

UNIVERSIDADE DE SÃO PAULO
ESCOLA DE ENGENHARIA DE SÃO CARLOS
PÓS-GRADUAÇÃO EM ENGENHARIA HIDRÁULICA E SANEAMENTO

ARLAN SCORTEGAGNA ALMEIDA

Modelagem estocástica de precipitação por processos de Poisson

Versão Corrigida

São Carlos (SP)

2018

ARLAN SCORTEGAGNA ALMEIDA

Modelagem estocástica de precipitação por processos de Poisson

Class.
Curr.
Tempo
Sistema

Dissertação apresentada à Escola de Engenharia de São Carlos, da Universidade de São Paulo, como parte dos requisitos para obtenção do título de Mestre em Ciências: Engenharia Hidráulica e Saneamento.

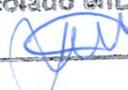
Orientadora: Profa. Dra. Luisa Fernanda Ribeiro Reis



São Carlos (SP)

2018

EESC/USP
Serviço de Pós Graduação
Protocolado em 06/08/2018



AUTORIZO A REPRODUÇÃO TOTAL OU PARCIAL DESTA TRABALHO,
POR QUALQUER MEIO CONVENCIONAL OU ELETRÔNICO, PARA FINS
DE ESTUDO E PESQUISA, DESDE QUE CITADA A FONTE.

Class.	TESE ✓
Cutt.	20.056
Tombo	T141/18
Sysno	2897546

31 Dec 21/18

16.08.18

Ficha catalográfica elaborada pela Biblioteca Prof. Dr. Sérgio Rodrigues Fontes da
EESC/USP com os dados inseridos pelo(a) autor(a).

A447m Almeida, Arlan Scortegagna
Modelagem estocástica de precipitação por processos
de Poisson / Arlan Scortegagna Almeida; orientador
Luisa Fernanda Ribeiro Reis. São Carlos, 2018.

Dissertação (Mestrado) - Programa de Pós-Graduação
em Engenharia Hidráulica e Saneamento e Área de
Concentração em Hidráulica e Saneamento -- Escola de
Engenharia de São Carlos da Universidade de São Paulo,
2018.

1. Precipitação. 2. Modelagem estocástica. 3.
Processos de Poisson. I. Título.

Eduardo Graziosi Silva - CRB - 8/8907



FOLHA DE JULGAMENTO

Candidato: Engenheiro **ARLAN SCORTEGAGNA ALMEIDA**.

Título da dissertação: "Modelagem estocástica de precipitação por processos de Poisson".

Data da defesa: 25/05/2018.

Comissão Julgadora:

Profa. Titular **Luisa Fernanda Ribeiro Reis**
(Orientadora)
(Escola de Engenharia de São Carlos/EESC)

Prof. Dr. **Ademir Paceli Barbassa**
(Universidade Federal de São Carlos/UFSCar)

Prof. Dr. **Erich Kellner**
(Universidade Federal de São Carlos/UFSCar)

Resultado:

Aprovado

APROVADO

APROVADO

Coordenador do Programa de Pós-Graduação em Engenharia Hidráulica e Saneamento:

Prof. Dr. **Eduardo Mario Menciondo**

Presidente da Comissão de Pós-Graduação:
Prof. Associado **Luís Fernando Costa Alberto**

AGRADECIMENTOS

Agradeço, primeiramente, a Deus, por me guiar nos dois anos que residi em São Carlos e por me indicar a direção certa em todos os momentos.

À professora Luisa Fernanda Ribeiro Reis pela orientação e por todo o apoio e motivação prestados durante todas as etapas do desenvolvimento deste trabalho.

À minha família que sempre esteve presente, sobretudo nas horas mais difíceis.

À minha noiva Laila, pelo auxílio, carinho e paciência.

Ao corpo docente e aos funcionários do Departamento de Hidráulica e Saneamento.

Ao Alexandre Ortega e aos demais responsáveis pela operação do posto meteorológico da ESALQ - Piracicaba, cuja ajuda foi de grande importância na obtenção dos dados de chuva.

À Joana Kaczmarska, autora de um dos modelos aplicados, que prestou grande auxílio na etapa de calibração.

Ao meu amigo Morris, que muito me auxiliou na programação.

Aos colegas do Labsin: Tatiana Gomes, Maria Gamboa, Daysy Cavalcanti, Daniele Feitoza, Rafael Damiano e Rafael Novais.

Aos amigos de mestrado: Alana Grandra, Bruno Freitas, Lucas Maroubo, David Rosa, Danilo Ayabe, Ana Cláudia, Jamil Anache, Hiago Schiano, Murilo Esposto, Nicolas Finkler.

Aos professores da UFAL: Vladimir Caramori, Roberaldo Carvalho, Ruberto Fragoso e Marllus Neves. Incluo também meus agradecimentos ao Luis Gustavo Reis, da GAMA Engenharia.

Aos colegas da UFAL, especialmente aos amigos do Programa Especial de Capacitação Discente que muito contribuiu para minha formação em pesquisa.

Ao CNPq pelo apoio financeiro.

E a todos os demais que de alguma forma contribuíram para que eu pudesse realizar esta dissertação.

Muito obrigado.

RESUMO

ALMEIDA, Arlan Scortegagna. **Modelagem estocástica de precipitação por processos de Poisson**. 2018. 147 p. Dissertação (Mestrado em Engenharia Hidráulica e Saneamento) – Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2018.

Modelos estocásticos baseados no processo de Poisson consistem em tecnologias de geração de chuva que vêm sendo aprimoradas ao longo das últimas três décadas. Nesta dissertação, foram abordados os modelos de pulsos retangulares, em que a precipitação é representada como uma sobreposição de eventos e células de chuva. Dentre os modelos existentes na literatura, foram selecionados dois representantes proeminentes: o de Pulsos Retangulares de Bartlett Lewis (BLRP), consolidado e aplicado em diversos estudos, e um de seus aperfeiçoamentos, denotado pela sigla BLRPRx, introduzido em 2014. Eles foram aplicados aos dados de chuva da série de Piracicaba, São Paulo, que representa as condições meteorológicas vigentes em uma localidade brasileira. Os parâmetros foram estimados através do Método Generalizado dos Momentos (MGM). O método utilizado na calibração conta com dois desenvolvimentos recentes na literatura: uma aproximação da matriz de pesos ótima do MGM, calculada segundo os dados observados, e as expressões analíticas dos momentos de terceira ordem de ambos os modelos. O ajuste foi realizado mensalmente, de modo a conciliar a suposição de estacionariedade, necessária aos modelos de Poisson, com a sazonalidade do fenômeno real. A otimização foi executada através do algoritmo Shuffled Complex Evolution (SCE-UA), implementado em linguagem C++. Para as simulações, foram programadas rotinas para a geração de séries sintéticas segundo os modelos calibrados. O desempenho dos modelos foi avaliado por comparação entre as propriedades estatísticas simuladas e as propriedades estatísticas observadas. Os resultados permitiram concluir que ambos os modelos reproduzem os momentos estatísticos observados de primeira e segunda ordem. Na reprodução dos momentos de terceira ordem e de outras propriedades estatísticas observadas (proporção de períodos secos e probabilidade de transição úmido-úmido), o BLRPRx apresentou desempenho superior sendo, portanto, mais recomendado para a modelagem da precipitação característica do local escolhido. Uma análise de valores extremos foi realizada, em que ambos os modelos reproduziram bem as lâminas máximas para tempos de retorno inferiores a 12 anos. Para tempos de retorno superiores, avalia-se que seja necessária uma análise baseada em uma série histórica de maior duração.

Palavras-chave: precipitação, modelagem estocástica, processos de Poisson.

ABSTRACT

ALMEIDA, Arlan Scortegagna. **Stochastic modelling of rainfall with Poisson processes.** 2018. 147 p. Dissertation (Master in Hydraulic Engineering and Sanitation) – Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2018.

Poisson process-based models represent a class of rainfall generation technologies that have been improved over the last three decades. This study focused on the rectangular pulses models, in which rainfall is represented as a superposition of storms and rain cells. Among existing models, two prominent representatives were selected: the Bartlett-Lewis Rectangular Pulses (BLRP), a consolidated model which has been applied in several studies, and one of its variants, denoted by the acronym BLRPRx, introduced in 2014. Both were applied to data from a rainfall series gauged in Piracicaba, São Paulo, which characterizes meteorological conditions from a Brazilian site. Model parameters were estimated according to Generalized Method of Moments (GMM). Calibration method incorporates two recent developments from related literature: an approximation to the optimal matrix of MGM, calculated from observed data, and the third order analytical moments of both models. The models were fitted in a monthly basis, in order to conciliate an assumption of stationarity, necessary to the application of Poisson models, with seasonality of real phenomenon. Optimization was performed with the Shuffled Complex Evolution (SCE-UA) algorithm, implemented in C++11 language with small modifications from the original algorithm. For simulation, routines were coded for the generation of synthetic series with the calibrated models. Model performance was assessed by comparing simulated statistical properties with the observed ones. The obtained results allowed to conclude that both models reproduced the first and second order observed moments. Regarding third-order moments and other observed statistics (proportion of dry periods and wet spell transition probabilities), BLRPRx outperformed BLRP and, therefore, is more recommended for modelling the characteristic rainfall of the chosen site. An extreme value analysis was performed, in which both models reproduced well the maximum depths from return periods smaller than 12 years. For higher return periods, an assessment based on a longer historical rainfall series is required.

Keywords: rainfall, stochastic modelling, Poisson processes.

LISTA DE ABREVIATURAS

BLIP	Modelo de Pulsos Instantâneos de Bartlett-Lewis
BLIPc	Modelo de Pulsos Instantâneos de Bartlett-Lewis Condicional
BLRP	Modelo de Pulsos Retangulares de Bartlett-Lewis
BLRPR	Modelo de Pulsos Retangulares de Bartlett-Lewis Randomizado
BLRPRG	Modelo de Pulsos Retangulares de Bartlett-Lewis Randomizado Gama
BLRPRx	Modelo de Pulsos Retangulares de Bartlett-Lewis Randomizado "x
CCE	Algoritmo <i>Competitive Complex Evolution</i>
DE	Algoritmo <i>Differential Evolution</i>
DRIP	Modelo Desagregado de Pulso Retangulares de Intensidade
DSPP	Modelo Estocástico Duplo do Processo de Poisson
EMA	Estação Meteorológica Automática
EMC	Estação Meteorológica Convencional
FAP	Função Acumulada de Probabilidades
FDP	Função Densidade de Probabilidades
GEV	Distribuição Generalizada de Valores Extremos
GNSRP	Modelo Generalizado de Pulsos Retangulares de Neyman-Scott
IDF	Intensidade – Duração – Frequência
i.i.d	Variáveis Aleatórias Independentes e Iguamente Distribuídas
MGM	Método Generalizado dos Momentos
MVS	Método da Máxima Verossimilhança
NSRP	Modelo de Pulsos Retangulares de Neyman-Scott
PPS	Proporção de Períodos Secos
PTuu	Probabilidade de Transição Úmido-Úmido
SCE-UA	Algoritmo <i>Shuffled Complex Evolution – University of Arizona</i>
STNSRP	Modelo Espaço-Temporal de Pulsos Retangulares de Neyman-Scott

LISTA DE SÍMBOLOS

$L_{célula}$	Duração de uma célula de chuva
L_{evento}	Duração de um evento de chuva
X	Intensidade de uma célula de chuva
μ_X	Valor esperado da intensidade de uma célula de chuva
λ	Taxa de ocorrência/chegada dos eventos de chuva (intensidade de Poisson)
β	Taxa de ocorrência/chegada das células de chuva (intensidade de Poisson)
β^*	Número de vezes que a seleção de soluções pais é realizada no algoritmo CCE
C	Número de células geradas em um evento
μ_C	Valor esperado do número de células de um evento
γ	Parâmetro de taxa da distribuição Exponencial relativa à duração de um evento
η	Parâmetro de taxa da distribuição Exponencial relativa à duração de uma célula
α	Parâmetro de forma da distribuição Gama relativa ao parâmetro η
α^*	Número de vezes que cada <i>simplex</i> é evoluído no algoritmo CCE
ν	Parâmetro de escala da distribuição Gama relativa ao parâmetro η
ϕ	Coefficiente adimensional correspondente à razão γ/η
κ	Coefficiente adimensional correspondente à razão β/η
ι	Coefficiente adimensional correspondente à razão μ_X/η
$Y(t)$	Intensidade de chuva no instante t
Y_i^h	Lâmina precipitada no i ésimo intervalo de duração arbitrária h
$\Gamma(.)$	Função Gama
Θ	Estimador dos parâmetros segundo o Método Generalizado dos Momentos
θ	Vetor contendo o conjunto de parâmetros populacionais do modelo
$L(\theta)$	Função de verossimilhança
$\hat{\theta}$	Vetor contendo as estimativas de θ obtido a partir da amostra
T	Vetor contendo os valores das propriedades estatísticas amostrais
$\tau(\theta)$	Vetor contendo as expressões do modelo para as propriedades estatísticas em T
W	Matriz de pesos do Método Generalizado dos Momentos
Ω	Espaço de soluções viáveis para geração da população inicial do SCE-UA
f_{min}	Valor da função objetivo obtido como resultado da otimização (minimização)
$g(T_r)$	Variável reduzida da distribuição de Gumbel para o tempo de retorno T_r

SUMÁRIO

1	INTRODUÇÃO	9
2	OBJETIVOS	13
3	REVISÃO BIBLIOGRÁFICA	15
3.1	Modelos de precipitação de Poisson	15
3.2	Estudos recentes	22
4	MÉTODOS	26
4.1	Seleção dos modelos	26
4.2	Concepção dos modelos	30
4.3	Expressões analíticas dos momentos estatísticos	36
4.4	Aquisição dos dados observados	41
4.4.1	Série histórica de precipitação	42
4.4.2	Estatísticas amostrais	43
4.4.3	Análise de falhas	48
4.5	Calibração	56
4.5.1	Métodos baseados em verossimilhança	57
4.5.2	Método Generalizado dos Momentos	58
4.5.3	Otimização	63
4.6	Simulação	68
5	RESULTADOS E DISCUSSÃO	71
5.1	Estatísticas amostrais	71
5.2	Calibração	77
5.2.1	BLRP.....	78
5.2.2	BLRPRx.....	81
5.2.3	Considerações sobre a calibração do BLRP/BLRPRx.....	82
5.3	Valores esperados dos momentos estatísticos	84
5.4	Resultados das simulações	90
5.5	Validação e desempenho dos modelos	95
5.6	Análise de valores extremos	99
6	CONCLUSÃO	105
	REFERÊNCIAS¹	107
	APÊNDICE A	113
	APÊNDICE B	117

1 INTRODUÇÃO

A precipitação é um fenômeno de natureza complexa, devido à grande quantidade de variáveis envolvidas e às relações de interdependência existentes entre elas. A compreensão, ao menos parcial, do mecanismo gerador das chuvas e a manipulação de seus resultados, que ocorrem na forma das séries temporais de precipitação observadas, são indispensáveis para muitas das aplicações em recursos hídricos.

Devido à sua natureza, o fenômeno precipitação pode ser modelado como um processo estocástico. Modelos estocásticos fundamentam-se no princípio de que o fenômeno em estudo, devido à sua complexidade, produz resultados aleatórios que podem ser descritos através de distribuições probabilísticas. No caso da precipitação, o objetivo primário da modelagem estocástica é a geração de séries sintéticas com propriedades estatísticas idênticas às da série observada. As séries sintéticas representam, portanto, possíveis realizações do fenômeno e permitem estimar dados de chuva para tempos ou locais desprovidos de observações.

Nos estudos hidrológicos de avaliação de risco, é comum a abordagem baseada em eventos chuvosos individuais para a quantificação do escoamento superficial. Esse é o caso dos métodos baseados na chuva de projeto, em que são feitas suposições arbitrárias referentes à duração crítica do evento, à sua distribuição temporal e às perdas iniciais e contínuas, decorrentes de infiltração, para o cálculo da precipitação efetiva. Verhoest et al. (2010) argumenta que uma desvantagem desse tipo de abordagem é que as condições antecedentes de umidade da bacia geralmente não são contabilizadas de forma adequada. A importância das condições antecedentes reside no fato de que eventos extremos de cheia frequentemente resultam de eventos menos extremos de chuva, ou seja, que têm menor tempo de retorno, mas que precipitaram em uma bacia com solo úmido (VERHOEST et al., 2010).

Uma forma de contornar esse problema e melhorar a qualidade das avaliações de risco é através dos modelos de simulação contínua, segundo os quais as condições antecedentes são contabilizadas continuamente e seleções arbitrárias, relativas, por exemplo, aos eventos individuais de chuva e suas características, podem ser eliminadas (BOUGHTON; DROOP, 2003). Geralmente empregados para estimar os hidrogramas de cheia de projeto, esses modelos simulam as condições de umidade da bacia, entre outras características físicas, e o escoamento superficial produzido em intervalos de tempo discretos da ordem de dias, horas e, ocasionalmente, minutos.

A simulação contínua pode ser aplicada no dimensionamento ou na avaliação probabilística de falha dos sistemas em que a chuva é uma das condicionantes. Esses sistemas incluem muitas das obras de infraestrutura, tais como estruturas de drenagem urbana, defesa contra enchentes, pontes, etc. Incluem até mesmo as redes de telecomunicação, em que o problema de “*rain fade*”, que requer estratégias de mitigação tais como a incorporação de diversidade de rota, é desencadeado por condições pluviométricas (KACZMARSKA; ISHAM; NORTHROP, 2015).

Para que os modelos de simulação contínua sejam representativos da realidade, eles precisam ser alimentados com séries de precipitação muito longas, de modo a simular tanto a ocorrência de eventos quanto os períodos entre eles (períodos secos). Séries muito longas geralmente não estão disponíveis nos registros históricos. Uma forma de suprir essa demanda é através de séries sintéticas com, possivelmente, milhares de anos, que podem ser produzidas como resposta dos modelos estocásticos. Em algumas aplicações, sobretudo em bacias urbanas, são necessárias séries longas e com alta resolução temporal, preferencialmente sub-horária, para capturar a variabilidade das vazões resultantes das simulações. Na grande maioria dos locais de monitoramento os registros pluviométricos ainda estão restritos à resolução diária. Os modelos estocásticos podem ser utilizados em combinação com técnicas de desagregação para a produção de séries sintéticas com maior resolução (KOUTSOYIANNIS; ONOF, 2001).

Face ao exposto, tecnologias de geração de chuva são de grande importância no contexto dos modelos de simulação contínua e o seu aprimoramento é benéfico para muitos estudos no campo dos recursos hídricos. Neste trabalho, essas tecnologias serão abordadas através de uma classe de modelos estocásticos, denominados de pulsos retangulares, cuja formulação conceitual e estatística fundamenta-se nos processos pontuais de Poisson. O foco é a modelagem dos dados de uma série de precipitação real e a discussão dos principais modelos e métodos de calibração que podem ser utilizados. As aplicações possíveis dos resultados decorrentes dessa modelagem são diversas, embora não constituam o escopo do presente trabalho.

Destacam-se três importantes propriedades que tornam os modelos de pulsos retangulares particularmente atraentes e populares entre os engenheiros hidrólogos, estatísticos e meteorologistas. A primeira é a representação do processo físico da precipitação de um modo realista e ao mesmo tempo simplificado. A segunda é a possibilidade de gerar séries sintéticas em diversas escalas de agregação temporal, inclusive de alta resolução (sub-horária). E a terceira é a possibilidade de estender a modelagem ao caso de sítios múltiplos,

em que a estimação dos parâmetros é feita para os dados de dois ou mais pluviômetros representativos da mesma região, e ao caso da modelagem espaço-temporal, em que a dimensão do espaço é introduzida nos modelos. Dadas às dificuldades na obtenção de séries históricas de chuva, que fossem ao mesmo tempo consistentes e de longa duração, o presente trabalho foi limitado à modelagem temporal de sítio único, em que as lâminas de apenas um pluviômetro foram consideradas. Não obstante, modelos estocásticos de sítio único permanecem em constante aprimoramento e são alvos de desenvolvimentos recentes da literatura.

Uma vez selecionado um modelo estocástico, sua aplicação aos dados de chuva é realizada, basicamente, em quatro etapas:

1. Seleção do local de modelagem e obtenção da série histórica: os dados de chuva são obtidos na forma de lâminas sequenciais agregadas na escala original, ou seja, na resolução temporal oferecida pelo instrumento de medição (neste caso, 15 minutos).
2. Agregação das lâminas e cálculo das estatísticas amostrais: os dados da série histórica de precipitação são agregados em diferentes escalas temporais (neste caso, 1 hora, 6 horas e 24 horas) e para cada escala, inclusive para a escala da resolução, são computados os momentos de primeira a terceira ordem e as demais propriedades estatísticas observadas: média, variância, autocovariâncias, momento central de terceira ordem, proporção de períodos secos (PPS) e probabilidades de transição entre um período úmido para outro período úmido (PTuu).
3. Calibração: os parâmetros do modelo são estimados de acordo com os momentos analíticos, calculados segundo expressões analíticas derivadas para cada modelo, que são equacionados com os respectivos momentos observados. Essa etapa exige a minimização de uma função objetivo formulada de acordo com um estimador.
4. Simulação: são produzidas séries sintéticas a partir dos modelos calibrados.

A qualidade da modelagem é avaliada através da comparação entre as propriedades estatísticas simuladas, computadas a partir das séries sintéticas, e as propriedades estatísticas observadas. Dada à importância das intensidades máximas em muitas aplicações de avaliação de risco, também deve ser feita uma análise comparativa dos valores extremos, ou seja, das lâminas máximas obtidas em diferentes escalas de agregação, produzidos pelo modelo e os respectivos valores observados.

2 OBJETIVOS

Este trabalho tem como objetivo realizar a modelagem estocástica da precipitação, segundo os dados de um pluviômetro de báscula, localizado em território nacional, através dos modelos de Poisson.

Os objetivos específicos são:

- Avaliar modelos estocásticos de precipitação, selecionados entre os mais eficientes, de acordo com resultados da literatura, e aplicá-los aos dados da série histórica.
- Estabelecer um método de calibração contemplando a formulação de uma função objetivo adequada (estimador) e a implementação de um algoritmo de otimização eficaz.
- Programar códigos computacionais para leitura e tratamento dos dados de entrada, calibração dos modelos e geração de séries sintéticas, e disponibilizá-los para desenvolvimentos futuros.
- Avaliar comparativamente o desempenho dos modelos em relação à reprodução das principais características estatísticas da série histórica.

3 REVISÃO BIBLIOGRÁFICA

3.1 Modelos de precipitação de Poisson

Os modelos utilizados no desenvolvimento deste trabalho têm como ponto de partida os processos estocásticos pontuais. Um processo pontual é definido como a realização de um grupo de pontos, ou ocorrências, cada qual tendo uma posição bem definida no espaço, ou no tempo. Nessa categoria se enquadra o processo de Poisson (revisado no Apêndice A). Trata-se do mais simples dentre os processos pontuais, que se caracteriza por considerar a ocorrência dos pontos de modo totalmente aleatório. O processo de Poisson desempenha um papel importante nos processos pontuais que na maioria dos aspectos é análogo ao da distribuição normal no estudo das variáveis aleatórias (COX; ISHAM, 1980).

O estudo de Waymire e Gupta (1981) figura entre os primeiros que descreveram a estrutura matemática básica de representação da precipitação. Essa estrutura é composta por duas sequências de variáveis aleatórias, mutuamente independentes, em que a primeira indica a ocorrência de eventos de chuva, e sua distribuição temporal, e a segunda associa uma altura pluviométrica a cada evento. Considera-se que o processo de contagem da primeira sequência, ou seja, a variável aleatória cujo valor realizado é o número de eventos, consiste em um processo de Poisson (PARZEN, 1962). Atendendo à definição desse processo, o tempo é representado como um intervalo contínuo no qual os eventos surgem instantaneamente.

Deve-se observar, no entanto, que as lâminas precipitadas durante os eventos de chuva não ocorrem de forma instantânea no tempo e tão pouco em uma escala que permita admitir a independência entre intervalos sucessivos de, por exemplo, 5 minutos. Para contornar essa dificuldade, os primeiros modelos estocásticos de processos pontuais consideravam a precipitação como uma sequência de “explosões instantâneas” (termo traduzido da expressão em inglês *bursts*), caracterizadas por uma altura pluviométrica aleatória.

Rodriguez-Iturbe, Gupta e Waymire (1984) descreveram o modelo Ruído Branco de Poisson, o mais simples de sua categoria, em que cada explosão instantânea é caracterizada por uma altura pluviométrica que representa a lâmina precipitada durante o evento. O número de explosões instantâneas, ou seja, de eventos, segue um processo de Poisson homogêneo. Trata-se de um modelo de baixa representatividade que desconsidera a duração da chuva e a estrutura de dependência existente entre as lâminas acumuladas em intervalos sucessivos de tempo. Para a modelagem dos dados de Denver (EUA) e Agua Fria (Venezuela), o Ruído Branco de Poisson resultou em consideráveis discrepâncias entre os parâmetros estimados

para a chuva agregada em diferentes escalas de tempo (horária e diária), o que indicou sua inadequação estrutural.

Contudo, essa simples concepção, embora tenha sofrido diversas modificações, ainda é utilizada na estrutura básica dos modelos de precipitação de Poisson. Conforme será mencionado adiante, o conceito de explosões instantâneas foi revisitado mais recentemente no modelo desenvolvido por Cowpertwait, Isham e Onof (2007), sendo que nesse caso as explosões instantâneas representam “pulsos” de chuva.

Rodriguez-Iturbe, Gupta e Waymire (1984) apresentaram também a aplicação de um modelo de pulsos retangulares, que conta com uma estrutura de dependência Markoviana, e do modelo de Ruído Branco de Neyman-Scott, introduzido anteriormente por Kavvas e Delleur (1975). Este, por sua vez, considera uma organização hierárquica da chuva em eventos, cada evento contendo um “aglomerado” (tradução do termo *cluster*) de explosões instantâneas, em que cada explosão é dotada de uma altura pluviométrica. Enquanto os resultados do modelo Markoviano foram similares ao que foi obtido para o Ruído Branco de Poisson, o modelo de Ruído Branco de Neyman-Scott se mostrou mais promissor.

Dando continuidade a esse desenvolvimento, Rodriguez-Iturbe, Cox e Isham (1987) introduziram dois modelos que também contam com uma estrutura hierárquica, de modo similar aos modelos baseados no processo de Neyman-Scott, mas contendo aglomerados de células em substituição aos aglomerados de explosões instantâneas. De acordo com essa concepção, as células de chuva consistem nas unidades básicas da precipitação, sendo que cada célula é caracterizada por uma duração e uma intensidade (constante ao longo da duração) que são variáveis aleatórias mutuamente independentes e independentes do processo de Poisson. Esses modelos foram denominados Pulsos Retangulares de Neyman-Scott (NSRP) e Pulsos Retangulares de Bartlett-Lewis (BLRP).

A publicação de Rodriguez-Iturbe, Cox e Isham (1987) é provavelmente a mais importante da literatura relacionada aos modelos de precipitação fundamentados no processo de Poisson. A estrutura hierárquica dos modelos BLRP e NSRP que considera a precipitação como uma sobreposição de eventos e células, com os eventos de chuva em um nível superior e as células de chuva em um nível inferior, foi preservada em praticamente todos os modelos da classe de pulsos retangulares desenvolvidos a partir de 1987.

O primeiro aprimoramento do modelo BLRP foi feito pelos mesmos autores que o introduziram (RODRIGUEZ-ITURBE; COX; ISHAM, 1988). Eles modificaram o parâmetro de taxa da função densidade de probabilidades (FDP) da duração das células, convertendo-o em uma variável randômica, de forma que para cada evento é atribuído um valor específico

desse parâmetro. Disso decorre que diferentes eventos possuem diferentes características, no que se refere à sua duração, às durações das células e à frequência de ocorrência de células (maiores detalhes podem ser encontrados na subseção 4.2). Isso conferiu maior flexibilidade ao modelo modificado, denominado Pulsos Retangulares de Bartlett-Lewis de Parâmetro η Randômico, ou simplesmente, Randomizado, denotado por BLRPR.

O BLRPR foi aplicado na modelagem dos dados de chuva de diversos países, a saber: Estados Unidos (RODRIGUEZ-ITURBE; COX; ISHAM, 1988; BO; ISLAM; ELTAHIR, 1994; VELGHE et al., 1994; KIM et al., 2017), Itália (ISLAM et al., 1990), Inglaterra (ONOF; WHEATER, 1993), Irlanda (KHALIQ; CUNNANE, 1996), Bélgica (VERHOEST et al., 2010; VANHAUTE et al., 2012), Coréia do Sul (KIM; KWON; KIM, 2017), Alemanha (KACZMARSKA; ISHAM; ONOF, 2014; CROSS et al., 2018), África do Sul (SMITHERS; PEGRAM; SCHULZE, 2002), Austrália (WASKO et al., 2015) e Brasil (BACK; DORFMAN; CLARKE, 1999; TEIXEIRA-GANDRA; DAMÉ, 2014).

Em uma tentativa de introduzir maior flexibilidade ao BLRPR, objetivando a melhor reprodução dos valores extremos das séries sintética, Onof e Wheeler (1994) implementaram uma distribuição Gama, de dois parâmetros, para caracterizar a FDP das intensidades das células, em substituição à FDP exponencial, com apenas um parâmetro, proposta originalmente. Essa versão foi denominada Pulsos Retangulares de Bartlett-Lewis Randomizado Gama (BLRPRG).

Nos modelos descritos até o momento (BLRP, BLRPR e BLRPRG), a estrutura da precipitação considera dois processos de Poisson hierárquicos: o processo primário em que surgem os eventos de chuva e os processos secundários, que são desencadeados a cada processo primário, em que surgem as células de chuva. Com o objetivo de promover um melhor ajuste às séries de precipitação de “escala fina”, ou seja, com alta resolução temporal de medição (sub-horária), Cowpertwait, Isham e Onof (2007) adicionaram um terceiro processo de Poisson à estrutura da precipitação. Nesse processo terciário, cada célula de chuva desencadeia um aglomerado de pulsos instantâneos, caracterizados somente por uma variável aleatória que representa a lâmina do pulso. Essa concepção é semelhante à de explosões instantâneas, com a diferença de que os pulsos ocorrem no terceiro nível da estrutura da chuva. O modelo foi denominado Pulsos Instantâneos de Bartlett-Lewis (BLIP).

O BLIP de Cowpertwait, Isham e Onof (2007) foi formulado, originalmente, como a superposição de dois modelos independentes, cada qual representando um tipo distinto de chuva (estratiforme e convectiva) com diferentes parâmetros. Embora esse procedimento tenha resultado em um número excessivo de parâmetros para a calibração, o modelo foi

validado na série histórica de precipitação com resolução de 5 minutos e 60 anos de dados de Wellington, Nova Zelândia, com algumas limitações na reprodução dos valores extremos.

Cowpertwait et al. (2011) estenderam o desenvolvimento do BLIP e implementaram uma estrutura em que os pulsos de uma mesma célula são dependentes, em termos de suas alturas pluviométricas, enquanto que os pulsos de células distintas são independentes. Isso foi feito por meio de uma FDP exponencial, para as alturas pluviométricas dos pulsos, condicionada a um parâmetro que varia entre as diferentes células geradas ao longo da trajetória da precipitação. Por parte dos autores, não foi feita uma denominação formal dessa versão aprimorada do BLIP. Neste trabalho ela será referida como BLIPc, em referência a essa distribuição exponencial condicional.

Kaczmarska, Isham e Onof (2014) avaliaram os principais desenvolvimentos dos modelos de Bartlett-Lewis e verificaram que a estrutura com apenas dois processos de Poisson seria suficiente para representar as características até mesmo das séries de escala fina, caso fosse introduzida uma relação de dependência entre a intensidade e a duração das células de um mesmo evento. De forma similar ao aprimoramento feito por Rodriguez-Iturbe, Cox e Isham (1988) que deu origem ao BLRPR, Kaczmarska, Isham e Onof (2014) desenvolveram o modelo BLRPRx que considera essa relação de dependência (a formulação detalhada do BLRPRx foi feita na subseção 4.2).

Na Tabela 3.1 apresenta-se uma síntese dos principais modelos desenvolvidos a partir do BLRP e os respectivos estudos que os introduziram.

Tabela 3.1 – Principais desenvolvimentos realizados a partir do modelo de pulsos retangulares de Bartlett-Lewis.

Modelo	Autores	Descrição
BLRP	Rodriguez-Iturbe, Cox e Isham (1987)	Modelo de pulsos retangulares original.
BLRPR	Rodriguez-Iturbe, Cox e Isham (1988)	Considera a randomização do parâmetro de taxa da FDP das durações das células.
BLRPRG	Onof e Wheeler (1994)	BLRP com a FDP Gama substituindo a FDP exponencial das intensidades das células.
BLIP	Cowpertwait, Isham e Onof (2007)	Considera um terceiro nível hierárquico, no qual, segundo um processo de Poisson, surgem pulsos caracterizados por alturas pluviométricas.
BLIPc	Cowpertwait et al. (2011)	BLIP com relação de dependência entre os pulsos de uma mesma célula.
BLRPRx	Kaczmarska, Isham e Onof (2014)	BLRPR com relação de dependência entre intensidade e duração das células de um mesmo evento.

Os modelos baseados no processo de Neyman-Scott também passaram por uma série de aprimoramentos.

O primeiro foi realizado por Entekhabi, Rodriguz-Iturbe e Eagleson (1989) que aplicaram um tratamento estatístico ao NSRP e modificaram o modelo a fim de flexibilizar o parâmetro de taxa da FDP da duração das células para que variasse entre eventos distintos, de modo similar ao que foi feito no desenvolvimento do BLRPR. Entretanto, para esse modelo, denominado simplesmente como NSRP modificado, não foi atribuído o mesmo desenvolvimento matemático, com a derivação de expressões analíticas para certas propriedades estatísticas que foram desenvolvidas para o BLRPR.

Como consequência, o modelo NSRP original recebeu mais desenvolvimentos do que o NSRP modificado, sobretudo com o tratamento matemático realizado por Cowpertwait (1991, 1998) que derivou expressões analíticas para a proporção de períodos secos e para o momento de terceira ordem do NSRP original, entre outras propriedades estatísticas.

Uma maior disponibilidade de expressões analíticas de certas propriedades estatísticas, tais como o momento de terceira ordem e a proporção de períodos secos, é extremamente desejável. A importância disso reside no fato de que um modelo mais simplista, no que se refere à sua estrutura de representação da precipitação, que foi calibrado com uma maior quantidade de informações estatísticas pode apresentar resultados melhores do que um modelo mais complexo que, por outro lado, foi calibrado somente com os momentos estatísticos de primeira e segunda ordem (WHEATER et al., 2005). Destaca-se que para os modelos mais complexos, expressões analíticas para o momento de terceira ordem, ou de ordens superiores, e para a proporção de períodos secos podem estar indisponíveis, pois isso exige um tratamento matemático e estatístico eventualmente impraticável. Além disso, geralmente os modelos mais complexos têm mais parâmetros, o que torna sua calibração mais difícil com possíveis problemas, como a existência de parâmetros distintos resultando no mesmo valor da função objetivo.

Dois importantes desenvolvimentos do NSRP original foram apresentados por Cowpertwait (1994) e Cowpertwait (1995). No primeiro, o autor implementou a possibilidade de ocorrência de diferentes tipos de células nos eventos de chuva. O modelo foi denominado NSRP Generalizado (GNSRP). No GNSRP, os parâmetros inseridos nas FDPs das durações e intensidades são específicos ao tipo de célula que é atribuído a cada pulso retangular. Essa atribuição é feita conforme probabilidades que são parâmetros de calibração do modelo. Cowpertwait (1994) descreveu uma aplicação do GNSRP com dois tipos de células: células “pesadas”, com origem em chuvas convectivas e com menor duração, e células “leves”, com

origem em sistemas estratiformes e com maior duração. O GNSRP foi aplicado em estudos de reabilitação de sistemas de drenagem pluvial no Reino Unido (COWPERTWAIT et al., 1996).

O GNSRP foi estendido à modelagem espacial por Cowpertwait (1995). Trata-se do modelo espaço-temporal de Neyman-Scott, que ficou conhecido pela sigla STNSRP. No STNSRP, a distribuição temporal dos eventos de chuva segue o processo de Poisson descrito para o NSRP original. A diferença é que no plano x-y ocorre um processo de Poisson, definido por um parâmetro de densidade, segundo o qual os centros de discos uniformes, representando as células de chuva, são distribuídos espacialmente. Os discos possuem duas características, um raio e uma intensidade, ambas são variáveis aleatórias. O modelo foi formulado de modo que pode ser ajustado a partir dos dados observados de uma única estação pluviométrica (sítio único) ou de duas ou mais estações pluviométricas (sítios múltiplos). A partir do GNSRP, podem ser simuladas séries de chuva para localidades sem dados observados.

Em relação à modelagem estocástica da precipitação pontual, as aplicações da modelagem espaço-temporal são mais raras na literatura (BURTON et al., 2008). Não obstante, o GNSRP espaço-temporal foi implementado na plataforma computacional *RainSim*, que tem capacidade de simular séries de precipitação para uma localidade ou para uma área com até 200 km de diâmetro. O *Rainsim* é utilizado pela *UK Water Industry* através de um software distribuído comercialmente (BURTON et al., 2008). Uma versão mais recente, lançada em 2008, o *Rainsim V3* conta com a utilização do momento de terceira ordem na calibração e com a implementação da modelagem espaço-temporal através do STNSRP.

Na Tabela 3.2 apresenta-se uma síntese dos principais modelos desenvolvidos a partir do BLRP e os respectivos estudos que os introduziram.

Tabela 3.2 - Principais desenvolvimentos realizados a partir do modelo de pulsos retangulares de Neyman-Scott.

Modelo	Autores	Descrição
NSRP	Rodrigues-Iturbe, Cox e Isham (1987)	Modelo de pulsos retangulares original.
NSRP modificado	Entekhabi, Rodriguez-Iturbe e Eagleson (1989)	Considera a randomização do parâmetro de taxa da FDP das durações das células.
GNSRP	Cowpertwait (1994)	Possibilita ocorrência de células de diferentes tipos.
STNSRP	Cowpertwait (1995)	Desenvolvido para a modelagem espaço-temporal.

Os modelos de pulsos retangulares foram também utilizados no desenvolvimento de métodos de desagregação, que permitem refinar uma série de precipitação com resolução diária, por exemplo, em uma série de dados horários. Nesse caso, primeiramente ajusta-se um

modelo BLRP ou NSRP aos dados de chuva diários. Séries sintéticas são produzidas, a partir dos modelos calibrados, na escala horária, sendo esse procedimento classificado como “redução de escala” (cujo termo em inglês é *downscaling*). A desagregação é realizada por meio de técnicas de ajuste iterativas que modificam a série sintética de modo a compatibilizar as lâminas agregadas da série sintética horária com as correspondentes lâminas diárias observadas. Os procedimentos de ajuste não afetam a estrutura estocástica produzida pelo modelo. Por definição, a redução de escala é diferente da desagregação, que tenta estabelecer consistência entre os dados horários produzidos e os dados diários observados (KOUTSOYIANNIS; ONOF, 2001). Durante as duas últimas décadas, os processos de Poisson foram utilizados sistematicamente em estudos que abordam o problema da desagregação da chuva. Nessa linha de pesquisa, destacam-se os trabalhos de Koutsoyiannis e Onof (2001) e de Kossieris et al. (2018).

Além dos modelos já citados, existem outras alternativas. O NSRP e o BLRP são baseados em aglomerados (*cluster based*). No contexto da modelagem de sítio único, notável é o modelo descrito por Heneker, Lamber e Kuczera (2001) denominado Modelo Desagregado de Pulso Retangular de Intensidade (DRIP), que por sua vez é baseado em eventos (*event based*). Ele considera a chuva como uma sequência de eventos e de tempos entre eventos. Cada evento é modelado em dois estágios: no primeiro, o evento ocorre na forma de um pulso retangular de intensidade, similar ao BLRP e NSRP, sendo atribuídas distribuições para a intensidade e duração, e também para o tempo entre eventos; e no segundo estágio ocorre a desagregação do pulso por meio de um passeio aleatório (tradução literal de *random walk*). A calibração do DRIP pode ser feita com o método da máxima verossimilhança.

Frost, Srikanthan e Cowpertwait (2004) compararam os resultados do DRIP aos do NSRP para 10 estações na Austrália. Ambos os modelos preservaram os momentos estatísticos de primeira, segunda e terceira ordem, sendo, portanto, igualmente recomendados. Observou-se, no entanto, que o NSRP foi mais eficiente em preservar as características referentes aos valores extremos, pois resultou em uma maior aproximação das curvas de intensidade – duração – frequência (IDF), sobretudo para eventos de curta duração. O DRIP, contudo, foi mais eficiente em preservar as estatísticas referentes aos períodos secos das séries originais. O DRIP foi considerado como um modelo menos suscetível ao viés produzido por falhas na série original e cujos parâmetros possuem um significado físico mais bem definido (HENEKER; LAMBERT; KUCZERA, 2001; JENNINGS; LAMBERT; KUCZERA, 2010). Recentemente o DRIP foi utilizado na geração de séries de precipitação com alta resolução

temporal para análise de sensibilidade de um modelo de transporte de sedimentos (HANCOCK; VERDON-KIDD; LOWRY, 2017).

3.2 Estudos recentes

Nesta subseção, serão apresentados alguns trabalhos que foram publicados recentemente e que utilizaram os modelos de pulsos retangulares como parte de seus métodos. O objetivo é contextualizar, de modo não exaustivo, o cenário atual dos modelos estocásticos de precipitação e somar informações para futuros desenvolvimentos.

Atualmente, os modelos estocásticos de precipitação vêm sendo aplicados no contexto das mudanças climáticas. O estudo de Wasko e Sharma (2017) pode ser citado como exemplo. No método proposto, o modelo NSRP original foi utilizado para simular séries de chuva em um cenário de aquecimento climático. Em uma primeira etapa, propriedades estatísticas dos dados observados foram calculadas, em uma base mensal, para os diferentes anos disponíveis na série histórica. Por meio de modelos de regressão, foram estabelecidas relações funcionais entre essas propriedades e as respectivas temperaturas médias mensais. As propriedades foram então extrapoladas para temperaturas mais altas e utilizadas na calibração do modelo NSRP. Ao final, foram geradas séries sintéticas representativas do clima mais quente. O estudo foi validado a partir dos dados de duas estações meteorológicas da Austrália.

Ritschel et al. (2017) avaliaram se o modelo BLRP é capaz de reproduzir as estatísticas decorrentes da relação de dependência entre intensidade e duração existente nos dados de precipitação, que geralmente são sintetizadas pelas curvas IDF. O BLRP foi ajustado aos dados de uma série com 13 anos de duração e resolução de 1 minuto. A série histórica contempla um evento raro, cujo tempo de retorno para duração de 1 hora foi estimado em 1790 anos. Foram construídas curvas IDF a partir dos dados observados e a partir dos dados de séries sintéticas produzidas com o modelo calibrado. Em ambos os casos, as curvas IDF foram construídas segundo uma extensão paramétrica da distribuição generalizada de valores extremos (GEV), que permite o ajuste simultâneo aos extremos observados de todas as durações consideradas na análise. Desconsiderando o evento raro, foram obtidas curvas IDF dos dados observados e simulados muito próximas, possibilitando concluir que o BLRP reproduz as estatísticas contidas nessas curvas. Verificou-se, no entanto, que se o evento raro for considerado na construção das curvas, elas apresentam discrepâncias. Nesse caso, as IDF dos dados simulados são menos suscetíveis à influência do evento raro do que as IDF dos dados observados. Melhores resultados poderiam ter sido obtidos com a utilização do

BLRPRx, embora seja pouco provável que mesmo o modelo aprimorado consiga reproduzir um evento raro. Isso decorre do fato de os valores extremos não serem utilizados no ajuste dos parâmetros dos modelos, ao passo que são diretamente utilizados no ajuste das curvas IDF.

Kossieris et al. (2018) abordaram o problema da desagregação de séries com resolução temporal diária. A proposta dos autores é um método baseado no modelo BLRPRx cuja estrutura, que conta com a dependência entre intensidade e duração das células de chuva, supostamente é mais apropriada para desagregar os dados originais até a escala de 1 minuto. Para cada sequência de dias chuvosos, verificada na série original, são produzidas iterativamente séries com menor escala, de 5 minutos, por exemplo, até que se obtenha uma sequência sintética com mesmo número de dias chuvosos que a sequência observada. A seguir, as intensidades são atribuídas iterativamente aos eventos e células gerados estocasticamente na sequência sintética, até que as lâminas diárias obtidas a partir da agregação dos dados sintéticos sejam compatíveis com as lâminas observadas correspondentes, dentro de uma tolerância pré-estabelecida. Finalmente, as lâminas desagregadas são ajustadas por meio de uma simples relação linear. Esse processo, que parece ser extremamente intensivo em termos computacionais, é repetido para as demais sequências de dias chuvosos da série observada. Ao final, obtém-se uma série desagregada a partir da série diária.

Esse esquema de desagregação foi implementado em um pacote computacional denominado *HyetosMinute*, com linguagem R, que é de livre utilização. Segundo o método proposto por Kossieris et al. (2018), seria possível utilizar séries de precipitação diária, que são mais longas e amplamente disponíveis, em modelos de simulação contínua que necessitam de séries com resolução muito maior, da ordem de alguns minutos, e cuja disponibilidade é extremamente baixa. No entanto, os resultados empíricos no estudo de Kossieris et al. (2018) apresentam uma deficiência. A desagregação foi aplicada a uma série histórica agregada na escala diária, denominada escala maior, cuja resolução original de medição era de 5 minutos, denominada escala menor. Os resultados sintéticos produzidos com o BLRPRx na escala menor foram comparados com os resultados observados correspondentes, que nesse caso eram conhecidos, e se mostraram muito promissores. Entretanto, a calibração do modelo foi feita com a informação estatística dos dados da escala menor, que em situações reais não estariam disponíveis. Os autores mencionaram que o cenário de desagregação “totalmente operacional”, ou seja, em que os dados da escala menor são desconhecidos, será abordado em estudos futuros.

Cross et al. (2018) abordaram o problema da reprodução dos valores extremos, que geralmente são subestimados nas séries sintéticas dos modelos de Poisson agregadas nas menores escalas (horária ou sub-horária). A hipótese dos autores é que isso ocorre devido às lâminas muito pequenas, inclusive nulas ou inferiores à resolução de medição dos pluviômetros (geralmente 0,1 mm), que dominam o hietograma observado cujos momentos estatísticos centrais são utilizados na calibração. Foram implementados “censores”, ou seja, limites a partir dos quais as lâminas foram removidas da série original, resultando na precipitação “censurada”, que efetivamente foi utilizada na calibração. Para o censor de 0,1 mm, por exemplo, todas as lâminas inferiores a 0,1 mm são convertidas para zero e todas as lâminas superiores são diminuídas de 0,1 mm. Dessa forma, o ajuste dos parâmetros é feito para a parcela “pesada” da precipitação. O método foi aplicado em duas séries históricas, da Inglaterra e Alemanha, e séries sintéticas foram produzidas, nas quais os censores foram restaurados, segundo as principais versões do modelo de Bartlett-Lewis (BLRP, BLRPR e BLRPRx). Os resultados possibilitaram concluir que as séries censuradas reproduzem muito melhor os extremos observados e que as séries geradas com o BLRPRx têm um desempenho superior em relação aos demais modelos avaliados. No entanto, ainda há uma tendência em subestimar os extremos com tempo de retorno superior a 10 anos, embora as discrepâncias sejam muito menores do que as obtidas com as séries não censuradas. Duas fraquezas desse método são a subjetividade na determinação dos valores dos censores, o que pode exigir um número excessivo de simulações, e o fato de que o hietograma observado é reproduzido apenas parcialmente (somente a parcela não censurada é efetivamente modelada) e não há sentido em utilizar séries sintéticas censuradas em modelos de simulação contínua.

No Brasil, foi relatada uma aplicação recente do modelo de Bartlett-Lewis, em sua versão modificada (BLRPR). Teixeira-Gandra e Damé (2014) realizaram a modelagem de 17 anos de dados de chuva da cidade de Pelotas, RS, com resolução de 15 minutos, tendo produzido resultados adequados. A qualidade dos modelos foi avaliada mediante a reprodução dos momentos estatísticos de primeira e segunda ordem e da proporção de períodos secos, simulados e observados. O momento de terceira ordem não foi utilizado na calibração dos modelos e uma análise de valores extremos não foi realizada.

4 MÉTODOS

Nesta seção foram apresentados e discutidos os métodos aplicados à modelagem da série de precipitações observadas.

Na subseção 4.1 foi feita uma discussão para justificar a seleção dos modelos efetivamente utilizados no desenvolvimento deste estudo. Nas subseções 4.2 e 4.3 foram apresentadas as especificações conceituais e estatísticas juntamente à formulação matemática dos modelos selecionados.

Na subseção 4.4 foram descritos os procedimentos para o cálculo das propriedades estatísticas amostrais, posteriormente utilizadas na calibração. Esses procedimentos abrangem a leitura de dados de chuva, o tratamento da série histórica com análise de falhas e a programação desenvolvida para agregar as séries e computar as estatísticas amostrais.

Na subseção 4.5 foram discutidos os principais métodos de calibração dos modelos de precipitação de Poisson. Na subseção 4.5.2 foi apresentado o método efetivamente utilizado na formulação de uma função objetivo (estimador). Para tanto, foram utilizados o equacionamento analítico apresentado na subseção 4.3 e as propriedades estatísticas amostrais, obtidas segundo a descrição na subseção 4.4. Na subseção 4.5.3 foi apresentado o algoritmo de otimização SCE-UA, aplicado à minimização da função objetivo para a estimação dos parâmetros dos modelos selecionados.

Na subseção 4.6 foram descritos os procedimentos realizados no âmbito das simulações, com o desenvolvimento de programas para a geração de séries sintéticas segundo a concepção dos modelos e seus parâmetros calibrados.

Para a execução dos métodos foram desenvolvidos programas em linguagem C++11. Os códigos foram escritos em um editor de texto e compilados no *GNU Compiler Collection* (GCC), ambos softwares livres. As rotinas foram comentadas, de forma elucidativa, e os códigos construídos foram disponibilizados nos Apêndices B.

4.1 Seleção dos modelos

Ao longo das últimas três décadas, foram desenvolvidos vários modelos de precipitação fundamentados nos processos de Poisson. Tratam-se, em sua grande maioria, de versões aprimoradas dos modelos de pulsos retangulares baseados nos processos de Bartlett-Lewis e Neyman-Scott, que foram originalmente propostos por Rodriguez-Iturbe, Cox e Isham (1987).

Conforme mencionado na subseção 3.1, o primeiro aprimoramento do modelo original de Bartlett-Lewis, o BLRPR introduzido por Rodriguez-Iturbe, Cox e Isham (1988), foi abordado em diversos estudos com resultados empíricos de diferentes países. O consenso geral é que o BLRPR reproduz bem os momentos estatísticos de primeira e segunda ordem das séries observadas e tem um desempenho melhor que o BLRP na reprodução da proporção de períodos secos (ONOF et al., 2000; VERHOEST et al., 2010).

Os modelos de Neyman-Scott também foram desenvolvidos e avaliados extensivamente, mas ao contrário dos modelos de Bartlett-Lewis, a sua versão não modificada, o NSRP, recebeu maior atenção. De fato, o NSRP constitui o padrão utilizado nos softwares de projeto de drenagem industrial urbana do Reino Unido (KILSBY et al., 2007; BURTON et al., 2010). Essa priorização se deve em parte à estrutura do processo de Neyman-Scott, que é menos suscetível ao tratamento matemático do que a estrutura do processo de Bartlett-Lewis. Como consequência, a derivação de expressões analíticas para as propriedades estatísticas do NSRP foi menos ampla do que a derivação análoga realizada para o BLRPR (ENTEKHABI; RODRIGUEZ-ITURBE; EAGLESON, 1989). Disso decorre que uma expressão analítica para o cálculo da proporção de períodos secos foi desenvolvida somente para o modelo não modificado (COWPERTWAIT, 1991). O uso do NSRP original, em detrimento da versão modificada, também foi alavancado em três momentos:

1. desenvolvimento do modelo generalizado, GNSRP, cuja concepção considera a sobreposição de dois processos, representando, por exemplo, chuvas dos tipo convectiva e estratiforme (COWPERTWAIT, 1994);
2. extensão à modelagem espaço-temporal (STNSRP) (COWPERTWAIT, 1995);
3. derivação de uma expressão analítica para o momento de terceira ordem do NSRP (COWPERTWAIT, 1998).

No abrangente estudo conduzido por Wheeler et al. (2005) foram avaliadas comparativamente as versões mais proeminentes dos modelos de Bartlett-Lewis e Neyman-Scott. Os autores concluíram que as diferenças entre o BLRP e o NSRP são insignificantes. Além disso, recomendaram o uso do BLRPR, pois verificaram seu bom desempenho e constataram que ele proporciona uma maior estabilidade na identificação dos parâmetros, o que é uma consequência de sua estrutura simplificada de apenas 6 parâmetros.

Com base nessas considerações, optou-se pelos modelos de Bartlett-Lewis para a modelagem realizada neste estudo. Essa escolha também foi motivada por desenvolvimentos recentes da literatura em que foram utilizados os modelos de Bartlett-Lewis (RITSCHHEL et

al., 2017; CROSS et al., 2018; KOSSIERIS et al., 2018). Portanto, a discussão remanescente nesta seção limita-se ao modelo BLRP e seus aprimoramentos.

Uma notável versão surgiu da constatação, de alguns estudos iniciais, que o BLRPR tem um desempenho insatisfatório na reprodução de valores extremos das lâminas observadas (ONOF; WHEATER, 1993). Para superar esse problema, Onof e Wheeler (1994) substituíram a distribuição exponencial das intensidades das células pela distribuição Gama e introduziram o BLRPR Gama (BLRPRG). Supostamente, a distribuição Gama é mais robusta na modelagem de intensidades extremas, devido à sua cauda mais longa, e confere maior flexibilidade à calibração do modelo, devido aos dois parâmetros que a definem. Entretanto, Verhoest, Troch e De Troch (1997) avaliaram o desempenho do BLRPRG em uma série com 27 anos de dados de chuva de Uccle, Bélgica, e não constataram um aprimoramento na reprodução de valores extremos em relação aos modelos BLRP e BLRPR.

Uma deficiência comum aos estudos citados nesta subseção, que talvez não tenha sido visualizada na época das publicações, foi a não inclusão do momento de terceira ordem na estimação dos parâmetros e validação dos modelos. É provável que isso tenha ocorrido porque a expressão analítica do momento de terceira ordem do BLRP só foi publicada em 2006 (WHEATER et al., 2006) e, segundo Vanhaute et al. (2012), até o ano 2012 não havia sido publicada uma expressão análoga para o BLRPR. O momento central de terceira ordem, cuja forma adimensional é o coeficiente de assimetria, consiste na estatística descritiva que reflete a ocorrência de desvios muito grandes na série observada (NAGHETTINI; PINTO, 2007). É pouco provável que os valores extremos de uma série histórica sejam reproduzidos sem a inclusão do momento de terceira ordem, ou de ordem superior, na calibração dos modelos (COWPERTWAIT, 1998).

Em sua tese de doutorado, Kaczmarska (2013) investigou o desempenho do BLRPR e comparou esse modelo com o BLRP original, levando em conta a reprodução do coeficiente de assimetria. Foi verificado que uma melhor reprodução da proporção de períodos secos é de fato oferecida pelo BLRPR, conforme demonstrado por diversos autores, mas que isso ocorre às custas de uma deterioração na reprodução do coeficiente de assimetria, que por sua vez não foi incluído no ajuste dos modelos. Kaczmarska (2013) concluiu então que o BLRPR não oferece um melhoramento real na modelagem da série observada em relação à versão original. Pelo contrário: caso o coeficiente de assimetria não seja incluído na calibração, o que é comum nos estudos com o BLRPR devido à indisponibilidade de uma expressão analítica, esse momento pode ser altamente superestimado nos meses chuvosos. Por esse motivo, optou-se pela modelagem com o BLRP em detrimento da versão modificada (BLRPR).

No estudo de Kaczmarska, Isham e Onof (2014), foram avaliados comparativamente o BLRP e os desenvolvimentos mais recentes dos modelos de Bartlett-Lewis, entre eles os modelos que consideram a sobreposição de dois processos de chuva, representando a ocorrência de precipitações estratiforme e convectiva, de modo análogo ao GNSRP, e o modelo BLIP desenvolvido por Cowpertwait, Isham e Onof (2007) no qual foi implementado um terceiro processo de Poisson com o objetivo de modelar séries de chuva com resolução sub-horária (5 minutos, por exemplo). Os resultados mostraram que os modelos com sobreposição de dois processos contêm um número excessivo de parâmetros (12 no total) que não puderem ser identificados de forma realista e estável mesmo a partir de uma série longa, com 69 anos de dados de chuva de Bochum, Alemanha. Constatação semelhante havia sido feita no estudo de Wheeler et al. (2005). Com relação ao BLIP, Kaczmarska, Isham e Onof (2014) avaliaram uma versão aprimorada, denotada por BLIPR, que considera uma randomização similar à implementada no BLRPR. Os resultados demonstraram que o terceiro processo de Poisson contido no BLIPR produziu um número extremamente alto de pulsos dentro das células, de modo que a estrutura final obtida é semelhante às células retangulares dos modelos BLRP e BLRPR.

Kaczmarska, Isham e Onof (2014) concluíram então que a estrutura de células retangulares é satisfatória e que uma melhoria real no desempenho dos modelos viria pela implementação da relação de dependência entre duração e intensidade das células. Os autores propuseram um novo modelo, denotado por BLRPRx, que considera essa dependência e que foi aplicado à série de chuva de Bochum. Os resultados demonstraram o desempenho superior do BLRPRx em relação ao BLRP original e o BLIP de Cowpertwait, Isham e Onof (2007).

Além de ser um avanço recente na literatura da modelagem de sítio único, duas importantes vantagens do BLRPRx podem ser apontadas:

- parcimônia: o modelo tem apenas 6 parâmetros, um a mais do que o BLRP e o mesmo número de parâmetros que o BLRPR, o que facilita a calibração e atende às recomendações de Onof e Wheeler (1994);
- disponibilidade de uma expressão analítica para o momento de terceira ordem, o que permite a utilização do coeficiente de assimetria na calibração e, possivelmente, uma melhor reprodução dos valores extremos, conforme já mencionado.

Diante da revisão realizada nesta subseção, os modelos BLRP e o BLRPRx foram selecionados para a modelagem deste trabalho.

Deve-se, contudo, ressaltar que o método desenvolvido, em termos de aquisição de dados, calibração, simulação e análise de resultados, e os códigos programados podem ser adaptados para a aplicação de outros modelos de sítio único, a exemplo do NSRP, sendo necessários somente pequenos ajustes na formulação da função objetivo e no algoritmo de simulação.

4.2 Concepção dos modelos

Os modelos da classe de pulsos retangulares, na qual se encaixam o BLRP e BLRPRx, partem do princípio que a precipitação pode ser representada como uma sobreposição de células de chuva. Cada célula é caracterizada por uma duração ($L_{célula}$) e uma intensidade (X). A intensidade é constante ao longo da duração, de modo que as células assumem a forma de “pulsos retangulares”, o que justifica a terminologia adotada.

Os modelos baseados nos processos de Bartlett-Lewis e Neyman-Scott consideram que a precipitação tem uma estrutura hierárquica, segundo a qual eventos de chuva são responsáveis pela geração de “aglomerados” (*clusters*) de células retangulares.

A descrição desses modelos tem como ponto de partida uma série temporal, definida em intervalo contínuo de tempo. Nessa série, a ocorrência de eventos de chuva é governada por um processo de Poisson primário com intensidade, ou taxa, denotada por λ . A formulação conceitual e matemática do processo de Poisson foi revisada no Apêndice A. Neste trabalho, a unidade de tempo adotada foi de horas (hr), de modo que λ tem dimensão hr^{-1} . Como consequência do processo de Poisson, os intervalos de tempo entre os eventos de chuva consistem em variáveis aleatórias contínuas independentes e identicamente distribuídas (i.i.d), que são caracterizadas por uma distribuição exponencial com parâmetro de taxa λ .

No processo de Bartlett-Lewis, cada evento desencadeia um processo de Poisson secundário com taxa β (hr^{-1}), que governa a ocorrência das células de chuva. Os intervalos de tempo entre as células, portanto, também são variáveis aleatórias i.i.d que seguem uma distribuição exponencial com parâmetro de taxa β . A origem da primeira célula sobrepõe a origem do evento, e as células subsequentes são geradas após cada intervalo distribuído segundo a FDP exponencial. A geração de células encerra-se após o tempo que define a duração do evento, denotada por L_{evento} , que também é distribuída exponencialmente com parâmetro de taxa γ (hr^{-1}). De acordo com essas definições, o número de células C geradas

para um evento qualquer é uma variável aleatória caracterizada por uma distribuição geométrica com média $\mu_c = 1 + \beta/\gamma$ (RODRIGUEZ-ITURBE; COX; ISHAM, 1988).

A única diferença entre os modelos de Bartlett-Lewis e Neyman-Scott reside no processo secundário de geração de células. Enquanto que no Bartlett-Lewis os intervalos de tempo entre células sucessivas são variáveis exponenciais i.i.d, no Neyman-Scott são as distâncias temporais entre as origens das células e a origem do evento que são variáveis exponenciais i.i.d. Contrariamente ao Bartlett-Lewis, no processo de Neyman-Scott a primeira célula não sobrepõe a origem do evento. Além disso, não é definido um tempo de duração para os eventos, de modo que alguma distribuição discreta (Poisson ou geométrica) deve ser arbitrada para o número de células C dos eventos (VELGHE et al., 1994).

O BLRP consiste na versão mais simples dos modelos de Bartlett-Lewis. Segundo esse modelo, as durações ($L_{célula}$) e intensidades (X) das células são variáveis aleatórias i.i.d caracterizadas por distribuições exponenciais com parâmetros η (hr^{-1}) e $1/\mu_X$ ($\text{mm}^{-1}.\text{hr}$), respectivamente. Esse modelo tem, portanto, 5 parâmetros ($\lambda, \beta, \gamma, \eta, \mu_X$).

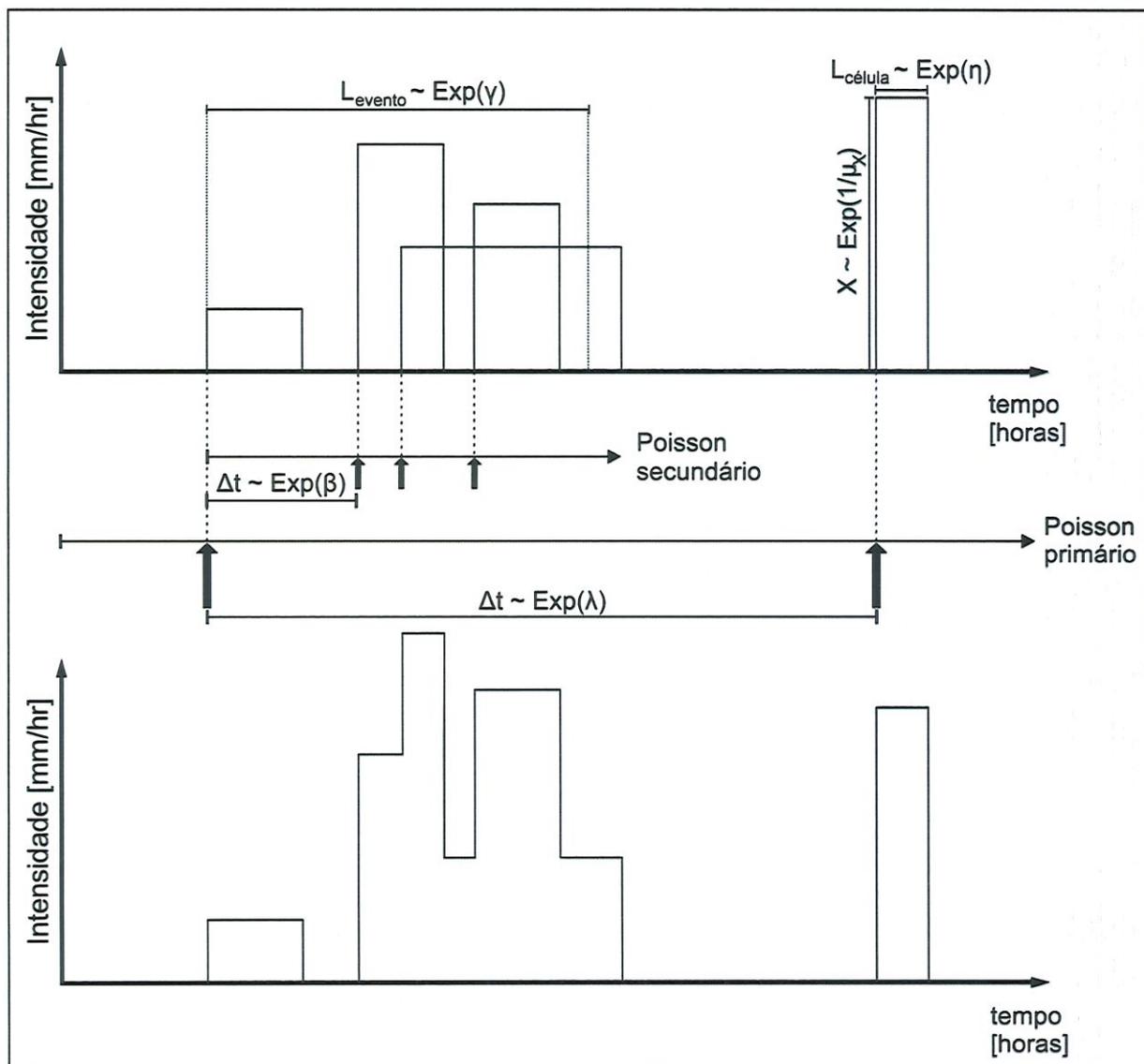
Segundo Rodriguez-Iturbe, Cox e Isham (1987), a intensidade total de chuva $Y(t)$ ($\text{mm}.\text{hr}^{-1}$) é dada pelo somatório de todos os pulsos “ativos” no instante t . Formalmente:

$$Y(t) = \int_{u=0}^{\infty} X_{t-u}(u) dN(t-u), \quad (4.1)$$

em que $X_{t-u}(u)$ é a intensidade do pulso que se origina no tempo $(t-u)$ medida após o tempo u ; e $dN(t-u)$ refere-se ao processo de contagem de Poisson, $N(t)$, em que $dN(t)$ equivale a 1 quando há ocorrência de algum pulso no intervalo $(t, t+dt)$ e a 0 em caso contrário. O desenvolvimento da integral na Equação 4.1 deve considerar a sobreposição tanto das células que têm origem em um mesmo evento, quanto das células com origem em eventos distintos, e isso dificulta o tratamento analítico de $Y(t)$.

A representação esquemática da Figura 4.1 ilustra a concepção do modelo BLRP.

Figura 4.1 – Representação esquemática da concepção do BLRP e variáveis características. O gráfico ilustrado na parte inferior corresponde às intensidades instantâneas totais, obtidas como resultado da sobreposição de eventos e células preconizadas no modelo.



Fonte: próprio autor.

Na primeira modificação do BLRP que originou o BLRPR, foi idealizada a randomização do parâmetro η , que especifica a distribuição exponencial da duração das células, de modo que η variasse entre os diferentes eventos de chuva. No BLRPR, cada evento de chuva recebe um valor específico de η segundo uma distribuição Gama com parâmetro de forma α e de escala ν . Além disso, os parâmetros β e γ foram reparametrizados através dos coeficientes adimensionais constantes $\phi = \gamma/\eta$ e $\kappa = \beta/\eta$.

Dessa forma, no BLRPR o parâmetro $\eta \sim \text{Gama}(\alpha, \nu)$ consiste em uma variável aleatória que, por sua vez, é utilizada como parâmetro nas distribuições de outras três variáveis aleatórias, a saber:

- duração das células ($L_{célula} \sim Exp(\eta)$);
- duração dos eventos ($L_{evento} \sim Exp(\gamma = \phi \cdot \eta)$);
- tempo entre origens de células sucessivas $\sim Exp(\beta = \kappa \cdot \eta)$.

No modelo BLRPR, cada evento de chuva possui valores específicos de η , γ e β , o que permite que os eventos tenham características distintas. Isso confere maior flexibilidade, em relação ao BLRP. Nessa concepção preserva-se uma característica comum a todos os eventos de chuva, por meio dos parâmetros α e ν , o que pode ser o resultado de regionalidade, e ao mesmo tempo permite-se que cada evento tenha suas especificidades.

No BLRPRx, desenvolvido por Kaczmarzka, Isham e Onof (2014), a randomização de η foi estendida ao parâmetro $1/\mu_X$, que no BLRP corresponde à taxa da FDP das intensidades das células. Nesse caso, a reparametrização foi feita através do coeficiente constante $\iota = \mu_X/\eta$ (mm). Em sua configuração mais simples, o BLRPRx adota uma distribuição exponencial para a intensidades das células, de modo que:

$$X \sim Exp\left(\frac{1}{\mu_X} = \frac{1}{\iota \cdot [\eta \sim Gama(\alpha, \nu)]}\right). \quad (4.2)$$

De acordo com essa formulação, há uma dependência estatística implícita entre a intensidade das células e as durações das células e dos eventos.

A versão que adota a distribuição exponencial conforme a Equação 4.2, com apenas um parâmetro para as intensidades, foi utilizada neste estudo. Essa versão tem 6 parâmetros de calibração ($\lambda, \iota, \alpha, \nu, \kappa, \phi$). Versões alternativas que adotam as distribuições de dois parâmetros de Weibull ou Gama também foram testadas pelos autores do BLRPRx, mas os resultados não justificaram a maior dificuldade decorrente da adição de um parâmetro extra na calibração (KACZMARSKA; ISHAM; ONOF, 2014).

No BLRPRx, as durações dos eventos (L_{evento}) e células ($L_{célula}$) e as intensidades (X) têm suas distribuições exponenciais condicionadas ao parâmetro η . Dessa forma, para um evento de chuva específico ao qual foi atribuído o valor η_0 , por exemplo, os valores esperados dessas variáveis são:

$$E[L_{evento}] = \frac{1}{\gamma} = \frac{1}{\phi \eta_0}. \quad (4.3)$$

$$E[L_{célula}] = \frac{1}{\eta_0}. \quad (4.4)$$

$$E[X] = \mu_x = \iota \cdot \eta_0. \quad (4.5)$$

Nota-se que quanto maior for o valor de η_0 , menor será a duração dos eventos (ϕ da Equação 4.3 é constante) e as durações de suas respectivas células, e maior será a intensidade dessas células (ι da Equação 4.5 também é constante). Por meio dessa análise simplificada, fica evidente que o BLRPRx traduz a dependência existente entre intensidade e duração da precipitação, que é coerente com a prática de Hidrologia.

Conforme já mencionado, L_{evento} e $L_{célula}$ seguem distribuições exponenciais condicionadas a η . As distribuições não condicionais dessas variáveis, que dependem dos parâmetros de calibração α e ν , constantes para toda a série temporal, são de Pareto do tipo II (KACZMARSKA; ISHAM; ONOF, 2014). Portanto, para a duração das células, por exemplo, a FDP característica é dada por:

$$f_{L_{célula}}(l_{célula}) = \frac{\alpha \nu^\alpha}{(\nu + l)^{\alpha+1}}, l \geq 0. \quad (4.6)$$

Essa função tem média infinita se o parâmetro α for menor do que 1 e variância infinita se α for menor do que 2. Isso exige cautela na etapa de calibração: se for obtido $\alpha < 2$ no ajuste do modelo aos dados observados, é possível que as séries sintéticas geradas na etapa subsequente de simulação tenham períodos de chuva muito longos, que não atendem à realidade. Kaczmarska, Isham e Onof (2014) sugerem que α seja restrito ao intervalo $[2, +\infty)$ durante a calibração.

As Tabelas 4.1 e 4.2 representam uma síntese das variáveis características da precipitação e suas respectivas distribuições segundo a concepção dos modelos BLRP e BLRPRx, respectivamente.

Tabela 4.1 – BLRP: variáveis características e suas distribuições.

Variável	Distribuição	Parâmetro
tempo entre eventos	Exponencial	λ
duração dos eventos (L_{evento})	Exponencial	γ
tempo entre células	Exponencial	β
duração das células ($L_{célula}$)	Exponencial	η
intensidade das células (X)	Exponencial	$1/\mu_x$

Tabela 4.2 – BLRPRx: variáveis características e suas distribuições.

Variável	Distribuição	Parâmetros
tempo entre eventos	Exponencial	λ
parâmetro η (específico de cada evento)	Gama	α, ν
duração dos eventos (L_{evento})	Exponencial (condicionada a η)	$\phi. [\eta \sim Gama(\alpha, \nu)]$
tempo entre células	Exponencial (condicionada a η)	$\kappa. [\eta \sim Gama(\alpha, \nu)]$
duração das células ($L_{célula}$)	Exponencial (condicionada a η)	$\eta \sim Gama(\alpha, \nu)$
intensidade das células (X)	Exponencial (condicionada a η)	$\frac{1}{\iota. [\eta \sim Gama(\alpha, \nu)]}$

Uma vez esclarecidas as variáveis características dos modelos e suas respectivas distribuições de probabilidades, retorna-se a discussão da Equação 4.1.

Como resultado das variáveis aleatórias inseridas no BLRP e no BLRPRx (tempo de origem dos eventos, parâmetros η , duração das células, intensidade das células, duração dos eventos, etc.), é evidente que as intensidades instantâneas totais $Y(t)$ também são variáveis aleatórias. De acordo com a concepção teórica de cada modelo, a partir da Equação 4.1 foram derivadas expressões analíticas para os valores esperados dos momentos estatísticos de $Y(t)$. Esses momentos correspondem à média, variância, autocovariâncias de lag k ($k \in Z$) e ao momento central de ordem 3 populacionais.

No entanto, a precipitação não é medida através das intensidades instantâneas, mas sim através das lâminas, que podem efetivamente ser quantificadas nos pluviômetros. Portanto, em sua maioria, os modelos de precipitação de Poisson são calibrados segundo as propriedades estatísticas das lâminas observadas. A partir dos momentos analíticos de $Y(t)$, foram derivadas expressões analíticas dos valores esperados dos momentos de até a terceira ordem das lâminas para o BLRP e o BLRPRx, que foram transcritas na subseção 4.3.

Julga-se importante definir o termo “lâmina” como a precipitação total acumulada em intervalos de tempo disjuntos de duração fixa, denotada por h . A dimensão característica das lâminas é o comprimento, geralmente milímetros (mm). Sob a ótica dos modelos estocásticos de precipitação, as lâminas consistem na realização de variáveis aleatórias, denotadas por Y_i^h , que caracterizam o “processo agregado” (conforme denominação utilizada na literatura). As lâminas sequenciais, precipitadas em intervalos discretos de duração h , relacionam-se com as intensidades instantâneas por meio da seguinte integral:

$$Y_i^h = \int_{(i-1)h}^h Y(t) dt. \quad (4.7)$$

Esses esclarecimentos são pertinentes porque os termos “lâmina” e “processo agregado” foram amplamente utilizados neste manuscrito.

Conforme mencionado, a quase totalidade dos modelos de Poisson utiliza as propriedades estatísticas das lâminas observadas na calibração. Contudo, existem exceções, a exemplo dos modelos baseados no processo estocástico duplo de Poisson (DSPP) (RAMESH; ONOF; XIE, 2012). A modelagem do DSPP, diferentemente do BLRP ou NSRP, é feita diretamente com os dados brutos dos pluviômetros de balança, que são os instantes de tempo em que ocorre cada basculada, e não com as lâminas acumuladas. O DSPP apresenta vantagens, sobretudo no método de calibração, mas tem sua aplicação limitada à disponibilidade desses dados brutos. Séries de tamanho considerável contendo dados desse tipo ainda são escassas no Brasil e o monitoramento instantâneo da chuva ainda é incipiente na grande maioria dos países.

4.3 Expressões analíticas dos momentos estatísticos

Nesta subseção foram apresentados os valores esperados dos momentos estatísticos de Y_i^h . Tratam-se de funções dos parâmetros que foram utilizadas no equacionamento do estimador formulado na subseção 4.5.2. A derivação de expressões analíticas para esses momentos exige um tratamento matemático bastante complexo, com base em técnicas estatísticas avançadas, que não será abordado em detalhes.

Para as intensidades das células, X , foram adotadas distribuições exponenciais com parâmetro de taxa $1/\mu_X$. As esperanças matemáticas, decorrentes dessa distribuição, são:

$$E(X^2) = 2\mu_X^2. \quad (4.8)$$

$$E(X^3) = 6\mu_X^3. \quad (4.9)$$

Essas relações foram utilizadas no cálculo dos momentos analíticos do BLRP e BLRPRx.

Em algumas equações do BLRPRx foram utilizadas ainda os quocientes $f_1 = E(X^2)/(\mu_X^2) = 2$ e $f_2 = E(X^3)/(\mu_X^3) = 6$, que são constantes caso a distribuição exponencial seja mantida para a variável aleatória X .

Para o BLRP, define-se o valor esperado do número de células por evento, μ_C , segundo:

$$\mu_C = 1 + \frac{\beta}{\gamma}. \quad (4.10)$$

As expressões analíticas dos momentos de primeira e segunda ordem do modelo BLRP foram derivadas por (RODRIGUEZ-ITURBE; COX; ISHAM, 1987) e estão apresentadas na forma das Equações 4.11 – 4.13.

Média

$$E[Y_i^h] = \frac{\lambda \mu_X \mu_C h}{\eta}. \quad (4.11)$$

Variância

$$\text{var}[Y_i^h] = \frac{2\lambda\mu_C}{\eta} \left[\frac{h}{\eta} \left(E(X^2) + \frac{\beta\mu_X^2}{\gamma} \right) + \frac{\mu_X^2\beta\eta(1-e^{-\gamma h})}{\gamma^2(\gamma^2-\eta^2)} - \left(E(X^2) + \frac{\beta\gamma\mu_X^2}{\gamma^2-\eta^2} \right) \left(\frac{1-e^{-\eta h}}{\eta^2} \right) \right]. \quad (4.12)$$

Autocovariância de lag $k \geq 1$

$$\text{cov}[Y_i^h, Y_{i+k}^h] = \frac{\lambda\mu_C}{\eta} \left[\left(E(X^2) + \frac{\beta\gamma\mu_X^2}{\gamma^2-\eta^2} \right) \left(\frac{(1-e^{-\eta h})^2 e^{-\eta(k-1)h}}{\eta^2} \right) - \frac{\mu_X^2\beta\eta(1-e^{-\gamma h})^2 e^{-\gamma(k-1)h}}{\gamma^2(\gamma^2-\eta^2)} \right]. \quad (4.13)$$

A Equação 4.14 é a expressão analítica do momento central de terceira ordem do BLRP, derivada por Wheater et al. (2006):

Momento central de ordem 3

$$E \left[\left(Y_i^h - E(Y_i^h) \right)^3 \right] \quad (4.14)$$

$$= \frac{\lambda \mu_c}{(\eta^2 + 2\gamma\eta + \gamma^2)(\gamma^4 - 2\eta\gamma^3 - 3\eta^2\gamma^2 + 8\eta^3\gamma - 4\eta^4)\eta^4\gamma^3}$$

$$\times \left\{ \begin{aligned} & 12\gamma^7\mu_X^3\beta^2e^{-\eta h} - 48\mu_X^3e^{-\gamma h}\beta^2\eta^7 + 72\gamma^7E(X^3)\eta^2 + 48\gamma\mu_XE(X^2)\beta\eta^7 \\ & - 24\gamma h\mu_X^3\eta^7\beta^2 + 24\mu_XE(X^2)\gamma^4he^{-\eta h}\beta\eta^5 + 24\mu_XE(X^2)\gamma^2\beta\eta^6 - 36\mu_XE(X^2)\gamma^3\beta\eta^5 \\ & - 72\mu_XE(X^2)\gamma^6h\beta\eta^3 - 84\gamma^2\mu_X^3\beta^2\eta^5 - 36\gamma^5h\mu_X^3\beta^2\eta^3 + 54\gamma^3h\mu_X^3\eta^5\beta^2 \\ & + 6\mu_XE(X^2)\gamma^5\beta\eta^3 + 6\gamma^7h\mu_X^3\eta\beta^2 - 24\mu_XE(X^2)\gamma^2e^{-\eta h}\beta\eta^6 + 117\mu_XE(X^2)\gamma^6\beta\eta^2 \\ & - 18\gamma^4\mu_X^3e^{-\eta h}\beta^2\eta^3 + 54\gamma^5hE(X^3)\eta^5 + 39\gamma^5\mu_X^3\beta^2\eta^2 - 36\gamma^7hE(X^3)\eta^3 \\ & - 24\gamma^3hE(X^3)\eta^7 - 12\gamma^9E(X^3) + 6\eta\gamma^9hE(X^3)e^{-\eta h} \\ & + 6\mu_XE(X^2)\gamma^4e^{-\gamma h}\beta\eta^4 - 30\mu_XE(X^2)\gamma^6he^{-\eta h}\beta\eta^3 - 48\mu_XE(X^2)\gamma^2h\beta\eta^7 \\ & - 48\gamma\mu_XE(X^2)e^{-\gamma h}\beta\eta^7 - 24\gamma h\mu_X^3e^{-\gamma h}\eta^7\beta^2 + 6\gamma^5\mu_X^3e^{-\gamma h}\beta^2\eta^2 \\ & + 6\mu_XE(X^2)\gamma^8he^{-\eta h}\beta\eta - 138\mu_XE(X^2)\gamma^4\beta\eta^4 + 48\mu_X^3\beta^2\eta^7 + 30\gamma^3h\mu_X^3e^{-\gamma h}\eta^5\beta^2 \\ & - 24\mu_XE(X^2)\gamma^2e^{-\gamma h}\beta\eta^6 + 36\mu_XE(X^2)\gamma^3e^{-\gamma h}\beta\eta^5 + 24\mu_XE(X^2)\gamma^3e^{-\eta h}\beta\eta^5 \\ & - 132\mu_XE(X^2)\gamma^6e^{-\eta h}\beta\eta^2 - 6\mu_XE(X^2)\gamma^5e^{-\gamma h}\beta\eta^3 + 150\mu_XE(X^2)\gamma^4e^{-\eta h}\beta\eta^4 \\ & - 42\gamma^5\mu_X^3e^{-\eta h}\beta^2\eta^2 - 6\gamma^5h\mu_X^3e^{-\gamma h}\beta^2\eta^3 + 12\mu_XE(X^2)\gamma^8h\beta\eta \\ & - 6\mu_XE(X^2)\gamma^5e^{-\eta h}\beta\eta^3 - 24E(X^2)\mu_X\beta\gamma^3e^{-h(\eta+\gamma)}\eta^5 - 12E(X^2)\mu_X\gamma^4\beta e^{-2\eta h}\eta^4 \\ & - 6E(X^2)\mu_X\gamma^4\beta e^{-h(\eta+\gamma)}\eta^4 + 6\gamma^5E(X^2)\mu_X\beta e^{-h(\eta+\gamma)}\eta^3 - 3E(X^2)\mu_X\gamma^8\beta e^{-2\eta h} \\ & + 24E(X^2)\mu_X\beta\gamma^2e^{-h(\eta+\gamma)}\eta^6 + 15E(X^2)\mu_X\gamma^6\beta e^{-2\eta h}\eta^2 - 3\gamma^7\mu_X^3\beta^2e^{-2\eta h} \\ & + 18\gamma^4\mu_X^3\beta^2e^{-h(\eta+\gamma)}\eta^3 - 12\gamma^3\mu_X^3\beta^2e^{-h(\eta+\gamma)}\eta^4 - 6\gamma^5\mu_X^3\beta^2e^{-h(\eta+\gamma)}\eta^2 \\ & + 3\gamma^5\mu_X^3\beta^2e^{-2\eta h}\eta^2 - 9\gamma^7\mu_X^3\beta^2 + 108\gamma^5E(X^3)e^{-\eta h}\eta^4 + 48\gamma^3E(X^3)\eta^6 \\ & - 72\gamma^7E(X^3)e^{-\eta h}\eta^2 - 48E(X^3)\gamma^3e^{-\eta h}\eta^6 + 84\gamma^2\mu_X^3e^{-\gamma h}\beta^2\eta^5 + 18\gamma^4\mu_X^3\beta^2\eta^3 \\ & + 24E(X^2)\mu_X\gamma^8\beta e^{-\eta h} + 54\gamma^5hE(X^3)e^{-\eta h}\eta^5 - 24\eta^7hE(X^3)\gamma^3e^{-\eta h} \\ & - 36\eta^3\gamma^7hE(X^3)e^{-\eta h} - 21\gamma^8E(X^2)\mu_X\beta + 6\gamma^9hE(X^3)\eta + 12\gamma^3\mu_X^3e^{-\eta h}\beta^2\eta^4 \\ & + 12\gamma^3\mu_X^3e^{-\gamma h}\beta^2\eta^4 - 18\gamma^4\mu_X^3e^{-\gamma h}\beta^2\eta^3 - 12\gamma^3\mu_X^3\beta^2\eta^4 - 108\gamma^5E(X^3)\eta^4 \\ & + 108\mu_XE(X^2)\gamma^4h\beta\eta^5 + 12\gamma^9E(X^3)e^{-\eta h} \end{aligned} \right\}.$$

As Equações 4.17 – 4.20 correspondem aos valores esperados dos momentos de primeira a terceira ordem do BLRPRx que foram derivados por Kaczmarzka (2013).

As esperanças matemáticas foram mantidas na forma $E_{\eta}[\eta^{-k}e^{-\eta s}]$ para valores arbitrários de k e s . Essa expressão é avaliada segundo:

$$E_{\eta}[\eta^{-k}e^{-\eta s}] = \frac{\nu^{\alpha}}{\Gamma(\alpha)} \times \frac{\Gamma(\alpha - k)}{(\nu + s)^{\alpha - k}}, \text{ para } \alpha > k, \quad (4.15)$$

em que $\Gamma(\alpha)$ é a função Gama para α . De acordo com Naghettini e Pinto (2007), a função Gama foi aproximada por:

$$\Gamma(\alpha) = \left[\frac{\sqrt{2\pi}}{\alpha} \left(p_0 + \sum_{i=1}^6 \frac{p_i}{\alpha + i} \right) \right] (\alpha + 5,5)^{\alpha + 0,5} e^{-(\alpha + 5,5)}, \text{ com}$$

$$\begin{aligned} p_0 &= 1,000000000190015, \\ p_1 &= 76,18009172947146, \\ p_2 &= -86,50532032941677, \\ p_3 &= 24,01409824083091, \\ p_4 &= -1,231739572450155, \\ p_5 &= 1,208650973866179 \times 10^{-3}, \\ p_6 &= -5,395239384953 \times 10^{-6}. \end{aligned} \quad (4.16)$$

Média

$$E[Y_i^h] = \lambda h \mu_c. \quad (4.17)$$

Variância

$$\begin{aligned} \text{var}[Y_i^h] &= 2\lambda\mu_c l^2 \left\{ \left[f_1 + \frac{\kappa}{\phi} \right] h + E_{\eta}(\eta^{-1}) \left[\frac{\kappa(1 - \phi^3)}{\phi^2(\phi^2 - 1)} - f_1 \right] \right. \\ &\quad \left. - E_{\eta}(\eta^{-1}e^{-\phi\eta h}) \frac{\kappa}{\phi^2(\phi^2 - 1)} + E_{\eta}(\eta^{-1}e^{-\eta h}) \left[f_1 + \frac{\kappa\phi}{\phi^2 - 1} \right] \right\}. \end{aligned} \quad (4.18)$$

Autocovariância de lag $k \geq 1$

$$\begin{aligned} \text{cov}[Y_i^h, Y_{i+k}^h] &= \lambda\mu_c l^2 \left\{ \left(f_1 + \frac{\kappa\phi}{\phi^2 - 1} \right) [E_{\eta}(\eta^{-1}e^{-\eta(k-1)h}) - 2E_{\eta}(\eta^{-1}e^{-\eta kh}) \right. \\ &\quad \left. + E_{\eta}(\eta^{-1}e^{-\eta(k+1)h})] - \frac{\kappa}{\phi^2(\phi^2 - 1)} [E_{\eta}(\eta^{-1}e^{-\phi\eta(k-1)h}) \right. \\ &\quad \left. - 2E_{\eta}(\eta^{-1}e^{-\phi\eta kh}) + E_{\eta}(\eta^{-1}e^{-\phi\eta(k+1)h})] \right\}. \end{aligned} \quad (4.19)$$

Momento central de ordem 3

$$\begin{aligned}
 & E \left[\left(Y_i^h - E(Y_i^h) \right)^3 \right] \tag{4.20} \\
 &= \frac{\lambda \mu_c t^3}{(1 + 2\phi + \phi^2)(\phi^4 - 2\phi^3 - 3\phi^2 + 8\phi - 4)\phi^3} \\
 &\times \left\{ E_\eta \left[\eta^{-1} e^{-\eta h} \right] \left(12\phi^7 \kappa^2 - 24f_1 \phi^2 \kappa - 18\phi^4 \kappa^2 + 24f_1 \phi^3 \kappa - 132f_1 \phi^6 \kappa + 150f_1 \phi^4 \kappa \right. \right. \\
 &\quad \left. \left. - 42\phi^5 \kappa^2 - 6f_1 \phi^5 \kappa + 108\phi^5 f_2 - 72\phi^7 f_2 - 48\phi^3 f_2 + 24f_1 \mu_x \phi^8 \kappa + 12\phi^3 \kappa^2 + 12\phi^9 f_2 \right) \right. \\
 &\quad + E_\eta \left[e^{-\eta h} \right] \left(24f_1 \phi^4 h \kappa + 6\phi^9 h f_2 - 30f_1 \phi^6 h \kappa + 6f_1 \phi^8 h \kappa + 54\phi^5 h f_2 - 24h f_2 \phi^3 - 36\phi^7 h f_2 \right) \\
 &\quad + E_\eta \left[\eta^{-1} e^{-\eta \phi h} \right] \left(-48\kappa^2 + 6f_1 \phi^4 \kappa - 48\phi f_1 \kappa + 6\phi^5 \kappa^2 - 24f_1 \phi^2 \kappa + 36f_1 \phi^3 \kappa \right. \\
 &\quad \left. - 6f_1 \phi^5 \kappa + 84\phi^2 \kappa^2 + 12\phi^3 \kappa^2 - 18\phi^4 \kappa^2 \right) \\
 &\quad + E_\eta \left[e^{-\eta \phi h} \right] \left(-24\phi h \kappa^2 + 30\phi^3 h \kappa^2 - 6\phi^5 h \kappa^2 \right) \\
 &\quad + E_\eta \left[\eta^{-1} \right] \left(72\phi^7 f_2 + 48\phi f_1 \kappa + 24f_1 \phi^2 \kappa - 36f_1 \phi^3 \kappa - 84\phi^2 \kappa^2 + 6f_1 \phi^5 \kappa + 117f_1 \phi^6 \kappa \right. \\
 &\quad \left. + 39\phi^5 \kappa^2 - 12\phi^9 f_2 - 138f_1 \phi^4 \kappa + 48\kappa^2 - 9\phi^7 \kappa^2 + 48\phi^3 f_2 + 18\phi^4 \kappa^2 - 21\phi^8 f_1 \kappa \right. \\
 &\quad \left. - 12\phi^3 \kappa^2 - 108\phi^5 f_2 \right) \\
 &\quad + \left(-24\phi h \kappa^2 - 72f_1 \phi^6 h \kappa - 36\phi^5 h \kappa^2 + 54\phi^3 h \kappa^2 + 6\phi^7 h \kappa^2 + 54\phi^5 h f_2 - 36\phi^7 h f_2 \right. \\
 &\quad \left. - 24\phi^3 h f_2 - 48f_1 \phi^2 h \kappa + 12f_1 \phi^8 h \kappa + 6\phi^9 h f_2 + 108f_1 \phi^4 h \kappa \right) \\
 &\quad + E_\eta \left[\eta^{-1} e^{-2\eta h} \right] \left(-12f_1 \phi^4 \kappa - 3f_1 \phi^8 \kappa + 15f_1 \phi^6 \kappa - 3\phi^7 \kappa^2 + 3\phi^5 \kappa^2 \right) \\
 &\quad + E_\eta \left[\eta^{-1} e^{-\eta h(1+\phi)} \right] \left(-24f_1 \kappa \phi^3 - 6f_1 \phi^4 \kappa + 6\phi^5 f_1 \kappa + 24f_1 \kappa \phi^2 + 18\phi^4 \kappa^2 \right. \\
 &\quad \left. - 12\phi^3 \kappa^2 - 6\phi^5 \kappa^2 \right) \left. \right\}.
 \end{aligned}$$

As Equações 4.11 – 4.14 e 4.17 – 4.20 foram utilizadas juntamente com os respectivos momentos amostrais no equacionamento dos estimadores do BLRP e do BLRPRx, respectivamente. Os parâmetros de cada modelo foram estimados através da minimização desse estimador. Esse procedimento será descrito na subseção 4.5.

4.4 Aquisição dos dados observados

O primeiro passo para a aquisição dos dados de chuva foi a escolha de uma série histórica de precipitação com resolução temporal sub-horária de longa duração.

A resolução sub-horária foi estabelecida como critério de escolha, em um primeiro momento, porque o estudo de Kaczmarska, Isham e Onof (2014), que introduziu o BLRPRx, é feito para uma série histórica com resolução de 5 minutos.

A Tabela 4.3 apresenta as características das séries históricas utilizadas em estudos representativos. Com base nessas informações, foi estabelecido que a série utilizada na modelagem deveria ter uma duração superior a 10 anos, o que corresponde a menor duração verificada nos estudos anteriores.

Tabela 4.3 – Duração e resolução das séries históricas utilizadas em estudos representativos.

Estudo	Local	Duração (anos)	Resolução
Rodriguez-Iturbe, Cox e Isham (1987)	Denver, EUA	28	1 hr
Cowpertwait (1991)	Blackpool, Inglaterra	10	1 hr
Onof e Wheeler (1994)	Birmingham, Inglaterra	38,5	1 hr
Vanhaute et al. (2012)	Uccle, Bélgica	105	10 min
Kaczmarska, Isham e Onof (2014)	Bochum, Alemanha	69	5 min

A proposta inicial era modelar as séries de chuva de São Carlos, São Paulo, provenientes de uma rede de monitoramento instalada no âmbito do projeto “Manejo de Águas Pluviais em Meio Urbano” (MAPLU). Essa rede conta com 8 pluviômetros de balsa distribuídos em diferentes localidades do município. Os dados estão disponíveis na forma bruta, ou seja, na forma dos instantes de ocorrência das basculadas. Isso permite que as séries de precipitação geradas a partir dos dados de São Carlos tenham resoluções arbitrárias, limitadas, evidentemente, à resolução de leitura do instrumento de medição que, por sua vez, é da ordem de segundos.

Apesar de as séries contarem com uma resolução sub-horária adequada, elas têm durações muito curtas, que eram inferiores a cinco anos na época de início das atividades deste estudo. A opção pelas séries de São Carlos do MAPLU foi descartada, embora o contínuo monitoramento possa permitir a utilização desses dados em desenvolvimentos futuros. Uma vez que os dados brutos dos pluviômetros estão disponíveis, um modelo de intensidades instantâneas como o DSPP pode ser utilizado na modelagem.

Verificou-se a existência de séries longas com lâminas de resolução diária disponíveis no território nacional, enquanto que séries longas com lâminas em resolução horária ou sub-horária são escassas. Em Piracicaba, São Paulo, há uma série de precipitação com 21 anos de duração e que conta com resolução sub-horária de 15 minutos, atendendo, portanto, aos critérios estabelecidos.

A série de Piracicaba foi escolhida para a modelagem devido à facilidade de aquisição dos dados e à proximidade da EESC-USP.

4.4.1 Série histórica de precipitação

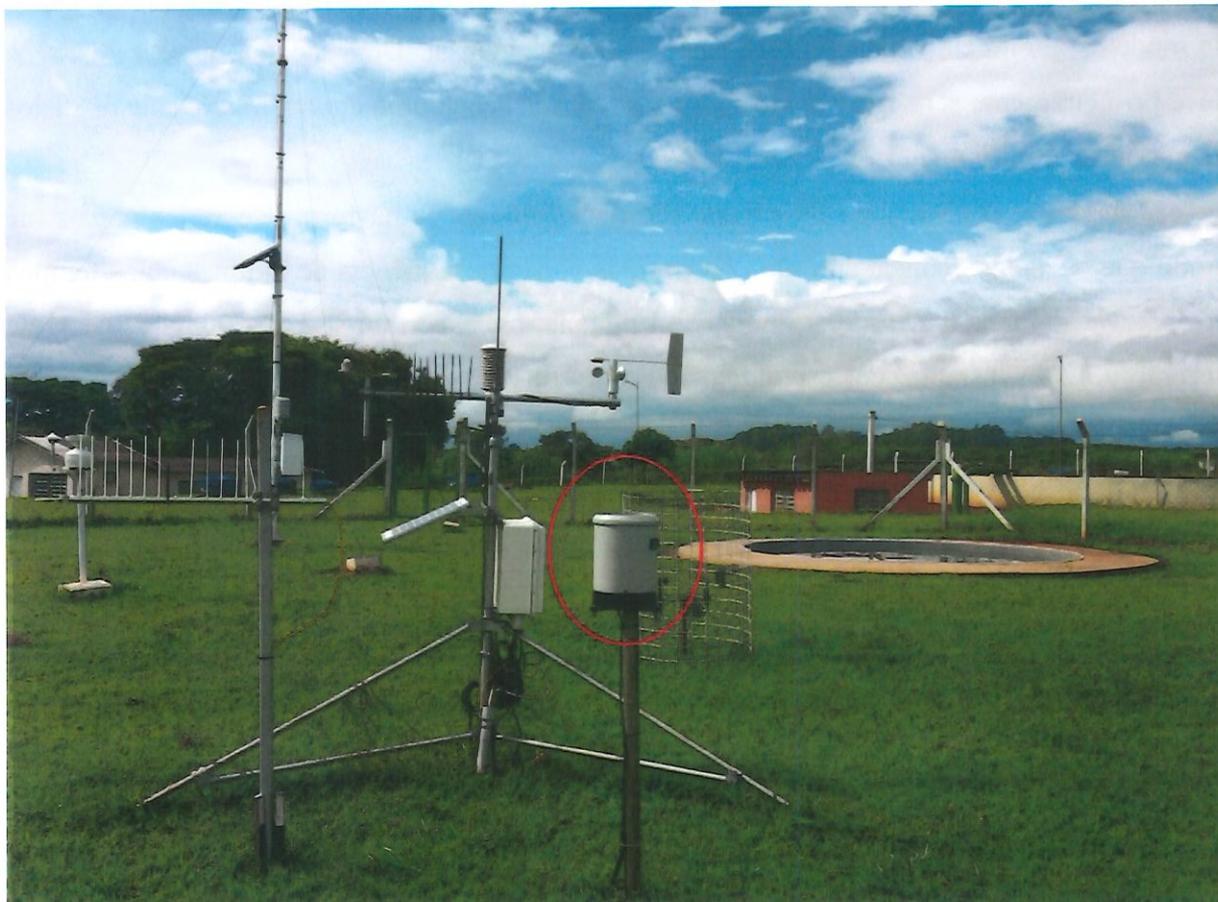
A série de chuva foi adquirida no posto meteorológico da Escola Superior de Agricultura Luiz de Queiroz (ESALQ-USP) em Piracicaba. O posto localiza-se nas coordenadas geográficas 22°42,5 Sul e 47°38,0 Oeste, na altitude 546 metros. Ele é mantido pelo Departamento de Engenharia de Biosistemas (LEB) e conta com duas estações meteorológicas:

- Estação Meteorológica Convencional (EMC), que desde 1917 registra diariamente as precipitações através de um pluviômetro;
- Estação Meteorológica Automática (EMA), que desde janeiro de 1997 registra as precipitações a cada 15 minutos medidas por um pluviômetro de balança.

Somente as lâminas registradas pela EMA no período de janeiro/1997 a outubro/2017 foram utilizadas na composição da série temporal de dados observados.

No dia 5 de maio de 2017 foi feita uma visita técnica ao posto meteorológico da ESALQ-USP. A Figura 4.2 apresenta uma fotografia da EMA e do pluviômetro de balança cujos dados foram utilizados na modelagem.

Figura 4.2 – Estação Meteorológica Automática (EMA) da ESALQ-USP em Piracicaba-SP. O pluviômetro de balsa utilizado na modelagem é apresentado com destaque em vermelho. Fotografia de 05/05/2017.



Fonte: próprio autor.

No caso da EMA, o pluviômetro converte automaticamente o número de basculadas observado a cada 15 minutos nas lâminas que, por sua vez, são salvas em um *data logger* e disponibilizadas diariamente para consulta. Para economizar memória, os dados brutos do pluviômetro são eliminados após a conversão. Nota-se que essa operação torna impossível refinar a série para uma resolução maior ou modelar os dados com o DSPP.

Os dados da EMA são de livre uso e podem ser acessados diretamente na página web¹ do posto meteorológico.

4.4.2 Estatísticas amostrais

O processamento dos dados observados foi feito com o auxílio de programas codificados na linguagem C++11. Cada programa consiste em um arquivo com formato *.cpp* pronto para compilação e execução. Os códigos desenvolvidos no âmbito da aquisição dos dados observados estão disponíveis no Apêndice B-1.

¹ <http://www.esalq.usp.br/departamentos/leb/postoaut.html> (Acesso em 21 de março de 2018).

Os modelos de Poisson têm como pré-requisito a estacionariedade das séries temporais. Para conciliar essa estacionariedade com a sazonalidade, inerente à precipitação, os modelos geralmente são calibrados em uma base mensal. Dessa forma, a série histórica deve primeiramente ser separada em 12 séries mensais, cada uma contendo somente os dados de um mês específico.

Os dados do posto de Piracicaba são disponibilizados na forma de séries anuais, de modo que os dados de cada ano são constituídos por uma sequência de lâminas cujos tempos de observação iniciam-se às 00:15 horas de 1 de janeiro e encerram-se às 24:00 horas de 31 de dezembro. Esses instantes correspondem ao limite superior do intervalo de 15 minutos em que ocorre o acúmulo de precipitação. No caso da série original, esse intervalo é de 15 minutos. 21 séries anuais contemplando o período 1997-2017 foram adquiridas diretamente do site do posto meteorológico e processadas no programa “gerar_series.cpp”.

O algoritmo contido nesse programa realiza um número de iterações equivalente ao número de anos disponíveis na série histórica. Neste caso foram realizadas 21 iterações para o processamento dos 21 anos disponíveis. A cada iteração, o algoritmo percorre a sequência e identifica o mês ao qual o registro pertence. O registro contém basicamente o instante de observação e a respectiva lâmina acumulada. Após a identificação do mês, o registro é salvo em um arquivo separado, com formato .csv, que recebe o nome abreviado do mês e a escala de agregação temporal: “Fev_15min.csv”, por exemplo. Considera-se na execução do algoritmo a possibilidade de o ano ser bissexto. Como resultado do programa “gerar_series.cpp”, são produzidas 12 séries mensais contendo registros no formato ilustrado na Figura 4.3. A inclusão de dias julianos, dias do mês e horas no formato HHMM deve-se ao formato original adotado pelo posto meteorológico, que foi reproduzido a fim de conferir a eventual existência de discrepâncias nos resultados do programa “gerar_series.cpp”.

Figura 4.3 – Exemplo de registro da série “Jan_15min.csv” produzida com ao programa “gerar_series.cpp”.

Ano	Dia juliano	Hora (10:45)	Dia do mês	Mês	Hora decimal
1997	2	1045	2	Jan	10.75

Lâmina acumulada entre às 10:30 e 10:45

A etapa subsequente consiste na agregação das séries mensais. Para calibrar os modelos de Poisson, utilizam-se propriedades estatísticas calculadas tanto para a série com a resolução original, neste caso 15 minutos, quanto para as séries agregadas, em que as lâminas são acumuladas em intervalos de tempo maiores que a resolução. Os intervalos devem ser múltiplos da resolução e a sua duração é denominada escala de agregação temporal, ou simplesmente escala de agregação. Dessa forma, a série agregada na escala de 1 hora corresponde à sequência constituída pela acumulação das lâminas pertencentes a quatro intervalos consecutivos de uma série com resolução de 15 minutos. As séries mensais foram agregadas através do programa “agregar_series.cpp”.

Antes da execução desse programa alguns meses foram eliminados das séries mensais porque apresentaram falhas, conforme foi discutido na subseção 4.4.3.

O programa “agregar_series.cpp” foi configurado para as 5 escalas de agregação de 1, 2, 6, 12 e 24 horas. O algoritmo contido em “agregar_series.cpp” realiza 12 iterações, referentes aos meses do calendário. Em cada iteração, a respectiva série mensal gerada na etapa anterior (“gerar_series.cpp”) é acessada e as lâminas são acumuladas produzindo um número de séries agregadas equivalente ao tamanho do vetor “escalas”. Como resultado desse processamento, foram produzidas $12 \times 5 = 60$ séries agregadas. Cada série é salva em um arquivo .csv contendo o registro de tempos de observação e as lâminas acumuladas. O nome do arquivo que recebe a série de fevereiro na escala de 2 horas, por exemplo, é “Fev_2_hr.csv”.

Finalmente, as 72 séries mensais (de 15 min e agregadas) foram utilizadas no cálculo das estatísticas amostrais que, por sua vez, foram utilizadas na formulação da função objetivo descrita em detalhes na subseção 4.5.2. Para tanto, foi desenvolvido o programa “calcular_estatisticas.cpp”, que acessa todas as séries mensais e computa média, variância, autocovariâncias de *lag* k ($k = 1, 2, 3$) e momento central de ordem 3 das lâminas agregadas, com base nas Equações 4.21- 4.24. Nessas equações, y_i^h representa uma lâmina, com dimensão de comprimento, da escala de agregação arbitrária h . Para o cálculo dos momentos amostrais (observados) de cada mês, foram utilizadas as lâminas de todos os anos disponíveis. Dessa forma, as estatísticas de janeiro, por exemplo, reúnem a informação de janeiro de 1997, janeiro de 1999, ..., e janeiro de 2017.

Média
$$E[y_i^h] = \frac{1}{n} \sum_{i=1}^n y_i^h, \quad (4.21)$$

$$\text{Variância} \quad \text{var}[y_i^h] = \frac{1}{n-1} \sum_{i=1}^n (y_i^h - E[y_i^h])^2, \quad (4.22)$$

$$\text{Autocovariância de lag } k \quad \text{cov}[y_i^h, y_{i+k}^h] = \frac{1}{n-1} \sum_{i=1}^{n-1} (y_i^h - E[y_i^h]) (y_{i+k}^h - E[y_{i+k}^h]), \quad (4.23)$$

$$\text{Momento central de ordem 3} \quad E[(y_i^h - E[y_i^h])^3] = \frac{n}{(n-1)(n-2)} \sum_{i=1}^n (y_i^h - E[y_i^h])^3. \quad (4.24)$$

Além desses momentos, o programa computa a proporção de períodos secos (PPS) e a probabilidade de transição úmido-úmido (PTuu).

A PPS foi extraída das séries histórica segundo a relação entre o número intervalos com lâminas menores que 0,1 mm e o número total de intervalos. A PTuu, por sua vez, foi calculada segundo:

$$PT_{uu} = P(y_i^h \geq 0.1 | y_{i-1}^h \geq 0.1) = \frac{P(y_i^h \geq 0.1 \cap y_{i-1}^h \geq 0.1)}{P(y_i^h \geq 0.1)}, \quad (4.25)$$

em que o quociente da Equação 4.25 corresponde à razão entre o número de dois períodos úmidos consecutivos, com lâminas maiores do que 0,1 mm, e o número total de períodos úmidos. Note-se que como critério para considerar um período úmido utilizou-se o limite inferior nas lâminas de 0,1 mm, o que representa a resolução do instrumento de medição. Esse critério foi aplicado tanto à série histórica, quanto às séries simuladas, nas quais lâminas menores que 0,1 mm são possíveis.

Os resultados são impressos sequencialmente no arquivo “estatísticas_observadas.csv”. Para os 12 meses, 6 escalas (incluindo resolução) e 8 propriedades estatísticas, foram produzidos 576 dados que caracterizam a série histórica de Piracicaba. A sub-rotina “organizar_estatísticas” classifica os dados de forma hierárquica por propriedade e, em um nível inferior, por escala seguindo a ordem crescente: 15 min, 1 hr, ..., 24 hr. Em seguida, todas as informações são agrupadas em uma matriz de 12 linhas, cada linha representando um mês do calendário. Essa matriz é impressa no arquivo “estatísticas_observadas_ORGANIZADO.csv”.

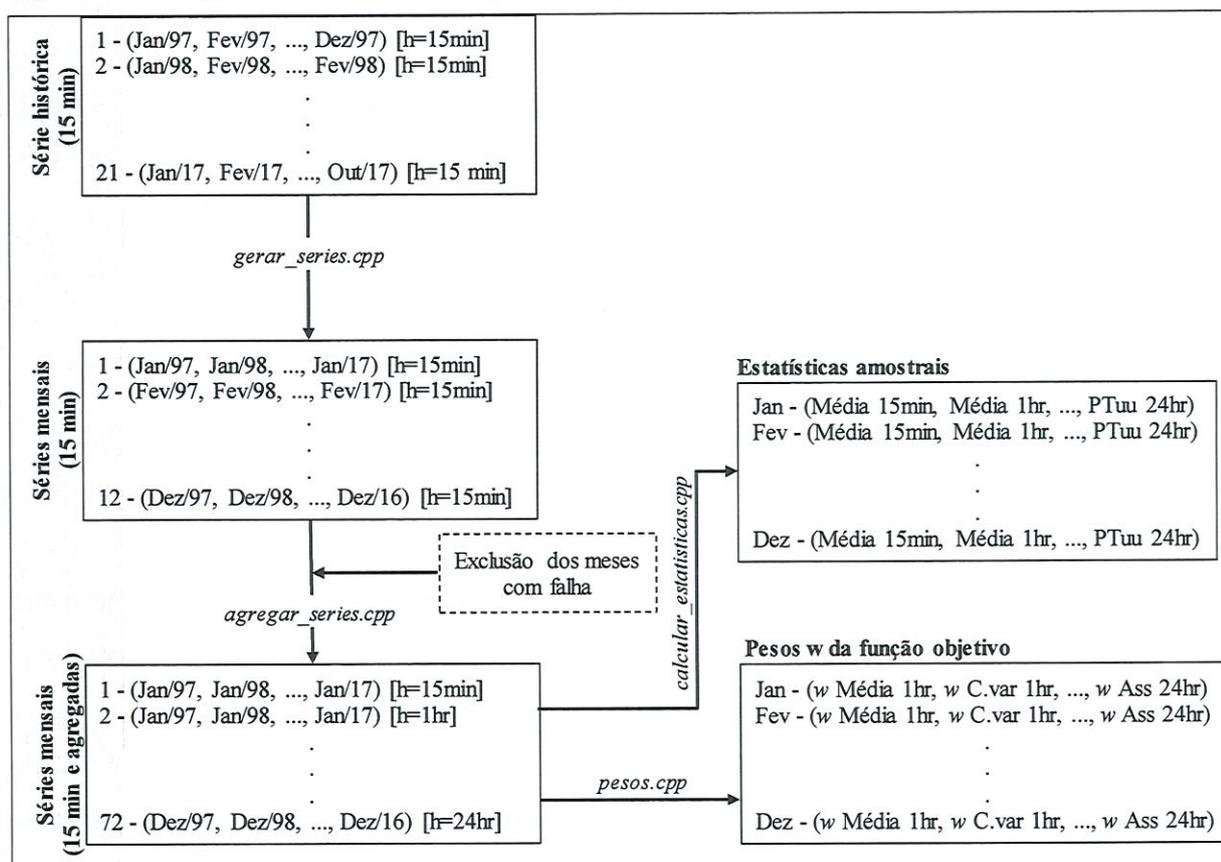
Na formulação da função objetivo foram utilizados pesos que, segundo a formulação adotada para o calibrador descrito na subseção 4.5.2, devem ser a partir dos momentos amostrais de cada ano. Para tanto, foi desenvolvido o programa “pesos.cpp”.

Diferentemente do programa “calcular_estatisticas.cpp”, o programa “pesos.cpp” percorre as séries mensais e computa separadamente os momentos estatísticos de cada ano. Se 21 anos de dados estão disponíveis para o mês de janeiro, a média das lâminas de 1 hora, por exemplo, é computada separadamente para janeiro de 1997, janeiro de 1998, ..., e janeiro de 2017, resultando em uma sequência de 21 repetições interanuais dessa propriedade. O peso que é associado à média de 1 hora corresponde ao inverso da variância amostral dessas 21 repetições. Da mesma forma, foram calculadas as variâncias das demais propriedades estatísticas.

Nem todos os momentos e escalas de agregação foram considerados na calibração dos modelos. Dessa forma, somente os pesos associados aos momentos efetivamente inseridos na função objetivo são computados por *pesos.cpp*. Esse programa produz um relatório para cada mês contendo os momentos interanuais e os respectivos pesos.

Os procedimentos realizados na etapa de aquisição estão ilustrados no fluxograma da Figura 4.4.

Figura 4.4 – Fluxograma ilustrativo dos procedimentos e programas executados na etapa de aquisição.



Fonte: próprio autor.

Nota-se que o grande número de dados estatísticos extraídos da série de um único posto meteorológico torna indispensável o uso de uma programação robusta para a modelagem estocástica.

Alternativamente, os códigos em “gerar_series.cpp”, “agregar_series.cpp”, “calcular_estatisticas.cpp” e “pesos_cpp” podem ser convertidos para um único programa, de modo que com apenas uma execução sejam obtidas as estatísticas amostrais e os pesos, que consistem na informação efetivamente usada na calibração. No entanto, o encapsulamento proposto e a execução sequencial dos procedimentos de aquisição trazem vantagens. Uma delas foi a possibilidade de intervir nas séries mensais de 15 minutos, produzidas por “gerar_series.cpp”, para a exclusão dos meses com falhas, conforme será discutido na próxima subseção.

4.4.3 Análise de falhas

As séries anuais da EMA, disponibilizadas em sua forma original no site do LEB-ESALQ-USP, apresentaram dados duvidosos ou falhas em diversos registros. Para cada registro da série histórica contendo falhas ou inconsistências foi executada uma análise visando à correção ou à eliminação desses dados duvidosos. Isso preveniu que lâminas de magnitudes inconsistentes e períodos secos ou chuvosos duvidosos, supostamente observados, introduzissem erros na calibração, sobretudo nos parâmetros implícitos das distribuições das variáveis envolvidas no modelo estocástico (*e. g.* número de eventos chuvosos, intensidade e duração dos pulsos, etc.).

Segundo o observado, os dados duvidosos podem ser classificados em três tipos:

1. registros com informações evidentemente ilógicas tais como precipitações negativas e registros duplicados que especificam um mesmo intervalo de tempo (estes últimos sem qualquer prejuízo à informação contida no intervalo);
2. registros que especificam intervalos de tempo para os quais não há informação das lâminas precipitadas, que se manifestam como células vazias na coluna referente às lâminas ou que simplesmente não constam no arquivo da série adquirida;
3. inconsistências entre os totais precipitados diários da EMA, obtidos mediante agregação dos registros de 15 minutos em escala de 24 horas (por meio do programa “agregar_series.cpp”), e os dados correspondentes que foram publicados no site do posto meteorológico como os totais diários observados na estação automática.

Os dados duvidosos classificados como tipo 1 foram corrigidos manualmente. Os registros duplicados foram primeiramente identificados por meio de uma rotina especialmente programada com esse propósito e que, portanto, não será disponibilizada nos Apêndices. Destaca-se que as séries contêm muitos registros, de modo que uma verificação visual dessas inconsistências seria por demais dispendiosa. Após a identificação, os registros duplicados foram excluídos. Já as precipitações negativas foram substituídas por lâminas nulas (0,0 mm). Esse procedimento é consistente, pois foi verificado nos registros da EMC que nos dias correspondentes aos intervalos com precipitações negativas não houve chuva.

Procedimento semelhante foi adotado para lidar com os dados do tipo 2. Para cada dia correspondente aos intervalos com falhas foi verificada a ocorrência de precipitação nos registros da EMC, atentando-se para o fato de que na EMC a anotação de um determinado dia refere-se, mais precisamente, ao período entre as sete horas desse dia e as sete horas do dia subsequente. Em grande parte dos casos não foi constatada precipitação, de modo que as células vazias (ou inexistentes) foram preenchidas com lâminas nulas e esses dados foram utilizados na modelagem. Nos casos em que houve chuva, os meses contendo os intervalos com falha foram removidos das séries mensais, segundo o procedimento ilustrado na Figura 4.4. A Tabela 4.4 mostra em maiores detalhes os resultados dessa análise de falhas.

Tabela 4.4 – Falhas do tipo 2 identificadas na série histórica.

(continua)				
Ano	Intervalo	Duração	Constatação no registro da EMC	Ação relativa à falha
1997	5/Nov	1,25 horas	Sem precipitação	Preenchimento com 0,0 mm
	12-16/Mar	~4 dias	Sem precipitação	Preenchimento com 0,0 mm
	10/Maio	1 hora	Sem precipitação	Preenchimento com 0,0 mm
1998	18/Maio	14,25 horas	Sem precipitação	Preenchimento com 0,0 mm
	28/Maio	1 hora	Lâmina de 0,1 mm	Preenchimento com 0,0 mm
	1-4/Jun	~3 dias	Sem precipitação	Preenchimento com 0,0 mm
	16-17/Jul	18,5 horas	Sem precipitação	Preenchimento com 0,0 mm
2000	24-26/Jan	~3 dias	Precipitação de 36,8 mm	Eliminação de janeiro/2000
	7/Abr	1,25 horas	Sem precipitação	Preenchimento com 0,0 mm
	24-25/Abr	~2 dias	Sem precipitação	Preenchimento com 0,0 mm
	28/Jul	3 horas	Sem precipitação	Preenchimento com 0,0 mm
2001	1/Fev	2,5 horas	Sem precipitação	Preenchimento com 0,0 mm
	7/Maio	4,75 horas	Sem precipitação	Preenchimento com 0,0 mm
2002	23/Abr	45 minutos	Sem precipitação	Preenchimento com 0,0 mm
	9-11/Out	~3 dias	Sem precipitação	Preenchimento com 0,0 mm
2004	7-15/Jun	~9 dias	Precipitação de 9,1 mm	Eliminação de junho/2004
	16-23/Ago	~8 dias	Sem precipitação	Preenchimento com 0,0 mm

(conclusão)

Ano	Intervalo	Duração	Constatação no registro da EMC	Ação relativa à falha
2006	28-29/Mar	6,5 horas	Precipitação de 30,5 mm	Eliminação de março/2006
	20-31/Dez	~12 dias	Precipitação de 170,2 mm	Eliminação de dezembro/2006
2007	1-2/Jan	~2 dias	Precipitação de 12,4 mm	Eliminação de janeiro/2007
2008	9-11/Abr	~3 dias	Precipitação de 61,4 mm	Eliminação de abril/2008
	12-15/Set	~3 dias	Precipitação de 13,7 mm	Eliminação de setembro/2008
2009	26/Out	6,75 horas	Sem precipitação*	Preenchimento com 0,0 mm
2010	7/Jan	4 horas	Precipitação de 9,0 mm	Eliminação de janeiro/2010
	2/Set	6,75 horas	Sem precipitação	Preenchimento com 0,0 mm
	21/Set	4 horas	Precipitação de 38,0 mm	Eliminação de setembro/2010
	13/Dez	3,5 horas	Precipitação de 27,6 mm	Eliminação de dezembro/2010
2011	8/Jan	3,25 horas	Sem precipitação*	Preenchimento com 0,0 mm
	14/Fev	3,5 horas	Precipitação de 10,7 mm	Eliminação de fevereiro/2011
	5/Abr	3,5 horas	Sem precipitação*	Preenchimento com 0,0 mm
	12/Abr	3,75 horas	Sem precipitação*	Preenchimento com 0,0 mm
	25/Abr	1 hora	Sem precipitação	Preenchimento com 0,0 mm
	17/Dez	6,5 horas	Sem precipitação	Preenchimento com 0,0 mm
2016	4/Mar-6/Abr	~33 dias	Precipitação de 126,7 mm em março Sem precipitação no período de abril	Eliminação de março/2016
	3/Maio	8 horas	Sem precipitação	Preenchimento com 0,0 mm
	5/Ago	10 horas	Sem precipitação	Preenchimento com 0,0 mm

*Ocorre precipitação no dia, mas em um intervalo fora do período da falha observada no registro da EMA

De acordo com as informações do funcionário responsável pela coleta dos dados do posto meteorológico, as células vazias se devem ao entupimento do funil receptor do pluviômetro. Além disso, consta no registro um período sem dados de aproximadamente um mês, entre março e abril de 2016, devido à troca do pluviômetro da EMA. Nenhuma explicação nos foi dada a respeito dos intervalos de tempo sem a devida anotação do registro, a exemplo dos dias entre 12 e 15 de setembro de 2008, para os quais o registro salta das 7:15 horas do primeiro dia para as 8:00 horas do segundo.

Os dados duvidosos classificados como falhas do tipo 3 foram constatados após a execução do programa “agregar_series.cpp”. As séries mensais de 15 minutos foram agregadas em várias escalas, dentre as quais inclui-se a escala de 24 horas, o que permitiu comparar os totais diários precipitados com os “dados diários” disponibilizados na página web do posto meteorológico, na seção referente à estação automática: <http://www.esalq.usp.br/departamentos/leb/automatica/pagina5> (acesso em 22/03/2018). Obviamente esperava-se que a execução do programa produzisse lâminas diárias iguais

àquelas publicadas em “dados diários”. Contudo, foram verificadas discrepâncias significativas entre o que foi obtido por meio do programa de agregação das séries de 15 minutos e o que se encontra disponibilizado pelos responsáveis pela coleta e gerenciamento dos dados.

Apresentam-se, a seguir, quatro exemplos de meses contendo dados duvidosos do terceiro grupo, a fim de justificar os procedimentos de correção adotados.

- Julho de 2002

Após a agregação dos registros de 15 minutos, esse mês apresentou uma precipitação total de 199,7 mm, o que é muito superior à média mensal calculada segundo os dados remanescentes da EMA para julho (33,8 mm). Analisando a série original de 15 minutos, foram verificadas precipitações com intensidade média de até 333,6 mm/hr que totalizam lâminas de 24,2 e 147,6 mm nos dias 23 e 24, respectivamente. Conforme exibido na Tabela 4.5, os “dados diários” da EMA, disponíveis no site do posto meteorológico, apresentam lâminas nulas nos cinco dias duvidosos grifados, enquanto que nos demais dias os resultados são idênticos ao que foi produzido pelo algoritmo de agregação. Coincidentemente, os dados diários da EMC indicam que não houve precipitação nesses cinco dias.

- Junho de 2006

Na série original consta que no mês de junho de 2006 não houve precipitação. Os “dados diários” da EMA, contudo, indicam que no dia 26 de junho de 2006 precipitaram 18,6 mm. Essa lâmina é idêntica à precipitação verificada no registro da EMC para o mesmo dia. Alerta-se que, no caso da modelagem estocástica em estudo, um longo período contínuo seco configura uma informação relevante na calibração dos parâmetros.

- Fevereiro de 2013

De forma similar ao mês julho de 2002, os dados de fevereiro de 2013 apresentam diferenças entre o que foi produzido por agregação da série original e o que foi publicado nos “dados diários” da EMA, em dias específicos. Segundo a Tabela 4.6, verifica-se que nos dias duvidosos (grifados) as lâminas diárias supostamente observadas na EMA são idênticas às observadas na EMC. Nos demais dias, há consistência entre os totais produzidos por agregação e os “dados diários” disponibilizados para a EMA.

- Maio de 2016

Para o dia 23 desse mês a lâmina precipitada segundo agregação da série original foi de 5,4 mm. Entretanto, os dados diários publicados tanto da EMA quanto da EMC acusam que não houve precipitação nesse dia.

Esses exemplos são representativos de vários outros casos em que há inconsistência entre as lâminas agregadas a partir da série original e as lâminas disponibilizadas nos “dados diários”. A forma que encontramos para identificar essas inconsistências foi comparar os totais precipitados em cada mês referentes aos resultados da agregação dos “dados diários” divulgados. É visível que em certos dias os dados da estação convencional foram utilizados para substituir os dados da estação automática. Os motivos que levaram os responsáveis pela coleta a realizarem tal procedimento são desconhecidos. Não obstante, os registros com essas inconsistências foram considerados duvidosos, até mesmo porque as informações diárias da EMA e da EMC têm horários de coleta distintos: no último caso a coleta ocorre às 7:00 horas da manhã. Dessa forma, os meses em que foram verificadas essas inconsistências foram eliminados das séries utilizadas na calibração.

Tabela 4.5 – Análise de falhas para o mês julho de 2002. Os registros duvidosos foram grifados em cinza.

Dia	Lâmina (mm)		
	Resultado da agregação (série original de 15 minutos)	Dados diários (EMA)	Dados diários (EMC)
1/jul/2002	0.0	0.0	0.0
2/jul/2002	0.0	0.0	0.0
3/jul/2002	0.0	0.0	0.0
4/jul/2002	0.0	0.0	0.0
5/jul/2002	0.1	0.1	0.0
6/jul/2002	0.0	0.0	0.0
7/jul/2002	0.0	0.0	0.0
8/jul/2002	0.0	0.0	0.0
9/jul/2002	0.1	0.1	0.0
10/jul/2002	0.1	0.1	0.0
11/jul/2002	2.1	2.1	3.2
12/jul/2002	0.4	0.4	0.4
13/jul/2002	2.4	2.4	2.4
14/jul/2002	0.7	0.0	0.0
15/jul/2002	2.2	0.0	0.0
16/jul/2002	3.7	0.0	0.0
17/jul/2002	0.0	0.0	0.0
18/jul/2002	0.2	0.2	0.0
19/jul/2002	0.0	0.0	0.0
20/jul/2002	0.0	0.0	0.0
21/jul/2002	0.9	0.9	2.5
22/jul/2002	4.4	4.4	2.1
23/jul/2002	24.2	0.0	0.0
24/jul/2002	147.6	0.0	0.0
25/jul/2002	0.0	0.0	0.0
26/jul/2002	0.0	0.0	0.0
27/jul/2002	0.0	0.0	0.0
28/jul/2002	0.0	0.0	0.0
29/jul/2002	0.0	0.0	0.0
30/jul/2002	0.0	0.0	0.0
31/jul/2002	10.6	10.6	12.8
Total	199.7	21.3	23.4

Tabela 4.6 – Análise de falhas para o mês fevereiro de 2003. Os registros duvidosos foram grifados em cinza.

Dia	Lâmina (mm)		
	Resultado da agregação (série original de 15 minutos)	Dados diários (EMA)	Dados diários (EMC)
1/fev/2013	2.9	2.9	3.0
2/fev/2013	7.6	7.6	7.5
3/fev/2013	6.9	6.9	9.3
4/fev/2013	6.3	6.3	4.5
5/fev/2013	0.0	0.0	0.0
6/fev/2013	0.7	0.7	0.9
7/fev/2013	0.7	0.7	2.5
8/fev/2013	14.5	14.5	14.5
9/fev/2013	8.5	7.6	7.6
10/fev/2013	2.6	2.8	2.8
11/fev/2013	7.7	34.8	34.8
12/fev/2013	0.0	1.2	1.2
13/fev/2013	0.0	0.0	0.0
14/fev/2013	0.2	0.2	0.2
15/fev/2013	3.1	3.1	3.1
16/fev/2013	2.4	2.4	2.7
17/fev/2013	0.3	0.3	0.0
18/fev/2013	0.1	0.1	0.8
19/fev/2013	0.6	0.6	0.0
20/fev/2013	3.5	3.5	3.8
21/fev/2013	2.0	2.0	1.9
22/fev/2013	0.0	0.0	0.0
23/fev/2013	6.1	6.1	6.2
24/fev/2013	0.1	0.1	0.0
25/fev/2013	0.0	0.0	0.0
26/fev/2013	6.3	6.3	7.0
27/fev/2013	0.0	0.0	0.0
28/fev/2013	0.0	0.0	0.0
Total	83.1	110.7	114.3

Alguns dos meses eliminados apresentaram registros com falhas classificadas no segundo e terceiro grupos. Células vazias ou inexistentes podem justificar algumas das inconsistências entre os dados diários supracitadas, mas não todas, uma vez que vários meses apresentaram falhas somente de um desses grupos.

No Quadro 4.1 são apresentados os resultados da análise de falhas, com as lâminas totais (mm) precipitadas em cada mês e ano. Os meses grifados, de acordo com o tipo de falhas que apresentaram, foram eliminados das séries mensais e desconsiderados no cálculo dos momentos amostrais. Disso decorre que cada série mensal tem uma duração que varia entre 16 e 19 anos.

Quadro 4.1 – Resultado da análise de falhas. Os valores nesse quadro correspondem às lâminas totais em milímetros precipitadas no mês.

Ano	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez
1997	309.0	34.8	0.0	21.1	55.2	108.6	19.0	16.5	90.9	59.6	262.0	183.8
1998	108.6	362.2	122.7	70.6	93.2	24.8	15.1	22.1	87.9	155.4	25.4	251.3
1999	371.4	203.9	179.8	76.6	46.3	69.1	3.1	4.1	78.6	27.8	51.6	261.5
2000	207.1	117.1	166.5	1.1	4.8	5.2	61.0	78.4	88.9	102.7	228.5	187.5
2001	214.0	87.7	160.8	22.6	60.8	19.0	24.9	39.8	48.3	153.7	138.7	196.8
2002	275.1	167.8	259.5	28.4	104.7	0.3	199.7	63.0	42.8	48.9	166.1	169.1
2003	297.6	52.5	177.5	55.0	54.2	8.9	16.4	17.6	12.0	88.6	138.3	133.1
2004	214.3	193.7	79.0	87.7	107.9	41.3	77.4	0.4	7.1	151.2	105.5	101.4
2005	234.3	66.3	112.0	30.7	23.4	42.2	8.1	23.6	35.2	120.4	116.8	122.4
2006	225.2	164.8	121.7	17.1	0.1	0.0	19.7	9.4	60.3	74.9	195.3	56.6
2007	252.0	228.3	85.9	36.4	55.7	22.3	160.7	0.0	1.9	87.6	120.5	179.1
2008	406.7	131.5	137.7	111.4	54.7	60.6	0.3	65.5	32.3	104.2	24.3	143.6
2009	159.3	135.9	116.8	20.0	11.1	55.8	61.9	59.5	136.5	55.8	195.5	263.5
2010	230.0	140.3	96.8	76.0	24.6	16.9	59.2	0.0	80.5	86.6	38.3	225.6
2011	400.7	128.2	218.3	106.2	29.0	48.8	3.0	30.8	1.7	193.9	155.3	153.4
2012	214.9	138.7	61.5	159.2	57.8	158.0	24.7	0.0	40.9	70.3	97.9	191.4
2013	224.7	83.1	135.8	161.4	78.0	121.9	42.0	2.6	30.1	148.7	105.6	106.6
2014	77.4	51.1	114.5	51.3	34.4	1.7	24.1	7.6	87.2	17.5	121.6	255.9
2015	89.7	174.1	95.6	11.2	79.4	1.9	39.6	33.0	181.4	49.4	211.2	218.3
2016	244.8	170.2	13.3	5.7	99.8	161.9	2.6	46.5	19.8	103.1	190.8	191.2
2017	336.8	88.7	136.9	131.8	161.3	17.8	0.0	51.3	46.7	81.5		
Tamanho da série (anos)	17	18	18	16	19	19	19	19	19	18	18	18

Legenda

	Meses com falhas do tipo 2
	Meses com falhas do tipo 3
	Meses com falhas do tipo 2 e 3

4.5 Calibração

A calibração, ou ajuste, de um modelo estocástico consiste nos procedimentos realizados para a estimação dos parâmetros que o definem. O conjunto de parâmetros populacionais pode ser representado convenientemente pelo vetor $\theta = (\theta_1, \theta_1, \dots, \theta_p)$. No caso dos modelos de precipitação de Poisson, θ reflete as condições meteorológicas vigentes no local de interesse. O número de parâmetros p é uma característica específica do modelo, que varia de acordo com a sua concepção, sendo $p = 5$ no BLRP e $p = 6$ no BLRPRx.

A inferência paramétrica tem como ponto de partida uma amostra de tamanho n que também é representada por um vetor, denotado por $y = (y_1, y_2, \dots, y_n)$. Essa amostra corresponde a uma possível realização do vetor de variáveis aleatórias $Y = (Y_1, Y_2, \dots, Y_n)$. Disso decorre que a amostra y foi gerada a partir de uma distribuição cuja FDP conjunta é definida pela função $f_{Y_1, Y_2, \dots, Y_n}(y_1, y_2, \dots, y_n; \theta_1, \theta_2, \dots, \theta_p) = f_Y(y; \theta)$. Uma vez que a população não pode ser amostrada em sua totalidade, a inferência de θ a partir das informações contidas na amostra y permite apenas aproximar os valores “verdadeiros” por meio de suas estimativas, denotadas por $\hat{\theta}$.

Trazendo essas definições para o contexto em estudo, é apropriado designar os vetores das variáveis aleatórias como $Y^h = (Y_1^h, Y_2^h, \dots, Y_n^h)$ e das suas respectivas realizações como $y^h = (y_1^h, y_2^h, \dots, y_n^h)$. Reitera-se que h corresponde à escala de agregação temporal, isto é, o intervalo de tempo arbitrado para a acumulação da precipitação (15 minutos, 1 hora, 24 horas, etc.). Assim, torna-se evidente que uma série de n lâminas sucessivas y^h consiste em uma realização de Y^h . Designa-se a FDP conjunta de Y^h como $f_{Y^h}(y^h; \theta)$.

Os métodos de estimação de θ , no caso dos modelos de precipitação de Poisson, podem ser classificados genericamente em método dos momentos, métodos de verossimilhança e métodos Bayesianos (WHEATER et al., 2005).

Os métodos de verossimilhança e Bayesianos requerem a formulação de uma função de verossimilhança e, por esse motivo, sua aplicação nos processos de Poisson encontra dificuldades e restrições. Os principais trabalhos levantados que utilizam esses métodos na calibração foram brevemente revisados e serão apresentados na subseção 4.5.1.

Prevaecem na literatura os estudos em que a calibração é realizada por uma extensão do Método dos Momentos, denominada Método Generalizado dos Momentos (MGM). O MGM pode ser aplicado a qualquer modelo de Poisson, pois utiliza apenas as expressões analíticas dos momentos e, eventualmente, de alguma outra propriedade estatística (e.g. PPS),

sem a necessidade de um tratamento analítico para a formulação de uma distribuição marginal ou função de verossimilhança. Adicionalmente, o MGM permite a utilização de pesos na otimização e a escolha de propriedades estatísticas prioritárias que, a depender da capacidade estrutural do modelo, são reproduzidas com melhor qualidade nas séries sintéticas. Por esses motivos e por ser o método mais utilizado, o MGM foi escolhido para a calibração dos modelos selecionados neste trabalho e será apresentado na subseção 4.5.2.

4.5.1 Métodos baseados em verossimilhança

Na estimação de parâmetros, o método mais eficiente é o Método da Máxima Verossimilhança (MVS) (NAGHETTINI; PINTO, 2007). O MVS consiste na maximização da função de verossimilhança $L(\theta|y^h)$, que por sua vez é proporcional à FDP conjunta da amostra:

$$L(\theta|y^h) \propto f_{y^h}(y^h; \theta). \quad (4.26)$$

$L(\theta|y^h)$ pode ser interpretada como a probabilidade de que a amostra observada y^h tenha origem no modelo estatístico definido pelos valores atribuídos a θ . O conjunto de valores atribuídos a θ que maximizam a função de verossimilhança é também o conjunto que melhor explica os dados observados e, dessa forma, constitui a estimativa $\hat{\theta}$ dos parâmetros populacionais.

No âmbito dos processos de Poisson, a utilização do MVS em sua configuração formal é impraticável, sobretudo no que concerne às variáveis aleatórias do processo agregado (Y^h). Basicamente, isso se deve a dois motivos:

- complexa estrutura de dependência existente entre lâminas sucessivas ($Y_1^h, Y_2^h, \dots, Y_n^h$);
- dificuldades na formulação da distribuição marginal de Y_i^h .

Disso decorre que qualquer tentativa de calcular a FDP conjunta do processo agregado, $f_{y^h}(y^h; \theta)$, e conseqüentemente a função de verossimilhança, está essencialmente fadada ao insucesso (JESUS; CHANDLER, 2011). Não obstante, em alguns estudos foram feitas aproximações da função de verossimilhança para o ajuste dos modelos. Reitera-se que esses estudos constituem uma minoria na literatura relacionada aos modelos de precipitação de Poisson.

Chandler (1997) propôs a utilização da transformada de Fourier dos dados observados, partindo do princípio que, assintoticamente, os coeficientes de Fourier são independentes e normalmente distribuídos. Com a aplicação de técnicas estatísticas avançadas, envolvendo teoria espectral, foi aproximada uma função de verossimilhança em termos dos coeficientes de Fourier. Nesse caso, os dados modelados correspondem às intensidades médias observadas em intervalos equivalentes à resolução do pluviômetro, ao invés das lâminas y^h . Salim e Pawitan (2003) utilizaram uma abordagem de *quasi-verossimilhança* para a calibração de modelos baseados no processo de Bartlett-Lewis. Em ambos os estudos se verificou um mau desempenho na reprodução da proporção de períodos de secos (PPS).

Os métodos baseados em verossimilhança têm como vantagem um uso mais eficiente dos dados, em relação ao método dos momentos e suas variações, de modo que a calibração pode ser executada a partir de séries históricas de menor duração. Entretanto, o ajuste é baseado somente na média e nos momentos de segunda ordem, o que resultou em um mau desempenho na reprodução da proporção de períodos secos nos estudos de Chandler (1997) e Salim e Pawitan (2003). Embora não tenham sido abordadas nos estudos, é improvável que essa classe de métodos permita uma reprodução razoável de outras propriedades estatísticas, além dos momentos de primeira e segunda ordem (KACZMARSKA, 2013).

Rodriguez-Iturbe, Cox e Isham (1988) questionaram se as técnicas derivadas do MVS são de fato a melhor opção para a calibração dos modelos de Poisson. Os autores alegam que os modelos são demasiadamente simplistas: a chuva é idealizada como uma sucessão de pulsos retangulares em trajetórias contínuas no tempo. Isso implica que os métodos de verossimilhança tendem a dar uma ênfase indevida para características que são irreais.

4.5.2 Método Generalizado dos Momentos

O Método Generalizado dos Momentos (MGM) é uma extensão do Método dos Momentos (MOM) que, por sua vez, consiste no equacionamento de expressões analíticas dos momentos populacionais e dos respectivos valores amostrais. No MGM, o número de expressões analíticas utilizadas pode ser superior ao número de parâmetros desconhecidos, de modo que a estimação de θ deve ser feita através da minimização de alguma medida de discrepância, geralmente o quadrado dos desvios, entre os valores esperados do modelo e os valores observados. A teoria formal do MGM foi introduzida no artigo de Hansen (1982).

A seguir, apresenta-se a aplicação do MGM na formulação da função objetivo utilizada na calibração dos modelos BLRP e BLRPRx.

Para o modelo estocástico definido pelo vetor de parâmetros populacionais $\theta = (\theta_1, \theta_1, \dots, \theta_p)$, designa-se um vetor de propriedades estatísticas amostrais $T = (T_1, T_2, \dots, T_k)$, com $k \geq p$, computadas a partir da série observada para diferentes escalas de agregação. Essas propriedades sintetizam as características da série e, na prática, correspondem à média, variância, autocovariância, etc. Devem existir expressões analíticas que permitam o cálculo dos valores esperados das propriedades estatísticas em T . Essas expressões consistem em funções de θ e compõem o vetor de valores esperados que é designado por $\tau(\theta) = [\tau_1(\theta), \tau_2(\theta), \dots, \tau_k(\theta)]$. Com bases nessas definições, é formulada a expressão generalizada do estimador Θ segundo o MGM:

$$\Theta = \operatorname{argmin}_{\theta} \{[T - \tau(\theta)]^T \times W \times [T - \tau(\theta)]\}, \quad (4.27)$$

em que $\operatorname{argmin}_{\theta}$ representa o valor de θ que minimiza a expressão entre chaves, valor este que correspondente à estimativa $\hat{\theta}$, e W consiste em uma matriz de pesos $k \times k$ definida positiva. Nota-se que $\hat{\theta}$ deve ser obtida por otimização.

Geralmente, os pesos são atribuídos individualmente às propriedades estatísticas em T . Isso significa que a matriz W na Equação 4.27 assume uma configuração diagonal, $W = \operatorname{diag}(w_1, w_2, \dots, w_k)$, de modo que o estimador é reduzido a:

$$\Theta = \operatorname{argmin}_{\theta} \left\{ \sum_{i=1}^k w_i [T_i - \tau_i(\theta)]^2 \right\}, \quad (4.28)$$

em que w_i corresponde ao peso atribuído à i -ésima propriedade estatística em T .

Durante a primeira década de desenvolvimentos (1987–1997), os modelos eram calibrados sem a atribuição de pesos, de modo que todas as propriedades estatísticas em T tinham a mesma prioridade na otimização (RODRIGUEZ-ITURBE; COX; ISHAM, 1987, 1988; ISLAM et al., 1990; ONOF; WHEATER, 1993; BO; ISLAM; ELTAHIR, 1994; VELGHE et al., 1994; VERHOEST; TROCH; DE TROCH, 1997). Como resultado, alguns estudos relataram problemas de identificabilidade dos parâmetros θ , ou seja, a obtenção de parâmetros ótimos com magnitudes consideravelmente diferentes (ENTEKHABI; RODRIGUEZ-ITURBE; EAGLESON, 1989).

A escolha subjetiva de pesos também foi relatada em alguns estudos. Cowpertwait et al. (1996) utilizaram $w_i = 100$ para o peso associado à média e $w_i = 1$ para os demais pesos.

Wheater et al. (2005) utilizaram pesos $w_i = 100$ para as propriedades estatísticas agregadas na escala horária (média, variância e proporção de períodos secos) e pesos $w_i = 1$ para as demais propriedades em escalas de 6, 12 e 24 horas. Os autores justificaram que as propriedades horárias têm uma maior importância em aplicações hidrológicas. Recentemente, Wasko et al. (2015) executaram uma calibração com $w_i = 10$ para a média e $w_i = 1$ para outras 11 estatísticas utilizadas na otimização.

Apesar dessas atribuições subjetivas, segundo a teoria do MGM existe uma matriz de pesos ótima, em termos de minimização da variância do estimador Θ , que corresponde à inversa da matriz de covariâncias das propriedades estatísticas em T , denotada por S^{-1} . Dada a complexidade dos modelos baseados no processo de Poisson, expressões analíticas formuladas para as covariâncias esperadas em S não estão disponíveis. Por exemplo, a variância dos momentos de segunda ordem (e.g. variância e autocovariâncias de y^h) depende das propriedades teóricas de quarta ordem do processo agregado, cuja derivação analítica é proibitiva até mesmo para o mais simplista dos modelos baseados no processo Poisson (JESUS; CHANDLER, 2011).

Jesus e Chandler (2011) sugerem que uma aproximação razoável de S pode ser extraída dos dados observados por meio de um procedimento de duas etapas: uma estimativa inicial $\widehat{\theta}_1$, obtida por otimização do estimador sem a matriz ótima, é utilizada na geração de um grande número de séries sintéticas, 1000 por exemplo, a partir das quais computam-se as propriedades estatísticas amostrais e suas respectivas covariâncias, considerando cada ano como uma realização independente. Dessa forma, a matriz de covariâncias médias constitui uma estimativa da matriz teórica S . No segundo estágio, a matriz W do estimador (Equação 2) é substituída pela aproximação de S^{-1} , uma nova otimização é executada e o resultado $\widehat{\theta}_2$ corresponde a uma estimativa aprimorada dos parâmetros populacionais (JESUS; CHANDLER, 2011). Em termos computacionais, esse procedimento é intensivo e deve ser evitado.

Jesus e Chandler (2011) avaliaram a otimização de 1000 séries sintéticas com duração de 20 anos segundo diferentes formulações da matriz de pesos W . Os dados foram produzidos com um modelo de Poisson simplificado, que não considera o processo de Poisson secundário relativo à geração das células de chuva. A grande quantidade de dados permitiu computar a matriz ótima teórica S^{-1} do MGM com base na teoria assintótica. Para cada série de 20 anos, foram calculadas as 20 repetições das propriedades estatísticas em T . Em seguida, foram calculadas as variâncias e covariâncias dessas propriedades. Esse procedimento é equivalente

ao que foi descrito na subseção 4.4.2, realizado neste trabalho durante a execução do programa “pesos.cpp”. Com base no estimador da Equação 4.27, foram executadas otimizações, para cada série sintética, com as seguintes formulações da matriz de pesos:

- W_0 – inversa da matriz de covariâncias amostrais interanuais (consiste em uma aproximação de S^{-1} calculada somente com os dados de uma série sintética, o que na prática corresponderia a melhor aproximação de S^{-1} possível de se obter com base nos dados observados);
- W_1 – matriz identidade (pesos iguais em todas as propriedades);
- W_2 – diagonal com pesos de 100 para as estatísticas de 1 hora e pesos de 1 para as demais escalas de agregação (trata-se da escolha subjetiva de pesos sugerida por Wheeler et al., 2005);
- W_3 – diagonal com pesos equivalentes ao inverso das variâncias amostrais interanuais das propriedades estatísticas (trata-se de uma abordagem semelhante à matriz W_0 em que ignoram-se as covariâncias entre as diferentes propriedades estatísticas).

Os resultados desse estudo demonstraram que as matrizes W_3 e W_0 têm um desempenho similar e proporcionam estimadores com variâncias substancialmente menores do que as matrizes W_1 e W_2 . Foram também executadas otimizações com a matriz aproximada de S^{-1} , obtida segundo o procedimento de duas etapas descrito anteriormente. Os resultados foram comparados com os parâmetros estimados segundo a matriz ótima teórica. O desempenho da matriz W_3 foi semelhante ao da matriz de duas etapas, e ambas proporcionaram estimadores com resultados próximos ao do ótimo teórico. Por esses motivos, este estudo adotou a configuração W_3 para a matriz de pesos em Θ .

O trabalho de Jesus e Chandler (2011) é particularmente importante porque muitos estudos que relatam a calibração dos modelos de Poisson utilizaram os estimadores menos eficientes formulados segundo as configurações W_1 ou W_2 . Somente os estudos mais recentes, posteriores à publicação de Jesus e Chandler (2011), têm utilizado uma abordagem baseada no inverso das variâncias dos momentos estatísticos interanuais (VANHAUTE et al., 2012; PHAM et al., 2013; KACZMARSKA; ISHAM; ONOF, 2014; KACZMARSKA; ISHAM; NORTHROP, 2015; THAYAKARAN; RAMESH, 2017).

Face a essas considerações, o estimador foi reformulado para:



$$\Theta = \operatorname{argmin}_{\theta} \left\{ \sum_{i=1}^k \frac{[T_i - \tau_i(\theta)]^2}{\operatorname{Var}[T_i]} \right\}, \quad (4.29)$$

em que $\operatorname{Var}[T_i]$ é a variância amostral das repetições interanuais da i -ésima propriedade estatística T_i .

Intuitivamente, a utilização de $w_i = 1/\operatorname{Var}[T_i]$ é consistente, pois uma propriedade estatística com menor variação entre os anos é mais informativa do que uma com maior variação e, conforme o estimador Θ da Equação 4.29, recebe uma maior prioridade durante a otimização (KACZMARSKA, 2013). Além disso, as propriedades estatísticas em T que têm maior magnitude, em termos absolutos, naturalmente têm uma maior variância o que de certa forma compensa os maiores desvios entre T_i e $\tau_i(\theta)$ dessas propriedades.

Com relação às propriedades estatísticas que compõem os vetores T e $\tau(\theta)$, seguimos Cowpertwait, Isham e Onof (2007) e Kaczmarska, Isham e Onof (2014) que utilizaram:

- média de 1 hora,
- coeficiente de variação de 15 minutos, 1 hora, 6 horas e 24 horas,
- autocorrelações de *lag* 1 de 15 minutos, 1 hora, 6 horas e 24 horas,
- e coeficientes de assimetria de 15 minutos, 1 hora, 6 horas e 24 horas,

totalizando 13 propriedades estatísticas para a calibração dos modelos. Para a média, apenas uma escala foi selecionada, pois essa propriedade reflete o total precipitado ao longo da duração da série que não depende da escala de agregação temporal. A escala de 1 hora foi escolhida por conveniência.

Nota-se que a composição adotada para os vetores T e $\tau(\theta)$ considera as variâncias, autocovariâncias de *lag* 1 e momentos centrais de terceira ordem em suas formas adimensionais, representadas, respectivamente, pelos coeficientes de variação, autocorrelações de *lag* 1 e coeficientes de assimetria.

As propriedades estatísticas em T foram computadas a partir dos momentos amostrais da série de Piracicaba, calculados segundo as Equações 4.21 – 4.24, com o programa “calcular_estatisticas.cpp” (ver subseção 4.4.2). Esses momentos foram então utilizados no cálculo dos coeficientes adimensionais:

$$\text{Coeficiente de variação} = \frac{\sqrt{\operatorname{var}[y_i^h]}}{E[y_i^h]}, \quad (4.30)$$



$$\text{Autocorrelação de lag 1} = \frac{\text{cov}(y_i^h, y_{i+1}^h)}{\text{var}[y_i^h]}, \quad (4.31)$$

$$\text{Coeficiente de assimetria} = \frac{E[(y_i^h - E[y_i^h])^3]}{\text{var}[y_i^h]^{3/2}}. \quad (4.32)$$

Por sua vez, os pesos $w_i = 1/\text{var}[T_i]$ foram extraídos através do programa “pesos.cpp” (subseção 4.4.2).

Os elementos de T e os pesos w_i representam a parte amostral do estimador Θ que é comum aos modelos BLRP e BLRPRx. Os elementos de $\tau(\theta)$, por sua vez, dependem das expressões analíticas dos valores esperados dos momentos, que são específicas para cada modelo. No BLRP, $\tau(\theta)$ contém as Equações 4.11 – 4.14, e no BLRPRx, contém as Equações 4.17 – 4.20. Da mesma forma que os momentos amostrais, essas expressões, à exceção daquelas referentes à média, foram convertidas aos coeficientes adimensionais através das Equações 4.30 – 4.32.

Dessa forma, foram equacionados dois estimadores, um para cada modelo. Diante dessa formulação, cada estimador Θ pode ser interpretado como uma função objetivo, na qual o vetor de parâmetros θ representa o conjunto de variáveis de decisão. A otimização das funções objetivo permite a obtenção das estimativas $\hat{\theta}$, que melhor aproximam os parâmetros populacionais dos modelos.

4.5.3 Otimização

O algoritmo escolhido para a etapa de otimização foi o *Shuffled Complex Evolution* (SCE-UA), desenvolvido na Universidade do Arizona por Duan, Sorooshian e Gupta (1992). O SCE-UA é um algoritmo de otimização global eficiente e robusto, adequado para os problemas de objetivo único. Neste trabalho, o SCE-UA foi programado em linguagem C++11, com algumas alterações periféricas em relação ao algoritmo proposto originalmente (DUAN; SOROOSHIAN; GUPTA, 1992, 1994; DUAN; GUPTA; SOROOSHIAN, 1993).

Os códigos desenvolvidos foram disponibilizados no Apêndice B-2 e comentados a fim de elucidar o funcionamento do algoritmo.

A chamada do SCE-UA é feita pelo programa “calibracao.cpp”. Nele, é acessada a rotina “sce.h” que, por sua vez, exige dois argumentos: um inteiro de 1 a 12, indicando o mês a ser calibrado, e um objeto de saída da classe *ofstream*, indicando o caminho de um arquivo

que irá receber o relatório detalhado dos resultados de cada iteração. A configuração do SCE-UA é feita no arquivo “parametros.txt” e o espaço de soluções viáveis é definido no arquivo “Omega.txt”, ambos localizados em uma subpasta denominada “entradas”.

O programa “calibracao.cpp” deve ser executado em uma pasta contendo a função objetivo (“f_objetivo.h”) específica do modelo (foram desenvolvidas duas funções objetivo, uma para cada modelo selecionado). Essa função acessa os arquivos “M_obs.txt” e “W.txt”, localizados em “entradas”, que contêm respectivamente os momentos amostrais e os pesos de cada mês, e equaciona esses elementos com as expressões dos valores esperados. Nos códigos, essas expressões consistem em sub-funções das variáveis de decisão, isto é, dos parâmetros θ , designadas por meio da variável vetorial “x”, e da escala de agregação temporal, designada pela variável “h”. No caso do BLRP, o vetor “x” tem dimensão 5 e no BLRPRx tem dimensão 6, de acordo com o número de parâmetros de cada modelo. A cada avaliação de “x”, a função objetivo retorna um único valor, correspondente ao somatório da Equação 4.29.

Após uma série de avaliações, o SCE-UA retorna uma solução ótima, isto é, o vetor “x” que produziu o menor resultado da função objetivo e que, portanto, representa a melhor estimativa de θ . Isso é feito de forma estruturada através das etapas:

1. geração da população inicial de soluções dentro do espaço de soluções viáveis, denotado por Ω ;
2. atribuição de p complexos, cada um contendo m soluções;
3. evolução dos complexos pelo algoritmo *Competitive Complex Evolution* (CCE);
4. embaralhamento (tradução do termo *shuffling*) das soluções dos complexos evoluídos;
5. repetição das etapas 2-3 até que se atinja convergência ou um número máximo de iterações.

No algoritmo implementado, o conjunto de $p \times m$ soluções da população inicial foi gerada do espaço de soluções viáveis Ω por meio de amostragem por Hipercubo Latino. A rotina “amostragem_hipercubo_latino.h” foi desenvolvida especialmente para essa tarefa.

No SCE-UA, as soluções da população inicial e dos resultados de cada iteração são atribuídas aos complexos, por meio de uma lógica sequencial, de modo que eles contenham tanto soluções “boas”, com menor valor da função objetivo, quanto soluções “ruins”, com maior valor da função objetivo. Isso promove uma busca muito ampla no espaço de soluções viáveis, que confere ao algoritmo sua capacidade de otimização global.

O SCE-UA utiliza os princípios básicos dos algoritmos evolucionários. Cada complexo representa uma comunidade que evolui independentemente por meio do algoritmo

CCE, implementado na rotina “cce.h”. Após um certo número de gerações, produzidas dentro do CCE, as soluções aprimoradas retornam à função principal “sce.h”, são embaralhadas e novamente designadas aos complexos. Esse embaralhamento, técnica que na literatura é denominada *shuffling*, promove a mistura da informação adquirida sobre a superfície de resposta.

A execução do CCE começa com a atribuição de uma distribuição arbitrária de probabilidades às soluções do complexo que é utilizada para a seleção das soluções “pais” que farão parte da evolução. Evidentemente, as melhores soluções devem ter maiores probabilidades de seleção, mas a possibilidade de escolha das piores soluções deve ser mantida para que o algoritmo não se torne demasiadamente elitista. Dentre as m soluções do complexo, são selecionadas q soluções pais. Esse processo é repetido β^* vezes. As m soluções constituem um “simplex” que pode então ser evoluído por meio do algoritmo de Nelder e Mead (NELDER; MEAD, 1965). Cada simplex é evoluído α^* vezes. Verificou-se que a forma original do SCE-UA utiliza uma versão simplificada do algoritmo de Nelder e Mead (DUAN; SOROOSHIAN; GUPTA, 1992). Optamos por utilizar a proposta original desse algoritmo que foi implementado na rotina “nelder_mead.h”.

A evolução das q soluções ocorre por uma série de etapas que, resumidamente, envolve reflexão e expansão, em torno de um centroide, ou contração e redução do espaço abrangido pelos vértices do simplex. A cada iteração, a pior solução é substituída por uma solução mais promissora ou por uma solução que restringe o espaço de busca. Se durante as etapas de reflexão e expansão algum vértice incidir fora de Ω , uma nova solução é gerada aleatoriamente dentro do menor hipercubo H que contempla as soluções do complexo inicial. Esse procedimento representa a etapa de mutação, no contexto do SCE-UA (DUAN; SOROOSHIAN; GUPTA, 1992). Após α^* evoluções, as soluções “filhas” retornam ao complexo, substituindo somente as respectivas soluções pais. Desse modo, tanto as melhores quanto as piores soluções, que geralmente não são escolhidas para a evolução, são preservadas e retornam aos complexos para o embaralhamento e execução da próxima iteração. Evidentemente, a preservação das piores soluções atrapalha a convergência, mas também dificulta a atração para uma região de mínimos locais.

Para a evolução dos simplexes, uma alternativa ao algoritmo de Nelder e Mead, que foi abordada no estudo de Mariani et al. (2011), é o *Differential Evolution* (DE) introduzido por Storn e Price (1997). O DE é provavelmente o mais simples dentre os algoritmos evolucionários e ainda assim é eficiente. Uma rotina baseada no DE foi implementada, em substituição à rotina “nelder_mead.h”, e o SCE-UA modificado foi avaliado para um conjunto

de funções de teste, mas os resultados finais foram praticamente idênticos, de modo que o método de Nelder e Mead foi mantido na otimização.

O SCE-UA foi implementado com dois critérios de parada, que são avaliados simultaneamente ao final de cada iteração:

- contagem do número iterações $it > 100$;
- coeficiente de variação $cv < 10^{-6}$, em termos dos valores da função objetivo das soluções contidas na população da respectiva iteração.

O segundo critério é atingido quando as soluções convergem para um ótimo, seja ele local ou global. Quando não há convergência, o algoritmo para após a execução de 100 iterações.

Na Tabela 4.7 apresenta-se uma síntese das rotinas desenvolvidas durante a programação do SCE-UA e uma breve descrição de suas funções. Reitera-se que os códigos dessas rotinas e do programa “calibracao.cpp” foram disponibilizados no Apêndice B-2.

Tabela 4.7 – Rotinas programadas para implementação do algoritmo SCE-UA e descrição de suas funções.

Rotina	Função
<i>f_objetivo.h</i> (BLRP ou BLRPRx)	Acessa os momentos amostrais e pesos de cada mês, em arquivos separados, e os equaciona com as expressões analíticas dos valores esperados. Retorna o valor do estimador, calculado segundo o vetor "x" de variáveis de decisão.
<i>sce.h</i>	Governa as iterações principais, atribui as soluções aos complexos, direciona os complexos para o CCE (<i>cce.h</i>), executa o embaralhamento e checa os critérios de parada.
<i>cce.h</i>	Computa o hipercubo H , define as probabilidades de seleção, realiza seleção aleatória de pais β^* vezes, atribui as soluções “pais” aos simplexes, direciona os simplexes para <i>nelder_mead.h</i> , substitui as soluções pais pelas soluções filhas.
<i>nelder_mead.h</i>	Realiza α^* evoluções dos simplexes com o algoritmo de Nelder e Mead segundo os procedimentos de reflexão, expansão, contração, redução ou mutação (geração de uma solução aleatória em H).
<i>amostragem_hipercubo_latino.h</i>	Gera a população de soluções iniciais por meio de amostragem com Hipercubo Latino do espaço viável Ω .
<i>escolha_aleatoria.h</i>	Realiza a seleção aleatória, sem reposição, das q soluções pais, dentre as m disponíveis no complexo, que irão compor o simplex.
<i>gerar_ponto_z.h</i>	Referente ao procedimento de mutação. Gera uma solução aleatória dentro do menor hipercubo H contendo as soluções do complexo inicial.
<i>ordenar_matriz.h</i>	Ao longo da execução do SCE são produzidas diversas matrizes contendo as soluções “x” e os respectivos valores da função objetivo. Essa rotina ordena os elementos nessas matrizes, segundo o resultado da função objetivo, em ordem crescente (da melhor solução para a pior).

Ao longo dessa descrição, foram citados os parâmetros internos do SCE-UA, a saber:

- p – número de complexos;

- m – número de soluções em cada complexo;
- q – número de soluções em cada simplex;
- β^* – número de repetições da seleção de soluções “pais” para a evolução;
- α^* – número de iterações da evolução, por Nelder e Mead, de cada simplex.

Exceto o número de complexos, os parâmetros foram atribuídos conforme as recomendações de Duan, Sorooshian e Gupta (1994), em que os parâmetros dependem do número de variáveis de decisão n (dimensão do problema):

- $m = 2n + 1$: $m = 11$ no BLRP e $m = 13$ no BLRPRx;
- $q = n + 1$: $q = 6$ no BLRP e $q = 7$ no BLRPRx;
- $\beta^* = 2n + 1$: $\beta^* = 11$ no BLRP e $\beta^* = 13$ no BLRPRx;
- $\alpha^* = 1$, em ambos os modelos.

O asterisco em β^* e α^* tem a função de distinguir esses parâmetros de β e α das distribuições populacionais dos modelos BLRP e BLRPRx que, por sua vez, fazem parte do vetor de variáveis de decisão e não dos parâmetros internos do SCE-UA.

Em um primeiro momento, foram utilizados 10 complexos ($p = 10$). Essa escolha representa uma estratégia inicial para a otimização dos modelos.

A atribuição desses parâmetros deve ser feita no arquivo “parametros.txt”. A rotina “sce.h” foi desenvolvida especialmente para adquirir os parâmetros nesse arquivo, de modo que não é necessário recompilar o código caso algum parâmetro seja alterado.

Da mesma forma, a atribuição do espaço de soluções viáveis ocorre no arquivo “Omega.txt”. Nele são definidos os limites superior e inferior de cada variável de decisão. Teoricamente, os parâmetros dos modelos de Poisson podem assumir valores no intervalo $[0, +\infty)$. Entretanto, foram feitas as seguintes restrições:

- O limite superior do parâmetro λ (h^{-1}), que representa a taxa de ocorrência de eventos de chuva, foi fixado em 0,1 para ambos os modelos. Isso representa um valor esperado de aproximadamente 74 eventos de chuva em um mês, o que é muito superior ao que ocorre na realidade, mesmo nos meses mais chuvosos. Esse limite também foi utilizado por Vanhaute et al. (2012). Aproveita-se esse comentário para alertar que os parâmetros não são diretamente observáveis. Por exemplo, um evento de chuva que é observado na realidade pode, sob a ótica dos modelos de Poisson, ser o resultado da sobreposição de dois ou mais eventos.
- O limite inferior do parâmetro α do BLRPRx foi fixado em 2, com o intuito de prevenir a eventual geração de eventos de chuva com duração irreal (muito longa) na

etapa de simulação. Essa abordagem foi sugerida por Kaczmarek, Isham e Onof (2014);

- O limite superior do parâmetro ϕ do BLRPRx foi fixado em 1. ϕ é o quociente entre γ e η , ou seja, o quociente entre a duração esperada das células ($1/\eta$) e a duração esperada do evento gerador ($1/\gamma$). Obviamente γ deve ser menor do que η .

Nos demais casos foram estipulados, inicialmente, os limites inferior e superior de 0 e 20. As Tabelas 4.8 e 4.9 apresentam o espaço de soluções viáveis definido para cada modelo.

Tabela 4.8 – Espaço de soluções viáveis das variáveis de decisão (parâmetros) do modelo BLRP.

Parâmetro	λ (hr ⁻¹)	β (hr ⁻¹)	γ (hr ⁻¹)	η (hr ⁻¹)	μ_x (mm. hr ⁻¹)
Limite inferior	0	0	2	0	0
Limite superior	0,1	20	20	20	20

Tabela 4.9 – Espaço de soluções viáveis das variáveis de decisão (parâmetros) do modelo BLRPRx.

Parâmetro	λ (hr ⁻¹)	ι (mm)	α	ν	κ	ϕ
Limite inferior	0	0	2	0	0	0
Limite superior	0,1	20	20	20	20	1

Os resultados da otimização com o programa “calibracao.cpp” são impressos no arquivo “relatorio.csv”, criado na subpasta “resultados”. Para cada iteração, imprime-se a melhor solução e o respectivo valor da função objetivo. Além disso, imprime-se o máximo (f_{max}), médio (f_{med}) e mínimo (f_{min}), além do coeficiente de variação (cv), em termos dos valores da função objetivo de todas as soluções da população obtida como resultado da iteração. Os valores do vetor “x” da última iteração correspondem à estimativa $\hat{\theta}$, que é o resultado efetivo da etapa de otimização.

4.6 Simulação

Com os parâmetros calibrados, os modelos foram utilizados na simulação de séries sintéticas. Isso foi feito através das rotinas “blrp.h” e “blrpx.h”, disponibilizadas no Apêndice B-3.

Essas rotinas devem ser declaradas em um programa externo. A chamada é feita com 5 argumentos:

- vetor de parâmetros com a forma $\{\lambda, \beta, \gamma, \eta, \mu_x\}$ para o BLRP e com a forma $\{\lambda, \iota, \alpha, \nu, \kappa, \phi\}$ para o BLRPRx;

- o número de dias da série (28, 30 ou 31, a depender do mês);
- o número de anos;
- a resolução h em horas (0,25; 1,0; 2,0; ...; 24,0; etc.);
- e uma semente para geração de números aleatórios (devem ser atribuídas sementes diferentes a cada execução. Isso pode ser feito utilizando os milissegundos dos relógio).

Primeiramente, o algoritmo de simulação cria uma matriz R com duas colunas, que representa a série sintética, sendo que a primeira coluna contém os instantes dos intervalos de tempo sucessivos (limite superior do intervalo), com duração h , e a segunda contém células inicialmente vazias que irão receber as lâminas geradas no processo.

Admite-se que um evento de chuva ocorre no primeiro intervalo e os tempos de origem dos próximos eventos são calculados por meio de incrementos gerados em uma distribuição exponencial que utiliza λ como parâmetro de taxa. A biblioteca `<random>` do C++ permite a utilização de distribuições exponenciais ou Gama, no último caso para o parâmetro η do BLRPRx, na geração de números aleatórios segundo a FDP dessas distribuições.

A cada evento de chuva é iniciado um *loop* no qual serão geradas as células. No BLRPRx, para cada evento também é gerado um valor aleatório de η por meio de uma distribuição Gama com α e ν , segundo o qual são gerados os valores do demais parâmetros condicionados a η (γ , β e ι).

Em cada *loop*, a duração do evento é produzida por uma distribuição exponencial cujo parâmetro de taxa é γ . O tempo de origem da primeira célula sobrepõe a origem do evento e por meio de incrementos gerados em uma distribuição exponencial com parâmetro de taxa β são produzidos os instantes de origem das demais células. O processo dura até que esses instantes superem o instante de encerramento do evento, definido pela soma entre o instante de origem e o tempo de duração do evento.

Além do instante de origem, cada célula recebe uma duração e uma intensidade, segundo distribuições exponenciais com η e ι , respectivamente.

Uma função externa, definida na rotina “superposicao.h”, é chamada para cada célula gerada na série sintética. Ela utiliza os seguintes elementos: instante de origem da célula, intensidade da célula e um marcador. A função percorre os elementos da primeira coluna da matriz R , encontra o primeiro intervalo cujo instante superior contempla a lâmina e atribui o total precipitado de acordo com a intensidade da célula nesse e nos demais intervalos

correspondentes à duração da mesma. Essas lâminas são atribuídas na segunda coluna da matriz R.

A função “superposicao.h” tem uma importante atribuição: como os instantes de origem e de encerramento das células não coincidem com os limites dos intervalos de tempo em R, a atribuição das lâminas no primeiro e último intervalos da primeira coluna de R, que contemplam a extensão da célula, é feita de forma proporcional. Dessa forma, para uma célula com origem no instante 12,5 horas e intensidade de 1 mm/hr, por exemplo, a função de superposição atribui 0,5 mm ao intervalo compreendido entre as 12 e 13 horas, considerando a resolução horária da série sintética. Para as sucessivas células, caso ocorra sobreposição, a função “superposicao.h” executa a soma das lâminas que são atribuídas a intervalos coincidentes.

Foi mencionado que “superposicao.h” utiliza um marcador. Esse marcador indica o intervalo de partida para a atribuição de lâminas na coluna de tempos da matriz R e é incrementado sucessivamente, de modo que o processo de superposição das células é otimizado e o tempo de execução das rotinas é consideravelmente reduzido.

Os *loops* relativos à geração de cada evento se encerram quanto o tempo de origem do último evento supera a duração pré-estabelecida da série sintética.

Portanto, a matriz R consiste de uma série sintética e uma possível realização do processo estocástico preconizado nos modelos BLRP e BLRPRx. No formato final, R contém os registros de tempo e as respectivas lâminas produzidas de forma análoga à série histórica de precipitações. Essa matriz pode ser impressa em um arquivo separado ou, como foi feito neste trabalho, utilizada no cálculo dos momentos estatísticos simulados dentro de um programa desenvolvido especialmente para executar diversas simulações. Esse programa não foi disponibilizado, pois seu formato foi modificado inúmeras vezes, de acordo com a análise de resultados desejada.

Para verificar a consistência dos programas desenvolvidos foram produzidas séries sintéticas para o BLRP e o BLRPRx, de acordo com conjuntos de parâmetros pré-estabelecidos. Os valores esperados dos momentos analíticos, calculados segundo as Equações 4.11 – 4.14 (BLRP) e as Equações 4.17 – 4.20 (BLRPRx) foram comparados com os respectivos momentos extraídos das séries sintéticas, ou simplesmente, momentos simulados. Os resultados foram muito próximos, para todos os momentos considerados na comparação, demonstrando que as rotinas “blrp.h” e “blrprx.h” reproduzem as concepções dos modelos discutidas na subseção 4.2.

5 RESULTADOS E DISCUSSÃO

Nesta seção apresentam-se os principais resultados produzidos no trabalho e a discussão pertinente, seguindo as etapas definidas na seção de métodos.

Na subseção 5.1 foram disponibilizados os momentos e demais estatísticas amostrais da série histórica de Piracicaba, além dos pesos utilizados na função objetivo, resultados dos programas desenvolvidos para a aquisição de dados observados.

Na subseção 5.2 foram apresentados os parâmetros calibrados, como resultado da aplicação do algoritmo de otimização SCE-UA. Também foram feitas considerações sobre os parâmetros usados pelo algoritmo otimizador e a eficiência da calibração dos modelos BLRPR e BLRPRx.

Os resultados dos momentos analíticos, calculados segundo as Equações 4.11 – 4.14 para o BLRP e segundo as Equações 4.17 – 4.20 para o BLRPRx, com os parâmetros calibrados, foram apresentados na subseção 5.3, enquanto que os resultados das simulações, na forma dos momentos e demais propriedades estatísticas, realizadas com os modelos calibrados foram apresentados na subseção 5.4.

Na subseção 5.5, foram avaliados o desempenho e a validação dos modelos mediante a reprodução, nas séries sintéticas, dos momentos amostrais e demais propriedades estatísticas observados.

Finalmente, na subseção 5.6 foi feita uma análise de valores extremos em que foram comparados os extremos observados, de meses representativos, com os extremos produzidos pelos modelos nas séries sintéticas.

5.1 Estatísticas amostrais

Os momentos de primeira a terceira ordem e as demais propriedades estatísticas observadas que foram utilizados neste trabalho estão disponibilizados, com no mínimo três algarismos significativos, nas Tabelas 5.1 e 5.2. Esses resultados foram obtidos a partir das séries agregadas, separadas por mês, através do programa *calcular_estatisticas.cpp*.

Na

Tabela 5.3 são apresentados os momentos adimensionais nas escalas de agregação de 15 minutos, 1 hora, 6 horas e 24 horas. Esses dados compõem a matriz de momentos observados (M_{obs}), cujos valores foram efetivamente inseridos na função objetivo. Na

Tabela 5.4 são apresentados os pesos (w) e os momentos aos quais foram associados durante a otimização. Esses pesos compõem a matriz W , que é acessada pelo algoritmo de otimização para a avaliação da função objetivo. Isso também ocorre com a matriz M_{obs} . Reitera-se que os pesos correspondem ao inverso das variâncias dos momentos amostrais interanuais. Os resultados detalhados do programa de aquisição dos pesos (“*pesos.cpp*”) consistem nos valores dos 13 momentos utilizados na calibração, em diferentes escalas, calculados separadamente para os dados de cada ano. Esse procedimento foi repetido para os doze meses do ano, de modo que os resultados têm um número excessivo de dados que, por essa razão, não foram reportados.

Na Tabela 5.5 são apresentadas as lâminas máximas anuais observadas, em escala de 15 minutos (resolução original da série de Piracicaba), identificadas separadamente para cada mês. As células com traço na Tabela 5.5 correspondem aos meses que apresentaram falhas, conforme descrito na subseção 4.4.3. Esses dados foram utilizados na análise de valores extremos descrita na subseção 5.6.

Considera-se que as informações das Tabelas 5.1 – 5.5 consistem de importantes resultados deste trabalho. Primeiramente porque são o fruto de uma análise que resultou em uma série de precipitação a qual, apesar das limitações apontadas na subseção 4.4.3, aproxima-se de modo satisfatório de uma série livre de falhas. Em segundo lugar, porque a partir dessas informações é possível a aplicação direta das características meteorológicas locais, na forma dos momentos amostrais e demais propriedades estatísticas, em desenvolvimentos futuros que objetivem o aprimoramento ou validação de outros modelos estocásticos de precipitação, ou ainda a expansão deste estudo para o caso de múltiplos sítios.

Tabela 5.1 – Média e momentos centrais de ordem 2 amostrais de Piracicaba.

Mês / Escala	Média (mm)						Variância (mm ²)						Autocovariância lag 1 (mm ²)					
	15 min	1 hr	2 hr	6 hr	12 hr	24 hr	15 min	1 hr	2 hr	6 hr	12 hr	24 hr	15 min	1 hr	2 hr	6 hr	12 hr	24 hr
Jan	0,0797	0,319	0,638	1,913	3,826	7,653	0,356	3,028	7,717	32,629	76,122	155,905	0,190	0,901	2,047	5,477	7,868	32,764
Fev	0,0547	0,219	0,438	1,314	2,627	5,255	0,249	2,069	5,329	23,683	54,888	107,283	0,134	0,612	1,508	2,244	0,853	17,121
Mar	0,0459	0,184	0,367	1,101	2,202	4,404	0,246	1,971	5,566	23,502	56,735	128,749	0,133	0,718	1,312	2,974	5,993	11,525
Abr	0,0216	0,0866	0,173	0,519	1,039	2,078	0,0955	0,808	2,211	9,779	25,280	59,672	0,0506	0,300	0,623	2,132	2,989	2,074
Maio	0,0197	0,0787	0,157	0,472	0,945	1,890	0,0499	0,436	1,214	5,948	14,328	39,808	0,0229	0,159	0,453	1,626	4,524	6,159
Jun	0,0173	0,0691	0,138	0,415	0,830	1,659	0,0426	0,384	1,100	5,383	13,942	36,880	0,0215	0,165	0,435	1,705	3,875	9,838
Jul	0,0114	0,0455	0,0910	0,273	0,546	1,092	0,0194	0,168	0,498	2,629	7,119	20,455	0,00902	0,0778	0,241	1,233	2,997	5,856
Ago	0,00865	0,0346	0,0692	0,208	0,415	0,831	0,0139	0,114	0,332	1,719	4,275	10,529	0,00629	0,0479	0,131	0,482	0,965	2,899
Set	0,0201	0,0803	0,161	0,482	0,963	1,927	0,055	0,453	1,164	5,364	14,243	38,348	0,0277	0,167	0,427	1,715	4,055	6,479
Out	0,0319	0,128	0,255	0,766	1,531	3,063	0,097	0,895	2,553	11,269	24,071	60,426	0,0535	0,359	0,738	2,064	4,475	6,930
Nov	0,0455	0,182	0,364	1,091	2,182	4,364	0,198	1,667	4,649	19,191	43,580	98,881	0,103	0,552	1,005	2,031	4,897	8,959
Dez	0,0618	0,247	0,494	1,483	2,966	5,932	0,322	2,518	6,991	26,669	59,133	121,133	0,153	0,752	1,208	2,099	2,783	7,651

Tabela 5.2 – Momento central de ordem 3 e propriedades de períodos úmidos/secos amostrais de Piracicaba.

Mês / Escala	Momento central de ordem 3 (mm ³)						Proporção de períodos secos $P(Y^h < 0,1 \text{ mm})$						Probabilidade de transição úmido-úmido $P(Y_t^h \geq 0,1 \text{ mm} Y_{t-1}^h \geq 0,1 \text{ mm})$					
	15 min	1 hr	2 hr	6 hr	12 hr	24 hr	15 min	1 hr	2 hr	6 hr	12 hr	24 hr	15 min	1 hr	2 hr	6 hr	12 hr	24 hr
Jan	3,41	53,25	163,22	931,76	2574,60	4880,86	0,915	0,858	0,805	0,674	0,539	0,380	0,720	0,635	0,607	0,590	0,676	0,742
Fev	2,24	37,27	126,55	941,39	2384,60	4359,63	0,939	0,894	0,851	0,729	0,600	0,428	0,697	0,594	0,545	0,543	0,639	0,739
Mar	2,62	38,28	148,71	891,44	2662,48	7555,47	0,954	0,920	0,891	0,797	0,705	0,556	0,711	0,607	0,544	0,528	0,565	0,641
Abr	0,888	13,78	48,03	291,75	977,56	2524,54	0,975	0,955	0,934	0,882	0,826	0,735	0,685	0,580	0,553	0,533	0,533	0,543
Maio	0,317	4,27	14,37	107,84	290,65	1291,19	0,970	0,948	0,930	0,878	0,819	0,708	0,714	0,635	0,584	0,470	0,455	0,541
Jun	0,246	3,74	13,74	97,13	335,97	1134,12	0,974	0,950	0,928	0,858	0,795	0,661	0,654	0,564	0,510	0,477	0,385	0,580
Jul	0,0805	1,05	4,21	33,83	123,53	511,87	0,977	0,960	0,944	0,900	0,851	0,750	0,682	0,601	0,559	0,455	0,381	0,497
Ago	0,0693	0,716	2,67	20,57	59,12	159,66	0,982	0,968	0,956	0,925	0,889	0,834	0,662	0,642	0,586	0,500	0,511	0,531
Set	0,344	4,62	12,97	86,02	288,85	1100,29	0,970	0,949	0,932	0,887	0,839	0,744	0,707	0,662	0,624	0,543	0,489	0,552
Out	0,642	13,20	48,58	300,38	630,11	2275,35	0,956	0,927	0,901	0,832	0,760	0,640	0,717	0,652	0,611	0,551	0,541	0,540
Nov	1,75	26,58	98,08	520,55	1251,15	3026,83	0,951	0,917	0,883	0,798	0,708	0,554	0,718	0,611	0,576	0,519	0,527	0,598
Dez	3,89	51,86	197,82	809,17	1799,75	3348,75	0,938	0,894	0,856	0,740	0,622	0,443	0,715	0,618	0,555	0,536	0,594	0,661

Tabela 5.3 – Momentos adimensionais utilizados na calibração.

Mês / Escala	Média (mm) 1 hr	Coeficiente de variação				Autocorrelação de lag 1				Coeficiente de assimetria			
		15 min	1 hr	6 hr	24 hr	15 min	1 hr	6 hr	24 hr	15 min	1 hr	6 hr	24 hr
Jan	0,319	7,486	5,457	2,986	1,632	0,534	0,298	0,168	0,210	16,037	10,107	4,999	2,507
Fev	0,219	9,111	6,570	3,704	1,971	0,541	0,296	0,0948	0,160	18,086	12,523	8,168	3,923
Mar	0,184	10,809	7,651	4,403	2,576	0,539	0,364	0,127	0,0895	21,491	13,830	7,824	5,172
Abr	0,0866	14,280	10,385	6,020	3,718	0,530	0,371	0,218	0,0348	30,074	18,959	9,541	5,477
Maio	0,0787	11,342	8,389	5,162	3,338	0,459	0,365	0,273	0,155	28,444	14,804	7,433	5,141
Jun	0,0691	11,949	8,969	5,594	3,660	0,504	0,429	0,317	0,267	27,938	15,686	7,777	5,064
Jul	0,0455	12,247	9,021	5,940	4,142	0,465	0,462	0,469	0,286	29,793	15,241	7,935	5,533
Ago	0,0346	13,613	9,744	6,314	3,907	0,453	0,422	0,281	0,275	42,450	18,673	9,128	4,673
Set	0,0803	11,670	8,386	4,808	3,214	0,504	0,368	0,320	0,169	26,741	15,146	6,924	4,633
Out	0,128	9,759	7,416	4,384	2,538	0,552	0,401	0,183	0,115	21,263	15,583	7,941	4,844
Nov	0,182	9,801	7,101	4,016	2,279	0,518	0,331	0,106	0,0906	19,842	12,354	6,192	3,078
Dez	0,247	9,178	6,420	3,482	1,855	0,477	0,299	0,079	0,0632	21,335	12,981	5,875	2,512

Tabela 5.4 – Pesos w , inseridos na função objetivo, associados aos momentos apropriados.

Mês / Escala	Média (mm) 1 hr	Coeficiente de variação				Autocorrelação de lag 1				Coeficiente de assimetria			
		15 min	1 hr	6 hr	24 hr	15 min	1 hr	6 hr	24 hr	15 min	1 hr	6 hr	24 hr
Jan	0,319	7,486	5,457	2,986	1,632	0,534	0,298	0,168	0,210	16,037	10,107	4,999	2,507
Fev	0,219	9,111	6,570	3,704	1,971	0,541	0,296	0,0948	0,160	18,086	12,523	8,168	3,923
Mar	0,184	10,809	7,651	4,403	2,576	0,539	0,364	0,127	0,0895	21,491	13,830	7,824	5,172
Abr	0,0866	14,280	10,385	6,020	3,718	0,530	0,371	0,218	0,0348	30,074	18,959	9,541	5,477
Mai	0,0787	11,342	8,389	5,162	3,338	0,459	0,365	0,273	0,155	28,444	14,804	7,433	5,141
Jun	0,0691	11,949	8,969	5,594	3,660	0,504	0,429	0,317	0,267	27,938	15,686	7,777	5,064
Jul	0,0455	12,247	9,021	5,940	4,142	0,465	0,462	0,469	0,286	29,793	15,241	7,935	5,533
Ago	0,0346	13,613	9,744	6,314	3,907	0,453	0,422	0,281	0,275	42,450	18,673	9,128	4,673
Set	0,0803	11,670	8,386	4,808	3,214	0,504	0,368	0,320	0,169	26,741	15,146	6,924	4,633
Out	0,128	9,759	7,416	4,384	2,538	0,552	0,401	0,183	0,115	21,263	15,583	7,941	4,844
Nov	0,182	9,801	7,101	4,016	2,279	0,518	0,331	0,106	0,0906	19,842	12,354	6,192	3,078
Dez	0,247	9,178	6,420	3,482	1,855	0,477	0,299	0,079	0,0632	21,335	12,981	5,875	2,512

Tabela 5.5 – Lâminas máximas anuais, observadas, na escala de agregação de 15 minutos.

Ano	Lâmina máxima (mm)											
	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez
1997	11,9	-	-	-	5,7	5,0	2,1	5,4	8,9	3,7	15,2	21,1
1998	27,8	18,7	20,6	12,5	4,9	4,5	1,7	1,9	15,9	8,9	4,9	13,5
1999	-	12,5	20,1	11,6	5,8	4,4	1,6	-	5,1	3,3	4,2	13,7
2000	-	4,9	13,2	0,4	0,6	0,9	4,5	-	6,3	11,8	8,4	11,7
2001	12,8	5,4	21,3	2,0	11,1	5,1	4,6	13,4	7,6	15,2	13,6	39,5
2002	22,3	13,5	16,9	-	13,6	0,1	-	3,4	6,8	4,1	14,4	19,4
2003	20,6	6,2	17,5	16,2	9,4	2,7	2,9	0,8	1,5	7,3	9,3	9,1
2004	8,2	17,6	9,2	14,3	10,0	-	4,4	0,1	2,6	10,4	12,5	7,8
2005	9,4	13,3	5,9	5,9	-	5,9	2,5	3,1	3,5	-	12,3	7,6
2006	10,4	13,4	-	-	0,1	-	-	3,0	4,0	4,8	11,1	-
2007	-	16,7	11,4	6,7	4,1	3,0	7,0	0,0	0,7	5,9	10,1	16,5
2008	15,2	14,7	16,6	-	7,7	6,8	0,1	7,3	-	9,3	5,8	10,7
2009	8,8	9,3	12,7	8,9	2,1	4,3	2,3	2,0	10,6	6,3	25,2	13,7
2010	-	10,8	16,9	6,4	2,5	4,1	6,4	0,0	-	10,2	6,9	-
2011	15,2	-	10,3	-	7,0	6,8	0,9	3,3	0,4	12,0	8,4	19,7
2012	8,5	14,8	10,6	13,5	9,0	7,0	2,8	0,0	10,3	6,2	11,5	9,7
2013	11,2	-	13,1	16,7	2,9	11,3	10,9	0,6	2,1	15,5	16,9	15,0
2014	13,8	6,2	17,3	3,0	2,8	0,5	1,4	0,8	5,4	2,1	-	8,8
2015	7,1	9,9	9,7	3,2	9,1	0,4	4,2	3,3	10,7	-	-	17,9
2016	13,1	10,5	-	0,5	-	16,0	0,9	2,8	1,8	-	10,9	12,2
2017	16,0	16,3	19,8	14,7	14,2	1,8	0,0	3,0	6,4	7,6	-	-

Os traços correspondem aos meses com falhas.

5.2 Calibração

Os resultados da etapa de calibração correspondem às soluções ótimas de cada modelo que permitiram a minimização da função objetivo, segundo os critérios de convergência estabelecidos para a execução do algoritmo SCE-UA. Em linguagem estatística, essas soluções consistem em vetores de valores que, por sua vez, correspondem às estimativas) dos parâmetros populacionais. Na análise de resultados a formalidade estatística foi dispensada e o “parâmetros calibrados” foi utilizado como referência do vetor de estimativas $\hat{\theta}$.

O programa *calibracao.cpp* foi utilizado na otimização dos modelos BLRP e BLRPRx. A mesma função objetivo foi utilizada em ambos os modelos e as diferenças nas rotinas programas limitam-se basicamente ao número de parâmetros de cada caso: 5 no BLRP e 6 no BLRPRx. Os resultados dessa etapa, na forma dos parâmetros calibrados, foram produzidos para cada mês do calendário de modo a conciliar a sazonalidade do fenômeno real com a estacionariedade, que é um pré-requisito dos modelos de Bartlett-Lewis.

5.2.1 BLRP

A calibração do BLRP seguindo a estratégia inicial, em que o SCE-UA foi configurado para rodar com apenas 10 complexos, resultou nos parâmetros apresentados na Tabela 5.6.

Tabela 5.6 – Parâmetros calibrados do BLRP. Resultados preliminares do SCE-UA executado com apenas 10 complexos.

Mês / Parâmetro	$\hat{\lambda}$ (hr^{-1})	$\hat{\beta}$ (hr^{-1})	$\hat{\gamma}$ (hr^{-1})	$\hat{\eta}$ (hr^{-1})	$\hat{\mu}_x$ ($mm. hr^{-1}$)	f_{min}
Jan	0,02426	0,24701	0,06676	4,19723	11,73770	1,13
Fev	0,04778	1,83242	20	2,76783	11,62010	3,59
Mar	0,02840	9,17519	20	2,43725	10,79840	6,24
Abr	0,00666	1,93087	0,34760	6,52913	12,93900	0,46
Mai	0,00667	2,07608	0,21836	8,94853	10,05200	0,64
Jun	0,00485	1,49134	0,13109	7,63474	8,79078	1,46
Julho	0,00302	2,11523	0,09433	9,50085	6,10239	0,80
Ago	0,00587	2,06132	0,25954	15,42170	10,16760	3,90
Set	0,00688	1,31897	0,15326	6,35957	7,72884	0,86
Out	0,03127	6,11803	20	2,52536	7,89027	9,20
Nov	0,02010	3,02181	0,65666	7,27778	11,75440	0,75
Dez	0,03015	2,49321	0,81954	7,70044	15,61680	1,40

Para os meses de fevereiro, março e outubro foram obtidos resultados duvidosos. Nesses casos, as estimativas de γ equivalem ao limite superior fixado para esse parâmetro no espaço de soluções viáveis do algoritmo de otimização. Nos demais meses foram obtidos resultados consistentes, tanto no que se refere às estimativas dos parâmetros, quanto aos valores da função objetivo (f_{min}). Nota-se que nos meses não duvidosos, excetuando-se agosto, f_{min} resultou em valores menores que 2,00, o que pode ser considerado, em uma análise preliminar, como um bom ajuste.

Foi feita uma investigação aprofundada a respeito da calibração do BLRP, objetivando esclarecer, entre outros, os fatores que ocasionaram as soluções duvidosas dos meses supracitados. Os resultados dessa investigação foram utilizados na formulação de considerações discutidas na subseção 5.2.3.

Em uma primeira tentativa, o limite superior do espaço de soluções viáveis de todos os parâmetros foi ampliado para 100, e uma nova calibração foi executada para os meses duvidosos. Os resultados obtidos estão apresentados na Tabela 5.7.

Tabela 5.7 – Resultados da calibração dos meses duvidosos com ampliação do espaço de soluções viáveis.

Mês / Parâmetro	$\hat{\lambda}$ (hr^{-1})	$\hat{\beta}$ (hr^{-1})	$\hat{\gamma}$ (hr^{-1})	$\hat{\eta}$ (hr^{-1})	$\hat{\mu}_x$ ($mm. hr^{-1}$)	f_{min}
Fev	0,04620	13,90880	100	2,75658	11,46990	3,56
Mar	0,02743	53,34190	100	2,39726	10,45920	6,12
Out	0,03005	37,82630	100	2,47643	7,62937	9,10

Independentemente da amplitude do espaço de soluções viáveis, os resultados persistiram com γ no limite superior do parâmetro. O parâmetro β também foi afetado, atingindo valores muito elevados em comparação com os resultados da calibração anterior e às estimativas de β dos demais meses. Embora sejam consideráveis as diferenças de magnitude de $\hat{\gamma}$ e $\hat{\beta}$ entre as respectivas soluções das Tabelas 5.6 e 5.7, a redução de f_{min} foi praticamente desprezível em relação à primeira tentativa de calibração ($< 2\%$).

Uma segunda tentativa foi realizada mantendo-se o espaço de soluções viáveis estabelecido inicialmente e incrementando progressivamente o número de complexos do SCE-UA para 20, 40, 60, 80 e 100. A calibração foi executada para todos os meses, incluindo aqueles não duvidosos. Eficácia em todos os meses, no que se refere à convergência para soluções ótimas dentro do espaço viável, só foi obtida na tentativa com 100 complexos. Nesse caso, o procedimento resultou em soluções consistentes, obtendo-se definitivamente os parâmetros calibrados do BLRP, apresentados na Tabela 5.8.

Tabela 5.8 – Parâmetros calibrados do BLR. Foram utilizados 100 complexos no SCE-UA.

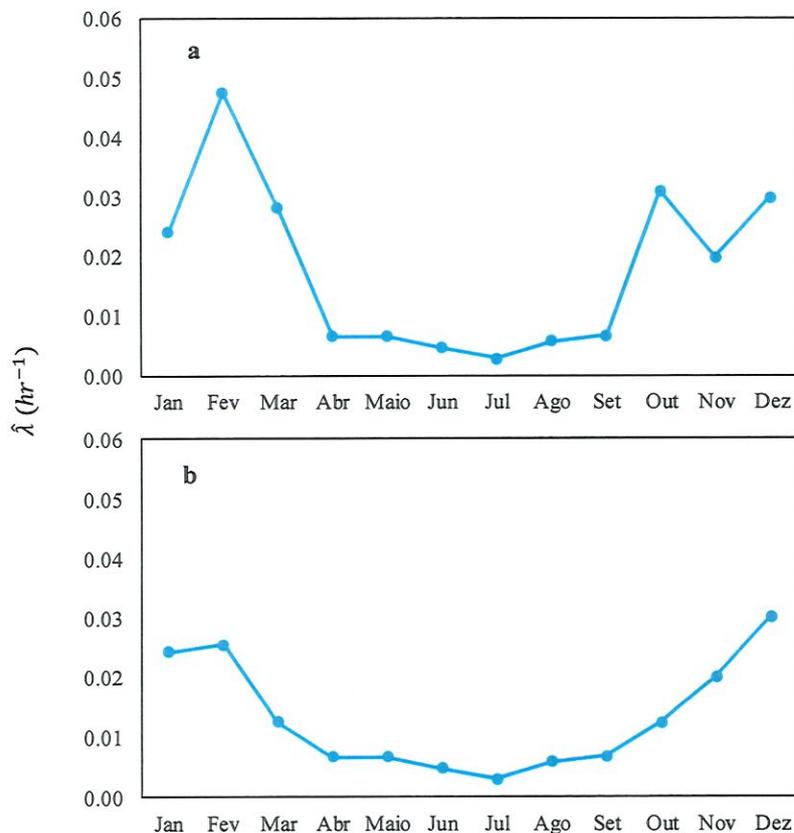
Mês / Parâmetro	$\hat{\lambda}$ (hr^{-1})	$\hat{\beta}$ (hr^{-1})	$\hat{\gamma}$ (hr^{-1})	$\hat{\eta}$ (hr^{-1})	$\hat{\mu}_x$ ($mm. hr^{-1}$)	f_{min}
Jan	0,02426	0,24701	0,06676	4,19723	11,73770	1,13
Fev	0,02549	0,65655	0,40445	3,92223	12,84280	2,39
Mar	0,01255	1,96490	0,41656	5,50608	14,08490	2,47
Abr	0,00666	1,93087	0,34760	6,52913	12,93900	0,46
Mai	0,00667	2,07608	0,21836	8,94853	10,05210	0,64
Jun	0,00485	1,49135	0,13109	7,63474	8,79076	1,46
Julho	0,00302	2,11523	0,09433	9,50086	6,10237	0,80
Ago	0,00587	2,06135	0,25955	15,42170	10,16760	3,90
Set	0,00688	1,31896	0,15326	6,35957	7,72885	0,86
Out	0,01230	0,99295	0,25033	4,48665	9,37319	3,95
Nov	0,02010	3,02183	0,65666	7,27779	11,75440	0,75
Dez	0,03015	2,49321	0,81954	7,70044	15,61680	1,40

Observou-se que, apesar do maior número de complexos, somente as soluções ótimas de fevereiro, março e outubro foram alteradas na segunda tentativa de calibração. Isso demonstra que a calibração dos demais meses além de eficaz e também mais eficiente, uma

vez que o mínimo global pôde ser obtido com um número menor de complexos, o que, geralmente, representa também um menor número de avaliações da função objetivo e processamento computacional.

Além da não ocorrência do limite superior como valor para o parâmetro $\hat{\gamma}$, a consistência dos resultados definitivos pode ser verificada por comparação das estimativas de um determinado parâmetro obtidas entre os diferentes meses do calendário. No caso de λ , que representa a taxa de ocorrência de eventos, foi verificado que as inconsistências para março e outubro manifestaram-se como picos incoerentes, em relação aos meses adjacentes, em um gráfico onde foram plotados $\hat{\lambda}$, da primeira tentativa de calibração, e os respectivos meses (Figura 5.1 – a). Nota-se que esse gráfico traduz a tendência sazonal do fenômeno em estudo. No gráfico da Figura 5.1 – b, em que foram plotados os parâmetros calibrados mais consistentes, definitivos, essas incoerências desapareceram e a transição entre as estimativas de λ ocorreu de forma suave entre os meses, demonstrando um resultado mais próximo da realidade.

Figura 5.1 – Parâmetro λ calibrado com o SCE-UA configurado para executar com 10 complexos (a) e 100 complexos (b).



O parâmetro λ foi escolhido, dentre os cinco parâmetros do modelo BLRP, para avaliar a consistência dos resultados por se tratar daquele que representa de forma mais direta as características sazonais e que pode ser avaliado individualmente. A tendência sazonal, ilustrada na Figura 5.1 – b, sugere que aos meses de dezembro, janeiro e março foram atribuídos os maiores valores de $\hat{\lambda}$, ao passo que os menores foram atribuídos aos meses de junho, julho e agosto. Isso está de acordo com a realidade do fenômeno em estudo, pois Piracicaba tem nos meses de dezembro, janeiro e março as maiores precipitações totais, de modo que o esperado é que esses meses tenham também as maiores taxas de ocorrência de eventos. O oposto ocorre nos meses de junho, julho e agosto e condiz com as menores taxas resultantes da calibração. Cuidados devem ser tomados para analisar dessa mesma forma outros parâmetros como, por exemplo, β (hr^{-1}) que representa a taxa de ocorrência de células de chuva. Esse parâmetro parece não atender à mesma lógica sazonal de λ , pois tanto o mínimo quanto o máximo valor de $\hat{\beta}$ ocorrem em meses chuvosos (janeiro e novembro, respectivamente). Uma análise de consistência de β deve ser feita em conjunto com outros parâmetros.

Embora não tenha sido reportado neste trabalho, verificou-se que removendo da calibração os coeficientes de assimetria nas escalas de 1, 6 e 24 horas, e preservando os coeficientes de 15 minutos, o SCE-UA executado com apenas 10 complexos converge para soluções consistentes e próximas dos resultados da Tabela 5.8. Isso reflete a deficiência conhecida da calibração dos modelos de Poisson com o Método Generalizado dos Momentos: a qualidade e o sucesso do ajuste estão condicionados à escolha (subjetiva) das propriedades estatísticas e escalas de agregação utilizadas na função objetivo.

5.2.2 BLRPRx

A calibração do BLRPRx foi mais eficiente do que a do BLRP. Embora se tenha notado um número de avaliações da função objetivo levemente superior, por meio do contador inserido nas rotinas *sce.h* e *f_objetivo.h*, com apenas 10 complexos configurados no SCE-UA foram obtidas estimativas consistentes dos para os parâmetros do BLRPRx para todos os meses. Além disso, o maior f_{min} resultante foi de aproximadamente 3,00 (Tabela 5.9). Para comprovar a inexistência de mínimos locais nessas soluções, foi executada uma calibração com 100 complexos, que produziu como resultado as mesmas soluções ótimas.

Tabela 5.9 – Parâmetros calibrados do BLRPRx. Foram utilizados 100 complexos no SCE-UA.

Mês / Parâmetro	$\hat{\lambda}$ (hr^{-1})	\hat{i} (mm)	$\hat{\alpha}$	$\hat{\nu}$	$\hat{\kappa}$	$\hat{\phi}$	f_{min}
Jan	0,02333	2,84237	2,27849	0,60097	0,05755	0,01511	0.58
Fev	0,01529	4,01810	3,01793	1,05574	0,01921	0,00749	1.41
Mar	0,01235	2,73477	2,49634	0,51663	0,36209	0,08169	1.87
Abr	0,00644	1,77888	2,19466	0,33334	0,39445	0,06019	0.15
Mai	0,00628	1,07776	2	0,22614	0,22930	0,02154	0.38
Jun	0,00489	1,09148	2	0,29569	0,22620	0,01891	0.79
Julho	0,00315	0,61601	3,19239	0,34666	0,22925	0,01020	0.62
Ago	0,00553	0,71247	2	0,19344	0,15439	0,01985	2.99
Set	0,00705	1,17166	2	0,33195	0,21628	0,02482	0.46
Out	0,01148	2,27158	2	0,58457	0,23276	0,05981	0.99
Nov	0,01890	1,19824	2	0,20899	0,73391	0,10440	0.40
Dez	0,02946	1,99133	2	0,26409	0,45043	0,14019	0.79

Para sete meses $\hat{\alpha}$ resultou no limite inferior do respectivo parâmetro, no espaço de soluções viáveis. Nesse caso, não se trata de uma inconsistência, mas de uma restrição atribuída ao parâmetro α para que os eventos e células gerados pelas séries sintéticas não tenham tempos de duração irrealistas (ver subseção 4.2). Resultados semelhantes foram obtidos por Kaczmarzka, Isham e Onof (2014) para a série histórica de Bochum.

Os valores de f_{min} do BLRPRx foram menores, para todos os meses, em comparação aos do BLRPR Tabela 5.8. Considerando somente os momentos utilizados na calibração, isso representa um melhor ajuste. Entre os resultados do BLRP e do BLRPRx, houve uma redução variando entre 23% de f_{min} , nos meses de julho e agosto, e 75% no mês de outubro. Esse último resultado particularmente indica um melhor desempenho devido à flexibilidade proporcionada pelo BLRPRx, uma vez que outubro foi um dos meses de difícil calibração reportados no caso do BLRP.

5.2.3 Considerações sobre a calibração do BLRP/BLRPRx

A calibração dos modelos BLRP e BLRPRx não é uma tarefa trivial. O sucesso da otimização na convergência para soluções consistentes está condicionado à uma série de fatores. Neste trabalho, o sucesso da calibração foi atribuído principalmente a três fatores:

1. escolha adequada das propriedades estatísticas que compõe a função objetivo (momentos de primeira a terceira ordem em escalas de agregação de 15 minutos, 1 hora, 6 horas e 24 horas);

2. utilização de pesos não subjetivos na otimização, calculados segundo a variância interanual das respectivas propriedades estatísticas;
3. e o emprego de um algoritmo de otimização global robusto.

Essa afirmação decorre de várias tentativas frustradas de calibração dos modelos de Bartlett-Lewis, em que esses três fatores não foram verificados simultaneamente e os resultados foram parâmetros instáveis; não convergência do SCE-UA (dentro do máximo estabelecido de 100 iterações); convergência de um ou mais parâmetros para a fronteira do espaço de soluções viáveis; ou soluções irrealistas.

A experiência adquirida na calibração e os resultados apresentados até o momento permitem elencar as seguintes considerações sobre a otimização dos modelos de Bartlett-Lewis:

1. os meses apresentam superfícies de resposta com diferentes comportamentos e, portanto, distintos graus de dificuldade de minimização;
2. a convergência de um parâmetro para seu respectivo limite pode indicar a existência de um mínimo local na fronteira do espaço de soluções viáveis;
3. nesses casos, é provável a existência de um mínimo consistente, dentro do espaço de soluções viáveis;
4. não temos informações suficientes para afirmar se esse mínimo consistente será sempre um mínimo global e isso deve ser considerado nas tentativas de otimização;
5. o mínimo local na fronteira parece refletir uma incapacidade estrutural do modelo em reproduzir simultaneamente as diversas estatísticas e escalas de agregação utilizadas;
6. o mínimo consistente pode ser atingido com o aprimoramento da busca (através da configuração do algoritmo de otimização); se ainda assim a solução ótima persistir na fronteira, pode ser necessário remover da função objetivo alguma propriedade estatística de determinada escala de agregação (e.g. coeficiente de assimetria em escalas superiores a 15 minutos);

A consideração 5 decorre dos resultados obtidos inicialmente para os meses de fevereiro, março e outubro. Os parâmetros γ e β estão relacionados, respectivamente, à duração dos eventos e à taxa de ocorrência de células de chuva. Quanto maiores forem as estimativas desses parâmetros, mais detalhada serão as trajetórias do processo estocástico, pois a série sintética terá eventos de menor duração e contendo mais células. O modelo BLRP preconiza uma mesma distribuição para as intensidades de todas as células, com valor

esperado μ_X . Dessa forma, na tentativa de reproduzir simultaneamente todas as estatísticas inseridas na calibração, sobretudo aquelas que refletem os períodos com maior intensidade de chuva, que no presente estudo são os momentos de terceira ordem, o modelo compensa suas limitações estruturais refinando os pulsos retangulares da série e forçando não só $\hat{\gamma}$ e $\hat{\beta}$ para o limite superior, mas também $\hat{\lambda}$. Isso foi evidenciado nos meses duvidosos, para os quais os valores de $\hat{\lambda}$ são maiores nos resultados preliminares (Figura 5.1-a) do que nos resultados definitivos (Figura 5.1-b). Nota-se ainda que o valor esperado das intensidades, μ_X manteve-se relativamente estável e, portanto, o incremento ocorreu para os outros parâmetros, seguindo esse mecanismo de compensação.

Face ao exposto, justifica-se o porquê da utilização do momento de terceira ordem ter dificultado a calibração do BLRP em determinados meses: a manutenção de uma FDP comum às intensidades de todas as células dificulta a reprodução de lâminas com distintas magnitudes. Essas constatações apontam que a calibração pode ser facilitada com o modelo BLRPRx, devido à sua maior flexibilidade em relação à atribuição de distribuições com diferentes parâmetros para as intensidades e durações das células.

5.3 Valores esperados dos momentos estatísticos

Nesta subseção, foram apresentados os resultados dos valores esperados dos momentos de até terceira ordem, ou seja, os momentos estatísticos calculados analiticamente para o BLRP e BLRPRx através das Equações 4.11 – 4.14 e Equações 4.17 – 4.20, respectivamente, nas quais foram utilizados os parâmetros calibrados (Tabelas 5.8 e 5.9). Os valores esperados foram referidos no texto como momentos analíticos, e são distintos momentos simulados que, por sua vez, são computados a partir das séries sintéticas produzidas como resultado das simulações.

As Figuras 5.2 – 5.5 apresentam os momentos que foram inseridos na função objetivo: coeficiente de variação, autocorrelação de *lag* 1 e coeficiente de assimetria nas escalas de agregação de 15 minutos, 1 hora, 6 horas e 24 horas, e a média das lâminas na escala de 1 hora. Os momentos analíticos de cada mês foram plotados juntamente com os momentos observados correspondentes da série histórica de Piracicaba. Os gráficos nas Figuras 5.2 – 5.5, e demais figuras similares, foram agrupados por propriedade estatística e confeccionados para as diferentes escalas e, portanto, devem ser analisados individualmente.

Figura 5.2 – Média de 1 hora por mês, momentos analíticos e observados.

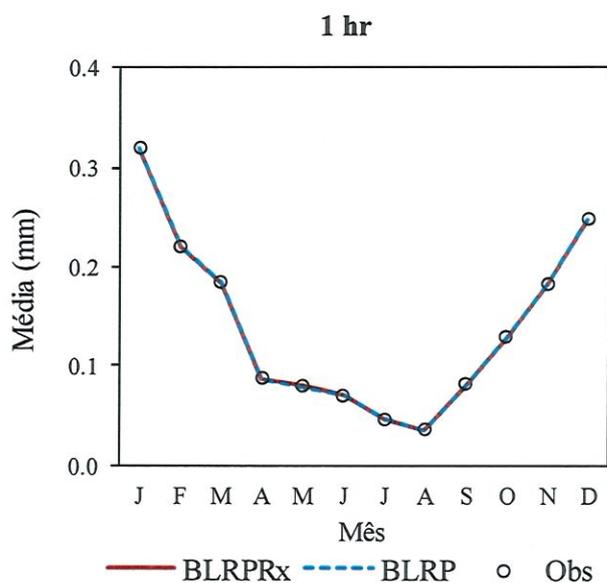


Figura 5.3 -- Coeficiente de variação por mês, momentos analíticos e observados.

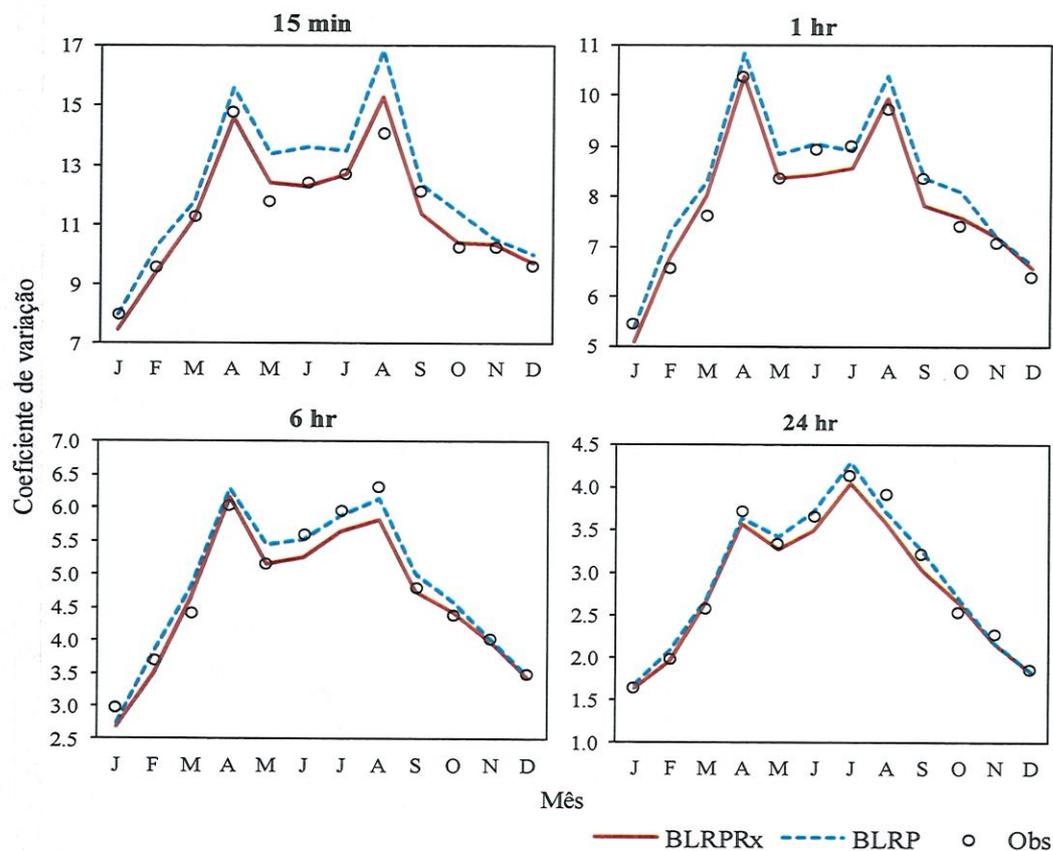


Figura 5.4 – Autocorrelação lag 1 por mês, momentos analíticos e observados.

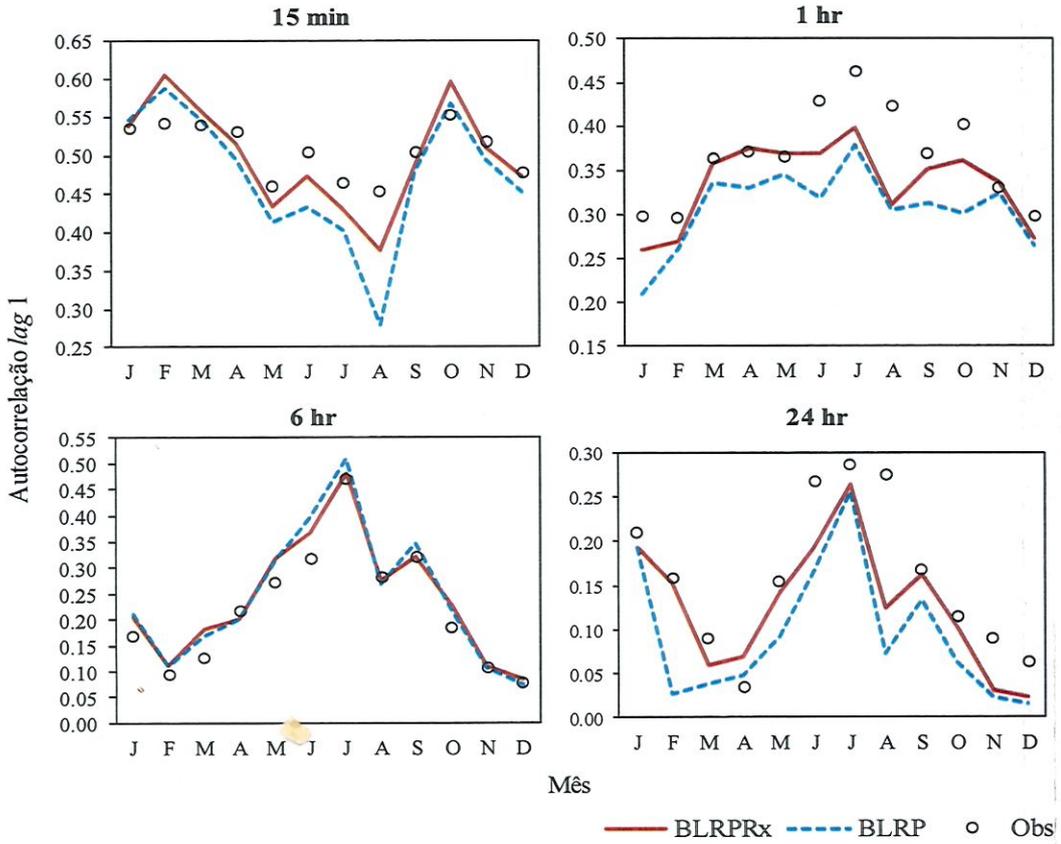
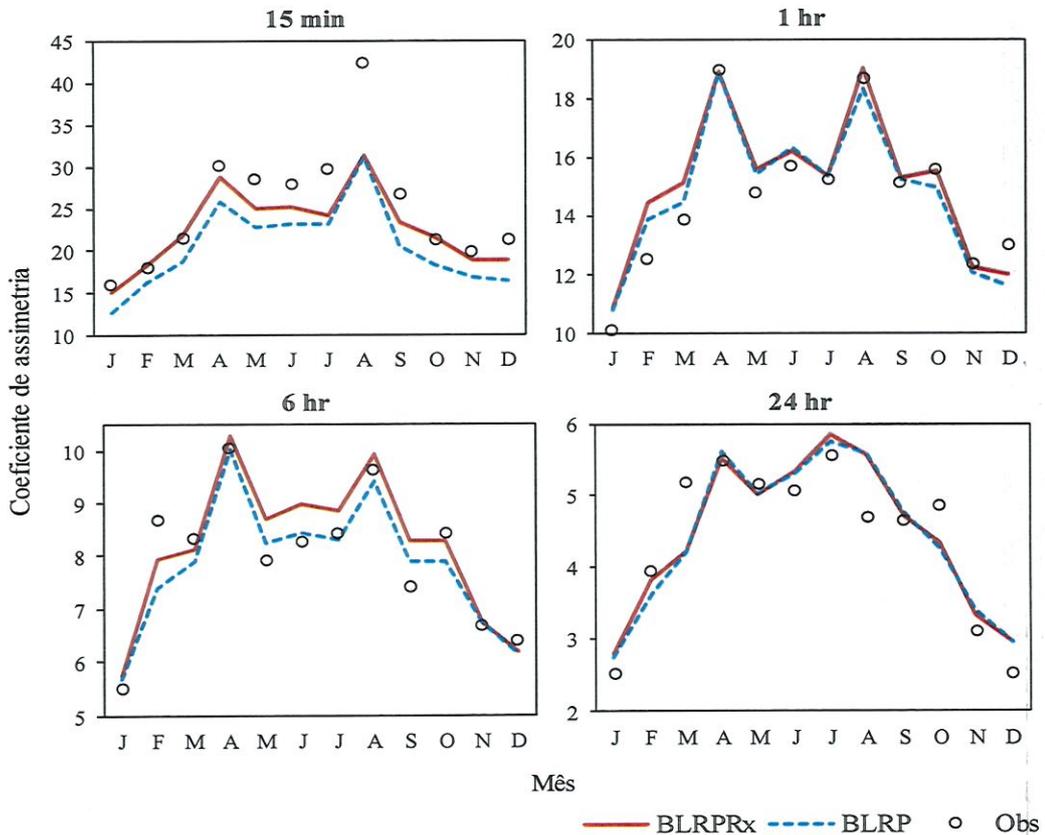


Figura 5.5 – Coeficientes de assimetria por mês, momentos analíticos e observados.



A Figura 5.2 possibilita a conclusão de que foi obtido um ajuste exato à média de 1 hora para todos os meses em ambos os modelos. Isso ocorreu, em parte, porque o peso na função objetivo associado à média é elevado em comparação aos pesos dos demais momentos. Contudo, esse não é o único fator responsável por esses resultados. Em alguns meses, os pesos associados às autocorrelações de *lag* 1, sobretudo nas escalas de 15 minutos e 1 hora, superam o peso da média sem que, no entanto, os modelos reproduzam esses momentos com a mesma qualidade (Figura 5.4).

Os resultados da Figura 5.3 possibilitam a constatação do ótimo ajuste obtido para a variância analítica, na forma adimensional do coeficiente de variação. Os desvios entre os coeficientes de variação analíticos e observados são da ordem de 5%, tanto no BLRP quanto no BLRPRx. Analisando somente a Figura 5.3 não é possível afirmar de forma categórica qual modelo apresentou um melhor desempenho. Nos meses de junho, julho e agosto, por exemplo, o BLRPRx reproduziu melhor os coeficientes de variação de 15 minutos do que o BLRP. No entanto, o inverso foi verificado na escala de 6 hr, em que o BLRP apresentou um melhor desempenho. Aparentemente há uma tendência, que não é global, segundo a qual o BLRPRx atenua aos desvios entre os momentos analíticos e observados quando estes são superiores a 10% no BLRPR. Isso pode ser evidenciado para os coeficientes de variação de 15 minutos de maio, agosto e outubro e nos coeficientes de 1 hora de fevereiro e outubro.

Em relação às autocorrelações de *lag* 1 apresentadas na Figura 5.4, o BLRPRx mostrou um ajuste superficialmente melhor, evidenciado principalmente no gráfico da escala de 1 hora. Na escala de 15 minutos, obteve-se um melhor ajuste com o BLRPRx para a maioria dos meses, com agosto e março mostrando uma atenuação semelhante à mencionada no parágrafo anterior. Ainda assim, as discrepâncias entre as autocorrelações analíticas e observadas são notavelmente baixas para ambos os modelos (da ordem de 0,04 no BLRPRx e 0,06 no BLRP). Assim como foi observado no estudo de Kaczmarska, Isham e Onof (2014) há uma tendência de ambos os modelos em subestimar a autocorrelação de *lag* 1 na escala de 24 horas, embora isso não tenha ocorrido no mês de abril.

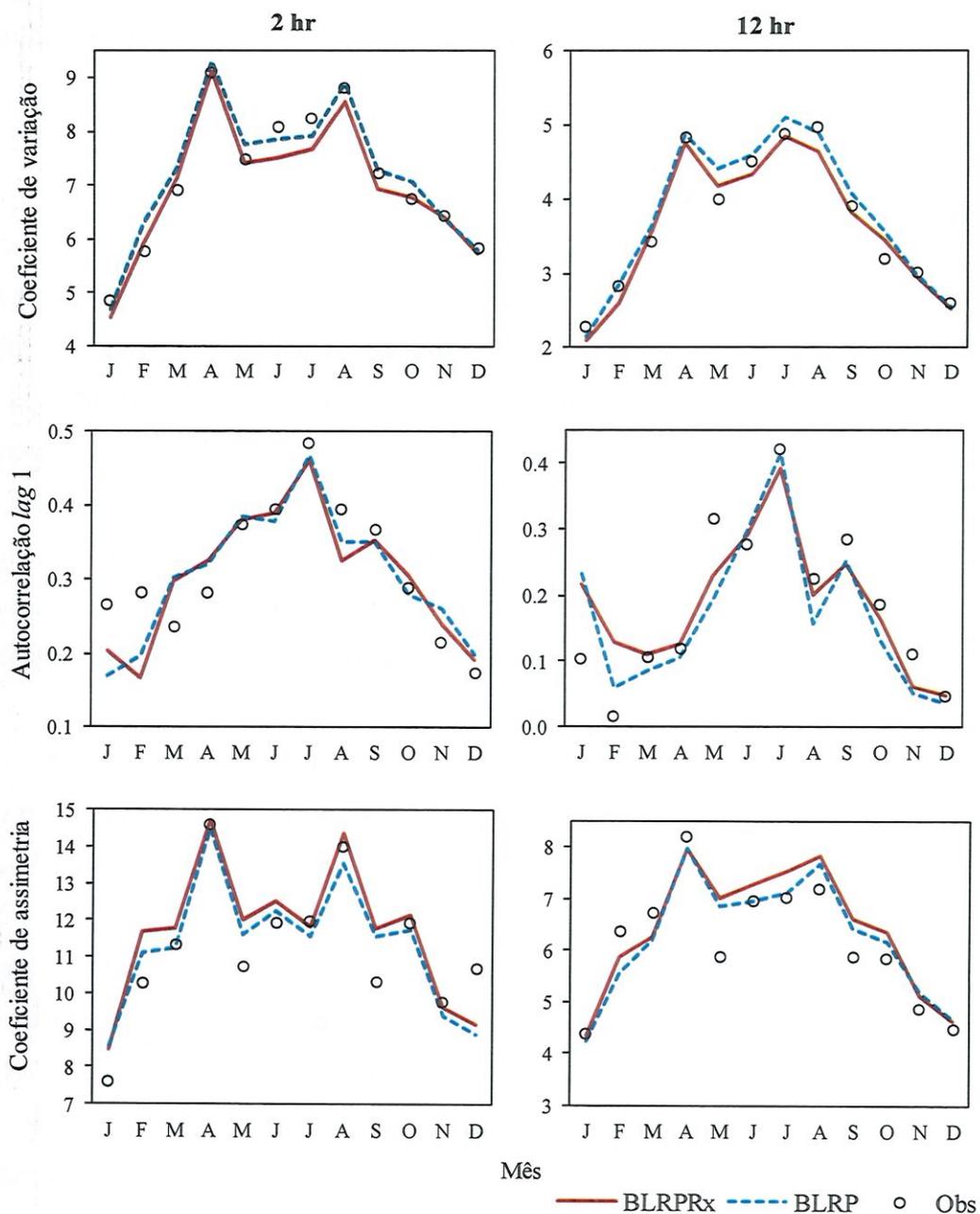
O ajuste relativo à reprodução dos coeficientes de assimetria mostrou diferenças notáveis na escala de 15 minutos, em que o BLRPRx teve um desempenho superior. Os desvios de 15 minutos são da ordem de 10% para o BLRPRx e de 20% para BLRP. O mês de agosto foi uma exceção a esse resultado e será discutido nesta subseção. Na escala de 24 horas, o BLRPRx e o BLRP resultaram em momentos aproximadamente equivalentes com desvios da ordem de 10%. Nas escalas de 1 e 6 horas os desvios são menores, da ordem de 5%, e a comparação de desempenho entre os modelos é inconclusiva.

Um comentário deve ser feito sobre os coeficientes de assimetria na escala de 15 minutos (Figura 5.5). Há uma notável diferença entre o valor observado no mês de agosto e os valores correspondentes dos demais meses. Investigando a série histórica, verificou-se que isso ocorre devido às lâminas registradas em agosto de 2001, uma vez que removendo os dados de 2001 do cálculo de momentos, o coeficiente de assimetria de 15 minutos resulta em 24,45, o que é muito próximo dos valores observados dos demais meses. Os dados registrados em agosto de 2001, contudo, não apresentaram falhas seguindo os critérios estabelecidos na seção 4.4.3. Até o momento, não foi identificado um procedimento consolidado para lidar com a variabilidade interanual dos momentos estatísticos, de modo que os dados de 2001 foram preservados nesta análise. Os efeitos dessa variabilidade são parcialmente atenuados com a utilização dos pesos (w) na função objetivo. Os pesos correspondem ao inverso da variância dos momentos interanuais, de modo que ao coeficiente de assimetria de 15 minutos é atribuída uma menor prioridade na otimização. Isso justifica, em parte, a maior discrepância observada no gráfico da Figura 5.5. Valores extremos presentes na série histórica também contribuem para essa discrepância. A discussão sobre os valores extremos será discutida na subseção 5.6.

Em condições ideais, um modelo de precipitação deveria reproduzir, com a mesma qualidade, as características da série histórica em quaisquer escalas de agregação. Na prática, isso não acontece por diversos motivos, a saber: simplificação do fenômeno por meio de pulsos retangulares, ocorrência de chuvas de naturezas fenomenológicas distintas, incerteza de registro, entre outros. Dessa forma, propriedades estatísticas agregadas em escalas que não foram utilizadas na calibração podem ser úteis para validar a aplicabilidade de um modelo estocástico.

Face ao exposto, foram selecionadas as escalas de agregação intermediárias de 2 e 12 horas, que não foram utilizadas na calibração. Os momentos estatísticos esperados nessas escalas, segundo os modelos calibrados, foram computados e comparados com os respectivos momentos observados Figura 5.6.

Figura 5.6 – Propriedades estatísticas nas escalas de 2 horas (esquerda) e 12 horas (direita), não utilizadas na calibração, momentos analíticos e observados.



Ambos os modelos reproduziram notavelmente bem os coeficientes de variação e de assimetria nas escalas intermediárias. Embora diferenças tenham sido observadas nas autocorrelações de *lag 1*, especificamente para os meses de fevereiro e janeiro, as discrepâncias médias permanecem baixas e semelhantes ao que foi verificado nas escalas de calibração: entre 0,04 (BLRP) e 0,05 (BLRPRx).

O mês de agosto, que apresentou as maiores diferenças em termos das autocorrelações, foi melhor reproduzido pelo BLRP na escala de 2 horas e, contrariamente, pelo BLRPRx na

escala de 12 horas. Constatações similares também foram observadas para outros meses, o que inviabiliza inferir qual modelo apresentou o melhor desempenho. A avaliação final dos resultados da Figura 5.6 é que a qualidade do ajuste nas escalas intermediárias é essencialmente idêntica à qualidade das escalas utilizadas na calibração.

5.4 Resultados das simulações

O objetivo da modelagem estocástica da precipitação é a geração de séries sintéticas como possíveis realizações da série histórica. Portanto, é adequado validar os modelos segundo os momentos simulados, que, por sua vez, são calculados diretamente das séries sintéticas. Dessa forma, nas próximas subseções, a análise dos resultados será feita com base nos momentos simulados, diferentemente da análise realizada na subseção 5.3, na qual foram avaliados os desempenhos dos modelos segundo os momentos analíticos. Algumas propriedades estatísticas dos modelos de Bartlett-Lewis só podem ser obtidas através de simulações, sendo esse o caso dos valores extremos e, no presente trabalho, das propriedades de períodos úmidos/secos.

Os momentos analíticos podem divergir dos momentos simulados correspondentes. Isso ocorre devido às incertezas inerentes às realizações do processo estocástico, às aproximações feitas no tratamento matemático dos momentos analíticos e, possivelmente, à existência de inconsistências lógicas nos algoritmos programados para executar as simulações. Não obstante, foram verificadas diferenças menores que 10% entre os coeficientes de variação e de assimetria analíticos e simulados. As autocorrelações de 15 minutos, 1 hora e 6 horas apresentaram resultados semelhantes. Somente as autocorrelações de 24 horas, que tem pequena magnitude, apresentaram diferenças relativas superiores a 10%. Isso será discutido na subseção 5.5. Essas diferenças são aceitáveis e demonstram a consistência das rotinas de simulação programadas, conforme descrito na subseção 4.6.

Nesta análise, foram produzidas séries sintéticas com mesma duração que a série observada. Desse modo, se janeiro, por exemplo, tem 17 anos de dados disponíveis (livres de falhas), cada série sintética gerada para esse mês tem também 17 anos de duração. É evidente que, na ausência de inconsistências, os momentos simulados de séries sintéticas infinitamente longas convergem para os momentos analíticos. Contudo, acredita-se que uma análise feita com base na repetição de séries sintéticas geradas com a mesma duração da série histórica é apropriada, até mesmo por conta de limitações computacionais. Dessa forma, os modelos BLRPR e BLRPRx foram validados através das medianas das repetições de um momento, ou

propriedade estatística, calculado para cada série sintética. Essa abordagem é semelhante à utilizada por Vanhaute et al. (2012) e Kaczmarska, Isham e Onof (2014).

Foram realizadas 100 simulações para cada mês nas escalas de agregação de 15 minutos, 1 hora, 6 horas e 24 horas, totalizando 4.800 séries sintéticas. Para cada uma das séries foram calculadas seis propriedades estatísticas: média, coeficiente de variação, autocorrelação *lag* 1, coeficiente de assimetria, proporção de períodos secos (PPS) e probabilidade de transição úmido-úmido (PTuu). Essas propriedades foram computadas da mesma forma que as estatísticas amostrais observadas, descritas na subseção 4.4.2. No âmbito do método adotado e seguido neste trabalho, as propriedades PPS e PTuu só podem ser obtidas via simulação. A mediana das 100 realizações de um mês e escala específicos foi utilizada como valor representativo dos momentos e demais propriedades estatísticas. Esse procedimento foi realizado tanto para o BLRPRx quanto para o BLRPR.

As Figuras 5.7 e 5.8 ilustram os resultados das simulações. As curvas representam as funções acumuladas de probabilidades (FAP) empíricas das 100 realizações, cada realização correspondendo a uma série sintética gerada com o BLRPRx, de uma propriedade estatística específica: precipitação média de 15 minutos, no caso da Figura 5.7, e coeficiente de assimetria de 6 horas de setembro, no caso da Figura 5.8. As linhas verticais tracejadas representam os valores observados. As medianas correspondem ao valor das FAPs que intercepta o eixo das ordenadas em 0,5. Verifica-se em ambos os casos que as medianas se aproximam razoavelmente bem dos valores observados, com melhor resultado para a média de 15 minutos de janeiro. Nota-se ainda a variabilidade, em termos absolutos, dos resultados simulados: os valores obtidos para a média variaram num intervalo de amplitude 0,05 (0,11 – 0,06), com mediana 0,081 e valor observado 0,079; e o coeficiente de assimetria oscilou num intervalo de amplitude 6,0 (5,0 – 11,0) com mediana 7,06 e valor observado 6,9.

Figura 5.7 – FAP empírica dos 100 valores simulados das médias de 15 minutos para janeiro com o BLRPRx. A linha vertical indica o valor observado.

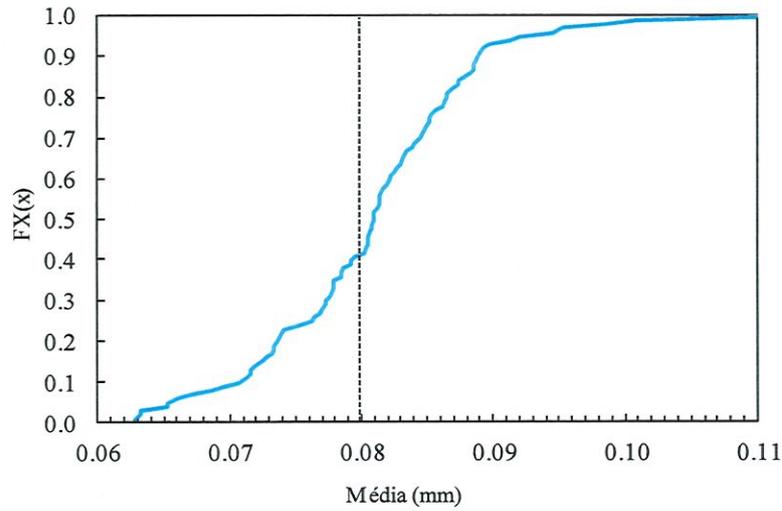
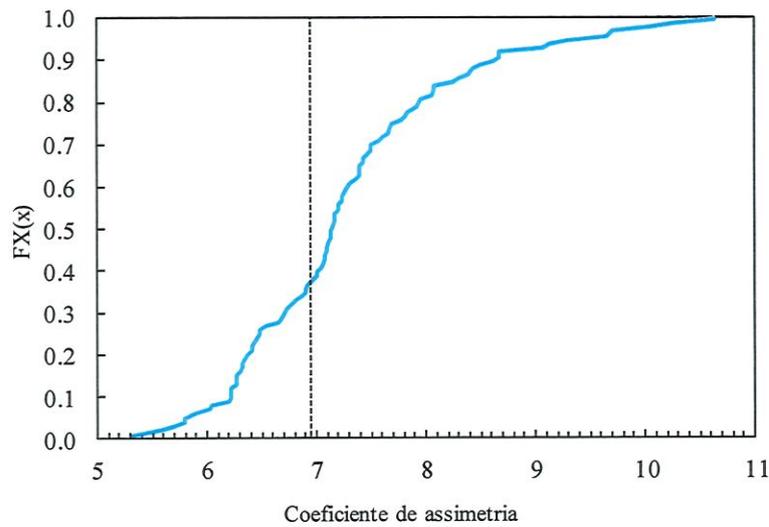


Figura 5.8 – FAP empírica dos 100 valores simulados dos coeficientes de assimetria de 15 minutos para setembro com o BLRPRx. A linha vertical indica o valor observado.



As Tabelas 5.10 e 5.11 contêm os resultados das simulações. As propriedades apresentadas nessas tabelas foram utilizadas na validação e análise de desempenho dos modelos.

Tabela 5.10 – Resultados das simulações com o BLRP. Os valores correspondem às medianas de 100 realizações.

Escala	Propriedade	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez
15 min	Média	0,079	0,054	0,046	0,021	0,020	0,018	0,011	0,008	0,020	0,032	0,046	0,062
	C.var	7,405	9,806	11,174	15,236	12,899	13,006	12,898	16,288	11,810	10,812	9,899	9,563
	Corr(1)	0,546	0,587	0,548	0,492	0,414	0,430	0,399	0,284	0,478	0,563	0,496	0,456
	Ass	12,455	16,070	18,387	24,623	22,107	22,377	22,674	29,313	20,190	17,629	16,107	16,485
	PPS	0,949	0,971	0,979	0,988	0,982	0,982	0,982	0,988	0,979	0,977	0,973	0,969
	PTuu	0,525	0,578	0,632	0,592	0,547	0,520	0,516	0,436	0,537	0,584	0,634	0,583
1 hr	Média	0,321	0,215	0,179	0,088	0,082	0,072	0,044	0,036	0,080	0,127	0,181	0,247
	C.var	5,374	7,184	8,211	10,536	8,626	8,879	8,850	10,046	8,180	8,135	7,186	6,654
	Corr(1)	0,208	0,261	0,334	0,331	0,339	0,334	0,381	0,316	0,313	0,296	0,325	0,264
	Ass	10,269	13,035	13,571	17,160	14,248	14,584	14,953	16,658	14,068	14,060	11,781	11,219
	PPS	0,886	0,941	0,962	0,977	0,967	0,963	0,969	0,974	0,959	0,955	0,953	0,940
	PTuu	0,398	0,466	0,625	0,633	0,695	0,670	0,749	0,631	0,636	0,566	0,562	0,504
6 hr	Média	1,916	1,314	1,064	0,524	0,465	0,428	0,272	0,201	0,468	0,765	1,071	1,496
	C.var	2,734	3,773	4,792	6,133	5,430	5,416	5,811	6,047	4,996	4,519	3,982	3,422
	Corr(1)	0,210	0,109	0,163	0,196	0,316	0,395	0,507	0,257	0,350	0,222	0,107	0,075
	Ass	4,944	6,298	6,861	8,785	7,538	7,617	7,539	8,418	7,055	6,826	6,017	5,419
	PPS	0,671	0,818	0,903	0,944	0,932	0,937	0,952	0,945	0,921	0,890	0,863	0,807
	PTuu	0,635	0,339	0,323	0,321	0,431	0,556	0,640	0,375	0,511	0,411	0,281	0,295
1 hr	Média	7,598	5,226	4,445	2,123	1,918	1,669	1,122	0,847	1,984	3,041	4,379	6,040
	C.var	1,675	2,074	2,668	3,540	3,309	3,692	4,191	3,560	3,203	2,614	2,144	1,805
	Corr(1)	0,190	0,013	0,036	0,042	0,072	0,151	0,249	0,067	0,115	0,050	0,013	0,029
	Ass	2,605	3,455	3,907	5,073	4,616	5,135	5,505	4,970	4,446	3,969	3,237	2,795
	PPS	0,413	0,522	0,710	0,835	0,829	0,855	0,901	0,850	0,813	0,711	0,600	0,470
	PTuu	0,671	0,489	0,318	0,237	0,275	0,309	0,345	0,241	0,323	0,350	0,415	0,525

Tabela 5.11 – Resultados das simulações com o BLRPRx. Os valores correspondem às medianas de 100 realizações.

Escala	Propriedade	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez
15 min	Média	0,079	0,056	0,046	0,021	0,020	0,017	0,011	0,009	0,020	0,032	0,046	0,063
	C.var	6,941	8,772	10,649	14,003	11,853	11,954	12,364	14,615	10,788	9,901	9,734	9,125
	Corr(1)	0,543	0,609	0,561	0,524	0,430	0,471	0,431	0,379	0,484	0,589	0,506	0,473
	Ass	14,356	17,045	20,703	26,667	23,588	23,626	22,150	28,114	21,383	20,407	18,177	17,647
	PPS	0,928	0,959	0,971	0,983	0,975	0,974	0,979	0,983	0,969	0,963	0,963	0,958
	PTuu	0,665	0,673	0,717	0,708	0,613	0,638	0,558	0,532	0,646	0,735	0,744	0,692
1 hr	Média	0,320	0,213	0,183	0,086	0,080	0,069	0,046	0,035	0,081	0,127	0,184	0,245
	C.var	5,037	6,770	7,981	10,309	8,093	8,380	8,510	9,861	7,619	7,559	7,152	6,578
	Corr(1)	0,268	0,260	0,357	0,374	0,370	0,371	0,392	0,305	0,363	0,358	0,333	0,279
	Ass	9,972	13,248	14,261	17,249	14,542	15,288	14,411	17,424	13,679	13,844	11,659	11,665
	PPS	0,864	0,924	0,953	0,972	0,955	0,954	0,960	0,967	0,945	0,940	0,948	0,930
	PTuu	0,479	0,397	0,627	0,652	0,641	0,636	0,679	0,553	0,615	0,606	0,605	0,542
6 hr	Média	1,919	1,301	1,102	0,527	0,490	0,404	0,268	0,208	0,489	0,765	1,108	1,490
	C.var	2,649	3,422	4,590	5,965	5,107	5,250	5,554	5,767	4,637	4,418	3,894	3,406
	Corr(1)	0,206	0,108	0,175	0,199	0,317	0,358	0,477	0,266	0,309	0,229	0,112	0,087
	Ass	4,787	6,396	7,123	9,003	7,772	8,300	8,010	8,759	7,383	7,198	5,920	5,372
	PPS	0,618	0,734	0,886	0,935	0,910	0,913	0,937	0,923	0,890	0,862	0,859	0,797
	PTuu	0,613	0,433	0,418	0,416	0,566	0,626	0,695	0,511	0,603	0,515	0,329	0,340
1 hr	Média	7,610	5,277	4,299	2,050	1,931	1,699	1,098	0,828	1,926	2,973	4,357	5,868
	C.var	1,604	1,920	2,618	3,451	3,191	3,383	3,939	3,553	3,000	2,629	2,126	1,812
	Corr(1)	0,189	0,147	0,056	0,055	0,137	0,188	0,262	0,107	0,152	0,089	0,028	0,024
	Ass	2,593	3,298	3,881	4,986	4,491	4,860	5,429	5,169	4,322	4,006	3,113	2,782
	PPS	0,346	0,424	0,710	0,832	0,806	0,829	0,885	0,834	0,774	0,694	0,605	0,470
	PTuu	0,740	0,686	0,362	0,263	0,387	0,439	0,420	0,346	0,431	0,430	0,419	0,538

5.5 Validação e desempenho dos modelos

Os modelos foram validados segundo sua capacidade em reproduzir as propriedades estatísticas observadas na série de precipitação de Piracicaba. Reitera-se que isso foi feito com base nas propriedades extraídas das séries sintéticas, comparando as medianas de 100 simulações com os respectivos valores observados.

Uma forma alternativa de validar os modelos, digna de menção, foi utilizada por Cowpertwait, 1991. O autor realizou testes *t* de Student para as diferenças entre os resultados simulados e o valor observado de cada propriedade, em diferentes escalas, para os doze meses do calendário. Intervalos de significância de 5% foram traçados e, ao constatar que os resultados foram obtidos em sua maioria dentro dos limites desses intervalos, o modelo foi considerado validado. Nesta análise, optou-se pela validação com base nas medianas dos resultados simulados, sobretudo porque isso promove a comparação direta do desempenho dos modelos.

As propriedades de períodos úmidos/secos, avaliadas segundo a PPS e a PTuu, são extremamente relevantes na validação dos modelos, uma vez que, ao contrário das demais propriedades, não foram inseridas na função objetivo. A proporção de períodos secos poderia ser aproveitada na calibração, pois foi desenvolvida uma expressão analítica aproximada para essa propriedade no caso do BLRP (RODRIGUEZ-ITURBE; COX; ISHAM, 1987). No entanto, expressão semelhante não foi derivada por Kaczmariska, Isham e Onof (2014) no desenvolvimento do BLRPRx, de modo que a PPS deve ser obtidas por simulação no caso desse modelo. Os autores do BLRPRx defendem que a PPS seja reservada para validar os modelos e testar a adequação de sua estrutura, e essa mesma abordagem foi seguida neste trabalho.

Na Figura 5.9 foram plotadas as PPS resultantes das simulações e os valores observados nas escalas de agregação de 15 minutos, 1, 6 e 24 horas. Verifica-se o desempenho superior do BLRPRx para as quatro escalas. Ainda assim, ambos os modelos reproduziram notavelmente bem as PPS nas escalas de 15 minutos, 1 hora e 6 horas, com desvios médios inferiores à 5%.

Na Figura 5.10 foram plotadas as PTuu resultantes das simulações e os valores observados de 15 minutos, 1, 6 e 24 horas. Novamente verifica-se o desempenho superior do BLRPRx, sobretudo para a escala de 15 minutos, com desvios médios da ordem de 7% para o BLRPRx e 20% para o BLRPR. Esses desvios aumentam progressivamente nos resultados de ambos os modelos até a escala de 24 horas, atingindo 25% (BLRPRx) e 38% (BLRPR).

Figura 5.9 – Proporção de períodos secos (PPS) por mês, resultados simulados e observados.

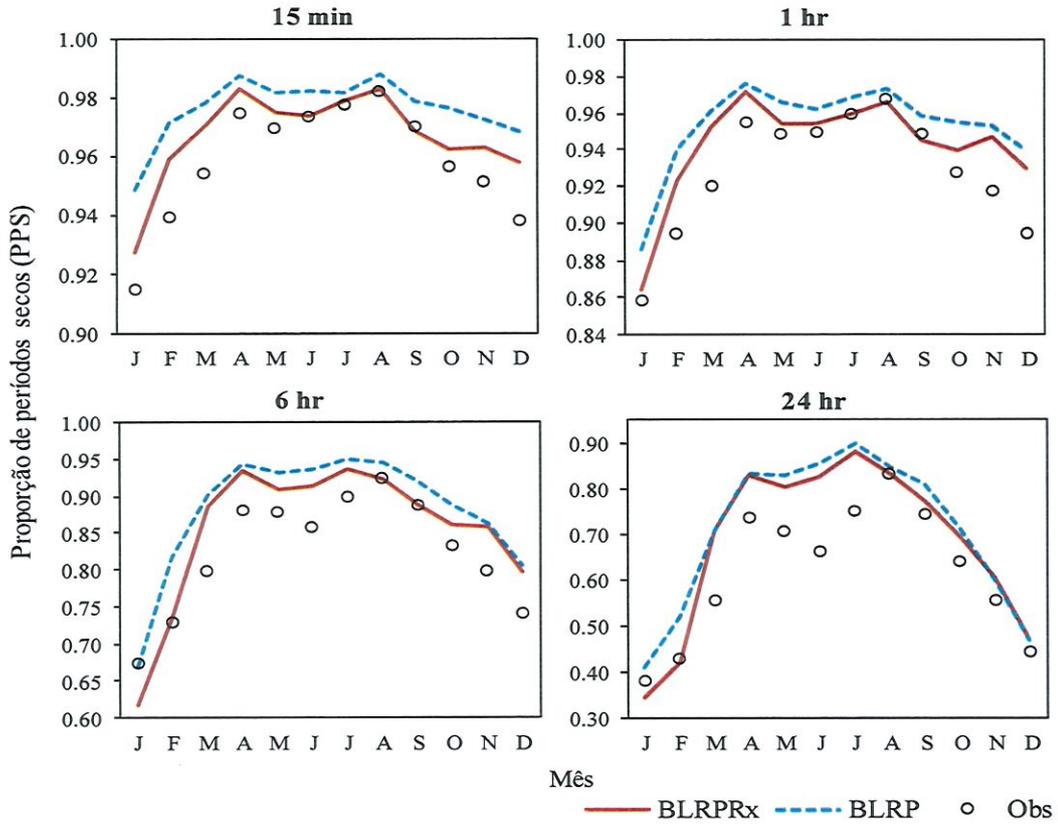
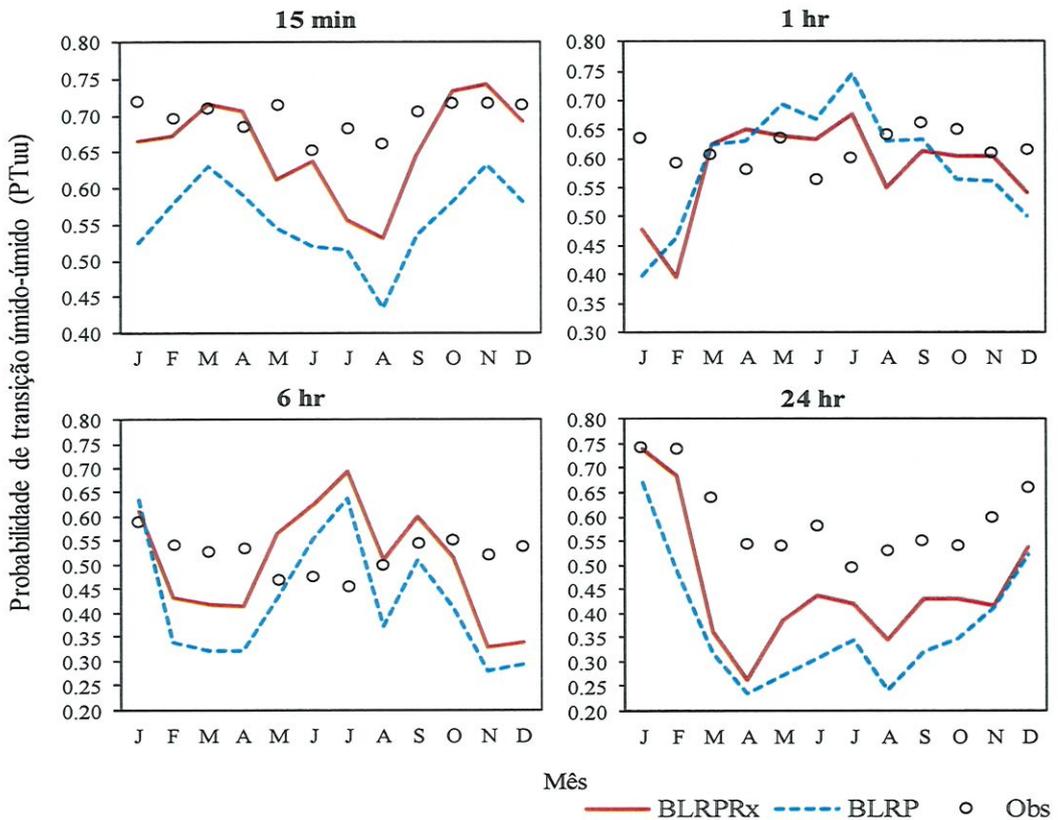


Figura 5.10 – Probabilidades de transição úmido-úmido (PTuu) por mês, resultados simulados e observados.

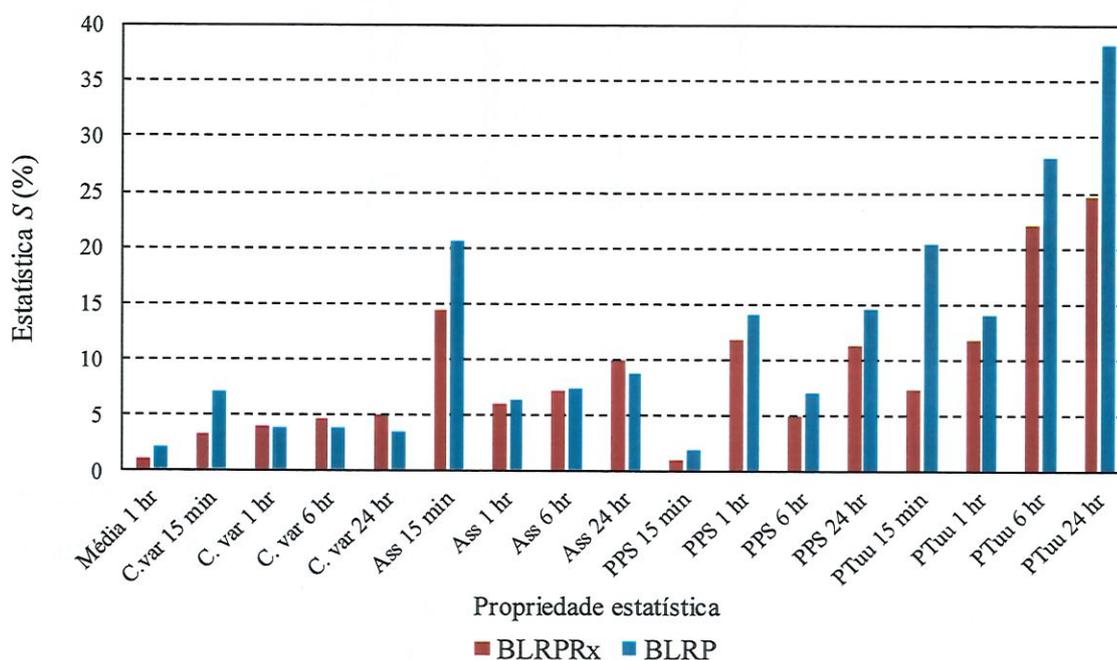


Para uma análise genérica dos resultados apresentados até o momento utilizou-se a estatística S , referente à qualidade de reprodução das propriedades, sugerida inicialmente por Velghe et al. (1994):

$$S = \frac{1}{m} \sum_{i=1}^{m=12} \left| 1 - \frac{X_i}{X_{i,hist}} \right| \times 100, \quad (5.1)$$

em que X_i é o valor simulado (mediana) de uma propriedade estatística, em uma escala de agregação específica, no mês i ; $X_{his,i}$ é o valor histórico correspondente; e $m = 12$ é o número de meses para os quais os modelos foram calibrados. A estatística S representa, portanto, a formalização de um desvio relativo médio para os meses. Trata-se de uma forma eficiente de validar e comparar os modelos. A Figura 5.11 apresenta os resultados de S obtidos para o BLRP e o BLRPRx para todas as propriedades abordadas, exceto as autocorrelações de lag 1.

Figura 5.11 – Estatística S para as propriedades analisadas, exceto a autocorrelação de lag 1.



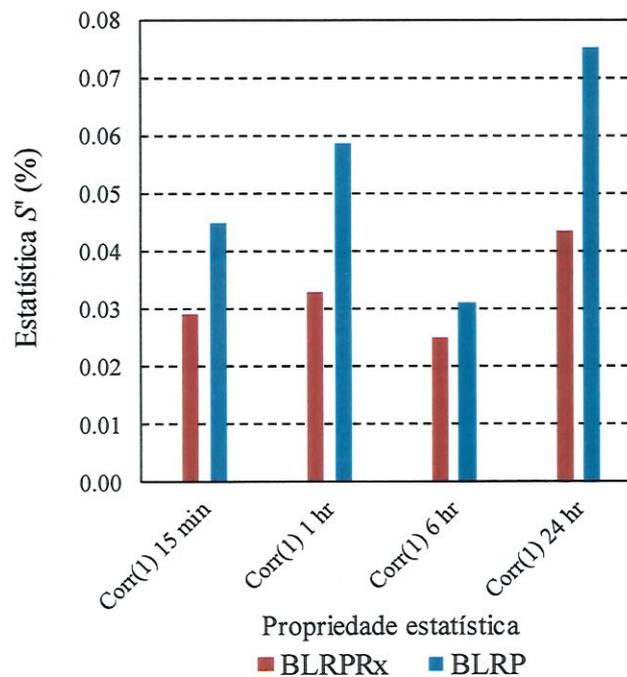
Algumas das autocorrelações de lag 1, especialmente na escala de 24 horas, tiveram valores simulados e observados muito baixos. No caso de setembro, por exemplo, as autocorrelações de 24 horas observada e simulada foram 0,091 e 0,013, respectivamente. Consequentemente, a estatística S computada foi muito alta (86%). Obviamente, uma

correlação tão baixa quanto 0,09 ou 0,01 indica que as lâminas com defasagem de 1 intervalo são, de fato, não correlacionadas. Optou-se por apresentar esses momentos na forma apropriada dos desvios não normalizados, de modo que a fórmula da estatística S utilizada no caso específico das autocorrelações de lag 1 foi:

$$S' = \frac{1}{m} \sum_{i=1}^{m=12} |X_{i,hist} - X_i| \times 100. \quad (5.2)$$

Os resultados estão apresentados na Figura 5.12.

Figura 5.12 – Estatística S' das autocorrelações de lag 1.



Os gráficos das Figuras 5.11 e 5.12 reforçam o desempenho superior do BLRPRx, que produziu diferenças notavelmente menores do que o BLRPR nas estatísticas de 15 minutos e nas propriedades de períodos secos/úmidos, em todas as escalas.

Em se tratando da validação, as simulações com o BLRPRx produziram diferenças da ordem de 5% para as médias, os coeficientes de variação e as PPSs de escalas inferiores a 24 horas. As diferenças nas autocorrelações de lag 1 foram menores do que 0,05. Com base nessas observações e na avaliação de Onof et al. (2000), que resumiram os resultados empíricos de estudos realizados na Escócia e Bélgica, pode-se afirmar que o ajuste obtido com o BLRPRx para Piracicaba é de muito boa qualidade.

Evidentemente existem algumas restrições nos modelos, sobretudo quanto à reprodução dos coeficientes de assimetria de 15 minutos e 24 horas e das probabilidades de transição úmido-úmido (PTuu). As causas do primeiro são atribuídas à maior variabilidade interanual dos coeficientes de assimetria, o que consistentemente diminui sua prioridade na otimização, e à presença de valores extremos na série histórica. No que se refere à PTuu, verifica-se que a incapacidade dos modelos em reproduzir essa propriedade aumenta conforme aumenta a escala de agregação. Portanto, nos estudos de simulação contínua em que a PTuu é uma propriedade prioritária, os modelos de Bartlett-Lewis devem ser aplicados com cautela.

Análises dos modelos de precipitação de Poisson considerando os momentos de terceira ordem, tanto na calibração quanto na validação, e a probabilidade de transição úmido-úmido são relativamente escassas na literatura. Nas aplicações dos modelos de Bartlett-Lewis realizadas no Brasil, a análise estatística se restringiu aos momentos de primeira e segunda ordem e à PPS, que foi utilizada em calibrações do modelo BLRPR (BACK; DORFMAN; CLARKE, 1999; BACK; UGGIONI; VIEIRA, 2011; TEIXEIRA-GANDRA; DAMÉ, 2014). Conforme mencionado na subseção 4.2, há evidência de que a utilização da PPS na função objetivo produz um melhor ajuste para essa propriedade às custas de uma deterioração nos momentos de terceira ordem, o que tem impacto na reprodução dos valores extremos. Entretanto, nos estudos mencionados, não foram feitas análises de valores extremos.

5.6 Análise de valores extremos

A reprodução fidedigna dos valores extremos observados, nas séries sintéticas, é extremamente desejável e em alguns casos imprescindível, como nos estudos que têm por objetivo quantificar o risco de falha de estruturas e sistemas hidráulicos. Ao longo de décadas, esforços vêm sendo realizados para aprimorar a reprodução dos extremos nos modelos estocásticos de Poisson e esse problema continua em aberto. A própria formulação diferenciada do BLRPR_x, cuja estrutura tem intensidades dependentes da duração das células, e a proposta de uso dos momentos de terceira ordem na calibração são alguns exemplos de tais esforços.

Existem diversos estudos com resultados empíricos em que os valores extremos são analisados por comparação entre os máximos anuais sintéticos, produzidos como resultado das simulações, e os máximos anuais da série histórica. Usualmente, essa comparação é feita por análise gráfica em papel de probabilidades de Gumbel (KOUTSOYIANNIS; ONOF,

2001; COWPERTWAIT et al., 2011; KACZMARSKA; ISHAM; ONOF, 2014). Na literatura consultada, não foram identificados estudos que fazem inferência analítica sobre a distribuição de probabilidades dos máximos dos modelos de Bartlett-Lewis ou Neyman-Scott.

Uma análise gráfica foi considerada apropriada para avaliar os resultados empíricos de Piracicaba. A série histórica tem 19 anos de duração, mas apresenta vários meses contendo falhas em anos diferentes (subseção 4.4.3). Considerando somente os anos em que não há meses com falhas, a série é reduzida para 5 anos, o que inviabiliza uma análise confiável dos máximos anuais. Como alternativa, foi feita uma análise para meses separados de modo que os máximos de cada ano foram identificados utilizando somente os dados de dezembro, por exemplo. Esse procedimento é consistente porque na prática, para conciliar estacionaridade e sazonalidade, foi ajustado um modelo diferente para cada mês com um conjunto específico de parâmetros. Considera-se ainda que a análise baseada nos máximos anuais pode omitir eventuais discrepâncias que são verificadas somente com a separação dos meses (COWPERTWAIT et al., 2011).

Inicialmente, foram selecionados para investigação os meses representativos contendo o menor número de falhas e distribuídos ao longo do ano: dezembro, março, julho e setembro. A disponibilidade de dados sem falhas varia entre 18 (dezembro e março) e 19 anos (julho e setembro). Para esses meses, foram coletadas as maiores lâminas de 15 minutos de cada ano, compondo assim a série de extremos observados. A análise foi limitada à escala de agregação de 15 minutos, equivalente à resolução da série histórica, admitindo-se que nessa escala são observadas de forma mais detalhada as intensidades médias.

Os procedimentos descritos a seguir, relativos à análise de valores extremos, foram executados em paralelo para o BLRP e o BLRPRx. Para cada mês, foram geradas 100 séries sintéticas com duração igual ao número de anos e dias do mês correspondente. Para cada série, foram identificadas as maiores lâminas de 15 minutos que compõem a sequência de máximos anuais. Dessa forma, para dezembro, por exemplo, foram identificados os máximos gerados em intervalos sucessivos de 31 dias. As sequências de máximos de cada mês foram classificadas em ordem crescente e, respeitando essa ordem, extraiu-se a mediana e os percentis inferior e superior, correspondentes a um intervalo de abrangência de 95% dos 100 valores realizados.

Alternativamente, os resultados das simulações poderiam ser avaliados através dos intervalos de predição, ou seja, intervalos calculados por inferência estatística que contêm os possíveis valores amostrados da população (NAVIDI, 2011). No entanto, os extremos de diferentes realizações, considerando a mesma posição dentro da amostra, não necessariamente

seguem distribuições normais. Portanto, intervalos limitados inferiormente pelo percentil 2,5 e superiormente 97,5, aqui denominados intervalos de 95% de abrangência, foram considerados apropriados para representar a incerteza nas simulações.

Uma vez estabelecidas as três sequências de máximos (mediana, percentil 2,5 e percentil 97,5), foram calculadas as posições de plotagem p . Essas posições equivalem a uma estimativa empírica das probabilidades de superação $P(Y > y)$, calculadas segundo a fórmula de Gringorten:

$$p = \frac{k - 0,44}{n + 0,12}, \quad (5.1)$$

em que k é a ordem do valor dentro da série de máximos e n é número de anos que foram considerados na composição dessa série. O valor esperado do tempo de recorrência entre superações, ou seja, o tempo de retorno (T_r), é dado pelo inverso de p :

$$T_r = \frac{1}{p}. \quad (5.2)$$

A probabilidade empírica de não superação, $P(Y \leq y)$, que pode ser efetivamente utilizada na FAP de Gumbel em sua forma convencional, é:

$$P(Y \leq y) = 1 - p = 1 - \frac{1}{T_r}. \quad (5.3)$$

Os valores extremos observados e simulados foram plotados segundo as variáveis reduzidas da distribuição de Gumbel, aqui denotadas por g , calculadas conforme:

$$g(T_r) = -\ln \left[-\ln \left(1 - \frac{1}{T_r} \right) \right], \quad (5.4)$$

de modo que tempos de retorno $T_r = 1; 5$ e $33,6$ anos correspondem a $g = -1,5; 1,5$ e $3,5$, respectivamente.

As sequências de máximos anuais contêm entre 18 e 19 dados, de modo que os maiores tempos de retorno são da ordem de 30 anos. Uma alternativa para avaliar os modelos para tempos de retorno maiores que esse limite seria ajustar alguma distribuição de máximos

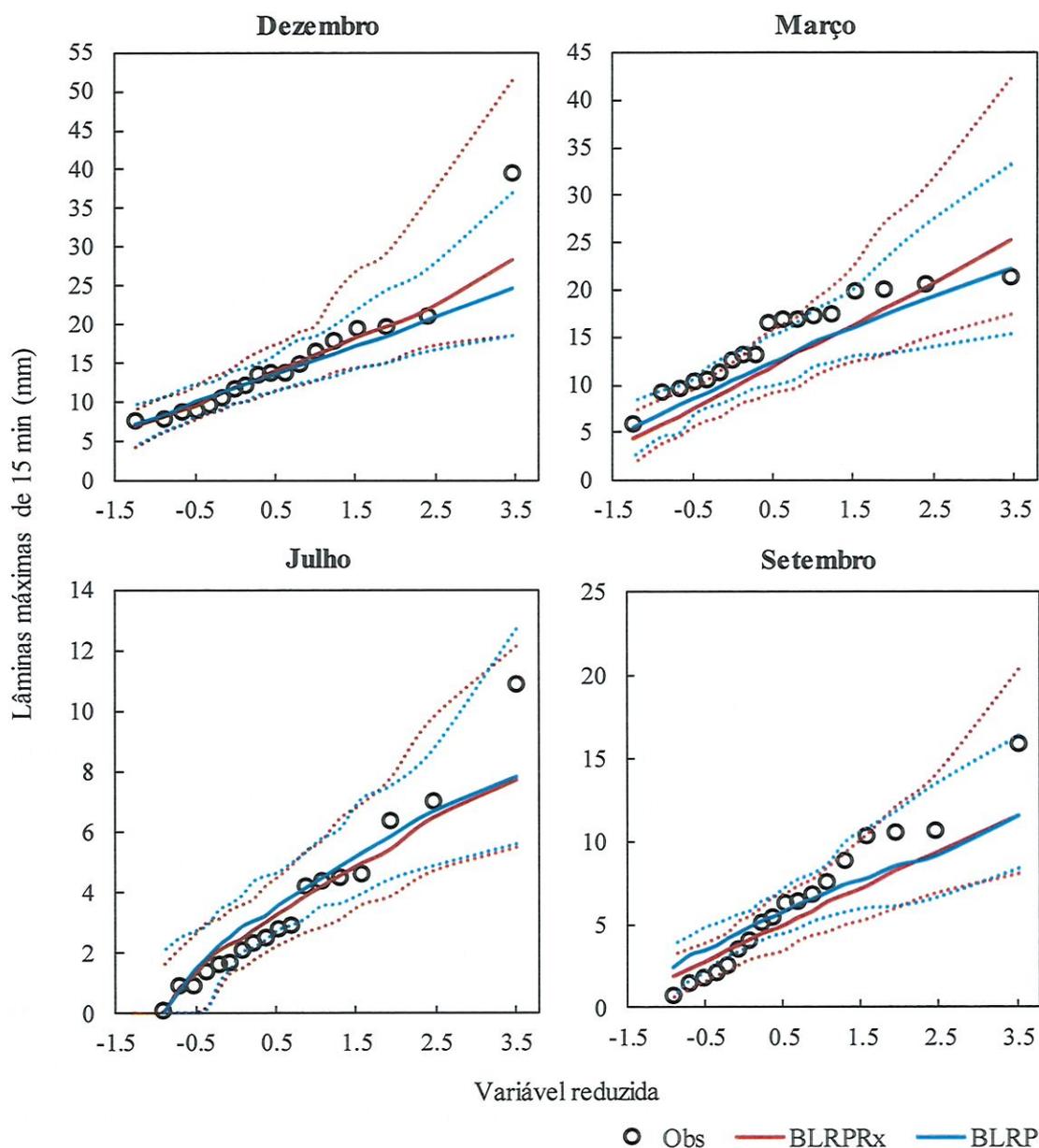
aos dados observados, extrapolar os valores e executar novas simulações com séries sintéticas de maior duração. Contudo, avaliou-se que o ajuste de uma distribuição e a extrapolação dos dados somente adicionaria mais incerteza à análise de valores extremos, que já conta com a incerteza inerente às posições de plotagem.

Na Figura 5.13 são apresentados os gráficos para a análise de valores extremos dos meses representativos. Os máximos observados e os resultados das simulações com o BLRP e o BLRPRx foram plotados juntamente com os valores da variável reduzida de Gumbel. Devido às aproximações no cálculo das posições de plotagem, alguns extremos da amostra podem exibir um comportamento atípico em relação aos outros pontos amostrais (NAGHETTINI; PINTO, 2007). Entretanto, a função dos gráficos é comparativa e as distorções nesses extremos devem afetar igualmente os valores observados e simulados.

Nos meses mais secos de julho e setembro, os extremos foram consistentemente reproduzidos, com os valores observados incidindo dentro dos intervalos de 95% de abrangência de ambos os modelos. Devido às incertezas, não é possível afirmar qual modelo apresentou um melhor desempenho para julho. Em setembro parece que o BLRPRx acompanha melhor a tendência dos extremos observados.

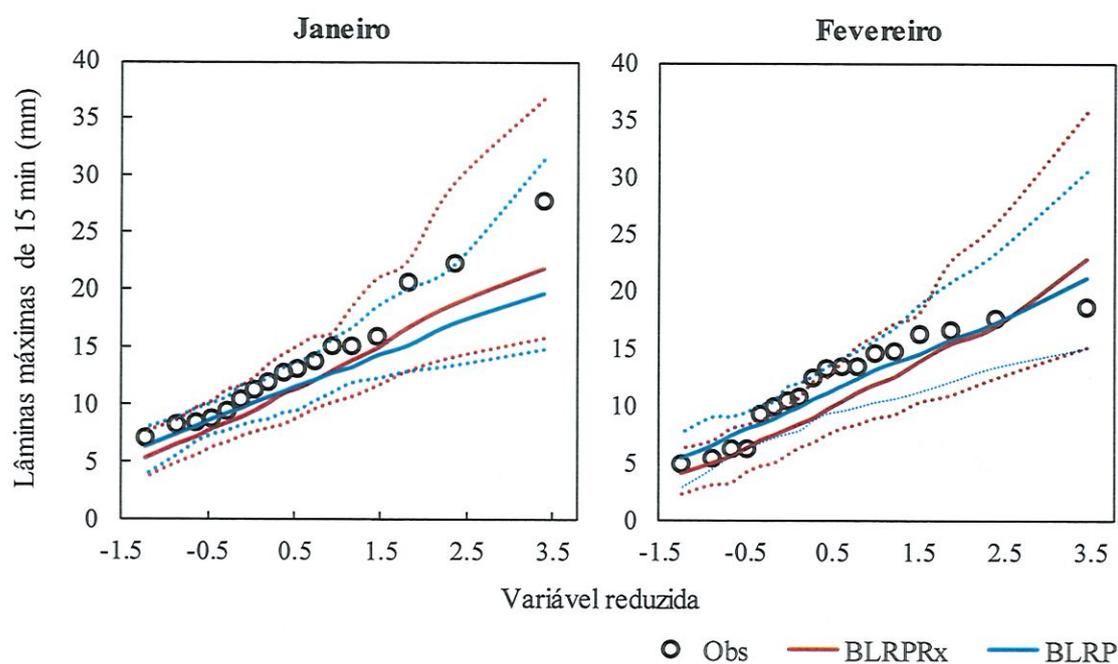
No mês de dezembro, em que foram registradas as maiores intensidades, ambos os modelos reproduziram notavelmente bem os extremos, com os valores observados incidindo próximos às medianas e dentro dos intervalos de 95% de abrangência. Observa-se que o BLRPRx reproduziu dentro desse intervalo inclusive o extremo com maior tempo de retorno, ao passo que o BLRP falhou nesse ponto. Evidentemente isso ocorreu às custas de uma maior incerteza: a amplitude do intervalo de 95% de abrangência do BLRPRx é visivelmente maior. Observa-se que as amplitudes de ambos os modelos guardam uma relação crescente com T_r , e que o aumento ocorre em maior escala no BLRPRx. Isso é atribuído a maior flexibilidade estrutural do modelo. As distribuições das intensidades e durações das células, condicionadas a η , abrangem mais valores no caso do BLRPRx, o que contribui para a geração de uma maior variedade de lâminas extremas.

Figura 5.13 – Máximos anuais, separados por mês, e variável reduzida de Gumbel: resultados simulados e observados. As linhas contínuas correspondem às medianas de 100 realizações ordenadas dos máximos anuais. As linhas pontilhadas correspondem aos percentis 2,5 e 97,5.



Para março verificou-se uma tendência dos modelos em subestimar os extremos observados, embora isso não tenha ocorrido para o valor com o maior tempo de retorno. Nos gráficos de dezembro e março, verificou-se ainda outra tendência: para os tempos de retorno superiores, os extremos do BLRPRx são maiores do que os do BLRP. Disso decorre que, caso ambas as tendências possam ser confirmadas, a subestimação dos extremos mais críticos ocorre com menor intensidade no BLRPRx. Avançando nessa questão, a análise de extremos foi estendida para os demais meses chuvosos de janeiro e fevereiro (Figura 5.14).

Figura 5.14 – Máximos anuais, de janeiro e fevereiro, e variável reduzida de Gumbel: resultados simulados e observado. As linhas contínuas correspondem às medianas de 100 realizações ordenadas dos máximos anuais. As linhas pontilhadas correspondem aos percentis 2,5 e 97,5.



A tendência de subestimação dos extremos foi observada somente em janeiro, ao passo que a tendência do BLRPRx em reproduzir extremos maiores para T_r de ordem superior foi observada em ambos os meses, janeiro e fevereiro. Essas considerações reforçam a necessidade de uma série histórica de maior duração para confirmar o melhor desempenho do BLRPRx.

Avaliando de forma genérica os resultados dessa subseção, pode-se afirmar que os extremos com tempos de retorno menores que 12,7 anos ($g < 2,5$) são bem reproduzidos por ambos os modelos, ao passo que para tempos de retorno maiores, parece ocorrer subestimação dos extremos observados. Embora a análise de valores extremos tenha produzido fraca evidência de que o BLRPRx tem um melhor desempenho, uma série mais longa é necessária para consolidar essa constatação. Essas conclusões condizem com os trabalhos de diversos autores (VERHOEST et al., 2010; HANAISH; IBRAHIM; JEMAIN, 2013; KACZMARSKA; ISHAM; ONOF, 2014; KOSSIERIS et al., 2018).

6 CONCLUSÃO

Este trabalho aplicou, em paralelo, dois modelos estocásticos aos dados de chuva de um local representativo do clima do sudeste brasileiro. O primeiro, o BLRP, é um modelo consolidado, ao passo que o segundo, o BLRPRx, consiste em um desenvolvimento recente da literatura.

Ao longo do desenvolvimento foram superadas diversas dificuldades, sobretudo no que se refere à disponibilidade de uma série histórica adequada à modelagem estocástica e aos fatores combinados que possibilitaram o sucesso na estimação de parâmetros consistentes. Para tanto, o método de calibração foi formulado com duas novidades, em relação às aplicações anteriores dos modelos de Bartlett-Lewis no Brasil, a saber: a utilização dos momentos de terceira ordem na função objetivo e da matriz de pesos aproximadamente ótima do MGM.

Os resultados indicam que ambos os modelos reproduzem bem os momentos de primeira a terceira ordem observados, enquanto que o BLRPRx tem um desempenho superior na reprodução desses momentos e da proporção de períodos secos. Portanto, o BLRPRx é recomendado para modelagem de chuvas caracterizadas por condições meteorológicas semelhantes às de Piracicaba, cujos dados foram utilizados no estudo.

É evidente que a modelagem de um fenômeno tão complexo quanto a precipitação tem limitações, sobretudo se os modelos são caracterizados por um pequeno número de parâmetros, sendo esse o caso dos modelos aplicados. Os reflexos disso foram evidenciados na reprodução das probabilidades de transição úmido-úmido, em que ambos os modelos falharam, embora o BLRPRx tenha produzido resultados menos discrepantes dos observados.

A análise de extremos indicou que os modelos podem ser utilizados em aplicações hidrológicas de baixo risco, com tempos de retorno inferiores a 12 anos. Entretanto, a duração limitada da série histórica inviabilizou constatações sobre o comportamento dos modelos para tempos de retorno superiores.

A aplicação do BLRPRx às condições pluviométricas de Piracicaba contribuiu com mais resultados empíricos para a literatura relacionada aos modelos de precipitação de Poisson.

REFERÊNCIAS¹

- BACK, Á.; DORFMAN, R.; CLARKE, R. Modelagem da Precipitação Horária por meio do Modelo de Pulsos Retangulares de Bartlett-Lewis Modificado. **Revista Brasileira de Recursos Hídricos**, v. 4, n. 1, p. 5–17, 1999.
- BACK, Á. J.; UGGIONI, E.; VIEIRA, H. J. Modeling precipitation of short duration by means of the modified Bartlett-Lewis rectangular pulse model. **Revista Brasileira de Meteorologia**, v. 26, n. 3, p. 461–472, 2011.
- BO, Z.; ISLAM, S.; ELTAHIR, E. A. B. Aggregation-disaggregation properties of a stochastic rainfall model. **Water Resources Research**, v. 30, n. 12, p. 3423–3435, 1994.
- BOUGHTON, W.; DROOP, O. Continuous simulation for design flood estimation - a review. **Environmental Modelling & Software**, v. 18, n. 4, p. 309–318, 2003.
- BURTON, A.; FOWLER, H. J.; BLENKINSOP, S.; KILSBY, C. G. Downscaling transient climate change using a Neyman–Scott Rectangular Pulses stochastic rainfall model. **Journal of Hydrology**, v. 381, n. 1-2, p. 18–32, 2010.
- BURTON, A.; KILSBY, C. G.; FOWLER, H. J.; COWPERTWAIT, P. S. P.; O’CONNELL, P. E. RainSim: A spatial–temporal stochastic rainfall modelling system. **Environmental Modelling & Software**, v. 23, n. 12, p. 1356–1369, 2008.
- CHANDLER, R. E. A Spectral Method for Estimating Parameters in Rainfall Models. **Bernoulli**, v. 3, n. 3, p. 301, 1997.
- COWPERTWAIT, P.; ISHAM, V.; ONOF, C. Point process models of rainfall: developments for fine-scale structure. **Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences**, v. 463, n. 2086, p. 2569–2587, 2007.
- COWPERTWAIT, P. S. P. Further developments of the neyman-scott clustered point process for modeling rainfall. **Water Resources Research**, v. 27, n. 7, p. 1431–1438, 1991.
- COWPERTWAIT, P. S. P. A Generalized Point Process Model for Rainfall. **Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences**, v. 447, n. 1929, p. 23–37, 1994.
- COWPERTWAIT, P. S. P. A Generalized Spatial-Temporal Model of Rainfall Based on a Clustered Point Process. **Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences**, v. 450, n. 1938, p. 163–175, 1995.
- COWPERTWAIT, P. S. P. A Poisson-cluster model of rainfall: some high-order moments and extreme values. **Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences**, v. 454, n. 1971, p. 885–898, 1998.
- COWPERTWAIT, P. S. P.; O’CONNELL, P. E.; METCALFE, A. V.; MAWDSLEY, J. A. Stochastic point process modelling of rainfall. I. Single-site fitting and validation. **Journal of Hydrology**, v. 175, n. 1-4, p. 17–46, 1996.

¹ De acordo com a Associação Brasileira de Normas Técnicas (ABNT NBR 6023).

COWPERTWAIT, P. S. P.; XIE, G.; ISHAM, V.; ONOF, C.; WALSH, D. C. I. A fine-scale point process model of rainfall with dependent pulse depths within cells. **Hydrological Sciences Journal**, v. 56, n. 7, p. 1110–1117, 2011.

COX, D. R.; ISHAM, V. **Point Processes**. CRC Press, 1980.

CROSS, D.; ONOF, C.; WINTER, H.; BERNARDARA, P. Censored rainfall modelling for estimation of fine-scale extremes. **Hydrology and Earth System Sciences**, v. 22, n. 1, p. 727–756, 2018.

DUAN, Q.; SOROOSHIAN, S.; GUPTA, V. Effective and efficient global optimization for conceptual rainfall-runoff models. **Water Resources Research**, v. 28, n. 4, p. 1015–1031, 1992.

DUAN, Q.; SOROOSHIAN, S.; GUPTA, V. K. Optimal use of the SCE-UA global optimization method for calibrating watershed models. **Journal of Hydrology**, v. 158, n. 3-4, p. 265–284, 1994.

DUAN, Q. Y.; GUPTA, V. K.; SOROOSHIAN, S. Shuffled complex evolution approach for effective and efficient global minimization. **Journal of Optimization Theory and Applications**, v. 76, n. 3, p. 501–521, 1993.

ENTEKHABI, D.; RODRIGUEZ-ITURBE, I.; EAGLESON, P. S. Probabilistic representation of the temporal rainfall process by a modified Neyman-Scott Rectangular Pulses Model: Parameter estimation and validation. **Water Resources Research**, v. 25, n. 2, p. 295–302, 1989.

FROST, A. J.; SRIKANTHAN, R.; COWPERTWAIT, P. S. P. **Stochastic generation of point rainfall data at sub-daily timescales: A comparison of DRIP and NSRP**. Relatório Técnico. Cooperative Research Centre, 2004.

HANAISH, I. S.; IBRAHIM, K.; JEMAIN, A. A. On the potential of Bartlett Lewis rectangular pulse models for simulating rainfall in Peninsular Malaysia. **Hydrological Sciences Journal**, v. 58, n. 8, p. 1690–1703, 2013.

HANCOCK, G. R.; VERDON-KIDD, D.; LOWRY, J. B. C. Sediment output from a post-mining catchment – Centennial impacts using stochastically generated rainfall. **Journal of Hydrology**, v. 544, p. 180–194, 2017.

HANSEN, L. P. Large Sample Properties of Generalized Method of Moments Estimators. **Econometrica**, v. 50, n. 4, p. 1029, 1982.

HENEKER, T. M.; LAMBERT, M. F.; KUCZERA, G. A point rainfall model for risk-based design. **Journal of Hydrology**, v. 247, n. 1-2, p. 54–71, 2001.

ISLAM, S.; ENTEKHABI, D.; BRAS, R. L.; RODRIGUEZ-ITURBE, I. Parameter estimation and sensitivity analysis for the modified Bartlett-Lewis rectangular pulses model of rainfall. **Journal of Geophysical Research**, v. 95, n. D3, p. 2093, 1990.

JENNINGS, S. A.; LAMBERT, M. F.; KUCZERA, G. Generating synthetic high resolution rainfall time series at sites with only daily rainfall using a master–target scaling approach. **Journal of Hydrology**, v. 393, n. 3-4, p. 163–173, 2010.

JESUS, J.; CHANDLER, R. E. Estimating functions and the generalized method of moments. **Interface Focus**, v. 1, n. 6, p. 871–885, 2011.

KACZMARSKA, J. M.; ISHAM, V.; ONOF, C. Point process models for fine-resolution rainfall. **Hydrological Sciences Journal**, v. 59, n. 11, p. 1972–1991, 2014.

KACZMARSKA, J. M. **Single-site point process-based rainfall models in a nonstationary climate**. 2013. 203 f. Tese de Doutorado - Department of Statistical Science, University College of London, Londres, 2013.

KACZMARSKA, J. M.; ISHAM, V. S.; NORTHROP, P. Local generalised method of moments: an application to point process-based rainfall models. **Environmetrics**, v. 26, n. 4, p. 312–325, 2015.

KAVVAS, M. L.; DELLEUR, J. W. **The Stochastic and Chronologic Structure Of Rainfall Sequences: Application To Indiana**. Relatório Técnico. Purdue University, 1975.

KHALIQ, M. N.; CUNNANE, C. Modelling point rainfall occurrences with the modified Bartlett-Lewis rectangular pulses model. **Journal of Hydrology**, v. 180, n. 1-4, p. 109–138, 1996.

KILSBY, C. G.; JONES, P. D.; BURTON, A.; FORD, A. C.; FOWLER, H. J.; HARPHAM, C.; JAMES, P.; SMITH, A.; WILBY, R. L. A daily weather generator for use in climate change studies. **Environmental Modelling and Software**, v. 22, n. 12, p. 1705–1719, 2007.

KIM, D.; CHO, H.; ONOF, C.; CHOI, M. Let-It-Rain: a web application for stochastic point rainfall generation at ungaged basins and its applicability in runoff and flood modeling. **Stochastic Environmental Research and Risk Assessment**, v. 31, n. 4, p. 1023–1043, 2017.

KIM, J. G.; KWON, H. H.; KIM, D. A hierarchical Bayesian approach to the modified Bartlett-Lewis rectangular pulse model for a joint estimation of model parameters across stations. **Journal of Hydrology**, v. 544, p. 210–223, 2017.

KOSSIERIS, P.; MAKROPOULOS, C.; ONOF, C.; KOUTSOYIANNIS, D. A rainfall disaggregation scheme for sub-hourly time scales: Coupling a Bartlett-Lewis based model with adjusting procedures. **Journal of Hydrology**, v. 556, p. 980–992, 2018.

KOUTSOYIANNIS, D.; ONOF, C. Rainfall disaggregation using adjusting procedures on a Poisson cluster model. **Journal of Hydrology**, v. 246, n. 1-4, p. 109–122, 2001.

MARIANI, V. C.; JUSTI LUVIZOTTO, L. G.; GUERRA, F. A.; DOS SANTOS COELHO, L. A hybrid shuffled complex evolution approach based on differential evolution for unconstrained optimization. **Applied Mathematics and Computation**, v. 217, n. 12, p. 5822–5829, 2011.

NAGHETTINI, M.; PINTO, É. J. de A. **Hidrologia Estatística**. CPRM, 2007.

- NAVIDI, W. **Statistics for Engineers and Scientists**. 3^a. ed. McGraw-Hill, 2011.
- NELDER, J. A.; MEAD, R. A Simplex Method for Function Minimization. **The Computer Journal**, v. 7, n. 4, p. 308–313, 1965.
- ONOF, C.; CHANDLER, R. E.; KAKOU, A.; NORTHROP, P.; WHEATER, H. S.; ISHAM, V. Rainfall modelling using Poisson-cluster processes: a review of developments. **Stochastic Environmental Research and Risk Assessment**, v. 14, n. 6, p. 384–411, 2000.
- ONOF, C.; WHEATER, H. S. Modelling of British rainfall using a random parameter Bartlett-Lewis Rectangular Pulse Model. **Journal of Hydrology**, v. 149, n. 1-4, p. 67–95, 1993.
- ONOF, C.; WHEATER, H. S. Improvements to the modelling of British rainfall using a modified Random Parameter Bartlett-Lewis Rectangular Pulse Model. **Journal of Hydrology**, v. 157, n. 1-4, p. 177–195, 1994.
- PARZEN, E. **Stochastic Processes**. 1^a. ed. Philadelphia: Society for Industrial and Applied Mathematics, 1962.
- PHAM, M. T.; VANHAUTE, W. J.; VANDENBERGHE, S.; DE BAETS, B.; VERHOEST, N. E. C. An assessment of the ability of Bartlett-Lewis type of rainfall models to reproduce drought statistics. **Hydrology and Earth System Sciences**, v. 17, n. 12, p. 5167–5183, 2013.
- RAMESH, N. I.; ONOF, C.; XIE, D. Doubly stochastic Poisson process models for precipitation at fine time-scales. **Advances in Water Resources**, v. 45, p. 58–64, 2012.
- RITSCHER, C.; ULBRICH, U.; NÉVIR, P.; RUST, H. W. Precipitation extremes on multiple timescales – Bartlett–Lewis rectangular pulse model and intensity–duration–frequency curves. **Hydrology and Earth System Sciences**, v. 21, n. 12, p. 6501–6517, 2017.
- RODRIGUEZ-ITURBE, I.; COX, D. R.; ISHAM, V. Some Models for Rainfall Based on Stochastic Point Processes. **Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences**, v. 410, n. 1839, p. 269–288, 1987.
- RODRIGUEZ-ITURBE, I.; COX, D. R.; ISHAM, V. A Point Process Model for Rainfall: Further Developments. **Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences**, v. 417, n. 1853, p. 283–298, 1988.
- RODRIGUEZ-ITURBE, I.; GUPTA, V. K.; WAYMIRE, E. Scale considerations in the modeling of temporal rainfall. **Water Resources Research**, v. 20, n. 11, p. 1611–1619, 1984.
- SALIM, A.; PAWITAN, Y. Extensions of the Bartlett-Lewis model for rainfall processes. **Statistical Modelling: An International Journal**, v. 3, n. 2, p. 79–98, 2003.
- SMITHERS, J. C.; PEGRAM, G. G. S.; SCHULZE, R. E. Design rainfall estimation in South Africa using Bartlett-Lewis rectangular pulse rainfall models. **Journal of Hydrology**, v. 258, n. 1-4, p. 83–99, 2002.

STORN, R.; PRICE, K. Differential Evolution – A Simple and Efficient Heuristic for global Optimization over Continuous Spaces. **Journal of Global Optimization**, v. 11, n. 4, p. 341–359, 1997.

TEIXEIRA-GANDRA, C. F. A.; DAMÉ, R. de C. F. Bartlett-Lewis of rectangular pulse modified model: estimate of parameters for simulation of precipitation in sub-hourly duration. **Engenharia Agrícola**, v. 34, n. 5, p. 925–934, 2014.

THAYAKARAN, R.; RAMESH, N. I. Doubly stochastic Poisson pulse model for fine-scale rainfall. **Stochastic Environmental Research and Risk Assessment**, v. 31, n. 3, p. 705–724, 2017.

VANHAUTE, W. J.; VANDENBERGHE, S.; SCHEERLINCK, K.; DE BAETS, B.; VERHOEST, N. E. C. Calibration of the modified Bartlett-Lewis model using global optimization techniques and alternative objective functions. **Hydrology and Earth System Sciences**, v. 16, n. 3, p. 873–891, 2012.

VELGHE, T.; TROCH, P. A.; DE TROCH, F. P.; VAN DE VELDE, J. Evaluation of cluster-based rectangular pulses point process models for rainfall. **Water Resources Research**, v. 30, n. 10, p. 2847–2857, 1994.

VERHOEST, N. E. C.; VANDENBERGHE, S.; CABUS, P.; ONOF, C.; MECA-FIGUERAS, T.; JAMELEDDINE, S. Are stochastic point rainfall models able to preserve extreme flood statistics? **Hydrological Processes**, v. 24, n. 23, p. 3439–3445, 2010.

VERHOEST, N.; TROCH, P. A.; DE TROCH, F. P. On the applicability of Bartlett-Lewis rectangular pulses models in the modeling of design storms at a point. **Journal of Hydrology**, v. 202, n. 1-4, p. 108–120, 1997.

WASKO, C.; PUI, A.; SHARMA, A.; MEHROTRA, R.; JEREMIAH, E. Representing low-frequency variability in continuous rainfall simulations: A hierarchical random Bartlett Lewis continuous rainfall generation model. **Water Resources Research**, v. 51, n. 12, p. 9995–10007, 2015.

WASKO, C.; SHARMA, A. Continuous rainfall generation for a warmer climate using observed temperature sensitivities. **Journal of Hydrology**, v. 544, p. 575–590, 2017.

WAYMIRE, E.; GUPTA, V. K. The mathematical structure of rainfall representations: 1. A review of the stochastic rainfall models. **Water Resources Research**, v. 17, n. 5, p. 1261–1272, 1981.

WHEATER, H. S.; CHANDLER, R. E.; ONOF, C. J.; ISHAM, V. S.; BELLONE, E.; YANG, C.; LEKKAS, D.; LOURMAS, G.; SEGOND, M.-L. Spatial-temporal rainfall modelling for flood risk estimation. **Stochastic Environmental Research and Risk Assessment**, v. 19, n. 6, p. 403–416, 2005.

WHEATER, H. S.; ISHAM, V. S.; CHANDLER, R. E.; ONOF, C. J.; STEWART, E. J. **Improved methods for national spatial-temporal rainfall and evaporation modelling for BSM**. Relatório Técnico. Joint Defra/EA Flood and Coastal Erosion Risk Management R&D Programme, 2006.

APÊNDICE A

Revisão do Processo de Poisson

O processo de Poisson consiste em um caso limite do processo binomial, que se desenvolve em uma escala contínua de tempo (NAVIDI, 2011). O processo binomial, por sua vez, representa o número de sucessos (S) de uma série de N experimentos binários independentes (testes de Bernoulli), com probabilidade de sucesso p e de falha $1 - p$, sendo caracterizado por uma distribuição discreta do tipo:

$$P(S = s) = \frac{N!}{s!(N - s)!} p^s (1 - p)^{N-s}, s = 0, 1, \dots, N. \quad (\text{A.1})$$

No caso de Poisson, admite-se que um intervalo contínuo de tempo $(0, T)$ possa ser discretizado em n subintervalos iguais de tamanho Δt , de modo que todos os subintervalos sejam submetidos a testes de Bernoulli. São feitas as seguintes considerações: (i) cada subintervalo é pequeno o bastante para comportar apenas uma ocorrência (sucesso) do evento em estudo; (ii) a probabilidade p de ocorrência de sucesso em cada subintervalo é constante; e (iii) o número médio de ocorrências em um intervalo de tempo arbitrário é proporcional ao comprimento de tal intervalo.

A partir da última consideração, é introduzida uma constante de proporcionalidade denotada por λ , com dimensão tempo⁻¹, que pode ser interpretada como a taxa segundo a qual os eventos ocorrem no tempo e que é denominada intensidade de Poisson (NAGHETTINI; PINTO, 2007).

Seja k o número total de ocorrências do evento em um intervalo arbitrário (t_1, t_2) , com $0 < t_1 < t_2 < T$, a intensidade de Poisson é estimada como $\lambda = k/(t_2 - t_1)$. A probabilidade de ocorrência do evento nesse intervalo é dada pela razão entre a frequência de subintervalos com ocorrência (sucessos) e o número total de subintervalos de forma que $p = k/n = \lambda(t_2 - t_1)/n$. Admitindo que essa probabilidade seja suficientemente pequena ($p \rightarrow 0$), o produto np converge para o parâmetro denotado por $\nu = \lambda(t_2 - t_1)$, que representa o número esperado de ocorrências de sucessos no intervalo (t_2, t_1) . Deve-se destacar que esse parâmetro é uma função do tempo, uma vez que a taxa $\lambda(t)$ pode variar ao longo do tempo. Dessa forma, o parâmetro ν é generalizado de acordo com:

$$v(t) = \int_0^t \lambda(s) ds. \quad (\text{A.2})$$

Se ao longo do intervalo $(0, T)$ o processo de Poisson é estacionário, λ e v são constantes, e o último parâmetro é dado pelo produto $v = \lambda T$. No caso dos processos não homogêneos, a integral deve ser computada para obter v .

Aplicando essas considerações, a distribuição binomial (Equação A.1) converge para a distribuição discreta de Poisson:

$$P(N(T) = k) = \frac{v^k}{k!} e^{-v}, \quad k = 0, 1, \dots \quad (\text{A.3})$$

Embora essa distribuição seja discreta, deve-se destacar que a formulação realizada de acordo com as considerações i, ii e iii implica que ela seja válida somente para um intervalo de tempo contínuo.

Os eventos de chuva, no entanto, não ocorrem de forma instantânea no tempo, pelo menos não em uma escala que permita validar a hipótese de independência entre subintervalos sucessivos. Os modelos estocásticos consideram então que apenas o instante de início dos eventos segue a distribuição de Poisson, de forma que o processo de contagem $N(t)$ refere-se somente à quantidade de origens de eventos de chuva no intervalo considerado (WAYMIRE; GUPTA, 1981).

Dando continuidade a essa análise, considera-se que a precipitação é representada por um processo estocástico estacionário, em que o número de eventos de chuva no intervalo de interesse $(0, T)$, denotado simplesmente por N , segue a distribuição de Poisson com intensidade λ . Tem-se a seguinte função de densidades acumuladas:

$$P[N = n] = \frac{(\lambda T)^n}{n!} e^{-\lambda T}. \quad (\text{A.4})$$

Considera-se ainda uma variável aleatória, denotada por ΔT_o , que representa o intervalo de tempo entre o instante de origem de um evento e o instante de origem do evento subsequente, ou seja, o “tempo entre eventos”. A ocorrência dessa variável aleatória está associada à condição em que $N = 0$ para $T = \Delta t_o$, ou seja, nenhum evento ocorra no tempo

equivalente ao tempo entre eventos. Essa curiosa constatação lógica é necessária para a seguinte formulação:

$$P[\Delta T_o > \Delta t_o] = P[N = 0 \mid T = \Delta t_o] = e^{-\lambda \Delta t_o}. \quad (\text{A.5})$$

A Equação A.5 representa a probabilidade de superação do tempo entre eventos. Por definição, a função acumulada de densidades (FAD) da variável aleatória ΔT_o é:

$$F_{\Delta T_o} = P[\Delta T_o \leq \Delta t_o] = 1 - P[\Delta T_o > \Delta t_o] = 1 - e^{-\lambda \Delta t_o}. \quad (\text{A.6})$$

A função densidade de probabilidades (FDP) do tempo entre eventos é obtida por derivação da Equação A.6:

$$f_{\Delta T_o}(\Delta t_o) = \frac{dF_{\Delta T_o}(\Delta t_o)}{d\Delta t_o} = \lambda e^{-\lambda \Delta t_o}, \quad (\text{A.7})$$

que equivale à FDP característica da distribuição exponencial. A Equação A.7 representa, portanto, que o tempo entre eventos consiste em uma variável aleatória distribuída exponencialmente com parâmetro de taxa λ .

APÊNDICE B

B1 – Programas desenvolvidos na etapa de aquisição

“gerar_series.cpp”

```

#include <iostream>
#include <algorithm>
#include <vector>
#include <string>
#include <fstream>
#include <sstream>
#include <cmath>
#include <cstdlib>
#include <cstdio>
#include <stdio.h>
#include <dirent.h>
#include "deletar_arquivos.h"
using namespace std;
using matriz2d = vector<vector<double>>;
int gerar_series(string, int, vector<string>);
int main()
{
    deletar_arquivos(".\\series_mensais"); //Limpa a pasta que contem as series de saida
    int ano, mes;
    vector<string> nome_meses =
{"Jan", "Fev", "Mar", "Abr", "Mai", "Jun", "Jul", "Ago", "Set", "Out", "Nov", "Dez", "Jan"};
    string ano_string, mes_string, ano_nome_arquivo;
    for (ano=1997; ano<=2017; ano++){
        ano_string = to_string(ano);
        ano_nome_arquivo=".\\dados_entrada\\diario";
        ano_nome_arquivo.append(ano_string);
        ano_nome_arquivo.append("_ok.csv"); //Aquisicao dos arquivos de entrada
        gerar_series(ano_nome_arquivo, ano, nome_meses);
        ano_string.clear();
    }
}
int gerar_series(string ano_nome_arquivo, int ano, vector<string> nome_meses)
{
    //Processa sucessivamente cada ano de forma individual, separando os registros de cada mes em uma serie
    mensal continua
    ifstream ano_arquivo;
    ano_arquivo.open(ano_nome_arquivo, ios::in);

    if(ano_arquivo.fail()!=1)
        cout << "Processando ano " << ano << endl;
    else
        cout << "Falha ao abrir o ano " << ano << endl;

    vector<int> bissextos={2000, 2004, 2008, 2012, 2016, 2020, 2024, 2028, 2032};
    vector<int> dias_julianos;
    if(find(bissextos.begin(), bissextos.end(), ano) != bissextos.end()) //Para anos bissextos
        dias_julianos={0,31,60,91,121,152,182,213,244,274,305,335,366};
    else //Para anos normais
        dias_julianos={0,31,59,90,120,151,181,212,243,273,304,334,365};
    ofstream mes_arquivo;
    string linha, dia_juliano_string, hora_esalq_string, lamina_string, mes;

```

```

int dia_juliano, dia;
int hora_esalq; //No formato da ESALQ: 13:45 - 1345
double hora; //No formato desejado: 13:45 = 13.75
double lamina;
int ponteiro_mes=0; //Vai de 0 a 11 (Janeiro a Dezembro)
string mes_nome_arquivo=".\\series_mensais\\";
mes_nome_arquivo.append(nome_meses.at(ponteiro_mes));
mes_nome_arquivo.append("_15min.csv");
mes_arquivo.open(mes_nome_arquivo,ios::app);
while(getline(ano_arquivo,linha)){
    stringstream ss(linha);
        getline(ss,dia_juliano_string,',');
            dia_juliano=stoi(dia_juliano_string);
        getline(ss,hora_esalq_string,',');
            hora_esalq=stoi(hora_esalq_string);
            hora=(int)hora_esalq/100;
            hora=double(hora)+(hora_esalq % 100)/60.0;
        getline(ss,lamina_string,',');
            lamina=stod(lamina_string);
        dia=dia_juliano-dias_julianos.at(ponteiro_mes);
        mes=nome_meses.at(ponteiro_mes);
        if(dia_juliano>dias_julianos.at(ponteiro_mes+1))
        { //Condicao satisfeita sempre que o mes termina
            dia=1;
            mes=nome_meses.at(ponteiro_mes+1);
        }
        if(dia_juliano==1 && ponteiro_mes==11)
        { //Condicao satisfeita sempre que o ano termina (meses de Dezembro)
            dia=1;
            mes=nome_meses.at(ponteiro_mes+1);
            ano++;
        }
        //Insere os dados na serie do mes em questao
        mes_arquivo << ano << "," << dia_juliano << "," << hora_esalq << "," << dia << "," <<
endl;

        mes_arquivo << mes << "," << hora << "," << lamina << "," << endl;
        if(dia_juliano>dias_julianos.at(ponteiro_mes+1)){
            mes_arquivo.close();
            ponteiro_mes++;
            mes_nome_arquivo=".\\series_mensais\\";
            mes_nome_arquivo.append(nome_meses.at(ponteiro_mes));
            mes_nome_arquivo.append("_15min.csv");
            mes_arquivo.open(mes_nome_arquivo,ios::app);
        }
        ss.clear();
    }
    mes_arquivo.close();
    ano_arquivo.close();
    return 0;
}

```

“agregar_series.cpp”

```

#include <iostream>
#include <algorithm>
#include <vector>
#include <string>
#include <fstream>
#include <sstream>
#include <cmath>
#include <cstdlib>
#include <cstdio>
#include <stdio.h>
#include <dirent.h>
#include "deletar_arquivos.h"
using namespace std;
using matriz2d = vector<vector<double>>;
int agregar_series(string, vector<double>);
int main()
{
    deletar_arquivos(".\\series_agregadas"); //Limpa a pasta que contem as series de saida
    int mes;
    vector<string> nome_meses = {"Jan", "Fev", "Mar", "Abr", "Maio", "Jun", "Jul", "Ago", "Set", "Out",
"Nov", "Dez", "Jan"};
    string mes_string;
    vector<double> escalas={1.0, 2.0, 6.0, 12.0, 24.0}; //Definir as escalas em horas (nao incluir a da
resolucao);
    for(mes=1;mes<=12;mes++){
        mes_string = nome_meses.at(mes-1);
        agregar_series(mes_string, escalas);
        mes_string.clear();
    }
}
int agregar_series(string mes_string, vector<double> escalas)
{ //Processa as series mensais geradas na funcao anterior, agregando os dados em escalas pre-definidas
    double resolucao=0.25; //As escalas devem ser multiplos de 0.25 hora - o que e passivel de ser alterado,
caso a aquisicao seja realizada em outra escala
    string mes_nome_arquivo = ".\\dados_entrada\\";
    mes_nome_arquivo.append(mes_string);
    mes_nome_arquivo.append("_15min.csv");
    //Campos a serem coletados do registro com a resolucao
    string linha, desprezar, ano_string, dia_juliano_string, hora_esalq_string,
    double horario, lamina;
    double horario_acumulado=0.0, tempo_acumulado=0.0, lamina_acumulada=0.0;
    ifstream mes_arquivo;
    mes_arquivo.open(mes_nome_arquivo,ios::in);
    if(mes_arquivo.fail()!=1)
        cout << "Processando a serie de " << mes_string << endl;
    else{
        cout << "Falha ao abrir serie de " << mes_string << endl;
        exit(1);
    }
}
string mes_agregado_nome_arquivo;
ofstream mes_agregado_arquivo;
for(int i=0; i<escalas.size();i++){
    cout << "Agregando para " << escalas.at(i) << " hora(s)" << endl;
    mes_agregado_nome_arquivo = ".\\series_agregadas\\";
    mes_agregado_nome_arquivo.append(mes_string);
    mes_agregado_nome_arquivo.append("_");
    mes_agregado_nome_arquivo.append(to_string(int(escalas.at(i))));
    mes_agregado_nome_arquivo.append("_hr.csv");
    mes_agregado_arquivo.open(mes_agregado_nome_arquivo,ios::out);
}
}

```

```

while(getline(mes_arquivo, linha)){
    stringstream ss(linha);
    getline(ss,ano_string,',');
    getline(ss,dia_juliano_string,',');
    getline(ss,hora_esalq_string,',');
    getline(ss,dia_string,',');
    getline(ss,mes_string2,',');
    getline(ss,hora_string,',');
    getline(ss,lamina_string);
        horario=stod(hora_string);
        lamina=stod(lamina_string);
        if(horario-horario_acumulado==resolucao){ //Verifica se ha erros no registro
            horario_acumulado=horario;
        }
        else if(horario-horario_acumulado==(resolucao-24.0)){
            horario_acumulado=0.0;
        }
        else{ //Termina a execucao caso verifique alguma inconsistencia
            cout << "Erro no registro!" << endl;
            cout << "Ano = " << ano_string << endl;
            cout << "Mes = " << mes_string2 << endl;
            cout << "Dia = " << dia_string << endl;
            cout << "Hora = " << hora_string << endl;
            exit(1);
        }
        tempo_acumulado=tempo_acumulado+resolucao;
        lamina_acumulada=lamina_acumulada+lamina;
        if(tempo_acumulado==escalas.at(i)){
            mes_agregado_arquivo << ano_string << "," << dia_juliano_string <<
            "," << hora_esalq_string << "," << dia_string << "," << mes_string2 << "," << hora_string << "," <<
            lamina_acumulada << endl;
            tempo_acumulado=0.0;
            lamina_acumulada=0.0;
        }
        ss.clear();
    }
    mes_arquivo.clear();
    mes_arquivo.seekg(0,ios::beg);
    mes_agregado_arquivo.close();
}
return 0;
}

```

“*calcular_estatisticas.cpp*”

```

#include <iostream>
#include <algorithm>
#include <vector>
#include <string>
#include <fstream>
#include <sstream>
#include <cmath>
#include <cstdlib>
#include <cstdio>
#include <stdio.h>
#include <dirent.h>
#include "deletar_arquivos.h"
using namespace std;
int calcular_estatisticas(string, string, string);
int organizar_estatisticas();
int main()
{
    deletar_arquivos(".\\estatisticas");
    vector<string> meses = {"Jan", "Fev", "Mar", "Abr", "Maio", "Jun", "Jul", "Ago", "Set", "Out", "Nov",
"Dez"};
    vector<string> escalas={"15min", "1_hr", "2_hr", "6_hr", "12_hr", "24_hr"};
    string caminho = ".\\dados_entrada\\";
    string arquivo;
    int i, j;
    for(i=0;i<12;i++)
    {
        cout << "Processando mes " << meses.at(i) << endl;
        for(j=0;j<escalas.size();j++){
            cout << "Escala de " << escalas.at(j) << endl;
            arquivo = caminho + meses.at(i) + "_" + escalas.at(j) + ".csv";
            calcular_estatisticas(arquivo, meses.at(i), escalas.at(j));
        }
    }
    organizar_estatisticas();
    return 0;
}
int calcular_estatisticas(string serie_nome_arquivo, string mes, string escala)
{
    //Para cada serie de um determinado mes, em uma determinada escala, imprime media, variancia, momento de
ordem3, autocovariancia de lag 1, 2, 3, PPS e PTuu – anteriormente, os termo usados para a PPS e a PTuu era
em inglês: zdp (zero depth probability) e pww (probability wet-wet)
    ofstream estatisticas_arquivo;
    estatisticas_arquivo.open(".\\estatisticas\\estatisticas_observadas.csv",ios::app);
    ifstream serie_arquivo;
    serie_arquivo.open(serie_nome_arquivo,ios::in);
    int i, k, N;
    vector<double> laminas;
    double media, variancia, momento_ordem3, autocovar1, autocovar2, autocovar3, zdp, pww;
    double acumulado=0, acumulado2=0, acumulado3=0;
    int sem_chuva=0, com_chuva=0, com_chuva_com_chuva=0;
    string linha, desprezar, lamina_string;
    while(getline(serie_arquivo, linha))
    {
        stringstream ss(linha);
        getline(ss,desprezar,',');
        getline(ss,desprezar,',');
        getline(ss,desprezar,',');
        getline(ss,desprezar,',');
        getline(ss,desprezar,',');
    }
}

```

```

        getline(ss,desprezar,',');
        getline(ss,lamina_string,',');
        laminas.push_back(stod(lamina_string));
    }
    N=laminas.size();
    for(i=0;i<N;i++){
        acumulado=acumulado+laminas.at(i);
        if(laminas.at(i)<0.1)
            sem_chuva++;
    }
    media=acumulado/N;
    zdp=double(sem_chuva)/N;
    for(i=0;i<N-1;i++){
        if(laminas.at(i)>=0.1)
        {
            com_chuva++; //Pode ser diferente de N - sem_chuva, dependendo do ultimo intervalo da
serie
            if(laminas.at(i+1)>=0.1)
                com_chuva_com_chuva++;
        }
    }
    pww = double(com_chuva_com_chuva)/com_chuva;
    for(i=0;i<N;i++){
        acumulado2=acumulado2+pow(laminas.at(i)-media,2);
        acumulado3=acumulado3+pow(laminas.at(i)-media,3);
    }
    variancia=acumulado2/(N-1.0);
    momento_ordem3=acumulado3*N/((N-1.0)*(N-2.0));
    acumulado=0.0;
    k=1;
    for(i=0;i<N-k;i++){
        acumulado=acumulado+(laminas.at(i)-media)*(laminas.at(i+k)-media);
    }
    autocovar1=acumulado/(N-1);
    acumulado=0.0;
    k=2;
    for(i=0;i<N-k;i++){
        acumulado=acumulado+(laminas.at(i)-media)*(laminas.at(i+k)-media);
    }
    autocovar2=acumulado/(N-1);
    acumulado=0.0;
    k=3;
    for(i=0;i<N-k;i++){
        acumulado=acumulado+(laminas.at(i)-media)*(laminas.at(i+k)-media);
    }
    autocovar3=acumulado/(N-1);

    estatisticas_arquivo << mes+"_"+"media"+"_"+"escala << " << media << endl;
    estatisticas_arquivo << mes+"_"+"variancia"+"_"+"escala << " << variancia << endl;
    estatisticas_arquivo << mes+"_"+"momento_ordem3"+"_"+"escala << " << momento_ordem3 << endl;
    estatisticas_arquivo << mes+"_"+"autocovar1"+"_"+"escala << " << autocovar1 << endl;
    estatisticas_arquivo << mes+"_"+"autocovar2"+"_"+"escala << " << autocovar2 << endl;
    estatisticas_arquivo << mes+"_"+"autocovar3"+"_"+"escala << " << autocovar3 << endl;
    estatisticas_arquivo << mes+"_"+"zdp"+"_"+"escala << " << zdp << endl;
    estatisticas_arquivo << mes+"_"+"pww"+"_"+"escala << " << pww << endl;
    serie_arquivo.close();
    estatisticas_arquivo.close();
    return 0;
}

```

```

int organizar_estatisticas()
{
//Organiza os resultados produzidos na funcao anterior, facilitando a visualizacao das propriedades estatisticas
    ifstream estatisticas_arquivo;
    estatisticas_arquivo.open("..\estatisticas\\estatisticas_observadas.csv",ios::in);
    string desprezar;
    string dado;
    string linha;
    int mes=1, p=1, e=1;
    int indice;
    vector<vector<double>> estatisticas(12, vector<double>(8*6)); //8 propriedades em 6 escalas
    while(getline(estatisticas_arquivo,linha)){
        stringstream ss(linha);
        getline(ss,desprezar,',');
        getline(ss,dado,',');
        indice=(p-1)*6+e;
        estatisticas.at(mes-1).at(indice-1)=stod(dado);
        p++;
        if(p>8 && e==6)
        {
            p=1;
            e=1;
            mes++;
        }
        if(p>8)
        {
            p=1;
            e++;
        }
    }
    estatisticas_arquivo.close();
    ofstream estatisticas_observadas_ORGANIZADO;
    estatisticas_observadas_ORGANIZADO.open("..\estatisticas\\estatisticas_observadas_ORGANIZADO.csv",ios::out);
    for(auto i:estatisticas){
        for(auto j:i){
            estatisticas_observadas_ORGANIZADO << j << ",";
        }
        estatisticas_observadas_ORGANIZADO << endl;
    }
    return 0;
}

```

“pesos.cpp”

```

#include <fstream>
#include <sstream>
#include <vector>
#include <iostream>
#include <iomanip>
#include <cmath>
#include <string>
#include <algorithm>
#include <cstdlib>
#include <cstdio>
#include <stdio.h>
#include "deletar_arquivos.h"
using namespace std;
vector<double> estatisticas(vector<double>);
int calcular_pesos(string);
int organizar_relatorio(string);
int main()
{
    deletar_arquivos(".\\pesos");
    vector<string> meses = {"Jan", "Fev", "Mar", "Abr", "Maio", "Jun", "Jul", "Ago", "Set", "Out", "Nov",
"Dez"};
    for(int i=0;i<12;i++)
    {
        cout << "Processando o mes de " << meses.at(i) << endl;
        calcular_pesos(meses.at(i));
    }
    return 0;
}
int calcular_pesos(string mes){
    ofstream relatorio_pesos;
    ifstream arquivo;
    vector<double> resultados(4); //Para cada ano, em uma determinada escala
    string relatorio_arquivo = ".\\pesos\\" + mes + "_relatorio.csv";
    relatorio_pesos.open(relatorio_arquivo,ios::app);
    vector<string> nome_arquivo = {"_15min.csv", "_1_hr.csv", "_6_hr.csv", "_12_hr.csv", "_24_hr.csv"};
    for(int i=0;i<5;i++)
    {
        string linha, desprezar, ano_string, dia_string, lamina_string;
        int ano, dia1, dia2=0;
        double lamina;
        vector<double> laminas;
        arquivo.open(".\\dados_entrada\\" + mes + nome_arquivo.at(i));
        while(getline(arquivo, linha))
        {
            stringstream ss(linha);
            getline(ss,ano_string,',');
            ano=stoi(ano_string);
            getline(ss,desprezar,',');
            getline(ss,desprezar,',');
            getline(ss,dia_string,',');
            dia1=stoi(dia_string);
            getline(ss,desprezar,',');
            getline(ss,desprezar,',');
            getline(ss,lamina_string,',');
            lamina=stod(lamina_string);
            laminas.push_back(lamina);
            if(dia1<dia2)
            {
                if(mes!="Dez")

```

```

        relatorio_pesos << ano << ",";
    else if(mes=="Dez")
        relatorio_pesos << ano-1 << ",";
    resultados=estatisticas(laminas);
    for(auto i:resultados)
        relatorio_pesos << setprecision(10) << i << ",";
    relatorio_pesos << endl;
    laminas.clear();
    dia2=0;
}
else
    dia2=dia1;
}
arquivo.close();
}
organizar_relatorio(relatorio_arquivo);
relatorio_pesos.close();
return 0;
}
vector<double> estatisticas(vector<double> laminas) //Produz 4 estatisticas (media, coeficiente de variacao,
autocorrelacao de lag 1 e coeficiente de assimetria) para cada ano em uma determinada escala
{
    double acumulado=0.0, acumulado2=0.0, acumulado3=0.0;
    double media, variancia, autocovar_1, momento_ordem_3;
    double coef_var, autocorr_1, coef_ass;
    int i, k=1;
    int N=laminas.size();
    for(i=0;i<N;i++){
        acumulado=acumulado+laminas.at(i);
    }
    media=acumulado/N;
    for(i=0;i<N;i++){
        acumulado2=acumulado2+pow(laminas.at(i)-media,2);
        acumulado3=acumulado3+pow(laminas.at(i)-media,3);
    }
    variancia=acumulado2/(N-1.0);
    momento_ordem_3=acumulado3*N/((N-1.0)*(N-2.0));
    acumulado=0.0;
    for(i=0;i<N-k;i++){
        acumulado=acumulado+(laminas.at(i)-media)*(laminas.at(i+k)-media);
    }
    autocovar_1=acumulado/(N-1);
    coef_var=pow(variancia,1.0/2)/media;
    autocorr_1=autocovar_1/variancia;
    coef_ass=momento_ordem_3/pow(variancia,3.0/2);
    return {media,coef_var,autocorr_1,coef_ass};
}
int organizar_relatorio(string relatorio_arquivo)
{
    ifstream relatorio_pesos;
    relatorio_pesos.open(relatorio_arquivo,ios::in);
    vector<vector<double>> matriz2d;

    string linha, ano_string, media_string, coef_var_string, autocorr_1_string, coef_ass_string;
    int ano1, ano2=0;
    vector<double> anos, medias, coefs_var, autocorrs_1, coefs_ass;

    while(getline(relatorio_pesos, linha))
    {
        stringstream ss(linha);

```

```

getline(ss,ano_string,');
ano1=stoi(ano_string);
if(ano1<ano2)
{
    matriz2d.push_back(anos);
    matriz2d.push_back(medias);
    matriz2d.push_back(coefs_var);
    matriz2d.push_back(autocorr_1);
    matriz2d.push_back(coefs_ass);
    anos.clear();
    medias.clear();
    coefs_var.clear();
    autocorr_1.clear();
    coefs_ass.clear();
    ano2=0;
}
else
    ano2=ano1;
anos.push_back(stod(ano_string));
getline(ss,media_string,');
medias.push_back(stod(media_string));
getline(ss,coef_var_string,');
coefs_var.push_back(stod(coef_var_string));
getline(ss,autocorr_1_string,');
autocorr_1.push_back(stod(autocorr_1_string));
getline(ss,coef_ass_string,');
coefs_ass.push_back(stod(coef_ass_string));
}
matriz2d.push_back(anos);
matriz2d.push_back(medias);
matriz2d.push_back(coefs_var);
matriz2d.push_back(autocorr_1);
matriz2d.push_back(coefs_ass);
ofstream relatorio_pesos_organizado;
relatorio_pesos_organizado.open(".\\pesos\\relatorio_definitivo.csv",ios::app);
int i, j, m, n;
n=matriz2d.size();
m=matriz2d.at(0).size();
vector<vector<double>> matriz2d_ajustada(m, vector<double>(n));
for(i=0;i<n;i++)
{
    for(j=0;j<m;j++)
        matriz2d_ajustada.at(j).at(i)=matriz2d.at(i).at(j);
}
for(i=0;i<m;i++)
{
    for(j=0;j<n;j++)
    {
        if(j==0 || j%5!=0)
            relatorio_pesos_organizado << matriz2d_ajustada.at(i).at(j) << ",";
    }
    relatorio_pesos_organizado << endl;
}
relatorio_pesos_organizado << "\n\n";
relatorio_pesos_organizado.close();
return 0;
}

```

B2 – programas e rotinas desenvolvidos na etapa de otimização

“f_objetivo.h” (BLRP)

```
//Sera computada f(x) = f_objetivo(x) para o modelo original
#include <vector>
#include <cmath>
#include <iostream>
#include <fstream>
#include <sstream>
#pragma once
using namespace std;
double c_variacao(double, double, double, double, double, double);
double autocorrelacao(int, double, double, double, double, double, double);
double c_assimetria(double, double, double, double, double, double);
double f_objetivo(int mes, int& contador, vector<double> x){
    double lambda=x.at(0);
    double beta=x.at(1);
    double gama=x.at(2);
    double eta=x.at(3);
    double mux=x.at(4);
    double muc=1+beta/gama;
    //Aquisicao dos momentos observados no arquivo M_obs.txt (dados de entrada)
    ifstream matriz_M_obs;
    matriz_M_obs.open("..\entradas\M_obs.txt");
    vector<vector<double>> M_obs;
    vector<double> m_obs;
    string linha, dado_string;
    while(getline(matriz_M_obs,linha))
    {
        stringstream ss(linha);
        for(int i=0;i<13;i++)
        {
            getline(ss,dado_string,',');
            m_obs.push_back(stod(dado_string));
        }
        M_obs.push_back(m_obs);
        m_obs.clear();
    }
    //Aquisicao dos pesos de calibracao no arquivo Wi.txt (dados de entrada)
    ifstream matriz_Wi;
    matriz_Wi.open("..\entradas\W.txt");
    vector<vector<double>> Wi;
    vector<double> wi;
    while(getline(matriz_Wi,linha))
    {
        stringstream ss(linha);
        for(int i=0;i<13;i++)
        {
            getline(ss,dado_string,',');
            wi.push_back(stod(dado_string));
        }
        Wi.push_back(wi);
        wi.clear();
    }
    //Momentos calculados pelo modelo
    vector<double> M_calc;
    M_calc.push_back(1*lambda*mux*muc/eta); //Media de 1 hora
    M_calc.push_back(c_variacao(15.0/60,lambda,mux,beta,gama,eta)); //15 minutos
```

```

M_calc.push_back(autocorrelacao(1,15.0/60,lambda,mux,beta,gama,eta));
M_calc.push_back(c_assimetria(15.0/60,lambda,mux,beta,gama,eta));
M_calc.push_back(c_variacao(1,lambda,mux,beta,gama,eta)); //1 hora
M_calc.push_back(autocorrelacao(1,1,lambda,mux,beta,gama,eta));
M_calc.push_back(c_assimetria(1,lambda,mux,beta,gama,eta));
M_calc.push_back(c_variacao(6,lambda,mux,beta,gama,eta)); //6 horas
M_calc.push_back(autocorrelacao(1,6,lambda,mux,beta,gama,eta));
M_calc.push_back(c_assimetria(6,lambda,mux,beta,gama,eta));
M_calc.push_back(c_variacao(24,lambda,mux,beta,gama,eta)); //24 horas
M_calc.push_back(autocorrelacao(1,24,lambda,mux,beta,gama,eta));
M_calc.push_back(c_assimetria(24,lambda,mux,beta,gama,eta));
//Atribuicao da funcao objetivo
double f_obj=0;
for(int i=0;i<M_calc.size();i++)
{
    f_obj=f_obj+Wi.at(mes-1).at(i)*pow(M_calc.at(i)-M_obs.at(mes-1).at(i),2);
}
contador++;
return f_obj;
}

double c_variacao(double h,double lambda,double mux,double beta,double gama,double eta) //Coeficiente de
variacao
{
    double muc=1+beta/gama;
    double media=h*lambda*mux*muc/eta;
    double variancia=(2*pow(mux,2)+beta*pow(mux,2)/gama)*h*1/eta+pow(mux,2)*beta*eta*(1-exp(-
gama*h))/(pow(gama,2)*(pow(gama,2)-pow(eta,2)))
    -(2*pow(mux,2)+beta*gama*pow(mux,2)/(pow(gama,2)-pow(eta,2)))*(1-exp(-
eta*h))/pow(eta,2);
    variancia=2*lambda*muc*1/eta*variancia;
    return sqrt(variancia)/media;
}

double autocorrelacao(int k,double h,double lambda,double mux,double beta,double gama,double eta)
//Autocorrelacao de lag k
{
    double muc=1+beta/gama;
    double variancia=(2*pow(mux,2)+beta*pow(mux,2)/gama)*h*1/eta+pow(mux,2)*beta*eta*(1-exp(-
gama*h))/(pow(gama,2)*(pow(gama,2)-pow(eta,2)))
    -(2*pow(mux,2)+beta*gama*pow(mux,2)/(pow(gama,2)-pow(eta,2)))*(1-exp(-
eta*h))/pow(eta,2);
    variancia=2*lambda*muc*1/eta*variancia;
    double autocovariancia_lag1=(2*pow(mux,2)+beta*gama*pow(mux,2)/(pow(gama,2)-
pow(eta,2)))*(pow(1-exp(-eta*h),2)*exp(-eta*(k-1)/h)/pow(eta,2))
    -pow(mux,2)*beta*eta*pow(1-exp(-gama*h),2)*exp(-gama*(k-
1)*h)/(pow(gama,2)*(pow(gama,2)-pow(eta,2)));
    autocovariancia_lag1=lambda*muc*1/eta*autocovariancia_lag1;
    return autocovariancia_lag1/variancia;
}

double c_assimetria(double h,double lambda,double mux,double beta,double gama,double eta)
{
    double muc=1+beta/gama;
    double variancia=(2*pow(mux,2)+beta*pow(mux,2)/gama)*h*1/eta+pow(mux,2)*beta*eta*(1-exp(-
gama*h))/(pow(gama,2)*(pow(gama,2)-pow(eta,2)))
    -(2*pow(mux,2)+beta*gama*pow(mux,2)/(pow(gama,2)-pow(eta,2)))*(1-exp(-
eta*h))/pow(eta,2);
    variancia=2*lambda*muc*1/eta*variancia;
    double momento_ordem_3=

```



```

-36*pow(eta,3)*pow(gama,7)*h*6*pow(mux,3)*exp(-eta*h)-
21*pow(gama,8)*2*pow(mux,2)*mux*beta+6*pow(gama,9)*h*6*pow(mux,3)*eta+12*pow(gama,3)*pow(mu
x,3)*exp(-eta*h)*pow(beta,2)*pow(eta,4)
+12*pow(gama,3)*pow(mux,3)*exp(-gama*h)*pow(beta,2)*pow(eta,4)-
18*pow(gama,4)*pow(mux,3)*exp(-gama*h)*pow(beta,2)*pow(eta,3)-
12*pow(gama,3)*pow(mux,3)*pow(beta,2)*pow(eta,4)-108*pow(gama,5)*6*pow(mux,3)*pow(eta,4)

+108*mux*2*pow(mux,2)*pow(gama,4)*h*beta*pow(eta,5)+12*pow(gama,9)*6*pow(mux,3)*exp(-eta*h);
momento_ordem_3=momento_ordem_3*lambda*muc*1/((pow(eta,2)+2*gama*eta+pow(gama,2))*(pow
(gama,4)-2*eta*pow(gama,3)-3*pow(eta,2)*pow(gama,2)+8*pow(eta,3)*gama-
4*pow(eta,4))*pow(eta,4)*pow(gama,3));
return momento_ordem_3/pow(variancia,1.5);
}

```

“f_objetivo.h” (BLRPRx)

```

//Sera computada f(x) = f_objetivo(x) para o modelo BLRPRx

#include <vector>
#include <cmath>
#include <iostream>
#include <fstream>
#include <sstream>
#pragma once
using namespace std;
double funcao_gamma(double z);
double E_eta(int, double, double, double);
double c_variacao(double, double, double, double, double, double, double);
double autocorrelacao(int, double, double, double, double, double, double, double);
double c_assimetria(double, double, double, double, double, double, double, double);
double f_objetivo(int mes, int& contador, vector<double> x){
    double lambda=x.at(0);
    double iota=x.at(1);
    double alpha=x.at(2);
    double nu=x.at(3);
    double kappa=x.at(4);
    double phi=x.at(5);
    double muc=1+kappa/phi;
    double f1=2.0;
    double f2=6.0;
    //Aquisicao dos momentos observados no arquivo M_obs.txt (dados de entrada)
    ifstream matriz_M_obs;
    matriz_M_obs.open("..\entradas\M_obs.txt");
    vector<vector<double>> M_obs;
    vector<double> m_obs;
    string linha, dado_string;
    while(getline(matriz_M_obs,linha))
    {
        stringstream ss(linha);
        for(int i=0;i<13;i++)
        {
            getline(ss,dado_string,',');
            m_obs.push_back(stod(dado_string));
        }
        M_obs.push_back(m_obs);
        m_obs.clear();
    }
    //Aquisicao dos pesos de calibracao no arquivo Wi.txt (dados de entrada)
    ifstream matriz_Wi;
    matriz_Wi.open("..\entradas\W.txt");
    vector<vector<double>> Wi;
    vector<double> wi;
    while(getline(matriz_Wi,linha))
    {
        stringstream ss(linha);
        for(int i=0;i<13;i++)
        {
            getline(ss,dado_string,',');
            wi.push_back(stod(dado_string));
        }
        Wi.push_back(wi);
        wi.clear();
    }
    //Momentos calculados pelo modelo
    vector<double> M_calc;

```

```

M_calc.push_back(1*lambda*iota*muc); //Media de 1 hora
M_calc.push_back(c_variacao(15.0/60,lambda,iota,alpha,nu,kappa,phi)); //15 minutos
M_calc.push_back(autocorrelacao(1,15.0/60,lambda,iota,alpha,nu,kappa,phi));
M_calc.push_back(c_assimetria(15.0/60,lambda,iota,alpha,nu,kappa,phi));
M_calc.push_back(c_variacao(1,lambda,iota,alpha,nu,kappa,phi)); //1 hora
M_calc.push_back(autocorrelacao(1,1,lambda,iota,alpha,nu,kappa,phi));
M_calc.push_back(c_assimetria(1,lambda,iota,alpha,nu,kappa,phi));
M_calc.push_back(c_variacao(6,lambda,iota,alpha,nu,kappa,phi)); //6 horas
M_calc.push_back(autocorrelacao(1,6,lambda,iota,alpha,nu,kappa,phi));
M_calc.push_back(c_assimetria(6,lambda,iota,alpha,nu,kappa,phi));
M_calc.push_back(c_variacao(24,lambda,iota,alpha,nu,kappa,phi)); //24 horas
M_calc.push_back(autocorrelacao(1,24,lambda,iota,alpha,nu,kappa,phi));
M_calc.push_back(c_assimetria(24,lambda,iota,alpha,nu,kappa,phi));
//Atribuicao da funcao objetivo
double f_obj=0;
for(int i=0;i<M_calc.size();i++)
{
    f_obj=f_obj+Wi.at(mes-1).at(i)*pow(M_calc.at(i)-M_obs.at(mes-1).at(i),2);
}
contador++;
return f_obj;
}
double c_variacao(double h,double lambda,double iota,double alpha,double nu,double kappa,double phi)
//Coeficiente de variacao
{
    double muc=1+kappa/phi;
    double f1=2.0;
    double f2=6.0;
    double media=h*lambda*iota*muc;
    double
        variancia=(f1+kappa/phi)*h+E_eta(1,0,nu,alpha)*(kappa/pow(phi,2))*(1-
pow(phi,3))/(pow(phi,2)-1)-f1
        -E_eta(1,phi*h,nu,alpha)*kappa/(pow(phi,2)*(pow(phi,2)-
1))+E_eta(1,h,nu,alpha)*(f1+kappa*phi/(pow(phi,2)-1));
    variancia=2*lambda*muc*pow(iota,2)*variancia;
    return sqrt(variancia)/media;
}

double autocorrelacao(int k,double h,double lambda,double iota,double alpha,double nu,double kappa,double
phi) //Autocorrelacao de lag k
{
    double muc=1+kappa/phi;
    double f1=2.0;
    double f2=6.0;
    double media=h*lambda*iota*muc;
    double
        variancia=(f1+kappa/phi)*h+E_eta(1,0,nu,alpha)*(kappa/pow(phi,2))*(1-
pow(phi,3))/(pow(phi,2)-1)-f1
        -E_eta(1,phi*h,nu,alpha)*kappa/(pow(phi,2)*(pow(phi,2)-
1))+E_eta(1,h,nu,alpha)*(f1+kappa*phi/(pow(phi,2)-1));
    variancia=2*lambda*muc*pow(iota,2)*variancia;
    double
        autocovariancia_lag1=(f1+kappa*phi/(pow(phi,2)-1))*(E_eta(1,(k-1)*h,nu,alpha)-
2*E_eta(1,k*h,nu,alpha)+E_eta(1,(k+1)*h,nu,alpha))
        -kappa/(pow(phi,2)*(pow(phi,2)-1))*(E_eta(1,phi*(k-
1)*h,nu,alpha)-2*E_eta(1,phi*k*h,nu,alpha)+E_eta(1,phi*(k+1)*h,nu,alpha));
    autocovariancia_lag1=lambda*muc*pow(iota,2)*autocovariancia_lag1;
    return autocovariancia_lag1/variancia;
}
double c_assimetria(double h,double lambda,double iota,double alpha,double nu,double kappa,double phi)
//Coeficiente de assimetria
{
    double muc=1+kappa/phi;

```

```

double f1=2.0;
double f2=6.0;
double media=h*lambda*iota*muc;
double
    variancia=(f1+kappa/phi)*h+E_eta(1,0,nu,alpha)*(kappa/pow(phi,2)*(1-
pow(phi,3))/(pow(phi,2)-1)-f1)
    -E_eta(1,phi*h,nu,alpha)*kappa/(pow(phi,2)*(pow(phi,2)-
1))+E_eta(1,h,nu,alpha)*(f1+kappa*phi/(pow(phi,2)-1));
variancia=2*lambda*muc*pow(iota,2)*variancia;
double
    momento_ordem_3=E_eta(1,h,nu,alpha)*(12*pow(phi,7)*pow(kappa,2)-
24*f1*pow(phi,2)*kappa-18*pow(phi,4)*pow(kappa,2)+24*f1*pow(phi,3)*kappa-
132*f1*pow(phi,6)*kappa+150*f1*pow(phi,4)*kappa
    -42*pow(phi,5)*pow(kappa,2)-
6*f1*pow(phi,5)*kappa+108*pow(phi,5)*f2-72*pow(phi,7)*f2-
48*pow(phi,3)*f2+24*f1*pow(phi,8)*kappa+12*pow(phi,3)*pow(kappa,2)+12*pow(phi,9)*f2)
    +E_eta(0,h,nu,alpha)*(24*f1*pow(phi,4)*h*kappa+6*pow(phi,9)*h*f2-
30*f1*pow(phi,6)*h*kappa+6*f1*pow(phi,8)*h*kappa+54*pow(phi,5)*h*f2-24*h*f2*pow(phi,3)-
36*pow(phi,7)*h*f2)
    +E_eta(1,phi*h,nu,alpha)*(-48*pow(kappa,2)+6*f1*pow(phi,4)*kappa-
48*phi*f1*kappa+6*pow(phi,5)*pow(kappa,2)-24*f1*pow(phi,2)*kappa+36*f1*pow(phi,3)*kappa-
6*f1*pow(phi,5)*kappa+84*pow(phi,2)*pow(kappa,2)+12*pow(phi,3)*pow(kappa,2)-
18*pow(phi,4)*pow(kappa,2))
    +E_eta(0,phi*h,nu,alpha)*(-
24*phi*h*pow(kappa,2)+30*pow(phi,3)*h*pow(kappa,2)-6*pow(phi,5)*h*pow(kappa,2))

    +E_eta(1,0,nu,alpha)*(72*pow(phi,7)*f2+48*phi*f1*kappa+24*f1*pow(phi,2)*kappa-
36*f1*pow(phi,3)*kappa-84*pow(phi,2)*pow(kappa,2)+6*f1*pow(phi,5)*kappa+117*f1*pow(phi,6)*kappa
    +39*pow(phi,5)*pow(kappa,2)-12*pow(phi,9)*f2-
138*f1*pow(phi,4)*kappa+48*pow(kappa,2)-
9*pow(phi,7)*pow(kappa,2)+48*pow(phi,3)*f2+18*pow(phi,4)*pow(kappa,2)-21*pow(phi,8)*f1*kappa-
12*pow(phi,3)*pow(kappa,2)-108*pow(phi,5)*f2)
    -24*phi*h*pow(kappa,2)-72*f1*pow(phi,6)*h*kappa-
36*pow(phi,5)*h*pow(kappa,2)+54*pow(phi,3)*h*pow(kappa,2)+6*pow(phi,7)*h*pow(kappa,2)+54*pow(phi
,5)*h*f2-36*pow(phi,7)*h*f2
    -24*pow(phi,3)*h*f2-
48*f1*pow(phi,2)*h*kappa+12*f1*pow(phi,8)*h*kappa+6*pow(phi,9)*h*f2+108*f1*pow(phi,4)*h*kappa
    +E_eta(1,2*h,nu,alpha)*(-12*f1*pow(phi,4)*kappa-
3*f1*pow(phi,8)*kappa+15*f1*pow(phi,6)*kappa-3*pow(phi,7)*pow(kappa,2)+3*pow(phi,5)*pow(kappa,2))
    +E_eta(1,h*(1+phi),nu,alpha)*(-24*f1*kappa*pow(phi,3)-
6*f1*pow(phi,4)*kappa+6*pow(phi,5)*f1*kappa+24*f1*kappa*pow(phi,2)+18*pow(phi,4)*pow(kappa,2)-
12*pow(phi,3)*pow(kappa,2)-6*pow(phi,5)*pow(kappa,2));
    momento_ordem_3=lambda*muc*pow(iota,3)/((1+2*phi+pow(phi,2))*(pow(phi,4)-2*pow(phi,3)-
3*pow(phi,2)+8*phi-4)*pow(phi,3)); //momento de terceira ordem
    return momento_ordem_3/pow(variancia,1.5);
}
double funcao_gamma(double z) //Aproximacao numerica da funcao Gama
{
    if(z<=0)
    {
        cout << "Erro na Funcao Gama!";
        exit(1);
    }
    double pi= 3.14159265358979323846;
    vector<double> p;
    p.push_back(1.000000000190015);
    p.push_back(76.18009172947146);
    p.push_back(-86.50532032941677);
    p.push_back(24.01409824083091);
    p.push_back(-1.231739572450155);
    p.push_back(1.208650973866179*pow(10,-3.0));
    p.push_back(5.395239384953*pow(10,-6.0));

```

```

    double somatorio;
    somatorio=p.at(0);
    for(int i=1;i<=6;i++)
        somatorio=somatorio+p.at(i)/(z+i);
    return sqrt(2*pi)/z*somatorio*pow(z+5.5,z+0.5)*exp(-1*(z+5.5));
}
double E_eta(int k, double s, double nu, double alpha) //Valor esperado, assim como definido no paper
{
    if(alpha<=k)
    {
        cout << "Erro no calculo do valor esperado!" << endl;
        exit(1);
    }
    return pow(nu,alpha)/funcao_gamma(alpha)*funcao_gamma(alpha-k)/pow(nu+s,alpha-k);
}

```

“calibracao.cpp”

```

include "sce.h"
#include "ordenar_matriz.h"
#include "f_objetivo.h"
#include <iostream>
#include <string>
using namespace std;
int main()
{
    ofstream resultados;
    resultados.open(".\\resultados\\relatorio.csv",ios::app);
    resultados << "it,x1min,x2min,x3min,x4min,x5min,x6min,f_min,f_max,f_med,cv,cont" << endl;
    vector<double> solucao;
    for(int mes=1;mes<=12;mes++)
    {
        cout << "Calibrando o mes " << mes << "..." << endl;
        resultados << "Calibracao do mes " << mes << endl;
        sce(mes, resultados);
    }
    resultados.close();
    return 0;
}

```

“amostragem_hipercubo_latino.h”

```

#include <vector>
#include <random>
//Realiza a amostragem do espaco viavel (Omega) por meio do Metodo de Hipercubo Latino
using namespace std;
using matriz2d = vector<vector<double>>;
matriz2d amostragem_hipercubo_latino(matriz2d Omega, int s, mt19937_64& gen){
    int n=Omega.size(); //Numero de variaveis de decisao
    matriz2d amostra(s,vector<double>(n));
    uniform_int_distribution<int> dist_int(1,s); //Distribuicao uniforme para numeros inteiros
    uniform_real_distribution<double> dist(0.0,1.0); //Distribuicao uniforme para numeros reais
    vector<int> marcador(s); //Aponta se o intervalo sorteado ja se encontra ocupado (uma vez fixada a
dimensao)
    double dx, lim_sup, lim_inf;
    int int_aleatorio;
    for(int j=0;j<n;j++){//Percorre cada dimensao
        for(auto& m:marcador){
            m=0;
        }
        dx=(Omega.at(j).at(1)-Omega.at(j).at(0))/double(s);
        for(int i=0;i<s;i++){//Percorre cada solucao
            do{
                int_aleatorio=dist_int(gen);
            }while(marcador.at(int_aleatorio-1)==1);
            marcador.at(int_aleatorio-1)=1;
            lim_inf=Omega.at(j).at(0)+(int_aleatorio-1)*dx;
            lim_sup=Omega.at(j).at(0)+(int_aleatorio)*dx;
            amostra.at(i).at(j)=Omega.at(j).at(0)+(int_aleatorio-1)*dx+dx*dist(gen);
        }
    }
    return amostra;
}

```

“ordenar_matriz.h”

```

#include <vector>
#include <algorithm>
#pragma once
using namespace std;
using matrix2d = vector<vector<double>>;
matrix2d ordenar_matriz(matrix2d matriz_entrada, int coluna){
    vector<pair<double,int>> valores_f_objetivo;
    int k=0;
    for(auto i:matriz_entrada){
        valores_f_objetivo.push_back(make_pair(matriz_entrada.at(k).at(coluna),k));
        k++;
    }
    sort(valores_f_objetivo.begin(),valores_f_objetivo.end());
    int colunas=matriz_entrada.at(0).size();
    int linhas=matriz_entrada.size();

    matrix2d matriz_ordenada(linhas, vector<double>(colunas));
    for(int m=0;m<linhas;m++){
        for(int n=0;n<colunas;n++){
            matriz_ordenada.at(m).at(n)=matriz_entrada.at(valores_f_objetivo.at(m).second).at(n);
        }
    }
    return matriz_ordenada;
}

```

“escolha_aleatoria.h”

```

//Realiza q escolhas aleatorias sem reposicao a partir de um vetor de probabilidades discretas (pi)
//Retorna um vetor com os respectivos indices dos elementos escolhidos
#include <vector>
#include <algorithm>
#include <random>
#pragma once
using namespace std;
using matriz2d = vector<vector<double>>;
vector<int> escolha_aleatoria(vector<double> pi, int q, mt19937_64& gen){
    uniform_real_distribution<double> dist(0.0,1.0);
    int i, j;
    double somatorio=0.0, u;
    vector<double> pi_acum;
    for(int i=0;i<pi.size();i++){
        somatorio=somatorio+pi.at(i);
        pi_acum.push_back(somatorio);
    }
    vector<int> L(q);
    for(i=0;i<q;i++){
        L.at(i)=-1;
    }
    i=0;
    while(i<q){
        j=0;
        u=dist(gen);
        while(u>pi_acum.at(j))
            j++;
        if(find(L.begin(),L.end(),j) == L.end()){ //Se o elemento nao for encontrado em L, find retorna o
iterador relativo a L.end()
            L.at(i)=j;
            i++;
        }
    }
    return L;
}

```

“gerar_ponto_z.h”

```

#include <vector>
#pragma once
using namespace std;
using matriz2d = vector<vector<double>>;

vector<double> gerar_ponto_z(matriz2d H, mt19937_64& gen){
    uniform_real_distribution<double> dist(0.0,1.0);
    int j;
    int n=H.size();
    vector<double> z(n);
    for(j=0;j<n;j++){
        z.at(j)=(H.at(j).at(1)-H.at(j).at(0))*dist(gen)+H.at(j).at(0);
    }
    return z;
}

```

“sce.h”

```

//ALGORITMO SHUFFLED COMPLEX EVOLUTION (SCE) (Duan, Sorooshian e Gupta, 1992)
#include <iostream>
#include <vector>
#include <fstream>
#include <sstream>
#include <random>
#include <chrono>
#include <string>
#include "f_objetivo.h"
#include "amostragem_hipercubo_latino.h"
#include "ordenar_matriz.h"
#include "cce.h"
using namespace std;
using matriz2d = vector<vector<double>>;
int sce(int mes, ofstream& resultados_relatorio)
{
    int contador=0;
    //DECLARACAO DE CONTROLADORES DO SCE-UA
    int itmax=100; //Numero maximo de iteracoes
    double cv_max=pow(10,-6); //Coeficiente de variacao maximo admissivel (tolerancia)
    //PARAMETROS DO SCE-UA - Aquisicao no arquivo .\\entradas\\parametros.csv
    vector<int> parametros(6); //Parametros do SCE
    ifstream parametros_arquivo;
    parametros_arquivo.open(".\\entradas\\parametros.txt", ios::in);
    string linha, desprezar;
    string parametro;
    for(auto& i:parametros)
    {
        getline(parametros_arquivo,linha);
        stringstream ss1(linha);
        getline(ss1,desprezar,',');
        getline(ss1,parametro);
        i=stoi(parametro);
    }
    int n=parametros.at(0); //Numero de variaveis de decisao - dimensao do problema
    int p=parametros.at(1); //Numero de complexos Ak (arbitrar)
    int m=parametros.at(2); //Numero de pontos em cada complexo m>=n+1 (recomenda-se 2n+1)
    int q=parametros.at(3); //Numero de pontos em cada simplex 2<=q<=m (recomenda-se n+1)
    int beta=parametros.at(4); //Numero de vezes que serao escolhidas solucoes de Ak para compor os
    simplexes beta>=1 (recomenda-se 2n+1)
    int alfa=parametros.at(5); //Numero de vezes que os simplexes serao evolidos por Nelder-Mead ou outro
    algoritmo (recomenda-se alfa>=1)
    //ESPACO VIAVEL - Aquisicao no arquivo localizado em “.\\entradas\\Omega.csv”
    matriz2d Omega(n,vector<double>(2));
    ifstream Omega_arquivo;
    Omega_arquivo.open(".\\entradas\\Omega.txt", ios::in);
    linha.clear();
    string lim_inf, lim_sup;
    for(auto& i:Omega)
    {
        getline(Omega_arquivo,linha);
        stringstream ss2(linha);
        getline(ss2,desprezar,',');
        getline(ss2,lim_inf,',');
        i.at(0)=stod(lim_inf);
        getline(ss2,lim_sup,',');
        i.at(1)=stod(lim_sup);
    }
    //DECLARACAO DAS VARIAVEIS DO CODIGO

```

```

int it; //Referente as iteracoes do SCE: it < itmax
int k; //Referente aos Complexos "Ak": - k <= p
int i, j, i1, i2; //Iteradores para matrizes e vetores (i1 para a matriz D - i1<=s e i2 para os complexos Ak -
i2 <=m)
int s=p*m; //Tamanho das populacoes (e fixo, uma vez determinado o numero de complexos e o numero
de solucoes em cada complexo)
//DECLARACAO DE VARIAVEIS PARA GERACAO DE ALEATORIOS
mt19937_64 gen; //Declaracao do gerador referente a curva Mersenne Twister
unsigned long seed = chrono::system_clock::now().time_since_epoch()/chrono::milliseconds(1);
//Semente - milissegundos do relógio
gen.seed(seed); //Geracao de semente para acesso a curva Mersenne Twister
// POPULACAO INICIAL
matriz2d D(s,vector<double>(n+1)); //Matriz contendo as solucoes, atualizada a cada iteracao do SCE-
UA
D=amostragem_hipercubo_latino(Omega,s,gen); //Amostragem de Omega pelo Metodo de Hipercubo
Latino
vector<double> x(n);
//Expandir a matriz para conter a coluna com os VFOs
for(auto& i:D)
{
    i.push_back(0);
}
//Calcular os VFOs e ordenar a populacao inicial
for(i=0;i<s;i++)
{
    x.assign(D.at(i).begin(),D.at(i).begin()+n);
    D.at(i).back()=f_objetivo(mes, contador, x);
}
D=ordenar_matriz(D,n);
matriz2d A(m,vector<double>(n+1)); //Matriz contendo os complexos Ak, atualizada a cada nova
execucao do CCE
vector<int> vinculos(m); //Vetor "vinculo" contendo os indices de acesso a matriz D a partir de cada
complexo Ak
// ===== INICIO DO PROCESSO ITERATIVO (governado por itmax) ===== //
resultados_relatorio << "it,";
for(i=1;i<=n;i++){
    resultados_relatorio << "x" << i << "min,";
}
resultados_relatorio << "f_min" << "," << "f_max" << "," << "f_med" << "," << "cv" << endl;
vector<double> solucao;
double f_med, var, cv;
it=0;
do{
    cout << "...Iteracao " << it+1 << endl;
    i2=0;
    k=1;
    do{
        vinculos.at(i2)=k+p*i2-1;
        for(j=0;j<n+1;j++){
            A.at(i2).at(j)=D.at(vinculos.at(i2)).at(j);
        }
        if(i2+1==m){ //Esse condicional implica que o complexo k ja esta contido na matriz Ak
            A=cce(mes, contador, A, k, Omega, q, beta, alfa, gen); //Executar o algoritmo CCE
            para evolucao do complexo "Ak"
            for(i2=0;i2<m;i2++){
                for(j=0;j<n+1;j++){ //Retornar as solucoes evoluídas para a matriz D
                    D.at(vinculos.at(i2)).at(j)=A.at(i2).at(j);
                }
            }
        }
        i2=0;
    }
}

```

```

        k++;
    }
    else
        i2++;
}while(k<=p);
D=ordenar_matriz(D,n);
f_med=0;
for(i=0;i<s;i++){
    f_med=f_med+D.at(i).back();
}
f_med=f_med/i;
for(i=0;i<s;i++){
    var=pow(D.at(i).back()-f_med,2);
}
var=var/i;
cv=sqrt(var)/abs(f_med);
it++;
resultados_relatorio << it << ",";
//Imprime a solucao otima da iteracao
for(i=0;i<n;i++){
    resultados_relatorio << D.at(0).at(i) << ",";
}
//Imprime o valor de f_min
resultados_relatorio << D.at(0).back() << ",";
//Imprime o valor de f_max
resultados_relatorio << D.at(s-1).back() << ",";
//Imprime o valor de f_max
resultados_relatorio << f_med << ",";
//Imprime o valor de cv
resultados_relatorio << cv << ",";
//Imprime o numero de avaliacoes da funcao objetivo
resultados_relatorio << contador << endl;
//Analise do criterio de parada (Coeficiente de Variacao)
}while(it<itmax && cv>cv_max);
return 0;
}

```

"cce.h"

//ALGORITMO COMPETITIVE COMPLEX EVOLUTION (CCE) (Duan, Sorooshian e Gupta, 1992)

```

#include <vector>
#include <algorithm>
#include <iostream>
#include <fstream>
#include "escolha_aleatoria.h"
#include "nelder_mead.h"
#pragma once
using namespace std;
using matriz2d = vector<vector<double>>;
matriz2d cce(int mes, int& contador, matriz2d A, int k, matriz2d Omega, int q, int BETA, int ALFA,
mt19937_64& gen){
    int i, j; //Iteradores para linhas (i) e colunas (j) das matrizes
    int beta=0; //Iteracoes governadas pelo parametro beta
    int m=A.size();
    int n=A.at(0).size()-1;
    matriz2d H(n,vector<double>(2)); H - menor hipercubo contendo as solucoes do complexo Ak inicial
    double min, max;
    for(j=0;j<n;j++){
        min=numeric_limits<double>::max();
        max=-min;
        for(i=0;i<m;i++){
            if(A.at(i).at(j)<min)
                min=A.at(i).at(j);
            if(A.at(i).at(j)>max)
                max=A.at(i).at(j);
        }
        H.at(j).at(0)=min;
        H.at(j).at(1)=max;
    }
    //Atribuir distribuicao "triangular" para as solucoes de Ak pontos do complexo
    vector<double> pi(m);
    for(j=0;j<m;j++){
        pi.at(j)=2.0*(m+1.0-(j+1.0))/double((m*(m+1.0)));
    }
    //EVOLUCAO - ESCOLHA ALEATORIA (controlado por beta)
    //Selecionar "q" solucoes pais do complexo para executar a evolucao
    vector<int> L(q);
    matriz2d B(q, vector<double>(n+1));
    do{
        L = escolha_aleatoria(pi,q,gen);
        for(i=0;i<q;i++){
            for(j=0;j<n+1;j++){
                B.at(i).at(j)=A.at(L.at(i)).at(j);
            }
        }
        B=nelder_mead(mes, contador, B, Omega, H, ALFA, gen); //Entra no N&M
        //Substituir as solucoes no complexo Ak
        for(i=0;i<q;i++){
            for(j=0;j<n+1;j++){
                A.at(L.at(i)).at(j)=B.at(i).at(j);
            }
        }
        A=ordenar_matriz(A,n);
        beta++;
    }while(beta<BETA);
    return A;
}

```

"nelder_mead.h"

```
//ALGORITMO DE NELDER E MEAD (Nelder e Mead, 1965)
#include <iostream>
#include <vector>
#include <algorithm>
#include <fstream>
#include <random>
#include <iomanip>
#include <cmath>
#include <string>
#include "f_objetivo.h"
#include "ordenar_matriz.h"
#include "gerar_ponto_z.h"
#pragma once
using namespace std;
using matriz2d = vector<vector<double>>;
matriz2d nelder_mead(int mes, int& contador, matriz2d B, matriz2d Omega, matriz2d H, int ALFA,
mt19937_64& gen){
    int i, j; //Iteradores para linhas (i) e colunas (j) das matrizes
    int n=B.at(0).size()-1;
    int q=B.size();
    int alfa=0; //Iteracoes governadas pelo parametro ALPHA
    vector<double> x(n), h(n), l(n), g(n), r(n), e(n), c(n); //Vetores contendo os pontos de expansao,
    contracao, etc.
    double f_h, f_l, f_r, f_e, f_c, f_med; //Valores da funcao objetivo para os pontos high(pior), low(melhor),
    refletido, expandido e contraido
    B=ordenar_matriz(B,n);
    do{
//Calcular o centroide - g
        for(auto& j:g){
            j=0;
        }
        for(j=0;j<n;j++){
            for(i=0;i<q-1;i++){
                g.at(j)=g.at(j)+B.at(i).at(j);
            }
            g.at(j)=g.at(j)/double(q-1);
        }
//Atribuir ao ponto h ("high" - o pior) o valor da funcao objetivo - f_h
        f_h=B.at(q-1).at(n);
//Atribuir ao ponto l ("low" - o melhor) o valor da funcao objetivo - f_l
        f_l=B.at(0).at(n);
//Calcular o ponto de reflexao - r=2g-h
        for(j=0;j<n;j++){
            r.at(j)=2*g.at(j)-B.at(q-1).at(j);
        }
//Verificar se "r" esta dentro do espaco viavel (Omega); se nao estiver, gerar aleatorio em H
        for(j=0;j<n;j++){
            if(r.at(j)<Omega.at(j).at(0) || r.at(j)>Omega.at(j).at(1)){
                r=gerar_ponto_z(H, gen);
                break;
            }
        }
        f_r=f_objetivo(mes, contador, r);
        if(f_r<=f_l){
//Se houve melhora na reflexao, calcular o ponto de expansao - e=2*r-g
            for(j=0;j<n;j++){
                e.at(j)=2*r.at(j)-g.at(j);
            }
//Verificar se "e" esta dentro do espaco viavel (Omega); se nao estiver, gerar aleatorio em H
```

```

    for(j=0;j<n;j++){
        if(e.at(j)<Omega.at(j).at(0) || e.at(j)>Omega.at(j).at(1)){
            e=gerar_ponto_z(H, gen);
            break;
        }
    }
    f_e=f_objetivo(mes, contador, e);
    if(f_e<=f_l){
//Se houve melhora na expansao, substituir o ponto "high" pelo expandido (PROXIMA ITERACAO)
        for(j=0;j<n;j++){
            B.at(q-1).at(j)=e.at(j);
        }
        B.at(q-1).at(n)=f_e;
    }
    else{
//Se nao houve melhora na expansao, substituir o pior "high" pelo refletido (PROXIMA ITERACAO)
        for(j=0;j<n;j++){
            B.at(q-1).at(j)=r.at(j);
        }
        B.at(q-1).at(n)=f_r;
    }
}
else if(f_r>=B.at(q-2).at(n)){
    if(f_r<f_h){
//Se o refletido e melhor do que o h, substituir h pelo refletido e contrair na direcao do mesmo (outside
contraction)
//Do contrario, automaticamente vai manter o h anterior e contrair na direcao dele (inside contraction)
        for(j=0;j<n;j++){
            B.at(q-1).at(j)=r.at(j);
        }
        B.at(q-1).at(n)=f_r;
        f_h=f_r;
    }
}
//Calcular o ponto de contracao - c=1/2(h+g)
    for(j=0;j<n;j++){
        c.at(j)=0.5*(g.at(j)+B.at(q-1).at(j));
    }
    f_c=f_objetivo(mes, contador, c);
    if(f_c>=f_h){
//Se nao houve melhora na contracao, realizar a reducao de todos os pontos em direcao ao ponto low
(PROXIMA ITERACAO)
        for(i=1;i<=q-1;i++){
            for(j=0;j<n;j++){
                B.at(i).at(j)=(B.at(i).at(j)+B.at(0).at(j))/2.0;
                x.at(j)=B.at(i).at(j);
            }
            B.at(i).at(n)=f_objetivo(mes, contador, x);
        }
    }
    else{
//Se houve melhora na contracao, substituir o pior ponto pelo ponto de contracao (PROXIMA ITERACAO)
        for(j=0;j<n;j++){
            B.at(q-1).at(j)=c.at(j);
        }
        B.at(q-1).at(n)=f_c;
    }
}
else{
//O ponto refletido produziu uma melhora parcial (existem pontos piores do que ele no simplex da iteracao) -
substituir o pior pelo refletido (PROXIMA ITERACAO)

```

```
        for(j=0;j<n;j++){
            B.at(q-1).at(j)=r.at(j);
        }
        B.at(q-1).at(n)=f_r;
    }
//Incrementar alfa, ordenar a matriz B para a proxima iteracao, imprimir o simplex resulante da iteracao atual e
avaliar o criterio de parada
    alfa++;
    B=ordenar_matriz(B,n);
}while(alfa<ALFA);
return B;
}
```

B3 – programas e rotinas desenvolvidos para a etapa de simulação

“blrp.h” (Geração de séries sintéticas com o BLRP)

```

#include <random>
#include "superposicao.h"
#include "estatisticas.h"
using namespace std;
using matriz2d = vector<vector<double>>;
vector<double> blrp(vector<double> parametros, int anos, int dias, double h, mt19937_64& gen)
{
    double lambda=parametros.at(0);
    double beta=parametros.at(1);
    double gama=parametros.at(2);
    double eta=parametros.at(3);
    double mux=parametros.at(4);
    double to_evento, to_celula; //Tempos de origem dos eventos e das celulas
    double L_evento, L_celula; //Duracao dos eventos e das celulas
    double X; //Intensidade das celulas
    matriz2d R(anos*dias*int(24/h),vector<double>(2)); //Matriz que comporta a serie sintetica (tempos
    finais de cada subintervalo e laminas precipitadas)
    double t_acum=h;
    for(auto& i:R)
    {
        i.at(0)=t_acum;
        i.at(1)=0;
        t_acum=t_acum+h;
    }
    exponential_distribution<double> tempo_entre_eventos(lambda); //Processo de Poisson relativo a
    geracao de eventos
    exponential_distribution<double> duracao_eventos(gama); //Duraçã dos eventos segue distribuicao
    exponencial com taxa gama
    exponential_distribution<double> tempo_entre_celulas(beta); //Processo de Poisson relativo a geracao de
    celulas
    exponential_distribution<double> duracao_celulas(eta); //Duracao das celulas segue distribuicao
    exponencial com taxa eta (nesse caso constante)
    exponential_distribution<double> intensidade_celulas(1.0/mux); //Intensidade das celulas segue
    distribuicao exponencial com taxa 1/mux (nesse caso constante)
    int marcador_ev=0, marcador_cel; //Ponto a partir do qual deve ocorrer a busca do tempo de origem de
    celula especifica na matriz R
    to_evento=0;
    do //Geracao de eventos
    {
        L_evento=duracao_eventos(gen);
        to_celula=to_evento; //No caso dos modelos BL o tempo de origem da primeira celula coincide
        com o tempo de origem do evento
        while(to_evento>=R.at(marcador_ev).at(0)) //Otimiza a funcao de superposicao
        {
            marcador_ev++;
        }
        marcador_cel=marcador_ev;
        do //Geracao de celulas
        {
            L_celula=duracao_celulas(gen);
            X=intensidade_celulas(gen);
            if(to_celula>=R.back().at(0)) //Evita a superposicao de celulas cujo instante de origem
            ocorre depois do final do evento
                break;
        }
    }
}

```

```
        else
            superposicao(R,h,to_celula,L_celula,X,marcador_cel);
            to_celula=to_celula+tempo_entre_celulas(gen);
        }while(to_celula<to_evento+L_evento);
        to_evento=to_evento+tempo_entre_eventos(gen);
    }while(to_evento < R.back().at(0));
    vector<double> resultados=estatisticas(R);
    return resultados;
}
```

“blrprx.h” (Geração de séries sintéticas com o BLRPRx)

```

#include <random>
#include "superposicao.h"
#include "estatisticas.h"
using namespace std;
using matriz2d = vector<vector<double>>;
vector<double> blrprx(vector<double> parametros, int anos, int dias, double h, mt19937_64& gen)
{
    double lambda=parametros.at(0);
    double iota=parametros.at(1);
    double alfa=parametros.at(2);
    double ni=parametros.at(3);
    double capa=parametros.at(4);
    double fi=parametros.at(5);
    double eta, beta, gama, mux; //Esses parametros de evento para evento
    double to_evento, to_celula; //Tempos de origem dos eventos e das celulas
    double L_evento, L_celula; //Duracao dos eventos e das celulas
    double X; //Intensidade das celulas
    matriz2d R(anos*dias*int(24/h),vector<double>(2)); //Matriz que comporta a serie sintetica (tempos
    finais de cada subintervalo e laminas precipitadas)
    double t_acum=h;
    for(auto& i:R)
    {
        i.at(0)=t_acum;
        i.at(1)=0;
        t_acum=t_acum+h;
    }
    exponential_distribution<double> tempo_entre_eventos(lambda); //Processo de Poisson relativo a
    geracao de eventos
    gamma_distribution<double> dist_gama(alfa,1/ni); //Distribuicao utilizada na geracao do eta de cada
    evento
    int marcador_ev=0, marcador_cel; //Ponto a partir do qual deve ocorrer a busca do tempo de origem de
    celula especifica na matriz R
    to_evento=0;
    do //Geracao de eventos
    {
        eta=dist_gama(gen); //Definicao de eta do evento
        //Duracao do evento
        gama=eta*fi;
        exponential_distribution<double> duracao_eventos(gama);
        L_evento=duracao_eventos(gen);
        //Demais distribuicoes condicionadas a eta
        exponential_distribution<double> duracao_celulas(eta); //Duracao das celulas
        beta=eta*capa;
        exponential_distribution<double> tempo_entre_celulas(beta); //Processo de Poisson relativo a
        geracao de celulas
        mux=eta*iota;
        exponential_distribution<double> intensidade_celulas(1.0/mux); //Intensidade das celulas
        to_celula=to_evento; //No caso dos modelos BL o tempo de origem da primeira celula coincide
        com o tempo de origem do evento
        while(to_evento>=R.at(marcador_ev).at(0)) //Otimiza a funcao de superposicao
        {
            marcador_ev++;
        }
        marcador_cel=marcador_ev;
        do //Geracao de celulas
        {
            L_celula=duracao_celulas(gen);
            X=intensidade_celulas(gen);

```

```

evento
    if(to_celula>=R.back().at(0)) //Evita a superposicao de celulas com origem apos o final do
        break;
    else
        superposicao(R,h,to_celula,L_celula,X,marcador_cel);
        to_celula=to_celula+tempo_entre_celulas(gen);
    }while(to_celula<to_evento+L_evento);
    to_evento=to_evento+tempo_entre_eventos(gen);
}while(to_evento < R.back().at(0));
vector<double> resultados=estatisticas(R);
return resultados;
}

```

“superposicao.h”

```

#include <vector>
using namespace std;
int superposicao(vector<vector<double>>& R, double h, double to_celula, double L, double X, int&
marcador)
{
    int i=marcador; //Indice referente ao intervalo a partir do qual ocorre a busca na coluna de tempos da
matriz R
    if(to_celula+L>R.back().at(0)) //Encurtando a duracao pulsos que extrapolam o tamanho da serie
        L=R.back().at(0)-to_celula;
    //Identificacao do subintervalo contendo o tempo de origem da celula (to_celula)
    while(to_celula>=R.at(i).at(0))
    {
        i++;
    }
    marcador=i;
    //Identificacao do subintervalo contendo o tempo de encerramento da celula (to_celula+L)
    if(to_celula+L<=R.at(i).at(0)) //Encerramento dentro do subintervalo que tambem contem a origem
    {
        R.at(i).at(1)=R.at(i).at(1)+X*L; //Superposicao do pulso no primeiro subintervalo
    }
    else
    {
        R.at(i).at(1)=R.at(i).at(1)+X*(R.at(i).at(0)-to_celula); //Superposicao do pulso no primeiro
subintervalo
        if(i!=R.size()-1)
        {
            i++;
            while(to_celula+L>R.at(i).at(0))
            {
                R.at(i).at(1)=R.at(i).at(1)+X*h; //Superposicao do pulso nos subintervalos
intermediarios
                i++;
            }
            R.at(i).at(1)=R.at(i).at(1)+X*(to_celula+L-R.at(i-1).at(0)); //Superposicao do pulso no
subintervalo de encerramento
        }
    }
    return 0;
}

```