



1920 | 2020

**Universidade  
Federal do  
Rio de Janeiro**

## **Análise de precipitação pluviométrica no município de Niterói via modelos dinâmicos bayesianos**

Deborah Cholodovsky Barbedo Pereira

**Orientadora:** Prof<sup>a</sup>.Dr<sup>a</sup>. Kelly Cristina Mota Gonçalves

Rio de Janeiro - RJ, Brasil

Junho de 2022



Resumo de monografia: A análise de precipitações pluviiais ao longo do tempo pode ser relevante tanto para uma simples conferência do esperado para um determinado período de tempo, como para o planejamento de atividades vulneráveis a sua ocorrência, podendo evitar danos e prejuízos consequentes de desastres naturais. Por se tratar de séries temporais, a metodologia de modelos dinâmico pode ser empregada na análise destes tipos de dados. Modelos lineares dinâmicos fazem parte de uma ampla classe de modelos com parâmetros que variam no tempo, úteis para modelagem e previsão de séries temporais. São representados por um par de equações que caracterizam a distribuição das observações e a evolução dos parâmetros ao longo do tempo, de forma probabilística. A inferência feita sob o paradigma bayesiano permite incorporação de conhecimento a priori e realização de previsão considerando a incerteza presente no passo estimação. Neste trabalho há interesse em aplicar modelos lineares dinâmicos a séries temporais de precipitação para o município de Niterói, usando diferentes formas de agregação no tempo.

Análise de precipitação pluviométrica no município de Niterói via modelos dinâmicos bayesianos

Deborah Cholodoysky Barbedo Pereira

Junho de 2022

Orientadora: Prof<sup>ª</sup>.Dr<sup>ª</sup>. Kelly Cristina Mota Gonçalves

Curso: Pós-graduação – Especialização em Ciência de Dados

Palavras-chave: Modelos lineares dinâmicos, Séries temporais, Precipitação pluviométrica

# Sumário

## Lista de Figuras

## Lista de Tabelas

<b>1</b>	<b>Introdução</b>	p. 9
1.1	Estrutura . . . . .	p. 10
<b>2</b>	<b>Metodologia</b>	p. 11
2.1	Modelo polinomial de primeira ordem . . . . .	p. 12
2.2	Modelo polinomial de segunda ordem . . . . .	p. 12
2.3	Filtro de Kalman . . . . .	p. 13
2.4	Suavizador de Kalman . . . . .	p. 14
<b>3</b>	<b>Aplicação</b>	p. 15
3.1	Análise descritiva . . . . .	p. 15
3.1.1	Acumulado . . . . .	p. 16
3.1.2	Máximo . . . . .	p. 18
3.1.3	Média diária de chuva . . . . .	p. 20
3.1.4	Logaritmo da média diária de chuva . . . . .	p. 23
3.1.5	Comparação geral das medidas propostas . . . . .	p. 25
3.2	Resultados . . . . .	p. 26
3.2.1	Acumulado . . . . .	p. 26
3.2.2	Máximo . . . . .	p. 31

3.2.3	Média dos dias de chuva . . . . .	p. 34
3.2.4	Logaritmo da média diária de chuva . . . . .	p. 38
<b>4</b>	<b>Conclusão</b>	p. 42
	<b>Referências</b>	p. 43

# Lista de Figuras

1	Gráfico do volume de chuva acumulado ao longo dos meses observados.	p. 16
2	Histograma do volume de chuva acumulado ao longo dos meses observados.	p. 17
3	Boxplot do volume de chuva acumulado ao longo dos meses observados.	p. 17
4	Gráfico do volume máximo de chuva ao longo dos meses observados. . .	p. 18
5	Histograma do volume máximo de chuva ao longo dos meses observados.	p. 19
6	Boxplot do volume máximo de chuva ao longo dos meses observados. .	p. 20
7	Gráfico do volume médio dos dias de chuva ao longo dos meses observados.	p. 21
8	Histograma do volume médio de chuva dos meses observados. . . . .	p. 22
9	Boxplot do volume médio de chuva ao longo dos meses observados. . . .	p. 22
10	Gráfico do logaritmo do volume médio dos dias de chuva ao longo dos meses observados. . . . .	p. 23
11	Histograma do volume logarítmico da média de chuva dos meses observados.	p. 24
12	Boxplot do volume logarítmico médio de chuva ao longo dos meses observados. . . . .	p. 25
13	Nível pelo filtro de Kalman para a série de acumulados de chuva dos meses observados. . . . .	p. 27
14	Tendência pelo filtro de Kalman a série de acumulados de chuva dos meses observados. . . . .	p. 28
15	Suavização para a série de acumulados de chuva dos meses observados.	p. 29
16	Previsão pontual e intervalar 6 meses a frente para a série de acumulados de chuva. . . . .	p. 30
17	Nível pelo filtro de Kalman para a série de máximo de chuva dos meses observados. . . . .	p. 31

18	Tendência pelo filtro de Kalman para a série de máximo de chuva dos meses observados. . . . .	p. 32
19	Suavização para a série de máximo de chuva dos meses observados. . .	p. 33
20	Previsão pontual e intervalar 6 meses a frente para a série de máximo de chuva. . . . .	p. 33
21	Nível pelo filtro de Kalman para a série de média de chuva dos meses observados. . . . .	p. 35
22	Tendência pelo filtro de Kalman para a série de média de chuva dos meses observados. . . . .	p. 36
23	Suavização para a série de média de chuva dos meses observados. . . .	p. 36
24	Previsão pontual e intervalar 6 meses a frente para a série de média de chuva. . . . .	p. 37
25	Nível pelo filtro de Kalman para a série de logaritmo da média de chuva dos meses observados. . . . .	p. 38
26	Tendência pelo filtro de Kalman para a série de logaritmo da média de chuva dos meses observados. . . . .	p. 39
27	Suavização para a série de logaritmo da média de chuva dos meses observados. . . . .	p. 40
28	Previsão pontual e intervalar 6 meses a frente para a série de logaritmo da média de chuva. . . . .	p. 40

# Lista de Tabelas

1	Tabela com os erros de previsão para a série com o acumulado de chuva no mês, para cada modelo ajustado. . . . .	p.30
2	Tabela com os erros de previsão para a série com o máximo de chuva no mês, para cada modelo ajustado. . . . .	p.34
3	Tabela com os erros de previsão para a série com a média de chuva no mês, para cada modelo ajustado. . . . .	p.37
4	Tabela com os erros de previsão para a série com o logaritmo da média de chuva no mês, para cada modelo ajustado. . . . .	p.41



# 1 Introdução

A análise das chuvas ao longo do tempo pode ser relevante tanto para uma simples conferência do que se espera para um determinado período, quanto para o planejamento de atividades vulneráveis à sua ocorrência, podendo evitar danos e prejuízos decorrentes de desastres naturais. Por serem séries temporais, a metodologia de modelos dinâmicos pode ser utilizada na análise destes tipos de dados. Modelos lineares dinâmicos fazem parte de uma ampla classe de modelos com parâmetros variantes no tempo, úteis para modelagem e previsão de séries temporais. Eles são representados por um par de equações que caracterizam a distribuição das observações e a evolução dos parâmetros ao longo do tempo de forma probabilística. A inferência feita sob o paradigma Bayesiano permite a incorporação de conhecimento a priori e a realização da previsão considerando a incerteza presente na etapa de estimação.

Alguns desafios são encontrados ao estudar os índices pluviométricos, incluindo o melhor período a ser utilizado para a previsão e como contornar a grande quantidade de valores zero em medições que ocorrem a cada 15 minutos. Em [1], para contornar o problema dos períodos secos, adotou-se a análise do acumulado em um período fixo de tempo, com isso foram utilizados modelos lineares dinâmicos para estimar o volume de chuva e períodos de seca. Já em [2], foi utilizada a precipitação máxima anual para a previsão de precipitações pluviais extremas.

Neste trabalho, há interesse em aplicar modelos lineares dinâmicos a séries temporais de precipitação para a cidade de Niterói, utilizando diferentes formas de agregação no tempo. Para a análise, utiliza-se o índice pluviométrico, que é a medida do volume de chuva que cai em determinada área em determinado período de tempo, a menor granularidade disponível é de 15 minutos, sua unidade de medida é milímetros. Portanto, é possível estimar a quantidade de litros por metro quadrado em um determinado espaço de tempo. Esses dados estão disponíveis no site do CEMADEN [3] e INMET [4].

## 1.1 Estrutura

O texto está dividido em introdução, metodologia, aplicação e conclusão. Na metodologia, são apresentados os modelos lineares dinâmicos, bem como a estrutura e ferramentas utilizadas na monografia.

A aplicação está dividida em duas partes, a análise descritiva, onde é exposto o comportamento dos dados e explicadas as transformações aplicadas, e os resultados, onde a metodologia é aplicada aos dados.

Na conclusão, são discutidos os pontos a serem desenvolvidos e a discussão dos resultados aqui obtidos.

## 2 Metodologia

Um modelo dinâmico é formado por uma série não observável de parâmetros que evoluem no tempo  $(\theta_t)_{t \geq 0}$  definida nos reais de dimensão  $p$  e uma série temporal em  $(Y_t)_{t \geq 1}$  definida nos reais com dimensão  $n$ . Onde os parâmetros sucessivos estão intimamente relacionados e mudam ao longo do tempo, em geral, um parâmetro é igual ao seu antecessor mais uma perturbação do sistema [5].

O modelo linear dinâmico é uma classe específica dos modelos dinâmicos, da qual é completamente especificada pela quádrupla  $\{F_t, G_t, V_t, W_t\}$ . As equações que especificam o MLD são as seguintes:

$$Y_t = F_t \theta_t + v_t, \text{ onde } v_t \sim N(0, V_t) \quad (2.1)$$

$$\theta_t = G_t \theta_{t-1} + w_t, \text{ onde } w_t \sim N(0, W_t) \quad (2.2)$$

Sob o paradigma bayesiano a incerteza inicial é incorporada ao modelo por meio de uma distribuição a priori especificada. No caso de um MLD essa incerteza prévia é caracterizada por  $\theta_0 \sim N(m_0, C_0)$  quando  $t = 0$  [6]. Os valores de  $m_0$  e  $C_0$  estão ligados ao conhecimento prévio do comportamento da série, ou seja, em um cenário de grande incerteza, pode-se supor que  $m_0$  é um vetor nulo e  $C_0$  possui elementos da diagonal com valores grandes e fora da diagonal apresenta o valor 0 ou alguma estrutura de covariância diferente.

A variável  $Y_t$  é um vetor  $n$ -dimensional, representando a observação no tempo  $t$ , enquanto  $\theta_t$  é um vetor  $p$ -dimensional não observável, chamado de parâmetro de estado, que representa o estado do sistema no tempo  $t$ . A matriz de covariância  $n \times n$   $V_t$  está associada ao erro observacional e a matriz de covariância  $p \times p$   $W_t$  está associado ao erro de evolução dos parâmetros de estado. As matrizes  $F_t$  e  $G_t$  possuem dimensões  $n \times p$  e  $p \times p$  respectivamente, sendo que  $F_t$  é uma matriz de regressão e  $G_t$  é uma matriz que

descreve a evolução temporal dos parâmetros de estado. Em geral ambas as matrizes  $F_t$  e  $G_t$  são conhecidas e elicítadas de acordo com o comportamento dinâmico das séries.

Normalmente, a variação das observações em torno dos níveis  $V_t$  é muito maior do que as variações temporais do nível ao longo do tempo  $W_t$ . De forma a facilitar, neste trabalho, será considerado que  $V_t = V$ ,  $W_t = W$ ,  $F_t = F$  e  $G_t = G$ ,  $\forall t > 0$ . Também iremos abordar apenas séries univariadas, ou seja,  $n = 1$ .

## 2.1 Modelo polinomial de primeira ordem

O modelo dinâmico mais simples, composto apenas por um nível que varia de acordo com um passeio aleatório mais ruído, é também chamado de modelo de nível local. É um MLD que assume que  $F_t = G_t = 1$  para  $t \geq 1$  nas Equações 2.1 e 2.2. De acordo com este modelo, o nível permanece localmente constante, mas varia por longos períodos de tempo. Os parâmetros do modelo dados pelas variâncias observacional e de evolução são estimadas a partir dos dados usando técnicas de máxima verossimilhança ou bayesianas.

$$Y_t = \theta_t + v_t, \text{ onde } v_t \sim N(0, V) \quad (2.3)$$

$$\theta_t = \theta_{t-1} + w_t, \text{ onde } w_t \sim N(0, W) \quad (2.4)$$

## 2.2 Modelo polinomial de segunda ordem

Este modelo permite o crescimento do nível, pois considera na estrutura dinâmica a inclinação do nível, que varia ao longo do tempo. Também chamado de modelo de tendência linear, é obtido com  $F_t = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$  e  $G_t = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$  e  $V_t = V$  e  $W_t = \begin{bmatrix} \sigma_{t1}^2 & \sigma_{t3}^2 \\ \sigma_{t3}^2 & \sigma_{t2}^2 \end{bmatrix}$ , estas duas últimas invariantes no tempo, já que estão sendo assumidos modelos constantes e com comportamento limitados []

$$Y_t = \theta_t + v_t, \text{ onde } v_t \sim N(0, V) \quad (2.5)$$

$$\theta_{1,t} = \theta_{1,t-1} + \theta_{2,t-1} + w_{1,t}, \text{ onde } w_{1,t} \sim N(0, \sigma_{t1}^2) \quad (2.6)$$

$$\theta_{2,t} = \theta_{2,t-1} + w_{2,t}, \text{ onde } w_{2,t} \sim N(0, \sigma_{t2}^2) \quad (2.7)$$

## 2.3 Filtro de Kalman

Utilizado para estimar os parâmetros de estados do MLD, onde as estimativas para o instante futuro  $t + 1$  são obtidas através da estimativa no instante anterior  $t$ , acrescida de um erro devido a alterações no sistema.

Nos modelos de espaço de estados, as informações futuras estão relacionadas às informações do passado. Definindo  $D_t$ , como o total de informações obtidas até o instante  $t$ , ou seja,  $D_t = \{D_0, y_1, \dots, y_t\}$ , temos que  $D_t$  fornece informações sobre  $\theta_t$ , que por sua vez fornece informações sobre o parâmetro de estado de evolução futuro até  $\theta_{t+k}$ , para qualquer  $k > 0$ . Portanto, é computacionalmente viável descrever recursivamente da seguinte forma:

1. Estima o parâmetro em um passo à frente para  $\theta_t$ , dado  $D_{t-1}$ , com base na densidade  $f(\theta_{t-1} | D_{t-1})$  e no modelo de transição.
2. Calcula a distribuição  $f(Y_t | D_t - 1)$  de um passo para o próximo estado
3. Calcula a distribuição  $f(\theta_t | D_t)$ , usando a regra de Bayes com base na distribuição a priori  $f(\theta_t | D_{t-1})$  e a verossimilhança  $f(y_t | \theta_t)$

O MLD é um caso que simplifica consideravelmente a recursão, pois os resultados da distribuição normal multivariada permitem provar que o vetor aleatório  $(\theta_0, \theta_1, \dots, Y_1, \dots, Y_t)$  é normalmente distribuído para  $t \geq 1$  [7].

Em suma, o filtro de Kalman torna possível calcular recursivamente as distribuições preditivas e filtradas, começando com  $\theta_0 \sim N(m_0, C_0)$ , e então calculando a distribuição de  $\theta_1 | y_1$  e assim por diante. Pois é possível chegar a priori do tempo  $t$  a partir da posteriori de  $t - 1$  pela estrutura markoviana do modelo dinâmico, e pelo método Bayesiano, com a priori de tempo  $t$ , é possível obter a distribuição a posteriori dos estados no tempo  $t$ . Onde a distribuição filtrada  $f(\theta_t | D_t)$  resume as informações contidas nas observações passadas  $D_t$ , que são suficientes para prever  $Y_{t+k}$ . As previsões no modelo dinâmico são obtidas combinando as informações a priori  $(\theta_t | D_{t-1})$  com a equação de observações  $Y_t = F_t \theta_t + v_t$ , apresentado em Equação 2.1.

## 2.4 Suavizador de Kalman

A suavização é um algoritmo recursivo que vem do Filtro de Kalman para calcular as densidades condicionais de  $\theta_t | D_T$ , para  $t \leq T$ , partindo da densidade de filtragem  $f(\theta_T | D_T)$  e estimando para trás todo o histórico de parâmetros de estado [7].

Esta operação de passagem de informação retrospectiva no tempo usa toda a série observada para reavaliar a inferência feita durante o procedimento sequencial [8], com intuito de reconstruir retrospectivamente o comportamento e estrutura da série.

## 3 Aplicação

Os dados disponíveis referem-se ao índice pluviométrico de medições a cada quinze minutos da cidade de Niterói, de janeiro de 2014 a abril de 2022. Devido a esse intervalo pequeno entre as observações, existem vários registros com o valor igual a zero. Para uma análise de séries temporais, alguns sumários mensais foram propostos com base em artigos. Em [9], a análise é realizada com os dados agrupados por mês, realizando a média mensal apenas dos dias em que ocorreram chuva, desta forma contorna o problema de ter muitos valores zerados. Já em [1] foram utilizados diferentes períodos de tempo, adotando-se a análise de valores acumulados. Para a análise da precipitação extrema, em [2] foi utilizada a precipitação máxima anual.

Neste trabalho, quatro agrupamentos mensais foram adotados, são eles:

- Volume pluviométrico acumulado, que possui o índice pluviométrico total no mês de referência.
- Máximo de precipitação, tomando como referência a quantidade de chuva diária.
- Média diária de chuva, com base apenas nos dias que choveram no mês de referência.
- Logaritmo da média diária de chuva.

### 3.1 Análise descritiva

Nesta seção, o objetivo é resumir e explorar o comportamento dos dados tratados, e assim entender como as transformações afetaram a distribuição, comportamento e ordem de grandeza dos dados observados.

### 3.1.1 Acumulado

O tratamento aqui realizado foi adotado para contornar o desafio de ter muitos zeros, que inclusive ocorrem em agrupamentos diários. Por outro lado, ao analisar o acumulado total de chuvas do mês de referência, perdemos informações sobre o comportamento das chuvas, por exemplo, não conseguimos definir se tivemos vários dias chuvosos ou alguns dias com chuvas torrenciais.

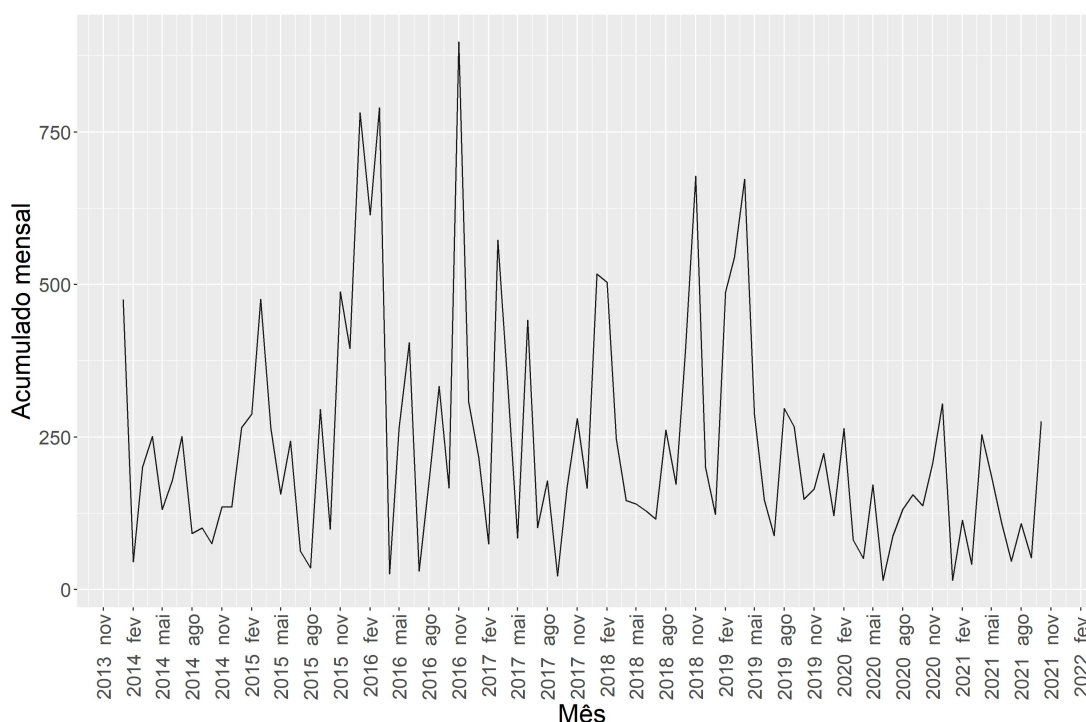


Figura 1: Gráfico do volume de chuva acumulado ao longo dos meses observados.

Na Figura 1 é possível observar um pico no mês de novembro de 2016, atingindo o valor de 898,03 mm acumulados no mês. O que se caracteriza como um valor muito alto. Analisando as informações deste mês e ano específico, as notícias relatam fortes chuvas e inundações, incluindo distúrbios nos hospitais [10], [11].

A Figura 2 mostra um histograma da distribuição do volume acumulado mensal ao longo dos meses observados. O volume acumulado apresenta uma distribuição assimétrica, o que pode ser esperado pois valores muito altos tornam-se menos comuns.



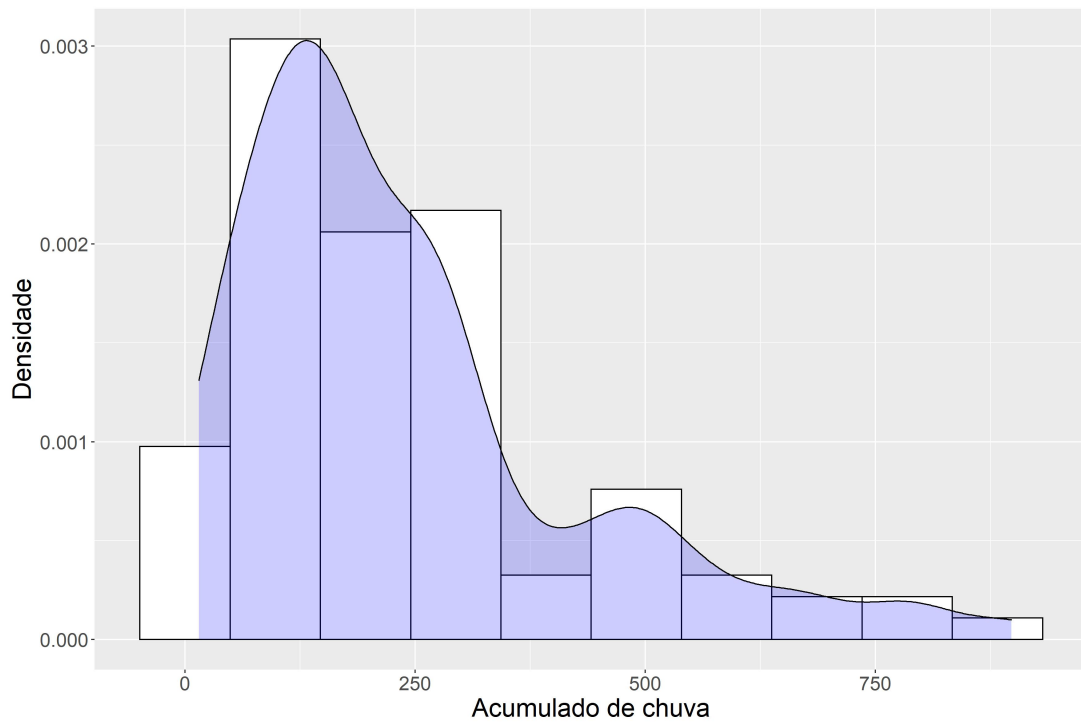


Figura 2: Histograma do volume de chuva acumulado ao longo dos meses observados.

A Figura 3 apresenta um boxplot para melhor visualização e interpretação da distribuição do volume de chuva acumulado.

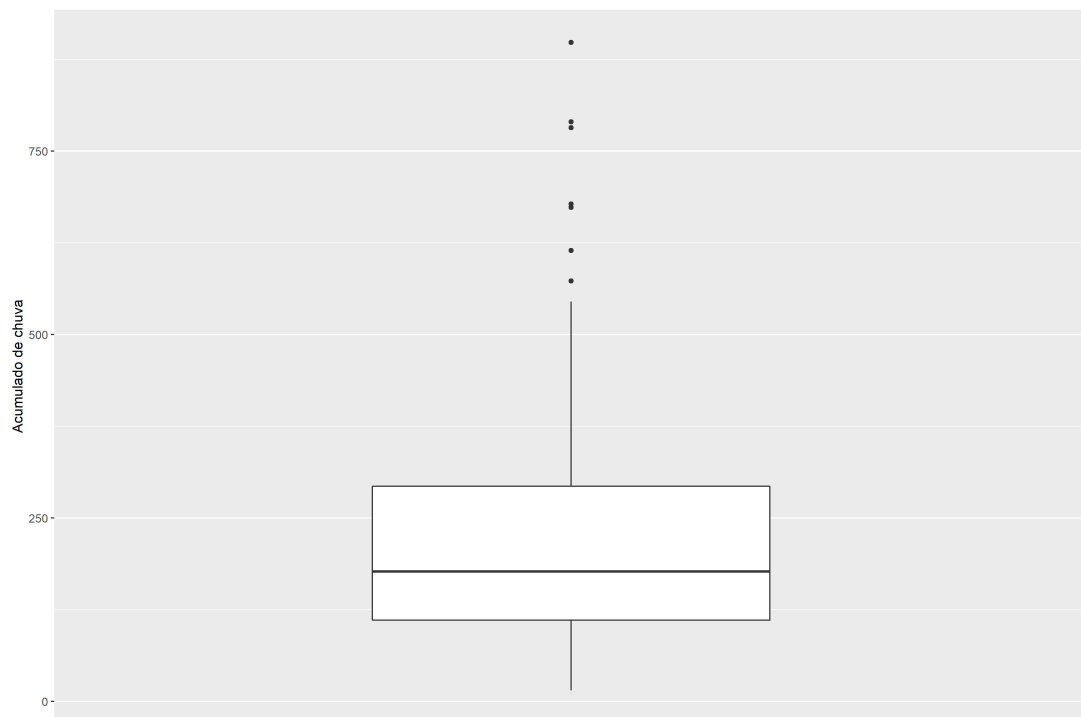


Figura 3: Boxplot do volume de chuva acumulado ao longo dos meses observados.

O boxplot (Figura 3) confirma a assimetria presente nos dados e presença de outliers na cauda à direita (Figura 2) .

### 3.1.2 Máximo

Outro tratamento para contornar a grande quantidade de zeros diários, é a análise de precipitação máxima, como realizada em [2]. Uma análise envolvendo a precipitação máxima diária dentro de um mês de referência permite classificar se o volume previsto é violento, forte, moderado ou fraco e se há necessidade do uso de mecanismos de alerta à população de antemão.

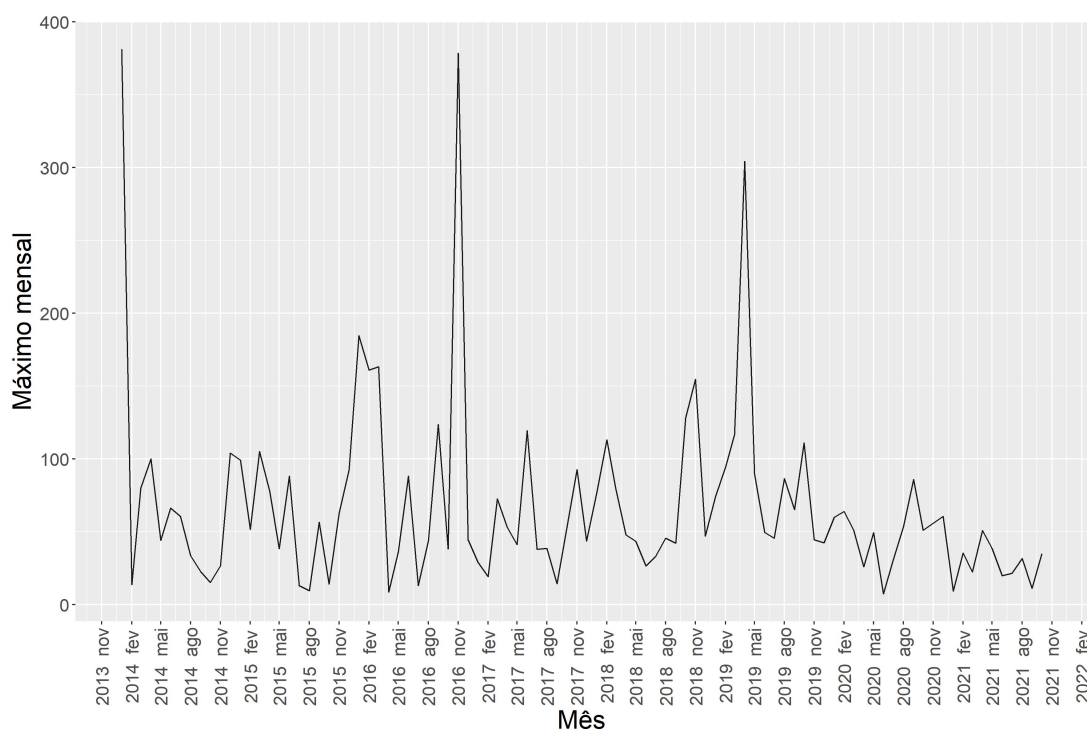


Figura 4: Gráfico do volume máximo de chuva ao longo dos meses observados.

A Figura 4 apresenta a série temporal com o máximo volume de chuva diário por mês. Ao observá-la é possível verificar dois importantes picos. Um ocorreu no primeiro mês de análise, janeiro de 2014, com 381,31 mm, ou seja, aproximadamente 15,89 por hora se tivesse chovido todas as vinte e quatro horas do dia. E outro pico em novembro de 2016, com 378,48 mm no dia, aproximadamente 15,77 mm por hora. Segundo [12] a chuva é classificada como forte quando sua intensidade é igual ou superior a 10 mm por hora, e violenta quando maior que 50 mm por hora.

O terceiro menor pico que ocorre em março de 2019, foi relatado estado de alerta em

Niteró [13].

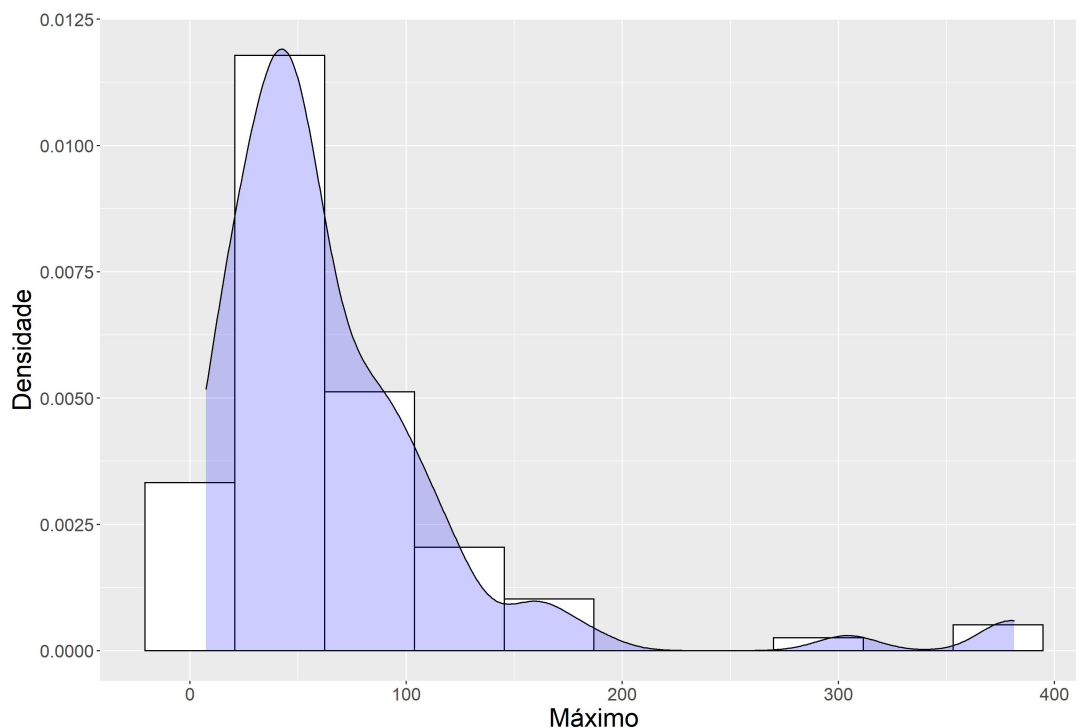


Figura 5: Histograma do volume máximo de chuva ao longo dos meses observados.

O histograma da Figura 5 mostra a distribuição do volume máximo diário de chuva ao longo dos meses observados. Com esta transformação, a série tem uma distribuição assimétrica. Novamente o esperado, assim como com a acumulada, os valores extremos tendem a ser raros, e o volume de chuvas é maior ou igual a zero.

Adicionalmente, o boxplot da Figura 6 é capaz de representar algumas estatísticas do volume máximo para os meses observados.



Figura 6: Boxplot do volume máximo de chuva ao longo dos meses observados.

A precipitação máxima diária, do mês de estudo, está muito longe do terceiro quartil, é possível constatar o quão atípico é este evento máximo.

### 3.1.3 Média diária de chuva

Foi adotada a média mensal apenas dos dias que choveram devido ao impacto dos dias sem chuva. O artigo [9] mostra bem esses impactos na média, e como obtemos mais informações adotando apenas os dias em que choveu para a média.

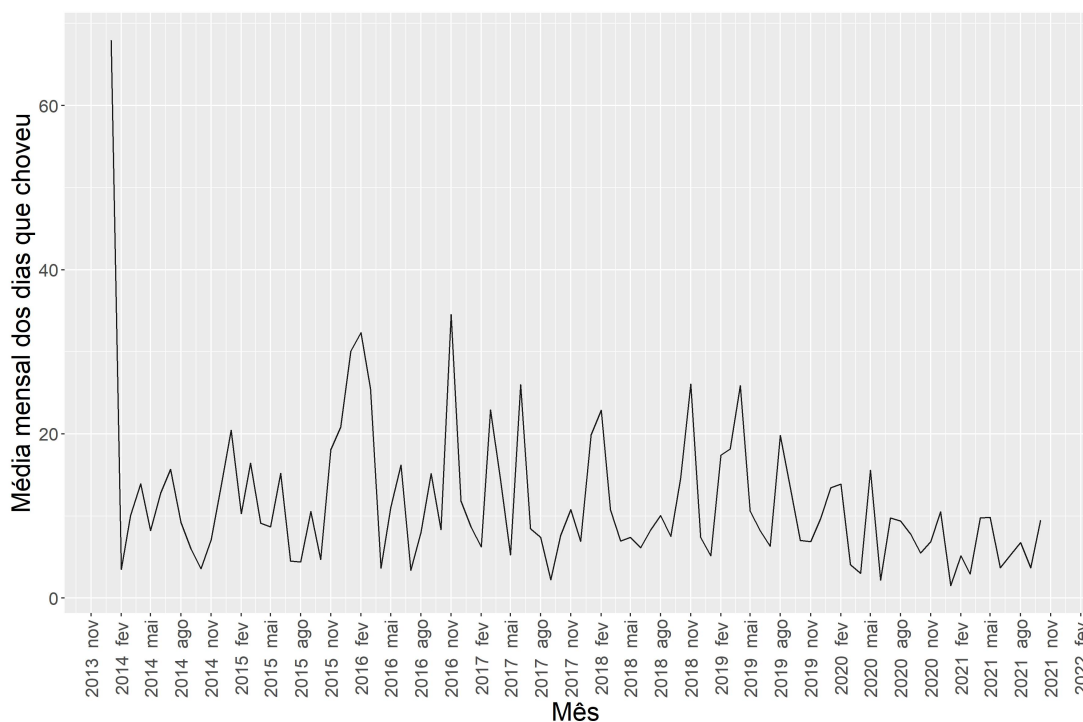


Figura 7: Gráfico do volume médio dos dias de chuva ao longo dos meses observados.

Na Figura 7 está representada a série temporal com a transformação da média dos dias em que ocorreu chuva no mês de referência. O pico, muito maior do que nos meses seguintes, ocorre com um evento atípico, que teria ocorrido após muitos dias sem chuva no estado do Rio de Janeiro [14]. Desta forma, é possível verificar a perda de informações importantes pelo uso da média, pois não é possível saber se foram poucos dias com chuvas torrenciais ou muitos dias com comportamento similar de chuva.

Na Figura 8 está representado o histograma do volume de chuva médio dos dias em que houve de chuva no mês de referência. A distribuição é assimétrica, porém menos que nos tratamentos feitos anteriormente, tendo apenas um valor muito alto na cauda direita da distribuição.

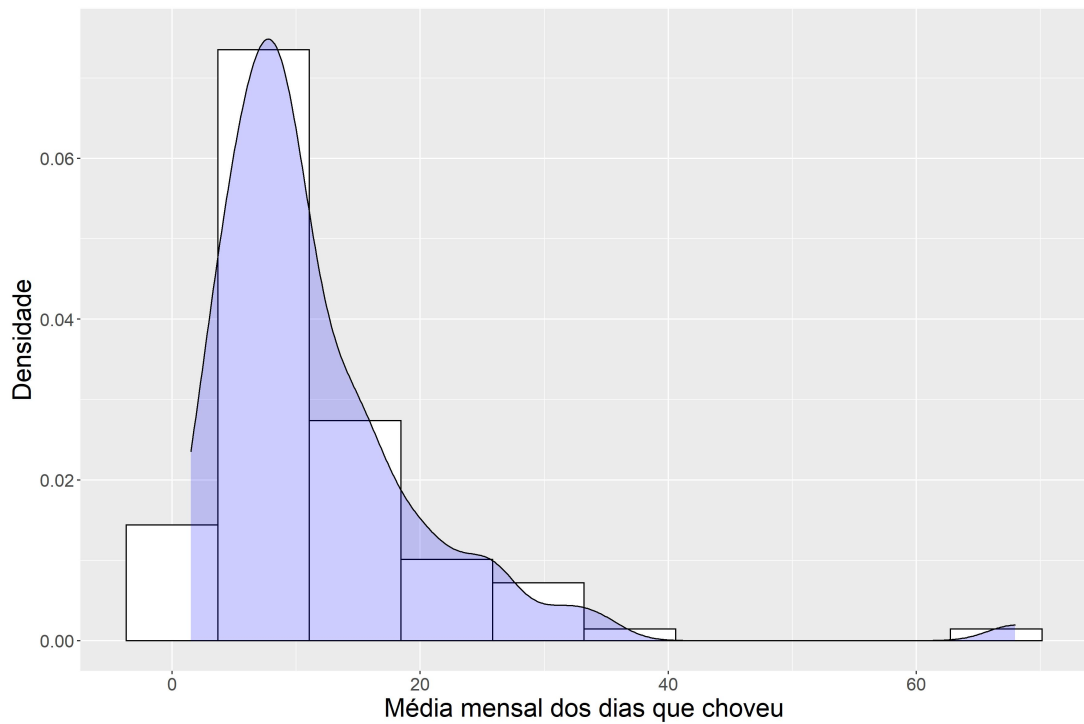


Figura 8: Histograma do volume médio de chuva dos meses observados.

O boxplot na Figura 9, por sua vez, mostra que outros três meses também apresentaram um volume médio de chuva considerado atípico.



Figura 9: Boxplot do volume médio de chuva ao longo dos meses observados.

Na Figura 9, é possível verificar que o valor máximo está muito distante do terceiro quartil e até mesmo dos demais outliers, este evento de janeiro de 2014 também é visivelmente destacado no gráfico da Figura 7 .

### 3.1.4 Logaritmo da média diária de chuva

A intenção de aplicar o logaritmo na média dos dias em que ocorreu chuva, é fazer com que a distribuição dos dados se torne mais simétrica. A média utilizada aqui é a mesma do Capítulo 3.1.3 anterior , que foi tratada para que seja estudado apenas os dias em que houve chuva.

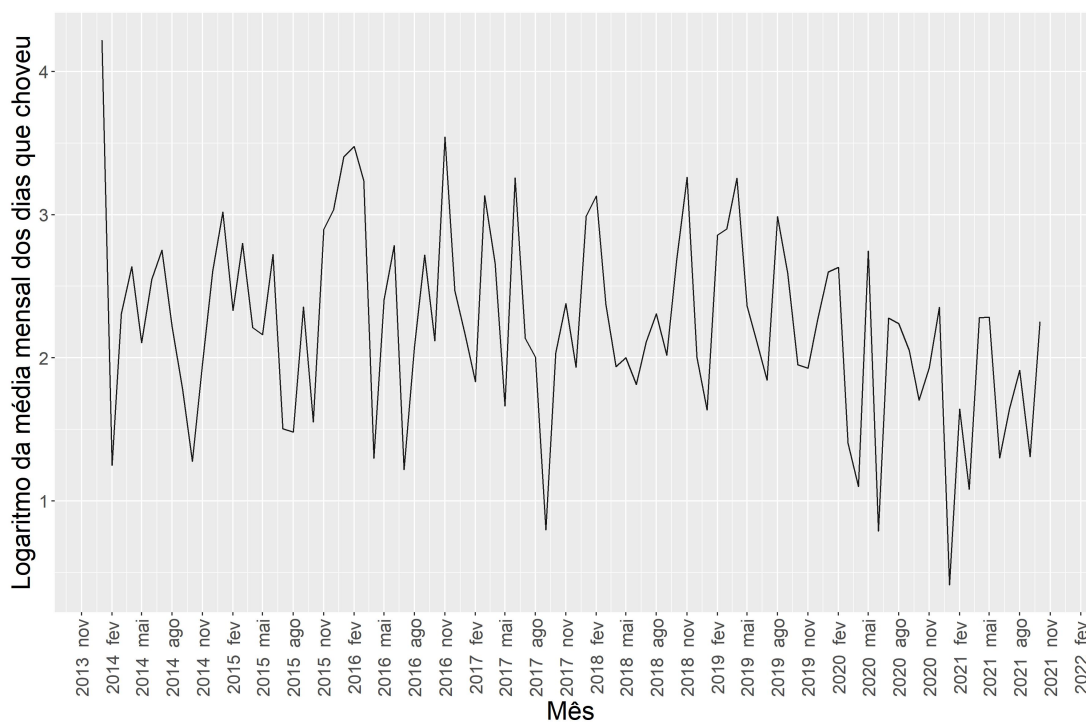


Figura 10: Gráfico do logaritmo do volume médio dos dias de chuva ao longo dos meses observados.

A Figura 10 mostra a série temporal dos dados com a transformação logarítmica da média dos dias chuvosos. Com a transformação aplicada, o pico que ocorre no gráfico da Figura 7 é mais controlado, mas mesmo assim se comporta de forma diferente do restante da série temporal.

Na Figura 11, onde é representado o histograma do volume de chuva com a transformação logarítmica da média dos dias de chuva do mês, é possível concluir que esse tratamento dos dados foi o único em que a distribuição parece ser simétrica, com a maior

frequência ocorrendo próxima à média dos dados.

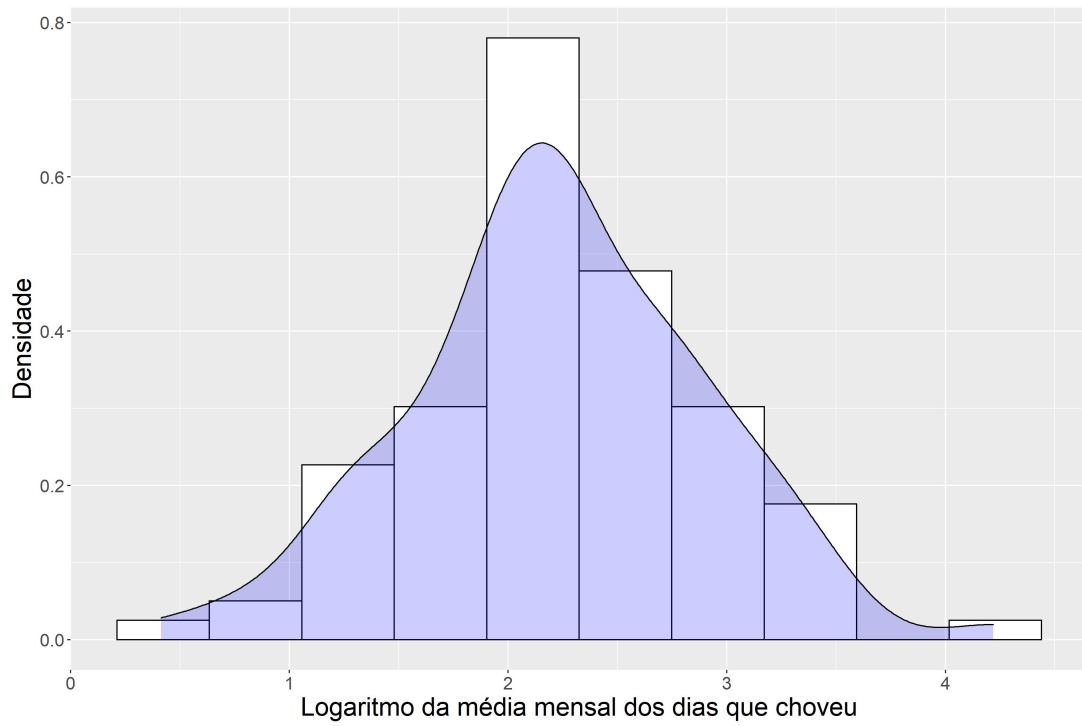


Figura 11: Histograma do volume logarítmico da média de chuva dos meses observados.

Na Figura 12 com a representação da distribuição do volume de chuva, com a análise dos quartis, a simetria do comportamento é comprovada, apesar de ter outliers.



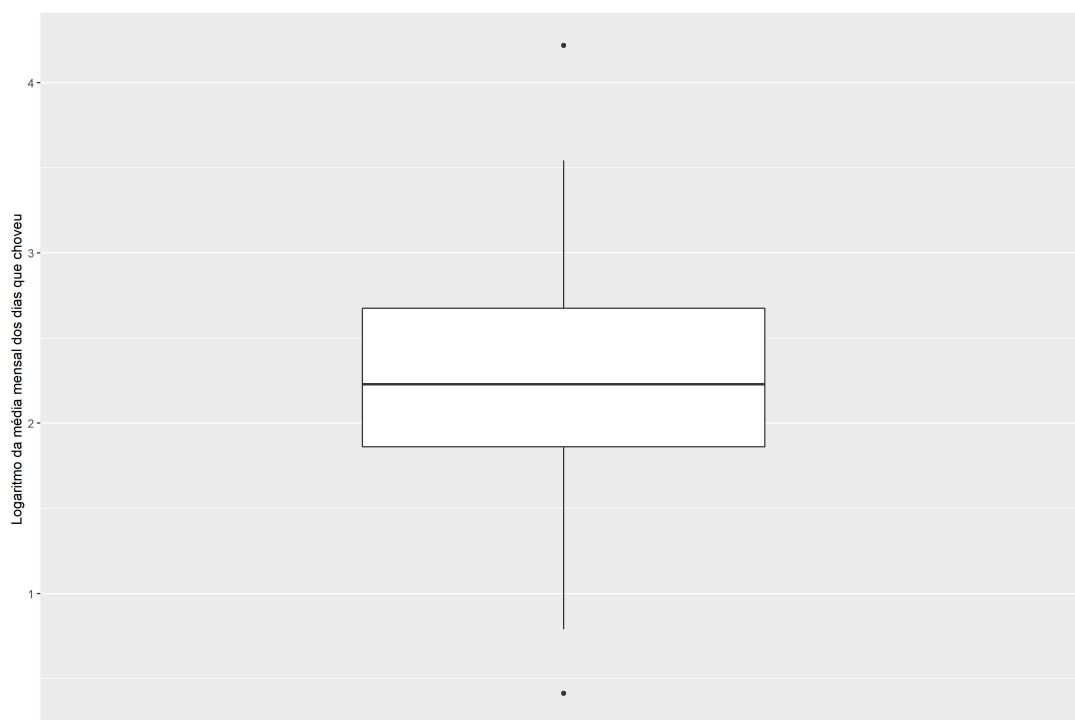


Figura 12: Boxplot do volume logarítmico médio de chuva ao longo dos meses observados.

Mas ao usar a transformação logarítmica, a escala original dos dados é modificada o que torna mais complexa a interpretabilidade e necessidade de uso de transformação inversa para retornar à escala original.

### 3.1.5 Comparação geral das medidas propostas

Ao comparar as Figuras 4 e 1, com as séries temporais referentes ao volume máximo e acumulado mensal, respectivamente, é possível perceber que o agrupamento usando o máximo gera picos ainda mais afastados do nível geral da série.

Ao adotar a média, é possível que dias altamente chuvosos se dispersem com dias com pouca chuva no mês, fazendo com que picos deixem de aparecer nas séries, e gerando perda de informação relevante para tomadas de decisão.

Da mesma forma ao aplicar a transformação logarítmica, apesar de haver ganhos na estrutura da série permitindo assim o uso de modelos mais simples, perdemos o referencial dos dados para interpretabilidade, por não ser analisado na escala original.

## 3.2 Resultados

O ajuste dos modelos e previsões foram realizadas em R [15], utilizando o pacote de modelos lineares dinâmicos [16] [7], tendo sua visualização com o pacote ggplot2 [17]. Para a validação dos modelos, os últimos 6 meses da série temporal foram separados.

Os códigos usados para a análise e os dados extraídos e processados estão localizados no repositório [18] do GitHub.

### 3.2.1 Acumulado

Primeiramente, começamos com o modelo dinâmico mais simples, que é o modelo polinomial de primeira ordem (Capítulo 2.1), um modelo composto por apenas um nível. E então passamos para um modelo polinomial de segunda ordem (Capítulo 2.2), adicionando a tendência.

Para definir os parâmetros de entrada desconhecidos no DLM, foram estimados os parâmetros  $V$  e  $W$ , utilizando o método da máxima verossimilhança, cujos valores são mostrados abaixo.

- Modelo polinomial de primeira ordem:

$$V = [29833, 74]$$

$$W = [614, 9078]$$

- Modelo polinomial de segunda ordem:

$$V = [32135, 09]$$

$$W = \begin{bmatrix} 0,3920903 & 0,000000 \\ 0,000000 & 1,012726 \end{bmatrix}$$

Valores grandes em  $W$  implicam incerteza na evolução da equação do sistema. Sabendo que os valores variam de quinze a quase novecentos, como visto na Figura 3, o valor da variância associada ao erro observacional,  $V$ , é muito grande em ambos os casos.

Tomando os parâmetros estimados como conhecidos, o `dmlFilter` foi usado para o Filtro de Kalman, explicado no Capítulo 2.3.

Na Figura 13 as linhas pontilhadas são as médias a posteriori obtidas com base no Filtro de Kalman, enquanto a linha contínua são os dados reais sem os últimos 6 meses que

foram separados para validação do modelo. Em verde está representado o nível estimado a partir do uso do Filtro de Kalman no modelo polinomial de primeira ordem, enquanto o azul é referente ao de segunda ordem. É possível notar que as estimativas se comportam suavemente com os dados, e os picos não afetam muito a mudança de nível pelo filtro.

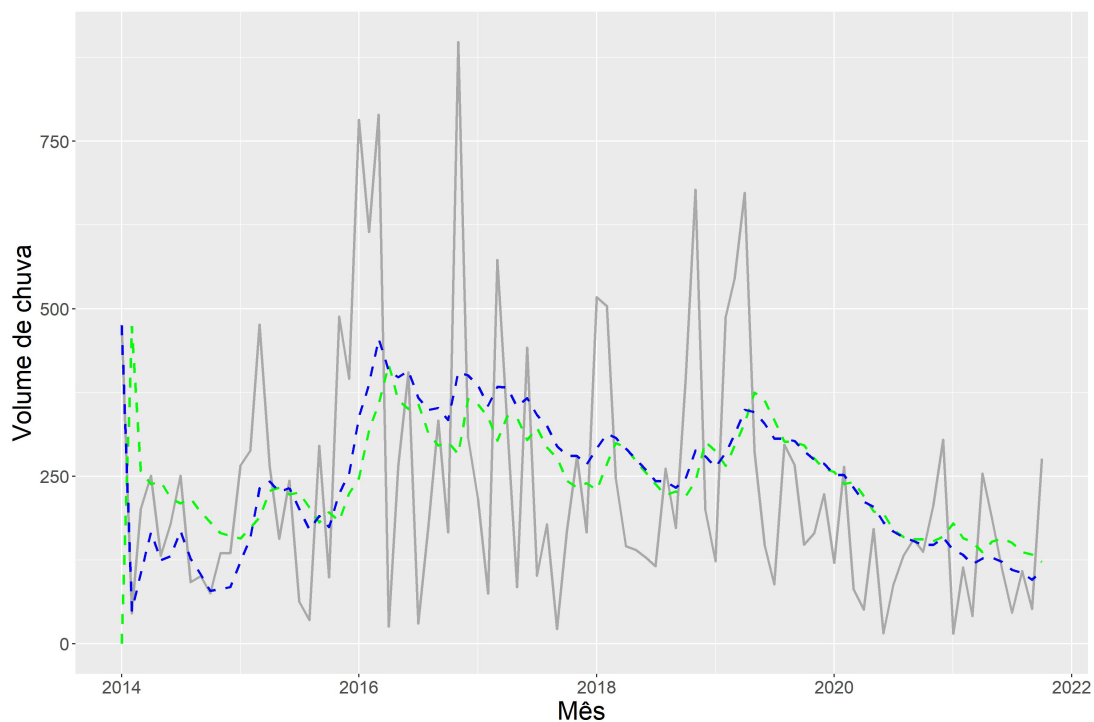


Figura 13: Nível pelo filtro de Kalman para a série de acumulados de chuva dos meses observados.

A Figura 14 mostra a tendência do modelo polinomial de segunda ordem, tendo sido retirados os dois primeiros meses devido ao fato da incerteza inicial ser muito grande, dificultando a visualização dos meses seguintes no gráfico.

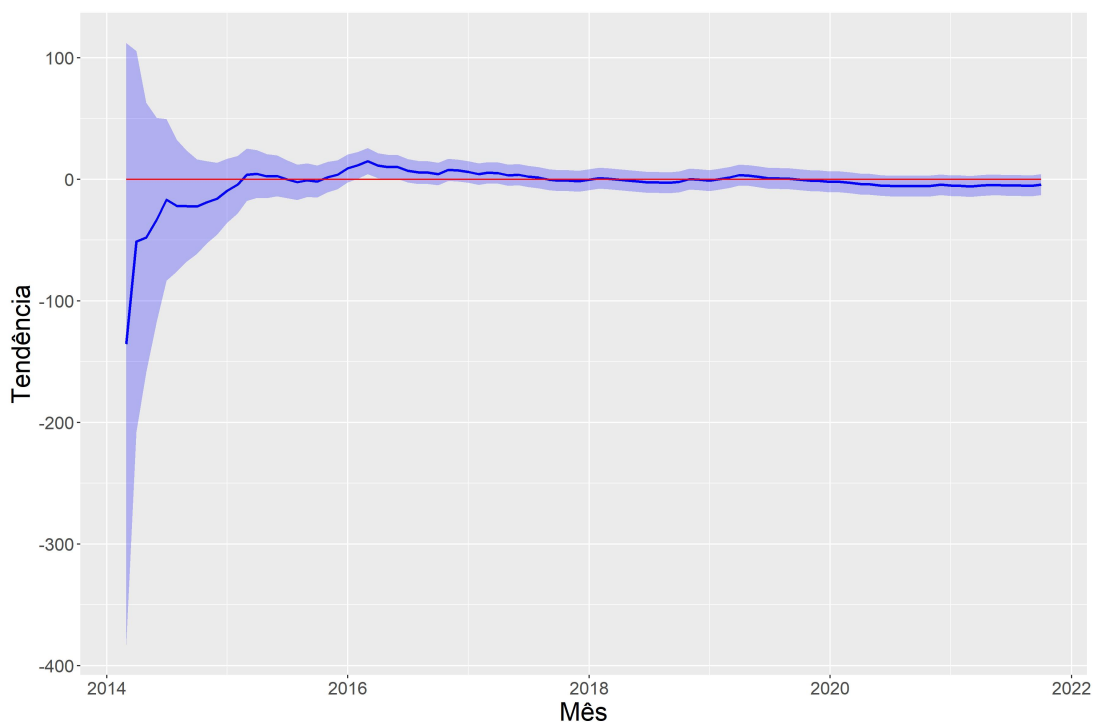


Figura 14: Tendência pelo filtro de Kalman a série de acumulados de chuva dos meses observados.

Em vermelho está uma linha para guiar o valor zero, em azul o valor estimado da tendência e em roxo o intervalo de credibilidade de 95%. É possível observar que a tendência é próxima de zero, geralmente assumindo valores negativos, uma vez que a variação ao longo do tempo para o modelo polinomial de segunda ordem está mais presente no nível. Portanto, visualmente, a série com a transformação acumulada parece não ter um componente de tendência significativo, com apenas alguns picos aleatórios.

Também foi aplicada a suavização que, através do filtro de Kalman, estima as distribuições através dos dados obtidos pela filtragem e previsão um passo à frente [19] [7].

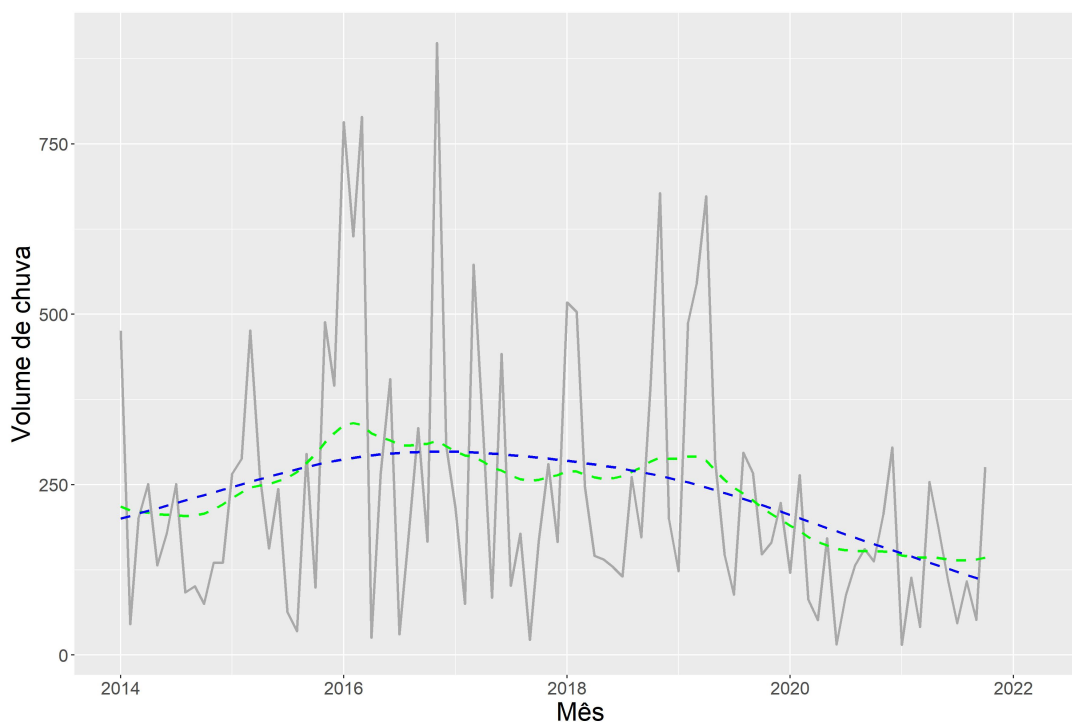
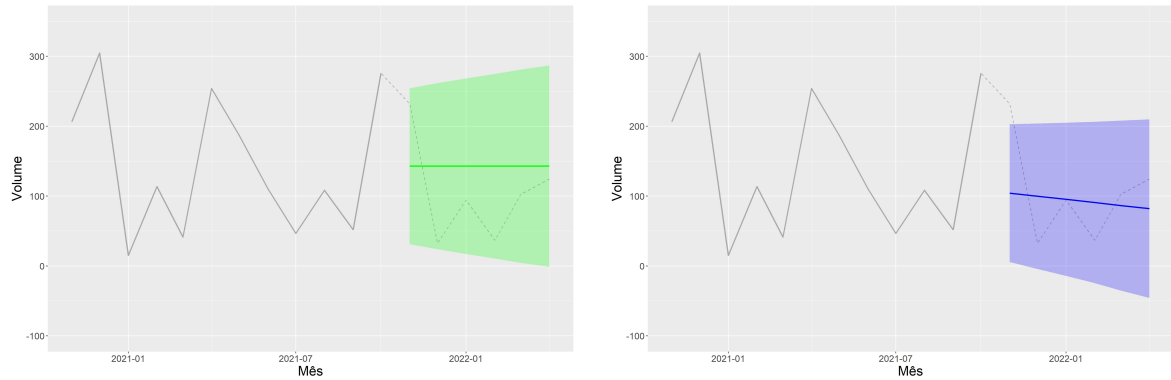


Figura 15: Suavização para a série de acumulados de chuva dos meses observados.

Na Figura 15, a série é representada pela linha contínua, pontilhada são as estimativas pelo suavizador de Kalman. O ajuste de suavização de Kalman do modelo polinomial de primeira ordem, em verde, apresenta mais variações de comportamento durante a série do que o modelo polinomial de segunda ordem, em azul.

Foi realizada uma previsão de seis meses, o argumento usado na função `dlmForecast`, do pacote `dml` [16], é o objeto de classe `dmlFilter`, ou seja, o Filtro de Kalman mostrado anteriormente.

Na Figura 16 está a previsão da transformação acumulada 6 meses à frente. A linha azul escura contínua refere-se à previsão do modelo de segunda ordem, e a sombra roxa representa o intervalo de credibilidade de 95%. Em verde, o modelo polinomial de primeira ordem é representado. Em ambos os modelos, todos os resultados futuros tiveram seus valores dentro do intervalo de credibilidade. Sendo que o modelo polinomial de segunda ordem, além de apresentar previsões pontuais mais próximas dos valores verdadeiros, apresentou menor incerteza associada.



(a) Modelo polinomial de primeira ordem

(b) Modelo polinomial de segunda ordem

Figura 16: Previsão pontual e intervalar 6 meses a frente para a série de acumulados de chuva.

Para avaliar as previsões, três métricas foram utilizadas para analisar o erro, são elas: o Erro médio absoluto (MAE),  $\frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$ , onde é realizada a média da diferença absoluta entre a predição do modelo e o valor real; a raiz quadrada da média das diferenças quadráticas entre os valores preditos e reais (RMSE),  $\sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$ ; o erro percentual absoluto médio (MAPE),  $\frac{1}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i| \times 100\%}{y_i}$ , o qual é uma medida relativa, retornando valores em unidades de porcentagem em vez da dimensão da variável de estudo.

Modelo	Erro		
	MAE	RMSE	MAPE (%)
Polinomial de primeira ordem	68,74	77,07	128,47
Polinomial de segunda ordem	51,40	65,58	76,62

Tabela 1: Tabela com os erros de previsão para a série com o acumulado de chuva no mês, para cada modelo ajustado.

A Tabela 1 informa as medidas de erro dos modelos. Como MAE e RMSE possuem a mesma dimensão que as variáveis de estudo, elas só podem ser utilizadas para avaliar modelos de mesma dimensão, ou seja, avaliação entre os modelos de primeira e segunda ordem da análise pluviométrica acumulada mensal.

Ao contrário do que se esperava do comportamento aparentemente de nível da transformação acumulada, ao adicionar a tendência, os erros de previsão do modelo foram menores do que no modelo apenas com nível.

### 3.2.2 Máximo

Na transformação de máximo, também foram ajustados dois modelos, os polinomiais de primeira e segunda ordem. E os parâmetros  $V$  e  $W$  foram estimados pelo método da máxima verossimilhança.

- Modelo polinomial de primeira ordem:

$$V = [4032, 973]$$

$$W = [11, 09326]$$

- Modelo polinomial de segunda ordem:

$$V = [3939, 285]$$

$$W = \begin{bmatrix} 26,77531 & 0,000000 \\ 0,000000 & 1,784857e - 07 \end{bmatrix}$$

Conforme mencionado no Capítulo 2,  $V$  é a matriz de variância do modelo observacional, enquanto  $W$  é a variância associada ao erro de evolução. Como o volume máximo diário de precipitação mensal é de  $7,44mm$  a  $381,31mm$ , as variações estimadas estão próximas das expectativas.

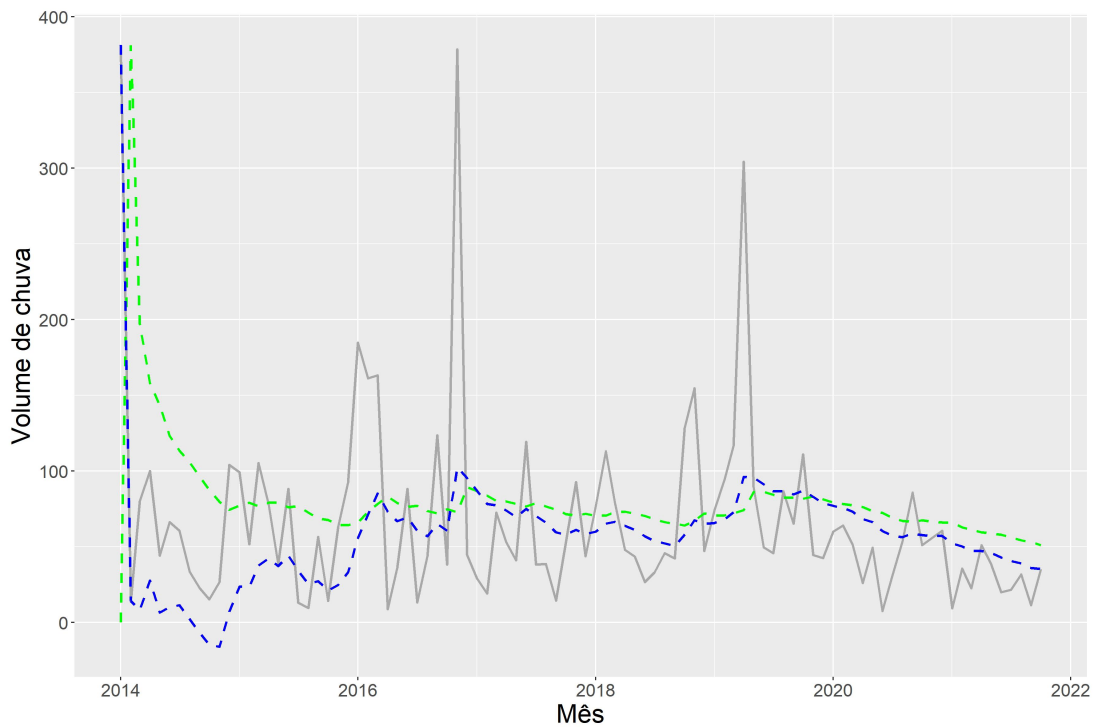


Figura 17: Nível pelo filtro de Kalman para a série de máximo de chuva dos meses observados.

No gráfico da Figura 17 os níveis de primeira ordem são representados em verde, e os de segunda ordem em azul, em cinza são os valores dos dados reais.

Para o Filtro de Kalman, foram utilizados os parâmetros estimados na função `dlm-Filter`. Após o primeiro pico que ocorre bem no início da série, em ambos os modelos, o Filtro de Kalman parece se ajustar bem, e não ser tão afetado pelos outros picos que ocorrem ao longo da série.

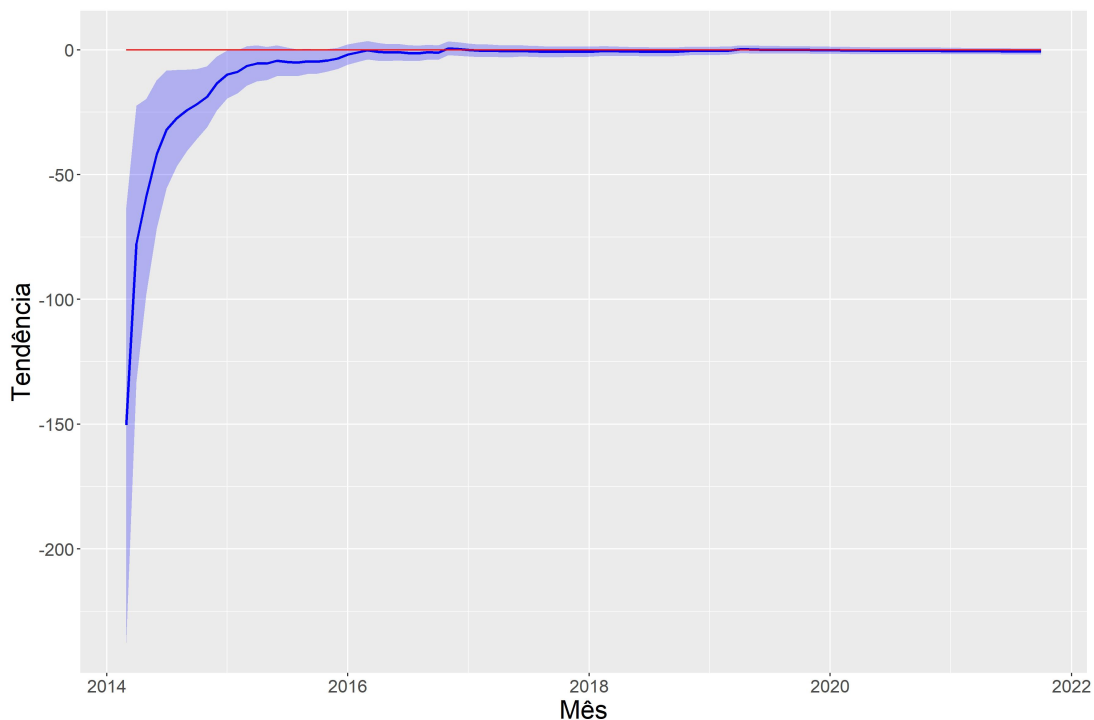


Figura 18: Tendência pelo filtro de Kalman para a série de máximo de chuva dos meses observados.

Na Figura 18 está a tendência do modelo polinomial de segunda ordem, também foi necessário remover os dois primeiros meses devido ao fato da incerteza inicial ser muito grande. O sombreamento roxo é o intervalo de credibilidade de 95%, e a linha vermelha contínua ajuda a visualizar a tendência igual a zero. Assim, é possível verificar que a tendência é negativa até 2016, após este ano, mantém-se muito próxima de zero.



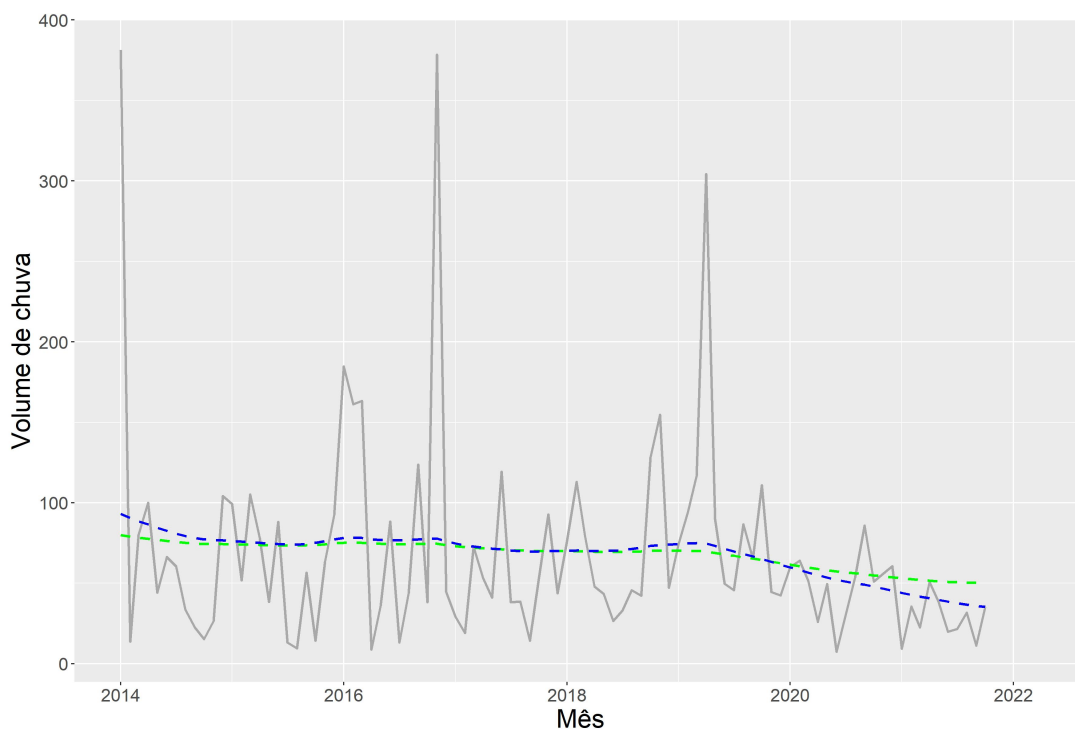


Figura 19: Suavização para a série de máximo de chuva dos meses observados.

A Figura 19 mostra a suavização dos dados com o agrupamento mensal de máximo. Aqui o modelo de primeira ordem, em verde, e o modelo de segunda ordem, em azul, estão muito próximos.

Embora a suavização tenha sido aplicada após o Filtro de Kalman, a previsão para os próximos seis meses foi realizada utilizando o Filtro de Kalman.



(a) Modelo polinomial de primeira ordem

(b) Modelo polinomial de segunda ordem

Figura 20: Previsão pontual e intervalar 6 meses a frente para a série de máximo de chuva.

As linhas contínuas azul e verde referem-se às médias a posteriori da previsão e as sombras representam o intervalo de credibilidade de 95%. No modelo polinomial de pri-

meira ordem, Figura 20(a), os valores reais para alguns meses estão fora do intervalo de credibilidade, enquanto no modelo polinomial de segunda ordem, representado na Figura 20(b), os resultados futuros têm seu valor dentro do intervalo de credibilidade.

Modelo	Erro		
	MAE	RMSE	MAPE (%)
Polinomial de primeira ordem	14,63	19,28	68,41
Polinomial de segunda ordem	13,07	14,66	43,03

Tabela 2: Tabela com os erros de previsão para a série com o máximo de chuva no mês, para cada modelo ajustado.

Ao observar a Tabela 2, constata-se que o modelo com tendência teve um desempenho melhor do que o modelo somente de nível. Considerando pelo MAPE, ambos os modelos realizados a partir da transformação máxima tiveram resultados melhores que os modelos realizados a partir da transformação acumulada.

### 3.2.3 Média dos dias de chuva

Na transformação da média, também foram realizados dois modelos. Começando com o modelo composto apenas por nível, e então um novo modelo acrescentando a tendência.

Novamente, a matriz de variância do modelo e a matriz de variância observacional foram estimadas pelo método da máxima verossimilhança. Porque esses parâmetros desconhecidos são necessários como parâmetros de entrada no DLM.

- Modelo polinomial de primeira ordem:

$$V = [78, 03]$$

$$W = [0, 341723]$$

- Modelo polinomial de segunda ordem:

$$V = [78, 54434]$$

$$W = \begin{bmatrix} 1,965528e - 06 & 0,000000 \\ 0,000000 & 1,539565e - 24 \end{bmatrix}$$

Como o mínimo é  $1,51mm$ , o máximo é  $67,96mm$ , e o terceiro quartil é  $14,52mm$  na

Figura 9, a variância estimada não parece ser muito grande para o comportamento dos dados.

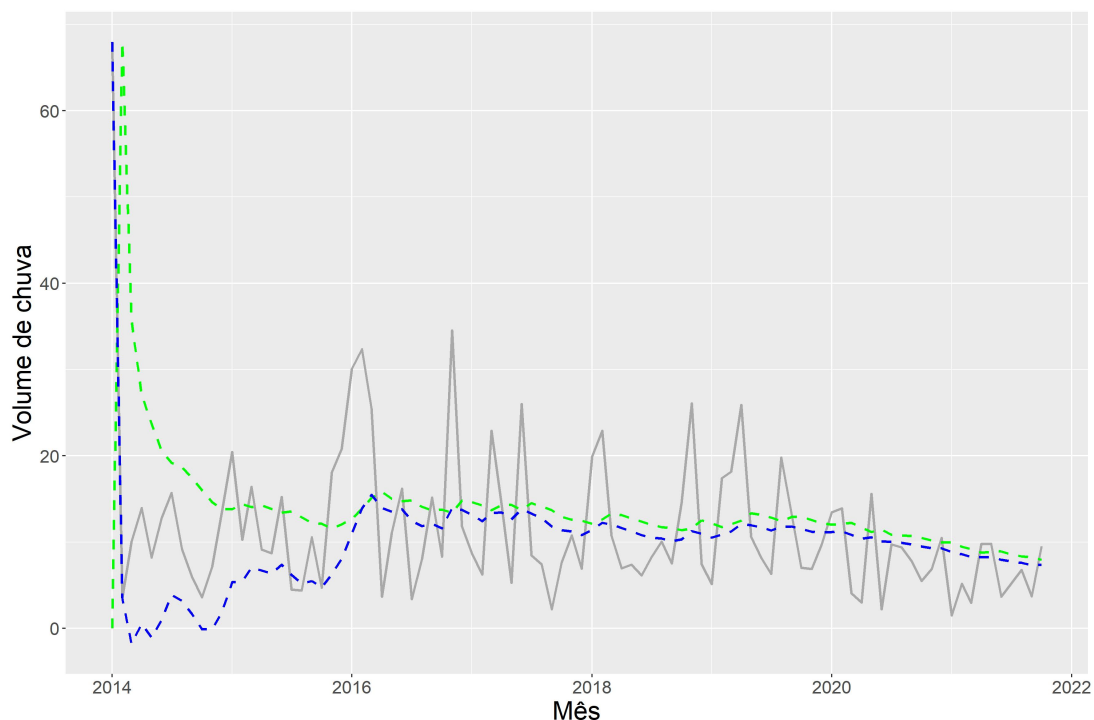


Figura 21: Nível pelo filtro de Kalman para a série de média de chuva dos meses observados.

Na Figura 21, o nível do modelo de primeira ordem que está representado em verde, e o de segunda ordem em azul, começam a se comportar de maneira similar após 2016. Lembrando que os dados transformados estão em cinza.

Os parâmetros estimados foram utilizados no Filtro de Kalman. Devido ao valor muito alto de 2014, vemos uma grande queda na estimativa pontual apenas no início, depois disso, os valores tanto para os modelos de primeira ordem quanto para os de segunda ordem parecem ficar mais estáveis.

A Figura 22 representa a tendência do modelo de segunda ordem, em vermelho está representado o valor da tendência igual a zero. Para melhor visualização, os dois primeiros meses da imagem foram removidos. É possível verificar que os valores de tendência são negativos até aproximadamente 2015, e o intervalo de credibilidade também está abaixo de zero. Mas depois deste ano, tanto o intervalo de credibilidade quanto os valores de tendência estão muito próximos de zero.

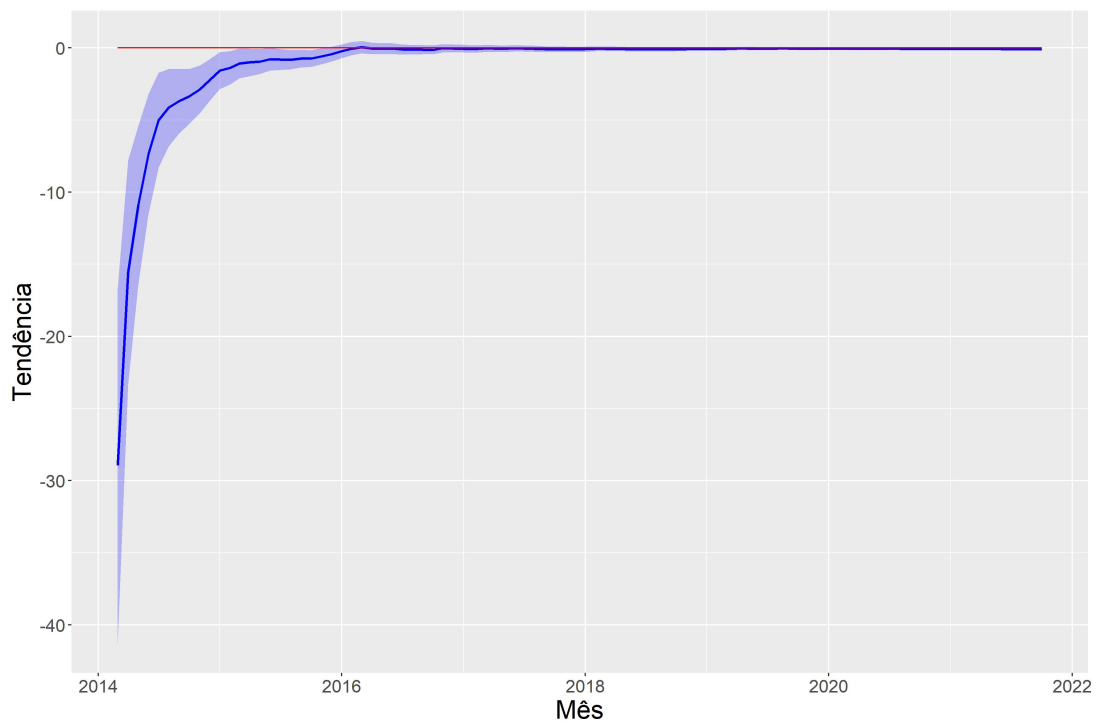


Figura 22: Tendência pelo filtro de Kalman para a série de média de chuva dos meses observados.

Após o Filtro de Kalman, foi aplicada a suavização, Figura 23.

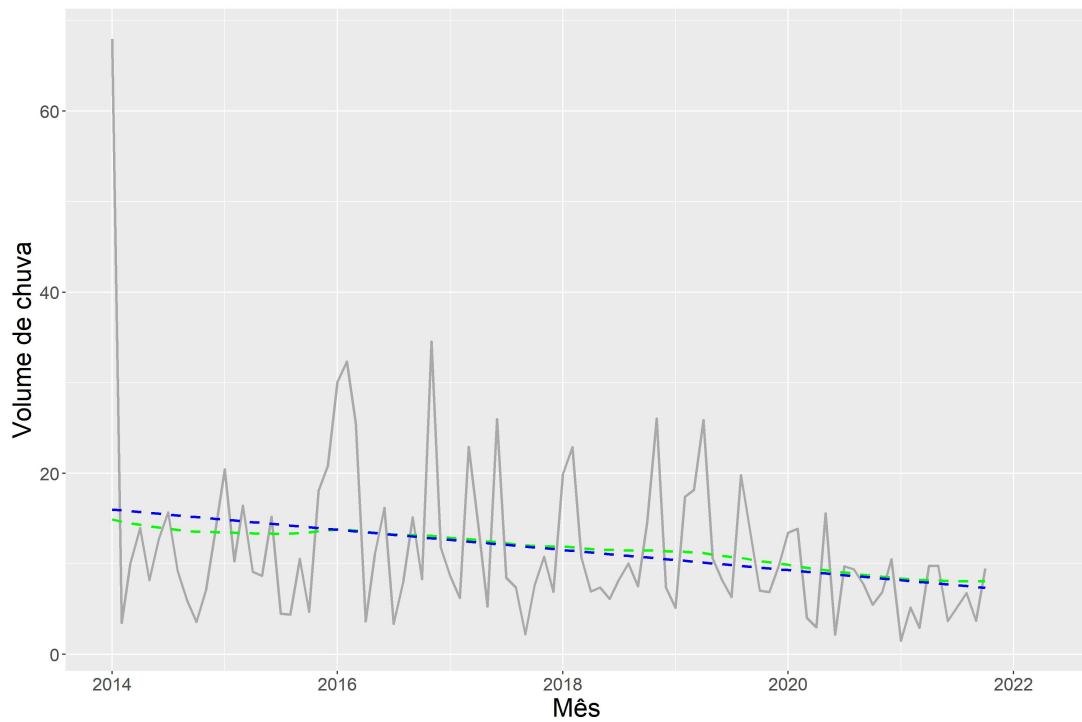
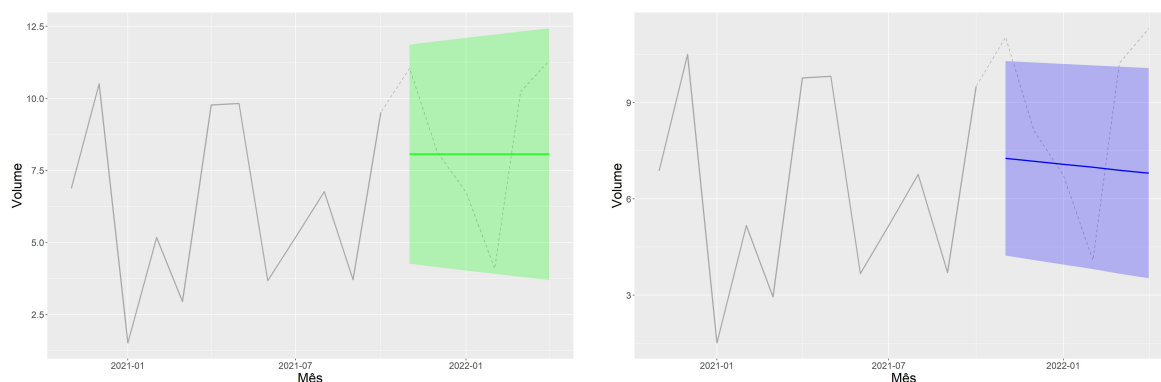


Figura 23: Suavização para a série de média de chuva dos meses observados.

Representado pela linha pontilhada azul está a suavização do modelo de segunda ordem, e em verde a de primeira ordem, é possível verificar que são muito semelhantes ao longo da série.

Para a previsão de seis meses à frente, foi utilizado o filtro de Kalman. A Figura 24(a) mostra a série em cinza, a previsão em uma linha verde contínua e o intervalo de credibilidade 95%. A Figura 24(b) mostra o modelo polinomial de segunda ordem, com a série em cinza e a previsão em azul.



(a) Modelo polinomial de primeira ordem,

(b) Modelo polinomial de segunda ordem,

Figura 24: Previsão pontual e intervalar 6 meses a frente para a série de média de chuva.

Todos os valores reais ficaram dentro do intervalo de credibilidade do modelo de primeira ordem, enquanto o mesmo não ocorreu para o modelo de segunda ordem.

Modelo	Erro		
	MAE	RMSE	MAPE (%)
Polinomial de primeira ordem	2,29	2,64	32,43
Polinomial de segunda ordem	2,64	3,03	32,40

Tabela 3: Tabela com os erros de previsão para a série com a média de chuva no mês, para cada modelo ajustado.

Na Tabela 3 de erros, para MAE e RMSE, o modelo de primeira ordem foi o mais acertado. No entanto para MAPE, o de segunda ordem foi um pouco melhor, mas não é prudente tirar conclusões sobre qual modelo é o mais razoável com essa pequena diferença.

### 3.2.4 Logaritmo da média diária de chuva

Dois modelos foram ajustados, os polinômios de primeira e segunda ordem. A estimação dos parâmetros  $W$ , variância associada ao erro, e  $V$ , variância do modelo observacional, necessários como parâmetros de entrada do MLD, foi realizada pelo método de máxima verossimilhança.

Tendo na Figura 12 o mínimo igual a  $0,41mm$  e o máximo de  $4,22mm$ , a variância estimada não parece ser grande para o comportamento simétrico dos dados estudados.

- Modelo polinomial de primeira ordem:

$$V = [0,4097235]$$

$$W = [0,002676958]$$

- Modelo polinomial de segunda ordem:

$$V = [0,4138383]$$

$$W = \begin{bmatrix} 1,231042e-09 & 0,000000 \\ 0,000000 & 4,804338e-06 \end{bmatrix}$$

Para o Filtro de Kalman, foram utilizados os parâmetros estimados.

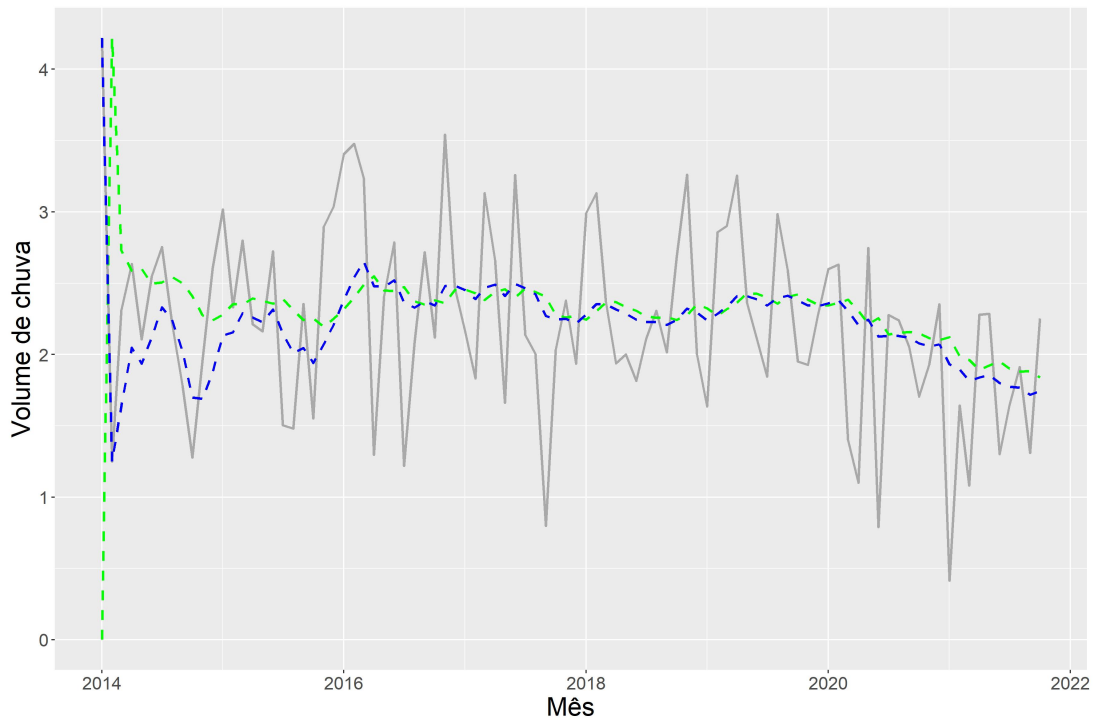


Figura 25: Nível pelo filtro de Kalman para a série de logaritmo da média de chuva dos meses observados.

A Figura 25 é composta pelo nível do modelo de primeira ordem em verde, o modelo de segunda ordem em azul e os dados transformados em cinza. A partir de 2016, ambos os níveis parecem se aproximar.

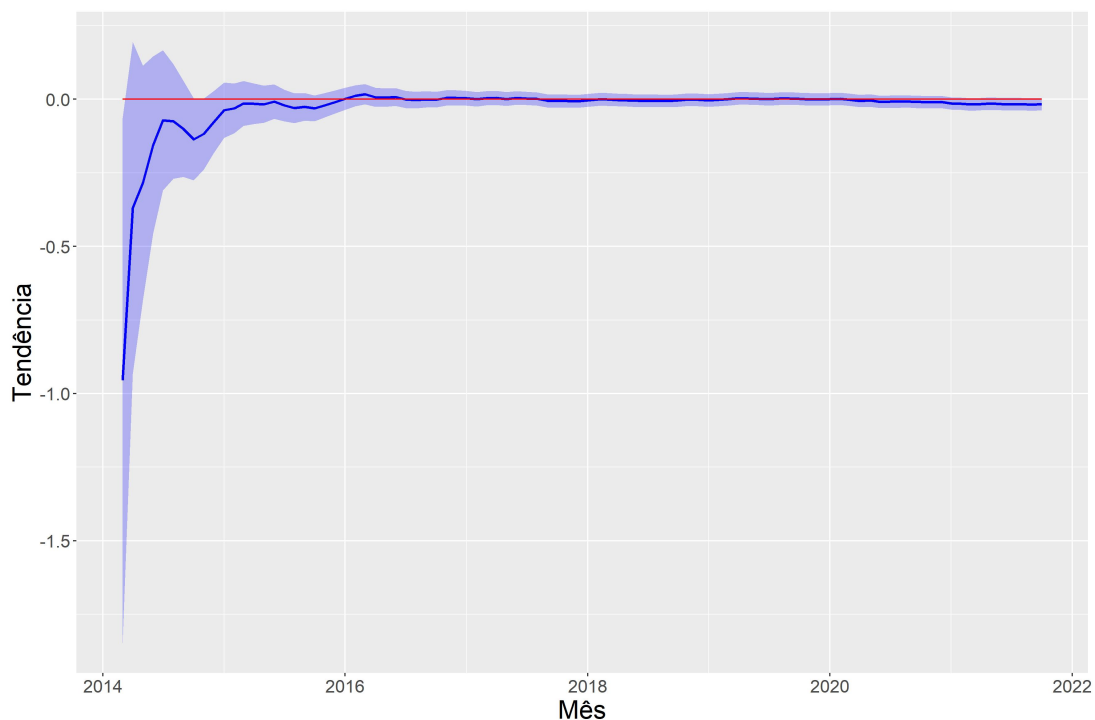


Figura 26: Tendência pelo filtro de Kalman para a série de logaritmo da média de chuva dos meses observados.

No modelo de segunda ordem, a tendência é negativa até aproximadamente 2015, após esta data permanece muito próxima de zero, Figura 26. Em nenhum estante de tempo a tendência se mostrou significativa ao nível de 95%.

A partir do Filtro de Kalman foi possível realizar a suavização. Onde é possível, pela Figura 27, verificar que o comportamento de ambos os modelos é semelhante ao longo da série.

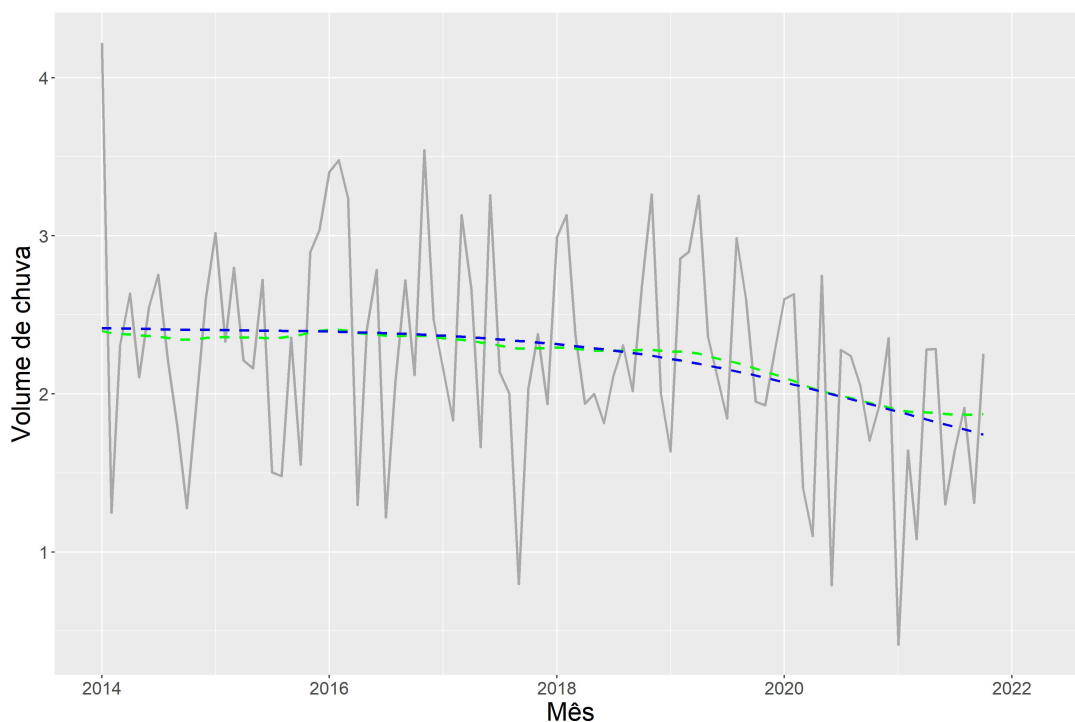
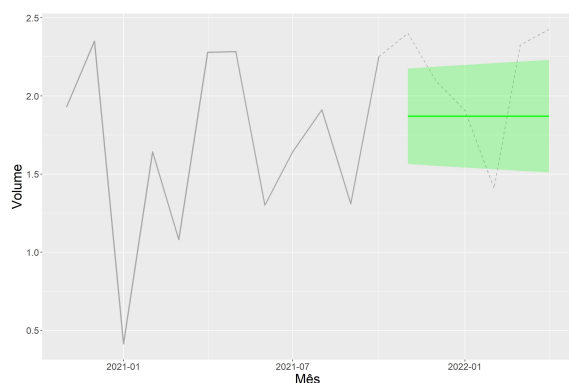
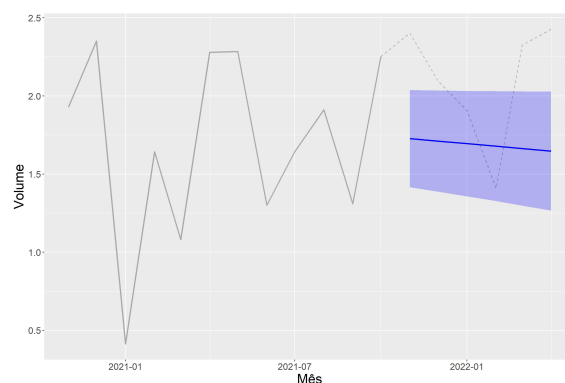


Figura 27: Suavização para a série de logaritmo da média de chuva dos meses observados.

Foram guardados seis meses dos dados para a validação do modelo. Para ambos modelos foram usados os resultados do Filtro de Kalman, mostrados anteriormente.



(a) Modelo polinomial de primeira ordem



(b) Modelo polinomial de segunda ordem

Figura 28: Previsão pontual e intervalar 6 meses a frente para a série de logaritmo da média de chuva.

Na Figura 28(b), a linha azul escura contínua refere-se ao modelo de segunda ordem, e a sombra azul representa o intervalo de credibilidade de 95%. Os resultados reais, em sua maioria, não estão dentro do intervalo de credibilidade.

Já na Figura 28(a), a linha verde contínua refere-se ao modelo de primeira ordem e a



sombra o intervalo de credibilidade de 95%. Os resultados reais também não se enquadram no intervalo de credibilidade.

Modelo	Erro		
	MAE	RMSE	MAPE (%)
Polinomial de primeira ordem	0,38	0,42	18,34
Polinomial de segunda ordem	0,50	0,54	22,90

Tabela 4: Tabela com os erros de previsão para a série com o logaritmo da média de chuva no mês, para cada modelo ajustado.

Pela Tabela 4, para MAE, RMSE e MAPE, o modelo de primeira ordem foi o melhor, mesmo entre todos os cenários estudados neste trabalho. Mostrando que a transformação do logaritmo da média forneceu dados mais bem comportados, como pode ser visto na Figura 11, então um modelo mais parcimonioso foi capaz de obter boas previsões.

## 4 Conclusão

Dentre os agrupamentos aplicados aos dados do estudo, verificou-se que a proposta de trabalhar com a transformação logarítmica da média mensal com os dias em que houve chuva, foi a transformação que melhor se comportou e retornou os melhores resultados ao analisar os erros dos modelos aplicados.

O modelo mais simples, modelo polinomial de primeira ordem, resultou nos menores erros quando aplicado a transformação do logaritmo da média mensal de chuvas. Mas não foi o melhor quando comparado ao modelo de segunda ordem para as outras transformações aplicadas nos dados pluviométricos.

Neste trabalho foram considerados apenas modelos lineares dinâmicos com distribuição Normal para as observações, por serem modelos mais simples. No entanto, o comportamento assimétrico e valores atípicos presente nas séries revelam a necessidade do uso de modelos que capturem melhor esse tipo de comportamento, quando comparados ao modelo normal. O uso da transformação logarítmica neste trabalho corrobora com este fato. Uma possibilidade neste caso é o uso do modelo Gama ou Teoria de Valores Extremos [20]. Por outro lado, a grande quantidade de valores iguais a zero poderia ser acomodada se modelos inflacionados de zero fossem utilizados [21]. Além disso, adicionar parâmetros relacionados à meteorologia, como informações auxiliares no modelo o tornaria mais completo.

## Referências

- [1] SANSO, B.; GUENNI, L. A stochastic model for tropical rainfall at a single location. *Journal of Hydrology*, v. 214, n. 1, p. 64–73, 1999. ISSN 0022-1694. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0022169498002418>>.
- [2] SILVA, S. V. C. da. *Análise Bayesiana da precipitação máxima de Petrópolis-RJ*. 07 2021. <https://bdtd.unifal-mg.edu.br:8443/handle/tede/1926>. (Accessed on 06/29/2022).
- [3] CEMADEN. *Cemaden - Centro Nacional de Monitoramento e Alertas de Desastres Naturais*. <http://www2.cemaden.gov.br/mapainterativo/>.
- [4] INMET. *INMET - Instituto Nacional de Meteorologia do Brasil*. <https://tempo.inmet.gov.br/TabelaEstacoes/A001>. (Accessed on 05/26/2022).
- [5] WEST, M.; HARRISON, J. *Bayesian forecasting and dynamic models*. 2. ed. New York, NY: Springer, 1997. (Springer Series in Statistics).
- [6] GAMERMAN, D. *Curso de Modelos Dinâmicos*. 8 1997. <http://www.dme.ufrj.br/dani/pdf/IPEA.pdf>.
- [7] PETRIS, G.; PETRONE, S.; CAMPAGNOLI, P. *Dynamic Linear Models with R*. [S.l.]: Springer-Verlag, New York, 2009. (useR!).
- [8] MIGON DANI GAMERMAN, R. R. H. *Introdução aos Modelos Dinâmicos Bayesianos*. 11 2004. <http://www.din.uem.br/sbpo/sbpo2004/pdf/arq0278.pdf>. XXVI SBPO - Simposio Brasileiro de Pesquisa Operacional - Sao João del Rey.
- [9] SIDIQ, M. Forecasting rainfall with time series model. *IOP Conference Series: Materials Science and Engineering*, IOP Publishing, v. 407, p. 012154, sep 2018. Disponível em: <<https://doi.org/10.1088/1757-899x/407/1/012154>>.
- [10] ALAGAMENTOS em Niterói e SG. <https://www.ofluminense.com.br/pt-br/cidades/alagamentos-em-niter%C3%B3i-e-sg.html>. (Accessed on 06/29/2022).
- [11] FORTE chuva em Niterói alaga hospitais Getulinho e Azevedo Lima - Jornal O Globo. <https://oglobo.globo.com/rio/bairros/forte-chuva-em-niteroi-alaga-hospitais-getulinho-azevedo-lima-20495612>. (Accessed on 06/29/2022).
- [12] CONTI, J. B. *Clima e meio ambiente*. 7. ed. [S.l.]: Atual, 2019.
- [13] G1. *Chuva forte deixa Rio em estágio de atenção; Niterói entra em estágio de alerta — Rio de Janeiro — G1*. 03 2019. <https://g1.globo.com/rj/rio-de-janeiro/noticia/2019/03/01/>

- chuva-forte-alaga-varios-bairros-do-rio-e-coloca-cidade-em-estagio-de-atencao. ghtml.
- [14] G1 - Temporal alaga ruas, fecha aeroporto e causa falta de luz no Rio - notícias em Rio de Janeiro. <https://g1.globo.com/rio-de-janeiro/noticia/2014/01/temporal-cai-e-deixa-rio-e-baixada-fluminense-em-estagio-de-atencao.html>. (Accessed on 06/29/2022).
- [15] R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2021. Disponível em: <<https://www.R-project.org/>>.
- [16] PETRIS, G. An R package for dynamic linear models. *Journal of Statistical Software*, v. 36, n. 12, p. 1–16, 2010. Disponível em: <<http://www.jstatsoft.org/v36/i12/>>.
- [17] WICKHAM, H. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016. ISBN 978-3-319-24277-4. Disponível em: <<https://ggplot2.tidyverse.org>>.
- [18] PEREIRA, D. C. B. *Análise de precipitação pluviométrica no município de Niterói via modelos dinâmicos bayesianos*. 8 2022. <https://github.com/D3B0R4H/PGCD/tree/master>. Repositório com os códigos utilizados na monografia.
- [19] CUTRIM, I. A. *Uma Revisão de Modelos Lineares Dinâmicos*. 7 2014. <https://www.ufjf.br/cursoestatistica/files/2014/04/Uma-Revis~ao-de-MLD-final.pdf>.
- [20] SOUZA, L. d. S. et al. Modelando a chuva máxima diária no município de João Pessoa-PB por meio da Teoria dos Valores Extremos. *Revista Brasileira de Climatologia*, v. 30, p. 488–503, mar. 2022. ISSN 2237-8642, 2237-8642. Disponível em: <<https://ojs.ufgd.edu.br/index.php/rbclima/article/view/14886>>.
- [21] LEE, C.-E.; KIM, S. Applicability of Zero-Inflated Models to Fit the Torrential Rainfall Count Data with Extra Zeros in South Korea. *Water*, v. 9, n. 2, p. 123, fev. 2017. ISSN 2073-4441. Disponível em: <<http://www.mdpi.com/2073-4441/9/2/123>>.