Revista Brasileira de Climatologia
Brazilian Journal of Climatology
OPEN ACCESS

# COMPARAÇÃO DE REGIÕES HOMOGÊNEAS DE PRECIPITAÇÃO A PARTIR DE DUAS FONTES DE DADOS DISTINTAS PARA O ESTADO DO PARÁ-BRASIL

## *Comparison of homogeneous regions of precipitation from two distinct data source for the state of Pará-Brazil*

## *Comparación de regiones homogéneas de precipitación a partir de dos fuentes de datos distintas para el estado de Pará-Brasil*

**David Figueiredo Ferreira Filho**
Programa de Pós-Graduação em Engenharia Civil pela Universidade Federal do Pará – PPGEC/UFPA
davydferreira@gmail.com

**Jéssica Ramos Abreu Ferreira**
Programa de Pós-Graduação em Engenharia Civil pela Universidade Federal do Pará – PPGEC/UFPA
jessicaramosabreu@gmail.com

**Francisco Carlos Lira Pessoa**
Docente no Programa de Pós-Graduação em Engenharia Civil pela Universidade Federal do Pará – PPGEC/UFPA
fclpessoa@ufpa.br

**Resumo:** O objetivo desse estudo foi comparar a regionalização da precipitação realizada através da técnica de agrupamento Fuzzy C-Means, com duas fontes de dados distintas, uma fornecida pela Agência Nacional da Água e Saneamento Básico (ANA) e outra obtida através do satélite meteorológico Global Precipitation Climatology Centre (GPCC) fornecido pelo Serviço Meteorológico Nacional Alemão (DWD), durante 30 anos (1986 a 2015), com o objetivo de verificar, através de técnicas estatísticas, qual será a representatividade e as diferenças das regiões formadas por dados tradicionais e satélite. A técnica não-hierárquica de Fuzzy C-Means foi aplicada à formação das regiões para os dois dados, a fim de agrupar as estações, sendo aplicadas técnicas de validação (Dunn, Silhouette e PBM), com o objetivo de formar o melhor agrupamento para análise dos dados. Análises de desempenho também foram feitas, usando métodos estatísticos. Como resultados, 2 regiões homogêneas de precipitação foram encontradas após os cálculos dos índices de validação, sendo especializadas em ambiente SIG. A parte sudoeste do estado foi onde houve maior divergência entre os dados analisados, de tal maneira que, na região homogênea de chuvas formada pelo GPCC, houve uma maior concentração da região

2, enquanto na análise formada pelos dados da ANA, houve fragmentos da região 2. Os resultados dos testes estatísticos mostraram que as comparações entre as duas regiões são aceitáveis, com pequenas diferenças, mas de grande valor para estudos hidrológicos na região.
**Palavras-chave:** GPCC. ANA. Fuzzy C-Means. Índices de Validações. Regionalização.

**Abstract:** The aim of this study was to compare the regionalization of precipitation carried out using the Fuzzy C-Means grouping technique, with two distinct data sources, one provided by the National Water Agency (ANA) and the other obtained through the Global Precipitation Climatology Centre (GPCC) meteorological satellite provided by the German National Meteorological Service (DWD), for 30 years (1986 to 2015), with the aim of verifying through statistical techniques, what will be the representativeness and differences of the regions formed by traditional and satellite. The non-hierarchical technique of Fuzzy C-Means was applied to the formation of the regions for the two data, in order to group the stations, later owned by the groupings, validation techniques (Dunn, Silhouette and PBM) were applied, with the aim of forming the best cluster for data analysis. Performance analyses were also performed, using statistical methods. As results, 2 homogeneous rainfall regions were found after the calculations of the validation indices, in which they were specialized in GIS environment. The southwestern portion of the state was where there was the greatest divergence between the analyzed data, in such a way that, in the homogeneous rainfall region formed by GPCC, there was a greater concentration of region 2, while in the analysis formed by ANA data, there were fragments of region 2. The results of the statistical tests showed that comparisons between the two regions are acceptable, with small differences, but of great value for hydrological studies in the region.
**Keywords:** GPCC. ANA. Fuzzy C-Means. Validation Indices. Regionalization.

**Resumen:** El objetivo de este estudio fue comparar la regionalización de la precipitación realizada a través de la técnica de agrupación Fuzzy C-Means, con dos fuentes de datos diferentes, una proporcionada por la Agencia Nacional de Agua y Saneamiento Básico (ANA) y otra obtenida a través del satélite meteorológico Global. Centro de Climatología de Precipitaciones (GPCC) proporcionado por el Servicio Meteorológico Nacional Alemán (DWD), durante 30 años (1986 a 2015), con el objetivo de verificar, a través de técnicas estadísticas, cuál será la representatividad y diferencias de las regiones formadas por tradicionales y datos satelitales. Se aplicó la técnica no jerárquica de Fuzzy C-Means a la formación de las regiones para los dos datos, con el fin de agrupar las estaciones, aplicándose técnicas de validación (Dunn, Silhouette y PBM), con el objetivo de formar la mejor agrupación. para el análisis de datos. También se realizaron análisis de rendimiento utilizando métodos estadísticos. Como resultado, se encontraron 2 regiones de precipitación homogéneas después de los cálculos del índice de validación, siendo especializadas en el entorno GIS. La zona suroeste del estado fue donde hubo mayor divergencia entre los datos analizados, de tal forma que, en la región de lluvia homogénea formada por el GPCC, hubo una mayor concentración de la región 2, mientras que en el análisis formado por la ANA datos, hubo fragmentos de la región 2. Los resultados de las pruebas estadísticas mostraron que las comparaciones entre las dos regiones son aceptables, con pequeñas diferencias, pero de gran valor para los estudios hidrológicos en la región.
**Palabras clave:** GPCC. ANA. Fuzzy C-Means. Índices de Validações. Regionalización.

Submetido em: 14/10/2022
Aceito para publicação em: 08/06/2023
Publicado em: 23/06/2023

# 1. INTRODUÇÃO

Para estudos hidrológicos, a precipitação se destaca como uma variável climática com grande variabilidade espaço-temporal, de grande impacto sobre os ciclos hidrológicos, fornecendo informações sobre o clima, tornando-se essencial para o planejamento das atividades humanas, promovendo o desenvolvimento local (DEZFULI, 2011; SHI et al., 2013).

Entretanto, estudos recentes descobriram que a variabilidade da precipitação foi alterada como consequência de mudanças climáticas (ÁVILA et al., 2014; SILVEIRA et al., 2016), sendo associada a essas mudanças a dificuldade de dados precisos, baixas densidades de estações, problemas com medições, erros apresentados nos dados, dificuldade de acesso às estações em algumas áreas, bem como a manutenção das existentes (FITZJARRALD et al., 2008; YAMANA; ELTAHIR, 2011; NOBRE et al., 2011), o que torna os estudos hidrológicos difíceis.

Tais variabilidades na Amazônia também estão associadas ao aumento do desmatamento, e as possíveis mudanças no clima regional e global tem motivado uma série de experimentos e modelos climáticos, pois há uma grande preocupação com as mudanças no uso do solo amazônico pela interferência antrópica, cuja tendência aponta para alterações e consequências climatológicas e ambientais em escala local e global (DAVIDSON et al., 2012).

Segundo Menezes e Fernandes (2016), no estado do Pará há uma grande variabilidade espacial e temporal da precipitação no estado do Pará relacionada à atuação de fenômenos climáticos como o El Niño, e no estudo, apenas 6 estações (Acará, Brasil Novo, Tracambeua, Cajueiro, Santo Antonio e Uruará) apresentaram tendências significativas a um nível de significância de 5%, que podem estar relacionadas a diversos motivos, incluindo falhas nos registros.

Esses problemas abriram à comunidade científica estudos sobre a busca de alternativas às medidas pluviográficas tradicionais, como o uso de satélites meteorológicos para estudos hidrológicos (SARMADI; SHOKOOHI, 2014; SULOCHANA; CHANDRIKA; RAO, 2014; CAI et al., 2015).

As falhas nas medições ligadas aos avanços nos algoritmos e na coleta de dados usados pelos satélites meteorológicos abrem estudos sobre métodos de obtenção de informações hidrológicas. Tais informações complementares podem ser obtidas através da formação de

regiões homogêneas, cujo objetivo principal é delimitar áreas com comportamentos hidrológicos semelhantes, como explorado por diversos trabalhos (ARELLANO; ESCALANTE, 2014; ASONG; KHALIQ; WHEATER, 2015; FAZEL et al., 2018).

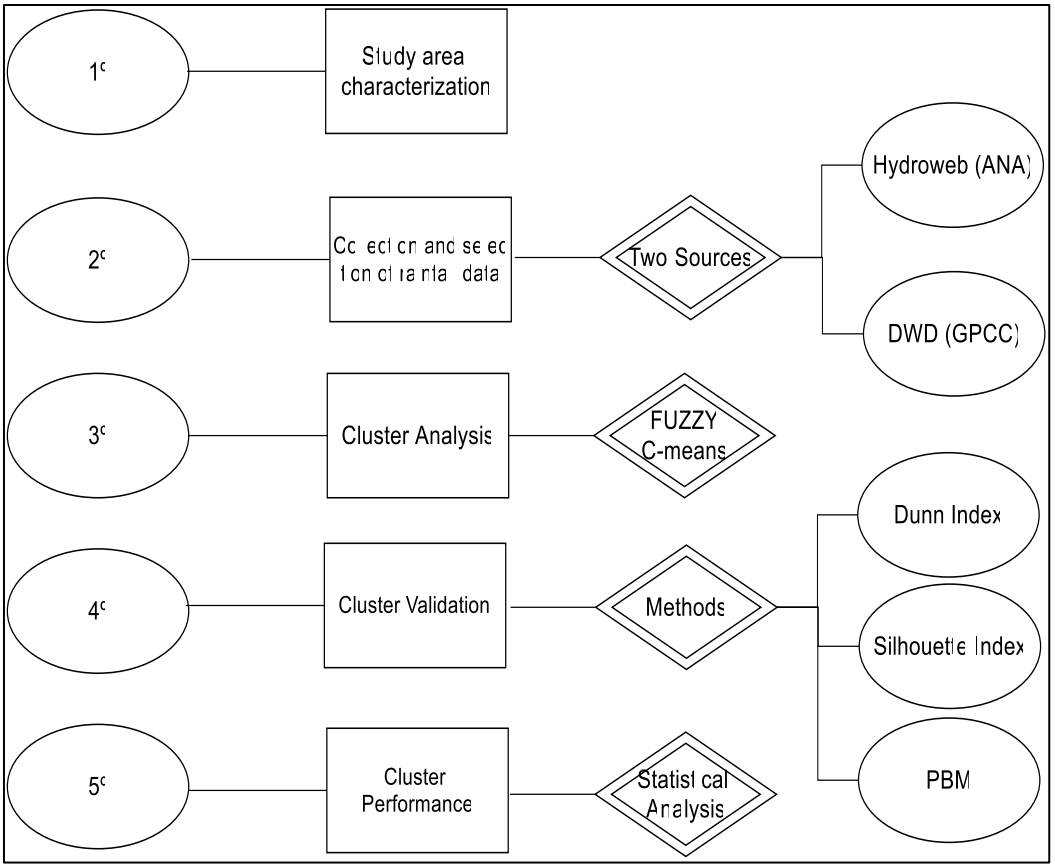
Uma das técnicas de agrupamento de regiões homogêneas é a formação de grupos por Fuzzy C-Means (AHANI; NADOUSHANI, 2016; NADOUSHANI; DEHGHANIAN; BAHRAM SAGHAFIAN, 2018), a qual é uma metodologia bem conhecida e sua importância está relacionada à obtenção de informações hidrológicas em locais não monitorados. Essa técnica é conhecida como nebulosa (AGUADO; CANTANHEDE, 2010), e forma grupos de livre escolha e precisa ser reaplicada várias vezes a fim de evitar locais mínimos e perguntas sobre os resultados, usando assim índices de validação (HALKIDI; BATISTAKIS; VAZIRGIANNIS, 2002), dos quais se encontram os índices mais diversos (PAKHIRA et al., 2004).

Portanto, o objetivo principal desse trabalho foi comparar a regionalização realizada através da técnica de agrupamento Fuzzy C-Means, mas com duas fontes de dados diferentes, uma fornecida pela Agência Nacional da Água e Saneamento Básico (ANA) e outra obtida através do satélite meteorológico Global Precipitation Climatology Centre (GPCC) fornecido pelo Serviço Meteorológico Nacional Alemão (DWD), durante 30 anos (1986 a 2015), com o objetivo de verificar por meio de técnicas estatísticas, que apesar das medições de erros nas estações tradicionais, qual será a representatividade e as diferenças das regiões formadas pelo satélite tradicional e pelo satélite.

## 2. MATERIAIS E MÉTODOS

Para o desenvolvimento deste trabalho, a metodologia foi subdividida de acordo com a Figura 1.

**Figura 1 –** Esquema Metodológico.



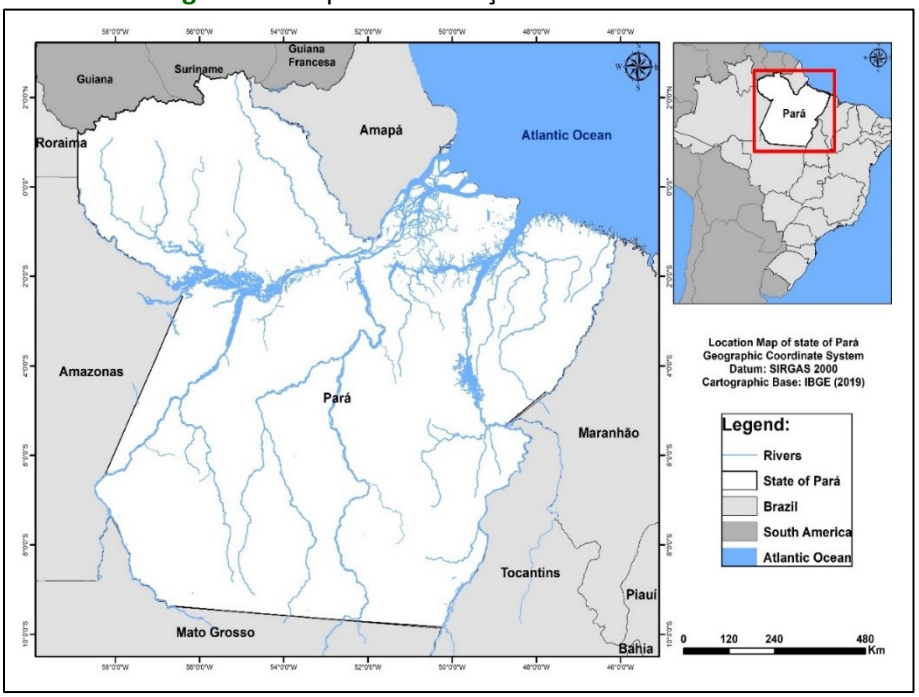

**Fonte:** Autores.

Como primeiro passo, a seleção da área de estudo, estado do Pará, foi feita a fim de posteriormente coletar os dados médios anuais de precipitação para ela. A área foi escolhida em virtude do estado apresentar estações com registros falhos e/ou ausência de dados em uma parte. A segunda etapa, a seleção dos dados, consistiu em obter dados de duas fontes diferentes, uma fornecida pela Agência Nacional da Água e Saneamento Básico - ANA através da plataforma HidroWeb, e outra fornecida pelo Centro Global de Climatologia de Precipitação - GPCC. Em posse dos dados, foi utilizado o método não hierárquico de Fuzzy C-Means, onde o número de grupos definidos variou de 2 a 10. Nesta fase, as variáveis Latitude, Longitude e Média anual de precipitação de cada estação foram agrupadas para cada fonte de dados. Posteriormente, com os agrupamentos formados, foram utilizados índices de validação, Dunn, Silhouette e PBM, para validar qual grupo foi melhor formado, para melhores análises. Finalmente, testes estatísticos foram aplicados para analisar o desempenho dos grupos formados, a fim de ter uma discussão mais rica e ver a correlação entre as diferentes fontes de dados.

## 2.1. Área de estudo (Study Area)

A área de estudo foi a região norte do Brasil, mais precisamente o estado do Pará (Figura 2), onde, segundo o IBGE (2010), tem uma população de aproximadamente 8,5 milhões de habitantes. De acordo com a classificação Köppen-Geiger (1928), a região pertence ao clima tipo "A", designado a climas tropicais com alta pluviosidade e temperaturas médias superiores a 18 °C. Na subclasse equatorial (Af), de acordo com Menezes et al. (2015), as temperaturas médias anuais estão entre 24 °C e 26 °C, com alta pluviosidade, próxima a 2.000 mm por ano em algumas análises.

**Figura 2 –** Mapa de localização da área de estudo.



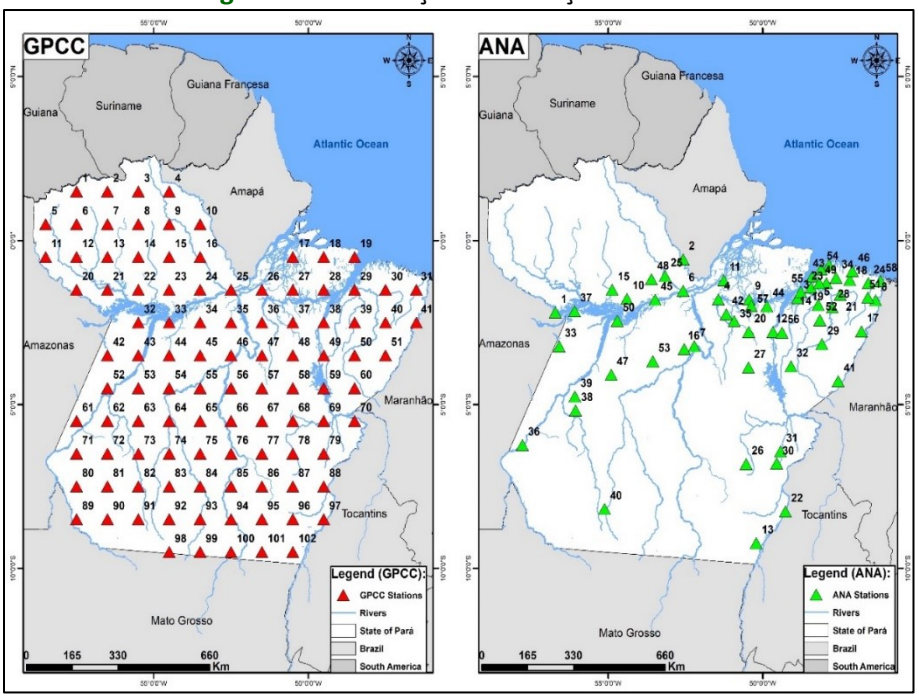

**Fonte:** Autores.

## 2.2. Seleção dos dados (Data Selection)

A fim de comparar o agrupamento, foram selecionadas duas fontes diferentes de dados de precipitação, da Agência Nacional da Água e Saneamento Básico - ANA e do Centro Global de Climatologia de Precipitação - GPCC, de acordo com a Figura 3. De acordo com a Organização Meteorológica Mundial (WMO, 1984), para o cálculo dos padrões climatológicos é necessário um período de estudo de pelo menos três décadas consecutivas (30 anos). Portanto, o período de estudo selecionado foi de janeiro de 1986 a dezembro de 2015 (30

anos), o que corresponde ao período hidrológico da região (janeiro a dezembro) para as duas fontes de dados, usando a precipitação média anual, para subsequentemente aplicar os métodos.

**Figura 3 –** Localização das estações da ANA e GPCC.




**Fonte:** Autores.

## 2.2.1. Global Precipitation Climatology Centre – GPCC

Em 1989 a WMO estabeleceu o GPCC, operado pelo Serviço Meteorológico Nacional Alemão (DWD), para fornecer análises globais de precipitação para ajudar a monitorar e pesquisar o clima da Terra. O centro é uma contribuição alemã ao Programa Mundial de Pesquisa Climática (WCRP) e ao Sistema Global de Observação Climática (GCOS), e é importante em estudos hidroclimáticos e meteorológicos, especialmente em áreas sem fontes de dados confiáveis e precisas (DINKU et al. 2008; RAZIEI et al. 2014).

O banco de dados de monitoramento GPCC é o resultado da união de três fontes de dados, os totais mensais precipitados derivados de relatórios meteorológicos sinópticos (SYNOP) do DWD da Alemanha, da Administração Nacional Oceânica e Atmosférica dos EUA (NOOA) e dos boletins climáticos do DWD da Agência Meteorológica do Japão (JMA) e do Serviço Meteorológico Nacional do Reino Unido (Met. Office UK).

Estes são dados de re-análise de modelos climáticos que se ajustam aos dados de campo e interpolam os campos pluviométricos. De acordo com Schneider et al. (2011), um dos objetivos do GPCC é fornecer ao público dados mensais e anuais de precipitação, disponíveis com resolução espacial de 1° x 1° e 2,5° x 2,5° de latitude por longitude, de 1901 até os dias de hoje, interpolados e calculados por 13 métodos estatísticos, disponíveis no site da NOOA.

## 2.2.2. Agência Nacional de Águas e Saneamento Básico – ANA

No portal HidroWeb da Agência Nacional de Águas e Saneamento Básico (ANA), estações com dados da série histórica foram selecionadas dentro do período de estudo. O resultado foi 57 estações pluviométricas localizadas no estado do Pará, com os códigos e coordenadas geográficas descritos na Tabela 1.

**Tabela 1 -** Estações pluviométricas selecionadas no Estado do Pará.

| Cod | Lat (°) | Long (°) | Cod | Lat (°) | Long (°) | Cod | Lat (°) | Long (°) |
|---|---|---|---|---|---|---|---|---|
| **148010** | -1,75 | -48,87 | **849000** | -8,26 | -49,26 | **855000** | -8,19 | -55,12 |
| **151001** | -1,79 | -51,43 | **47003** | -0,74 | -47,85 | **447001** | -4,29 | -47,56 |
| **148009** | -1,96 | -48,21 | **146010** | -1,29 | -46,58 | **251000** | -2,25 | -51,18 |
| **152005** | -1,53 | -52,58 | **152006** | -1,08 | -53,16 | **148012** | -1,09 | -48,40 |
| **352001** | -3,21 | -52,21 | **650001** | -6,82 | -50,54 | **149003** | -2,00 | -49,86 |
| **146008** | -1,82 | -46,34 | **350000** | -3,87 | -50,46 | **153000** | -1,80 | -53,48 |
| **150003** | -1,79 | -50,43 | **247003** | -2,04 | -47,75 | **47004** | -0,93 | -47,10 |
| **154000** | -1,77 | -54,40 | **348001** | -3,15 | -48,09 | **455004** | -4,09 | -54,90 |
| **151002** | -1,21 | -51,26 | **649001** | -6,79 | -49,55 | **253000** | -1,18 | -53,60 |
| **249003** | -2,79 | -49,67 | **649000** | -6,43 | -49,42 | **148003** | -1,30 | -48,17 |
| **950001** | -9,23 | -50,21 | **349002** | -3,83 | -49,09 | **254000** | -2,44 | -54,71 |
| **148002** | -1,44 | -48,44 | **356002** | -3,23 | -56,59 | **146005** | -1,73 | -46,60 |
| **154001** | -1,50 | -54,87 | **147010** | -1,13 | -47,63 | **248003** | -2,42 | -48,15 |
| **352005** | -3,31 | -52,54 | **250002** | -2,45 | -50,92 | **353000** | -3,68 | -53,55 |
| **247005** | -2,77 | -46,80 | **657000** | -6,24 | -57,78 | **48006** | -0,87 | -48,11 |
| **147002** | -1,20 | -47,18 | **151000** | -1,12 | -52,00 | **148011** | -1,57 | -48,77 |
| **147007** | -1,30 | -47,94 | **256001** | -2,15 | -56,09 | **249002** | -2,80 | -49,38 |
| **250000** | -2,79 | -50,45 | **555000** | -5,18 | -56,06 | **250001** | -1,99 | -50,37 |
| **147011** | -1,66 | -47,49 | **455003** | -4,75 | -56,08 | **146009** | -1,23 | -46,19 |

**Fonte:** Autores.

Entretanto, a série apresentou falhas, de modo que para viabilizar a pesquisa, as falhas foram preenchidas utilizando o método de ponderação regional. De acordo com Ishihara et al. (2014), o método é simplificado, mas seus resultados são bastante representativos quando aplicados a regiões climáticas homogêneas.

## 2.3. Análises de Agrupamentos por Fuzzy C-Means - FCM

A identificação de regiões pluviométricas homogêneas através da aplicação de técnicas de agrupamento contribui para a compreensão da variação espaço-temporal da precipitação (AMANAJÁS e BRAGA, 2012). Nesta pesquisa, foi utilizado o método nebuloso não-hierárquico Fuzzy C-Means proposto por Dunn (1973) e depois generalizado por Bezdek (1981).

Fuzzy Clustering, como é conhecido, é caracterizado pela ideia básica de que um conjunto de dados X = {x1, x2, ..., xn} é dividido em grupos p, e o resultado do agrupamento é expresso por graus de relevância, de tal forma que cada elemento pode pertencer a um único grupo ou mais.

O algoritmo é formado com uma suposição, de tal forma, que é necessário um banco de dados X = {x1,x2,x3,...,xn), no qual cada ponto xk, k=(1,2,3,...,n) é um vetor em $\Re p$, n é o total de dados do banco de dados X e $\Re p$ representa um espaço p-dimensional dos números reais. Assim, a matriz de partição para o domínio X é disposta na Equação 1:

$$Mfnc = \left\{ U \epsilon Ucn : Uik \epsilon [0,1], \sum_{i=1}^{c} Uik = 1{,}0 < \sum_{k=1}^{n} Uik < n \right\} \quad \text{Equação (1)}$$

Onde, Ucn é o verdadeiro grupo matricial c x n, c é o número de grupos que serão encontrados, dispostos em 2≤c≤n, U é a matriz de partição Fuzzy para o domínio X e U$ik$ é o grau de relevância de xk em um grupo i.

Assim, a tarefa de gerar um indicador para ajudar a verificar a convergência é atribuída à função objetiva (J), definida através da Equação 2 abaixo:

$$J = \sum_{i=1}^{n} \sum_{j=1}^{p} (Uik)^m * d(Xk, Cj)^2 \quad \text{Equação (2)}$$

Onde n é o número de dados, p é o número de grupos, m é o parâmetro fuzzy, $d$ é a distância euclidiana entre $X$k,$Cj$, $X$k é o vetor de dados, onde i=1, 2,..., n, representa um atributo de dados e Cj é o centro de um agrupamento Fuzzy.

Então a função objetiva J é minimizada e os graus de relevância U$ik$ são gerados de acordo com a Equação 3:

$$\mathrm{Uik} = \left[\sum_{\mathrm{k=l}}^{\mathrm{c}} \left(\frac{\mathrm{d(XkCj)}}{\mathrm{d(Xk, Cj}}\right)^{2/(\mathrm{m}-1)}\right] \quad \text{Equação (3)}$$

E Cj é um vetor chamado centroide ou protótipos (PEDRYCZ; VUKOVICH, 2004), que pode ser obtido através da Equação 4:

$$Cj = \frac{\Sigma(\mathrm{Uik})^{\mathrm{m}}\ \mathrm{Xk}}{\Sigma(\mathrm{Uik})^{\mathrm{m}}} \quad \text{Equação (4)}$$

## 2.4. Validação dos grupos (Groups Validation)

Todos os processos de agrupamento produzem uma solução, mesmo quando os dados originais não têm subestruturas (TAN; STEINBACH); KUMAR, 2005), principalmente no método Fuzzy C-Means, onde pode gerar várias soluções porque é um método de livre escolha, e para isso, no presente estudo, foram utilizados 3 índices de validações para verificar qual é a melhor disposição dos agrupamentos para o estado do Pará, ou seja, aquele que melhor representa o comportamento de acordo com os dados de entrada das duas fontes de dados, a fim de evitar maus agrupamentos para as análises, e, entre os escolhidos para avaliá-los estão os índices de Dunn (PAKHIRA et al. , 2004), Silhouette (ROUSSEEUW, 1987) e PBM (PAKHIRA et al, 2004).

### 2.4.1. Índice Dunn

O método Dunn (PAKHIRA et al., 2004) é definido como S e T dois subconjuntos não vazios no RN. Então, o diâmetro Δ(S) é definido como a distância δ entre S e T:

$$\Delta(\mathrm{S}) = \max_{x,y\ E\ S}\{d(x,y)\} \quad \text{Equação (5)}$$

$$\delta(\mathrm{S,T}) = \min_{xES,yET}\{d(x,y)\} \quad \text{Equação (6)}$$

Onde d (x; y) é a distância entre os pontos x e y. Para qualquer divisória Dunn, você definiu o seguinte índice:

$$Vd = \min_{1\le s\le K}\{\min_{1\le t\le K, t\neq s}\{\frac{\delta i\{Cs, Ci)}{\max_{1\le k\le K}\Delta j(Ck)}\}\}$$ Equação (7)

Assim, quanto maior o Vd, o correspondente é o número de clusters que maximiza o Vd é considerado o número ideal de clusters (PAKHIRA et al., 2004).

## 2.4.2. Índice Silhouette

O método de silhueta, desenvolvido por Rousseeuw (1987), afirma que a largura da silhueta avalia a qualidade de uma solução de agrupamento, considerando tanto a compacidade quanto a separação, a distância entre os pontos de dados em dois grupos vizinhos. Este método permite determinar o número apropriado de grupos, de modo que o valor de k é escolhido de modo a fornecer o melhor valor médio da Silhueta (GIL et al., 2015), assim de acordo com a seguinte equação:

$$s(i) = \frac{b_i - w_i}{max(b_i, w_i)}$$ Equação (8)

Para $b_i = min_k(B_{i,k})$. Onde Wi é a distância média do i-ésimo ponto aos outros pontos do mesmo grupo e B_(i,k) é a distância média do i-ésimo ponto aos pontos de outro grupo k. Assim, ao apresentar um valor unitário positivo, os pontos são dispostos de forma correta, por outro lado, ao ser caracterizado pelo valor zero, torna-se impossível identificar a que grupo pertencem e, finalmente, ao apresentar um valor unitário negativo, os pontos são provavelmente alocados aos grupos errados.

Assim, a largura média da silhueta de um grupo $\bar{}$(k) para todos i em um dado grupo é definida como a média de todas as silhuetas individuais, onde n é o número de objetos no conjunto de dados, de acordo com a seguinte equação:

$$\bar{s}(k) = \frac{1}{n}\sum_{i=1}^{n} s(i)$$

Equação (9)

### 2.4.3. Índice PBM

O índice PBM é definido como o produto de três fatores (Equação 10), onde o objetivo é maximizar o mesmo para obter o número real de agrupamentos, em outras palavras, o valor máximo para a melhor partição (PAKHIRA et al., 2004).

$$PBM(k) = (\frac{1}{k} * \frac{E1}{Ek} * Dk)^2 \quad \text{Equação (10)}$$

Onde, k é o número de clusters.

O primeiro fator diminui conforme o valor de K aumenta, reduzindo o valor do índice. O segundo fator consiste na razão E1 (ponto central do conjunto de dados), que é constante para um dado conjunto de dados, e EK, que diminui à medida que o valor de K aumenta. Então, o valor do índice PBM aumenta à medida que o valor de EK diminui. Isto, por outro lado, indica que a formação de mais clusters, que são naturalmente comprimidos, deve ser encorajada. Finalmente, o terceiro fator (DK), que mede a separação máxima entre um par de aglomerados, aumenta com o valor de K.

Assim, o fator E1 é a soma das distâncias das unidades de amostra do centro geométrico de todas as amostras (Equação 11), este fator não depende do número de agrupamentos iniciais.

$$E1 = \sum_{t-1}^{n} d(x(t), Wo) \quad \text{Equação (11)}$$

Onde, d é a Distância Euclidiana, x(t) é o vetor de dados, Wo é o centro de um cluster. Com E1, o valor de Ek (Equação 12) é calculado, determinado pela distância entre os grupos de grupos de K e ponderado pelo valor de pertinência correspondente a cada amostra para o agrupamento.

$$Ek = \sum_{t=1}^{n} \sum_{i=1}^{k} Ui(t) d[x(t) x\ Wi]^2 \quad \text{Equação (12)}$$

Onde Ui(t) = Graus de relevância, e Dk representa a separação máxima de cada par de agrupamentos (Equação 13).

$$Dk = \max_{i,j=1\ldots k} (d\ (Wi, Wj)) \quad \text{Equação (13)}$$

Para selecionar o melhor agrupamento, o valor máximo do índice PBM (Equação 14), ou seja, a melhor partição (argmax(PBM(k)) deve ser alcançada).

$$K = argmax\ (PBM(k)) \qquad \text{Equação (14)}$$

## 2.5. Performance dos agrupamentos (Group Performance)

Para analisar o desempenho das estimativas de regiões homogêneas formadas entre as estações ANA e o satélite GPCC, foi utilizada uma malha de mil pontos gerados nos mapas de interpolação para a correlação. Além disso, alguns critérios estatísticos foram aplicados, a saber: o Coeficiente de Correlação (R) e Determinação ($R^2$) usados para medir a concordância geral; o Erro Quadrático Médio (MSE) e Raiz do Erro Quadrático Médio (RMSE) usados para verificar erro e predisposição de dados; e o Coeficiente Nash-Sutcliffe (NASH) usado para avaliar a eficiência entre as estimativas (Tabela 2).

**Tabela 2 -** Estatísticas para estimar o desempenho.

| Método | Equação | Intervalo | Valor bom | Nº Equação |
|---|---|---|---|---|
| R | $R = \frac{\sum(Pi - \overline{Pi})(Piy - \overline{Piy})^2}{\sqrt{\sum(Pi - \overline{P})^2}\sqrt{\sum(Piy - \overline{Piy})^2}}$ | -1 to 1 | 1 | (15) |
| $R^2$ | $R^2 = \frac{\sum(Piy - \bar{P})^2}{\sum(Pi - \bar{P})^2}$ | 0 to 1 | 1 | (16) |
| MSE | $MSE = \frac{\mathbf{1}}{\mathbf{N}}\sum(Pi - \boldsymbol{Piy})^2$ | 0 to ∞ | 0 | (17) |
| RMSE | $RMSE = \left[\frac{\mathbf{1}}{\mathbf{N}}\sum(Pi - \boldsymbol{Piy})^2\right]^{\mathbf{1}/\mathbf{2}}$ | 0 to ∞ | 0 | (18) |
| NASH | $NASH = 1 - \frac{\sum_{i=1}^{n}(Pi - Piy)^2}{\sum_{i-1}^{n}(Pi - \bar{P})^2}$ | - ∞ to 1 | 1 | (19) |

Nota: Onde Pi é precipitação ANA, Piy é precipitação GPCC.

**Fonte:** Autores.

O coeficiente R de Pearson foi determinado por (Eq. 1) e pode variar de -1 (correlação perfeita negativa), +1 (correlação perfeita positiva) a 0 (sem correlação). Assim, como sugerido por Rencher e Christensen (2012), as correlações R podem ser: 0,00 a 0,19 (muito fraca); 0,20 a 0,39 (fraca); 0,40 a 0,69 (moderada); 0,70 a 0,89 (forte); 0,90 a 1,0 (muito forte). O coeficiente de determinação ($R^2$) é um critério estatístico que mede a proporção da mudança

em Y (variável dependente) que pode ser explicada pela variável X (variável independente). O valor de $R^2$ pode variar de 0 a 1, portanto $0 \leq R^2 \leq 1$ (NASH; SUTCLIFFE, 1970).

Outro critério utilizado para o ajuste dos resultados foi o coeficiente Nash-Sutcliffe (NASH) (NASH; SUTCLIFFE, 1970). Assim, foi utilizada uma classificação sugerida por Pereira et al. (2016) e Costa et al. (2019): bons valores (NASH > 0,75); aceitáveis (0,36 ≤ NASH ≤ 0,75); inadequados (NASH < 0,36). O RMSE foi aplicado para retratar a diferença entre os valores de precipitação (COSTA et al., 2019).

## 3. RESULTADOS E DISCUSSÃO

A partir de uma comparação de simulação usando a precipitação média anual como variável dependente gerada a partir de duas fontes de dados, ANA e GPCC, para o estado do Pará, com a aplicação de testes não hierárquicos Fuzzy C-Means de 2 a 10 grupos, e tomando como dados de entrada as variáveis independentes - latitudes, longitudes e precipitação, para cada estação, obtivemos as regiões homogêneas (ou grupos homogêneos) de precipitação para o estado.

A formação destes grupos foi validada com os índices de validação, já que a técnica não hierárquica aplicada pode gerar resultados sem subestruturas para análise (Tan; Steinbach; Kumar, 2005), na qual foram gerados coeficientes como os apresentados na Tabela 3, também aplicados por Parchure e Gerdam (2019) em Mumbai, na Índia, e diferentes dos aplicados por Menezes, Fernandes e Rocha (2015) para o estado do Pará.

**Tabela 3 -** Resultado dos índices de validações dos dados GPCC e ANA.

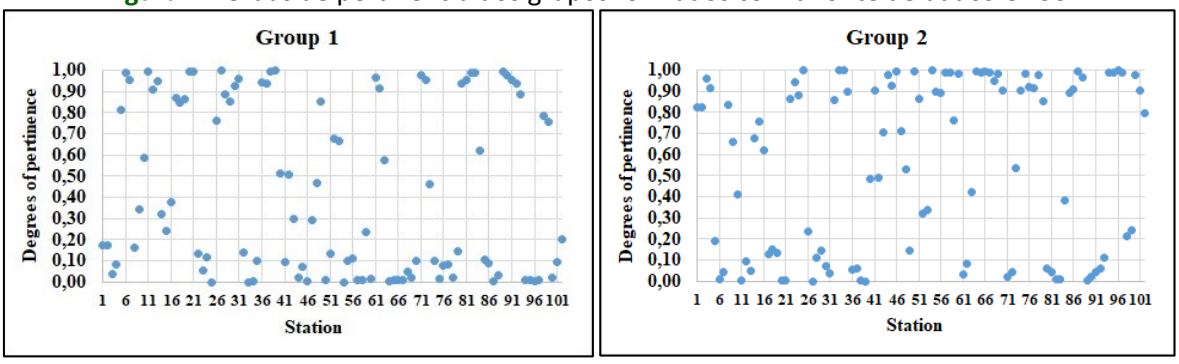
| Nº de Grupos | GPCC | | | ANA | | |
|---|---|---|---|---|---|---|
| | Dunn | Silhouette | PBM | Dunn | Silhouette | PBM |
| **2** | **0.13** | 0.33 | **2.98** | 0.17 | **0.42** | **3.46** |
| 3 | 0.11 | 0.33 | 2.41 | **0.19** | 0.37 | 2.51 |
| 4 | 0.11 | **0.40** | 2.14 | 0.09 | 0.27 | 1.83 |
| 5 | 0.09 | 0.35 | 2.16 | 0.09 | 0.36 | 2.90 |
| 6 | 0.11 | 0.33 | 2.2 | 0.14 | 0.35 | 2.57 |
| 7 | 0.11 | 0.32 | 2.02 | 0.15 | 0.41 | 2.55 |
| 8 | 0.11 | 0.31 | 1.71 | 0.15 | 0.39 | 2.28 |
| 9 | 0.11 | 31 | 1.71 | 0.15 | 0.39 | 2.28 |
| 10 | 0.11 | 0.26 | 1.28 | 0.17 | 0.36 | 1.96 |

**Fonte:** Autores.

A escolha do melhor cluster foi definida pelos valores mais altos dos coeficientes, já que todos tendem a maximizar o melhor resultado, assim, para os dados GPCC, os índices Dunn e PBM geraram um cluster ideal para 2 grupos, com valores respectivos de 0,13 e 2,98, e no índice Silhouette, 0,40, com formação ideal para 4 grupos. Para os dados fornecidos pela ANA, o índice Dunn resultou em um valor de coeficiente de 0,19, Silhouette e PBM apresentaram um cluster melhor para 2 clusters, com índices de 0,42 e 3,46 respectivamente. Portanto, embora o índice Dunn e Silhouette apresentassem formações diferentes para suas fontes de dados, o índice PBM apontava para 2 grupos, porém, como nos testes a melhor formação independentemente do método era para 2 clusters, logo se chega à conclusão de que a formação ideal era para 2 arranjos.

Em posse do número ideal para a formação dos grupos, para os dados do GPCC, eles foram organizados de acordo com o maior grau de relevância de cada estação, ou seja, as estações foram subdivididas em grupos com maior afinidade entre si. A Figura 4 ilustra todas as estações GPCC com seus respectivos graus de relevância para os dois grupos, na qual pode-se ver que para o grupo 1 algumas estações têm maior afinidade com o grupo 2 e vice-versa, portanto quanto mais próximas de 1, maior sua relação com o grupo formado.

**Figura 4 -** Graus de pertinência dos grupos formados com a fonte de dados GPCC.




**Fonte:** Autores.

Com a formação dos respectivos grupos e, consequentemente, com a subdivisão das estações de acordo com seu grau de relevância (Tabela 4), dois grupos distintos foram formados entre si, mas com indivíduos internos (estações) semelhantes, para mais tarde espacializá-los no ambiente GIS com a aplicação da técnica de Kriging.

Para o grupo 1 do GPCC, 44 estações foram agrupadas, com uma precipitação média anual de 2321,00 mm e para o grupo 2, 58 estações com uma precipitação média anual de 1920,98 mm.

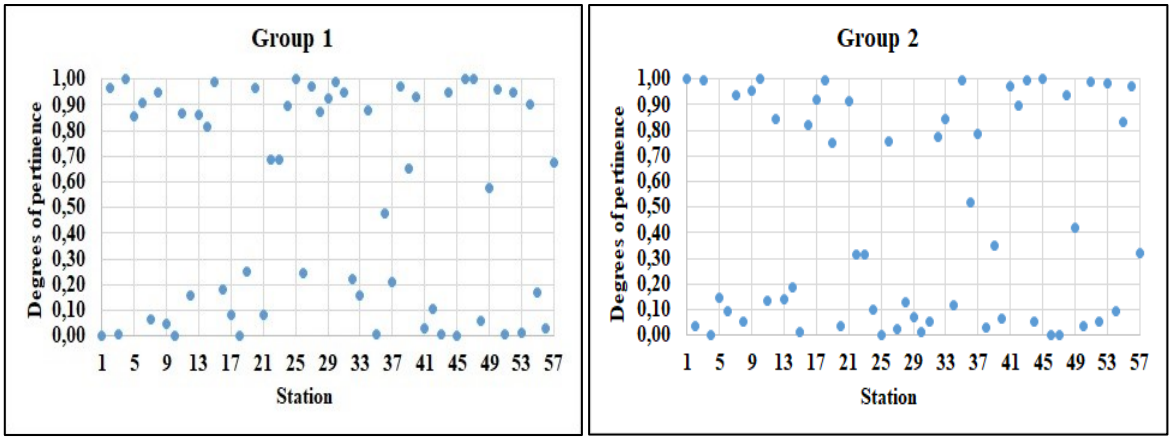
**Tabela 4 -** Graus de relevância de cada estação GPCC com a formação de seu respectivo grupo.

| Grupo 1 - GPCC | | | | Grupo 2 - GPCC | | | |
|---|---|---|---|---|---|---|---|
| **Id** | **Graus** | **Id** | **Graus** | **Id** | **Graus** | **Id** | **Graus** |
| 5 | 0.8115 | 63 | 0.5759 | 1 | 0.8270 | 57 | 0.9919 |
| 6 | 0.9883 | 71 | 0.9770 | 2 | 0.8276 | 58 | 0.9905 |
| 7 | 0.9545 | 72 | 0.9554 | 3 | 0.9592 | 59 | 0.7656 |
| 10 | 0.5869 | 80 | 0.9393 | 4 | 0.9154 | 60 | 0.9825 |
| 11 | 0.9957 | 81 | 0.9537 | 8 | 0.8392 | 64 | 0.9949 |
| 12 | 0.9080 | 82 | 0.9919 | 9 | 0.6588 | 65 | 0.9919 |
| 13 | 0.9490 | 83 | 0.9919 | 14 | 0.6784 | 66 | 0.9924 |
| 17 | 0.8723 | 84 | 0.6188 | 15 | 0.7596 | 67 | 0.9883 |
| 18 | 0.8486 | 89 | 0.9934 | 16 | 0.6225 | 68 | 0.9511 |
| 19 | 0.8658 | 90 | 0.9775 | 22 | 0.8664 | 69 | 0.9809 |
| 20 | 0.9938 | 91 | 0.9558 | 23 | 0.9463 | 70 | 0.9021 |
| 21 | 0.9946 | 92 | 0.9379 | 24 | 0.8822 | 73 | 0.5350 |
| 26 | 0.7613 | 93 | 0.8859 | 25 | 0.9993 | 74 | 0.9022 |
| 27 | 0.9999 | 98 | 0.7871 | 32 | 0.8614 | 75 | 0.9848 |
| 28 | 0.8893 | 99 | 0.7578 | 33 | 0.9989 | 76 | 0.9232 |
| 29 | 0.8519 | - | - | 34 | 0.9978 | 77 | 0.9146 |
| 30 | 0.9270 | - | - | 35 | 0.8975 | 78 | 0.9769 |
| 31 | 0.9621 | - | - | 41 | 0.9041 | 79 | 0.8548 |
| 36 | 0.9450 | - | - | 43 | 0.7036 | 85 | 0.8944 |
| 37 | 0.9393 | - | - | 44 | 0.9793 | 86 | 0.9120 |
| 38 | 0.9932 | - | - | 45 | 0.9270 | 87 | 0.9953 |
| 39 | 0.9998 | - | - | 46 | 0.9963 | 88 | 0.9691 |
| 40 | 0.5152 | - | - | 47 | 0.7094 | 94 | 0.9915 |
| 42 | 0.5089 | - | - | 48 | 0.5315 | 95 | 0.9876 |
| 49 | 0.8526 | - | - | 50 | 0.9925 | 96 | 0.9985 |
| 52 | 0.6770 | - | - | 51 | 0.8647 | 97 | 0.9909 |
| 53 | 0.6643 | - | - | 54 | 0.9997 | 100 | 0.9784 |
| 61 | 0.9690 | - | - | 55 | 0.8980 | 101 | 0.9040 |
| 62 | 0.9138 | - | - | 56 | 0.8909 | 102 | 0.7988 |

**Fonte:** Autores.

O mesmo procedimento foi aplicado aos dados das estações ANA, Figura 5, onde eles foram mostrados com seus respectivos graus de relevância para os dois grupos formados, e quanto mais próximo de 1, maior a relação entre este e o grupo.

**Figure 5 -** Graus de pertinência dos grupos formados com a fonte de dados ANA.




**Fonte:** Autores.

Assim, também foram formados dois grupos distintos, mas com indivíduos internos (estações) semelhantes, para mais tarde espacializá-los, no qual o grupo 1 tinha 31 estações, com precipitação média anual de 2169,00 mm, e no grupo 2, 26 estações foram agrupadas, com precipitação média anual de 2640,16 mm.

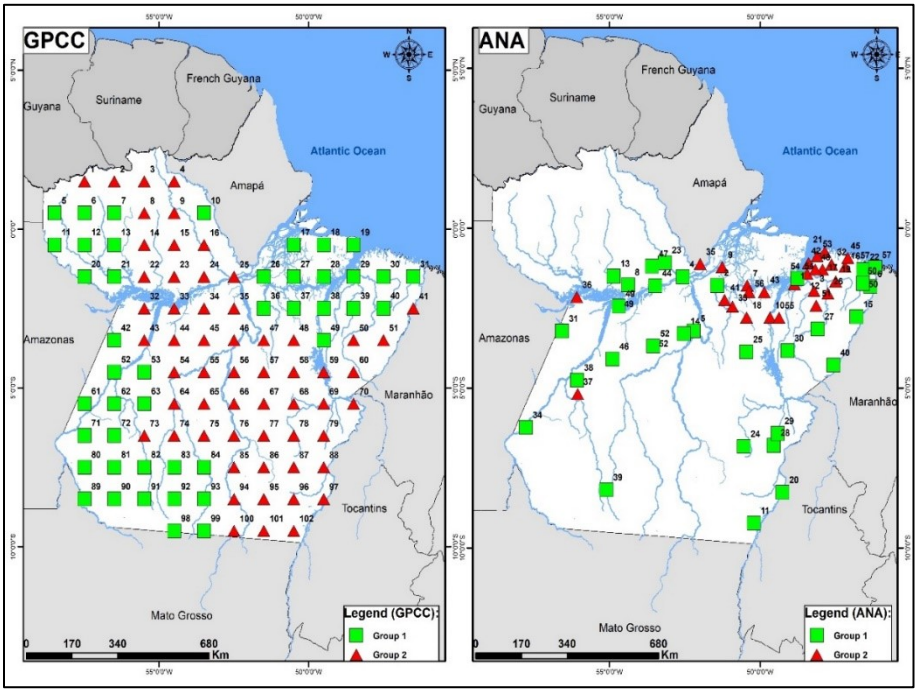
**Tabela 5 -** Graus de relevância de cada estação ANA com a formação de seu respectivo grupo.

| Grupo 1 - ANA | | | | Grupo 2- ANA | | | |
|---|---|---|---|---|---|---|---|
| **Id** | **Graus** | **Id** | **Graus** | **Id** | **Graus** | **Id** | **Graus** |
| 2 | 0.9662 | 29 | 0.9276 | 1 | 0.9997 | 37 | 0.7872 |
| 4 | 0.9999 | 30 | 0.9862 | 3 | 0.9927 | 41 | 0.9706 |
| 5 | 0.8548 | 31 | 0.9475 | 7 | 0.9368 | 42 | 0.8958 |
| 6 | 0.9059 | 34 | 0.8806 | 9 | 0.9512 | 43 | 0.9943 |
| 8 | 0.9462 | 38 | 0.9697 | 10 | 0.9997 | 45 | 1.00 |
| 11 | 0.8657 | 39 | 0.6511 | 12 | 0.8426 | 48 | 0.9384 |
| 13 | 0.8617 | 40 | 0.9322 | 16 | 0.8204 | 51 | 0.9909 |
| 14 | 0.8144 | 44 | 0.9464 | 17 | 0.9166 | 53 | 0.9845 |
| 15 | 0.9896 | 46 | 0.9980 | 18 | 0.9965 | 55 | 0.8330 |
| 20 | 0.9645 | 47 | 0.9998 | 19 | 0.7508 | 56 | 0.9725 |
| 22 | 0.6869 | 49 | 0.5778 | 21 | 0.9150 | - | - |
| 23 | 0.6870 | 50 | 0.9616 | 26 | 0.7549 | - | - |
| 24 | 0.8981 | 52 | 0.9462 | 32 | 0.7758 | - | - |
| 25 | 0.9977 | 54 | 0.9040 | 33 | 0.8420 | - | - |
| 27 | 0.9735 | 57 | 0.6774 | 35 | 0.9955 | - | - |
| 28 | 0.8714 | - | - | 36 | 0.5204 | - | - |

**Fonte:** Autores.

Plotar estes resultados em um ambiente SIG, sem espacializá-los, resultou na formação de dois grupos, para duas fontes de dados distintas, para o estado do Pará (Figura 6). Claramente a grade de estações GPCC propõe uma região mais central homogênea, dividindo na porção ocidental e parte do nordeste outra região homogênea. Para os dados da ANA, uma parte da região nordeste é destacada como homogênea, e o restante é considerado como outra região.

**Figura 6 -** Formação de agrupamentos pelo mais alto grau de pertinência de cada estação.



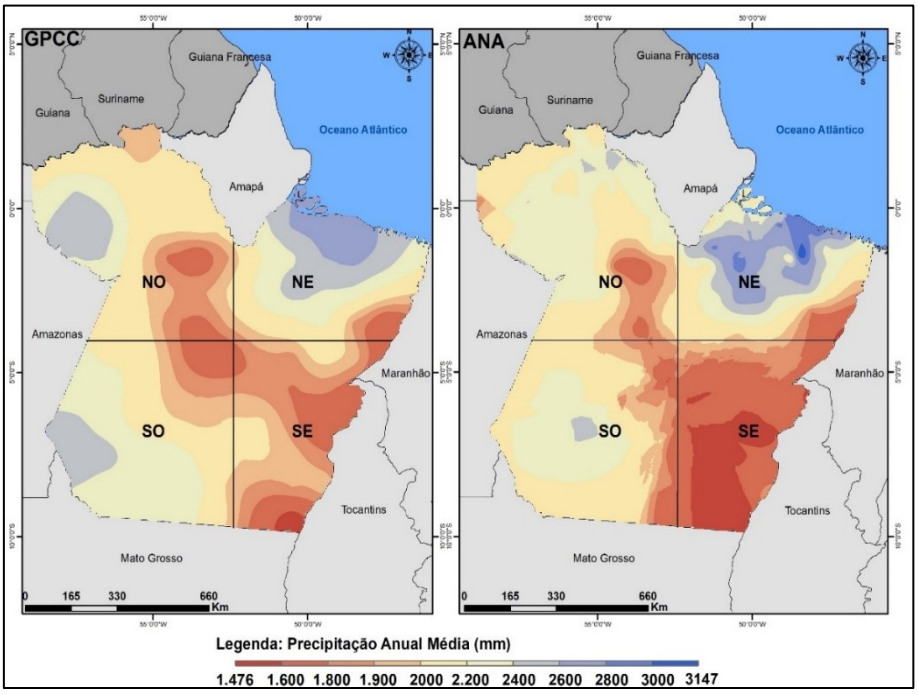

**Fonte:** Autores.

Quando estes dados foram espacializados para a formação de regiões homogêneas de precipitação, observou-se que ambos estão bem distribuídos em toda a área de estudo (Figura 7), com maior semelhança na parte nordeste do estado, onde, segundo Albuquerque et al. (2010), este fato se deve ao desempenho de fenômenos climatológicos, como a Zona de Convergência Intertropical (ZCIT), que opera nesta região.

Ferreira Filho et al. (2019) fizeram análises comparativas da variabilidade pluviométrica (Figura 7) para as mesmas duas fontes de dados e encontraram resultados semelhantes aos obtidos por Amaral et al. (2016) e Albuquerque et al. (2010), os quais

constataram que a maior precipitação média anual foi na parte nordeste do estado, cobrindo áreas dentro destas três mesorregiões, bem como uma porção na parte baixa da Amazônia.

**Figure 7 -** Precipitação anual Média de longo prazo do Pará (1986 a 2015) para ambas as fontes de dados.



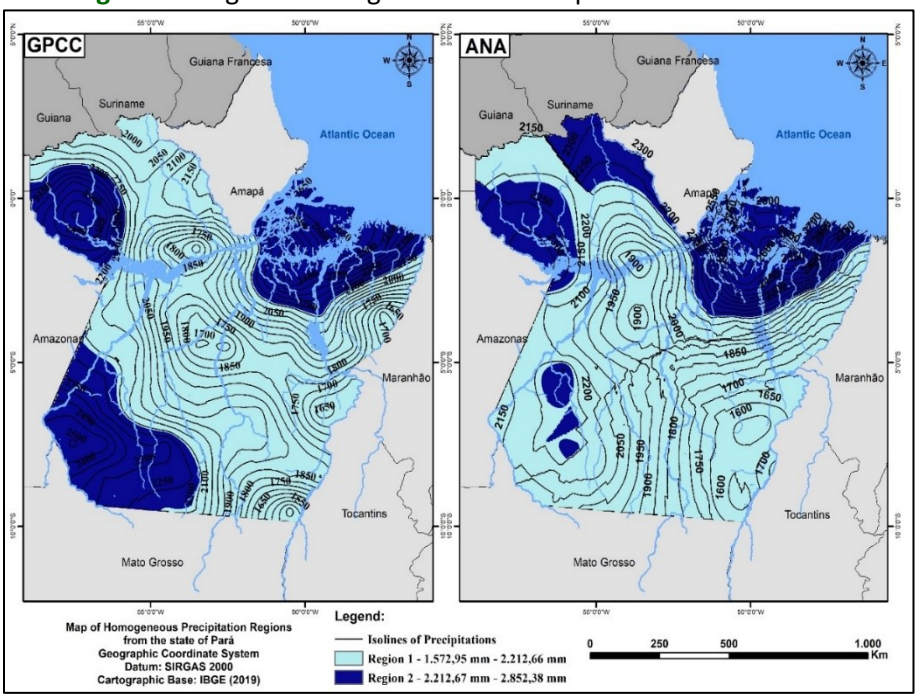

**Fonte:** Autores.

A região norte do estado, localizada na porção noroeste, é considerada uma região homogênea de chuvas (PESSOA; BLANCO; MARTINS, 2011; FERREIRA FILHO et al., 2019a), por este motivo, os comportamentos em ambas as fontes de dados ocorrem de maneira semelhante. Pessoa, Blanco e Martins (2011) ainda afirmaram que este comportamento está diretamente relacionado à vegetação densa da área, uma vez que é considerada uma região pouco explorada, contribuindo para altas taxas de umidade e, consequentemente, altas taxas de pluviosidade.

A porção sudoeste do estado foi onde houve maior divergência entre os dados analisados, de modo que, na região homogênea de chuvas formada pelo GPCC, houve uma maior concentração da região 2, enquanto na análise formada pelos dados da ANA, houve fragmentos da região 2. Este fato pode ser justificado em um estudo de Ferreira Filho et al. (2020), que afirmaram que esta porção do estado tem áreas de baixa densidade de estações

ANA, enquanto que o GPCC tem uma distribuição uniforme da rede de dados na região. Apesar das diferenças de comportamento das regiões homogêneas nesta área, elas apresentam índices pluviométricos bastante semelhantes, com diferenças de apenas 34 mm entre elas. Com a falta de dados para esta porção, a grade de dados do GPCC torna-se mais completa para a obtenção de informações nesta porção do estado.

**Figura 8 -** Regiões Homogêneas de Chuva para o Estado do Pará.




**Fonte:** Autores.

Em um estudo desenvolvido por Silva et al. (2018) na bacia do Tapajós no estado do Pará, foi observada a formação de 3 regiões pluviométricas homogêneas, embora utilizassem um método diferente do presente estudo, mas que se aproxima do resultado obtido utilizando os dados do GPCC (2 regiões homogêneas).

Visualmente a porção central do estado é melhor representada em ambos os casos, já que as densidades de estação nas duas fontes de dados estão distribuídas nesta região, podendo assim afirmar que é a região com as maiores informações de precipitação. Apesar destas distinções, ambas apresentam comportamentos semelhantes.

Um resultado semelhante foi obtido por Limberger; Silva (2018) no qual utilizaram dados do GPCC e observaram que seguiram a variabilidade de precipitação mensal dos dados de ANA, sendo mais evidente nas sub-regiões Norte e Central da região Amazônica. Em geral,

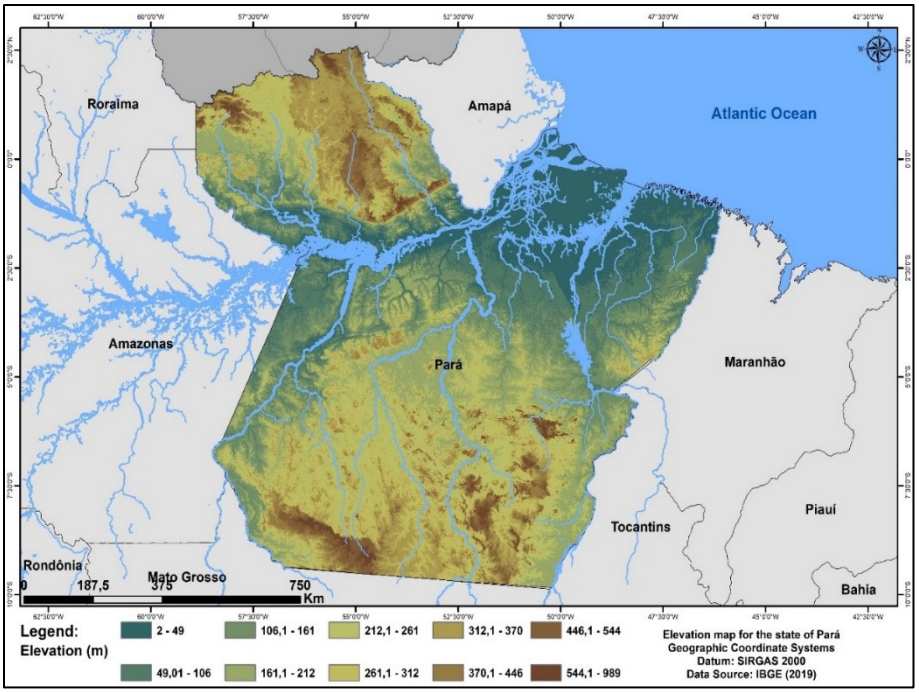
nesta área estudada, os produtos GPCC mostram um desempenho hidrológico comparável ao produto de ANA após o tempo real, apresentando o grande potencial para aplicação em tempo real.

Ao relacionar as regiões formadas com a vegetação presente na área de estudo, pode-se observar que na região 1, há uma diversidade na vegetação, onde na porção sudeste, concentram-se as menores taxas de precipitação juntamente com o Arco do Desmatamento, com a presença de vegetação esparsa, esparsa, e em alguns pontos, sem vegetação, fato devido ao aumento da pecuária, da agricultura relacionada com a expansão da produção de soja e madeira (COSTA; PIRES, 2010; LEMOS;SILVA, 2011), outro fator que pode associar este índice é devido à presença de fenômenos climáticos, El Niño e La Niña (GONÇALVES et al. , 2016). Quando se analisa a porção central do estado, nota-se uma vegetação intermediária e densa, indo para a região noroeste, outra característica intrínseca a este grupo formado, que deve ser observada, que apesar da variabilidade da precipitação neste grupo, nota-se também a variabilidade da vegetação.

Em relação ao grupo homogêneo 2, a presença de vegetação intermediária e densa, em sua maioria territorial, pode justificar a formação de chuvas com maiores índices de precipitação, sendo o principal destaque a região da Calha Norte, como um ataque anterior. Observa-se a concentração de estações ANA na porção sul do estado, devido a áreas de difícil acesso, presença de vegetação densa e altitudes elevadas (Figura 10), assim, a implantação de estações de monitoramento nesta área torna-se difícil, outro fato que justifica a utilização dos dados da grade GPCC para estudos hidrológicos.

Ao comparar regiões com elevação (Figura 9), esta foi uma variável de entrada no modelo de agrupamento, portanto, também influenciou os resultados; portanto, a região 1 foi a de maiores altitudes, outro fato que justifica que tenha menores taxas de precipitação, e tendo em vista a região homogênea 2, pode-se observar que, devido à presença de baixas altitudes, com exceção da porção sudoeste, o agrupamento ocorre em áreas mais próximas ao nível do mar.

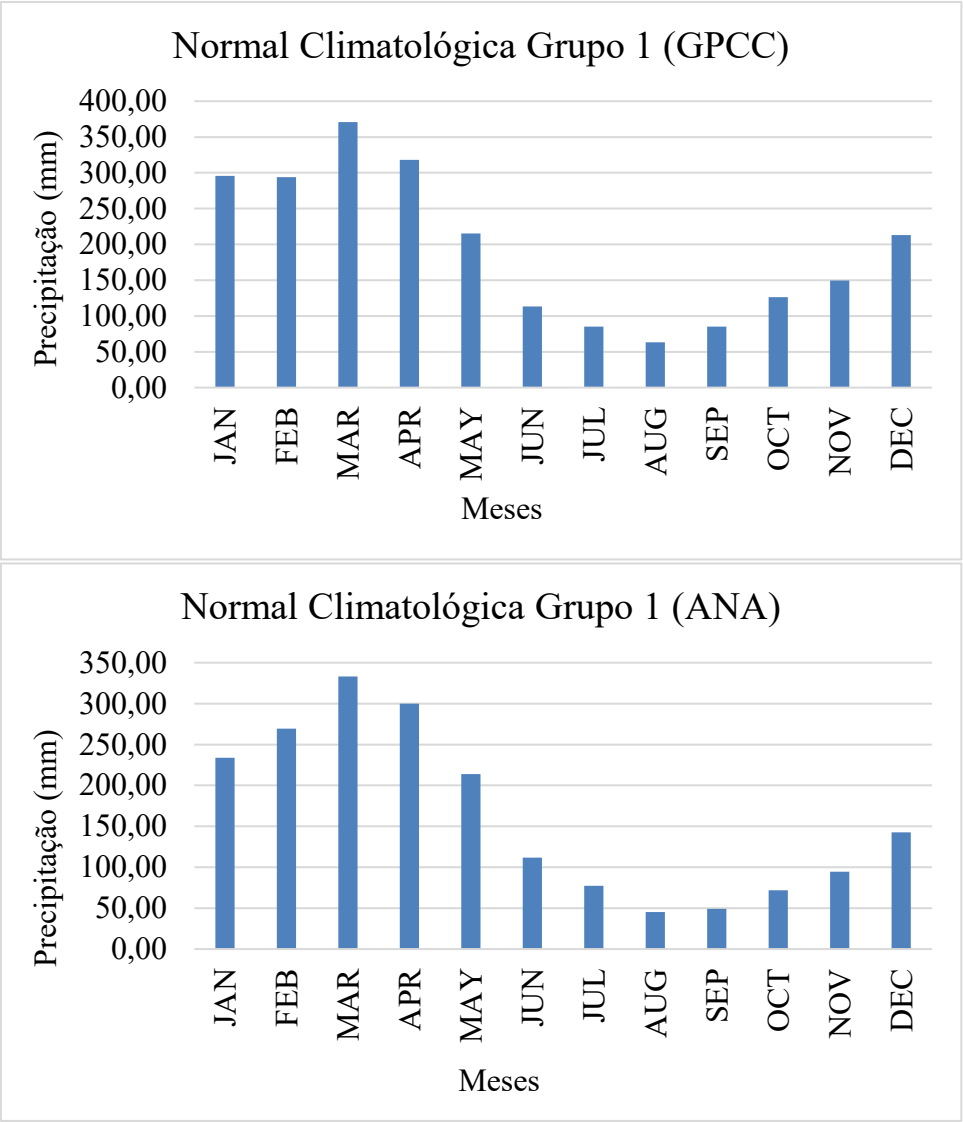
**Figura 9 –** Mapa de elevação em metros para o estado do Pará.




**Fonte:** Autores.

Na análise da distribuição pluviométrica da região homogênea 1 (Figura 10), ao longo dos meses, pode-se observar que em ambos os meses o mês com maior índice pluviométrico é registrado em março, semelhante ao encontrado por Amanajás e Braga (2012); Menezes, Fernandes e Rocha (2015) e Ferreira Filho et al. (2020), com o início de uma estação chuvosa de novembro a abril, e uma estação menos chuvosa de maio a outubro. Uma das poucas diferenças entre os dados formados pelas duas fontes foi para o mês de janeiro, onde o GPCC obteve um índice pluviométrico maior do que os dados da ANA, fato que se justifica porque o GPCC superestimou entre janeiro e agosto (FERREIRA FILHO et al., 2020).

**Figura 10 -** Normas climáticas GPCC e ANA respectivamente para região homogênea 1.



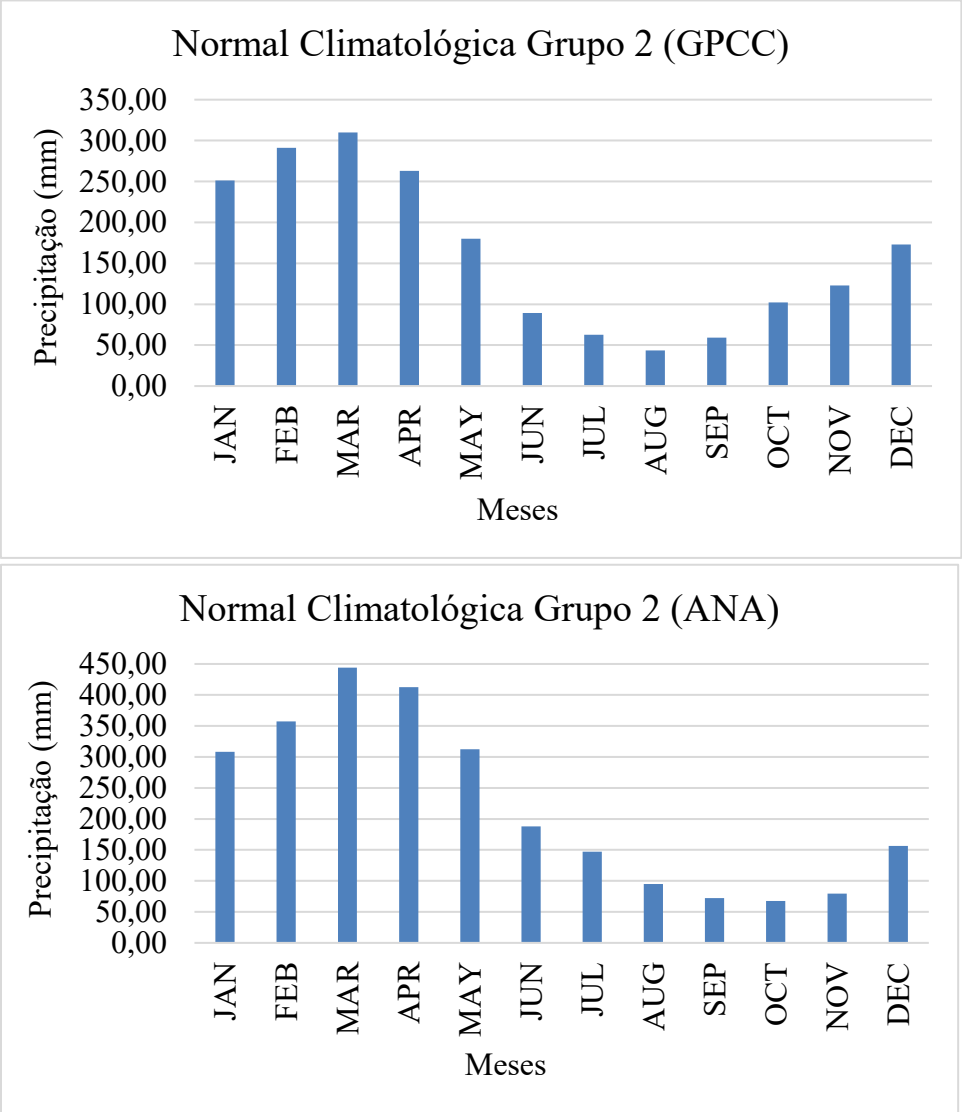

**Fonte:** Autores.

Em um estudo aplicado por Menezes, Fernandes e Rocha (2015), o comportamento das distribuições climatológicas normais mostrou semelhanças com o da pesquisa, confirmando assim que o estudo era válido.

A mesma análise foi feita para a região homogênea 2, representada pela Figura 11, onde março também foi registrado como o mês mais chuvoso. Nesta região, a divergência entre as normais ocorreu no topo do mês menos chuvoso, para o GPCC, foi o mês de agosto, idêntico ao estudo de Ferreira Filho et al., (2019), e para os dados da ANA, o mês de outubro. Vale notar que esta região divergiu mais na porção sudoeste do estado, devido à baixa densidade de informações e estações da ANA em relação à grade do GPCC, portanto esta pode

ter sido a maior influência no comportamento da região formada, juntamente com o ZCIT (RIBEIRO et al., 1996).

**Figura 11–** Normas climáticas GPCC e ANA respectivamente para região homogênea 2.




**Fonte:** Autores.

Assim, para verificar o desempenho das estimativas das regiões formadas pela ANA e GPCC, a partir da espacialização, foi gerada uma malha de 1.000 pontos aleatórios e foram extraídos os valores dos pontos de precipitação, nos quais os testes estatísticos foram analisados e aplicados (Tabela 6).

**Tabela 6 -** Resultados dos testes estatísticos de desempenho das regiões formadas.

| Métodos | Resultados | Intervalo | Bom valor |
|---|---|---|---|
| **R** | 0,82 | -1 to 1 | 1 |
| **$R^2$** | 0,67 | 0 to 1 | 1 |
| **MSE** | 24936,40 | 0 to ∞ | 0 |
| **RMSE** | 157,91 | 0 to ∞ | 0 |
| **NASH** | 0,66 | - ∞ to 1 | 1 |
| **Standard Error** | 143,10 | - | - |

**Fonte:** Autores.

Vários índices estatísticos amplamente utilizados foram usados para avaliar quantitativamente a precisão dos valores dos dados gerados por ANA e GPCC com relação à precipitação, mostrando que os resultados de desempenho apresentaram uma boa correlação entre os dados, nos quais o coeficiente R apresentou um valor considerado positivo e bom, $R^2$ como valor moderado, com bons resultados entre os dados. Quanto aos erros, foi obtido um valor de 143,10 mm como erro padrão, o que é aceitável, considerando que se trata de dados médios anuais. Os métodos de MSE e RMSE apresentaram valores de 24.936,0 e 157,91. O MSE é sensível a grandes erros, porque quadratura as diferenças individuais, mas é sempre positivo, e quanto mais próximo de zero, melhor indica uma simulação perfeita. Neste caso, como trabalhamos com dados anuais, o MSE apresentou um valor considerado bom em termos estatísticos, evidenciado no teste estatístico do RMSE, com um valor raiz de 157,91.

Finalmente, o coeficiente NASH, 0,66, de acordo com a classificação de Pereira et al. (2016) e Costa et al. (2019), apresentou um valor aceitável, para análise de desempenho estatístico. Portanto, os coeficientes mostraram que as regiões formadas têm uma boa correlação entre os dados analisados, finalmente, ilustrados no gráfico de probabilidade normal dos dados.

## 4. CONCLUSÕES

O estudo fez contribuições importantes para a formação de regiões pluviométricas homogêneas para o estado do Pará. Constatou-se que o monitoramento hidrometeorológico

na região apresenta problemas como pequena cobertura de dados, baixa densidade de estações e outras com séries incompletas, podendo ser uma das principais influências na distinção das regiões apresentadas.

Como os dados do GPCC apresentam uma grade mais completa de informações, sugere-se, como pesquisado e visto em outros trabalhos, o uso destes para obter dados de precipitação em áreas com poucos dados disponíveis pela Agência Nacional de Águas e Saneamento Básico- ANA.

Este fator foi evidenciado através de técnicas estatísticas (R, $R^2$, MSE, RMSE e NASH), cujos valores apresentaram uma boa correlação entre as regiões formadas, com os dados utilizados para a reanálise (GPCC) e ANA, analisando principalmente a porção sudoeste do estado, com baixa informação das estações de monitoramento.

Tanto para as regiões 1 e 2 formadas, em ambas as fontes de dados, as normas climatológicas obedecem a sua variabilidade conforme verificado em outros estudos apresentados, demonstrando que apesar do estado ser considerado grande, as comparações entre as duas fontes de dados se tornaram ferramentas importantes sobre estudos hidrológicos.

Vale ressaltar que quaisquer conjuntos de dados, observados ou simulados, sejam ou não de reanálise, podem apresentar problemas de consistência de dados devido a falhas na obtenção ou registro, assim como as ferramentas estatísticas empregadas, entre outros fatores. Por esta razão, a comparação entre as duas fontes torna-se de grande valor.

Em geral, os resultados para o estado do Pará indicaram que os dados do GPCC tiveram um bom desempenho em relação aos dados observados, levando-o a indicar o uso dos dados do GPCC como uma fonte alternativa de dados de precipitação em locais sem estações de chuva ou séries longas de dados.

## REFERÊNCIAS

AGUADO, A. G.; CANTANHEDE, Marco André. Lógica **Fuzzy. 2010.** Disponível em: < http://www.sysrad.com.br/redmine/attachments/1843/Artigo_logicaFuzzi.pdf >.

AHANI, A.; NADOUSHANI, S. M. Avaliação de algumas combinações de técnicas de agrupamento hard e fuzzy para regionalização de bacias hidrográficas na bacia de Sefidroud. **Journal of Hydroinformatics,** v. 18, n. 6, pág. 1033–1054. 2016.

de ALBUQUERQUE, M. F.; DE SOUZA, E. B.; DE OLIVEIRA, M.D.C.F.; DE SOUZA JÚNIOR, J. A. Precipitação nas mesorregiões do estado do Pará: climatologia, variabilidade e tendências nas últimas décadas (1978-2008). **Revista Brasileira de Climatologia**, n. 6. 2010.

AMANAJÁS, J. C.; BRAGA, C. C. Padrões espaço-temporais de precipitação na Amazônia oriental usando análise multivariada. **Revista Brasileira de Meteorologia,** v. 27, n. 4, p.423 – 434. 2012.

ARELLANO-LARA, F.; ESCALANTE-SANDOVAL, C. A. Delineamento multivariado de regiões homogêneas de precipitação para estimar quantis de precipitação máxima diária: um estudo de caso do noroeste do México. **Atmosfera,** v. 27, n. 1, pág. 47-60. 2014.

ASONG, Z. E.; KHALIQ, M. N.; WHEATER, H. S. Regionalização das características de precipitação nas províncias das pradarias canadenses usando covariáveis atmosféricas de larga escala e atributos geofísicos. **Estocástico Env. Res. e Risk Asses.,** v. 29, n. 3, pág. 875-892. 2015.

ÁVILA, P. L. R.; DE SOUZA, E. B.; PINHEIRO, A. N.; FIGUEIRA, W. S. Análise da precipitação sazonal simulada usando regcm4 sobre o estado do Pará em anos de extremos climáticos. **Revista Brasileira de Climatologia,** n. 14. 2014.

BEZDEK, James C. Reconhecimento de padrões com algoritmos fuzzy de função objetivo. **Plenum Press,** Nova York. 1981.

CAI, Y.; JIN, C.; WANG, A.; GUAN, D.; WU, J.; YUAN, F.; XU, L. Análise espaço-temporal da precisão da análise de precipitação multissatélite tropical 3B42 dados de precipitação em altas latitudes médias na China. **PloS One,** v. 10, n. 4, e0120026. 2015.

COSTA, J.C.; PEREIRA, G.; SIQUEIRA, M. E.; DA SILVA CARDOZO, F.; DA SILVA, V. V. Validação dos dados pluviométricos estimados pelo CHIRPS para o Brasil. **Revista Brasileira de Climatologia**, v. 24, p. 228-243. 2019.

COSTA, M.H.; PIRES, G. F. Efeitos dos cenários de desmatamento da Amazônia e do Brasil Central sobre a duração da estação seca no arco do desmatamento. **Jornal Internacional de Climatologia**, v. 30, n. 13, pág. 1970-1979. 2010.

DAVIDSON, E. A.; DE ARAUJO, A. C.; ARTAXO, P.; BATCH, J. K.; BROWN, I. F.; BUSTAMANTE, M. M.; e outros A bacia amazônica em transição. **Natureza,** v. 481, pág. 321-328. 2012.

DEZFULI, A. K. Variabilidade espaço-temporal da precipitação sazonal na África Equatorial Ocidental. **Climatologia teórica e aplicada**, v. 104, n. 1-2, pág. 57-69. 2011.

DINKU, T.; CONNOR, S.J.; CECCATO, P.; ROPELEWSKI, C.F. Comparação de produtos globais de precipitação em grade em uma região montanhosa da África. **Climatologia Internacional**, v. 11, p. 2960–2979. 2008.

Do AMARAL, M.A.C.M.; JOSÉ, J. V.; FOLEGATTI, M. V.; COELHO, R. D.; BARROS, T. H. S. Distribuição espacial da precipitação em relação à topografia no estado do Pará. **Irriga,** v.1, n. 1, p.1-10. 2016.

DUNN, J. C. Um parente difuso do processo ISODATA e seu uso na detecção de clusters compactos bem separados. **Cibernética e Sistemas,** v. 3, p. 32-5. 1973.

FAZEL, N.; BERNDTSSON, R.; UVO, C.B.; MADANI, K.; KLØVE, B. Regionalização das características de precipitação na bacia do Lago Urmia no Irã**. Climatologia Teórica e Aplicada,** v. 132, n. 1-2, pág. 363-373. 2018.

FERREIRA FILHO, D. F.; BEZERRA, P. E. S..; SILVA, M. de N. A. da; RODRIGUES, R. S. S. .; de FIGUEIREDO, N. M. Aplicação de técnicas de interpolação para espacialização das chuvas na região hidrográfica da Calha Norte, Pará. **Revista Brasileira de Climatologia,** v. 24, p. 341-363. 2019.

FERREIRA FILHO, D. F.; LIRA, B. R. P. .; CRISPIM, D. L.; PESSOA, F. C. L. .; FERNANDES, L. L. Análise de chuvas no estado do Pará: comparação entre dados obtidos de estações pluviométricas e do satélite GPCC. **Revista Brasileira de Climatologia,** v. 26. 2020. See More

FITZJARRALD, D. R; SAKAI, R. K.; MORAES, O. L.; COSME DE OLIVEIRA, R.; ACEVEDO, O. C.; CZIKOWSKY, M. J.; BELDINI, T. Variabilidade espacial e temporal da precipitação próxima à confluência Amazon-Tapajós. **Jornal de Geophys. Res.,** v. 113. 2008.

De OLIVEIRA GIL, V.; FERRARI, F.; EMMENDORFER, L. Pesquisa sobre a aplicação de algoritmos de agrupamento para o problema astrofísico de classificação de galáxias. **Revista Brasileira de Computação Aplicada, v**. 7, n. 2, pág. 52-61. 2015.

GONÇALVES, M. F.; BLANCO, C.J.C.; DOS SANTOS, V. C.; DOS SANTOS OLIVEIRA, L. L.; PESSOA, F. C. L. Identificação de Regiões Homogêneas de Precipitação considerando El Niño e La Niña e Redução de Precipitação no Estado do Pará, Amazônia Brasileira. **Acta Scientiarum. Tecnologia,** v. 38, n. 2, pág. 209-216. 2016.

HALKIDI, M.; BATISTAKIS, Y.; VARGIANNIS, M. **Métodos de validade de agrupamento: Parte. I**. Registro ACM SIGMOD, v. 31, n. 2. 2002.

IBGE (2010). Censo 2010. https://cidades.ibge.gov.br. Acesso em 10 de agosto de 2022.

ISHIHARA, J. H.; FERNANDES, L. L.; DUARTE, A. A. A. M.; DUARTE, A.R.C.L.M.; PONTE, M. X.; LOUREIRO, G. E. Avaliação Quantitativa e Espacial da Precipitação na Amazônia Brasileira (Amazônia Legal) - (1978 a 2007). **Revista Brasileira de Recursos Hídricos.,** Porto Alegre, v. 19, p. 29-39. 2014.

KÖPPEN, W.; GEIGER, R. Klimate der Erde. Gotha: Verlagcondicionadas. Justo Perthes, 1928.

LEMOS, A. L. F.; SILVA, J. de A. Desmatamento na Amazônia Legal: evolução, causas, monitoramento e possibilidades de mitigação por meio do Fundo Amazônia. **Floresta e Meio Ambiente,** v. 18, n. 1, pág. 98-108. 2012.

LIMBERGER, L.; SILVA, M. E. S. Precipitação observada na Amazônia brasileira: redes convencionais e dados da Reanálise I do NCEP/NCAR, CRU e GPCC. **Revista Brasileira de Climatologia,** v. 22, ed. Jan/Jun. 2018.

MENEZES, F.P.; FERNANDES, L. L.; DA ROCHA, E. J. P.. O Uso da Estatística para a Regionalização da Precipitação no Estado do Pará, Brasil. **Revista Brasileira de Climatologia,** v. 16, p. 64-71. 2015.

MENEZES, F.; FERNANDES, L. Análise de tendência e variabilidade da precipitação no Estado do Pará. **Enciclopédia da Biosfera,** v. 13, no. 24 de 2016.

NADOUSHANI, S. S. M.; DEHGHANIAN, N.; SAGHAFIAN, B. Um método de agrupamento híbrido difuso para identificar regiões hidrológicas homogêneas. **Journal of Hydroinformatics,** v. 20, n.6. 2018.

NASH, J. E.; SUTCLIFFE, J. V. Previsão de vazão de rios por meio de modelos conceituais parte I – uma discussão de princípios. **Journal of Hydrology (Amsterdã),** v. 10, n. 3, pág. 282-290. 1970.

NOBRE, C.; JOVEM, A. F.; SALDIVA, P. H. N.; ORSINI, J. A. M.; NOBRE, A. D.; OGURA, A. T.; e outros. **Vulnerabilidade das Megacidades Brasileiras às Mudanças Climáticas: a Região Metropolitana de São Paulo (RMSP).** Mudanças Climáticas no Brasil: aspectos econômicos, sociais e regulatórios. Brasília: IPEA., v., p. 197-219. 2011.

PAKHIRA, M. K.; BANDYOPADHYAY, S.; MAULIK, U. Índice de validade para agrupamentos nítidos e difusos. **Reconhecimento de padrões,** n. 37, p.481-501.2004.

PARCHURE, A. S.; GEDAM, S. K. Regionalização homogênea via momentos L para a cidade de Mumbai, Índia. **Meteorologia Hidrologia e Gestão da Água,** v. 7, n. 2, pág. 73 – 83. 2019.

PEDRYCZ, W.; VUKOVICH, G. Agrupamento difuso com supervisão. Reconhecimento de padrões. **O Jornal da Sociedade de Reconhecimento de Padrões**, v.37, p. 1339-1349. 2004.

PEREIRA, D. D. R.; ULIANA, E. M.; MARTINEZ, M.A; DA SILVA, D. D. Desempenho de um modelo hidrológico concentrado e semidistribuído na previsão de vazões diárias. **Irriga, Botucatu,** v. 21, n. 2, p.409-424. 2016.

PESSOA, F.C.L.; BLANCO, C.J.C.; MARTINS, J. R. Regionalização de curvas de permanência de vazão na região da Calha Norte no Estado do Pará. **Revista Brasileira de Recursos Hídricos,** v. 16, p. 65-74. 2011.

RAZIEI, T.; DARYABARI, J.; BORDI, I.; PEREIRA, L. S. Padrões espaciais e tendências temporais de precipitação no Irã. **Theor Appl Climatol,** v. 115, p. 531–540. 2014.

RENCHER, A. C.; CHRISTENSEN, W. F. Métodos de análise multivariada. **Nova Jersey: John Wiley and Sons,** 2012. 768 p.

RIBEIRO, A.; VITÓRIA, R. L.; PEREIRA, A. R.; VILLA NOVA, N. A.; MARTINELLI, L. A.; MORTATTI, J. Análise do regime pluviométrico da Amazônia a partir de dados de onze localidades. **Revista Brasileira de Meteorologia,** v. 11, p. 25 – 35. 1996.

ROUSSEEUW, P. J. Silhuetas: uma ajuda gráfica para a interpretação e validação da análise de cluster. **Journal of Computational and Applied Mathematics,** v. 20, p. 53–65. 1987.

SARMADI, F.; SHOKOOHI, A. Regionalizando a precipitação no Irã usando dados em grade GPCC por meio de análise multivariada e métodos de momento L. **Theor Appl Climatol,** p. 122:121–128. 2014.

SCHNEIDER, U.; FUCHS, T.; MEYER-CHRISTOFFER, A.; RUDOLF, B. **Produtos de Análise Global de Precipitação do GPCC.** Centro Global de Climatologia de Precipitação (GPCC) Deutscher Wetterdienst, Offenbach a. M., 2011, Alemanha.

SHI, W.; YU, X.; LIAO, W.; WANG, Y.; JIA, B. Variabilidade espacial e temporal da concentração diária de precipitação na bacia do rio Lancang, China. **Journal of Hydrology,** v. 495, p. 197–207. 2013.

SILVA, M. D. N. A. D.; PESSOA, F.C.L.; SILVEIRA, R. N. P. D. O.; ROCHA, G. S.; MESQUITA, D. A. Determinação da Homogeneidade e Tendência das Precipitações na Bacia do Rio Tapajós. **Revista Brasileira de Meteorologia,** v. 33, v. 4, p. 665 675. 2018.

SILVEIRA, C. D. S.; SOUZA FILHO, F. D. A. D.; MARTINS, E. S. P. R.; OLIVEIRA, J.L.; COSTA, A. C.; NÓBREGA, M. T.; e outros Mudanças climáticas na bacia do rio São Francisco: uma análise para precipitação e temperatura. **Revista Brasileira de Recursos Hídricos**, v. 21, 416-428. 2016.

SULOCHANA, Y.; CHANDRIKA, P.; BHASKARA RAO, S. V. Taxa de chuva e estatísticas de atenuação da chuva para diferentes regiões homogêneas da Índia. **Indian Journal of Radio & Space Physics**, v. 43, p. 301-314. 2014.

TAN, P. N.; STEINBACH, M.; KUMAR, V. **Introdução à Mineração de Dados.** Addison Wesley. 2005.

WWO. Organização Meteorológica Mundial. **Comissão de Instrumentos e Métodos de Observação.** Comitê Organizador Internacional para a Intercomparação de Medição de Precipitação Sólida da WMO, relatório final da primeira sessão, 31 pp., Genebra, 1985.

YAMANA, T. K.; ELTAHIR, E. A. B. Sobre o uso de estimativas baseadas em satélite da distribuição temporal das chuvas para simular o potencial de transmissão da malária na África rural. **Recursos Hídricos. Res.,** v. 47. 2011.