

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

**GRÁFICOS DE CONTROLE PARA DADOS DO
TIPO TAXAS E PROPORÇÕES
AUTOCORRELACIONADOS**

DISSERTAÇÃO DE MESTRADO

Cátia Michele Tondolo

Santa Maria, RS, Brasil

2016

GRÁFICOS DE CONTROLE PARA DADOS DO TIPO TAXAS E PROPORÇÕES AUTOCORRELACIONADOS

Cátia Michele Tondolo

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

Orientador: Prof. Dr. Fábio M. Bayer

Santa Maria, RS, Brasil

2016

Ficha catalográfica elaborada através do Programa de Geração Automática da Biblioteca Central da UFSM, com os dados fornecidos pelo(a) autor(a).

Tondolo, Cátia Michele
Gráficos de controle para dados do tipo taxas e
proporções autocorrelacionados / Cátia Michele Tondolo.-
2016.
82 p. ; 30cm

Orientador: Fábio Mariano Bayer
Dissertação (mestrado) - Universidade Federal de Santa
Maria, Centro de Tecnologia, Programa de Pós-Graduação em
Engenharia de Produção, RS, 2016

1. Gráficos de controle 2. Autocorrelação 3. Beta ARMA
I. Bayer, Fábio Mariano II. Título.

**Universidade Federal de Santa Maria
Centro de Tecnologia
Programa de Pós-Graduação em Engenharia de Produção**

A Comissão Examinadora, abaixo assinada,
aprova a Dissertação de Mestrado

**GRÁFICOS DE CONTROLE PARA DADOS DO TIPO TAXAS E
PROPORÇÕES AUTOCORRELACIONADOS**

elaborada por
Cátia Michele Tondolo

como requisito parcial para obtenção do grau de
Mestre em Engenharia de Produção

COMISSÃO EXAMINADORA:

Fábio M. Bayer, Dr.
(Presidente/Orientador)

Denis A. de O. Moraes, Dr. (UFSM)

Helton S. B. dos Santos, Dr. (UFG)

Santa Maria, 28 de Março de 2016.

AGRADECIMENTOS

Ao meu orientador, Professor Fábio Mariano Bayer, pelo apoio e disponibilidade demonstrada em fazer parte deste trabalho.

Aos meus colegas e amigos que sempre me apoiaram e contribuíram para a elaboração deste trabalho, em especial à Fernanda Maria Muller e Bruna Gregory Palm pela preciosa ajuda na exposição e discussão de ideias e pelo companheirismo.

A toda a minha família, em especial à minha irmã e aos meus pais pela oportunidade e apoio incondicional durante a elaboração deste trabalho e um agradecimento especial a Carmem Rita S. Farias pela sala de estudos disponibilizada.

Aos membros da banca examinadora, pelas sugestões.

À CAPES, pelo apoio financeiro.

*“E ele se erguerá contra todo mal de Daburá,
Ide somente com Ele
E as Esferas do Dragão encontrarão.”*
— NAMEKUSEI 10:2

RESUMO

Dissertação de Mestrado
Programa de Pós-Graduação em Engenharia de Produção
Universidade Federal de Santa Maria

GRÁFICOS DE CONTROLE PARA DADOS DO TIPO TAXAS E PROPORÇÕES AUTOCORRELACIONADOS

AUTORA: CÁTIA MICHELE TONDOLO

ORIENTADOR: FÁBIO M. BAYER

Local da Defesa e Data: Santa Maria, 28 de Março de 2016.

Este trabalho aborda o controle estatístico de processos (CEP), que é uma das áreas para melhoria da qualidade. Uma das ferramentas mais empregadas do CEP é o gráfico de controle, sendo utilizado com o objetivo de monitorar parâmetros de um processo. Em geral, esses gráficos são construídos sob as suposições de normalidade e de independência das observações. No entanto, por vezes essas suposições não se verificam. Os gráficos de controle usuais funcionam razoavelmente bem se a hipótese de normalidade for violada com moderação, mas a violação da hipótese de independência das observações reduz o desempenho dos mesmos. Quando os dados possuem autocorrelação uma das abordagens é o emprego de gráficos de controle utilizando resíduos de modelos, usualmente da classe ARIMA. Os resíduos são utilizados para confeccionar gráficos de controle usuais como de Shewhart, CUSUM e EWMA. Ainda, quando as variáveis são do tipo taxas e proporções, restritas ao intervalo $(0,1)$, é natural assumir distribuição beta. Com isso, propomos para a modelagem e monitoramento de processos beta distribuídos autocorrelacionados a confecção de gráficos de controle utilizando diferentes resíduos do modelo β ARMA. Os desempenhos dos gráficos propostos foram avaliados por meio de simulações de Monte Carlo. Foi analisada a medida de desempenho de gráficos de controle, ARL (average run length), ou comprimento médio de corrida, sob controle e fora de controle. Comparou-se os desempenhos destes com as abordagens tradicionais para dados que possuem autocorrelação. Para evidenciar a importância prática dos gráficos de controle propostos, foram feitas duas aplicações em dados reais, em que foram monitorados o volume de energia armazenada na Região Sul do Brasil e os níveis dos mananciais do Sistema Cantareira (São Paulo). Os gráficos propostos obtiveram bom desempenho para dados do tipo taxas e proporções, apresentando melhor poder de detecção de causas especiais do que as alternativas usuais.

Palavras-chave: Gráficos de controle. Autocorrelação. β ARMA.

ABSTRACT

Master's Dissertation
Post-Graduate Program in Production Engineering
Federal University of Santa Maria

CONTROL CHARTS FOR RATES AND PROPORTIONS DATA AUTOCORRELATED.

AUTHOR: CÁTIA MICHELE TONDOLO

ADVISOR: FÁBIO M. BAYER

Defense Place and Date: Santa Maria, March 28st, 2016.

This work discusses one of the areas for quality improvement, defined as statistical process control (SPC). One of the most used tools in SPC is the control chart, which is used to monitor parameters of a process. In general, these charts are built under normality and independence assumptions of observations. However, sometimes these suppositions do not occur. The usual control charts work reasonably well if normal distribution assumption is moderately violated, but the violation of independence assumption reduces the applicability of them. When the data are autocorrelated, it is adequate to use residuals control charts, usually from ARIMA class. The residuals are used to produce the usual control charts, like the Shewhart, CUSUM and EWMA. In addition, variables restricted to the interval (0,1), such as rates and proportions, are naturally assumed to follow beta distribution. Thus, we propose the use of control chart with different residuals of the model β ARMA to model and monitor autocorrelated beta distributed processes. The performance measures of the proposed control charts were evaluated by Monte Carlo simulations and the ARL (average run length) was analyzed under control and out of control. Proposed and traditional models were compared for autocorrelated data adjustment. Two applications were performed using real data associated to the volume of energy stored in southern Brazil and the levels of the sources of the Cantareira System (São Paulo, Brazil). The proposed control charts showed good performance for rates and proportions data, getting a better detection of special causes than usual modeling.

Keywords: Control charts. Autocorrelation. β ARMA.

LISTA DE TABELAS

Tabela 4.1 – Cenários simulados.	37
Tabela 5.1 – Modelo ARMA(1,1) ajustado para os dados da EA.	48
Tabela 5.2 – Modelo β ARMA(1,1) ajustado para os dados da EA.	48
Tabela 5.3 – Modelo ARIMA para o volume dos mananciais do Sistema Cantareira.	53
Tabela 5.4 – Modelo β ARMA 5.2 para o volume dos mananciais do Sistema Cantareira. .	54
Tabela A.1 – Apresentação dos modelos de regressão lineares aos ARL's do cenário com $\phi = 40$ e $\mu = 0,2$	67
Tabela A.2 – Apresentação dos modelos de regressão lineares aos ARL's do cenário com $\phi = 100$ e $\mu = 0,2$	68
Tabela A.3 – Apresentação dos modelos de regressão lineares aos ARL's do cenário com $\phi = 200$ e $\mu = 0,2$	69
Tabela A.4 – Apresentação dos modelos de regressão lineares aos ARL's do cenário com $\phi = 40$ e $\mu = 0,5$	70
Tabela A.5 – Apresentação dos modelos de regressão lineares aos ARL's do cenário com $\phi = 100$ e $\mu = 0,5$	71
Tabela A.6 – Apresentação dos modelos de regressão lineares aos ARL's do cenário com $\phi = 200$ e $\mu = 0,5$	72
Tabela A.7 – Apresentação dos modelos de regressão lineares aos ARL's do cenário com $\phi = 40$ e $\mu = 0,8$	73
Tabela A.8 – Apresentação dos modelos de regressão lineares aos ARL's do cenário com $\phi = 100$ e $\mu = 0,8$	74
Tabela A.9 – Apresentação dos modelos de regressão lineares aos ARL's do cenário com $\phi = 200$ e $\mu = 0,8$	75
Tabela B.1 – Apresentação dos resultados para o ARL_1 dos GC baseados nos resíduos do modelo de suavização exponencial simples do cenário para $\mu = 0,2$	77
Tabela B.2 – Apresentação dos resultados para o ARL_1 dos GC baseados nos resíduos do modelo de suavização exponencial simples do cenário para $\mu = 0,5$	78
Tabela B.3 – Apresentação dos resultados para o ARL_1 dos GC baseados nos resíduos do modelo de suavização exponencial simples do cenário para $\mu = 0,8$	79

LISTA DE APÊNDICES

APÊNDICE A – Calibração	67
APÊNDICE B – Resultados para suavização exponencial simples	76
APÊNDICE C – Resultados do ARL_1 para os cenários 7, 8 e 9	80

LISTA DE ABREVIATURAS E SIGLAS

ARIMA	<i>Autoregressive integrated moving averages</i> , ou autorregressivo integrado de médias móveis
ARL	<i>Average run length</i> , ou comprimento médio de corrida
β ARMA	<i>Beta autoregressive moving average</i> , ou beta autorregressivo de médias móveis
CCC	<i>Common control chart</i> , ou gráfico de causas comuns
CEP	Controle estatístico de processo
CQ	Característica da qualidade
CUSUM	<i>Cumulative sum</i> ou soma acumulada
EMV	Estimador de máxima verossimilhança
EWMA	<i>Exponentially weighted moving average</i> , ou média móvel ponderada exponencialmente
FAC	Função de autocorrelação
FACP	Função de autocorrelação parcial
GC	Gráfico de controle
<i>iid</i>	Independente e identicamente distribuído
LC	Linha central
LIC	Limite inferior de controle
LSC	Limite superior de controle
RL	<i>Run length</i> , ou comprimento de corrida
SCC	<i>Special control chart</i> , ou gráfico de causas especiais
SES	Suavização exponencial simples

SUMÁRIO

1 INTRODUÇÃO	13
1.1 Objetivos e justificativa	15
1.2 Estrutura do trabalho	16
2 GRÁFICOS DE CONTROLE	17
2.1 Controle estatístico de processos	17
2.2 Gráficos de controle usuais	18
2.2.1 Gráficos de controle de Shewhart.....	20
2.2.2 Gráficos de controle CUSUM	21
2.2.3 Gráficos de controle EWMA	22
2.3 Gráficos de controle para dados autocorrelacionados	22
2.3.1 Modelos ARIMA	24
2.3.2 Modelos EWMA	26
3 GRÁFICOS DE CONTROLE PARA DADOS DO TIPO TAXAS E PROPORÇÕES 28	
3.1 Gráfico de controle beta	28
3.2 Gráficos de controle propostos	29
3.2.1 Modelo β ARMA	30
3.2.2 Gráfico β ARMA Shewhart	34
3.2.3 Gráfico β ARMA CUSUM	34
3.2.4 Gráfico β ARMA EWMA	35
4 AVALIAÇÃO NUMÉRICA	36
4.1 Simulação de Monte Carlo e análise de desempenho	36
4.2 Resultados	39
5 APLICAÇÕES	46
5.1 Percentual de energia armazenada	46
5.2 Sistema Cantareira	51
6 CONCLUSÕES	58
REFERÊNCIAS	60
APÊNDICES	66

1 INTRODUÇÃO

As técnicas de controle estatístico de processos (CEP) são amplamente utilizadas na indústria e nos mais diversos processos com o objetivo de monitorar e obter melhorias da qualidade (ZHANG, 1998; LIU et al., 2015). Cabe aos gestores uma busca continuada para o controle e o aperfeiçoamento dos seus serviços e/ou produtos, seja pelo monitoramento ou pela adequação dos seus processos (JURAN, 1951).

Cada produto ou processo possui elementos que os descrevem que são chamados de características da qualidade (CQ). Produzir produtos ou processos com características idênticas de uma unidade a outra é difícil, dispendioso, ou mesmo impossível, devido à variabilidade inerente a qualquer processo. Contudo, esta variabilidade pode ser descrita e controlada através de métodos estatísticos. Estes métodos podem ser aplicados em diferentes áreas como manufatura, desenvolvimento de processos, finanças, epidemiologia, entre outras (MONTGOMERY, 2009).

O CEP possui ferramentas importantes na resolução de problemas para determinar a estabilidade de processos e obter melhorias através da redução da variabilidade (TEOH et al., 2014). Neste trabalho o foco será nos gráficos de controle (GC), que são umas das ferramentas mais importantes e aplicadas do CEP, sendo utilizados para o monitoramento de processos (KRAMER; SCHMID, 1997). Os GC acompanham uma característica do processo e são compostos por um limite superior de controle (LSC), uma linha central (LC) e um limite inferior de controle (LIC) (MONTGOMERY, 2009).

Os modelos tradicionais do CEP assumem que existem duas possíveis causas de variabilidade: as causas comuns que são as causas aleatórias; e as causas especiais que provém de falha humana, de máquinas com má calibração, ou seja, causas detectáveis que podem ser removidas do processo (MUKHERJEE, 2015). Diversos estudos comparam a capacidade de detecção de causas especiais destas ferramentas com relação aos métodos usuais. Alguns trabalhos nessa linha são Acosta-Mejia, Pignatiello e Venkateshwara Rao (1999), Amirzadeh, Mashinchi e Parchami (2009), Colosimo e Pacella (2010) e Sant'Anna e Ten Caten (2012).

No CEP, um processo sob controle é identificado quando gera variáveis aleatórias independentes e identicamente distribuídas (*i.i.d.*) (ALWAN; ROBERTS, 1988). Ou seja, tradicionalmente assume-se que observações consecutivas de um processo são independentes e com mesma distribuição de probabilidade (ZHANG, 1998). No entanto, na prática, em processos contínuos, as observações podem ser correlacionadas ao longo do tempo (MONTGOMERY,

2009).

Gráficos de controle, incluindo os de Shewhart (1931), soma cumulativa (CUSUM) (PAGE, 1954) e gráfico de média móvel exponencialmente ponderada (EWMA) (ROBERTS, 1959) são amplamente utilizados para o monitoramento de processos. Esses gráficos são construídos sob a suposição de normalidade da variável associada à característica que está sendo avaliada e a independência entre as unidades amostrais (WARDELL; MOSKOWITZ; PLANTE, 1992). Contudo, muitas variáveis de interesse em diversos processos não satisfazem essas suposições restritivas.

Ao desconsiderar a autocorrelação, por exemplo, pode-se deixar de perceber fatores que contribuem para efeitos defasados (ALWAN; ROBERTS, 1988). Ao não considerar essa correlação serial entre as unidades amostrais, quando ela existe, o desempenho dos gráficos de controle é diretamente afetado, podendo ocorrer conclusões inapropriadas, uma vez que a taxa de alarmes falsos e o poder dos gráficos podem ser afetados (MINGOTI; FIDELIS, 2001; CLARO; COSTA; MACHADO, 2007).

Uma ação simples para atenuar a questão da autocorrelação é diminuir a frequência de amostragem dos dados. Entretanto, esta técnica pode levar ao não conhecimento do processo, de modo que a detecção de mudanças no processo pode ser atrasada, não sendo aconselhável utilizar esta alternativa (MONTGOMERY, 2009). Nesses casos, necessita-se utilizar diferentes técnicas para modificar os gráficos de controle usuais para que a situação de autocorrelação seja considerada adequadamente. Primeiramente, é necessário modelar os dados para eliminar o efeito da autocorrelação e então controlar estatisticamente a parte não sistemática do processo (ALWAN; ROBERTS, 1988). Desta forma, algumas técnicas estatísticas são necessárias para abordar os dados, dentre as quais se destaca, segundo Montgomery (2009), o ajuste de modelos autorregressivos integrados de médias móveis (ARIMA) (MORETTIN; TOLOI, 2004; BOX; JENKINS; REINSEL, 2008). Nesses casos, o monitoramento é feito através do gráfico de controle de resíduos, assumindo que a série de observações do processo é uma série temporal.

A justificativa do uso de gráficos de controle residuais é que, assumindo que o modelo de séries temporais está corretamente ajustado, os resíduos deste modelo não são autocorrelacionados. Portanto, gráficos de controle usuais, como Shewhart, CUSUM e o EWMA, podem então ser aplicados ao processo residual (ZHANG, 1997).

Ainda, quando a distribuição dos dados no processo industrial é assimétrica ou é do tipo fração, taxa ou proporção, a frequência de falsos alarmes dos gráficos de controle usuais

aumenta devido a discrepância entre a forma da distribuição dos dados e da distribuição normal (SANT'ANNA; TEN CATEN, 2012). Para os casos em que os dados estão restritos ao intervalo contínuo $(0,1)$, Sant'Anna e Ten Caten (2012) propõem um gráfico de controle beta destinado ao monitoramento de frações e proporções. Este gráfico de controle assume distribuição beta para os dados e define novos limites de controle baseados nos quantis dessa distribuição. Contudo, eles ainda consideram dados não autocorrelacionados.

Com isso, o objetivo principal deste trabalho é estender o gráfico de controle beta de Sant'Anna e Ten Caten (2012) para o caso não *i.i.d.*, assumindo que os dados possuem autocorrelação. Para isso, modela-se o processo por meio de um modelo beta autorregressivo de médias móveis (β ARMA) (ROCHA; CRIBARI-NETO, 2009) e então aplica-se gráficos de controles usuais nos resíduos desse modelo ajustado. O modelo β ARMA é um modelo de séries temporais recentemente proposto para modelar dados do tipo taxas e proporções ao longo do tempo.

1.1 Objetivos e justificativa

Usualmente, no controle estatístico de processo, um estado sob controle é identificado quando um processo de geração é *i.i.d.*. Entretanto, algumas vezes, a suposição de independência é violada. Para os casos em que a autocorrelação está presente é necessário o tratamento dos dados antes do uso dos gráficos de controle. Uma abordagem utilizada nestes casos é o uso de resíduos de modelos de séries temporais, usualmente os modelos ARIMA. Com isso a relevância deste trabalho é utilizar e analisar o desempenho dos gráficos de controle utilizando os resíduos do modelo β ARMA, mais indicado para dados do tipo taxas e proporções (ROCHA; CRIBARI-NETO, 2009).

Este trabalho contempla os gráficos de controle para o monitoramento de variáveis, cujas características de qualidade são do tipo fração ou proporção, que apresentam valores no intervalo $(0,1)$ e são autocorrelacionadas. Será assumido que os dados possuem distribuição de probabilidade beta e o processo será modelado utilizando os modelos β ARMA.

Pontualmente, vislumbram-se os seguintes objetivos específicos:

- Revisar os conceitos relativos ao controle estatístico de processos, gráficos de controle, séries temporais, modelos para frações e proporções e simulação de Monte Carlo;

- Propor a utilização de resíduos do modelo β ARMA para a construção dos gráficos de controle para dados do tipo frações e proporções autocorrelacionados;
- Utilizar simulações de Monte Carlo para calibrar os gráficos de controle e avaliar o desempenho dos gráficos de controle propostos em comparação à modelagem usual;
- Considerar aplicações dos gráficos de controle propostos em dados reais.

1.2 Estrutura do trabalho

Para atingir aos objetivos propostos, esta dissertação foi estruturada em seis capítulos que são descritos a seguir. O Capítulo 1 aborda os aspectos gerais do estudo, o tema, os objetivos gerais e específicos, e a estrutura da pesquisa. No Capítulo 2 é apresentada a fundamentação teórica do controle estatístico de processos (CEP), os principais gráficos de controle utilizados, o método de utilização de gráficos de controle para dados autocorrelacionados e os modelos usuais de ajuste utilizados na literatura. O Capítulo 3 apresenta o modelo β ARMA e introduz os gráficos de controle residuais propostos. No Capítulo 4 encontram-se os resultados numéricos obtidos através das simulações de Monte Carlo e a comparação entre os gráficos de controle considerados. O Capítulo 5 apresenta duas aplicações para o monitoramento de dados do volume de energia armazenada na Região Sul do Brasil e dos níveis dos mananciais do Sistema Cantareira. O Capítulo 6 traz as conclusões do trabalho.

2 GRÁFICOS DE CONTROLE

Neste capítulo são apresentados alguns conceitos relacionados ao controle estatístico de processos e ao uso de uma de suas principais ferramentas, os gráficos de controle. Serão abordados os tópicos de gráficos de controle usuais e gráficos de controle para séries temporais.

2.1 Controle estatístico de processos

O CEP é a utilização de métodos estatísticos com o objetivo de melhorar a qualidade de produtos e/ou serviços. Um objetivo importante do monitoramento estatístico de processo é a detecção rápida da variação no sistema de produção para permitir que ações corretivas necessárias sejam tomadas de forma rápida e satisfatória (SHEU et al., 2009).

A qualidade de um produto ou processo possui conceitos amplos, onde o mais usual está relacionado às características desejáveis que este deve conter. A tradicional definição seria que os produtos e processos devem apresentar as especificações exigidas por aqueles que o usam, possuindo adequação ao uso; sendo a qualidade inversamente proporcional à variabilidade. Se a variabilidade das características que compõe um produto ou processo decresce, então a qualidade tende a aumentar. Dessa forma, a melhoria da qualidade é a redução da variabilidade, pois significa menos retrabalho e redução de gastos em tempo, esforço e dinheiro (MONTGOMERY, 2009).

Para controlar estatisticamente um processo, garantir qualidade e satisfazer as especificações, uma das mais importantes ferramentas do CEP são os gráficos de controle. Estes visam monitorar um processo para controlar ou reduzir sua variabilidade e também estimam os parâmetros de qualidade do produto e do processo. A finalidade básica de implementação de gráficos de controle é detectar alterações anormais no processo, tanto nos parâmetros de escala como nos de localização, buscando eliminar a variabilidade do mesmo (ABBASI, 2010).

Em processos produtivos existem basicamente dois tipos de causas de variabilidade. O primeiro tipo, a variabilidade natural, é o efeito cumulativo de causas pequenas, essencialmente inevitáveis, em geral, chamada sistema estável de causas aleatórias. Um processo que opera apenas com as causas aleatórias da variação está sob controle estatístico. Essas causas aleatórias são parte inerente ao processo, de forma que, independentemente do planejamento, uma certa

quantidade de variabilidade natural sempre existirá (REYNOLDS JR; LU, 1997; MUKHERJEE, 2015). Outro tipo são as fontes de variabilidade, que não fazem parte do padrão de causas aleatórias, como causas atribuíveis. Um processo que opera na presença de causas atribuíveis está fora de controle. Para detecção dessa segunda causa de variabilidade é que os gráficos de controle são amplamente utilizados (MONTGOMERY, 2009).

Para construção dos gráficos de controle são coletados dados a partir de um processo, sendo que estes dados devem representar níveis das características de qualidade associadas. Se essas características podem ser mensuráveis em escalas numéricas, então são usados os gráficos de controle para variáveis. Se as características não podem ser representadas de forma numérica, julga-se cada unidade como conforme ou não-conforme, ou conta-se o número de não-conformidades (defeitos), e então são utilizados os gráficos de controle para atributos (AMIRZADEH; MASHINCHI; PARCHAMI, 2009). As amostras podem ser divididas em subgrupos racionais ou em observações individuais. Neste trabalho as amostras serão dadas como no último caso, medidas individuais em uma sequência temporal.

Quando os efeitos de séries temporais sistemáticos estão presentes, há graves problemas de não-detecção das causas especiais que realmente existem, assim como de perceber causas especiais aparentes que não existem. Nestes casos, a detecção de causas especiais por métodos convencionais pode não ser confiável (ALWAN, 1992).

Alwan e Roberts (1988) sugerem o uso de modelagem de séries temporais para ajudar a detectar a existência de variação sistemática e para obter uma detecção precisa de causas especiais. A ideia básica é a de associar as causas comuns com os valores ajustados por um modelo de séries temporais e as causas especiais devem ser analisadas a partir dos resíduos, em vez de uma aplicação ineficaz dos métodos usuais diretamente aos dados originais.

2.2 Gráficos de controle usuais

Shewhart (1931) desenvolveu o gráfico de controle com o objetivo de detectar a presença de causas especiais de variação em um dado processo. Utilizando a ideia de que se o processo está sob controle estatístico, então pode-se controlar ou identificar observações atípicas, baseando-se nas observações passadas. Desta forma, um gráfico de controle usual possui uma linha central, que representa o valor médio da característica do processo, e limites superior de controle e inferior de controle, que são determinados utilizando a distribuição de probabili-

dade da característica da qualidade. Na prática, os parâmetros de média e desvio-padrão, úteis para a determinação desses limites, são estimados baseados na amostra observada.

Montgomery (2009) afirma que os gráficos de controle equivalem a um teste de hipóteses:

$$\begin{cases} H_0 & : \text{processo sob controle} \\ H_1 & : \text{processo fora de controle.} \end{cases}$$

Observações da amostra entre os limites de controle equivalem a não rejeitar a hipótese nula de que o processo está sob controle estatístico e os pontos fora equivalem a rejeitar esta hipótese, ou seja, indica que o processo está fora de controle estatístico.

O objetivo primário do gráfico de controle é criar limiares que facilitem o monitoramento da qualidade de um processo, verificando se este é estável em relação à parâmetros pré-estabelecidos, e se existem causas especiais que atuam sobre ele. Quando causas especiais são detectadas nos gráficos de controle (pontos fora dos limiares) o processo deve sofrer modificações ou ações corretivas para que volte à condição de controle estatístico (MONTGOMERY, 2009).

Os gráficos de controle são principalmente classificados em duas categorias no que diz respeito à sua estrutura, nomeado como sem memória (tipo Shewhart) e gráficos de controle de memória. Gráficos de controle CUSUM e EWMA são os dois gráficos de controle com memória mais utilizados (ABBAS; RIAZ; DOES, 2014).

A construção de gráficos de controle pode utilizar duas fases, nomeadamente: Fase I e Fase II. Os principais interesses na Fase I são analisar o conjunto de dados do processo para entender a variação, estimar os parâmetros e determinar a estabilidade do processo e os limites de controle. Uma vez que as amostras associadas com causas atribuíveis são removidas, estima-se os valores em controle dos parâmetros do processo. Em contrapartida, a preocupação na análise da Fase II é detectar rapidamente mudanças no processo a partir das estimativas dos parâmetros obtidas na Fase I (WOODALL, 2000).

Vining (2009) mostra a importância da Fase I, pois esta afeta o desempenho do gráfico de controle e depende de quão bem os parâmetros do gráfico são estimados, sendo que quanto maior o número de observações melhor a estimação. Nesta etapa o processo é estável, previsível e repetitivo e os limites devem ser reavaliados periodicamente. Na Fase II, segundo Mahmoud e Woodal (2004), se compara os métodos concorrentes em termos da distribuição do comprimento médio de corrida, ou ARL (*average run length*). Este comprimento de execução é definido como o número de observações amostradas até que uma observação fora de controle no processo seja

detectada.

Para avaliar os gráficos de controle após estimados os limiares na Fase I (sob controle) usualmente são criados cenários fora de controle para representar diferentes situações (WODALL, 2000). O desempenho do gráfico é medido pela capacidade para detectar padrões anormais nos dados durante a fase operacional do gráfico de controle (Fase II) (COLOSIMO; PACELLA, 2010).

2.2.1 Gráficos de controle de Shewhart

O conceito fundamental de gráficos de controle introduzidos por Shewhart (1931) assume que uma observação obtida de um processo sob controle (Y_t) pode ser representada por:

$$Y_t = \mu + \epsilon_t, \quad t = 1, 2, \dots, n, \quad (2.1)$$

em que μ é média do processo, supondo que seja fixa e constante, ϵ_t são os termos de erro, supostos independentes, com distribuição normal de média zero e desvio padrão constante σ , i.e., $\epsilon_t \sim \mathcal{N}(0, \sigma^2)$ e n o tamanho amostral. Desde que seja assumido que apenas uma observação está disponível em cada período de tempo, o gráfico de Shewhart é na verdade um gráfico de controle para observações individuais (WARDELL; MOSKOWITZ; PLANTE, 1992).

Seja w uma estatística amostral que mede alguma característica da qualidade de interesse. Suponha que a média de w seja μ_w e o desvio padrão σ_w . Com isso, os limites superior (LSC), central (LC) e inferior de controle (LIC) do gráfico de controle são dados por:

$$\text{LSC} = \mu_w + k\sigma_w, \quad (2.2)$$

$$\text{LC} = \mu_w, \quad (2.3)$$

$$\text{LIC} = \mu_w - k\sigma_w, \quad (2.4)$$

em que k é a distância dos limites de controle superior e inferior do limite central, expressa em unidades de desvio padrão. Geralmente utiliza-se $k = 3$, para uma porcentagem de ocorrência de pontos amostrais fora dos limites de controle, quando o processo está sob controle estatístico, delimitada em 0,27%, obtendo um $\text{ARL}_0 \approx 370$ (MINGOTI; YASSUKAWA, 2008; BERGQUIST; SÖDERHOLM, 2015). Esta construção está diretamente relacionada com a suposição de normalidade da estatística w .

2.2.2 Gráficos de controle CUSUM

Outro importante gráfico que pode ser utilizado é o gráfico de controle CUSUM, proposto inicialmente por Page (1954). O gráfico CUSUM tem por característica uma maior sensibilidade para detectar pequenas mudanças na média do processo, pois incorpora toda a informação das observações. Este gráfico utiliza somas acumuladas dos desvios das observações. O gráfico de controle CUSUM pode ser utilizado para monitorar a média do processo e incorpora diretamente a informação da sequência de valores da amostra, plotando as somas cumulativas dos desvios dos valores da amostra de um valor alvo (média). O CUSUM tabular trabalha acumulando desvios de μ_w que estão acima do valor alvo, com uma estatística C^+ , e acumulando desvios de μ_w que estão abaixo do valor alvo, com outra estatística C^- (MONTGOMERY, 2009). As estatísticas C^+ e C^- são chamadas CUSUM's unilaterais superior e inferior e são determinadas, respectivamente, por:

$$C_t^+ = \max[0, Y_t - (\mu_w + K) + C_{t-1}^+], \quad (2.5)$$

$$C_t^- = \max[0, (\mu_w - K) - Y_t + C_{t-1}^-]. \quad (2.6)$$

em que os valores iniciais são $C_0^+ = C_0^- = 0$, e $t = 1, 2, \dots, n$, sendo n o tamanho da amostra.

Nas Equações (2.5) e (2.6) K é chamado de valor de referência e representa o desvio da média que se pretende detectar. Se os desvios são expressos em unidades de desvio padrão, o valor de K é normalmente escolhido como a metade do desvio que se deseja detectar. O intervalo de decisão, valor de H , é normalmente atribuído em números de desvio padrão. Montgomery (2009) recomenda cinco desvios padrão para o intervalo de decisão e 0,5 desvios padrão para o valor de referência (K). Assim, desvios individuais menores que meio desvio padrão não são considerados para as somas C_t^+ ou C_t^- .

Os limites para o gráfico CUSUM são dados por:

$$\text{LSC} = H, \quad (2.7)$$

$$\text{LIC} = -H. \quad (2.8)$$

Se C_t^+ ou C_t^- excederem o intervalo de decisão $[-H, H]$ o processo é considerado fora de controle.

2.2.3 Gráficos de controle EWMA

Com um desempenho similar ao gráfico de controle CUSUM está o gráfico de controle EWMA, introduzido por Roberts (1959). Os gráficos de controle EWMA foram apresentados principalmente com o intuito de incorporar as limitações dos gráficos de Shewhart. No gráfico EWMA o peso das observações decai à medida que ficam mais antigas.

O GC EWMA é definido por:

$$z_t = \lambda Y_t + (1 - \lambda)z_{t-1}, \quad (2.9)$$

em que $0 < \lambda \leq 1$ é uma constante de suavização. O valor de partida do EWMA é o alvo do processo, então $z_0 = \mu_w$ (média alvo) e Y_t são os valores observados. Os limites de controle são dados por:

$$\text{LSC} = \mu_w + L\sigma_w \sqrt{\frac{\lambda}{(2 - \lambda)} [1 - (1 - \lambda)^{2t}]}, \quad (2.10)$$

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

$$\text{LIC} = \mu_w - L\sigma_w \sqrt{\frac{\lambda}{(2 - \lambda)} [1 - (1 - \lambda)^{2t}]}, \quad (2.11)$$

em que L é a largura dos limites de controle também expressos em unidades de desvio padrão.

Montgomery (2009) mostra que para grandes valores de t o termo $1 - (1 - \lambda)^{2t}$ se aproxima de 1 e os limites (2.10) e (2.11) podem ser considerados como:

$$\text{LSC} = \mu_w + L\sigma_w \sqrt{\frac{\lambda}{(2 - \lambda)}}, \quad (2.12)$$

$$\text{LIC} = \mu_w - L\sigma_w \sqrt{\frac{\lambda}{(2 - \lambda)}}. \quad (2.13)$$

Se os valores de z_t excederem os limites de controle o processo é considerado fora de controle.

2.3 Gráficos de controle para dados autocorrelacionados

Segundo Montgomery (2009), uma abordagem útil quando temos dados autocorrelacionados consiste em modelar diretamente a estrutura de dependência temporal com um modelo

apropriado de séries temporais. Este modelo é usado para remover a autocorrelação dos dados e então aplica-se os gráficos de controle usuais aos resíduos do modelo. Tradicionalmente, considera-se modelos da classe ARIMA (BOX; JENKINS; REINSEL, 2008) para a etapa de modelagem dos dados.

Portanto, quando há autocorrelação entre as unidades amostrais a alternativa é tratar a série de observações sequenciais do processo como uma série temporal. A partir daí, busca-se o melhor modelo estatístico, usualmente da classe ARIMA, para descrever o comportamento da série de observações do processo. Uma vez ajustado o modelo, os seus resíduos são utilizados para monitorar a média do processo, já que uma mudança na média da característica da qualidade Y se reflete diretamente nos resíduos do modelo (BOX; LUCENO, 1997). Considerando que, quando o modelo está corretamente ajustado, os resíduos devem ser independentes, identicamente distribuídos e com distribuição normal, assim os gráficos de controle usuais, descritos na Seção 2.2, podem ser aplicados diretamente à série de resíduos (PADONOU; ROUSTANT; LUTZ, 2015). O uso de um gráfico residual tem a vantagem de que ele pode ser aplicado a quaisquer dados autocorrelacionados mesmo quando não estacionários (ZHANG, 1998).

O CEP para processos autocorrelacionados tem recebido grande atenção na literatura (ALWAN, 1991, 1992; REYNOLDS JR; LU, 1997; LOREDO; JEARKPAPORN; BORROR, 2002; CHENG; THAGA, 2005; NOOROSSANA; VAGHEFI, 2006; CLARO; COSTA; MACHADO, 2007; CROUX; GELPER; MAHIEU, 2011). Isso se deve, em parte, às melhorias na medição e coleta de dados que permitem que os processos sejam amostrados em taxas de frequência mais elevadas e, portanto, autocorrelacionados (YU; LIU, 2011). Quando a suposição de que as observações são independentes é violada (mesmo em baixos níveis de correlação com o tempo), os GC usuais possuem desempenho pobre. Normalmente, esses gráficos de controle terão as taxas de falso alarme mais elevadas do que quando os dados não estão autocorrelacionados (LOREDO; JEARKPAPORN; BORROR, 2002). Alguns autores têm descrito o efeito negativo da autocorrelação em GC sob diferentes abordagens (ALWAN; ROBERTS, 1988; ALWAN, 1992; NOOROSSANA; VAGHEFI, 2006).

Wardell, Moskowitz e Plante (1992) estudam o desempenho dos gráficos de controle usuais de Shewhart, CUSUM e EWMA utilizando resíduos de modelos ajustados. Estes são chamados de gráficos de causas especiais (*SCC-special control chart*). E o gráfico de causas comuns (*CCC-common control charts*) quando é utilizado os valores estimados pelo modelo em dados que possuem um comportamento ARMA(1,1).

Cheng e Thaga (2005) propõem um gráfico CUSUM para dados autocorrelacionados de ordem AR(1). Chang e Wu (2011) estudam as propriedades do ARL de gráficos de controle Shewhart, CUSUM e EWMA para processos AR(1) e AR(2) usando Cadeias de Markov finitas.

Seja $Y = (Y_1, Y_2, \dots, Y_n)$ um processo estocástico (MORETTIN; TOLOI, 2004) e $y = (y_1, y_2, \dots, y_n)$ uma ocorrência de Y . A autocorrelação de Y é uma medida de correlação entre os valores dessa mesma variável aleatória defasados ao longo do tempo t , com $t = 1, 2, \dots, n$. A função de autocorrelação (FAC) é dada por:

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{\text{Cov}(Y_t, Y_{t-k})}{\text{Var}(Y_t)}, \quad k = 0, 1, \dots, \quad (2.14)$$

em que $\text{Cov}(Y_t, Y_{t-k})$ é a covariância das variáveis separadas por k períodos de tempo e supõe-se que Y_t tenha variância constante $\text{Var}(Y_t)$. Estima-se os valores de ρ_k com a FAC amostral:

$$\hat{\rho}_k = \frac{\sum_{t=1}^{n-k} (y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=1}^n (y_t - \bar{y})^2}, \quad (2.15)$$

em que \bar{y} é a média aritmética de y . Os valores de $\hat{\rho}_k$ estão entre -1 e 1 e quanto maior seu valor em módulo, maior a possibilidade de existência de autocorrelação serial no processo estocástico (MORETTIN; TOLOI, 2004).

A autocorrelação é uma medida de dependência entre observações da mesma série separadas por um determinado intervalo k . A FAC pode ter uma conveniente representação gráfica dessa correlação, a qual permite que se entenda melhor o comportamento da dependência estatística entre os dados, sendo útil para determinar qual modelo utilizar para ajustar os dados (SHUMWAY; STOFFER, 2006). Para detectar a presença de autocorrelação pode-se utilizar diferentes métodos: (i) o gráfico de correlograma, baseado no cálculo da FAC e da função de autocorrelação parcial (FACP), (ii) o teste de Box-Pierce (BOX; PIERCE, 1970) ou o teste de Ljung-Box (LJUNG; BOX, 1978), entre outros.

2.3.1 Modelos ARIMA

A metodologia de Box e Jenkins (1976) é representada por um conjunto de características estocásticas, definidas como modelo ARIMA, que incorpora termos autorregressivos (AR), filtro de integração (I) e termos de médias móveis (MA). A implementação dessa metodologia requer que sejam cumpridas as seguintes etapas: identificação, estimação, verificação e previsão. Outras informações podem ser encontradas mais detalhadamente em Morettin e Toloi (2004) e Box, Jenkins e Reinsel (2008).

No modelo autorregressivo de ordem p , $AR(p)$, cada observação corrente Y_t será constituída pela soma ponderada de seus valores passados e de um termo de perturbação aleatória no período corrente. Esse processo $AR(p)$ pode ser descrito como:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + e_t, \quad (2.16)$$

em que ϕ_k , com $k = 1, 2, \dots, p$, são os parâmetros autorregressivos e e_t assumido como $e_t \sim RB(0, \sigma^2)$, em que RB denota ruído branco, ou seja, uma sequência de variáveis aleatórias não-autocorrelacionadas, com média zero e variância finita e constante. Se $e_t \sim \mathcal{N}(0, \sigma^2)$ temos um ruído branco Gaussiano (GUJARATI, 2000).

No modelo de médias móveis de ordem q , $MA(q)$, cada observação Y_t será obtida pela soma ponderada de perturbações aleatórias corrente e defasadas. Tal processo $MA(q)$ pode ser representado por:

$$Y_t = e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q}, \quad (2.17)$$

em que θ_k , com $k = 1, 2, \dots, q$, são os parâmetros de médias móveis.

No modelo misto autorregressivo de médias móveis, $ARMA(p, q)$, cada observação Y_t será uma função de seus valores passados e de perturbações aleatórias defasadas, assim como de um termo de perturbação corrente. Esse processo $ARMA(p, q)$ pode ser expresso por:

$$Y_t = \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + e_t - \theta_1 e_{t-1} - \cdots - \theta_q e_{t-q}, \quad (2.18)$$

em que ϕ_k , com $k = 1, 2, \dots, p$, e θ_k , com $k = 1, 2, \dots, q$, são os parâmetros do modelo.

Se o modelo tiver os componentes autorregressivos, de médias móveis (2.18) e possuir filtro de integração, então se refere ao processo $ARIMA(p, d, q)$. Neste caso, a série original sofre d diferenciações para se tornar estacionária e pode ser representado por:

$$W_t = \phi_1 W_{t-1} + \cdots + \phi_p W_{t-p} + e_t - \theta_1 e_{t-1} - \cdots - \theta_q e_{t-q}, \quad (2.19)$$

em que W_t representa a série diferenciada, $W_t = Y_t - Y_{t-1} = \Delta Y_t$ e então $W_t^d = \Delta^d Y_t$ é a série diferenciada d vezes.

Para a estimação dos parâmetros o método usualmente considerado é o de máxima verossimilhança. Esse método de estimação é de uso extremamente amplo, pois possibilita realizar inferências com propriedades altamente desejáveis (MORETTIN; TOLOI, 2004; KEDEM; FOKIANOS, 2005).

A identificação da ordem do modelo pode ser feita mediante a análise da FAC e FACP amostrais. Aliado a isso, os critérios de informação de Akaike (AIC) (AKAIKE, 1974) e o

critério de informação de Schwarz (BIC) (SCHWARZ, 1978) podem auxiliar na identificação do modelo. O AIC e o BIC são dados, respectivamente, por:

$$\text{AIC} = -2\hat{\ell} + 2(k + 1), \quad (2.20)$$

$$\text{BIC} = -2\hat{\ell} + (k + 1) \log(n), \quad (2.21)$$

em que k é o número de parâmetros do modelo e $\hat{\ell}$ a log-verossimilhança maximizada. O melhor modelo será aquele que obtiver o menor valor para os critérios AIC e BIC.

Contudo, para a estimação dos parâmetros dos modelos ARIMA assume-se normalidade dos dados (BOX; JENKINS, 1976; MORETTIN; TOLOI, 2004). Quando não for razoável assumir normalidade para a distribuição das observações do processo, particularmente quando estes forem do tipo frações e proporções, a modelagem ARIMA pode não ser adequada. Com isso, o modelo β ARMA (ROCHA; CRIBARI-NETO, 2009) pode ser mais indicado.

2.3.2 Modelos EWMA

O modelo EWMA, ou também conhecido como suavização exponencial simples (SES), pondera as observações passadas com pesos decrescentes exponencialmente para previsão de valores futuros, sendo uma alternativa não-paramétrica, não considerando qualquer distribuição de probabilidade.

Sendo Y_1, Y_2, \dots, Y_n uma série estacionária, o modelo de suavização exponencial simples pode ser descrito por (MORETTIN; TOLOI, 2004):

$$\bar{Y}_t = \lambda Y_t + (1 - \lambda)\bar{Y}_{t-1}, \quad \bar{Y}_0 = Y_1, \quad t = 1, 2, \dots, n, \quad (2.22)$$

ou

$$\bar{Y}_t = \lambda \sum_{k=0}^{t-1} (1 - \lambda)^k Y_{t-k} + (1 - \lambda)^t \bar{Y}_0, \quad (2.23)$$

em que \bar{Y}_t é denominado de valor exponencialmente suavizado, λ a constante de suavização e $0 \leq \lambda \leq 1$.

Expandindo a Equação (2.23), temos:

$$\bar{Y}_t = \lambda Y_t + \lambda(1 - \lambda)Y_{t-1} + \lambda(1 - \lambda)^2 Y_{t-2} + \dots + \lambda(1 - \lambda)^k Y_{t-k}. \quad (2.24)$$

Isso significa que o alisamento exponencial simples é uma média ponderada que dá pesos maiores às observações mais recentes (MORETTIN; TOLOI, 2004).

A previsão dos valores futuros é dada pelo último valor exponencialmente alisado:

$$\hat{Y}_{t+1}(h) = Y_t \quad \forall h > 0, \quad (2.25)$$

ou

$$\hat{Y}_t(h) = \lambda Y_t + \lambda(1 - \lambda)\hat{Y}_{t-1}(h + 1), \quad (2.26)$$

que pode ser interpretada como uma atualização de previsão, dado que tem-se uma nova observação. Para $h = 1$, pode-se notar que a Equação (2.26) se reduz a:

$$\hat{Y}_t(1) = \lambda e_t + \hat{Y}_{t-1}(1), \quad (2.27)$$

em que $e_t = Y_t - \hat{Y}_{t-1}(1)$ é o erro de previsão a um passo.

Quanto menor for o valor de λ , mais estáveis serão suas previsões. Um baixo valor de λ implicaria que pesos maiores sejam dados as observações passadas e conseqüentemente, qualquer flutuação aleatória, no presente, exercerá um menor peso no cálculo da previsão (ABBAS; RIAZ; DOES, 2014). O valor da constante de alisamento pode ser escolhida também através da minimização de uma das funções dos erros de previsão na amostra, ou seja, minimizando $\sum_{t=1}^n e_t^2$.

3 GRÁFICOS DE CONTROLE PARA DADOS DO TIPO TAXAS E PROPORÇÕES

Neste capítulo são apresentados conceitos e métodos necessários para a proposição dos gráficos de controle utilizando os resíduos do modelo β ARMA.

3.1 Gráfico de controle beta

Sant'Anna e Ten Caten (2012) propõem o gráfico de controle beta, no qual assumem que a variável do tipo fração ou proporção, restrita no intervalo $(0,1)$, pode ser modelada utilizando a distribuição beta. O gráfico de controle beta acomoda naturalmente a assimetria de dados do tipo fração e proporção e os limites de controle ficam restritos ao intervalo $(0,1)$, o que pode não acontecer nos gráficos usuais que assumem aproximação pela normal.

A distribuição beta é utilizada para modelar dados nos quais a variável de interesse é distribuída continuamente no intervalo de uma unidade padrão $(0,1)$. Sendo Y uma variável aleatória com distribuição beta, sua função densidade de probabilidade (fdp) é dada por:

$$f(y; \varrho_1, \varrho_2) = \frac{\Gamma(\varrho_1 + \varrho_2)}{\Gamma(\varrho_1)\Gamma(\varrho_2)} y^{(\varrho_1-1)}(1-y)^{(\varrho_2-1)}, 0 < y < 1, \quad (3.1)$$

em que $\varrho_1 > 0$, $\varrho_2 > 0$ e $\Gamma(\cdot)$ é a função gama, em que $\Gamma(z) = \int_0^{\infty} t^{z-1} e^{-t} dt$. A esperança e a variância de Y são, respectivamente:

$$E(Y) = \frac{\varrho_1}{\varrho_1 + \varrho_2}, \quad (3.2)$$

$$\text{Var}(Y) = \frac{\varrho_1 \varrho_2}{(\varrho_1 + \varrho_2)^2 (\varrho_1 + \varrho_2 + 1)}. \quad (3.3)$$

Os limites de controle do gráfico de controle beta proposto por Sant'Anna e Ten Caten (2012) são definidos como:

$$\text{LSC} = \tilde{\pi} + w_2 \sqrt{s^2(\tilde{\pi})}, \quad (3.4)$$

$$\text{LIC} = \tilde{\pi} - w_1 \sqrt{s^2(\tilde{\pi})}, \quad (3.5)$$

em que $\tilde{\pi}$ e $s^2(\tilde{\pi})$ são a média e a variância da variável fração que está sendo monitorada, w_1 e w_2 constantes tabeladas que definem a largura dos limites de controle. Esses valores w_1 e w_2

podem ser dados por (SANT'ANNA; TEN CATEN, 2012):

$$w_1 = \frac{\tilde{\pi} - Q([\alpha/2], \varrho_1, \varrho_2)}{\sqrt{s^2(\tilde{\pi})}}, \quad (3.6)$$

$$w_2 = \frac{Q([1 - \alpha/2], \varrho_1, \varrho_2) - \tilde{\pi}}{\sqrt{s^2(\tilde{\pi})}}, \quad (3.7)$$

em que Q é a função quantil da distribuição beta de parâmetros ϱ_1 e ϱ_2 . Isto é, $Q(\alpha; \varrho_1, \varrho_2) = F^{-1}(\alpha; \varrho_1, \varrho_2)$, em que F é a função distribuição, $F(y; \varrho_1, \varrho_2) = \int_0^y f(z; \varrho_1, \varrho_2) dz$.

Contudo, substituindo (3.6) em (3.5) e (3.7) em (3.4) obtemos diretamente os seguintes limites de controle:

$$\text{LSC} = Q([\alpha/2], \varrho_1, \varrho_2), \quad (3.8)$$

$$\text{LIC} = Q([1 - \alpha/2], \varrho_1, \varrho_2), \quad (3.9)$$

em que, neste trabalho, assumiremos que ϱ_1 e ϱ_2 podem ser substituídos por suas respectivas estimativas de máxima verossimilhança.

O gráfico de controle beta proposto pelos autores Sant'Anna e Ten Caten (2012) assume a independência dos dados. A partir da ideia do uso de uma distribuição mais adequada para as variáveis do tipo taxas e proporções, percebeu-se a necessidade de ser utilizada uma mudança na metodologia para dados que sejam beta distribuídos e que possuam autocorrelação.

3.2 Gráficos de controle propostos

A suposição para construção de gráficos de controle é de que um processo sob controle é identificado quando gera variáveis aleatórias *i.i.d.* Na prática, no entanto, é difícil de atingir um estado de controle estatístico no sentido de um processo *i.i.d.*, pois autocorrelações e outros efeitos de séries temporais são muitas vezes substanciais em alguns processos (ALWAN, 1992).

Liu e Tang (1996) afirmam que, geralmente, as observações de um processo de fabricação, por exemplo, são geradas a intervalos de tempo espaçados igualmente, existindo a correlação cruzada entre as observações vizinhas. Quando as observações são dependentes, os métodos não são válidos, porque geralmente se subestima (ou superestima) a variação da média da amostra se as observações são positivamente (ou negativamente) correlacionadas.

Variáveis resposta contínuas duplamente limitadas, como frações e proporções, são frequentemente encontradas em muitas áreas práticas. Quando o processo não é *i.i.d.*, este tipo

de dados é geralmente examinado por meio de modelos ARIMA. Apesar da viabilidade, esta modelagem pode conter deficiências, pois os modelos ARIMA assumem normalidade dos dados, podendo gerar valores previstos ou limites de controle fora do intervalo $(0,1)$. Dados do tipo frações e proporções podem ser tipicamente melhor modelados assumindo a distribuição beta, pois esta, dependendo dos valores dos seus parâmetros, pode assumir diferentes formas (FERRARI; CRIBARI-NETO, 2004). Por esse motivo, propomos gráficos de controle residuais baseados no modelo β ARMA.

3.2.1 Modelo β ARMA

O modelo de regressão beta foi proposto por Ferrari e Cribari-Neto (2004) para modelar variáveis contínuas em um intervalo $(0,1)$, tais como dados do tipo taxas e proporções. Ferrari e Cribari-Neto (2004) propuseram uma reparametrização para a densidade beta, onde $\mu = \frac{\varrho_1}{\varrho_1 + \varrho_2}$ e $\varphi = \varrho_1 + \varrho_2$. Com essa parametrização, a função densidade de probabilidade, a esperança e a variância de Y são dadas, respectivamente, por:

$$f(y; \mu, \varphi) = \frac{\Gamma(\varphi)}{\Gamma(\mu\varphi)\Gamma((1-\mu)\varphi)} y^{(\mu\varphi-1)}(1-y)^{((1-\mu)\varphi-1)}, 0 < y < 1, \quad (3.10)$$

$$E(Y) = \mu, \quad (3.11)$$

$$\text{Var}(Y) = \frac{V(\mu)}{(1 + \varphi)}, \quad (3.12)$$

em que $V(\mu) = \mu(1 - \mu)$ é a função de variância, μ é a média e φ o parâmetro de precisão.

Kieschnick e McCullough (2003) mostram que a utilização da distribuição beta é mais apropriada para dados do tipo taxas e proporções, devido à variedade de forma (simétrica ou assimétrica) que a densidade pode assumir. Algumas das diferentes formas da distribuição beta podem ser visualizadas na Figura 3.1.

Rocha e Cribari-Neto (2009) afirmam que as distribuições de taxas e proporções são tipicamente assimétricas e, portanto, inferir com base na distribuição normal não é apropriado. O modelo β ARMA naturalmente acomoda assimetrias e também dispersão não-constante, assumindo distribuição beta para a variável de interesse e uma estrutura ARMA para sua média.

Tendo como objetivo definir um modelo para séries temporais para dados do tipo fração e/ou proporção, Rocha e Cribari-Neto (2009) desenvolveram o modelo β ARMA como uma extensão dinâmica do modelo de regressão beta. Este modelo assume que a variável de interesse Y_t possui distribuição beta com densidade condicional às observações prévias definida pela

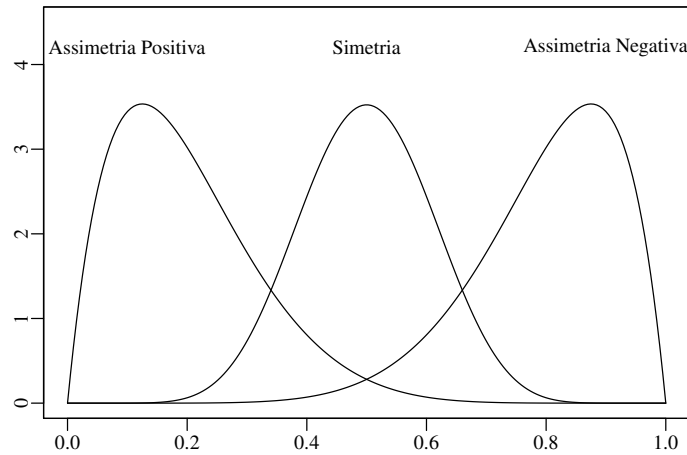


Figura 3.1: Diferentes formas da densidade beta.

Equação (3.10). Ao considerar que a variável de interesse possui distribuição beta, o suporte dessa variável consiste no intervalo unitário contínuo $(0,1)$. Um caso particular do modelo $\beta\text{ARMA}(p,q)$, quando não há variáveis explicativas, é definido por:

$$g(\mu_t) = \alpha + \sum_{i=1}^p \phi_i g(y_{t-i}) + \sum_{j=1}^q \theta_j e_{t-j}, \quad (3.13)$$

em que o termo α é o valor do intercepto, os índices p e q são as ordens autorregressiva e de médias móveis do modelo em que os parâmetros são, respectivamente, $\phi = (\phi_1, \phi_2, \dots, \phi_p)^\top$ e $\theta = (\theta_1, \theta_2, \dots, \theta_q)^\top$. Além disso, $e_t = y_t - \mu_t$ representa os termos de erro e $g(\cdot)$ representa a função de ligação como, por exemplo, as funções logit e probit.

Utiliza-se o método de máxima verossimilhança para estimação do vetor paramétrico $\gamma = (\alpha, \varphi, \phi^\top, \theta^\top)^\top$. A função de log-verossimilhança condicional às primeiras m observações, onde $m = \max\{p, q\}$, do vetor de parâmetros γ pode ser escrita como $\ell(y; \gamma) = \sum_{t=m+1}^n \ell_t(\mu_t, \varphi)$, em que:

$$\begin{aligned} \ell_t(\mu_t, \varphi) &= \log \Gamma(\varphi) - \log \Gamma(\mu_t \varphi) - \log \Gamma((1 - \mu_t) \varphi) \\ &+ (\mu_t \varphi - 1) \log y_t + \{(1 - \mu_t) \varphi - 1\} \log(1 - y_t). \end{aligned} \quad (3.14)$$

O vetor escore $U(\gamma)$ é construído a partir das derivadas parciais da log-verossimilhança condicional em relação a cada um dos parâmetros desconhecidos α , φ , ϕ e θ . Assim, se

$$y_t^* = \log \left\{ \frac{y_t}{(1 - y_t)} \right\}, \quad (3.15)$$

$$\mu_t^* = \psi(\mu_t, \varphi) - \psi((1 - \mu_t) \varphi), \quad (3.16)$$

e $\psi(\cdot)$ é a função digama, ou seja, $\psi(z) = d \log \Gamma(z)/dz$, para $z > 0$, temos:

$$U_\alpha(\gamma) = \frac{\partial \ell(y; \gamma)}{\partial \alpha} = \varphi \mathbf{1}^\top T(y^* - \mu^*), \quad (3.17)$$

$$U_\varphi(\gamma) = \frac{\partial \ell(y; \gamma)}{\partial \varphi} = \sum_{t=m+1}^n \{\mu_t(y_t^* - \mu_t^*) + \log(1 - y_t) - \psi((1 - \mu_t)\varphi) + \psi(\varphi)\}, \quad (3.18)$$

$$U_\phi(\gamma) = \frac{\partial \ell(y; \gamma)}{\partial \phi} = \varphi P^\top T(y^* - \mu^*), \quad (3.19)$$

$$U_\theta(\gamma) = \frac{\partial \ell(y; \gamma)}{\partial \theta} = \varphi R^\top T(y^* - \mu^*), \quad (3.20)$$

em que $y^* = (y_{m+1}^*, \dots, y_n^*)^\top$, $\mu^* = (\mu_{m+1}^*, \dots, \mu_n^*)^\top$, $T = \text{diag}(1/g(\mu_{m+1}), \dots, 1/g(\mu_n))$, $\mathbf{1}$ é um vetor unitário, P é uma matriz $(n - m) \times p$ cujo elemento (t, j) é dado por $g(y_{t+m-j})$ e R é uma matriz $(n - m) \times q$ em que o elemento da t -ésima linha e j -ésima coluna é dado por e_{t+m-j} .

Os estimadores de máxima verossimilhança (EMV) são obtidos pela resolução do seguinte sistema:

$$\begin{cases} U_\alpha = 0 \\ U_\varphi = 0 \\ U_\phi = 0 \\ U_\theta = 0 \end{cases}. \quad (3.21)$$

A solução desse sistema não possui uma forma fechada, podendo ser numericamente obtido através da maximização da função de log-verossimilhança condicional utilizando algoritmos de otimização não-linear, como o algoritmo quasi-Newton BFGS (NOCEDAL; WRIGHT, 1999).

A matriz de informação de Fisher conjunta ($K(\gamma)$) pode ser expressa por:

$$K = K(\gamma) = \begin{bmatrix} K_{\alpha\alpha} & K_{\alpha\varphi} & K_{\alpha\phi} & K_{\alpha\theta} \\ K_{\varphi\alpha} & K_{\varphi\varphi} & K_{\varphi\phi} & K_{\varphi\theta} \\ K_{\phi\alpha} & K_{\phi\varphi} & K_{\phi\phi} & K_{\phi\theta} \\ K_{\theta\alpha} & K_{\theta\varphi} & K_{\theta\phi} & K_{\theta\theta} \end{bmatrix},$$

em que $K_{\alpha\alpha} = \varphi \text{tr}(W)$, $K_{\alpha\varphi} = K_{\varphi\alpha}^\top = \mathbf{1}^\top Tc$, $K_{\alpha\phi} = K_{\phi\alpha}^\top = \varphi P^\top W \mathbf{1}$, $K_{\alpha\theta} = K_{\theta\alpha}^\top = \varphi R^\top W \mathbf{1}$, $K_{\varphi\varphi} = \text{tr}(D)$, $K_{\varphi\phi} = K_{\phi\varphi}^\top = P^\top Tc$, $K_{\varphi\theta} = K_{\theta\varphi}^\top = R^\top Tc$, $K_{\phi\phi} = \varphi P^\top W P$, $K_{\phi\theta} = K_{\theta\phi}^\top = R^\top W P$ e $K_{\theta\theta} = \varphi R^\top W R$. Ainda, $W = \text{diag}\{w_{m+1}, \dots, w_n\}$, com $w_t = \varphi \{\psi(\mu_t \varphi) + \psi'((1 - \mu_t)\varphi)\} / g'(\mu_t)^2$, $c = (c_{m+1}, \dots, c_n)^\top$, em que $c_t = \varphi \{\psi'(\mu_t \varphi) \mu_t - \psi'((1 - \mu_t)\varphi)(1 - \mu_t)\}$, e $D = \text{diag}\{d_{m+1}, \dots, d_n\}$, com $d_t = \psi'(\mu_t \varphi) \mu_t^2 - \psi'((1 - \mu_t)\varphi)(1 - \mu_t)^2 - \psi'(\varphi)$.

Sob certas condições de regularidade e tamanhos amostrais grandes, a distribuição conjunta dos estimadores dos parâmetros α , φ , ϕ e θ é aproximadamente normal k -dimensional, de

modo que:

$$\begin{pmatrix} \hat{\alpha} \\ \hat{\varphi} \\ \hat{\phi} \\ \hat{\theta} \end{pmatrix} \sim \mathcal{N} \left(\begin{pmatrix} \alpha \\ \varphi \\ \phi \\ \theta \end{pmatrix}, K^{-1} \right),$$

em que $\hat{\alpha}$, $\hat{\varphi}$, $\hat{\phi}$ e $\hat{\theta}$ são os EMV's de α , φ , ϕ e θ , respectivamente, e $k = p + q + 2$ é o número de parâmetros do modelo. Para auxiliar no processo de seleção de modelos podem ser utilizados critérios de seleção de modelos como AIC e BIC, bem como suas versões corrigidas (GUERRA, 2015).

A análise de resíduos é uma etapa importante e visa identificar observações atípicas e erros de especificação do modelo. Os resíduos são medidas de conformidade entre os dados e o modelo ajustado (ESPINHEIRA; FERRARI; CRIBARI-NETO, 2008). Para o β ARMA diferentes resíduos podem ser considerados. Por exemplo, os resíduos para regressão beta podem ser diretamente empregados. Ferrari e Cribari-Neto (2004) definem o resíduo ordinário padronizado para a regressão beta, dado por:

$$r_t^1 = \frac{y_t - g(\hat{\mu}_t)}{\sqrt{\widehat{\text{Var}}(y_t)}}; \quad (3.22)$$

neste trabalho denominado resíduo 1. Rocha e Cribari-Neto (2009) apresentam a medida dos termos de erro em função da escala do preditor, e este é dado por:

$$r_t^2 = \frac{g(y_t) - g(\hat{\mu}_t)}{\sqrt{g'(\hat{\mu}_t)V(\hat{\mu}_t)(1 + \varphi)}} = \frac{g(y_t) - g(\hat{\mu}_t)}{\approx \text{Var}(g(\mu_t))}, \quad (3.23)$$

neste trabalho denominado por resíduo 2.

Espinheira, Ferrari e Cribari-Neto (2008) definem o resíduo ponderado padronizado, dado por:

$$r_t^3 = \frac{y_t^* - \hat{\mu}_t^*}{\sqrt{v_t}}, \quad (3.24)$$

em que $v_t = \{\psi'(\mu_t\varphi) + \psi'((1 - \mu_t)\varphi)\}$. Neste trabalho este resíduo será denominado resíduo 3.

Ferrari e Cribari-Neto (2004) também propõe o resíduo *deviance*, dado por:

$$r_t^4 = \text{sign}(y_t - \hat{\mu}_t) \{2(\ell_t(y_t, \hat{\varphi}) - \ell_t(\hat{\mu}_t, \hat{\varphi}))\}^{\frac{1}{2}}, \quad (3.25)$$

em que $\ell_t(y_t, \hat{\varphi})$ e $\ell_t(\hat{\mu}_t, \hat{\varphi})$, são respectivamente as contribuições para a função de log-verossimilhança das observações y_t e das estimativas $\hat{\mu}_t$. Neste trabalho esse resíduo será denominado resíduo 4.

O gráfico de controle proposto consiste em encontrar o melhor modelo β ARMA explicativo da variável de interesse e então definir limites de controle utilizando gráficos de controle usuais definidos na Seção 2.2 para os diferentes resíduos. Espera-se que a modelagem utilizando o β ARMA traga melhores resultados pelo suporte da variável resposta do modelo consistir no intervalo unitário contínuo $(0,1)$, encaixando-se adequadamente às características de dados do tipo frações e proporções.

O gráfico de resíduos é obtido pela plotagem da sequência dos resíduos r_1, r_2, \dots, r_n , juntamente com os limites de controle definidos nas subseções a seguir. Para esses gráficos qualquer um dos quatro resíduos definidos anteriormente podem ser considerados.

3.2.2 Gráfico β ARMA Shewhart

O gráfico β ARMA Shewhart é obtido utilizando o GC usual de Shewhart nos resíduos do modelo β ARMA. Os limites serão dados por:

$$\text{LSC} = \bar{r} + k\sigma_r, \quad (3.26)$$

$$\text{LC} = \bar{r}, \quad (3.27)$$

$$\text{LIC} = \bar{r} - k\sigma_r, \quad (3.28)$$

em que \bar{r} e σ_r são, respectivamente, a média e o desvio padrão do resíduo considerado, k é a distância dos limites de controle superior e inferior do limite central expressa em unidades de desvio padrão. Quando r_t estiver fora dos limites de controle, então o processo será considerado fora de controle.

3.2.3 Gráfico β ARMA CUSUM

Considerando o GC CUSUM nos resíduos do β ARMA propomos o GC β ARMA CUSUM. Após as devidas adequações, dado um resíduo r_t , as estatísticas CUSUM's unilaterais superiores e inferiores são dadas por:

$$C_t^+ = \max[0, C_{t-1}^+ + r_t - K], \quad (3.29)$$

$$C_t^- = \max[0, C_{t-1}^- - r_t - K], \quad (3.30)$$

em que os valores iniciais são $C_0^+ = C_0^- = 0$. Da mesma forma que no CUSUM usual, os limites são dados pelo valor H do valor de referência do intervalo de decisão:

$$\text{LSC} = H, \quad (3.31)$$

$$\text{LIC} = -H. \quad (3.32)$$

Dessa forma, nos instantes t em que C_t^+ e C_t^- estiverem fora dos limites o processo será considerado fora de controle.

3.2.4 Gráfico β ARMA EWMA

Adaptando o GC EWMA usual para os resíduos do β ARMA, temos os seguintes limites de controle:

$$\text{LSC} = \bar{r} + L\sigma_r \sqrt{\frac{\lambda}{(2-\lambda)} [1 - (1-\lambda)^{2t}]}, \quad (3.33)$$

$$\text{LC} = \bar{r},$$

$$\text{LIC} = \bar{r} - L\sigma_r \sqrt{\frac{\lambda}{(2-\lambda)} [1 - (1-\lambda)^{2t}]}, \quad (3.34)$$

em que L é a largura dos limites de controle também em unidades de desvio padrão. Quando r_t estiver fora dos limites de controle, então o processo será considerado fora de controle.

Cada um dos quatro resíduos do β ARMA será utilizado para análise de desempenho e implicará em um GC alternativo em todas as versões (Shewhart, CUSUM e EWMA), totalizando doze gráficos de controle propostos.

4 AVALIAÇÃO NUMÉRICA

4.1 Simulação de Monte Carlo e análise de desempenho

Para realizar uma análise de desempenho em gráficos de controle uma das formas é analisar o seu ARL, ou comprimento médio de corrida, que é o número médio de amostras até que uma causa especial seja detectada (MONTGOMERY, 2009). Duas medidas de desempenho foram utilizadas para avaliar a qualidade dos gráficos de controle, nomeadamente: o ARL sob controle (ARL_0) e o ARL fora de controle (ARL_1). O ARL sob controle representa o número de amostras observadas do processo até que um alarme falso seja detectado. Já o ARL fora de controle representa o número de amostras observadas do processo até que um alarme verdadeiro seja detectado.

Segundo Montgomery (2009), o ARL_0 pode ser interpretado em função do erro do tipo I de um teste de hipóteses. Já o ARL_1 é interpretado como função do erro do tipo II. Essas quantidades são descritas, matematicamente, por:

$$ARL_0 = \frac{1}{\alpha}, \quad (4.1)$$

$$ARL_1 = \frac{1}{[1 - \beta]}, \quad (4.2)$$

em que $\alpha = P(Y_t \notin [LIC, LSC] | \mu_t = \mu_w)$ é a probabilidade de alarmes falsos, $\beta = P(Y_t \in [LIC, LSC] | \mu_t = \mu' = \mu_w + \delta)$, $1 - \beta$ é a probabilidade de alarmes verdadeiros e δ representa uma mudança no processo.

Normalmente, deseja-se que o ARL seja grande quando nenhuma causa atribuível ocorreu (ARL_0), indicando uma baixa frequência de falsos alarmes, e pequeno se ocorreu (ARL_1), indicando maior poder na detecção da mudança (CHEN et al., 2015). O conhecimento da ARL para uma causa atribuível particular (normalmente uma mudança no processo) nos permite projetar gráficos de controle mais eficazes (WARDELL; MOSKOWITZ; PLANTE, 1992, 1994; ZHANG, 1997). Para um mesmo ARL_0 fixado, o melhor gráfico de controle será aquele que tenha menor ARL_1 .

No presente trabalho utilizamos como critério de sinalização de processo fora de controle a existência de pontos fora dos limites. Os demais critérios e regras de sensibilidade existentes, como a presença de uma sequência de pontos acima ou abaixo da linha central ou uma

tendência crescente ou decrescente de pontos, por questões de brevidade, não serão abordados. Para maiores detalhes, ver Montgomery (2009).

A análise de desempenho foi feita através de simulação computacional utilizando o método de simulação de Monte Carlo. Uma avaliação numérica da performance (em termos do ARL sob controle e fora de controle) dos gráficos de controle para processos com um comportamento β ARMA(1,1) foi feita. A simulação foi desenvolvida por meio de linguagem de programação R (R Core Team, 2014). Utilizou-se também funções disponíveis nos pacotes `qcc` (SCRUCCA, 2004) e `spc` (KNOTH, 2016).

Através de algoritmo computacional, seqüências de dados pseudo-aleatórios foram gerados de ocorrências de um processo β ARMA(1,1), com diferentes níveis de dispersão. O parâmetro de precisão assumiu valores iguais a 40,100 e 200, representando, respectivamente, grande dispersão, dispersão média e pequena dispersão. Foram considerados diferentes valores para os parâmetros α , ϕ_1 e θ_1 , conforme cenários apresentados na Tabela 4.1. Os cenários foram desenvolvidos para contemplar diferentes características possíveis dos dados ao longo do intervalo (0,1), abrangendo as possibilidades de assimetria e simetria, além de diferentes níveis de dispersão, totalizando nove cenários.

Tabela 4.1: Cenários simulados.

μ	Parâmetros	φ	Cenários	Tipo
$\approx 0,2$	$\alpha = -0,8$	40	Cenário 1	Assimetria positiva
	$\phi_1 = 0,5$	100	Cenário 2	
	$\theta_1 = 0,45$	200	Cenário 3	
$\approx 0,5$	$\alpha = 0$	40	Cenário 4	Simetria
	$\phi_1 = 0,5$	100	Cenário 5	
	$\theta_1 = 0,45$	200	Cenário 6	
$\approx 0,8$	$\alpha = 0,8$	40	Cenário 7	Assimetria negativa
	$\phi_1 = 0,5$	100	Cenário 8	
	$\theta_1 = 0,45$	200	Cenário 9	

Foram conduzidos dois estudos de simulação de Monte Carlo. Na primeira etapa, ou Fase I, buscou-se calibrar os gráficos de controle em termos de ARL sob controle, ou seja, os gráficos foram equiparados para um mesmo valor de $ARL_0 = 200$, garantindo um mesmo nível de erro do tipo I. Para a calibração foram ajustados modelos de regressão lineares, como em Moraes D.A.O. et al. (2014), em que a variável dependente foi o valor nominal da probabilidade de alarmes falsos e a independente foi o ARL_0 empírico retornado pela simulação. Detalhes

desses ajustes estão no Apêndice A.

Para os GC EWMA, Psarakis, Vyniou e Castagliola (2014) mostraram que para valores pequenos do λ a taxa de alarme falso para o gráfico com parâmetros estimados é superior à do gráfico com parâmetros conhecidos. Isto poderia indicar a utilização de valores maiores de λ , embora a escolha da constante de alisamento também dependa da sensibilidade do gráfico desejado na detecção de mudanças médias de uma magnitude específica. O parâmetro λ determina a velocidade a que a informação passada entra no cálculo da estatística EWMA. Quanto maior o valor de λ mais peso terão as informações atuais e menor peso as informações passadas, ao passo que um pequeno valor de λ dá mais peso às informações passadas e menos peso para a informação atual. Quando $\lambda = 1$, apenas a observação mais recente é usada para o cálculo da estatística EWMA e, em consequência, o GC EWMA coincide com o gráfico Shewhart (ABBAS; RIAZ; DOES, 2014). Montgomery (2009) sugere para λ valores no intervalo (0,05; 0,25). Dessa forma, assim como em Zhang (1998), para o GC EWMA fixamos $\lambda = 0,2$. Para o GC CUSUM, como sugerido em Loredo, Jearpaporn e Borrer (2002), fixamos $K = 0,5$.

Schaffer e Kim (2007) mostram que simulações de Monte Carlo têm sido amplamente utilizadas no estudo do desempenho de GC e que em diversos trabalhos os autores utilizam diferentes números de repetições em seus estudos. Eles analisaram seis estudos publicados para propor recomendações para o número mínimo de repetições necessárias e obtiveram como resultados a indicação de 10.000 réplicas de Monte Carlo. Até mesmo um menor número de repetições pode ser utilizado, em muitos casos, foram necessárias apenas 5.000 repetições. Com isso, o número de réplicas de Monte Carlo utilizadas neste estudo foram de 10.000. Encontrou-se os valores de ARL_0 para cada gráfico para uma sequência de diferentes valores da probabilidade de erro do tipo I, variando entre 1,0 e 3,8 em incrementos de tamanho 0,2, e foi feito o ajuste de regressão linear para facilitar o cálculo do ARL_0 através de um modelo. Detalhes no Apêndice A.

Na segunda etapa foram simuladas novas ocorrências do processo β ARMA(1,1) com uma perturbação aditiva na média, a fim de identificar o efeito da detecção do ARL_1 dos gráficos de controle quando submetidos à alguma mudança no processo, ou Fase II. Neste trabalho, foi considerada uma mudança aditiva da seguinte forma: $g(\mu_t) = \alpha + \delta + \phi_1 g(y_{t-1}) + \theta_1 e_{t-1}$, para diferentes magnitudes de δ . Quando $\delta = 0$ o processo está sob controle.

Para a análise do ARL_1 são geradas amostras de tamanho 2.200. As 200 primeiras observações do processo em controle e nas 2.000 últimas são feitas mudanças na média de diferentes

magnitudes, com $\delta = (-0,20; -0,15; -0,10; -0,5; 0; 0,5; 0,10; 0,15; 0,20)$. Em cada réplica de Monte Carlo os diferentes modelos são ajustados baseados nas 200 primeiras observações (ARIMA, EWMA e β ARMA) e os resíduos destes modelos nas 2000 observações fora de controle são plotados juntamente com os limites de controle ajustados. Em cada réplica é guardado o valor do número de observações (RL) (*run length*) até que ocorra a primeira observação fora dos limites de controle. Ao final das 10.000 réplicas, calcula-se a média dos valores (RL) de cada réplica, obtendo uma estimativa de Monte Carlo do ARL. Em nossas simulações, os GC propostos, utilizando os resíduos do modelo β ARMA, foram avaliados comparativamente aos GC baseados nos resíduos dos modelos ARIMA e EWMA.

4.2 Resultados

Esta seção apresenta os resultados e as discussões da avaliação numérica descrita na seção anterior. Pelo fraco poder de detecção dos gráficos baseados nos resíduos do modelo EWMA, distorcendo as análises gráficas, esses resultados não foram incluídos nas discussões a seguir, mas seu comportamento pode ser observado no Apêndice B. Os resultados para os Cenários 7, 8, e 9 que possuem como característica assimetria negativa, com média em torno de 0,9, por serem semelhantes aos resultados do Cenários 1, 2 e 3, foram suprimidos da discussão, mas encontram-se no Apêndice C.

Os resultados apresentados pela Figura 4.1 mostram o desempenho dos GC residuais de Shewhart, para os Cenários 1, 2, 3, 4, 5 e 6. Nestes gráficos, nos Cenários 1, 2 e 3, em que a média do processo está próxima de zero, observa-se visualmente o fraco poder de detecção dos GC baseados nos resíduos do modelo ARMA, quando $\delta < 0$. Já para os GC baseados nos resíduos do modelo β ARMA, o baseado no resíduo 1, para $\delta < 0$, possui um comportamento semelhante ao baseado no resíduo ARMA. Já os GC baseados nos resíduos 2, 3 e 4 têm bom poder de detecção no extremo do intervalo em todos os níveis de precisão. Para o Cenário 1, com maior dispersão, o ARL_1 dos GC baseados nos resíduos 2, 3 e 4, para $\delta = -0,2$ foram, respectivamente, 33, 54 e 100, enquanto que para o GC ARMA foi de 728. Para o Cenário 2, com grau médio de dispersão, o ARL_1 dos GC baseados nos resíduos 2, 3 e 4, para $\delta = -0,2$ foram, respectivamente, 21, 37 e 58. Para o Cenário 3, com pouca dispersão, o ARL_1 dos GC baseados nos resíduos 2, 3 e 4, para $\delta = -0,2$ foram, respectivamente, 14, 24 e 34. Para $\delta > 0$ os GC baseados nos resíduos 2 e 3 ficam um pouco instáveis, ou seja, perdem seu poder de

deteccoo assim como o GC ARMA para $\delta < 0$. Porm, o GC baseado no resduo 4 mantm

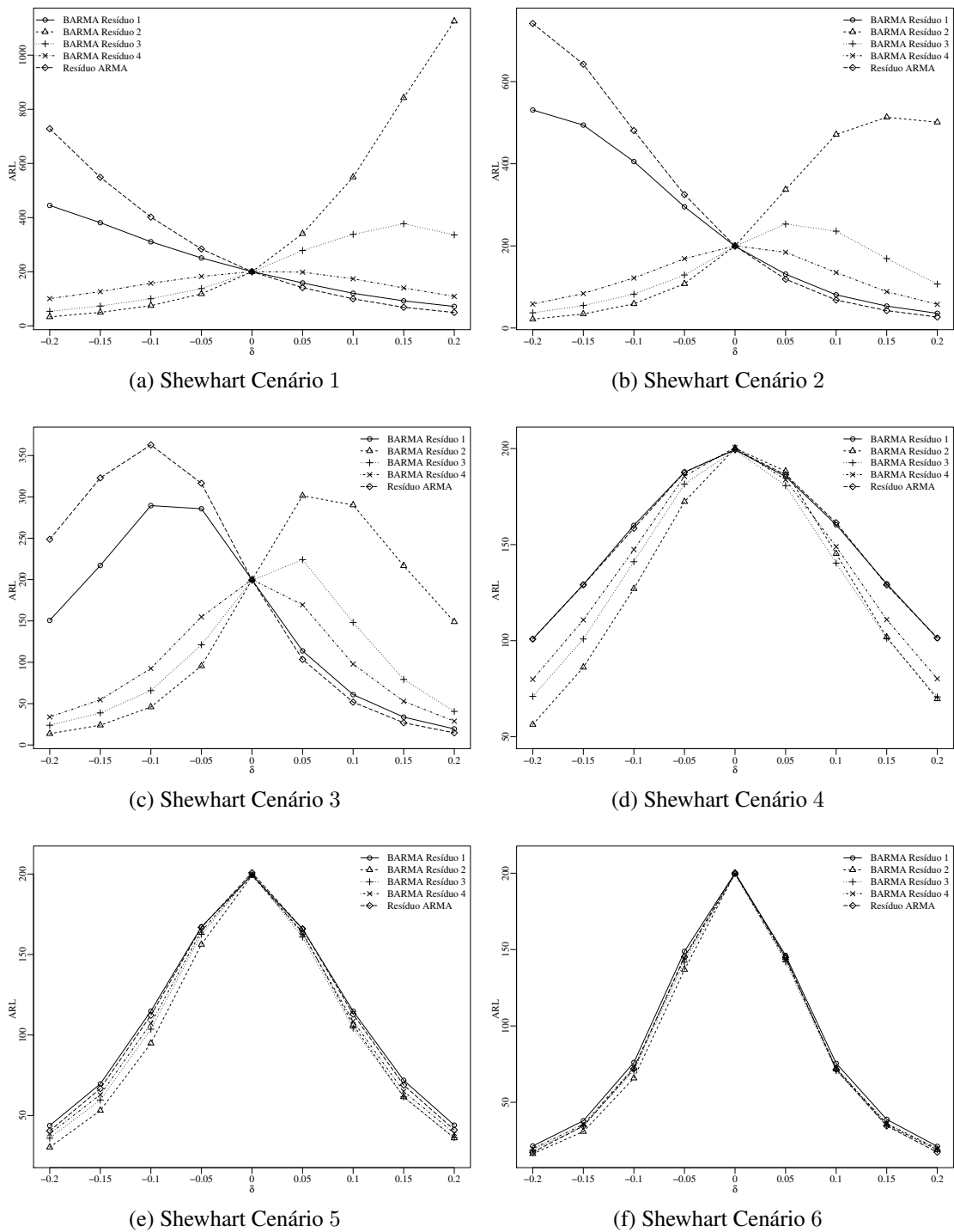


Figura 4.1: ARL_1 para os grficos de controle residuais de Shewhart.

seu poder de deteccoo regular, para todos os cenrios. Para os Cenrios 4, 5 e 6, que possuem caracterstica de simetria, em que a mdia est em torno de 0,5, os GC baseados nos resduos 2, 3 e 4 do modelo β ARMA possuem um melhor desempenho detectando mais rapidamente o estado de fora de controle. Apenas o GC baseado no resduo 1 tem um comportamento mais po-

bre que o baseado no resíduo do modelo ARMA. O GC β ARMA Shewhart baseado no resíduo 4 obteve um bom comportamento geral, enquanto que os outros resíduos em alguns cenários apresentaram comportamento irregular, principalmente o baseado no ARMA, que apresentou um comportamento errático ($ARL_1 > ARL_0$) comprometendo o poder de detecção. A seguir são brevemente discutidos esses comportamentos atípicos.

Resultados que apresentam $ARL_1 > ARL_0$ são inesperados e podem fazer pouco sentido. Quando isso ocorre, a probabilidade de detecção de pontos fora dos limites de controle é maior quando o processo está sob controle do que quando está fora de controle. Para melhor entender o que acontece nessas situações, foi simulada uma série fora de controle com as características do Cenário 1 com uma magnitude de mudança $\delta = -0,2$. A Figura 4.2 mostra uma ocorrência dessa série fora de controle traçada juntamente com os limites dos gráficos baseados nos resíduos do modelo ARMA, no GC β ARMA com resíduo 1 e no GC β ARMA com resíduo 4. Para esses gráficos o ARL_0 é de 200, então seria esperado mais de 10 pontos fora dos limites de controle em cada gráfico, uma vez que há 2000 observações de um processo fora de controle. Percebe-se que para os GC baseados no resíduo ARMA e no resíduo 1, há bem menos que 10 pontos fora dos limites de controle, caracterizando ARL_1 maior do que o $ARL_0 = 200$. O comportamento atípico dos dois primeiros é justificado por uma característica da distribuição beta. Como pode ser visto na Equação (3.12), para um parâmetro de precisão φ fixo, a variância de Y diminui a medida que a média de Y também diminui. Ou seja, nos cenários em que $\delta = -0,2$, a média e a variância do processo fora de controle são menores que a média e a variância do processo sob controle. Com menor variância de Y , resíduos que consideram explicitamente a diferença entre y_t e $\hat{\mu}_t$ terão também menor variância, apesar de média diferente de zero, conforme Figura 4.2, ficando majoritariamente dentro dos limites sob controle. Por outro lado, os GC baseados no resíduo 4 não possuem esse comportamento atípico pelo fato de levar em conta a diferença das funções de log-verossimilhanças avaliadas em y_t e $\hat{\mu}_t$, conforme Equação (3.25), e não dependem diretamente da diferença de y_t e $\hat{\mu}_t$ como os demais resíduos.

Os resultados apresentados pela Figura 4.3 mostram o desempenho dos GC residuais CUSUM para os Cenários 1, 2, 3, 4, 5 e 6. Todos os GC propostos para o Cenário 1 com $\delta < 0$ apresentam melhor desempenho do que o GC baseado no resíduo do modelo ARMA. Os GC baseados nos resíduos 2, 3 e 4 apresentam ARL_1 de 29, 34 e 38, para $\delta = -0,2$, respectivamente. Para $\delta > 0$ os GC propostos baseados nos resíduos 1, 3 e 4 possuem um bom desempenho, similar ao GC baseado no resíduo ARMA. Nota-se que conforme a dispersão diminui o poder

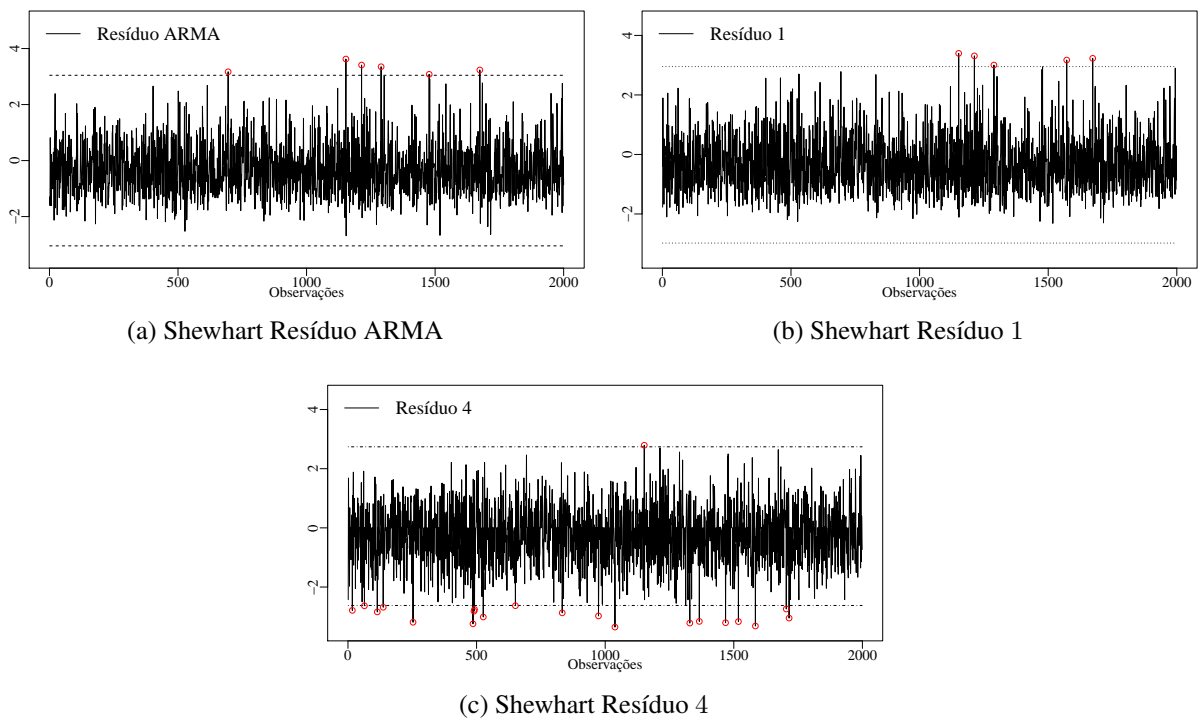


Figura 4.2: Gráficos de controle residuais de Shewhart para uma ocorrência do Cenário 1 fora de controle.

de detecção aumenta. Para o Cenário 2, para $\delta = -0,2$, o GC baseado no resíduo ARMA possui um ARL_1 de 19 e os baseados nos resíduos β ARMA 1, 2, 3 e 4, respectivamente, 20, 16, 17 e 18. Para o $\delta > 0$ o GC baseado no resíduo ARMA e os GC propostos possuem desempenho semelhantes. Para o Cenário 3, para $\delta = -0,2$, o GC baseado no resíduo ARMA possui um ARL_1 de 11 e os GC baseados nos resíduos β ARMA 1, 2, 3 e 4, respectivamente, 12, 10, 11 e 12. Para os Cenários 4, 5 e 6, os GC residuais CUSUM de todos os modelos possuem um comportamento regular e percebe-se a estabilidade destes na detecção dos pontos de fora de controle tendo um rápido decaimento. Para esses gráficos, no Cenário 4, os ARL_1 para todos os resíduos, com $\delta = -0,2$ e $\delta = 0,2$, foram aproximadamente 22. Já para o Cenário 5 os ARL_1 foram aproximadamente 12; e para o Cenário 6, os ARL_1 foram aproximadamente de 8 observações. O GC residual CUSUM baseado no resíduo 4 do modelo β ARMA obteve um bom comportamento geral. Percebe-se que o GC CUSUM teve bom desempenho na sensibilidade para detecção das diferentes magnitudes de mudança.

Os resultados apresentados pela Figura 4.4 mostram o desempenho dos GC residuais EWMA. Para os Cenários 1, 2 e 3, que possuem característica de assimetria positiva, estes apresentam um comportamento semelhante ao do GC residual de Shewhart, porém o gráfico é mais sensível às mudanças, ou seja, detecta pontos fora de controle mais rapidamente. Percebe-

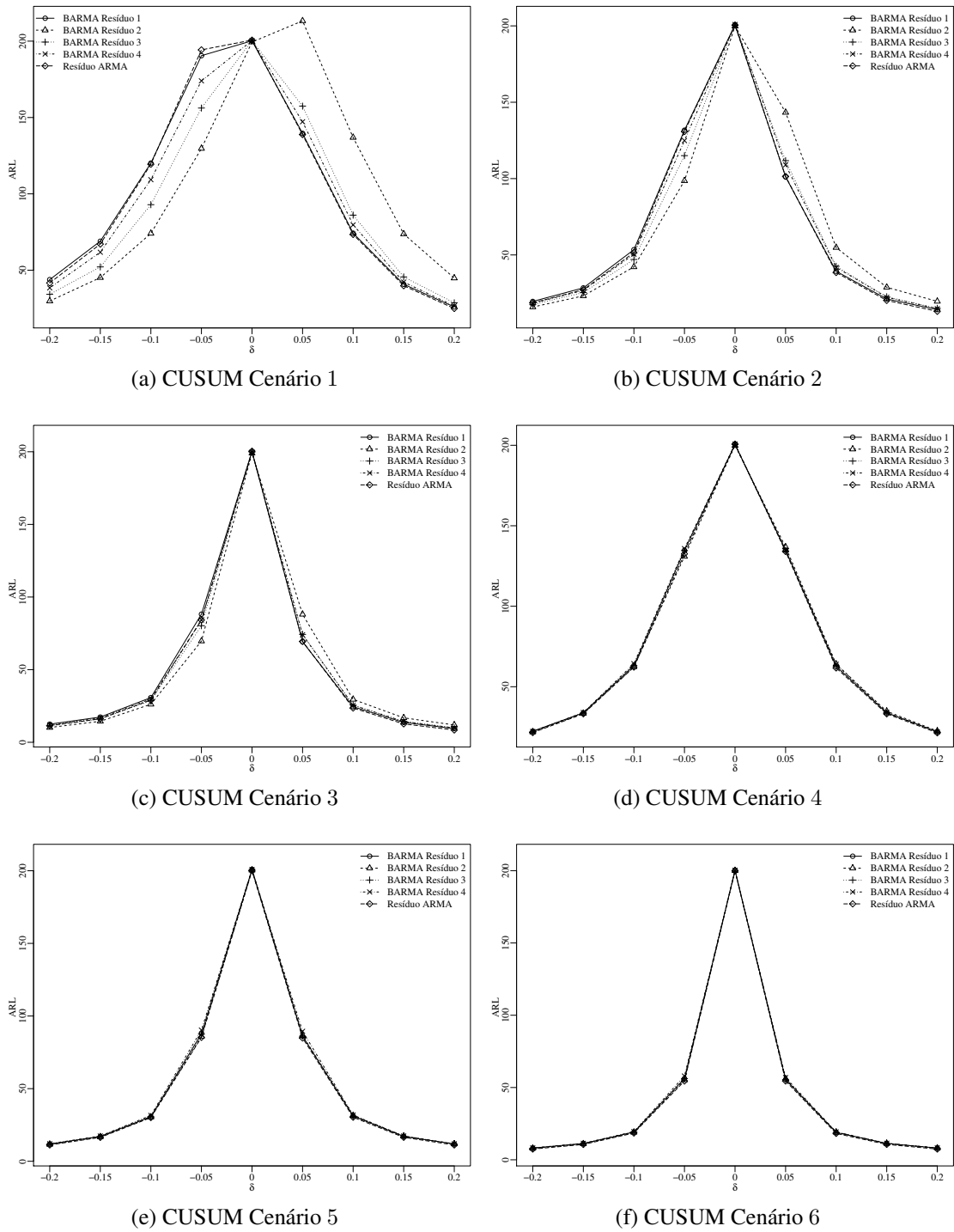


Figura 4.3: ARL_1 para os gráficos de controle residuais CUSUM.

se também bom comportamento geral de detecção dos gráficos baseados nos resíduos 3 e 4 do β ARMA. Para o Cenário 1, com $\delta = -0,2$, os gráficos baseados nos resíduos β ARMA 2, 3 e 4 possuem ARL_1 , respectivamente, de 24, 34 e 51, enquanto o resíduo ARMA possui ARL_1 de 290. Para o Cenário 2, com $\delta = -0,2$, o GC baseado no resíduo ARMA possui um ARL_1 de 27 e os GC baseados nos resíduos β ARMA 1, 2, 3 e 4, respectivamente, 23, 11, 16 e 17.

Para o Cenário 3, com $\delta = -0,2$, o GC baseado no resíduo ARMA possui ARL_1 de 10 e o GC baseado nos resíduos β ARMA 1, 2, 3 e 4, respectivamente, 10, 7, 9 e 9. Para $\delta > 0$, nos Cenários 1, 2 e 3, os GC residuais possuem um desempenho semelhante, com exceção do resíduo 2 que apresenta comportamento errático. Para os Cenários 4, 5 e 6, os resíduos de todos

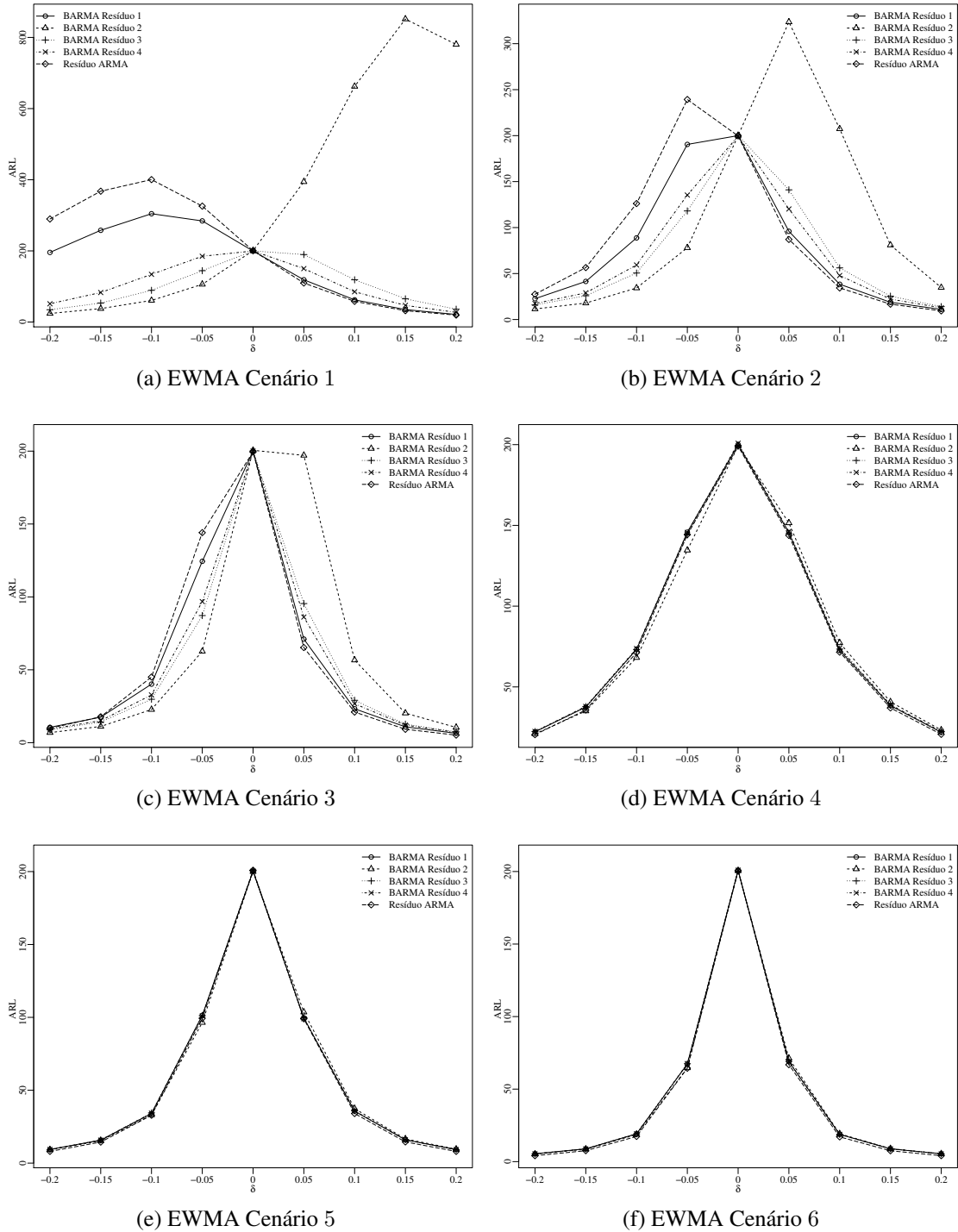


Figura 4.4: ARL_1 para os gráficos de controle residuais EWMA.

os modelos possuem um comportamento regular e percebe-se a estabilidade destes na detecção

dos pontos de fora de controle. Para estes GC residuais EWMA, para $\delta = -0,2$ e $\delta = 0,2$, no Cenário 4, os ARL_1 para todos os resíduos foram aproximadamente 22. Para o Cenário 5, os ARL_1 foram de aproximadamente 12; e para o Cenário 6, os ARL_1 foram aproximadamente de 5 observações. O GC residual EWMA baseado no resíduo 4 do modelo β ARMA obteve um bom comportamento geral, mesmo nos cenários onde a assimetria está presente.

Como muitos processos podem apresentar autocorrelação entre as observações, o uso de GC para verificar a estabilidade do processo fica condicionado a aplicação de modelos matemáticos para remover a autocorrelação, com a utilização da monitoração através dos resíduos dos modelos. Através dos resultados encontrados nas simulações desenvolvidas neste trabalho, sugere-se a utilização do GC residual CUSUM, pois este possui o melhor desempenho e melhor sensibilidade para a detecção de pontos fora de controle, corroborando o encontrado em Montgomery (2009). Indica-se o uso da modelagem β ARMA para os dados restritos ao intervalo $(0,1)$, principalmente em situações de fortes assimetrias, positivas ou negativas. Dentre as propostas consideradas destacou-se o GC baseado no resíduo 4 do modelo β ARMA, ou também chamado de resíduo *deviance*. Este resíduo apresenta melhor desempenho pelo fato de utilizar as contribuições para a função de log-verossimilhança das observações y_t e das estimativas $\hat{\mu}_t$, ou seja, da função de log-verossimilhança observada e ajustada para cada observação e não apenas a variância de y_t . Representa portanto a distância entre a observação y_t e o valor ajustado μ_t . A partir do exame do resíduo 4 pode-se observar a presença de resíduos não usuais (demasiadamente grandes ou pequenos), bem como a presença de *outliers* ou, ainda, padrões sistemáticos de variação indicando, possivelmente, o estado de fora de controle com maior sensibilidade. Esses GC mostraram bons resultados gerais e, em nenhum cenário, apresentou ARL_1 maior que o ARL_0 . Gráficos de controle que apresentam $ARL_1 > ARL_0$ podem não ser confiáveis, perdendo poder de detecção de pontos fora de controle, podendo levar a conclusões errôneas sobre a estabilidade do processo.

5 APLICAÇÕES

Nesta seção serão apresentadas duas aplicações em dados reais para demonstrar a utilidade prática dos métodos propostos. Serão considerados os GC β ARMA Shewhart, GC β ARMA CUSUM e GC β ARMA EWMA baseados no resíduo 4. Comparativamente também serão considerados o GC beta (SANT'ANNA; TEN CATEN, 2012) e o GC baseado no resíduo do modelo ARMA. Para a confecção dos gráficos em ambas aplicações utilizou-se $\lambda = 0,2$ e $L = 2,36$ no EWMA, $K = 0,5$ e $H = 3,51$ para o CUSUM, $k = 2,578$ para o Shewhart e $\alpha = 0.01$ para o gráfico de controle beta equiparando todos os gráficos em um $ARL_0 = 100$.

5.1 Percentual de energia armazenada

A primeira aplicação foi desenvolvida utilizando-se observações da série do percentual mensal do volume de energia armazenada (EA) na Região Sul do Brasil, dados publicados pelo Operador Nacional do Sistema Elétrico (ONS, 2014a). Os dados contemplam uma série temporal com 220 observações entre janeiro de 1996 e abril de 2014. Estes mesmos dados foram considerados em Palm, Bayer e Bayer (2015).

A EA é uma variável restrita ao intervalo $(0,1)$ importante para o setor elétrico brasileiro, sendo monitorada diariamente pelo Operador Nacional do Sistema Elétrico (ONS) através das energias naturais afluentes e das energias armazenadas. Essas informações são de grande relevância para a realização de previsões e para o planejamento operacional do Sistema Interligado Nacional (SIN) (ONS, 2014b).

Palm, Bayer e Bayer (2015) discutem que a obtenção de previsões de variáveis como o percentual de energia armazenada, que está diretamente relacionado à fonte hidrelétrica de produção de energia, se faz importante para auxiliar a gestão e a tomada de decisão pelo ONS no SIN e utilizam a modelagem β ARMA para as previsões. Pereira Filho (2015) afirma que cerca de 70% da energia elétrica do Brasil provém da geração por meio de hidroelétricas, que sofre drástica influência da estiagem, sugerindo ao setor elétrico brasileiro investir mais em sistemas de monitoramento e previsão para haver melhor planejamento e gestão dos recursos hídricos. Nesse contexto, utilizou-se gráficos de controle para monitorar os volumes de energia armazenada.

O gráfico da série de observações, o histograma, a FAC e a FACP amostrais podem ser vistas na Figura 5.1. Uma análise descritiva preliminar da série foi feita e mostra que a média não condicional das observações é igual a 0,7185, o mínimo e o máximo são, respectivamente, 0,2060 e 0,9889 com um desvio padrão de 0,1980. A série possui, portanto, valores muito próximos ao extremo, ou seja, muito próximos de 1, e possui grande dispersão.

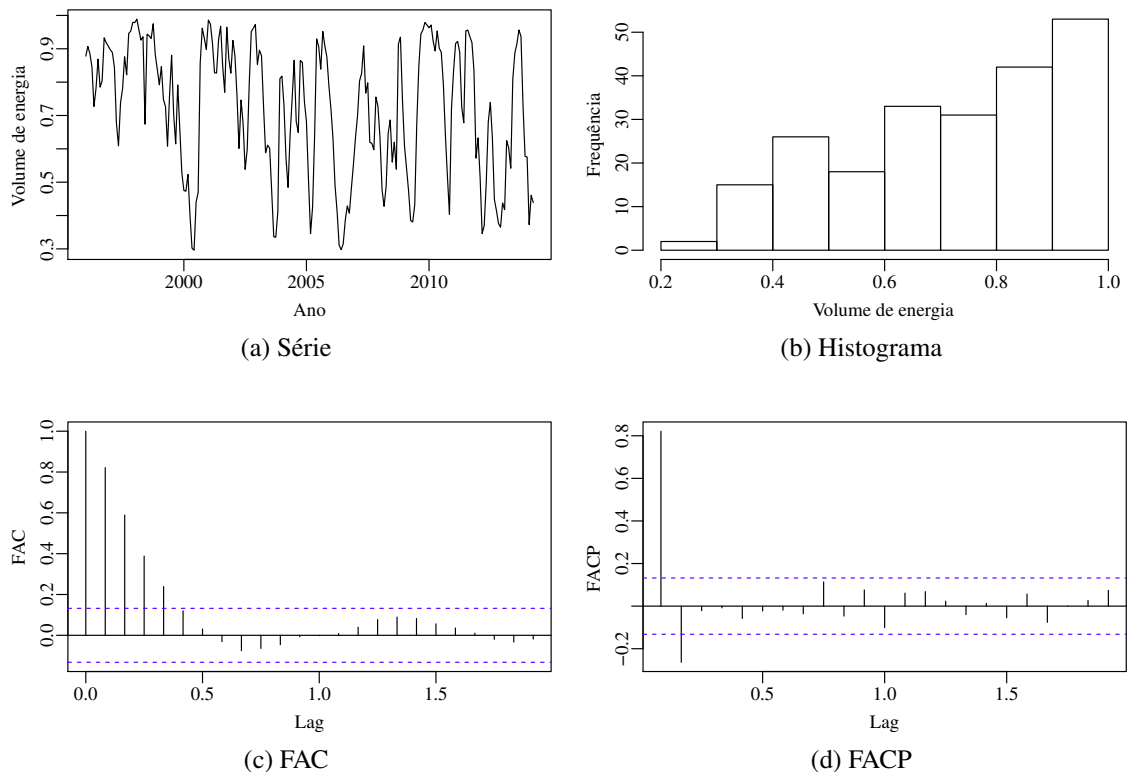


Figura 5.1: Gráficos da série do volume de EA.

Primeiramente confeccionou-se o gráfico de controle beta (SANT'ANNA; TEN CATTEN, 2012) para os dados da EA, apresentado na Figura 5.2. O gráfico de controle beta seria uma das alternativas apresentadas na literatura para dados contínuos restritos ao intervalo $(0,1)$, porém este não leva em consideração a dependência temporal entre as observações. Isso pode fazer com que as causas especiais não sejam detectadas. A Figura 5.2 mostra que os limites de controle são alargados e não detectam nenhum ponto fora de controle. A partir disto, ajustou-se os modelos $ARMA(1,1)$ e $\beta ARMA(1,1)$, mesmos modelos utilizados em Palm, Bayer e Bayer (2015), para considerar GC residuais. O ajuste pelo modelo de suavização exponencial simples não foi considerado pelo seu desempenho pobre nas simulações. O modelo $ARMA$ ajustado é apresentado na Tabela 5.1 e o $\beta ARMA$ na Tabela 5.2.

Os resultados dos testes de Ljung-Box e Box-Pierce não rejeitaram a hipótese nula dos

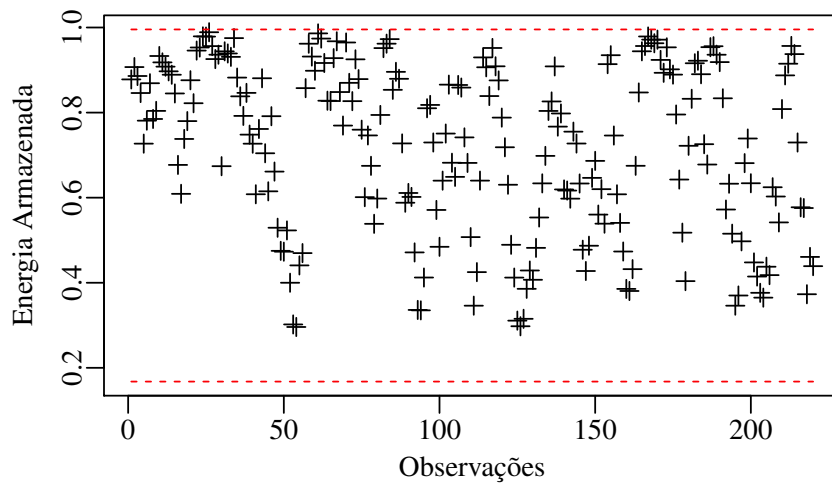


Figura 5.2: Gráfico de controle beta para os dados da EA.

Tabela 5.1: Modelo ARMA(1,1) ajustado para os dados da EA.

	Estimativa	Erro-Padrão
Intercepto	0,7165	0,0360
ϕ_1	0,7442	0,0519
θ_1	0,2880	0,0705
σ^2	0,01156	
AIC	-347,57	
Box-Pierce		p -valor = 0,8036
Ljung-Box		p -valor = 0,7556

Tabela 5.2: Modelo β ARMA(1,1) ajustado para os dados da EA.

	Estimativa	Erro-Padrão
α	0,2967	0,0322
ϕ_1	0,5961	0,0498
θ_1	0,2294	0,0536
φ	13,8705	1,3083
AIC	-383,5783	
		Resíduo 4
Box-Pierce		p -valor = 0,8169
Ljung-Box		p -valor = 0,7709

resíduos serem *i.i.d.*, ou seja, os resíduos não possuem autocorrelação significativa em nenhuma das modelagens. A FAC e a FACP dos resíduos dos modelos ajustados podem ser observadas nas Figuras 5.3 e 5.4. A partir destes resultados há evidências de que os modelos estão bem

ajustados e de que os gráficos de controle residuais podem ser confeccionados.

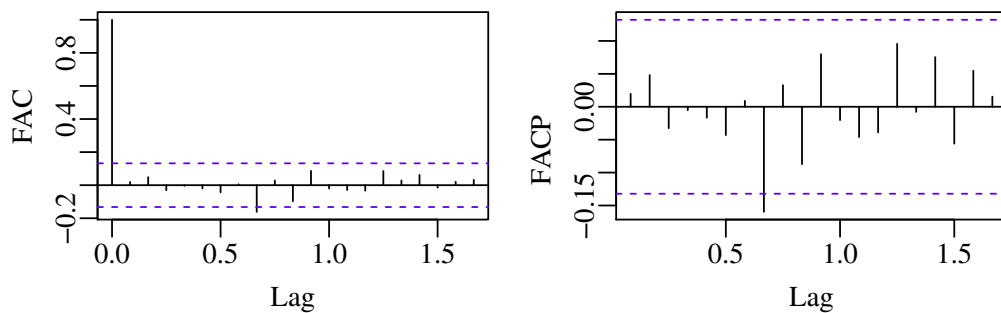


Figura 5.3: FAC e FACP dos resíduos do modelo ARMA(1,1) ajustado aos dados de EA.

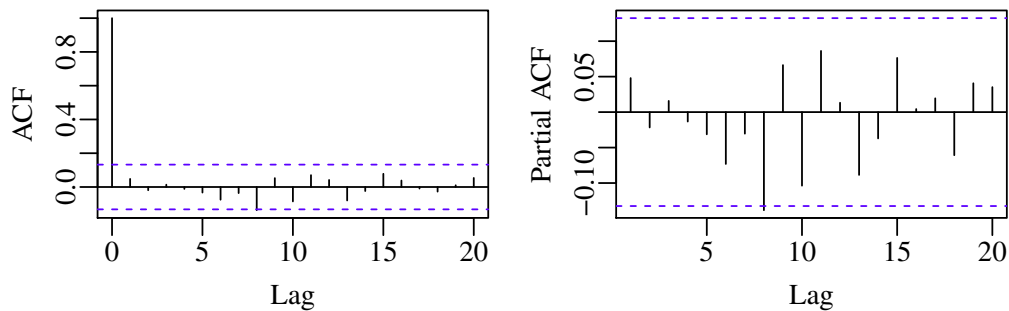
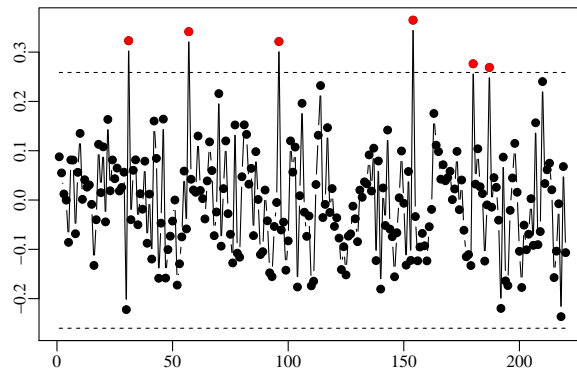


Figura 5.4: FAC e FACP do resíduo 4 do modelo β ARMA(1,1) ajustado aos dados da EA.

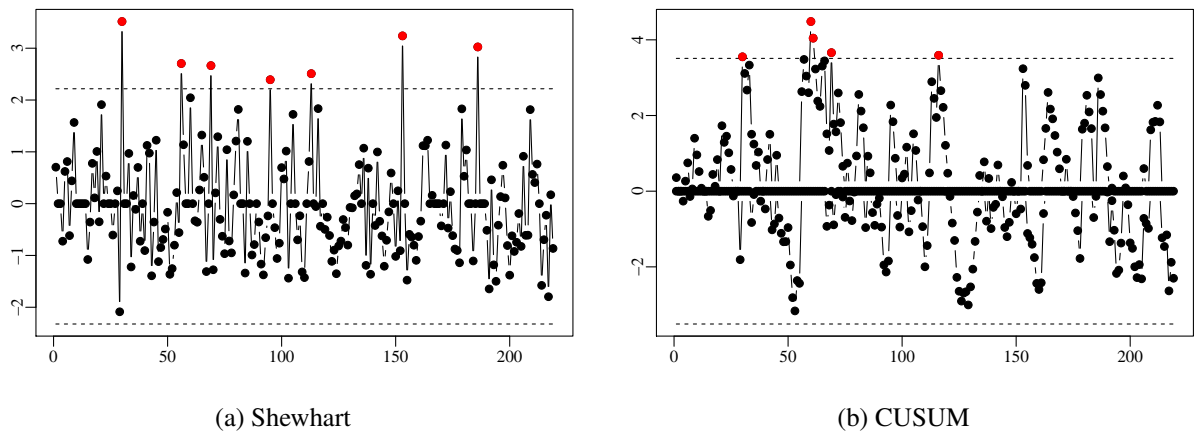
Em ambas as modelagens o gráfico de controle residual EWMA não foi sensível a detecção de pontos fora de controle e também o CUSUM residual ARMA, por esse motivo não serão apresentados. No gráfico utilizando os resíduos do modelo ARMA, apresentado na Figura 5.5, pode-se notar que o gráfico de controle residual de Shewhart foi sensível marcando alguns pontos de descontrole. Contudo, há uma possível inadequação no uso desta modelagem, pois os dados possuem distribuição assimétrica. Os pontos detectados no gráfico de Shewhart residual ARMA foram 31, 57, 96, 154, 180 e 187, que correspondem aos períodos de julho de 1998, setembro de 2000, dezembro de 2003, outubro de 2008, dezembro de 2010 e julho de 2011.

A Figura 5.6 apresenta os GC propostos quando aplicados ao monitoramento da EA. Ao analisar a série foi visto que seu valor de máximo era aproximado de 1, ou seja, a série possui valores perto do extremo superior. Nesses casos, os resultados das simulações evidenciaram melhores resultados do GC β ARMA. O gráfico de controle residual de Shewhart utilizando o resíduo 4 apontou que os pontos 31, 57, 70, 96, 114, 154 e 187, correspondentes a julho de 1998, setembro de 2000, outubro de 2001, dezembro de 2003, junho de 2005, outubro de 2008 e julho



(a) Shewhart

Figura 5.5: Gráfico de controle com resíduos do modelo $ARMA(1,1)$ para a série de EA.



(a) Shewhart

(b) CUSUM

Figura 5.6: Gráficos de controle com o resíduo 4 do modelo $\beta ARMA(1,1)$ para a série de EA.

de 2011, respectivamente, estão fora de controle. No gráfico de controle residual CUSUM os pontos assinalados foram 31, 61, 62, 70 e 117, que correspondem aos meses de julho de 1998, janeiro e fevereiro de 2001 e setembro de 2005.

Alguns desses pontos identificados pelos GC $\beta ARMA$ podem ser identificados por períodos de ocorrências dos fenômenos El Niño e La Niña. No ano de 1998 houve o fenômeno El Niño de intensidade forte e em 2003 com intensidade moderada. Nos anos de 2008 ocorreu o La Niña com intensidade forte. Em 2000 e 2001 aconteceu o La Niña com intensidade moderada, sendo alguns dos pontos detectados pelo gráfico residual $\beta ARMA$. O El Niño é caracterizado através do aumento da temperatura da superfície do mar (TSM) do Oceano Pacífico Equatorial e para a região Sul do Brasil o regime de chuvas tende a valores acima da média climatológica entre os meses de novembro a fevereiro o que interfere na armazenagem dos reservatórios (OLIVEIRA, 1999). A La Niña é caracterizada pela diminuição da TSM, através

do resfriamento das águas do Oceano Pacífico Equatorial e para a região Sul do Brasil as precipitações tendem a ser abaixo da média entre os meses de junho a dezembro (OLIVEIRA, 1999). As datas de ocorrência e a intensidade do El Niño e La Niña podem ser visualizados em <http://enos.cptec.inpe.br/>. Percebe-se que, como evidenciado nas simulações, o resíduo 4 teve um bom poder de detecção de causas especiais em dados assimétricos.

A utilização de gráficos de controle como ferramenta aplicada ao controle e monitoramento dos volumes da energia armazenada pode ser muito útil, pois uma vez produzida, a energia elétrica deve ser consumida assim que gerada. Como a demanda sofre alterações ao longo do ano, da semana e mesmo no decorrer de um dia, manter o equilíbrio entre a geração e o consumo é importante ao SIN para o controle da geração de energia.

5.2 Sistema Cantareira

Para o desenvolvimento desta aplicação foi utilizada a série histórica dos níveis dos volumes armazenados dos mananciais do Sistema Cantareira em observações mensais, com um total de 151 observações no período de janeiro de 2003 a julho de 2015. Trata-se de um dos principais mananciais da cidade de São Paulo e um dos maiores do mundo, sendo utilizado para abastecer 8,8 milhões de clientes da Companhia de Água de São Paulo (SABESP), responsável pelo abastecimento de metade da Região Metropolitana e de municípios vizinhos como Barueri, Caieiras, Carapicuíba, Franco da Rocha, Francisco Morato, Guarulhos, Osasco, Santo André, São Caetano do Sul e Taboão da Serra (CÔRTEZ et al., 2015). Entre os recursos naturais disponíveis na terra, a água é provavelmente o mais importante e valioso, especialmente quando se trata de água doce e água adequada para o consumo humano (SANCHO et al., 2015).

Os dados para a série temporal do volume de água armazenada no reservatório principal do sistema do Estado de São Paulo (Cantareira), foram obtidos a partir da base de dados da SABESP. Os dados estão disponíveis e podem ser acessados em <http://www2.sabesp.com.br/mananciais/DivulgacaoSiteSabesp.aspx>. Primeiramente, foi feita a análise descritiva dos dados que compõem a série. A média não condicional dos níveis dos mananciais foi de 0,5304, com mínimo e máximo, respectivamente, de 0,0606 e 0,9966, com desvio padrão de 0,2429. Por meio desta análise preliminar pode ser visto um grande nível de dispersão na série em estudo.

O reservatório da Cantareira enfrentou uma crise no fornecimento de água como o ocor-

rido no verão 2013-2014, com uma seca prolongada e altas temperaturas (CÔRTEZ et al., 2015). Esta situação excepcional pode ser vista a partir do decaimento do volume de água armazenado do sistema Cantareira, na Figura 5.7. Pode-se observar no gráfico da série os períodos de cheia que ocorreram no ano de 2010, onde as represas chegaram perto do nível máximo de armazenamento, e o período de seca extrema ocorrido no ano de 2014. O histograma e as FAC e FACP da série também são observadas na Figura 5.7.

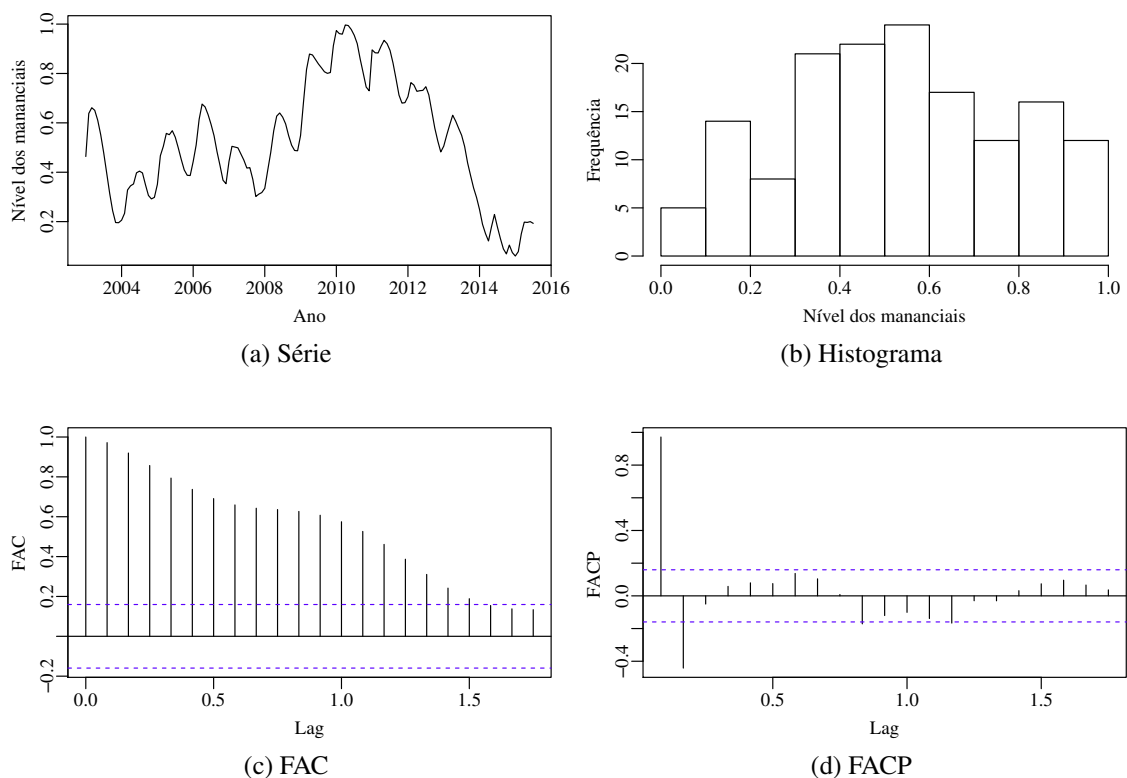


Figura 5.7: Gráficos da série do volume dos mananciais do sistema Cantareira..

O volume de água diminuiu acentuadamente a partir de meados de 2013 e a capacidade operacional do reservatório foi praticamente esvaziada em julho de 2014. Desde então a retirada de água é feita por bombeamento da chamada "reserva estratégica", ou "volume morto". A SABESP começou a reduzir as retiradas em janeiro de 2014, e até maio de 2015 o fluxo total foi de 40% dos valores médios (COUTINHO; KRAENKEL; PRADO, 2015). Novamente, foi confeccionado o gráfico de controle beta para os dados dos níveis dos mananciais do Sistema Cantareira, conforme Figura 5.8. O gráfico de controle beta não é satisfatório para o estudo da série, detectando apenas dois pontos de descontrole e não foi capaz de detectar o período crítico de abastecimento de água no ano de 2014. Isso deve acontecer pelo fato do GC beta não levar em consideração a dependência temporal das observações, que estão evidentes nas

Figuras 5.7 (c) e (d). Com isso, com o objetivo de considerar GC residuais, após exaustivos ajustes e testes, foram selecionados os modelos ARIMA(5,1,5) e o modelo β ARMA dado por $\eta_t = \alpha + \varphi_1 g(y_{t-1}) + \varphi_9 g(y_{t-9}) + \varphi_{10} g(y_{t-10}) + \varphi_{11} g(y_{t-11}) + \varphi_{18} g(y_{t-18}) + \theta_1 r_{t-1} + \theta_{12} r_{t-12}$. O ajuste exponencial simples também não foi considerado nesta aplicação pelo seu desempenho pobre nas simulações. Os modelos ARMA e β ARMA ajustados à série de observações são representados nas Tabelas 5.3 e 5.4, respectivamente.

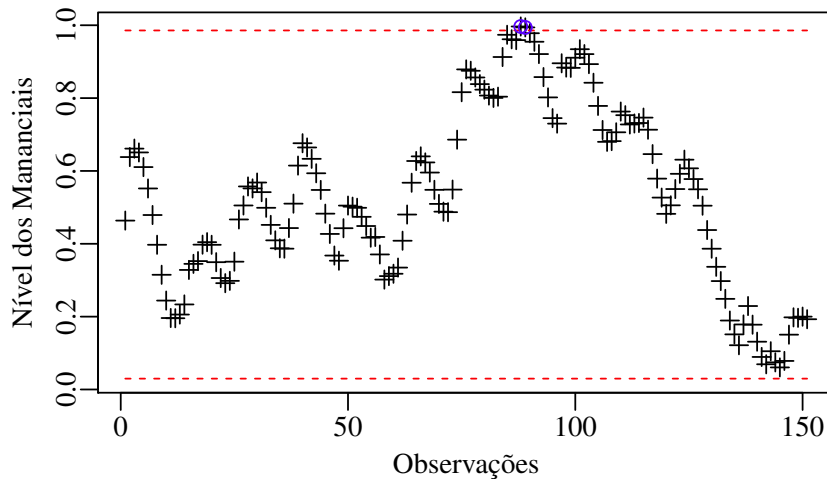


Figura 5.8: Gráfico de controle beta níveis dos mananciais do Sistema Cantareira

Tabela 5.3: Modelo ARIMA para o volume dos mananciais do Sistema Cantareira.

	Estimativa	Erro-Padrão
ϕ_1	0.5226	0.1661
ϕ_2	0.3027	0.0781
ϕ_3	0.3084	0.0589
ϕ_4	-1.0186	0.0728
ϕ_5	0.1250	0.1642
θ_1	-0.0094	0.1553
θ_2	-0.5042	0.0668
θ_3	-0.5792	0.0609
θ_4	0.8067	0.0823
θ_5	0.4348	0.1546
σ^2	0.0011	
AIC	-569.43	
Box-Pierce		p -valor = 0.7241
Ljung-Box		p -valor = 0.6296

Para testar os resíduos foram realizados os testes de Ljung-Box e Box Pierce, os quais

Tabela 5.4: Modelo β ARMA 5.2 para o volume dos mananciais do Sistema Cantareira.

	Estimativa	Erro-Padrão
α	0,0227	0,0102
φ_1	0,8908	0,0138
φ_9	0,2333	0,0445
φ_{10}	-0,1118	0,0194
φ_{18}	-0,0905	0,0207
θ_1	0,3760	0,0445
θ_{12}	0,1879	0,0485
ϕ	66,7705	
AIC	-529,6873	
		Resíduo 4
Box-Pierce		p -valor = 0,1136
Ljung-Box		p -valor = 0,1069

não rejeitaram a hipótese nula dos resíduos serem *i.i.d.*, indicando que os gráficos de controle tradicionais podem ser confeccionados. As FAC e FACP dos resíduos dos modelos ajustados podem ser observadas pelas Figuras 5.9 e 5.10.

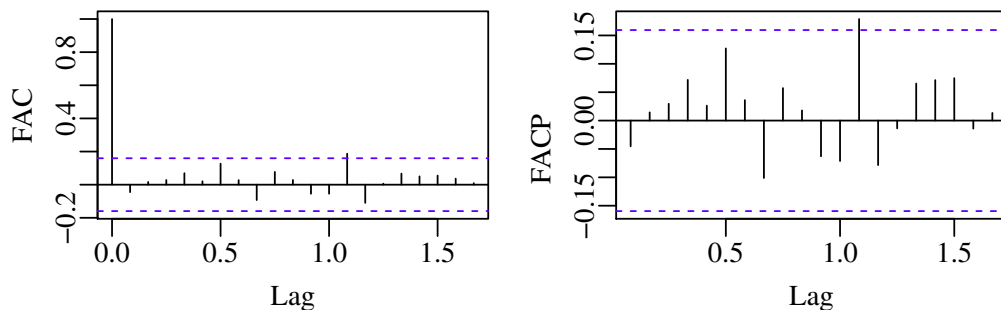


Figura 5.9: FAC e FACP dos resíduos do modelo ARIMA(5,1,5) para os níveis dos mananciais do Sistema Cantareira.

Os GC residuais do modelo ARMA podem ser vistos na Figura 5.11 e foram sensíveis para captar alguns pontos da série. O gráfico de controle residual de Shewhart assinalou os pontos 2, 84, 97, 98 que correspondem a fevereiro de 2003, dezembro de 2009, janeiro e fevereiro de 2011. O gráfico residual CUSUM apresentou os pontos 84, 85, 97, 135 e 136 os dois últimos correspondendo a março e abril de 2014 e o gráfico de controle residual EWMA apresentou os pontos 74, 75, 84, 85, 97, 134, 135 e 136 fora de controle, sendo os dois primeiros correspondentes a fevereiro e março de 2009. Os GC CUSUM e EWMA foram capazes de identificar somente março e abril de 2014. O período de 2014 deveria ser melhor identificado

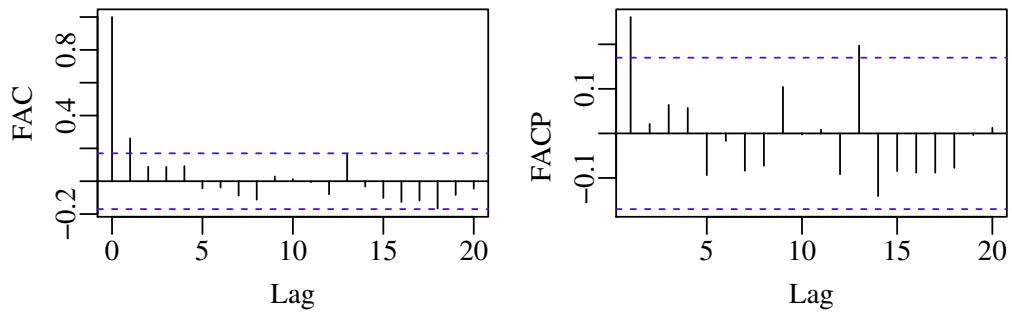


Figura 5.10: FAC e FACP do resíduo 4 do modelo β ARMA para os níveis dos mananciais do Sistema Cantareira.

fora de controle pelos GC, pois é evidente o problema de abastecimento na grande São Paulo.

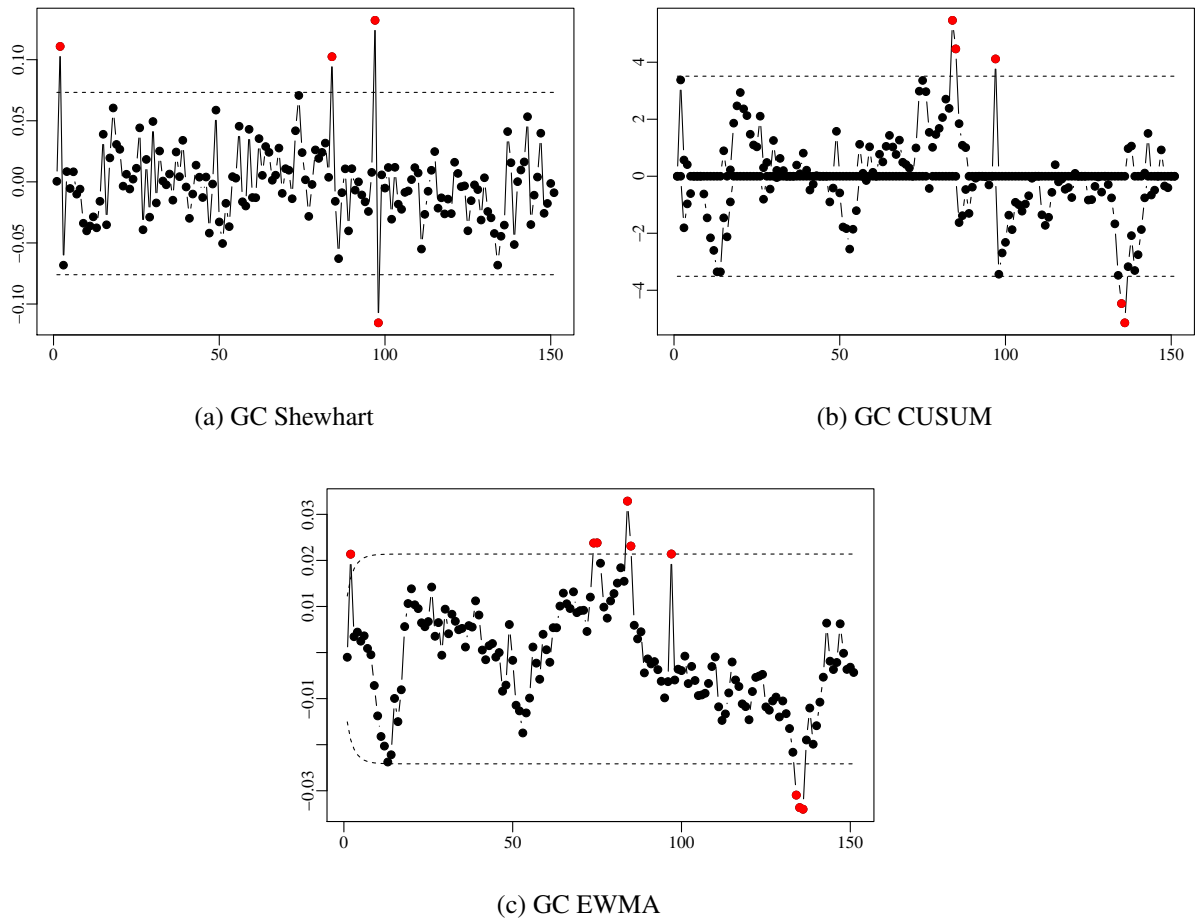


Figura 5.11: Gráficos de controle com resíduos do modelo ARIMA para o volume dos mananciais do Sistema Cantareira.

Os GC gerados com o resíduo 4 do modelo β ARMA obtiveram bons desempenhos na detecção de pontos fora de controle. O GC residual de β ARMA Shewhart assinalou os pontos 57, 66, 70, 75, 83 e 129, que correspondem a março e dezembro de 2009, abril e setembro

de 2010, maio de 2011 e março de 2014. O gráfico de controle residual β ARMA CUSUM apresentou os pontos 46, 47, 48, 49, 50, 57, 58, 59, 60, 61, 66, 67, 68, 69, 70, 71, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 116, 117, 118, 119, 121, 122, 123, 124 e 126 que correspondem de abril a agosto de 2008, março a julho de 2009, dezembro de 2009 a maio de 2010, julho de 2010 a setembro de 2011, fevereiro a maio de 2014, julho a outubro de 2014 e dezembro de 2014. O gráfico de controle residual β ARMA EWMA apresentou os pontos 8, 47, 57, 58, 66, 67, 70, 75, 76, 77, 78, 80, 115, 116, 117, 118 e 123 fora de controle que correspondem a fevereiro de 2005, março de 2008, março e abril de 2009, dezembro de 2009 a abril de 2010, outubro a dezembro de 2010, fevereiro e março de 2011, janeiro a abril de 2014 e agosto a outubro de 2014.

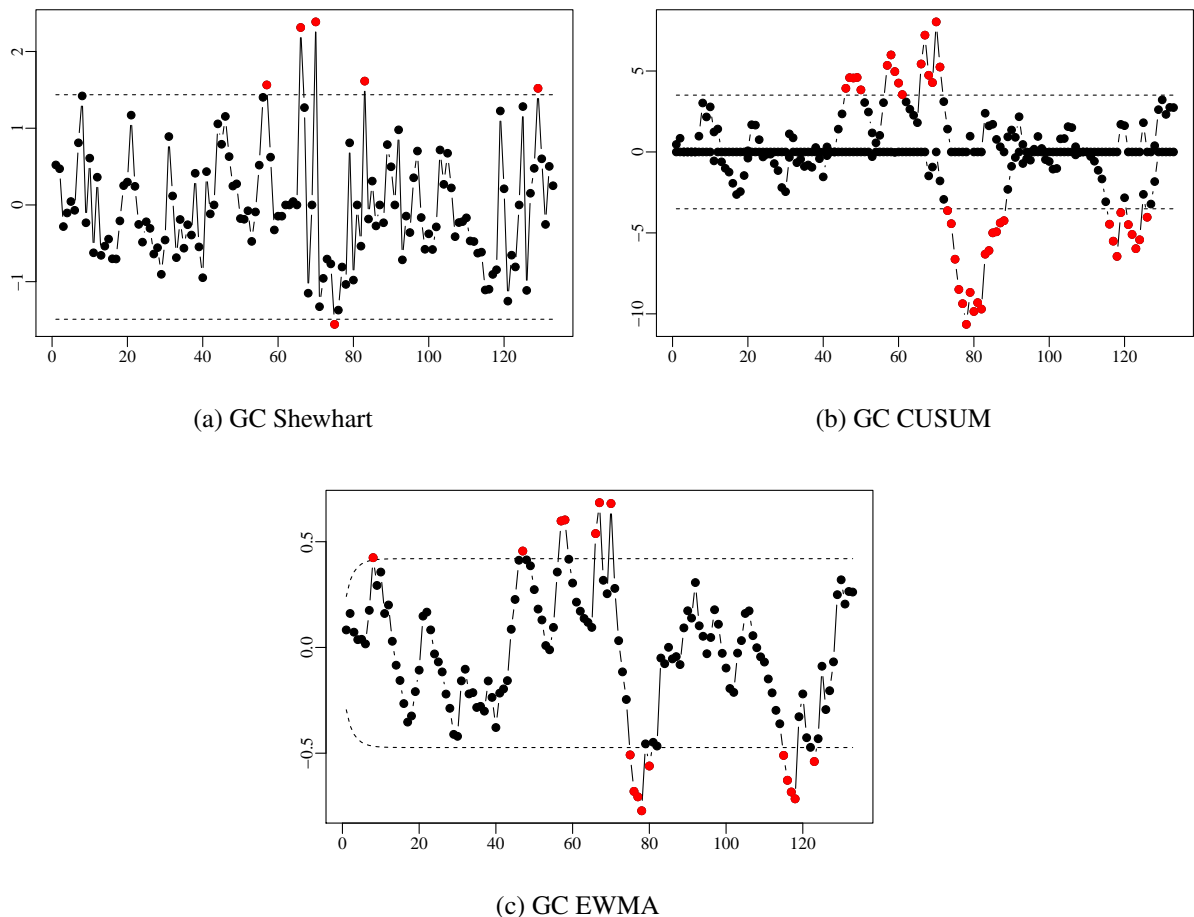


Figura 5.12: Gráficos de controle com resíduos 4 do modelo β ARMA para o volume dos mananciais do Sistema Cantareira.

Os GC gerados com os resíduos do modelo β ARMA obtiveram um bom e rápido desempenho na detecção de pontos aberrantes na série do Sistema Cantareira detectando, por exemplo, o período de 2010 que correspondeu a cheia do reservatório e a grande crise da água

no Sistema Cantareira que ocorreu no ano de 2014. O gráfico β ARMA CUSUM mostra o forte descontrole, detectando as quebras da série e os desvios de média que podem ser observados no gráfico da série. Ele detecta tanto períodos de seca como o período de cheia mais rapidamente do que outras alternativas da literatura. Desde o princípio de 2014 divulgam-se notícias relativas ao baixo nível pluviométrico em algumas das represas que abastecem o Sistema Cantareira. Ainda que a redução da disponibilidade de água seja relacionada aos períodos de estiagem e das temperaturas acima das normais para estas épocas do ano, ela é o reflexo da falta de planejamento estratégico que afeta o sistema de abastecimento da Região nos últimos dez anos (CÔRTEZ et al., 2015). Com a utilização dos gráficos de controle propostos seria possível detectar quando interferir adequadamente no processo e auxiliar nas tomadas de decisões sobre a utilização da água contida nos reservatórios. O uso de gráficos de controle pode ser uma ferramenta eficaz e muito importante para detectar as quebras na série e desenvolver planejamentos e políticas de economia e uso racional da água, para o enfrentamento de épocas de possível seca.

6 CONCLUSÕES

Nesta dissertação foi abordado o problema de monitorar dados do tipo taxas e proporções autocorrelacionados. Nestes casos, as abordagens tradicionais de confecção de gráficos de controle, que supõe a normalidade ou independência das observações, não podem ser utilizadas e é necessário o tratamento das observações para eliminar a autocorrelação. Com isto, o objetivo deste trabalho foi propor novos gráficos de controle para dados contínuos duplamente limitados autocorrelacionados. Utilizou-se o modelo de séries temporais β ARMA para o ajuste da série de dados e foram confeccionados os gráficos de controle de Shewhart, CUSUM e EWMA, aplicados à diferentes resíduos deste modelo.

Este trabalho tratou o problema de gráficos de controle para taxas e proporções, generalizando o gráfico de controle beta proposto por Sant'Anna e Ten Caten (2012) para o caso de dependência serial. Quando os dados são autocorrelacionados uma das técnicas tradicionalmente utilizadas é a modelagem ARIMA, modelando a autocorrelação e aplicando os gráficos de controle nos resíduos deste modelo. Porém, esta abordagem pode ser inadequada por supor a normalidade dos dados. Pela distribuição beta ser mais adequada para a modelagem deste tipo de dados, foi proposta a utilização da modelagem β ARMA e de seus resíduos.

Para avaliação dos GC propostos, foram simuladas diferentes séries temporais representativas de processos autocorrelacionados contemplando diferentes características possíveis do intervalo em que as observações são delimitadas. Os gráficos de controle residuais foram avaliados através de simulações de Monte Carlo pelo número médio de amostras até que um falso alarme ou uma causa especial fosse detectada (ARL_0 e ARL_1). Os resultados mostraram que para dados do tipo taxas e proporções restritos ao intervalo $(0,1)$, principalmente quando a série possui forte assimetria, os GC baseados nos resíduos do modelo β ARMA apresentam um melhor poder de detecção do que a modelagem usual, sendo o resíduo 4 o mais indicado.

Nas aplicações para a energia armazenada na Região Sul e para o nível dos mananciais do Sistema Cantareira se pode confirmar a aplicabilidade prática dos GC propostos. Para a energia armazenada, por ela precisar ser consumida assim que gerada e haver um sistema interligando este consumo em âmbito nacional, é importante ter um controle rigoroso da produção desta energia. Já para o Sistema Cantareira, a detecção e o controle dos períodos de cheia e de seca são importantes para a criação de políticas de economia e uso racional da água. Os GC propostos foram eficazes em identificar os pontos fora de controle das séries, principalmente

os períodos críticos do Sistema Cantareira, enquanto que as alternativas usuais da literatura tiveram um menor poder de detecção.

REFERÊNCIAS

- ABBAS, N.; RIAZ, M.; DOES, R. J. M. M. An EWMA-Type control chart for monitoring the process mean using auxiliary information. **Communications in Statistics - Theory and Methods**, v.43, n.16, p.3485–3498, 2014.
- ABBASI, S. A. On the performance of EWMA chart in presence of two component measurement error. **Quality Engineering**, v.v22 i3., p.199–213., 2010.
- ACOSTA-MEJIA, C. A.; PIGNATIELLO, J. J.; VENKATESHWARA RAO, B. A comparison of control charting procedures for monitoring process dispersion. **IIE Transactions**, v.31, n.6, p.569–579, 1999.
- AKAIKE, H. A new look at the statistical model identification. **IEEE Transactions on Automatic Control**, v.19, n.6, p.716–723, 1974.
- ALWAN, L. C. Autocorrelation: fixed versus variable control limits. **Quality Engineering**, v.4, n.2, p.167–188, 1991.
- ALWAN, L. C. Effects of autocorrelation on control chart performance. **Communications in Statistics-Theory and Methods**, v.21, n.4, p.1025–1049, 1992.
- 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.
- AMIRZADEH, V.; MASHINCHI, M.; PARCHAMI, A. Construction of p-charts using degree of nonconformity. **Information Sciences: an International Journal**, New York, NY, USA, v.179, n.1-2, p.150–160, Jan. 2009.
- BERGQUIST, B.; SÖDERHOLM, P. Data analysis for condition-based railway infrastructure maintenance. **Quality and Reliability Engineering International**, v.31, n.5, p.773–781, 2015.
- BOX, G. E. P.; JENKINS, G. M. **Time series analysis forecasting and control**. Holden-Day, 1976.
- BOX, G. E. P.; PIERCE, D. A. Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. **Journal of the American Statistical Association**, v.65, n.332, p.1509–1526, 1970.

- BOX, G.; JENKINS, G.; REINSEL, G. **Time series analysis: forecasting and control**. Wiley Series in Probability and Statistics, 2008.
- BOX, G.; LUCENO, A. **Statistical control by monitoring and feedback adjustment**. New York: Wiley & Sons., 1997.
- CHANG, Y.-M.; WU, T.-L. On average run lengths of control charts for autocorrelated processes. **Methodology and Computing in Applied Probability**, v.13, n.2, p.419–431, 2011.
- CHEN, H. et al. Symmetric X Charts: sensitivity to nonnormality and control-limit estimation. **Communications in Statistics - Simulation and Computation**, v.0, n.0, p.1–25, 2015.
- CHENG, S. W.; THAGA, K. Max-CUSUM chart for autocorrelated processes. **Statistica Sinica**, v.15, n.2, p.527–546, 2005.
- CLARO, F. A. E.; COSTA, A. F. B.; MACHADO, M. A. G. Gráficos de controle de EWMA e de X-barra para monitoramento de processos autocorrelacionados. **Revista Produção Online**, v.17, n.3, p.536–546, 2007.
- COLOSIMO, B. M.; PACELLA, M. A comparison study of control charts for statistical monitoring of functional data. **International Journal of Production Research**, v.48, n.6, p.1575–1601, 2010.
- CÔRTEZ, P. L. et al. Crise de abastecimento de água em São Paulo e falta de planejamento estratégico. **Estudos Avançados**, v.29, p.7 – 26, 08 2015.
- COUTINHO, R. M.; KRAENKEL, R. A.; PRADO, P. I. Catastrophic regime shift in water reservoirs and São Paulo water supply crisis. **PLoS ONE**, v.10, n.9, p.e0138278, 09 2015.
- CROUX, C.; GELPER, S.; MAHIEU, K. Robust control charts for time series data. **Expert Systems with Applications**, v.38, p.13810–13815, 2011.
- 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.
- FERRARI, S.; CRIBARI-NETO, F. Beta regression for modeling rates and proportions. **Journal of Applied Statistics**, v.31, n.7, 2004.

FILHO, A. J. P. Análise da escassez hídrica brasileira em 2014. **Revista USP, Brasil**, n.104, p.125–132, 2015.

GUERRA, R. R. **Avaliação de critérios de seleção no modelo BARMA**. 2015. Mestrado em Engenharia de Produção — Universidade Federal de Santa Maria.

GUJARATI, D. N. **Econometria Básica**. Pearson Education Brasil, 2000.

JURAN, J. **Quality control handbook**. McGraw-Hill, 1951.

KEDEM, B.; FOKIANOS, K. **Regression models for time series analysis**. Wiley, 2005. (Wiley Series in Probability and Statistics).

KIESCHNICK, R.; MCCULLOUGH, B. D. Regression analysis of variates observed on $(0, 1)$: percentages, proportions and fractions. **Statistical Modelling**, v.3, n.3, p.193–213, 2003.

KNOTH, S. **spc**: statistical process control – collection of some useful functions. 2016. R package version 0.5.3.

KRAMER, H.; SCHMID, W. Control charts for time series. **Nonlinear Analysis: Theory, Methods & Applications**, v.30, n.7, p.4007–4016, 1997.

LIU, L. et al. Adaptive phase II nonparametric EWMA control chart with variable sampling interval. **Quality and Reliability Engineering International**, v.31, n.1, p.15–26, 2015.

LIU, R. Y.; TANG, J. Control charts for dependent and independent measurements based on bootstrap methods. **Journal of the American Statistical Association**, v.91, n.436, p.1694–1700, 1996.

LJUNG, G. M.; BOX, G. E. P. On a measure of lack of fit in time series models. **Biometrika**, v.65, n.2, p.297–303, 1978.

LOREDO, E. N.; JEARKPAPORN, D.; BORROR, C. M. Model-based control chart for autoregressive and correlated data. **Quality and Reliability Engineering International**, v.18, n.6, p.489–496, 2002.

MAHMOUD, M. A.; WOODALL, W. H. Phase I analysis of linear profiles with calibration applications. **Technometrics**, v.46, n.4, p.380–391, 2004.

- MINGOTI, S. A.; FIDELIS, M. Aplicando geoestatística no controle estatístico de processos. **Revista Produto & Produção**, v.5, n.2, p.55–70, 2001.
- MINGOTI, S. A.; YASSUKAWA, F. R. Uma comparação de gráficos de controle para a média de processos autocorrelacionados. **Revista Eletrônica Sistemas & Gestão**, v.3, n.1, p.55–73, 2008.
- MONTGOMERY, D. C. **Introduction to Statistical Quality Control**. 6.ed. Wiley, 2009.
- MORAES, D. A. O. et al. Self-oriented control charts for efficient monitoring of mean vectors. **Computers & Industrial Engineering**, v.75, p.102 – 115, 2014.
- MORETTIN, P.; TOLOI, C. **Análise de séries temporais**. Edgar Blucher Ltda. São Paulo., 2004.
- MUKHERJEE, P. S. On phase II monitoring of the probability distributions of univariate continuous processes. **Statistical Papers**, p.1–24, 2015.
- NOCEDAL, J.; WRIGHT, S. **Numerical optimization**. Springer, New York, 1999.
- NOOROSSANA, R.; VAGHEFI, S. Effect of autocorrelation on performance of the MCUSUM control chart. **Quality and Reliability Engineering International**, v.22, n.2, p.191–197, 2006.
- OLIVEIRA, G. S. de. **O El niño e você o fenômeno climático**. Transtec Editorial, 1999.
- ONS. **Operador nacional do sistema elétrico - energia armazenada**. 2014a.
- ONS. **O que é o SIN - Sistema Interligado Nacional**. 2014b.
- PADONOU, E.; ROUSTANT, O.; LUTZ, M. Robust monitoring of an industrial IT System in the presence of structural change. **Quality and Reliability Engineering International**, v.31, n.6, p.949–962, 2015.
- PAGE, E. S. Continuous inspection schemes. **Biometrika**, v.41, n.1-2, p.100–115, 1954.
- PALM, B. G.; BAYER, F. M.; BAYER, D. M. Comparação de diferentes modelagens para percentual de energia armazenada. **Revista Eletrônica em Gestão, Educação e Tecnologia Ambiental**, v.19, n.3, p.1–12, 2015.

PSARAKIS, S.; VYNIYOU, A. K.; CASTAGLIOLA, P. Some recent developments on the effects of parameter estimation on control charts. **Quality and Reliability Engineering International**, v.30, n.8, p.1113–1129, 2014.

R Core Team. **R: a language and environment for statistical computing**. Vienna, Austria: R Foundation for Statistical Computing, 2014.

REYNOLDS JR, M. R.; LU, C.-W. Control charts for monitoring processes with autocorrelated data. **Nonlinear Analysis: Theory, Methods & Applications**, v.30, n.7, p.4059–4067, 1997.

ROBERTS, S. W. Control chart tests based on geometric moving averages. **Technometrics**, v.1, n.3, p.239–250, 1959.

ROCHA, A. V.; CRIBARI-NETO, F. Beta autoregressive moving average models. **Test**, v.18, n.3, p.529–545, 2009.

SANCHO, J. et al. Study of water quality in a Spanish river based on statistical process control and functional data analysis. **Mathematical Geosciences**, p.1–24, 2015.

SANT'ANNA, A. M. O.; TEN CATEN, C. S. Beta control charts for monitoring fraction data. **Expert Systems with Applications**, v.39, n.11, p.10236–10243, 2012.

SCHAFFER, J. R.; KIM, M.-J. Number of replications required in control chart Monte Carlo simulation studies. **Communications in Statistics - Simulation and Computation**, v.36, n.5, p.1075–1087, 2007.

SCHWARZ, G. Estimating the dimension of a model. **The Annals of Statistics**, v.6, n.2, p.461–464, 1978.

SCRUCCA, L. qcc: an R package for quality control charting and statistical process control. **R News**, v.4/1, p.11–17, 2004.

SHEU, S.-H. et al. Monitoring process mean and variability with generally weighted moving average control charts. **Computers & Industrial Engineering**, v.57, n.1, p.401 – 407, 2009.

SHEWHART, W. A. **Economic control of quality of manufactured product**. New York: Van-
Nostrand Reinhold, 1931.

SHUMWAY, R. H.; STOFFER, D. S. **Time series analysis and its applications** : with R examples. New York: Springer, 2006. (Springer texts in statistics).

TEOH, W. et al. Optimal design of the double sampling chart with estimated parameters based on median run length. **Computers & Industrial Engineering**, v.67, n.0, p.104 – 115, 2014.

VINING, G. Technical advice: phase I and phase II control charts. **Quality Engineering**, v.21, n.4, p.478–479, 2009.

WARDELL, D. G.; MOSKOWITZ, H.; PLANTE, R. D. Control charts in the presence of data correlation. **Management Science**, v.v. 38, n.n. 8, p.p. 1084–1100, 1992.

WARDELL, D. G.; MOSKOWITZ, H.; PLANTE, R. D. Run-Length distributions of special-cause control charts for correlated processes. **Technometrics**, v.36, n.1, p.3–17, 1994.

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

YU, J.; LIU, J. LRProb control chart based on logistic regression for monitoring mean shifts of auto-correlated manufacturing processes. **International Journal of Production Research**, v.49, n.8, p.2301–2326, 2011.

ZHANG, N. F. Detection capability of residual control chart for stationary process data. **Journal of Applied Statistics**, v.24, n.4, p.475–492, 1997.

ZHANG, N. F. A statistical control chart for stationary process data. **Technometrics**, v.40, n.1, p.24–38, 1998.

APÊNDICES

APÊNDICE A – Calibração

Neste apêndice encontra-se as tabelas que apresentam os ajustes de regressão linear feitos para facilitar o cálculo do ARL_0 através de um modelo.

Tabela A.1: Apresentação dos modelos de regressão lineares aos ARL's do cenário com $\phi = 40$ e $\mu = 0,2$.

Precisão	μ	GC	Ajuste	Modelo	R^2
40	0,2	Shewhart	β ARMA Res1	$\log(ARL) \times 0,4949 + 0,4413$	0,9994
			β ARMA Res2	$\log(ARL) \times 0,5708 + 0,1794$	0,9926
			β ARMA Res3	$\log(ARL) \times 0,4617 + 0,5387$	0,9976
			β ARMA Res4	$\log(ARL) \times 0,4119 + 0,6722$	0,9921
			ARIMA	$\log(ARL) \times 0,4923 + 0,4369$	0,9996
			EWMA	$\log(ARL) \times 0,4619 + 0,5190$	0,9976
			β ARMA Res1	$\log(ARL) \times 0,8040 - 0,8423$	0,9971
			β ARMA Res2	$\log(ARL) \times 0,9699 - 1,4326$	0,9854
			β ARMA Res3	$\log(ARL) \times 0,8044 - 0,8631$	0,9970
		β ARMA Res4	$\log(ARL) \times 0,7827 - 0,8236$	0,9939	
		ARIMA	$\log(ARL) \times 0,8112 - 0,8917$	0,9969	
		EWMA	$\sqrt{ARL} \times 0,0607 + 1,1447$	0,9671	
		EWMA	β ARMA Res1	$\log(ARL) \times 0,4642 + 0,4357$	0,9969
			β ARMA Res2	$\sqrt{ARL} \times 0,1811 + 0,8851$	0,9837
			β ARMA Res3	$\log(ARL) \times 0,4923 + 0,3806$	0,9988
			β ARMA Res4	$\log(ARL) \times 0,4596 + 0,4519$	0,9970
			ARIMA	$\log(ARL) \times 0,4929 + 0,3432$	0,9977
			EWMA	$\sqrt{ARL} \times 0,0613 + 1,3218$	0,9458

Tabela A.2: Apresentação dos modelos de regressão lineares aos ARL's do cenário com $\phi = 100$ e $\mu = 0,2$.

Precisão	μ	GC	Ajuste	Modelo	R^2
100	0,2	Shewhart	β ARMA Res1	$\log(\text{ARL}) \times 0,4470 + 0,5646$	0,9974
			β ARMA Res2	$\log(\text{ARL}) \times 0,4646 + 0,5211$	0,9980
			β ARMA Res3	$\log(\text{ARL}) \times 0,4286 + 0,6201$	0,9952
			β ARMA Res4	$\log(\text{ARL}) \times 0,4096 + 0,6700$	0,9922
			ARIMA	$\log(\text{ARL}) \times 0,4473 + 0,5497$	0,9976
			EWMA	$\log(\text{ARL}) \times 0,4315 + 0,5899$	0,9959
		CUSUM	β ARMA Res1	$\log(\text{ARL}) \times 0,8051 - 0,8449$	0,9961
			β ARMA Res2	$\log(\text{ARL}) \times 0,8678 - 1,0360$	0,9938
			β ARMA Res3	$\log(\text{ARL}) \times 0,8070 - 0,8601$	0,9964
			β ARMA Res4	$\log(\text{ARL}) \times 0,8107 - 0,8774$	0,9937
			ARIMA	$\log(\text{ARL}) \times 0,7912 - 0,8368$	0,9966
			EWMA	$\text{sqrt}(\text{ARL}) \times 0,0595 + 1,1447$	0,9664
		EWMA	β ARMA Res1	$\log(\text{ARL}) \times 0,4676 + 0,4426$	0,9978
			β ARMA Res2	$\log(\text{ARL}) \times 0,5419 + 0,2480$	0,9972
			β ARMA Res3	$\log(\text{ARL}) \times 0,4776 + 0,220$	0,9984
			β ARMA Res4	$\log(\text{ARL}) \times 0,4681 + 0,4432$	0,9978
			ARIMA	$\log(\text{ARL}) \times 0,4746 + 0,4006$	0,9981
			EWMA	$\text{sqrt}(\text{ARL}) \times 0,0575 + 0,4413$	0,9994

Tabela A.3: Apresentação dos modelos de regressão lineares aos ARL's do cenário com $\phi = 200$ e $\mu = 0,2$.

Precisão	μ	GC	Ajuste	Modelo	R^2
200	0,2	Shewhart	β ARMA Res1	$\log(\text{ARL}) \times 0,4283 + 0,6148$	0,9954
			β ARMA Res2	$\log(\text{ARL}) \times 0,4352 + 0,6003$	0,9957
			β ARMA Res3	$\log(\text{ARL}) \times 0,4186 + 0,6438$	0,9937
			β ARMA Res4	$\log(\text{ARL}) \times 0,4092 + 0,6675$	0,9923
			ARIMA	$\log(\text{ARL}) \times 0,4265 + 0,5985$	0,9953
			EWMA	$\log(\text{ARL}) \times 0,4188 + 0,6158$	0,9944
		CUSUM	β ARMA Res1	$\log(\text{ARL}) \times 0,8074 - 0,8628$	0,9964
			β ARMA Res2	$\log(\text{ARL}) \times 0,8432 - 0,9714$	0,9952
			β ARMA Res3	$\log(\text{ARL}) \times 0,8082 - 0,8718$	0,9964
			β ARMA Res4	$\log(\text{ARL}) \times 0,8194 - 0,9107$	0,9939
			ARIMA	$\log(\text{ARL}) \times 0,7769 - 0,8210$	0,9969
			EWMA	$\text{sqrt}(\text{ARL}) \times 0,0582 + 1,1343$	0,9666
EWMA	β ARMA Res1	$\log(\text{ARL}) \times 0,4681 + 0,4460$	0,9980		
	β ARMA Res2	$\log(\text{ARL}) \times 0,5068 + 0,3441$	0,9987		
	β ARMA Res3	$\log(\text{ARL}) \times 0,4735 + 0,4348$	0,9982		
	β ARMA Res4	$\log(\text{ARL}) \times 0,4691 + 0,4438$	0,9980		
	ARIMA	$\log(\text{ARL}) \times 0,4646 + 0,4191$	0,9976		
	EWMA	$\text{sqrt}(\text{ARL}) \times 0,0554 + 1,2936$	0,9537		

Tabela A.4: Apresentação dos modelos de regressão lineares aos ARL's do cenário com $\phi = 40$ e $\mu = 0,5$.

Precisão	μ	GC	Ajuste	Modelo	R^2
40	0,5	Shewhart	β ARMA Res1	$\log(\text{ARL}) \times 0,3871 + 0,7084$	0,9878
			β ARMA Res2	$\log(\text{ARL}) \times 0,4202 + 0,6397$	0,9943
			β ARMA Res3	$\log(\text{ARL}) \times 0,4173 + 0,6515$	0,9939
			β ARMA Res4	$\log(\text{ARL}) \times 0,4082 + 0,6700$	0,9927
			ARIMA	$\log(\text{ARL}) \times 0,3867 + 0,7134$	0,9878
			EWMA	$\log(\text{ARL}) \times 0,3983 + 0,6780$	0,9906
		CUSUM	β ARMA Res1	$\log(\text{ARL}) \times 0,7979 - 0,8191$	0,9967
			β ARMA Res2	$\log(\text{ARL}) \times 0,8558 - 0,9848$	0,9956
			β ARMA Res3	$\log(\text{ARL}) \times 0,8078 - 0,8677$	0,9965
			β ARMA Res4	$\log(\text{ARL}) \times 0,8079 - 0,8774$	0,9939
			ARIMA	$\log(\text{ARL}) \times 0,7994 - 0,8234$	0,9963
			EWMA	$\text{sqrt}(\text{ARL}) \times 0,0556 + 1,1194$	0,9598
		EWMA	β ARMA Res1	$\log(\text{ARL}) \times 0,4493 + 0,4973$	0,9970
			β ARMA Res2	$\log(\text{ARL}) \times 0,5061 + 0,3692$	0,9995
			β ARMA Res3	$\log(\text{ARL}) \times 0,4792 + 0,4223$	0,9987
			β ARMA Res4	$\log(\text{ARL}) \times 0,4653 + 0,4503$	0,9981
			ARIMA	$\log(\text{ARL}) \times 0,4486 + 0,4974$	0,9970
			EWMA	$\text{sqrt}(\text{ARL}) \times 0,0511 + 1,2556$	0,9464

Tabela A.5: Apresentação dos modelos de regressão lineares aos ARL's do cenário com $\phi = 100$ e $\mu = 0,5$.

Precisão	μ	GC	Ajuste	Modelo	R^2
100	0,5	Shewhart	β ARMA Res1	$\log(\text{ARL}) \times 0,3987 + 0,6864$	0,9902
			β ARMA Res2	$\log(\text{ARL}) \times 0,4124 + 0,6572$	0,9927
			β ARMA Res3	$\log(\text{ARL}) \times 0,4117 + 0,6608$	0,9927
			β ARMA Res4	$\log(\text{ARL}) \times 0,4082 + 0,6684$	0,9920
			ARIMA	$\log(\text{ARL}) \times 0,3990 + 0,6886$	0,9903
			EWMA	$\log(\text{ARL}) \times 0,4031 + 0,6673$	0,9915
		CUSUM	β ARMA Res1	$\log(\text{ARL}) \times 0,7898 - 0,8074$	0,9972
			β ARMA Res2	$\log(\text{ARL}) \times 0,8129 - 0,8749$	0,9969
			β ARMA Res3	$\log(\text{ARL}) \times 0,7931 - 0,8251$	0,9971
			β ARMA Res4	$\log(\text{ARL}) \times 0,8021 - 0,8569$	0,9949
			ARIMA	$\log(\text{ARL}) \times 0,8045 - 0,8459$	0,9968
			EWMA	$\text{sqrt}(\text{ARL}) \times 0,0570 + 1,1278$	0,9635
EWMA	β ARMA Res1	$\log(\text{ARL}) \times 0,4592 + 0,4665$	0,9973		
	β ARMA Res2	$\log(\text{ARL}) \times 0,4807 + 0,4181$	0,9985		
	β ARMA Res3	$\log(\text{ARL}) \times 0,4709 + 0,4362$	0,9981		
	β ARMA Res4	$\log(\text{ARL}) \times 0,4656 + 0,4476$	0,9978		
	ARIMA	$\log(\text{ARL}) \times 0,4626 + 0,4625$	0,9976		
	EWMA	$\text{sqrt}(\text{ARL}) \times 0,0522 + 1,2699$	0,9493		

Tabela A.6: Apresentação dos modelos de regressão lineares aos ARL's do cenário com $\phi = 200$ e $\mu = 0,5$.

Precisão	μ	GC	Ajuste	Modelo	R^2
200	0,5	Shewhart	β ARMA Res1	$\log(\text{ARL}) \times 0,4034 + 0,6745$	0,9916
			β ARMA Res2	$\log(\text{ARL}) \times 0,4105 + 0,6588$	0,9928
			β ARMA Res3	$\log(\text{ARL}) \times 0,4102 + 0,6609$	0,9928
			β ARMA Res4	$\log(\text{ARL}) \times 0,4084 + 0,6644$	0,9925
			ARIMA	$\log(\text{ARL}) \times 0,4043 + 0,6744$	0,9917
			EWMA	$\log(\text{ARL}) \times 0,4051 + 0,6594$	0,9921
		CUSUM	β ARMA Res1	$\log(\text{ARL}) \times 0,7869 - 0,8039$	0,9964
			β ARMA Res2	$\log(\text{ARL}) \times 0,7983 - 0,8369$	0,9963
			β ARMA Res3	$\log(\text{ARL}) \times 0,7888 - 0,8139$	0,9963
			β ARMA Res4	$\log(\text{ARL}) \times 0,8007 - 0,8528$	0,9938
			ARIMA	$\log(\text{ARL}) \times 0,8075 - 0,8562$	0,9963
			EWMA	$\text{sqrt}(\text{ARL}) \times 0,0576 + 1,1287$	0,9649
		EWMA	β ARMA Res1	$\log(\text{ARL}) \times 0,4600 + 0,4646$	0,9974
			β ARMA Res2	$\log(\text{ARL}) \times 0,4711 + 0,4401$	0,9980
			β ARMA Res3	$\log(\text{ARL}) \times 0,4663 + 0,4483$	0,9978
			β ARMA Res4	$\log(\text{ARL}) \times 0,4636 + 0,4541$	0,9976
			ARIMA	$\log(\text{ARL}) \times 0,4653 + 0,4557$	0,9978
			EWMA	$\text{sqrt}(\text{ARL}) \times 0,0528 + 1,1273$	0,9510

Tabela A.7: Apresentação dos modelos de regressão lineares aos ARL's do cenário com $\phi = 40$ e $\mu = 0,8$.

Precisão	μ	GC	Ajuste	Modelo	R^2
40	0,8	Shewhart	β ARMA Res1	$\log(\text{ARL}) \times 0,4950 + 0,4413$	0,9994
			β ARMA Res2	$\log(\text{ARL}) \times 0,5562 + 0,2342$	0,9934
			β ARMA Res3	$\log(\text{ARL}) \times 0,4616 + 0,5388$	0,9976
			β ARMA Res4	$\log(\text{ARL}) \times 0,4120 + 0,6719$	0,9920
			ARIMA	$\log(\text{ARL}) \times 0,4924 + 0,4363$	0,9996
			EWMA	$\log(\text{ARL}) \times 0,4619 + 0,5190$	0,9976
		CUSUM	β ARMA Res1	$\log(\text{ARL}) \times 0,8056 - 0,8456$	0,9971
			β ARMA Res2	$\log(\text{ARL}) \times 0,9654 - 1,3886$	0,9871
			β ARMA Res3	$\log(\text{ARL}) \times 0,8069 - 0,8701$	0,9970
			β ARMA Res4	$\log(\text{ARL}) \times 0,7855 - 0,8305$	0,9939
			ARIMA	$\log(\text{ARL}) \times 0,8111 - 0,8914$	0,9970
			EWMA	$\text{sqrt}(\text{ARL}) \times 0,0607 + 1,1448$	0,9670
		EWMA	β ARMA Res1	$\log(\text{ARL}) \times 0,4648 + 0,4348$	0,9969
			β ARMA Res2	$\log(\text{ARL}) \times 0,6328 - 0,0319$	0,9808
			β ARMA Res3	$\log(\text{ARL}) \times 0,3987 + 0,6864$	0,9988
			β ARMA Res4	$\log(\text{ARL}) \times 0,4604 + 0,4506$	0,9971
			ARIMA	$\log(\text{ARL}) \times 0,4930 + 0,3429$	0,9977
			EWMA	$\text{sqrt}(\text{ARL}) \times 0,0613 + 1,3216$	0,9458

Tabela A.8: Apresentação dos modelos de regressão lineares aos ARL's do cenário com $\phi = 100$ e $\mu = 0,8$.

Precisão	μ	GC	Ajuste	Modelo	R^2
100	0,8	Shewhart	β ARMA Res1	$\log(\text{ARL}) \times 0,4469 + 0,5652$	0,9974
			β ARMA Res2	$\log(\text{ARL}) \times 0,4587 + 0,5364$	0,9977
			β ARMA Res3	$\log(\text{ARL}) \times 0,4286 + 0,6207$	0,9951
			β ARMA Res4	$\log(\text{ARL}) \times 0,4096 + 0,6701$	0,9922
			ARIMA	$\log(\text{ARL}) \times 0,4473 + 0,5497$	0,9976
			EWMA	$\log(\text{ARL}) \times 0,4316 + 0,5899$	0,9959
		CUSUM	β ARMA Res1	$\log(\text{ARL}) \times 0,8050 - 0,8448$	0,9961
			β ARMA Res2	$\log(\text{ARL}) \times 0,8657 - 1,0273$	0,9939
			β ARMA Res3	$\log(\text{ARL}) \times 0,8063 - 0,8585$	0,9964
			β ARMA Res4	$\log(\text{ARL}) \times 0,8105 - 0,8778$	0,9935
			ARIMA	$\log(\text{ARL}) \times 0,7912 - 0,8368$	0,9966
			EWMA	$\text{sqrt}(\text{ARL}) \times 0,0595 + 1,1447$	0,9664
		EWMA	β ARMA Res1	$\log(\text{ARL}) \times 0,4675 + 0,4433$	0,9978
			β ARMA Res2	$\log(\text{ARL}) \times 0,5362 + 0,2659$	0,9978
			β ARMA Res3	$\log(\text{ARL}) \times 0,4777 + 0,4216$	0,9983
			β ARMA Res4	$\log(\text{ARL}) \times 0,4682 + 0,4431$	0,9978
			ARIMA	$\log(\text{ARL}) \times 0,4746 + 0,4006$	0,9981
			EWMA	$\text{sqrt}(\text{ARL}) \times 0,0575 + 1,3087$	0,9507

Tabela A.9: Apresentação dos modelos de regressão lineares aos ARL's do cenário com $\phi = 200$ e $\mu = 0,8$.

Precisão	μ	GC	Ajuste	Modelo	R^2
200	0,8	Shewhart	β ARMA Res1	$\log(\text{ARL}) \times 0,4282 + 0,6149$	0,9954
			β ARMA Res2	$\log(\text{ARL}) \times 0,4324 + 0,6073$	0,9954
			β ARMA Res3	$\log(\text{ARL}) \times 0,4186 + 0,6435$	0,9937
			β ARMA Res4	$\log(\text{ARL}) \times 0,4091 + 0,6681$	0,9923
			ARIMA	$\log(\text{ARL}) \times 0,4266 + 0,5985$	0,9953
			EWMA	$\log(\text{ARL}) \times 0,4188 + 0,6158$	0,9944
		CUSUM	β ARMA Res1	$\log(\text{ARL}) \times 0,8083 - 0,8654$	0,9965
			β ARMA Res2	$\log(\text{ARL}) \times 0,8423 - 0,9677$	0,9954
			β ARMA Res3	$\log(\text{ARL}) \times 0,8093 - 0,8746$	0,9965
			β ARMA Res4	$\log(\text{ARL}) \times 0,8197 - 0,9110$	0,9940
			ARIMA	$\log(\text{ARL}) \times 0,7769 - 0,8210$	0,9969
			EWMA	$\text{sqrt}(\text{ARL}) \times 0,0582 + 1,1343$	0,9666
EWMA	β ARMA Res1	$\log(\text{ARL}) \times 0,4682 + 0,4460$	0,9980		
	β ARMA Res2	$\log(\text{ARL}) \times 0,5038 + 0,3542$	0,9989		
	β ARMA Res3	$\log(\text{ARL}) \times 0,4737 + 0,4347$	0,9982		
	β ARMA Res4	$\log(\text{ARL}) \times 0,4693 + 0,4439$	0,9980		
	ARIMA	$\log(\text{ARL}) \times 0,4646 + 0,4191$	0,9976		
	EWMA	$\text{sqrt}(\text{ARL}) \times 0,0554 + 1,2936$	0,9537		

APÊNDICE B – Resultados para suavização exponencial simples

Neste apêndice são apresentadas em forma de tabelas as avaliações numéricas dos gráficos de controle baseados nos resíduos do modelo SES.

Tabela B.1: Apresentação dos resultados para o ARL_1 dos GC baseados nos resíduos do modelo de suavização exponencial simples do cenário para $\mu = 0,2$.

ϕ	δ	40			100			200		
		Shewhart	CUSUM	EWMA	Shewhart	CUSUM	EWMA	Shewhart	CUSUM	EWMA
	-0,20	431,2145	583,6854	418,1718	493,9592	657,9652	496,2553	585,5947	724,3408	543,3143
	-0,15	352,9570	452,1440	342,1852	403,7277	488,7576	399,8363	455,6483	512,6680	410,3717
	-0,10	297,7939	343,1138	286,4247	317,9233	351,9027	309,7085	344,2531	364,7364	322,9087
	-0,05	243,7544	262,0293	238,4104	249,0059	276,5520	255,4884	264,5199	268,7323	257,4258
	0,00	200,5029	200,6291	199,4597	200,1304	199,0944	200,2901	199,7790	199,9520	200,2627
	0,05	171,7696	159,0449	168,5939	161,1399	156,6344	163,4613	159,0926	159,2498	159,3482
	0,10	139,8530	126,7010	141,0114	127,5632	123,8157	131,2781	125,0666	127,1063	131,7201
	0,15	117,4347	103,4341	119,0137	104,7796	101,2300	109,0452	103,0793	102,3328	107,3723
	0,20	100,4617	86,1966	101,0405	87,7615	86,0134	94,2507	85,0445	88,0865	91,0893

Tabela B.2: Apresentação dos resultados para o ARL_1 dos GC baseados nos resíduos do modelo de suavização exponencial simples do cenário para $\mu = 0,5$.

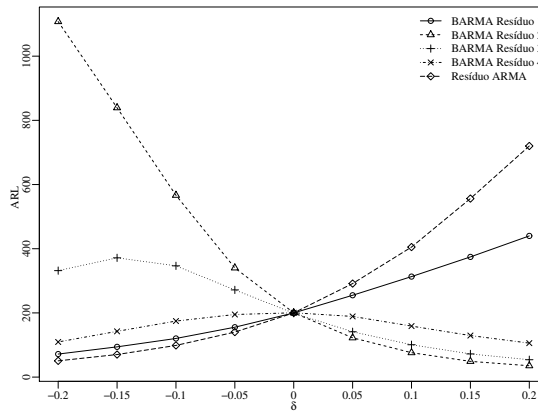
ϕ	δ	40			100			200		
		Shewhart	CUSUM	EWMA	Shewhart	CUSUM	EWMA	Shewhart	CUSUM	EWMA
	-0,20	235,8840	246,5272	228,3180	242,9098	260,4761	255,0245	253,1757	271,2270	265,3821
	-0,15	220,0790	228,1747	215,0717	223,9602	238,9639	228,4864	227,0595	232,7664	226,8817
	-0,10	205,4833	212,7049	203,9028	210,3386	214,8784	213,2619	212,2668	214,5041	207,9807
	-0,05	198,0899	207,4620	198,9607	200,1877	203,7834	203,3147	207,6346	203,1628	203,0824
	0,00	199,5980	200,5169	200,4914	201,8442	200,0404	200,5793	200,4519	200,5902	200,9462
	0,05	201,1147	205,5948	193,3594	202,4933	198,1785	198,0003	202,9980	204,4340	198,5328
	0,10	208,4700	212,3220	202,7681	208,7691	217,9616	207,7402	217,0253	212,8103	204,4709
	0,15	216,8232	220,3553	217,4170	225,2728	235,8674	223,7485	229,6871	235,9057	228,0381
	0,20	232,4448	250,3876	235,2432	242,0521	260,6579	248,7476	254,4494	263,7141	252,0368

Tabela B.3: Apresentação dos resultados para o ARL_1 dos GC baseados nos resíduos do modelo de suavização exponencial simples do cenário para $\mu = 0,8$.

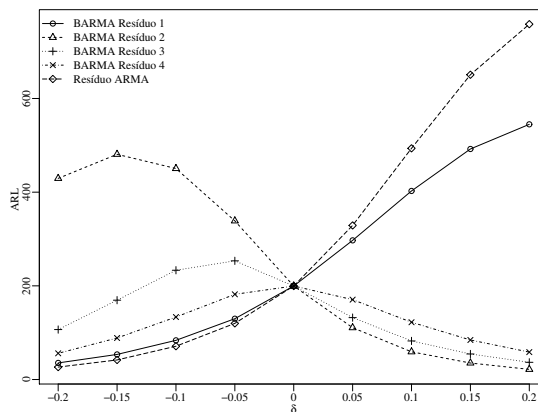
ϕ	δ	40			100			200		
		Shewhart	CUSUM	EWMA	Shewhart	CUSUM	EWMA	Shewhart	CUSUM	EWMA
	-0,20	101,2532	89,1773	103,4605	86,5872	84,5511	91,9807	82,1976	88,1648	95,0495
	-0,15	118,1330	109,3787	124,9365	103,8880	99,5671	107,7177	97,7076	99,9218	108,0727
	-0,10	137,8131	128,1646	140,8131	127,3043	124,5161	135,3316	121,8198	122,7330	134,0465
	-0,05	168,7284	163,9591	176,4834	156,6763	152,1563	163,9878	154,4190	153,7369	162,2711
	0,00	200,0444	199,5894	200,2118	199,9147	200,7173	200,7098	199,3744	200,5655	200,1807
	0,05	245,0379	268,7140	250,3950	251,6702	254,9455	246,8147	254,0109	265,2162	263,1481
	0,10	297,6754	354,0418	297,1715	321,1526	343,1789	312,7222	333,9869	358,3464	328,1463
	0,15	359,2450	467,2769	366,0934	399,2989	480,9513	398,5420	439,7483	501,7907	427,3961
	0,20	428,8273	604,2170	425,4628	505,5749	638,8279	501,8392	571,2124	704,9219	554,7220

APÊNDICE C – Resultados do ARL_1 para os cenários 7, 8 e 9

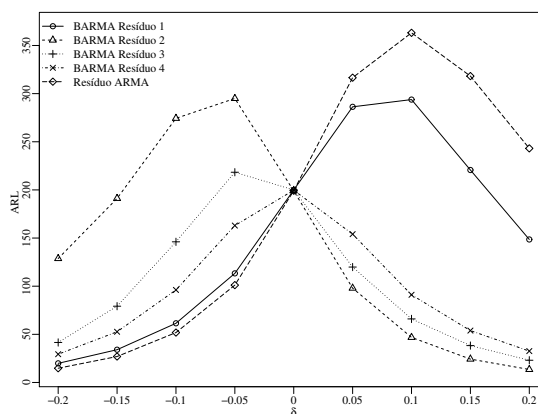
Neste apêndice encontram-se os resultados da avaliação numérica dos GC residuais para os Cenários 7, 8, e 9.



(a) Shewhart Cenário 7

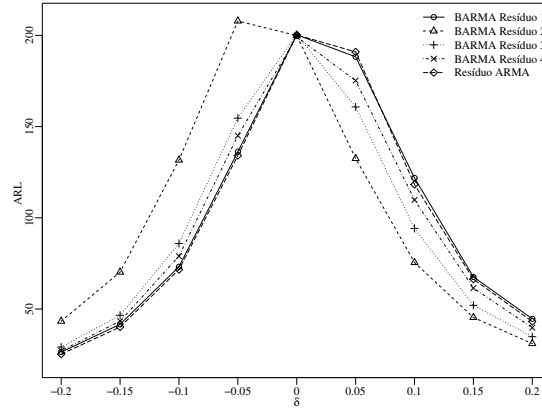


(b) Shewhart Cenário 8

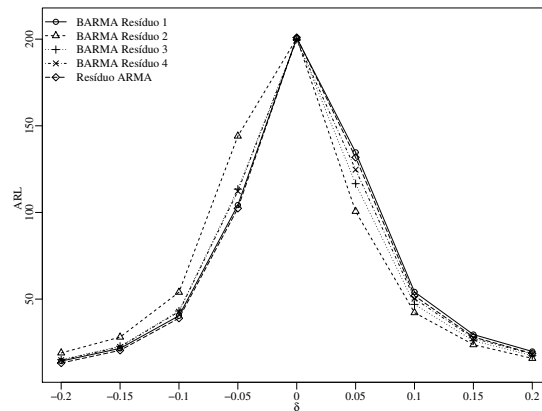


(c) Shewhart Cenário 9

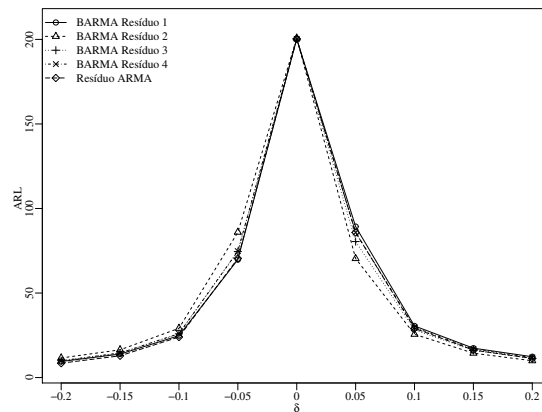
Figura C.1: ARL_1 para os gráficos de controle residuais de Shewhart para os cenários 7,8 e 9.



(a) CUSUM Cenário 7

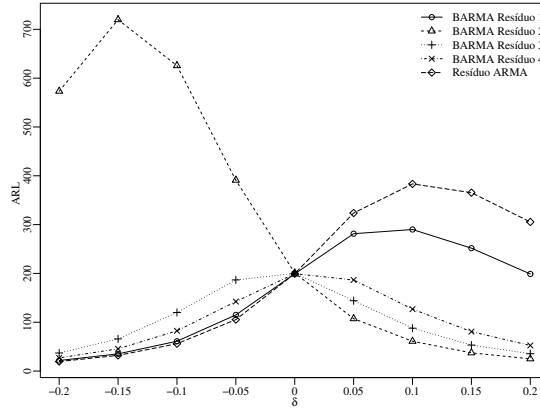


(b) CUSUM Cenário 8

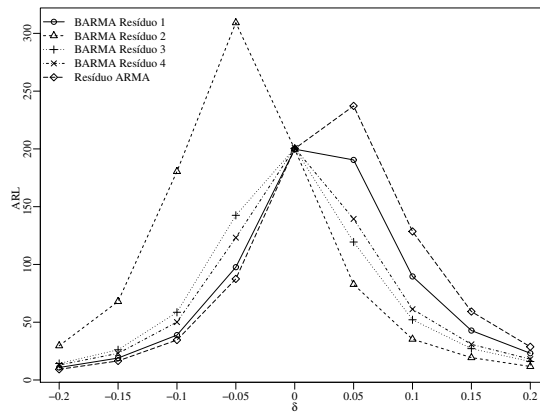


(c) CUSUM Cenário 9

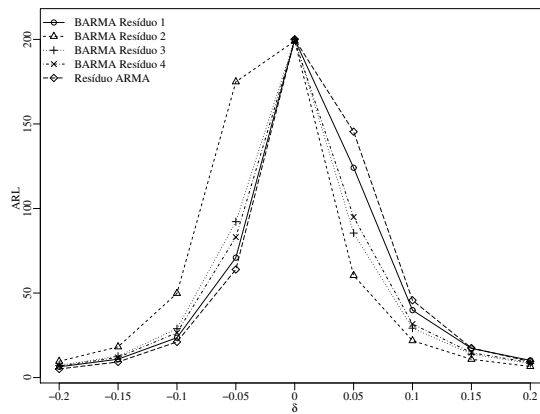
Figura C.2: ARL_1 para os gráficos de controle residuais CUSUM para os cenários 7,8 e 9.



(a) EWMA Cenário 7



(b) EWMA Cenário 8



(c) EWMA Cenário 9

Figura C.3: ARL_1 para os gráficos de controle residuais EWMA para os cenários 7,8 e 9.