UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

INSTITUTO DE PESQUISAS HIDRÁULICAS

PROGRAMA DE PÓS-GRADUAÇÃO EM RECURSOS HÍDRICOS E SANEAMENTO AMBIENTAL

LUCAS RANGEL MARTINS

EFEITO EM TOTAIS MENSAIS DE PRECIPITAÇÃO DO NÃO PREENCHIMENTO DE FALHAS DE DADOS DIÁRIOS

PORTO ALEGRE

2019

LUCAS RANGEL MARTINS

EFEITO EM TOTAIS MENSAIS DE PRECIPITAÇÃO DO NÃO PREENCHIMENTO DE FALHAS DE DADOS DIÁRIOS

Dissertação apresentada ao Programa de Pós-graduação em Recursos Hídricos e Saneamento Ambiental da Universidade Federal do Rio Grande do Sul, como requisito parcial à obtenção do grau de mestre.

Orientador: Fernando Dornelles

PORTO ALEGRE

2019

CIP - Catalogação na Publicação

Martins, Lucas Rangel
Efeito em Totais Mensais de Precipitação do Não Preenchimento de Falhas de Dados Diários / Lucas Rangel Martins. -- 2019.
97 f.
Orientador: Fernando Dornelles.

Dissertação (Mestrado) -- Universidade Federal do Rio Grande do Sul, Instituto de Pesquisas Hidráulicas, Programa de Pós-Graduação em Recursos Hídricos e Saneamento Ambiental, Porto Alegre, BR-RS, 2019.

1. Hidrologia. 2. Análise de consistência. 3. Precipitação. 4. Pluviometria. 5. Preenchimento de dados. I. Dornelles, Fernando, orient. II. Título.

Elaborada pelo Sistema de Geração Automática de Ficha Catalográfica da UFRGS com os dados fornecidos pelo(a) autor(a).

LUCAS RANGEL MARTINS

EFEITO EM TOTAIS MENSAIS DE PRECIPITAÇÃO DO NÃO PREENCHIMENTO DE FALHAS DE DADOS DIÁRIOS

Dissertação apresentada ao Programa de Pós-graduação em Recursos Hídricos e Saneamento Ambiental da Universidade Federal do Rio Grande do Sul, como requisito parcial à obtenção do grau de mestre.

Aprovado em: Porto Alegre, 17 de junho de 2019.

Dr. Fernando Dorneles – IPH/UFRGS
Orientador

Dr. Francisco Fernando Noronha Marcuzzo – CPRM/SGB
Examinador

Prof. Dr. André Luiz Lopes da Silveira – IPH/UFRGS
Examinador

Prof. Dr. Olavo Correa Pedrollo – IPH/UFRGS
Examinador

Dedico este trabalho a meus pais, Renato e Corina e a minha noiva Pietra, que sempre me apoiaram e especialmente durante o período do meu Curso de Pós-Graduação estiveram ao meu lado.

AGRADECIMENTOS

Agradeço ao Professor Fernando Dornelles, orientador deste trabalho, por ter demonstrado interesse em auxiliar o desenvolvimento deste estudo, disponibilizando, sempre com paciência, o seu tempo para o debate e a reflexão de ideias, guiando o rumo desta pesquisa.

Agradeço aos meus pais, Renato e Corina, por terem tornado a minha trajetória acadêmica possível, fornecendo todos os recursos necessários para que esta etapa fosse realizada com êxito. Através de palavras de motivação e acreditando na minha capacidade, me apoiaram e me guiaram em todas as etapas que me trouxeram a este momento.

Agradeço a minha noiva Pietra pelas revisões textuais do presente trabalho, assim como, toda a paciência e compreensão nos momentos de empenho e tensão que se passaram ao longo da elaboração desta pesquisa e ao longo dos demais anos de academia.

Agradeço a todos familiares e amigos pelas palavras de motivação e carinho e pelos conselhos nos momentos decisivos durante minha vida acadêmica e profissional.

Sábio é aquele que conhece os limites da própria ignorância.

Sócrates

RESUMO

Para o apropriado conhecimento da disponibilidade hídrica e de eventos extremos para a gestão dos recursos hídricos e a elaboração de estudos e projetos na área, é essencial o monitoramento do fenômeno da precipitação. Para que seja feito de maneira apropriada, deve se assegurar a qualidade dos registros pluviométricos, assim como, das análises e tratamentos dessas séries anteriormente à realização de estudos hidrológicos. Portanto, buscou-se nesse estudo compreender melhor os limites do uso de técnicas de preenchimento convencionais de falhas, utilizadas de forma abrangente em estudos de hidrologia, e comparar os erros de tais técnicas aos erros que falhas aleatórias diárias de extensão variada podem causar às séries mensais de precipitação, calculando, então, o total mensal sem o total de registros do mês. Para isso, foram testadas as técnicas de preenchimento através de regressões lineares, ponderação regional, vetor regional, inverso da distância ao quadrado e ponderações regionais com base em regressões lineares. Tais técnicas foram testadas em 12 grupos distintos, sendo uma estação principal em cada grupo, e pelo menos outras cinco estações de apoio. A técnica com melhor desempenho em geral foi a regressão linear, enquanto as piores foram o vetor regional e o inverso da distância ao quadrado. Além disso, nas estações principais, foram comparados os erros de estimativa da série original através das séries sintéticas das técnicas de preenchimento aos erros que falhas diárias causam ao total mensal precipitado da série original. Foram geradas 1.000 séries sintéticas com falhas aleatórias, variando a sua extensão. Observou-se que a média dos erros tende a ser a precipitação média diária multiplicada pelo número de falhas criadas no mês. Para as diferentes regiões, observou-se que o erro das técnicas de preenchimento tende a ser equivalente a 4 a 12 dias de falhas em termos medianos, dependendo da região e da época do ano. Tais valores levam a crer que em muitos casos, mantendo-se os valores originais diários registrados, mesmo havendo falhas naquele mês, tem-se um valor total mensal mais próximo da realidade do que através das técnicas de preenchimento tradicionalmente utilizadas. Observou-se que é viável a utilização da alternativa de não-preenchimento, uma vez observados os critérios de limite de falhas, devendo ser aplicado a um maior número de estações para que se possa validar de forma mais conclusiva tal observação.

Palavras chave: chuva; análise de consistência; preenchimento de dados; homogeneização; análise de precipitação; pluviometria.

ABSTRACT

For the adequate knowledge of water availability and extreme flood and drought events for water resources management as for studies and projects in this area, it is essential to measure and register rainfall data. In order to do it properly, the quality of the rainfall records, as well as the analyzes and treatments of these series prior to the hydrological studies must be ensured. Therefore, this study sought to better understand the limits of the use of conventional filling gaps techniques, widely used in hydrology studies, and to compare the errors of such techniques to the errors that daily random failures of varied extent can cause to the monthly series of rainfall. To assess it, the filling techniques were tested through linear regressions, regional weighting, regional vector, inverse squared distance weighting and regional weights based on linear regressions. These techniques were tested in twelve distinct groups, one main station in each group, and at least five other support stations. The best performing technique in general terms was the linear regression, while the worst was the regional vector and the inverse squared distance weighting. In addition, in the main stations, the estimation errors through synthetic series generated by filling techniques of the original series were compared to errors that daily failures cause to the original series. A thousand synthetic series were generated with random faults, varying their length. It was observed that the mean of the errors tends to be the average daily precipitation multiplied by the number of failures created in the series. For the different regions, it was observed that the error of the filling techniques tends to be equivalent to 4 to 12 days of failure in medium terms. Depending on the region and the time of year. These values lead to believe that in many cases, keeping the original daily values recorded, even if there were failures in that month, there is a total monthly value closer to reality than through the traditional use of filling techniques. It was observed that it is feasible to use the no-fill technique, once the failure limit criteria have been observed, and it should be applied to a larger number of stations in order to allow a more conclusive validation of this observation.

Keywords: precipitation; consistency analysis; filling in data; homogenization; rainfall analysis; rainfall measure.

LISTA DE SÍMBOLOS

$\rho_{X,Y}$ – coeficiente de correlação linear entre as variáveis X e Y.

$\sigma_{x,y}$ – covariância entre as variáveis X e Y.

σ_x , σ_y – desvios-padrão das variáveis X e Y.

[C] – vetor linha com n colunas, representando n postos.

[L] – vetor coluna com m linhas, representando m observações.

$[P_e]$ – matriz de precipitações do método vetor regional.

a_0, a_j, α, β,... – coeficientes de ajuste.

C_j – valor do índice j do vetor coluna C do método vetor regional.

d_i – distância da estação i com a estação a ser estimada.

EAM – erro absoluto médio.

k – número de falhas diárias em 1 mês.

L_i – valor do índice i do vetor coluna L do método vetor regional.

m_1 e m_2 – coeficientes angulares da reta nos períodos recente e antigo.

m_i – número de meses em que ocorreu sobreposição de dados.

n – número de estações.

n_s – número de séries sintéticas.

P_A, P_B,... – precipitações nas estações A, B, ...

P_{est} – precipitação estimada no período a ser corrigido.

P_i – precipitação das n estações.

P_{ij} – estimativa da precipitação anual ou mensal no ano “i” na estação “j” pelo método vetor regional.

$P_{m\,i}$ – precipitação média mensal nas n estações.

$P_{m\,x}$ – precipitação média mensal na estação X a ser estimada.

P_{obs} – precipitação observada no período a ser corrigido.

REQM – Raiz do erro quadrático médio.

S_x – desvio padrão amostral.

$S_{x,y}$ – covariância amostral entre as variáveis X e Y.

V – viés (diferença entre o valor médio de um estimador estatístico e o valor que pretende estimar.

VP – viés percentual.

W_i – peso da estação i ponderado pela correlação linear.

ρ_i – coeficiente de correlação linear da estação i com a estação X.

SUMÁRIO

1 INTRODUÇÃO 16
2 DIRETRIZES DA PESQUISA 18
2.1 QUESTÃO DE PESQUISA 18
2.2 OBJETIVOS DA PESQUISA 18
2.3 HIPÓTESE 18
2.4 PRESSUPOSTO 18
2.5 DELIMITAÇÕES 19
2.6 LIMITAÇÕES 19
2.7 DELINEAMENTO 19
3 REVISÃO BIBLIOGRÁFICA 20
3.1 PRECIPITAÇÃO 20
3.2 PLUVIOMETRIA 21
3.2.1 Climatologia e Pluviometria Brasileira 22
3.3 ANÁLISE DE CONSISTÊNCIA 26
3.3.1 Tipos de erros 26
3.3.2 Análise dos dados de precipitação 28
3.3.3 Análise preliminar dos dados hidrológicos e técnicas de preenchimento de falhas 30
3.3.4 Resultados de avaliações dos métodos de preenchimento apresentados por diferentes autores 37
3.4 IMPACTO DE FALHAS DE DADOS NAS ESTATÍSTICAS DAS SÉRIES E ESTUDOS HIDROLÓGICOS 39
4 MATERIAIS E MÉTODOS 43
4.1 MÉTODOS 43
4.1.1 Criação de Falhas Sintéticas 44
4.1.2 Critério para Escolha das Séries Históricas e Análise dos Efeitos das Falhas 45
4.1.3 Comparação da Técnica do Não Preenchimento às Técnicas Tradicionalmente Utilizadas 46
4.2 MATERIAIS 47
5 APLICAÇÃO DO MÉTODO E RESULTADOS 53
5.1 AVALIAÇÃO POR MÉTODO DE DUPLA-MASSA E ESCOLHA DAS ESTAÇÕES PRINCIPAIS 53
5.2 GERAÇÃO DE SÉRIES POR PREENCHIMENTO 56
5.3 GERAÇÃO DE SÉRIES SINTÉTICAS COM FALHAS 72

5.4 ANÁLISE GERAL DOS RESULTADOS 87
5.5 RECOMENDAÇÕES PARA APLICAÇÃO PRÁTICA DA METODOLOGIA 88
6 CONCLUSÃO E RECOMENDAÇÕES 90
7 REFERÊNCIAS 92

FIGURAS

Figura 1. Classificação climatológica do Brasil de acordo com os critérios de Koppen (1936) 23
Figura 2. Mapa a partir das isoietas de Precipitação Média Anual Total no Brasil (1977-2006) 25
Figura 3. Fluxograma de procedimentos para a análise proposta 44
Figura 4. Localização das estações estudadas 49
Figura 5. Ampliação da localização das estações estudadas. 50
Figura 6. Diagramas de disponibilidade temporal das estações analisadas em cada região (continua). 51
Figura 7. Curvas dupla massa das estações com melhor ajuste de cada região (milímetros acumulados) (continua) 54
Figura 8. Média histórica dos erros das técnicas de preenchimento Norte G1 (8361000). 57
Figura 9. Média histórica dos erros das técnicas de preenchimento Norte G2 (260009). 57
Figura 10. Média histórica dos erros das téc. de preenchimento Centro-Oeste G1 (1656004). 57
Figura 11. Média histórica dos erros das téc. de preenchimento Centro-Oeste G2 (2054009). 58
Figura 12. Média histórica dos erros das téc. de preenchimento Nordeste G1 (340105). 58
Figura 13. Média histórica dos erros das téc. de preenchimento Nordeste G2 (638090). 58
Figura 14. Média histórica dos erros das téc. de preench. Sudeste RJ-ES G1 (2242009). 59
Figura 15. Média histórica dos erros das téc. de preench. Sudeste RJ-ES G2 (2041019). 59
Figura 16. Média histórica dos erros das téc. de preench. Sudeste SP-MG G1 (2247182). 59
Figura 17. Média histórica dos erros das téc. de preench. Sudeste SP-MG G2 (2044041). 60
Figura 18. Média histórica dos erros das técnicas de preenchimento Sul G1 (2853035). 60
Figura 19. Média histórica dos erros das técnicas de preenchimento Sul G2 (2750010). 60
Figura 20. *Box-plot* dos erros das téc. preenchimento mensal Norte G1 (8361000). 62
Figura 21. *Box-plot* dos erros das téc. preenchimento mensal Norte G2 (260009)... 62
Figura 22. *Box-plot* dos erros das téc. preenchimento mensal Centro-Oeste G1 (1656004). 62
Figura 23. *Box-plot* dos erros das téc. preenchimento mensal Centro-Oeste G2 (2054009). 63
Figura 24. *Box-plot* dos erros das téc. preenchimento mensal Nordeste G1 (340105). 63
Figura 25. *Box-plot* dos erros das téc. preenchimento mensal Nordeste G2 (638090). 63

Figura 26. *Box-plot* dos erros das téc. preenchimento mensal Sudeste RJ-ES G1 (2242009). 64
Figura 27. *Box-plot* dos erros das téc. preenchimento mensal Sudeste RJ-ES G2 (2041019). 64
Figura 28. *Box-plot* dos erros das téc. preench. mensal Sudeste SP-MG G1 (2247182). 64
Figura 29. *Box-plot* dos erros das téc. preench. mensal Sudeste SP-MG G2 (2044041). 65
Figura 30. *Box-plot* dos erros das téc. preenchimento mensal Sul G1 (2853035). ...65
Figura 31. *Box-plot* dos erros das téc. preenchimento mensal Sul G2 (2750010). ...65
Figura 32. Correlação linear entre a precipitação média diária e a média dos erros medianos das técnicas de preenchimento. 70
Figura 33. Correlação linear entre o percentual de dias chuvosos e a média dos erros medianos das técnicas de preenchimento. 70
Figura 34. Razão entre a média dos erros medianos das técnicas de preenchimento e a média diária por região. 72
Figura 35. Gráfico do erro médio mensal causado por falhas de extensão variada (dia) nas diferentes regiões de estudo. 73
Figura 36. Convergência do erro médio diário de 1.000 séries sintéticas com uma falha diária. 76
Figura 37. Número de dias máximo de falhas diárias com erro mensal inferior ao das técnicas de preenchimento para cada região. 79
Figura 38. Gráfico *box-plot* do número de dias máximo de falhas diárias com erro mensal inferior ao das técnicas de preenchimento para cada região. 79
Figura 39. *Box-plot* do máximo de falhas diárias com erro mensal inferior ao das técnicas de preenchimento (esq.) e precipitação média histórica mensal (dir.) na região Norte G1. 82
Figura 40. *Box-plot* do máximo de falhas diárias com erro mensal inferior ao das técnicas de preenchimento (esq.) e precipitação média histórica mensal (dir.) na região Norte G2. 83
Figura 41. *Box-plot* do máximo de falhas diárias com erro mensal inferior ao das técnicas de preenchimento (esq.) e precipitação média histórica mensal (dir.) na reg. Centro-Oeste G1. 83
Figura 42. *Box-plot* do máximo de falhas diárias com erro mensal inferior ao das técnicas de preenchimento (esq.) e precipitação média histórica mensal (dir.) na reg. Centro-Oeste G2. 83
Figura 43. *Box-plot* do máximo de falhas diárias com erro mensal inferior ao das técnicas de preenchimento (esq.) e precipitação média histórica mensal (dir.) na região Nordeste G1. 84
Figura 44. *Box-plot* do máximo de falhas diárias com erro mensal inferior ao das técnicas de preenchimento (esq.) e precipitação média histórica mensal (dir.) na região Nordeste G2. 84
Figura 45. *Box-plot* do máximo de falhas diárias com erro mensal inferior ao das técnicas de preenchimento (esq.) e precipitação média histórica mensal (dir.) na região RJ-ES G1. 84

Figura 46. *Box-plot* do máximo de falhas diárias com erro mensal inferior ao das técnicas de preenchimento (esq.) e precipitação média histórica mensal (dir.) na região RJ-ES G2. 85
Figura 47. *Box-plot* do máximo de falhas diárias com erro mensal inferior ao das técnicas de preenchimento (esq.) e precipitação média histórica mensal (dir.) na região SP-MG G1. 85
Figura 48. *Box-plot* do máximo de falhas diárias com erro mensal inferior ao das técnicas de preenchimento (esq.) e precipitação média histórica mensal (dir.) na região SP-MG G2. 85
Figura 49. *Box-plot* do máximo de falhas diárias com erro mensal inferior ao das técnicas de preenchimento (esq.) e precipitação média histórica mensal (dir.) na região Sul G1. 86
Figura 50.*Box-plot* do máximo de falhas diárias com erro mensal inferior ao das técnicas de preenchimento (esq.) e precipitação média histórica mensal (dir.) na região Sul G2. 86

QUADROS

Quadro 1. Dados das estações analisadas (continua) 47

TABELAS

Tabela 1. Extensão das séries e coeficiente de determinação do ajuste linear da curva dupla massa de cada estação (em destaque os melhores ajustes por região) (continua). 53
Tabela 2. Estatísticas Descritivas dos erros de preenchimento (continua) 66
Tabela 3. Percentual de dias chuvosos de cada estação analisada em cada região 74
Tabela 4. Estatísticas Descritivas de 1 falha diária em 1 mês qualquer em cada região analisada 75

1 INTRODUÇÃO

A água é um recurso essencial à manutenção da vida e ao progresso da humanidade. Apesar de ser um recurso renovável, é crescente a sua demanda dado o crescimento populacional, industrial e agrícola, o que pode levar a conflitos caso não haja uma gestão apropriada do recurso.

O conhecimento da disponibilidade hídrica é diretamente relacionado à ciência da hidrologia que, por sua vez, estuda os fenômenos que compõem o ciclo hidrológico. Para que seja desenvolvida qualquer pesquisa hidrológica, ou que sejam desenvolvidos estudos e projetos na área, é fundamental o monitoramento sistemático dos fenômenos naturais, mais especificamente, os que compõem o ciclo hidrológico.

Dentre eles, os principais são a precipitação, a evapotranspiração, a infiltração, o fluxo subterrâneo, as vazões nas calhas dos rios e o transporte de substâncias nos fluxos entre as diferentes fases do ciclo. Portanto, é essencial que sejam observados e medidos tais fenômenos para o apropriado entendimento e gestão dos recursos hídricos.

A observação sistemática de tais fenômenos exige custos elevados já que a manutenção e operação de uma rede demanda uma grande quantidade de equipamentos e recursos humanos capacitados, sobretudo num país de dimensões continentais como o Brasil. A instalação de redes de monitoramento hidrológico no Brasil é uma atividade que vem se desenvolvendo ao longo das últimas décadas e, ainda assim, há necessidade constante de expansão da rede. Não bastasse a necessidade de expansão territorial e adensamento, é necessária também, a operação constante da rede para que os históricos das séries hidrológicas sejam cada vez mais representativos do seu comportamento estocástico.

É notório que diante da complexidade logística e técnica existente no monitoramento de variáveis hidrometeorológicas, em uma área tão abrangente e durante períodos tão longos, a falha na observação de tais variáveis é esperada e inevitável. São diversos os fatores que influenciam para que haja erros no monitoramento, sejam humanos ou tecnológicos.

Dado que um dos principais fenômenos dentro do ciclo hidrológico terrestre é a precipitação, e que falhas são inerentes no processo de monitoramento, o presente

estudo tem interesse de aprofundar o conhecimento dos efeitos que falhas no monitoramento da precipitação geram nas estatísticas da série. Almeja-se compreender quais são as variações que as falhas causam nas estatísticas da série de estações em diferentes localizações climáticas e em diferentes épocas do ano.

Além disso, o presente estudo entende que é praxe no desenvolvimento de estudos e projetos hídricos a adoção de métodos de preenchimento de falhas em séries de precipitação sem o devido conhecimento de sua eficácia. Portanto, pretende-se avaliar a viabilidade da utilização da alternativa de *não preenchimento* (ou preenchimento com 0 mm), uma vez constatada a sua eficácia em cada caso.

2 DIRETRIZES DA PESQUISA

2.1 QUESTÃO DE PESQUISA

Qual o efeito do não preenchimento de falhas diárias em totais mensais de precipitação em comparação às metodologias de preenchimento do total mensal usualmente utilizadas?

2.2 OBJETIVOS DA PESQUISA

Avaliar o efeito de falhas em séries diárias de precipitação no cálculo do total mensal (atribuindo zero precipitação às falhas) em comparação a utilizar formulações matemáticas empíricas lineares usualmente utilizadas para a estimativa do total mensal precipitado.

2.3 HIPÓTESE

A hipótese do trabalho é que, para determinados meses do ano, em determinadas regiões e dependendo da extensão da falha (dias), tem-se um menor erro no total mensal precipitado ao não se realizar o preenchimento do mês em comparação ao preenchimento pelos métodos usuais tendo em vista o conhecimento da estimativa do erro de não preenchimento, ou seja, assumir que nos períodos de falha diária não ocorreu precipitação.

2.4 PRESSUPOSTO

As séries de dados a serem utilizadas nas análises serão obtidas a partir da plataforma Hidroweb (AGÊNCIA NACIONAL DE ÁGUAS, 2019), buscando-se estações que apresentem um histórico de dados brutos diários, sem falhas com uma extensão maior ou igual a 10 anos. Pressupõe-se que essas séries de dados não foram preenchidas, assim como não apresentam problemas de tendenciosidade, não tiveram mudanças significativas de localização durante sua operação e são representativas de sua região.

2.5 DELIMITAÇÕES

O trabalho delimita-se à definição da metodologia e aplicação da técnica de avaliação do impacto das falhas diárias aos totais mensais precipitados em um determinado número de estações pluviométricas para a sua validação.

2.6 LIMITAÇÕES

São limitações do trabalho as técnicas de monitoramento e observação do fenômeno que apresentam erros do processo de observação, e que por melhor que sejam operadas as estações, poderão apresentar tendências que distanciam as observações da realidade, por menor que sejam.

Outra limitação deste estudo é o uso de séries de dados brutos, que apesar de avaliadas e escolhidas a partir de sua melhor qualidade, são utilizadas como parâmetro de avaliação das técnicas de preenchimento, podendo ainda assim conter erros que afetarão a avaliação do desempenho das técnicas de preenchimento.

2.7 DELINEAMENTO

O trabalho será realizado em etapas, as quais são representadas a seguir:

a) Pesquisa bibliográfica;
b) Definição da metodologia de análise de não-preenchimento (preenchimento com zero precipitação);
c) Aplicação da metodologia em determinado número de estações;
d) Análise dos resultados com a avaliação e possível validação da metodologia;
e) Considerações finais e recomendações.

3 REVISÃO BIBLIOGRÁFICA

3.1 PRECIPITAÇÃO

Um dos principais fenômenos estudados na hidrologia é a precipitação. A precipitação compreende toda a água proveniente da atmosfera que atinge a superfície terrestre, podendo ser classificada em suas diferentes formas que dependem do seu estado físico: neblina, chuva, granizo, saraiva e neve.

A hidrologia estuda a precipitação em suas principais características, sendo analisada a duração (t) do fenômeno, o total precipitado (P) e a distribuição temporal e espacial. Decorrem dessas características a intensidade do fenômeno (i = P/t), a frequência de ocorrência e demais estatísticas das séries. No estudo da precipitação, por ser um fenômeno aleatório, se pressupõe que seu tratamento seja feito a partir de análises estatísticas.

A atmosfera, por ser constituída por vapor de água, armazena um volume de água, que ao condensar, permite a ocorrência de precipitações. Diferentes processos regem a sua ocorrência, portanto diferentes condições devem ser satisfeitas para que ocorra o crescimento de gotículas na atmosfera e que ocorra a precipitação. As precipitações podem ser (SMITH,1992):

- Convectivas: São chuvas de pequena duração e grande intensidade, abrangendo áreas menores.
- Orográficas: São chuvas de pequena intensidade e de grande duração, que cobrem pequenas áreas.
- Frontais ou ciclônicas: São chuvas de grande duração, atingindo grandes áreas com intensidade média.

Dependendo do tipo de precipitação que ocorre em uma determinada região, maiores ou menores são as correlações entre os dados de precipitação de estações meteorológicas vizinhas. Portanto, dependendo da região e da época do ano, podem ocorrer mais ou menos precipitações do tipo frontais ou convectivas, e em determinados casos, orográficas. Dessa forma, é importante entender os fenômenos de formação das precipitações e quais os tipos de precipitações que ocorrem em cada região para possibilitar ter-se uma avaliação preliminar da potencial qualidade dos métodos de preenchimento.

3.2 PLUVIOMETRIA

Para que haja o conhecimento das grandezas observadas na hidrologia, devem ser realizadas observações ou medições dos fenômenos em campo. A pluviometria se constitui na coleta dos dados hidrológicos que caracterizam a pluviosidade de uma região, e é usualmente realizada a partir de estações pluviométricas convencionais ou automáticas. Segundo Bertoni e Tucci (2007), as grandezas observadas que caracterizam a precipitação são as seguintes:

Altura pluviométrica (P): é a altura média da lâmina de água precipitada que recobriria a região onde ocorreu a precipitação não admitindo quaisquer perdas de água do sistema, como por exemplo, por infiltração, evaporação ou escoamento. A unidade de medida utilizada pelo sistema métrico é o milímetro.

Duração (t): é o intervalo de tempo durante o qual ocorre a precipitação. As unidades utilizadas são o minuto ou a hora.

Intensidade (i): é o quociente entre a precipitação pelo intervalo de tempo. As unidades de medidas usuais são o mm/h ou o mm/min.

Os aparelhos usualmente utilizados são os pluviômetros para medidas diárias e os pluviógrafos que permitem medidas no tempo de forma automática em intervalos de minutos, os disdrômetro, e, por fim, o radar meteorológico que mede a precipitação no tempo e no espaço. O radar possui erros ainda relativamente grandes, não permitindo uma quantificação exata, no entanto por terem uma espacialização contínua permitem a interpolação da informação (CALVETTI; BENETI; PEREIRA FILHO, 2003).

Os pluviômetros são aparelhos dotados de uma superfície de captação horizontal delimitada por um anel metálico e de um reservatório para acumular a água recolhida. Ele totaliza a altura de precipitação acumulada entre os intervalos de observação que é usualmente um por dia (AGÊNCIA NACIONAL DE ÁGUAS, 2012. É um equipamento operado por um observador que verifica o total precipitado e anota em caderneta de campo suas observações.

O pluviógrafo, da mesma forma que o pluviômetro, também é composto por uma superfície de captação, necessitando ou não acumular a água precipitada. O mais comum nos dias de hoje é o pluviógrafo de cubas basculantes, mas também

existem outros modelos como o de boia e o de balança. O pluviógrafo de cubas basculantes tem a água precipitada dirigida a um conjunto de duas cubas suspensas por um eixo. Quando uma das cubas atinge um determinado peso, o conjunto rotaciona e a cuba oposta passa a se posicionar debaixo do tubo proveniente da superfície coletora, ocorrendo o movimento de forma sucessiva. A cada movimento basculante é realizado um registro de forma analógica ou digital do valor precipitado referente a um movimento (SANTOS et al., 2001).

Os pluviógrafos registram as variações de precipitação ao longo do tempo de forma automática permitindo a escolha da resolução temporal através do ajuste do equipamento. Os aparelhos analógicos gráficos estão em desuso, enquanto os mais utilizados hoje em dia são os aparelhos digitais que armazenam os dados dos sensores, podendo ser conectados a um sistema de transmissão via telemetria, permitindo o acompanhamento do monitoramento em tempo real.

Dentre os métodos de estimativa da precipitação por sensoriamento remoto, o satélite TRMM (*Tropical Rainfall Measuring Mission*), lançado em 1990, permite obter dados de precipitação pluvial com uma resolução espacial de 28 x 28 km e resolução temporal em escala mensal e diária desde o ano 1997 (PEREIRA et al., 2016). O TRMM é formado por diversos produtos de estimativas de chuvas, cujo objetivo é o melhoramento da medição da precipitação tropical por sensoriamento remoto (SODRÉ et al. 2012).

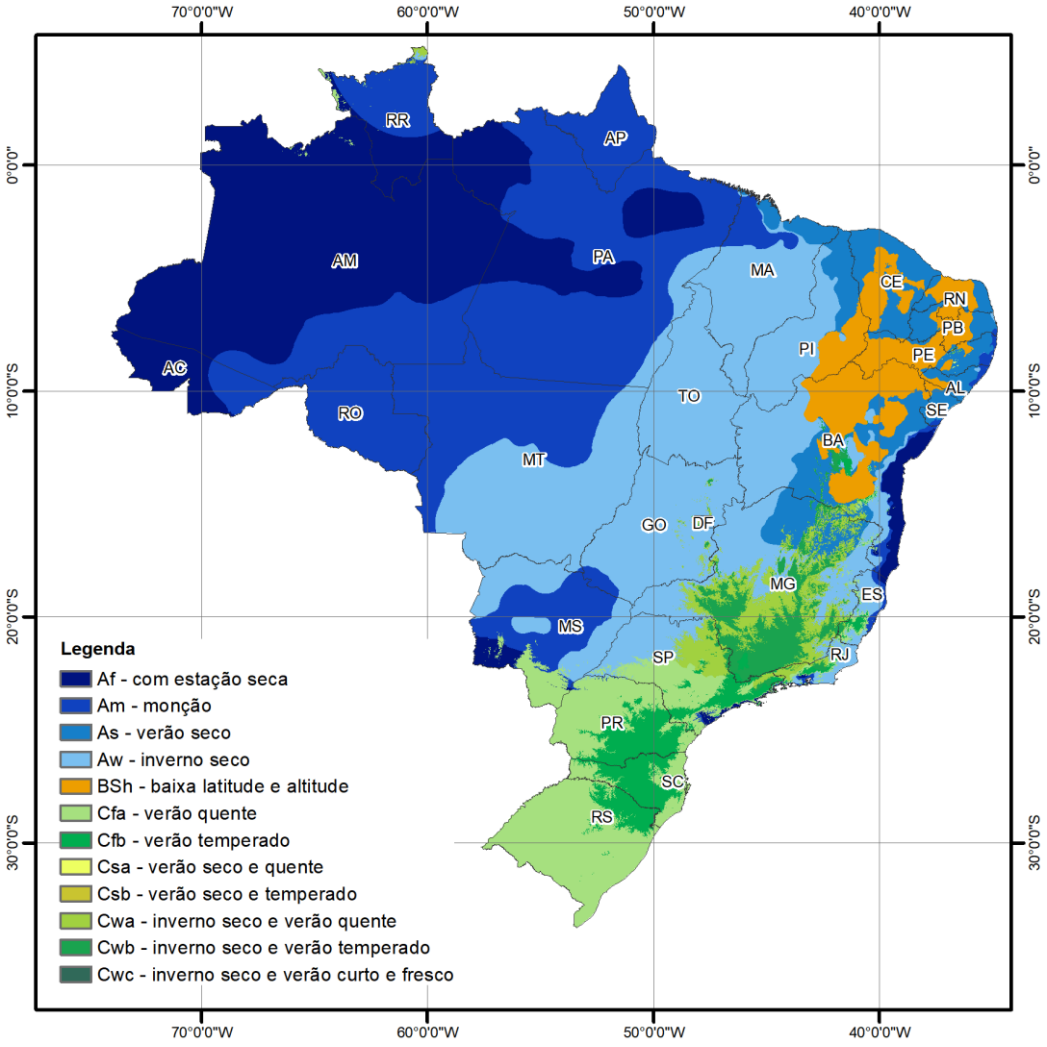
Dentre as fontes de dados de precipitação, as mais utilizadas em trabalhos de hidrologia ainda são os pluviômetros convencionais e automáticos, apesar de haver um grande crescimento do uso de dados oriundos de radar para a alimentação de modelos hidrológicos. Portanto, cabe um estudo mais aprofundado dos erros provenientes desses equipamentos mais difundidos, dadas as limitações inerentes no processo de observação de um fenômeno heterogêneo tanto temporalmente, quanto espacialmente e, por isso, são apresentados os tipos de erros deste processo mais a diante.

3.2.1 Climatologia e Pluviometria Brasileira

A classificação dos diferentes climas do Brasil permitiu realizar a primeira análise de como se comportam as precipitações temporal e espacialmente em todo o território brasileiro. Os tipos de climas, segundo Köppen, são simbolizados por dois

ou três caracteres. O primeiro indica a zona climática e é definido pela temperatura e pela precipitação e o segundo considera a distribuição da precipitação. O terceiro é a variação sazonal da temperatura. A Figura 1 apresenta os diferentes climas presentes no território brasileiro. São três zonas distintas: Grupo A, zona tropical; Grupo B, zona seca; e Grupo C, zona temperada. Estas zonas são sub-classificadas entre os diferentes climas, conforme os aspectos mencionados de precipitação e temperatura.

Figura 1. Classificação climatológica do Brasil de acordo com os critérios de Koppen (1936)



Adaptado de Alvares et al. (2014).

O grupo A apresenta 12 grupos diferentes de clima no Brasil. A maior parte do território é do grupo tropical (81,4%), sendo que o clima Af (Tropical Equatorial) cobre 22,6% do território. Neste tipo de clima, a sazonalidade não é marcante, com grande homogeneidade durante o ano quanto às precipitações (ALVARES et al., 2014).

Já o clima Am (Tropical Monçônico) é o mais representativo do território brasileiro, cobrindo 27,5%. Se apresenta em diversas regiões do Brasil, com volumes anuais precipitados marcantes no estado do Amapá, Rondônia e Roraima. Estende-se também a regiões do Pará, estando presente inclusive no Mato Grosso do Sul. Este clima está presente em diversas outras pequenas áreas do país, variando de 1.200 mm por ano a 3.300 mm (ALVARES et al., 2014).

O clima Aw (Tropical Savânico) está presente também em boa parte do território brasileiro (25,8%) com precipitações anuais entre 1.300 e 1.900 mm. Está localizado na região central do país e apresenta sazonalidade bastante marcada com as precipitações entre dezembro e fevereiro e o inverno entre maio e setembro com o tempo bastante seco, especialmente de junho a agosto. O clima do tipo As (também chamado de Tropical Savânico) está presente em apenas 5,5% do território, apresentando o inverno chuvoso com grande variação de precipitação dependendo do local entre 700 mm e 1.500 mm (ALVARES et al., 2014).

O grupo B, zona semi-árida, é um grupo característico do nordeste do Brasil com precipitações abaixo de 800 mm anuais. O único clima do grupo B encontrado no Brasil é o BSh (Seco Semiárido Quente), localizado majoritariamente no interior do território, com exceção de 150 km na costa do Rio Grande do Norte (abaixo de 650 mm anuais) (ALVARES et al., 2014).

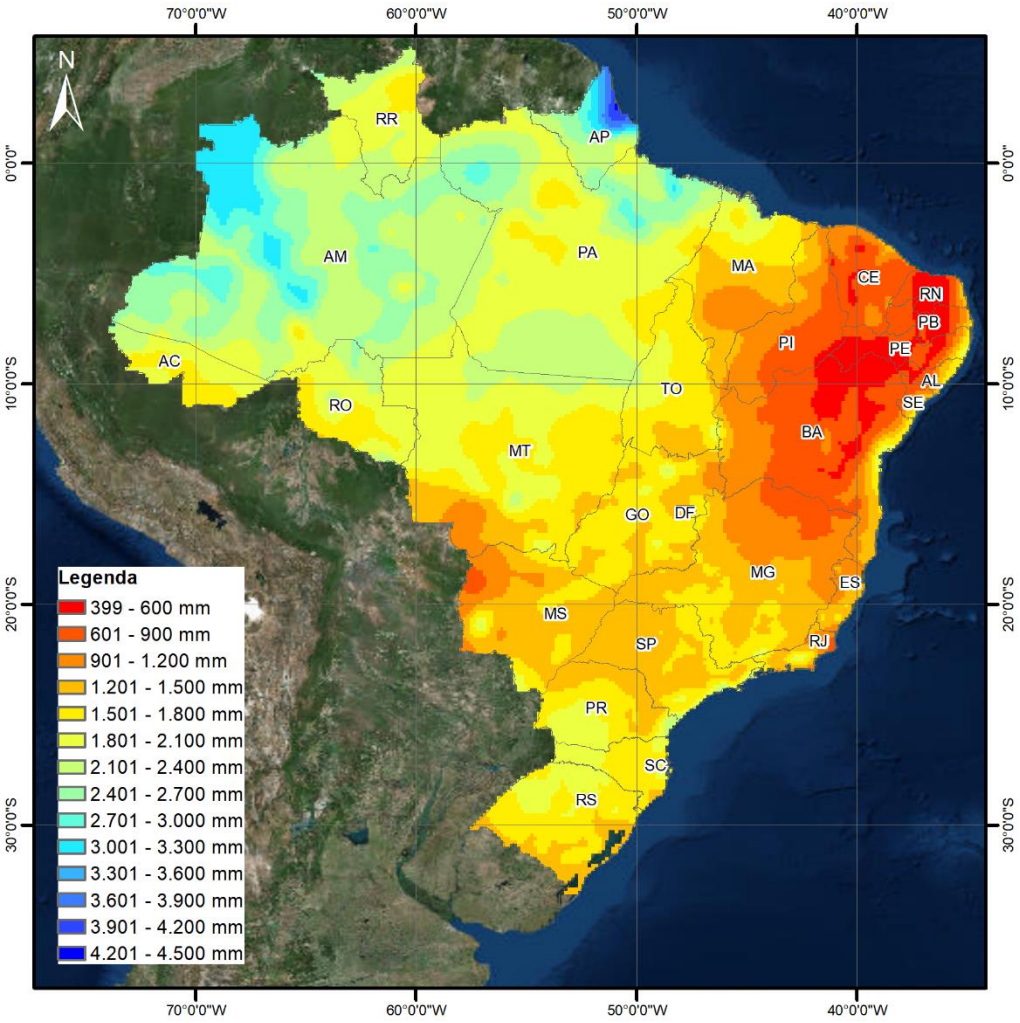
O grupo C, zona temperada ou sub-tropical, está presente em 13,7% do território, predominantemente na região sul do país. Porto Alegre pode ser considerada uma típica localidade do clima Cfa (Temperado sem Estação Seca com Verão Quente), com precipitações mensais bem distribuídas entre 100 e 170 mm (ALVARES et al., 2014).

Os climas Cwa e Cwb se distinguem pelo primeiro ter um verão quente, enquanto o segundo um verão fresco, sendo ambos característicos por terem invernos secos. O clima Cwa é característico da região sudeste do Brasil, enquanto o clima Cwb é encontrado do estado de São Paulo ao da Bahia.

O clima Csa e Csb (Temperado com Verão Seco) é observado em poucos lugares em Pernambuco e Alagoas, com precipitações anuais que chegam a 710 mm concentradas no outono entre abril e junho.

A Figura 2 apresenta a precipitação anual total no Brasil ilustrando a distribuição das características pluviométricas acima descritas através dos diferentes climas do país. Observam-se as maiores precipitações anuais totais no Norte do país, enquanto as menores estão localizadas no Nordeste. Ao Sul, as precipitações variam de 1.000 mm a 2.200 com um gradiente Norte-Sul.

Figura 2. Mapa a partir das isoietas de Precipitação Média Anual Total no Brasil (1977-2006)



Adaptado de Pinto et al. (2011)

Importante é o entendimento da distribuição espacial e sazonal da precipitação no território brasileiro para se poder entender as capacidades das técnicas de preenchimento de dados mensais de precipitação com o uso de estações vizinhas e adaptar a distância máxima para que haja boa correlação entre os dados, assim como avaliar a possibilidade do uso de tais técnicas.

3.3 ANÁLISE DE CONSISTÊNCIA

3.3.1 Tipos de erros

Nas estações de monitoramento hidrometeorológico, o fenômeno observado por instrumentos e em função das deficiências da medida, em geral difere do verdadeiro valor e essa diferença é chamada de “erro de observação”. Durante a manipulação dos dados, podem ocorrer equívocos na transcrição das informações, denominando-se os “erros de transcrição” (SANTOS et al., 2001).

Smith (1992) explica que, virtualmente, todos os pluviômetros e pluviógrafos incorrem em erros devido às próprias modificações do campo de ventos causadas pelo aparelho. A magnitude dos erros depende da velocidade do vento, da topografia local e do tipo de precipitação. Realizar medidas de precipitação pode ser uma tarefa difícil dependendo do local a ser monitorado.

Os erros de observação podem ser classificados em três categorias de acordo com a teoria dos erros: erros grosseiros, erros sistemáticos, e erros fortuitos ou aleatórios (JOINT COMMITTEE FOR GUIDES IN METROLOGY, 2010).

Os erros grosseiros derivam de erros humanos e falhas nos sensores e registradores. Geralmente, têm maior magnitude que a exatidão[1] dos aparelhos não obedecendo a um padrão preestabelecido. Medidas de incerteza não são projetadas para levar em conta tais erros. Segundo Santos et al. (2001), os erros grosseiros mais comuns da pluviometria são:

- Derramamento de água na coleta – pluviômetro;
- Erro de contagem das provetas cheias (leitura) – pluviômetro;
- Vazamento no mecanismo da torneira – pluviômetro;
- Falta de coleta em certos dias (p. ex.: domingos e feriados) – pluviômetro;
- Boia presa (soltando-se sozinha depois) – pluviógrafo de boia;
- Correções por conta do observador, para compensar água derramada – pluviômetro;
- Leitura em horários diferentes – pluviômetro;

[1] Exatidão é a aptidão de um instrumento para dar respostas próximas ao valor verdadeiro do mensurando. É a capacidade que o instrumento de medição tem de fornecer um resultado correto (JOINT COMMITTEE FOR GUIDES IN METROLOGY, 2010).

- Transbordamento– pluviômetro/pluviógrafo de balança;
- Preenchimento de registro sem leitura (dado inventado) – pluviômetro; e
- Retirada da água por estranhos ou animais pluviômetro/pluviógrafo de balança.

A maioria dos erros grosseiros listados são erros humanos, ocorrendo em sua maioria nos pluviômetros, cuja operação depende de observadores. Esse tipo de erro pode ser difícil de detectar e, dependendo de sua magnitude, se for pequeno, pode ser mascarado por variações aleatórias (JOINT COMMITTEE FOR GUIDES IN METROLOGY, 2010). Comparando os dados aos de estações vizinhas, há um risco de correção de valores corretos, mas aparentemente incoerentes (SANTOS et al., 2001).

Os erros sistemáticos são associados ao instrumento e têm como característica principal a apresentação destes seguindo um determinado comportamento dependendo das condições de medidas. Se um erro sistemático se origina de um efeito reconhecido de uma grandeza de influência em um resultado de medição, o efeito pode ser quantificado e, se for significativo com relação à exatidão requerida da medição, uma correção pode ser aplicada para compensar o efeito. Os erros sistemáticos não podem ser eliminados, no entanto podem ser reduzidos (JOINT COMMITTEE FOR GUIDES IN METROLOGY, 2010).

Ainda, esses erros podem ser resultado de calibrações defeituosas, instalação inadequada ou mudança nas características dos aparelhos registradores. Um exemplo de situação pode ser um aparelho instalado próximo a um obstáculo, como uma parede ou uma árvore: sempre que a precipitação ocorrer no sentido do obstáculo em direção ao aparelho, o valor medido será inferior ao real, enquanto em outras situações o mesmo não ocorre. Enquadram-se nesse tipo de erro na pluviometria (SANTOS et al., 2001):

- Falta de regulagem no relógio do pluviógrafo;
- Proveta trocada;
- Exposição inadequada (debaixo de árvores, telhados, etc.);
- Mudança nas condições de exposição (troca de local); e
- Desnivelamento da superfície de exposição.

Por fim, os erros aleatórios, ou fortuitos, são erros de observação que estão dentro da margem de exatidão dos aparelhos ou da capacidade humana de observar. Esses erros podem estar sujeitos a variações estocásticas seguindo a teoria dos erros ou imprevisíveis. Além disso, são em geral independentes da magnitude do valor observado, que levam ao alargamento da curva de distribuição dos resultados (aumento da curtose), relacionando-se à inexatidão. Embora não possam ser compensados, podem ser reduzidos com o aumento do número de observações. São erros aleatórios ou fortuitos na pluviometria (SANTOS et al., 2001):

- Evaporação da água no pluviômetro ou no pluviógrafo de balança;
- Influência do vento na quantidade de chuva captada; e
- Variações aleatórias do relógio.

3.3.2 Análise dos dados de precipitação

Uma estação pluviométrica para monitoramento tem como objetivo a obtenção de uma série ininterrupta de registros de precipitação ao longo do tempo. Devido às dificuldades já expostas, podem ocorrer períodos sem informações ou com falhas constatadas de observação. São dois os tipos de lacunas geradas na série de dados: (a) a inexistência do registro por falhas humanas ou dos aparelhos de registro; ou (b) a lacuna gerada devido ao descarte e necessidade de substituição de dados suspeitos.

Dadas tais circunstâncias, há a necessidade de submeter as informações coletadas à análise, anteriormente a sua utilização. A análise de dados brutos coletados no ramo da hidrologia, com eventual detecção e correção de erros é chamada de análise de consistência.

A análise de consistência deve preceder todo estudo hidrológico dadas as dificuldades das instituições responsáveis por tal trabalho na eliminação de erros de observação e posterior tratamento dos dados. É a tarefa inicial em todo trabalho de hidrologia, a busca por informações e posterior análise de consistência, definindo de antemão a potencial qualidade do estudo hidrológico (MINISTÉRIO DAS MINAS E ENERGIA, 1984).

A análise de consistência tradicionalmente é realizada com uma seleção de estações situadas em uma mesma bacia hidrográfica, ou no caso das precipitações, não necessariamente na mesma bacia hidrográfica, mas em uma área homogênea quanto ao regime pluviométrico.

A sistemática inicia com a seleção das estações alvo de análise sendo separadas em duas classes: estações principais e estações de apoio. As estações principais são os pontos de monitoramento cujos resultados do estudo hidrológico serão baseados, sendo validados pelas estações de apoio (MINISTÉRIO DAS MINAS E ENERGIA, 1984). Segundo Santos (2001) e o Ministério de Minas e Energia (1984), as principais características a serem observadas na seleção das estações pluviométricas são:

- Localização na bacia alvo dos estudos;
- Suficiente extensão do período de observação;
- Escolha de estações com pouca diferença de altitude;
- A região para ser considerada homogênea não deve apresentar variações abruptas de relevo que possam causar algum distúrbio na circulação do fenômeno, como é o caso de duas estações separadas por cadeias de montanhas onde efeitos orográficos na precipitação são significativos;
- Poucos períodos sem informações;
- Com pouca ou nenhuma alteração no equipamento e no local de instalação.

As estações de apoio têm como finalidade principal verificar a existência de erros nos históricos das estações principais e possibilitar a correção ou validar os dados existentes. A detecção e a correção de erros grosseiros de observações pluviométricas são realizadas em uma análise preliminar com a conferência de boletins de campo e checagem dos dados transcritos ao computador. Devido à maior independência entre eventos chuvosos, a continuidade do fenômeno, como é no caso dos históricos de vazões, não pode ser verificada, o que torna o processo de análise mais complexo (MINISTÉRIO DAS MINAS E ENERGIA, 1984; SANTOS et al., 2001; BERTONI; TUCCI., 2007).

A comparação de dados entre estações próximas é a técnica utilizada para o levantamento de suspeitas de erros, permitindo eventualmente a descoberta de erros grosseiros. No entanto, apesar de ser esperada certa homogeneidade numa região, esta dependerá do clima e da época do ano, em que as precipitações podem ter diferentes características, o que pode aumentar a heterogeneidade da região. Isso ocorre, por exemplo, na região sul do Brasil quando se realiza uma comparação entre chuvas frontais, menos concentradas e mais homogêneas, com maior ocorrência no inverno e as chuvas convectivas, que são mais concentradas no tempo e no espaço.

Os erros sistemáticos, geralmente quando passam a ocorrer durante a operação da estação apresentam uma mudança na tendência observada. Caso o erro sistemático persista desde o início da operação do posto, esse pode ser de mais difícil descoberta. Em geral, os erros durante a operação ocorrem na mudança das condições de exposição, como o crescimento de uma árvore ao lado da estação ou a mudança de seu local. Para a constatação desse erro, compara-se o seu histórico acumulado ao de outras estações através da curva duplo-acumulada (SANTOS et al., 2001).

De acordo com as orientações do Ministério de Minas e Energia (1984), a recuperação de totais anuais para anos incompletos deve ser feita apenas quando os totais mensais faltantes representam menos de 15% do total anual. Apesar das possibilidades de correção, através dos métodos apresentados a seguir, deve se ter em mente que cada série possui características próprias. Portanto, os desvios que ocorrem entre a série analisada e a série de apoio são normais e aceitáveis, desde que não sejam desvios sistemáticos ou não sejam resultados de erros grosseiros de observação.

3.3.3 Análise preliminar dos dados hidrológicos e técnicas de preenchimento de falhas

Para realizar a análise preliminar dos dados hidrológicos disponíveis, podem se utilizar diversas técnicas de representação gráfica ou descritiva a partir do cálculo das estatísticas das séries. Algumas formas de apresentação são descritas abaixo (NAGHETTINI; PINTO, 2007).

Sumário numérico e estatísticas descritivas: medidas de tendência central; medidas de dispersão; medidas de assimetria e curtose.

- Apresentação gráfica: diagrama de linha; diagrama uniaxial dos pontos; histograma; polígono de frequências; diagrama de frequências relativas acumuladas; curva de permanência.
- Métodos exploratórios: diagrama *box-plot*; diagrama ramo-e-folha.
- Associação entre variáveis: diagrama de dispersão; diagrama quantis-quantis (Q-Q).

Para se verificar e validar uma possível homogeneidade e correlação entre duas estações próximas é possível utilizar o coeficiente de correlação linear ou correlação de Pearson. Duas variáveis apresentam uma correlação linear quando os pontos do diagrama de dispersão se aproximam de uma reta.

No caso específico, o total precipitado mensal de duas estações analisadas para uma série mensal deve se aproximar de uma reta, caso haja alta correlação entre os dois (grau de associação linear). Essa correlação pode ser positiva, caso seja diretamente proporcional ou negativa caso seja inversamente (domínio [-1;1]) (NAGHETTINI; PINTO, 2007).

O coeficiente de correlação linear ρ também chamado de covariância normalizada é obtido por:

$$\rho_{X,Y} = \frac{\sigma_{x,y}}{\sigma_x \sigma_y} \tag{1}$$

Em que,

$\rho_{X,Y}$ – coeficiente de correlação linear entre as variáveis X e Y;

$\sigma_{x,y}$ – covariância entre as variáveis X e Y;

σ_x, σ_y – desvios-padrão das variáveis X e Y, respectivamente.

A covariância amostral entre duas variáveis pode ser estimada pela equação abaixo. Ela pode representar uma medida do grau e do sinal da correlação:

$$S_{X,Y} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{n-1} \tag{2}$$

Em que,

$S_{x,y}$ – covariância amostral entre as variáveis X e Y.

Os desvios-padrão das amostras podem ser calculados pela equação seguinte:

$$S_X = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}} \quad (3)$$

Em que,

S_x – desvio padrão amostral.

Método da dupla-massa

Para a constatação de erros na série analisada, compara-se o seu histórico acumulado ao de outras estações através da curva do gráfico de dispersão entre os valores continuamente acumulados de cada estação (ou estação média fictícia) e os da estação analisada, chamado de dupla-massa ou duplo-acumulada. Segundo Santos (2001), caso se observe um desvio na declividade da reta duplo-acumulativa originada de duas estações, as observações mais antigas podem ser corrigidas pela expressão:

$$P_{est} = \frac{m_1}{m_2} P_{obs} \quad (4)$$

Em que,

P_{obs} – precipitação observada no período a ser corrigido;

P_{est} – precipitação estimada no período a ser corrigido; e

m_1 e m_2 – coeficientes angulares da reta nos períodos recente e antigo.

Método do Vetor Regional

Outro método de análise de consistência pluviométrica é o “método do vetor regional”. Desenvolvido por Hiez (1977) e aplicado à chuva por Hiez e Rancan (1983), o vetor regional foi definido como “uma série cronológica, sintética, de índices pluviométricos anuais (ou mensais), oriundos da extração por um método de máxima verossimilhança da informação contida nos dados de um conjunto de estações agrupadas regionalmente” (GOVERNO DO ESTADO DE MINAS GERAIS, 2015).

O método consiste na determinação de dois vetores [L] (vetor coluna com *m* linhas, representando *m* observações) e [C] (vetor linha com *n* colunas, representando *n* postos), cuja multiplicação resulta numa matriz de precipitações

[P_e]. Os índices são únicos para toda a região e estão relacionados às alturas precipitadas em cada estação através dos coeficientes. A estimativa da altura precipitada (anual ou mensal), no ano “i”, na estação “j”, resulta $P_{ij} = L_i * C_j$.

O método é adequado apenas para totais anuais ou mensais. Não é usual a utilização do método para tratamento de precipitações diárias, devido às variações espaciais e temporais de chuvas diárias (BERTONI; TUCCI, 2007; SANTOS et al., 2001).

Regressão Linear Múltipla

A técnica da regressão linear múltipla também pode ser utilizada para o preenchimento de falhas no histórico analisado ou também a estimativa do valor correto caso seja constatado um erro:

$$P_{est} = a_0 + \alpha P_A + \beta P_B + \cdots \quad (5)$$

Em que,

P_A, P_B,... – precipitações nas estações A, B, ...;

a_0, α, β,... – coeficientes de ajuste.

A partir dos resultados de correlação obtidos para cada estação de apoio, estas devem ser classificadas em ordem de importância. A classificação terá um peso atribuído não só à correlação, mas também ao período de sobreposição dos dados entre as estações utilizadas.

Ponderação Regional

A ponderação regional relaciona o dado mensal faltante P_{est} a ser estimado da estação analisada, com o dado ocorrido na estação de apoio P_i no mês de ocorrência da falha na estação analisada. A relação é feita através do quociente do dado médio mensal da estação analisada $P_{m\,x}$, pelo dado médio mensal da estação de apoio $P_{m\,i}$, sendo ponderado pelo número de estações de apoio utilizadas (*n*). A expressão a seguir apresenta o cálculo a ser realizado para o preenchimento da falha mensal (ALVES et al., 2006).

$$P_{est} = \frac{1}{n}\sum_{i=1}^{n}\frac{P_{m\,x}}{P_{m\,i}}P_i \quad (6)$$

Em que,

P_i – precipitação mensal das *n* estações;

$P_{m\,x}$ – precipitação média mensal na estação X a ser estimada;

$P_{m\,i}$ – precipitação média mensal nas *n* estações; e

n – número de estações.

Método da razão normal

Paulhus e Kohler (1952) propõem a metodologia, que então é modificada por Young (1992). A metodologia consiste na estimativa do dado mensal faltante P_{est} por meio da relação com o dado ocorrido na estação de apoio no mês de ocorrência da falha pelo quociente do peso ponderado atribuído a cada estação e apoio W_i.

$$P_{est} = \frac{\sum_{i=1}^{n} P_i W_i}{\sum_{i=1}^{n} P_i} \tag{7}$$

Em que,

W_i – peso da estação *i* ponderado pela correlação linear e pela extensão da série; e

n – número de estações.

O peso de cada estação de apoio é calculado a partir da correlação linear ρ_i entre as estações analisada e de apoio. O cálculo é feito com o auxílio do número de meses em que ocorreu sobreposição de dados entre a estação analisada e a estação de apoio a ter o seu peso definido. Dessa forma, é possível levar em consideração o tamanho de dados da série de apoio.

$$W_i = {\rho_i}^2 \left(\frac{m_i - 2}{1 - {\rho_i}^2}\right) \tag{8}$$

Em que,

ρ_i – coeficiente de correlação linear da estação *i* com a estação X;

m_i – número de meses em que ocorreu sobreposição de dados.

Interpolação do inverso da distância

Hubbard (1994) apresenta o método da interpolação do inverso da distância em que o dado mensal faltante a ser preenchido na estação analisada P_{est}, é calculado com base no dado observado na estação de apoio P no mês da falha ponderado pela relação do inverso da distância de cada estação de apoio d_i. Quanto

maior a distância da estação de apoio, menor o seu peso no cálculo do dado faltante. A equação abaixo apresenta a forma de cálculo pelo método do inverso da distância.

$$P_{est} = \frac{\sum_{i=1}^{n} {P_i}/{d_i}}{\sum_{i=1}^{n} {1}/{d_i}} \quad (9)$$

Em que,

d_i – distância da estação *i* com a estação a ser estimada.

A metodologia pode ser alterada para que a ponderação seja feita pelo quadrado da distância.

$$P_{est} = \frac{\sum_{i=1}^{n} {P_i}/{d_i^{\,2}}}{\sum_{i=1}^{n} {1}/{d_i^{\,2}}} \quad (10)$$

Método Tradicional do Reino Unido

O método tradicional do Reino Unido é baseado em comparações com uma única estação vizinha. Ela é escolhida com base na correlação com a estação analisada. É feita a ponderação das médias do mês em que houve a falha para a estação analisada e para a estação de apoio. É assumida uma razão constante para cada mês entre as duas estações, assumindo-se, que a estimativa de precipitação da falha em um determinado mês é um percentual fixo da média da estação de apoio (TABONY, 1983).

Método do Potencial Múltiplo

O método Potencial Múltiplo é uma alternativa à regressão linear múltipla apresentada por Bertoni e Tucci (2007). A expressão utiliza também coeficientes de ajuste, que podem ser obtidos pelo método dos mínimos quadrados, sendo relacionados segundo a expressão abaixo.

$$P_{est} = a_0 \prod_{i=1}^{n} P_i^{\,a_j} \quad (11)$$

Em que,

a_0, a_j – coeficientes de ajuste.

Média aritmética simples

A média aritmética simples soma os totais precipitados nas estações vizinhas P_i no mês de ocorrência da falha na estação analisada P_{est} e divide pelo número de estações de apoio utilizadas no cálculo (*n*). A expressão que representa o método é apresentada abaixo.

$$P_{est} = \frac{1}{n}\sum_{i=1}^{n} P_i \tag{12}$$

Erros Estatísticos

Para a avaliação da eficácia e do desempenho dos métodos aqui descritos, quando comparados os dados estimados aos valores reais podem ser utilizados os seguintes desvios (DÉQUÉ, 2003):

- Viés (V)

$$V = \frac{1}{n}\sum_{i=1}^{n}\left(P_{est_i} - P_{obs_i}\right) \tag{13}$$

- Viés percentual (VP)

$$VP = \frac{\sum_{i=1}^{n}\left(P_{est_i} - P_{obs_i}\right)}{\sum_{i=1}^{n} P_{obs_i}} \tag{14}$$

- Erro absoluto médio (EAM)

$$EAM = \frac{1}{n}\sum_{i=1}^{n}\left|P_{est_i} - P_{obs_i}\right| \tag{15}$$

- Raiz do erro quadrático médio (REQM)

$$REQM = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(P_{est_i} - P_{obs_i}\right)^2} \tag{16}$$

3.3.4 Resultados de avaliações dos métodos de preenchimento apresentados por diferentes autores

Bier e Ferraz (2017) testaram os diferentes métodos descritos acima e apresentaram os resultados da aplicação em estações localizadas no estado do Rio Grande do Sul. Os métodos testados pelos autores foram os seguintes:

- Regressão linear múltipla;
- Ponderação regional;
- Interpolação do inverso da distância;
- Método da razão normal;
- Método tradicional do Reino Unido;
- Média aritmética simples.

Foi criada uma falha sintética de um ano de dados mensais de estações pluviométricas e por meio dos métodos de preenchimento de falhas foram estimados os valores e comparados aos valores reais observados omitidos. Na média entre todas as estações, o método que apresentou os melhores resultados foi o da Ponderação Regional. Os métodos da Interpolação pelo Inverso da Distância, da Média Aritmética e da Razão Normal apresentaram resultados próximos ao da Ponderação Regional quando analisadas as métricas para o cálculo dos erros.

Segundo Bier e Ferraz (2017), outros autores também chegaram a resultados que indicam o método da Ponderação Regional como o mais exato para a estimativa de precipitação (17,1% de erro absoluto médio), no entanto os demais métodos testados tiveram desempenho próximo (PAULHUS; KOHLER, 1952; OLIVEIRA, 2010).

Oliveira et al. (2010) testaram o preenchimento de falhas pela ponderação regional, regressão linear, ponderação regional com base em regressões lineares, vetor regional e alternativamente a regressão linear, testaram também o modelo potencial múltiplo. Foram realizadas diferentes combinações de métodos. Os autores identificaram que os métodos com menores desvios relativos foram obtidos na seguinte ordem:

1. Regressão linear múltipla;
2. Vetor regional combinado com a regressão potencial múltipla;

3. Ponderação regional;
4. Vetor regional combinado com a regressão linear múltipla;
5. Regressão potencial;
6. Ponderação regional com base em regressões lineares;
7. Vetor regional combinado com a ponderação regional; e
8. Vetor regional.

Em estudos realizados por Fernandez (2007), foram analisados os desempenhos de diferentes técnicas de preenchimento em falhas de séries temporais de temperatura máxima, temperatura mínima, temperatura média, umidade relativa e precipitação. Foram criadas falhas sintéticas e aplicados os seguintes métodos estatísticos: regressão múltipla, média simples, Steurer, média de três estações, proporção normal e análise harmônica. Os métodos que apresentaram menores erros e melhores resultados foram os de análise de regressão múltipla, Steurer e média de três estações.

É possível perceber que há diversos estudos que tentam entender e propor quais são as melhores técnicas de preenchimento, onde se identifica preliminarmente dentre todos os estudos, melhores desempenhos utilizando a técnica de regressão linear múltipla, no entanto sem conclusões definitivas quando avaliados em conjunto.

Destaca-se também as inúmeras possibilidades de preenchimento combinando as técnicas existentes, que basicamente são o desdobramento de médias ponderadas, ajuste de equações, sendo majoritariamente relações lineares entre as diferentes técnicas. No entanto, observa-se um baixo grau de complexidade dentre as técnicas mais usuais.

3.4 IMPACTO DE FALHAS DE DADOS NAS ESTATÍSTICAS DAS SÉRIES E ESTUDOS HIDROLÓGICOS

Apesar de haver diversos estudos avaliando o desempenho das diferentes técnicas de preenchimento de falhas de séries de precipitação, são raros os estudos que avaliam o impacto ou o efeito das falhas nas estatísticas das séries, relacionando tal impacto à sua localização geográfica, consequentemente ao clima da região e à época do ano em que a falha ocorre.

São menos raros os estudos que avaliam o impacto das falhas de dados em algoritmos de aprendizado de máquina e em modelos de simulação hidrológica. Como é o caso do estudo de Gill et al. (2007) que avaliaram o efeito de falhas de dados na performance do aprendizado de máquina para a predição hidrológica e Mamédio (2014) que avaliou o impacto de erros dos dados de entrada na eficiência de um modelo hidrológico.

A maioria dos estudos encontrados que avaliam técnicas mais atuais são os estudos sobre Redes Neurais Artificiais (RNA). No entanto, o objetivo do presente trabalho é avaliar o impacto de falhas de dados diários em totais mensais de precipitação e compará-los às metodologias tradicionalmente utilizadas por serem mais usuais e por terem maior abrangência. Portanto, não serão abordadas aqui técnicas mais sofisticadas como o caso das RNA’s.

A premissa por trás do presente estudo é que, é praxe na análise de consistência preliminar a qualquer estudo hidrológico, a identificação de falhas e imediata correção. No entanto, não é avaliada a hipótese de se manter a falha (entendida no presente contexto como falta de dados) como está, na condição de se obter resultados do total mensal precipitado naquele mês mais próximos da realidade frente à alternativa da utilização de técnicas de preenchimento que excluem todo o restante da série no mês em que houve a falha.

Acredita-se que em determinadas situações, o erro que incorre em se utilizar uma estação distante, numa determinada estação do ano em que as chuvas são mais heterogêneas serão maiores do que se manter a falha. Ainda restam dúvidas quanto ao número de falhas diárias necessárias para a decisão entre o preenchimento ou não de um determinado mês. Ficando a critério do hidrólogo a tomada de tal decisão de forma intuitiva.

Nesse sentido, a busca por trabalhos com esse viés não foi bem sucedida, no entanto foi o que motivou o presente estudo a seguir essa linha de pesquisa. Foi encontrado um trabalho publicado pela *World Meteorological Organization* (2007) em que se avaliou o impacto de falhas de dados em normais climatológicas:

Estudo *World Meteorological Organization (WMO)*, 2007

Foram avaliados dois tipos de falhas de dados:

a) Falhas de dados diários que contribuem para o valor do mês de um mês/ano específico; e

b) Falha de dado mensal, incluindo valores não calculados devido à presença de dados diários insuficientes, durante o período em que as normais climatológicas foram calculadas.

O impacto das falhas de dados foi avaliado nos dois casos a partir da estimativa da incerteza na média daquele período provocada ao se apagar de forma aleatória um determinado número de observações. Os parâmetros avaliados foram a média diária máxima e mínima da temperatura, a média da pressão no nível do mar, a média da pressão de vapor e a média diária de duração da incidência solar diária.

Para cada série de dados foram geradas 100 novas séries modificadas. As séries foram modificadas ao se apagar m observações de forma aleatória, com a condição de que pelo menos n de m observações deveriam ser consecutivas.

A média das 100 séries modificadas foi calculada e o intervalo de confiança de 95% para a média das séries modificadas. Para cada m e n, a largura do intervalo de confiança de 95% foi dividida pelo desvio padrão. Esses valores tiveram as médias calculadas entre todas as estações, meses e, no primeiro tipo de avaliação (a), anos. O desvio padrão das séries de dados é definido pelo desvio padrão de todos os dados diários para o mês em questão no caso da análise (a), ou para todos os dados mensais do mês em questão no caso (b).

Os estudos indicam que, no caso estudado da Austrália, o desvio padrão típico de dados diários é de 3ºC para a temperatura máxima e mínima. A mesma avaliação foi feita para as outras normais analisadas. Concluiu-se que a largura do

intervalo de confiança de 95% quando aplicada a falha sintética na série é similar, ou até menor que o erro médio absoluto quando utilizada a normal entre 1961-1990 para prever os valores de 1991-2000.

O estudo apontou que a incerteza gerada pela falta de dados mensais num período médio é ainda menor do que a incerteza gerada pela falta de dados diários num valor mensal, refletindo baixos desvios padrões dos valores mensais. Em geral, séries de dados diários com um número específico de valores consecutivos com falhas mostram uma incerteza um pouco maior em suas médias do que as séries sem valores consecutivos. Isso reflete uma auto-correlação entre a maioria das séries de dados analisadas.

Em resumo, o estudo acima conclui que falhas de dados diários que contribuem para o cálculo mensal, assim como meses faltantes no histórico, geram incertezas pequenas nas normais climatológicas estudadas, podendo ser menores até do que a predição da normal climatológica por meio de normais de períodos anteriores.

De maneira similar ao estudo acima, a expectativa é que ao se gerar falhas diárias sintéticas em históricos de precipitação, a erro absoluto médio gerado no cálculo mensal da precipitação é inferior ao erro absoluto médio gerado pela estimativa de totais mensais de precipitação através de técnicas usuais de preenchimento a partir de formulações matemáticas empíricas lineares.

Enquanto o estudo da WMO trata de dados faltantes históricos para o cálculo da **média** histórica, podendo haver incertezas tanto para mais quanto para menos em tal cálculo, a metodologia proposta neste trabalho implica, necessariamente, que os dados faltantes para o cálculo do **total** (soma) precipitado seja inferior ao valor real. Ou seja, as incertezas geradas por dados faltantes no estudo da WMO podem, eventualmente, se equilibrar dado que os erros são negativos ou positivos, enquanto no presente estudo, os erros gerados sempre subestimaram a precipitação mensal.

Desta forma, as falhas deste estudo são consideradas como precipitação 0 mm. Portanto, espera-se que o impacto que uma determinada falha gerada seja

superior ao do estudo apresentado acima, uma vez que a perda de dados implica em uma subestimativa do valor amostrado.

4 MATERIAIS E MÉTODOS

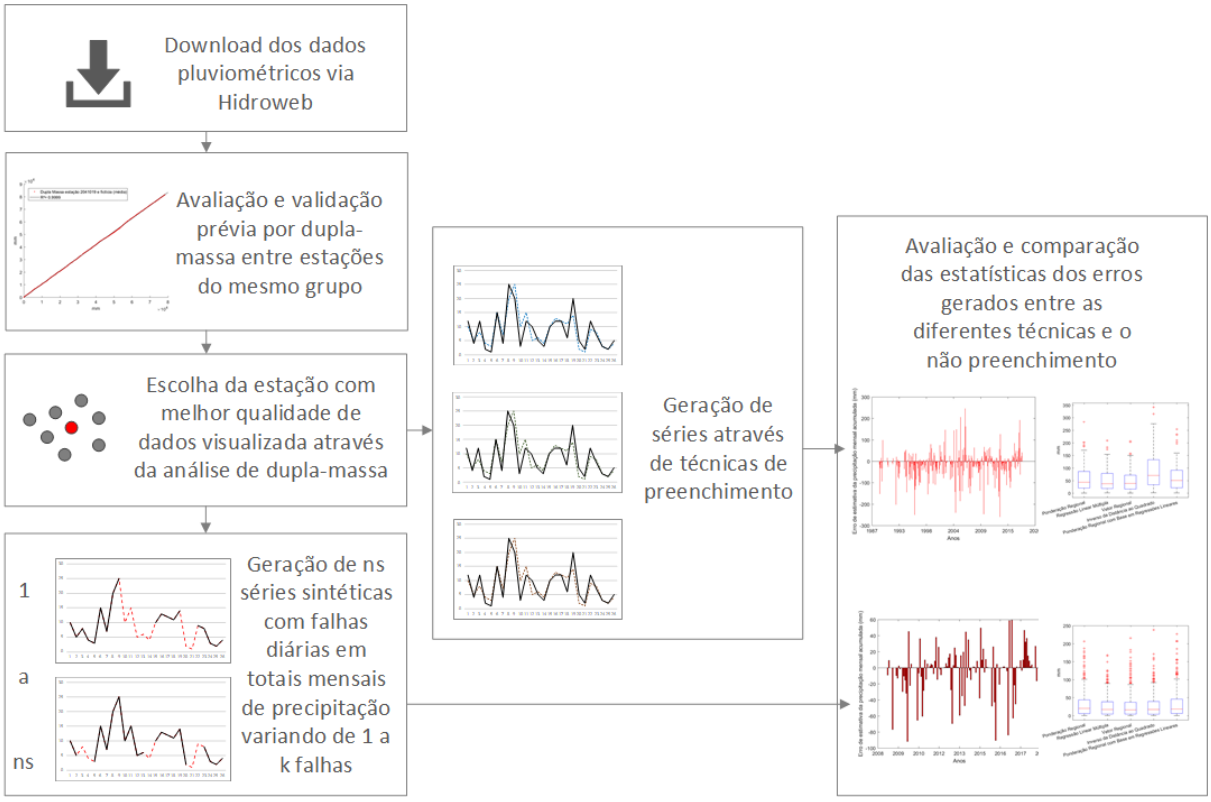
O presente estudo pretende avaliar os efeitos que uma falha com determinadas características gera em uma série de dados de precipitação. Para tanto, será utilizada metodologia semelhante à descrita anteriormente no estudo da *World Meteorological Organization* (2007), com as adaptações necessárias para a aplicação em dados de precipitação e de acordo com a disponibilidade dos dados. Os resultados indicarão a possibilidade do não preenchimento (ou preenchimento com zero precipitação) das séries com melhores resultados que a prática comum de preenchimento de séries apresenta.

4.1 MÉTODOS

Conforme apresentado no delineamento da pesquisa, foi desenvolvida uma ferramenta computacional que gera séries com falhas sintéticas diárias a partir das séries originais. Dessa forma, foi possível comparar a média das n_s séries com falhas sintéticas, e assim, determinar os efeitos que as falhas causam à série original. Foram avaliadas falhas que variam em número de dias no mês e suas posições. Analisando o efeito nas estatísticas mensais da série e a localização geográfica da estação, foi possível determinar o tamanho do “dano” e sua relação com a sazonalidade da precipitação e a sua localização.

O fluxograma da Figura 3 apresenta a sequência que compõe os processos da análise proposta para a comparação das técnicas de preenchimento às eventuais falhas diárias.

Figura 3. Fluxograma de procedimentos para a análise proposta



Obs.: data do download dos dados no Hidroweb: 22 de out. de 2018 (AGÊNCIA NACIONAL DE ÁGUAS, 2019).

4.1.1 Criação de Falhas Sintéticas

Para a criação de falhas sintéticas usou-se a técnica de amostragem pseudoaleatória, tradicionalmente utilizada na computação, que é gerada por algoritmos determinísticos. A técnica se assemelha à amostragem aleatória ou probabilística, uma vez que os números pseudoaleatórios são aprovados em testes de aleatoriedade. A utilização de amostragem pseudoaleatória possibilita a repetição de sequências, seguem distribuição uniforme e são geradas de forma rápida com baixo custo computacional.

Para cada mês da série, foram inseridas falhas diárias que variam de 1 a *k* dias (*k*), sendo que para cada *k* foram geradas n_s séries sintéticas com falhas. O número de séries sintéticas (n_s) para cada *k* de cada mês foi definido a partir do momento em que se observou a estacionariedade do erro médio de forma visual das n_s séries sintéticas em relação à original.

4.1.2 Critério para Escolha das Séries Históricas e Análise dos Efeitos das Falhas

Inicialmente, foi escolhida uma estação piloto que apresentou determinadas características, que demonstrassem confiabilidade em seus dados **brutos**. Tal confiabilidade foi demonstrada através da análise gráfica entre a estação analisada e a média da precipitação das estações vizinhas (curva de dupla-massa). Para isso, foram utilizados apenas os períodos em que não havia falhas nos dados da estação escolhida, portanto, a série de dados efetivamente analisada foi uma série descontínua mensal.

A ideia é que essa série de dados brutos fosse validada por tais estações de apoio com alto coeficiente de correlação, sem que fosse necessária qualquer correção/imputação de dados. Sendo assim, almejou-se validar tal série de dados brutos e considerá-los consistidos, assumindo que tal série seja representativa das características pluviométricas da região quando avaliadas as estatísticas extraídas da série.

A avaliação/validação da estação escolhida foi a que apresentou melhor coeficiente de determinação (r^2) quando construido o gráfico duplo-acumulativo entre a estação analisada e a média das estações vizinhas e ajustada a regressão linear. Uma vez calculados os coeficientes de determinação, a que apresentou melhor coeficiente de determinação, passou a ser a estação de análise.

A opção por utilizar uma estação com o melhor coeficiente de determinação da curva duplo-acumulativa (dados do posto contra a média dos postos vizinhos) foi feita a partir da premissa que esta seria a estação com a melhor série de dados disponível entre todas as estações do grupo. Portanto, na avaliação de desempenho das técnicas de preenchimento, os erros apresentados seriam oriundos das próprias técnicas de preenchimento e da utilização de estações de apoio com séries menos confiáveis, do que da própria série em análise.

De qualquer forma, conforme relatado no item 3.3.1 e nas limitações do estudo (2.6), a análise do desempenho do preenchimento das séries e do efeito das falhas geradas é influenciada pelos erros inerentes ao processo de medição do fenômeno.

Buscou-se assim, reduzir tais interferências filtrando as estações que apresentaram maior homogeneidade com as estações vizinhas.

Uma vez validada a série original, esta serviu de série base para comparação das séries modificadas através das falhas e se pode avaliar o efeito da extensão das falhas, conforme objetivo proposto. Caso a série avaliada não fosse validada dentre os demais critérios (extensão da série, baixo coeficiente de determinação no teste de dupla massa, etc.), seria escolhida uma nova estação para a avaliação.

Cada estação escolhida em cada grupo de cada região obedeceu principalmente aos critérios de distância entre as estações do mesmo grupo (não maior do que 120 km), sendo estas estações todas convencionais, ou seja, com dados com discretização diária. Os dados utilizados foram os dados brutos, evitando-se assim, a utilização de séries que já tivessem passado por tratamentos e homogeneização, o que afetaria os resultados.

A partir da análise de estações em diferentes regiões, almejou-se descobrir determinados padrões para a região no que tange à variação sazonal do efeito das falhas diárias nos totais mensais precipitados. Dessa forma, se obteve uma estimativa da magnitude do impacto que determinadas falhas geram no total mensal precipitado. A partir daí, se pode comparar às técnicas de preenchimento tradicionalmente utilizadas.

4.1.3 Comparação da Técnica do Não Preenchimento às Técnicas Tradicionalmente Utilizadas

Uma vez quantificados os desvios da série devido ao não preenchimento, se pode, da mesma forma comparar o preenchimento das técnicas tradicionalmente utilizadas e os desvios que os preenchimentos causaram em relação à série original.

Foram testadas as técnicas mencionadas anteriormente e revisitadas abaixo:

- Método do Vetor Regional.
- Regressão Linear Múltipla.
- Ponderação Regional.
- Interpolação do inverso da distância ao quadrado.
- Regressão Linear Múltipla com base na Ponderação Regional.

4.2 MATERIAIS

Para a aplicação da metodologia apresentada, foram selecionadas seis regiões no Brasil, com dois grupos de estações para cada região, de forma a testar as hipóteses apresentadas anteriormente em climas com diferentes características. Foram testados dois grupos para cada região de forma a ter certa redundância para possibilitar a verificação de possíveis padrões.

O Quadro 1 apresenta os dados das estações utilizadas enquanto a Figura 4 apresenta o mapa com a localização destas estações. Cada grupo conta com pelo menos seis estações, permitindo que se apliquem as diferentes metodologias de preenchimento à estação escolhida dentre elas.

Quadro 1. Dados das estações analisadas (continua)

Código	Nome	Bacia	Sub-B.	Estado	Município	Operadora	Latitude	Longitude	Altitude
Norte - G1									
8360007	Três Corações	1	14	RR	Amajari	CPRM	3,6483	-60,97	110*
8361000	Fazenda São João	1	14	RR	Amajari	CPRM	3,6608	-61,38	95*
8361001	Boqueirão	1	14	RR	Alto Alegre	CPRM	3,2906	-61,29	108*
8361004	Colônia Do Taiano	1	14	RR	Alto Alegre	CPRM	3,2872	-61,09	147*
8361005	Tepequem	1	14	RR	Amajari	CPRM	3,7592	-61,72	710*
8361007	Maracá	1	14	RR	Alto Alegre	CPRM	3,3528	-61,42	94*
Norte - G2									
259004	Rio Preto Da Eva	1	15	AM	Rio Preto da Eva	CPRM	-2,7003	-59,6997	29*
259005	Rio Urubú	1	16	AM	Rio Preto da Eva	CPRM	-2,66	-59,3608	22*
260008	Lago Do Ipixuna	1	14	AM	Iranduba	CPRM	-2,8964	-60,6883	49*
260009	Rio Cuieiras	1	14	AM	Manaus	CPRM	-2,6656	-60,3275	37*
358002	Balsa Do Rio Urubú	1	16	AM	Itacoatiara	CPRM	-2,9131	-59,0433	16*
359008	Comunidade Bom Sucesso	1	15	AM	Manaus	CPRM	-3,1306	-59,4392	9*
360001	Manacapuru	1	14	AM	Manacapuru	CPRM	-3,3083	-60,6094	11*
360005	Comunidade São João	1	14	AM	Manaus	CPRM	-3,0472	-60,2558	36*
Centro-Oeste - G1									
1655004	Santa Lúcia	6	66	MT	Barão de Melgaço	CPRM	-16,8939	-55,9072	298
1656001	Porto Cercado	6	66	MT	Poconé	CPRM	-16,5119	-56,3756	119
1656003	São José Do Boriréu	6	66	MT	Barão de Melgaço	CPRM	-16,9211	-56,2236	160
1656004	São João	6	66	MT	Poconé	CPRM	-16,9442	-56,6319	116
1755003	São Jerônimo	6	66	MT	Barão de Melgaço	CPRM	-17,2017	-56,0086	112
1756000	Ilha Camargo	6	66	MT	Barão de Melgaço	CPRM	-17,0572	-56,5856	113
1756001	São José Do Piquiri	6	66	MT	Barão de Melgaço	CPRM	-17,2914	-56,3847	180
Centro-Oeste - G2									
1954002	Rochedo	6	66	MS	Rochedo	CPRM	-19,9533	-54,8786	259
1955000	Iguaçu	6	66	MS	Aquidauana	CPRM	-19,9447	-55,7944	130
2054005	Jaraguá	6	63	MS	Terenos	CPRM	-20,4931	-54,8114	518
2054009	Santa Elisa	6	66	MS	Terenos	CPRM	-20,4906	-54,8669	411
2055001	Cipolândia	6	66	MS	Aquidauana	CPRM	-20,1278	-55,3933	171
2055002	Palmeiras	6	66	MS	Dois Irmãos do Buriti	CPRM	-20,4489	-55,4308	175
2055003	Fazenda Lajeado	6	66	MS	Dois Irmãos do Buriti	CPRM	-20,2922	-55,4453	204
2055004	Taboco	6	66	MS	Aquidauana	CPRM	-20,0703	-55,6453	146
Nordeste - G1									
340064	Sobral	3	35	CE	Sobral	CPRM	-3,6967	-40,34	65*
340105	Ararius	3	35	CE	Cariré	CPRM	-3,89	-40,6047	115*
340106	Fazenda Poço Verde	3	35	CE	Sobral	CPRM	-3,7506	-40,5064	114*
340118	Curral Velho	3	35	CE	Groairas	CPRM	-3,9975	-40,3703	102*
440042	Fazenda Cajazeiras	3	35	CE	Hidrolândia	CPRM	-4,3678	-40,5356	157*
440043	Fazenda Paraná	3	35	CE	Santa Quitéria	CPRM	-4,0969	-40,1264	148*
440070	Trapiá	3	35	CE	Santa Quitéria	CPRM	-4,1894	-40,3186	142*
440084	Várzea Do Grosso	3	35	CE	Varjota	CPRM	-4,1378	-40,4397	111*

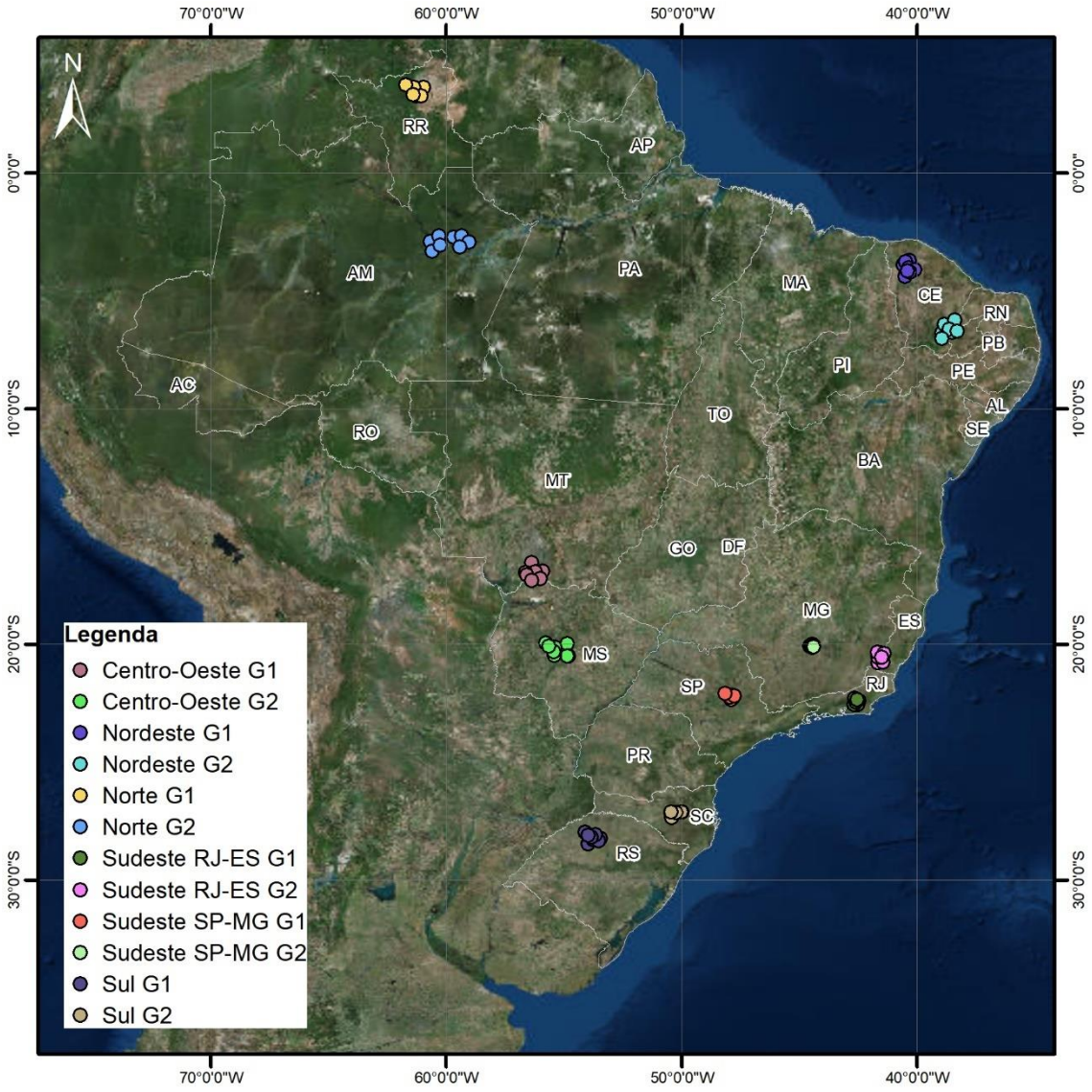
Quadro 1. Dados das estações analisadas (continuação/conclusão)

Código	Nome	Bacia	Sub-B.	Estado	Município	Operadora	Latitude	Longitude	Altitude
Nordeste - G2									
638007	Lavras Da Mangabeira	3	36	CE	Lavras da Mangabeira	CPRM	-6,7628	-38,9606	247
638014	Icó	3	36	CE	Icó	CPRM	-6,4089	-38,8636	160
638044	Açude Bonito	3	37	RN	São Miguel	CPRM	-6,2128	-38,4256	500
638088	Fazenda Santo Antônio	3	37	PB	Santa Helena	CPRM	-6,7403	-38,5297	259*
638089	Fazenda Boa Vista	3	36	CE	Umari	CPRM	-6,6261	-38,6617	295*
638090	Serrote Dos Letreiros	3	37	PB	Sousa	CPRM	-6,6994	-38,3111	240*
738063	Sítio Santa Cruz	3	36	CE	Aurora	CPRM	-7,0142	-38,9581	283*
738067	Sítio Lapinha	3	36	CE	Aurora	CPRM	-7,0139	-38,9581	283*
Sudeste RJ-ES - G1									
2242005	Fazenda São João	5	59	RJ	Nova Friburgo	CPRM	-22,3903	-42,4947	960
2242008	Gaviões	5	59	RJ	Silva Jardim	CPRM	-22,5489	-42,5461	1620
2242009	São Lourenço	5	58	RJ	Nova Friburgo	CPRM	-22,3494	-42,6242	1130
2242014	Japuiba	5	59	RJ	Cachoeiras de Macacu	CPRM	-22,5564	-42,69	27
2242016	Fazenda São Joaquim	5	59	RJ	Cachoeiras de Macacu	CPRM	-22,4411	-42,6219	275
2242022	Fazenda Mendes	5	58	RJ	Nova Friburgo	CPRM	-22,2858	-42,66	1010
2242024	Teodoro De Oliveira	5	58	RJ	Nova Friburgo	CPRM	-22,3772	-42,5517	1105
2242025	Cascatinha Do Conego	5	58	RJ	Nova Friburgo	CPRM	-22,3425	-42,5572	1210
Sudeste RJ-ES – G2									
2041001	Guaçuí	5	57	ES	Guaçuí	CPRM	-20,7736	-41,6817	576
2041003	Rive	5	57	ES	Alegre	CPRM	-20,7469	-41,4661	127
2041013	Iúna	5	57	ES	Iúna	CPRM	-20,3458	-41,5375	615
2041016	Ibitirama	5	57	ES	Alegre	CPRM	-20,5406	-41,6656	794
2041017	Santa Cruz - Caparaó	5	57	ES	Iúna	CPRM	-20,3228	-41,7042	920
2041018	Usina Fortaleza	5	57	ES	Muniz Freire	CPRM	-20,3714	-41,4089	580
2041019	Itaici	5	57	ES	Muniz Freire	CPRM	-20,5283	-41,5114	380
Sudeste SP-MG - G1									
2247180	Itirapina	6	62	SP	Itirapina	Construf.	-22,2492	-47,8267	739
2247182	Ribeirão Do Feijão	6	62	SP	São Carlos	Construf.	-22,1522	-47,8864	676
2247184	Itaqueri Da Serra	6	62	SP	Itirapina	Construf.	-22,3422	-47,9178	894
2247185	Fazenda Campo Alegre	6	62	SP	Brotas	Construf.	-22,1483	-48,0006	733
2247196	Crhea-Broa	6	62	SP	Itirapina	Construf.	-22,17	-47,8989	732
2247197	Est, Ferroviária Campo Alegre	6	62	SP	Brotas	Construf.	-22,2358	-47,9581	747
2247198	Fazenda Pequena Holanda	6	62	SP	Itirapina	Construf.	-22,1797	-47,7842	780
2248110	Ribeirão Bonito	6	62	SP	Ribeirão Bonito	Construf.	-22,0778	-48,1803	486
Sudeste SP-MG – G2									
2044002	Itaúna - Montante	4	40	MG	Itaúna	CPRM	-20,0714	-44,5703	859
2044019	Fazenda Vista Alegre	4	40	MG	Mateus Leme	CPRM	-20,0514	-44,4517	913,4
2044020	Calambau	4	40	MG	Itaúna	CPRM	-20,0683	-44,4922	943
2044026	Fazenda Coqueiros	4	40	MG	Itaúna	CPRM	-20,1297	-44,4744	974,7
2044041	Fazenda Laranjeiras - Jusante	4	40	MG	Itaúna	CPRM	-20,1022	-44,4847	894,7
2044043	Estiva	4	40	MG	Mateus Leme	CPRM	-20,0003	-44,4617	806,8
2044052	Jardim	4	40	MG	Mateus Leme	CPRM	-20,045	-44,4078	806
2044054	Serra Azul	4	40	MG	Mateus Leme	CPRM	-20,0867	-44,4272	817,4
Sul – G1									
2754010	Esquina Araujo	7	74	RS	Independência	CPRM	-27,9681	-54,1164	400
2853003	Conceição	7	75	RS	Ijuí	CPRM	-28,4561	-53,9717	160
2853010	Passo Faxinal	7	75	RS	Ajuricaba	CPRM	-28,2894	-53,7794	200
2853023	Condor	7	75	RS	Condor	CPRM	-28,2256	-53,4703	440
2853034	Panambi	7	75	RS	Panambi	CPRM	-28,3444	-53,5417	480
2853035	Nova Ramada	7	75	RS	Nova Ramada	CPRM	-28,0667	-53,6833	533
2853040	Mauá	7	75	RS	Ijuí	CPRM	-28,1792	-53,8578	256
2854001	Boa Vista	7	75	RS	Catuipe	CPRM	-28,1111	-53,9931	447
Sul – G2									
2749003	Taió	8	83	SC	Taió	EPAGRI	-27,1131	-49,9944	360
2750010	Ponte Alta Do Norte	7	71	SC	Ponte Alta Do Norte	CPRM	-27,1611	-50,4689	980
2750012	Ponte Do Rio Antinhas	7	71	SC	São Cristóvão do Sul	CPRM	-27,3453	-50,4358	940
2750014	Barragem Oeste	8	83	SC	Taió	EPAGRI	-27,0972	-50,0339	370
2750021	Cabeceira Ribeirão Caetano	8	83	SC	Mirim Doce	EPAGRI	-27,1411	-50,2644	900
2750022	Ponte Alta Do Norte - Cifsul	7	71	SC	Ponte Alta Do Norte	ANA	-27,1208	-50,4567	980

Obs.: Dados oriundos do inventário da Agência Nacional de Águas, com codificação padronizada (AGÊNCIA NACIONAL DE ÁGUAS, 2019).

* Altitude estimada a partir dos dados da adaptação do Modelo Digital de Elevação da Missão Topográfica Radar Shuttle (SRTM) para o sistema oficial brasileiro (WEBER; HASENACK; FERREIRA, 2004).

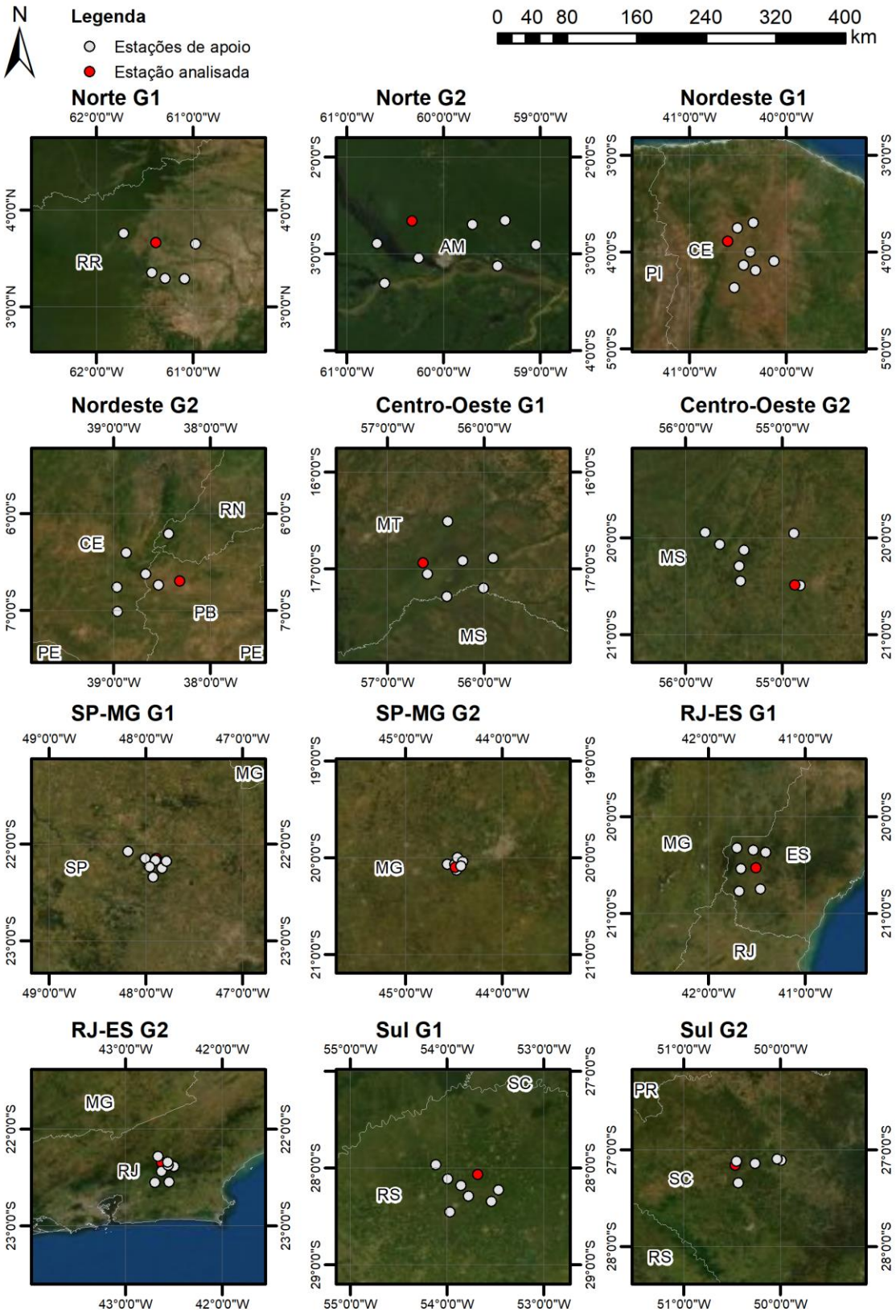
Figura 4. Localização das estações estudadas



A Figura 5 apresenta em maior escala a localização das estações dos diferentes grupos analisados neste estudo. A escala de todas as regiões é a mesma, portanto, pode-se observar que a distância entre estações de cada grupo varia. Regiões como a Norte G2 tem um menor grupo de estações e por isso, é mais difícil encontrar grupos de estações vizinhas, ao contrário de grupos como de São Paulo e Minas Gerais, o que pode influenciar nos resultados obtidos.

A Figura 6 apresenta os diagramas temporais de disponibilidade das séries das estações de cada grupo. Os diagramas temporais apresentam toda a série de dados disponível, variando a escala de cores, dependendo do número de dias disponível em cada mês, dependendo da extensão de eventuais falhas diárias.

Figura 5. Ampliação da localização das estações estudadas.

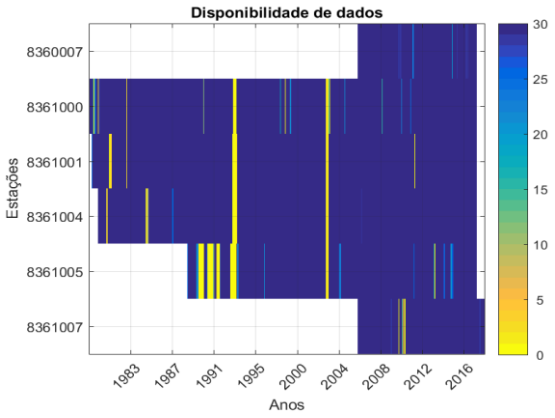


As análises realizadas filtraram apenas os meses das estações que não apresentaram falhas diárias, excluindo-se assim erros (falhas) já constatados nas

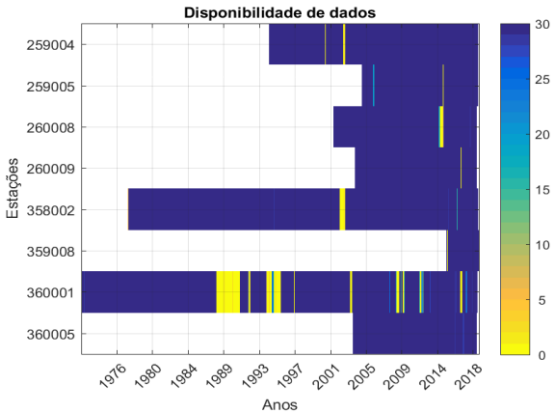
séries. Além disso, para que as técnicas de preenchimento se fizessem consistentes, as séries de dados analisadas também foram filtradas para que houvesse simultaneidade entre pelo menos quatro estações para cada mês de análise.

Figura 6. Diagramas de disponibilidade temporal das estações analisadas em cada região (continua).

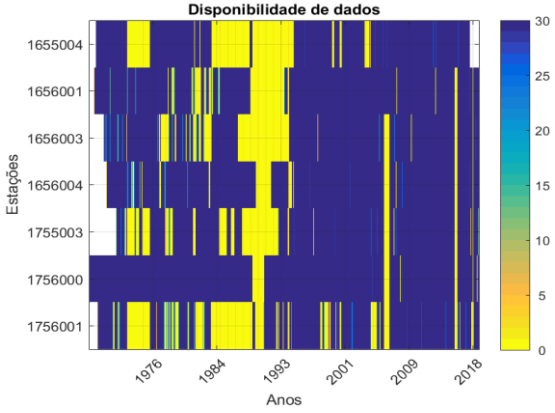
a) Norte G1 (dias/mês)



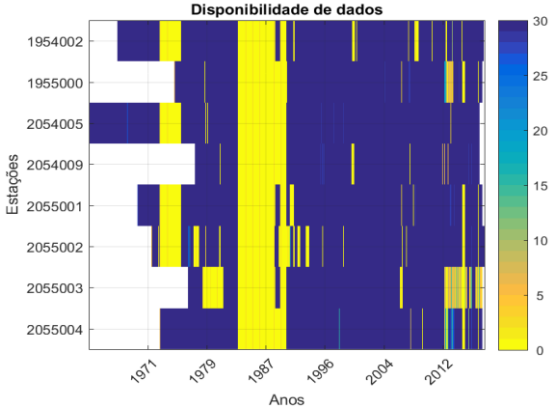
b) Norte G2 (dias/mês)



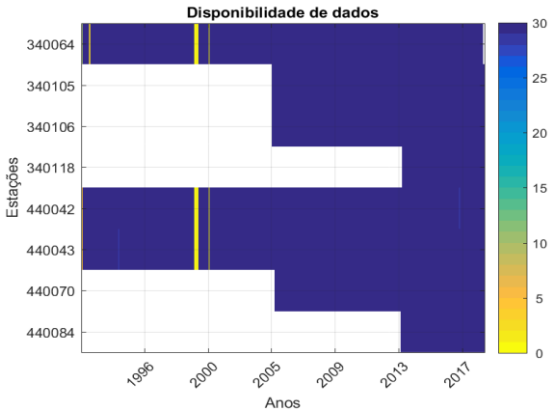
c) Centro-Oeste G1 (dias/mês)



d) Centro-Oeste G2 (dias/mês)



e) Nordeste G1 (dias/mês)



f) Nordeste G2 (dias/mês)

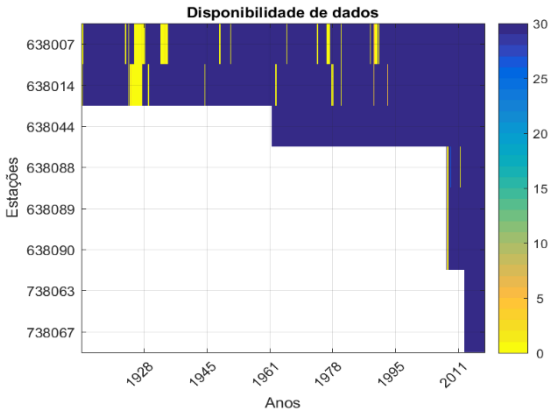
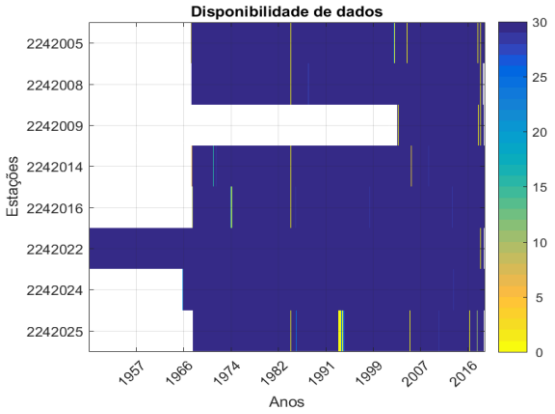
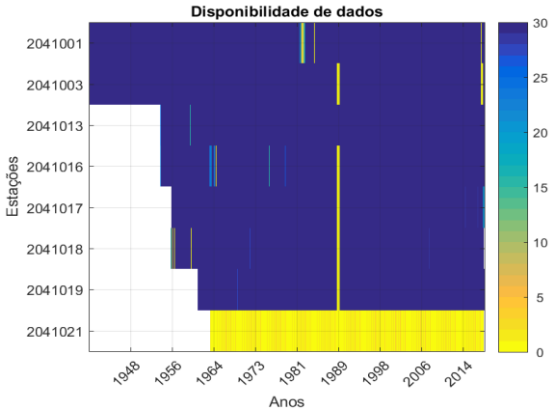


Figura 6. Diagramas de disponibilidade temporal das estações analisadas em cada região (continuação/conclusão).

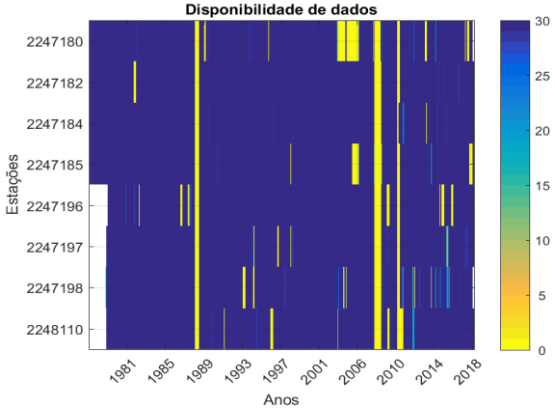
g) Sudeste RJ-ES G1 (dias/mês)



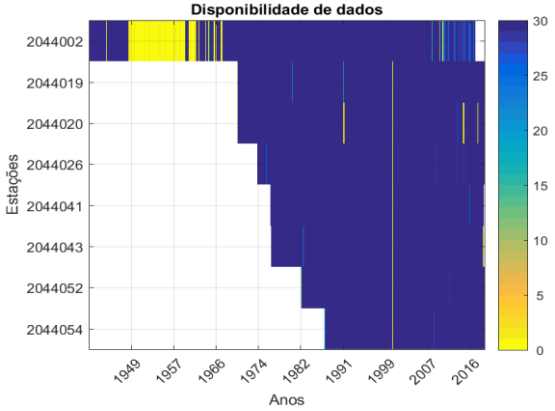
h) Sudeste RJ-ES G2 (dias/mês)



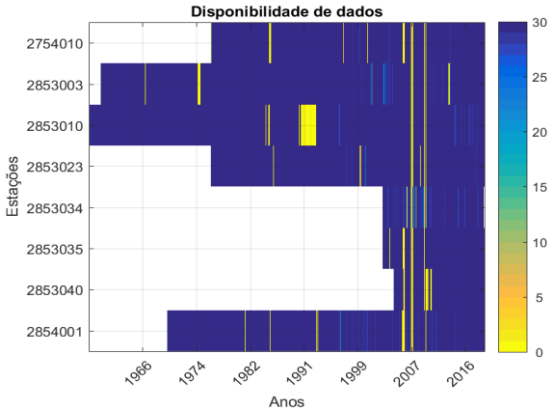
a) Sudeste SP-MG G1 (dias/mês)



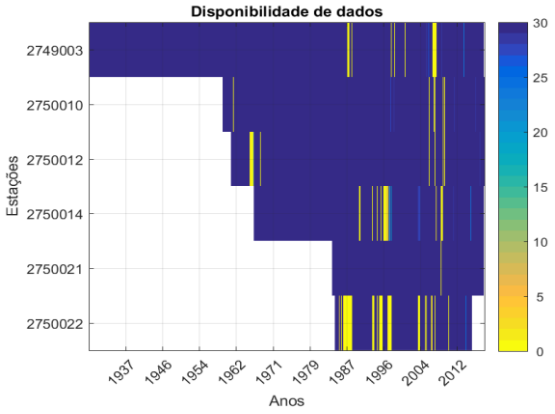
b) Sudeste SP-MG G2 (dias/mês)



c) Sul G1 (dias/mês)



d) Sul G2 (dias/mês)



5 APLICAÇÃO DO MÉTODO E RESULTADOS

5.1 AVALIAÇÃO POR MÉTODO DE DUPLA-MASSA E ESCOLHA DAS ESTAÇÕES PRINCIPAIS

A Tabela 1 e os gráficos da Figura 7 apresentam os resultados da análise das curvas duplo-acumulativas de cada região. O menor período analisado (estações escolhidas) contava com 15 anos de dados interruptos (não contínuos, mas sem falhas diárias) e o maior de 57 anos

Tabela 1. Extensão das séries e coeficiente de determinação do ajuste linear da curva dupla massa de cada estação (em destaque os melhores ajustes por região) (continua).

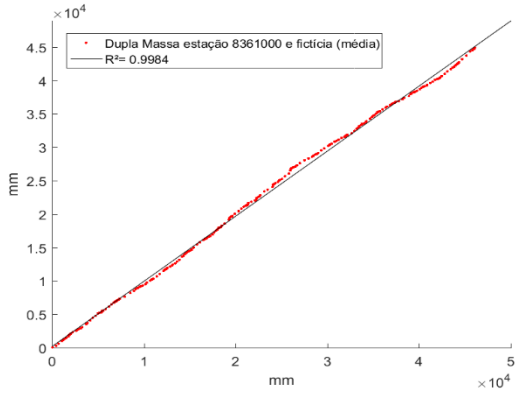
Grupo	Estação	Extensão de análise (anos)	r^2 ajuste linear curva dupla -massa
Norte G1	8360007	11	0,9884
Norte G1	8361000	25	**0,9984**
Norte G1	8361001	25	0,9976
Norte G1	8361004	25	0,9982
Norte G1	8361005	24	0,9839
Norte G1	8361007	12	0,9921
Norte G2	259004	16	0,9978
Norte G2	259005	13	0,9964
Norte G2	260008	15	0,9877
Norte G2	260009	14	**0,9988**
Norte G2	358002	16	0,9971
Norte G2	359008	4	0,9922
Norte G2	360001	14	0,9981
Norte G2	360005	14	0,975
Centro-Oeste G1	1655004	31	0,9987
Centro-Oeste G1	1656001	38	0,9989
Centro-Oeste G1	1656003	35	0,9968
Centro-Oeste G1	1656004	39	**0,9995**
Centro-Oeste G1	1755003	34	0,9991
Centro-Oeste G1	1756000	42	0,9995
Centro-Oeste G1	1756001	28	0,9986
Centro-Oeste G2	1954002	35	0,9417
Centro-Oeste G2	1955000	33	0,9984
Centro-Oeste G2	2054005	35	0,9972
Centro-Oeste G2	2054009	32	**0,9995**
Centro-Oeste G2	2055001	36	0,9976
Centro-Oeste G2	2055002	33	0,996
Centro-Oeste G2	2055003	27	0,9996
Centro-Oeste G2	2055004	34	0,9955
Nordeste G1	340064	14	0,9954
Nordeste G1	340105	14	**0,9987**
Nordeste G1	340106	14	0,9985
Nordeste G1	340118	5	0,9975
Nordeste G1	440042	14	0,995
Nordeste G1	440043	14	0,9981
Nordeste G1	440070	14	0,9983
Nordeste G1	440084	6	0,9967
Nordeste G2	638007	10	0,9984
Nordeste G2	638014	10	0,9982
Nordeste G2	638044	10	0,9975
Nordeste G2	638088	9	0,997
Nordeste G2	638089	10	0,9968
Nordeste G2	638090	10	**0,9985**
Nordeste G2	738063	5	0,9938
Nordeste G2	738067	5	0,997
Sudeste RJ-ES G1	2242005	51	0,9991
Sudeste RJ-ES G1	2242008	51	0,9952
Sudeste RJ-ES G1	2242009	15	**0,9996**
Sudeste RJ-ES G1	2242014	51	0,9991

Tabela 1. Extensão das séries estudadas e coeficiente de determinação do ajuste linear da curva dupla massa de cada estação (em destaque os melhores ajustes por região) (continuação/conclusão).

Grupo	Estação	Extensão de análise (anos)	r2 ajuste linear curva dupla -massa
Sudeste RJ-ES G1	2242016	50	0,999
Sudeste RJ-ES G1	2242022	51	0,9981
Sudeste RJ-ES G1	2242024	51	0,9984
Sudeste RJ-ES G1	2242025	49	0,9973
Sudeste RJ-ES G2	2041001	63	0,9998
Sudeste RJ-ES G2	2041003	64	0,9999
Sudeste RJ-ES G2	2041013	64	0,9986
Sudeste RJ-ES G2	2041016	63	0,9997
Sudeste RJ-ES G2	2041017	62	0,9998
Sudeste RJ-ES G2	2041018	61	0,9994
Sudeste RJ-ES G2	2041019	57	**0,9999**
Sudeste SP-MG G1	2247180	36	0,9987
Sudeste SP-MG G1	2247182	40	**0,9997**
Sudeste SP-MG G1	2247184	40	0,9996
Sudeste SP-MG G1	2247185	39	0,9991
Sudeste SP-MG G1	2247196	36	0,998
Sudeste SP-MG G1	2247197	37	0,9992
Sudeste SP-MG G1	2247198	36	0,994
Sudeste SP-MG G1	2248110	37	0,9996
Sudeste SP-MG G2	2044002	39	0,9997
Sudeste SP-MG G2	2044019	44	0,9997
Sudeste SP-MG G2	2044020	43	0,9994
Sudeste SP-MG G2	2044026	44	0,9992
Sudeste SP-MG G2	2044041	42	**0,9999**
Sudeste SP-MG G2	2044043	42	0,9997
Sudeste SP-MG G2	2044052	36	0,9997
Sudeste SP-MG G2	2044054	31	0,9997
Sul G1	2754010	40	0,9998
Sul G1	2853003	39	0,9998
Sul G1	2853010	37	0,9999
Sul G1	2853023	40	0,9993
Sul G1	2853034	12	0,9998
Sul G1	2853035	15	**0,9999**
Sul G1	2853040	13	0,9999
Sul G1	2854001	38	0,9998
Sul G2	2749003	49	0,9995
Sul G2	2750010	50	**0,9999**
Sul G2	2750012	51	0,9998
Sul G2	2750014	48	0,9987
Sul G2	2750021	34	0,9996
Sul G2	2750022	25	0,9996

Figura 7. Curvas dupla massa das estações com melhor ajuste de cada região (milímetros acumulados) (continua)

a) Norte G1 (8361000)



b) Norte G2 (260009)

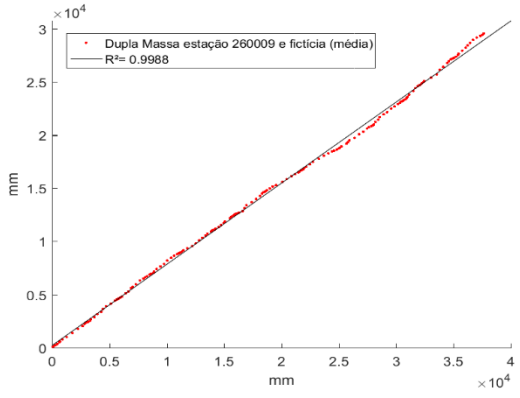
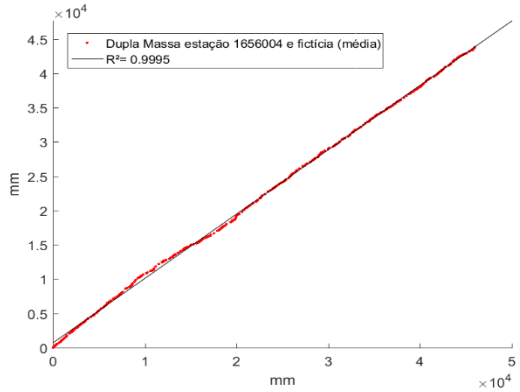
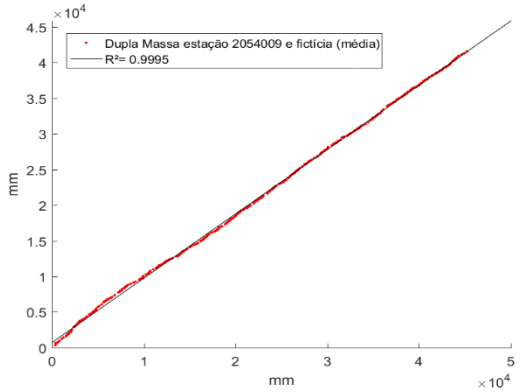


Figura 7. Curvas dupla massa das estações com melhor ajuste de cada região (milímetros acumulados) (continuação)

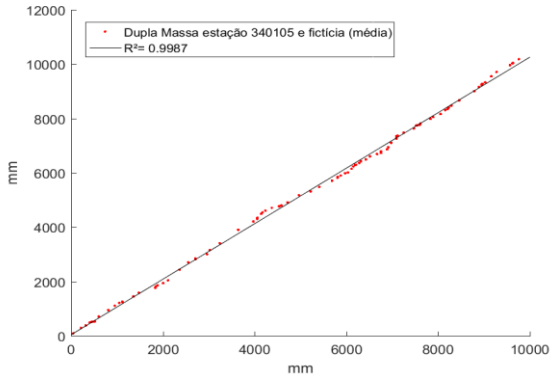
a) Centro-Oeste G1 (1656004)



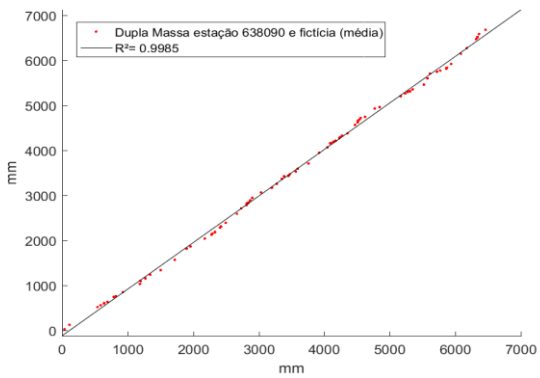
b) Centro-Oeste G2 (2054009)



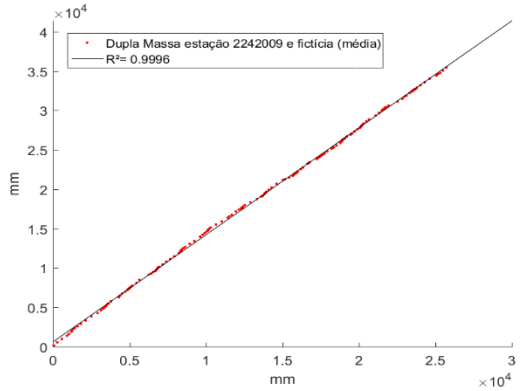
c) Nordeste G1 (340105)



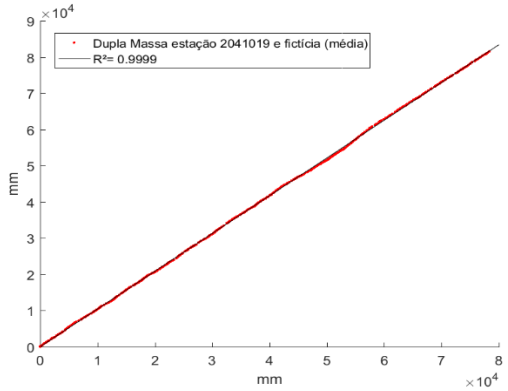
d) Nordeste G2 (638090)



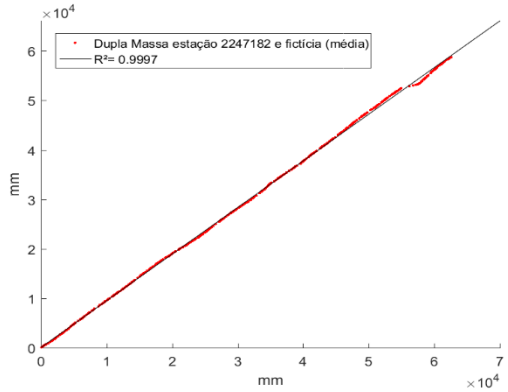
e) Sudeste RJ-ES G1 (2242009)



f) Sudeste RJ-ES G2 (2041019)



g) Sudeste SP-MG G1 (2247182)



h) Sudeste SP-MG G2 (2044041)

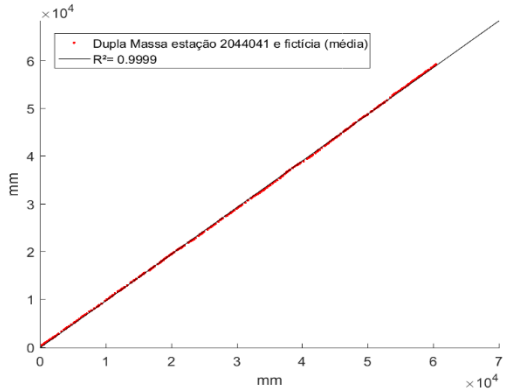
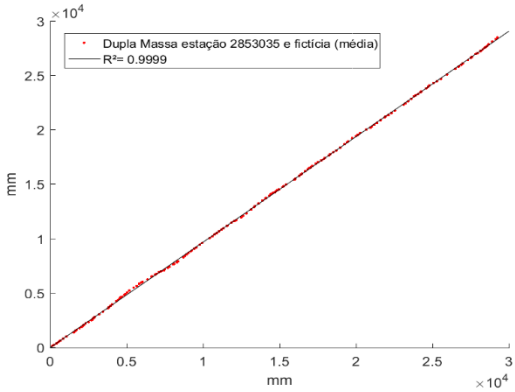
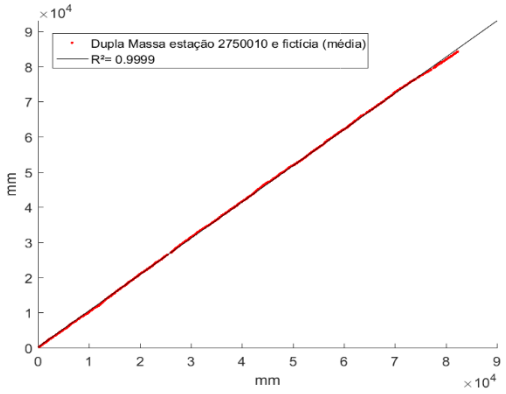


Figura 7. Curvas dupla massa das estações com melhor ajuste de cada região (milímetros acumulados) (conclusão)

a) Sul G1 (2853035)



b) Sul G2 (2750010)



5.2 GERAÇÃO DE SÉRIES POR PREENCHIMENTO

Com o auxílio das séries históricas das estações de apoio, foram geradas séries sintéticas através das técnicas de preenchimento. Os meses escolhidos para compor a série original e as séries sintéticas foram os meses que pelo menos quatro estações de apoio apresentavam dados. Dessa forma, as séries não são ininterruptas, e não necessariamente apresentam a mesma quantidade de cada mês do ano, variando, portanto, o tamanho das amostras para efeito de análise das diferenças sazonais.

Para cada mês, a série original foi comparada com as séries sintéticas geradas a partir de cada técnica de preenchimento e foram avaliados os histogramas dos erros absolutos, assim como demais estatísticas descritivas das séries de erros geradas. Da Figura 8 à 19 são apresentados os resultados da média das técnicas de preenchimento em termos de erro absoluto ao longo do tempo nas estações de cada região.

Figura 8. Média histórica dos erros das técnicas de preenchimento Norte G1 (8361000).

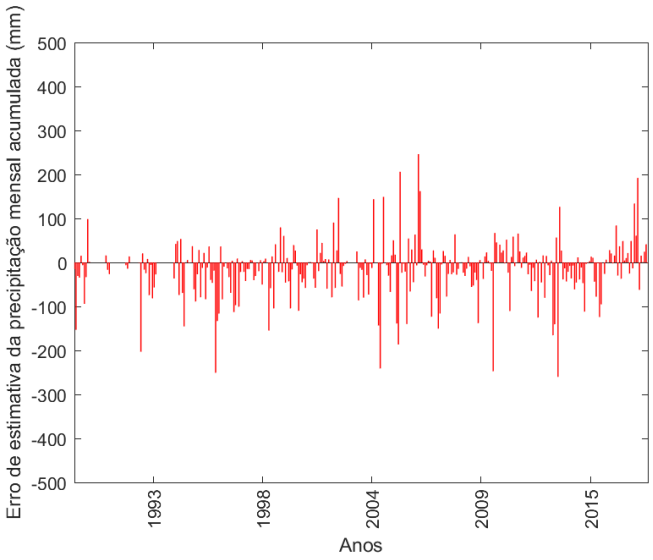


Figura 9. Média histórica dos erros das técnicas de preenchimento Norte G2 (260009).

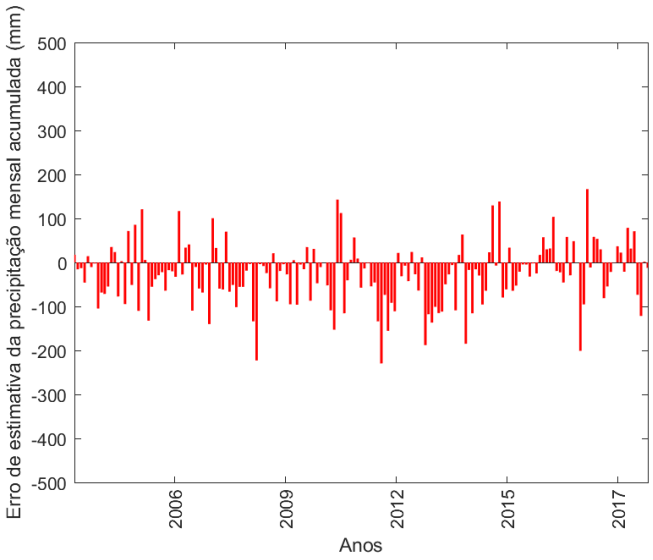


Figura 10. Média histórica dos erros das téc. de preenchimento Centro-Oeste G1 (1656004).

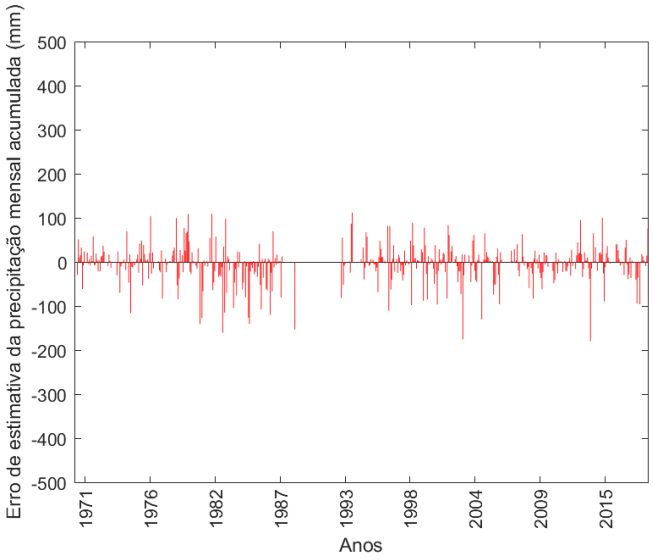


Figura 11. Média histórica dos erros das téc. de preenchimento Centro-Oeste G2 (2054009).

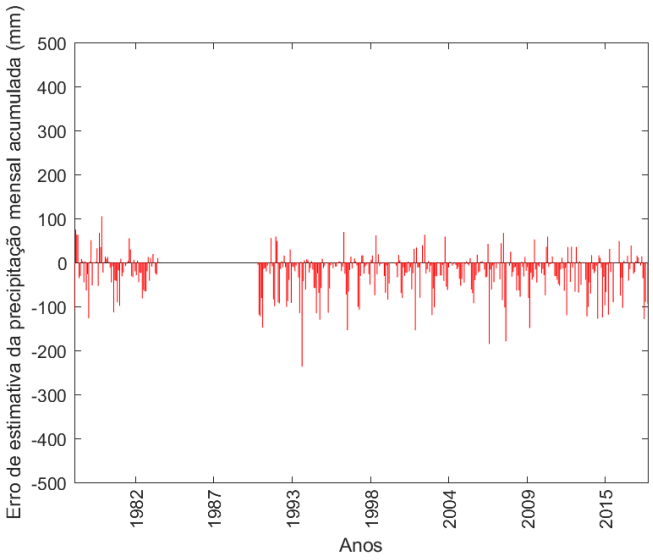


Figura 12. Média histórica dos erros das téc. de preenchimento Nordeste G1 (340105).

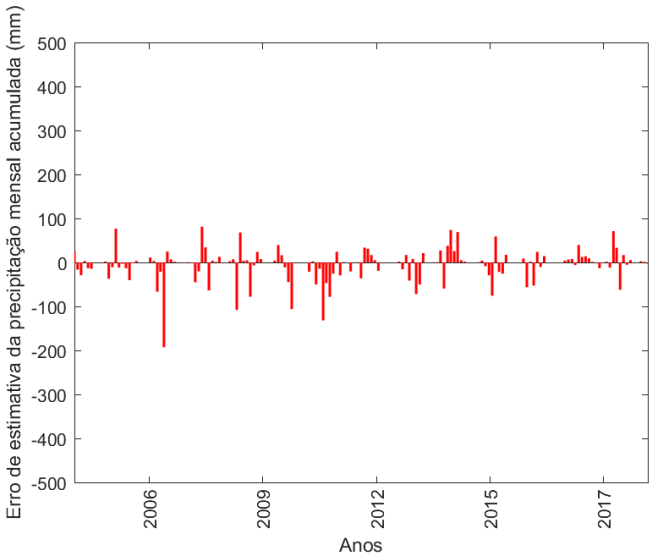


Figura 13. Média histórica dos erros das téc. de preenchimento Nordeste G2 (638090).

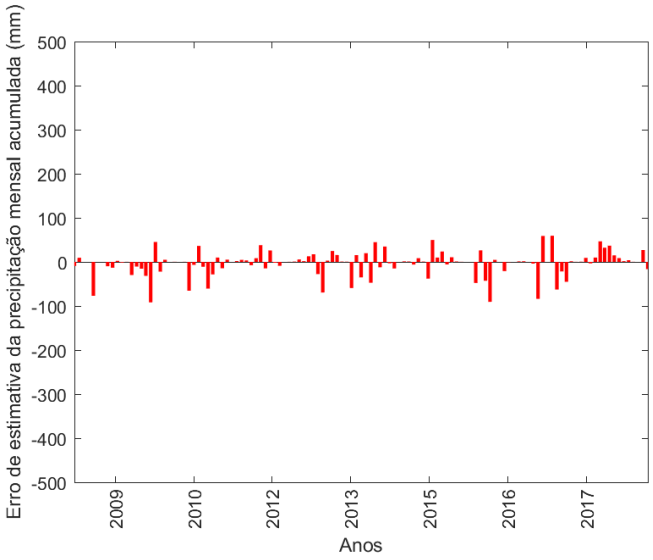


Figura 14. Média histórica dos erros das téc. de preench. Sudeste RJ-ES G1 (2242009).

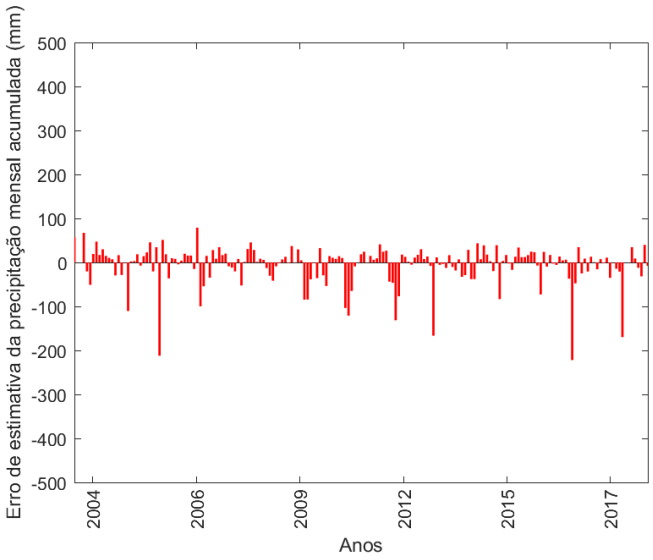


Figura 15. Média histórica dos erros das téc. de preench. Sudeste RJ-ES G2 (2041019).

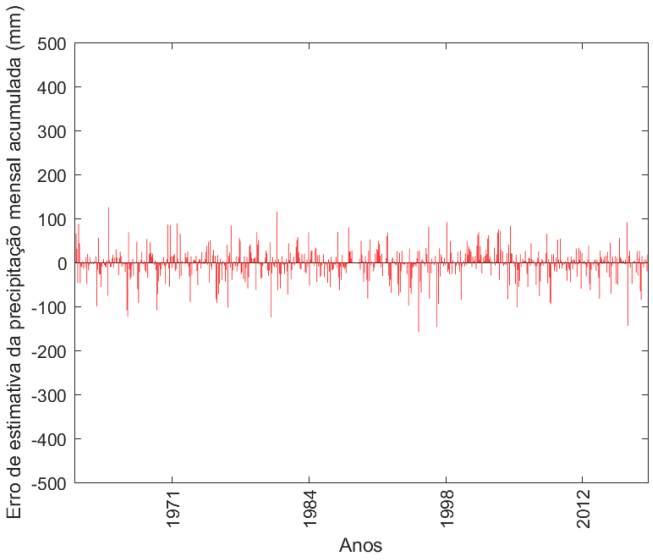


Figura 16. Média histórica dos erros das téc. de preench. Sudeste SP-MG G1 (2247182).

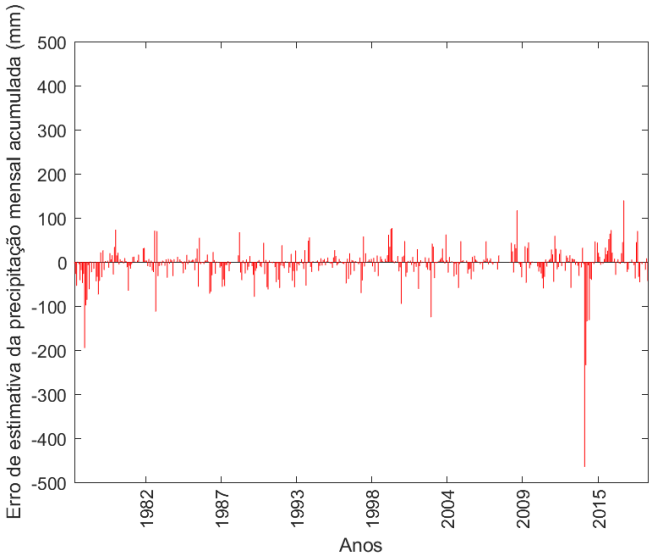


Figura 17. Média histórica dos erros das téc. de preench. Sudeste SP-MG G2 (2044041).

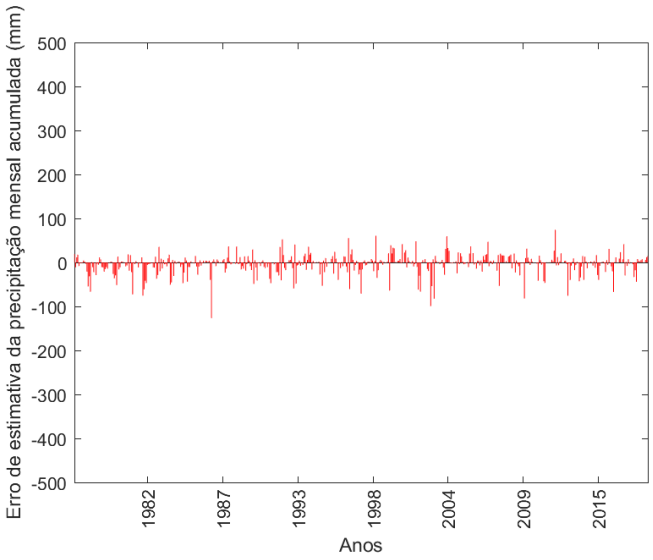


Figura 18. Média histórica dos erros das técnicas de preenchimento Sul G1 (2853035).

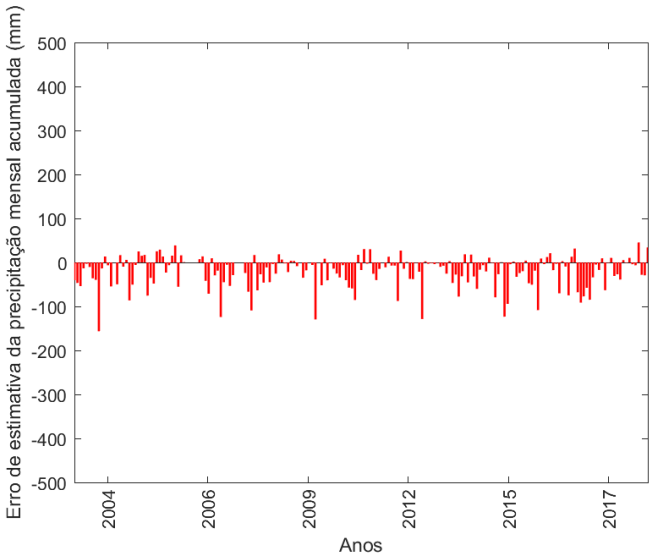
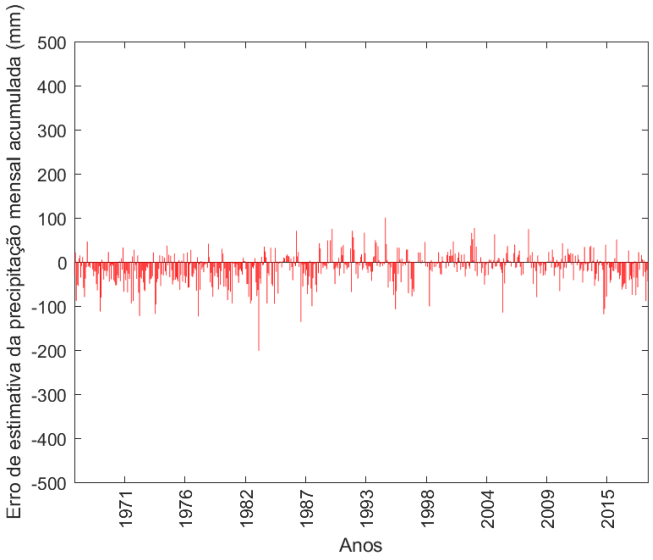


Figura 19. Média histórica dos erros das técnicas de preenchimento Sul G2 (2750010).



As Figuras (8 à 19) acima demonstram uma grande variabilidade (amplitude) da média dos erros das técnicas de preenchimento ao longo do tempo. Em muitos casos observa-se uma subestimativa da precipitação na maior parte do tempo, com diversos valores próximos de zero, mas também uma quantidade significativa de erros se aproximando de 100 mm a 200 mm. Em algumas estações, observa-se uma menor amplitude e menor assimetria (sempre a direita) como é o caso das duas estações das regiões Sudeste SP-MG e outras maior, como é o caso da duas estações das regiões Centro-Oeste.

Os gráficos (Figura 20 à 31) a seguir apresentam o módulo dos erros, portanto a subestimativa em muitas das regiões só é possível de ser visualizada nas figuras anteriormente apresentadas (Figura 8 à 19). Para fins de análise do desempenho das técnicas de preenchimento frente ao não preenchimento, é um detalhe menos relevante, no entanto vale destacar tais características. Uma hipótese de tal fato, é que as estações escolhidas, por terem sido validadas como tendo um melhor histórico, tendem a ter menos problemas no registro da precipitação em comparação com as estações vizinhas, apesar de terem sido utilizados apenas meses em que não houve falhas em nenhuma das estações nas análises.

Da Figura 20 à 31 são apresentados os gráficos *box-plot* do erro absoluto das diferentes técnicas de preenchimento das estações escolhidas de cada região. São os mesmos resultados apresentados nos gráficos anteriores, mas descritos não de forma temporal, mas dividindo os resultados calculados em diferentes “classes”.

Interpretando as Figuras abaixo (20 a 31), a linha horizontal vermelha representa a mediana de todos os erros absolutos calculados entre o preenchimento utilizando a técnica de preenchimento e a série original. Já o retângulo azul representa a amplitude interquartil que é limitada entre o 25º e o 75º percentil. As linhas horizontais pretas separam os extremos do restante da amostra, sendo estes graficados com o sinal de cruz em vermelho. Os extremos são calculados em função do intervalo interquartil e do desvio-padrão da amostra.

Destaca-se que para determinadas regiões aparentemente há um maior valor de extremos, no entanto isso se deve ao fato dos extremos serem graficados individualmente e o tamanho das séries de dados analisadas serem distintos.

Figura 20. *Box-plot* dos erros das téc. preenchimento mensal Norte G1 (8361000).

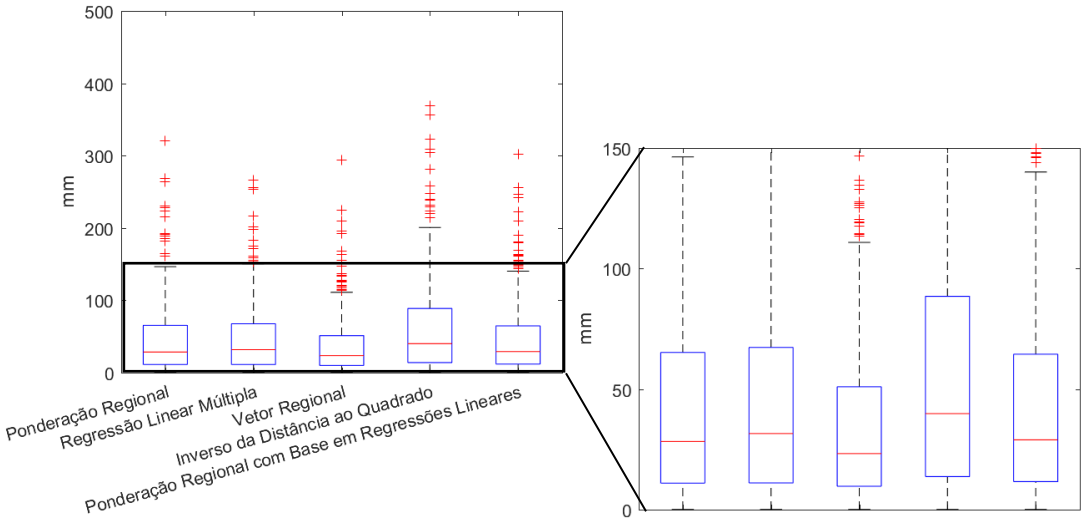


Figura 21. *Box-plot* dos erros das téc. preenchimento mensal Norte G2 (260009).

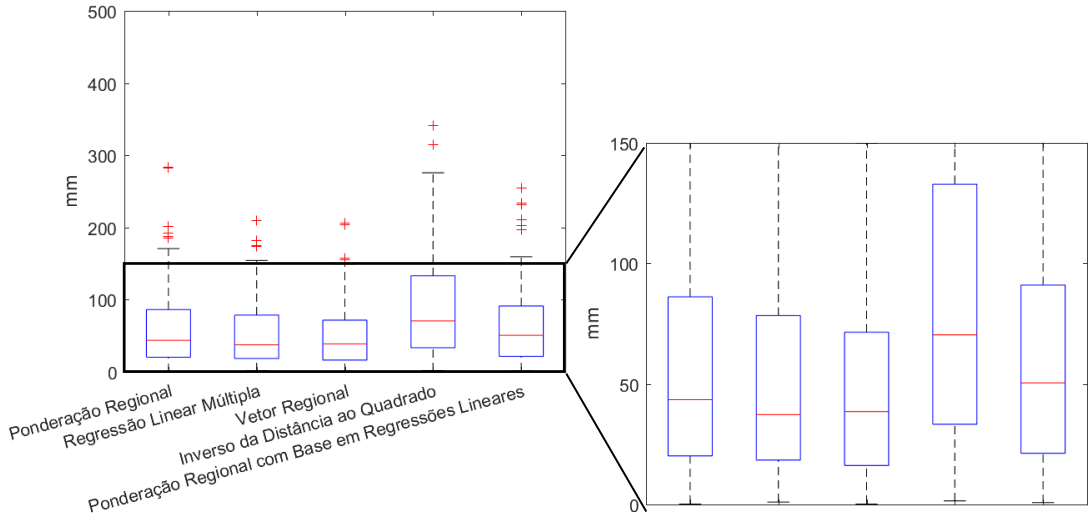


Figura 22. *Box-plot* dos erros das téc. preenchimento mensal Centro-Oeste G1 (1656004).

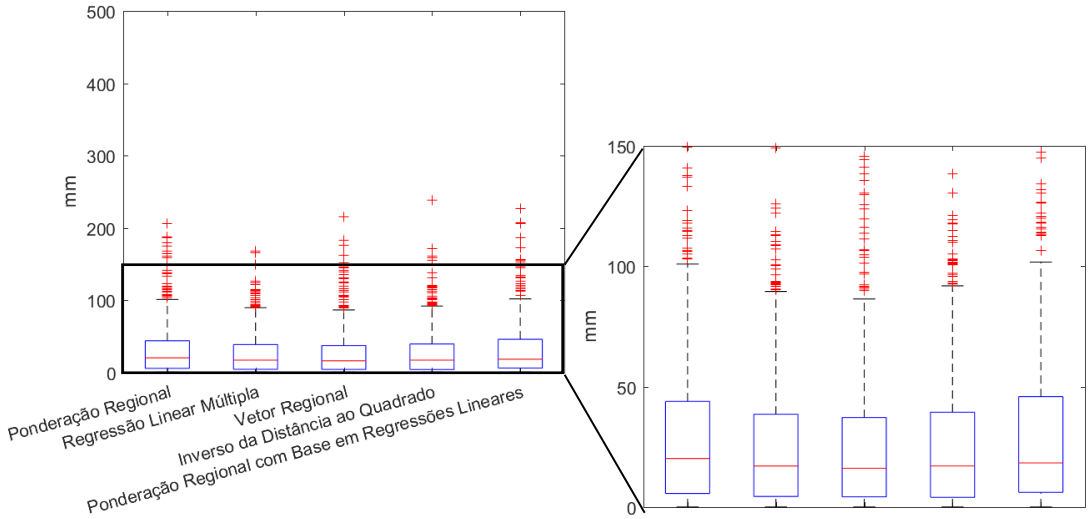


Figura 23. *Box-plot* dos erros das téc. preenchimento mensal Centro-Oeste G2 (2054009).

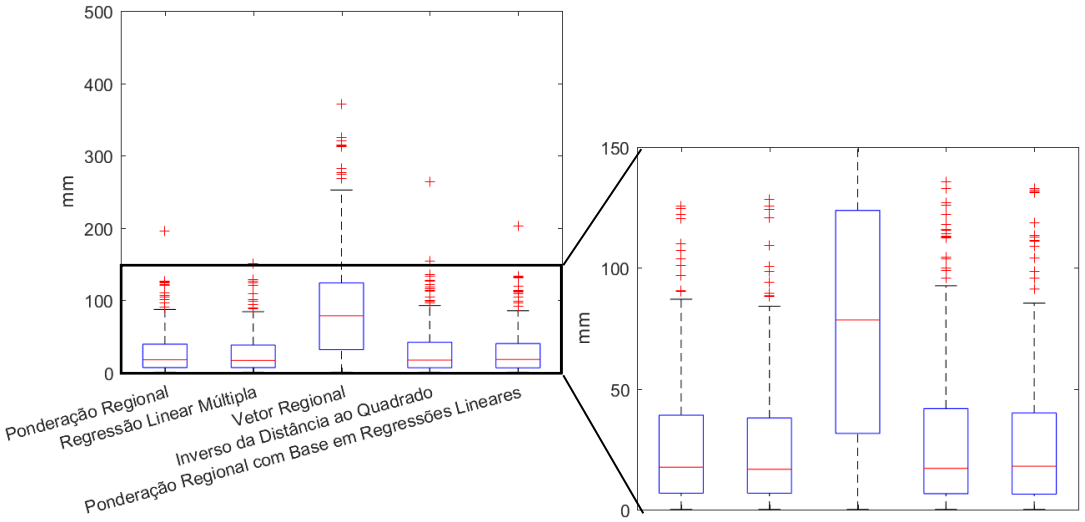


Figura 24. *Box-plot* dos erros das téc. preenchimento mensal Nordeste G1 (340105).

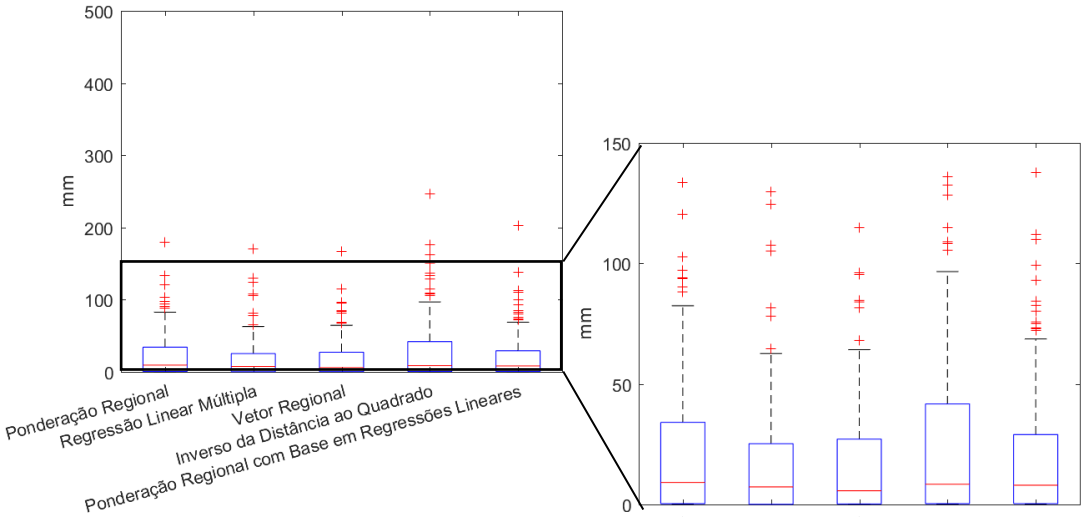


Figura 25. *Box-plot* dos erros das téc. preenchimento mensal Nordeste G2 (638090).

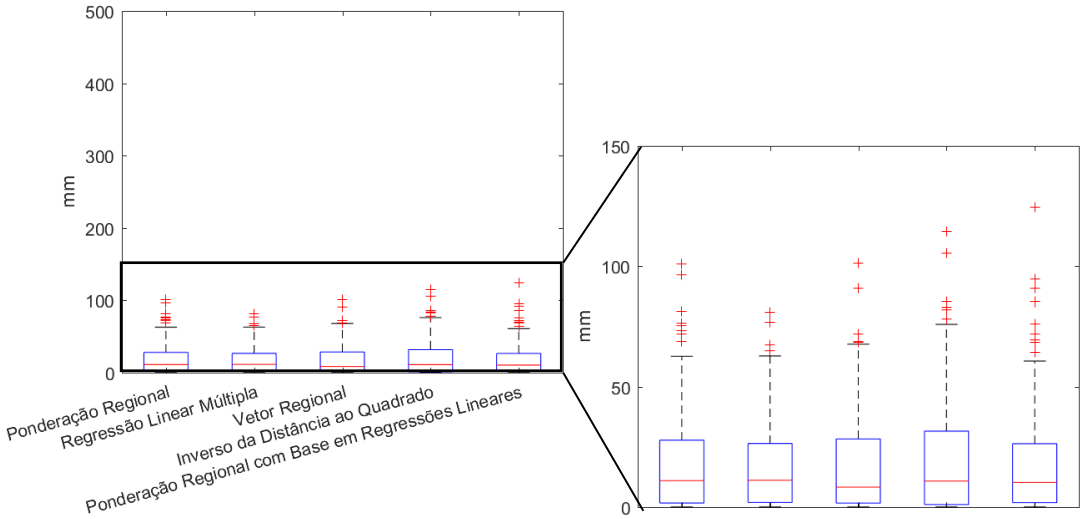


Figura 26. *Box-plot* dos erros das téc. preenchimento mensal Sudeste RJ-ES G1 (2242009).

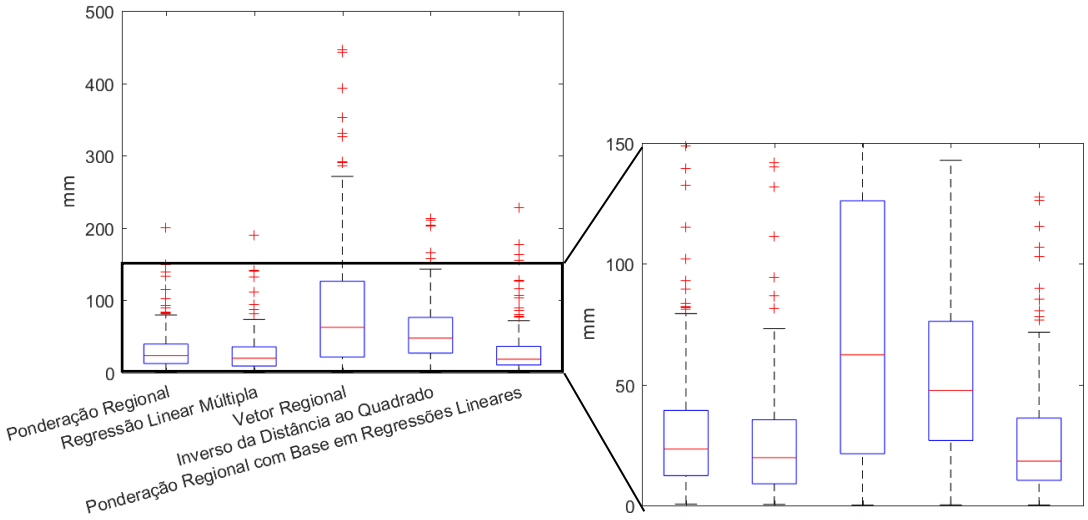


Figura 27. *Box-plot* dos erros das téc. preenchimento mensal Sudeste RJ-ES G2 (2041019).

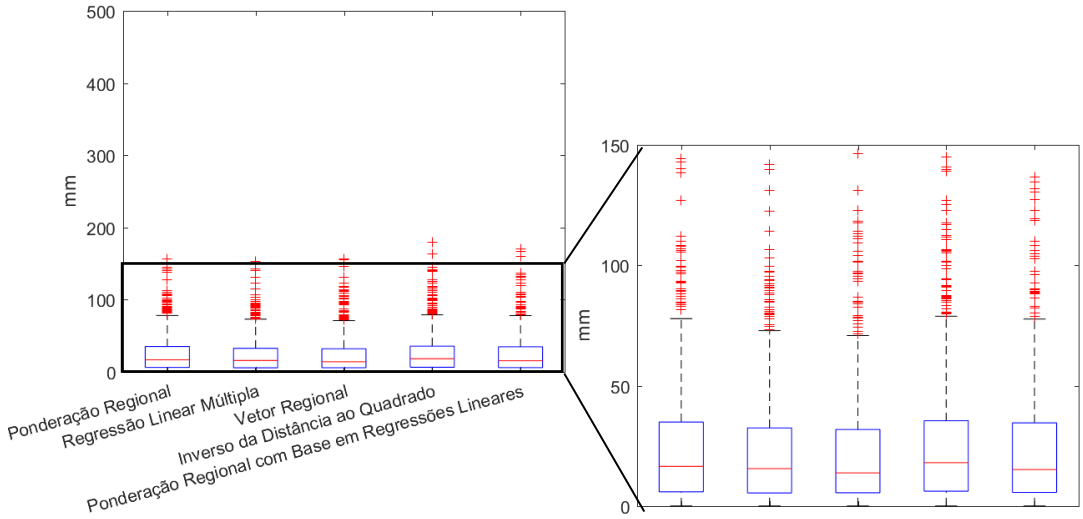


Figura 28. *Box-plot* dos erros das téc. preench. mensal Sudeste SP-MG G1 (2247182).

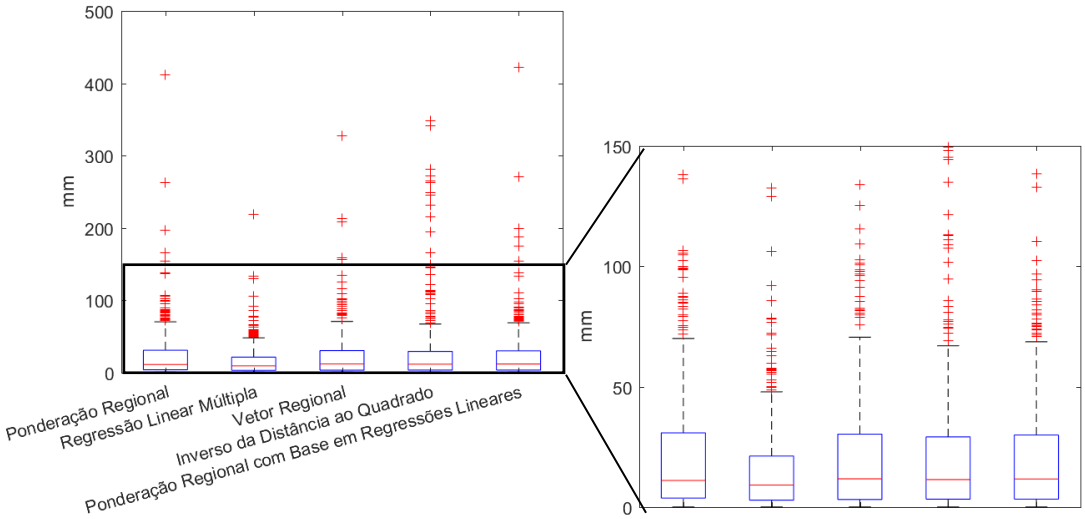


Figura 29. *Box-plot* dos erros das téc. preench. mensal Sudeste SP-MG G2 (2044041).

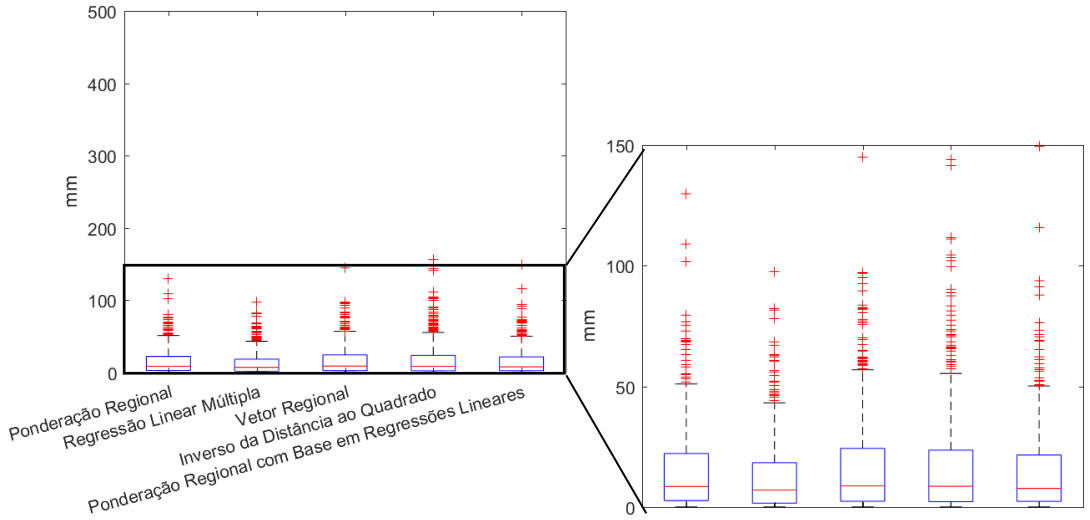


Figura 30. *Box-plot* dos erros das téc. preenchimento mensal Sul G1 (2853035).

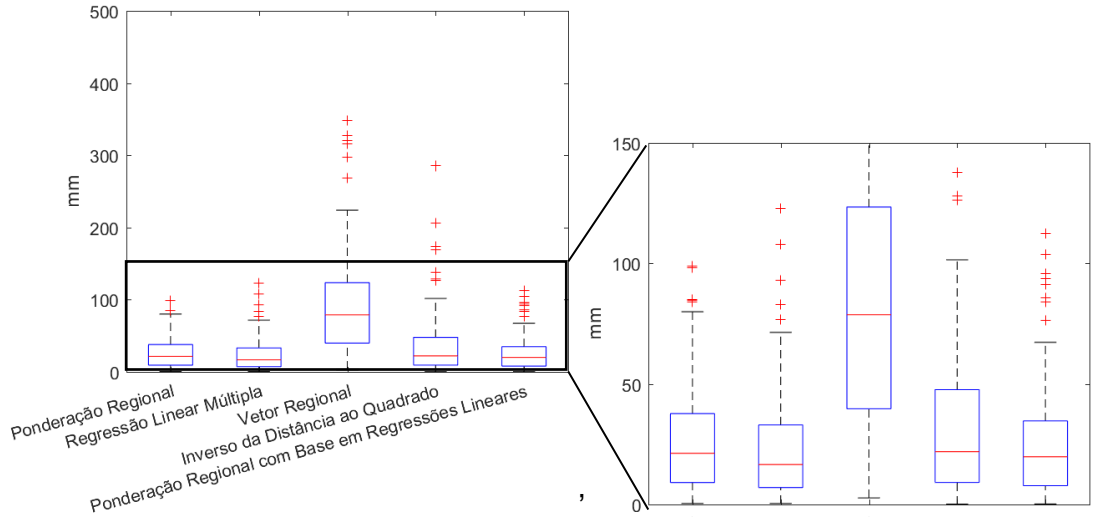
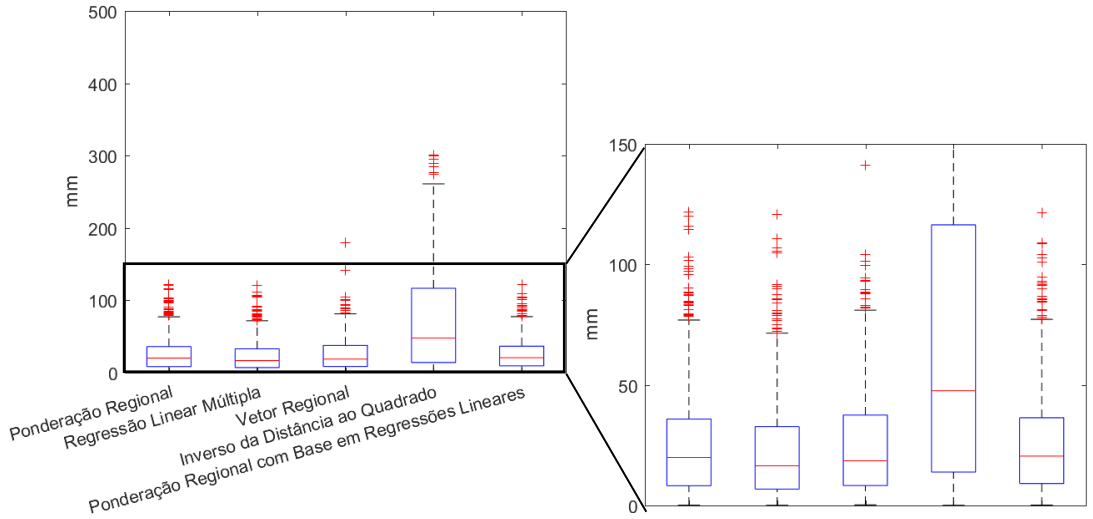


Figura 31. *Box-plot* dos erros das téc. preenchimento mensal Sul G2 (2750010).



Observou-se, nas Figuras anteriores (20 à 31) que em termos gerais a técnica de preenchimento por Regressão Linear Múltipla apresentou o melhor desempenho dentre todos os grupos de análise, quando observada a média aritmética, a mediana, a moda, a amplitude, dentre outras estatísticas. Enquanto isso, a técnica de preenchimento pelo método do vetor regional e do inverso da distância ao quadrado apresentam os piores desempenhos de forma equivalente, sendo o Vetor Regional bastante pior em algumas regiões e com desempenho satisfatório em outras.

O estudo de Bier e Ferraz (2017) apresentou como melhor técnica a Ponderação Regional, que apresenta neste estudo resultados satisfatórios e dentro da média das diferentes técnicas, mas não é a melhor delas. Já o estudo de Oliveira et al. (2010) demonstra que para as estações estudas a melhor técnica foi a Regressão Linear Múltipla, convergindo para os resultados do presente trabalho.

O estudo de Campos et al. (2016) foi realizado para identificar também o desempenho dos melhores métodos de preenchimento de dados pluviométricos, no entanto para totais **anuais** de precipitação, e assim como o presente estudo, também encontraram a regressão linear múltipla como melhor alternativa.

O desempenho das técnicas entre todas as regiões variou, quando avaliada a mediana, entre aproximadamente 9 mm e 78 mm de erro absoluto. A região que apresentou maiores erros médios foi região Norte G2, com média de 62 mm de erro entre a média das diferentes técnicas de preenchimento enquanto os menores erros médios foram da região Sudeste SP-MG G2 de 16 mm.

A Tabela 2 apresenta as estatísticas descritivas das técnicas de preenchimento analisadas em cada estação de cada região.

Tabela 2. Estatísticas Descritivas dos erros de preenchimento (continua)

Estatísticas (mm)	PR	RLM	VR	IDQ	PRRL
Norte G1 (8361000)					
Média Aritmética	47,6	46,6	38,8	63	47,9
Mediana	28,2	31,5	**23,2**	39,8	28,9
Moda	6,6	2,2	3,4	3,5	6,6
Amplitude	320,5	265,7	294,1	368,8	301,3
Amplitude Interquartil	54,2	56,3	41,3	74,8	52,9
Desvio Médio Absoluto	38,8	35,6	31,9	51,1	38,9
Variância	2782,3	2341,1	1988,2	4685,4	2786,7
Desvio Padrão	52,7	48,4	44,6	68,5	52,8
Coeficiente de Variação	1,11	1,04	1,15	1,09	1,10
Curtose	7,7	6,8	8,8	6,5	6,9
Assimetria	2	1,8	2,2	1,8	1,9

Tabela 2. Estatísticas Descritivas dos erros de preenchimento (continuação)

Estatísticas (mm)	PR	RLM	VR	IDQ	PRRL
Norte G2 (260009)					
Média Aritmética	58,6	53,0	48,9	86,4	61,7
Mediana	43,4	**37,2**	38,5	70,4	50,4
Moda	2,8	8,4	12,1	7,0	3,8
Amplitude	283,2	208,2	206,2	340,0	253,6
Amplitude Interquartil	66,0	60,0	55,3	99,7	69,9
Desvio Médio Absoluto	40,3	34,9	33,6	55,1	41,5
Variância	2642,6	1902,3	1785,2	4704,9	2763,7
Desvio Padrão	51,4	43,6	42,3	68,6	52,6
Coeficiente de Variação	0,88	0,82	0,87	0,79	0,85
Curtose	4,9	3,8	4,2	4,0	4,3
Assimetria	1,3	1,1	1,2	1,1	1,2
Centro-Oeste G1 (1656004)					
Média Aritmética	32,4	27,4	28,1	28,7	32,2
Mediana	20,1	17	**16,1**	17	18,3
Moda	0	0	0	0	0
Amplitude	206	168,1	215,5	238,1	226,7
Amplitude Interquartil	38,2	34,2	32,9	35,4	39,8
Desvio Médio Absoluto	27,2	23,2	24,8	24,9	28,1
Variância	1381,1	926,9	1166,6	1127,1	1478,9
Desvio Padrão	37,2	30,4	34,2	33,6	38,5
Coeficiente de Variação	1,15	1,11	1,22	1,17	1,20
Curtose	6,9	5,6	7,7	7,9	7,8
Assimetria	1,9	1,6	2	1,9	2
Centro-Oeste G2 (2054009)					
Média Aritmética	27	25,6	88,8	29,6	27,5
Mediana	17,5	**16,6**	78,4	17	17,9
Moda	0	0	0	0	0
Amplitude	196,1	150,8	371,1	263,8	202,4
Amplitude Interquartil	32,3	31,1	92,3	35,2	33,6
Desvio Médio Absoluto	21,1	19,6	56,4	24,5	21,8
Variância	798,5	673,1	5313,4	1120,1	870,3
Desvio Padrão	28,3	25,9	72,9	33,5	29,5
Coeficiente de Variação	1,05	1,01	0,82	1,13	1,07
Curtose	8	6,6	4,2	11	8,2
Assimetria	1,9	1,7	1,1	2,2	2
Nordeste G1 (340105)					
Média Aritmética	21,2	18,2	17,8	27,4	20,8
Mediana	9,1	7,2	**5,6**	8,3	7,9
Moda	0	0	0	0	0
Amplitude	179,4	169,6	166,9	246,9	202,2
Amplitude Interquartil	33,8	25,1	26,9	41,4	28,7
Desvio Médio Absoluto	21,6	19,2	19	29,6	21,8
Variância	877,1	765,9	677,8	1652,9	936,1
Desvio Padrão	29,6	27,7	26	40,7	30,6
Coeficiente de Variação	1,40	1,52	1,46	1,49	1,47
Curtose	8,9	10,9	10,1	9,1	11,1
Assimetria	2,2	2,5	2,3	2,2	2,4
Nordeste G2 (638090)					
Média Aritmética	19,8	17,8	18,2	21	19,1
Mediana	10,9	11,1	**8,2**	10,7	10,2
Moda	0	0	0	0	0
Amplitude	101	81	101,3	114,6	124,4
Amplitude Interquartil	26,1	24,5	26,7	30,5	24,5
Desvio Médio Absoluto	18,6	15,7	17,5	20	17,8
Variância	565,7	401,6	503,2	653,5	577,8
Desvio Padrão	23,8	20	22,4	25,6	24
Coeficiente de Variação	1,20	1,12	1,23	1,22	1,26
Curtose	4,4	3,8	4,6	4,7	6,6
Assimetria	1,5	1,3	1,5	1,5	1,9
Sudeste RJ-ES G1 (2242009)					
Média Aritmética	31,1	26,5	91,2	56	29,1
Mediana	23,3	**19,7**	62,3	47,6	18,3
Moda	7,4	1,3	1,5	2,7	1,6
Amplitude	199,7	189,5	446,5	213,1	228,3
Amplitude Interquartil	27	26,5	104,8	49,4	25,8
Desvio Médio Absoluto	20,1	18,5	69,6	31	21,6
Variância	860,2	766,5	8278,5	1712,6	1130,4
Desvio Padrão	29,3	27,7	91	41,4	33,6

Tabela 2. Estatísticas Descritivas dos erros de preenchimento (conclusão)

Coeficiente de Variação	0,942	1,045	0,998	0,739	1,155
Curtose	11	12,8	5,4	5,7	13,5
Assimetria	2,4	2,7	1,5	1,5	2,9
Sudeste RJ-ES G2 (2041019)					
Média Aritmética	25,3	23,9	23	26,6	25
Mediana	16,4	15,5	**13,7**	18	15,1
Moda	0,1	1,6	6,3	1,8	2,8
Amplitude	156,4	152,5	156,7	179,7	169,8
Amplitude Interquartil	29	27	26,3	29,3	28,9
Desvio Médio Absoluto	20,4	19,1	18,8	20,8	20,5
Variância	732,7	645,7	656,7	803,6	768,7
Desvio Padrão	27,1	25,4	25,6	28,3	27,7
Coeficiente de Variação	1,07	1,06	1,11	1,06	1,11
Curtose	6,1	6,3	7,7	6,9	7,4
Assimetria	1,7	1,7	2	1,9	1,9
Sudeste SP-MG G1 (2247182)					
Média Aritmética	24,5	15,9	24,1	26,3	24,3
Mediana	11	**9,1**	11,7	11,4	11,6
Moda	0	0,1	0,6	0	0,1
Amplitude	777,1	218,3	722,6	348,4	799,8
Amplitude Interquartil	27,1	18,3	27,2	25,9	26,7
Desvio Médio Absoluto	23,7	13,7	22,6	26	23
Variância	2391	435	2033	2127,2	2476,6
Desvio Padrão	48,9	20,9	45,1	46,1	49,8
Coeficiente de Variação	2,00	1,31	1,87	1,75	2,05
Curtose	128,5	25,6	127,6	21,4	135
Assimetria	9,3	3,6	9,1	3,9	9,7
Sudeste SP-MG G2 (2044041)					
Média Aritmética	15,3	12,8	16,8	17,7	15,1
Mediana	8,5	**7**	8,7	8,5	7,7
Moda	0	0	0	0	0
Amplitude	129,8	97,5	145	156,8	149,6
Amplitude Interquartil	19,5	16,8	21,9	21,3	19,2
Desvio Médio Absoluto	13,4	11,4	15,6	16,7	13,6
Variância	327,4	241	445	564,3	346
Desvio Padrão	18,1	15,5	21,1	23,8	18,6
Coeficiente de Variação	1,18	1,21	1,26	1,34	1,23
Curtose	8,9	7,7	7,5	9,9	11,3
Assimetria	2,1	2	2	2,4	2,3
Sul G1 (2853035)					
Média Aritmética	25,8	21,8	90	35,9	24,7
Mediana	21,2	**16,5**	78,7	21,9	19,7
Moda	6,2	1,3	32,9	1,8	1,8
Amplitude	98,7	122,5	346	285,2	112,1
Estatísticas (mm)	**PR**	**RLM**	**VR**	**IDQ**	**PRRL**
Amplitude Interquartil	28,6	26	83,8	38,5	26,9
Desvio Médio Absoluto	17	15,4	50,2	28,3	16,8
Variância	440	419,5	4329,1	1622,7	498,1
Desvio Padrão	21	20,5	65,8	40,3	22,3
Coeficiente de Variação	0,81	0,94	0,73	1,12	0,90
Curtose	3,8	7,7	5,7	12,7	5,4
Assimetria	1	1,8	1,4	2,6	1,5
Sul G2 (2750010)					
Média Aritmética	26,1	22,7	25,6	73,3	26,1
Mediana	19,8	**16,4**	18,5	47,5	20,4
Moda	2,3	0,5	6,5	7,8	1,1
Amplitude	121,7	120,7	179	651,4	121,2
Amplitude Interquartil	27,7	25,9	29,3	102,6	27,3
Desvio Médio Absoluto	17,8	16,2	18	58,2	17,3
Variância	540,6	444,5	539,6	5313,5	508
Desvio Padrão	23,3	21,1	23,2	72,9	22,5
Coeficiente de Variação	0,89	0,93	0,91	0,99	0,86
Curtose	4,8	5,4	7,3	9,2	4,7
Assimetria	1,4	1,5	1,6	1,7	1,4

*PR: Ponderação Regional; RLM: Regressão Linear Múltipla; VR: Vetor Regional: IDQ: Inverso da Distância ao Quadrado; PRRL: Ponderação Regional com Base em Regressões Lineares.

Alguns pontos em comum dentre as técnicas e regiões é a moda sempre próxima a zero, ou seja, há mais casos em que o valor do erro fica próximo de zero do que qualquer outro valor. Além disso, percebe-se uma maior persistência do valor da mediana quando comparada à média, sendo sempre inferior por ser menos influenciada pelos valores extremos dos erros.

A assimetria dos histogramas de erros é sempre positiva, uma vez que a cauda do histograma sempre tende a ser maior para a direita, para os valores extremos, estando a média mais próxima do extremo inferior dos erros.

É possível traçar uma correlação linear direta entre a precipitação média diária e a média dos erros medianos das técnicas de preenchimento. Da mesma forma, com menor intensidade, é possível visualizar uma tendência semelhante entre o percentual de dias chuvosos e a média dos erros medianos das técnicas de preenchimento. A Figura 32 e a Figura 33 ilustram tais correlações respectivamente.

Quanto maior a precipitação média diária e o percentual de dias chuvosos, maior é o erro mediano entre as técnicas de preenchimento, o que possivelmente deve-se à grande variação espacial e temporal do fenômeno o que impede que seja traçada uma boa correlação linear entre estações, mesmo que próximas. Com um maior número de estações analisadas, se poderá afirmar com maior grau de segurança tal constatação.

Segundo Einfalt, Johann e Pfister (1998), quanto maior o volume de chuva, maior é a variabilidade espacial do fenômeno. Tal argumento corrobora com a tese acima exposta de que quanto maior é a precipitação média diária, maiores são as incertezas das técnicas de preenchimento através de estações vizinhas, uma vez que quanto maiores são os fenômenos chuvosos (maiores volumes), maior é a variabilidade espacial, o que influi diretamente na correlação entre estações.

Figura 32. Correlação linear entre a precipitação média diária e a média dos erros medianos das técnicas de preenchimento.

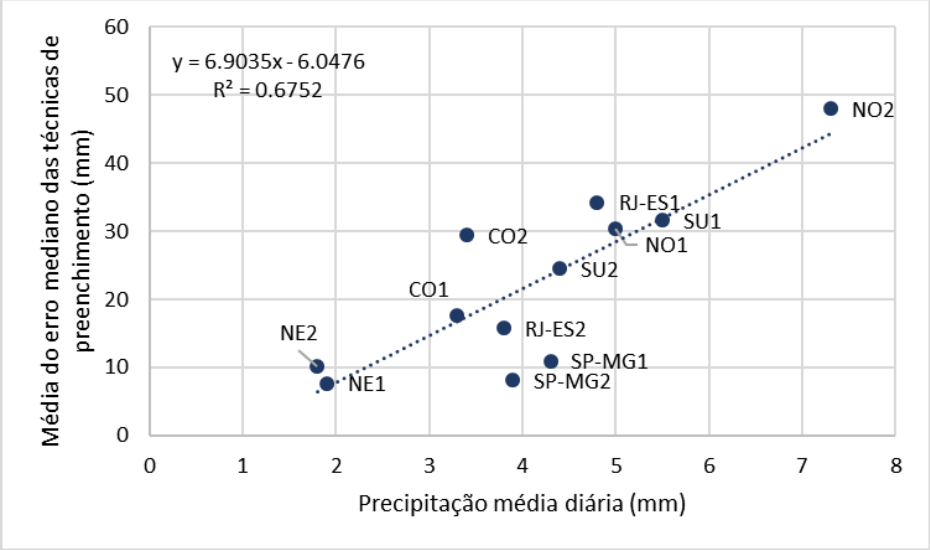
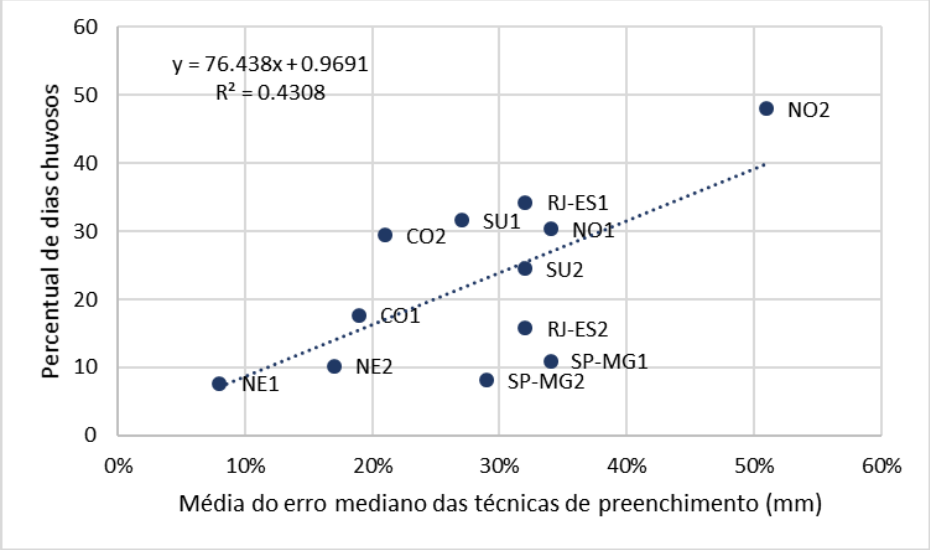


Figura 33. Correlação linear entre o percentual de dias chuvosos e a média dos erros medianos das técnicas de preenchimento.



Outros aspectos influenciam o eventual erro nos preenchimentos como a heterogeneidade da região ou a qualidade das estações analisadas. Por exemplo, estações mais próximas a centros urbanos deveriam tender a ter melhor qualidade de operação dada a facilidade de acesso, portanto, um grupo de estações vizinhas pode melhor explicar o comportamento entre elas.

A região com maior heterogeneidade de chuvas, como é o caso da região Norte G2 e também por apresentar um maior total precipitado apresentou os maiores erros medianos dentre as regiões. Já regiões com menor precipitação

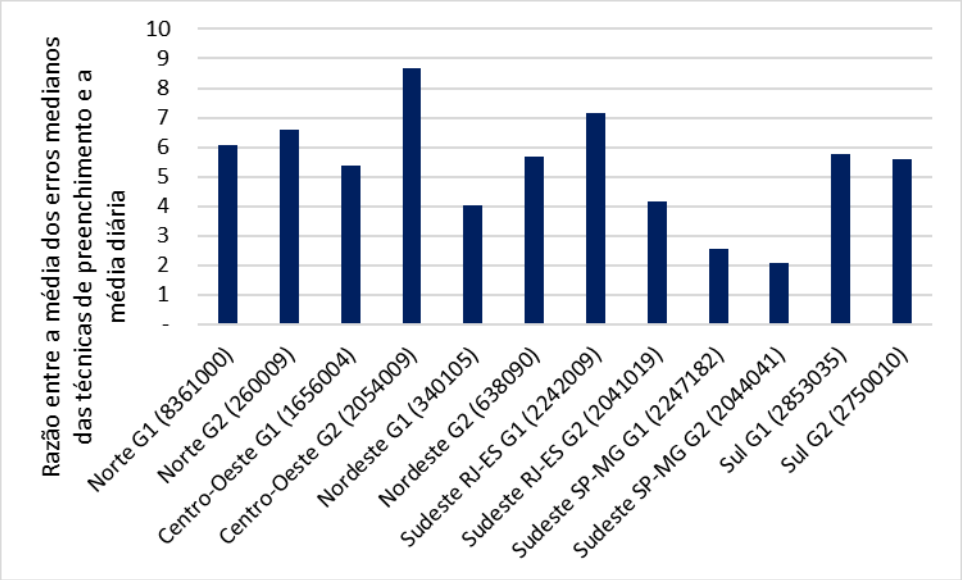
média diária, como é o caso das regiões Nordeste, apresentaram a menor média dos erros de técnicas de preenchimento. Tal fato, de quanto maior os totais precipitados, maiores são os erros, foi demonstrado e ilustrado na correlação realizada na Figura 32, e balizado pelos estudos de Einfalt, Johann e Pfister (1998).

A diferença entre as regiões Sudeste SP-MG e as regiões Nordeste, que apesar de apresentarem as médias dos erros medianos entre as técnicas de preenchimento similares, é que as regiões Sudeste SP-MG apresentam uma precipitação média diária superior (e percentual de dias chuvosos) às regiões Nordeste. Tal fato se deve provavelmente pela heterogeneidade entre as chuvas das duas regiões ou então pela qualidade do monitoramento dos grupos analisados.

Quando analisado o erro médio relativo à precipitação média diária das diferentes técnicas, ele tende a estar em um patamar entre 4 e 12 vezes a precipitação média diária. Aparentemente, a região sudeste apresenta os menores valores enquanto as regiões Centro-Oeste e Norte apresentam os maiores valores.

Quando analisado sobre o ponto de vista da razão entre a média dos erros medianos das técnicas de preenchimento e a média de precipitação diária total de cada região, nota-se de forma mais acentuada os menores valores da região Sudeste SP-MG, sem se observar outros padrões dentre as outras regiões como se observa na Figura 34.

Figura 34. Razão entre a média dos erros medianos das técnicas de preenchimento e a média diária por região.

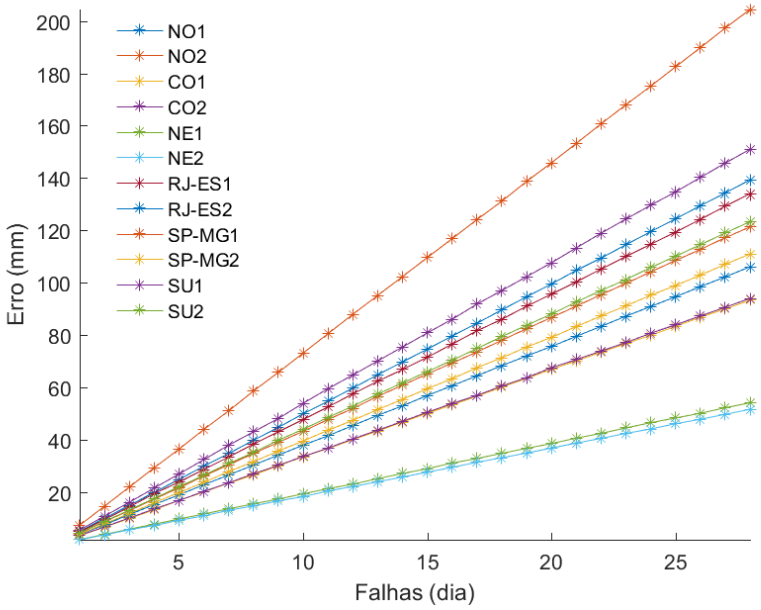


5.3 GERAÇÃO DE SÉRIES SINTÉTICAS COM FALHAS

As séries com falhas sintéticas foram geradas com o intuito de se avaliar o efeito que falhas diárias geram em totais mensais de precipitação. Foram avaliadas diferentes extensões (1 a 28 dias), gerando-se 1.000 séries sintéticas para a estação escolhida de cada região. A partir dessas séries, pôde-se também gerar séries dos erros absolutos criados pelas falhas sintéticas e avaliar os histogramas de erros e suas estatísticas descritivas.

A Figura 35 apresenta os resultados dos erros médios para cada estação escolhida de cada região com a variação da extensão da falha (distribuída aleatoriamente ao longo de cada mês) de 1 a 28 dias.

Figura 35. Gráfico do erro médio mensal causado por falhas de extensão variada (dia) nas diferentes regiões de estudo.



*NO: Norte; CO: Centro-Oeste; NE: Nordeste; RJ-ES: Rio de Janeiro e Espírito Santo; SP-MG: São Paulo e Minas Gerais; SU: Sul.

Observou-se, conforme esperado, que em média o erro que uma falha diária causa no total mensal precipitado é igual à média de precipitação diária daquela estação. Como se pode observar, o crescimento ocorre de forma linear, sendo acrescida a média de precipitação diária a cada acréscimo de falha diária introduzido. A região Norte G2 apresentou um erro absoluto do não preenchimento acentuadamente superior às demais regiões, devido aos maiores registros pluviométricos observados, já as regiões Nordeste apresentaram os menores valores.

Também puderam ser avaliados os desvios padrão dos erros, a amplitude, o percentual das vezes que o erro era igual a zero, ou seja, em que não houve chuva, sendo este igual a probabilidade de ser um dia não chuvoso, sendo apresentados mais à frente.

A Tabela 3 apresenta o percentual de dias chuvosos de cada estação analisada de cada região. Novamente, nota-se que a estação da região Norte G2, localizada no Amazonas é a que apresenta o maior percentual de dias chuvosos em contraponto à estação localizada na região Nordeste G1 no estado do Ceará que apresenta o menor percentual de dias chuvosos.

Tabela 3. Percentual de dias chuvosos de cada estação analisada em cada região

Região	Percentual de dias chuvosos
Norte G1 (8361000)	34%
Norte G2 (260009)	51%
Centro-Oeste G1 (1656004)	19%
Centro-Oeste G2 (2054009)	21%
Nordeste G1 (340105)	8%
Nordeste G2 (638090)	17%
Sudeste RJ-ES G1 (2242009)	32%
Sudeste RJ-ES G2 (2041019)	32%
Sudeste SP-MG G1 (2247182)	34%
Sudeste SP-MG G2 (2044041)	29%
Sul G1 (2853035)	27%
Sul G2 (2750010)	32%

Tais estatísticas de dias chuvosos podem ser relacionadas às estatísticas descritivas do prejuízo das falhas diárias na série mensal, quanto maior o percentual de dias chuvosos, tende a ser maior a precipitação diária média, logo a perda de registros diários acarreta em menores totais mensais precipitados.

Outra forma de interpretar as informações de percentual de dias chuvosos, é que, por exemplo, na região Centro-Oeste G1 há 81% de chances (extrapolando as estatísticas da amostra para a população) de que uma falha de registro diário não cause prejuízo ao total mensal precipitado sem distinção do mês do ano, dado que apenas 19% dos dias são chuvosos.

A Tabela 4 apresenta as estatísticas descritivas que a falha de um dia em um mês qualquer das estações analisadas em cada região gera de perda no total precipitado mensal. Estas estatísticas levam em consideração os dias chuvosos e não chuvosos. Aumentando-se o número de dias de falha, cresce o valor de precipitação não medido, linearmente. Foram geradas séries sintéticas suficientes com falhas sintéticas até que as estatísticas descritivas das falhas se igualassem às estatísticas das séries diárias das estações.

Percebe-se que a moda, em todos os casos, iguala-se a zero uma vez que as falhas diárias tendem a ter mais valores zerados do que qualquer outro valor (vide Tabela 3). Da mesma forma, a mediana, diferente da média, também resiste às influências de valores acima da amplitude interquartil. Em todas as regiões, com exceção da Norte G2, a mediana dos erros também foi igual a zero, ou seja, mais da metade dos valores que uma falha de dados diária causa é igual a zero. Em regiões

como o Centro-Oeste e o Nordeste, os valores das amplitudes interquartis também se igualaram a zero.

Tabela 4. Estatísticas Descritivas de 1 falha diária em 1 mês qualquer em cada região analisada

Estatística	NO1	NO2	CO1	CO2	NE1	NE2	RJ-ES1	RJ-ES2	SP-MG1	SP-MG2	SU1	SU2
Média Aritmética [mm]	5,0	7,3	3,3	3,4	1,9	1,8	4,8	3,8	4,3	3,9	5,5	4,4
Mediana [mm]	0,0	0,4	0	0	0	0	0	0	0	0	0	0
Moda [mm]	0,0	0	0	0	0	0	0	0	0	0	0	0
Amplitude [mm]	180,3	135,3	180	155	114	135,5	164,5	202,4	353	190,2	151,8	119
Amplitude Interquartil [mm]	3,0	10	0	0	0	0	3,7	1,7	1,9	1,3	0,9	3
Desvio Médio Absoluto [mm]	7,3	8,8	5,5	5,5	3,5	3,1	6,9	5,7	6,5	6	8,5	6,4
Variância [mm^2]	143,8	174,5	107,2	113,6	73,5	53	136,6	105,6	141	108,3	218,5	107,5
Desvio Padrão [mm]	12,0	13,2	10,4	10,7	8,6	7,3	11,7	10,3	11,9	10,4	14,8	10,4
Coeficiente de Variação [-]	2,4	1,81	3,15	3,15	4,53	4,06	2,44	2,71	2,77	2,67	2,69	2,36
Curtose [-]	21,8	17,5	36,6	36,1	47,1	68,7	27,3	34,1	100,9	25,2	23,6	18,4
Assimetria [-]	3,7	3,2	4,9	5	6,1	6,9	4,1	4,6	6,6	4,1	4	3,5

*NO: Norte; CO: Centro-Oeste; NE: Nordeste; RJ-ES: Rio de Janeiro e Espírito Santo; SP-MG: São Paulo e Minas Gerais; SU: Sul.

De forma segura, é possível afirmar que **em média** o erro que uma falha diária causa de prejuízo no total mensal precipitado é significativamente inferior ao erro que técnicas de preenchimento tradicionais em nível mensal causam na série, variando entre todas as regiões entre 4 e 10 vezes aproximadamente.

A mediana dos erros na maioria dos casos é igual a 0, ou seja, na maioria dos dias em que houve falha, a falha era igual a zero, independente da época do ano, portanto a moda em todas as regiões também é igual a 0. Estes são indícios relevantes, de que a probabilidade de acerto, ao não se preencher dados faltantes (ou se preencher com zeros) para o cálculo do total mensal precipitado é uma solução trivial, mas eficiente. Em outras palavras, a técnica de preenchimento nada mais é do que o preenchimento das falhas diárias com o valor de precipitação diária mais frequente que é **0 mm**.

As regiões com menores médias aritméticas das falhas foram as regiões Nordeste, enquanto as com maior foram a região Norte G2 e a região Sul G1. As regiões Nordeste, conforme mencionado por estarem localizadas em regiões de clima As (Tropical Savânico) ou BSh (Seco Semiárido Quente) apresentam precipitações anuais menores que em outras regiões do Brasil (BSh, menores que

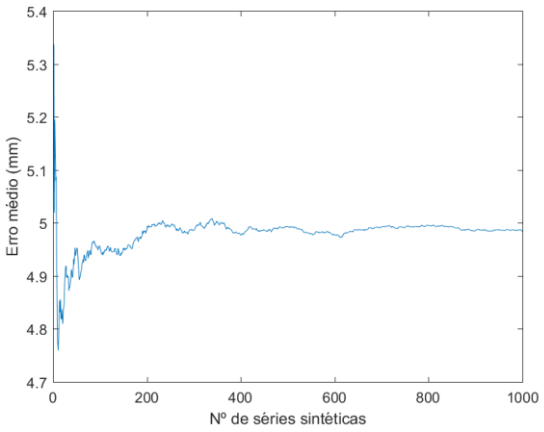
650 mm e As entre 700 e 1.500 mm) e portanto, quando há falhas, tendem a serem 0 mm ou próximas, diminuindo o erro médio causado por 1 dia de falha.

Já regiões como a Norte G2, que apresentam um clima do tipo Am (Tropical Monçônico) variam a sua precipitação em 1.200 e 3.300 mm o que podem ser valores considerados altos de precipitação, e por isso quando ocorrem falhas, a probabilidade de que cause erros no total mensal precipitado é maior. Da mesma forma, a região Sul G1, apresentou o segundo maior valor de erro, não só por apresentar um volume anual de precipitações entre 1.200 e 2.040 mm, mas também por apresentar precipitações mensais bem distribuídas, o que incorre em maiores dias de chuva, aumentando a probabilidade de que um dia de falha, apresente um erro superior a 0 mm.

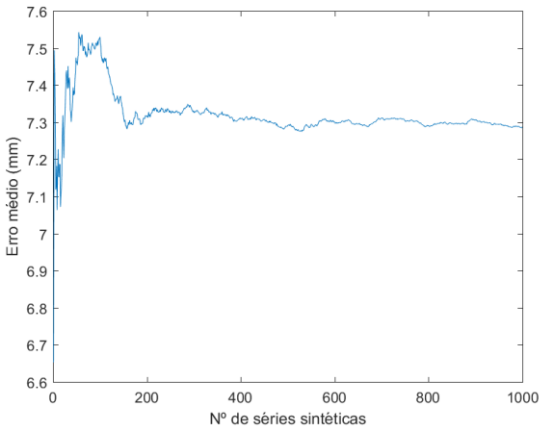
Conforme relatado anteriormente, ao se aumentar o número de séries sintéticas e se avaliar o erro médio para cada extensão de falhas, observa-se conforme exposto nos gráficos da Figura 36 a convergência do erro médio para a estabilização no entorno da precipitação média diária.

Figura 36. Convergência do erro médio diário de 1.000 séries sintéticas com uma falha diária.

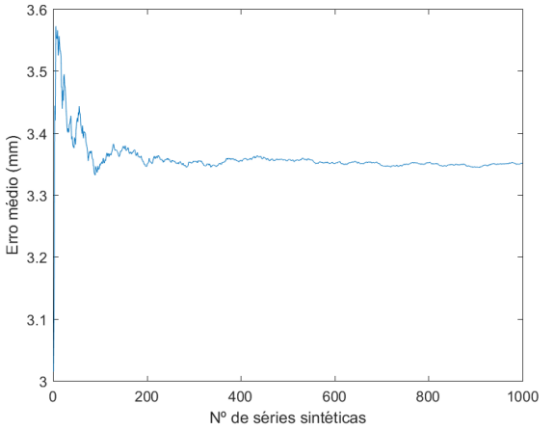
a) Norte G1



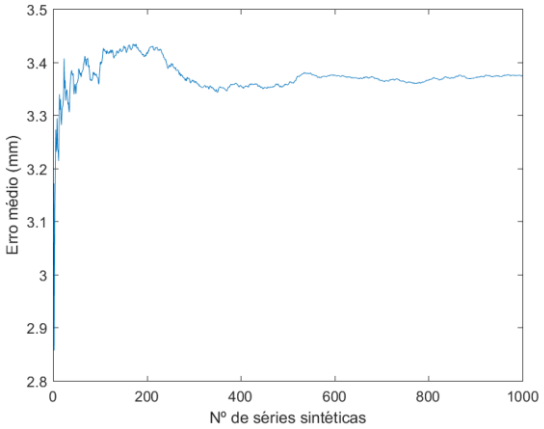
b) Norte G2



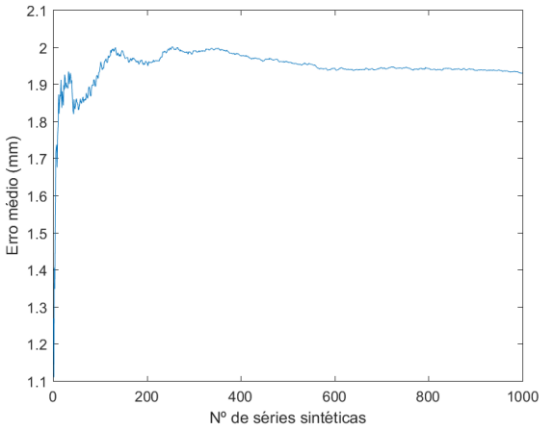
c) Centro-Oeste G1



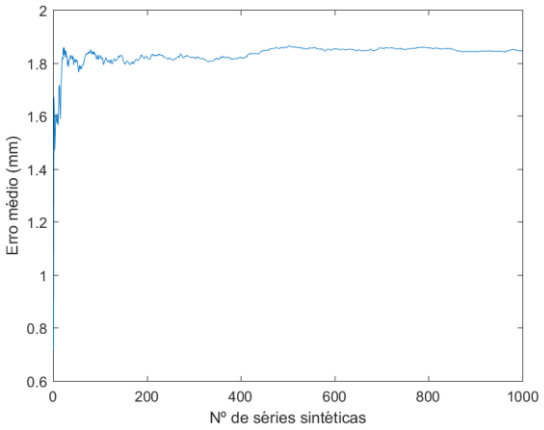
d) Centro-Oeste G2



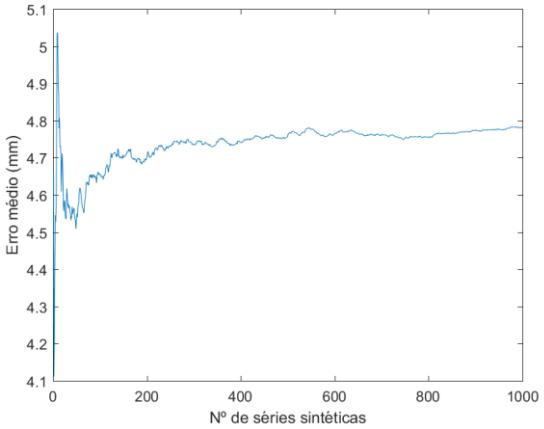
e) Nordeste G1



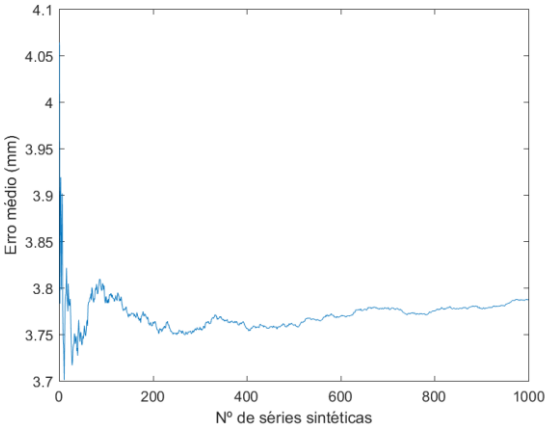
f) Nordeste G2



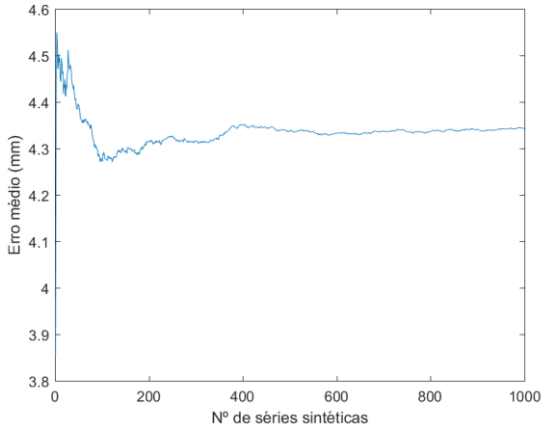
g) Sudeste RJ-ES G1



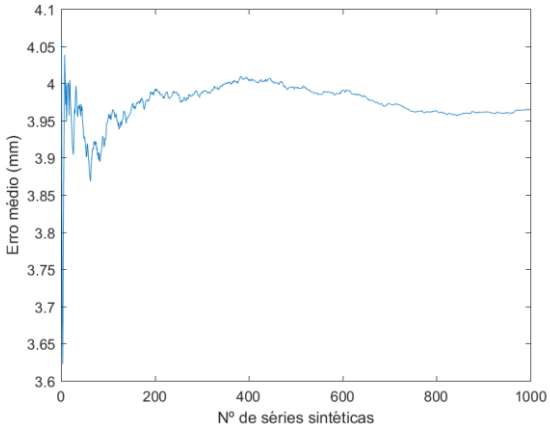
h) Sudeste RJ-ES G2



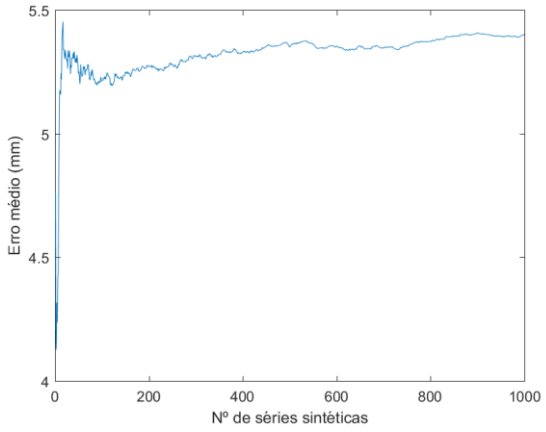
i) Sudeste SP-MG G1



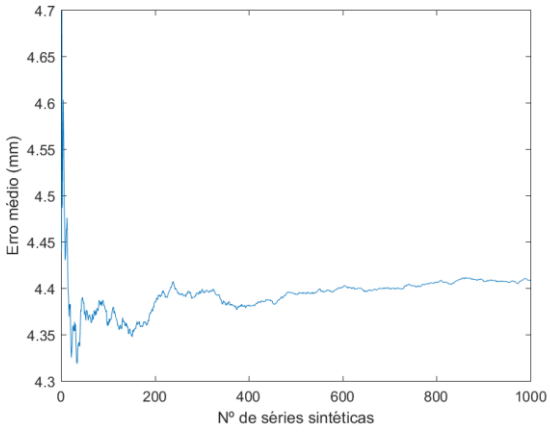
j) Sudeste SP-MG G2



k) Sul G1



l) Sul G2



De forma análoga aos gráficos acima expostos, foram geradas também 1.000 séries sintéticas aleatórias para falhas variando de 2 a 28 dias. Foram observadas as mesmas tendências, com a média dos erros tendendo à precipitação média diária multiplicada pelo número de falhas diárias.

A partir desta comparação de resultados é possível então traçar limites do número falhas diárias que *em média* o erro das falhas é inferior ao erro das técnicas de preenchimento utilizadas. O limite de falhas (dias) pode ser calculado como sendo igual ao erro médio das técnicas de preenchimento dividido pela precipitação média diária da região. Tal conclusão, apesar de simples, pode ser extrapolada para tantas estações quantas forem necessárias, criando-se, assim, mapas com regiões homogêneas quanto aos critérios de falhas limites sem necessidade de descarte do total mensal precipitado e posterior preenchimento.

A Figura 37 apresenta o número de dias máximo de falhas diárias que em média acarretará erros inferiores aos erros ocasionados pelo preenchimento

tradicional para cada região. A Figura 38 apresenta o *box-plot* do máximo de falhas diárias que em média terá um erro inferior ao do preenchimento.

Figura 37. Número de dias máximo de falhas diárias com erro mensal inferior ao das técnicas de preenchimento para cada região.

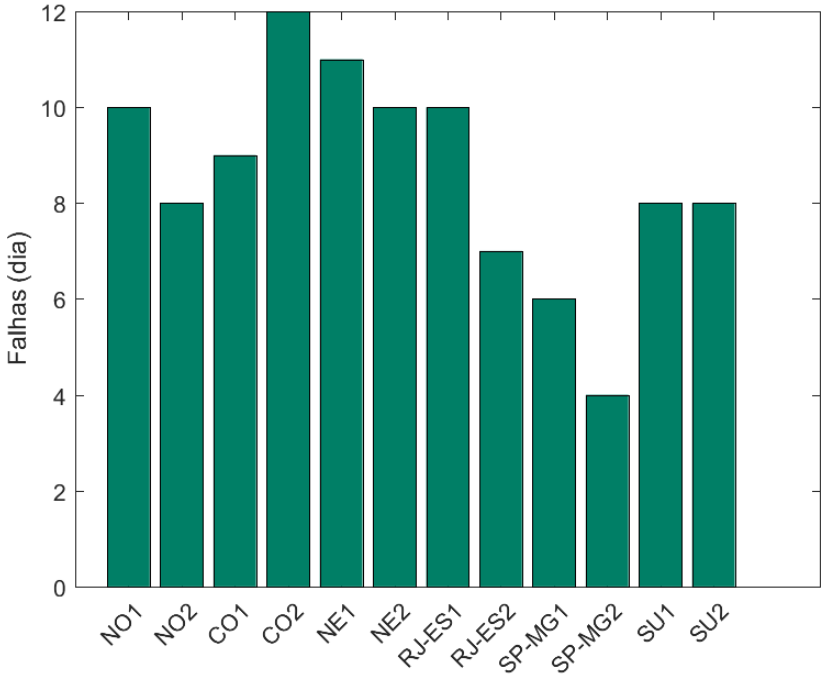
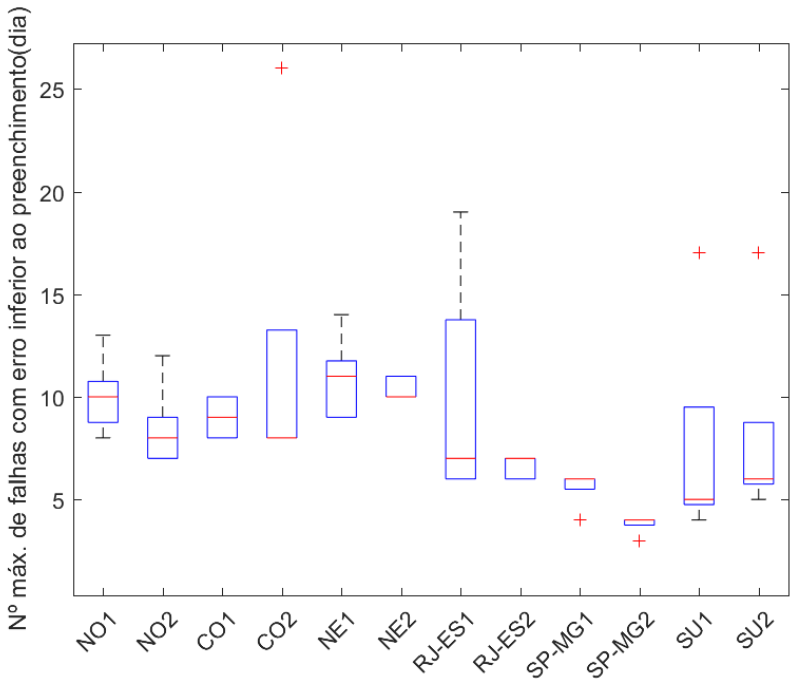


Figura 38. Gráfico *box-plot* do número de dias máximo de falhas diárias com erro mensal inferior ao das técnicas de preenchimento para cada região.



Analisando os dois gráficos acima, é possível perceber pelo primeiro que a média dentre todas as regiões analisadas no estudo do número máximo de falhas diárias que o prejuízo causado por tais falhas causa ao total mensal precipitado, de forma a ser inferior às técnicas tradicionais de preenchimento flutua em torno de um valor médio próximo de oito dias, variando de 4 a 12 dias, dependendo da região. Já a segunda figura permite observar que em determinadas regiões a dispersão de resultados varia mais em algumas do que outras, com poucos valores extremos.

Traçando um paralelo à média dos erros médios absolutos das técnicas de preenchimento e analisando o gráfico da Figura 38 percebe-se que as regiões SP-MG1 e SP-MG2, que apresentam os menores limites de falhas são as que apresentaram menores erros de técnicas de preenchimento dentre todas as regiões. O limite de falhas diário aceitável é reduzido devido à melhor assertividade dos métodos de preenchimento nessas duas regiões enquanto o oposto ocorre em regiões como a AM1, AM2 e NE1.

Retomando a análise de percentual de dias chuvosos nas estações de cada região, percebe-se quando analisados os extremos (Norte G2 com 51% de dias chuvosos e Nordeste G1 com apenas 8%) e relacionando com o *box-plot* da Figura 38 não é possível perceber uma relação linear entre um maior percentual de dias chuvosos e maior ou menor aceitabilidade de falhas. Novamente, esse fato se deve mais à qualidade dos preenchimentos realizados, podendo ser influenciado pela qualidade dos dados utilizados, mas também por se tratar de erros absolutos, que quanto maior forem os totais mensais precipitados, maiores poderão ser os erros absolutos conforme correlação exposta anteriormente na Figura 33.

Nos gráficos subsequentes que exploram graficamente essa distribuição a nível mensal, fica visível a homogeneidade ao longo do ano da estação localizada no Amazonas (Norte G2) em contraponto à estação localizada no Ceará (Nordeste G1). Esta última apresenta diferenças bastante marcantes dependendo da época do ano quanto a aceitabilidade de falhas devido à heterogeneidade temporal anual.

Vale lembrar que os resultados dos gráficos da Figura 38 levam em consideração a média do erro médio das técnicas de preenchimento para todos os meses disponíveis das séries de dados das estações escolhidas de cada região. Cabe destacar que são análises pontuais, não representando em si a região em que as estações estão situadas, uma vez que seriam necessários diversos pontos

geográficos de uma determinada região para se poder descrever as características pluviométricas de uma determinada região.

De qualquer forma, pôde-se observar características semelhantes entre as estações localizadas em regiões próximas, como é o caso de São Paulo e Minas Gerais (SP-MG G1 e SP-MG G2 respectivamente) que apresentaram um baixo valor de falha diária permitida, em contraponto às estações localizadas na região Nordeste que apresentaram os maiores valores (10 e 11 dias).

A maior ou menor amplitude de falhas aceitáveis diárias dentre as diferentes regiões pode ser função de diferentes fatores:

1. Heterogeneidade das precipitações na região da estação analisada e das estações de apoio em determinadas épocas do ano;
2. Problemas nas séries de apoio, o que repercute na qualidade do preenchimento.
3. Problemas na série de dados analisada, o que repercute na quantificação do suposto erro de preenchimento, quando na verdade pode ser um erro inerente da estação analisada.

O primeiro aspecto é o que se pretende compreender melhor nesse estudo, no entanto os dois últimos são inerentes do processo de observação do fenômeno. Quanto ao segundo aspecto, não se teve a preocupação de eliminar eventuais problemas nas séries de apoio (além de excluir da análise os meses com falhas), uma vez que é um aspecto ao qual o processo de consistência e homogeneização de dados pluviométricos está sujeito.

O terceiro aspecto é o mais preocupante na presente análise, uma vez que interfere nos resultados observados, mas também de difícil eliminação uma vez que a presente análise se propõe a observar os erros da técnica de preenchimento e compará-la à original, quando na verdade a própria série original pode conter erros. Por esse motivo, vale lembrar que se buscou, através da análise das curvas duplo-acumulativas, as estações que apresentaram melhores séries dentre os grupos de forma a diminuir tais erros.

Levando-se em consideração também estatísticas como a mediana, menos influenciada por tais flutuações ocasionadas devido a erros de medição, é possível

inferir que uma vez replicada tal técnica para diversos postos de uma determinada região, será possível observar os padrões e propor faixas limites para a tomada de decisão entre a correção de um mês com falhas de dados ou não.

Para tanto, também podem ser geradas faixas limites de decisão dependendo do mês em que se deve decidir entre o preenchimento ou não. Para ilustrar tal análise, foram gerados também gráficos do tipo *box-plot* para cada região e para cada mês do ano, sendo apresentados da Figura 39 à 50.

Cabe destacar que em alguns casos, o critério de número de falhas permitidas supera o valor de dias em um mês. Dado que o valor apresentado é uma **razão** entre a média dos erros das técnicas de preenchimento **sobre** a precipitação média diária. Em outras palavras, o erro de técnicas de preenchimento pode ser superior ao total mensal precipitado em um determinado mês do ano (ser maior do que 100% do total mensal precipitado).

Figura 39. *Box-plot* do máximo de falhas diárias com erro mensal inferior ao das técnicas de preenchimento (esq.) e precipitação média histórica mensal (dir.) na região Norte G1.

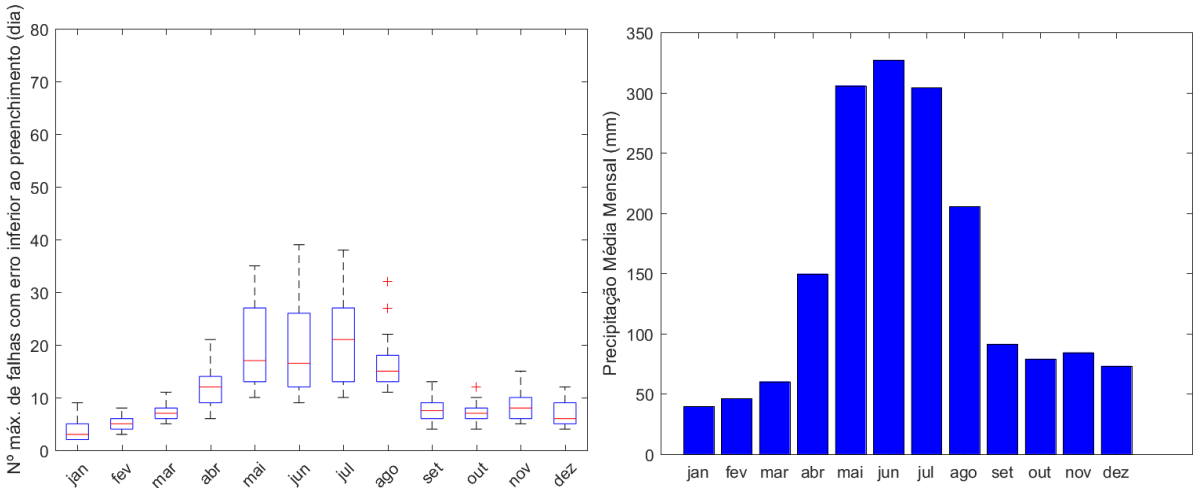


Figura 40. *Box-plot* do máximo de falhas diárias com erro mensal inferior ao das técnicas de preenchimento (esq.) e precipitação média histórica mensal (dir.) na região Norte G2.

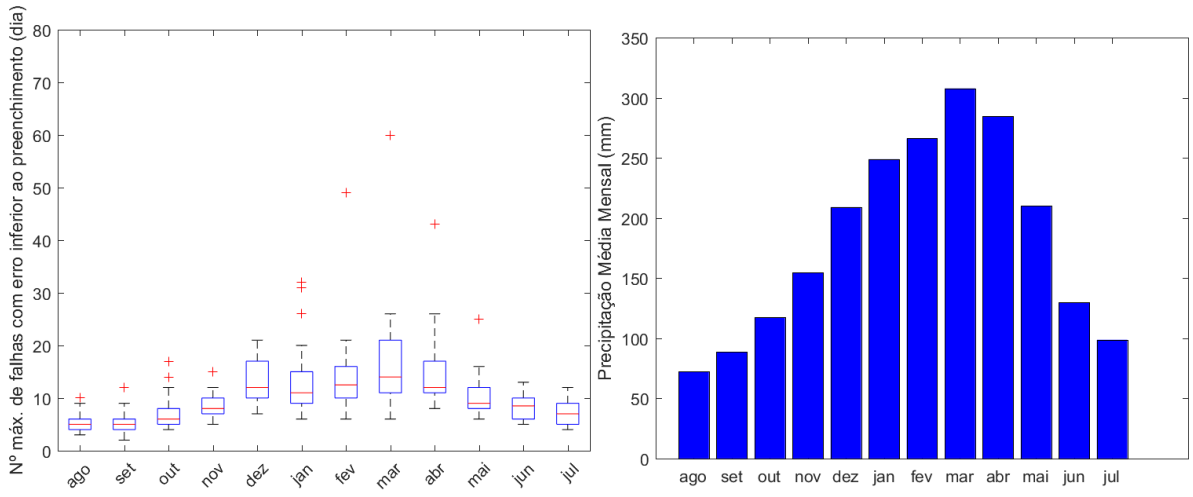


Figura 41. *Box-plot* do máximo de falhas diárias com erro mensal inferior ao das técnicas de preenchimento (esq.) e precipitação média histórica mensal (dir.) na reg. Centro-Oeste G1.

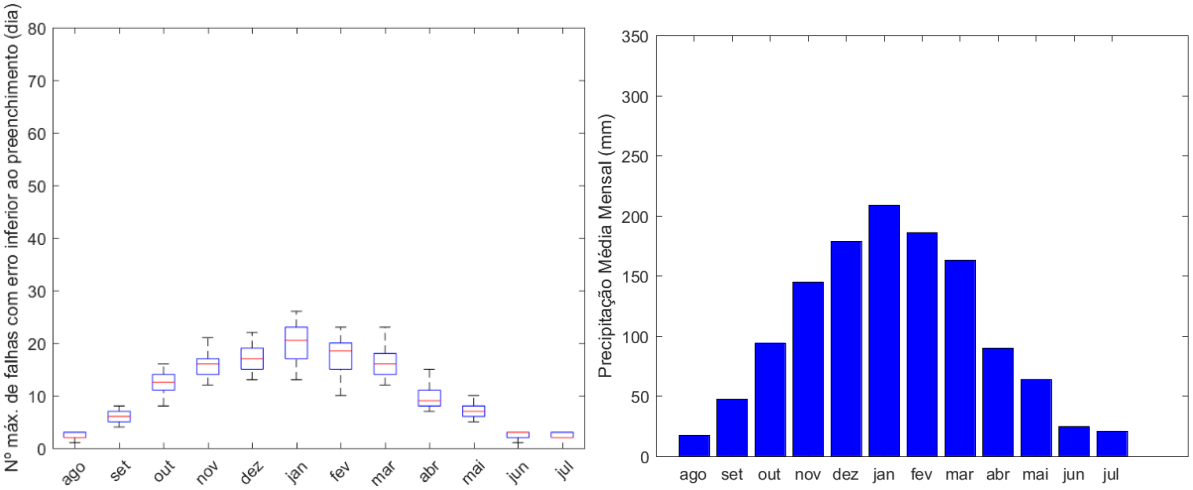


Figura 42. *Box-plot* do máximo de falhas diárias com erro mensal inferior ao das técnicas de preenchimento (esq.) e precipitação média histórica mensal (dir.) na reg. Centro-Oeste G2.

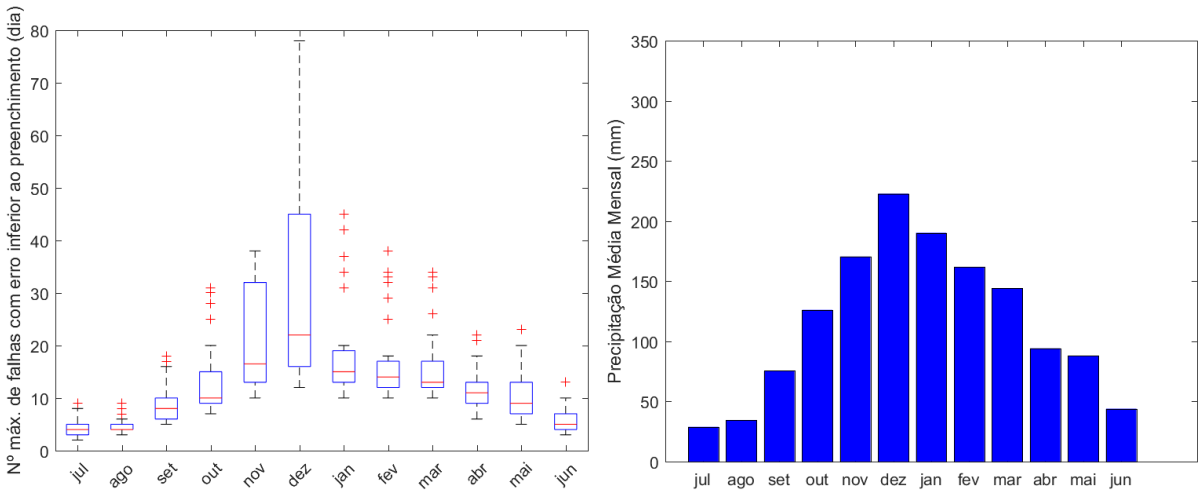


Figura 43. *Box-plot* do máximo de falhas diárias com erro mensal inferior ao das técnicas de preenchimento (esq.) e precipitação média histórica mensal (dir.) na região Nordeste G1.

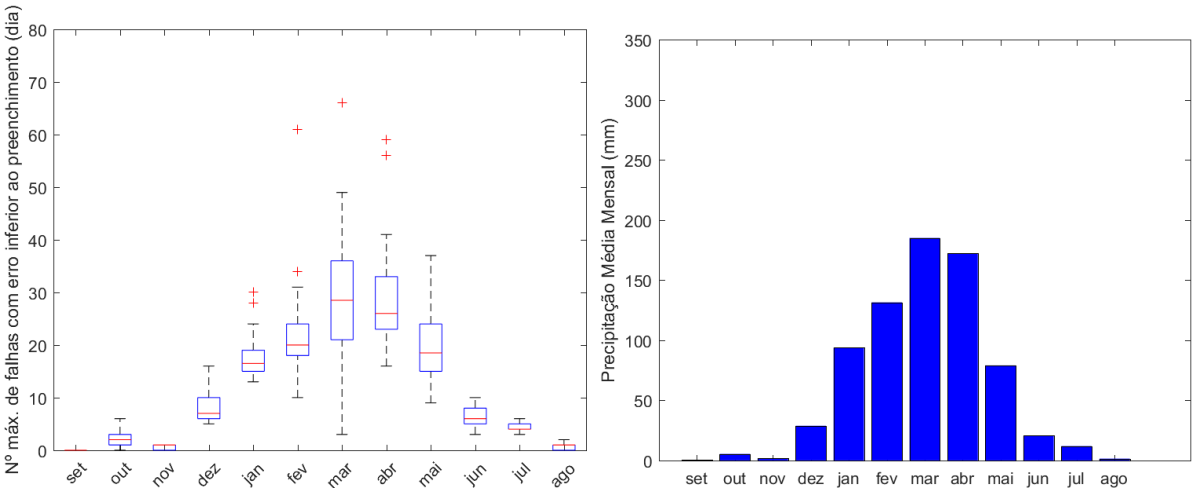


Figura 44. *Box-plot* do máximo de falhas diárias com erro mensal inferior ao das técnicas de preenchimento (esq.) e precipitação média histórica mensal (dir.) na região Nordeste G2.

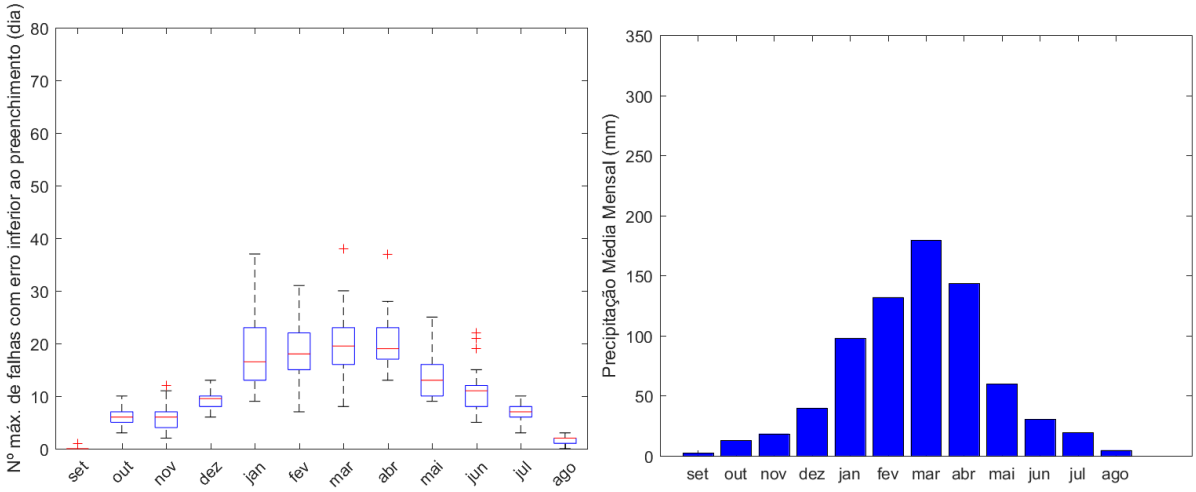


Figura 45. *Box-plot* do máximo de falhas diárias com erro mensal inferior ao das técnicas de preenchimento (esq.) e precipitação média histórica mensal (dir.) na região RJ-ES G1.

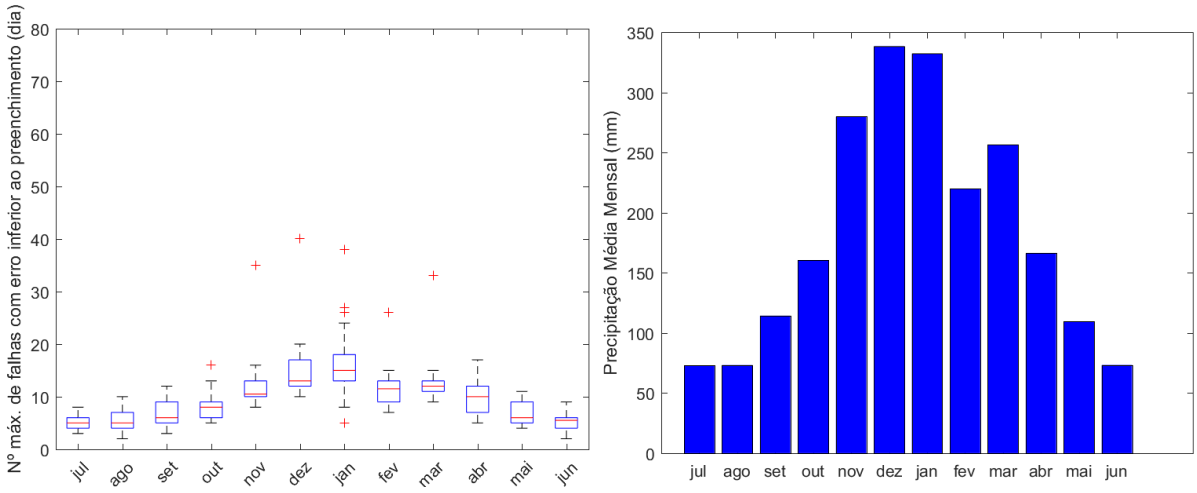


Figura 46. *Box-plot* do máximo de falhas diárias com erro mensal inferior ao das técnicas de preenchimento (esq.) e precipitação média histórica mensal (dir.) na região RJ-ES G2.

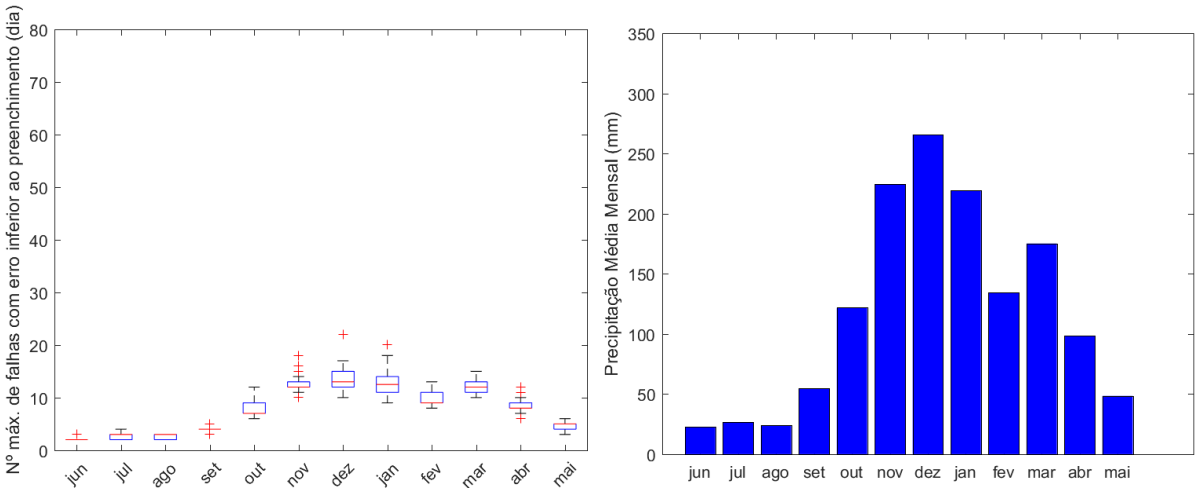


Figura 47. *Box-plot* do máximo de falhas diárias com erro mensal inferior ao das técnicas de preenchimento (esq.) e precipitação média histórica mensal (dir.) na região SP-MG G1.

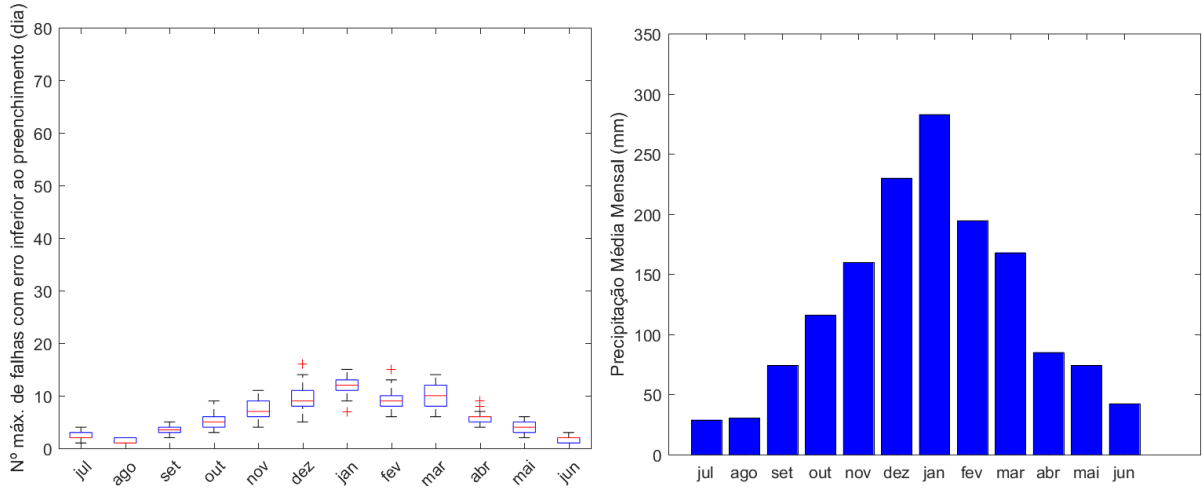


Figura 48. *Box-plot* do máximo de falhas diárias com erro mensal inferior ao das técnicas de preenchimento (esq.) e precipitação média histórica mensal (dir.) na região SP-MG G2.

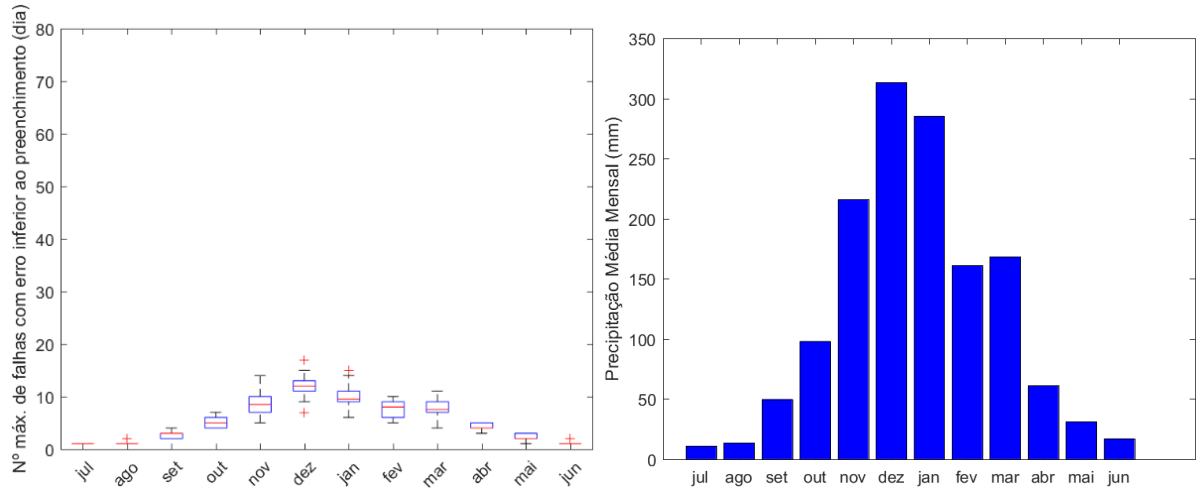


Figura 49. *Box-plot* do máximo de falhas diárias com erro mensal inferior ao das técnicas de preenchimento (esq.) e precipitação média histórica mensal (dir.) na região Sul G1.

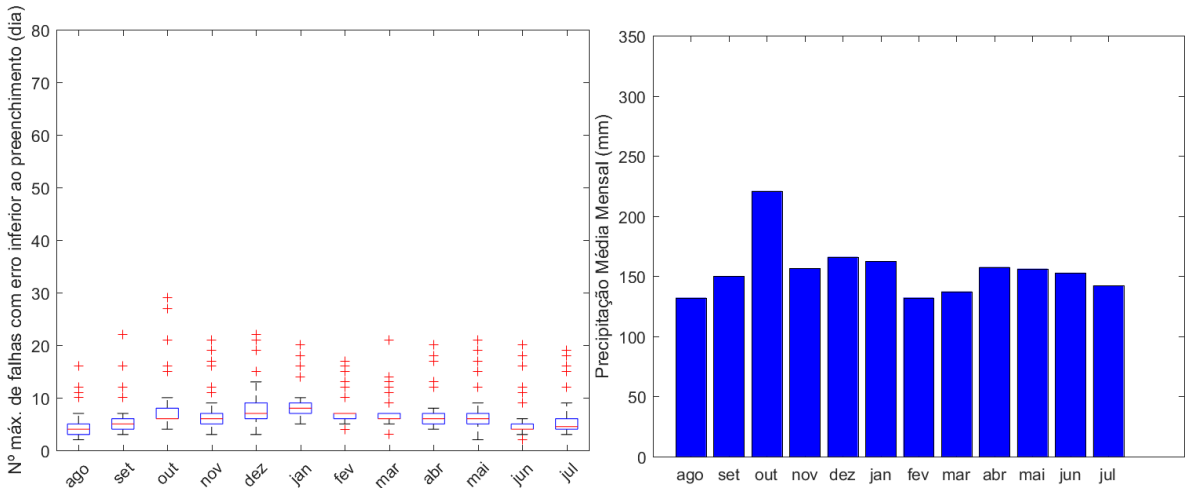
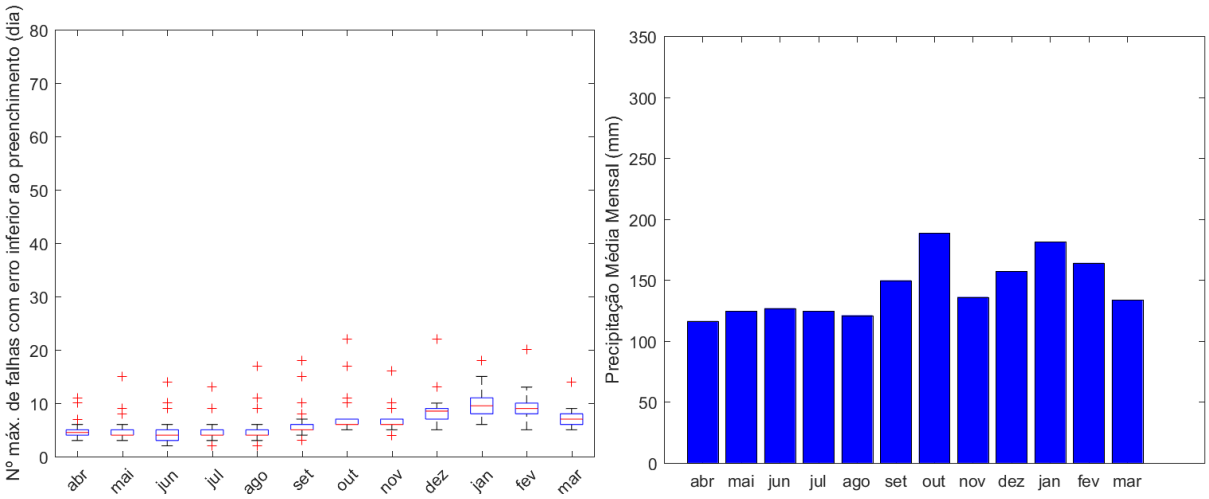


Figura 50.*Box-plot* do máximo de falhas diárias com erro mensal inferior ao das técnicas de preenchimento (esq.) e precipitação média histórica mensal (dir.) na região Sul G2.



Observando-se os gráficos acima expostos, é possível perceber uma tendência de queda do limite de falhas diárias nos meses menos pluviosos de determinadas regiões, como é o caso do inverno seco da região sudeste que, por ter menores precipitações e de forma homogênea, o preenchimento tende a ser mais assertivo.

Já na região Sul G1 (Rio Grande do Sul) o limite tende a ser menor e mais homogêneo. Isso se deve ao fato de a região mais homogênea permitir que se possa extrapolar dados de estações vizinhas à estação analisada sem grandes erros, o que diminui o erro médio das técnicas de preenchimento. Assim como, regiões com menos dias de chuva igual a 0 mm favorece menos o tipo de preenchimento por 0 mm.

Observou-se padrões semelhantes entre a das regiões do Brasil quando delineada uma curva entre as medianas dos diferentes meses do ano, sendo algumas mais lineares, como é o caso da região Norte G2 (Amazonas) e Sul G1 (Rio Grande do Sul) e ocorrendo o oposto na região Norte G1 (Roraima) em que os maiores limites de falhas “aceitáveis” estavam no inverno (meses de Abril à Agosto).

Roraima (Norte G1) apresentou os maiores limites de falhas diárias no meio do ano, exatamente nos meses mais chuvosos. Tal constatação, mais uma vez, leva a crer que há uma relação direta entre o total precipitado no mês e o erro absoluto das técnicas de preenchimento. Uma vez elevado o erro dos preenchimentos, eleva-se também as falhas diárias aceitáveis.

Cabe destacar, novamente, que são análises realizadas de forma pontual (em apenas 12 estações). No entanto, buscou-se relacionar tais análises ao clima que tais estações estão sujeitas, para se poder entender o comportamento das técnicas de preenchimento e correlacionar as características pluviométricas das regiões. A partir da validação da metodologia proposta, será possível então, a aplicação em um maior número de estações, podendo-se realizar análises e proposições regionais.

5.4 ANÁLISE GERAL DOS RESULTADOS

Expôs-se, até o momento, avaliações pontuais dos dados gerados a partir da análise estatística de técnicas de preenchimento tradicionais e comparando-as com os totais diários precipitados que eventualmente podem não ser registrados, gerando as chamadas falhas no registro dos dados. Esta técnica de comparação teve como objetivo avaliar o grau de magnitude entre os erros entre as falhas de registro de dados e as falhas geradas por preenchimentos pouco assertivos.

Embora a análise da metodologia de comparação tenha sido efetuada com base em apenas 12 estações (escolhidas em um grupo de 91 estações), os resultados foram considerados satisfatórios e parcialmente conclusivos.

Os destaques da análise efetuada, que pode ser ampliada para mais estações, são os seguintes:

1. Quanto maior o total mensal precipitado médio em uma estação, maiores são os erros médios de preenchimento.

2. Quanto mais heterogêneas são as precipitações de uma época do ano em uma determinada localidade, maiores são os erros observados nas técnicas de preenchimento (em acordo com Einfalt, Johann e Pfister, 1998).
3. Quanto maiores as precipitações médias diárias, maiores são os erros médios causados por falhas diárias no total mensal precipitado, ou maior é o erro do preenchimento assumindo-se 0 mm de precipitação diária.
4. Quanto maior o percentual de dias chuvosos em um determinado local, maiores são os erros médios das técnicas de preenchimento, dada a maior variabilidade espacial de fenômenos maiores.
5. Quanto maiores são os erros das técnicas de preenchimento, maiores são os limites aceitáveis de falhas diárias nos totais mensais precipitados.
6. Quanto menor o percentual de dias chuvosos em um determinado local em uma determinada época do ano, maiores são os limites aceitáveis de falhas diárias nos totais mensais precipitados.

Tal metodologia, além de avaliar o limite de falhas diárias aceitáveis também pode ser utilizada para a avaliação da efetividade de preenchimento de estações sob análise, utilizando-se de séries sintéticas de preenchimento e comparando-as com a série original. A comparação do erro das técnicas pode indicar a técnica mais indicada para cada caso.

5.5 RECOMENDAÇÕES PARA APLICAÇÃO PRÁTICA DA METODOLOGIA

A seguir apresenta-se um roteiro para a tomada de decisão entre a utilização do não preenchimento de falhas (ou preenchimento com valores 0 mm) e como realizá-lo.

1. Escolha de um número suficiente de estações de apoio no entorno da estação analisada, conforme metodologia tradicional de preenchimento.
2. Aplicação de diferentes técnicas de preenchimento a nível mensal, preferencialmente com o auxílio de ferramentas computacionais para acelerar o tempo de processamento gerando-se séries sintéticas para todo o histórico da estação.

3. Cálculo do erro mediano dentre as diferentes técnicas de preenchimento e avaliação da melhor técnica de preenchimento para cada mês do ano (janeiro a dezembro).
4. Cálculo da precipitação média diária histórica de cada mês do ano da estação escolhida.
5. Divisão entre o erro mediano de cada mês do ano da série sintética gerada pela técnica de preenchimento que mais se assemelha à série original e a precipitação média diária histórica de cada mês.
6. Para avaliações mais criteriosas, podem ser utilizados o 5º e o 95º percentis do histograma dos erros da melhor série de preenchimento e o 5º e o 95º percentis da precipitação média diária de cada mês. Dessa forma, pode-se entender melhor o comportamento da distribuição (dispersão) dos erros da técnica de preenchimento ou de eventuais falhas na série.

6 CONCLUSÃO E RECOMENDAÇÕES

Com vistas a atingir o objetivo deste trabalho que era o de avaliar o efeito que falhas aleatórias com extensão variada têm sobre os totais mensais, propôs-se uma metodologia para avaliação de tal efeito, assim como, uma metodologia para escolha entre o preenchimento convencional e o não preenchimento (ou preenchimento com 0 mm), baseando-se em critérios como os meses do ano e as diferentes regiões climáticas.

A motivação deste estudo foi o de questionar o uso de técnicas de preenchimento de forma indiscriminada, na hipótese de que podem ser obtidos resultados mais condizentes com a realidade caso se mantenha o registro original com falhas, ao invés de se substituir os valores originais.

Para isso, propôs a criação de séries sintéticas a partir de técnicas de preenchimento convencionalmente utilizadas, comparando-as à série original e avaliando a distribuição dos erros. Tais erros foram comparados aos erros causados por falhas sintéticas criadas nas séries originais, sendo proposto como critério entre o preenchimento ou não (preenchimento trivial com 0 mm) a razão entre tais erros.

Tal metodologia foi aplicada a 12 estações localizadas em diversas regiões do Brasil, tendo sido utilizadas estações de apoio para a aplicação das técnicas de preenchimento. Cada grupo contou com pelo menos seis estações, e cada mês analisado contou com pelo menos quatro estações (sendo uma delas a principal).

A partir das análises realizadas, pode-se observar um indício positivo para a aceitabilidade de falhas diárias sem maiores prejuízos para os totais mensais precipitados que as técnicas de preenchimento tradicionalmente utilizadas. Tal aceitabilidade varia de acordo com as regiões em que estão localizadas as estações, assim como, as épocas do ano em que ocorre a falha ou o eventual preenchimento.

Observou-se determinados padrões entre regiões próximas como, por exemplo, a menor aceitabilidade de falhas em épocas com maior assertividade no preenchimento como é o caso do inverno no Sudeste que tem menores precipitações. Em contraponto, as épocas mais chuvosas como são os meses de março a maio no Nordeste e no verão do Centro-Oeste apresentaram maiores

limites de falhas aceitáveis, uma vez que as técnicas de preenchimento tendem a ser menos assertivas.

Recomenda-se que tal metodologia seja aplicada a um maior número de estações de forma sistemática, para que se possa validar os resultados preliminares aqui expostos e se possa mapear tais padrões. Dessa forma, entende-se que se poderá propor critérios regionais sazonais para a tomada de decisão quanto ao preenchimento de séries de dados mensais de precipitação.

Cabe ressaltar que as comparações realizadas foram na hipótese de que o preenchimento haveria de ser realizado, quando em outros casos, por exemplo do cálculo da média mensal histórica, o mês em que houve a falha pode ser excluído do histórico. Dessa forma o prejuízo, mesmo que menor do que um eventual preenchimento, não venha a influenciar a média mensal histórica. Cabe a futuros estudos, avaliar os impactos que falhas diárias em determinados meses de um histórico causam à média mensal histórica e qual a aceitabilidade destas falhas.

7 REFERÊNCIAS

AGÊNCIA NACIONAL DE ÁGUAS. **HidroWeb**: sistemas de informações hidrológicas. Disponível em: <http://hidroweb.ana.gov.br/HidroWeb>. Acesso em: 31 maio 2019.

AGÊNCIA NACIONAL DE ÁGUAS. **Orientações para a Operação das Estações Hidrométricas**. Brasília: Superintendência de Gestão da Rede Hidrometeorológica, 2012. Disponível em: <http://arquivos.ana.gov.br/infohidrologicas/cadastro/OrientacoesParaOperacaoDeEstacoesHidrometricas-VersaoJun12.pdf>. Acesso em: 10 jun. 2019.

ALVARES, C. A.; STAPE, J. L. SENTELHAS, P. C.; GONÇALVES, J. L. DE M.; SPAROVEK G. **Köppen's climate classification map for Brazil**. Gebrüder Borntraeger, Meteorologische Zeitschrift, 2014. Vol. 22, pp. 711–728.

ALVES, A.; KOBIYAMA, M.; SILVA, R. V. DA; CHECCHIA, T. **Análise de dados hidrológicos na região do município de Alfredo Wagner/SC**. Simpósio de Recursos Hídricos do Sul. p. Anais... Porto Alegre: ABRH, 2006.

BERTONI, J. C.; TUCCI, C. E. M. Precipitação. In: TUCCI, C. E. M. (Org.). **Hidrologia**: Ciência e Aplicação. 4. ed. Porto Alegre, UFRGS/ABRH, 2007. p. 177-241.

BIER, A. A.; FERRAZ, S. E. T. **Comparação de Metodologias de Preenchimento de Falhas em Dados Meteorológicos para Estações no Sul do Brasil**. Revista Brasileira de Meteorologia, 2017. Vol. 32, pp. 215-226.

CALVETTI, L.; BENETI, C.; PEREIRA FILHO, A. J. **Integração do radar meteorológico doppler do SIMEPAR e uma rede de pluviômetros para a estimativa da precipitação**. Simpósio Brasileiro de Sensoriamento Remoto. São José dos Campos: s.n., 2003.

CAMPOS, L.; LEITÃO, F. T.; SAMPAIO, A.; DE CARVALHO, T. S. **Análise Intercomparativa De Métodos De Preenchimento De Falhas De Precipitações Anuais Para O Estado Do Ceará**. 2016.

DÉQUÉ, M. Continuous variables. JOLLIFE, I. T.; STEPHENSON, D. B. (Org.). **Forecast Verification**: A Practitioner's Guide in Atmospheric Science. s.l.: Willey, 2003. pp. 97-119.

EINFALT, T.; JOHANN, G.; PFISTER, A. **On the spatial validity of heavy point rainfall measurements**. Water Science and Technology. Volume 37, Issue 11, 1998, Pages 21-28.

FERNANDEZ, M. N. **Preenchimento de Falhas em Séries Temporais**. Rio Grande: FURG, 2007.

GILL, M. K.; ASEFA, T.; KAHEIL Y.; MCKEE, M. **Effect of missing data on performance of learning algorithms for hydrologic predictions**: Implications to an imputation technique. Water Resources Research, 2007.

GOVERNO DO ESTADO DE MINAS GERAIS. **Diretrizes e Análises Recomendadas para a Consistência de Dados Fluviométricos e Pluviométricos**. Belo Horizonte: Instituto Mineiro de Gestão das Águas, 2015.

HIEZ, G. **L'homogenéité des données pluviométriques. Cahiers ORSTOM**. Série Hydrologie. 1977. Vol. 14, 2, pp. 129-172.

HIEZ, G.; RANCAN, L. **Aplicação do Método Vetor Regional no Brasil**. V Simpósio Brasileiro de Recursos Hídricos. Blumenau: ABRH, 1983. pp. 242-266.

HUBBARD, K. G. **Spatial variability of daily weather variables in the high plains of the USA**. Agricultural and Forest Meteorology, 1994. Vol. 68, pp. 29-41.

JOINT COMMITTEE FOR GUIDES IN METROLOGY. **Evaluation of measurement data - Guide to the expression of uncertainty in measurement**. s.l: JCGM, 2010.

MAMÉDIO, F. M. A. **Impacto de erros nos dados de entrada na eficiência de um modelo hidrológico**. Porto Alegre: UFRGS, 2014.

MINISTÉRIO DAS MINAS E ENERGIA. **Sistemática para Análise de Consistência e Homogeneização de Dados Pluviométricos**. Brasília: Departamento Nacional de Águas e Energia Elétrica, 1984.

NAGHETTINI, M.; PINTO, E. J. DE A. **Hidrologia Estatística**. Belo Horizonte: CPRM, 2007.

OLIVEIRA, L. F. C.; FIOREZE, A. P.; MEDEIROS, A. M. M.; SILVA, M. A. S. **Comparação de metodologias de preenchimento de falhas de séries históricas de precipitação pluvial anual**. Revista Brasileira de Engenharia Agrícola e Ambiental, 2010. Vol. 14, pp. 1186-1192.

PAULHUS, J. L. H.; KOHLER, M. A. **Interpolation of missing precipitation records**. Monthly Weather Review, 1952. Vol. 80, pp. 129-133.

PEREIRA, D. R.; ULIANA, E. M.; MARTINEZ, M. A.; SILVA, D. D. **Desempenho de um modelo hidrológico concentrado e de um semidistribuído na predição de vazões diárias**. Irriga, 2016. Vol. 21, pp. 398-413.

PINTO, E. J. de A.; AZAMBUJA, A. M. S. de; FARIAS, J. A. M.; SALGUEIRO, J. P.de B.; PICKBRENNER, K. (Coords.). **Atlas pluviométrico do Brasil**: isoietas mensais, isoietas trimestrais, isoietas anuais, meses mais secos, meses mais chuvosos, trimestres mais secos, trimestres mais chuvosos. Brasília: CPRM, 2011. 1 DVD. Escala 1.5:000.000. Equipe Executora: Da Costa, Margarida Regueira; Dantas, Carlos Eduardo de Oliveira; Melo, De Azambuja, Andressa Macêdo Silva; De Rezende, Denise C.; Do Nascimento, Jean Ricardo da Silva; Dos Santos, André Luis M. Real; Farias, José Alexandre Moreira; Machado, Érica C.; Marcuzzo, Francisco Fernando Noronha; Medeiros, Vanesca Sartorelli; Rodrigues, Paulo de Tarso R.; Weschenfelder, Adriana Burin; SIG - versão 2.0 - atualizada em 11/2011; Levantamento da Geodiversidade. Disponível em:<http://www.cprm.gov.br/publique/Hidrologia/Mapas-e-Publicacoes/Atlas-Pluviometrico-do-Brasil-1351.html>. Acesso em: 20 jan. 2019.

SANTOS, I. DOS; FILL, H. D.; SUGAI, M. R. v. B.; BUBA, H.; KISHI, R. T. MARONE, E.; LAUTERT, L. F. **Hidrometria Aplicada**. Curitiba: CEHPAR, 2001.

SMITH, J. A. Precipitation. In: MAIDMENT D. R. (Org.). **Handbook of Hydrology**. Michigan: McGrall-Hill, 1992. p. 1424.

SODRÉ, G. R.; BORGES, L.; VENTURA, J.; MELO, F. DE A. C.. **Análise quantitativa e Qualitativa Horária de Dados do Radar, TRMM e Estações Meteorológicas na Região Amazônica**. Revista Brasileira de Geografia Física, 2012. Vol. 6, pp. 110-117.

TABONY, R. C. **The estimation of missing climatological data. Journal of Climatology**, 1983. Vol. 3, pp. 297-314.

WEBER, E.; HASENACK, H.; FERREIRA, C.J.S. **Adaptação do modelo digital de elevação do SRTM para o sistema de referência oficial brasileiro e recorte por unidade da federação**. Porto Alegre: UFRGS Centro de Ecologia, 2004. ISBN 978-85-63843-02-9. Disponível em: <http://www.ecologia.ufrgs.br/labgeo>. Acesso em 10 out. 2018.

WORLD METEOROLOGICAL ORGANIZATION. **The Role of Climatological Normals in a Changing Climate**. Geneva: World Climate Data and Monitoring Programme, 2007.

YOUNG, K. C. **A three-way model for interpolating for monthly precipitation values**. Monthly Weather Review, 1992. Vol. 120, pp. 2561-2569.