

Elielma Santana de Jesus

**Análise espaço-temporal da velocidade do vento em
Pernambuco: aplicação de Entropia Multiescala e Métodos de Agrupamento**

Recife - PE

2024



**UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOMETRIA E ESTATÍSTICA APLICADA**

**Análise espaço-temporal da velocidade do vento em
Pernambuco: aplicação de Entropia Multiescala e Métodos de Agrupamento**

Tese apresentada ao Programa de Pós-Graduação em Biometria e Estatística Aplicada como requisito parcial para obtenção do título de Doutora em Biometria e Estatística Aplicada.

Área de concentração:
Biometria e Estatística Aplicada

Orientador:
Prof. Dr. Guilherme Rocha Moreira

Coorientadores:
Prof. Dr. Jader da Silva Jale
Prof.^a Dra. Cristiane Rocha Albuquerque

Recife - PE

2024

Dados Internacionais de Catalogação na Publicação
Sistema Integrado de Bibliotecas da UFRPE
Bibliotecário(a): Ana Catarina Macêdo – CRB-4 1781

J58a Jesus, Elielma Santana de.

Análise espaço-temporal da velocidade do vento em Pernambuco: aplicação de entropia multiescala e métodos de agrupamento / Elielma Santana de Jesus. – Recife, 2024.

56 f.; il.

Orientador(a): Guilherme Rocha Moreira.

Co-orientador(a): Jader da Silva Jale.

Co-orientador(a): Cristiane Rocha Albuquerque.

Tese (Doutorado) – Universidade Federal Rural de Pernambuco, Programa de Pós-Graduação em Biometria e Estatística Aplicada, Recife, BR-PE, 2024.

Inclui referências.

1. Complexidade (Ciência). 2. Energia eólica. 3. Análise multivariada. I. Moreira, Guilherme Rocha, orient. II. Jale, Jader da Silva, coorient. III. Albuquerque, Cristiane Rocha, coorient. IV. Título

CDD 519.5

Elielma Santana de Jesus

**Análise espaço-temporal da velocidade do vento em
Pernambuco: aplicação de Entropia Multiescala e Métodos de Agrupamento**

Tese apresentada ao Programa de Pós-Graduação em Biometria e Estatística Aplicada como requisito parcial para obtenção do título de Doutora, defendida e aprovada em 30/08/2024.

Banca Examinadora:

Prof. Dr. Guilherme Rocha Moreira - Orientador
Universidade Federal Rural de Pernambuco – UFRPE

Prof. Dr. Jader da Silva Jale - Coorientador
Universidade Federal Rural de Pernambuco – UFRPE

Prof.^a Dra. Cristiane Rocha Albuquerque - Coorientadora
Universidade Federal Rural de Pernambuco – UFRPE

Prof. Dr. Josimar Mendes de Vasconcelos
Universidade Federal Rural de Pernambuco – UFRPE

Prof.^a Dra. Kerolly Kedma Felix do Nascimento
Universidade Regional do Cariri – URCA

Dedico aos meus pais Maria e Antônio.

AGRADECIMENTOS

Agradeço a Deus pelo dom da vida e por me permitir chegar até aqui.

A todos os familiares, em especial, meus pais, que sempre me apoiaram.

Aos amigos que fiz ao longo de toda minha jornada acadêmica, em que compartilhamos conhecimentos, perrengues estudantis, boas risadas e são presentes mesmo distantes fisicamente.

Aos meus orientadores, Guilherme, Jader e Cristiane, pela dedicação, paciência e compreensão, no período mais difícil.

Aos Professores e Secretário do Programa de Pós-Graduação em Biometria e Estatística Aplicada, pela excelência no que fazem.

A todos que direta ou indiretamente contribuíram durante o processo de pós-graduanda e sonharam junto comigo o título de Doutora.

A Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001, pelo apoio financeiro para o desenvolvimento desta pesquisa.

RESUMO

As fontes de energia renováveis vêm ganhando espaço no mercado internacional, e isso se deve à crescente demanda da sociedade por maior quantidade de energia e pela necessidade da redução do uso de fontes que causam poluição ambiental. Nestas circunstâncias, a energia eólica apresenta um expressivo crescimento, com o Brasil ocupando a sexta posição em capacidade instalada, sendo a região Nordeste a maior produtora do país. Este estudo analisou o potencial eólico e as séries temporais mensais da velocidade do vento na região Nordeste do Brasil, no período de 1979 a 2020. Foram empregados métodos de agrupamento hierárquicos e de entropia multiescala. Observou-se que a velocidade dos ventos não se comportou de forma homogênea ao longo dos meses ou do espaço territorial, com as maiores médias registradas no inverno e na primavera, nas mesorregiões Sertão e São Francisco Pernambucano. As séries temporais mais regulares foram observadas nas escalas de tempo maiores, ou seja, apresentaram os menores valores de entropia em todas as mesorregiões. A pesquisa contribui para melhor aproveitamento dessa fonte renovável de energia, auxiliando na tomada de decisões quanto ao direcionamento de recursos financeiros, formulação de políticas e busca pelo desenvolvimento sustentável do estado.

Palavras-chave: complexidade; energia eólica; multivariada.

ABSTRACT

Renewable energy sources have been gaining ground in the international market, due to society's growing demand for more energy and the need to reduce the use of sources that cause environmental pollution. Under these circumstances, wind energy has shown significant growth, with Brazil occupying the sixth position in installed capacity, with the Northeast region being the largest producer in the country. This study analyzed the wind potential and the monthly time series of wind speed in the Northeast region of Brazil, from 1979 to 2020. Hierarchical clustering and multiscale entropy methods were used. It was observed that wind speeds did not behave homogeneously throughout the months or across the territorial space, with the highest averages recorded in winter and spring, in the Sertão and São Francisco Pernambucano mesoregions. The most regular time series were observed on larger time scales, that is, they presented the lowest entropy values in all mesoregions. The research contributes to better use of this renewable energy source, assisting in decision-making regarding the allocation of financial resources, policy formulation and the search for sustainable development in the state.

Keywords: complexity; wind energy; multivariate.

LISTA DE FIGURAS

Figura 2.1 - Localização geográfica do estado de Pernambuco, Brasil, distribuição dos pontos de grade ERA5 observados nas cinco mesorregiões do estado (Metropolitana de Recife, Zona da Mata, Agreste Pernambucano, Sertão Pernambucano e São Francisco Pernambucano) e hipsometria estadual.	16
Figura 2.2 - Boxplots mensais da velocidade média do vento para os pontos observados no estado de Pernambuco (ponto vermelho indica a média do mês).....	22
Figura 2.3 - Localização dos grupos com velocidade do vento mensal e interpolação da média para o território estadual.	24
Figura 2.4 - Localização dos pontos representantes dos grupos para os meses de junho, julho e agosto.	25
Figura 2.5 - Série de velocidade do vento mensal original (a) e anomalias (b) para três estações representativas dos grupos.	26
Figura 2.6 - EAMM para séries mensais de velocidade do vento (original e aleatória) para três estações representativas dos grupos.....	27
Figura 2.7 - EAMM para séries mensais de anomalias e aleatória para três estações representativas dos grupos.....	28
Figura 2.8 - Distribuição espacial EAMM ($\tau = 1, 2, \dots, 12$) para as séries originais da velocidade do vento sobre o estado de Pernambuco.	30
Figura 2.9 - Distribuição espacial EAMM ($\tau = 1, 2, \dots, 12$) para as séries de anomalias da velocidade do vento sobre o estado de Pernambuco.	31
Figura 2.10 - Boxplots para valores EAMM (série original) para G1, G2 E G3, obtidos com a análise de agrupamento, para $\tau = 1, 2, \dots, 12$. “*” indica que há diferença significativa entre os grupos ao nível de significância de 5% e “ns” representa não significativo.....	32
Figura 2.11 - Boxplots para valores EAMM (série de anomalia) para G1, G2 E G3, obtidos com a análise de agrupamento, $\tau = 1, 2, \dots, 12$. “*” indica que há diferença significativa entre os grupos ao nível de significância de 5% e “ns” representa não significativo.....	33
Figura 3.1 - Localização geográfica do estado de Pernambuco, Brasil, e distribuição dos pontos de grade da reanálise ERA5. Os círculos vermelhos indicam a localização dos pontos representativos das cinco mesorregiões do estado (Sertão Pernambucano, São Francisco Pernambucano, Agreste Pernambucano, Mata Pernambucana e Metropolitana do Recife). ...	37
Figura 3.2 - Séries temporais originais da velocidade média mensal do vento para pontos representativos das mesorregiões de Pernambuco.....	42

Figura 3.3 - Séries temporais das anomalias da velocidade média mensal do vento para pontos representativos das mesorregiões de Pernambuco.....	43
Figura 3.4 - Dados médios mensais da velocidade do vento, para o período de 1979 a 2020, interpolados sobre o estado de Pernambuco.....	44
Figura 3.5 - EAMM para as séries de médias mensais de velocidade do vento (original e anomalia) para cinco estações representativas das mesorregiões do estado de Pernambuco...	45
Figura 3.6 - Distribuição espacial do EAMM ($\tau = 1,2, \dots, 12$) no estado de Pernambuco para as séries originais da velocidade do vento.....	46
Figura 3.7 - Grupos formados a partir dos valores de EAMM.....	48

LISTA DE TABELAS

Tabela 2.1 - Principais medidas de distância para calcular a matriz de dissimilaridade.	17
Tabela 2.2 - Medidas de distância e métodos de agrupamento selecionados, por apresentarem maior coeficiente de correlação cofenética para cada mês.	23
Tabela 3.1 – Métodos de agrupamento aglomerativos.	40
Tabela 3.2 - Medidas de distância e métodos de agrupamento selecionados, por apresentarem maior coeficiente de correlação cofenética para cada mês.	47

SUMÁRIO

1	INTRODUÇÃO GERAL	11
2	ANÁLISE DE AGRUPAMENTO E DE ENTROPIA MULTIESCALA PARA A VELOCIDADE DO VENTO EM PERNAMBUCO, BRASIL.....	13
	2.1 INTRODUÇÃO.....	13
	2.2 MATERIAL E MÉTODOS.....	15
	2.3 RESULTADOS E DISCUSSÃO	21
	2.4 CONSIDERAÇÕES FINAIS	33
3	ESTUDO DA VELOCIDADE DO VENTO COM MÉTODOS DE ENTROPIA E ANÁLISE MULTIVARIADA	35
	3.1 INTRODUÇÃO.....	35
	3.2 MATERIAL E MÉTODOS.....	36
	3.3 RESULTADOS E DISCUSSÃO	41
	3.4 CONSIDERAÇÕES FINAIS	48
4	CONSIDERAÇÕES GERAIS	49
	REFERÊNCIAS	50

1 INTRODUÇÃO GERAL

Diante dos efeitos negativos das mudanças climáticas e da crescente demanda por energia no cenário global, os países são impulsionados a buscarem recursos sustentáveis e diversifiquem suas matrizes energéticas, além de cumprirem com as metas de redução do efeito estufa do Acordo de Paris (Kim; Hur, 2021). Nesse contexto, a energia eólica é uma das que mais cresce e se desenvolve (IRENA, 2020). Essa fonte apresenta vantagens como ser altamente renovável, livre de poluição, de baixo custo e geradora de renda, além de proporcionar melhorias no contexto financeiro para as populações onde os parques eólicos são instalados (Liu; Song; Bi, 2021).

No Brasil, o mercado eólico vem apresentando crescimento expressivo. Em menos de dez anos, o país passou da 15ª para 6ª posição no *ranking* internacional de capacidade instalada em 2021 e continua ocupando essa mesma posição. A região Nordeste (NE) tem sido responsável por abrigar grande parte dos parques eólicos, pois apresenta características favoráveis para essa produção devido à regularidade, constância, intensidade e direção praticamente unidirecional dos ventos (ABEEólica, 2020).

O vento é a segunda maior fonte de geração de energia no país, respondendo por 15,4% na matriz energética. A energia eólica consegue suprir a demanda do NE e complementa a demanda nacional, considerando especialmente que as mudanças climáticas comprometem os níveis dos reservatórios hidrelétricos, e conseqüentemente a geração de energia (GWEC, 2024).

A expansão do mercado requer pesquisas que apoiem a tomada de decisões quanto ao direcionamento de recursos financeiros, formulação de políticas e busca pelo desenvolvimento sustentável. Para obter resultados sólidos a respeito desses vieses utilizam-se ferramentas estatísticas em áreas como Séries Temporais, Análise Multivariada, Análise de Regressão, Estatística Espacial, entre outras.

A abordagem de agrupamento visa auxiliar na discussão de fenômenos meteorológicos, sendo empregada para definir a regionalidade (Ratto; Berri; Maronna, 2014). Seguindo a ideia de que o estudo climatológico de uma variável física deve ter como objetivo a compreensão de seu estado médio, bem como a caracterização de sua dinâmica, a análise de agrupamento foi aplicada para estudar o clima eólico da Córsega (França), para identificar as regiões anemológicas (estado médio) e os regimes de vento (variabilidade do tempo) que caracterizam as suas zonas costeiras (Burlando; Antonelli; Ratto, 2008). Já Siraj-Ud-Douh e Islam

(2019), identificaram sete zonas climáticas em Bangladesh, ao aplicarem métodos de agrupamento hierárquico aglomerativos em dados de precipitação.

Para atenuar a difícil tarefa em prever a velocidade do vento e a produção de energia eólica, as séries temporais são estudadas através de métodos como dimensão fractal em diferentes condições de terreno (Yan *et al.*, 2020), método de Fisher-Shannon (Guignard *et al.*, 2019b), aprendizado de máquina (Amato *et al.*, 2022), redes neurais (Robert; Foresti; Kanevski, 2012), informação mútua (Laib *et al.*, 2019), multifractal (Laib *et al.*, 2018) e entropia multiescala (Li *et al.*, 2011).

Para o NE, Santos *et al.*, 2021 empregaram mistura de distribuições e análise multifractal a dados de 136 estações no período de 2004 a 2018, observando que todas as estações apresentaram comportamento persistente para as séries de velocidade do vento. A multifractalidade causada pelas correlações de longo alcance e pela função densidade de probabilidade foi observada para a maioria das estações. A correlação de longo alcance foi avaliada em dados de Fernando de Noronha – PE (Dos Anjos *et al.*, 2015) e em diferentes regiões do estado da Bahia e do Arquipélago de Abrolhos (Santos *et al.*, 2019) por meio da técnica DFA (*Detrended Fluctuation Analysis*). Em ambos os estudos se verificou a persistência nas séries temporais da velocidade do vento.

O método de entropia multiescala tem sido pouco explorado nesta área, embora seja importante conhecer a regularidade e complexidade das séries nas diversas escalas.

Este trabalho está estruturado em capítulos. Neste primeiro foi apresentada a introdução geral. No capítulo seguinte, encontra-se o Artigo 1 intitulado “Análise de agrupamento e de entropia multiescala para a velocidade do vento em Pernambuco, Brasil”. Em seguida, é apresentado o Artigo 2 intitulado “Estudo da velocidade do vento com métodos de entropia e análise multivariada”. Por fim, são expostas as considerações gerais e as referências.

2 ANÁLISE DE AGRUPAMENTO E DE ENTROPIA MULTIESCALA PARA A VELOCIDADE DO VENTO EM PERNAMBUCO, BRASIL

Resumo: Neste estudo, investigou-se o potencial eólico e a complexidade das séries temporais mensais da velocidade do vento no período de 1979 a 2020, em 133 pontos observados no estado de Pernambuco, nordeste do Brasil. Foram empregados métodos de agrupamento hierárquicos para identificar regiões homogêneas, e o método de entropia amostral em múltiplas escalas modificado para análise da complexidade e regularidade da velocidade do vento em escala de um mês a um ano. As maiores médias foram observadas para os meses de junho a outubro e, nas mesorregiões Sertão e São Francisco Pernambucano. Foram encontrados três grupos distintos para todos os meses com diferentes métodos de agrupamento, em junho, julho e agosto os grupos permaneceram com a mesma formação. À medida que a escala de tempo aumentou, a entropia diminuiu tanto para as séries temporais originais da velocidade do vento quanto para as de anomalias. Nas escalas baixas e intermediárias observou-se diminuição gradual da entropia no sentido Leste-Oeste. Com o teste de significância estatística foi possível diferenciar a regularidade dos ventos no grupo que tem seus pontos concentrados no Sertão e São Francisco, dos grupos do Agreste e Zona da Mata. Os resultados evidenciam a importância de técnicas exploratórias e de teoria da informação em estudos relacionados a fenômenos atmosféricos, e podem auxiliar em trabalhos futuros e para tomada de decisões quanto à investimentos na região para captação de energia eólica.

Palavras-chave: agrupamento hierárquico; complexidade; energia eólica; teoria da informação.

2.1 INTRODUÇÃO

Com o avanço da tecnologia e industrialização a demanda da sociedade por energia tem aumentado, e essa energia é oriunda de fontes renováveis e não renováveis. Já são amplamente conhecidos os impactos negativos causados pela utilização das fontes não renováveis, e a possível solução para este problema é aumentar a participação de recursos renováveis e sustentáveis nas matrizes energéticas (Azam *et al.*, 2021).

No Brasil, a utilização da força dos ventos para gerar energia tem mostrado um crescimento expressivo quando comparado a outras fontes renováveis não hídricas. Segundo o relatório mundial do setor eólico divulgado pelo *Global Wind Energy Council* - (GWEC,

2024), o país alcançou o recorde de 4,8 *gigawatts* (GW) em novas instalações no último ano e, atualmente, ocupa a sexta posição no *ranking* internacional de capacidade total instalada. Parte significativa das instalações e produção eólica provém da região Nordeste (NE) que possui características favoráveis, como ventos estáveis, intensos e sem mudanças bruscas de velocidade ou direção, o que contribui para a expansão (ABEEólica, 2020).

A região NE é afetada pela escassez de chuvas, e as previsões a longo prazo indicam que os níveis médios de precipitação devem diminuir, devido às mudanças climáticas, o que conseqüentemente, comprometerá o volume dos reservatórios hidrelétricos e a produção de energia (Jong *et al.*, 2013, 2017). A energia eólica pode atuar como uma solução complementar à demanda nacional, visto que já consegue suprir o consumo do Nordeste brasileiro, especialmente nos períodos de seca do ano (GWEC, 2022).

Para que este setor continue crescendo e mais usinas eólicas sejam implantadas, são necessários estudos que visem estimar as características da velocidade do vento e identificar regiões e períodos com maior potencial energético.

Uma das técnicas que podem ser utilizadas para a análise da velocidade do vento é a análise de agrupamento (Carreón-Sierra *et al.*, 2015), que tem vasta aplicação em climatologia, seja na determinação de regiões homogêneas, seja no agrupamento de meses e/ou estações do ano, de acordo com o fenômeno de interesse (Alves *et al.*, 2016; Gebert; Kist; Virgens Filho, 2018).

Outra ferramenta estatística utilizada na avaliação do potencial eólico é o ajuste de modelos de distribuição de probabilidade das séries históricas da velocidade do vento (Liu *et al.*, 2019; Mahmood; Resen; Khamees, 2020; Rocha *et al.*, 2012; Sumair *et al.*, 2021). Além disso, por se tratar de uma variável atmosférica complexa, métodos fractais e multifractais também são aplicados (Laib *et al.*, 2018; Telesca; Lovallo; Kanevski, 2016; Yan *et al.*, 2020).

A abordagem baseada em entropia, inicialmente proposta para sistemas biológicos e pesquisas clínicas, também vem ganhando espaço com êxito em estudos sobre a complexidade das séries de velocidade do vento (Li *et al.*, 2011), temperatura (Shuangcheng *et al.*, 2006) e precipitação (Hu; Liu; Sang, 2019; Xavier *et al.*, 2021; Zhou; Lei, 2020). Contudo, a entropia tradicional é medida em escala única (Richman; Moorman, 2000). A complexidade dos sistemas dinâmicos não-lineares foi mais bem compreendida quando analisada para diversas escalas de tempo, o que levou ao desenvolvimento da entropia amostral em múltiplas escalas (EAM) para preencher essa lacuna (Costa; Goldberger; Peng, 2002). Mais tarde, Wu *et al.* (2013) apresentaram o método EAM modificado (EAMM) para superar a imprecisão das estimativas obtidas devido ao processo de granulação para séries temporais curtas.

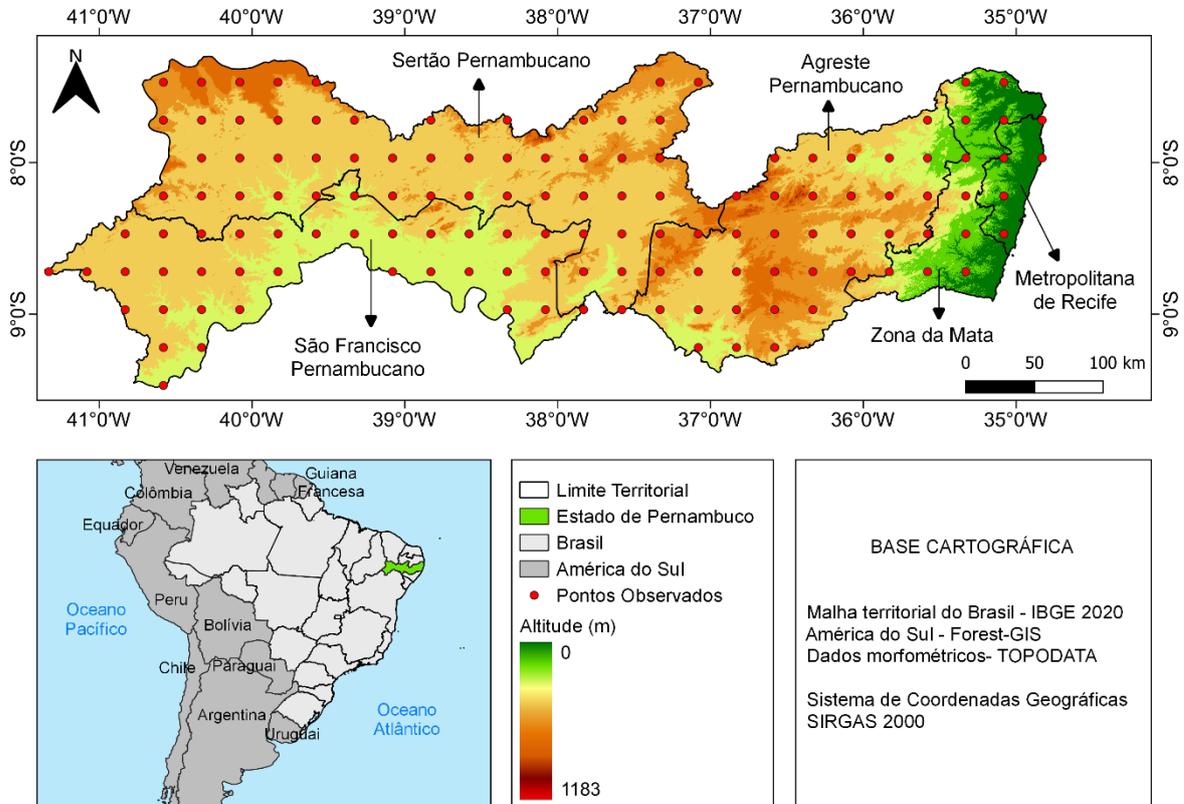
Pesquisas sobre a complexidade em múltiplas escalas para a velocidade do vento ainda são escassas, inclusive para o Brasil. Por isso, o objetivo deste estudo foi contribuir para a compreensão da dinâmica do vento, temporal e espacialmente, por meio dos métodos de análise de agrupamento hierárquico e EAMM, aplicados às séries mensais de velocidade do vento, no período de 1979 a 2020 para o estado de Pernambuco, região NE.

2.2 MATERIAL E MÉTODOS

2.2.1 Área de estudo e conjunto de dados

O estado de Pernambuco está localizado no centro-leste da região NE, entre os paralelos $7^{\circ}15'45''$ S e $9^{\circ}28'18''$ S e os meridianos $34^{\circ}48'33''$ W e $41^{\circ}19'54''$ W. Tem como limites os estados da Paraíba e Ceará (norte), Piauí (oeste), Alagoas e Bahia (sul) e ao leste o Oceano Atlântico. A área territorial é de aproximadamente $98.067,877$ km² (IBGE, 2020), são 184 municípios e o Distrito Estadual Arquipélago Fernando de Noronha, distribuídos em cinco mesorregiões: Metropolitana de Recife, Zona da Mata, Agreste Pernambucano, Sertão Pernambucano e São Francisco Pernambucano (Figura 2.1).

Figura 2.1 - Localização geográfica do estado de Pernambuco, Brasil, distribuição dos pontos de grade ERA5 observados nas cinco mesorregiões do estado (Metropolitana de Recife, Zona da Mata, Agreste Pernambucano, Sertão Pernambucano e São Francisco Pernambucano) e hipsometria estadual.



Fonte: Autor (2022).

Os dados utilizados neste trabalho são provenientes da recente reanálise ERA5 (Hersbach *et al.*, 2020), produzida pelo *European Centre for Medium-Range Weather Forecasts* (ECMWF) e disponibilizada através do *Copernicus Climate Data Store* (<https://cds.climate.copernicus.eu/>). Foram coletadas informações horárias das componentes do vento, zonal (u) e meridional (v), medidas a uma altura de 10 m do solo, para 133 pontos de grade com resolução espacial de $0,25^\circ$, aproximadamente 31 km, distribuídos no estado de Pernambuco (Figura 2.1) para o período de 1979 a 2020, e foram agregadas como médias mensais.

A partir das componentes, a velocidade escalar do vento (m/s) de cada ponto foi determinada como em Molina, Gutiérrez e Sánchez (2021):

$$x(t) = \sqrt{u^2 + v^2},$$

em que u e v representam a direção do vento no sentido Oeste-Leste e Norte-Sul, respectivamente.

Dada a série original da velocidade do vento $x(t)$, as anomalias mensais foram calculadas para retirar o efeito sazonal, conforme (Bhatia *et al.*, 2020; Kantelhardt *et al.*, 2006):

$$X(t) = \frac{x(t) - \mu_\tau}{\sigma_\tau},$$

em que μ_τ e σ_τ são a média e o desvio padrão, respectivamente, da velocidade do vento mensal $x(t)$ calculada para cada mês de todos os anos.

2.2.2 Análise de agrupamento

Foram utilizados métodos hierárquicos aglomerativos de acordo com Hair *et al.* (2019), em que cada ponto começou como seu próprio grupo e foram unidos sucessivamente com os dois mais semelhantes, até que se formaram um único grupo contendo todos os pontos observados.

O algoritmo consistiu nos seguintes passos: primeiro foi calculada a matriz de dissimilaridade, que expressou a diferença entre os pares de observações, através das medidas de distância (Tabela 2.1).

Tabela 2.1 - Principais medidas de distância para calcular a matriz de dissimilaridade.

Distância	Equação	Referência
Euclidiana	$d(x_i, x_j) = \sqrt{\sum_{l=1}^n (x_{il} - x_{jl})^2}$	Kumar; Chhabra; Kumar, (2014); Li <i>et al.</i> (2019); Modiri; Bárdossy, (2021)
Canberra	$d(x_i, x_j) = \sum_{l=1}^n \frac{ x_{il} - x_{jl} }{ x_{il} + x_{jl} }$	Kumar; Chhabra; Kumar, (2014)
Manhattan	$d(x_i, x_j) = \sum_{l=1}^n x_{il} - x_{jl} $	Kumar; Chhabra; Kumar, (2014); Strauss; von Maltitz, (2017); Zelerman (2015)
Máxima	$d(x_i, x_j) = \max_{1 \leq j \leq n} x_{il} - x_{jl} $	Zelerman (2015)
Minkowski	$d(x_i, x_j) = \left(\sum_{l=1}^n x_{il} - x_{jl} ^p \right)^{1/p},$ $p \geq 1$	Strauss; von Maltitz, (2017); Xu; Tian, (2015); Zelerman (2015)

$d(x_i, x_j)$ é a distância entre os pontos x_i e x_j ; x_{il} e x_{jl} são os valores da l -ésima velocidade nos pontos x_i e x_j , respectivamente. A distância Minkowski é uma generalização das distâncias de Manhattan ($p = 1$) e euclidiana ($p = 2$), sendo p qualquer número real maior que um.

Em seguida, foram aplicados cinco métodos de ligação aglomerativos para a formação dos grupos:

- Ligação simples - distância entre pontos mais próximos de grupos diferentes, também chamado de método do vizinho mais próximo (Li; Rezaeipanh; Tag El Din, 2022; Modiri; Bárdossy, 2021). Calculada como:

$$d_{XY} = \min\{d_{ij} : i \in X; j \in Y\},$$

em que d_{XY} é a distância entre os grupos X e Y ; d_{ij} é a distância entre os pontos i e j .

- Ligação completa - distância máxima entre pontos pertencentes a grupos diferentes, também chamado de método do vizinho mais distante (Li; Rezaeipanh; Tag El Din, 2022; Modiri; Bárdossy, 2021). Calculada como:

$$d_{XY} = \max\{d_{ij} : i \in X; j \in Y\},$$

em que d_{XY} é a distância entre os grupos X e Y ; d_{ij} é a distância entre os pontos i e j .

- Ligação média - distância média de todos os pontos pertencentes a grupos diferentes (Li; Rezaeipanh; Tag El Din, 2022; Modiri; Bárdossy, 2021; Santos et al., 2019). Calculada como:

$$d_{XY} = \frac{1}{n_X n_Y} \sum_{i \in X} \sum_{j \in Y} d_{ij},$$

em que d_{XY} é a distância entre os grupos X e Y ; n_X e n_Y representam o número de pontos nos grupos X e Y , respectivamente; d_{ij} é a distância entre os pontos i e j .

- Método do centroide - distância entre os centroides, ou seja, a distância entre os valores médios dos pontos de cada grupo (Li; Rezaeipanh; Tag El Din, 2022). O centroide é alterado quando novos pontos são incorporados ao grupo. Calculada como:

$$d_{XY} = d_{\bar{x}\bar{y}},$$

em que d_{XY} é a distância entre os grupos X e Y ; \bar{x} e \bar{y} são os centroides (médias) dos grupos X e Y , respectivamente.

- Método Ward - minimiza a soma quadrática dos desvios dentro de cada agrupamento (Modiri; Bárdossy, 2021; Ward, 1963). Também é conhecido como método da mínima variância e é calculado como:

$$d_{XY} = \sqrt{\frac{2n_X n_Y}{(n_X + n_Y)}} \|\bar{x} - \bar{y}\|^2,$$

em que d_{XY} é a distância entre os grupos X e Y ; n_X e n_Y representam o número de pontos e, \bar{x} e \bar{y} são as médias dos grupos X e Y , respectivamente.

2.2.2.1 Comparação dos métodos

Para avaliar o método que melhor descreveu os agrupamentos formados utilizou-se o coeficiente de correlação cofenética (r) proposto por Sokal e Rohlf (1962) e expresso como:

$$r = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 (y_i - \bar{y})^2}},$$

em que x_i é o valor da variável na matriz de dissimilaridade, y_i é o valor da variável na matriz cofenética, \bar{x} e \bar{y} são as médias das variáveis nas matrizes de dissimilaridade e cofenética, respectivamente. Valores de r acima de 0,7 indicam boa representatividade entre as distâncias (Bussad; Miazaki; Andrade, 1990).

A média mensal da velocidade do vento foi considerada na análise de agrupamento.

2.2.2.2 Número de grupos

Para a determinação do número ideal de grupos foi utilizado o pacote *NbClust* do *software* R Core Team, (2021). Esse pacote emprega índices que combinam informações sobre a união de objetos dentro de um mesmo grupo e a separação entre os grupos, além de considerar as medidas de similaridade ou dissimilaridade (Charrad *et al.*, 2014).

Foram observadas as seguintes medidas: ch; ball; beale; ccc; cindex; db; duda; dunn; friedman; gamma; gplus; hartigan; kl; marriot; mcclain; pseudot2; ptbiserial; ratkowsky; rubin; scott; sdbw; sdindex; silhouette; tau e tracew, para selecionar o número e a organização mais adequados dos grupos, de acordo com o critério da maioria.

2.2.3 Entropia amostral em múltiplas escalas modificada

As séries de velocidade do vento apresentam comportamento complexo não linear (Rind, 1999) e os métodos de entropia são utilizados para medir a complexidade e a regularidade de sistemas dinâmicos (Li *et al.*, 2011). Neste estudo, foi aplicado o algoritmo de EAMM (Wu *et al.*, 2013), uma versão aprimorada de EAM (Costa; Goldberger; Peng, 2002), que depende do cálculo de entropia amostral (EA) (Richman; Moorman, 2000) para múltiplas escalas. Por sua vez, EA é o logaritmo natural negativo da probabilidade condicional, de que duas séries de dados de comprimento N sejam semelhantes dentro de um nível de tolerância r para m pontos e permaneçam assim para $m + 1$ pontos, nas quais as auto correspondências não são incluídas no cálculo da probabilidade (Richman; Moorman, 2000).

Inicialmente, para uma série de velocidade do vento $x(i)$, $i = 1, \dots, N$ foi gerada a nova série $x_\tau(j)$ com o procedimento de média móvel para cada fator de escala τ , como segue (Wu *et al.*, 2013):

$$x_\tau(j) = \frac{1}{\tau} \sum_{i=j}^{j+\tau-1} x(i) \quad \text{e} \quad j = 1, \dots, N - \tau + 1.$$

Em seguida foram definidos os vetores de tamanho m , em que $x_m^\tau(i) = \{x(i), x(i+1), \dots, x(i+m)\}$, $i = 1, \dots, N - \tau + 1 - m\tau$, com atraso de tempo τ . A distância entre os vetores $x_m(i)$ e $x_m(j)$ foi calculada como a máxima diferença entre seus componentes escalares correspondentes (Yin *et al.*, 2018):

$$d_m[x_i, x_j] = \max_{0 \leq k \leq m-1} |x(i+k) - x(j+k)|, \quad 1 \leq i, j \leq N - \tau + 1 - m\tau, i \neq j. \quad (15)$$

Os vetores $x_m(i)$ e $x_m(j)$ são semelhantes dentro de um nível de tolerância r , se $d_m[x_i, x_j] \leq r$ ($r = r\sigma$, em que σ é o desvio-padrão de $x_\tau(j)$). O número de vetores semelhantes para m pontos e para $m+1$ pontos é representado, respectivamente, por B e A . Assim, a entropia amostral é estimada por:

$$S_E(m, r, N, \tau) = -\ln \frac{A}{B}.$$

As séries mais regulares são menos complexas e levam a valores de EA mais baixos. No presente trabalho, foram utilizados os parâmetros $m = 2,0$ e $r = 0,2$, são valores comumente empregados nos estudos de entropia (Pincus, 1991; Richman; Moorman, 2000) e, o fator de escala $\tau = 1, \dots, 12$ (representa a escala de um a doze meses). Foi utilizado o pacote *TSEntropies* (Tomcala, 2018) do *software* R Core Team (2021).

Os resultados de EAMM foram apresentados para pontos selecionados de forma aleatória para representar os grupos obtidos com a análise de agrupamento.

2.2.4 Ponderação pelo inverso da distância

Foi realizada a espacialização das médias da velocidade do vento e dos valores de entropia utilizando o método Inverso da Distância Ponderada (IDP), proposto por Shepard (1968). Esse método baseia-se na suposição de que pontos próximos têm maior correlação e semelhança do que observações distantes. Assim, os pesos das amostras diminuem com o aumento da distância entre as amostras conhecidas e os pontos estimados (Ouabo; Sango-doyin; Ogundiran, 2020).

A expressão matemática é dada por:

$$\hat{Z}_{s_0} = \sum_{i=1}^N \lambda_i Z_{(s_i)},$$

em que \hat{Z}_{s_0} é o valor estimado para cada local s_0 , N é o número de pontos observados, λ_i é o peso atribuído ao i -ésimo ponto conhecido e $Z_{(s_i)}$ é o valor observado. O peso é determinado pelas equações:

$$\lambda_i = \frac{d_{i0}^{-p}}{\sum_{i=1}^N d_{i0}^{-p}} \quad \text{e} \quad \sum_{i=1}^N \lambda_i = 1,$$

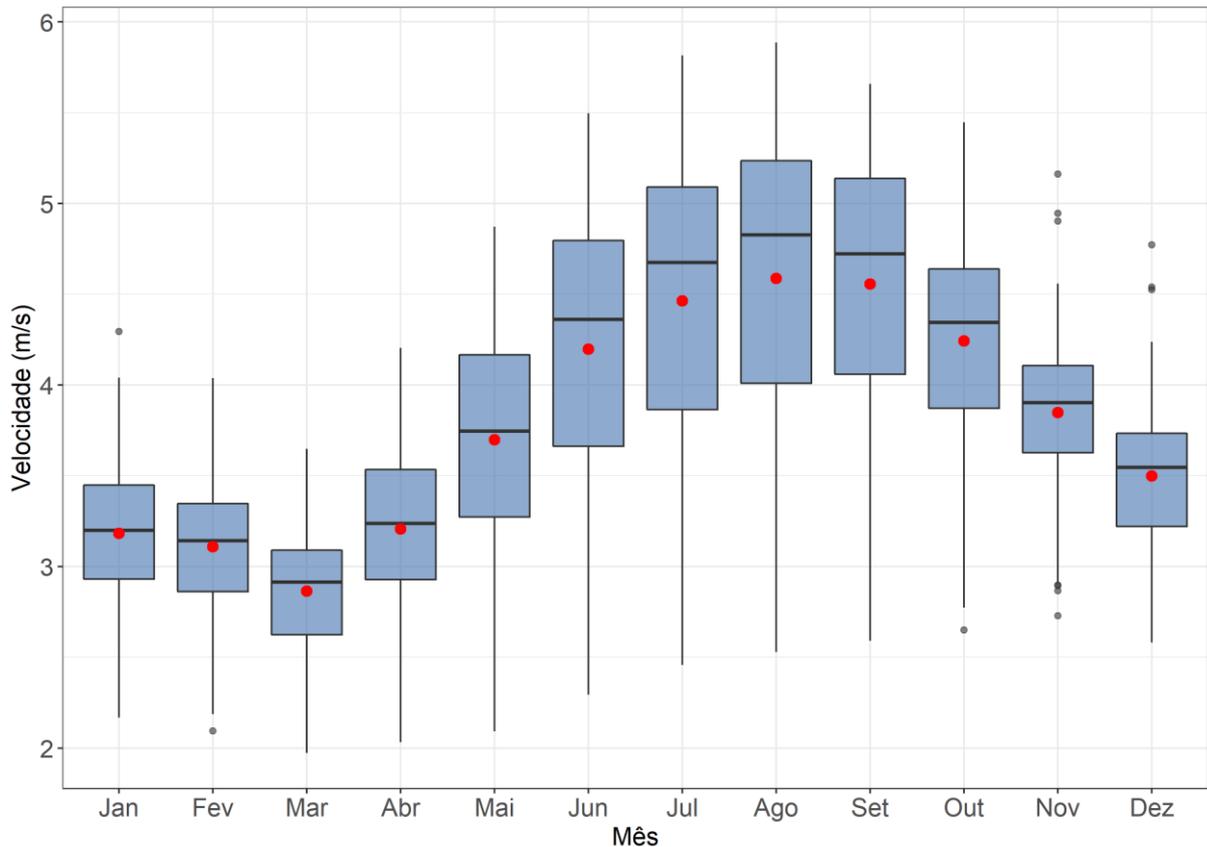
em que d_{i0} é a distância entre o local de previsão s_0 e os locais medidos s_i , p é o parâmetro de potência exponencial e que usualmente é definido como 2,0 (Ly; Charles; Degré, 2011).

A análise foi realizada no *software* R Core Team (2021). Para trabalhar com dados espaciais foram utilizados os pacotes *sp* (Pebesma; Bivand, 2005) e *spatstat* (Baddeley; Turner, 2005), além da função “*idw*” do pacote *gstat* (Pebesma, 2015) para a predição.

2.3 RESULTADOS E DISCUSSÃO

A Figura 2.2 apresenta os boxplots da velocidade média do vento para os meses de janeiro a dezembro, nos anos de 1979 a 2020, com base nos pontos observados no estado de Pernambuco. Nota-se que os ventos mais fortes ocorreram nos meses de junho a outubro, com médias (representadas com pontos vermelhos nos gráficos) acima de 4,0 m/s. A partir de setembro a intensidade diminui, continua a baixar até março quando a velocidade média fica abaixo de 3,0 m/s, o que segundo Santos et al. (2021) são insuficientes para interligar os subsistemas de geração de energia.

Figura 2.2 - Boxplots mensais da velocidade média do vento para os pontos observados no estado de Pernambuco (ponto vermelho indica a média do mês).



Fonte: Autor (2022).

Nas estações de inverno (junho, julho, agosto) e primavera (setembro, outubro, novembro) observou-se maior disponibilidade do recurso eólico e menor, nas estações de verão (dezembro, janeiro, fevereiro) e outono (março, abril, maio), devido as circulações atmosféricas. A Zona de Convergência Intertropical e o Anticiclone do Atlântico Sul atuam no Brasil, e suas posições exercem influências sobre as velocidades médias sazonais do vento no país (Gilliland; Keim, 2017). Essa tendência sazonal está em consonância com estudos anteriores sobre o potencial eólico e as contribuições desta fonte energética para geração de energia limpa e altamente renovável na Suíça (Kruyt; Lehning; Kahl, 2017), na China (Liu *et al.*, 2019), na costa litorânea e na região Sudeste do Brasil (Correia Filho *et al.*, 2021; Pimenta *et al.*, 2019) e na Coréia do Sul (Kim; Hur, 2021).

Para a escolha da medida de distância e método de agrupamento, foi avaliado o coeficiente de correlação cofenética (r). Quanto mais próximo de 1,0, melhor o ajuste entre a matriz de distância e a matriz cofenética. A Tabela 2.2 apresenta a combinação das métricas que obtiveram o maior valor de r para cada mês, e os respectivos coeficientes. Com exceção do mês de dezembro, em que a distância euclidiana foi a que apresentou melhor desempenho

combinada com o método centroide, para os demais meses a distância Canberra obteve melhores resultados com diferentes métodos.

Tabela 2.2 - Medidas de distância e métodos de agrupamento selecionados, por apresentarem maior coeficiente de correlação cofenética para cada mês.

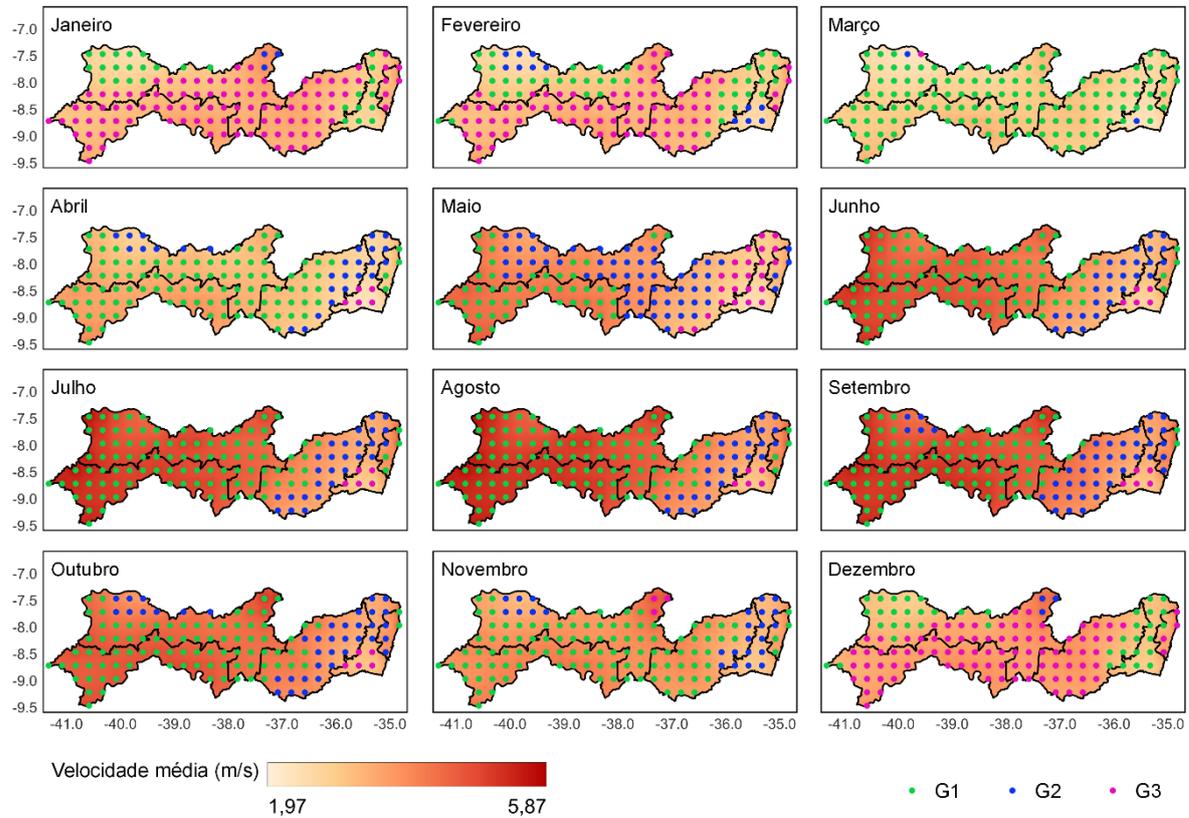
Mês	Distância	Método	<i>r</i>
Janeiro	Canberra	média	0,750
Fevereiro	Canberra	centroide	0,766
Março	Canberra	simples	0,736
Abril	Canberra	centroide	0,795
Maiο	Canberra	completa	0,774
Junho	Canberra	completa	0,807
Julho	Canberra	centroide	0,832
Agosto	Canberra	centroide	0,830
Setembro	Canberra	média	0,832
Outubro	Canberra	média	0,817
Novembro	Canberra	centroide	0,820
Dezembro	euclidiana	centroide	0,772

Fonte: Autor (2022).

O método Ward não se mostrou adequado para nenhum dos meses; no entanto, foi o que alcançou maior correlação cofenética ao agrupar regiões homogêneas de precipitação extrema no Reino Unido, em comparação com outros métodos hierárquicos (Darwish *et al.*, 2020). Não existe um método específico que funcione melhor para todos os problemas (Li; Rezaeiapanah; Tag El Din, 2022), e não faz parte do escopo deste trabalho discutir o desempenho dos métodos.

A aplicação dos índices do pacote *NbClust* indicou três como o número satisfatório de grupos para todos os meses (G1, G2 e G3). A Figura 2.3 apresenta a distribuição geográfica dos grupos e a interpolação pelo método IDP para todo o território do estado, com base nas médias da velocidade mensal do vento para os pontos observados.

Figura 2.3 - Localização dos grupos com velocidade do vento mensal e interpolação da média para o território estadual.



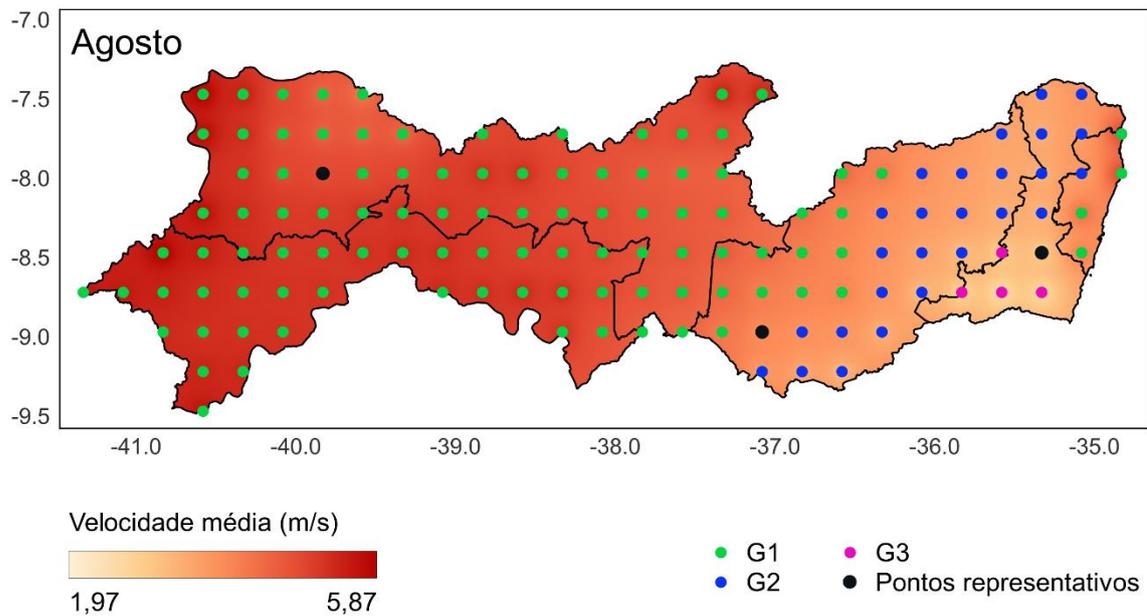
Fonte: Autor (2022).

Para os meses no qual os ventos apresentaram maior intensidade média, de junho a outubro, os grupos continuaram com a mesma configuração em junho, julho e agosto, com poucas alterações em setembro e outubro. Durante esse período em que os grupos foram iguais, observou-se um gradiente para a média da velocidade do vento que foi do litoral para o interior, em que as mesorregiões Sertão Pernambucano e São Francisco Pernambucano apresentaram as maiores médias, destacando-se como regiões com grande potencial eólico. Conforme o Atlas Eólico e Solar de Pernambuco (2015), o potencial eólico na costa litorânea é explicado pela circulação de brisa marinha e ventos alísios; já no interior, é devido ao relevo e a baixa rugosidade. Em novembro, dezembro e janeiro, houve semelhança na composição dos grupos, sendo um grupo formado somente por três pontos localizados na região norte do Sertão Pernambucano. A intensidade da cor nos mapas indica que essa área apresentou uma velocidade média superior às demais. O grupo com maior número de pontos abrange as mesorregiões do São Francisco, Sertão, Agreste e Metropolitana de Recife enquanto o último grupo é formado por pontos da Zona da Mata e Sertão. A Figura 2.3 mostrou, ainda, que a análise de agrupa-

mento identificou similaridade entre pontos observados no litoral (mesorregião Metropolitana de Recife) e no interior do estado, colocando-os no mesmo grupo em todos os meses.

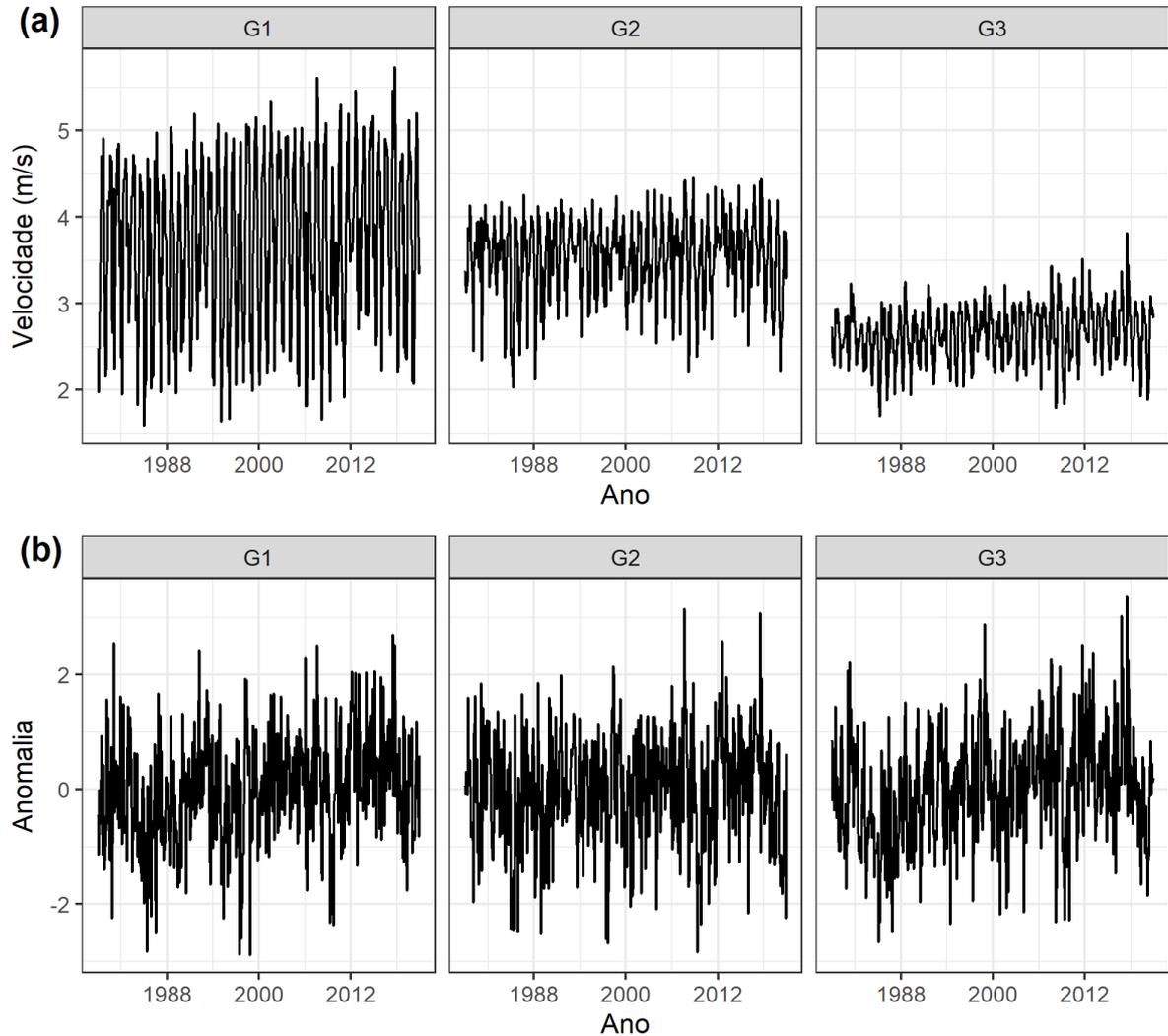
A complexidade e regularidade das séries temporais mensais, tanto as originais quanto as de anomalias, da velocidade do vento foram calculadas para todos os pontos. Os resultados são apresentados para três pontos escolhidos aleatoriamente para representar os grupos do período em que houve vasta disponibilidade do recurso eólico e que se mantiveram inalterados (junho, julho e agosto). A Figura 2.4 indica a localização dos pontos representativos, e as séries originais e de anomalias para os grupos G1, G2, e G3 são apresentadas na Figura 2.5. Observou-se grande variabilidade na série original (Figura 2.5 (a)) do grupo G1, em comparação aos grupos G2 e G3, sendo que G3, localizado na mesorregião Zona da Mata, apresentou menor velocidade média ao longo do período analisado. Para as séries de anomalias (Figura 2.5 (b)) houve comportamento semelhante para todos os grupos.

Figura 2.4 - Localização dos pontos representantes dos grupos para os meses de junho, julho e agosto.



Fonte: Autor (2022).

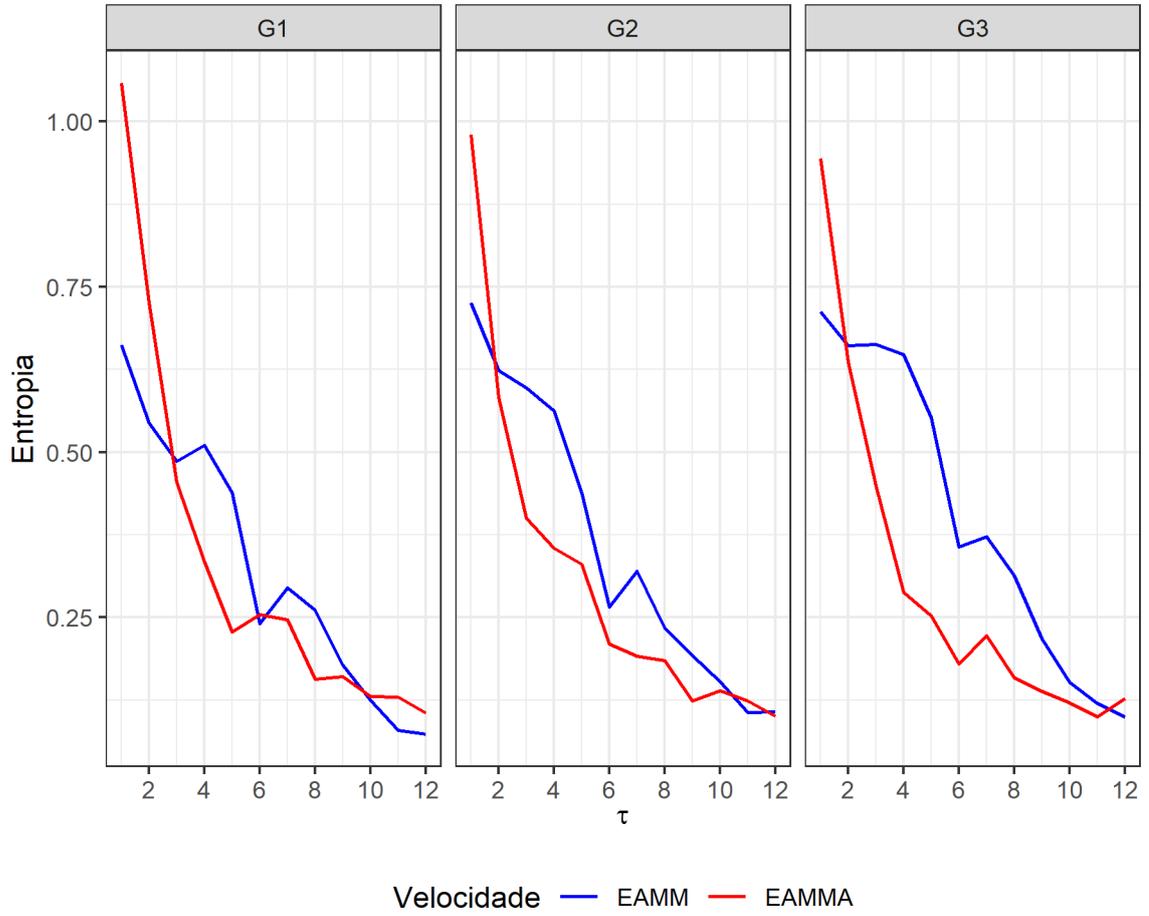
Figura 2.5 - Série de velocidade do vento mensal original (a) e anomalias (b) para três estações representativas dos grupos.



Fonte: Autor (2022).

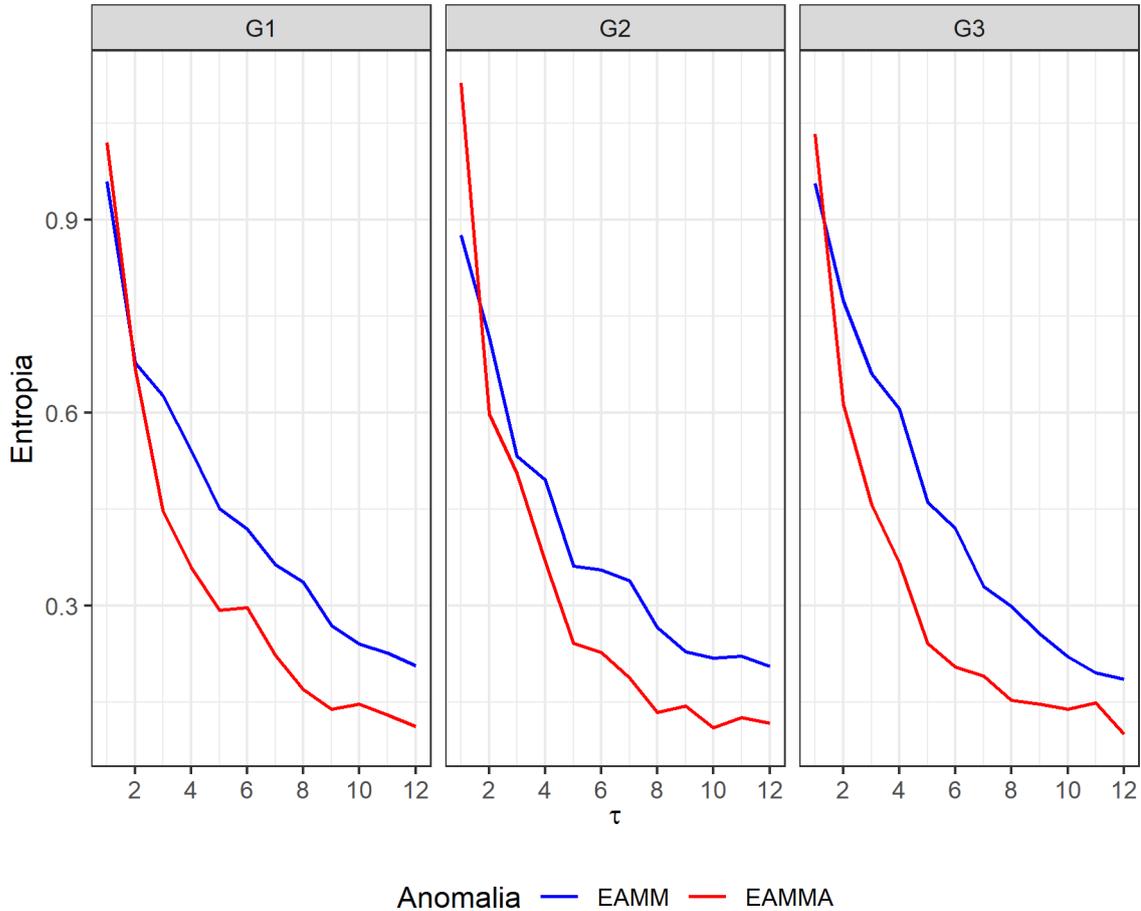
Os valores de entropia também foram estimados para séries aleatórias (EAMMA), randomizadas com $10.000 * N$ transposições de pares de valores para cada série temporal, sendo N o comprimento da série. Os resultados para $\tau = 1, \dots, 12$ são mostrados na Figura 2.6 (original) e Figura 2.7 (anomalia). De maneira geral, as séries originais apresentaram valores de entropia menores que as séries de anomalias. À medida que a escala de tempo aumentou, o EAMM diminuiu em ambos os conjuntos, indicando maior regularidade nas escalas mais altas, sendo que para os três grupos nas séries originais, observou-se padrão de crescimento da entropia entre $\tau = 6$ e $\tau = 7$, e que diminuiu nas escalas posteriores. Para G1, notou-se aumento entre as escalas de tempo 3 e 4 meses, enquanto no G3 manteve-se constante.

Figura 2.6 - EAMM para séries mensais de velocidade do vento (original e aleatória) para três estações representativas dos grupos.



Fonte: Autor (2022).

Figura 2.7 - EAMM para séries mensais de anomalias e aleatória para três estações representativas dos grupos.



Fonte: Autor (2022).

Os valores de EAMMA foram maiores na escala temporal inicial, seguiram o padrão de decrescimento e voltaram a ser maiores ou iguais de que as séries originais (Figura 2.6) nas escalas finais. Para as anomalias (Figura 2.7) EAMMA continuou a diminuir até as escalas finais. O fato de a entropia ser menor para as séries aleatórias é devido ao comportamento destes dados, que são mais regulares, ou seja, menos complexos de que os dados originais e o objetivo da análise de entropia é exatamente estimar a complexidade contida nas séries temporais (Costa; Goldberger; Peng, 2002).

A Figura 2.8 e Figura 2.9 mostram a distribuição espacial dos valores de EAMM para a série original e de anomalia, respectivamente, para o estado de Pernambuco. Estes resultados concordaram com os apresentados anteriormente (Figura 2.6 e Figura 2.7) para os pontos representativos.

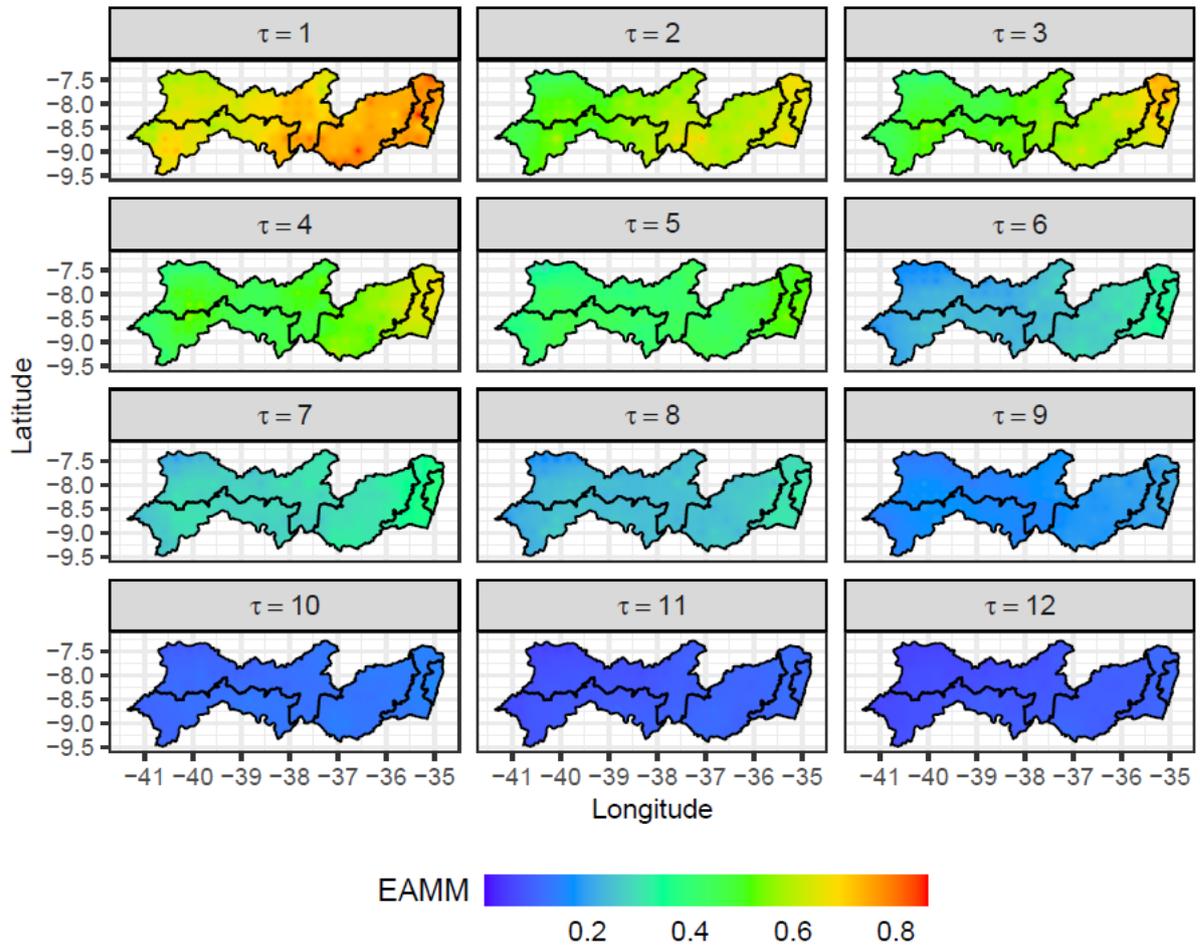
Nas escalas iniciais, a entropia foi alta, enquanto nas escalas finais, foi baixa. Para $\tau = 7$, observou-se um aumento no valor de EAMM para todo o território nas séries originais,

como indicado pelas cores nos mapas (Figura 2.8). Também se verificou que existiu gradiente de leste a oeste para as escalas baixas e intermediárias, ou seja, na região litorânea os ventos não ocorreram de forma regular, o que torna difícil sua previsibilidade. Por outro lado, nas escalas mais altas, houve maior regularidade, e o mesmo comportamento para EAMM foi observado em todo o estado.

Nas séries de anomalias (Figura 2.9), a entropia se apresentou de forma semelhante em todas as regiões, independentemente dos grupos, para cada valor de τ . No entanto, para a menor escala de tempo, observou-se a maior EAMM, e a partir de $\tau = 8$, a série tornou-se menos complexa e mais regular.

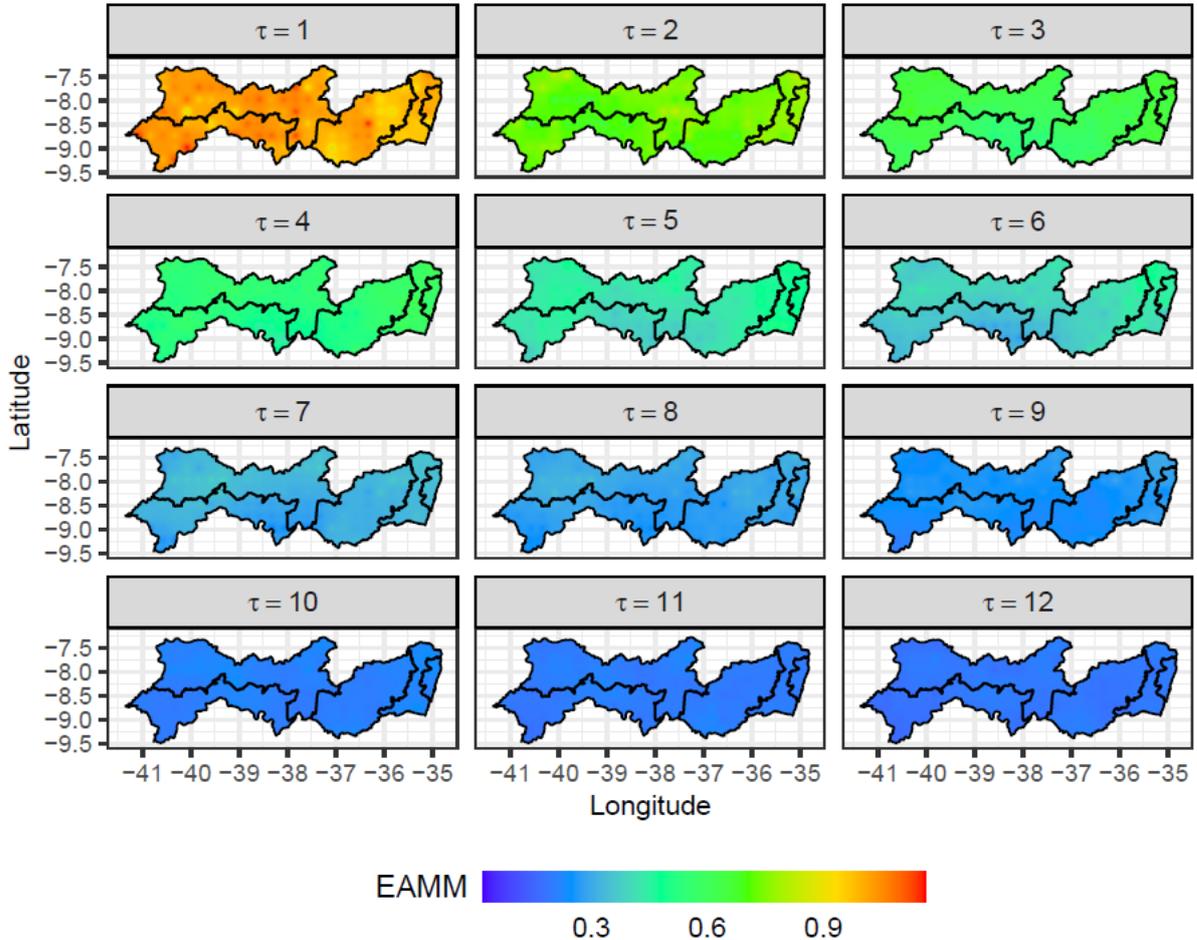
Li et al. (2011) encontraram em seus estudos que a entropia e a velocidade do vento estão correlacionadas positivamente ao analisar a dinâmica do vento no plano horizontal próximo a superfície.

Figura 2.8 - Distribuição espacial EAMM ($\tau = 1, 2, \dots, 12$) para as séries originais da velocidade do vento sobre o estado de Pernambuco.



Fonte: Autor (2022).

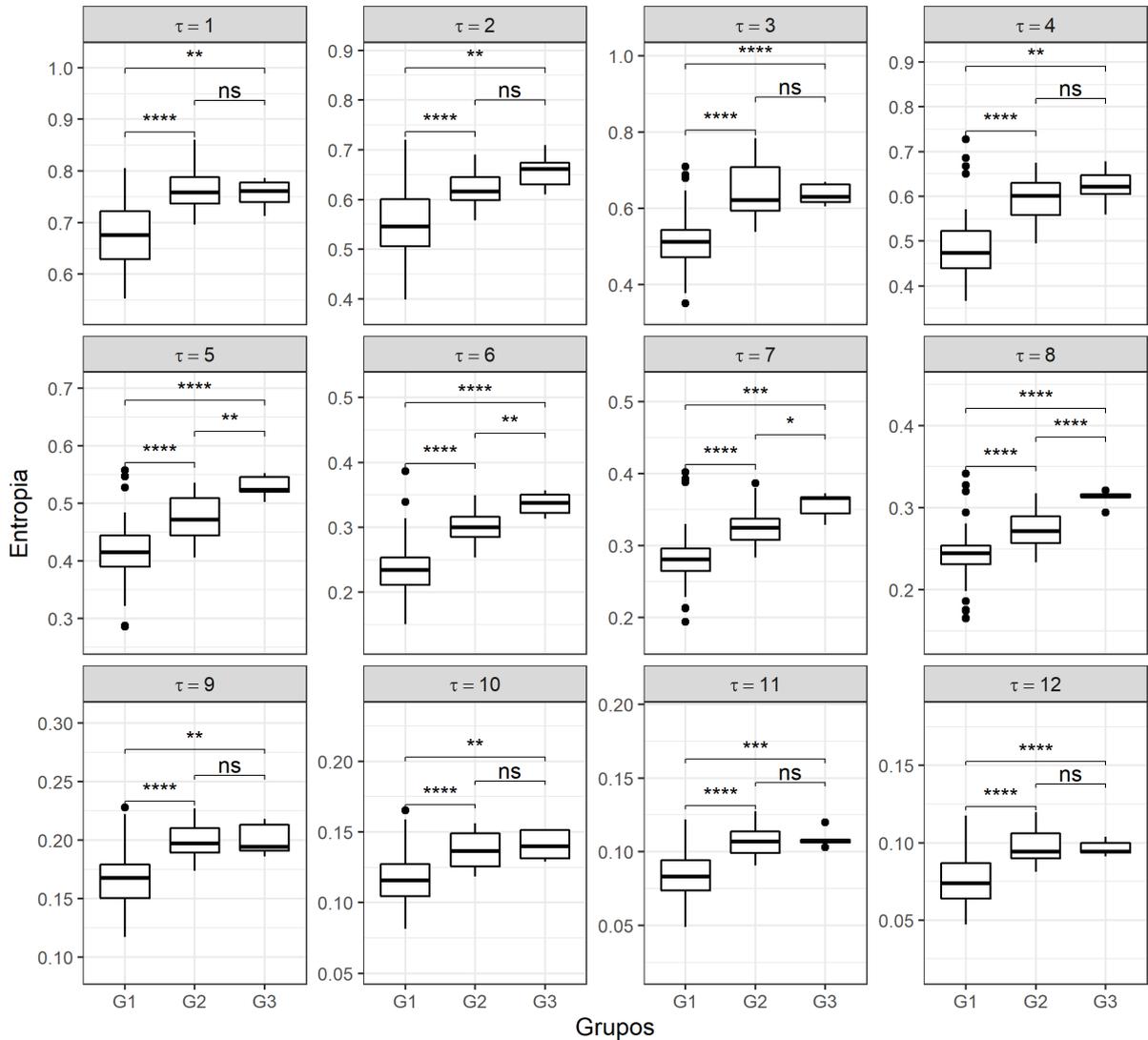
Figura 2.9 - Distribuição espacial EAMM ($\tau = 1, 2, \dots, 12$) para as séries de anomalias da velocidade do vento sobre o estado de Pernambuco.



Fonte: Autor (2022).

O teste Wilcoxon-Mann-Whitney (Mann; Whitney, 1947) foi aplicado para avaliar a diferença nos valores de EAMM entre os três grupos, ao nível de significância de 5%, para as escalas temporais $\tau = 1, 2, \dots, 12$. A Figura 2.10 mostra que, para a série original, houve diferença significativa entre G1 e os outros grupos em todas as escalas de tempo, o que indica uma variabilidade na regularidade da velocidade do vento presente em G1. Entre G2 e G3, foi observada uma diferença significativa somente para $\tau = 5, 6$ e 8 , indicando que esses grupos apresentaram valores de EAMM semelhantes.

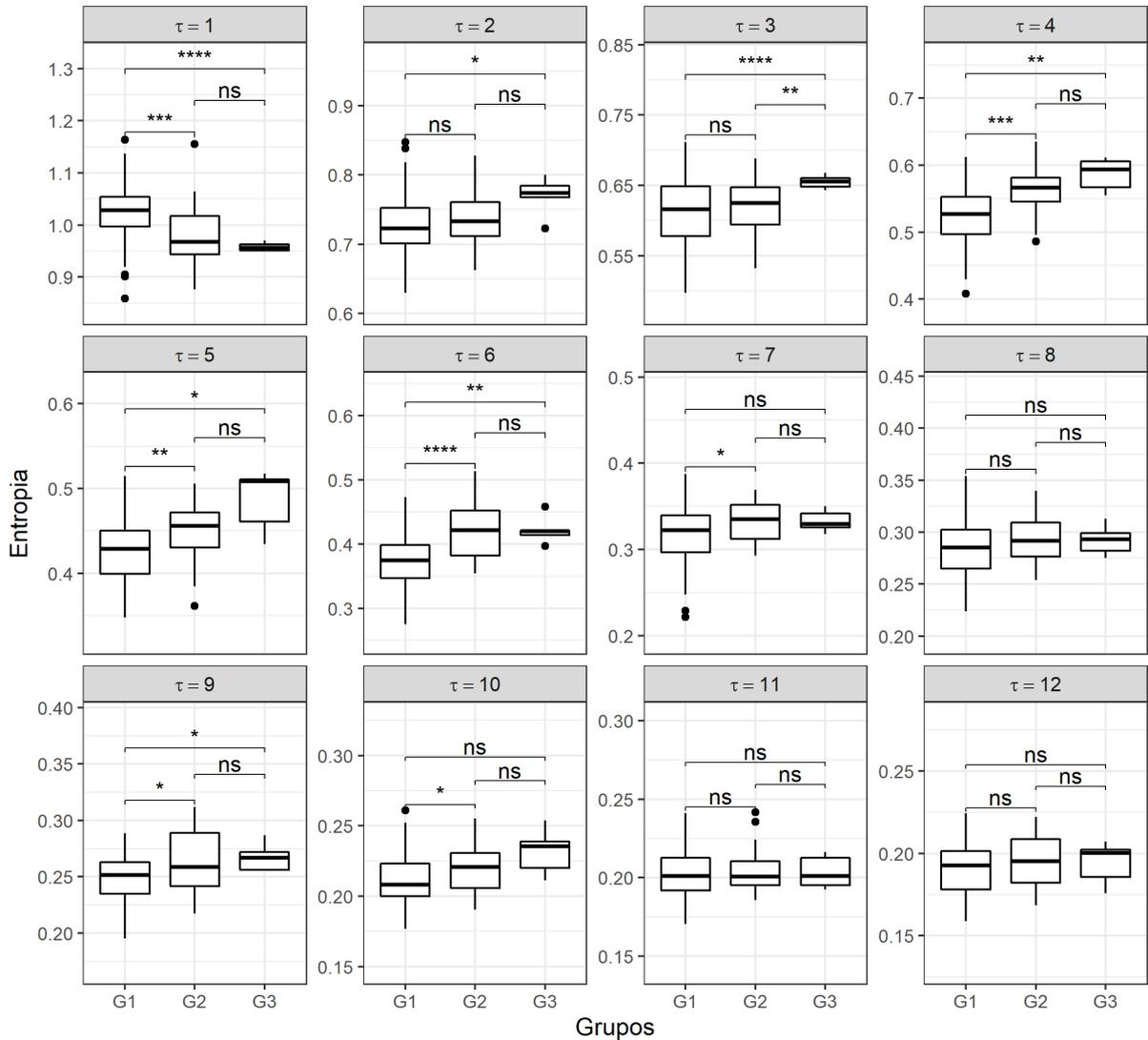
Figura 2.10 - Boxplots para valores EAMM (série original) para G1, G2 E G3, obtidos com a análise de agrupamento, para $\tau = 1, 2, \dots, 12$. “*” indica que há diferença significativa entre os grupos ao nível de significância de 5% e “ns” representa não significativo.



Fonte: Autor (2022).

No caso das anomalias (Figura 2.11), não foi observada diferença significativa entre os grupos G2 e G3, padrão similar ao encontrado nas séries originais, em que a diferença existiu apenas em três escalas. No entanto, para G1 e G2, a diferença significativa ocorreu nas escalas $\tau = 1, 4, 5, 6, 9$ e 10 e para G1 e G3, a diferença ocorreu nas escalas $\tau = 1, 2, 3, 4, 5, 6, 9$ e 10.

Figura 2.11 - Boxplots para valores EAMM (série de anomalia) para G1, G2 E G3, obtidos com a análise de agrupamento, $\tau = 1, 2, \dots, 12$. “*” indica que há diferença significativa entre os grupos ao nível de significância de 5% e “ns” representa não significativo.



Fonte: Autor (2022).

2.4 CONSIDERAÇÕES FINAIS

Com a aplicação da análise de agrupamento foi possível identificar regiões homogêneas em relação à circulação da velocidade média do vento. Constatou-se que os ventos não são uniformemente distribuídos no espaço e no tempo para todas as mesorregiões. As maiores médias foram registadas em um período de quatro meses, durante as estações de inverno e primavera (de junho a setembro), assim como em parte da costa litorânea e no interior do estado. As áreas com maior potencial eólico estão localizadas no Sertão Pernambucano e no São Francisco Pernambucano.

Foi observada maior regularidade nas séries originais da velocidade do vento em relação às séries de anomalias, sendo que, em ambos os conjuntos de dados, os valores de entropia são maiores nas escalas de tempo menores. Em dois dos meses de maior velocidade média, que ocorrem em escalas intermediárias (6 e 7 meses), verificou-se uma queda na entropia para os três pontos representativos dos grupos.

Há evidências de diferença significativa para os valores médios de EAMM entre todos os grupos e escalas de tempo analisadas, com exceção dos grupos G1 e G3 em que não apresentaram diferença significativa nas escalas médias.

Os resultados mostraram que a análise de agrupamento foi adequada para o estudo inicial do padrão da velocidade média do vento no tempo e espaço. Além disso, a abordagem de entropia em múltiplas escalas foi útil para avaliar o quão regular são as séries temporais nas diferentes escalas, o que contribui para o entendimento da dinâmica dessa fonte renovável de energia. Esses resultados podem servir também como base para novos estudos e em diferentes regiões, assim como apoiar decisões sobre investimentos provenientes de setores públicos e/ou privados, formulação de políticas e busca pelo desenvolvimento sustentável.

3 ESTUDO DA VELOCIDADE DO VENTO COM MÉTODOS DE ENTROPIA E ANÁLISE MULTIVARIADA

Resumo: Neste trabalho, foi analisada a complexidade das séries temporais mensais e de anomalia da velocidade do vento no estado de Pernambuco, Brasil, para o período de 1979 a 2020. A análise de entropia multiescala modificada (EAMM) foi empregada para investigar o quão complexas e irregulares são as séries de velocidade do vento em escala variando de um a doze meses. Em seguida, aplicou-se a análise multivariada por meio de métodos de agrupamento para avaliar a homogeneidade dos valores de entropia em cada escala. Observou-se que a regularidade das séries aumentou conforme a escala temporal aumentou, em todas as mesor-regiões. Em relação aos grupos formados, foram identificados três grupos nas escalas entre cinco e oito meses, com o primeiro grupo sendo incorporado ao segundo, e dois grupos bem separados para as demais escalas.

Palavras-chave: agrupamento; complexidade; eólica; homogeneidade; série temporal.

3.1 INTRODUÇÃO

A necessidade de controlar os efeitos das mudanças climáticas fez a sociedade buscar formas para aumentar a produção de energia limpa. Uma destas formas é a conversão da energia cinética do vento em energia elétrica (Chitura; Mukumba; Lethole, 2024). É notável o progresso do mercado eólico, em 2023 foram adicionados 117 GW à capacidade mundial instalada, que já ultrapassa 1 TW (GWEC, 2024).

No cenário brasileiro, o país aumentou sua capacidade instalada em 4,8 GW no último ano ocupando o sexto lugar no *ranking* internacional de capacidade eólica acumulada (GWEC, 2024). A região Nordeste é a principal responsável pelo crescimento e desenvolvimento deste setor, o estado de Pernambuco abriga 43 dos 1.043 parques eólicos distribuídos pelo Brasil (ABEEólica, 2024).

Para compreender a dinâmica deste recurso energético, bem como sua disponibilidade, formas para melhor aproveitá-lo e fazer novos investimentos, são utilizadas técnicas estatísticas. A variável velocidade do vento é complexa e não linear (Rind, 1999), exigindo o uso de métodos que capturem tal comportamento. Abordagens baseadas em entropia têm ganhado destaque por quantificar a regularidade contida em séries temporais (Chen *et al.*, 2024). Guignard et al. (2019a) aplicaram a análise de Fisher-Shannon para obter clareza sobre o compor-

tamento complexo de séries temporais de velocidade do vento em uma área urbana e, Sivhugwana e Ranganai (2024), utilizaram a entropia amostral (EA) para avaliar a complexidade de subséries de velocidade do vento para previsão de curto prazo.

Os algoritmos tradicionais de entropia são medidos em escala única, ou seja, não consideram a variabilidade temporal presente em sistemas complexos, desse modo, Costa, Goldberger e Peng (2002) propuseram o método de entropia amostral em múltiplas escalas (EAM) para avaliar a desordem em séries temporais, considerando as múltiplas escalas de tempo. Porém, para séries curtas, a EAM pode gerar estimativas equivocadas. Para solucionar essa falha, Wu et al. (2013) apresentaram o método de entropia amostral em múltiplas escalas modificada (EAMM). Ainda é pouco explorada a aplicação do método modificado em pesquisas voltadas para o setor eólico e análises de velocidade do vento.

Em pesquisas climatológicas e meteorológicas a análise de agrupamento apresenta bom desempenho para identificar padrões e estruturas dentro da base de dados (Pampuch *et al.*, 2023). Wang *et al.* (2024) realizaram um estudo sobre as características do perfil do vento do Super Tufão Lekima 2019, usando a distância euclidiana e métodos de agrupamento hierárquico.

Este estudo teve por objetivo descobrir possíveis padrões existentes nos valores de EAMM por meio dos métodos de agrupamento hierárquico, para séries mensais de velocidade do vento para o estado de Pernambuco, nordeste do Brasil, no período de 1979 a 2020. A pesquisa é especialmente importante no contexto da redução dos efeitos das mudanças climáticas.

3.2 MATERIAL E MÉTODOS

3.2.1 Base de dados e localização da área de estudo

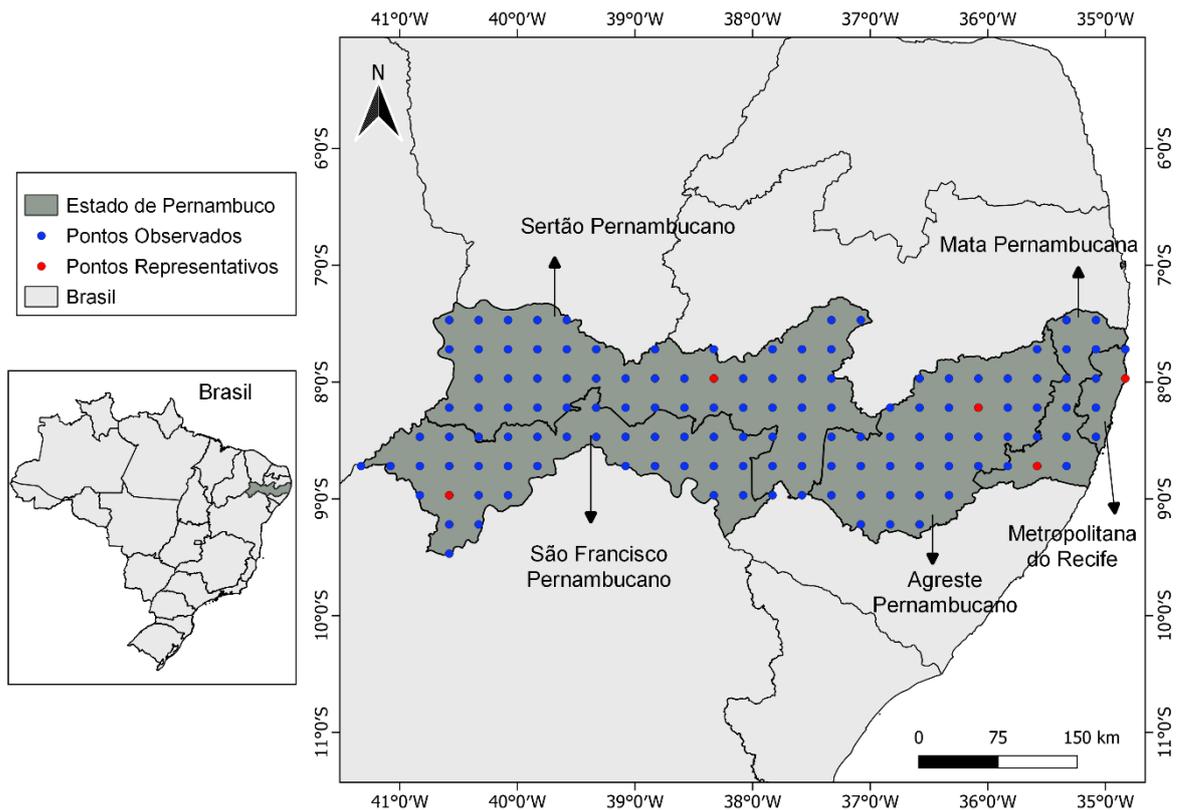
As médias mensais dos dados horários da velocidade do vento utilizados neste trabalho, referem-se ao período de 1979 a 2020 e são provenientes da ERA5 (Hersbach *et al.*, 2020), reanálise atmosférica de quinta geração do *European Centre for Medium-Range Weather Forecasts* (ECMWF). Esses dados estão disponíveis por meio do *Copernicus Climate Data Store* (Hersbach *et al.*, 2021), como componentes do vento, zonal (u) e meridional (v), medidas a 10 m do solo, com resolução espacial de aproximadamente 31 km. A partir destas componentes a velocidade escalar do vento (m/s) foi calculada como em Molina, Gutiérrez e Sánchez (2021):

$$x(t) = \sqrt{u^2 + v^2},$$

em que u e v representam a direção do vento no sentido Oeste-Leste e Norte-Sul, respectivamente.

Foram observados 133 pontos de grade distribuídos no estado de Pernambuco, localizado na região Nordeste do Brasil, precisamente entre os paralelos $7^{\circ}15'45''$ S e $9^{\circ}28'18''$ S e os meridianos $34^{\circ}48'33''$ W e $41^{\circ}19'54''$ W. E, tem como limites os estados da Paraíba e Ceará (norte), Piauí (oeste), Alagoas e Bahia (sul) e ao leste, o Oceano Atlântico. Pernambuco é dividido em cinco mesorregiões: Sertão Pernambucano, São Francisco Pernambucano, Agreste Pernambucano, Mata Pernambucana e Metropolitana do Recife (Figura 3.1).

Figura 3.1 - Localização geográfica do estado de Pernambuco, Brasil, e distribuição dos pontos de grade da reanálise ERA5. Os círculos vermelhos indicam a localização dos pontos representativos das cinco mesorregiões do estado (Sertão Pernambucano, São Francisco Pernambucano, Agreste Pernambucano, Mata Pernambucana e Metropolitana do Recife).



Fonte: Autor (2023).

3.2.2 Entropia amostral em múltiplas escalas modificada

Entropia amostral (EA) mede a complexidade e regularidade de séries temporais em escala única. Entropia amostral (EA) é o logaritmo natural negativo da probabilidade condi-

onal, de que duas séries de dados de comprimento N sejam semelhantes dentro de um nível de tolerância r para m pontos e permaneçam assim para $m + 1$ pontos, nas quais as auto correspondências não são incluídas no cálculo da probabilidade (Richman; Moorman, 2000). Neste trabalho foi utilizado o algoritmo EAMM, proposto por Wu *et al.* (2013) como melhoria da estimativa de EAM (Costa; Goldberger; Peng, 2002), ou seja, estimar EA para múltiplas escalas da velocidade do vento, permitindo uma análise mais robusta da complexidade dessas séries em diferentes escalas de tempo.

Foi gerada uma nova série $x_\tau(j)$ para a série de velocidade do vento $x(i)$, $i = 1, \dots, N$, com o procedimento de média móvel para cada fator de escala τ , como segue (Wu *et al.*, 2013):

$$x_\tau(j) = \frac{1}{\tau} \sum_{i=j}^{j+\tau-1} x(i) \quad \text{e} \quad j = 1, \dots, N - \tau + 1.$$

Posteriormente, foram definidos os vetores de tamanho m , em que $x_m^\tau(i) = \{x(i), x(i + 1), \dots, x(i + m)\}$, $i = 1, \dots, N - \tau + 1 - m\tau$, com atraso de tempo τ . A distância entre os vetores $x_m(i)$ e $x_m(j)$ foi calculada como a máxima diferença entre seus componentes escalares correspondentes (Yin *et al.*, 2018):

$$d_m[x_i, x_j] = \max_{0 \leq k \leq m-1} |x(i+k) - x(j+k)|, 1 \leq i, j \leq N - \tau + 1 - m\tau, i \neq j.$$

Os vetores $x_m(i)$ e $x_m(j)$ são semelhantes dentro de um nível de tolerância r , se $d_m[x_i, x_j] \leq r$ ($r = r\sigma$, em que σ é o desvio-padrão de $x_\tau(j)$). O número de vetores semelhantes para m pontos e para $m + 1$ pontos é representado, respectivamente, por B e A . Assim, a entropia amostral é estimada por:

$$S_E(m, r, N, \tau) = -\ln \frac{A}{B}.$$

A regularidade está associada aos valores de EA, quanto menor este valor, mais regulares e menos complexas são as séries. Para a análise, foram utilizados os parâmetros padrão para entropia, que são $m = 2,0$ e $r = 0,2$, conforme recomendado por Pincus, (1991) e Richman; Moorman, (2000). O fator de escala foi de um a doze meses, ou seja, $\tau = 1, \dots, 12$, e o *software* utilizado foi o R Core Team (2021) com o pacote *TSEntropies* (Tomcala, 2018).

A EAMM foi calculada para as séries de médias mensais de velocidade do vento original e de anomalia para todos os pontos. Foi dado destaque para cinco pontos, escolhidos de forma aleatória para representar cada uma das mesorregiões do estado de Pernambuco.

3.2.3 Análise de agrupamento

Métodos hierárquicos aglomerativos foram empregados para agrupar os pontos de grade estudados de acordo com os valores de EAMM das séries temporais da velocidade do vento. De acordo com Hair *et al.* (2019), com esta técnica cada ponto começou como seu próprio grupo e foram unidos sucessivamente com os dois mais semelhantes, até que se formaram um único grupo contendo todos os pontos observados.

As seguintes medidas de distância foram utilizadas no cálculo da matriz de dissimilaridade, que expressou a diferença entre os pares de observações, ou seja, entre os pares de valores de EAMM. Assim, x_{il} e x_{jl} são os valores da l -ésima velocidade nos pontos x_i e x_j , respectivamente:

- Euclidiana – Kumar; Chhabra; Kumar, (2014); Li *et al.*, (2019); Modiri; Bárdossy, (2021):

$$d(x_i, x_j) = \sqrt{\sum_{l=1}^n (x_{il} - x_{jl})^2}.$$

- Canberra - Kumar; Chhabra; Kumar (2014):

$$d(x_i, x_j) = \sum_{l=1}^n \frac{|x_{il} - x_{jl}|}{|x_{il}| + |x_{jl}|}.$$

- Manhattan - Kumar; Chhabra; Kumar, (2014); Strauss; von Maltitz, (2017); Zelterman, (2015):

$$d(x_i, x_j) = \sum_{l=1}^n |x_{il} - x_{jl}|.$$

- Máxima - Zelterman, (2015):

$$d(x_i, x_j) = \max_{1 \leq j \leq n} |x_{il} - x_{jl}|.$$

- Minkowski - Strauss; von Maltitz, (2017); Xu; Tian, (2015); Zelterman, (2015):

$$d(x_i, x_j) = \left(\sum_{l=1}^n |x_{il} - x_{jl}|^p \right)^{1/p}, p \geq 1.$$

Esta distância é uma generalização das distâncias de Manhattan ($p = 1$) e euclidiana ($p = 2$), sendo p qualquer número real maior que um.

Os métodos aglomerativos utilizados para a formação dos grupos são mostrados na Tabela 3.1, em que d_{XY} é a distância entre os grupos X e Y ; d_{ij} é a distância entre os pontos i

e j ; n_X e n_Y representam o número de pontos nos grupos X e Y , respectivamente; \bar{x} e \bar{y} são as médias dos grupos X e Y , respectivamente.

Tabela 3.1 – Métodos de agrupamento aglomerativos.

Método	Equação	Referência
Ligação simples	$d_{XY} = \min\{d_{ij} : i \in X; j \in Y\}$	Li; Rezaeipannah; Tag El Din, (2022); Modiri; Bárdossy, (2021)
Ligação completa	$d_{XY} = \max\{d_{ij} : i \in X; j \in Y\}$	Li; Rezaeipannah; Tag El Din, (2022); Modiri; Bárdossy, (2021)
Ligação média	$d_{XY} = \frac{1}{n_X n_Y} \sum_{i \in X} \sum_{j \in Y} d_{ij}$	Li; Rezaeipannah; Tag El Din, (2022); Modiri; Bárdossy, (2021); Santos <i>et al.</i> (2019)
Centroide	$d_{XY} = d_{\bar{x}\bar{y}}$	Li; Rezaeipannah; Tag El Din, (2022)
Ward	$d_{XY} = \sqrt{\frac{2n_X n_Y}{(n_X + n_Y)}} \ \bar{x} - \bar{y}\ ^2$	Modiri; Bárdossy, (2021); Ward (1963)

3.2.3.1 Seleção de métodos

O coeficiente de correlação cofenética (r) (Sokal; Rohlf, 1962) foi avaliado para selecionar a medida de distância e o método de agrupamento, dentre as combinações obtidas que melhor representa os agrupamentos formados pelos valores de EAMM. Este coeficiente é expresso por:

$$r = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 (y_i - \bar{y})^2}},$$

em que x_i é o valor da variável na matriz de dissimilaridade, y_i é o valor da variável na matriz cofenética, \bar{x} e \bar{y} são as médias das variáveis nas matrizes de dissimilaridade e cofenética, respectivamente. Segundo Bussad, Miazaki e Andrade (1990), r deve apresentar valores acima de 0,7 para indicar boa representatividade entre as distâncias.

3.2.3.2 Número de grupos

O pacote *NbClust* do *software* R Core Team (2021) foi utilizado para determinar o número ideal de grupos para os pontos de grade estudados, de acordo com os valores de EAMM. Este pacote trabalha com índices que combinam informações sobre a união de objetos dentro de um mesmo grupo e a separação entre os grupos, além de considerar as medidas de similaridade ou dissimilaridade (Charrad *et al.*, 2014).

Foi utilizado o critério da maioria dos índices para selecionar o número e a distribuição adequada dos pontos dentro dos grupos. As seguintes medidas contidas no pacote foram observadas: ch; ball; beale; ccc; cindex; db; duda; dunn; friedman; gamma; gplus; hartigan; kl; marriot; mcclain; pseudot2; ptbserial; ratkowsky; rubin; scott; sdbw; sdindex; silhouette; tau e tracew.

3.3 RESULTADOS E DISCUSSÃO

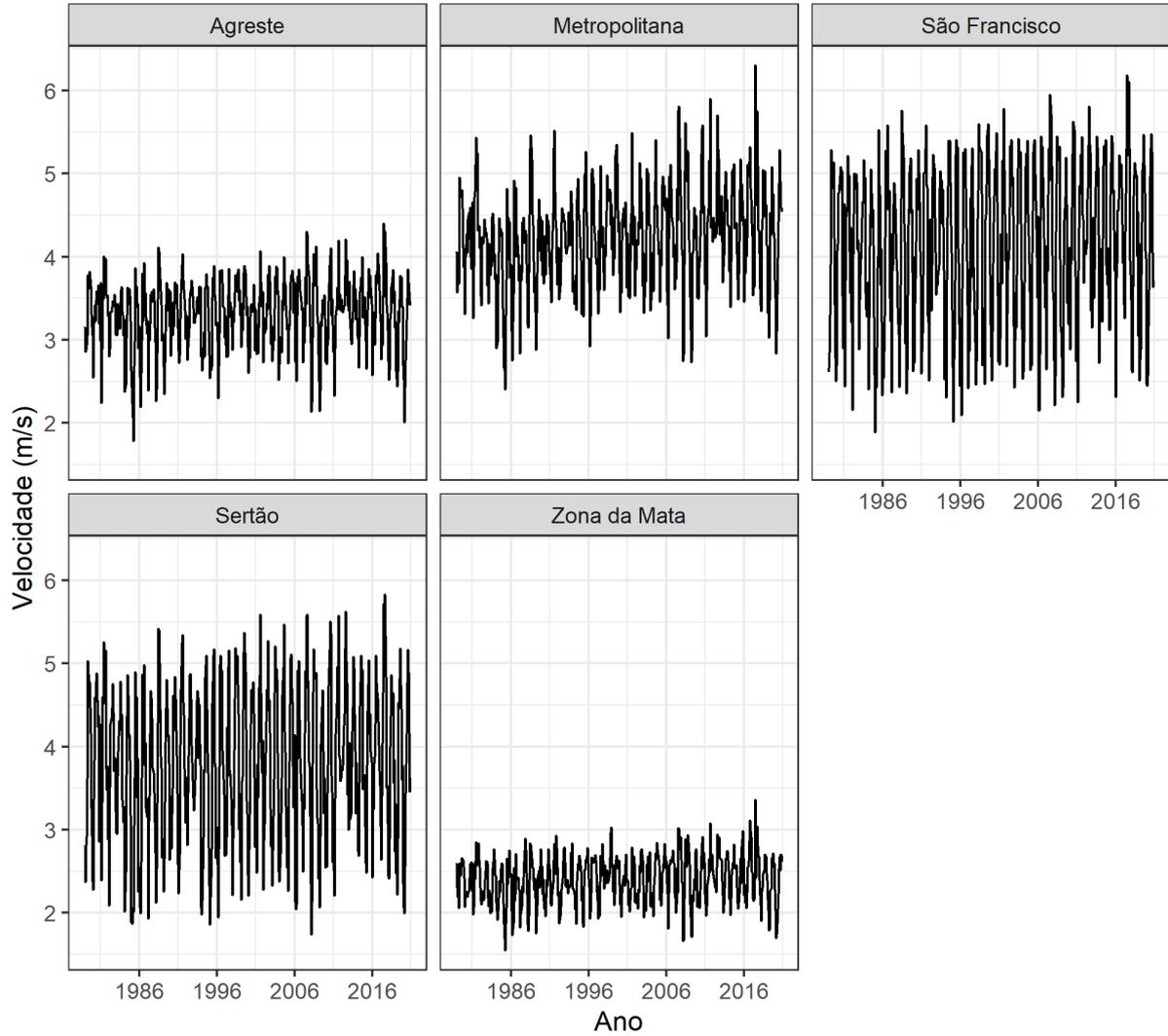
Foram analisadas as séries originais da velocidade do vento e as séries de anomalias para os 133 pontos de grade observados no estado de Pernambuco, durante o período de 1979 a 2020. Dada a série original $x(t)$, as anomalias mensais foram calculadas conforme Bhatia et al. (2020) e Kantelhardt et al. (2006):

$$X(t) = \frac{x(t) - \mu_t}{\sigma_t},$$

em que μ_t e σ_t são a média e o desvio padrão, respectivamente, da velocidade do vento mensal $x(t)$ calculada para cada mês de todos os anos.

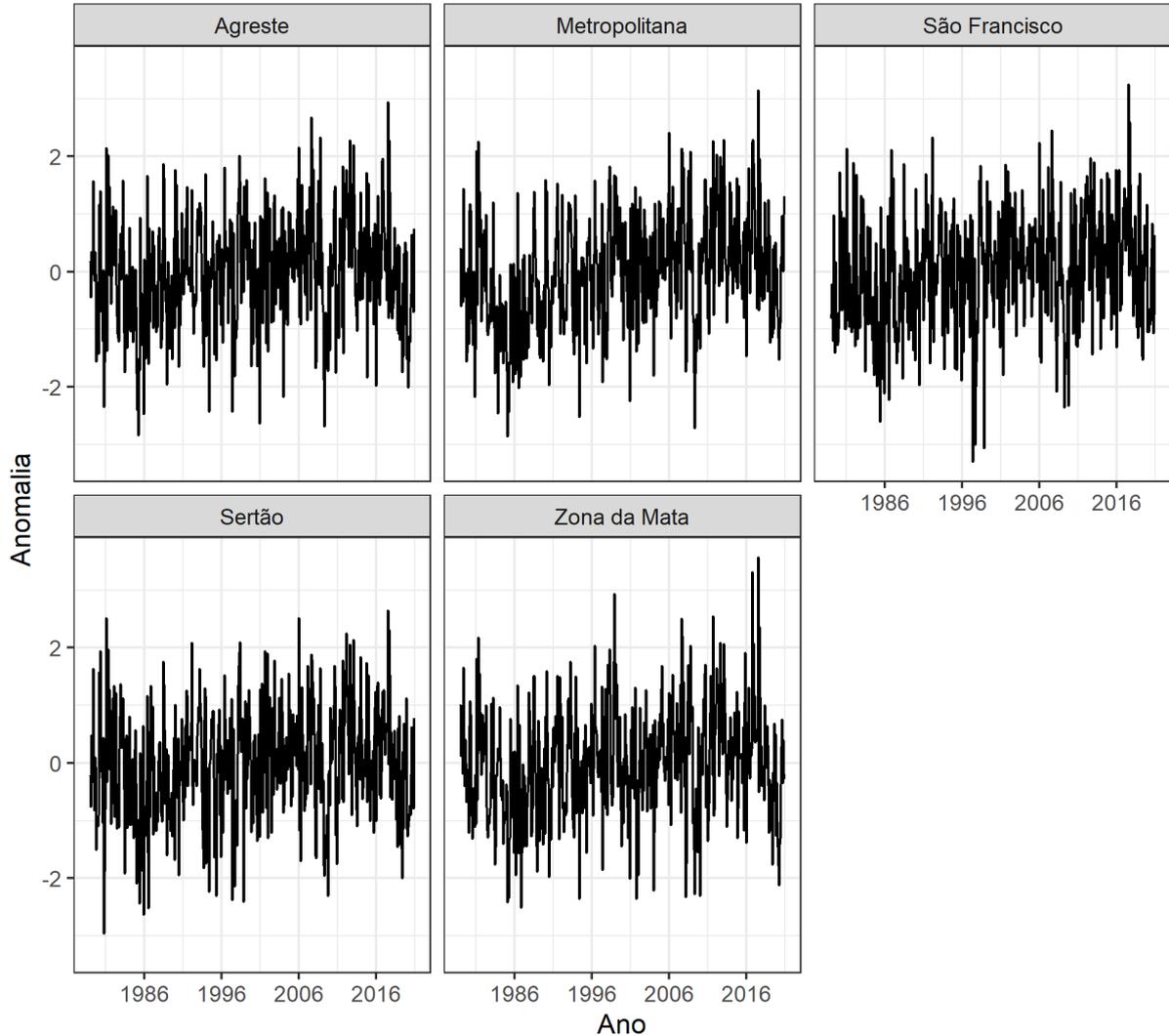
Na Figura 3.2 e Figura 3.3 são apresentadas as séries temporais mensais, originais e de anomalias, respectivamente, para cinco pontos representativos localizados nas mesorregiões: Agreste (lat = -8,22, lon = -36,08), Metropolitana do Recife (lat = -7,97, lon = -34,83), São Francisco (lat = -8,97, lon = -40,58), Sertão (lat = -7,79, lon = -38,33) e Zona de Mata (lat = -8,72, lon = -35,58).

Figura 3.2 - Séries temporais originais da velocidade média mensal do vento para pontos representativos das mesorregiões de Pernambuco.



Fonte: Autor (2023).

Figura 3.3 - Séries temporais das anomalias da velocidade média mensal do vento para pontos representativos das mesorregiões de Pernambuco.



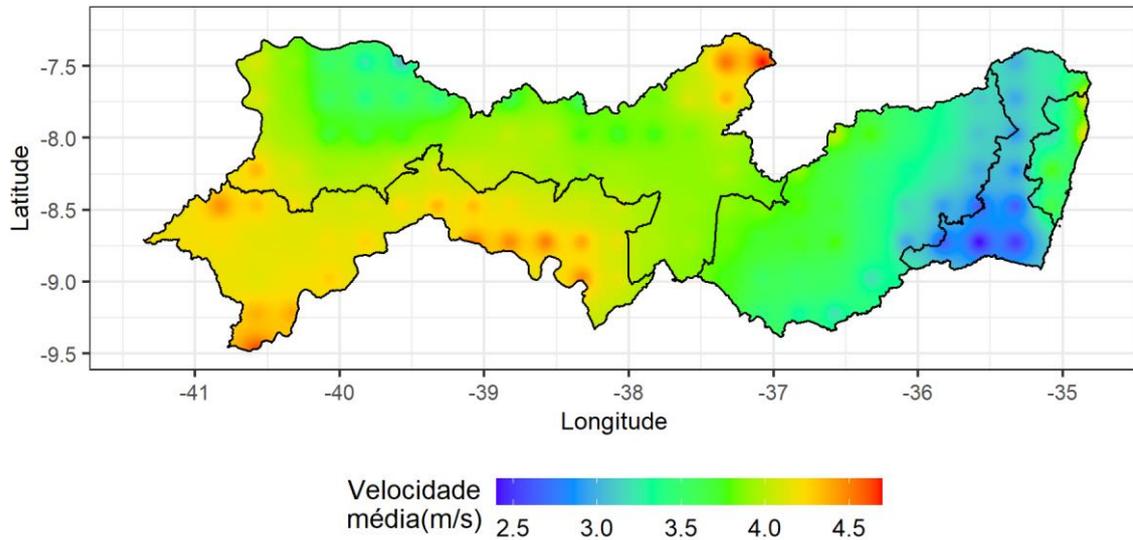
Fonte: Autor (2023).

Nas séries originais, as médias mais baixas foram observadas para Zona da Mata (ou Mata Pernambucana), mesorregião que também apresentou a menor variabilidade. Por outro lado, as mesorregiões São Francisco e Sertão apresentaram alta variabilidade e médias elevadas, já a Metropolitana exibiu altas médias e variabilidade intermediária em relação as demais regiões (Figura 3.2). Com a retirada do efeito de sazonalidade, as séries de anomalias se mostraram com comportamento semelhante em todos os pontos, com alta variabilidade (Figura 3.3)

A distribuição espacial da velocidade média sobre toda a área de estudo, interpolada pelo método inverso da distância ponderada – IDP (Shepard, 1968), é apresentada na Figura 3.4, em que se observou ventos menos intensos concentrados nas mesorregiões Mata e Agreste e, os mais intensos em São Francisco, Sertão e em parte do litoral da mesorregião Metro-

litana do Recife. Este padrão é similar ao observado nas séries originais dos pontos representativos na Figura 3.2.

Figura 3.4 - Dados médios mensais da velocidade do vento, para o período de 1979 a 2020, interpolados sobre o estado de Pernambuco.

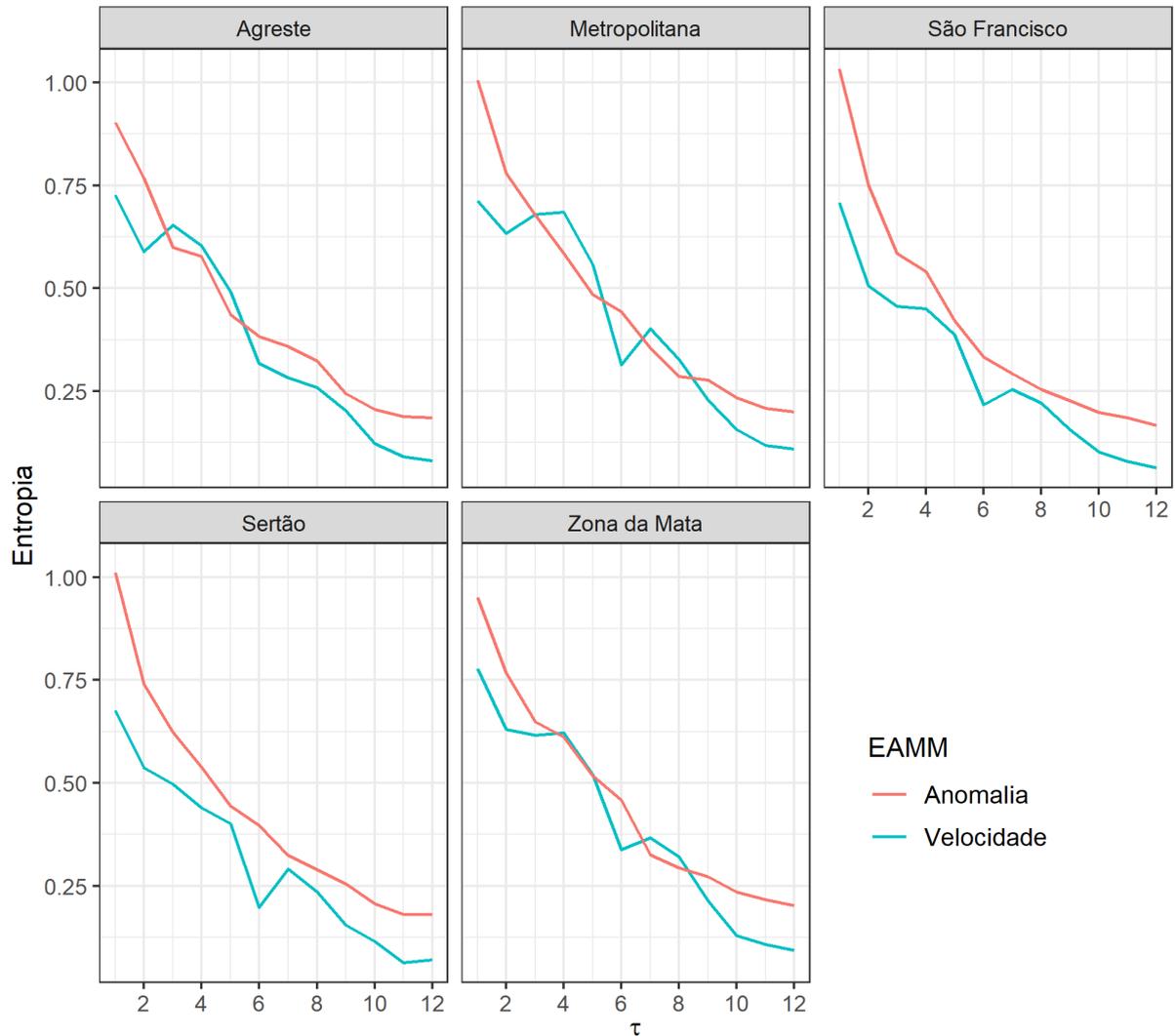


Fonte: Autor (2023).

Os valores de EAMM foram calculados para $\tau = 1, 2, \dots, 12$ tanto para as séries originais quanto para as de anomalias da velocidade média do vento para todos os pontos. Na Figura 3.5 são apresentados os resultados para os cinco pontos representativos das mesorregiões do estado. De maneira geral, a EAMM foi menor nas séries originais devido ao efeito sazonal do regime dos ventos, indicando maior regularidade e previsibilidade em comparação com as séries de anomalias. Somente para escalas entre $\tau = 3$ e $\tau = 8$ a entropia da série original foi maior do que a da anomalia para as mesorregiões Agreste, Mata e Metropolitana do Recife.

Com o aumento da escala temporal, a entropia apresentou um padrão decrescente em todas as mesorregiões e para ambas as séries, mostrando que os ventos foram mais regulares quando observados em escalas mais altas. Com exceção do Agreste, região de transição climática, todas as demais mesorregiões mostraram uma mudança de comportamento nas séries originais, em que exibiram tendência de queda nos valores de EAMM, com um aumento observado entre as escalas $\tau = 6$ e $\tau = 7$.

Figura 3.5 - EAMM para as séries de médias mensais de velocidade do vento (original e anomalia) para cinco estações representativas das mesorregiões do estado de Pernambuco.

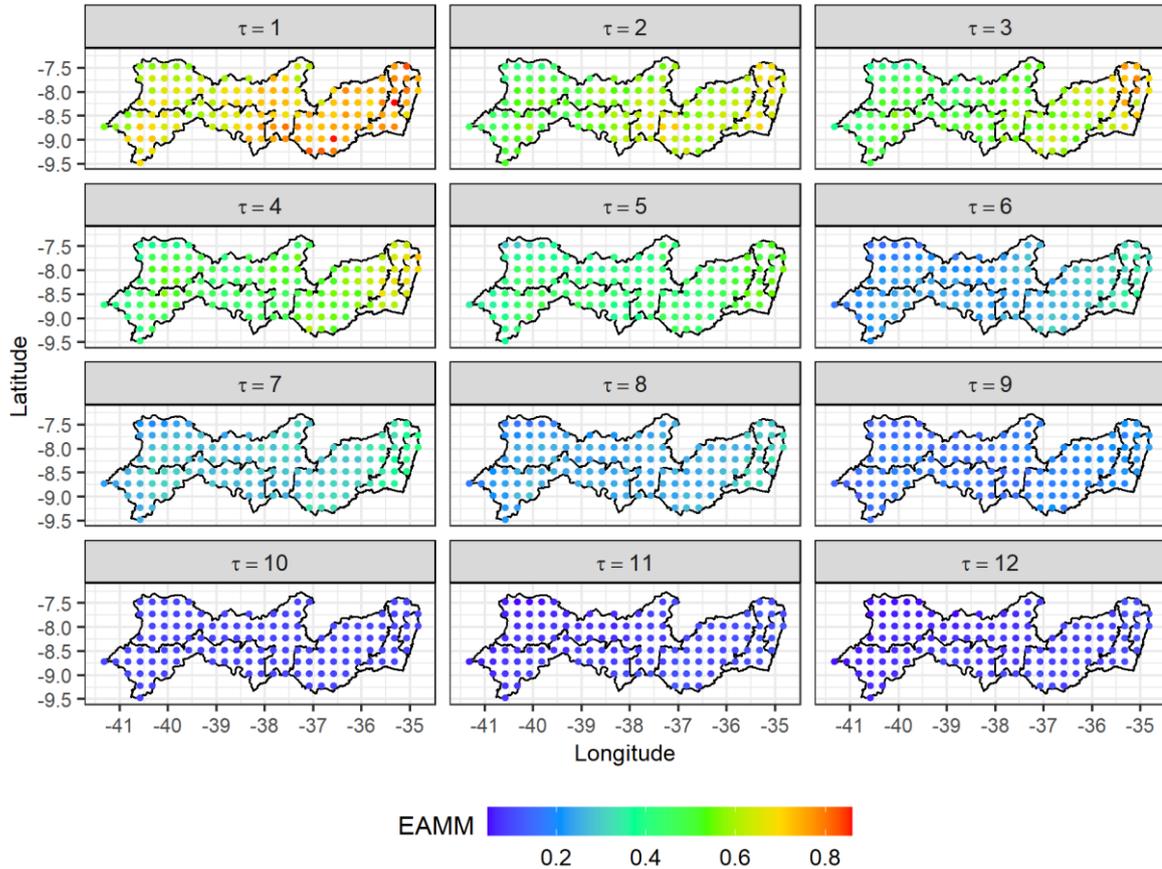


Fonte: Autor (2023).

Na Figura 3.6, é apresentada a distribuição espacial dos valores de EAMM da série original para os 133 pontos de grade, observados em todo o estado de Pernambuco. Corroborando com os resultados dos pontos representativos (Figura 3.5), a entropia diminuiu com o aumento da escala temporal em todas as mesorregiões.

Foi observado que, até a escala de oito meses ($\tau = 8$) que os valores de EAMM aumentaram no sentido do interior para o litoral do Estado. Já nas escalas mais altas a entropia apresentou os valores mais baixos em toda a área de estudo. Agreste, Zona da Mata e a Metropolitana do Recife são as mesorregiões que se destacaram com a maior entropia para $\tau = 1$, indicando que nessas áreas e para a menor escala, a dinâmica dos ventos ocorre de forma irregular e menos previsível.

Figura 3.6 - Distribuição espacial do EAMM ($\tau = 1, 2, \dots, 12$) no estado de Pernambuco para as séries originais da velocidade do vento.



Fonte: Autor (2023).

O coeficiente de correlação cofenética (r) foi avaliado para selecionar a combinação da medida de distância e método de agrupamento que melhor representa os grupos formados pelos valores de entropia. Na Tabela 3.2, é apresentado o arranjo das métricas que atingiu o maior valor de r em cada escala de tempo, seus respectivos coeficientes e o número de grupos considerado ideal, de acordo com o método da maioria para os índices do pacote *NbClust*.

Para as escalas de tempo de cinco a oito meses foram estruturados três grupos, todos com a distância Canberra. Utilizou-se o método centroide para $\tau = 5, 6$ e 7 e o método de ligação média para $\tau = 8$, sendo esta última escala a que apresentou o maior valor de r (0,829) indicando maior adequação do método para o agrupamento das observações. Nas demais escalas foram formados dois grupos, utilizando as distâncias máxima, euclidiana e Canberra, e os métodos de ligação média e ligação completa.

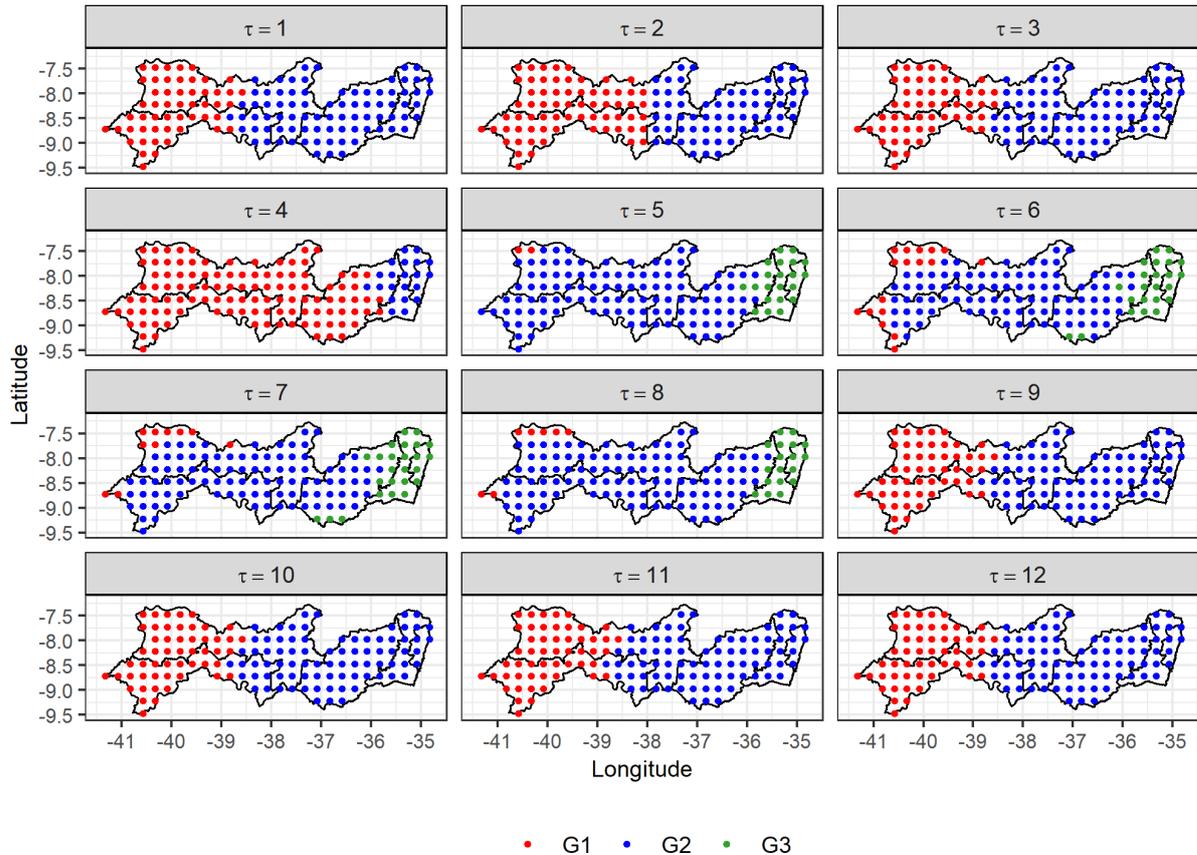
Tabela 3.2 - Medidas de distância e métodos de agrupamento selecionados, por apresentarem maior coeficiente de correlação cofenética para cada mês.

τ	Distância	Método	r	Nº de grupos
1	máxima	ligação média	0,725	2
2	euclidiana	ligação completa	0,730	2
3	euclidiana	ligação completa	0,733	2
4	Canberra	ligação média	0,732	2
5	Canberra	centroide	0,739	3
6	Canberra	centroide	0,740	3
7	Canberra	centroide	0,734	3
8	Canberra	ligação média	0,829	3
9	máxima	ligação média	0,725	2
10	máxima	ligação média	0,725	2
11	máxima	ligação média	0,725	2
12	máxima	ligação média	0,725	2

Fonte: Autor (2023).

A Figura 3.7 mostra a configuração dos grupos formados em cada escala. Como observado na Figura 3.5 e Figura 3.6, os maiores valores de entropia ocorreram na primeira escala, enquanto os menores foram registrados nas escalas finais. Com a análise de agrupamento foram constituídos dois grupos idênticos para as escalas $\tau = 1, 9, 10, 11, 12$. A medida de distância e o método que melhor aglomerou as observações foram distância máxima e ligação média, respectivamente, com o menor coeficiente de correlação cofenética (0,725) dentre todas as escalas.

Figura 3.7 - Grupos formados a partir dos valores de EAMM.



Fonte: Autor (2023).

3.4 CONSIDERAÇÕES FINAIS

A complexidade das séries temporais da velocidade do vento e de anomalias foi avaliada por meio da EAMM. Observou-se que, para ambas as séries, a irregularidade diminui com o aumento da escala temporal. Os menores valores de entropia foram encontrados nas escalas finais, e a análise de agrupamento formou grupos idênticos nas referidas escalas.

É relevante destacar a abordagem empregada, com o cálculo da entropia em múltiplas escalas para períodos curtos, aplicada a fenômenos climáticos, especificamente para velocidade do vento. Estudos desse tipo são escassos. Além disso, conhecer a homogeneidade dos valores de EAMM distribuídos na área de estudo e dentro de um intervalo de escalas temporais é fundamental para planejamento de investimentos e novas instalações eólicas.

4 CONSIDERAÇÕES GERAIS

Este estudo analisou o comportamento espaço-temporal da velocidade do vento no estado de Pernambuco, nordeste do Brasil, de acordo com as informações do período de 1979 a 2020, obtidas da reanálise ERA-5. Os resultados mostraram que a aplicação da análise de agrupamento foi adequada para tanto o estudo inicial a respeito do padrão da velocidade média do vento no tempo e espaço, assim como, para revelar padrão na entropia para diversas escalas temporais.

Nas séries de velocidade do vento, foram identificados três grupos para todos os meses, com igualdade na composição para junho, julho e agosto, meses em que se concentram as maiores médias de velocidade do vento. Em relação, a EAMM, foi observada a existência de dois grupos para as quatro primeiras e para as quatro últimas escalas de tempo, enquanto para as escalas intermediárias os valores de entropia foram classificados em três grupos.

A regularidade das séries foi maior para a velocidade média, comparada com a anomalia, assim como, para as séries aleatórias. Ao observar o aumento das escalas temporais, verificou-se a diminuição nos valores de EAMM para todas as mesorregiões do estado.

Com método IDP foi observado que as mesorregiões de Sertão e de São Francisco apresentaram as maiores velocidades médias. Em contrapartida, com a espacialização da entropia foi verificada instabilidade dos ventos na região litorânea e nas escalas iniciais.

O presente estudo, contribui para um melhor entendimento do comportamento do regime dos ventos. Como trabalhos futuros, pode expandir a aplicação para os outros estados do Nordeste; encontrar padrões de escala cruzando com outras séries de variáveis, como temperatura, umidade e direção do vento; e ainda realizar a aplicação da entropia cruzada multiescala.

REFERÊNCIAS

- ABEEÓLICA. **Associação Brasileira de Energia Eólica. Boletim Anual 2020.** [S. l.], 2020. Disponível em: http://abeeolica.org.br/wp-content/uploads/2020/06/InfoventoEN_16.pdf. Acesso em: 19 ago. 2020.
- ABEEÓLICA. **Associação Brasileira de Energia Eólica. Boletim Anual 2023.** [S. l.: s. n.], 2024. Disponível em: https://abeeolica.org.br/wp-content/uploads/2024/07/424_ABEEOLICA_BOLETIM-ANUAL-2024_DIGITAL_PT_V3.pdf. Acesso em: 28 jul. 2024.
- ALVES, E. D. L. *et al.* **Revista Brasileira de Geografia Física v.09.** [S. l.: s. n.], 2016. Disponível em: www.ufpe.br/rbgfe. .
- AMATO, F. *et al.* Spatio-temporal estimation of wind speed and wind power using extreme learning machines: predictions, uncertainty and technical potential. **Stochastic Environmental Research and Risk Assessment**, [s. l.], v. 36, n. 8, p. 2049–2069, 2022.
- ATLAS EÓLICO E SOLAR DE PERNAMBUCO. [S. l.], 2015. Disponível em: <http://www.atlaseolicosolar.pe.gov.br/>. Acesso em: 4 set. 2020.
- AZAM, A. *et al.* Analyzing the relationship between economic growth and electricity consumption from renewable and non-renewable sources: Fresh evidence from newly industrialized countries. **Sustainable Energy Technologies and Assessments**, [s. l.], v. 44, 2021.
- BADDELEY, A.; TURNER, R. spatstat: An R Package for Analyzing Spatial Point Patterns. **Journal of Statistical Software**, [s. l.], v. 12, n. 6, p. 1–42, 2005.
- BHATIA, N. *et al.* Role of Cluster Validity Indices in Delineation of Precipitation Regions. **Water**, [s. l.], v. 12, n. 5, p. 1372, 2020.
- BURLANDO, M.; ANTONELLI, M.; RATTO, C. F. Mesoscale wind climate analysis: Identification of anemological regions and wind regimes. **International Journal of Climatology**, [s. l.], v. 28, n. 5, p. 629–641, 2008.
- BUSSAD, W. O.; MIAZAKI, E. S.; ANDRADE, D. F. **Introdução à análise de agrupamentos.** ABEd. [S. l.: s. n.], 1990.
- CARREÓN-SIERRA, S. *et al.* Cluster analysis of the wind events and seasonal wind circulation patterns in the Mexico City region. **Atmosphere**, [s. l.], v. 6, n. 8, p. 1006–1031, 2015.
- CHARRAD, M. *et al.* **NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set** Véronique Boiteau **JSS Journal of Statistical.** [S. l.: s. n.], 2014. Disponível em: <http://www.jstatsoft.org/>. .
- CHEN, Z. *et al.* Tsallis Entropy-Based Complexity-IPE Casualty Plane: A Novel Method for Complex Time Series Analysis. **Entropy**, [s. l.], v. 26, n. 6, p. 521, 2024.

CHITURA, A. G.; MUKUMBA, P.; LETHOLE, N. Enhancing the Performance of Savonius Wind Turbines: A Review of Advances Using Multiple Parameters. **Energies**, [s. l.], v. 17, n. 15, p. 3708, 2024.

CORREIA FILHO, W. L. F. *et al.* The wind regime over the Brazilian Southeast: Spatial and temporal characterization using multivariate analysis. **International Journal of Climatology**, [s. l.], p. 1–22, 2021.

COSTA, M.; GOLDBERGER, A. L.; PENG, C. K. Multiscale Entropy Analysis of Complex Physiologic Time Series. **Physical Review Letters**, [s. l.], v. 89, n. 6, 2002.

DARWISH, M. M. *et al.* New hourly extreme precipitation regions and regional annual probability estimates for the UK. **International Journal of Climatology**, [s. l.], v. 41, n. 1, p. 582–600, 2020.

DOS ANJOS, P. S. *et al.* Long-term correlations and cross-correlations in wind speed and solar radiation temporal series from Fernando de Noronha Island, Brazil. **Physica A: Statistical Mechanics and its Applications**, [s. l.], v. 424, p. 90–96, 2015.

GEBERT, D. M. P.; KIST, A.; VIRGENS FILHO, J. S. das. Determinação de regiões homogêneas de precipitação pluviométrica no estado do Paraná por meio de técnicas de análise multivariada e geoestatística. **Revista Brasileira de Climatologia**, [s. l.], v. 14, 2018.

GILLILAND, J. M.; KEIM, B. D. Surface wind speed: trend and climatology of Brazil from 1980–2014. **International Journal of Climatology**, [s. l.], v. 38, n. 2, p. 1060–1073, 2017.

GUIGNARD, F. *et al.* Fisher–Shannon Complexity Analysis of High-Frequency Urban Wind Speed Time Series. **Entropy**, [s. l.], v. 21, n. 1, p. 47, 2019a.

GUIGNARD, F. *et al.* Investigating the time dynamics of wind speed in complex terrains by using the Fisher–Shannon method. **Physica A: Statistical Mechanics and its Applications**, [s. l.], v. 523, p. 611–621, 2019b.

GWEC. **Global Wind Energy Council**. [S. l.: s. n.], 2022. Disponível em: <https://gwec.net/global-wind-report-2022>. Acesso em: 19 mar. 2022.

GWEC. **Global Wind Energy Council**. [S. l.: s. n.], 2024. Disponível em: https://gwec.net/wp-content/uploads/2024/05/GWR-2024_digital-version_final-2.pdf. Acesso em: 28 jul. 2024.

HAIR, J. F. *et al.* **Multivariate Data Analysis**. [S. l.: s. n.], 2019. Disponível em: www.cengage.com/highered.

HERSBACH, H. *et al.* **ERA5 hourly data on single levels from 1940 to present**. [S. l.], 2021. Disponível em: <https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels?tab=overview>. Acesso em: 18 nov. 2021.

HERSBACH, H. *et al.* The ERA5 global reanalysis. **Quarterly Journal of the Royal Meteorological Society**, [s. l.], v. 146, n. 730, p. 1999–2049, 2020.

- HU, J.; LIU, Y.; SANG, Y.-F. Precipitation Complexity and its Spatial Difference in the Taihu Lake Basin, China. **Entropy**, [s. l.], v. 21, n. 1, p. 48, 2019.
- IBGE. **Instituto Brasileiro de Geografia e Estatística**. [S. l.], 2020. Disponível em: <https://cidades.ibge.gov.br/brasil/pe/panorama>. Acesso em: 17 mar. 2022.
- IRENA. **International Renewable Energy Agency**. [S. l.], 2020. Disponível em: <https://www.irena.org/wind>. Acesso em: 15 jan. 2020.
- JONG, P. *et al.* Forecasting high proportions of wind energy supplying the Brazilian Northeast electricity grid. **Applied Energy**, [s. l.], v. 195, p. 538–555, 2017.
- JONG, P. *et al.* Solar and wind energy production in relation to the electricity load curve and hydroelectricity in the northeast region of Brazil. **Renewable and Sustainable Energy Reviews**, [s. l.], v. 23, p. 526–535, 2013.
- KANTELHARDT, J. W. *et al.* Long-term persistence and multifractality of precipitation and river runoff records. **Journal of Geophysical Research**, [s. l.], v. 111, p. 1–13, 2006.
- KIM, G.; HUR, J. Probabilistic modeling of wind energy potential for power grid expansion planning. **Energy**, [s. l.], v. 230, 2021.
- KRUYT, B.; LEHNING, M.; KAHL, A. Potential contributions of wind power to a stable and highly renewable Swiss power supply. **Applied Energy**, [s. l.], v. 192, p. 1–11, 2017.
- KUMAR, V.; CHHABRA, J. K.; KUMAR, D. **Performance Evaluation of Distance Metrics in the Clustering Algorithms**. [S. l.: s. n.], 2014.
- LAIB, M. *et al.* Community detection analysis in wind speed-monitoring systems using mutual information-based complex network. **Chaos**, [s. l.], v. 29, n. 4, 2019.
- LAIB, M. *et al.* Multifractal analysis of the time series of daily means of wind speed in complex regions. **Chaos, Solitons and Fractals**, [s. l.], v. 109, p. 118–127, 2018.
- LI, K. *et al.* An agglomerative hierarchical clustering-based strategy using Shared Nearest Neighbours and multiple dissimilarity measures to identify typical daily electricity usage profiles of university library buildings. **Energy**, [s. l.], v. 174, p. 735–748, 2019.
- LI, H. *et al.* Multi-scale Entropy Analysis of Single-point Wind Speed in Outdoor near-surface Environments. *In:* , 2011. **International Conference on Electrical and Control Engineering (ICECE)**. [S. l.: s. n.], 2011. p. 4579–4582.
- LI, T.; REZAEIPANAH, A.; TAG EL DIN, E. M. An ensemble agglomerative hierarchical clustering algorithm based on clusters clustering technique and the novel similarity measurement. **Journal of King Saud University - Computer and Information Sciences**, [s. l.], v. 34, n. 6, p. 3828–3842, 2022.
- LIU, F. *et al.* On wind speed pattern and energy potential in China. **Applied Energy**, [s. l.], v. 236, p. 867–876, 2019.

- LIU, W.; SONG, Y.; BI, K. Exploring the patent collaboration network of China's wind energy industry: A study based on patent data from CNIPA. **Renewable and Sustainable Energy Reviews**, [s. l.], v. 144, 2021.
- LY, S.; CHARLES, C.; DEGRÉ, A. Geostatistical interpolation of daily rainfall at catchment scale: the use of several variogram models in the Ourthe and Ambleve catchments, Belgium. **Hydrology and Earth System Sciences**, [s. l.], v. 15, n. 7, p. 2259–2274, 2011.
- MAHMOOD, F. H.; RESEN, A. K.; KHAMEES, A. B. Wind characteristic analysis based on Weibull distribution of Al-Salman site, Iraq. **Energy Reports**, [s. l.], p. 79–87, 2020.
- MANN, H. B.; WHITNEY, D. R. On a test of whether one of two random variables is stochastically larger than the other. **The Annals of Mathematical Statistics**, [s. l.], 1947. p. 50–60.
- MODIRI, E.; BÁRDOSSY, A. Clustering Simultaneous Occurrences of the Extreme Floods in the Neckar Catchment. **Water**, [s. l.], v. 13, n. 4, p. 399, 2021.
- MOLINA, M. O.; GUTIÉRREZ, C.; SÁNCHEZ, E. Comparison of ERA5 surface wind speed climatologies over Europe with observations from the HadISD dataset. **International Journal of Climatology**, [s. l.], v. 41, n. 10, p. 4864–4878, 2021.
- OUABO, R. E.; SANGODOYIN, A. Y.; OGUNDIRAN, M. B. Assessment of Ordinary Kriging and Inverse Distance Weighting Methods for Modeling Chromium and Cadmium Soil Pollution in E-Waste Sites in Douala, Cameroon. **Journal of Health and Pollution**, [s. l.], v. 10, n. 26, 2020.
- PAMPUCH, L. A. *et al.* A Review on Clustering Methods for Climatology Analysis and Its Application over South America. **International Journal of Geosciences**, [s. l.], v. 14, n. 09, p. 877–894, 2023.
- PEBESMA, E. **The meuse data set: a brief tutorial for the gstat r package**. [S. l.], 2015. Disponível em: <https://cran.r-project.org/web/packages/gstat/vignettes/gstat.pdf>. Acesso em: 24 fev. 2022.
- PEBESMA, E. J.; BIVAND, R. S. **Classes and Methods for Spatial Data: the sp Package**. [S. l.], 2005. Disponível em: https://cran.r-project.org/web/packages/sp/vignettes/intro_sp.pdf. Acesso em: 24 fev. 2022.
- PIMENTA, F. M. *et al.* Brazil Offshore Wind Resources and Atmospheric Surface Layer Stability. **Energies**, [s. l.], v. 12, n. 21, p. 4195, 2019.
- PINCUS, S. M. Approximate entropy as a measure of system complexity. *In:* , 1991. **Proceedings of the National Academy of Sciences USA**. [S. l.: s. n.], 1991. p. 2297–2301. Disponível em: <https://www.pnas.org>.
- R CORE TEAM. **R: A language and environment for statistical computing**. Vienna: [s. n.], 2021. Disponível em: <https://www.R-project.org/>.

RATTO, G.; BERRI, G. J.; MARONNA, R. On the application of hierarchical cluster analysis for synthesizing low-level wind fields obtained with a mesoscale boundary layer model. **Meteorological Applications**, [s. l.], v. 21, n. 3, p. 708–716, 2014.

RICHMAN, J. S.; MOORMAN, J. R. Physiological time-series analysis using approximate entropy and sample entropy. **American Journal of Physiology-Heart and Circulatory Physiology**, [s. l.], v. 278, n. 6, p. H2039–H2049, 2000.

RIND, D. Complexity and Climate. **Science**, [s. l.], v. 284, p. 105–107, 1999. Disponível em: www.sciencemag.org.

ROBERT, S.; FORESTI, L.; KANEVSKI, M. Spatial prediction of monthly wind speeds in complex terrain with adaptive general regression neural networks. **International Journal of Climatology**, [s. l.], v. 33, n. 7, p. 1793–1804, 2012.

ROCHA, P. A. C. *et al.* Comparison of seven numerical methods for determining Weibull parameters for wind energy generation in the northeast region of Brazil. **Applied Energy**, [s. l.], v. 89, n. 1, p. 395–400, 2012.

SANTOS, J. V. C. *et al.* Analysis of long-range correlations of wind speed in different regions of Bahia and the Abrolhos Archipelago, Brazil. **Energy**, [s. l.], v. 167, p. 680–687, 2019.

SANTOS, C. A. G. *et al.* Cluster Analysis Applied to Spatiotemporal Variability of Monthly Precipitation over Paraíba State Using Tropical Rainfall Measuring Mission (TRMM) Data. **Remote Sensing**, [s. l.], v. 11, n. 6, p. 637, 2019.

SANTOS, F. S. *et al.* Mixture distribution and multifractal analysis applied to wind speed in the Brazilian Northeast region. **Chaos, Solitons and Fractals**, [s. l.], v. 144, 2021.

SHEPARD, D. A two-dimensional interpolation function for irregularly-spaced data. *In:* , 1968, New York, New York, USA. **Proceedings of the 1968 23rd ACM National Conference**. New York, New York, USA: ACM Press, 1968. p. 517–524.

SHUANGCHENG, L. *et al.* Measurement of climate complexity using sample entropy. **International Journal of Climatology**, [s. l.], v. 26, n. 15, p. 2131–2139, 2006.

SIRAJ-UD-DOULAH, M.; ISLAM, M. N. Defining Homogenous Climate zones of Bangladesh using Cluster Analysis. **International Journal of Statistics and Mathematics**, [s. l.], v. 6, p. 119–129, 2019.

SIVHUGWANA, K. S.; RANGANAI, E. Short-Term Wind Speed Prediction via Sample Entropy: A Hybridisation Approach against Gradient Disappearance and Explosion. **Computation**, [s. l.], v. 12, n. 8, p. 163, 2024.

SOKAL, R. R.; ROHLF, F. J. The Comparison of Dendrograms by Objective Methods. **Taxon**, [s. l.], v. 11, n. 2, p. 33–40, 1962.

STRAUSS, T.; VON MALTITZ, M. J. Generalising Ward's Method for Use with Manhattan Distances. **PLoS ONE**, [s. l.], v. 12, n. 1, p. e0168288, 2017.

SUMAIR, M. *et al.* Wind potential estimation and proposed energy production in Southern Punjab using Weibull probability density function and surface measured data. **Energy Exploration and Exploitation**, [s. l.], v. 39, n. 6, p. 2150–2168, 2021.

TELESCA, L.; LOVALLO, M.; KANEVSKI, M. Power spectrum and multifractal detrended fluctuation analysis of high-frequency wind measurements in mountainous regions. **Applied Energy**, [s. l.], v. 162, p. 1052–1061, 2016.

TOMCALA, J. **TSEntropies: Time Series Entropies**. [S. l.], 2018. Disponível em: <https://CRAN.R-project.org/package=TSEntropies>. Acesso em: 19 fev. 2022.

WANG, Y. *et al.* Study on Wind Profile Characteristics Using Cluster Analysis. **Atmosphere**, [s. l.], v. 15, n. 6, p. 708, 2024.

WARD, J. H. Hierarchical Grouping to Optimize an Objective Function. **Journal of the American Statistical Association**, [s. l.], v. 58, n. 301, p. 236–244, 1963.

WU, S. De *et al.* Modified multiscale entropy for short-term time series analysis. **Physica A: Statistical Mechanics and its Applications**, [s. l.], v. 392, n. 23, p. 5865–5873, 2013.

XAVIER, S. F. A. *et al.* Multiscale entropy analysis of monthly rainfall time series in Paraíba, Brazil. **Chaos, Solitons and Fractals**, [s. l.], v. 151, 2021.

XU, D.; TIAN, Y. A. A Comprehensive Survey of Clustering Algorithms. **Annals of Data Science**, [s. l.], v. 2, n. 2, p. 165–193, 2015.

YAN, B. *et al.* Characterising the fractal dimension of wind speed time series under different terrain conditions. **Journal of Wind Engineering and Industrial Aerodynamics**, [s. l.], v. 201, 2020.

YIN, J. *et al.* Complexity Analysis of Carbon Market Using the Modified Multi-Scale Entropy. **Entropy**, [s. l.], v. 20, n. 6, p. 434, 2018.

ZELTERMAN, D. **Applied Multivariate Statistics with R**. New York: Springer International Publishing, 2015.

ZHOU, X.; LEI, W. Spatial patterns of sample entropy based on daily precipitation time series in China and their implications for land surface hydrological interactions. **International Journal of Climatology**, [s. l.], v. 40, n. 3, p. 1669–1685, 2020.