

UNIVERSIDADE FEDERAL DE CAMPINA GRANDE
CENTRO DE TECNOLOGIA E RECURSOS NATURAIS
UNIDADE ACADÊMICA DE CIÊNCIAS ATMOSFÉRICAS

PROGRAMA DE PÓS-GRADUAÇÃO EM
METEOROLOGIA



Tese de Doutorado

Leandro Fontes de Sousa

**Formulação e Avaliação de Índices Preditores em Área
Semiárida do Brasil**

Campina Grande, outubro de 2019



UNIVERSIDADE FEDERAL DE CAMPINA GRANDE
CENTRO DE TECNOLOGIA E RECURSOS NATURAIS
UNIDADE ACADÊMICA DE CIÊNCIAS ATMOSFÉRICAS
PROGRAMA DE PÓS-GRADUAÇÃO EM METEOROLOGIA

Leandro Fontes de Sousa

Formulação e Avaliação de Índices Preditores em Área Semiárida do Brasil

Campina Grande
2019

Leandro Fontes de Sousa

Formulação e Avaliação de Índices Preditores em Área Semiárida do Brasil

Tese submetida ao Programa de Pós-Graduação em Meteorologia da Universidade Federal de Campina Grande para a obtenção do título de doutor em Meteorologia.

Orientadora: Prof^a. Dr^a. Maria Regina da Silva Aragão

Área de concentração: Meteorologia de Meso e Grande Escalas
Subárea: Sinótica-Dinâmica da Atmosfera Tropical

Campina Grande

2019

S725f Sousa, Leandro Fontes de.
Formulação e avaliação de índices preditores em área semiárida do Brasil / Leandro Fontes de Sousa. – Campina Grande, 2021.
147 f. : il. color.

Tese (Doutorado em Meteorologia) – Universidade Federal de Campina Grande, Centro de Tecnologia e Recursos Naturais, 2019.
"Orientação: Profa. Dra. Maria Regina da Silva Aragão". Referências.

1. Meteorologia de Meso e Grande Escalas. 2. Chuva - Petrolina.
3. Sinótica-Dinâmica da Atmosfera Tropical. 4. Índices de Estabilidade.
5. Análise Multivariada. 6. Percentil. 7. Normalidade Multivariada. I. Aragão, Maria Regina da Silva. II. Título.

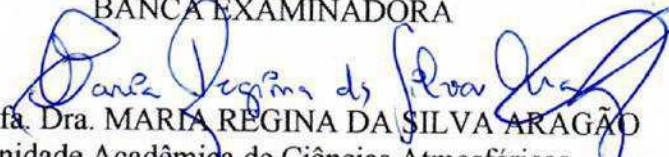
CDU 551.515.6(813.4)(043)


LEANDRO FONTES DE SOUSA

FORMULAÇÃO E AVALIAÇÃO DE ÍNDICES PREDITORES EM ÁREA
SEMIÁRIDA DO BRASIL


TESE APROVADA EM 25/10/2019

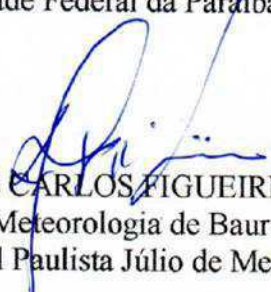
BANCA EXAMINADORA


Profa. Dra. MARIA REGINA DA SILVA ARAGÃO
Unidade Acadêmica de Ciências Atmosféricas
Universidade Federal de Campina Grande


Prof. Dr. ENILSON PALMEIRA CAVALCANTI
Unidade Acadêmica de Ciências Atmosféricas
Universidade Federal de Campina Grande


Profa. Dra. MAGALY DE FATIMA CORREIA
Unidade Acadêmica de Ciências Atmosféricas
Universidade Federal de Campina Grande


Prof. Dr. LINCOLN ELOI DE ARAÚJO
Departamento de Engenharia e Meio Ambiente
Centro de Ciências Aplicadas e Educação
Universidade Federal da Paraíba


Prof. Dr. JOSÉ CARLOS FIGUEIREDO
Centro de Meteorologia de Bauru
Universidade Estadual Paulista Júlio de Mesquita Filho

O melhor de mim é dedicado a:
meus pais, Agostinho (*in memoriam*) e Socorro;
irmãos;
demais familiares e amigos;
professores.

AGRADECIMENTOS

Agradeço primeiramente a Deus pela oportunidade de alcançar mais um objetivo da minha vida.

Aos meus pais, José Agostinho (*in memoriam*) e Maria do Socorro, por todo apoio, incentivo e ensinamentos, os quais levarei comigo pelo resto da vida.

Aos meus irmãos, por toda e qualquer forma de apoio que sempre me deram.

À Prof.^a Dr.^a Maria Regina da Silva Aragão, pela ética e profissionalismo com que conduziu o processo de orientação e pelas contribuições que tornaram possível o êxito desta pesquisa.

Aos professores e funcionários do Programa de Pós-Graduação em Meteorologia da UFCG, que contribuíram de forma direta ou indireta para realização da pesquisa.

À Coordenação de Aperfeiçoamento de Pessoal de Ensino Superior (CAPES) e ao Programa de Pós-Graduação em Meteorologia da Universidade Federal de Campina Grande (UFCG), pela concessão de bolsa de estudos.

À equipe da Biblioteca Central da UFCG, pelos serviços prestados à comunidade e que são essenciais para realização dos nossos estudos.

Sou filho do homem que conhecia o amor e o deu incondicionalmente a todos que ele amou. Saudade da minha luz. (SOUSA, 2020)

RESUMO

Índices de estabilidade são parâmetros numéricos obtidos de sondagens de ar superior, amplamente utilizados nos centros operacionais como ferramentas auxiliares para a previsão do Tempo, posto que evidenciam relações entre o estado termodinâmico da atmosfera e a precipitação. Nesta pesquisa, um conjunto de índices de estabilidade para a região de Petrolina é analisado com foco para meses de janeiro, mês da pré-estação chuvosa, marcado pela atuação de diversos sistemas meteorológicos e contrastes que caracterizam as condições de tempo dessa região. A base de dados utilizada é formada por sondagens diárias de ar superior de quinze meses de janeiro, observados no período de 1999 a 2016, juntamente com dados de precipitação. Técnicas de análise meteorológica são empregadas para diagnosticar as condições termodinâmicas locais, com foco nos dias antecedentes à chuva e nos dias com chuva na estação de Petrolina. Aplicou-se a transformação de Yeo-Johnson aos índices de estabilidade, com o objetivo de obter a normalização do conjunto de dados. O conjunto de índices transformados foi analisado por meio de ferramentas estatísticas, tais como a técnica dos quantis e a análise multivariada, com vistas a investigar a habilidade desses índices em detectar condições favoráveis a chuvas. Os resultados indicam que a técnica dos quantis facilita a compreensão dos índices como indicadores do grau de instabilidade e da ocorrência de chuva na área de Petrolina, através das ordens quantílicas de 20, 40, 60 e 80%. A análise de desempenho dos índices de estabilidade nos dias antecedentes à chuva e nos dias com chuva indica que os índices K e SWEAT são os melhores indicadores de instabilidade. A transformação de Yeo-Johnson foi eficiente para aproximar a distribuição dos índices de distribuição normal e mostrou que a variável transformada amplia os valores indicadores de instabilidade dos índices K, SWEAT e IC. O TMED* representou bem o efeito temporário da inversão térmica como condição favorável para a ocorrência de chuva (rompimento da camada estável). A Análise Discriminante do primeiro fator permitiu concluir que os grupos obtidos da análise de agrupamentos estavam bem discriminados e que os índices SWEAT*, K* TNCL* e IC* foram os mais importantes para a determinação da função discriminante padronizada.

Palavras-chave: Petrolina. Chuva. Índices de estabilidade. Análise multivariada. Percentil. Normalidade multivariada.

ABSTRACT

Stability indexes are numeric parameters obtained from upper-air reports in many operational stations and are widely used as auxiliary tools in weather forecasts, given that they state the relationship between atmospheric thermodynamics and precipitation. In this study, we analyze a range of stability indexes at play in the region of Petrolina (9 24'S, 40 30'W), with a special focus on January months, a pre-rainy season month that is locally determined by diverse meteorological systems that shape the weather conditions in the area. Our data base is formed by daily upper-air reports throughout the period of 15 January months, from 1999 to 2016, alongside with precipitation data. Weather forecasting techniques are used as to assess local thermodynamic conditions, by focusing on both the days prior to rain events and rain days catalogued by Petrolina rain-gauge station. Yeo-Johnson transformation was applied to the stability indexes, aiming at the normalization of data. The transformed indexes were then analyzed with the aid of statistical tools, such as Quantis techniques and multivariate analysis, as a way to evaluate whether such indexes are useful to determine conditions that precede rainfall days. By means of the quantile orders of 20, 40, 60 and 80%, our results indicate that both quantis analysis and multivariate analysis pave the way to a better understanding of such indexes as sound indicators for the assessment of instability levels and the prediction of rain occurrence in the Petrolina area. The performance analysis of stability indexes ranging from rain days to days prior to rain events indicates that K and SWEAT indexes are the best stability indicators. Yeo-Johnson transformation was efficient enough to approximate indexes distribution to that of normal distribution and have shown that the transformed variable amplify the instability indicator values of K, SWEAT and IC indexes. TMED* has satisfactorily addressed the temporary effects of thermal inversion as a favorable condition to the rain occurrence (by means of a disruption of the stable layer). The discriminant analysis of the preliminary factor allowed us to conclude that the groups we have obtained from groupment analysis have been efficiently discriminated and that SWEAT*, K* TNCL* and IC* were the most important indexes for the determination of standardized discriminant function.

Keywords: Petrolina. Rain. Stability indexes. Multivariate analysis. Percentage. Multivariate normality.

LISTA DE FIGURAS

Figura 1 – Perfis verticais para um dia com atmosfera estável (a) e para um dia com condição de instabilidade (b).	32
Figura 2 – Perfis verticais com indicação do TMED para um dia com atmosfera estável (a) e para um dia com condição de instabilidade (b).	35
Figura 3 – Perfis verticais com indicação do TNCL para um dia com atmosfera estável (a) e para um dia com atmosfera instável (b).....	36
Figura 4 – Perfis verticais com indicação do DPM para um dia com atmosfera estável (a) e para um dia com atmosfera instável (b).....	37
Figura 5 – Relação entre a variável original e a variável obtida pela transformação de Yeo-Johnson.	40
Figura 6 – Representação de um <i>boxplot</i> esquemático	52
Figura 7 – Série temporal do índice TNCL para o mês de janeiro de 2013. No eixo das abscissas (lado esquerdo), constam os valores do TNCL; no eixo à direita do gráfico, constam os valores das ordens quantílicas de 20%, 40%, 60%, 80% do TNCL.	53
Figura 8 – Distribuição dos grupos resultantes da ACP do primeiro fator. Linhas horizontais indicam as ordens quantílicas de 20%, 40%, 60% e 80%.	54
Figura 9 – Variáveis transformadas agrupadas de acordo com a AA de uma CP.....	55
Figura 10 – Distribuição do número de dias de chuva na Estação Meteorológica de Petrolina no período de janeiro de 1963 a dezembro de 2016.	57
Figura 11 – Distribuição do volume de precipitação diária na Estação de Petrolina.	58
Figura 12 – Distribuição temporal dos totais mensais de chuva na região de Petrolina. O asterisco “*” representa valores extremos e a bola “•” representa a média da distribuição.	59
Figura 13 – Séries temporais do DPM. Linhas tracejadas horizontais correspondem aos percentis 20%, 40%, 60% e 80 %. Dias com ocorrência de precipitação são sinalizados com linhas verticais pontilhadas.	61
Figura 14 – Relação entre o Total Totals e o Índice de Showalter.....	63
Figura 15 – Séries temporais do IC (TT-IS). Linhas tracejadas horizontais correspondem aos percentis 20%, 40%, 60% e 80 %. Dias com ocorrência de precipitação são sinalizados com linhas verticais pontilhadas.	65

Figura 16 – Séries mensais do IK. Linhas tracejadas horizontais correspondem aos percentis 20%, 40%, 60% e 80 %. Dias com ocorrência de precipitação são sinalizados com linhas verticais pontilhadas.....	68
Figura 17 – Séries mensais do Índice TMED. Linhas tracejadas horizontais correspondem aos percentis 20%, 40%, 60% e 80 %. Dias com ocorrência de precipitação são sinalizados com linhas verticais pontilhadas.	71
Figura 18 – Séries temporais do Índice TNCL. Linhas tracejadas horizontais correspondem aos percentis 20%, 40%, 60% e 80 %. Dias com ocorrência de precipitação são sinalizados com linhas verticais pontilhadas.	74
Figura 19 – Séries mensais do PII. Linhas tracejadas horizontais correspondem aos percentis 20%, 40%, 60% e 80 %. Dias com ocorrência de precipitação são sinalizados com linhas verticais pontilhadas.....	78
Figura 20 – Séries temporais do Índice DPM. Linhas tracejadas horizontais correspondem aos percentis 20%, 40%, 60% e 80 %. Dias com ocorrência de precipitação são sinalizados com linhas verticais pontilhadas.	81
Figura 21 – Séries temporais do KO. Linhas tracejadas horizontais correspondem aos percentis 20%, 40%, 60% e 80 %. Dias com ocorrência de precipitação são sinalizados com linhas verticais pontilhadas.....	83
Figura 22 – Séries temporais do SWEAT. Linhas tracejadas horizontais correspondem aos percentis 20%, 40%, 60% e 80 %. Dias com ocorrência de precipitação são sinalizados com linhas verticais pontilhadas.	86
Figura 23 – Séries temporais do DTel. Linhas tracejadas horizontais correspondem aos percentis 20%, 40%, 60% e 80 %. Dias com ocorrência de precipitação são sinalizados com linhas verticais pontilhadas.	88
Figura 24 – Indicativo de instabilidade dos índices nos dias sem chuva (s), dias de véspera de chuva (v), dias de chuva (c) e dias precedentes à chuva (pc).	92
Figura 25 – Gráficos de probabilidade normal (Q-Q plot) dos índices não transformados. A linha reta em azul representa a distribuição normal; e os marcadores em preto indicam como os índices em estudo estão distribuídos.	95
Figura 26 – Relação entre os índices originais e os índices transformados.	98
Figura 27 – Gráficos de probabilidade normal (Q-Q plot) dos índices transformados. A linha reta em azul representa a distribuição normal; e os pontos em preto indicam a distribuição dos índices.	102

Figura 28 – Diagrama <i>Scree Plot</i> , determinação no número de componentes principais adequadas para os índices.....	105
Figura 29 – Variância explicada pelas componentes principais retidas.....	105
Figura 30 – (a) Curva de inércia do agrupamento do primeiro fator e (b) dendrograma com a identificação dos grupos.....	107
Figura 31 – Distribuição das cargas fatoriais por grupo do primeiro fator: as linhas horizontais representam os percentis 20%, 40%, 60% e 80% dos escores da primeira CP e seus valores são apresentados no eixo vertical à direita da figura. .	108
Figura 32 – Variáveis transformadas relacionadas com a primeira CP, organizadas conforme o agrupamento dessa CP.....	110
Figura 33 – Distribuição dos dias de véspera de chuva por grupo do primeiro fator.....	112
Figura 34 – Distribuição dos dias de chuva por grupo do primeiro fator.....	114
Figura 35 – (a) Curva de inércia e (b) dendrograma do agrupamento do segundo fator.....	115
Figura 36 – Distribuição das cargas fatoriais por grupo do segundo fator: as linhas horizontais representam os percentis 20%, 40%, 60% e 80% dos escores da segunda CP e seus valores são apresentados no eixo vertical à esquerda da Figura.....	116
Figura 37 – Variáveis transformadas relacionadas com a segunda CP agrupadas conforme a AA desta CP.....	117
Figura 38 – Distribuição dos dias de véspera de chuva por grupo no segundo fator.....	118
Figura 39 – Distribuição dos dias de chuva por grupo no segundo fator.....	119
Figura 40 – (a) Curva de inércia e (b) dendrograma do agrupamento do terceiro fator.....	120
Figura 41 – Distribuição das cargas fatoriais por grupo do terceiro fator: as linhas horizontais representam os percentis 20%, 40%, 60% e 80% dos escores da terceira CP e seus valores são apresentados no eixo vertical à direita da figura. .	121
Figura 42 – Distribuição do indicativo de estabilidade por grupo no terceiro fator.....	122
Figura 43 – Distribuição dos dias de véspera de chuva por grupo no terceiro fator.....	122
Figura 44 – Distribuição dos dias de chuva por grupo no terceiro fator.....	123
Figura 45 – Q-Q plot multivariado para as variáveis relacionadas com a primeira CP. A linha vermelha representa a distribuição normal multivariada; a linha pontilhada representa um intervalo de confiança de onde os pontos nele contidos têm 95% de chance de pertencerem a uma distribuição normal.....	125
Figura 46 – Separação dos grupos pela primeira função discriminante. Os centroides são representados por números na cor preta.....	131

LISTA DE ABREVIATURAS

AA	Análise de Agrupamentos
ACP	Análise de Componentes Principais
AD	Análise Discriminante
CAPE	Energia Potencial Convectiva Disponível
CP	Componentes Principais
DMAP	Diferença Média entre Temperatura do Ambiente e a Parcela na camada 700-300 hPa
DPM	Depressão Psicrométrica Média da camada 700-500 hPa
DTel	Índice Delta-TetaE
IC	Índice Combinado
IK	Índice K
INMET	Instituto Nacional de Meteorologia
KO	Índice KO
LI	Índice de Levantamento
PII	Índice de Instabilidade Potencial
CCM	Complexo Convectivo de Mesoescala
SF	Sistema Frontal
SI	Índice de Showalter
SWEAT	<i>Severe Weather Threat Index</i>
TMED	Temperatura Média da Camada 700-500 hPa
TNCL	Temperatura ao Nível de Condensação por Levantamento
TT	Índice Total Totals
VCAN	Vórtice Ciclônico de Altos Níveis
ZCIT	Zona de Convergência Intertropical

LISTA DE SÍMBOLOS

*	Asterisco indica os índices transformados e, nos <i>boxplots</i> , os <i>outliers</i>
v	Véspera de dia com ocorrência de chuva
s	Dia seco
c	Dia com ocorrência de chuva
pc	Dia que sucede a ocorrência de chuva
ψ	Transformação de Yeo-Johnson
λ	Parâmetro da transformação de Yeo-Johnson
%	Porcentagem
°C	Grau Celsius
hPa	hectopascal
km ²	Quilômetros quadrados
N	Norte
S	Sul
E	Leste
W	Oeste
m	Metros
T	Temperatura do ar
Td	Temperatura do ponto de orvalho
Ur	Umidade relativa

SUMÁRIO

1	INTRODUÇÃO	15
1.1	OBJETIVOS	16
1.1.1	Objetivo geral.....	16
1.1.2	Objetivos específicos.....	17
2	REVISÃO DA LITERATURA.....	18
2.1	SISTEMAS METEOROLÓGICOS ATUANTES NA REGIÃO SEMIÁRIDA..	18
2.1.1	Zona de Convergência Intertropical.....	18
2.1.2	Vórtices Ciclônicos de Altos Níveis/ Cavado de Ar Superior	19
2.1.3	Sistemas Convectivos de Mesoescala	21
2.1.4	Sistemas Frontais.....	22
2.1.5	Índices de estabilidade.....	23
2.1.6	Aplicações da análise multivariada.....	25
3	DADOS E MÉTODOS	29
3.1	DADOS	29
3.2	METODOLOGIA.....	29
3.2.1	Seleção dos dados de altitude.....	29
3.2.2	Dados de precipitação	30
3.2.3	Estabilidade termodinâmica da atmosfera.....	30
3.2.4	Transformação de Yeo-Johnson.....	39
3.2.5	Análise multivariada	40
3.2.6	Gráficos <i>boxplot</i>	52
3.2.7	Método dos quantis.....	52
4	RESULTADOS E DISCUSSÕES.....	56
4.1	ESTUDO CLIMATOLÓGICO DA PRECIPITAÇÃO.....	56
4.1.1	Variabilidade do número de dias de chuva	56
4.1.2	Variabilidade dos volumes diários de precipitação	57

4.1.3	Variabilidade dos totais mensais	58
4.2	SÉRIES MENSAIS DE ÍNDICES DE ESTABILIDADE	60
4.2.1	Diferença média da temperatura do ambiente e da parcela	60
4.2.2	Índice Combinado	63
4.2.3	Séries mensais do Índice K	67
4.2.4	Índice da temperatura média da camada 700-500 hPa	70
4.2.5	Índice da temperatura do nível de condensação por levantamento	73
4.2.6	Índice de instabilidade potencial	76
4.2.7	Índice da depressão psicrométrica média da camada 700-500	80
4.2.8	Índice KO	83
4.2.9	Índice SWEAT	85
4.2.10	Índice Delta Teta-E	88
4.2.11	Desempenho dos índices de estabilidade	90
4.3	PRESSUPOSTOS DA ANÁLISE MULTIVARIADA.....	93
4.3.1	Análise dos <i>outliers</i>	93
4.3.2	Verificação da normalidade dos índices de estabilidade	95
4.3.3	Normalização pela transformação de Yeo-Johnson	97
4.3.4	Índices transformados	100
4.3.5	Estudo de normalidade dos índices transformados	102
4.4	ANÁLISE MULTIVARIADA I: ANÁLISE DE COMPONENTES PRINCIPAIS E ANÁLISE DE AGRUPAMENTOS	104
4.4.1	Análise de componentes principais dos índices transformados	104
4.4.2	Análise de agrupamentos do primeiro fator	106
4.4.3	Análise de agrupamentos do segundo fator	115
4.4.4	Análise de agrupamentos do terceiro fator	119
4.5	ANÁLISE MULTIVARIADA II: ANÁLISE DISCRIMINANTE.....	124
4.5.1	Análise Discriminante das variáveis relacionadas com o primeiro fator	124

4.5.2	Avaliação das técnicas estatísticas multivariadas.....	132
5	CONCLUSÕES.....	134
	REFERÊNCIAS.....	137

1 INTRODUÇÃO

Situado no interior da região Nordeste e em parte do norte da região Sudeste, o semiárido brasileiro é uma das regiões mais importantes do Brasil para estudos meteorológicos, já que é densamente povoada e está sujeita aos efeitos das variabilidades climáticas. Conforme o censo demográfico de 2010, essa região tem uma população de 22.598.318 habitantes e ocupa uma área de 980.133,079 km², o equivalente a 12% do território nacional (IBGE, 2010). Na região semiárida nordestina, o bioma nativo predominante é a Caatinga, um dos mais afetados pelas mudanças decorrentes da ocupação humana.

Um dos principais focos de pesquisa na região semiárida é o município de Petrolina, situado no extremo sudoeste de Pernambuco. Petrolina ocupa uma área de 4.561,872 km² e está inserida na mesorregião do São Francisco Pernambucano (IBGE – DGC/DEGEO, 1990). Conforme o Censo Demográfico 2010, o município, que tem 293.962 habitantes, está entre os mais populosos da porção submédia do vale do São Francisco (IBGE 2010).

A alta variabilidade da precipitação, uma das características marcantes do clima da região semiárida nordestina, tem motivado vários pesquisadores a estudar os fenômenos que atuam na região. Tais estudos são de grande significância para o conhecimento dos sistemas meteorológicos de grande e mesoescala que influenciam as condições do Tempo nessa área, mas ainda não explicam tais fenômenos na totalidade. Além disso, uma série de fatores justifica a realização de estudos meteorológicos para essa área do semiárido, tais como: uma melhor disponibilidade de dados meteorológicos; a contribuição do agronegócio para a economia do estado de Pernambuco, onde as culturas permanentes de manga e da uva possuem uma participação representativa no PIB do estado, contribuindo de forma expressiva para a geração de empregos na região; a densidade demográfica elevada, associada a dois grandes centros urbanos próximos (Petrolina-PE e Juazeiro-BA), o que torna essa região vulnerável aos efeitos de grandes acumulados pluviométricos e tempestades severas.

Nesse contexto, o surgimento de sofisticados modelos físico-matemáticos tem possibilitado um avanço significativo na previsibilidade das chuvas. Entretanto, o ajuste de tais modelos depende principalmente do conhecimento de processos energéticos, dinâmicos, termodinâmicos e também de padrões atmosféricos associados a eventos de chuva em várias escalas espaciais. Dessa forma, a busca de indicadores termodinâmicos representativos das situações atmosféricas associadas à ocorrência de chuva é fundamental para atingir uma melhor previsibilidade desses fenômenos.

Diante disto, o desenvolvimento desta pesquisa se justifica no aspecto científico, social e econômico, com o objetivo de avaliar o potencial preditor de índices de estabilidade atmosférica para a região semiárida nordestina. Investigou-se um conjunto de dados composto de índices de estabilidade já conhecidos na literatura, juntamente com novos índices formulados para a região semiárida nordestina em meses de janeiro, mês da pré-estação chuvosa caracterizado pela diversidade de sistemas meteorológicos que influenciam as condições do Tempo na área de Petrolina. A série de dados de ar superior compreende os meses de janeiro de 1999 a 2016. Dados de precipitação do período de 1963 a 2016 foram utilizados para o estudo da climatologia e para o estudo da relação da instabilidade atmosférica com a precipitação nos meses das sondagens de ar superior (meses de janeiro de 1999 a 2016). Desse modo, uma série de ferramentas estatísticas foram empregadas com o intuito de avaliar a relação entre índices de estabilidade e a ocorrência de precipitação na região de Petrolina.

1.1 OBJETIVOS

Nas subseções abaixo, apresenta-se o objetivo geral e os objetivos específicos desta pesquisa.

1.1.1 Objetivo geral

O presente estudo tem o objetivo de avaliar e formular índices de estabilidade para a região semiárida nordestina, bem como, investigar a habilidade de tais índices em representar o estado termodinâmico da atmosfera nos dias sem chuva, nos dias antecedentes à ocorrência de chuva e nos dias com ocorrência de chuva, utilizando técnicas de correção de normalidade de dados, técnicas dos quantis e técnicas estatísticas multivariadas, bem como avaliar o emprego dessas técnicas na identificação de índices de estabilidade que melhor representem a instabilidade atmosférica nas situações acima elencadas.

1.1.2 Objetivos específicos

- a. Analisar séries temporais de índices de instabilidade para meses de janeiro e sua relação como indicadores de ocorrência de chuvas na área Petrolina;
- b. introduzir novos índices de instabilidade para a área de Petrolina e investigar o desempenho desses índices quando nas ocorrências de chuvas em Petrolina;
- c. avaliar limiares dos índices de instabilidade para a região de Petrolina, de acordo com faixas de valores identificadas pela técnica dos Quantis;
- d. aplicar a transformação de Yeo-Johnson aos índices de instabilidade para obter a normalização do conjunto de dados e propor equações para a obtenção de índices de estabilidade transformados;
- e. utilizar técnicas estatísticas multivariadas para identificar grupos de dias de acordo com a similaridade dos indicadores de instabilidade dos índices transformados;
- f. utilizar técnicas estatísticas multivariadas para avaliar a relação entre a instabilidade atmosférica e a ocorrência de chuvas na área de Petrolina;
- g. avaliar se as técnicas estatísticas foram eficientes em filtrar os índices de estabilidade que melhor representem a instabilidade da atmosfera na região de Petrolina.

2 REVISÃO DA LITERATURA

Nesta seção, é feita uma revisão dos principais sistemas meteorológicos que atuam no semiárido brasileiro e também dos principais índices de estabilidade utilizados na previsão do Tempo, bem como do emprego de técnicas estatísticas multivariadas na Meteorologia.

2.1 SISTEMAS METEOROLÓGICOS ATUANTES NA REGIÃO SEMIÁRIDA

Numa revisão dos mecanismos dinâmicos que provocam chuva na região Nordeste, Molion e Bernardo (2002) destacaram a Zona de Convergência Intertropical (ZCIT), a Zona de Convergência do Atlântico Sul (ZCAS) e os sistemas frontais. A região também é influenciada por Sistemas Convectivos de Mesoescala (SCM), circulação de brisa e convecção local. Reboita *et al.* (2010) destacaram a ZCIT, os Vórtices Ciclônicos de Altos Níveis (VCAN) e os Sistemas Frontais (SF) como sistemas meteorológicos produtores de chuva na área do sertão nordestino.

2.1.1 Zona de Convergência Intertropical (ZCIT)

A ZCIT é uma banda de nebulosidade (contínua ou não) identificada nas imagens de satélite e que se estende na direção zonal na faixa equatorial do globo. A região de atuação da ZCIT coincide com a zona de confluência dos ventos alísios, onde a convergência de umidade é máxima com a área de maior temperatura da superfície do mar e com a área do cavado equatorial. A ZCIT é formada pela confluência entre os ventos alísios de nordeste e os ventos alísios de sudeste, o que resulta em movimento ascendente do ar com alto teor de vapor d'água. Ao ascender, o ar esfria, podendo formar nuvens. A ZCIT é um dos principais sistemas produtores de chuva no norte do Nordeste, atuando de fevereiro a maio, principalmente no Ceará, oeste do Rio Grande do Norte, interior da Paraíba, Pernambuco e norte da Bahia (ARAÚJO *et al.*, 2008).

Alguns estudos realizados para a região semiárida nordestina sugerem que a atuação simultânea da ZCIT e de sistemas meteorológicos de escala sinótica é frequente na estação chuvosa da região, principalmente nas situações de chuva acima da média. Moscati e Gan (2007) analisaram o vento no nível de 850 hPa e a precipitação no semiárido nordestino, com o objetivo de investigar os mecanismos responsáveis pela ocorrência de eventos extremos de chuva nessa área do território brasileiro. Os autores verificaram que a ocorrência de eventos extremos de precipitação pode ser substancialmente modulada pela atuação simultânea de dois

ou mais sistemas de grande escala – ZCIT, SF, ZCAS, VCAN, cavados de altos níveis, Alta da Bolívia, entre outros – e que os mesmos dependem da frequência e das interações entre esses sistemas. Seus resultados também sugerem que o deslocamento meridional da ZCIT não é o único fator determinante para a definição da qualidade da estação chuvosa na região semiárida, mesmo que a ZCIT mais próxima do equador possa afetar a precipitação na região. A relação entre o vento (à superfície) e a chuva em estações chuvosas contrastantes no Submédio São Francisco também foi analisada por Silva Aragão *et al.* (1997). As autoras verificaram que, nos meses com precipitação inferior à climatologia, a região era dominada por ventos associados à alta subtropical do Atlântico Sul (direção leste/sudeste). Por outro lado, nos meses com chuva acima da média, o vento apresentava outras direções devido à atuação de sistemas atmosféricos como a ZCIT, os VCAN e os SF oriundos de latitudes mais ao sul.

A interação da ZCIT com um Sistema Frontal (SF) foi estudada por Lyra *et al.* (2019). Os autores analisaram a atuação da extremidade de um SF sobre o Nordeste do Brasil (NEB) e sua interação com a ZCIT, resultando na formação de um Complexo Convectivo de Mesoescala (CCM). Conforme os autores, o ciclone frontal começou a se desenvolver no dia 10 de maio de 2014 e se intensificou até o dia 13 do mesmo mês. A banda frontal se estendeu até a região nordeste, favorecendo com que a ZCIT se deslocasse meridionalmente para sul ao longo dos dias, conectando sua nebulosidade à frente fria no dia 11 de maio de 2014. A convecção gerada pela interação entre os dois sistemas favoreceu a formação de um CCM nas primeiras horas do dia 12, com área aproximada de 57.000km², cuja propagação fora modulada pela corrente da ZCIT, fazendo com que o sistema se deslocasse para oeste.

2.1.2 Vórtices Ciclônicos de Altos Níveis (VCAN) / Cavado de Ar Superior

Os VCAN que atuam na região Nordeste são sistemas atmosféricos caracterizados por um centro de baixa pressão em altitude, cuja área central é mais fria do que a periferia. Esses sistemas pertencem à escala sinótica e geralmente se formam sobre o oceano Atlântico. Eles atuam principalmente nos meses de primavera, verão e outono, havendo uma maior frequência de ocorrência no mês de janeiro. Os VCAN apresentam uma área de nebulosidade em forma de arco, que indica a existência de movimento descendente no centro e movimento ascendente na periferia. No decorrer do seu ciclo de vida, eles podem se deslocar para leste, para oeste ou permanecer estacionários por alguns dias. Em um estudo sobre VCAN de origem tropical, Kousky e Gan (1981) propuseram um mecanismo para sua formação relacionado com a

amplificação do cavado de ar superior sobre o Nordeste-Atlântico Sul. Eles formularam a hipótese de que a penetração de sistemas frontais austrais está diretamente ligada à amplificação do cavado e que, portanto, ela contribui indiretamente para a formação do VCAN. Anjos (1995) diagnosticou que sistemas frontais austrais exerceram um papel importante no desenvolvimento de um VCAN sobre o Nordeste-Atlântico Sul, no período de 15 a 25 de novembro de 1986, corroborando essa hipótese.

Os VCAN são conhecidos como sistemas meteorológicos com dupla atuação; se, por um lado, são capazes de produzir grandes volumes acumulados de chuva devido à convecção ao longo de sua região periférica, por outro, podem inibir drasticamente a nebulosidade devido à subsidência de ar frio e seco na sua área central. Costa *et al.* (2014) simularam um evento de VCAN na região Nordeste, aplicando a teoria da máquina térmica baseada na segunda lei da termodinâmica, com o objetivo de analisar mecanismos de manutenção desse tipo de sistema. O modelo *Brazilian Developments on the Regional Atmospheric Modeling System* (BRAMS) simulou bem o evento ocorrido no período de 10 a 21 de janeiro de 2011, o qual mostrou sua maior intensidade no período de 16 a 19 de janeiro.

Dando continuidade a trabalhos anteriores (CORREIA, 1989; ARAÚJO, 1996; SILVA ARAGÃO *et al.*, 2000), Silva Aragão *et al.* (2007) investigaram as características em mesoescala e em escala sinótica de um dia chuvoso em janeiro de 1985 na região de Petrolina. Foram utilizados dados do radar meteorológico em banda-C instalado em Petrolina, dados de reanálise e dados observacionais de superfície e de ar superior. As autoras identificaram áreas de precipitação na escala meso- β , constituídas por núcleos convectivos profundos inseridos em áreas estratiformes. Na escala sinótica, elas mostraram que a área de cobertura do radar foi influenciada por um cavado de grande amplitude na alta troposfera e pela alta subtropical do Atlântico Sul enfraquecida, conforme evidenciado pela inversão de subsidência menos intensa e pelo alto teor de umidade e vento fraco com direção variável, na baixa e na média troposfera.

Utilizando sondagens atmosféricas realizadas no mês de janeiro no período de 2000 a 2014 na estação de altitude de Petrolina-PE, Correia *et al.* (2018) estudaram como a atuação de VCAN pode influenciar a estrutura da camada limite atmosférica e gerar condições favoráveis à formação de Duto de Superfície (DS). Seus resultados indicam que o número de DS é consideravelmente maior em meses com escassez de chuva e que, na ausência de atividade convectiva intensa, a influência das condições atmosféricas locais torna-se dominante na formação de DS. A persistência e a posição geográfica de VCAN foram fatores predominantes na formação de DS, mas não houve evidências de uma relação entre a frequência de VCAN na

região de estudo e as mudanças significativas na espessura e/ou intensidade dos DS.

2.1.3 Sistemas Convectivos de Mesoescala (SCM)

Os SCM também atuam no semiárido da região Nordeste. Existem dois mecanismos dinâmicos pelos quais os SCM podem se formar: o primeiro atua em mesoescala, quando uma parcela de ar potencialmente instável, situada próxima à superfície, é mecanicamente forçada a ascender em uma zona frontal ou, alternativamente, ascende devido ao deslocamento de uma corrente de jato nos altos níveis. Isto é mais frequente em regiões baroclínicas. O segundo mecanismo ocorre nas áreas onde o ambiente é barotrópico e se deve à dinâmica do próprio sistema convectivo. Após atingir a maturidade, surgem correntes descendentes frias (rajadas de vento da nuvem) nos SCM, formando então uma “bolha” de ar frio que funciona como uma rampa para as correntes ascendentes quentes e úmidas que alimentam o processo convectivo. Isto resulta em um movimento ascendente em mesoescala e na formação de uma grande área com nuvens estratiformes características (ANABOR, 2009). Ainda, de acordo com Anabor (2009), os SCM que se formam conforme a primeira situação são sistemas de grandes dimensões, favorecidos pelas características da topografia, situações sinóticas e circulações locais. Os SCM que obedecem à segunda condição de formação estão associados à própria convecção e à flutuabilidade das parcelas potencialmente instáveis próximas do SCM, assemelhando-se ao tipo de convecção que se desenvolve localmente ou aos núcleos que realimentam o sistema. Conforme Orlanski (1975), os SCM tropicais são fenômenos classificados entre as escalas meso- α e meso- β . Frequentemente, a evolução desses sistemas está associada à convecção profunda, cujo desenvolvimento ocorre normalmente devido ao aquecimento nos baixos níveis da atmosfera, ao umedecimento na camada limite e/ou ao movimento ascendente em grande escala.

Silva Aragão *et al.* (2000) analisaram ecos do radar meteorológico em banda-C instalado em Petrolina, com o objetivo de gerar informações sobre a estrutura de pequena escala e mesoescala dos sistemas causadores de chuva na região semiárida da região Nordeste. A análise de correlação entre a energia convectiva da parcela e o número de ecos em diferentes classes de área horizontal resultou em uma correlação linear positiva para os ecos com área menor ou igual a 400 km².

Silva *et al.* (2008) realizaram um estudo sobre sistemas convectivos linearmente organizados dentro de um círculo com raio de 250 km centrado em Petrolina. Os autores analisaram três dias (de 27 a 29 de março de 1985), investigando o papel desempenhado por

fatores locais e de grande escala no desenvolvimento desses sistemas. No cenário de grande escala, a área de cobertura do radar foi influenciada por um cavado de ar superior austral no dia 27 e por um Vórtice Ciclônico de Altos Níveis (VCAN) no dia 29. A convergência de umidade próxima à superfície foi favorável ao desenvolvimento de atividade convectiva nos dias 27 e 29, enquanto que a divergência de umidade próxima à superfície inibiu a atividade convectiva no dia 28. No cenário de mesoescala, foi observado que o aquecimento diurno é um fator importante para a formação de células convectivas, somando-se a ele o papel determinante da topografia.

A revisão apresentada nesta subseção mostra a diversidade de sistemas meteorológicos que influencia as condições de tempo na região Nordeste, evidenciando mecanismos envolvidos na sua formação, bem como as principais técnicas utilizadas no seu diagnóstico.

2.1.4 Sistemas Frontais (SF)

Um SF geralmente se forma quando o gradiente horizontal de temperatura se intensifica na região de transição entre duas massas de ar com características distintas. Na frente fria, na fronteira entre as massas de ar, o ar quente e úmido é forçado a ascender sobre a massa de ar frio do anticiclone situado na retaguarda da frente. Essa ascendência do ar quente e úmido produz uma banda de nebulosidade característica na área da frente fria, que pode ser facilmente identificada em imagens de satélite. As frentes frias são um dos principais mecanismos de mudança do Tempo na América do Sul. Elas atuam desde o sul do continente (onde têm influência mais marcante), até as latitudes equatoriais, onde geralmente chegam enfraquecidas, mas capazes de modificar o Tempo local, principalmente no que diz respeito à precipitação. Na revisão dos mecanismos dinâmicos que provocam chuva na região Nordeste, Molion e Bernardo (2002) destacaram que a penetração de SF ou de SF em dissipação, entre as latitudes de 5°S e 18°S, é um importante mecanismo de produção de chuva no sul e no leste da região Nordeste.

De acordo com Kousky (1979), incursões de SF em latitudes baixas produzem um efeito pronunciado na atividade convectiva da faixa tropical da América do Sul. O autor estudou incursões de sistemas frontais na região Nordeste, utilizando três observações diárias de superfície do período de 1961 a 1970. Ele mostrou que o avanço de SF favorece o deslocamento do cavado equatorial para latitudes mais ao sul, o que aumenta a precipitação no setor leste da região nos meses de outono e inverno; e que ocorre diminuição da pressão à superfície nas latitudes baixas em associação à penetração desses sistemas.

Souza (2003) investigou as formas de atuação de um sistema de latitudes médias que avançou sobre a região semiárida nordestina no início de junho de 1985. A autora diagnosticou, em um cenário de grande escala, que uma ciclogênese iniciada no dia 3 ao longo de uma banda frontal, situada a leste da região Sudeste, desencadeou uma sequência de eventos: (a) o avanço, até próximo de 7°S, de um cavado de latitudes médias com o eixo meridionalmente alinhado na média troposfera em torno de 45°W, estendendo-se desde a região Nordeste até o Atlântico Sul; (b) o enfraquecimento dos ventos alísios de leste-sudeste a partir do dia 4; (c) a convergência de umidade em baixos níveis sobre grande parte da região Nordeste no dia 6; e (d) uma extensa cobertura de nuvens nos dias 5 e 6, em contraste com os dias anteriores.

2.1.5 Índices de estabilidade

Vários indicadores obtidos de sondagens de ar superior são convenientemente expressos na forma de um valor numérico denominado índice de estabilidade. Tais índices foram introduzidos principalmente como ferramentas auxiliares na previsão de tempo, juntamente com determinadas técnicas de estudo, estando os valores de cada índice associados a uma característica de tempo significativo (AWS, 1979). Existem vários trabalhos na literatura utilizando índices de estabilidade, desenvolvidos para diversas regiões do globo.

Zawadzki *et al.* (1981), analisando sistemas convectivos extratropicais, encontraram uma alta correlação entre a quantidade de energia convectiva disponível (CAPE) e as taxas máximas de precipitação convectiva em um período de 24 horas. Entretanto, encontraram também um número significativo de dias com valores altos de CAPE sem ocorrência de precipitação convectiva, uma indicação de que a forçante dinâmica exerceu um controle maior na formação e na manutenção dos sistemas.

Diniz *et al.* (2004) investigaram as condições de tempo no aeroporto de Petrolina e a sua relação com a atividade convectiva intensa. As autoras utilizaram dados do radar meteorológico banda-C instalado no aeródromo e de observações de superfície e de ar superior. O grau de estabilidade da atmosfera foi avaliado por meio de índices de estabilidade e de perfis verticais das temperaturas potencial, potencial equivalente e potencial equivalente de saturação. Segundo as autoras, os resultados obtidos com os índices são satisfatórios no sentido de indicar situações favoráveis ao desenvolvimento de tempestades. Entretanto, são insuficientes para determinar a intensidade e a duração desses sistemas. Os índices são mais eficientes como previsores de tempestades quando usados em conjunto com outras informações meteorológicas e considerações de grande escala. Na maioria dos casos analisados, foi encontrado que eventos

significativos de cisalhamento vertical do vento estiveram associados com inversões térmicas e que valores altos da CAPE são uma condição necessária, mas não suficiente para o desenvolvimento de sistemas convectivos intensos.

Com o objetivo de ampliar a discussão sobre a previsibilidade de tempo severo no Brasil, Nascimento (2005) fez uma descrição de índices de estabilidade. Esses índices foram o K, SI, SWEAT, Total Totals (TT), IEH (Índice Energia-helicidade), Fawbush-Miller e Lifted. O autor concluiu, em uma primeira análise, que os índices estudados têm potencial para aplicação na previsão da atividade convectiva no setor subtropical do Brasil que abrange a região Sul, parte do Sudeste e o Mato Grosso do Sul.

Com o objetivo de avaliar a natureza de sistemas convectivos intensos detectados na região de alcance do radar de Petrolina-PE, Barbosa e Correia (2005) consideraram aspectos de grande e meso escalas e analisaram o efeito do cisalhamento vertical do vento e da CAPE na organização dos sistemas convectivos profundos. As autoras verificaram que valores da CAPE pouco acima de 1000 J kg⁻¹ podem ser um forte indicativo de atividade convectiva intensa nessa região e que, em situações semelhantes, deve-se considerar fatores de grande escala na análise de tempo significativo.

Tinmaker *et al.* (2017) utilizaram dados de satélite em ponto de grade, provenientes de um sensor de imagens de raios, para avaliar a relação entre precipitação e índices de estabilidade na região da Planície Indo-Ganjética na Índia. Os autores concluíram que o intervalo de tempo de dois meses do pico da densidade da taxa de raios pode estar atribuído a um *lapse-rate* do ambiente de mais de 7°C no período de pré-moção. A alta correlação estatística entre os dados do sensor e a temperatura do ar em 700 mb ajuda a confirmar essa hipótese. A distribuição anual do LI apresenta valores negativos de março a setembro e valores mais altos do IK. O atraso de fase com a monção deve-se à insolação intensa, à disponibilidade de umidade e a um elevado valor da CAPE.

Em estudo para a região da Finlândia, Ukkonen *et al.* (2017) utilizaram índices de estabilidade derivados de dados de reanálises do projeto ERA INTERIN para treinar redes neurais artificiais a identificarem a ocorrência de tempestades. Os autores concluíram que a aplicação das Redes Neurais para prever convecção e, em particular, o início da convecção, é um caminho promissor para pesquisas futuras, podendo-se utilizar saídas de modelo NWP para alimentar RNA. Os autores também concluíram que as variáveis relacionadas com a umidade se mostraram preditoras eficazes da ocorrência de raios e que a umidade relativa média da camada 800-600 hPa foi a segunda melhor entrada para as Redes Neurais.

Chakraborty *et al.* (2018) realizaram um estudo climatológico da instabilidade atmosférica usando dados diários de sondagens de alta resolução na região da Gadanki, de abril de 2006 a abril de 2017. Os autores abordaram quatro aspectos para avaliar os parâmetros da instabilidade atmosférica: (a) variação diurna; (b) variação mensal e sazonal; (c) variação pré-moção e transição; e, finalmente, (d) as tendências de longo prazo. Foi utilizado um conjunto de dez parâmetros de instabilidade tais como: Nível de Condensação por Levantamento (NCL), Nível de Convecção Livre (LFC), Nível de Equilíbrio (EL), Energia Potencial Convectiva Disponível (CAPE), Inibição Convectiva (CINE). Também foram utilizados índices de instabilidade tradicionais, tais como: Índice de Levantamento (LI), IK, TT, Índice Vertical Totals (VT) e Cross Totals (CT). As variações sazonais representam uma forte variabilidade na instabilidade, a qual é mascarada por uma variação diurna mais forte com o NCL e o LFC decrescente entre 11h e 18h (hora local), resultando em altos valores CAPE e chuvas fortes. No aspecto sazonal, os índices relacionados com as parcelas são altos durante, no final e após a monção, enquanto os parâmetros de instabilidade, como TT e VT, mostram os valores mais altos no período pré-moção, associados a forte convecção, ao passo que LFC e LCL diminuem juntamente com EL. Na variação de longo prazo (11 anos), os parâmetros LFC e LCL apresentam valores ligeiramente crescentes, estando os valores de EL em declínio, o que indica uma diminuição da instabilidade através do declínio do CAPE e K (KI), bem como o aumento no Índice de Levantamento (LI) e na Inibição Convectiva (CIN).

Umakanth *et al.* (2019) utilizaram índices de estabilidade provenientes de dados remotos dos satélites MODIS TERRA/AQUA e INSAT 3D para avaliar produtos e a previsibilidade de trovoadas na região de Andhra Pradesh, considerada uma região vulnerável da Índia no período pré-moção. Também foram utilizados dados de precipitação em pontos de grade do Departamento de Meteorologia da Índia e selecionadas as tempestades com mais de 20mm de chuva. Os autores estimaram limiares dos índices de estabilidade derivados dos satélites MODIS AQUA/TERRA e do INSAT 3D, os quais revelaram os limites associados à convecção severa, e concluíram que os resultados obtidos neste estudo forneceram uma evolução da previsibilidade da convecção, desempenhando um importante papel para a gestão de desastres.

2.1.6 Aplicações da análise multivariada

A análise multivariada é uma técnica muito usada na Meteorologia por permitir reduzir um grande conjunto de dados para um conjunto menor, sem que haja perda significativa da

informação contida nos mesmos. É grande o número de estudos realizados em todo o mundo, com diferentes finalidades, que fazem uso dessa técnica estatística.

Dalezios e Papamanolis (1991) avaliaram o desempenho de diversos índices de instabilidade atmosférica na previsão operacional da ocorrência de granizo no *Greek National Hail Suppression Program* (GNHSP). A avaliação dos índices foi realizada objetivamente através da análise multivariada. A ACP agrupou os índices em três fatores, tendo os índices com carga fatorial e escores elevados sido os melhores para prever convecção com formação e queda de granizo no *GNHSP*.

Ruiz (1997) relacionou a ocorrência de precipitação (preeditando), em várias estações meteorológicas na Argentina, com campos de vorticidade relativa na média troposfera (variável preditora) por meio de equações de diagnóstico obtidas utilizando a Análise de Discriminante Múltipla e a regressão de probabilidades de um evento. Seus resultados indicam que as porcentagens de acerto variam entre 60% e 70% e que este esquema estatístico possibilita um maior grau de acerto na previsão da ocorrência de chuvas quando comparado com a climatologia. Assim, ela concluiu que a implementação de métodos estatísticos objetivos de diagnóstico e/ou prognóstico de probabilidade de precipitação nos centros operacionais pode auxiliar na realização de prognósticos.

Mason (1998) desenvolveu um modelo de Análise Discriminante quadrática para obter previsões de chuva na África do Sul para períodos trimestrais, correspondentes às estações do ano (primavera, verão outono, inverno), e também para períodos semestrais. O modelo constitui um complemento útil para os modelos baseados em regressão utilizados por outros grupos de previsão sazonal na região da África do Sul. Seu modelo relaciona a precipitação em oito regiões da África do Sul com as Componentes Principais (CP) da temperatura da superfície do mar nos oceanos Índico, Atlântico Sul e Pacífico. O modelo apresenta níveis elevados de acerto em grande parte do país durante todo o ano. Após a validação para um período de 15 anos (1981-1995), foi constatado que o modelo pode ser utilizado com sucesso no ambiente operacional.

Tendo como objetivo determinar padrões de precipitação na região da Tessália (Grécia central), Stathis e Myronidis (2009) aplicaram a ACP a precipitações médias mensais de 75 estações meteorológicas para um período de 35 anos. As duas primeiras CP descreveram 87,7% da variância total dos dados. A primeira CP está relacionada com a frequência de ciclones que atingem a região e a segunda está ligada à frequência de anticiclones e tipos de tempo não-gradientes (que não têm padrão bem definido ou associado a qualquer sistema de

escala sinótica). Ou seja, a primeira componente explica o período chuvoso da região e a segunda componente, as chuvas de verão.

Santos (2013) identificou padrões espaciais e temporais e regiões homogêneas da precipitação e do Índice de Precipitação Normalizado (SPI) no Estado do Pará através da ACP e da AA. Foram utilizados dados mensais de 72 postos pluviométricos localizados no estado. As três primeiras CP explicaram 92,91% da variância total dos acumulados mensais de precipitação. A técnica da AA aplicada aos principais fatores comuns dividiu o Pará em três regiões homogêneas de precipitação. A primeira, situada no norte do estado, apresenta o regime pluviométrico influenciado principalmente pelo deslocamento norte-sul da ZCIT; a segunda está localizada na sua área central, onde a convecção local e a ZCAS são os principais mecanismos atmosféricos com chuva associada; e a terceira região está situada no sul do estado, onde as chuvas estão relacionadas com sistemas frontais e com convecção local.

Sousa (2015) fez uma análise das condições atmosféricas na área de Petrolina no mês chuvoso de abril de 2014. A aplicação da ACP em um conjunto de nove variáveis resultou em três CP, que retêm acima de 90% da variância total dos dados em cada nível isobárico analisado. A primeira CP retida explica as variações da temperatura do ar nos níveis de 925 hPa, 850 hPa e 200 hPa, bem como as variações do teor de umidade no nível de 500 hPa. A aplicação da AA aos fatores retidos identificou seis grupos de dias homogêneos em termos das condições atmosféricas, em cada nível isobárico. O autor também aplicou a ACP a um conjunto de dez índices de estabilidade que também resultaram em três CPs que retêm aproximadamente 83% da variância total dos dados. Neste caso, a aplicação da AA aos fatores retidos possibilitou identificar grupos de dias nos quais os índices representam alta ou baixa condição de instabilidade. O autor também constatou que os índices CAPE, LI, SI, K, TT e SWEAT são os que melhor representam características termodinâmicas e dinâmicas da atmosfera no mês de estudo, apresentando as maiores variações nos dias que antecedem os eventos de chuva, uma indicação de que eles podem ser úteis na previsão desses eventos na região semiárida do Nordeste do Brasil.

Lopes *et al.* (2019) utilizaram análise ACP e a análise de correlação de Pearson para avaliar o comportamento espacial da precipitação e da produtividade do milho no semiárido brasileiro para o período de 1990 a 2014. Os autores constataram uma alta variabilidade dos dados em escala interanual, revelando as áreas mais propícias ao cultivo do milho e as áreas de menor produtividade que eram, portanto, as mais afetadas pelo estresse hídrico. A variabilidade do rendimento do milho nas últimas décadas está intimamente relacionada com a variabilidade

climática, haja vista que se trata de uma agricultura de subsistência e em sistema sequeiro e que a eficiência da produção do grão é dependente das precipitações pluviais locais.

3 DADOS E MÉTODOS

3.1 DADOS

No presente estudo, são utilizados totais diários de precipitação registrados na estação meteorológica de superfície do Instituto Nacional de Meteorologia (INMET), localizada em Petrolina (9,38°S; 40,48°W; 370,46 m) (Figura 4.1). A série histórica de precipitação abrange o período de 01 de janeiro de 1963 a 31 de dezembro de 2016. Esses dados foram empregados em duas abordagens: a) estudo climatológico do período que abrange a série histórica (1963-2016); b) estudo da relação entre a instabilidade da atmosfera e a ocorrência de chuvas na região de Petrolina nos meses que contemplam os dados de altitude, como veremos abaixo.

Os dados de ar superior utilizados nesta pesquisa são medições de sondagens de altitude realizadas diariamente às 12 UTC na estação do INMET em Petrolina (9°24'S, 40°30'W, 370m). Eles estão disponíveis na página eletrônica da Universidade de Wyoming. Foram selecionados os meses de janeiro, de 1999 a 2016. Optou-se por estudar os meses de janeiro, pois este é um mês da estação chuvosa de Petrolina que é caracterizado pela atuação de diversos sistemas produtores de chuva nessa Região, cujos contrastes podem fornecer sinais significativos tanto da variabilidade dos índices como da eficácia desses. As sondagens fornecem dados de pressão, altitude, temperatura do ar (T), temperatura do ponto de orvalho (Td), umidade relativa (Ur), direção e velocidade do vento. Esses dados permitiram obter o conjunto de índices de estabilidade utilizados nesta pesquisa, o qual será apresentado mais adiante.

3.2 METODOLOGIA

A seguir, apresenta-se a descrição dos métodos utilizados neste estudo, tais como: a) os critérios para a escolha dos meses com as sondagens de ar superior; b) a utilização dos dados de precipitação, índices e estabilidade; c) a normalização dos índices com a transformação de Yeo-Johnson; e, por fim, d) a utilização das técnicas estatísticas multivariadas.

3.2.1 Seleção dos dados de altitude

São utilizados dados de ar superior para meses de janeiro, início da estação chuvosa de Petrolina. Ao todo, foram escolhidos 15 meses de janeiro, entre 1999 e 2016. Os meses

selecionados para análise são aqueles que estão completos ou com, no mínimo, 70% dos dias com observações de altitude. O mês de janeiro de 2004 também foi incluído em função da disponibilidade de estudos realizados na região para esse mês, apesar de ter menos de 70% de dias com sondagens de altitude disponíveis. O conjunto de dados total possui 408 sondagens de ar superior, o que representa 87,74% dos dias com observação dos presentes dados. Os janeiros dos anos de 2002, 2006 e 2010 não foram considerados na análise devido a baixa disponibilidade de sondagens de altitude.

3.2.2 Dados de precipitação

Os dados de precipitação analisados nesta pesquisa são utilizados em duas abordagens. A primeira é o estudo climatológico da precipitação na Estação de Petrolina. A série de dados diários de precipitação considerada neste estudo abrange o período de janeiro de 1963 a dezembro de 2016. Foram contabilizados o número de dias de chuva em cada mês e os totais mensais de cada ano. Para avaliar a variabilidade anual da precipitação na Região de Petrolina, foram confeccionados gráficos do tipo *Boxplot* contendo o número de dias de chuva, os totais mensais e os totais diários acima de 1mm. Os dados de precipitação também foram utilizados para investigar a relação entre os índices de estabilidades derivados das sondagens de altitude e a ocorrência de precipitação. De posse da planilha com os índices de estabilidade compreendendo os meses de janeiro de 1999 a 2016, adicionou-se uma coluna com a precipitação correspondente a cada dia com sondagem e outra coluna com a classificação dos dias com base na ocorrência de precipitação. Os dias sem chuvas são classificados como “seco (s)”; os dias antecedentes à ocorrência de chuvas são classificados como “véspera(v)”; os dias com chuva são classificados como “chuvosos (c)”; e os dias posteriores às chuvas ou sequências de dias com chuva são considerados “dias pós-chuva (pc)”.

3.2.3 Estabilidade termodinâmica da atmosfera

3.2.3.1 Diagramas termodinâmicos

Nesta pesquisa, a estabilidade da atmosfera foi analisada com o objetivo de identificar e prever situações favoráveis às chuvas no semiárido brasileiro, onde a precipitação é predominantemente de natureza convectiva e o conhecimento dos processos termodinâmicos é fundamental (CORREIA, 1989; ARAÚJO, 1996; SILVA ARAGÃO *et al.*, 2000).

O surgimento dos diagramas termodinâmicos possibilitou o desenvolvimento de indicadores para analisar as condições de estabilidade da atmosfera. Sob a denominação de índices de estabilidade, eles são amplamente utilizados na previsão meteorológica. Nesta pesquisa, os índices são obtidos de um diagrama Skew (T - Log P). Em Tsonis (2007), encontramos uma abordagem para a obtenção desse diagrama.

Nesta pesquisa, foi utilizado o pacote *RadioSonde* do R, que permite a plotagem de perfis verticais de sondagens de ar superior em um diagrama Skew (T - Log P) (NYCHKA *et al.*, 2016). Algumas equações dos *scripts* contidos neste pacote foram modificadas para possibilitar o cálculo dos índices de estabilidade utilizados neste estudo e, mais especificamente, daqueles que dependem da temperatura da parcela.

3.2.3.2 Índices de estabilidade

De acordo com Fragoso (1998), a maioria dos índices de estabilidade é obtida a partir de sondagens de ar superior, tendo seus limites críticos sido, em geral, determinados para uma região específica. Consequentemente, para utilizá-los em outra região, pode ser necessário adotar novos limiares. A estabilidade da atmosfera pode ser inferida por meio de índices que levam em consideração: a diferença das condições de temperatura e/ou umidade entre dois ou mais níveis isobáricos; índices que associam essas diferenças de temperatura e/ou umidade com as características do vento na troposfera; bem como índices que se baseiam na determinação da quantidade de energia necessária ao desenvolvimento de convecção.

O conjunto de índices de estabilidade utilizados nesta pesquisa é composto de índices já conhecidos na literatura — IK, SWEAT, DTeI, PII, KO — juntamente com outros índices que estão sendo introduzidos pela primeira vez neste tipo de análise e que foram obtidos através da observação e análise prévia de vários perfis verticais para Petrolina — DMAP, TNCL, DPM, TMED e IC.

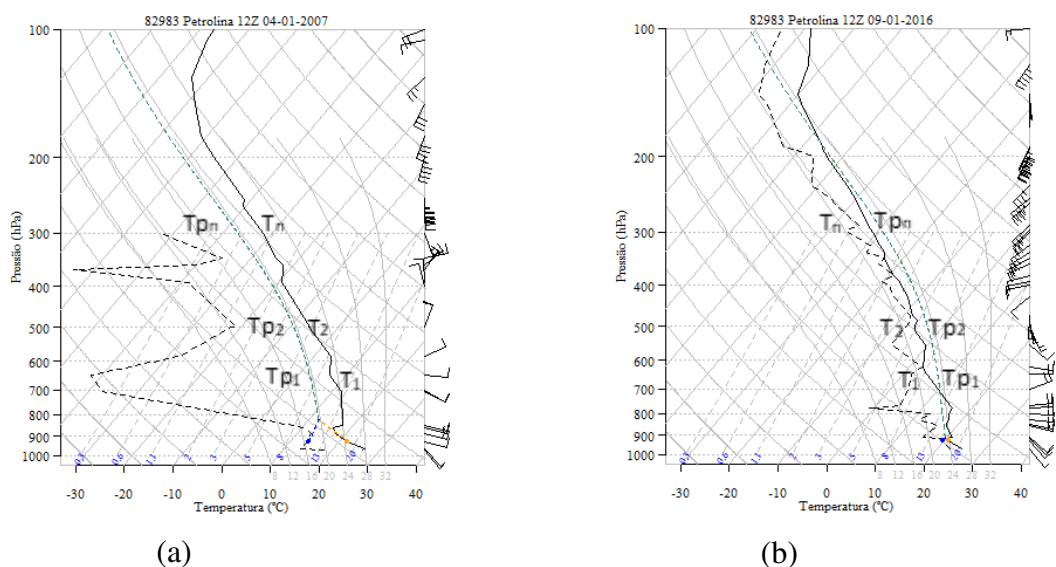
3.2.3.2.1 Diferença média entre o ambiente e a parcela (DMAP).

Esse índice está sendo introduzido pela primeira vez. Seu cálculo consiste em obter a diferença média entre a temperatura do ar e a temperatura da parcela ($T - TP$) em cada nível da sondagem contido na camada 700-300 hPa. Uma vez encontrada esta diferença para cada nível, aplica-se a média conforme a equação abaixo:

$$DMAP = \frac{1}{n} \sum_{i=1}^n (T - TP)_i \quad (1)$$

em que n é o número de níveis contidos na camada 700-300 hPa, T é a temperatura do ar e TP , a temperatura da parcela. O uso deste índice se justifica porque, nos dias associados à estabilidade, a temperatura de uma parcela em ascensão é menor do que a temperatura do ambiente, enquanto que nos dias com condição de instabilidade, a temperatura da parcela costuma ser maior do que a do ambiente. Assim, os valores mais baixos desse índice estão associados com estabilidade e os valores mais altos desse índice associados com instabilidade (Figura 1).

Figura 1 – Perfis verticais para um dia com atmosfera estável (a) e para um dia com condição de instabilidade (b).



3.2.3.2.2 Índice combinado (IC)

Também introduzido pela primeira vez, este índice corresponde a uma combinação entre o TT e o Índice de Showalter (SI). O TT é dado pela seguinte expressão:

$$TT = T_{850} + T_{d\ 850} - 2T_{500} \quad (2)$$

em que T_{850} e $T_{d\ 850}$ são as temperaturas do ar e do ponto de orvalho no nível de 850 hPa, respectivamente, e T_{500} a temperatura do ar no nível de 500 hPa. Já o SI é dado pela seguinte equação:

$$IS = T_{500} + (T_{p500})_{850} \quad (3)$$

em que T_{500} e $(T_{p500})_{850}$ são, respectivamente, a temperatura do ar e a temperatura de uma parcela que ascende adiabaticamente úmida a partir do nível de 850 hPa.

Por fim, o IC é a diferença entre o SI e o TT.

$$IC = TT - IS \quad (4)$$

O uso do IC se justifica porque os índices IS e TT são bastante influenciados pelo teor de umidade nos baixos níveis. Assim, é esperado que o comportamento de ambos seja semelhante. Ademais, a combinação dos índices TT e IS como uma única variável pode ser justificada como uma forma de reduzir o problema da multicolinearidade no conjunto de dados em que a análise multivariada será aplicada. Além disso, em uma análise prévia do conjunto de dados, foi observado que os índices TT e IS estão altamente correlacionados entre si na região de Petrolina, sendo, portanto, viável escrevê-los como uma única variável.

3.2.3.2.3 Índice K (IK)

O IK consiste em uma combinação da diferença de temperatura entre os níveis de 850 e 500 hPa, a temperatura do ponto de orvalho em 850 hPa e a depressão do ponto de orvalho em 700 hPa, que é uma medida do teor de umidade (KUNZ *et al.*, 2009). O principal objetivo

da sua formulação é prever chuvas convectivas e tempestades elétricas. Algebricamente, o IK pode ser determinado através de:

$$K = [T_{(850 \text{ hPa})} - T_{(500 \text{ hPa})}] + Td_{(850 \text{ hPa})} - [T_{(700 \text{ hPa})} - Td_{(700 \text{ hPa})}] \quad (5)$$

Os símbolos seguem os critérios usados na equação anterior:

$T_{850 \text{ hPa}}$ – Temperatura no nível de 850 hpa;

$T_{500 \text{ hPa}}$ – Temperatura no nível de 500 hpa;

$Td_{850 \text{ hPa}}$ – Temperatura no ponto de orvalho no nível de 850 hpa;

$T_{700 \text{ hPa}} - Td_{700 \text{ hPa}}$ é a depressão psicrométrica no nível de 700 hpa.

O Quadro 1 indica os valores que o IK pode assumir e a condição de tempestade a ele associada.

Quadro 1. Valores do IK e a respectiva condição de tempestade associada (Fonte: AWS, 1979).

K (°C)	Condição de Tempestade
< 20	Sem tempestades
20 a 25	Tempestades isoladas
26 a 30	Tempestades muito dispersas
31 a 35	Tempestades dispersas
> 35	Tempestades numerosas

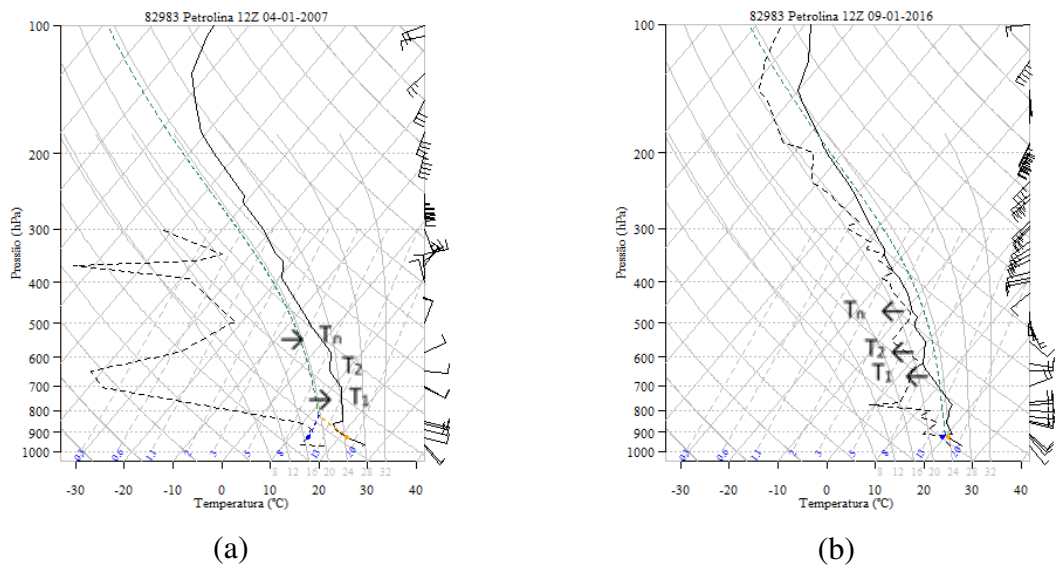
3.2.3.2.4 Temperatura média da camada 700-500 (TMED)

Também introduzido pela primeira vez, este índice é obtido pela média da temperatura do ar dos níveis da sondagem compreendidos entre a camada 700-500hPa. Com base na observação de vários perfis verticais para a área de Petrolina, foi constatado que, nos dias com condições de instabilidade atmosférica, houve uma redução da temperatura do ar nos níveis dessa camada, quando comparados com perfis verticais associados à condição de estabilidade. O estudo dessa camada é importante devido à atuação da inversão térmica de subsidência, que é mais intensa no período seco e menos evidente no período chuvoso. O TMED pode ser obtido pela seguinte expressão:

$$TMED = \frac{1}{n} \sum_{i=1}^n (T)_i \quad (6)$$

em que n é o número de níveis da sondagem contidos na camada 700-500 hPa e T é a temperatura do ambiente em cada nível compreendido nesta camada, conforme evidenciado na Figura 2.

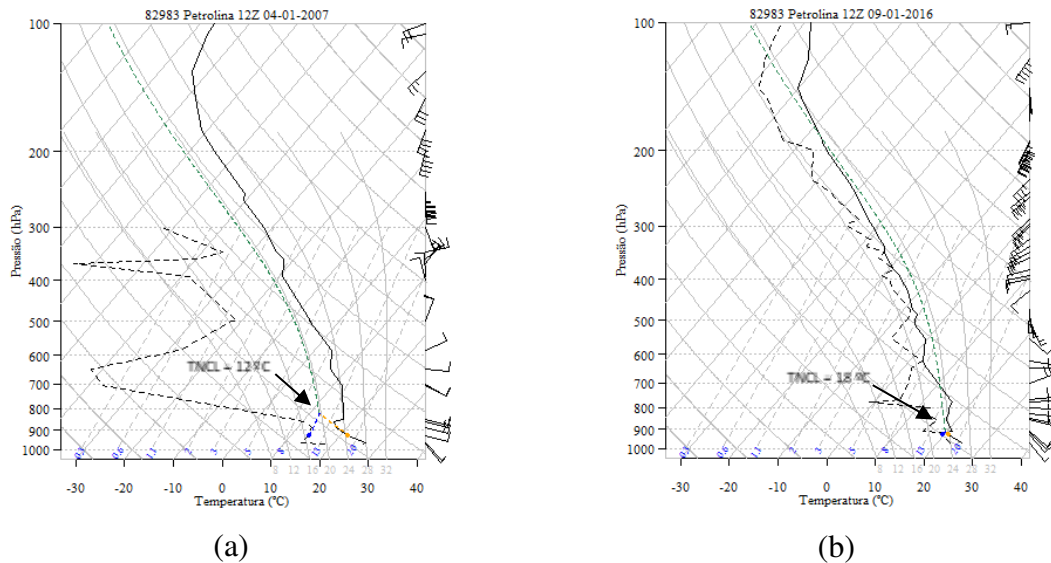
Figura 2 – Perfis verticais com indicação do TMED para um dia com atmosfera estável (a) e para um dia com condição de instabilidade (b).



3.2.3.2.5 Temperatura do nível de condensação por levantamento a partir de 925 hPa (TNCL)

Também introduzido pela primeira vez como um índice preditor de chuva, o índice *TNCL* é simplesmente a temperatura no nível de condensação por levantamento de uma parcela que ascende a partir do nível de 925 hPa. Seu uso se justifica devido ao fato de que o comportamento do *TNCL* ajuda a distinguir dias com condição de estabilidade atmosférica de dias com condição de instabilidade atmosférica, em função do teor de umidade nos baixos níveis. Assim, quanto mais instável o ambiente, maior será o valor de *TNCL*; e quanto mais estável, menor o valor de *TNCL*. Abaixo, um exemplo para um dia com atmosfera estável e com atmosfera instável (Figura 3).

Figura 3 – Perfis verticais com indicação do TNCL para um dia com atmosfera estável (a) e para um dia com atmosfera instável (b).



3.2.3.2.6 Índice de instabilidade potencial (PII)

Este índice é calculado pela diferença entre a temperatura potencial equivalente do nível de 925 hPa e a do nível de 500 hPa, dividido pela diferença entre as alturas dos níveis de 500 e 925 hPa (VAN DELDEN, 2001). O PII é obtido por:

$$PII = \frac{(\theta e_{925} - \theta e_{500})}{(Z_{500} - Z_{925})} \quad (7)$$

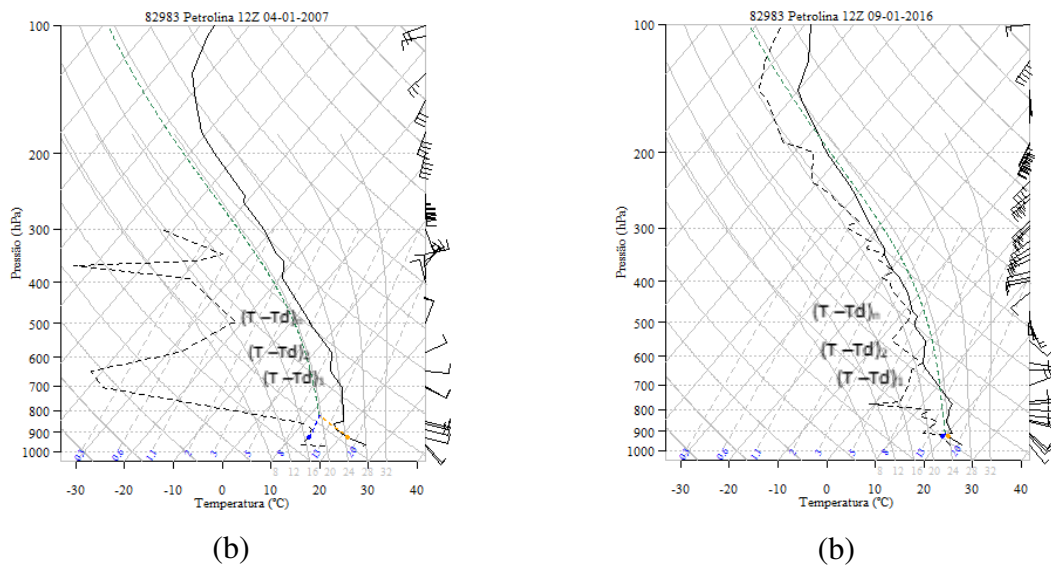
3.2.3.2.7 Depressão Psicrométrica Média da camada 700-500 (DPM)

Este índice é obtido pela média da DPM. Também introduzido pela primeira vez, seu uso se justifica a partir do entendimento de que dias com condição de instabilidade atmosférica costumam apresentar um teor de umidade alto da baixa até a média troposfera. Dias com condição de estabilidade atmosférica apresentam um teor de umidade baixo nessa camada. Seu cálculo é realizado pela seguinte expressão:

$$DPM = \frac{1}{n} \sum_{i=1}^n (T - Td)_i \quad (8)$$

em que n é o número de níveis contidos na camada 700-500hPa, T é a temperatura do ar e Td , a temperatura do ponto de orvalho. Na Figura 4, podemos identificar a principal diferença entre um dia com atmosfera estável e um dia com atmosfera instável, usando este índice como referência.

Figura 4 – Perfis verticais com indicação do DPM para um dia com atmosfera estável (a) e para um dia com atmosfera instável (b).



3.2.3.2.8 Índice KO

O KO pode ser obtido conforme a equação abaixo. Para tanto, modificou-se o modelo de DWD (2013) com a substituição de θe_{1000} por θe_{970} , que constitui o nível mais próximo da superfície em Petrolina:

$$KO = 0,5(\theta e_{700} + \theta e_{500} - \theta e_{970} - \theta e_{850}) \quad (9)$$

na qual θe_{500} , θe_{700} , θe_{850} , θe_{970} são a temperatura potencial equivalente nos níveis de 500, 700, 850 e 970 hPa, respectivamente. As condições de tempestades atribuídas ao KO são apresentadas no Quadro 4.7 a seguir.

Quadro 4.1. Valores do KO e a respectiva condição de tempestade associada.

KO (K)	Condição de tempestade
$KO > 6$	Sem tempestades
$6 \geq KO > 2$	Tempestades elétricas isoladas
$KO \leq 2$	Tempestades numerosas

Fonte: DWD, 2013

3.2.3.2.9 *Severe Weather Threat* (SWEAT)

O SWEAT combina parâmetros importantes como indicadores do potencial de tempestades severas. Os cinco parâmetros descrevem a estabilidade da coluna atmosférica, o jato de baixos níveis, o jato de níveis médios e o cisalhamento direcional entre o vento dos níveis baixos e médios. Entretanto, tal formulação é válida apenas para o hemisfério norte, devido aos termos que envolvem o vento. Nascimento (2005) apresenta a seguinte equação, considerando as definições e adaptações desse índice para o hemisfério sul:

$$SWEAT = 20 (TT - 49) + 12 Td_{(850 \text{ hPa})} + 2 V_{(850 \text{ hPa})} + V_{(500 \text{ hPa})} + 125[sen||diffdir|| + 0,2] \quad (10)$$

na qual TT é o TT, $Td_{(850 \text{ hPa})}$ é a temperatura do ponto de orvalho em 850 hPa, $V_{(850 \text{ hPa})}$ e $V_{(500 \text{ hPa})}$ são a velocidade do vento em nós para os níveis isobáricos de 850 e 500 hPa, respectivamente, e $||diffdir||$ é o valor absoluto da diferença entre as direções do vento (em graus) entre 500 hPa e 850 hPa ($||diffdir|| = ||[direção(500 \text{ hPa}) - direção(850 \text{ hPa})]||$). Se o TT for menor do que 49, este termo deve ser igualado a zero. Se $Td_{(850 \text{ hPa})}$ for menor do que zero, este termo também deve ser igualado a zero. O último termo da equação deve ser desprezado caso pelo menos uma das condições a seguir não seja satisfeita: (i) a direção do vento em 850 hPa deve estar entre 290° e 360° ou entre 0° e 50° ; (ii) a direção do vento em 500 hPa deve estar entre 230° e 330° ; (iii) $diffdir$ deve ser negativo, com a consideração importante de que, caso a direção do vento em 850 hPa esteja entre 0° e 50° , será necessário somar 360 ao valor da direção em 850 hPa antes de calcular $diffdir$; (iv) tanto $V_{(850 \text{ hPa})}$ quanto $V_{(500 \text{ hPa})}$ devem ser pelo menos 15 nós. Destas quatro condições, apenas a (iv) é mantida igual à do hemisfério norte.

3.2.3.2.10 Índice Delta Teta-E (DTeI)

Outro índice de estabilidade utilizado é o Delta Teta-E (DTeI) (ATKINS e WAKIMOTO, 1991). Esse índice é baseado no gradiente da temperatura potencial equivalente entre a superfície e a média troposfera. Ele pode ser obtido através de:

$$DTeI = \theta e_{MÁX} - \theta e_{MIN} \quad (11)$$

em que $\theta e_{MÁX}$ é a temperatura potencial equivalente máxima, obtida próximo à superfície, e θe_{MIN} é a temperatura potencial equivalente mínima, geralmente encontrada entre os níveis de 650 e 500 hPa, ou seja, nos níveis médios.

3.2.4 Transformação de Yeo-Johnson

Muitas vezes, o uso de estatística multivariada em estudos meteorológicos pode ser prejudicado por possível falta de normalidade dos dados. Mas a normalidade multivariada é uma das principais exigências para uma boa utilização de técnicas estatísticas multivariadas, ou seja, a forma da distribuição de uma variável métrica individual deve corresponder à sua distribuição normal. Se a discrepância em relação à distribuição normal é suficientemente grande, não é possível comprovar a significância estatística dos resultados. Caso os dados não sigam uma distribuição normal, ações corretivas podem ser adotadas para contornar o desvio de normalidade. Dessa forma, os índices passarão por transformações matemáticas que ajudam a corrigir possíveis desvios da normalidade.

Yeo e Johnson (2000) introduziram uma nova família de transformações que está definida para todos os números reais, revelando-se adequada para reduzir a assimetria e aproximar a distribuição dos dados originais de uma distribuição normal. Essa transformação é uma extensão da transformação de Box-Cox, que não está definida para dados negativos.

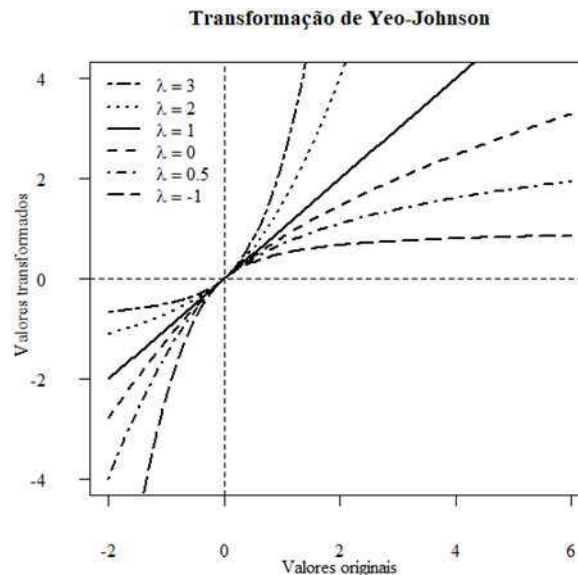
A família de transformações propostas por Yeo e Johnson é apresentada a seguir:

$$\psi(y, \lambda) \begin{cases} \frac{[y + 1]^\lambda - 1}{\lambda}, & \text{para } \lambda \neq 0 \text{ e } y \geq 0 \\ \log[y + 1], & \text{para } \lambda = 0 \text{ e } y \geq 0 \\ -\frac{[1 - \lambda]^{2-\lambda} - 1}{\lambda}, & \text{para } \lambda \neq 2 \text{ e } y < 0 \\ -\log[-y + 1], & \text{para } \lambda = 0 \text{ e } y < 0 \end{cases} \quad (12)$$

em que y é a variável a ser transformada, $\psi(y, \lambda)$ é a variável transformada e λ , um parâmetro a ser estimado. A estimativa do parâmetro dessa transformação consiste em encontrar o λ que maximiza o perfil de verossimilhança entre a variável transformada e a distribuição normal. Uma demonstração matemática deste processo é apresentada por Li (2005) em uma abordagem sobre as transformações de Yeo-Johnson e Box-Cox. Nesta pesquisa, o parâmetro λ da transformação de Yeo-Johnson é calculado com o auxílio do pacote de rotinas “*best normalize*” do Software R, que estima o parâmetro por meio da metodologia demonstrada por Yeo e Johnson (2000) e apresentada por Li (2005).

A Figura 5 demonstra o efeito dessa transformação sobre os dados. Se $\lambda > 1$, os maiores valores da variável a ser transformada são ampliados com mais potência, enquanto os valores menores sofrem uma diminuição. Quando $\lambda = 1$, a transformação não altera o valor da variável original. E quando $\lambda < 1$, os valores menores da variável original são ampliados com mais potência, enquanto os valores maiores são reduzidos.

Figura 5 – Relação entre a variável original e a variável obtida pela transformação de Yeo-Johnson.



3.2.5 Análise multivariada

Técnicas de estatística multivariada são amplamente utilizadas em Meteorologia. Neste estudo, aplica-se técnicas de estatística multivariada a um conjunto de dez índices de estabilidade transformados. Espera-se, com a aplicação dessas técnicas, compreender a relação entre os índices, identificar grupos de dias de acordo com sua

similaridade, bem como avaliar seu comportamento nas vésperas de dias com ocorrência de chuva e nos dias com chuva.

3.2.5.1 *Análise Fatorial de Componentes Principais*

A Análise Fatorial (AF) de ACP é uma técnica de análise multivariada, que tem como objetivo principal resumir as informações contidas em um grande número de variáveis em um conjunto mais restrito de variáveis denominado de fatores. O princípio matemático consiste em fazer a transformação linear e ortogonal do conjunto das variáveis iniciais sob o critério de maximizar as variâncias. Ou seja, a ACP consiste em projetar uma nuvem de indivíduos situados em um espaço afim de dimensão p , em um subespaço afim de dimensão q , de tal forma que sua projeção sofra a menor deformação possível. Dessa forma, é possível reduzir o número de variáveis sem perder as informações contidas nos dados originais. Cada componente principal (CP) explica uma porção da variância total dos dados das variáveis a ela relacionadas, as quais são ordenadas por ordem decrescente dos autovalores mais significativos. Assim, a primeira componente principal retém mais informação sobre os dados do que a segunda, e assim sucessivamente. Antecedendo a ACP, o teste de Kaiser-Meyer-Olkin (KMO) deve ser aplicado ao conjunto de variáveis para verificar se existe uma boa adequação dos dados para a aplicação da técnica. O KMO deve apresentar valores no intervalo de 0,5 a 1 para que a realização da ACP seja viável (HAIR *et al.*, 2009).

Neste trabalho, a ACP foi aplicada aos dados obtidos das sondagens de ar superior, usando-se uma matriz de dados no modo P (RICHMAN, 1986). Ou seja, haverá uma matriz constituída pelas seguintes variáveis transformadas: DMAP*, IC*, K*, TMED*, TNCL*, PII*, DPm*, KO*, SWEAT* e DTel*, nas quais o símbolo “*” denota que a variável foi transformada pela equação de Yeo-Johnson.

A ACP também envolve a rotação das CP. O objetivo da rotação dos fatores é aumentar o poder explicativo dos mesmos. Os eixos podem ser girados sem alterar a distância entre os pontos. Todavia, as coordenadas dos pontos em relação aos eixos são modificadas, ou seja, as cargas fatoriais (relação entre fator e variável) são alteradas na rotação. No entanto, a rotação não muda o total de variância obtida na etapa anterior; o que ocorre é um rearranjo dos autovetores.

Seja $[X]$ uma matriz de dados de dimensão $(m \times n)$, na qual m representa os dias de observações do conjunto de meses de janeiro e n os índices transformados dos quais se deseja

extrair as CPs. Os elementos da matriz $[\mathbf{X}]$ são normalizados com a média zero e a variância igual a um. A normalização dos desvios garante que cada observação dentro do conjunto estudado tenha a mesma importância na determinação dos padrões temporais.

A matriz de correlação $[\mathbf{R}]$ é obtida a partir da matriz $[\mathbf{X}]$ das variáveis padronizadas e de sua transposta $[\mathbf{X}]^T$, que é dada por:

$$[\mathbf{R}] = \frac{1}{n} [\mathbf{X}]^T [\mathbf{X}] \quad (13)$$

A partir da matriz de correlações $[\mathbf{R}]$, determina-se os autovalores λ e autovetores associados a esta matriz.

Os autovalores de $[\mathbf{R}]$ são as raízes de seu polinômio característico. Eles são obtidos pela seguinte expressão:

$$[\mathbf{R}] - \lambda[\mathbf{I}] = \mathbf{0} \quad (14)$$

$$\det([\mathbf{R}] - \lambda[\mathbf{I}]) = 0 \quad (15)$$

Já os autovetores da matriz $[\mathbf{R}]$ são n escalares de um vetor não nulo $[\mathbf{E}]$ de dimensão $(n \times 1)$, que satisfaz a relação:

$$[\mathbf{R}][\mathbf{E}] = \lambda[\mathbf{E}] \quad (16)$$

$$([\mathbf{R}] - \lambda[\mathbf{I}]) \cdot [\mathbf{E}] = [\mathbf{0}] \quad (17)$$

Como a matriz $[\mathbf{R}]$ é simétrica, os autovalores e autovetores a ela associados são ortogonais, ou seja:

$$[\mathbf{E}]^T [\mathbf{E}] = [\mathbf{I}] \quad (18)$$

A soma dos autovalores define a variância total dos dados. A variância explicada por cada fator é calculada pela razão entre o autovalor e a soma dos autovalores. Pode-se obter uma nova matriz de dados a partir da combinação linear entre a matriz dos dados originais normalizados e a matriz transposta dos autovalores $[\mathbf{E}]^T$. Os elementos dessa nova matriz serão as CPs, a qual é dada por:

$$[C] = [E]^T [X] \quad (19)$$

As CPs são calculadas de modo que a primeira CP (C_1) tenha variância máxima. A segunda componente C_2 também é calculada visando a maximizar sua variância, e assim por diante. Assim, a variância de C_1, C_2, \dots, C_n decresce, sendo grande parte da variância dos dados frequentemente explicada por um número reduzido de componentes. Assim, as CPs que não retêm uma variância significativa podem ser descartadas, sem perda significativa das informações contidas na matriz $[X]$.

3.2.5.2 *Análise de Agrupamentos ou Análise de Cluster*

Os fatores obtidos através da ACP formarão a base de dados para a aplicação de outra técnica de análise multivariada: a Análise de Agrupamentos (AA). O objetivo da AA é agrupar os indivíduos — observações — em um número restrito de classes, de modo que os indivíduos pertencentes a cada classe sejam semelhantes em algum aspecto que o distingue das demais classes. Essas classes podem ser obtidas por meio de algoritmos formalizados e de métodos subjetivos ou visuais. Em síntese, esta técnica evidencia os grupos homogêneos e heterogêneos no meio de uma população (EVERITT, 1993).

Na obtenção do agrupamento, destacam-se dois métodos de classificação: os hierárquicos e os não hierárquicos. Nos métodos não hierárquicos, o número de classes K é determinado *a priori*. O principal inconveniente deste método é que a partição final depende do ponto de partida, razão pela qual é necessário efetuar várias classificações, mudando-se o ponto de partida inicial. A partição escolhida neste trabalho constitui os K centros de aglomerações. O ponto de partida desses algoritmos não hierárquicos pode ser determinado de diferentes maneiras: triagem aleatória, cálculo dos indivíduos mais próximos, cálculo dos indivíduos mais distantes e análise baseada no conhecimento dos dados.

Nos métodos hierárquicos, o número de classes não é fornecido inicialmente, ou seja, este método gera o número de classes. A aglomeração hierárquica se caracteriza pelo estabelecimento de uma hierarquia ou estrutura em forma de árvore. Os grupos nos métodos hierárquicos são geralmente representados por um gráfico denominado de dendrograma. Por meio dele, do conhecimento prévio a respeito dos dados e da curva de inércia, é escolhida uma distância de corte para definir quais grupos serão formados. Essa decisão é subjetiva e deve ser feita de acordo com o

objetivo da análise e com o número de grupos desejados. Nesta pesquisa, o agrupamento será feito a partir de um método aglomerativo hierárquico denominado Método de Ward.

O método de Ward propõe que, em qualquer fase da análise, a perda de informação que resulta do agrupamento de elementos entre grupos seja medida pela soma dos quadrados dos desvios (SQD) de cada ponto em relação à média do grupo ao qual pertence (EVERITT, 1993). Assim, para n observações ordenadas em um número de classes K , o SQD é dado pela seguinte expressão:

$$SQD = \sum_{i=1}^n \left[x_i^2 - \frac{1}{n} \left(\sum x_i \right)^2 \right] \quad (20)$$

onde, x_i é a observação pertencente a i -ésima classe.

O método da distância euclidiana quadrática é utilizado para medir a similaridade ou a dissimilaridade entre os indivíduos de acordo com a seguinte expressão:

$$d_{ij} = \left[\sum_{q=1}^p (x_{iq} - x_{jq})^2 \right]^{\frac{1}{2}} \quad (21)$$

na qual q são as variáveis ($q = 1, 2, 3, \dots, p$), x_{iq} é a q -ésima característica da i -ésima observação e x_{jq} é a q -ésima característica da j -ésima observação.

3.2.5.3 Análise Discriminante Múltipla

A Análise Discriminante Múltipla (ADM), outra técnica estatística multivariada, será aplicada para avaliar as diferenças dos grupos obtidos com a AA. A AD é utilizada para discriminar e classificar objetos. Segundo Khattree e Naik (2000), esta técnica separa objetos de uma população em duas ou mais classes. A parte exploratória da análise é a primeira das etapas envolvidas nesse processo. A discriminação ou separação consiste em identificar características que possam ser utilizadas para alocar objetos em diferentes grupos definidos previamente. Essa classificação ou alocação pode ser definida através de um conjunto de regras (funções), que serão usadas para alocar novos objetos.

Conforme Hair *et al.* (2009), o uso da ADM é adequado quando a única variável dependente é não-métrica. A ADM pode ser aplicada em situações em que a amostra total está

dividida em grupos, baseada em uma variável dependente que caracteriza diversas classes conhecidas. Conforme os autores, a ADM tem como objetivos primários entender diferenças de grupos e prever a probabilidade com que um indivíduo ou objeto pertencerá a uma classe ou grupo em particular, com base em diversas variáveis independentes métricas.

Conforme Wilks (2006), o problema básico da ADM consiste em alocar K vetores de dados \mathbf{x} , de dimensão k em um vetor de dois ou mais grupos, baseados em $j = \min(G-1, K)$ vetores discriminantes \mathbf{a}_j , com $j=1,2,3,\dots,J$. A projeção dos dados sobre esses vetores permite obter as funções discriminantes. Nesta pesquisa, \mathbf{x} é a matriz formada pelos índices transformados, relacionada com cada um dos fatores obtidos da ACP e G , os grupos obtidos com a AA. Em Wilks (2006), encontramos a seguinte abordagem sobre a ADM:

$$\delta_j = \mathbf{a}_j^T \mathbf{x} \quad (22)$$

em que, $j=1, 2,3,\dots,j$

As funções discriminantes são obtidas a partir de um conjunto de G matrizes de dados $[x_1]$, $[x_2]$, $[x_3]$, ..., $[x_g]$, de dimensão $(n_g \times K)$, respectivamente. A matriz de variância-covariância para cada um dos G grupos, $[S_1]$, $[S_2]$, $[S_3]$, ..., $[S_g]$ é obtida de acordo com:

$$[S] = \frac{1}{n-1} [\mathbf{x}']^T [\mathbf{x}'] \quad (23)$$

em que $[\mathbf{x}'] = [\mathbf{x}] - \frac{1}{n} [\mathbf{1}][\mathbf{x}]$ é a matriz dos desvios, na qual o segundo termo é uma matriz de dimensão $(n \times k)$ que contém a média da amostra, onde $[\mathbf{1}]$ é uma matriz de ordem $(n \times n)$, cujos elementos equivalem a 1. Um dos pré-requisitos da AD é assumir a igualdade das matrizes de variância-covariância para os G grupos, de modo que a matriz de covariância dentro dos grupos possa ser estimada através da seguinte média ponderada (WILKS, 2006):

$$[S_{dg}] = \frac{1}{n-G} \sum_{g=1}^G (n_g - 1) [S_g] \quad (24)$$

em que há n_g observações dentro de cada grupo e o tamanho total é:

$$n = \sum_{g=1}^G (n_g) \quad (25)$$

Ademais, o cálculo da função discriminante também envolve o cálculo da matriz covariância entre os grupos $[S_b]$. Essa matriz descreve a dispersão das médias dos G grupos em relação à média geral de \mathbf{x} (WILKS, 2006).

$$[S_{eg}] = \frac{1}{G-1} \sum_{g=1}^G [\bar{\mathbf{x}}_g - \bar{\mathbf{x}}_{\bullet}] [\bar{\mathbf{x}}_g - \bar{\mathbf{x}}_{\bullet}]^T \quad (26)$$

onde a média para cada grupo individual é dada por:

$$\bar{\mathbf{x}}_g = \frac{1}{n_g} [\mathbf{x}_g]^T \cdot \mathbf{1} = \begin{bmatrix} \frac{1}{n_g} \sum_{i=1}^{n_g} [x_{i,1}] \\ \frac{1}{n_g} \sum_{i=1}^{n_g} [x_{i,2}] \\ \vdots \\ \frac{1}{n_g} \sum_{i=1}^{n_g} [x_{i,G}] \end{bmatrix}, g = 1, 2, \dots, G \quad (27)$$

e $\bar{\mathbf{x}}_{\bullet}$ é a média geral das n observações de uma variável \mathbf{x} :

$$\bar{\mathbf{x}}_{\bullet} = \frac{1}{n} \sum_{g=1}^G n_g \cdot \mathbf{x}_g \quad (28)$$

Então, as j funções discriminantes são obtidas para o menor valor entre o número de grupos menos um ($G-1$) e o número de variáveis (K). Elas são oriundas dos primeiros autovalores da matriz abaixo:

$$[S_{dg}]^{-1} [S_{eg}] \quad (29)$$

Essa matriz de ordem $K \times K$ geralmente não é simétrica. Os vetores discriminantes \mathbf{a}_j estão relacionados com os autovalores desta matriz e muitas vezes são padronizados para que tenham variância 1.

$$\mathbf{a}_j^T [\mathbf{S}_{dg}] \mathbf{a}_j = 1, \quad j = 1, 2, 3, \dots, J \quad (30)$$

Mas a condição da equação 39 também pode ser obtida por:

$$\mathbf{a}_j = \frac{\mathbf{e}_j}{(\mathbf{e}_j^T [\mathbf{S}_{dg}] \mathbf{e}_j)^{1/2}} \quad (31)$$

A contribuição dos vetores discriminante de $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_j$ para discriminar os G grupos é decrescente, de modo que o primeiro autovetor \mathbf{a}_1 tem a maior contribuição para discriminar os grupos, o autovetor \mathbf{a}_2 tem uma contribuição menor e assim sucessivamente.

Para classificar os indivíduos de acordo com as regras de classificação, é necessário avaliar as distâncias euclidianas no espaço discriminante entre o vetor candidato \mathbf{x}_0 e a média de cada um dos G grupos, a fim de descobrir qual é o mais próximo. Dessa forma, \mathbf{x}_0 será atribuído a um grupo g se:

$$\sum_{j=1}^J [\mathbf{a}_j (x_0 - \bar{x}_g)]^2 \leq \sum_{j=1}^J [\mathbf{a}_j (x_0 - \bar{x}_h)]^2, \text{ para todo } g \neq h \quad (32)$$

Kilca (2009) empregou a AD com o objetivo de desenvolver um esquema de classificação utilizando características estruturais florestais para diferenciar tipos de fitofisionomias florestais no bioma Pampa, do Rio Grande do Sul. Em sua pesquisa, o autor descreveu o significado das saídas do *Software SPSS* durante a realização da AD. A ADM se divide em três etapas: (a) a validação do modelo, onde são apresentados a análise descritiva dos dados (como as médias e desvios padrões dentro dos grupos para cada variável do estudo), a matriz de correlação para avaliar se existe multicolinearidade nos dados, o teste de igualdade de médias e o teste de igualdade das matrizes de covariância; (b) o cálculo das funções discriminantes, etapa que envolve a escolha do método computacional utilizado para realização da AD, a apresentação das funções discriminantes, a apresentação da importância relativa das funções discriminantes, o teste do poder discriminatório das funções, os coeficientes da função discriminante canônica e a estimativa da função nos centroides dos grupos; (c) por fim, é

realizada a etapa de classificação, na qual são apresentados os coeficientes das funções e a classificação dos casos.

As etapas para a aplicação da AD na presente pesquisa estão descritas a seguir.

3.2.5.3.1 Análise descritiva:

Nesta fase da AD, o propósito é avaliar a viabilidade da aplicação da técnica aos dados.

A) Normalidade multivariada

Foi aplicado o teste de Normalidade Multivariada de Mardia (MARDIA, 1970). Este teste se baseia na análise de extensões multivariadas dos coeficientes de assimetria ($\hat{\gamma}_{1,p}$) e curtose ($\hat{\gamma}_{2,p}$). Em Korkmaz (2014), encontra-se a seguinte abordagem sobre esse teste:

$$\hat{\gamma}_{1,p} = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n m_{i,j}^3 \quad \text{e} \quad \hat{\gamma}_{2,p} = \frac{1}{n^2} \sum_{i=1}^n m_{ii} \quad (33 \text{ e } 34)$$

onde $m_{i,j} = (x_i - \bar{x})^T S^{-1} (x_j - \bar{x})$ é o quadrado da Distância de Mahalanobis e p , o número de variáveis. O teste estatístico de assimetria $(n/6)\hat{\gamma}_{1,p}$ é aproximadamente a distribuição X^2 com $p(p+1)(p+2)/6$ graus de liberdade. Similarmente, o teste estatístico para a curtose é, de maneira aproximada, normalmente distribuída com média $p(p+2)$ e variância $8p(p+2)/n$.

O Pacote ‘MVN’ (KORKMAZ *et. al.*, 2019) do software R implementou o teste de Mardia para normalidade multivariada, onde é possível testar as estatísticas $\hat{\gamma}_{1,p}$ e $\hat{\gamma}_{2,p}$. Conforme o autor, os dados multivariados estarão normalmente distribuídos se o p-valor dessas estatísticas for maior do que 0,05.

Outra forma de avaliar a normalidade multivariada dos dados é por meio do gráfico Q-Q plot. Em Melo (2016), encontra-se as seguintes etapas da construção desse gráfico:

i- Calcular as distâncias de Mahalanobis

$$D_j^2 = (x_i - \bar{x})^T S^{-1} (x_j - \bar{x}) \quad (35)$$

ii- Ordenar as distâncias

$$D_{(1)}^2, D_{(2)}^2, \dots, D_{(j)}^2 \quad (36)$$

iii-para cada J, obter:

$$X_{(j)}^2 = F_{(x_{\frac{v}{2}}^2)}^{-1} \left[\frac{j - 0,5}{n} \right], \text{ com } j = 1, 2, \dots, n \quad (37)$$

onde: $F_{(x_{\frac{v}{2}}^2)}^{-1}$ é a função inversa de uma distribuição Qui-Quadrado com v graus de liberdade.

iv-Plotar as coordenadas $(D_{(j)}^2, X_{(j)}^2)$

No entanto, no seu estudo, a autora propõe e analisa uma alternativa ao uso de métodos fundamentados em distâncias quadráticas, pois os $D_{(j)}^2$ não são independentes e identicamente distribuídos. Eles são dependentes mesmo que as variáveis em estudo sejam independentes.

A biblioteca “heplots” (FRIENDLY *et al.*, 2018) do R possui uma função que gera o gráfico QQ-plot utilizado nesta pesquisa.

B) Homogeneidade das matrizes de variância-covariância dos grupos

Conforme Hair *et al.* (2009), o pressuposto da igualdade das matrizes de covariância — homocedasticidade — afeta tanto a estimação como a classificação das regras de classificação. No caso em que as matrizes de covariâncias dos grupos são desiguais, as observações podem ser super classificadas nos grupos com matrizes de covariância maiores, mas este efeito pode ser minimizado aumentando-se o tamanho da amostra. O teste de homogeneidade das matrizes de variância/covariância M de Box é um dos mais usados para fazer a verificação deste pressuposto. Todavia, ele é bastante sensível a desvios de normalidade e a variações do tamanho da amostra dos grupos. Nesta pesquisa, foi utilizada outra abordagem para verificar questões de homocedasticidade. O pacote ‘Vegan’ do R contém a função “betadisper”, que implementa o método de Marti Anderson para análise de homogeneidade de variância nos grupos multivariados. Para avaliar se as dispersões dentro de um ou mais grupos são homogêneas, são calculadas as distâncias de cada observação aos centroides dos grupos e, em seguida, essas distâncias são submetidas ao teste de significância ANOVA. Quando as

distâncias das observações aos centroides dos grupos são obtidas pela distância euclidiana, este procedimento se torna análogo ao teste de Levene para a homogeneidade de variâncias multivariada (FRIENDLY *et al.*, 2018). Nesta pesquisa, investigou-se a dispersão nos grupos, por meio da distância euclidiana e da distância de Mahalanobis.

C) Multicolinearidade

A verificação deste pressuposto é necessária uma vez que a multicolinearidade pode causar impactos sobre a interpretação das funções discriminantes. Segundo Hair *et al.* (2009), a multicolinearidade denota que duas ou mais variáveis independentes estão altamente correlacionadas, de modo que uma variável pode ser altamente explicada ou prevista pela(s) outra(s), o que acrescenta pouco poder explicativo ao conjunto.

Uma das formas de avaliar a multicolinearidade é através do determinante da matriz de correlação do conjunto de dados. O determinante varia entre 0 e 1, sendo que, quanto mais próximo de 1, mais baixa será a colinearidade entre os dados multivariados. Um nível adequado de colinearidade também pode ser obtido selecionando-se as variáveis cujas correlações são menores que 0,7 na matriz de correlação (DORMAN *et al.*, 2012).

D) Teste de igualdade de médias dos grupos.

Esse teste permite avaliar se as populações advêm de um mesmo vetor médio. Uma abordagem sobre esse teste está descrita em Kilca (2009). A estatística lambda de Wilks é dada pela seguinte equação:

$$\Lambda = \frac{|W|}{|T|} \quad (38)$$

onde $|W|$ é o determinante da matriz da soma dos quadrados de produtos cruzados dentro dos grupos e $|T|$ é o determinante da matriz da soma dos quadrados de produtos totais. Quanto mais próximo de 1 lambda estiver, menor será o poder discriminatório da variável.

As hipóteses a serem testadas são:

$$\begin{cases} H_0: \text{As médias dos grupos são iguais} \\ H_1: \text{A média dos grupos diferem} \end{cases}$$

3.2.5.3.2 Obtenção e análise das regras de classificação

Esta etapa envolve a escolha do método computacional utilizado para calcular as funções discriminantes. Nesta pesquisa, foi utilizada a estimação simultânea, onde o cálculo das funções discriminante ocorre de modo que todas as variáveis são consideradas juntas na análise. Neste caso, não se leva em consideração o poder discriminatório individual das variáveis independentes na estimação das funções (Hair *et al.*, 2009).

A importância relativa das funções discriminantes é verificada através dos autovalores da matriz da equação 39, onde o percentual do poder discriminatório explicado pelas funções pode ser estimado pela equação:

$$\% \text{ var} = \frac{e_j}{\sum e} \quad (39)$$

O teste de significância das funções discriminantes deve ser realizado em seguida, de modo a avaliar o poder de cada função em discriminar os grupos. Foi realizado o teste lambda de Wilks para identificar o poder discriminantes das funções. Um maior detalhamento deste teste é encontrado em Kilca (2009). A função “Wilks.test” do pacote de rotinas “rrcov” do software R realiza esse teste de significância das funções (TODOROV, 2018).

Também são analisados os coeficientes padronizados da função discriminante. Conforme Wilks (2006), para interpretar as funções discriminantes, deve-se examinar o sinal e a magnitude dos coeficientes padronizados das funções. Caso o sinal seja ignorado, cada peso representará uma medida da contribuição relativa das variáveis associadas àquela função. Dessa forma, as variáveis com maior peso exercem maior contribuição para o poder discriminatório da função.

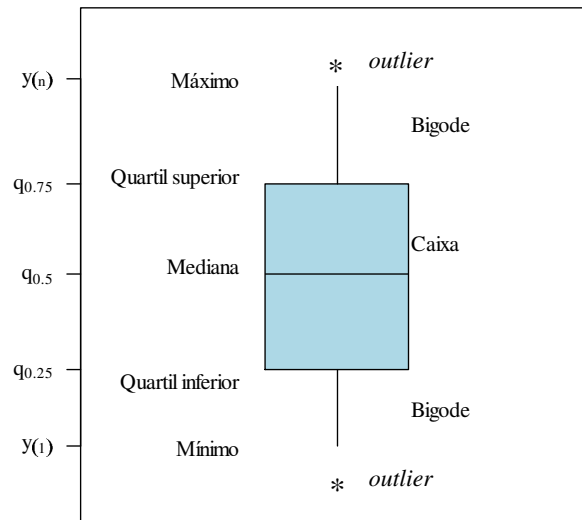
3.2.5.3.3 Análise da classificação

Nesta etapa, foi realizada a análise da classificação obtida pelas funções discriminantes. A análise da tabela de classificação dos coeficientes das funções é útil para avaliar a contribuição das variáveis em cada grupo. Por fim, foi feita a análise da matriz de confronto, que permite identificar a quantidade de observações que foram mal classificadas pela AA. Conseqüentemente, estas foram realocadas pelas funções discriminantes para outro grupo.

3.2.6 Gráficos *boxplot*

O *Boxplot* é um gráfico composto por uma caixa, que concentra a maior parte da distribuição dos dados e pelas extremidades chamadas de bigodes (*whiskers*), que ligam as extremidades da caixa até os valores máximos e mínimos da distribuição. Conforme Wilks (2006), o *boxplot* é composto por cinco quantidades: $y_{(1)}$ valor mínimo, $q_{(0,25)}$ primeiro quartil, $q_{(0,50)}$ mediana, $q_{(0,75)}$ quartil superior e $y_{(n)}$ valor máximo (Figura 6).

Figura 6 – Representação de um *boxplot* esquemático



3.2.7 Método dos quantis

Neste estudo, o conjunto de índices de estabilidade obtido para os meses de janeiro, o conjunto de índices transformados e os escores dos fatores obtidos da ACP foram submetidos ao método dos quantis. Este tem por objetivo identificar faixas de valores que permitem diferenciar/comparar, respectivamente, o indicativo de instabilidade dos índices e a magnitude das cargas fatoriais. O método dos quantis consiste em ordenar os dados em ordem crescente $\{x_1, x_2, x_3, \dots, x_n\}$, em que x_1 representa o menor valor de precipitação; x_n , o maior valor da série; e o índice n representa o tamanho da amostra. Em seguida, são calculados os percentis, que são medidas (quantis) que dividem a amostra em cem partes iguais. O k -ésimo percentil, P_k , é o valor x (x_k) correspondente à frequência cumulativa de $nk/100$ e que ocupa a posição $k = p(n+1)/100$ arredondada para o inteiro mais próximo.

A técnica dos quantis pode ser usada para classificar vários tipos de dados meteorológicos. Santos (2013) utilizou as ordens quantílicas $q_{0,15}$; $q_{0,35}$; $q_{0,65}$ e $q_{0,85}$ com a finalidade de classificar totais mensais de precipitação em faixas – “muito seco”, “seco”,

“normal”, “chuvoso” e “muito chuvoso” – com o objetivo de caracterizar os períodos de anomalia de acordo com a intensidade dos eventos climáticos.

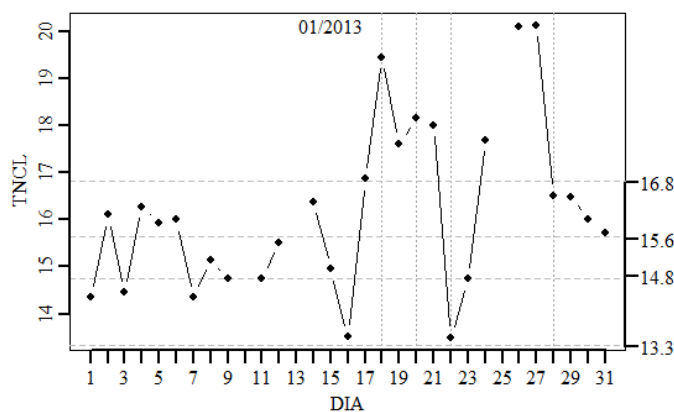
Com o objetivo de construir uma metodologia para monitoramento mensal da precipitação e das temperaturas do ar máxima e mínima da região Sul do Brasil usando a técnica dos quantis, Farias (2012) calculou os valores limites dos quantis de ordem 15, 35, 50, 65 e 85 da precipitação e temperaturas do ar máxima e mínima mensais da região em questão. O autor mostrou que o monitoramento com a técnicas dos quantis indica que as classificações de períodos secos, normais, chuvosos, bem como de períodos quentes, normais e frios, na região Sul do Brasil, são decorrentes de fatores tanto de grande escala como de escala regional, ou, ainda, de uma combinação destas em fenômenos de ambas as escalas.

Nesta pesquisa utiliza-se a técnica dos quantis com as seguintes finalidades:

1- Auxiliar a análise das séries temporais dos índices de estabilidade (Figura 7).

As ordens quantílicas de 20%, 40%, 60% e 80% do conjunto de índices dos meses de janeiro são estimadas e apresentadas no gráfico juntamente com as séries mensais desses índices, como no exemplo abaixo. Isto permite avaliar o comportamento dos índices de estabilidade na presença ou não de dias de chuva e confrontar seus valores em diferentes situações meteorológicas com os quantis desses índices. Na Figura 7, as linhas horizontais representam as ordens quantílicas e as linhas verticais, os dias de chuva na estação de Petrolina. A linha preta com marcadores ilustra o comportamento do índice de estabilidade ao longo do mês.

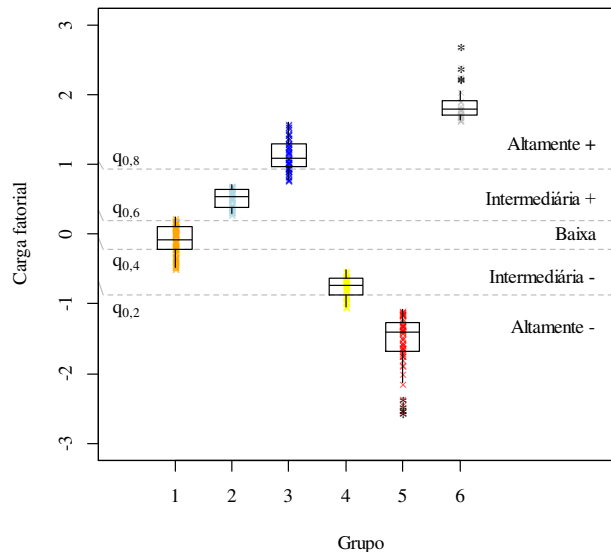
Figura 7 – Série temporal do Índice TNCL para o mês de janeiro de 2013. No eixo das abscissas (lado esquerdo), constam os valores do TNCL; no eixo à direita do gráfico, constam os valores das ordens quantílicas de 20%, 40%, 60%, 80% do TNCL.



2- Classificar as cargas fatoriais dos fatores em faixas de valores — altamente negativas, intermediárias negativas, baixas, intermediárias positivas e altamente positivas — com o objetivo de avaliar o comportamento dos grupos nessas faixas de valores. Também são utilizadas as ordens quantílicas de 20%, 40%, 60% e 80%.

No exemplo da Figura 8, verifica-se que o grupo 5 possui valores inferiores ao quantil 20%; esta faixa de valores indica cargas fatoriais *altamente negativas*. Já os valores do grupo 4 estão situados entre as ordens quantílicas 20% e 40%; esta faixa serve para classificar as cargas fatoriais como *intermediárias negativas*. O grupo 1 está situado entre os quantis 40% e 60%, que está associado com cargas fatoriais *baixas*. O grupo 2 concentra valores entre os quantis 60% e 80%, faixa que classifica os escores como *intermediários positivos*. E o grupo 3 corresponde à maioria dos dias. Por fim, o grupo 6 possui valores maiores que o percentil 80%, sendo atribuído, portanto, a cargas fatoriais *altamente positivas*.

Figura 8 – Distribuição dos grupos resultantes da análise de CP do primeiro fator. As linhas horizontais indicam as ordens quantílicas de 20%, 40%, 60% e 80%.

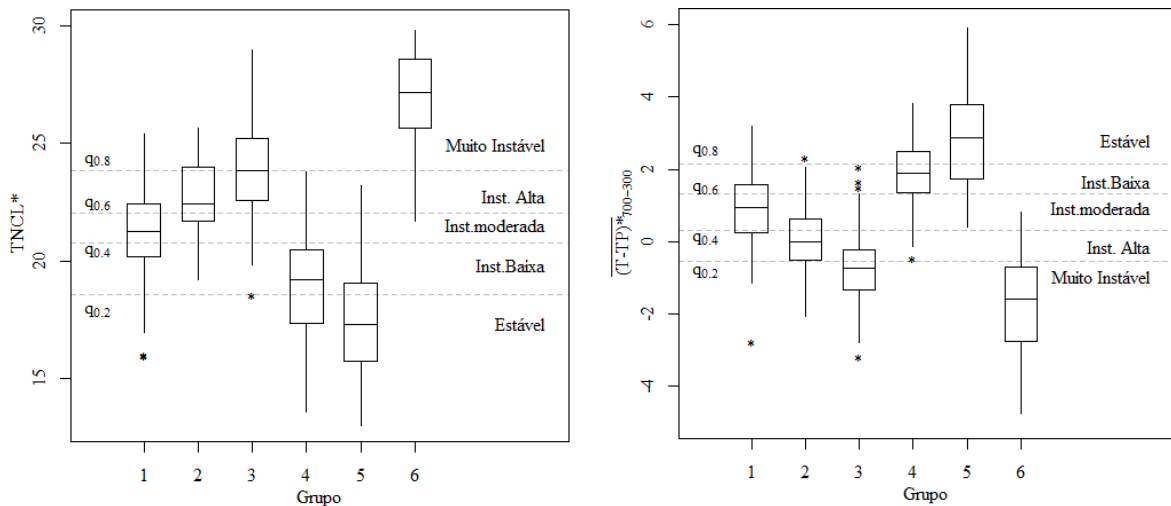


3- Classificar o indicativo de instabilidade dos índices transformados em faixas de valores — estável, instabilidade baixa, instabilidade moderada e instabilidade alta — para avaliar o comportamento dos índices organizados de acordo com o agrupamento dos escores fatoriais. Isto ajudará a identificar relações entre o indicativo de estabilidade e a magnitude dos escores fatoriais. Também são utilizadas as ordens quantílicas de 20%, 40%, 60% e 80% dos índices transformados.

Na Figura 9a, apresenta-se os índices transformados arranjados de acordo com o agrupamento dos escores de uma componente principal. A maior parte dos dias reunidos no

grupo 5 possui valores do TNCL transformado inferiores ao quantil 20%. Esta faixa de valores classifica o indicativo dos índices como *estável*. Já a maior parte dos valores do Grupo 4 está situada entre as ordens quantílicas 20% e 40%, estando o indicativo de instabilidade dessa faixa classificado como *instabilidade baixa*. Se os valores dos índices estiverem entre os quantis 40% e 60%, o indicativo de estabilidade é classificado como *instabilidade moderada*. O grupo 1 encontra-se muito disperso, mas seus valores tendem a se concentrar na faixa de valores situada entre o percentil 40% e 60%, podendo ser atribuído à classe *instabilidade moderada*. O grupo 2 tende a concentrar valores entre os quantis 60% e 80%, faixa que classifica o indicativo de estabilidade dos índices como *instabilidade alta*. O grupo 6 concentra a maioria dos dias com valores superiores ao percentil 80%, sendo classificado, portanto, como *muito instável*.

Figura 9 – Variáveis transformadas agrupadas de acordo com a AA de uma CP.



4 RESULTADOS E DISCUSSÕES

Neste capítulo, realiza-se um estudo sobre o comportamento das chuvas na estação meteorológica de Petrolina. Considera-se na análise a variabilidade temporal do número de dias de chuva, do volume diário e do total mensal de precipitação.

4.1 ESTUDO CLIMATOLÓGICO DA PRECIPITAÇÃO

4.1.1 Variabilidade do número de dias de chuva

Na Figura 10, apresenta-se o *boxplot* do número de dias de chuva na estação de Petrolina. Nos meses de novembro a maio, verifica-se que a ocorrência de dias chuvosos na estação de Petrolina ocorre sempre de maneira muito dispersa, pois esta é uma característica da precipitação no semiárido nordestino. No mês de janeiro, a quantidade de dias chuvosos está mais concentrada na faixa de valores entre 3 e 9 dias com ocorrência de chuva. Os valores acima do terceiro quartil encontram-se mais dispersos, haja vista que a distância entre o valor mínimo e o terceiro quartil é alta. A quantidade mediana de dias com ocorrência de chuva no mês de janeiro é de aproximadamente 6 dias.

Para o mês de fevereiro, os dados também são bastante dispersos, porém, esta dispersão é menor na faixa de valores entre o primeiro e o terceiro quartil. Isso indica que a quantidade de dias chuvosos no mês de fevereiro tende a se concentrar entre 5 e 10 dias, e que a quantidade mediana é igual a 7 dias.

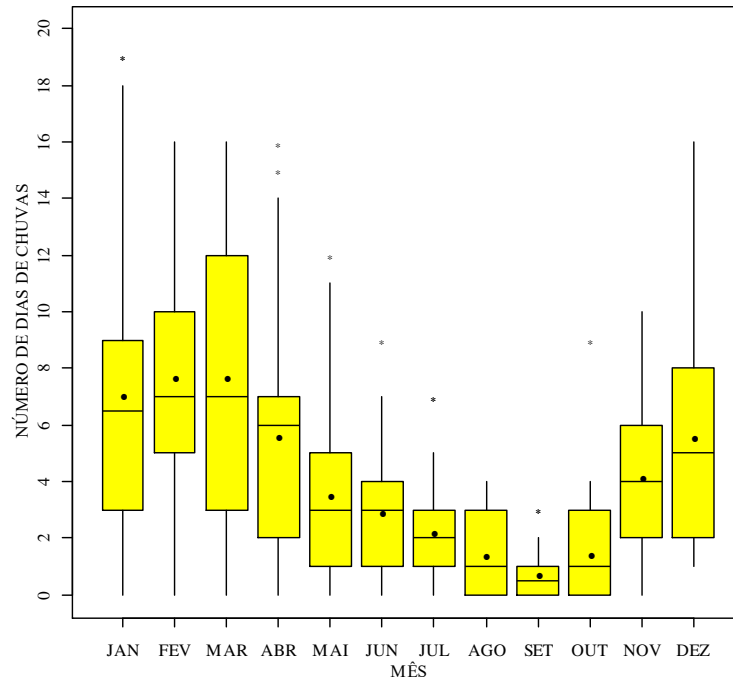
O mês com a maior variabilidade dos dias de chuva é o mês de março. Este mês é o que apresenta a maior distância entre o primeiro e terceiro quartil, indicando uma alta variabilidade da precipitação. No mês de abril, há uma alta variabilidade dos dias com chuva entre o primeiro quartil e a mediana, e uma pequena variabilidade entre a mediana e o terceiro quartil. Valores acima do terceiro quartil também apresentam alta variabilidade. Isso evidencia uma maior probabilidade da quantidade de dias com chuva se concentrarem entre 6 e 7 dias.

O mês de maio já é considerado um mês da estação seca, mas a alta variabilidade dos dias com chuva neste mês ilustra bem a transição da estação chuvosa para a estação seca. 75% dos dias concentram-se de 0 a 5 dias com chuva. Acima do terceiro quartil, verifica-se uma alta variabilidade dos dias com chuva, podendo chegar a 14 dias com precipitação no mês de maio.

Nos meses de junho a outubro, a quantidade de dias com chuva é baixa e a variabilidade também é menor, conforme mostra a Figura 10. Nos meses de novembro e

dezembro, que correspondem ao início da estação chuvosa, a variabilidade do número de dias com precipitação volta a aumentar.

Figura 10 – Distribuição do número de dias de chuva na Estação Meteorológica de Petrolina no período de janeiro de 1963 a dezembro de 2016.



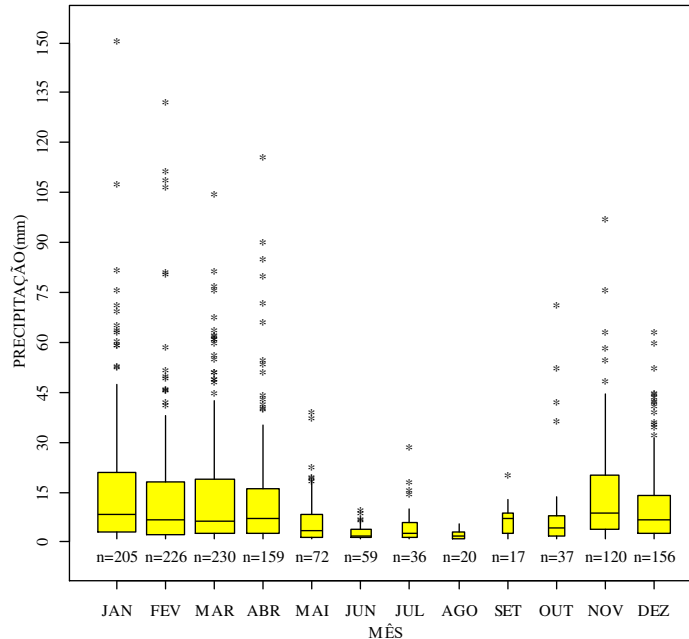
4.1.2 Variabilidade dos volumes diários de precipitação

Na Figura 11, apresenta-se o *boxplot* dos volumes de precipitação diários identificados na série histórica da estação de Petrolina. Foram selecionados os dias com volumes de precipitação maior ou igual a 1 mm. O número de dias chuvosos, contabilizados em cada mês durante o período de janeiro de 1963 a dezembro de 2016, varia de 159 a 230 dias na estação chuvosa (meses de novembro a abril) e entre 17 e 72 na estação de estiagem (maio a outubro). A espessura dos *boxplots* é apresentada proporcionalmente ao tamanho do número de registros de chuva em cada mês. Assim, quanto maior a quantidade de dias com registro de chuva em um dado mês, maior a espessura do *boxplot*.

A quantidade de chuva mediana varia entre 2 mm na estação seca e 8,6 mm na estação chuvosa, ou seja, a distância entre os valores medianos do volume de precipitação diária na estação seca e na estação chuvosa é de apenas 6,6 mm. A assimetria negativa na distribuição dos dados indica que a maior parte dos volumes de precipitação se concentra na faixa de valores inferiores a mediana ~ 8mm. Na estação chuvosa, a precipitação acima da mediana, que

corresponde a 50% do total de casos, apresenta alta variabilidade, oscilando de 8,6 a 47,6 mm. A magnitude dos *outliers* da estação chuvosa mostra que a maioria dos dias com precipitação acima de 50 mm, pode estar associada a eventos extremos.

Figura 11 – Distribuição do volume de precipitação diária na Estação de Petrolina



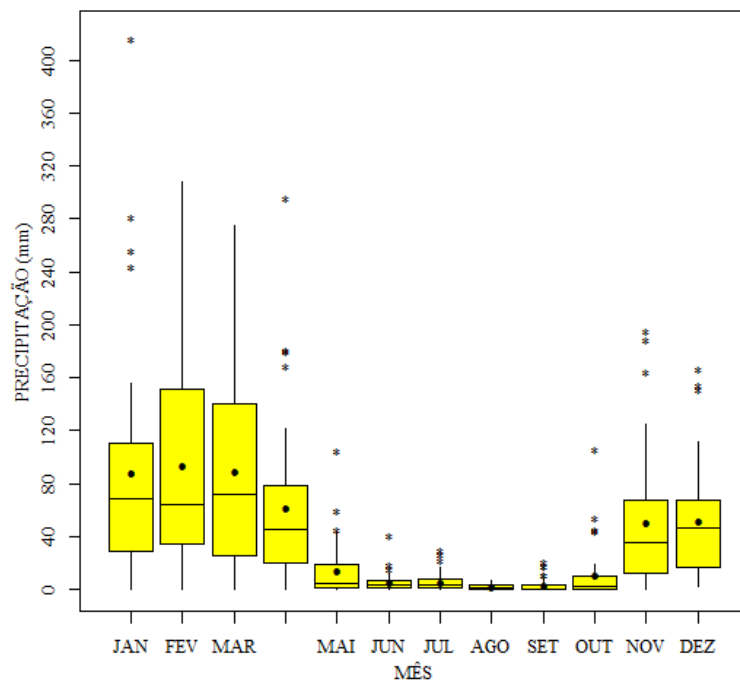
4.1.3 Variabilidade dos totais mensais

Na Figura 12, apresenta-se a distribuição dos totais mensais de precipitação. Neste gráfico, o período chuvoso e o período de estiagem de Petrolina ficam mais evidentes quando comparados ao gráfico do número de dias com chuva e ao do volume de precipitação diário. A variabilidade da precipitação mensal nos meses de novembro, dezembro, janeiro e abril são menores em comparação aos meses de fevereiro e março. A alta variabilidade nesses dois últimos meses ilustra bem a irregularidade das chuvas no semiárido nordestino na estação chuvosa.

A alta variabilidade da precipitação na estação chuvosa de Petrolina, demonstrada nos gráficos *boxplot* abaixo, também é demonstrada em outros estudos conduzidos para a região nordeste do Brasil. Essa alta variabilidade está associada a variações de mecanismos em escala interanual e intrasazonal, bem como interdecenal. Vasconcelos Júnior (2014) fez uma profunda análise da pré-estação chuvosa (PESQ) e da estação chuvosa (EST) do norte da região Nordeste do Brasil, com o objetivo investigar variações interanuais e intrasazonais no período de 1979 a 2010. O autor determinou datas de início e término da PESQ/EST e verificou que os desvios

padrão no início da PESQ/EST são menores em relação aos do término dessas estações. Isto sugere que os sistemas precipitantes no início da PESQ/EST têm uma regularidade maior do que no término. Em relação aos comportamentos anômalos — início antecipado ou término tardio / início tardio e término antecipado — do início/fim da pré-estação chuvosa e da estação chuvosa, tais variações estão associadas à variabilidade de índices de anomalias de temperatura da superfície do mar (TSM) nos oceanos Atlântico e Pacífico.

Figura 12 – Distribuição temporal dos totais mensais de chuva na região de Petrolina. O asterisco “*” representa valores extremos e a bola “•” representa a média da distribuição.



Corrêa (2016) concluiu que, na precipitação no Estado do Ceará, variabilidades interanuais decorrentes de eventos de El-niño Oscilação Sul (ENOS), Dipolo do Atlântico e interdecenais associadas à oscilação decadal do Pacífico (ODP) e à oscilação multidecadal do Atlântico (OMA) são capazes de produzir variabilidades interanuais e interdecenais na classificação do clima daquele estado. A autora aplicou a análise de ondeletas e a análise harmônica a séries do índice efetivo de umidade (Im) e do índice de aridez da UNEP (IaUNEP), tendo identificado ciclos de 3 a 6 anos na variabilidade dos índices em todas as regiões homogêneas de precipitação do Ceará, os quais estão associados com anos de atuação do ENOS. Em algumas regiões homogêneas, a autora identificou ciclos de 10 a 12 anos com um padrão mais forte, estes últimos associados com variações da ODP e da OMA.

4.2 SÉRIES MENSAIS DE ÍNDICES DE ESTABILIDADE

Nesta seção, são analisadas séries mensais de índices de estabilidade para o conjunto de janeiro, onde é possível observar o indicativo de instabilidade dos índices em situações de dias secos, véspera de dias de chuva e dias de chuva. O objetivo aqui é verificar o desempenho desses índices na detecção de condições favoráveis à precipitação na área de Petrolina.

4.2.1 Diferença média da temperatura do ambiente e da parcela (DMAP) ou

$$\overline{(T - TP)}_{700-300}$$

O índice analisado na Figura 13 é o DMAP. As séries mensais do DMAP exibidas na Figura 13 têm o propósito de avaliar o comportamento deste índice nas vésperas e nos dias de chuva.

Em janeiro de 1999, houve registro de chuva nos dias 7, 11, 12 e 15. O índice DMAP apresentou maior indicativo de instabilidade — valores abaixo do percentil 40% — no dia 7, no dia 10, que é véspera de dia de chuva, e no dia 12 [Figura 13 (a)].

No mês de janeiro do ano de 2000, o índice DMAP não indicou instabilidade alta nos dias de chuva 9, 10 e 14, mas indicou com dois dias de antecedência ao dia de chuva 20 e manteve o indicativo de instabilidade alta até o dia 26, que sucede o dia de chuva 25. O índice também apresenta indicativo de instabilidade alta no dia 28, que também antecede um dia de chuva [Figura 13 (b)]. Em janeiro de 2001, o DMAP apresentou sinal de instabilidade no dia 4, que antecede o dia de chuva 5. Mesmo comportamento é observado no dia 11, véspera da chuva do dia 12 na estação de Petrolina [Figura 13 (c)].²

No mês de janeiro de 2003, o DMAP possui valores abaixo do percentil 40% no dia de chuva 2 e na véspera dos dias de chuva 9, 14, 22, 25 e 31, indicando instabilidade alta. Nos demais dias com registro de chuva, o índice não indicou instabilidade [Figura 13 (d)]. No mês de janeiro de 2004, o DMAP não apresentou bom desempenho nos dias antecedentes aos dias de chuva, já que o indicativo de instabilidade variou para valores acima do percentil 60% [Figura 13 (e)]. Já no mês de janeiro de 2005, o DMAP apresentou um bom desempenho para a previsão de chuva, indicando instabilidade alta nas vésperas dos dias 11, 16, 21, 23, 26, e 31 [Figura 13 (f)]. Em janeiro de 2007, só houve registro de chuva no dia 14, ocasião em que o DMAP indicou condição de instabilidade na sua véspera [Figura 13 (g)]. Para janeiro de 2008,

² Dados indisponíveis para os meses de janeiro dos anos de 2002, 2006 e 2010 (ver Cap.3, Seção 3).

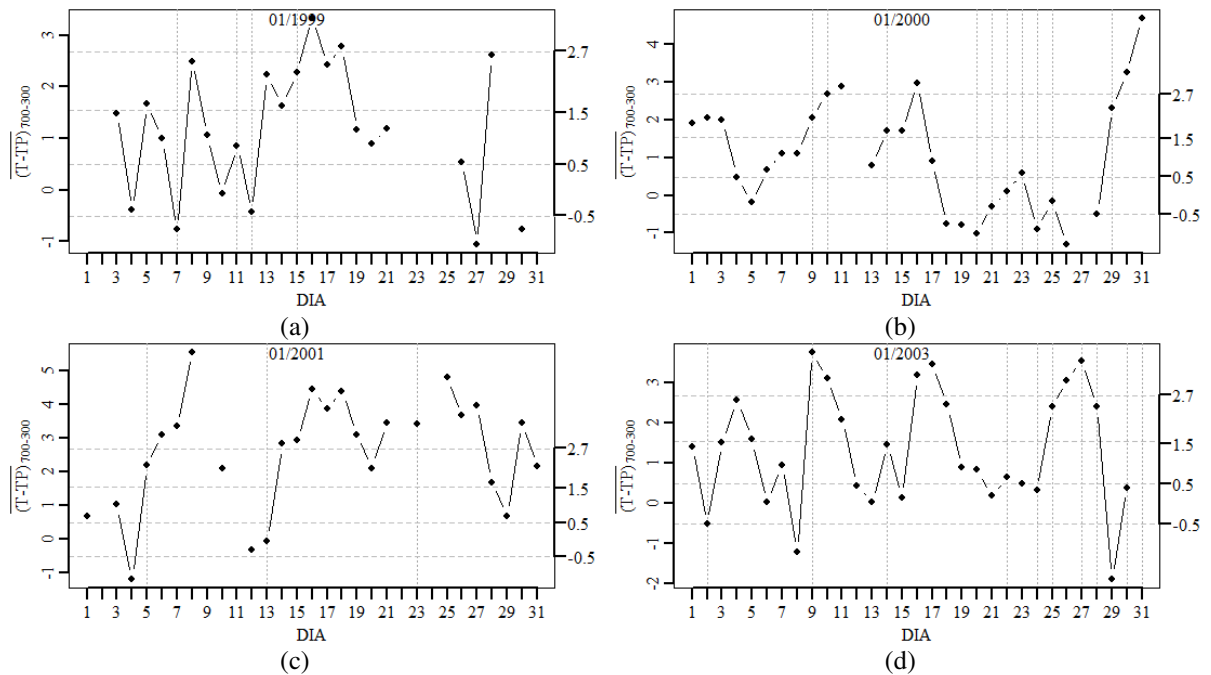
o DMAP indicou instabilidade alta na antevéspera do dia de chuva 22 e no dia 31 [Figura 13 (h)].

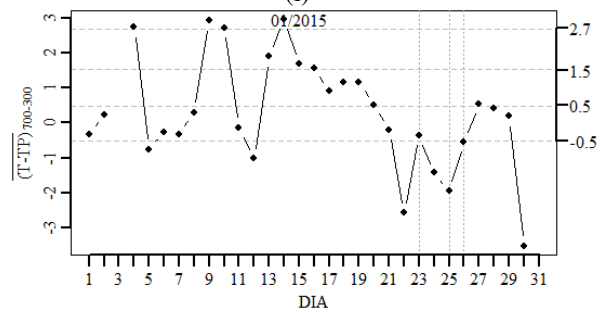
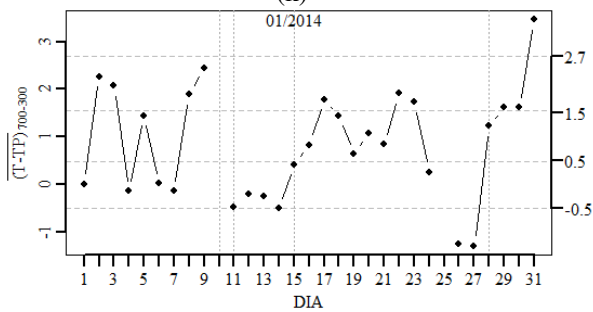
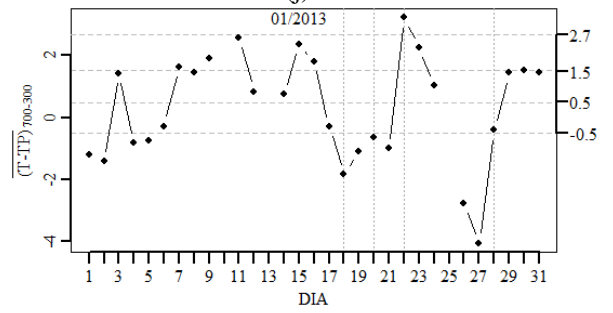
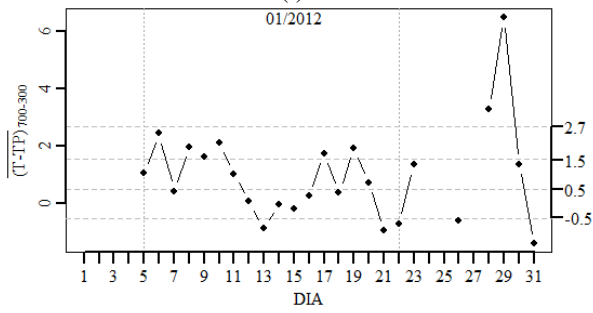
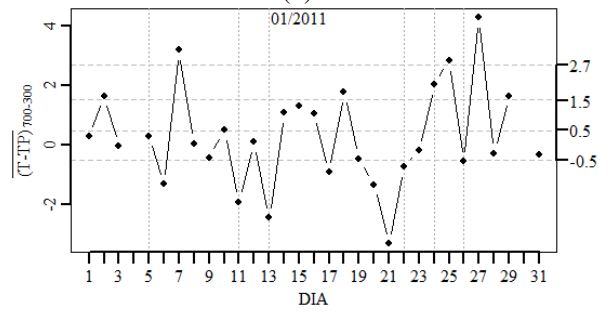
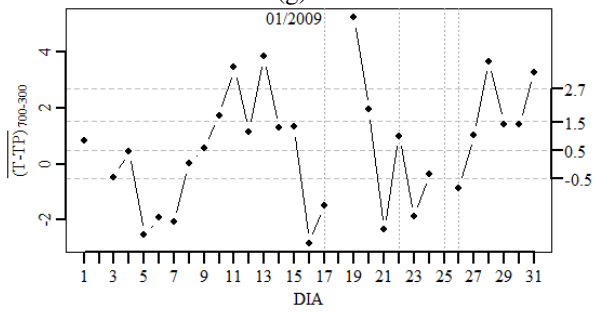
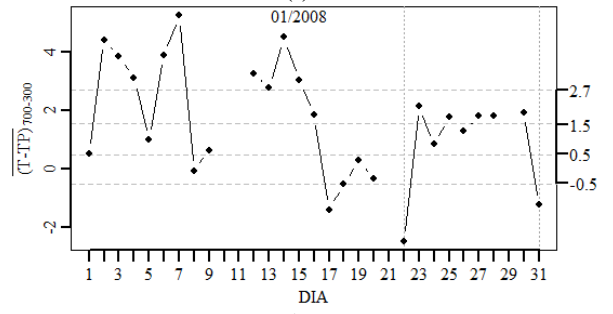
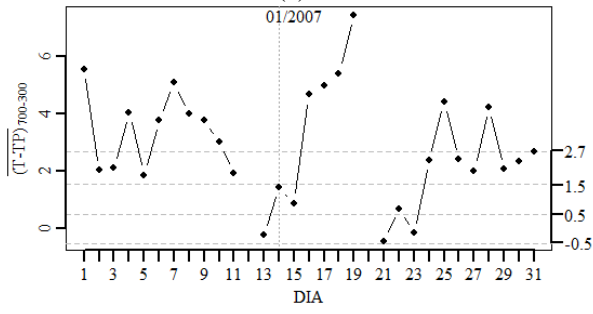
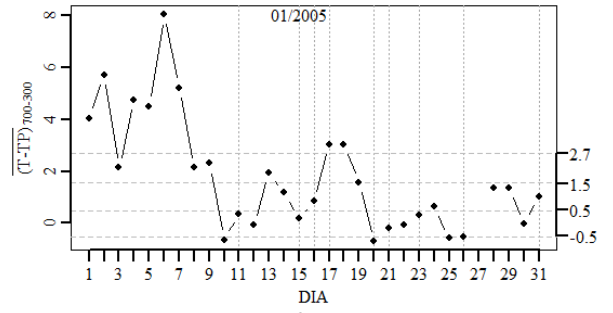
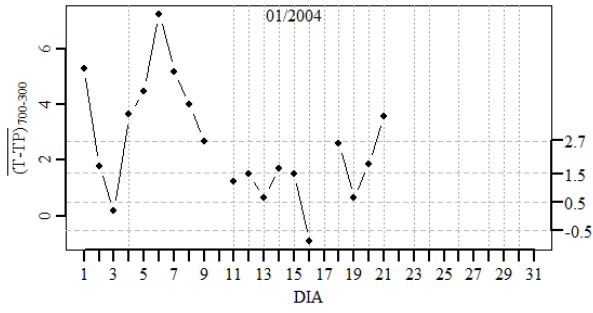
No mês de janeiro de 2009, o DMAP indica condição de instabilidade na véspera dos dias de chuva 16 e 22, bem como na antevéspera do dia 26 [Figura 13 (i)]. Em janeiro de 2011, o $\overline{(T - TP)}_{700-300}$ possui indicativo de instabilidade alto na antevéspera do dia de chuva 5, na antevéspera e na véspera do dia 11 e indica condição de instabilidade para as vésperas dos dias de chuva 13, 22 e 24 [Figura 13 (j)].

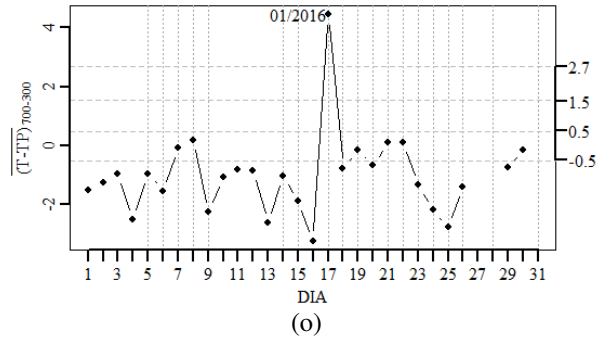
No mês de janeiro de 2012, houve registro de chuva apenas nos dias 5 e 22. O indicativo de instabilidade do DMAP é alto no dia 21 [Figura 13 (k)]. Em janeiro de 2013, o DMAP possui indicativo de instabilidade alto na véspera dos dias de chuva 18, 20, 22 e 28 [Figura 13 (l)]. Em janeiro de 2014, o DMAP indica condição de instabilidade na véspera dos dias de chuva 15 e 27 [Figura 13 (m)]. No mês de janeiro de 2015, o DMAP também indica instabilidade em todos os dias de chuva desse mês (23, 25 e 26) [Figura 13 (n)].

Por fim, no mês de janeiro de 2016, o DMAP indica condição de instabilidade em praticamente todo o mês, que teve ocorrência de chuva em 18 dias. O índice apresentou alto indicativo de instabilidade na véspera ou na antevéspera desses dias de chuva [Figura 13 (o)].

Figura 13 – Séries temporais do DPM. Linhas tracejadas horizontais correspondem aos percentis 20%, 40%, 60% e 80 %. Dias com ocorrência de precipitação são sinalizados com linhas verticais pontilhadas.



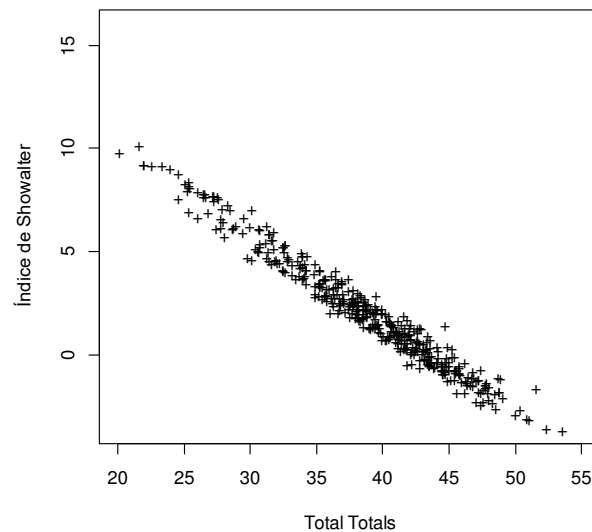




4.2.2 Índice Combinado (IC)

O índice analisado na Figura 15 é uma combinação do SI com o TT. O primeiro considera o teor de umidade em 850hPa e o *lapse rate* entre os níveis de 850 e 500 hPa, sendo o SI uma diferença algébrica da temperatura da parcela de ar e do ambiente em 500 hPa. Esses índices estão altamente correlacionados entre si, mostrando que a temperatura e o teor de umidade nos baixos níveis têm relação com o levantamento da parcela (Figura 14). Devido à alta correlação entre esses índices, é conveniente escrevê-los como uma combinação, de modo a serem utilizados posteriormente pelas técnicas estatísticas multivariadas, evitando, assim, o aumento da colinearidade dos dados. Chamaremos, portanto, o índice resultante da diferença entre o TT e IS de IC.

Figura 14 – Relação entre o Total Totals e o Índice de Showalter



Em janeiro de 1999, o IC indica instabilidade alta (com valores acima do percentil 60%) nas vésperas dos dias 7, 11 e 12 [Figura 15 (a)]. No mês de janeiro de 2000, o IC não

indicou instabilidade alta nos eventos chuvosos dos dias 9 e 10, mas indicou para outros dias de chuva do mês (14, 20, 21, 22, 23, 24 e 25) [Figura 15 (a)].

Em janeiro de 2001, o IC indicou condição de instabilidade alta nas vésperas dos dias de chuva 4 e 13, com valor acima do percentil 60% [Figura 15 (c)].³

Em janeiro de 2003, o IC indica condição de instabilidade nas vésperas dos dias de chuva 8, 22 e 30, sendo que, nos dias de chuva 2 e 28, o IC também apresenta indicativo de instabilidade alto. Já para os dias 14, 24, 25 e 26, o IC não indicou instabilidade alta [Figura 15 (d)]. No mês de janeiro de 2004, o IC também não apresentou bom desempenho na previsão de dias de chuva, indicando condição de instabilidade alta apenas para os dias 2 e 21. Ressalta-se que o mês de janeiro de 2004 apresentou um número elevado de dias de chuva, e que os últimos nove dias do mês não tiveram disponibilidade de dados de altitude [Figura 15 (e)]. Já no mês de janeiro de 2005, o IC indicou condição de instabilidade alta nas vésperas dos dias 11, 16, 17, 20, 21, 23, 26 e 31 [Figura 15 (f)]. Em janeiro de 2007, o IC indicou instabilidade alta na véspera do dia de chuva 13 [Figura 15 (g)]. Em janeiro de 2008, o IC indicou condição de instabilidade alta na antevéspera do dia 22, o qual teve registro de chuva, mas não acertou a previsão para o dia 31 [Figura 15 (h)].

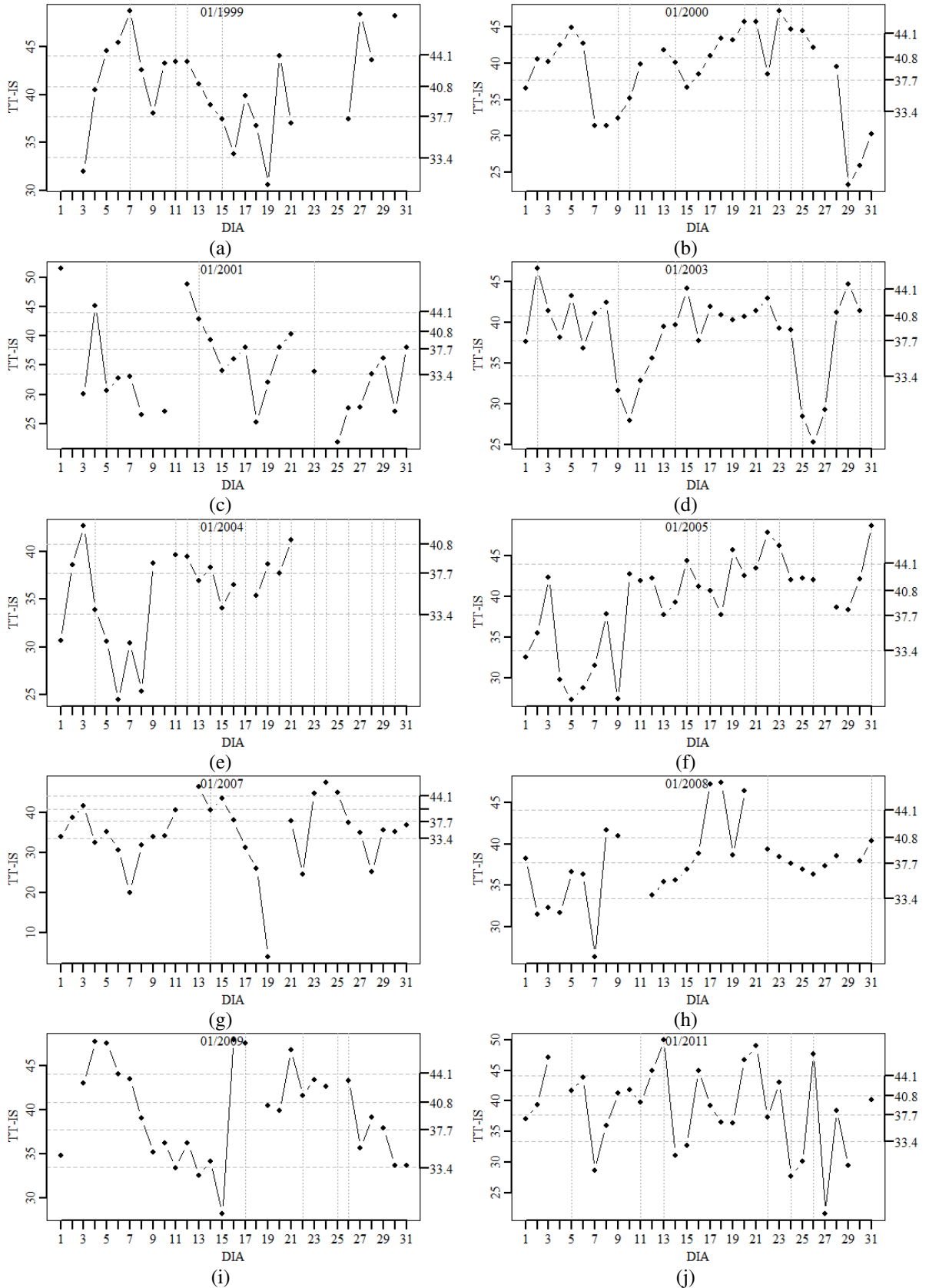
No mês de janeiro de 2009, o IC apresentou bom desempenho na previsão dos dias de chuva, indicando instabilidade alta nos dias antecedentes aos dias de chuva 17, 22 e 26 [Figura 15 (i)]. Em janeiro de 2011 o IC também apresentou um bom desempenho na previsão de dias de chuva, apresentando indicativo de instabilidade superior ao percentil 60% nas antevésperas e vésperas dos eventos chuvosos (dias 5, 11 e 13) [Figura 15 (k)].

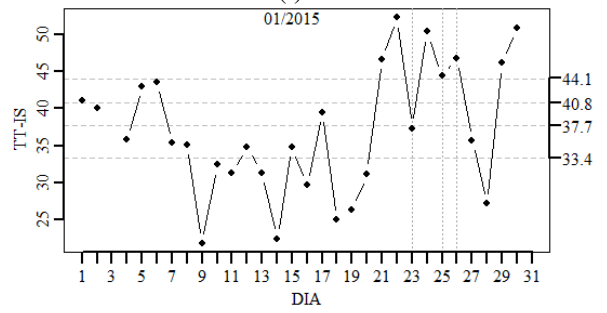
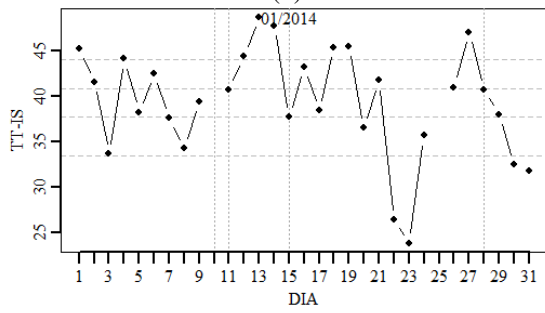
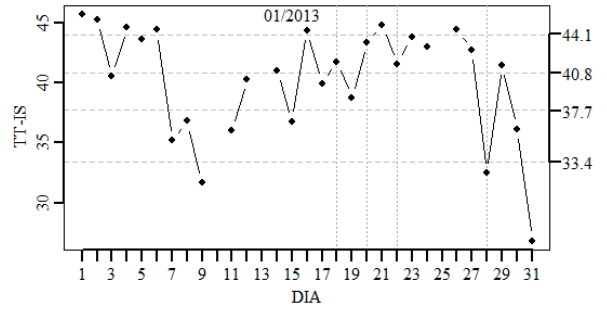
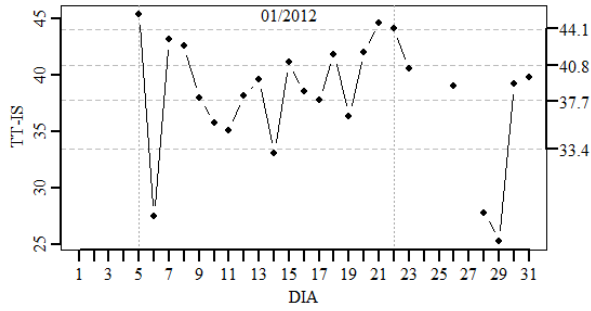
No mês de janeiro de 2012, só houve registro de chuva nos dias 5 e 22. O valor do IC também é alto na véspera do dia 21 [Figura 15 (l)]. Em janeiro de 2013, o valor do IC é alto nas vésperas dos dias de chuva 22 e 28 [Figura 15 (m)]. Em janeiro de 2014, o IC indica instabilidade alta na véspera dos dias de chuva 15 e 28, comportamento que também é visto na véspera de todos os dias de chuva de janeiro de 2015 (23, 25 e 26) [Figura 15 (n)].

Para o mês de janeiro de 2016, o índice indicou condição de instabilidade na véspera ou antevéspera da maioria dias, com exceção dos 17, 18, 19, 23 e 28 [Figura 15 (o)].

³ Dados indisponíveis para os meses de janeiro dos anos de 2002, 2006 e 2010 (ver Cap.3, Seção 3).

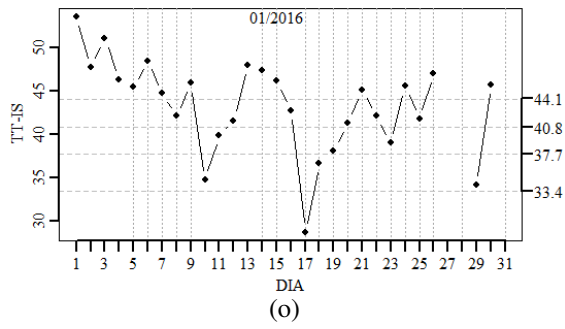
Figura 15 – Séries temporais do IC (TT-IS). Linhas tracejadas horizontais correspondem aos percentis 20%, 40%, 60% e 80 %. Dias com ocorrência de precipitação são sinalizados com linhas verticais pontilhadas.





(m)

(n)



(o)

4.2.3 Séries mensais do Índice K (IK)

O próximo índice analisado é o IK (Tabela 2), que é uma função do *lapse rate* entre os níveis 850 e 500 hPa, a temperatura do ponto de orvalho no nível de 850 hPa e a DPM (Figura 16).

Em janeiro de 1999, o IK possui valor acima do percentil 60% para os 3 dias de chuva do mês (7, 11 e 12) e também indica condição de instabilidade nas vésperas desses dias [Figura 16 (a)]. O IK indica instabilidade alta para a sequência de dias de chuva de 20 a 24 de janeiro de 2000 e também para o dia 28 do mesmo mês, mas não indicou para os dias 9, 10 e 14 [Figura 16 (b)]. Em janeiro de 2001, o IK possui valor acima do percentil 60% na antevéspera do dia de chuva 23, mas, para os dias de chuva 5 e 13, seu valor fica entre os percentis 40% e 60% [Figura 16 (c)].⁴

Em janeiro de 2003, o IK indica instabilidade alta nas vésperas dos dias de chuva 2, 9, 14 e 30 [Figura 16 (d)]. No mês de janeiro de 2004, o IK apresentou um desempenho melhor que o DMAP e o IC, indicando instabilidade alta para os dias de chuva 11, 13, 14, 16, 18, 19 e 21 [Figura 16 (e)]. No mês de janeiro de 2005, o IK apresentou um bom desempenho nas vésperas dos dias 11, 15, 16, 20, 23, 26 e 31, com valores variando acima do percentil 60%. Para três dias deste mês, no entanto, o IK não indicou instabilidade alta [Figura 16 (f)]. O IK também indicou instabilidade alta na véspera do dia 7, em janeiro de 2007 [Figura 16 (g)]. Em janeiro de 2008, o IK indicou instabilidade alta na antevéspera do dia de chuva 22 e também na véspera e no dia de chuva 31 [Figura 16 (h)].

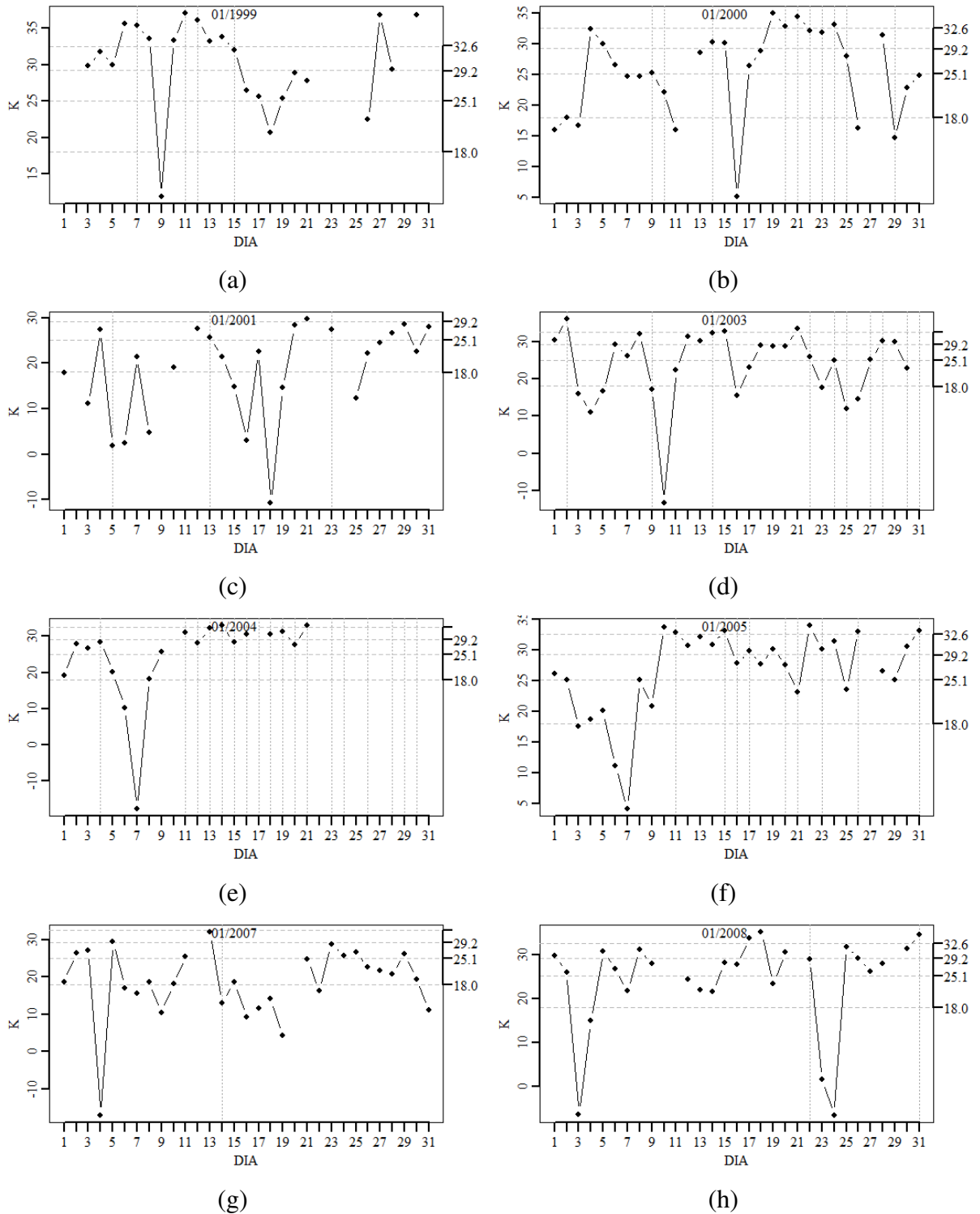
Em janeiro de 2009, o IK apresentou bom desempenho na previsão dos dias de chuva 17, 22 e 24, com indicativo de instabilidade variando acima do percentil 60% nos dias antecedentes aos dias de chuva [Figura 16 (i)]. Em janeiro de 2011, o IK também apresentou um bom desempenho na previsão de dias de chuva, indicando instabilidade alta nas antevésperas e vésperas dos dias de chuva (5, 11 e 13) [Figura 16 (j)].

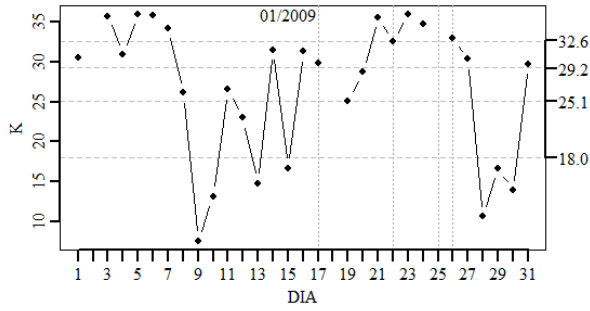
No mês de janeiro de 2012, o valor do IK é superior ao percentil 60% no dia de chuva 5, bem como na sequência de dias 21, 22 e 23 [Figura 16 (k)]. Em janeiro de 2013, o indicativo de instabilidade do IK ultrapassa o percentil 60% nas vésperas dos dias de chuva 22 e 28 [Figura 16 (l)]. Em janeiro de 2014, o IK acerta a previsão para os dias de chuva 10, 11 e 15 [Figura 16 (m)]. No mês de janeiro de 2015, o IK também apresentou indicativo de instabilidade alto na sequência de dias de 22 a 26, sendo que houve registro de chuva nos dias 23, 25 e 26 [Figura

⁴ Dados indisponíveis para os meses de janeiro dos anos de 2002, 2006 e 2010 (ver Cap.3, Seção 3).

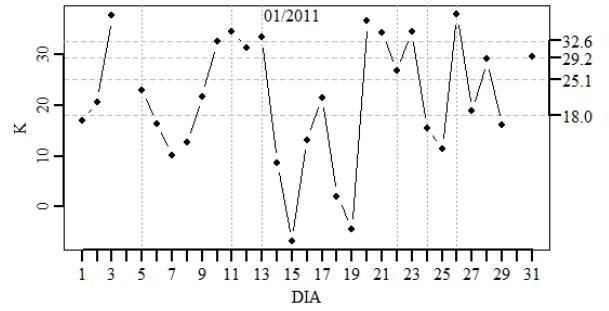
16 (n)]. Em janeiro de 2016, o indicativo de instabilidade do IK ultrapassou o percentil 60% para a maioria dos dias de chuva com sondagens disponíveis [Figura 16 (o)].

Figura 16 – Séries mensais do IK. Linhas tracejadas horizontais correspondem aos percentis 20%, 40%, 60% e 80%. Dias com ocorrência de precipitação são sinalizados com linhas verticais pontilhadas.

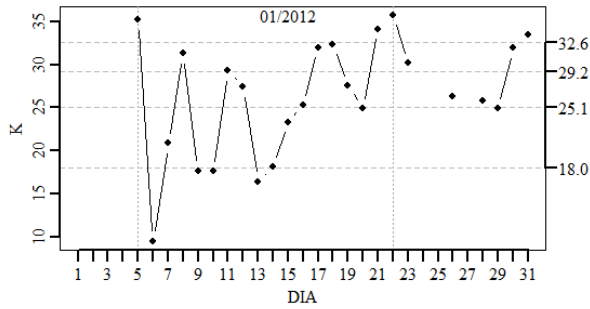




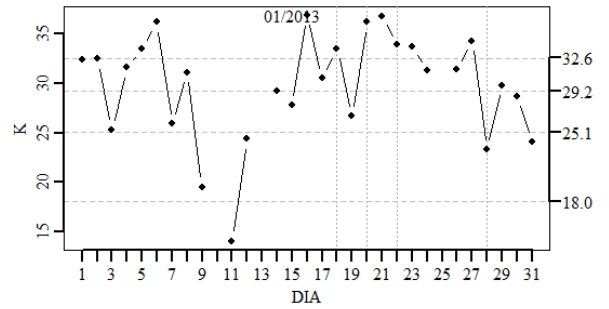
(i)



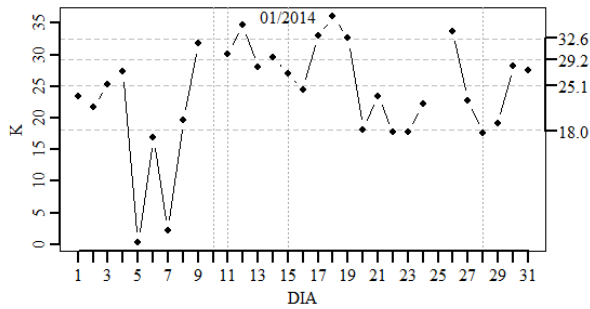
(j)



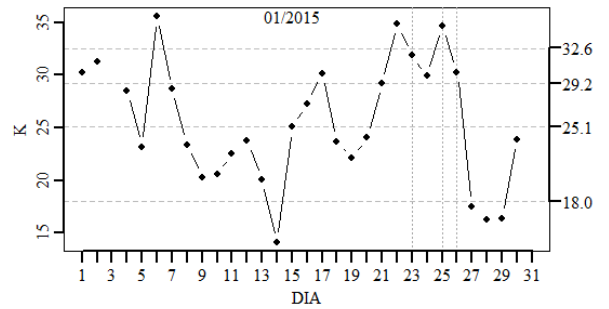
(k)



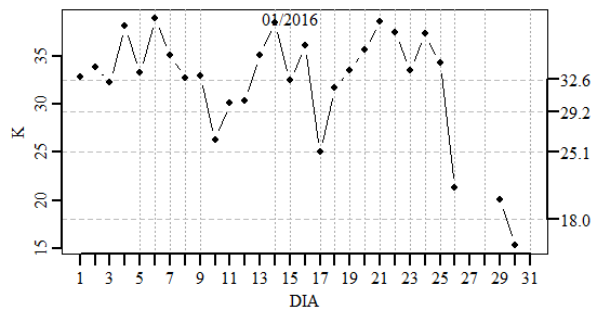
(l)



(m)



(n)



(o)

4.2.4 Índice da temperatura média da camada 700-500 hPa (TMED).

O comportamento da TMED também é analisado no presente estudo. Seu uso se justifica uma vez que há indício de resfriamento dessa camada nos dias com condições favoráveis à ocorrência de precipitação nos perfis verticais da região de Petrolina. Chamaremos este índice de TMED. Ele simula a quebra da inversão térmica de subsidência visível nos perfis verticais dessa região.

Em janeiro de 1999, o TMED acerta a previsão para três (7, 11 e 12) de quatro dias chuva do mês, apresentando valor inferior ao percentil 40% nas vésperas desses dias [Figura 17 (a)]. No mês de janeiro de 2000, o TMED acerta a previsão para sete dos 10 dias de chuva do mês, com indicativo de instabilidade alta nas vésperas dos dias 9, 10, 21, 22, 24, 25 e 29 [Figura 17 (b)]. No mês de janeiro de 2001, o TMED indica instabilidade alta nas vésperas dos dias de chuva 5 e 13, assim como na antevéspera do dia 23. Não houve observação de altitude para o dia 22 [Figura 17 (c)].⁵

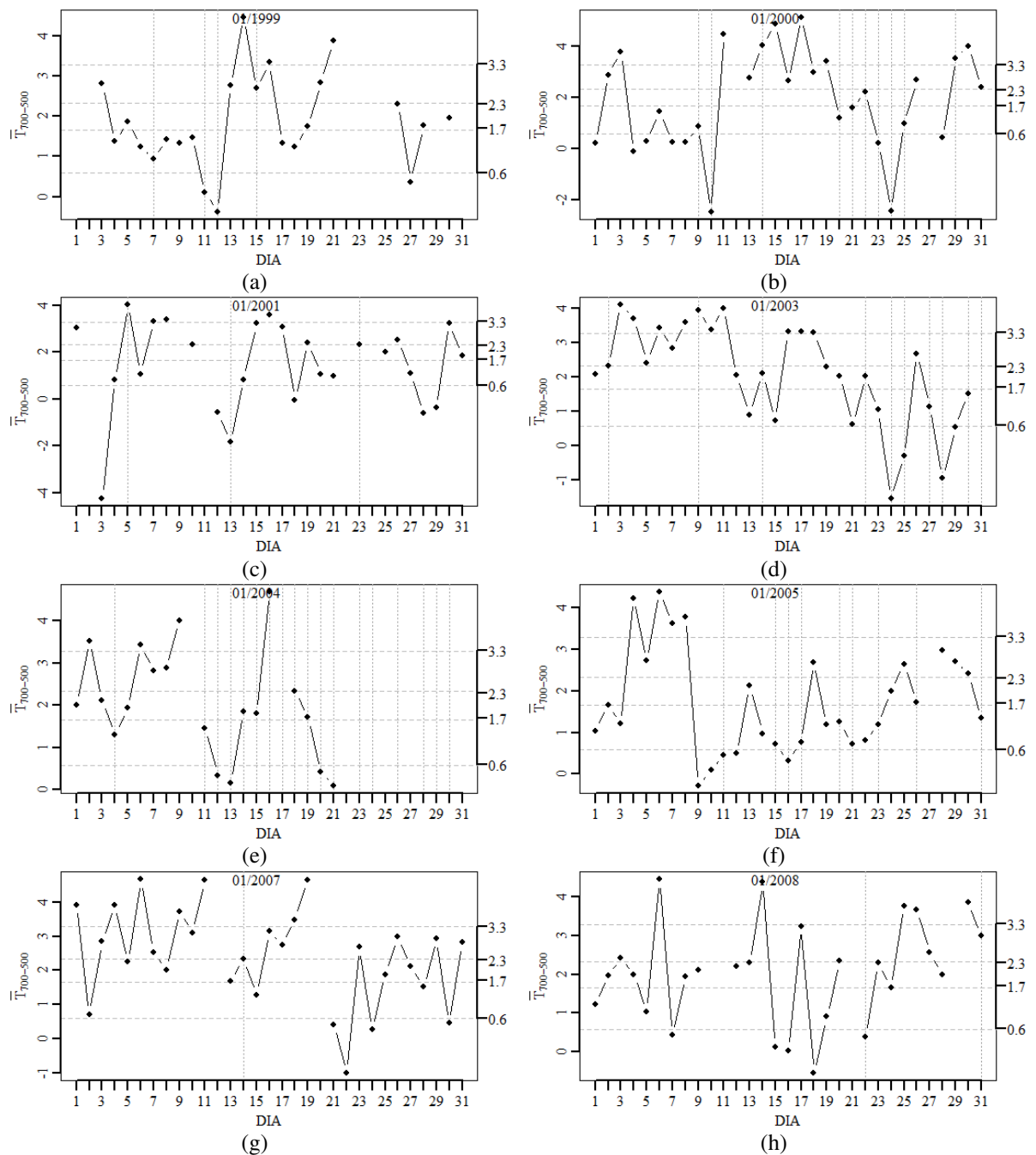
Em janeiro de 2003, o comportamento do TMED também é satisfatório para a previsão de dias de chuva. O TMED indica instabilidade alta nas vésperas de sete dos 10 dias com registro de chuva. O índice só não teve bom desempenho para os dias de chuva 2 e 9, no primeiro decêndio do mês [Figura 17 (d)]. Para o mês de janeiro de 2004, o TMED apresentou um desempenho inferior ao do IK (Tabela 3), porém superior ao DMAP (Figura 14) e ao IC, com indicativo de instabilidade alto para seis dos 12 dias de chuva com observação de altitude [Figura 17 (e)]. Para janeiro de 2005, o TMED indica instabilidade alta em oito dos nove dias com ocorrência de chuva, com indicativo de instabilidade alto na véspera dos dias 11, 15, 16, 17, 20, 21 e 23 [Figura 17 (f)]. Em 2007, o TMED não indicou instabilidade alta para o único dia de chuva do mês de janeiro [Figura 17 (g)], não tendo também um bom desempenho nas vésperas dos dias de chuva de 2008 [Figura 17 (h)]. O TMED indicou instabilidade alta na véspera de dois dias de chuva (17 e 21) de um total de quatro dias no mês de janeiro de 2009 [Figura 17 (i)]. No mês de janeiro de 2011, [Figura 17 (j)] o TMED também indicou instabilidade alta nas vésperas ou nos dias de chuva do mês, comportamento que também é visto para janeiro do ano de 2012 [Figura 17 (k)].

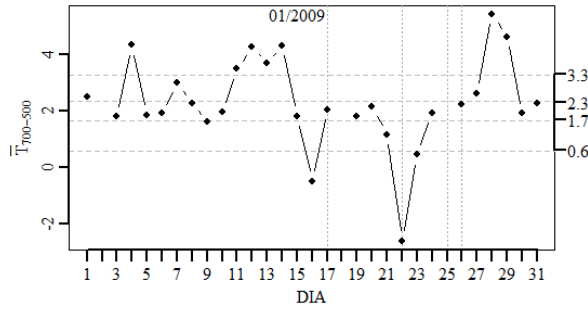
Em janeiro de 2013, o TMED apresenta indicativo de instabilidade alta em dois dos três dias chuvosos do mês (20 e 22) [Figura 17 (l)]. O índice não apresentou bom desempenho nos dias antecedentes aos eventos chuvosos. Já em janeiro de 2014, o comportamento do TMED

⁵ Dados indisponíveis para os meses de janeiro dos anos de 2002, 2006 e 2010 (ver Cap.3, Seção 3).

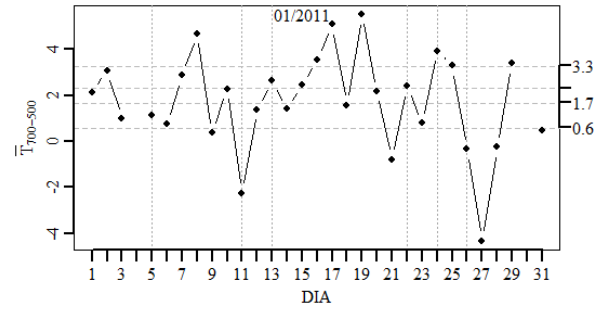
foi satisfatório nos dias antecedentes aos dias de chuva, já que o indicativo de instabilidade era alto nas vésperas dos dias 10, 15 e 29 [Figura 17 (m)]. Este comportamento também é visto em janeiro de 2015 [Figura 17 (n)]. No mês de janeiro de 2016, o desempenho do indicativo do TMED foi inferior aos do DMAP, IC e IK (Figura 14, Figura 15 e Figura 16), indicando instabilidade alta em apenas cinco dos 18 dias de chuva do mês [Figura 17 (o)].

Figura 17 – Séries mensais do Índice TMED. Linhas tracejadas horizontais correspondem aos percentis 20%, 40%, 60% e 80 %. Dias com ocorrência de precipitação são sinalizados com linhas verticais pontilhadas.

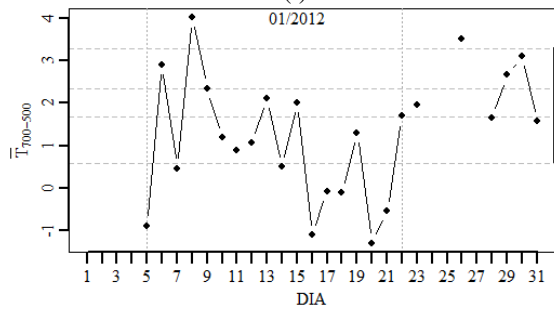




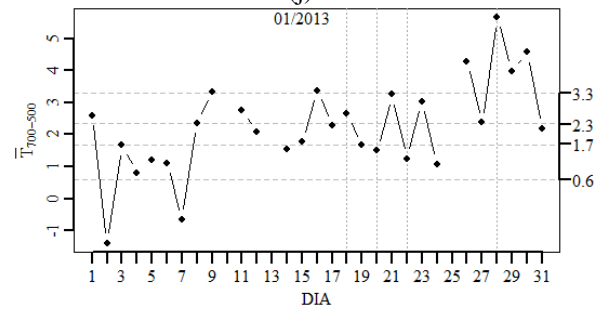
(i)



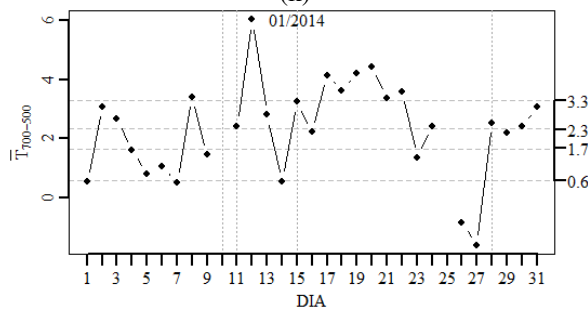
(j)



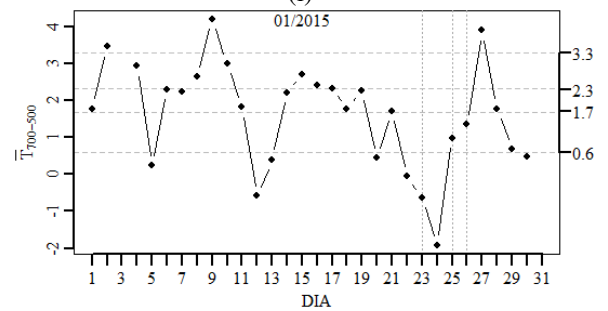
(k)



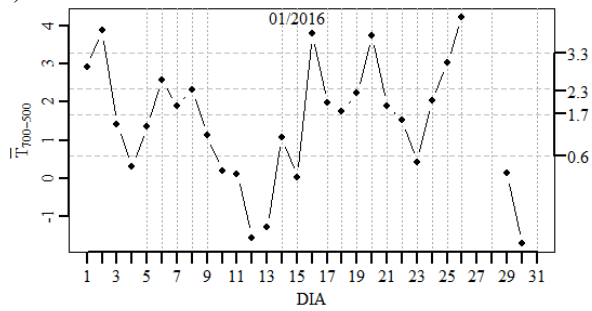
(l)



(m)



(n)



(o)

4.2.5 Índice da temperatura do nível de condensação por levantamento (TNCL).

O próximo índice analisado é o TNCL do nível de 925 hPa. Os dias favoráveis à ocorrência de chuva apresentam valores altos do TNCL, quando comparados aos dias sem condições favoráveis à ocorrência de chuva na área de Petrolina. Em um estudo sobre o mês extremo chuvoso de janeiro de 2004 na região de Petrolina, Correia *et al.* (2012) enfatizaram a relação entre o teor de umidade elevado nos níveis baixos e os valores altos do TNCL.

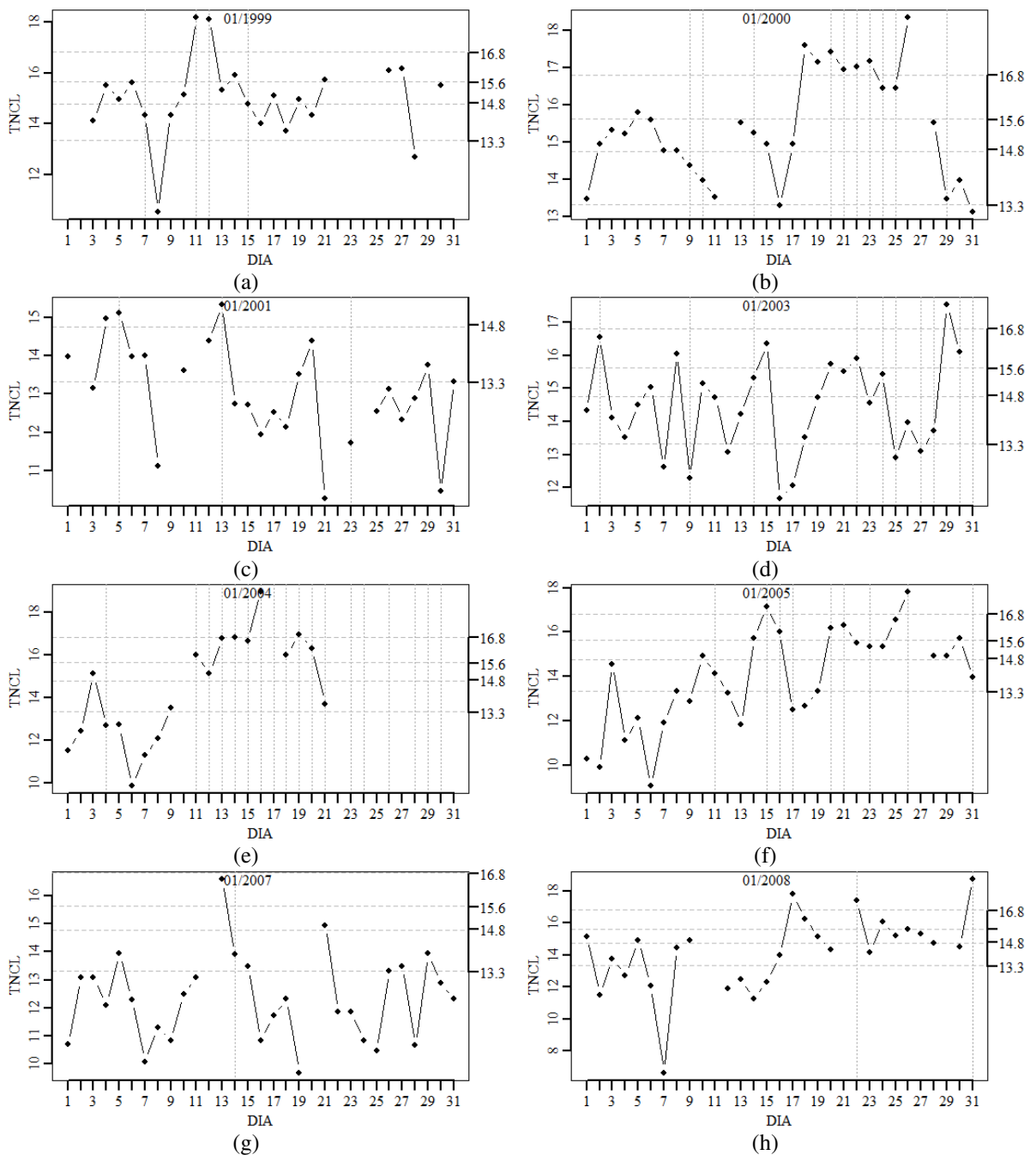
No mês de janeiro de 1999, o TNCL apresenta valores acima do percentil 60% nos dias de chuva 11 e 12, assim como na véspera do dia de chuva 15 [Figura 18 (a)]. Para o ano 2000, o TNCL não indica instabilidade alta para os dias de chuva 9, 10 e 14, mas indica instabilidade alta na sequência de dias de chuva de 20 a 25 do mês de janeiro daquele ano [Figura 18 (b)].

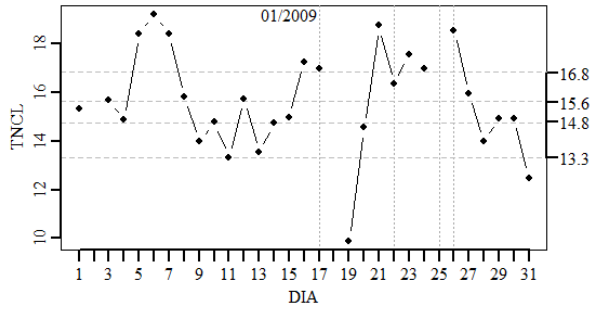
Em 2001 [Figura 18 (c)], o TNCL não indicou instabilidade alta em nenhum dos dias de chuva do mês, enquanto que, em janeiro de 2003, apresenta indicativo de instabilidade alto para os dias 2, 15 e 30.⁶ Para os dias de chuva entre 24 e 28, no entanto, o TNCL não indicou instabilidade alta [Figura 18 (d)]. Já para 2004 o TNCL apresenta um desempenho melhor que os demais índices já analisados. O valor do TNCL é alto em oito dos 12 dias de chuva com observação de ar superior [Figura 18 (e)]. Para janeiro de 2005, o TNCL indica instabilidade alta em oito dos nove dias com ocorrência de chuva [Figura 18 (f)]. O indicativo de instabilidade do TNCL é alto na véspera dos dias de chuva 15, 16, 21, 26 e 31. No mês de janeiro de 2007, o TNCL é alto na véspera do único dia de chuva registrado nesse mês (dia 14) [Figura 18 (g)]. Em janeiro de 2008, o TNCL é alto nos dias de chuva 22 e 31 [Figura 18 (h)]. Em janeiro de 2009, o TNCL apresenta melhor desempenho em relação aos outros meses de janeiro analisados, apresentando valor alto do TNCL nas vésperas dos dias de chuva. O TNCL indica instabilidade com um dia de antecedência para os dias de chuva 17, 22 e 25, da mesma forma como também apresenta um valor alto para o dia 26, cuja véspera não teve observação de altitude [Figura 18 (i)]. Em janeiro de 2011, o valor do TNCL é alto na maior parte do mês, incluindo os dias de chuva e suas respectivas vésperas [Figura 18 (j)]. Em janeiro de 2012, o TNCL apresentou indicativo de instabilidade alta na véspera e no dia de chuva 21 [Figura 18 (k)]. Para janeiro de 2013, o TNCL acerta a previsão dos dias de chuva 18, 20, 22 e 28, já que o valor é alto nas vésperas desses dias [Figura 18 (l)]. Para 2014, o TNCL não possui valor alto na véspera do dia de chuva 10, mas, a partir do dia 11, o TNCL indicou instabilidade alta nos

⁶ Dados indisponíveis para os meses de janeiro dos anos de 2002, 2006 e 2010 (ver Cap.3, Seção 3).

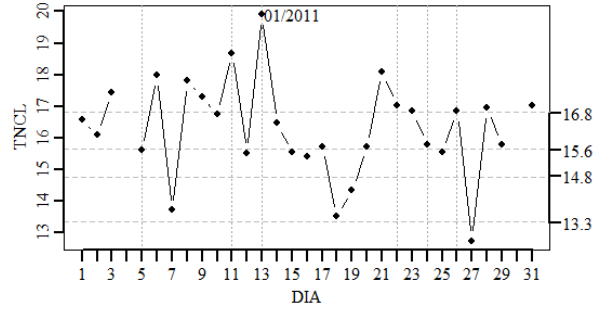
dias de chuva 15 e 28 [Figura 18 (m)]. Para janeiro de 2015, a relação do TNCL com a chuva também pode ser observada nas vésperas dos dias de chuva 23, 25 e 27 [Figura 18 (n)]. O índice se mantém acima do percentil 60 %, indicando instabilidade mais alta. Já no mês de janeiro de 2016, o TNCL apresentou indicativo de instabilidade alta em praticamente todo o mês, mantendo-se acima do percentil 80% em 26 dias do mês.

Figura 18 – Séries temporais do Índice TNCL. Linhas tracejadas horizontais correspondem aos percentis 20%, 40%, 60% e 80 %. Dias com ocorrência de precipitação são sinalizados com linhas verticais pontilhadas.

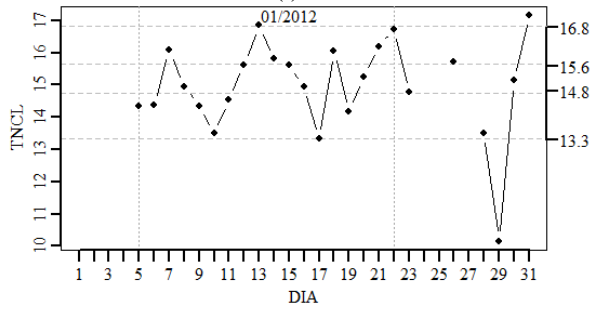




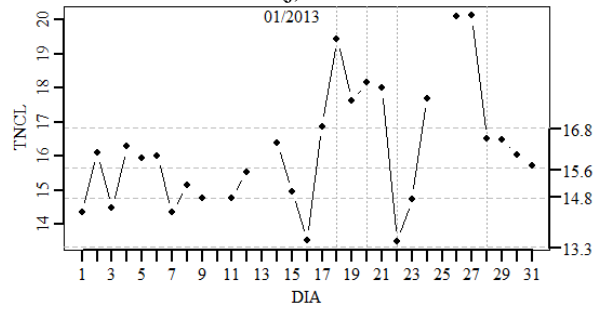
(i)



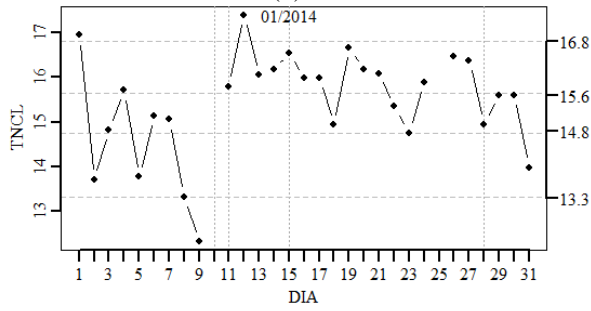
(j)



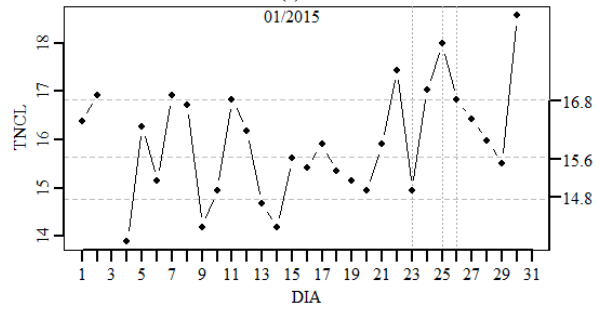
(k)



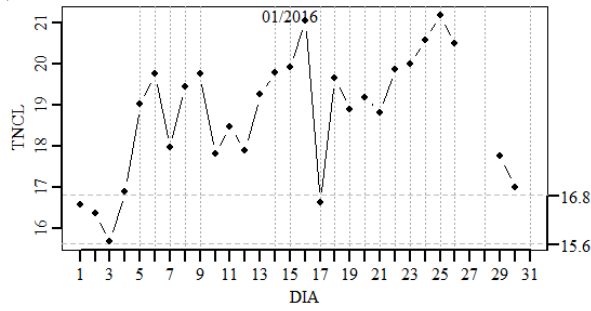
(l)



(m)



(n)



(o)

4.2.6 Índice de instabilidade potencial (PII)

O PII é outro índice analisado neste estudo. Conforme Vandelden (2001), a atividade de tempestades está associada à instabilidade potencial. Os fatores que mais contribuem para o aumento de instabilidade potencial são advecção de ar quente na baixa troposfera, bloqueio de advecção de ar frio por montanhas, acúmulo de umidade em baixos níveis e o aquecimento da superfície da terra pelo sol.

Para o mês de janeiro de 1999, o PII indica instabilidade potencial alta na véspera dos dias de chuva 7 e 15. Para os dias 11 e 12 de chuva, o PII indica instabilidade potencial baixa, com valor abaixo do percentil 20% [Figura 19 (a)].

Para janeiro de 2000, o PII indica instabilidade potencial alta — acima do percentil 60% — na véspera dos dias de chuva 7 e 15, na antevéspera do dia de chuva 11, sendo que este índice indica um aumento da instabilidade potencial do dia 17 ao dia 19, dias estes que antecedem à sequência de dias de chuva que vai de 19 a 23 [Figura 19 (b)].

Em janeiro de 2001, o PII indica instabilidade potencial alta na véspera do dia de chuva 13 [Figura 19 (c)].⁷ No mês de janeiro de 2003, é possível verificar que a instabilidade potencial aumenta nos dias antecedentes aos dias de chuva e diminui nos dias de chuva. A série do PII indica aumento da instabilidade potencial nos dias antecedentes aos dias de chuva ou nos dias antecedentes às sequências de dias de chuva. Podemos ver este aumento do PII nos intervalos de 5 a 9, de 10 a 14 de janeiro, estando os valores máximos do PII localizados de 17 a 21 de janeiro de 2003 [Figura 19 (d)].

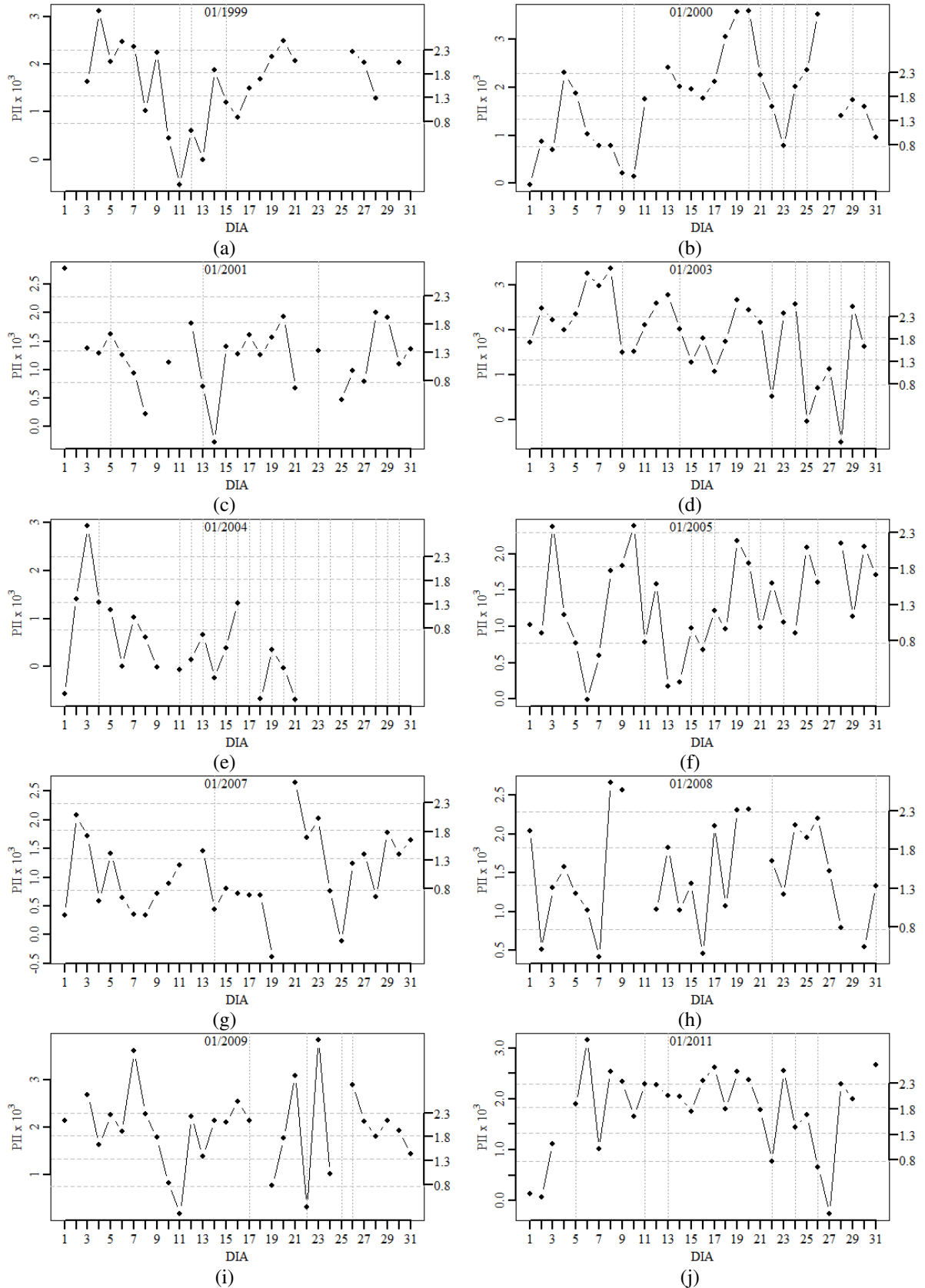
No mês de janeiro de 2004, o PII aumenta do dia 1 ao dia 3, dia que antecede o dia de chuva 4. Nos demais dias do mês, a instabilidade potencial foi baixa [Figura 19 (e)]. O aumento do indicativo de instabilidade do índice PII também ocorre nos dias antecedentes a dias de chuva no mês de janeiro de 2005 [Figura 19 (f)]. Em janeiro de 2007, o PII não indicou instabilidade alta na véspera do dia 7, mas é possível observar o aumento do PII de valores abaixo do percentil 20% para valores da ordem do percentil 40% na véspera do dia de chuva 7 [Figura 19 (g)]. Em janeiro de 2008, o PII indica alta instabilidade potencial no dia 19, antevéspera do dia de chuva 22. Não houve observação de altitude para o dia 20. O PII indica condição de instabilidade potencial alta nas vésperas dos dias de chuva 17 e 22 do mês de janeiro de 2009 [Figura 19 (h)]. Para janeiro de 2011 o PII indica instabilidade potencial alta, em torno do percentil 60%, durante a maior parte do mês. Há instabilidade potencial alta no dia

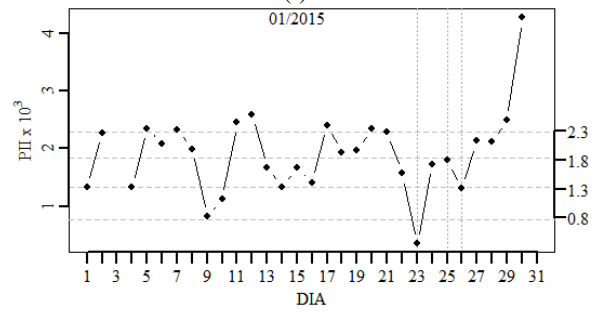
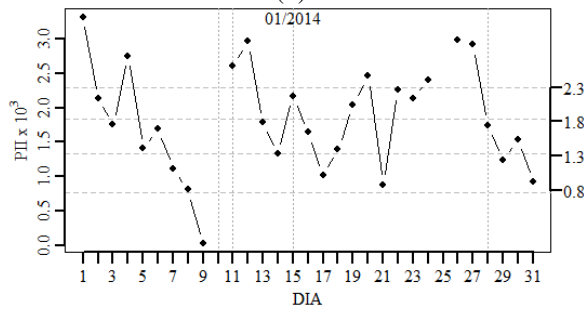
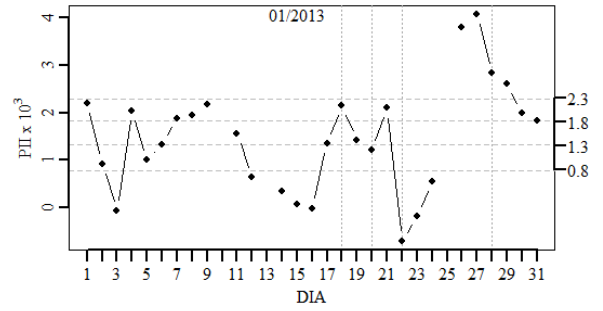
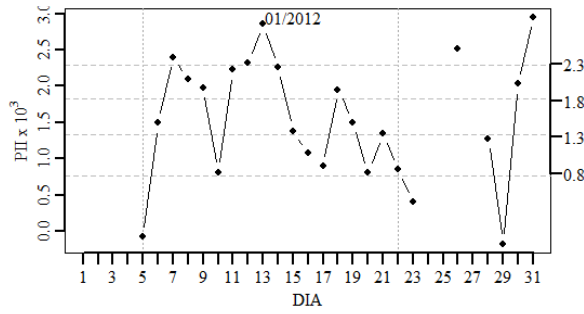
⁷ Dados indisponíveis para os meses de janeiro dos anos de 2002, 2006 e 2010 (ver Cap.3, Seção 3).

de chuva 5, 11, 13 e na véspera do dia de chuva 24 [Figura 19 (i)]. Para janeiro de 2012, que só teve dois dias de chuva no mês, o indicativo de instabilidade do PII é baixo nos dias de chuva 5 e 22 [Figura 19 (j)].

No mês de janeiro de 2013, o PII teve bom desempenho nas vésperas dos dias de chuva 18, 22 e 28 [Figura 19 (k)]. Já no mês de janeiro de 2014, o PII indicou instabilidade potencial alta nos dias de chuva 11 e 15, bem como no dia 27, que antecede o dia de chuva 28 [Figura 19 (m)]. No mês de janeiro de 2015, o PII não indicou instabilidade alta nos dias de chuva do mês [Figura 19 (n)]. Já em janeiro de 2016, o índice indica instabilidade potencial baixa para a maioria dos dias de chuva do mês, mas, nas vésperas dos dias de chuva, houve instabilidade potencial alta [Figura 19 (o)].

Figura 19 – Séries mensais do PII. Linhas tracejadas horizontais correspondem aos percentis 20%, 40%, 60% e 80 %. Dias com ocorrência de precipitação são sinalizados com linhas verticais pontilhadas.



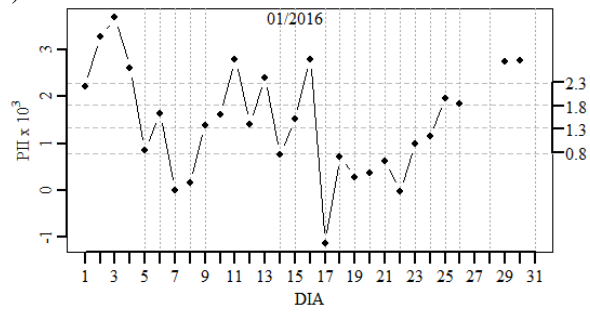


(k)

(l)

(m)

(n)



(o)

4.2.7 Índice da depressão psicrométrica média da camada 700-500 (DPM)

Outro índice analisado neste estudo é o Índice da DPM. Seu uso se justifica devido à variação do teor de umidade na média troposfera ter relação com a ocorrência de chuva na área de Petrolina. Nos dias com condições favoráveis à ocorrência de chuva, o teor de umidade na baixa média troposfera é alto, enquanto ele é baixo nos dias sem condições favoráveis à ocorrência de chuva. Este índice indica instabilidade alta para valores baixos. Assim, quanto mais próximo de zero, maior será o teor de umidade na baixa média troposfera e, portanto, maior a possibilidade de chuva.

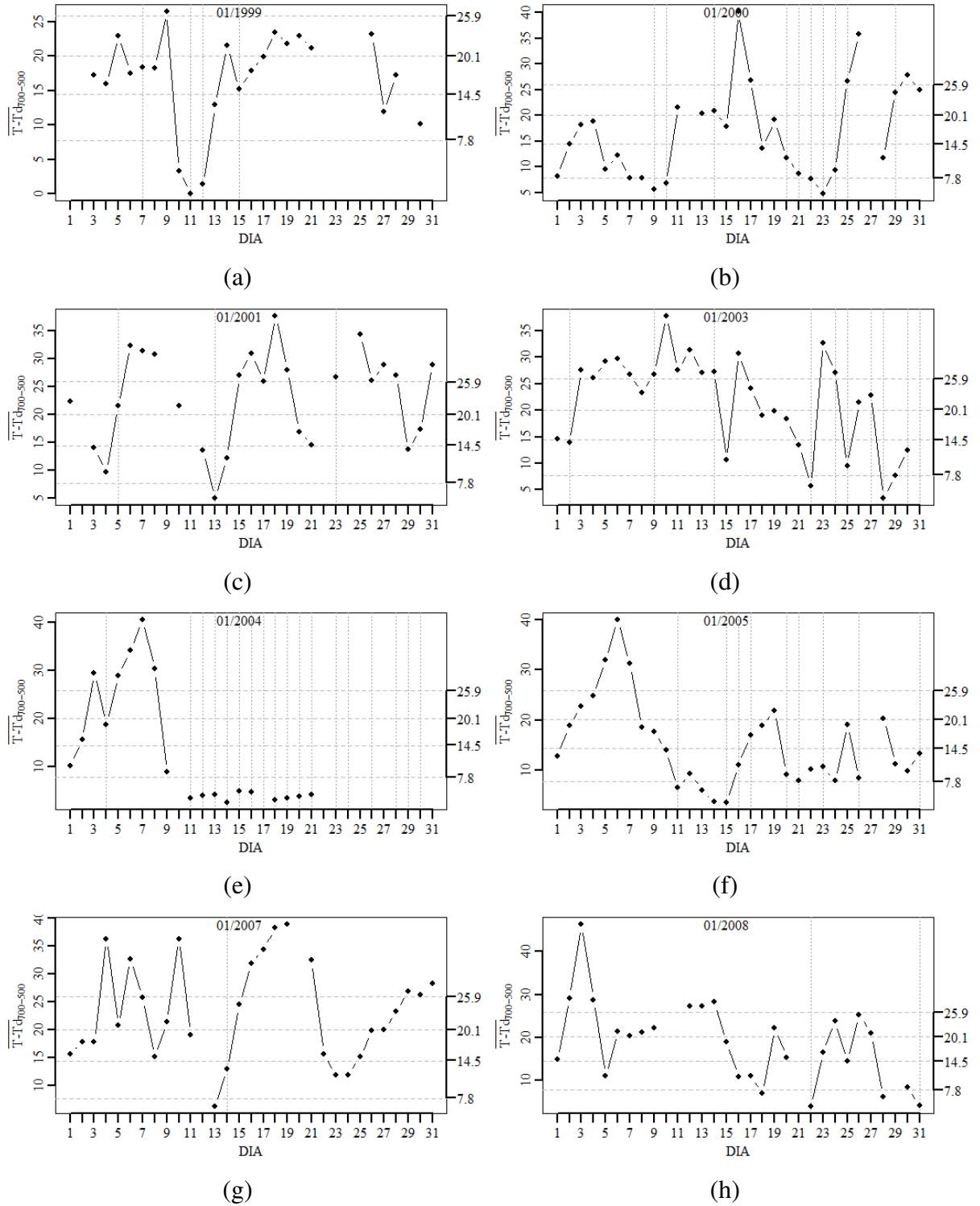
Em janeiro de 1999, o DPM indica instabilidade alta na véspera e nos dias de chuva 11 e 13. Nos dias de chuva 7 e 15, o DPM varia em torno dos percentis 50 e 60%, indicando que o levantamento de umidade não foi tão forte nesses dias [Figura 20 (a)]. Para o ano de 2000, o DPM indica forte levantamento de umidade nos dias antecedentes aos dias de chuva 9 e 10, na sequência de dias de chuva de 21 à 24 e na véspera do dia 29 [Figura 20 (b)]. Na série do DPM para o mês de janeiro de 2001, é possível constatar a diminuição do valor do DPM nos dias antecedentes aos dias de chuva 5, 13 e 23, indicando um bom desempenho do índice na detecção de condições favoráveis à precipitação [Figura 20 (c)].⁸ Em janeiro de 2003, no entanto, o comportamento observado para janeiro de 2001 só é visto para os dias 1, 21 e 28 [Figura 20 (d)]. Para o mês de janeiro de 2004, o DPM indicou forte levantamento de umidade para os dias de chuva a partir do dia 11 [Figura 20 (e)]. Mesmo comportamento também é visto para o mês de janeiro de 2005 [Figura 20 (f)].

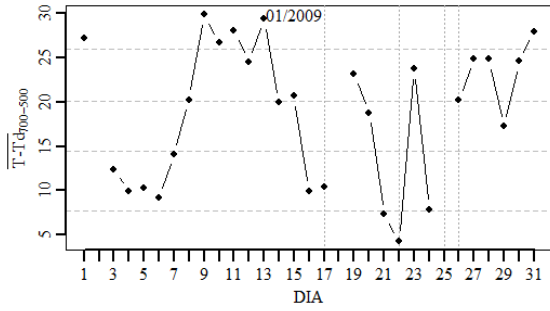
Em janeiro de 2007, observa-se que o DPM indica umidade alta na véspera do dia de chuva 14, com valor inferior ao percentil 20% [Figura 20 (g)]. Também verificamos esse comportamento para os dias de chuva 22 e 31 do mês de janeiro de 2008 [Figura 20 (h)].

O indicativo de instabilidade dos dias DPM também é alto nas vésperas dos demais meses estudados. [Figura 20 (i), (j), (k), (m), (n) e (o)].

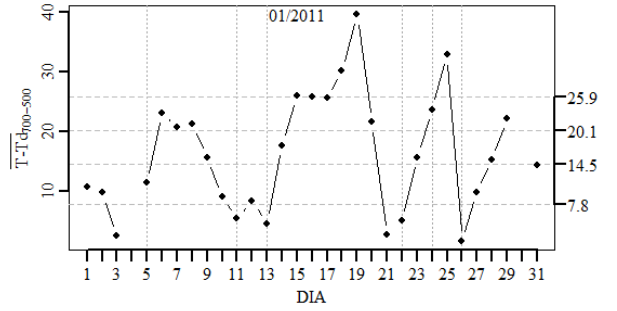
⁸ Dados indisponíveis para os meses de janeiro dos anos de 2002, 2006 e 2010 (ver Cap.3, Seção 3).

Figura 20 – Séries temporais do Índice DPM. Linhas tracejadas horizontais correspondem aos percentis 20%, 40%, 60% e 80 %. Dias com ocorrência de precipitação são sinalizados com linhas verticais pontilhadas.

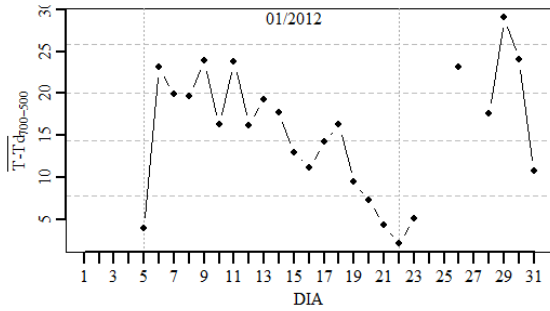




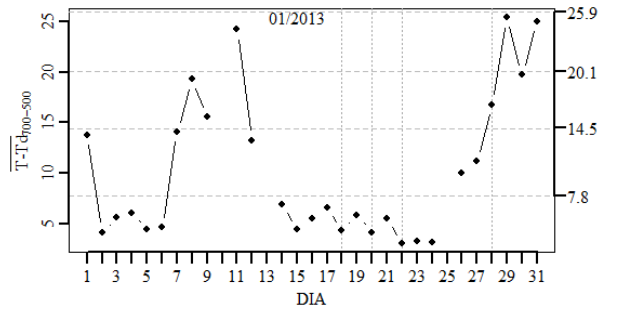
(i)



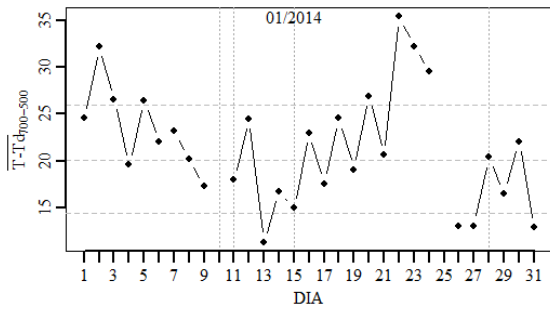
(j)



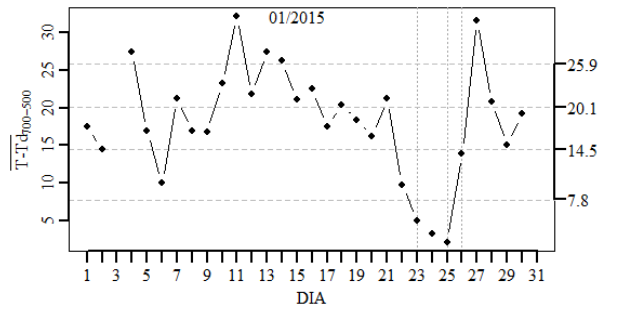
(k)



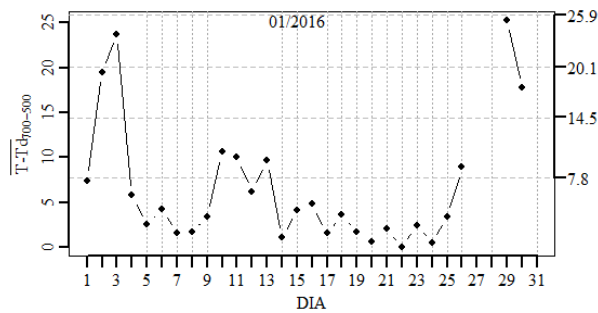
(l)



(m)



(n)

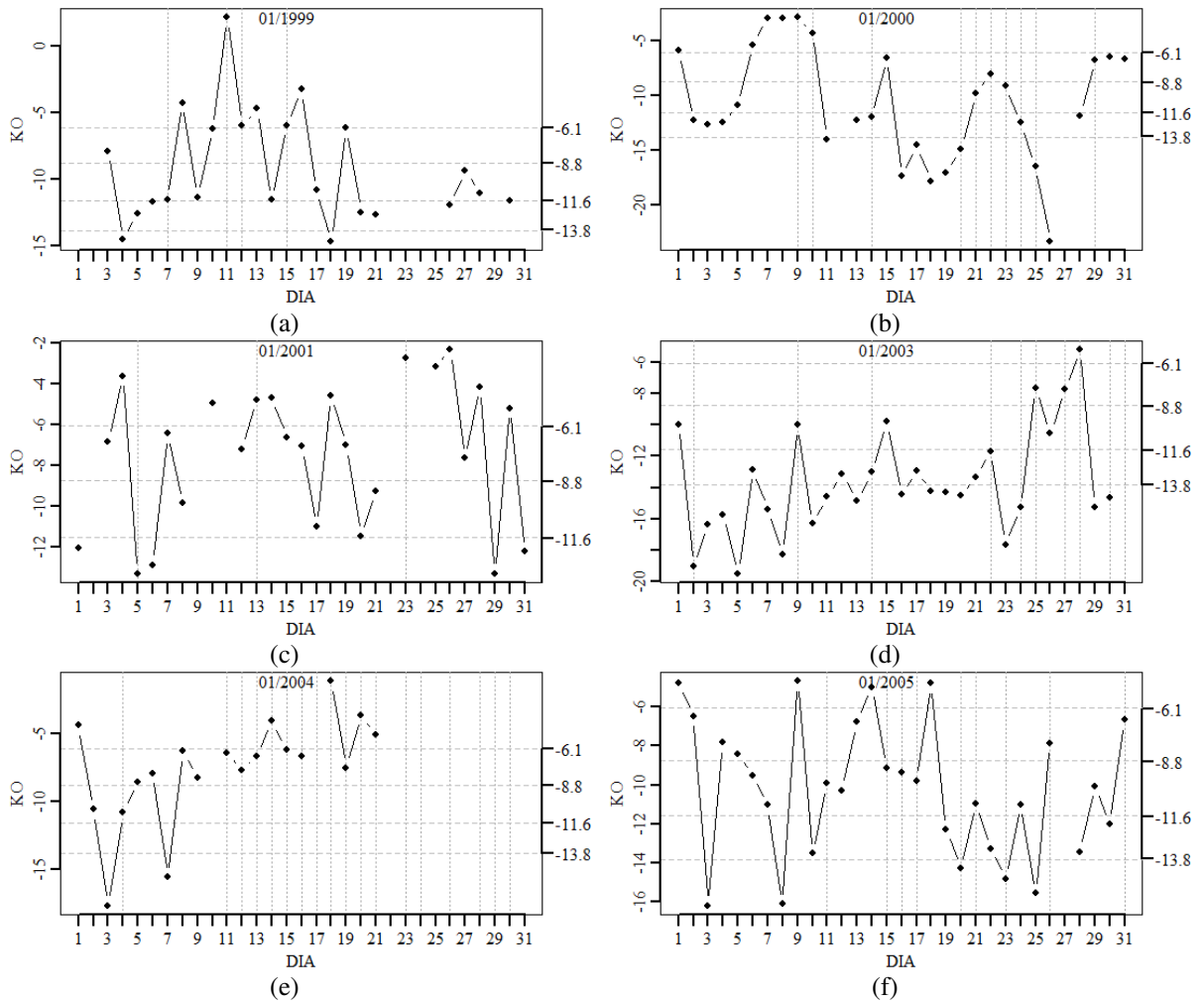


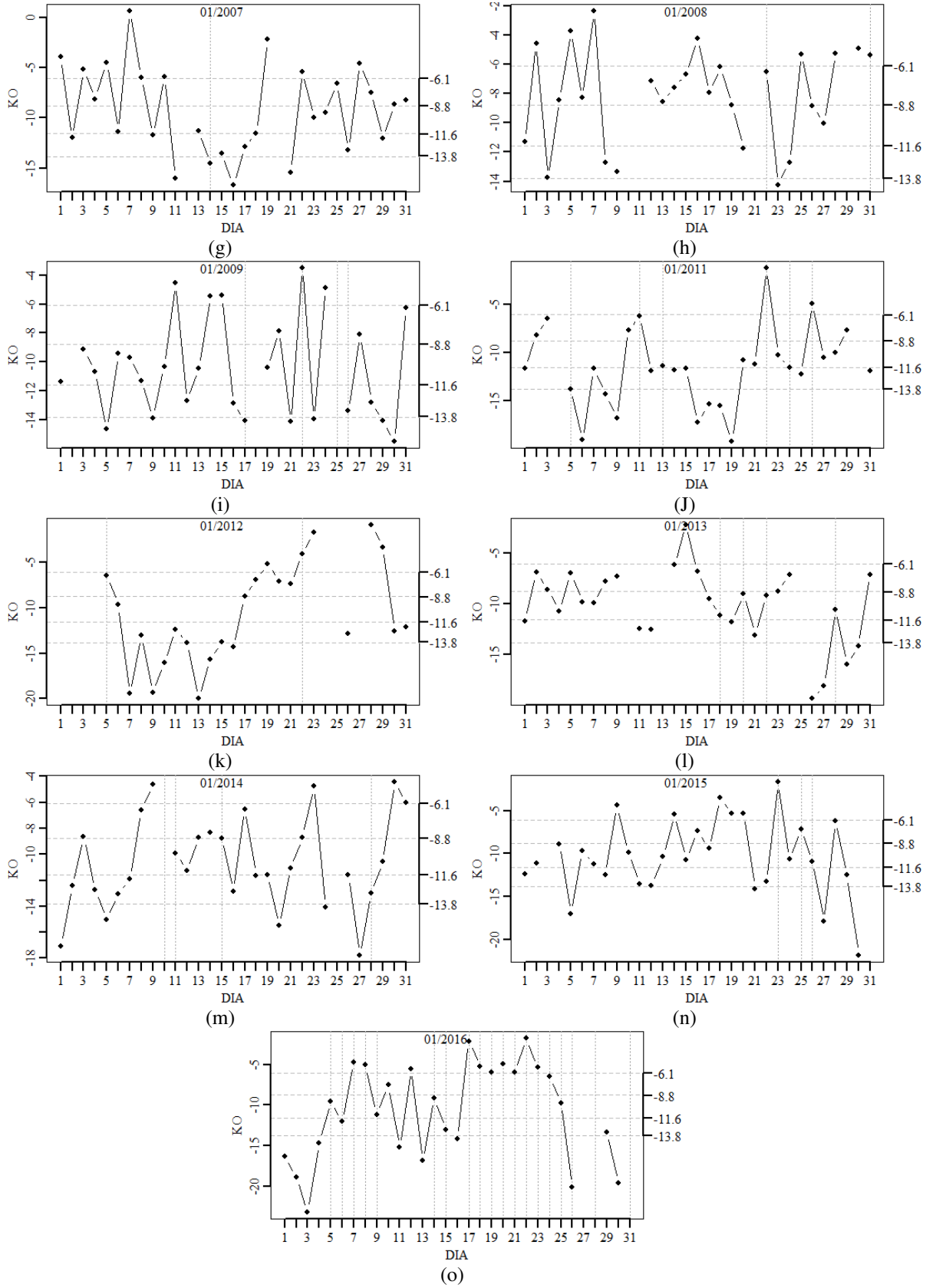
(o)

4.2.8 Índice KO

O KO é um índice que mede a instabilidade potencial entre a superfície e a média troposfera. Como está evidente na Figura 21 este índice não representa bem a instabilidade atmosférica da região de Petrolina. Em algumas ocasiões, o índice apresenta indicativo de instabilidade alto em alguns dias antecedentes a ocorrência de chuva, mas na maioria das vezes o KO não apresenta um bom desempenho como indicador da ocorrência de chuvas. Sua variabilidade pode estar associada com a natureza dos sistemas atuantes na região.

Figura 21 – Séries temporais do KO. Linhas tracejadas horizontais correspondem aos percentis 20%, 40%, 60% e 80 %. Dias com ocorrência de precipitação são sinalizados com linhas verticais pontilhadas.





4.2.9 Índice SWEAT

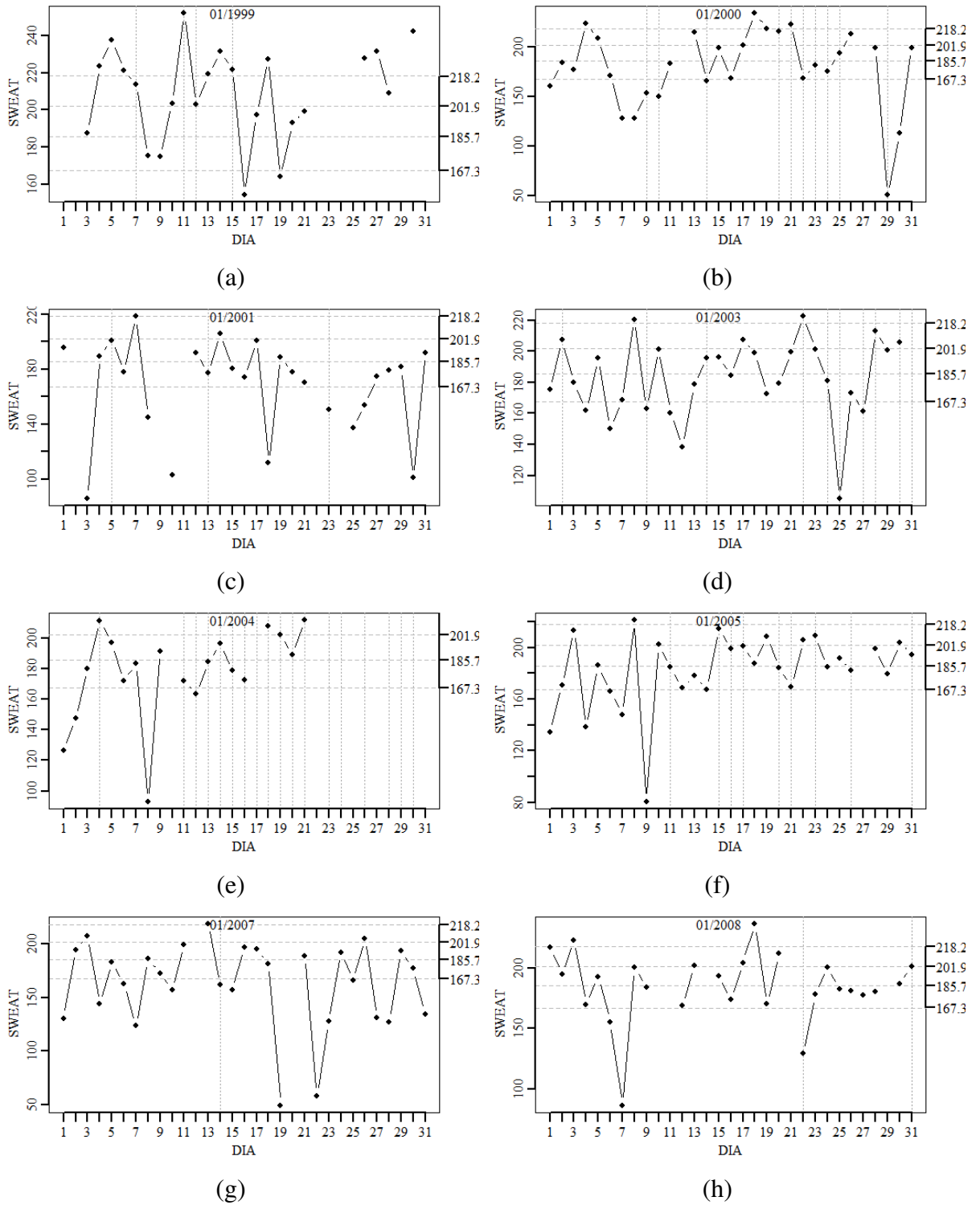
O SWEAT é um índice associado ao teor de umidade nos baixos níveis e à direção e velocidade do vento na baixa e média troposfera.

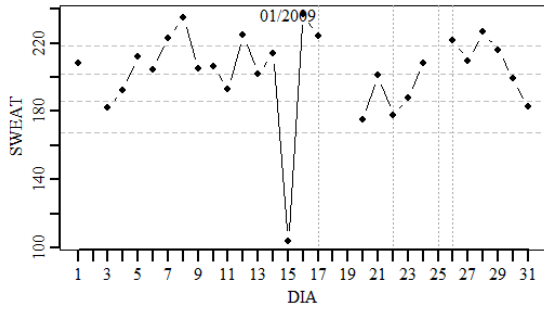
Em janeiro de 1999, o SWEAT possui valor acima do percentil 60% para os 3 dias de chuva do mês (7, 11 e 12) e também indica condição de instabilidade nas vésperas desses dias [Figura 22 (a)]. O SWEAT indica instabilidade alta para a sequência de dias de chuva de 20 a 22 de janeiro de 2000 e também para o dia 28, mas não indicou para os dias 9 e 10 [Figura 22 (b)]. Em janeiro de 2001, o SWEAT possui valor entre os percentis 20% e 40% na antevéspera do dia de chuva 23; e entre os percentis 40% e 60% na véspera dos dias de chuva 5 e 13 [Figura 22 (c)].⁹

De janeiro de 2003 a janeiro de 2009, o SWEAT indica instabilidade moderada nos dias antecedentes à chuva. O indicativo de instabilidade desse índice tende a oscilar em torno dos percentis 40% e 60%. De 2010 a 2016, o SWEAT apresenta um bom desempenho nas vésperas e nos dias de chuva, tendendo a se concentrar com grande frequência acima do percentil 80%, configurando-se como um índice com bom desempenho para a previsão das chuvas.

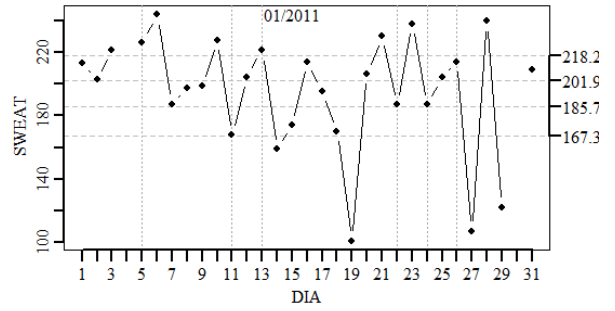
⁹ Dados indisponíveis para os meses de janeiro dos anos de 2002, 2006 e 2010 (ver Cap.3, Seção 3).

Figura 22 – Séries temporais do SWEAT. Linhas tracejadas horizontais correspondem aos percentis 20%, 40%, 60% e 80 %. Dias com ocorrência de precipitação são sinalizados com linhas verticais pontilhadas.

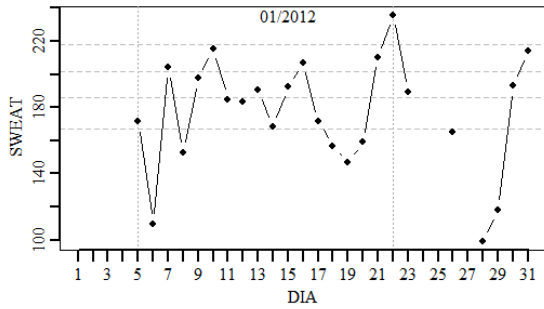




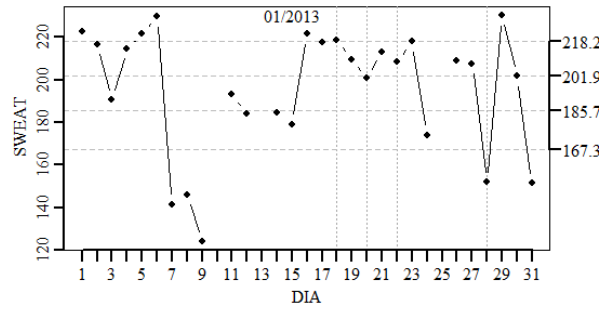
(i)



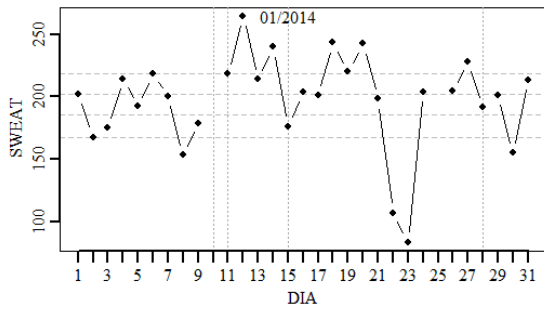
(j)



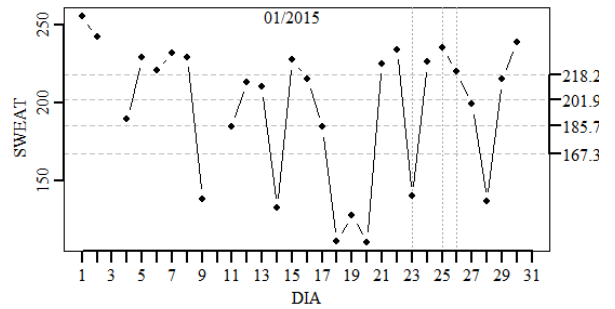
(k)



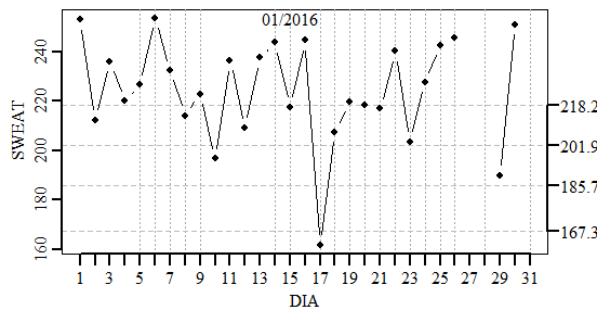
(l)



(m)



(n)

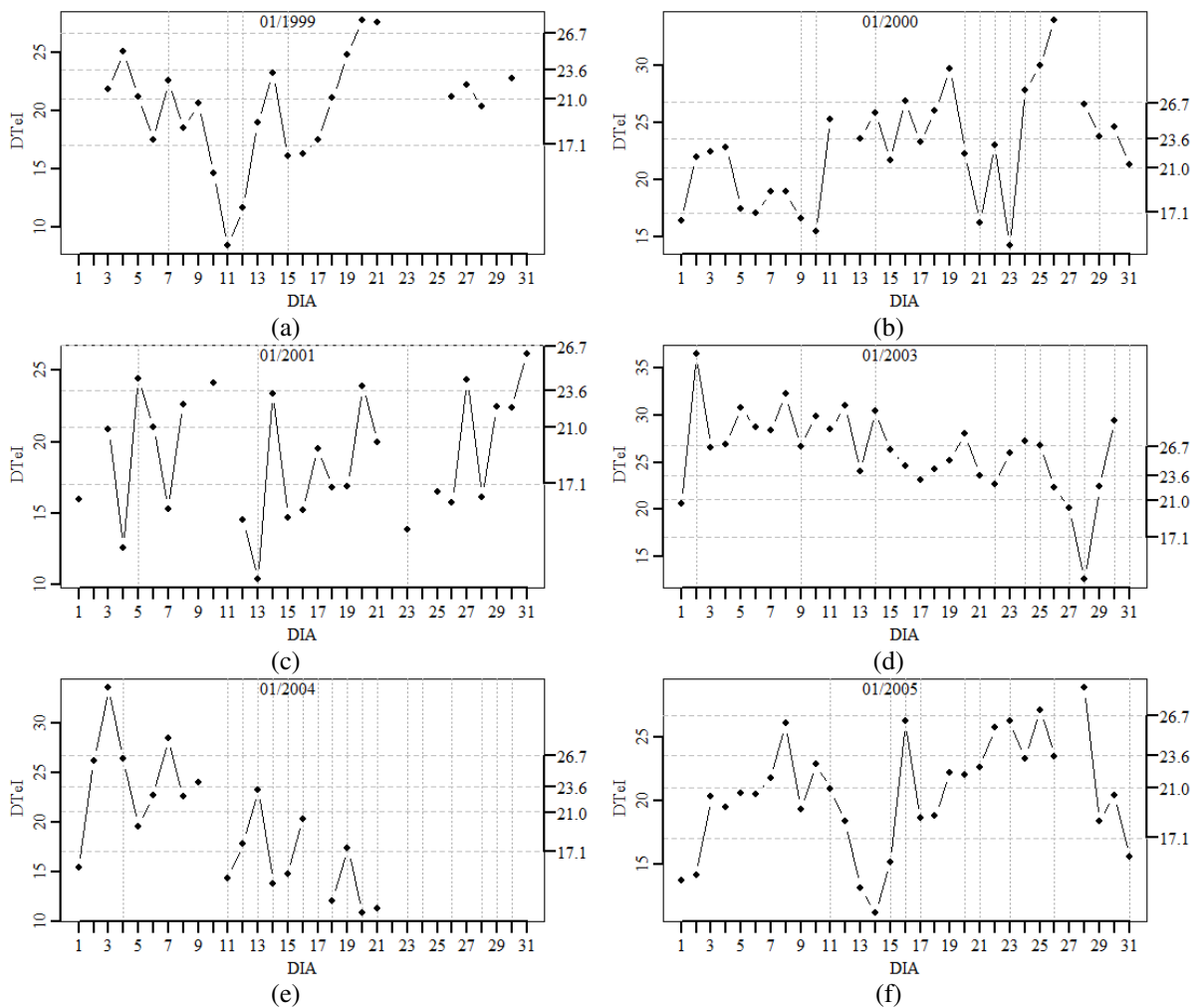


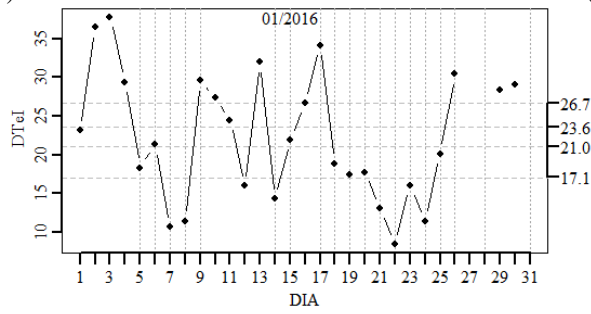
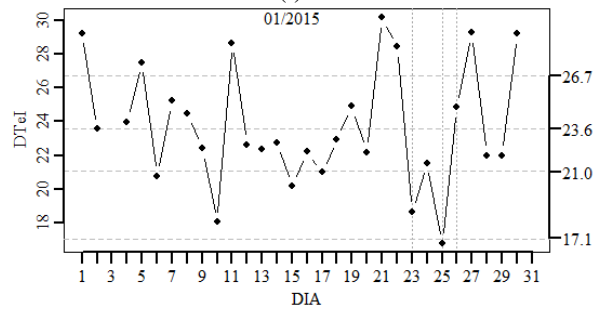
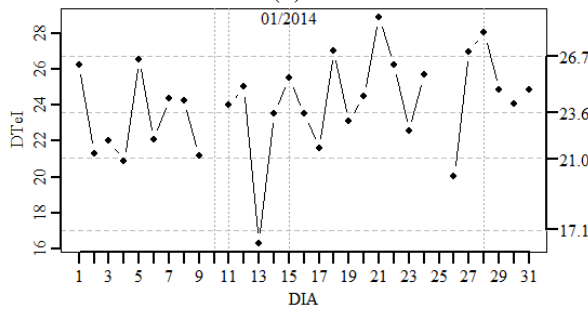
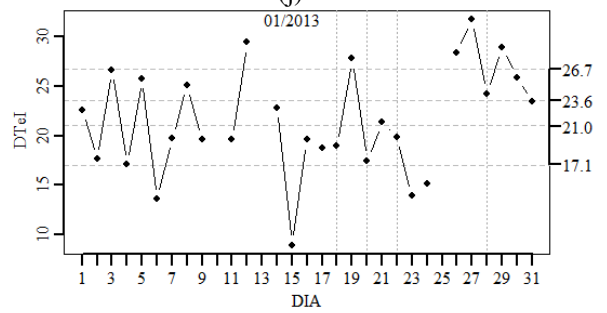
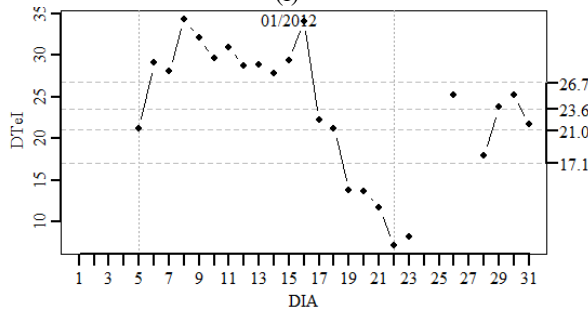
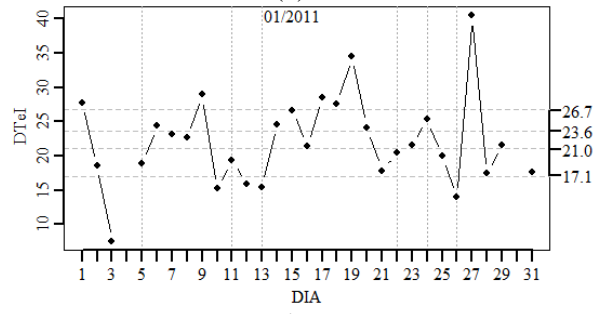
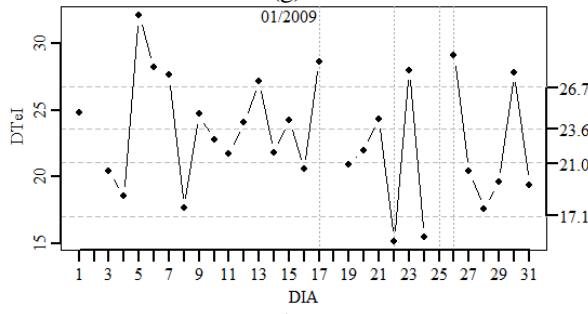
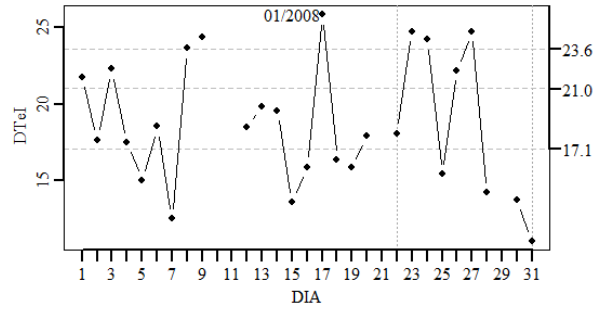
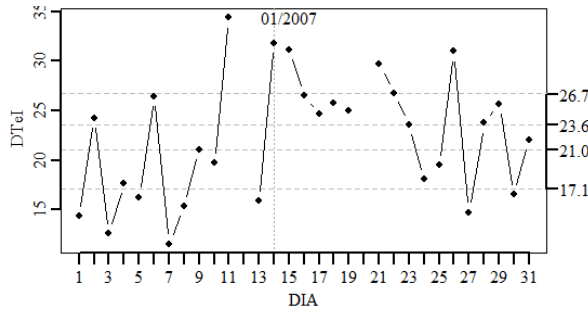
(o)

4.2.10 Índice Delta Teta-E (DTel)

O DTel pode ser entendido como um gradiente do teor de umidade entre a superfície e a média troposfera, obtido a partir da temperatura potencial equivalente. Como está evidente na Figura 22 este índice também não representa bem a instabilidade atmosférica da região de Petrolina nas situações antecedentes a ocorrência de chuva. Em algumas ocasiões, o índice apresenta comportamento semelhante ao DPM, mas na maioria das vezes, o DTel não apresenta bom desempenho como índice preditor da ocorrência de chuvas. Sua variabilidade pode estar associada com a natureza dos sistemas atuantes na região.

Figura 23 – Séries temporais do DTel. Linhas tracejadas horizontais correspondem aos percentis 20%, 40%, 60% e 80%. Dias com ocorrência de precipitação são sinalizados com linhas verticais pontilhadas.





(m)

(n)

(o)

4.2.11 Desempenho dos índices de estabilidade

No conjunto de índices de meses de janeiro, foram identificados dias sem chuva, vésperas de chuva, dias de chuva e dias sucedente aos dias de chuva. Para ilustrar o comportamento geral do indicativo de instabilidade dos índices, foram confeccionados gráficos *boxplots* desses conjuntos. Em seguida, os índices são comparados aos valores das ordens quantílicas de 20%, 40%, 60% e 80%. Estes índices estão representados nos gráficos *boxplots* da Figura 24 por linhas horizontais que terminam no eixo vertical da direita com os valores dos percentis do conjunto de dados.

Os índices que indicam instabilidade para valores negativos indicam maior probabilidade de chuva quando o valor desse índice é menor que o percentil 40%. Já os índices que indicam instabilidade para valores positivos indicam maior probabilidade de chuva quando o seu valor está acima do percentil 60%.

O índice da distância média entre a temperatura do ar e a temperatura da parcela nos níveis entre 700 e 300 hPa indica alta variabilidade no grupo de dias secos. Este índice indica instabilidade para valores negativos. 75% dos dias secos se concentram acima do percentil 40%, o que indica instabilidade de moderada a estável [Figura 24 (a)]. Dias situados nas faixas de valores com instabilidade alta ou muito instável (associadas com o percentil 40% e 20%, respectivamente) podem estar associadas a dias com atuação de sistemas meteorológicos nas proximidades de Petrolina e, portanto, com condições favoráveis a chuvas.

Para os dias de véspera de chuva “v”, a maioria dos valores do DMAP são inferiores a 0,5 °C (percentil 40%), concentrando dias com instabilidade alta ou muito alta. Já aqueles dias concentrados nas faixas de valores que indicam estabilidade podem, de um lado, estar associados a situações meteorológicas em que o índice não apresentou sensibilidade à aproximação do sistema meteorológico; ou, de outro lado, é possível que o mecanismo de formação do sistema meteorológico não tenha sido associado à energia obtida por aquecimento nos níveis baixos, energia esta necessária à convecção [Figura 24 (a)]. O *boxplot* dos dias de chuva “c” indica uma maior dispersão em relação aos dias de véspera dos dias de chuva. Neste grupo, aproximadamente 75% dos dias estão concentrados abaixo de 1,5 — isto é, abaixo do percentil 60%. Os casos reunidos na faixa de valores que indica estabilidade ou instabilidade baixa podem indicar situações em que o índice não foi sensível à aproximação do sistema meteorológico; ou, alternativamente, podem indicar situações em que o mecanismo de formação do sistema meteorológico não esteve associado à energia obtida por aquecimento nos níveis baixos, a qual é necessária à convecção [Figura 24 (a)]. O conjunto de dias precedentes

às chuvas, representados pela sigla “pc”, tem uma variação mais próxima à dos dias secos. Aproximadamente 80% dos dias possuem valores associados à estabilidade, instabilidade moderada ou instabilidade baixa.

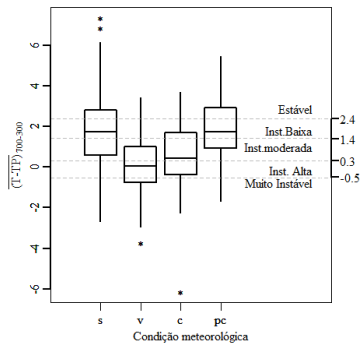
O IC (Figura 11 b), que indica instabilidade para valores positivos, possui alta variabilidade no grupo de dias secos, apresentando a maior parte dos dias com indicativo de instabilidade de moderada a estável; e um percentual de 25% com instabilidade alta a muito alta [Figura 24 (b)]. Nos dias de véspera de chuva (v), o indicativo de instabilidade do IC é maior que nos demais grupos. A maioria dos dias está concentrada acima do percentil 60%, indicando instabilidade alta ou muito instável. Para os dias de chuva, o IC também indica a tendência dos dados se concentrarem nas faixas de valores com indicativo de instabilidade mais alta. Os dias precedentes de chuva (“pc”) possuem indicativo de instabilidade do IC um pouco mais alta que o grupo dos dias secos (“s”). Apesar disso, a maioria dos dados se concentram abaixo do percentil 60%, que concentram as faixas de valores com menor indicativo de instabilidade [Figura 24 (b)].

Os índices que indicam instabilidade para valores baixos — TMED [Figura 24 (d)], DPM [Figura 24 (g)], KO [Figura 24 (h)] e DTel [Figura 24 (j)] — possuem uma análise semelhante àquela do índice DMAP [Figura 24 (a)]. Os índices que indicam instabilidade para valores altos são: K [Figura 24 (c)], TNCL [Figura 24 (e)], PII [Figura 24 (f)] e SWEAT [Figura 24 (i)].

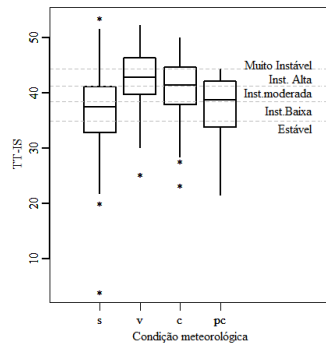
O índice DPM possui uma alta variabilidade em todos os conjuntos de dias. Ele também segue o comportamento do DMAP. O aspecto assimétrico da distribuição do DPM nas vésperas e nos dias de chuva, nos quais a mediana se encontra entre as ordens quantílicas que indicam instabilidade alta, é um indicativo de que esse índice pode ser um bom indicador das chuvas na área de Petrolina. O DTel tem um comportamento parecido com o do DPM. No entanto, a diferença entre dias secos e vésperas de chuvas, dias de chuva e precedentes de chuva é menor.

Dentre os índices que indicam instabilidade para valores altos, o que mais distingue as vésperas e dias de chuva dos dias secos é o IK, apresentando-se bastante concentrado nas faixas de maior instabilidade — acima do percentil 60%. O TNCL do nível de 925 hPa também apresenta um bom desempenho, assim como o SWEAT, que tem relação com o Total Totals e com o vento em 500 e 850 hPa.

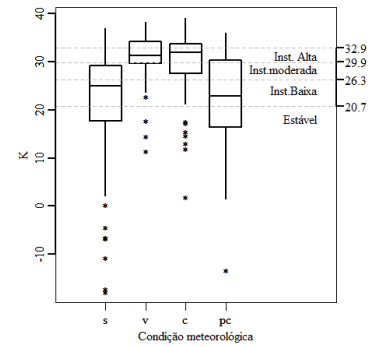
Figura 24 – Indicativo de instabilidade dos índices nos dias sem chuva (s), dias de véspera de chuva (v) dias de chuva (c) e dias precedentes à chuva (pc).



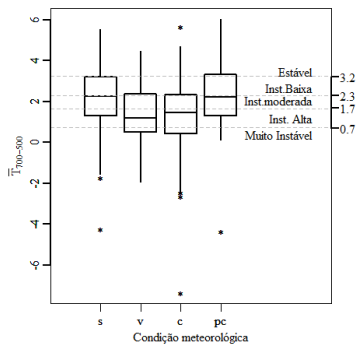
(a)



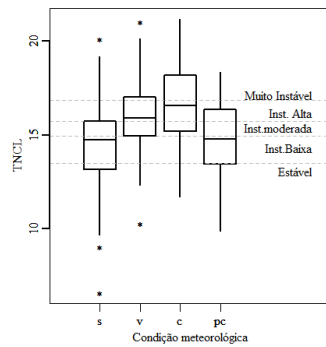
(b)



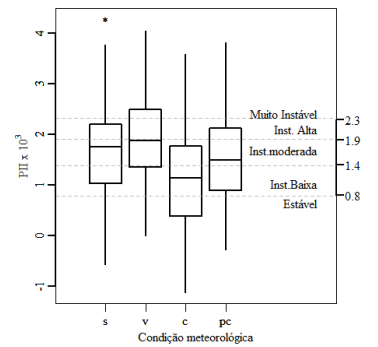
(c)



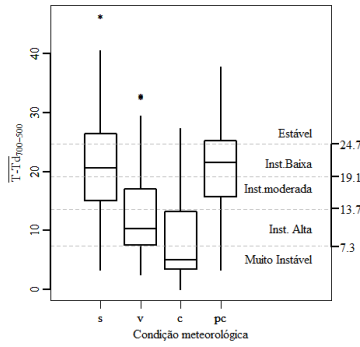
(d)



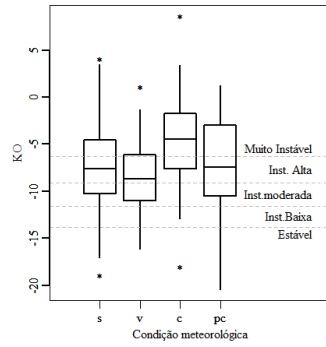
(e)



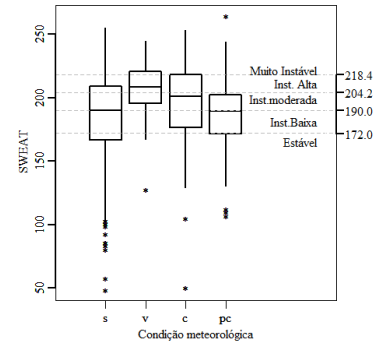
(f)



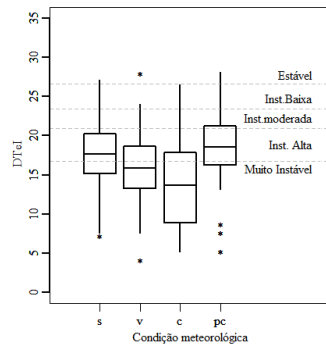
(g)



(h)



(i)



(j)

4.3 PRESSUPOSTOS DA ANÁLISE MULTIVARIADA

Dentre os principais pressupostos para a aplicação bem-sucedida das técnicas estatísticas multivariadas estão a normalidade multivariada dos dados e a ausência de *outliers*. Neste capítulo, é feita uma análise descritiva dos dados, com o objetivo de avaliar os requisitos de técnicas estatísticas multivariadas e as correções a problemas de normalidade.

4.3.1 Análise dos *outliers*

Os *outliers* foram excluídos do conjunto de dados, pois um dos requisitos da análise multivariada consiste na ausência de *outliers*. Outro fator que reforça a viabilidade da exclusão dos *outliers* é o de que os índices de estabilidade são bem representativos da condição de estabilidade atmosférica nos valores extremos. No entanto, optou-se por avaliar os *outliers* separadamente, uma vez que eles podem explicar algumas situações meteorológicas.

Os *outliers* do Índice DMAP que foram identificados estão associados à condição de estabilidade na região de Petrolina, o que também é observado para os índices IC, K, DPM e SWEAT. Já os índices TNCL e TMED apresentaram maioria dos dias com *outliers* associados às condições de estabilidade e de alta instabilidade, evidenciados pela ocorrência de chuvas na região de Petrolina (Tabela 1).

De um total de 408 dias, foram excluídos da análise 12,25% dos dias por conterem *outliers* em algum dos dez índices utilizados no estudo, o que corresponde a um total de 50 dias.

Tabela 1. Identificação dos *outliers* do conjunto meses de janeiro. Os *outliers* estão circulos e em grinto.

<i>DMAP</i>	IC	K	TMED	TNCL	PII	DPM	KO	SWEAT	DTeI	Data	chuva
2,69	35,15	22,10	-2,46	13,98	0,15	6,77	-0,13	149,80	8,69	10/01/2000	0,8
-0,89	44,74	33,20	-2,42	16,47	2,02	9,40	-6,26	175,20	15,29	24/01/2000	7,2
-1,30	42,15	16,30	2,68	18,36	3,52	35,87	-20,45	213,00	28,07	26/01/2000	0
2,32	23,31	14,70	3,54	13,49	1,74	24,51	-2,72	50,80	19,80	29/01/2000	11,4
1,02	30,03	11,10	-4,23	13,15	1,38	14,06	-2,89	86,00	12,97	03/01/2001	0
2,18	30,64	1,90	4,03	15,12	1,63	21,58	-10,81	201,00	19,39	05/01/2001	4,4
3,11	32,75	2,50	1,08	13,98	1,26	32,44	-10,53	177,60	16,30	06/01/2001	0
2,09	27,14	19,10	2,32	13,60	1,12	21,55	-0,45	103,20	15,11	10/01/2001	0
4,45	36,04	3,10	3,60	11,94	1,28	31,04	-7,68	174,20	15,24	16/01/2001	0
4,39	25,28	-10,70	-0,04	12,14	1,26	37,70	-5,93	111,80	16,79	18/01/2001	0
4,80	21,87	12,30	2,02	12,55	0,46	34,41	-3,09	137,40	16,50	25/01/2001	0
3,44	27,16	22,50	3,25	10,46	1,09	17,37	0,96	101,00	10,40	30/01/2001	0
3,09	27,95	-13,30	3,40	15,15	1,51	37,73	-12,85	201,40	23,13	10/01/2003	0
2,39	28,43	11,90	-0,29	12,91	-0,04	9,52	-2,17	105,20	17,79	25/01/2003	0,6
7,22	24,50	10,30	3,43	9,88	0,02	34,27	-3,16	172,00	19,18	06/01/2004	0
5,18	30,45	-17,70	2,83	11,29	1,03	40,60	-10,73	183,60	22,39	07/01/2004	0
3,98	25,39	18,30	2,89	12,10	0,61	30,41	-1,21	93,00	18,49	08/01/2004	0
8,05	28,81	11,10	4,38	9,06	-0,01	39,91	-5,54	166,00	20,49	06/01/2005	0
2,31	27,42	20,90	-0,28	12,88	1,84	17,76	-1,83	80,60	17,15	09/01/2005	0
1,16	39,34	30,90	0,98	15,73	0,24	3,75	-1,40	167,60	4,12	14/01/2005	0
4,05	32,39	-17,10	3,93	12,10	0,59	36,25	-8,12	144,00	17,69	04/01/2007	0
5,09	20,07	15,70	2,53	10,08	0,35	25,72	0,20	123,40	11,51	07/01/2007	0
7,44	4,07	4,30	4,64	9,68	-0,39	38,88	3,44	49,00	19,58	19/01/2007	0
0,69	24,50	16,30	-1,00	11,86	1,69	15,72	0,11	58,00	15,83	22/01/2007	0
3,83	32,36	-6,50	2,43	13,78	1,30	46,39	-12,40	223,60	22,32	03/01/2008	0
5,24	26,43	21,70	0,43	6,62	0,41	20,28	1,18	86,00	10,73	07/01/2008	0
2,15	38,45	1,50	2,29	14,16	1,22	16,56	-11,24	178,40	19,36	23/01/2008	0
0,82	37,63	-6,70	1,66	16,10	2,12	23,89	-15,48	201,20	24,22	24/01/2008	0
1,34	28,20	16,70	1,82	14,95	2,10	20,65	-3,83	103,80	21,70	15/01/2009	0
1,00	41,60	32,50	-2,62	16,33	0,33	4,29	-0,65	177,60	12,60	22/01/2009	35
-1,92	39,84	34,40	-2,25	18,68	2,29	5,35	-3,50	167,80	13,91	11/01/2011	11,6
1,31	32,72	-6,70	2,45	15,54	1,76	26,03	-10,73	174,00	24,79	15/01/2011	0
1,78	36,55	2,10	1,59	13,52	1,80	30,19	-12,04	170,20	20,56	18/01/2011	0
-0,46	36,41	-4,30	5,53	14,35	2,53	39,67	-14,16	100,80	24,44	19/01/2011	0
4,28	21,54	18,80	-4,34	12,73	-0,25	9,76	1,22	107,20	22,91	27/01/2011	0
2,44	27,50	9,50	2,91	14,37	1,49	23,10	-3,95	109,80	22,61	06/01/2012	0
3,30	27,79	25,80	1,65	13,52	1,28	17,65	0,09	99,20	17,94	28/01/2012	0
6,49	25,28	24,90	2,66	10,15	-0,17	29,07	4,11	118,20	18,11	29/01/2012	0
1,44	38,28	0,30	0,83	13,78	1,41	26,41	-11,01	192,20	18,48	05/01/2014	0
-0,15	37,57	2,20	0,53	15,06	1,11	23,17	-10,77	200,20	22,09	07/01/2014	0
-0,20	44,43	34,80	6,03	17,39	2,97	24,48	-10,41	264,40	23,31	12/01/2014	0
1,91	26,46	17,70	3,60	15,35	2,26	35,41	-6,30	107,00	24,66	22/01/2014	0
1,73	23,87	17,70	1,37	14,76	2,14	32,21	-3,30	83,40	19,66	23/01/2014	0
2,92	21,87	20,30	4,20	14,19	0,81	16,75	0,64	138,40	12,32	09/01/2015	0
2,96	22,51	14,10	2,22	14,18	1,34	26,34	-2,61	132,40	19,74	14/01/2015	0
-3,52	50,81	23,90	0,49	18,58	4,29	19,22	-18,93	239,00	23,41	30/01/2015	0
-3,25	42,74	36,20	3,80	21,04	2,79	4,84	-10,89	244,60	20,19	16/01/2016	0
4,46	28,70	25,00	1,98	16,64	-1,13	1,55	8,60	161,60	23,18	17/01/2016	5,4
-2,20	45,52	37,40	2,03	20,58	1,14	0,46	-5,02	227,80	8,56	24/01/2016	1,8
-2,78	41,77	34,30	3,03	21,17	1,96	3,43	-6,40	242,40	13,44	25/01/2016	1,8

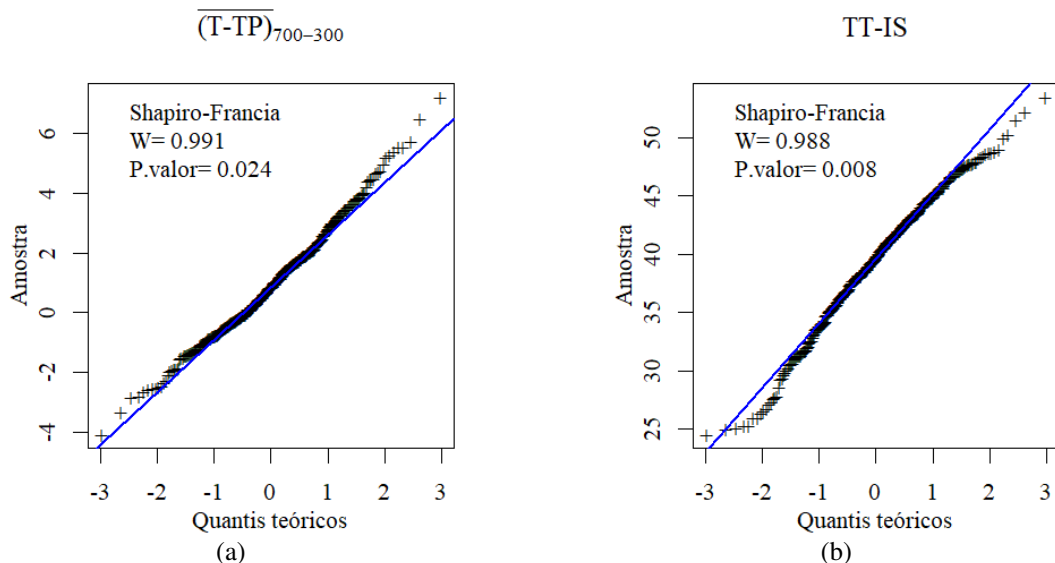
(Conclusão)

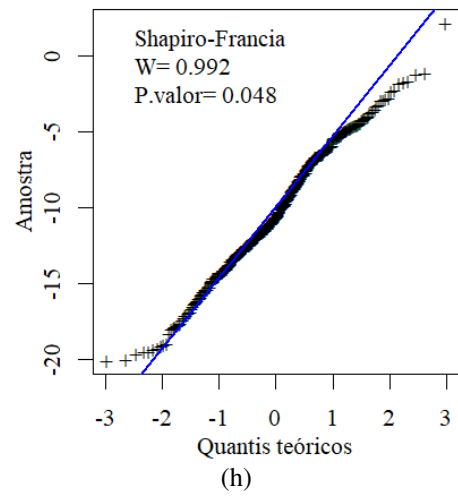
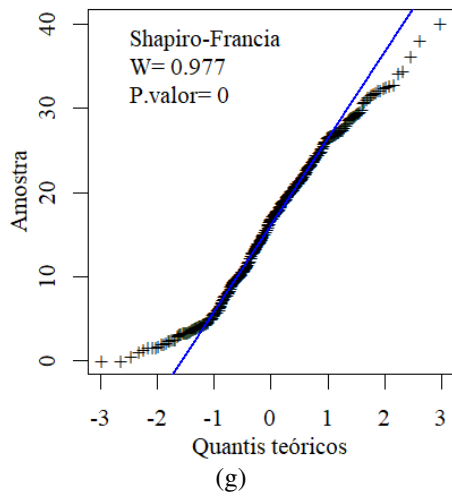
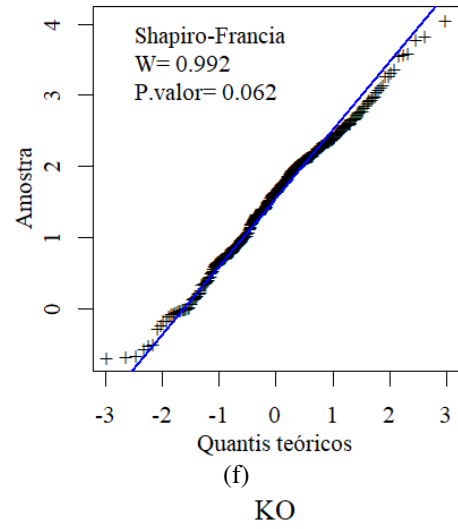
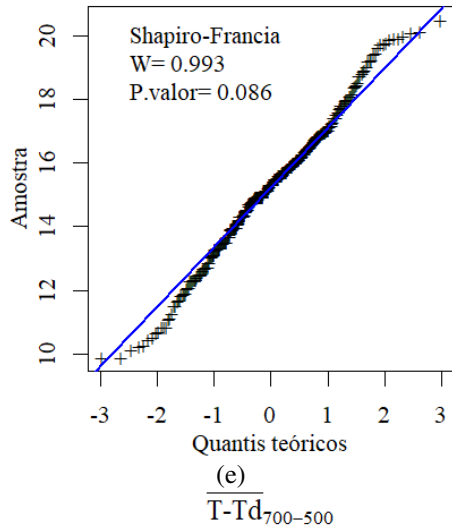
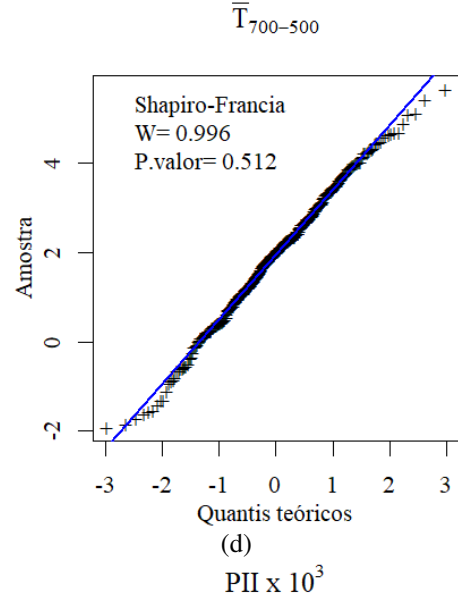
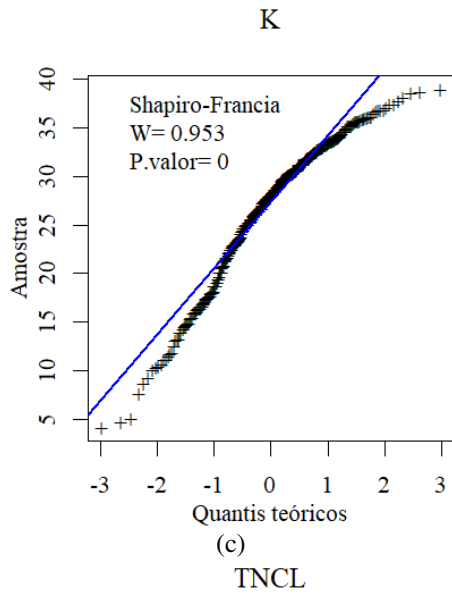
4.3.2 Verificação da normalidade dos índices de estabilidade

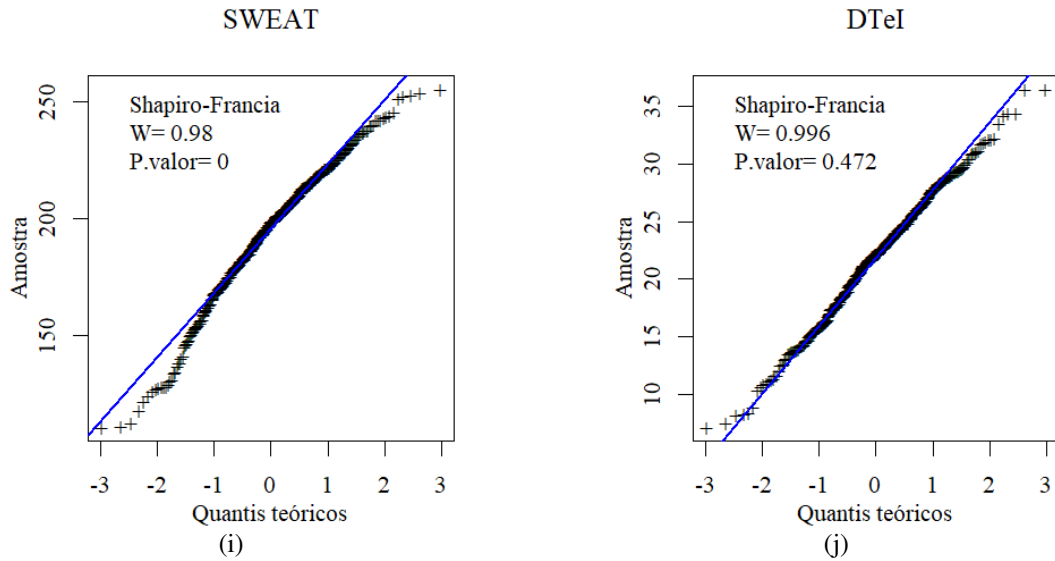
No presente estudo, o diagnóstico de normalidade dos dados foi feito utilizando um diagrama de probabilidade normal: o Q-Q plot. Também aplicou-se o teste de aderência de Shapiro-Francia nas distribuições dos dados, o qual testa a hipótese nula de normalidade dos dados.

Na Figura 11 de “a” a “j”, são apresentados os gráficos de probabilidade normal e o teste de Shapiro–Francia para os índices utilizados neste estudo. A linha azul representa a distribuição normal e os marcadores em preto representam a distribuição dos índices. A análise do QQ-plot juntamente com o teste Shapiro-Francia permite a identificação de seis índices que não possuem distribuição normal: DMAP [Figura 25 (a)], IC [Figura 25 (b)], K [Figura 25 (c)], DPM [Figura 25 (g)], SWEAT [Figura 25 (i)] e DTel [Figura 25 (j)]. Isto deve-se aos maiores desvios em relação à normal e ao p-valor do teste ser altamente significativo, o que indica evidências suficientes para rejeitar a hipótese nula de normalidade dos dados. A análise gráfica indica que os índices TMED [Figura 25 (d)], TNCL [Figura 25 (e)], PII [Figura 25 (f)] e KO [Figura 25 (g)] são os que possuem os menores desvios em relação à distribuição normal. Já o teste Shapiro-Francia recomenda a aceitação da hipótese de normalidade desses índices, uma vez que não há evidências suficiente para rejeitar essa hipótese a um nível de significância $\alpha=0,05$.

Figura 25 – Gráficos de probabilidade normal (Q-Q plot) dos índices não transformados. A linha reta em azul representa a distribuição normal; e os marcadores em preto indicam como os índices em estudo estão distribuídos.







4.3.3 Normalização pela transformação de Yeo-Johnson

Por meio da análise dos gráficos qq-plot dos dados, foi possível identificar índices que se desviam da distribuição normal. Conforme Hair *et al.* (2009), a falta de normalidade pode ocasionar sérios efeitos em pequenas amostras (com menos de 50 casos), mas o impacto diminui efetivamente quando a amostra atinge 200 casos ou mais. Segundo os autores, a maioria dos casos de heteroscedasticidade (não homogeneidade de variância) resultam da falta de normalidade em uma ou mais variáveis. Dessa forma, a depender do tamanho da amostra, pode não ser necessário corrigir a normalidade, mas, por outro lado, pode ser necessário para homogeneizar a variância. Há vários métodos de transformações para obter a normalidade e homogeneidade das variâncias. Neste estudo, aplicou-se a transformação de Yeo-Johnson.

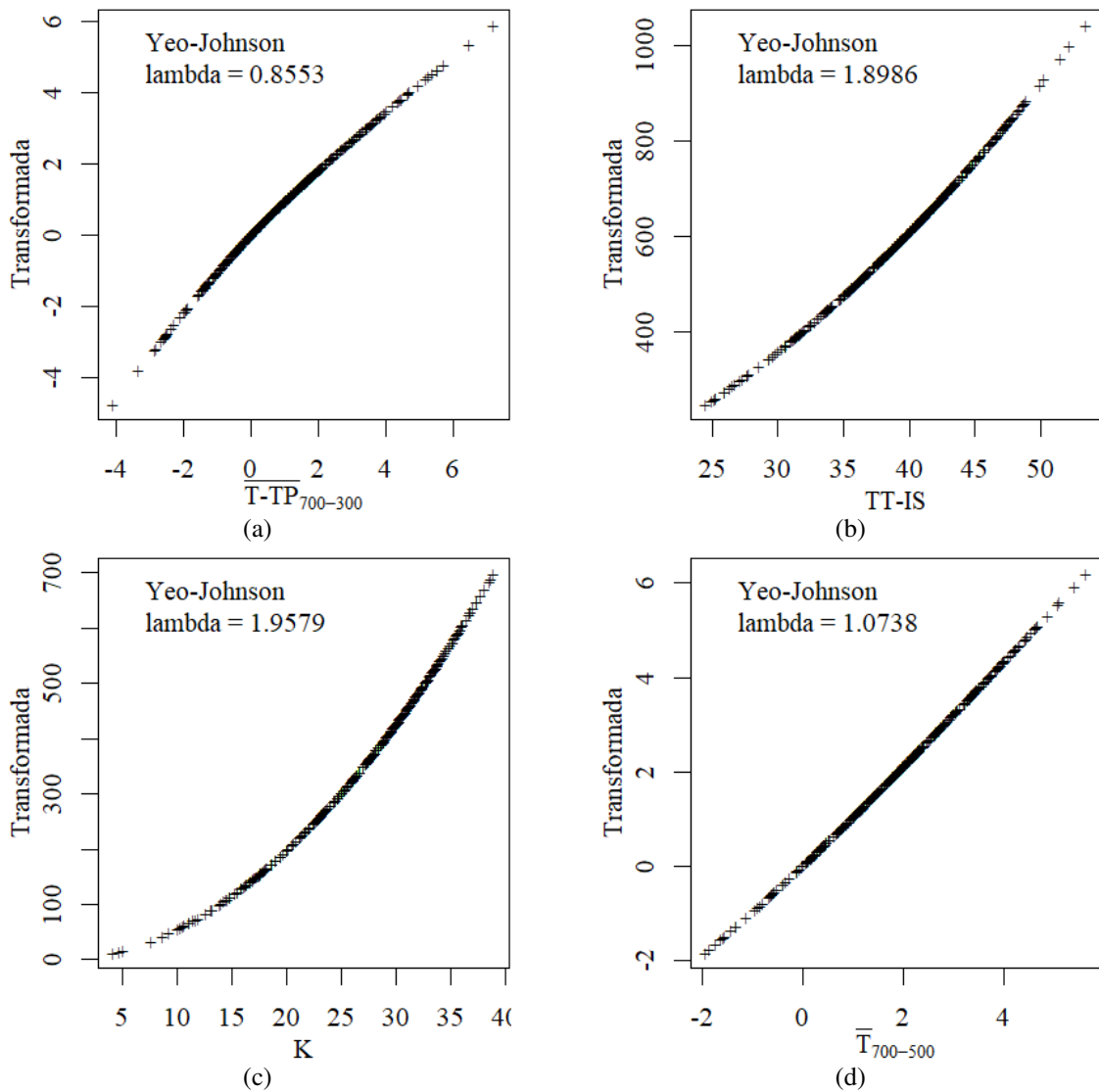
A relação entre as variáveis originais e suas transformadas é apresentada na Figura 26. A análise gráfica da transformação de Yeo-Johnson, mostra que a interpretação física dos índices não é comprometida pela transformação. Pelo contrário, a aplicação desta transformação nos índices de estabilidade representa um efeito benéfico para a interpretação, já que os valores mais extremos sofrem uma influência mais significativa do que os valores centrais.

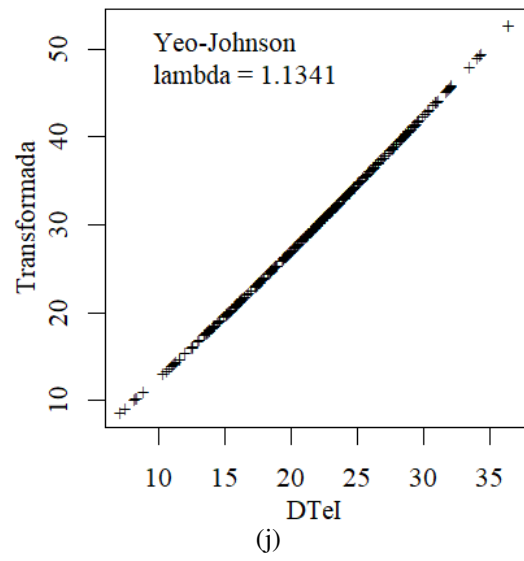
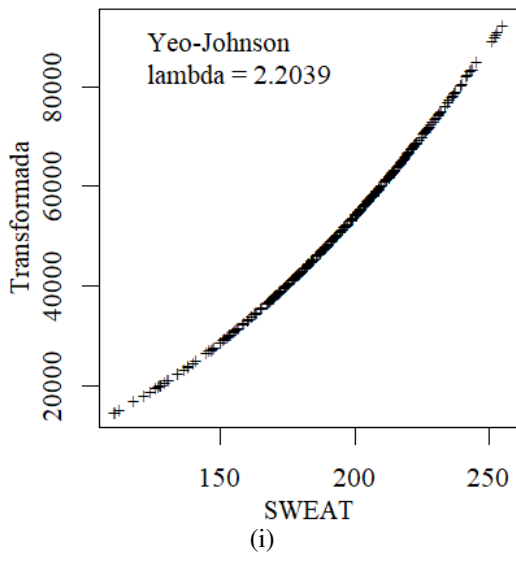
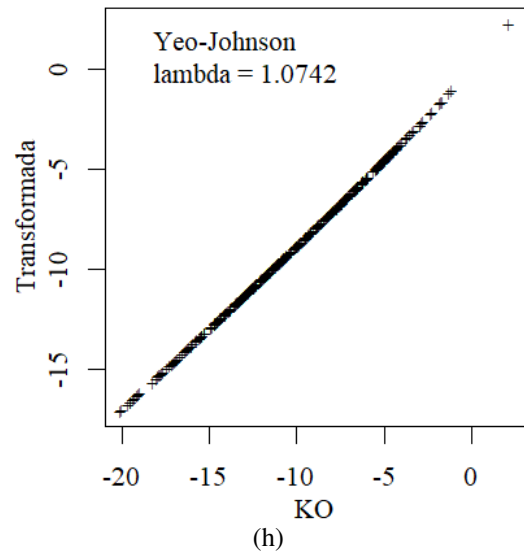
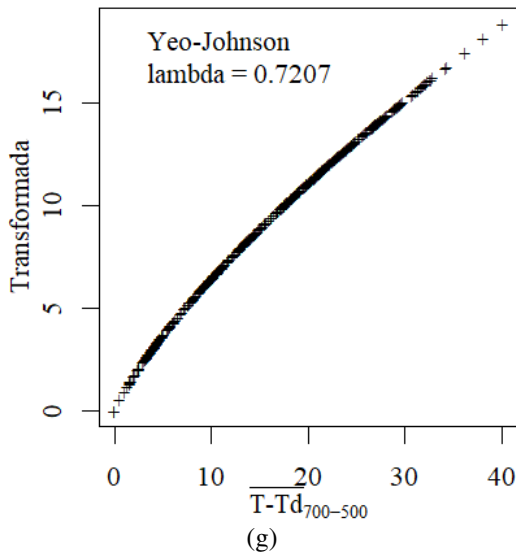
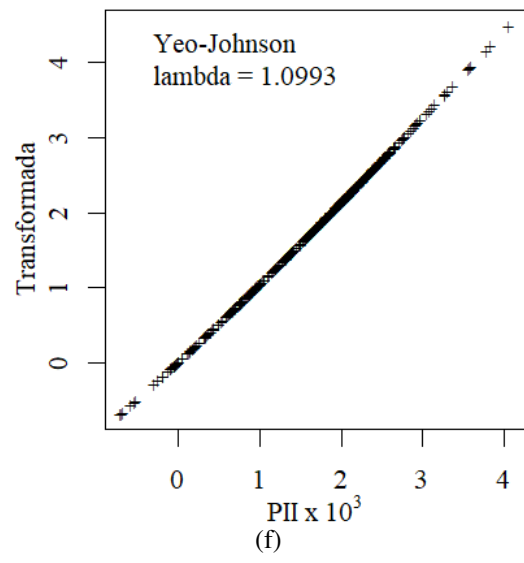
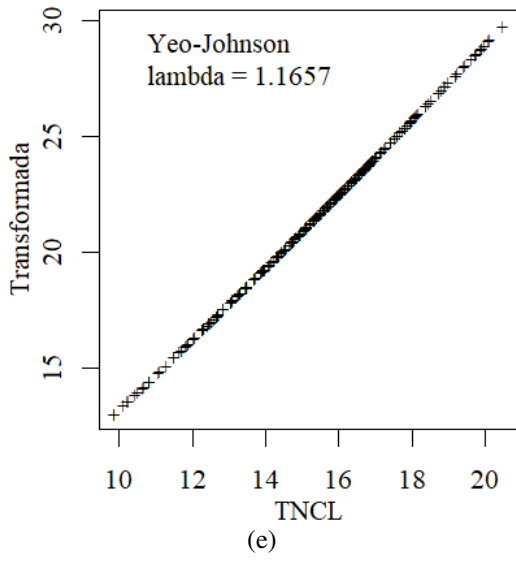
Os índices associados com o parâmetro $\lambda > 1$ da transformação de Yeo-Johnson têm os valores superiores aumentados mais significativamente do que os valores inferiores. Os índices que sofreram esse tipo de influência foram: IC [Figura 26 (b)], IK [Figura 26 (c)], SWEAT [Figura 26 (i)] e DTel [Figura 26 (j)].

Os índices associados com o parâmetro $\lambda < 1$ têm os valores inferiores diminuídos mais significativamente pela transformação do que os valores superiores. Os índices que sofreram esse tipo de influência foram DMAP [Figura 26 (a)] e DPM [Figura 26 (g)].

Quando $\lambda = 1$, a transformação não altera o valor das variáveis. Os índices TMED [Figura 26 (d)], TNCL [Figura 26 (e)], PII [Figura 26 (f)], KO [Figura 26 (h)] não sofreram alterações já que $\lambda \cong 1$.

Figura 26 – Relação entre os índices originais e os índices transformados





4.3.4 Índices transformados

4.3.4.1 Equações dos índices transformados

A vantagem de se utilizar a transformação de Yeo-Johnson é que esta transformação está definida para todos os números reais. Assim, de posse do conjunto de índices de estabilidade para os meses de janeiro, obteve-se o parâmetro lambda da transformação de Yeo-Johnson que melhor aproxima os índices de uma distribuição normal. Tal estimativa foi realizada com o suporte do pacote *bestNormalize* do software R. Uma vez estimado o parâmetro da transformação de Yeo-Johnson para cada índice, foi possível escrever as equações gerais que permitem obter índices transformados, mais próximos de uma distribuição normal. Tais equações são apresentadas a seguir:

1. Índice DMAP*

$$\psi(\text{DMAP}, \lambda) \begin{cases} \frac{(\text{DMAP} + 1)^{0,8553} - 1}{0,8553}, & \text{para } \lambda \neq 0 \text{ e } x \geq 0 \\ -\frac{[1 - \text{DMAP}]^{2-0,8553} - 1}{0,8553}, & \text{para } \lambda \neq 2 \text{ e } x < 0 \end{cases} \quad (40)$$

2. IC*

$$\psi(\text{IC}, \lambda) \begin{cases} \frac{[\text{IC} + 1]^{1,8986} - 1}{1,8986}, & \text{para } \lambda \neq 0 \text{ e } x \geq 0 \\ -\frac{[1 - \text{IC}]^{2-1,8986} - 1}{1,8986}, & \text{para } \lambda \neq 2 \text{ e } x < 0 \end{cases} \quad (41)$$

3. IK*

$$\psi(\text{K}, \lambda) \begin{cases} \frac{[\text{K} + 1]^{1,9579} - 1}{1,9579}, & \text{para } \lambda \neq 0 \text{ e } x \geq 0 \\ -\frac{[1 - \text{K}]^{2-1,9579} - 1}{1,9579}, & \text{para } \lambda \neq 2 \text{ e } x < 0 \end{cases} \quad (42)$$

4. Índice TMED*

$$\psi(\text{TMED}, \lambda) \begin{cases} \frac{[\text{TMED} + 1]^{1,0738} - 1}{1,0738}, & \text{para } \lambda \neq 0 \text{ e } x \geq 0 \\ -\frac{[1 - \text{TMED}]^{2-1,0738} - 1}{1,0738}, & \text{para } \lambda \neq 2 \text{ e } x < 0 \end{cases} \quad (43)$$

5. Índice TNCL*

$$\psi(\text{TNCL}, \lambda) \begin{cases} \frac{[\text{TMED} + 1]^{1,1657} - 1}{1,1657}, & \text{para } \lambda \neq 0 \text{ e } x \geq 0 \\ -\frac{[1 - \text{TMED}]^{2-1,1657} - 1}{1,1657}, & \text{para } \lambda \neq 2 \text{ e } x < 0 \end{cases} \quad (44)$$

6. PII*

$$\psi(\text{PII}, \lambda) \begin{cases} \frac{[\text{PII} + 1]^{1,0993} - 1}{1,0993}, & \text{para } \lambda \neq 0 \text{ e } x \geq 0 \\ -\frac{[1 - \text{PII}]^{2-1,0993} - 1}{1,0993}, & \text{para } \lambda \neq 2 \text{ e } x < 0 \end{cases} \quad (45)$$

7. Índice DPM

$$\psi(\text{DPM}, \lambda) \begin{cases} \frac{[\text{DPM} + 1]^{0,7207} - 1}{0,7207}, & \text{para } \lambda \neq 0 \text{ e } x \geq 0 \\ -\frac{[1 - \text{DPM}]^{2-0,7345} - 1}{0,7207}, & \text{para } \lambda \neq 2 \text{ e } x < 0 \end{cases} \quad (46)$$

8. KO*

$$\psi(\text{KO}, \lambda) \begin{cases} \frac{[\text{KO} + 1]^{1,0742} - 1}{1,0742}, & \text{para } \lambda \neq 0 \text{ e } x \geq 0 \\ -\frac{[1 - \text{KO}]^{2-1,0742} - 1}{1,0742}, & \text{para } \lambda \neq 2 \text{ e } x < 0 \end{cases} = \quad (47)$$

9. SWEAT*

$$\psi(\text{SWEAT}, \lambda) \begin{cases} \frac{[\text{SWEAT} + 1]^{2,2039} - 1}{2,2039}, & \text{para } \lambda \neq 0 \text{ e } x \geq 0 \\ -\frac{[1 - \text{SWEAT}]^{2-2,2039} - 1}{5,2039}, & \text{para } \lambda \neq 2 \text{ e } x < 0 \end{cases} \quad (48)$$

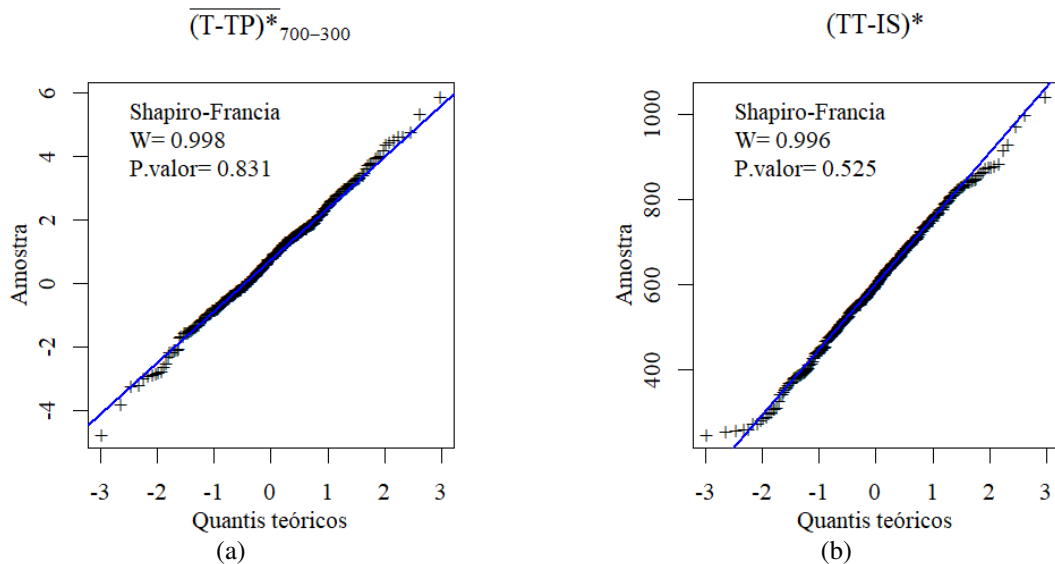
10. DTeI*

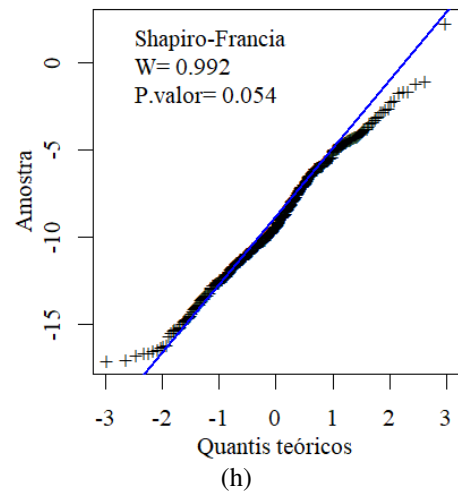
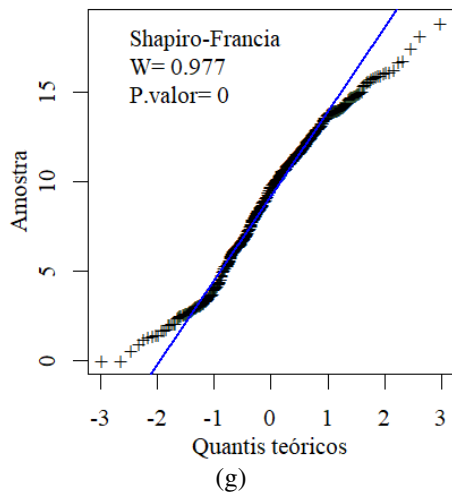
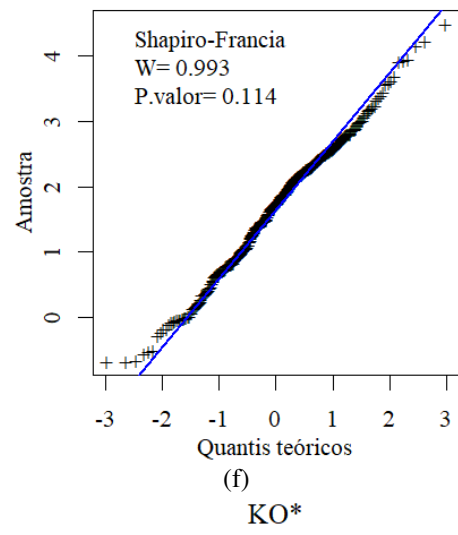
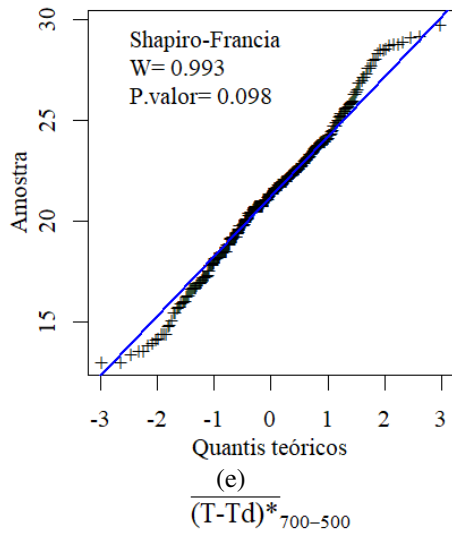
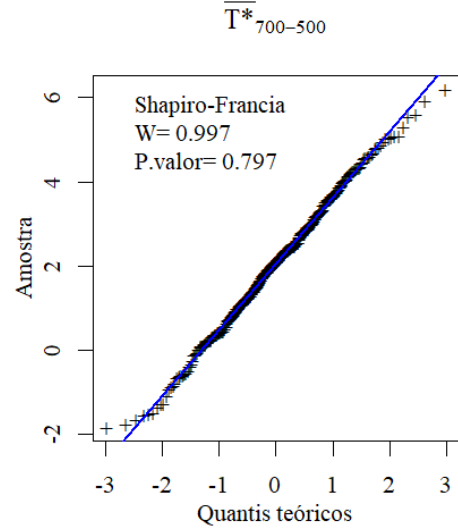
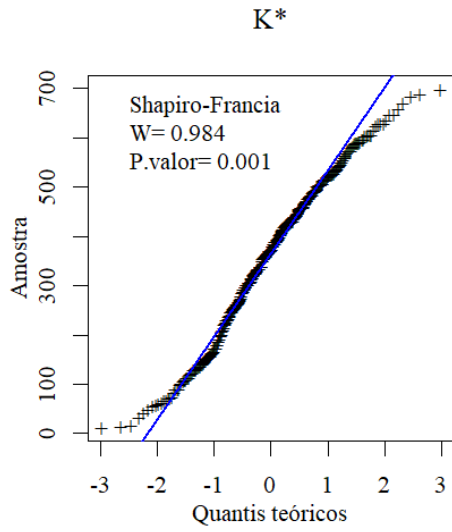
$$\psi(\text{DTeI}, \lambda) \begin{cases} \frac{[\text{DTeI} + 1]^{1,1341} - 1}{1,1341}, & \text{para } \lambda \neq 0 \text{ e } x \geq 0 \\ -\frac{[1 - \text{DTeI}]^{2-1,1341} - 1}{1,1341}, & \text{para } \lambda \neq 2 \text{ e } x < 0 \end{cases} \quad (49)$$

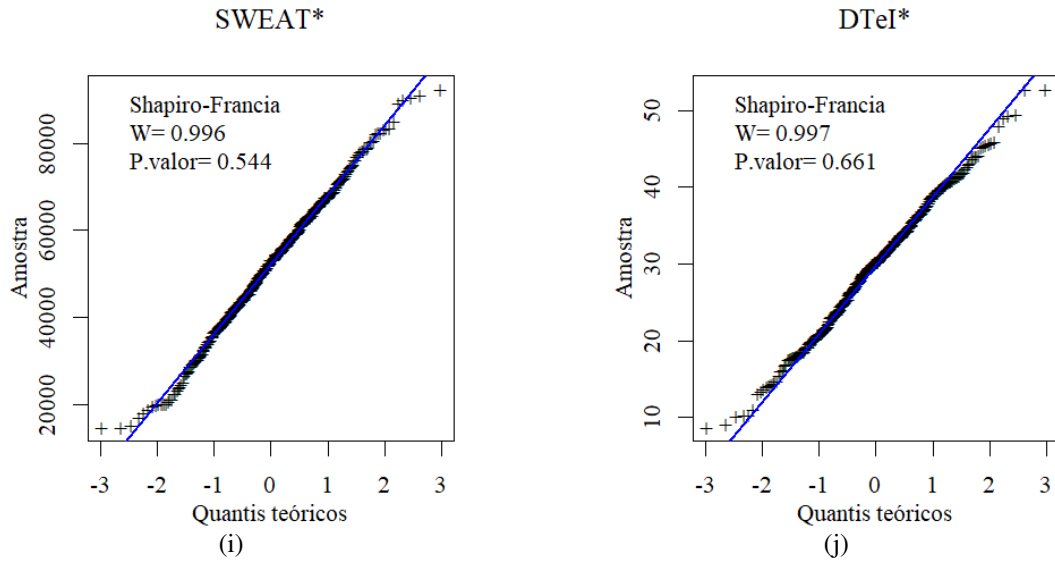
4.3.5 Estudo de normalidade dos índices transformados

Os gráficos de probabilidade normal da Figura 27 mostram a eficiência da transformação de Yeo-Johnson para a normalização dos índices utilizados neste estudo. O parâmetro estatístico usado para avaliar a normalidade dos índices transformados é o p-valor do teste de Shapiro-Francia, que varia de 0 a 1. A hipótese de normalidade dos dados é rejeitada quando o p-valor for menor que 0,05. Como está evidente na Figura 28, todos os índices apresentaram melhoras com o teste de aderência Shapiro-Francia. No entanto, os índices K e DPM não atingiram a normalidade.

Figura 27 – Gráficos de probabilidade normal (Q-Q plot) dos índices transformados. A linha reta em azul representa a distribuição normal; e os pontos em preto indicam a distribuição dos índices.







4.4 ANÁLISE MULTIVARIADA I: ANÁLISE DE COMPONENTES PRINCIPAIS E ANÁLISE DE AGRUPAMENTOS

Uma série de técnicas estatísticas multivariadas são utilizadas neste estudo. A ACP foi aplicada ao conjunto de índices transformados, com o objetivo de obter um número reduzido de variáveis, denominado fatores. A AA, que foi aplicada em seguida aos escores fatoriais por meio do Método Aglomerativo Hierárquico de Ward, tem como objetivo obter uma variável categórica formada pelos grupos de dia reunidos de acordo com a dissimilaridade. Em seguida, os índices transformados relacionados com cada fator foram organizados conforme o agrupamento obtido na AA, permitindo avaliar a sua relação com a ocorrência de chuva na área de Petrolina. Por fim, a AD é aplicada com o objetivo de identificar as variáveis com maior poder de discriminação, bem como avaliar a qualidade do agrupamento obtidos com a AA.

4.4.1 Análise de componentes principais (ACP) dos índices transformados

O conjunto de índices transformados foi submetido à ACP. Para isto, verificou-se primeiramente a adequabilidade desse conjunto de dados à aplicação dessa técnica estatística multivariada. Os pressupostos da ACP foram verificados pelo teste Kaiser-Meyer-Olkin (KMO), que fornece uma medida da adequabilidade do conjunto de dados à aplicação da ACP. O KMO obteve um valor de 0.605, indicando um grau de adequabilidade razoável. Outro teste aplicado foi o de esfericidade de Bartlett, teste que avalia a hipótese nula de que as variáveis estão correlacionadas entre si. Caso a hipótese nula seja rejeitada, a aplicação da técnica de análise fatorial é viável. De acordo com o teste de esfericidade de Bartlett, o conjunto de dados

de meses de janeiro é adequado, já que o p.valor é altamente significativo, rejeitando a hipótese nula de que as variáveis estão altamente correlacionadas (Figura 28).

O número de CP retidas foi obtido de forma subjetiva, por meio da análise do gráfico *Scree Plot*. Por meio deste método, foi possível selecionar três fatores com o maior poder de explicação (Figura 28). As três CP juntas explicam 76.7% da variância total dos dados, onde a primeira componente explica 38,8%, a segunda componente explica 29,2 % e a terceira componente, 8,7% da variância total dos dados (Figura 29).

Figura 28 – Diagrama *Scree Plot*, determinação no número de componentes principais adequadas para os índices.

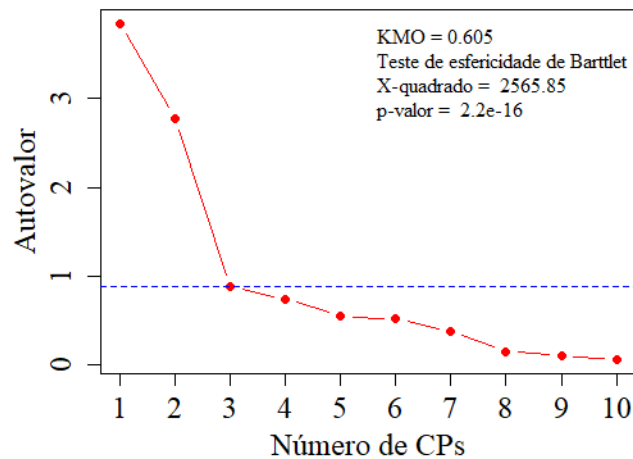
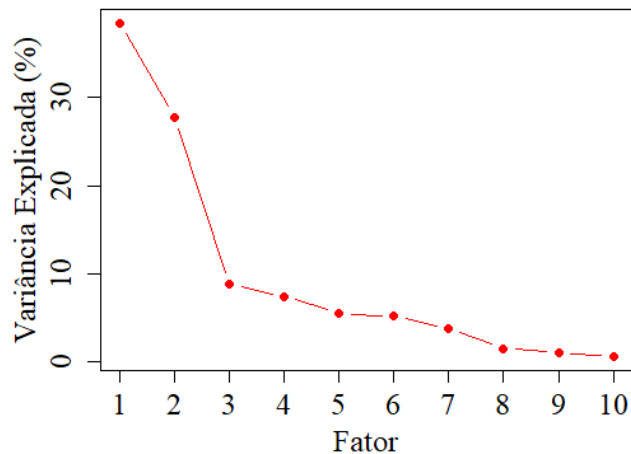


Figura 29 – Variância explicada pelas componentes principais retidas



A primeira componente apresenta forte correlação com os índices transformados TNCL*, DMAP*, IC*, K* DPM* e SW*. Estes índices possuem forte relação com o teor de umidade nos níveis baixos e na baixa média troposfera. A segunda componente está relacionada com os índices de instabilidade potencial, KO* e DTeI*, enquanto a terceira componente está

altamente correlacionada com a transformada da TMED. A temperatura do nível de condensação por levantamento tem a maior correlação com o primeiro fator; o KO*, com a segunda componente; e o TMED* com a terceira componente.

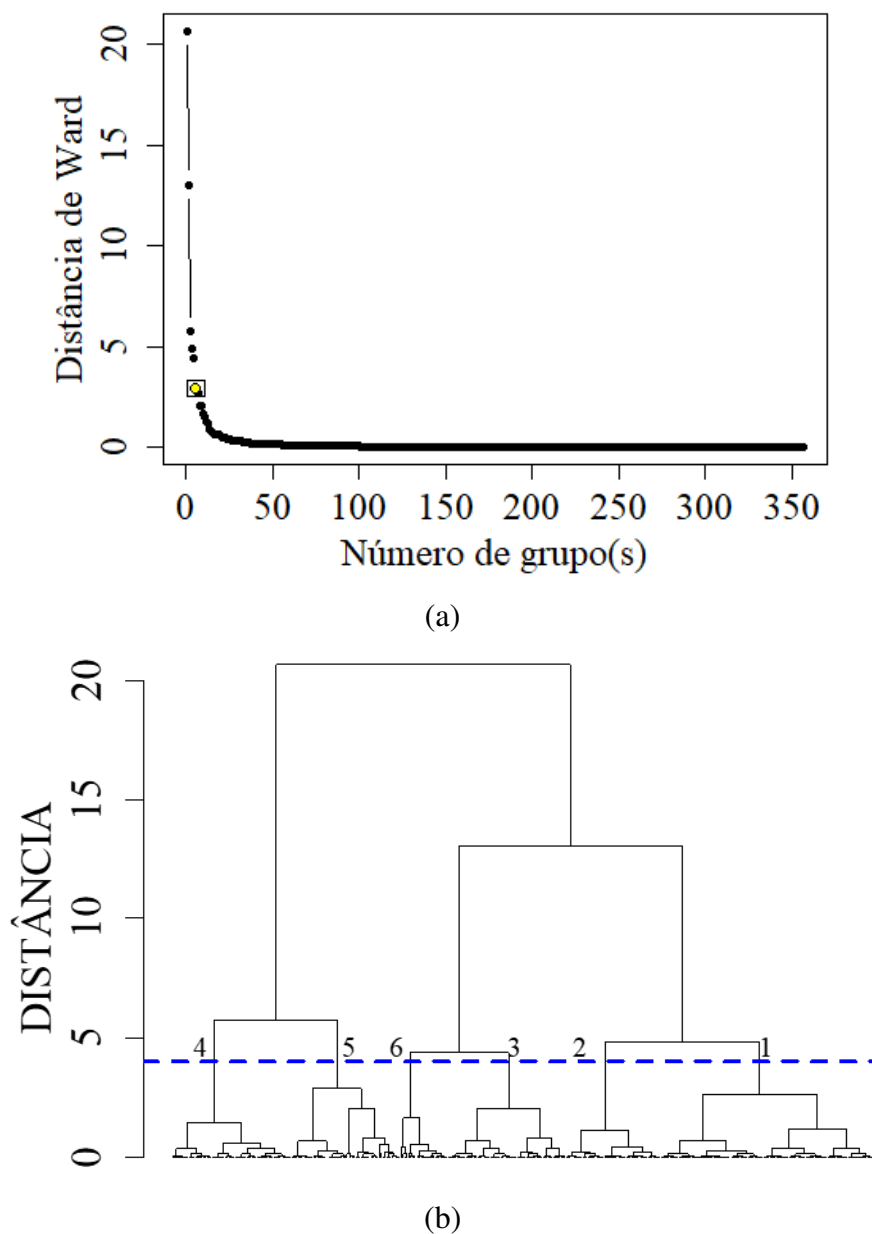
Tabela 2 – Matriz de estrutura obtida pela ACP

Índice Transformado	CP 1	CP 2	CP 3
DMAP*	-0,81	-0,32	0,34
IC*	0,79	0,12	-0,16
K*	0,75	-0,28	-0,13
<u>TMED*</u>	-0,16	0,11	<u>0,93</u>
TNCL*	0,82	0,14	-0,11
PII*	0,22	0,82	-0,05
DPM*	-0,66	0,55	0,29
KO*	-0,14	-0,9	-0,08
SW*	0,74	0,14	0,2
DTel*	-0,08	0,83	0,09

4.4.2 Análise de agrupamentos do primeiro fator

Por meio do método aglomerativo hierárquico de Ward, a AA foi aplicada aos escores do primeiro fator. O número de grupos foi determinado por meio da análise da curva de inércia [Figura 30 (a)]. Seis grupos foram suficientes para separar os dias de acordo com a dissimilaridade. Estes grupos podem ser visualizados no Dendrograma [Figura 30 (b)].

Figura 30 – (a) Curva de inércia do agrupamento do primeiro fator e (b) dendrograma com a identificação dos grupos

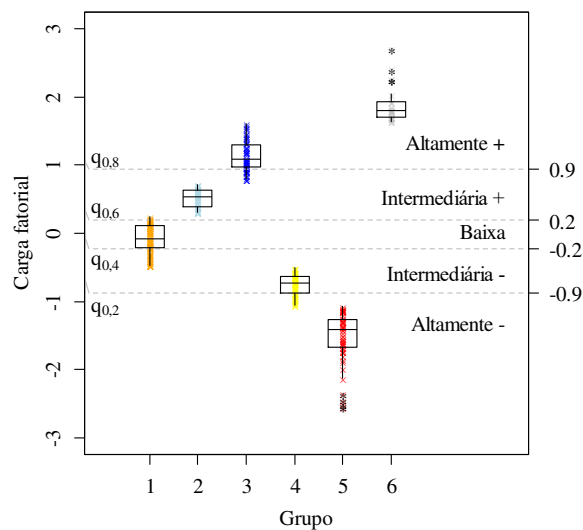


A Figura 31 apresenta o agrupamento de acordo com a magnitude das cargas fatoriais. O Método do Percentil também é aplicado para classificar os grupos por faixas de valores. Quando o grupo está situado entre os valores $-0,2$ e $0,2$, isto é, entre os percentis 40% e 60%, ele está associado a cargas fatoriais baixas da primeira CP. Esta faixa de valores concentra os dias com características meteorológicas que ocorrem com mais frequência nos meses de janeiro, porque reúne o maior número de casos. Quando o grupo está associado a cargas intermediárias positivas, ele fica situado entre a faixa de valores de $0,2$ a $0,9$ (entre os percentis 60% e 80%). Se o grupo estiver associado a cargas intermediárias negativas, ele estará situado na faixa de

valores entre -0,2 e -0,9 (entre os percentis 20% e 40%). Grupos associados a cargas altas da primeira CP — positivas ou negativas — são aqueles que ficam acima do percentil 0,8 e abaixo do percentil 0,2, respectivamente (Figura 31).

Assim, os dias reunidos no grupo 1 estão associados a cargas fatoriais baixas da primeira CP, já que pertencem à faixa de valores situadas entre o percentil 40% e 60% (-0,2 e 0,2). Alguns dias reunidos neste grupo possuem cargas fatoriais intermediárias negativas. Já os dias reunidos no grupo 2 apresentam cargas fatoriais intermediárias positivas (situadas entre o percentil 60% e 80%). Os dias concentrados no grupo 3 estão associados a cargas fatoriais altamente positivas; e o grupo 4, a dias com cargas fatoriais intermediárias negativas da primeira CP. Os grupos 5 e 6 são os que reúnem as cargas com magnitudes mais elevadas, estando o grupo 5 associado com cargas altamente negativas e o grupo 6 com cargas altamente positivas (Figura 31).

Figura 31 – Distribuição das cargas fatoriais por grupo do primeiro fator: as linhas horizontais representam os percentis 20%, 40%, 60% e 80% dos escores da primeira CP e seus valores são apresentados no eixo vertical à direita da figura.

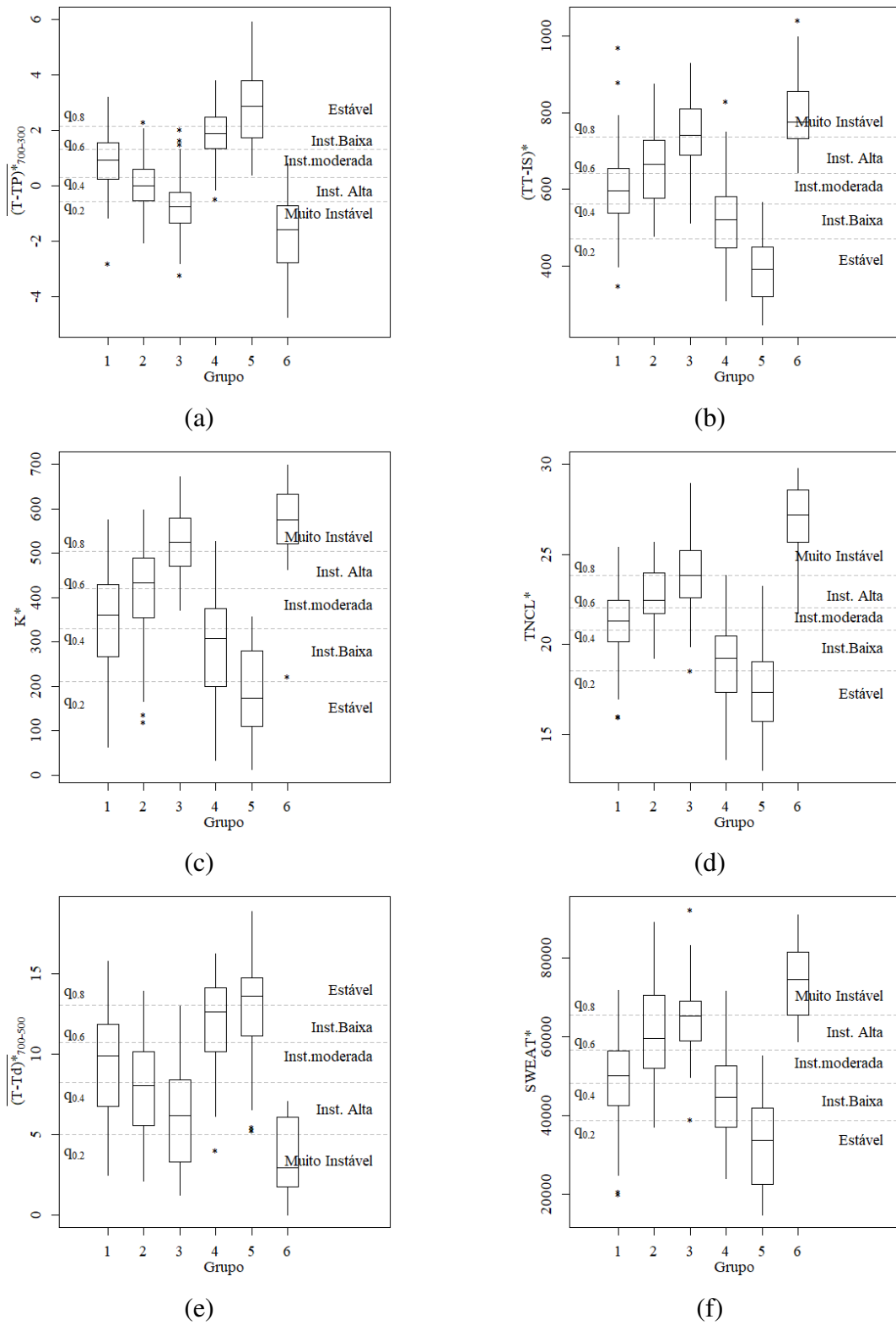


4.4.2.1 Relação dos índices com o primeiro fator.

Para ilustrar a relação das variáveis relacionadas com a primeira CP, os índices transformados foram organizados de acordo com o agrupamento dessa CP. Os grupos associados a cargas fatoriais positivas altas reúnem dias associados com instabilidade atmosférica alta, enquanto os grupos associados com cargas elevadas negativas reúnem uma quantidade maior de dias associados à condição de estabilidade.

O grupo 1, que está associado a cargas baixas do primeiro fator (Figura 31), reúne o maior número de dias em que os índices indicam instabilidade atmosférica moderada ou baixa (Figura 32). O grupo 2 possui indicativo de instabilidade mais alto; no entanto, este grupo contém uma quantidade entre 30% e 50% de dias concentrados na faixa de valores com menor instabilidade (Figura 32). Este grupo reúne dias com cargas intermediárias positivas e cargas baixas do primeiro fator (Figura 31). Já o grupo 3 reúne dias com instabilidade alta (Figura 32) e está associado a cargas fatoriais positivas altas do primeiro fator (Figura 31). O grupo 4, que está associado com cargas fatoriais negativas intermediárias e elevadas do primeiro fator, concentra dias com instabilidade baixa ou estável (Figura 32). Já o grupo 5, associado com cargas altamente negativas do primeiro fator (Figura 31), reúne dias associados com estabilidade, enquanto que o grupo 6 reúne dias associados ao indicativo de instabilidade muito alta dos índices (Figura 32).

Figura 32 – Variáveis transformadas relacionadas com a primeira CP, organizadas conforme o agrupamento dessa CP.

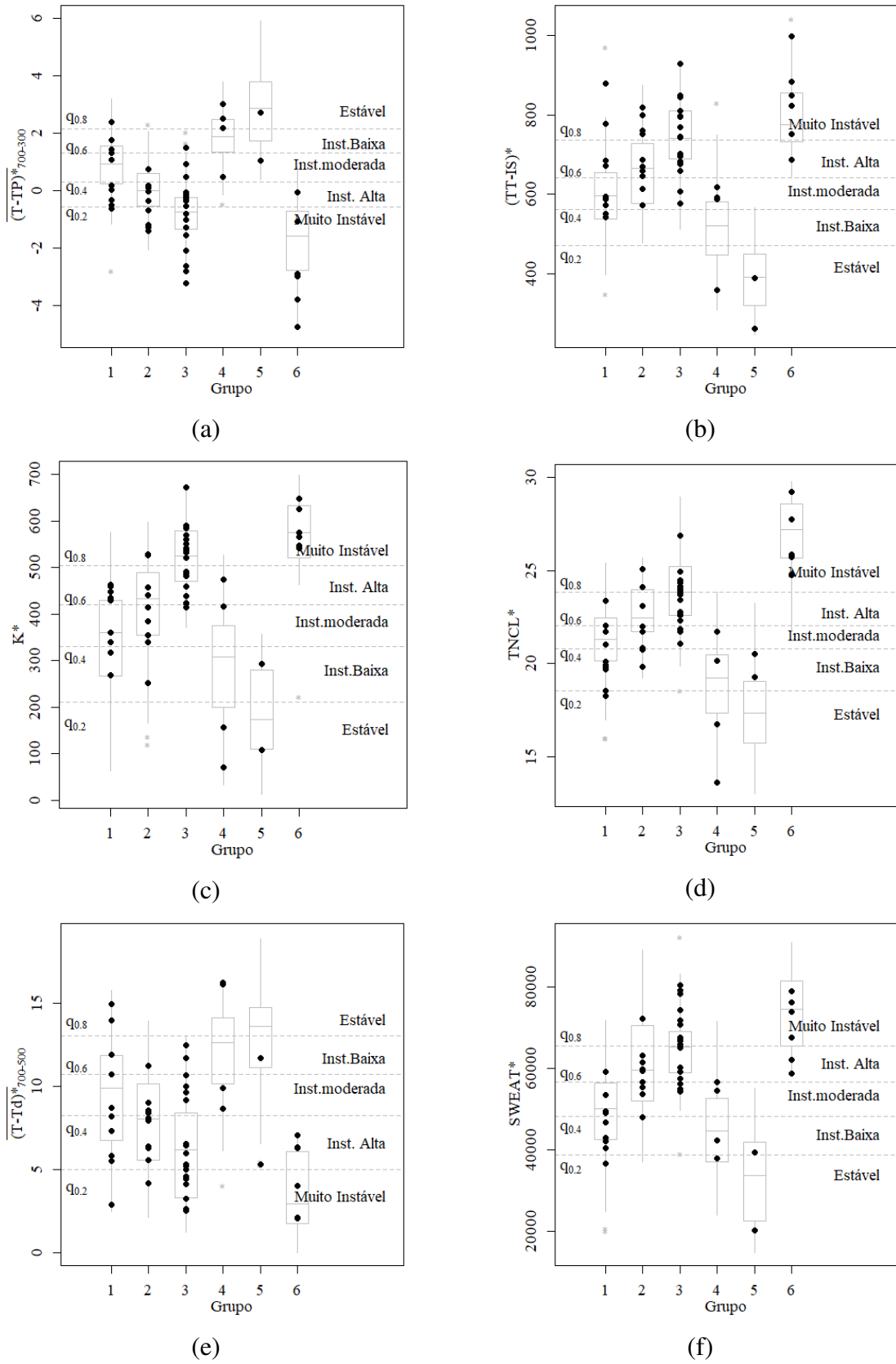


4.4.2.2 *Instabilidade dos índices relacionados com a primeira CP nos dias antecedentes à chuva.*

Dentre os dias pertencentes a cada grupo, foram identificados os dias antecedentes à ocorrência de chuva para verificar se os índices de estabilidade são capazes de indicar condições favoráveis à chuva na área de Petrolina. Os dias de véspera de ocorrência de chuva estão representados por pontos sobrepostos aos *boxplot* dos grupos identificados para cada índice.

Como está evidente, o limiar da maioria dos dias de véspera de chuva apresenta o indicativo de instabilidade acima do percentil 40% para aquelas variáveis que indicam instabilidade para valores negativos (DMAP* e DPM*); e acima do percentil 60% para os índices que indicam instabilidade atmosférica para valores altos (IC*, K*, TNCL* e SWEAT*). A maior parte dos dias antecedentes à ocorrência de chuva estão reunidos nos grupos 2 e 3, sendo que a maior parte se concentra abaixo do percentil 40% [Figura 33 (a – f)].

Figura 33 – Distribuição dos dias de véspera de chuva por grupo do primeiro fator.

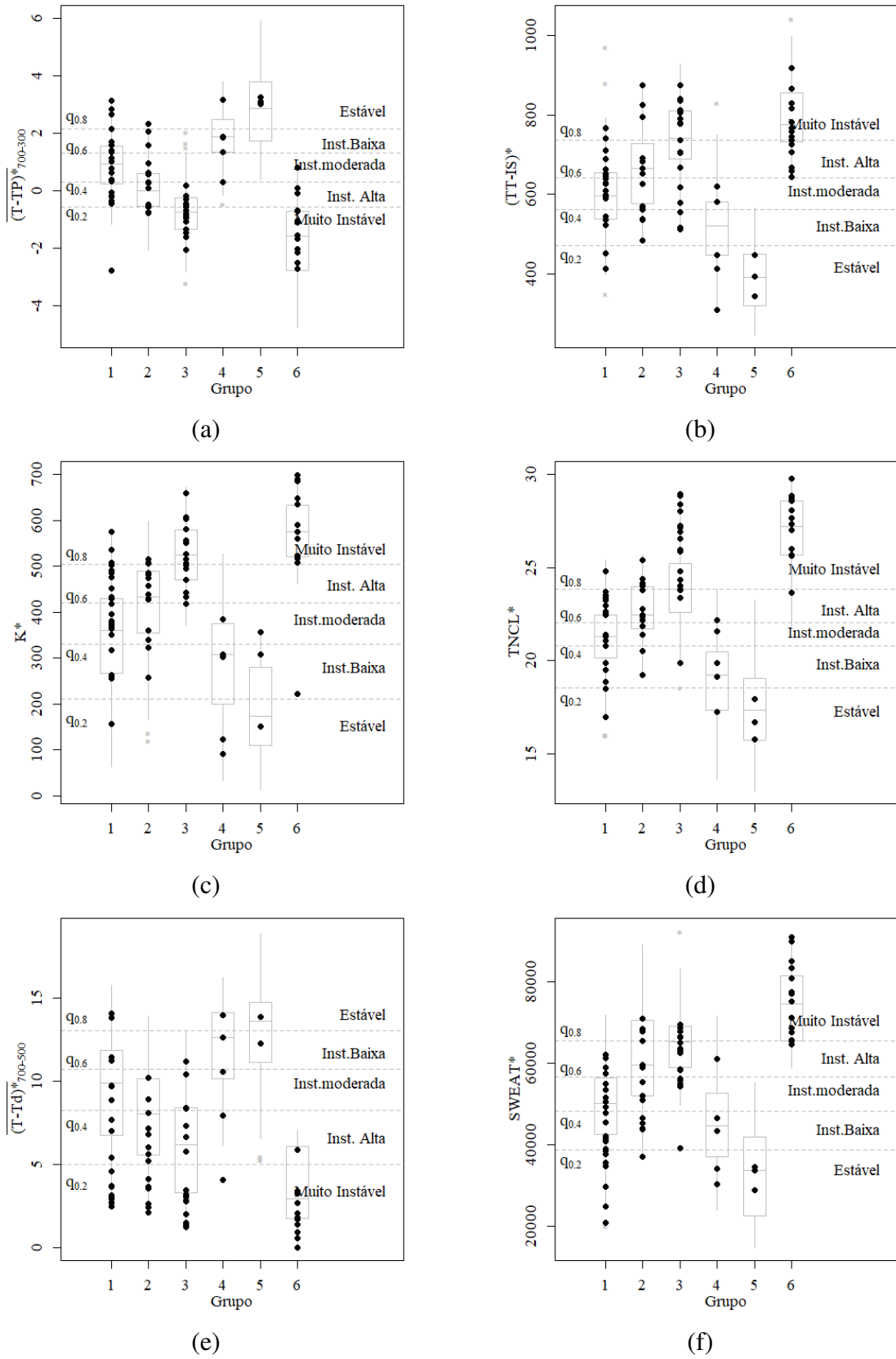


4.4.2.4 *Instabilidade dos índices relacionados com a primeira CP nos dias de chuva.*

A maior parte dos dias de chuva estão reunidos nos grupos 1, 2, 3 e 6. Os valores dos índices que indicam instabilidade para valores negativos (DMAP* e DPM*) tendem a se concentrar abaixo do percentil 40%. Já os índices que indicam instabilidade atmosférica para valores altos (K*, TT-IS*, TNCL* e SWEAT*) tendem a se concentrar acima do percentil 60%. Grupos associados a cargas fatoriais baixas ou altamente negativas (grupos 4 e 5) reúnem poucos de dias de chuva (Figura 34).

A análise apresentada acima permite verificar a relação entre as cargas fatoriais da primeira CP e o indicativo de instabilidade dos índices retidos nela. Cargas altas estão atribuídas à condição de alta instabilidade; cargas baixas estão associadas com condições de estabilidade moderada, sendo esta a condição de estabilidade mais frequente no mês de janeiro, uma vez que o grupo 1 retém o maior número de dias entre os demais grupos. Por fim, cargas altamente negativas estão associadas com condição de estabilidade dos índices relacionados com a primeira CP.

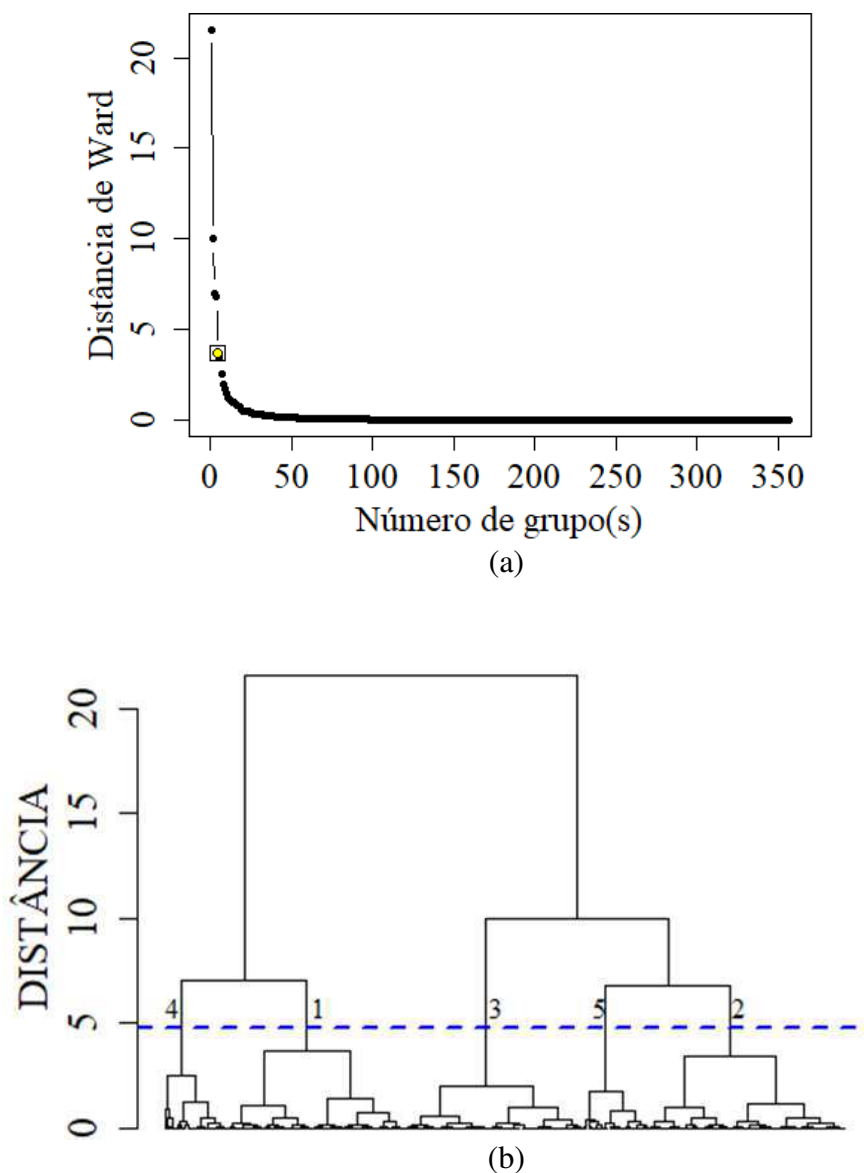
Figura 34 – Distribuição dos dias de chuva por grupo do primeiro fator.



4.4.3 Análise de agrupamentos do segundo fator

De acordo com a análise da curva de inércia para o agrupamento do segundo fator, foi possível obter cinco grupos de dias de acordo com a similaridade das cargas fatoriais. A curva de inércia auxilia na escolha do ponto de corte no Dendrograma para a obtenção dos grupos (Figura 35).

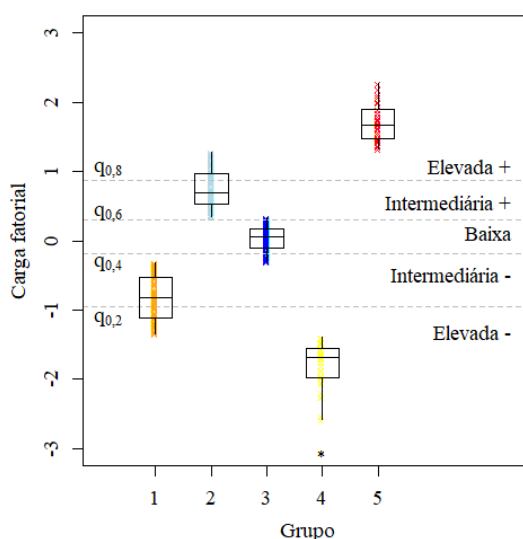
Figura 35 – (a) Curva de inércia e (b) dendrograma do agrupamento do segundo fator



A distribuição dos grupos de acordo com as cargas fatoriais é apresentada na Figura 36. No grupo 1, predominam cargas fatoriais intermediárias negativas da segunda CP,

intermediárias positivas no grupo 2; cargas fatoriais baixas no grupo 3; cargas altamente negativas no Grupo 4; e cargas altamente positivas no Grupo 5 (Figura 36).

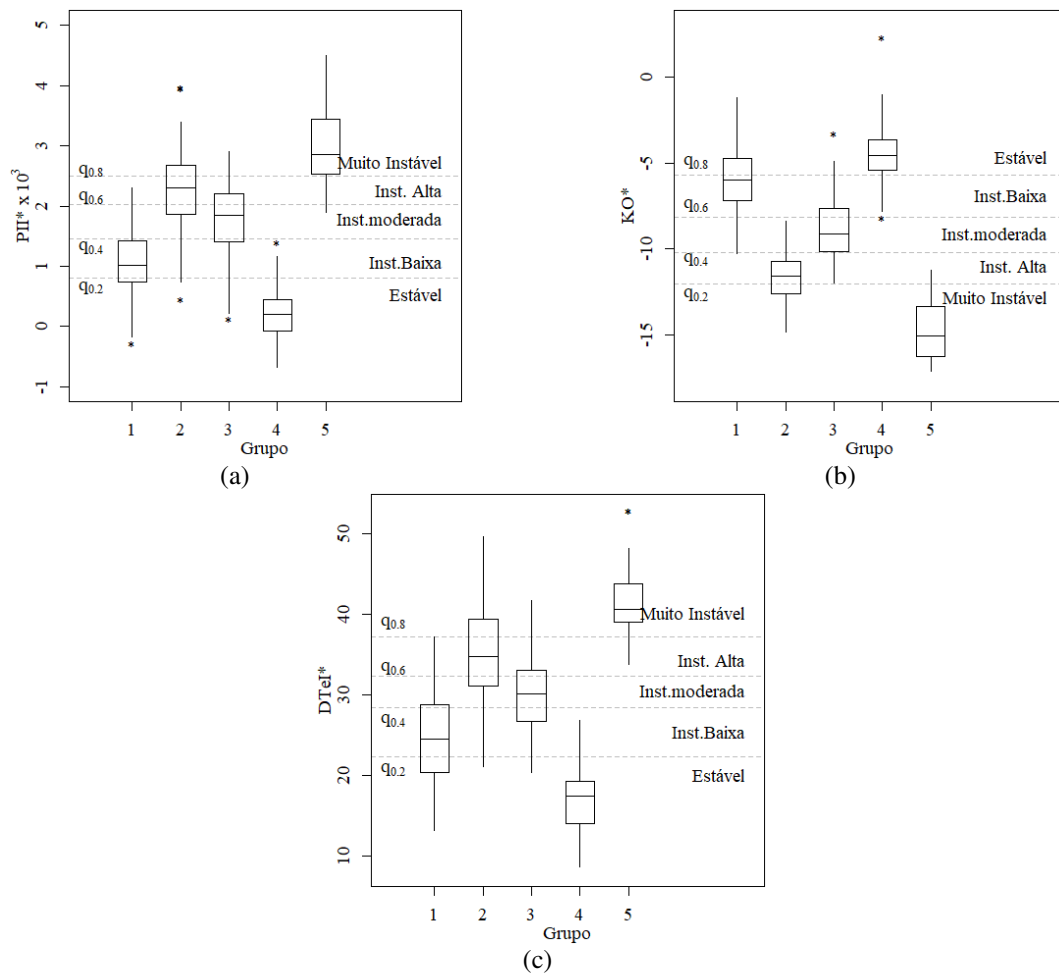
Figura 36 – Distribuição das cargas fatoriais por grupo do segundo fator: as linhas horizontais representam os percentis 20%, 40%, 60% e 80% dos escores da segunda CP e seus valores são apresentados no eixo vertical à esquerda da Figura.



4.4.3.1 Relação dos índices com o segundo fator

No grupo 1, que está associado a cargas intermediárias negativas do primeiro fator (Figura 36), predominam dias com indicativo de instabilidade baixo ou estável dos índices PII*, KO* e DTel* (Figura 37). No grupo 2, associado a escores intermediários positivos da segunda CP (Figura 36), predominam dias com indicativo de instabilidade alto ou muito instável de acordo com os índices PII*, KO* e DTel* (Figura 37). Já no grupo 3, que está associado com cargas fatoriais baixas da segunda CP (Figura 36), predominam dias com indicativo de instabilidade moderado a alto dos índices PII*, KO* e DTel* (Figura 37). O grupo 4 reúne dias associados à condição de estabilidade e o grupo 5, a dias muito instáveis de acordo com os índices PII*, KO* e DTel* (Figura 37).

Figura 37 – Variáveis transformadas relacionadas com a segunda CP, agrupadas conforme a AA desta CP.



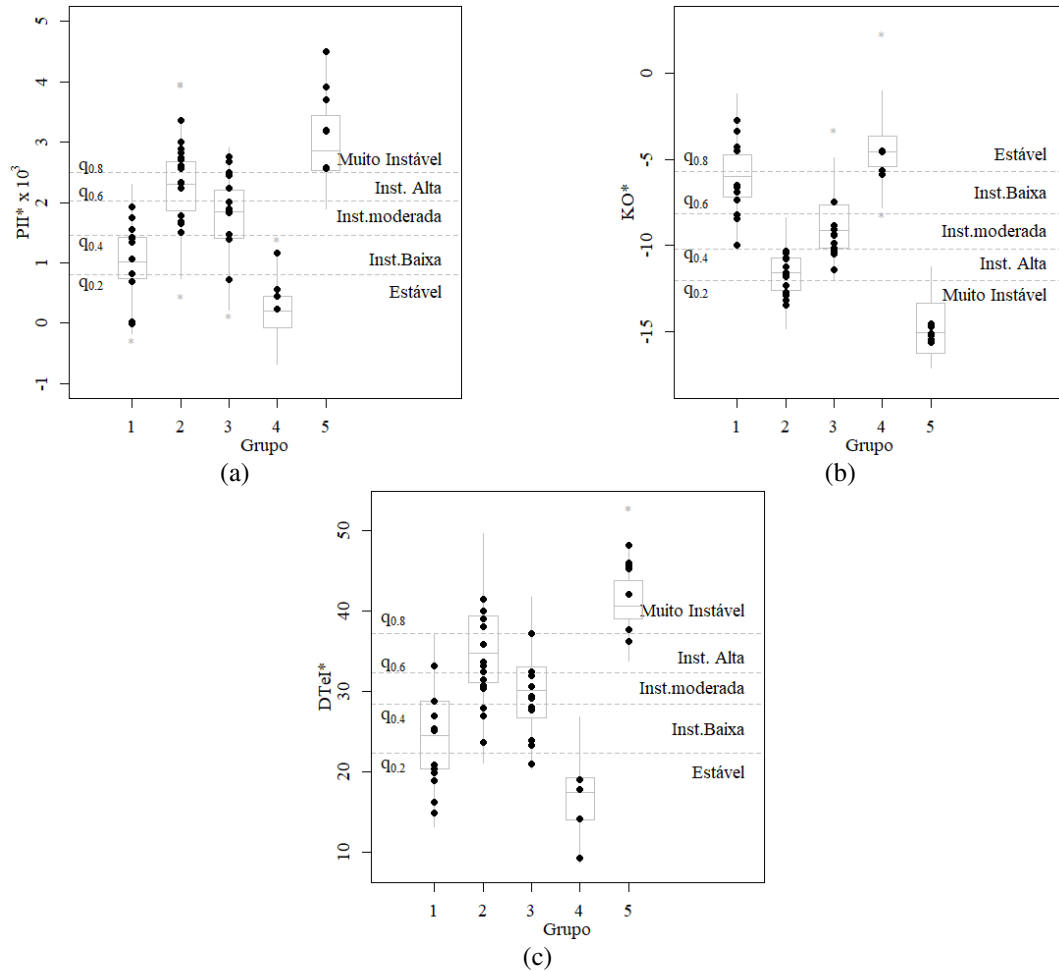
4.4.3.2 Instabilidade dos índices relacionados com a segunda CP nos dias de véspera de chuva.

Os dias de véspera de chuva pertencentes a cada grupo da segunda CP foram identificados de forma a verificar se os índices de estabilidade são capazes de indicar condições favoráveis à chuva na área de Petrolina (APÊNDICES C e D).

A maior parte dos dias de véspera de chuva estão reunidos nos grupos 1, 2, 3 e 5. No grupo 1, a maioria desses dias indicam instabilidade potencial moderada ou baixa dos índices PII*, KO* e DTel* (Figura 38). Nos grupos 2, 3, e 5, a maioria dos dias de véspera de chuva possuem valores acima do percentil 60% com relação aos índices PII*, KO* e DTel*, indicando instabilidade alta ou muito alta (Figura 38). Os dias reunidos no grupo 4 indicam estabilidade de acordo com os índices relacionados com a segunda CP.

A análise da Figura 38 indica que os índices relacionados com instabilidade potencial (PII^* e KO^* e $DTel^*$), mostraram-se menos sensíveis à identificação de condições favoráveis à chuva nos dias de véspera em relação aos índices relacionados com o primeiro fator.

Figura 38 – Distribuição dos dias de véspera de chuva por grupo no segundo fator.



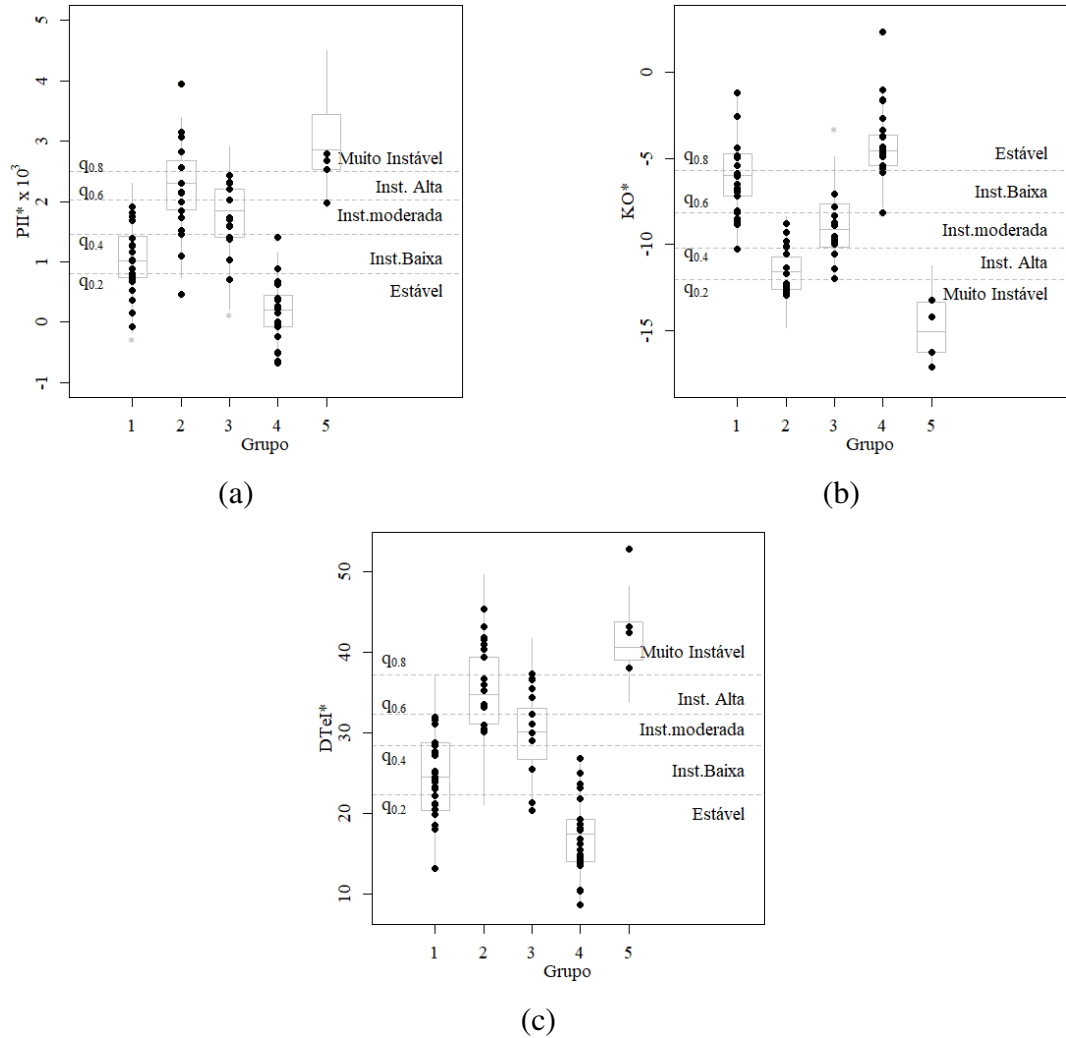
4.4.3.3 Instabilidade dos índices relacionados com a segunda CP nos dias de chuva.

Os dias de chuva pertencentes a cada grupo da segunda CP foram identificados para verificar se os índices de estabilidade são capazes de indicar condições favoráveis à chuva na área de Petrolina.

A maior parte dos dias de chuva estão reunidos nos grupos 1, 2, 3 e 4. No grupo 1, a maioria desses dias indicam instabilidade potencial moderada ou baixa para os índices PII^* , KO^* e $DTel^*$ (Figura 39); e, no Grupo 4, a maioria indica estabilidade. Nos grupos 2, 3, e 5, a maioria dos dias de chuva possuem valores acima do percentil 40% para os índices PII^* , KO^*

e DTeI *, indicando instabilidade de moderada a alta (Figura 39). Em geral, foi identificada uma redução da instabilidade em relação às vésperas de chuva.

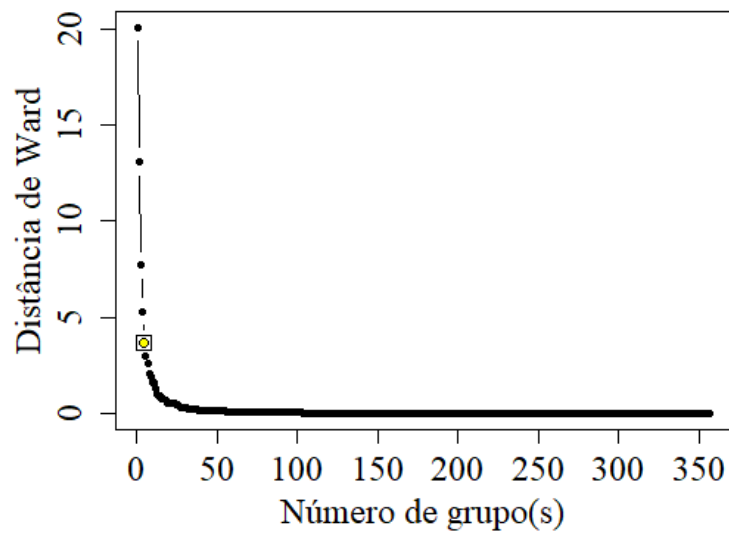
Figura 39 – Distribuição dos dias de chuva por grupo no segundo fator



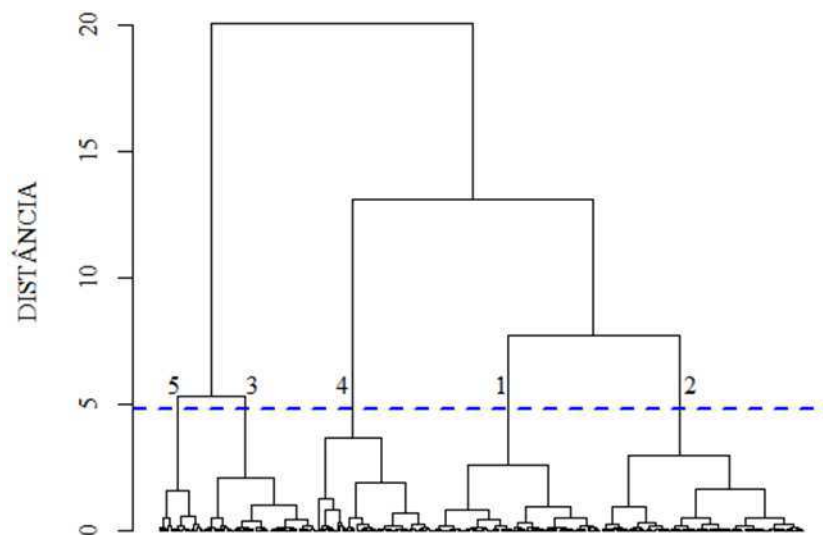
4.4.4 Análise de agrupamentos do terceiro fator

A análise da curva de inércia para o agrupamento do terceiro fator permitiu identificar cinco grupos de dias de acordo com a similaridade das cargas fatoriais. A curva de inércia auxiliou na escolha do ponto de corte no dendrograma para obtenção dos grupos (Figura 40).

Figura 40 – (a) Curva de inércia e (b) dendrograma do agrupamento do terceiro fator.



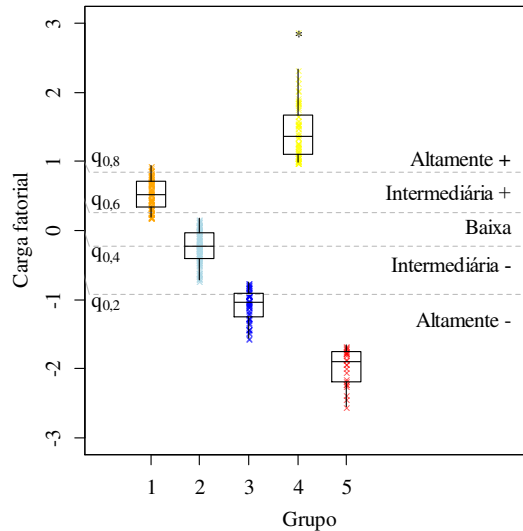
(a)



(b)

A distribuição dos grupos de acordo com as cargas fatoriais é apresentada na Figura 41. No grupo 1, predominam cargas fatoriais intermediárias negativas da primeira CP. Já o grupo 2 está associado com cargas fatoriais intermediárias baixas e intermediárias negativas da terceira CP. Nos grupos 3 e 5, predominam cargas fatoriais altamente negativas, enquanto que, no grupo 4, predominam cargas fatoriais altamente positivas da terceira CP (Figura 41).

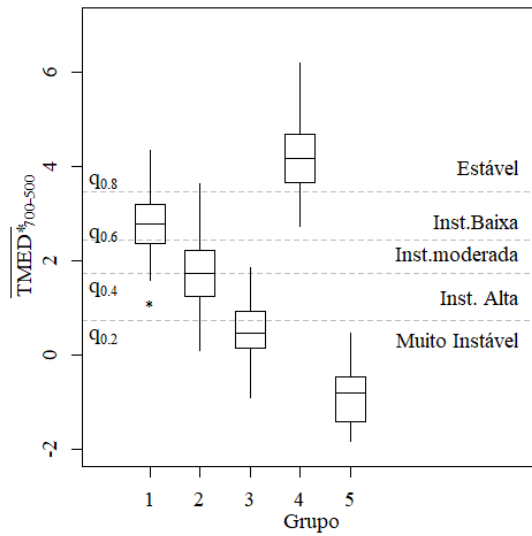
Figura 41 – Distribuição das cargas fatoriais por grupo do terceiro fator: as linhas horizontais representam os percentis 20%, 40%, 60% e 80% dos escores da terceira CP e seus valores são apresentados no eixo vertical à direita da figura.



4.4.4.1 Relação do índice TMED com o terceiro fator

A maior parte dos dias reunidos no grupo 1 da primeira CP está associado com instabilidade baixa ou estabilidade, de acordo com o TMED*. Já o grupo 2 está associado a cargas fatoriais baixas e intermediárias negativas do terceiro fator, reunindo dias com indicativo de instabilidade moderada a alta do TMED*. No grupo 3, que está associado a escores altamente negativos da terceira CP, o TMED* indica instabilidade atmosférica alta ou muito instável. O grupo 4 está associado com escores altamente positivos da terceira CP, reunindo dias em que o TMED* indica estabilidade. O grupo 5 está associado com cargas fatoriais altamente negativas e reúne dias com instabilidade atmosférica muito alta de acordo com o TMED (Figura 42).

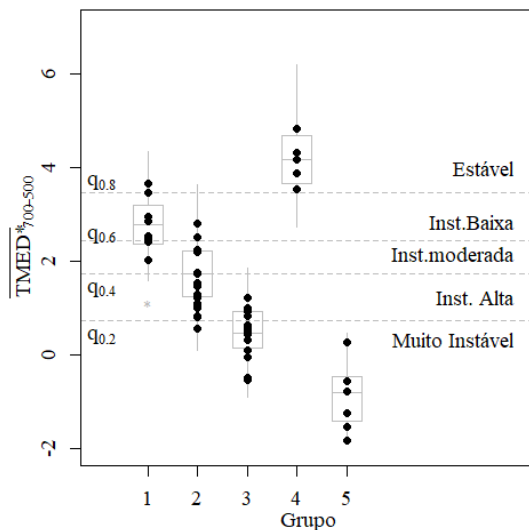
Figura 42 – Distribuição do indicativo de estabilidade por grupo no terceiro fator



4.4.4.2 Instabilidade do índice *TMED* em dias de véspera de chuva

A maior parte dos dias de véspera de chuva estão concentrados nos grupos 2, 3 e 5, estando a maioria desses dias associada com indicativo de instabilidade alto a muito alto do *TMED**. Os grupos 1 e 4 reúnem poucos dias de véspera de chuva, sendo que o grupo 1 está relacionado com instabilidade baixa e o grupo 4 está associado à estabilidade (Figura 43).

Figura 43 – Distribuição dos dias de véspera de chuva por grupo no terceiro fator.

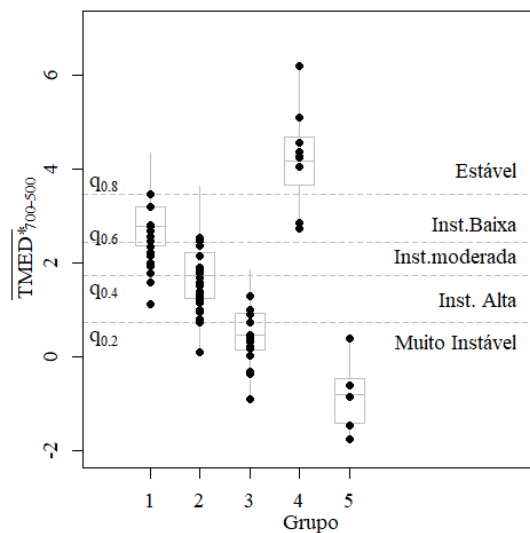


4.4.4.4 Instabilidade do TMED em dias de chuva

A Figura 56 admite uma análise igual à da Figura 55 (a qual se refere aos dias de véspera de chuva). A diferença é que, para os dias de chuva (Figura 44), o indicativo de instabilidade do TMED* é ampliado, uma vez que o valor do índice tende a se concentrar nas faixas de valores que indicam instabilidade mais alta.

Da mesma forma, a maioria dos dias de chuva estão concentrados nos grupos 2, 3 e 5, estando a maioria desses dias associada com indicativo de instabilidade alto a muito alto do TMED*(Figura 44). Os grupos 1 e 4 reúnem poucos dias de véspera de chuva, sendo que o grupo 1 está relacionado com instabilidade baixa a moderada; e o grupo 4 está associado à estabilidade (Figura 44).

Figura 44 – Distribuição dos dias de chuva por grupo no terceiro fator



4.5 ANÁLISE MULTIVARIADA II: ANÁLISE DISCRIMINANTE

Na seção 4.4, o conjunto de índices transformados foram submetidos à ACP, da qual resultou três novas variáveis denominadas CP, que são combinações lineares das variáveis relacionadas com esses fatores. Os escores da primeira CP foram submetidos à AA, o que permitiu obter uma variável que contém a classificação dos dias (casos) em seis grupos. Cada um desses grupos reúne características termodinâmicas similares, agrupados conforme o método de Ward e a distância euclidiana quadrática. Essa variável com a classificação é a variável dependente do modelo, enquanto os índices transformados relacionados com a primeira CP (DMP*, IC*, K*, TNCL*, DPm* e SWEAT*) são as variáveis independentes, ou seja, o conjunto de variáveis que explicam as variáveis dependentes.

Nesta seção, o objetivo é avaliar a qualidade do agrupamento e constatar se há diferenças significativas entre os grupos obtidos pela AA, bem como identificar as variáveis que melhor discriminam os grupos. A técnica estatística multivariada adequada para esse problema é a AD.

4.5.1 Análise Discriminante das variáveis relacionadas com o primeiro fator

4.5.1.1 Análise descritiva dos índices relacionados com o primeiro fator

Nesta etapa da AD, é realizada a análise descritiva das variáveis independentes. O objetivo aqui é verificar se o corpus atende aos pressupostos da técnica.

4.5.1.1.1 Verificação de normalidade multivariada

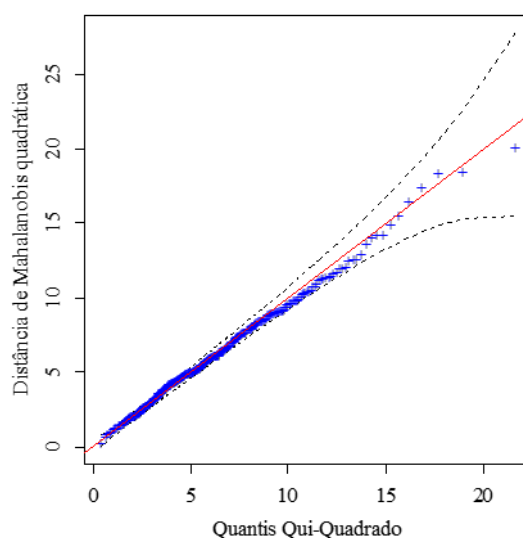
Foi aplicado o teste de normalidade multivariada de Mardia, que leva em consideração os coeficientes de assimetria e curtose multivariados. O teste rejeita a hipótese nula de normalidade multivariada dos dados, já que apenas o coeficiente de curtose possui valor não significativo.

Tabela 3. Estatísticas de normalidade multivariada de Mardia

Teste de Normalidade Multivariada			
Teste	Estatística	P-valor	Resultado
Assimetria	145,853	6,10E+00	NÃO
Curtose	-1,159	0,246	SIM

O gráfico Q-Q plot é então usado para avaliar a discrepância da distribuição normal (Figura 45). A análise do gráfico indica que há indícios de normalidade multivariada. Portanto, prosseguimos com a análise.

Figura 45 – Q-Q plot multivariado para as variáveis relacionadas com a primeira CP. A linha vermelha representa a distribuição normal multivariada; a linha pontilhada representa um intervalo de confiança de onde os pontos nele contidos têm 95% de chance de pertencerem a uma distribuição normal.



4.5.1.1.2 Multicolinearidade

Na Tabela 3, apresenta-se a matriz de correlação das variáveis relacionadas com a primeira CP. Conforme Dorman *et. al.* (2013), as correlações entre cada par de variáveis devem ser inferiores a 0,7 para que o nível de multicolinearidade seja adequada. Já Hair *et. al.* (2009) afirmam que se a correlação for superior a 0,9, a multicolinearidade trará sérios impactos para a interpretação da AD. Na Tabela 3, apresenta-se a matriz de correção das variáveis relacionadas com a primeira CP. Verifica-se que apenas um par de variáveis (DMP e TNCL) apresenta correlação superior a 0,7, permanecendo, no entanto, inferior a 0,9.

Tabela 4. Matriz de correlação das variáveis relacionadas com a terceira CP

	DMP*	IC*	K*	TNCL*	DPm*	SWEAT*
DMP*	1.00	-0.65	-0.53	-0.83	0.48	-0.46
IC*		1.00	0.57	0.46	-0.48	0.57
K*			1.00	0.48	-0.62	0.39
TNCL*				1.00	-0.53	0.51
DPm*					1.00	-0.26
SWEAT*						1.00

4.5.1.1.3 Teste de igualdade de médias

O teste de significância Lambda de Wilks testa a hipótese nula de igualdade de média dos grupos. Logo, o objetivo é rejeitar essa hipótese (NÓBREGA, 2012). O p-valor do teste possui alta significância estatística, indicando que as médias dos grupos diferem. O valor do Lambda de Wilks possui valores intermediários baixos, variando entre 0,39 e 5,7. Ele também indica que as variáveis discriminam bem os grupos, sendo que o índice TNCL* é o que possui maior poder discriminante (menor valor do lambda de Wilks), seguido do IC e do DMP*. O índice DPM* é a variável que menos discrimina os grupos, já que o Lambda de Wilks correspondente é o maior entre as variáveis em estudo (Tabela 1).

Tabela 5. Teste de igualdade de média dos grupos para as variáveis preditoras

Índice	Lambda de Wilks	Estística F	p-valor
DMP*	0,408	101,963	0,00
IC*	0,399	105,948	0,00
K*	0,436	90,932	0,00
TNCL*	0,392	108,882	0,00
DPm*	0,576	51,592	0,00
SWEAT*	0,453	84,602	0,00

4.5.1.1.4 Homogeneidade de dispersão dos grupos multivariados

Aqui queremos testar a hipótese nula de igualdade de dispersão entre os grupos. O método consiste em determinar as distâncias (Euclidiana ou de Mahalanobis) de cada observação ao centro do grupo (média) e avaliar a homogeneidade de dispersão através do teste de significância ANOVA. Quando a distância das observações aos centroides dos grupos é obtida pela distância de Mahalanobis, o teste indica homogeneidade na dispersão entre os grupos, já que o p-valor é não significativo. O mesmo é verificado quando a distância é obtida pelo método euclidiano, considerando-se um nível de significância $\alpha=0,05$.

Tabela 6. Teste de significância ANOVA realizado com o pacote *vegan* do R para teste de análise de variância.

Método	Df	Soma dos Quadrados	Média dos Quadrados	Valor F	P-valor
Mahalanobis	5,000	3,004	0,601	13,966	0,225
Dist. Euclidiana	5,000	4,12 x 10 ¹²	82483845	2,094	0,066

4.5.1.2 Análise das funções discriminantes

Prosseguindo a análise, apresenta-se, a partir desta seção, a análise das regras de classificação (funções discriminantes).

Ao todo, foram obtidas cinco funções discriminantes, das quais a capacidade da primeira função em discriminar os grupos corresponde a 99,2%; a da segunda função equivale a 0,51% em discriminar os grupos; a terceira função tem habilidade de 0,19%; e as duas últimas funções discriminantes (4 e 5) possuem habilidade de discriminar apenas 0,1% dos grupos (Tabela 7).

Tabela 7. Autovalores e proporção da capacidade de discriminação das cinco funções discriminantes

Função	Autovalores	Proporção (%)	Cumulativa (%)
1	12,97	99,21	99,21
2	0,07	0,51	99,72
3	0,02	0,19	99,91
4	0,01	0,08	99,99
5	0,00	0,01	100,00

A Tabela 8 apresenta os testes de Significância Lambda de Wilks para as funções discriminantes. Esse teste avalia a hipótese nula de igualdade dos centroides dos grupos. Logo, para comprovar a capacidade das funções em discriminar os grupos, essa hipótese deve ser rejeitada. Assim, se o p-valor do teste for menor do que o nível de significância $\alpha=0,05$, teremos evidências suficientes para afastar a hipótese de que os centroides dos grupos são iguais (NOBREGA, 2010; KILCA, 2009).

A Tabela 8 permite concluir que apenas a primeira função discriminante é suficiente para discriminar os grupos, pois ela é altamente significativa; e que o valor do lambda de Wilks é próximo de zero. A segunda função discriminante, apesar de ser significativa, possui o valor do lambda de Wilks próximo de 1, ou seja, tem pouco poder de discriminar os grupos.

Tabela 8. Teste de significância das funções discriminantes

Teste de Funções	Wilks' Lambda	Chi2-Value	DF	p. valor
1 até 5	0,065	959,490	25	$2,2 \times 10^{-16}$
2 até 5	0,904	35,415	20	0,018
3 até 5	0,964	12,802	15	0,618
4 até 5	0,988	41,829	10	0,939

A próxima etapa consiste em analisar os coeficientes da função discriminante canônica padronizada (Tabela 9). Eles permitem verificar a relação entre as funções e as variáveis predictoras a elas associadas. A primeira função foi escolhida por ser responsável por praticamente a totalidade do poder discriminatório das funções. A Tabela 9 contém os coeficientes para as cinco funções discriminantes canônicas padronizadas, bastando analisar a primeira para identificar as variáveis mais importantes para discriminar os grupos.

A análise da Tabela 9 permite concluir que as variáveis que tiveram a maior importância para a função discriminante foram os índices SWEAT*, TNCL* e K*. O IC* também foi importante para a determinação da função, porém, sua contribuição foi inferior àquela dos índices SWEAT*, TNCL* e K*.

Tabela 9. Coeficientes da função discriminante canônica padronizada

	F1	F2	F3	F4	F5
DMP*	-0,13814	-0,40453	0,327398	-0,07469	0,903932
IC*	0,558696	0,037927	0,960816	-0,06371	0,165659
K*	0,69018	0,508327	-0,36638	-0,5479	0,397597
TNCL*	0,692486	-0,66051	0,339367	-0,03275	1,067963
DPm*	-0,30324	0,619217	0,267584	0,060819	0,783088
SWEAT*	0,790333	0,209435	-0,51017	0,639892	-0,20371

O cálculo dos centroides dos grupos para os escores discriminantes também permite visualizar a função que melhor discrimina os grupos. Na Tabela 10, é evidente que a primeira função apresenta as maiores distâncias entre os centroides dos grupos, quando comparada à segunda e às demais funções.

Tabela 10. Centroides dos grupos para os escores discriminantes

	F1	F2	F3	F4	F5
Grupo 1	0,368	0,163	0,203	-0,033	-0,009
Grupo 2	-2,047	-0,020	-0,064	0,221	-0,038
Grupo 3	-4,179	-0,246	-0,105	-0,136	-0,022
Grupo 4	2,649	-0,399	0,047	0,044	0,041
Grupo 5	5,642	0,213	-0,226	-0,042	-0,007
Grupo 6	-6,795	0,440	-0,134	0,032	0,082

4.5.1.3 Análise da classificação dos grupos

Utilizou-se saídas do Pacote MASS para a construção da matriz de confronto dos grupos originais versus os grupos previstos pela função (RIPLEY *et al.*, 2019). A matriz indica uma taxa alta de acerto dentro dos grupos, indicando que poucas observações precisaram ser alocadas para os grupos previstos (Tabela 11). Isso indica que os grupos obtidos pela AA são bem discriminados.

Tabela 11. Matriz de confronto para os grupos originais e para os grupos previstos.

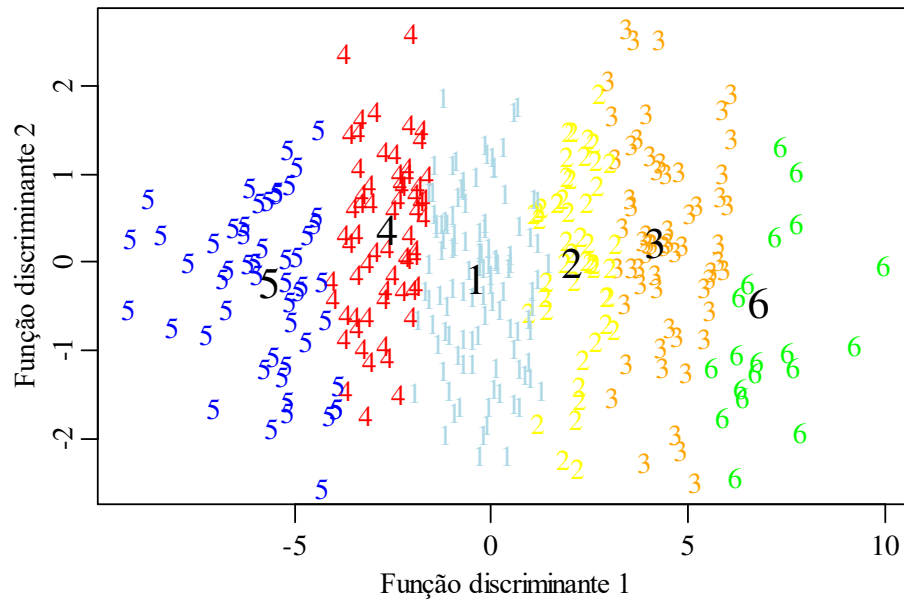
Previstos/Originais	1	2	3	4	5	6
1	98	7	0	6	0	0
2	4	38	7	0	0	0
3	0	3	52	0	0	6
4	7	0	0	53	4	0
5	0	0	0	2	51	0
6	0	0	1	0	0	18

O percentual de acerto dentro dos grupos é obtido fazendo divisão dos valores da diagonal principal pelo número de elementos de cada grupo, isto é, a soma das colunas. Esses valores são exibidos na Tabela 12. Já o acerto global é obtido através da soma da diagonal principal pelo tamanho da amostra, o que equivale a 86,8%. A Figura 46 mostra graficamente o resultado da classificação.

Tabela 12. Porcentagem de classificações agrupadas corretamente.

Grupo	Acerto (%)
1	89,91
2	79,17
3	86,67
4	86,89
5	92,73
6	75,00

Figura 46 – Separação dos grupos pela primeira função discriminante. Os centroides são representados por números na cor preta.



4.5.2 Avaliação das técnicas estatísticas multivariadas

A aproximação de um sistema meteorológico provoca a instabilização do ambiente, que é evidenciada pelo aumento dos valores dos índices de estabilidade. Assim, os grupos identificados a partir dos escores das CP são grupos de dias reunidos de acordo com a dissimilaridade dos indicadores de instabilidade dos índices. Caso o objetivo de aplicar as técnicas estatísticas multivariadas fosse investigar a assinatura que um determinado sistema meteorológico poderia deixar no perfil vertical, seria necessário incluir no conjunto de dados, além de variáveis termodinâmicas, variáveis dinâmicas, tais como as componentes do vento.

Os grupos associados com cargas fatoriais baixas representam situações que ocorrem com mais frequência durante os meses de janeiro. No entanto, a dispersão elevada dos índices de estabilidade, evidenciada pela amplitude elevada dos *boxplots*, pode estar associada a eventuais erros na classificação do dia (como véspera ou dia de ocorrência de chuva). Tal erro pode ter sido induzido devido a uma defasagem entre o horário da sondagem e o da ocorrência da chuva. Os dados de precipitação utilizados nesta pesquisa são provenientes de uma estação meteorológica convencional, ou seja, as medições da precipitação foram realizadas diariamente para 12 UTC. Assim, caso a chuva tivesse ocorrido após o horário da medição do dia anterior, essa chuva apenas seria registrada no dia seguinte. Este raciocínio explica o fato de haver a presença de alguns dias com registro de chuva, associados à instabilidade baixa ou estável nos grupos da primeira e terceira CP, como também dias de véspera de chuva associados à alta instabilidade, que podem estar associados à ocorrência de chuva, mas terem sido classificados erroneamente como véspera.

Deste ponto de vista, a utilização da técnica estatísticas AA não apresentou um resultado satisfatório no sentido de distinguir os dias com condições favoráveis à chuva dos dias associados à condição de estiagem, a partir dos dados de entrada utilizados neste estudo, o qual é composto apenas por variáveis termodinâmicas. Um fato que reforça essa conclusão é o de que os dias de véspera de chuva e os dias de chuva foram alocados em vários grupos, quando o desejável seria ficarem alocados em um ou dois grupos. Por outro lado, isso não compromete a utilidade dos índices como variáveis capazes de prever ou de identificar condições favoráveis à chuva, já que é visível a tendência dos seus valores indicadores de instabilidade se concentrarem nas faixas quantílicas associadas a instabilidade de moderada a alta, tanto nos dias antecedentes à chuva como nos dias de chuva.

Com relação à aplicação da ACP, podemos concluir que essa técnica apresentou bom desempenho em reter as variáveis com maior potencial de detectar a ocorrência das chuvas.

Essas variáveis ficaram retidas no primeiro e no terceiro fator. As variáveis relacionadas com o segundo fator não apresentaram bom desempenho como preditoras de chuva na região de Petrolina, já que há a concentração de muitos dias de véspera de chuva e dias de chuva nos grupos associados à menor condição de estabilidade. Já os grupos associados ao primeiro e terceiro fator apresentam a tendência a se concentrarem nas faixas associadas à maior condição de instabilidade nos dias de véspera e nos dias com ocorrência de chuva. Isto é, os grupos tendem a se concentrar acima do percentil 40% para as variáveis que indicam instabilidade para valores positivos; e abaixo do percentil 60% para as variáveis que indicam instabilidade em direção a valores negativos.

Com relação à técnica da AD, o seu resultado foi complementar ao da AA, porém, o emprego da primeira técnica apresentou um caráter redundante, permitindo apenas constatar que os grupos identificados a partir da AA estavam bem discriminados. Assim, nas próximas pesquisas, sugere-se que o emprego da AD não esteja vinculado à AA; e que a variável descritiva a ser utilizada para a construção das regras de classificação seja definida previamente por meio de outra técnica estatística como, por exemplo, a técnica dos quantis.

5 CONCLUSÕES

Nesta pesquisa, foi realizada a análise climatológica da precipitação e de um conjunto de índices de estabilidade obtidos de uma amostra formada por observações de altitude para 15 meses de janeiro, na região de Petrolina.

A precipitação na região de Petrolina apresenta alta variabilidade, associada a mecanismos em escala interanual, intrasazonal e interdecenal. A variabilidade no número de dias chuvosos é alta, principalmente nos meses da estação chuvosa, padrão que também é observado nos volumes mensais. Já a variabilidade dos volumes diários também é alta, porém, os volumes em que a faixa a precipitação ocorre com ampla frequência são faixas de valores inferiores a 10 mm. Os maiores volumes de precipitação tendem a se concentrar em torno de 50 mm. Eventos acima disto, são considerados extremos, sendo classificados como *outliers* nos gráficos *boxplot*.

Na análise das séries mensais dos índices de estabilidade, é possível concluir que a análise da relação entre o indicativo de instabilidade dos índices e a ocorrência de precipitação na área de Petrolina foi beneficiada pela técnica dos quantis, permitindo comparar o indicativo de instabilidade dos índices nos dias secos, dias de chuva e vésperas de chuva com faixas de valores estabelecidas pelas ordens quantílicas de 20, 40, 60 e 80% para cada índice.

Na análise do desempenho dos índices de estabilidade para dias secos, dias de véspera de chuva, dias de chuva e dias pós chuva para a região de Petrolina, os índices K e SWEAT se apresentaram como os melhores indicadores da chuva para a região de Petrolina. A dispersão do indicativo de instabilidade nos dias de chuva e nas vésperas é mais baixa do que a apresentada pelos demais índices nesses dias.

O indicativo de instabilidade dos índices para dias de véspera de chuva e dias de chuva se concentrou com maior frequência acima do percentil 60%, permitindo concluir que existe uma relação entre esta faixa de valores dos índices e a ocorrência de precipitação na área de Petrolina.

A transformação de Yeo-Johnson se mostrou eficiente para aproximar a distribuição dos índices da distribuição normal, melhorando os pressupostos para a aplicação de técnicas estatísticas multivariadas. Além disso, os resultados gráficos dessa transformação indicam que aqueles casos cujo parâmetro da transformação foi afastado de 1, a transformação de Yeo-Johnson foi igualmente benéfica para a interpretação, fazendo com que os indicadores de instabilidade dos índices transformados sejam um ganho em comparação com os índices

originais. Os índices K, SWEAT, IC e DPM foram influenciados mais significativamente pela transformação.

Na ACP, os índices associados ao teor de umidade nos baixos níveis e levantamento da parcela (TNCL*, IC* e DMP*) foram reunidos na primeira CP e melhor relacionados com essa mesma CP, indicando que o teor de umidade na baixa troposfera é fator preponderante para o aumento de instabilidade atmosférica na área de Petrolina. O segundo fator reuniu os índices relacionados com a instabilidade potencial (PII*, KO* e DTel*). Estes índices não apresentaram uma relação evidente com a ocorrência de precipitação, podendo não ser bons previsores da chuva na área de Petrolina.

O TMED* isolado na terceira CP, simulou bem a quebra da inversão térmica de subsidência na área de Petrolina. A análise dos valores desses índices nos dias de véspera de chuva e nos dias de chuva identificou uma concentração nas faixas de maior instabilidade, concentração esta maior do que a dos índices reunidos nas outras duas CP. Dessa forma, a relação indica um resfriamento da camada 700-500 hpa nos dias favoráveis à chuva.

A aplicação da AA aos escores das CP permitiu reunir grupos homogêneos de índices associados a diferentes faixas de valores de instabilidade. A análise dos grupos para os dias de véspera de chuva e de chuva corrobora a análise feita para o desempenho dos índices, reafirmando a tendência de os valores dos índices se concentrarem acima do percentil 60% (aqueles que indicam instabilidade em direção a valores positivos) e abaixo do percentil 40% (para os índices que indicam instabilidade em direção a valores negativos).

A AD do primeiro fator permite concluir que os grupos obtidos a partir da AA estavam bem discriminados e que apenas uma função foi suficiente para discriminar praticamente a totalidade dos grupos. Os índices SWEAT*, K* TNCL* e IC* foram os mais importantes para a determinação da função discriminante padronizada. Esses índices levam em consideração informações de umidade e vento na baixa e média troposfera, podendo auxiliar na detecção de condições favoráveis à chuva na área de Petrolina.

Na avaliação das técnicas estatísticas multivariadas, constatou-se que a AA, apesar de obter resultados coerentes com outras análises desenvolvidas no presente estudo, não apresentou desempenho satisfatório no sentido de distinguir dias associados à condição de instabilidade dos dias associados à estabilidade, uma vez que os dias de chuva e os dias de véspera de chuva ficaram distribuídos em vários grupos.

A técnica ACP apresentou o melhor desempenho dentre as técnicas estatísticas multivariadas, reunindo nas CP os índices de acordo com seu potencial preditor de chuvas na região de Petrolina.

A aplicação da AD foi redundante, já que a separação obtida pela AA já seria suficiente para discriminar os dias reunidos de acordo com o indicativo de instabilidade dos índices. Deste modo, a AD permitiu apenas constatar que os grupos identificados a partir da AA estavam bem discriminados. Assim, recomenda-se, para pesquisas futuras, que a variável descritiva utilizada para a construção das regras de classificação seja independente da variável de classificação obtida da AA.

REFERÊNCIAS

- ANABOR, V. **Análise descritiva dos sistemas convectivos de escala meso- α através de imagens de satélite GOES-8**. 2004. 78f. Dissertação (Mestrado em Sensoriamento Remoto)- Universidade Federal do Rio Grande Sul, Porto Alegre, 2004.
- ANJOS, B. L. **Conexões entre a circulação do hemisfério norte e os vórtices ciclônicos da alta troposfera na Região Nordeste do Brasil: Um estudo de caso**. 1995. 119f. Dissertação (Mestrado em Meteorologia) - Universidade Federal da Paraíba, Campina Grande, 1995.
- ARAÚJO, H. A. **Padrões de ecos de radar meteorológico no semi-árido brasileiro e condições atmosféricas associadas**. 1996. 130f. Dissertação (Mestrado em Meteorologia) - Universidade Federal da Paraíba, Campina Grande, 1996.
- ARAÚJO, L. E.; SOUSA, F. A. S.; RIBEIRO, M. A. F. M.; SANTOS, A. S.; MEDEIROS, P. C. Análise estatística de chuvas intensas na bacia hidrográfica do Rio Paraíba. **Revista Brasileira de Meteorologia**, v. 23, n. 2, p. 162-169, 2008.
- AWS - AIR WEATHER SERVICE (Estados Unidos). **The use of the Skew T, Log P diagram in analysis and forecasting**. Illinois, Dezembro de 1979 (revisado em março de 1990).
- ATKINS, N.; WAKIMOTO R. Wet microburst activity over the southeastern United States: Implications for forecasting. **Weather and Forecasting**, v. 6, p. 470-482, 1991.
- BARBOSA, T. F.; CORREIA, M. F. Sistemas convectivos intensos no semiárido brasileiro: O controle da grande escala. **Revista Brasileira de Meteorologia**, v. 20, n. 3, p. 395-410, 2005.
- RIPLEY, B.; VENABLES, B.; BATES, M. d.; HORNIK, K.; GEBHARDT, A.; FIRTH, D. MASS: **Package ‘MASS’.version, 7.3-51.4** , 2019. Disponível em: <https://cran.r-project.org/web/packages/MASS/MASS.pdf>
- BOLTON, D. The computation of equivalent potential temperature. **Monthly Weather Review**, v. 108, p. 1046-1053, 1980.
- CHAKRABORTY, R., BASHA G., RATNAM, M.V.. Diurnal and long-term variation of instability indices over a tropical region in India. **Atmospheric Research**, v. 207, p. 145-154, 2018.
- COSTA, R. L.; SOUZA, E. P.; SILVA, F. D. S. Aplicação de uma teoria termodinâmica no estudo de um vórtice ciclônico de altos níveis sobre o Nordeste do Brasil. **Revista Brasileira de Meteorologia**, v. 29, n. 1, p. 96-104, 2014.
- CORREIA, M. F. **Diagnóstico via radar dos sistemas precipitantes do semi-árido brasileiro: O evento de 1985**. 1989. Dissertação (Mestrado em Meteorologia) - Universidade de São Paulo, São Paulo, 1989.
- CORREIA, M. F., SILVA ARAGÃO, M. R., PENAFORTE, A. G., MEDEIROS, R. L B., Características de Dutos Atmosféricos de Superfície no Semiárido Brasileiro, **Anuário do Instituto de Geociências – UFRJ**, v.41, n.3, p. 375-381. 2018.
- DIALEZIOS, N. R.; PAPAMANOLIS, N. K. Objective assessment of indices for operational hail forecasting in Greece. **Meteorology and Atmospheric Physics**, v. 45, p.87-100, 1991.

DINIZ, M. C. S.; CORREIA, M. F, SILVA ARAGÃO, M. R. Convecção profunda no submédio do Rio São Francisco: Uma avaliação das condições atmosféricas na área do aeroporto de Petrolina-PE. **Revista Brasileira de Meteorologia**, v. 19, n. 3, p. 273-285, 2004

EVERITT, B. S. **Cluster Analysis**. London: Edward Arnold,1993.

FRAGOSO, M. A Utilização de índices de estabilidade no estudo de situações convectivas: O exemplo dos temporais do outono de 1997, em Lisboa. **Finisterra Revista Portuguesa de Geografia**, v.XXXIII, n. 66, p. 59-76, 1998.

FRIENDLY, M.; FOX, J; MONETTE, **Package ‘heplots’**. **Version 1.3-5**. 2018. Disponível em <https://cran.r-project.org/web/packages/heplots/heplots.pdf>

HAIR Jr., J.F.; BLACK, W.C.; BABIN, B.J.; ANDERSON, R.E. & TATHAM, R.L. **Análise multivariada de dados**. 6.ed. Porto Alegre, Bookman, 2009. 688p

IBGE - INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA 2010, **Censo Demográfico 2010**. Disponível em: <http://www.censo2010.ibge.gov.br/amostra/> Acesso em 21, mai. 2017.

IBGE – DGC/DEGEO - INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA - DIRETORIA DE GEOCIÊNCIAS / DEPARTAMENTO DE GEOGRAFIA 1990, **Divisão regional do Brasil em mesorregiões e microrregiões geográficas**. Disponível em: http://biblioteca.ibge.gov.br/visualizacao/monografias/GEBIS%20-%20RJ/DRB/Divisao%20regional_v01.pdf Acesso em 21, mai. 2014.

LOPES, J. R. F.; DANTAS, M. P.; FERREIRA, F. E. P. Variabilidade da precipitação pluvial e produtividade do milho no semiárido brasileiro através da análise multivariada. **Nativa**, v. 7, n. 1, p. 77-83, 2019.

LYRA, M.J.A., CAVALCANTE, L.C.V.; LEVIT, V.; FEDOROVA, N. Ligação entre Extremidade Frontal e Zona de Convergência Intertropical sobre a Região Nordeste do Brasil. **Anuário do Instituto de Geociências**. v. 42, n. 1, p. 413-424, 2019.

KANAMITSU, M.; EBISUZAKI, W.; WOOLLEN, J.; YANG, S-K.; HNILO, J. J.; FIORINO, M.; POTTER, G. L. NCEP-DOE AMIP-II Reanalysis (R-2). **Bulletin of the American Meteorological Society**, v. 83, p. 1631-1643, 2002.

KHATTREE, R.; NAIK, D.N. **Multivariate data reduction and discrimination with SAS software**. Cary, NC, USA: SAS Institute Inc., 2000.

KOUSKY, V. E. Frontal influences on Northeast Brazil. **Monthly Weather Review**, v. 107, p. 1140-1153, 1979

KOUSKY, V. E.; GAN, M. A. Upper tropospheric cyclonic vortices in the tropical south Atlantic. **Tellus**, v. 33, p. 538-551, 1981.

KUNZ, M.; SANDER, J.; KOTTMEIER, Ch. Recent trends of thunderstorms and hailstorm frequency and their relation to atmospheric characteristics in southwest Germany. **International Journal of Climate**, v. 29, p. 2283–2297, 2009.

KILCA, R.V.; **Emprego da análise discriminante de Fischer para classificar tipos de florestas no bioma Pampa, RS**, Brasil, 2009, Trabalho de Conclusão de Curso (Especialização em Estatística) - Universidade Federal de Santa Maria, Santa Maria, 2009.

MASON, S. J. Seasonal forecasting of south african rainfall using a non-linear discriminant analysis model. **International Journal of Climatology**, v. 18, p. 147-164, 1998.

MELO, M. J, **Proposta de um teste exato para avaliar a normalidade multivariada baseado em uma transformação tde Studen**, 2016, 86f, Dissertação (Mestrado em Estatística),- Universidade Federal de Lavras, Lavras, 2016.

MOLION, L. C. B.; BERNARDO, S. O. Uma revisão da dinâmica das chuvas no nordeste brasileiro. **Revista Brasileira de Meteorologia**, v. 17, n. 1, p. 1-10, 2002

MOSCATI, M. C. L.; GAN, M. A. Rainfall variability in the rainy season of semiarid zone of Northeast Brazil (NEB) and its relation to wind regime. **International Journal of Climatology**, v. 27, p. 493–512, 2007.

NASCIMENTO, E. L. Previsão de tempestades severas utilizando-se parâmetros convectivos e modelos de mesoescala: Uma estratégia operacional adotável no Brasil? **Revista Brasileira de Meteorologia**, v. 20, n. 1, p. 121-140, 2005.

NÓBREGA, D. M. **Análise discriminante utilizando o Software SPSS**. 2010. 53 f. Trabalho de conclusão de curso (Graduação em Estatística) –Centro de Ciências e Tecnologias da Universidade Estadual da Paraíba, 2010

NYCHKA, D.; GILLELAND, E.; ZHANG, L; HOAR, T: **Package ‘RadioSonde’ , Version 1.4**, 2016, Disponível em: <https://cran.r-project.org/web/packages/RadioSonde/RadioSonde.pdf>, acesso em 21, mai. 2018.

ORLANSKI, I. A rational subdivision of scales for atmospheric processes. **Bulletin of the American Meteorological Society**, v. 56, p. 529-530, 1975.

REBOITA, M. S.; GAN, M. A. S; ROCHA, R. P.; AMBRIZZI, T. Regimes de precipitação na América do Sul: Uma revisão bibliográfica. **Revista Brasileira de Meteorologia**, v. 25, n. 2, p. 185-204, 2010.

RICHMAN, M. B. Rotation of principal components. **Journal of Climatology**, v. 6, p. 293-335, 1986.

RUIZ, N. E. **Análisis climático-sinóptico de la precipitación mediante campos meteorológicos provenientes del análisis numérico y su aplicación al pronóstico estadístico** 1997. 162f. Tese (Doutorado em Ciências Meteorológicas) - Universidad de Buenos Aires, Buenos Aires, 1997.

SANTOS, A. P. P. **Precipitação na cidade de Salvador: Classificação em quantis e análise de um evento climático extremo** 2013. 100f. Tese (Doutorado em Meteorologia) - Universidade Federal da Paraíba, Campina Grande, 2013.

SANTOS, S. R. Q. **Determinação de regiões homogêneas do índice de precipitação normalizada (SPI) no Estado do Pará por meio de análise multivariada**. 2013, 97f. Dissertação (Mestrado em Meteorologia) - Universidade Federal de Campina Grande, Campina Grande, 2013.

SILVA ARAGÃO, M. R.; CORREIA, M. F.; ARAÚJO, H. A. Characteristics of C-band meteorological radar echoes at Petrolina, Northeast Brazil. **International Journal of Climatology**, v. 20, p. 279-298, 2000.

SILVA ARAGÃO, M. R.; CORREIA, M. F., SOBRAL, Z. R. Vento à superfície e chuva em anos contrastantes no Submédio São Francisco. **Boletim Climatológico**. Presidente Prudente-SP, n. 3, p. 213-217, 1997.

SILVA ARAGÃO, M. R.; MENDES, M. C. D.; CAVALCANTI, I. F. A.; CORREIA, M. F. observational study of a rainy January day in the Northeast Brazil semi-arid region: Synoptic and mesoscale characteristics. **Quarterly Journal of the Royal Meteorological Society**, v. 133, p. 1127-1141, 2007.

SILVA, F. D. S.; CORREIA, M. F.; SILVA ARAGÃO, M. R.; SILVA, J. M. Convecção linearmente organizada na área de Petrolina, semi-árido do Nordeste do Brasil: Aspectos de meso e grande escala. **Revista Brasileira de Meteorologia**, v. 23, n. 2, p. 292-304, 2008.

SOUSA, L. F. **Análise multivariada das condições atmosféricas em mês chuvoso extremo na área de Petrolina, semiárido do Nordeste do Brasil**. 2015. 88f. Dissertação (Mestrado em Meteorologia) - Universidade Federal de Campina Grande, Campina Grande, 2015.

SOUZA, P. P. **Estudo da atuação de um sistema de latitudes médias no semi-árido do Nordeste do Brasil: Aspectos de meso e grandes escalas**. 2003. 102f. Dissertação (Mestrado em Meteorologia) – Universidade Federal de Campina Grande, Campina Grande, 2003.

STATHIS, D.; MYRONIDIS, D. Principal component analysis of precipitation in Thessaly Region (Central Greece). **Journal NEST Global**, v. 11, n. 4, p. 467-476, 2009.

TSONIS, A. A. **An introduction to atmospheric thermodynamics**. 2. ed. Cambridge University Press, 2007.

TODOROV, V.; **Package rrcov. Version1.4-7**. 2018. Disponível em <https://cran.r-project.org/web/packages/rrcov/rrcov.pdf>

Blanche, V.; **Package rrcov. Version1.4-7**. 2018. Disponível em <https://cran.r-project.org/web/packages/rrcov/rrcov.pdf>

TINMAKER, M. I. R, ASLAM, M.Y. CHATE, D. M. Association of rainfall and stability index with lightning parameter over the Indo-Gangetic Plains. **American Journal of Climate Change**, v. 06, n. 3, p. 443-454, 2019.

YEO, K.W.,JOHNSON, R. A new family of power transformations to improve normality or symmetry. **Biometrika**,v 87, p. 954-959, 2000.

UMAKANTH, N. SATYANARAYANA, C. BHASKAR RAO, D. V. Analysis and prediction of thunderstorms over Andhra Pradesh using INSAT 3D and MODIS satellite data. **VayuMandal**, v. 45, n. 1, p. 37-59, 2017.

UKKONEN, P., MANZATO, A., MÄKELÄ, A. Evaluation of thunderstorm predictors for Finland using reanalyses and neural networks.**Journal of Applied Meteorology and Climatology**, v.56, n.8, p.2335–2352, 2017.

ZAWADSKI, I.; TORLARCHI, E.; SAUVAGEAU, R. The relationship between mesoscale thermodynamic variables and convective precipitation. **Journal of the Atmospheric Sciences**, v. 38, p. 1535-1540, 1981.

APÊNDICE A– Dias de véspera de chuva organizados conforme agrupamento da primeira CP

Tabela 13 – Dias de véspera de chuva, organizados de acordo com o agrupamento da primeira CP

Grupo 1	Grupo 2	Grupo 3	Grupo 4	Grupo 5	Grupo 6
28/01/2000	13/01/2000	06/01/1999	21/01/2001	08/01/2000	03/01/2011
12/01/2001	04/01/2001	10/01/1999	23/01/2003	26/01/2003	21/01/2011
01/01/2003	21/01/2003	14/01/1999	25/01/2011	-	21/01/2013
13/01/2003	10/01/2005	19/01/2000	09/01/2014	-	27/01/2013
03/01/2004	30/01/2005	08/01/2003	-	-	22/01/2015
09/01/2004	20/01/2008	29/01/2003	-	-	13/01/2016
14/01/2005	12/01/2011	22/01/2005	-	-	-
19/01/2005	31/01/2011	13/01/2007	-	-	-
25/01/2005	19/01/2013	16/01/2009	-	-	-
30/01/2008	27/01/2014	21/01/2009	-	-	-
-	-	24/01/2009	-	-	-
-	-	10/01/2011	-	-	-
-	-	23/01/2011	-	-	-
-	-	21/01/2012	-	-	-
-	-	17/01/2013	-	-	-
-	-	14/01/2014	-	-	-
-	-	24/01/2015	-	-	-
-	-	04/01/2016	-	-	-

APÊNDICE B – Dias de chuva organizados conforme agrupamento da primeira CP

Tabela 14 – Dias de chuva organizados de acordo com o agrupamento da primeira CP

Grupo 1	Grupo 2	Grupo 3	Grupo 4	Grupo 5	Grupo 6
14/01/2000	15/01/1999	07/01/1999	09/01/2000	23/01/2001	11/01/1999
25/01/2000	22/01/2000	12/01/1999	24/01/2003	09/01/2003	13/01/2011
13/01/2001	23/01/2000	20/01/2000	04/01/2004	27/01/2003	22/01/2012
14/01/2003	22/01/2003	21/01/2000	14/01/2007	-	18/01/2013
28/01/2003	14/01/2004	02/01/2003	24/01/2011	-	25/01/2015
30/01/2003	18/01/2004	16/01/2004	-	-	05/01/2016
11/01/2004	19/01/2004	15/01/2005	-	-	06/01/2016
12/01/2004	20/01/2005	31/01/2008	-	-	07/01/2016
13/01/2004	23/01/2005	17/01/2009	-	-	09/01/2016
15/01/2004	26/01/2005	26/01/2009	-	-	14/01/2016
20/01/2004	31/01/2005	26/01/2011	-	-	20/01/2016
21/01/2004	05/01/2011	20/01/2013	-	-	21/01/2016
11/01/2005	22/01/2011	26/01/2015	-	-	22/01/2016
16/01/2005	11/01/2014	08/01/2016	-	-	26/01/2016
17/01/2005	-	15/01/2016	-	-	-
21/01/2005	-	18/01/2016	-	-	-
22/01/2008	-	19/01/2016	-	-	-
05/01/2012	-	23/01/2016	-	-	-
22/01/2013	-	-	-	-	-
28/01/2013	-	-	-	-	-
15/01/2014	-	-	-	-	-
28/01/2014	-	-	-	-	-
23/01/2015	-	-	-	-	-

APÊNDICE C – Dias de véspera de chuva organizados conforme agrupamento da segunda CP

Tabela 15 – Dias de chuva organizados de acordo com o agrupamento da segunda CP

Grupo 1	Grupo 2	Grupo 3	Grupo 4	Grupo 5
08/01/2000	13/01/2000	06/01/1999	10/01/1999	19/01/2000
04/01/2001	13/01/2003	14/01/1999	14/01/2005	08/01/2003
12/01/2001	21/01/2003	28/01/2000	30/01/2008	23/01/2003
21/01/2001	29/01/2003	01/01/2003	03/01/2011	03/01/2004
09/01/2004	10/01/2005	26/01/2003	-	27/01/2013
13/01/2007	19/01/2005	30/01/2005	-	27/01/2014
24/01/2009	22/01/2005	20/01/2008	-	13/01/2016
10/01/2011	25/01/2005	12/01/2011	-	-
21/01/2012	16/01/2009	21/01/2011	-	-
17/01/2013	21/01/2009	23/01/2011	-	-
09/01/2014	25/01/2011	21/01/2013	-	-
-	31/01/2011	14/01/2014	-	-
-	19/01/2013	24/01/2015	-	-
-	22/01/2015	-	-	-
-	04/01/2016	-	-	-

APÊNDICE D – Dias de chuva organizados conforme agrupamento da segunda CP

Tabela 16 – Dias de chuva organizados de acordo com o agrupamento da segunda CP

Grupo 1	Grupo 2	Grupo 3	Grupo 4	Grupo 5
15/01/1999	07/01/1999	21/01/2000	11/01/1999	25/01/2000
22/01/2000	14/01/2000	09/01/2003	12/01/1999	02/01/2003
23/01/2000	20/01/2000	04/01/2004	09/01/2000	24/01/2003
13/01/2001	14/01/2003	16/01/2005	28/01/2003	26/01/2016
23/01/2001	30/01/2003	21/01/2005	11/01/2004	-
22/01/2003	20/01/2005	26/01/2005	14/01/2004	-
27/01/2003	23/01/2005	05/01/2011	15/01/2004	-
12/01/2004	14/01/2007	13/01/2011	18/01/2004	-
13/01/2004	17/01/2009	18/01/2013	20/01/2004	-
16/01/2004	26/01/2009	15/01/2014	21/01/2004	-
19/01/2004	24/01/2011	26/01/2015	31/01/2008	-
11/01/2005	28/01/2013	06/01/2016	26/01/2011	-
15/01/2005	11/01/2014	15/01/2016	22/01/2012	-
17/01/2005	28/01/2014	-	22/01/2013	-
31/01/2005	09/01/2016	-	23/01/2015	-
22/01/2008	-	-	07/01/2016	-
22/01/2011	-	-	08/01/2016	-
05/01/2012	-	-	19/01/2016	-
20/01/2013	-	-	20/01/2016	-
25/01/2015	-	-	21/01/2016	-
05/01/2016	-	-	22/01/2016	-
14/01/2016	-	-	-	-
18/01/2016	-	-	-	-
23/01/2016	-	-	-	-

APÊNDICE E – Dias de véspera de chuva organizados conforme agrupamento da terceira CP

Tabela 17 – Dias de véspera de chuva organizados de acordo com o agrupamento da terceira CP

Grupo 1	Grupo 2	Grupo 3	Grupo 4	Grupo 5
13/01/2000	06/01/1999	28/01/2000	14/01/1999	08/01/2000
19/01/2000	10/01/1999	04/01/2001	08/01/2003	12/01/2001
26/01/2003	21/01/2001	13/01/2003	09/01/2004	21/01/2011
30/01/2005	01/01/2003	21/01/2003	30/01/2008	27/01/2014
20/01/2008	23/01/2003	29/01/2003	25/01/2011	24/01/2015
24/01/2009	03/01/2004	10/01/2005	-	13/01/2016
10/01/2011	19/01/2005	14/01/2005	-	-
17/01/2013	22/01/2005	16/01/2009	-	-
21/01/2013	25/01/2005	21/01/2009	-	-
-	13/01/2007	31/01/2011	-	-
-	03/01/2011	21/01/2012	-	-
-	12/01/2011	22/01/2015	-	-
-	23/01/2011	04/01/2016	-	-
-	19/01/2013	-	-	-
-	27/01/2013	-	-	-
-	09/01/2014	-	-	-
-	14/01/2014	-	-	-

APÊNDICE F – Dias de chuva organizados conforme agrupamento da terceira CP

Tabela 18 – Dias de véspera de chuva organizados conforme o agrupamento da terceira CP

Grupo 1	Grupo 2	Grupo 3	Grupo 4	Grupo 5
14/01/2003	07/01/1999	12/01/1999	15/01/1999	13/01/2001
22/01/2003	11/01/1999	09/01/2000	14/01/2000	24/01/2003
14/01/2004	20/01/2000	23/01/2000	09/01/2003	22/01/2008
18/01/2004	21/01/2000	25/01/2000	16/01/2004	05/01/2012
31/01/2008	22/01/2000	28/01/2003	24/01/2011	23/01/2015
26/01/2009	23/01/2001	12/01/2004	28/01/2013	-
13/01/2011	02/01/2003	13/01/2004	06/01/2016	-
22/01/2012	27/01/2003	20/01/2004	20/01/2016	-
18/01/2013	30/01/2003	11/01/2005	26/01/2016	-
11/01/2014	04/01/2004	16/01/2005	-	-
15/01/2014	11/01/2004	20/01/2005	-	-
28/01/2014	15/01/2004	21/01/2005	-	-
07/01/2016	19/01/2004	26/01/2011	-	-
08/01/2016	21/01/2004	15/01/2016	-	-
14/01/2016	15/01/2005	23/01/2016	-	-
19/01/2016	17/01/2005	-	-	-
21/01/2016	23/01/2005	-	-	-
22/01/2016	26/01/2005	-	-	-
-	31/01/2005	-	-	-
-	14/01/2007	-	-	-
-	17/01/2009	-	-	-
-	05/01/2011	-	-	-
-	22/01/2011	-	-	-
-	20/01/2013	-	-	-
-	22/01/2013	-	-	-
-	25/01/2015	-	-	-
-	26/01/2015	-	-	-
-	05/01/2016	-	-	-
-	09/01/2016	-	-	-
-	18/01/2016	-	-	-



UFCC - CTRN - UACA - PPGMet
AV. Aprígio Veloso, 882 - Bairro Universitário
58.429-140 Campina Grande - Paraíba
Fone: (83)2101-1054 - Fax: (83)2101-1650