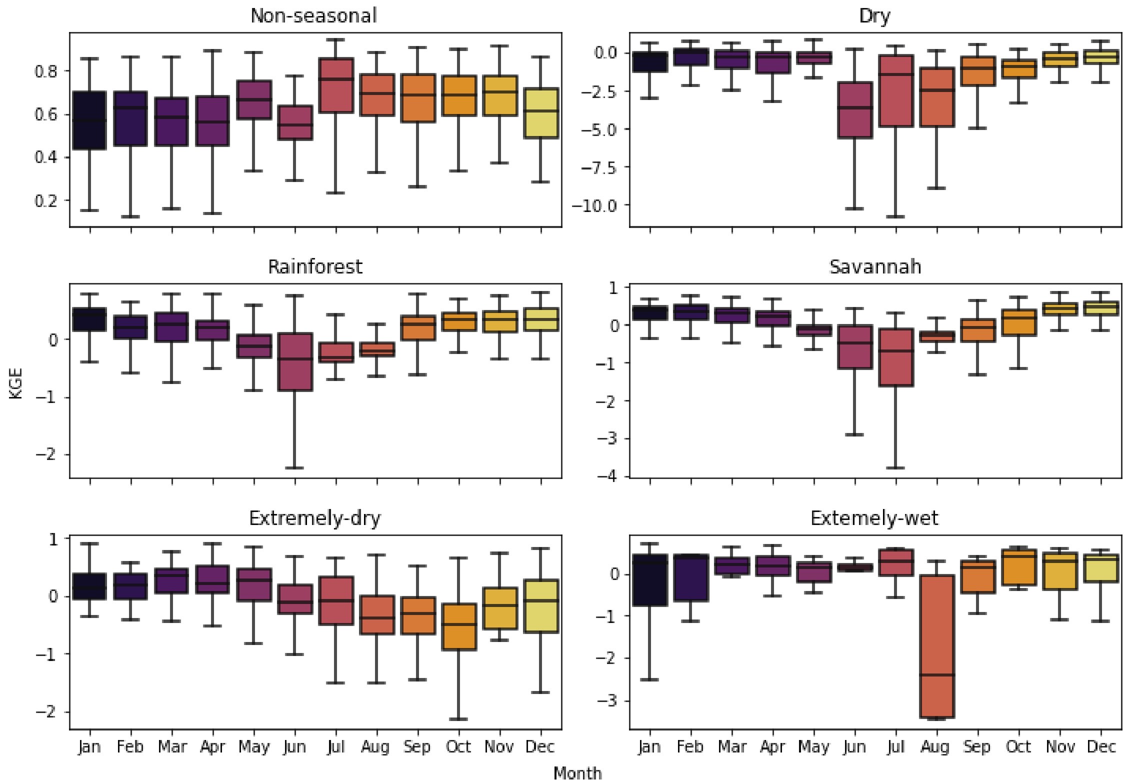

Figure 10: Boxplot with KGE from regionalized monthly runoff prediction for each month and group from the 716 CABra catchments.

The groups with the lowest values of PBIAS and highest values of KGE are Non-seasonal and Extremely-wet, given a more constant rainfall behavior on an intra-annual scale (Almagro, 2021). The worse performance was the Dry group, with median monthly KGE < 0 for all months, and PBIAS > 100% from June to August. The regionalization for a monthly scale appears to have worse results than the interannual scale for all groups, indicating the influence of interannual and intra-annual behavior of streamflow on the method’s performance. The results appear to be similar to the predictions by Arheimer *et al.*, 2020, which showed relatively good results in Brazilian North region catchments when predicting monthly runoff, with median monthly KGE > 0.60, and worse performance in the North-east region, with KGE < -1. Dry and Extremely-dry groups might have higher variance on an interannual scale, and particularly on an intra-annual scale due to the influence of soil water content, *and high* rates of evapotranspiration (Pinheiro *et al.*, 2016). display boxplots for mean, median and MMSE α values, respectively, for each month and hydrological group. **Erro! Fonte de referência não encontrada.** shows the spatialized KGE and PBIAS values from the local analysis on interannual scale. Figure S25 presents the spatial distribution for MMSE α value on the Brazilian territory for the monthly scale, and Figure S26, the respective PBIAS values. Table S1 contains the median values for MMSE α for each month and group. Figure S27 exhibits the boxplots with PBIAS from regionalized monthly runoff prediction for each month and group from the 716 CABra catchments.

2. Grunsky method plot

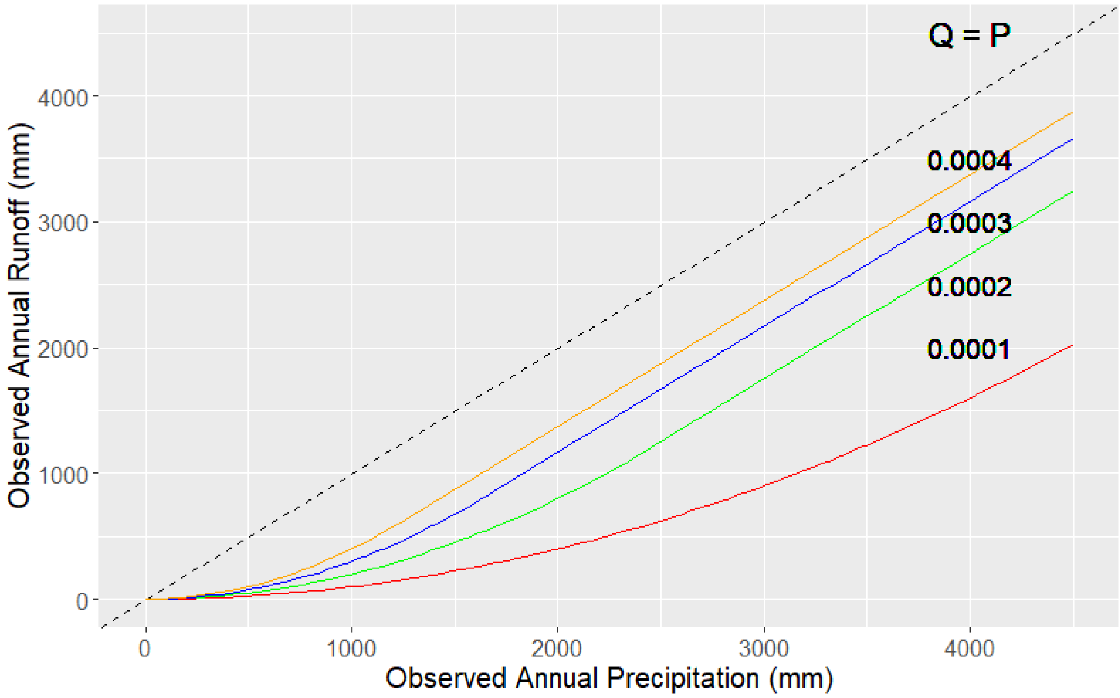

Figure S17: Grunsky's equation plot for various values of α (mm^{-1}), and lines with relations between $\bar{Q}$ (annual runoff) and $\bar{P}$ (annual precipitation).

3. Boxplots of α coefficient by group

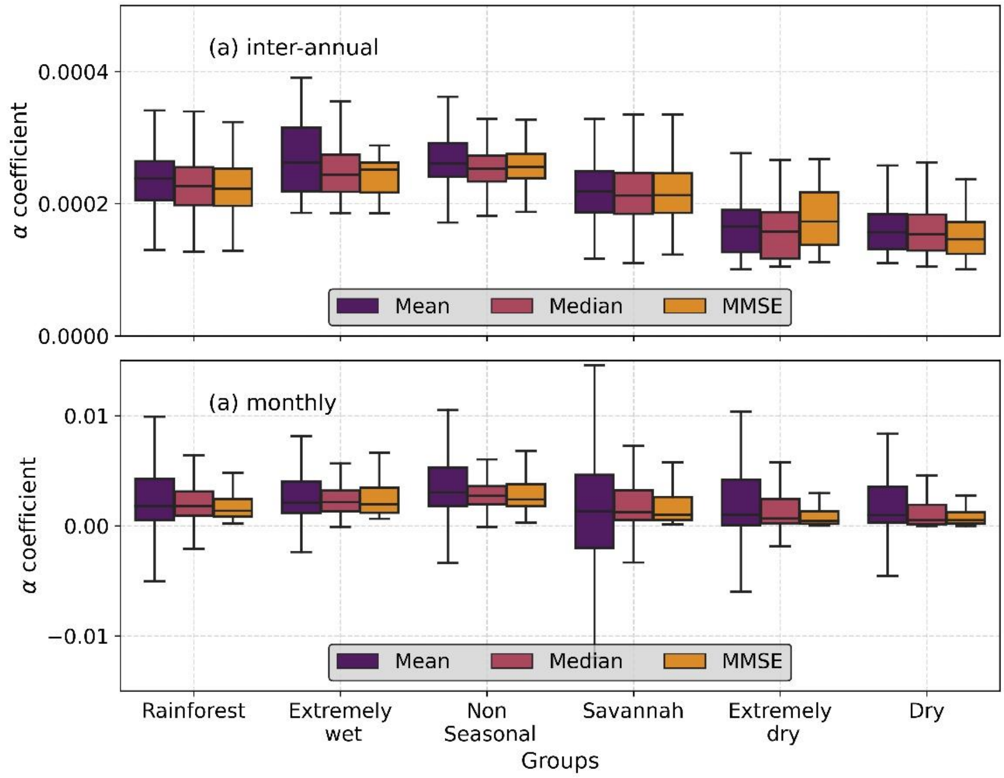

Figure S18: α coefficient of all 716 evaluated catchments, divided according to their hydrological groups for both (a) interannual and (b) monthly temporal scales.

4. Grunsky curve for hydrological groups on annual scale

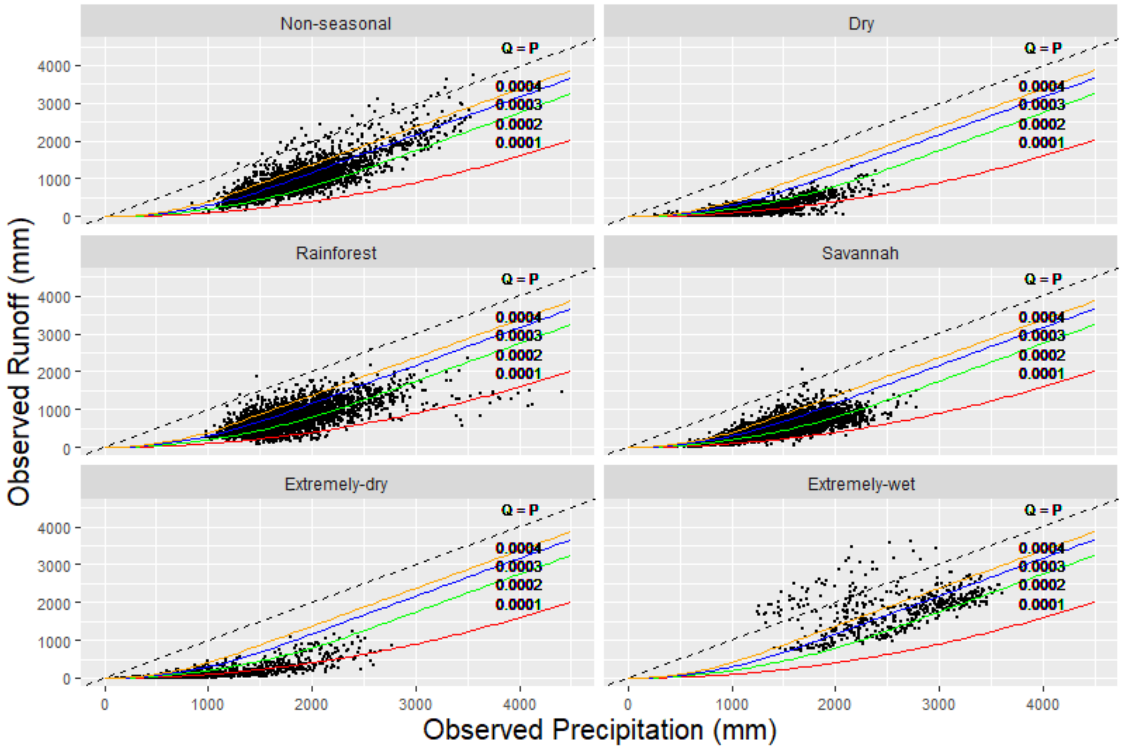

Figure S19: Scatter plot with observed annual runoff versus observed annual precipitation for each year of the 30 years series from the 735 CABra catchments, and Grunsky's equation plots for multiple α values (mm^{-1}), divided by cluster.

As it can be seen, Group 1 (Non-seasonal) shows a large range of precipitation and runoff, resulting in a high variance of α values. Although, the point cloud indicates higher density of points for α between 0.0002 and 0.0003. Group 2 (Dry) shows values of α between 0.0001 and 0.0002, which means that runoff represent a minor part of precipitation and a high rate of infiltration. The majority of catchments show precipitations below 1500 mm $year^{-1}$. Group 3 (Rainforest) presents high density of points with precipitation between 1000 and 2000 mm $year^{-1}$, and a great variety of α values. Although, it is noted that there is a high density between 0.0002 and 0.0003. An interesting fact is that, for some catchments with high values of P_{obs}, α is approximate 0.0001, indicating high rate of infiltration.

Group 4 (Savannah) indicates a majority of points with α very close to 0.0002 curve, and very few points with α > 0.0004. Group 5 (Extremely-dry), as expected, shows great density of points with low values of precipitation, and as observed in the biome analysis, have a relatively good fit in the Grunsky's Equation, with α between 0.0001 and 0.0002. Group 6 (Extremely-wet) clearly shows the highest variability of α values, with higher density of points between 0.0002 and 0.0004. Points with $Q_{obs} > P_{obs}$ may indicate that the runoff values of these catchments have a high percentage of water importation (Almagro, 2021).

The groups within Amazon and Atlantic Forest biomes, with wetter climate, show great variability of α values, mainly for higher P_{obs} and Q_{obs}. Pampa biome is located in the south region, with temperate climate, plain topography and grass vegetation, and shows a better fit for α = 0.0002 to 0.0004. Pantanal biome, containing the world's largest tropical wetland area, is hardly represented in the plot, but the few points show a great fit for α = 0.0001. While Cerrado biome shows a higher variability of α values, the scatter plot for Caatinga biome seems to follow the curve for α = 0.0001, which indicates that a great part of the precipitation does not turn into runoff, but instead, it is lost by infiltration or evapotranspiration, agreeing with findings of Pinheiro *et al.* (2016), which found out that approximately 75% (±17%) of precipitation in Caatinga catchments is lost by evapotranspiration.

5. Boxplots of α coefficient by month

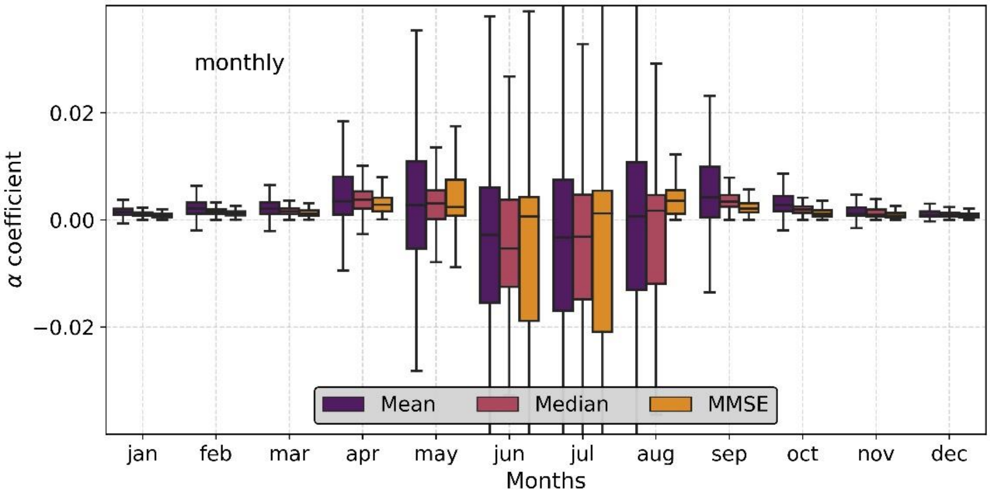

Figure S20: Boxplot of (a) Mean, (b) Median, and (c) MMSE α values, for the 12 months of the year, from the 716 catchments.

6. Interannual scale analysis

The MMSE α values were plotted within the Brazilian territory in order to observe any spatial distribution of its values, and is presented on Figure S21.

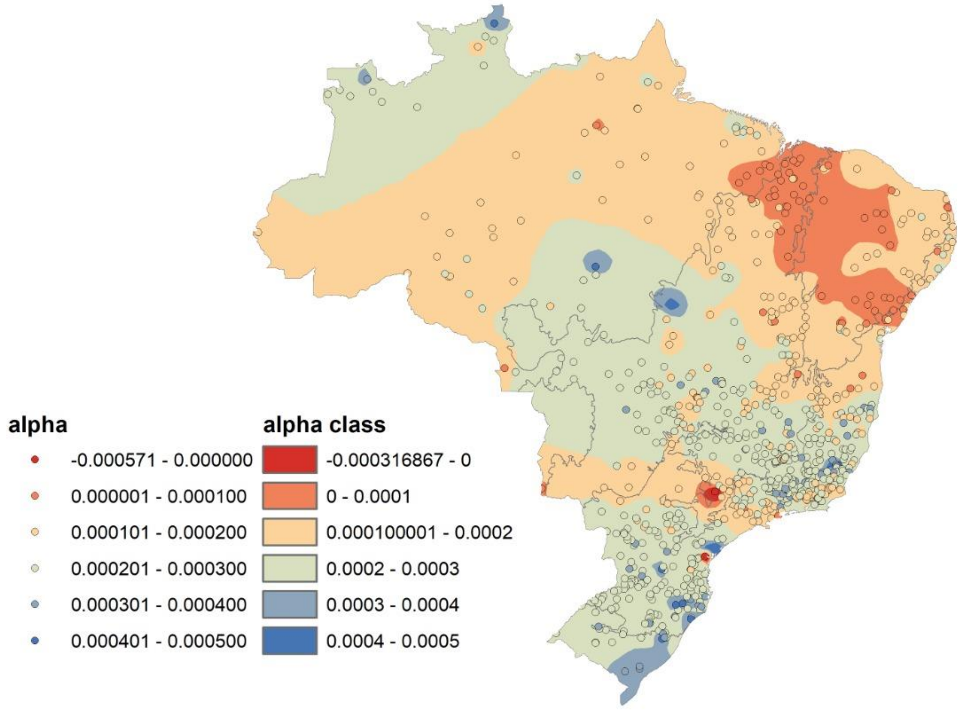

Figure S21: Map of α values correspondent to the MSE for each catchment.

As it can be seen, α values are lower in the northeast region, where groups 2 (Dry) and 5 (Extremely-dry) are predominant, as already verified on the boxplot of Figure S2. In addition, the highest values are presented on the south and northwest regions, where the Group 1 (Non-seasonal) and 6 (Extremely-wet) are located, respectively. In the center and eastern regions, the values vary more, as this region corresponds to groups 3 and 4 (Rainforest and Savannah, respectively), confirming a higher variability. It is noted that the behavior is similar to hydrological groups spatial distribution.

7. Monthly scale analysis

i) Mean value

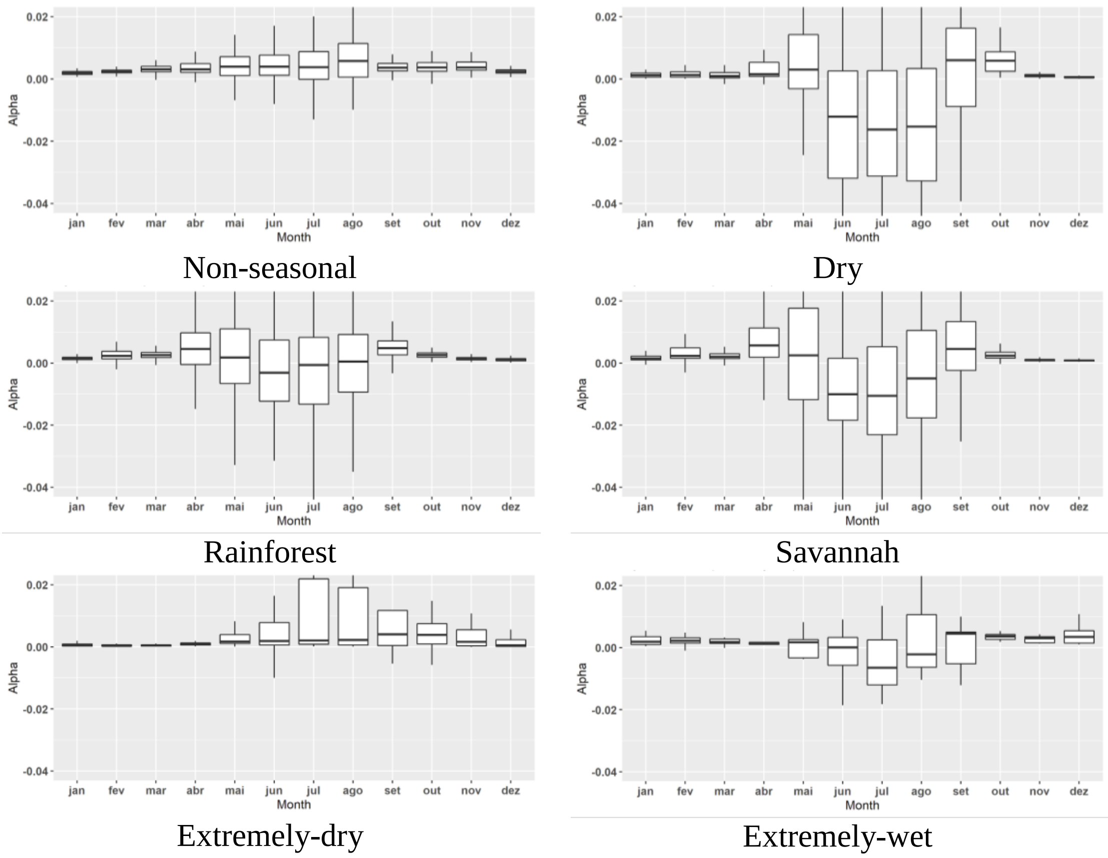

Figure S22: Boxplots α Mean Monthly Values by hydrological group.

According to Figure S22, the division of boxplots with the value of α for each month separated by hydrological group agrees with the behavior already seen in the all-catchments analysis: higher variation on drier months, and more cohesive values in other months. Although, the α values are different among the groups. The highest variabilities are found on Dry and Savannah groups, while the lowest ones are presented on Non-seasonal group.

Negative α values are less common to find in Non-seasonal and Extremely-dry groups. According to Generalized Grunsky's Equation (Equation 1), a negative α value is only possible when runoff is higher than precipitation (the opposite is not valid, i.e., runoff higher than precipitation not only leads to negative α value). This observation might indicate that runoff is intrinsically dependent on precipitation, or precipitation is almost constant during all year, and its relation with runoff hardly changes.

ii) Median value

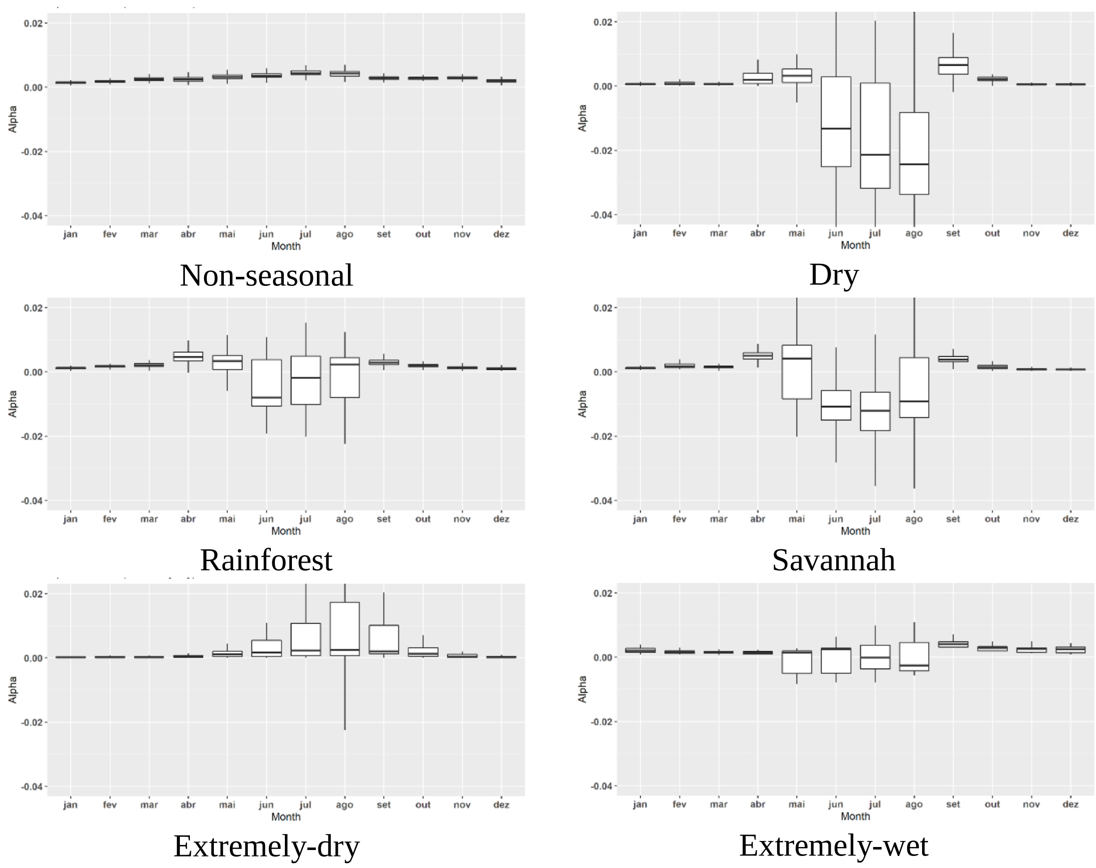

Figure S23: Boxplots α Median Monthly α values by hydrological group.

By analyzing Figure S23, it is clear again that by calculating α value for each catchment by the median of the 30-years series results in lower variability on values. Dry and Savannah group still present higher scatter, though it has decreased in magnitude and number of months with high variability, comparing to the method i) mean value. Non-seasonal has decreased even more its variability.

iii) MMSE value

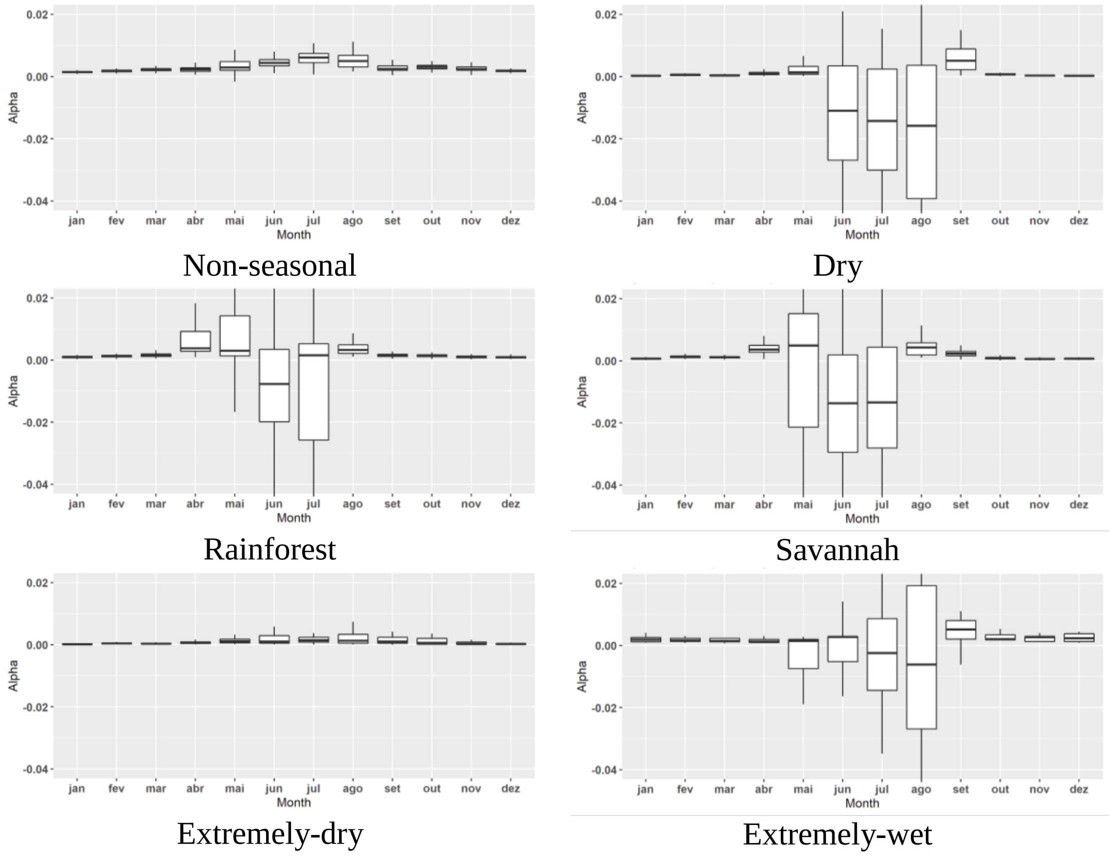

Figure S24: Boxplots α MMSE Monthly Values by hydrological group.

The regionalization on a monthly scale was done by assuming one α coefficient for each group and month, from the median values from the catchments. The boxplots of KGE values for the 12 months divided by each group are presented in Figure 10. These α values and the boxplots of PBIAS values are exhibited in the Supplementary Material (Table S1 and Figure S12, respectively).

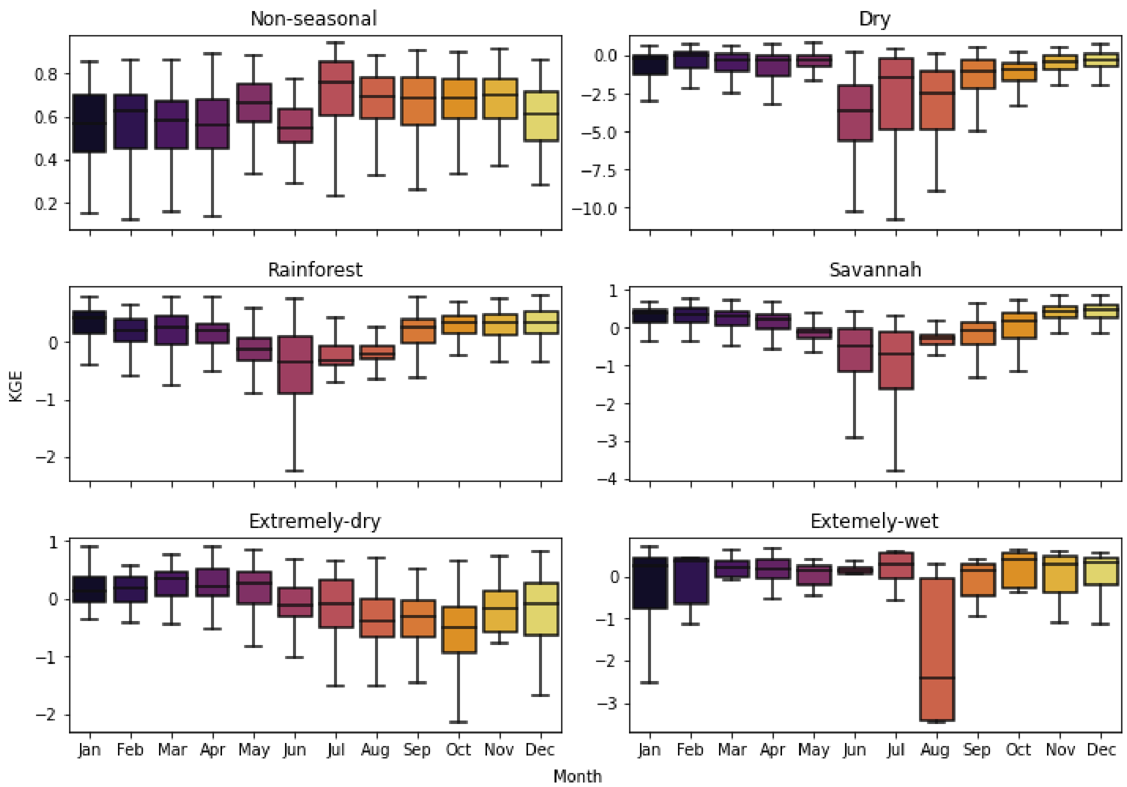

Figure 10: Boxplot with KGE from regionalized monthly runoff prediction for each month and group from the 716 CABra catchments.

The groups with the lowest values of PBIAS and highest values of KGE are Non-seasonal and Extremely-wet, given a more constant rainfall behavior on an intra-annual scale (Almagro, 2021). The worse performance was the Dry group, with median monthly KGE < 0 for all months, and PBIAS > 100% from June to August. The regionalization for a monthly scale appears to have worse results than the interannual scale for all groups, indicating the influence of interannual and intra-annual behavior of streamflow on the method's performance. The results appear to be similar to the predictions by Arheimer *et al.*, 2020, which showed relatively good results in Brazilian North region catchments when predicting monthly runoff, with median monthly KGE > 0.60, and worse performance in the North-east region, with KGE < -1. Dry and Extremely-dry groups might have higher variance on an interannual scale, and particularly on an intra-annual scale due to the influence of soil water content, *and high* rates of evapotranspiration (Pinheiro *et al.*, 2016). presents the boxplots of MMSE α values, with decrease of variability compared to the method of item i) mean value. Although, its results seem a little difference from the item ii) median value, as shown below. Apparently, the groups with the lowest variabilities are Non-seasonal and Extremely-dry, while the other groups show similar magnitudes of variation to the mean α values. In addition, negative α values are less common to find in Non-seasonal and Extremely-dry groups.

8. Spatial distribution of local analysis on a monthly scale

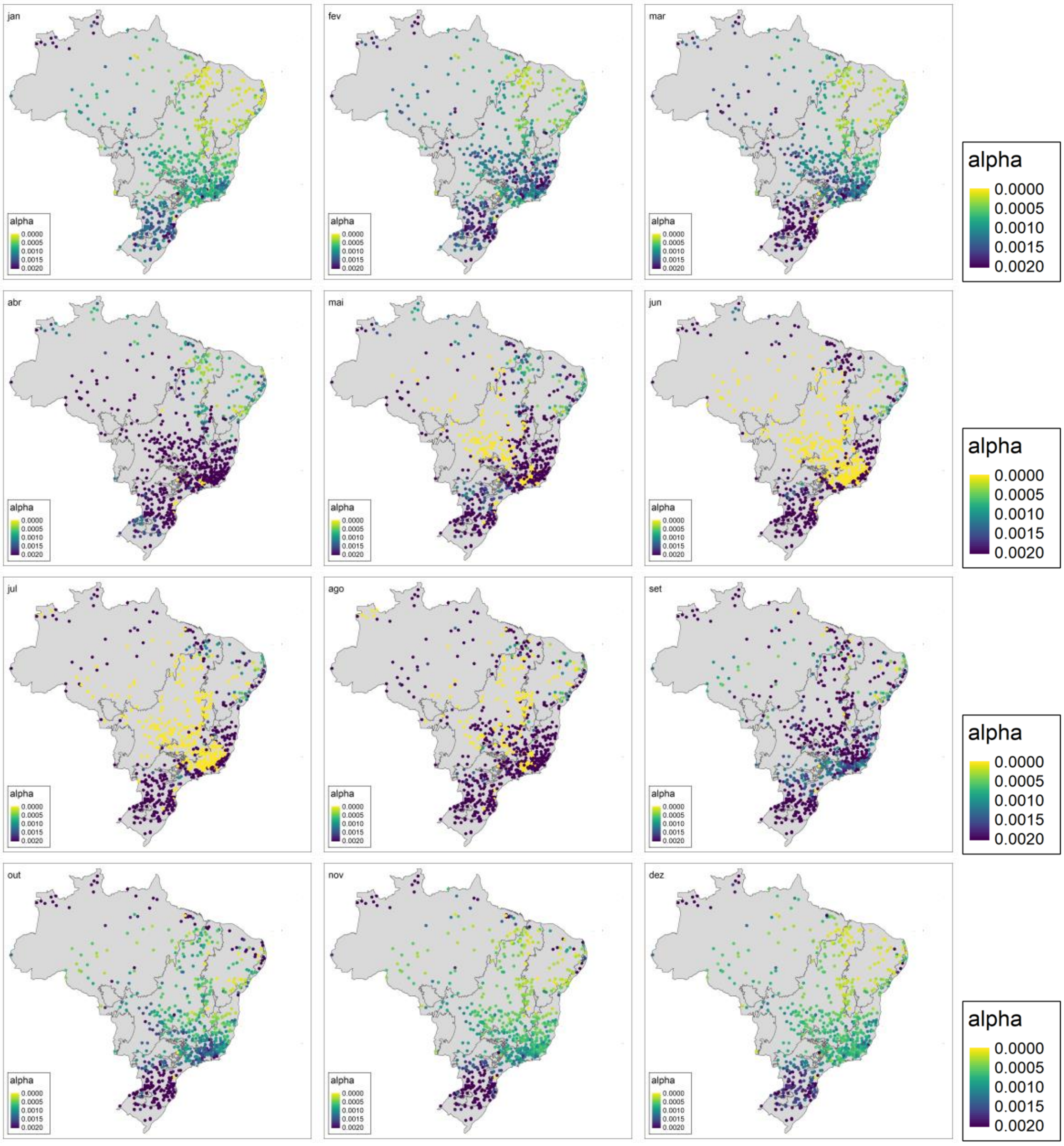

Figure S25: Maps of MMSE α value for each group for each month, from January (top left) do December (bottom right).

The spatial distribution for MMSE α value is presented on Figure S25. The lowest α values are exhibited in May, June and July for few catchments in the central region The values go from nearly 0.002 in April to close to or below zero from May to August. For Northeast region, α values are lower from October to March, varying from 0.0005 to 0.002. PBIAS spatial distribution is shown in Figure S26.

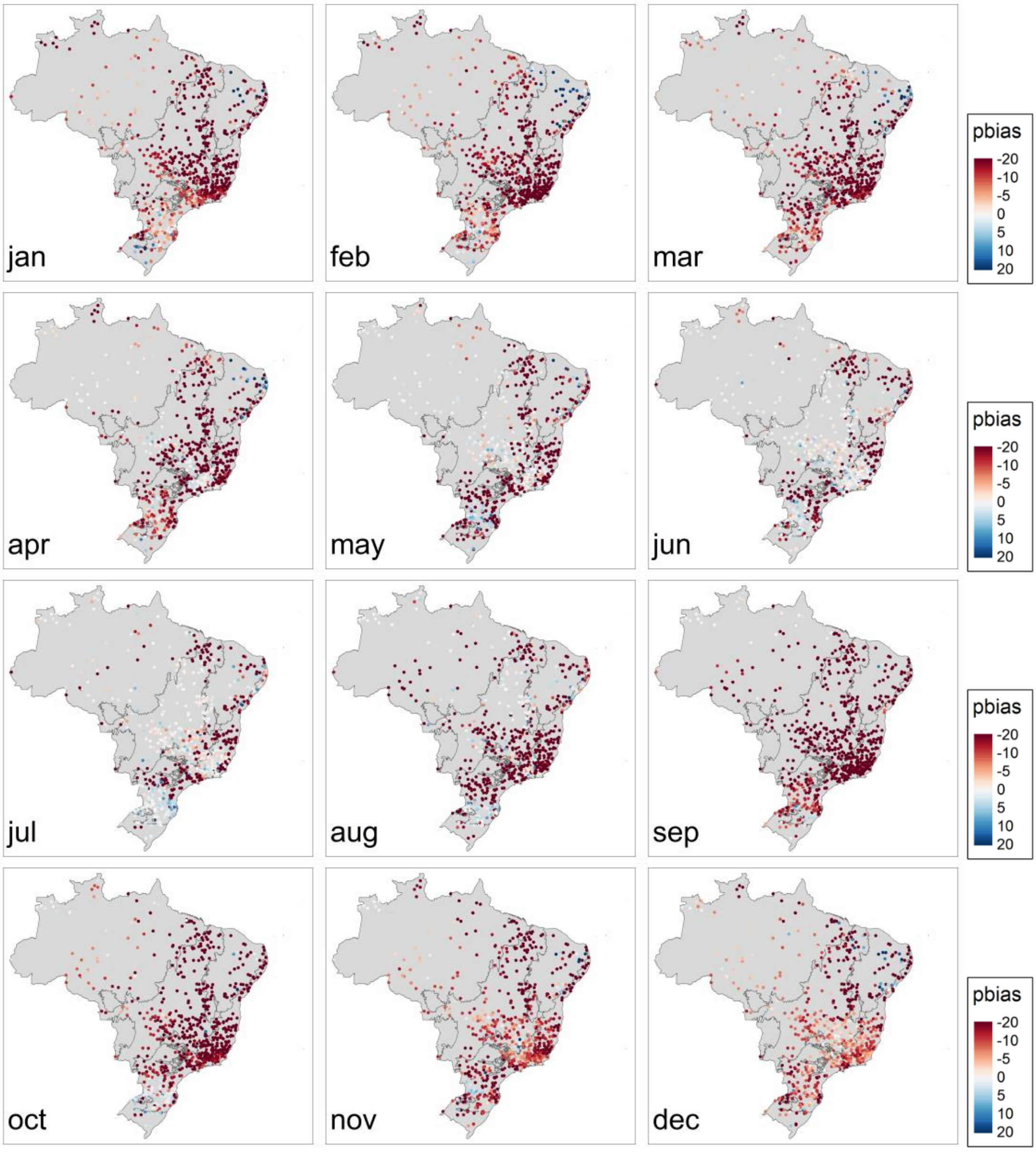

Figure S26: Maps of PBIAS from MMSE α value for each group for each month, from January (top left) to December (bottom right).

9. Regionalization results for the monthy scale

By regionalizing the α coefficients by month and group, Table S1 presents the results.

Table S3: Median alpha values from the regionalization on a monthly scale.

Group	Month	Median α
1	1	0.001441779
	2	0.001783571
	3	0.002155377
	4	0.002281761
	5	0.002870428
	6	0.004385096
	7	0.006076298
	8	0.005008543
	9	0.002480084
	10	0.003042482
	11	0.002377738
	12	0.001776446
2	1	0.000227684
	2	0.000505705
	3	0.000315033
	4	0.000867098
	5	0.001370199
	6	-0.010999635
	7	-0.014310653
	8	-0.015854241
	9	0.005079955
	10	0.000645941
	11	0.000303946
	12	0.000211039
3	1	0.000920937
	2	0.001291106
	3	0.001552959
	4	0.003792639
	5	0.002971881
	6	-0.007725215
	7	0.001540382
	8	0.003255095
	9	0.001472394
	10	0.001346934
	11	0.000976754
	12	0.000795681
4	1	0.000696192
	2	0.001229448
	3	0.001121238
	4	0.003586983
	5	0.004868715
	6	-0.013652271
	7	-0.013398232
	8	0.004242868
	9	0.002285178
	10	0.000823669
	11	0.000609988
	12	0.00064883
5	1	0.000135825
	2	0.000377692
	3	0.00030591
	4	0.000538056
	5	0.001049056
	6	0.000988222
	7	0.00140903
	8	0.001248355
	9	0.001026862
	10	0.00053345
	11	0.000257496
	12	0.000199181
6	1	0.001902146
	2	0.001571838
	3	0.001453796
	4	0.001238644
	5	0.001455963
	6	0.00270638
	7	-0.002427728
	8	-0.006144618
	9	0.005166935
	10	0.002012781
	11	0.002627158
	12	0.002319432

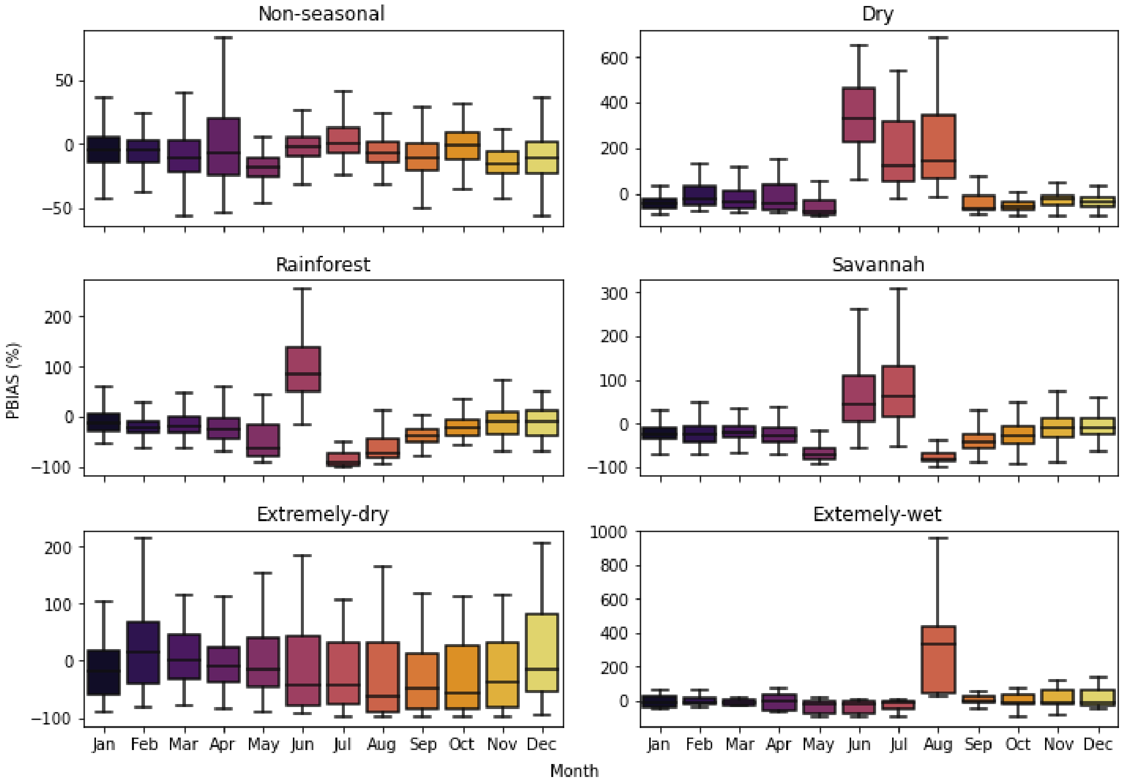

Figure S27: Boxplots with PBIAS from regionalized monthly runoff prediction for each month and group from the 716 CABra catchments.

EESC • USP