

## **Contribuição de técnicas multivariadas a índices de seca na compreensão das otto-regiões da bacia do São Francisco**

**Contribution of multivariate techniques to drought rates in understanding the otto-regions of the São Francisco basin**

**Contribución de técnicas multivariadas a las tasas de sequía en la comprensión de las otto-regiones de la cuenca de São Francisco**

Recebido: 19/02/2021 | Revisado: 22/02/2021 | Aceito: 24/02/2021 | Publicado: 06/03/2021

**Eucymara França Nunes Santos**

ORCID: <https://orcid.org/0000-0001-5674-0911>

Universidade Federal de Sergipe, Brasil

E-mail: [eucymara@gmail.com](mailto:eucymara@gmail.com)

**Ikaro Daniel de Carvalho Barreto**

ORCID: <https://orcid.org/0000-0001-7253-806X>

Universidade Federal Rural de Pernambuco, Brasil

E-mail: [daniel.carvalho.ib@gmail.com](mailto:daniel.carvalho.ib@gmail.com)

**Esdras Adriano Santos Barbosa**

ORCID: <https://orcid.org/0000-0002-3621-1913>

Universidade Federal de Sergipe, Brasil

E-mail: [esdras.adriano@gmail.com](mailto:esdras.adriano@gmail.com)

**Laélia Campos**

ORCID: <https://orcid.org/0000-0001-5701-9851>

Universidade Federal de Sergipe, Brasil

E-mail: [lpcampos@gmail.com](mailto:lpcampos@gmail.com)

**Antônio Samuel Alves da Silva**

ORCID: <https://orcid.org/0000-0002-8759-0036>

Universidade Federal Rural de Pernambuco, Brasil

E-mail: [samuelpmathematical@gmail.com](mailto:samuelpmathematical@gmail.com)

### **Resumo**

**Objetivo:** Utilizar as técnicas de análise multivariada, em particular a análise de componentes principais, para encontrar a melhor descrição dos índices de seca, e em seguida usar a análise de agrupamentos para determinar as regiões homogêneas do Rio São Francisco. **Método:** Foram calculados os índices para caracterizar a seca: Média anual de precipitação máxima diária (annualmax), média anual de precipitação acumulada (amount), média anual de dias sem chuva (<1 mm) (noprec), média anual de dias consecutivos sem chuva (<1 mm) (consecdry), média anual de dias consecutivos com chuva (>=1 mm) (consecwet) e a média anual de dias chuvosos com Precipitação excedendo o percentil 90% (prec90). Esses índices foram ortogonalizados por meio do método de componentes principais e posteriormente agrupados por meio do método de K-médias. **Resultados:** As variáveis amount e prec90 são as mais importantes, e juntas no primeiro componente são responsáveis por 40,56%, e as variáveis noprec e consecwet foram importantes para explicar 31,04% na segunda componente, e juntas explicam 71,60% da variabilidade total dos dados. Através da variabilidade de 86,40% nas três primeiras componentes principais retidas, a técnica de agrupamentos K-médias possibilitou a divisão de quatro áreas homogêneas na bacia do São Francisco. **Conclusões:** Foram observadas 4 regiões que são compostas pelas regiões do baixo e Sub-Médio São Francisco, Alto São Francisco e o Médio São Francisco, dividindo-se em duas partes que não ocorre uma perfeita correspondência com as otto-regiões estabelecidas.

**Palavras-chave:** Índices pluviométricos; Estatística multivariada; Rio São Francisco.

### **Abstract**

**Objective:** Use multivariate analysis techniques, in particular principal component analysis, to find the best description of drought rates, and then use cluster analysis to determine the homogeneous regions of the São Francisco River. **Method:** The indices to characterize the drought were calculated: Average annual maximum daily precipitation (annual max), annual average accumulated precipitation (amount), annual average of days without rain (<1 mm) (noprec), annual average of consecutive days without rain (<1 mm) (consecdry), annual average of consecutive days with rain (>=1 mm) (consecwet) and the annual average of Rainy Days with Precipitation exceeding the 90% percentile (prec90). These indices were orthogonalized using the principal component method and later grouped using the K-means method. **Results:** The variables amount and prec90 are the most important, and together in the first component they are responsible for 40.56%, and the variables noprec and consecwet were important to explain 31.04% in the second component, and together they explain

71.60% the total variability of the data. Through the variability of 86.40% in the first three main components retained, the technique of K-means clusters allowed the division of four homogeneous areas in the São Francisco basin. Conclusions: Four regions were observed, which are composed of the regions of the Lower and Sub-Middle São Francisco, Alto São Francisco and the Middle São Francisco, dividing into two parts that there is no perfect correspondence with the established otto-regions.

**Keywords:** Pluviometric indices; Multivariate statistics; São Francisco river.

### Resumen

Objetivo: Usar técnicas de análisis multivariante, en particular análisis de componentes principales, para encontrar la mejor descripción de las tasas de sequía, y luego usar el análisis de conglomerados para determinar las regiones homogéneas del río São Francisco. Método: Se calcularon los índices para caracterizar la sequía: promedio anual máximo de precipitación diaria (máximo anual), promedio anual de precipitación acumulada (cantidad), promedio anual de días sin lluvia (<1 mm) (noprec), promedio anual de días consecutivos sin lluvia (<1 mm) (consecuente), promedio anual de días consecutivos con lluvia (>=1 mm) (consecwet) y el promedio anual de días lluviosos con precipitación superior al percentil 90% (prec90). Estos índices se ortogonalizaron utilizando el método de componentes principales y luego se agruparon utilizando el método de K-medias. Resultados: Las variables cantidad y prec90 son las más importantes, y juntas en el primer componente son responsables del 40,56%, y las variables noprec y consecwet fueron importantes para explicar el 31,04% en el segundo componente, y juntas explican el 71,60% del total. variabilidad de los datos. A través de la variabilidad del 86,40% en los tres primeros componentes principales retenidos, la técnica de conglomerados K-médias permitió la división de cuatro áreas homogéneas en la cuenca del São Francisco. Conclusiones: Se observaron cuatro regiones, que están compuestas por las regiones de São Francisco Bajo y Sub-Medio, Alto São Francisco y São Francisco Medio, dividiéndose en dos partes que no existe una correspondencia perfecta con las otto-regiones establecidas.

**Palabras clave:** Índices pluviométricos; Estadística multivariada; Río São Francisco.

## 1. Introdução

O conhecimento do regime de precipitação de dias chuvosos consecutivos, acerca da magnitude, frequência e duração da precipitação, desempenha um papel essencial para as políticas públicas de prevenção a catástrofes naturais (Ahmad et al., 2018; Zhang et al., 2012). Este conhecimento permite avaliar os riscos de enchentes e secas, possibilitando o planejamento de estruturas de proteção, bem como o planejamento da agricultura (Koutsoyiannis, 2020; Zhang et al., 2014).

Dessa forma, hoje é bem aceito que as mudanças climáticas aceleram o ciclo hidrológico, que envolve vazão de rios, precipitação, escoamento e umidade do solo (Koutsoyiannis, 2020; Zhang et al., 2012; Ziegler et al., 2003).

O ciclo hidrológico quando acelerado pode levar a distribuição desigual de recursos hídricos e conseqüentemente modificar o padrão de ocorrência dos eventos climáticos extremos tanto no tempo como no espaço (Wang et al., 2017; Zhang et al., 2012). Mudanças no ciclo hidrológico trarão impactos severos na quantidade, periodicidade e duração das chuvas, evaporação, temperatura e escoamento, que trarão conseqüências na disponibilidade e competição por recursos hídricos (de Oliveira et al., 2017; Mishra et al., 2009).

O sistema energético brasileiro é vulnerável a mudanças climáticas. Diversas atividades são afetadas por estas mudanças, tais como a agricultura, o abastecimento de água e produção de energia elétrica, que são influenciadas pela diminuição da vazão. Neste contexto, a região nordeste sofre influência de seu clima semiárido, sendo observado um processo de desertificação mais evidente neste cenário (Andrade et al., 2012; de Jong et al., 2018).

Nesta região, encontra-se a importante bacia do Rio São Francisco, que é totalmente distribuída no território brasileiro, nas regiões sudeste e nordeste correndo de norte a sul com uma área de drenagem de 630.000 km<sup>2</sup>, abastecendo aproximadamente 13 milhões de habitantes (Bezerra et al., 2019; Maneta et al., 2009).

Sendo assim, diante de tantas variáveis meteorológicas, as ferramentas estatísticas vêm contribuir para as análises dos dados coletados, na compreensão do comportamento dos fenômenos climatológicos, aprimoramento da análise de previsão, gestão de recursos hídricos, evolução das séries ao longo do tempo, identificação de tendências significativas, buscando entender a dinâmica do clima para um melhor planejamento e desenvolvimento de uma determinada região (Mingoti, 2017).

A estatística multivariada corresponde a um conjunto de técnicas utilizadas para estudar diversas variáveis simultaneamente, com o objetivo de sintetizar ou simplificar a estrutura de variabilidade dos dados (Mingoti, 2017).

A técnica multivariada de análise de componentes principais (ACP) foi introduzida por Karl Pearson (1901), e fundamentada no artigo de Hotelling (1933), que usou com o propósito determinado de analisar a matriz de correlação (Hu et al., 2020). O método consiste em transformar um conjunto de variáveis originais em combinações lineares, através da estrutura de variância e covariância, com o objetivo de reduzir a extensão do conjunto de dados, e identificar novas variáveis significantes.

As chamadas componentes principais são as novas variáveis não correlacionadas entre si, que são capazes de reterem o máximo possível de informações, e de explicarem a maior parte da variabilidade total, sem que as informações relevantes dos dados brutos sejam perdidas.

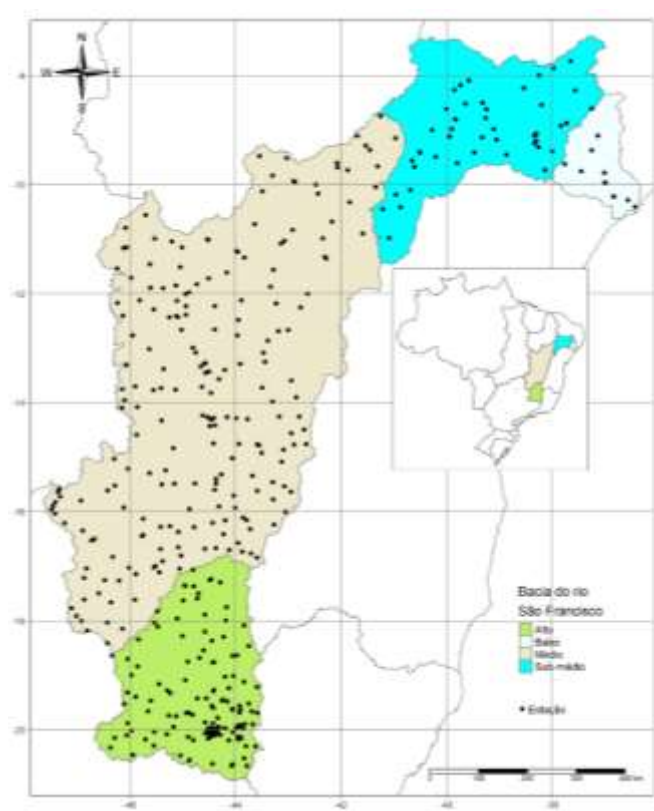
Já a técnica classificatória de análise de agrupamentos é utilizada para explorar as similaridades entre variáveis, definindo-os em grupos, considerando simultaneamente todas as medidas em cada elemento. O intuito é dividir indivíduos da amostra, em grupos, de forma que os elementos pertencentes a um mesmo grupo sejam similares entre si, e os de grupos diferentes sejam heterogêneos, em relação as variáveis que neles foram medidas (Mingoti, 2017).

Nesse sentido, o objetivo deste trabalho é utilizar as técnicas de análise multivariada, em particular a análise de componentes principais, para encontrar a melhor descrição dos índices de seca, e em seguida usar a análise de agrupamentos para determinar as regiões homogêneas do Rio São Francisco.

## 2. Metodologia

Os dados secundários de precipitação diária de 01/01/1970 a 31/12/2015 em 421 estações dentro da área de drenagem da bacia do Rio São Francisco foram coletados na Agência Nacional de Águas (ANA), conforme mapa mostrado na Figura 1.

**Figura 1** - Bacia do Rio São Francisco e Estações Pluviométricas, 2010.



Fonte: Autores.

Os índices pluviométricos foram calculados através da caracterização dos eventos extremos de precipitação e do regime de precipitação da bacia, conforme a Tabela 1 (Zhang et al., 2014). Diversos autores (Christensen & Christensen, 2004; Easterling et al., 2000; Zhang et al., 2011; Zolina et al., 2008) utilizaram esses índices para avaliar as propriedades estatísticas de eventos extremos de precipitação.

**Tabela 1** – Índices pluviométricos, definições e unidades.

<b>Índice</b>	<b>Definições</b>	<b>Unidade</b>
<b>ANUALMAX</b>	Média anual de precipitação máxima diária	mm
<b>AMOUNT</b>	Média anual de precipitação acumulada	mm
<b>NOPREC</b>	Média anual de dias sem chuva (<1 mm)	ias
<b>CONSECDRY</b>	Média anual de dias consecutivos sem chuva (<1 mm)	dias
<b>CONSECWET</b>	Média anual de dias consecutivos com chuva (>=1 mm)	dias
<b>PREC90</b>	Média anual de Dias Chuvosos com Precipitação excedendo o percentil 90%.	dias

Fonte: Autores.

A técnica multivariada de análise de componentes principais foi utilizada nos índices pluviométricos supracitados, a partir da matriz de correlação, para amenizar possíveis discrepâncias acentuadas entre as variâncias (Härdle & Simar, 2013). O objetivo deste procedimento é de que os primeiros componentes principais contam com grandes proporções de variabilidade dos dados, que poderão ser usados em outras análises no lugar das variáveis originais sem perder muita informação, além disso será possível identificar os índices mais importantes utilizados nesta análise.

O critério de Jolliffe foi utilizado na escolha desses componentes principais, que sugere reter os autovalores, que representam a variância de no mínimo 0,7 (Charleton et al., 2010; Marston & Ellis, 2020). Os primeiros componentes podem explicar de 80% a 90% da variabilidade total. A variável que apresenta maior coeficiente em valor absoluto com a componente principal de maior autovalor deverá ser a mais importante na representação dos dados, e assim segue a análise para as principais componentes retidas (Härdle & Simar, 2013).

Os gráficos utilizados são um valioso auxílio para análise exploratória dos resultados, permitem que as informações sejam exibidas graficamente, assim como também representar os níveis das variáveis. O teste de Bartlett (Aslam, 2020) será utilizado para verificar a homogeneidade de variância.

Os resultados das componentes principais serão utilizados para formar grupos em relação aos índices pluviométricos gerados, através da análise de agrupamentos não-hierárquicos com a finalidade de identificar as regiões que se relacionam com as variáveis, buscando mostrar a homogeneidade dentro dos grupos, e heterogeneidade fora dos grupos (Mingoti, 2007).

A medida de dissimilaridade, distância euclidiana, apropriada para variáveis quantitativas, foi utilizada para calcular as distâncias entre as amostras. A aplicação do método não hierárquico tem o objetivo buscar diretamente uma partição de dados em grupos, de modo a preservar a coesão interna e isolamento entre os grupos. Em seguida, utilizou-se o método não hierárquico K-médias, a fim de se identificar grupos distintos na bacia. Os métodos não hierárquicos possuem a vantagem de ser computacionalmente menos intensivos quando o número de objetos é grande (Mingoti, 2007).

Para uma melhor compreensão das diferenças observadas nos agrupamentos, foram apresentadas medidas descritivas (média, mediana, desvio padrão, mínimo, máximo, assimetria e curtose). A hipótese de normalidade das variáveis contínuas foi testada por meio do teste de Shapiro-Wilks (Razali & Wah, 2011). Sendo esta rejeitada, as diferenças nas medidas de tendência central foram testadas por meio do teste de Kruskal-Wallis (Hecke, 2012), bem como as múltiplas comparações avaliadas por meio do teste de Mann-Whitney com correção de Dunn (García et al., 2010).

Em todos os testes de hipótese, o nível de significância adotado foi de 5%. O software R Core Team 2021 (Team, 2021) foi utilizado em todas as análises.

### 3. Resultados e Discussão

As estimativas de correlação entre os índices de precipitação encontram-se apresentadas na Tabela 2. A maior correlação observada foi entre *prec90* e *amount*, razoável entre *consecwet* e *noprec90*, e baixa entre as demais. Devido a pequena quantidade de variáveis e a baixa correlação existente entre elas será necessário um grande número de componentes. A alta correlação entre as variáveis faz com que a maior proporção de variância seja explicada pela primeira componente.

**Tabela 2** – Matriz de correlação entre os índices utilizados na análise de componentes principais.

Variáveis	<i>anualmax</i>	<i>amount</i>	<i>noprec</i>	<i>prec90</i>	<i>consecwet</i>	<i>consecdry</i>
<i>anualmax</i>	1,00					
<i>amount</i>	0,33	1,00				
<i>noprec</i>	0,30	0,06	1,00			
<i>prec90</i>	0,29	0,98	0,01	1,00		
<i>consecwet</i>	-0,22	0,16	-0,70	0,18	1,00	
<i>consecdry</i>	-0,03	-0,43	0,12	-0,49	0,04	1,00

Fonte: Autores.

O teste de Bartlett mostrou-se significativo para a homogeneidade das variâncias. Os resultados obtidos para as componentes principais, referem-se aos índices utilizados nesta análise, e assim a Tabela 3 apresenta as estimativas dos autovalores, com suas respectivas variâncias individuais e acumuladas. Conforme observa-se na Tabela 3, foram obtidos seis componentes principais, podendo reduzir as variáveis originais em três componentes, por apresentarem autovalor inferior a 0,7, os quais retiveram 86,40% da variância total nos índices pluviométricos.

**Tabela 3** – Componentes, autovalores, proporção individual e acumulada da variação dos dados através da análise dos componentes principais.

Componente	Autovalor	Proporção Individual	Proporção Acumulada
1	2,4338	0,4056	0,4056
2	1,8621	0,3104	0,7160
3	0,8879	0,1480	0,8640
4	0,5842	0,0974	0,9614
5	0,2154	0,0359	0,9973
6	0,0162	0,0027	1,0000

Fonte: Autores.

No nordeste, a variação espacial e temporal da precipitação dos índices climáticos dos postos pluviométricos foi analisado utilizando análise de componentes principais e análise de agrupamentos (Araújo et al., 2012). Os resultados retiveram três componentes para a média pluvial, e para cada análise dos nove índices, todos explicaram mais de 85% da variância total do sistema, sendo que apenas dois deles (dias extremamente úmidos e índice de intensidade diária) foram representados por quatro componentes, que explicaram menos de 70%, índices não utilizados neste trabalho. No Estado do Piauí, autores consideraram apenas duas componentes principais, que explicaram 88% da variação dos dados de precipitação pluvial (Guedes et al., 2010).

Num estudo em Uberlândia no período de 2008 a 2012, por meio da análise de componentes principais conseguiu explicar 70,59% da variação total das características climáticas no município (Prado et al., 2016). Os meses úmidos, chuvosos e com pouca variação de temperatura foram representados por um único componente. Em outro estudo, autores encontraram cinco componentes fatoriais para explicar mais de 87% do índice de chuva, e 83,37% da variância para o índice de intensidade de precipitação diária (dos Santos, 2015). Além disso, foram necessários nove componentes para explicar 81,69% do índice de dias

consecutivos chuvosos, enquanto que para o índice de dias consecutivos secos, quatro componentes foram necessários para explicar 88%.

Com o objetivo de se entender a importância das variáveis nas componentes, e utiliza-las para auxiliar na formação dos grupos, a Tabela 4 mostra os coeficientes de ponderação de cada índice. As variáveis de maior peso no primeiro vetor (CP1), que explica 40,56% da variabilidade dos dados, foram amount e prec90, os índices mais correlacionados. O segundo vetor (CP2) explica 31,04%, e apresentam os índices noprec e consecwet com maiores associações. Já as variáveis consecdry e anualmax obtiveram maiores ponderações no terceiro e quatro componentes (CP3 e CP4).

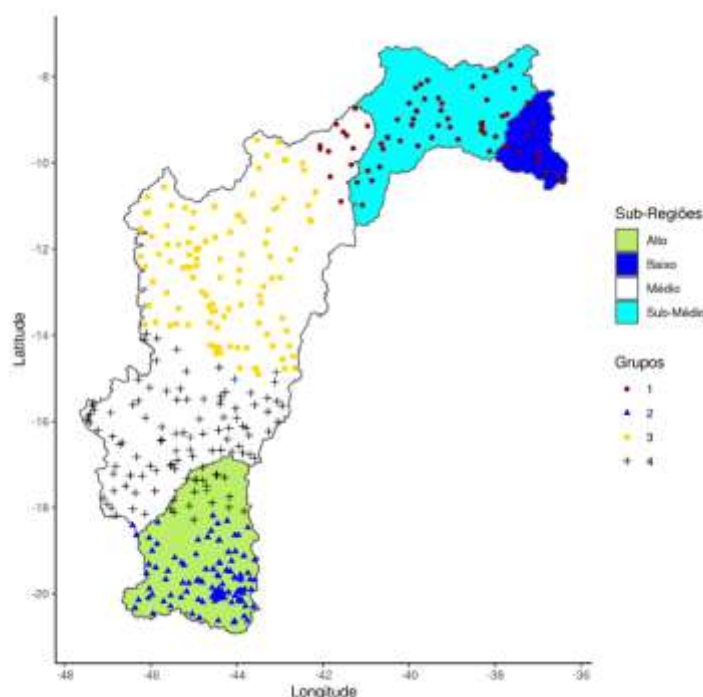
**Tabela 4** – Coeficientes de ponderação (autovetores) dos índices em relação aos componentes principais.

Variáveis	CP1	CP2	CP3	CP4	CP5	CP6
anualmax	-0,2616	-0,3859	0,5763	0,6672	-0,0700	0,0208
amount	-0,6151	-0,0048	0,1125	-0,2998	0,1865	-0,6959
noprec	-0,0051	-0,6618	-0,0494	-0,4082	-0,6267	0,0090
prec90	-0,6211	0,0326	0,0489	-0,2729	0,1548	0,7158
consecwet	-0,0934	0,6381	0,3728	-0,0588	-0,6644	-0,0143
consecdry	0,3984	-0,0689	0,7151	-0,4694	0,3195	0,0518

Fonte: Autores.

Assim, as cinco componentes que representam a maior variabilidade dos dados foram utilizadas para agrupar as estações pluviométricas em grupos. Devido ao grande número de estações foi utilizado o método não-hierárquico K-médias, onde a técnica tende a encontrar clusters de extensão espacial comparáveis, possibilitando extrair quatro grupos que estão representados na Figura 2. O mesmo algoritmo foi utilizado por outro grupo que permitiu a identificação de cinco diferentes zonas pluviométricas homogêneas no estado da Bahia (Dourado et al., 2013). Autores encontraram seis regiões que representam uma estratificação adequada da distribuição da precipitação das estações meteorológicas do estado do Rio de Janeiro, pelo mesmo método de agrupamento (André et al., 2008).

**Figura 2** - Agrupamentos dos índices pluviométricos da bacia do Rio São Francisco.



Fonte: Autores.



É possível observar que o grupo 1 está mais presente na região do baixo e sub-médio, os grupos 3 e 4 na região do médio São Francisco e o grupo 2 está mais presente nas regiões do alto São Francisco. Assim, é identificada certa correspondência entre os grupos encontrados e as otto-regiões previamente propostas.

O objetivo de identificar regiões pluviometricamente homogêneas foi temática de vários trabalhos. Para o estado da Bahia autores definiram um novo zoneamento com 10 agrupamentos levando em consideração a principal característica climática da localidade, utilizando os parâmetros das altas variabilidades das precipitações, e as grandezas estatísticas das distribuições espacial e temporal das chuvas (Simões et al., 2018). No estado do Paraíba, o método de agrupamento de Ward foi utilizado para agrupar quatro e cinco regiões homogêneas, dos índices climáticos (umidade, aridez e hídrico), ajustando-se a realidade da climatologia da chuva (De Freitas et al., 2013).

A estatística descritiva dos grupos está apresentada na Tabela 5. A respeito dos índices, é possível destacar que existe uma heterogeneidade entre os grupos em todos os índices, uma vez que o teste de Kruskal-Wallis rejeita a hipótese nula em todos os índices. Foi aplicado o teste de Mann-Whitney ajustado por Dunn para identificar as diferenças intergrupo e dessa forma corrigir a taxa de erro de família.

**Tabela 5** – Estatística descritiva dos índices pluviométricos por agrupamento.

	<b>anualmax</b>	<b>amount</b>	<b>noprec</b>	<b>prec90</b>	<b>consecwet</b>	<b>consecdry</b>
<b>Grupo 1 (n=105)</b>						
<b>Baixo/Sub-Médio São Francisco</b>						
<b>Média</b>	141,96	944,01	147,39	7,40	16,25	115,58
<b>DP</b>	47,92	129,58	104,33	1,80	6,86	12,46
<b>Mediana</b>	131,40	933,03	92,20	7,21	15,83	114,91
<b>Q1</b>	112,40	858,60	57,85	6,39	9,61	107,17
<b>Q3</b>	152,20	1033,19	265,65	8,85	21,65	123,85
<b>K</b>	1,66	0,01	0,26	-0,22	0,35	0,32
<b>A</b>	3,13	-0,37	-1,68	0,04	-1,00	0,31
<b>TMC</b>	b	b	a,b	b	b	d
<b>Grupo 2 (n=95)</b>						
<b>Alto São Francisco</b>						
<b>Média</b>	149,16	1157,38	159,77	9,85	19,01	98,24
<b>DP</b>	45,73	264,52	91,00	2,45	7,58	12,18
<b>Mediana</b>	137,00	1131,84	142,44	9,63	18,44	97,67
<b>Q1</b>	113,45	1025,33	87,79	8,45	12,07	89,55
<b>Q3</b>	174,10	1264,31	255,79	11,14	24,67	105,21
<b>K</b>	1,21	3,17	0,10	1,11	0,24	0,58
<b>A</b>	1,54	16,69	-1,49	3,13	-1,08	0,11
<b>TMC</b>	b,c	c	a,b	c	b	c
<b>Grupo 3 (n=83)</b>						
<b>Médio São Francisco</b>						
<b>Média</b>	122,90	620,57	137,66	3,93	12,68	82,78
<b>DP</b>	36,05	158,14	101,43	1,94	6,20	27,84
<b>Mediana</b>	120,40	604,51	92,59	3,88	11,28	89,07
<b>Q1</b>	96,60	501,33	75,92	2,75	7,70	55,24
<b>Q3</b>	141,35	684,90	228,13	4,83	18,46	103,05
<b>K</b>	0,94	1,67	0,64	0,81	0,27	0,15
<b>A</b>	1,28	3,60	-1,05	1,21	-1,19	-0,81
<b>TMC</b>	a	a	a	a	a	b
<b>Grupo 4 (n=138)</b>						
<b>Alto/Médio São Francisco</b>						
<b>Média</b>	159,10	1400,89	178,40	12,97	13,46	64,96
<b>DP</b>	39,10	138,45	92,23	1,52	4,52	9,27
<b>Mediana</b>	150,10	1409,79	220,90	13,15	11,90	61,22
<b>Q1</b>	137,60	1319,62	79,87	12,01	10,55	58,31
<b>Q3</b>	177,25	1480,72	254,38	13,86	14,83	70,90
<b>K</b>	1,44	-0,01	-0,68	-0,21	1,67	0,96
<b>A</b>	4,15	0,09	-1,14	0,33	2,58	0,12
<b>TMC</b>	c	d	b	d	a	a
<b>KW</b>	<0,001	<0,001	<0,001	<0,001	<0,001	<0,001

Legenda: DP – Desvio Padrão. Q1 – Primeiro Quartil. Q3 – Terceiro Quartil. K – Curtose. A – Assimetria. KW – Teste de Kruksal-Wallis. TMC – Teste de Múltiplas Comparações (Mann-Whitney-Dunn). <sup>a,b,c,d</sup> Subgrupos distintos ao nível de 5%.

Fonte: Autores.

Quanto à anualmax, diferenças significativas foram observadas entre os grupos 1, 3 e 4. O grupo 2 apresentou diferença apenas com o grupo 3. Quanto a noprec, foram observadas diferenças apenas entre os grupos 3 e 4. Quanto a amount, prec90 e consecdry os 4 grupos diferiram entre si. Quanto a consecwet, os grupos 1 e 2 foram diferentes dos grupos 3 e 4, não sendo diferentes entre si.

Dessa forma, as 4 regiões propostas podem ser caracterizadas da seguinte forma:

Região 1 (grupo 1): maior média anual de número de dias consecutivos sem chuva com posições intermediárias nos demais índices.

Região 2 (grupo 2): maior média anual de precipitação acumulada e maior média anual de número de dias consecutivos com chuva com posições intermediárias nos demais índices.



Região 3 (grupo 3): menor média anual de máxima precipitação diária, menor média anual de precipitação acumulada, menor média anual de número de dias sem chuva, menor média anual de número de dias com precipitação acima do percentil 90 e menor média anual de número de dias consecutivos com chuva.

Região 4 (grupo 4): maior média anual de máxima precipitação diária, maior média anual de precipitação acumulada, maior média anual de número de dias sem chuva, maior média anual de número de dias com precipitação acima do percentil 90 e menor média anual de número de dias consecutivos com chuva e sem chuva.

#### 4. Conclusão

As técnicas de análise multivariada foram eficientes para extrair informações dos índices pluviométricos e identificar as regiões do Rio São Francisco. As componentes principais permitiram reduzir a dimensão dos dados sem perda de informação, e identificou as variáveis significativas com as informações mais relevantes. Em seguida, através dessa sintetização dos dados, a análise de agrupamentos foi utilizada para propor uma regionalização baseado nos índices de precipitação. Foram observadas 4 regiões que são compostas pelas regiões do baixo e Sub-Médio São Francisco, Alto São Francisco e o Médio São Francisco, dividindo-se em duas partes que não ocorre uma perfeita correspondência com as oito-regiões estabelecidas.

#### Referências

- Ahmad, I., Zhang, F., Tayyab, M., Anjum, M. N., Zaman, M., Liu, J., Farid, H. U., & Saddique, Q. (2018). Spatiotemporal analysis of precipitation variability in annual, seasonal and extreme values over upper Indus River basin. *Atmospheric Research*, 213, 346–360.
- Andrade, E. M., Paulo Cosenza, J., Pinguelli Rosa, L., & Lacerda, G. (2012). The vulnerability of hydroelectric generation in the Northeast of Brazil: The environmental and business risks for CHESF. In *Renewable and Sustainable Energy Reviews*. 16(8), 5760–5769. Elsevier. <https://doi.org/10.1016/j.rser.2012.06.028>
- André, R. G. B., Marques, V. da S., Pinheiro, F. M. A., & Ferraud, A. S. (2008). Identificação de regiões pluviometricamente homogêneas no Estado do Rio de Janeiro, utilizando-se valores mensais. *Revista Brasileira de Meteorologia*, 23(4), 501–509. <https://doi.org/10.1590/s0102-77862008000400009>
- Araújo, W. S., Saviano Souza, F. A., Brito, J. I. B. de, & Lima, L. M. (2012). Estudo Pluvial no Nordeste do Brasil Utilizando Análise Multivariada (Rain Study in Northeast Brazil Using Multivariate Analysis). *Revista Brasileira de Geografia Física*, 5(3), 448. <https://doi.org/10.26848/rbgf.v5i3.232781>
- Aslam, M. (2020). Design of the Bartlett and Hartley tests for homogeneity of variances under indeterminacy environment. *Journal of Taibah University for Science*, 14(1), 6–10.
- Bezerra, B. G., Silva, L. L., Santos e Silva, C. M., & de Carvalho, G. G. (2019). Changes of precipitation extremes indices in São Francisco River Basin, Brazil from 1947 to 2012. *Theoretical and Applied Climatology*, 135(1–2), 565–576. <https://doi.org/10.1007/s00704-018-2396-6>
- Charleton, M. E., Brunson, C., Demšar, U., Harris, P., & Fotheringham, A. S. (2010). Principal Components Analysis: from Global to Local. *13th AGILE International Conference on Geographic Information Science*, 1–10.
- Christensen, O. B., & Christensen, J. H. (2004). Intensification of extreme European summer precipitation in a warmer climate. *Global and Planetary Change*, 44(1–4), 107–117. <https://doi.org/10.1016/j.gloplacha.2004.06.013>
- Freitas, J. C., De Andrade, A. R. S., Braga, C. C., Honorato, A., Neto, G., & De Almeida, T. F. (2013). Análise de agrupamentos na identificação de regiões homogêneas de índices climáticos no Estado da Paraíba, PB – Brasil. *Revista Brasileira de Geografia Física*, 6(4), 732–748.
- Jong, P., Tanajura, C. A. S., Sánchez, A. S., Dargaville, R., Kiperstok, A., & Torres, E. A. (2018). Hydroelectric production from Brazil's São Francisco River could cease due to climate change and inter-annual variability. *Science of The Total Environment*, 634, 1540–1553. <https://doi.org/10.1016/j.scitotenv.2018.03.256>
- Dourado, C. da S., Oliveira, S. R. de M., & de Avila, A. M. H. (2013). Análise de zonas homogêneas em séries temporais de precipitação no Estado da Bahia. *Bragantia*, 72(2), 192–198. <https://doi.org/10.1590/S0006-87052013000200012>
- Easterling, D. R., Meehl, G. A., Parmesan, C., Changnon, S. A., Karl, T. R., & Mearns, L. O. (2000). Climate extremes: observations, modeling, and impacts. *Science*, 289(5487), 2068–2074.
- García, S., Fernández, A., Luengo, J., & Herrera, F. (2010). Advanced nonparametric tests for multiple comparisons in the design of experiments in computational intelligence and data mining: Experimental analysis of power. *Information Sciences*, 180(10), 2044–2064.
- Guedes, R. S., Lopes, F. J., Amanajás, J. C., & Braga, C. C. (2010). Aplicação Da Análise Fatorial Em Componentes Principais A Dados De Precipitação No Estado Do Amapá. *Revista de Geografia (Recife)*, 27(1), 107–119–119.
- Härdle, W. K., & Simar, L. (2013). Applied multivariate statistical analysis. In *Applied Multivariate Statistical Analysis* (Vol. 5, Issue 8). Prentice hall Upper Saddle River, NJ. <https://doi.org/10.1007/978-3-642-17229-8>

- Hecke, T. Van. (2012). Power study of anova versus Kruskal-Wallis test. *Journal of Statistics and Management Systems*, 15(2–3), 241–247.
- Hu, Y., Yao, L., & Hu, Q. (2020). Evaluation of Water Resources Carrying Capacity in Jiangxi Province Based on Principal Component Analysis. *Journal of Coastal Research*, 105(sp1), 147–150. <https://doi.org/10.2112/JCR-SI105-031.1>
- Koutsoyiannis, D. (2020). Revisiting the global hydrological cycle: is it intensifying? *Hydrology and Earth System Sciences*, 24(8), 3899–3932. <https://doi.org/10.5194/hess-24-3899-2020>
- Maneta, M. P., Torres, M., Wallender, W. W., Vosti, S., Kirby, M., Bassoi, L. H., & Rodrigues, L. N. (2009). Water demand and flows in the São Francisco River Basin (Brazil) with increased irrigation. *Agricultural Water Management*, 96(8), 1191–1200. <https://doi.org/10.1016/j.agwat.2009.03.008>
- Marston, M. L., & Ellis, A. W. (2020). Delineating Precipitation Regions of the Contiguous United States from Cluster Analyzed Gridded Data. *Annals of the American Association of Geographers*, 1–19. <https://doi.org/10.1080/24694452.2020.1828803>
- Mingoti, S. A. (2007). Análise de dados através de métodos de estatística multivariada – uma abordagem aplicada. In *Análise de dados através de métodos estatística multivariada: uma abordagem aplicada*.
- Mishra, A. K., Özger, M., & Singh, V. P. (2009). An entropy-based investigation into the variability of precipitation. *Journal of Hydrology*, 370(1–4), 139–154. <https://doi.org/10.1016/j.jhydrol.2009.03.006>
- Oliveira, V. A., de Mello, C. R., Viola, M. R., & Srinivasan, R. (2017). Assessment of climate change impacts on streamflow and hydropower potential in the headwater region of the Grande river basin, Southeastern Brazil. *International Journal of Climatology*, 37(15), 5005–5023. <https://doi.org/10.1002/joc.5138>
- Prado, B. Q. de M., Fernandes, H. R., Araújo, T. G., Laia, G. A., & Biase, N. G. (2016). Avaliação de variáveis climatológicas da cidade de Uberlândia (MG) por meio da análise de componentes principais. *Engenharia Sanitaria e Ambiental*, 21(2), 407–413. <https://doi.org/10.1590/S1413-41522016147040>
- Razali, N. M., & Wah, Y. B. (2011). Power comparisons of shapiro-wilk, kolmogorov-smirnov, lilliefors and anderson-darling tests. *Journal of Statistical Modeling and Analytics*, 2(1), 21–33.
- Santos, D. C. (2015). Análise multivariada de índices climáticos na amazônia ocidental. *Revista Brasileira de Climatologia*, 15.
- Simões, Y. de S., Silva, E. H. B. C., & Araújo, H. A. de. (2018). Rainfall zoning of Bahia State, Brazil: an update proposal. *Ambiente e Agua - An Interdisciplinary Journal of Applied Science*, 13(1), 1. <https://doi.org/10.4136/ambi-agua.2171>
- Team, R. C. (2021). *R: A language and environment for statistical computing*. Vienna, Austria.
- Wang, H., Shao, Z., Gao, T., Zou, T., Liu, J., & Yuan, H. (2017). Extreme precipitation event over the Yellow Sea western coast: Is there a trend? *Quaternary International*, 441, 1–17. <https://doi.org/10.1016/j.quaint.2016.08.014>
- Zhang, Q., Peng, J., Xu, C. Y., & Singh, V. P. (2014). Spatiotemporal variations of precipitation regimes across Yangtze River Basin, China. *Theoretical and Applied Climatology*, 115(3–4), 703–712. <https://doi.org/10.1007/s00704-013-0916-y>
- Zhang, Q., Singh, V. P., Peng, J., Chen, Y. D., & Li, J. (2012). Spatial-temporal changes of precipitation structure across the Pearl River basin, China. *Journal of Hydrology*, 440–441, 113–122. <https://doi.org/10.1016/j.jhydrol.2012.03.037>
- Zhang, Q., Xu, C. Y., Chen, X., & Zhang, Z. (2011). Statistical behaviours of precipitation regimes China and their links with atmospheric circulation 1960–2005. *International Journal of Climatology*, 31(11), 1665–1678. <https://doi.org/10.1002/joc.2193>
- Ziegler, A. D., Sheffield, J., Maurer, E. P., Nijssen, B., Wood, E. F., & Lettenmaier, D. P. (2003). Detection of intensification in global- and continental-scale hydrological cycles: Temporal scale of evaluation. *Journal of Climate*, 16(3), 535–547. [https://doi.org/10.1175/1520-0442\(2003\)016<0535:DOIIGA>2.0.CO;2](https://doi.org/10.1175/1520-0442(2003)016<0535:DOIIGA>2.0.CO;2)
- Zolina, O., Simmer, C., Kapala, A., Bachner, S., Gulev, S., & Maechel, H. (2008). Seasonally dependent changes of precipitation extremes over Germany since 1950 from a very dense observational network. *Journal of Geophysical Research Atmospheres*, 113(6). <https://doi.org/10.1029/2007JD008393>