



Avaliação de análise dos tipos de tempo através de técnicas de machine learning

Assessment of weather types analysis through machine learning techniques

Francisco Jablinski Castelhano^{*1} ; Dayvid Carlos Medeiros¹ ; Leonardo Lima de Oliveira²

¹Departamento de Geografia, Universidade Federal do Rio Grande do Norte, Natal, RN, Brasil.

Recebido (Received): 13/05/2024

²Instituto de Políticas Públicas, Universidade Federal do Rio Grande do Norte, Natal, RN, Brasil.

ACEITO (Accepted): 03/04/2025

E-mail: dayvid.medeiros.123@ufrn.edu.br (DCM); leonardo.oliveira.n21@gmail.com (LLO).

*Email para correspondência: francisco.castelhano@ufrn.br

Resumo: A presente pesquisa propõe uma ferramenta metodológica para Análise dos Tipos de Tempo que atuam em dada localidade com base em modelos de Machine Learning. Para este estudo, foram coletados dados de oito variáveis climáticas em escala diária da cidade de Natal, Rio Grande do Norte, nordeste brasileiro, para o ano de 2022 como dados de entrada. Foi realizada uma Análise de Componentes Principais junto ao *dataset* em questão, o qual, sumarizaram-se as oito variáveis de entrada em uma única. Na sequência, os valores desta única variável passaram por um processo de clusterização, originando agrupamentos que denotam os diferentes tipos de tempo que atuaram na localidade analisada. A título de comparação, foi realizada uma análise sinóptica utilizando a técnica da Análise Rítmica para o mesmo *dataset*. Os resultados do modelo apontaram a ocorrência de quatro tipos de tempo meteorológico em Natal no período analisado. Embora tenha apresentado incongruências com a Análise Rítmica, a proposta automatizada mostrou-se satisfatória dada a facilidade de aplicação e possibilidade de reprodução em séries longas. Em tempos de mudanças climáticas, grandes volumes de dados meteorológicos possibilitam novas formas de analisar o clima presente e interpretar suas interações com o espaço geográfico.

Palavras-chave: Componentes principais; Climatologia dinâmica; Big data; Análise rítmica.

Abstract: The present research proposes a methodological technique for the Synoptic Analysis of the Types of Weather in a given location based on Machine Learning models. For this study, climatic data of eight variables were collected on a daily scale for the city of Natal, Rio Grande do Norte, northeastern Brazil, for the year 2022 as input. A Principal Component Analysis was carried out with the organized dataset, which, the eight input variables were combined into one. After that, the values of this single variable underwent a clustering process, originating clusters that denote the different types of weather that acted in the analyzed location. By way of comparison, a synoptic analysis was performed using the Rhythmic Analysis technique for the same dataset. The model results indicated the occurrence of four types of meteorological weather in Natal during the analyzed period. Although it presented inconsistencies with Rhythmic Analysis, the automated proposal proved to be satisfactory given the ease of application and possibility of application in large volume datasets. In times of climate change, large volumes of meteorological data enable new ways of analyzing the present climate and interpreting its interactions with geographic space.

Keywords: Principal components; Dynamic climatology; Big data; Rhythmic analysis.

1. Introdução

Compreendemos a Climatologia Geográfica como uma área de estudo da Geografia, distinta da Meteorologia por promover análises de fenômenos atmosféricos associados a dimensão espacial. Até meados dos anos 1950, em um período em que a climatologia é chamada de “clássica”, seu foco residia em estudos voltados para a descrição dos distintos elementos que formam o clima (temperatura, umidade relativa do ar, ventos, pressão atmosférica etc.) de maneira separada e isolada. Neste período, as análises propostas não propõem relações entre os elementos climáticos, tampouco discutem sua relação com a circulação atmosférica, sendo por tal fato, muito criticada (ZAVATINNI, 1998; CUNHA; VECCHIA, 2007).

Uma concepção de climatologia, chamada de “dinâmica”, surge então como um novo paradigma nesta ciência, partindo para uma análise não mais isolada, mas sim conectada do clima e de seus elementos e suas repercussões na sociedade. Cunha e Vecchia (2007) apontam que, nesta visão, a climatologia estudará não apenas cada elemento de forma isolada, mas sim o conjunto climático na figura dos sistemas atmosféricos atuantes sobre as localidades, se baseando no movimento sucessivo das massas de ar para estabelecer as características dos tipos de tempo atuante nas localidades estudadas.

Fez-se necessário portanto, o desenvolvimento de técnicas que permitam a análise integrada dos elementos do clima como temperatura, umidade, precipitação, com os sistemas atmosféricos atuantes na respectiva localidade, convergindo no que passa a ser denominado de Análise Sinóptica, acha vista que seu foco serão os sistemas atmosféricos em escala sinótica.

Segundo Yarnal *et al.* (2001) as técnicas de análise sinótica podem ser classificadas utilizando como critério os tipos de processos realizados. Assim, os autores pontuam a existência de técnicas de análise sinóticas manuais, semi-automáticas ou híbridas e automáticas.

De acordo com Fontão *et al.* (2018), diversas foram as técnicas desenvolvidas para análise sinótica dos tipos de tempo, citando como exemplos a “Spatial Synoptic Classification” proposta por Sheridan (2002) e a Análise Rítmica proposta por Monteiro (1971) no âmbito do Brasil.

Ambas as técnicas citadas se utilizam de abordagens híbridas, no que tange seus processos, alternando em técnicas automatizadas e análises manuais. Em ambos os casos, trabalha-se com dados meteorológicos em escala diária visando uma classificação sinótica do sistema atmosférico atuante na localidade analisada. Na primeira, por exemplo, os dados meteorológicos são manualmente agrupados em tipos de tempo conforme suas especificações e observações do pesquisador. No segundo, realiza-se leituras diárias de cartas sinóticas em conjunto a análise dos elementos meteorológicos para a definição do tipo de tempo para cada dia (FONTÃO *et al.*, 2018).

Reforça-se a proposta da Análise Rítmica de Monteiro (1971) como a técnica de Análise Sinóptica dos tipos de tempo atuantes mais utilizadas pela ciência climatológica brasileira e precursora de uma escola própria de climatologia dinâmica no país (SANT’ANNA NETO, 2014).

No âmbito das técnicas automáticas, Wanderley e Nobrega (2022) propõe a utilização de classificações sinóticas a partir da técnica de Análise de Componentes Principais [PCA], para classificar os tipos de tempo atuantes na região Nordeste do Brasil. Segundo os autores, o PCA seria uma dentre várias técnicas de “Machine Learning” que podem ser aplicadas para classificar os sistemas atmosféricos atuantes em uma determinada localidade, apontando como sua principal vantagem a facilidade na realização da análise e a possibilidade de aplicação de tal técnica em grandes volumes de dados, fato este que torna-se difícil quando dá aplicação de uma técnica híbrida como a Análise Rítmica mencionada anteriormente.

Trabalhos recentes têm apontado para o uso de técnicas de “Machine Learning” como “Random Forest”, Arvores de Decisão e Redes neurais como novas possibilidades na classificação e “forecasting” de eventos climáticos, jogando luz a novas possibilidades de análises sinóticas automatizadas (ZHOU *et al.*, 2019; MCGOVERN *et al.*, 2010, ZHANG *et al.*, 2022).

Pretende-se, portanto, neste trabalho, comparar métodos alternativos de análise sinótica dos tipos de tempo, realizando uma análise híbrida em comparação a aplicação de técnicas automáticas. Tal trabalho permitirá que pesquisadores possam definir, com base em estatística, quais as melhores técnicas para realizar uma análise sinótica dos tipos de tempo.

Desta forma serão aplicadas duas técnicas de análise sinótica dos tipos de tempo para um mesmo banco de dados climáticos, sendo uma automatizada (Análise de Componentes Principais) e uma técnica híbrida (Análise Rítmica dos Tipos de Tempo). O objetivo geral, portanto, será comparar os resultados das análises automáticas com a híbrida, buscando apresentar qual das técnicas automáticas empregadas aproxima-se mais da híbrida.

2. Material e métodos

A metodologia utilizada constitui-se de Quatro etapas a serem descritas a seguir: 1) Coleta e análise exploratória de dados; 2) Construção e Aplicação do modelo automatizado de Análise Sinóptica; 3) Construção e Aplicação do modelo híbrido de Análise Sinóptica e 4) Análise comparativa dos resultados obtidos.

2.1. Coleta e análise exploratória dos dados

Para a construção e comparação de diferentes modelos de análise sinóptica faz-se necessária a utilização de uma base de dados meteorológicos robusta e única, a ser aplicada para ambas as técnicas e que permita que as conclusões propostas tenham elevado grau de significância. Igualmente, faz-se necessário que cenários climáticos diferentes sejam analisados, o que novamente permitirá a construção de conclusões com mais relevância estatística.

Deste modo, foi utilizado como entrada para os modelos, dados meteorológicos das cidades de Natal, região nordeste do mesmo país. Os modelos foram gerados com os dados referentes ao ano de 2022, em escala diária, tendo sido coletado junto a página do Banco de Dados Meteorológicos do Instituto Nacional de Meteorologia.

Para a realização das análises sinópticas foram utilizados apenas dias com registros completos das seguintes variáveis: Precipitação Diária Acumulada (mm), Insolação (horas), Temperatura Máxima (°C), Temperatura Mínima (°C), Temperatura Média (°C), Umidade Relativa do Ar (%), Velocidade do Vento (m/s) e Direção Predominante do Vento. Os dias cujas variáveis encontravam-se indisponíveis foram descartados da análise, assim, foi utilizado ao total um banco de dados com oito variáveis e 334 observações.

Além destes dados, foram utilizadas cartas sinópticas fornecidas pela marinha para a realização e aplicação da Análise Rítmica dos tipos de tempo segundo a metodologia de Monteiro (1973).

Natal situa-se no regime climático tropical litorâneo, com duas estações de chuva e seca bem definidas e pouca variação térmica entre seus meses (MENDONÇA; DANNI-OLIVEIRA, 2007). Segundo Novais (2023), Natal encontra-se na chamada zona térrida, no domínio equatorial, subdomínio do litoral Leste do Brasil, especificando-o como Equatorial semiseco, litorâneo leste do Brasil. Segundo o autor, esse clima caracteriza-se por temperaturas médias entre 22,5 e 25,1°C, e precipitações acumuladas em 6 meses, variando entre 683 e 1461 mm anuais. Dubreuil *et al.* (2019) classifica o clima de Natal como As segundo a classificação de Koppen, portanto como tropical de savana, tendo como principal característica as estações seca e chuvosas bem definidas. Além disso, os autores apontam que, diferente de outras partes do país, ao longo dos últimos 50 anos, o clima da área de estudo manteve-se dentro da mesma classificação, sem apresentar modificações relevantes.

2.2. Construção e aplicação do modelo automatizado de análise dos tipos de tempo

Para a construção do modelo automatizado de análise dos tipos de tempo optou-se pelo uso de duas técnicas de Machine Learning, em um primeiro momento pela técnica da Análise dos Componentes Principais aplicada ao banco de dados meteorológicos seguido pela técnica de agrupamentos não-hierárquicos “Kmeans” em que se buscou a quantidade de agrupamentos em que a soma dos quadrados intra-clusters (ou do inglês “within-clusters sum-of-squares”), seja a menor possível.

A PCA consiste em uma técnica não supervisionada de “machine learning” que permite a combinação de diversas variáveis de um determinado banco de dados em novas variáveis denominadas Componentes Principais e com dimensões iguais as das variáveis originais. Tal técnica permite a redução de um banco de dados com múltiplas variáveis sem que ocorra a perda de informações, agrupando as variáveis conforme sua variância. De forma simplificada, é possível agrupar variáveis em um novo fator principal de modo a simplificar o banco de dados sem que informações originais sejam perdidas (VARELLA, 2008).

Levando em conta que a PCA agrupa as variáveis a partir de semelhanças na variância, é fundamental que exista um mínimo de correlação entre as variáveis originais que permita que as mesmas possam vir a ser agrupadas. Assim, as primeiras etapas para realização da Análise de Componentes Principais, e definição de ditos componentes, iniciam-se pela criação de uma matriz de correlação envolvendo as variáveis originais (**Figura 1**), seguido pela aplicação do teste de esfericidade de Bartlet que compara as matrizes de correlação com uma matriz de identidade, atestando a possibilidade da criação de componentes principais mediante a constatação da existência de correlações entre as variáveis.

Uma vez que o teste de Bartlet atesta a possibilidade de criação de componentes principais, passou-se a Análise Fatorial, para em um primeiro momento, calcular-se a variância compartilhada em cada fator (autovalores) e em seguida calcular-se os “scores” fatoriais que são os parâmetros que relacionam os fatores gerados com as variáveis originais dos bancos de dados. Após a criação dos fatores, aqueles que foram utilizados na análise foram selecionados seguindo o critério de Kaiser, portanto utilizando apenas os fatores cujos autovalores superam o valor de 1.

Em um terceiro momento, foram calculadas as cargas fatoriais dos fatores utilizados, de modo que foi possível observar a correlação entre as variáveis originais e os fatores principais criados, facilitando a interpretação da importância de cada variável para cada novo fator.

Por fim, os fatores selecionados passaram por um processo de ranqueamento, utilizando o critério da soma ponderada e ordenamento, para cada observação da amostra, em síntese deve-se multiplicar o resultado obtido de cada fator por seu percentual de variância compartilhada e depois é realizado o ordenamento deste resultado. Este ranking final situa-se como uma síntese dos fatores principais criados e com maiores autovalores, de modo que, chega-se a um único valor final que possa sintetizar as variáveis contidas no banco de dados original.

Uma vez que o “ranking” tenha sido calculado, aplicou-se um cálculo de agrupamentos pelo método não-hierárquico “Kmeans”, possibilitando o agrupamento do valor previamente ranqueado em novos grupos. A definição do número de agrupamentos se deu pela técnica da mínima soma dos quadrados das distâncias intracusters.

O método de agrupamento não-hierárquico “Kmeans” agrupa observações que apresentem certo grau de homogeneidade. Para isso, após a determinação do número de clusters, determina-se um valor central para cada agrupamento. O algoritmo “Kmeans” processa com o pensamento de que os clusters devem ser formados de acordo com a distância entre pontos e centro dos clusters (GULAGIZ; SAHIN, 2017).

Assim, cada cluster irá agrupar os valores da variável resultante da aplicação da PCA de acordo com sua proximidade. Considerando a variável síntese da PCA como uma síntese das variáveis climáticas imputadas inicialmente no modelo (Temperatura máxima, média e mínima, insolação, umidade relativa do ar, precipitação) assumiu-se que cada “cluster” calculado para os valores de rankings obtidos a partir da análise dos componentes principais representaria um tipo de tempo diferente atuando sobre a cidade de Natal.

O uso da PCA para análise climática não é novidade, já tendo sido aplicado no Japão (KOJIMA, 1973), nos Estados Unidos (DENTON; BARNES, 1988), Canadá (NICHOLSON; BRYANT, 1988) e Gana (2022) por exemplo.

No Brasil, análises semelhantes da aplicação da técnica de agrupamento não-hierárquico “Kmeans” podem ser vistas em Machado (2014), e aplicadas a classificação climática de Minas Gerais, onde a autora estabelece quatro agrupamentos com base nas variações dos elementos climáticos, e assume cada agrupamento como um tipo climático observado no estado.

Wanderley e Nobrega (2022) aplicam técnica semelhante na classificação climática da região nordeste, aplicando o método de componentes principais seguido por um método de agrupamento. Diferentemente do proposto neste artigo, os artigos optaram pelo método de agrupamento hierárquico de Ward. Igualmente, os agrupamentos encontrados foram interpretados como tipos climáticos individuais,

A presente proposta, todavia, objetiva agrupar o resultado de uma PCA para uma única estação, associando-a aos tipos de tempo que atuam na localidade em questão, dando-lhe assim um caráter inovador.

A construção destas etapas sucedeu-se através do software R e da utilização dos pacotes *psych* e *factoextra*.

2.3. Construção e aplicação do modelo híbrido de análise sinóptica

A terceira etapa consistiu na aplicação da análise rítmica dos tipos de tempo sobre o banco de dados originais. A análise rítmica situa-se como uma técnica de análise sinóptica proposta por Monteiro (1976), situando-se como uma ferramenta chave na compreensão do clima enquanto um fenômeno dinâmico. A climatologia dinâmica não separa os elementos, tendo como princípio os tipos de tempo, com cada tipo sendo analisado em seus elementos constitutivos.

Desta forma, a abordagem dinâmica propõe considerar os elementos climáticos de forma integrada, desconsiderando o tratamento desintegrado na busca da compreensão da realidade climática (CUNHA; VECCHIA, 2007).

Baseado em Pédelaborde e Sorre, Monteiro (1976) propôs a abordagem da atmosfera a partir da análise do ritmo dos tipos de tempos, ou sucessão dos estados atmosféricos, sobre um determinado lugar. Assim, o ritmo é a estratégia espacial e temporal que Monteiro propõe para entender o clima no cotidiano da sociedade.

A proposta de Monteiro permite visualizar a variação dos elementos climáticos e acompanhar a sua génesis, e sua evolução de modo a individualizar na superfície os impactos causados.

Para o conhecimento do ritmo, deve-se utilizar dados climáticos diários em sequência e cartas sinópticas e imagens de satélites, em uma análise associativa e qualitativa.

O ritmo retrata os encadeamentos dos estados atmosféricos sobre um lugar, sendo os mais frequentes considerados representativos do ritmo habitual, de modo que suas variações e desvios, atinjam padrões extremos.

A proposta de Monteiro (1976) segue três princípios: O ritmo climático só poderá ser compreendido através da representação concomitante dos elementos fundamentais do clima em unidades de tempo pelo menos diárias, compatíveis com a representação da circulação atmosférica regional, geradora dos estados atmosféricos que se sucedem e constituem o fundamento do ritmo; Somente a análise rítmica detalhada a nível de tempo a revelando a gênese dos fenómenos climáticos pela interação dos elementos e fatores, dentro de uma realidade parâmetros regional, é capaz de oferecer válidos à consideração dos diferentes e variados problemas geográficos desta região; as expressões quantitativas dos elementos climáticos estão indissoluvelmente ligadas. A qualidade dos mesmos, à gênese ou e os parâmetros resultantes desta análise devem ser considerados levando em conta a posição no espaço geográfico em que se definem.

Na prática, a técnica consiste na representação simultânea dos elementos climáticos básicos e dos mecanismos da circulação secundária, com o objetivo de definir os tipos de tempo - Na montagem do gráfico, os elementos (temperatura diária, máxima e mínima, umidade relativa, pressão atmosférica, radiação solar, precipitações, insolação, nebulosidade, velocidade e direção do vento) são dispostos simultaneamente e sincronamente, o que permite a comparação, de forma integrada, da variação diária/horária dos mesmos.

A realização e aplicação desta técnica deu-se utilizando o software *excel* a partir de onde os elementos climáticos foram organizados, e com o auxílio das cartas sinópticas, os tipos de tempo atuantes em cada dia foi definido.

2.4. Análise comparativa

Uma vez que ambos os modelos de análise sinóptica tenham sido realizados, passou-se a etapa da comparação de resultados. Admitiu-se que cada “cluster” do rankeamento resultado da Análise de Componentes Principais situava-se como um tipo de tempo diferente. Igualmente, os tipos de tempo diferentes observados na análise rítmica foram comparados com os “clusters” gerados, de modo a validar os dados da análise sinóptica automática (PCA) a partir dos resultados da análise sinóptica híbrida (Análise Rítmica).

3. Resultados e discussões

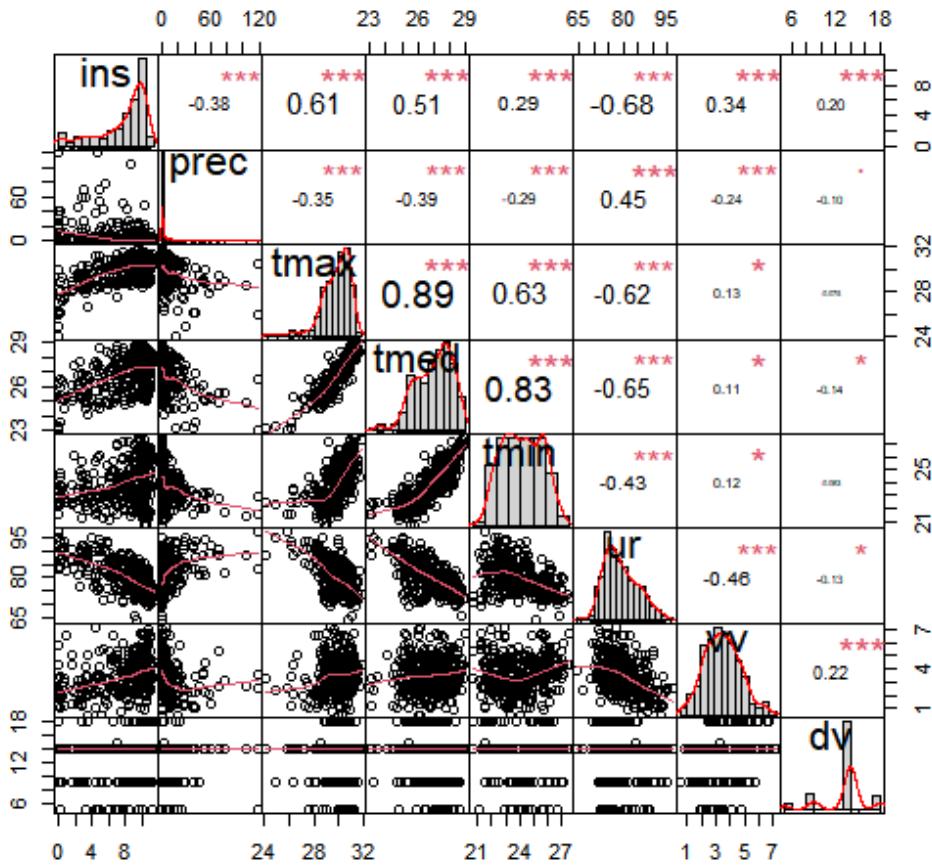
3.1. Análise automatizada dos tipos de tempo

Conforme mencionado anteriormente, a primeira etapa para realização da análise sinóptica automatizada se deu pela construção de uma matriz de correlações envolvendo as variáveis originais do banco de dados utilizado no estudo. A mesma pode ser visualizada no “scatterplot” apresentado junto a **Figura 1**, reforçando as relações positivas e fortes entre as variáveis térmicas (insolação, temperatura máxima, média e mínima) além de relação negativas e moderadas entre umidade relativa do ar e as variáveis térmicas.

Uma vez construída a matriz de correlações, realizou-se o teste de esferecideade de Bartlet, cujos resultados são apresentados na **Tabela 1**. O *p*-valor inferior a 0,05 levou-nos a rejeitar a hipótese zero que apontaria ausência de correlação entre as variáveis o que impediria o prosseguimento da Análise de Componentes Principais. Neste caso, pelo *p*-valor totalizando zero, a hipótese H1 foi aceita, admitindo a possibilidade de realizarmos a PCA.

Após o teste de esferecideade de Bartlet, partiu-se para a elaboração dos Fatores Principais. Nesta primeira etapa foram criados um número de fatores respectivos ao número de variáveis originais, oito. A **Tabela 2** apresenta os autovalores, a variância compartilhada e a variância acumulada de cada fator principal [PC] gerado nesta etapa.

O primeiro fator gerado apresentou a maior variância compartilhada, totalizando 48,1% de variância em relação as variáveis originárias. O segundo fator por sua vez, apresentou 18,6% de modo que, com os dois primeiros fatores, chegou-se a um valor total de 66,7% de variância acumulada.



*Nota: ins = Insolação (horas); prec = Precipitação (mm); tmax = Temperatura Máxima (°C); tmed = Temperatura Média (°C); tmin = Temperatura Mínima (°C); ur = Umidade Relativa do Ar (%); vv = Velocidade do Vento (m/s); dv = Direção Predominante do Vento (graus).

Figura 1: Scatterplot das Correlações de Pearson envolvendo as variáveis originais. Fonte: Resultados originais da pesquisa.

Tabela 1: Resultado do Teste de Esferecideade de Bartlet;

Qui-Quadrado	P-valor	Graus de Liberdade
1870,073	0	28

Fonte: Resultados originais da pesquisa

Tabela 2: Fatores Principais, Autovalores e Variâncias.

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
Auto valor	3,852	1,488	0,755	0,712	0,630	0,278	0,239	0,043
Prop. Da Variância	0,481	0,186	0,094	0,089	0,078	0,034	0,029	0,005
Variância Acumulada	0,481	0,667	0,762	0,851	0,929	0,964	0,994	1

Fonte: Resultados originais da pesquisa

A **Tabela 3** apresenta os resultados dos “scores” fatoriais de cada variável, demonstrando como cada variável original irá contribuir para o fator principal gerado estatisticamente. Percebe-se que no primeiro fator, as variáveis Temperatura Máxima, Média, Mínima e Insolação foram aquelas com maior contribuição, enquanto no segundo fator, as variáveis Insolação, Velocidade do Vento e Direção do Vento apresentaram contribuição maior.

Visando interpretar os tipos de tempo de Natal com menos variáveis, optou-se pelo critério de Kaiser para selecionar um número menor de fatores principais que possam representar o “dataset” em questão. Por este critério, são mantidos, dentre os oito fatores criados, apenas aqueles com autovalores superiores a 1. Neste caso, apenas os fatores 1 e 2 foram mantidos por apresentarem respectivamente autovalores de 3.852 e 1.488.

Tabela 3: Scores Fatoriais das Variáveis Originais.

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
Insolação	0,194	0,209	0,090	0,174	-0,733	0,888	-0,913	0,318
Precipitação	-0,149	-0,144	0,031	1,107	-0,071	-0,138	-0,074	0,097
Temp. Máx.	0,226	-0,174	0,148	0,159	-0,217	0,348	1,384	-2,159
Temp. Méd.	0,234	-0,248	0,106	0,087	0,144	-0,140	0,311	3,858
Temp. Min.	0,190	-0,262	0,145	0,104	0,754	0,268	-1,011	-1,613
Umid. Rel.	-0,218	-0,166	0,171	-0,118	0,226	1,535	0,316	0,867
Vel. Vento	0,100	0,422	-0,728	0,266	0,485	0,458	0,379	0,320
Dir. Vento	0,014	0,491	0,837	0,138	0,347	-0,097	0,257	0,234

Fonte: Resultados originais da pesquisa

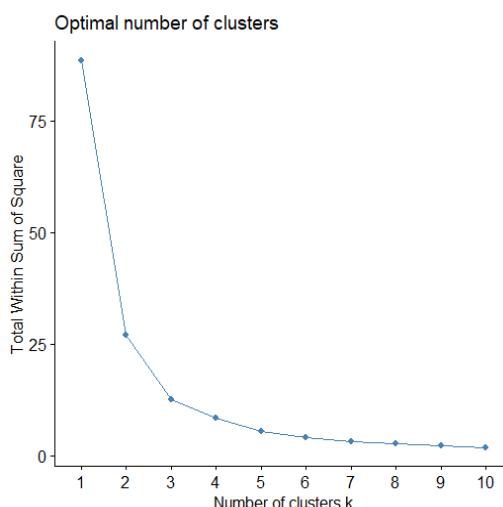
Neste momento, temos então dois fatores principais que representam 66,7% da variância do “dataset” completo. Ambos estes fatores passaram por um processo de rankeamento, de modo que, seguiu-se o critério de soma ponderada e ordenamento dos dois fatores principais mantidos, resultando na criação de um único valor que pudesse representar ambos. Neste caso, multiplicou-se o resultado de cada fator pelo seu respectivo percentual de variância compartilhada seguido pela soma destes valores.

Este processo resulta na criação de um único componente principal, que, no banco de dados utilizado, variou entre -2,084 e 0,918.

Admite-se, portanto, que os valores deste *ranking* criado a partir dos fatores principais com maior variância compartilhada, indicam de forma genérica a variação dos elementos meteorológicos de Natal de forma conjunta. Assim, passou-se a um processo de *clusterização* desta variável de forma a agrupar as observações com valores mais semelhantes. Este processo foi possível, mediante aplicação do método não-hierárquico “Kmeans” e pela definição de um total de quatro “clusters” mediante o cálculo da mínima soma dos quadrados das distâncias intraclusters cuja representação gráfica pode ser visualizada na **Figura 2**. Nesta figura, destaca-se que o número ideal de clusters seria aquele a partir do qual a soma dos quadrados intraclusters passa a apresentar estabilidade, isto é, após o fim de sua curvatura na disposição gráfica.

Assim, definiu-se um total de quatro tipos de tempo atuando em Natal durante o ano de 2022. O tipo 1 ocorreu em 37,2% do tempo observado, o tipo 2 foi registrado em 48,3%, o tipo 3 ocorre em 13,5% e por fim o tipo 4 registrado em apenas 0,9% das observações.

A **Tabela 4** revela as características dos elementos climáticos dos “clusters” formados. Em relação ao “cluster 1”, percebe-se uma oscilação nos valores de Insolação entre 6,2 e 11,5 horas diárias. A precipitação apresentou pouca variação neste tipo climático, com um registro máximo de 15,6mm em 24 horas, mas predomínio de dias sem chuva. A temperatura máxima mostrou-se elevada, oscilando entre 28,9 e 31,90°C com média de 30,5°C. A temperatura média apresentou oscilação semelhante, variando entre 25,1°C e 29,0°C com média de 27,6°C. A temperatura mínima variou de 21,2 °C até 27,5°C com médias registradas em 25,2°C. Em relação a umidade relativa do ar, o menor valor encontrado para este tipo climático foi 65,7%, com um valor médio de 75% e máximo de 84,7%. As variáveis referentes ao vento apontam velocidade do vento média de 4,4 m/s e direção do vento predominante sentido Sudeste e Sul.

**Figura 2:** Soma dos Quadrados Intraclusters x Número de Clusters. Fonte: Resultados originais da pesquisa.

O “cluster” 2, por sua vez, apresenta semelhanças com o “cluster” 1 no âmbito das temperaturas, insolação, e umidade do ar, todavia, apresenta diferenças no predomínio da direção dos ventos (SE), velocidade dos ventos com médias menores, e sobretudo, maiores valores de precipitação. Dos 124 dias classificados como tipo 1, em apenas 21 deles foram registrados valores de chuva, os quais, em apenas um dia, o valor excede 10 milímetros, registrando 15.6mm em 24 horas. Já no tipo 2, dos 161 dias assim classificados, em 83 foram registrados valores de chuva, portanto em mais da metade. Destes, em 21 dias os valores superam 10 mm.

Tabela 4: Scores Fatoriais das Variáveis Originais.

	Cluster 1			Cluster 2			Cluster 3			Cluster 4		
	Min	Méd	Max	Min	Méd	Max	Min	Méd	Max	Min	Méd	Max
Prec. (mm)	0	0,4	15,6	0	4,6	52,2	0	24,9	119,1	29,0	62,4	119,0
Umidade Relativa (%)	65,7	75	84,7	71	80,3	89,7	81,2	88,5	95	91,7	94,6	97
Temp. Máx (°C)	28,9	30,5	31,9	28,1	29,9	31,7	24,1	28,4	30,5	24,9	25,6	26,1
Temp. Méd (°C)	25,1	27,6	28,9	24,5	23,5	28,6	23,1	25,5	27,0	22,9	23,1	23,3
Temp. Min. (°C)	21,2	25,3	27,5	20,7	23,5	27,1	21,1	23,0	25,6	21,5	21,8	22,3
Insolação (horas)	6,2	10,0	11,5	0	8,3	11,2	0	3,1	8,6	0	0	0
Vel.Vento (m/s)	1,5	4,3	7,0	0,8	3,3	7,2	0,5	2,4	5,4	1,2	2,1	2,6
Dir. Vento	-	Se/S	-	Se	-	-	Se/E	-	-	Se	-	-

Fonte: Resultados originais da pesquisa

No “cluster” 3 percebem-se quedas significativas nos valores de temperatura, além de diminuições nos valores de insolação. A velocidade do vento igualmente apresentou quedas, e a umidade relativa do ar apresentou significativo incremento. Todavia, novamente a variável com maior diferença em relação as demais foram as chuvas. Dos 45 dias classificados como tipo 3, apenas em 4 não foram constatados valores de chuva, com 27 dias de precipitação acima de 10 mm, e o valor máximo registrado na série, 119,1 mm em 24 horas.

O tipo climático com menos registros na série analisada, o tipo 4, contando com apenas 3 dias assim classificados, apresentou os menores valores de temperatura e de velocidade dos ventos. A umidade relativa do ar mostrou-se como a mais elevada da série e a insolação não foi registrada, apontando para alta nebulosidade nestes dias. Igualmente, observou-se a ocorrência de precipitação em todos os dias, com o menor valor em 29 mm em 24 horas, e o maior registrando 119mm no mesmo período.

A concentração temporal também é um fator a ser levado em conta para analisar as características dos quatro tipos de tempo levantados pela técnica da Análise Síntética Automatizada. A **Figura 3** revela os padrões de dispersão ao longo do ano analisado.

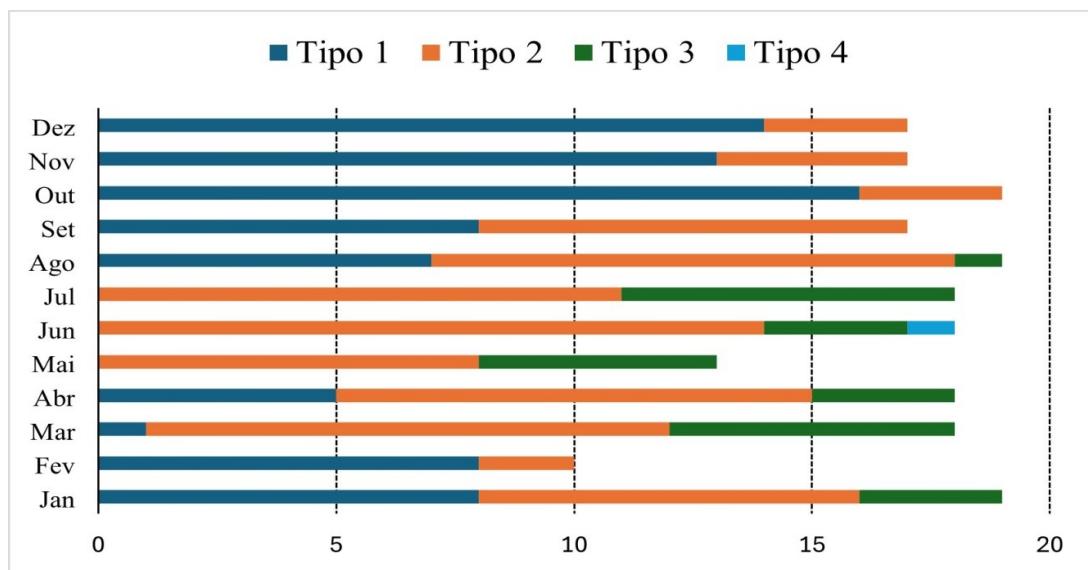


Figura 3: Distribuição Temporal dos Tipos Climáticos da Análise Automatizada. Fonte: Resultados originais da pesquisa.

Percebe-se uma maior concentração dos tipos 3 e 4 nos meses típicos de inverno, entre o final de Maio e início de Agosto. O tipo 4 em específico ocorreu apenas entre os meses de Junho e Julho.

O tipo 2, por outro lado, mostrou-se com ocorrência mais concentrada entre os meses de Março e Setembro, ocorrendo porém, ao longo de todo ano. Por fim, o tipo 1, com temperaturas mais elevadas, é registrado entre Janeiro e Abril, e entre Agosto e Dezembro, ocorrendo com mais frequência nos meses de verão.

3.2. Análise rítmica dos tipos de tempo

A primeira etapa para aplicação da Análise Rítmica dos tipos de tempo sobre o banco de dados situou-se na organização gráfica dos dados de elementos climáticos para cada mês. Tal fato permitiu a análise conjuntural da variação diária dos elementos utilizados. Em um segundo momento, em conjunto aos dados de cada dia, foram analisadas cartas sinópticas disponibilizadas pela marinha e imagens do satélite GOES-16 na banda do infravermelho que permitiram, de forma qualitativa, a definição do sistema atmosférico atuante em cada dia da série temporal do local de estudo.

A **Figura 4** revela um exemplo de gráfico de Análise Rítmica aplicado para o mês de janeiro de 2022. O gráfico em questão apresenta-nos concomitantemente a variação de temperatura máxima, média e mínima, Umidade relativa do ar, precipitação, direção predominante do vento e velocidade média do vento para cada dia mês. Ao final, é apresentado o sistema atmosférico ou tipo de tempo para cada dia. No caso de janeiro, apresentado como exemplo, observa-se um predomínio do sistema atmosférico referente a Massa Tropical Atlântica, sistema este originado no Anticiclone do Atlântico Sul [ASAS] cujas principais características no Nordeste brasileiro são temperaturas e umidade elevadas e nebulosidade considerável.

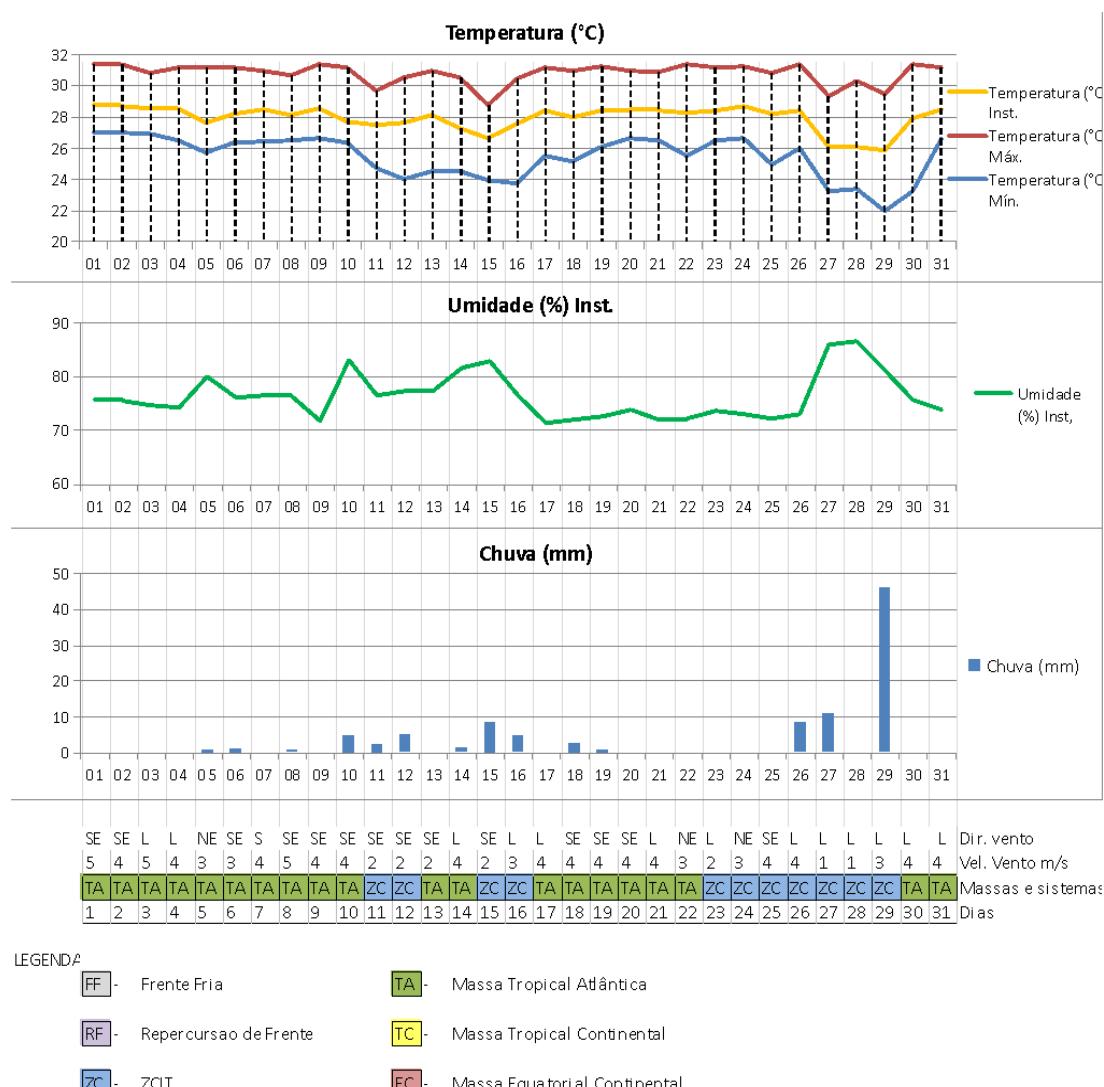


Figura 4: Análise rítmica dos tipos de tempo - Natal, Janeiro de 2022. Fonte: Resultados originais da pesquisa.

Ainda no exemplo de janeiro, observou-se que os momentos com chuva foram associados a presença da Zona de Convergência Intertropical (ZCIT), sistema global gerado pelo encontro dos ventos alísio de Norte e Sul na região equatorial e que traz consigo alta nebulosidade e precipitação para os locais de baixas latitudes.

A ZCIT atua normalmente nos primeiros meses do ano na região de estudo, trazendo consigo valores elevados de precipitação e nebulosidade e mantendo as temperaturas elevadas, aja vista que o sistema se origina em uma região naturalmente quente. Sua identificação, além das características citadas, dá-se pela constatação de uma extensa faixa de nebulosidade nas baixas latitudes. A **Figura 5** apresenta a situação sinóptica no dia 06/03/2022 onde constatou-se a atuação da ZCIT sobre a cidade de Natal. Percebe-se uma extensa banda horizontal de nebulosidade ao longo da linha do Equador representando tal sistema.

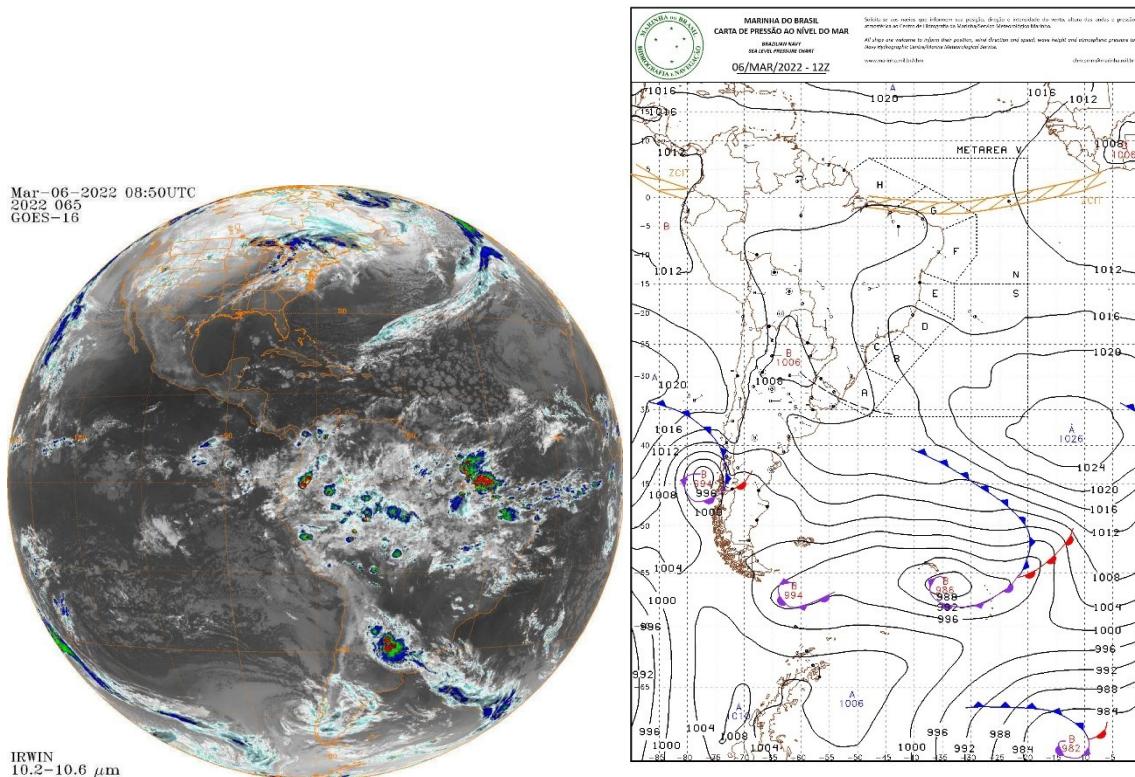


Figura 5: Atuação da ZCIT sobre Norte da América do Sul, Março de 2022. Fonte: Satélite GOES-16 e Marinha.

Ao longo de 2022, a Análise Rítmica atestou um predomínio de atuação do Sistema Tropical Atlântico e da ZCIT sobre o local de estudo. Além dos dois citados, observou-se também a atuação de Distúrbios Ondulatórios de Leste, este sistema de menor escala está associado a episódios de precipitação, com ocorrência principalmente nos meses de outono e inverno. Wanderley e Nobrega (2022) o descrevem como sistemas que modulam a convecção de umidade, trazendo episódios de chuva concentradas na região leste do nordeste brasileiro. Os autores indicam, todavia, que tal sistema está relacionado a atuação da Massa Tropical Atlântica (MTA), haja vista que, igualmente tem sua origem no deslocamento do Anticiclone do Atlântico Sul. Assim, nos dias de ocorrência dos Distúrbios Ondulatórios de Leste, o mesmo foi tratado como uma variação da atuação da Massa Tropical Atlântica.

Ao final portanto, foram constatados a atuação de três sistemas atmosféricos diferentes, cujas características foram sintetizadas na **Tabela 5**. Durante os dias em que se identificou a ocorrência da MTA, os registros de precipitação foram escassos, com média de 0,8mm/dia. A umidade relativa média foi igualmente a mais baixa em comparação aos demais sistemas e as temperaturas foram as maiores, com médias de temperatura máxima superando os 30°C. Durante a atuação deste sistema, observou-se também uma maior quantidade de insolação, com média de 9,5 horas/dia, o que vem de acordo com as características desta massa de ar em relação as demais.

A presença do DOL por outro lado, aumentou significativamente os valores de precipitação, com média de 19,4/mm. A insolação foi a menor, com apenas 5,4 horas de sol/dia em média, o que resulta em valores menores de temperatura, com temperatura máxima média de 28,7°C, a menor dentre os três sistemas

identificados. A umidade relativa do ar nos dias com atuação do DOL, por outro lado, foram as mais elevadas, com média de 86%.

Tabela 5: Características dos sistemas atmosféricos atuantes em Natal - 2022.

	Massa Tropical Atlântica			Massa Tropical Atlântica / DOL			Zona de Convergência Intertropical		
	Min	Méd	Max	Min	Méd	Max	Min	Méd	Max
Prec. (mm)	0	0,8	22,7	0	19,4	119	0	10,4	119,1
Umidade Relativa (%)	65,7	77,2	95	71,7	86,0	97	71,2	80,4	90,5
Temp. Máx (°C)	28,1	30,1	31,9	24,1	28,7	31,1	26,3	30,2	31,3
Temp. Méd (°C)	24,6	27,1	28,9	22,9	25,5	27,6	25,1	27,3	28,6
Temp. Min. (°C)	20,7	24,4	27,5	21,3	22,8	25,9	22	24,4	27,1
Insolação (horas)	0,8	9,5	11,5	0	5,4	10,8	0	6,5	11,2
Vel.Vento (m/s)	0,8	4,0	7,0	0,5	3,1	7,2	0,5	2,6	4,8
Dir. Vento	-	Se/S	-	-	Se	-	-	Se/E	-

Fonte: Resultados originais da pesquisa.

Durante os dias de atuação da ZCIT tivemos uma insolação inferior aos dias de MTA, mas superior aos dias com ocorrência de DOL, com uma média de 6,5 horas de sol/dia. A temperatura máxima oscilou entre 26,3°C e 31,3°C, com média de 30,2°C. Salienta-se que nos três sistemas atuantes, os ventos foram predominantes de direção Sudeste, embora sob MTA também tenham sido registrados ventos sentido Sul e sob a ZCIT tenham ocorrido ventos sentido Leste.

A **Figura 6** apresenta a disposição temporal da atuação dos sistemas detectados durante o procedimento de análise rítmica. Durante o período analisado, o sistema predominante foi a Massa Tropical Atlântica, ocorrendo em 61,5% dos dias analisados. Tal sistema mostrou-se atuante durante o ano inteiro, com períodos menores apenas nos meses de outono, entre março e junho.

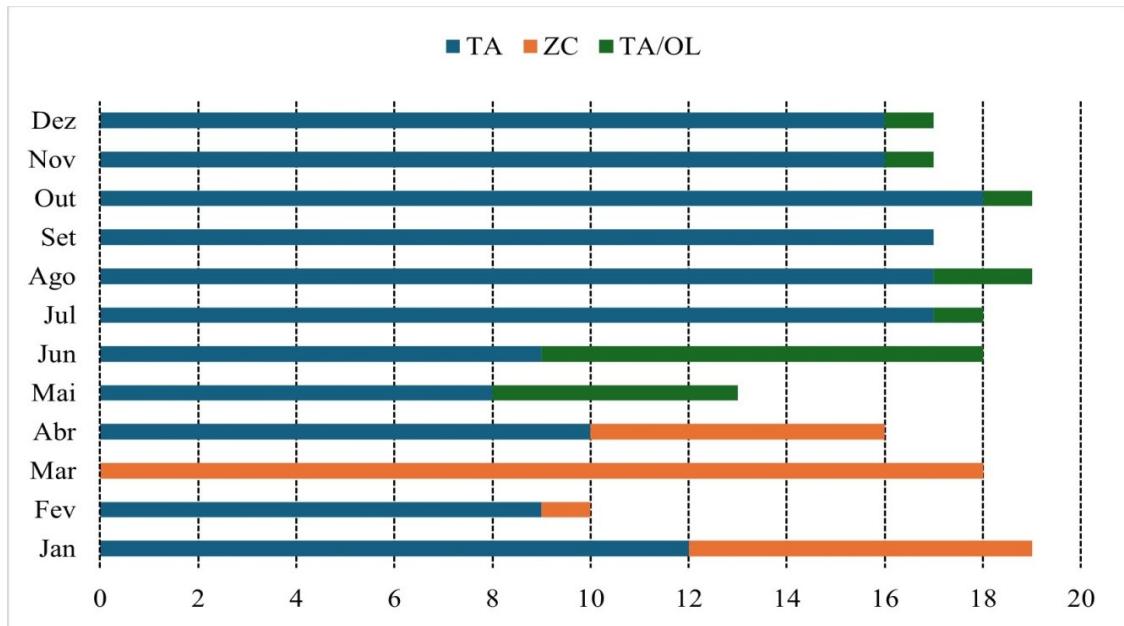


Figura 6: Distribuição Temporal dos Tipos Climáticos da Análise Rítmica dos Tipos de Tempo. Fonte: Resultados originais da pesquisa

Em seguida, temos a atuação do sistema tropical atlântico aliado aos Distúrbios Ondulatórios de Leste (TA/OL) com 19,8% de atuação, e presença mais significativa nos meses de maio, junho e julho.

Por fim, a atuação da ZCIT se limitou a 18,6% do tempo em Natal no ano de 2022, atuando predominantemente no primeiro semestre do ano, destaque para os meses de fevereiro, março e abril. Os resultados são similares aos de Wanderley e Nobrega (2022). Os autores reforçam que para o tipo climático de Natal, tem-se o predomínio de sistemas de origem Tropical, com maior ocorrência dos Sistemas tropicais úmidos com ocorrência dos Distúrbios Ondulatórios de Leste, e uma menor ocorrência de sistemas de origem equatorial (ZCIT).

3.3. Comparação entre análises sinóptica automatizada e híbrida

Embora ambas as técnicas aplicadas tivessem o mesmo objetivo, classificar os tipos de tempo atuantes na cidade analisada com base nos seus elementos meteorológicos, fica claro pelos resultados obtidos que ambas tiveram divergências em seus resultados.

A técnica automatizada agrupou os tipos de tempo em quatro classes enquanto, pela análise rítmica, detectaram-se apenas três tipos de tempo atuantes na cidade.

Enquanto na classificação automatizada perceberam-se nítidas diferenças nos valores dos elementos climáticos em cada classe, na análise rítmica tal fato mostrou-se com menos clareza, com diferenças menos perceptíveis entre os elementos analisados.

Comparando as classes obtidas nos dois métodos, percebe-se uma semelhança entre a classe 1 da classificação automatizada e a atuação da MTA na análise rítmica.

Ambas foram as classes com menores valores de precipitação, e maiores temperaturas e insolação em ambos os cenários. De fato, constatou-se que em 95,9% dos dias classificados como tipo 1 pela classificação automatizada, tivemos a ocorrência da Massa Tropical Atlântica.

Todavia, na classe climática do tipo 2, observou-se que a MTA ocupou cerca de 50,9% do tempo, seguido pela ZCIT que ocupou 26,1% do tempo dessa classe, e pôr fim a MTA atuando com o DOL com 23,6%. A classe 2 mostrou-se como uma transição entre um tipo climático mais estável (classe 1) e tipos de tempo com mais nebulosidade. Nesta classe, foram observados menores valores de insolação em relação a classe 1 e igualmente mais eventos de chuva, embora as temperaturas tenham se mantido elevadas.

Na classe 3, 55,5% do tempo foi classificado como atuação de Distúrbios Ondulatórios de Leste associados a MTA, em seguida tivemos 33,3% do tempo classificados como atuação da ZCIT, e apenas 13,3% do tempo classificado como atuação da MTA. Neste caso, trata-se de uma classe climática (classe 3) com eventos mais intensos de chuva, causando menos insolação e menores temperaturas, características típicas de dias com atuação da ZCIT e dos DOL.

Por fim, a classe 4, cuja ocorrência se limitou a apenas três dias do período analisado, situou-se como um tipo de tempo com baixa insolação e elevadas precipitações, além de alta umidade e temperaturas mais amenas. Em todos os dias classificados como classe 4 foram constatados a atuação dos Distúrbios Ondulatórios de Leste junto a sistemas atlânticos. A comparação entre ambos esta apresentada na **Tabela 6**.

Tabela 6: Ocorrência dos sistemas atmosféricos detectados nas classes definidas pelo modelo automatizado.

	Massa Tropical Atlântica		Massa Tropical Atlântica / DOL		Zona de Convergência Intertropical	
	Dias	% de ocorrência	Dias	% de ocorrência	Dias	% de ocorrência
Classe 1	118	95,9	0	0	4	4,1
Classe 2	82	50,9	38	23,6	42	26,1
Classe 3	6	13,3	25	55,5	15	33,3
Classe 4	0	0	4	100	0	0

Fonte: Resultados originais da pesquisa

4. Considerações finais

O presente trabalho propôs uma metodologia automatizada para análise e classificação dos tipos de tempo atuante em dada localidade a partir de técnicas de “Machine Learning”, especificamente Análise de Componentes Principais e *Clusterização* pelo método k-means. Ademais da aplicação do método, realizou-se uma comparação do mesmo com a técnica híbrida de análise sinóptica da análise rítmica. Os resultados obtidos pelo modelo automatizado mostraram-se válidos quando se percebe que cada classe criada apresenta individualidades no comportamento de seus elementos climáticos e certo grau de similaridade com as características dos tipos de tempo detectados pela análise rítmica. Percebe-se todavia, que as classes 2 e 3 do modelo foram aquelas com maiores incongruências em relação aos resultados da análise rítmica, demonstrando dificuldade do modelo automatizado em definir características de sistemas atmosféricos instáveis como a ZCIT e os DOL.

A análise rítmica, apesar de ser uma técnica precisa e já muito utilizada no Brasil, mostra-se como uma ferramenta que despende muito tempo de seu usuário, aja vista sua demanda qualitativa. Tal fato limita sua

utilização a séries temporais curtas e análises episódicas. A análise automatizada por outro lado, abre um novo leque de possibilidades na análise climática, uma vez que permite ao usuário trabalhar com grandes volumes de dados e com séries mais longas. Em tempos de mudanças climáticas e *big data*, uma técnica que consiga unir ambos pode propiciar uma nova forma de analisar o clima e suas dinâmicas.

Referências

- BESSAH, E.; AMPONSAH, W.; ANSAH, S. O.; AFRIFA, A.; YAHAYA, B., WEMEGAH, C. S.; TANU, M.; AMEKUDZI, L. K.; AGYARE, W. A. Climatic zoning of Ghana using selected meteorological variables for the period 1976–2018. **Meteorological Applications**, 29(1), e2049, 2022.
- CUNHA, D. G. F.; VECCHIA, F. As Abordagens Clássica e Dinâmica de Clima: Uma Revisão Bibliográfica Aplicada ao Tema da Compreensão da Realidade Climática. **Ciência e Natura**, v. 29, n. 1, p. 137–149, 2007
- DENTON, S. R.; BARNES, B. V., An Ecological Climatic Classification of Michigan: A Quantitative Approach, **Forest Science**, Volume 34, Issue 1, Pages 119–138, 1988
- DUBREUIL V.; FANTE K.P.; PLANCHON O.; SANT'ANNA NETO J.L., Climate change evidence in Brazil from Köppen's climate annual types frequency. **Int J Climatol**. V.39, 39, p. 1446–1456, 2019
- FONTÃO, P. A. B.; ZAVATINNI, J. A.; SHERIDAN, S. C.; ARMOND, N. B. Gênese Das Chuvas Em São Paulo (Sp): Estudo Comparativo entre a "Spatial Synoptic Classification" e a "Análise Rítmica Em Climatologia". **Revista Brasileira de Climatologia**, v. 23, 2021
- GÜLAGIZ, F. K.; SAHIN, S., Comparison of hierarchical and non-hierarchical clustering algorithms. **International Journal of Computer Engineering and Information Technology**, 9 (1), 6-14, 2017
- KOJIMA, C., Detailed Climatic Classification of Tohoku District by Principal Component Analysis, **Journal of Agricultural Meteorology**, Volume 29, Issue 3, Pages 165-172, 1973
- NICHOLSON, J.; BRYANT, D. G., Climatic zones of insular Newfoundland: a principal component analysis, **Canadian Forestry Service**, No. 1299, iv, 13 p, 1972
- MACHADO, L. A., Classificação Climática Para Minas Gerais Por Meio Do Método De Agrupamento Não Hierárquico De K-Means. **Cadernos do Leste**, [S. l.], v. 14, n. 14, 2014
- MCGOVERN, A.; SUPINIE, T.; GAGNE, J.D.; TROUTMAN, N.; COLLIER, M.; BROWN, R.; BASARA, J.; WILLIANS, J., Understanding Severe Weather Processes Through Spatiotemporal Relational Random Forests, In: CONFERENCE ON INTELIGENT DATA UNDERSTANDING, 2010.
- MONTEIRO, C. A. F. A análise rítmica em climatologia: problemas da atualidade climática em São Paulo e alegrias para um programa de trabalho. **Climatologia**, n. 01, p. 01-21, 1971
- NOVAIS, G.T.; MACHADO, L.A., Os climas do Brasil: segundo a classificação climática de Novais. **Revista Brasileira de Climatologia**, [S. l.], v. 32, n. 19, p. 1–39, 2023
- SANT'ANNA NETO, J. L. Medir o tempo, ler, interpretar e sentir o clima: uma climatologia a serviço do entendimento das relações entre os homens e seus lugares, entre as sociedades e seus territórios. In: SILVA, C. A.; FIALHO, E. S.; STEINKE, E. T. **Experimentos em Climatologia Geográfica**. Dourados-MS: UFGD, 2014. p. 23-32
- SHERIDAN, S. C. The redevelopment of a weather-type classification scheme for North America. **International Journal of Climatology**, v. 22, n. 1, p. 51-68, 2002
- WANDERLEY, L. S. A.; NÓBREGA, R. S. Desenvolvimento de um novo sistema de classificação climática com base na metodologia dos tipos de tempo sinóticos para a região Nordeste do Brasil. **Geousp**, v. 26, n. 1, e-175243, abr. 2022

ZAVATINNI, J. A. A Climatologia Geográfica Brasileira, o Enfoque Dinâmico e a Noção de Ritmo Climático. **Geografia (Rio Claro)**, v. 23, n. 3, p. 5-24, 1998

ZHANG,X.; MAGGIONI,V.;Houser,P.;XUE,Y.;MEI, Y.,The impact of weather condition and social activity on COVID-19 transmission in the United States, **Journal of Environmental Management**, Volume 302, Part B, 2022,

ZHOU, K.; ZHENG, Y.; LI, B., Forecasting Different Types of Convective Weather: A Deep Learning Approach. **J Meteorol Res** 33, 797–809, 2019.



Este artigo é distribuído nos termos e condições do *Creative Commons Attributions/Atribuição-NãoComercial-Compartilhamento* (CC BY-NC-SA).