

MINISTÉRIO DE MINAS E ENERGIA
SECRETARIA DE MINAS E METALURGIA

CPRM – Serviço Geológico do Brasil

ESTUDO GEOAMBIENTAL DO ESTADO DO RIO DE JANEIRO

ESTUDO DE CHUVAS INTENSAS NO ESTADO DO RIO DE JANEIRO

*Elisabeth Guelman Davis e
Mauro Cunha Naghettini*

BRASÍLIA, Dezembro 2000

CRÉDITOS DE AUTORIA

Elisabeth Guelman Davis e Mauro Cunha Naghettini (Prof. PHD da Escola de Engenharia da UFMG)

Equipe Técnica:

Adriana Coelho Jacome, Alessandro José da Silva, Agnaldo Francisco Teixeira de Freitas, Gláucio André da Silva, Heloiza Moraes Pinheiro, Ivete Souza de Almeida, Keli Regina R. Pedroza, Lucia Maria Silva, Luciene Magalhães da Fonseca, Magda Cristina Ferreira Pinto, Margareth Marques dos Santos, Marina Vilele Bastos, Monica Rejane Dias, Sonia Aparecida Pessoa, Éber José Andrade Pinto, Ligia Maria N. Araujo, Márcio de Oliveira Cândido, Marcelo Jorge Medeiros, Maria Leticia Rabelo Alves Patrus

Copidescagem/Revisão:
Sueli Cardoso de Araújo e
Sérgio Artur Giaquinto

Coordenação:
Cássio Roberto da Silva
Antonio Ivo de Menezes Medina

Executado pela Companhia de Pesquisa de Recursos Minerais – CPRM
Superintendência Regional de Belo Horizonte e
Divisão de Hidrologia Aplicada – DIHAPI do
Departamento de Hidrologia – DEHID

Coordenação Editorial a cargo da
Divisão de Editoração Geral – DIEDIG
Departamento de Apoio Técnico – DEPAT

Davis, Elisabeth Guelman

D261

Estudo de Chuvas Intensas no Estado do Rio de Janeiro. 2ª ed. revista e ampliada. / Elisabeth Guelman Davis e Mauro Cunha Naghettini. – Brasília : CPRM, 2000.

1 CD-ROM.

Estudo Geoambiental do Estado do Rio de Janeiro.

Executado pela CPRM – Serviço Geológico do Brasil. Superintendência Regional de Belo Horizonte.

1. Chuvas – Rio de Janeiro. 2. Meteorologia – Rio de Janeiro. 3. Hidrologia. – I. CPRM – Serviço Geológico do Brasil. II. Naghettini, Mauro Cunha. III. Título.

CDD 551.577

PREFÁCIO

Nas últimas décadas temos nos defrontado com profundas modificações tecnológicas, sociais, econômicas e principalmente ambientais. Entretanto, essas mudanças não têm propiciado a melhoria da qualidade de vida da maioria da população mundial. O homem, necessitando de alimento, proteção, qualidade de vida, vem se esmerando em preparar, consumir e transformar os bens retirados da natureza, em sua busca incessante de promoção do bem-estar social.

Esses bens, ao sofrerem vários processos de transformação para consumo da sociedade, impactam o meio ambiente sob diferentes formas. Isso tem provocado o aumento e o adensamento populacional e, conseqüentemente, o excessivo consumo de recursos naturais.

O ato de consumir e apropriar os recursos naturais, sem qualquer preocupação com as conseqüências futuras, tem promovido uma série de problemas globais, não perceptíveis aos nossos olhos, mas que estão danificando a biosfera e a vida humana de uma maneira alarmante, e que podem se tornar irreversíveis.

Ao analisarmos esses problemas, verificamos que não podemos tratá-los isoladamente, mas de forma sistêmica, pois as variáveis do meio físico, biótico, social, econômico e cultural estão interligadas e são interdependentes, sendo fundamental o entendimento de suas relações e conexões. Entendemos que o enfrentamento e a busca de soluções passam necessariamente pela integração dos profissionais das diversas áreas do conhecimento com visões e atitudes transdisciplinares.

Nesse sentido, a cooperação e a parceria proporcionam a dinâmica para mudanças de comportamento e para o fortalecimento da consciência ao adequado manejo do meio ambiente, pois quando as mesmas se processam, os parceiros (governo, ong, empresa privada, sociedade civil) passam a entender melhor as necessidades dos outros, aprendem, mudam e, como resultado dessa interação, coevoluem.

Na medida que adentramos no novo milênio, a sobrevivência da humanidade dependerá do nosso comportamento em relação à utilização dos recursos naturais. Assim, faz-se necessário adotar os princípios básicos da ecologia: interdependência, reciclagem, parceria, flexibilidade, diversidade e, em decorrência, sustentabilidade.

De forma inédita, o Serviço Geológico do Brasil – CPRM, associando-se a importantes parceiros, em especial a Empresa Brasileira de Agropecuária – EMBRAPA e o Departamento de Recursos Minerais – DRM/RJ, efetivou estudos multidisciplinares, principalmente do meio físico, visando fornecer informações e conhecimentos que possibilitem aos responsáveis pela gestão territorial tomarem decisões para proporcionar o desenvolvimento sustentável do Estado do Rio de Janeiro.

É com grata satisfação que apresentamos à sociedade brasileira, particularmente à comunidade fluminense, o Projeto Rio de Janeiro, que trata de estudos relacionados à geologia, geomorfologia, pedologia, geofísica, geoquímica ambiental, hidrologia, hidrogeologia, recursos minerais, economia mineral, inventário de escorregamentos e diagnóstico geoambiental, dentro de uma abordagem sistêmica.

Esse projeto constitui importante fonte de informações de interesse para múltiplos usuários (mineração, energia, agricultura, saúde pública, urbanismo, saneamento básico, moradia, defesa civil, transportes, turismo e meio ambiente) e instrumento de grande utilidade para subsidiar a gestão ambiental e, principalmente, para o estabelecimento de macrodiretrizes de planejamento com base nas potencialidades e limitações naturais do território, podendo, assim, orientar as políticas de desenvolvimento, levando em consideração a capacidade de suporte de cada região.

Esperamos que este exemplo prospere e que se torne motivo para avançarmos cada vez mais na busca da melhor convivência, possível, entre a exploração dos recursos naturais e a preservação do meio ambiente.

Se pretendemos deixar um mundo melhor para nossos descendentes, temos que olhar o mundo como um sistema vivo, onde tudo ao nosso redor tem o seu papel e sua relativa importância nas complexas relações e conexões. Portanto, enfatizamos a necessidade de mudança comportamental do homem, adotando uma nova forma de pensar e novos valores. Devemos dar vazão aos valores da consciência ecológica, para que seja mantido o adequado equilíbrio com o desenvolvimento econômico. Acreditamos ser esse o procedimento compatível com uma sociedade que deseja estabelecer o desenvolvimento sustentável para o nosso planeta.

Thales de Queiroz Sampaio
Diretor de Hidrologia e Gestão Territorial

APRESENTAÇÃO

O Projeto Rio de Janeiro consiste de estudos multitemáticos do meio físico realizados através do Programa Informações para Gestão Territorial – GATE, da Diretoria de Hidrologia e Gestão Territorial do Serviço Geológico do Brasil – CPRM, na escala 1:250.000, em todo o Estado do Rio de Janeiro, abrangendo uma área de 44.000 km².

O objetivo principal é fornecer subsídios técnicos às administrações estadual e municipais e às entidades privadas, para o planejamento do desenvolvimento sustentado do território fluminense, assim como para o Programa Brasileiro de Zoneamento Ecológico-Econômico, em atendimento à Agenda 21.

O Projeto foi desenvolvido em parceria com a EMBRAPA – Empresa Brasileira de Pesquisa Agropecuária, DRM-RJ – Departamento de Recursos Minerais, CIDE – Centro de Informações e Dados do Rio de Janeiro, SERLA – Superintendência Estadual de Rios e Lagos, EMOP – Empresa de Obras Públicas do Estado do Rio de Janeiro, PUC/RJ – Pontifícia Universidade Católica, UFF – Universidade Federal Fluminense, UERJ – Universidade do Estado do Rio de Janeiro, UFRRJ – Universidade Federal Rural do Rio de Janeiro, UFRJ – Universidade Federal do Estado do Rio de Janeiro, RESUB – Rede de Geotecnologia em Águas Subterrâneas, ANEEL – Agência Nacional de Energia Elétrica e INMET – Instituto Nacional de Meteorologia.

Os temas executados foram litoestrutura, recursos minerais, economia mineral, geomorfologia, inventário de escorregamentos, uso do solo e cobertura vegetal, solos, aptidão agrícola, geofísica e geoquímica ambiental, estudo de chuvas intensas e caracterização hidrogeológica.

A integração das informações levantadas dos diversos temas estudados, propiciou a espacialização de áreas com potencial natural para determinado desenvolvimento, proteção, conservação e recuperação, concluindo por um diagnóstico das potencialidades e das vulnerabilidades naturais ou induzidas de todo o estado do Rio de Janeiro, representadas, em escala 1:500.000, no Mapa Geoambiental.

Os produtos e dados resultantes dos estudos do Projeto estão disponíveis em CD-ROM e mapas plotados em papel na escala 1:500.000.

AGRADECIMENTOS

Agradecemos à Agência Nacional de Energia Elétrica (ANEEL), ao 6º Distrito de Meteorologia do Instituto Nacional de Meteorologia (DISME/INMET) e à Fundação Superintendência Estadual de Rios e Lagoas (SERLA) pela colaboração na obtenção e liberação dos dados imprescindíveis ao estudo.

SUMÁRIO

1 CARACTERIZAÇÃO DO PROBLEMA	1
2 ASPECTOS FÍSICOS E CLIMATOLÓGICOS.....	3
2.1 Localização Geográfica	3
2.2 Aspectos Socioeconômicos.....	3
2.3 Relevo	3
2.4 Hidrografia	4
2.5 Geologia.....	5
2.6 Vegetação.....	5
2.7 Aspectos Climatológicos.....	6
2.7.1 Fatores Estáticos.....	6
2.7.2 Fatores Dinâmicos	6
2.7.3 Distribuição e Regime das Precipitações Pluviométricas	6
3 REDE PLUVIOGRÁFICA EXISTENTE NO ESTADO DO RIO DE JANEIRO	9
3.1 Rede de Dados Pluviográficos.....	9
4 COLETA E ANÁLISE DOS DADOS	13
4.1 Geral.....	13
4.2 Banco de Dados.....	14
5 METODOLOGIA UTILIZADA	15
5.1 Geral.....	15
5.2 Escolha da Metodologia.....	15
5.3 Séries de Duração Parcial.....	16
5.4 Etapas Para a Análise Regional do Estudo de Chuvas Intensas.....	17

6. RESULTADOS.....	19
6.1 Definição Preliminar das Regiões Homogêneas	19
6.2 Análise Regional das Intensidades de Precipitações.....	20
6.2.1 Montagem das Séries Históricas de Duração Parcial	20
6.2.2 Análise de Consistência dos Dados	21
6.2.3 Verificação do Grau de Heterogeneidade Regional	21
6.2.4 Estimativa dos Parâmetros das Distribuições Regionais	23
7 EQUAÇÕES IDF REGIONAIS	41
7.1 Forma Básica	41
7.2 Regressão Múltipla	41
7.3 Equações IDF Regionais Finais	42
8 CONCLUSÃO	45
9 REFERÊNCIAS BIBLIOGRÁFICAS	47
ANEXO I	
• Análise Regional de Frequência de Variáveis Hidrológicas e Hidrometeorológicas: Síntese da Metodologia dos Momentos-L	
ANEXO II	
• Análise Comparativa dos Resultados de Outros Estudos de Chuvas Intensas	
ANEXO III	
• Exemplos Práticos de Aplicação das Equações IDF Regionais	

1

CARACTERIZAÇÃO DO PROBLEMA

A precipitação é uma das variáveis meteorológicas mais importantes para os estudos climáticos de qualquer região. Tal importância deve-se às conseqüências que o excesso de precipitação pode ocasionar, principalmente em eventos de chuvas intensas.

Chuvas intensas são aquelas que registram um grande volume de água precipitado em um curto espaço de tempo. Essas chuvas, freqüentemente, causam prejuízos materiais e humanos consideráveis.

Para evitar esses prejuízos, o dimensionamento de pequenas estruturas hidráulicas (galerias pluviais, bueiros, sistemas de drenagem, bacias de retenção) e o planejamento da ocupação das planícies de inundação precisam levar em consideração os riscos associados à ocorrência de eventos de chuvas intensas. Esses riscos são estimados a partir da análise estatística de séries históricas de dados pluviográficos, através dos quais se pode estimar a intensidade e duração dos eventos.

O presente estudo foi motivado pela ocorrência recente de grandes prejuízos decorrentes de chuvas intensas no estado do Rio de Janeiro e pela constatação de que o último estudo estatístico de âmbito estadual foi realizado na década de 50. Além disso, a qualidade do estudo anterior foi limitada pelas metodologias de inferência estatística da época e pela inexistência de um número expressivo de estações pluviográficas. Para este estudo estão disponíveis dados de 44 estações pluviográficas, distribuídas por todo o território estadual. Além desse expressivo volume de dados, as metodologias contemporâneas de análise regional de freqüência, baseadas em momentos-L, permitem a obtenção de estimativas mais robustas e confiáveis de parâmetros e quantis associados às distribuições de probabilidades das intensidades de eventos chuvosos extremos. A partir da definição de regiões homogêneas no estado do Rio de Janeiro, são estimadas as curvas intensidade-duração-freqüência (IDF) para cada região, com as quais se pode definir as variáveis características de projeto para o dimensionamento de estruturas hidráulicas e sistemas de drenagem.

2

ASPECTOS FÍSICOS E CLIMATOLÓGICOS

2.1 Localização Geográfica

Situado na Região Sudeste do Brasil, o Estado do Rio de Janeiro possui uma área de aproximadamente 43.900km². Faz fronteira ao norte e noroeste com o estado de Minas Gerais, a nordeste com o estado do Espírito Santo e a sudoeste com o estado de São Paulo. Suas costas leste e sul são banhadas pelo Oceano Atlântico. O estado do Rio de Janeiro encontra-se compreendido entre as latitudes 20° 45' 56" S, 23° 22' 08" S e longitudes 40° 57' 23" W, 44° 53' 19" W.

2.2 Aspectos Socioeconômicos

O estado do Rio de Janeiro possui 13.064.296 habitantes, distribuídos entre 81 municípios. A população urbana representa 95,25% desse total, enquanto no meio rural vivem os 4,75% restantes. A cidade do Rio de Janeiro, capital do estado, é a mais populosa, com 5.473.909 habitantes.

O estado do Rio de Janeiro disputa com Minas Gerais a condição de segunda maior economia entre os estados brasileiros. Possui grande potencial turístico, além de relevante parque industrial. É ainda onde se concentram as maiores jazidas de petróleo do país, localizadas na plataforma continental a nordeste, no município de Campos. Destacam-se no estado as indústrias metalúrgica, siderúrgica, química, alimentícia, mecânica, editorial e gráfica, de papel e celulose, de extração mineral, de derivados de petróleo e naval.

2.3 Relevo

O estado do Rio de Janeiro apresenta um relevo bastante acidentado. No sentido sudoeste-nordeste, de maneira aproximada, sucedem-se em faixas estreitas os três componentes principais do quadro morfológico do estado: as baixadas, os maciços litorâneos e o planalto.

As baixadas litorâneas são planas, apenas entalhadas pelos rios, e mais para o interior dão lugar a morros argilosos, onde as rochas do cristalino se acham profundamente decompostas.

Os maciços litorâneos estendem-se desde Cabo Frio até a porção oriental da Baía de Guanabara em meio à Baixada Fluminense, formando um alinhamento com 200 a 500m de altitude.

O planalto ocupa a maior parte do território e seu rebordo oriental é formado pela Serra do Mar, que atravessa o estado do Rio de Janeiro acompanhando o seu litoral. Inicia-se ao norte do estado de Santa Catarina e se estende por mais de 1.000km até o norte do estado do Rio de Janeiro. Ao longo desse percurso, recebe diferentes denominações, como Serra da Bocaina, ao sul do estado, Serra da Estrela e serra dos Órgãos, ao fundo da baía de Guanabara. Na Serra dos Órgãos, as elevações chegam a mais de 1.000m. O planalto decai suavemente para o interior até o vale do rio Paraíba do Sul. Para além desse vale e seguindo aproximadamente a divisa com Minas Gerais, encontra-se a Serra da Mantiqueira, que apresenta as maiores altitudes do estado.

2.4 Hidrografia

O estado do Rio de Janeiro apresenta uma rede hidrográfica particularmente rica, composta de vários rios navegáveis, responsáveis pela conquista e ocupação do interior.

A hidrografia fluminense está dividida em dez unidades, de acordo com a deliberação nº 804 da Comissão Estadual de Controle Ambiental (CECA), de 20.02.1986:

Região Hidrográfica 0 Rio Itabapoana	Abrange os municípios das regiões noroeste e norte fluminense, tendo o Itabapoana como rio principal.
Região Hidrográfica 1 Bacias contribuintes à lagoa Feia	Abrange os corpos d'água compreendidos entre as desembocaduras dos rios Paraíba do Sul (margem direita) e Macaé (margem esquerda).
Região Hidrográfica 2 Bacias contribuintes aos rios Macaé, São João e Una	Compreende as bacias dos rios Macaé e São João, oriundos das encostas da Serra do Mar, e do rio Una, das colinas cristalinas.
Regiões Hidrográficas 3 e 4 Complexos lagunares de Araruama, Saquarema, Maricá, Piratininga/Itaipu	Compreendem as lagunas e tributários localizados entre Cabo Frio e Niterói.
Região Hidrográfica 5 Bacias contribuintes à Baía de Guanabara	Compreende as bacias dos rios que nascem nas encostas da Serra do Mar, nas colinas e nos maciços costeiros e que deságuam na baía de Guanabara. Abrange quase toda a área metropolitana, além de Cachoeiras de Macacu, Rio Bonito e Petrópolis.
Região Hidrográfica 6 Bacia do rio Paraíba do Sul	Compreende a maior bacia hidrográfica do estado, drenando várias regiões (médio Paraíba, Centro-Sul fluminense, Serrana, Noroeste e Norte Fluminense). O rio Paraíba do Sul nasce em São Paulo e se estende do município de Resende até o de São Fidélis, segue a Baixada dos Goytacazes, desembocando no oceano Atlântico. Desempenha importante papel na produção de energia hidrelétrica do estado do Rio de Janeiro, além de abastecer os vários municípios que atravessa.
Região Hidrográfica 7 Contribuinte ao sistema de lagunas de Jacarepaguá	Compreende as lagunas Rodrigo de Freitas, Marapendi, Jacarepaguá, Mearim, Tijuca e rios do sul do município do Rio de Janeiro, desde o Canal do Leblon (Jardim de Alá) até o extremo-oeste da Baixada de Jacarepaguá.

Região Hidrográfica 8 Contribuinte à baía de Sepetiba	Compreende os rios que deságuam na Baía de Sepetiba, desde Barra de Guaratiba (no município do Rio de Janeiro) até o limite entre Mangaratiba e Angra dos Reis.
Região Hidrográfica 9 Contribuinte à baía da ilha Grande	Compreende os rios que nascem na encosta da serra do Mar, próximos ao litoral, e deságuam na baía da ilha Grande, nos municípios de Angra dos Reis e Parati.

2.5 Geologia

O estado do Rio de Janeiro pode ser compartimentado em dois domínios morfoestruturais: o Cinturão Orogênico do Atlântico e as Bacias Sedimentares Cenozóicas.

O Cinturão Orogênico do Atlântico pode ser subdividido nos seguintes domínios morfoestruturais: Maciços Costeiros e Interiores, Maciços Alcalinos Intrusivos, Superfícies Aplainadas nas Baixadas Litorâneas, Escarpas Serranas, Planaltos Residuais, Depressões Interplanálticas, Depressões Interplanálticas com Alinhamentos Serranos.

O soerguimento de blocos decorrentes da tectônica sustentou algumas zonas planálticas, remanescentes de antigas superfícies de aplainamento não-dissecadas no Cenozóico Superior, tais como os planaltos da Bocaina e do Varre-Sai, mas produziu, em contrapartida, depressões interplanálticas, tais como o médio vale do rio Paraíba do Sul.

Junto à zona das baixadas litorâneas sobressaem terrenos colinosos de baixa amplitude topográfica na Região dos Lagos e entre Macaé e Campos, que se caracterizam por superfícies de aplainamento que foram modeladas até o Terciário Superior.

As Bacias Sedimentares Cenozóicas podem ser subdivididas nos seguintes domínios morfoestruturais: Tabuleiros de Bacias Sedimentares, Planícies Fluvio-marinhas (Baixadas) e Planícies Costeiras.

Essas rochas sedimentares estão armazenadas em bacias tafrogênicas continentais resultantes da tectônica extensional gerada no início do Cenozóico. Nesse contexto, foram formadas as bacias de Itaboraí e Resende. Os Tabuleiros da Formação Barreiras, presente no norte do estado, também são tentativamente correlacionados ao mesmo período de sedimentação das bacias continentais. Por fim, os sedimentos inconsolidados das baixadas costeiras foram gerados ao longo dos ciclos transgressivos e regressivos da linha de costa durante o Quaternário, que promoveram o afogamento generalizado do relevo junto à atual linha de costa, registrado nas atuais rias, baías e lagunas nas colinas e morros isolados no recôncavo das baixadas.

2.6 Vegetação

A vegetação do estado do Rio de Janeiro apresenta grande diversidade devida à configuração de seu relevo e de fatores climáticos. Sendo assim, encontra-se distribuída pelo território fluminense, conforme se segue.

Região da floresta ombrófila densa	Vegetação característica das regiões altas da serra do Mar, cujos representantes são os Parques Nacionais, da Serra dos Órgãos e da Bocaina e o Parque Estadual da Ilha Grande.
Região da floresta ombrófila mista	Compõe a floresta pluvial subtropical. Possui representantes da floresta tropical amazônica e da flora australásica, caracterizados pelos planaltos da Bocaina e do Itatiaia (em altitudes de mais de 800m).
Região da floresta estacional semidecidual	Área com mais de 60 dias secos. Cobre a região do médio Paraíba até a do Norte Fluminense.
Região da savana (cerrado)	Restringe-se a uma pequena área do município de Resende, hoje substituída por pastagens.

Região da estepe (caatinga)	Vegetação constituída por plantas cactáceas e espinhosas. Localiza-se nos maciços costeiros próximos a Cabo Frio.
Formações pioneiras	Referem-se às vegetações concentradas em áreas cujos solos estão submetidos às influências marinha, fluvial e fluviomarina. Seus principais representantes são as restingas, os mangues e demais ambientes com acúmulo d'água e sujeitos a inundações periódicas.

No território fluminense, as restingas ocupam trechos do litoral desde a Marambaia, a oeste do município do Rio de Janeiro, até o norte do estado. Já os manguezais se situam no entorno das baías de Guanabara e de Sepetiba.

2.7 Aspectos Climatológicos

De acordo com Nimer (1989, *apud* Goulart, 1999), os fatores responsáveis pelo comportamento pluviométrico da Região Sudeste são: estáticos (localização geográfica e topografia) e dinâmicos (massas de ar).

2.7.1 Fatores Estáticos

A posição na borda oriental do continente sul-americano expõe a Região Sudeste ao fluxo meridional de ar frio oriundo do Pólo Sul sobre as águas quentes do oceano. Assegura-lhe ainda maior frequência de invasões de frentes frias e de linhas de instabilidade tropicais, uma vez que ela está sob a trajetória preferida por tais correntes.

Sua posição marítima confere às camadas de ar que lhe estão superpostas maior densidade de núcleos de condensação.

2.7.2 Fatores Dinâmicos

Do ponto de vista da circulação normal, a Região Sudeste do Brasil permanece, a maior parte do ano, sob o domínio da massa Tropical Atlântica.

Essa massa, de divergência anticiclônica, possui elevada temperatura, fornecida pela intensa radiação solar das latitudes tropicais, e forte umidade específica, fornecida pela intensa evaporação marítima. Em condições normais, essa massa de ar atua na região com ventos geralmente de NE na superfície do solo, com inversão térmica superior, sobre a qual sopram ventos descendentes, também de NE, ambos se dirigindo para as baixas subpolares. Nessas condições, o tempo permanece estável e com pouca nebulosidade. No entanto, essa circulação zonal é freqüentemente perturbada pela circulação meridiana dos anticiclones móveis de origem polar.

2.7.3 Distribuição e Regime das Precipitações Pluviométricas

Determinada pelos fatores dinâmicos e estáticos, nenhuma região brasileira apresenta uma distribuição espacial de precipitações pluviométricas tão diferenciada quanto a Região Sudeste. O volume da precipitação anual em cada local exprime muito bem essa característica.

Dois fatores estáticos devem ser destacados: a posição latitudinal e a localização na borda do oceano. Essa posição confere ao estado do Rio de Janeiro forte radiação solar.

A posição tropical e a proximidade da superfície oceânica, com o conseqüente processo de radiação e evaporação intensas, não explicam por si só a elevada pluviosidade do Sudeste brasileiro e, em particular, do estado do Rio de Janeiro. Esses fatores criam apenas pré-condições à precipitação que, associados aos fatores dinâmicos, determinam os elevados índices pluviométricos da região.

A localização do estado do Rio de Janeiro, na borda oriental da América do Sul, na zona tropical, sob a trajetória preferida das correntes perturbadas, assegura-lhe boa frequência de chuvas ao longo do ano. Porém, sua distribuição espacial é determinada por dois fatores: a orografia e o mecanismo dinâmico.

O paralelismo das escarpas das serras do Mar e da Mantiqueira, opondo-se frontalmente à direção dos ventos das correntes de circulação atmosférica perturbada, representada sobretudo pelas descontinuidades polares, exerce uma sensível influência nesse caso: as precipitações pluviométricas crescem na proporção direta da altitude.

As serras do Mar e da Mantiqueira são bem mais pluviosas que o litoral e o vale do Paraíba do Sul. Nenhuma outra serra do território nacional exerce tanta influência no sentido do acréscimo de precipitações quanto elas.

Nas situações de chuvas generalizadas por toda a região, quase sempre os índices mais elevados se dão nas referidas serras. Muitas vezes, enquanto na baixada Litorânea e no vale do Paraíba do Sul as chuvas são insignificantes, e até mesmo inexistentes, em largos trechos dessas serras se verificam intensos aguaceiros.

É notável o papel orográfico na pluviosidade dessa região. Acrescente-se que em virtude de o maior número de precipitações estar ligado às instabilidades frontais e pós-frontais, a Serra do Mar – por ser, na maioria das vezes, a primeira a ser atingida pelas correntes perturbadas de origem subpolar – é mais pluviosa que a Serra da Mantiqueira. No entanto, chuvas de frentes semi-estacionárias ou quentes podem levar à Mantiqueira precipitações mais copiosas que na Serra do Mar, bem como as de frentes frias de orientação quase N/S. Enquanto os níveis mais elevados da Mantiqueira recebem em média entre 2.000 a 2.500mm de chuva durante o ano, a Serra do Mar acolhe de 2.000 a 4.500mm.

Chama-se a atenção ainda para a diferença entre os índices pluviométricos de um lado e de outro das referidas serras. As encostas meridionais, na maioria das vezes a barlavento, são bem mais chuvosas que as encostas a sotavento. Na encosta a barlavento da Serra do Mar, por exemplo, as isoietas de valores mais elevados que passam na crista da serra descem aos níveis médios dessa encosta, enquanto que a sotavento as precipitações diminuem bruscamente da crista para o vale do Paraíba do Sul.

Em violento contraste, aparece o Vale do Paraíba. A subtração de umidade do ar realizada pelas serras da Mantiqueira e do Mar e a dessecação adiabática tornam essa depressão topográfica bem menos úmida e chuvosa. Aí, a pluviometria anual varia de 750 a 1.000mm.

Outra área igualmente menos chuvosa é a Baixada Litorânea. Seu trecho do Espírito Santo à lagoa de Araruama no estado do Rio de Janeiro apresenta as mais baixas alturas de precipitação, inferiores a 1.000mm.

Outro fator igualmente importante na repartição da pluviosidade nessa área é constituído pelo mecanismo atmosférico regional. O mapa das isoietas das precipitações médias anuais apresentado na figura 1 ilustra a distribuição da chuva ao longo do estado do Rio de Janeiro.

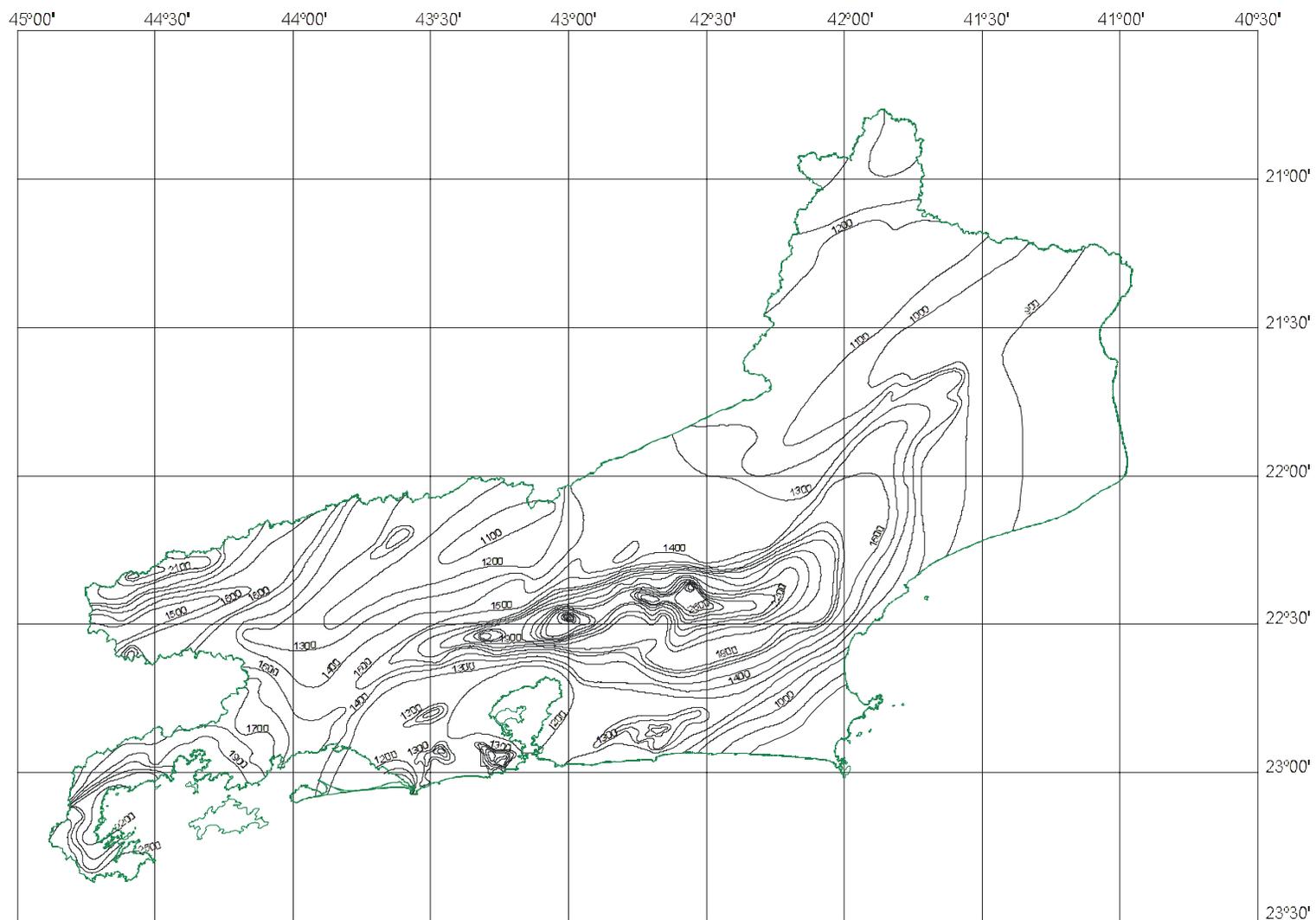


Figura 1 – Mapa de isoietas das precipitações médias anuais.

3

REDE PLUVIOGRÁFICA EXISTENTE NO ESTADO DO RIO DE JANEIRO

3.1 Rede de Dados Pluviográficos

Conforme o *Inventário das Estações Pluviométricas* (ANEEL, 1996), a rede pluviográfica é composta de 80 estações em operação e 43 estações desativadas. A tabela 1 mostra as entidades operadoras e o número de estações com período de observação maior e menor que dez anos, em atividade e desativadas.

Tabela 1 – Rede Pluviográfica.

Entidades	Estações (séries históricas maiores que 10 anos)		Estações (séries históricas menores que 10 anos)		Estações Total
	Ativadas	Desativadas	Ativadas	Desativadas	
ANEEL	16	0	0	0	16
SERLA	39	16	2	12	69
INMET	14	1	0	0	15
Outros	5	5	4	9	23
Total	74	22	6	21	123

Na elaboração dos estudos de chuvas intensas, foram consultados os registros pluviográficos disponíveis no arquivo das entidades operadoras: Agência Nacional de Energia Elétrica (ANEEL), Fundação Superintendência Estadual de Rios e Lagoas (SERLA) e 6º Distrito de Meteorologia do Instituto Nacional de Meteorologia (DISME/INMET). Selecionaram-se apenas as estações com período de dados superior a dez anos.

No total, foram coletados dados de 44 estações, cuja distribuição geográfica nos limites do estado é apresentada no mapa da figura 2. As estações pertencem às seguintes entidades:

- ANEEL: registros de oito estações, que se encontram distribuídas na região centro-sul do estado, concentrando-se próximo à divisa com o estado de Minas Gerais.
- SERLA: registros de 14 estações, as quais se encontram distribuídas aproximadamente entre a região metropolitana da cidade do Rio de Janeiro e a Serra dos Órgãos.
- INMET: registros de 22 estações, que se encontram espacialmente bem distribuídas pelo território do estado do Rio de Janeiro.

A figura 3 apresenta os diagramas de barras das estações disponíveis por entidade, com os períodos de dados disponíveis, localização geográfica e altitude.

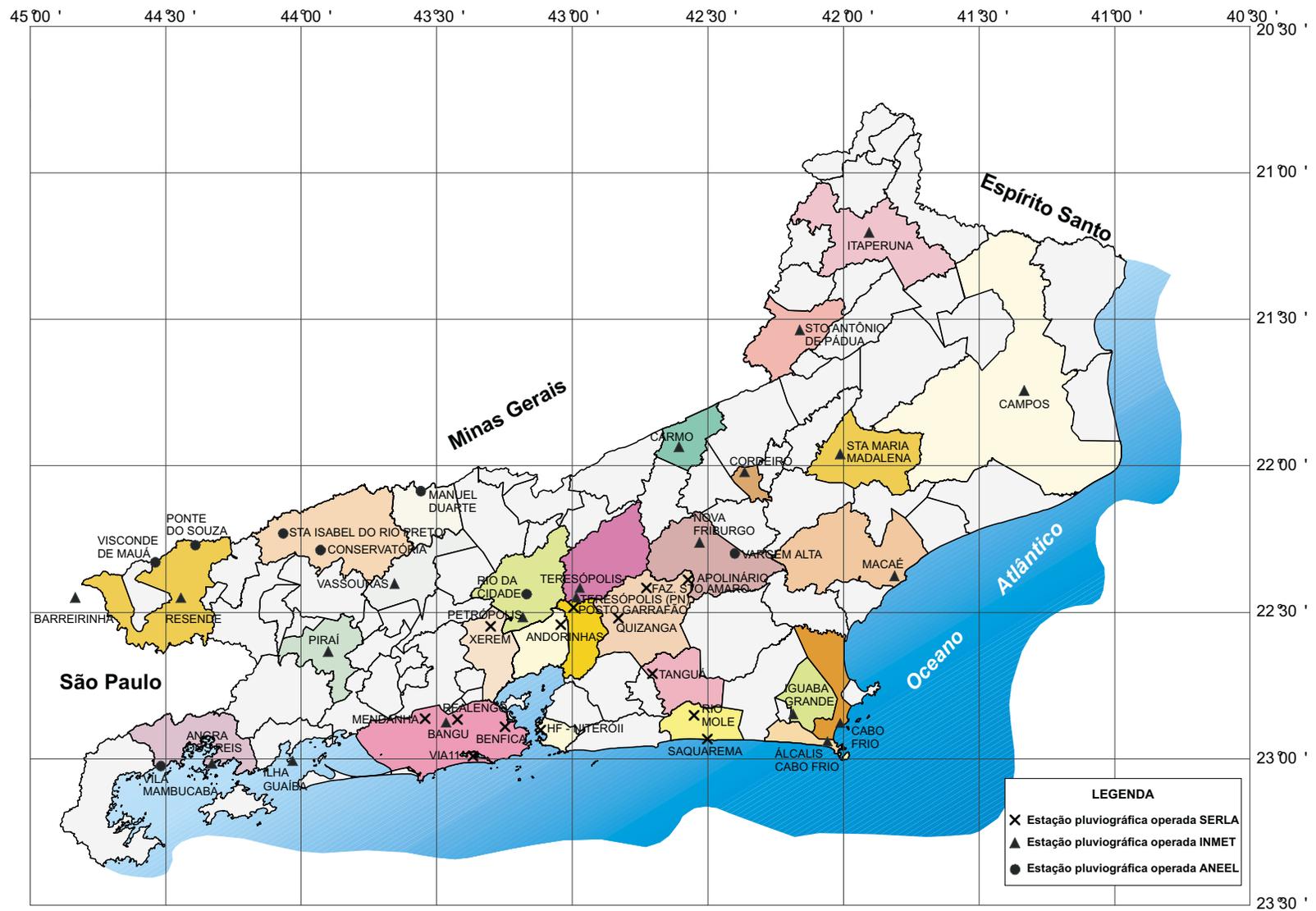


Figura 2 – Localização das estações pluviográficas.

4

COLETA E ANÁLISE DOS DADOS

4.1 Geral

Os dados pluviográficos levantados, no geral, estão registrados em pluviogramas, cujos comprimento e altura correspondem a 24 horas de observação e a 10mm de precipitação, respectivamente. Na graduação horizontal, os intervalos de tempo são de dez minutos e, na vertical, de 0,2mm.

Na análise dos pluviogramas, fez-se a seleção das chuvas consideradas intensas, ou seja, aquelas cuja intensidade excede um limite mínimo estipulado para cada duração. Foram adotados os limites mínimos propostos por Pfafstetter (1957), complementando com limites propostos por Wilken (1978), referentes a durações não consideradas no estudo de Pfafstetter. A tabela 2 mostra os valores mínimos recomendados pelos autores citados e os limites utilizados no estudo de chuvas intensas do estado do Rio de Janeiro.

Tabela 2 – Limites mínimos para chuvas intensas.

Autor	Pfafstetter (1957)		Wilken (1978)		Valores Adotados (1999)	
	P (mm)	I (mm/h)	P (mm)	I (mm/h)	P (mm)	I (mm/h)
Duração						
5 min	8	96	10	120	8	96
10 min			12	72	12	72
15 min	15	60	15	60	15	60
30 min	20	40	20	40	20	40
45 min			23	30,6	23	30,6
1 h	25	25	25	25	25	25
2 h	30	15	30	15	30	15
3 h			33	11	33	11
4 h	35	8,8	35	8,8	35	8,8
8 h	40	5			40	5
14 h	47	3,4			47	3,4
24 h	55	2,3			55	2,3

A pesquisa dos eventos de chuvas intensas registrados nos pluviogramas foi realizada manualmente. Leram-se todos os eventos que apresentavam intensidades superiores às da tabela, para cada uma das durações. As leituras foram feitas em intervalos de cinco minutos e armazenadas em planilhas de cálculo. Cada evento de chuvas intensas foi gravado em uma planilha separada, armazenando-se as horas de início e de fim do evento, bem como os valores de precipitação a cada cinco minutos, o que possibilita a posterior reconstituição do evento.

Nessa leitura, foram encontradas diversas dificuldades:

- os dados levantados no INMET foram obtidos através de cópias xerox, o que dificultou as leituras e provocou imprecisões;
- os registros pluviométricos diários, que deveriam ser utilizados para efeito comparativo, só foram liberados pelo INMET após a fase de leitura;
- em diversos eventos de chuvas intensas, os pluviógrafos de várias estações apresentaram problemas, tais como: pena emperrada, não registrando o evento completo; leitura confusa, devido à não-substituição diária do papel (chuvas de vários dias registradas no mesmo formulário); registros em papel fora do formato padrão, sem as linhas de escala; pluviogramas manchados de tinta, provavelmente, devido à própria chuva; registros ilegíveis, devido à deficiência de tinta; defasagem temporal em seus registros.

Em várias estações, e durante alguns períodos de observação, o volume de sifonagem do pluviógrafo não estava bem calibrado e o traço da pena não coincidia com as indicações 10mm e 0mm. Nesses casos, fizeram-se as devidas correções na leitura da precipitação.

Neste estudo, levantaram-se 8.996 eventos distintos de chuvas intensas, assim considerados aqueles que apresentaram um período mínimo de seis horas sem chuva entre eles. Esse critério acompanha autores como Huff (1967) e Pinheiro (1997).

Com o auxílio de um programa desenvolvido em Visual Basic e executado a partir do Microsoft Excel, para cada evento de chuvas intensas levantado pesquisaram-se os maiores valores de precipitações para durações de 5min, 10min, 15min, 30min, 45min, 1h, 2h, 3h, 4h, 8h, 14h e 24h.

4.2 Banco de Dados

Criou-se um banco de dados, no formato do Microsoft Excel, composto de todos os eventos de chuvas intensas analisados, com durações entre cinco minutos e 24 horas, para todas as estações selecionadas (tabela 3).

Tabela 3 – Amostra do banco de dados.

Estação	Evento	5min	10min	15min	30min	45min	1 h	2 h	3 h	4 h	8 h	14 h	24 h
S.ISABEL	25/12/1981	8,0	12,8	16,8	30,5	32,3	33,6	37,2	39,0	39,3	39,5	39,6	39,6
S.ISABEL	18/12/1981	5,8	11,1	16,4	17,4	17,5	17,5	17,6	17,7	21,8	25,9	29,4	42,0
S.ISABEL	12/12/1981	8,7	17,0	23,0	34,3	39,4	44,2	45,2	45,3	45,3	46,0	46,4	46,4
S.ISABEL	11/12/1981	3,0	5,5	6,4	7,4	7,6	7,6	10,7	11,7	11,9	16,0	21,3	23,3
S.ISABEL	08/12/1981	3,6	5,6	7,6	13,0	17,0	18,8	22,4	24,0	27,7	40,1	47,9	54,4
S.ISABEL	04/12/1981	1,6	3,0	3,8	6,6	8,4	8,9	9,9	11,3	17,0	22,5	28,9	35,4
S.ISABEL	02/12/1981	8,5	15,3	22,7	29,7	30,6	30,7	31,6	32,4	32,7	32,8	33,0	33,0

5

METODOLOGIA UTILIZADA

5.1 Geral

A análise de frequência dos eventos extremos pode ser local, ou seja, utilizando-se a série de dados de uma única estação, ou regional, fazendo uso dos dados de várias estações de uma região.

Neste estudo, adotou-se a análise de frequência regional. A aplicação de métodos de regionalização de funções estatísticas de variáveis hidrológicas consiste basicamente na utilização de vários postos de medição em uma região, transferindo informações espacialmente.

A vantagem em se utilizar a análise de frequência regional é a possibilidade de estimar os parâmetros ou variáveis hidrológicas em locais que não possuam uma coleta sistemática de informações. Além disso, essa técnica permite melhorar a estimativa dos parâmetros e quantis de uma distribuição de probabilidades teórica, identificar a ausência de postos de observação em uma região e verificar a consistência das séries hidrológicas.

Com a regionalização, é possível aumentar consideravelmente o número de informações, que passam a ter caráter regional. De acordo com Pires (1994), essa técnica possibilita reduzir o erro amostral. Além disso, os modelos de regionalização permitem melhorar as estimativas de eventos extremos, substituindo o tempo por espaço.

5.2 Escolha da Metodologia

Tucci (1993) classifica os métodos de regionalização em três classes: métodos de regionalização dos parâmetros da distribuição de probabilidades, métodos de regionalização do evento com um determinado risco e os métodos de regionalização da curva adimensional de frequências, denominados métodos de regionalização *index-flood*.

De acordo com Pinto & Naghettini (1999), independentemente do método de regionalização a ser utilizado, um dos pontos cruciais é a definição de regiões estatisticamente homogêneas, ou seja, aquelas contendo várias estações cujas séries sejam oriundas de populações regidas pela mesma distribuição de probabilidades, com os parâmetros de posição e escala variando entre as estações.

O trabalho elaborado por Wallis (1989) sobre técnicas de regionalização *index-flood*, utilizando momentos-L, representou um avanço nos estudos de regionalização, permitindo reduzir a subjetividade das etapas necessárias ao seu desenvolvimento. Hosking & Wallis (1993; 1995) desenvolveram três estatísticas-L, que são aplicadas nas etapas de consistência de dados, identificação de regiões homogêneas e na definição da distribuição de frequência regional.

Em sua revisão sobre os avanços recentes da pesquisa na área de análise de frequência, Bobée & Rasmussen (1995) consideram a contribuição de Hosking & Wallis como a mais relevante para a obtenção de melhores estimativas das probabilidades de eventos raros.

No presente estudo, optou-se pela utilização da metodologia desenvolvida por Hosking & Wallis para delimitação de regiões homogêneas e regionalização dos eventos de chuvas intensas das estações estudadas, através do método de regionalização *index-flood* com a utilização dos momentos-L e do emprego das estatísticas-L. Detalhes sobre essa metodologia são apresentados no Capítulo 1 do Anexo I.

No método de regionalização *index-flood*, os postos localizados nas regiões estatisticamente homogêneas seguem as mesmas distribuições de frequência, à exceção de um fator dependente de cada posto. Esse fator é por isso chamado *index-flood*.

Os momentos-L são interpretados como as medidas das características de uma distribuição de probabilidades. Essas medidas são análogas aos momentos convencionais, mas são estimadas por combinações lineares dos elementos de uma amostra ordenada. A vantagem dos momentos-L sobre os momentos convencionais é que eles possibilitam a caracterização de um grande número de distribuições, permitindo maior robustez na estimação dos parâmetros. De fato, quando estimados a partir de uma amostra, são mais robustos à presença de *outliers*. Os momentos-L são descritos com maior profundidade no Capítulo 2 do Anexo I).

5.3 Séries de Duração Parcial

A diferença básica entre a série de duração parcial e a série de máximos anuais é que a primeira compreende somente os maiores eventos de chuvas superiores a uma determinada precipitação de referência, e a segunda compreende um único valor para cada ano, o valor máximo anual.

Segundo Wilken (1978), para fins práticos, as séries de duração parcial e as séries anuais não diferem muito, exceto para baixos períodos de retorno. O relatório dos estudos de vazões do NERC (1975) recomenda o uso de séries de valores máximos anuais somente quando se dispõe de mais de 25 anos de observações.

Bertoni & Tucci (1993) afirmam que a metodologia de séries parciais é utilizada quando o número de anos de dados é pequeno (menos de 12 anos) e os tempos de retorno que serão utilizados são inferiores a cinco anos. Wang (1991, *apud* Laura, 1998) mostrou que os modelos de séries parciais também são eficientes para grandes períodos de retorno.

Devido às falhas existentes e aos curtos períodos de observação disponíveis em algumas estações, optou-se pelo uso das séries de duração parcial.

O uso de séries de duração parcial na análise regional de frequência é apresentado com maiores detalhes no Capítulo 7 do Anexo I.

5.4 Etapas para a Análise Regional do Estudo de Chuvas Intensas

As etapas da análise regional de consistência de dados, identificação e delimitação de regiões homogêneas, seleção da distribuição regional de frequência e estimação da distribuição regional de frequência são detalhadas nos capítulos 3 a 6 do Anexo I.

O diagrama da figura 4 sintetiza as diversas etapas da metodologia utilizada para a análise regional de frequências dos eventos de chuvas intensas no estado do Rio de Janeiro.

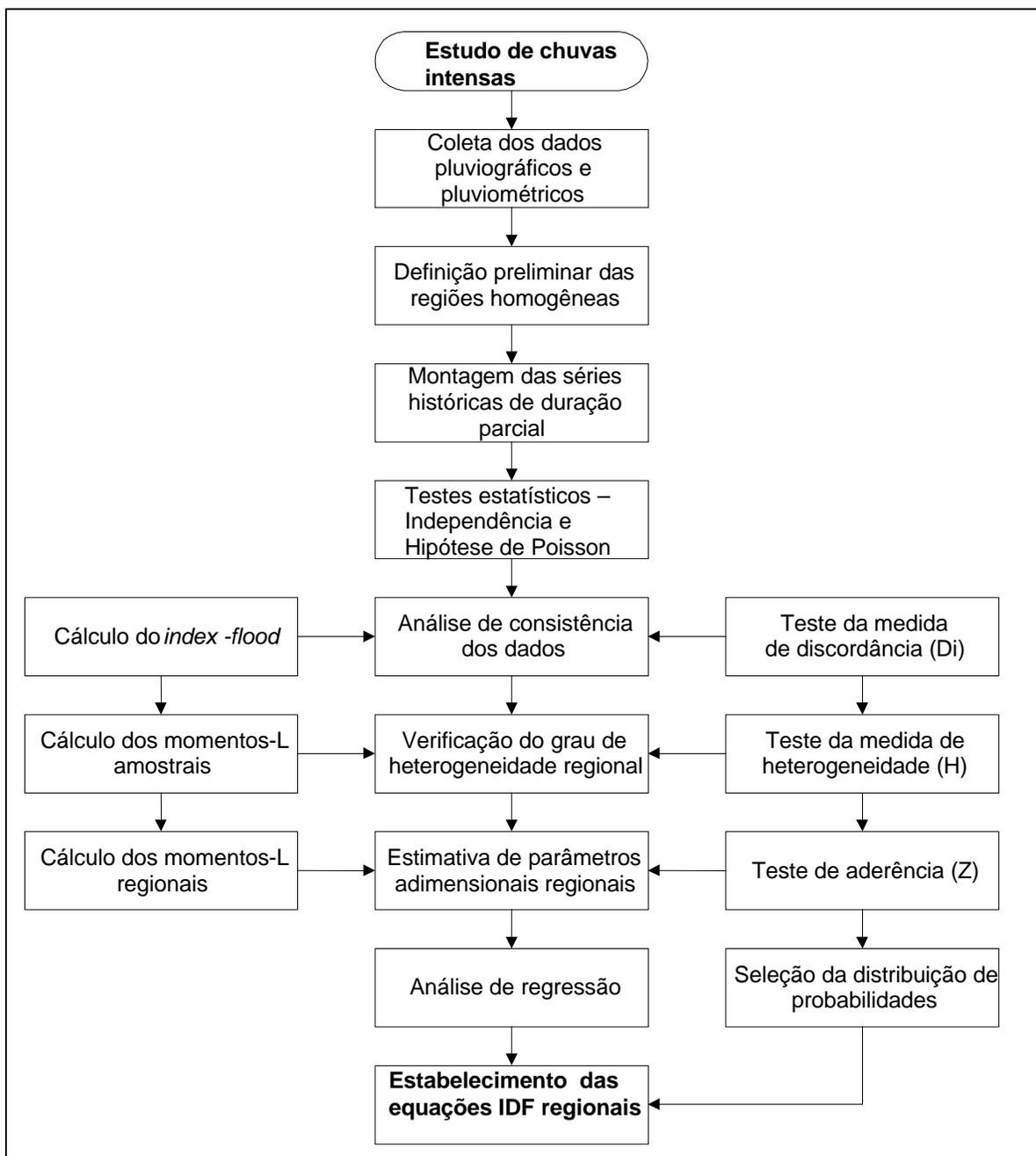


Figura 4 – Síntese metodológica para análise regional de chuvas intensas.

6

RESULTADOS

6.1 Definição Preliminar das Regiões Homogêneas

Na identificação preliminar das regiões homogêneas, foram analisadas as informações fisiográficas, climatológicas e o mapa isoietal de precipitação média anual do estado do Rio de Janeiro. Para determinar a divisão preliminar das estações em regiões homogêneas, utilizou-se a técnica de análise de *clusters* recomendada por Hosking & Wallis (1997) (vide item 4.2 do Anexo I).

Na análise foram utilizadas 50 estações pluviométricas distribuídas espacialmente ao longo do estado do Rio de Janeiro.

De acordo com essa técnica, os postos são agrupados em regiões consonantes com a variabilidade espacial de algumas características locais associadas a um grau de importância ou ponderação. No presente estudo, as características locais usadas na divisão em *clusters* foram:

- precipitação média anual (peso 3);
- latitude e longitude (peso 2);
- altitude (peso 1).

De acordo com o algoritmo *K-means clustering*, foram identificados quatro *clusters*, ou seja, quatro regiões homogêneas. Os limites dos *clusters* foram definidos com base na altimetria e no mapa isoietal de precipitação anual.

Baseadas na análise isoietal e no método de *clustering*, definiram-se preliminarmente quatro regiões homogêneas. As estações pertencentes a cada uma delas são:

- **Região 1:** Álcalis Cabo Frio, Cabo Frio, Campos, Carmo, Iguaba Grande, Itaperuna, Macaé, Ordinária do Carmo, Rio Mole, Santa Maria Madalena, Santo Antônio de Pádua, Saquarema.
- **Região 2:** Andorinhas, Apolinário, Fazenda Santo Amaro, Nova Friburgo, Posto Garrafão, Quizanga, Teresópolis, Teresópolis PN, Petrópolis.

- **Região 3:** Bangu, Benfica, Conservatória, Cordeiro, Horto Florestal, Manuel Duarte, Mendanha, Pirai, Realengo, Rio Cidade, Tanguá, Vargem Alta, Vassouras, Via 11, Xerém.
- **Região 4:** Angra dos Reis, Barreirinha, Ilha Guaíba, Ponte do Souza, Resende, Santa Isabel do Rio Preto, Vila Mambucaba, Visconde de Mauá.

A delimitação final das regiões foi realizada com base nos dados obtidos dos registros pluviográficos, utilizando-se as sub-rotinas desenvolvidas por Hosking & Wallis (1997), que permitem o cálculo da medida de heterogeneidade. Os resultados estão apresentados no próximo item.

6.2 Análise Regional das Intensidades de Precipitações

6.2.1 Montagem das Séries Históricas de Duração Parcial

Na montagem dessas séries, selecionaram-se os maiores eventos de intensidade de chuva independentes, com duração entre cinco minutos e 24 horas. Nessas séries, o número de excedências em cada estação em relação a um valor limiar μ seguiu a distribuição de Poisson com intensidade anual $\Lambda=1$ ou 1,5 ou 2.

Portanto, para cada estação i com n_i anos hidrológicos de registro, o tamanho da série foi λn_i eventos de chuva independentes observados para cada duração.

Para verificar a aderência da distribuição de Poisson ao número de excedências anuais (Estatística de Dispersão γ – ver subitem 7.3.2 do Anexo I), foi desenvolvido um programa em Pascal.

A partir dos resultados obtidos dessa análise, fez-se a montagem das séries dos postos (tabela 4).

Tabela 4 – Tamanho da série parcial.

Região	Λ	$\alpha(\%)$	Tamanho da Série Parcial em cada Estação
1	1	2,50	n
2	2	2,50	2n
3	1,5	2,50	1,5n
4	1	2,50	n

Λ : razão de ocorrência média das excedências anuais; α : nível de significância (teste do χ^2); n: número de anos hidrológicos com dados disponíveis

Nessa primeira análise, as seguintes estações não passaram no teste de aderência à distribuição de Poisson e foram excluídas:

- Ordinária do Carmo (região 1);
- Teresópolis (região 2);
- Realengo, Rio Cidade, Vassouras (região 3).

Cabe ressaltar que em cada região os momentos-L amostrais foram calculados com base nas séries parciais de cada duração.

No presente estudo, o *index-flood* é o valor médio da série parcial utilizada em cada duração. Nas etapas da análise regional, descritas nos próximos subitens, as intensidades foram divididas pelo *index-flood*.

6.2.2 Análise de Consistência dos Dados

Em uma análise preliminar, foram excluídas as seguintes estações:

- Região 1: Santa Maria Madalena (período de dados curto) e Carmo (má qualidade dos pluviogramas);
- Região 4: Barreirinha, Visconde de Mauá e Vila Mambucaba (séries de dados com muitas falhas e má qualidade dos pluviogramas).

Hosking & Wallis (1995) desenvolveram uma estatística baseada em momentos-L, denominada medida de discordância (D_i) da estação i em relação ao padrão regional (vide item 3.1 do Anexo I), sugerindo os valores críticos apresentados na tabela 3.1 (Anexo I). O objetivo é identificar as estações que possuem dados grosseiramente discordantes dos observados na região.

Os resultados obtidos nessa análise foram:

- Na região 1, foram avaliadas dez estações. Conforme Hosking & Wallis (1995), o valor crítico da medida de discordância (D_i) é 2,491. Em todas as estações, o valor calculado de D_i foi inferior ao valor crítico. Apenas a estação de Saquarema (duração cinco minutos) apresentou-se discordante.
- Na região 2, foram avaliadas oito estações. Conforme Hosking & Wallis (1995), o valor crítico da medida de discordância (D_i) é 2,140. Em todas as estações, o valor calculado de D_i foi inferior ao valor crítico. Apenas as estações Petrópolis (durações 8 e 14 horas) e Apolinário (duração 24 horas) apresentaram-se discordantes.
- Na região 3, foram avaliadas 11 estações. Conforme Hosking & Wallis (1995), o valor crítico da medida de discordância (D_i) é 2,632. Em todas as estações, o valor calculado de D_i foi inferior ao valor crítico.
- Na região 4, foram avaliadas cinco estações. Conforme Hosking & Wallis (1995), o valor crítico da medida de discordância (D_i) é 1,333. Em todas as estações, o valor calculado de D_i foi inferior ao valor crítico.

Observação: As estações que se apresentaram discordantes não foram excluídas, porque sua retirada não alterou significativamente a medida de heterogeneidade (ver próximo subitem).

6.2.3 Verificação do Grau de Heterogeneidade Regional

A medida de heterogeneidade (H) tem por objetivo verificar o grau de heterogeneidade de uma região através da comparação da variabilidade da razão dos momentos-L amostrais de uma estação com aquela que seria esperada de uma região homogênea resultante da simulação (vide item 4.3 do Anexo I).

Nessa etapa, a estação Manuel Duarte foi excluída da região 3 e incluída na região 1.

Os valores da medida de heterogeneidade H foram satisfatórios. Essa análise comprovou a hipótese inicial de quatro regiões homogêneas, tal como definidas preliminarmente.

As regiões homogêneas finais e suas respectivas estações pluviográficas são mostradas na figura 5.

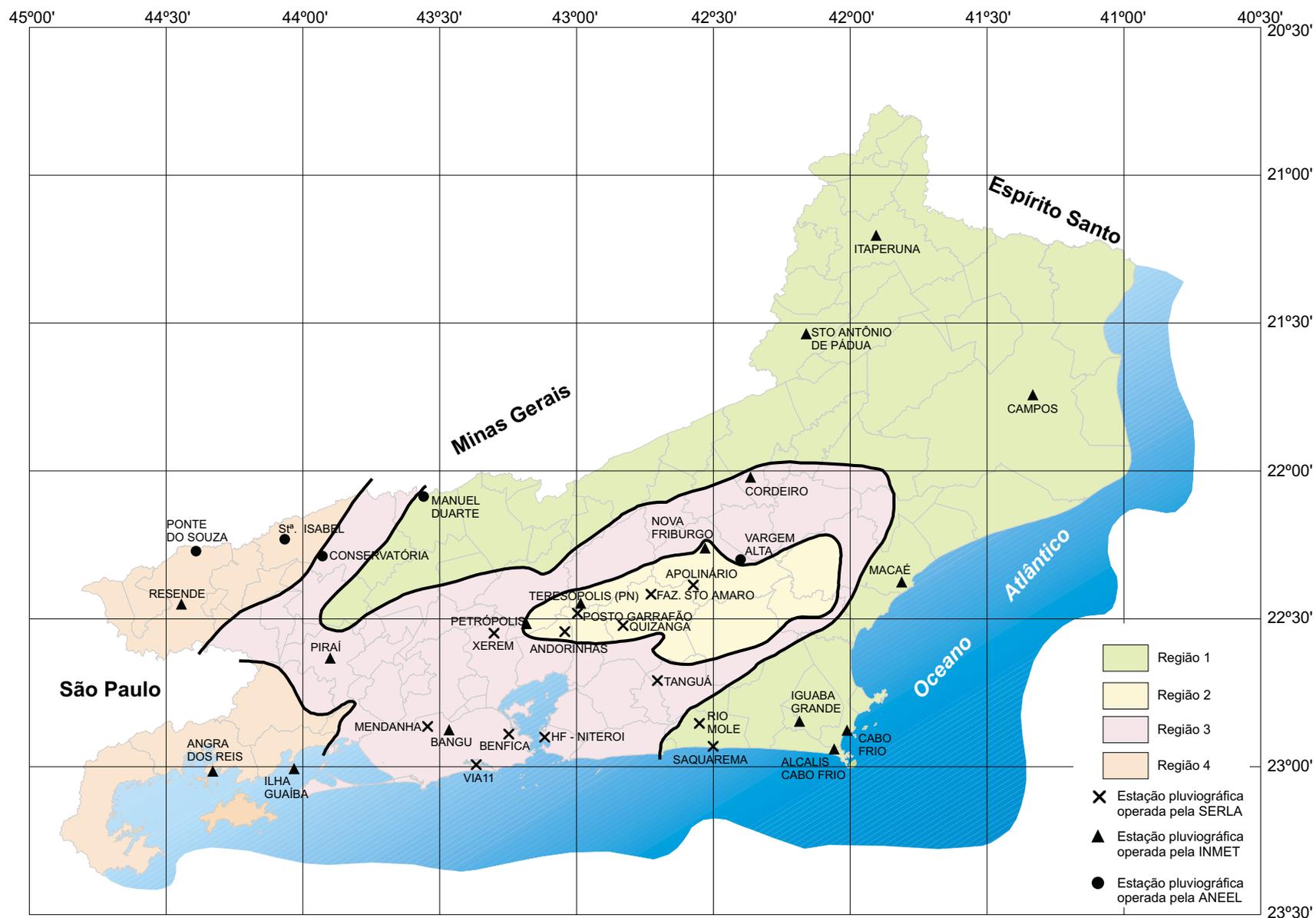


Figura 5 – Mapa das regiões homogêneas.

6.2.4 Estimativa dos Parâmetros das Distribuições Regionais

A seleção das distribuições de frequências regionais foi efetuada com as sub-rotinas descritas por Hosking (1991), implementadas em linguagem Fortran-77. Esse programa faz o ajuste das distribuições Logística Generalizada, Generalizada de Valores Extremos, Generalizada de Pareto, Generalizada Normal e Pearson tipo III, estimando-se os seus parâmetros a partir dos momentos-L regionais, além de aplicar o teste de aderência (Z^{DIST}) para verificar o ajuste entre a distribuição candidata e os dados regionais (vide Capítulo 5 do Anexo I).

No presente estudo, para analisar as distribuições que apresentaram o ajuste adequado em cada região, seguiram-se os critérios do teste de aderência de Hosking & Wallis (1993), ou seja, se $|Z^{\text{DIST}}| \leq 1,64$.

As distribuições de três parâmetros que apresentaram melhor ajuste foram a Logística Generalizada nas regiões 1, 3 e 4 e GEV na região 2.

Os resultados obtidos nessa análise foram:

- Na região 1, os valores de $|Z^{\text{DIST}}|$ calculados para todas as durações foi inferior ao limite. Apenas para a duração de cinco minutos, o valor de $|Z^{\text{DIST}}|$ calculado foi $-3,48$.
- Na região 2, os valores de $|Z^{\text{DIST}}|$ calculados para todas as durações foi inferior ao limite. Apenas para a duração de dez minutos, o valor de $|Z^{\text{DIST}}|$ calculado foi $-3,86$ e para duração de 15 minutos foi de $-2,27$.
- Na região 3, os valores de $|Z^{\text{DIST}}|$ calculados para todas as durações foi inferior ao limite. Sendo que:
 - para a duração de dez minutos, as estações Conservatória e Vargem Alta foram excluídas;
 - para a duração de 45 minutos, a estação de Xerém foi excluída;
 - para a duração de uma hora, a estação Via 11 foi excluída;
 - para as durações de duas horas e três horas, a estação Vargem Alta foi excluída.

Na região 4, os valores de $|Z^{\text{DIST}}|$ calculados para todas as durações foi inferior ao limite. Apenas para a duração de dez minutos, o valor de $|Z^{\text{DIST}}|$ calculado foi $-2,30$ e para duração de 15 minutos foi de $-2,44$.

Os parâmetros das distribuições regionais foram calculados a partir da série parcial adimensional. Os resultados estão apresentados nas tabelas 5 a 8.

As figuras 6 a 9 apresentam alguns exemplos de ajustes das distribuições em cada região homogênea. Na elaboração dos gráficos das referidas figuras, utilizou-se a ordem de plotagem de Gringorten. Para $-\ln(P(X \leq x))$, adotou-se a fórmula $(n + 0,12) / (i - 0,44)$, onde:

n – número de anos de registros;

i – posição dos dados em ordem decrescente.

Tabela 5 – Parâmetros estimados da distribuição Logística (região 1).

Duração	Parâmetros Estimados			
	Posição x	Escala a	Forma k	L
5 min	0,938	0,106	-0,315	1
10 min	0,932	0,118	-0,310	1
15 min	0,926	0,116	-0,337	1
30 min	0,928	0,128	-0,305	1
45 min	0,928	0,130	-0,302	1
1 h	0,924	0,128	-0,319	1
2 h	0,909	0,125	-0,373	1
3 h	0,908	0,124	-0,377	1
4 h	0,911	0,128	-0,362	1
8 h	0,917	0,128	-0,342	1
14 h	0,920	0,130	-0,329	1
24 h	0,922	0,138	-0,307	1

Tabela 6 – Parâmetros estimados da distribuição GEV (região 2).

Duração	Parâmetros Estimados			
	Posição x	Escala a	Forma k	L
5 min	0,912	0,084	-0,326	2
10 min	0,928	0,080	-0,246	2
15 min	0,925	0,117	-0,057	2
30 min	0,911	0,122	-0,133	2
45 min	0,897	0,135	-0,158	2
1 h	0,892	0,139	-0,168	2
2 h	0,891	0,159	-0,100	2
3 h	0,885	0,148	-0,170	2
4 h	0,879	0,150	-0,189	2
8 h	0,867	0,146	-0,254	2
14 h	0,863	0,150	-0,254	2
24 h	0,871	0,154	-0,211	2

Tabela 7 – Parâmetros estimados da distribuição Logística (região 3).

Duração	Parâmetros Estimados			
	Posição x	Escala a	Forma k	L
5 min	0,937	0,069	-0,436	1,5
10 min	0,959	0,078	-0,286	1,5
15 min	0,951	0,087	-0,304	1,5
30 min	0,940	0,096	-0,332	1,5
45 min	0,931	0,107	-0,340	1,5
1 h	0,937	0,104	-0,325	1,5
2 h	0,921	0,107	-0,375	1,5
3 h	0,917	0,108	-0,389	1,5
4 h	0,914	0,108	-0,397	1,5
8 h	0,900	0,112	-0,429	1,5
14 h	0,896	0,116	-0,433	1,5
24 h	0,895	0,125	-0,413	1,5

Tabela 8 – Parâmetros estimados da distribuição Logística (região 4).

Duração	Parâmetros Estimados			
	Posição x	Escala a	Forma k	L
5 min	0,943	0,086	-0,345	1
10 min	0,946	0,076	-0,365	1
15 min	0,951	0,083	-0,318	1
30 min	0,950	0,085	-0,318	1
45 min	0,954	0,097	-0,264	1
1 h	0,947	0,101	-0,290	1
2 h	0,950	0,110	-0,257	1
3 h	0,936	0,120	-0,291	1
4 h	0,930	0,114	-0,327	1
8 h	0,934	0,117	-0,307	1
14 h	0,938	0,111	-0,305	1
24 h	0,939	0,109	-0,304	1

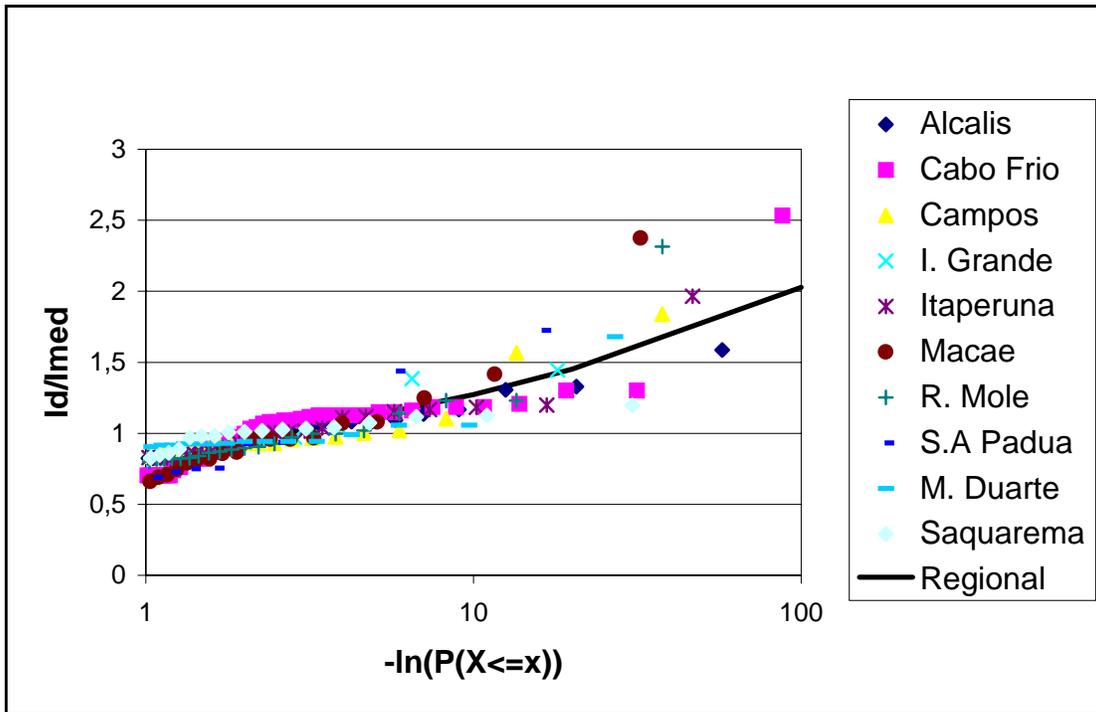


Figura 6A – Região homogênea 1 – duração 5min: ajuste da Distribuição Logística Generalizada para série parcial.

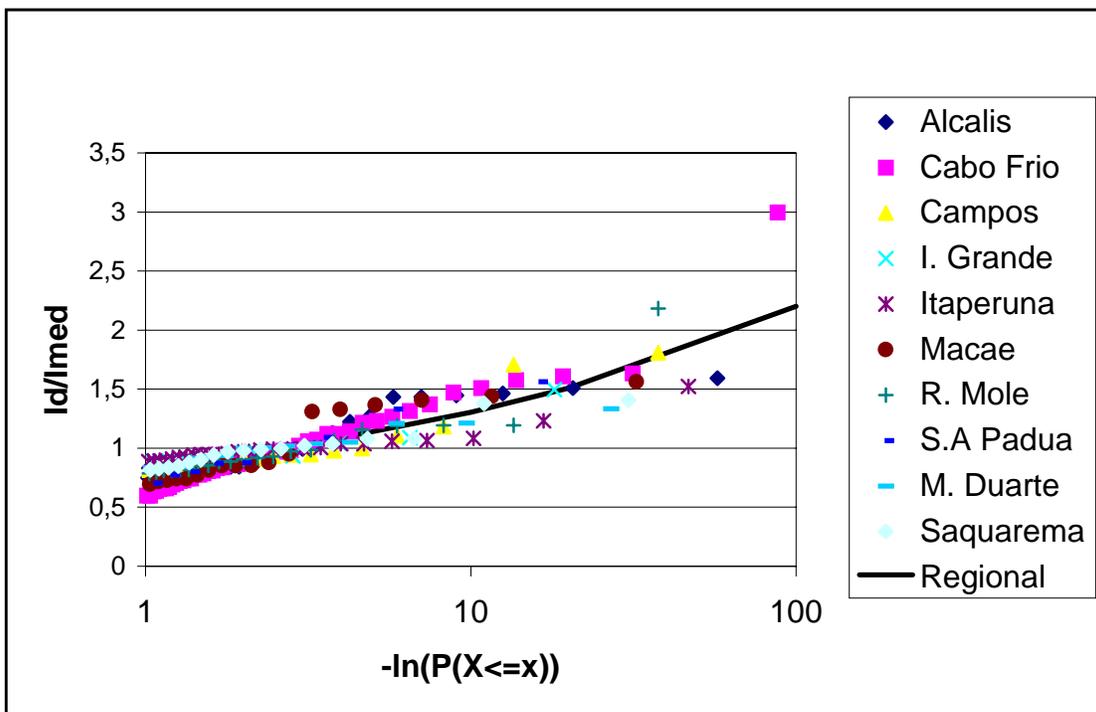


Figura 6B – Região homogênea 1 – duração 15min: ajuste da Distribuição Logística Generalizada para série parcial.

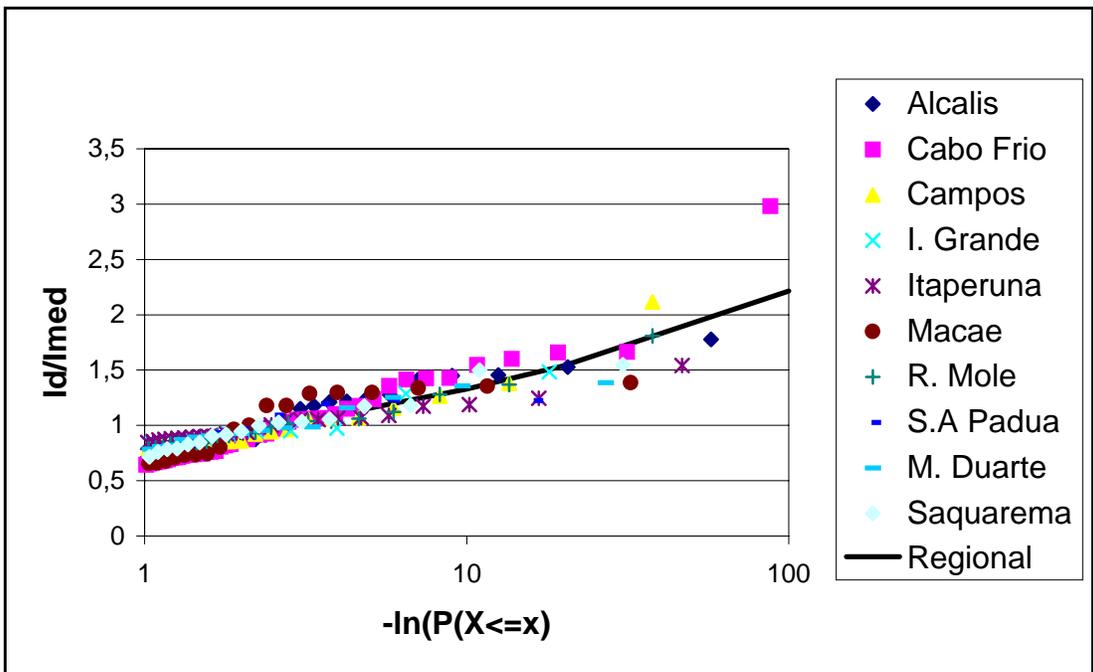


Figura 6C – Região homogênea 1 – duração 30min: ajuste da Distribuição Logística Generalizada para série parcial.

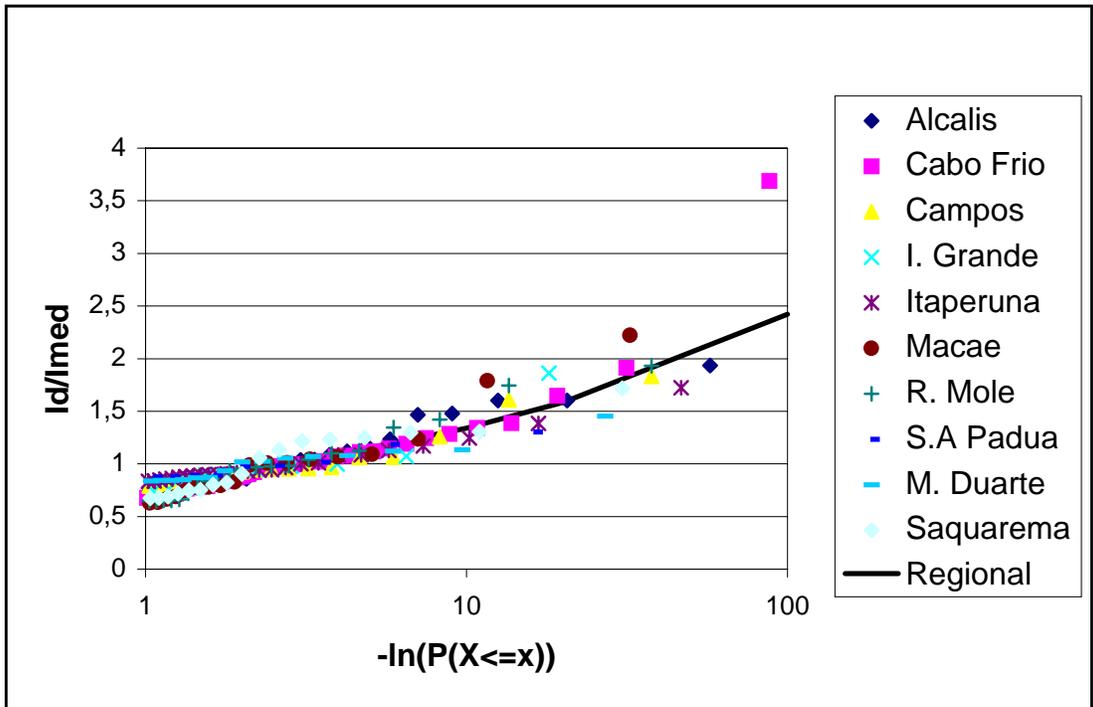


Figura 6D – Região homogênea 1 – duração 4h: ajuste da Distribuição Logística Generalizada para série parcial.

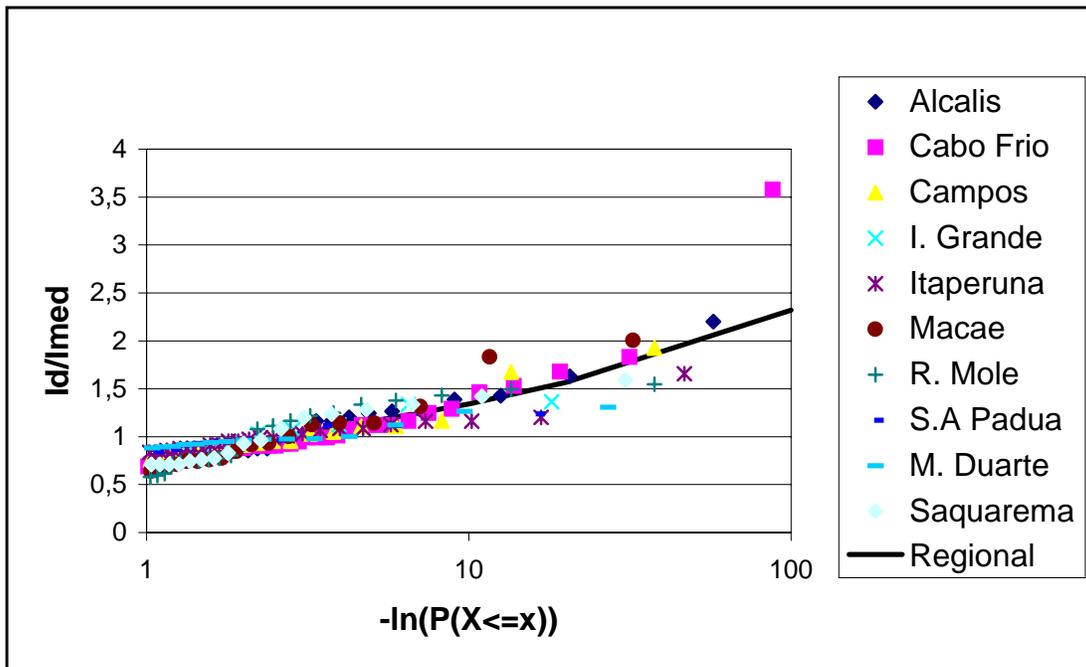


Figura 6E – Região homogênea 1 – duração 14h: ajuste da Distribuição Logística Generalizada para série parcial.

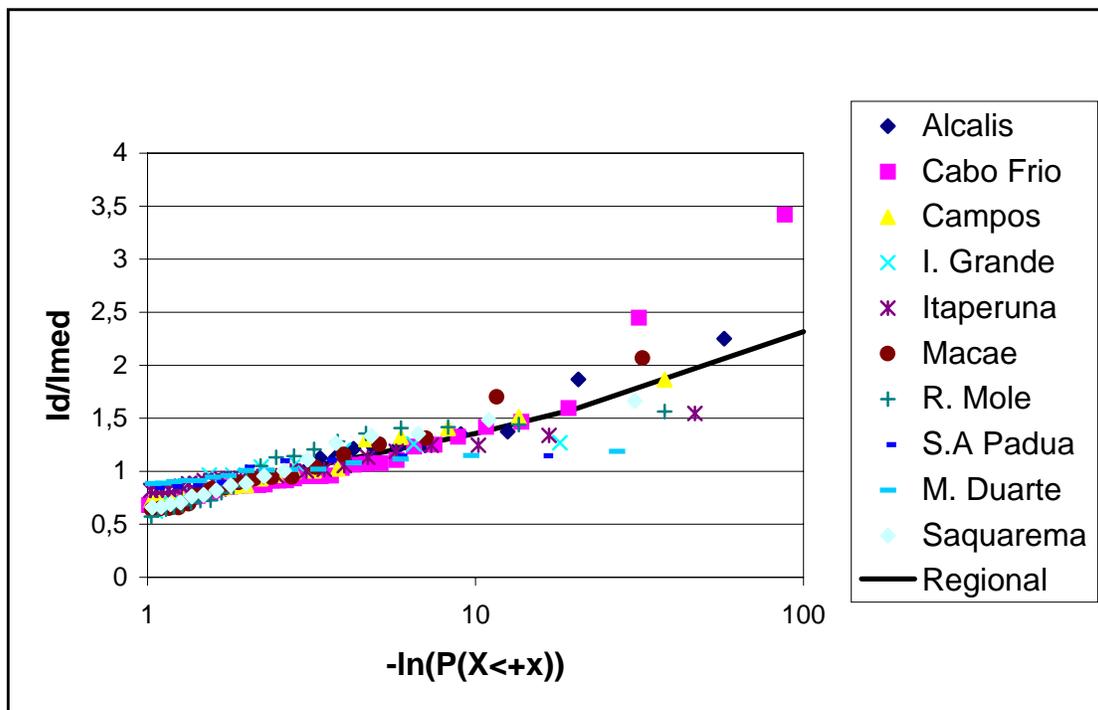


Figura 6F – Região homogênea 1 – duração 24h: ajuste da Distribuição Logística Generalizada para série parcial.

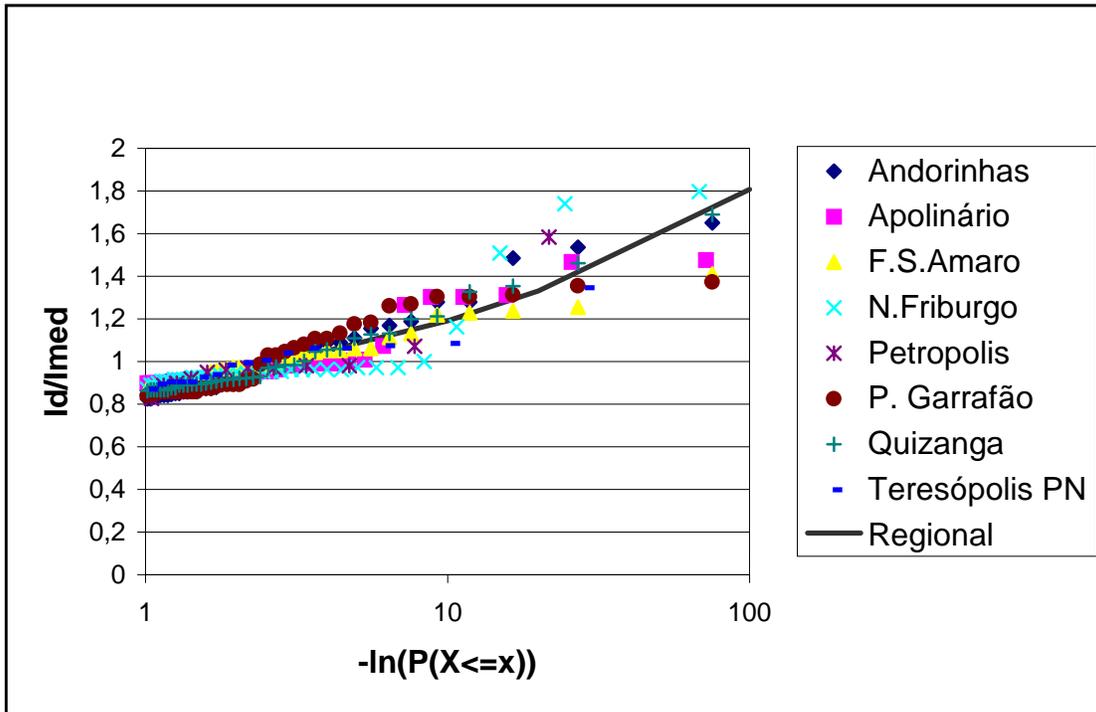


Figura 7A – Região homogênea 2 – duração 5min: ajuste da Distribuição Generalizada de Valores Extremos para série parcial.

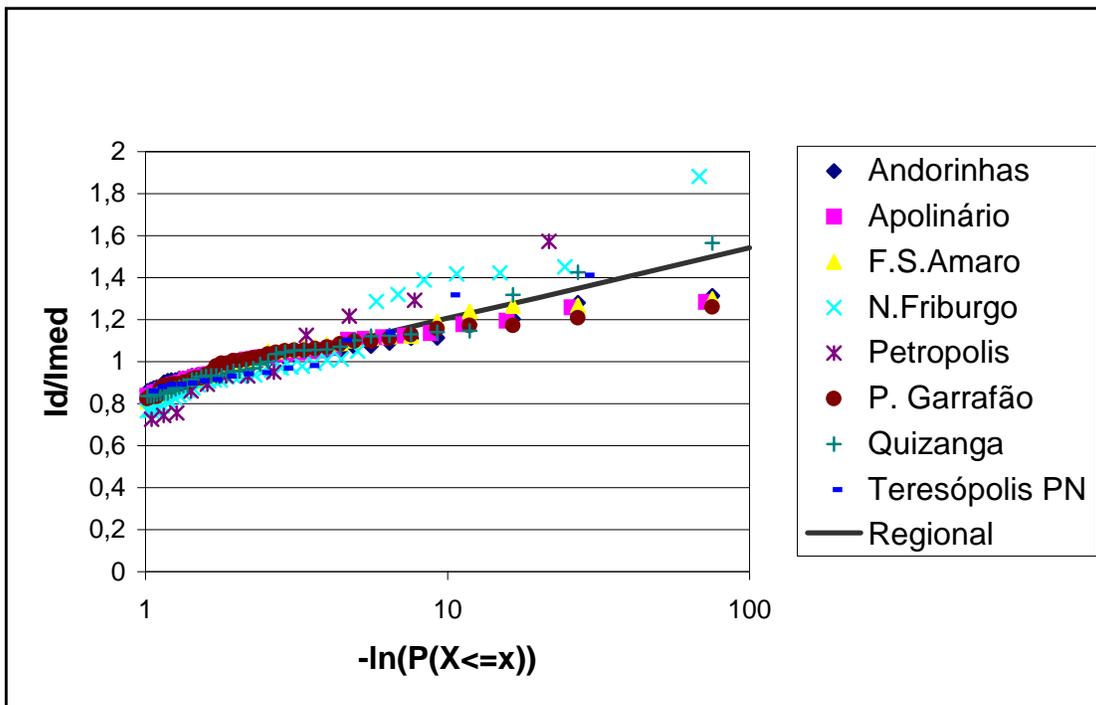


Figura 7B – Região homogênea 2 – duração 15min: ajuste da Distribuição Generalizada de Valores Extremos para série parcial.

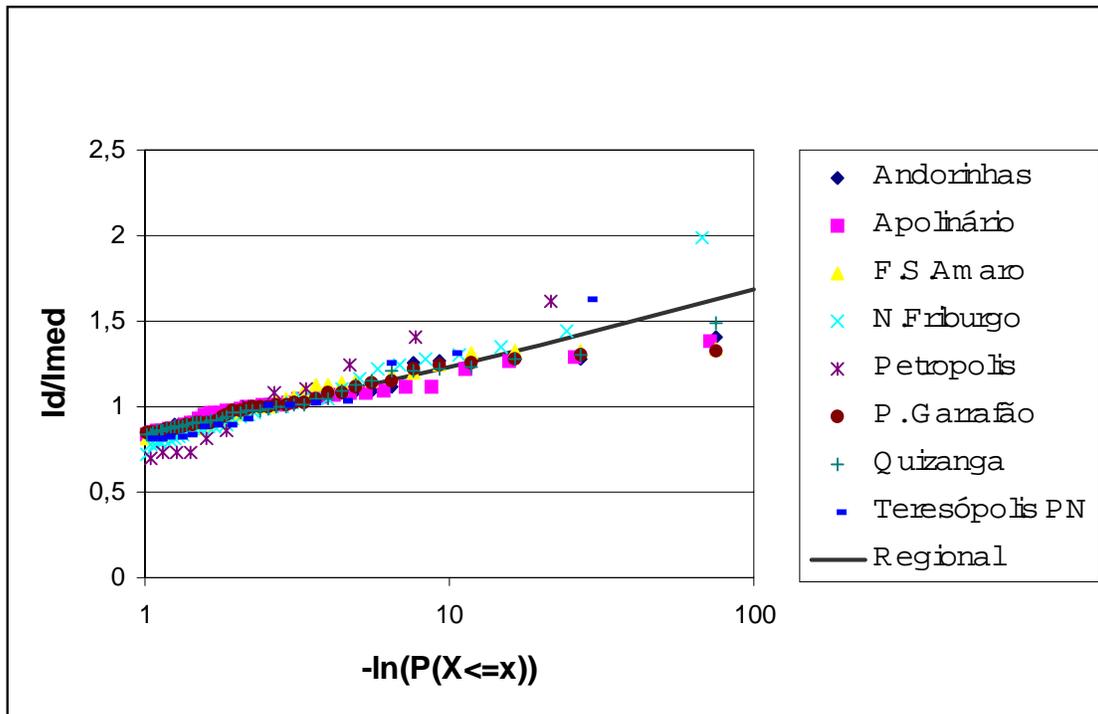


Figura 7C – Região homogênea 2 – duração 30min: ajuste da Distribuição Generalizada de Valores Extremos para série parcial.

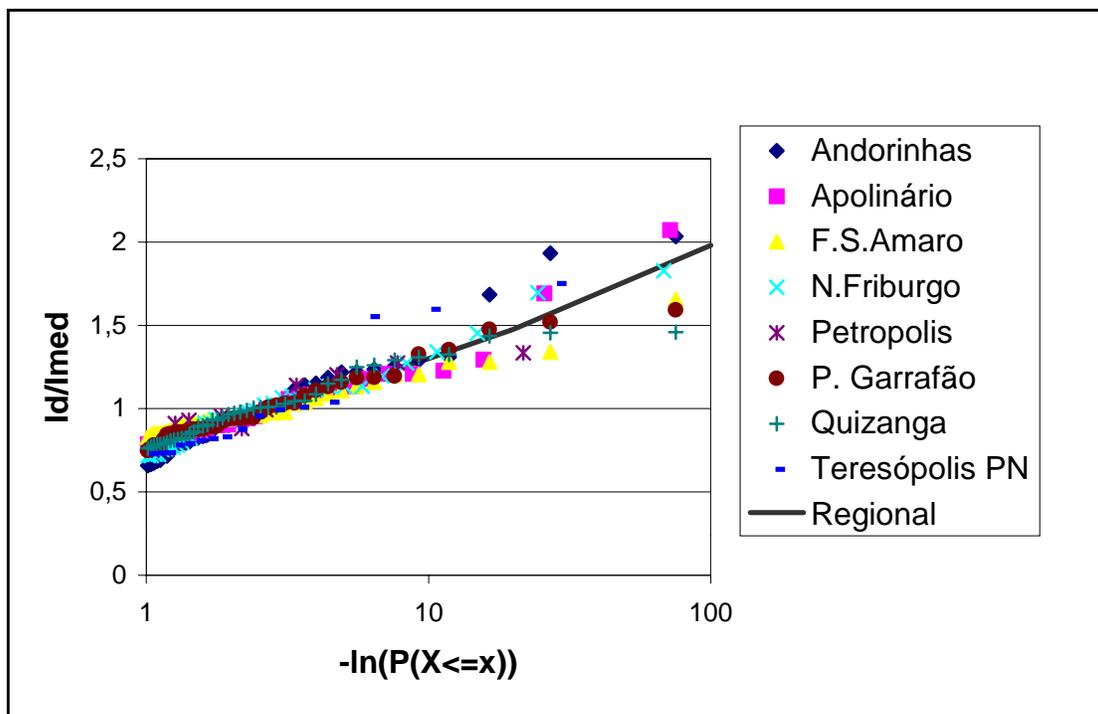


Figura 7D – Região homogênea 2 – duração 4h: ajuste da Distribuição Generalizada de Valores Extremos para série parcial.

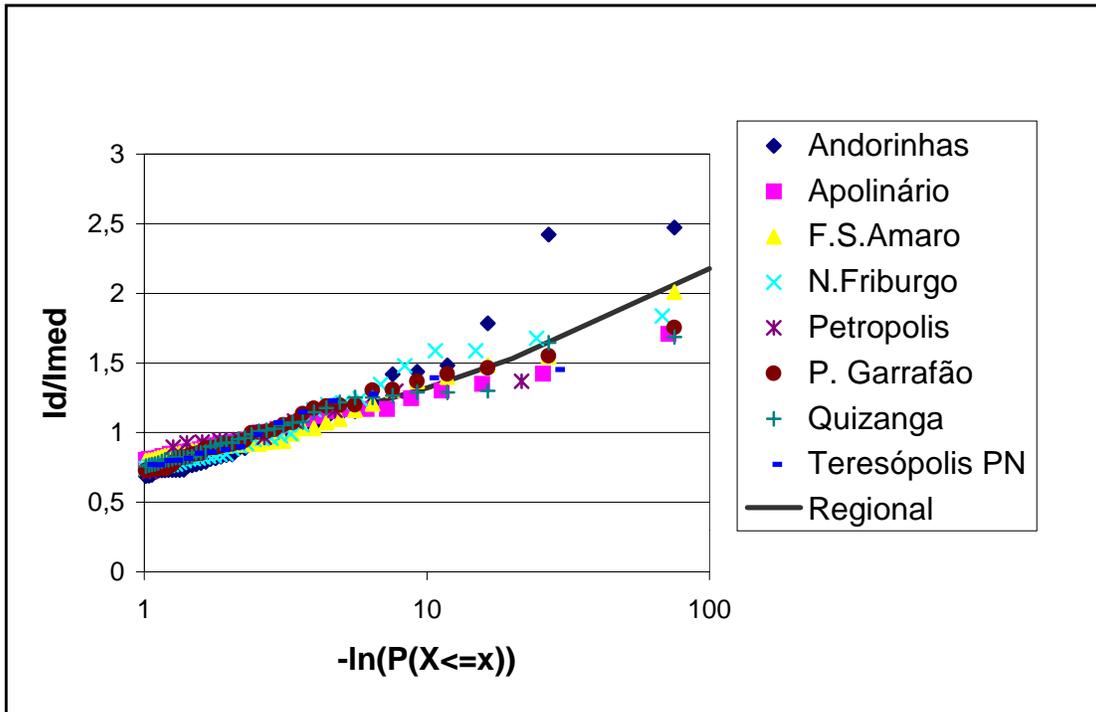


Figura 7E – Região homogênea 2 – duração 14h: ajuste da Distribuição Generalizada de Valores Extremos para série parcial.

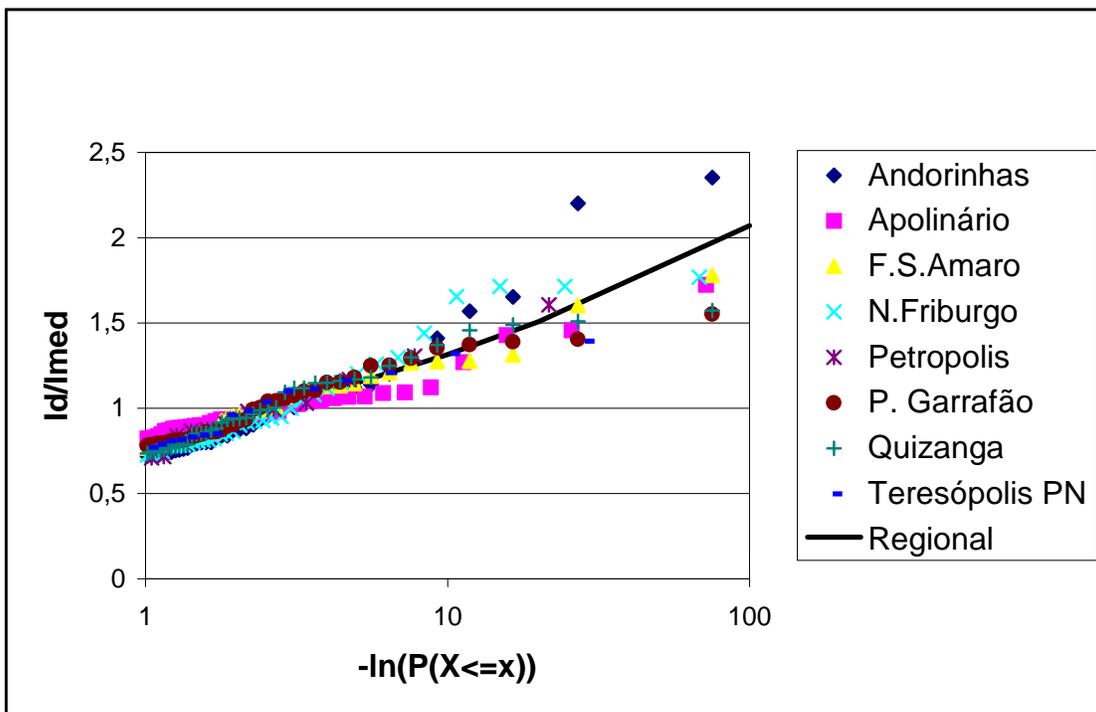


Figura 7F – Região homogênea 2 – duração 24h: ajuste da Distribuição Generalizada de Valores Extremos para série parcial.

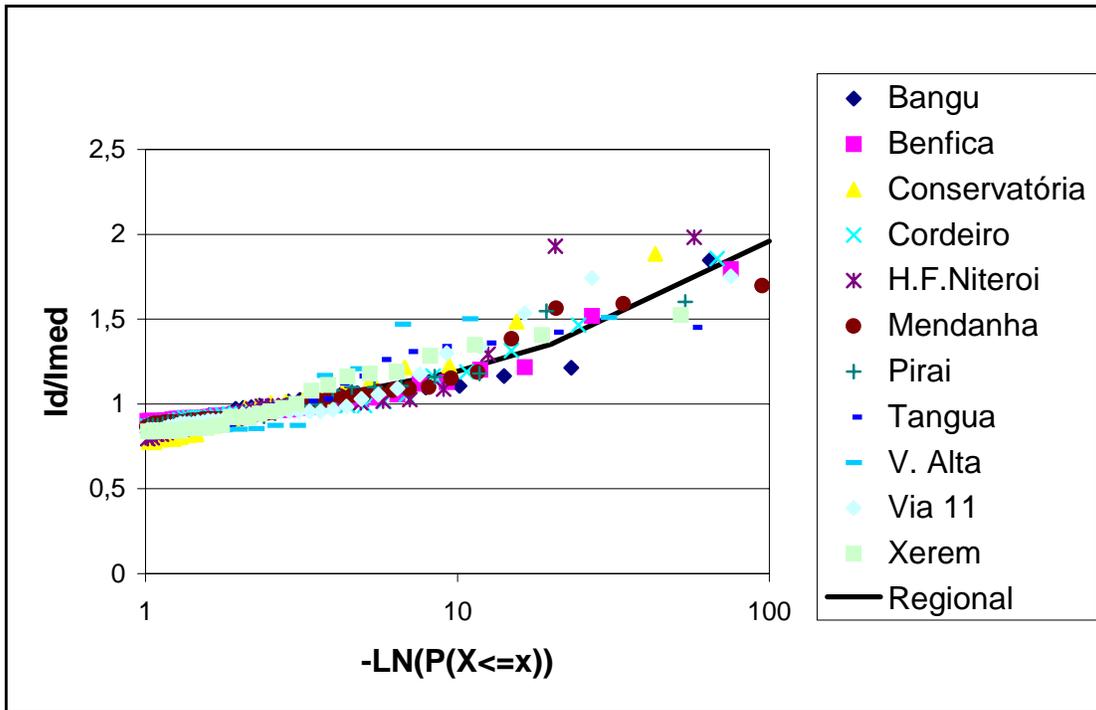


Figura 8A – Região homogênea 3 – duração 5min: ajuste da Distribuição Logística Generalizada para série parcial.

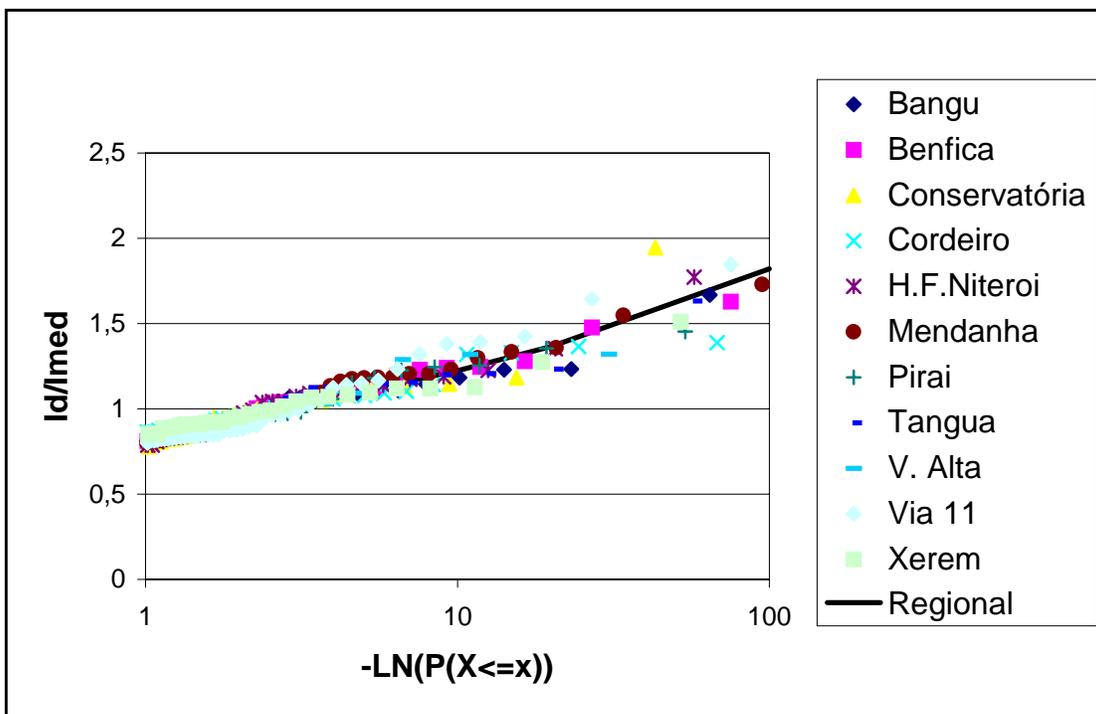


Figura 8B – Região homogênea 3 – duração 15min: ajuste da Distribuição Logística Generalizada para série parcial.

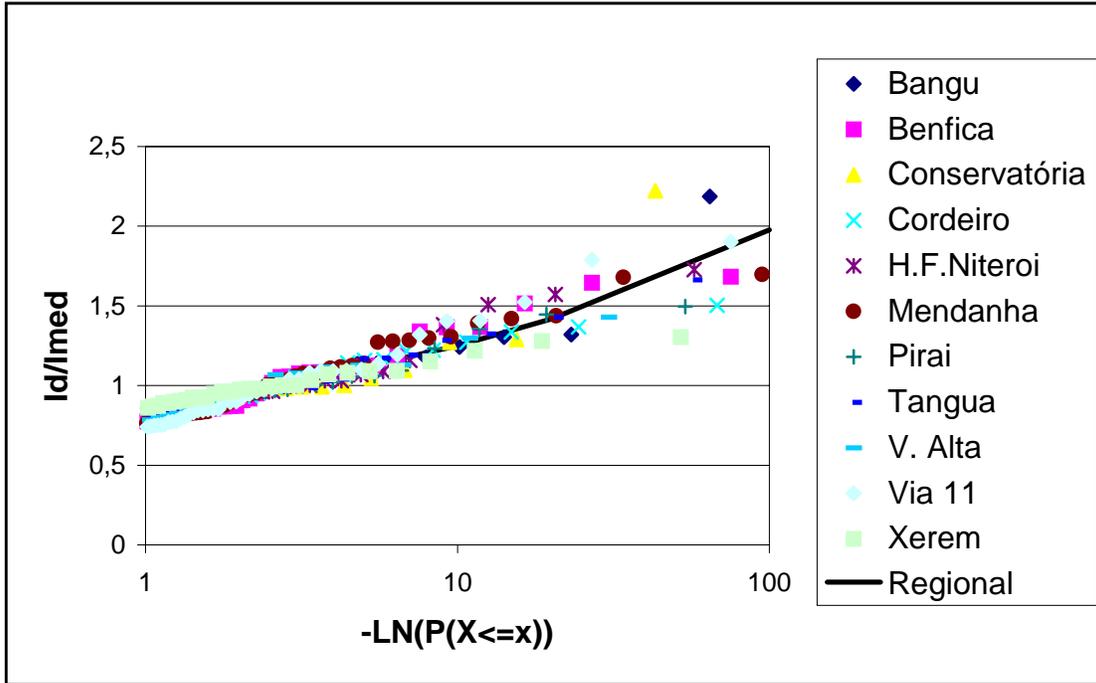


Figura 8C – Região homogênea 3 – duração 30min: ajuste da Distribuição Logística Generalizada para série parcial.

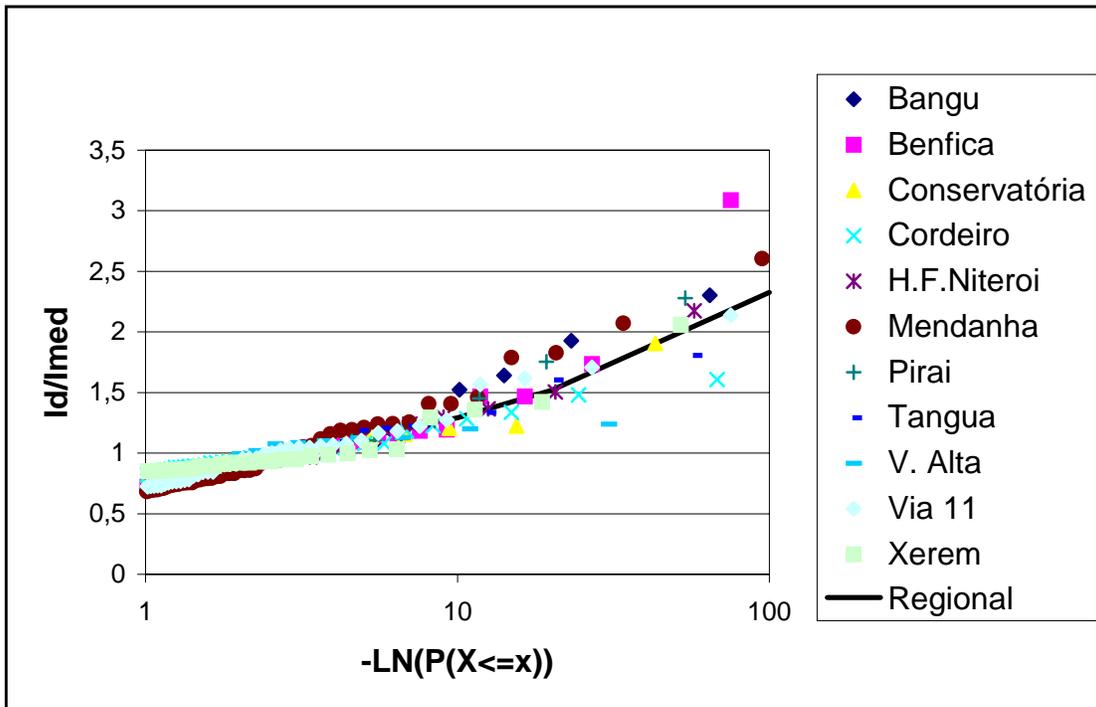


Figura 8D – Região homogênea 3 – duração 4h: ajuste da Distribuição Logística Generalizada para série parcial.

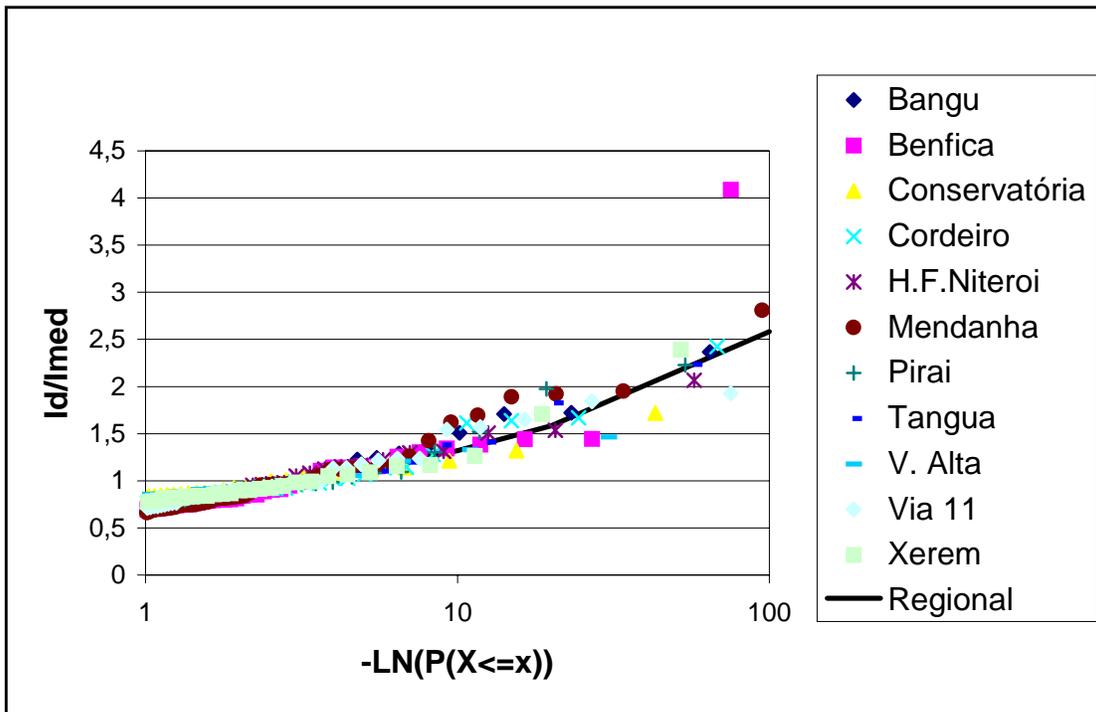


Figura 8E – Região homogênea 3 – duração 14h: ajuste da Distribuição Logística Generalizada para série parcial.

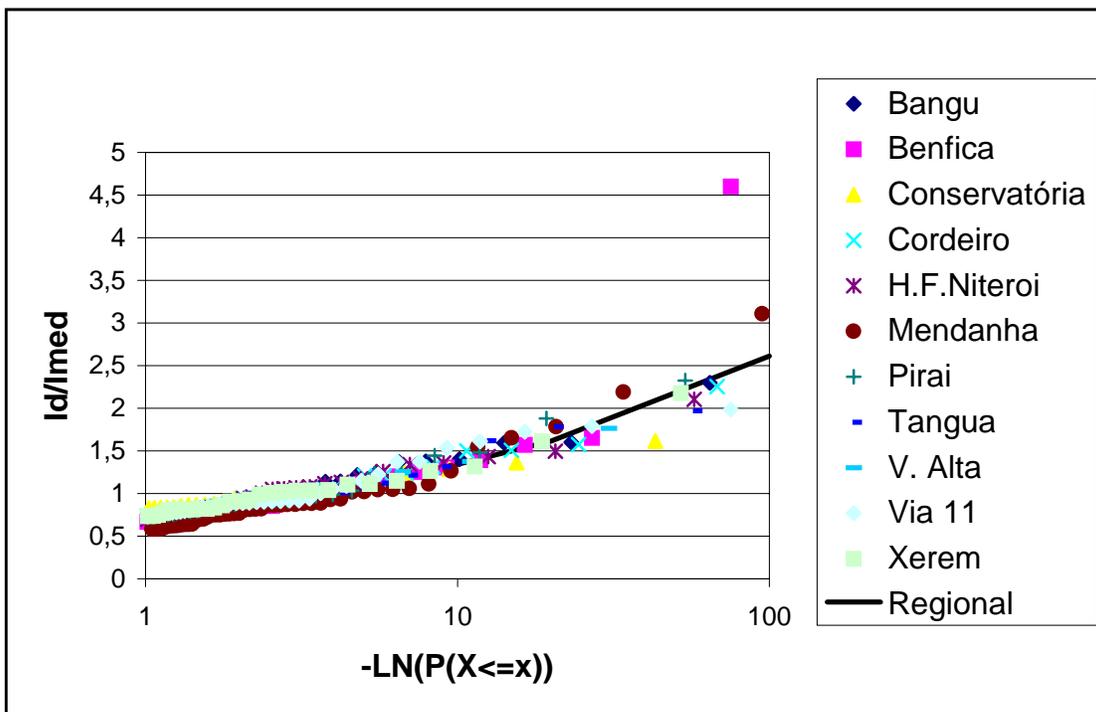


Figura 8F – Região homogênea 3 – duração 24h: ajuste da Distribuição Logística Generalizada para série parcial.

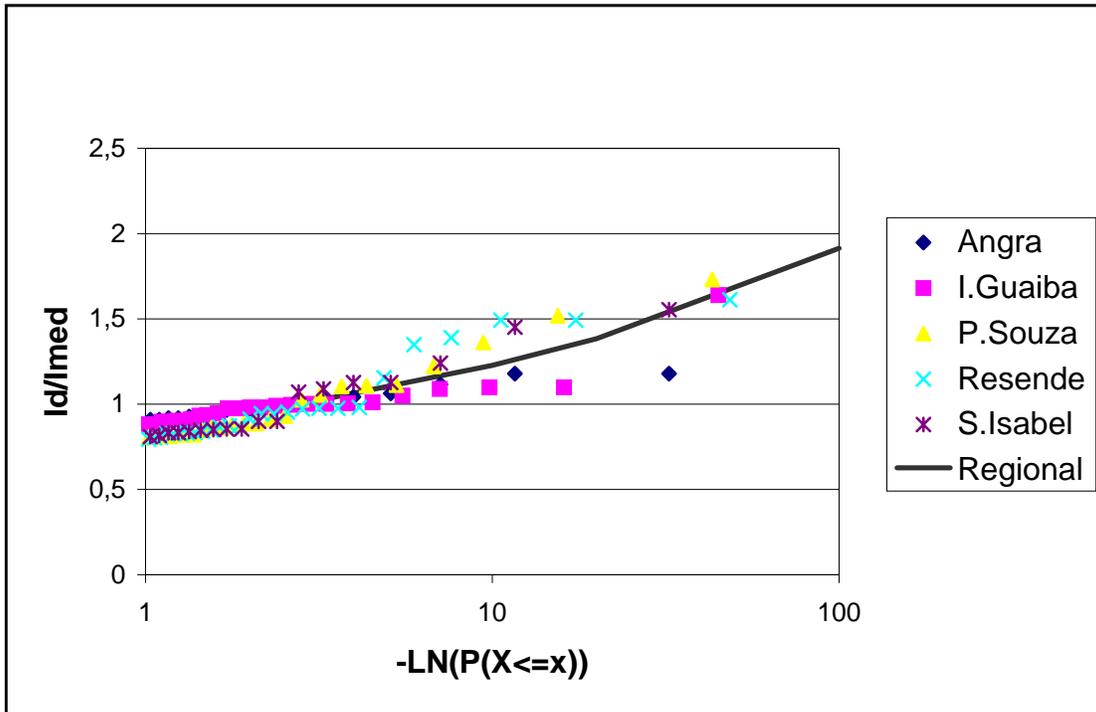


Figura 9A – Região homogênea 4 – duração 5min: ajuste da Distribuição Logística Generalizada para série parcial.

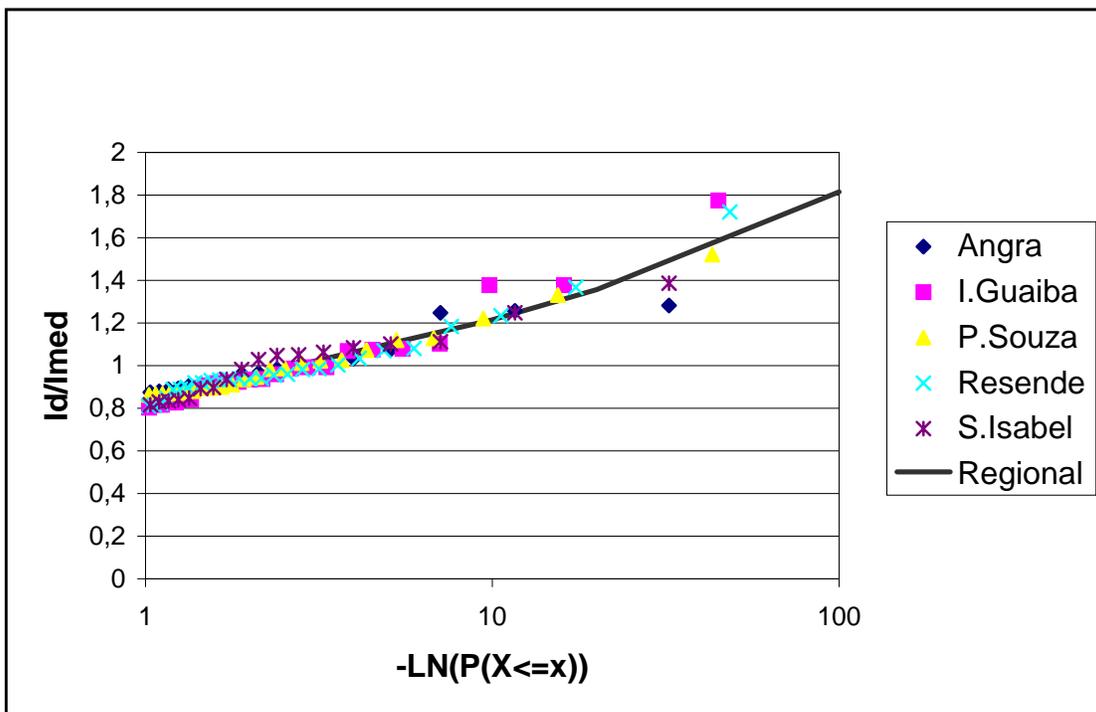


Figura 9B – Região homogênea 4 – duração 15min: ajuste da Distribuição Logística Generalizada para série parcial.

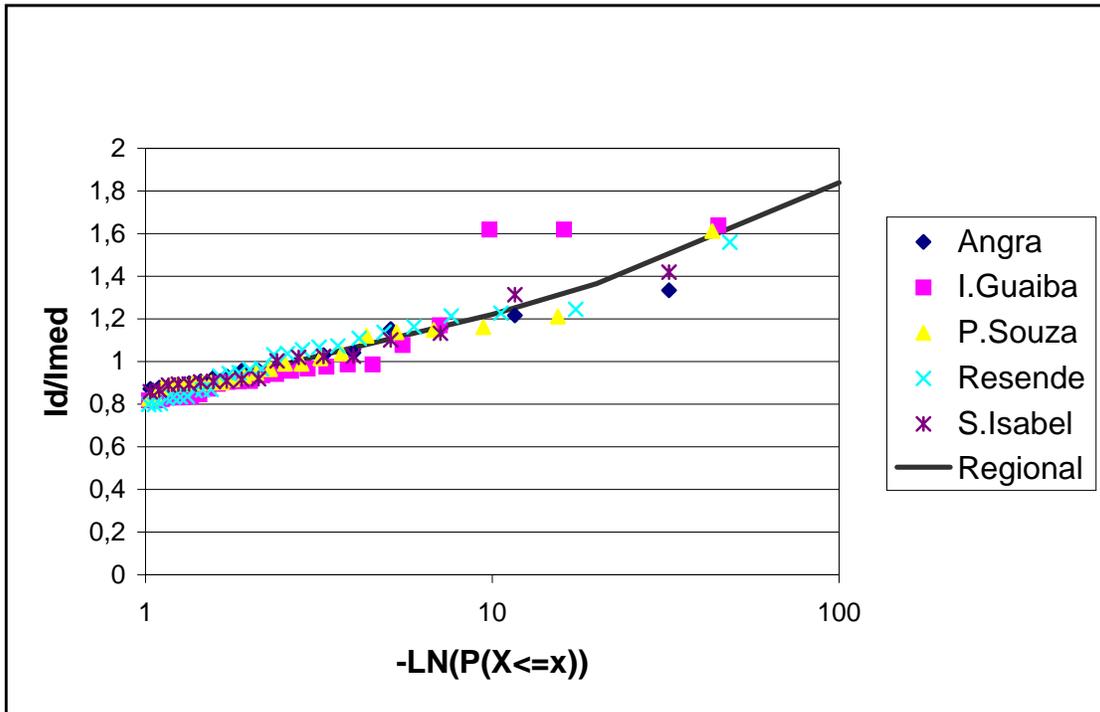


Figura 9C – Região homogênea 4 – duração 30min: ajuste da Distribuição Logística Generalizada para série parcial.

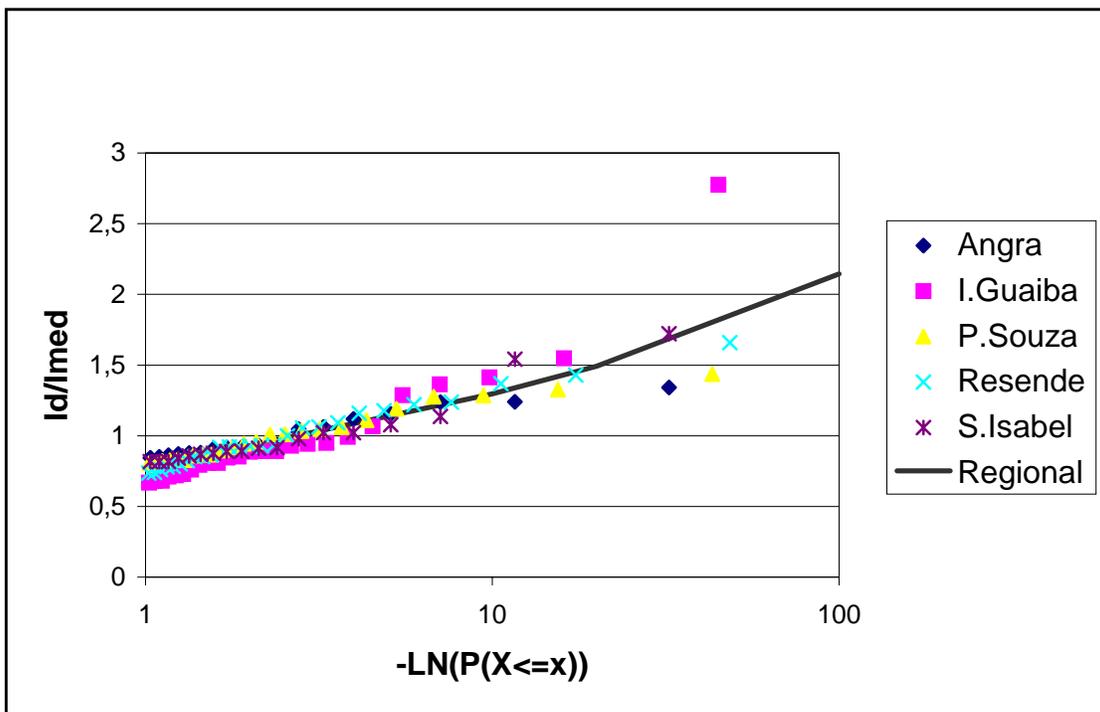


Figura 9D – Região homogênea 4 – duração 4h: ajuste da Distribuição Logística Generalizada para série parcial.

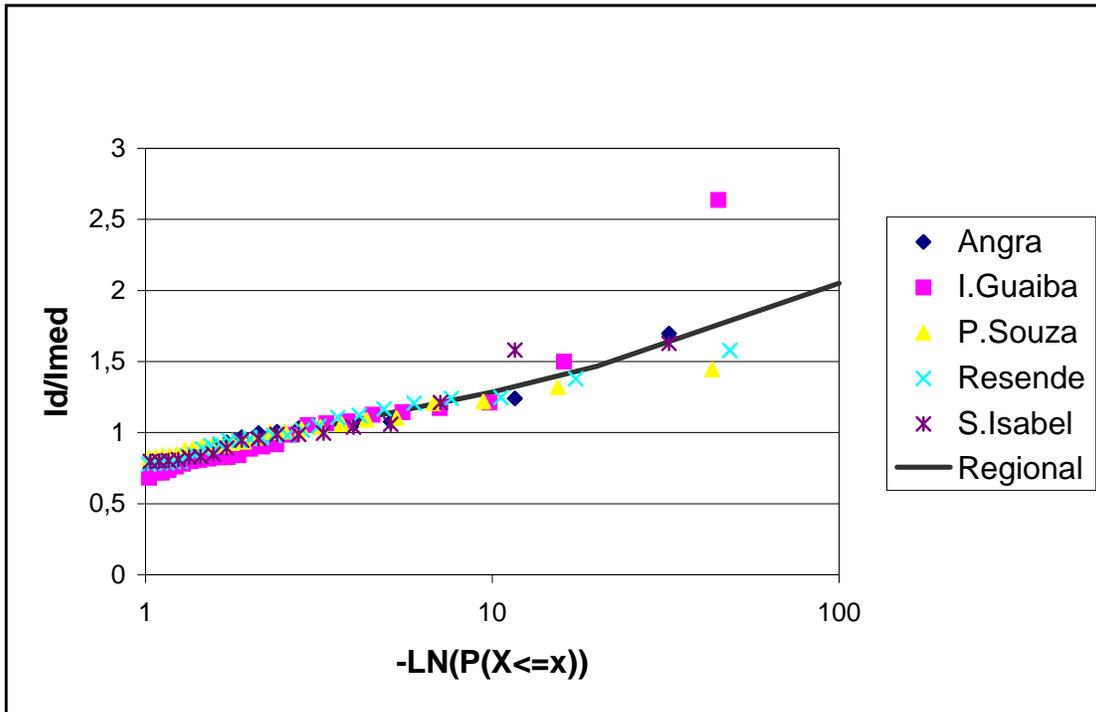


Figura 9E – Região homogênea 4 – duração 14h: ajuste da Distribuição Logística Generalizada para série parcial.

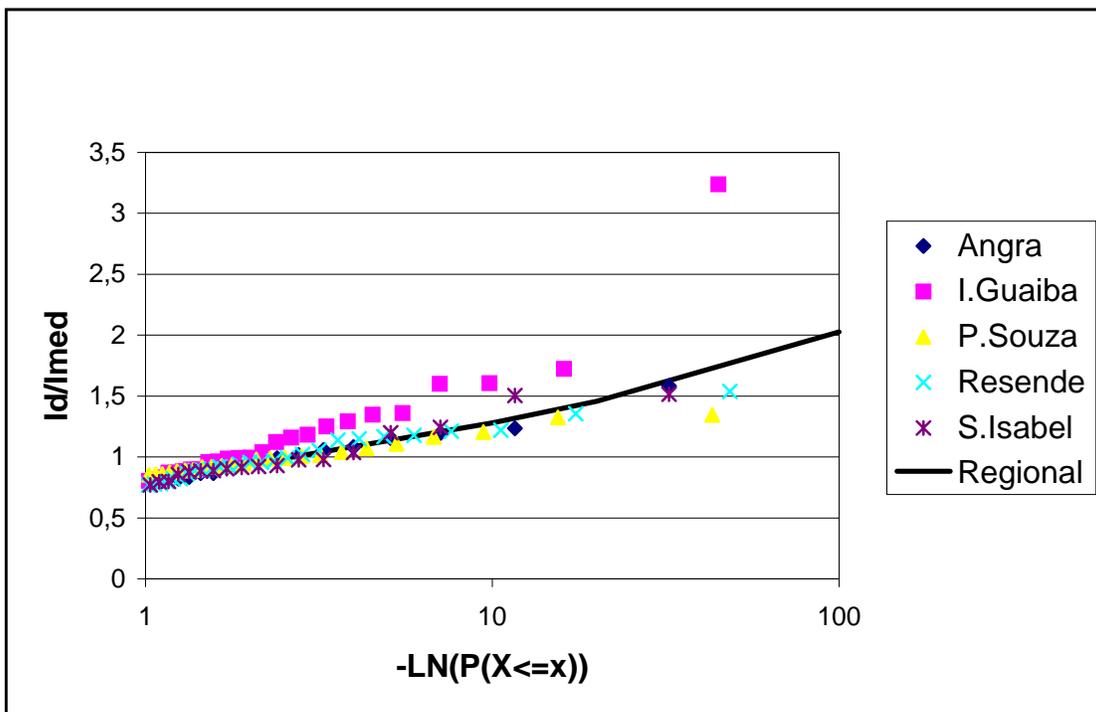


Figura 9F – Região homogênea 4 – duração 24h: ajuste da Distribuição Logística Generalizada para série parcial.

Cabe ressaltar que as distribuições regionais ajustadas referem-se a probabilidades associadas a eventos que excedem a um dado limite. Para estimar a distribuição de probabilidade regional anual, aplicou-se a equação:

$$F(y) = \exp \{-\Lambda [1 - H(y)]\} \quad (5.1)$$

Onde:

- Λ – razão de ocorrência média das excedências anuais;
- $H(y)$ – distribuição selecionada a partir da série parcial adimensional;
- $F(y)$ – distribuição regional anual.

A dedução dessa equação é detalhada no item 7.2 do Anexo I.

A partir de $F(y)$, foi calculada a função $\mu_{T,d}$ de quantis adimensionais anuais para as durações d levantadas e períodos de retorno T desejados.

Para a região 2, os valores $\mu_{T,d}$ foram obtidos a partir da transformação da equação 5.1 para o modelo Poisson-GEV. A equação resultante é:

$$F(y) = \mathbf{x} + \frac{\mathbf{a}}{k} \left\{ 1 - \left[-\ln \left(\frac{\Lambda + \ln(1 - (1/T(y)))}{\Lambda} \right) \right]^k \right\} \text{ se } k \neq 0$$

Para as regiões 1, 3 e 4, os valores $\mu_{T,d}$ foram obtidos a partir da transformação da equação 5.1 para o modelo Poisson-Logística. A equação resultante é:

$$F(y) = \mathbf{x} + \frac{\mathbf{a}}{k} \left[1 - \left(-\frac{\ln(1 - (1/T(y)))}{\Lambda + \ln(1 - (1/T(y)))} \right)^k \right]$$

Os valores de $\mu_{T,d}$ (quantis adimensionais de frequência), calculados para diferentes períodos de retorno T , são apresentados nas Tabelas 9 a 12.

Tabela 9 – Região 1: quantis anuais adimensionais regionais $m_{r,d}$.

Período de Retorno (Anos)	2	5	10	20	50	75	100
Duração							
5 minutos	0,8618	1,1000	1,2616	1,4451	1,7444	1,9043	2,0301
10 minutos	0,8470	1,1117	1,2901	1,4918	1,8193	1,9937	2,1307
15 minutos	0,8433	1,1059	1,2896	1,5019	1,8550	2,0465	2,1984
30 minutos	0,8356	1,1223	1,3142	1,5301	1,8794	2,0647	2,2101
45 minutos	0,8341	1,1249	1,3188	1,5365	1,8876	2,0735	2,2192
1 hora	0,8322	1,1201	1,3166	1,5404	1,9068	2,1031	2,2578
2 horas	0,8212	1,1076	1,3181	1,5689	1,9994	2,2385	2,4306
3 horas	0,8210	1,1055	1,3158	1,5671	2,0001	2,2412	2,4352
4 horas	0,8207	1,1128	1,3244	1,5741	1,9986	2,2327	2,4200
8 horas	0,8260	1,1161	1,3206	1,5579	1,9543	2,1700	2,3413
14 horas	0,8271	1,1205	1,3236	1,5567	1,9418	2,1495	2,3138
24 horas	0,8225	1,1318	1,3393	1,5733	1,9525	2,1540	2,3121

Tabela 10 – Região 2: quantis anuais adimensionais regionais $m_{r,d}$.

Período de retorno (Anos)	2	5	10	20	50	75	100
Duração							
5 minutos	0,9948	1,1711	1,3211	1,5013	1,8049	1,9697	2,1002
10 minutos	1,0041	1,1526	1,2692	1,4011	1,6087	1,7155	1,7979
15 minutos	1,0274	1,1906	1,2962	1,3998	1,5388	1,6020	1,6475
30 minutos	1,0214	1,2121	1,3457	1,4843	1,6827	1,7775	1,8477
45 minutos	1,0205	1,2397	1,3972	1,5637	1,8071	1,9254	2,0137
1 hora	1,0197	1,2489	1,4151	1,5923	1,8536	1,9813	2,0771
2 horas	1,0328	1,2693	1,4294	1,5915	1,8172	1,9226	1,9998
3 horas	1,0211	1,2658	1,4438	1,6337	1,9142	2,0515	2,1545
4 horas	1,0181	1,2734	1,4626	1,6675	1,9750	2,1275	2,2426
8 horas	1,0063	1,2807	1,4979	1,7450	2,1366	2,3393	2,4959
14 horas	1,0061	1,2880	1,5112	1,7650	2,1673	2,3756	2,5365
24 horas	1,0152	1,2862	1,4917	1,7178	2,0636	2,2375	2,3699

Tabela 11 – Região 3: quantis anuais adimensionais regionais $m_{r,d}$.

Período de retorno (Anos)	2	5	10	20	50	75	100
Duração							
5 minutos	0,9478	1,1173	1,2668	1,4579	1,8077	2,0110	2,1781
10 minutos	0,9711	1,1354	1,2572	1,3953	1,6175	1,7344	1,8256
15 minutos	0,9645	1,1512	1,2924	1,4550	1,7205	1,8619	1,9729
30 minutos	0,9550	1,1668	1,3325	1,5275	1,8538	2,0308	2,1710
45 minutos	0,9477	1,1858	1,3737	1,5963	1,9713	2,1758	2,3383
1 hora	0,9532	1,1811	1,3578	1,5648	1,9088	2,0946	2,2415
2 horas	0,9377	1,1845	1,3873	1,6344	2,0634	2,3028	2,4955
3 horas	0,9339	1,1866	1,3977	1,6577	2,1147	2,3721	2,5802
4 horas	0,9309	1,1857	1,4005	1,6668	2,1380	2,4049	2,6213
8 horas	0,9176	1,1907	1,4296	1,7334	2,2862	2,6058	2,8680
14 horas	0,9142	1,1983	1,4479	1,7663	2,3477	2,6848	2,9616
24 horas	0,9146	1,2144	1,4719	1,7951	2,3752	2,7071	2,9778

Tabela 12 – Região 4: quantis adimensionais regionais anuais $m_{r,d}$.

Período de retorno (anos)	2	5	10	20	50	75	100
Duração							
5 minutos	0,8819	1,0771	1,2151	1,3758	1,6449	1,7916	1,9083
10 minutos	0,8924	1,0661	1,1923	1,3417	1,5964	1,7371	1,8498
15 minutos	0,8914	1,0781	1,2053	1,3500	1,5868	1,7136	1,8135
30 minutos	0,8890	1,0801	1,2104	1,3587	1,6011	1,7310	1,8332
45 minutos	0,8829	1,0973	1,2328	1,3803	1,6103	1,7291	1,8209
1 hora	0,8737	1,0988	1,2464	1,4104	1,6722	1,8097	1,9170
2 horas	0,8691	1,1118	1,2636	1,4279	1,6826	1,8135	1,9144
3 horas	0,8489	1,1165	1,2921	1,4875	1,7996	1,9637	2,0917
4 horas	0,8485	1,1056	1,2830	1,4865	1,8219	2,0025	2,1453
8 horas	0,8496	1,1118	1,2878	1,4862	1,8077	1,9785	2,1126
14 horas	0,8579	1,1065	1,2729	1,4602	1,7630	1,9238	2,0498
24 horas	0,8603	1,1043	1,2675	1,4509	1,7473	1,9045	2,0277

7

EQUAÇÕES IDF REGIONAIS

7.1 Forma Básica

O objetivo final do estudo foi definir as equações do tipo IDF (intensidade-duração-freqüência) para cada região homogênea:

$$\hat{i}_{T,d,j} = \bar{i}_d \mu_{T,d}$$

Onde:

- $\hat{i}_{T,d,j}$ é a estimativa de chuva (mm/h), de duração d (min), no local j , associada ao período de retorno T (anos);
- \bar{i}_d é o *index-flood* de cada estação. Para estimá-lo em locais desprovidos de informações, foi utilizado um modelo de regressão de \bar{i}_d com variáveis externas, tais como clima e características fisiográficas locais (ver item 7.2).
- $\mu_{T,d}$ representa os quantis adimensionais de freqüência, de validade regional, associados a d e T (ver tabelas 9 a 12).

7.2 Regressão Múltipla

Para explicar a variabilidade espacial de \bar{i}_d , vários modelos de regressão foram ajustados para diferentes variáveis climáticas e fisiográficas independentes. A seleção das variáveis independentes foi baseada no coeficiente de correlação ajustado, que é diretamente proporcional ao coeficiente de correlação e inversamente proporcional ao número de variáveis incluídas no modelo de regressão.

Em uma primeira etapa, foram consideradas a duração da chuva (d), a precipitação média anual (P) e a altitude (ALT) nas localidades das estações pluviográficas.. Dado que o objetivo desse trabalho é o de estabelecer a relação IDF regional, uma vez incluída a duração da precipitação, a atenção voltou-se para a inclusão das outras variáveis que efetivamente viessem a explicar a variabilidade espacial das intensidades médias para as diversas durações. A tabela 13 ilustra os resultados obtidos com os diversos modelos de regressão ajustados.

Tabela 13 – Comparação dos resultados com d, P e ALT

Região	Intervalo	Variáveis	Coefficiente de Correlação	Coefficiente de Correlação Ajustado
(1 e 4)	5min≤d<1h	d, P, ALT	0,9602	0,9201
(1 e 4)	5min≤d<1h	d, P	0,9384	0,8793
(1 e 4)	1h≤d≤24h	d, P, ALT	0,9694	0,9383
(1 e 4)	1h≤d≤24h	d, P	0,9691	0,9385
2	5min≤d<1h	d, P, ALT	0,9727	0,9434
2	5min≤d<1h	d, P	0,9596	0,9189
2	1h≤d≤24h	d, P, ALT	0,9887	0,9765
2	1h≤d≤24h	d, P	0,9756	0,9507
3	5min≤d<1h	d, P, ALT	0,9657	0,9303
3	5min≤d<1h	d, P	0,9654	0,9309
3	1h≤d≤24h	d, P, ALT	0,9933	0,9862
3	1h≤d≤24h	d, P	0,9899	0,9795

Em todos os modelos testados, a inclusão da variável independente ALT não proporcionou uma melhoria significativa na regressão de \bar{i}_d . Provavelmente esse fato vem refletir a condensação expressiva, que provoca intensidades de chuvas maiores, e que depende não só da altitude mas também da direção dominante das massas de ar úmidas na região. Esse efeito combinado de relevo e altitude parece poder ser explicado através da variável P, a qual os sintetiza de maneira indireta.

Em uma segunda etapa, foram incluídas a duração da chuva (d), a precipitação média anual (P) e a média de longo termo das precipitações diárias máximas anuais (Pdmax). A tabela 14 ilustra os resultados obtidos com os diversos modelos de regressão ajustados.

Tabela 14 – Comparação dos resultados com d, P, Pdmax

Região	Duração	Variáveis	Coefficiente de Correlação	Coefficiente de Correlação ajustado
(1 e 4)	5min≤d<1h	d, P, Pdmax	0,9431	0,8868
(1 e 4)	5min≤d<1h	d, Pdmax	0,9174	0,8398
(1 e 4)	1h≤d≤24h	d, P, Pdmax	0,9733	0,9462
(1 e 4)	1h≤d≤24h	d, Pdmax	0,9674	0,9351
2	5min≤d<1h	d, P, Pdmax	0,9738	0,9457
2	5min≤d<1h	d, Pdmax	0,9711	0,9415
2	1h≤d≤24h	d, P, Pdmax	0,9979	0,9957
2	1h≤d≤24h	d, Pdmax	0,9949	0,9895
3	5min≤d<1h	d, P, Pdmax	0,9704	0,9397
3	5min≤d<1h	d, Pdmax	0,9527	0,9061
3	1h≤d≤24h	d, P, Pdmax	0,9906	0,9806
3	1h≤d≤24h	d, Pdmax	0,9877	0,9752

Observa-se que a inclusão da variável Pdmax elevou o coeficiente de correlação ajustado, quando utilizada em conjunto com P, porém não de forma convincente o bastante para justificar a inclusão de mais uma variável independente no modelo de regressão, tendo-se em conta não só o preceito da parcimônia estatística, como também as aplicações práticas do presente trabalho, considerando que incluir Pdmax representaria um complicador adicional na utilização das equações.

A complicação resultaria de dois aspectos: da estimativa da variável, a partir de séries diárias, de preferência completas e longas (séries com falhas são muito freqüentes e de difícil preenchimento), com menor disponibilidade se comparadas com as séries de totais anuais; e da necessidade de elaboração do mapa de isoietas da variável, que certamente não estaria disponível, sendo mais provável a existência do mapa de isoietas totais anuais.

Essas considerações representam, a nosso ver, razões suficientes para se utilizar somente a variável P no modelo de regressão, com a finalidade de explicar a variabilidade espacial das intensidades médias, para as diversas durações. A tabela 15 mostra o resultado do modelo de regressão selecionado.

Tabela 15 - Modelo de regressão da variável dependente \bar{i}_d

Região	Modelo de regressão para o <i>index flood</i> \bar{i}_d
1	$\bar{i}_d = 44,888d^{-0,385}P^{0,244}$ (para 5min≤d<1h)
	$\bar{i}_d = 81,432d^{-0,771}P^{0,371}$ (para 1h≤d≤24h)
2	$\bar{i}_d = 39,445d^{-0,339}P^{0,234}$ (para 5min≤d<1h)
	$\bar{i}_d = 16,204d^{-0,761}P^{0,564}$ (para 1h≤d≤24h)
3	$\bar{i}_d = 36,301d^{-0,392}P^{0,276}$ (para 5hmin≤d<1h)
	$\bar{i}_d = 85,264d^{-0,789}P^{0,367}$ (para 1h≤d≤24h)
4	$\bar{i}_d = 44,888d^{-0,385}P^{0,244}$ (para 5min≤d<1h)
	$\bar{i}_d = 81,432d^{-0,771}P^{0,371}$ (para 1h≤d≤24h)

7.3 Equações IDF Regionais Finais

Combinando os modelos de regressão da tabela 13 com os quantis regionais das tabelas 9 a 12, as relações IDF válidas para o estado do Rio de Janeiro resultantes são:

$$\text{Região 1: } \hat{i}_{T,d,j} = 44,888d^{-0,385}P_j^{0,244} \mu_{T,d} \text{ para } T \leq 100 \text{ e } 5\text{min} \leq d < 1\text{h}$$

$$\hat{i}_{T,d,j} = 81,432d^{-0,771}P_j^{0,371} \mu_{T,d} \text{ para } T \leq 100 \text{ e } 1\text{h} \leq d \leq 24\text{h}$$

$$\text{Região 2: } \hat{i}_{T,d,j} = 39,445d^{-0,339}P_j^{0,234} \mu_{T,d} \text{ para } T \leq 100 \text{ e } 5\text{min} \leq d < 1\text{h}$$

$$\hat{i}_{T,d,j} = 16,204d^{-0,761}P_j^{0,564} \mu_{T,d} \text{ para } T \leq 100 \text{ e } 1\text{h} \leq d \leq 24\text{h}$$

Região 3: $\hat{i}_{T,d,j} = 36,301d^{-0,392}P_j^{0,276}\mu_{T,d}$ para $T \leq 100$ e $5\text{min} \leq d < 1\text{h}$

$$\hat{i}_{T,d,j} = 85,264d^{-0,789}P_j^{0,367}\mu_{T,d} \text{ para } T \leq 100 \text{ e } 1\text{h} \leq d \leq 24\text{h}$$

Região 4: $\hat{i}_{T,d,j} = 44,888d^{-0,385}P_j^{0,244}\mu_{T,d}$ para $T \leq 100$ e $5\text{min} \leq d < 1\text{h}$

$$\hat{i}_{T,d,j} = 81,432d^{-0,771}P_j^{0,371}\mu_{T,d} \text{ para } T \leq 100 \text{ e } 1\text{h} \leq d \leq 24\text{h}$$

Onde:

- $\hat{i}_{T,d,j}$ é a estimativa da intensidade da chuva de duração d associada a um período de retorno T em um local j dentro de uma região homogênea do estado do Rio de Janeiro (mm/h).
- d é a duração da precipitação (min).
- P_j é a precipitação média anual (mm) no local j , dentro de cada região homogênea. Para locais que não possuem estações pluviométricas e pluviográficas, os valores de P_j podem ser obtidos a partir do mapa isoietal (ver figura 1).
- $\mu_{T,d}$ é o quantil adimensional regional. Esse valor é obtido a partir das tabelas 9 a 12.

O Anexo II apresenta uma análise comparativa dos resultados obtidos a partir das equações apresentadas com as de outros estudos de chuvas intensas disponíveis para o estado do Rio de Janeiro.

8

CONCLUSÃO

Conforme descrito nos subitens 6.2.1 e 6.2.2, das 44 estações pluviográficas inicialmente disponíveis, foram efetivamente utilizadas 34 para estimar as relações IDF para o estado do Rio de Janeiro. O número de anos hidrológicos em cada estação variou de 8 a 49 anos, com um número médio de 21 anos de registro. Pelo menos em duas regiões, em que foram utilizados os modelos das distribuições de Poisson-Logística e Poisson-GEV, com excedências anuais independentes acima de um limite, os tamanhos das amostras foram aumentados, o que certamente produziu parâmetros e quantis mais robustos e confiáveis. Além disso, a estimativa baseada em séries de duração parcial utiliza os maiores valores da amostra, que são mais representativos na cauda superior das distribuições de probabilidade. O agrupamento dos dados de locais diferentes dentro de uma área homogênea permite que a análise de frequência regional baseada em momentos-L seja usada não somente para estimar quantis em locais sem informação, mas também para melhorar a estimativa local. Essas razões são suficientes para recomendar o uso das equações regionais apresentadas neste trabalho em lugar das equações IDF locais desenvolvidas anteriormente.

A variabilidade espacial da intensidade média de chuva para cada duração (o fator de escala – *index-flood*) sobre o estado do Rio de Janeiro aparece associada com a variabilidade espacial da precipitação média anual. A explicação física para tal associação está provavelmente relacionada aos efeitos de longo período das altas taxas de condensação a barlavento das encostas refletidas pela distribuição geográfica das precipitações anuais. A variabilidade espacial das equações deve afetar os futuros projetos de estruturas hidráulicas no estado.

9

REFERÊNCIAS BIBLIOGRÁFICAS

- BERTONI, J.C. & TUCCI, C.E. Precipitação. *In: Hidrologia: Ciência e Aplicação*. Porto Alegre : Ed. UFRGS/ABRH/EDUSP, p. 177-241, 1993.
- BOBÉE, B. The log Pearson type 3 distribution and its application in hydrology. *Water Resources Research*, v. 11, nº 5, p. 681-689, 1975.
- BOBÉE, B. & RASMUSSEN, P. Recent advances in flood frequency analysis. U.S. National Report to IUGG, 1991-1994, *Rev. Geophysics*, v. 33 Suppl. (<http://earth.agu.org/revgeophys/bobee01/bobee01.htm>), 1995.
- BOUGHTON, W.C. A frequency distribution for annual floods. *Water Resources Research*, v. 16, p. 347-354, 1980.
- BUIHAND, T.A. Statistics of extremes in climatology. *Statistica Neerlandica*, v. 43, p. 1-30, 1989.
- BUIHAND, T.A. Extreme rainfall estimation by combining data from several sites. *Hydrological Sciences Journal*, v. 36, nº 4, p. 345-365, 1991.
- BURN, D.H. Cluster analysis as applied to regional flood frequency. *Journal of Water Resources Planning and Management*, v. 115, p. 567-582, 1989.
- CAVADIAS, G.S. The canonical correlation approach to regional flood estimation. *In: Regionalization in Hydrology*, IAHS Publication 191, Ed. M. Beran, M. Brilly, A. Becker & O. Bonacci, IAHS, Wallingford, Reino Unido, p. 171-178, 1980.
- CHOW, V.T. The log probability law and its engineering applications. *Proceedings ASCE*, v. 80(536), p. 1-25, 1954.
- CORREIA, F.N. *Métodos de análise e determinação de caudais de cheia*. Tese de concurso para Investigador Auxiliar do LNEC, Laboratório Nacional de Engenharia Civil, Lisboa, 380 pp., 1983.
- CRAMÉR, H. & LEADBETTER, M.R. *Stationary and related stochastic processes*, John Wiley, New York, 1967.

- CUNNANE, C. A note on the Poisson assumption in partial duration series models. *Water Resources Research*, v. 15, nº 2, p. 489-494, 1979.
- CUNNANE, C. A particular comparison of annual maximum and partial duration series methods of flood frequency prediction. *Journal of Hydrology*, v. 18, p. 257-271, 1973.
- DALRYMPLE, T. *Flood-frequency analyses, Manual of Hydrology: Part.3. Flood-flow Techniques*, Geological Survey Water Supply Paper 1543-A, U.S. Government Printing Office, Washington, D.C., 80p., 1960.
- DANTAS, M.E. *et al. Domínios Geomorfológicos do Estado do Rio de Janeiro*. SIMPÓSIO NACIONAL DE GEOMORFOLOGIA, 2, UGB, Florianópolis-SC, p. 593-596, 1998
- ELETOBRÁS. *Metodologia para Regionalização de Vazões*. Vol. 1, Eletrobrás, DPE, Departamento de Recursos Energéticos, Rio de Janeiro, 203 pp., 1985.
- GINGRAS, D. & ADAMOWSKI, K. Homogeneous region delineation based on annual flood generation mechanisms. *Hydrological Sciences Journal*, v. 38, nº 1, p. 103-121, 1993.
- GOULART, R.D. Diagnóstico da suscetibilidade à erosão dos solos da bacia hidrográfica do Bonfim-Correas. Petrópolis-RJ, PPGG-UFRJ, 129 p. (dissertação de mestrado), 1999.
- GREENWOOD, J.A., LANDWEHR, J.M., MATALAS, N.C., 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.
- GUMBEL, E.J. *Statistics of Extremes*. Columbia University Press, New York, 375 pp., 1958.
- GUPTA, V.K., DUCKSTEIN, L., PEEBLES, R.W. On the joint distribution of the largest flood and its occurrence time. *Water Resources Research*, v. 12, nº 2, p. 295-304, 1976.
- GUTTMAN, N.B. The use of L-moments in the determination of regional precipitation climates. *Journal of Climate*, v. 6, p. 2309-2325, 1993.
- HARTIGAN, J.A. *Clustering Algorithms*, Wiley, New York, 1975, *apud* Statsoft Inc., Electronic Statistics Textbook, Statsoft, Tulsa, OK, Estados Unidos (<http://www.statsoft.com/textbook/stathome.html>), 1997.
- HOSKING, J.R.M. Fortran routines for use with method of L-moments. Version 2. In: *IBM Research Report*, New York, IBM Research Division, RC 17097, 117 p., ago., 1991.
- HOSKING, J.R.M. L-moments: analysis and estimation of distributions using linear combination of order statistics. *Journal of the Royal Statistical Society, Series B*, v. 52, p. 105-124, 1990.
- HOSKING, J.R.M. Some theoretical results concerning L-moments. *IBM Research Report*, RC 14492, IBM Research Division, Yorktown Heights, NY, EUA, 1989.
- HOSKING, J.R.M. The theory of probability weighted moments. *IBM Research Report*, RC 12210, IBM Research Division, Yorktown Heights, NY, EUA, 1986.
- HOSKING, J.R.M. & WALLIS, J.R. *Regional Frequency Analysis – An Approach Based on L-Moments*. 224 p. Cambridge University Press, Cambridge, Reino Unido, 1997.
- HOSKING, J.R.M. & WALLIS, J.R. Correction to “some statistics useful in regional frequency analysis”. *Water Resources Research*, v. 31, nº 1, p. 251, 1995.
- HOSKING, J.R.M. & WALLIS, J.R. Some statistics useful in regional frequency analysis. *Water Resources Research*, v. 29, nº 1, p. 271-281, 1993.
- HOSKING, J.R.M. & WALLIS, J.R. Parameter and quantile estimation for the generalized Pareto distribution. *Technometrics*, v. 29, p. 339-349, 1987.
- HUFF, F.A. Time distribution of rainfall in heavy storms. *Water Resources Research*, v. 3, nº 4, p. 1007-1019, 1967.
- INSTITUTION OF ENGINEERS AUSTRALIA. *Australian rainfall and runoff: a guide to flood estimation*. V. 1, Institution of Engineers Australia, Canberra, Austrália, 374p., 1987.
- KOTTEGODA, N.T. & ROSSO, R. *Statistics, Probability and Reliability for Civil and Environmental Engineers*, McGraw-Hill, New York, 735 pp., 1997.

LANDWEHR, J.M., MATALAS, N.C., WALLIS, J.R. Probability-weighted moments compared with some traditional techniques in estimating Gumbel parameters and quantiles. *Water Resources Research*, v. 15, nº 5, p. 1.055-1.046, 1979.

LAURA, A.A. & FERREIRA FILHO, W.M. Uso de séries parciais para análise de frequência de chuvas intensas. In: XII SIMPÓSIO BRASILEIRO DE RECURSOS HÍDRICOS. Vitória, nov., 1997. CD-ROM.

LAURSEN, E.M. Comment on "Paleohydrology of southwestern Texas" by KOCHER, R.C., BAKER, V.R., PATTON, P.C. *Water Resources Research*, v. 19, p. 1339, 1983.

LEADBETTER, M.R., LINDGREN, G., ROOTZÉN, H. *Extremes and related properties of random sequences and processes*, Springer-Verlag, New York, 335 pp., 1983.

MADSEN, H., ROSBJERG, D., HARREMOES, P. Application of the partial duration series approach in the analysis of extreme rainfalls in extreme hydrological events: precipitation, floods and droughts. *Proceedings of the Yokohama Symposium*, I.A.S.H. Publication 213, p. 257-266, 1993.

MINISTÉRIO DAS RELAÇÕES EXTERIORES. O Estado do Rio de Janeiro. Disponível em <http://www.mre.gov.br/ndsg/textos/rjest-p.htm>.

NAGHETTINI, M., POTTER, K.W., ILLANGASEKARE, T. Estimating the upper-tail of flood-peak frequency distributions using hydrometeorological information. *Water Resources Research*, v. 32, nº 6, p. 1.729-1.740, 1996.

NATHAN, R.J. & MCMAHON, T. Identification of homogeneous regions for the purpose of regionalisation. *Journal of Hydrology*, v. 121, p. 217-238, 1990.

NERC. *Flood Studies Report, vol. 1, Hydrological Studies*. National Environment Research Council, Londres, Reino Unido, 1975.

NORTH, M. Time-dependent stochastic model of floods. *Journal of Hydraulics Division*, ASCE, v. 106, nº 05, p. 717-731, 1980.

NRC. *Estimating Probabilities of Extreme Floods*. National Research Council, National Academy Press, Washington, 141, pp., 1987.

PEARSON, C.P. Regional flood frequency for small New Zealand basins 2: flood frequency groups. *Journal of Hydrology* (Nova Zelândia), v. 30, p. 53-64, 1991.

PERICHI, L.R. & RODRÍGUEZ-ITURBE, I. On the statistical analysis of floods. In: *A Celebration of Statistics*, ed. ATKINSON, A.C. & FIENBERG, S.E. Springer-Verlag, New York, p. 511-541, 1985.

PFADFSTETTER, O. *Chuvas intensas no Brasil*. Rio de Janeiro. Departamento Nacional de Obras de Saneamento, 2ª ed., 1982. 1ª ed. 1957.

PINHEIRO, M.M.G. *Estudo de chuvas intensas na região metropolitana de Belo Horizonte – RMBH*. Belo Horizonte: EE-UFMG. Dissertação de Mestrado, 216 p., 1997.

PINHEIRO, M.M.G. & NAGHETTINI, M. Análise regional de frequência e distribuição temporal das tempestades na região metropolitana de Belo Horizonte-RMBH. In: *Revista Brasileira de Recursos Hídricos*, v. 3 nº 4, out./dez., p.73-88, 1998.

PINTO, E.J.A. *Determinação de curvas e equações intensidade-duração-freqüência de estações pluviográficas*. Publicação interna da CPRM-SUREG/BH. Belo Horizonte, set., 1998.

PINTO, E.J.A. & NAGHETTINI, M. *Definição de regiões homogêneas e regionalização de freqüência das precipitações diárias máximas anuais da bacia do alto rio São Francisco*. Anais... 13º SIMPÓSIO BRASILEIRO DE RECURSOS HÍDRICOS (CD-ROM), Belo Horizonte, 1999.

PIRES, C.L.F. Análise de Freqüência – Revisão Metodológica. In: *A Água em Revista*, nº 3. CPRM, Belo Horizonte, out., p. 13-22, 1994.

PIRES, C.L.F. & KELMAN, J. Regionalização de precipitações máximas diárias em uma região no vale do Paraíba do Sul usando momentos-L. In: *A Água em Revista*, nº 11. CPRM, Belo Horizonte, nov., p. 49-56, 1999.

POTTER, K.W. Research on flood frequency analysis: 1983-1986. *Rev. Geophys.*, v. 26, nº 3, p. 113-118, 1987.

REICH, B.M. Lysenkoism. *In: U.S. flood determinations, AGU Surface Runoff Committee – Session on flood frequency methods*, San Francisco, CA, 13 p., 1977.

ROSBJERG, D. Estimation in partial duration series with independent and dependent peak values. *Journal of Hydrology*, v. 76, p. 183-195, 1984.

ROSBJERG, D. & MADSEN, H. On the choice of threshold level in partial duration series. *Proceedings of the Nordic Hydrological Conference, Alta (Noruega)*, NHP Report 30, pp. 604-615, 1992.

SCHAEFER, M.C. Regional analyses of precipitation annual maxima in Washington State. *Water Resources Research*, v. 26, n° 1, p. 119-131, 1990.

SMITH, J.A. Regional flood frequency analysis using extreme order statistics of the annual peak record. *Water Resources Research*, v. 25, n° 2, p. 311-317, 1989.

SMITH, R.L. Threshold models for sample extremes. *In: Statistical Extremes and Applications*, ed. J. Tiago de Oliveira, 621-638, D. Reidel, Hingham, Ma., EUA, 1984.

STEDINGER, J.R., VOGEL, R.M., & FOUFOULA-GEORGIU, E. Frequency analysis of extreme events. Chapter 18. *In: Handbook of Hydrology*, ed. D.R. Maidment, McGraw-Hill, New York, 18.1-18.66, 1993.

TAESOMBUT, V. & YEVJEVICH, V. *Use of partial flood series for estimating distributions of maximum annual flood peak*, Hydrology Paper 82, Colorado State University, Fort Collins, CO, EUA, 1978.

TASKER, G.D. Simplified testing of hydrologic regression regions. *Journal of Hydraulics Division, ASCE*, v. 108, n° 10, p. 1218-1222, 1982.

TODOROVIC, P. Stochastic models of floods. *Water Resources Research*, v. 14, n° 2, p. 345-356, 1978.

TODOROVIC, P. & ZELENHASIC, E. A stochastic model for flood analysis. *Water Resources Research*, v. 6, n° 6, p. 411-424, 1970.

TRYON, R.C. *Cluster Analysis*. Edwards Brothers, Ann Arbor, MI, EUA, 1939, *apud* Statsoft Inc., Electronic Statistics Textbook, Statsoft, Tulsa, OK, Estados Unidos (<http://www.statsoft.com/textbook/stathome.html>), 1997.

TUCCI, C.E. Regionalização de vazões. *In: Hidrologia: Ciência e Aplicação*. Porto Alegre: UFRG/ABRH/EDUSP, p. 573-611, 1993.

VAN MONTFORT, M.A.J. & WITTER, J.V. The generalized Pareto distribution applied to rainfall depths. *Hydrological Sciences Journal*, v. 31, n° 2, p. 151-162, 1986.

WALLIS, J.R. Regional frequency studies using L-moments. *In: IBM Research Report*, New York, IBM Research Division, RC 14597, p. 469-476, mar., 1989.

WARD, J.H. Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, v. 58, p. 236, 1963, *apud* Statsoft Inc., Electronic Statistics Textbook, Statsoft, Tulsa, OK, Estados Unidos (<http://www.statsoft.com/textbook/stathome.html>), 1997.

WAYLEN P.R. & WOO, M.K. Regionalization and prediction of floods in the Fraser river catchment. *Water Resources Bulletin*, v. 20, n° 6, p. 941-949, 1984.

WHITE, E.L. Factor analysis of drainage basin properties: classification of flood behavior in terms of basin geomorphology. *Water Resources Bulletin*, v. 11, n° 4, p. 676-687, 1975.

WILKEN, P.S. *Engenharia de Drenagem Superficial*. São Paulo: CETESB – Companhia de Tecnologia de Saneamento Ambiental, 477p., 1978.

WILTSHIRE, S.E. Identification of homogeneous regions for flood frequency analysis. *Journal of Hydrology*, v. 84, p. 287-302, 1986.

WILTSHIRE, S.E. Grouping basins for regional flood frequency analysis. *Hydrological Sciences Journal*, v. 30, n° 1, p. 151-159, 1985.

ANEXOS

Anexo I

ANÁLISE REGIONAL DE FREQUÊNCIA DE VARIÁVEIS HIDROLÓGICAS E HIDROMETEOROLÓGICAS: SÍNTESE DA METODOLOGIA DOS MOMENTOS-L

1. Introdução à Análise Regional de Frequência

1.1 Preâmbulo

A análise de frequência refere-se à quantificação do número esperado de ocorrências de um evento de certa magnitude. Em se tratando de ocorrências naturais raras, como cheias, estiagens e tempestades extremas, a análise de frequência adquire especial importância devido, principalmente, às consequências desastrosas que esses eventos podem acarretar para a sociedade humana. Em particular, o planejamento, projeto e operação racionais de sistemas de recursos hídricos assentam-se no conhecimento da frequência com que ocorrem os eventos hidrometeorológicos e hidrológicos extremos.

Como decorrência do fato de existirem inúmeras incertezas associadas à quantificação dos processos físicos causais de um evento de dada magnitude, a análise de frequência de variáveis do ciclo hidrológico faz uso intensivo dos métodos da estatística matemática. Esses métodos buscam extrair inferências quanto à *probabilidade* com que uma variável irá igualar ou superar um certo valor ou *quantil*, a partir de um conjunto amostral de ocorrências daquela variável. Se as ocorrências referem-se a observações tomadas unicamente em um ponto específico do espaço geográfico (por exemplo, uma estação pluviométrica, em uma dada bacia hidrográfica), a análise de frequência é dita *local*. Contrari-

amente, se um número maior de observações da variável em questão, tomadas em diferentes pontos de uma certa região, for empregado conjuntamente para a inferência estatística, a análise de frequência é dita *regional*.

A análise local de frequência de variáveis hidrológicas dispõe de um conjunto de técnicas de inferência estatística e de modelos probabilísticos, os quais têm sido objeto freqüente de investigação, visando, principalmente, à obtenção de estimativas cada vez mais eficientes e confiáveis. Entretanto, a inexistência de amostras suficientemente longas impõe um limite superior ao grau de sofisticação estatística a ser empregado na análise local de frequência. Nesse sentido, a análise regional de frequência representa uma alternativa que procura compensar a insuficiente caracterização temporal do comportamento de eventos extremos por uma coerente caracterização espacial da variável hidrológica em questão. Em linhas gerais, a análise regional de frequência utiliza um grande conjunto de dados espacialmente disseminados de uma certa variável, observados em pontos distintos de uma região considerada homogênea do ponto de vista dos processos físicos ali envolvidos, adimensionaliza-os por um fator de escala local e produz a estimativa da curva adimensional de frequência, válida em qualquer ponto da região em estudo. A operação inversa da adimensionalização, ou seja, o produto da curva regional de frequência pelo fator de escala local, permite a obtenção de estimativas de frequência em pontos providos ou desprovidos de observações.

Os princípios da análise regional de frequência são há muito conhecidos e empregados em diversas metodologias de uso corrente em hidrologia [ver, por exemplo, Dalrymple (1960), NERC (1975) e Eletrobrás (1985)]. Entretanto, as subjetividades presentes em algumas etapas das metodologias existentes, bem como o aparecimento de novas técnicas de inferência estatística, como os momentos ponderados por probabilidades (MPP) apresentados por Greenwood *et al.* (1979), motivaram os pesquisadores J.R.M. Hosking, do Centro de Investigações Thomas J. Watson da IBM, e J.R. Wallis, da Universidade Yale, a proporem um conjunto unificado de procedimentos para a análise regional de frequência de diversos tipos de variáveis, com destaque para as hidrológicas, meteorológicas e ambientais. Em sua revisão sobre os avanços recentes da pesquisa na área de análise de frequência, Bobée & Rasmussen (1995) consideram a contribuição de Hosking & Wallis como a mais relevante para a obtenção de melhores estimativas das probabilidades de eventos raros.

O conjunto metodológico mencionado, resultado de diversos desenvolvimentos parciais publicados entre 1980 e 1995, encontra-se descrito em detalhes no livro *Regional Frequency Analysis – An Approach Based on L-Moments* (Hosking & Wallis, 1997), do qual foi extraída grande parte das argumentações e formulações teóricas que se encontram expostas na presente síntese. Em linha gerais, a metodologia descrita por Hosking & Wallis (1997) baseia-se nos princípios do *index-flood*, ou “cheia-índice”, primeiramente enunciados por Dalrymple (1960), e utiliza os momentos-L, quantidades deduzidas dos MPPs e descritas por Hosking (1986; 1990), não só para estimar parâmetros e quantis da distribuição regional de probabilidade, como também para construir estatísticas capazes de tornar menos subjetivas algumas etapas da análise regional de frequência. Neste capítulo, os itens subsequentes procuram apresentar uma panorâmica das etapas da metodologia descrita por Hosking & Wallis (1997). Os capítulos seguintes as abordam de forma mais detalhada.

1.2 Princípios do Método *Index-Flood* (“Cheia-Índice”)

O termo *index-flood* (“cheia-índice”) foi introduzido por Dalrymple (1960), dentro de um contexto de regionalização de vazões de cheia. Trata-se de um expediente para adimensionalizar quaisquer dados obtidos em pontos distintos de uma região considerada homogênea, com a finalidade de utilizá-los como um conjunto amostral único. Apesar de possuir referências a enchentes, o método e o termo *index-flood* têm uso consagrado em estudos de regionalização de frequência de qualquer tipo de variável.

Seja o caso de se regionalizarem as frequências de uma variável genérica X , cuja variabilidade espaço-temporal foi amostrada em N locais ou *postos* de observação, situados em uma certa região

geográfica. As observações, tomadas nos postos indexados por i , formam amostras de tamanho variável n_i e são denotadas por X_{ij} , $i=1, \dots, N$; $j=1, \dots, n_i$. Se F , $0 < F < 1$, representa a distribuição de frequências da variável X no posto i , então, a função de *quantis* nesse local é simbolizada por $X_i(F)$. A hipótese básica do método *index-flood* é a de que os postos formam uma região *homogênea*, ou seja, que as distribuições de frequência nos N pontos são idênticas, à exceção de um *fator de escala local*, denominado *index-flood*. Formalmente,

$$X_i(F) = \mu_i x(F), \quad i = 1, \dots, N \quad (1.1)$$

onde μ_i é o *index-flood* do local i e $x(F)$ representa a *curva regional de quantis adimensionais*, algumas vezes denominada *curva regional de crescimento*, comum a todos os postos.

O fator de escala μ_i pode ser estimado por qualquer medida de posição ou tendência central da amostra de observações $\{X_{i1}, X_{i2}, \dots, X_{in_i}\}$. Por conveniência matemática, Hosking & Wallis (1997) utilizam como estimador do *index-flood* a média aritmética das observações no posto i , ou seja, $\hat{\mu}_i = \bar{X}_i$.

Os *dados adimensionais padronizados* $x_{i,j} = X_{i,j} / \hat{\mu}_i$, $j = 1, \dots, n_i$; $i = 1, \dots, N$ formam a base para se estimar a curva regional de quantis adimensionais $x(F)$. A forma de $x(F)$ é supostamente conhecida, à exceção dos p parâmetros $\theta_1, \dots, \theta_p$, os quais são próprios da distribuição F , e, em geral, funções das características populacionais de posição central, dispersão e assimetria. Hosking & Wallis (1997) propõem que os parâmetros da curva regional de quantis adimensionais, agora denotada por $x(F; \theta_1, \dots, \theta_p)$, sejam os resultados da ponderação dos parâmetros locais $\hat{\theta}_k^{(i)}$, $k = 1, \dots, p$ estimados separadamente para cada posto i . Portanto, a estimativa do parâmetro regional θ_k^R é dada por:

$$\hat{\theta}_k^R = \frac{\sum_{i=1}^N n_i \hat{\theta}_k^{(i)}}{\sum_{i=1}^N n_i} \quad (1.2)$$

Com essas estimativas em $x(F)$, pode-se obter a estimativa da curva regional de quantis adimensionais $\hat{x}(F) = x(F; \hat{\theta}_1^R, \dots, \hat{\theta}_p^R)$. Inversamente, as estimativas dos quantis para o posto i podem ser obtidas pelo produto de $\hat{x}(F)$ por $\hat{\mu}_i$, ou seja:

$$\hat{X}_i(F) = \hat{\mu}_i \hat{x}(F) \quad (1.3)$$

As premissas inerentes ao método *index-flood* são:

- a) as observações em um posto qualquer são identicamente distribuídas;
- b) as observações em um posto qualquer não apresentam dependência estatística serial;
- c) as observações em diferentes postos são estatisticamente independentes;
- d) as distribuições de frequência em diferentes postos são idênticas, a menos de um fator de escala;
- e) a forma matemática da curva regional de quantis adimensionalizados foi corretamente especificada.

Segundo Hosking & Wallis (1997), as premissas (a) e (b) são plausíveis para diversos tipos de variáveis, principalmente aquelas relacionadas a totais ou máximos anuais. Entretanto, é improvável que as três últimas premissas possam ser empiricamente verificadas por dados hidrológicos, meteorológicos ou ambientais. Sabe-se, por exemplo, que precipitações frontais ou estiagens severas são eventos que afetam extensas áreas. Como essas áreas podem conter vários postos de observação da variável em questão, é provável que as amostras, coletadas em pontos distintos, apresentem um grau de correlação significativo. Ainda segundo Hosking & Wallis (1997), na prática, as premissas (d) e (e) jamais são verificadas com exatidão. Apesar dessas restrições, esses autores sugerem que as premissas do método *index-flood* podem ser *razoavelmente aproximadas* tanto pela escolha criteriosa dos postos componentes de uma região, como também pela seleção apropriada de uma distribuição de frequência que apresente consistência com os dados amostrais.

1.3 Etapas da Análise Regional de Frequência

A metodologia para análise regional de frequência, proposta por Hosking & Wallis (1997), fundamenta-se tanto nos princípios enunciados no item 1.2, como também em algumas estatísticas construídas a partir dos chamados momentos-L, cuja descrição é objeto do Capítulo 2, a seguir. Essas estatísticas, a serem detalhadas nos capítulos 3 a 5, constituem instrumentos valiosos para diminuir o grau de subjetividade presente nas quatro etapas usuais da análise regional de frequência. Essas etapas encontram-se sumarizadas a seguir.

Etapa 1: Análise Regional de Consistência de Dados

Essa etapa refere-se à detecção e eliminação de erros grosseiros e/ou sistemáticos eventualmente existentes nas amostras individuais dos vários postos de observação. Além das técnicas usuais de análise de consistência, como as curvas de dupla acumulação, por exemplo, Hosking & Wallis (1997) sugerem o uso de uma estatística auxiliar, denominada *medida de discordância* (ver Capítulo 3), a qual se fundamenta na comparação das características estatísticas do conjunto de postos com as apresentadas pela amostra individual em questão.

Etapa 2: Identificação de Regiões Homogêneas

Uma região homogênea consiste em um agrupamento de postos de observação, cujas curvas de quantis adimensionalizados podem ser aproximadas por uma única curva regional. Para se determinar a correta divisão dos postos em regiões homogêneas, Hosking & Wallis (1997) sugerem o emprego da técnica de análise de *clusters*. De acordo com essa técnica, os postos são agrupados em regiões consonantes com a variabilidade espacial de algumas características locais, estas selecionadas entre aquelas que supostamente podem influir sobre a variável a ser regionalizada. Depois dos postos terem sido convenientemente agrupados em regiões, Hosking & Wallis (1997) sugerem a *medida de heterogeneidade* para testar a correção dos agrupamentos efetuados. Essa medida baseia-se na comparação da variabilidade grupal das características estatísticas dos postos de observação com a variabilidade esperada dessas mesmas características em uma região homogênea. A descrição da técnica de análise de *clusters* e o teste da medida de heterogeneidade são abordados no Capítulo 4 desta síntese.

Etapa 3: Seleção da Função Regional de Distribuição de Probabilidades

Depois de os erros grosseiros e sistemáticos terem sido eliminados das amostras individuais e as regiões homogêneas haverem sido identificadas, a etapa seguinte é a correta prescrição do modelo probabilístico, representado por $x(F)$ na equação 1.3. Para a seleção da função regional de distribuição de probabilidades entre diversos modelos candidatos, Hosking & Wallis (1997) sugerem o emprego do teste da *medida de aderência* (ver Capítulo 5). Esse teste é construído de forma a se poder comparar algumas características estatísticas regionais com aquelas que se espera obter de uma amostra aleatória simples retirada de uma população, cujas propriedades distributivas são as mesmas do modelo candidato.

Etapa 4: Estimação de Parâmetros e Quantis da Função Regional de Distribuição de Probabilidades

Identificado o modelo probabilístico regional, representado por $\hat{x}(F) = x(F; \hat{\theta}_1^R, \dots, \hat{\theta}_p^R)$, os parâmetros locais $\hat{\theta}_k^{(i)}$, $k = 1, \dots, p$ são estimados separadamente para cada posto i e, em seguida, ponderados, conforme a equação 1.2, para produzir a curva regional de quantis adimensionais. Hosking & Wallis (1997) sugerem a utilização dos chamados momentos-L para a estimação de parâmetros e quantis da função regional de distribuição de probabilidades (ver Capítulo 6).

Hosking & Wallis (1997) codificaram um conjunto de rotinas em Linguagem Fortran-77 para automatização das quatro etapas da metodologia proposta para análise regional de frequência. Esse conjunto de rotinas encontra-se disponibilizado ao público no repositório de programas Statlib, acessível via Internet através da URL <http://lib.stat.cmu.edu/general/lmoments>.

2. Momentos-L

Funções de Distribuição de Probabilidades

Seja X uma variável aleatória contínua, cuja função de distribuição de probabilidades acumuladas é dada por:

$$F_X(x) = \mathbf{P}(X \leq x) \quad (2.1)$$

A função densidade de probabilidades, denotada por $f_X(x)$, é definida como a derivada primeira de $F_X(x)$ em relação a X , enquanto $x(p)$ representa a função dos quantis x tal que a probabilidade de a variável não exceder o valor $x(p)$ é igual a p .

O valor esperado, ou esperança matemática da variável aleatória X , denotado por $\mathbf{E}(X)$, é um operador definido por:

$$\mathbf{E}(X) = \int_{-\infty}^{\infty} x f_X(x) dx \quad (2.2)$$

Considerando a transformação $p = F(x)$, pode-se reescrever a equação 2.2 da seguinte forma:

$$\mathbf{E}(X) = \int_0^1 x(p) dp \quad (2.3)$$

Do mesmo modo, a função de variável aleatória $g(X)$ é também uma variável aleatória e sua esperança matemática é dada por:

$$\mathbf{E}[g(x)] = \int_{-\infty}^{\infty} g(x) f_X(x) dx = \int_0^1 g[x(p)] dp \quad (2.4)$$

A variância de X , simbolizada por $\text{var}(X)$, representa uma medida da dispersão dos valores de X em torno do valor central $\mathbf{E}(X)$ e é definida pela seguinte expressão:

$$\text{var}(X) = \mathbf{E}\{[X - \mathbf{E}(X)]^2\} = \mathbf{E}(X^2) - [\mathbf{E}(X)]^2 \quad (2.5)$$

2.1.1 Estimadores

A distribuição da variável aleatória X é completamente conhecida se também o for o conjunto de parâmetros $\theta_1, \theta_2, \dots, \theta_k$, associado à definição das funções $f_X(x; \theta_1, \theta_2, \dots, \theta_k)$ ou $x(p; \theta_1, \theta_2, \dots, \theta_k)$. A maioria das funções de distribuição de probabilidades requer a definição dos parâmetros de posição e de escala. O parâmetro de posição ξ de uma distribuição é o número real que satisfaz:

$$x(p; \xi, \theta_2, \dots, \theta_k) = \xi + x(p; 0, \theta_2, \dots, \theta_k) \quad (2.6)$$

O parâmetro de escala α de uma distribuição, cujo parâmetro de posição é ξ , é dito de escala se:

$$x(p; \xi, \alpha, \theta_3, \dots, \theta_k) = \xi + \alpha x(p; 0, 1, \theta_3, \dots, \theta_k) \quad (2.7)$$

Os parâmetros de uma distribuição devem ser estimados a partir de uma amostra de dados observados. O estimador de um certo parâmetro θ é representado por $\hat{\theta}$, o qual é uma função dos dados amostrais e, portanto, uma variável aleatória. A qualidade do estimador depende de quanto $\hat{\theta}$ desvia-se do verdadeiro valor de θ . Esse desvio pode ser decomposto em um viés e uma variabilidade. O viés representa o desvio sistemático para maior ou para menor, ou seja:

$$\text{viés}(\hat{\theta}) = \mathbf{E}(\hat{\theta} - \theta) \quad (2.8)$$

A variabilidade diz respeito aos desvios aleatórios em relação ao valor populacional de θ . Essa variabilidade pode ser quantificada pela variância do estimador, simbolizada por $\text{var}(\hat{\theta})$. Outra medida que combina o viés e a variabilidade do estimador é dada pela raiz quadrada do erro quadrático médio (REQM) definido por:

$$\text{REQM}(\hat{\theta}) = \sqrt{\mathbf{E}(\hat{\theta} - \theta)^2} = \sqrt{[\text{viés}(\hat{\theta})]^2 + \text{var}(\hat{\theta})} \quad (2.9)$$

Para estimativas com base em amostras de tamanho n , o viés e a variância de $\hat{\theta}$ são assintoticamente proporcionais ao inverso de n . Conseqüentemente, REQM é inversamente proporcional a \sqrt{n} . Como essas quantidades possuem as unidades do parâmetro a ser estimado, Hosking & Wallis (1997) sugerem as razões $\text{var}(\hat{\theta})/\theta$ e $\text{REQM}(\hat{\theta})/\theta$, respectivamente o viés e REQM relativos, como medidas mais convenientes e representativas.

2.1.2 Momentos

As características das distribuições de probabilidades podem ser sumarizadas pelos momentos populacionais. O momento de ordem 1, em relação à origem dos X , representa a média populacional, ou seja:

$$\mu = \mathbf{E}(X) \quad (2.10)$$

Em decorrência da equação 2.3, os momentos centrais de ordem superior podem ser calculados como valores esperados das r -ésimas potências dos desvios da variável em relação ao centro da distribuição μ . Em termos formais:

$$\mu_r = \mathbf{E}(X - \mu)^r ; r = 2,3,\dots \quad (2.11)$$

Alguns momentos centrais de particular interesse são os de ordem 2, 3 e 4. O momento central de ordem 2 é por definição a variância de X , geralmente simbolizada por $\text{var}(X)$ ou σ^2 . As quantidades que podem ser deduzidas do momento central de ordem 2 são o desvio-padrão σ e o coeficiente de variação C_v que são definidas por:

$$\sigma = \sqrt{\mu_2} = \sqrt{\sigma^2} \quad (2.12)$$

$$C_v = \frac{\sigma}{\mu} \quad (2.13)$$

Para $r > 2$, é usual descreverem-se as características da função de distribuição através das razões adimensionais $\mu_r / \mu_2^{r/2}$, das quais se destacam o coeficiente de assimetria:

$$\gamma = \frac{\mu_3}{\mu_2^{3/2}} \quad (2.14)$$

E a curtose:

$$\kappa = \frac{\mu_4}{\mu_2^2} \quad (2.15)$$

Os momentos populacionais podem ser estimados por quantidades similares, calculadas a partir dos dados de uma amostra de tamanho n . O estimador natural de μ é a média aritmética ou momento amostral de 1ª ordem:

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} \quad (2.16)$$

Enquanto os momentos amostrais de ordem superior:

$$m_r = \frac{\sum_{i=1}^n (x_i - \bar{x})^r}{n} \quad (2.17)$$

São estimadores enviesados dos momentos populacionais de mesma ordem. Entretanto, os momentos amostrais m_r podem ser corrigidos para produzirem estimadores sem viés. Por exemplo, as seguintes quantidades são respectivamente os estimadores sem viés da variância e dos coeficientes de variação, assimetria e curtose:

$$\hat{\sigma}^2 = s^2 = \frac{n}{n-1} m_2 \quad (2.18)$$

$$\hat{C}_v = \frac{s}{\bar{x}} \quad (2.19)$$

$$\hat{\gamma} = g = \frac{n^2}{(n-1)(n-2)} \frac{m_3}{s^3} \quad (2.20)$$

$$\hat{\kappa} = k = \frac{n^2}{s^4 (n-2)(n-3)} \left[\left(\frac{n+1}{n-1} \right) m_4 - 3m_2^2 \right] + 3 \quad (2.21)$$

2.2 Momentos-L: Conceitos Básicos

Além de dependentes de n , as estimativas com base em momentos amostrais convencionais envolvem potências sucessivas dos desvios dos dados em relação ao valor central. Em consequência, pequenas amostras tendem a produzir estimativas não confiáveis, particularmente para as funções de momentos de ordem superior como a assimetria e a curtose. Os momentos-L, a serem abordados a seguir, compõem um sistema de medidas estatísticas mais confiáveis para a descrição das características das distribuições de probabilidades.

Os momentos-L são derivados dos "momentos ponderados por probabilidades", ou simplesmente MPPs, os quais foram introduzidos na literatura científica por Greenwood *et al.* (1979). Os MPPs de uma variável aleatória X , variável essa descrita pela função de probabilidades acumuladas $F_X(x)$, são as quantidades definidas por:

$$M_{p,r,s} = \mathbf{E} \left\{ X^p [F_X(x)]^r [1 - F_X(x)]^s \right\} \quad (2.22)$$

Os MPPs $\alpha_r = M_{1,0,r}$ e $\beta_r = M_{1,r,0}$ representam casos especiais de relevância particular para a inferência estatística. Com efeito, considerando uma distribuição cuja função de quantis seja dada por $x(p)$, pode-se combinar as equações 2.4 e 2.22 para expressar α_r e β_r da seguinte forma:

$$\alpha_r = \int_0^1 x(p) (1-p)^r dp \quad , \quad \beta_r = \int_0^1 x(p) p^r dp \quad (2.23)$$

Contrastando as equações acima com a definição de momentos convencionais, ou seja, $\mathbf{E}(X) = \int_0^1 [x(p)]^r dp$, observa-se que esses implicam potências sucessivamente crescentes da função de quantis $x(p)$, enquanto que α_r e β_r implicam potências sucessivamente crescentes de p ou $(1-p)$; dessa forma, os MPPs α_r e β_r podem ser vistos como integrais de $x(p)$, ponderadas pelos polinômios p^r ou $(1-p)^r$.

Diversos autores, como Landwehr *et al.* (1979) e Hosking & Wallis (1986), utilizaram os MPPs α_r e β_r como base para a estimação de parâmetros de distribuições de probabilidades. Hosking & Wallis (1997) ponderam, entretanto, que α_r e β_r são de interpretação difícil, em termos das medidas de escala e forma de uma distribuição de probabilidades, e sugerem, para esse efeito, certas combinações lineares de α_r e β_r . Ainda segundo esses últimos autores, essas combinações advêm da ponderação

das integrais de $x(p)$ por um conjunto de polinômios ortogonais, denotados por $P_r^*(p), r = 0, 1, 2, \dots$, definidos pelas seguintes condições:

(i) $P_r^*(p)$ é um polinômio de grau r em p .

(ii) $P_r^*(1) = 1$

(iii) $\int_0^1 P_r^*(p) P_s^*(p) dp = 0$, para $r \neq s$ (condição de ortogonalidade)

Essas condições definem os polinômios de Legendre, devidamente modificados para a condição de ortogonalidade no intervalo $0 \leq p \leq 1$ e não $-1 \leq p \leq 1$, como em sua formulação original. Formalmente, esses polinômios são dados por:

$$P_r^*(p) = \sum_{k=0}^r l_{r,k}^* p^k \quad (2.24)$$

$$\text{Onde: } l_{r,k}^* = (-1)^{r-k} \binom{r}{k} \binom{r+k}{k} = \frac{(-1)^{r-k} (r+k)!}{(k!)^2 (r-k)!}.$$

De posse dessas definições, os momentos-L de uma variável aleatória X podem ser agora conceituados como sendo as quantidades:

$$\lambda_r = \int_0^1 x(p) P_{r-1}^*(p) dp \quad (2.25)$$

Em termos dos MPPs, os momentos-L são dados por:

$$\lambda_{r+1} = (-1)^r \sum_{k=0}^r l_{r,k}^* \alpha_k = \sum_{k=0}^r l_{r,k}^* \beta_k \quad (2.26)$$

Os primeiros quatro momentos-L são, portanto:

$$\lambda_1 = \alpha_0 = \beta_0 \quad (\text{média ou momento-L de posição}) \quad (2.27)$$

$$\lambda_2 = \alpha_0 - 2\alpha_1 = 2\beta_1 - \beta_0 \quad (\text{momento-L de escala}) \quad (2.28)$$

$$\lambda_3 = \alpha_0 - 6\alpha_1 + 6\alpha_2 = 6\beta_2 - 6\beta_1 + \beta_0 \quad (2.29)$$

$$\lambda_4 = \alpha_0 - 12\alpha_1 + 30\alpha_2 - 20\alpha_3 = 20\beta_3 - 30\beta_2 + 12\beta_1 - \beta_0 \quad (2.30)$$

Em termos de medidas de forma das distribuições, torna-se mais conveniente que os momentos-L sejam expressos em quantidades adimensionais. Estas são representadas pelos quocientes de momentos-L, dados por:

$$\tau_r = \frac{\lambda_r}{\lambda_2}, r = 3, 4, \dots \quad (2.31)$$

Dessa forma, τ_3 e τ_4 são, respectivamente, as medidas de assimetria e curtose, independentes da escala da distribuição de probabilidades. Pode-se definir, também em termos de momentos-L, uma medida análoga ao coeficiente de variação, qual seja:

$$CV - L = \tau = \frac{\lambda_2}{\lambda_1} \quad (2.32)$$

2.2.1 Momentos-L e Estatísticas de Ordem

Os momentos-L podem ser expressos como combinações lineares das estatísticas de ordem de uma amostra. Para esse efeito, considere uma amostra de tamanho n , disposta em ordem crescente $X_{1:n} \leq X_{2:n} \leq \dots \leq X_{n:n}$, de forma que a k -ésima menor observação, ou estatística de ordem k , seja denotada por $X_{k:n}$. De forma consistente com a equação 2.25, os momentos-L da distribuição de probabilidades da qual a amostra foi retirada são dados por:

$$\lambda_r = r^{-1} \sum_{j=0}^{r-1} (-1)^j \binom{r-1}{j} \mathbf{E}(X_{r-j:r}) \quad (2.33)$$

Onde a esperança matemática $\mathbf{E}(\cdot)$ de uma estatística de ordem r é o operador definido por:

$$\mathbf{E}(X_{r:n}) = \frac{n!}{(r-1)!(n-r)!} \int_0^1 x(p) p^{r-1} (1-p)^{n-r} dp \quad (2.34)$$

Dessa forma, os quatro primeiros momentos-L podem ter as seguintes expressões:

$$\lambda_1 = \mathbf{E}(X_{1:1}) \quad (2.35)$$

$$\lambda_2 = \frac{1}{2} \mathbf{E}(X_{2:2} - X_{1:2}) \quad (2.36)$$

$$\lambda_3 = \frac{1}{3} \mathbf{E}(X_{3:3} - 2X_{2:3} + X_{1:3}) \quad (2.37)$$

$$\lambda_4 = \frac{1}{4} \mathbf{E}(X_{4:4} - 3X_{3:4} + 3X_{2:4} - X_{1:4}) \quad (2.38)$$

2.2.2 Propriedades dos Momentos-L

Hosking (1989; 1990) apresenta as provas matemáticas para as seguintes propriedades dos momentos-L:

- Existência: se a média de uma distribuição existe, então todos os momentos-L existem.
- Singularidade: se a média de uma distribuição existe, então os momentos-L a definem singularmente.
- Valores-Limites:

$$-\infty \leq \lambda_1 \leq \infty .$$

$$\lambda_2 \geq 0 .$$

Se a distribuição é definida somente para $X \geq 0 \Rightarrow 0 \leq \tau \leq 1$.

$$|\tau_r| < 1 \text{ para } r \geq 3 .$$

$$\frac{1}{4}(5\tau_3^2 - 1) \leq \tau_4 \leq 1 .$$

Se a distribuição é definida somente para $X \geq 0 \Rightarrow 2\tau - 1 \leq \tau_3 \leq 1$.

- Transformações Lineares: se X e $Y = aX + b$ são duas variáveis aleatórias de momentos-L λ_r e λ_r^* , respectivamente, então são válidas as seguintes relações:

$$\lambda_1^* = a\lambda_1 + b ;$$

$$\lambda_2^* = |a|\lambda_2 ;$$

$$\tau_r^* = (\langle \text{sinal de } a \rangle 1)^r \tau_r , r \geq 3 .$$

- Simetria: se X é uma variável aleatória, descrita por uma distribuição de probabilidades simétrica, então todos os quocientes de momentos-L de ordem ímpar (τ_r , $r=3,5, \dots$) serão nulos.

2.3 Momentos-L de algumas Distribuições de Probabilidades

As propriedades, os momentos-L, os quocientes de momentos-L e os parâmetros de diversas distribuições de probabilidades foram objeto de extenso estudo por Hosking (1986). Essas características encontram-se transcritas a seguir, para algumas das principais distribuições usadas em hidrologia.

Uma dessas distribuições é a Pearson tipo III, a qual pode ser completamente descrita pelos parâmetros μ , de posição, σ , de escala, e γ , de forma. Se $\gamma \neq 0$, essa distribuição pode ser reparametrizada em termos de $\alpha = 4/\gamma^2$, $\beta = \sigma|\gamma|/2$ e $\xi = \mu - 2\sigma/\gamma$. Se $\gamma > 0$, o domínio da variável aleatória é $\xi \leq x < \infty$ e as respectivas funções, densidade e acumulada, são dadas por:

$$f(x) = \frac{(x - \xi)^{\alpha-1} \exp\left(-\frac{x - \xi}{\beta}\right)}{\beta^\alpha \Gamma(\alpha)} \quad (2.39)$$

$$F(x) = \frac{G[\alpha, (x - \xi)/\beta]}{\Gamma(\alpha)} \quad (2.40)$$

Onde $\Gamma(\cdot)$ representa a função gama, definida por $\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt$, e $G(\cdot)$ a função gama incompleta, dada por $G(\alpha, x) = \int_0^x t^{\alpha-1} e^{-t} dt$. Se $\gamma < 0$, o domínio da variável aleatória passa a ser $-\infty \leq x \leq \xi$ e

$$f(x) = \frac{(\xi - x)^{\alpha-1} \exp\left(-\frac{\xi - x}{\beta}\right)}{\beta^\alpha \Gamma(\alpha)} \quad (2.41)$$

$$F(x) = 1 - \frac{G[\alpha, (\xi - x)/\beta]}{\Gamma(\alpha)} \quad (2.42)$$

A distribuição Normal, no domínio $-\infty < x < \infty$, representa um caso particular da Pearson III, quando $\gamma = 0$. Os momentos-L e seus quocientes são dados pelas seguintes expressões:

$$\lambda_1 = \xi + \alpha\beta \quad (2.43)$$

$$\lambda_2 = \beta\Gamma(\alpha + 1/2)/\sqrt{\pi}\Gamma(\alpha) \quad (2.44)$$

$$\tau_3 = 6I_{1/3}(\alpha, 2\alpha) - 3 \quad (2.45)$$

Onde $I_x(p, q)$ representa a razão da função beta incompleta, definida por $I_x(p, q) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} \int_0^x t^{p-1} (1-t)^{q-1} dt$. Não há expressão simples para τ_4 ; o leitor interessado em expressões aproximadas para τ_4 deve consultar Hosking & Wallis (1997, p. 201). Dado α , os parâmetros da distribuição Pearson tipo III são dados por $\gamma = \langle \text{ sinal de } \tau_3 \rangle 2/\sqrt{\alpha}$, $\sigma = \lambda_2 \sqrt{\pi\alpha}\Gamma(\alpha)/\Gamma(\alpha + 1/2)$ e $\mu = \lambda_1$.

As Tabelas 2.1 e 2.2 resumizam as propriedades, os momentos-L, os quocientes de momentos-L e os parâmetros de outras distribuições de probabilidades de 2 e 3 parâmetros, respectivamente.

Tabela 2.1 – Momentos-L para algumas distribuições de 2 parâmetros.

Distribuição	Uniforme	Exponencial	Gumbel	Normal
$f(x)$	$\frac{1}{\beta-\alpha}$	$\frac{\exp\left(-\frac{x-\xi}{\alpha}\right)}{\alpha}$	$\frac{\exp\left(-\frac{x-\xi}{\alpha}\right)\exp\left[-\exp\left(-\frac{x-\xi}{\alpha}\right)\right]}{\alpha}$	$\frac{\phi\left(\frac{x-\mu}{\sigma}\right)}{\sigma}$, onde $\phi(x) = \frac{\exp\left(-\frac{x^2}{2}\right)}{\sqrt{2\pi}}$
$F(x)$	$\frac{x-\alpha}{\beta-\alpha}$	$1-\exp\left(-\frac{x-\xi}{\alpha}\right)$	$\exp\left[-\exp\left(-\frac{x-\xi}{\alpha}\right)\right]$	$\Phi\left(\frac{x-\mu}{\sigma}\right)$, onde $\Phi(x) = \int_{-\infty}^x \phi(t)dt$
$x(F)$	$\alpha+(\beta-\alpha)F$	$\xi-\alpha\ln(1-F)$	$\xi-\alpha\ln(-\ln F)$	não tem forma analítica explícita
λ_1	$\frac{\alpha+\beta}{2}$	$\xi+\alpha$	$\xi+\alpha\gamma$, onde $\gamma = 0,5772\dots$ (c ^{te} de Euler)	μ
λ_2	$\frac{\beta-\alpha}{6}$	$\frac{\alpha}{2}$	$\alpha\ln(2)$	0,5642 σ
τ_3	0	$\frac{1}{3}$	0,1699	0
τ_4	0	$\frac{1}{6}$	0,1504	0,1226
Parâmetro de Posição	-	$\xi = \lambda_1 - \alpha$	$\xi = \lambda_1 - \gamma\alpha$	$\mu = \lambda_1$
Parâmetro de Escala	-	$\alpha = 2\lambda_2$	$\alpha = \frac{\lambda_2}{\ln(2)}$	$\sigma = \sqrt{\pi}\lambda_2$

Fonte: Hosking & Wallis (1997).

Tabela 2.2 – Momentos-L para algumas distribuições de probabilidades de 3 parâmetros.

Distribuição	Generalizada de Pareto	Generalizada de Valores Extremos	Logística Generalizada	Log-Normal
$f(x)$	$\frac{\exp[-(1-k)y]}{\alpha}$, onde $y = -\frac{\ln\left[1 - \frac{k(x-\xi)}{\alpha}\right]}{k}$, $k \neq 0$ $y = \frac{x-\xi}{\alpha}$, $k=0$	$\frac{\exp[-(1-k)y - \exp(-y)]}{\alpha}$, onde $y = -\frac{\ln\left[1 - \frac{k(x-\xi)}{\alpha}\right]}{k}$, $k \neq 0$ $y = \frac{x-\xi}{\alpha}$, $k=0$	$\frac{\exp[-(1-k)y]}{\alpha[1 + \exp(-y)]^2}$, onde $y = -\frac{\ln\left[1 - \frac{k(x-\xi)}{\alpha}\right]}{k}$, $k \neq 0$ $y = \frac{x-\xi}{\alpha}$, $k=0$	$\frac{\exp[ky - y^2/2]}{\alpha\sqrt{2\pi}}$, onde $y = -\frac{\ln\left[1 - \frac{k(x-\xi)}{\alpha}\right]}{k}$, $k \neq 0$ $y = \frac{x-\xi}{\alpha}$, $k=0$
$F(x)$	$1 - \exp(-y)$	$\exp[-\exp(-y)]$	$\frac{1}{1 + \exp(-y)}$	$\Phi(y)$ (ver Tabela 2.1)
$x(F)$	$\xi + \frac{\alpha[1 - (1-F)^k]}{k}$, $k \neq 0$ $\xi - \alpha \ln(1-F)$, $k=0$	$\xi + \frac{\alpha[1 - (-\ln F)^k]}{k}$, $k \neq 0$ $\xi - \alpha \ln(-\ln F)$, $k=0$	$\xi + \frac{\alpha}{k} \left[1 - \left(\frac{1-F}{F}\right)^k\right]$, $k \neq 0$ $\xi - \alpha \ln[(1-F)/F]$, $k=0$	não tem forma analítica explícita
λ_1	$\xi + \frac{\alpha}{1+k}$	$\xi + \frac{\alpha[1 - \Gamma(1+k)]}{k}$	$\xi + \alpha \left[\frac{1}{k} - \frac{\pi}{\text{sen}(k\pi)}\right]$	$\xi + \alpha \frac{1 - \exp(k^2/2)}{k}$
λ_2	$\frac{\alpha}{(1+k)(2+k)}$	$\frac{\alpha(1 - 2^{-k})\Gamma(1+k)}{k}$	$\frac{\alpha k \pi}{\text{sen}(k\pi)}$	$\frac{\alpha e^{k^2/2}}{k} \left[1 - 2\Phi\left(-\frac{k}{\sqrt{2}}\right)\right]$
τ_3	$\frac{1-k}{3+k}$	$\frac{2(1-3^{-k})}{1-2^{-k}} - 3$	-k	$-k \frac{A_0 + A_1 k^2 + A_2 k^4 + A_3 k^6}{1 + B_1 k^2 + B_2 k^4 + B_3 k^6}$
τ_4	$\frac{(1-k)(2-k)}{(3+k)(4+k)}$	$\frac{5(1-4^{-k}) - 10(1-3^{-k}) + 6(1-2^{-k})}{1-2^{-k}}$	$\frac{1+5k^2}{6}$	$\tau_4^0 + k^2 \frac{C_0 + C_1 k^2 + C_2 k^4 + C_3 k^6}{1 + D_1 k^2 + D_2 k^4 + D_3 k^6}$
Parâmetro de Posição	$\xi = \lambda_1 - (2+k)\lambda_2$	$\xi = \lambda_1 - \frac{\alpha[1 - \Gamma(1+k)]}{k}$	$\xi = \lambda_1 - \alpha \left[\frac{1}{k} - \frac{\pi}{\text{sen}(k\pi)}\right]$	$\xi = \lambda_1 - \frac{\alpha}{k} \left(1 - e^{k^2/2}\right)$
Parâmetro de Escala	$\alpha = (1+k)(2+k)\lambda_2$	$\alpha = \frac{\lambda_2 k}{(1-2^{-k})\Gamma(1+k)}$	$\alpha = \frac{\lambda_2 \text{sen}(k\pi)}{k\pi}$	$\alpha = \frac{\lambda_2 k \exp(-k^2/2)}{1 - 2\Phi(-k/\sqrt{2})}$
Parâmetro de Forma	$k = \frac{1-3\tau_3}{1+\tau_3}$	$k \approx 7,8590c + 2,9554c^2$	$k = -\tau_3$	$k = -\tau_3 \frac{E_0 + E_1 \tau_3^2 + E_2 \tau_3^4 + E_3 \tau_3^6}{1 + F_1 \tau_3^2 + F_2 \tau_3^4 + F_3 \tau_3^6}$
Observações	-	$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt$, $c = \frac{2}{3+\tau_3} - \frac{\ln 2}{\ln 3}$	-	Ver Tabela 2.3

Fonte: Hosking & Wallis (1997).

Tabela 2.3 – Coeficientes para o cálculo dos momentos-L da distribuição Log-Normal.

$i \rightarrow$	0	1	2	3
Coeficiente ↓				
τ_4^i	$1,2260172 \cdot 10^{-1}$	-	-	-
A_i	$4,8860251 \cdot 10^{-1}$	$4,4493076 \cdot 10^{-3}$	$8,8027039 \cdot 10^{-4}$	$1,1507084 \cdot 10^{-6}$
B_i	$6,4662924 \cdot 10^{-2}$	$3,3090406 \cdot 10^{-3}$	$7,4290680 \cdot 10^{-5}$	-
C_i	$1,8756590 \cdot 10^{-1}$	$-2,5352147 \cdot 10^{-3}$	$2,6995102 \cdot 10^{-4}$	$-1,8446680 \cdot 10^{-6}$
D_i	-	$8,2325617 \cdot 10^{-2}$	$4,2681448 \cdot 10^{-3}$	$1,1653690 \cdot 10^{-4}$
E_i	2,0466534	-3,6544371	1,8396733	-0,20360244
F_i	-2,0182173	1,2420401	-0,21741801	-

Fonte: Hosking & Wallis (1997).

Um modo conveniente de representação dos momentos-L das diversas distribuições de probabilidades é o diagrama de quocientes de momentos-L, exemplificado pela figura 2.1, onde são lançadas a assimetria-L em ordenadas e a curtose-L em abcissas. Nesse diagrama, uma distribuição de 2 parâmetros (posição e escala) será grafada como um ponto, em decorrência da propriedade das transformações lineares dos momentos-L (ver subitem 2.2.2). Quanto às distribuições de 3 parâmetros (posição, escala e forma), estas serão grafadas como curvas, cujos pontos irão corresponder aos diferentes valores do parâmetro de forma.

2.4 Momentos-L Amostrais

A estimação dos MPPs e momentos-L, a partir de uma amostra finita de tamanho n , inicia-se com a ordenação de seus elementos constituintes em ordem crescente, ou seja, $x_{1:n} \leq x_{2:n} \leq \dots \leq x_{n:n}$. Um estimador não-enviesado do MPP β_r pode ser escrito como:

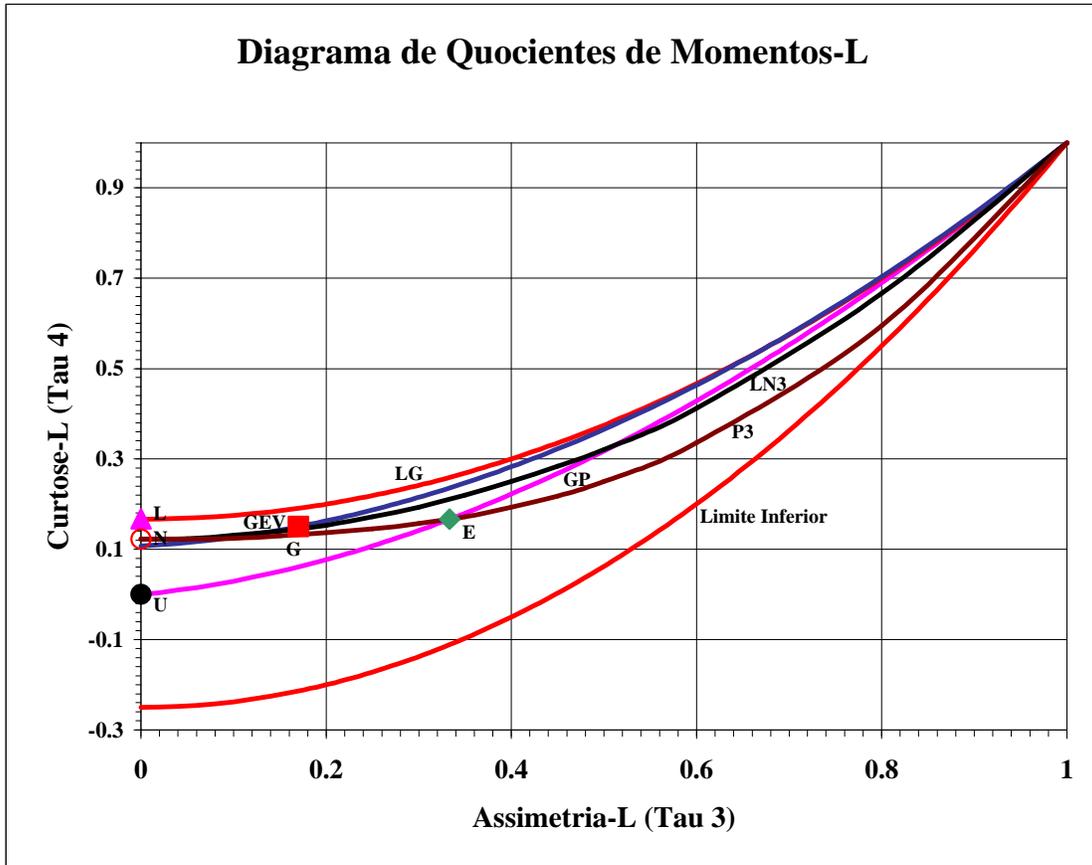
$$b_r = \hat{\beta}_r = \frac{1}{n} \sum_{j=r+1}^n \frac{(j-1)(j-2)\dots(j-r)}{(n-1)(n-2)\dots(n-r)} x_{j:n} \quad (2.46)$$

Dessa forma, os estimadores de β_r , $r \leq 2$ são dados por:

$$b_0 = \frac{1}{n} \sum_{j=1}^n x_{j:n} \quad (2.47)$$

$$b_1 = \frac{1}{n} \sum_{j=2}^n \frac{(j-1)}{(n-1)} x_{j:n} \quad (2.48)$$

$$b_2 = \frac{1}{n} \sum_{j=3}^n \frac{(j-1)(j-2)}{(n-1)(n-2)} x_{j:n} \quad (2.49)$$



(L: Logística; LG: Logística Generalizada; G: Gumbel; GEV: Generalizada de Valores Extremos; N: Normal; LN3: Log-Normal 3 Parâmetros; P3: Pearson Tipo III; E: Exponencial; GP: Generalizada de Pareto; U: Uniforme)

Figura 2.1 – Diagrama de momentos-L para algumas distribuições de probabilidades.

Analogamente às equações 2.27 a 2.30, os estimadores não-enviesados de λ_r são os momentos-L amostrais, estes definidos pelas seguintes expressões:

$$\ell_1 = b_0 \tag{2.50}$$

$$\ell_2 = 2b_1 - b_0 \tag{2.51}$$

$$\ell_3 = 6b_2 - 6b_1 + b_0 \tag{2.52}$$

$$\ell_4 = 20b_3 - 30b_2 + 12b_1 - b_0 \tag{2.53}$$

$$\ell_{r+1} = \sum_{k=0}^r l_{r,k}^* b_k ; \quad r = 0, 1, \dots, n-1 \tag{2.54}$$

Na equação 2.54, os coeficientes $l_{r,k}^*$ são definidos tal como na equação 2.24. Da mesma forma, os quocientes de momentos-L amostrais são dados por:

$$t_r = \frac{\ell_r}{\ell_2}; \quad r \geq 3 \quad (2.55)$$

Enquanto o CV-L amostral calcula-se através de:

$$t = \frac{\ell_2}{\ell_1} \quad (2.56)$$

Os estimadores de τ_r , fornecidos pelas equações 2.55 e 2.56, são muito pouco enviesados quando calculados para amostras de tamanho moderado a grande. Hosking (1990, pág. 116) utilizou a teoria assintótica para calcular o viés para amostras grandes; para a distribuição Gumbel, por exemplo, o viés assintótico de t_3 é $0,19n^{-1}$, enquanto o de t_4 , para a distribuição Normal, é $0,03n^{-1}$, onde n representa o tamanho da amostra. Para amostras de pequeno tamanho, o viés pode ser avaliado por simulação. Segundo Hosking & Wallis (1997, pág. 28) e para uma gama variada de distribuições, o viés de t pode ser considerado desprezível para $n \geq 20$. Ainda segundo esses autores, mesmo em se tratando de amostras de tamanho em torno de 20, o viés de t_3 e o viés de t_4 são considerados relativamente pequenos e definitivamente menores do que os produzidos por estimadores convencionais de assimetria e curtose.

2.5 Estimação de Parâmetros através de Momentos-L

Um problema clássico da inferência estatística refere-se à estimação, a partir de uma amostra de tamanho n , dos $\theta_1, \theta_2, \dots, \theta_p$ parâmetros que especificam uma determinada distribuição de probabilidades. Equivalentemente ao chamado método dos momentos, o método dos momentos-L consiste em se obter as estimativas dos parâmetros igualando-se os primeiros p momentos-L amostrais aos seus correspondentes populacionais; esse procedimento irá resultar em um sistema de p equações e p incógnitas, cujas soluções pressupõem que os parâmetros sejam expressões de momentos-L. Essas expressões encontram-se sumarizadas nas tabelas 2.1 e 2.2 para algumas distribuições de 2 e 3 parâmetros, respectivamente.

Hosking & Wallis (1997) mostram que os estimadores de parâmetros e quantis, obtidos por momentos-L para as distribuições mais comumente utilizadas, são assintoticamente distribuídos como uma distribuição Normal, a partir da qual podem ser calculados erros-padrão das estimativas e intervalos de confiança. Além disso, mostram que, para amostras de tamanho pequeno a moderado, o método dos momentos-L é geralmente mais eficiente do que o da máxima verossimilhança.

3 Análise Regional de Consistência de Dados

A primeira etapa da análise regional de frequência de variáveis aleatórias é certificar-se (i) de que os dados coletados em qualquer dos postos de observação estão isentos de erros grosseiros e (ii) de que todos os dados individuais provêm de uma mesma distribuição de frequências.

No caso de dados hidrológicos ou hidrometeorológicos, os erros grosseiros devem-se principalmente à leitura, à transcrição ou ao processamento incorretos. São muito frequentes nas leituras lini-

métricas e pluviométricas, nas quais a intervenção humana é mais presente e, em consequência, a probabilidade de erro é maior. Em alguns casos, a identificação e eliminação dos erros grosseiros presentes nas séries hidrológicas/hidrometeorológicas não são tarefas de fácil execução.

Quando são alteradas as circunstâncias (localização, regime, equipamento de medição) sob as quais os dados são coletados, as séries hidrológicas/hidrometeorológicas podem vir a apresentar tendências e não-estacionariedade. Nesses casos, a distribuição de frequência dos dados coletados passa a não ser constante no tempo e a série hidrológica/hidrometeorológica, como uma amostra única, não pode ser considerada *homogênea* e nem utilizada para a inferência estatística. São exemplos pertinentes: (a) a relocação de um posto pluviométrico para local com características de vento muito diferentes daquelas apresentadas na instalação de origem; (b) a alteração do regime hidrológico causada pela implantação de reservatório de acumulação a montante de um posto fluviométrico; (c) a utilização de equipamentos não aferidos, defeituosos ou incompatíveis com a sistemática-padrão de coleta de dados primários.

As técnicas mais usuais para a identificação de erros e heterogeneidades nas séries hidrológicas/hidrometeorológicas são:

- a) comparação de cotogramas e/ou fluviogramas de postos fluviométricos próximos;
- b) comparação entre totais mensais de precipitação entre postos pluviométricos próximos ou entre um posto e a média de postos vizinhos;
- c) curvas de dupla acumulação de séries mensais/anuais do posto em questão e do “padrão regional”, este tomado como a média de vários postos das proximidades;
- d) testes estatísticos convencionais para verificação de independência, homogeneidade e pontos atípicos (Spearman, Mann-Whitney, Grubbs-Beck dentre outros).

Além dessas técnicas de uso corrente em hidrologia, Hosking & Wallis (1997) sugerem também a comparação entre os quocientes de momentos-L amostrais calculados para os diferentes postos de observação. Segundo esses autores, os quocientes de momentos-L amostrais são capazes de refletir erros, pontos atípicos e heterogeneidades eventualmente presentes em uma série de observações. Isso pode ser efetuado através de uma estatística-sumário, a qual representa a *medida da discordância* entre os quocientes de momentos-L amostrais de um dado local e a média dos quocientes de momentos-L dos vários postos da região.

3.1 A Medida de Discordância

3.1.1 Descrição

Em um grupo de amostras, a *medida de discordância* tem por objetivo identificar aquelas que apresentam características estatísticas muito discrepantes das grupais. A medida de discordância é expressa como uma estatística única envolvendo as estimativas dos quocientes de momentos-L CV-L (ou τ), assimetria-L (ou τ_3) e a curtose-L (ou τ_4). Em um espaço tridimensional de variação desses quocientes de momentos-L, a idéia é assinalar como discordantes as amostras cujos valores $\{\hat{\tau}, \hat{\tau}_3, \hat{\tau}_4\}$, representados por um ponto no espaço, afastam-se “demasiadamente” do núcleo de concentração das amostras do grupo. Para melhor visualização do significado dessa estatística, considere o plano definido pelos limites de variação das estimativas do CV-L e da assimetria-L para diversos postos de observação de uma região geográfica (figura 3.1). Nessa figura, as médias grupais encontram-se no ponto assinalado pelo símbolo +, em torno do qual constroem-se elipses concêntricas cujos eixos maiores e menores são funções da matriz de covariância amostral dos quocientes de momentos-L. Os pontos considerados discordantes são aqueles que se encontram fora da área definida pela elipse mais externa.

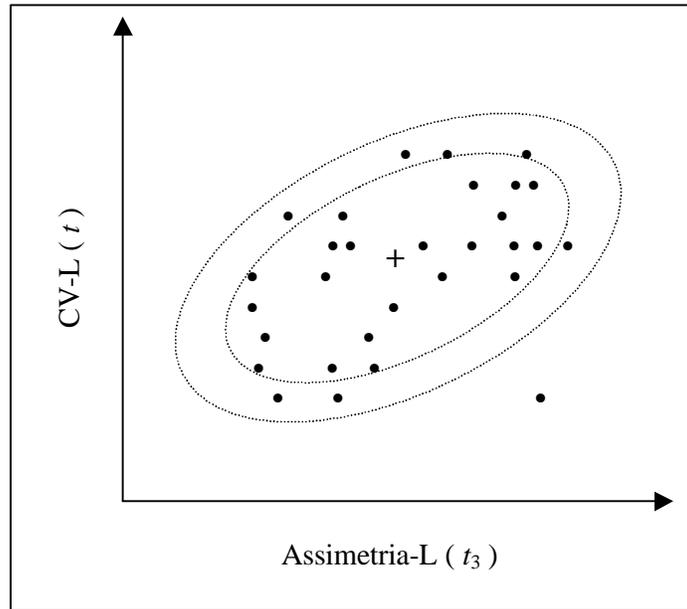


Figura 3.1 – Descrição esquemática da medida de discordância.

3.1.2 Definição Formal

Os quocientes de momentos-L de um local i , a saber CV-L, assimetria-L e curtose-L, são considerados como um ponto em um espaço tridimensional. Em termos formais, considere que u_i representa um vetor (3x1) contendo esses quocientes de momentos-L, dado por:

$$u_i = (t^i \ t_3^i \ t_4^i)^T \quad (3.1)$$

Onde t , t_3 e t_4 denotam CV-L, assimetria-L e curtose-L, respectivamente, e o símbolo T indica matriz transposta. Seja \bar{u} um vetor (3x1), contendo a média grupal ou regional dos quocientes de momentos-L, tomada como a média aritmética simples de u_i para todos os postos estudados, ou seja:

$$\bar{u} = \frac{\sum_{i=1}^N u_i}{N} = (t^R \ t_3^R \ t_4^R)^T \quad (3.2)$$

Onde N representa o número de postos de observação do grupo ou região R em questão. Dada a matriz de covariância amostral S , definida por:

$$S = (N-1)^{-1} \sum_{i=1}^N (u_i - \bar{u})(u_i - \bar{u})^T \quad (3.3)$$

Hosking & Wallis (1995) definem a medida de discordância Di , para o local i pela expressão:

$$Di = \frac{N}{3(N-1)} (u_i - \bar{u})^T S^{-1} (u_i - \bar{u}) \quad (3.4)$$

Em trabalhos anteriores, Hosking & Wallis (1993) sugeriram o valor-limite $Di = 3$ como critério para decidir se a amostra é discordante das características grupais. Por exemplo, quando uma certa amostra produz $Di \geq 3$, isso significa que ela pode conter erros grosseiros e/ou sistemáticos, ou mesmo pontos atípicos, que a tornam discordante ou discrepante das demais do grupo de amostras. Posteriormente, Hosking & Wallis (1995) apresentaram novos valores críticos para Di , para grupos ou regiões com menos de 15 postos de observação. Esses valores críticos para Di encontram-se listados na tabela 3.1.

Tabela 3.1 – Valores críticos da medida de discordância – Di .

Nº de Postos da Região	Di_{crit}
5	1,333
6	1,648
7	1,917
8	2,140
9	2,329
10	2,491
11	2,632
12	2,757
13	2,869
14	2,971
≥ 5	3

Fonte: Hosking & Wallis (1995).

De acordo com Hosking & Wallis (1995), para grupos ou regiões com número muito reduzido de postos de observação, a estatística Di não é informativa. Por exemplo, para $N \leq 3$, a matriz de covariância S é singular e o valor de Di não pode ser calculado. Para $N = 4$, $Di = 1$ e, para $N = 5$ ou $N = 6$, os valores de Di , como indicados na tabela 3.1, são bastante próximos do limite algébrico da estatística, definido por $Di \leq (N - 1)/3$. Em consequência, os autores sugerem o uso da medida de discordância Di somente para $N \geq 7$.

3.2 Discussão

Hosking & Wallis (1997) fazem as seguintes recomendações para o uso da medida de discordância Di :

- a) A análise regional de consistência de dados inicia-se com o cálculo das Di s individuais de todos os postos de uma grande região geográfica, sem considerações preliminares relativas à homogeneidade regional. Aqueles postos assinalados como discordantes devem ser submetidos a cuidadosa análise individual (testes estatísticos, curva de dupla acumulação, comparação com postos vizinhos), visando à identificação/eliminação de eventuais inconsistências em seus dados.

- b) Em seguida, quando a homogeneidade regional (ver Capítulo 4) já houver sido definida, as medidas de discordância devem ser recalculadas, desta feita com os postos devidamente agrupados em suas respectivas regiões homogêneas. Se um certo posto se apresentar discordante em uma região, deve ser considerada a possibilidade de sua transferência para outra.
- c) Ao longo de toda a análise regional de consistência de dados, deve-se ter em conta que os quocientes de momentos-L amostrais podem apresentar diferenças naturalmente possíveis, mesmo entre postos similares do ponto de vista dos processos físicos em questão. Hosking & Wallis (1997) exemplificam que um evento extremo, porém localizado, pode ter afetado somente alguns postos em uma região. Entretanto, se é provável que um evento como esse pode afetar qualquer posto da região, então a providência mais sensata seria a de tratar todo o grupo de postos como uma única região homogênea, mesmo que alguns possam apresentar medidas de discordância superiores aos valores-limites estabelecidos.

4 Identificação e Delimitação de Regiões Homogêneas

Das etapas que compõem a análise regional de frequência de variáveis aleatórias, a identificação e delimitação de regiões homogêneas é considerada a mais difícil e mais sujeita a subjetividades. Uma região é homogênea se existem evidências suficientes de que as diferentes amostras do grupo possuem a mesma distribuição de frequências, a menos, é claro, do fator de escala local. Potter (1987) considera que essa etapa é crucial por exigir do analista e da metodologia empregada a capacidade de discernir se observações anômalas, eventualmente existentes em uma ou mais amostras do grupo, devem-se a diferenças populacionais em relação ao modelo probabilístico proposto ou a meras flutuações amostrais. Embora diversas técnicas tenham sido propostas para a identificação e delimitação de regiões homogêneas, nenhuma delas constitui um critério estritamente objetivo ou uma solução consensual para o problema. De fato, Bobée & Rasmussen (1995) reconhecem que, por si, a análise regional de frequência e, em particular a delimitação de regiões homogêneas, são construídas com base em premissas difíceis de serem tratadas com rigor matemático. Concluem enfatizando que esse fato deve ser visto como um desafio a ser vencido por futuras investigações pertinentes à área de análise de frequência.

Uma primeira fonte de controvérsias quanto à correta abordagem para a identificação de regiões homogêneas diz respeito ao tipo de dado local a ser utilizado. Faz-se distinção entre *estatísticas locais* e *características locais*. As estatísticas locais referem-se, por exemplo, a estimadores das medidas de dispersão e assimetria, tais como CV-L e assimetria-L, calculados diretamente a partir das amostras dos dados objetos da análise regional de frequência. Por outro lado, as características locais são, em princípio, quantidades previamente conhecidas e não dedutíveis ou estimadas a partir das amostras pontuais. Como exemplos de características locais para o caso de variáveis hidrológicas/hidrometeorológicas, podem ser citadas a latitude, a longitude, a altitude e outras propriedades físicas relacionadas a um certo local específico. Podem ser incluídas também outras características indiretamente relacionadas à amostra, tais como a altura média de precipitação anual, o mês mais freqüente de ocorrência de cheias ou o volume médio anual do escoamento-base. Alguns autores, nominalmente Wiltshire (1986), Burn (1989) e Pearson (1991), propuseram técnicas que fazem uso somente das estatísticas locais para definir regiões homogêneas de vazões de enchentes na Inglaterra, Estados Unidos e Nova Zelândia, respectivamente. Contrariamente, Hosking & Wallis (1997) recomendam que a identificação de regiões homogêneas se faça em duas etapas consecutivas: a primeira consistindo em uma delimitação preliminar baseada unicamente nas *características* locais e a segunda consistindo num teste estatístico, construído com base somente nas *estatísticas* locais, cujo objetivo é o de verificar os resultados preliminarmente obtidos.

De fato, dentro da construção proposta por Hosking & Wallis (1997), tratar-se-ia de um raciocínio circular usar os mesmos dados tanto para identificar as regiões como para testar a sua correção. Além, evidentemente, de agregar novas informações independentes, o processo de identificação de regiões homogê-

neas em duas etapas, tal como recomendado por Hosking & Wallis (1997), é reforçado por outros argumentos. Tome-se, por exemplo, o caso em que uma estatística, como o CV-L local, é empregada como critério único para agrupar as amostras e identificar regiões homogêneas. Nesse contexto, existirá sempre uma tendência de agrupar aquelas amostras com valores atípicos (*outliers*) altos (conseqüentemente, com elevadas estimativas locais de CV-L), muito embora esses *outliers* possam dever-se a meras flutuações de uma amostra, as quais podem não estar presentes em locais vizinhos.

A identificação de regiões homogêneas em duas etapas, proposta por Hosking & Wallis (1997), encontra-se sintetizada a seguir. Inicialmente, são apresentados alguns dos métodos existentes para identificação preliminar de regiões homogêneas, seguidos de uma descrição mais detalhada da técnica de *clusters*. Na seqüência, apresenta-se o teste estatístico, materializado pela *medida de heterogeneidade*, e construído com base nos quocientes de momentos-L amostrais.

4.1 Identificação Preliminar de Regiões Homogêneas: Métodos Existentes

De acordo com Hosking & Wallis (1997), os diversos métodos e técnicas de agrupamento de locais similares em regiões homogêneas podem ser categorizados como se segue.

- Conveniência Geográfica

Dentro dessa categoria, encontram-se todas as experiências de identificação de regiões homogêneas que se baseiam no agrupamento subjetivo e/ou conveniente dos postos de observação, geralmente contíguos, em áreas administrativas ou em zonas previamente definidas segundo limites arbitrários. Dentre os inúmeros trabalhos que fizeram uso da conveniência geográfica, podem ser citadas as regionalizações de vazões de enchentes das Ilhas Britânicas (NERC, 1975) e da Austrália (Institution of Engineers Australia, 1987).

- Agrupamento Subjetivo

Nessa categoria, a delimitação subjetiva das regiões homogêneas 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. Schaefer (1990), por exemplo, utilizou alturas similares de precipitação anual para delimitar regiões homogêneas de chuvas máximas anuais no estado americano de Washington. Da mesma forma, Pinto & Naghettini (1999) utilizaram de modo combinado as conformações de relevo, clima e isoietas anuais para a delimitação preliminar de regiões homogêneas de alturas diárias de chuva máximas anuais na bacia do alto rio São Francisco. Embora um grau considerável de subjetividade esteja presente nessas experiências, os seus resultados podem ser objetivamente verificados através do teste estatístico da medida de heterogeneidade, a ser descrito no item 4.3.

- Agrupamento Objetivo

Nesse caso, as regiões são formadas pelo agrupamento dos postos de observação em um ou mais conjuntos, de modo que uma dada estatística não exceda um valor limiar previamente selecionado. Esse valor limiar é arbitrado de forma a minimizar critérios variados de heterogeneidade. Por exemplo, Wiltshire (1985) utilizou como critério a razão de verossimilhança e, posteriormente, Wiltshire (1986) e Pearson (1991) empregaram as variabilidades intragrupos de estatísticas locais como os coeficientes de variação e assimetria. Na seqüência, os grupos são subdivididos iterativamente até que se satisfaça o critério de homogeneidade proposto. Hosking & Wallis (1997) apontam como uma desvantagem dessa técnica o fato de que as iterações sucessivas de reagrupamento dos postos de observação nem sempre conduzem a uma solução final otimizada. Apontam também para o fato de que as estatísticas intragrupos empregadas podem ser influenciadas, em grau indeterminado, pela eventual existência de dependência estatística entre as amostras consideradas.

- Análise de Clusters

Trata-se de um método usual de análise estatística multivariada, no qual se associa a cada posto um vetor de dados contendo as características e/ou estatísticas locais. Em seguida, os postos são agrupados e reagrupados de forma que seja possível identificar a maior ou menor similaridade entre os seus vetores de dados. Hosking & Wallis (1997) citam diversos estudos (Burn, 1989 e Guttman, 1993, dentre outros), nos quais a análise de *clusters* foi empregada com sucesso para a regionalização de freqüências de precipitação, vazões de enchentes e outras variáveis. Esses autores consideram a análise de *clusters* como o método mais prático, porém ainda sujeito a subjetividades, para a identificação preliminar de regiões homogêneas. Por constituir-se em um método preferencial, apresentam-se, no item 4.2, uma descrição da técnica de análise de *clusters* e recomendações para o seu emprego na identificação preliminar de regiões homogêneas.

- Outros Métodos

Além dos mencionados anteriormente, outros métodos têm sido empregados para a identificação e delimitação de regiões homogêneas. No contexto de variáveis hidrológicas/hidrometeorológicas, podem ser citados os seguintes exemplos: (a) análise de resíduos de regressão (Tasker, 1982); (b) análise de componentes principais (Nathan & McMahon, 1990); (c) análise fatorial (White, 1975); (d) correlação canônica (Cavadias, 1990); (e) análise de discriminante (Waylen & Woo, 1984); (f) análise de formas das funções densidade de probabilidade (Gingras & Adamowski, 1993). Da mesma forma que os anteriores, esses métodos também apresentam elementos subjetivos e limitações.

4.2 Análise de *Clusters* – Noções

O termo “análise de *clusters*” foi empregado pela primeira vez por Tryon (1939) e engloba um grande número de diferentes algoritmos de classificação em grupos ou taxonomias estruturalmente similares. Essencialmente, a análise de *clusters* é a aglomeração seqüencial de indivíduos a grupos cada vez maiores, de acordo com algum critério, distância ou medida de dissimilaridade. Um indivíduo pode ter diversos atributos ou características, os quais são quantificados e representados pelo vetor de dados locais $\{Z_1, Z_2, \dots, Z_p\}$. As medidas ou distâncias de dissimilaridade entre dois indivíduos devem ser representativas da variação mútua das características locais em um espaço p -dimensional. A medida mais usada é a *distância Euclidiana generalizada*, a qual é simplesmente a distância geométrica tomada em um espaço de p dimensões. Por exemplo, a distância Euclidiana entre dois indivíduos i e j é dada por:

$$d_{ij} = \sqrt{\sum_{k=1}^p (Z_{ik} - Z_{jk})^2} \quad (4.1)$$

Para efeito de entendimento da lógica inerente à análise de *clusters*, tomemos um de seus métodos de aglomeração mais simples, que é conhecido como o do *vizinho mais próximo*. A aglomeração em *clusters* inicia-se pelo cálculo das distâncias d entre um certo indivíduo e todos os outros do grupo, para cada um deles. Inicialmente, existem tantos grupos quanto numerosos forem os indivíduos. O primeiro *cluster* se forma com o par de indivíduos mais próximos (ou de menor distância Euclidiana); se a distância para outros indivíduos for a mesma da anterior, estes também farão parte do *cluster*. Em seguida, forma-se o *cluster* seguinte com o par (ou grupo, ou *cluster*) de menor distância Euclidiana e assim sucessivamente até que, ao final, todos os indivíduos estejam aglomerados. Considere o exemplo hipotético da figura 4.1, no qual dez indivíduos, assinalados em abcissas, tiveram calculadas suas distâncias Euclidianas, em ordenadas, de acordo com um certo número de atributos. Se forem consi-

derados somente dois *clusters*, o primeiro seria formado pelo indivíduo 1 e o segundo pelos nove indivíduos restantes. Na seqüência, o segundo *cluster* poderia ser dividido em dois: um, formado pelos indivíduos 8, 9 e 10, enquanto o outro o seria pelos indivíduos restantes; dessa forma, teríamos um total de três *clusters*. Se agora seis *clusters* são necessários, então os indivíduos 1 a 4 formariam quatro *clusters* e os seis indivíduos restantes se agrupariam tal como se apresenta no dendograma da figura 4.1. Dessa maneira, pode-se ler em ordenadas a distância em que os indivíduos se aglomeram para formar um *cluster* e pode-se, através das distintas ramificações do dendograma, interpretar a estrutura de similaridade dos dados.

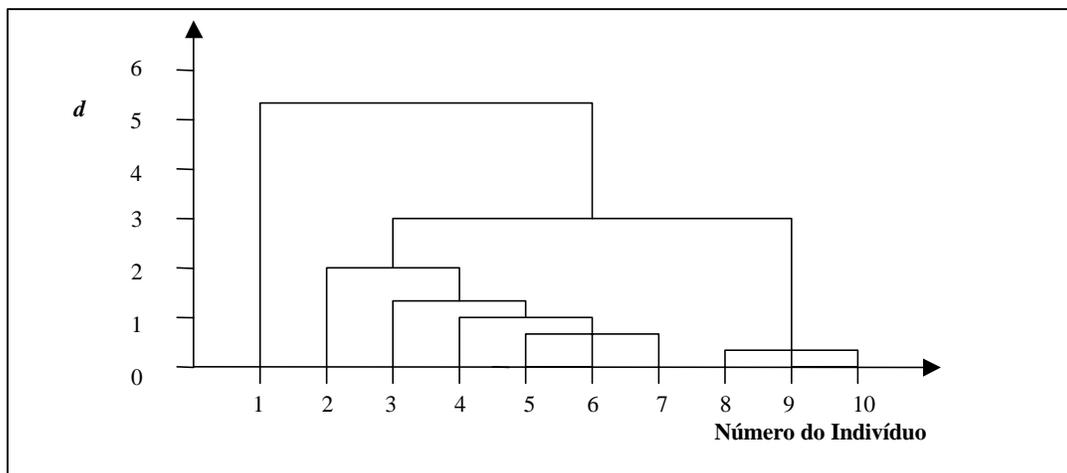


Figura 4.1 – Dendrograma hipotético – 10 indivíduos (adap. de Kottegodda & Rosso, 1997).

Inicialmente, quando cada indivíduo constitui o seu próprio *cluster*, as distâncias entre indivíduos são definidas por d , tal como calculado pela equação 4.1. Entretanto, a partir do momento em que vários indivíduos formam um ou mais *clusters*, põe-se a questão de como serão determinadas as distâncias de dissimilaridade entre esses novos *clusters*. Em outras palavras, faz-se necessária uma *regra de aglomeração* para definir quando dois *clusters* são suficientemente similares para se juntarem. Uma das várias possibilidades para se definir essa regra foi exemplificada na figura 4.1; nesse caso, usou-se o critério do *vizinho mais próximo*, segundo o qual, a distância entre dois *clusters* é determinada pela distância entre os seus dois respectivos indivíduos que mais se aproximam. Esse critério pode conduzir à formação de extensos *clusters* que se aglomeram meramente porque contêm indivíduos próximos.

Um método alternativo e muito utilizado como regra de aglomeração é o descrito por Ward (1963). Em linhas gerais, o método de Ward emprega a análise de variância para determinar as distâncias entre *clusters* e aglomerá-los de forma a minimizar a soma dos quadrados de quaisquer pares de dois *clusters* hipotéticos, a cada iteração. O método de Ward é considerado como eficiente e, em geral, tende a produzir *clusters* pouco extensos e de igual número de indivíduos.

Outro método muito empregado é o devido a Hartigan (1975) e conhecido como o das *K-médias* (*K-means clustering*). O princípio desse método é o de que o analista pode, *a priori*, ter indícios ou hipóteses relativas ao número "correto" de *clusters* a ser considerado. Dessa forma, o método das *K-médias* irá produzir K *clusters*, os quais deverão ser os mais distintos entre si. Para fazê-lo, o método começa com a formação de K *clusters* iniciais, cujos membros são escolhidos aleatoriamente entre os indivíduos a serem agrupados. Em seguida, os indivíduos são movidos iterativamente de um *cluster* para outro, de forma a (1) minimizar a variabilidade intra*cluster* e (2) maximizar a variabilidade entre os *clusters*. Essa lógica é análoga a se proceder a uma análise de variância ao revés, no sentido de que, ao testar a hipótese nula de que as médias grupais são diferentes entre si, a análise de variância con-

fronta a variabilidade entre grupos com a variabilidade intragrupos. Em geral, os resultados do método das K -médias devem ser examinados de forma a se avaliar quão distintas são as médias dos K *clusters* obtidos.

Quando aplicada à identificação preliminar de regiões homogêneas para estudos regionais de frequência de variáveis hidrológicas/hidrometeorológicas, a análise de *clusters* requer algumas considerações específicas. Hosking & Wallis (1997) recomendam atenção para os seguintes pontos:

1. Muitos algoritmos para a aglomeração em *clusters* utilizam o recíproco da distância Euclidiana como medida de similaridade. Nesse caso, é usual padronizar os elementos do vetor das características, dividindo-os pela sua amplitude ou desvio-padrão, de forma que passem a ter aproximadamente a mesma variabilidade. Essa padronização implica atribuir ponderações iguais às diferentes características locais, o que pode ocultar a maior ou menor influência relativa de uma delas na forma da curva regional de frequências. Pode-se compensar essa deficiência pela atribuição direta de diferentes ponderações às características locais consideradas.
2. Os métodos como o das K -médias requerem a definição do número de *clusters* a se considerar; objetivamente, não se tem, *a priori*, o número "correto" de *clusters*. Na prática, deve-se buscar um equilíbrio entre regiões demasiadamente grandes ou demasiadamente pequenas, com muitos ou poucos postos de observação. Para as metodologias de análise regional de frequências que utilizam o princípio do *index-flood*, existe muito pouca vantagem em se empregar regiões muito extensas. Segundo Hosking & Wallis (1997), ganha-se muito pouca precisão nas estimativas de quantis ao se usar mais de 20 postos em uma região. Portanto, não há razão premente para se juntar regiões extensas cujas estimativas das distribuições de frequências são similares.
3. Os resultados da análise de *clusters* devem ser considerados como preliminares. Em geral, são necessários ajustes, muitas vezes subjetivos, cuja finalidade é a de tornar fisicamente coerente a delimitação das regiões, assim como a de reduzir a *medida de heterogeneidade* a ser descrita no item 4.3, a seguir. Os ajustes mencionados podem ser obtidos pelas seguintes providências:
 - mover um ou mais postos de uma região para outra;
 - desconsiderar ou remover um ou mais postos;
 - subdividir uma região;
 - abandonar uma região e realocar os seus postos para outras regiões;
 - combinar uma região com outra ou outras;
 - combinar duas ou mais regiões e redefini-las;
 - obter mais dados e redefinir as regiões.

4.3 Medida de Heterogeneidade Regional

4.3.1 Descrição

Em uma região homogênea, todos os indivíduos possuem os mesmos quocientes de momentos-L populacionais. Entretanto, as suas estimativas, ou sejam, os quocientes de momentos-L calculados a partir das amostras, serão diferentes devido às flutuações amostrais. Portanto, é natural questionar se a dispersão dos quocientes de momentos-L amostrais, calculados para um certo conjunto de postos, é maior do que aquela que se esperaria encontrar em uma região homogênea. Essencialmente, é essa a lógica empregada para a construção da medida de heterogeneidade regional.

Pode-se visualizar o significado da medida de heterogeneidade através de diagramas de quocientes de momentos-L, como o da figura 4.2. Embora também se possam usar outras estatísticas, no exemplo hipotético dessa figura encontram-se grafados o CV-L e a assimetria-L amostrais de um lado, enquanto que do outro estão os seus correspondentes tais como obtidos a partir de simulações de amostras de mesmo tamanho das originais, localizadas, por hipótese, em uma região homogênea. Em diagramas como esses, uma região possivelmente heterogênea apresentaria, por exemplo, CV-Ls amostrais muito mais dispersos do que aqueles obtidos por simulação. Em termos quantitativos, essa idéia básica pode ser traduzida pela diferença relativa centrada entre as dispersões, observada e simulada, ou seja, pela razão $\frac{(\text{dispersão observada}) - (\text{média das simulações})}{\text{desvio - padrão das simulações}}$.

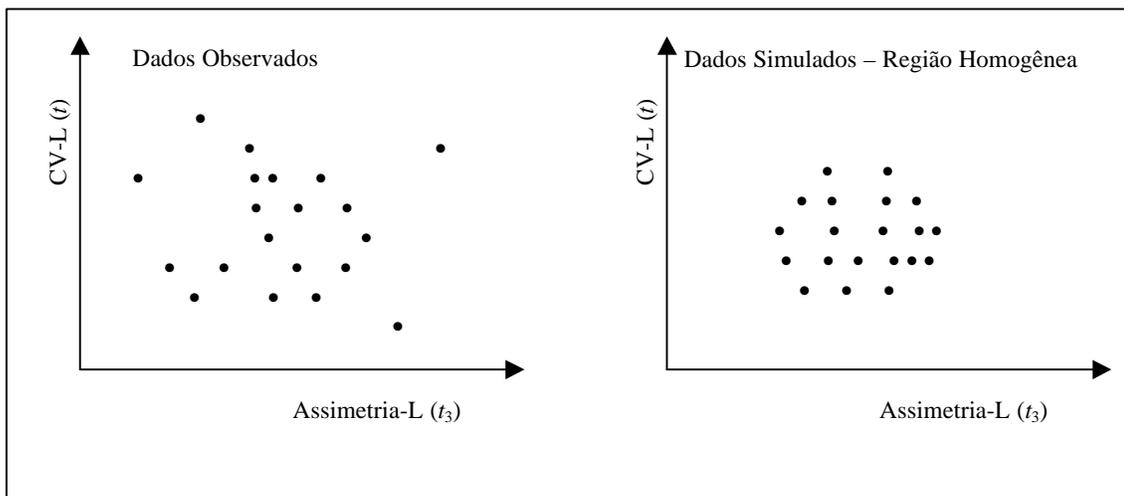


Figura 4.2 – Descrição esquemática do significado de heterogeneidade regional.

Para tornar possível o cálculo das estatísticas simuladas para a região homogênea, é necessário especificar uma função de distribuição de probabilidades para a população de onde serão extraídas as amostras. Hosking & Wallis (1997) recomendam o emprego da distribuição Kapa de 4 parâmetros, a ser formalmente definida no subitem 4.3.2, e justificam que essa recomendação prende-se à preocupação de não assumir, *a priori*, nenhum comprometimento com distribuições de 2 e/ou 3 parâmetros. Os momentos-L da distribuição Kapa populacional devem reproduzir as médias grupais dos quocientes CV-L, assimetria-L e curtose-L, calculados para os dados observados.

4.3.2 Definição Formal

Considere que uma dada região contenha N postos de observação, cada um deles indexado por i , com amostra de tamanho n_i e quocientes de momentos-L amostrais representados por t^i, t_3^i e t_4^i . Considere também que t^R, t_3^R e t_4^R denotam, respectivamente, as médias regionais dos quocientes CV-L, assimetria-L e curtose-L, ponderados, de forma análoga à especificada pela equação 1.2, pelos tamanhos das amostras individuais. Hosking & Wallis (1997) recomendam que a *medida de heterogeneidade*, denotada por H , baseie-se preferencialmente no cálculo da dispersão de t , ou seja, o CV-L para as regiões proposta e simulada. Inicialmente, efetua-se o cálculo do desvio-padrão ponderado V dos CV-Ls das amostras observadas, através da seguinte expressão:

$$V = \left[\frac{\sum_{i=1}^N n_i (t^i - t^R)^2}{\sum_{i=1}^N n_i} \right]^{\frac{1}{2}} \quad (4.2)$$

Em seguida, para a simulação da região homogênea, Hosking & Wallis (1997) sugerem, conforme menção anterior, a utilização da distribuição Kapa de 4 parâmetros. Essa distribuição é definida pelos parâmetros ξ , α , k e h e inclui, como casos particulares, as distribuições Logística, Generalizada de Valores Extremos e Generalizada de Pareto, sendo, portanto, teoricamente capaz de representar variáveis hidrológicas/hidrometeorológicas. As funções densidade, acumulada de probabilidades e de quantis da distribuição Kapa são dadas, respectivamente, por:

$$f(x) = \frac{1}{\alpha} \left[i - \frac{k(x - \xi)}{\alpha} \right]^{\frac{1}{k-1}} [F(x)]^{1-h} \quad (4.3)$$

$$F(x) = \left\{ 1 - h \left[1 - \frac{k(x - \xi)}{\alpha} \right]^{\frac{1}{k}} \right\}^{\frac{1}{h}} \quad (4.4)$$

$$x(F) = \xi + \frac{\alpha}{k} \left[1 - \left(\frac{1 - F^h}{h} \right)^k \right] \quad (4.5)$$

Se $k > 0$, x tem o limite superior de $\xi + \alpha/k$; se $k \leq 0$, x é ilimitado superiormente; x tem o limite inferior de $\xi + \alpha(1 - h^{-k})/k$ se $h > 0$, de $\xi + \alpha/k$ se $h \leq 0$ e $k < 0$ e de $-\infty$ se $h \leq 0$ e $k \geq 0$. Os momentos-L da distribuição Kapa são definidos para $h \geq 0$ e $k > -1$ ou para $h < 0$ e $-1 < k < -1/h$ e dados pelas seguintes expressões:

$$\lambda_1 = \xi + \frac{\alpha(1 - g_1)}{k} \quad (4.6)$$

$$\lambda_2 = \frac{\alpha(g_1 - g_2)}{k} \quad (4.7)$$

$$\tau_3 = \frac{(-g_1 + 3g_2 - 2g_3)}{g_1 - g_2} \quad (4.8)$$

$$\tau_4 = \frac{(-g_1 + 6g_2 - 10g_3 + 5g_4)}{g_1 - g_2} \quad (4.9)$$

Onde:

$$g_r = \begin{cases} \frac{r\Gamma(1+k)\Gamma\left(\frac{r}{h}\right)}{h^{1+k}\Gamma\left(1+k+\frac{r}{h}\right)} & \text{se } h > 0 \\ \frac{r\Gamma(1+k)\Gamma\left(-k-\frac{r}{h}\right)}{(-h)^{1+k}\Gamma\left(1-\frac{r}{h}\right)} & \text{se } h < 0 \end{cases} \quad (4.10)$$

Onde $\Gamma(\cdot)$ representa a função gama, tal como anteriormente definida.

Os parâmetros da população Kapa são estimados de forma a reproduzir os quocientes de momentos-L regionais $\{1, t^R, t_3^R, t_4^R\}$. Com os parâmetros populacionais, são simuladas N_{SIM} regiões homogêneas, sem correlações cruzada e serial, contendo N amostras individuais, cada qual com n_i valores da variável normalizada. Na seqüência, as estatísticas V_j ($j=1, 2, \dots, N_{SIM}$) são calculadas para todas as regiões homogêneas através da equação 4.2. A sugestão é que se faça o número de simulações, N_{SIM} , igual a 500.

A média aritmética das estatísticas V_j obtidas por simulação, fornecerá a dispersão média esperada para a região homogênea:

$$\mu_v = \frac{\sum_{j=1}^{N_{SIM}} V_j}{N_{SIM}} \quad (4.11)$$

A medida de heterogeneidade H compara a dispersão observada com a simulada:

$$H = \frac{(V - \mu_v)}{\sigma_v} \quad (4.12)$$

Onde, σ_v é o desvio-padrão entre os N_{SIM} valores da medida de dispersão V_j , ou seja:

$$\sigma_v = \sqrt{\frac{\sum_{j=1}^{N_{SIM}} (V_j - \mu_v)^2}{N_{SIM} - 1}} \quad (4.13)$$

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 heterogênea" e, finalmente, se $H \geq 2$, a região deve ser classificada como "definitivamente heterogênea".

4.3.3 Discussão

Conforme mencionado anteriormente, alguns ajustes subjetivos, como remoção ou reagrupamento de postos de uma ou mais regiões, podem se tornar necessários para fazer com que a medida de he-

terogeneidade se adeqüe aos limites propostos. Entretanto, é possível que, em alguns casos, a heterogeneidade aparente se deva à presença de um pequeno número de postos “atípicos” na região. Uma alternativa é reagrupá-los em outra região, na qual sejam “mais típicos”, muito embora não exista nenhuma razão física evidente de que esse pequeno grupo de postos tenha comportamento distinto do restante dos postos da região de origem. Hosking & Wallis (1997) argumentam que, nesses casos, as razões de natureza física devem ter precedência sobre as de natureza estatística e recomendam a alternativa de manter o grupo de postos “atípicos” na região originalmente proposta. Continuam a argumentação tomando, como exemplo, a situação em que uma certa combinação de eventos meteorológicos extremos seja passível de ocorrer em qualquer ponto de uma região, mas que, de fato, tenha sido registrada em somente alguns de seus postos, durante o período disponível de observações. Os verdadeiros benefícios potenciais da regionalização poderiam ser atingidos em situações como a exemplificada, na qual o conhecimento dos mecanismos físicos associados à ocorrência de eventos extremos permite agrupar todos os postos em uma única região homogênea. Para esse exemplo, os dados locais encontram-se indevidamente influenciados pela *presença ou ausência* de eventos raros e a curva regional de freqüências, construída como a média das curvas individuais, constitui certamente o melhor instrumento para se estimar os riscos de futuras ocorrências dessa natureza.

A medida de heterogeneidade é construída como um teste de significância da hipótese nula de que a região é homogênea. Entretanto, Hosking & Wallis (1997) argumentam que não se deve interpretá-lo rigorosamente como tal, porque um teste de homogeneidade exata só seria válido sob as premissas de que os dados não possuem correlações cruzada e/ou serial e que a função Kapa representa a verdadeira distribuição regional. Mesmo se fosse possível construir um rigoroso teste de significância, ele teria utilidade duvidosa pois, na prática, mesmo uma região moderadamente heterogênea pode produzir melhores estimativas de quantis do que aquelas produzidas pela análise exclusiva de dados locais.

Os critérios $H=1$ e $H=2$, embora arbitrários, representam indicadores úteis. Se a medida de heterogeneidade fosse interpretada como um teste de significância e supondo que a estatística V possuísse uma distribuição Normal, o critério de rejeição da hipótese nula de homogeneidade, no nível $\alpha=10\%$, seria $H=1,28$. Nesse contexto, o critério arbitrário de $H=1$ pode parecer muito rigoroso; entretanto, conforme argumentação anterior, não se quer interpretar a medida H como um teste de significância exato. A partir de resultados de simulação, Hosking & Wallis (1997) demonstraram que, em média, $H\approx 1$ para uma região suficientemente heterogênea, na qual as estimativas de quantis são 20 a 40% menos precisas do que as obtidas para uma região homogênea. Assim sendo, o limite $H=1$ é visto como o ponto a partir do qual a redefinição da região pode apresentar vantagens. Analogamente, o limite $H=2$ é visto como o ponto a partir do qual redefinir a região é definitivamente vantajoso.

Em alguns casos, H pode apresentar valores negativos. Eles indicam que há menos dispersão entre os valores amostrais de CV-L do que se esperaria de uma região homogênea com *distribuições individuais de freqüência independentes*. A causa mais provável para esses valores negativos é a presença de correlação positiva entre os dados dos diferentes postos. Se valores muito negativos, como $H<-2$, são observados durante a regionalização, isso pode ser uma indicação de que há muita correlação cruzada entre as distribuições individuais de freqüência ou de que há uma regularidade excessiva dos valores amostrais de CV-L. Para esses casos, Hosking & Wallis (1997) recomendam reexaminar os dados de forma mais cuidadosa.

5 Seleção da Distribuição Regional de Freqüências

5.1 Seleção das Distribuições Candidatas – Propriedades Gerais

Existem diversas famílias de distribuições de probabilidade que podem ser consideradas candidatas a modelar um conjunto de dados regionais. A sua adequação como distribuições candidatas depende de sua capacidade de reproduzir algumas características amostrais relevantes. Apresentam-

se, a seguir, as principais propriedades a serem consideradas quando da seleção de um modelo probabilístico regional.

Limite Superior

Algumas quantidades físicas possuem limites superiores inerentemente definidos; é o caso, por exemplo, da concentração de oxigênio dissolvido em um corpo d'água, limitado fisicamente em um valor entre 9 a 10mg/l, a depender da temperatura ambiente. Outras quantidades podem possuir um limite superior; entretanto, esse limite não é conhecido *a priori*, fato decorrente da insuficiente compreensão e/ou quantificação de todos os processos físicos causais envolvidos. A esse respeito, é bastante conhecida a controvérsia quanto à existência da *Precipitação Máxima Provável* (PMP), originalmente formulada como um limite superior de produção de precipitação pelo ar atmosférico; se de fato existe a PMP, a determinação desse limite superior fica comprometida pela insuficiente quantificação da variabilidade espaço-temporal das variáveis que lhe dão origem. Entretanto, pode-se conjecturar que seria *fisicamente impossível* a ocorrência de uma vazão, digamos de 100.000m³/s, em uma pequena bacia hidrográfica, por exemplo, da ordem de 100km² de área de drenagem. Por essa razão, alguns pesquisadores, como Boughton (1980) e Laursen (1983), recomendam que *somente* distribuições limitadas superiormente devem ser usadas para modelar variáveis com essas características. Hosking & Wallis (1997) consideram errônea essa recomendação e sustentam que, se o objetivo da análise de frequência é o de estimar o quantil de tempo de retorno de 100 anos, é irrelevante considerar como "fisicamente impossível" a ocorrência do quantil de 100.000 anos. Acrescentam que impor um limite superior ao modelo probabilístico pode comprometer a obtenção de boas estimativas de quantis para os tempos de retorno que realmente interessam. Esses autores concluem afirmando que, ao se empregar uma distribuição ilimitada superiormente, as premissas implícitas são (i) que o limite superior não é conhecido e nem pode ser estimado com a precisão necessária e (ii) que no intervalo de tempos de retorno de interesse do estudo a distribuição de probabilidades da população pode ser melhor aproximada por uma função ilimitada do que por uma que possua um limite superior. Evidentemente, quando existem evidências empíricas de que a distribuição populacional possui um limite superior, ela deve ser aproximada por uma distribuição limitada superiormente. Seria o caso, por exemplo, do ajuste da distribuição generalizada de valores extremos a uma certa amostra, cuja tendência de possuir um limite superior estaria refletida na estimativa de um valor positivo para o parâmetro de forma k .

Cauda Superior

O peso da cauda superior de uma função distribuição de probabilidades determina a intensidade com que os quantis aumentam, à medida que os tempos de retorno tendem para valores muito elevados. Em outras palavras, o peso da cauda superior é proporcional às probabilidades de excedência associadas a quantis elevados e é reflexo da intensidade com que a função densidade $f(x)$ decresce quando x tende para valores muito elevados. Os pesos das caudas superiores de algumas das principais funções de distribuição de probabilidades encontram-se relativizados na tabela 5.1.

Para a maioria das aplicações envolvendo variáveis hidrológicas/hidrometeorológicas, a correta prescrição da cauda superior de uma distribuição de probabilidades é de importância fundamental e, em muitos casos, representa a motivação primeira da análise de frequência. Entretanto, os tamanhos das amostras disponíveis para essas aplicações são invariavelmente insuficientes para se determinar com exatidão a forma da cauda superior do modelo probabilístico. Segundo Hosking & Wallis (1997), não havendo razões suficientes para se recomendar o emprego exclusivo de somente um tipo de cauda superior, é aconselhável utilizar um grande conjunto de distribuições candidatas cujos pesos de suas caudas superiores se estendam por um amplo espectro. A *medida de aderência*, a ser descrita no item 5.3, provê os meios para se decidir quais distribuições e, portanto, quais pesos de caudas superiores, melhor *aproximam* o comportamento do conjunto de dados da região homogênea considerada.

Tabela 5.1 – Pesos das caudas superiores de algumas distribuições de probabilidade.

Cauda Superior	Forma de $f(x)$ para Valores Elevados de x	Distribuição
Pesada	x^{-A}	Generalizada de Valores Extremos, Generalizada de Pareto e Logística Generalizada com parâmetro de forma $k < 0$.
↑	$x^{-A \ln x}$	Log-Normal com assimetria positiva.
	$\exp(-x^A)$ $0 < A < 1$	Weibull com parâmetro de forma $\lambda < 1$.
	$x^A \exp(-Bx)$	Pearson tipo III com assimetria positiva.
	$\exp(-x)$	Exponencial, Gumbel.
↓	$\exp(-x^A), A > 1$	Weibull com parâmetro de forma $\lambda < 1$.
Leve	Limite superior	Generalizada de Valores Extremos, Generalizada de Pareto e Logística Generalizada com parâmetro de forma $k > 0$, Log-Normal e Pearson tipo III com assimetria negativa.

Nota: A e B representam constantes positivas (adap. de Hosking & Wallis, 1997, p. 75).

Cauda Inferior

Considerações semelhantes às do subitem anterior se aplicam à cauda inferior: é necessário utilizar um grande conjunto de distribuições candidatas cujos pesos de suas caudas inferiores se estendam por um amplo espectro. Entretanto, se o interesse do estudo encontra-se centrado em se prescrever a melhor aproximação da cauda superior, a forma da cauda inferior é irrelevante. Em alguns casos, conforme enfatizado no relatório *Estimating Probabilities of Extreme Floods, Methods and Recommended Research*, do National Research Council (NRC, 1987), a presença de *outliers* baixos em uma dada amostra pode inclusive vir a comprometer a correta estimação das características da cauda superior.

Limite Inferior

Considerações semelhantes às do limite superior também se aplicam ao limite inferior. Contudo, diferentemente do limite superior, o inferior é, em geral, conhecido ou pode ser igualado a zero; algumas distribuições, como as generalizadas de Pareto e Logística, permitem com facilidade o ajuste do parâmetro de posição, quando se conhece ou se prescreve o limite inferior. Hosking & Wallis (1997) ressaltam, entretanto, que, em diversos casos, a prescrição de limite inferior nulo é inútil e que melhores resultados podem ser obtidos sem nenhuma prescrição *a priori*. Exemplificam afirmando que os totais anuais de precipitação em regiões úmidas, apesar de números positivos, são muito superiores a zero; para esse exemplo, uma distribuição de probabilidades realista deve ter um limite inferior muito maior do que zero.

5.2 Seleção de Distribuições Candidatas – Considerações Adicionais

As distribuições oriundas da teoria clássica de valores extremos (Gumbel, 1958), ou sejam, os modelos Gumbel, Fréchet e Weibull, são as únicas para as quais existem justificativas teóricas para seu emprego na modelação de valores máximos (ou mínimos) de dados empíricos. Por exemplo, o modelo de valores extremos do tipo I para máximos (EV1 ou Gumbel) é a distribuição assintótica do maior valor de uma seqüência ilimitada de variáveis aleatórias independentes e igualmente distribuídas (*iid*), a distribuição das quais possui uma cauda superior do tipo exponencial. Analogamente, a distribuição do tipo II para máximos (EV2 ou Fréchet) relaciona-se à variáveis *iid* com cauda superior do tipo polinomial, enquanto a distribuição do tipo III (EV3 ou Weibull) refere-se a variáveis *iid* que possuem um limite superior finito. Sob as premissas da teoria de valores extremos, a distribuição de probabilidades, por exemplo das vazões médias diárias máximas anuais de uma certa bacia hidrográfica, depende da distribuição inicial única dos valores diários considerados independentes. A maior objeção ao uso das distribuições oriundas da teoria de valores extremos em hidrologia refere-se à premissa de variáveis iniciais *iid*, a qual muito dificilmente é satisfeita por variáveis hidrológicas/hidrometeorológicas. A esse respeito, transcreve-se o seguinte comentário escrito por Perichi & Rodriguez-Iturbe (1985, pág. 515):

Presumir que duas vazões médias diárias, observadas digamos no dia 15 de maio e em 20 de dezembro, são variáveis aleatórias identicamente distribuídas, é uma clara violação da realidade hidrológica. Essa premissa 'regulariza' as distribuições históricas iniciais afirmando não só que elas são do mesmo tipo, mas também que elas possuem os mesmos parâmetros (e.g. média e variância) para qualquer dia do ano. Sob essa premissa, não se pode admitir o fato de que se uma mesma vazão média diária foi observada em dois dias diferentes, é mais provável que aquele que possui a maior variância produzirá cheias maiores do que aquele de menor variância. A realidade hidrológica é que a combinação da média e da variância de um dado mês faz com que alguns meses do ano sejam mais suscetíveis à ocorrência de cheias do que outros.

Além dessas considerações, a seqüência de variáveis hidrológicas/hidrometeorológicas, amostradas em intervalos horários ou diários ao longo de um ano, apresenta correlação serial significativa e não pode ser considerada suficientemente grande em termos assintóticos.

O fato de que variáveis hidrológicas/hidrometeorológicas dificilmente satisfazem às premissas da teoria clássica de valores extremos vem justificar o uso de distribuições não-extremais, tais como a Log-Normal, na análise local de frequência de eventos máximos anuais. Chow (1954) apresenta a seguinte justificativa para o emprego da distribuição Log-Normal: os fatores causais de várias variáveis hidrológicas agem de forma multiplicativa, ao invés de aditiva, e a soma dos logaritmos desses fatores, em consequência do *teorema central limite* da teoria de probabilidades, tende a ser normalmente distribuída. Stedinger *et al.* (1993) afirmam que algumas variáveis, como a diluição, por exemplo, podem resultar do produto de fatores causais. Entretanto, para o caso de enchentes ou precipitações máximas, a interpretação dessa ação multiplicativa não é evidente.

Essas objeções referem-se às justificativas teóricas inerentes à distribuição Log-Normal, bem como às distribuições oriundas da teoria clássica de valores extremos, porém, não têm o objetivo de excluí-las do elenco de distribuições candidatas à modelação de variáveis hidrológicas/hidrometeorológicas. No contexto da metodologia para análise regional de frequência, proposta por Hosking & Wallis (1997), elas devem ser consideradas candidatas como quaisquer outras distribuições e devem ter suas *medidas de aderência* (ver item 5.3) comparadas com as demais.

Com relação ao número de parâmetros desconhecidos de uma distribuição de probabilidades, Hosking & Wallis (1997) afirmam que as distribuições de dois parâmetros produzem estimativas precisas de quantis quando as características distributivas populacionais a elas se assemelham. Entretanto, quando isso não ocorre, pode-se produzir estimativas dos quantis seriamente enviesadas. A busca de um modelo probabilístico mais geral e flexível levou as agências do governo norte-americano a reconhecerem o uso da distribuição Log-Pearson do tipo III para a análise local de frequência de cheias má-

ximas anuais em projetos com participação federal. Conforme definição formal anterior, trata-se de uma distribuição de 3 parâmetros, resultante da transformação logarítmica de variáveis aleatórias distribuídas de acordo com Gama ou Pearson do tipo III. Embora os seus 3 parâmetros confirmem flexibilidade de forma a essa distribuição, a sua estimação, com base exclusiva em dados locais, é uma fonte de controvérsias. Bobée (1975) reporta situações em que a simples alteração do método de inferência estatística faz com que o parâmetro de forma dessa distribuição passe de negativo a positivo, o que a torna limitada superiormente ou inferiormente de acordo com o sinal do parâmetro. São essas características indesejáveis da distribuição Log-Pearson do tipo III que levaram, por exemplo, Reich (1977) a argumentar contra a sua utilização na análise local de frequência de cheias máximas anuais. Entretanto, no contexto de regionalização, Hosking & Wallis (1997) observam que a grande vantagem potencial da análise regional de frequência é justamente a de se poder estimar as distribuições de mais de 2 parâmetros de forma mais confiável do que o seria a partir de uma única amostra local. Seguem adiante afirmando que, obedecido o preceito da *parcimônia estatística*, recomenda-se o uso de distribuições de mais de 2 parâmetros por produzirem estimativas menos enviesadas de quantis nas caudas superior e inferior. Concluem dizendo que para as aplicações da análise regional de frequência, as distribuições de 3 a 5 parâmetros são mais apropriadas.

Em geral, a seleção da “melhor” distribuição de probabilidades baseia-se na qualidade e consistência de seu ajuste aos dados disponíveis. Entretanto, Hosking & Wallis (1997) ponderam que o objetivo da análise regional de frequência não é o de ajustar uma distribuição a uma amostra em particular. De fato, o que se objetiva é a obtenção de estimativas de quantis de uma distribuição de probabilidades da qual se espera serem extraídos futuros valores amostrais. Em outras palavras, o que se preconiza é a seleção, entre diversas candidatas, da distribuição mais *robusta*, ou seja, aquela mais capaz de produzir boas estimativas de quantis, mesmo que os futuros valores amostrais possam ser extraídos de uma outra distribuição, algo diferente da que foi ajustada.

Existem diversos testes de aderência de uma distribuição aos dados amostrais que são passíveis de serem adaptados ao contexto da análise regional de frequência; os seguintes exemplos podem ser citados: gráficos quantil-quantil, testes do Qui-Quadrado, de Kolmogorov-Smirnov e Filliben, bem como diagramas de momentos ou de quocientes de momentos-L. Hosking & Wallis (1997) consideram uma escolha natural tomar como base para um teste de aderência as médias regionais de estatísticas de momentos-L, como por exemplo a assimetria-L e a curtose-L, e compará-las às características teóricas das diferentes distribuições candidatas. Essa é a idéia básica da *medida de aderência Z*, a ser descrita no item que se segue.

5.3 Medida de Aderência

5.3.1 Descrição

Em uma região homogênea, os quocientes de momentos-L individuais flutuam em torno de suas médias regionais. Na maioria dos casos, as distribuições de probabilidades, candidatas a modelar o comportamento da variável em estudo, possuem parâmetros de posição e escala que reproduzem a média e o L-CV regionais. Portanto, a aderência de uma certa distribuição aos dados regionais deve se basear necessariamente em momentos-L de ordem superior; Hosking & Wallis (1997) consideram suficientes a assimetria-L e a curtose-L. Logo, pode-se julgar a aderência pelo grau com que uma certa distribuição aproxima as médias regionais de assimetria-L e curtose-L. Por exemplo, suponha que a distribuição candidata é a Generalizada de Valores Extremos (GEV) de 3 parâmetros. Quando ajustada aos dados da região pelo método dos momentos-L, essa distribuição irá reproduzir a média regional de assimetria-L. Pode-se julgar o grau de ajuste, portanto, pela diferença entre a curtose-L τ_4^{GEV} da distribuição e a média regional correspondente t_4^R , tal como esquematizado na figura 5.1.

Entretanto, essa diferença deve levar em conta a variabilidade amostral de t_4^R . Esta pode ser quantificada através de σ_4 , ou seja, o desvio-padrão de t_4^R , o qual é obtido por simulação de um grande número de regiões homogêneas, todas extraídas de uma população de valores distribuídos conforme uma GEV, contendo os mesmos indivíduos e tamanhos de amostras dos dados observados. Nesse caso, portanto, a medida de aderência da distribuição GEV pode ser calculada como $Z^{GEV} = (t_4^R - \tau_4^{GEV}) / \sigma_4$.

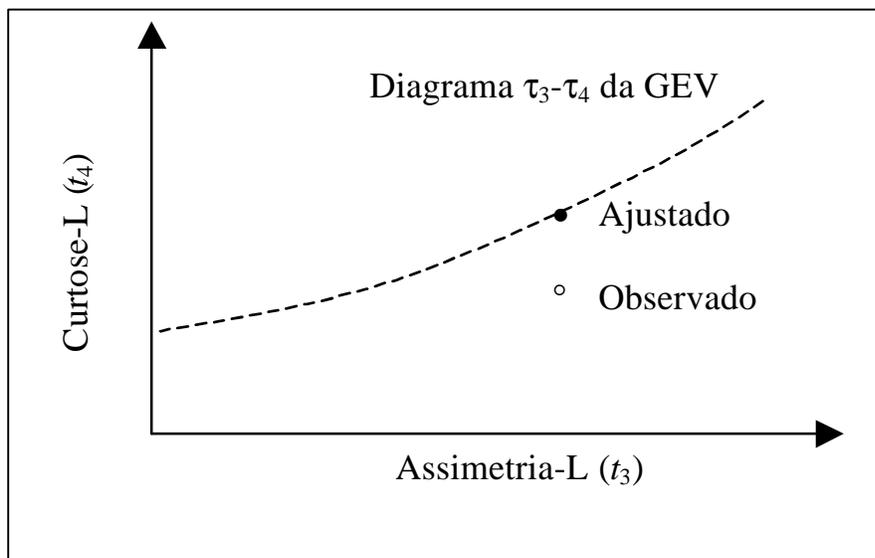


Figura 5.1 – Descrição esquemática da medida de aderência Z.

Hosking & Wallis (1997) reportam as seguintes dificuldades relacionadas ao procedimento de cálculo da medida de aderência, tal como anteriormente descrito:

- Para obter os valores corretos de σ_4 , é necessário um conjunto de simulações específico para cada distribuição candidata. Entretanto, na prática, Hosking & Wallis (1997) consideram que é suficiente supor que σ_4 tem o mesmo valor para todas as distribuições candidatas de 3 parâmetros. Justificam afirmando que, como todas as distribuições ajustadas têm a mesma assimetria-L, é razoável supor que elas também se assemelham com relação a outras características. Assim sendo, também é razoável supor que uma distribuição Kapa de 4 parâmetros, ajustada aos dados regionais, terá um valor de σ_4 próximo ao das distribuições candidatas. Portanto, σ_4 pode ser obtido a partir da simulação de um grande número de regiões homogêneas extraídas de uma população Kapa; para esse objetivo, podem ser empregadas as mesmas simulações usadas no cálculo da medida de heterogeneidade, conforme descrito no subitem 4.3.2.
- As estatísticas aqui mencionadas pressupõem a inexistência de qualquer viés no cálculo dos momentos-L amostrais. Hosking & Wallis (1997) observam que essa suposição é válida para t_3 mas não o é para t_4 , sob as condições de amostras de pequeno tamanho ($n \leq 20$) ou de populações de grande assimetria ($\tau_3 \geq 0,4$). A solução desse problema é feita por uma correção de viés para t_4 . Essa correção, denotada por B_4 , pode ser calculada através dos mesmos resultados de simulação usados para se calcular σ_4 .

- A medida de aderência Z refere-se a distribuições candidatas de 3 parâmetros. Embora seja possível construir procedimento semelhante para as distribuições de 2 parâmetros, elas possuem valores populacionais fixos de τ_3 e τ_4 e, em consequência, tornam problemática a estimação de σ_4 . Apesar de haverem sugerido algumas adaptações plausíveis, Hosking & Wallis (1997) desaconselham o uso da medida de aderência para distribuições de apenas 2 parâmetros.

5.3.2 Definição Formal

Considere que uma dada região contenha N postos de observação, cada um deles indexado por i , com amostra de tamanho n_i e quocientes de momentos-L amostrais representados por t^i, t_3^i e t_4^i . Considere também que t^R, t_3^R e t_4^R denotam, respectivamente, as médias regionais dos quocientes CV-L, assimetria-L e curtose-L, ponderados, de forma análoga à especificada pela equação 1.2, pelos tamanhos das amostras individuais.

Considere também um conjunto de distribuições candidatas de 3 parâmetros. Hosking & Wallis (1997) propõem o seguinte conjunto de distribuições candidatas: Logística Generalizada (LG), Generalizada de Valores Extremos (GEV), Generalizada de Pareto (GP), Log-Normal (LN3) e Pearson do tipo III (P3). Em seguida, cada distribuição candidata deve ter seus parâmetros ajustados ao grupo de quocientes de momentos-L regionais $\{1, t^R, t_3^R, t_4^R\}$. Denota-se por τ_4^{DIST} a curtose-L da distribuição ajustada, onde $DIST$ poderá ser qualquer uma das distribuições (por exemplo, LG, GEV, LN3).

Na seqüência, deve-se ajustar a distribuição Kapa ao grupo de quocientes de momentos-L regionais e proceder à simulação de grande número, N_{SIM} de regiões homogêneas, cada qual tendo a Kapa como distribuição de freqüência. Essa simulação deverá ser efetuada exatamente da mesma forma como a apresentada para o cálculo da medida de heterogeneidade (ver subitem 4.3.2). Em seguida, calculam-se as médias regionais t_3^m e t_4^m da assimetria-L e curtose-L da $m^{\text{ésima}}$ região simulada. O viés de t_4^R é dado por:

$$B_4 = \frac{\sum_{m=1}^{N_{SIM}} (t_4^m - t_4^R)}{N_{SIM}} \quad (5.1)$$

Enquanto o desvio-padrão de t_4^R o é pela expressão:

$$\sigma_4 = \sqrt{\frac{\sum_{m=1}^{N_{SIM}} (t_4^m - t_4^R)^2 - N_{SIM} B_4^2}{N_{SIM} - 1}} \quad (5.2)$$

A medida de aderência Z de cada distribuição candidata pode ser calculada pela equação:

$$Z^{DIST} = \frac{\tau_4^{DIST} - t_4^R + B_4}{\sigma_4} \quad (5.3)$$

A hipótese de um ajuste adequado é mais verdadeira quanto mais próxima de zero for a medida de aderência; Hosking & Wallis (1997) sugerem como critério razoável o limite $|Z^{DIST}| \leq 1,64$.

5.3.3 Discussão

A estatística Z é especificada sob a forma de um teste de significância e, segundo Hosking & Wallis (1997), possui distribuição que se aproxima da Normal padrão, sob as premissas de que a região é perfeitamente homogênea e de que não há correlação cruzada entre os seus indivíduos. Se a distribuição de Z é de fato a Normal, o critério $|Z^{DIST}| \leq 1,64$ corresponde à aceitação da hipótese de que os dados provêm da distribuição candidata, com um nível de confiança de 90%. Entretanto, as premissas, necessárias para se aproximar a distribuição de Z pela Normal padrão, dificilmente são completamente satisfeitas na prática. Assim sendo, o critério $|Z^{DIST}| \leq 1,64$ é simplesmente um indicador de boa aderência e não uma estatística de teste formal. Hosking & Wallis (1997) relatam que o critério $|Z^{DIST}| \leq 1,64$ é particularmente inconsistente se os dados apresentarem correlação serial e/ou correlação cruzada. Tanto uma quanto a outra tendem a fazer aumentar a variabilidade de t_4^R ; como não há correlação para as regiões simuladas de população Kapa, a estimativa de σ_4 resulta ser excessivamente pequena e a estatística Z excessivamente grande, conduzindo a uma falsa indicação de falta de aderência.

Se, ao se aplicar o teste da medida de aderência a uma região homogênea, resultar que várias distribuições são consideradas candidatas, Hosking & Wallis (1997) recomendam o exame das curvas de quantis adimensionais. Se estas fornecem resultados aproximadamente iguais, qualquer das distribuições candidatas pode ser selecionada. Entretanto, se os resultados diferem significativamente, a escolha deve tender para o modelo probabilístico que apresentar maior *robustez*. Nesses casos, ao invés de um modelo probabilístico de 3 parâmetros, recomenda-se a seleção da distribuição Kapa de 4 parâmetros ou da Wakeby de 5 parâmetros, as quais são mais robustas à incorreta especificação da curva regional de frequência. A mesma recomendação se aplica aos casos em que nenhuma das distribuições de 3 parâmetros atendeu ao critério $|Z^{DIST}| \leq 1,64$ ou aos casos de regiões "possivelmente heterogêneas" ou "definitivamente heterogêneas".

Além da verificação da medida de aderência Z , recomenda-se grafar as médias regionais da assimetria-L e curtose-L $\{t_3^R, t_4^R\}$ em um diagrama de quocientes de momentos-L, tal como o da Figura 2.1. Hosking & Wallis (1993) sugerem que, se o ponto $\{t_3^R, t_4^R\}$ se localizar acima da curva da distribuição Logística Generalizada, nenhuma distribuição de 2 ou 3 parâmetros se ajustará aos dados, devendo possivelmente se adotar uma distribuição Kapa de 4 parâmetros ou Wakeby de 5 parâmetros. Finalmente, ao se analisar uma grande área geográfica, sujeita à divisão em várias regiões homogêneas, a especificação da distribuição de frequência de uma região pode afetar a das outras. Se uma determinada distribuição se ajusta bem aos dados da maioria das regiões, é de bom senso utilizá-la para todas, muito embora ela possa não ser a distribuição que particularmente melhor se ajuste aos dados de uma ou de algumas das regiões.

6 Estimação da Distribuição Regional de Frequências

6.1 Justificativas

Depois que os dados dos diferentes postos da área em estudo foram submetidos às etapas descritas nos capítulos 3 a 5, tem-se como resultado a partição da área em regiões aproximadamente homogêneas, nas quais as distribuições de frequência de seus indivíduos são idênticas, à exceção de um fator de escala local, e podem ser modeladas por uma única distribuição de probabilidades regional, selecionada entre diversas funções candidatas. Essa relação entre as distribuições de frequência dos diversos locais representa a própria justificativa para a análise regional de frequência, permitindo

a obtenção de melhores estimativas de parâmetros e quantis a partir da combinação de dados espacialmente disseminados.

Diversos métodos podem ser utilizados para se ajustar uma distribuição de probabilidades aos dados de uma região homogênea. Para descrevê-los, considere inicialmente uma variável aleatória X , cuja variabilidade foi amostrada em N locais ou postos de observação situados em uma região homogênea. As observações, tomadas nos postos indexados por i , formam amostras de tamanho variável n_i e são denotadas por $X_{i,j}$, $i=1, \dots, N$; $j=1, \dots, n_i$. Se F , $0 < F < 1$, representa a distribuição de freqüências da variável X no posto i , então, a função de *quantis* nesse local é simbolizada por $X_i(F)$. Em uma região homogênea, as distribuições de freqüências nos N pontos são idênticas, à exceção de um *fator de escala local* μ_i , o *index-flood*, ou seja:

$$X_i(F) = \mu_i x(F), \quad i = 1, \dots, N \quad (6.1)$$

Se $\hat{\mu}_i$ denota a estimativa do fator de escala no local i , pode-se representar os *dados adimensionais padronizados* por $x_{i,j} = X_{i,j} / \hat{\mu}_i$, $j = 1, \dots, n_i$; $i = 1, \dots, N$.

O método mais simples e antigo para se combinar os dados locais, com o objetivo de se estimar os parâmetros e quantis da distribuição regional, é conhecido como o da *estação-ano*. Esse método simplesmente agrupa todos os dados adimensionais padronizados em única amostra, considerada aleatória simples, a qual é em seguida usada para se ajustar a distribuição regional. Hosking & Wallis (1997) consideram que, na atualidade, esse método é raramente empregado, principalmente porque não é correto tratar os dados adimensionais padronizados como uma *amostra aleatória simples*, ou seja, uma realização de variáveis aleatórias independentes e igualmente distribuídas. De fato, como os fatores de escala locais $\hat{\mu}_i$ são, em geral, estimativas obtidas a partir de amostras de diferentes tamanhos, os dados adimensionais padronizados dos diversos postos considerados não serão igualmente distribuídos.

Em outro extremo, encontra-se o *método de estimação através do máximo da função de verossimilhança*, tal como aplicado aos N fatores de escala locais μ_i e aos p parâmetros de $x(F; \theta_1, \dots, \theta_p)$, contidos na equação 6.1. O modelo estatístico procura encontrar, em geral de forma iterativa, as $N+p$ soluções de um sistema de $N+p$ equações que visam a maximizar a função de verossimilhança [ver, por exemplo, Buishand (1989)]. Esse método pode ser usado também para situações em que os fatores de escala são considerados parâmetros dependentes de informações covariadas, ou seja, $\mu_i = h(z_i, \omega)$, onde z_i representa um vetor de características ou informações covariadas no local i , h uma função matemática convenientemente escolhida e ω um vetor de parâmetros a serem estimados. Exemplos recentes de utilização dessa abordagem podem ser encontrados nos trabalhos de Smith (1989) e de Naghettini *et al.* (1996).

O método *index-flood* utiliza as estatísticas características dos dados locais para obter as estimativas regionais, ponderando-as através da equação:

$$\hat{\lambda}_k^R = \frac{\sum_{i=1}^N n_i \hat{\lambda}_k^{(i)}}{\sum_{i=1}^N n_i} \quad (6.2)$$

Onde $\hat{\lambda}_k^R$ denota a estimativa regional e $\hat{\lambda}_k^{(i)}$, $k = 1, \dots, p$ representam as estatísticas locais. Se estas se baseiam nos quocientes de momentos-L, Hosking & Wallis (1997) definem a metodologia de esti-

mação como a do *algoritmo dos momentos-L regionais*. Apesar de reconhecerem não haver nenhuma superioridade teórica da metodologia proposta, em relação à do máximo de verossimilhança, justificam o seu emprego pela maior simplicidade de cálculo. O algoritmo dos momentos-L regionais será descrito nos subitens que se seguem, tomando como premissa a inexistência de correlação cruzada entre as observações dos diferentes indivíduos de uma região homogênea ou de correlação serial entre as observações de um dado local.

6.2 Algoritmo dos Momentos-L Regionais

6.2.1 Descrição

O objetivo é o de ajustar uma única distribuição de frequência aos dados adimensionais padronizados, observados em diferentes locais de uma região considerada aproximadamente homogênea. O ajuste é efetuado através do *método dos momentos-L*, o qual consiste em igualar os momentos-L populacionais da distribuição em questão aos respectivos momentos-L amostrais. De forma mais conveniente, os *quocientes* de momentos-L locais são ponderados pelos seus respectivos tamanhos de amostra, de forma a produzir as estimativas regionais dos *quocientes* de momentos-L, as quais são, em seguida, empregadas para a inferência estatística. Se o *index-flood* é representado pela média da distribuição local de frequências, cuja estimativa é dada pela média amostral dos dados individuais, então a média dos dados adimensionais padronizados, bem como da ponderação regional, é 1. Isso faz com que os quocientes de momentos-L amostrais t e t_r , para $r \geq 3$, sejam os mesmos, não importando se foram calculados a partir dos dados originais $\{X_{i,j}\}$ ou pelos dados adimensionais padronizados $\{x_{i,j}\}$.

6.2.2 Definição Formal

Considere que uma dada região contenha N postos de observação, cada um deles indexado por i , com amostra de tamanho n_i e quocientes de momentos-L amostrais representados por t^i, t_3^i, t_4^i, \dots . Considere também que t^R, t_3^R, t_4^R, \dots denotam as médias regionais dos quocientes de momentos-L ponderados, de forma análoga à especificada pela equação 6.1, pelos tamanhos das amostras individuais. Conforme justificativa anterior, a média regional é 1, ou seja, $\ell_1^R = 1$.

Efetua-se o ajuste da distribuição regional, igualando-se os seus quocientes de momentos-L populacionais $\lambda_1, \tau, \tau_3, \tau_4, \dots$ às médias regionais $1, t^R, t_3^R, t_4^R, \dots$. Se F_i ou seja, a distribuição a ser ajustada, é definida por p parâmetros $\theta_k, k = 1, \dots, p$, resultará um sistema de p equações e p incógnitas, cujas soluções serão as estimativas $\hat{\theta}_k, k = 1, \dots, p$. Com estas, pode-se obter a estimativa da curva regional de quantis adimensionais $\hat{x}(F) = x(F; \hat{\theta}_1, \dots, \hat{\theta}_p)$. Inversamente, as estimativas dos quantis para o posto i são obtidas pelo produto de $\hat{x}(F)$ por $\hat{\mu}_i$, ou seja:

$$\hat{X}_i(F) = \ell_1^i \hat{x}(F) \quad (6.3)$$

6.2.3 Discussão

Os resultados obtidos por qualquer análise estatística possuem uma incerteza inerente, a qual pode ser avaliada por métodos tradicionais como, por exemplo, a construção de intervalos de confiança para as estimativas de parâmetros e quantis. Em geral, a construção de intervalos de confiança pressupõe que todas as premissas do modelo estatístico empregado sejam satisfeitas, o que, em termos da análise regional de

freqüência, equivale a dizer que as seguintes hipóteses tenham que ser rigorosamente verdadeiras: (a) a região é exatamente homogênea, (b) o modelo probabilístico foi especificado com exatidão e (c) não há correlação cruzada ou serial entre as observações. Por essa razão, Hosking & Wallis (1997) consideram que, no contexto da análise regional de freqüência, a construção de intervalos de confiança para parâmetros e quantis é de utilidade limitada. Como alternativa, propõem uma abordagem de avaliação da precisão das estimativas de quantis, com base em simulação de Monte Carlo, na qual se leva em consideração a possibilidade de heterogeneidade regional e existência de correlação cruzada e/ou serial, bem como da incorreta especificação do modelo probabilístico regional. A descrição do experimento de Monte Carlo encontra-se fora do escopo da presente síntese; a seguir, estão transcritas as principais conclusões do estudo levado a termo por Hosking & Wallis (1997).

- Mesmo em regiões com grau moderado de heterogeneidade, presença de correlação cruzada e incorreta especificação do modelo probabilístico regional, os resultados da análise regional de freqüência são mais confiáveis do que os obtidos pela análise local.
- A regionalização é particularmente útil para a estimação de quantis muito altos ou baixos das respectivas caudas superior e inferior das distribuições de freqüência.
- Em se tratando de regiões heterogêneas com muitos postos, os erros das estimativas de quantis e da curva regional de quantis adimensionalizados decrescem lentamente em função do número de postos. Como conclusão, pode-se afirmar que, em geral, o ganho em precisão é pequeno quando se usam regiões com mais de 20 postos.
- As amostras maiores tornam de menor utilidade a análise regional de freqüência, relativamente à análise local. Entretanto, as amostras maiores facilitam a identificação de heterogeneidade regional. Como conclusão, pode-se afirmar que, em geral, quando os tamanhos das amostras são grandes, as regiões devem conter poucos postos.
- Não se recomenda o uso de distribuições de 2 parâmetros para a análise regional de freqüência. Preconiza-se o seu emprego somente se o analista está completamente seguro de que a assimetria-L e a curtose-L da distribuição são precisamente reproduzidas pelas estimativas amostrais. Caso contrário, as estimativas de quantis estarão fortemente enviesadas.
- Os erros provenientes da incorreta especificação da distribuição de freqüência são importantes somente para quantis muito altos ou baixos das respectivas caudas superior e inferior. Por exemplo, para a cauda superior, ocorrem erros significativos somente para $F > 0,99$.
- Certas distribuições robustas, como a Kapa e Wakeby, produzem estimativas de quantis razoavelmente precisas para uma ampla variedade de distribuições locais.
- A heterogeneidade regional introduz um viés nas estimativas de quantis dos postos considerados atípicos, em relação à região como um todo.
- A dependência estatística entre os postos aumenta a variabilidade das estimativas de quantis, mas tem pouca influência sobre o viés. Um pequeno grau de correlação cruzada não invalida os resultados da estimação regional.
- Para quantis extremos ($F \geq 0,999$), a vantagem da análise regional sobre a local é muito maior. Para quantis dessa ordem de grandeza, a heterogeneidade é menos importante como fonte de erros, ao passo que a incorreta especificação do modelo probabilístico é mais significativa.

7 Uso de Séries de Duração Parcial na Análise Regional de Frequência

7.1 Introdução

A metodologia para análise regional de frequência, proposta por Hosking & Wallis (1997), pode ser aplicada a quaisquer variáveis aleatórias que, em um ponto do espaço geográfico de uma região homogênea, sejam independentes e identicamente distribuídas. É o caso, por exemplo, de *eventos máximos anuais* associados a variáveis hidrológicas/hidrometeorológicas, objeto de grande parte das aplicações da metodologia proposta. Entretanto, é possível adaptar a seqüência metodológica para que as chamadas *séries de duração parcial* possam ser utilizadas, respeitadas certas condições relativas às observações locais que as constituem. Uma série de duração parcial é constituída somente pelos eventos que excedem um certo valor limiar, os quais permitem, na maioria dos casos, a obtenção de uma amostra de tamanho superior ao número de anos de registros. Dessa forma, se um dado posto possui n anos de registros, a série de duração parcial correspondente poderia ser constituída pelos $2n$ ou $3n$ maiores valores independentes que excedem o valor limiar estabelecido. Isso equivale a dizer que um determinado ano pode não conter nenhum valor superior ao limiar, enquanto outros podem conter vários. A alternativa de séries de duração parcial é particularmente útil para a situação em que os registros são muito curtos, como é geralmente exemplificado pelas séries de observações de postos pluviográficos.

Além de permitir a utilização de uma amostra de tamanho maior do que o número de anos de registros, a série de duração parcial contém os maiores eventos observados, independentemente do requisito de terem sido os máximos em um dado ano. Considere, por exemplo, as observações tomadas em um ano relativamente muito seco e, contrariamente, em um ano muito chuvoso, que compõem a série de vazões médias diárias de um posto fluviométrico. Somente o valor máximo de cada ano teria que ser selecionado para compor a série de máximos anuais, não importando a possibilidade de que a máxima observação do ano seco fosse 15 ou 20 vezes menor do que o evento máximo do ano chuvoso; nesse caso, também não importaria a possibilidade de que o ano chuvoso pudesse ter três ou quatro eventos independentes apenas ligeiramente inferiores ao valor máximo daquele ano. Portanto, o uso das séries de duração parcial apresenta também a vantagem potencial de se poder contar com as maiores observações, certamente mais representativas das características da cauda superior da distribuição de probabilidades a ser ajustada. Em se tratando de eventos raros, a inferência quanto às características da cauda superior de uma distribuição de probabilidades representa a própria razão de ser da análise de frequência. Nesse sentido, o emprego das séries de duração parcial faz com que a inferência estatística leve em consideração somente as maiores observações disponíveis, em detrimento daquelas muito frequentes e de menor importância, as quais poderiam vir a distorcer ou prejudicar a estimação da distribuição de probabilidades dos eventos raros.

Este capítulo apresenta as bases teóricas da modelação de séries de duração parcial, algumas condicionantes necessárias para o seu uso e, finalmente, as deduções das distribuições de probabilidade anuais a partir dos principais modelos probabilísticos regionais propostos na metodologia de Hosking & Wallis (1997).

7.2 Bases Teóricas

Diversas variáveis hidrológicas/hidrometeorológicas variam no tempo de forma a constituir períodos de curta duração em que seus valores são muito elevados em relação à média, separados por períodos de valores inferiores à média ou mesmo nulos. Esse fato confere a essas variáveis a configuração característica de uma sucessão de *excedências*, em relação a um certo valor limiar de referência. A magnitude e o número dessas excedências são naturalmente aleatórios e elas são passíveis de serem modeladas por um *processo estocástico bivariado*. Para maior clareza, considere que a figura 7.1 re-

presente um trecho da variação temporal de uma variável hidrológica Y , ao longo do qual são identificadas todas as ocorrências superiores a um certo valor limiar u . Dessa forma, a i -ésima ocorrência de Y superior a u terá o seu *valor máximo* denotado por Y_i , resultado da soma de u e da excedência z_i , enquanto o tempo a ela associado será representado por T_i . Essa representação constitui o processo estocástico bivariado $\{T_i, Y_i; i = 1, 2, \dots\}$, a modelação do qual tem sido objeto de diversos estudos e investigações, entre as quais podem ser citadas as referências clássicas de Todorovic & Zelenhasic (1970), Gupta *et al.* (1976), Todorovic (1978) e North (1980). Outras referências importantes são os trabalhos de Taesombut & Yevjevich (1978), Smith (1984), Rosbjerg (1984) e Van Montfort & Witter (1986).

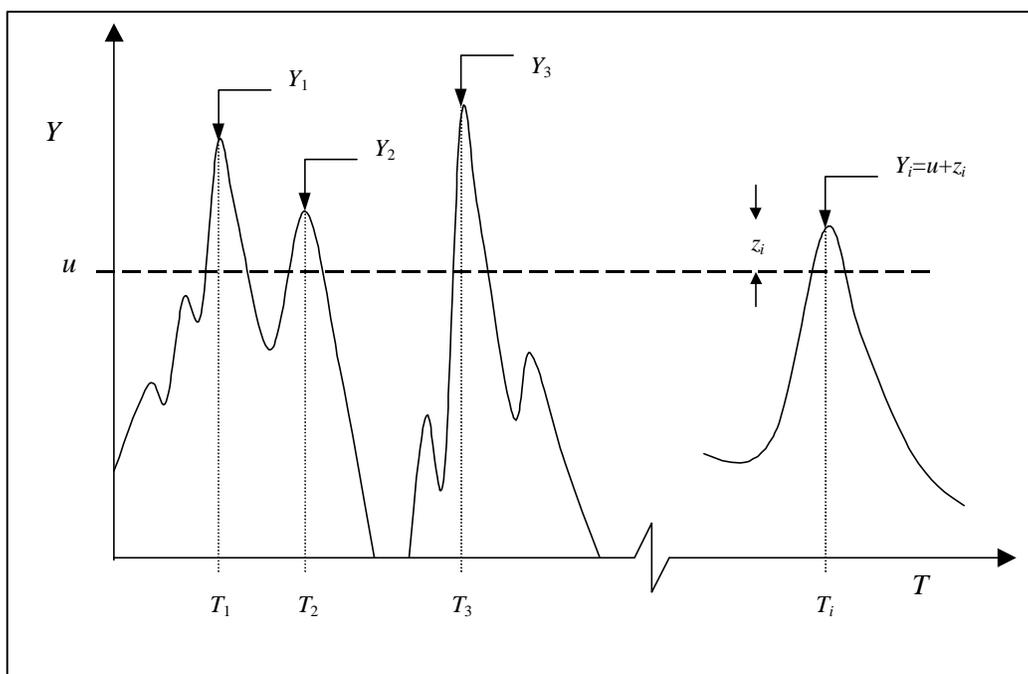


Figura 7.1 – Representação gráfica de processos estocásticos bivariados.

Sob condições gerais, os eventos $\{T_i, Y_i; i = 1, 2, \dots\}$ podem ser representados pela classe de processos estocásticos compostos e não-homogêneos de Poisson. Para isso, dois requisitos são necessários:

- 1 O número $N_{\alpha\beta}$ de excedências em um intervalo de tempo $[\alpha, \beta]$ é uma variável aleatória discreta, cuja função massa de probabilidades é a de Poisson com intensidade ou taxa de ocorrência $\lambda(t)$ dependente do tempo. Logo, por definição,

$$P(N_{\alpha\beta} = n) = \frac{\left[\int_{\alpha}^{\beta} \lambda(t) dt \right]^n \exp\left[- \int_{\alpha}^{\beta} \lambda(t) dt \right]}{n!} \tag{7.1}$$

- 2 $\{Y_i\}$ é uma seqüência de variáveis aleatórias mutuamente independentes com distribuição de probabilidades dependente do tempo de ocorrência T_i .

Suponha que o intervalo $[\alpha, \beta]$ possa ser dividido em k_0 subintervalos, dentro de cada qual a distribuição de $\{Y_i\}$ não dependa do tempo. Denotando o número de ocorrências dentro do j -ésimo subintervalo por N_j e o máximo de Y correspondente por M_j , pode-se escrever:

$$P(M_j \leq y) = P(N_j = 0) + \sum_{n=1}^{\infty} P\left[\bigcap_{i=1}^n (Y_{i,j} \leq y) \cap (N_j = n)\right] \quad (7.2)$$

Onde Y_{ij} denota a i -ésima ocorrência superior ao valor limiar u , dentro do j -ésimo subintervalo, e \cap representa a simultaneidade ou interseção dos eventos indicados. Pela condição de independência mútua, imposta pelo requisito 2, segue-se que:

$$P(M_j \leq y) = P(N_j = 0) + \sum_{n=1}^{\infty} P(N_j = n) [H_{u,j}(y)]^n \quad (7.3)$$

Nessa equação, $H_{u,j}$ representa a função de distribuição de probabilidades das ocorrências de Y que excedem u , dentro do j -ésimo subintervalo. Substituindo a equação 7.1 na expressão 7.3, segue-se que:

$$P(M_j \leq y) = \exp\left\{-[1 - H_{u,j}(y)] \int_j \lambda(t) dt\right\} \quad (7.4)$$

Conforme North (1980), pode-se deduzir a distribuição do máximo $M_{\alpha\beta}$ ao longo do intervalo $[\alpha, \beta]$, da seguinte forma:

$$P(M_{\alpha\beta} \leq y) = P\left(\bigcap_{j=1}^{k_0} M_j \leq y\right) \quad (7.5)$$

Ou, pela condição expressa pelo requisito 2:

$$P(M_{\alpha\beta} \leq y) = \prod_{j=1}^{k_0} P(M_j \leq y) \quad (7.6)$$

Onde o símbolo Π indica o produto das probabilidades indicadas. Combinando as equações 7.6 e 7.4, resulta que:

$$P(M_{\alpha\beta} \leq y) = \exp\left\{-\sum_{j=1}^{k_0} [1 - H_{u,j}(y)] \int_j \lambda(t) dt\right\} \quad (7.7)$$

Quando $k_0 \rightarrow \infty$, a equação 7.7 torna-se:

$$P(M_{\alpha\beta} \leq y) = \exp\left\{-\int_{\alpha}^{\beta} [1 - H_u(y/t)] \lambda(t) dt\right\} \quad (7.8)$$

Essa equação permite o cálculo da probabilidade do máximo $M_{\alpha\beta}$ dentro de qualquer intervalo de tempo $[\alpha, \beta]$. Em geral, como o interesse se volta para a obtenção da distribuição dos *máximos anuais* $F_M(y)$, faz-se com que os limites $\alpha=0$ e $\beta=1$ representem, respectivamente, o início e o fim do ano, e a equação 7.8 torna-se:

$$F_M(y) = \exp\left\{-\int_0^1 \lambda(t)[1 - H_u(y/t)]dt\right\} \quad (7.9)$$

Nessa equação, a distribuição de probabilidades das ocorrências de Y que excedem o valor limiar u , representada por $H_u(y/t)$, depende do tempo. Em geral, os diversos estudos e aplicações das séries de duração parcial sugerem não haver evidências empíricas suficientemente fortes para rejeitar a hipótese de que a distribuição $H_u(y/t)$ não depende do tempo. Se essa dependência não é considerada, a equação 7.9 pode ser muito simplificada e a distribuição dos máximos anuais passa a ser:

$$F_M(y) = \exp\left\{-[1 - H_u(y)]\int_0^1 \lambda(t)dt\right\} = \exp\{-\Lambda[1 - H_u(y)]\} \quad (7.10)$$

Onde Λ indica a *intensidade anual de ocorrências*. A equação 7.10 é a base para o emprego de séries de duração parcial e requer a estimação de Λ e da função de distribuição $H_u(y)$. A intensidade ou taxa anual de ocorrências pode ser estimada pelo número médio anual de eventos que superam o valor limiar u , por exemplo, se houver n anos de registros e forem selecionados os $2n$ maiores valores de Y , a estimativa de Λ é 2. A função de distribuição $H_u(y)$ está associada aos eventos que superaram o valor limiar u e pode ser prescrita pelo modelo paramétrico que melhor se ajustar aos dados amostrais.

No contexto da metodologia para análise regional de freqüência, proposta por Hosking & Wallis (1997), a aplicação da equação 7.10 não apresenta dificuldades adicionais. Nesse caso, após a identificação preliminar das regiões homogêneas através da análise de *clusters*, as medidas de discordância e heterogeneidade devem ser calculadas com base nos n_i , $2n_i$ ou $3n_i$ maiores valores da variável a ser regionalizada, observados nos n_i anos de registros do posto i ; dessa forma, garante-se a mesma estimativa da intensidade anual de ocorrências Λ para todos os postos da região considerada, muito embora cada um deles tenha o seu próprio valor limiar u_i . Em seguida, a medida de aderência Z irá permitir a seleção do modelo probabilístico regional $H_u(y)$, válido para as ocorrências que superarem os respectivos valores limiares u_i , e o algoritmo dos momentos-L regionais permitirá obter as estimativas de seus parâmetros e quantis. Finalmente, a aplicação da equação 7.10 culminará na estimação da *distribuição regional dos máximos anuais* da variável Y .

7.3 Condicionantes

A construção teórica da equação 7.10, descrita no item 7.2, pressupõe que as ocorrências superiores ao valor limiar u sejam independentes entre si e que o número dessas excedências seja uma variável de Poisson. Tais pressupostos são as condicionantes fundamentais para a correta modelação de séries de duração parcial e serão objeto de discussão nos subitens que se seguem.

7.3.1 Independência Serial

A independência serial das ocorrências superiores ao valor limiar u é um pressuposto importante e sua confirmação empírica deve anteceder o uso do modelo estocástico bivariado, desenvolvido no

item 7.2. Entretanto, algumas características próprias dos processos hidrológicos/hidrometeorológicos, bem como diversos estudos empíricos, indicam certas condições gerais sob as quais a hipótese de independência pode ser aceita. Embora não se possam estabelecer regras gerais, em se tratando de hidrogramas de cheia, os eventos devem ser selecionados de forma que estejam separados por um período de recessão suficientemente grande para que sejam considerados oriundos de episódios de chuva distintos. Da mesma forma, a seleção de eventos chuvosos deve ser condicionada à existência de um período significativo sem precipitação; no caso de chuvas intensas, por exemplo, é usual selecionar eventos separados por um mínimo de seis horas sem precipitação. Por se tratar de processos estocásticos contínuos, é de se esperar que a dependência serial contida nas séries hidrológicas/hidrometeorológicas de duração parcial irá decrescer com o aumento do valor limiar u ou, contrariamente, irá crescer com o acréscimo da intensidade anual Λ . De fato, um valor limiar suficientemente elevado faz com que o número de excedências se torne relativamente pequeno, enquanto o período entre os eventos se torna relativamente grande; em consequência, as excedências tendem a se tornar independentes entre si. Taesombut & Yevjevich (1978) estudaram a variação do coeficiente de correlação serial de primeira ordem com o valor médio do número de excedências $\hat{\Lambda}$ para as vazões observadas em 17 estações fluviométricas dos Estados Unidos; concluíram que esse coeficiente cresce com $\hat{\Lambda}$, mantendo-se dentro do limite de tolerância de 95% para $\hat{\Lambda} \leq 4,5$. Conclusões semelhantes foram obtidas por Madsen *et al.* (1993) a partir de séries de duração parcial de precipitação, observadas em diversas estações pluviométricas da Dinamarca.

7.3.2 Distribuição de Frequência do Número de Excedências

Para as variáveis hidrológicas/hidrometeorológicas, a premissa de que o número de excedências em relação a um valor limiar é uma variável de Poisson tem justificativas empíricas e teóricas. Do ponto de vista empírico, são inúmeros os estudos e aplicações em que essa premissa se verifica para valores limiares elevados [por exemplo, Todorovic (1978), Taesombut & Yevjevich (1978), Correia (1983), Rosbjerg & Madsen (1992) e Madsen *et al.* (1993)]. As justificativas teóricas de se usar um processo de Poisson para modelar excedências mutuamente independentes provêm dos trabalhos de Cramér & Leadbetter (1967) e Leadbetter *et al.* (1983). Em particular, Cramér & Leadbetter (1967, pág. 256) demonstraram que se um processo estocástico é Gaussiano, então, sob condições gerais, pode-se afirmar que o número de excedências em relação a um valor limiar u converge para um processo de Poisson, quando u tende para o infinito. Em relação a esse estudo, Todorovic (1978) argumenta que não há razão para presumir que essa conclusão estaria incorreta se o processo não for Gaussiano. Posteriormente, Leadbetter *et al.* (1983, pág. 282) demonstraram que as excedências de alguns outros processos não Gaussianos também convergem para um processo de Poisson quando u aumenta.

Apesar das justificativas teóricas mencionadas, resta, do ponto de vista prático, perguntar quão elevado deve ser o valor limiar para que as excedências possam ser consideradas independentes e aproximadas por um processo de Poisson. Langbein (1949, pág. 879) propôs o critério prático de se escolher o valor limiar de modo que, em média, não mais de duas ou três excedências anuais sejam selecionadas; em outras palavras, $\hat{\Lambda} \leq 3$. Por outro lado, Taesombut & Yevjevich (1978) concluíram pela aceitação da hipótese de Poisson nos casos em que a relação entre a média e a variância de Y é aproximadamente igual a 1. Outros resultados obtidos por Taesombut & Yevjevich (1978) mostram também que, quando comparadas às séries de máximos anuais, as de duração parcial conduzem a menores erros de estimação de quantis de Gumbel *apenas* quando $\Lambda \geq 1,65$; concluem pela recomendação das séries de duração parcial para um número médio anual de excedências igual ou superior a 1,95. Cunnane (1973), por sua vez, já recomenda sem reservas o uso das séries de duração parcial, principalmente para amostras com menos de dez anos de registros. Apesar da dificuldade de

se propor um critério geral, a experiência indica que especificar $\hat{\Lambda}$ entre 2 e 3 parece ser suficiente para auferir as vantagens de uso das séries de duração parcial e, ao mesmo tempo, garantir a independência serial dos eventos selecionados e, em muitos casos, a hipótese de Poisson. Entretanto, tal recomendação deve ser sempre sujeita a teste estatístico para verificar a sua adequação. O teste apropriado para se averiguar a veracidade da hipótese de Poisson foi primeiramente formulado por Cunnane (1979) e baseia-se na aproximação da distribuição de Poisson pela distribuição Normal. Considera-se que o número de excedências que ocorrem no ano k , denotado por m_k , segue uma distribuição Normal com média $\hat{\Lambda}$ e desvio-padrão $\hat{\Lambda}$. Nessas condições, pode-se afirmar que a estatística:

$$\gamma = \sum_{k=1}^N \left(\frac{m_k - \hat{\Lambda}}{\hat{\Lambda}} \right)^2 \quad (7.11)$$

Segue uma distribuição do Qui-Quadrado com $(N-1)$ graus de liberdade, onde N indica o número de anos de registros. Esse teste é considerado válido para os valores de $\hat{\Lambda}$ correntemente empregados e para tamanhos de amostra superiores a 5.

7.4 Funções de Distribuição de Probabilidades Anuais

Conforme menção anterior, o interesse último é o de obter a estimativa da distribuição dos *máximos anuais* $F_M(y)$ a partir da distribuição $H_u(y)$. No contexto da metodologia para análise regional de frequência, proposta por Hosking & Wallis (1997), esse interesse é o de obter a estimativa da *distribuição regional dos máximos anuais* $F_M(y)$ [ou $F(x)$, por simplicidade] a partir do modelo probabilístico regional $H_u(y)$ [ou $H(x)$], selecionado para as ocorrências que superarem os respectivos valores limiares u_i . No presente subitem, a equação 7.10 será empregada para deduzir a distribuição regional dos máximos anuais e sua respectiva função de quantis, considerando os seguintes modelos paramétricos para $H(x)$: Generalizada de Pareto, Generalizada de Valores Extremos e Logística Generalizada. Os resultados serão aqui referidos como os modelos Poisson-Pareto, Poisson-GEV e Poisson-Logística, respectivamente.

7.4.1 Modelo Poisson-Pareto

O modelo Poisson-Pareto refere-se ao processo estocástico bivariado em que o número de excedências é Poissoniano e as ocorrências que excedem o valor limiar, com intensidade anual Λ , seguem uma distribuição Generalizada de Pareto. Portanto,

$$H(x) = 1 - \exp(-y) \quad (7.12)$$

Onde:

$$y = -\frac{\ln \left[1 - \frac{k}{\alpha} (x - \xi) \right]}{k} \text{ para } k \neq 0, \quad y = \frac{x - \xi}{\alpha} \text{ para } k = 0 \text{ e } \xi, \alpha \text{ e } k \text{ são, respectivamente, os parâmetros}$$

de posição, escala e forma. A distribuição Generalizada de Pareto é ilimitada superiormente para $k \leq 0$ e possui limite superior para $k > 0$. Observe que, quando $k = 0$, ela se reduz à distribuição exponencial com parâmetros ξ e α .

Conforme dedução do item 7.2, a distribuição das probabilidades anuais da variável X é dada por:

$$F(x) = \exp\{-\Lambda[1 - H(x)]\} \quad (7.13)$$

Explicitando $H(x)$ nessa equação, tem-se que:

$$H(x) = 1 + \frac{1}{\Lambda} \ln[F(x)] \quad (7.14)$$

Igualando as equações 7.12 e 7.14, pode-se obter a expressão da função de quantis anuais do modelo Poisson-Pareto, ou seja:

$$x(F) = \xi + \frac{\alpha}{k} \left\{ 1 - [-\ln F]^{k/\Lambda} \right\} \text{ se } k \neq 0 \quad (7.15)$$

Ou:

$$x(F) = \xi + \alpha \{-\ln[F]\}^{1/\Lambda} \text{ se } k = 0 \quad (7.16)$$

Note que, ao fazer $\Lambda=1$ nas equações 7.15 e 7.16, elas se tornam as expressões das funções de quantis das distribuições Generalizada de Valores Extremos e Gumbel, respectivamente. As equações 7.15 e 7.16 podem ser expressas em função do período de retorno T , em anos, bastando, para isso, substituir $F(x)$ por $\left[1 - \frac{1}{T(x)}\right]$, conforme definição.

7.4.2 Modelo Poisson-GEV

Para o modelo Poisson-GEV, a função $H(x)$ corresponde à distribuição Generalizada de Valores Extremos (GEV) e é dada por:

$$H(x) = \exp[-\exp(-y)] \quad (7.17)$$

Onde:

$$y = -\frac{\ln\left[1 - \frac{k}{\alpha}(x - \xi)\right]}{k} \text{ para } k \neq 0, \quad y = \frac{x - \xi}{\alpha} \text{ para } k=0 \text{ e } \xi, \alpha \text{ e } k \text{ são, respectivamente, os parâmetros}$$

de posição, escala e forma. A distribuição GEV é ilimitada superiormente para $k \leq 0$ e possui limite superior para $k > 0$. Observe que, quando $k=0$, ela se reduz à distribuição de Gumbel com parâmetros ξ e α .

Nesse caso, igualando-se as expressões 7.14 e 7.17, pode-se obter a expressão da função de quantis anuais do modelo Poisson-GEV, ou seja:

$$x(F) = \xi + \frac{\alpha}{k} \left\{ 1 - \left[-\ln\left(\frac{\Lambda + \ln F}{\Lambda}\right) \right]^k \right\} \text{ se } k \neq 0 \quad (7.18)$$

Ou:

$$x(F) = \xi + \alpha \ln \left[-\ln \left(\frac{\Lambda + \ln F}{\Lambda} \right) \right] \text{ se } k = 0 \quad (7.19)$$

Da mesma forma, as equações 7.18 e 7.19 podem ser expressas em função do período de retorno T , em anos, bastando, para isso, substituir $F(x)$ por $\left[1 - \frac{1}{T(x)} \right]$, conforme definição.

7.4.3 Modelo Poisson-Logística

Nesse caso, a função $H(x)$ corresponde à distribuição Logística Generalizada, a qual, conforme parametrização de Hosking & Wallis (1997), é dada por:

$$H(x) = \frac{1}{1 + \exp(-y)} \quad (7.20)$$

Onde:

$y = -\frac{\ln \left[1 - \frac{k}{\alpha} (x - \xi) \right]}{k}$ para $k \neq 0$, $y = \frac{x - \xi}{\alpha}$ para $k = 0$ e ξ , α e k são, respectivamente, os parâmetros de posição, escala e forma. A distribuição Logística Generalizada é ilimitada superiormente para $k \leq 0$ e possui limite superior para $k > 0$.

Mais uma vez, igualando-se as expressões 7.14 e 7.20, pode-se obter a expressão da função de quantis anuais do modelo Poisson-Logística, ou seja:

$$x(F) = \xi + \frac{\alpha}{k} \left[1 - \left(-\frac{\ln F}{\Lambda + \ln F} \right)^k \right] \quad (7.21)$$

A equação 7.21 pode ser expressa em função do período de retorno T , em anos, bastando, para isso, substituir $F(x)$ por $\left[1 - \frac{1}{T(x)} \right]$, conforme definição.

Anexo II

ANÁLISE COMPARATIVA DOS RESULTADOS DE OUTROS ESTUDOS DE CHUVAS INTENSAS

1. Introdução

Preliminarmente à realização deste trabalho, pesquisou-se a existência de estudos de chuvas do estado do Rio de Janeiro. Foram encontrados e disponibilizados os trabalhos realizados por: Departamento de Estradas de Rodagem (DER/RJ), Fundação Superintendência Estadual de Rios e Lagoas (SERLA), Universidade Federal de Viçosa (UFV) e Otto Pfafstetter – *Estudo de Chuvas Intensas no Brasil*.

Quanto aos dados utilizados nesses estudos, ressaltam-se:

- *Estudo de Chuvas Intensas do Estado do Rio de Janeiro* – DER/RJ

Foram coletados os dados de precipitação diária das entidades INMET e ANEEL. O cálculo das precipitações com durações menores que 24 horas para TR = 5, 10 e 25 anos foi realizado seguindo a metodologia apresentada em 1974 pelo engenheiro Jaime Taborga Torrido em seu livro *Práticas Hidrológicas*.

- Projeto e Obras de Drenagem e Canalização das Bacias dos Rios Iguaçu e Barra Bota – Estudos Hidrológicos (revisão final em 08/91), realizado pela Tecnosolo para a SERLA.

Selecionaram-se os eventos máximos anuais de chuva dos pluviogramas para as durações de 10min, 20min, 30min, 1h, 2h, 6h e 12h. Os eventos de 24h foram extraídos dos registros pluviométricos. Na análise de freqüência, os dados foram ajustados à distribuição de probabilidades de Gumbel.

- Equações de Intensidade-Duração-Freqüência da Precipitação para os Estados do Rio de Janeiro e Espírito Santo: Estimativa e Espacialização – Tese de mestrado de Francisco Regis Liberato Pinto, desenvolvida na Universidade Federal de Viçosa (UFV), em junho/99.

Foram selecionados os eventos máximos anuais de chuva para as durações de 10min, 20min, 30min, 40min, 50min, 1h, 2h, 3h, 4h, 6h, 12h e 24h. Utilizaram-se as estações pluviográficas operadas pelo 6º Distrito de Meteorologia do Instituto Nacional de Meteorologia (DISMET/INMET). O período médio das séries históricas abrange os anos de 1974 a 1989. Na análise de freqüência, os dados foram ajustados à distribuição de probabilidade de Gumbel. As equações Intensidade-Duração-Freqüência foram ajustadas pelo método de regressão não-linear Gauss-Newton.

- *Estudo de Chuvas Intensas no Brasil* – Otto Pfafstetter

Os dados foram coletados a partir dos registros pluviográficos das estações do INMET. As estações levantadas têm um período médio de aproximadamente 22 anos de observação. Considerando que esse trabalho foi editado em 1957, acredita-se que os dados utilizados no estudo cobrem o período de observação de aproximadamente 1933 a 1955.

As figuras 1 a 20 apresentam uma análise comparativa das equações IDF do estado do Rio de Janeiro.

As figuras 21 a 50 apresentam os ajustes das curvas das distribuições de freqüências regionais, com as séries de máximas anuais das estações utilizadas na análise comparativa. Cabe ressaltar que as curvas de freqüência obtidas a partir do estudo de regionalização aderiram convenientemente à série de máximas anuais.

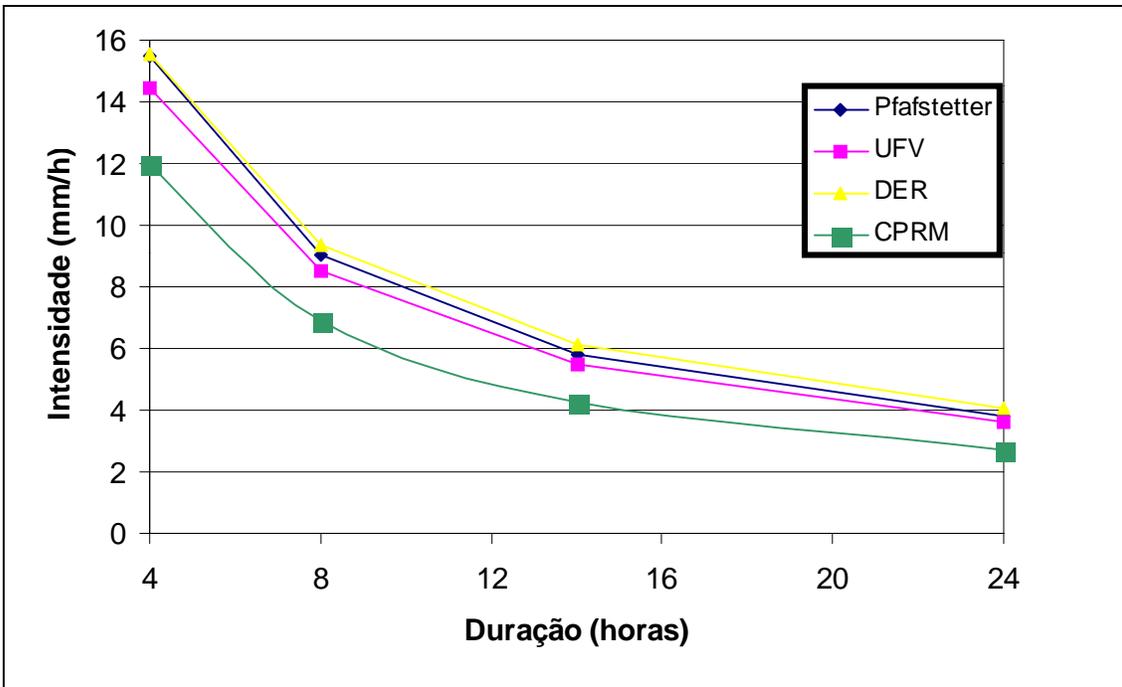
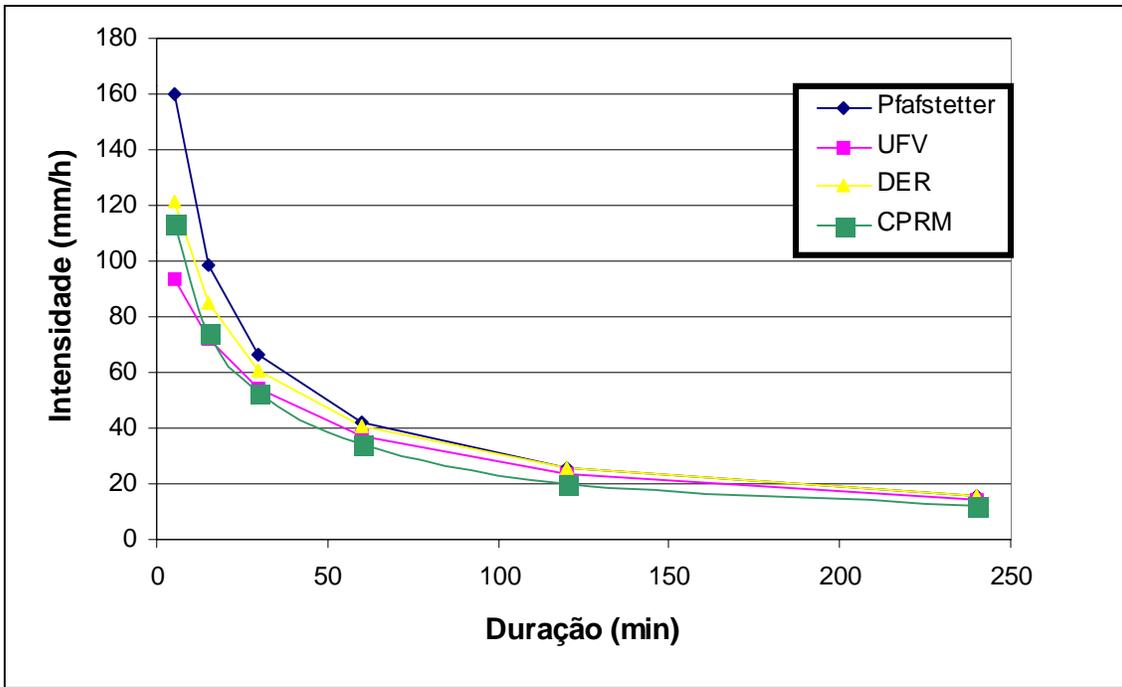


Figura 1 – Diagrama comparativo das equações IDF de Campos (região 1) TR = 2 anos.

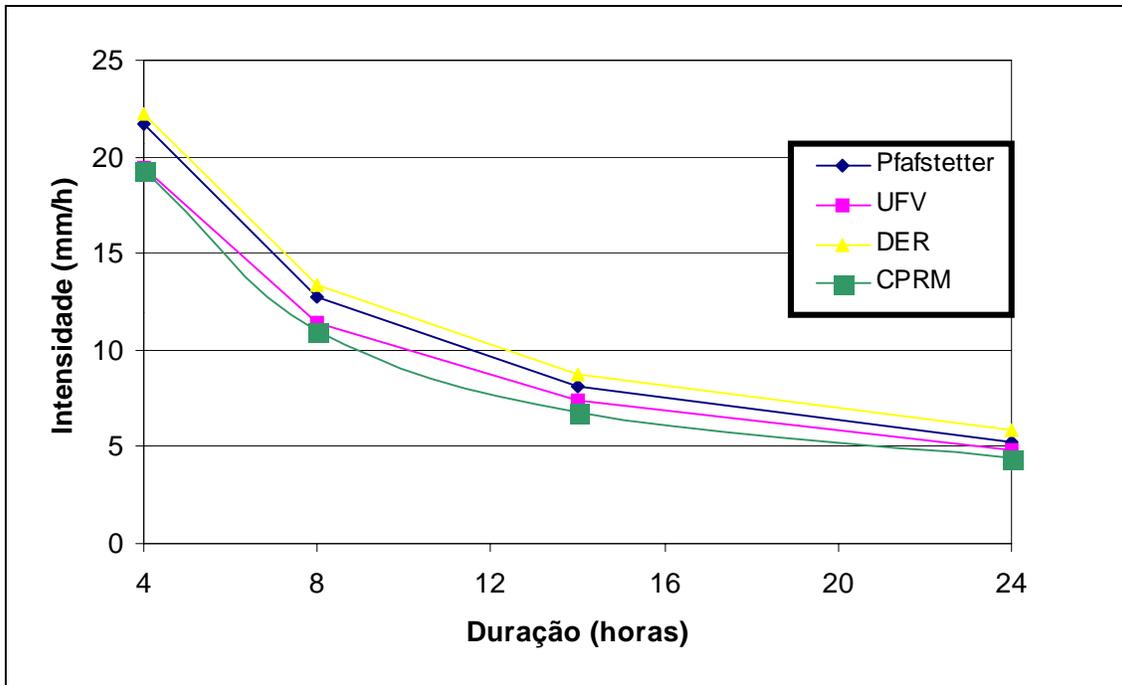
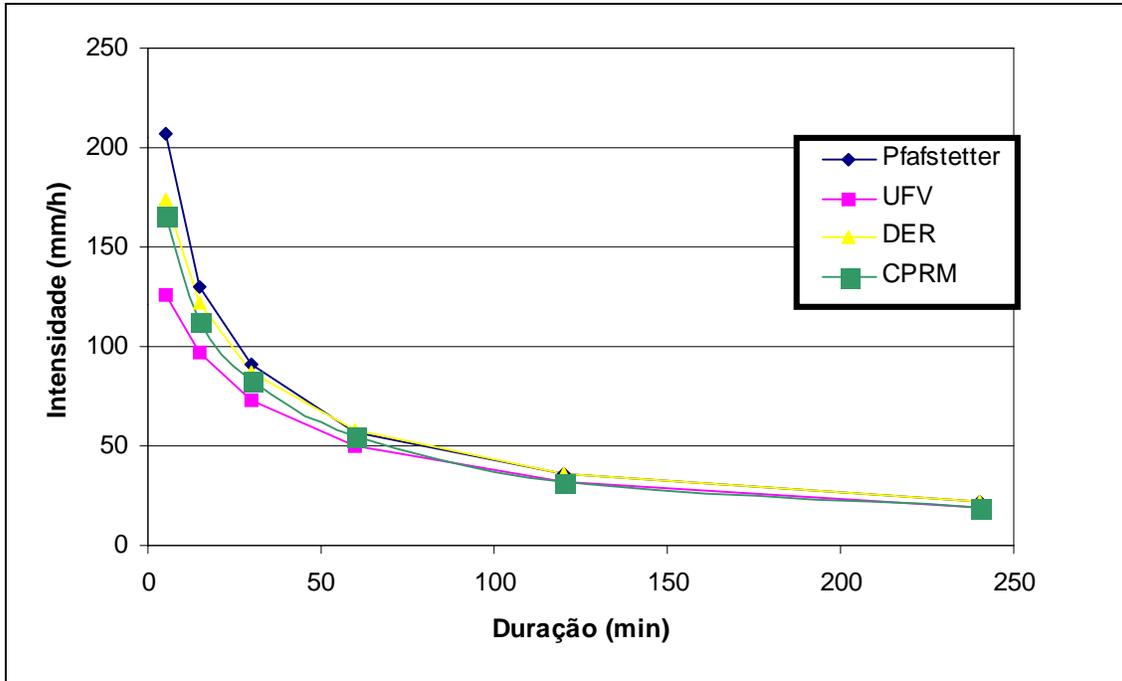


Figura 2 – Diagrama comparativo das equações IDF de Campos (região 1) TR = 10 anos.

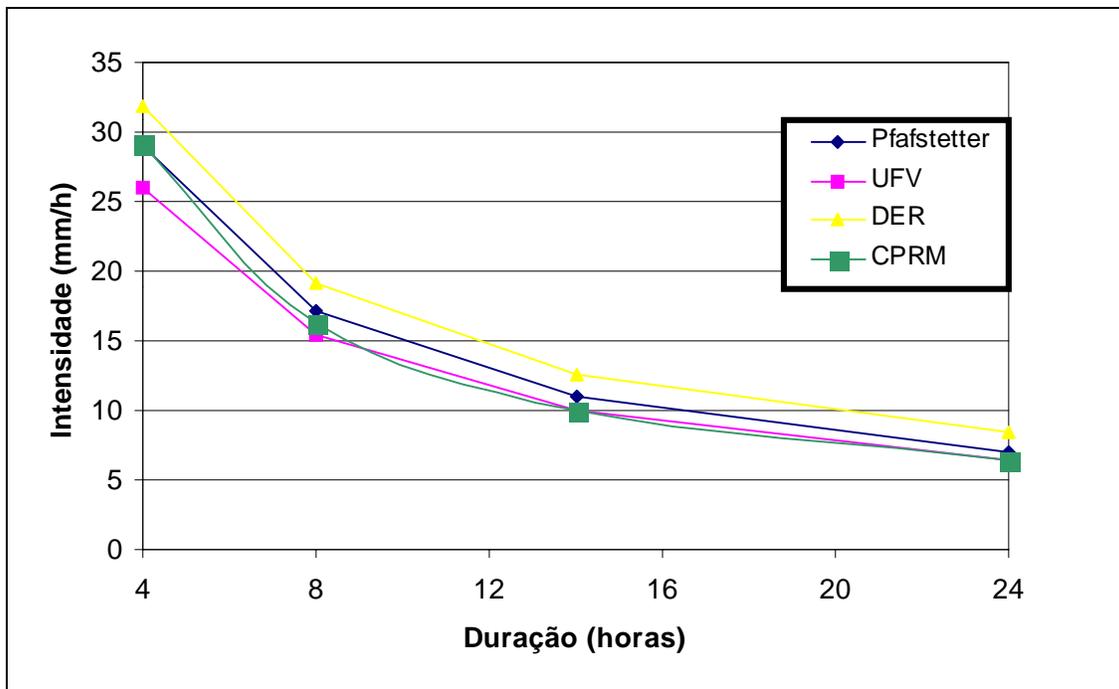
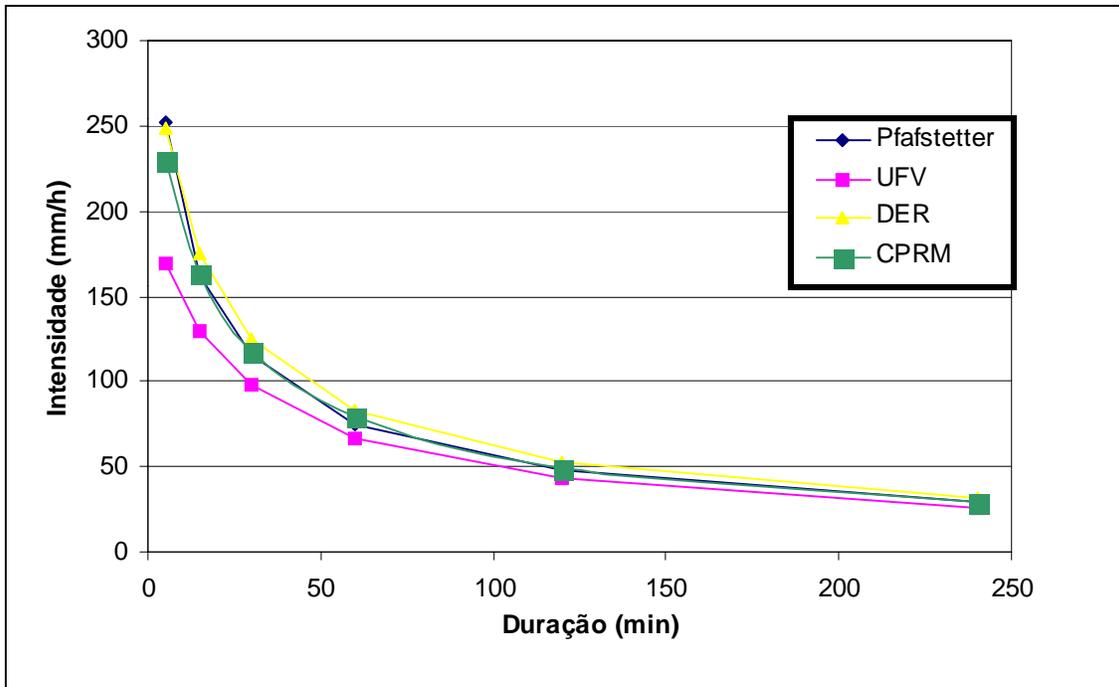


Figura 3 – Diagrama comparativo das equações IDF de Campos (região 1) TR = 50 anos.

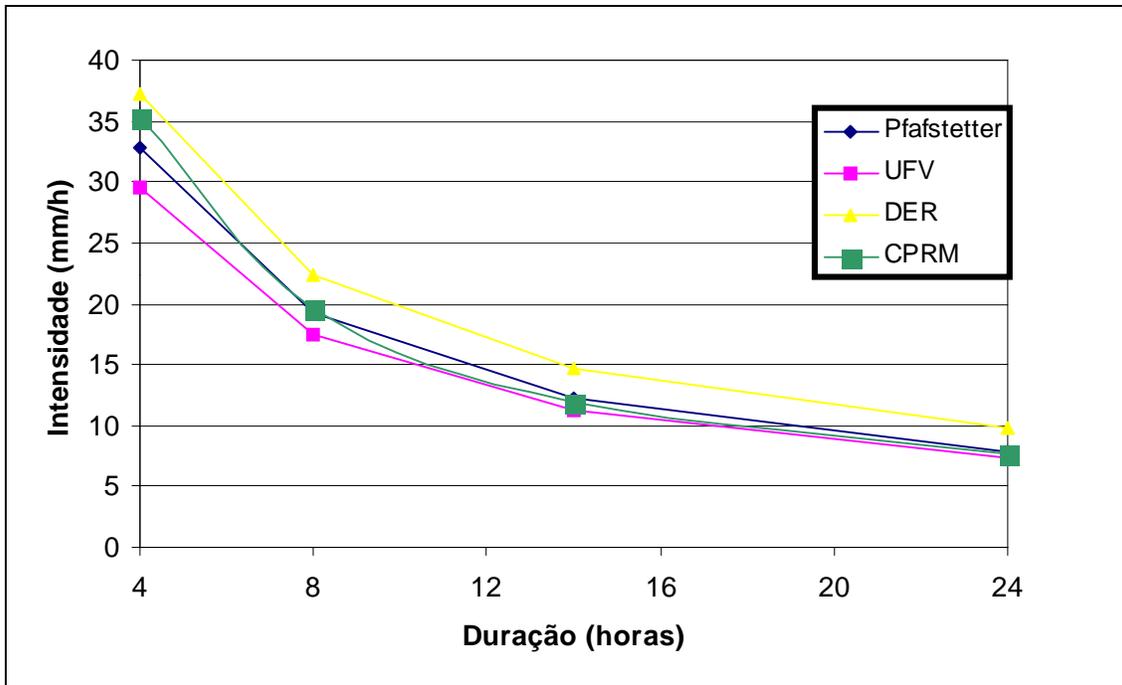
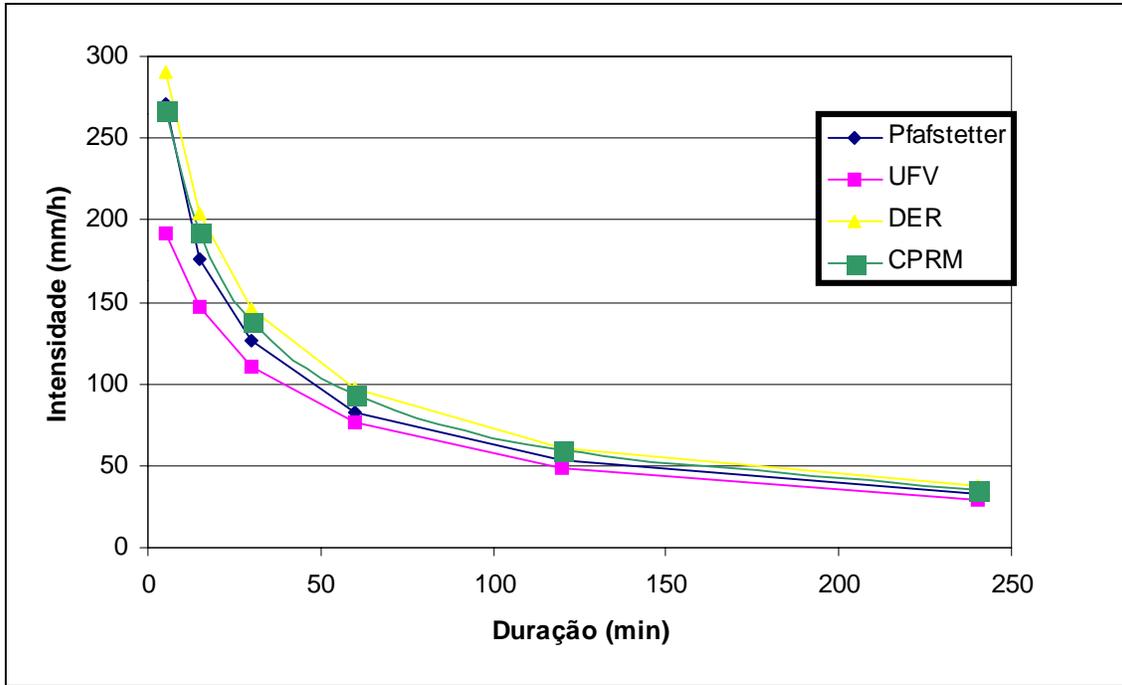


Figura 4 – Diagrama comparativo das equações IDF de Campos (região 1) TR = 100 anos.

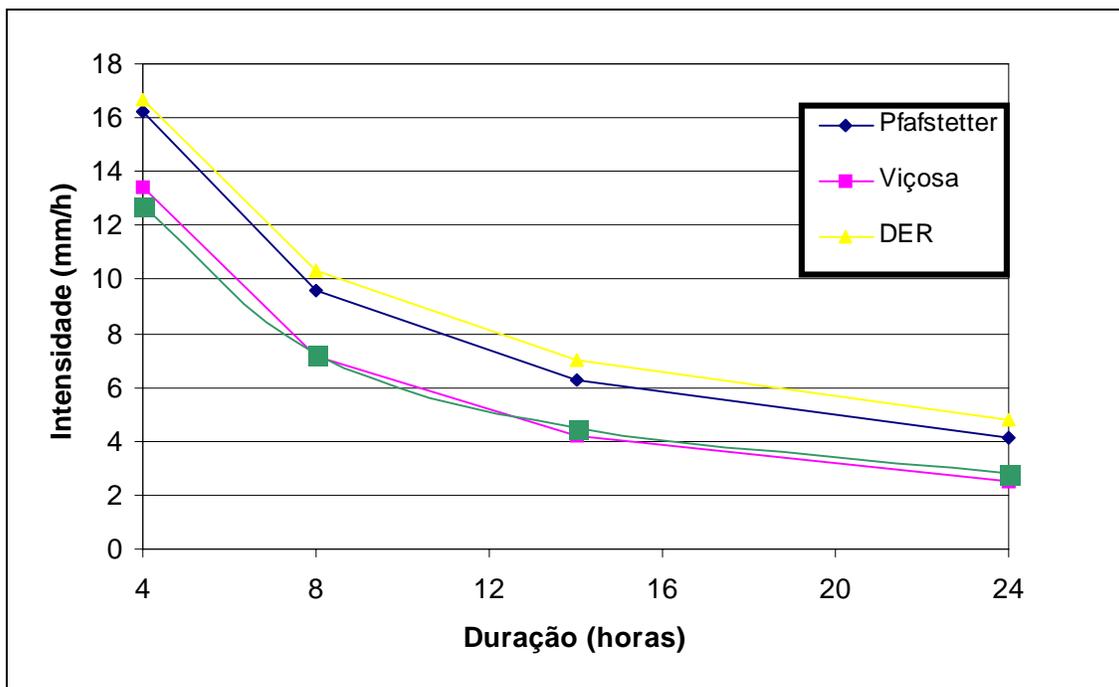
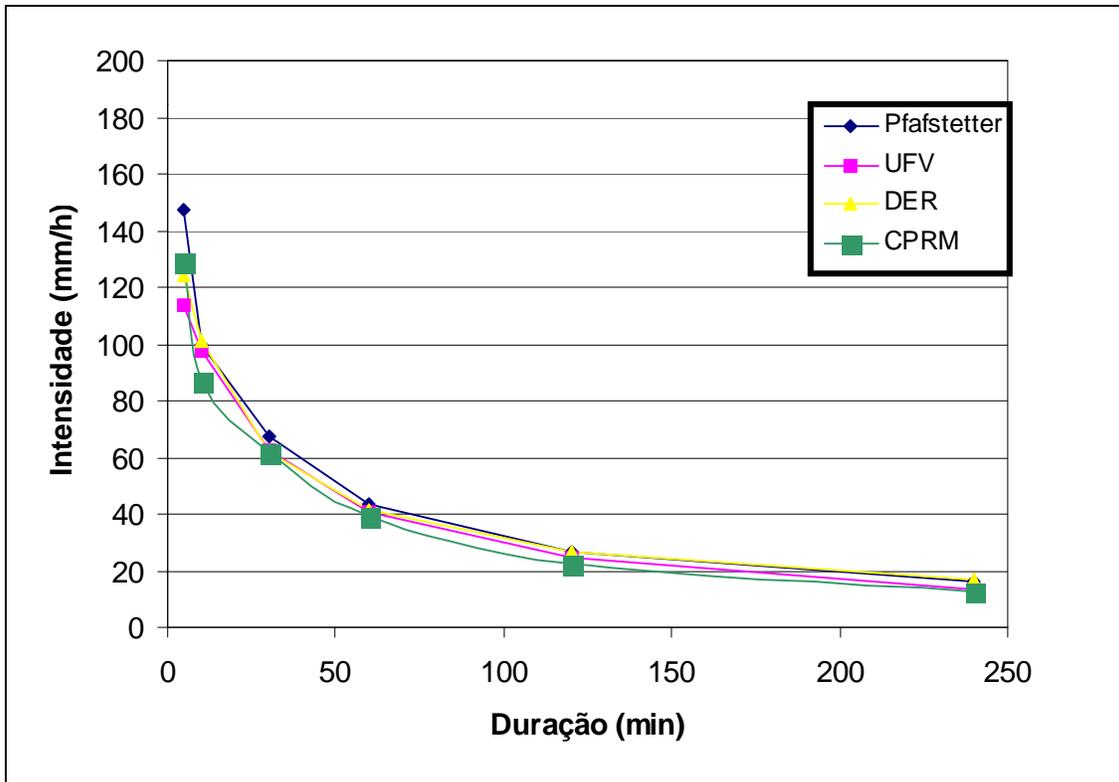


Figura 5 – Diagrama comparativo das equações IDF de Nova Friburgo (região 2) TR = 2 anos.

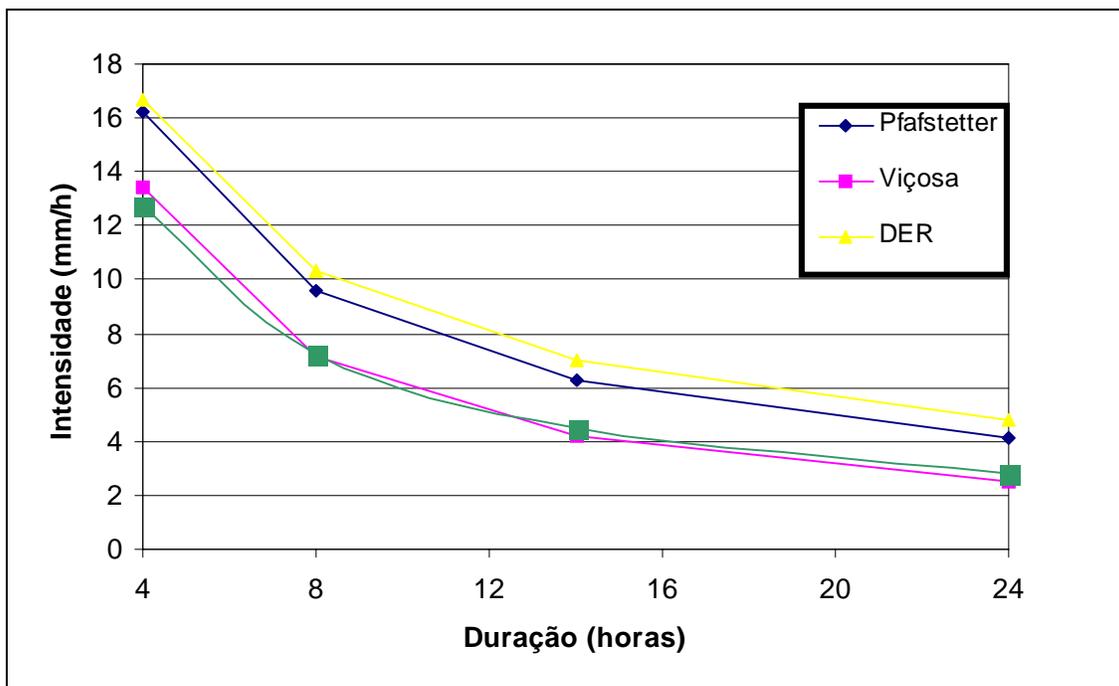
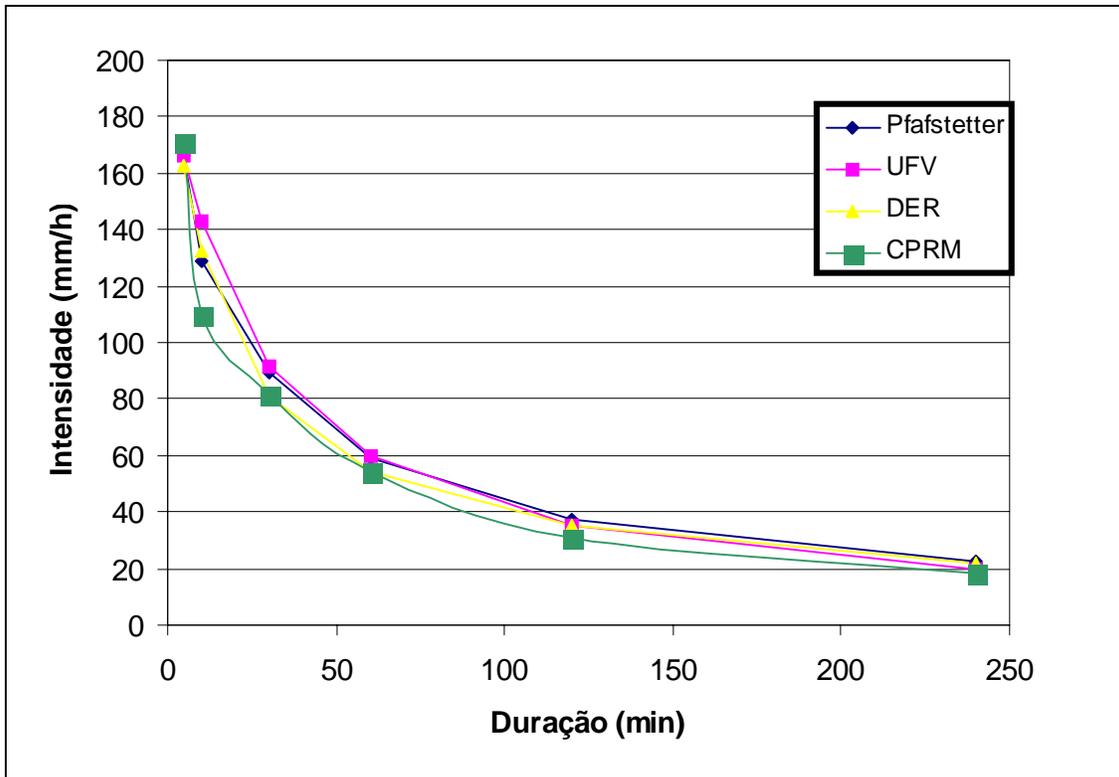


Figura 6 – Diagrama comparativo das equações IDF de Nova Friburgo (região 2) TR = 10 anos.

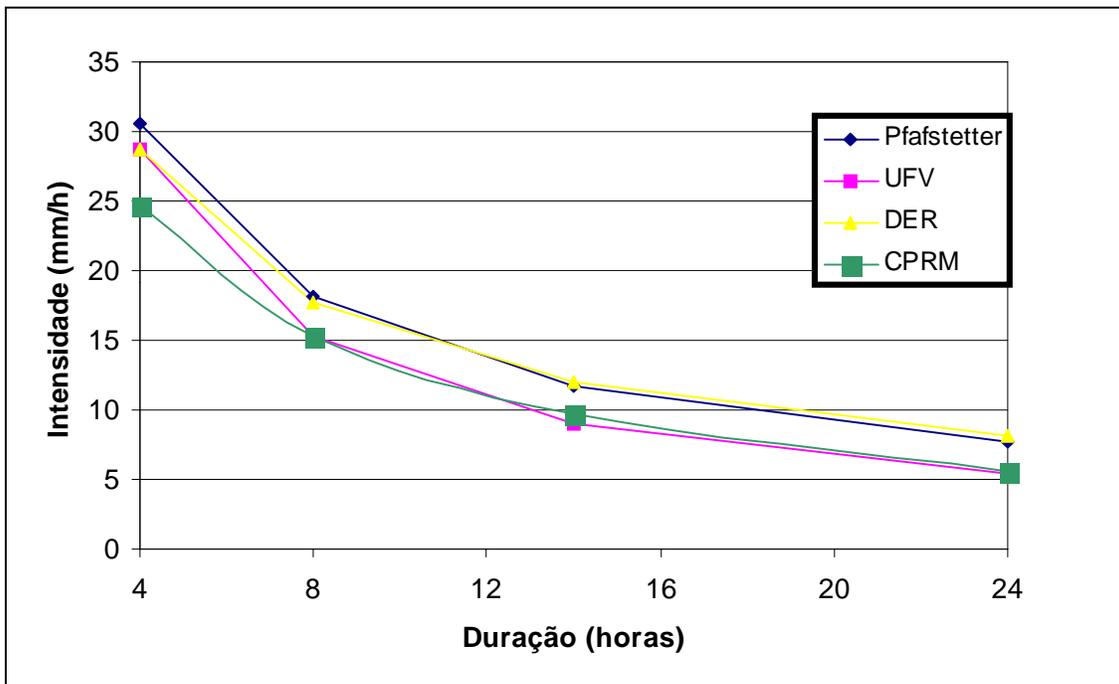
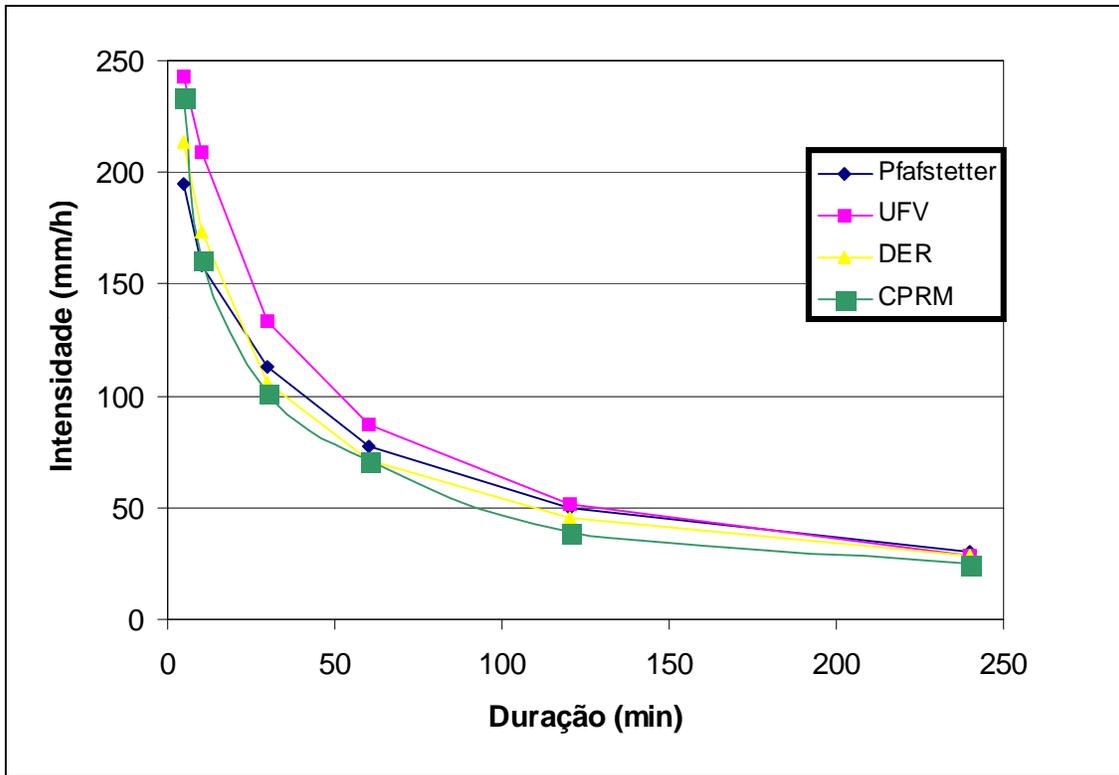


Figura 7 – Diagrama comparativo das equações IDF de Nova Friburgo (região 2) TR = 50 anos.

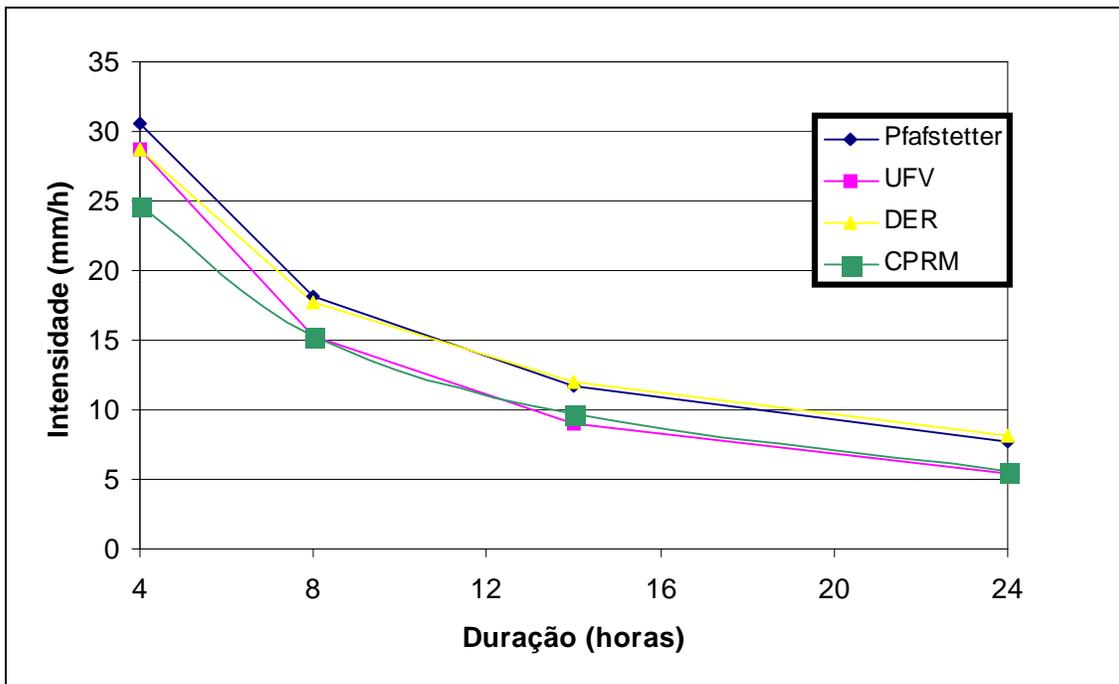
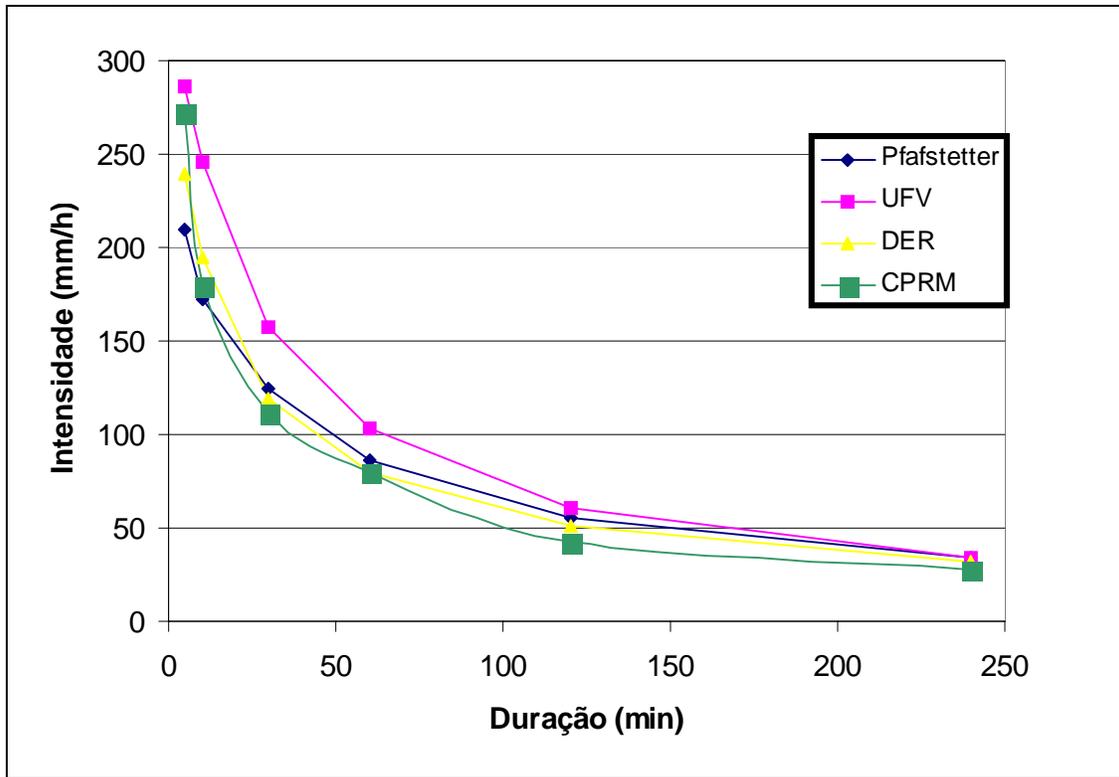


Figura 8 – Diagrama comparativo das equações IDF de Nova Friburgo (região 2) TR = 100 anos.

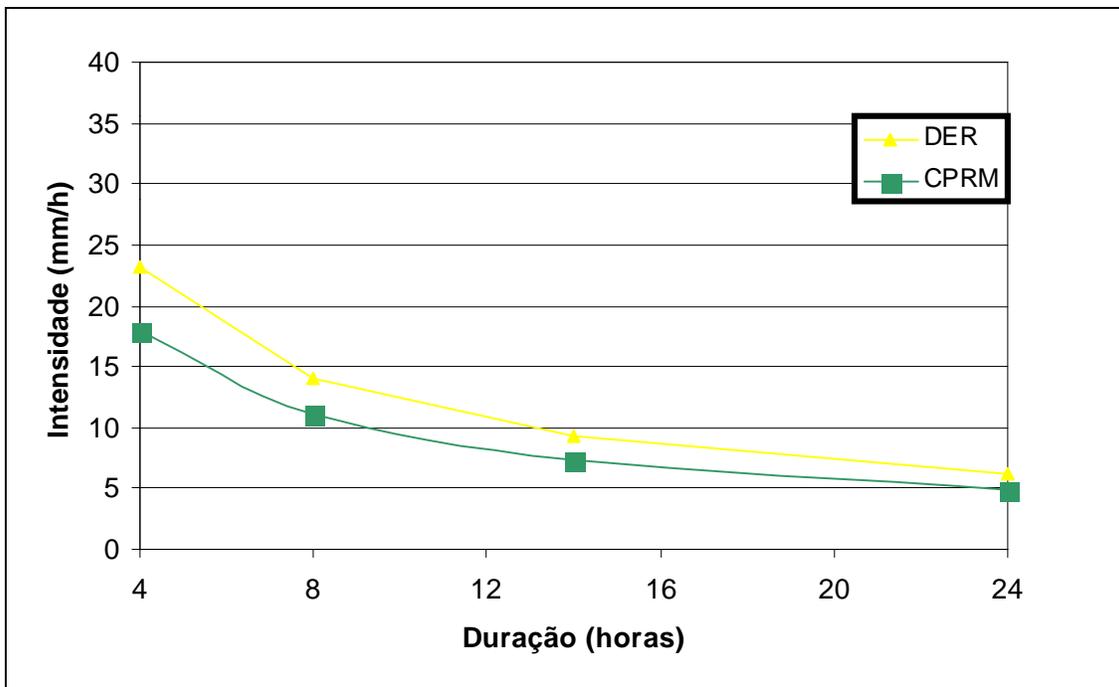
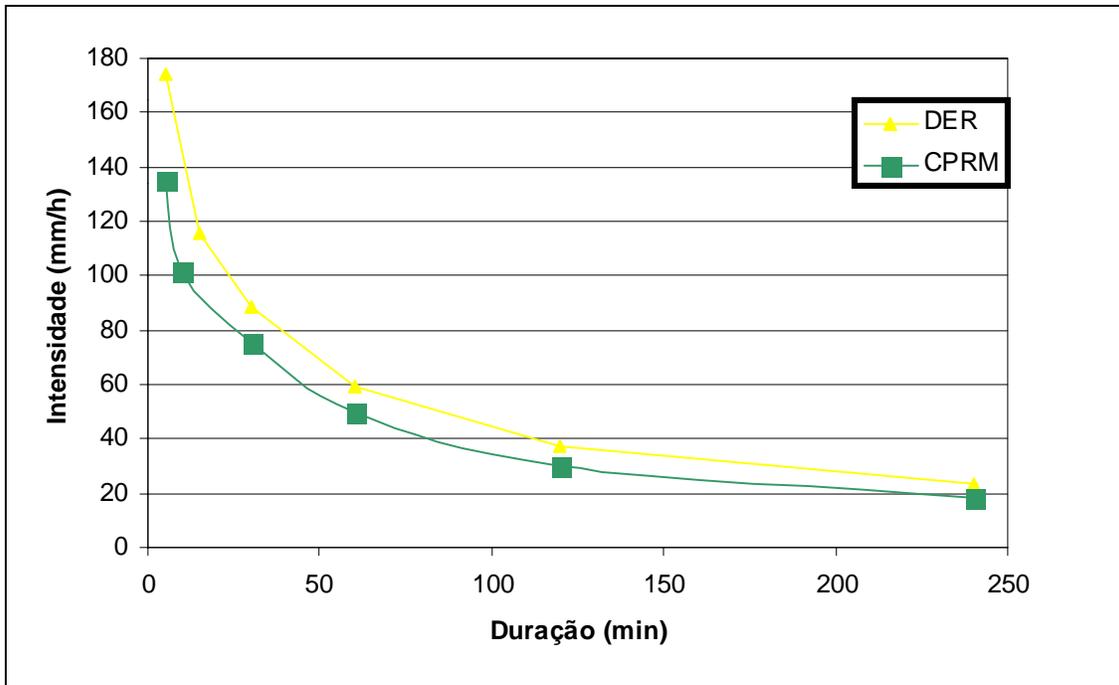


Figura 9 – Diagrama comparativo das equações IDF de Xerém (região 3) TR = 2 anos.

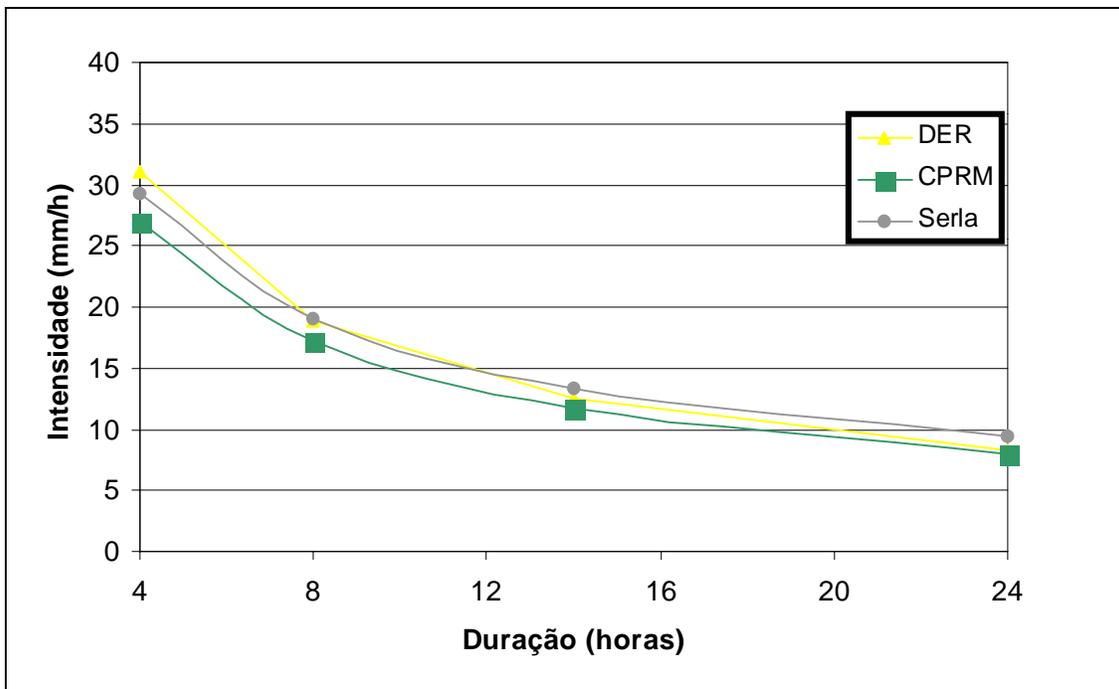
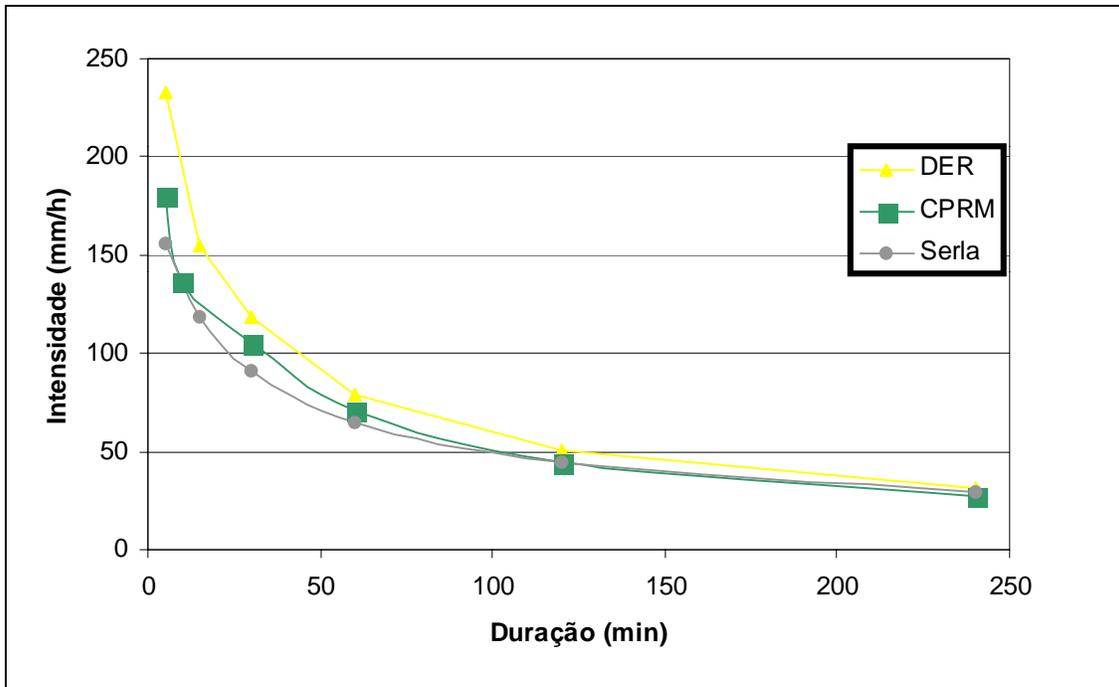


Figura 10 – Diagrama comparativo das equações IDF de Xerém (região 3) TR = 10 anos.

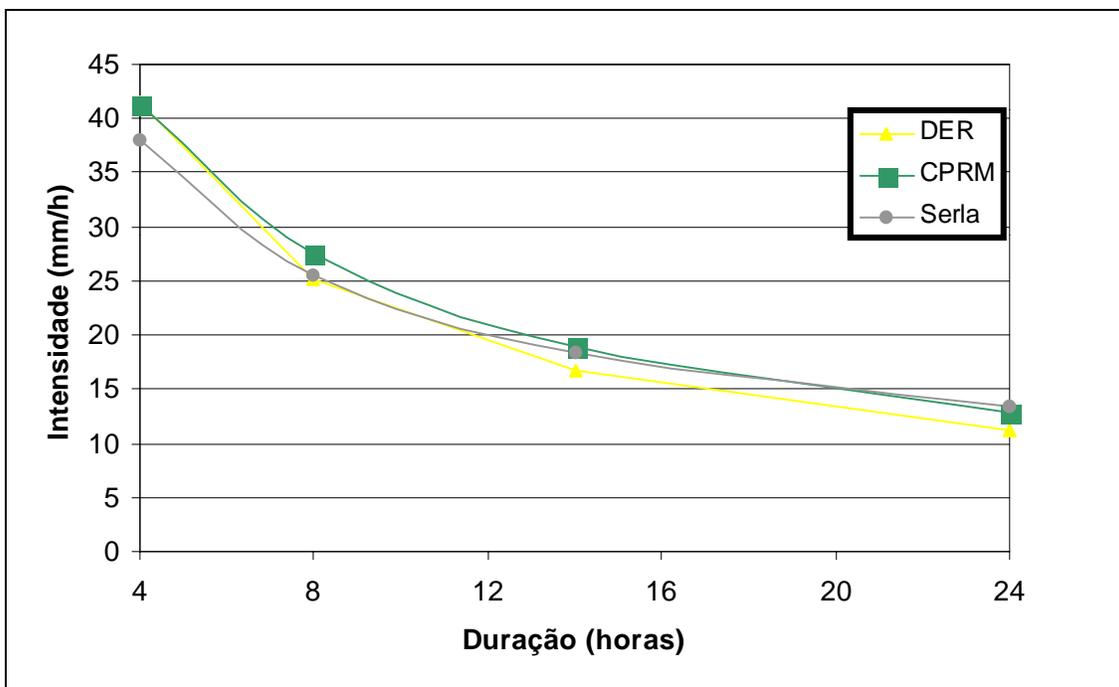
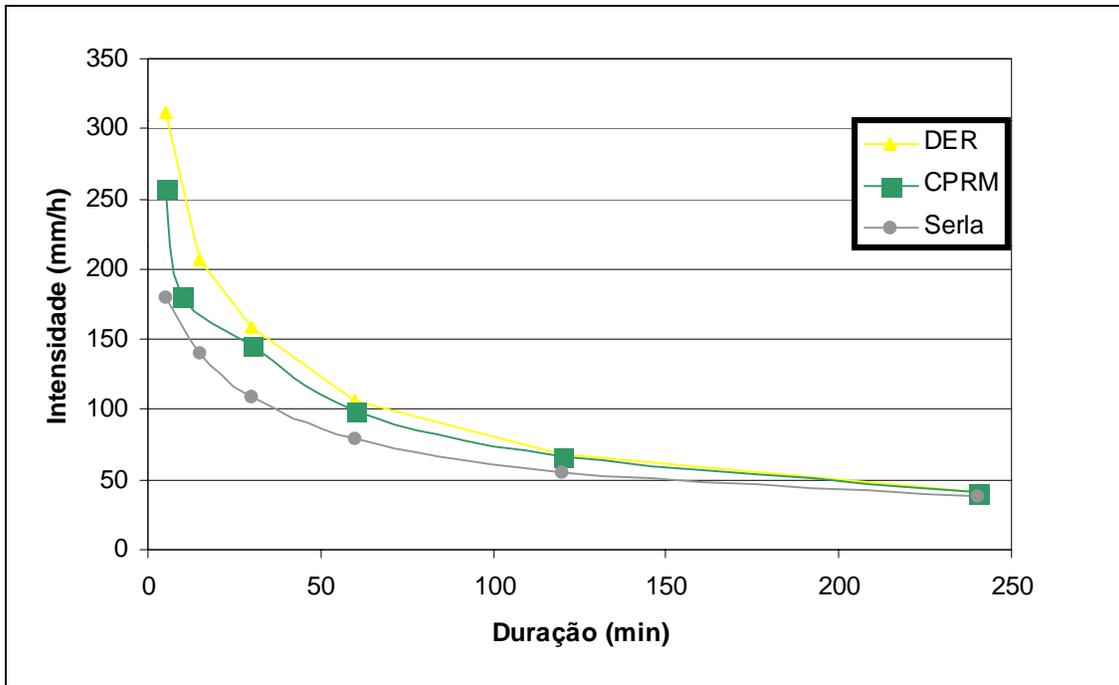


Figura 11 – Diagrama comparativo das equações IDF de Xerém (região 3) TR = 50 anos.

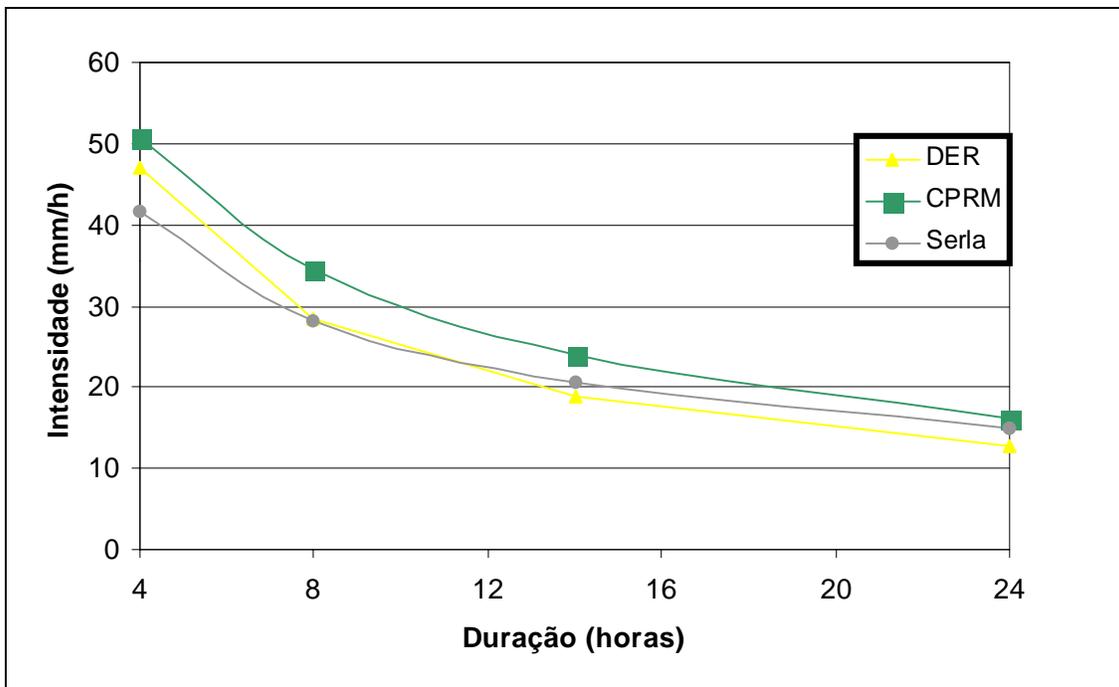
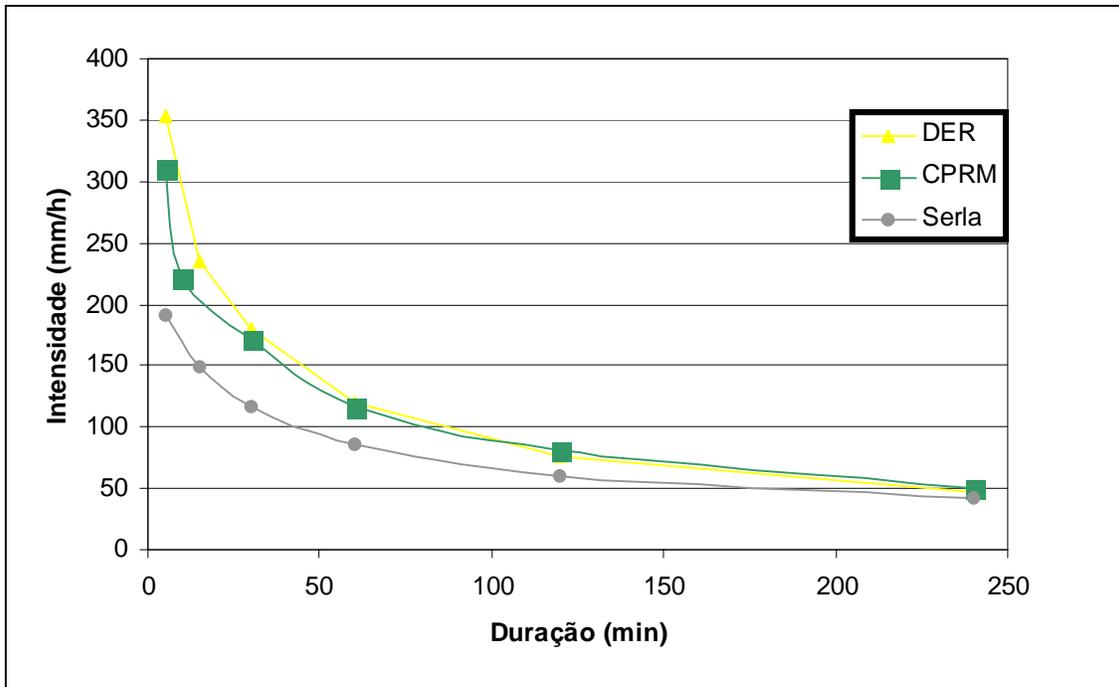


Figura 12 – Diagrama comparativo das equações IDF de Xerém (região 3) TR = 100 anos.

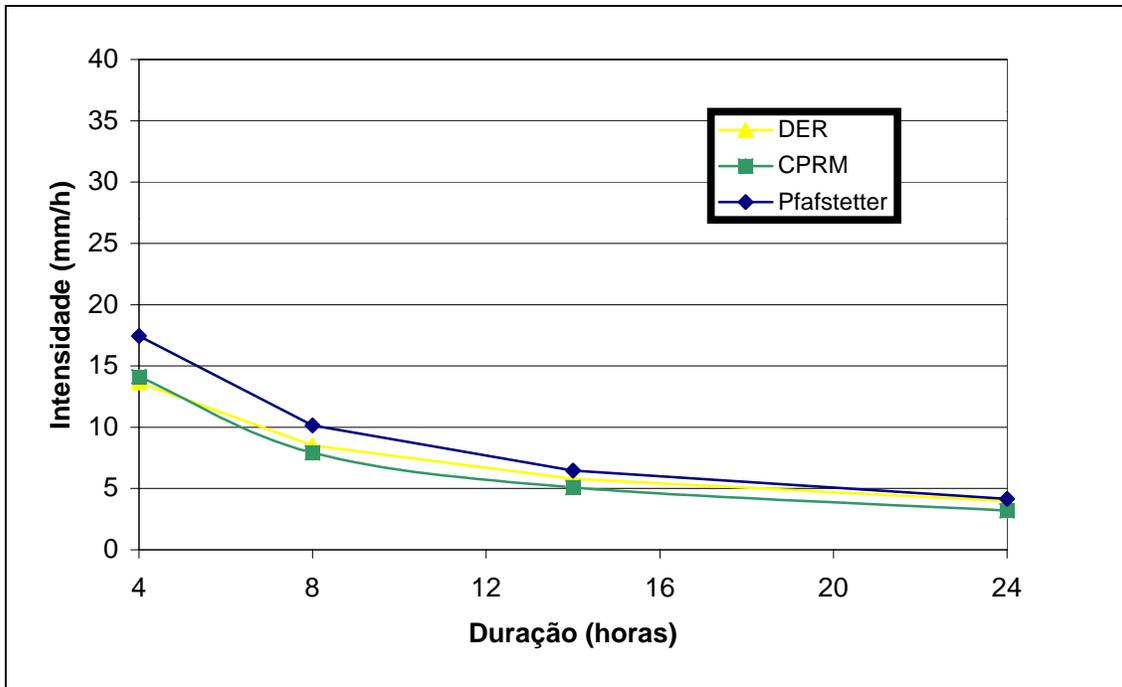
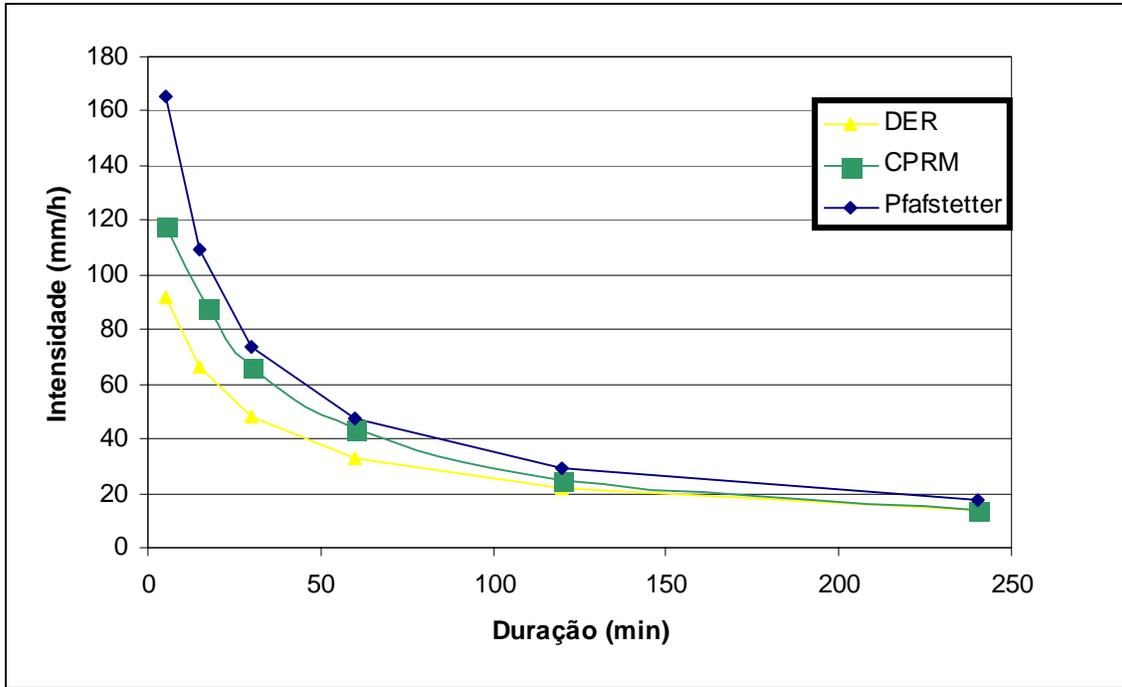


Figura 13 – Diagrama comparativo das equações IDF de Bangu (região 3) TR = 2 anos.

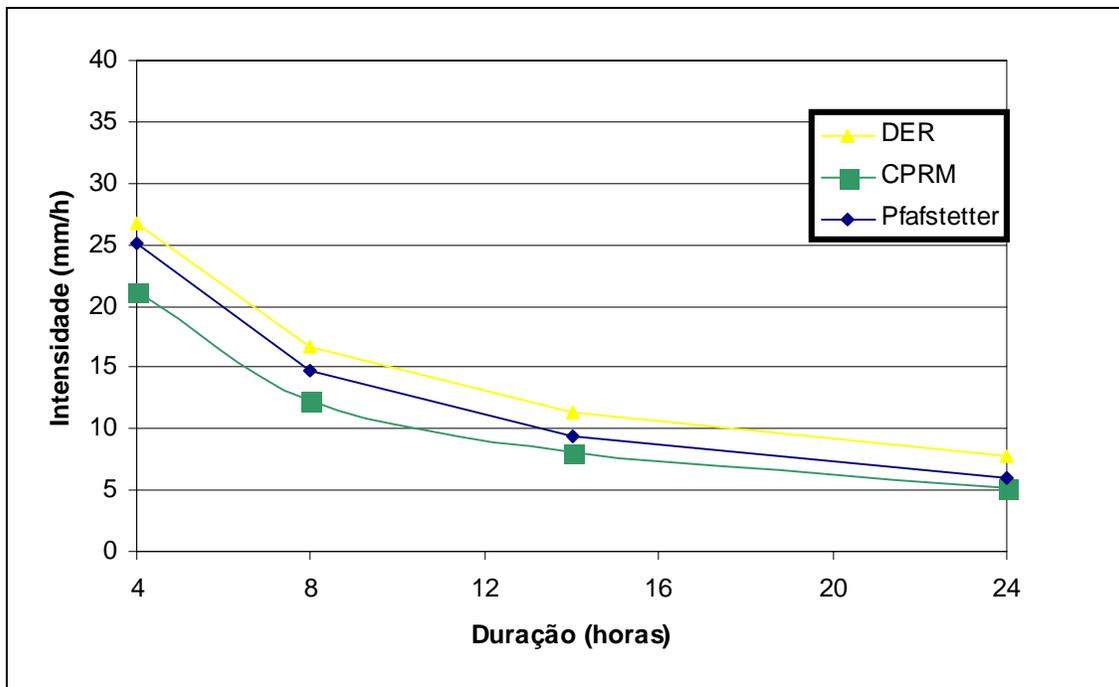
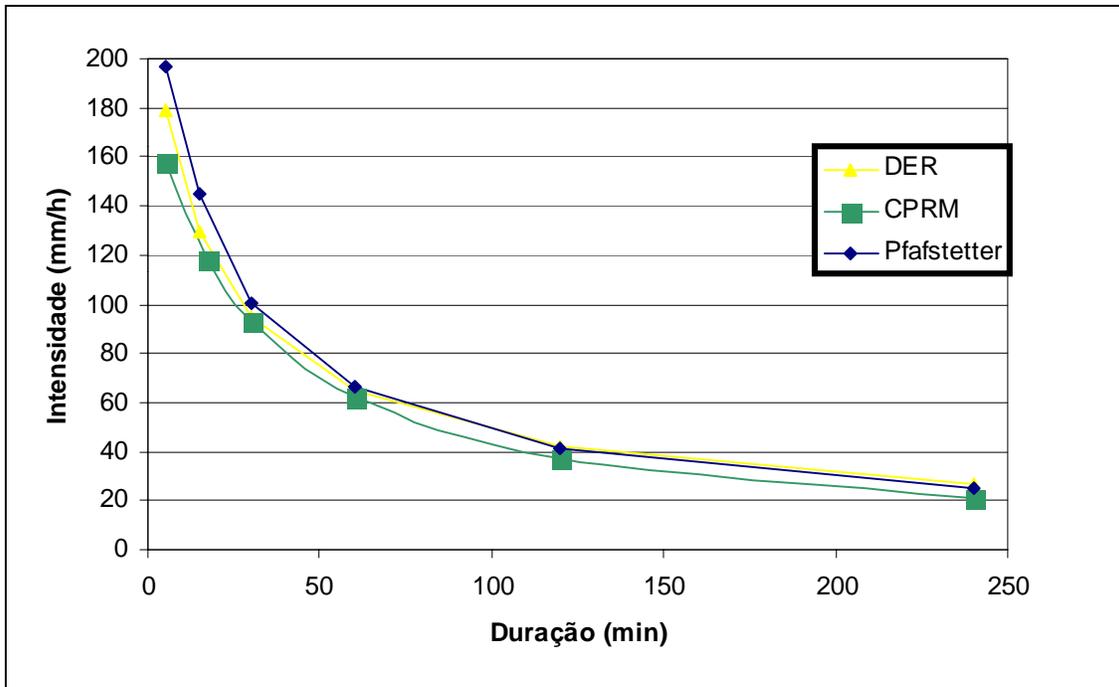


Figura 14 – Diagrama comparativo das equações IDF de Bangu (região 3) TR = 10 anos.

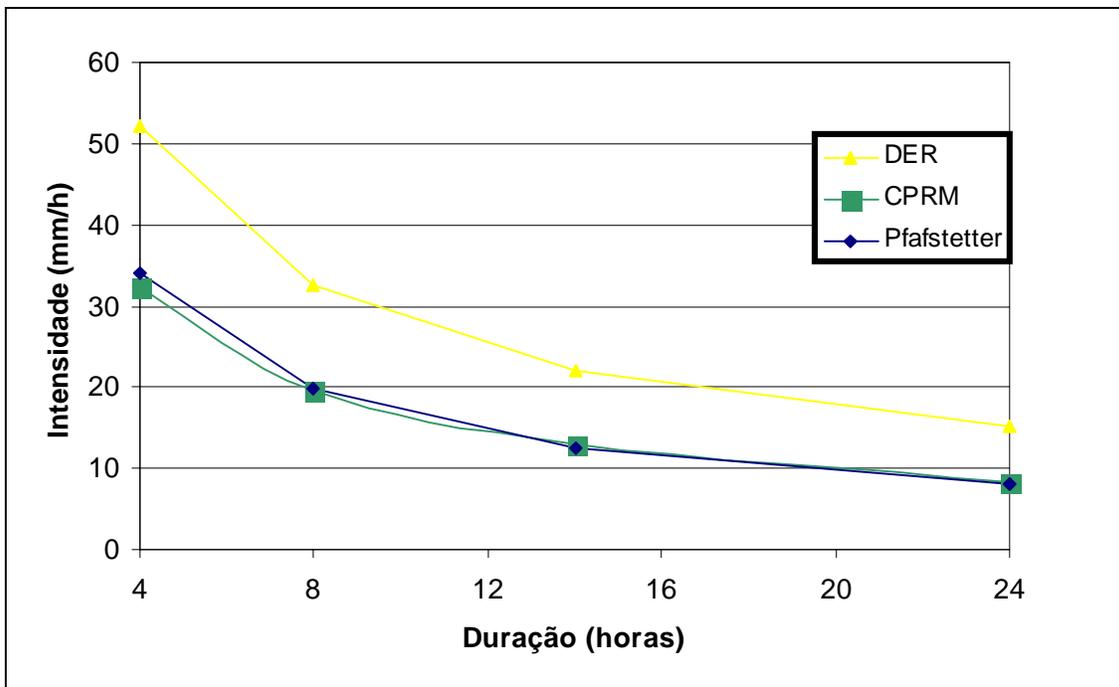
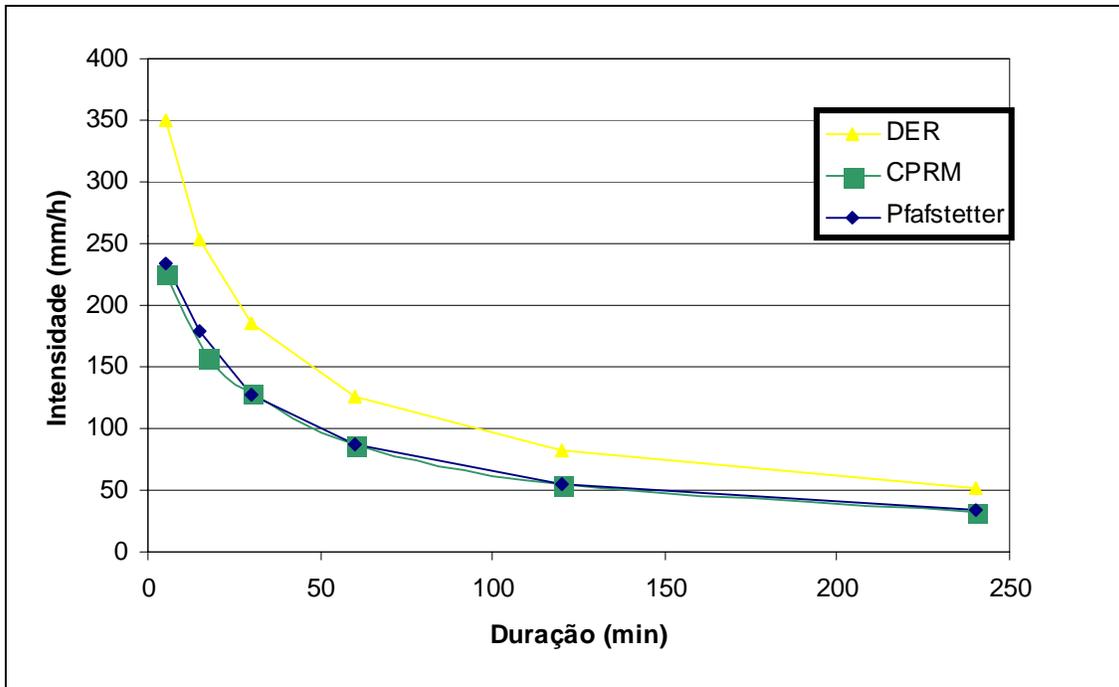


Figura 15 – Diagrama comparativo das equações IDF de Bangu (região 3) TR = 50 anos.

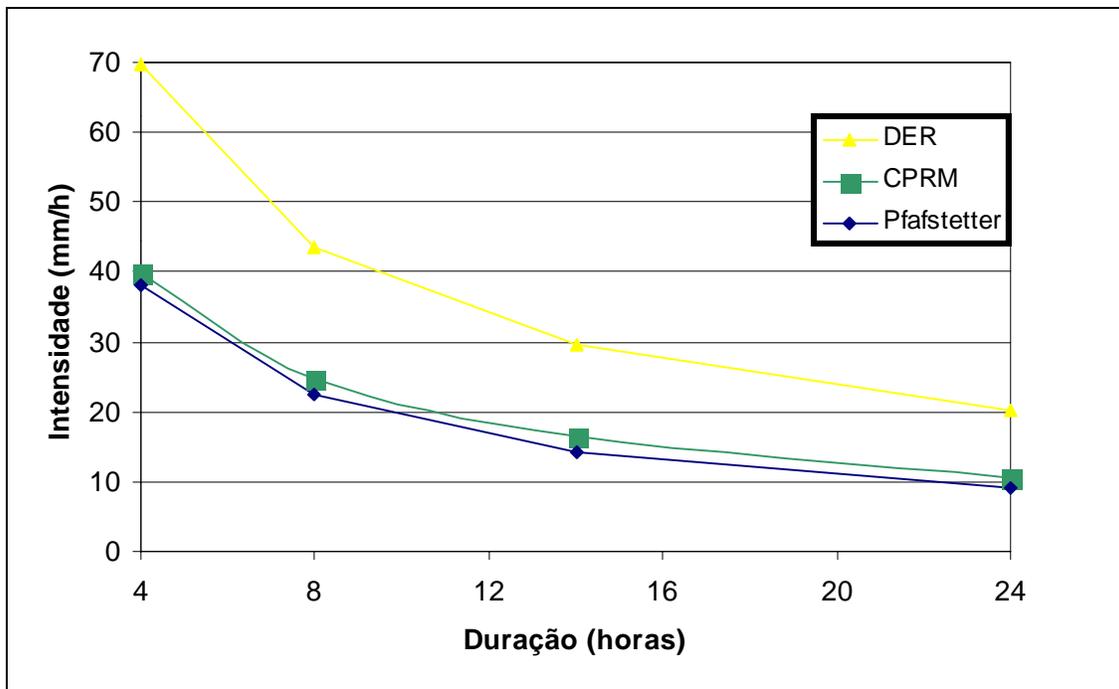
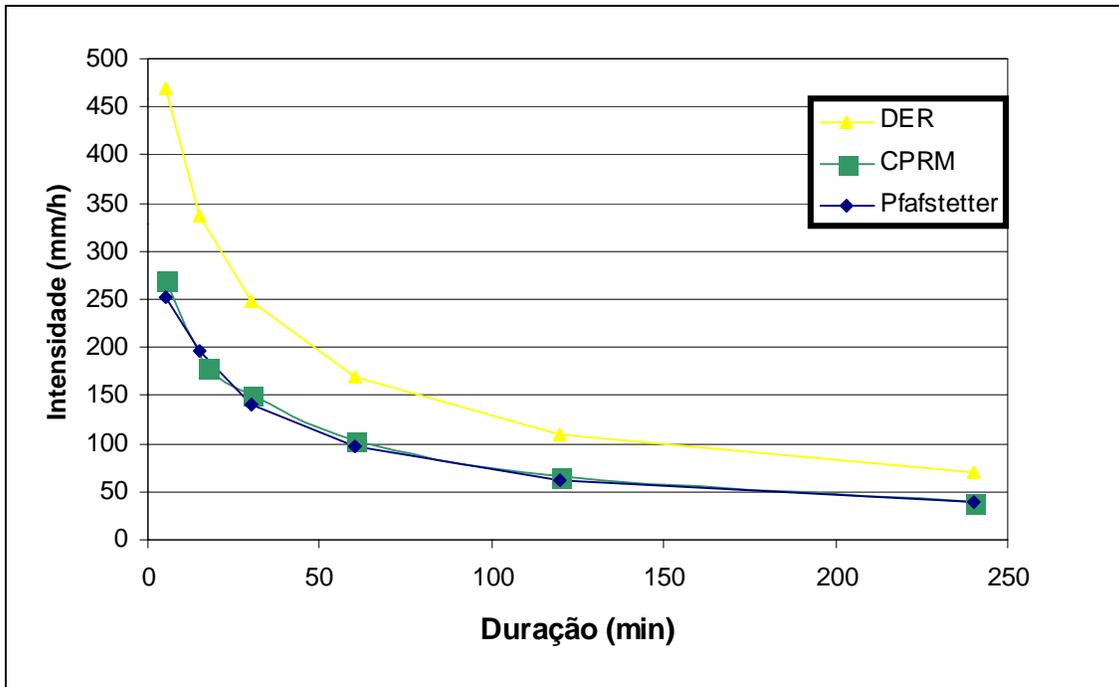


Figura 16 – Diagrama comparativo das equações IDF de Bangu (região 3) TR = 100 anos.

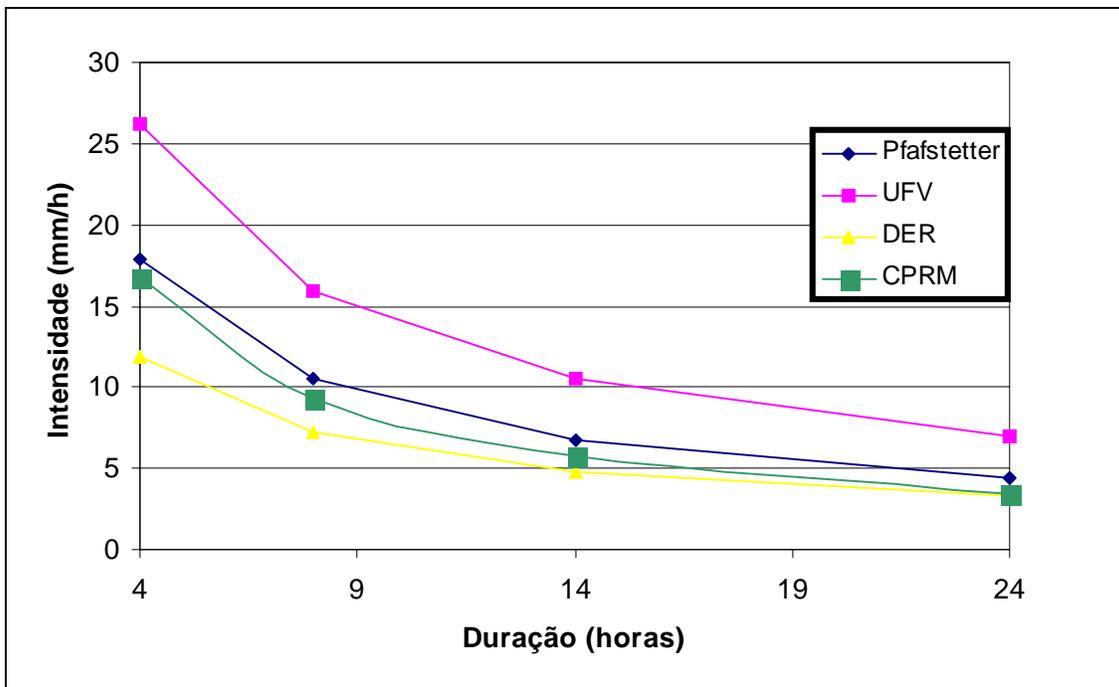
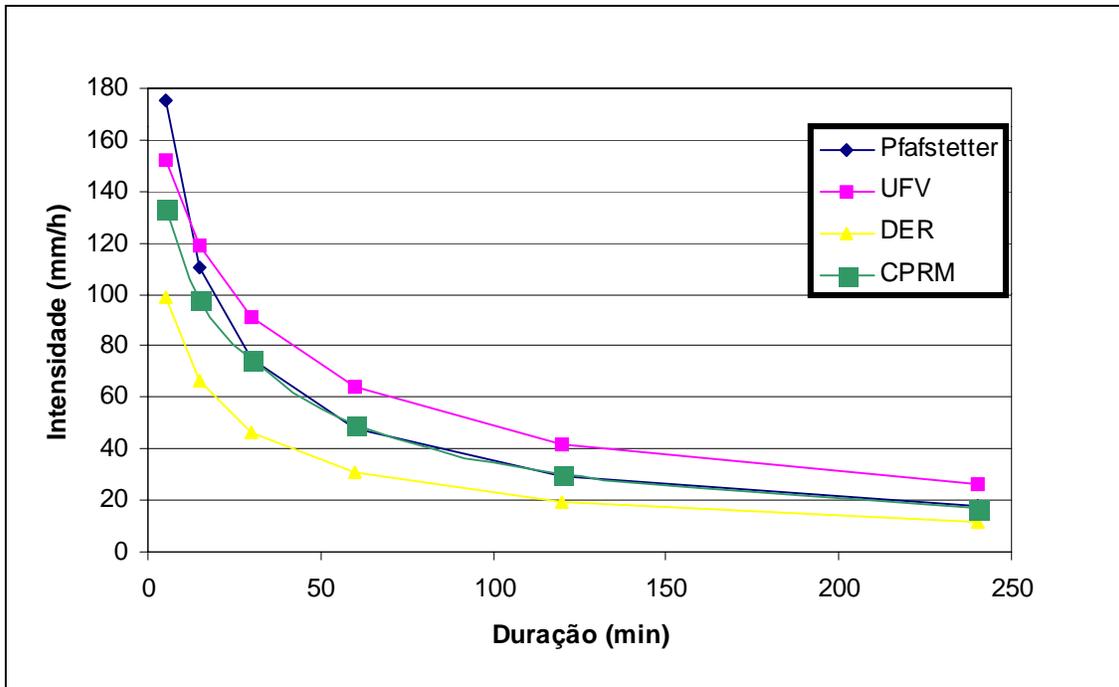


Figura 17 – Diagrama comparativo das equações IDF de Resende (região 4) TR = 2 anos.

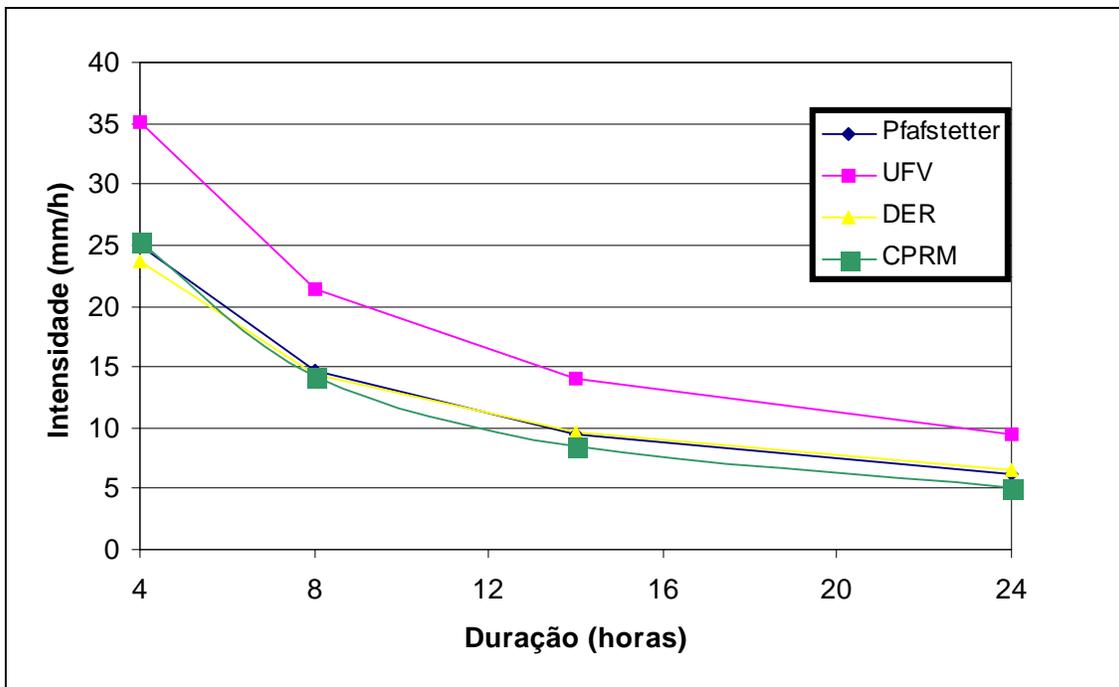
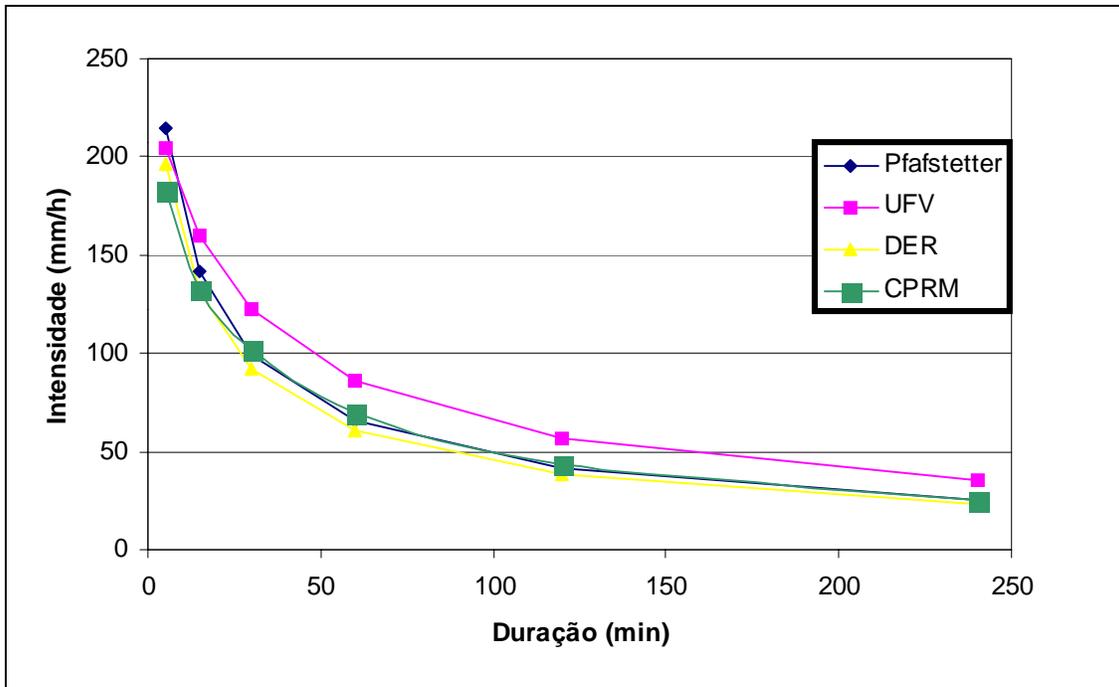


Figura 18 – Diagrama comparativo das equações IDF de Resende (região 4) TR = 10 anos.

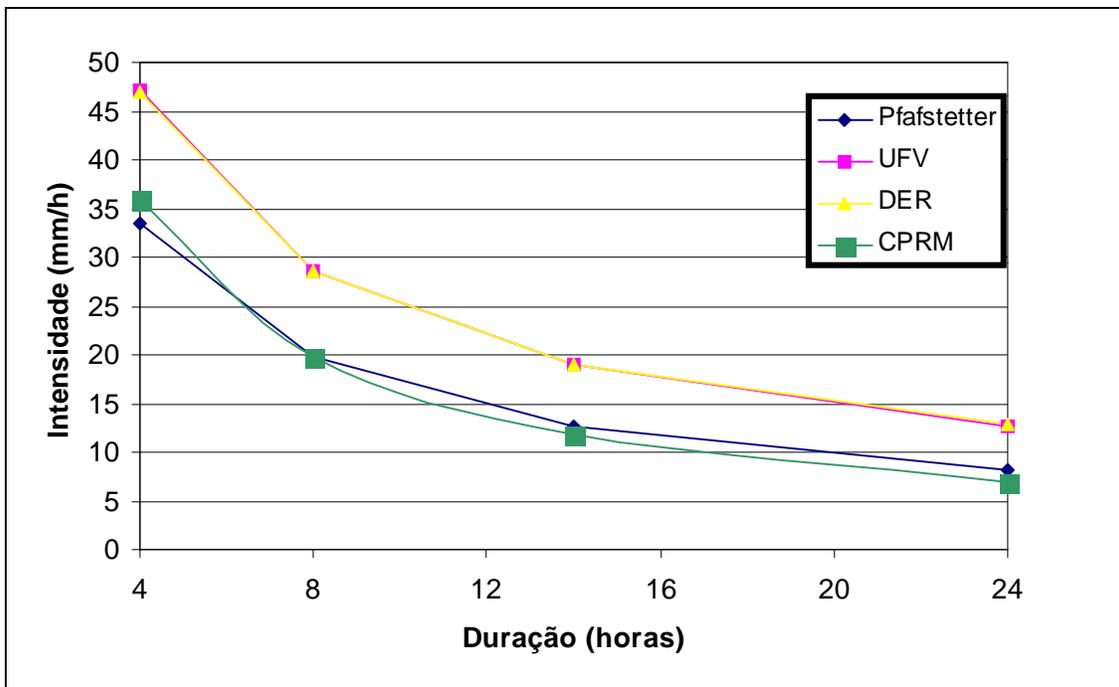
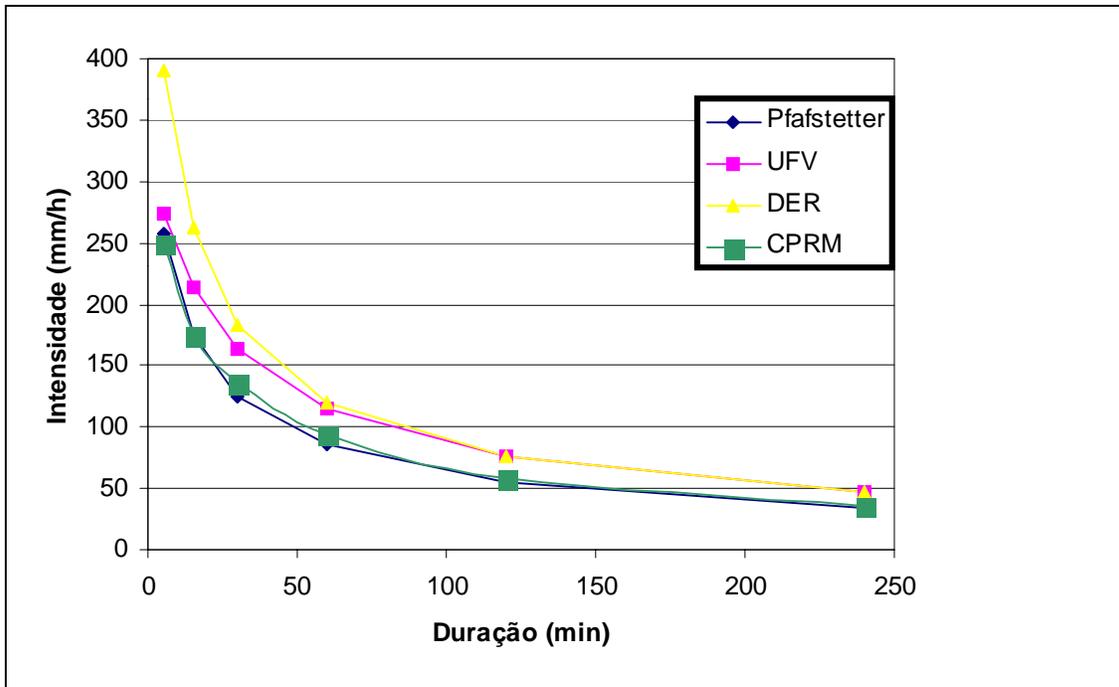


Figura 19 – Diagrama comparativo das equações IDF de Resende (região 4) TR = 50 anos.

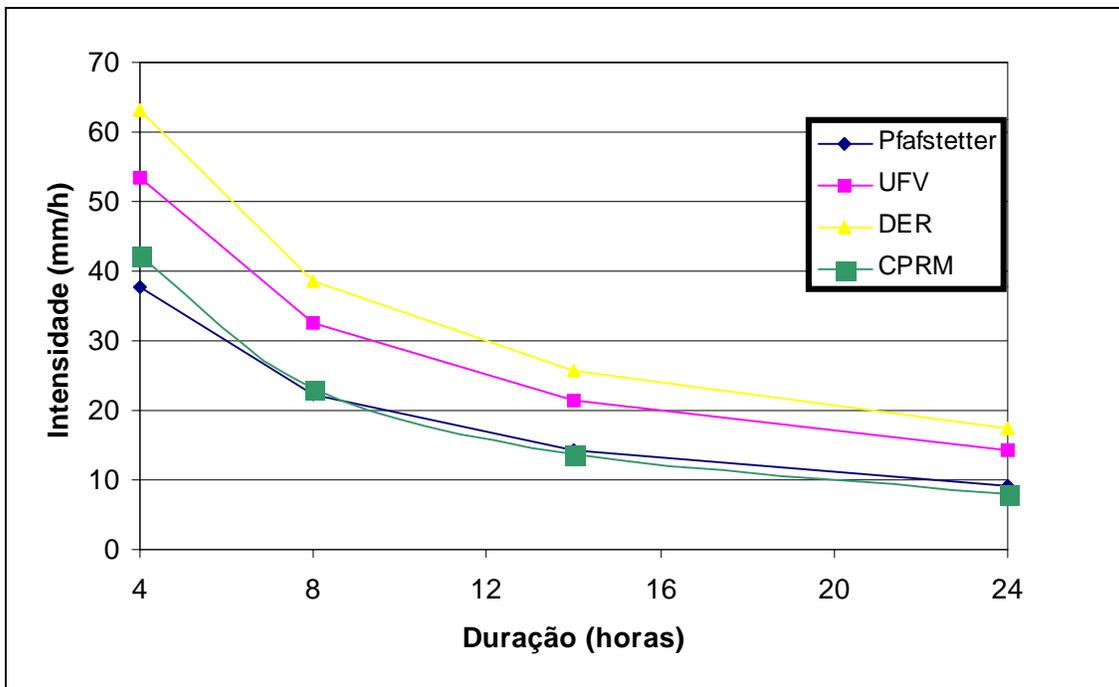
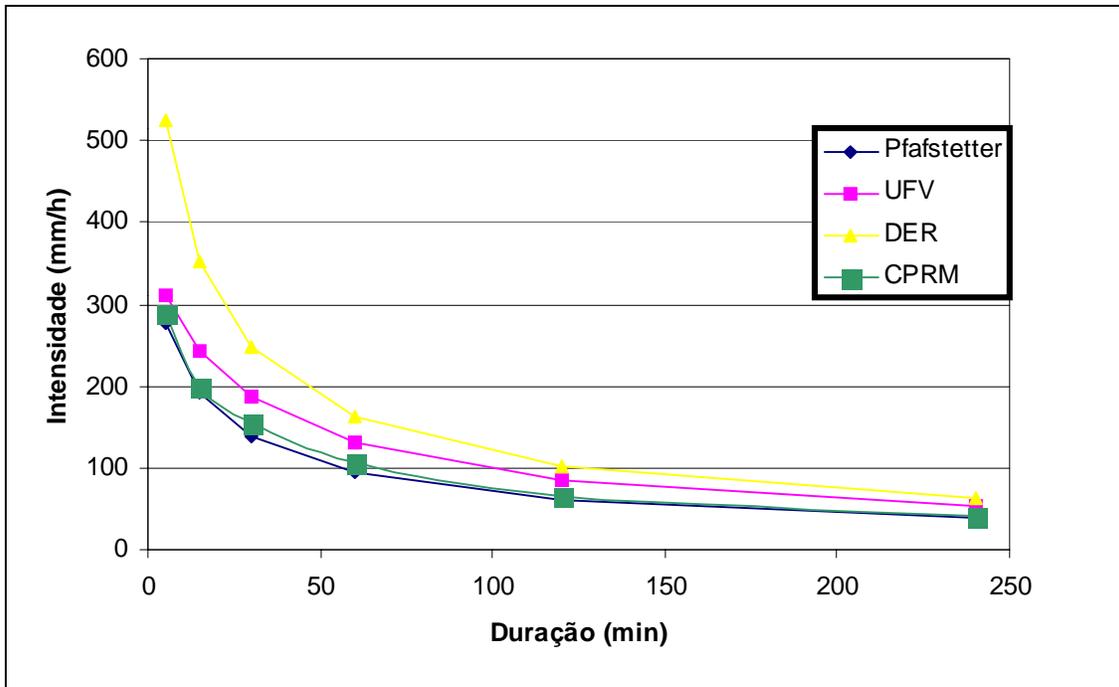


Figura 20 – Diagrama comparativo das equações IDF de Resende (região 4) TR = 100 anos.

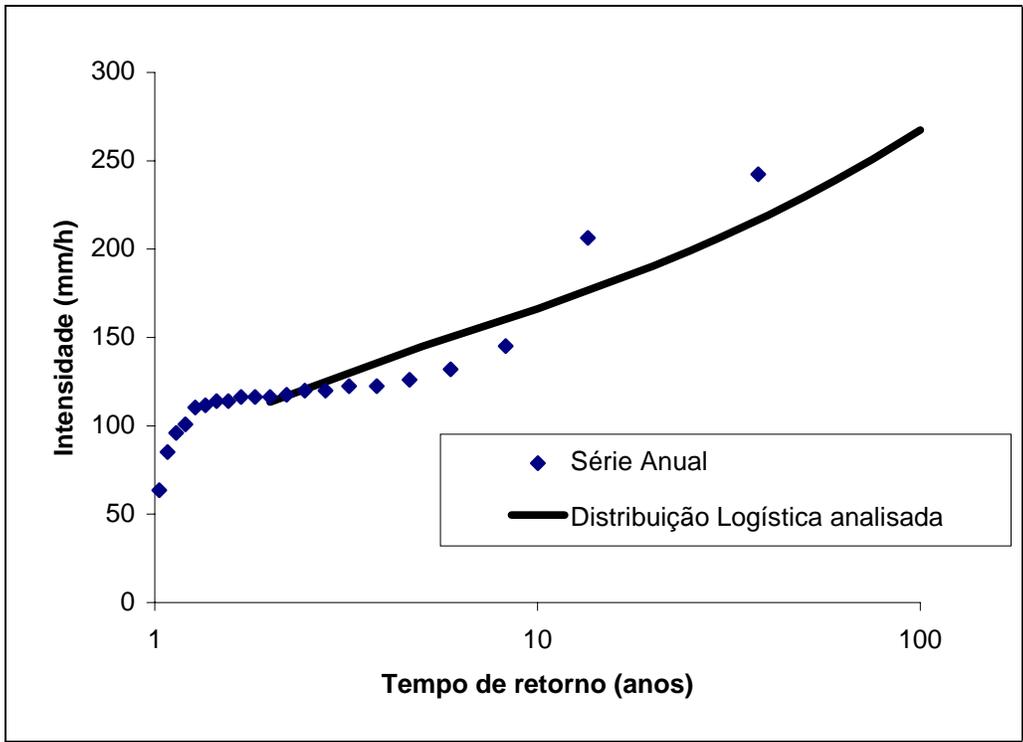


Figura 21 – Curva de frequência para a duração 5 minutos – Campos.

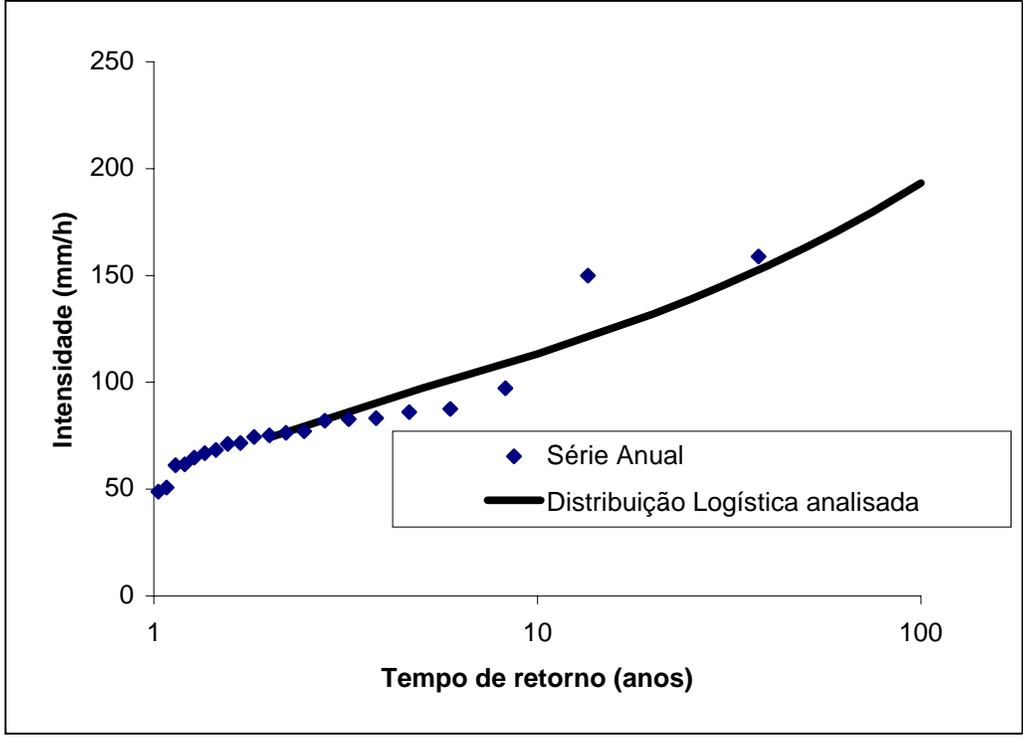


Figura 22 – Curva de frequência para a duração 15 minutos – Campos.

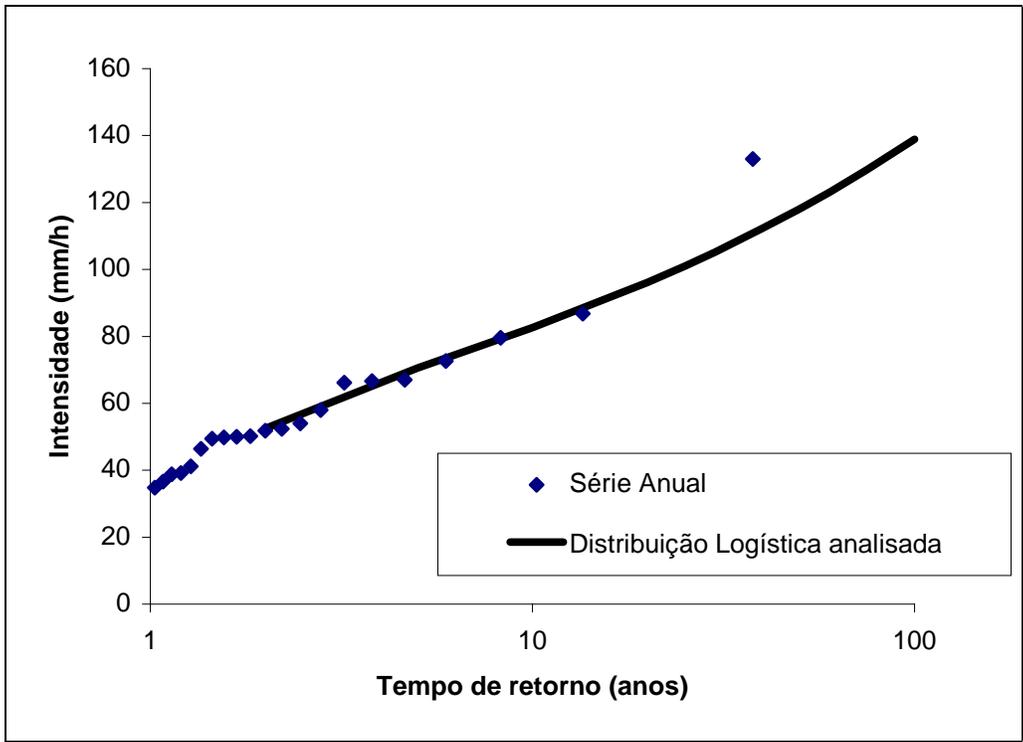


Figura 23 – Curva de frequência para a duração 30 minutos – Campos.

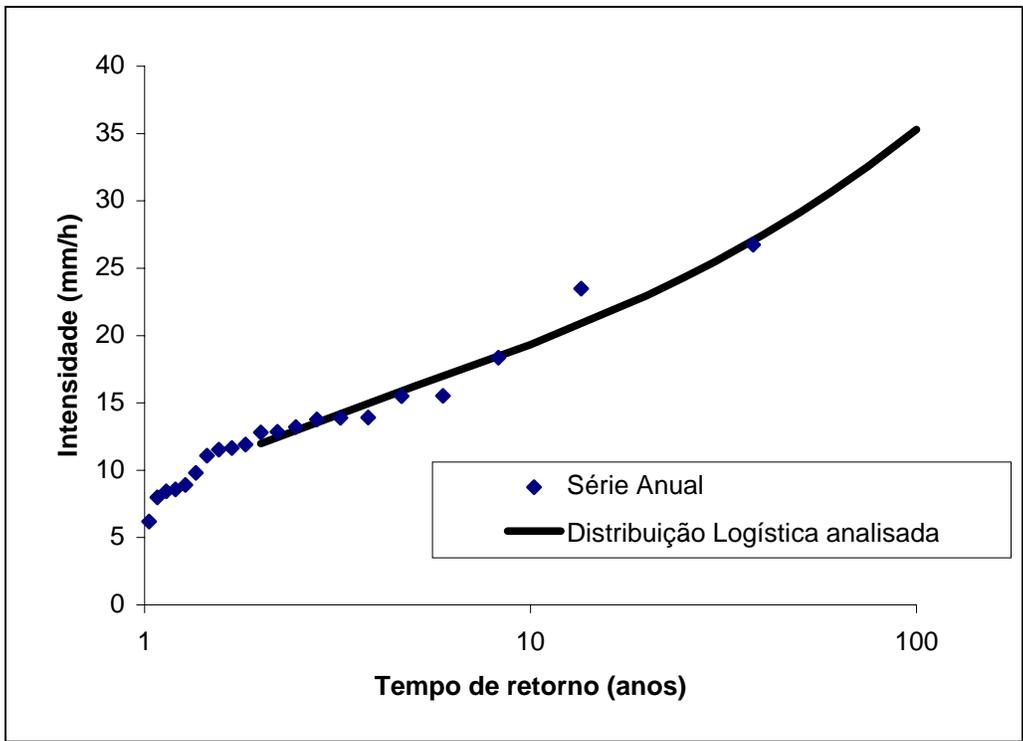


Figura 24 – Curva de frequência para a duração 4 horas – Campos.

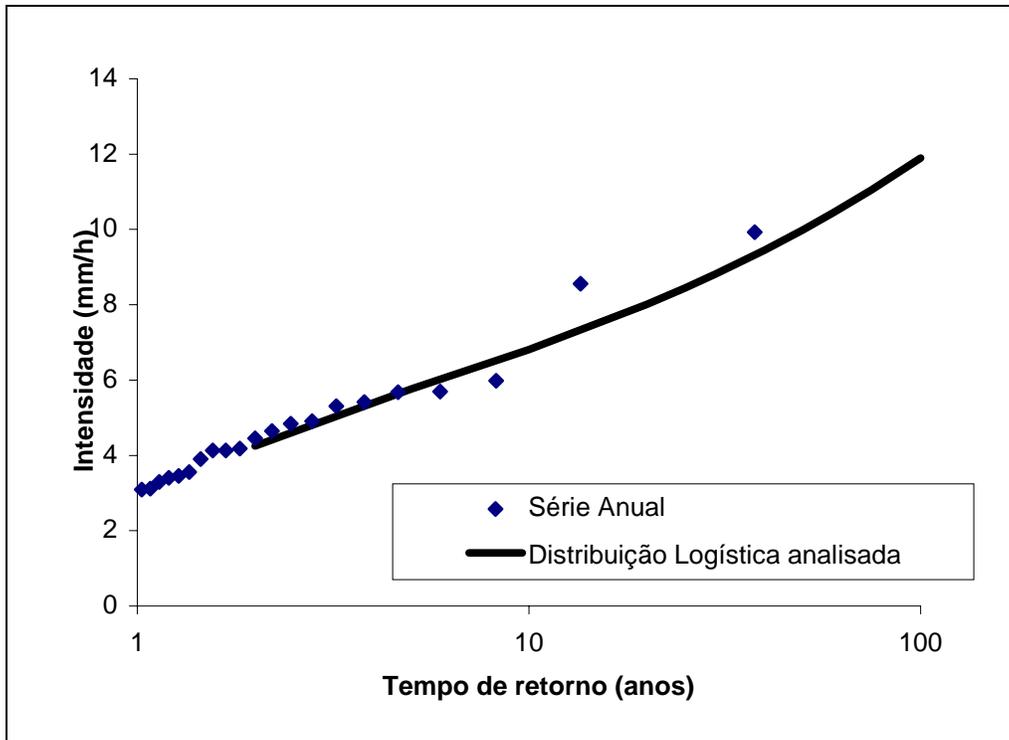


Figura 25 – Curva de freqüência para a duração 14 horas – Campos.

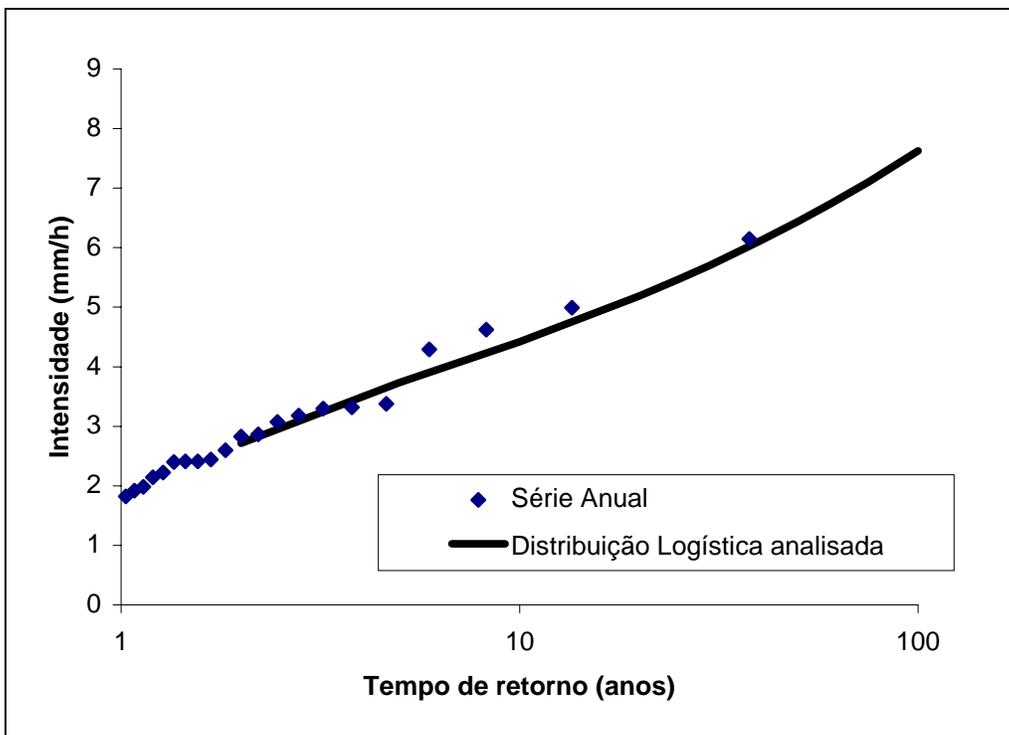


Figura 26 – Curva de freqüência para a duração 24 horas – Campos.

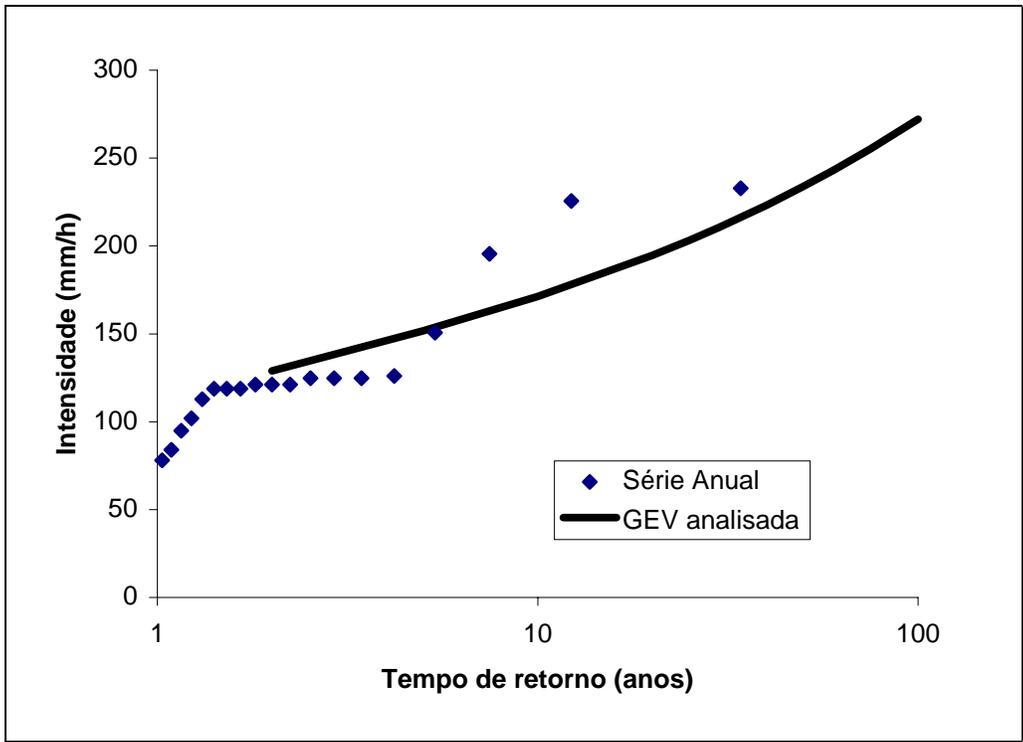


Figura 27 – Curva de frequência para a duração 5 minutos – Nova Friburgo.

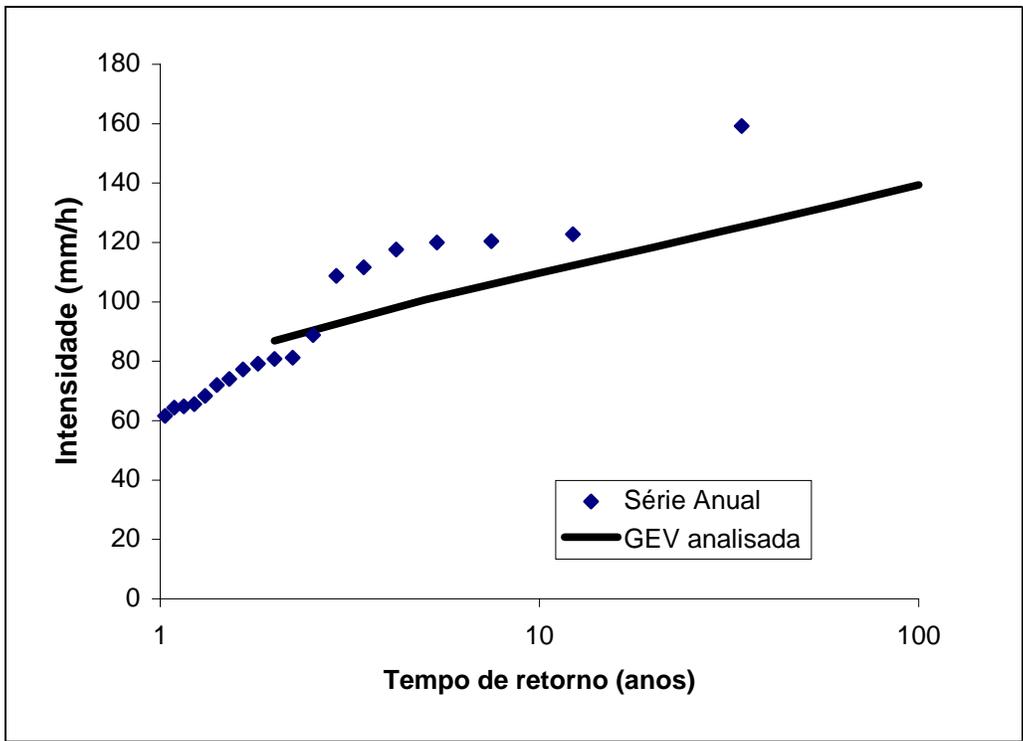


Figura 28 – Curva de frequência para a duração 15 minutos – Nova Friburgo.

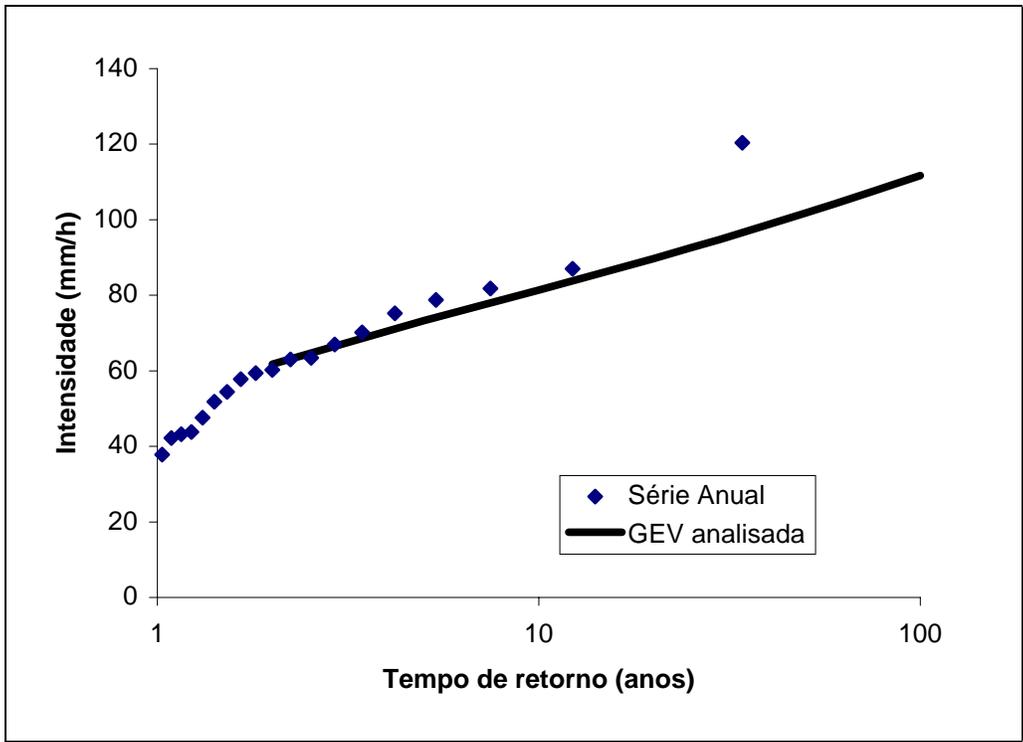


Figura 29 – Curva de frequência para a duração 30 minutos – Nova Friburgo.

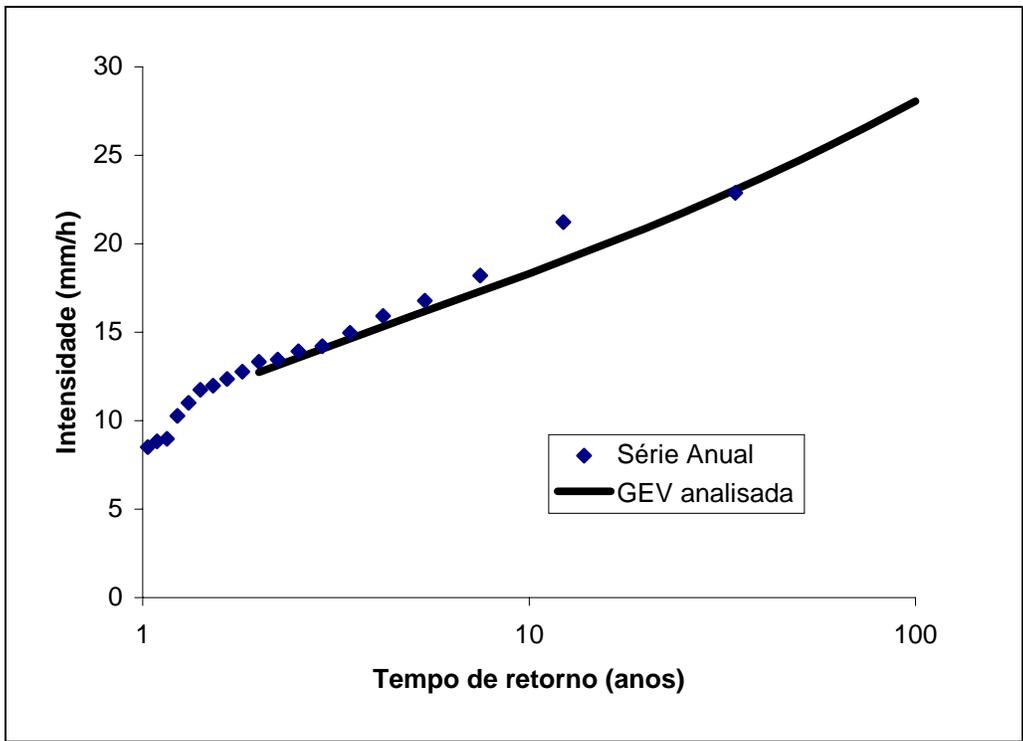


Figura 30 – Curva de frequência para a duração 4 horas – Nova Friburgo.

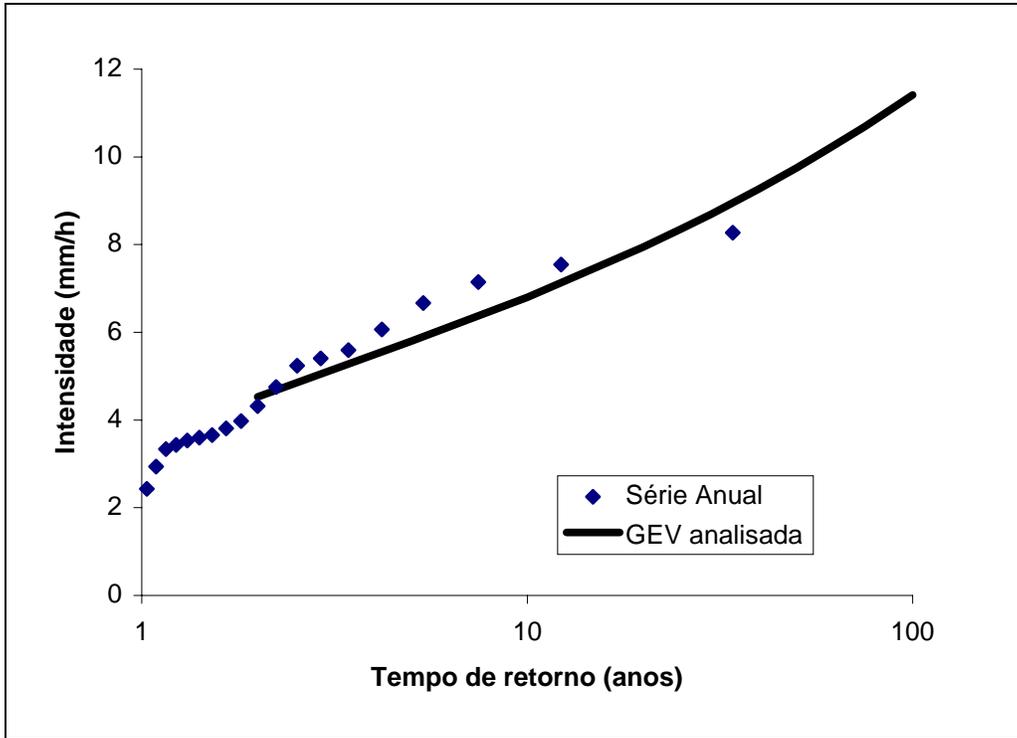


Figura 31 – Curva de frequência para a duração 14 horas – Nova Friburgo.

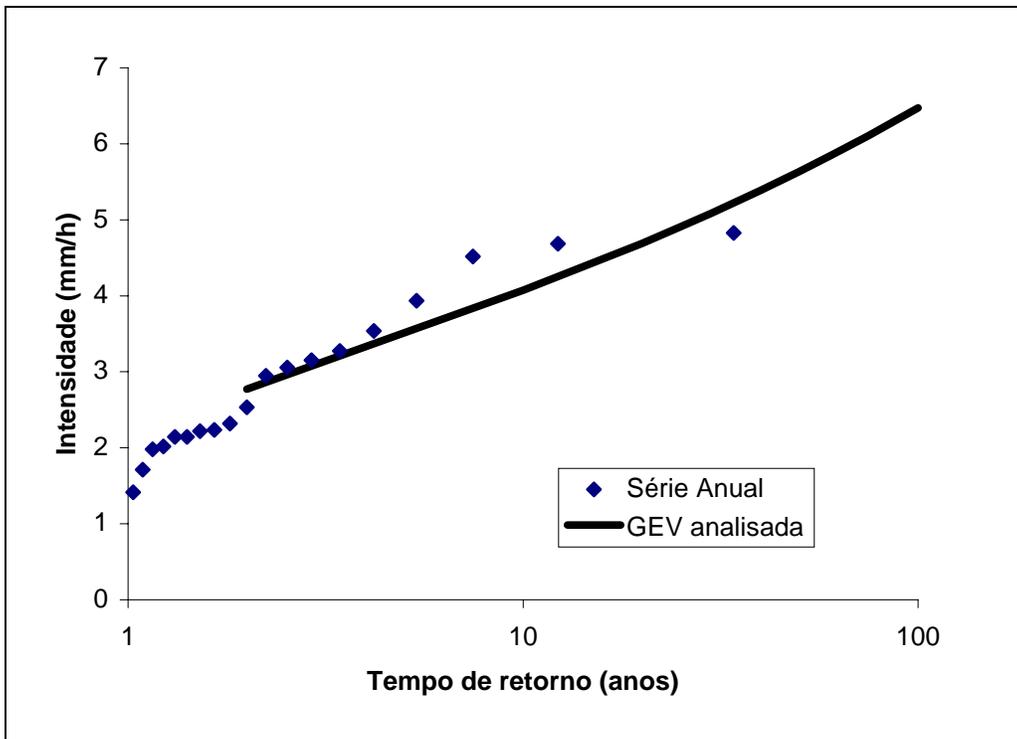


Figura 32 – Curva de frequência para a duração 24 horas – Nova Friburgo.

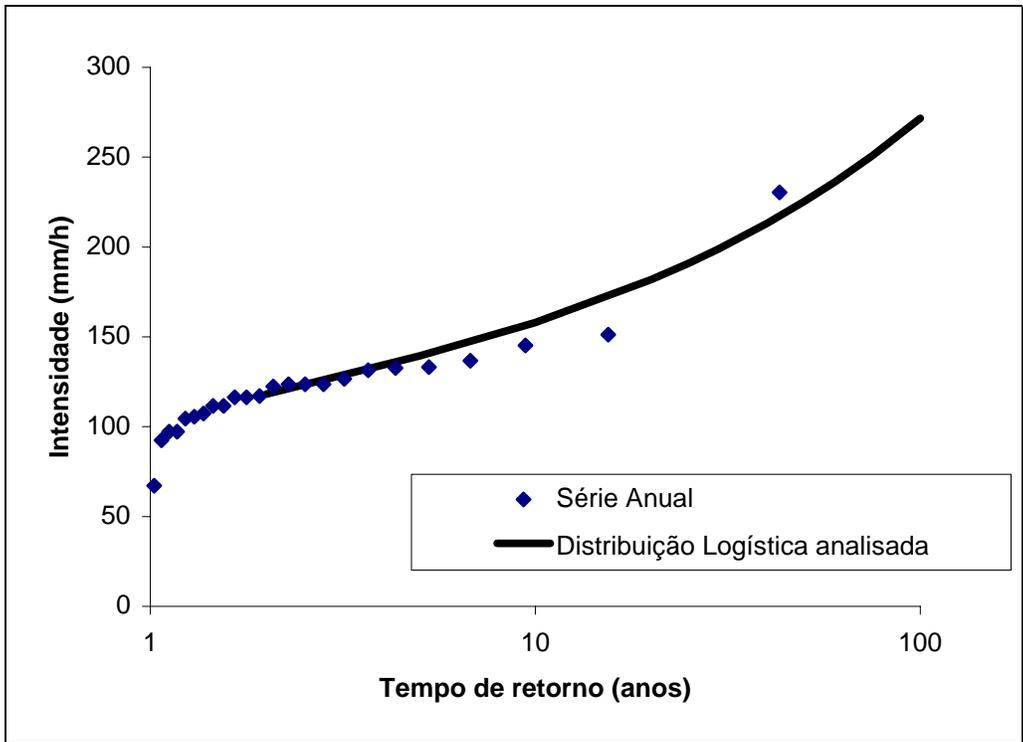


Figura 33 – Curva de frequência para a duração 5 minutos – Bangu.

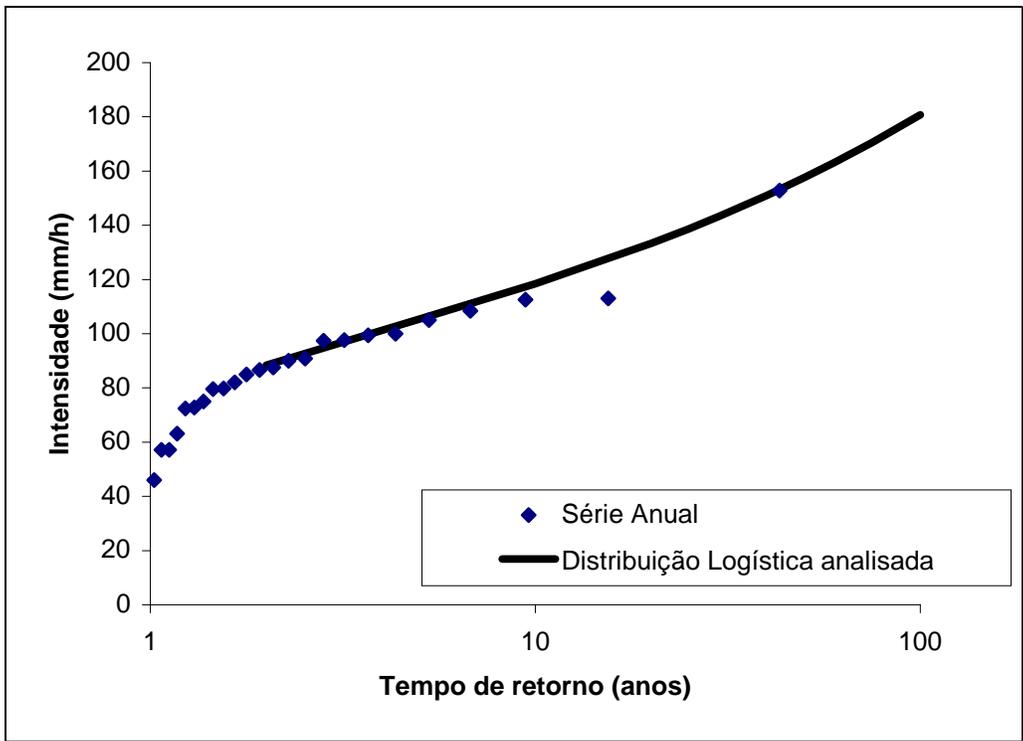


Figura 34 – Curva de frequência para a duração 15 minutos – Bangu.

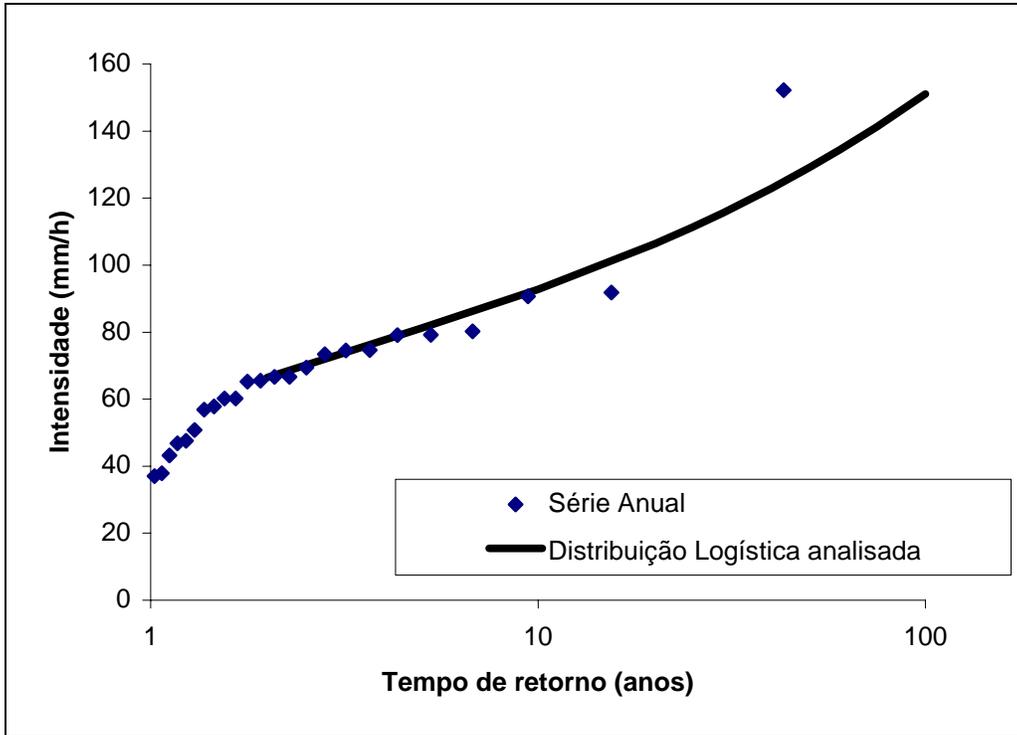


Figura 35 – Curva de frequência para a duração 30 minutos – Bangu.

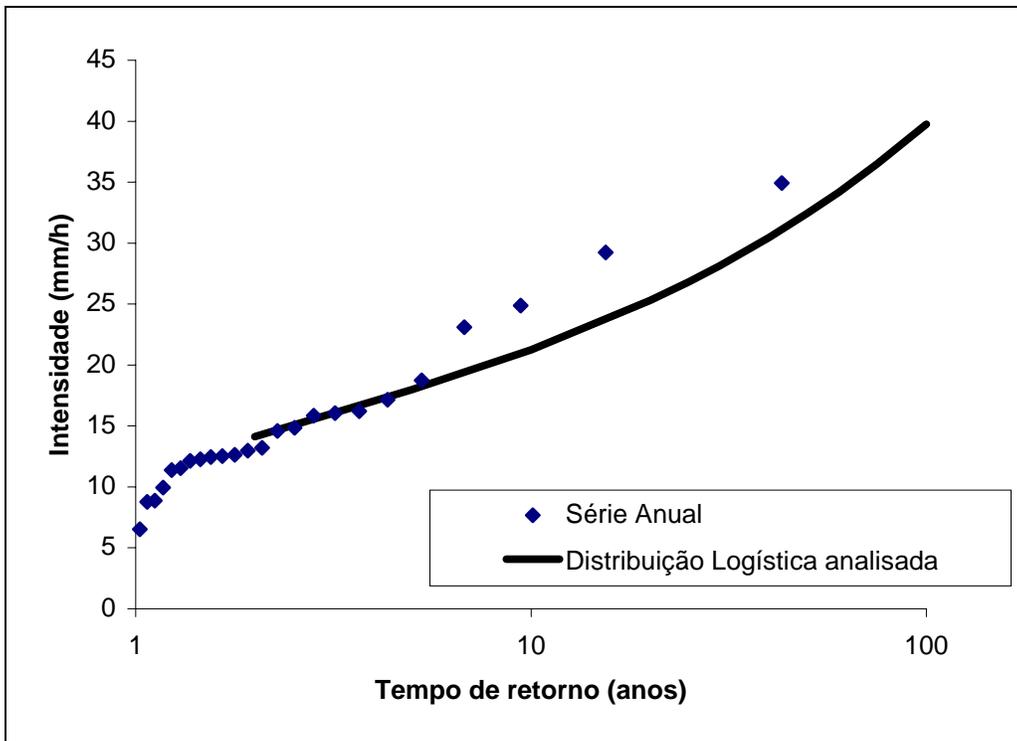


Figura 36 – Curva de frequência para a duração 4 horas – Bangu.

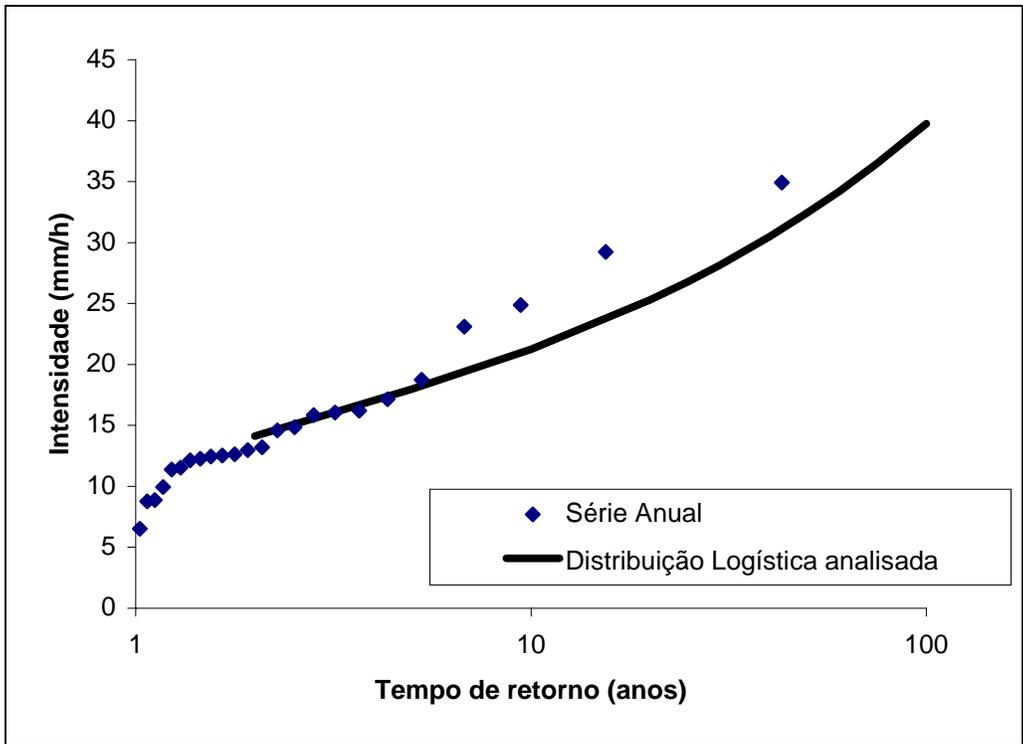


Figura 37 – Curva de frequência para a duração 14 horas – Bangu.

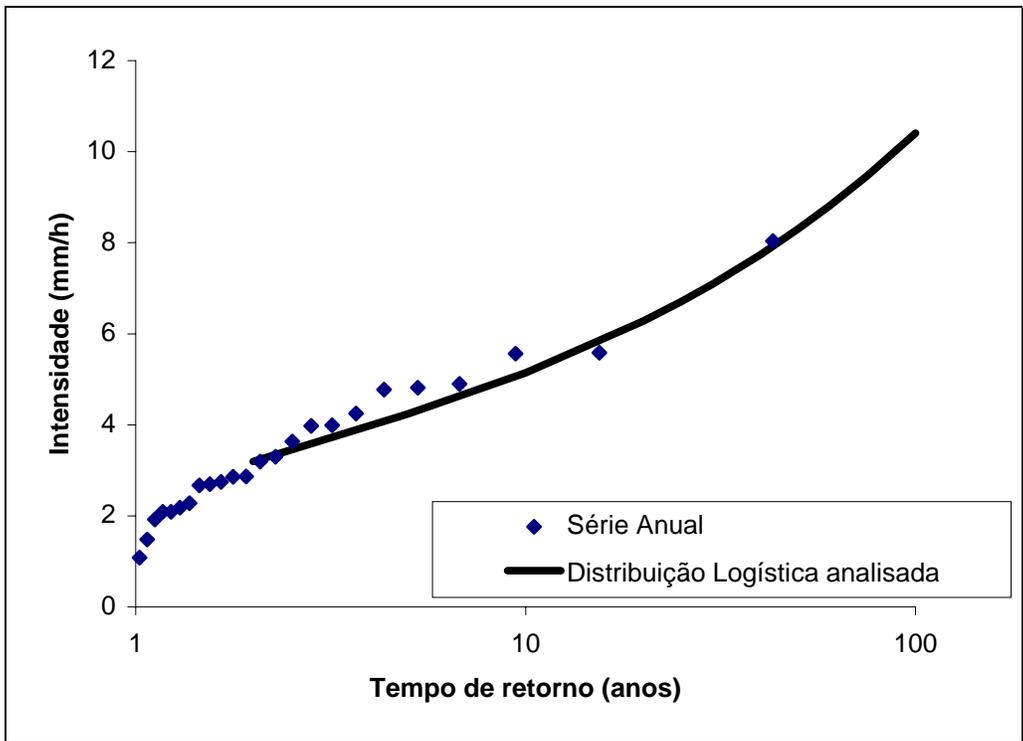


Figura 38 – Curva de frequência para a duração 24 horas – Bangu.

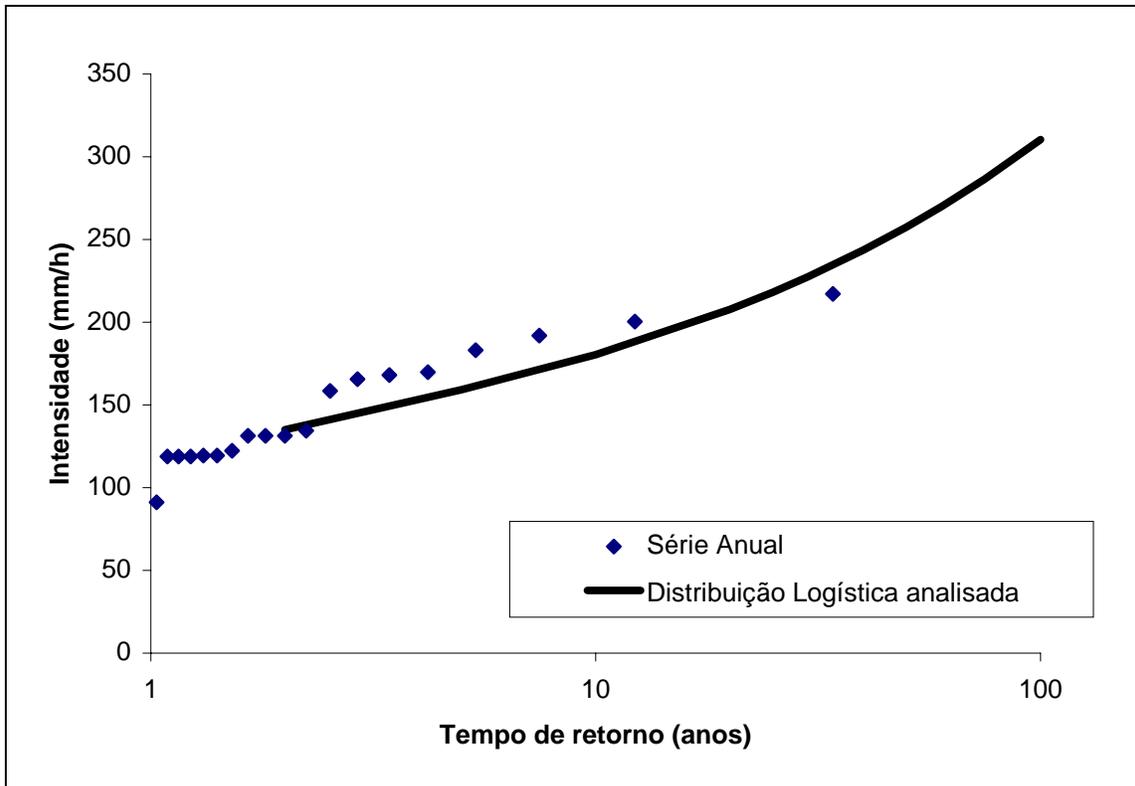


Figura 39 – Curva de frequência para a duração 5 minutos – Xerém.

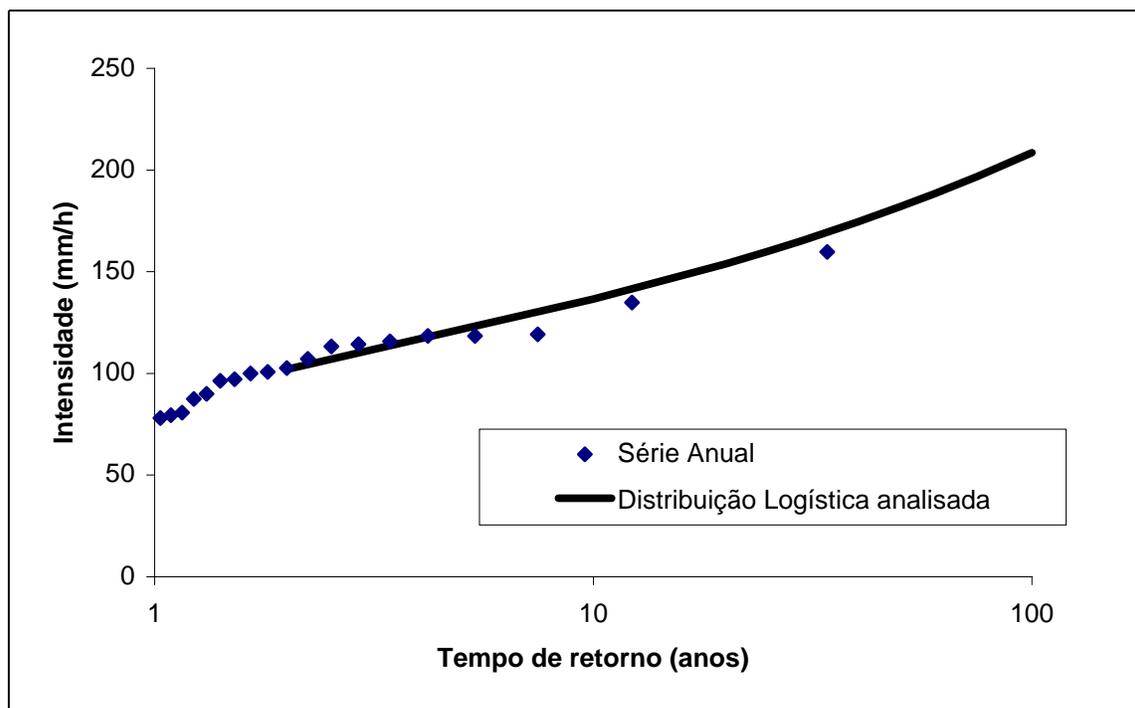


Figura 40 – Curva de frequência para a duração 15 minutos – Xerém.

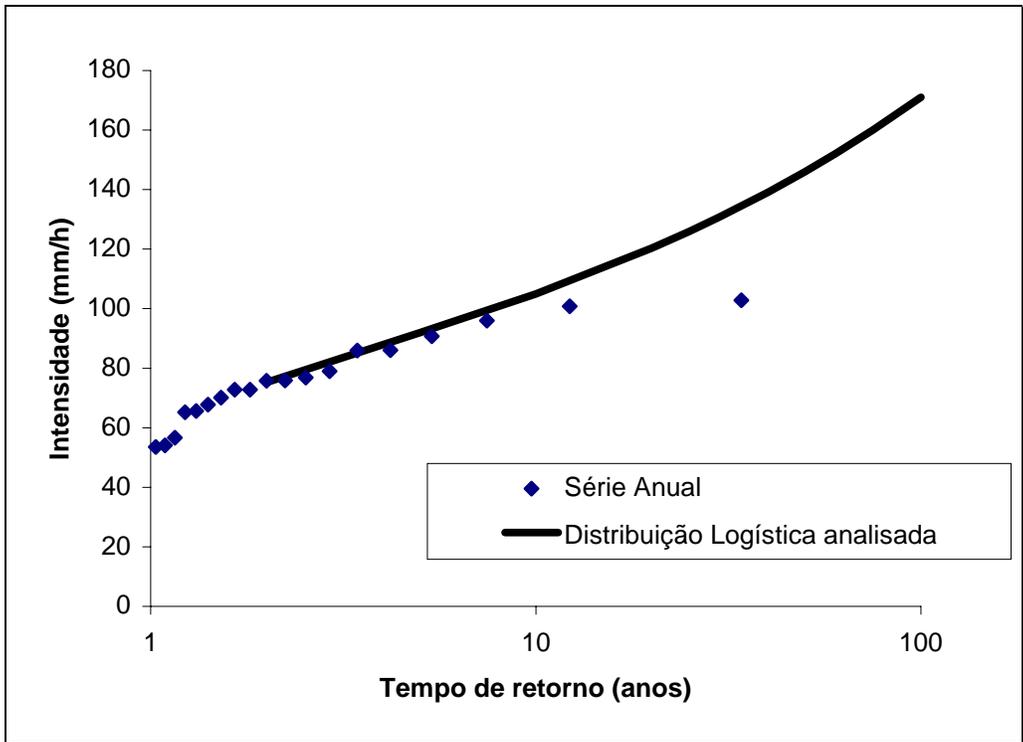


Figura 41 – Curva de frequência para a duração 30 minutos – Xerém.

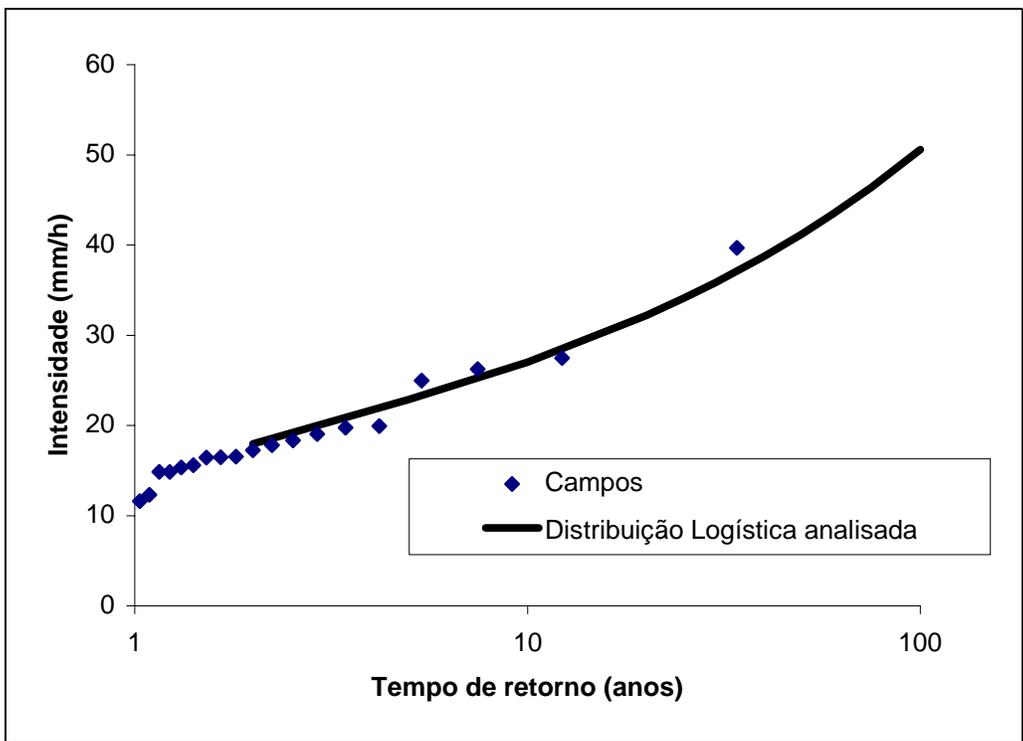


Figura 42 – Curva de frequência para a duração 4 horas – Xerém.

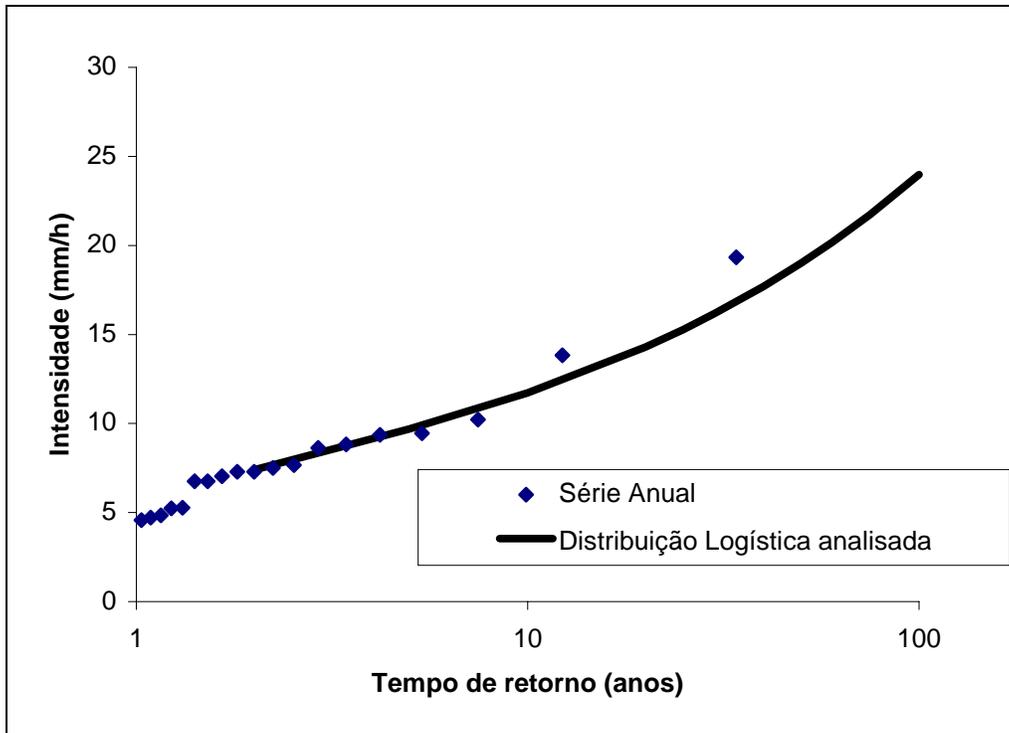


Figura 43 – Curva de frequência para a duração 14 horas – Xerém.

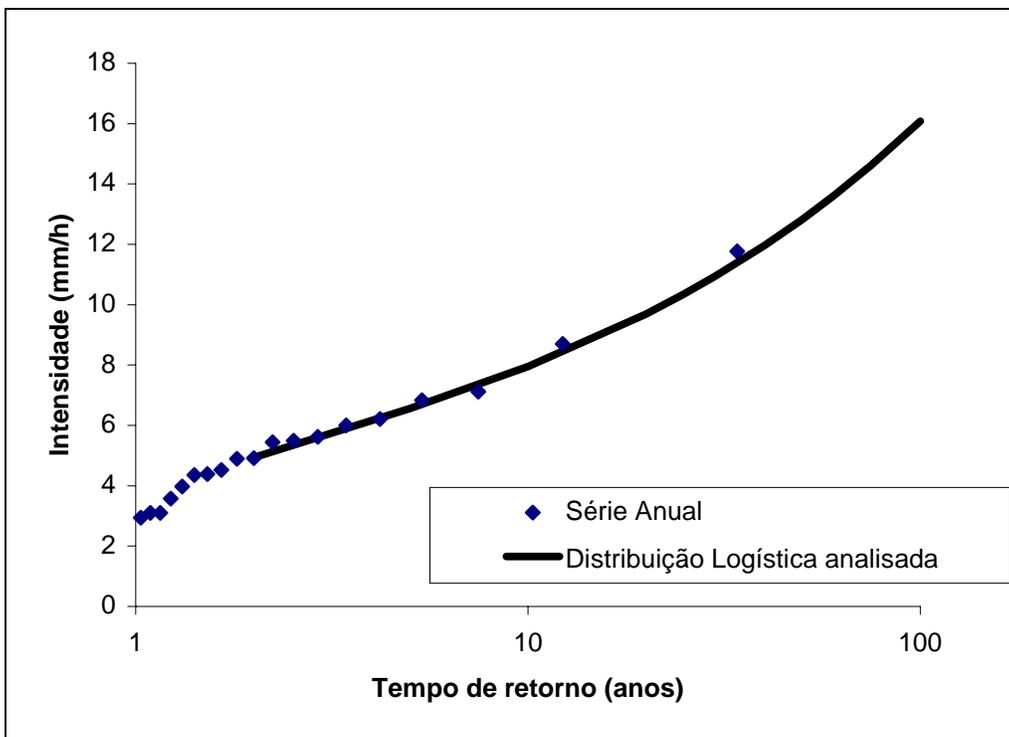


Figura 44 – Curva de frequência para a duração 24 horas – Xerém.

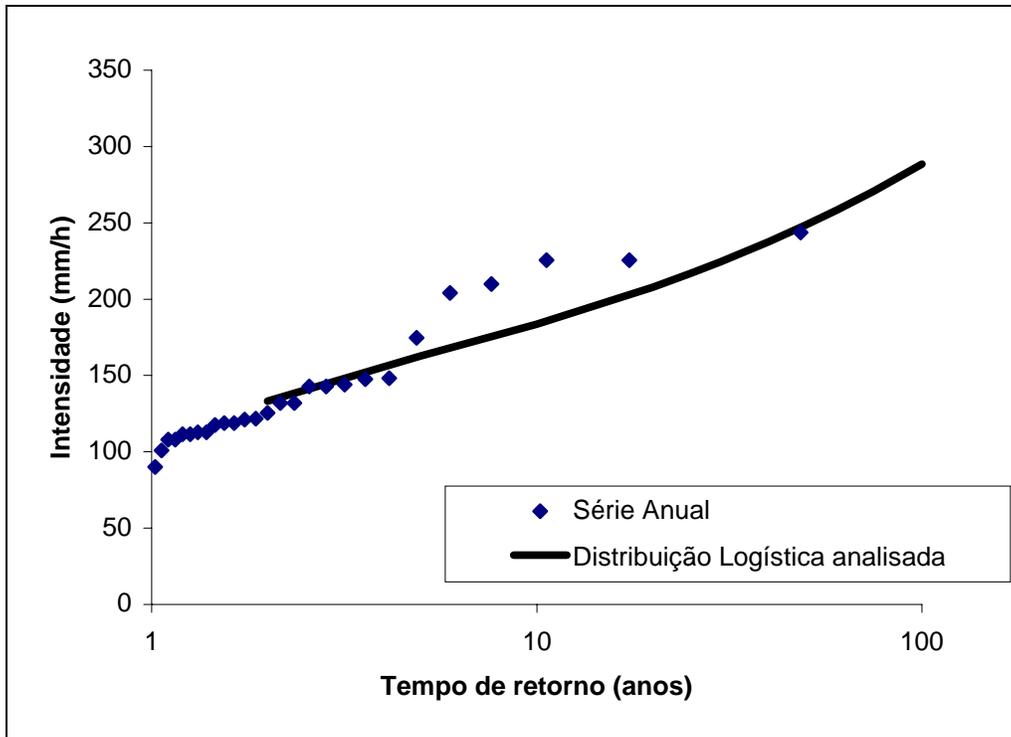


Figura 45 - Curva de frequência para a duração 5 minutos - Resende.

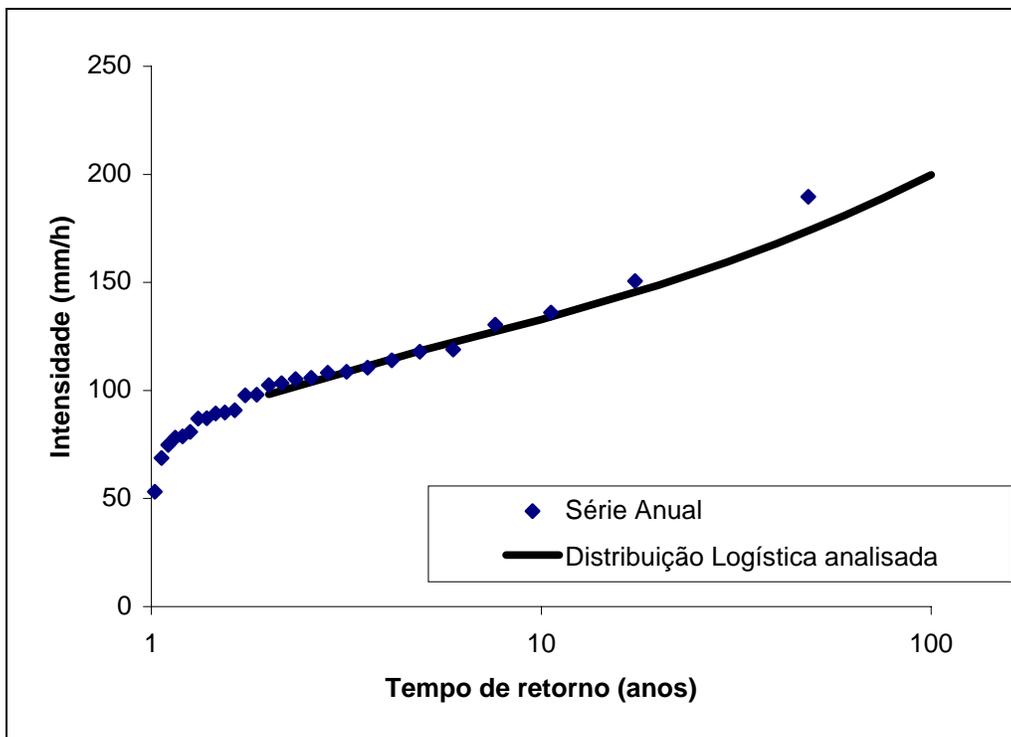


Figura 46 - Curva de frequência para a duração 15 minutos - Resende.

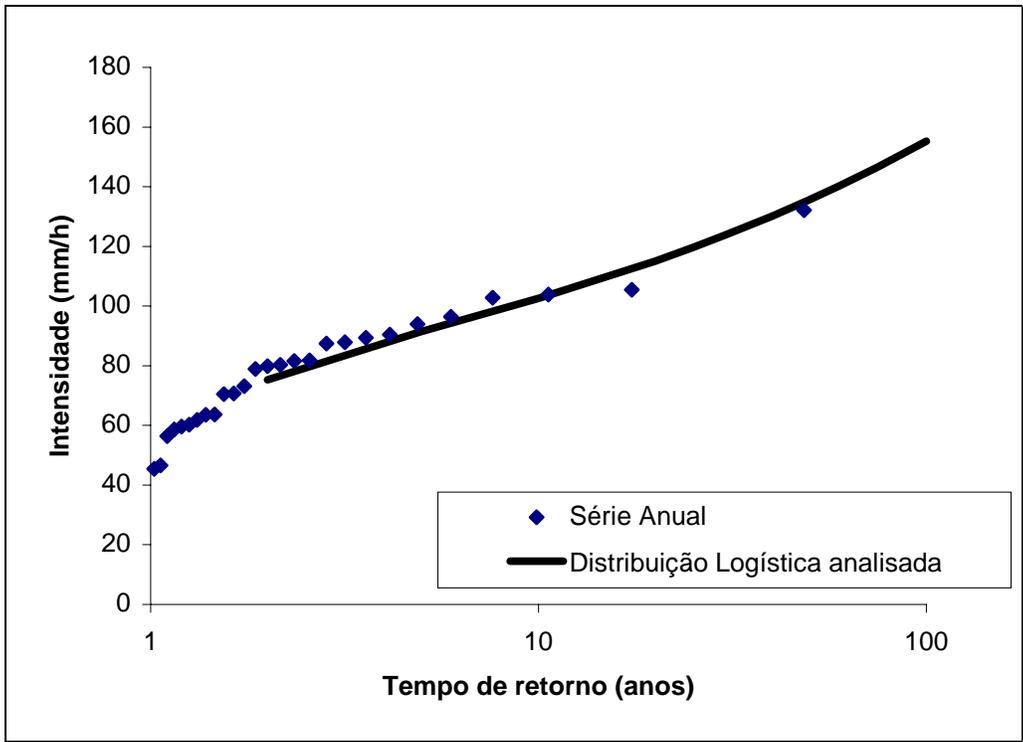


Figura 47 – Curva de frequência para a duração 30 minutos – Resende.

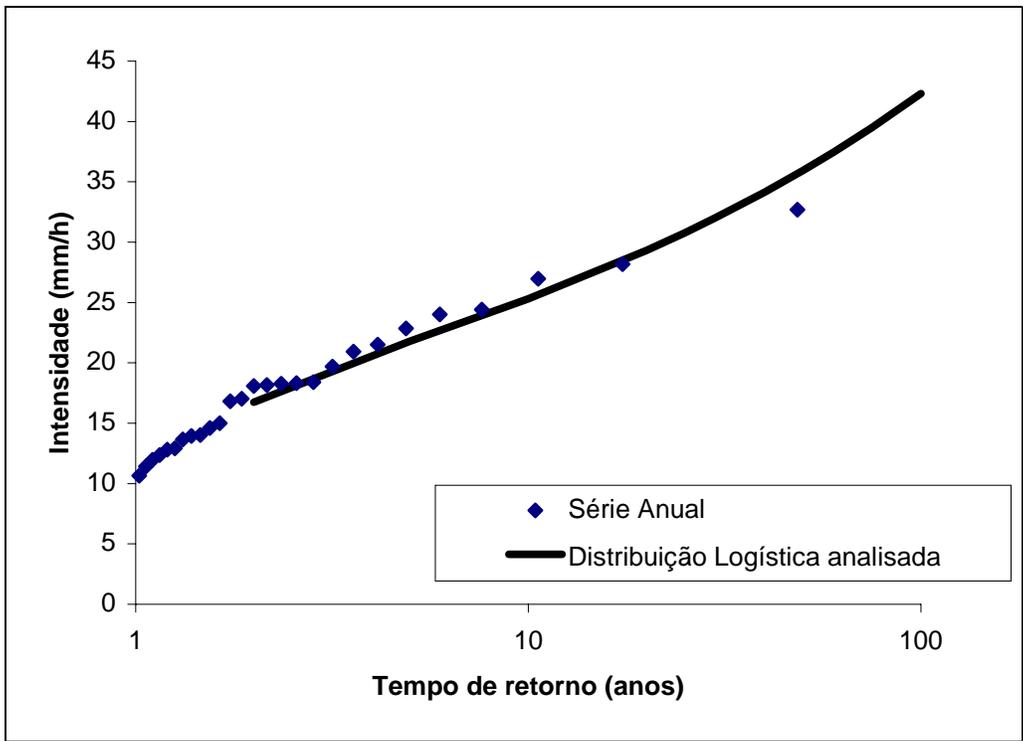


Figura 48 – Curva de frequência para a duração 4 horas – Resende.

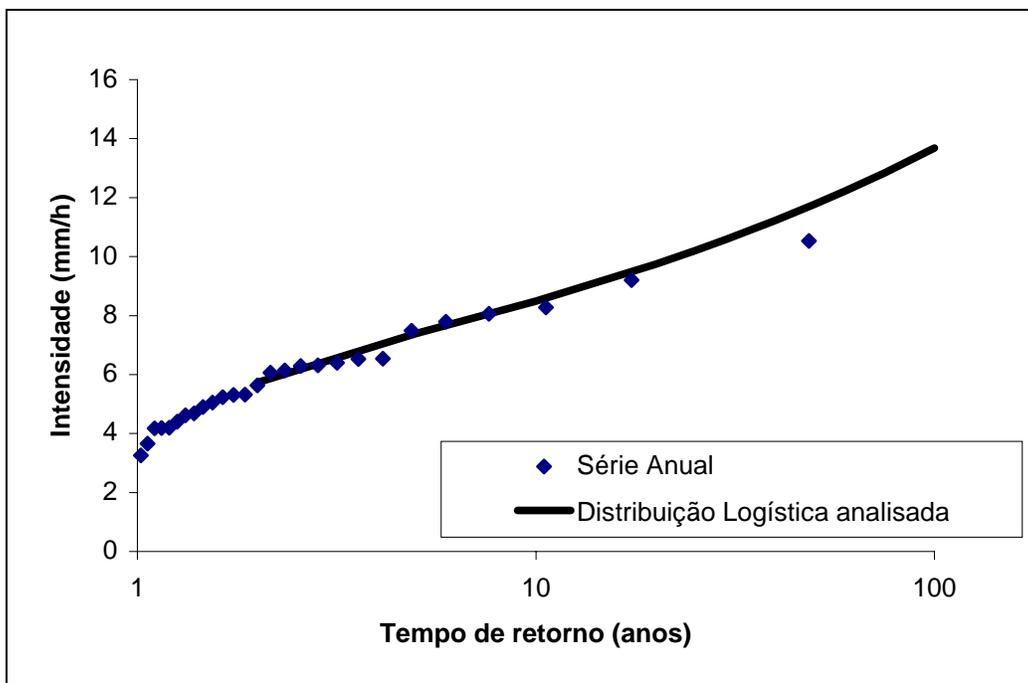


Figura 49 – Curva de freqüência para a duração 14 horas – Resende.

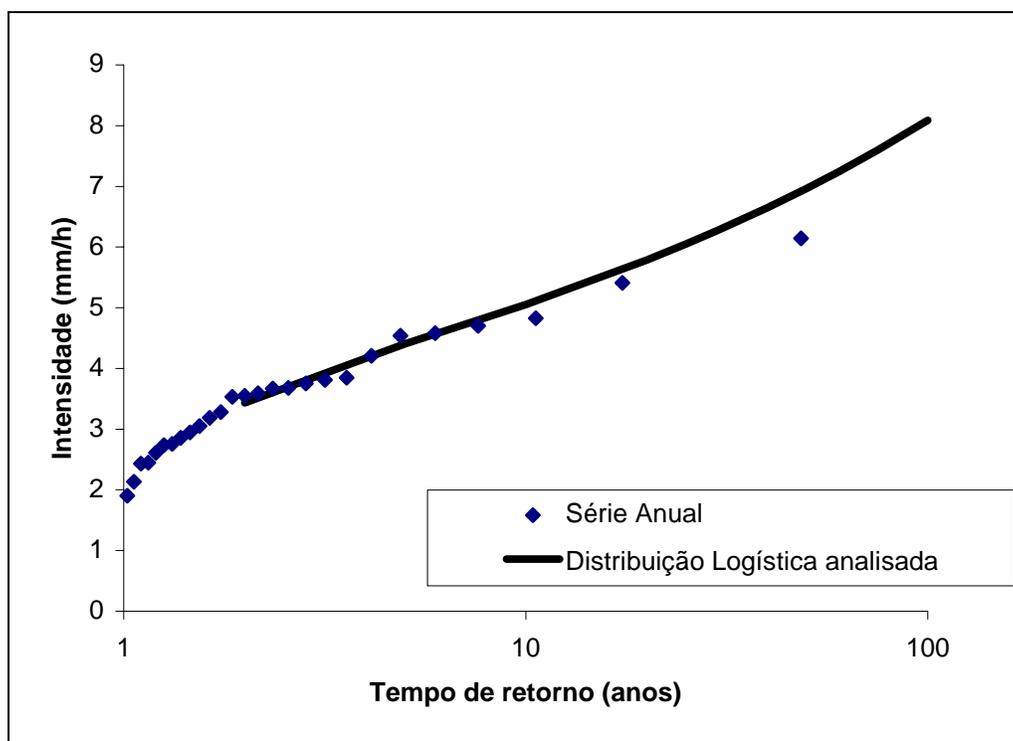


Figura 50 – Curva de freqüência para a duração 24 horas – Resende.

Anexo III

EXEMPLOS PRÁTICOS DE APLICAÇÃO DAS EQUAÇÕES IDF REGIONAIS

Exemplo (fictício) nº.1

Para o projeto de uma rede de drenagem pluvial em um novo loteamento na cidade de Nova Friburgo, verificou-se a necessidade de estimar qual é o valor relativo da intensidade de uma precipitação de 15 minutos de duração e um tempo de retorno de 10 anos.

- a) Levantamento das informações básicas disponíveis no estudo referentes à cidade de Nova Friburgo.
Precipitação média anual (P)
P=1390mm (figura 1 – Mapa de isoietas das precipitações médias anuais)
Identificação da região homogênea
Região 2 (figura 5 – Mapa das regiões homogêneas)
- b) Determinação dos quantis adimensionais de frequência ($\mu_{T,d}$).
Com o tempo de retorno (T=10 anos) e duração (d=15 minutos), obtemos da tabela 10 (Região 2: quantis anuais adimensionais regionais $\mu_{T,d}$) o valor de $\mu_{T,d} = 1,2962$.

c) Estimativa do *index-flood* (\bar{i}_d) para o local.

Para a região 2 e duração de 15 minutos utilizamos a seguinte equação:

$$\bar{i}_d = 39,445 d^{-0,339} P^{0,234}$$

(tabela 13- Modelo de regressão da variável dependente \bar{i}_d)

$$\bar{i}_d = 39,445 (15^{-0,339}) (1390^{0,234})$$

$$\bar{i}_d = 85,7 \text{ mm/h}$$

d) Estimativa da intensidade da precipitação para duração (d) de 15 minutos associada ao período de retorno (T) de 10 anos usando a seguinte equação:

$$\hat{i}_{T,d,j} = \bar{i}_d \mu_{T,d}$$

$$\hat{i}_{T,d,j} = 85,7 * 1,2962$$

$$\hat{i}_{T,d,j} = 111 \text{ mm/h}$$

Departamento de Apoio Técnico

Giuseppina Giaquinto de Araújo

Divisão de Cartografia

Paulo Roberto Macedo Bastos

Divisão de Editoração Geral

Maria da Conceição C. Jinho

EQUIPES DE PRODUÇÃO

Cartografia Digital

Afonso Lobo	José Pacheco Rabelo
Carlos Alberto da Silva Copolillo	Julimar de Araújo
Carlos Alberto Ramos	Leila Maria Rosa de Alcantara
Elaine de Souza Cerdeira	Luiz Guilherme Araújo Frazão
Elcio Rosa de Lima	Marco Antonio de Souza
Hélio Tomassini de O. Filho	Maria José Cabral Cezar
Ivan Soares dos Santos	Maria Luiza Poucinho
Ivanilde Muniz Caetano	Marília Santos Salinas do Rosário
João Batista Silva dos Santos	Paulo José da Costa Zilves
João Bosco de Azevedo	Regina de Sousa Ribeiro
João Carlos de Souza Albuquerque	Risonaldo Pereira da Silva
Jorge de Vasconcelos Oliveira	Sueli Mendes Sathler
José Barbosa de Souza	Valter Alvarenga Barradas
José Carlos Ferreira da Silva	Wilhelm Petter de Freire Bernard
José de Arimathéia dos Santos	

Editoração

Antonio Lagarde	Marília Asfura Turano
Edalair Rizzo	Pedro da Silva
Jean Pierre Souza Cruz	Sandro José Castro
José Luiz Coelho	Sergio Artur Giaquinto
Laura Maria Rigoni Dias	

MINISTÉRIO DE MINAS E ENERGIA
SECRETARIA DE MINAS E METALURGIA

<i>Ministro de Estado</i>	Rodolpho Tourinho Neto
<i>Secretário Executivo</i>	Helio Vitor Ramos Filho
<i>Secretário de Minas e Metalurgia</i>	Luciano de Freitas Borges

COMPANHIA DE PESQUISA DE RECURSOS MINERAIS – CPRM
Serviço Geológico do Brasil

<i>Diretor-Presidente</i>	Umberto Raimundo Costa
<i>Diretor de Hidrologia e Gestão Territorial</i>	Thales de Queiroz Sampaio
<i>Diretor de Geologia e Recursos Minerais</i>	Luiz Augusto Bizzi
<i>Diretor de Administração e Finanças</i>	José de Sampaio Portela Nunes
<i>Diretor de Relações Institucionais e Desenvolvimento</i>	Paulo Antonio Carneiro Dias
<i>Chefe do Departamento de Hidrologia</i>	Frederico Cláudio Peixinho

SUPERINTENDÊNCIAS REGIONAIS

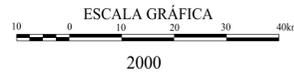
<i>Superintendente de Belém</i>	Xafi da Silva Jorge João
<i>Superintendente de Belo Horizonte</i>	Oswaldo Castanheira
<i>Superintendente de Goiânia</i>	Mário de Carvalho
<i>Superintendente de Manaus</i>	Fernando Pereira de Carvalho
<i>Superintendente de Porto Alegre</i>	Cladis Antonio Presotto
<i>Superintendente de Recife</i>	Marcelo Soares Bezerra
<i>Superintendente de Salvador</i>	José Carlos Vieira Gonçalves da Silva
<i>Superintendente de São Paulo</i>	José Carlos Garcia Ferreira
<i>Chefe da Residência de Fortaleza</i>	Clodionor Carvalho de Araújo
<i>Chefe da Residência de Porto Velho</i>	Rommel da Silva Sousa

PROJETO RIO DE JANEIRO

MAPA DE ISOIETAS TOTAIS ANUAIS

BACIAS DOS RIOS PARAÍBA DO SUL, ITABAPOANA E LITORÂNEAS DO ESTADO DO RIO DE JANEIRO

PERÍODO 1968-1995

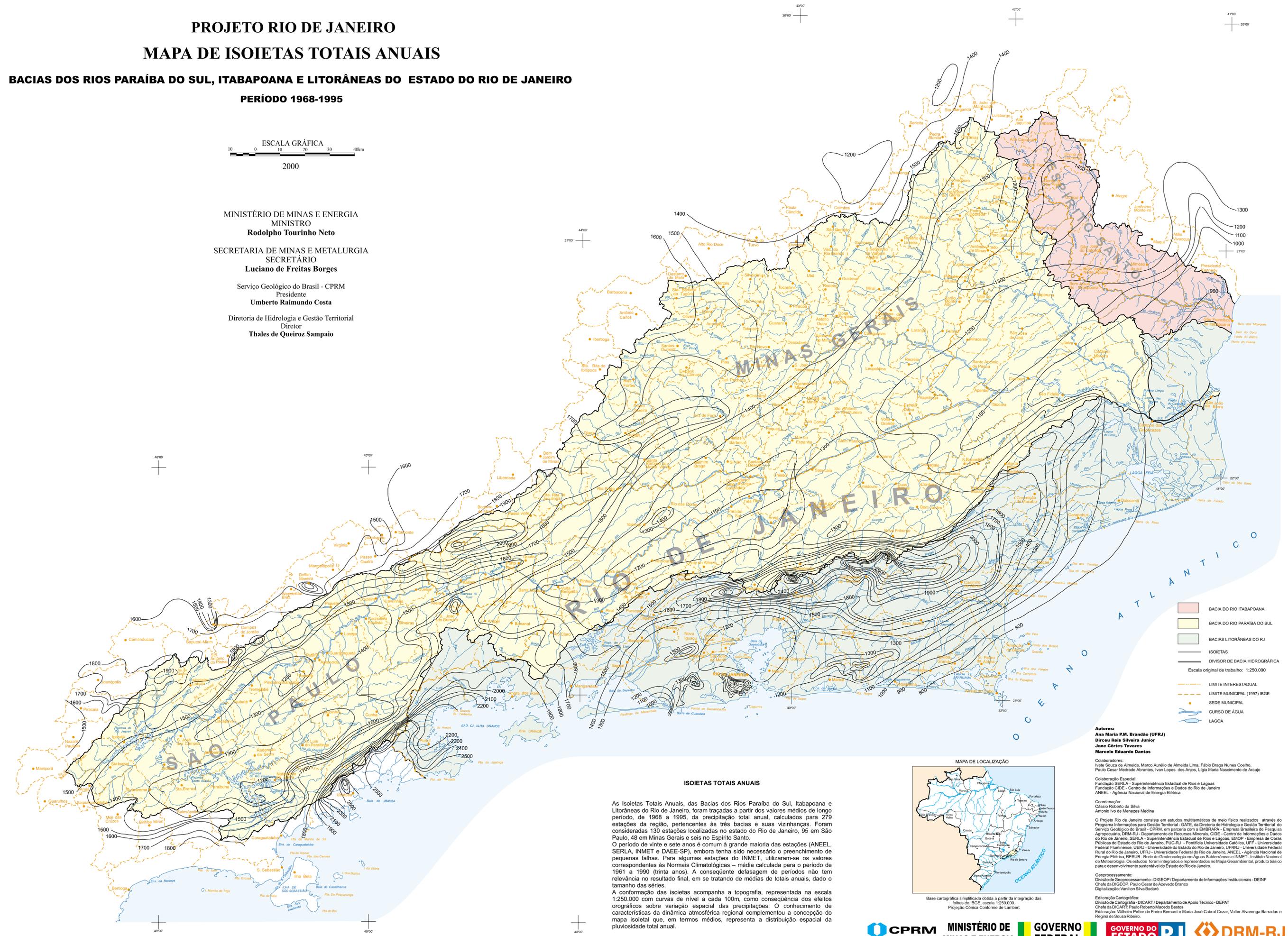


MINISTÉRIO DE MINAS E ENERGIA
MINISTRO
Rodolpho Tourinho Neto

SECRETARIA DE MINAS E METALURGIA
SECRETÁRIO
Luciano de Freitas Borges

Serviço Geológico do Brasil - CPRM
Presidente
Umberto Raimundo Costa

Diretoria de Hidrologia e Gestão Territorial
Diretor
Thales de Queiroz Sampaio



- BACIA DO RIO ITABAPOANA
- BACIA DO RIO PARAÍBA DO SUL
- BACIAS LITORÂNEAS DO RJ
- ISOIETAS
- DIVISOR DE BACIA HIDROGRÁFICA
- Escala original de trabalho: 1:250.000
- LIMITE INTERESTADUAL
- LIMITE MUNICIPAL (1997) IBGE
- SEDE MUNICIPAL
- CURSO DE ÁGUA
- LAGOA

Autores:
Ana Maria P.M. Brandão (UFRRJ)
Dirceu Reis Silveira Junior
Jane Carlos Tavares
Marcelo Eduardo Dantas

Colaboradores:
Ivete Souza de Almeida, Marco Aurélio de Almeida Lima, Fábio Braga Nunes Coelho, Paulo Cesar Medrado Abrantes, Ivan Lopes dos Anjos, Lígia Maria Nascimento de Araujo

Colaboração Especial:
Fundação SERLA - Superintendência Estadual de Rios e Lagos
Fundação CIDE - Centro de Informações e Dados do Rio de Janeiro
ANEEL - Agência Nacional de Energia Elétrica

Coordenação:
Cássio Roberto da Silva
Antonio Ivo de Menezes Medina



ISOIETAS TOTAIS ANUAIS

As Isoietas Totais Anuais, das Bacias dos Rios Paraíba do Sul, Itabapoana e Litorânea do Rio de Janeiro, foram traçadas a partir dos valores médios de longo período, de 1968 a 1995, da precipitação total anual, calculados para 279 estações da região, pertencentes às três bacias e suas vizinhanças. Foram consideradas 130 estações localizadas no estado do Rio de Janeiro, 95 em São Paulo, 48 em Minas Gerais e seis no Espírito Santo. O período de vinte e sete anos é comum à grande maioria das estações (ANEEL, SERLA, INMET e DAEE-SP), embora tenha sido necessário o preenchimento de pequenas falhas. Para algumas estações do INMET, utilizaram-se os valores correspondentes às Normais Climatológicas – média calculada para o período de 1961 a 1990 (trinta anos). A consequente defasagem de períodos não tem relevância no resultado final, em se tratando de médias de totais anuais, dado o tamanho das séries.

A conformação das isoietas acompanha a topografia, representada na escala 1:250.000 com curvas de nível a cada 100m, como consequência dos efeitos orográficos sobre variação espacial das precipitações. O conhecimento de características da dinâmica atmosférica regional complementou a concepção do mapa isoieta que, em termos médios, representa a distribuição espacial da pluviosidade total anual.