

UFPA

PPGEC

Universidade Federal do Pará



Evanice Pinheiro Gomes

Regionalização de Precipitações via Fuzzy C-means

DISSERTAÇÃO DE MESTRADO

Instituto de Tecnologia
Programa de Pós-Graduação em Engenharia Civil

Dissertação orientada pelo Professor Claudio José Cavalcante Blanco

Belém – Pará – Brasil

2017

REGIONALIZAÇÃO DE PRECIPITAÇÕES VIA FUZZY C-MEANS

Evanice Pinheiro Gomes

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Civil da Universidade Federal do Pará.

Linha de Pesquisa: Recursos Hídricos e Saneamento Ambiental.

Orientador: Prof. Claudio José Cavalcante Blanco Ph.D.

Coorientador: Prof. Dr. Francisco Carlos Lira Pessoa.

BELÉM/PA
2017

REGIONALIZAÇÃO DE PRECIPITAÇÕES VIA FUZZY C-MEANS

Evanice Pinheiro Gomes

Dissertação de Mestrado submetida ao Corpo Docente do Programa de Pós-Graduação em Engenharia Civil da Universidade Federal do Pará, como parte dos requisitos à obtenção do título de Mestre em Engenharia Civil.

Aprovado em: __/__/2017

Banca Examinadora:

Prof. Claudio José Cavalcante Blanco Ph.D.
(FAESA/UFPA - Orientador)

Prof. Dr. Francisco Carlos Lira Pessoa
(FAESA/UFPA - Coorientador)

Prof. Dr Lindemberg Lima Fernandes
(FAESA/UFPA - Examinador Interno)

Prof. Dr Adílson Pinheiro
(FURB - Examinador Externo)

Dados Internacionais de Catalogação na Publicação (CIP)

Gomes, Evanice Pinheiro
Regionalização de precipitações via Fuzzy C-means
/Evanice Pinheiro Gomes.- 2017.

Orientador: Claudio José Cavalcante Blanco;
Coorientador: Francisco Carlos Lira Pessoa

Dissertação (Mestrado) - Universidade Federal do Pará.
Instituto de Tecnologia. Programa de Pós-Graduação em
Engenharia Civil, Belém, 2017

1.Chuvas - Pará 2. Precipitação (meteorologia)- Pará
3.Lógica Fuzzy I. Título

CDD 22.ed.551.7728115

AGRADECIMENTOS

Agradeço a Deus acima de tudo, por sempre está ao meu lado, proporcionando-me perseverança, sabedoria e determinação na realização deste trabalho.

Aos meus pais que me ensinaram a viver com dignidade: a minha gratidão, respeito, e eterno amor.

A minha família pelo amor, incentivo e compreensão em todos os momentos da minha vida, especialmente as minhas irmãs: Antônia, Ilma e Ruthilene, por todo apoio nas horas difíceis do dia-a-dia.

Ao meu companheiro Edson por todo amor, apoio, incentivo e compreensão.

Ao meu orientador Claudio José Cavalcante Blanco, Ph.D. pela orientação, esclarecimentos, dedicação e apoio oferecidos ao longo do desenvolvimento deste trabalho.

Ao meu Co-orientador, Francisco Carlos Lira Pessoa, pela sua orientação e amizade no caminho desta etapa da vida.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq pela concessão da bolsa de estudo.

Aos colegas da Pós-Graduação em Engenharia Civil, pelo agradável convívio e apoio durante o curso.

Finalmente, a todos que de alguma maneira contribuíram para a realização dessa dissertação.

RESUMO

GOMES, Evanice Pinheiro. **Regionalização de Precipitações via Fuzzy C-means**. Dissertação de Mestrado em Recursos Hídricos e Saneamento Ambiental – Programa de Pós-Graduação em Engenharia Civil da Universidade Federal do Pará, Belém, 2017.

O conhecimento do comportamento da precipitação é indispensável, pois qualquer alteração na sua quantidade e distribuições espaciais e temporais, têm um impacto importante sobre a natureza e conseqüentemente nas diversas atividades humanas. Contudo em estudos de precipitação, a falta de monitoramento pluviométrico, gerando a ausência de informação ao longo do tempo e espacialmente nas bacias hidrográficas é um problema. Com o intuito de contornar esse problema, no presente trabalho, foi proposto a aplicação do método de regionalização de precipitações. Nesse contexto, a ideia principal foi dividir a região hidrográfica Tocantins Araguaia – RHTA em regiões homogêneas, definidas pelo método Fuzzy C-means, e para cada uma, definir modelos de probabilidades de ocorrência de chuvas e modelos regionais de estimativa de precipitações mensais e anuais. Para o desenvolvimento do método Fuzzy C-means, adotou-se a distância euclidiana, como medida de similaridade e o parâmetro de fuzificação, variando de 1,2 a 2,0, e as variáveis explicativas de precipitação (altitude, latitude e longitude), foram usadas como dados de entrada. Foram obtidas 3 regiões homogêneas por intermédio do método Fuzzy C-means, as quais foram validadas pelo índice PBM e o teste de heterogeneidade. As frequências de ocorrências de chuvas observadas foram geradas para cada uma das 83 estações pluviométricas, distribuídas em suas respectivas regiões, e calibradas pelas funções de probabilidade Normal, Log-Normal, Gama, Gumbel, Exponencial, Logarítmica e Weibull. Com a aplicação do teste do qui-quadrado, definiu-se a melhor função de probabilidade na ocorrência de chuvas de cada região homogênea. A validação das funções de probabilidades foi realizada em 9 estações alvo, a partir do teste qui-quadrado. Nesta etapa, foi observado que para precipitação média anual, ocorreu aderência dos dados a todas as estações pluviométricas alvo, pois apresentaram resultados da aplicação do teste qui-quadrado inferior a 5,99 (para funções de distribuição Log-normal). Também se constatou que para precipitação média mensal, houve aderência dos dados a todas as estações pluviométricas alvo, com as funções Gama e Weibull. Para simulação de alturas de precipitação, foram testados na calibração, modelos Linear, Potencial, Exponencial e Logaritmo, através do método de

regressão múltipla, adotando como variáveis independentes, a altitude, latitude e longitude. Como critério de desempenho dos modelos, foi utilizado o R^2 , R^2_a , $E\%$, $\epsilon\%$, $NASH$ e $RMSE$. Na simulação de médias anuais o modelo Linear apresentou os melhores índices de desempenho. Na estimativa de precipitações médias mensais, todos os modelos de regressão múltipla não apresentaram bom desempenho, com erros acima de 50%, fato que provocou a estimativa de chuvas médias mensais para períodos chuvosos e secos. Nessa nova abordagem os modelos de regressão apresentaram critérios de desempenho adequados e com erros abaixo de 10%. Os índices de desempenho encontrados permitiram concluir que os modelos regionais desenvolvidos para as regiões homogêneas de precipitações, definidas por meio do método Fuzzy C-Means, apresentam-se como uma opção adequada na estimativa de precipitações anuais e em períodos secos e chuvosos, e são importantes para um melhor entendimento do regime pluviométrico na RHTA, podendo servir como ferramenta para um melhor planejamento dos recursos hídricos na região.

Palavras Chave: Precipitação, Regiões Homogêneas, Regionalização

Abstract of the dissertation presented to the Graduate Program in Civil Engineering of the Federal University of Pará as part of the requirements to obtain the title of Master in Civil Engineering.

PRECIPITATION REGIONALIZATION USING FUZZY C-MEANS

Evanice Pinheiro Gomes

March/2017

Advisor: Claudio José Cavalcante Blanco

Research Area: Water Engineering

The knowledge of the precipitation behavior is indispensable, since any change in its quantity and spatial and temporal distributions have an important impact on nature and consequently on the various human activities. However, in precipitation studies, the lack of rainfall monitoring, generating the lack of information over time and spatially in the river basins is a problem for the understanding of this variable. In order to overcome this problem, the rainfall regionalization method was proposed. The main idea was to divide the Tocantins Araguaia - RHTA hydrographic region into homogeneous regions, defined by the Fuzzy C-means method. The Euclidean distance was adopted as a measure of similarity and the fuzzification parameter, ranging from 1.2 to 2.0, and the explanatory variables of rainfall (altitude, latitude and longitude) were used as input data. Three homogeneous regions were obtained, which were validated by the PBM index and the heterogeneity test. The frequencies of observed rainfall events were generated for the 83 rain gauge stations, distributed in their respective regions, and calibrated by the Normal, Log-Normal, Gama, Gumbel, Exponential, Logarithmic and Weibull probability functions. With the application of the chi-square test, we defined the best probability function in the occurrence of rainfall in each homogeneous region. The validation of the probabilities functions was performed in 9 target stations, using the chi-square test. In this stage, it was observed that for annual average precipitation, data adherence occurred to all the rain gauge stations, since they presented results of the chi-square test of less than 5.99 (for Log-normal distribution functions). It was also observed that for monthly average precipitation, data were adhered to all the rainfall stations with the Gama and Weibull functions. For the simulation of rain depth, Linear, Potential, Exponential and Logarithm models were tested through the multiple regression method, using as independent variables, altitude, latitude and longitude. As performance criterion of the models, the R^2 , R^2_a , $E\%$, $\varepsilon\%$, NASH and RMSE were used. In the simulation of annual averages, the Linear model presented the best performance indices. In the estimation of

monthly averages, all multiple regression models did not perform well, with errors above 50%, which motivated the estimation of monthly rainfall for rainy and dry periods. In this new approach the regression models presented excellent performance criteria with errors below 10%. The performance indexes allowed us to conclude that the regional models developed for the homogeneous regions of rainfall, defined by the Fuzzy C-Means method, are a good option in the estimation of annual and monthly average rainfall and are important for a better understanding of the rainfall regime in RHTA, and can serve as a tool for better planning of water resources in the region.

Keyword: Precipitation, Homogeneous Regions, Amazon Region.

LISTA DE FIGURAS

Figura 1 – Esquema do Ciclo Hidrológico.....	19
Figura 2 – Aparelhos de Medição de Chuva.....	22
Figura 3 – Comportamento da Precipitação conforme latitude.....	24
Figura 4 – Variação Espacial da Precipitação no mundo.....	24
Figura 5 - Precipitação mensal média de 1961-1990 no Brasil.....	25
Figura 6 – Região Hidrográfica Tocantins Araguaia - RHTA.....	34
Figura 7 – Sub bacias da RHTA.....	34
Figura 8 – Estações Pluviométricas.....	38
Figura 9 – Etapas da Metodologia.....	39
Figura 10 – Estrutura do Algoritmo Fuzzy C-means.....	44
Figura 11 – Índice PBM dos agrupamentos.....	70
Figura 12 - Graus de Pertinência das estações.....	71
Figura 13 - Regiões Homogêneas de Chuva na RHTA.....	73
Figura 14 - Gráfico de dispersão dos quocientes CV-L e Assimetria-L.....	75
Figura 15 – Funções de distribuição de Probabilidade da simulação de médias anuais de precipitação para a Região Homogênea I.....	77
Figura 16 - Funções de distribuição de Probabilidade na simulação de médias anuais de precipitação para a Região Homogênea II.....	78
Figura 17 - Funções de distribuição de Probabilidade na simulação de médias anuais de precipitação para a Região Homogênea III.....	79
Figura 18 – Distribuições de Probabilidade observada e simulada das médias mensais da Região Homogênea I.....	82
Figura 19 – Distribuições de Probabilidade observada e simulada das médias mensais da Região Homogênea II.....	84
Figura 20 – Distribuições de Probabilidade observada e simulada das médias mensais da Região Homogênea III.....	85
Figura 21 – Localização das Estações Alvo.....	88
Figura 22 – Probabilidade de ocorrência de chuva observada e a simulada das Estações Alvos – Região Homogênea I.....	89
Figura 23 - Probabilidade de ocorrência de chuva observada e a simulada das Estações Alvos – Região Homogênea II.....	90

Figura 24 - Probabilidade de ocorrência de chuva observada e a simulada das Estações Alvos – Região Homogênea III.....	90
Figura 25 – FDP's ajustadas.....	93
Figura 26– Erros percentuais por estação alvo das chuvas médias mensais, utilizando modelo Linear.....	105
Figura 27 Erros percentuais por estação alvo das chuvas médias para período chuvoso, utilizando modelo Linear.....	112
Figura 28 – Erros percentuais por estação alvo das chuvas médias para período seco, utilizando modelo.....	113
Figura 29 - Regiões Homogêneas de Precipitação e Modelos Probabilísticos e Matemáticos para estimativa de chuvas na RHTA.....	115

LISTA DE TABELAS

Tabela 1 – Características dos principais satélites usados em estimativas de chuva.....	31
Tabela 2 – Informações das Estações Pluviométricas.....	38
Tabela 3 – ANOVA de regressão múltipla.....	59
Tabela 4 – Matrix de Correlação de Pearson.....	63
Tabela 5 – Índice PBM para cada agrupamento e parâmetro de fuzificação.....	68
Tabela 6 – Valores máximos do Índice PBM.....	70
Tabela 7 – Comportamento das Variáveis nas Regiões Homogêneas.....	73
Tabela 8 – Quocientes de Momentos-L para as estações estudadas.....	74
Tabela 9 – Teste X^2 para as Funções de Probabilidade.....	76
Tabela 10 - Valores do X^2 das simulações das séries mensais -Região Homogênea I...80	80
Tabela 11 - Valores do X^2 das simulações das séries mensais-Região Homogênea II...81	81
Tabela 12 - Valores do X^2 das simulações das séries mensais- Região Homogênea III..82	82
Tabela 13 – Estações Pluviométricas Alvo adotadas para a calibração.....	87
Tabela 14 - Valores do X^2 na validação dos modelos probabilísticos para as séries anuais.....	88
Tabela 15 - Valores do X^2 na validação dos modelos probabilísticos para as séries mensais.....	91
Tabela 16 - Matriz de correlação entre as variáveis explicativas das regiões homogêneas definidas por meio do método <i>fuzzy c-means</i>	102
Tabela 17 - Coeficiente de desempenho dos Modelos de Regressão.....	103
Tabela 18 - Validação do Modelo de Regressão Linear nas estimativas de chuva média anual.....	104
Tabela 19 - Coeficiente de Determinação e o Erro relativo percentual obtido para cada mês na Região Homogênea I.....	106
Tabela 20 - Coeficiente de Determinação e o Erro relativo percentual obtido para cada mês na Região Homogênea II.....	107
Tabela 21 - Coeficiente de Determinação e o Erro relativo percentual obtido para cada mês na Região Homogênea III.....	108
Tabela 22 - Coeficientes de Desempenho dos modelos para os período chuvoso.....	109
Tabela 23 - Coeficientes de Desempenho dos modelos para os período seco.....	110
Tabela 24 - Validação do Modelo de Regressão Linear para o Período Chuvoso.....	111
Tabela 25 - Validação do Modelo de Regressão Potencial para o Período Seco.....	113

Tabela 26 - Modelos de Regressão para estimativa de totais de precipitação média anual.....	113
Tabela 27 - Modelos de Regressão para estimativa de totais de precipitação média mensal para período chuvoso.....	114
Tabela 28 - Modelos de Regressão para estimativa de totais de precipitação média mensal para período seco.....	114

NOMENCLATURA

ANA	Agência Nacional de Águas
ARA	Araguaia
EMBRAPA	Empresa Brasileira de Pesquisas Agropecuárias
ENOS	El nino oscilação sul
FCM	Fuzzy C-means
FCP	Função Cumulativa de Probabilidade
FDP	Função de Distribuição de Probabilidade
GMS	Geostationary Meteorological Satélite
GOES	Geostacionary Operational Enverionmental Satélite
IOD	Dipolo Oceano Índico
IBGE	Instituto Brasileiro de Geografia e Estatística
MMA	Ministério do Meio Ambiente
MML	Método dos Momentos-L
MPP	Momentos Ponderados de Probabilidade
NASA	National Aeronautics and Space Administration
NOAA	National Oceanic and Atmospheric Administration
OMJ	Oscilação Madden Julian
RHTA	Região Hidrográfica Tocantins Araguaia
TOA	Tocantins Alto
TOB	Tocantins Baixo
ZCTI	Zona de Convergência Intertropical
ZCAS	Zona de Convergência Atlântico Sul

SUMÁRIO

1 INTRODUÇÃO.....	17
1.1 OBJETIVO GERAL.....	18
1.2 OBJETIVOS ESPECÍFICOS.....	18
2 FUNDAMENTAÇÃO TEÓRICA.....	19
2.1 CICLO HIDROLÓGICO.....	19
2.2 PRECIPITAÇÃO.....	20
2.2.1 Formas e Tipos.....	20
2.2.2 Grandezas e Medição.....	21
2.2.3 Fatores que influenciam na precipitação.....	23
2.3 REGIONALIZAÇÃO HIDROLÓGICA.....	27
2.3.1 Regiões Homogêneas.....	28
2.4 MODELOS DE ESTIMATIVAS DE CHUVA.....	30
2.4.1 Funções de Distribuição de Probabilidade – FDP.....	31
2.4.2 Modelos Regionais de Regressão Múltipla.....	31
3 ÁREA DE ESTUDO.....	33
3.1 LOCALIZAÇÃO.....	33
3.2 CARACTERÍSTICAS CLIMÁTICAS.....	35
3.3 HIDROGRAFIA E VEGETAÇÃO.....	35
3.4 USO E OCUPAÇÃO DO SOLO.....	36
3.5 ASPECTOS ECONÔMICOS.....	37
3.6 BANCO DE DADOS.....	37
4 METODOLOGIA.....	39
4.1 ANÁLISE DE AGRUPAMENTO.....	40
4.2 FUZZY C-MEANS.....	41
4.3 MEDIDAS DE SIMILARIDADE.....	44

4.4	INDICE DE VALIDAÇÃO PBM.....	45
4.5	MOMENTOS – L.....	47
4.6	TESTE DE HETEROGENEIDADE (H).....	50
4.7	DISTRIBUIÇÕES DE PROBABILIDADE.....	52
4.7.1	Teste de Aderência – Qui-quadrado.....	57
4.7.2	Calibração dos Modelos de probabilidade.....	58
4.7.3	Validação dos Modelos de probabilidade.....	58
4.8	REGIONALIZAÇÃO POR REGRESSÃO MÚLTIPLA.....	58
4.8.1	Modelos de Regressão Múltipla.....	61
4.8.2	Teste de Multicolinearidade.....	62
4.8.3	Desempenho dos Modelos de Regressão.....	63
4.8.4	Calibração da Regressão Múltipla.....	65
4.8.5	Validação dos Modelos de Regressão Múltipla.....	66
5	RESULTADOS E DISCUSSÕES.....	67
5.1	FORMAÇÃO DAS REGIÕES HOMOGÊNEAS.....	67
5.1.1	Método Fuzzy c-means.....	67
5.1.2	Formação das Regiões homogêneas.....	72
5.1.3	Teste de Heterogeneidade (H).....	74
5.2	ESTIMATIVAS DE CHUVAS POR FUNÇÕES DE PROBABILIDADE.....	76
5.2.1	Calibração das funções de probabilidade para Chuva média Anual.....	76
5.2.2	Calibração das funções de probabilidade para Chuva média Mensal.....	80
5.2.3	Validação dos Modelos de Probabilidade.....	87
5.3	REGIONALIZAÇÃO POR REGRESSÃO MÚLTIPLA.....	102
5.3.1	Análise de Multicolinearidade.....	102
5.3.2	Calibração dos Modelos de Regressão.....	103

5.3.3 Validação dos Modelos de Regressão	104
5.3.4 Calibração dos Modelos de Regressão para Precipitações Médias Mensais	105
5.3.5 Estimativas de Chuvas para Período Chuvoso e Seco	108
5.3.6 Calibração dos Modelos de Regressão para período Chuvoso e Seco	108
5.3.7 Validação do Modelo de Regressão para período Chuvoso e Seco	111
6 CONCLUSÃO E SUGESTÕES	116
6.1 SUGESTÕES E RECOMENDAÇÕES.....	116
REFERÊNCIAS	118
APENDICE A Estações Pluviométricas utilizadas no estudo	124
APENDICE B - Algoritmo Fuzzy C-means FCM.....	128
APENDICE C - Algoritmo FDP.....	131
APENDICE D - Graus de Pertinência das Estações.....	135
APENDICE E - Quocientes de Momentos-L.....	136
APENDICE F – Valores dos Índices de Desempenho obtidos na calibração dos modelos de Regressão para as médias mensais.....	138
ANEXO A – Tabela com valores do Qui-quadrado X^2 tabelado.....	141

1 INTRODUÇÃO

A precipitação é uma variável chave em estudos climáticos, pois com as mudanças no clima, altera-se a sua quantidade e distribuições espaciais e temporais, que têm um impacto importante sobre ambas as atividades humanas (como a agricultura e abastecimento de água potável), e os eventos extremos (tais como secas e inundações) (Lopez-Moreno *et al.* 2008).

Por toda a sua importância, o conhecimento da precipitação é indispensável, no entanto, um dos problemas que se apresenta em estudos de precipitação, é a falta de monitoramento pluviométrico, gerando a ausência de informação ao longo do tempo e espacialmente nas bacias hidrográficas. Assim, para tentar preencher essa falta de dados, o desenvolvimento de modelos de estimativas de chuvas, em bacias sem monitoramento, tem se tornado muito comum. A regionalização é uma alternativa que contribui para a solução do problema.

A regionalização hidrológica consiste na utilização de um conjunto de ferramentas que buscam formas de transferência de informações de um local para outro, em bacias hidrográficas, com o intuito de suprir a carência de informação, em uma dada região considerada homogênea (SAMUEL *et al.* 2011).

Para a aplicação da regionalização de chuvas, procedimentos matemáticos e estatísticos são aplicados às séries de dados históricos de precipitação e às características físicas e climáticas das bacias hidrográficas, através de modelos hidrológicos, que depois de calibrados e validados são capazes de estimar chuvas em regiões homogêneas.

Uma das condições para a aplicação da regionalização é a identificação de regiões homogêneas, considerada etapa inicial do processo. De acordo com Hosking & Wallis (1997), a identificação dessas regiões pode ser realizada de várias maneiras, como por exemplo, através de convergência geográfica, análise fatorial, análise de formas das funções de probabilidade, análise de agrupamento, entre outros.

Diversos trabalhos de identificação de regiões homogêneas têm sido realizados no mundo inteiro, objetivando a aplicação da regionalização, como os trabalhos desenvolvidos por RAO & SIRVINAS (2006); COSTA *et al.* (2011); SATYANARAYANA & SIRVINAS (2011); DIKBAS *et al.* (2012); ZHANG (2012); SANTOS (2014) e PARRACHO & ROCHA (2015).

No presente estudo, adotou-se o método de análise de agrupamento, por meio da técnica Fuzzy c-means, para identificar regiões homogêneas de precipitação na Região Hidrográfica Tocantins-Araguaia – RHTA. Essa metodologia tem como vantagens, sua

flexibilidade no uso de inúmeras características dos elementos a serem agrupados, uso de parâmetros específicos adotados, além da subjetividade na formação de regiões homogêneas.

Assim, para a definição das regiões homogêneas de precipitação, adotou-se algumas características físico-climáticas da bacia, como a localização geográfica (latitude, longitude), altitude e precipitação média anual. O número de agrupamentos, via fuzzy c-means, é definido pelo índice PBM e os grupos formados são testados quanto a sua heterogeneidade pelo teste H. Definidas as regiões, aplicam-se os modelos de função de probabilidade para estimativas de lâminas de chuvas anuais e mensais, e ao final aplica-se a regressão múltipla para definir os modelos de regionalização.

Os modelos de regionalização de chuvas podem contribuir na quantificação dos dados hidrológicos, inclusive nos dados de vazão, que são mais escassos, mas que com o desenvolvimento dessa técnica, poderão ser estimados através dos modelos chuva-vazão, sendo a chuva oriunda da regionalização, e ainda servirão de apoio na gestão e planejamento dos recursos hídricos, fornecendo informações sobre o comportamento da precipitação na RHTA.

1.1 OBJETIVO GERAL

Desenvolver um modelo de regionalização para estimativas de precipitações mensais e anuais para a Região Hidrográfica do Tocantins Araguaia – RHTA.

1.2 OBJETIVOS ESPECÍFICOS

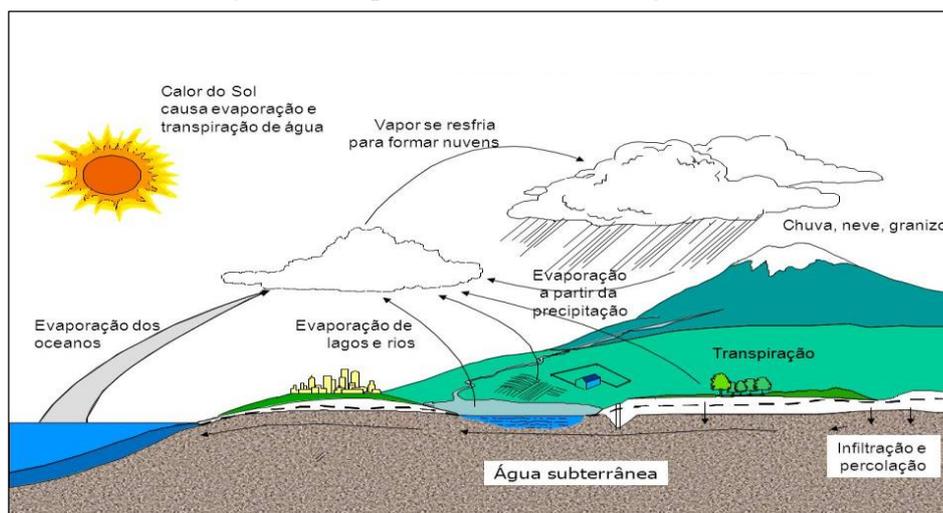
- Aplicar o método Fuzzy c-means para identificação de regiões homogêneas de precipitação;
- Aplicar a estatística de momento-L, através do teste H para verificar a homogeneidade das regiões homogêneas de chuvas formadas;
- Determinar a probabilidade de ocorrência de precipitações médias mensais e anuais através das funções de probabilidade;
- Avaliar modelos de regressão múltipla para determinação das precipitações médias mensais e anuais nas regiões homogêneas de precipitação.

2 FUNDAMENTAÇÃO TEÓRICA

2.1 CICLO HIDROLÓGICO

A água distribuída no planeta está em contínua movimentação. A quantidade existente na atmosfera, em um rio e nas geleiras não é constante. Essa dinâmica de circulação da água explica-se pelo ciclo hidrológico, o qual envolve os fenômenos de evaporação, precipitação, transpiração, infiltração e escoamento superficial (Figura 1).

Figura 1 – Esquema do Ciclo Hidrológico.



Fonte: <http://www.meioambiente.culturamix.com>.

Pode-se considerar, inicialmente, que ocorre a evaporação, que é a mudança da água no estado líquido para vapor devido à radiação solar e aos processos de difusão molecular e turbulenta. Neste caso as águas dos oceanos, mares, rios e lagos evaporam, ficando retidas na atmosfera. A fração da água evaporada, somada àquela liberada pela atividade biológica de organismos, principalmente as plantas, através da transpiração é chamada de evapotranspiração. Num segundo momento, tem-se a precipitação, que é entendida como toda água proveniente do meio atmosférico, que atinge a superfície terrestre, como por exemplo, neblina, chuva, granizo, saraiva, orvalho, geada e neve. A água advinda das precipitações que chega ao solo pode escoar sobre a superfície encaminhando-se para os cursos d'água ou infiltrar-se no solo. A parte de água que infiltra no solo pode ser absorvida pelas raízes de plantas e voltar para a atmosfera por meio da respiração ou percolar para o lençol freático e contribuir para o escoamento de base dos rios.

O Ciclo hidrológico manifesta-se de diferentes maneiras nas regiões da terra, isso se deve a desuniformidade com que a energia solar atinge a superfície, o diferente

comportamento térmico dos continentes, a quantidade de CO₂ e ozônio na atmosfera, a variabilidade espacial de solos, cobertura vegetal, a influência de rotação e inclinação do eixo terrestre (SILVEIRA, 2012).

2.2 PRECIPITAÇÃO

2.2.1 Formas e Tipos

A importância do estudo da distribuição e dos modos de ocorrência da precipitação está no fato dela se constituir na principal entrada do balanço hídrico em uma dada região hidrológica. Segundo Torres e Machado (2011), a precipitação é o processo pelo qual a água condensada na atmosfera atinge a superfície terrestre na forma líquida (chuva ou chuveiro, ou orvalho) ou sólida (granizo, saraiva ou neve), apresentando as seguintes características:

A Neblina ou chuveiro é a precipitação de gotas de água muito pequenas, com diâmetros inferiores a 0,5 mm, que se dispersam uniformemente e tem baixa intensidade.

A chuva é um tipo de precipitação de partículas de água líquida na forma de gotas com diâmetro mínimo de 0,5 mm e velocidade de queda de 3 m/s.

O Granizo é a precipitação de grãos redondos ou cônicos de gelo que atingem tamanho grande (diâmetro ≥ 5 mm).

A Saraiva é a precipitação de pedras de gelo mais ou menos ovais, com diâmetro entre 0 mm e 50 mm ou mais.

O Orvalho é a precipitação de gotículas de água que se formam por haver a condensação do vapor de água do ar, dos objetos que se esfriam durante a noite.

A Geada é a deposição de cristais de gelo, fenômeno semelhante ao da formação de orvalho, mas que ocorre quando a temperatura é inferior a 0°C.

A Neve é a precipitação em forma de cristais de gelo que durante a queda coalescem formando flocos de dimensões variáveis.

Torres e Machado (2011), explicam que as chuvas podem ser classificadas em três tipos principais, de acordo com sua gênese: convectivas ou de convecção, frontais ou ciclônicas e orográficas ou de relevo.

a) Chuvas Convectivas

As nuvens de convecção (grandes cúmulos ou cúmulos-nimbos) são formados com a ascensão de uma massa de ar úmido em regiões quentes; são comuns em áreas quentes e úmidas. Com o aumento da concentração de vapor de água ou com o resfriamento da massa de ar (seja por causa da altitude, seja pela presença de ventos mais

frios), ocorre a saturação do ar, resultando em chuvas pesadas e intensas, embora de duração mais curta. Nas regiões equatoriais, onde há baixa pressão e a evaporação é constante e intensa em razão das elevadas temperaturas, ocorrem, comumente, chuvas de convecção, também provocadas pela ação dos ventos alísios oriundos das áreas de alta pressão nas latitudes de 30°.

b) Chuvas Frontais

Esse tipo de chuva, também chamado de ciclônica, está associado à instabilidade causada pelo encontro de duas massas de ar com características térmicas diferentes, uma massa de ar quente e outra de ar frio. É uma precipitação moderadamente intensa, contínua, que afeta áreas bastante extensas. São comuns nas áreas de médias latitudes, onde ocorre, normalmente (principalmente no período de inverno), o encontro de massas de ar com características oposta. Com o lento resfriamento do ar, ocorre a saturação e posterior condensação do vapor de água, como resultado, acontecem as chuvas frontais.

c) Chuvas Orográficas

As chuvas orográficas, orogênicas ou de relevo, ocorrem em razão da ascensão forçada dos ventos úmidos ante um obstáculo do relevo. O ar, obrigado a se elevar para transpor o obstáculo, resfria-se (com a altitude) e pode ficar saturado. As vertentes do obstáculo voltadas para o vento ficam cobertas de nuvens das quais cai a chuva.

2.2.2 Grandezas e Medição

As principais características da precipitação são: a Duração, a Intensidade, a Frequência, o Tempo de retorno; e a Altura pluviométrica.

A Duração (t) é o período de tempo durante o qual a chuva cai. As unidades normalmente utilizadas são o minuto ou a hora.

A Intensidade (i) é a precipitação por unidade de tempo, obtida pela razão entre a precipitação e o tempo ($i = P/t$). É expressa em mm/h ou mm/mim. A intensidade de uma precipitação apresenta variabilidade temporal e nas análises hidrológicas, geralmente são definidos intervalos de tempo.

A Frequência (f) está relacionada com a probabilidade de um fenômeno igual ou superior ao analisado, se apresentar em um certo intervalo de tempo.

O Tempo de retorno (Tr) é interpretado como o número médio de anos durante o qual espera-se que a precipitação analisada seja igualada ou superada. O seu inverso é a frequência e vice-versa, sendo muito utilizada no controle e prevenção de enchentes.

A Altura pluviométrica (P) está associada à espessura média da lâmina de água precipitada que recobriria a região atingida pela precipitação admitindo-se que não houvesse perdas para fora dos limites da região. A unidade de medida habitual é o milímetro de chuva, definido como a quantidade de precipitação correspondente ao volume de 1 litro por metro quadrado de superfície e esta pode ser determinada através de aparelhos conhecidos como pluviógrafos e pluviômetros. Na Figura 2, pode-se observar alguns aparelhos de medição de chuva.

Fonte: www.inmet.com.br

Figura 2 – Aparelhos de Medição de Chuva.



(a) Pluviômetro



(b) Pluviógrafo



(c) Pluviômetro digital

Fonte: www.inmet.com.br

O pluviômetro acumula a precipitação em um recipiente e necessita de leituras manuais em intervalos de tempo fixos. Já o pluviógrafo registra automaticamente, num suporte, os dados durante os intervalos de tempo.

As medidas realizadas nos pluviômetros são periódicas, feitas em geral em intervalos de 24 horas. O recipiente do pluviômetro deve apresentar um volume suficiente para conter as maiores precipitações dentro do intervalo de tempo definido para as observações. Acima desse recipiente é colocado um funil com um anel receptor biselado que define a área de interceptação. O anel deve ficar bem na horizontal. Pode-se determinar a altura da lâmina precipitada por meio da Equação 2.1.

$$P = 10 * Vol/A \quad (2.1)$$

Em que P é a precipitação acumulada em mm; Vol é o volume recolhido em cm^3 (ou mL); A é a área de interceptação do anel em cm^2 .

Atualmente, existem inúmeros aparelhos usados para medir e registrar continuamente as precipitações, os quais podem até transmitir os dados via satélite para uma fonte receptora, dispensando um observador diário na estação e agilizando o monitoramento hidrológico.

2.2.3 Fatores que influenciam na precipitação

Do ponto de vista da climatologia, a precipitação é classificada como um elemento climático, juntamente com a temperatura, a umidade do ar, a pressão atmosférica, ventos, nebulosidade, insolação e radiação solar. Sendo estes, os elementos principais que compõe o clima e o tempo, ou seja aqueles que se conjugam para formar o tempo atmosférico e o clima propriamente dito (TORRES & MACHADO, 2011).

Em climatologia e meteorologia, o tempo é considerado um estado momentâneo da atmosfera em determinado lugar, enquanto que o clima pode ser definido como a sucessão ou o conjunto de variações desses estados médios que vai caracterizar a atmosfera de um determinado local.

Todos os elementos climáticos, bem como a precipitação, sofrem alterações, provocadas pelos fatores climáticos ou geográficos. Esses fatores provocam mudanças por vezes bastantes significativas no clima.

No caso da precipitação, os fatores geográficos que influenciam na sua distribuição, são: a latitude, a altitude, a continentalidade, a disposição e orientação do relevo.

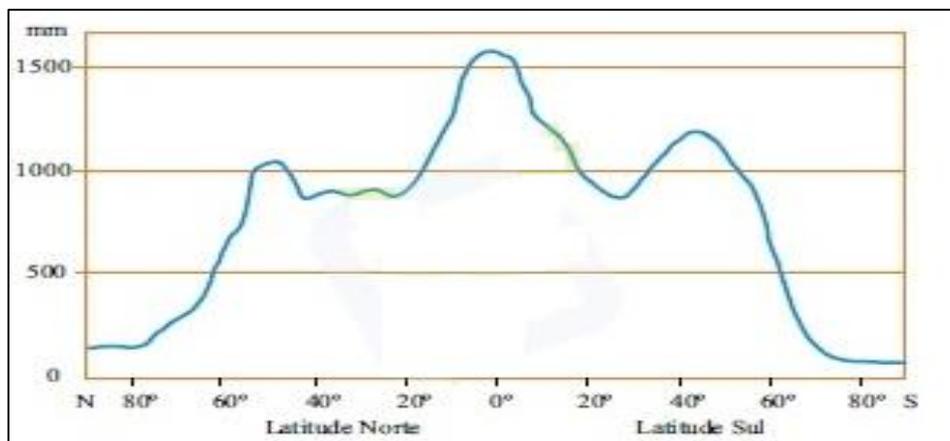
A latitude, que é a distância ao Equador medida ao longo do meridiano de Greenwich, influencia a distribuição desigual das pressões e temperaturas do globo e na circulação geral da atmosfera. Quanto a precipitação, verificam-se máximos valores de precipitação no equador (0°), valores médios nas latitudes 50° ao norte e sul e mínimos valores, nas latitudes em torno dos 30° (norte e sul) e nos polos (Figura 3).

A precipitação aumenta com a altitude. O ar, ao ascender ao longo de uma montanha arrefece e condensa, originando a formação de nuvens e precipitação. No entanto, a pluviosidade aumenta com a altitude até um certo ponto, passando a reduzir a partir daí. A altitude de máxima precipitação é em torno de 2500 m nos Alpes.

Em relação à influência da continentalidade, pode-se dizer que as regiões do litoral são, em geral, mais chuvosas do que as do interior. A proximidade de correntes marítimas quentes faz aumentar ainda mais a precipitação, pois a água em temperaturas mais

elevadas evapora com mais facilidade para a atmosfera, formando mais nuvens e assim aumenta a chance de precipitação (ROSS, 2005).

Figura 3 – Comportamento da Precipitação conforme latitude.

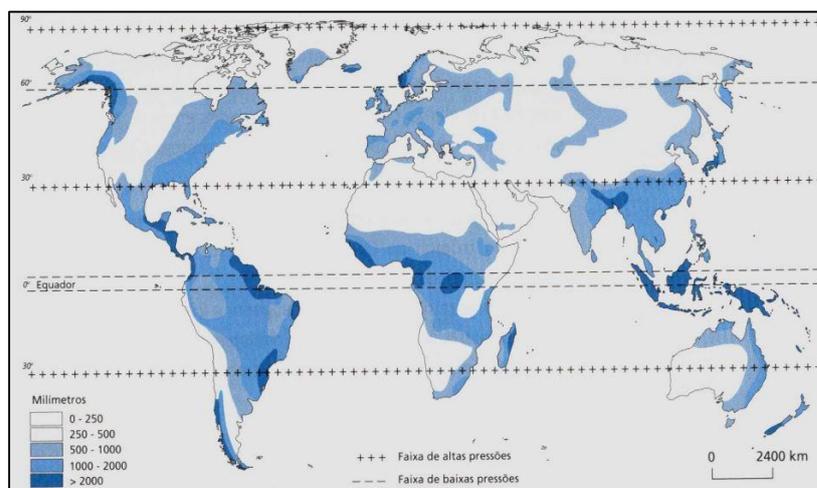


Fonte: Adaptado de: Ross (1995).

A disposição do relevo, além de estar associado à altitude, influencia na temperatura e umidade, ao facilitar ou dificultar a circulação de massas de ar, com isso pode facilitar ou impedir o processo de precipitação. Do ponto de vista geográfico, a precipitação é máxima na região equatorial e decresce com o aumento da latitude em direção aos polos.

Os elementos climáticos como a temperatura, a latitude, a umidade e a pressão, que são influenciados pela diversidade geográfica, manifestam-se por meio da precipitação, vento, nebulosidade, etc., de forma diferente em cada parte do globo, gerando distintos tipos, formas e alturas de precipitação (Figura 4).

Figura 4 – Variação Espacial da Precipitação média anual no mundo.

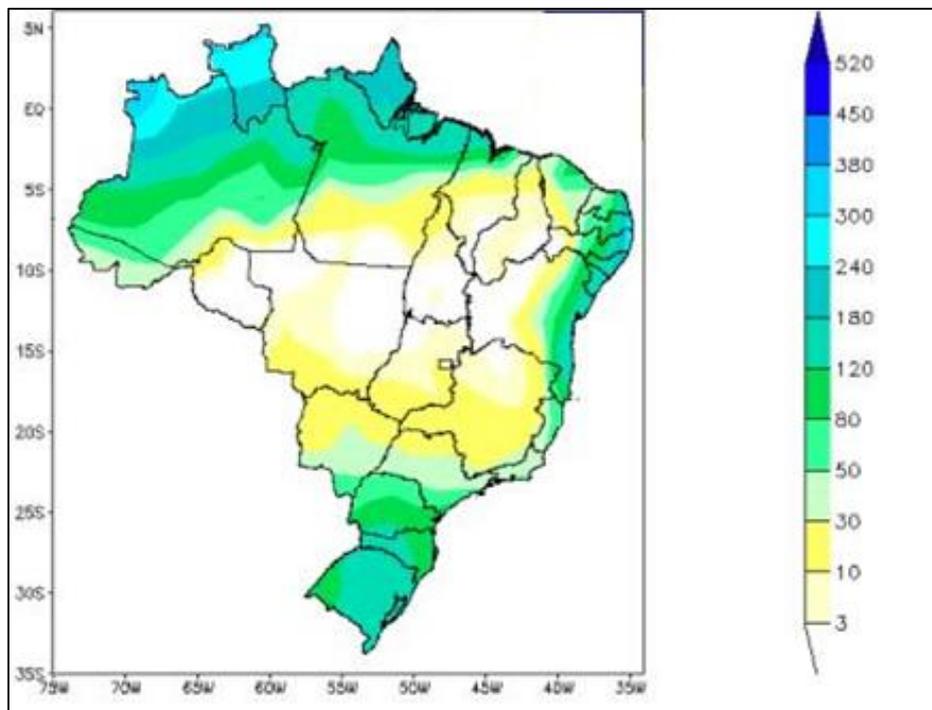


Fonte: Earth Fórum, 2008.

No Brasil, a distribuição e a variabilidade das chuvas estão associadas à atuação e à sazonalidade dos sistemas convectivos de macro e mesoescala e, em especial, da frente polar atlântica. A sazonalidade das chuvas mantém-se na região Centro-Oeste, embora seus valores sejam significativamente superiores aos do nordeste. Na região Sudeste e Sul, as chuvas são bem distribuídas (Figura 5) ao longo do ano, e na Amazônia, os índices pluviométricos são bem superiores das demais regiões do país (MENDONÇA, 2007).

Os fenômenos meteorológicos, também influenciam no regime de chuvas de uma região. Por exemplo, a região tropical é influenciada por diversos sistemas meteorológicos que atuam em diferentes escalas espaciais e que podem levar a ocorrência de precipitação como: a) Zona de Convergência Intertropical (ZCIT); b) Zona de Convergência do Atlântico Sul (ZCAS); c) Sistemas Frontais (SF); d) Linhas de Instabilidade; e) Oscilação Madden-Julian (OMJ). A atuação desses sistemas determinam os principais períodos chuvosos bem como a distribuição espacial da precipitação na região norte, nordeste e centro oeste, na qual, situa-se a RHTA (COHEN *et al.*, 1989; MOLION & BERNARDO, 2002).

Figura 5 - Precipitação mensal média de 1961-1990 no Brasil.



Fonte: INMET, 2015.

Existem também as anomalias climáticas (El Niño e La Niña), as quais têm provocado mudanças no regime de chuva de diversas regiões, conhecido como ENOS -

El Niño Oscilação Sul. Este representa um fenômeno de interação atmosfera-oceano, associado a alterações dos padrões normais de temperatura da superfície do mar e dos ventos alísios na região do Pacífico Equatorial, entre a costa peruana e o Pacífico Oeste, próximo a Austrália (Oliveira, 2001). O El Niño é caracterizado pelo aquecimento incomum das águas superficiais do pacífico, atingindo temperaturas acima da média térmica que é de 23°C. O La Niña, é representado pelo resfriamento atípico das águas do pacífico. Ainda segundo Oliveira (2001), o fenômeno El Niño, tem provocado diminuição das chuvas na região norte e aumento na região sul do Brasil.

Além dos fatores climáticos e dos eventos especiais, a intervenção antrópica pode provocar certas mudanças no clima, inclusive na precipitação de uma determinada região. Para Torres (2003), as alterações ambientais causadas por processos antrópicos tendem a produzir modificações em alguns elementos climáticos, originando fenômenos como o da “*Ilha de Calor*”, responsáveis por temperaturas mais elevadas na área central das cidades, além de pluviosidades quantitativamente maiores nas áreas urbanizadas.

Apesar da precipitação ter um comportamento aleatório, deve-se convir que algumas condições do clima e geografia da terra, podem contribuir para a sua distribuição. Assim, para estudos de precipitação, é preciso considerar os sistemas hídrico, climático, geológico, biológico e até mesmo a ação antrópica. É claro que um estudo dessa natureza se tornaria um tanto complexo, mas conhecer a influência de fatores geográficos, climáticos e antrópicos, é essencial para qualquer estudo desta variável.

2.3 REGIONALIZAÇÃO HIDROLÓGICA

Segundo Tucci (2002), o termo regionalização tem sido utilizado em hidrologia para denominar a transferência de informação de um local para outro dentro de uma área com comportamento hidrológico semelhante. Essa informação pode ocorrer em forma de uma variável, função, parâmetro ou indicador regional:

- a) *regionalização de variável*: quando uma variável pode ser determinada numa região com base em relações estabelecidas através dos dados pontuais existentes;
- b) *regionalização de função hidrológica*: quando a curva de uma função pode ser determinada em uma região com base em dados hidrológicos existentes;
- c) *regionalização de parâmetros*: quando o parâmetro de uma função ou de um modelo matemático pode ser determinado pela sua relação com características físicas das bacias; e
- d) *regionalização de indicadores regionais*: quando os indicadores regionais são definidos como valores médios de variáveis ou proporções que permitam uma rápida estimativa de uma variável ou entendimento do seu comportamento.

Dependendo do tipo de variável, função, parâmetro ou indicador, que se queira estimar, é necessário escolher o tipo de regionalização mais conveniente para representar a variável hidrológica estudada. Para a regionalização de chuvas, pode-se optar por modelos estatísticos e ou matemáticos, para estimativas das lâminas de chuvas.

O principal propósito da regionalização, segundo Franchini & Suppo (1996), é a transferência e troca de conhecimentos entre bacias hidrográficas, ou seja, informações hidrológicas de uma bacia monitorada podem ser transferidas para uma outra não monitorada, considerando a similaridade hidrológica existente entre elas.

Para Samuel *et al.* (2011), a regionalização é uma técnica utilizada para suprir a falta de informação hidrológica em locais com pouca ou nenhuma disponibilidade de dados. Assim, pode-se dizer que o objetivo da regionalização é justamente conseguir estimar dados de uma variável, função ou parâmetro, dentro de uma região homogênea, a partir de dados existentes dessa região.

Segundo Tucci (2012), na aplicação da regionalização é necessário que sejam realizadas as seguintes etapas: (1) definição dos limites da área a ser estudada; (2) determinação da variável dependente e possíveis variáveis explicativas; (3) seleção de dados e cálculo de variáveis; e (4) definição das áreas com comportamento hidrológico semelhante e estimação das relações regionais.

A regionalização pode ser aplicada na estimativa de vazões máximas e mínimas, em curvas de regularização de vazão e de permanência e também na estimativa de precipitações.

Na regionalização para estimativas de chuvas, diversos trabalhos foram desenvolvidos, em diversas partes do mundo, como: Satyanarayana & Sirvinas (2011), conseguiram identificar e regionalizar vinte e três grupos homogêneos de precipitação no território indiano; Zhang *et al.* (2012), através do método de regionalização de Análise frequência regional, conseguiu regionalizar cinco regiões homogêneas de chuva na bacia do Rio das Pérolas, no território Chinês; Parracho *et al.* (2015) identificaram quatro grupos regionais de precipitação, através do método de análise de agrupamento K-means, na Península Ibérica; Gonçalves *et al.* (2015) identificaram seis regiões homogêneas de chuvas, na Amazônia brasileira, especificamente no estado do Pará, por meio do método de análise de agrupamento via Ward.

A regionalização de precipitação tem amplos aproveitamentos, que incluem planejamento agrícola, auxílio no dimensionamento de sistemas de drenagem e infraestruturas hidráulicas e na estimativa de frequência e extensão da área de secas meteorológicas (SATYANARAVA & SIRVINAS, 2011).

É preciso considerar que os resultados advindos da regionalização, nunca irão substituir o monitoramento hidrológico com a observação feita em campo, mas as decisões feitas com base nesses resultados, serão sempre superiores àqueles executados por técnicas que se baseiam na subjetividade.

Para que a Regionalização seja aplicada, ou seja, para que ocorra a transferência de informação de um local para outro dentro de uma área com comportamento hidrológico semelhante, é necessário, inicialmente, identificar as regiões hidrologicamente homogêneas. Assim, a identificação de regiões homogêneas é uma das etapas mais importante da Regionalização e merece devida atenção.

2.3.1 Regiões Homogêneas

Uma Região hidrologicamente homogênea é entendida como aquela que possui similaridade hidrológica. Para Lanna (2004), essa similaridade inclui fatores físicos, climáticos, biológicos, geológicos e efeitos antrópicos. Segundo Tucci (2002), na regionalização hidrológica, a homogeneidade é entendida como a semelhança na resposta das funções hidrológicas, e o ideal é reconhecer a variabilidade das condições hidrológicas e determinar as variáveis físicas da bacia hidrográfica. Para Naghettini &

Pinto (2007), do ponto de vista estatístico, uma região é homogênea se existem evidências suficientes de que as diferentes amostras do grupo possuem as mesmas distribuições de frequência.

Pode-se dizer que a identificação de regiões hidrologicamente homogêneas possui duas finalidades: uma descritiva, cujo objetivo é a caracterização hidrológica, e outra como subsídio essencial na regionalização.

A definição de regiões homogêneas, além de auxiliar em estudos de regionalização, contribui também para o entendimento do comportamento hidrológico de uma bacia hidrográfica, tendo em vista que a formação dessas regiões depende de variáveis que explicam os fenômenos hidrológicos.

Existem diversos métodos de identificação de regiões homogêneas. De acordo com Hosking & Wallis (1997), os principais métodos de formação de regiões homogêneas são:

- Convergência geográfica

São métodos de identificação de regiões homogêneas, baseados na subjetividade e ou conveniência dos postos de observação, geralmente contíguos, em áreas administrativas, definidas segundo limites arbitrários.

- Agrupamento subjetivo

Nessa categoria, a delimitação das regiões é feita por agrupamento dos postos de observação em conformidade com a similaridade de algumas características locais, tais como classificação climática, relevo ou conformação das isoietas anuais.

-Agrupamento objetivo

Nesse caso, as regiões são formadas pelo agrupamento dos postos de observação em um ou mais conjuntos, de maneira que uma dada estatística não exceda um valor limiar previamente estabelecido.

-Análise de Clusters

Trata-se do método de análise estatística multivariada, na qual se associa a cada posto um vetor de dados contendo as características locais. Em seguida, os postos são agrupados e reagrupados, de forma que seja possível identificar a maior e a menor similaridade entre os vetores de dados.

Além dos métodos propostos anteriormente, existem também os métodos de análise de resíduo, análise fatorial, correlação canônica e métodos baseados no geoprocessamento e redes neurais.

A aplicação da Análise de agrupamento pelo método Fuzzy c-means, tem sido utilizado com frequência na identificação de regiões hidrologicamente homogêneas.

Satyarayana e Sirvinas (2011) apresentaram uma abordagem baseada na análise de agrupamento Fuzzy c-means através da qual foi possível identificar regiões homogêneas de chuva, empregando um estudo de regionalização de precipitação em áreas esparsas de dados, usando variáveis atmosféricas de larga escala, atributos de localização e sazonalidade de chuvas.

Goyal & Gupta (2014) compararam os métodos fuzzy c-means e o k-means para definir regiões homogêneas de chuva nas bacias hidrográficas da região nordeste da Índia. Eles concluíram que as regiões definidas pelo método fuzzy c-means, são suficientemente homogêneas para análise de frequência regional.

Gomes *et al.* (2014) aplicaram a técnica de análise de agrupamento fuzzy c-means para classificar estações fluviométricas em grupos homogêneos, utilizando variáveis explicativas de vazão. Por meio do método, foi possível definir quatro regiões consideradas hidrologicamente homogêneas de vazão no Estado do Pará.

Pessoa (2015) desenvolveu uma metodologia de estimativa de curvas de permanência de vazão na Amazônia Legal, onde identificou 12 regiões homogêneas de vazão na região em estudo, através do agrupamento fuzzy c-means, adotando variáveis físico climáticas da região.

2.4 MODELOS DE ESTIMATIVAS DE CHUVA

Os modelos mais usuais de estimativas de lâminas de chuva são os métodos de interpolação espacial, modelos estatísticos e os de estimativa por satélite.

Dentre os modelos de interpolação espacial, pode-se citar o do polígono de Thiessen e método das isoietas. Entre os modelos estatísticos, destacam-se a distribuição de probabilidades e a análise de regressão, sendo estes os mais aplicados em estudos de estimativas de chuva.

Loureiro *et al.* (2015), por exemplo, formulou mapas de isoietas de chuvas com a magnitude do volume precipitado por ano na RHTA. Ishihara *et al.* (2014) produziu mapas de isoietas, através de interpolação espacial na Amazônia Legal.

Existem também, os modelos de estimativas de chuvas por satélite. Os satélites encontram-se espalhados na órbita terrestre e são capazes de estimar as lâminas de chuvas, em diversas resoluções temporais e espaciais. Na Tabela 1, alguns exemplos de satélites usados para estimar chuvas.

Tabela 1 – Características dos principais satélites usados em estimativas de chuva.

Satélite	Tipo de órbita	Sensores	Operadora	Abrangência
GOES	Geoestacionária	VISSR	NOAA	1/3 do globo
NOAA	Polar	AVHRR	NASA	Global
GMS	Geoestacionária	VISSR	AMJ	Hemisfério Norte

Fonte: Adaptado de www.sat.cpnm.embrapa.br.

2.4.1 Funções de Distribuição de Probabilidade – FDP

Os modelos ou as funções de probabilidade são utilizados para estimar a probabilidade de ocorrência de chuva em uma determinada escala temporal (diária, mensal, anual, etc.) e verificar o comportamento e a variabilidade das chuvas.

Para Lyra *et al.* (2006), várias funções de distribuição de probabilidade têm sido utilizadas para verificar o comportamento e a variabilidade das chuvas ao longo dos anos, dentre elas estão as distribuições: Normal, Gama, Gama mista, Gumbel, Exponencial, Log-normal, Weibull e Pearson.

No trabalho de Silva *et al.* (2007), aplicaram-se as funções de distribuição de probabilidade Gama, Weibull, Normal, Log-normal e Exponencial para analisar a distribuição da probabilidade de ocorrência de precipitação média diária, durante os meses do ano, em Santa Maria-RS. Nessa observação, as funções Gama e Weibull, foram as que melhor descreveram a variação da probabilidade de ocorrência de precipitação diária, na referida pesquisa.

O trabalho de Oliveira *et al.* (2011), aplicou a função de probabilidade Gumbel, para estimar chuvas intensas para diferentes períodos de retorno, no estado do Mato Grosso. Os resultados provaram que a função Gumbel estima de forma adequada as chuvas intensas naquela região.

2.4.2 Modelos Regionais - Regressão Múltipla

A regressão múltipla é usada para verificar a relação entre uma única variável dependente e várias independentes, ou seja, ela utiliza-se das variáveis independentes cujos valores são conhecidos para prever os valores da variável dependente selecionada.

Das diversas aplicações da regressão múltipla a mais usual em hidrologia, é a estimativa de variáveis hidrológicas, como a vazão e a precipitação. Para estimar valores de precipitação, alguns trabalhos têm sido desenvolvidos.

Costa *et al.* (2012) utilizaram a regressão linear múltipla para a geração de uma superfície interpolada com os valores de precipitação máxima média para a Bacia Hidrográfica do Rio Doce, com período de retorno variando em 10, 25 e 50 anos e duração de precipitação definida em 30 min, com base nos parâmetros regionalizados por FREITAS *et al.* (2001) para as estações pluviométricas de Minas Gerais.

Cruz *et al.* (2013) propuseram uma equação através de regressão linear múltipla que correlacionou os valores anuais de precipitação e fatores de localização espacial para 58 postos pluviométricos no estado de Sergipe. Neste trabalho, obteve-se uma equação com coeficiente de determinação de 0,844, considerado bom, com erros médios percentuais entre -7,5% a 5,5%.

Mekanik *et al.* (2013) realizaram estimativas de chuvas de primavera sazonal de longo prazo em Victoria, na Austrália, utilizando os métodos de Redes Neurais Artificiais (RNA) e análise de regressão múltipla (RM). Os autores pesquisaram, também, o defasado El Niño Oscilação Sul (ENOS) e Dipolo do Oceano Índico (IOD) como potenciais preditores. A RNA e a MR foram avaliados estatisticamente por meio do erro quadrático, erro médio absoluto, correlação de Pearson (r) e índice de concordância.

3 ÁREA DE ESTUDO

A Região Hidrográfica Tocantins-Araguaia – RHTA, destaca-se por ser a segunda maior do país em termos de área e vazão, inferior apenas à do Amazonas, e a maior do país com área de drenagem situada integralmente em território nacional. As dimensões equivalem a 1,5 vezes a bacia do Rio São Francisco e a vazão média, é de 13.799 m³/s (8% do total do país) (ANA, 2009).

A RHTA possui uma área de 918.822 km², disposta na porção central do Brasil e abrange os Estados de Goiás, Tocantins, Pará, Maranhão, Mato Grosso e o Distrito Federal. Parte considerável dela situa-se na Região Centro-Oeste, desde as nascentes dos Rios Araguaia e Tocantins até a sua confluência, e daí, para jusante, adentra na Região Norte até sua foz.

Atualmente a região destaca-se nas atividades agropecuárias e no fornecimento de energia elétrica, sendo o potencial hidrelétrico da região de 23.825 MW. Assim, a localização, a abundância e a sua potencialidade hídrica, agropecuária, mineral, para navegação e geração de energia, conferem à região um relevante papel no desenvolvimento do Brasil.

3.1 LOCALIZAÇÃO

A Região Hidrográfica Tocantins-Araguaia (RHTA) localiza-se entre o paralelo sul 0° 30' e 18° 05' e os meridianos de longitude oeste 45° 45' e 56° 20'. Sua configuração é alongada, com sentido Sul-Norte, seguindo a direção predominante dos cursos d'água principais, os rios Tocantins e o Araguaia, que se unem na parte setentrional da região, a partir de onde é denominado apenas de Rio Tocantins, que segue até desaguar na Baía do Marajó.

A área total de drenagem da RHTA (Figura 6), abrange parte das regiões Centro-Oeste, Norte e Nordeste. Ocupa 11% do território nacional, incluindo as áreas dos Estados de Goiás (21,4% da RHTA), Tocantins (30,2%), Pará (30,3%), Maranhão (3,3%), Mato Grosso (14,7%) e o Distrito Federal (0,1%), totalizando 409 municípios e integrando totalmente o estado do Tocantins.

O Conselho Nacional de Recursos Hídricos instituiu a Divisão Hidrográfica Nacional, segundo a Resolução nº 32, de 15 de outubro de 2003, dividindo o território brasileiro em 12 regiões hidrográficas (Nível um). Para fins de planejamento e com a finalidade de estudar, detalhar e caracterizar as regiões hidrográficas, as 12 regiões foram divididas em 83 unidades (Nível 2) e 332 unidades hidrográficas de referência (Nível 3).

3.2 CARACTERÍSTICAS CLIMÁTICAS

A RHTA possui clima tropical, com temperatura média anual de 26° C, e dois períodos climáticos bem definidos: o chuvoso, de outubro a abril, com mais de 90% de precipitação, com a existência de alguns dias secos entre janeiro e fevereiro, formando o chamado veranico; e o seco, de maio a setembro, com baixa umidade relativa do ar (ANA/MMA, 2006).

A precipitação média na região é de 1.744 mm com os totais pluviométricos crescendo de Sul para o Norte – valores próximos a 1.500 mm (Brasília) a 3.000 mm (Belém) – e decrescendo no sentido de Oeste para Leste – valores da ordem de 1.800 mm a 1.200 mm (ANA, 2009).

O regime pluviométrico é caracterizado pela ocorrência de aumento das precipitações com a diminuição da latitude (sentido sul-norte). A zona de menor precipitação ocorre na faixa a oeste do município de Paranã (GO), na divisa com a Região Nordeste, com valores médios de 1mm/mês entre junho e agosto (ANA/MMA, 2006). No extremo norte o período seco reduz-se a três meses do ano (setembro, outubro e novembro) e no restante da região atinge entre cinco e seis meses (maio a outubro).

3.3 HIDROGRAFIA E VEGETAÇÃO

A RHTA tem como vias hidrográficas principais, o rio Tocantins e o rio Araguaia. O rio Tocantins nasce no Planalto de Goiás, a cerca de 1.000 m de altitude, sendo formado pelos rios das Almas e Maranhão, e com extensão total aproximada de 1.960 km até a sua foz no Oceano Atlântico. Seu principal tributário é o rio Araguaia (2.600 km de extensão), onde se encontra a Ilha do Bananal, maior ilha fluvial do mundo (350 km de comprimento e 80 km de largura). Na margem direita do Tocantins destacam-se os rios Bagagem, Tocantinszinho, Paranã, dos Sonos, Manoel Alves Grande e Farinha; na margem esquerda, os rios Santa Teresa e Itacaiunas (ANA/MMA, 2006).

Segundo Sano *et al.* (2007) e Brun *et al.* (2013), na RHTA há a predominância dos biomas Amazônia e Cerrado. O bioma Amazônia é formado pelos ecossistemas de floresta ombrófila densa, floresta ombrófila aberta, floresta estacional decidual e semidecidual, campinarana, formações pioneiras, refúgios montanos, savana amazônica, matas de terra firme, matas de várzeas e matas de igapó. Na RHTA, nesse bioma, o ecossistema de maior expressão é o da floresta ombrófila densa. No Bioma de cerrado, predominam as formações fisionômicas que são o cerradão, o cerrado típico, o campo

cerrado, o campo sujo cerrado e o campo limpo, com características de clima quente e úmido megatérmico, com chuvas de verão e estação seca no inverno (ANA, 2009).

Segundo ainda, ANA (2009), a RHTA apresenta duas zonas de transição entre os biomas, os chamados ecótonos. Nessas áreas, normalmente, verifica-se uma fauna mais diversificada e abundante do que nos biomas, o denominado “efeito de borda”.

3.4 USO E OCUPAÇÃO DO SOLO

A ocupação da RHTA sofreu um grande impulso com a abertura da rodovia Belém-Brasília e, na década de 60, com a construção de Brasília, que resultaram na proliferação de novos municípios em volta da capital federal e ao longo do eixo viário que passou a atravessar a região no sentido norte-sul. Na década de 70 e 80, a construção da Transamazônica e a criação de programas de integração, atraíram mais habitantes para a região.

Dentre os diversos usos do solo da região, destacam-se os projetos de irrigação, mineração, garimpos, os aproveitamentos hidrelétricos, processamento de produtos agropecuários como a soja, o arroz, o milho e a carne bovina. Atualmente, o principal eixo industrial da Região Hidrográfica do Tocantins Araguaia localiza-se na área de influência do complexo mineral da Serra do Carajás, onde se encontram os principais polos e projetos industriais nos setores mineral e madeireiro (ANA, 2009).

O aspecto crítico do processo de ocupação do bioma Amazônico na região é o desmatamento descontrolado. Os principais eixos de desmatamento situam-se no norte da RHTA, desde a região sudoeste da cidade de Marabá até Belém, onde ocorre a floresta ombrófila densa. Também apresentam expressões os eixos de desmatamentos que ocorrem em algumas porções isoladas situadas mais ao sul, até próximo à Conceição do Araguaia, em áreas de transição para o Cerrado (ANA, 2009).

A região possui Unidades de Conservação, correspondentes a territórios constituídos legalmente pelo poder público com o objetivo de proteção e conservação dos recursos naturais. Na RHTA, totalizam 82.321 km², que correspondem a 9% da área total. A região abriga também, 25 distintas etnias em 53 terras indígenas, distribuídas no estado de Goiás (7), no estado do Mato Grosso (16), no estado do Tocantins (8), no estado do Pará (19) e no estado do Maranhão (2) (ANA, 2009).

3.5 ASPECTOS ECONÔMICOS

A RHTA concentra diversas atividades que alavancam a economia da região. Destacam-se, agricultura, pecuária, mineração e produção de energia elétrica.

A pecuária, com base na criação de bovinos, é uma atividade que vem crescendo ao longo dos anos. Segundo dados do IBGE (2014), só no estado do Tocantins e no município de Marabá – PA, no ano de 2014, o rebanho bovino somado, estava em aproximadamente 1.122.700 cabeças. Essa atividade vem se destacando principalmente no centro norte da RHTA (leste e sudeste do Pará, norte do Tocantins e oeste do Maranhão) e o no sul (centro oeste de Goiás) com ligações adensadas entre elas, principalmente ao longo da rodovia Belém-Brasília. Outros tipos de atividades ligadas a agropecuárias, como a silvicultura, também são desenvolvidas na região.

A indústria é outra atividade que se destaca na região, devido à existência do Complexo Alunorte-Albrás, em Barcarena-PA, o Projeto Carajás, em Parauapebas-PA, as usinas hidrelétricas de Tucuruí-PA, Serra da Mesa em Minaçu-GO, Cana Brava em Cavalcante-GO e Lajeado em Lajeado-GO; a cadeia industrial do agronegócio (abatedouros, frigoríficos, laticínios, madeiras e a usina de biodiesel em Porto Nacional, TO), o polo de ferro gusa de Marabá e Açailândia-PA, além de indústrias minerais diversas (ANA, 2009).

Por toda sua dinâmica sócia econômica, a região apresenta expressivo desenvolvimento, nos diversos setores da economia, tornando a região, ainda mais atrativa à imigração.

3.6 BANCO DE DADOS

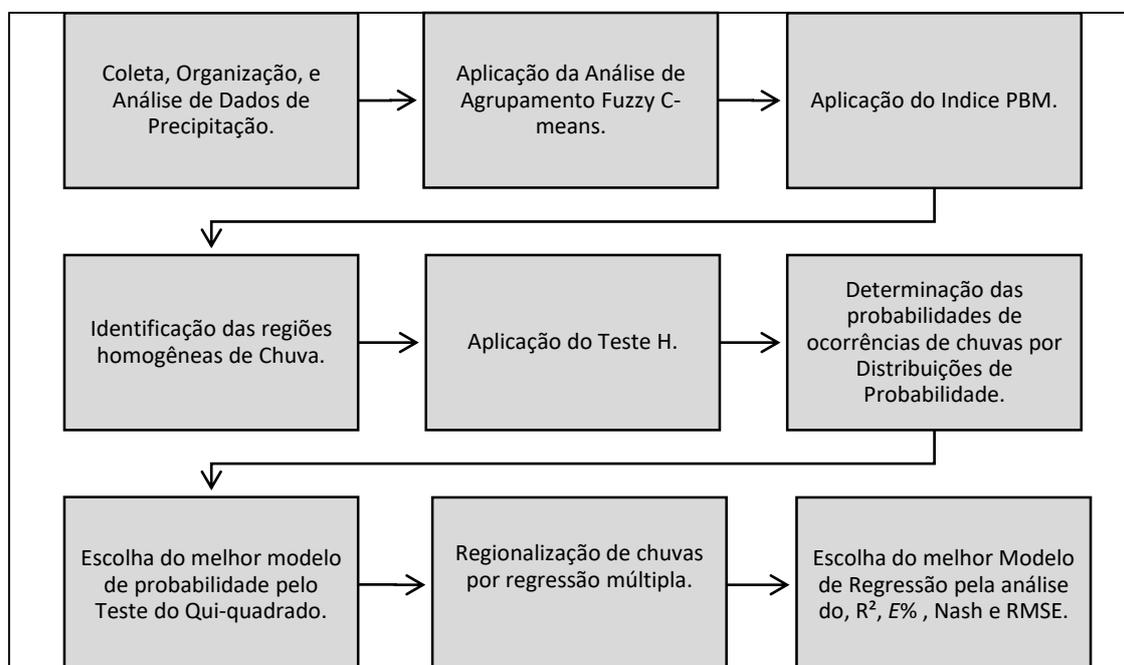
Foram adotadas as séries históricas de chuvas de 83 estações pluviométricas do banco de dados da Agência Nacional de Águas - ANA, localizadas na RHTA (Figura 8). Essas séries de precipitação foram organizadas em um banco de dados e calculou-se a precipitação média anual e mensal de cada estação. As informações sobre a altitude e a localização, foram extraídas, também, do banco de dados da ANA. Foi utilizado a precipitação anual média, a altitude e a localização (latitude e longitude) de cada posto pluviométrico, para aplicar a análise de agrupamento e assim identificar as regiões homogêneas de chuva. A Tabela 2 apresenta algumas estações usadas neste estudo, sendo que a lista completa das estações encontra-se disponível no Apêndice A. As estações foram escolhidas com base na série histórica de dados e optou-se pelas estações que possuíam uma maior série dados consistidos e sem falhas na observação. Das 83 estações

4 METODOLOGIA

A metodologia adotada neste trabalho pretende, primeiramente, classificar ou agrupar as estações pluviométricas em grupos homogêneos de chuva baseado na técnica de análise de agrupamento por meio do método Fuzzy c-means. Este método serviu para identificar os grupos homogêneos de chuvas na RHTA, adotando características regionais, de cada posto pluviométrico, como a altitude, precipitação média anual, posição geográfica (latitude e longitude) e parâmetros específicos do método Fuzzy c-means, como a medida de similaridade (euclidiana), o número de grupos e o parâmetro de fuzificação.

Toda a metodologia adotada neste estudo é apresentada, com mais detalhes, nos tópicos seguintes e no esquema da Figura 9.

Figura 9 – Etapas da Metodologia.



A medida de similaridade, adotada pela distância euclidiana, fornece a similaridade existente entre os elementos envolvidos e entre os agrupamentos, e assim auxilia no processo de formação dos grupos homogêneos de precipitação.

O Índice PBM consiste em selecionar o melhor número de agrupamento, obtido pelo método Fuzzy c-means, adotando o critério de maximização do referido índice e assim tornar a escolha do número de grupos homogêneos sem subjetividade.

O método dos momentos-L (MML) é uma metodologia estatística, baseada na teoria dos momentos ponderados por probabilidade (MPP), introduzida por Greenwood *et al.* (1979), para determinar os momentos-L de uma distribuição de dados

(precipitações), através de combinações lineares que fornecem como resultado final, os quocientes de variação, assimetria e curtose, estes por sua vez, foram adotados na aplicação do teste de heterogeneidade H.

O teste de heterogeneidade H tem a finalidade de verificar o grau de heterogeneidade de uma região, baseado no cálculo da dispersão dos quocientes de variação e assimetria. Portanto, este teste, serve para indicar se uma região dita homogênea pode ser considerada como tal, indicando possíveis erros e pontos atípicos em uma série de observações.

As funções de distribuição de probabilidade foram utilizadas para estimar a probabilidade de ocorrência de lâminas de chuva em frequência mensal e anual e verificar o comportamento e a variabilidade das chuvas nas regiões homogêneas definidas.

O emprego da regressão múltipla determinou as estimativas de lâminas de chuvas, através dos modelos de regressão múltiplas proposto para cada região homogênea.

4.1 ANÁLISE DE AGRUPAMENTO

Análise de agrupamento é o nome dado para um grupo de técnicas multivariadas, cuja finalidade primária é agregar objetos com base nas características que eles possuem. A análise de agrupamento classifica objetos de modo que cada objeto é muito semelhante aos outros no agrupamento em relação a algum critério de seleção predeterminado. Os agrupamentos resultantes de objetos devem então exibir elevada homogeneidade interna (dentro dos agrupamentos) e elevada heterogeneidade externa (entre grupos). Assim, se a classificação for bem sucedida, os objetos dentro dos agrupamentos estarão próximos quando representados graficamente e diferentes agrupamentos estarão distantes (HAIR *et al.*, 2005).

As técnicas de agrupamento ou cluster são frequentemente classificadas em dois tipos: técnicas hierárquicas e não hierárquicas. As técnicas hierárquicas são utilizadas em análises exploratórias dos dados com o intuito de identificar possíveis agrupamentos e o valor provável do número de grupos. No uso de técnicas não hierárquicas, é necessário que o valor do número de grupos já esteja pré especificado pelo operador.

Os métodos hierárquicos podem ser aglomerativos ou divisivos e se caracterizam por estabelecerem um diagrama de árvore ou dendograma composto pela sucessão hierárquica dos n grupos formados, tal que n , representa o número total de indivíduos do conjunto de dados.

Na abordagem aglomerativa, o processo inicia-se com n grupos unitários formados por cada um dos elementos e, a cada iteração, os dois grupos mais próximos unem-se até que todos os indivíduos constituam um único grupo. Como exemplo de técnicas hierárquicas, cita-se o método da Ligação Simples (Single Linkage), o método da Ligação Completa (Complete Linkage) e o método de Ward (MINGOTI, 2005).

Diferentemente dos métodos hierárquicos, os procedimentos não hierárquicos, não envolvem o processo de construção em árvore. Em vez disso, designam objetos a agrupamentos, assim que o número de agregados a serem formados tenha sido especificado. Os métodos K-Means, Redes Neurais e o Fuzzy C-means são alguns exemplos de métodos não hierárquicos.

O método Fuzzy C-means pode ser executado para diferentes números de grupos, com a vantagem de sua versatilidade no uso de inúmeras características dos elementos a serem agrupados, e dos parâmetros, específicos da metodologia, que permitem várias opções de agrupamento, tornando-o mais flexível e subjetivo na formação dos grupos.

Para Malhotra (2006), três questões fundamentais devem ser consideradas na aplicação da análise de agrupamento: primeira, como será medida a similaridade dos dados; segunda, como formar os agrupamentos; e por fim como decidir quantos grupos formar.

4.2 FUZZY C-MEANS

Para a determinação das regiões hidrologicamente homogêneas de chuva, foi utilizado o método de análise de agrupamento Fuzzy c-means, do tipo não hierárquico proposto por DUNN (1973) e depois generalizado por BEZDEK (1981). Conhecido como agrupamento nebuloso, “*fuzzy clustering*”, caracterizado pela ideia básica de que um conjunto $X = \{x_1, x_2, \dots, x_n\}$ seja dividido em p grupos, e o resultado do agrupamento seja expresso pelos graus de pertinência, de modo que cada elemento passe a pertencer a um ou mais conjuntos, diferente da configuração binária da teoria clássica dos conjuntos.

Este tipo de agrupamento é realizado através de um algoritmo conhecido como FCM, que corresponde a abreviatura de fuzzy c-means. Este tem como objetivo encontrar grupos *fuzzy* (baseados no grau de pertinência) para um conjunto de dados. Para alcançar este objetivo, o algoritmo precisa minimizar uma função que diz respeito à minimização das distâncias entre os dados e os centros dos grupos aos quais tais dados pertencem com algum grau de pertinência (XU & WUNSCH, 2005).

No agrupamento fuzzy c-means, a partição é gerada através da minimização de uma função, equacionada por meio de um algoritmo iterativo, o qual indica o grau de pertinência de um elemento pertencer a um determinado cluster (grupo). Portanto, é uma técnica em que cada elemento pertence a um grupo com um determinado grau de pertinência e requer a pré especificação do número de grupos que se deseja formar.

O agrupamento Fuzzy c-means procura a partição que minimiza a função objetivo (Equação 4.1).

$$J = \sum_{i=1}^n \sum_{j=1}^p (u_{ij})^m d(X_i, C_j)^2 \quad (4.1)$$

Em que:

n - o número de dados;

p - o número de grupos;

u_{ij} - o grau de pertinência da amostra X_i ao j -ésimo cluster;

m - o parâmetro de fuzificação;

d - a distância euclidiana entre X_i, C_j ;

X_i - vetor de dados, em que $i=1, 2, \dots, n$, representa um atributo de dados;

C_j - é o centro de um agrupamento Fuzzy.

A função objetivo J é minimizada e os graus de pertinência u_{ij} são gerados conforme a Equação 4.2.

$$u_{ij} = \left[\sum_{k=1}^c \left(\frac{d(X_i, C_k)}{d(X_i, C_j)} \right)^{2/(m-1)} \right]^{-1} \quad (4.2)$$

Em que C_j , chamado de centroide, pode ser obtido pela Equação 4.3.

$$C_j = \frac{\sum_{i=1}^n (u_{ij})^m X_i}{\sum_{i=1}^n (u_{ij})^m} \quad (4.3)$$

Os graus de pertinência, que representam as probabilidades, são gerados de uma distribuição uniforme no intervalo $[0,1]$. Os centroides vão se modificando a cada iteração e o algoritmo é interrompido, através de duas situações:

1º) Quando a distância entre os centroides de uma iteração em relação à anterior é menor ou igual a certo valor de erro. Em que: V_t e V_{t+l} são os vetores que guardam os centroides nas iterações t (iteração atual) e $t+l$ (iteração anterior), respectivamente.

2º) Quando o número máximo de iterações, pré-fixado, for executado. Então, repete-se o processo.

O método fuzzy c-means é equacionado por meio de um algoritmo iterativo, baseado na minimização de um índice de desempenho, atribuído à função objetivo, que indica a adequabilidade da repartição gerada. Segundo BEZDEK (1992), o desempenho do algoritmo é influenciado pela escolha do número de grupos c , dos centros de clusters (grupos) iniciais, da ordem na qual os vetores de dados são processados, da medida de distância, do critério de parada e do parâmetro de fuzificação.

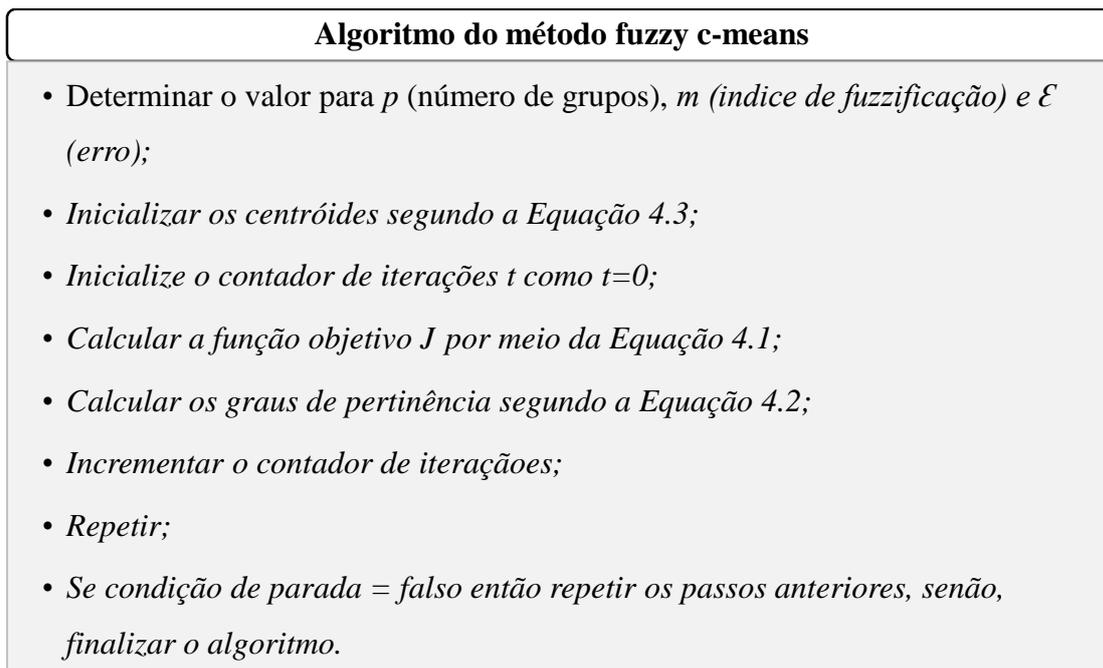
O parâmetro de fuzificação (m) é conhecido, também, como o expoente de peso Fuzzy. Para HALL e MINS (1999), o índice de fuzificação (m) é um parâmetro que controla o nível de difusividade no processo de classificação. Desta forma, para $m = 1$, os agrupamentos apresentam limites rígidos equivalentes aos do K-means e, à medida que o valor cresce, os limites tornam-se mais difusos. Segundo COX (2005), usualmente, m está no intervalo de 1,25 a 2,0.

Em linguagem de programação, o algoritmo FCM (Apêndice B), funciona seguindo rotinas pré estabelecidas que permitem sua inicialização, seu processamento, geração de resultados e a sua parada. A estrutura do algoritmo Fuzzy c-means é apresentada na Figura 10, conforme, proposto por BEZDEK (1992).

A decisão do agrupamento é definida pelo maior grau de pertinência apresentado para cada elemento analisado. Assim para cada elemento, o seu maior grau de pertinência u_{ij} , determinará a qual agrupamento este X_i pertencerá. Desta forma, todos os dados serão agrupados evitando equívocos e rigidez na formação dos agrupamentos.

Segundo Ferreira (2012), uma das vantagens de se trabalhar com o método Fuzzy c-means, é que se consegue melhorar o detalhamento e a descrição dos dados, permitindo uma concisa representação da variável ou sistema estudado.

Figura 10 - Estrutura do algoritmo Fuzzy c-means.



Fonte: Bezdek, (1992).

4.3 MEDIDAS DE SIMILARIDADE

Normalmente, os algoritmos que executam tarefas de dados, utilizam alguma medida de similaridade entre os vetores em seu processo de execução. Essas medidas servem para guiar o processo de construção da superfície de decisão que determinará qual é a região de abrangência de uma classe de dados no caso da tarefa de classificação, ou quais dados pertencem aos grupos, no caso da tarefa de agrupamento.

Para Hair *et al.* (2005), as características de cada objeto são combinadas em uma medida de semelhança, que pode ser de similaridade ou dissimilaridade, calculada para todos os pares de objetos, possibilitando a comparação de qualquer objeto com outro pela medida de similaridade e associação dos objetos semelhantes por meio da análise de agrupamento.

Para Rocha *et al.* (2012), a similaridade entre dois vetores corresponde a uma medida que compara a igualdade dos mesmos, podendo também ser utilizada como uma forma de medir a distinção entre eles. Normalmente, a similaridade é medida com base em alguma medida de distância entre dois vetores.

Existem diferentes medidas de distâncias que podem ser aplicadas nos processos de análise de dados, como Distância Euclidiana, Distância de Hamming, Distância de Chebyshev, Distância City-Block, Métrica de Tanimoto, Métrica de Mahalanobis etc.

Para avaliar a similaridade entre as estações a serem agrupadas, neste trabalho, utilizou-se a medida de distância Euclidiana (Equação 4.4). Ela mede a distância geométrica determinada por um segmento de reta entre dois objetos. Para MINGOTI (2005), é a medida mais comumente usada de similaridade, principalmente, porque se aplica muito bem em variáveis quantitativas.

$$d_{ii'} = \sqrt{\sum_{j=1}^p (Z_{ij} - Z_{i'j})^2} \quad (4.4)$$

Em que Z_{ij} é a variável padronizada; e $d_{ii'}$ é a distância euclidiana.

Na análise de agrupamento, a distância euclidiana fornece a similaridade existente entre os elementos envolvidos e entre os agrupamentos, e assim auxilia no processo de formação dos grupos homogêneos de precipitação.

4.4 ÍNDICE DE VALIDAÇÃO PBM

Um dos grandes questionamentos na análise de agrupamento, trata-se da validação dos grupos formados. Para que se possa alcançar um bom resultado, é necessário avaliar qual a partição mais adequada para os dados e se a partição gerada pelo algoritmo é de boa qualidade. Para explicar essas questões, existem na literatura diversos índices de validação, os quais se baseiam na função objetivo e nos dados de cada agrupamento. Como exemplo, podem ser citados:

- Índice de validação de Partição de Coeficiente (PC), proposto por BEZDEK (1981);
- Entropia de Partição (PE), proposto por BEZDEK (1981);
- Expoente Proporcional, proposto por WINDHAN (1981);
- Índice VFS, proposto por FUKWYAMA E SUGENO (1989);
- Índice VXB, proposto por XIE e BENI (1991); e
- Índice PBM, proposto por PAKHIRA *et al.* (2004).

No presente estudo será adotado o índice PBM para validação dos grupos, devido seu julgamento considerar tanto as distancias entre os grupos formados, como também, as distancias dos elementos aos centros dos grupos formados, tornando a validação mais consistente.

O índice PBM é usado para avaliar o número de agrupamentos ou subconjuntos formados a partir de um conjunto de dados. É definido como o produto de três fatores

(Equação 4.5), dos quais a maximização assegura que a partição tem um pequeno número de grupos compactos com grande separação entre pelo menos dois deles PAKHIRA *et al.* (2004).

$$PBM(k) = \left(\frac{1}{k} \cdot \frac{E_1}{E_k} \cdot D_k \right)^2 \quad (4.5)$$

Em que K é o número de agrupamentos.

O fator E_1 (Equação 4.6) é a soma das distâncias de cada amostra ao centro geométrico de todas as amostras. Esse fator não depende do número de agrupamentos.

$$E_1 = \sum_{t=1..n} d(x(t), W_0) \quad (4.6)$$

Em que d é a distancia euclidiana; $x(t)$ é um vetor de dados e W_0 é um centro de agrupamento.

O fator E_k (Equação 4.7) é a soma das distâncias entre grupos de K agrupamentos, ponderada pelo valor de pertinência correspondente de cada amostra ao agrupamento.

$$E_k = \sum_{t=1..n} \sum_{i=1..k} u_i(t) d[x(t), W_i]^2 \quad (4.7)$$

Em que $u_i(t)$ são os graus de pertinência.

D_k representa a máxima separação de cada par de agrupamentos (Equação 4.8).

$$D_k = \max_{i,j=1..k} (d(w_i, w_j)) \quad (4.8)$$

O procedimento para calcular o índice PBM pode ser descrito como a seguir, segundo Pakhira *et al.* (2004).

- 1) Selecione o número máximo de agrupamentos M ;
- 2) Calcule o fator E_1 pela Equação 4.6;
- 3) Para $K=2$ a $K= M$, faça:
 - a. Execute o algoritmo FCM;
 - b. Calcule os fatores E_k e D_k ;

c. Calcule o índice PBM(k).

4) Selecione o melhor número de agrupamentos K (Equação 4.9) como:

$$K = \arg \max(PBM(k)) \quad (4.9)$$

Em que $\arg \max(PBM(k))$ representa o máximo valor do índice PBM.

Quanto maior o índice PBM, melhor a partição. Assim como outros, o índice PBM é um índice de otimização, de forma que para se obter a melhor partição, deve-se processar o algoritmo para diversos valores de K e escolher aquele que resultar no maior valor de índice PBM (MODENESI, 2008).

4.5 MOMENTOS – L

Quando os fenômenos hidrológicos são assumidos como aleatórios, eles passam a ter caráter probabilístico. Com relação a esse caráter, deve-se ressaltar que há probabilidade de um dado evento hidrológico ser inferior ou superior a um registro histórico. Assim, para o entendimento das variáveis hidrológicas, torna-se essencial observar os eventos e modelar as frequências de ocorrência, permitindo que sejam feitas previsões hidrológicas.

Para Wolf (2013), as previsões de variáveis hidrológicas, baseiam-se na condição destas serem contínuas, ou seja, aquelas que em conceito físicos existem sucessivamente no tempo. Sendo assim, é possível trabalhar com cálculos de funções de distribuição de probabilidade, que permitem determinar a probabilidade de um evento ocorrer em um intervalo de tempo definido.

Uma variável contínua como a precipitação, pode ter sua distribuição de probabilidade, apresentada na forma de uma função densidade de probabilidade (FDP). Em vez de ser estabelecida a probabilidade para a ocorrência da precipitação em dado tempo, é estabelecida uma função que integrada em um intervalo, determina a sua probabilidade de ocorrência. No entanto, para o ajuste de uma FDP é necessário estabelecer parâmetros que possam gerar distribuições de probabilidade bem próximas do valor observado.

Para a estimativa de parâmetros das distribuições de probabilidade, tem sido adotado o método dos momentos-L, que são considerados como medidas de posição,

escala e forma das distribuições de probabilidade, similares aos momentos convencionais, porém estimadas por combinações lineares da assimetria, da curtose e do coeficiente de variação. Eles compõem um sistema de medidas estatísticas mais confiáveis para descrição das características das distribuições de probabilidades e são derivados dos momentos ponderados por probabilidade (MPP) (HOSKING & WALLIS, 1993).

Os momentos-L foram apresentados por Hosking (1986) como sendo combinações lineares dos momentos ponderados por probabilidades (MPPs) de uma variável aleatória X . Esses momentos foram definidos mais tarde por Greenwood *et al.* (1979), conforme Equação 4.10.

$$\beta_r = E\{X[Fx(x)]\} \quad (4.10)$$

Em que β_r é o momento ponderado de probabilidade; E é a esperança de ocorrência da variável; e $Fx(x)$ é a função de distribuição cumulativa de X .

Segundo Naghettini e Pinto (2007), a estimativa dos MPPs, a partir de uma amostra finita de tamanho n , inicia-se com a ordenação de seus elementos constituintes em ordem crescente. Um estimador não tendencioso do MPPs (β_r) pode ser obtido pela Equação 4.11.

$$\beta_r = \frac{1}{n} \sum_{j=r+1}^n \frac{(j-1)(j-2)\dots(j-r)}{(n-1)(n-2)\dots(n-r)} x_{j:n} \quad (4.11)$$

Os estimadores de $\beta_r \leq 2$, são obtidos pelas Equações 4.12 a 4.14.

$$\beta_0 = \frac{1}{n} \sum_{j=1}^n x_{j:n} \quad (4.12)$$

$$\beta_1 = \frac{1}{n} \sum_{j=2}^n \frac{(j-1)}{(n-1)} x_{j:n} \quad (4.13)$$

$$\beta_2 = \frac{1}{n} \sum_{j=1}^n \frac{(j-1)(j-2)}{(n-1)(n-2)} x_{j:n} \quad (4.14)$$

Em que X_j representa as amostras e n o número de amostras.

Esses estimadores servem para calcular os quatros primeiros momentos-L de uma distribuição de determinada amostra, $\lambda_1, \lambda_2, \lambda_3, \lambda_4$, que são obtidos pelas Equações 4.15, 4.16, 4.17 e 4.18.

$$\lambda_1 = \beta_0 \quad (4.15)$$

$$\lambda_2 = 2\beta_1 - \beta_0 \quad (4.16)$$

$$\lambda_3 = 6\beta_2 - 6\beta_1 + \beta_0 \quad (4.17)$$

$$\lambda_4 = 20\beta_3 - 30\beta_2 + 12\beta_1 - \beta_0 \quad (4.18)$$

O λ_1 corresponde ao primeiro momento, chamado de momento de posição ou média. O λ_2 é segundo momento e corresponde a uma medida de escala. O λ_3 e o λ_4 são os demais momentos amostrais correspondentes. Segundo Greenwood *et al.* (1979), em termos de medidas de forma das distribuições, torna-se mais conveniente que os momentos-L sejam expressos em quantidades adimensionais, obtidas pela razão entre eles, onde essa razão, dar origem aos quocientes de assimetria e curtose.

Os quocientes de momentos-L, representados pelo coeficiente de variação, assimetria e curtose são obtidos pelas Equações 4.19 a 4.21, simultaneamente, e estão em função dos momentos-L.

$$Cv - L = \tau_2 = \frac{\lambda_2}{\lambda_1} \quad (4.19)$$

$$\gamma - L = \tau_3 = \frac{\lambda_3}{\lambda_2} \quad (4.20)$$

$$K - L = \tau_4 = \frac{\lambda_4}{\lambda_2} \quad (4.21)$$

A determinação dos momentos-L (MML) e dos quocientes de momentos-L em estudos hidrológicos, de uma determinada região, podem auxiliar no tratamento de consistência de dados, na análise regional e na identificação de regiões homogêneas. A

vantagem deste método, consiste em exigir um menor esforço computacional para a solução de sistemas de equações (NAGHETTINI & PINTO, 2007).

O emprego desta metodologia visa permitir a aplicação do teste H, que utiliza os momentos-L, para testar a homogeneidade de regiões classificadas em homogêneas.

4.6 TESTE DE HETEROGENEIDADE (H)

A finalidade deste teste estatístico é verificar o grau de heterogeneidade de uma região. De acordo com Hosking e Wallis (1997), em uma região homogênea, todos os postos com as séries de valores de alguma variável hidrológica na forma adimensional dispõem dos mesmos momentos populacionais. Contudo, devido à variabilidade amostral, os seus momentos amostrais diferem. A estatística H compara essa variabilidade amostral observada com uma variabilidade esperada para uma região homogênea, questionado nesse caso, se a dispersão dos quocientes de momentos-L amostrais, calculados para um certo conjunto de postos, é diferente daquela que se esperaria encontrar em uma região homogênea.

Hosking e Wallis (1997) aconselham o cálculo da dispersão das regiões, proposta e simulada, por intermédio do coeficiente de variação-L amostral τ . Consideram, portanto, a medida de variância V , como uma medida de dispersão, ponderada pelo tamanho das séries, por meio da Equação 4.22.

$$V = \left[\frac{\sum_{i=1}^n n_i (t^i - t^R)^2}{\sum_{i=1}^n n_i} \right] \quad (4.22)$$

Considera-se que a região avaliada contenha n postos de observação, cada um deles, indexados por i , com amostra de tamanho n_i e quocientes de momentos-L amostrais representados por t^i , t_4^i , e que t^R , t_3^R , t_4^R , denotando, respectivamente, as médias regionais dos quocientes $CV-L$, assimetria-L e curtose-L, calculados a partir da Equação 4.11 a 4.21.

Para simular uma região homogênea Hosking e Wallis (1997), aconselham a utilização da distribuição Kapa de quatro parâmetros, pela sua flexibilidade e, ainda para se desviar do comprometimento prévio com uma distribuição particular de dois e três parâmetros. Por outro lado, se não for possível ajustar essa distribuição aos momentos-L médios, isto é, se τ_4 for muito maior que τ_3 , recomenda-se o uso da distribuição Logística Generalizada na simulação, sendo esta distribuição um caso particular da distribuição

Kapa. A distribuição de probabilidades acumulada da distribuição Kapa é dada pela Equação 4.23.

$$F(x) = \left\{ 1 - h \left[1 - \frac{k(x-\xi)}{\alpha} \right]^{\frac{1}{k}} \right\}^{1/h} \quad (4.23)$$

Em que x é a variável estudada, ξ é o parâmetro de posição, α o parâmetro de escala e k e h os parâmetros de forma. Fazendo-se $h=1$, têm-se a Equação da distribuição Logística Generalizada.

Os parâmetros da população Kapa são estimados por intermédio dos momentos-L regionais, $(\lambda_1, \lambda_2, \lambda_3)$, através das Equações 4.24, 4.25 e 4.26.

$$\xi = \lambda_1 - \alpha = \left[\frac{1}{k} - \frac{\pi}{\text{sen}(k\pi)} \right] \quad (4.24)$$

$$\alpha = \frac{\lambda_2 \text{sen}(k\pi)}{k\pi} \quad (4.25)$$

$$k = -\tau_3 = -\frac{\lambda_3}{\lambda_2} \quad (4.26)$$

Na sequência, são gerados N_{sim} conjuntos de séries de n_i valores para a variável normalizada em cada local, a partir dos parâmetros Kapa; posteriormente, calculam-se as estatísticas V_j ($j=1, 2, \dots, N_{sim}$) para cada conjunto, por meio da Equação 22. Hosking e Wallis (1997) indicam que se faça um número mínimo de simulações igual a 500.

A média aritmética das estatísticas V_j obtidas por simulação fornecerá a dispersão média (Equação 4.27) esperada na região.

$$\mu_v = \frac{\sum_{j=1}^{N_{sim}} V_j}{N_{sim}} \quad (4.27)$$

Segundo Hosking e Wallis (1997), a medida de Heterogeneidade H (Equação 4.28) confronta a dispersão observada com a simulada.

$$H = \frac{(V - \mu_v)}{\sigma_v} \quad (4.28)$$

Em que V é o desvio padrão ponderado amostral para $CV-L$; μ_v a média aritmética das estatísticas V_j , obtidas por simulação e σ_v (Equação 4.29) é o desvio padrão entre os valores da medida de dispersão das amostras simuladas (N_{sim}).

$$\sigma_v = \sqrt{\frac{\sum_{j=1}^{n_{sim}} (V_j - \mu_v)^2}{N_{sim}^{-1}}} \quad (4.29)$$

De acordo com o teste de significância, proposto por Hosking & Wallis (1997), se $H < 1$, considera-se a região como “aceitavelmente homogênea”, se $1 \leq H < 2$, a região é “possivelmente homogênea” e, finalmente, se $H \geq 2$, a região deve ser classificada como “definitivamente heterogênea”.

Nesse teste H , busca-se avaliar as regiões homogêneas, analisando somente a variável precipitação. O comportamento dessa variável nos grupos, a partir do emprego da estatística H vai confirmar ou não a homogeneidade da região formada. Assim, pode-se inferir que, especificamente neste estudo, o teste de heterogeneidade é uma metodologia auxiliar na identificação de região homogênea, obtidas pela análise de agrupamento, via fuzzy c-means.

4.7 DISTRIBUIÇÕES DE PROBABILIDADE

As funções ou modelos de densidades de probabilidades são comumente usadas em engenharia para descrever sistemas físicos. Na área de hidrologia, esses modelos, buscam uma projeção do que ocorrerá no futuro, baseados na frequência de ocorrências passadas de chuvas e vazões. Assim, para se modelar a frequência de dados hidrológicos é preciso fazer um estudo de sua ocorrência, em que se estabelece um percentual que a variável hidrológica pode ser maior que um dado valor. Isso é chamado de frequência de excedência e pode ser obtida diretamente de uma série histórica de dados. Contudo pode-se trabalhar com a frequência de não excedência, a qual corresponde ao percentual da variável ser menor ou igual a um dado valor.

A forma mais usual de estimar as frequências observadas de uma variável é por meio da fórmula empírica de Weibull (Equação 4.30).

$$f_{obs} = \frac{i}{N+1} \quad (4.30)$$

Em que i é a ordem que determinado dado ocupa; N refere-se ao tamanho da série histórica.

Com base nos estudos das frequências de ocorrência, ajusta-se uma distribuição de probabilidade, e aquela que obtiver o melhor ajuste deve ser adotada. Dentre os modelos ou funções de distribuição de probabilidade muito utilizadas em recursos hídricos, estão as distribuições: Gama, Gumbel, Normal, Weibull, Exponencial e Log-normal, as quais são apresentadas a seguir.

a) Gama

A distribuição Gama é utilizada para o ajuste de totais de chuva de períodos mensais ou menores (ASSIS *et al.* 1996; SILVA & ASSAD, 1998). Sua função de densidade de probabilidade (FDP) Gama, para 2 parâmetros, é dada pela Equação 4.31.

$$FDP: f(X) = \frac{1}{\Gamma(\gamma)\beta^\gamma} X^{\gamma-1} e^{-\frac{X}{\beta}} \quad (4.31)$$

Em que β é dado pela Equação 4.32, e γ dado pela Equação 4.33, sendo β o parâmetro de escala, γ o parâmetro de forma, e $\Gamma(\gamma)$ a função gama do parâmetro.

$$\beta = \frac{s^2}{\bar{x}} \quad (4.32)$$

$$\gamma = \frac{\bar{x}^2}{s^2} \quad (4.33)$$

Em que, s^2 a variância e \bar{x} a média dos valores observados. O valor de Γ pode ser encontrado em manuais de fórmulas e tabelas matemáticas.

b) Gumbel

A distribuição Gumbel é bastante aplicada a acontecimentos extremos, como por exemplo, vazões máximas anuais ou a distribuição das precipitações totais anuais (MELLO, 2013). Sua FCP de não excelência é dada pela Equação 4.34.

$$FCP: P(Y \leq y) = e^{-\alpha \cdot (y - \mu)} \quad (4.34)$$

A qual apresenta como função de densidade de probabilidade a Equação 4.35.

$$FDP: f(y) = \alpha \cdot e^{\{-\alpha \cdot (y-\mu) - e^{-\alpha(y-\mu)}\}} \quad (4.35)$$

Em que α (Equação 4.36) e μ (Equação 4.37) são os parâmetros de distribuição, sendo α , o parâmetro de escala e μ é o parâmetro de posição (localização).

$$\alpha = 1,2826/s \quad (4.36)$$

$$\bar{\mu} = \bar{x} - 0,451 \cdot s \quad (4.37)$$

c) Normal

A distribuição Normal é utilizada para cálculo da propriedade acumulada de ocorrências totais de precipitação (HASTINGS e PEACOCK, 1975). A distribuição normal é uma distribuição de dois parâmetros. Sua função de probabilidade pode ser observada na Equação 4.38.

$$FDP: f(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-0,5 \left(\frac{x-\mu}{\sigma}\right)^2} \quad (4.38)$$

Em que μ é a média dos valores das variáveis observadas e σ é o desvio padrão dos valores das variáveis observadas.

Definindo uma variável $y = (x - \mu)/\sigma$ pode ser verificado que ela tem média igual a 0 e desvio padrão igual a 1. Logo, ela distribui-se segundo a Normal (Equação 4.39).

$$FDP: f(Y) = \frac{1}{\sqrt{2\pi}} e^{-0,5 \cdot y^2} \quad (4.39)$$

Essa é chamada de distribuição Normal padrão, sendo y a variável padronizada. A vantagem do estudo dessa distribuição é que qualquer distribuição normal pode ser reduzida a ela, se for efetuada a transformação de não-excedência relacionada com $f(x)$, sendo dada pela Equação 4.40.

$$FCP: F(Y \leq y) = \int_{-\infty}^y \frac{1}{\sqrt{2\pi}} e^{-0,5 \cdot \tau^2} \cdot d\tau \quad (4.40)$$

Em que a integração não tem solução analítica, então para seu cálculo pode ser utilizada uma aproximação numérica dada pela Equação 4.41 (TUCCI, 2012).

$$F(Y \leq y) = 1 - f(y) \cdot (a_1 \cdot q + a_2 \cdot q^2 + a_3 \cdot q^3) \quad (4.41)$$

Em que y corresponde aos valores da variável padronizada e os coeficientes assumem os seguintes valores $a_0 = 0,33267$; $a_1 = 0,43618$; $a_2 = 0,12017$ $a_3 = 0,93730$ e $q = (1 + a_0 \cdot Y)^{-1}$.

d) Weibull

A distribuição de probabilidade Weibull pode ser utilizada em eventos extremos, principalmente para estudos hidrológicos, assim como a distribuição Gumbel (DUAN et al., 1998). Nesse caso, as funções cumulativas de probabilidade de não-excedência e excedência são dadas através das Equações 4.42 e 4.43 respectivamente.

$$FCP: P(Z \leq z) = 1 - e^{-\left[\frac{z-\gamma}{\mu-\gamma}\right]^\lambda} \quad (4.42)$$

$$FCP: P(Z \geq z) = e^{-\left[\frac{z-\gamma}{\mu-\gamma}\right]^\lambda} \quad (4.43)$$

Com sua função densidade de probabilidade dada pela Equação 4.44.

$$FDP: f(z) = \left[\frac{\lambda}{\mu-\gamma}\right] \cdot \left[\frac{z-\gamma}{\mu-\gamma}\right]^{\lambda-1} P(Z \leq z) \quad (4.44)$$

Os parâmetros λ , μ e γ , são obtidos pelas Equações 4.45, 4.46 e 4.47 que representam respectivamente esperança matemática, variância e assimetria populacional.

$$E(Z) = \gamma + (\mu - \gamma) \Gamma\left(1 + \frac{1}{\lambda}\right) \quad (4.45)$$

$$VAR(Z) = (\mu - \gamma)^2 \left[\Gamma\left(1 + \frac{2}{\lambda}\right) - \Gamma^2\left(1 + \frac{1}{\lambda}\right) \right] \quad (4.46)$$

$$A(Z) = \frac{\Gamma(1+3/\lambda) - 3\Gamma(1+2/\lambda)\Gamma(1+1/\lambda) + 2\Gamma^3(1+1/\lambda)}{[\Gamma(1+2/\lambda) - \Gamma^2(1+1/\lambda)]^{1,5}} \quad (4.47)$$

Verificando-se que a assimetria populacional depende apenas do parâmetro λ , pode-se obter uma estimativa de λ , conhecida como estimativa amostral (Equação 4.48).

$$\frac{1}{\lambda} = [C_0 + C_1 \cdot \alpha + C_2 \cdot \alpha^2 + C_3 \cdot \alpha^3 + C_4 \cdot \alpha^4] \quad (4.48)$$

Em que $C_0 = 0,2777757913$, $C_1 = 0,3132617714$, $C_2 = 0,057567091$, $C_3 = -0,0013038566$, $C_4 = -0,0081523408$; e α a estimativa amostral da assimetria, que deverá apresentar valores entre -1,02 e 2 para validade da Equação 4.48 (TUCCI, 2012).

e) Exponencial

É uma função de distribuição que possui inúmeras aplicações, em diversas áreas do conhecimento humano e, em particular, em hidrologia. Segundo Tucci (2012), a importância da distribuição Exponencial vem de que ela, assim como outras, é definida para valores contínuos de tempo.

Sua função de Probabilidade é dada pela Equação 4.49.

$$FDP: f(x) = a \cdot e^{-\alpha \cdot x} \quad (4.49)$$

A sua FCP de não excedência pode ser calculada pela integração da Equação 4.49, resultando na Equação 4.50.

$$FCP: P(X \leq x) = \int_{\tau=0}^x \alpha \cdot e^{-\alpha \cdot t} d\tau = 1 - e^{-\alpha \cdot x} \quad (4.50)$$

Em que x é a variável aleatória e α o inverso da média aritmética dos valores de x observados em relação ao tempo t .

f) Log-Normal

Segundo Mello (2013), a distribuição Log-Normal com dois ou três parâmetros busca, por meio de transformação logarítmica dos dados da série histórica, aproximar os dados, que originalmente apresentam-se com distribuição de frequência assimétrica, a uma distribuição normal. A função densidade de probabilidade é dada pela Equação 4.51.

$$FDP: P(x) = \frac{1}{x \cdot \sigma_n \sqrt{2\pi}} e^{-0,5 \left(\frac{\ln(x) - \mu_n}{\sigma_n} \right)^2} \quad (4.51)$$

Com $x \geq 0$, μ_n , a média dos valores das variáveis observadas e σ_n o desvio padrão dos valores das variáveis observadas, os quais podem ser calculados através das Equações 4.52 e 4.53.

$$\mu_n = E(X_n) = E(\ln(X)) \quad (4.52)$$

$$\sigma_n^2 = VAR(X_n) = VAR(\ln(X_n)) \quad (4.53)$$

Em que η é o coeficiente de variação dado pela Equação 4.54.

$$\eta = \frac{\sigma}{\mu} = [e^{\sigma_n^2} - 1]^{1/2} \quad (4.54)$$

Em que μ é a média dos logaritmos e σ é o desvio padrão dos logaritmos das variáveis observadas.

4.7.1 Teste de Aderência – Qui-quadrado

As distribuições de probabilidades foram aplicadas às séries anuais e mensais das estações pluviométricas, as quais já pertenciam a um determinada região homogênea. O Teste Qui-Quadrado é aplicado para auxiliar na seleção da melhor função de probabilidade. Neste teste é feito a comparação entre a soma do quadrado dos desvios entre as frequências observadas e simuladas e o valor obtido em tabelas.

O valor do teste Qui-Quadrado pode ser determinado através da Equação 4.55.

$$X^2 = \sum \left[\frac{(f_o - f_e)^2}{f_e} \right] \quad (4.55)$$

Em que:

f_o - frequência observada (mm);

f_e - frequência estimada (mm) pela função de probabilidade;

O teste do Qui-quadrado pode ser considerado como um teste de adequação, no qual, pretende-se determinar se uma dada distribuição teórica é razoável diante dos dados disponíveis. Assim pode-se realizar o teste de hipótese para verificar se:

1. Se $H=0$, e $f_o=f_e$, então não há associação entre os grupos;
2. Se $H\neq 0$, e $f_o\neq f_e$, então há associação entre os grupos.

A tomada de decisão é dada comparando-se os valores de X^2 calculado e os valores de X^2 tabelado, assim tem-se as seguintes hipóteses:

1. Se X^2 calculado $\geq X^2$ tabelado, então se rejeita a $H=0$, $f_o=f_e$;
2. Se X^2 calculado $< X^2$ tabelado, então se aceita a hipótese $H=0$.

A estatística deste teste é obtida por meio da tabela do Qui-quadrado (Anexo A), adotando-se o valor tabelado com base em graus de liberdade e nível de significância. Para que a distribuição de probabilidade seja considerada adequada (aceitação da hipótese nula H_0), o valor X^2 , calculado, deve ser menor que o de tabela.

4.7.2 Calibração dos Modelos de Probabilidade

Na calibração de funções de probabilidade foram feitas simulações, procurando estimar a probabilidade de ocorrência de alturas de precipitações médias anuais e mensais, das regiões homogêneas encontradas através da análise de agrupamento Fuzzy c-means. Nesta etapa, foram consideradas as séries históricas das estações de cada região homogênea e um código computacional denominado de FDP (Apêndice C), elaborado para geração das frequências de ocorrência de chuva. Assim, foram geradas as distribuições de frequência e a determinação do melhor ajuste através do teste Qui-quadrado.

4.7.3 Validação dos Modelos de Probabilidades

Nesta etapa, buscou-se avaliar o ajuste das funções de probabilidades selecionadas na calibração em estações alvo, previamente selecionadas, localizadas nas regiões homogêneas formadas, mas que não foram adotadas na etapa da calibração. Determinou-se a distribuição de frequência para as estações alvo e analisou-se o melhor ajuste, considerando o resultado gráfico e o teste do Qui-quadrado.

4.8 REGIONALIZAÇÃO POR REGRESSÃO MÚLTIPLA

A análise de regressão é um dos inúmeros modelos estatísticos explanatórios causais referentes ao tratamento de séries temporais de dados. É uma metodologia de

previsão de valores de uma ou mais variáveis de resposta (dependentes) através de um conjunto de variáveis explicativas (independentes).

Segundo HAIR *et al.* (2005), a análise de regressão múltipla é uma técnica estatística que pode ser usada para verificar a relação entre uma única variável dependente e várias independentes. O objetivo da análise de regressão múltipla é usar as variáveis independentes cujos valores são conhecidos para prever os valores da variável dependente selecionada pelo pesquisador. A relação entre a variável dependente e as demais variáveis independentes pode ser formada de acordo com um modelo linear dado pela Equação 4.56.

$$Y = \beta_0 + \beta_1 \cdot X_1 + \beta_2 \cdot X_2 + \beta_i \cdot X_i + \varepsilon \quad (4.56)$$

Em que Y é a variável dependente ou prevista, X_1, X_2, \dots, X_i são as variáveis independentes ou explicativas. $\beta_0, \beta_1, \beta_2, \dots, \beta_i$, são os coeficientes de regressão e ε denota os resíduos da regressão.

O coeficiente β_0 é conhecido como a interseção do plano ou coeficiente linear, e os coeficientes $\beta_1, \beta_2, \dots, \beta_i$ são chamados de coeficientes parciais de regressão, porque (no caso de duas variáveis independentes) β_1 mede a variação esperada em Y por unidade de variação em X_1 , quando X_2 for constante, e β_2 , mede a variação esperada em Y por unidade de variação em X_2 , quando X_1 for constante. Para determinação dos coeficientes $\beta_0, \beta_1, \beta_2, \dots, \beta_i$, usa-se o método dos mínimos quadrados.

O modelo da regressão múltipla é representado em notação matricial pela Equação 4.57.

$$[Y] = [X] \cdot [\beta] \quad (4.57)$$

Em que $[Y]$ é um vetor ($n \times 1$) das observações da variável dependente; $[X]$ é uma matrix ($n \times P$) com as n observações de cada uma das P variáveis independentes, e $[\beta]$ é um vetor ($P \times 1$) com os coeficientes desconhecidos (HAIR *et al.*, 2005). Na forma matricial tem-se então (Equação 4.58):

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \dots \\ Y_n \end{pmatrix}, X = \begin{pmatrix} 1 & X_{1,1} & X_{1,2} & X_{1,p} \\ 1 & X_{2,1} & X_{2,2} & X_{2,p} \\ \dots & \dots & \dots & \dots \\ 1 & X_{n,1} & X_{n,2} & X_{n,p} \end{pmatrix}, \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \dots \\ \beta_n \end{pmatrix} \quad (4.58)$$

As equações normais de regressão são formadas pelo sistema (Equação 4.59) e podem ser obtidas pela multiplicação de ambos os membros da Equação 4.58 por 1, X_1 , X_2, \dots, X_i sucessivamente, e a soma membro a membro das expressões resultantes.

$$\begin{cases} \sum Y = \beta_0 \cdot N + \beta_1 \cdot \sum X_1 + \beta_2 \cdot \sum X_2 + \dots + \beta_i \cdot \sum X_i \\ \sum Y \cdot X_1 = \beta_0 \cdot \sum X_1 + \beta_1 \cdot \sum X_1^2 + \beta_2 \cdot \sum X_2 \cdot X_1 + \dots + \beta_i \cdot \sum X_i \cdot X_1 \\ \sum Y \cdot X_2 = \beta_0 \cdot \sum X_2 + \beta_1 \cdot \sum X_1 \cdot X_2 + \beta_2 \cdot \sum X_2^2 + \dots + \beta_i \cdot \sum X_1 \cdot X_2 \\ \sum Y \cdot X_i = \beta_0 \cdot \sum X_i + \beta_1 \cdot \sum X_1 X_i + \beta_2 \cdot \sum X_2 \cdot X_i + \dots + \beta_i \cdot \sum X_i^2 \end{cases} \quad (4.59)$$

Pode-se dizer que as soluções das equações normais de regressão são encontradas pela multiplicação dos termos da equação por $([X]^T \cdot [X])^{-1}$. Assim, a solução do vetor $[\hat{\beta}]$ corresponde ao estimador de mínimos quadrados de β dado pela Equação 4.60.

$$[\hat{\beta}] = ([X]^T \cdot [X])^{-1} \cdot [X]^T \cdot [Y] \quad (4.60)$$

O somatório total dos quadrados pode ser representado pela Equação 4.61.

$$[Y]^T [Y] = n \cdot \hat{Y}^2 + \left([\hat{\beta}]^T [X^T] [Y] - n \cdot \hat{Y}^2 \right) + ([X^T] [Y] + ([\hat{\beta}]^T [X^T] [Y]) \quad (4.61)$$

Essas parcelas dos somatórios dos quadrados são calculados por planilhas eletrônicas na forma de uma tabela de análise de variância (ANOVA), tal como ilustra a Tabela 3. Essas tabelas são comumente usadas em diversos estudos estatísticos.

Tabela 3 – ANOVA de regressão múltipla.

Fonte	Somatório dos quadrados	Quadrados médios	Graus de liberdade
Regressão	$SQReg = [\hat{\beta}]^T [Y]^T [Y] - n \cdot \hat{Y}^2$	$QMReg = \frac{SQReg}{P}$	P
Resíduos	$SQRes = [X^T] [Y] + ([\hat{\beta}]^T [X^T] [Y])$	$QMRes = \frac{SQRes}{n - P - 1}$	n - P - 1
Total	$SQT = [Y]^T [Y] - n \cdot \hat{Y}^2$		n - 1

Fonte: HAIR, *et al.* (2005).

Em que:

n = tamanho da amostra;

P = número de variáveis independentes;

$SQReg$ = somatório dos quadrados da regressão;

$QMReg$ = quadrado médio da regressão;

$QMRes$ = quadrado médio dos resíduos;

SQT = somatório dos quadrados total.

4.8.1 Modelos de Regressão Múltipla

Na determinação da variável dependente Y (precipitação - P), foi aplicada a regressão múltipla entre as variáveis independentes (Altitude - H , Latitude - la , Longitude - lo) para a determinação dos parâmetros, β_0 , β_1 , β_2 , e β_3 (equações. 4.58 a 4.61), que são as constantes da regressão, determinadas via método dos mínimos quadrados; e as variáveis independentes, por meio dos seguintes modelos estatísticos, a saber: modelo Linear (Equação 4.62), Potencial (Equação 4.63), Exponencial (Equação 4.64) e Logaritmo (Equação 4.65).

$$P = \beta_0 + \beta_1 \cdot H + \beta_2 \cdot la + \beta_3 \cdot lo \quad (4.62)$$

$$P = \beta_0 + H^{\beta_1} + la^{\beta_2} + lo^{\beta_3} \quad (4.63)$$

$$P = e^{\beta_0 + \beta_1 \cdot H + \beta_2 \cdot la + \beta_3 \cdot lo} \quad (4.64)$$

$$P = \beta_0 + \beta_1 \cdot \ln(H) + \beta_2 \cdot \ln(la) + \beta_3 \cdot \ln(lo) \quad (4.65)$$

Em que

P = precipitação simulada;

H - altitude;

la – latitude;

lo – longitude;

$\beta_0, \beta_1, \beta_2,$ e β_3 são os coeficientes de regressão.

A escolha dos modelos Linear, Potencial, Exponencial e Logaritmo, deve-se ao fato destes serem muito bem empregados em hidrologia e em estudos para simulação de precipitação.

Na maioria dos estudos, envolvendo modelos de regressão, observa-se, que, apenas com o uso das variáveis Latitude, Longitude e Altitude; que são as mais fáceis de serem adquiridas, foram obtidos bons resultados com modelos satisfatórios na estimativa da precipitação, como por exemplo nos trabalhos de Mello & Silva (2009), Teixeira *et al.* (2013) e Teixeira-Gandra *et al.* (2015).

4.8.2 Teste de Multicolinearidade

A multicolinearidade está relacionada, a condição em que uma certa variável explicativa possui alta correlação com uma ou mais variáveis explicativas, implicando em uma superparametrização do modelo de regressão, afetando a qualidade dos modelos de regressão. Segundo Naghettini e Pinto, (2007), para se evitar a multicolinearidade elimina-se uma entre cada conjunto de duas variáveis independentes que apresentarem coeficiente de correlação superior a 0,85. Para Helsel e Hirsch (1992), algumas das consequências da multicolinearidade são:

- equações aceitáveis em termo do teste de F_{total} cujos coeficientes possuem escalas não realistas;
- coeficientes podem ter sinais não realistas; e
- coeficientes instáveis: uma pequena mudança em um ou poucos dados de entrada pode provocar grandes mudanças nos coeficientes.

Em geral, as etapas e os critérios de seleção dos melhores modelos de regressão e do melhor conjunto de variáveis explicativas são: (a) definição da matriz de correlação simples entre as variáveis; (b) cálculo do coeficiente de determinação ajustado (R^2_a); e (c) testes de estatísticas F_{total} , para a verificação da significância do modelo de regressão como um todo.

A matriz de correlação é construída a partir do cálculo dos coeficientes de correlação simples entre as variáveis do modelo. A correlação entre duas variáveis X_1 e X_2 é determinada pelo coeficiente de regressão simples r , conforme Equação 4.66, definida por:

$$r = \frac{\sum_{i=1}^n X_{1,i} X_{2,i} - \sum_{i=1}^n X_{1,i} \cdot \sum_{i=1}^n X_{2,i}}{\sqrt{[\sum_{i=1}^n X_{1,i}^2 - (\sum_{i=1}^n X_{1,i})^2] \cdot [\sum_{i=1}^n X_{2,i}^2 - (\sum_{i=1}^n X_{2,i})^2]}} \quad (4.66)$$

O coeficiente r varia de -1 a 1. Quando r é positivo, indica uma tendência de crescimento conjunto de X_1 e X_2 . Quando r é negativo, maiores valores de X_1 são associados a menores valores de X_2 . Quanto mais próximo da unidade, melhor a correlação entre X_1 e X_2 .

Considerando X_1 , X_2 e X_3 as variáveis explicativas e r o coeficiente simples entre as variáveis, pode-se descrever a matriz de correlação como:

Tabela 4 – Matriz de correlação Pearson.

	X_1	X_1	X_1
X_1	1		
X_1	$r_{X_1X_2}$	1	
X_1	$r_{X_1X_3}$	$r_{X_2X_3}$	1

4.8.3 Desempenho dos Modelos de Regressão

Com o intuito de analisar o desempenho e a eficiência dos modelos matemáticos aos dados observados de precipitação, foram adotados critérios de avaliação como o coeficiente de determinação - R^2 , o coeficiente de determinação ajustado - R^2_a , o erro relativo percentual - $E\%$, o erro quadrático relativo médio percentual - $\epsilon\%$, o coeficiente de Nash-Sutcliffe – NASH e a raiz quadrada da média dos erros quadráticos (The Root Mean Square Error – RMSE).

Coefficiente de Determinação

Na aplicação da regressão múltipla, o uso do coeficiente de determinação R^2 é indispensável, pois este é capaz de explicar o quanto o modelo é representativo na estimação da variável. O mesmo varia entre 0 e 1. Quanto mais próximo de 1, melhor a representação do modelo.

O coeficiente de regressão R^2 pode ser obtido pela Equação 4.67.

$$R^2 = \frac{SQReg}{SQT} = \frac{[\hat{\beta}]^T [Y]^T [Y] - n \bar{y}^2}{[Y]^T [Y] - n \bar{y}^2} \quad (4.67)$$

Coeficiente de Determinação ajustado

O coeficiente de determinação ajustado (R^2_a), é calculado, considerando o número de variáveis independentes da equação de regressão e pode ser obtido pela Equação 4.68, e assim como o R^2 , quanto mais próximo de 1, melhor é a representação do modelo.

$$R^2_a = 1 - \frac{n-1}{n-p-1} (1 - R^2) \quad (4.68)$$

Em que n é o número de valores observados, p é o número de variáveis independentes e R^2 é o coeficiente de regressão.

Erro Relativo Percentual

O erro relativo percentual ($E\%$) avalia o desempenho do modelo, considerando a diferença percentual entre os valores da variável observada e a simulada, conforme Equação 4.69.

$$E\% = \left(\frac{P_{obs} - P_{sim}}{P_{obs}} \right) * 100 \quad (4.69)$$

Em que: P_{obs} é a precipitação observada e P_{sim} é precipitação simulada pelo modelo.

Erro Quadrático Relativo Médio Percentual

O erro quadrático relativo médio percentual ($\epsilon\%$) prioriza o ajuste dos valores relativos, retirando o peso de precipitações maiores ou menores. Pode ser calculado, utilizando a Equação 4.70.

$$\epsilon = N^{-1} \left[\sum_{i=1}^N \left(\frac{p_i - \hat{p}_i}{p_i} \right)^2 \right]^{1/2} \quad (4.70)$$

Em que p_i é a precipitação observada (m^3/s), \hat{p}_i é a precipitação estimada pelo modelo (m^3/s) e N corresponde ao número total de precipitações observadas.

Coefficiente de Nash-Sutcliffe (NASH)

A estatística desse coeficiente compara a redução do desvio quadrático do erro do modelo com o desvio quadrático do modelo alternativo de prever sempre a média dos valores. Podemos dizer, então, que para cada estação “s” =1, 2, 3, ..., n, o coeficiente NASH é dado pela Equação 4.71:

$$NASH = 1 - \frac{\sum(Y_{obs} - Y_{cal})^2}{\sum(Y_{obs} - \bar{X})^2} \quad (4.71)$$

Em que Y_{obs} é a precipitação observada, Y_{cal} é a precipitação simulada pelo modelo e \bar{X} é a média das precipitações observadas.

O coeficiente de NASH pode variar de $-\infty$ a 1, sendo o valor de 1 indicativo de um perfeito ajuste, enquanto valores inferiores a 0,36 fazem com que o modelo seja julgado como inaceitável (COLLISCHONN, 2001; SILVA, *et. al.*, 2008).

Índice de desempenho RMSE

A raiz quadrada da média dos erros quadráticos (RMSE), é utilizada para estimar a qualidade de um modelo. O uso do RMSE tem grande vantagem em relação ao coeficiente múltiplo (R^2); contudo, não estabelece um percentual ou tipo de erro como critério de comparação. Menores valores nesses índices evidenciam melhor qualidade de ajuste e menores erros. A obtenção desse índice pode ser feita através da Equação 4.72.

$$RMSE = \sqrt{\frac{1}{n} * \sum(Y_{obs} - Y_{sim})^2} \quad (4.72)$$

4.8.4 Calibração da Regressão Múltipla

A Regressão Múltipla foi realizada com o auxílio de planilhas eletrônicas, nas quais foram organizadas tabelas, contendo as informações das variáveis independentes (altitude, latitude e longitude) e dependente (precipitação).

Na calibração dos modelos, utilizaram-se as precipitações totais médias anuais e mensais das estações pluviométricas de cada um dos grupos formados pelo método *fuzzy c-means*.

Utilizou-se quatro modelos, o Linear, Potencial, Exponencial e Logaritmo. Os parâmetros de regressão β_0 , β_1 , β_2 , e β_3 , foram calculados pelo método dos mínimos

quadrados. Tais parâmetros representam as informações físicas das estações. Essas informações, através dos modelos, podem ser transferidas para locais sem dados de precipitação, permitindo a simulação dos valores de precipitação. Assim, a regionalização é efetuada por meio da regressão dos parâmetros em relação às características geográficas e de altitude locais das estações pluviométricas utilizadas na calibração. O desempenho dos modelos de regressão foi avaliado e o melhor modelo foi posteriormente validado.

4.8.5 Validação dos Modelos de Regressão Múltipla

Na etapa da validação foram selecionadas 9 estações pluviométricas chamadas de estações alvo. Essas estações foram escolhidas conforme sua localização, preferencialmente as que se encontravam nas regiões homogêneas definidas e que possuísem dados de precipitação. Desta forma, com base nos seus dados de localização e a altitude, e aplicando o modelo de regressão escolhido, estimou-se a precipitação. Assim, adotando o erro relativo percentual ($E\%$), comparou-se o valor observado e simulado dos totais médios anuais e mensais. Se os erros encontrados, forem baixos, então se confirma o bom desempenho do modelo para estimar as precipitações nas regiões homogêneas.

5 RESULTADOS E DISCUSSÃO

5.1 FORMAÇÃO DAS REGIÕES HOMOGÊNEAS DE PRECIPITAÇÃO

5.1.1 Método Fuzzy c-means

A aplicação do método Fuzzy c-means, foi realizada em ambiente computacional, por intermédio da função *fcm*, que funciona como um algoritmo, organizado em uma rotina de programação (Apêndice B), capaz de gerar o processamento dessa função. Assim foi criado um arquivo de texto (*matrix.txt*) composto pelos valores das variáveis adotadas das 83 estações pluviométricas. Essa matriz é um dos dados de entrada do algoritmo, juntamente com o número de grupos (p), o parâmetro de fuzificação (m) e o critério de parada (ϵ). Após a aplicação da análise de agrupamento pelo método Fuzzy c-means, pelo qual foram testados 7 agrupamentos, variando o número de grupos, de 2 a 8, obtiveram-se os graus de pertinência de cada estação.

Na realização dos agrupamentos foram utilizadas variáveis de precipitação média anual, altitude e localização (latitude e longitude), verificou-se que, quanto maior o número de grupos, menor seria o valor do índice PBM. Desta forma, foram realizados testes até 8 grupos, tornando a pesquisa mais objetiva, já que o índice PBM, tenderia a diminuir com agrupamentos maiores que 8 e assim não poderia assegurar uma partição boa entre os grupos formados.

Além do número de grupos, foram testados os parâmetros de fuzificação, no intervalo de 1,2 a 2,0, indicado por ROSS (1995) e COX (2005), para verificar qual agrupamento apresentaria o melhor resultado, analisando tanto o número de grupos, como o parâmetro de fuzificação, sendo adotado o número máximo de iterações (t_{max}) igual a 200 e o critério de parada igual a um erro mínimo (ϵ) de 0,001.

Pelo fato da composição da série histórica de 15 estações pluviométricas, não estarem com dados de 30 anos completo, optou-se em realizar uma análise de sensibilidade nesta etapa do agrupamento, para verificar se as médias de chuva de estações com períodos diferentes influenciaria no resultado da formação dos grupos. Assim foram calculadas as médias anuais para séries com 5, 10, 15, 20, 25 e 30 anos, e a média anual de toda a série histórica de cada estação e assim poder avaliar se essa composição da série influenciaria no agrupamento.

Aplicando a metodologia fuzzy c-means, obteve-se como resposta do algoritmo, os graus de pertinência de cada elemento, o número de iterações e o valor da função

objetivo, para as diferentes situações, considerando a composição da série histórica. Realizou-se, portanto 441 simulações para encontrar o melhor agrupamento (Tabela 5).

Para definir entre os 441 agrupamentos, qual seria o mais adequado, foi aplicado o índice PBM aos resultados do agrupamento, procurando assegurar que a partição gerada, tenha grupos compactos com grande separação entre eles. Este índice indica, como melhor agrupamento, o que apresentar o maior valor de $K = \arg \max(PBM(k))$.

Tabela 5 - Índice PBM para cada agrupamento P e parâmetro de fuzificação.

PBM-Referente a dados de médias anuais de chuva de 5 anos de observação									
P	$m=1,2$	$m=1,3$	$m=1,4$	$m=1,5$	$m=1,6$	$m=1,7$	$m=1,8$	$m=1,9$	$m=2,0$
2	2E+05	2E+05	2E+05	2E+05	3E+04	1E+05	1E+05	3E+04	3E+04
3	1E+06	4E+06	1E+06	4E+06	8E+05	1E+06	1E+06	9E+05	5E+05
4	5E+04	9E+04	6E+04	4E+04	2E+05	6E+04	1E+05	1E+04	5E+04
5	5E+04	2E+04	5E+04	4E+04	4E+04	4E+04	6E+04	1E+05	7E+04
6	7E+04	3E+04	2E+04	3E+04	2E+04	6E+04	4E+04	5E+04	2E+04
7	6E+04	2E+04	3E+04	2E+04	2E+04	4E+04	2E+04	2E+04	2E+04
8	4E+04	1E+04	2E+04	2E+04	2E+04	3E+04	2E+04	4E+04	1E+04

PBM-Referente a dados de médias anuais de chuva de 10 anos de observação									
P	$m=1,2$	$m=1,3$	$m=1,4$	$m=1,5$	$m=1,6$	$m=1,7$	$m=1,8$	$m=1,9$	$m=2,0$
2	3E+04	3E+04	2E+05	3E+04	3E+04	3E+04	2E+05	1E+05	1E+05
3	1E+06	3E+06	8E+06	1E+06	3E+06	3E+06	9E+04	1E+06	2E+06
4	1E+05	5E+04	1E+05	4E+04	4E+04	5E+04	4E+04	4E+04	3E+04
5	3E+04	5E+04	4E+04	2E+04	2E+04	8E+04	2E+04	3E+04	4E+04
6	3E+04	1E+04	2E+04	2E+04	2E+04	3E+04	4E+04	3E+04	1E+04
7	1E+04	1E+04	2E+04	3E+04	2E+04	1E+04	2E+04	4E+04	1E+04
8	2E+04	1E+04	3E+02	3E+04	2E+04	2E+04	2E+04	3E+04	2E+04

PBM-Referente a dados de médias anuais de chuva de 15 anos de observação									
P	$m=1,2$	$m=1,3$	$m=1,4$	$m=1,5$	$m=1,6$	$m=1,7$	$m=1,8$	$m=1,9$	$m=2,0$
2	1E+05	7E+04	3E+04	3E+04	3E+04	1E+05	3E+04	5E+04	3E+04
3	1E+06	1E+06	2E+06	1E+06	1E+06	2E+06	8E+05	2E+06	4E+04
4	1E+05	2E+04	5E+04	3E+04	7E+04	2E+04	4E+04	2E+04	4E+04
5	4E+04	1E+04	2E+04	2E+04	2E+04	5E+04	4E+04	3E+04	4E+04
6	2E+04	2E+04	2E+04	3E+04	4E+04	2E+04	4E+04	3E+04	4E+04
7	2E+04	2E+04	2E+04	2E+04	1E+04	2E+04	2E+04	2E+04	1E+04
8	2E+04	1E+04	2E+04	5E+03	2E+04	2E+04	7E+03	5E+04	9E+03

PBM-Referente a dados de médias anuais de chuva de 20 anos de observação									
P	$m=1,2$	$m=1,3$	$m=1,4$	$m=1,5$	$m=1,6$	$m=1,7$	$m=1,8$	$m=1,9$	$m=2,0$
2	3E+04	3E+04	3E+04	2E+04	1E+05	1E+05	3E+04	1E+05	1E+05

3	9E+05	1E+06	1E+06	1E+06	8E+05	8E+05	2E+05	2E+07	2E+06
4	3E+04	3E+04	3E+04	3E+04	2E+04	8E+04	4E+04	6E+04	2E+05
5	3E+04	1E+04	3E+04	3E+04	3E+04	5E+04	2E+04	3E+04	2E+04
6	2E+04	2E+04	2E+04	3E+04	3E+04	1E+04	2E+04	1E+04	3E+04
7	4E+04	2E+04	1E+04	3E+04	2E+04	2E+04	1E+04	2E+04	3E+04
8	5E+04	9E+03	2E+04	2E+04	1E+04	2E+04	2E+04	2E+04	1E+04

PBM-Referente a dados de médias anuais de chuva de 25 anos de observação

<i>P</i>	<i>m</i> =1,2	<i>m</i> =1,3	<i>m</i> =1,4	<i>m</i> =1,5	<i>m</i> =1,6	<i>m</i> =1,7	<i>m</i> =1,8	<i>m</i> =1,9	<i>m</i> =2,0
2	1E+05	3E+04	1E+05	3E+04	3E+04	1E+05	1E+05	1E+05	2E+04
3	9E+05	1E+06	2E+06	2E+06	9E+05	9E+05	9E+05	3E+06	2E+06
4	8E+04	4E+04	2E+04	2E+04	9E+04	4E+04	4E+04	3E+04	4E+04
5	3E+04	2E+04	2E+04	6E+04	2E+04	2E+04	2E+04	7E+03	2E+04
6	3E+04	4E+04	1E+04	2E+04	1E+04	3E+04	3E+04	2E+04	1E+04
7	2E+04	1E+04	2E+04	1E+04	2E+04	2E+04	2E+04	9E+03	2E+04
8	2E+04	2E+04	2E+04	2E+04	1E+04	1E+04	1E+04	1E+04	1E+04

PBM-Referente a dados de médias anuais de chuva de 30 anos de observação

<i>P</i>	<i>m</i> =1,2	<i>m</i> =1,3	<i>m</i> =1,4	<i>m</i> =1,5	<i>m</i> =1,6	<i>m</i> =1,7	<i>m</i> =1,8	<i>m</i> =1,9	<i>m</i> =2,0
2	6E+04	6E+04	2E+04	1E+04	5E+04	1E+04	5E+04	5E+04	1E+04
3	8E+05	8E+05	3E+05						
4	2E+04	1E+04	8E+03	2E+04	2E+04	2E+04	2E+04	1E+04	2E+04
5	1E+04	9E+03	3E+04	3E+04	4E+04	2E+04	2E+04	2E+04	2E+04
6	9E+03	6E+03	4E+04	1E+04	6E+03	1E+04	8E+03	9E+03	7E+03
7	1E+04	8E+03	9E+03	8E+03	8E+03	1E+04	1E+04	5E+03	7E+03
8	5E+03	6E+03	6E+03	5E+03	6E+03	1E+04	8E+03	1E+04	5E+03

PBM-Referente a dados de médias anuais de chuva de toda a série observada

<i>P</i>	<i>m</i> =1,2	<i>m</i> =1,3	<i>m</i> =1,4	<i>m</i> =1,5	<i>m</i> =1,6	<i>m</i> =1,7	<i>m</i> =1,8	<i>m</i> =1,9	<i>m</i> =2,0
2	2E+03	1E+04	8E+03	2E+04	2E+04	2E+04	2E+04	1E+04	2E+04
3	1E+04	1E+06	3E+04	3E+04	4E+04	2E+04	2E+04	2E+04	2E+04
4	9E+03	6E+03	4E+04	1E+04	6E+03	1E+04	8E+03	9E+03	7E+03
5	4E+04	1E+04	2E+04	2E+04	2E+04	5E+04	4E+04	3E+04	4E+04
6	2E+04	2E+04	2E+04	3E+04	4E+04	2E+04	4E+04	3E+04	4E+04
7	2E+04	2E+04	2E+04	2E+04	1E+04	2E+04	2E+04	2E+04	1E+04
8	2E+04	1E+04	2E+04	5E+03	2E+04	2E+04	7E+03	5E+04	9E+03

Pelos valores de *K* expostos da Tabela 5, verifica-se que a formação com três grupos (*p*) obteve o maior índice PBM, em todas as sete situações consideradas (5, 10, 15, 20, 25, 30 anos e toda a série). Esse resultado indica que a composição de séries histórica, com períodos não uniformes, não provoca resultados diferentes na formação dos grupos para essa região.

O melhor agrupamento, indicado pelo índice *PBM*, entre as 441 simulações foi para a formação de três agrupamentos, tendo como $K = 2E+07$ e parâmetro de fuzificação (m) de 1,9, para os totais médios anuais das séries históricas com 20 anos de dados.

Nota-se que para o agrupamento desses postos pluviométricos, o índice *PBM* foi maior para todos os agrupamentos com três grupos ($p = 3$) e menores para os demais grupos, indicando que o agrupamento mais adequado seria o formado por três grupos com o máximo índice *PBM*.

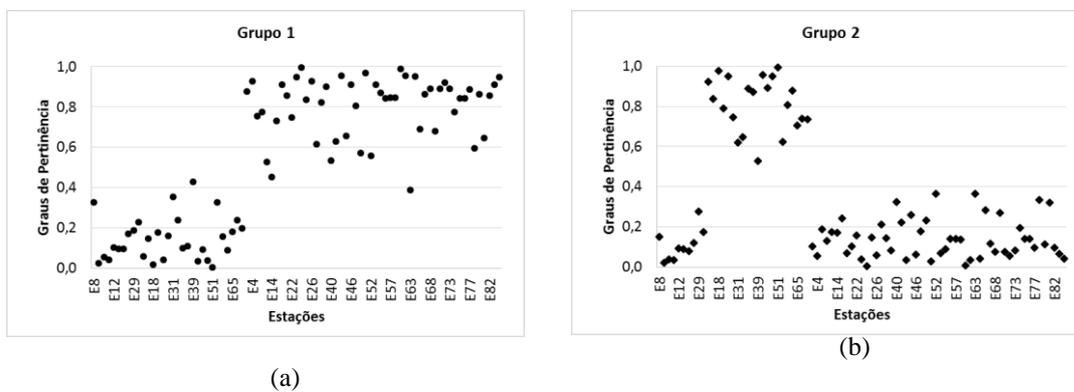
Os grupos formados representam as regiões homogêneas de chuva. O grupo I é formado por 52 estações, o grupo II é formado por 21 estações e o grupo III é formado por 10 estações de acordo com seus graus de pertinência.

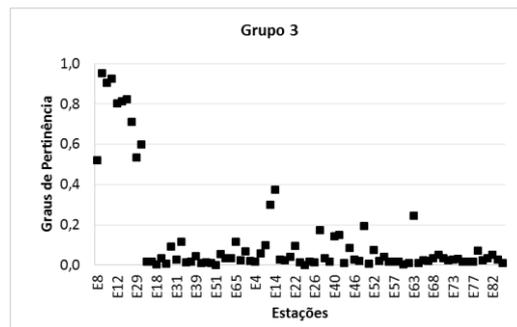
Um dos resultados do algoritmo FCM são os graus de pertinência dos elementos agrupados. Esse grau de pertinência refere-se a probabilidade de um elemento pertencer ou não a um determinado grupo. Quanto mais próximo de 1, maior a chance de pertencer a um grupo. Assim, todas as estações pluviométricas, representadas por suas características de chuva anual média, altitude e localização, apresentaram um grau de pertinência para cada grupo.

A estação Alô Brasil (01251000), por exemplo, localizada no estado do Mato Grosso, apresenta para o Grupo 1, grau de pertinência de 0,88, para o grupo 2 de 0,10 e para o grupo 3 de 0,02. Para Mingoti (2005), isso significa que essa estação possui aproximadamente 88% de probabilidade de pertencer ao grupo 1.

Na Figura 12 apresenta-se o comportamento dos graus de pertinência das estações. A lista completa dos graus de pertinência de cada estação encontra-se no Apêndice D.

Figura 12 – Graus de Pertinência das Estações.





(c)

5.1.2 Formação das Regiões Homogêneas

Os grupos formados pela análise de agrupamento *fuzzy c-means*, são considerados como as regiões homogêneas de chuva (Figura 13). A Região I é formada pelas estações que apresentam chuvas anuais com média de 1.625 mm, mínima de 1.187 mm e máxima de 1.990 mm. Essas estações estão concentradas na porção central e sudoeste da RHTA, mais especificamente nas sub bacias do Alto Tocantins e Araguaia, onde predomina o bioma do cerrado, de clima tropical sazonal.

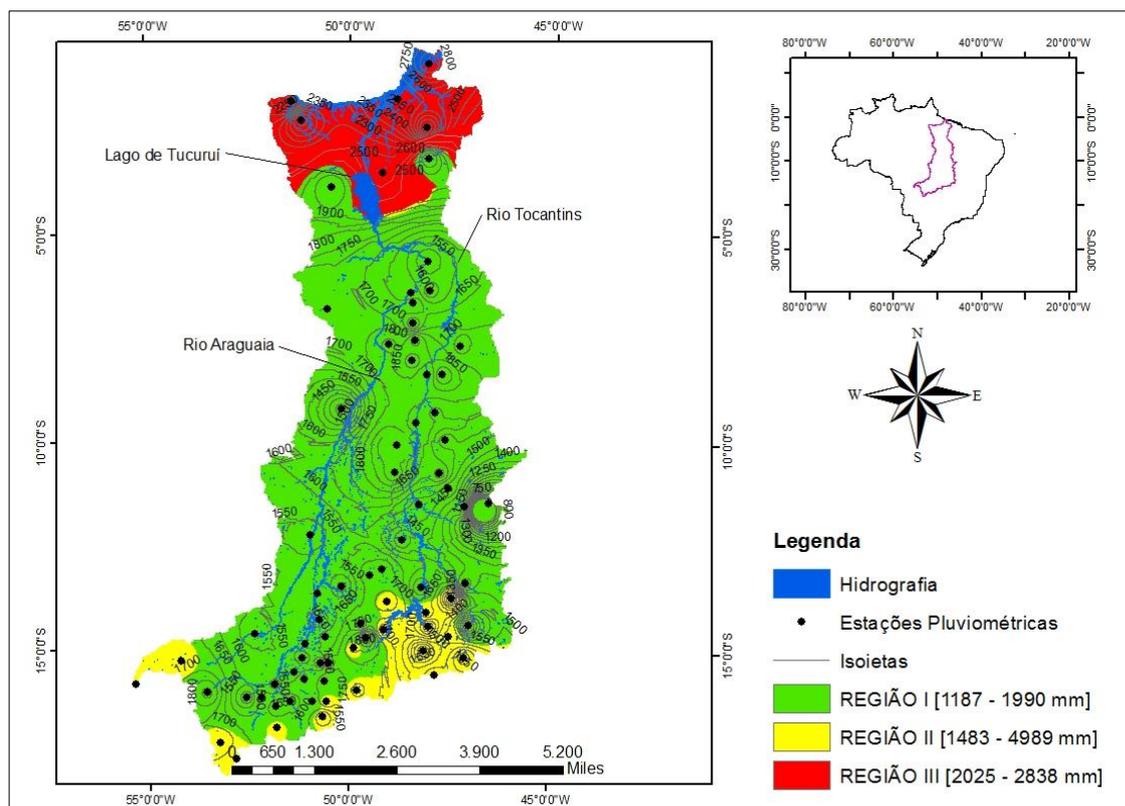
A Região II é formada por estações com médias de chuvas anuais de aproximadamente 1.700 mm, com mínima de 1.349 mm e máxima de 1.989 mm. A maior parte das estações desse grupo, estão distribuídas na porção sul e sudeste da RHTA, especificamente em parte dos estados de Goiás e o Distrito Federal. O bioma predominante nesta região, também é o cerrado.

A Região III é formada por estações que apresentam chuvas de maior intensidade, com média de 2.400 mm, mínima de 2.025 mm e máxima de 2.843 mm. As estações deste grupo estão concentradas na porção norte da RHTA, na sub-bacia Baixo Tocantins, onde predomina o bioma Amazônico, com clima quente e úmido e alto índice pluviométrico.

As Regiões I e II podem ser caracterizadas por regiões com volumes aproximados de precipitações anuais médios de 1600 e 1700 mm, respectivamente, enquanto que a Região III apresenta volumes de precipitações de 2400 mm.

O agrupamento (Figura 13) representou de forma satisfatória o comportamento da precipitação nesta região, pois estudos já existentes de precipitação, realizados pela Embrapa (1994), pela Agência Nacional de Águas - ANA (2009) e por Loureiro *et. al.* (2014), nesta região, já haviam identificado os totais pluviométricos semelhantes nestas regiões, através de métodos de interpolação geoestatístico.

Figura 13 - Regiões Homogêneas de Precipitação da RHTA.



O comportamento da precipitação apresentado nestas regiões (Tabela 7), estão sob influência de diferentes fatores climáticos, como a temperatura, a latitude, a umidade e a pressão, que juntos, condicionam a distribuição de chuva nas regiões formadas.

Tabela 7 – Comportamento das Variáveis nas Regiões Homogêneas.

Regiões	NE	%	Precipitação (mm)	Altitude (m)	Média Precipitação	Média Altitude
I	52	62,7	1187 - 1990	20 - 450	1625	281
II	21	25,3	1349 - 1989	467 - 1242	1700	709
III	10	12,0	2025 - 2838	11 - 683	2400	154
Total	83	100				

NE-Número de Estações.

5.1.3 Teste de Heterogeneidade (H)

O cálculo da medida de heterogeneidade, conforme estabelecido no item 4.6, foi feito pela comparação das variâncias entre os Cv-L observado e simulado. Dessa forma, procedeu-se ao cálculo da medida de heterogeneidade de acordo com a Equação 4.28.

Os valores do quocientes de momentos-L, para algumas estações estão apresentados na Tabela 8, os valores completos estão dispostos no Apêndice D.

Tabela 8 – Quocientes de Momentos-L para as estações estudadas.

ID	N _{dados}	CV-L	Assimetria-L	Curtose-L
E1	19	0,130	-1,595	-4,635
E2	17	0,072	-2,767	-5,087
E3	18	0,081	-2,417	-4,844
⋮	⋮	⋮	⋮	⋮
E21	31	0,105	-0,074	-5,934

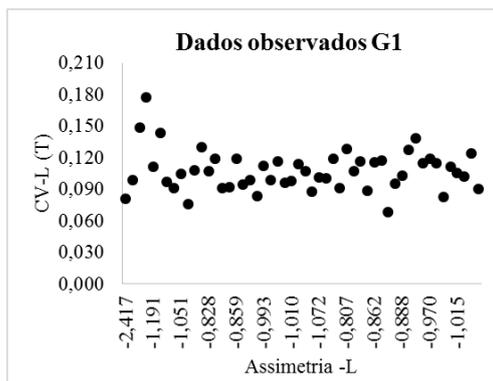
Os valores da Tabela 9 permitem o cálculo dos quocientes-L regionais por meio da média ponderada tomando o tamanho da série, conforme mencionado no tópico 4.5 Na verificação do Teste de Heterogeneidade (H), foi obtido o valor de 0,047 para a Região I, -0,0049 para a Região II e -0,7874 para a Região III, conferindo regiões aceitavelmente Homogêneas, pois $H < 1$.

Hosking & Wallis (1997), afirmam que quando H apresenta valores negativos, significa que há menos dispersão entre os valores amostrais de Cv-L do que se esperaria de uma região homogênea.

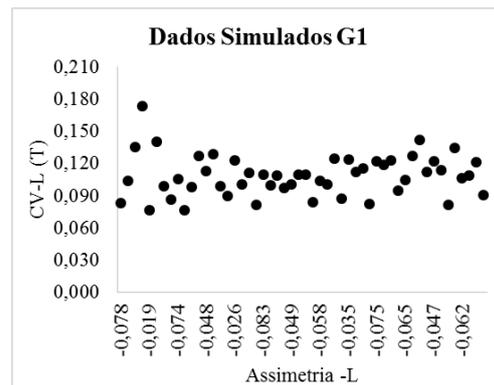
O significado da medida de heterogeneidade pode ser visualizado através dos diagramas de quocientes de momentos-L (Figura 14). Em diagramas como esses, uma região possivelmente homogênea apresentaria, por exemplo, Cv-Ls amostrais, menos dispersos possíveis, do que aqueles obtidos por simulação. Em termos quantitativos essa ideia básica pode ser traduzida pela diferença centrada entre as dispersões, observadas e simuladas.

A dispersão nas regiões simuladas, para os quocientes de momento-L, demonstra que não ocorreu dispersão dos dados, e, portanto, não existem postos com média muito acima ou abaixo dos valores já esperados. Assim, as dispersões simulada e observada são semelhantes, configurando uma região aceitavelmente homogênea.

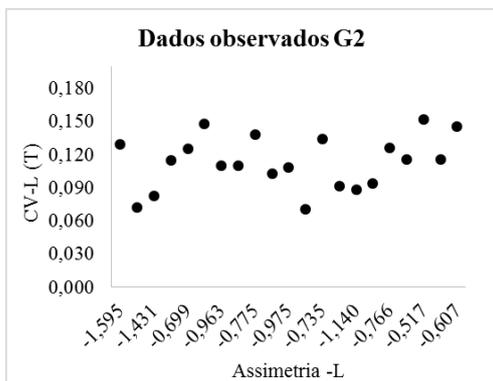
Figura 14 – Dispersão dos quocientes CV-L e Assimetria-L.



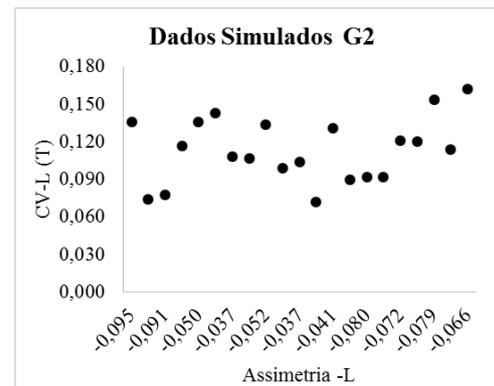
(a)



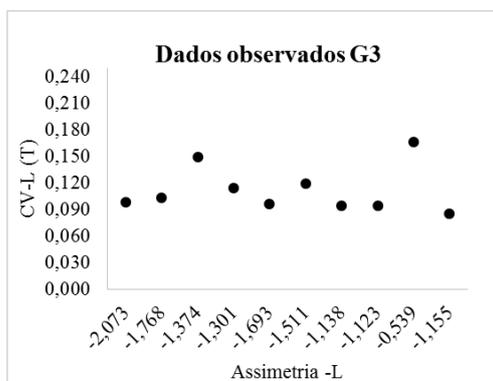
(b)



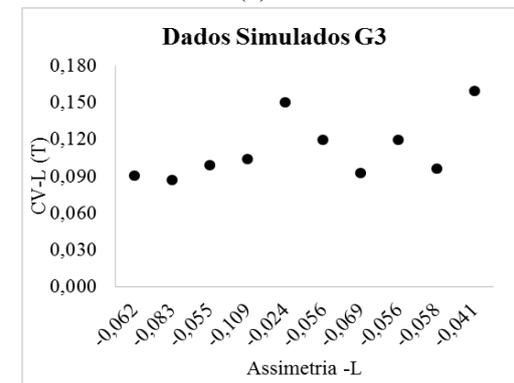
(c)



(d)



(e)



(f)

5.2 ESTIMATIVAS DE CHUVAS POR FUNÇÕES DE PROBABILIDADE

Com o objetivo de determinar a melhor função de distribuição de probabilidade, para a estimativa de chuva nas regiões homogêneas formadas, foi utilizado o código computacional FDP (APENDICE C), para a geração das funções de probabilidade Normal, Log-normal, Gama, Gumbel, Exponencial e Weibull.

Para a calibração, foram adotadas as médias de precipitação anual e mensal de cada uma das 83 estações pluviométricas utilizadas no agrupamento *fuzzy c-means*. A partir desses dados foram geradas as funções de probabilidade para as séries de precipitação média anual e mensal de cada região homogênea. Seguidamente, para a calibração da melhor função foi realizado o teste do Qui Quadrado (X^2) para avaliar o ajuste ou aderência das frequências das funções de distribuição de Probabilidade Normal, Log-normal, Gama, Gumbel, Exponencial e Weibull.

Na aplicação do teste Qui-quadrado, o número de graus de liberdade será igual a dois e o nível de significância adotado, é o de 5%, por ser o valor mais usual na aplicação deste teste. Sendo assim, o valor do X^2 tabelado é igual a 5,99 para todas as funções.

5.2.1 Calibração da Precipitação Média Anual

Para as médias anuais de precipitação, dentre os modelos de probabilidades testados, as funções Gumbel e Exponencial não se ajustaram adequadamente, isso pode ser comprovado pelos valores do X^2 (Tabela 9), que foram maiores do que o tabelado, e pelo comportamento das funções de probabilidade, de acordo com as Figuras 15, 16 e 17. As funções de distribuição de probabilidade Normal, Gama e Weibull, tiveram boa aderência, pois seus valores do X^2 foram todos abaixo do valor tabelado de 5,99, porém a melhor foi a Log-Normal, conferindo a ela um bom modelo de probabilidade de ocorrência de precipitação anual nas referidas regiões.

Tabela 9 – Teste do X^2 para as funções de probabilidade.

R. Homogênea	Resultado do X^2					
	Normal	Log-Normal	Gama	Gumbel	Exponencial	Weibull
I	0,79	0,54	3,57	38,15*	28,87*	2,93
II	0,46	0,36	4,39	28,21*	56,23*	3,95
III	0,09	0,04	5,18	29,52*	42,84*	3,08

(*) valores inadequados.

Figura 15 – Funções de distribuição de Probabilidade na simulação de médias anuais de precipitação para a Região Homogênea I

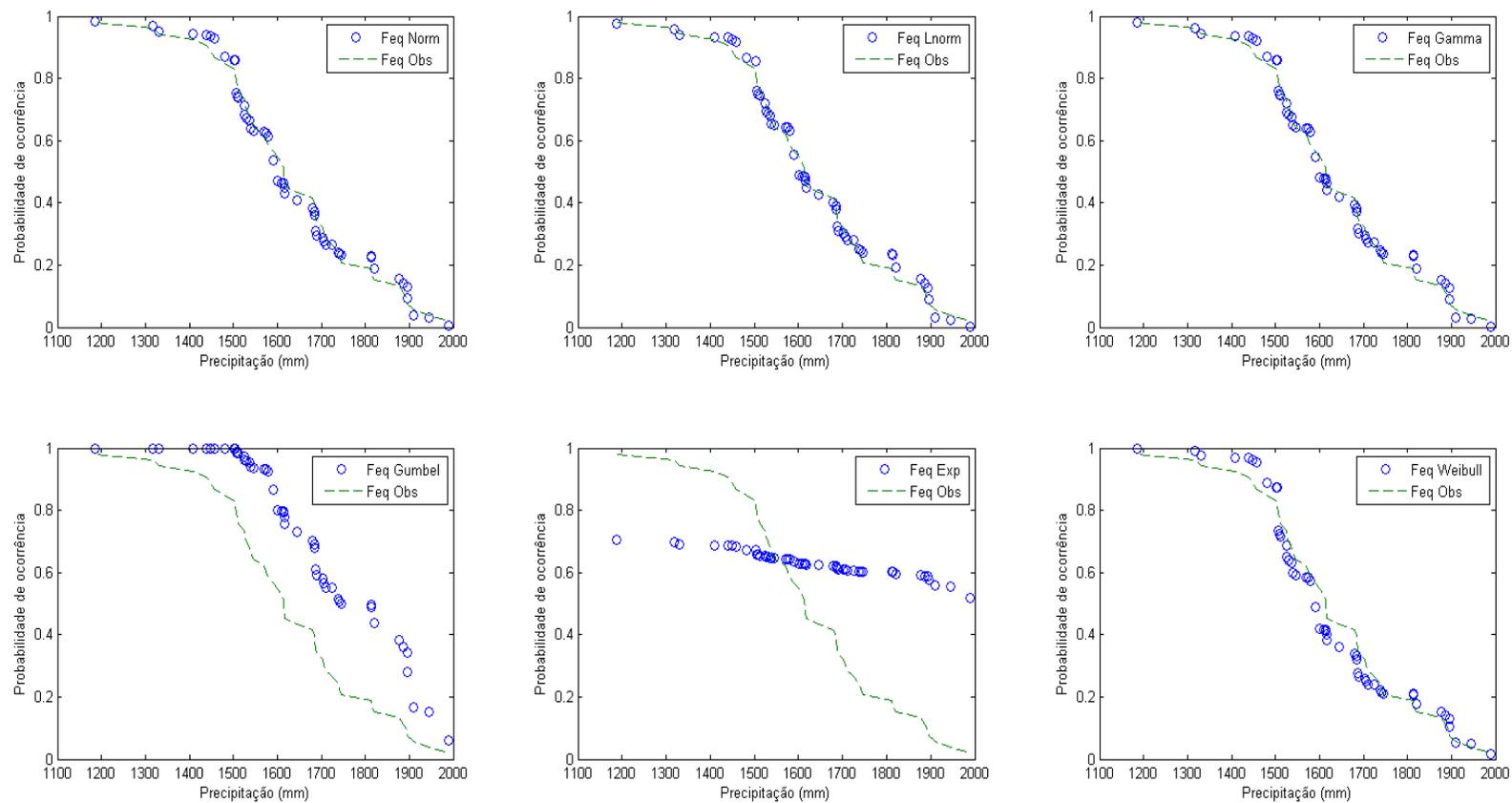


Figura 16 – Funções de distribuição de Probabilidade na simulação de médias anuais de precipitação para a Região Homogênea II

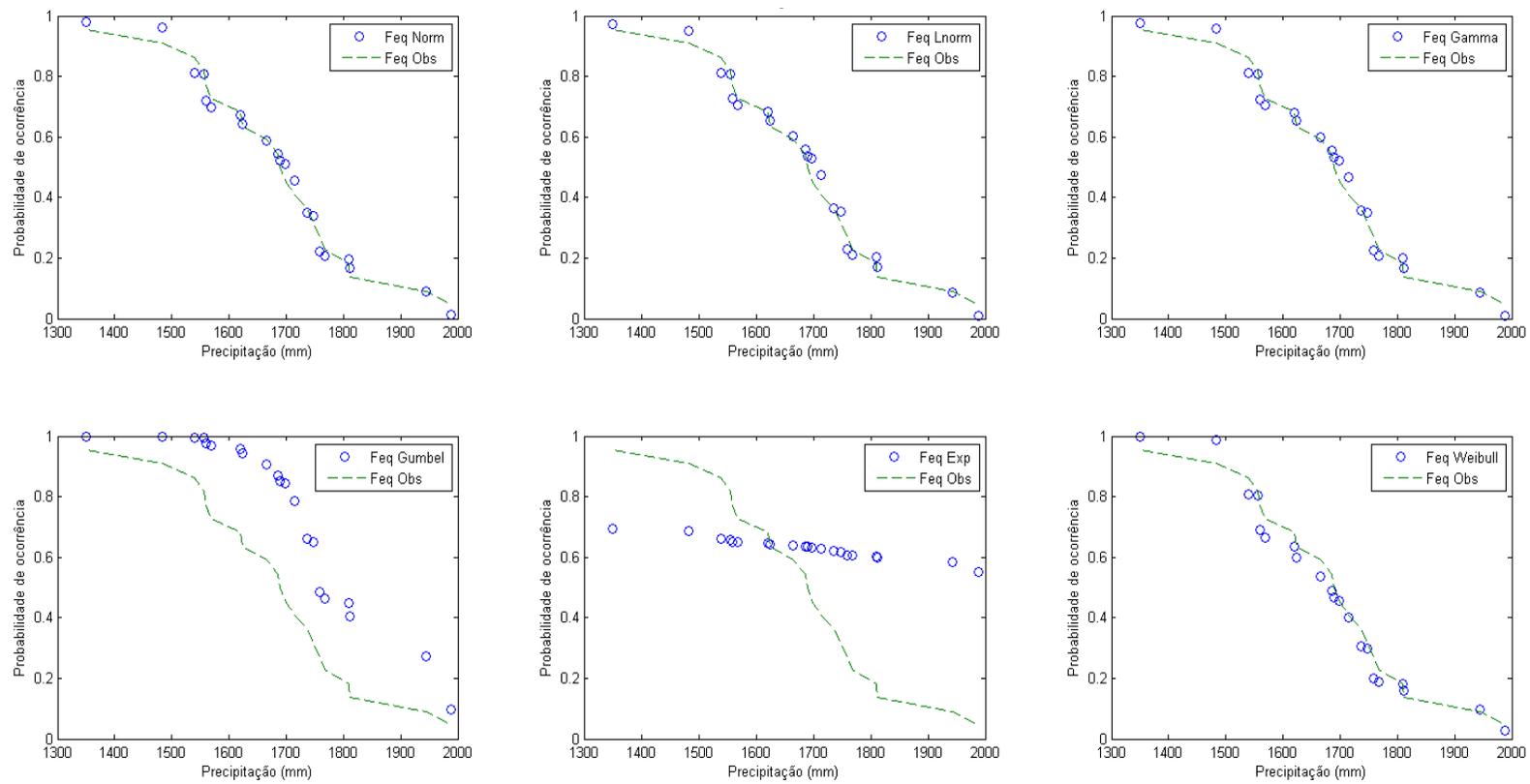
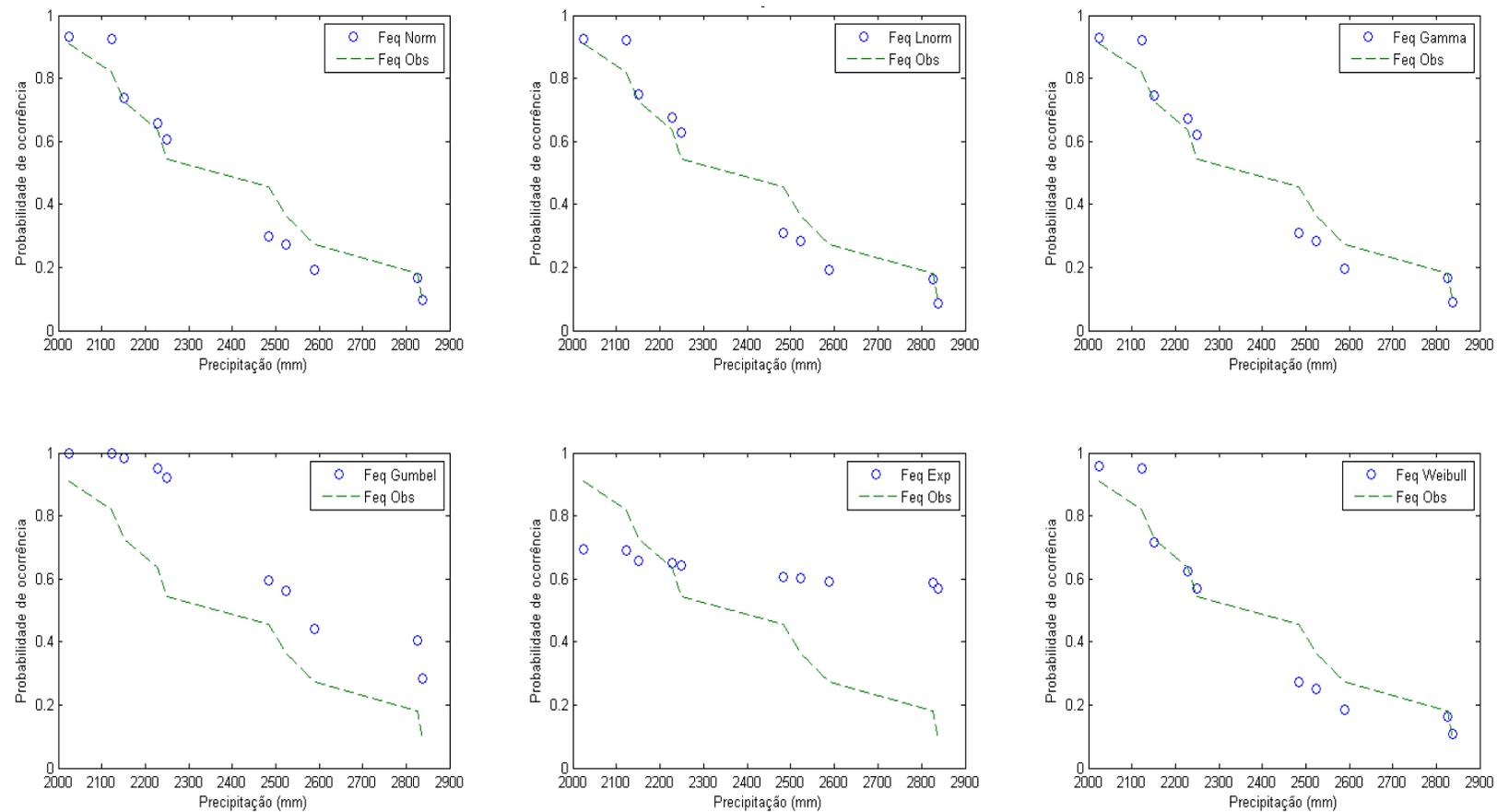


Figura 17 – Funções de distribuição de Probabilidade na simulação de médias anuais de precipitação para a Região Homogênea III



Dentre todas as funções de probabilidade usadas na calibração das médias anuais, a Log-normal foi a que apresentou o melhor ajuste pelo teste de aderência, ratificando a boa aplicação da distribuição Log-Normal para séries associadas a totais máximos.

Segundo Mello (2013) e Tucci (2012), as FDP's Gumbel e Exponencial tem melhor desempenho na estimativa de valores extremos para máximos ou mínimos, fato que ficou comprovado pelo seu desempenho não adequado, na estimativa de totais médios de precipitação.

5.2.2 Calibração da Precipitação Média Mensal

Como a proposta do trabalho, é também estudar a precipitação média mensal das regiões homogêneas da RHTA, foram realizadas simulações das funções de probabilidade para a média mensal de precipitação de cada estação para as Regiões I, II e III.

A Tabela 10 apresenta os resultados do teste Qui-quadrado, os quais foram obtidos com a simulação das funções de probabilidades para as médias mensais. Observa-se que as funções Gumbel e Exponencial não apresentaram um bom ajuste para nenhum dos meses em nenhuma região, pois os valores do X^2 foram superiores ao valor tabelado (5,99).

Tabela 10 – Valores do X^2 das simulações das séries mensais – Região Homogênea I.

	Valores do X^2					
	Normal	Log-normal	Gama	Gumbel	Exponencial	weibull
Jan	0,89	0,52	0,72	12,8	19,3	2,32
Fev	3,85	3,1	3,31	78,23	27,4	6,7
Mar	1,06	1,61	1,02	45,23	12,35	1,31
Abr	10,11	6,15	5,67	19,23	5,01	5,74
Mai	15	4,64	1,16	68,21	1,28	1,63
Jun	46,58	2,23	2,14	75,52	2,43	1,44
Jul	49,58	2,52	2,02	88,17	2,79	1,35
Ago	12,15	1,61	5,11	62,32	18,12	6,49
Set	7,75	2,11	5,19	56,31	16,2	8,1
Out	2,88	5,18	2,52	17,32	13,62	3,86
Nov	4,86	15,39	6,09	23,67	17,65	4,33
Dez	1,09	3,19	2,41	24,12	15,25	0,13
Média	13,0	4,0	3,1	47,6	12,6	3,6

Na Região I a função Gama apresentou um bom ajuste, apesar de em alguns meses o valor do X^2 ser maior que o tabelado. Porém a função Gama, apresentou,

somente para o mês de novembro, valor do X^2 superior ao tabelado, enquanto as demais funções apresentaram valores do X^2 superior em mais de um mês. Sendo assim a função de distribuição de probabilidade Gama é indicada como a mais adequada para a simulação das probabilidades de ocorrência das precipitações médias mensais na região homogênea I.

Na Região II as funções Normal, Log-normal, Gama e Weibull, apresentaram valores menores do X^2 tabelado, conferindo um bom ajuste (Tabela 11). Somente para as séries do mês de julho, o valor do X^2 foi superior para todas as funções. A função que apresentou menores valores do X^2 foi a Gama, e em seguida a Weibull, com média do X^2 de 1,99 e 2,09 respectivamente. Logo, indica-se para a região homogênea II, a função de probabilidade Gama, como modelo mais adequado para a simulação das probabilidades de ocorrência de precipitação média mensal.

Tabela 11 – Valores do X^2 das simulações das séries mensais – Região Homogênea II

	.Valores do X^2					
	Normal	Log-Normal	Gama	Gumbel	Exponencial	Weibull
Jan	0,85	0,52	0,56	28,23	17,08	2,16
Fev	2,17	2,43	2,36	42,12	15,23	1,97
Mar	1,29	1,14	1,21	28,24	16,31	1,92
Abr	3,01	2,76	2,48	34,21	31,27	3,86
Mai	1,68	2,54	2,14	36,23	10,92	1,3
Jun	3,41	1,84	1,83	18,36	1,02	1,53
Jul	7,59	7,94	6,94	26,35	7,12	6,14
Ago	0,75	1,41	0,89	13,1	4,42	0,58
Set	2,91	2,05	2,026	16,41	11,63	2,57
Out	0,91	1,03	0,96	19,27	14,23	0,88
Nov	0,37	0,36	0,37	24,24	10,63	0,65
Dez	1,07	1,2	1,1	19,63	80,14	1,57
Med	2,2	2,1	1,9	25,5	18,3	2,1

Na Região Homogênea III a funções Normal, Log-normal e Gama, apesar de não se ajustarem bem em alguns meses (abril, maio, junho e julho), obtiveram um bom ajuste nos demais meses. A função Weibull, foi a que apresentou menores valores do X^2 em relação ao tabelado. Assim essa função pode ser indicada como modelo adequado para simular a probabilidade de ocorrência de chuvas mensais na Região homogênea III (Tabela 12).

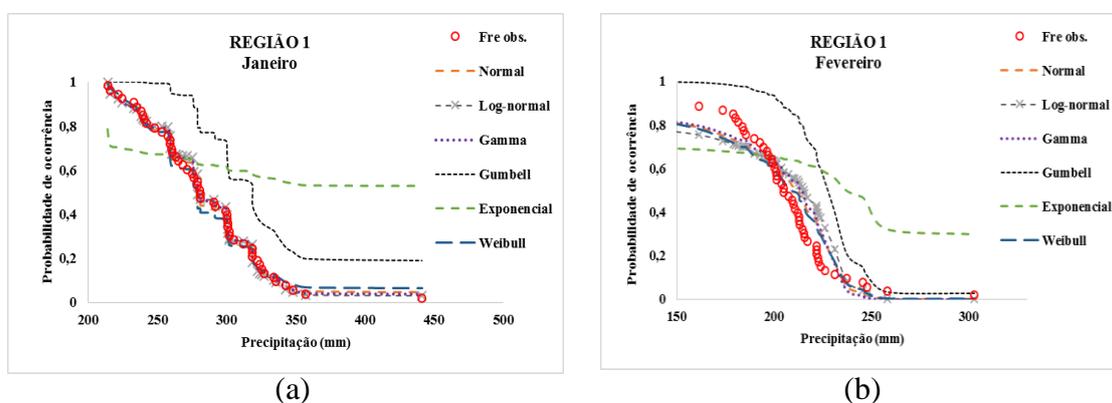
A função Weibull, geralmente, é usada em séries históricas de valores mínimos. No caso das séries médias mensais apresentarem valores menores nos meses, principalmente na estiagem, pode ser uma justificativa para essa função ter se adequado muito bem, para os dados mensais de precipitação, na Região homogênea III.

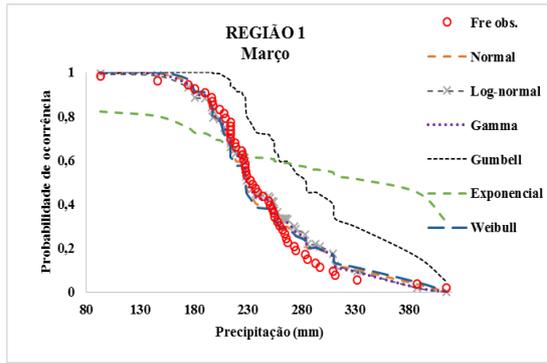
Tabela 12 – Valores do X^2 das simulações das séries mensais – Região Homogênea III

	Valores do X^2					
	Normal	Log-normal	Gama	Gumbel	Exponencial	Weibull
Jan	0,083	0,22	0,17	17,24	25,86	0,005
Fev	0,016	0,06	0,04	16,37	21,54	1,23
Mar	0,05	0,23	0,14	12,93	19,29	0,09
Abr	6,5	8,07	7,67	11,7	16,32	5,94
Mai	0,37	5,58	1,9	4,8	5,5	1,52
Jun	0,196	8,26	1,76	5,96	2,59	1,56
Jul	0,38	12,54	3,22	15,56	3,53	3,04
Ago	3,9	0,42	0,43	2,54	0,32	0,47
Set	2,53	1,43	1,62	3,49	1,3	2,08
Out	0,16	0,75	0,31	13,5	1,96	0,12
Nov	0,12	1,17	0,55	3,31	1,7	0,3
Dez	0,13	1,2	0,73	3,5	3,08	0,3
Med	1,2	3,3	1,5	9,2	8,6	1,4

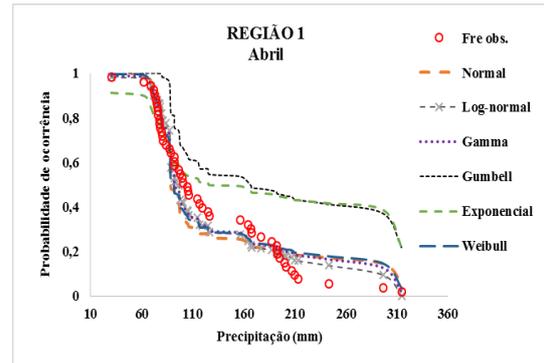
Nas Figuras 18, 19 e 20, observa-se o comportamento das frequências observadas e das funções de distribuições de probabilidade (FDP) simuladas para as precipitações médias mensais de cada região homogênea.

Figura 18 – Distribuições de Probabilidade observada e simulada das médias mensais da Região Homogênea I.

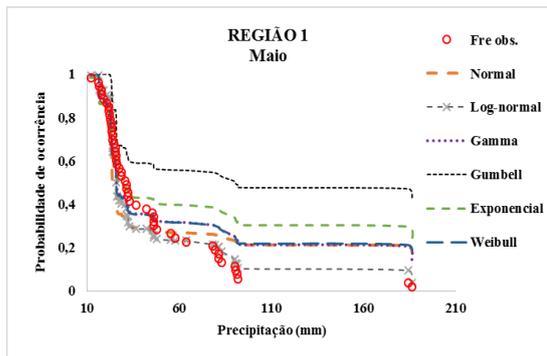




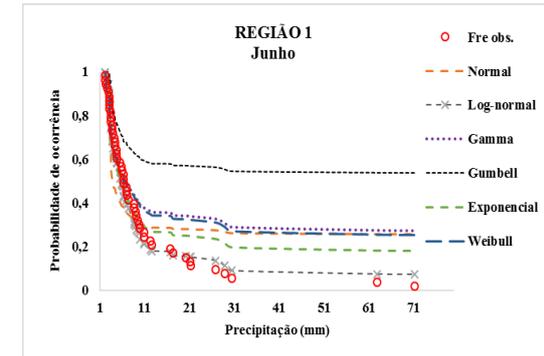
(c)



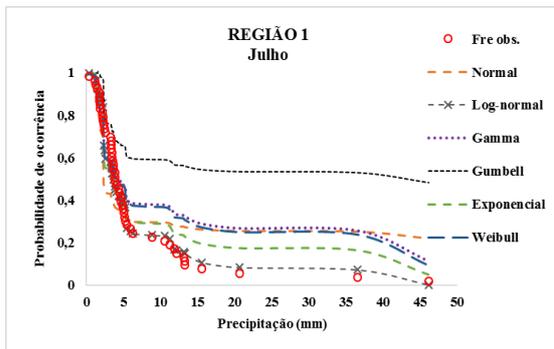
(d)



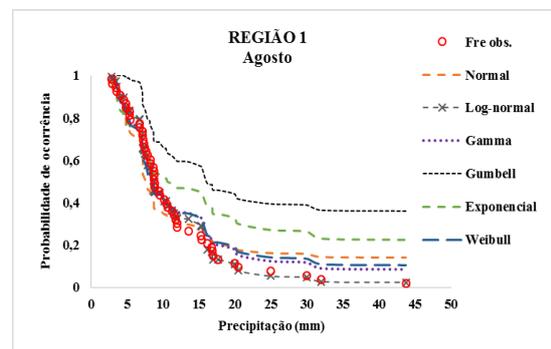
(e)



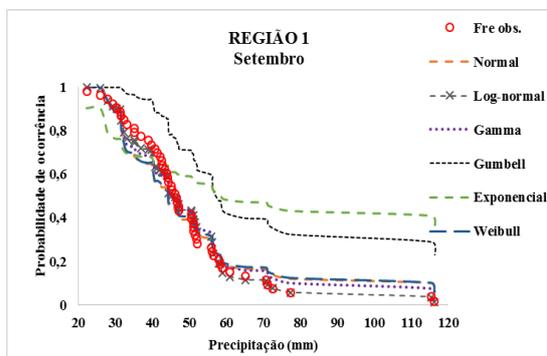
(f)



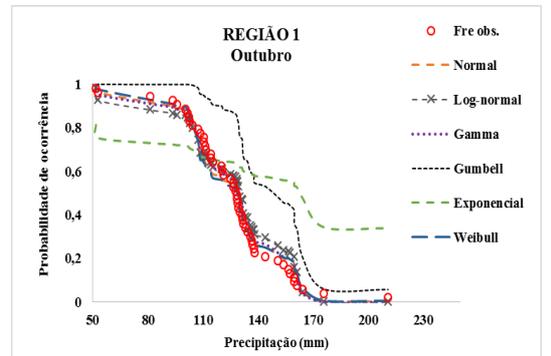
(g)



(h)



(i)



(j)

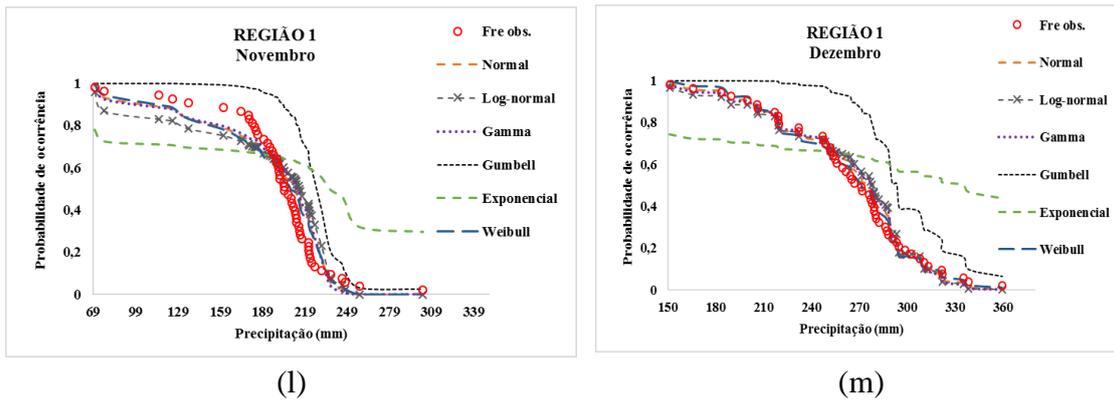
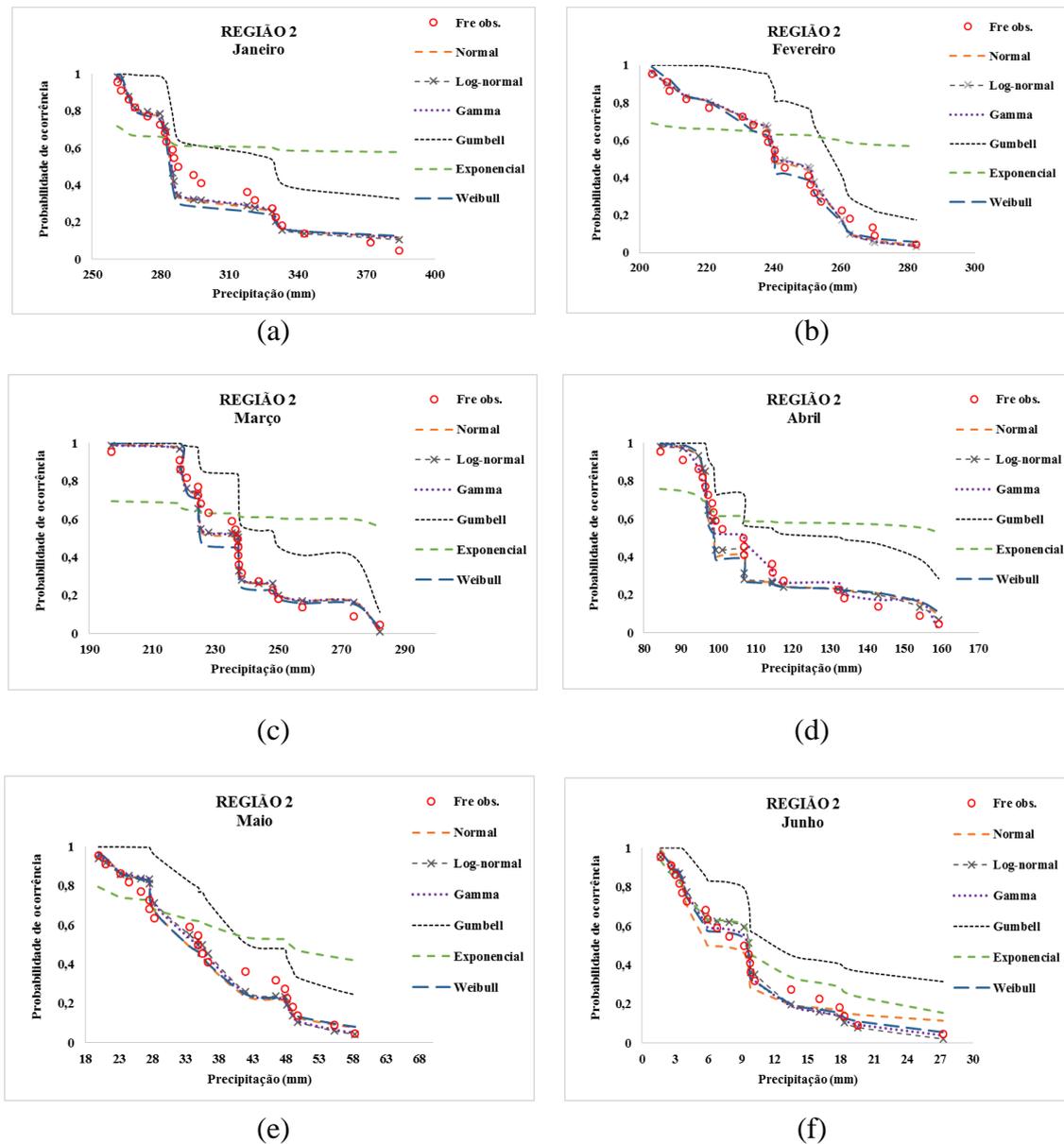
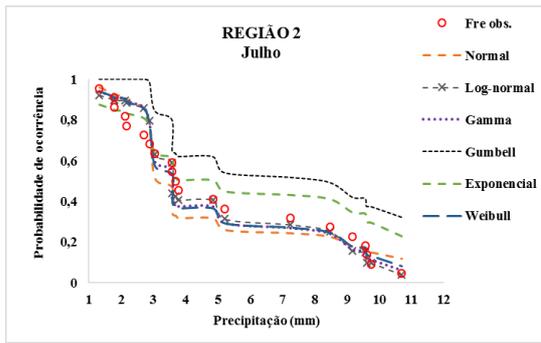
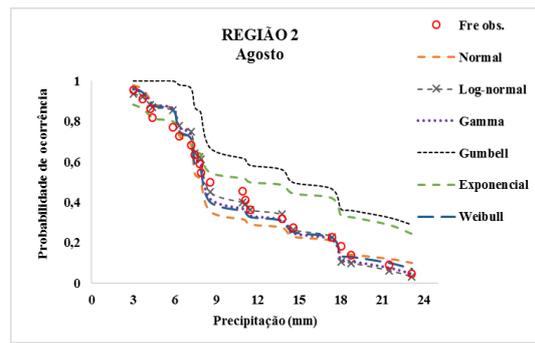


Figura 19 – Distribuições de Probabilidade observada e simulada das médias mensais da Região Homogênea II.

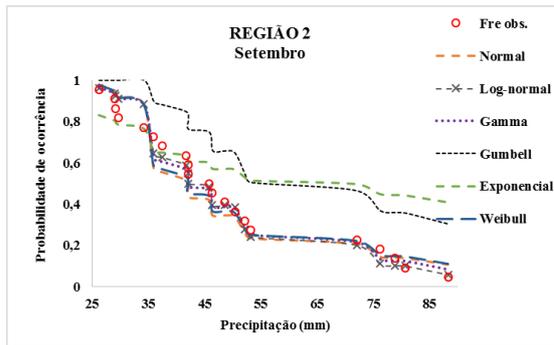




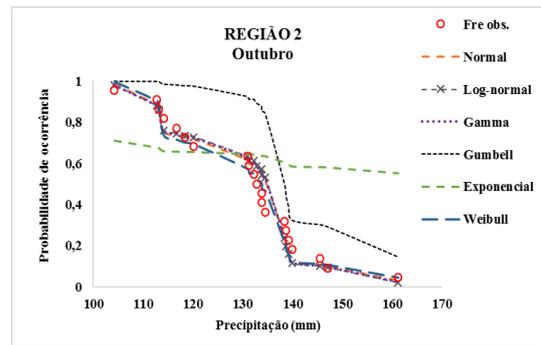
(g)



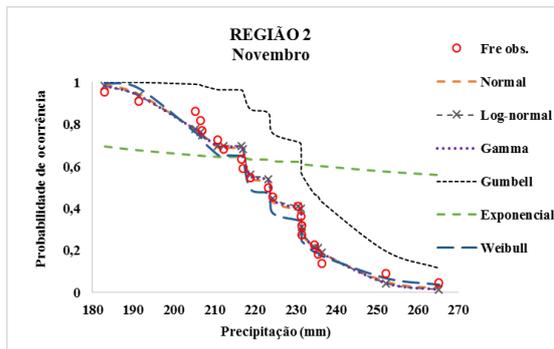
(h)



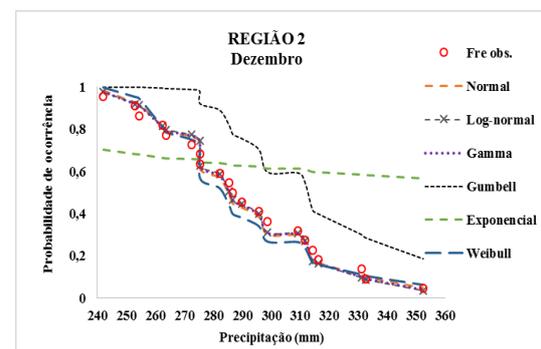
(i)



(j)

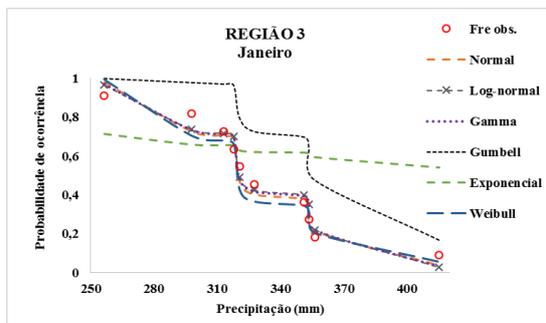


(l)

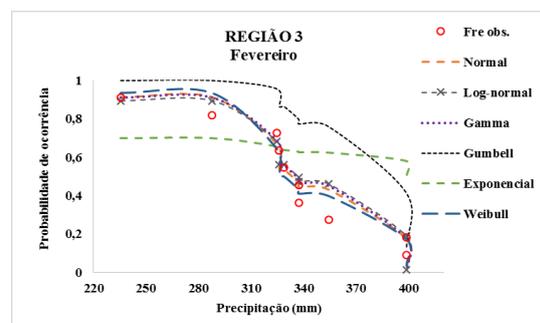


(m)

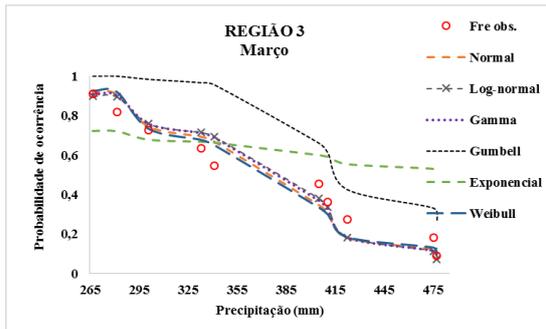
Figura 20 – Distribuições de Probabilidade observada e simulada das médias mensais da Região Homogênea III.



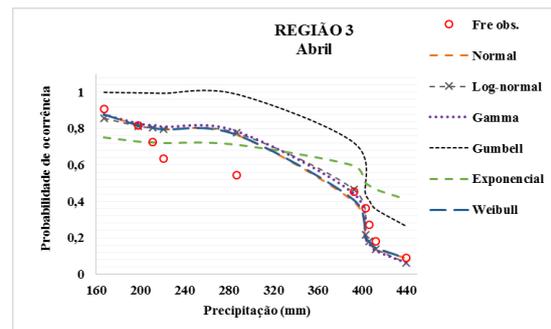
(a)



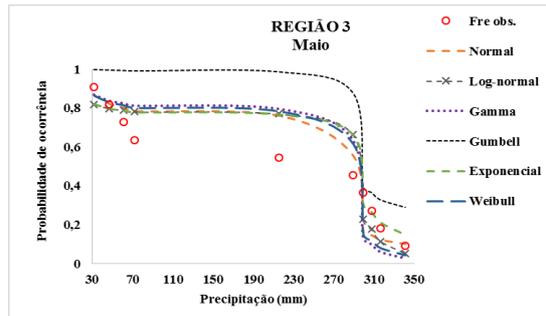
(b)



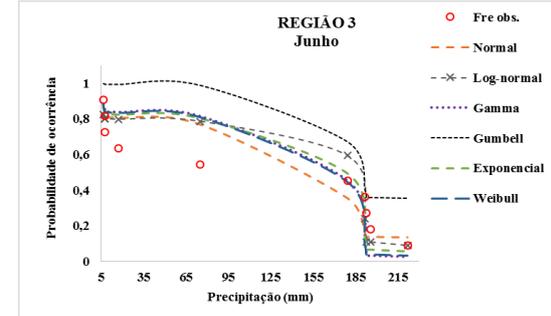
(c)



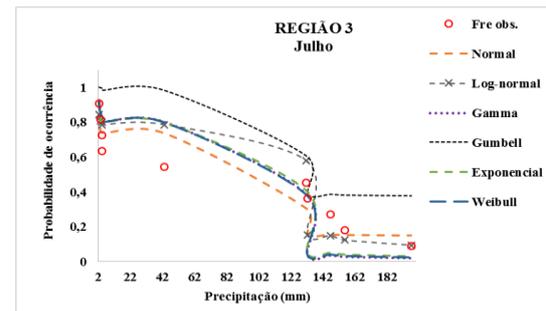
(d)



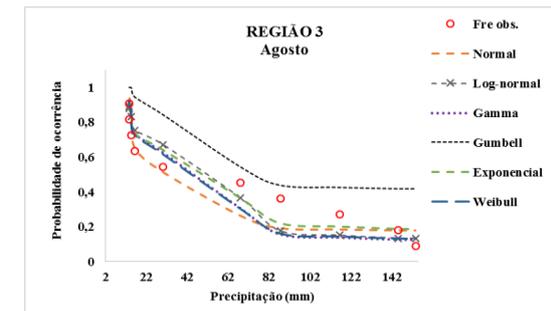
(e)



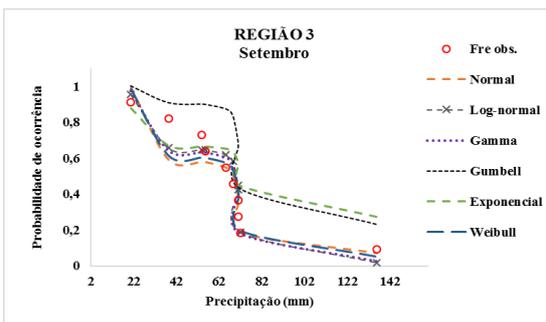
(f)



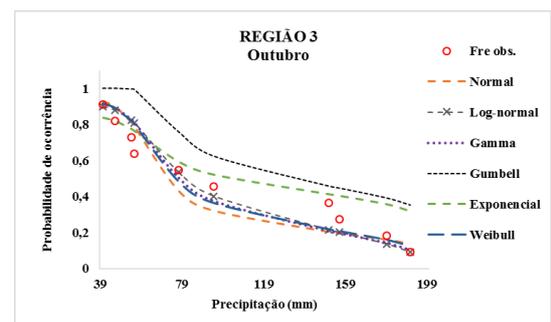
(g)



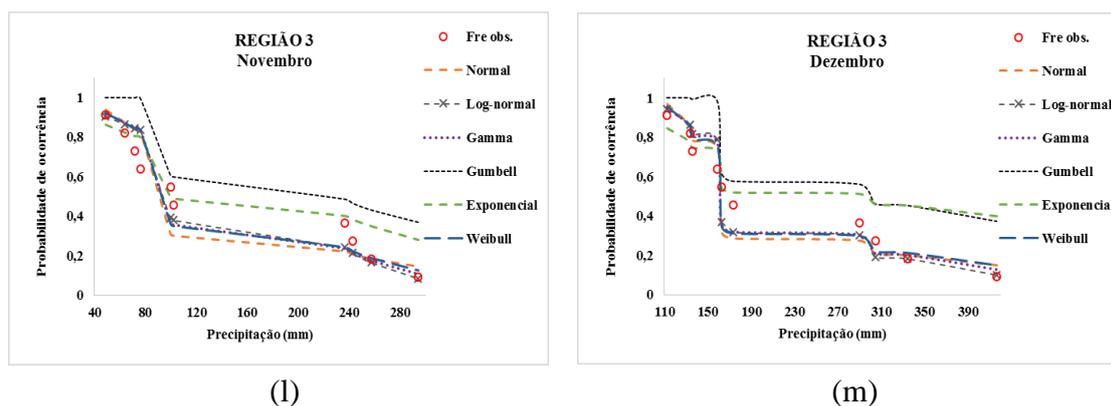
(h)



(i)



(j)



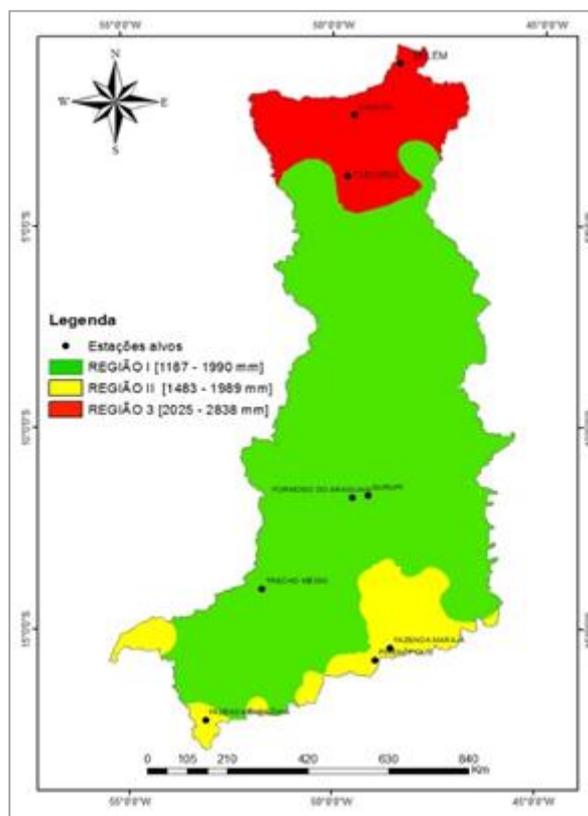
5.2.3 Validação dos Modelos de Probabilidade

Para validar os melhores modelos de probabilidade, obtidos na calibração, foram selecionados nove estações pluviométricas. A escolha das estações basearam-se na sua localização, considerando que as mesmas estivessem localizadas nas regiões homogêneas e possuísem série histórica mínima de 10 anos de dados. A Tabela 13 e a Figura 21 exibem as estações utilizadas para a Validação.

Tabela 13 – Estações Pluviométricas Alvo adotadas na Validação.

Estação	Código	Latitude	Longitude	Região
Tucuruí	00349000	-3,7603	-49,667	III
Cametá	00249001	-2,2394	-49,4997	
Belém	00148001	-1,45	-48,5	
Trecho Médio	01351000	-14,0867	-51,6964	II
Gurupi	01149002	-11,7372	-49,1361	
Formosa do Araguaia	01149001	-11,7972	-49,525	
Faz. Marajá	01548011	-15,5531	-48,5772	I
Pirenópolis	01548004	-15,85	-48,95	
Faz. Babilônia	01753002	-17,3528	-53,0919	

Figura 21 – Localização das Estações Alvo.



O modelo usado na validação para a simulação da probabilidade de ocorrência de chuva média anual, na Região homogênea I, II e III foi a função de distribuição Log-Normal, devido ao seu melhor ajuste na etapa da calibração. Na Tabela 14 são apresentados os valores do Qui-Quadrado obtidos para cada estação alvo, das séries de precipitações médias anuais, obtidas pela função Log_Normal.

Tabela 14 – Valores do X^2 na validação dos modelos probabilísticos para as séries anuais.

Região	Estações	X^2
		Log-Normal
III	Tucuruí	2,37
	Cametá	0,59
	Belém	2,96
II	Trecho Médio	2,95
	Gurupi	0,76
	Formosa do Araguaia	1,8
I	Faz. Marajá	1,7
	Pirenópolis	0,82
	Faz. Babilônia	1,12

O teste Qui-quadrado realizado com as estações alvo foram todos adequados, para a função de distribuição Log-normal, com valores abaixo de 5,99 (tabelado). A aderência adequada desse modelos para as séries de chuvas anuais, pode ser justificada pela sua boa aplicação em séries totais de precipitação, especialmente anuais, conferidas as funções de probabilidade Log-normal.

Nas Figuras 22, 23 e 24 pode ser observado o comportamento da frequência observada e da frequência simulada pela função de probabilidade Log-Normal ajustada as séries médias anuais das estações alvos na Região Homogênea I, II e III, respectivamente.

Figura 22 – Probabilidade de ocorrência de chuva observada e a simulada das Estações Alvos – Região Homogênea I.

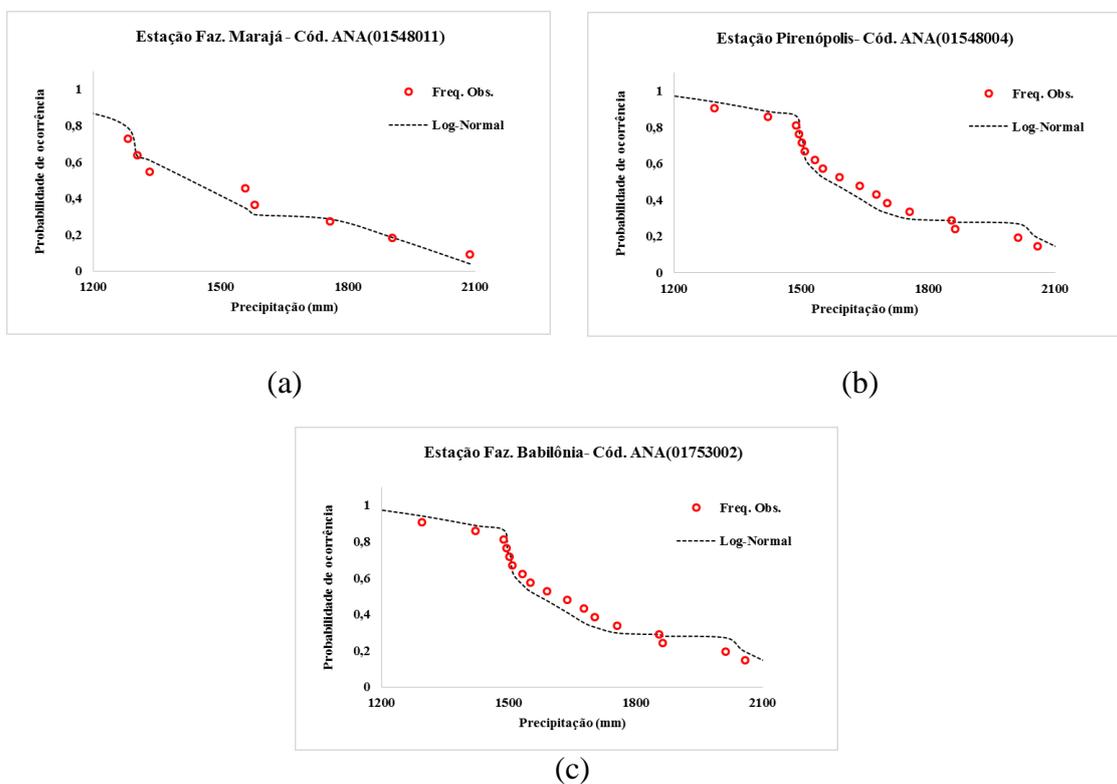
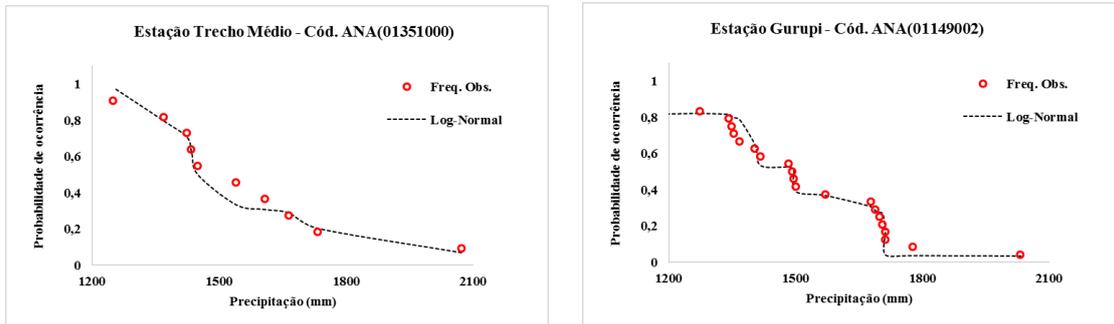
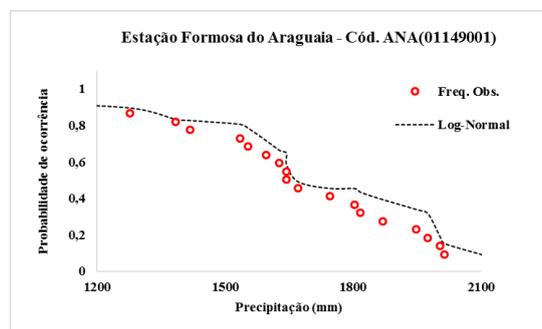


Figura 23 – Probabilidade de ocorrência de chuva observada e simulada das Estações Alvos – Região Homogênea II.



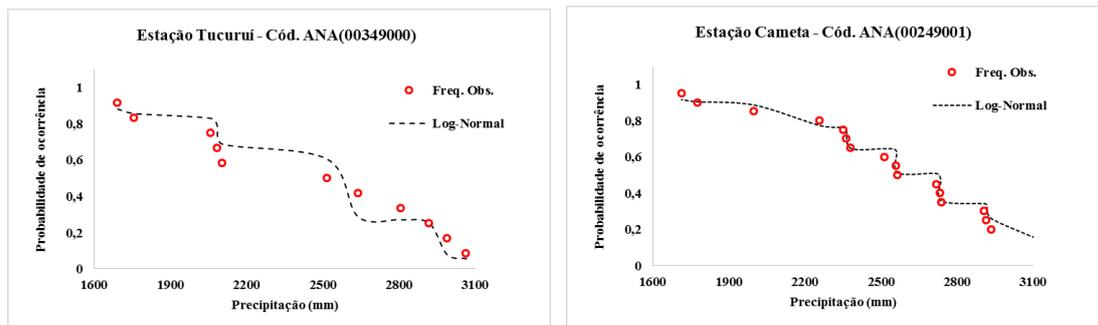
(a)

(b)



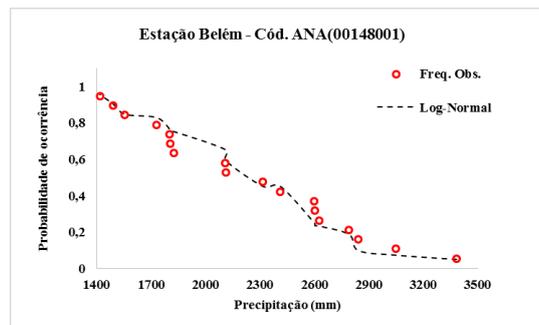
(c)

Figura 24 – Probabilidade de ocorrência de chuva observada e simulada das Estações Alvos – Região Homogênea III



(a)

(b)



(c)

Na Validação das funções de probabilidade para as séries médias mensais, foram adotadas as funções Gama nas regiões I e II e Weibull na região III, pelo seu melhor ajuste apresentado na etapa da calibração. Na Tabela 15 são apresentados os valores do X^2 da etapa de validação para as séries de precipitação média mensal.

Tabela 15 - Valores do X^2 na validação dos modelos probabilísticos para as séries mensais.

Região	Estações	Messes											
		Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez
III	Tucuruí	3,69	1,52	2,23	3,61	4,38	3,59	1,35	1,99	0,44	1,08	3,38	2,65
	Cametá	1,13	1,12	0,79	1,07	2,14	3,35	0,19	1,52	0,43	1,21	0,23	2,00
	Belém	0,21	1,66	1,63	1,67	1,86	3,39	1,97	4,58	2,4	0,29	0,17	0,32
II	Trecho Médio	2,19	0,26	2,31	3,21	1,74	0,04	2,12	3,52	1,63	0,11	2,8	2,25
	Gurupi	0,28	1,17	2,64	0,16	0,59	0,57	0,26	2,52	0,30	0,31	0,61	1,23
	Formosa do Araguaia	3,19	0,92	0,98	3,28	2,39	0,09	3,12	0,75	3,4	3,36	2,58	2,36
I	Faz. Marajá	0,07	0,2	1,46	0,63	0,52	0,53	2,71	1,64	0,14	1,13	2,71	0,06
	Pirenópolis	3,8	2,84	0,03	1,36	0,96	0,46	2,21	1,04	0,19	0,77	0,81	0,50
	Faz. Babilônia	0,57	0,71	0,83	1,10	2,41	0,54	0,29	3,6	0,95	0,07	0,78	0,65

A função de distribuição de probabilidade Gama é muito bem aplicada na estimativa da precipitação e a valores extremos (máximos ou mínimos), fato que pode ter contribuído para o ajuste dessa função, já que os dados de chuvas mensais, ora apresentam valores mínimos nos meses de estiagem e outrora valores máximos no período chuvoso. Em relação à função Weibull, o seu ajuste é melhor em séries com valores mínimos, o que ocorre nas séries mensais, com valores inferiores, principalmente nos períodos menos chuvosos. O gráfico das funções de probabilidade ajustados, para as séries de precipitação média mensal das estações alvos podem ser observadas nas Figuras 25(a) a 25(i).

Os melhores ajustes às distribuições Gama, observados na Região Homogênea I, foram encontrados na simulação da estação alvo Pirenópolis, em Dezembro. Porém, observou-se nas estações alvo Fazenda Marajá, Pirenópolis e Fazenda Babilônia, no mês de agosto, um menor ajuste às distribuições de frequência Gama.

Na Região Homogênea II, a estação que apresentou maior e menor aderência às distribuições Gama foi a estação Gurupi, no mês de Janeiro e a Estação Trecho Médio, no mês de abril, respectivamente.

Na Região Homogênea III, entre as estações alvo observadas, a Estação Cametá, apresentou maior e menor aderência nas distribuições Weibull, no mês de Janeiro e Junho, respectivamente.

Em uma avaliação geral dos gráficos ajustados na validação e comparando as diferenças entre as distribuições de probabilidade simuladas e observadas, nos meses de novembro, dezembro e janeiro, ocorreram os melhores ajustes, enquanto que nos meses de Junho, Julho e Agosto, ocorreu maior diferença entre elas.

Em termos estatísticos, as distribuições de probabilidades ajustadas as estações alvo, representam os modelos que podem prever o comportamento da variável precipitação, considerando principalmente seu caráter contínuo, ou seja, aquele que ocorrerá continuamente no tempo.

Os resultados do teste de ajustamento demonstraram que a metodologia adotada pode ser uma boa opção quando utilizada como ferramenta numérica para determinação de probabilidade de precipitações médias mensais e anuais, e sua aplicação consegue estimar as frequências de eventos que ainda não foram registrados.

Figura 25 (a) – FDP's ajustadas
Estação Fazenda Marajá Cód. ANA 01548011 (Região Homogênea I)

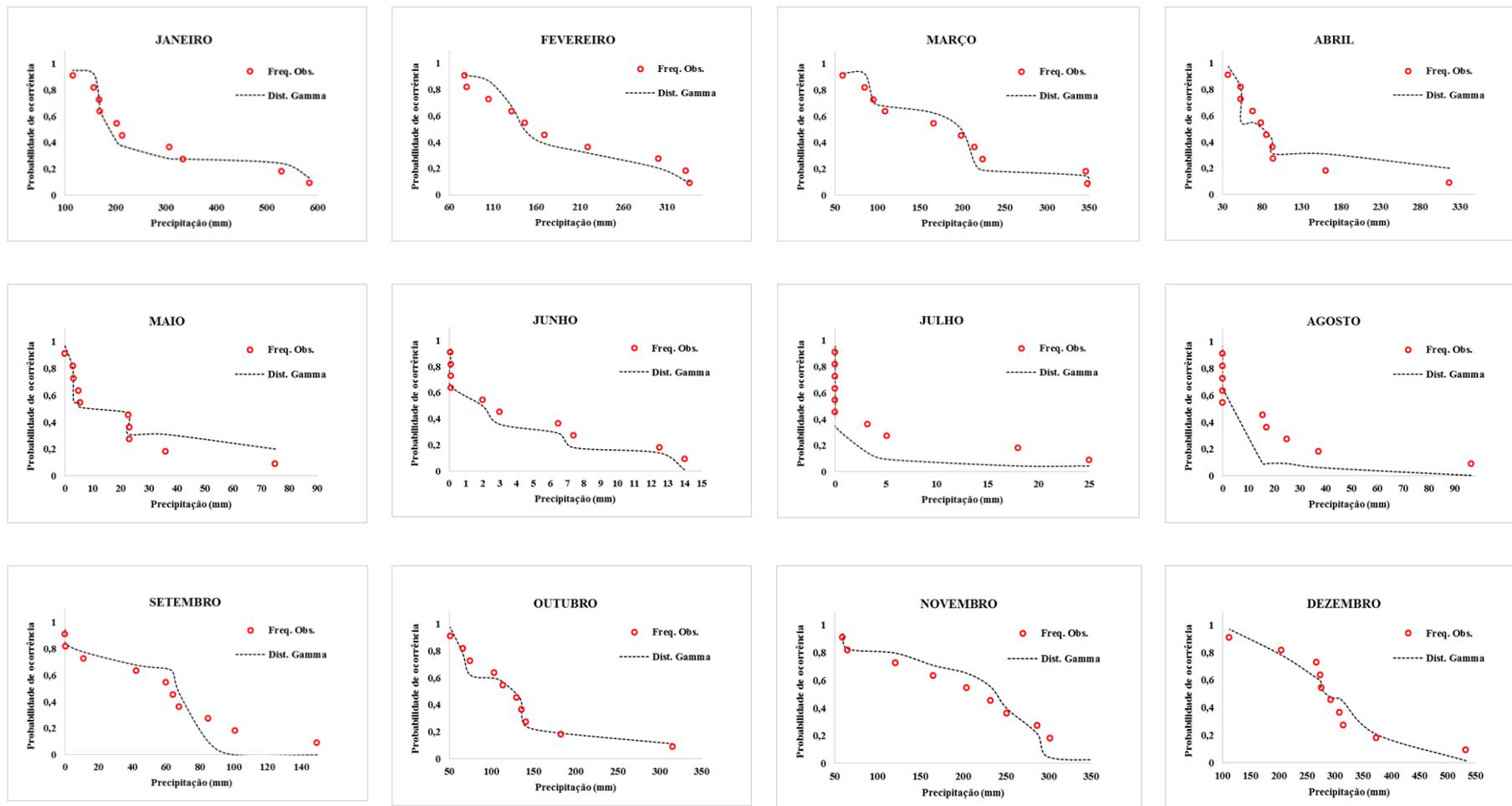


Figura 25 (b) - FDP's ajustadas
Estação Pirenópolis Cód. ANA 01548004 (Região Homogênea I)

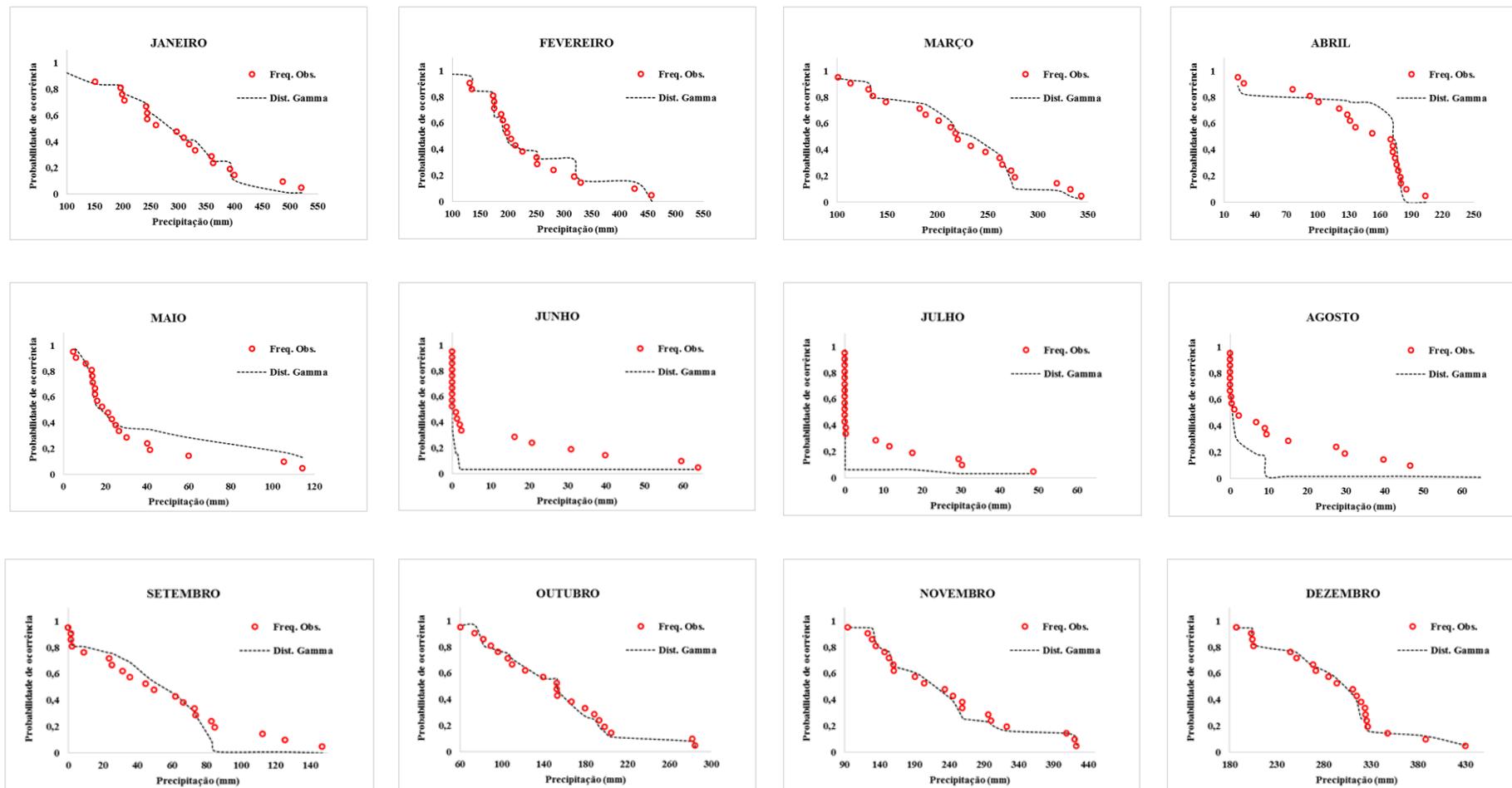


Figura 25 (c) - FDP's ajustadas
 Estação Fazenda Babilônia Cód. ANA 01753002 (Região Homogênea I)

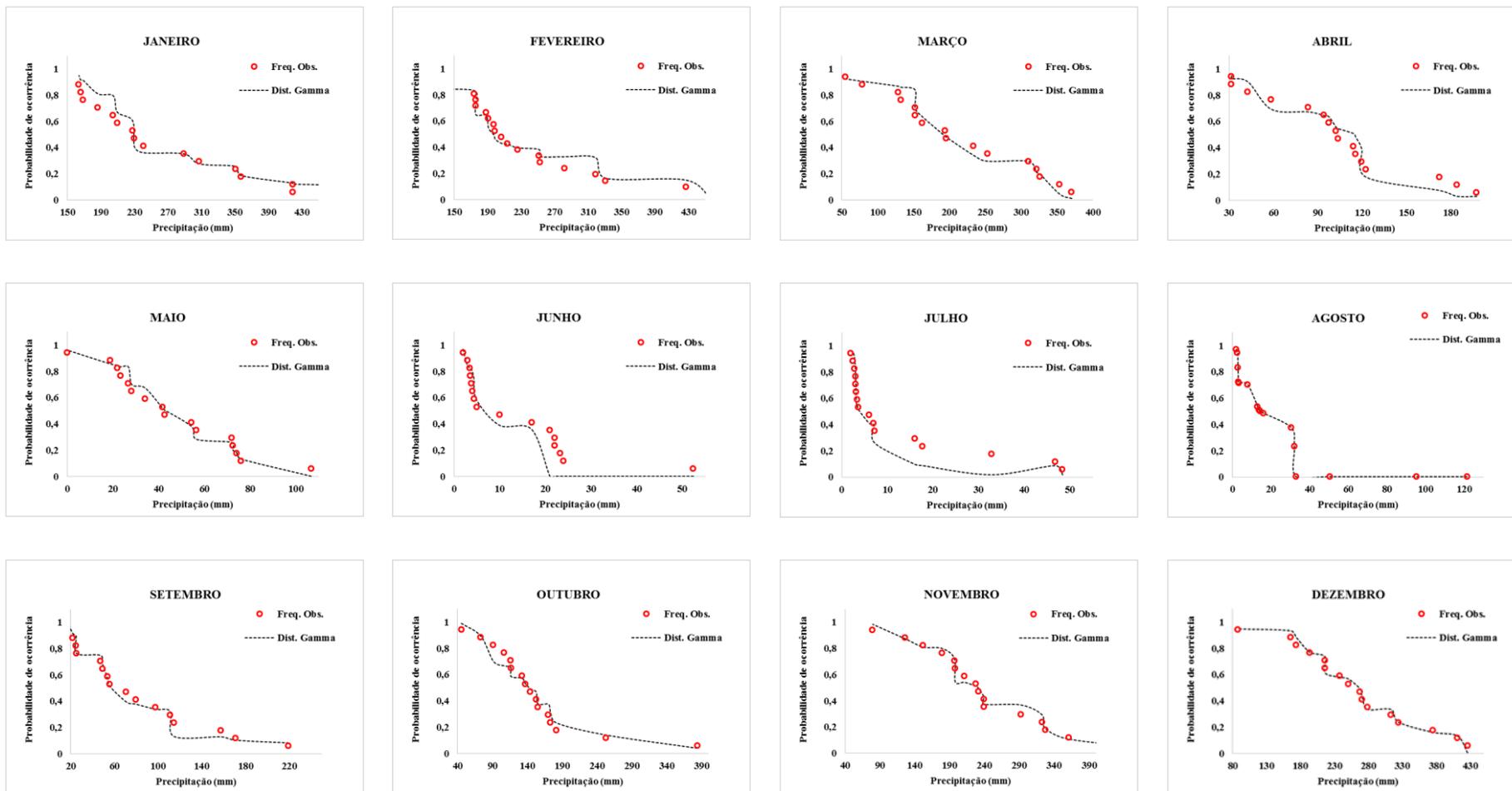


Figura 25 (d) - FDP's ajustadas
 Estação Trecho Médio Cód. ANA 01351000 (Região Homogênea II)

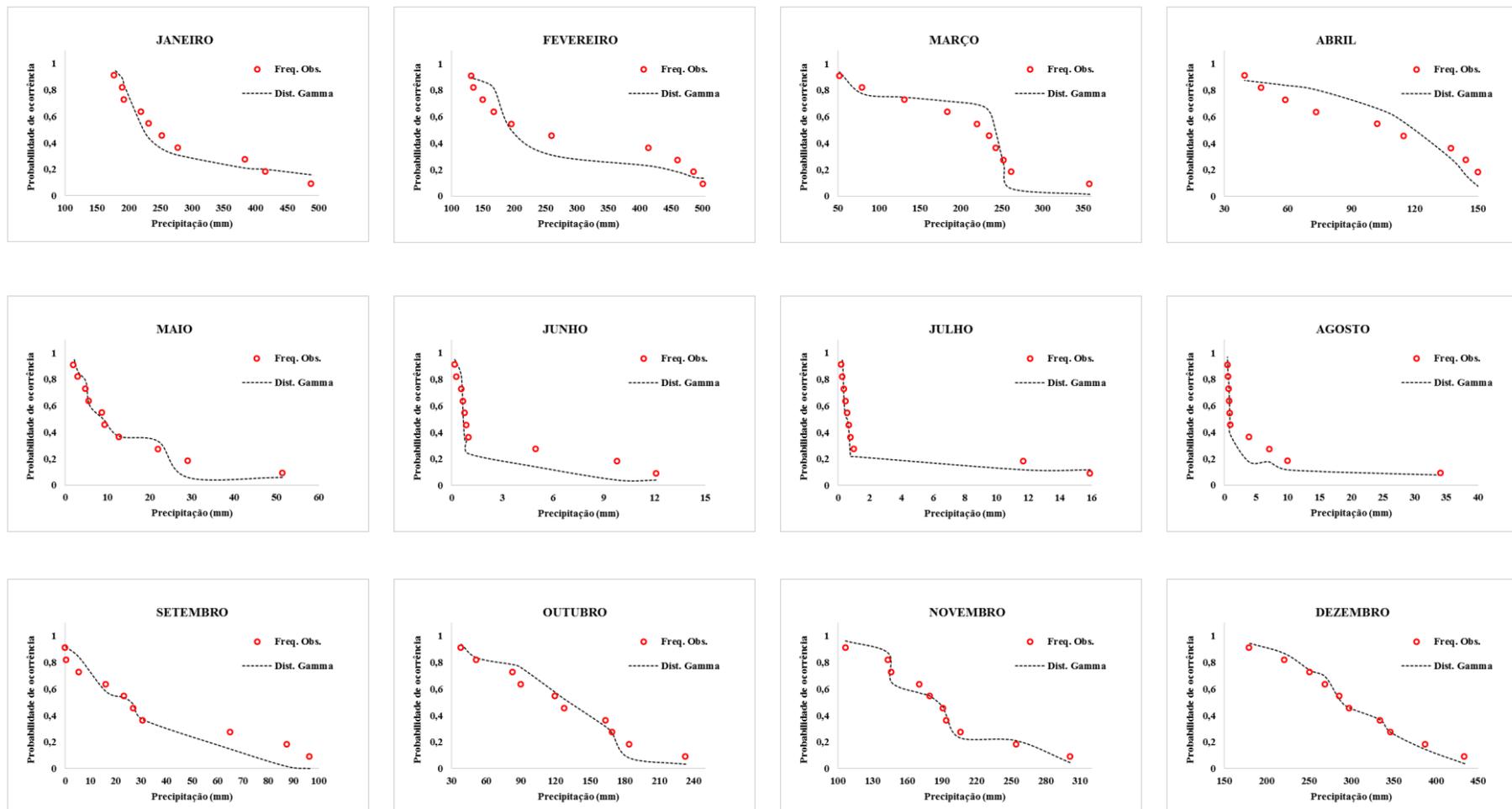


Figura 25 (e) - FDP's ajustadas
Estação Gurupi Cód. ANA 01149002 Região Homogênea II

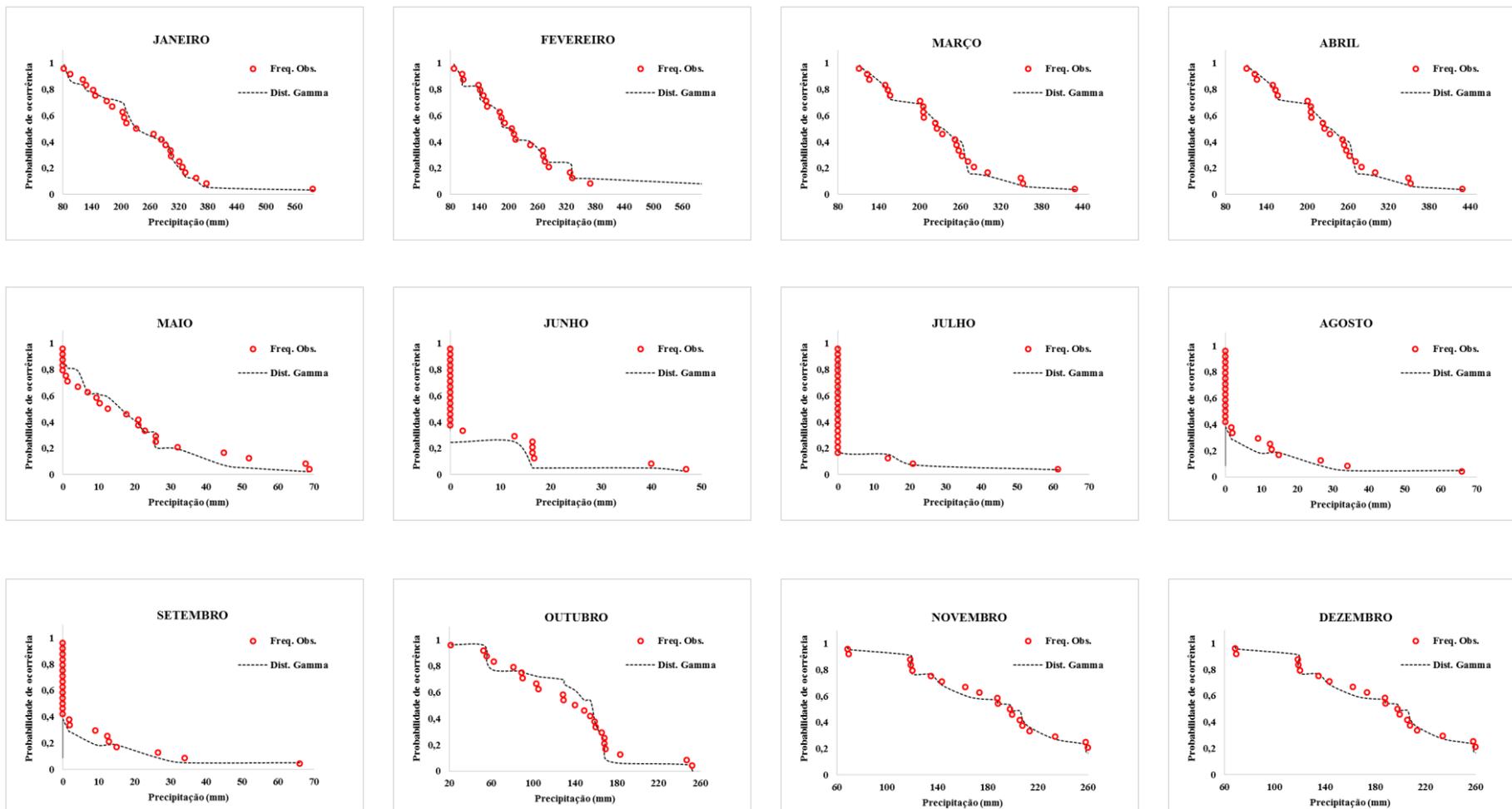


Figura 25 (f) - FDP's ajustadas
Estação Formosa do Araguaia Cód. ANA 01149001 (Região Homogênea II)

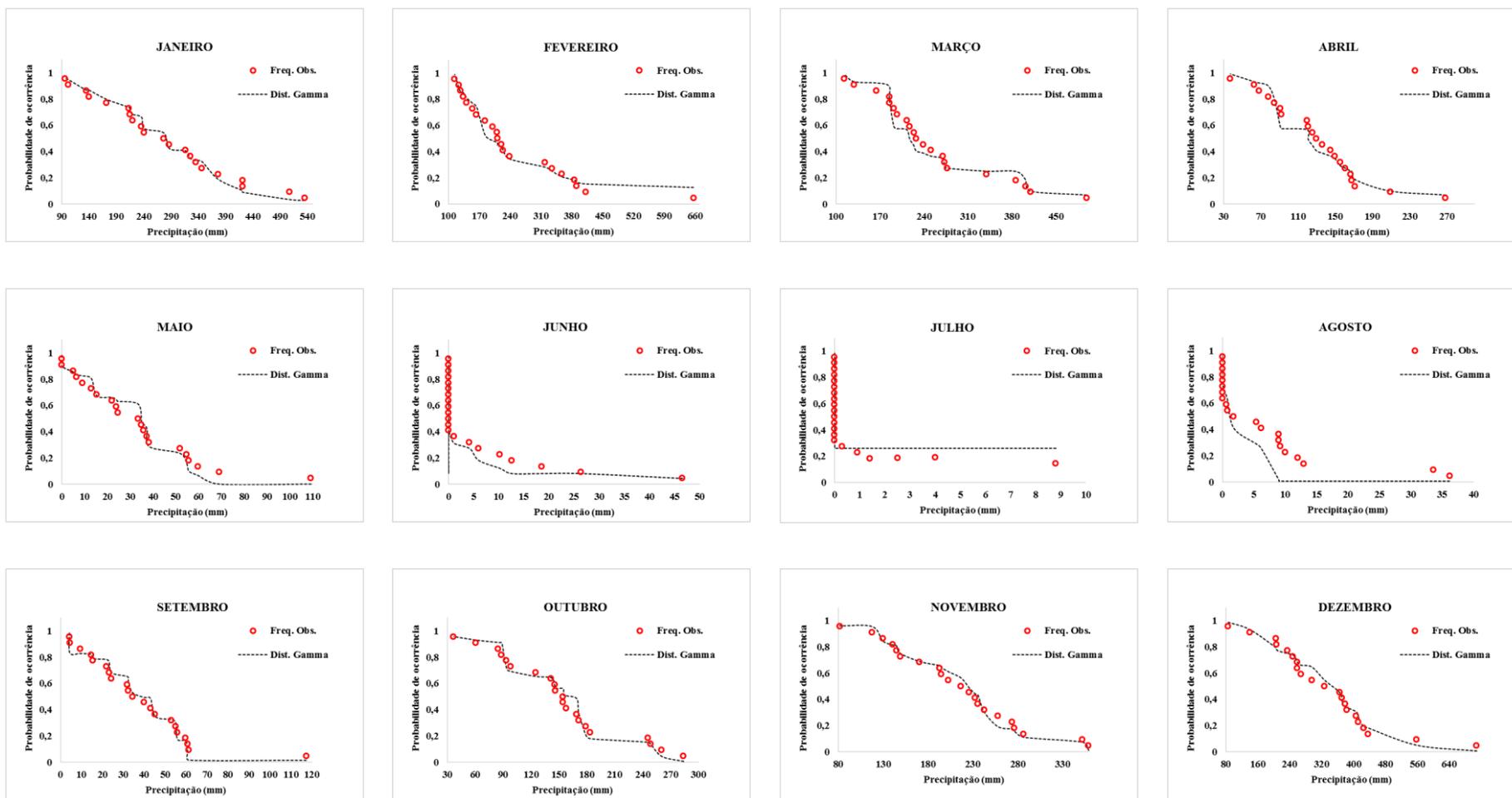


Figura 25 (g) - FDP's ajustadas
 Estação Tucuruí Cód. ANA 00349000 (Região Homogênea III)

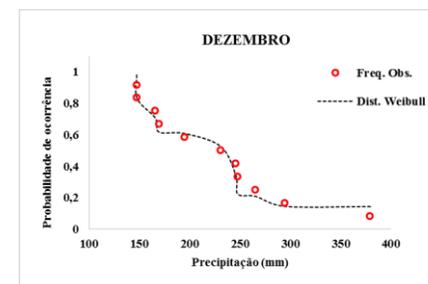
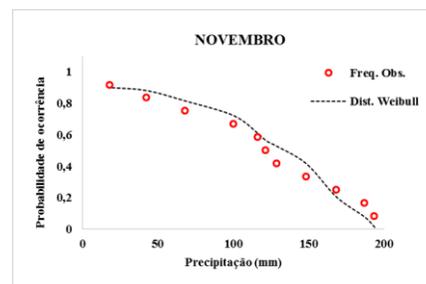
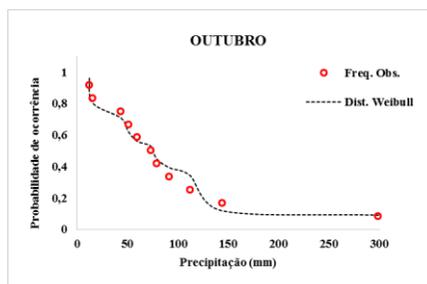
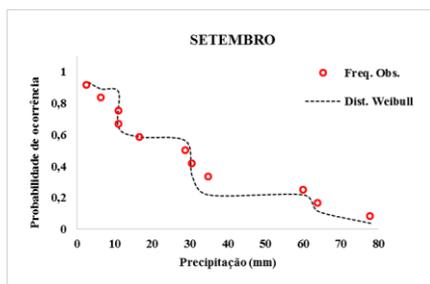
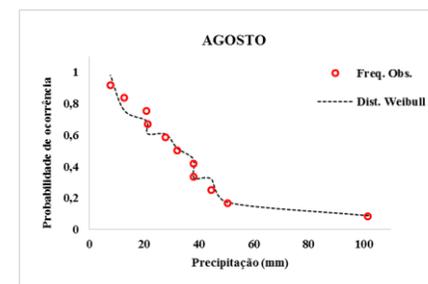
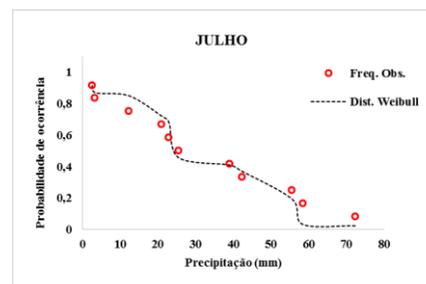
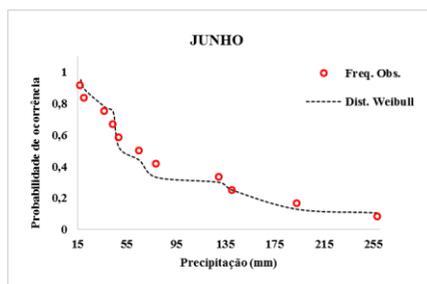
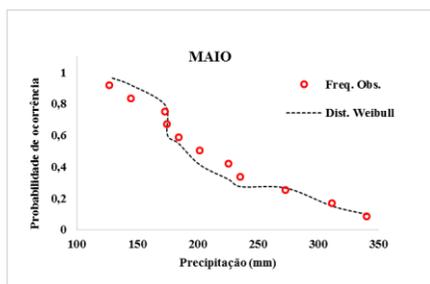
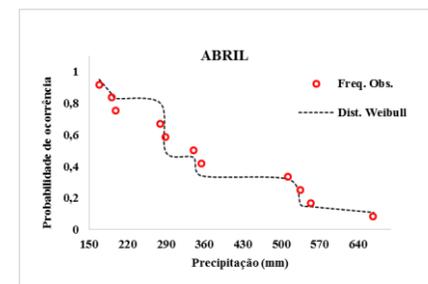
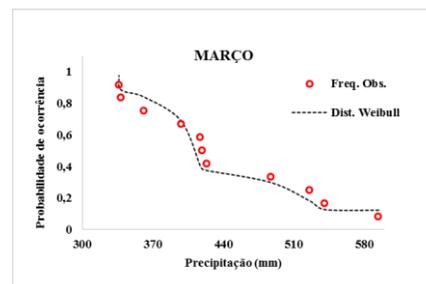
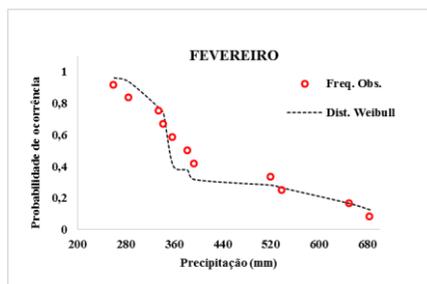
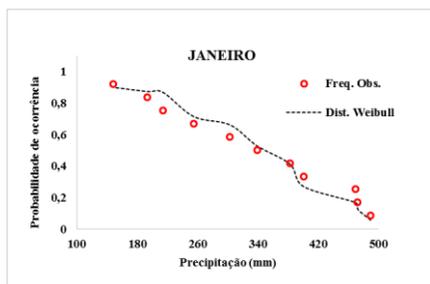


Figura 25 (h) - FDP's ajustadas
 Estação Cameté Cód. ANA 00249001 (Região Homogênea III)

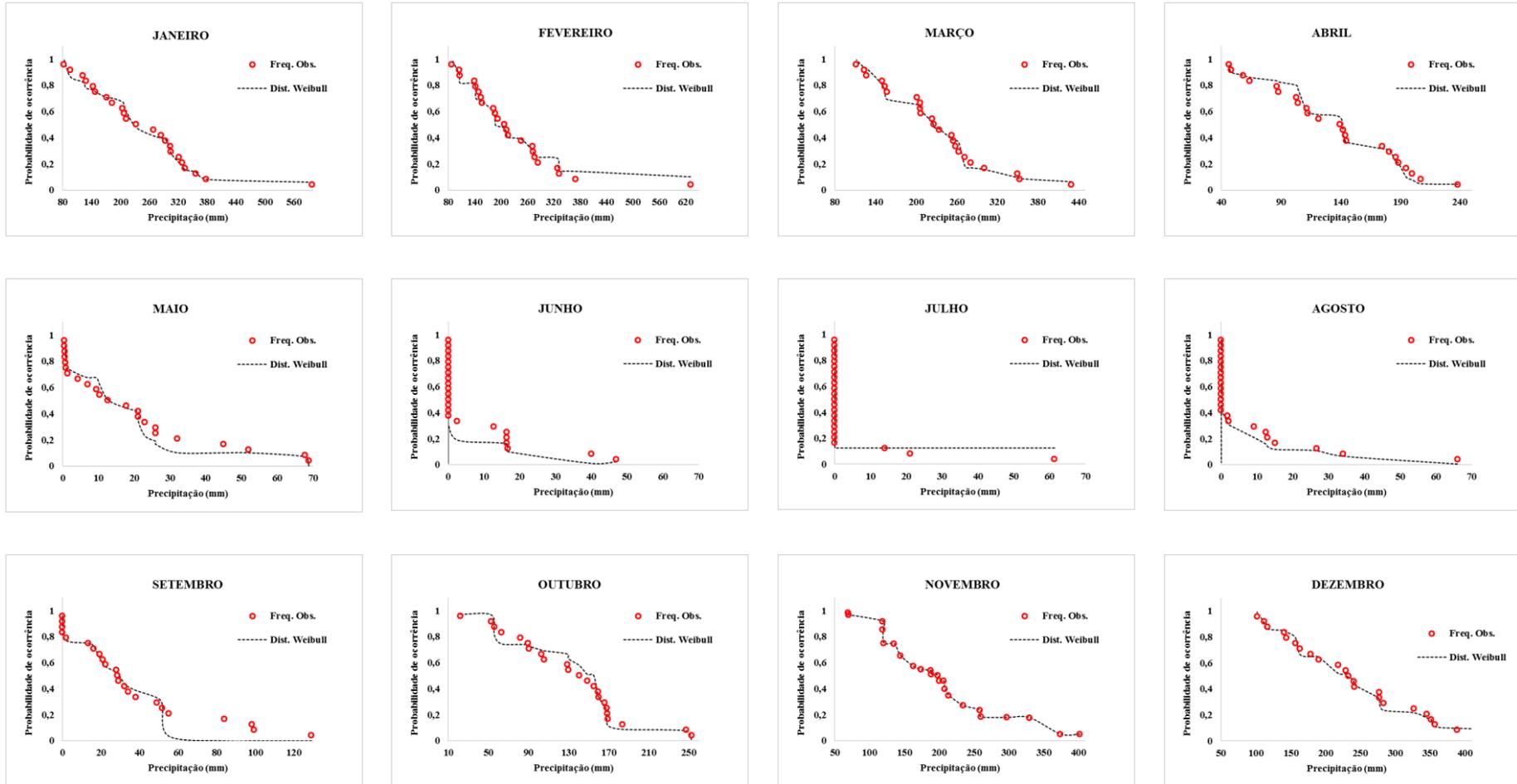
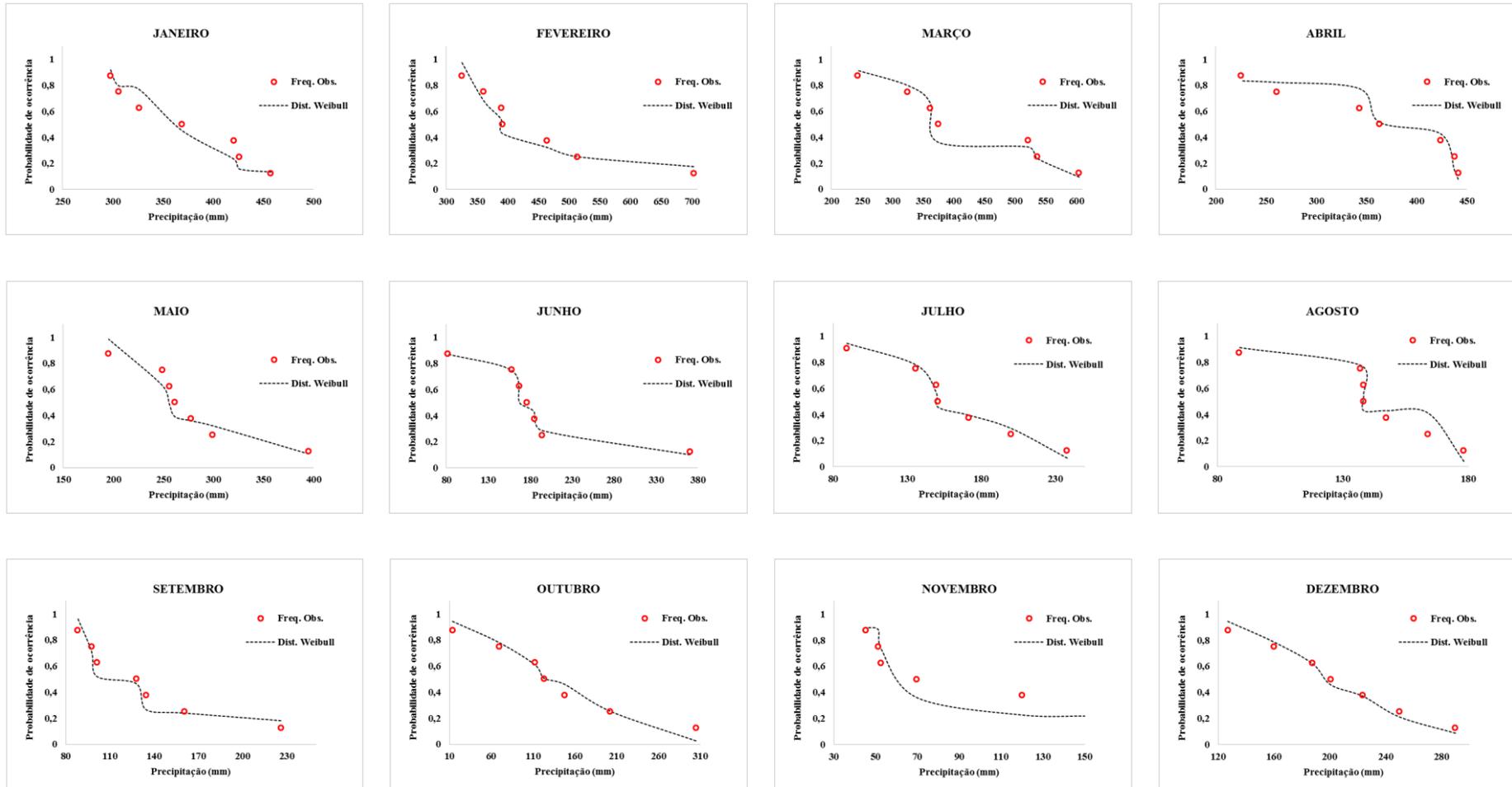


Figura 25 (i) - FDP's ajustadas
 Estação Belém Cód. ANA 00148001 (Região Homôgenea III)



5.3 REGIONALIZAÇÃO POR REGRESSÃO MÚLTIPLA

Na aplicação da regressão múltipla foram usadas as variáveis independentes cujos valores são conhecidos (Altitude, latitude e longitude) para prever os valores da variável dependente, que neste estudo é a precipitação média anual e a precipitação média mensal.

Inicialmente foi realizado a análise de multicolinearidade e em seguida foi aplicada a calibração e validação dos modelos de regressão, conforme são descritos nos tópicos a seguir.

5.3.1 Análise de Multicolinearidade

Com a aplicação da matriz de correlação, utilizando os dados das variáveis independentes da regressão, não houve correlação superior a 0,85 entre as variáveis explicativas em nenhuma das regiões homogêneas. Assim, não há necessidade de aplicar o teste do F-parcial para determinar a significância de cada variável explicativa na regressão. As matrizes de correlação entre as variáveis explicativas construídas para cada uma das regiões homogêneas, podem ser visualizadas na Tabela 16.

Tabela 16 - Matriz de correlação entre as variáveis explicativas das regiões homogêneas definidas por meio do método *fuzzy c-means*.

Região Homogênea 1				Região Homogênea 2			
	<i>Alt.(mm)</i>	<i>lat.</i>	<i>long.</i>		<i>Alt.(mm)</i>	<i>lat.</i>	<i>long.</i>
<i>Alt.(mm)</i>	1			<i>Alt.(mm)</i>	1		
<i>lat.</i>	-0,68992	1		<i>lat.</i>	-0,13564	1	
<i>long.</i>	-0,25493	0,617298	1	<i>long.</i>	0,120639	0,659567	1

Região Homogênea 3			
	<i>Alt.(m)</i>	<i>lat.</i>	<i>Long.</i>
<i>Alt.(m)</i>	1		
<i>lat.</i>	-0,92977	1	
<i>long</i>	0,280483	-0,21164	1

5.3.2 Calibração dos Modelos de Regressão Múltipla para Chuvas Totais Anuais

Os modelos de regressão foram aplicados as variáveis dependentes e independentes e o resultado dos coeficientes de desempenho sugerem os modelos que apresentaram os melhores resultados, conforme Tabela 17.

Tabela 17 – Coeficientes de Desempenho dos Modelos de Regressão.

Região Homogênea I						
Modelos	R ²	R ² _a	E(%)	ε (%)	Nash	RMSE
Linear	0,46	0,32	6,09	0,138	0,46	0,00017
Potencial	0,42	0,28	5,61	0,138	0,42	0,00017
Exponencial	0,45	0,30	5,61	0,139	0,45	0,00017
Logaritmo	0,42	0,28	6,08	0,137	0,42	0,00017
Região Homogênea II						
Modelos	R ²	R ² _a	E(%)	ε (%)	Nash	RMSE
Linear	0,40	0,26	3,89	0,58	0,41	0,00075
Potencial	0,39	0,29	3,81	0,58	0,40	0,00076
Exponencial	0,40	0,28	3,90	0,58	0,40	0,00075
Logaritmo	0,39	0,22	3,80	0,58	0,39	0,00076
Região Homogênea III						
Modelos	R ²	R ² _a	E(%)	ε (%)	Nash	RMSE
Linear	0,74	0,61	4,73	0,64	0,74	0,00071
Potencial	0,67	0,51	5,75	0,63	0,68	0,00064
Exponencial	0,72	0,58	4,51	0,65	0,75	0,00072
Logaritmo	0,70	0,55	5,64	0,63	0,70	0,00046

R² - coeficiente de determinação; R²_a - coeficiente de determinação ajustado; E(%) – erro relativo percentual; ε (%) – erro quadrático médio percentual; NASH – coeficiente de Nash Sutcliffe; RMSE – raiz quadrada da média dos erros quadrados.

Nas Regiões homogêneas I e II os coeficientes de determinação, R² e R²_a, foram menores que 0,5, em todos os modelos adotados, indicando que os mesmos não seriam tão representativos. Porém analisando o erro relativo percentual, o erro quadrático relativo médio percentual e o RMSE, o modelo teria um bom desempenho na estimativa de chuvas, pois os erros obtidos são menores que 10% e o RMSE apresentou valores mínimos. Em relação ao coeficiente de NASH, valores acima de 0,36, conferem ao modelo a condição de aceitável para a estimativa da variável dependente. Assim entre os quatro modelos, o Linear é o melhor, considerando os maiores coeficientes R² e NASH, para estimar chuvas médias anuais nas Regiões Homogêneas I e II.

Na Região Homogênea III, para todos os modelos analisados, os coeficientes de desempenho, R², R²_a e NASH apresentaram valores satisfatórios, porque quanto mais próximo de 1, os seus valores, melhor é a sua representação na estimação da variável dependente. Os erros foram bem inferiores a 10%, assim como o RMSE foi mínimo. No entanto, o modelo que apresentou o maior coeficiente de determinação - R² e maior coeficiente de NASH foi o Linear. Assim, o modelo Linear é o mais adequado na estimativa de chuva média anual da Região Homogênea III.

5.3.3 Validação dos Modelos de Regressão para Precipitações Totais Anuais

Foram selecionados nove postos pluviométricos, os quais foram chamados de estações Alvo (Tabela 13), para validar os modelos de regressão. Utilizaram-se as informações das estações (altitude, latitude e longitude), para estimar os totais médios de chuva anual. A validação foi realizada com o modelo Linear, que apresentou melhores resultados na etapa da calibração, adotando os seus respectivos parâmetros de regressão e as informações das estações alvo.

Foi calculado o erro relativo percentual (Equação 4.67), o qual compara o valor da precipitação observada com a simulada, e assim verifica-se, o modelo em questão é satisfatório na estimativa de chuvas médias anuais nas regiões homogêneas estudadas.

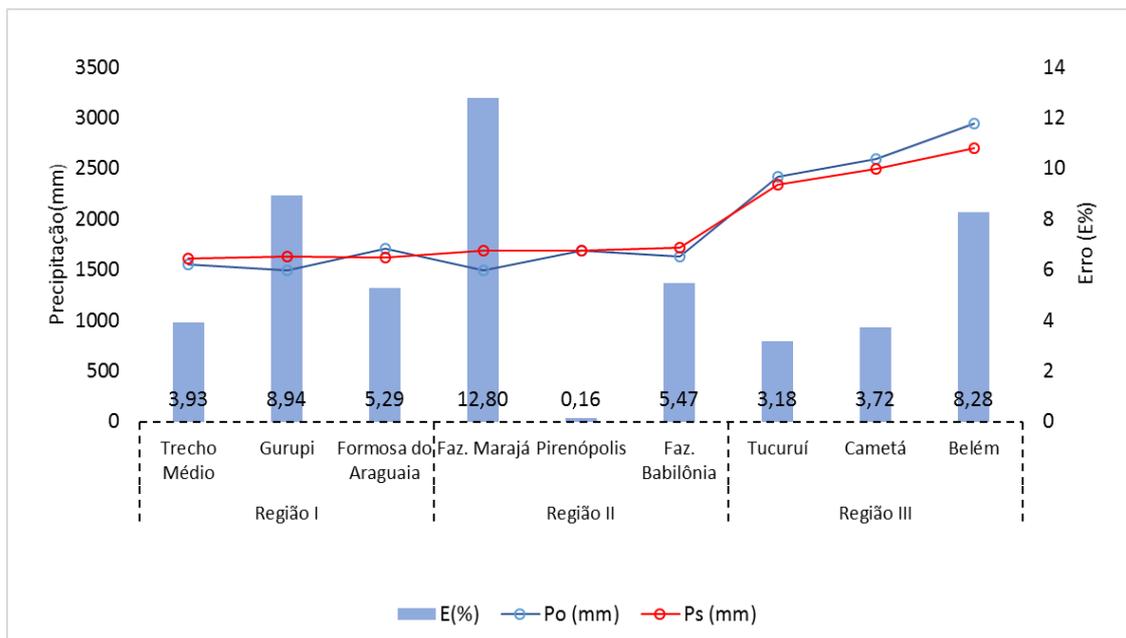
Pelos resultados apresentados na Tabela 18, o modelo Linear apresentou erros relativos percentuais menores que 10% em quase todas as estações alvos. Somente a estação Fazenda Marajá, pertencente a região homogênea II, o erro foi maior que 10%. No entanto a estação Pirenópolis, localizada na região II, o erro foi mínimo de 0,16%. A média dos erros foi de 6,10%, garantindo um bom desempenho do modelo Linear para a previsão de média de chuvas anuais nas regiões homogêneas formadas.

Tabela 18 - Validação do Modelo de Regressão Linear nas estimativas de chuva média anual.

Região	Estação Alvo	Po (mm)	Ps (mm)	E(%)
Região I	Trecho Médio	1554,74	1615,819	3,93
	Gurupi	1496,69	1630,496	8,94
	Formosa do Araguaia	1707,96	1617,634	5,29
Região II	Faz. Marajá	1497,85	1689,618	12,80
	Pirenópolis	1687,47	1684,773	0,16
	Faz. Babilônia	1632,19	1721,407	5,47
Região III	Tucuruí	2421,9	2344,829	3,18
	Cametá	2590,132	2493,893	3,72
	Belém	2942,843	2699,125	8,28
			Média E%	6,10

Utilizando o modelo Linear, considerado o mais adequado, foram determinadas as chuvas médias anuais simulados (Ps) e comparadas com a precipitação observada (Po), considerando os erros percentuais (E%) obtidos entre os valores de precipitação (Figura 26).

Figura 26 - Erros percentuais por estação alvo das chuvas médias mensais utilizando modelo Linear.



Os resultados obtidos dos erros médios demonstraram que a metodologia pode ter uma boa aplicação para estimativas de precipitações médias anuais na RHTA, principalmente em locais com carência de dados de chuvas nas referidas regiões homogêneas.

5.3.4 Calibração dos Modelos de Regressão para Precipitações Médias Mensais

O desempenho dos modelos de regressão obtidos na calibração das médias mensais de chuva, serão realizados por todos os coeficientes de desempenho, já citados. A regressão múltipla para as médias mensais será realizada, adotando-se as médias mensais de cada mês e de cada estação nas diferentes regiões homogêneas. As variáveis independentes continuam sendo a altitude a latitude e a longitude. A variável dependente, agora corresponde a precipitação média mensal.

Pela análise dos resultados do coeficiente de determinação e do erro relativo percentual obtidos (Tabela 19), nos meses de dezembro, janeiro e março, o modelo Linear apresentou erros percentuais abaixo dos 10% e os coeficientes de determinação não ficaram tão distantes de 1. Porém nos demais meses, e principalmente nos meses mais secos e com ausência de precipitação, os erros foram superiores a 10%, chegando a 100% de erro no mês de Julho, o qual é o menos chuvoso. Portanto, esses modelos não se aplicariam para as estimativas médias mensais de precipitação de forma adequada

nessa região. No Apêndice F encontram-se todos os valores dos coeficientes de desempenho utilizados.

Tabela 19 - Coeficiente de Determinação e o Erro relativo percentual obtido para cada mês na Região Homogênea I.

Região Homogênea I								
Mês	Linear		Potencial		Exponencial		Logaritmo	
	R ²	E(%)	R ²	E(%)	R ²	E(%)	R ²	E(%)
Jan	0,4	7,7	0,4	7,7	0,5	7,7	0,4	7,7
fer	0,5	17,1	0,5	6,8	0,5	7,1	0,1	7,2
mar	0,7	8,3	0,7	8,2	0,7	8,2	0,7	8,3
abr	0,9	12,7	0,9	12,7	0,9	11,3	0,9	11,3
mai	0,8	26,2	0,9	14,2	0,9	14,4	0,9	20,8
jun	0,6	96,2	0,7	41,3	0,6	49,0	0,4	81,4
jul	0,6	115,5	0,6	50,5	0,5	60,5	0,5	96,4
ago	0,6	45,2	0,6	40,8	0,6	34,3	0,7	40,3
set	0,3	21,4	0,5	17,8	0,3	19,9	0,4	20,0
out	0,1	17,0	0,4	14,7	0,2	17,4	0,3	14,3
nov	0,5	12,1	0,7	9,9	0,5	13,3	0,7	8,8
dez	0,5	7,2	0,8	6,5	0,7	7,5	0,7	6,9
Média	0,6	31,4	0,6	18,4	0,6	20,9	0,6	26,9

Na Região Homogênea II, as médias dos erros relativos percentuais foram todas elevadas, sendo que nos meses menos chuvosos (maio a setembro) o erro em todos os modelos ficaram acima de 10 %. Os resultados dos coeficientes de determinação foram baixos, próximos de 0, sendo que os erros foram elevados, como por exemplo no mês de fevereiro o R² foi de 0,06 e o erro foi de 100%, no modelo exponencial. Os modelos de regressão adotados não são representativos na estimativa de chuvas médias mensais para essa região porque em todos os modelos adotados, nem todos os meses apresentam resultados simulados adequados. O ideal seria o modelo apresentar mínimos erros em todos os meses analisados.

Na Tabela 20 encontra-se os valores dos coeficientes de determinação e do erro relativo percentual obtidos na calibração dos modelos de regressão para as médias mensais da Região Homogênea II.

Tabela 20 - Coeficiente de Determinação e o Erro relativo percentual obtido para cada mês na Região Homogênea II.

Região Homogênea II								
Mês	Linear		Potencial		Exponencial		Logaritmo	
	R ²	E(%)						
Jan	0,05	9,71	0,05	9,62	0,06	9,46	0,04	9,92
fer	0,07	8,95	0,06	90,03	0,06	100	0,07	97,52
mar	0,07	7,21	0,07	7,32	0,07	7,29	0,07	7,35
abr	0,62	8,47	0,58	8,49	0,58	58,37	0,63	18,68
mai	0,76	80,9	0,73	12,57	0,72	12,02	0,76	14,38
jun	0,78	36,45	0,83	22,82	0,84	22,11	0,76	37,2
jul	0,81	60,11	0,8	54,64	0,82	23,24	0,79	31,79
ago	0,77	57,25	0,78	15,97	0,79	15,67	0,76	17,91
set	0,83	10,49	0,82	11,19	0,82	41,32	0,83	75,8
out	0,17	11,64	0,18	12,73	0,17	7,68	0,18	67,69
nov	0,09	11,05	0,07	82,2	0,09	96,99	0,07	27,24
dez	0,06	10,35	0,26	7,43	0,27	7,43	0,25	7,46
Média	0,4	26,0	0,4	27,9	0,4	33,5	0,4	34,4

Na calibração da Região Homogênea III, apesar do coeficiente de determinação – R² ter uma média próxima a 0,7 na maioria dos meses (Tabela 21), as médias dos erros relativos foram acima de 20%, com erros elevados, acima de 90%, no mês de julho, nos modelos linear, potencial e logaritmo, conferindo aos modelos de regressão a condição de não adequados nas estimativas de chuvas para essa região.

Os valores do R²_a e do Nash ficaram próximos do valor do R² e os valores do ϵ (%) e do RMSE foram mínimos, em todos os modelos. Os valores dos índices de desempenho dos modelos de regressão da Região Homogênea I, II e III, das médias mensais, estão disponíveis no Apêndice F.

Pelos erros percentuais elevados e pelos valores do R², obtidos na calibração, os modelos de regressão para as médias mensais não são bons estimadores de chuvas nas referidas regiões homogêneas, devido principalmente, não apresentarem bons resultados para todos os meses. Nos meses de maio, junho e julho, por exemplo, ocorreram erros percentuais elevados, nas três regiões. Em virtude desses resultados, será realizado uma outra abordagem na estimativa de chuvas médias mensais, indicado pelos tópicos seguintes.

Tabela 21 - Coeficiente de Determinação e o Erro relativo percentual obtido para cada mês na Região Homogênea III.

Região Homogênea III									
Mês	Linear		Potencial		Exponencial		Logaritmo		
	R ²	E(%)							
Jan	0,63	6,91	0,62	16,68	0,59	6,87	0,62	16,75	
fer	0,4	92	0,41	87,89	0,39	8,86	0,41	27,95	
mar	0,78	60,62	0,63	10,74	0,77	6,7	0,67	10,43	
abr	0,9	72,01	0,82	12,61	0,92	7,41	0,8	11,4	
mai	0,97	13,27	0,93	20,33	0,98	11,46	0,92	24,73	
jun	0,94	49,98	0,92	37,89	0,97	109,9	0,93	87,89	
jul	0,91	105	0,92	95,72	0,48	22,37	0,53	124	
ago	0,86	40,21	0,88	28,99	0,92	21,34	0,87	32,1	
set	0,25	35,8	0,2	32,62	0,2	32,25	0,25	36,26	
out	0,58	20,99	0,74	23,27	0,8	19,23	0,78	27,6	
nov	0,99	7,09	0,17	19,11	0,97	8,14	0,96	17,78	
dez	0,98	4,73	0,25	7,91	0,98	4,91	0,92	12,1	
Média	0,77	42,38	0,62	32,81	0,79	21,62	0,72	35,75	

5.3.5 Estimativas de Chuvas para Período Chuvoso e Seco

Como foi observado na calibração das médias mensais, que os modelos de regressão múltipla não tiveram um bom desempenho, nas estimativas de chuvas médias mensais e que os maiores erros relativos percentuais ocorreram nos meses menos chuvosos e os menores erros nos períodos mais chuvosos, optou-se em realizar a regressão múltipla considerando períodos secos e chuvosos, na tentativa de obter modelos mais representativos e adequados na estimativa de chuvas médias mensais. Desta forma foram considerados meses chuvosos os meses de novembro, dezembro, janeiro, fevereiro, março e abril. Os meses secos compreendem, então, os meses de maio, junho, julho, agosto, setembro e outubro, conforme observado na série histórica das estações.

5.3.6 Calibração dos Modelos de Regressão para período Chuvoso e Seco

Essa nova análise foi realizada utilizando as médias mensais dos meses mais chuvosos e dos meses secos de cada estação, das referidas regiões homogêneas formadas no agrupamento *fuzzy c-means*. Assim, aplicou-se a regressão linear múltipla para os modelos Linear, Potencial, Exponencial e Logaritmo, usando as mesmas variáveis dependentes e a variável independente, sendo a precipitação média do período chuvoso e a precipitação média do período seco.

Na calibração dos modelos de regressão tanto para o período chuvoso como o seco, o critério de avaliação de desempenho dos modelos foram os mesmos usados anteriormente (R^2 , R^2_a , $E\%$, $\epsilon\%$, $NASH$ e $RMSE$).

Pretendendo analisar os modelos de regressão, a Tabela 22 foi organizada para melhor avaliação.

Tabela 22 – Coeficientes de Desempenho dos modelos para o período chuvoso.

Região Homogênea I						
Modelo	R^2	R^2_a	$E(\%)$	$\epsilon(\%)$	Nash	RMSE
Linear	0,38	0,23	6,09	0,11	0,38	0,00110
Potencial	0,32	0,18	6,32	0,10	0,32	0,00104
Exponencial	0,29	0,25	6,12	0,11	0,26	0,00109
Logaritmo	0,21	0,16	6,24	0,10	0,21	0,00105
Região Homogênea II						
Modelo	R^2	R^2_a	$E(\%)$	$\epsilon(\%)$	Nash	RMSE
Linear	0,37	0,34	4,77	0,44	0,37	0,00341
Potencial	0,17	0,027	4,80	0,44	0,17	0,00338
Exponencial	0,18	0,037	4,76	0,45	0,18	0,00340
Logaritmo	0,17	0,024	4,82	0,44	0,17	0,00339
Região Homogênea III						
Modelo	R^2	R^2_a	$E(\%)$	$\epsilon(\%)$	Nash	RMSE
Linear	0,73	0,44	4,60	0,60	0,996	0,00568
Potencial	0,66	0,49	4,61	0,62	0,997	0,00600
Exponencial	0,64	0,41	4,63	0,60	0,996	0,00562
Logaritmo	0,69	0,53	4,64	0,62	0,997	0,00427

R^2 - coeficiente de determinação; R^2_a - coeficiente de determinação ajustado; $E(\%)$ – erro relativo percentual; $\epsilon(\%)$ – erro quadrático médio percentual; $NASH$ – coeficiente de Nash Sutcliffe; $RMSE$ – raiz quadrada da média dos erros quadrados.

Com base nos resultados da Tabela 23, para o período chuvoso, o modelo Linear apresentou melhor desempenho, apesar de apresentar um R^2 baixo. Ainda assim, o coeficiente de Nash (acima de 0,36), o caracteriza como um modelo aceitável. Os erros percentuais foram bons, abaixo de 10%, conferindo um modelo ajustado para estimar chuvas médias nos períodos chuvosos para a Região Homogênea I, II e III.

Na Região III todos os modelos de regressão múltipla testados apresentaram um ótimo resultado do coeficiente de Nash que ficou bem próximo do valor 1, porém o modelo Linear apresentou o R^2 mais próximo de 1, sendo este modelo, o indicado para simulação de chuvas mensais no período chuvoso. Para a avaliação do período seco analisaremos a Tabela 23.

Tabela 23 – Coeficiente de Desempenho dos modelos para o período Seco.

Região Homogênea I						
Modelo	R ²	R ² _a	E(%)	ε (%)	Nash	RMSE
Linear	0,80	0,78	9,65	0,03	0,80	0,00391
Potencial	0,80	0,78	9,22	0,04	0,79	0,00387
Exponencial	0,80	0,79	9,57	0,04	0,81	0,00403
Logaritmo	0,80	0,78	9,58	0,03	0,80	0,00391
Região Homogênea II						
Modelo	R ²	R ² _a	E(%)	ε (%)	Nash	RMSE
Linear	0,60	0,53	11,09	0,17	0,60	0,00953
Potencial	0,62	0,61	10,94	0,17	0,61	0,00948
Exponencial	0,60	0,53	10,85	0,18	0,60	0,00965
Logaritmo	0,59	0,52	11,21	0,16	0,59	0,00935
Região Homogênea III						
Modelo	R ²	R ² _a	E(%)	ε (%)	Nash	RMSE
Linear	0,85	0,77	15,73	0,16	0,85	0,00527
Potencial	0,89	0,87	14,08	0,27	0,80	0,00457
Exponencial	0,87	0,82	13,78	0,23	0,87	0,00561
Logaritmo	0,85	0,77	15,11	0,21	0,85	0,00360

R² - coeficiente de determinação; R²_a - coeficiente de determinação ajustado; E(%) – erro relativo percentual; ε (%) – erro quadrático médio percentual; NASH – coeficiente de Nash Sutcliffe; RMSE – raiz quadrada da média dos erros quadrados.

Na região homogênea I, todos os modelos estudados apresentaram uma boa performance, devido ao resultados dos coeficientes de desempenho serem muito satisfatórios, destacando os valores do R², R²_a e o Nash. Porém o modelo Potencial foi o melhor, devido seu erro relativo percentual e o RMSE serem menores que nos demais modelos.

Nas regiões homogêneas II e III, apesar do erro relativo percentual ser maior que 10%, o resultado do coeficiente R², por ser o maior, no modelo Potencial, permite que tal modelo seja considerado mais adequado na estimativas de chuvas médias mensais para períodos secos nessas regiões, respectivamente.

A avaliação feita considerando períodos secos e chuvosos, apresentou melhores resultados na estimativa de chuvas médias mensais do que avaliando as médias em cada mês. Esse bom resultado pode ser explicado por esses modelos matemáticos simularem melhor, dados mais homogêneos com características bem similares, que agregam características estatísticas favoráveis ao desempenho do modelo, como a média e o desvio padrão mais ajustado.

5.3.7 Validação do Modelo de Regressão para período Chuvoso e Seco

Na validação do período chuvoso, adotou-se os seus respectivos parâmetros de regressão obtidos na calibração com o modelo Linear e foi calculado o erro relativo percentual da precipitação observada com a simulada. A Tabela 24 mostra o resultado da validação do modelo de regressão Linear para os totais de precipitação médio no período chuvoso, adotando as estações alvos.

Apesar da estação Tucuruí, apresentar erro maior que 10%, na estimativa de chuvas médias para o período chuvoso, o modelo Linear simulou muito bem a precipitação, pois a média de erros foi de 5,6, garantindo um bom desempenho do modelo para estimar chuvas médias mensais no período chuvoso.

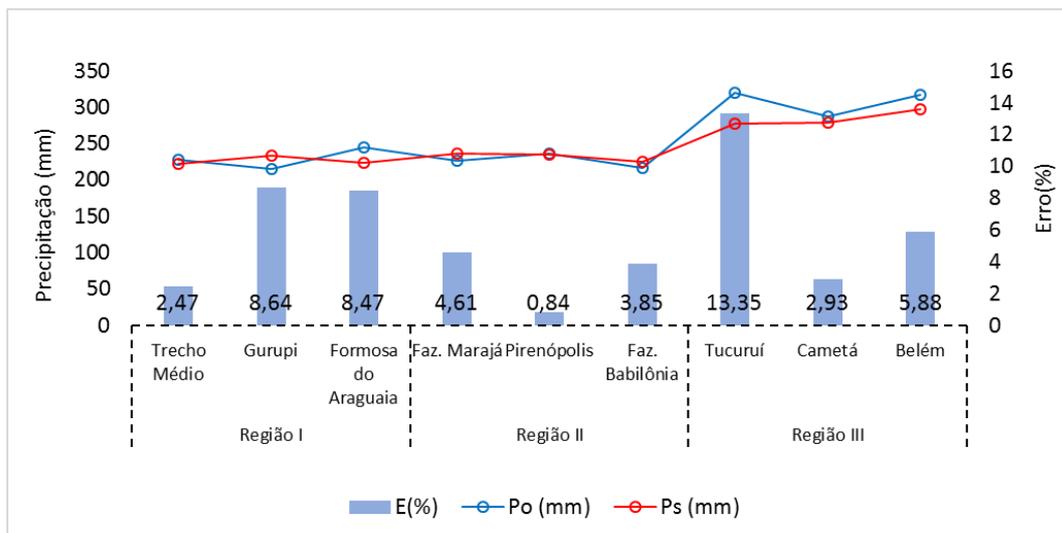
Foram determinadas as precipitações médias, do período chuvoso, simulados (Ps) pelo modelo Linear e comparadas com a precipitação média observada (Po), no período chuvoso, considerando os erros percentuais ($E\%$) obtidos entre os valores das precipitações, conforme pode ser observado na Figura 27.

Tabela 24 - Validação do Modelo de Regressão Linear para o Período Chuvoso.

Região	Estação Alvo	Po (mm)	Ps (mm)	E(%)
Região I	Trecho Médio	227,983	222,352	2,47
	Gurupi	214,78	233,341	8,64
	Formosa do Araguaia	244,33	223,634	8,47
Região II	Faz. Marajá	226,42	236,869	4,61
	Pirenópolis	236,62	234,618	0,84
	Faz. Babilônia	216,56	224,9	3,85
Região III	Tucuruí	320,271	277,515	13,35
	Cametá	287,99	279,547	2,93
	Belém	316,586	297,961	5,88
			Média $E\%$	5,67

De acordo com os resultados do erro percentual (5,67), o modelo de Regressão Linear é satisfatório na estimativa de chuvas médias no período chuvoso nas regiões homogêneas I, II e III.

Figura 27 - Erros percentuais por estação alvo das chuvas médias para período chuvoso, utilizando modelo Linear



Para o período seco, o modelo de regressão Potencial foi testado, apresentando erros médios abaixo de 10%. Apesar da estação Faz. Babilônia e Cametá, apresentarem erro 12% e 14%, respectivamente na estimativa de chuvas médias para o período seco, o modelo Potencial apresentou um bom desempenho na estimativa de chuvas médias mensais, com erro médio de 6,86% para as três regiões homogêneas (Tabela 25).

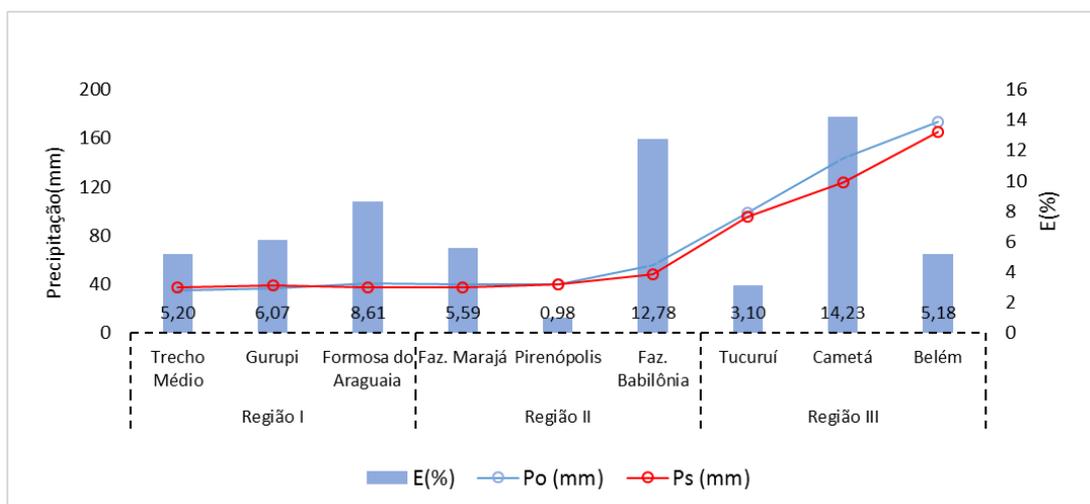
Foram determinadas as precipitações médias, do período seco simulados (Ps) pelo modelo Potencial e comparadas com a precipitação média observada (Po), no período seco, considerando os erros percentuais ($E\%$) obtidos entre os valores das precipitações, conforme pode ser observado na Figura 28. De acordo com os resultados do erro percentual, o modelo de Regressão Potencial é satisfatório na estimativa de chuvas médias no período seco nas regiões homogêneas I, II e III.

Pelos resultados obtidos na calibração e validação dos modelos de regressão, de estimativas de totais de chuvas médias anuais e de totais de médias mensais para períodos secos e chuvosos, principalmente, considerando o erro relativo percentual encontrado, pode-se considerar que a técnica de regionalização, empregada neste trabalho, é uma boa alternativa na estimativa de precipitações médias anuais e nas médias totais mensais de chuva em locais desprovidos de estações pluviométricas da RHTA.

Tabela 25 - Validação do Modelo de Regressão Potencial para o Período Seco.

Região	Estação Alvo	Po (mm)	Ps (mm)	E(%)
Região I	Trecho Médio	35,2	36,98	5,20
	Gurupi	36,7	38,88	6,07
	Formosa do Araguaia	40,4	36,94	8,61
Região II	Faz. Marajá	39,6	37,38	5,59
	Pirenópolis	40,0	39,63	0,98
	Faz. Babilônia	55,5	48,38	12,78
Região III	Tucuruí	98,95	95,50	3,10
	Cametá	143,7	123,26	14,23
	Belém	173,9	164,88	5,18
Média E%				6,86

Figura 28 - Erros percentuais por estação alvo das chuvas médias para período seco, utilizando modelo Potencial



Nas Tabelas 26, 27 e 28 são apresentadas, respectivamente, as equações dos Modelos de Regressão propostos neste trabalho para as regiões homogêneas formadas na RHTA, para estimativas de totais de chuvas anuais e mensais.

Tabela 26 – Modelos de Regressão para estimativa de totais de precipitação média anual.

Região Homogênea	Modelos de Regressão Linear	R ²	ε (%)
I	$P = 272,3 + 0,27 * H + 31 * la + 33,4 * lo$	0,5	0,13
II	$P = 1475 - 0,15 * H + 21,4 * la + 0,38 * lo$	0,4	0,58
III	$P = 100063 - 1,5 * H + 97 * la + 149 * lo$	0,7	0,64

Tabela 27 – Modelos de Regressão para estimativa de totais de precipitação média mensal para período chuvoso.

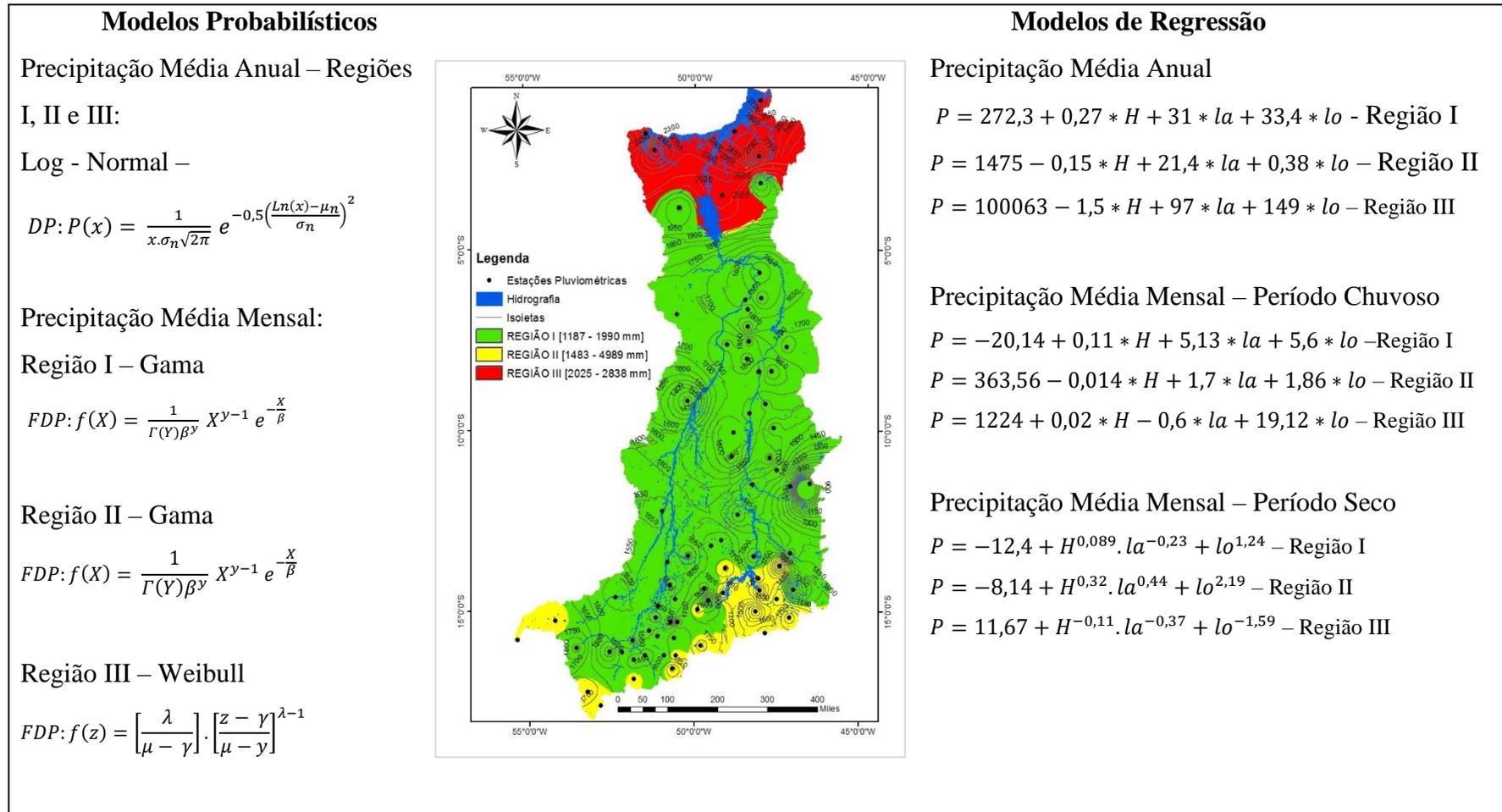
Região Homogênea	Modelos de Regressão Linear	R ²	ε (%)
I	$P = -20,14 + 0,11 * H + 5,13 * la + 5,6 * lo$	0,4	0,11
II	$P = 363,56 - 0,014 * H + 1,7 * la + 1,86 * lo$	0,4	0,44
III	$P = 1224 + 0,02 * H - 0,6 * la + 19,12 * lo$	0,7	0,6

Tabela 28 – Modelos de Regressão para estimativa de totais de precipitação média mensal para período seco.

Região Homogênea	Modelos de Regressão Potencial	R ²	ε (%)
I	$P = -12,4 + H^{0,089} . la^{-0,23} + lo^{1,24}$	0,8	0,04
II	$P = -8,14 + H^{0,32} . la^{0,44} + lo^{2,19}$	0,6	0,17
III	$P = 11,67 + H^{-0,11} . la^{-0,37} + lo^{-1,59}$	0,87	0,27

Na Figura 29 é apresentado um resumo dos modelos de estimativas de chuvas para a região estudada.

Figura 29 - Regiões Homogêneas de Precipitação e Modelos Probabilísticos e Matemáticos para estimativa de chuvas na RHTA.



6 CONCLUSÕES E SUGESTÕES

As técnicas usadas na formação das regiões homogêneas (fuzzy c-means, Índice PBM e Teste H), foram satisfatórias, porque conseguiram formar grupos distintos, com médias de precipitação bem definidas e com uma espacialização das regiões bem apropriadas e condizentes com o volume de chuvas registrado nas referidas regiões, onde as Regiões homogêneas I e II, formada ao sudoeste e centro oeste da RHTA, respectivamente possuem menores índices pluviométricos, característicos do bioma de cerrado e a Região homogênea III, localizada ao norte, possui maior índice pluviométrico, característico do bioma Amazônico.

Na estimativa da probabilidade de ocorrência de Precipitação Média Anual, a função de distribuição Log-Normal apresentou melhor aderência aos dados, nas três regiões homogêneas formadas (I, II e III) em comparação com as outras funções estudadas. Porém, constatou-se que para se estimar Precipitações Médias Mensais, nas Regiões homogêneas I e II, a função de distribuição Gama é a mais indicada, devido ter apresentado melhores ajustes aos testes de aderência e às frequências observadas. Já na Região Homogênea III a distribuição Weibull foi a mais adequada, visto que apresentou melhor ajuste aos testes de aderência e às frequências observadas.

Na estimativa da Precipitação Média Anual, por meio dos modelos de regressão múltipla, nas três Regiões Homogêneas de chuva da RHTA, o modelo Linear é o mais adequado, pois conseguiu simular valores bem próximos do observado.

Apesar dos modelos de estimativas de Precipitação Média Mensal não serem satisfatórios, foram selecionados modelos de regressão para simular Precipitação Média para os períodos Chuvosos e Secos das Regiões homogêneas na RHTA, baseados em médias mensais. Para o período Chuvoso, o modelo Linear foi o mais apropriado e para o Período Seco o modelo Potencial foi o melhor na simulação da Precipitação Média.

O emprego dos modelos estatístico e matemático, resultou em um desempenho satisfatório tanto para funções de distribuição de probabilidade quanto para regressão múltipla. No entanto, somente, para simulação de médias mensais, com os modelos de regressão, os resultados obtidos não foram adequados, sendo descartada a presente metodologia para simulação de chuvas média mensais, sem a consideração de períodos secos e chuvosos.

6.1 SUGESTÃO E RECOMENDAÇÃO

Como sugestões para trabalhos futuros, deixa-se o seguinte:

- Proposição de outro modelo para estimativa de alturas de chuva, para os casos em que a presente metodologia falhou, por exemplo, métodos híbridos, integrando métodos de regressão e de interpolação.

- Aplicar o método de Redes Neurais como ferramenta de agrupamento, com a finalidade de identificar regiões homogêneas de precipitação e comparar seus resultados ao método Fuzzy C-Means.

REFERÊNCIAS

- ANA, Agência Nacional de Águas (2013). *Meta Dados: Bacias Hidrográficas e Hidrografia*. Disponível: <http://metadados.ana.gov.br/geonetwork/srv/pt/main>.
- ANA, Agência Nacional de Águas (2009). *Plano Nacional de Recursos Hídricos da Região Hidrográfica Tocantins Araguaia*. Disponível: <http://www.ana.gov.br>.
- ANA, Agência Nacional de Águas (2006). *Caderno da Região Hidrográfica Tocantins Araguaia*. Ministério do Meio Ambiente, Secretaria de Recursos Hídricos. – Brasília ANA, MMA, 2006.
- ASSIS, F.N. de; ARRUDA, H.V. de; PEREIRA, A.R. *Aplicações de estatística a climatologia: teoria e prática*. Pelotas: Universitária, 1996. 161p.
- BEZDEK, J. *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum Press, New York 1981.
- BEZDEK, J. C. & PAL, S. K. *Fuzzy models for pattern recognition: methods that search for structures in data*. Methods that search for structures in data. IEEE Press, New York 1992.
- COSTA F.F DA, OVLIVEIRA G. D., CAIXETA P.R.C, SANTOS G. R., 2012. *Espacialização de Precipitação Máxima Média com Interpolador Geoestatístico:Krigagem*. IV Simpósio Brasileiro de Ciências Geodésicas e Tecnologias da Geoinformação. Recife - PE, 06- 09 de Maio de 2012 p. 001 – 008.
- COSTA, A. S.; CARIELLO, B. L.; BLANCO, C. J. C; PESSOA, F. C. L. *Regionalização de Curvas de Permanência de Vazão para a Região Hidrográfica do Xingu no Estado do Pará*. In: 26º Congresso Brasileiro de Engenharia Sanitária e Ambiental, CBESA, Porto Alegre, 2011.
- COHEN, J. C. P.; SILVA DIAS, M. F.; NOBRE, C. A. *Aspectos climatológicos das linhas de instabilidade na Amazônia*. Climanálise – Boletim de Monitoramento e Análise Climática, v.4, n. 11, p. 34-40, nov. 1989. (INPE-6028-PRE/2143).
- COX, E. (2005) *Fuzzy modeling and genetic algorithms for data mining and exploration* [S.1]: Elsevier/Morgan Kaufmann, 2005. Hardcover. (Morgan Kaufmann series in data management systems).
- CRUZ M.A.S, SOUZA L.A.S, ARAGÃO R.DE, SILVA, R.R.S, MOTA P.V.M. DA., 2013. *Aplicação de Regressão Linear Múltipla para estimativa da Precipitação Média Anual considerando a variabilidade espacial no Estado de Sergipe*. XX Simpósio Brasileiro de Recursos Hídricos – ABRH. Bento Gonçalves/RS.
- DIKBAS, Fatin et al. *Classification of precipitation series using fuzzy cluster method*. Journal of Climatology. 32: 1596-1603. 2012.

DUAN, J.; SELKER, J.; GRANT, G.E. *Evaluation of probability density functions in precipitation models for the Pacific Northwest*. Journal of the American Water Resources Association, v.34, p.617-627, 1998.

DUNN, J. C. A. *A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters*. Cybernetics and Systems, v. 3, p. 32-57, 1973.

EARTH FORUM, 2008. Disponível em <http://www.theplatearthsociety.org>. Acessado em Janeiro de 2016.

EMBRAPA. Empresa Brasileira de Pesquisa Agropecuária. *Chuvas nos cerrados: análise e especialização*. Brasília, DF: EMBRAPA-CPAC / EMBRAPA-SPI, 1994. 423p. ISBN:85-7075-004-8.

FERREIRA, C. de C. M. *Aplicação da Lógica Nebulosa (Fuzzy Cluster) na Definição de Unidades Climáticas: Estudo de caso na Bacia do Rio Paraibuna – MG/RJ*. In: Revista Brasileira de Climatologia. ISSN:1980-005x. Vol. 10, Janeiro/Junho 2012.

FRANCHINI, M.; SUPPO, M. (1996) *Regional analysis of flow duration curves for a limestone region*. Water Resource Management, vol. 10 p. 199-218.

FREITAS, A. J., SILVA, D. D., PRUSKI, F. F. PINTO, F. A., PEREIRA, S. B., GOMES FILHO, R. R., TEIXEIRA, A. F., BAENA, L. G., MELLO, L. T. A., NOVAES, L. F. *Equações chuvas intensas no Estado de Minas Gerais*. Belo Horizonte; Companhia de Saneamento de Minas Gerais; Viçosa: Universidade Federal de Viçosa, 2001. 65p.

FUKWYAMA Y., SUGENO M. 1989. *A New method of choosing the number of clusters for the fuzzy c-means method*. In: Proc. Fifth Fuzzy Systems Symp., 1989, pp. 247-250.

GONÇALVES, M. F., BLANCO, C. J. C., SANTOS V. C., SANTOS L. L. *Regiões homogêneas de Precipitação Pluvial no Estado do Pará, considerando a ocorrência de El Niño e La Niña*. In: Anais do XXI Simpósio Brasileiro de Recursos Hídricos. Brasília – DF, novembro de 2015.

GOMES, E. P.; BLANCO, C. J. C.; PESSOA, F. C. L. *Identificação de Regiões Homogêneas via Agrupamento Fuzzy C-means no Estado do Pará*. IV Simpósio de Métodos Numéricos Computacionais da Universidade Federal do Paraná – Curitiba, (IVSMNC), 2014.

GOYAL, M. K., GUPTA V., *Identification of Homogeneous Rainfall Regimes in Northeast Region of India using Fuzzy Cluster Analysis*. In: Water Resour Manage. DOI 10.1007/s11269-014-0699-7. Agosto de 2014.

GREENWOOD, J. A., LANDWEHR, J. M., MATALAS, N. C., e WALLIS, J. R. *Probability weighted moments: definition and relation to parameters expressible in inverse form*. Water Resources Research, v.15, n. 5, p.1049-1054, 1979.

HALL, M. J.; MINNS, W. A. *The classification of hydrologically homogeneous regions*. Hydrological Sciences Journal, 44(5), October, 1999.

- HAIR, J.F. et al. (2005). *Análise multivariada de dados*. 5.ed. Porto Alegre: Bookman, 593p.
- HASTINGS, N.A.J.; PEACOCK, J.B. *Statistical distributions: a handbook for students and practitioners*. New York: J. Wiley, 1975. 130p.
- HOSKING, J., WALLIS J., 1993. *Some Statistic useful in regional frequency analysis*. Water Resour. Res., 29, 271-28.
- HOSKING, J., WALLIS J. *Regional Frequency Analysis: Na Approach Based on L-moments*. Cambridge Univ. Pr. 1997.
- IBGE, (2014). *Instituto Brasileiro de Geografia e Estatística*. Acesso em nov. 2015. Disponível em: <http://www.ibge.gov.br/home>.
- INMET, Instituto Nacional de Meteorologia. *Boletim Agroclimatológico Mensal*. N. 7, Vol. 49, Brasília DF, junho de 2015.
- ISHIHARA J. H., FERNANDES L.L., DUARTE A.A.A.M., DUARTE A.R.C.L.M., PONTE M.X., LOREIRO G.E., 2014, *Quantitative and Spatial Assessment of Precipitation in the Brazilian Amazon (Legal Amazon) - (1978 to 2007)*. RBRH - Revista Brasileira de Recursos Hídricos Volume 19 n.1 —Jan/Mar 2014, pg.29-39.
- LANNA, A. E. L., Gestão de Recursos Hídricos. In: Carlos Eduardo Morelli Tucci (org.). *Hidrologia: Ciência e Aplicação*. Porto Alegre: EDUFRS, EDUSP, ABRH, 2004.
- LYRA, G.B; GARCIA, B.I.L; PIEDADE, S.M.S; SEDIYAMA, G.C; SENTELHAS, P.C. *Regiões homogêneas e funções de distribuição de probabilidade de precipitação pluvial no Estado de Táchira, Venezuela*. Pesq.agropec.brás, Brasília, v.41, n.2, p.205-215, fev.2006.
- LÓPEZ-MORENO, J. L, GARCIA-RUIZ, J. M., BENISTON, M. 2008. *Environmental Change and Water Management in the pyrenees. Facts and Future perspectives for Mediterranean mountains*. Glob. Planet. Change 66, 300-312.
- LOUREIRO, G. E., FERNANDES, L. L., ISHIHARA J. H., *Spatial and temporal variability of rainfall in the Tocantins-Araguaia hydrographic region*. In Acta Scientiarum, v. 37, n. 1., p. 89-98, Jan-Mar., 2015.
- MALHOTRA, N. *Pesquisa de marketing: uma orientação aplicada*. Trad. Laura Bocco. 4ª ed. Porto Alegre: Bookman, 2006.
- MELLO, C. R. & SILVA, A M. *Modelagem estatística da precipitação mensal e anual e no período seco para o estado de Minas Gerais*. In: Revista Brasileira de Engenharia Agrícola e Ambiental. V. 13, n.1, p.68-74, 2009.
- MENDONÇA, F. (2007). *Climatologia: Noções Básicas e Climas no Brasil*. São Paulo: Oficina de textos.

MEKANIK F., IMTEAZ M.A., GATO-TRINIDAD S., ELMAHDI A., 2013. Multiple regression and Artificial Neural Network for long-term rainfall forecasting using large scale climate modes. *Journal of Hydrology* 503 (2013) 11–21.

MINGOTI, S. A. *Análise de dados através de métodos de estatística multivariada*. Editora UFMG, Belo Horizonte, 2005.

MONDENESI M. V. *Análise de Agrupamentos FCM utilizando processamento Paralelo*. COOPE/UFRJ. Rio de Janeiro, 2008.

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

NAGHETTINI, M. PINTO, E. J. A. *Hidrologia Estatística*. Belo Horizonte, Ed. CPRM, 552 p. 2007.

OLIVEIRA, L. F. C., VIOLA M. R., PEREIRA S., MORAES N. R. *Modelo de Predição de Chuvas intensas para o Estado do Mato Grosso, Brasil*. *Revista Ambiente e Água*, An interdisciplinary journal of Applied Science, v. 6, n. 3, 2011.

PAKHIRA M. K., BANDYOPADHYAY S., MAULIK K. *Validity index for crisp and fuzzy clusters*, *Pattern Recognition* 37, p.481-501, 2004.

PAL N. R., BEZDEK J. C. *On cluster validity the fuzzy c-means model*. *IEEE Trans Fuzzy Syst* 3(3):370-379, 1995.

PARRACHO, A. C.; ROCHA, MELO GONÇALVES, P., ROCHA, A. *Regionalization of precipitation for the Iberian Peninsula and climate change*. In: *Physics and Chemistry of the Earth*. Julho 2015.

PESSOA, F. C. L. *Desenvolvimento de uma metodologia para Regionalização de Curvas de Permanência de Vazões da Amazônia*. PRODERNA/UFPA, Belém, 2015.

ROCHA, T., PERES, S. M., BÍSCARO, H. H., MACEDO, R. C. B., BOSCARIOLI, C. *Tutorial sobre Fuzzy c-means e Fuzzy Learning Vector Quantization: Abordagem híbridas para Tarefas de Agrupamento e Classificação*. *Revista de Informática Teórica e Aplicada*, volume 19, n. 1, p. 120 – 163, 2012.

RAO, A. R.; SRINIVAS, V. V. *Regionalization of watersheds by hybrid-cluster analysis*. *Journal of Hydrology*, v. 318, p. 37-56, 2006.

ROSS, J. L. S. In: (Org.) *Geografia do Brasil*. São Paulo: Edusp, 2005.

SANTOS, et al. *Precipitation regionalization of the Brazilian Amazon*. In: *Atmospheric Science Letters*. 16: 185-192. 2014.

SAMUEL, J.; COULIBALY, P.; METCALFE, R. A. *Estimation of continuous streamflow in Ontario ungauged basins: comparison of regionalization methods*. Journal of Hydrologic Engineering, v. 16, n. 5, p. 447-459, 2011.

SATYANARAYANA, P. & SIRVINAS V. V. *Regionalization of precipitation in data sparse áreas using large scale atmospheric variables – A fuzzy clustering approach*. In: Journal of Hidrology. 405: 462-473, 2011.

SILVA, C. J. et al. *Análise de distribuição de chuva para Santa Maria, RS*. Revista Brasileira de Engenharia Agrícola e Ambiental. Campina Grande, PB, v.11, n.1, p. 67-72, 2007.

SILVA, F.A.M. da; ASSAD, E.D. *Análise temporal do potencial hídrico climático do Estado de Goiás*. In: ASSAD, E.D.; SANO, E.E. (Org.). Sistema de informações geográficas: aplicações na agricultura. 2.ed. Brasília: Embrapa SPI, 1998. p.273-309.

SILVEIRA, A. L. L., *Ciclo Hidrológico e Bacia Hidrográfica*. In: Carlos Eduardo Morelli Tucci (org.). Hidrologia: Ciência e Aplicação, 4ª edição. Porto Alegre: EDUFERS, EDUSP, ABRH, 2012.

TEIXEIRA, C. F. A, DAMÉ, R. C., DISCONZI, P. B., PINTO, M. A. B., WINKLER, A. S., SANTOS J. P. *Estatística de Mallows na seleção de modelos de predição da precipitação média mensal e anual no Rio Grande do Sul*. In: Revista Agro@ambiente On-line, v. 7 n. 2, p. 145-153, mai-ago 2013. Disponível em: <http://www.agroambiente.ufr.br>.

TEIXEIRA-GANDRA, C. F. A., DAMÉ, R. C. F., SIMONETE, M. A. *Predição da precipitação a partir das coordenadas geográficas no Estado do Rio Grande do Sul*. In: Revista Brasileira de Geografia Física. V.8 n. 3, p. 848-856, 2015.

TORRES, F.T. P., MACHADO, P. J. de O. *Introdução a Climatologia*. São Paulo: Cengage Learning, 2011.

TORRES, F.T. P. A qualidade do ar em juiz de fora. Universidade Federal de Juiz de Fora, Minas Gerais, 2003, 121 f.

TUCCI, C. E. M. . *Regionalizações de Vazões*. In: Introdução à Regionalização. Editora Universidade – UFRGS, 2002. Capítulo 1, p. 13 – 22.

TUCCI, C. E. M. . *Hidrologia: ciência e aplicação* / organizado por Carlos E. M. Tucci; André L. L. da Silveira...[et al.]. 4ª edição reimp. Porto Alegre. Editora da UFRGS/ABRH. 2012.

XU, R. E WUNSCH, D. *II survey of clustering algorithms*. IEEE Transactions on Neural Networks 16, 3(may 2005), 645-678.

XIE X. L., BENI G. *A validity measure for fuzzy clustering*, IEEE Trans. Patter Anal. Mach. Intell. 13 (1991) 841-847.

WINDHAN, M. P. *Cluster Validity for fuzzy clustering algorithms*, Fuzzy sets and systems 5, 177-185, 1981.

WOLF, W. *Avaliação e nova proposta de regionalização hidrológica para o estado de São Paulo*. ESALQ/USP, Piracicaba 2013.

ZHANG et. al. *Regionalization and spatial changing properties of droughts across the Pearl River basin, China*. In: Journal of Hydrology. 472-473: 355-366, 2012.

APÊNDICE A – Estações Pluviométricas utilizadas.

Código	Estação	Latitude	Longitude	Altitude (m)	P_anua_med	Município	Estado	Período
01555005	São José da Serra	-15,8364	-55,3219	797	1613,78	Campo Verde	Mato Grosso	1977-1996
01554005	Rio das Mortes	-15,3147	-54,1758	551	1689,7	Primavera do Leste	Mato Grosso	1977-1994
01251000	Alô Brasil	-12,1642	-51,6969	339	1657,61	Bom Jesus do Araguaia	Mato Grosso	1983-2001
01250001	Santo Antonio Leverger	-12,2917	-50,9631	205	1581	Novo Santo Antônio	Mato Grosso	1970-1994
00950001	Barreira do Campo	-9,2275	-50,2108	195	1417,3	Santana Araguaia	Pará	1981-2005
00650001	Fazenda Caiçara	-6,8153	-50,5389	95	1730,1	Marabá	Pará	1985-2005
00350000	Faz. Estrela Norte	-3,8539	-50,4625	22	1931,4	Portel	Pará	1979-2005
00151001	Acampamento IBDF	-1,7922	-51,4344	11	2011,56	Portel	Pará	1981-1997
00251000	Maracacuera Florestal	-2,245	-51,1794	20	2619,68	Portel	Pará	1983-2002
00349001	Cachoeira Tracambeua	-3,51	-49,2092	50	2381,51	Moju	Pará	1983-2000
00148010	Abaetetuba	-1,7381	-48,8589	13	2583,9	Abaetetuba	Pará	1981-2004
00048006	Vigia	-0,8681	-48,1103	15	2843,3	Vigia	Para	1982-2004
00148001	Belém	-1,45	-48,5	10	2810,79	Belém	Pará	1967-1980
00348001	Faz. Maringa	-3,1606	-48,1	20	1933,3	São Domingos Capim	Pará	1983-2004
00248003	Tomé-Açu	-2,4214	-48,1492	45	2552,8	Tomé-Açu	Pará	1984-2004
00949000	Abreulandia	-9:37:28	-49:9:19	240	1941,12	Abreulândia	Tocantins	1975-2005
01147000	Almas	-11:34:44	-47:10:26	427	1524,14	Almas	Tocantins	1975-2005
01753000	Alto Araguaia	-17:18:0	-53:13:10	659	1681,33	Alto Araguaia	Mato Grosso	1975-2005
00648001	Ananas	-6:21:50	-48:4:21	191	1561,50	Ananás	Tocantins	1975-2005
00849002	Araguacema	-8:48:8	-49:33:20	203	2003,77	Araguacema	Tocantins	1975-2005
00548000	Araguatins	-5:38:54	-48:7:30	122	1551,70	Araguatins	Tocantins	1975-2005
00749000	Arapoema	-7:39:19	-49:3:54	215	1866,98	Arapoema	Tocantins	1975-2005
01451000	Aruanã	-14:55:39	-51:4:58	200	1536,86	Aruanã	Goiás	1975-2005
01350000	Bandeirantes	-13:41:22	-50:47:46	276	1456,25	Nova Crixás	Goiás	1975-2005

Continua Apêndice – A

01652000	Bom Jardim de Goiás	-16:12:33	-52:10:7	402	1649,99	Bom Jardim de Goiás	Goiás	1975-2005
01551000	Britânia	-15:14:46	-51:9:48	297	1416,60	Britânia	Goiás	1975-2005
01650000	Cachoeira GO	-16:40:10	-50:38:57	766	1513,54	Cachoeira de Goiás	Goiás	1975-2005
01651000	Caiapônia	-16:56:59	-51:47:58	713	1631,64	Caiapônia	Goiás	1975-2005
01348000	Campinaçu	-13:47:24	-48:34:1	683	2483,06	Campinaçu	Goiás	1975-2005
1347000	Cavalcante	-13:47:49	-47:27:42	821	1849,89	Cavalcante	Goiás	1975-2005
01448000	Colinas do Sul	-14:9:2	-48:4:42	530	1573,37	Colinas do Sul	Goiás	1975-2005
00848000	Colinas TO	-8:3:10	-48:28:54	229	1801,25	Colinas do Tocantins	Tocantins	1975-2005
01248001	Colonha	-12:23:15	-48:42:39	264	1419,7	Peixe	Tocantins	1975-2005
01547010	Contagem	-15:39:11	-47:52:44	1242	1570,26	Brasília	Distrito Federal	1975-2005
01650001	Córrego do ouro	-16:17:54	-50:33:24	569	1544,32	Córrego do Ouro	Goiás	1975-2005
01146000	Dianópolis	-11:37:31	-46:48:38	679	1449,21	Dianópolis	Tocantins	1975-2005
00949001	Dois Irmãos Tocantins	-9:15:26	-49:3:51	264	1706,89	Dois Irmãos Tocantins	Tocantins	1975-2005
01349003	Entroncamento S M	-13:6:10	-49:12:5	345	1634,60	Porangatu	Goiás	1975-2005
01349000	Estrela do Norte	-13:52:21	-49:4:17	467	1750,67	Estrela do Norte	Goiás	1975-2005
01048000	Fátima	-10:45:49	-48:54:8	352	1897,41	Fátima	Tocantins	1975-2005
00748002	Faz Primavera	-7:33:38	-48:25:22	257	1816,08	Nova Olinda	Tocantins	1975-2005
01752002	Faz São Bernardo	-17:41:20	-52:59:0	750	1674,44	Mineiros	Goiás	1975-2005
01148000	Faz. Lobeira	-11:31:53	-48:17:41	243	1556,30	São Valério da Natividade	Tocantins	1975-2005
01547001	Faz. Santa sé	-15:12:58	-47:9:25	573	1684,06	Formosa	Goiás	1975-2005
01447001	Flores GO	-14:27:0	-47:2:45	200	1144,44	Flores Goiás	Goiás	1975-2005
00747001	Goiantins	-7:42:53	-47:18:54	185	1572,48	Goiantins	Tocantins	1975-2005
01650002	Israelândia	-16:18:59	-50:54:22	406	1597,32	Israelândia	Goiás	1975-2005
01649007	Itaberaí	-16:1:39	-49:48:1	726	1827,62	Itaberaí	Goiás	1975-2005
00847001	Itacaja	-8:23:30	-47:45:46	250	1844,53	Itacajá	Tocantins	1975-2005

Continuação Apêndice A

01550000	Itapirapua	-15:48:57	-50:36:31	343	1589,17	Itapirapuã	Goiás	1975-2005
01549002	Itapuranga	-15:33:51	-49:56:37	646	1644,69	Itapuranga	Goiás	1975-2005
01550001	Jeroaquara	-15:22:23	-50:30:0	400	1779,80	Faina	Goiás	1975-2005
01450000	Lagoa da Flexa	-14:19:58	-50:43:49	200	1436,41	Nova Crixás	Goiás	1975-2005
01548001	Mimoso	-15:3:47	-48:9:59	687	1307,59	Mimoso de Goiás	Goiás	1975-2005
00948000	Miracema Tocantins	-9:33:51	-48:23:15	210	1706,89	Miracema do Tocantins	Tocantins	1975-2005
01551001	Monte Carlos GO	-15:56:27	-51:21:26	400	1542,97	Monte Carlos Goiás	Goiás	1975-2005
01450001	Mozarlandia	-14:44:29	-50:34:38	400	1654,07	Mozarlândia	Goiás	1975-2005
00748003	Muricilandia	-7:9:16	-48:36:11	393	1671,4	Muricilândia	Tocantins	1975-2005
01448001	Niquelândia	-14:28:30	-48:27:30	568	1703,61	Niquelândia	Goiás	1975-2005
01549004	Nova América	-15:1:14	-49:53:30	800	1605,70	Nova América	Goiás	1975-2005
01047001	Novo Acordo	-9:57:38	-47:40:29	300	1598,19	Novo Acordo	Tocantins	1975-2005
01349001	Novo Planalto	-13:14:41	-49:30:6	286	1588,39	Novo Planalto	Goiás	1975-2005
01048001	Paraíso do TO	-10:9:55	-48:53:26	390	2280,65	Paraíso do Tocantins	Tocantins	1975-2005
01551002	Perez	-15:53:24	-51:51:13	299	1499,26	Aragarças	Goiás	1975-2005
01449000	Pilar de Goiás	-14:45:40	-49:34:49	765	1948,00	Pilar de Goiás	Goiás	1975-2005
01147002	Pindorama do Tocantins	-11:8:25	-47:34:36	444	1614,57	Pindorama Tocantins	Tocantins	1975-2005
01651002	Piranhas	-16:25:24	-51:49:22	356	1582,92	Piranhas	Goiás	1975-2005
00648002	Piraquê	-6:40:20	-48:28:13	184	1760,80	Araguanã	Tocantins	1975-2005
01347001	Ponte Paranã	-13:25:31	-47:8:21	363	1245,44	Teresina de Goiás	Goiás	1975-2005
01047002	Porto Gilandia	-10:45:19	-47:45:53	220	1656,37	Monte do Carmo	Tocantins	1975-2005
00948001	Porto Real	-9:18:25	-47:55:45	200	1599,03	Pedro Afonso	Tocantins	1975-2005
01449001	Porto Uruaçu	-14:33:19	-49:8:22	572	1468,39	Uruaçu	Goiás	1975-2005
01350001	Rio Pintado	-13:31:43	-50:11:17	200	1443,73	São Miguel do Araguaia	Goiás	1975-2005
01348001	Sama	-13:31:59	-48:13:37	375	1410,71	Minaçu	Goiás	1975-2005

Continuação Apêndice A

01551003	Santa fé	-15:46:2	-51:6:16	400	1614,51	Santa fé de Goiás	Goiás	1975-2005
01449002	Santa Terezinha GO	-14:26:1	-49:43:0	400	1504,88	Santa Terezinha de Goiás	Goiás	1975-2005
01651003	São Ferreira	-16:18:21	-51:28:14	361	1673,12	Arenópolis	Goiás	1975-2005
01447002	São João Aliança	-14:42:26	-47:31:25	1009	1498,55	São João D'aliança	Goiás	1975-2005
01653000	Tesouro	-16:4:40	-53:32:58	389	1714,94	Tesouro	Mato Grosso	1975-2005
01652002	Torixoreu	-16:12:5	-52:33:0	307	1405,73	Torixoréu	Mato Grosso	1975-2005
01550002	Travessão	-15:32:8	-50:42:10	450	1516,55	Matrinchã	Goiás	1975-2005
00848003	Tupiratins	-8:23:57	-48:7:49	192	1740,09	Tupiratins	Tocantins	1975-2005
00648000	Xambioá	-6:24:47	-48:32:0	148	1694,99	Xambioá	Tocantins	1975-2005
01452000	Xavantina	-14:40:20	-52:21:17	263	1526,45	Nova Xavantina	Mato Grosso	1975-2005

APÊNDICE B - Algoritmo Fuzzy C-means – FCM.

```

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%-----FCM-----%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function [center, U, obj_fcn] = (data, cluster_n, options)
%-----
%VARIÁVEIS
%-----
%data= matrix de dados que contém as características
fisiográficas de cada estação fluviométrica.
%cluster_n = número de grupos pré-determinado para a
classificação.
data = load('matrix.txt');
cluster_n = 8;
expo=2.0; %parâmetro de fuzificação.
max_iter= 200; %número máximo de iterações.
min_impro=1e-5; %mínima diferença entre os elementos.
display= 1;
%-----
FUNÇÃO FCM
%-----
[center,U,obj_fcn] = fcm(data,cluster_n,options);
    maxU = max(U);
%-----
%PROCESSAMENTO DOS CLUSTER
%-----
U(1,:);
U(2,:);
U(3,:);
U(4,:);
U(5,:);
U(6,:);
U(7,:);
U(8,:);
%-----
%INICIALIZAÇÃO - CALCULA GRAUS DE PERTINÊNCIA DOS CLUSTERS
%-----
index1 = find(U(1,:) == maxU);
index2 = find(U(2,:) == maxU);
index3 = find(U(3,:) == maxU);
index4 = find(U(4,:) == maxU);
index5 = find(U(5,:) == maxU);
index6 = find(U(6,:) == maxU);
index7 = find(U(7,:) == maxU);
index8 = find(U(8,:) == maxU);
line(data(index1,1),data(index1,2),
data(index1,3),'marker','*','color','g');
line(data(index2,1),
data(index2,3),'marker','*','color','r');
line(data(index3,1),
data(index3,3),'marker','*','color','b');
data(index2,2),
data(index3,2),

```

```

line(data(index4,1), data(index4,2),
data(index4,3), 'marker', '*', 'color', 'k');
line(data(index5,1), data(index5,2),
data(index5,3), 'marker', '*', 'color', 'm');
line(data(index6,1), data(index6,2),
data(index6,3), 'marker', '*', 'color', 'y');
line(data(index7,1), data(index7,2),
data(index7,3), 'marker', '*', 'color', 'c');
line(data(index8,1), data(index8,2),
data(index8,3), 'marker', '*', 'color', 'c');
plot((center([1 8],1)), (center([1 8],2)), (center([1
8],3)), (center([1 8],4)), 'marker', 'o', 'color', 'k');
hold on;

```

```

%-----
%CONDIÇÕES DO ALGORITMO
%-----
if nargin < 2
    error(message);
end

data_n = size(data, 1);
in_n = size(data, 2);

if nargin == 2,
    options = [2.0; 100; 1e-5; 1];
else

    if length(options) < 4,
        tmp = options;
        tmp(1:length(options)) = options;
        options = tmp;
    end

    nan_index = find(isnan(options)==1);
    options(nan_index) = options(nan_index);
    if options(1) <= 1,
        error('O expoente deve ser maior que 1');
    end
end

obj_fcn = zeros(max_iter, 1);

U = initfcm(cluster_n, data_n);
fprintf('U = %d, U= %f\n', cluster_n);

for i = 1:max_iter,
    [U, center, obj_fcn(i)] = stepfcm(data, U, cluster_n,
expo);
    if display,

```

```

        fprintf('Iteration count = %d, obj. fcn = %f\n', i,
obj_fcn(i));
    end

    if i > 1,
        if abs(obj_fcn(i) - obj_fcn(i-1)) < min_impro,
break; end,
        end
    end

    iter_n = i;
    obj_fcn(iter_n+1:max_iter) = [0.1 0.5 0.9];
    out = zeros(size(center, 1), size(data, 1));

    %Distância de Similaridade
    if size(center, 2) > 1,
        for k = 1:size(center, 1),
            out(k, :) = sqrt(SUM(((data-ones(size(data, 1),
1)*center(k, :)).^2)'));
        end
    else
        for k = 1:size(center, 1),
            out(k, :) = abs(center(k)-data)';
        end
    end

```

APÊNDICE C – Código Computacional FDP

```

%-----Funções de Distribuições de Probabilidade - FDP--%
Essa programação tem como objetivo, gerar as funções de
distribuição de probabilidade Normal, Log-Normal, Gama 2
parâmetros, Weibull, Exponencial e Gumbel. Para sua execução
é necessário fornecer a matriz de dados contendo as séries
da variável estudada (precipitação). Os Resultados obtidos
são as frequências observadas e as simuladas pelas FDP.
%-----%
Series = xlsread('E29.xlsx',1,'A2:L32'); %(Séries de chuva)

col = size(Series,2);
for i = 1:col

    M = Series(~isnan(Series(:,i)),i);

    M_Ord = sort(M);
    N = size(M,1);

    media = mean(M);
    dp = std(M);

    Logn_P = lognfit(M);      %Parâmetros da Lognormal

    a_gam = (media/dp)^2;     %Parâmetros da distribuição
Gama
    b_gam = (dp^2)/media;

    mu_exp = expfit(M);      %Parâmetro da distribuição
Exponencial

    a_gum = 1.2826/dp;       %Parâmetros da distribuição de
Gumbel
    mu_gum = media-0.45*dp;

    a_wei = 1.0122*(dp/media)^(-1.0779); %Parâmetros da
distribuição de Weibull
    b_wei = media/(0.9982-
0.4419*dp/media+0.436*(dp/media)^2);

    [F_Obs, F_Gumbel] = deal(zeros(N,1));

    for j = 1:N
        F_Obs(j) = j/(N+1);
        F_Gumbel(j) = 1-exp(-exp(a_gum*(M_Ord(j)-mu_gum)));
    end

    F_Norm = normcdf(M_Ord,media,dp);
    F_Lnorm = logncdf(M_Ord,Logn_P(1),Logn_P(2));
    F_Gama = gamcdf(M_Ord,a_gam,b_gam);

```

```

F_Exp = expcdf(M_Ord,mu_exp);
F_Wei = wblcdf(M_Ord,b_wei,a_wei);

Diff_Norm = abs(F_Obs - F_Norm);
Diff_Lnorm = abs(F_Obs - F_Lnorm);
Diff_Gama = abs(F_Obs - F_Gama);
Diff_Gumbel = abs(F_Obs - F_Gumbel);
Diff_Exp = abs(F_Obs - F_Exp);
Diff_Wei = abs(F_Obs - F_Wei);

T_Norm = [M,M_Ord,F_Obs,F_Norm,Diff_Norm];
T_Lnorm = [M,M_Ord,F_Obs,F_Lnorm,Diff_Lnorm];
T_Gama = [M,M_Ord,F_Obs,F_Gama,Diff_Gama];
T_Gumbel = [M,M_Ord,F_Obs,F_Gumbel,Diff_Gumbel];
T_Exp = [M,M_Ord,F_Obs,F_Exp,Diff_Exp];
T_Wei = [M,M_Ord,F_Obs,F_Wei,Diff_Wei];

est = strcat('M',num2str(i)); %Obs.: Nome da estação ou
mês trabalhado;
arquivo = strcat(est,'.xlsx'); % Obs.: Nome do arquivo
gerado;

Enc_Norm = {'P Med Anual', 'P Ordenada', 'Freq Obs
(Prob P<=Pi)', 'Dist Normal (Prob P<=Pi)', '|Freq Emp - Freq
Teor|'};
Enc_Lnorm = {'P Med Anual', 'P Ordenada', 'Freq Obs
(Prob P<=Pi)', 'Dist Lognormal (Prob P<=Pi)', '|Freq Emp -
Freq Teor|'};
Enc_Gama = {'P Med Anual', 'P Ordenada', 'Freq Obs
(Prob P<=Pi)', 'Dist Gama (Prob P<=Pi)', '|Freq Emp - Freq
Teor|'};
Enc_Gumbel = {'P Med Anual', 'P Ordenada', 'Freq Obs
(Prob P<=Pi)', 'Dist Gumbel (Prob P<=Pi)', '|Freq Emp - Freq
Teor|'};
Enc_Exp = {'P Med Anual', 'P Ordenada', 'Freq Obs (Prob
P<=Pi)', 'Dist Exponencial (Prob P<=Pi)', '|Freq Emp - Freq
Teor|'};
Enc_Wei = {'P Med Anual', 'P Ordenada', 'Freq Obs (Prob
P<=Pi)', 'Dist Weibull (Prob P<=Pi)', '|Freq Emp - Freq
Teor|'};

set(gcf,'PaperUnits','inches','PaperPosition',[0 0 18
8])

subplot(2,3,1);
plot(flipud(M_Ord), F_Norm, 'o', flipud(M_Ord), F_Obs,
'--');

xlabel('Precipitação (mm)');
ylabel('Probabilidade de ocorrência');

```

```

legend('Feq Norm','Feq Obs');

subplot(2,3,2);
plot(flipud(M_Ord), F_Lnorm, 'o', flipud(M_Ord), F_Obs,
'--');
title(strcat('FDP - ',est));
xlabel('Precipitação (mm)');
ylabel('Probabilidade de ocorrência');
legend('Feq Lnorm','Feq Obs');

subplot(2,3,3);
plot(flipud(M_Ord), F_Gama, 'o', flipud(M_Ord), F_Obs,
'--');
xlabel('Precipitação (mm)');
ylabel('Probabilidade de ocorrência');
legend('Feq Gama','Feq Obs');

subplot(2,3,4);
plot(flipud(M_Ord), F_Gumbel, 'o', flipud(M_Ord),
F_Obs, '--');
xlabel('Precipitação (mm)');
ylabel('Probabilidade de ocorrência');
legend('Feq Gumbel','Feq Obs');

subplot(2,3,5);
plot(flipud(M_Ord), F_Exp, 'o', flipud(M_Ord), F_Obs,
'--');
xlabel('Precipitação (mm)');
ylabel('Probabilidade de ocorrência');
legend('Feq Exp','Feq Obs');

subplot(2,3,6);
plot(flipud(M_Ord), F_Wei, 'o', flipud(M_Ord), F_Obs,
'--');
xlabel('Precipitação (mm)');
ylabel('Probabilidade de ocorrência');
legend('Feq Weibull','Feq Obs');

print(est, '-dpng', '-r100');

close(figure);

xlswrite(arquivo,Enc_Norm,1,'A1');
xlswrite(arquivo,T_Norm,1,'A2');

xlswrite(arquivo,Enc_Lnorm,2,'A1');
xlswrite(arquivo,T_Lnorm,2,'A2');

xlswrite(arquivo,Enc_Gama,3,'A1');
xlswrite(arquivo,T_Gama,3,'A2');

```

```
xlswrite(arquivo,Enc_Gumbel,4,'A1');  
xlswrite(arquivo,T_Gumbel,4,'A2');  
  
xlswrite(arquivo,Enc_Exp,5,'A1');  
xlswrite(arquivo,T_Exp,5,'A2');  
  
xlswrite(arquivo,Enc_Wei,6,'A1');  
xlswrite(arquivo,T_Wei,6,'A2');  
end  
.
```

APÊNDICE D - Graus de Pertinência das Estações.

ID	G1	G2	G3	Decisão	ID	G1	G2	G3	Decisão	ID	G1	G2	G3	Decisão	ID	G1	G2	G3	Decisão
E1	0,0165	0,9246	0,0589	2	E24	0,0161	0,1475	0,8365	3	E46	0,0189	0,1764	0,8047	3	E68	0,0509	0,2686	0,6805	3
E2	0,0161	0,8366	0,1473	2	E25	0,0121	0,0590	0,9289	3	E47	0,0107	0,9511	0,0381	2	E69	0,0323	0,0759	0,8918	3
E3	0,0206	0,1012	0,8782	3	E26	0,0345	0,7889	0,1766	2	E48	0,1951	0,2331	0,5718	3	E70	0,0234	0,0558	0,9208	3
E4	0,0169	0,0545	0,9286	3	E27	0,0077	0,9514	0,0409	2	E49	0,0047	0,0281	0,9672	3	E71	0,0221	0,7382	0,2397	2
E5	0,0574	0,1868	0,7558	3	E28	0,5344	0,2775	0,1881	1	E50	0,0010	0,9935	0,0055	2	E72	0,0256	0,0835	0,8909	3
E6	0,0973	0,1282	0,7745	3	E29	0,0909	0,7477	0,1614	2	E51	0,0760	0,3648	0,5592	3	E73	0,0295	0,1949	0,7756	3
E7	0,2981	0,1745	0,5274	3	E30	0,0250	0,6215	0,3535	2	E52	0,0214	0,0688	0,9098	3	E74	0,0156	0,1408	0,8437	3
E8	0,5205	0,1505	0,3291	1	E31	0,1720	0,2116	0,6164	3	E53	0,0522	0,6219	0,3259	2	E75	0,0155	0,1404	0,8441	3
E10	0,9532	0,0210	0,0258	1	E32	0,0343	0,1418	0,8239	3	E54	0,0416	0,0893	0,8691	3	E76	0,0154	0,0959	0,8887	3
E11	0,9057	0,0379	0,0564	1	E33	0,1164	0,6457	0,2379	2	E55	0,0158	0,1412	0,8431	3	E77	0,0678	0,7346	0,1976	2
E12	0,9245	0,0345	0,0411	1	E34	0,0119	0,8883	0,0999	2	E56	0,0154	0,1391	0,8455	3	E78	0,0720	0,3333	0,5947	3
E13	0,3742	0,1718	0,4540	3	E35	0,0168	0,8732	0,1100	2	E57	0,0165	0,1374	0,8461	3	E79	0,0225	0,1126	0,8649	3
E14	0,8126	0,0898	0,0976	1	E36	0,5981	0,1732	0,2287	1	E58	0,0340	0,8084	0,1575	2	E80	0,0339	0,3208	0,6454	3
E15	0,8239	0,0793	0,0968	1	E37	0,0153	0,0829	0,9018	3	E59	0,0333	0,8772	0,0895	2	E81	0,0489	0,0958	0,8553	3
E16	0,0252	0,2434	0,7313	3	E38	0,0431	0,5284	0,4286	2	E60	0,0018	0,0086	0,9897	3	E82	0,0272	0,0633	0,9095	3
E17	0,0038	0,9788	0,0174	2	E39	0,1432	0,3233	0,5335	3	E61	0,0093	0,0350	0,9557	3	E83	0,0099	0,0407	0,9494	3
E18	0,0219	0,0675	0,9106	3	E40	0,1493	0,2214	0,6293	3	E62	0,2455	0,3656	0,3888	3					
E19	0,7096	0,1199	0,1706	1	E41	0,0080	0,9572	0,0348	2	E63	0,0082	0,0403	0,9515	3					
E20	0,0413	0,1007	0,8580	3	E42	0,0091	0,0340	0,9569	3	E64	0,1146	0,7038	0,1816	2					
E21	0,0946	0,1559	0,7495	3	E43	0,0142	0,8909	0,0949	2	E65	0,0247	0,2832	0,6921	3					
E22	0,0125	0,0375	0,9500	3	E44	0,0833	0,2601	0,6566	3	E66	0,0206	0,1168	0,8627	3					
E23	0,0008	0,0036	0,9956	3	E45	0,0271	0,0610	0,9118	3	E67	0,0351	0,0740	0,8909	3					

APENDICE E – Quocientes de Momentos-L.

Estações	Cv-L	γ -L	K-L
E3	0,081	-2,417	-4,844
E4	0,099	-1,592	-4,650
E5	0,148	-1,059	-4,999
E6	0,178	-0,961	-5,546
E7	0,111	-1,191	-5,167
E14	0,144	-1,171	-5,017
E17	0,097	-1,093	-5,400
E19	0,091	-1,355	-4,608
E21	0,105	-1,051	-5,284
E22	0,076	-1,488	-5,023
E23	0,108	-0,879	-6,089
E24	0,130	-0,911	-4,968
E25	0,107	-0,828	-6,697
E26	0,119	-0,776	-6,442
E32	0,091	-1,056	-6,039
E33	0,092	-1,072	-5,858
E38	0,119	-0,859	-5,646
E40	0,095	-0,972	-6,387
E41	0,099	-0,980	-6,078
E43	0,083	-1,366	-5,030
E45	0,113	-0,993	-5,165
E46	0,099	-0,934	-6,225
E47	0,116	-0,870	-5,900
E49	0,096	-1,067	-5,685
E50	0,098	-1,010	-5,886
E52	0,114	-0,969	-5,289
E53	0,108	-0,993	-5,552
E55	0,088	-1,172	-5,577
E56	0,101	-1,072	-5,339
E57	0,101	-1,045	-5,481
E58	0,119	-0,834	-5,890
E61	0,091	-1,105	-5,700
E62	0,129	-0,807	-5,753
E63	0,107	-0,792	-7,001
E64	0,116	-0,814	-6,205
E66	0,089	-1,231	-5,245
E67	0,116	-0,862	-5,949
E68	0,117	-0,748	-6,852
E69	0,068	-4,090	-4,447
E70	0,095	-1,093	-5,568
E71	0,103	-0,888	-6,400
E73	0,128	-0,901	-5,189

Continuação Apêndice E - quocientes de Momentos-L.

Estações	Cv-L	γ -L	K-L
E74	0,139	-0,834	-5,252
E75	0,115	-0,885	-5,802
E76	0,119	-0,970	-5,201
E77	0,115	-0,880	-5,928
E79	0,083	-1,441	-4,762
E80	0,112	-0,803	-6,616
E81	0,106	-1,015	-5,502
E82	0,102	-0,837	-6,875
E83	0,125	-0,812	-5,881
E84	0,090	-1,029	-6,255
E1	0,130	-1,595	-4,635
E2	0,072	-2,767	-5,087
E18	0,083	-1,431	-4,771
E27	0,115	-1,059	-4,804
E28	0,126	-0,699	-6,755
E30	0,148	-0,675	-6,129
E31	0,110	-0,963	-5,419
E34	0,111	-0,980	-5,347
E35	0,138	-0,775	-5,675
E36	0,103	-1,122	-5,071
E39	0,109	-0,975	-5,518
E42	0,071	-1,508	-5,308
E44	0,134	-0,735	-6,163
E48	0,092	-1,194	-5,346
E51	0,088	-1,140	-5,579
E54	0,094	-1,133	-5,382
E59	0,126	-0,766	-6,224
E60	0,116	-0,864	-5,806
E65	0,152	-0,517	-7,985
E72	0,116	-0,857	-5,938
E78	0,146	-0,607	-7,033
E8	0,098	-2,073	-5,457
E9	0,104	-1,768	-5,064
E10	0,150	-1,374	-5,257
E11	0,114	-1,301	-5,221
E12	0,097	-1,693	-3,367
E15	0,120	-1,511	-4,905
E16	0,095	-1,138	-5,365
E20	0,095	-1,123	-5,382
E29	0,167	-0,539	-6,992
E37	0,086	-1,155	-5,763

APÊNDICE F – Valores dos Índices de Desempenho obtidos na calibração dos modelos de Regressão para as médias mensais

Região Homogênea I

Região Homogênea I																								
Linear							Potencial						Exponencial						Logaritmo					
Mês	R ²	R ² _a	E(%)	ε (%)	Nash	RMSE	R ²	R ² _a	E(%)	ε (%)	Nash	RMSE	R ²	R ² _a	E(%)	ε (%)	Nash	RMSE	R ²	R ² _a	E(%)	ε (%)	Nash	RMSE
Jan	0,48	0,45	17,69	0,37	0,43	0,382	0,40	0,51	7,66	0,05	0,27	0,001	0,47	0,43	7,67	0,06	0,49	0,001	0,42	0,38	7,70	0,05	0,42	0,001
fer	0,51	0,48	18,09	0,08	0,51	0,001	0,51	0,48	6,84	0,08	0,27	0,001	0,51	0,47	7,15	0,08	0,50	0,001	0,14	0,09	7,20	0,06	0,34	0,001
mar	0,46	0,64	19,29	0,06	0,66	0,001	0,67	0,65	8,21	0,05	0,27	0,001	0,68	0,66	8,21	0,06	0,66	0,001	0,69	0,67	8,28	0,05	0,69	0,001
abr	0,38	0,88	12,69	0,01	0,88	0,001	0,87	0,86	12,73	0,01	0,27	0,001	0,88	0,87	11,30	0,01	0,90	0,001	0,91	0,91	11,29	0,01	0,91	0,001
mai	0,82	0,80	46,16	0,00	0,82	0,001	0,91	0,90	14,24	0,01	0,27	0,003	0,89	0,89	14,41	0,00	0,93	0,002	0,92	0,92	20,81	0,00	0,92	0,002
jun	0,57	0,57	96,15	0,00	0,57	0,002	0,67	0,67	41,25	0,00	0,27	0,004	0,57	0,57	49,00	0,00	0,67	0,003	0,78	0,78	81,35	0,00	0,78	0,003
jul	0,57	0,54	115,51	0,00	0,57	0,004	0,61	0,59	50,53	0,00	0,27	0,006	0,51	0,48	60,46	0,00	0,64	0,004	0,77	0,76	96,40	0,00	0,77	0,005
ago	0,58	0,56	45,15	0,00	0,58	0,005	0,64	0,62	30,81	0,00	0,27	0,005	0,57	0,54	34,26	0,00	0,59	0,005	0,67	0,65	40,29	0,00	0,67	0,005
set	0,32	0,28	21,44	0,00	0,32	0,001	0,46	0,43	17,77	0,00	0,27	0,002	0,34	0,30	19,85	0,00	0,30	0,001	0,39	0,35	19,98	0,00	0,39	0,002
out	0,12	0,07	16,96	0,00	0,12	0,001	0,44	0,41	14,66	0,01	0,27	0,001	0,15	0,10	17,36	0,00	0,07	0,001	0,34	0,30	14,26	0,01	0,34	0,001
nov	0,51	0,48	12,13	0,00	0,51	0,001	0,74	0,73	9,88	0,03	0,27	0,001	0,52	0,49	13,34	0,01	0,39	0,001	0,67	0,65	8,75	0,03	0,67	0,001
dez	0,67	0,65	7,23	0,07	0,67	0,001	0,75	0,74	6,52	0,09	0,27	0,001	0,69	0,67	7,55	0,07	0,64	0,001	0,70	0,68	6,87	0,08	0,70	0,001

Continuação do Apêndice F

Região Homogênea II

Região Homogênea I																								
Linear							Potencial					Exponencial					Logaritmo							
Mês	R ²	R ² _a	E(%)	ε (%)	Nash	RMSE	R ²	R ² _a	E(%)	ε (%)	Nash	RMSE	R ²	R ² _a	E(%)	ε (%)	Nash	RMSE	R ²	R ² _a	E(%)	ε (%)	Nash	RMSE
Jan	0,05	0,12	9,71	0,171	0,05	0,001	0,05	0,12	9,62	0,176	0,03	0,001	0,06	0,10	9,46	0,177	0,04	0,001	0,04	0,14	9,92	0,169	0,04	0,001
fer	0,07	0,09	8,95	0,212	0,07	0,002	0,06	0,10	90,03	0,214	0,06	0,002	0,06	0,10	100	0,000	0,08	0,0002	0,07	0,10	97,52	0,000	0,05	0,0002
mar	0,07	0,10	7,21	0,249	0,07	0,002	0,07	0,09	7,32	0,254	0,07	0,002	0,07	0,10	7,29	0,253	0,06	0,002	0,07	0,09	7,35	0,250	0,07	0,002
abr	0,62	0,56	8,47	0,225	0,62	0,005	0,58	0,50	8,49	0,235	0,64	0,005	0,58	0,50	58,37	0,235	0,63	0,005	0,63	0,56	18,68	0,224	0,63	0,005
mai	0,76	0,71	80,9	0,088	0,76	0,009	0,73	0,68	12,57	0,118	0,78	0,009	0,72	0,67	12,02	0,118	0,78	0,010	0,76	0,71	14,38	0,087	0,76	0,009
jun	0,78	0,74	36,45	0,000	0,78	0,016	0,83	0,79	22,82	0,013	0,83	0,018	0,84	0,81	22,11	0,017	0,85	0,020	0,76	0,72	37,2	0,000	0,76	0,015
jul	0,81	0,78	60,11	0,003	0,81	0,037	0,8	0,76	54,64	0,010	0,75	0,032	0,82	0,78	23,24	0,012	0,77	0,033	0,79	0,76	31,79	0,002	0,79	0,035
ago	0,77	0,73	57,25	0,051	0,77	0,019	0,78	0,75	15,97	0,065	0,77	0,019	0,79	0,75	15,67	0,070	0,79	0,020	0,76	0,71	17,91	0,046	0,76	0,019
set	0,83	0,80	10,49	0,157	0,83	0,007	0,82	0,79	11,19	0,152	0,82	0,007	0,82	0,79	41,32	0,150	0,82	0,007	0,83	0,80	75,8	0,154	0,83	0,007
out	0,17	0,02	11,64	0,257	0,17	0,004	0,18	0,04	12,73	0,261	0,17	0,004	0,17	0,03	7,68	0,260	0,16	0,004	0,18	0,03	67,69	0,258	0,18	0,004
nov	0,09	0,07	11,05	0,299	0,09	0,002	0,07	0,09	82,2	0,295	0,07	0,002	0,09	0,07	96,99	0,301	0,09	0,002	0,07	0,09	27,24	0,293	0,07	0,002
dez	0,06	0,13	10,35	0,239	0,26	0,002	0,26	0,12	7,43	0,242	0,24	0,002	0,27	0,14	7,43	0,241	0,26	0,002	0,25	0,12	7,46	0,238	0,25	0,002

Continuação do Apêndice F

Região Homogênea III

Região Homogênea III																									
Linear							Potencial						Exponencial						Logaritmo						
Mês	R ²	R ² _a	E(%)	€ (%)	Nash	RMSE	R ²	R ² _a	E(%)	€ (%)	Nash	RMSE	R ²	R ² _a	E(%)	€ (%)	Nash	RMSE	R ²	R ² _a	E(%)	€ (%)	Nash	RMSE	
Jan	0,63	0,44	6,91	0,57	0,63	0,004	0,62	0,43	16,68	0,58	0,63	0,004	0,59	0,39	6,87	0,57	0,63	0,004	0,62	0,43	16,75	0,58	0,62	0,003	
fer	0,40	0,07	92,00	0,43	0,38	0,003	0,41	0,12	87,89	0,45	0,39	0,003	0,39	0,08	8,86	0,44	0,37	0,003	0,41	0,12	27,95	0,44	0,41	0,002	
mar	0,78	0,68	60,62	0,49	0,78	0,003	0,63	0,44	10,74	0,40	0,65	0,002	0,77	0,65	6,70	0,50	0,78	0,003	0,67	0,50	10,43	0,40	0,67	0,002	
abr	0,90	0,85	72,01	0,50	0,90	0,003	0,82	0,73	12,61	0,34	0,73	0,002	0,92	0,88	7,41	0,48	0,88	0,003	0,80	0,71	11,40	0,38	0,80	0,002	
mai	0,97	0,95	13,27	0,26	0,97	0,005	0,93	0,89	20,33	0,19	0,75	0,002	0,98	0,97	11,46	0,36	0,95	0,003	0,92	0,88	24,73	0,09	0,92	0,002	
jun	0,94	0,91	49,98	0,00	0,94	0,005	0,92	0,88	37,89	0,04	0,60	0,002	0,97	0,96	109,90	0,18	0,93	0,004	0,93	0,90	87,89	0,00	0,93	0,003	
jul	0,91	0,87	105,00	0,00	0,91	0,005	0,92	0,87	95,72	0,02	0,29	0,002	0,98	0,97	22,37	0,13	0,89	0,004	0,53	0,90	124,00	0,00	0,93	0,004	
ago	0,86	0,79	40,21	0,01	0,86	0,005	0,88	0,83	28,99	0,09	0,63	0,003	0,92	0,89	21,34	0,09	0,88	0,005	0,87	0,80	32,10	0,03	0,87	0,004	
set	0,25	0,13	35,80	0,00	0,25	0,004	0,20	0,20	32,62	0,01	0,22	0,004	0,20	0,20	32,25	0,01	0,20	0,004	0,25	0,12	36,26	0,00	0,25	0,003	
out	0,58	0,82	20,99	0,05	0,88	0,005	0,74	0,60	23,27	0,05	0,78	0,004	0,80	0,70	19,23	0,10	0,90	0,006	0,78	0,67	27,60	0,02	0,78	0,003	
nov	0,99	0,99	7,09	0,48	0,99	0,012	0,17	0,96	19,11	0,46	0,98	0,007	0,97	0,95	8,14	0,43	0,99	0,010	0,96	0,94	17,78	0,19	0,96	0,004	
dez	0,98	0,98	4,73	0,63	0,98	0,008	0,25	0,93	7,91	0,49	0,96	0,005	0,98	0,98	4,91	0,67	0,99	0,009	0,92	0,88	12,10	0,33	0,92	0,002	

ANEXO A – Tabela dos Valores do Qui-quadrado, conforme graus de liberdade e nível de significância.

gl	0,990	0,950	0,900	0,500	0,100	0,050	0,025	0,010	0,005
1	0,0002	0,0039	0,0158	0,4549	2,7055	3,8415	5,0239	6,6349	7,8794
2	0,0201	0,1026	0,2107	1,3863	4,6052	5,9915	7,3778	9,2104	10,5965
3	0,1148	0,3518	0,5844	2,3660	6,2514	7,8147	9,3484	11,3449	12,8381
4	0,2971	0,7107	1,0636	3,3567	7,7794	9,4877	11,1433	13,2767	14,8602
5	0,5543	1,1455	1,6103	4,3515	9,2363	11,0705	12,8325	15,0863	16,7496
6	0,8721	1,6354	2,2041	5,3481	10,6446	12,5916	14,4494	16,8119	18,5475
7	1,2390	2,1673	2,8331	6,3458	12,0170	14,0671	16,0128	18,4753	20,2777
8	1,6465	2,7326	3,4895	7,3441	13,3616	15,5073	17,5345	20,0902	21,9549
9	2,0879	3,3251	4,1682	8,3428	14,6837	16,9190	19,0228	21,6660	23,5893
10	2,5582	3,9403	4,8652	9,3418	15,9872	18,3070	20,4832	23,2093	25,1881
11	3,0535	4,5748	5,5778	10,3410	17,2750	19,6752	21,9200	24,7250	26,7569
12	3,5706	5,2260	6,3038	11,3403	18,5493	21,0261	23,3367	26,2170	28,2997
13	4,1069	5,8919	7,0415	12,3398	19,8119	22,3620	24,7356	27,6882	29,8193
14	4,6604	6,5706	7,7895	13,3393	21,0641	23,6848	26,1189	29,1412	31,3194
15	5,2294	7,2609	8,5468	14,3389	22,3071	24,9958	27,4884	30,5780	32,8015
16	5,8122	7,9616	9,3122	15,3385	23,5418	26,2962	28,8453	31,9999	34,2671
17	6,4077	8,6718	10,0852	16,3382	24,7690	27,5871	30,1910	33,4087	35,7184
18	7,0149	9,3904	10,8649	17,3379	25,9894	28,8693	31,5264	34,8052	37,1564
19	7,6327	10,1170	11,6509	18,3376	27,2036	30,1435	32,8523	36,1908	38,5821
20	8,2604	10,8508	12,4426	19,3374	28,4120	31,4104	34,1696	37,5663	39,9969
21	8,8972	11,5913	13,2396	20,3372	29,6151	32,6706	35,4789	38,9322	41,4009
22	9,5425	12,3380	14,0415	21,3370	30,8133	33,9245	36,7807	40,2894	42,7957
23	10,1957	13,0905	14,8480	22,3369	32,0069	35,1725	38,0756	41,6383	44,1814
24	10,8563	13,8484	15,6587	23,3367	33,1962	36,4150	39,3641	42,9798	45,5584
25	11,5240	14,6114	16,4734	24,3366	34,3816	37,6525	40,6465	44,3140	46,9280
26	12,1982	15,3792	17,2919	25,3365	35,5632	38,8851	41,9231	45,6416	48,2898
27	12,8785	16,1514	18,1139	26,3363	36,7412	40,1133	43,1945	46,9628	49,6450
28	13,5647	16,9279	18,9392	27,3362	37,9159	41,3372	44,4608	48,2782	50,9936
29	14,2564	17,7084	19,7677	28,3361	39,0875	42,5569	45,7223	49,5878	52,3355
30	14,9535	18,4927	20,5992	29,3360	40,2560	43,7730	46,9792	50,8922	53,6719
35	18,5089	22,4650	24,7966	34,3356	46,0588	49,8018	53,2033	57,3420	60,2746
40	22,1642	26,5093	29,0505	39,3353	51,8050	55,7585	59,3417	63,6908	66,7660
45	25,9012	30,6123	33,3504	44,3351	57,5053	61,6562	65,4101	69,9569	73,1660
50	29,7067	34,7642	37,6886	49,3349	63,1671	67,5048	71,4202	76,1538	79,4898
55	33,5705	38,9581	42,0596	54,3348	68,7962	73,3115	77,3804	82,2920	85,7491
60	37,4848	43,1880	46,4589	59,3347	74,3970	79,0820	83,2977	88,3794	91,9518
70	45,4417	51,7393	55,3289	69,3345	85,5270	90,5313	95,0231	100,4251	104,2148
80	53,5400	60,3915	64,2778	79,3343	96,5782	101,8795	106,6285	112,3288	116,3209
90	61,7540	69,1260	73,2911	89,3342	107,5650	113,1452	118,1359	124,1162	128,2987
100	70,0650	77,9294	82,3581	99,3341	118,4980	124,3421	129,5613	135,8069	140,1697
110	78,4582	86,7916	91,4710	109,3341	129,3852	135,4802	140,9165	147,4143	151,9482
120	86,9233	95,7046	100,6236	119,3340	140,2326	146,5673	152,2113	158,9500	163,6485

Fonte: <http://www.ime.unicamp.br/>