

UNIVERSIDADE FEDERAL DE SANTA MARIA
CENTRO DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA CIVIL

Luiza Chiarelli Conte

**GRÁFICOS DE CONTROLE PARA VARIÁVEIS
AMBIENTAIS DUPLAMENTE LIMITADAS E
AUTOCORRELACIONADAS**

Santa Maria, RS, Brasil
2020

Luiza Chiarelli Conte

**GRÁFICOS DE CONTROLE PARA VARIÁVEIS AMBIENTAIS DUPLAMENTE
LIMITADAS E AUTOCORRELACIONADAS**

Dissertação apresentada ao Curso de Pós-Graduação em Engenharia Civil (PPGEC), da Universidade Federal de Santa Maria (UFSM, RS), como requisito parcial para obtenção do grau de **Mestre em Engenharia Civil**

Orientadora: Prof. Dr. Débora Missio Bayer
Co-orientador: Prof. Dr. Fábio Mariano Bayer

Santa Maria, RS, Brasil
2020

Chiarelli Conte, Luiza

GRÁFICOS DE CONTROLE PARA VARIÁVEIS AMBIENTAIS
DUPLAMENTE LIMITADAS E AUTOCORRELACIONADAS / por
Luiza Chiarelli Conte. – 2020.

98 f.: il.; 30 cm.

Orientadora: Débora Missio Bayer

Coorientador: Fábio Mariano Bayer

Dissertação (Mestrado) - Universidade Federal de Santa Maria,
Centro de Tecnologia, PPGEC, RS, 2020.

1. Gráficos de controle. 2. Hidrelétrica de Furnas. 3. Modelo
KARMA. 4. *OUTLIERS*. 5. Reservatórios de água. 6. Sistema
Cantareira. 7. Verossimilhança ponderada. I. Missio Bayer, Débora.
II. Bayer, Fábio Mariano. III. Título.

© 2020

Todos os direitos autorais reservados a Luiza Chiarelli Conte. A reprodução de partes ou do todo deste trabalho só poderá ser feita mediante a citação da fonte.

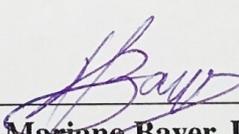
E-mail: cconteluiza@gmail.com

Luiza Chiarelli Conte

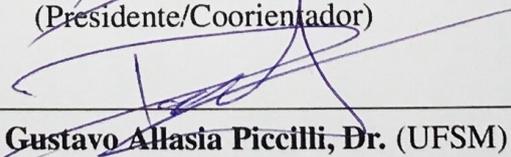
**GRÁFICOS DE CONTROLE PARA VARIÁVEIS AMBIENTAIS DUPLAMENTE
LIMITADAS E AUTOCORRELACIONADAS**

Dissertação apresentada ao Curso de Pós-Graduação em Engenharia Civil (PPGEC), da Universidade Federal de Santa Maria (UFSM, RS), como requisito parcial para obtenção do grau de **Mestre em Engenharia Civil**

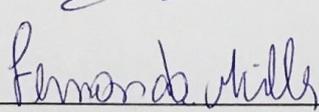
Aprovado em 28 de 02 de 2020:



Fábio Mariano Bayer, Dr. (UFSM)
(Presidente/Coorientador)



Daniel Gustavo Alhasia Picilli, Dr. (UFSM)



Fernanda Maria Müller, Dr. (UFRGS)

Santa Maria, RS
2020

AGRADECIMENTOS

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

RESUMO

GRÁFICOS DE CONTROLE PARA VARIÁVEIS AMBIENTAIS DUPLAMENTE LIMITADAS E AUTOCORRELACIONADAS

AUTORA: LUIZA CHIARELLI CONTE
ORIENTADORA: DÉBORA MISSIO BAYER
CO-ORIENTADOR: FÁBIO MARIANO BAYER

Na área ambiental, em especial na hidrológica, diversas atividades necessitam de ferramentas de monitoramento de variáveis que auxiliem nas tomadas de decisões. Os gráficos de controle são uma ferramenta do controle estatístico de processos (CEP) que podem ser utilizados para esse fim. No entanto, um dos pressupostos assumidos para seu uso é a independência entre as observações. Em alguns processos esse pressuposto pode não se verificar, como é o caso de séries temporais hidrológicas, prejudicando o desempenho dos gráficos de controle usuais. Uma forma de contornar essa situação pode ser dada por meio do monitoramento dos resíduos de um modelo de séries temporais ajustado aos dados de interesse, como, por exemplo, o modelo Kumaraswamy autorregressivo de médias móveis (KARMA), o qual foi proposto recentemente para modelagem de séries temporais ambientais duplamente limitadas. Nesse contexto, este trabalho propõe gráficos de controle para variáveis duplamente limitadas e autocorrelacionadas baseados no modelo KARMA. Os resultados foram avaliados numericamente por simulações de Monte Carlo, analisando o comprimento médio de corrida (*average run length* - ARL) da série sob controle (ARL_0) e fora de controle (ARL_1). Comparou-se os desempenhos dos gráficos propostos com outras metodologias da literatura, sob diferentes cenários. Os gráficos de controle KARMA se destacaram em quase todos os cenários, apresentando as menores distorções de ARL_0 e uma taxa de acerto superior ou próxima a dos outros modelos avaliados sob condições fora de controle. Em um segundo momento, é proposta uma metodologia robusta para minimizar o efeito da presença de *outliers*, comuns em séries históricas hidrológicas, no desempenho dos gráficos de controle, utilizando estimadores de máxima verossimilhança ponderada. Via simulação de Monte Carlo, os estimadores foram avaliados em termos de medidas específicas de robustez e comparado o desempenho dos gráficos de controle inicialmente propostos e a abordagem robusta. Identificou-se que os gráficos de controle robustos apresentam melhor desempenho quando na presença de *outliers*. Por fim, as técnicas desenvolvidas são utilizadas no monitoramento de séries hidrológicas reais e seus resultados são discutidos. Os gráficos de controle se mostraram uma ferramenta útil para o gerenciamento de sistemas de armazenamento de água, como é o caso do Sistema Cantareira e do reservatório da hidrelétrica de Furnas. Em 2014 foi noticiada uma crise no fornecimento de água nestes sistemas, e os gráficos propostos foram capazes de identificá-la. Desta forma, verifica-se a potencialidade da utilização dos gráficos propostos no monitoramento de níveis em reservatórios de água.

Palavras-chave: Gráficos de controle. Hidrelétrica de Furnas. Modelo KARMA. *Outliers*. Reservatórios de água. Sistema Cantareira. Verossimilhança ponderada.

ABSTRACT

CONTROL CHARTS FOR DOUBLE BOUNDED AND AUTOCORRELATED ENVIRONMENTAL DATA

AUTHOR: LUIZA CHIARELLI CONTE
ADVISOR: DÉBORA MISSIO BAYER
COADVISOR: FÁBIO MARIANO BAYER

In the environmental field, especially in hydrology, several activities require monitoring tools to assist in the process of decision making. Control chart is a statistical process control (SPC) tool that can be used for this purpose. However, one of the assumptions made for its use is the independence between different observations. In some processes, this assumption could not be verified, as in hydrological time series, which reduces the applicability of the usual control charts. A solution for this can be given by monitoring the residuals of a fitted time series model, such as, the Kumaraswamy autoregressive moving averages (KARMA) model, which was recently proposed for modeling double bounded environmental time series. In this context, this work proposes control charts for double bounded and autocorrelated data based on the KARMA model. The results were numerically evaluated using Monte Carlo simulations, by analyzing the average run length (ARL) of the series under control (ARL_0) and out of control (ARL_1). The performance of the proposed control charts were compared with other methodologies in the literature, under different scenarios. The KARMA control charts outperforms the competitors in several scenarios, presenting the smallest distortions for ARL_0 and the best power detection rates under out of control conditions. In a second part of this work, a robust methodology using weighted maximum likelihood estimators is proposed aiming to minimize the effect of outliers, which are typically present in historical hydrological series, in the performance of control charts. The estimators were evaluated with Monte Carlo simulations in terms of specific robustness measures and by comparing the performance of the control charts initially proposed versus the robust approach. It was identified that the robust control charts present better performance in the presence of outliers. Finally, the developed techniques are employed in real monitoring data of hydrological systems and the results are discussed. These control charts proved to be a useful tool for managing water storage, such as the Cantareira System and the reservoir of the Furnas hydroelectric power plant. In 2014, a crisis in the water supply of these systems was reported and the proposed charts were able to identify it. In this way, it is confirmed the potential of the proposed control charts to monitor water reservoir levels.

Keywords: Control Charts. Furnas hydropower plant. KARMA model. Outliers. Water reservoir. Cantareira System. Weighted likelihood..

SUMÁRIO

1	INTRODUÇÃO	9
1.1	OBJETIVOS	11
1.1.1	Objetivos específicos	11
1.2	ESTRUTURA DO TRABALHO	12
2	REVISÃO DE LITERATURA	13
2.1	GRÁFICOS DE CONTROLE	13
2.1.1	Gráficos de controle usuais	15
2.1.1.1	Gráfico de Shewhart	16
2.1.1.2	Gráfico CUSUM tabular	17
2.1.2	Gráficos de controle para dados duplamente limitados	19
2.1.3	Gráficos de controle para dados autocorrelacionados	20
3	GRÁFICOS DE CONTROLE KARMA	22
3.1	MODELO KARMA	22
3.2	GRÁFICO DE CONTROLE PROPOSTO	24
3.3	AVALIAÇÃO NUMÉRICA	26
3.3.1	Avaliação do ARL_0	26
3.3.2	Avaliação do ARL_1	28
3.4	RESULTADOS NUMÉRICOS	29
3.5	CONCLUSÃO	33
4	GRÁFICOS DE CONTROLE KARMA ROBUSTOS	38
4.1	GRÁFICOS DE CONTROLE PROPOSTOS	39
4.1.1	Estimadores de máxima verossimilhança ponderada	39
4.2	AVALIAÇÃO NUMÉRICA	42
4.2.1	Avaliação dos estimadores pontuais	43
4.2.1.1	Ponto de ruptura	43
4.2.1.2	Curva de sensibilidade	44
4.2.2	Análise de desempenho dos gráficos de controle	44
4.3	RESULTADOS NUMÉRICOS	45
4.4	CONCLUSÃO	52
5	GRÁFICOS DE CONTROLE KARMA ROBUSTOS APLICADOS À GESTÃO DE RESERVATÓRIOS	56
5.1	RESERVATÓRIOS DE ABASTECIMENTO DE ÁGUA	56
5.2	RESERVATÓRIOS PARA GERAÇÃO DE ENERGIA ELÉTRICA	57
5.3	MONITORAMENTO DE NÍVEL EM RESERVATÓRIOS	58
5.4	ESTUDO DE CASOS	60
5.4.1	Sistema Cantareira	60
5.4.2	Reservatório da usina hidrelétrica de Furnas	65
5.5	CONCLUSÃO	73
6	CONCLUSÕES GERAIS	74
	REFERÊNCIAS	76
	APÊNDICES	87

1 INTRODUÇÃO

A necessidade de compreender os mecanismos e os processos que envolvem as variáveis ambientais em determinadas atividades tem aproximado a ciência ambiental da estatística (MANLY, 2008). A partir de métodos estatísticos é possível realizar inferências, quantificar efeitos, presumir riscos e estabelecer padrões (MANLY, 2008). Essas análises são importantes em diversos estudos de monitoramento ambiental, tais como aquecimento global, níveis do mar, gases de efeito estufa e eventos extremos (WMO, 2016), auxiliando na gestão ambiental.

A gestão ambiental envolve a tomada de decisão, que é caracterizada pela incerteza quanto ao que irá ocorrer, quais serão os impactos e qual a chance de ocorrência (GOUGH; WARD, 1996). Processos hidrológicos, por exemplo, apresentam dependência do clima e de sua variabilidade, do uso e cobertura do solo e da heterogeneidade antrópica, tornando essa etapa ainda mais complexa (VUCETIC; SIMONOVIC, 2011). Frente a isso, destaca-se a importância do uso de boas ferramentas de apoio à tomada de decisão (VUCETIC; SIMONOVIC, 2011).

Uma forma de compreender a dinâmica dos processos hidrológicos pode ser dado por meio de métodos de monitoramento e controle de variáveis (MONTGOMERY, 2009), gerindo impactos oriundos de eventos extremos, como a cheia e a seca (BRAVO, 2010; YU et al., 2010; TONDOLO, 2016; PEREIRA; BATISTA; NEVES, 2018). Esses métodos costumavam ser aplicados majoritariamente no setor industrial, para controlar índices dos produtos ao longo do processo de produção. Entretanto, nos últimos anos, vêm sendo desenvolvidos estudos que utilizam destes princípios em diferentes âmbitos da ciência (GARDINER; MONTGOMERY, 1987; VARKEY; RELLER; RESAR, 2007; VRIES; RENEAU, 2010; EISSA; MAHMOUD; NOUBY, 2016; EISSA, 2019), inclusive na área ambiental (ANDERSON; THOMPSON, 2004; SMETI et al., 2007; SMETI; KORONAKIS; GOLFINOPOULOS, 2007; GOVE et al., 2013; MARCHANT et al., 2019).

Dentre as metodologias para monitoramento de variáveis, uma das mais aplicadas é o controle estatístico de processos (CEP) (MONTGOMERY, 2009). O CEP tem por objetivo, a partir de métodos estatísticos, identificar pontos amostrados que apresentem anomalias (MONTGOMERY, 2009). Dentre a gama de ferramentas disponíveis no CEP, estão os histogramas, gráficos de pareto, diagrama causa-e-efeito, diagrama de concentração de defeito, gráficos de controle, diagrama de dispersão e folha de verificação (MONTGOMERY; RUNGER, 2006). Dessas, destacamos os gráficos de controle, que são comumente utilizados em

função da sua fácil interpretação gráfica visual. Contudo, a aplicação desses gráficos, incluindo o de Shewhart (SHEWHART, 1931) e o de soma cumulativa (CUSUM) (PAGE, 1954), só é válida quando satisfeita a condição de variáveis independentes e identicamente distribuídas (i.i.d.) (WARDELL; MOSKOWITZ; PLANTE, 1992).

Usualmente, as variáveis hidrológicas não seguem o comportamento i.i.d., apresentando um fenômeno de autocorrelação, isto é, há uma tendência de um evento ser dependente do evento anterior (MACHIWAL; JHA, 2012). Esse fenômeno caracteriza a presença de autocorrelação na série, violando a suposição de independência dos dados requerida na aplicação de gráficos de controle usuais. Uma das abordagens utilizadas para situações em que se verifica a presença de autocorrelação na série de dados é dada pelo emprego de gráficos de controle utilizando resíduos de modelos (MONTGOMERY, 2009), usualmente, da classe autorregressivo integrado de médias móveis (ARIMA) (BOX; JENKINS, 1976). No entanto, a metodologia dada a partir de modelos ARIMA pode ser inadequada para diversos tipos de dados, uma vez que usualmente supõe-se normalidade para as inferências (BAYER; CINTRA; CRIBARI-NETO, 2018). Um exemplo disso seria uma série de dados duplamente limitados, tais como dados de volume útil de reservatórios, umidade relativa e energia armazenada, em que ao assumir normalidade os valores preditos poderiam estar fora do intervalo limitado, não fazendo qualquer sentido físico.

No monitoramento de reservatórios de água para abastecimento e energia, uma importante variável é o volume armazenado. Monitorar essa variável auxilia a tomada de decisões, possibilitando a gestão dos usos e redução dos transtornos provocados por eventos extremos. No entanto, o volume armazenado normalmente é dado a partir da relação de nível momentâneo e volume útil, podendo variar no intervalo $(0, 1)$. Alternativas vêm sendo desenvolvidas para suprir a necessidade de monitoramento de dados duplamente limitados (SANT'ANNA; CATEN, 2012; TONDOLO, 2016; BAYER; TONDOLO; MÜLLER, 2018; HO; FERNANDES; BOURGUIGNON, 2018; SILVA, 2019; ARAUJO LIMA FILHO; BAYER, 2019). Frequentemente, quando se trata de dados do tipo taxa e proporção ordenados ao longo do tempo, utiliza-se do modelo beta autorregressivo de médias móveis (β ARMA) (ROCHA; CRIBARI-NETO, 2008). O modelo β ARMA é um modelo proposto para séries continuamente distribuída no intervalo limitado (a, b) , sendo a e b conhecidos, com dinâmicas autorregressivas e de médias móveis (ROCHA; CRIBARI-NETO, 2008). Entretanto, os gráficos baseados no modelo β ARMA podem não ser tão acurados quando a distribuição da variável monitorada não se ajustar bem

a uma distribuição beta. Uma boa alternativa à essa distribuição, para análises hidrológicas, é a distribuição Kumaraswamy (KUMARASWAMY, 1980). Nesse sentido, recentemente foi proposto o modelo Kumaraswamy autorregressivo de médias móveis (KARMA) (BAYER; BAYER; PUMI, 2017) para modelagem de séries temporais ambientais limitadas no intervalo (a, b) , com distribuição Kumaraswamy.

Além da autocorrelação, a presença de *outliers* na fase de estimação dos parâmetros também pode vir a prejudicar o desempenho de gráficos de controle usuais (JONES-FARMER et al., 2014). Uma forma de mitigar os efeitos destas observações discrepantes é dada pelo uso de estimadores de máxima verossimilhança ponderada (FIELD; SMITH, 1994; MANCHINI; PORTELLA; BAYER, 2019). Com esse método pondera-se as observações discrepantes da série de dados estudada para minimizar seu efeito nas inferências sobre os parâmetros dos modelos ajustados.

Com o objetivo de contribuir para o monitoramento de variáveis ambientais e hidrológicas, este trabalho propõe gráficos de controle baseados no modelo KARMA. Ainda, assumindo que os dados possam apresentar *outliers* na fase de estimação, propõe-se uma melhoria aos gráficos de controle KARMA utilizando estimadores de máxima verossimilhança ponderada. Neste trabalho, buscou-se associar a terminologia tradicional da indústria aos conceitos hidrológicos.

1.1 OBJETIVOS

Oferecer bases suficientes para realizar monitoramento de dados ambientais duplamente limitados, autocorrelacionados e com presença de *outliers*. Ainda, mostrar a sua importância e aplicabilidade no monitoramento do nível de reservatórios.

1.1.1 Objetivos específicos

Os objetivos específicos deste trabalho são:

- Propor e avaliar numericamente, via simulação de Monte Carlo, gráficos de controle baseados nos resíduos do modelo KARMA;

- Propor e avaliar numericamente, via simulação de Monte Carlo, gráficos de controle KARMA robustos a *outliers*;
- Avaliar a viabilidade da utilização dos gráficos propostos como ferramentas de apoio na tomada de decisão a partir da aplicação a dados reais do Sistema Cantareira e de Furnas.

1.2 ESTRUTURA DO TRABALHO

Este trabalho está dividido em seis capítulos. O Capítulo 1 apresenta uma introdução ao tema desta dissertação, debatendo de forma objetiva os principais conceitos que nortearam a pesquisa, assim como os objetivos gerais e específicos. No Capítulo 2 é apresentada uma revisão de literatura sobre gráficos de controle. São apresentadas nesse capítulo as principais características dos gráficos de controle que serão utilizados neste estudo, quais sejam: gráficos de Shewhart e CUSUM. No Capítulo 3 são apresentados os gráficos de controle para dados duplamente limitados e autocorrelacionados, utilizando o modelo KARMA. Uma proposta de análise numérica para avaliar o desempenho dos gráficos de controle de Shewhart e CUSUM baseado nos resíduos do modelo KARMA é estruturada nesse capítulo. O Capítulo 4 apresenta uma segunda proposta, gráficos de controle robustos aplicados a séries de dados duplamente limitados, autocorrelacionados e com presença de *outliers*. Para avaliar o desempenho dos gráficos de controle KARMA robustos, apresenta-se ao final do Capítulo 4 uma avaliação numérica através de medidas específicas de robustez e da análise de desempenho. Já no Capítulo 5 é apresentada a aplicação dos gráficos de controle KARMA robustos a variáveis de volume armazenado de dois importantes sistemas de reservação do Brasil, a fim de verificar o desempenho desta ferramenta no monitoramento de dados hidrológicos. Por fim, no Capítulo 6 é apresentada uma conclusão para o estudo realizado nesta dissertação.

2 REVISÃO DE LITERATURA

A revisão de literatura deste trabalho versa acerca dos tópicos: gráficos de controle usuais, gráficos de controle para dados duplamente limitados e gráficos de controle para dados autocorrelacionados.

2.1 GRÁFICOS DE CONTROLE

Os gráficos de controle têm se tornado cada vez mais requeridos no monitoramento e controle de observações, pois, uma vez identificadas anomalias em um processo, é possível avaliar o risco atribuído às mesmas. A primeira proposta de gráficos de controle foi estabelecida por Shewhart (1931), que apresentou uma base estatística para o controle econômico da qualidade de produtos, por meio de limites de controle. Desta forma, tornou-se possível identificar, em todas as etapas do processo de produção, quando o processo variava mais do que seria técnico e economicamente desejável.

Atualmente, a aplicação de gráficos de controle já não se restringe apenas a processos industriais. Diferentes áreas vêm aplicando essa ferramenta no monitoramento das mais diversas variáveis. Na área ambiental, pode-se destacar alguns trabalhos no monitoramento da qualidade da água (NALLY; HART, 1997; SMETI et al., 2007; SMETI; KORONAKIS; GOLFINOPOULOS, 2007; LEE et al., 2013; PAROISSIN et al., 2016), da qualidade do ar (SCHNEIDER; HUI; PRUETT, 1992), no abastecimento de água (GOVE et al., 2013) e na fauna (MESNIL; PETITGAS, 2009; PETITGAS, 2009). O principal objetivo associado a essas aplicações é a identificação de ocorrências de variabilidade não aleatória no processo, ou seja, de anomalias. A partir dessa identificação, a investigação e adoção de ações corretivas podem ser tomadas para evitar que o problema perpetue ou se agrave.

A apresentação dos gráficos de controle é dada por uma série de dados variando ao longo das observações, delimitada por limites de controle (MONTGOMERY, 2009). Esses limites de controle, superior e inferior, controlam a faixa de variação desejada no sistema, ou seja, sempre que as observações estiverem dentro desses limites, o processo é considerado sob controle. Gráficos de controle usuais assumem o pressuposto de que a variável analisada advém de uma população i.i.d., tendo como padrão, a distribuição normal (FILHO, 1982; LAWSON; KLEINMAN, 2005; SMETI et al., 2007; SMETI; KORONAKIS; GOLFINOPOULOS, 2007).

Um gráfico de controle pode ser entendido como um teste de hipótese, no qual as hipóteses a serem testadas são:

$$\begin{cases} \mathcal{H}_0 : \text{processo está sob controle,} \\ \mathcal{H}_1 : \text{processo está fora de controle.} \end{cases}$$

Como em todo teste de hipótese, a rejeição ou a não rejeição da hipótese nula está sujeita a dois tipos de erros, a saber: erro tipo I e erro tipo II (MONTGOMERY, 2009). O erro tipo I é definido quando, sob controle, algum ponto amostral apresentar valor fora dos limites (falsos alarmes), levando a rejeição da hipótese nula, sendo ela verdadeira. Já o erro tipo II ocorre sempre que se assumir um processo sob controle, quando na verdade, ele não está. Assim, de forma intuitiva, identifica-se que limites de controle altos podem aumentar a propensão a erros tipo II, já limites de controle baixos podem aumentar a probabilidade de um erro tipo I (GOVE et al., 2013).

Na aplicação de gráficos de controle, para um bom monitoramento do processo, tem-se a subdivisão em duas fases, a Fase I e a Fase II (MONTGOMERY, 2009). A partir da Fase I, dado um conjunto de dados históricos assumidos sob controle, geram-se as estimativas dos parâmetros do modelo em controle para determinação dos limites (JONES-FARMER et al., 2014). Já na Fase II, tem-se a aplicação desses limites à sequência de observações, com o objetivo de identificar mudanças no processo de controle (WOODALL, 2000). Portanto, o monitoramento realizado na Fase II é baseado na ideia de que o processo de controle estatístico (Fase I) é conhecido. Assim, destaca-se a importância de realizar uma análise cautelosa e precisa dos dados a serem utilizados na Fase I, entendendo o processo de variabilidade da amostra sob controle (JONES-FARMER et al., 2014).

Dada uma série de observações sob controle, na Fase I, os limites são embasados no número médio previsto de observações percorridas até que um ponto seja identificado como fora de controle (MONTGOMERY, 2009). Define-se esse número a partir da média aritmética dos comprimentos de corrida (*run length* - RL) de um gráfico de controle, ou seja, ARL (*average run length*). Ao gráfico é atribuído um ARL proporcional ao erro tipo I assumido (MONTGOMERY, 2009). Assim, tem-se que:

$$ARL = 1/\alpha,$$

em que α é a probabilidade de que qualquer ponto ultrapasse os limites de controle (MONTGOMERY, 2009). Em processo sob controle tem-se o ARL_0 ; já em processos fora de controle tem-se o ARL_1 . Essa probabilidade α pode ser tratada como a proporção de observações fora

dos limites permitidos mesmo em um processo sob controle, ou seja, é a probabilidade do erro tipo I.

No monitoramento de variáveis hidrológicas, o tempo médio de corrida pode ser entendido como o tempo de recorrência de um evento (TR), ou seja, o tempo estimado entre ocorrências de determinado evento hidrológico com iguais magnitudes (CHOW, 1953). Define-se o TR como o inverso da probabilidade de ocorrência do evento, a partir da análise de uma série histórica (CHOW, 1953). O TR é estabelecido para determinação de eventos de chuvas ou de vazões de projeto, importantes no dimensionado de diversas obras, como barragens, sistemas de drenagem, entre outras, assegurando que, em eventos extremos, o sistema não falhe. Da mesma forma, esse conceito pode ser aplicado ao monitoramento e gerenciamento de sistemas envolvendo variáveis hidrológicas, como as de reservação. Nesse caso, assume-se uma probabilidade de ocorrência de erro tipo I (α) à análise estatística igual a probabilidade de ocorrência do evento desejado. O método assumirá que o sistema está sob controle, apesar de existirem observações além dos limites com probabilidade α de ocorrência.

Na seção a seguir são apresentados os gráficos de controle usualmente utilizados, gráfico de Shewhart e gráfico CUSUM, e os gráficos propostos na literatura para monitoramento de dados duplamente limitados e para dados autocorrelacionados.

2.1.1 Gráficos de controle usuais

Existem diversos gráficos de controle aplicados a diferentes comportamentos amostrais. Uma revisão mais completa pode ser encontrada em Montgomery (2009). Dentre esses, dois são amplamente utilizados, o gráfico de Shewhart e o CUSUM. Por muitos anos, foi utilizado apenas o gráfico de Shewhart, no entanto, com a crescente demanda por melhorias nos processos de produção, passaram a ser desenvolvidos outros métodos de controle. O gráfico CUSUM foi proposto como uma alternativa ao gráfico de Shewhart, sugerindo uma melhoria no desempenho. Pode-se destacar que o gráfico de Shewhart é apropriado para a análise de grandes mudanças nos processos, enquanto o gráfico CUSUM para a identificação de pequenas mudanças (SMETI et al., 2007).

Entretanto, independente do método escolhido para análise de controle, para se ter acurácia no monitoramento é necessário um cuidado especial com os limites de controle. Esses devem passar por revisões periódicas para adequada calibração (MONTGOMERY, 2009). Na

sequência é descrito, de forma particular para cada um dos gráficos de controle comentados, como são obtidos os seus limites de controle.

2.1.1.1 Gráfico de Shewhart

A apresentação do gráfico de controle de Shewhart é dada pela plotagem da variável monitorada *versus* o índice da observação, além do valor médio da variável representado por uma linha central (LC). Fixando-se um ARL desejável, para controlar a variação da série de dados, determina-se limites de controle superior (LSC), inferior (LIC) e LC que, de forma geral, podem ser definidos por (MONTGOMERY, 2009):

$$\text{LSC} = \mu + z\sigma,$$

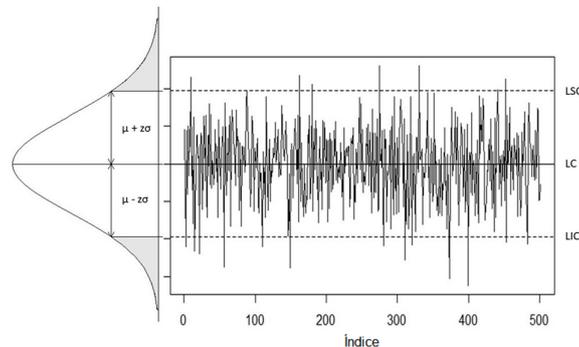
$$\text{LC} = \mu,$$

$$\text{LIC} = \mu - z\sigma,$$

em que μ é a média da variável analisada, σ é o desvio padrão da variável e z o coeficiente que estabelece a porcentagem de valores contidos no intervalo de controle, definido por $z = \Phi^{-1}(1 - \frac{\alpha}{2})$, em que Φ^{-1} é a função quantil da normal padrão e $\alpha = \frac{1}{\text{ARL}}$. A ideia proposta é baseada nas propriedades de uma distribuição de probabilidade normal com média μ e variância σ^2 ($N(\mu, \sigma^2)$). A área sob a curva da função densidade de probabilidade da distribuição normal representa a probabilidade de ocorrência de intervalos de valores, sendo que, quanto mais próximo da média μ , maior será a probabilidade de ocorrência. Portanto, sob distribuição normal, o comportamento da variável tende a ser em torno da média, onde aproximadamente 68% dos valores estão contidos no intervalo $\mu \pm 1\sigma$, 95% no $\mu \pm 2\sigma$ e 99,7% no $\mu \pm 3\sigma$ (FILHO, 1982). Para cada um dos intervalos, há um correspondente ARL, que, pela definição, é estabelecido pelo inverso da probabilidade da observação estar fora desse intervalo. Por exemplo, para uma probabilidade de 0,27% (3σ), o ARL correspondente é de 370, ou seja, uma observação estará fora dos limites de $\pm 3\sigma$ a cada 370 observações. A Figura 1 apresenta a relação entre a função densidade de probabilidade da distribuição normal e os limites de controle para o gráfico de Shewhart.

A média μ e o desvio padrão σ devem ser conhecidos, caso forem desconhecidos, deve-se utilizar as suas estimativas (FILHO, 1982). Borror, Montgomery e Runger (1999) mostraram que a performance dos gráficos de controle de Shewhart é muito sensível a dados que não pos-

Figura 1: Curva densidade de probabilidade da distribuição normal e limites de controle para o gráfico de Shewhart ($X \sim N(0,1)$).



Fonte: Elaborado pela autora.

suem distribuição normal, inibindo a utilidade prática do gráfico. Além disso, esses autores concluíram que quanto mais fortemente assimétrica é a distribuição, mais afetado será o processo.

2.1.1.2 Gráfico CUSUM tabular

O gráfico de controle CUSUM é uma boa alternativa quando o interesse do monitoramento é baseado em pequenas magnitudes de deslocamento da média, menores que $1,5\sigma$ (MONTGOMERY, 2009). Desde a sua proposição, esse gráfico vem sendo estudado por diversos pesquisadores (EWAN, 1963; GOEL; WU, 1971; FELLNER, 1990; MONTGOMERY, 2009). O objetivo principal do CUSUM é aumentar a sensibilidade na identificação de anomalias sem aumentar o erro tipo I, ou seja, falsos alarmes, por meio da soma cumulativa das observações. Uma das formas mais comuns de representar o gráfico CUSUM é por meio do CUSUM tabular, ou também conhecido como algorítmico.

O procedimento realizado no monitoramento de variáveis através do método de CUSUM tabular é dado a partir da acumulação dos desvios em relação à média (alvo) do processo, dada as estatísticas acumuladas acima, C^+ , e abaixo, C^- , do valor alvo (MONTGOMERY, 2009). Essas estatísticas recebem o nome de cusums unilaterais superior e inferior e são obtidas, respectivamente, da seguinte forma (MONTGOMERY, 2009):

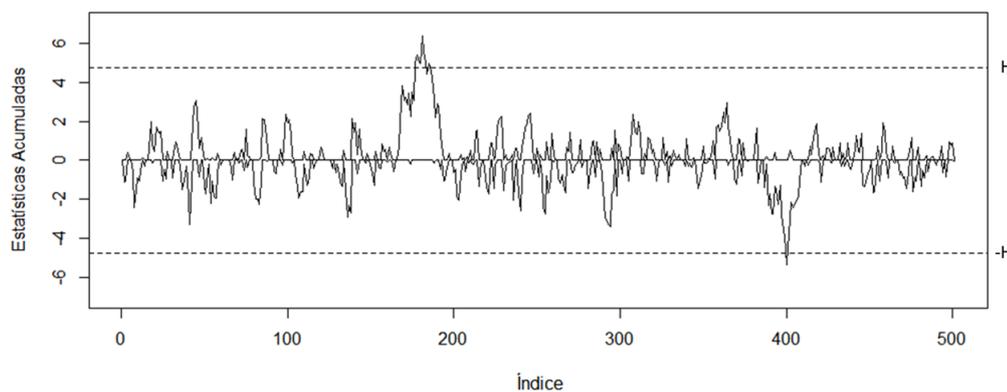
$$C_i^+ = \max[0, x_i - (\mu + K) + C_{i-1}^+],$$

$$C_i^- = \max[0, (\mu - K) - x_i + C_{i-1}^-],$$

para $i \geq 1$, sendo x_i a variável observada no instante i , K o valor de referência (de tolerância ou folga) e $\max[,]$ é a função máximo. A condição inicial de C_0^+ e C_0^- é zero.

C_i^+ e C_i^- acumulam os valores sempre que a relação $x_i - (\mu + K)$ ou $(\mu - K) - x_i$ for maior que zero, sendo possível a identificação de alterações no processo através do aumento do valor acumulado nas estatísticas. Esse valor acumulado das estatísticas é comparado, em cada iteração, com o intervalo de decisão H , pré definido. O processo é considerado fora de controle sempre que C_i^+ e C_i^- excederem H , portanto $LSC=H$ e $LIC=-H$ (MONTGOMERY, 2009). A Figura 2 apresenta um exemplo de gráfico de controle CUSUM tabular, onde o acumulado das estatísticas ultrapassam os limites, representando um processo fora de controle.

Figura 2: Exemplo gráficos de controle CUSUM ($X \sim N(0,1)$).



Fonte: Elaborado pela autora.

A estruturação do CUSUM tabular é dada por meio da escolha do valor de K e de H , definidos por $K = k\sigma$ e $H = h\sigma$ (MONTGOMERY, 2009), em que k e h são valores predefinidos. A recomendação é que atribua-se valores a esses parâmetros de modo a fornecer um ARL adequado ou o desejado (MONTGOMERY, 2009). O parâmetro K é responsável por aumentar a sensibilidade do método de CUSUM a mudanças de uma magnitude alvo. Desta forma, o parâmetro K deve estar entre a média sob controle e a média fora de controle que deseja-se identificar (HAWKINS, 1993). Uma vez que K tenha sido definido, H é ajustado de modo a fixar o ARL (HAWKINS, 1993). Hawkins (1993) apresenta uma relação de valores de k e os correspondentes valores h para um $ARL = 370$, o que coincide com a probabilidade obtida em um gráfico de controle de Shewhart para os habituais limites de 3σ (probabilidade erro tipo I igual a 0,27%). Diferente do método proposto por Shewhart, nos gráficos CUSUM não há uma forma de obtenção de ARL baseado em argumentos teóricos (CINAR; PALAZOGLU;

KAYIHAN, 2007).

2.1.2 Gráficos de controle para dados duplamente limitados

Diante da necessidade de monitorar variáveis com comportamento não gaussianos, adaptações vêm sendo propostas aos gráficos de controle usuais. Ali, Pievatolo e Göbc (2016) apresentam uma revisão com diversos trabalhos que propuseram novas versões aos gráficos de controle usuais para suprir o monitoramento de variáveis com distribuições diferentes da normal, destes, é apresentado apenas uma proposta de gráfico de controle para dados do tipo taxa e proporção.

Assim, no contexto de variáveis duplamente limitadas, pode-se destacar alguns trabalhos com a distribuição beta, como em Sant'Anna e Caten (2012) e em Bayer, Tondolo e Müller (2018), com a distribuição simplex e a gama unitária, como em Ho, Fernandes e Bourguignon (2018), e com distribuição Kumaraswamy, como em Lima Filho e Bayer (2019). Os gráficos propostos por Sant'Anna e ten Caten (2012) tiveram melhores resultados que os gráficos de controle usuais quando aplicados ao monitoramento de dados do tipo fração. Ainda, quando os parâmetros da distribuição beta estão relacionados a variáveis exógenas, pode-se utilizar os gráficos de controle considerando estruturas de regressão, como os propostos por Bayer, Tondolo e Müller (2018). No trabalho de Ho, Fernandes e Bourguignon (2018) é comparado o desempenho dos gráficos de controle para os modelos beta, apresenados em Sant'Anna e ten Caten (2012), com duas novas propostas, a simplex e a gama unitária. Essa comparação mostrou que os três modelos podem ser úteis para monitorar dados do tipo taxas e proporções.

Embora as propostas de adaptações nos gráficos de controle para distribuição beta, simplex e a gama unitária possam ser aplicadas a conjuntos de dados de fração, taxa ou proporção, o desempenho pode ser influenciado quando as variáveis não forem bem ajustadas a essas distribuições. Uma distribuição potencialmente utilizada para modelar dados de fração é a distribuição Kumaraswamy (KUMARASWAMY, 1980). Recentemente, Lima Filho e Bayer (2019) propuseram o gráfico de controle Kumaraswamy. Neste trabalho é evidenciando o melhor desempenho do gráfico proposto frente ao gráfico de controle beta de Sant'Anna e ten Caten (2012) no monitoramento de séries com distribuição Kumaraswamy. A distribuição de probabilidade Kumaraswamy pode acomodar diferentes comportamentos e formatos para densidade, como unimodal, uniantimodal, crescente, decrescente e constante (BAYER; BAYER;

PUMI, 2017). A família de distribuições Kumaraswamy pode ser aplicada para modelar processos aleatórios duplamente limitados, especialmente para aplicações de processos hidrológicos (SUNDAR; SUBBIAH, 1989; GANJI et al., 2006; WANG; WANG; YU, 2017). Pode-se destacar algumas vantagens da distribuição Kumaraswamy sobre a distribuição beta, tais como uma fórmula explícita simples para a sua função de distribuição acumulada e função quantil e, conseqüentemente, uma fórmula simples para geração de variáveis aleatórias e fórmulas mais simples para momentos de estatísticas de ordem (JONES, 2009). Quando se trata de processos hidrológicos, as distribuições clássicas, como por exemplo a distribuição beta, podem não se adequar bem, especialmente em curtos períodos (KUMARASWAMY, 1976), nesses casos, a distribuição Kumaraswamy pode ser considerada uma boa alternativa (NADARAJAH, 2008).

2.1.3 Gráficos de controle para dados autocorrelacionados

O monitoramento de séries temporais exige maiores cuidados quanto ao uso dos gráficos de controle (ALWAN; ROBERTS, 1988). Um dos pressupostos assumidos na aplicação de gráficos de controle usuais é a independência entre as observações, exigindo correlação nula entre as observações passadas e futuras na série de dados (MACHIWAL; JHA, 2012). Essa correlação serial, também chamada de autocorrelação, quando não nula, pode ser tratada como persistência ou memória da série (MACHIWAL; JHA, 2012). É comum encontrar essa característica em séries temporais (CHANDLER; SCOTT, 2011), como por exemplo, em uma série de vazões diárias de um rio. Na presença de autocorrelação, a estimação dos limites de controle dos gráficos tradicionais é prejudicada, aumentando a probabilidade de falsos alarmes (SMETI; KORONAKIS; GOLFINOPOULOS, 2007).

Na composição de uma série temporal admite-se, usualmente, uma componente estocástica dependente, representada por componentes autoregressivos de médias móveis (ARMA) (BOX; JENKINS, 1976), e uma componente independente, dada pelo erro aleatório (MACHIWAL; JHA, 2012). Essa forma de representar uma série temporal pode ser utilizada para dois fins, seja gerar dados sintéticos ou realizar previsões (SALAS, 1985). Gerar séries sintéticas a partir de uma série real possibilita, entre outros, o preenchimento de observações e a obtenção de séries i.i.d.. Portanto, uma forma de remover a autocorrelação de uma série temporal é aplicar modelos ARMA e trabalhar com o resíduo gerado (ALWAN; ROBERTS, 1988; SMETI; KORONAKIS; GOLFINOPOULOS, 2007; CHANDLER; SCOTT, 2011).

Para análises de dados duplamente limitados, a utilização de modelos ARMA pode não se adequar bem, uma vez que estimam parâmetros baseados em uma distribuição normal, podendo estender-se no conjunto de número reais (BAYER; CINTRA; CRIBARI-NETO, 2018). Neste contexto, Tondolo (2016) propôs uma alternativa ao uso destes resíduos, dada por gráficos de controle utilizando o modelo β ARMA (ROCHA; CRIBARI-NETO, 2008). Os gráficos de controle β ARMA (TONDOLO, 2016) apresentaram bom desempenho no monitoramento de processos beta distribuídos autocorrelacionados. No entanto, quando viola-se a hipótese da distribuição assumida no modelo aplicado (seja ARMA, β ARMA ou outro modelo), os gráficos podem ter seu desempenho prejudicado. Assim, introduzido por Bayer, Bayer e Pumi (2017), os modelos KARMA surgem como uma alternativa para monitoramento de variáveis hidrológicas, estimando parâmetros do modelo de séries temporais a partir de uma distribuição Kumaraswamy (KUMARASWAMY, 1980).

3 GRÁFICOS DE CONTROLE KARMA

Neste capítulo é introduzida e avaliada a proposta de gráficos de controle residuais utilizando os resíduos do modelo KARMA. Na seção a seguir é feita uma breve descrição dos modelos KARMA.

3.1 MODELO KARMA

O modelo KARMA utiliza a distribuição Kumaraswamy reparametrizada, modelando a mediana condicional, μ_t , de dados duplamente limitados no intervalo (a, b) , por meio de uma estrutura dinâmica. Assim, dada variável aleatória \tilde{Y} com distribuição Kumaraswamy, com parâmetro de precisão $\varphi > 0$ e mediana condicional $0 < \mu_t < 1$, e \mathcal{F}_{t-1} o conjunto de informações anteriores ao instante t , a densidade de probabilidade condicional $f_{\mu_t}(\tilde{y}_t | \mathcal{F}_{t-1})$ e a função acumulada $F_{\mu_t}(\tilde{y}_t | \mathcal{F}_{t-1})$ são obtidas por (BAYER; BAYER; PUMI, 2017):

$$f_{\mu_t}(\tilde{y}_t | \mathcal{F}_{t-1}) = \left(\frac{1}{b-a} \right) \frac{\varphi \log(0.5)}{\log(1-\mu_t^\varphi)} y_t^{\varphi-1} (1-y_t^\varphi)^{\frac{\log(0.5)}{\log(1-\mu_t^\varphi)}-1}$$

e

$$F_{\mu_t}(\tilde{y}_t | \mathcal{F}_{t-1}) = 1 - (1 - y_t^\varphi)^{\frac{\log(0.5)}{\log(1-\mu_t^\varphi)}},$$

em que $\log(\cdot)$ representa o logaritmo de base e e $y = \frac{(\tilde{y}-a)}{(b-a)}$. A esperança e a variância condicionais são calculadas, respectivamente, por:

$$E(\tilde{y}_t | \mathcal{F}_{t-1}) = a + (b-a) \frac{\log(0.5)}{\log(1-\mu_t^\varphi)} B \left(1 + \frac{1}{\varphi}, \frac{\log(0.5)}{\log(1-\mu_t^\varphi)} \right)$$

e

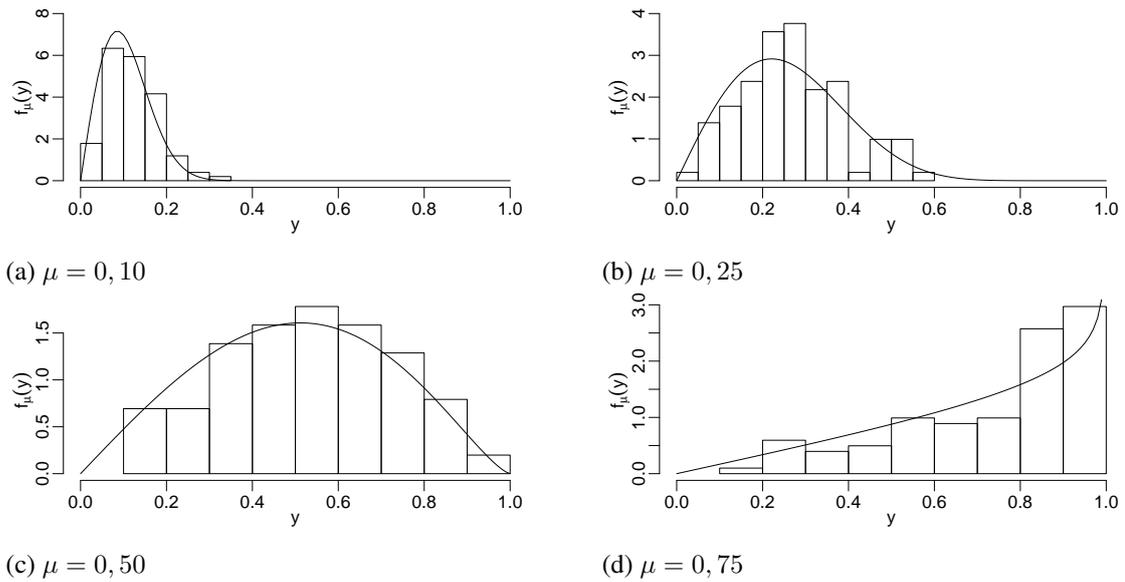
$$\text{Var}(\tilde{y}_t | \mathcal{F}_{t-1}) = \frac{1}{b-a} \left\{ \frac{\log(0.5)}{\log(1-\mu_t^\varphi)} B \left(1 + \frac{2}{\varphi}, \frac{\log(0.5)}{\log(1-\mu_t^\varphi)} \right) - \left[\frac{\log(0.5)}{\log(1-\mu_t^\varphi)} B \left(1 + \frac{1}{\varphi}, \frac{\log(0.5)}{\log(1-\mu_t^\varphi)} \right) \right]^2 \right\}.$$

sendo $B(\cdot)$ a função beta.

Diferentes formas da distribuição Kumaraswamy podem ser visualizadas na Figura 3. Essa distribuição de probabilidade apresenta um comportamento muito próximo da densidade

estimada de algumas variáveis ambientais, principalmente diárias e semanais. Na hidrologia este comportamento se ajustaria bem, entre outros, para os estudos de cheias e estiagens, para a análise de energia armazenada e para a análise de umidade relativa, diferente da distribuição normal que apresenta simetria em relação à média.

Figura 3: Função densidade Kumaraswamy para diferentes valores do parâmetro μ com $\varphi = 2$, $a = 0$ e $b = 1$.



Fonte: Elaborado pela autora.

Considerando um processo KARMA(p,q), a mediana condicional é definida pela seguinte estrutura dinâmica (BAYER; BAYER; PUMI, 2017):

$$\eta_t = g(\mu_t) = \alpha + x_t^\top \boldsymbol{\beta} + \sum_{i=1}^p \phi_i [g(y_{t-i}) - x_{t-i}^\top \boldsymbol{\beta}] + \sum_{j=1}^q \theta_j r_{t-j}, \quad (3.1)$$

em que μ_t é obtida por uma estrutura ARMA, com parâmetros autorregressivos ($\boldsymbol{\phi} = (\phi_1, \dots, \phi_p)^\top$) e de médias móveis ($\boldsymbol{\theta} = (\theta_1, \dots, \theta_q)^\top$), η_t é o preditor linear, $g(\cdot)$ é uma função de ligação, α uma constante, x_t é o vetor de covariáveis no tempo t e $\boldsymbol{\beta} = (\beta_1, \dots, \beta_r)^\top$ é o vetor de parâmetros associado às covariáveis. O erro de médias móveis do modelo é dado por $r_t = g(y_t) - g(\mu_t)$.

Um dos métodos para obter as estimativas dos parâmetros de modelos estatísticos é o de máxima verossimilhança (PAWITAN, 2001). No modelo KARMA, os estimadores são dados pelo valor do vetor paramétrico $\boldsymbol{\gamma} = (\alpha, \boldsymbol{\beta}^\top, \boldsymbol{\phi}^\top, \boldsymbol{\theta}^\top, \varphi)^\top$, com dimensão $(p + q + r + 2)$, que maximiza o logaritmo da função de verossimilhança condicional, dada por (BAYER; BAYER;

PUMI, 2017):

$$\ell = \ell(\gamma; \tilde{y}_t) = \sum_{t=m+1}^n \log(f(\tilde{y}_t | \mathcal{F}_{t-1})) = \sum_{t=m+1}^n \ell_t(\mu_t, \varphi), \quad (3.2)$$

em que

$$\begin{aligned} \ell_t(\mu_t, \varphi) &= \log(\varphi) - \log(b-a) + \log\left(\frac{\log(0.5)}{\log(1-\mu_t^\varphi)}\right) + (\varphi-1) \log(y_t) \\ &\quad + \left(\frac{\log(0.5)}{\log(1-\mu_t^\varphi)} - 1\right) \log(1-y_t^\varphi). \end{aligned}$$

A partir das derivadas parciais da função de log-verossimilhança condicional em relação a cada parâmetro do vetor γ , obtém-se o vetor escore $\mathbf{U}(\gamma)$:

$$\mathbf{U}(\gamma) = (\mathbf{U}_\alpha(\gamma), \mathbf{U}_\beta(\gamma)^\top, \mathbf{U}_\phi(\gamma)^\top, \mathbf{U}_\theta(\gamma)^\top, \mathbf{U}_\varphi(\gamma))^\top.$$

Detalhes e expressões matriciais fechadas para o vetor escore podem ser obtidos em Bayer, Bayer e Pumi (2017). O estimador de máxima verossimilhança condicional de γ é dado para a condição de $\mathbf{U}(\gamma) = \mathbf{0}$, em que $\mathbf{0}$ é um vetor nulo de dimensão $p + q + r + 2$. A solução desse sistema não possui forma fechada, portanto as estimativas de máxima verossimilhança condicional são obtidas utilizando o algoritmo de otimização quasi-Newton, conhecido como método Broyden-Fletcher-Goldfarb-Shanno (BFGS) (PRESS et al., 1992), utilizado em Bayer, Bayer e Pumi (2017).

3.2 GRÁFICO DE CONTROLE PROPOSTO

Em um processo sob controle, dadas as estimativas dos parâmetros, faz-se útil a análise de resíduos do modelo. Os resíduos representam o desvio existente entre a série estimada e a observada, sendo importante na avaliação da qualidade do modelo ajustado (KEDEM; FOKIANOS, 2005). Dessa forma, é condição que os resíduos não apresentem autocorrelação e possuam média aproximadamente zero e variância constante (KEDEM; FOKIANOS, 2005). Há diferentes formas de se obter o resíduo de um modelo. Pode-se destacar o resíduo padronizado, o resíduo *deviance* e o resíduo quantílico. Apesar de ser usual na avaliação de modelos da classe ARIMA, o resíduo padronizado não se mostra adequado para o modelo KARMA, uma vez que é definido por:

$$r_t^{(p)} = \frac{y_t - \hat{\mu}_t}{\sqrt{\widehat{Var}(y_t)}}, \quad (3.3)$$

sendo $\hat{\mu}_t$ a estimativa de μ_t . Para o modelo KARMA, a distribuição residual não possuirá média zero, pois, μ é a mediana condicional e não a média, logo, $E(y_t - \mu_t) \neq 0$. Já o resíduo *deviance* é dado, por:

$$r_t^{(d)} = \text{sign}(y_t - \hat{\mu}_t) 2(\ell_t(y_t, \hat{\varphi}) - \ell_t(\hat{\mu}_t, \hat{\varphi}))^{1/2}, \quad (3.4)$$

em que $\hat{\varphi}$ representa a estimativa de φ , e $\text{sign}(\cdot)$ é a função sinal. Este resíduo é muito utilizado em modelos não-normais (MCCULLAGH; NELDER, 1989), como KARMA e β ARMA. Por fim, o resíduo quantílico, considerado em Bayer, Bayer e Pumi (2017) para o modelo KARMA, é definido por:

$$r_t^{(q)} = \Phi^{-1}(F_{\mu_t}(\tilde{y}_t | \mathcal{F}_{t-1})), \quad (3.5)$$

sendo Φ^{-1} a função quantil da normal padrão. O resíduo quantílico, quando utilizado para o modelo ARMA, é equivalente ao padronizado dado que, sob normalidade, o resíduo quantílico é estabelecido por $\Phi^{-1}(F(\tilde{y}_t | \mu, \sigma^2))$, sendo $F(\tilde{y}_t | \mu, \sigma^2)$ a função acumulada da distribuição normal.

Esses resíduos podem ser utilizados para gerar os gráficos de controle, como Shewhart e CUSUM, uma vez que atendem as premissas necessárias. Assim, para cada série de resíduo, r_t , é gerado um gráfico de controle, ou seja, dos resíduos, $r_t^{(p)}$, $r_t^{(d)}$ e $r_t^{(q)}$, são gerados três gráficos de controle diferentes.

Nos gráficos de controle baseados em resíduos de modelos, os limites de controle também devem ser obtidos a partir dessa nova característica da variável monitorada, que passa a ser a sequência residual i.i.d., com média zero e variância constante. Assim, dada uma série de resíduos r_t , com $t = 1, 2, \dots, n$, a média da variável passa a ser representada por \bar{r} e o desvio padrão por σ_r . Os limites do gráfico de Shewhart passam a ser determinados por:

$$LSC = z,$$

$$LIC = -z.$$

Portanto, sempre que $r_t > LSC$ ou $r_t < LIC$, o processo estará fora de controle. Contudo, a literatura diz que se o modelo estiver corretamente ajustado, os resíduos terão uma distribuição aproximadamente normal com $\bar{r} = 0$ e $\sigma_r = 1$ (KEDEM; FOKIANOS, 2005; ESPINHEIRA; FERRARI; CRIBARI-NETO, 2008; BAYER; BAYER; PUMI, 2017), assim $LSC = z$ e $LIC = -z$.

Para o gráfico CUSUM, as estatísticas cusums passam a ser estimadas por:

$$C_i^+ = \max[0, r_t - K + C_{i-1}^+],$$

$$C_i^- = \max[0, -r_t - K + C_{i-1}^-].$$

Da mesma forma que para o CUSUM usual, o processo estará fora de controle, sempre que C_i^+ e C_i^- excederem H .

3.3 AVALIAÇÃO NUMÉRICA

Foi realizada uma avaliação numérica do desempenho do gráfico de controle KARMA via simulação de Monte Carlo. Nessa avaliação também foram considerados os gráficos de controle propostos por Tondolo (2016), por utilizar distribuição dos resíduos β ARMA, e os gráficos de controle com resíduos de modelos da classe ARIMA, por serem amplamente utilizados. Utilizou-se do programa computacional R (R Core Team, 2019), sendo esse um software livre. As funções do R utilizadas estão disponíveis no pacote *qcc* (SCRUCCA; SNOW; BLOOMFIELD, 2017) e *forecast* (HYNDMAN et al., 2019), além das implementações utilizadas no trabalho de Bayer, Bayer e Pumi (2017) sobre os modelos KARMA e β ARMA.

Na análise numérica gerou-se $R = 10.000$ réplicas da série de dados a partir de um processo KARMA (p, q) no intervalo $(0, 1)$, para cada um dos três cenários apresentados na Tabela 3.1. Estes cenários buscam representar comportamentos amostrais diferentes, com presença de simetria e assimetria, para diferentes medianas e precisões. A função de ligação utilizada foi a *logit*, ou seja, $g(\mu) = \log\left(\frac{\mu}{1-\mu}\right)$. Outros cenários também foram avaliados, mas, por brevidade e semelhança de resultados, não serão apresentados.

A medida de desempenho utilizada na avaliação numérica dos gráficos de controle é o ARL. O ARL de um gráfico de controle pode ser dado de duas formas: quando o processo está sob controle, através do ARL_0 , e quando o processo está fora de controle, por meio do ARL_1 .

3.3.1 Avaliação do ARL_0

Na avaliação do ARL_0 , o objetivo é entender a variabilidade do processo. A avaliação é composta por duas fases: na Fase I são obtidas as estimativas dos parâmetros para o modelo

Tabela 3.1: Cenários para avaliação do desempenho dos gráficos de controle KARMA.

Cenários	Modelo	Parâmetros	μ
Cenário 1	KARMA(1,1)	$\alpha = -1$	$\simeq 0,35$
		$\phi_1 = -0,7$	
		$\theta_1 = -0,5$	
		$\varphi = 10$	
Cenário 2	KARMA(1,1)	$\alpha = 0,5$	$\simeq 0,55$
		$\phi_1 = -0,7$	
		$\theta_1 = 0,3$	
		$\varphi = 5$	
Cenário 3	KARMA(1,1)	$\alpha = -1,5$	$\simeq 0,25$
		$\phi_1 = -0,4$	
		$\theta_1 = -0,5$	
		$\varphi = 10$	

Fonte: Elaborado pela autora.

desejado a partir de uma série de dados, enquanto que na Fase II são obtidos os resíduos e o gráfico de controle é aplicado. A fim de verificar a influência do tamanho amostral na Fase I, foram considerados dois tamanhos diferentes, $n_1 = 300$ e 500 . Já para avaliação da Fase II foi utilizado um tamanho amostral de $n_2 = 5.000$. A sequência de passos apresentada a seguir detalha, a partir de uma série de dados, como serão executadas as Fases I e II para avaliação dos gráficos de controle baseados no modelo KARMA:

1. Gera-se uma série sob controle com tamanho amostral $n = n_1 + n_2$ para cada um dos cenários da Tabela 3.1;
2. Estima-se os parâmetros do modelo KARMA para a série de dados composta pelas primeiras n_1 observações da série obtida no passo 1 (Fase I);
3. Utilizando as estimativas do passo 2 e as n_2 observações restantes geradas no passo 1, gera-se os resíduos do modelo (descritos nas Equações 3.3, 3.4 e 3.5) (Fase II);
4. Aplica-se os gráficos de controle à série de resíduos do passo 3 e registra-se o RL (Fase II);
5. Repete-se os passos 1 a 4 para cada uma das 10.000 réplicas de Monte Carlo;
6. Calcula-se o ARL empírico, por meio da média aritmética dos RL (registrados no passo 4) das 10.000 réplicas.

Os gráficos de controle foram avaliados considerando $z = 3$, para Shewhart, e $h = 4,77$ e $k = 0,5$, para CUSUM, valores recomendados na bibliografia para um $ARL_0 = 370$ (MONTGOMERY, 2009). Dessa forma, espera-se que o ARL_0 empírico, obtido no passo 6 seja $ARL_0 \approx 370$. O mesmo procedimento foi realizado para avaliação de desempenho dos gráficos de controle β ARMA e ARMA, em substituição ao modelo KARMA, no passo 2.

3.3.2 Avaliação do ARL_1

Dado que, a partir de uma série sob controle tenha sido obtida as estimativas do modelo (Fase I do Seção 3.3.1), a avaliação do ARL_1 é a Fase II aplicada a uma série fora de controle. O ARL_1 identifica o número médio de observações percorridas até que um alarme seja dado em um processo fora de controle. Contudo, para fins de comparação de ARL_1 , é necessário que o ARL_0 empírico dos gráficos avaliados sejam próximos, a fim de evitar a presença de distorções nos gráficos.

Para a avaliação numérica de ARL_1 , seguiu-se a metodologia descrita na Seção 3.3.1, no entanto, a série com tamanho n_2 foi gerada fora de controle. A série sintética fora de controle foi obtida adicionando à relação da mediana (Equação 3.1) uma mudança de nível (δ), como apresentado a seguir:

$$\eta_t = g(\mu_t) = \alpha + x_t^T \boldsymbol{\beta} + \sum_{i=1}^p \phi_i [g(y_{t-i}) - x_{t-i}^T \boldsymbol{\beta}] + \sum_{j=1}^q \theta_j r_{t-j} + \delta. \quad (3.6)$$

As mudanças de níveis consideradas foram $\delta = -0,8, -0,4, 0,4, 0,8$.

A fim de garantir a comparação de ARL_1 para os mesmos ARL_0 dos gráficos, utilizou-se da metodologia de Curva ROC (PETERSON; BIRDSALL; FOX, 1954). As curvas são utilizadas para apresentar a taxa de acertos dada uma taxa de falsos alarmes de classificadores. Assim, a escolha do melhor gráfico de controle residual pode ser dada a partir da relação entre a taxa de acertos e a taxa de falsos alarmes. A representação das taxas foi estabelecida nos gráficos pelo inverso do ARL_0 para taxa de falsos alarmes e o inverso do ARL_1 para taxa de acertos. Assim, para uma mesma taxa de falsos alarmes, destaca-se o gráfico com maior taxa de acertos, ou seja, menor ARL_1 .

3.4 RESULTADOS NUMÉRICOS

Esta seção apresenta os resultados obtidos na avaliação numérica de gráficos de controle KARMA e, comparativamente, β ARMA e ARMA, dadas a partir de simulações de Monte Carlo. A avaliação numérica destes gráficos foi dividida em duas etapas, avaliação do ARL_0 , seguido da avaliação do ARL_1 . Considerou-se o resíduo *deviance* e quantílico para o modelo KARMA e β ARMA e o resíduo padronizado na avaliação do modelo ARMA. Nas Tabelas 3.2, 3.3 e 3.4 são apresentados os resultados da análise numérica de ARL_0 , para os três cenários avaliados, nos dois diferentes tamanhos amostrais e para os resíduos considerados em cada modelo, em que os valores em negritos representam os melhores resultados. A análise de desempenho é dada pela comparação do ARL_0 numérico com o ARL_0 nominal igual a 370. O gráfico de controle com menor distorção, ou seja, com ARL_0 empírico mais próximo de 370, é o de melhor desempenho.

A partir da análise dos resultados apresentados nas Tabelas 3.2, 3.3 e 3.4 foi possível observar um desempenho superior dos gráficos de controle KARMA baseados no resíduo quantílico quando comparados com os gráficos KARMA baseados no resíduo *deviance*. Essa superioridade pode ser observada, principalmente, nos gráficos de controle KARMA CUSUM, como por exemplo, no Cenário 1 (Tabela 3.2), para um tamanho amostral de 500 observações, o gráfico referente ao resíduo *deviance* apresentou $ARL_0 = 87,8699$, enquanto que o gráfico baseado no resíduo quantílico o ARL_0 foi muito próximo do nominal ($ARL_0 = 368,3786$).

Foi possível observar que os gráficos de controle KARMA Shewhart e CUSUM baseados no resíduo quantílico apresentaram melhor desempenho quando comparados com os gráficos β ARMA e ARMA. Esse resultado justifica a importância da proposição de gráficos de controle utilizando o modelo KARMA para monitoramento de dados duplamente limitados, autocorrelacionados e com distribuição Kumaraswamy. Foram identificados cenários em que os gráficos β ARMA e ARMA apresentaram distorções de ARL_0 maiores que 60%, quando comparado com o ARL_0 nominal, como no Cenário 3 (Tabela 3.4) para os gráficos de controle ARMA Shewhart no tamanho amostral 500 em que $ARL_0 = 141,9536$ e nos gráficos de controle β ARMA Shewhart, para ambos resíduos e tamanhos amostrais, apresentando $ARL_0 < 110$. Por outro lado, nos gráficos KARMA quantílico nenhuma distorção superou este percentual, sendo a maior distorção apresentada inferior a 8% no Cenário 2 (Tabela 3.3) para o tamanho amostral 300 em que o gráfico de controle KARMA Shewhart quantílico obteve $ARL_0 = 340,9648$. A distorção de ARL_0 causa um aumento de detecção de observações

fora de controle, quando na verdade, estariam dentro de controle se bem ajustados os limites, ou seja, aumenta a probabilidade de falsos alarmes. Isso pode ser um problema para a gestão ambiental, uma vez que ações equivocadas poderiam ser tomadas. Um exemplo seria no monitoramento de reservatórios de água, dado falso alarme, medidas de gestão seriam tomadas para evitar uma crise no sistema. Assim, comportas poderiam ser abertas para escoamento da vazão ou restrições do uso da água poderiam ser aplicados sem necessidade, acarretando em perdas de produção.

De forma geral, os gráficos de controle KARMA com o resíduo quantílico apresentaram desempenho superior se comparados com o desempenho dos gráficos β ARMA e ARMA, ou seja, comparando o melhor resíduo para os gráficos KARMA com o melhor para os β ARMA e ARMA, o desempenho dos gráficos de controle KARMA é sempre superior. Portanto, a partir do resíduo quantílico é possível obter um comportamento dos gráficos de controle KARMA com menores distorções de ARL_0 , apresentando uma detecção de observações fora de controle mais próxima da que seria a real.

Quando comparado o desempenho dos gráficos de controle de Shewhart e CUSUM, foi possível verificar, de forma geral, uma tendência a redução da distorção do ARL_0 a medida que o tamanho amostral é aumentado. Ainda, identificou-se que, no geral, os gráficos CUSUM obtiveram melhor desempenho quando comparados aos gráficos de Shewhart, tanto para o modelo KARMA quanto para β ARMA e ARMA. De forma geral, fica evidente a superioridade do desempenho dos gráficos KARMA quantílico frente aos gráficos β ARMA e ARMA, em quase todos os cenários avaliados.

Tabela 3.2: Avaliação do ARL_0 para os gráficos de controle residuais de Shewhart e CUSUM no Cenário 1.

GC Shewhart			GC CUSUM		
Tamanho amostral 300					
Modelo	Resíduo	ARL_0	Modelo	Resíduo	ARL_0
KARMA(1, 1)	$r_t^{(d)}$	216,7693	KARMA(1, 1)	$r_t^{(d)}$	92,8248
	$r_t^{(q)}$	344,3415		$r_t^{(q)}$	384,1264
β ARMA(1, 1)	$r_t^{(d)}$	110,7783	β ARMA(1, 1)	$r_t^{(d)}$	246,9043
	$r_t^{(q)}$	108,4325		$r_t^{(q)}$	238,8665
ARMA(1, 1)	$r_t^{(p)}$	132,9791	ARMA(1, 1)	$r_t^{(p)}$	189,4188
Tamanho amostral 500					
Modelo	Resíduo	ARL_0	Modelo	Resíduo	ARL_0
KARMA(1, 1)	$r_t^{(d)}$	225,0013	KARMA(1, 1)	$r_t^{(d)}$	87,8699
	$r_t^{(q)}$	354,6760		$r_t^{(q)}$	368,3786
β ARMA(1, 1)	$r_t^{(d)}$	110,7768	β ARMA(1, 1)	$r_t^{(d)}$	243,0383
	$r_t^{(q)}$	108,5236		$r_t^{(q)}$	236,3413
ARMA(1, 1)	$r_t^{(p)}$	132,6631	ARMA(1, 1)	$r_t^{(p)}$	192,0291

Fonte: Elaborado pela autora.

Tabela 3.3: Avaliação do ARL_0 para os gráficos de controle residuais de Shewhart e CUSUM no Cenário 2.

GC Shewhart			GC CUSUM		
Tamanho amostral 300					
Modelo	Resíduo	ARL_0	Modelo	Resíduo	ARL_0
KARMA(1, 1)	$r_t^{(d)}$	218,9797	KARMA(1, 1)	$r_t^{(d)}$	91,8169
	$r_t^{(a)}$	340,9648		$r_t^{(a)}$	353,1031
β ARMA(1, 1)	$r_t^{(d)}$	221,2747	β ARMA(1, 1)	$r_t^{(d)}$	356,1937
	$r_t^{(a)}$	189,7643		$r_t^{(a)}$	317,9881
ARMA(1, 1)	$r_t^{(p)}$	290,3627	ARMA(1, 1)	$r_t^{(p)}$	342,0886
Tamanho amostral 500					
Modelo	Resíduo	ARL_0	Modelo	Resíduo	ARL_0
KARMA(1, 1)	$r_t^{(d)}$	225,8161	KARMA(1, 1)	$r_t^{(d)}$	87,6707
	$r_t^{(a)}$	351,5924		$r_t^{(a)}$	349,4611
β ARMA(1, 1)	$r_t^{(d)}$	221,7890	β ARMA(1, 1)	$r_t^{(d)}$	347,2678
	$r_t^{(a)}$	192,5443		$r_t^{(a)}$	309,8841
ARMA(1, 1)	$r_t^{(p)}$	282,4062	ARMA(1, 1)	$r_t^{(p)}$	330,9654

Fonte: Elaborado pela autora.

Tabela 3.4: Avaliação do ARL_0 para os gráficos de controle residuais de Shewhart e CUSUM no Cenário 3.

GC Shewhart			GC CUSUM		
Tamanho amostral 300					
Modelo	Resíduo	ARL_0	Modelo	Resíduo	ARL_0
KARMA(1, 1)	$r_t^{(d)}$	215,5792	KARMA(1, 1)	$r_t^{(d)}$	93,2710
	$r_t^{(a)}$	342,9300		$r_t^{(a)}$	381,4096
β ARMA(1, 1)	$r_t^{(d)}$	109,1173	β ARMA(1, 1)	$r_t^{(d)}$	253,3979
	$r_t^{(a)}$	105,9174		$r_t^{(a)}$	244,4084
ARMA(1, 1)	$r_t^{(p)}$	144,1035	ARMA(1, 1)	$r_t^{(p)}$	237,5069
Tamanho amostral 500					
Modelo	Resíduo	ARL_0	Modelo	Resíduo	ARL_0
KARMA(1, 1)	$r_t^{(d)}$	223,6458	KARMA(1, 1)	$r_t^{(d)}$	87,8313
	$r_t^{(a)}$	353,6704		$r_t^{(a)}$	366,9988
β ARMA(1, 1)	$r_t^{(d)}$	109,5441	β ARMA(1, 1)	$r_t^{(d)}$	251,4254
	$r_t^{(a)}$	106,7254		$r_t^{(a)}$	241,8567
ARMA(1, 1)	$r_t^{(p)}$	141,9536	ARMA(1, 1)	$r_t^{(p)}$	237,1055

Fonte: Elaborado pela autora.

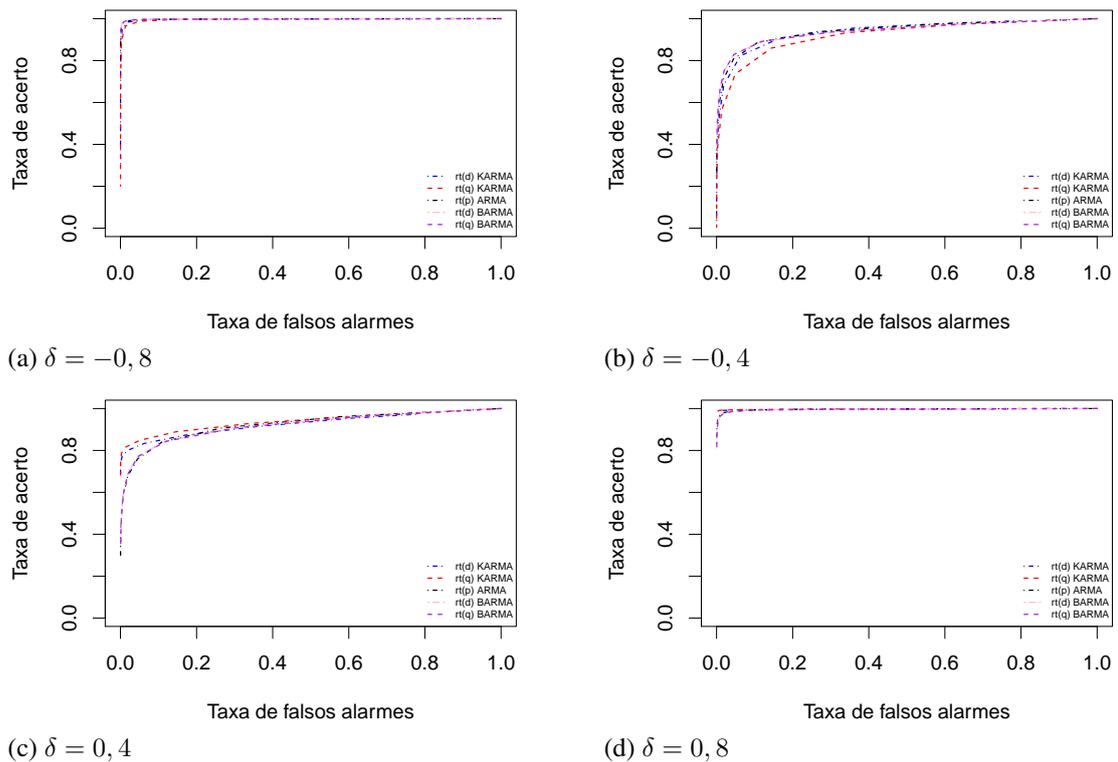
A avaliação do ARL_1 se deu por meio da metodologia de curva ROC. As Figuras 4, 5 e 6 apresentam o desempenho dos gráficos de controle de Shewhart e as Figuras 7, 8 e 9 o desempenho dos gráficos de controle CUSUM frente às diferentes magnitudes avaliadas para $n_1 = 300$ na Fase I.

Na avaliação de desempenho dos gráficos de controle, foi possível observar uma proximidade nos resultados das taxas de acertos. Destaca-se um melhor desempenho dos gráficos de controle KARMA para as mudanças de nível positivas ($\delta = 0,4$ e $0,8$) em todos os cenários. Além disso, no Cenário 2 (Figuras 5 e 8) também é verificada uma superioridade no desempenho dos gráficos KARMA para as mudanças de nível negativas ($\delta = -0,4$ e $-0,8$),

principalmente para o resíduo deviance.

Para os $\delta = -0,8$ e $0,8$ foi possível identificar um bom desempenho em todos os gráficos de controle, chegando a taxas de acerto próximas de 100% de forma mais acelerada que para as magnitudes medianas ($\delta = -0,4$ e $0,4$). Por exemplo, nos Cenários 1 e 3, para $\delta = \pm 0,8$, todos os gráficos alcançaram taxas de acerto igual a 100% para taxas de falsos alarmes inferior a 20%. Já para magnitudes de $\delta = \pm 0,4$, taxas de acerto próximas a esta foram alcançadas apenas para taxas de falsos alarmes próximos a 80%. Os resultados para Shewhart e CUSUM foram muito próximos, não apresentando diferenças na comparação de desempenho dos gráficos. Na comparação de desempenho dos gráficos quanto aos dois diferentes tamanhos amostrais avaliados, também identificou-se similaridade, no Apêndice A são apresentados os gráficos para o tamanho amostral $n_1 = 500$.

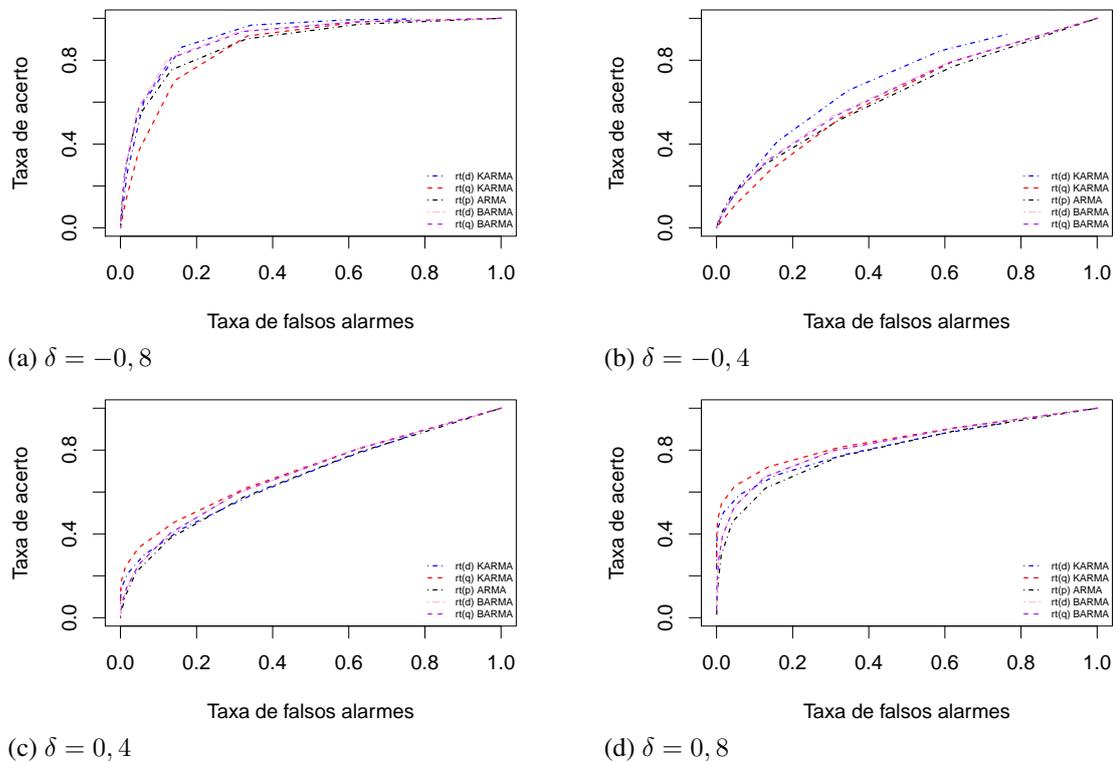
Figura 4: Curva ROC gráfico de controle de Shewhart - Cenário 1.



Fonte: Elaborado pela autora.

No geral, foi possível observar um bom desempenho dos gráficos de controle na identificação das mudanças de nível nos cenários avaliados. Verificou-se que os gráficos de controle KARMA apresentaram taxas de acerto sempre muito próximo ou superior aos demais.

Figura 5: Curva ROC gráfico de controle de Shewhart - Cenário 2.

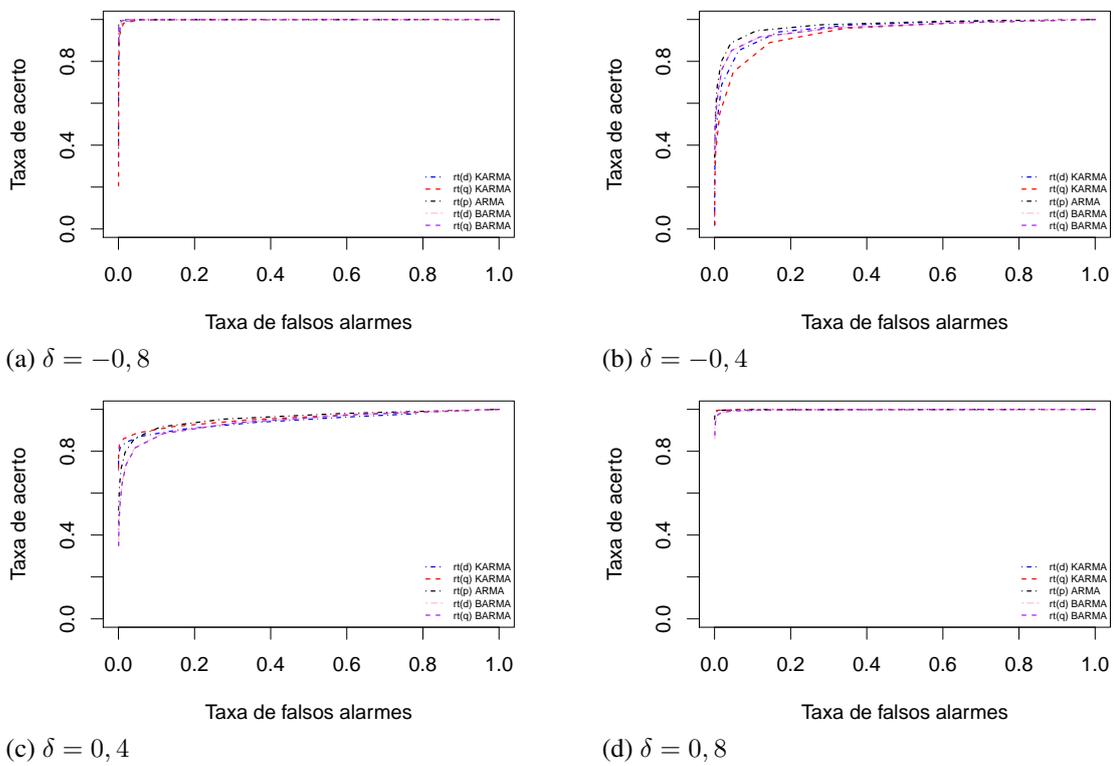


Fonte: Elaborado pela autora.

3.5 CONCLUSÃO

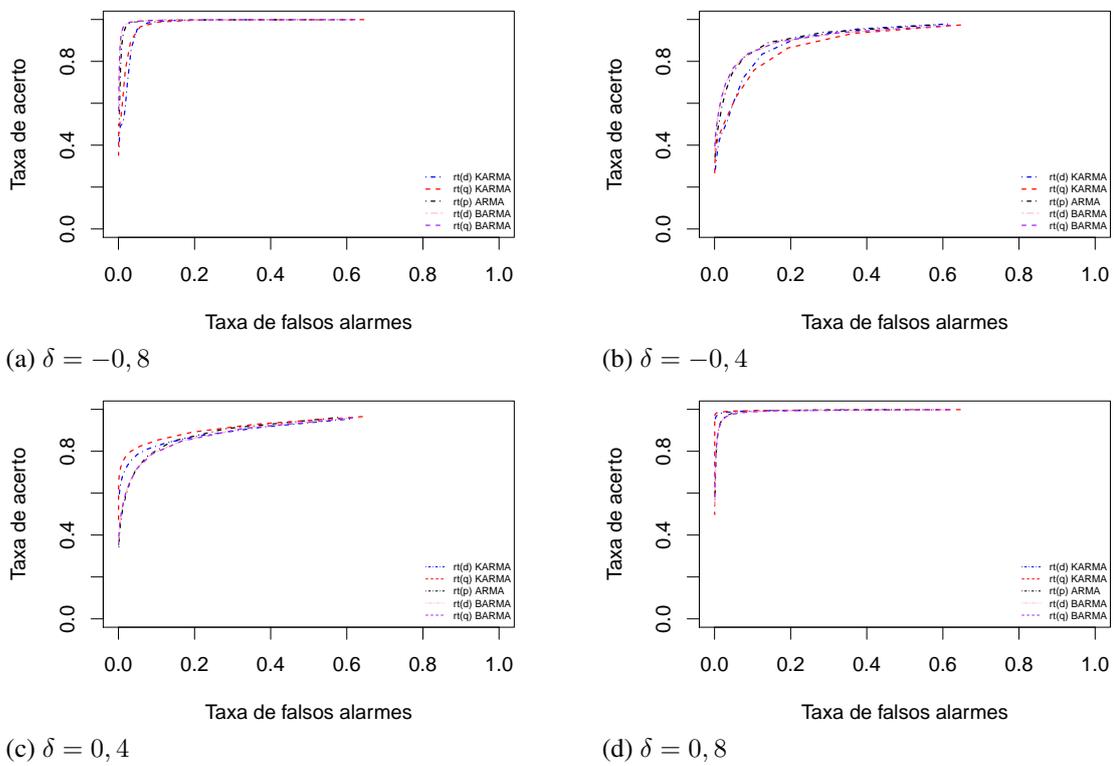
A partir da avaliação de ARL_0 e ARL_1 , foi possível evidenciar o melhor desempenho dos gráficos de Shewhart e CUSUM baseados no modelo KARMA, com destaque aos gráficos de controle CUSUM KARMA quantílico. Os gráficos de controle KARMA apresentaram pequenas distorções na avaliação de ARL_0 , enquanto os outros se mostram muito distorcidos. Na avaliação de ARL_1 os desempenhos são equiparáveis, com algum superioridade para o modelo KARMA. Assim, dadas avaliações de desempenho dos gráficos, fica constatado que os gráficos de controle KARMA podem ser uma boa alternativa no monitoramento de séries com comportamento próximo ao da distribuição Kumaraswamy. A utilização dos gráficos de controle KARMA garante a menor distorção de ARL_0 com altas taxas de acerto e, portanto, um monitoramento mais fidedigno a realidade.

Figura 6: Curva ROC gráfico de controle de Shewhart - Cenário 3.



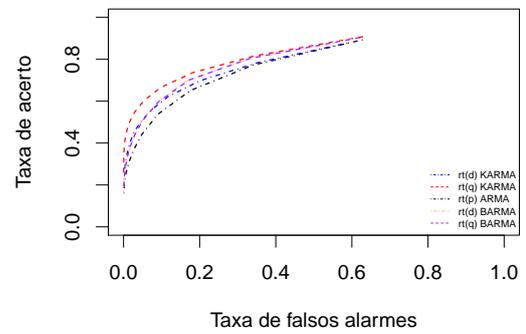
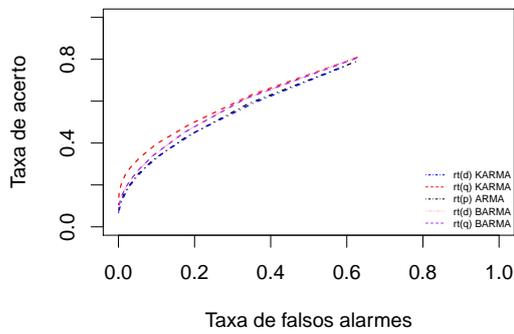
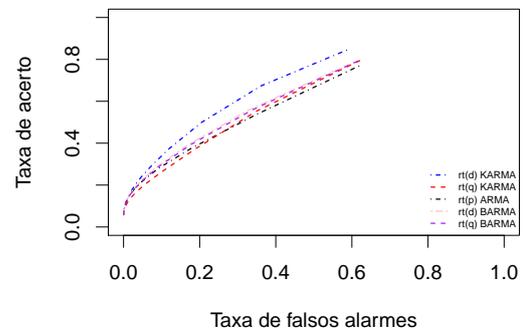
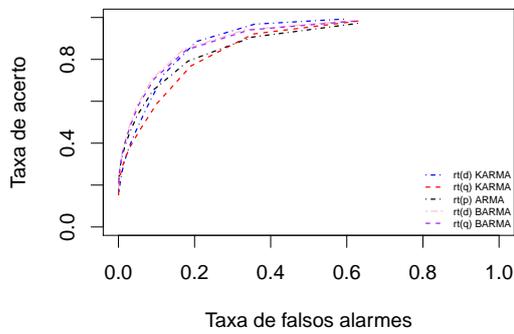
Fonte: Elaborado pela autora.

Figura 7: Curva ROC gráfico de controle CUSUM - Cenário 1.



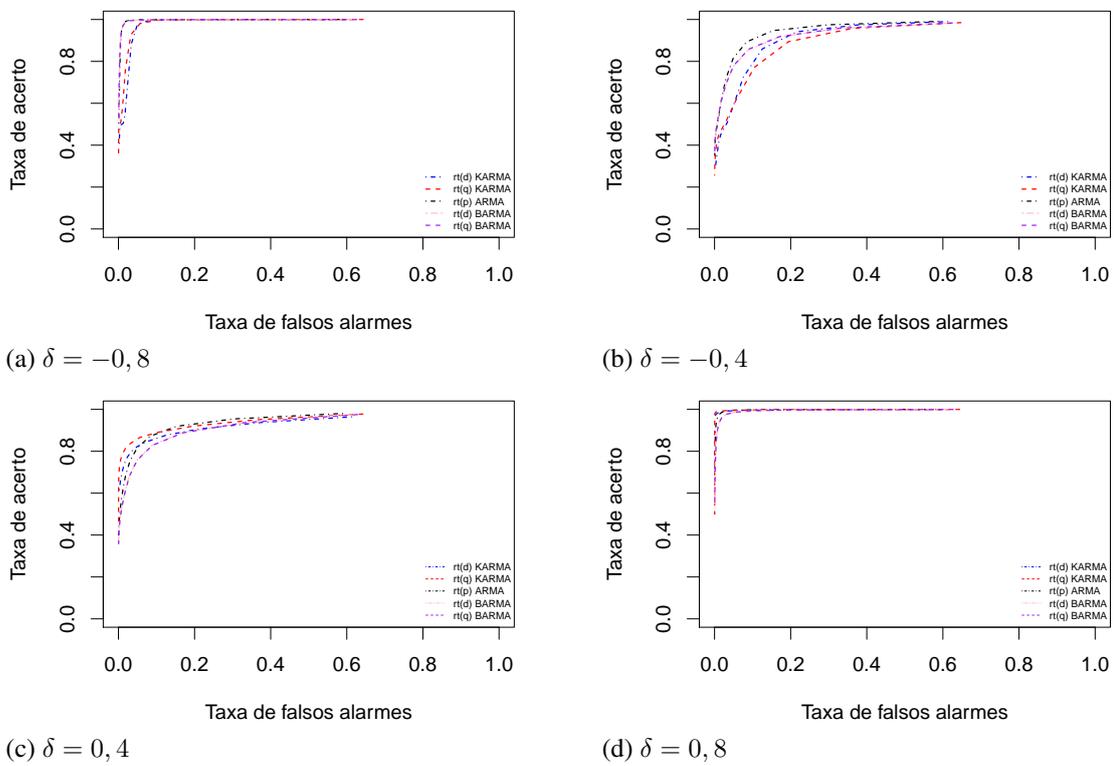
Fonte: Elaborado pela autora.

Figura 8: Curva ROC gráfico de controle CUSUM - Cenário 2.



Fonte: Elaborado pela autora.

Figura 9: Curva ROC gráfico de controle CUSUM - Cenário 3.



Fonte: Elaborado pela autora.

4 GRÁFICOS DE CONTROLE KARMA ROBUSTOS

Na análise de séries temporais, além da autocorrelação, é comum observar outros fenômenos, tais como sazonalidade, ciclos, tendência e *outliers* (observações discrepantes). A presença de *outliers* pode influenciar nas inferências sobre os parâmetros dos modelos ajustados à série de dados, e, conseqüentemente, na estimação e na previsão (FOX, 1972; ABRAHAM; BOX, 1979; DENBY; MARTIN, 1979; HILLMER, 1984; CHEN; LIU, 1993; BIANCO et al., 2001). *Outliers* podem ser definidos como, em um conjunto de dados, as observações que assumirem valores destoantes das demais (HYNDMAN; ATHANASOPOULOS, 2018). A origem dessas observações pode-se dar por erros grosseiros de catalogação de dados, falha de equipamento ou por influência de algum evento extremo (CHEN; LIU, 1993). Na área hidrológica, pode-se destacar os eventos extremos, como picos de cheia, secas ou chuvas máximas.

De acordo com sua estrutura temporal, os *outliers* podem aparecer de forma isolada, irregular, duplicada, ou ainda, causando mudanças de nível no valor médio (MARONNA; MARTIN; YOHAI, 2006). Na presença de valores discrepantes, a distribuição empírica dos dados apresenta caudas longas, uma vez que a probabilidade de ocorrência destes valores é baixa. Assim, ao ajustar uma distribuição teórica a uma série de dados contendo *outliers*, as estimativas geradas a partir dos estimadores de máxima verossimilhança passam a não ser mais acurados, mascaradas pela influência desses valores discrepantes, podendo apresentar rendimento estatístico com baixa aceitabilidade (MARONNA; MARTIN; YOHAI, 2006).

Uma alternativa para controlar a influência dos *outliers* na estimação pode ser dada com a retirada dessas observações discrepantes (JONES-FARMER et al., 2014). Entretanto, a presença de *outliers* em séries temporais requer maiores cuidados por apresentar dependência temporal (MARONNA; MARTIN; YOHAI, 2006), e também, por muitas vezes não se tratarem de erros grosseiros. Métodos de suavização dos efeitos de *outliers* vêm sendo discutidos em diferentes áreas, propondo ferramentas para minimizar seus efeitos nas estimativas (FIELD; SMITH, 1994; AGOSTINELLI, 2003; AGOSTINELLI; BISAGLIA, 2010; NGUYEN; WELSCH, 2010). Um dos métodos mais comentados e aceitos é a ponderação da verossimilhança, onde os estimadores se tornam robustos a desvios causados pelos *outliers*. Na presença de observações discrepantes, a ponderação evita que esses valores exerçam uma influência desproporcional na função de log-verossimilhança, uma vez que estes atraem os estimadores de máxima verossimilhança ao seu valor (NGUYEN; WELSCH, 2010).

Na ausência de qualquer tratamento prévio de um *outlier*, quando aplicado um gráfico de controle, é esperado que ocorra a redução da sensibilidade dos procedimentos de controle dados na Fase I (ROCKE, 1989; JONES-FARMER et al., 2014). Nesses casos, os limites ficam alargados, dificultando a identificação de qualquer comportamento fora de controle (ROCKE, 1989). Croux, Gelper e Mahieu (2011) propuseram gráfico de controle para séries temporais a fim de minimizar o efeito desse fenômeno. O gráfico é dado a partir dos erros de previsão pelo método de Holt-Winters. Além desse, Schoonhoven e Does (2013) apresentam em seu trabalho uma proposta de gráficos de controle robustos para Fase I, removendo uma pequena porcentagem dos maiores e menores valores antes de obter a média, como forma de identificar dados fora de controle.

Até o momento nenhum trabalho relacionado com os problemas de robustez dos modelos KARMA foi proposto. Ainda, sabendo da existência desse fenômeno em estudos com variáveis ambientais, ressalta-se a importância da acurácia das estimativas na Fase I de gráficos de controle para o efetivo controle estatístico dos processos. Sendo assim, a seguir são propostos os gráficos de controle de Shewhart e CUSUM robustos para o modelo KARMA.

4.1 GRÁFICOS DE CONTROLE PROPOSTOS

Os gráficos de controle KARMA robustos diferem dos apresentados no Capítulo 3 por basearem-se em estimadores robustos, conhecidos como estimadores de máxima verossimilhança ponderada. O método de ponderação utilizado é baseado no proposto por Manchini, Portella e Bayer (2019), adaptado para o modelo KARMA.

4.1.1 Estimadores de máxima verossimilhança ponderada

Os estimadores de máxima verossimilhança ponderada são uma adaptação dos estimadores de máxima verossimilhança usuais. A estimação torna-se robusta com a introdução de um vetor peso, que pondera a função de verossimilhança fazendo com que qualquer ponto discrepante em relação a massa de dados seja suavizado (FIELD; SMITH, 1994).

Dado um vetor peso conhecido, $\omega = (\omega_1, \omega_2, \dots, \omega_n)^\top$, a função de log-verossimilhança

condicional ponderada do modelo KARMA é definida por:

$$\ell_{\omega_t} = \sum_{t=m+1}^n \omega_t \log(f(\tilde{y}_t | \mathcal{F}_{t-1})) = \sum_{t=m+1}^n \ell_{\omega_t}(\mu_t, \varphi), \quad (4.1)$$

em que

$$\begin{aligned} \ell_{\omega_t}(\mu_t, \varphi) &= \omega_t \log(\varphi) - \omega_t \log(b-a) + \omega_t \log\left(\frac{\log(0.5)}{\log(1-\mu_t^\varphi)}\right) \\ &\quad + \omega_t(\varphi-1) \log(y_t) + \omega_t \left(\frac{\log(0.5)}{\log(1-\mu_t^\varphi)} - 1\right) \log(1-y_t^\varphi). \end{aligned}$$

As derivadas parciais da função de log-verossimilhança condicional ponderada, em relação a cada um dos elementos j no vetor de parâmetros γ (sendo $\gamma_j \neq \varphi$, para $j \in \{1, \dots, (p+q+r+1)\}$), são definidas, de forma geral, por:

$$\frac{\partial \ell_{\omega_t}}{\partial \gamma_j} = \sum_{t=m+1}^n \frac{\partial \ell_{\omega_t}(\mu_t, \varphi)}{\partial \mu_t} \frac{d\mu_t}{d\eta_t} \frac{\partial \eta_t}{\partial \gamma_j},$$

em que $\frac{d\mu_t}{d\eta_t} = \frac{1}{g'(\mu_t)}$,

$$\frac{\partial \ell_{\omega_t}(\mu_t, \varphi)}{\partial \mu_t} = \omega_t \varphi c_t,$$

sendo

$$c_t = \frac{\mu_t^{\varphi-1}}{(1-\mu_t^\varphi) \log(1-\mu_t^\varphi)} \left(\frac{\log(0.5)}{\log(1-\mu_t^\varphi)} \log(1-y_t^\varphi) + 1 \right).$$

A solução para a simplificação de $\frac{\partial \eta_t}{\partial \gamma_j}$ é obtida derivando η_t em relação a cada parâmetro do vetor γ_j . Assim, considerando $r_t = g(y_t) - g(\mu_t)$, a derivada de η_t em relação a α , β_l , ϕ_i , θ_j , respectivamente, são dadas por:

$$\begin{aligned} \frac{\partial \eta_t}{\partial \alpha} &= 1 + \sum_{j=1}^q \theta_j \frac{\partial r_{t-j}}{\partial \alpha} = 1 - \sum_{j=1}^q \theta_j \frac{\partial \eta_{t-j}}{\partial \alpha}, \\ \frac{\partial \eta_t}{\partial \beta_l} &= x_{tl} - \sum_{i=1}^p \phi_i x_{(t-i)l} - \sum_{j=1}^q \theta_j \frac{\partial \eta_{t-j}}{\partial \beta_l}, \\ \frac{\partial \eta_t}{\partial \phi_i} &= g(y_{t-1}) - x_{t-1}^\top \beta_l - \sum_{j=1}^q \theta_j \frac{\partial \eta_{t-j}}{\partial \phi_i}, \\ \frac{\partial \eta_t}{\partial \theta_j} &= r_{t-j} - \sum_{i=1}^q \theta_i \frac{\partial \eta_{t-i}}{\partial \theta_j}, \end{aligned}$$

em que $l = 1, \dots, r$, x_{tl} o l -ésimo elemento de x_t , $i = 1, \dots, p$ e $j = 1, \dots, q$.

Para a derivada de ℓ_{ω_t} em relação a φ , tem-se que:

$$\frac{\partial \ell_{\omega_t}}{\partial \varphi} = \frac{\omega_t}{\varphi} + \omega_t \log(y_t) + \omega_t c_t \mu_t \log(\mu_t) - \omega_t \left(\frac{\log(0,5)}{\log(1-\mu_t^\varphi)} - 1 \right) \frac{y_t^\varphi \log(y_t)}{(1-y_t^\varphi)}.$$

Dadas as derivadas e o vetor escore, $\mathbf{U}_\omega(\gamma) = (\mathbf{U}_{\omega\alpha}(\gamma), \mathbf{U}_{\omega\beta}(\gamma)^\top, \mathbf{U}_{\omega\phi}(\gamma)^\top, \mathbf{U}_{\omega\theta}(\gamma)^\top, \mathbf{U}_{\omega\varphi}(\gamma)^\top)^\top$, o estimador de máxima verossimilhança condicional ponderada é obtido pelo argumento que maximiza $\ell_{\omega_t}(\gamma)$, ou seja, $\mathbf{U}_\omega(\gamma) = \mathbf{0}$.

Um método utilizado para penalizar observações cujo valor é discrepante dos demais é apresentado por Field e Smith (1994), e utilizado no método iterativo de Manchini, Portella e Bayer (2019). Assim, tem-se que o vetor peso é dado por:

$$\omega_t = \begin{cases} F(y_t; \mu_t, \varphi)/p, & \text{se } F(y_t; \mu_t, \varphi) < p, \\ 1, & \text{se } p \leq F(y_t; \mu_t, \varphi) \leq 1 - p, \\ (1 - F(y_t; \mu_t, \varphi))/p, & \text{se } F(y_t; \mu_t, \varphi) > 1 - p, \end{cases} \quad (4.2)$$

sendo p qualquer valor do intervalo $(0, 1)$ responsável por gerar os limites da ponderação. Um valor usual para p , assim como em Manchini, Portella e Bayer (2019), é $p = 0,01$.

O processo de estimação ponderada dado pelo método iterativo de Manchini, Portella e Bayer (2019) propõe, a partir de uma estimação inicial dos parâmetros por um método sem ponderação, determinar a sequência do vetor peso ω_t e, desse, estimar os parâmetros pelo método de máxima verossimilhança ponderada. Define-se a sequência de passos do processo da seguinte forma:

1. Estimam-se os parâmetros pelo modelo KARMA;
2. Determina-se ω_t pela Equação 4.2;
3. Estimam-se novamente os parâmetros, porém através da maximização da função ponderada dada na Equação 4.1;
4. Verifica-se a diferença máxima entre as estimativas iniciais (passo 1) e ponderadas (passo 3).
5. Se a diferença obtida no passo 4 for menor que 1%, o processo se encerra e os estimadores são dados pela última estimativa realizada. Caso contrário, repete-se os passos 2 a 5, atribuindo às estimativas iniciais do passo 4 a estimativa obtida na iteração anterior.

Na proposição dos gráficos de controle KARMA robustos, a ponderação das observações discrepantes será dada a partir da versão iterativa de Manchini, Portella e Bayer (2019). No entanto, sabendo da existência do fenômeno de autocorrelação em dados hidrológicos, apenas essa ponderação não garante a minimização do efeito de uma observação discrepante nas previsões

futuras. A fim de filtrar esses *outliers* da série, substituiu-se as observações com valores discrepantes pela estimação da mediana realizada no passo 3, para assim determinar ω_t e estimar os parâmetros nos passos seguintes. Portanto, assim como em Bayer, Bayer e Gamba (2020), a série de observações y_t passa a ser definida por:

$$y_t = \begin{cases} \hat{\mu}_t, & \text{se } F(y_t; \mu_t, \varphi) < p, \\ y_t, & \text{se } p \leq F(y_t; \mu_t, \varphi) \leq 1 - p, \\ \hat{\mu}_t, & \text{se } F(y_t; \mu_t, \varphi) > 1 - p. \end{cases} \quad (4.3)$$

Assim, quando y_t apresentar valor discrepante, este não será considerado na estimação do passo seguinte, mas sim, $\hat{\mu}_t$.

4.2 AVALIAÇÃO NUMÉRICA

A avaliação numérica dos gráficos de controle KARMA robustos foi realizada em duas etapas. Em um primeiro momento avaliou-se os estimadores pontuais, isto é, medidas de robustez do ponto de ruptura das estimativas de máxima verossimilhança usual e ponderada à presença de *outliers* e da sensibilidade dos estimadores frente a diferentes magnitudes. Posteriormente, foi proposta uma análise de desempenho dos gráficos de controle KARMA robustos e não robustos à presença de *outliers*, de forma a avaliar a resistência dos gráficos à presença de *outliers*, em termos de falsos alarmes. Para ambas as etapas, introduziu-se *outliers* à série de dados adicionando à estrutura da mediana (descrita na Equação 3.1), em pontos aleatórios da série, um valor τ , da seguinte forma:

$$\eta_t = g(\mu_t) = \alpha + x_t^\top \beta + \sum_{i=1}^p \phi_i [g(y_{t-i}) - x_{t-i}^\top \beta] + \sum_{j=1}^q \theta_j r_{t-j} + \tau_u. \quad (4.4)$$

Ao parâmetro τ_u atribuiu-se valor diferente de zero quando inserido um *outlier* na posição u do comprimento da amostra de dados. A posição u será sorteada aleatoriamente entre 1 e n , sendo n o tamanho da série.

Assim como na avaliação numérica realizada no Capítulo 3, para avaliar numericamente os gráficos de controle KARMA robustos, utilizou-se de simulações de Monte Carlo e faz-se uso do software R (R Core Team, 2019) e dos pacotes citados no Capítulo anterior. Além disso, as séries geradas neste Capítulo seguiram o formato apresentado no Capítulo 3, Cenários 1 e 3, apresentados na Tabela 3.1. O Cenário 2 não foi utilizado nesta avaliação em função do baixo valor de precisão, resultando em séries com alta variabilidade. Assim, *outliers* inseridos seriam

mascarados pelos altos valores das observações presentes na série, não justificando a análise robusta.

4.2.1 Avaliação dos estimadores pontuais

Para análise de desempenho dos estimadores pontuais, gerou-se séries sintéticas Kumaraswamy com dois diferentes tamanhos amostrais: $n_1 = 300$ e 500 . Ajustou-se o modelo KARMA com estimadores usuais e estimadores robustos às séries de dados e avaliou-se as estimativas de cada estimador pelas medidas de robustez: ponto de ruptura e sensibilidade. Nas seções a seguir são descritas a análise de ponto de ruptura e curva de sensibilidade.

4.2.1.1 Ponto de ruptura

O ponto de ruptura da estimativa dos parâmetros é definido pela quantidade máxima de *outliers* inseridos na série, sem que haja prejuízos desproporcionais no viés do estimador (MARONNA; MARTIN; YOHAI, 2006). Matematicamente, esta relação pode ser expressa por m/n , sendo m o número de *outliers* em que o viés mantém-se aceitável na série com tamanho amostral n (MARONNA; MARTIN; YOHAI, 2006).

Para identificar o ponto de ruptura de um estimador, substituiu-se observações da série por *outliers* e avaliou-se o viés relativo do estimador a medida que aumentou-se o número de *outliers* inseridos, dado por:

$$VR(\%) = \frac{\text{valor estimado} - \text{valor do parâmetro}}{\text{valor do parâmetro}}.$$

Esta análise pode ser dada em termos médios, onde considera-se a relação agrupada do VR dos parâmetros dado através da média dos valores estimados e da média dos valores dos parâmetros. Assim, a partir do ponto de ruptura é possível medir qual a maior relação que o estimador suporta sem ser viesado. Portanto, quanto maior essa relação, mais robusto será o estimador.

Na identificação do ponto de ruptura, inseriu-se de 0 a 50 *outliers* na série de $n_1 = 500$ observações, e de 0 a 30 *outliers* na série de $n_1 = 300$ observações, todos com magnitude de $\tau = 0,65$ e em posições aleatórias. O comportamento dos estimadores em cada acréscimo de observação discrepante é avaliado em termos médios de viés relativo dos parâmetros.

4.2.1.2 Curva de sensibilidade

A curva de sensibilidade (CS) é usada para avaliar o efeito de um *outlier* em uma série de dados (MARONNA; MARTIN; YOHAI, 2006). Para avaliação da sensibilidade dos estimadores, foi introduzido um deslocamento τ em um ponto da série, escolhido de forma aleatória em cada réplica de Monte Carlo. Variou-se o valor de τ de -3 a 3 , a cada $0,5$, simulando diferentes magnitudes de *outliers*. Para cada magnitude avaliou-se o comportamento dos estimadores usual e ponderado através da seguinte métrica:

$$CS(\hat{\gamma}_j) = \hat{\gamma}_j(y_1, y_2, \dots, y_n) - \tilde{\gamma}_j(y_1, y_2, \dots, y_u, \dots, y_n),$$

em que as estimativas dos parâmetros $\hat{\gamma} = (\hat{\alpha}, \hat{\beta}^\top, \hat{\phi}^\top, \hat{\theta}^\top, \hat{\varphi})^\top$ foram obtidos da série sem *outliers* e as estimativas $\tilde{\gamma} = (\tilde{\alpha}, \tilde{\beta}^\top, \tilde{\phi}^\top, \tilde{\theta}^\top, \tilde{\varphi})^\top$ da série com *outlier* na posição u . Avaliando o comportamento dos estimadores em termos médios, a métrica passa a considerar a média de todas as estimativas dos parâmetros sem *outliers* subtraindo a média de todas as estimativas dos parâmetros com *outliers*. Assim, nesta avaliação foi considerada a sensibilidade das estimativas dos parâmetros de forma agregada.

4.2.2 Análise de desempenho dos gráficos de controle

Para esta avaliação, comparou-se o desempenho de gráficos de controle KARMA usuais na presença de *outliers* com o desempenho de gráficos de controle KARMA robustos a desvios causados pela presença de *outliers*. Considerou-se apenas o resíduo quantílico por ele ter apresentado o melhor desempenho na avaliação dos gráficos de controle KARMA do Capítulo 3. Essa avaliação é realizada apenas em termos de ARL_0 , pois é nesta etapa que evidenciam-se as distorções de tamanho dadas pelas distorções nas estimativas dos parâmetros do modelo na Fase I. Avaliações semelhantes para outras classes de gráficos de controle robustos podem ser observados em Croux, Gelper e Mahieu (2011), Schoonhoven e Does (2013) e Chiang et al. (2018).

Visando a identificação da influência dos *outliers* nas estimativas, o procedimento será semelhante ao que foi apresentado na Seção 3.3.1, com cenários simulados seguindo os descritos no Capítulo 3, Cenários 1 e 3, avaliando-os para um $ARL_0 = 370$. Assim, será adicionado à estrutura da mediana, em pontos aleatórios da série, um $\tau = 0,65$ para simular a presença de *outliers*, conforme a Equação 4.4. E em um segundo momento, será adicionando à estrutura da

mediana um único *outlier* com magnitude (τ) variando de $[-3, 3]$.

4.3 RESULTADOS NUMÉRICOS

A avaliação numérica dos estimadores pontuais foi realizada em duas etapas: ponto de ruptura e curva de sensibilidade. Os resultados desta avaliação são apresentados em termos médios, ou seja, a média do VR e a média da sensibilidade de todos parâmetros estimados. Assim, haverá um valor médio de VR e um de sensibilidade para cada estágio avaliado de interferência de *outlier* na série, seja pela quantidade de observações discrepantes inseridas ou pela intensidade atribuída às observações. Na avaliação numérica dos estimadores pontuais por meio do ponto de ruptura, o ajuste do modelo KARMA robusto apresentou melhores resultados. As Figuras 10 e 11 apresentam a análise de ponto de ruptura na estimação dos parâmetros do modelo KARMA para $n_1 = 300$ e $n_1 = 500$ no Cenário 1 e as Figuras 12 e 13 apresentam os resultados de ponto de ruptura para o Cenário 3.

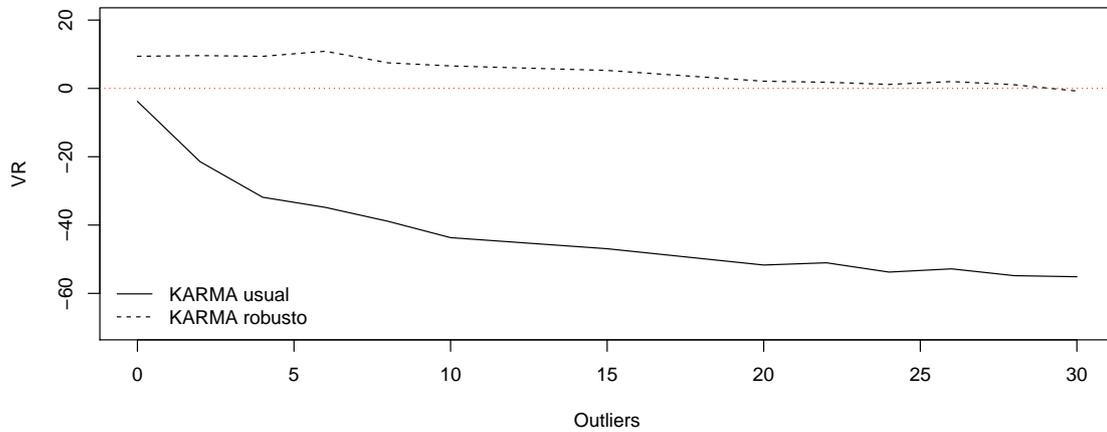
Nas Figuras 10, 11, 12 e 13, destaca-se o desempenho do estimador robusto, capaz de manter o viés com pouca variação ao aumento de *outliers* na série. Por outro lado, o estimador usual, apresentando uma influência crescente à medida que são adicionados mais *outliers*, atingindo valores de viés maiores que $| 50 | \%$. Este comportamento é verificado em ambos cenários avaliados, apesar de se mostrar mais evidente no Cenário 3, onde foram observados os valores de VR mais altos.

Em termos de tamanho da série, foi possível observar que no ajuste KARMA usual, para uma mesma quantidade de *outliers* inseridos, o viés apresentado na série com tamanho $n_1 = 300$ foi maior quando comparado ao viés apresentado no tamanho $n_1 = 500$. Isso já era esperado, uma vez que no tamanho amostral menor, haverá uma maior acumulação das observações discrepantes, prejudicando mais a estimação dos parâmetros do modelo. No tamanho amostral $n_1 = 500$, estes *outliers* são distribuídos ao longo da série, acumulando menor efeito sobre as demais observações. No KARMA robusto não foram observadas grandes diferenças de viés nos tamanhos amostrais avaliados.

As Figuras 14 e 15 apresentam a sensibilidade média na estimativa dos parâmetros do modelo KARMA a partir das metodologias usual e robusta para tamanhos amostrais de $n_1 = 300$ e $n_1 = 500$ no Cenário 1 e as Figuras 16 e 17 no Cenário 3.

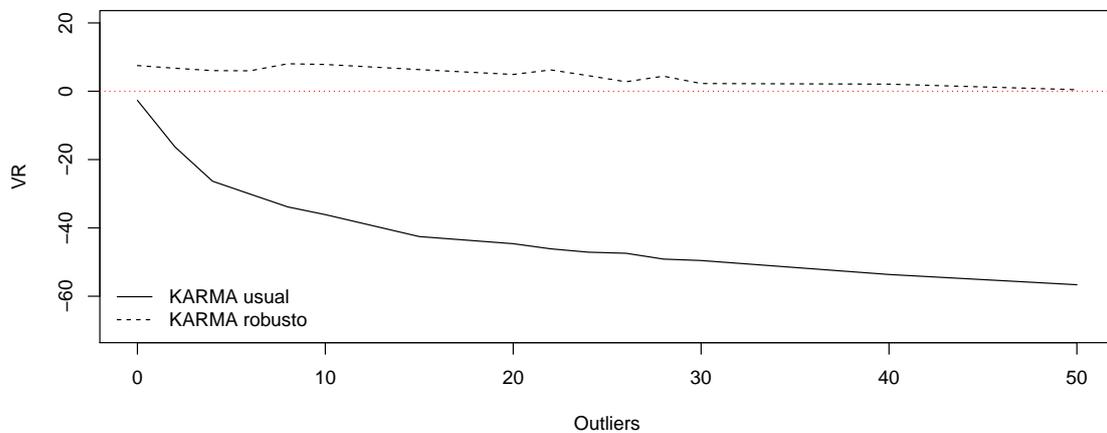
O ajuste KARMA, por modelar a mediana condicional, naturalmente é capaz de reduzir

Figura 10: Ponto de ruptura Cenário 1, $n_1 = 300$.



Fonte: Elaborado pela autora.

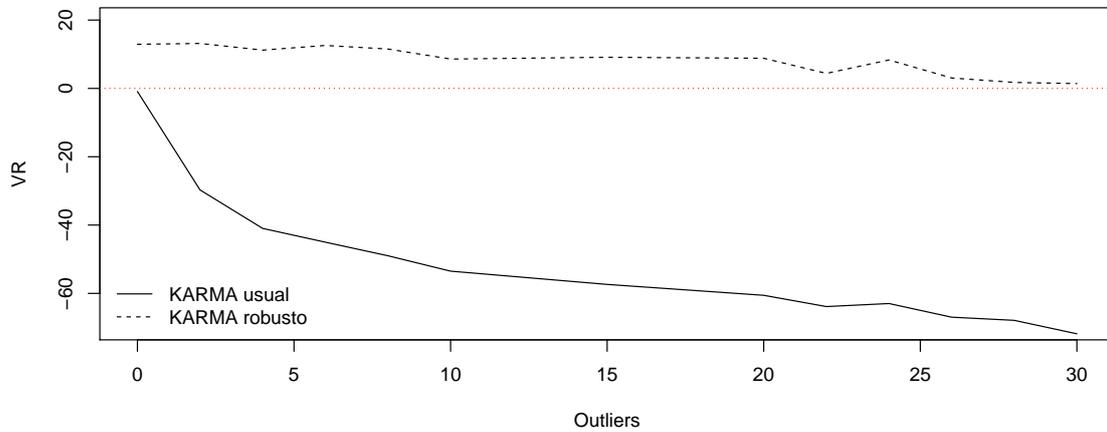
Figura 11: Ponto de ruptura Cenário 1, $n_1 = 500$.



Fonte: Elaborado pela autora.

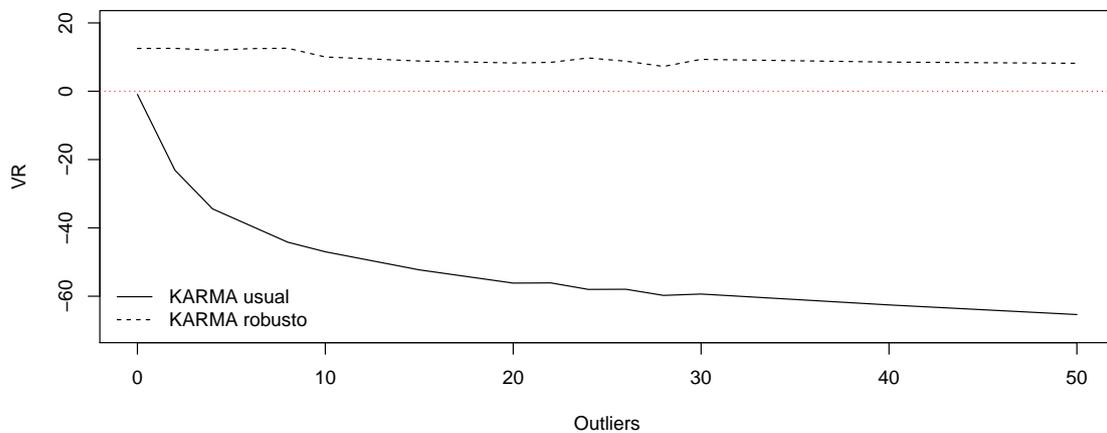
o efeito de *outliers* nas inferências sobre os parâmetros, fornecendo uma medida mais robusta na presença de valores discrepantes quando comparado com modelos que modelam a média, como o β ARMA. Na avaliação do ponto de ruptura, esta robustez do ajuste KARMA não foi capaz de suavizar o efeito da presença de *outliers* na série quando aumentada a frequência do valor da mesma, prejudicando as estimativas dos parâmetros. O fenômeno de autocorrelação, implícito na série de observações, corrobora para intensificação deste efeito, pois carrega a memória da

Figura 12: Ponto de ruptura Cenário 3, $n_1 = 300$.



Fonte: Elaborado pela autora.

Figura 13: Ponto de ruptura Cenário 3, $n_1 = 500$.

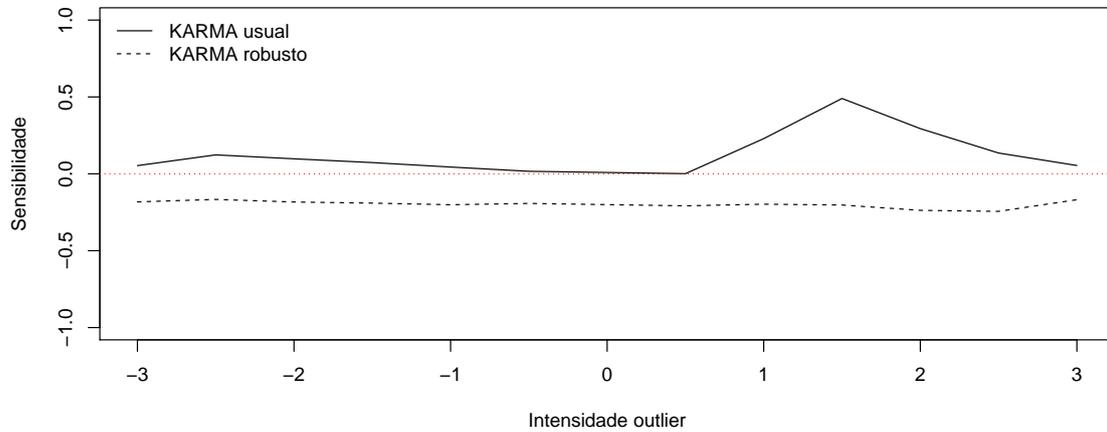


Fonte: Elaborado pela autora.

série e, portanto, o efeito das observações discrepantes. Já na avaliação de sensibilidade, dadas Figuras 14, 15, 16 e 17, por se tratar de uma única observação com valor discrepante, verificou-se, no geral, uma sensibilidade do estimador usual muito próxima à sensibilidade da curva do estimador robusto, consequência desta robustez do modelo.

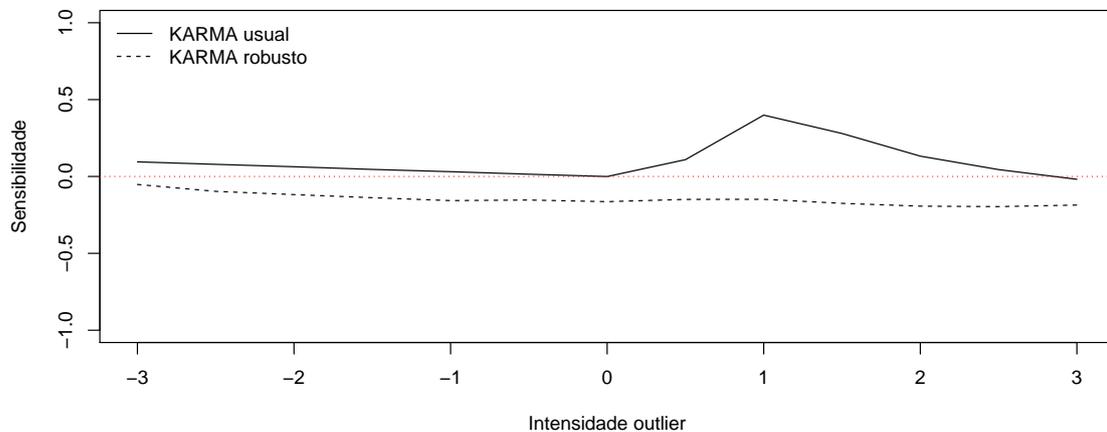
Nas Figuras 14, 15, 16 e 17 é possível observar um pico na sensibilidade do estimador usual próximo ao $\tau = 1$ para ambos cenários. Está contido nesse intervalo o valor da mediana

Figura 14: Curva de sensibilidade Cenário 1, $n_1 = 300$.



Fonte: Elaborado pela autora.

Figura 15: Curva de sensibilidade Cenário 1, $n_1 = 500$.

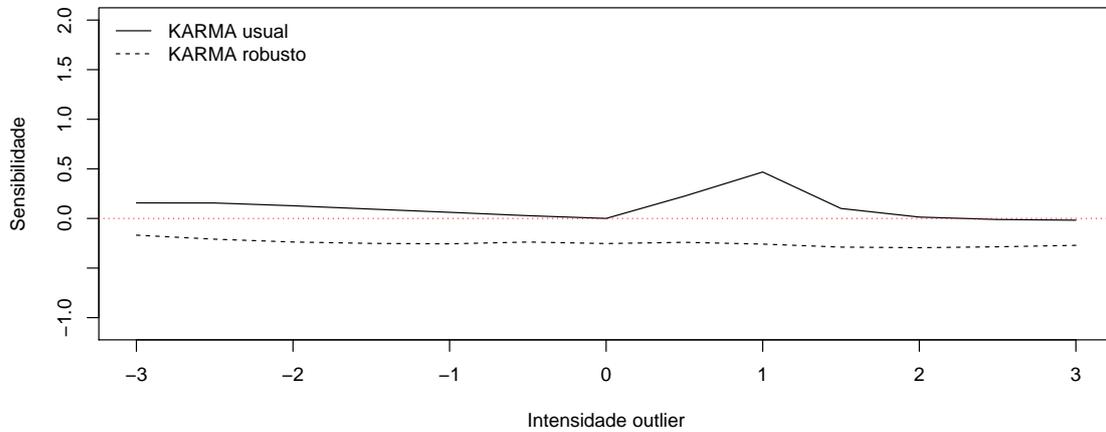


Fonte: Elaborado pela autora.

do Cenário 1 ($\mu = 0,35$) e da mediana do Cenário 3 ($\mu = 0,25$). Assim, verifica-se uma influência das observações inseridas na série apenas quando estas possuem valores próximos ao valor da mediana do cenário analisado, deslocando o valor da mediana e comprometendo as inferências sobre os parâmetros do ajuste KARMA usual. Por outro lado, o estimador robusto, a partir da ponderação w_t , minimiza o efeito dessas observações na estimativa dos parâmetros.

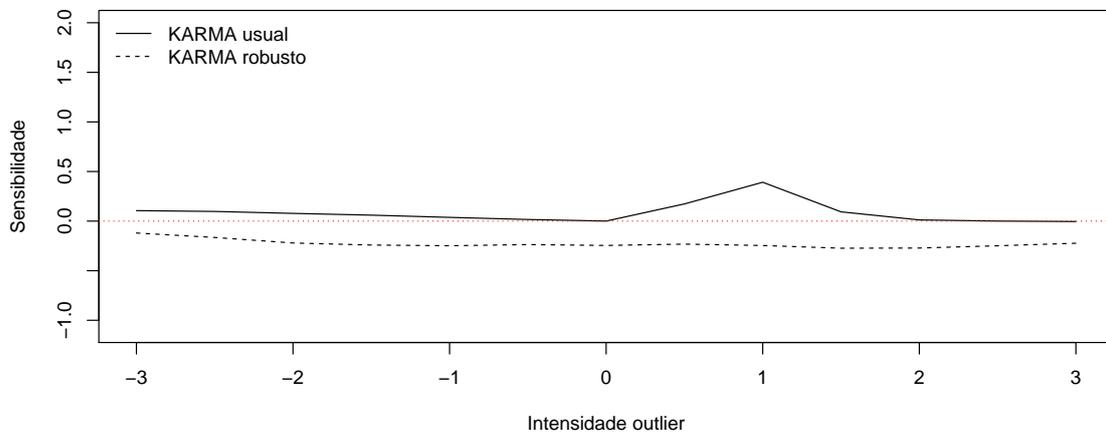
Comparando os resultados apresentados nas Figuras 14 e 15 e Figuras 16 e 17 quanto

Figura 16: Curva de sensibilidade Cenário 3, $n_1 = 300$.



Fonte: Elaborado pela autora.

Figura 17: Curva de sensibilidade Cenário 3, $n_1 = 500$.



Fonte: Elaborado pela autora.

ao número de observações da Fase I, poucas diferenças foram identificadas. Observou-se, no geral, valores de sensibilidade mais próximos a zero em $n_1 = 500$ para os Cenários 1 e 3, assim, quanto maior o tamanho da série na Fase I, melhores as inferências sobre os parâmetros do modelo ajustado.

Posterior à avaliação numérica dos estimadores pontuais, aplicou-se a metodologia dos gráficos apresentados neste Capítulo à série sob controle a fim de verificar a influência dos

outliers no monitoramento da série. Dada estimativa dos parâmetros na Fase I com presença de *outliers*, observou-se a confiabilidade do monitoramento na Fase II sob controle, avaliando o ARL_0 empírico.

Avaliou-se em um primeiro momento em termos de ponto de ruptura do ARL_0 , aumentando a frequência de *outliers* com magnitude $\tau = 0,65$ e, na sequência, avaliou-se em termos de sensibilidade do ARL_0 , aumentando a intensidade τ de um *outlier* inserido à série.

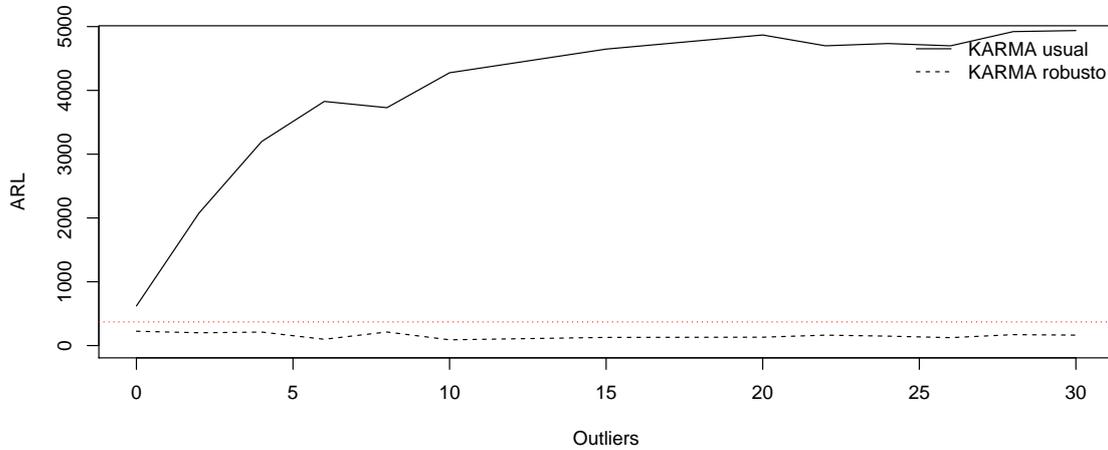
A avaliação dos gráficos de controle para as diferentes quantidade de *outliers* inseridos no tamanho amostral $n_1 = 300$ é apresentada nas Figuras 18 e 19 para o Cenário 1 e nas Figuras 20 e 21 para o Cenário 3. As análises realizadas para $n_1 = 500$ são apresentadas no Apêndice B deste documento por motivos de similaridade. Nesta avaliação, foi verificado um comportamento similar ao da avaliação numérica dos estimadores pontuais. Os gráficos de controle KARMA robustos mantiveram o ARL_0 empírico próximo ao valor nominal, sem sofrer influência do aumento de *outliers* na série. Por outro lado, desde a primeira observação inserida já é possível identificar uma ascensão da curva de ARL_0 dos gráficos KARMA usual.

Observa-se que os gráficos KARMA usual, em todos os cenários, atingem uma estabilidade, muito próxima ao valor de 5.000, após a adição de 15 *outliers* na série. Desta forma, é possível identificar a redução da sensibilidade dos procedimentos de controle na estimação usual dos parâmetros, causando o alargamento dos limites e prejudicando o monitoramento de séries fora de controle. No monitoramento de variáveis ambientais, especialmente hidrológicas, o alargamento dos limites de controle poderia causar prejuízos graves a tomada de decisões gerando extravasamento da vazão sobre a crista de barragens e estiagem em reservatórios. Nesse caso, teríamos uma taxa de acerto reduzida, diminuindo a identificação de qualquer comportamento fora de controle e tardando medidas de gestão que evitassem prejuízos nos sistemas monitorados.

As Figuras 22 e 23 apresentam os gráficos de controle de Shewhart e CUSUM dado Cenário 1 no tamanho amostral $n_1 = 300$ para análise de intensidade de *outlier*. Os resultados desta análise obtidos para o Cenário 3, no tamanho amostral $n_1 = 300$, são apresentados nas Figuras 24 e 25.

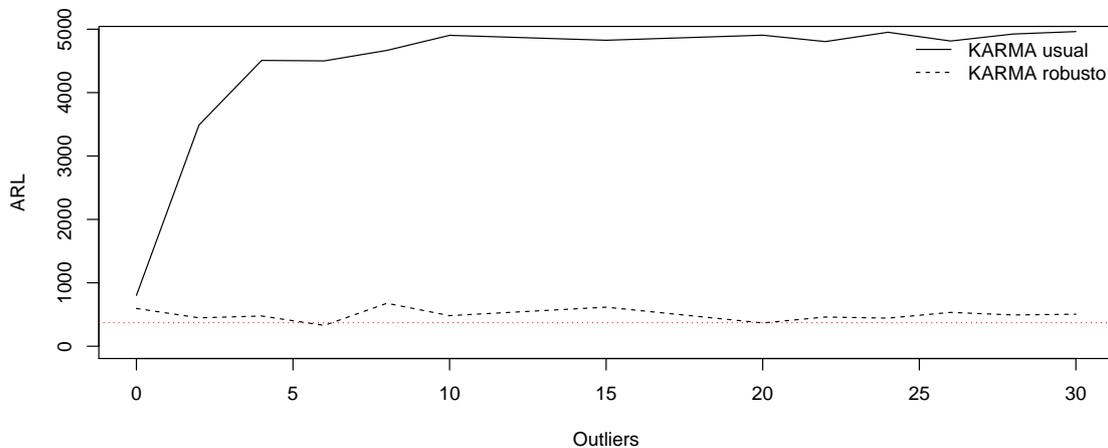
Das Figuras 22, 23, 24 e 25, assim como nos resultados de ponto de ruptura de ARL_0 , foi possível observar um comportamento semelhante ao da análise de curva de sensibilidade dos estimadores pontuais. Foi identificado um maior distorção de ARL_0 próxima ao intervalo de intensidade de observações com valores de $[0, 2]$, assim como na análise dos estimadores. Neste

Figura 18: Ponto de ruptura do ARL_0 do gráfico de controle de Shewhart, Cenário 1.



Fonte: Elaborado pela autora.

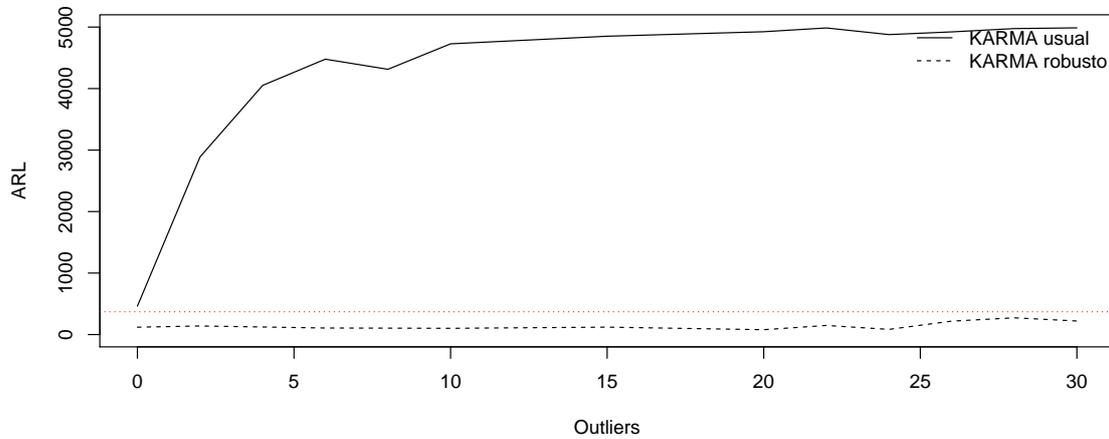
Figura 19: Ponto de ruptura do ARL_0 do gráfico de controle CUSUM, Cenário 1.



Fonte: Elaborado pela autora.

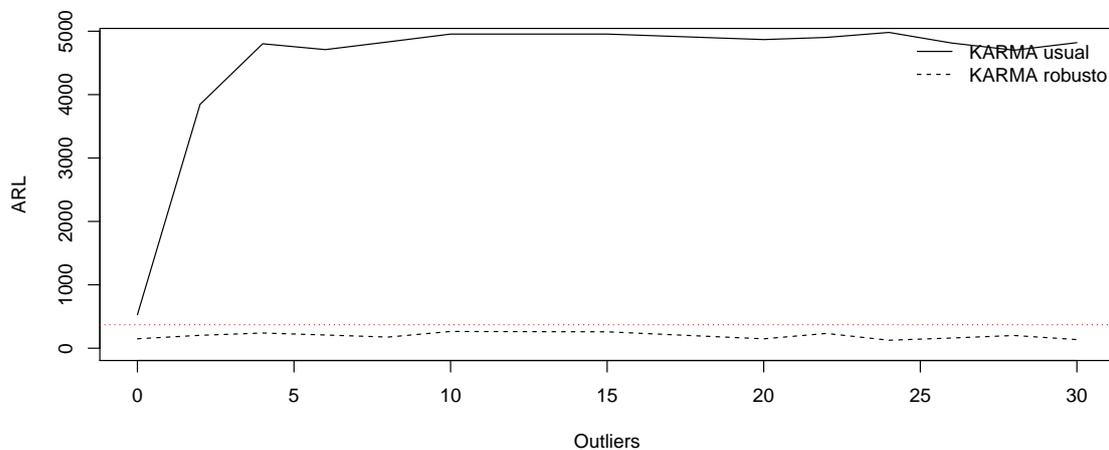
intervalo, em ambos cenários avaliados, os gráficos KARMA usual alcançaram ARL_0 maiores que 2.000 em Shewhart e maiores que 3.000 em CUSUM, enquanto que os gráficos KARMA robustos mantiveram o ARL_0 empírico próximo ao nominal ($ARL_0 = 370$). Portanto, neste intervalo, sob influência dessas observação, os limites de controle dos gráficos KARMA usuais foram alargados, aumentando a probabilidade de erro tipo II e dificultando a identificação de qualquer observação fora de controle.

Figura 20: Ponto de ruptura do ARL_0 do gráfico de controle de Shewhart, Cenário 3.



Fonte: Elaborado pela autora.

Figura 21: Ponto de ruptura do ARL_0 do gráfico de controle CUSUM, Cenário 3.

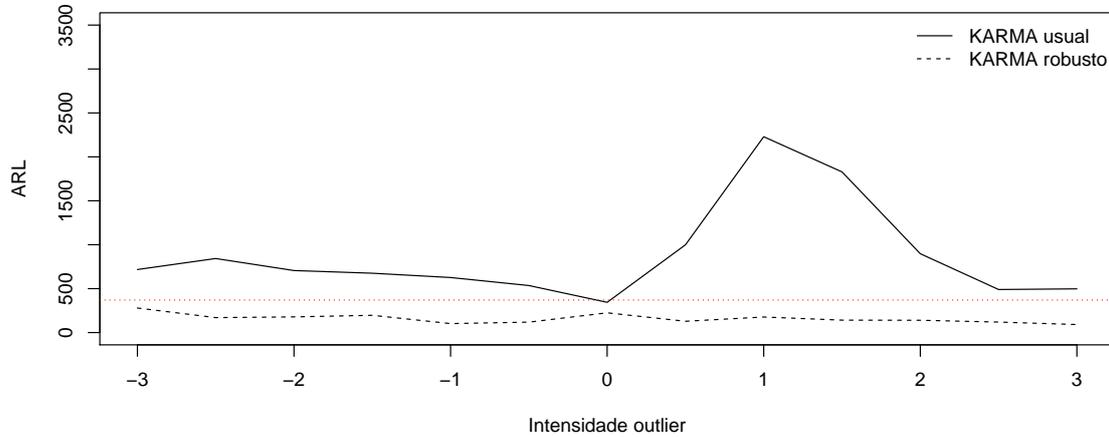


Fonte: Elaborado pela autora.

4.4 CONCLUSÃO

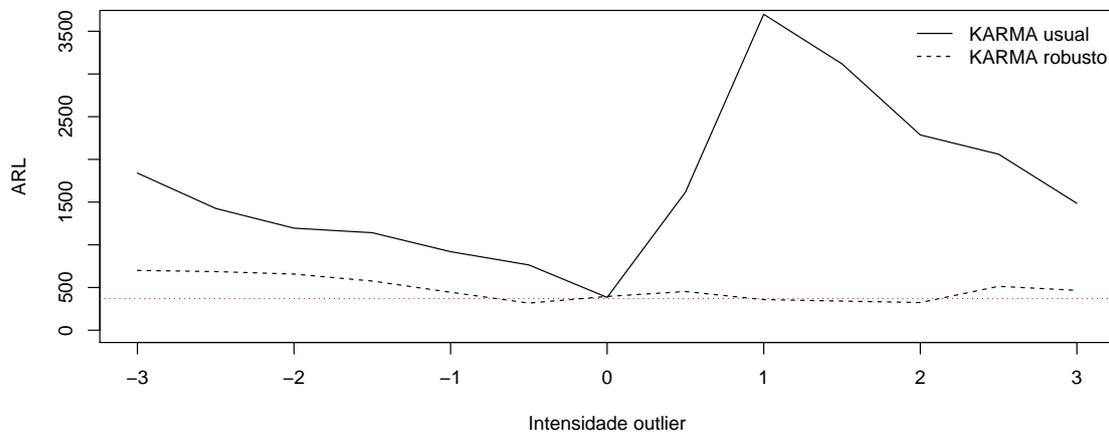
Na literatura, pouca atenção foi dedicada a robustez dos modelos KARMA na presença de *outliers*. Sabendo da importância de estimativas acuradas no monitoramento de gráficos de controle, nesse capítulo, propomos gráficos de controle robustos. Essa proposta foi dada a partir de um procedimento de ponderação dos estimadores, minimizando o efeito de obser-

Figura 22: Sensibilidade do ARL_0 do gráfico de controle de Shewhart, Cenário 1.



Fonte: Elaborado pela autora.

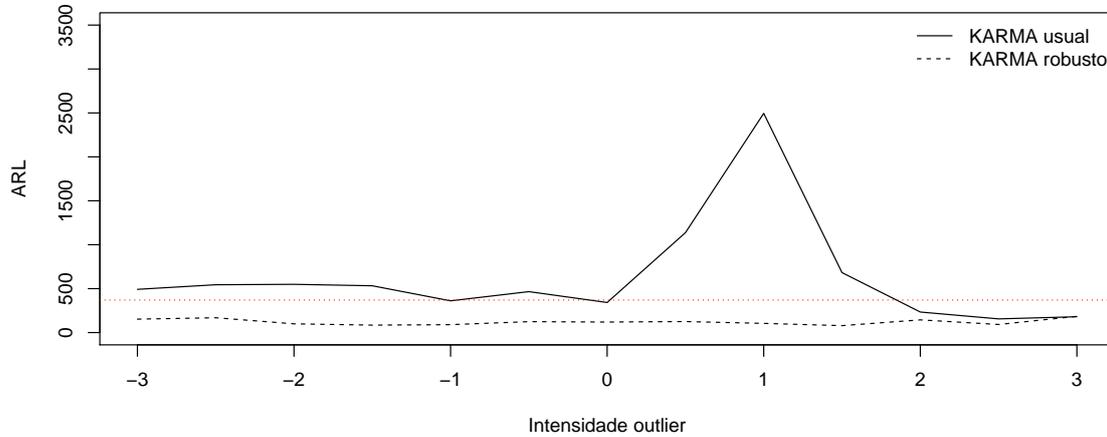
Figura 23: Sensibilidade do ARL_0 do gráfico de controle CUSUM, Cenário 1.



Fonte: Elaborado pela autora.

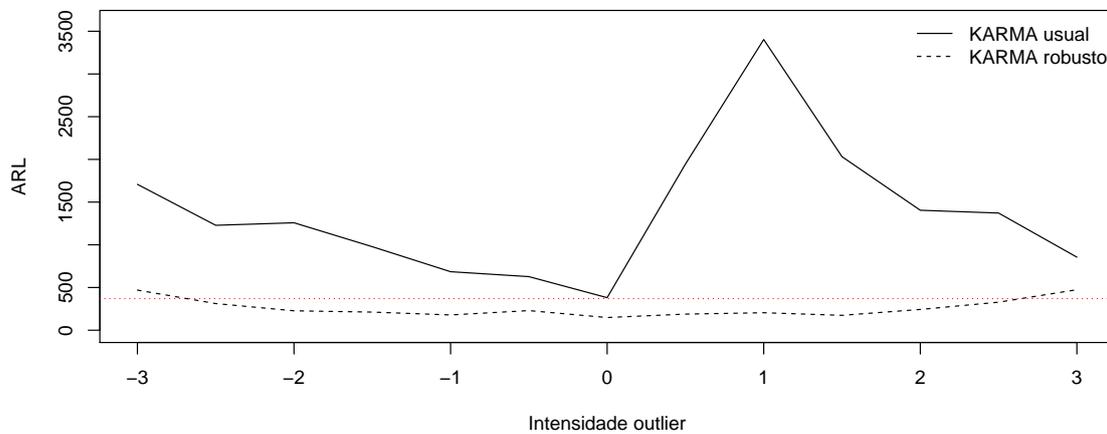
vações discrepantes nas estimativas dos parâmetros do modelo ajustado na Fase I dos gráficos de controle já discutidos no capítulo anterior. Nesse procedimento, o estimador de máxima verossimilhança ponderada foi capaz, simultaneamente, de classificar as observações, identificar observações suspeitas, atribuir pesos a essas e obter estimativas robustas dos parâmetros do modelo. Na avaliação do desempenho dos estimadores pontuais frente a diferentes intensidades e quantidades de *outliers* na série foi identificada uma superioridade nas estimativas robustas,

Figura 24: Sensibilidade do ARL_0 do gráfico de controle de Shewhart, Cenário 3.



Fonte: Elaborado pela autora.

Figura 25: Sensibilidade do ARL_0 do gráfico de controle CUSUM, Cenário 3.



Fonte: Elaborado pela autora.

apresentando menor sensibilidade e menor VR. Na avaliação de ponto de ruptura, a robustez implícita ao modelo KARMA não foi capaz de minimizar o efeito dos *outliers* na série de observações. Verificou-se um aumento crescente do VR médio dos estimadores usuais à medida que eram adicionados *outliers*. Apesar disso, na avaliação de ponto de ruptura dos estimadores robustos observou-se um VR médio constante e próximo a zero. Na análise de curva de sensibilidade, os estimadores robustos mantiveram uma sensibilidade aproximadamente constante para

todas intensidades de *outliers* avaliadas. Por outro lado, constatou-se maior sensibilidade nos estimadores usuais, principalmente quando adicionadas observações com valores próximos ao da mediana do cenário. Observações discrepantes com valores altos ou baixos, foram filtradas ao utilizar a mediana condicional no modelo.

Na avaliação de gráficos de controle, verificou-se um comportamento similar à avaliação dos estimadores pontuais. Os gráficos de controle KARMA robustos apresentaram um bom desempenho, mantendo um ARL_0 empírico constante e próximo ao ARL_0 nominal. No entanto, a presença de *outliers* na estimação dos parâmetros do modelo KARMA usual promoveu o alargamento dos limites, dificultando a identificação de observações fora de controle. Na análise de quantidade de *outliers*, este alargamento pode ser identificada, já nas primeiras observações adicionadas, uma grande influência e aumento do valor de ARL_0 empírico, consequência do alargamento dos limites de controle e aumento da probabilidade do erro tipo II. Já na avaliação de intensidades de *outlier* inserido, verificou-se certa distorção apenas no intervalo de $[0, 2]$, mesmo intervalo onde ocorreram os maiores valores de sensibilidade na análise dos estimadores pontuais.

Assim, a ponderação dos estimadores de máxima verossimilhança garantiu uma neutralização do efeito das observações discrepantes em todos os cenários analisados. A partir da neutralização dos *outliers*, as estimativas da Fase I dos gráficos de controle foram mais acuradas, garantindo um melhor ajuste dos limites para o monitoramento na Fase II.

5 GRÁFICOS DE CONTROLE KARMA ROBUSTOS APLICADOS À GESTÃO DE RESERVATÓRIOS

Diversas atividades envolvendo o uso da água necessitam de sistemas de reservação para operarem de maneira satisfatória o ano todo. Os reservatórios de água têm como principal função reservar a água em períodos de excesso hídrico para suprir a demanda de uso em períodos de escassez. Dentre as atividades que utilizam sistemas de reservação, destacam-se os reservatórios para abastecimento de água à população e à indústria e para geração de energia elétrica.

5.1 RESERVATÓRIOS DE ABASTECIMENTO DE ÁGUA

Apesar de haver diferentes formas de fornecimento de água para abastecimento público e industrial, de acordo com o último atlas de abastecimento urbano da Agência Nacional de Águas (ANA), aproximadamente 50% dos municípios brasileiros são abastecidos por mananciais superficiais, sendo a forma mais comum de captação dada através de reservatórios (ANA, 2010).

O Brasil é um dos países com maior disponibilidade de água doce do mundo, no entanto, a sua grande variabilidade climática torna desigual a distribuição territorial deste recurso (ANA, 2010). Um exemplo disso é de que aproximadamente 80% da água superficial do país está concentrada na bacia hidrográfica Amazônica, onde a demanda por abastecimento de água é baixa enquanto que no Semiárido brasileiro, onde vivem mais de 24 milhões de habitantes, ocorrem períodos críticos com prolongadas estiagens (ANA, 2017a). Visando o abastecimento contínuo da população, os reservatórios são utilizados como alternativas para compensar balanços hídricos negativos durante o ano. Outro desafio enfrentado na gestão de sistemas de abastecimento de água é a oferta hídrica não acompanhar o crescimento populacional de determinados locais. Nesses, muitas vezes, faz-se necessário a transferência de vazões entre reservatórios para reforçar o sistema defasado (ANA, 2010). Assim, para que ocorra este intercâmbio de vazões de maneira correta, é necessária uma gestão eficiente, com planejamento de operação e tomadas de decisões mais assertivas.

Em 2017 o número de reservatórios para abastecimento de água monitorados pela ANA ultrapassava o número de 670 (ANA, 2017a). Esses reservatórios são distribuídos pelo território

brasileiro e muitos possuem sistemas de interligação, sejam construídos através de dutos ou por canais naturais, podendo ou não interligar diferentes bacias hidrográficas.

Um dos sistemas de abastecimento interligados mais importante do Brasil é o Sistema Cantareira. Implantado em 1973, esse sistema é composto por cinco represas e operado pela Companhia de Saneamento Básico do Estado de São Paulo (Sabesp) (SABESP, 2015). O volume útil do Sistema Cantareira é representado pela soma dos volumes dos reservatórios interligados natural e artificialmente de cada represa, são eles: Jaguari-Jacaraí, Cachoeira, Atibainha e Paiva Castro (ANA, 2017b). Os três primeiros reservatórios pertencem à Bacia PCJ e o último à do Alto Tietê, logo, a interligação destes reservatórios acaba transferindo a água de uma bacia hidrográfica para outra. Para fins de gestão desse sistema, são estabelecidos períodos hidrológicos no ano, sendo o período úmido de 1º de Dezembro a 31 de Maio e o período seco de 1º de Junho a 30 de Novembro (ANA, 2017b). Essas informações hidrológicas, associadas a sistemas de monitoramento meteorológico e hidrológico, e em conjunto com ferramentas de gestão são essenciais para o funcionamento ideal dos sistemas de abastecimento. A operação de sistemas de reserva objetiva um uso efetivo do armazenamento (YOU, 2008), o que demanda ferramentas que auxiliem na tomada de decisão, reduzindo os riscos atribuídos frente a possíveis alterações hídricas.

5.2 RESERVATÓRIOS PARA GERAÇÃO DE ENERGIA ELÉTRICA

Além dos reservatórios para abastecimento de água, os reservatórios para geração de energia elétrica também são sistemas de extrema importância no país. O Brasil possui a terceira maior capacidade instalada de usinas hidrelétricas do mundo, atrás apenas da China e dos Estados Unidos (IHA, 2018). Dentre esses três maiores produtores de energia hidrelétrica, o Brasil é o que possui a maior dependência desta fonte energética. Isso pode ser observado através da representatividade da energia hidrelétrica no total de geração elétrica de cada país, enquanto a China e Estados Unidos apresentam taxas menores que 20%, no Brasil, 64% do total da capacidade instalada é de fonte hidrelétrica (IHA, 2018). Em 2019 o Brasil já possuía mais de 1.300 hidrelétricas em operação (ANEEL, 2019).

O sistema brasileiro de transmissão de energia elétrica é interligado, isto é, existe uma interconexão entre as unidades de produção, subestações e distribuição (IBGE, 2016). Assim, por exemplo, prevendo-se baixa na produção em sistemas hidrelétricos, solicita-se a inclusão

de uma fonte de produção complementar ao sistema (ANA, 2017a). O objetivo de um sistema interligado é suprir as necessidades técnicas do seu funcionamento, executando uma logística de energia (IBGE, 2016). Desta forma, é trivial que haja confiabilidade nas decisões quanto ao vertimento dos reservatórios e intercâmbios da energia produzida entre regiões. O sistema interligado nacional (SIN) é operado pelo operador nacional de sistemas elétricos (ONS).

A ONS divide o Brasil em quatro subsistemas hidrelétricos: Sudeste/Centro-oeste, Sul, Nordeste e Norte (ONS, 2019a). Estes subsistemas são definidos pelos limites das bacias hidrográficas que alimentam os reservatórios. O principal reservatório do subsistema Sudeste/Centro-oeste é o reservatório de Furnas, enquanto no subsistema Sul, o maior reservatório é o Governador Bento Munhoz (ONS, 2019a). Também faz parte do subsistema Sul uma das usinas hidrelétricas mais importantes do Brasil, a usina de Itaipu, que, apesar de possuir apenas o sétimo maior reservatório do país, é a usina com o melhor índice de aproveitamento de água (ITAIPU, 2015). No subsistema Nordeste o principal reservatório é o de Sobradinho e, por fim, no subsistema Norte, Tucuruí (ONS, 2019a).

O princípio básico de operação de uma hidrelétrica é a transformação de energia potencial em energia elétrica (HOUGHTALEN; AKAN; HWANG, 2010). A energia potencial é obtida a partir do desnível entre a cota máxima de água à montante da barragem e a turbina. Desta forma, a utilização de reservatórios de água à montante do sistema garante uma maior estabilidade, podendo manter a carga, estabelecida pelo volume armazenado, por longos períodos sem prejuízos à geração da usina (SOUZA; SANTOS; COSTA BORTONI, 2009). Dada relevância dos sistemas de reservação para geração de energia elétrica e a dependência desse tipo de fonte no país, é de suma importância possuir ferramentas que auxiliem nas tomadas de decisões e na gestão dos sistemas.

5.3 MONITORAMENTO DE NÍVEL EM RESERVATÓRIOS

Uma importante variável na operação de reservatórios é o nível de água. Por meio dessa variável é possível controlar a carga de água sobre a estrutura da barragem e monitorar o volume de água reservado, a fim de gerir adequadamente os usos e reduzir os transtornos provocados por eventos extremos (ONS, 2017). O Sistema Cantareira, por exemplo, opera sob a resolução conjunta ANA/DAEE nº 925 (ANA, 2017b). Esta normatização estabelece cinco faixas de captação de água autorizada do sistema, sendo a faixa normal estabelecida para um

volume acumulado igual ou superior a 60%, e faixa especial quando o volume for menor que 20% do volume útil (ANA, 2017b). No contexto de reservatórios para geração de energia, a usina hidrelétrica de Itaipu monitora o fluxo de água em vários pontos do rio Paraná, por meio de uma extensa rede de estações hidrometeorológicas (ITAIPU, 2015). Buscam por meio deste monitoramento uma maior previsibilidade de vazões afluentes à usina, de modo a garantir a segurança da barragem, o gerenciamento e o planejamento frente a eventuais eventos extremos. De forma similar, a usina hidrelétrica de Furnas monitora os níveis de água dos reservatórios, além das vazões afluentes e defluentes, como ferramenta de gestão e apoio a tomada de decisão na operação (ANA, 2018).

Vinculado ao monitoramento dos níveis de água nos reservatórios, podem ser aplicados modelos de previsão para antever um comportamento hidrológico (BRAVO, 2010; LOHANI; KUMAR; SINGH, 2012; ALLAWI et al., 2018; TIWARI; KUMAR, 2018). No entanto, ambos os métodos de monitoramento podem não ser suficientes, levando a tomadas de decisões equivocadas, como manobras em comportas e demais dispositivos de descarga em momentos inapropriados, por exemplo. Os riscos associados a cada uma dessas falhas na tomada de decisão podem vir a caracterizar um sistema ineficiente. Sistemas ineficientes acabam por não explorar todo o potencial do local e, muitas vezes, fazendo necessária a exploração de novas fontes hídricas, causando novos impactos à população, ao ambiente e à economia. Cabe aos gestores a decisão de como operar as comportas dos vertedouros durante os eventos extremos e acionar outras unidade de produção, no caso de sistemas interligados de energia elétrica.

Diversos estudos vêm sendo realizados com o intuito de auxiliar os tomadores de decisão na seleção da melhor estratégia de operação de reservatórios. Dentre a variedade de modelos matemáticos e ferramentas computacionais, pode-se citar métodos utilizando a lógica *fuzzi* para operação de comportas em tempo real (KARABOGA; BAGIS; HAKTANIR, 2004), mineração de dados para extrair regras de controle de reservatórios (BESSLER; SAVIC; WALTERS, 2003), modelos empíricos de chuva-vazão como subsídio à operação de reservatórios (BRAVO, 2010), métodos de simulação estocástica para análise de risco na operação de reservatórios (HUANG et al., 2018) e modelos de inteligência artificial empregados para simular a operação do reservatório (ZHANG et al., 2018). Independente do método utilizado, o objetivo comum em todas as propostas que vêm sendo desenvolvidas é tornar mais confiável a decisão dos gestores frente as adversidades hidrológicas.

Assim, verifica-se que a utilização de metodologias estatísticas como ferramentas de

suporte à gestão ambiental tem se tornado cada vez mais comum. Nesse sentido, de forma a complementar as metodologias existentes e já incorporadas aos sistemas, propõe-se a utilização dos gráficos de controle KARMA robustos no gerenciamento de variáveis hidrológicas. Esta ferramenta pode ser uma importante aliada, capaz de identificar crises hídricas antes mesmo do estabelecimento da mesma, garantindo um tempo maior que o usual para tomada de decisões.

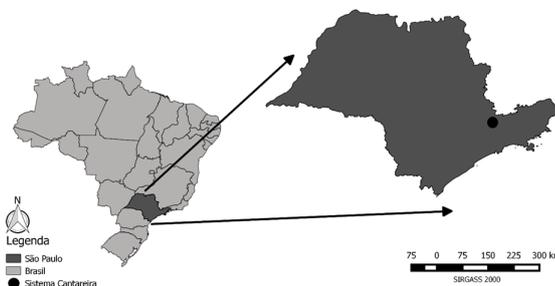
5.4 ESTUDO DE CASOS

Nesta seção serão apresentadas duas aplicações em séries de monitoramento de reservatórios, a fim de evidenciar a potencialidade dos gráficos de controle KARMA robustos como ferramenta de suporte às tomadas de decisões nestes sistemas. O primeiro estudo será realizado considerando o volume útil no Sistema Cantareira, destinado ao abastecimento público, enquanto o segundo no reservatório da usina hidrelétrica de Furnas.

5.4.1 Sistema Cantareira

Em 2014, o Sistema Cantareira, responsável pelo abastecimento da região metropolitana de São Paulo, enfrentou uma expressiva crise hídrica (SABESP, 2015). A Figura 26 apresenta a localização deste sistema.

Figura 26: Mapa de localização Sistema Cantareira.



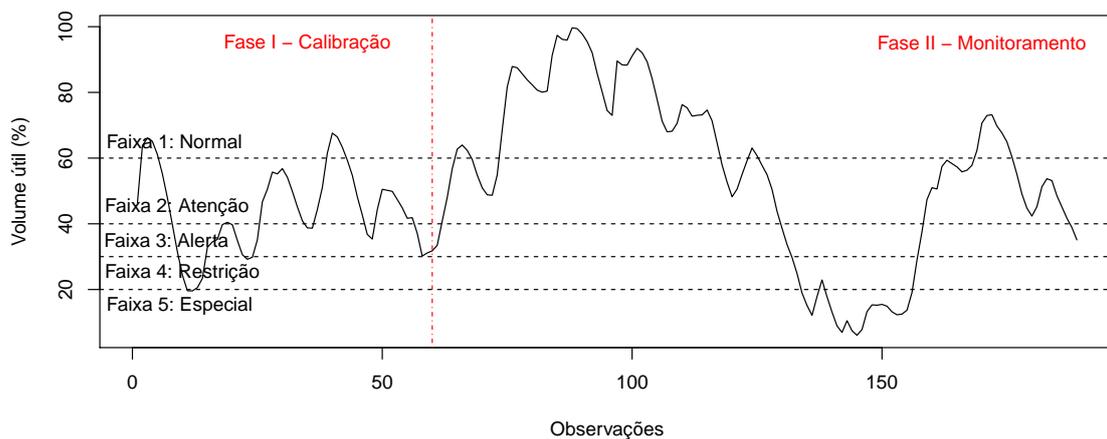
Fonte: Elaborado pela autora.

Como causas para esta crise hídrica apontou-se uma estiagem severa ocorrida na região Sudeste agravada pela fragilidade da gestão hídrica do sistema (ANA, 2017a). Foram registra-

das chuvas abaixo da média na região do Sistema Cantareira, representando uma queda de 55% em relação a média de longo termo (SABESP, 2015).

A região metropolitana de São Paulo é uma das principais regiões do país. De acordo com o último levantamento do Instituto Brasileiro de Geografia e Estatística (IBGE), feito em 2017, essa região respondia por 18% do Produto Interno Bruto (PIB) do país (IBGE, 2020) e concentrava aproximadamente 10% da população brasileira (IBGE, 2019), a tomada de decisão neste sistema de abastecimento é de suma importância. E nesse sentido, o uso de gráficos de controle KARMA robustos, para auxiliar o processo de tomada de decisão, se faz interessante. A aplicação versa sobre a crise hídrica ocorrida em 2014 no Sistema Cantareira, aplicando os gráficos de controle propostos aos percentuais de volume útil mensal obtidos do banco de dados mantido pela Sabesp (SABESP, 2018). O período analisado foi de Janeiro de 2003 a Setembro de 2018. A Figura 27 apresenta o gráfico da série de observações deste período e as faixas autorizadas para captação no sistema. Pode-se observar que próximo ao ano de 2015 os níveis de água alcançaram índices menores que 20% do volume útil, classificada como faixa especial.

Figura 27: Série temporal do volume útil mensal do Sistema Cantareira de Janeiro/03 a Setembro/18.

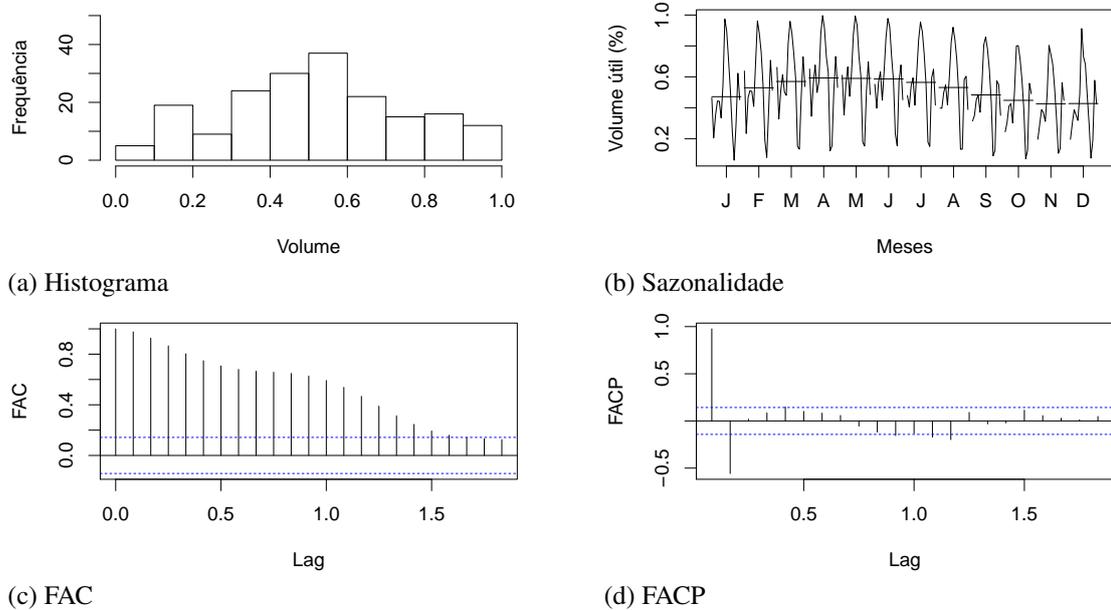


Fonte: Elaborado pela autora.

Como análise descritiva preliminar da série de dados, avaliou-se a média, mediana, máximo e mínimo não condicionais no período completo. A média da série de dados é aproximadamente 52%, a mediana 51%, os valores máximo e mínimo das observações são, respectivamente 99% e 6%. A série possui um desvio padrão próximo a 0,25. O histograma, a sazonalidade, a

função de autocorrelação (FAC) e a função autocorrelação parcial (FACP) podem ser analisadas na Figura 28.

Figura 28: Gráficos da série de observações do volume útil mensal do Sistema Cantareira.



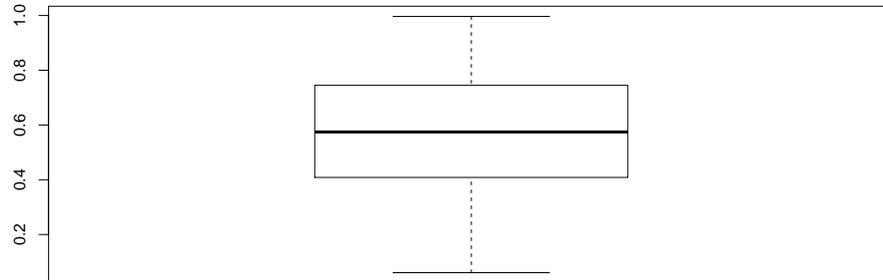
Fonte: Elaborado pela autora.

No histograma (Figura 28 (a)) é possível verificar uma grande amplitude nos valores das observações, com concentração próximo a média, onde a frequência apresentou valor maior que 30. Outro fenômeno identificado foi a presença de sazonalidade na série de dados, evidenciado na análise apresentada na Figura 28 (b), onde é possível verificar uma oscilação na média de observações em cada mês. Além disso, identificou-se, a partir dos gráficos FAC e FACP, a presença de autocorrelação na série de dados e, como consequência, a necessidade de utilização de gráficos de controle residuais baseadas em modelos de séries temporais.

Ajustou-se o modelo KARMA (1,1) com estimadores robusto e usual, com covariável para modelar a sazonalidade, na Fase I. O ajuste se deu nas primeiras 60 observações ($n_1 = 60$), representando 5 anos, e a modelagem da covariável β para modelar a sazonalidade foi dada a partir de um vetor cosseno com período igual a 12. A Figura 29 apresenta o diagrama de caixa da série com as 60 primeiras observações, onde pode ser observado uma certa simetria, com mediana próxima a média da série, sem identificar qualquer *outlier*.

Apesar disso, para a avaliação dos gráficos de controle KARMA robustos adotou-se $p = 0,01$ para gerar os limites da ponderação, em função do bom desempenho apresentado no

Figura 29: Diagrama de caixa - Fase I.



Fonte: Elaborado pela autora.

Capítulo 4. Dos modelos testados, o que apresentou menor valor para os critérios de Akaike (AIC) (AKAIKE, 1974) foi utilizado para estimar os parâmetros apresentados nas Tabelas 5.1 e 5.2. Verificou-se que alguns parâmetros estimados a partir do método usual apresentaram valores distantes dos parâmetros do estimador robusto. Por exemplo, o parâmetro α apresenta uma redução no valor estimado pelo método usual de quase 50% quando comparado com a estimativa robusta. Além do α , o φ , com um aumento de 30% quando comparada a estimativa usual com a robusta.

Tabela 5.1: Parâmetros do modelo KARMA usual (1,1) ajustado à série de observações.

Parâmetro	Estimativa	p-valor
α	0,0974	0,0194
ϕ_1	1,0302	0,0001
θ_1	0,6863	0,0001
φ	13,8579	0,0001
β_1	-0,6341	0,0001
AIC	-220,9874	

Fonte: Elaborado pela autora.

O gráfico de controle KARMA baseado no estimador de máxima verossimilhança ponderada identificou duas observações com valor fora do intervalo definido na Equação 4.2, atribuindo um peso ω_i capaz de minimizar seus efeitos nas inferências sobre os parâmetros do modelo ajustado à série de dados. A primeira observação ponderada recebeu $\omega = 0,376$ e a segunda, $\omega = 0,827$ no ajuste do modelo KARMA robusto.

Tendo em vista o bom desempenho dos gráficos de controle KARMA baseados no resí-

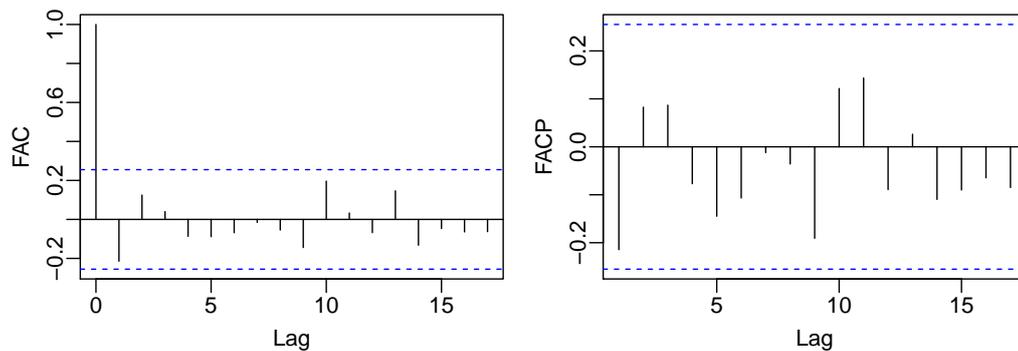
Tabela 5.2: Parâmetros do modelo KARMA robusto (1,1) ajustado à série de observações.

Parâmetro	Estimativa	p-valor
α	0,0459	0,2711
ϕ_1	0,9656	0,0001
θ_1	0,5729	0,0001
φ	18,0459	0,0001
β_1	-0,6375	0,0001
AIC	-250,8924	

Fonte: Elaborado pela autora.

duo quantílico, esse foi adotado para o estudo de caso. Os gráficos da FAC e FACP dos resíduos de cada ajuste são apresentados nas Figuras 30 e 31. Além disso, a fim de verificar o comportamento i.i.d. nas séries residuais, aplicou-se o teste Ljung - Box (LJUNG; BOX, 1978), o qual não rejeitou a hipótese de serem i.i.d. ao nível de 5% ($p\text{-valor} = 0,6089$). Assim, das análises de FAC e FACP dos resíduos e a aplicação do teste Ljung-Box, é possível concluir que as séries residuais são i.i.d. e não apresentam mais o fenômeno de sazonalidade, sendo possível a aplicação dos gráficos de controle de Shewhart e CUSUM para o monitoramento e a identificação de observações fora de controle.

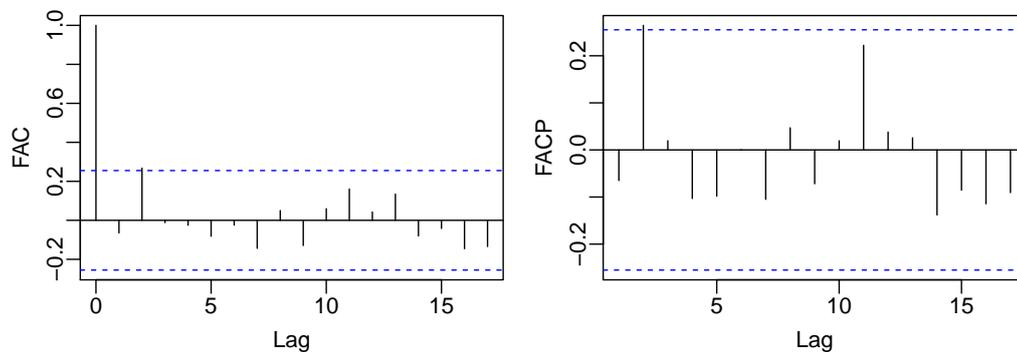
Figura 30: FAC e FACP dos resíduos do modelo KARMA com estimador usual.



Fonte: Elaborado pela autora.

Aplicou-se os gráficos de controle à série residual de $n_2 = 117$ observações, considerando os parâmetros dos limites de controle para um evento com TR de 370 meses, ou seja, $ARL_0 = 370$ e obteve-se os resultados apresentados nas Figuras 32 (b) e (c) para o modelo KARMA usual e nas Figuras 32 (d) e (e) para o modelo KARMA robusto. É possível observar que o gráfico de Shewhart KARMA usual identificou um ponto fora do limite inferior de controle apenas no tempo 141 (Setembro/2014), enquanto o robusto indicou a existência de alguma

Figura 31: FAC e FACP dos resíduos do modelo KARMA com estimador robusto.



Fonte: Elaborado pela autora.

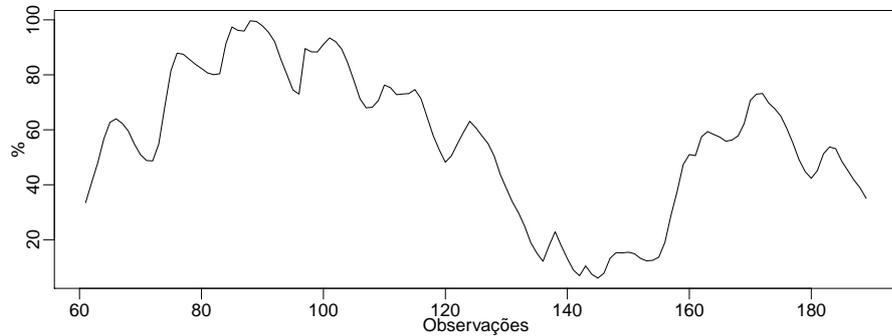
anomalia no tempo 133 (Janeiro/2014). Nos gráficos CUSUM isso também fica evidente, sendo que o gráfico robusto antecipou a tomada de decisão em um mês.

A partir da análise gráfica é possível verificar a permanência dos gráficos KARMA robustos fora dos limites de controle em 2014 e 2015, identificando a manutenção da situação dos baixos percentuais de volume de água nos reservatórios por este período, apesar da elevação da oferta hídrica no decorrer de 2015. Registros técnicos da operação do Sistema Cantareira mostram que em Maio de 2014 o sistema passou a operar com o volume morto (primeira reserva), em Outubro do mesmo ano, passou para a segunda reserva e apenas em Dezembro de 2015 foi possível deixar de utilizá-las (ANA, 2017a). Desta forma, fica evidente a maior habilidade dos gráficos de controle KARMA robustos, que foram capazes de identificar um comportamento mais próximo da realidade do Sistema Cantareira nos anos de 2014 e 2015. O modelo KARMA robusto apresentou resultados melhores em relação ao gráfico KARMA usual, adiantando 1 mês no caso dos gráficos de controle CUSUM e 8 meses nos gráficos de Shewhart a identificação da crise hídrica no sistema. Essa antecipação permite às equipes gestoras um melhor planejamento, auxiliando na tomada de decisões estratégicas.

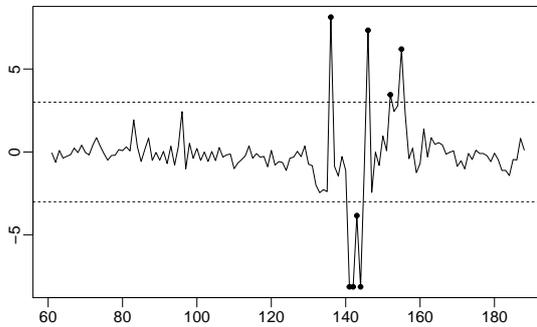
5.4.2 Reservatório da usina hidrelétrica de Furnas

No ano de 2014 também foram identificados períodos críticos na geração energética do reservatório de Furnas. Este reservatório possui um volume de operação de 17, 217 bilhões m^3 , que atinge mais de 30 municípios do país (ANA, 2018). Conforme o boletim de monitoramento de 2014 e 2015 (ANA, 2014, 2015) do Reservatório de Furnas, as precipitações registradas em

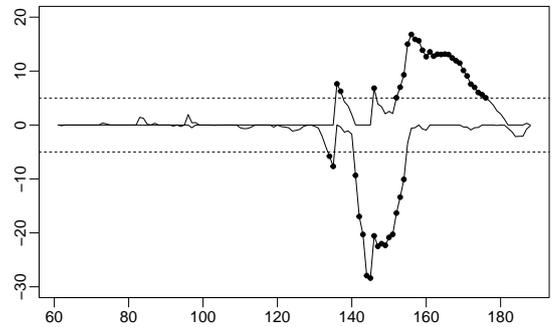
Figura 32: Gráficos de controle da série de observações do volume útil mensal do Sistema Cantareira.



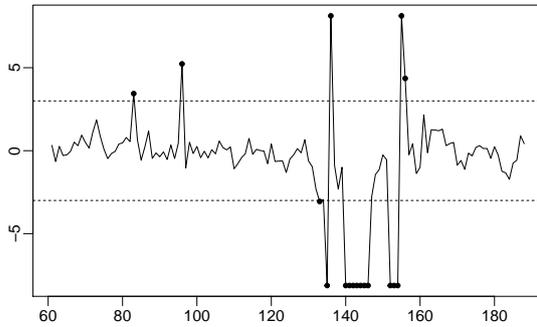
(a) Série monitorada.



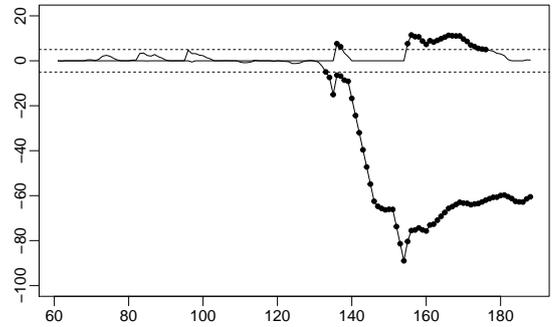
(b) Gráfico de controle Shewhart KARMA usual.



(c) Gráfico de controle CUSUM KARMA usual.



(d) Gráfico de controle de Shewhart KARMA robusto.



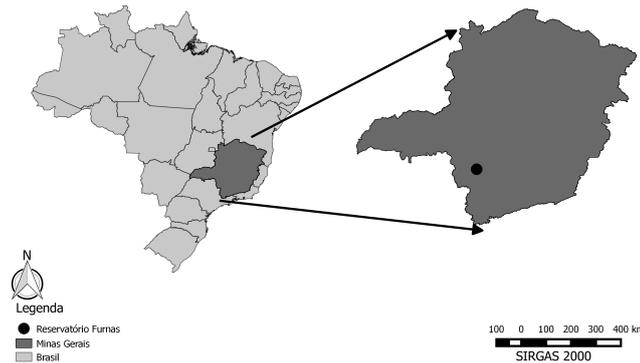
(e) Gráfico de controle CUSUM KARMA robusto.

Fonte: Elaborado pela autora.

praticamente todos os meses foram menores que a média de longo termo. As precipitações acumuladas mensais foram inferiores a 125mm durante o ano todo, exceto para Dezembro de 2014. Os meses de Dezembro e Janeiro, por exemplo, costumam ser os mais úmidos, com médias de longo termo de cerca de 260mm , porém, em Dezembro de 2013 a precipitação acumulada foi inferior a 170mm , enquanto a de Janeiro de 2014 não chegou a 80mm . Essa baixa nos índices de precipitação resultou em uma redução de mais de 25% da área alagada pelo reservatório entre 2013 e 2014 (GONÇALVES et al., 2015). Como consequência, a geração de

energia nesse período foi reduzida, em 2014 e em 2015 foi gerado, respectivamente, 59% e 28% da média de geração anual de 2000 a 2019 (ONS, 2020). A Figura 33 mostra a localização do reservatório de Furnas.

Figura 33: Mapa de localização reservatório Furnas.



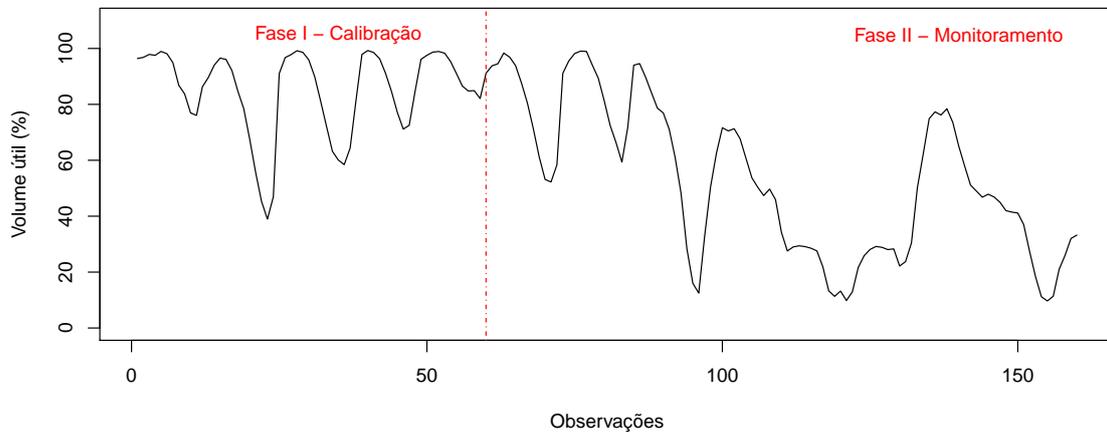
Fonte: Elaborado pela autora.

A fim de verificar a eficácia dos gráficos de controle propostos neste trabalho como ferramenta de auxílio na tomada de decisão em reservatórios destinados à produção de energia, os gráficos de controle KARMA robustos e usual foram aplicados à série de observações de percentual de volume armazenado do reservatório de Furnas. A série utilizada foi obtida do banco de dados mantidos pela ONS, do período de Janeiro 2005 a Abril de 2018 (ONS, 2019b). A Figura 34 apresenta o comportamento da série neste período.

Da Figura 34 é possível verificar uma baixa acentuada no mês de Dezembro do ano de 2012, chegando a um volume próximo de 10%. Além disso, durante todo o período de Dezembro de 2013 a Dezembro de 2015, foram observados índices menores que 50%. A média nesta série de dados foi de 65% e a mediana 71%, apresentando um desvio padrão de 0,28. O máximo e mínimo nesta série de dados foi, respectivamente, 99% e 9%. A Figura 35 apresenta o histograma, a análise de sazonalidade, a FAC e FACP da série do percentual de volume útil.

É possível verificar no histograma apresentado na Figura 35 (a) uma acumulação das observações no intervalo de 0,9 a 1, concentrando cerca de 27% dos dados, ou seja, com volume acima dos 90%. Além disso, identificou-se na série de observações certa dependência sazonal, evidenciada na Figura 35 (b), e das Figura 35 (c) e (d) foi possível verificar a presença do fenômeno de autocorrelação. Desta forma, faz-se necessária a aplicação de modelos estatísticos

Figura 34: Porcentagem do volume útil mensal reservatório Furnas no período de Janeiro/05 a Abril/18.



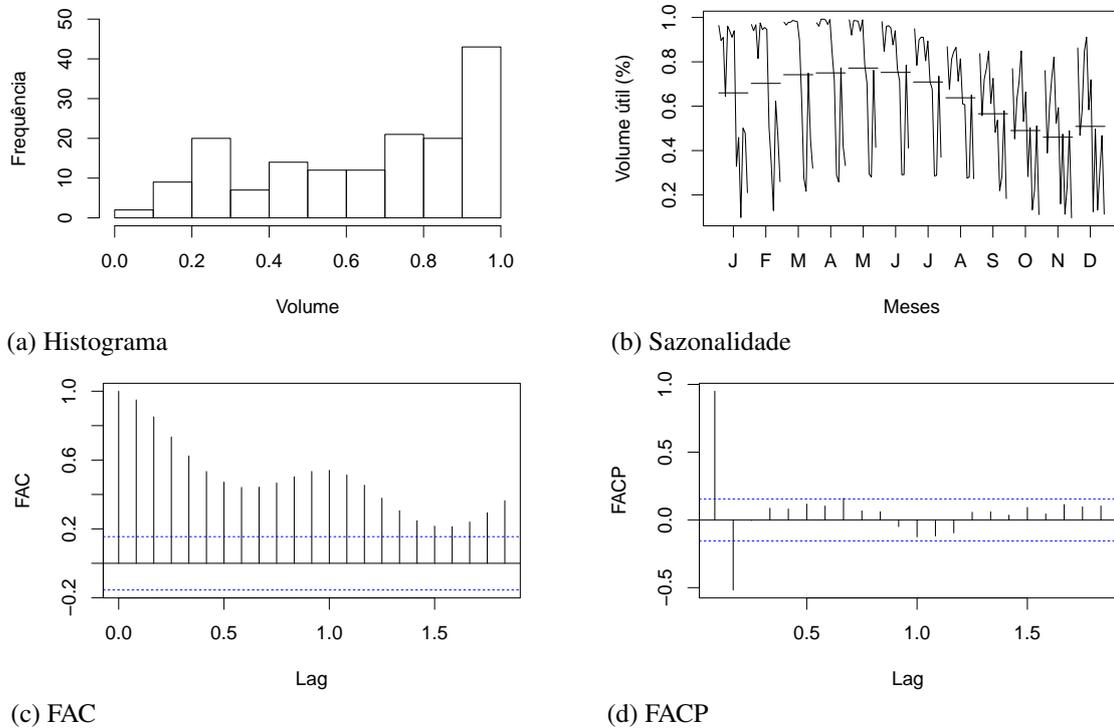
Fonte: Elaborado pela autora.

para obtenção de resíduos livres destes fenômenos. Assim, ajustou-se à série de dados o modelo KARMA (1,1) robusto e usual com covariável para modelar a sazonalidade na Fase I. O ajuste foi dado para as primeiras 60 observações ($n_1 = 60$) e a modelagem da covariável β foi dada a partir de um vetor seno com período de 12 meses. A Figura 36 apresenta o diagrama de caixa da série utilizada na Fase I, onde pode ser observado uma assimetria com mediana deslocada da média da série e presença de *outliers* próximo aos valores mínimos da série.

Assim, no caso dos gráficos de controle KARMA robustos, utilizou-se de um $p = 0,01$ para os limites de ponderação. Na escolha do melhor modelo, o que apresentou o menor valor de AIC foi considerado o modelo de melhor ajuste. As Tabelas 5.3 e 5.4 apresentam os parâmetros estimados para ambos modelos. É possível verificar uma certa diferença entre as estimativas usuais e robustas dos parâmetros do modelo KARMA. A maior diferença observada é na estimativa do parâmetro φ , sendo a estimativa robusta 200% maior que a estimativa usual.

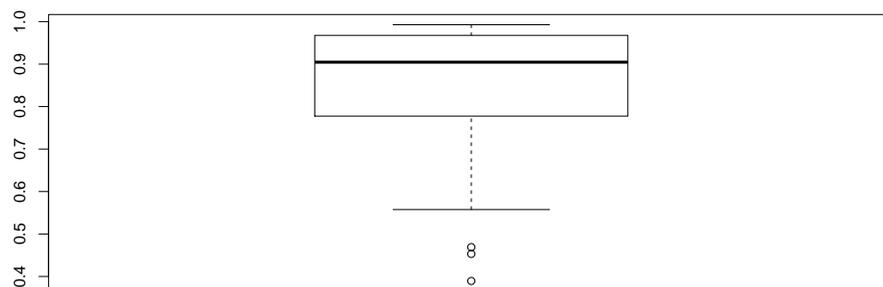
Para a estimativa dos parâmetros pelo método robusto, assim como no estudo de caso do Sistema Cantareira, identificou duas observações com valores fora do intervalo definido na Equação 4.2. Ambas observações foram suavizadas por um peso ω_i proporcional ao efeito causado por cada uma. A primeira recebeu $\omega = 0,526$ e a segunda $\omega = 0,095$. Essas ponderações são reflexo da subida acentuada ocorrida no período próximo a 2007, onde a porcentagem do volume útil passa de 46,9% para 91,0% no intervalo de um mês.

Figura 35: Gráficos da série de observações da porcentagem do volume útil mensal do reservatório de Furnas.



Fonte: Elaborado pela autora.

Figura 36: Diagrama de caixa - Fase I.



Fonte: Elaborado pela autora.

Após ajustado o modelo KARMA à série de dados, avaliou-se os resíduos por meio da FAC e FACP e teste de Ljung - Box. As Figuras 37 e 38 apresentam os gráficos da FAC e FACP residuais para o ajuste KARMA usual e KARMA robusto, respectivamente. Como

Tabela 5.3: Parâmetros estimados do modelo KARMA usual (1,1) ajustado à série de observações.

Parâmetro	Estimativa	p-valor
α	0,8564	0,0031
ϕ_1	0,4891	0,0017
θ_1	0,5391	0,0131
φ	11,7183	0,0001
β_1	1,4211	0,0001
AIC	-179,1074	

Fonte: Elaborado pela autora.

Tabela 5.4: Parâmetros estimados do modelo KARMA robusto (1,1) ajustado à série de observações.

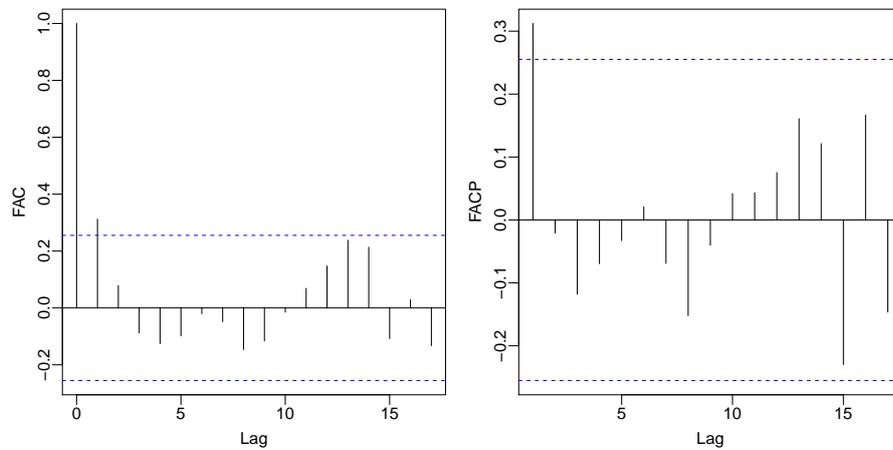
Parâmetro	Estimativa	p-valor
α	0,5885	0,0422
ϕ_1	0,6756	0,0001
θ_1	0,2311	0,2876
φ	40,1608	0,0001
β_1	0,9057	0,0043
AIC	-268,6976	

Fonte: Elaborado pela autora.

resultado da aplicação do teste Ljung - Box, não rejeitou-se a hipótese de serem i.i.d. ao nível de 5% ($p\text{-valor} = 0,6863$). Assim, dados resultados da análise residual, verifica-se a remoção da autocorrelação e sazonalidade nos resíduos tanto para KARMA usual quanto para KARMA robusto, possibilitando a aplicação dos gráficos de controle de Shewhart e CUSUM na série de resíduos para identificação de observações fora de controle.

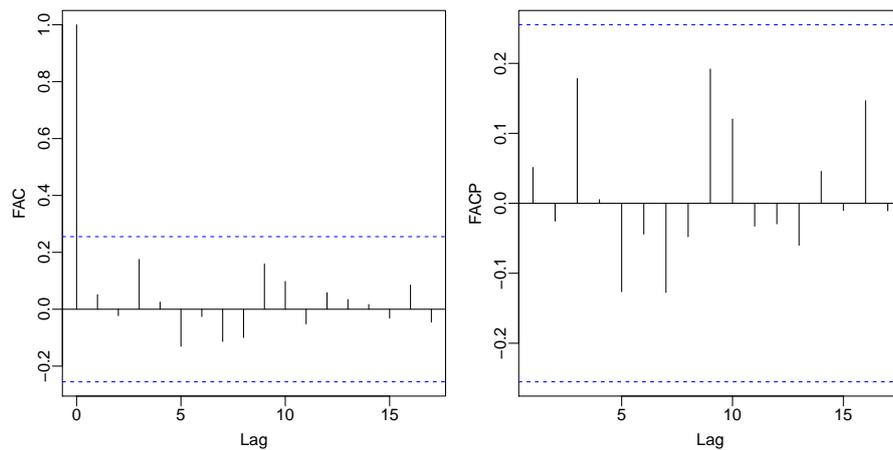
Aplicou-se os gráficos de controle à série residual do restante do período analisado ($n_2 = 100$) e avaliou-se a identificação das observações fora de controle considerando os limites estabelecidos para eventos com TR de 370 meses, isto é, $ARL_0 = 370$. As Figuras 39 (b) e (c) apresentam os gráficos para o modelo KARMA usual e as Figuras 39 (d) e (e) para o modelo KARMA robusto. Verificou-se que todos os gráficos identificaram anomalias no período de 2014 e 2015. No entanto, a crise estabelecida nestes anos parece ter iniciado nos anos anteriores. Foi possível verificar que as primeiras observações fora de controle identificadas no monitoramento dos gráficos foram no ano de 2012. Este comportamento é visível nos gráficos de controle KARMA robustos (Figuras 39 (d) e (e)) e no gráfico de controle CUSUM KARMA usual (Figura 39 (c)). O gráfico de controle de Shewhart KARMA usual (Figura 39 (b)) foi o que apresentou menor persistência das observações fora de controle, identificando apenas

Figura 37: FAC e FACP dos resíduos do modelo KARMA com estimador usual.



Fonte: Elaborado pela autora.

Figura 38: FAC e FACP dos resíduos do modelo KARMA com estimador robusto.

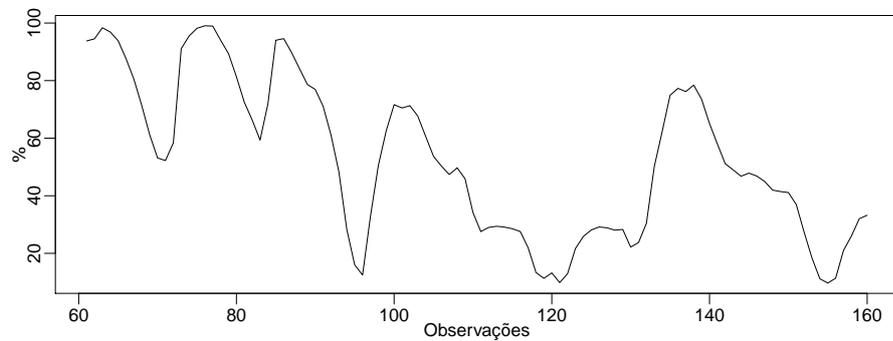


Fonte: Elaborado pela autora.

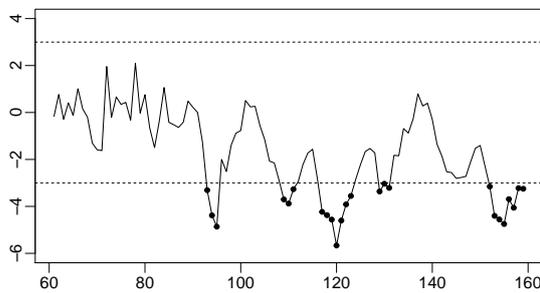
observações alternadas e não períodos fora de controle como os demais.

Apesar de todos os gráficos de controle terem identificado as anomalias na série de volume armazenado, foi verificado que os gráficos de controle KARMA robustos identificaram um mês antes o período fora de controle quando comparado ao gráfico de controle de Shewhart KARMA usual e dois meses antes quando comparado com o gráfico de controle CUSUM KARMA usual. A identificação de anomalias com antecedência, mesmo de poucos meses, pode fazer grandes diferença para a gestão de uma usina hidrelétrica, facilitando a tomada de decisão. No ano de 2012 a hidrelétrica de Furnas gerou aproximadamente $6.880GWh$, a maior geração de energia desde 2000 (ONS, 2020). Portanto, se medidas de gestão tivessem

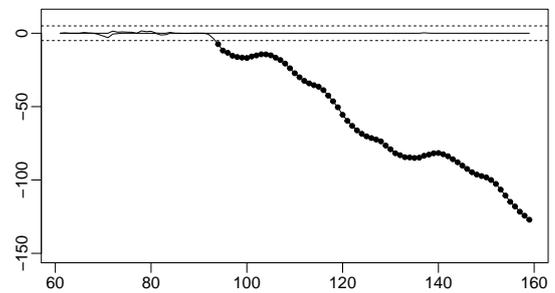
Figura 39: Gráficos de controle da série de observações do volume útil mensal do reservatório Furnas.



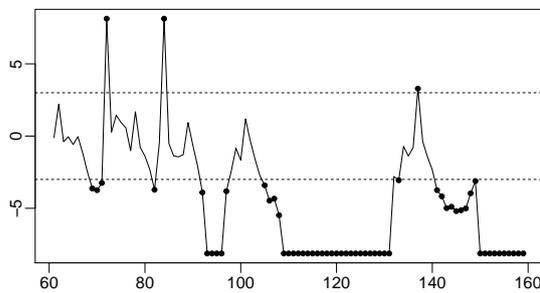
(a) Série monitorada.



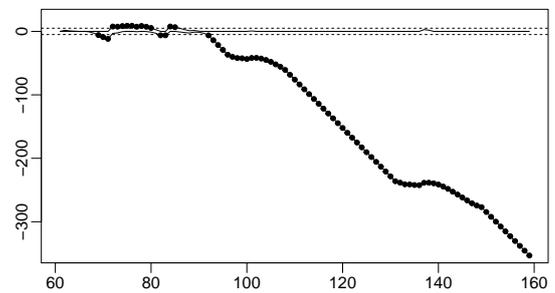
(b) Gráfico de controle Shewhart KARMA usual.



(c) Gráfico de controle CUSUM KARMA usual.



(d) Gráfico de controle de Shewhart KARMA robusto.



(e) Gráfico de controle CUSUM KARMA robusto.

Fonte: Elaborado pela autora.

sendo tomadas logo que identificadas tais anomalias no ano de 2012, possivelmente os reservatórios não teriam reduzido 25% da sua área alagada. Além disso, a situação teria sido normalizada com maior controle e hoje o sistema poderia estar operando com melhores condições.

5.5 CONCLUSÃO

As análises realizadas, baseada em séries de dados com fatos já estabelecidos, permitiram a identificação de pontos favoráveis e desfavoráveis nas metodologias utilizadas. Ambas metodologias apresentaram comportamentos satisfatórios, identificando observações fora de controle antes dos níveis de água chegarem a volumes críticos do sistema. No entanto, foi através dos gráficos de controle KARMA robustos que verificou-se maior agilidade na identificação de períodos fora de controle. Foram constatadas ponderações de observações discrepantes em ambos estudos garantindo estimativas mais acuradas ao modelo KARMA robusto. Assim, a partir desta aplicação, fica evidenciada a boa performance dos gráficos de controle KARMA robustos no monitoramento de séries hidrológicas.

6 CONCLUSÕES GERAIS

Esta dissertação buscou desenvolver ferramentas de monitoramento que auxiliem na gestão ambiental e tomada de decisão, principalmente na operação de reservatórios. Para isso, utilizou-se como base metodologias estatísticas de monitoramento de dados e modelos de séries temporais. Nesse sentido, foram propostos gráficos de controle baseados nos resíduos do modelo KARMA e avaliados via simulação de Monte Carlo em três diferentes cenários e dois tamanhos amostrais. Sabendo da presença de *outliers* em variáveis ambientais e hidrológicas e, ainda, do efeito que estas causam nas estimativas dos parâmetros dos modelos, foi proposta uma melhoria aos gráficos KARMA, utilizando estimadores de máxima verossimilhança ponderada. Os gráficos robustos a *outliers* foram avaliados via simulação de Monte Carlo em dois diferentes cenários e dois tamanhos amostrais. Por fim, aplicou-se e avaliou-se as ferramentas propostas nesta dissertação a dados de dois sistemas de reservação de grande importância no país, Sistema Cantareira e hidrelétrica de Furnas.

Foi possível verificar o melhor desempenho dos gráficos de controle KARMA quando comparados com o desempenho de outras metodologias já propostas. Os gráficos de controle KARMA apresentaram baixas distorções na avaliação de ARL_0 e desempenhos equiparáveis aos demais modelos analisados na avaliação de ARL_1 . No entanto, na presença de *outliers* os gráficos KARMA tiveram seus limites de controle afetados, apresentando maiores distorções quando comparado a proposta robusta. Os gráficos de controle KARMA robustos identificaram as observações discrepantes e as ponderaram de forma a suavizar seus efeitos nas inferências sobre o modelo ajustado. Portanto, a partir dos estimadores de máxima verossimilhança ponderada foi possível obter estimativas mais acuradas na Fase I dos gráficos e, como consequência, limites de controle no monitoramento da Fase II com menores distorções.

Na aplicação dos gráficos de controle propostos a dados reais do Sistema Cantareira e Furnas verificou-se uma maior agilidade dos gráficos robustos na identificação de observações fora de controle. Essa antecipação da identificação de situações fora de controle poderiam ter auxiliado a gestão e tomada de decisão desses sistemas a fim de evitar as crises hídricas que se estabeleceram. Os sistemas de reservação de água, tanto para abastecimento quanto para geração de energia, são de extrema importância no país, sendo trivial o desenvolvimento de ferramentas que auxiliem na tomada de decisões e minimizem os riscos frente a variações hídricas. Nesse sentido, os gráficos de controle KARMA robustos são apresentados como uma

promissora alternativa ao monitoramento de nível de reservatórios de água.

REFERÊNCIAS

- ABRAHAM, B.; BOX, G. E. P. Bayesian analysis of some outlier problems in time series. **Biometrika**, v.66, n.2, p.229–236, 1979.
- AGOSTINELLI, C. Robust time series estimation via weighted likelihood. In: **Developments in Robust Statistics**. : Springer. p.1–16, 2003.
- AGOSTINELLI, C.; BISAGLIA, L. ARFIMA processes and outliers: a weighted likelihood approach. **Journal of Applied Statistics**, v.37, n.9, p.1569–1584, 2010.
- AKAIKE, H. A new look at the statistical model identification. **IEEE Transactions on Automatic Control**, v.19, n.6, p.716–723, 1974.
- ALI, S.; PIEVATOLO, A.; GÖB, R. An overview of control charts for high-quality processes. **Quality and Reliability Engineering International**, v.32, n.7, p.2171–2189, 2016.
- ALLAWI, M. F.; OTHMAN, J.; HAMZAH, F. M.; MOHD, N. S.; DEO, R. C.; EL-SHAFIE, A. Reservoir inflow forecasting with a modified coactive neuro-fuzzy inference system: a case study for a semi-arid region. **Theoretical and Applied Climatology**, v.134, n.1, p.545–563, 2018.
- ALWAN, L. C.; ROBERTS, H. V. Time-series modeling for statistical process control. **Journal of Business & Economic Statistics**, v.6, n.1, p.87–95, 1988.
- ANA. **Atlas Brasil: abastecimento urbano de água: panorama nacional**. : Agência Nacional de Águas, 2010.
- ANA. **Boletim de Monitoramento do Reservatório de Furnas**. : Agência Nacional de Águas, 2014.
- ANA. **Boletim de Monitoramento do Reservatório de Furnas**. : Agência Nacional de Águas, 2015.
- ANA. **Conjuntura dos Recursos Hídricos no Brasil 2017**. : Agência Nacional de Águas, 2017.

ANA. **Resolução Conjunta ANA/DAEE nº 925**. <http://www3.ana.gov.br/portal/ANA/gestao-da-agua/resolucoes-e-normativos/resolucoes-de-uso>.

Acesso em 19 de Fevereiro de 2019.

ANA. **Boletim de Monitoramento do Reservatório de Furnas**. : Agência Nacional de Águas, 2018.

ANDERSON, M. J.; THOMPSON, A. A. Multivariate control charts for exological and environmental monitoring. **Ecological Applications**, v.14, n.6, p.1921–1935, 2004.

ANEEL. **Capacidade de Geração do Brasil**. <https://www2.aneel.gov.br/aplicacoes/capacidadebrasil/capacidadebrasil.cfm>. Acesso em 25 de Novembro de 2019.

ARAUJO LIMA FILHO, L. M. de; BAYER, F. M. Kumaraswamy control chart for monitoring double bounded environmental data. **Communications in Statistics - Simulation and Computation**, v.0, n.0, p.1–16, 2019.

BAYER, D. M.; BAYER, F. M.; GAMBA, P. **A Three-dimensional Spatio-temporal Model for Remote Sensing Data Cubes**. Preprint, 2020.

BAYER, F. M.; BAYER, D. M.; PUMI, G. Kumaraswamy autoregressive moving average models for double bounded environmental data. **Journal of Hydrology**, v.555, p.385–396, 2017.

BAYER, F. M.; CINTRA, R. J.; CRIBARI-NETO, F. Beta seasonal autoregressive moving average models. **Journal of Statistical Computation and Simulation**, v.88, n.15, p.1–26, 2018.

BAYER, F. M.; TONDOLO, C. M.; MÜLLER, F. M. Beta regression control chart for monitoring fractions and proportions. **Computers & Industrial Engineering**, v.119, p.416–426, 2018.

BESSLER, F. T.; SAVIC, D. A.; WALTERS, G. A. Water reservoir control with data mining. **Journal of Water Resources Planning and Management**, v.129, n.1, p.26–34, 2003.

BIANCO, A. M.; BEN, G.; MARTÍNEZ, E. J.; YOHAI, V. J. Outlier detection in regression models with ARIMA errors using robust estimates. **Journal of Forecasting**, v.20, n.8, p.565–579, 2001.

BORROR, C. M.; MONTGOMERY, D. C.; RUNGER, G. C. Robustness of the EWMA control chart to non-normality. **Journal of Quality Technology**, v.31, n.3, p.309–316, 1999.

BOX, G. E. P.; JENKINS, G. M. **Time Series Analysis: Forecasting and Control**. : Holden-Day, 1976.

BRAVO, J. M. **Subsídios à Operação de Reservatórios Baseada na Previsão de Variáveis Hidrológicas**. Tese (Doutorado em Engenharia de Produção) — Tese (Doutorado em Recursos Hídricos e Saneamento Ambiental), Universidade Federal do Rio Grande do Sul, Porto Alegre, 2010.

CHANDLER, R.; SCOTT, M. **Statistical Methods for Trend Detection and Analysis in the Environmental Sciences**. : John Wiley & Sons, 2011.

CHEN, C.; LIU, L.-M. Joint estimation of model parameters and outlier effects in time series. **Journal of the American Statistical Association**, v.88, n.421, p.284–297, 1993.

CHIANG, J. Y.; LIO, Y. L.; NG, H. K. T.; TSAI, T. R. Robust control charts for percentiles based on location-scale family of distributions. **Quality and Reliability Engineering International**, v.34, n.6, p.1059–1078, 2018.

CHOW, V. T. **Frequency Analysis of Hydrologic Data with Special Application to Rainfall Intensities**. : University of Illinois at Urbana Champaign, College of Engineering. Engineering Experiment Station., 1953.

CINAR, A.; PALAZOGLU, A.; KAYIHAN, F. **Chemical Process Performance Evaluation**. : CRC Press, 2007.

CROUX, C.; GELPER, S.; MAHIEU, K. Robust control charts for time series data. **Expert Systems with Applications**, v.38, n.11, p.13810–13815, 2011.

DENBY, L.; MARTIN, R. D. Robust estimation of the first-order autoregressive parameter. **Journal of the American Statistical Association**, v.74, n.365, p.140–146, 1979.

EISSA, M. E. A. The attribute control charts for outbreak trends of selected states in the USA: a brief report of the insight into the pattern. **International Medicine**, v.1, p.11–14, 2019.

- EISSA, M.; MAHMOUD, A.; NOUBY, A. Control chart in microbiological cleaning efficacy of pharmaceutical facility. **Dhaka University Journal of Pharmaceutical Sciences**, v.14, n.2, p.133–138, 2016.
- ESPINHEIRA, P. L.; FERRARI, S. L. P.; CRIBARI-NETO, F. On beta regression residuals. **Journal of Applied Statistics**, v.35, n.4, p.407–419, 2008.
- EWAN, W. D. When and how to use Cu-Sum charts. **Technometrics**, v.5, n.1, p.1 – 22, 1963.
- FELLNER, W. H. Average run lengths for cumulative sum schemes. **Journal of the Royal Statistical Society**, v.39, n.3, p.402 – 412, 1990.
- FIELD, C.; SMITH, B. Robust Estimation: a weighted maximum likelihood approach. **International Statistical Review**, v.62, n.3, p.405–424, 1994.
- FILHO, R. L. O. **Controle Estatístico de Qualidade**. : LTC-Livros técnicos e científicos Editora S.A., 1982.
- FOX, A. J. Outliers in time series. **Journal of the Royal Statistical Society. Series B (Methodological)**, v.34, n.3, p.350–363, 1972.
- GANJI, A.; PONNAMBALAM, K.; KHALILI, D.; KARAMOUZ, M. Grain yield reliability analysis with crop water demand uncertainty. **Stochastic Environmental Research and Risk Assessment**, v.20, n.4, p.259–277, 2006.
- GARDINER, J. S.; MONTGOMERY, D. C. Using statistical control charts for software quality control. **Quality and Reliability Engineering International**, v.3, p.15–20, 1987.
- GOEL, A. L.; WU, S. M. Determination of A.R.L. and a contour nomogram for CUSUM charts to control normal mean. **Technometrics**, v.13, n.2, p.221–230, 1971.
- GONÇALVES, T. G.; MACIEL, D. A.; ALVES, H. M. R.; VOLPATO, M. M. L.; VIEIRA, T. G. C.; DANRAS, M. F. Avaliação do nível de água do reservatório de Furnas nos anos de 2013 e 2014, utilizando imagens Landsat-8. In: XVII SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO. **Anais**. p.1502–1507, 2015.
- GOUGH, J. D.; WARD, J. C. Environmental Decision-Making and Lake Management. **Journal of Environmental Management**, v.48, n.1, p.1 – 15, 1996.

GOVE, A. D.; SADLER, R.; MATSUKI, M.; ARCHIBALD, R.; PEARSE, S.; GARKAKLIS, M. Control charts for improved decisions in environmental management: a case study of catchment water supply in South-West Western Australia. **Ecological Management & Restoration**, v.14, n.2, p.127–134, 2013.

HAWKINS, D. M. Cumulative SUM control charting: an underutilized SPC tool. **Quality Engineering**, v.5, n.3, p.463–477, 1993.

HILLMER, S. Monitoring and adjusting forecasts in the presence of additive outliers. **Journal of Forecasting**, v.3, n.2, p.205–215, 1984.

HO, L. L.; FERNANDES, F. H.; BOURGUIGNON, M. Control charts to monitor rates and proportions. **Quality and Reliability Engineering International**, v.0, n.0, p.1–10, 2018.

HOUGHTALEN, R. J.; AKAN, A. O. H.; HWANG, N. H. C. **Fundamentals of Hydraulic Engineering Systems**. : Pearson, 2010.

HUANG, K.; YE, L.; CHEN, L.; WANG, Q.; DAI, L.; ZHOU, J.; SINGH, V. S.; HUANG, M.; ZHANG, J. Risk analysis of flood control reservoir operation considering multiple uncertainties. **Journal of Hydrology**, v.565, p.672 – 684, 2018.

HYNDMAN, R.; ATHANASOPOULOS, G.; BERGMEIR, C.; CACERES, G.; CHHAY, L.; O'HARA-WILD, M.; PETROPOULOS, F.; RAZBASH, S.; WANG, E.; YASMEEN, F.; R CORE TEAM; IHAKA, R.; REID, D.; SHAUB, D.; TANG, Y.; ZHOU, Z. **Package “forecast”**. 2019.

HYNDMAN, R. J.; ATHANASOPOULOS, G. **Forecasting: Principles and Practice**. : OTexts, 2018.

IBGE. **Logística de Energia** : 2015. : Coordenação de Geografia, IBGE, 2016.

IBGE. **Estimativas da População**. <https://www.ibge.gov.br/estatisticas/sociais/populacao/9103-estimativas-de-populacao.html?=&t=resultados>. Acesso em 05 de Dezembro de 2019.

IBGE. **Produto Interno Bruto dos Municípios**. <https://www.ibge.gov.br/estatisticas/economicas/contas-nacionais/>

9088-produto-interno-bruto-dos-municipios.html?t=series-historicas&c=1100148. Acesso em 03 de Janeiro de 2020.

IHA. **2018 hydropower status report - sector trend and insights**. : International Hydropower Association Limited, 2018.

ITAIPU. **Relatório de Sustentabilidade 2015**. : Itaipu binacional, 2015.

JONES-FARMER, L. A.; WOODALL, W. H.; STEINER, S. H.; CHAMP, C. An overview of phase I analysis for process improvement and monitoring. **Journal of Quality Technology**, v.46, n.3, p.265–280, 2014.

JONES, M. C. Kumaraswamy's distribution: a beta-type distribution with some tractability advantages. **Statistical Methodology**, v.6, n.1, p.70–81, 2009.

KARABOGA, D.; BAGIS, A.; HAKTANIR, T. Fuzzy logic based operation of spillway gates of reservoirs during floods. **Journal of Hydrologic Engineering**, v.9, n.6, p.544–549, 2004.

KEDEM, B.; FOKIANOS, K. **Regression Models for Time Series Analysis**. : John Wiley & Sons. v.488, 2005.

KUMARASWAMY, P. Sinpower probability density function. **Journal of Hydrology**, v.31, n.1-2, p.181–184, 1976.

KUMARASWAMY, P. A generalized probability density function for double-bounded random processes. **Journal of Hydrology**, v.46, n.1-2, p.79–88, 1980.

LAWSON, A. B.; KLEINMAN, K. **Spatial and Syndromic Surveillance for Public Health**. : John Wiley & Sons, 2005.

LEE, P.-H.; HUANG, Y. -H; KUO, T. -I; WANG, C. -C. The effect of the individual chart with variable control limits on the river pollution monitoring. **Quality & Quantity**, v.47, n.4, p.1803–1812, 2013.

LJUNG, G. M.; BOX, G. E. P. On a measure of lack of fit in time series models. **Biometrika**, v.65, p.297–303, 1978.

LOHANI, A. K.; KUMAR, R.; SINGH, R. D. Hydrological time series modeling: a comparison between adaptive neuro-fuzzy, neural network and autoregressive techniques. **Journal of Hydrology**, v.442-443, p.23 – 35, 2012.

- MACHIWAL, D.; JHA, M. **Hydrologic Time Series Analysis: theory and practice.** : Springer Netherlands, 2012.
- MANCHINI, C. E. F.; PORTELLA, L.; BAYER, F. M. Estimação robusta dos parâmetros da distribuição beta. **Revista Brasileira de Biometria**, v.37, n.3, p.350–371, 2019.
- MANLY, B. F. J. **Statistics for Environmental Science and Management.** : CRC Press. (Chapman & Hall/CRC Applied Environmental Statistics), 2008.
- MARCHANT, C.; LEIVA, V.; CHRISTAKOS, G.; CAVIERES, M. F. Monitoring urban environmental pollution by bivariate control charts: new methodology and case study in santiago, chile. **Environmetrics**, v.30, n.5, p.e2551, 2019.
- MARONNA, R.; MARTIN, D.; YOHAI, V. **Robust Statistics: theory and methods.** : Wiley, 2006.
- MCCULLAGH, P.; NELDER, J. A. **Generalized linear models.** 2nd.ed. : London: Chapman and Hall, 1989.
- MESNIL, B.; PETITGAS, P. Detection of changes in time-series of indicators using CUSUM control charts. **Aquatic Living Resources**, v.22, n.2, p.187–192, 2009.
- MONTGOMERY, D. C. **Introduction to Statistical Quality Control.** 6.ed. : John Wiley & Sons, Inc., 2009.
- MONTGOMERY, D. C.; RUNGER, G. C. **Applied Statistics and Probability for Engineers.** 4th.ed. : Wiley, 2006.
- NADARAJAH, S. On the distribution of Kumaraswamy. **Journal of Hydrology**, v.348, p.568–569, 2008.
- NALLY, R. M.; HART, B. T. Use of CUSUM methods for water-quality monitoring in storages. **Environmental Science & Technology**, v.31, n.7, p.2114–2119, 1997.
- NGUYEN, T.-D.; WELSCH, R. E. Outlier detection and robust covariance estimation using mathematical programming. **Advances in Data Analysis and Classification**, v.4, n.4, p.301–334, 2010.

ONS. **Procedimentos de rede.** : ONS. Módulo 10 - Manual de Procedimentos da Operação, 2017.

ONS. **Reservatórios.** <http://www.ons.org.br/paginas/energia-agora/reservatorios>. Acesso em 03 de Dezembro de 2019.

ONS. **Histórico da operação.** <http://www.ons.org.br/paginas/resultados-da-operacao/historico-da-operacao>. Acesso em 10 de Agosto de 2019.

ONS. **Geração de energia.** http://www.ons.org.br/Paginas/resultados-da-operacao/historico-da-operacao/geracao_energia.aspx. Acesso em 12 de Janeiro de 2020.

PAGE, E. S. Continuous inspection schemes. **Biometrika**, v.41, n.1/2, p.100–115, 1954.

PAROISSIN, C.; PENALVA, L.; PÉTRAU, A.; VERDIER, G. New control chart for monitoring and classification of environmental data. **Environmetrics**, v.27, n.3, p.182–193, 2016.

PAWITAN, Y. **In All Likelihood: statistical modelling and inference using likelihood.** : Oxford Science Publication, 2001.

PEREIRA, R.; BATISTA, A. L.; NEVES, L. C. Probabilistic model for the representation of the reservoir water level of concrete dams during normal operation periods. **Water Resources Management**, v.32, n.9, p.3041–3052, 2018.

PETERSON, W.; BIRDSALL, T.; FOX, W. The theory of signal detectability. **Transactions of the IRE Professional Group on Information Theory**, v.4, n.4, p.171–212, 1954.

PETITGAS, P. The CUSUM out-of-control table to monitor changes in fish stock status using many indicators. **Aquatic Living Resources**, v.22, n.2, p.201–206, 2009.

PRESS, W. H.; TEUKOLSKY, S.; VETTERLING, W. T.; FLANNERY, B. P. **Numerical Recipes in C: The Art of Scientific Computing.** : Cambridge University Press, 1992.

R Core Team. **R: a language and environment for statistical computing.** Vienna, Austria: R Foundation for Statistical Computing, 2019.

- ROCHA, A. V.; CRIBARI-NETO, F. Beta autoregressive moving average models. **TEST**, v.18, n.3, p.529–545, 2008.
- ROCKE, D. M. Robust control charts. **Technometrics**, v.31, n.2, p.173–184, 1989.
- SABESP. **CHES**: crise hídrica, estratégias e soluções da SABESP para região metropolitana de São Paulo. : Sabesp, 2015.
- SABESP. **Situação dos Mananciais**. <http://www2.sabesp.com.br/mananciais/DivulgacaoSiteSabesp.aspx>. Acesso em 12 de Dezembro de 2018.
- SALAS, J. D. **Applied Modeling of Hydrologic Time Series**. : Water Resources Publications, 1985.
- SANT’ANNA, Â. M. O.; CATEN, C. S. ten. Beta control charts for monitoring fraction data. **Expert Systems with Applications**, v.39, n.11, p.10236–10243, 2012.
- SCHNEIDER, H.; HUI, Y.; PRUETT, J. M. Control Charts for Environmental Data. In: FRONTIERS IN STATISTICAL QUALITY CONTROL 4. **Anais...** Physica-Verlag HD p.216–226, 1992.
- SCHOONHOVEN, M.; DOES, R. J. M. M. A robust Xbar control chart. **Quality and Reliability Engineering International**, v.29, n.7, p.951–970, 2013.
- SCRUCCA, L.; SNOW, G.; BLOOMFIELD, P. **Package “qcc”**, 2017.
- SHEWHART, W. A. **Economic Control of Quality of Manufactured Product**. : New York: Van-Nostrand Reinhold, 1931.
- SILVA, L. P. da. **Gráficos de controle de regressão beta robustos**. Dissertação (Mestrado em Engenharia de Produção) — Dissertação (Mestrado em Engenharia de Produção), Universidade Federal de Santa Maria, Santa Maria, 2019.
- SMETI, E. M.; THANASOULIAS, N.; KOUSOURIS, L. P.; TZOUMERKAS, P. C. An approach for the application of statistical process control techniques for quality improvement of treated water. **Desalination**, v.213, p.273–281, 2007.
- SMETI, E. M.; KORONAKIS, D. E.; GOLFINOPOULOS, S. K. Control charts for the toxicity of finished water - modeling the structure of toxicity. **Water Research**, v.41, p.2679–2689, 2007.

SOUZA, Z. de; SANTOS, A. H. M.; COSTA BORTONI, E. da. **Centrais Hidrelétricas.** : Editora Interciência, 2009.

SUNDAR, V.; SUBBIAH, K. Application of double bounded probability density function for analysis of ocean waves. **Ocean Engineering**, v.16, n.2, p.193–200, 1989.

TIWARI, M. K.; KUMAR, S. Reservoir inflow forecasting using extreme learning machines. In: **Hydrologic Modeling.** : Springer p.565–585, 2018.

TONDOLO, C. M. **Gráficos de Controle para Dados do Tipo Taxas e Proporções Autocorrelacionados.** Dissertação (Mestrado em Engenharia de Produção) — Dissertação (Mestrado em Engenharia de Produção), Universidade Federal de Santa Maria, Santa Maria, 2016.

VARKEY, P.; RELLER, M. K.; RESAR, R. K. Basics of Quality Improvement in Health Care. **Mayo Clinic Proceedings**, v.82, n.6, p.735–739, 2007.

VRIES, A. D.; RENEAU, J. K. Application of statistical process control charts to monitor changes in animal production systems. **Journal of Animal Science**, v.88, n.13, p.E11–E24, 2010.

VUCETIC, D.; SIMONOVIC, S. P. **Water Resources Decision Making Under Uncertainty.** : Department of Civil and Environmental Engineering, The University of Western Ontario (73), 2011.

WANG, B. X.; WANG, X. K.; YU, K. Inference on the Kumaraswamy distribution. **Communications in Statistics - Theory and Methods**, v.46, n.5, p.2079–2090, 2017.

WARDELL, D. G.; MOSKOWITZ, H.; PLANTE, R. D. Control charts in the presence of data correlation. **Management Science**, v.38, n.8, p.1084–1105, 1992.

WMO. **WMO Statement on the State of the Global Climate in 2016.** : WMO (World Meteorological Organization), n.1189, 2016.

WOODALL, W. H. Controversies and contradictions in statistical process control. **Journal of Quality Technology**, v.32, n.4, p.341–350, 2000.

YOU, J.-Y. **Hedging Rule for Reservoir Operation:** How Much, When and How Long to Hedge. : University of Illinois at Urbana-Champaign, 2008.

YU, H.; WU, Z.; BAO, T.; ZHANG, L. Multivariate analysis in dam monitoring data with PCA. **Science China Technological Sciences**, v.53, n.4, p.1088–1097, 2010.

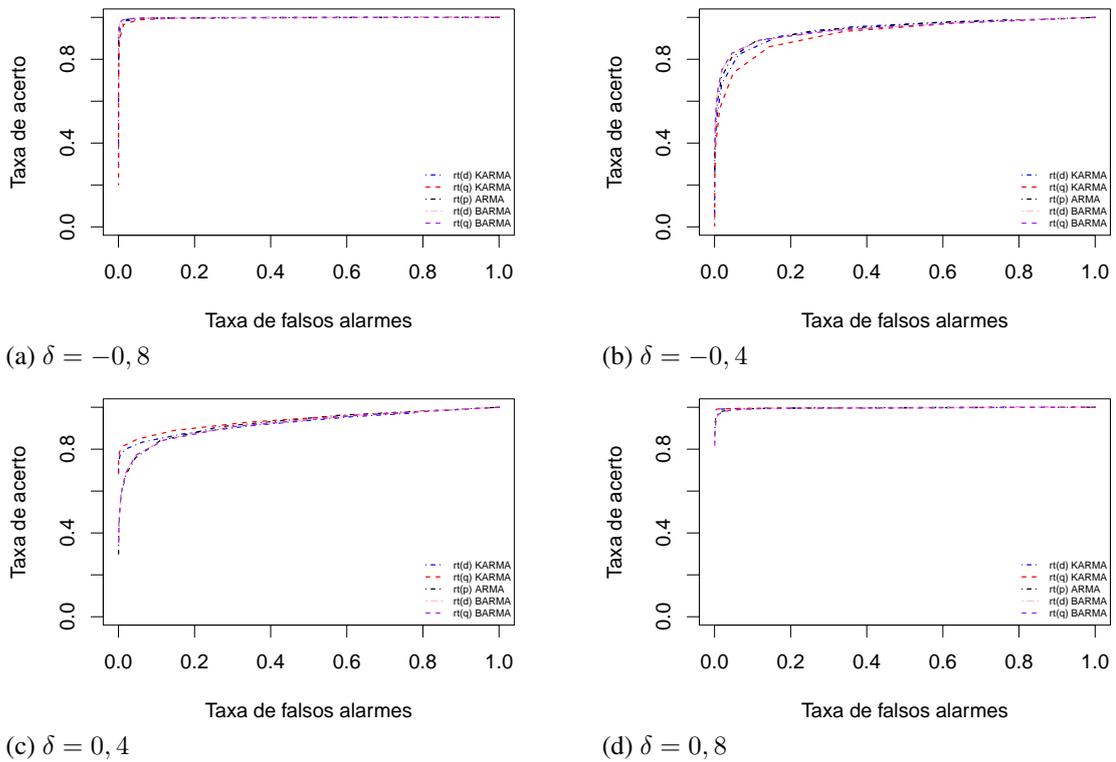
ZHANG, D.; LIN, J.; PENG, Q.; WANG, D.; YANG, T.; SOROOSHIAN, S.; LIU, X.; ZHU-ANG, J. Modeling and simulating of reservoir operation using the artificial neural network, support vector regression, deep learning algorithm. **Journal of Hydrology**, v.565, p.720 – 736, 2018.

APÊNDICES

APÊNDICE A – Resultados avaliação ARL_1 para $n_1 = 500$

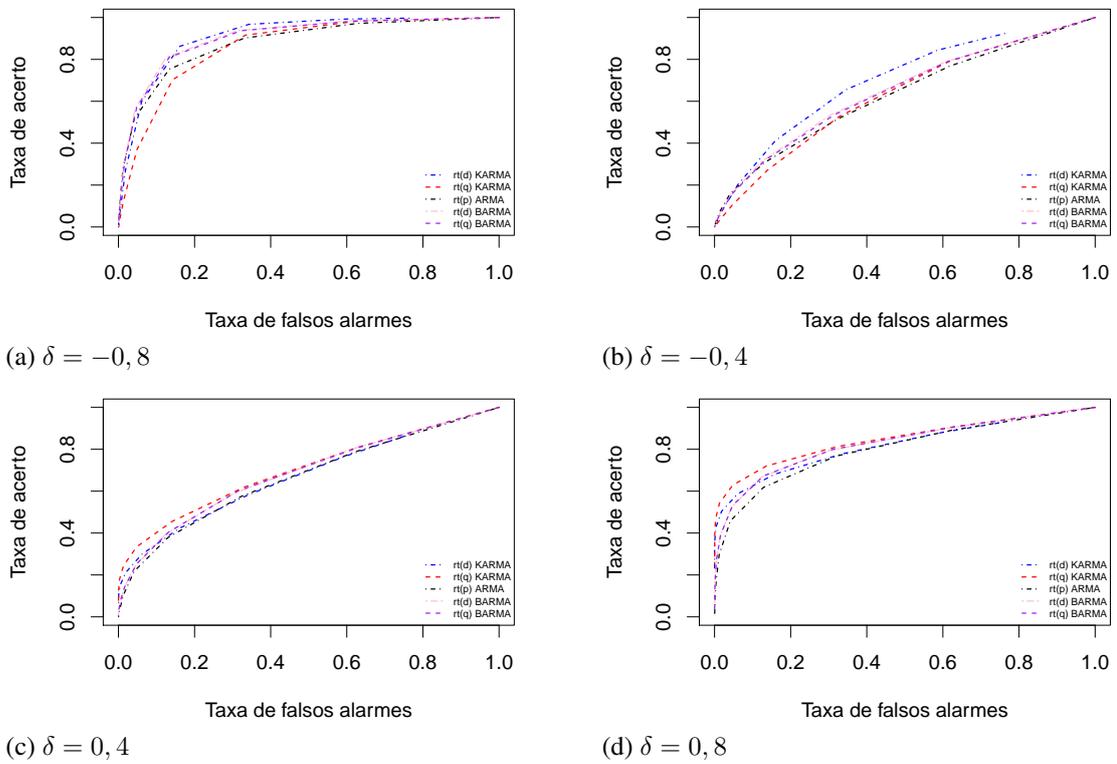
As Figuras 40, 41 e 42 apresentam o desempenho dos gráficos de controle de Shewhart e as Figuras 43, 44 e 45 o desempenho dos gráficos de controle CUSUM frente às diferentes magnitudes avaliadas para $n_1 = 500$ na Fase I.

Figura 40: Curva ROC gráfico de controle de Shewhart - Cenário 1.



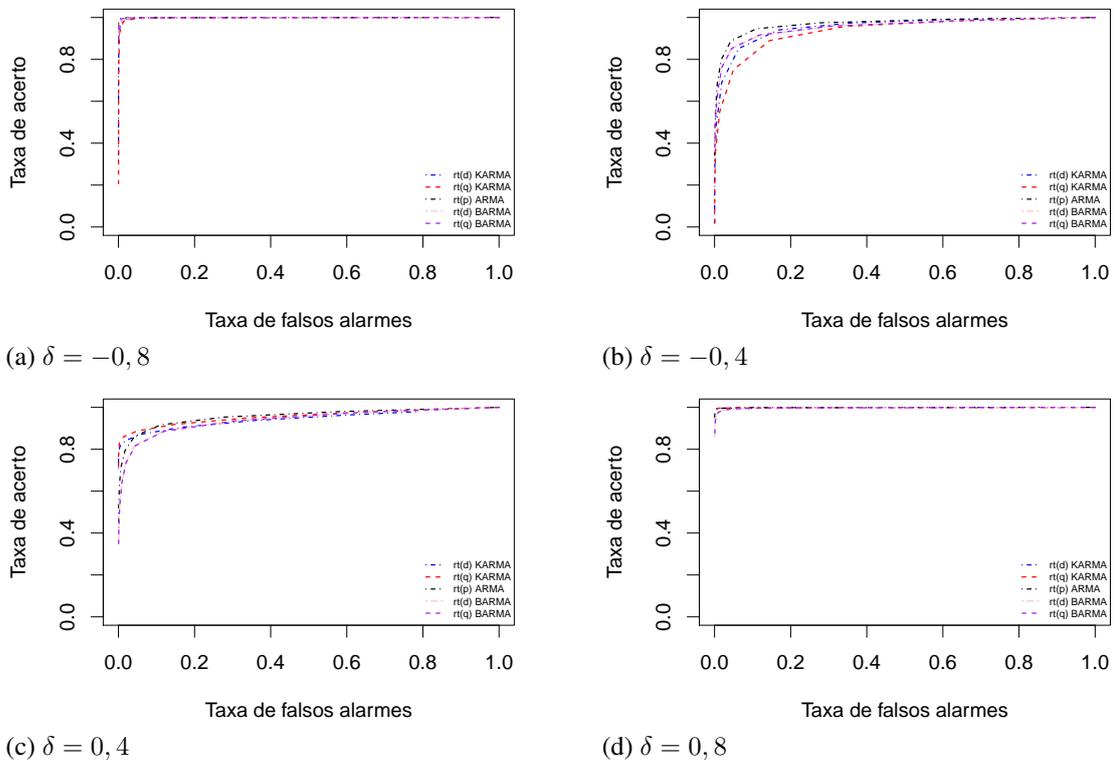
Fonte: Elaborado pela autora.

Figura 41: Curva ROC gráfico de controle de Shewhart - Cenário 2.



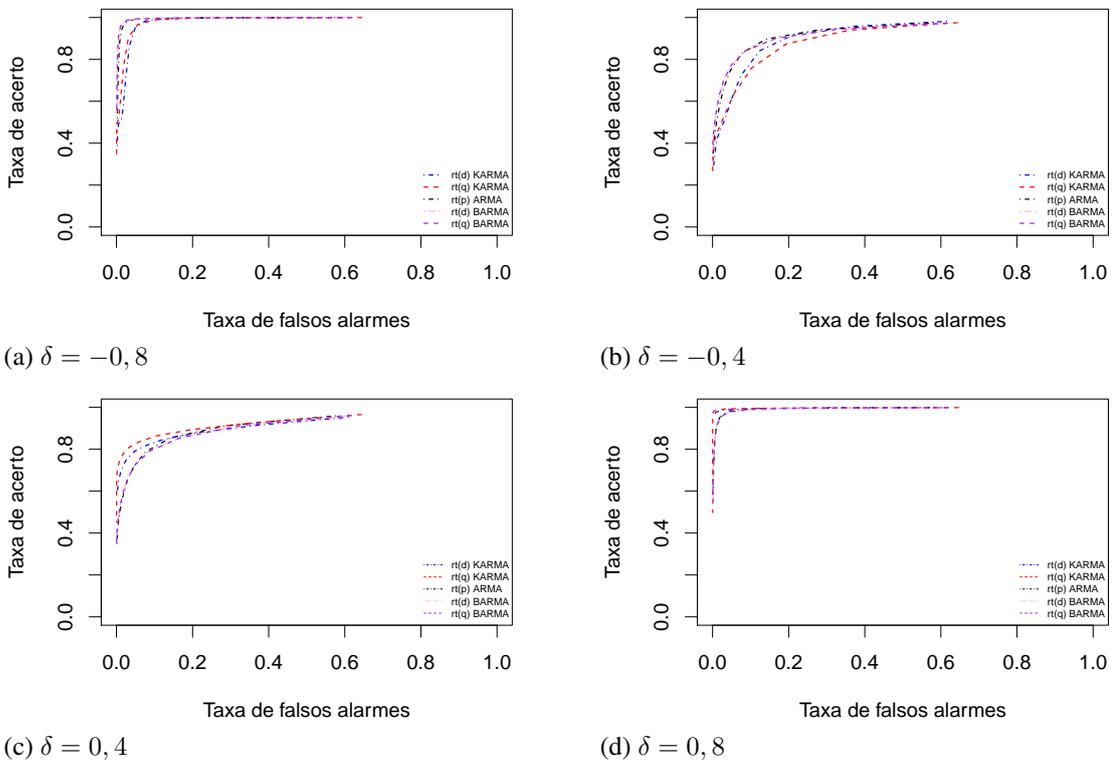
Fonte: Elaborado pela autora.

Figura 42: Curva ROC gráfico de controle de Shewhart - Cenário 3.



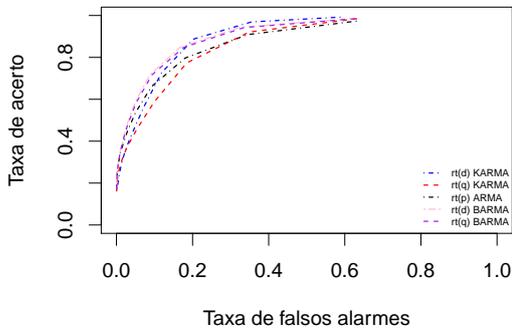
Fonte: Elaborado pela autora.

Figura 43: Curva ROC gráfico de controle CUSUM - Cenário 1.

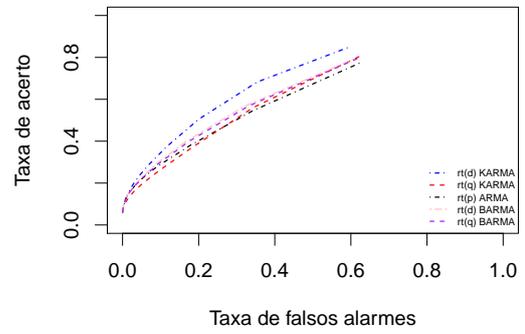


Fonte: Elaborado pela autora.

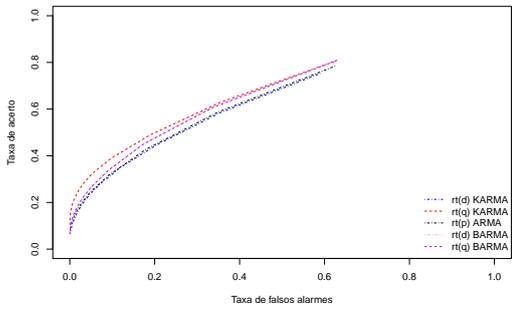
Figura 44: Curva ROC gráfico de controle CUSUM - Cenário 2.



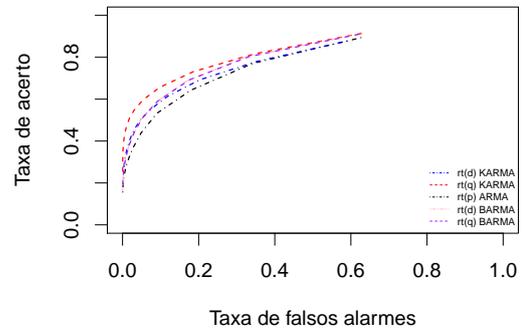
(a) $\delta = -0,8$



(b) $\delta = -0,4$



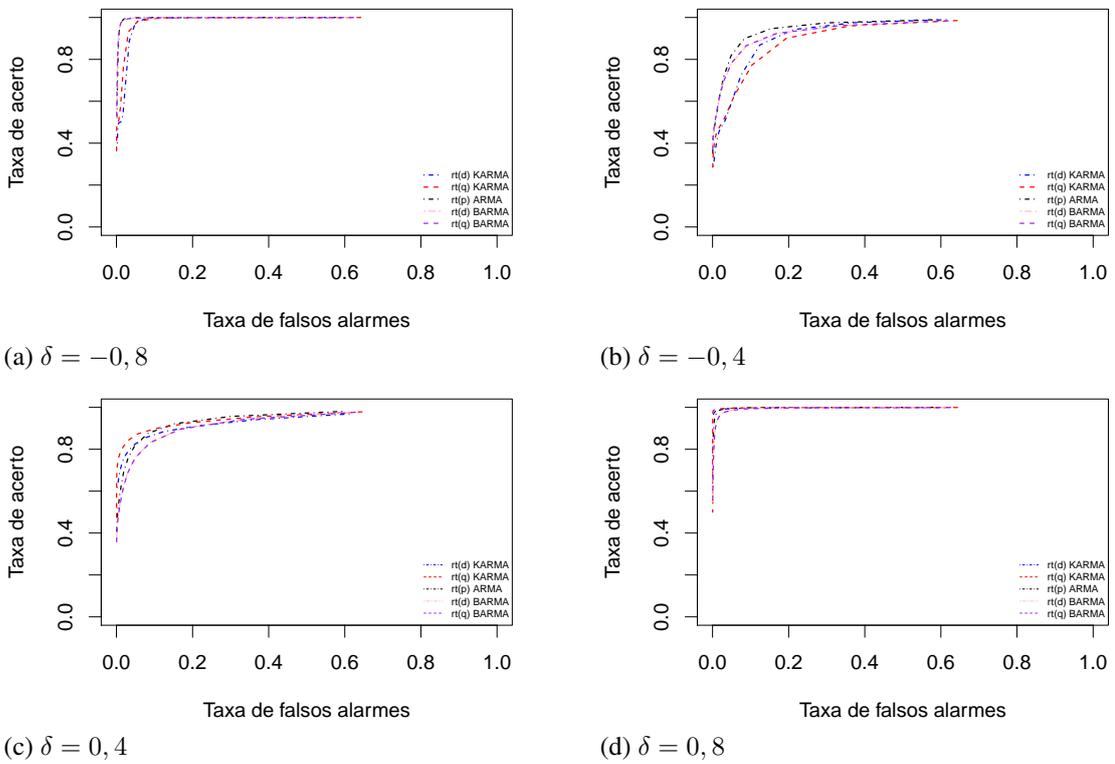
(c) $\delta = 0,4$



(d) $\delta = 0,8$

Fonte: Elaborado pela autora.

Figura 45: Curva ROC gráfico de controle CUSUM - Cenário 3.

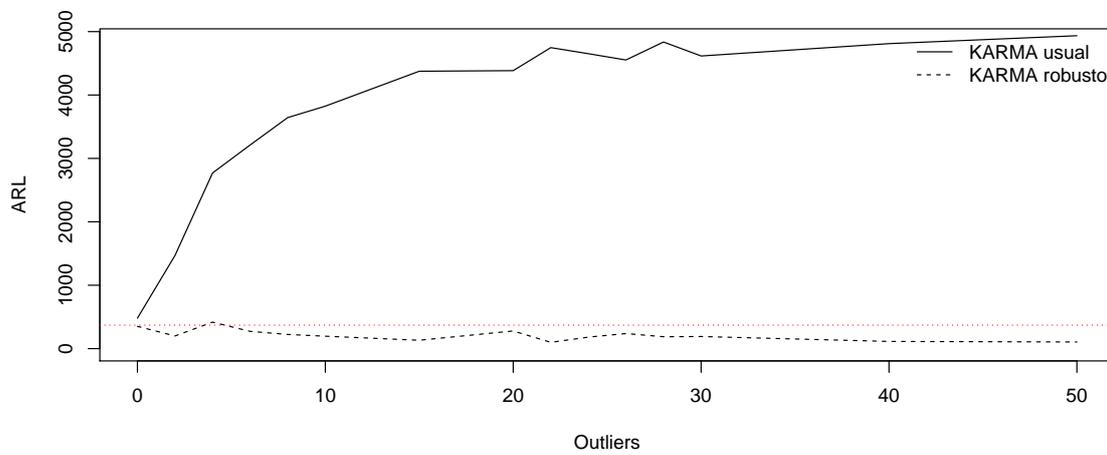


Fonte: Elaborado pela autora.

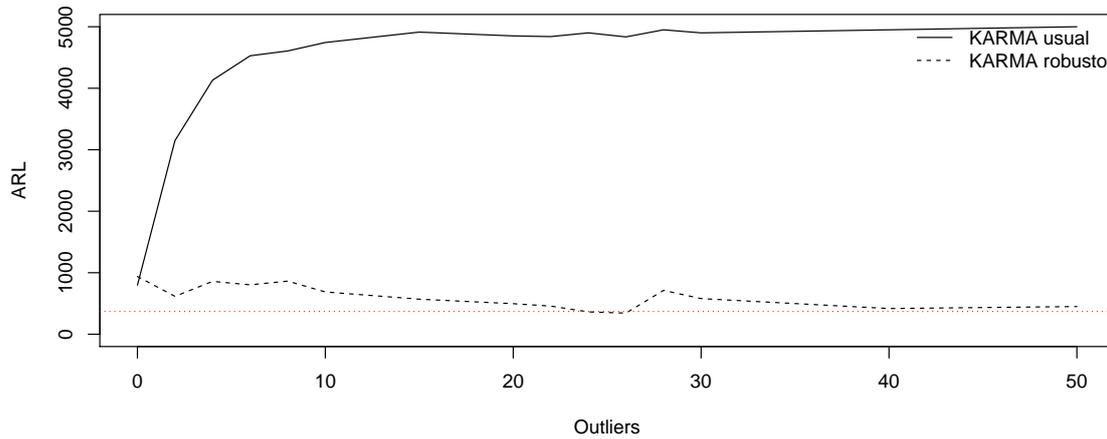
APÊNDICE B – Resultados avaliação de desempenho dos gráficos de controle KARMA robustos

Neste apêndice encontra-se as figuras que apresentam o desempenho dos gráficos de controle de Shewhart e CUSUM frente às diferentes intensidade e quantidades de *outliers* inseridos para $n_1 = 500$ na Fase I.

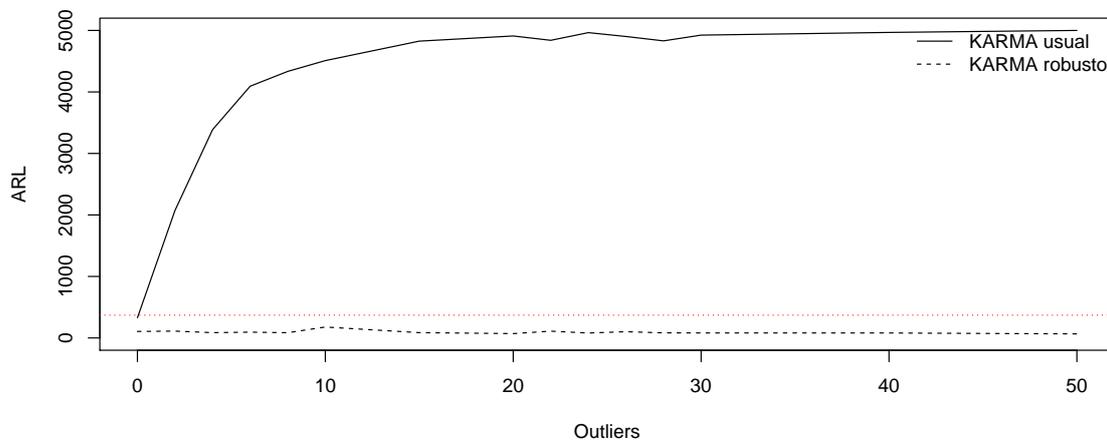
Figura 46: Ponto de ruptura do ARL_0 do gráfico de controle de Shewhart, Cenário 1.



Fonte: Elaborado pela autora.

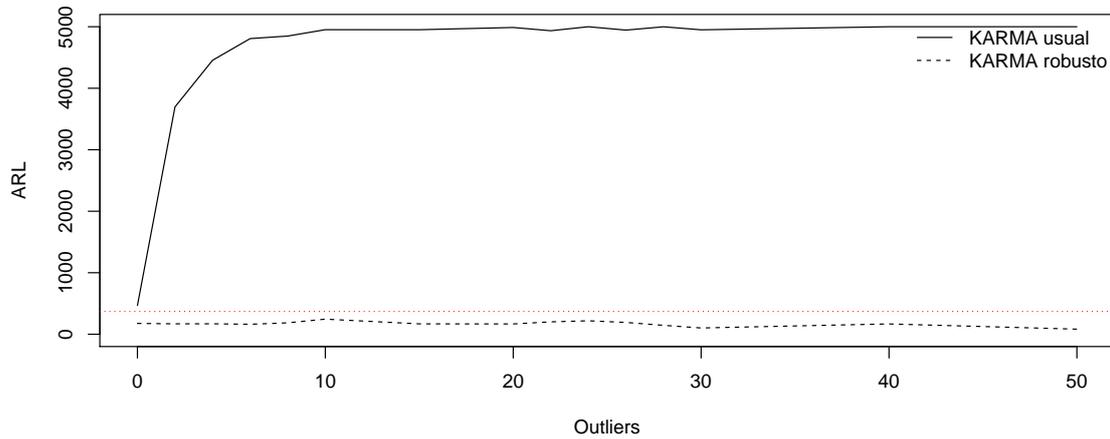
Figura 47: Ponto de ruptura do ARL_0 do gráfico de controle CUSUM, Cenário 1.

Fonte: Elaborado pela autora.

Figura 48: Ponto de ruptura do ARL_0 do gráfico de controle de Shewhart, Cenário 3.

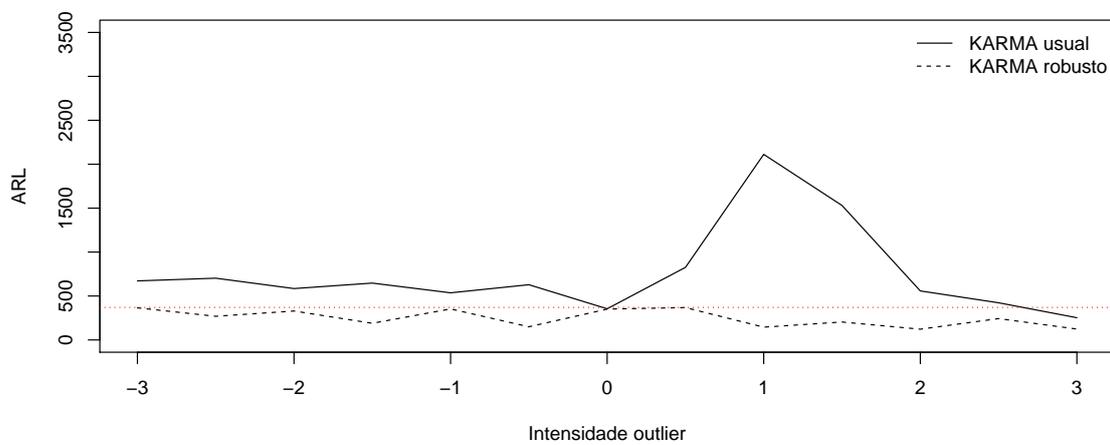
Fonte: Elaborado pela autora.

Figura 49: Ponto de ruptura do ARL_0 do gráfico de controle CUSUM, Cenário 3.



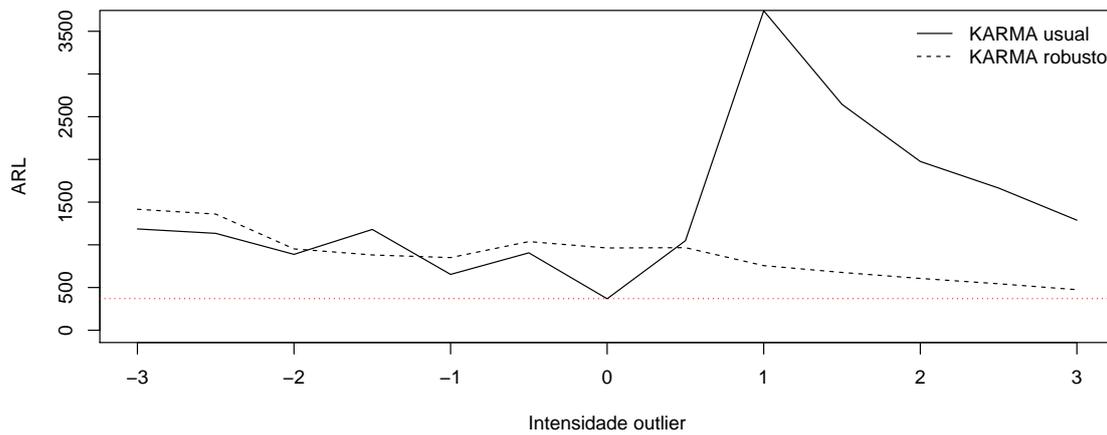
Fonte: Elaborado pela autora.

Figura 50: Sensibilidade do ARL_0 do gráfico de controle de Shewhart, Cenário 1.



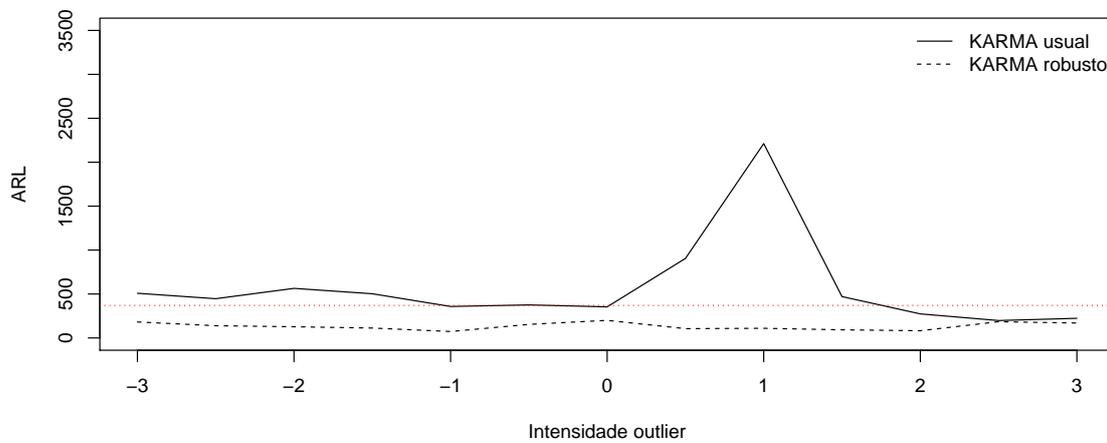
Fonte: Elaborado pela autora.

Figura 51: Sensibilidade do ARL_0 do gráfico de controle CUSUM, Cenário 1.



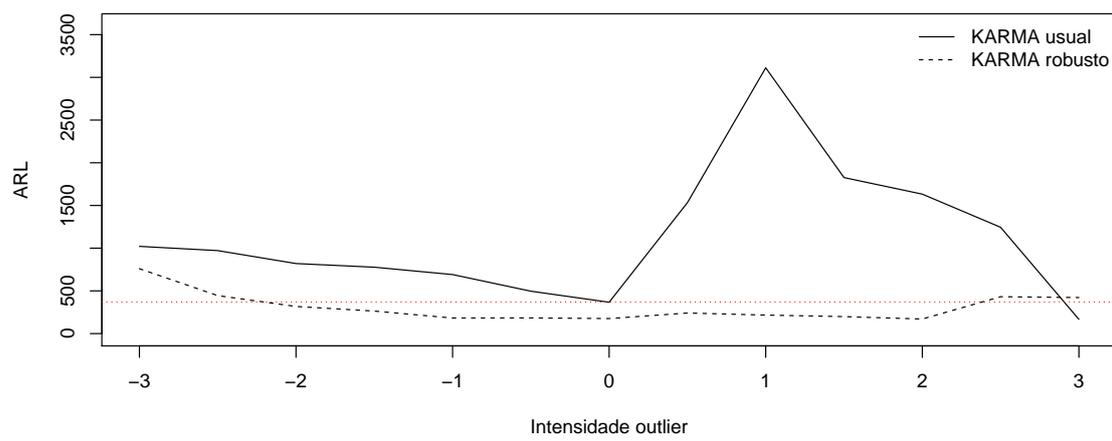
Fonte: Elaborado pela autora.

Figura 52: Sensibilidade do ARL_0 do gráfico de controle de Shewhart, Cenário 3.



Fonte: Elaborado pela autora.

Figura 53: Sensibilidade do ARL_0 do gráfico de controle CUSUM, Cenário 3.



Fonte: Elaborado pela autora.