

**UNIVERSIDADE FEDERAL DO TOCANTINS
CÂMPUS UNIVERSITÁRIO DE PALMAS
CURSO DE CIÊNCIA DA COMPUTAÇÃO**

**APLICAÇÃO DE TÉCNICAS DE MINERAÇÃO DE DADOS PARA
ANÁLISE CLIMÁTICA NO ESTADO DO TOCANTINS**

MARCOS GABRIEL PEREIRA ARAUJO

PALMAS (TO)

2021

MARCOS GABRIEL PEREIRA ARAUJO

APLICAÇÃO DE TÉCNICAS DE MINERAÇÃO DE DADOS PARA ANÁLISE
CLIMÁTICA NO ESTADO DO TOCANTINS

Trabalho de Conclusão de Curso II apresentado
à Universidade Federal do Tocantins para
obtenção do título de Bacharel em Ciência da
Computação, sob a orientação do(a) Prof.(a)
Dra. Glenda Michele Botelho.

Orientador: Dra. Glenda Michele Botelho

PALMAS (TO)

2021

MARCOS GABRIEL PEREIRA ARAUJO

APLICAÇÃO DE TÉCNICAS DE MINERAÇÃO DE DADOS PARA ANÁLISE
CLIMÁTICA NO ESTADO DO TOCANTINS

Trabalho de Conclusão de Curso II apresentado à UFT – Universidade Federal do Tocantins – Câmpus Universitário de Palmas, Curso de Ciência da Computação foi avaliado para a obtenção do título de Bacharel e aprovada em sua forma final pelo Orientador e pela Banca Examinadora.

Data de aprovação: 23 / 4 / 2021

Banca Examinadora:

Profa. Dra. Glenda Michele Botelho

Prof. Dr Alexandre Tadeu Rossini da Silva

Prof. Dr. Ary Henrique Morais de Oliveira

Dados Internacionais de Catalogação na Publicação (CIP)
Sistema de Bibliotecas da Universidade Federal do Tocantins

- A663a ARAUJO, MARCOS GABRIEL PEREIRA.
Aplicação de Técnicas de Mineração de Dados para Análise Climática no Estado do Tocantins. / MARCOS GABRIEL PEREIRA ARAUJO. – Palmas, TO, 2021.
83 f.

Monografia Graduação - Universidade Federal do Tocantins – Câmpus Universitário de Palmas - Curso de Ciências da Computação, 2021.
Orientadora : Glenda Michele Botelho

1. Descoberta de Conhecimento em Bases de Dados. 2. Mineração de Dados. 3. Análise Climática. 4. Tocantins. I. Título

CDD 004

TODOS OS DIREITOS RESERVADOS – A reprodução total ou parcial, de qualquer forma ou por qualquer meio deste documento é autorizado desde que citada a fonte. A violação dos direitos do autor (Lei nº 9.610/98) é crime estabelecido pelo artigo 184 do Código Penal.

Elaborado pelo sistema de geração automática de ficha catalográfica da UFT com os dados fornecidos pelo(a) autor(a).

*Dedico este trabalho de pesquisa
aos meus pais. Sua grande força
foi o que permitiu o meu
crescimento, mesmo durante os
momentos mais difíceis.
Agradeço do fundo do meu
coração.*

AGRADECIMENTOS

Gostaria de agradecer principalmente à minha mãe Jaimira Rodrigues Pereira, por todo o apoio e ajuda emocional, que muito contribuiu para a realização deste trabalho.

Aos amigos, que sempre estiveram ao meu lado, pela amizade incondicional e pelo apoio demonstrado ao longo de todo o período de tempo em que me dediquei a este trabalho.

A professora Glenda Michele Botelho, por ter sido minha orientadora e ter desempenhado tal função com dedicação.

Aos meus colegas de curso, com quem convivi intensamente durante os últimos anos, pelo companheirismo e pela troca de experiências que me permitiram crescer não só como pessoa, mas também como formando.

RESUMO

Mudanças climáticas são fenômenos com alto índice de variabilidade que, eventualmente, podem causar eventos extremos, como secas ou chuvas intensas. Esses eventos podem gerar impacto na socioeconomia, na agropecuária de uma região e até mesmo a perda de vidas humanas. Logo, ter conhecimento sobre as frequências, comportamento e mudanças climáticas é de suma importância para realizar um planejamento que busque minimizar danos causados pelo clima intenso. Atualmente, informações climáticas estão sendo coletadas e armazenadas em grandes bases de dados. A extração de conhecimento pode ser feita através de inteligência computacional como, por exemplo, o processo de Descoberta de Conhecimento em Bases de Dados (*Knowledge Discovery in Databases*), que possui um conjunto de técnicas de Mineração de Dados capazes de extrair padrões úteis. Baseado neste contexto, este trabalho tem por objetivo aplicar o processo de Descoberta de Conhecimento em dados climáticos do estado do Tocantins. Para isso, utilizou-se séries temporais históricas do Instituto Nacional de Meteorologia. Inicialmente, aplicou-se a tarefa de agrupamento para definir as zonas climáticas homogêneas para o primeiro vintênio (1980 a 1999) e para o segundo vintênio (2000 a 2019), ambos com a mesma quantidade de zonas. Para analisar as variações frequenciais e comportamentais climáticas, adotou-se a pesquisa conjunta entre os *clusters* para análise da variabilidade e a pesquisa individual em diferentes escalas (anual, mensal e diária), realizada por meio da Técnica dos Quantis e Análise de Sazonalidade sobre a variável de precipitação. Os resultados mostraram que as técnicas de mineração de dados e análise estatística são adequadas para estudos sobre comportamento e variações climáticas, reforçado que, por meio desta pesquisa, identificou-se as principais décadas com registro de eventos extremos no Tocantins.

Palavra-chave: Descoberta de Conhecimento em Bases de Dados. Mineração de Dados. Análise Climática. Tocantins.

ABSTRACT

Climatic changes are phenomena with a high rate of variability that, eventually, can cause extreme events, such as droughts or intense rains. These events can have an impact on socioeconomics, agriculture in a region and even the loss of human lives. Therefore, having knowledge about the frequencies, behavior and climate changes is extremely important to carry out a plan that seeks to minimize damage caused by the intense climate. Currently, climate information is being collected and stored in large databases. Knowledge extraction can be done through computational intelligence such as, for example, the Knowledge Discovery in Databases process, which has a set of data mining techniques capable of extracting useful patterns. Based on this context, this work aims to apply the Knowledge Discovery process to climatic data from the state of Tocantins. For this, historical time series from the National Institute of Meteorology was used. Initially, the grouping task was applied to define homogeneous climatic zones for the first twenty years (1980 to 1999) and for the second twenty years (2000 to 2019), both with the same number of zones. To analyze the frequency and climatic behavioral variations, a joint research between the clusters was adopted to analyze the variability and the individual research at different scales (annual, monthly and daily), carried out using the Quantis and Analysis Technique of Seasonality on the precipitation variable. The results showed that the techniques of data mining and statistical analysis are suitable for studies on behavior and climatic variations, reinforced that, through this research, the main decades were identified with record of extreme events in Tocantins.

Keywords: Knowledge Discovery in Databases. Datamining. Climate Analysis. Tocantins.

LISTA DE FIGURAS

Figura 1 – Safras da agricultura no estado do Tocantins (SEFAZ, 2020).	24
Figura 2 – Etapas do processo de KDD. Figura adaptada de Fayyad (1996). . .	26
Figura 3 – Etapas do modelo CRISP-DM. Figura adaptada de Chapman et al. (2000).	27
Figura 4 – Principais tarefas de mineração de dados.	29
Figura 5 – Distribuição espacial das 14 estações meteorológicas selecionadas no estado Tocantins.	36
Figura 6 – Sazonalidade da Precipitação para ano de 1980 do Cluster 0.	42
Figura 7 – Distribuição de Precipitação e Umidade Relativa do Ar do Primeiro Vintênio no estado do Tocantins.	44
Figura 8 – Distribuição da Temperatura Máxima e Mínima no Primeiro Vintê- nio no estado do Tocantins.	45
Figura 9 – Distribuição de Precipitação e Umidade Relativa do Ar do Segundo Vintênio no estado do Tocantins.	45
Figura 10 – Distribuição da Temperatura Máxima e Mínima no Segundo Vintênio no estado do Tocantins.	45
Figura 11 – Resultado da aplicação do algoritmo <i>k-means</i> nos dados do primeiro vintênio.	47
Figura 12 – Resultado da aplicação do algoritmo <i>k-means</i> nos dados do segundo vintênio.	47
Figura 13 – Precipitação média mensal dos <i>Clusters</i>	50
Figura 14 – Delimitação das Zonas Homogêneas comuns entre os Vintênios (1-3).	51
Figura 15 – Análise das variáveis considerando os dois vintênios presentes nas Zonas.	52
Figura 16 – Distribuição da precipitação anual para as zonas homogêneas no pri- meiro vintênio.	54

Figura 17 – Distribuição da precipitação anual para as zonas homogêneas no segundo vintênio.	55
Figura 18 – Sazonalidade da precipitação dos anos 1985 e 1984 para o <i>Cluster</i> 0.	58
Figura 19 – Sazonalidade da precipitação dos anos de 1985 e 1993 para o <i>Cluster</i> 1.	59
Figura 20 – Sazonalidade da precipitação dos anos de 1985 e 1993 para o <i>Cluster</i> 2.	60
Figura 21 – Sazonalidade da precipitação dos anos de 2000 e 2012 para o <i>Cluster</i> 3.	60
Figura 22 – Sazonalidade da precipitação dos anos de 2009 e 2012 para o <i>Cluster</i> 4	61
Figura 23 – Sazonalidade da precipitação dos anos de 2005 e 2012 para o <i>Cluster</i> 5.	61
Figura 24 – Distribuição da precipitação mensal para as zonas homogêneas do primeiro vintênio.	63
Figura 25 – Distribuição da precipitação mensal para as zonas homogêneas do segundo vintênio.	64
Figura 26 – Registro dos meses indicados como “muito secos” para o <i>Cluster</i> 0.	65
Figura 27 – Registro dos meses indicados como “muito chuvosos” para o <i>Cluster</i> 0.	66
Figura 28 – Registro dos meses indicados como “muito secos” para o <i>Cluster</i> 1.	66
Figura 29 – Registro dos meses indicados como “muito chuvosos” para o <i>Cluster</i> 1.	67
Figura 30 – Registro dos meses indicados como “muito secos” para o <i>Cluster</i> 2.	67
Figura 31 – Registro dos meses indicados como “muito chuvosos” para o <i>Cluster</i> 2.	68
Figura 32 – Registro dos meses indicados como “muito secos” para o <i>Cluster</i> 3.	68
Figura 33 – Registro dos meses indicados como “muito chuvosos” para o <i>Cluster</i> 3.	69
Figura 34 – Registro dos meses indicados como “muito secos” para o <i>Cluster</i> 4.	69
Figura 35 – Registro dos meses indicados como “muito chuvosos” para o <i>Cluster</i> 4.	70
Figura 36 – Registro dos meses indicados como “muito secos” para o <i>Cluster</i> 5	70
Figura 37 – Registro dos meses indicados como “muito secos” para o <i>Cluster</i> 5	71
Figura 38 – Distribuição da precipitação diária para as zonas homogêneas do primeiro e segundo vintênio.	74
Figura 39 – Distribuição da quantidade de dias extremamente secos e o número de sequências com dias consecutivos para os <i>clusters</i>	76

LISTA DE TABELAS

Tabela 1 – Relação das cidades com o tipo de estação.	35
Tabela 2 – Atributos referentes às bases de dados das estações.	36
Tabela 3 – Categorias da Técnica dos Quantis para as escala Anual e Mensal . .	41
Tabela 4 – Categorias da Técnica dos Quantis para a escala Diária.	41
Tabela 5 – Esquema para obter Sazonalidade utilizando a Precipitação Acumulada Mensal e Precipitação Média Anual do <i>Cluster</i> 0 (período de 1980 a 1999).	42
Tabela 6 – Esquema para obter a Média da Sazonalidade Mensal utilizando a sazonalidade de cada mês do <i>Cluster</i> 0 (período de 1980 a 1999). . .	42
Tabela 7 – Valores da Precipitação Anual, médias, desvio-padrão, máximo e mínimo de cada <i>Cluster</i>	48
Tabela 8 – Valores da Umidade Relativa do Ar Média Anual, médias, desvio-padrão, máximo e mínimo de cada <i>Cluster</i>	49
Tabela 9 – Valores da Temperatura Máxima Anual, médias, desvio-padrão, máximo e mínimo de cada <i>Cluster</i>	49
Tabela 10 – Valores da Temperatura Mínima Anual, médias, desvio-padrão, máximo e mínimo de cada <i>Cluster</i>	49
Tabela 11 – Tabela de desvios das variáveis climáticas do primeiro vintênio em relação ao segundo vintênio.	53
Tabela 12 – Registro da precipitação e classificação dos anos pelos quantis nos <i>clusters</i> . Categorias: MC = muito chuvoso, C = chuvoso, N = normal, MS = muito seco e S = seco.	56
Tabela 13 – Limiares de classificação anual para os eventos extremos muito seco e muito chuvoso relacionados com cada <i>cluster</i>	57
Tabela 14 – Relação dos anos que apresentaram ocorrência de eventos extremos (muito chuvoso e muito seco).	57
Tabela 15 – Período de ocorrência e intensidade dos fenômenos <i>El Niño</i> e <i>La Niña</i>	62

Tabela 16 – Limiares de classificação mensal para os eventos extremos muito seco e muito chuvoso relacionados com cada <i>cluster</i>	65
Tabela 17 – Registro de frequência dos eventos extremos para os meses de todos os anos de toda a serie temporal histórica.	72
Tabela 18 – Limiares de classificação diário para os eventos extremos “Extremamente Seco” e “Extremamente Chuvoso” relacionados com cada <i>cluster</i>	74

SUMÁRIO

1	INTRODUÇÃO	15
1.1	Justificativa	16
1.2	Hipótese	17
1.3	Objetivos	17
1.4	Organização do Trabalho	17
2	REVISÃO BIBLIOGRÁFICA	19
2.1	Ambientação da Área Pesquisada	19
2.1.1	Eventos Climáticos	19
2.1.2	Impactos Climatológicos	20
2.1.3	Estado do Tocantins	22
2.1.3.1	Clima no Estado do Tocantins	22
2.1.3.2	Conceitos e Impactos Econômicos no Estado do Tocantins	23
2.2	Processo de Descoberta de Conhecimento em Bases de Dados .	25
2.2.1	Modelos de Descoberta de Conhecimento	26
2.2.1.1	Compreensão do domínio	27
2.2.1.2	Entendimento dos dados	27
2.2.1.3	Preparação dos dados	28
2.2.1.4	Modelagem	28
2.2.1.5	Avaliação	28
2.2.1.6	Distribuição	29
2.2.2	Tarefas e Técnicas de Mineração de Dados	29
2.2.2.1	Predição	30
2.2.2.2	Descrição	30
2.2.3	Descrição das Técnicas de Mineração de Dados	31

2.2.3.1	Classificação	31
2.2.3.2	Clusterização	32
2.3	Aplicação do processo de Descoberta de Conhecimento na climatologia	32
3	MATERIAIS E MÉTODOS	35
3.1	Materiais	35
3.2	Métodos	36
3.2.1	Entendimento dos dados	37
3.2.2	Preparação dos dados	37
3.2.3	Modelagem	38
3.2.4	Análise dos Resultados	39
3.2.4.1	Análise dos Dados	39
3.2.4.2	Técnica dos Quantis	40
3.2.4.3	Análise Anual	41
3.2.4.4	Análise Mensal	43
3.2.4.5	Análise Diária	43
4	RESULTADOS E DISCUSSÃO	44
4.1	Análise da Distribuição Climática	44
4.2	Análise das Zonas Homogêneas	46
4.3	Análise Conjunta	50
4.4	Análise de Eventos Extremos de Precipitação	53
4.4.1	Análise Anual	53
4.4.2	Análise Mensal	63
4.4.3	Análise Diária	72
5	CONCLUSÃO	78
5.1	Trabalhos futuros	79
	REFERÊNCIAS	80

1 INTRODUÇÃO

Os termos mudança, variação e eventos extremos climáticos tiveram sua procura acentuada como objeto de estudo de diversos pesquisadores, tanto da iniciativa pública quanto privada, nos últimos tempos (ISP-GIFE, 2019). Um dos motivos que levam a esta busca se deve ao fato das alterações climáticas terem se intensificado nas últimas décadas, causando danos diretamente na biodiversidade da região afetada (CANHOS et al., 2010). O impacto é maior ao considerar as perdas econômicas e de vidas, quando os eventos são mais intensos, por conta da imprevisibilidade causada pela variabilidade climática.

De acordo com os relatórios do IPCC (*Intergovernmental Panel on Climate Change*), órgão que provê avaliações sobre as mudanças climáticas globais, nos próximos anos, as projeções sobre os eventos extremos estão relacionadas com cenários de chuvas intensas e secas prolongadas ao redor do planeta, além de outros eventos que se encontram ligados a desastres naturais, como furacões e tempestades (IPCC, 2012). A prevenção destes eventos e a sua previsão estão ligados a fatores e variáveis climáticas.

Atualmente, a era da informação utiliza sistemas receptores para captar e armazenar várias informações climáticas em grandes bases de dados, com as quais os sistemas de informação podem interagir. Os bancos de dados estão em constante crescimento e sua evolução tende para o aumento no número de registros e dos atributos armazenados. Baseado neste contexto, sabe-se que a partir desta enorme quantidade de informação armazenada podem existir padrões e relações importantes nos dados, que se encontram ocultos.

Contudo, verificar um alto quantitativo de dados de forma manual é praticamente impossível. Por isso, foram desenvolvidos meios tecnológicos para a análise de dados e a descoberta de padrões que se encontram invisíveis. O processo mais popular e com um grau elevado de eficiência para encontrar informações valiosas é a Descoberta de Conhecimento em Bases de Dados (*Knowledge Discovery in Database - KDD*).

O KDD faz parte do campo da inteligência computacional, o qual se mantém em constante expansão, sendo aplicado em diversas áreas como, por exemplo, negócios, *marketing*, medicina ou empresarial. Este processo é capaz de identificar padrões importantes, que não eram localizados por sistemas simples, interpretá-los e transformá-los em conhecimento útil (FAYYAD, 1996). O processo tem a capacidade de realizar uma análise em grandes bases de dados e por meio das técnicas de mineração escolhidas, consegue obter o objetivo proposto ao desenvolver a aplicação.

De acordo com Tan, Steinbach e Kumar (2016), a mineração de dados é um processo de descoberta automatizado de informações importantes em grandes bases de dados, de modo a descobrir padrões úteis que se encontravam ocultos. Basicamente, consiste em uma

etapa do processo de KDD, responsável pelo procedimento de extração de conhecimento, sendo, muitas vezes, tida como um sinônimo do processo completo. Com isso, sistemas podem ser desenvolvidos usando métodos e algoritmos de mineração de dados para o propósito de analisar dados climáticos, podendo prever as condições e eventos do clima.

Baseado neste contexto, esse trabalho busca aplicar técnicas de mineração de dados sobre os dados climáticos do estado do Tocantins, focado em transformar as séries históricas de dados em zonas climáticas homogêneas, por intermédio de agrupamento de dados. Sendo assim, a partir das zonas homogêneas, utilizando da Técnica dos Quantis, foi possível analisar e avaliar a ocorrência dos eventos extremos nas escalas anuais, mensais e diárias. A partir desse estudo espera-se que, com análise dos eventos extremos de cada zona climática, possa auxiliar no suporte à tomada de decisão, com medidas preventivas dos eventos extremos para o setor agropecuário do estado.

1.1 Justificativa

A região Norte do país está suscetível a sofrer com as mudanças climáticas (ONU BRASIL, 2016), especialmente a seca, que se apresenta com certa frequência, causando problemas no setor socioeconômico da região. Anexado a região norte, o estado do Tocantins, possui uma característica climática e de vegetação mais relacionada com a região centro-oeste, possuindo clima tropical e o bioma cerrado como vegetação predominante, mas ainda assim sofre com problemas climáticos que afetam a região norte do país.

Nos últimos anos, o estado do Tocantins sofreu com a ocorrência de secas prolongadas, que afetaram consideravelmente as produções agrícolas. No ano de 2016, os efeitos desse evento atingiram a produção de soja de alguns municípios do estado, devido a falta de umidade relacionada e condicionada com as temperaturas elevadas. Esse caso teve relação com o El Niño, evento que provoca grandes alterações no clima, que ocorreu durante o período, tendo como principal fator a baixa precipitação, que causa grande impacto nas atividades agrícolas (SEAGRO, 2016).

As atividades agrícolas dependem das variáveis climáticas, as quais podem variar com frequência. Caso as condições climáticas não sejam favoráveis, pode resultar em péssimos cultivos. Já as condições climáticas mais agradáveis tendem a tornar as safras mais produtivas. Com os avanços tecnológicos, a agricultura moderna teve seu crescimento ao longo do tempo, sendo fundamental o conhecimento do clima da região para que se possa indicar e planejar as práticas de plantio adequadas.

Diante deste cenário, surge a necessidade do desenvolvimento de sistemas computacionais que possam analisar os fatores e as variáveis climáticas, com o intuito de apresentar resultados confiáveis sobre as mudanças no clima que afetam uma determinada região. Com isso, pode-se minimizar problemas ambientais e socioeconômicos gerados diretamente por eventos climáticos. Neste contexto, técnicas sofisticadas de análise

de dados vem sendo desenvolvidas e aprimoradas. O KDD é um desses processos, que permite analisar comportamentos climáticos e transformar uma massiva quantidade de informações de grandes bases de dados em conhecimento, que podem ser utilizados para o processo de tomada de decisão (KOH; TAN et al., 2011).

1.2 Hipótese

A hipótese científica desta pesquisa se baseia na possibilidade de uso de técnicas de mineração de dados para realizar a análise climática no Estado do Tocantins, com o intuito de compreender o comportamento e frequência dos eventos extremos, em diferentes escalas (anual, mensal e diária), por meio da base de dados histórica climática.

1.3 Objetivos

O principal objetivo deste projeto é aplicar técnicas de mineração de dados e estatística para extrair informações e conhecimento do banco de dados do INMET, desta maneira, analisar o comportamento e a frequência de eventos climáticos no estado do Tocantins entre os anos de 1980 a 2019, podendo utilizar tal informação para tomar decisões a respeito de prevenções em impactos ambientais e econômicos causados por possíveis eventos climáticos.

- Identificar zonas climáticas semelhantes no estado do Tocantins, para os anos de 1980 até 2019, por meio da aplicação de tarefas e algoritmos de mineração de dados.
- Definir as características pluviométricas para cada zona homogênea, por intermédio de análise das séries históricas dos dados para escala anual, mensal e diária.
- Encontrar o limiar de intensidade de chuva e seca, que caracterizam um evento extremo dentro de cada zona de similaridade.
- Realizar análise frequencial e de comportamento da correlação entre a variação dos índices pluviométricos e os eventos extremos que ocorreram ao longo dos anos, para cada zona homogênea.

1.4 Organização do Trabalho

Este trabalho está dividido em 5 capítulos, que abrangem assuntos relacionados a análise de dados climáticos, assim discriminados:

- Capítulo 1: apresenta a introdução do presente trabalho, além das justificativas, hipótese e objetivos.

- Capítulo 2: apresenta uma revisão bibliográfica sobre os temas envolvidos no estudo, definindo os conceitos da área de pesquisa (climatologia), o processo de descoberta de conhecimento em bases de dados e os trabalhos relacionados com a área, os quais servem como diretriz para o desenvolvimento deste estudo.
- Capítulo 3: expõe a metodologia que será adotada no trabalho para a realização do processo de descoberta de conhecimento em base de dados climáticos do estado do Tocantins, bem como os materiais envolvidos no processo.
- Capítulo 4: está relacionado com apresentação e discussão dos resultados, para as análises climáticas conjuntas e as análises individuais de precipitação pluvial em escalas anuais, mensais e diárias.
- Capítulo 5: apresenta as conclusões referentes às análises conjunta e individual, buscando consolidar o conhecimento sobre a pesquisa do trabalho, assim como ideias para trabalhos futuros.

2 REVISÃO BIBLIOGRÁFICA

Este capítulo está dividido em duas seções principais. A primeira se refere a aplicação a ser trabalhada, contendo informações sobre climatologia, bem como a caracterização sobre a região do estado do Tocantins, com ênfase no clima e seus impactos na economia local. Já a segunda parte está relacionada ao processo de Descoberta de Conhecimento em Bases de Dados, destacando conceitos gerais e as tarefas de mineração de dados.

2.1 Ambientação da Área Pesquisada

Nesta seção serão apresentados conceitos referentes à aplicação estudada, ou seja, análise de dados climáticos. Logo, as subseções seguintes contêm informações referentes aos eventos e impactos climáticos, além de dados sobre a região estudada, ou seja, o estado do Tocantins.

2.1.1 Eventos Climáticos

O estudo dos eventos climáticos extremos vem ganhando importância significativa, tanto que órgãos mundiais como a Organização Mundial da Saúde (OMS) definiu que as mudanças climáticas são uma das ameaças a saúde global (FUNDO NACIONAL DA SAÚDE, 2019). Os problemas e desastres são de tamanha relevância que em 1988 aconteceu a consolidação global de pesquisas e estudos sobre mudanças climáticas, que gerou o Painel Intergovernamental sobre Mudanças Climáticas (*Intergovernmental Panel on Climate Change - IPCC*), que resultou em relatórios com informações sobre as variabilidades climáticas, impactos climatológicos, atividades humanas e cenários sobre a mudança climática global (BARBOSA et al., 2008).

As informações e definições de mudança climática resultantes do IPCC estão relacionadas com a forma de identificar a alteração de um estado do clima, sendo a mesma realizada por métodos estatísticos e observatórios, a partir de medições e percepções de variabilidade, as quais devem ser executadas em um período considerável, tendo em vista que a percepção de mudanças climáticas requer um tempo duradouro de análises (UNFCC, 2011).

Em relação ao tema mudanças climáticas existem outras organizações, tais como o Painel Internacional Não Governamental sobre Mudanças Climáticas (sigla em inglês NIPCC), bastante popular e contrastante com os ideais do IPCC. NIPCC é uma organização formada por cientistas e acadêmicos, a qual não possui influência de governos. A organização contesta as ideias do IPCC sobre as mudanças climáticas, mais precisamente sobre as causas das mudanças. Entretanto, mesmo discordando das causas, o grupo reafirma que as mudanças climáticas são impactantes com relação à saúde humana, causam

grandes alterações de temperatura no planeta e os climas extremos devem possuir uma maior atenção (IDSO et al., 2009).

Segundo o IPCC (2012), um evento climático extremo pode ser considerado como um evento que ocorre de forma rara sobre um determinado local em uma certa época do ano. Porém, esse nível de raridade pode variar dependendo da probabilidade estimada a partir de uma função de densidade retirada das análises de ocorrências dos eventos climáticos na região. Um clima considerado extremo possui um nível de variância, por causa das mudanças climáticas, que varia de região para região. Caso ocorra persistência de clima extremo por um tempo considerável, este por vez será entendido como um evento climático extremo e sua resultante pode ser identificada como seca ou chuvas fortes durante uma determinada estação.

Os eventos climáticos extremos são objetos intrigantes na pesquisa científica, pois tem como característica a imprevisibilidade e a variação climática. Fatores como frequência e intensidade podem variar de acordo com a mudança climatológica, que está correlacionada com diversas ações naturais como os processos bióticos, erupções vulcânicas e placas tectônicas, ou fatores antropogênicos, que envolvem as atividades humanas de queimadas e desmatamentos (MARENGO et al., 2009).

MARENGO et al. (2009) considera eventos climáticos extremos como ocorrências que geram estragos e provocam um alerta populacional como, por exemplo, enchentes, secas prolongadas, ondas de calor, tufões e tornados. Além disso, avalia certos eventos como grandes causadores de impactos destrutivos, pois estes causam perdas elevadas, como as chuvas intensas, vendavais e furacões, marés meteorológicas e grandes períodos de secas contínuos. Dourado (DOURADO et al., 2013) destaca os eventos extremos, no caso secas ou enchentes severas, como um fator de impacto geográfico que altera uma determinada região de forma física, movimentando massas, redistribuindo características da paisagem e gera transtornos sociais como, por exemplo, os casos de estiagem no sertão nordestino.

2.1.2 Impactos Climatológicos

Os impactos climáticos não são fáceis de se prever ou medir, pois os danos causados afetam todas as áreas de uma região. Os danos estão ligados a contextos socioeconômicos que vão desde prejuízos materiais até perdas de vidas, afetando diretamente a cultura e economia do local atingido. De acordo com IPCC os impactos projetados estão relacionados com os recursos hídricos e seu gerenciamento, ecossistemas, saúde, indústria e a sociedade. Sendo assim, eventos extremos podem causar perdas irreparáveis e devem ser estudados a ponto de serem previstos e evitados (UNFCCC, 2011).

Em se tratando de mudanças climáticas, o planejamento da Organização das Nações Unidas (ONU), referente aos Objetivos de Desenvolvimento Sustentável (ODS), propõe cinco metas sobre medidas imediatas para combater as mudanças climáticas (OLI-

VEIRA; SILVA, 2019). Uma das metas é integrar medidas de controle de mudanças climáticas nas políticas, estratégias e planejamentos nacionais. Com isso, propõe aos países que realizem um estudo em suas regiões e visem ao máximo minimizar as ações e impactos causados pelo clima.

No Brasil existe uma fragilidade em relação à previsão e tratamento de eventos climáticos extremos, os quais deveriam ter um alto nível de importância devido a dependência econômica dos recursos naturais que estão ligados diretamente com as condições climáticas. Segundo Barbosa et al. (2008), essa fragilidade se encontra na frequência e intensidade de enchentes e secas que causam perdas na agricultura e na biodiversidade, impactos no regime hidrológico, com altas elevações no nível do mar, entre outros inúmeros problemas socioeconômicos que são causados.

Chuvas intensas podem ocasionar o agravamento de problemas famosos no Brasil, como os alagamentos e escorregamentos de encostas. Além disso, grandes enchentes causam desastres, tanto biológicos quanto econômicos, os quais também podem ocasionar elevadas taxas de mortes. Outro extremo bastante comum é a seca, que pode tornar o fornecimento de água limitado ou esgotado por completo, fator agravante que causa impacto direto na sociedade podendo gerar migrações de pessoas. Por fim, pode ocasionar falhas e interrupção do fornecimento de eletricidade em regiões que possuem como fonte a água da chuva (RUIVO, 2012).

Já a seca causa grande impacto no setor agrícola, impedindo o desenvolvimento de culturas e matando milhares de plantações, como no caso relatado por MARENGO et al. (2009), que destacou como a seca afetou a região sul do Brasil e uma parte da Argentina, em meados de 2008, diminuindo a produção de soja e grãos, causando a queda no preço e no número de exportações, que tiveram uma redução de 8 a 9 bilhões de dólares em 2009.

Outro caso, relatado por Ruivo (2012), destaca como as chuvas fortes causaram inundações e deslizamentos afetando o estado de Santa Catarina no período de 22 a 24 de novembro de 2008. O evento afetou milhões de pessoas, deixando várias desabrigadas e um número considerável de mortes. Além disso, as fortes tempestades afetaram o fornecimento de energia de várias residências e interrompeu a circulação de veículos em estradas próximas à região.

De acordo com (RUIVO, 2012) existem fatores de anomalia na temperatura que agravam os eventos extremos, fazendo com que as temperaturas na superfície da água variem ocasionando chuvas fortes, que resultam em enchentes, ou chuvas fracas, podendo causar secas. Por exemplo, têm-se os eventos El Niño e La Niña, que ocorrem no Oceano Pacífico e afetam a disponibilidade de água de certas regiões, podendo também causar outros impactos socioeconômicos. Logo, deve-se prever o surgimento desses eventos, como descrito no trabalho (DOURADO et al., 2013), que relata que nos anos de 1998, 2001, 2002 e 2003 as temporadas de secas estavam relacionadas com o El Niño, destacando os períodos de estiagem no nordeste brasileiro.

2.1.3 Estado do Tocantins

A região do estado do Tocantins possui uma área total de 277.620,9 km^2 com relevo constituído basicamente de depressões, planaltos mais ao sul e no nordeste e planícies localizadas mais ao centro. Com uma altitude que não passa de 500 metros acima do nível do mar, os locais mais altos do estado têm predominância de serras.

O território do estado se situa em uma localidade englobada pela Amazônia Legal, a vegetação tem o cerrado como bioma predominante, ocupando cerca de 90% do total. O cerrado contém as características de grandes arbustos, árvores esparsas, galhos retorcidos e as raízes profundas, além de ser a vegetação na qual se situam a maioria das atividades agropecuárias do estado. Já o restante da região se encontra em uma área composta de vegetação amazônica, sendo esta situada ao norte do estado, podendo localizar também um certa constância de vegetações rasteiras e gramíneas.

2.1.3.1 Clima no Estado do Tocantins

O clima predominante no estado do Tocantins é o tropical, caracterizado basicamente por uma estação com chuva no verão e uma estação mais seca no inverno. Entretanto, com base no conhecimento da região, pode-se dividir o estado em duas metades, a primeira mais ao Norte, a qual possui uma região com cobertura de Floresta Pluvial Amazônica que influencia diretamente no clima, definindo-o como tropical úmido. A segunda metade, ao sul, tem o clima mais puro do estado, ou seja, tropical com um equilíbrio entre os períodos de chuva e seca.

Esse tipo de clima com um período seco e outro chuvoso representa um arranjo característico da região central brasileira. A estação quente (seca) é um período que se estende de início de abril até meados de outubro, com a precipitação girando em torno de 38% e as temperaturas médias variando de 19°C (mínima) até 31°C (máxima). A estação com maior nível de precipitação (chuvosa) se dá de meados de outubro até fim de março, possuindo uma precipitação de até 72%, podendo haver uma temperatura média de 14°C (mínima) até 27°C (máxima) (INMET, 2016).

O trabalho de (ROLDÃO; FERREIRA, 2019) realizou um estudo com os dados climáticos do Tocantins de 1985 até 2016 e constatou que as informações acima são verdadeiras, destacando que o estado possui índices de pluviosidade baixos e temperaturas elevadas nos meses de abril a outubro e o oposto nos meses restantes. Além disso, relata uma média pluviométrica anual de 1624,9 mm, com uma média de excedente hídrico de 687,5 mm.

De acordo com (ROLDÃO; FERREIRA, 2019) e com dados do Instituto Nacional de Meteorologia (INMET), o estado do Tocantins possui valores médios anuais de chuva variando de 1250 mm até 2050 mm, e os principais causadores de chuva são a Zona de Convergência Intertropical (ZCIT) e a Zona de Convergência do Atlântico Sul (ZCAS).

Ressalta-se que a ZCIT atua na região norte do Tocantins por ser afetada pelos sistemas de alta pressão do Oceano Atlântico e a ZCAS atua, na maioria das vezes, na região centro-sul do estado, podendo variar a distribuição de chuvas dependendo da variabilidade do seu eixo. Tanto o fenômeno El Niño quanto La Niña afetam as duas zonas de geração de chuvas fazendo com que se tenha períodos extremamente secos.

2.1.3.2 Conceitos e Impactos Econômicos no Estado do Tocantins

Em condições climáticas de seca a região está suscetível a longos períodos de estiagem, os quais de forma comum estão ligados com queimadas que atingem as serras e as florestas, causando impacto ambiental e afetando as condições de saúde e de vida dos cidadãos. Esses períodos geram perdas no setor primário e os danos podem ser maiores se os períodos de seca extrema estão ligados com veranicos, período de estiagem com calor intenso e umidade baixa. As atividades agrícolas são as principais impactadas devido a sua dependência de condições favoráveis.

A inclusão das atividades agropecuárias do Tocantins no mercado de produção brasileira está sendo feita de forma gradativa. O Produto Interno Bruto (PIB) do estado representa somente 0,5% de todo o Brasil (SEFAZ, 2019), porém a expansão comercial teve um crescimento que merece reconhecimento devido a grande quantidade de recursos exportados, sendo importante destacar o investimento na área de logística e mudanças no espaço urbano, com as áreas do cerrado sendo exploradas para fins comerciais.

Conforme (MIRANDA, 2016), o estado tem a agricultura, a área de serviços e o setor público como elementos que mantêm a maioria das pessoas ativas na economia local. Destes, o setor agropecuário se destaca como principal atividade econômica, correspondendo a 20% do PIB e, o restante é referente a área comercial, a indústria e a área de serviços. Logo, o destaque no setor rural demonstra que a economia local tem uma certa dependência das produções e recursos oriundos do setor agrícola.

A agropecuária possui dentro de seu conjunto uma extensa quantidade de atividades que dependem de condições climáticas favoráveis para se desenvolverem e se tornarem recursos utilizáveis. Pode-se destacar a temperatura, a pluviosidade, a umidade do solo e o nível de radiação solar como influenciadores em produções agrícolas Miranda (2016). Desta maneira, um período com potencial produtivo é fundamental para que se tenha uma colheita com muitos recursos. Por outro lado, eventos extremos podem provocar mudanças na colheita, pois modificam o clima, podendo causar o surgimento de pragas e doenças, as quais podem destruir as plantações (RODRIGUES; DINIZ, 2011).

Segundo (SILVA; ALMEIDA, 2007), que apresentaram dados da Secretaria da Agricultura, Pecuária e Aquicultura (SEAGRO), as principais atividades do Tocantins estão relacionadas com gado bovino, soja, arroz, fruticultura, piscicultura e milho. O trabalho ainda destacou que, em 2004, o estado era o nono em produção de gado bovino, sendo que no ano de 2009 foi registrado 63% de aumento no plantel, que corresponde a

uma das maiores taxas de crescimento do país (OGATA et al., 2009).

O último relatório da SEAGRO sobre as exportações agropecuárias demonstrou que o crescimento está sendo constante (SEAGRO, 2020). Segundo as informações, o número de exportações do primeiro trimestre de 2020 foi cerca de 34,4% mais elevado que o mesmo período de 2019. Os dados também demonstram que houve um superavit do período de janeiro a março de 2020, sendo este de 182 milhões de dólares. Os principais produtos exportados foram a soja (58%), a carne bovina (33%) e o milho (4,3%) e, o principal país consumidor é a China. Ressalta-se que é necessário se atentar para a situação atípica do início de 2020, pois um dos motivos que fizeram o número de exportações aumentarem foi a epidemia de Covid-19, que afetou os países asiáticos no começo do ano.

As safras agrícolas são de suma importância para a economia do estado, pois uma safra que possui rendimento alto gera mais valores do que uma que possui um rendimento abaixo do esperado e, geralmente, há uma influência climatológica por trás de uma safra com pouca produção. O gráfico apresentado na Figura 1 evidencia as safras de 2001 a 2020, além de relacionar os períodos que foram afetados por mudanças climáticas consideráveis, como na safra de 2005/06 que sofreu com períodos de secas. O trabalho (BARBOSA et al., 2016) evidencia que na safra de 2014/2015 foi plantada uma área de 849,6 mil hectares, com retorno de 2.475,7 mil toneladas. Mesmo a área plantada ter sido a maior dos últimos anos, a produção caiu cerca de 1.685,4 mil toneladas se comparada com os anos anteriores devido a redução das precipitações em decorrência do fenômeno El Niño que ocorreu no período da safra.

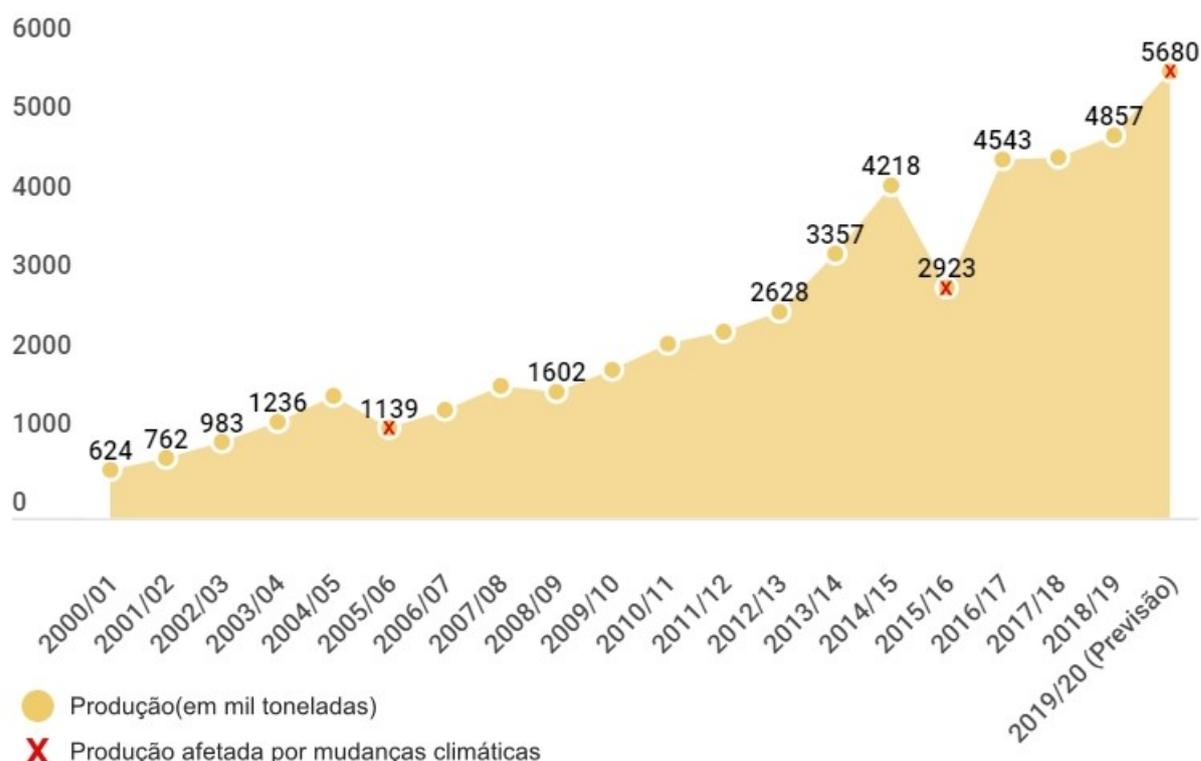


Figura 1 – Safras da agricultura no estado do Tocantins (SEFAZ, 2020).

Produções agrícolas como no ano de 2019/2020 acabam sendo atingidas por longos períodos de estiagem, neste caso cerca de 128 dias sem chuvas. Porém, por meio de estudos climáticos foi possível antecipar o período de seca e postegar a safra (CANAL RURAL, 2019). Portanto, demonstra que o estudo de mudanças climáticas é importante para minimizar as perdas na agropecuária.

2.2 Processo de Descoberta de Conhecimento em Bases de Dados

A Descoberta de Conhecimento em Bases de Dados (*Knowledge Discovery in Databases* - KDD) reúne um conjunto de técnicas e tarefas para que se possa descobrir informações importantes e relevantes em uma coleção de dados, sendo muito utilizado na tomada de decisão. De acordo com Fayyad (1996) “O KDD é o processo não trivial de identificar padrões válidos, novos, potencialmente úteis a partir de grandes conjuntos de dados”. Destaca-se que ao afirmar que o processo não é trivial está indicando a execução de diversos passos para que se possa, ao final do processo, obter o resultado esperado.

O processo de KDD surgiu para solucionar um problema da era da informação, que consiste em se ter uma quantidade massiva de dados e não conseguir identificar parte das informações contidas nesta base de dados pelo fato do processo manual ser demorado e não apresentar o resultado esperado (FAYYAD, 1996). Desta maneira, define-se a descoberta de conhecimento como um processo que visa analisar uma vasta quantidade de dados buscando identificar padrões e regularidades, muitas vezes ocultos, que possuam potencial de utilidade e possam ser de fácil compreensão.

Segundo Bucene et al. (2008) o processo de Descoberta de Conhecimento possui etapas muito complexas que devem ser executadas com muito rigor, já que cada etapa é descrita como fundamental e, caso uma fase seja mal executada ou não atinja o resultado esperado, irá afetar a etapa seguinte, fazendo com que o produto final não seja o desejado ou, de forma contrária, caso execute cada etapa da maneira correta, deverá alcançar os objetivos da aplicação. Desta maneira, o resultado de cada iteração deve ser analisado antes de avançar para a ação seguinte, para que se possa criar uma métrica de avaliação para verificar se o resultado de cada etapa está guiando o caminho correto. Assim, deve-se desenvolver uma ligação entre as etapas para caminhar entre as iterações, aprimorando-as ou refazendo a anterior com novas medidas. A Figura 2 apresenta as etapas do processo do KDD.

O processo de KDD está dividido em três etapas principais, nas quais estão contidos os procedimentos, tecnologias, técnicas e tarefas. Entretanto, antes de dar início às etapas do KDD, deve-se compreender o domínio no qual será aplicado o processo, ou seja, deve-se identificar o problema e quais objetivos deverão ser alcançados, do ponto de vista do usuário e, com isso, selecionar os tipos de dados que serão utilizados pelas etapas.

A etapa inicial do KDD corresponde ao pré-processamento, que consiste na reali-

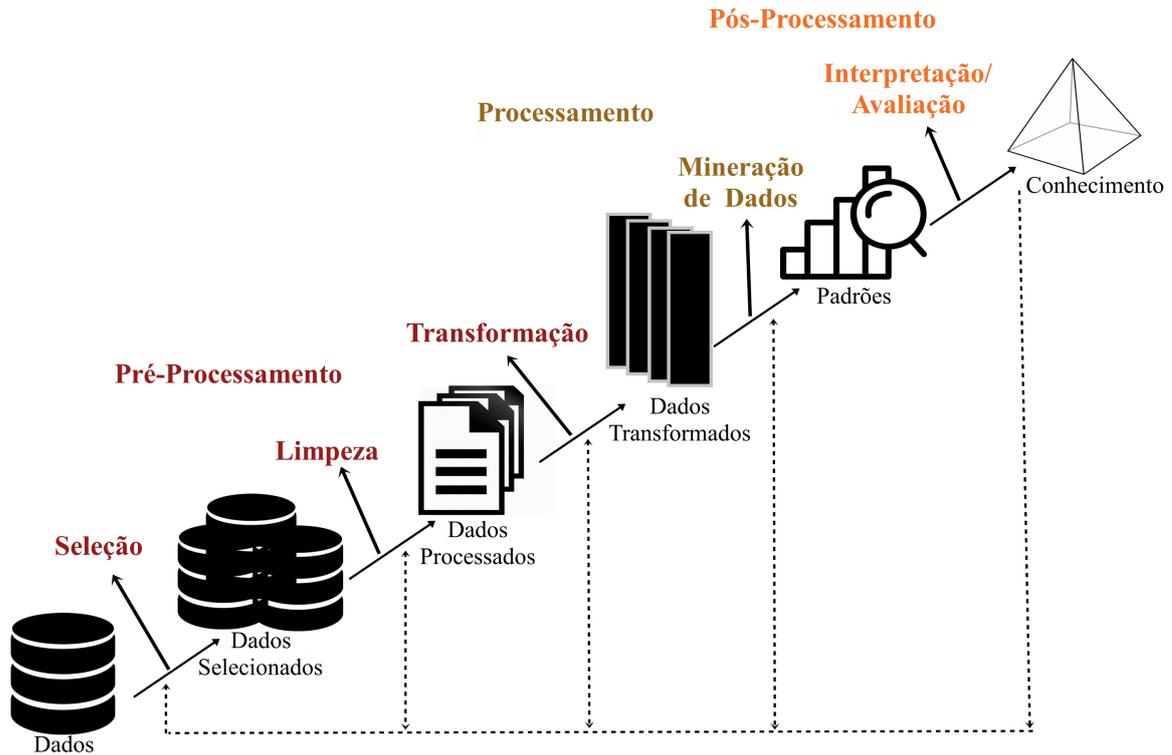


Figura 2 – Etapas do processo de KDD. Figura adaptada de Fayyad (1996).

zação de limpeza de ruídos, tratamento de inconsistências e transformação dos dados. Em seguida, os dados são encaminhados para etapa de processamento, conhecida comumente como a fase de mineração de dados, na qual serão aplicadas as tarefas e técnicas escolhidas para tratar os dados, resultando em informações que são utilizadas para criação de modelos de visualização de dados. Por fim, na etapa de pós-processamento deverão ser analisadas e interpretadas as informações que foram mineradas.

2.2.1 Modelos de Descoberta de Conhecimento

O KDD consiste em um processo que possui muitas tarefas e etapas cruciais para atingir os objetivos, que podem ser executadas de diversas maneiras e, caso sejam executadas de maneira incorreta, podem resultar em uma falha no projeto final. Alguns autores propuseram adaptações ao KDD com modelos que tinham o foco de padronizar o ciclo da mineração de dados como, por exemplo, o modelo *Cross-Industry Standard Process for Data Mining* (Processo Padrão da Indústria Cruzada para Mineração de Dados - CRISP-DM) (CHAPMAN et al., 2000), desenvolvido em 1996, presente em uma quantidade elevada de trabalhos, sendo considerado referência na área.

O CRISP-DM é um modelo que estrutura um ciclo de vida de um processo de Descoberta de Conhecimento, contendo seis etapas organizadas de forma cíclica. Ressalta-se que mesmo sendo uma sequência em forma de círculo, existe uma certa flexibilidade entre as etapas, logo, pode-se ir e voltar entre elas. A Figura 3 apresenta as seis etapas

do Crisp-DM: compreensão do domínio, entendimento dos dados, preparação dos dados, modelagem, avaliação e distribuição, que são detalhadas nas subseções abaixo.

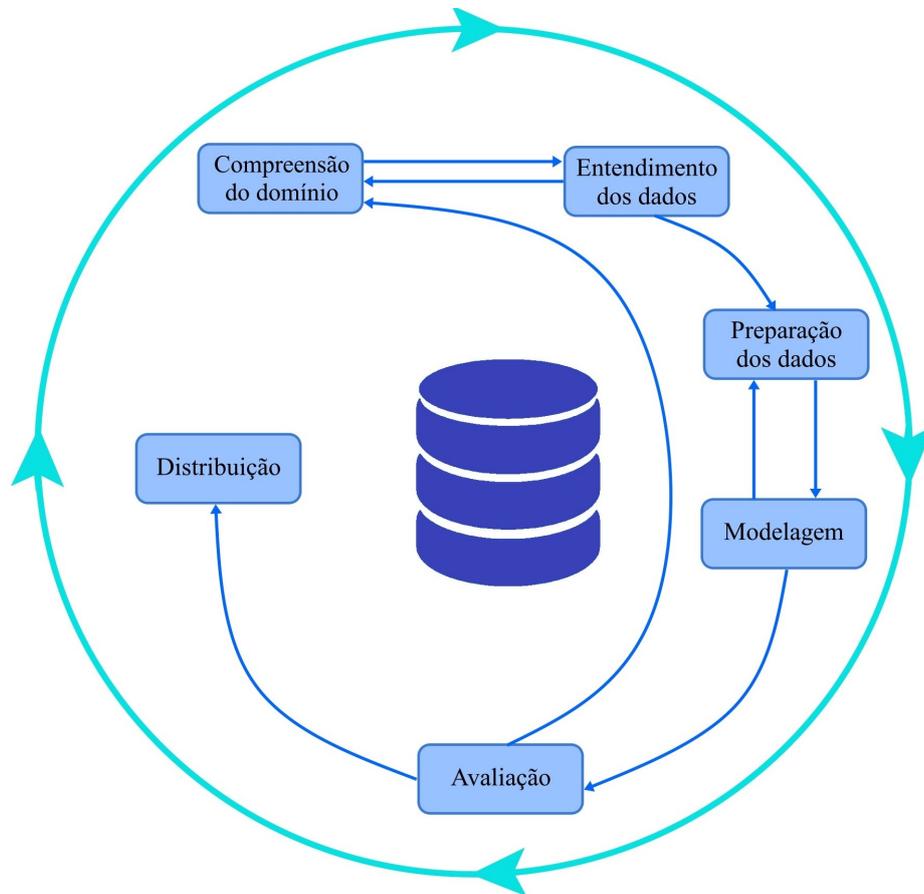


Figura 3 – Etapas do modelo CRISP-DM. Figura adaptada de Chapman et al. (2000).

2.2.1.1 Compreensão do domínio

Esta etapa tem como objetivo identificar e compreender o domínio, por meio da identificação dos requisitos e dos principais objetivos da aplicação. Deve-se extrair o máximo de conhecimento e depois identificar as metas a serem alcançadas. Desta maneira, é possível traduzir o problema de forma que seja possível traçar um planejamento sobre as etapas futuras e atribuir as tarefas que serão aplicadas na fase de mineração dos dados.

2.2.1.2 Entendimento dos dados

Nesta etapa, os dados contidos nas bases de dados são coletados e selecionados para iniciar o processo de exploração dos dados, ou seja, a fase de compreensão e familiarização dos dados visando uma descrição clara do problema, identificação de informações importantes, análise da qualidade dos dados, criação de hipóteses iniciais e, de forma eventual, identificação de quais conjuntos podem ter informações ocultas.

2.2.1.3 Preparação dos dados

Etapa fundamental para a eliminação e tratamento de inconsistência de dados. A fase inicia com os dados brutos pré-selecionados, que passam por tarefas de preparação, podendo haver uma análise prévia dos dados para detectar *outliers* e desvios (CÔRTEZ; PORCARO; LIFSCHITZ, 2002). As principais tarefas são referentes ao tratamento da qualidade dos dados, que possuem ações como a seleção, limpeza e transformação dos dados.

- **Seleção:** consiste em selecionar quais conjuntos de dados são importantes para produzir informações relevantes, que devem ser utilizados para a análise.
- **Limpeza:** fundamental para a eliminação de inconsistências, ausência de informação e correção de dados incorretos para garantir a qualidade nos modelos.
- **Transformação:** com os dados selecionados e limpos, as informações precisam estar armazenadas e em formatos adequados para que os algoritmos da etapa seguinte possam ser aplicados.

2.2.1.4 Modelagem

Nesta fase serão selecionadas e aplicadas tarefas e técnicas de mineração de dados com os parâmetros calibrados para os valores ideais. A escolha das técnicas depende dos objetivos que foram traçados. Por permitir a aplicação de diversas técnicas de mineração de dados, algumas vezes, pode ser necessário retornar a etapa de preparação de dados para adaptá-los para que a técnica possa ser aplicada.

Para realizar a mineração de dados são utilizadas técnicas de inteligência artificial para localizar similaridades ou divergências nos dados. O objetivo é encontrar padrões, dados irregulares e regras para gerar informações úteis, que antes se encontravam ocultas, para serem avaliadas na próxima fase.

2.2.1.5 Avaliação

Etapa correlacionada com o pós-processamento, na qual os resultados obtidos na fase anterior são avaliados considerando o modelo do problema. Dependendo da avaliação das informações, podem ocorrer casos que o resultado não foi satisfatório, sendo necessário iniciar o processo de avaliação das etapas anteriores, refazendo-as, com o intuito de gerar um resultado que atenda aos objetivos.

Ressalta-se que nesta fase deve-se aplicar técnicas para o tratamento do conhecimento que foi adquirido na fase de mineração de dados. O objetivo é tornar o conhecimento mais fácil de interpretar e avaliar por um usuário.

2.2.1.6 Distribuição

Na fase final deve-se ter um modelo construído, o qual foi avaliado e aprovado. Então, o conhecimento produzido deve ser organizado e apresentado para os usuários finais, de forma que eles possam utilizar o modelo para o processo de tomada de decisão.

2.2.2 Tarefas e Técnicas de Mineração de Dados

A mineração de dados consiste em um conjunto de tarefas e cada uma delas possui um grupo de técnicas e algoritmos a serem aplicados em uma base de dados, de modo que possa detectar padrões relevantes que, posteriormente, serão transformados em conhecimento. Basicamente, a mineração de dados é uma técnica de inteligência computacional que analisa uma quantidade massiva de dados e, por meio das metas estabelecidas e aplicadas, apresentará, ao final do processo, informações que se encontravam ocultas.

Segundo Bucene et al. (2008), na mineração de dados deve-se destacar a diferença entre tarefa e técnica. Basicamente, a tarefa é a especificação do que se deseja buscar nos dados, quais padrões devem ser encontrados ou quais dados poderiam gerar informações novas. Já a técnica de mineração consiste na especificação de métodos que deverão garantir como encontrar os padrões das tarefas selecionadas para a aplicação.

As tarefas de mineração de dados estão divididas em dois objetivos fundamentais: predição e descrição. Estes também se dividem em outras categorias principais, descritas como: classificação, regressão, associação e clusterização Bucene et al. (2008). A Figura 4 apresenta a representação das principais tarefas de mineração de dados, que serão melhor detalhadas nas subseções seguintes.

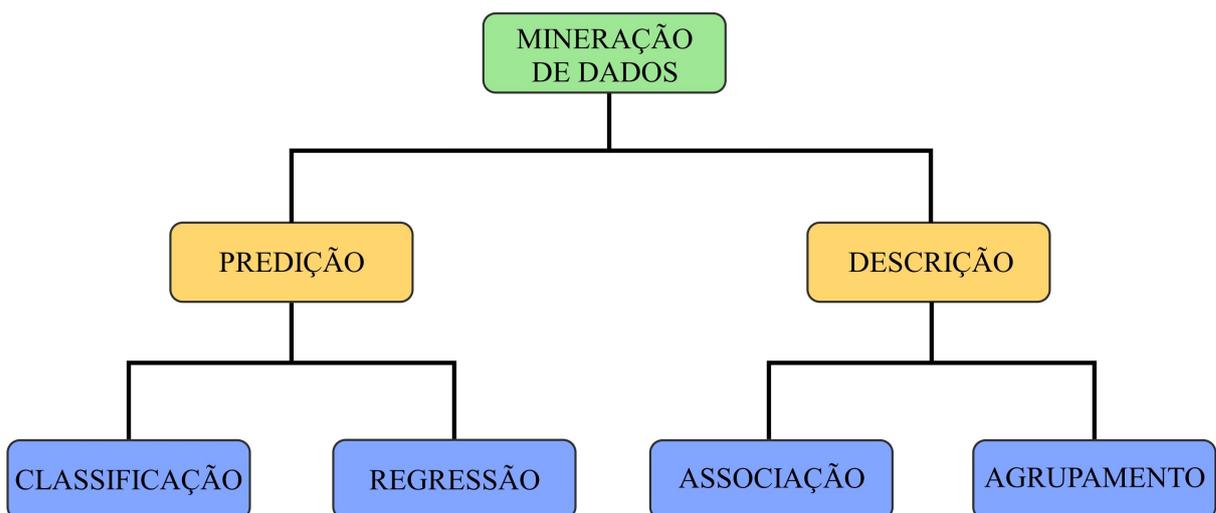


Figura 4 – Principais tarefas de mineração de dados.

2.2.2.1 Predição

A predição tem como principal característica analisar os atributos de um determinado conjunto de entidades e, por meio dos valores contidos nos atributos, indicar valores para uma nova entidade que deverá ser caracterizada. Resumindo, é um processo que pode prever uma ação futura baseado nos valores dos atributos avaliados. Seus principais subtipos são classificação e regressão.

- **Classificação:** é um processo que visa encontrar um modelo que irá descrever e diferenciar classes, as quais serão utilizadas na fase final para prever objetos que ainda não foram classificados. Esse tipo de classificação irá prever valores categóricos. Entre as técnicas de classificação destacam-se as: Redes Neurais Artificiais (*Artificial Neural Networks*) e Árvores de Decisão (*Decision Trees*).
- **Regressão:** abrange métodos que procuram prever um valor futuro de uma variável selecionada. Esta variável deve ser do tipo numérico, já que as previsões são realizadas somente com valores contínuos. Destacam-se algumas técnicas de regressão como: Regressão Linear (*Linear Regression*) e Regressão não Linear (*Nonlinear regression*).

2.2.2.2 Descrição

A descrição consiste em descrever padrões e tendências encontrados na análise de informações, aumentando o conhecimento sobre os objetos que geraram esses dados. Desta maneira, indica uma interpretação e identificação de comportamentos sobre os conjuntos de dados, mesmo que estes, por vez, não estivessem implícitos. Seus principais subtipos são associação e agrupamento.

- **Associação:** comumente conhecida como descoberta de regras de associação, consiste na tarefa de identificar relacionamentos entre atributos de uma coleção de dados. Assim, a tarefa pode indicar as preferências, afinidades e frequências com que uma regra é encontrada em uma base de dados e, também, a quantidade de acertos da regra. Sua forma padrão é apresentada da seguinte maneira SE (determinado atributo) X ENTÃO (determinado atributo) Y. As seguintes técnicas de associação apresentam maiores destaques: Análise da Cesta de Compras (*Market Basket Analysis*) e Mineração de Itens Frequentes (*Frequent Itemset Mining*).
- **Agrupamento :** esta tarefa tem como principal função identificar e aproximar atributos com valores semelhantes. Também conhecida como clusterização, basicamente divide um conjunto de dados em subgrupos, de acordo com o grau de similaridade, no qual cada subgrupo contém indivíduos com dados semelhantes. Ao final, deve-se indicar o melhor representante de cada conjunto para poder identificar

as características daquele subgrupo (CAMILO; SILVA, 2009). Destacam-se os métodos de agrupamento: K-means (*K-means clustering*) e Agrupamento Hierárquico (*Hierarchical clustering*).

2.2.3 Descrição das Técnicas de Mineração de Dados

A pesquisa adotou as tarefas de mineração de dados: Classificação e Clusterização (Agrupamento) para serem aplicadas sobre os dados. As próximas subseções iram complementar as informações sobre as tarefas e maior detalhamento das técnicas pertencentes a cada método.

2.2.3.1 Classificação

A classificação consiste, basicamente, em uma técnica de predição de valores categóricos, sendo necessário, para a definição de modelos e regras, um conjunto de exemplos já classificados anteriormente, pois com base nestes que os próximos exemplos serão classificados. Abaixo, descreve-se o funcionamento das principais técnicas:

- **Árvores de Decisão (*Decision Trees*):** é uma técnica de indução utilizada para descobrir as regras de classificação para um atributo a partir de uma subdivisão dos dados que estão contidos no conjunto de informações analisadas. Esta técnica utiliza como base a estratégia dividir para conquistar, que consiste em um problema complexo ser dividido em subproblemas menores e mais simples, sendo mais fácil verificar o problema. Com a árvore de decisão montada é possível classificar um novo registro de forma simples, analisando o fluxo da árvore a partir do nó raiz até chegar a um nó folha. As principais vantagens desta técnica são a facilidade de entendimento do modelo, a prontidão no segmento das regras de indução para classificação de um novo registro e um grau elevado de assertividade (BUCENE et al., 2008).
- **Redes Neurais Artificiais (*Artificial Neural Networks*):** são modelos computacionais com inspiração na psicologia e na neurobiologia, focando no sistema nervoso biológico, com o intuito de simular o comportamento dos neurônios. Uma rede neural possui um conjunto de unidades de processamento, que se encontram conectados por canais comunicadores, os quais estão ligados a pesos. Normalmente, uma rede neural é composta por uma camada de entrada, que irá receber os dados da entrada, camadas intermediárias, que deverão realizar o processamento, e uma camada de saída, responsável por apresentar o resultado do processo da rede neural. As redes em seu momento de aprendizagem devem ir ajustando os pesos para que se possa ter uma classificação de um objeto com maior precisão. É uma técnica que exige um período de treinamento longo e sua interpretação é tida como complexa.

Por outro lado, geralmente conseguem identificar padrões que ainda não tinham sido treinados (CAMILO; SILVA, 2009).

2.2.3.2 Clusterização

A Clusterização é um processo de agrupamento de dados em classes homogêneas. Diferente da classificação, essa técnica não aprende através de exemplos pré-definidos, ao invés disso, os objetos serão agrupados com base na semelhança dos seus atributos. Desta maneira, a clusterização é capaz de encontrar novos agrupamentos, que apresentem características semelhantes e agrupá-los. Abaixo, destacam-se as principais técnicas de agrupamento:

- **K-means:** é um método de particionamento que, dado um conjunto de dados com certo número de registros e o número de agrupamentos desejados (k), deverá particionar os objetos em k agrupamentos. Este método utiliza centróides para realizar os agrupamentos e, geralmente, o número de k centróides é definido de forma aleatória. Seu funcionamento é baseado no cálculo da similaridade entre cada registro e o centróide de todos os agrupamentos. Com isso, o registro é anexado ao agrupamento com maior similaridade. Por fim, o centróide de cada grupo é recalculado após a inclusão de todos os registros em um dado grupo.
- **Métodos Hierárquicos:** o procedimento padrão é organizar um conjunto de dados de forma que se tenha uma estrutura hierárquica baseada no grau de similaridade entre os indivíduos. O agrupamento pode ser feito por meio de métodos como aglomeração ou divisão. Geralmente, o resultado é apresentado como um dendograma, ou seja, uma espécie de gráfico em forma de árvore que divide o grupo de dados em subconjuntos menores, cuja raiz representa o conjunto de dados completo e os seus nós folhas são os objetos.

2.3 Aplicação do processo de Descoberta de Conhecimento na climatologia

A climatologia é um segmento amplo que possui muita variabilidade em suas informações. Entretanto, entender suas alterações e conseguir prever suas mudanças são de suma importância para uma região, pois afeta diretamente a sua socioeconomia. As mudanças climáticas, que em alguns casos podem ocasionar eventos extremos, produzem uma quantidade elevada de conjuntos de dados a cada dia e esses dados contém informações cada vez mais específicas. Logo, técnicas mais sofisticadas e precisas foram desenvolvidas para suprir a demanda de cálculos manuais, além de ferramentas que conseguem transformar a quantidade massiva e incontável de dados meteorológicos em conhecimento para tomada de decisão. Entre os trabalhos desenvolvidos na área de mineração de dados climáticos destacam-se algumas importantes contribuições.

Dourado et al. (2013) apresentaram um trabalho em que analisam as zonas pluviométricas homogêneas no estado da Bahia, utilizando dados mensais de 97 estações meteorológicas entre um período que varia entre 1981 a 2010 (30 anos). Foram aplicadas técnicas de mineração de dados baseadas em agrupamento, utilizando o algoritmo *k-means*, com o intuito de encontrar informações que pudessem relacionar alterações climáticas com a agricultura do estado. Ao segmentar os dados em escala mensal foi identificado cinco zonas homogêneas de similaridade pluvial na área de estudo, nomeadas de *cluster* com identificação de 0 ao 4. Em seguida, utilizaram técnicas de predição para classificar os totais de chuva anuais, conseguindo localizar alterações climáticas de falta e excesso de chuva em certas épocas do ano e correlacioná-las com a baixa produção agrícola da região.

O trabalho de Charles (2015) desenvolveu um modelo de classificação para a previsão de eventos extremos de precipitação na cidade de Nova Friburgo, no estado do Rio de Janeiro, utilizando algoritmos de Redes Neurais Artificiais e Árvores de Decisão para extrair padrões de acúmulo de precipitação acima de 10mm anterior a uma hora, por meio de medições das variáveis climáticas. Este trabalho utilizou, como treinamento, os dados de 2008 a 2012 (5 anos) e, para aperfeiçoar a aplicação usou dados históricos do Modelo ETA, criado para previsão de fenômenos meteorológicos na América do Sul com maiores detalhes e até 48 horas antes da ocorrência do fenômeno. Logo, ao cruzar a base do Modelo ETA com a base de dados de 5 anos encontrou-se a solução para o problema de falsos positivos. O resultado é um modelo que consegue realizar a previsão de uma chuva maior que 10mm com até uma hora de antecedência.

A pesquisa de Bucene et al. (2008) visava encontrar um padrão de comportamento para previsão de geada e deficiência hídrica no estado de São Paulo utilizando variáveis climáticas, com o intuito de auxiliar nos subsídios da agricultura da região, que estava tendo problemas com as variações bruscas de eventos climáticos. O modelo resultante do estudo consegue prever geada com até um dia de antecedência com resultados considerados confiáveis. Já para o caso de deficiência hídrica, o melhor resultado era referente há 3 dias antes, mas variava de acordo com o mês.

O trabalho de Boschi et al. (2010) buscou identificar zonas pluviométricas homogêneas no estado do Rio Grande do Sul para análise dos fenômenos climáticos veranicos na região. A análise foi dividida em dois decênios, o primeiro de 1987 – 1996 e o segundo de 1997 – 2006 e, depois se definiu as áreas pluviométricas homogêneas para cada decênio. Foram utilizadas técnicas de agrupamento, localizando seis zonas comuns aos decênios, as quais foram denominadas de A-F. Então, realizou-se uma análise das alterações de precipitação anual, sazonal e mensal das seis zonas, além das alterações na média de veranicos anuais. Os resultados mostraram que em determinados meses do ano tem uma redução de precipitação pluvial em algumas zonas e, a análise anual de veranico notou uma redução no número de veranicos maiores que 15 dias e um aumento nos menores que 15 dias.

Por fim, o trabalho de Kohail e El-Halees (2011) buscou extrair o máximo de

conhecimento de dados históricos diários da cidade de Gaza Strip, entre os anos de 1977-1985 (9 anos), aplicando tarefas de agrupamento, predição, classificação e associação. Na análise de agrupamento utilizou-se a técnica *k-means* sobre as informações das estações, gerando quatro grupos com informações sobre chuva, temperatura, umidade e velocidade do vento. A predição utilizou técnicas de Rede Neurais Artificiais e o método de Regressão Linear, que geraram modelos de previsão que podem ser usados para prever a temperatura diária para a cidade. A tarefa de classificação aplicou as técnicas *Naive Bayes*, *KNN*, *Árvore de Decisão* e *Redes Neurais* para classificar os registros de dados de três classes de temperatura e o algoritmo que apresentou melhores resultados foi o de *Redes Neurais Artificiais* com 85.77% de precisão. Ao aplicar associação, foram geradas doze regras de associação, das quais cinco poderiam ser utilizadas para prever chuva, além de regras que auxiliam na compreensão do clima da cidade.

3 MATERIAIS E MÉTODOS

O processo de Descoberta de Conhecimento em Bases de Dados (KDD) destaca a importância de compreender o ambiente de estudo para que, dessa forma, possa modelar o problema no qual serão aplicadas as técnicas de mineração de dados. Neste trabalho, a área de estudo compreende o estado do Tocantins, que se encontra entre as longitudes 45°W e 51°W e latitudes 5°S e 14°S. Além disso, os dados climáticos das séries históricas foram obtidos da base de dados do Instituto Nacional de Meteorologia INMET (2016) e compreendem um período de 59 anos (1961 a 2019). A escolha deste período se deu por ser a maior quantidade de tempo possível que pode ser filtrado na base de dados. As seções seguintes apresentam a base de dados e o processo de KDD aplicado.

3.1 Materiais

A base de dados utilizada está dividida em dois grupos, de acordo com o tipo da estação, que pode ser Convencional e Telemétrica. A Figura 5 apresenta a distribuição espacial das 14 estações meteorológicas selecionadas no Tocantins e a Tabela 1 mostra a relação das cidades com o tipo da estação implantada. Ressalta-se que as estações meteorológicas convencionais são dotadas de inúmeros instrumentos meteorológicos para captação de dados e as estações Telemétricas possuem apenas uma torre, com vários sensores automáticos de captação de informações meteorológicas.

Tabela 1 – Relação das cidades com o tipo de estação.

ESTAÇÃO	TIPO
Araguaína	Convencional
Palmas	Convencional
Pedro Afonso	Convencional
Peixe	Convencional
Porto Nacional	Convencional
Taguatinga	Convencional
Araguatins	Telemétrica
Campos Lindos	Telemétrica
Dianópolis	Telemétrica
Formoso do Araguaia	Telemétrica
Gurupi	Telemétrica
Marianópolis do Tocantins	Telemétrica
Mateiros	Telemétrica
Paraná	Telemétrica

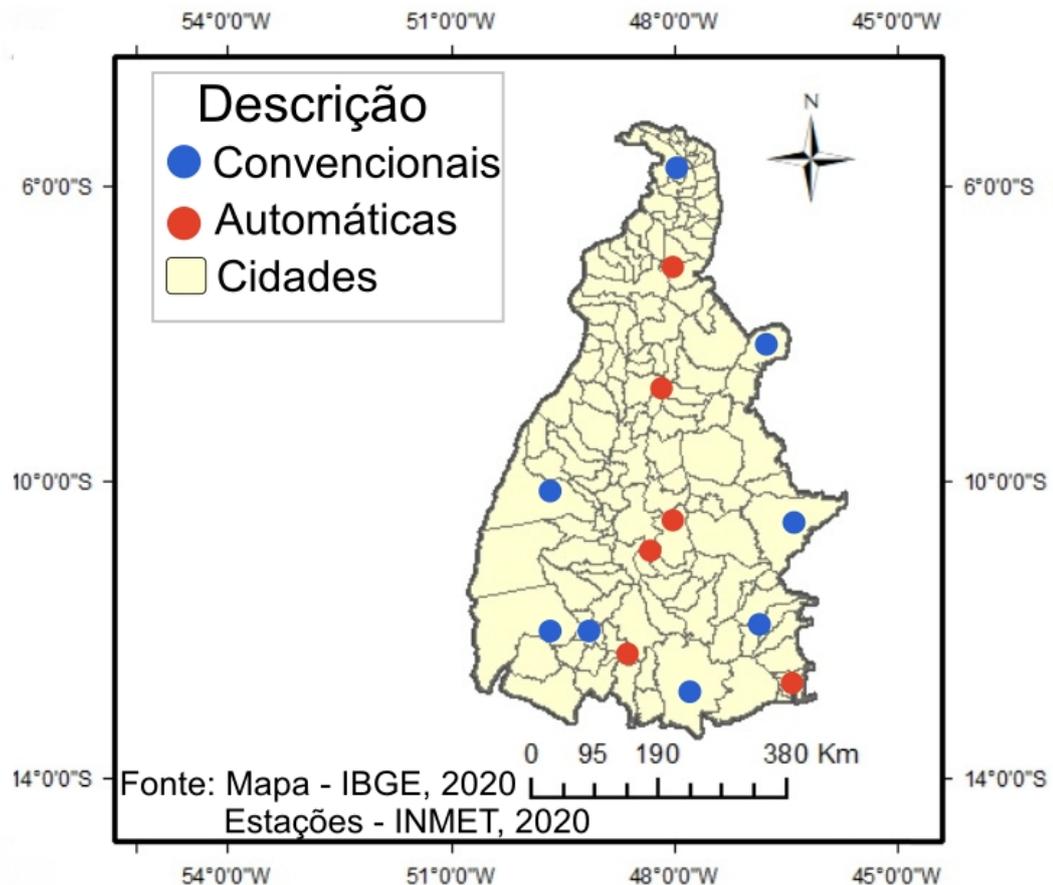


Figura 5 – Distribuição espacial das 14 estações meteorológicas selecionadas no estado Tocantins.

Os dois tipos de estações possuem atributos similares entre si, logo, foi realizada uma seleção de atributos em comum a serem trabalhados na pesquisa. As informações contidas em cada base de dados eram descritas na forma de registro diário, contendo informações diárias referentes aos dados pluviométricos, além da identificação da estação e data da coleta, conforme apresentado na Tabela 2.

Tabela 2 – Atributos referentes às bases de dados das estações.

ATRIBUTOS	
ID da Estação	Data
Precipitação (mm)	Temperatura Máxima (°C)
Temperatura Compensada Média (°C)	Temperatura Mínima (°C)
Umidade Relativa Média (%)	Velocidade do Vento Média (mps)

3.2 Métodos

A metodologia escolhida para aplicação do processo de descoberta de conhecimento em bases de dados foi a CRISP-DM (*Cross Industry Standard Process for Data Mining*)

proposta por (CHAPMAN et al., 2000). A escolha desta metodologia se deu devido à sua ampla utilização em diversos trabalhos de mineração de dados, alguns relacionados a análise climatológica, além de ser uma ferramenta completa de fácil replicação com todos os procedimentos definidos nas fases do processo. Cada etapa do processo foi definida e dividida da seguinte maneira: Fase 1 — entendimento dos dados; Fase 2 — preparação dos dados; Fase 3 — modelagem; Fase 4 — análise dos resultados. A seguir, detalha-se cada uma destas fases.

3.2.1 Entendimento dos dados

Os dados a serem trabalhados foram obtidos da base de dados do sistema de informação do Instituto Nacional de Meteorologia (<https://portal.inmet.gov.br/>) (INMET, 2016). Foram selecionados e recuperados dados pluviométricos diários referente às 14 principais postos meteorológicos de medição, que se encontram ao redor do estado, localizados nas cidades de Araguaína, Araguatins, Campos Lindos, Dianópolis, Formoso do Araguaia, Gurupi, Marianópolis, Mateiros, Palmas, Paranã, Pedro Afonso, Peixe, Porto Nacional e Taguatinga.

Uma escolha da pesquisa foi trabalhar com séries de dados homogêneas e contínuas sobre o maior período possível. Logo, foram selecionadas apenas estações que possuíam disponibilidade e qualidade em seus dados. O filtro destes dados se deu do período de 1961 a 2019, considerando as estações telemétricas e convencionais do estado. O conjunto de séries históricas escolhidas e filtradas do sistema do INMET se encontravam em um formato de CSV, onde cada arquivo representava uma estação meteorológica. As informações poderiam ser filtradas pelo tipo de estação.

3.2.2 Preparação dos dados

Nesta etapa, inicialmente, procurou-se eliminar e tratar as inconsistências das bases de dados, utilizando técnicas de limpeza e tratamento de dados. A limpeza tem como objetivo avaliar os dados e excluir aqueles que têm algum problema, como dados inseridos errados, duplicados ou em conflito. Logo, foram deletados registros que se encontravam duplicados ou células que apresentavam informações incoerentes com o atributo.

Com o intuito de tratar os campos que se encontravam em nulidade, foram aplicadas técnicas de tratamento visando tornar o banco de dados consistente. Foram utilizados métodos de atribuição múltipla, como substituição pela média e atribuição por regressão. Ao atribuir pela média, os campos nulos recebem um valor médio entre o atributo anterior e o atributo posterior, sendo atributos meteorológicos com dados diários, o valor médio atribuído se encontra em uma escala de relevância aceitável para o registro. Entretanto, foram detectadas algumas sequências de células vazias, nas quais não era possível aplicar a técnica de média, desta maneira, aplicou-se a atribuição por regressão, uma técnica

usada para prever valores nulos de um atributo com base em sua relação com os seus outros atributos. O algoritmo de regressão que apresentou melhor resultado, ou seja, menor quantidade de erros para o modelo treinado para preenchimento de dados nulos foi a Árvore de Regressão.

Após eliminar as inconsistências nas bases de dados, o próximo passo foi agrupar os arquivos das estações e as informações contidas em um arquivo apenas. Desta maneira, agrupou-se os dados em um conjunto, formando uma planilha que contém um total de 193.397 linhas (registros) e 8 colunas(atributos). A partir da planilha que continha os dados históricos em escala diária, foi possível construir um outro arquivo com os valores e atributos acumulados em uma escala mensal.

Um detalhe importante sobre a transformação da escala diária para mensal, está foi baseado na ideia apresentada no estudo de (DOURADO et al., 2013), que realizou testes de clusterização em escala diária e mensal, ao final conclui que o uso da escala mensal apresentou melhor desempenho, devido às particularidades referentes à alta variabilidade dos atributos meteorológicos. Pois, a técnica de agrupamento procura padrões com uma maior similaridade entre si, para que ao final sejam formados *clusters*.

A planilha de dados mensais foi dividida em dois conjuntos de séries históricas diferentes, considerando períodos de 20 anos. A escolha desse período foi motivada pela quantidade de anos restantes após a limpeza e a quantidade de registros de cada estação. O primeiro conjunto é formado por dados referentes aos anos de 1980 a 1999, contendo as 6 estações convencionais, já o segundo é do ano de 2000 a 2019 e contém todas as 14 estações. Em ambos os conjuntos, os atributos utilizados foram os mesmos, somente a seleção de estações foi distinta, pois as estações telemétricas não possuíam dados referentes ao período anterior ao ano 2000. Somente os dados meteorológicos foram utilizados para a formação dos *clusters*, não utilizando informações sobre localização geográfica, data ou tipo de estação.

De forma geral, técnicas de clusterização pressupõem a utilização de normalização nos dados em atributos com diferentes ordens de grandeza. Entretanto, para o algoritmo de clusterização, foi aplicada uma técnica de Análise de Componentes Principais (*PCA*), que consiste basicamente em um procedimento matemático de transformação ortogonal para conversão de um conjunto de atributos em componentes principais, menores ou iguais que as variáveis originais. Logo, a aplicação de normalização ou não nos dados, apresentou resultados similares em ambos os casos.

3.2.3 Modelagem

A ferramenta utilizada para extração de informações e aplicação de técnicas de inteligência computacional foi a biblioteca de aprendizado de máquina *Scikit-learn* (*SCIKIT-LEARN*...), criada em linguagem Python, suas características são de uma ferramenta simples, cheia de recursos e eficiente, além do que, foi desenvolvida para ser utilizada em

problemas de KDD, possuindo um arsenal vasto de algoritmos para a área de mineração de dados.

A técnica determinada para a seleção de zonas similares foi a clusterização por meio do algoritmo *k-means*. O algoritmo consiste basicamente em uma técnica de agrupamento de dados definidos por uma medida de distância euclidiana. O processo central da divisão é determinado pela escolha de *k-clusters*, onde cada elemento irá ser selecionado pela média com o *cluster* mais próximo.

Os *clusters* foram visualizados através das ferramentas de plotagem e espacialização do próprio *Scikit-learn*, para uma melhor análise e apresentação das informações. Os atributos que apresentaram melhor desempenho na aplicação da técnica de clusterização foram os que estão correlacionados com a ocorrência de chuvas, ou seja, Precipitação(mm), Temperatura Compensada Média(°C), Temperatura Mínima(°C), Temperatura Máxima(°C), Umidade Relativa do Ar Média(%). Desta maneira, essas cinco variáveis foram escolhidas para gerarem os *clusters* a serem analisados.

3.2.4 Análise dos Resultados

A avaliação do número de *clusters* foi determinada com base nos resultados obtidos para as duas séries trabalhadas de 1980 a 1999 e de 2000 a 2019, por meio de uma análise de mapas, que contém as informações das estações e das zonas de similaridades determinadas.

A espacialização das cidades e *clusters* foi realizada através da ferramenta ArcMap(ESRI, 2020) versão 10.6.1, uma aplicação de Sistema de Informação Geográficas (SIG) vetorial, onde é possível a referenciação geográfica e aplicação de diversas técnicas de análise de mapas. O sistema auxiliou na criação de mapas referentes ao estado do Tocantins, relacionando-os com suas variáveis climáticas, resultando em uma eficaz visualização das zonas homogêneas.

As informações geográficas acerca do estado foram retiradas diretamente do site do Instituto Brasileiro de Geografia e Estatística (IBGE) (IBGE, 2020). Os dados se encontravam em formato *shapefile*, que prontamente foram reconhecidos pelo programa ArcMap, permitindo a geração do mapa da região, contendo os indicativos acerca das cidades do estado.

3.2.4.1 Análise dos Dados

A fim de realizar uma avaliação acerca das alterações climáticas relacionadas com as duas séries temporais resultantes, primeiro vintênio (1980-1999) e segundo vintênio (2000-2019), resolveu-se empregar dois tipos de análises: uma conjunta e outra individual.

A Análise Conjunta se deu devido ao relacionamento entre as duas séries históricas, visto que há estações em comum entre elas. Desta maneira, por meio de uma sobreposição de *clusters* do primeiro período com *clusters* do segundo período, foi possível gerar zonas

em comum e através delas pode-se realizar uma constatação das alterações climáticas encontradas ao longo dos vintênios.

Além disso, procurou-se identificar zonas similares entre os resultados apresentados pela clusterização, com o intuito de realizar uma comparação com os dois períodos e avaliar as diferenças apresentadas em cada zona nos diferentes tempos. Realizada apenas em escala anual, foram analisadas as variações das seguintes variáveis climáticas: precipitação, temperatura máxima, temperatura mínima e umidade relativa.

Já a Análise Individual foi realizada com o intuito de avaliar de forma individual cada *cluster*, por meio da técnica dos Quantis para identificação de limiares referentes ao quantitativo de chuva no estado. A classificação se fez necessária para identificar os eventos extremos que ocorreram nos *clusters*. Ressalta-se que utilizou-se as escalas anual, mensal e diária.

Ao realizar uma avaliação em escala anual é possível compreender o regime climático da região, que é definido através dos quantitativos anuais. Por meio da análise mensal é possível observar a distribuição pluviométrica ao longo do ano, sendo assim, conforme se identifica um padrão nos meses é possível identificar os períodos que são afetados por eventos extremos. Por fim, a análise diária possibilita encontrar sequências de dias afetados pelo excesso ou escassez de chuva.

3.2.4.2 Técnica dos Quantis

Metodologia adaptada por (XAVIER; XAVIER, 2002), que consiste basicamente em definir uma variável \mathbf{X} , que representará os valores das variáveis climáticas, sendo $\mathbf{Q}_{(p)}$ o limite do intervalo do quantil para \mathbf{X} , p é a probabilidade de ocorrência e \mathbf{X} deverá ser um valor menor ou igual que $\mathbf{Q}_{(p)}$. Logo, deve-se escolher um intervalo a ser ligado a um quantil que satisfaça a seguinte interação: **Prob. ($\mathbf{X} \leq \mathbf{Q}_p$) = p** , para cada valor p que esteja no intervalo $0 < p < 1$.

Utilizando a técnica apresentada por (XAVIER; XAVIER, 2002) e com inspiração na mesma sendo aplicada em (DOURADO et al., 2013) para classificar eventos extremos na Bahia, adotou-se, para as escalas anual e mensal, as ordens de quantis p variando de 0,15 a 0,85, classificando os eventos como Muito Seco(MS), Seco(S), Normal(N), Chuvoso(C), Muito Chuvoso(MC), conforme demonstrado na Tabela 3. A análise em escala temporal diária utilizou a ordem quantílica p para os valores 0,05 e 0,95 e sua classificação está entre Extremamente Seco(ES) e Extremamente Chuvoso(EC), conforme apresentado na Tabela 4.

Tabela 3 – Categorias da Técnica dos Quantis para as escala Anual e Mensal

Classificação	Limites
Muito Seco (MS)	$X \leq Q_{(0,15)}$
Seco (S)	$Q_{(0,15)} < X \leq Q_{(0,35)}$
Normal (N)	$Q_{(0,35)} < X < Q_{(0,65)}$
Chuvoso	$Q_{(0,65)} \leq X < Q_{(0,85)}$
Muito Seco (MC)	$X \geq Q_{(0,85)}$

Tabela 4 – Categorias da Técnica dos Quantis para a escala Diária.

Classificação	Limites
Extremamente Seco (ES)	$X < Q_{(0,05)}$
Extremamente Chuvoso (EC)	$X > Q_{(0,95)}$

3.2.4.3 Análise Anual

Inicialmente, selecionou-se os dados em escalas mensais e os transformou em escalas anuais aplicando a média para cada base de dados. Em seguida, unificou-se os registros das estações que eram referentes a cada *clusters* e, então, foi calculada a média de todas as variáveis climáticas para cada zona com similaridade. A utilização da escala anual é de suma importância para a compreensão das tendências climáticas das zonas homogêneas.

A análise consistiu em avaliar as médias anuais geradas, para o primeiro vintênio referente ao período de 1980 a 1999, que contém os *clusters* 0, 1 e 2, e o segundo vintênio relacionado com os anos 2000 a 2019, que possui os *clusters* 3, 4 e 5. Por meio da utilização da análise dos quantis foi possível identificar os limiares superiores e inferiores, adotando como medida quantílica os valores de p sendo 0,15 (inferior) e 0,85 (superior). Além disso, após identificar os anos classificados como “Muito Seco” e “Muito Chuvoso”, foi realizada uma análise de Índice de Sazonalidade para cada um deles.

O cálculo da sazonalidade é uma técnica de comparação de valores médios de um determinado período, com base no seu padrão de ciclo sazonal. Seu índice é calculado a partir da média dos valores obtidos pela divisão entre o valor mensal pela sua média anual. Para fins de exemplificação, será feita uma avaliação do ano de 1980 do *cluster* 0 utilizando a variável com maior relação com as chuvas, ou seja, o índice de precipitação (mm). Os valores apresentados na Tabela 6 são obtidos a partir dos valores apresentados na Tabela 5, por meio da realização de uma racionalização do mês de um determinado ano, com a média anual desse mesmo ano. Por exemplo, considerando o mês de janeiro de 1980, sua precipitação é de 401,80 e a média anual é igual a 160,09, tem-se então $\frac{401,80}{160,09} \cong 2,51$, o resultado é a sazonalidade desse mês. Este mesmo cálculo aplica-se para

os outros meses, obtendo-se os índices de sazonalidade por meio de uma média entre os valores contidos na Tabela 5.

Tabela 5 – Esquema para obter Sazonalidade utilizando a Precipitação Acumulada Mensal e Precipitação Média Anual do *Cluster 0* (período de 1980 a 1999).

Ano	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dec	Méd. Anual
1980	401,8	487,1	87,7	78,89	0,3	0,1	0	0,5	79,1	120,3	356,7	308,6	160,09
...
1999	197,85	190,25	241,25	184,7	177,6	0,15	0	0	93,05	117,55	247,67	13,18	121,94

Tabela 6 – Esquema para obter a Média da Sazonalidade Mensal utilizando a sazonalidade de cada mês do *Cluster 0* (período de 1980 a 1999).

Ano	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dec
1980	2,51	3,04	0,55	0,49	0	0	0	0	0,49	0,75	2,23	1,93
...
1999	1,62	1,56	1,98	1,51	1,46	0	0	0	0,76	0,96	2,03	0,11
Med. Sazon.	1,93	1,7	1,98	1,27	0,39	0,09	0,03	0,09	0,37	1,03	1,49	1,61

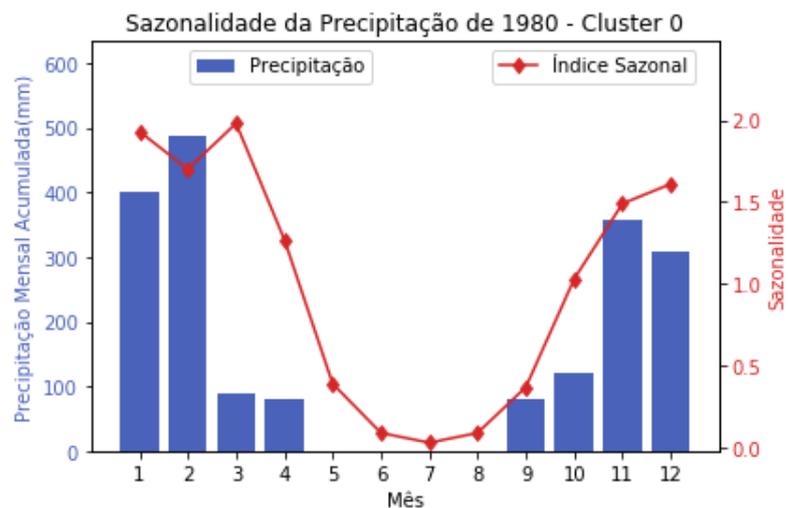


Figura 6 – Sazonalidade da Precipitação para ano de 1980 do Cluster 0.

A utilização do Índice de Sazonalidade foi empregado por (DOURADO et al., 2013) para uma análise de sazonalidade em todos os anos de sua pesquisa no estado da Bahia. Segunda a mesma, afirma que o método acaba por fornecer conhecimento acerca do comportamento pluviométrico de um determinado ano, em relação ao seu padrão sazonal,

possuindo como base toda uma série temporal. Relaciona-se aos resultados do exemplo a Figura 6, que representa uma comparação entre os valores médios de precipitação do ano de 1980 e a sazonalidade apresentada para aquele determinado *clusters*.

Ao final, foi empregada uma avaliação sobre as informações de ocorrência e intensidade dos principais fenômenos de oscilação de clima, atuantes na região estudada, o El Niño e o La Niña. Tendo como entendimento que seus efeitos acabam por afetar o comportamento climático “normal”, buscou-se explicar as mudanças climáticas influenciadas por eles ao longo de toda a série.

3.2.4.4 Análise Mensal

Nesta análise, semelhante a de escala anual, procurou-se utilizar a técnica do quantis, visando identificar, entre os meses de cada *cluster*, aqueles que eram tidos como mais chuvosos e mais secos. Os limiares foram obtidos utilizando a ordem quantílica de p sendo os valores de 0,85, para delimitar valores superiores, e 0,15, para delimitar os valores inferiores. A partir da aplicação da técnica foi possível identificar a quantidade de meses secos e muito chuvosos, para cada *cluster*, além de entender como está definida a distribuição das chuvas em cada mês.

3.2.4.5 Análise Diária

A última análise foi responsável por verificar os dados em sua forma inicial, pois as escalas mensais e anuais só foram possíveis através das médias convertidas utilizando esse tipo de dados.

A primeira etapa foi a aplicação da técnica de definição de limiares para identificação de eventos extremos, mas diferente das análises anteriores, esta foi realizada apenas para identificar e quantificar os eventos de precipitação que estão relacionados com extremas secas. Sendo assim, a ordem quantílica p , para os dias de extrema chuva foi de 0,95 e, considerou-se os de extrema seca os que apresentarem precipitação igual a 0,15 ou 0 mm.

Ao final, buscou-se localizar os dias de Extrema Seca (ES) e realizar uma contagem da quantidade de dias consecutivos que sofrem deste evento. Além disso, pode-se visualizar os anos que possuíram maiores sequências tidas como de extrema seca.

4 RESULTADOS E DISCUSSÃO

Este capítulo contém informações, resultados e avaliações sobre a aplicação da técnica de mineração de dados sobre a série temporal climática do Tocantins. As duas primeiras seções são referentes à análise dos dados sobre as principais variáveis climáticas e aplicação do algoritmo de clusterização. Enquanto as duas últimas seções estão relacionadas com avaliações dos *clusters* em diferentes aspectos (conjunta e individual) e em diferentes escalas (anual, mensal e diária).

4.1 Análise da Distribuição Climática

Por meio da ferramenta ArcMap foi possível obter a espacialização da distribuição climatológica das séries temporais analisadas, referente ao primeiro vintênio, que varia de 1980 até 1999, e o segundo vintênio, de 2000 a 2019, para a região do Tocantins. Além disso, utilizando o método geostático Densidade de Kernel (*Density Kernel*) presente na ferramenta, foi possível obter uma visualização da distribuição das variáveis climáticas, as quais foram utilizadas para obter as zonas homogêneas, que serão objeto de estudo para a próxima seção, na qual será analisado o comportamento das variáveis climáticas.

As variáveis que tiveram as suas distribuições representadas foram aquelas que apresentaram maior variabilidade, sendo possível identificá-las por meio de sua representação visual, são elas: Precipitação(mm), Umidade Relativa do Ar(%), Temperatura Máxima (°C) e Temperatura Mínima (°C). No primeiro vintênio as ilustrações das variáveis estão representadas nas Figuras 7, 8, no segundo vintênio são representadas nas Figuras 9, 10.

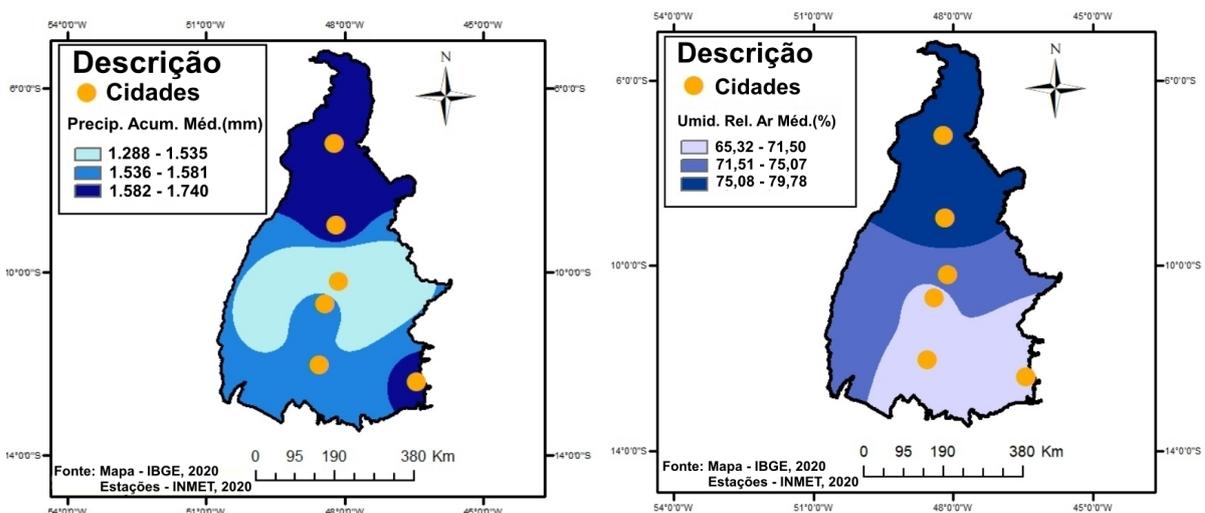


Figura 7 – Distribuição de Precipitação e Umidade Relativa do Ar do Primeiro Vintênio no estado do Tocantins.

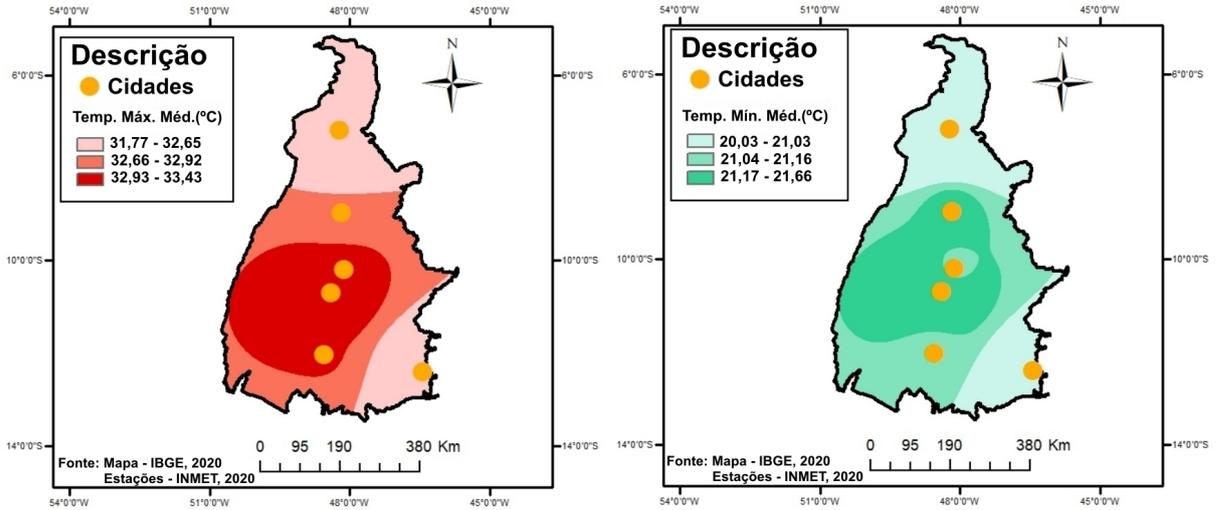


Figura 8 – Distribuição da Temperatura Máxima e Mínima no Primeiro Vintênio no estado do Tocantins.

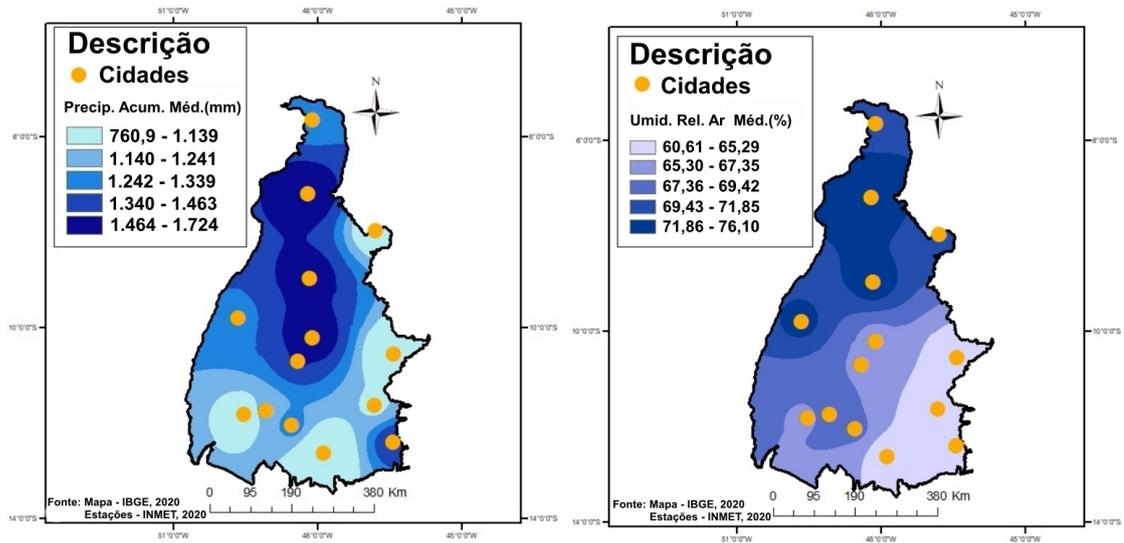


Figura 9 – Distribuição de Precipitação e Umidade Relativa do Ar do Segundo Vintênio no estado do Tocantins.

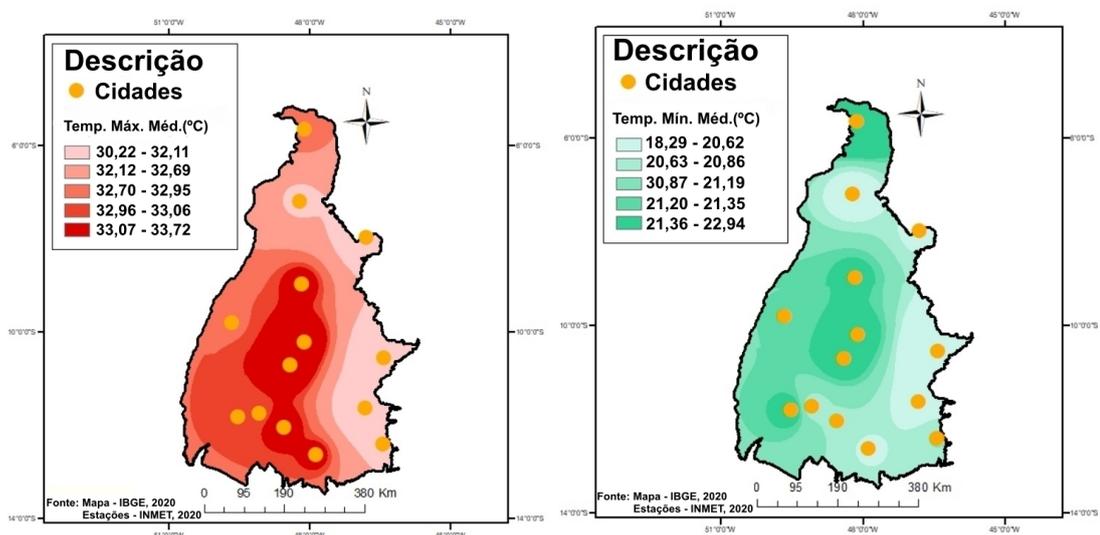


Figura 10 – Distribuição da Temperatura Máxima e Mínima no Segundo Vintênio no estado do Tocantins.

Analisando a distribuição das variáveis climáticas para o primeiro vintênio tem-se a precipitação, com a sua concentração em maioria na porção norte do estado e também no sudeste, enquanto a região central e sul registra valores menores. A variável umidade relativa do ar acompanhou a relação de precipitação, exceto para o sudeste que teve registro de valores semelhantes com a região sul. Em relação às temperaturas, a temperatura máxima teve maior concentração de calor na região norte e sul, enquanto a mínima registrou valores bastante variáveis em cada estação.

Já a distribuição do segundo vintênio, que possui uma quantidade maior de estações que o primeiro, apresentou maior variabilidade com relação a concentração de chuva, sendo esta maior na região central e centro-norte, com redução da chuva conforme se afasta do centro, principalmente, quando se aproxima da região sul e leste. A variável umidade concentra os maiores valores na região centro-norte, diminuindo quanto mais se aproxima da região sul-leste. A temperatura máxima do período registra os maiores valores na parte centro-sul, abaixando os valores quando se aproxima da região sul-leste e norte, enquanto as temperaturas mínimas, assim como no primeiro período, variam bastante, comparando com as outras variáveis, sua concentração está relacionada com a região do centro e no norte do estado.

4.2 Análise das Zonas Homogêneas

Os *clusters* foram obtidos a partir da aplicação do algoritmo de agrupamento *k-means*, que se baseiam nas similaridades encontradas entre as variáveis climáticas de cada estação. A técnica foi empregada em duas séries temporais, considerando o primeiro vintênio e o segundo vintênio, que resultaram em zonas de similaridade climática para o primeiro período e o segundo período. A aplicação deste método foi feita utilizando os dados em uma escala mensal.

Para a obtenção das zonas homogêneas, foram testados os valores de $k = 2, 3, 4, 5$ para o algoritmo *k-means*, considerando os atributos precipitação, umidade relativa do ar, temperatura máxima, temperatura média e temperatura mínima. Destes, $k = 3$ apresentou o melhor resultado em relação à convergência das zonas, o qual foi o único valor que apresentou menor tendência a convergir em apenas uma ou duas zonas. A clusterização do primeiro vintênio é apresentada na Figura 11, enquanto que a do segundo vintênio se encontra na Figura 12.

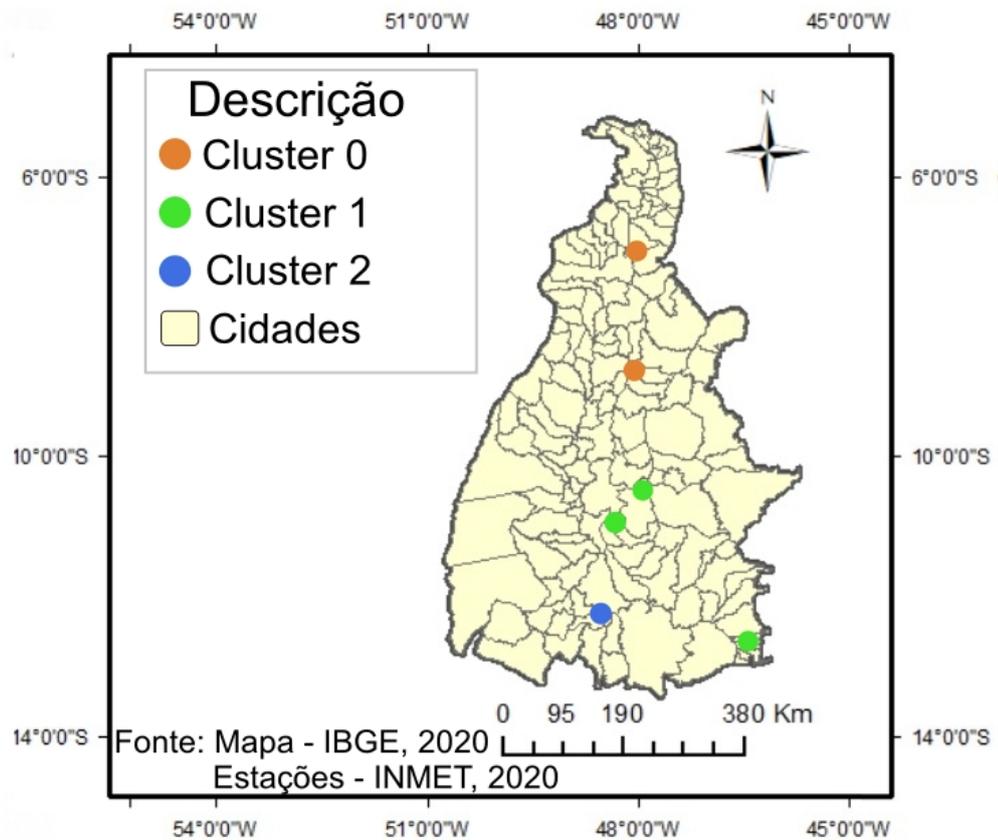


Figura 11 – Resultado da aplicação do algoritmo *k-means* nos dados do primeiro vintênio.

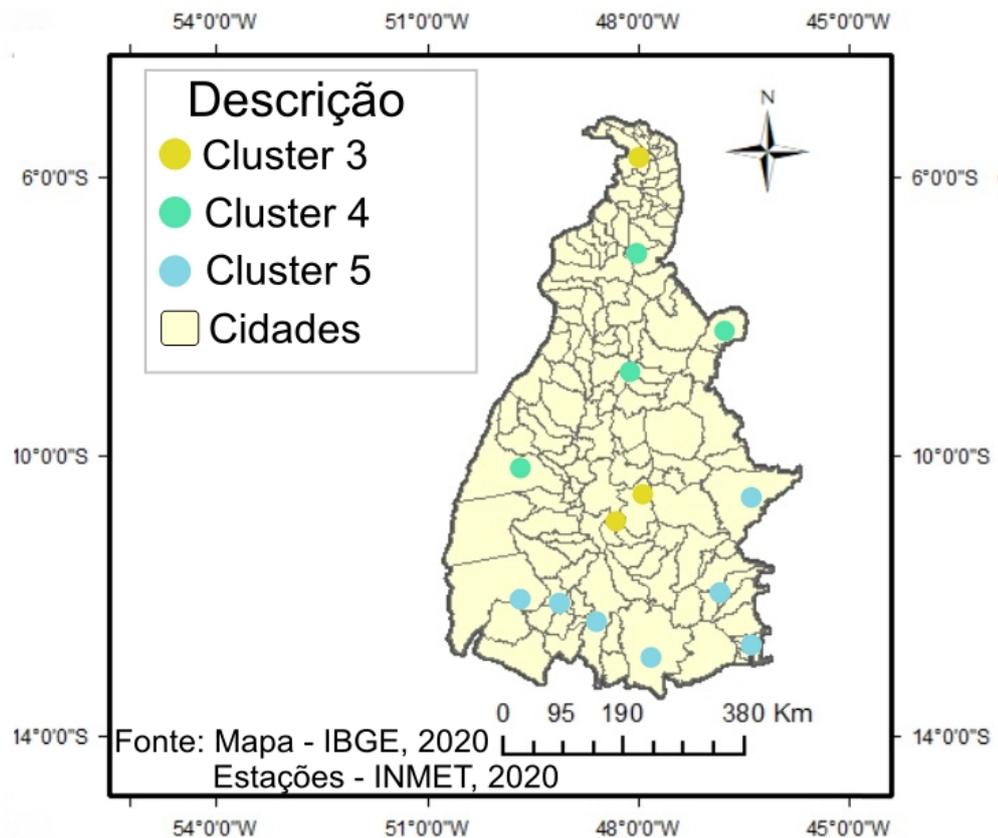


Figura 12 – Resultado da aplicação do algoritmo *k-means* nos dados do segundo vintênio.

O posicionamento dos *clusters* possivelmente teve influência da quantidade de precipitação, pois, de acordo com (CHARLES, 2015) o número de chuvas acaba afetando as outras variáveis climáticas, como temperatura e umidade, pois em cenários com muita chuva as temperaturas tendem a diminuir e a umidade a aumentar e, em cenários com seca, as temperaturas sobem e a umidade tendem a diminuir.

Observando a concentração da distribuição de chuva é notada que ela é maior pelo centro-norte, diminuindo conforme segue pelas extremidades do estado. Os sistemas que causam a variabilidade climática no Tocantins são a Zona de Convergência do Atlântico Sul (ZCAS), a Zona de Convergência de Umidade (ZCOU) e a Zona de Convergência Intertropical (ZCIT) que resultam em diferentes regimes de chuva sobre a região.

Analisando as influências de cada *cluster* percebe-se uma semelhança entre os presentes no primeiro e segundo vintênio, visto que o relacionamento se dá por serem afetados pelos mesmos sistemas de variabilidade. Os *clusters* 0 e 4 possuem duas e quatro estações relacionadas, respectivamente, e se encontram ligados com a região centro-norte e norte do estado. Com isso, a ZCAS possui uma atuação menor e acaba não gerando alta variabilidade nas chuvas, logo essa região possui o clima considerado mais "instável" da região.

Já os *clusters* 1 e 3, cada um com 3 estações, as quais estão presentes majoritariamente na região central do estado, que está relacionada com a ZCAS, que atua como gerador de chuvas na maior parte da região central do estado e sudeste. A região costuma receber altos volumes de chuvas quando ocorrem instabilidades nas zonas de ZCIT.

Por último, os *clusters* 2 e 5, o primeiro com uma estação e o segundo com 7, representam as cidades da região sul e sudeste. A região sudeste acaba por ter chuvas bem distribuídas causadas pela ZCAS e a região sul e sudoeste costuma sofrer com secas, pois é a mais afetada pela ZCOU, com oscilações e perturbações atmosféricas.

Com o intuito de entender a heterogeneidade da distribuição das variáveis climáticas, apresentou-se de forma visual as médias, desvios-padrão, medianas e valores máximos e mínimos de cada *cluster*. Os resultados estão em escala anual e podem ser vistos nas Tabelas 7, 8, 9, 10.

Tabela 7 – Valores da Precipitação Anual, médias, desvio-padrão, máximo e mínimo de cada *Cluster*.

Cluster	Média	Desv. Pad.	Mediana	Mínimo	Máximo
0	1674,99	376,02	1661,97	848,40	2702,56
1	1530,53	302,30	1557,34	818,01	2020,0
2	1551,66	352,90	1609,90	936,58	2288,45
3	1623,31	251,20	1659,77	1115,52	2005,12
4	1595,39	271,12	1653,00	936,59	1968,80
5	1352,84	331,61	1399,62	444,81	1817,20

Tabela 8 – Valores da Umidade Relativa do Ar Média Anual, médias, desvio-padrão, máximo e mínimo de cada *Cluster*.

Cluster	Média	Desv. Pad.	Mediana	Mínimo	Máximo
0	79,17	1,54	79,09	76,59	83,02
1	71,07	1,58	71,03	67,77	74,25
2	65,57	2,33	65,16	60,14	70,23
3	67,86	2,17	67,31	64,91	72,32
4	75,04	2,92	74,81	70,72	79,45
5	65,43	3,25	65,46	59,69	73,71

Tabela 9 – Valores da Temperatura Máxima Anual, médias, desvio-padrão, máximo e mínimo de cada *Cluster*.

Cluster	Média	Desv. Pad.	Mediana	Mínimo	Máximo
0	32,44	0,51	32,44	31,40	33,72
1	33,13	0,55	33,04	32,06	34,72
2	31,73	0,73	31,61	30,47	33,38
3	33,88	0,49	33,89	32,95	34,90
4	32,96	0,42	32,95	31,98	33,95
5	32,77	0,52	32,67	31,31	33,67

Tabela 10 – Valores da Temperatura Mínima Anual, médias, desvio-padrão, máximo e mínimo de cada *Cluster*.

Cluster	Média	Desv. Pad.	Mediana	Mínimo	Máximo
0	20,73	0,30	20,69	20,05	21,39
1	21,28	0,32	21,29	20,62	21,99
2	20,60	0,40	20,55	19,94	21,63
3	22,41	0,45	22,43	21,60	23,21
4	21,26	0,35	21,24	20,83	22,44
5	21,03	0,30	21,04	20,48	21,75

Por meio das tabelas é possível identificar as características entre as variáveis dos *clusters* como, por exemplo, as zonas homogêneas mais quentes são a 3 e 4, pois apresentam os valores mais altos tanto para temperatura máxima, quanto para temperatura mínima, e as suas médias nas duas variáveis são as maiores. O *cluster* mais frio é o 2, pois demonstra ter os valores mais baixos para temperaturas máxima e mínima e, assim como nas zonas 3 e 4, a sua média acompanha as temperaturas.

Verifica-se que existe uma correlação entre a precipitação e a umidade relativa do ar média, ou seja, o *cluster* 0 possui a maior quantidade de chuva entre todos e possui também a maior média. Além disso, observa-se que a variável umidade acompanha essa relação. De forma análoga, o *cluster* 5 registra o menor índice de chuva média e apresenta também a menor média de umidade relativa do ar.

A variável precipitação é a que mais se destaca entre todas e, com o objetivo de compreender melhor a sua variabilidade, foi analisada a sua distribuição em escala mensal, realizando uma média para os meses de todos os clusters, como mostrado no Figura 13.

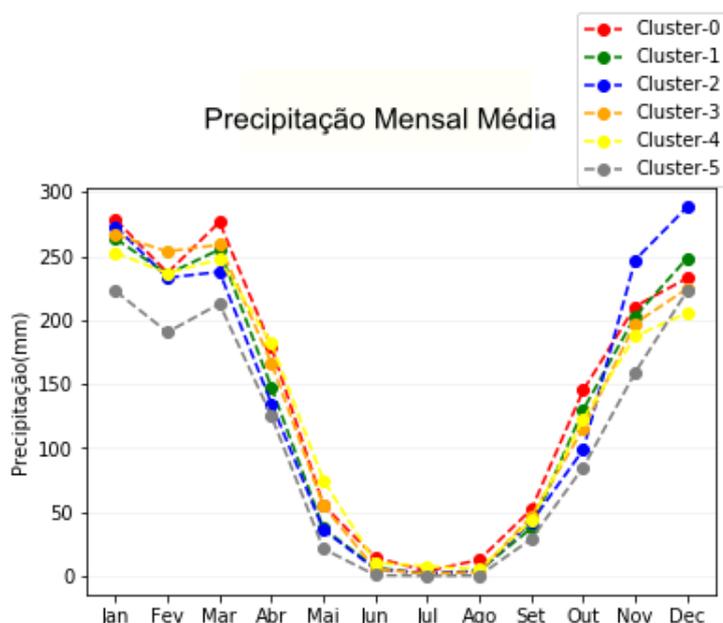


Figura 13 – Precipitação média mensal dos *Clusters*.

O gráfico da Figura 13 mostra que a distribuição pluviométrica do estado é bem definida para cada *clusters* com relação aos meses de pouca ou muita chuva. Destacam-se, inicialmente, os *clusters* 0 e 5, visto que o *cluster* 0, no início do ano, apresenta os maiores valores de precipitação, e o *cluster* 5, de forma contrária, é o pior com relação ao índice de chuva.

Durante todo o ciclo mensal o *cluster* 5, com exceção do mês de dezembro (neste caso o *cluster* 4 é o menor), apresenta os menores valores de precipitação, já o *cluster* 0 mantém os altos índices, exceto para os dois últimos meses do ano em que o *cluster* 2 apresenta as maiores médias de precipitação mensal.

4.3 Análise Conjunta

De modo a avaliar as mudanças e as alterações ocorridas ao longo das séries de forma agrupada, considerando o primeiro e o segundo vintênio, adotou-se um esquema de sobreposição das zonas homogêneas entre os dois períodos para que, desta maneira, fosse

possível representar as regiões em comum em forma de zonas. Baseado no posicionamento das estações de cada *cluster*, foram definidas três zonas comuns, como apresentadas na Figura 14. A zona 1, relacionada com a porção norte do estado, está ligada aos *clusters* 0 e 4. Já a zona 2 se encontra localizada na região central, representada pelos *clusters* 1 e 3. Por último a zona 3 posicionada na região sul do Tocantins, contém em seu grupo os *clusters* 2 e 5.

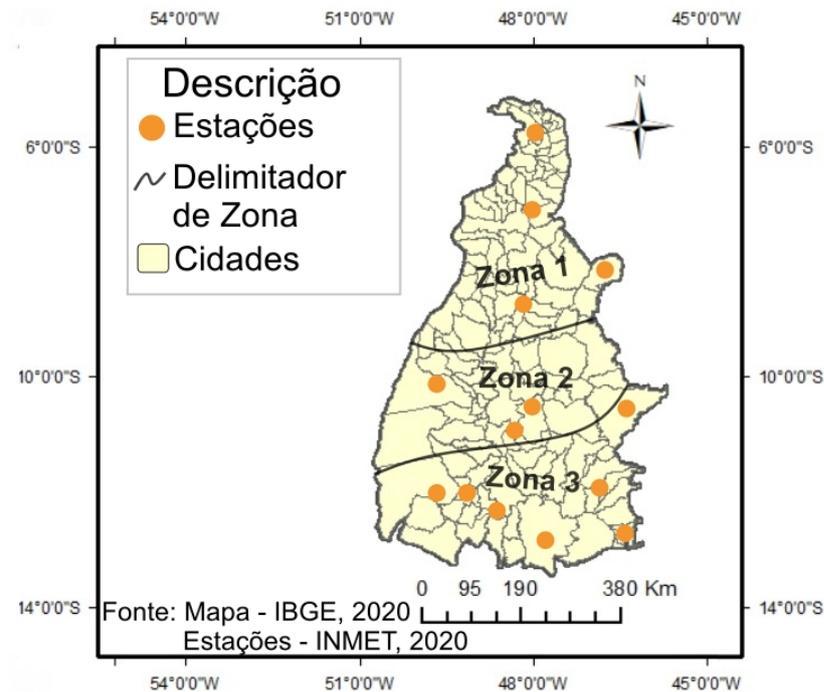
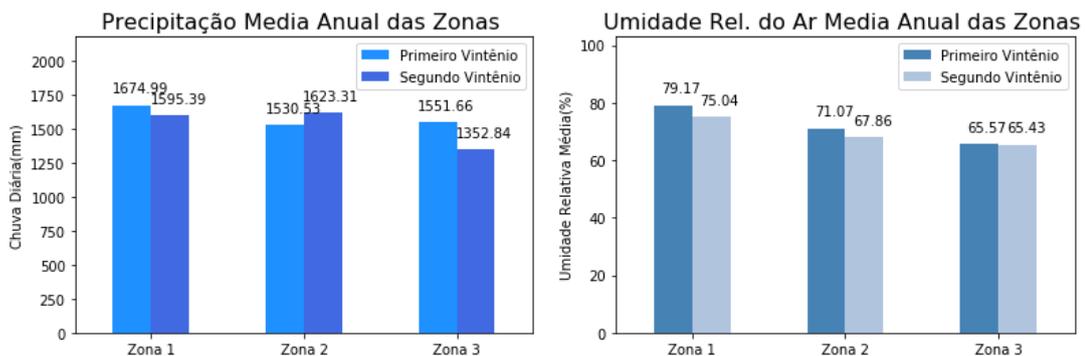


Figura 14 – Delimitação das Zonas Homogêneas comuns entre os Vintênios (1-3).

A partir da definição de zonas foi possível analisar os desvios do primeiro vintênio em relação ao segundo. Por meio das médias anuais das variáveis precipitação, umidade relativa do ar, temperatura máxima e temperatura mínima, apresentadas na Figura 15, conseguiu-se analisar as variações ao longo de toda a série temporal, as quais são apresentadas na Tabela 11.



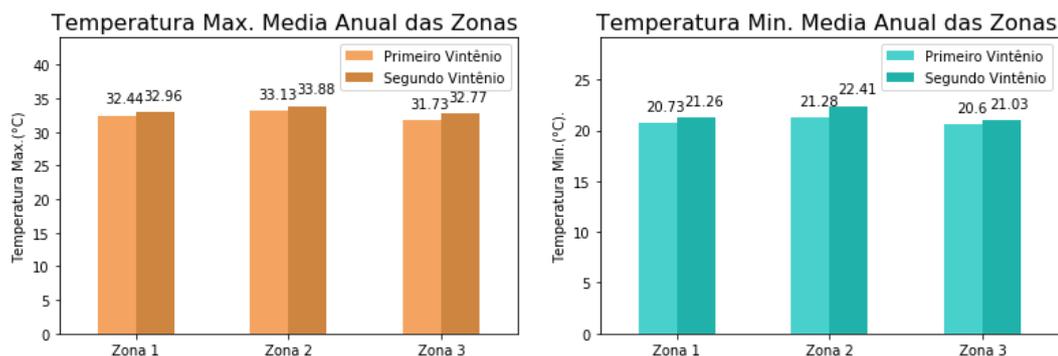


Figura 15 – Análise das variáveis considerando os dois vintênios presentes nas Zonas.

A zona que apresentou maior nível de precipitação no primeiro vintênio foi a Zona 1 com 1674,99 mm e, no segundo, a Zona 2 com 1623,31mm. De forma contrária a zona que apresentou o menor nível de precipitação foi a Zona 2 com 1530,53mm, no primeiro vintênio, e a Zona 3, com 1352,84mm, no segundo.

Em relação a umidade relativa do ar, a zona com maior porcentagem no primeiro período foi a Zona 1 com 79,17% e no segundo a Zona 2 com 75,04%. Já a Zona 3 apresentou a e menor umidade relativa do ar, tanto no primeiro quanto no segundo vintênio, com 65,57% e 65,43%, respectivamente.

Já a zona com maior temperatura máxima no primeiro vintênio foi a Zona 2 com 33,13°C e no segundo a Zona 1 com 33,88°C. E a de menor temperatura máxima, tanto no primeiro quanto no segundo vintênio, foi a Zona 3 com 31,73°C e 32,77°C, respectivamente.

Por fim, em termos de temperatura mínima, no primeiro vintênio, a Zona 2 apresentou a maior temperatura mínima com 21,28°C e, no segundo vintênio, a Zona 1 com 22,46°C. Em relação a menor temperatura mínima, a Zona 3 apresentou os menores valores no primeiro período, 20,60°C, e no segundo com 21,03°C.

A Tabela 11 apresenta as variações entre os *clusters* presentes no primeiro vintênio em relação ao segundo vintênio. Com base na precipitação somente a Zona 2 apresentou desvios positivos, a Zona 1 apresentou uma baixa de 79,60(mm) e a Zona 3, 198,82(mm) da precipitação anual. Esta última variação foi mais elevada, pois o primeiro vintênio possui apenas uma estação, enquanto o segundo tem sete estações, logo a tendência é a quantidade média ter impacto na quantidade de precipitação. Por outro lado, as variáveis de temperatura apresentaram ambas, uma subida do segundo vintênio com relação ao primeiro, o que demonstra que o segundo vintênio foi um período mais quente. Esta teoria pode ser confirmada analisando a variável umidade relativa do ar, que diferente das temperaturas teve diminuição em seus valores do segundo vintênio, o que pode ser relacionado com uma elevação nas temperaturas da região.

Tabela 11 – Tabela de desvios das variáveis climáticas do primeiro vintênio em relação ao segundo vintênio.

Zona	Desvio de Precipitação(mm)	Desvio de Umidade Rel. do Ar (%)
1	79,60	4,13
2	92,78	3,21
3	198,82	0,14
Zona	Desvio de Temp. Máx. (°C)	Desvio de Temp. Mín.(°C)
1	0.52	0.53
2	0.75	1.13
3	1.04	0.43

4.4 Análise de Eventos Extremos de Precipitação

Para compreender o comportamento e as tendências dos eventos extremos de precipitação foi realizada, por meio da aplicação da Técnica dos Quantis, considerando a escala anual, mensal e diária, a análise individual de cada *cluster*. As subseções seguintes apresentam os resultados.

4.4.1 Análise Anual

De acordo com análise conjunta, exceto para a Zona 2, todas as outras zonas apresentaram decréscimo em suas distribuições pluviométricas, evento que não aconteceu para os primeiros anos do segundo vintênio, representado pelos *clusters* 3, 4 e 5, porém nos últimos cinco anos houve uma diminuição considerável nos volumes de chuva nos *clusters* desse período. Segundo a Federação da Agricultura e Pecuária do Estado do Tocantins (FAET, 2016), esse decréscimo é consequência de um período de estiagem que ocorreu entre os anos de 2015 e 2016, devido a um fenômeno atmosférico-oceânico, El Niño, que deixou o volume de chuva do período abaixo do esperado, prejudicando as safras agrícolas, e acabou impactando o clima nos anos posteriores.

Sendo assim, resolveu-se analisar a distribuição da precipitação média anual, identificados nas Figuras 16 e 17, que apresenta os gráficos das médias anuais de cada *cluster* analisado. Desta maneira, as informações contidas neles serão avaliadas, visando destacar os anos com excesso de chuvas e os com escassez da mesma.

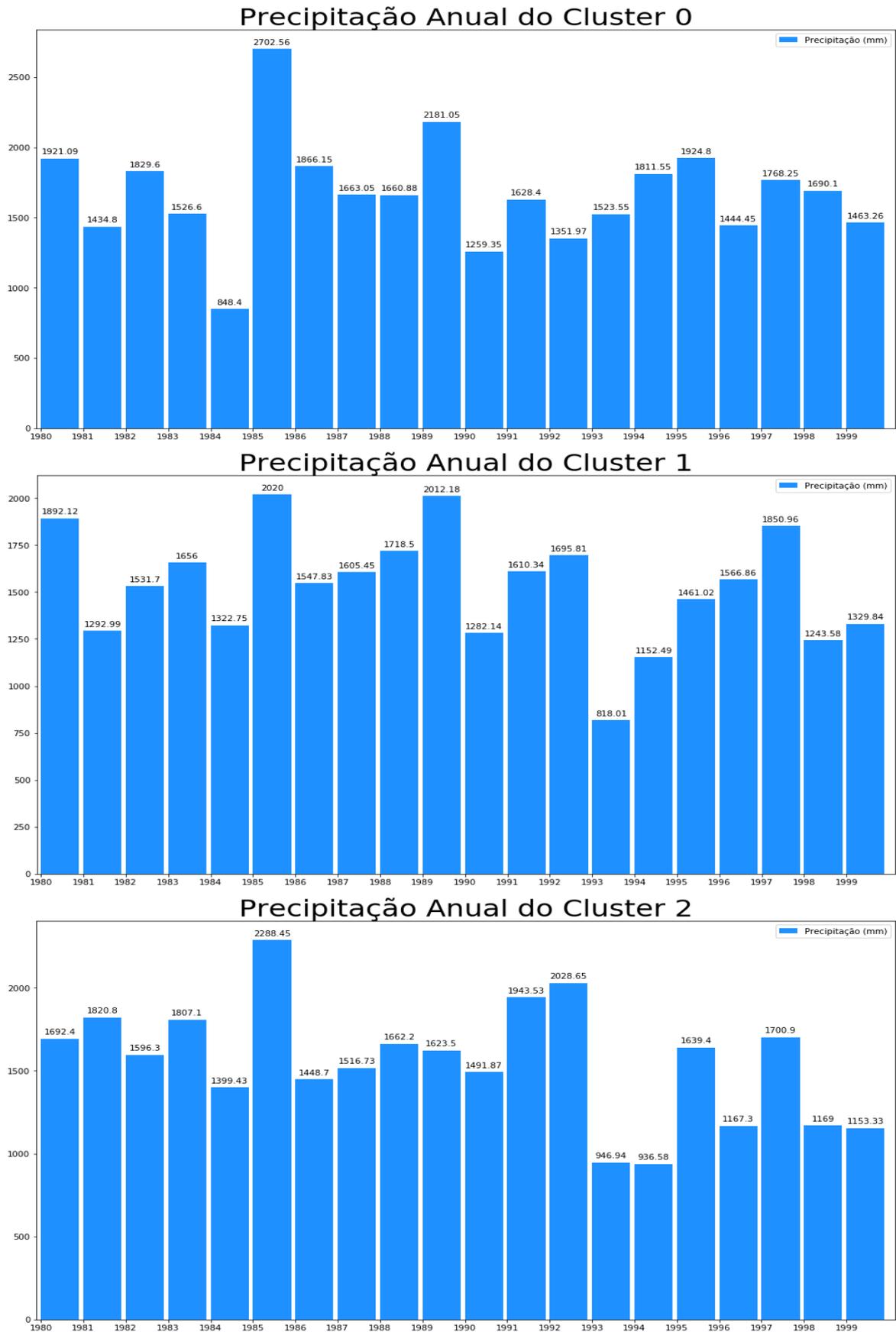


Figura 16 – Distribuição da precipitação anual para as zonas homogêneas no primeiro vintênio.

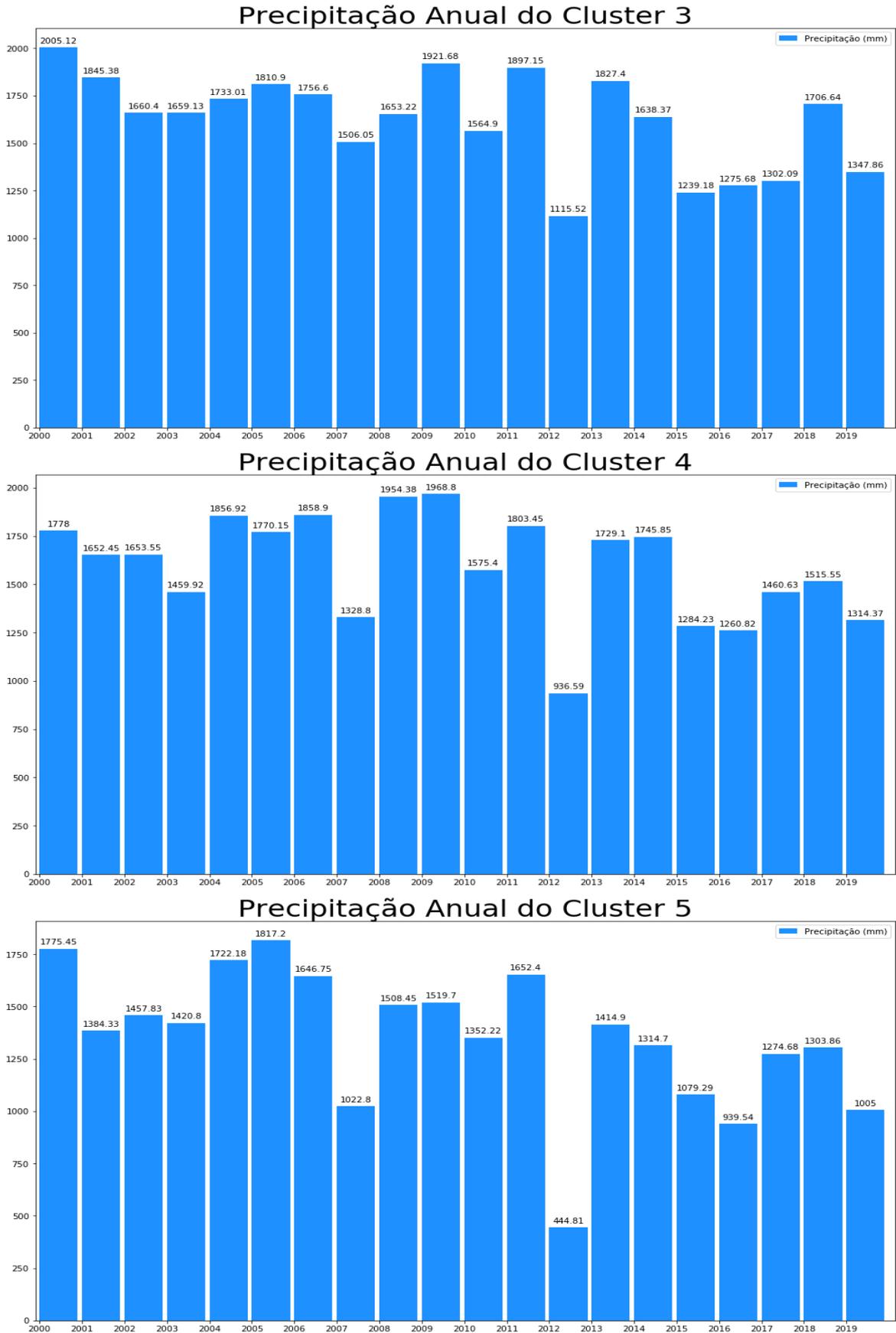


Figura 17 – Distribuição da precipitação anual para as zonas homogêneas no segundo vintênio.

A técnica dos quantis foi aplicada utilizando os dados pluviiais anuais e a classificação resultante dos registros dos anos é apresentado na Tabela 12, a qual apresenta a quantidade total pluviométrica do ano em conjunto com sua classificação de cada *cluster*.

Tabela 12 – Registro da precipitação e classificação dos anos pelos quantis nos *clusters*. Categorias: MC = muito chuvoso, C = chuvoso, N = normal, MS = muito seco e S = seco.

Ano	Cluster0	Cluster1	Cluster2	Ano	Cluster3	Cluster4	Cluster5
1980	1921,09 C	1892,12 MC	1692,4 C	2000	2005,12 MC	1778,0 C	1775,45 MC
1981	1434,8 S	1292,99 S	1820,8 C	2001	1845,38 C	1652,45 N	1384,33 N
1982	1829,6 C	1531,7 N	1596,3 N	2002	1660,4 N	1653,55 N	1457,83 N
1983	1526,6 N	1656,0 C	1807,1 C	2003	1659,13 N	1459,92 S	1420,8 N
1984	848,4 MS	1322,75 S	1399,43 S	2004	1733,01 N	1856,92 C	1722,18 MC
1985	2702,56 MC	2020,0 MC	2288,45 MC	2005	1810,9 C	1770,15 C	1817,2 MC
1986	1866,15 C	1547,83 N	1448,7 S	2006	1756,6 C	1858,9 MC	1646,75 C
1987	1663,05 N	1605,45 N	1516,73 N	2007	1506,05 S	1328,8 S	1022,8 S
1988	1660,88 N	1718,5 C	1662,2 N	2008	1653,22 N	1954,38 MC	1508,45 C
1989	2181,05 MC	2012,18 MC	1623,5 N	2009	1921,68 MC	1968,8 MC	1519,7 C
1990	1259,35 MS	1282,14 S	1491,87 N	2010	1564,9 S	1575,4 N	1352,22 N
1991	1628,4 N	1610,34 N	1943,53 MC	2011	1897,15 MC	1803,45 C	1652,4 C
1992	1351,97 MS	1695,81 C	2028,65 MC	2012	1115,52 MS	936,59 MS	444,81 MS
1993	1523,55 S	818,01 MS	946,94 MS	2013	1827,4 C	1729,1 N	1414,9 N
1994	1811,55 C	1152,49 MS	936,58 MS	2014	1638,37 N	1745,85 N	1314,7 N
1995	1924,8 MC	1461,02 N	1639,4 N	2015	1239,18 MS	1284,23 MS	1079,29 S
1996	1444,45 S	1566,86 N	1167,3 S	2016	1275,68 MS	1260,82 MS	939,54 MS
1997	1768,25 N	1850,96 C	1700,9 C	2017	1302,09 S	1460,63 S	1274,68 S
1998	1690,1 N	1243,58 MS	1169,0 S	2018	1706,64 N	1515,55 N	1303,86 S
1999	1463,26 S	1329,84 S	1153,33 MS	2019	1347,86 S	1314,37 S	1005,0 MS

Nota-se que os anos considerados secos e chuvosos variam de *cluster* para *cluster*, pois a rotulação ocorre com base nos registros individuais das zonas homogêneas. Logo, um ano considerado seco para uma região, pode ser considerado chuvoso para outra.

De acordo com (XAVIER; XAVIER, 2002) e confirmado por outros autores que aplicaram técnica semelhante, como (DOURADO et al., 2013), afirmam que os limiares de precipitação variam para um grupo de cada região. Analisando a Tabela 13 pode-se observar que a classificação dos limites extremos “muito seco” e “muito chuvoso” variam por *cluster*, independente de pertencer ao primeiro ou segundo vintênio.

Tabela 13 – Limiares de classificação anual para os eventos extremos muito seco e muito chuvoso relacionados com cada *cluster*.

Cluster	Muito Seco(MS)	Muito Chuvoso(MC)
0	X ≤ 1422,38	X ≥ 1921,65
1	X ≤ 1276,36	X ≥ 1857,13
2	X ≤ 1165,20	X ≥ 1839,21
3	X ≤ 1298,13	X ≥ 1853,15
4	X ≤ 1309,85	X ≥ 1857,22
5	X ≤ 1020,13	X ≥ 1662,87

Constata-se que os menores limiares para os dois extremos (muito seco e muito chuvoso) são encontrados no *cluster* 2, para o primeiro vintênio com 1165,20 e, no *cluster* 5, para o segundo vintênio com limiar anual de 1020,13, sendo estes localizados em uma região mais afetada por oscilações climáticas. Já os maiores limiares foram observados nos *cluster* 0 e 4 com 1921,65 e 1857,22, para o primeiro e segundo vintênio, respectivamente, e diferente dos menores limiares, esta é uma região menos afetada por oscilações climáticas, (ROLDÃO; FERREIRA, 2019).

A Tabela 14 apresenta os anos que foram considerados como “muito seco” e “muito chuvoso” para os *clusters* das séries temporais analisadas. Avaliando os anos identificados com ocorrência de eventos extremos destacam-se, para o primeiro vintênio, os anos representados por 1984, 1990, 1992, 1993, 1994, 1998 e 1999 como “muito seco”. Já para o segundo vintênio há uma predominância dos anos da década 20, os anos de 2012, 2015, 2016 e 2019. Para eventos considerados como “muito chuvoso” destacam-se, no primeiro vintênio, os anos de 1985, 1989, 1991, 1992 e 1995, enquanto que para o segundo vintênio tem-se os anos 2000, 2004, 2005, 2006, 2008, 2009 e 2011.

Tabela 14 – Relação dos anos que apresentaram ocorrência de eventos extremos (muito chuvoso e muito seco).

Cluster	Muito Seco (MS)	Muito Chuvoso (MC)
0	1984, 1990, 1992	1985, 1989, 1995
1	1993, 1994, 1998	1980, 1985, 1989
2	1993, 1994, 1999	1985, 1991, 1992
3	2012, 2015, 2016	2000, 2009, 2011
4	2012, 2015, 2016	2006, 2008, 2009
5	2012, 2016, 2019	2000, 2004, 2005

Entre os anos mais secos, aqueles que se destacam no primeiro vintênio são os anos de 1993 e 1994, que estão presentes nos *cluster* 1 e 2, possivelmente relacionados com a grande seca identificada na região nordeste e em parte da região norte, situada ao lado

leste. O período é registrado na história como o de grande seca para o setor agropecuário, o que foi constatado por pesquisas como as de Xavier (XAVIER; XAVIER, 2002), Barbosa (BARBOSA et al., 2008) e Dourado (DOURADO et al., 2013), que destacaram os anos como os que apresentaram piores índices pluviométricos.

Para o segundo vintênio destacam-se os anos 2012 e 2016, presentes em todos os *clusters*. Esses períodos estiveram relacionados com um período de estiagem e o ano de 2016 esteve relacionado com grandes perdas no setor agropecuário (FAET, 2016). Estudos como o de (FLEISCHMANN et al., 2017) e (CARVALHO et al.,) também relataram que esse ano (2016) foi um dos que registraram maiores perdas no setor agrícola, apresentando, tanto a curto prazo como a longo prazo, secas extremas relatando que, assim como no Tocantins, outros estados decretaram situação de emergência.

Com relação aos anos mais chuvosos, no primeiro vintênio, o ano de 1985, que está presente em todos os *clusters*, apresentou a maior precipitação. Já no segundo vintênio, houve uma maior dispersão entre os anos *clusters*, em relação à ocorrência de “muito chuvoso”, destacando o ano de 2009, presente nos *clusters* 3 e 4, enquanto o ano de 2000, esteve situado nos *clusters* 3 e 5.

Com isso, baseado no trabalho de (DOURADO et al., 2013), resolveu-se realizar uma análise de sazonalidade da precipitação. Deste modo, foi selecionado, entre os anos que apresentaram comportamento extremo, aqueles que possuíam maior discrepância entre os seus semelhantes, sendo possível avaliá-lo com relação ao comportamento sazonal da região a qual pertence.

A Figura 18 apresenta o relacionamento do índice de sazonalidade da precipitação mensal do *cluster* 0 com a distribuição de chuvas mensais nos anos de 1985 e 1984. Os dois anos demonstraram que os regimes pluviométricos acompanharam a sazonalidade, exceto, os meses de fevereiro e março que se inverteram, com base na precipitação esperada, e o mês de novembro, que diminuiu a chuva, para o segundo ano analisado.

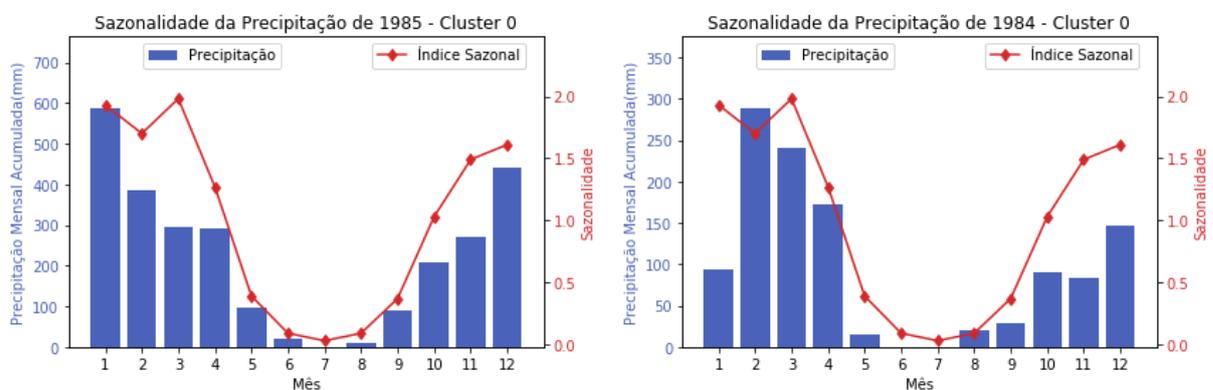


Figura 18 – Sazonalidade da precipitação dos anos 1985 e 1984 para o *Cluster* 0.

Em relação à precipitação, o ano 1985 apresentou no mês de janeiro, que é tido

como o mês mais chuvoso da série, um acúmulo de 587,33 mm, sendo este maior que o dobro da média esperada para o mês. Além disso, exceto o mês de julho, todos os meses do período apresentaram valores acima da média. O ano de 1984 registrou o mês de fevereiro, como uma anomalia, em relação aos outros meses, pois foi o único que não registrou valores abaixo da média.

Analisando a Figura 19, a qual apresenta os gráficos com a distribuição de precipitação nos anos 1985 e 1993, que foram, respectivamente, o ano mais chuvoso e o mais seco, do *cluster* 1, tem-se que, no ano de 1985, em relação aos registros de precipitação, com exceção dos meses de julho e novembro, todos os outros acompanham a curva de sazonalidade. Com isso, o mês de maior precipitação alinha-se ao mês que apresenta a maior precipitação sazonal, sendo este o de janeiro para ambos os casos.

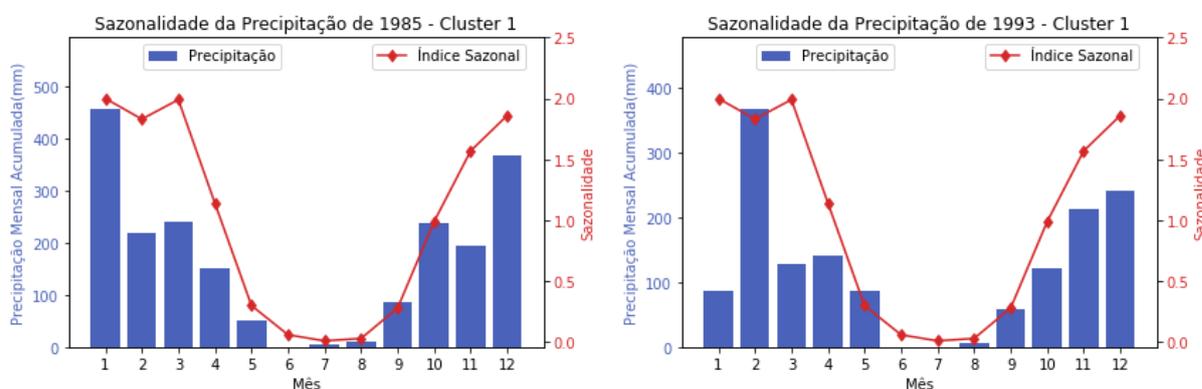


Figura 19 – Sazonalidade da precipitação dos anos de 1985 e 1993 para o *Cluster* 1.

Observando o ano de 1993, considerando o primeiro semestre, nota-se que este não acompanha a sazonalidade esperada, visto que os meses considerados chuvosos, com exceção do mês de fevereiro, apresentaram valores muito abaixo do esperado, como no mês de janeiro que teve 86,07 mm de precipitação, mesmo sendo esperado 263,54 mm.

A Figura 20 ilustra os gráficos das distribuições do ano de 1985, mais chuvoso, e 1993, mais seco, considerando a sazonalidade do *cluster* 2. Para o ano 1985, os meses de abril e novembro não acompanharam a sazonalidade, sendo que o primeiro apresentou valores altos e o segundo valores baixos, quando o esperado era o contrário. Destaca-se que os meses de janeiro e abril apresentaram, respectivamente, as precipitações acumuladas de 571,20 mm e 310,90 mm, valores duas vezes maior que a média esperada individualmente para cada mês. Nota-se que em 1993, em sua maioria os meses não acompanharam o índice de sazonalidade, pois os dois meses de maior precipitação para o *cluster*, janeiro e dezembro, mostraram índices de precipitação muito abaixo, assim como os meses de outubro e dezembro. Além disso, os meses que apresentaram registro “normais” para a precipitação foram os meses de seca, ou seja, fevereiro e novembro.

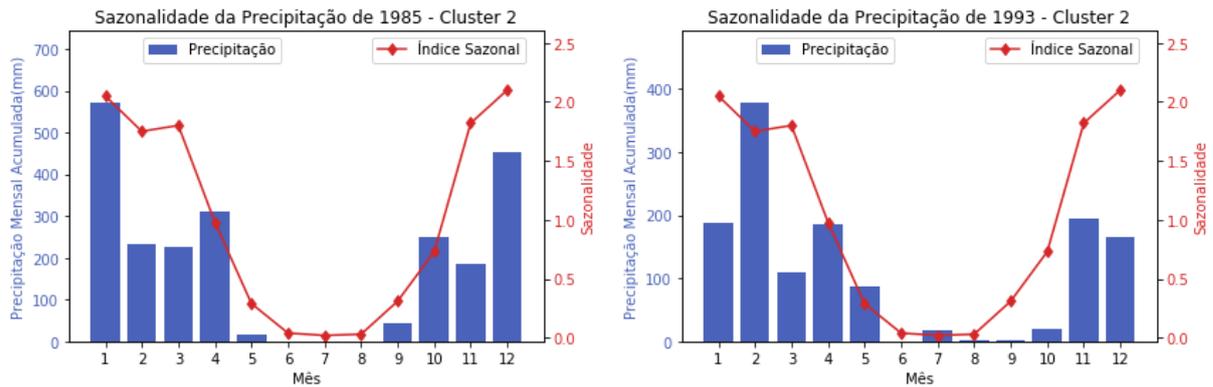


Figura 20 – Sazonalidade da precipitação dos anos de 1985 e 1993 para o *Cluster 2*.

A Figura 21 mostra a sazonalidade do *cluster 3*, considerando os anos 2000 e 2012, considerados os mais chuvosos e secos, respectivamente. Analisando a precipitação no ano 2000 percebe-se que a sazonalidade foi acompanhada, exceto para o mês de fevereiro que apresentou uma alta. Além disso, os meses de fevereiro e dezembro apresentaram valores muito superiores à distribuição pluviométrica esperada. Para o ano 2012 nota-se que os meses de janeiro e fevereiro estão relacionados com a estação chuvosa, já os outros meses que se esperava uma precipitação um pouco maior apresentaram quedas, como é o caso dos meses de março, outubro e dezembro.

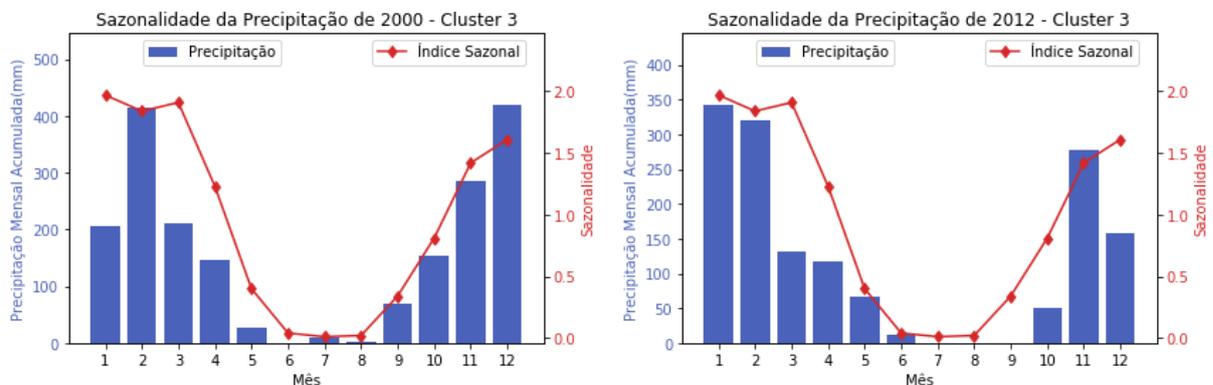


Figura 21 – Sazonalidade da precipitação dos anos de 2000 e 2012 para o *Cluster 3*.

A Figura 22 apresenta os gráficos da relação entre a sazonalidade e a distribuição pluvial do ano 2009, mais chuvoso, e 2012, mais seco. O ano 2009 apresentou distribuição de chuvas considerável para todos os meses, exceto os meses de janeiro e novembro que foram abaixo do esperado. Os meses de Abril e Maio se destacaram pela precipitação elevada, apresentando 303 mm e 181,50 mm de chuva, onde a média era para ser de valores 181,53 mm e 74,73 mm, respectivamente. Já o período mais seco, no ano 2012, demonstrou acompanhar a curva de variação da sazonalidade, apenas com os meses de março e dezembro apresentando valores divergentes do esperado.

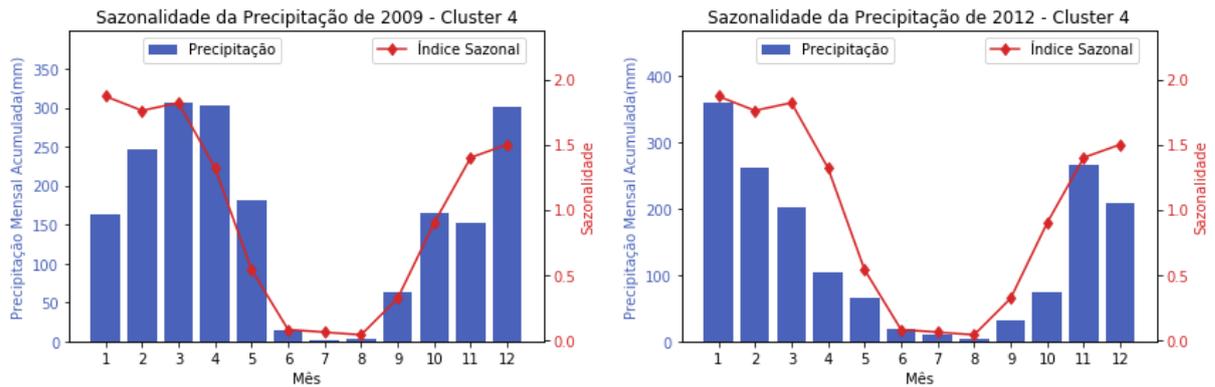


Figura 22 – Sazonalidade da precipitação dos anos de 2009 e 2012 para o *Cluster 4*

A Figura 23 representa a relação entre a distribuição pluviométrica dos anos mais chuvoso e mais seco, 2005 e 2012, com a sazonalidade mensal do *cluster 5*. A distribuição pluvial dos meses de 2005 respeitou a curva de sazonalidade do período, destacando o mês de janeiro que apresentou precipitação acumulada de 434,00 mm, quase o dobro do esperado para o mês. Já o ano 2012 apresentou os menores volumes de precipitação para toda a série histórica e, diferentemente da análise anterior, a maioria dos meses não respeitou o índice de sazonalidade. Por exemplo, os meses de abril e outubro foram os mais secos da série, com respectivamente, 12,00 mm e 23,00 mm, enquanto que a precipitação média esperada era de 124,85 mm e 84,86 mm.

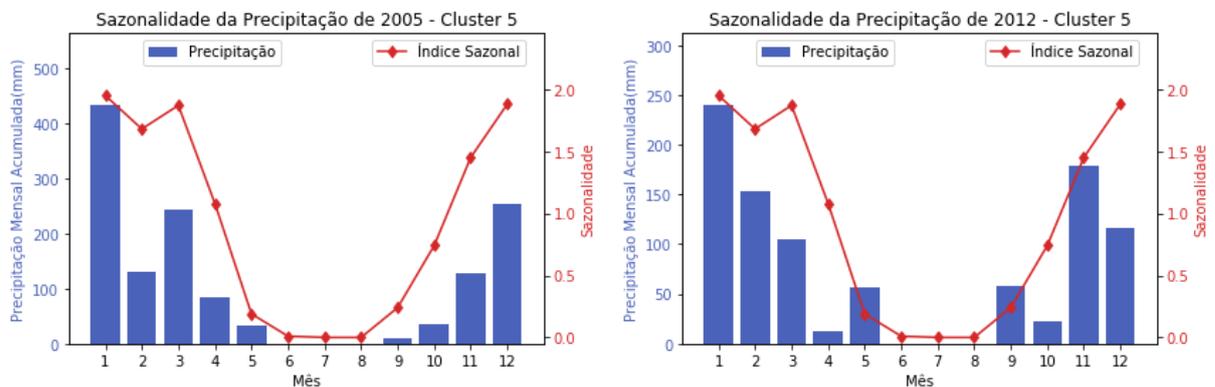


Figura 23 – Sazonalidade da precipitação dos anos de 2005 e 2012 para o *Cluster 5*.

Com o intuito de compreender as possíveis causas de influência na perturbação da distribuição das chuvas, selecionou-se os anos classificados como extremos, que foram indicados como muito secos e muito chuvosos, de toda a série temporal histórica de 1980 a 2019 do estado do Tocantins e comparou com os fenômenos atmosféricos-oceânicos (El Niño e La Niña) atuantes sobre a região.

A Tabela 15 apresenta uma relação dos anos de ocorrência, além da intensidade dos dois principais sistemas oceânicos atuantes para o período analisado. Analisando as

ocorrências, o El Niño apresentou variações de sete ocorrências moderadas, quatro fortes e apenas uma fraca. Já o La Niña apresentou menos registros do que o fenômeno anterior, com uma ocorrência fraca e todas as outras registradas como moderadas (INPE, 2020). É importante a análise destes fenômenos, pois o El Niño, responsável por forte seca, e o La Niña, causador de intensas chuvas, acabam intervindo na organização dos sistemas geradores de chuva do Tocantins (ROLDÃO; FERREIRA, 2019).

Tabela 15 – Período de ocorrência e intensidade dos fenômenos *El Niño* e *La Niña*.

Período	Ocorrência	Intensidade
1979-1980	El Niño	Moderada
1982-1983	El Niño	Moderada
1986-1987	El Niño	Moderada
1987-1988	El Niño	Forte
1990-1993	El Niño	Forte
1994-1995	El Niño	Fraca
1997-1998	El Niño	Forte
2002-2003	El Niño	Moderada
2006-2007	El Niño	Moderada
2009-2010	El Niño	Moderada
2015-2016	El Niño	Forte
2018-2019	El Niño	Moderada
1988-1989	La Niña	Moderada
1984-1989	La Niña	Fraca
1995-1996	La Niña	Moderada
1998-2001	La Niña	Moderada
2004-2007	La Niña	Fraca
2009-2011	La Niña	Moderada
2017-2018	La Niña	Moderada

Em relação aos anos classificados como “muito secos” identificou-se os seguintes anos: 1984, 1990, 1992, 1993, 1994, 1998, 1999, 2012, 2015, 2016 e 2019. os quais, estão em sua maioria relacionados com a ocorrência do El Niño, destacando os anos de 1990, 1992, 1993, 1994, 1998, 2015, 2016 e 2019. Com isso, identificou-se os anos noventa e a segunda década dos anos 2000 como os que apresentaram maiores registros de eventos de baixa precipitação. Além disso, estas foram as décadas que apresentaram o fenômeno El Niño com maiores ocorrências de intensidades moderadas e fortes.

Já os anos classificados como “muito chuvosos” variaram bastante, sendo identificados os seguintes: 1980, 1985, 1989, 1991, 1992, 1995, 2000, 2004, 2005, 2006, 2008, 2009 e

2011. Observa-se que alguns anos coincidem com os eventos do fenômeno La Niña como, por exemplo, os anos 1985, 1989, 1995, 2000, 2004, 2005, 2006 e 2011. Na primeira década do século atual, a distribuição da chuva de forma evidente teve relacionamento com o La Niña, que acabou aumentando a distribuição de precipitação do período com relação às outras décadas, assim como identificado por (KAYANO et al., 2016), que analisou a influência do fenômeno em relação aos últimos 30 anos.

4.4.2 Análise Mensal

A análise foi realizada com base nos resultados dos eventos extremos, utilizando a escala mensal, por meio da aplicação da técnica dos quantis, a partir das médias dos dados de precipitação pluvial dos meses correspondentes de cada *cluster*. Observa-se, nas Figuras 24 e 25, que as zonas homogêneas apresentam distribuição de precipitação, ao longo dos meses, bem definidas com relação às estações de chuva e seca.

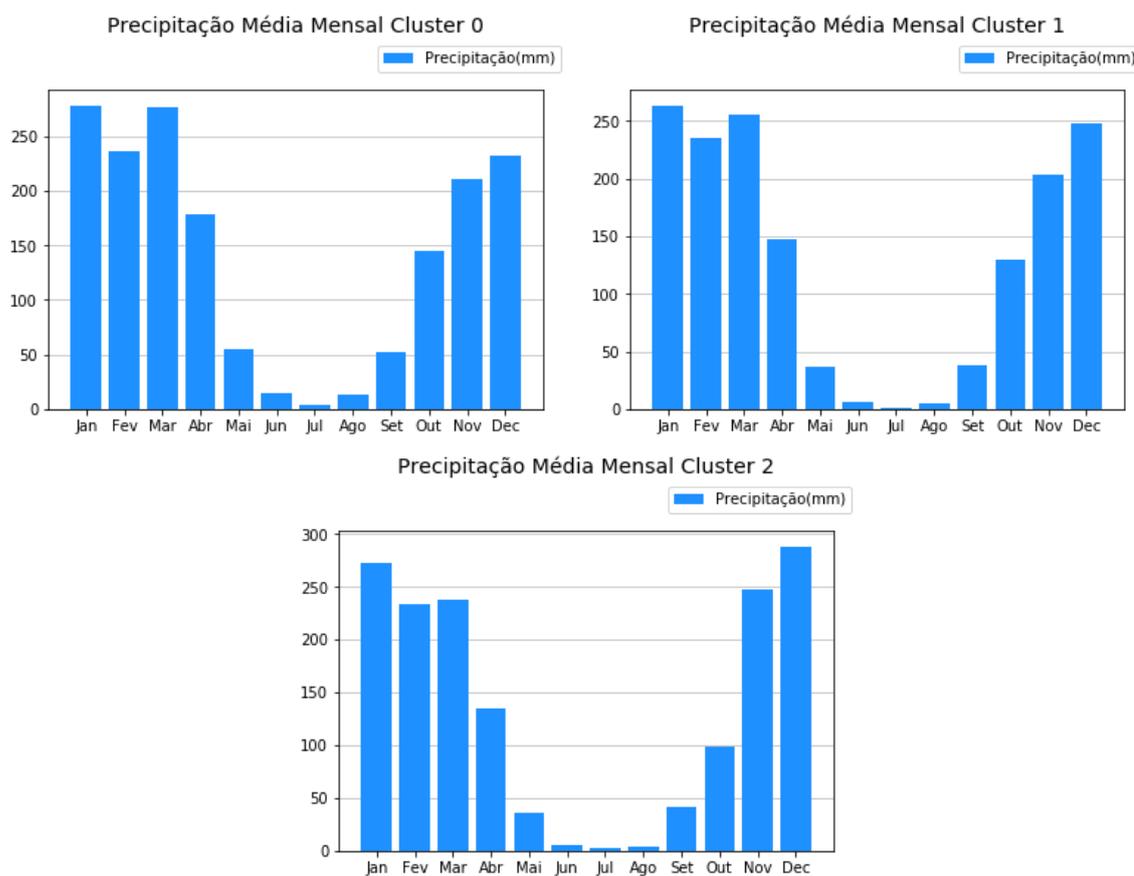


Figura 24 – Distribuição da precipitação mensal para as zonas homogêneas do primeiro vintênio.

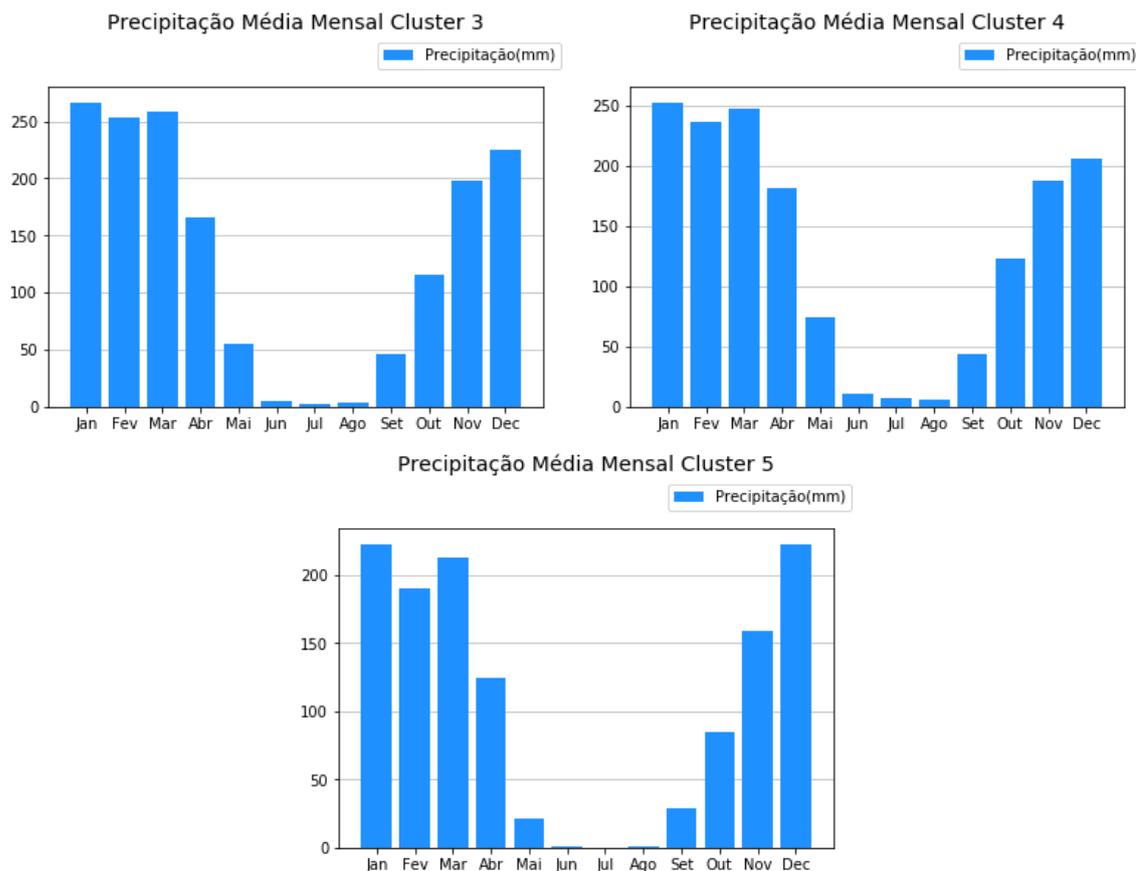


Figura 25 – Distribuição da precipitação mensal para as zonas homogêneas do segundo vintênio.

Os resultados sobre a distribuição pluviométrica, referentes as Figuras 24 e 25, mostram que a estação de seca varia de maio a outubro, enquanto a estação de chuva está relacionada com os meses de novembro até abril, sendo registrado esse regime para todos os *clusters*, assim como constatado na pesquisa de Roldão e Ferreira (2019), na qual os resultados demonstraram concordância em relação a distribuição de chuvas para o Tocantins.

Os limiares relacionados com os eventos extremos, para a classificação de “muito seco” e “muito chuvoso” se encontram presentes na Tabela 16. Com relação aos limites para o evento considerado “muito seco”, o menor limite inferior é o do *cluster* 5, com 0 mm, sendo a mesma região que na escala anual apresentou o menor valor, enquanto que o limite superior para esse evento foi igual a 5,46 mm, registrado pelo *cluster* 4. Já para o evento extremo considerado “muito chuvoso”, observou-se que o maior limite apresentado foi de 280,29 mm, que pertence ao *cluster* 2, e o que registrou o menor limite foi o *cluster* 5, assim como verificado no evento “muito seco”, com o registro de 224,66 mm.

Tabela 16 – Limiares de classificação mensal para os eventos extremos muito seco e muito chuvoso relacionados com cada *cluster*.

Cluster	Muito Seco(MS)	Muito Chuvoso(MC)
0	$X \leq 0,98$	$X \geq 273,49$
1	$X \leq 0,67$	$X \geq 261,18$
2	$X \leq 0,15$	$X \geq 280,29$
3	$X \leq 1,02$	$X \geq 273,09$
4	$X \leq 5,46$	$X \geq 261,07$
5	$X \leq 0,00$	$X \geq 224,66$

Os principais meses com ocorrência de eventos extremos “muito seco” e “muito chuvoso” foram avaliados visando compreender suas atuações sobre os *clusters*. A Figura 26 apresenta os meses do *cluster* 0 que foram classificados com “muito secos”, além do registro do ano, em que esses meses foram identificados. Analisando o gráfico em questão, nota-se que existe uma predominância de ocorrência desses eventos no mês de julho, que registrou 16 meses, o dobro do mês de junho, que é o segundo mês com maior quantidade de registros. Logo, constata-se que, para os anos dos *clusters* que sofreram com essa categoria de evento, a maioria registra precipitação abaixo de 0 mm.

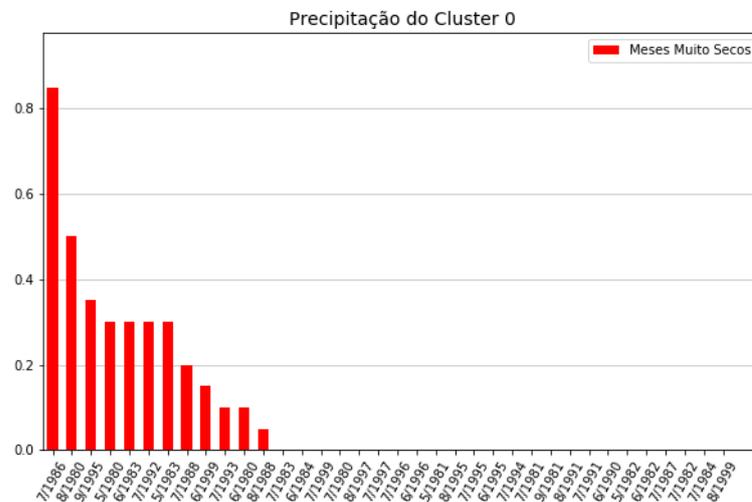


Figura 26 – Registro dos meses indicados como “muito secos” para o *Cluster* 0.

Para análise dos meses considerados como de maior precipitação do *cluster* 0, a Figura 27 apresenta os meses que foram classificados como “muito chuvoso”. Observa-se que aquele que apresentou maior chuva quantidade acumulada de chuva foi o mês de janeiro de 1985, com quase 600 mm de chuva, destacando que esse ano, assim como identificado na análise anual como o mais chuvoso do *cluster*, também apresentou cinco meses como os mais chuvosos no gráfico. Logo, o mês de março foi o que mais registrou

eventos de extrema precipitação, com 10 ocorrências, uma a mais que o segundo mês (janeiro).

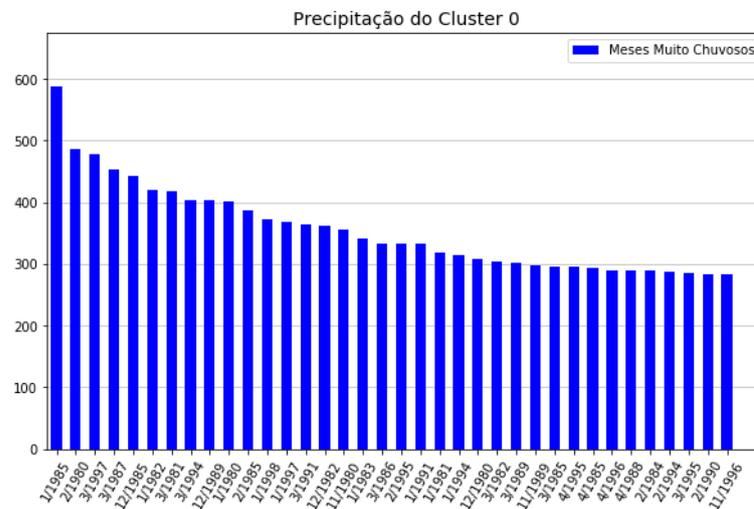


Figura 27 – Registro dos meses indicados como “muito chuvosos” para o *Cluster 0*.

A análise da Figura 28 é referente ao *cluster 1*, apresentando um aumento no número de meses com precipitação abaixo de 0 mm em comparação com análise anterior (*cluster 0*) para meses considerados como “muito seco”. Foi possível identificar que os meses de julho e junho acumularam juntos quase 70% dos registros dos anos considerados como de menor precipitação e, o mês de agosto ficou com 27,78% e o restante foi registrado para o mês de setembro.

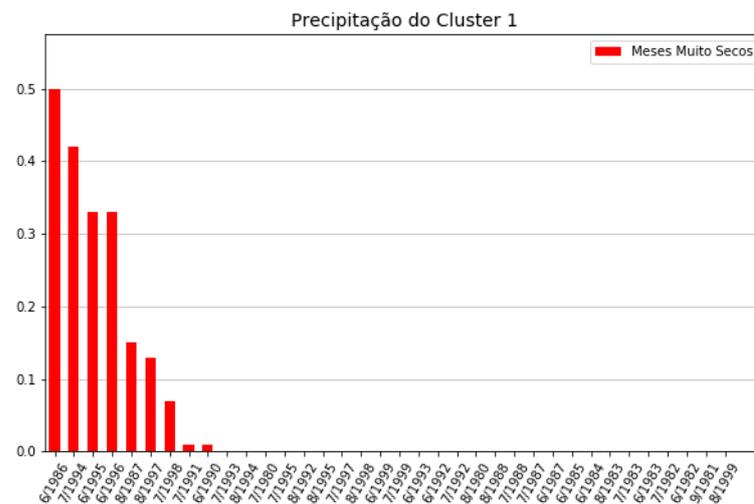


Figura 28 – Registro dos meses indicados como “muito secos” para o *Cluster 1*.

Ainda sobre o *cluster 1*, a Figura 29 ilustra os meses que foram considerados como “muito chuvosos”, cujos registros variam entre o mês mais chuvoso (dezembro de 1989),

com 620,68 mm, e o menos chuvoso (novembro de 1996), que possui 263,93 mm. Observa-se que entre os meses com maiores ocorrências de precipitação elevada, destacam-se janeiro e dezembro com, respectivamente, 10 e 8 meses presentes no gráfico.

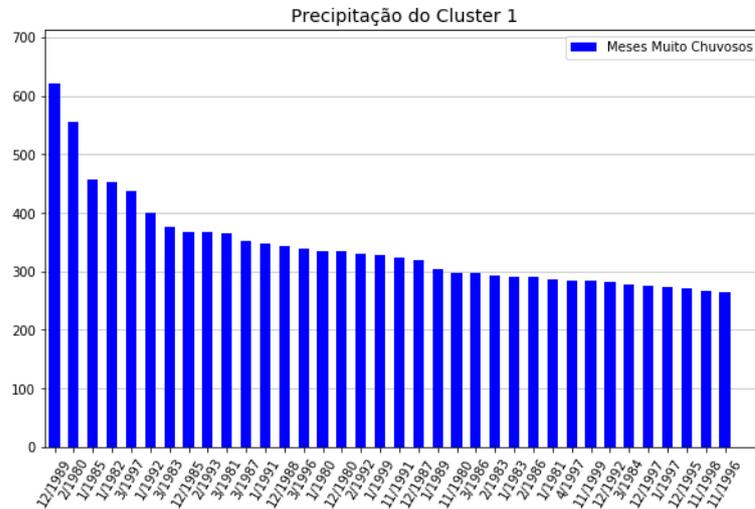


Figura 29 – Registro dos meses indicados como “muito chuvosos” para o *Cluster 1*.

O *cluster 2*, sendo este considerando o mais seco na análise mensal dos anos do primeiro vintênio, apresentou para os meses classificados como “muito secos” precipitação igual a 0 mm, conforme demonstrado na Figura 30. Destacam-se os meses junho, julho e agosto, pertencentes à temporada de seca, que representaram cerca de 92% dos registros da série. Além disso, eles aparecem em uma sequência de anos, de 1995 a 1998, onde o primeiro ano da sequência registrou quatro meses seguidos.

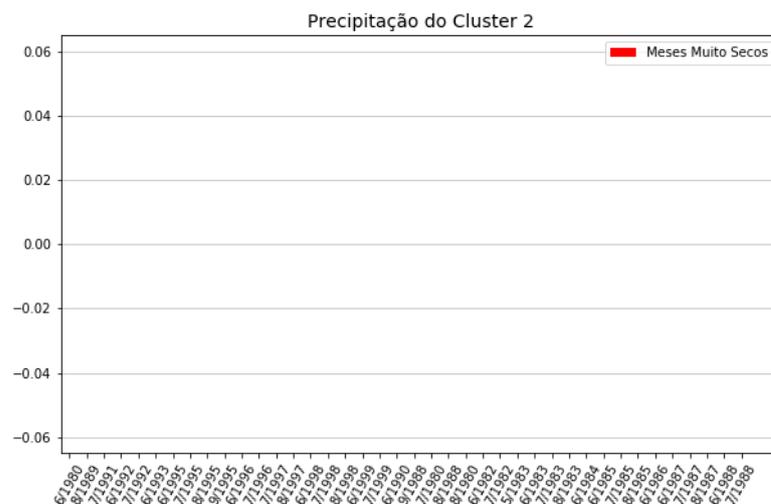


Figura 30 – Registro dos meses indicados como “muito secos” para o *Cluster 2*.

Com relação aos meses determinados como “muito chuvosos” para o *cluster 1*,

conforme apresentado na Figura 31, nota-se que novembro de 1992 apresenta a maior precipitação registrada entre todos os *clusters*, que é 766,54 mm, e fevereiro de 1991 apresenta o menor registro, que é igual a 281,94 mm. Além disso, a distribuição de ocorrências deste evento é bem distribuída entre os meses, havendo registro até mesmo para o mês de setembro, que se encontra presente no período de seca do estado.

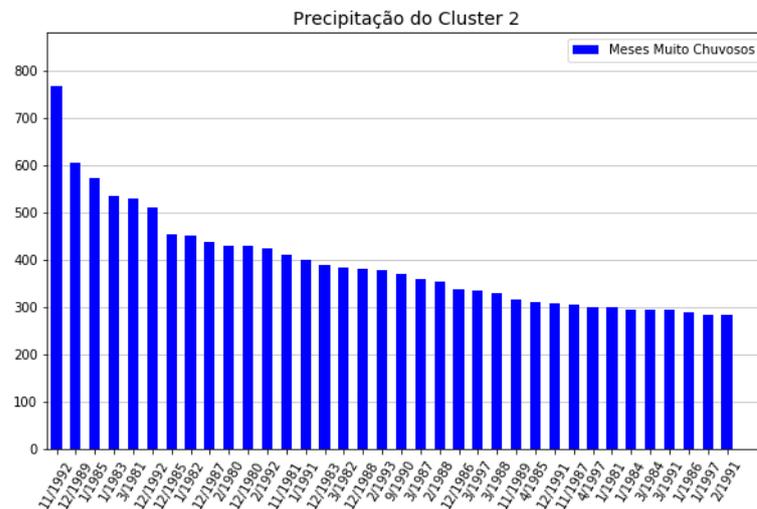


Figura 31 – Registro dos meses indicados como “muito chuvosos” para o *Cluster 2*.

Os índices de precipitação de chuva para o *cluster 3*, que está relacionado com o primeiro vintênio, para o período de baixa distribuição pluviométrica, se encontram registrados na Figura 32. Constata-se, inicialmente, que a maioria dos registros se encontram com precipitação zerada, sendo apenas 6 registros de precipitação acima de 0 mm. Além disso, o mês que apresentou maior quantidade de ocorrências foi julho, em comparação a análise dos *cluster* anteriores, possuindo 14 registros.

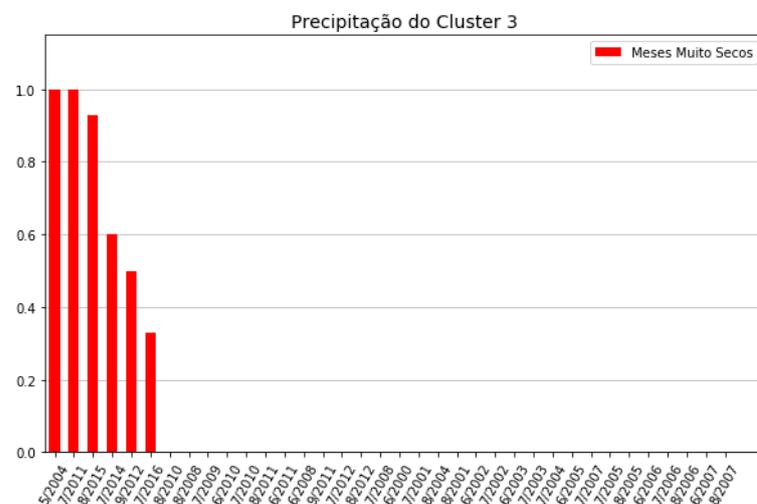


Figura 32 – Registro dos meses indicados como “muito secos” para o *Cluster 3*.

A Figura 33 corresponde ao gráfico dos meses classificados como “muito chuvosos” do *cluster* 3, onde a variação dos meses com a precipitação alta vai de 456,5 mm a 277 mm, respectivamente, para os meses de fevereiro de 2007 e novembro de 2012. Entre os meses identificados, destacam-se o conjunto entre os meses de janeiro, fevereiro e março, os quais representaram 72% da ocorrência desse evento extremo.

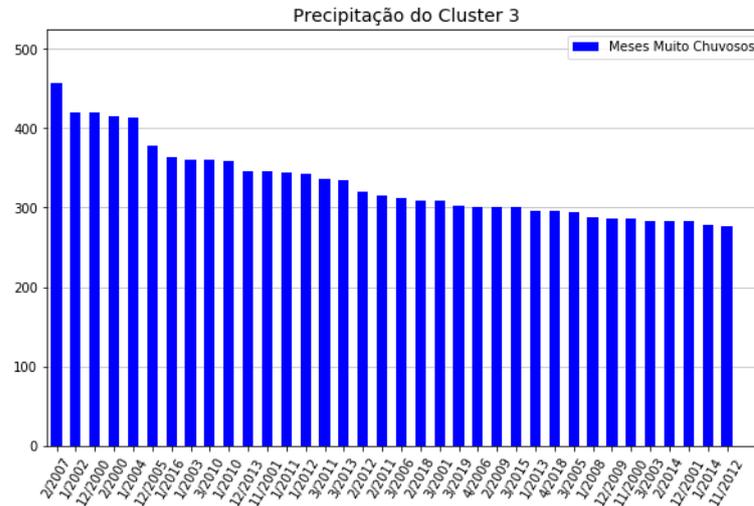


Figura 33 – Registro dos meses indicados como “muito chuvosos” para o *Cluster* 3.

O gráfico da Figura 34 contém registros dos meses considerados como “muito seco” encontrados nos *cluster* 4. Diferente dos outros *clusters*, esse registrou a maior quantidade de registros acima de 0 mm, ou seja, o limiar desse *cluster* foi o maior de todos. A quantidade de ocorrências desse evento extremo sobre o mês de agosto foi de quase 45%, com 16 registros. Na sequência tem-se os meses de junho e julho, ambos com 9 ocorrências e o restante está relacionado com o mês de setembro.

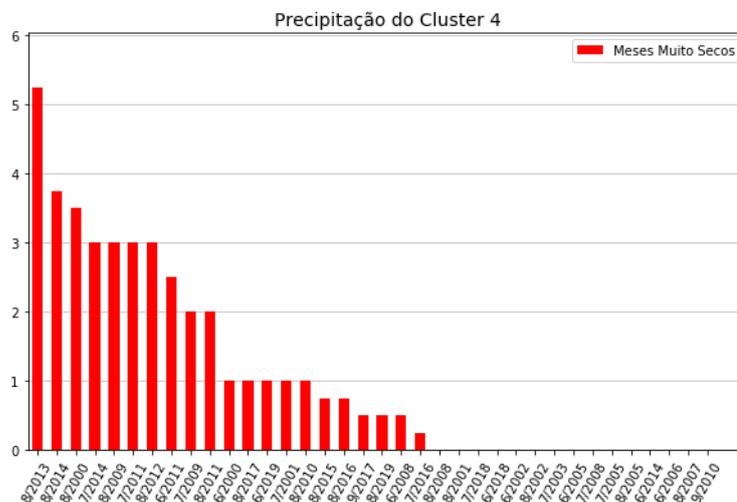


Figura 34 – Registro dos meses indicados como “muito secos” para o *Cluster* 4.

A Figura 35 apresenta as informações relacionadas com o registro de meses definidos como “muito chuvosos” para o *cluster* 4. Percebe-se que o mês considerado como mais chuvoso (fevereiro de 2007) foi o menor entre todos os *clusters*, com 392,50 mm registrados. Já o menor registro mensal do *cluster* foi em março de 2015 com 261,50 mm. Os casos de registro da ocorrência desse evento extremo foram bem distribuídos para os meses desta zona homogênea.

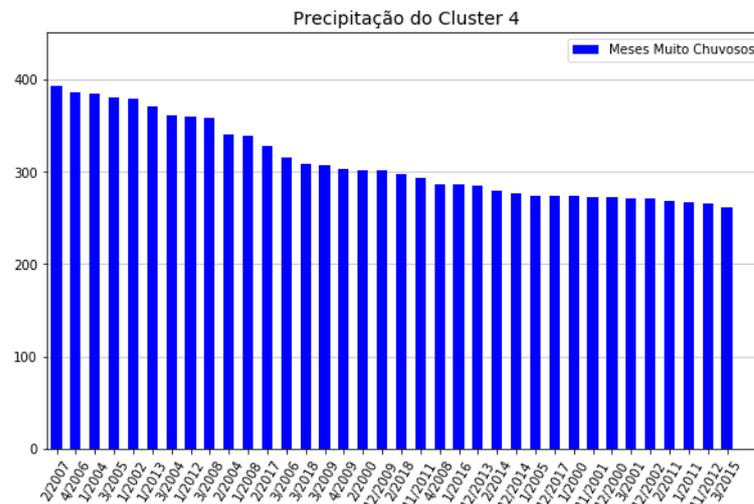


Figura 35 – Registro dos meses indicados como “muito chuvosos” para o *Cluster* 4.

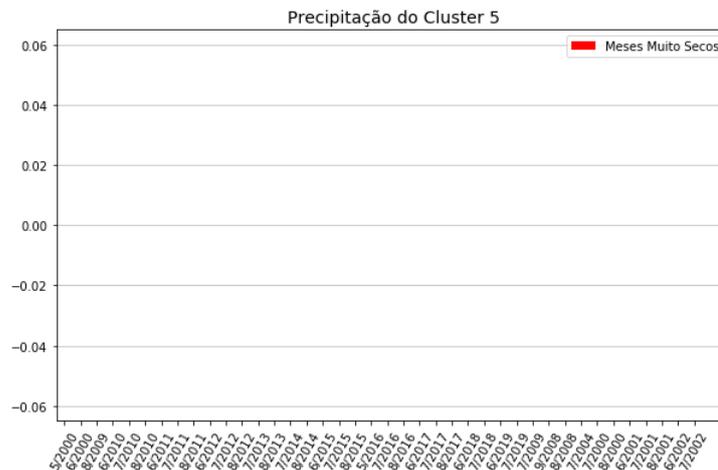


Figura 36 – Registro dos meses indicados como “muito secos” para o *Cluster* 5

A Figura 36 ilustra a variação da precipitação pluvial dos meses classificados como “muito secos” no *cluster* 5. Assim como no *cluster* 2, ambos estão presentes na mesma zona conjunta, o *cluster* 5 apresentou todos os registros abaixo de 0 mm, o que confirma a constatação da análise anual dos *cluster*, sendo esta a zona de menor precipitação. Analisando a distribuição do evento sobre os meses, verifica-se que os meses de junho,

julho e agosto unidos representam 91% de todas as ocorrências no *cluster*, sendo que os dois últimos estiveram presentes na sequência dos anos de 2010 a 2017.

Os dados presentes no gráfico da Figura 37 identificam os anos que foram considerados como “muito chuvosos” para o meses do *cluster* 5. Os registros variam da precipitação máxima de 518 mm do mês doze do ano 2000 à 228,43 mm do mês de novembro de 2018, sendo este o menor entre todos os meses classificados dos *clusters*.

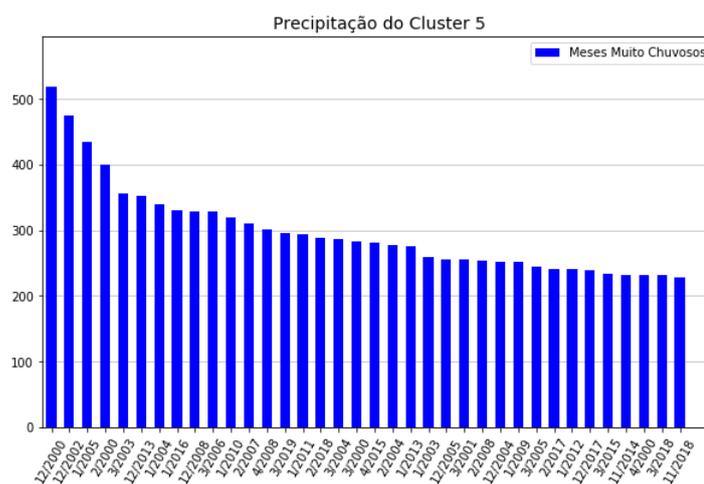


Figura 37 – Registro dos meses indicados como “muito secos” para o *Cluster* 5

Desta maneira, pode-se concluir que a partir da análise dos resultados dos gráficos referentes aos meses chuvosos e secos, que apontaram o regime de chuva como padronizado, com estações de chuva e seca bem estruturadas para os meses, seguindo um padrão de que o primeiro mês do ano possui alto índices de precipitação, em fevereiro a precipitação diminui, mas no mês seguinte sobe novamente, então quantidade de chuva decresce nos meses seguintes, quando chega em setembro o nível pluvial aumenta, seguindo essa relação para os meses seguintes até o último mês do ano.

Observa-se que, alguns meses da estação seca, apresentaram registros seguidos durante algumas sequências de anos, assim como alguns meses com excesso de chuva diferente do esperado, o que acaba impactando no setor que mais depende das condições climáticas, o agrícola, onde uma distribuição de chuva elevada ou mais escassa afeta o desempenho de uma safra (PORTAL AGRÍCOLA, 2021) (CANAL RURAL, 2019). Desta forma, resolveu-se avaliar a ocorrência de eventos extremos (“muito chuvoso” e “muito seco”), unificando todos os registros entre os *clusters* e calculando a porcentagem para cada mês, resultando na Tabela 17.

Tabela 17 – Registro de frequência dos eventos extremos para os meses de todos os anos de toda a serie temporal histórica.

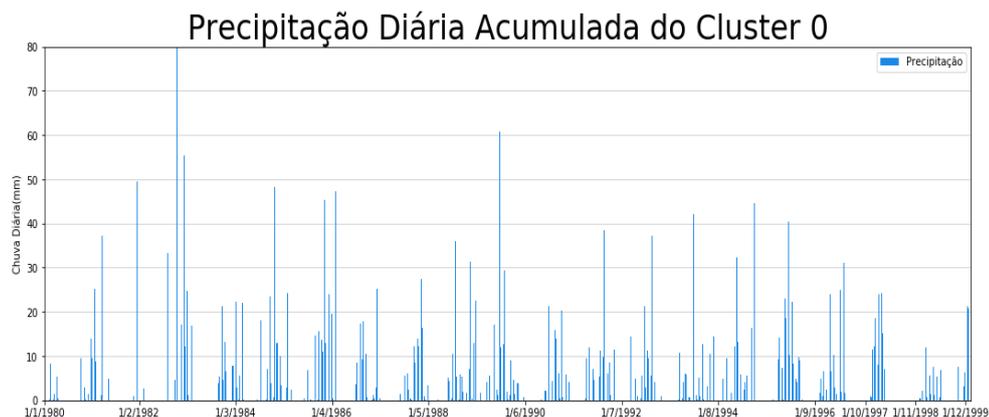
Meses	Muito Chuvoso(MC) Qtd.(%)	Meses	Muito Seco(MS) Qtd.(%)
Janeiro	25,58%	Maio	4,08%
Fevereiro	16,74%	Junho	28,16%
Março	22,79%	Julho	34,7%
Abril	6,98%	Agosto	28,98%
Setembro	0,47%	Setembro	4,08%
Novembro	9,3%		
Dezembro	18,14%		

Observa-se que, para o período chuvoso, a ocorrência de eventos de extrema precipitação é mais amplo sendo registrado em sete meses diferentes, com uma quantidade de registros totais iguais a 215. Ressalta-se que os meses de janeiro e março apresentaram mais ocorrências. Logo, pode-se entender como sendo os meses com maior tendência de ocorrer um evento extremo pluvial.

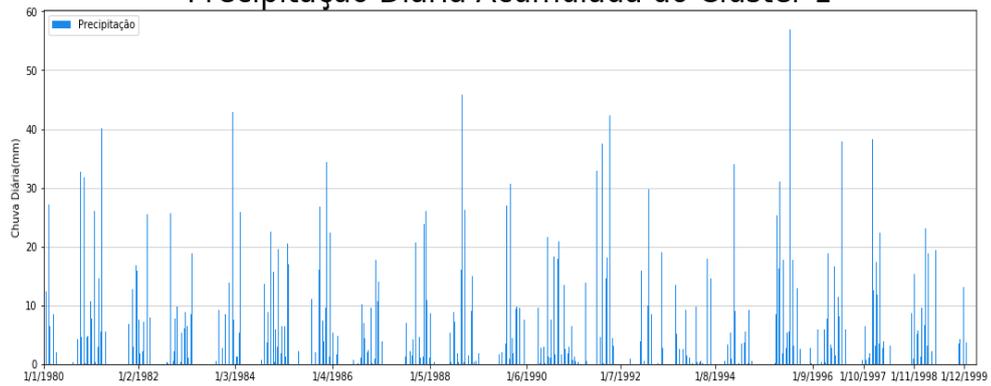
Por outro lado, a ocorrência de eventos extremos de seca apresentou uma quantidade inferior de meses afetados, sendo apenas cinco. Porém, obteve um maior quantitativo total de registros, 245. Nestes, os meses com mais registros foram julho, sendo este o de maior frequência, agosto e junho, com diferença mínima entre eles e, por fim, com a mesma quantidade de registros nos meses de maio e setembro.

4.4.3 Análise Diária

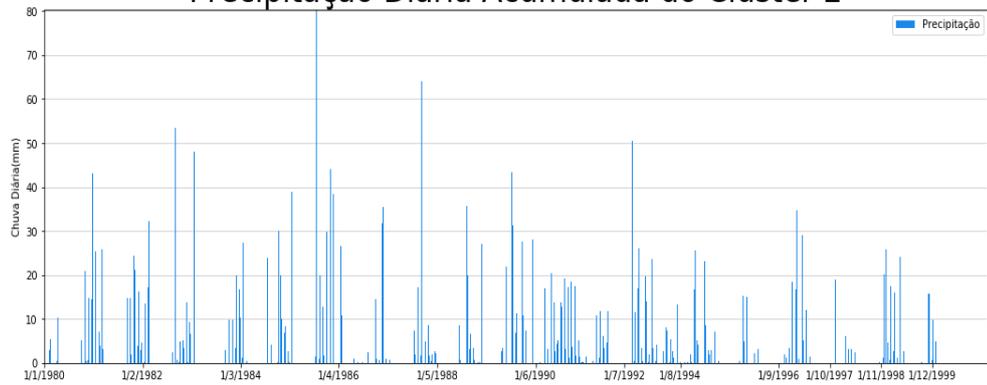
A Figura 38 apresenta a distribuição pluviométrica das zonas homogêneas presentes no primeiro e no segundo vintênio, sendo que estes se encontram em escala diária. Por meio da análise visual dos gráficos, percebe-se que os volumes de precipitação variaram muito em seu comportamento pluvial ao longo dos anos.



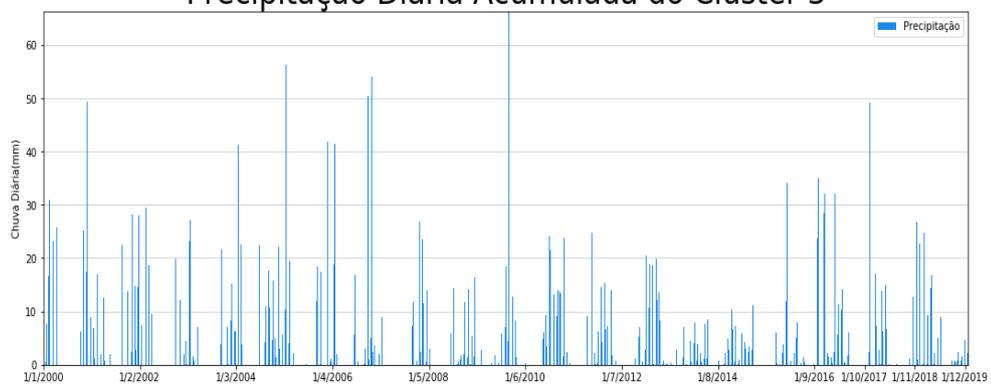
Precipitação Diária Acumulada do Cluster 1



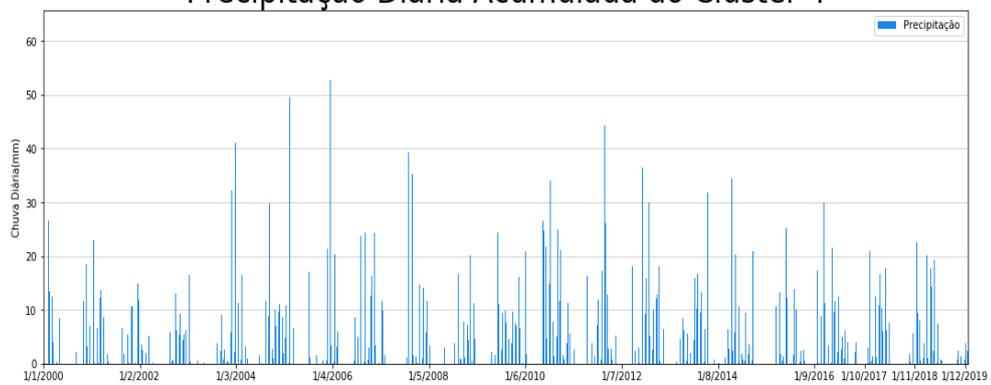
Precipitação Diária Acumulada do Cluster 2



Precipitação Diária Acumulada do Cluster 3



Precipitação Diária Acumulada do Cluster 4



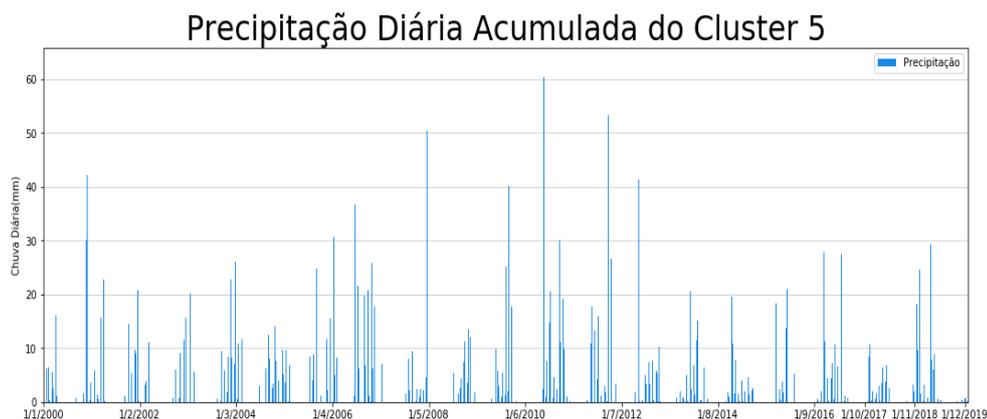


Figura 38 – Distribuição da precipitação diária para as zonas homogêneas do primeiro e segundo vintênio.

O resultado da aplicação da técnica dos quantis para os dados diários podem ser visualizados na Tabela 18, que mostra a ordem quantílica de classificação dos dias, onde diferentemente das análises anteriores, adotou-se uma escala diferente, na qual um evento extremo relacionado com a seca classifica-se pela ordem de $Q(0,05)$, o valor deve-se encontrar igual a 0mm, para todos os *clusters*. Já um evento extremo de muita chuva, a precipitação deve estar acima da quantidade estabelecida pelo quantil $Q(0,95)$, determinado por cada *cluster*. Desta vez, os eventos negativos foram classificados como “extremamente secos” e os eventos positivos foram descritos como “extremamente chuvosos”.

A quantidade de dias presente nos *clusters* do primeiro vintênio e do segundo foi igual, ou seja, um total de 7304 dias (vinte anos). Analisando o agrupamento de registros dos eventos extremos de todos os *clusters*, resulta em 23.570 registros, sendo que 21.405, cerca de 90,81%, equivalem aos eventos de extrema seca, enquanto 2.165 (9,16%) referem-se a eventos de extrema chuva.

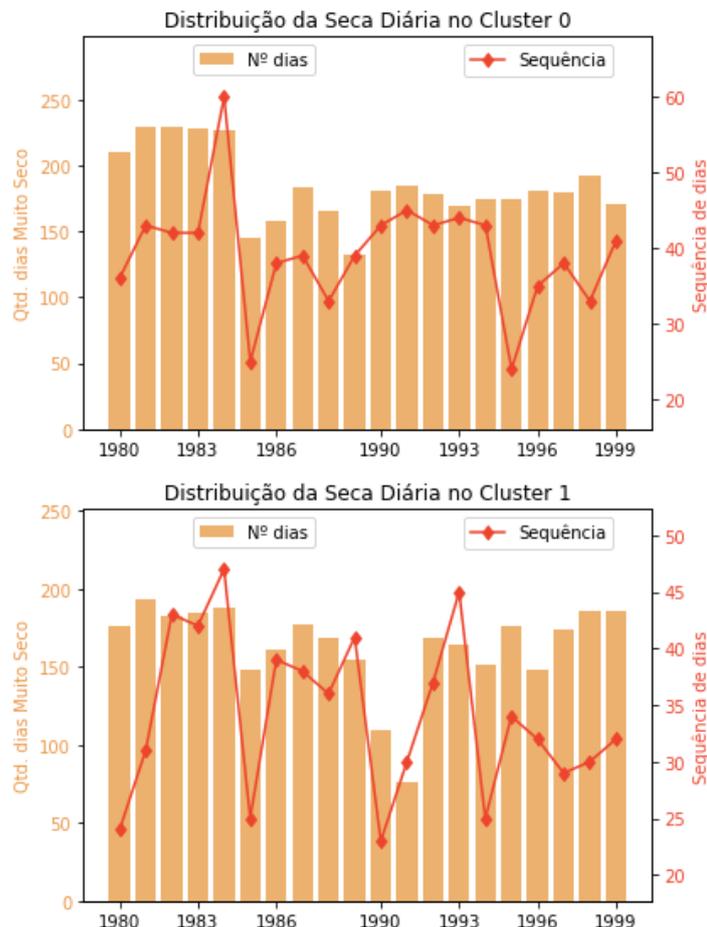
Tabela 18 – Limiares de classificação diário para os eventos extremos “Extremamente Seco” e “Extremamente Chuvoso” relacionados com cada *cluster*.

Cluster	Extremamente Seco(ES)	Extremamente Chuvoso(EC)
0	$X = 0,00$	$X \geq 24,55$
1	$X = 0,00$	$X \geq 21,60$
2	$X = 0,00$	$X \geq 26,00$
3	$X = 0,00$	$X \geq 22,89$
4	$X = 0,00$	$X \geq 21,59$
5	$X = 0,00$	$X \geq 20,79$

De forma individual, tem-se para o *cluster* 0, um total 4055 dias com eventos extremos, nos quais 3692 foram identificados como “extremamente seco” (91,05 %), e 363

dias estão relacionados com o evento “extremamente chuvoso” (8,95%). Para o *cluster* 1 foram identificados 3274 dias do evento de seca, o que representa 90,04%, enquanto que para o evento de chuva teve 362 (9,96%) dias. Já o *cluster* 2 apresentou a maior quantidade de eventos extremos 4358, sendo 4016 (92,15%) registros relacionados com seca e 342 dias de evento de chuva, correspondendo a 7,85%. Observa-se que no *cluster* 3 o evento “extremamente seco” registrou 3617 dias, o que representa 90,81% e 366 (9,19%) para o evento de “extremamente chuvoso”. O *cluster* 4 apresentou a menor quantidade de registros de eventos extremos, cerca de 3513, dos quais 89,58% e 10,42% representam, respectivamente, os eventos extremos de seca e de chuva. Ao final, no *cluster* 5 foram detectados 3659 dias, sendo 3147 (90,91%) classificados como “extremamente secos” e 366 (9,09%) como “extremamente chuvosos”.

Depois da quantificação dos registros de eventos extremos observa-se que a maioria dos registros superou a taxa de 90% de eventos de extrema seca, sendo que os *clusters* 2 e 5 apresentam os maiores índices deste evento e pertencem à mesma zona conjunta, além do *cluster* 0. Com o intuito de compreender o evento classificado como “extremamente seco” analisou-se a relação entre a quantidade de dias correspondentes a esta classificação com a sequência de registros de volumes abaixo de 0 mm dos anos referente ao primeiro e o segundo vintênio, conforme pode ser visualizado nas Figuras 39.



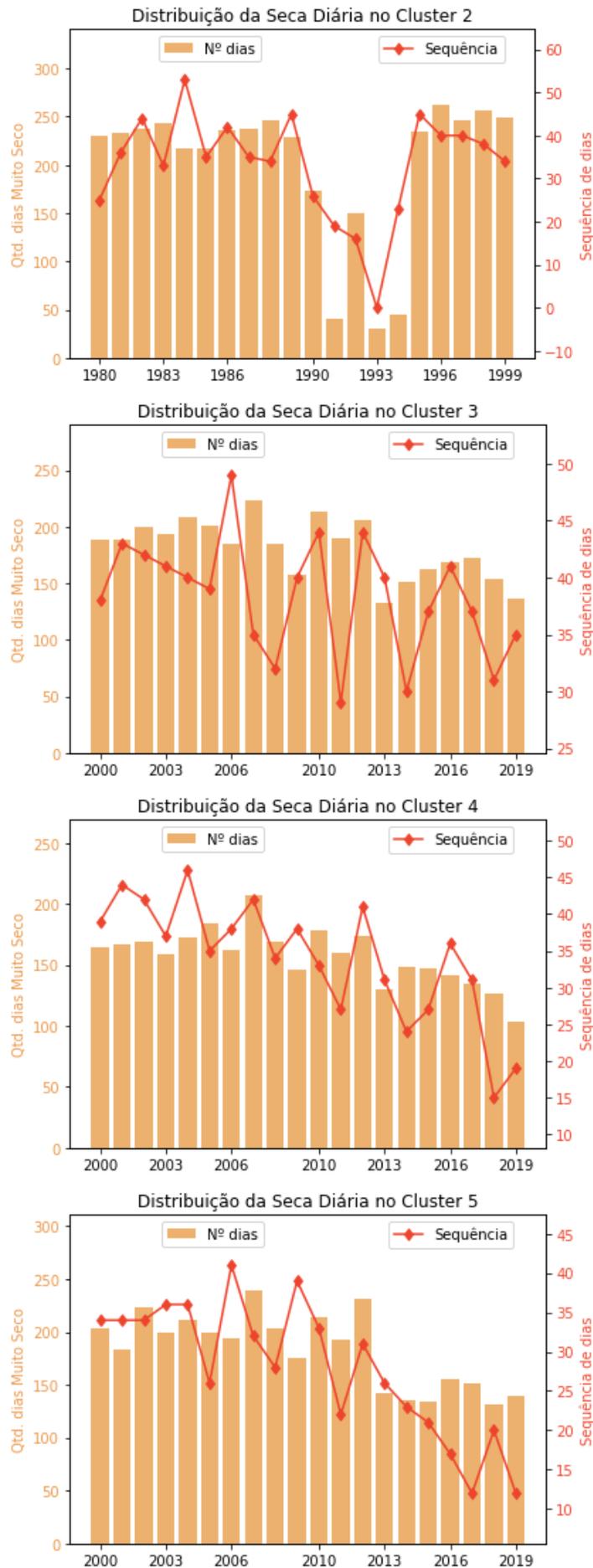


Figura 39 – Distribuição da quantidade de dias extremamente secos e o número de sequências com dias consecutivos para os *clusters*.

Comparando os resultados da análise diária com a análise anual pode-se observar os anos com maior quantidade de sequências ou maior número de dias consecutivos classificados como “muito seco” ou “seco”, em alguns casos, sendo que apresentam uma diferença mínima de volume para atingir a outra classificação. Considerando os gráficos acima, percebe-se que o *cluster* 0 apresentou a maior quantidade de sequências diárias (60 no total) no ano de 1984, o *cluster* 2 possui a maior quantidade de dias considerados “extremamente secos”, com um total de 262 registros correspondentes, sendo que na maior parte dos anos deste *cluster*, os registros apresentaram valores acima de 200 mm.

Com isso, não foi possível chegar a uma conclusão sobre a relação de sequência de dias e a quantidade de registros “extremamente secos”, pois, nem sempre um ano com o maior número de sequências registrou um número alto de dias de muita seca, o contrário também não ocorre. Resultados semelhantes foram encontrados por (DOURADO et al., 2013), ao analisar os dados referente a precipitação do estado da Bahia.

5 CONCLUSÃO

Com base nos resultados obtidos da pesquisa deste trabalho é possível validar a hipótese proposta inicialmente, que consiste na utilização de técnicas de mineração de dados para realizar uma análise climática no estado do Tocantins, a fim de obter resultados sobre o comportamento e frequência dos eventos climáticos extremos. Desta forma, por meio das variáveis climáticas de cada estação do estado do Tocantins, pode-se identificar padrões similares utilizando técnicas de mineração de dados e estatística, com o intuito de analisar o comportamento e a frequência dos eventos extremos que ocorrem durante toda a série de dados históricos.

Mesmo com estações com registros pertencentes a um período menor do que outras, foi possível realizar uma divisão em vintênios, sendo o primeiro referente ao período de 1980 a 1999 e o segundo referente aos anos de 2000 a 2010, ambos na mesma escala de 20 anos. Além disso, por meio da clusterização realizada pelo algoritmo *k-means* foram definidos três *clusters* para cada vintênio, atingindo, desta maneira, o primeiro objetivo específico proposto referente à identificação de zonas homogêneas.

A avaliação dos *clusters* foi realizada por meio de uma análise individual dos valores da variável chuva nas escalas anual, mensal e diária e uma análise conjunta, que buscou, a partir da sobreposição de todas os *clusters*, gerar três zonas (1-3), de modo a observar as variações das variáveis climatológicas ao longo dos vintênios. As variações do segundo vintênio em relação ao primeiro, em termos de precipitação média acumulada, mostra que somente a zona 2 apresentou crescimento, enquanto as outras zonas diminuíram, principalmente a zona 3. Nota-se que o segundo vintênio é considerado mais quente do que o primeiro, pois apresentou médias de temperatura máxima e mínima superiores aos do primeiro vintênio, além do que a umidade relativa do ar em todas as zonas no segundo vintênio foram inferiores.

Em relação à avaliação individual, por intermédio da técnica dos quantis, foi possível atingir os objetivos específicos de análise de comportamento e frequência em diferentes escalas (anual, mensal e diária), além de também, ter atingido o objetivo de encontrar os limiares de intensidade de chuva e seca. Logo, foram definidos limiares de classificação de eventos extremos para todas as escalas das zonas homogêneas de cada vintênio. Os limiares dos eventos extremos variaram em cada *cluster*, sendo identificados os anos com registros de precipitações extremas em todos *clusters*. Analisando a ocorrência dos eventos extremos “muito secos” constatou-se que os registros tiveram uma presença elevada nos anos 90 e na segunda década dos anos 2000. Por outro lado, os eventos considerados “muito chuvosos” estiveram presentes nos anos 80, e também na primeira década de 2000.

Analisando a correlação entre os eventos extremos de precipitação e os fenômenos atmosféricos influentes na região, observa-se que os anos caracterizados como “muito secos”

tiveram em sua maioria relação com o fenômeno El Niño, destacando os anos de 2016 e 1993. A ocorrência do fenômeno La Niña esteve ligada, também, com a maioria dos anos de eventos extremos “muito chuvosos”, o destaque é o ano de 1985, registrando os maiores volumes de precipitação em todos os anos dos *clusters* do primeiro vintênio. Com isso, ao final de todo o processo, o objetivo específico referente à avaliação e correlação das variações pluviométricas e os eventos extremos de cada zona foi atingido.

Na análise da escala mensal dos eventos extremos pode-se observar que, para todos os *clusters*, as estações tanto de seca quanto de chuvas encontram-se bem definidas. O evento extremo de baixa precipitação foi identificado em cerca de cinco meses de todas as séries temporais, sendo que os meses identificados com maior tendência a apresentar um evento extremo “muito seco” são junho, julho e agosto. Por outro lado, o evento extremo “muito chuvoso” esteve presente em sete meses, sendo os meses de janeiro e março aqueles que apresentaram maiores registros do evento.

Por fim, na análise em escala diária dos registros de eventos extremos acumulados de todos os *clusters*, mais de 90% estiveram relacionados com o evento de “extremamente seco”, enquanto que menos de 10% está relacionado com o evento “extremamente chuvoso”. Já em relação à quantidade de dias extremamente secos, o *cluster* 2 apresentou a maior quantidade de registros, além do que, os seus anos apresentaram, majoritariamente, o registro de 200 dias de seca.

5.1 Trabalhos futuros

Portanto, conclui-se que a aplicação das técnicas de mineração de dados apresentou resultados satisfatórios para análise climática, principalmente sobre a variável pluvial, no Estado do Tocantins, validando a hipótese e os objetivos deste trabalho. Ao final, sugerem-se para trabalhos futuros as seguintes ideias:

- Analisar as variáveis restantes (umidade relativa do ar, temperatura máxima, temperatura mínima e temperatura média) em diferentes escalas, assim como foi feito com a precipitação.
- Utilizar os dados em escala mensal, para encontrar padrões sequenciais da ocorrência dos eventos climáticos extremos em cada cluster.
- Realizar uma análise focada no fenômeno meteorológico verânico (período de estiagem na estação do inverno).
- Focar a pesquisa nos reais impactos na agropecuária do estado, relacionando os eventos identificados com os dados de produtividade.

REFERÊNCIAS

- BARBOSA, J. P. M. et al. Avaliação de técnicas empíricas e estatísticas de identificação de extremos de precipitação para o litoral paulista e entorno. [sn], 2008.
- BARBOSA, L. et al. Implicações climáticas de um modelo neodesenvolvimentista: Impactos, riscos e injustiças no estado do tocantins, brasil (climatic implications of a neo-developmental model: Impacts, risks and injustices in the tocantins state, brazil). **Revista Brasileira de Climatologia**, v. 19, 2016.
- BOSCHI, R. S. et al. Análise da precipitação pluvial e de veranicos no estado do rio grande do sul por meio de técnicas de mineração de dados. [sn], 2010.
- BUCENE, L. C. et al. Mineração de dados climáticos para previsão local de geada e deficiência hídrica. [sn], 2008.
- CAMILO, C. O.; SILVA, J. C. d. Mineração de dados: Conceitos, tarefas, métodos e ferramentas. **Universidade Federal de Goiás (UFG)**, p. 1–29, 2009.
- CANAL RURAL. **Soja: após 128 dias de seca e fim do vazio sanitário, Tocantins inicia plantio**. 2019. Disponível em: <<https://www.canalrural.com.br/programas/informacao/mercado-e-cia/soja-apos-128-dias-de-seca-e-fim-do-vazio-sanitario-tocantins-inicia-plantio/>>. Acesso em: 11 Jul. 2020.
- CANHOS, V. P. et al. Análise da vulnerabilidade da biodiversidade brasileira frente às mudanças climáticas globais. **Parcerias Estratégicas**, v. 13, n. 27, p. 113–148, 2010.
- CARVALHO, M. L. de et al. A seca no nordeste do brasil: Um estudo sobre as principais políticas públicas e métodos de previsão.
- CHAPMAN, P. et al. Crisp-dm 1.0: Step-by-step data mining guide. **SPSS inc**, v. 9, p. 13, 2000.
- CHARLES, A. C. **Mineração de dados para previsão de eventos extremos de precipitação**. Tese (Doutorado) — Universidade do Estado do Rio de Janeiro, 2015.
- CÔRTEZ, S. da C.; PORCARO, R. M.; LIFSCHITZ, S. **Mineração de dados-funcionalidades, técnicas e abordagens**. [S.l.]: PUC, 2002.
- DOURADO, C. d. S. et al. Mineração de dados climáticos para análise de eventos extremos de precipitação. [sn], 2013.
- ESRI. **ESRI – Gis and Mapping Software- Arc-Map**. 2020. Disponível em: <<https://www.esri.com/en-us/home>>. Acesso em: 11 Jul. 2020.
- FAET. **Tocantins decreta situação de emergência em decorrência da falta de chuvas**. 2016. Disponível em: <<http://www.faetrural.com.br/noticias-997-tocantins-decreta-situacao-de-emergencia-em-decorrenca-da-falta-de-chuvas.html>>. Acesso em: 11 Mar. 2021.

- FAYYAD. The kdd process for extracting useful knowledge from volumes of data. **Communications of the ACM**, ACM New York, NY, USA, v. 39, n. 11, p. 27–34, 1996.
- FLEISCHMANN, A. et al. Avaliação da seca de 2016 do rio javaés (bacia do rio araguaia) com uso de dados de múltiplos satélites. **Anais do XVII Simpósio Brasileiro de Sensoriamento Remoto**, INPE Santos, p. 5033–5040, 2017.
- FUNDO NACIONAL DA SAÚDE. **OMS lista as 10 principais ameaças para a saúde em 2019**. 2019. Disponível em: <<https://fns.saude.gov.br/ultimas-noticias/2375-oms-lista-as-10-principais-ameacas-para-a-saude-em-2019>>. Acesso em: 19 abr. 2020.
- IBGE. **Instituto Brasileiro de Geografia e Estatística**. 2020. Disponível em: <<https://www.ibge.gov.br/>>. Acesso em: 11 Jul. 2020.
- IDSO, C. et al. **Climate change reconsidered**. [S.l.]: The Heartland Institute, 2009.
- INMET. **INMET - Instituto Nacional de Meteorologia**. 2016. Disponível em: <<http://www.inmet.gov.br/portal/>>. Acesso em: 19 abr. 2020.
- INPE. **MONITORAMENTO DO EL NIÑO**. 2020. Disponível em: <<http://enos.cptec.inpe.br/>>. Acesso em: 11 Mar. 2021.
- IPCC. **IPCC – Intergovernmental Panel on Climate Change. Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation. A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change**. 2012. Disponível em: <<http://ipcc-wg2.gov/SREX/images/uploads/SREX-AllFINAL.pdf>>. Acesso em: 19 abr. 2020.
- ISP-GIFE. **MUDANÇAS CLIMÁTICAS**. 2019. Disponível em: <https://isppor.gife.org.br/isp_tema/mudancas-climaticas/>. Acesso em: 11 Ago. 2020.
- KAYANO, M. T. et al. El niño e la niña dos últimos 30 anos: diferentes tipos. **Instituto Nacional de Pesquisas Espaciais, INPE**, 2016.
- KOH, H. C.; TAN, G. et al. Data mining applications in healthcare. **Journal of healthcare information management**, Citeseer, v. 19, n. 2, p. 65, 2011.
- KOHAIL, S. N.; EL-HALEES, A. M. Implementation of data mining techniques for meteorological data analysis. **Implementation of data mining techniques for meteorological data analysis**, v. 1, n. 3, 2011.
- MARENGO, J. A. et al. Mudanças climáticas e eventos extremos no brasil. **Rio de Janeiro: FBDS**, 2009.
- MIRANDA, R. F. Miracema do tocantins: uma cidade em (des) construção. Universidade Federal do Tocantins, 2016.
- OGATA, R. A. et al. Situação epidemiológica da brucelose bovina no estado do tocantins. **Arquivo Brasileiro de Medicina Veterinária e Zootecnia**, SciELO Brasil, v. 61, p. 126–134, 2009.

OLIVEIRA, V. R. d.; SILVA, E. R. A. d. Ods 13: Tomar medidas urgentes para combater a mudança do clima e seus impactos: o que mostra o retrato do brasil? Instituto de Pesquisa Econômica Aplicada (Ipea), 2019.

ONU BRASIL. **Mudanças climáticas ameaçam agricultura familiar nas regiões Norte e Nordeste**. 2016. Disponível em: <<https://nacoesunidas.org/mudancas-climaticas-ameacam-agricultura-familiar-nas-regioes-norte-e-nordeste-diz-centro-da-onu/>>. Acesso em: 29 May. 2020.

PORTAL AGRÍCOLA. **Excesso de chuvas em importantes regiões reduz safra de soja do BR; fotos e vídeos**. 2021. Disponível em: <<https://www.noticiasagricolas.com.br/noticias/soja/281581-excesso-de-chuvas-em-importantes-regioes-reduz-safra-de-soja-do-br-fotos-e-videos.html>>. Acesso em: 11 Mar. 2021.

RODRIGUES, W.; DINIZ, B. C. Perspectivas de crescimento econômico no cenário amazônico: O caso do estado do tocantins. **Revista de Estudos Sociais**, v. 11, n. 22, p. 25–39, 2011.

ROLDÃO, A. de F.; FERREIRA, V. de O. Climatologia do estado do tocantins-brasil. **Caderno de Geografia**, v. 29, n. 59, p. 1161–1181, 2019.

RUIVO, H. M. Metodologias de mineração de dados em análise climática. **Doutorado em Computação Aplicada, Instituto Nacional de Pesquisas Espaciais, São José dos Campos (SP), Brasil**, 2012.

SCIKIT-LEARN: Machine learning em Python. v. 12, p. 2825–2830.

SEAGRO. **Condições climáticas podem frear crescimento expressivo da safra de grãos no Tocantins**. 2016. Disponível em: <<https://seagro.to.gov.br/noticia/2016/1/14/condicoes-climaticas-podem-frear-crescimento-expressivo-da-safra-de-graos-no-tocantins/>>. Acesso em: 19 abr. 2020.

SEAGRO. **SEAGRO - Exportações no primeiro trimestre de produtos agropecuários superam a marca dos 34% no Tocantins**. 2020. Disponível em: <<https://seagro.to.gov.br/noticia/2020/4/17/exportacoes-no-primeiro-trimestre-de-produtos-agropecuarios-superam-a-marca-dos-34-no-tocantins/>>. Acesso em: 19 abr. 2020.

SEFAZ. **SEFAZ - Produto Interno Bruto (PIB) do Tocantins**. 2019. Disponível em: <<http://www.sefaz.to.gov.br/estatistica/estatistica/produto-interno-bruto-pib-do-tocantins/>>. Acesso em: 05 may. 2020.

SEFAZ. **RELATORIO AGRICULTURA TOCANTINS ULTIMOS ANOS**. 2020. Disponível em: <<http://www.sefaz.to.gov.br/>>. Acesso em: 27 Jul. 2020.

SILVA, A. R. P. da; ALMEIDA, M. G. O agronegócio e o estado do tocantins: o atual estágio de consolidação. **Caminhos de Geografia**, v. 8, n. 21, 2007.

TAN, P.-N.; STEINBACH, M.; KUMAR, V. **Introduction to data mining**. [S.l.]: Pearson Education India, 2016.

UNFCCC. Fact sheet: Climate change science—the status of climate change science today. In: **United Nations Framework Convention on Climate Change**. [S.l.: s.n.], 2011.

XAVIER, T. d. M. B. S.; XAVIER, T. d. M. B. S. **A Técnica dos quantis e suas aplicações em Meteorologia, Climatologia e Hidrologia, com ênfase para as regiões brasileiras**. [S.l.]: Thesaurus Brasília, 2002.