

ADRIANO BURAN MOALA

**Monitoramento de Processos com Dependência Espaço-Temporal Utilizando
Gráficos de Controle**

São Paulo
2017

ADRIANO BURAN MOALA

**Monitoramento de Processos com Dependência Espaço-Temporal Utilizando
Gráficos de Controle**

Dissertação apresentada à Escola Politécnica da Universidade de São Paulo para obtenção do título de Mestre em Ciências

Área de Concentração: Engenharia de Sistemas Logísticos

Orientadora: Linda Lee Ho

São Paulo
2017

Catálogo-na-publicação

Moala, Adriano Buran
Monitoramento de Processos com Dependência Espaço-Temporal
Utilizando Gráficos de Controle / A. B. Moala -- São Paulo, 2017.
116 p.

Dissertação (Mestrado) - Escola Politécnica da Universidade de São Paulo. Departamento de Engenharia de Produção.

1.Logística 2.Gráficos de Controle 3.STARMA 4.MCUSM 5.MEWMA
I.Universidade de São Paulo. Escola Politécnica. Departamento de Engenharia de Produção II.t.

AGRADECIMENTOS

Agradeço imensamente a minha orientadora Linda Lee Ho por toda ajuda durante este trabalho.

Também gostaria de agradecer aos professores que participaram da banca, Roberto Quinino e Denise Botter.

Aos professores que tive o privilégio de ter aula, Claudio Barbieri (Programação Linear), Hugo Yoshizaki (Cadeia de Suprimentos), Rui Botter (Simulação de Sistemas Discretos), Orlando Strambi (Planejamento Urbano), Joaquim R. Santos (Sistemas Dinâmicos) e Débora Ronconi (Programação Linear Inteira).

Aos colegas de curso mais próximos pela parceria nos trabalhos desenvolvidos, inclusive nas diversas madrugadas, Adriana Kanashiro, Hamilton C. Carvalho, João Momesso, Leonardo Guerin, Mariana Flora, Paula Naomi, Raquel F. Buzogany e Samuel Neto.

Aos meus gestores na Porto Seguro que sempre valorizaram o conhecimento Felipe Prieto e Fabio Morita.

Agradeço também minha esposa Juliana Totti por entender todo o tempo de dedicação.

Aos meus pais Angela e Francisco Moala, e aos meus tios mais próximos que sempre deram conselhos valiosos Fernando A. Moala, Marcelo M. Moala e Waldomiro A. Moala.

RESUMO

O combate ao roubo de veículos requer monitoramento constante e ações policiais para alterar a logística de rondas. A proposta deste trabalho é apresentar uma aplicação de como monitorar o índice de roubo de veículos nos 93 distritos da cidade de São Paulo e estabelecer alertas quando houver um aumento da criminalidade que é considerado fora do padrão histórico. Para modelar a taxa de roubo em todos os distritos da cidade foi utilizado o modelo STARMA (*Spatio-Time Autoregressive Moving Average*) que incorpora dependência espaço-temporal. Já para os gráficos de controle foram utilizados o MEWMA (*Multivariate Exponentially Weighted Moving Average*) e o MCUSUM (*Multivariate Cumulative Sum*) direcionado para capturar aumentos. Os resultados indicaram que o MCUSUM teve um tempo de reação a aumentos da criminalidade menor que o MEWMA. Além disso, foi testado como seria o tempo de reação dessas estatísticas sem a presença da componente espacial do modelo STARMA e o resultado foi uma reação incorreta, com aumento de falsos positivos.

Palavras-chaves: logística, gráficos de controle, STARMA, MCUSUM, MEWMA.

ABSTRACT

A constant monitoring and police actions to change the routes of patrol vehicles are some requirements to act against the vehicle theft. The purpose of this dissertation is to present an application of monitoring the vehicle theft rates by control charts in the 93 police districts of the city of São Paulo. The control charts are built to detect increases in the crime rates, so a signal is triggered in regions where the crime rates are considered abnormal from the historical pattern. A STARMA (Spatio-Time Autoregressive Moving Average) model that incorporates space-time dependency is used to model the rate of robbery in all districts. MEWMA (Multivariate Exponentially Weighted Moving Average) and the MCUSUM (Multivariate Cumulative Sum) are built to meet some performance criteria. The results pointed out that MCUSUM outperforms MEWMA to capture increases in crime. Additionally earlier false alarms are observed in both charts as consequences when spatial components of STARMA model are wrongly omitted.

Key-words: logistic, control charts, STARMA, MCUSUM, MEWMA.

SUMÁRIO

1	Introdução	1
2	Objetivos Gerais e Específicos	4
3	Revisão Bibliográfica	5
3.1	Modelos para Dados com Correlação Espacial	11
3.1.1	SEM - Spatial Error Model	12
3.1.2	SAR - Spatial Lag ou Spatial Autoregressive Model	12
3.1.3	MRSAR - High-Order Mixed-Regressive Spatial Autoregressive Model	13
3.2	Modelos para Dados com Correlação Espaço-Temporal	14
3.2.1	Modelo em Painei	14
3.2.2	Modelo STARMA	15
3.3	Gráficos de Controle	19
3.3.1	Medida de Desempenho	21
3.3.2	Gráfico de Controle Shewhart	22
3.3.3	Exponentially Weighted Moving Average - EWMA	23
3.3.4	Cumulative Sum - CUSUM	24
4	Modelagem e Resultados	28
4.1	Ajuste do Modelo STARMA	29
4.2	Gráficos de Controle do MCUSUM e MEWMA	40
4.3	Aplicação do Gráfico de Controle	47
5	Conclusões	50
	REFERÊNCIAS	52

APÊNDICE A – TABELAS E FIGURAS	56
APÊNDICE B – STARMA	79
APÊNDICE C – CÓDIGOS R	82

LISTA DE SÍMBOLOS

N	número de locais
N_h	utilizado no cálculo do variograma, indica a quantidade de localidades cuja distância é h
T	quantidade de unidades de tempo
\mathbf{Z}_t	vetor de variável resposta das N localidades no instante t , com dimensão $N \times T$
$Z_{i,t}$	variável resposta da i -ésima localidade no instante t
\mathbf{Z}	vetor de variável resposta das N localidades nos modelos com ausência de componente temporal
Z_i	variável resposta da i -ésima localidade nos modelos com ausência de componente temporal
\mathbf{X}	matriz com dados exógenos, com dimensão $N \times M$
β	vetor de parâmetros das variáveis exógenas de \mathbf{X} , com dimensão $M \times 1$
\mathbf{W}	matriz de vizinhança, com dimensão $N \times N$
p	máximo lag autoregressivo
q	máximo lag de média móvel
s	defasagem espacial

ρ	parâmetro escalar associado a um único termo espacial
ρ_i	parâmetro escalar associado a ao i-ésimo termo espacial
γ_{rs}^k	covariância entre os lags espaciais r e s com lag temporal k
ρ_{rs}^k	correlação entre os lags espaciais r e s com lag temporal k
v_k	ordem espacial máxima da matriz \mathbf{W}_j ligada ao termo autoregressivo
u_k	ordem espacial máxima da matriz \mathbf{W}_j ligada ao termo de média móvel
λ	parâmetro do EWMA e do MEWMA

LISTA DE GRÁFICOS

Figura 1 – Quantidade de automóveis registrados pelo DENATRAN entre os anos de 2002 e 2016, em milhões de unidades, na cidade de São Paulo	2
Figura 2 – Vizinhança formada pelos movimentos do jogo de xadrez	6
Figura 3 – Exemplo de vizinhança segundo mapa de influência das delegacias policiais da cidade de São Paulo	7
Figura 4 – Exemplo de Variograma. Linha preta indica correlação e linhas cinzas indicam independência. Fonte: Próprio autor.	9
Figura 5 – Exemplo do Gráfico de Dispersão para Obtenção da Estatística \mathcal{I} . Fonte: Próprio autor.	11
Figura 6 – Estrutura do banco de dados	29
Figura 7 – Estrutura do banco de dados para modelagem	30
Figura 8 – Indicadores dos distritos policiais entre 2002 e 2013	31
Figura 9 – Taxa de Roubo para cada 1 milhão de veículos	32
Figura 10 – Variograma da taxa de roubo para com os meses de jan(1) até dez(12).	32
Figura 11 – Representação gráfica das matrizes de vizinhanças W_1 , W_2 e W_3	33
Figura 12 – Estatística \mathcal{I} de Moran para cada mês.	34
Figura 13 – Transformações para tornar os dados simétricos	35
Figura 14 – Transformação <i>Deviance Residual</i>	38
Figura 15 – Ajuste do modelo STARMA convertido para taxa de roubo	39
Figura 16 – Avaliação do números de réplicas para obter a distribuição do ARL	41
Figura 17 – Gráficos de Controle <i>MCUSUM</i> e <i>MEWMA</i> entre jan/14 e abr/16.	47

Figura 18 – Resultado da aplicação do <i>MCUSUM</i>	48
Figura 19 – Resultado da aplicação do <i>MEWMA</i>	49
Figura 20 – Taxa de Roubo de Veículos na Cidade de São Paulo em 2002 . . .	56
Figura 21 – Taxa de Roubo de Veículos na Cidade de São Paulo em 2003 . . .	56
Figura 22 – Taxa de Roubo de Veículos na Cidade de São Paulo em 2004 . . .	57
Figura 23 – Taxa de Roubo de Veículos na Cidade de São Paulo em 2005 . . .	57
Figura 24 – Taxa de Roubo de Veículos na Cidade de São Paulo em 2006 . . .	57
Figura 25 – Taxa de Roubo de Veículos na Cidade de São Paulo em 2007 . . .	58
Figura 26 – Taxa de Roubo de Veículos na Cidade de São Paulo em 2008 . . .	58
Figura 27 – Taxa de Roubo de Veículos na Cidade de São Paulo em 2009 . . .	58
Figura 28 – Taxa de Roubo de Veículos na Cidade de São Paulo em 2010 . . .	59
Figura 29 – Taxa de Roubo de Veículos na Cidade de São Paulo em 2011 . . .	59
Figura 30 – Taxa de Roubo de Veículos na Cidade de São Paulo em 2012 . . .	59
Figura 31 – Taxa de Roubo de Veículos na Cidade de São Paulo em 2013 . . .	60
Figura 32 – Taxa de Roubo de Veículos na Cidade de São Paulo em 2014 . . .	60
Figura 33 – Taxa de Roubo de Veículos na Cidade de São Paulo em 2015 . . .	60
Figura 34 – STACF de Ordem Espacial 0, 1, 2 e 3	61
Figura 35 – STPACF de Ordem Espacial 0, 1, 2 e 3	61
Figura 36 – STACF dos Resíduos do modelo $STAR(1,1)$, nas ordens espaciais 0, 1, 2 e 3	62
Figura 37 – STPACF dos Resíduos do modelo $STAR(1,1)$, nas ordens espaciais 0, 1, 2 e 3	62
Figura 38 – STACF dos Resíduos do modelo $STARMA(12,2,1,0)$ incompleto, nas ordens espaciais 0, 1, 2 e 3	63

Figura 39 – STPACF dos Resíduos do modelo STARMA(12,2,1,0) incompleto, nas ordens espaciais 0, 1, 2 e 3	63
Figura 40 – QQ-Normal dos Resíduos do Modelo STARMA Final - Parte 1/4 . . .	64
Figura 41 – QQ-Normal dos Resíduos do Modelo STARMA Final - Parte 2/4 . . .	65
Figura 42 – QQ-Normal dos Resíduos do Modelo STARMA Final - Parte 3/4 . . .	66
Figura 43 – QQ-Normal dos Resíduos do Modelo STARMA Final - Parte 4/4 . . .	67
Figura 44 – Comparação entre Observado e Estimado pelo Modelo STARMA - Parte 1/5.	68
Figura 45 – Comparação entre Observado e Estimado pelo Modelo STARMA - Parte 2/5.	69
Figura 46 – Comparação entre Observado e Estimado pelo Modelo STARMA - Parte 3/5.	70
Figura 47 – Comparação entre Observado e Estimado pelo Modelo STARMA - Parte 4/5.	71
Figura 48 – Comparação entre Observado e Estimado pelo Modelo STARMA - Parte 5/5.	72
Figura 49 – Mapa Numerado	78

LISTA DE TABELAS

Tabela 1 – Identificação dos modelos STARMA, com $G = \sigma^2 I$	19
Tabela 2 – Parâmetros do Modelo STARMA(12,2,1,0) Final	37
Tabela 3 – ARL para MCUSUM e MEWMA - segundo o modelo (4.7)	44
Tabela 4 – ARL para MCUSUM e MEWMA - STARMA sem componentes espaciais.	45
Tabela 5 – ARL e desvio padrão(DP) para o MCUSUM.	73
Tabela 6 – ARL e desvio padrão para o MEWMA.	74
Tabela 7 – ARL e desvio padrão para o MEWMA assintótico.	75
Tabela 8 – ARL e desvio padrão para o MCUSUM com STARMA Incorreto. . .	76
Tabela 9 – ARL e desvio padrão para o MEWMA com STARMA incorreto. . . .	77

1 Introdução

Frequentemente utilizam-se os procedimentos de controle de qualidade com os modelos de séries temporais para identificar o que seria um comportamento fora do padrão e então estabelecer estratégias corretivas (ALWAN; ROBERTS, 1988). Anderson e Diaz (1996) utilizaram essa abordagem para medir alterações na criminalidade em Houston, Texas. Alencar, Ho e Albarracín (2015) desenvolveram um gráfico de controle que avalia o número de internações diárias de pacientes, com mais de 65 anos, em um hospital devido a problemas respiratórios.

Alguns fenômenos apresentam não só uma dependência no tempo mas também no espaço. Um exemplo clássico desse efeito espacial foram as recentes crises de Dengue, no estado de São Paulo, e também a do Ebola, na África ocidental, ambas em 2014. Também na área da saúde pública o método proposto por Piroutek, Assunção e Paiva (2014) lida com as ocorrências para detectar focos de epidemia quando trata agrupamentos de doenças em janelas de tempo. Ou ainda na análise da água do oceano por causa da presença de metais pesados (CONSTANTIN; DOXARAN; CONSTANTINESCU, 2016).

No contexto econômico é possível pensar na dependência espaço-temporal de preços de imóveis (HOLLY; PESARAN; YAMAGATA, 2010; MAIUMUN et al., 2012; DUBÉ; LEGROS, 2013), no planejamento de expansão de cidades (KANTAKUMAR; KUMAR; SCHNEIDER, 2016), na qualidade do solo (SUN; ZHOU; ZHAO, 2003). Também há aplicações para medir o efeito do clima na indústria de seguros (SCHEEL et al., 2013; WOODARD et al., 2012). Kerry et al. (2010) mostraram uma aplicação da análise espacial ao roubo de veículos nos países bálticos.

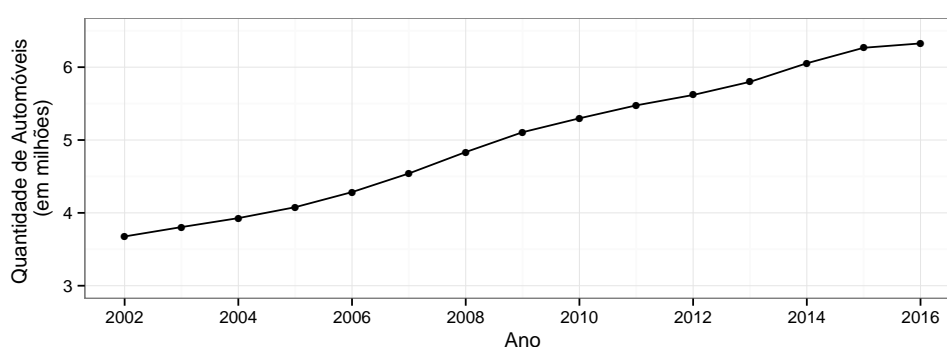
A criminalidade, que também tem essa característica espaço-temporal, trouxe um

custo estimado de R\$92 bilhões ao Brasil em 2004 (CERQUEIRA et al., 2007). Naquele ano essa perda representou aproximadamente 5% do PIB nacional. Se essa perda tiver a mesma representatividade no PIB de 2015 o valor corresponderia a quase R\$300 bilhões.

Além disso, estima-se que 70% (SUSEP, 2015) da frota brasileira não tenha seguro, isso indica que a perda do bem por roubo ou furto retira o dinheiro que iria para outros setores da economia. Já aqueles que possuem um seguro ainda terão que suportar um aumento desse serviço, pois ele também é influenciado pelos indicadores de criminalidade. Também encarece o custo de diversos itens devido ao roubo de caminhões que fazem o transporte de mercadorias e as empresas serão obrigadas a repassar as perdas através do custo do frete, que chegará ao consumidor final. Por fim, o maior custo de todos, a possibilidade de vítimas fatais decorrentes dessas ações criminosas.

Na Figura 1 a evolução da frota de automóveis cresceu aproximadamente 85% entre 2002 e 2016. Essa elevada quantidade de veículos também demandará um grande estoque de peças para reparos e é nesse ponto que o mercado paralelo é abastecido pelas peças de veículos roubados ou furtados.

Figura 1 – Quantidade de automóveis registrados pelo DENATRAN entre os anos de 2002 e 2016, em milhões de unidades, na cidade de São Paulo



O foco deste trabalho é apresentar uma abordagem com gráficos de controle,

que estabeleça um critério aos limites de estabilidade, para dados com dependência espaço-temporal. Ao final, será apresentada uma aplicação do método proposto com informações de roubo de veículos na cidade de São Paulo. Nesse contexto, essa aplicação pode ser benéfica sob o ponto de vista da logística de patrulhamento com potencial para afetar o bem estar da população ao manter a criminalidade em níveis baixos em cada uma das delegacias policiais de acordo com o orçamento estadual.

No Capítulo 3 serão apresentados os conceitos introdutórios para construção da modelagem espacial, como a definição de vizinhança e de estatísticas que indiquem a dependência espacial.

No Capítulo 3.1 serão exibidos três modelos tradicionais voltados apenas para modelagem espacial e no Capítulo 3.2 o foco será o modelo espaço-temporal da classe STARMA (*Spatio-Temporal Autoregressive Moving Average*).

Já no Capítulo 3.3 serão apresentados os gráficos de controle mais utilizados, entre eles o Shewhart, EWMA (*Exponentially Weighted Moving Average*) e CUSUM (*Cumulative Sum*) e as respectivas versões multivariadas denominadas MEWMA (*Multivariate Exponentially Weighted Moving Average*) e MCUSUM (*Multivariate Cumulative Sum*).

No Capítulo 4 será apresentada uma aplicação com os dados de roubo de veículos na cidade de São Paulo. O processo será construído em 3 etapas, na primeira a dependência dos dados será modelada pelo ajuste do STARMA, escolhido por ser um modelo parcimonioso e por ter estruturas que permitam identificar a correlação espaço-temporal. Já para obter o nível desejado de alarme será necessário determinar os parâmetros do MCUSUM e do MEWMA, mas como temos apenas uma realização da série histórica então o modelo STARMA será usado com o método de Monte

Carlo para simular novas trajetórias a fim de obter uma distribuição para o tempo até a sinalização de um alarme. Ao final, o modelo de monitoramento será aplicado aos dados de roubos mais recentes e que não foram utilizados na modelagem com intuito de verificar em quais meses e locais aconteceriam os alertas.

2 Objetivos Gerais e Específicos

O trabalho tem como objetivo apresentar o uso de gráficos de controle para dados com dependência espaço-temporal como um critério para emitir um sinal de alerta. Como exemplo serão utilizados dados de roubos de veículos que ocorreram na cidade de São Paulo entre 2002 e 2016.

Para atingir o objetivo será necessário a realização de 4 etapas:

- Ajuste do modelo STARMA (*Spatio-Temporal Autoregressive Moving Average*) aos dados de roubo das 93 delegacias policiais no período de 2002 até 2014;
- Determinar os parâmetros dos gráficos de controle:
 - MCUSUM, *Multivariate Cumulative Sum*;
 - MEWMA, *Multivariate Exponentially Weighted Moving Average*;
 - MEWMA com variância assintótica;
- Verificar o efeito da ausência da componente espacial do STARMA na sinalização do MCUSUM e também do MEWMA;
- Aplicação dos Gráficos de Controle aos dados de 2014 até 2016;

3 Revisão Bibliográfica

Neste Capítulo serão apresentados os conceitos que envolvem a dependência espacial como as matrizes de distâncias, o variograma e o índice de Moran.

Na modelagem de dados serão apresentados os modelos SAR (*Spatial Autoregressive*), SEM (*Spatial Error Model*), MRSAR (*High-Order Mixed-Regressive Spatial Autoregressivel Model*) e o STARMA (*Spatio-Temporal Autoregressive Moving Average*).

Ao final serão apresentados os gráficos de controle com intuito de monitorar a média com a suposição de variância constante, na versão univariada estão o Shewhart, EWMA e CUSUM. Já na versão multivariada estão o MEWMA e o MCUSUM que capturam flutuações bilaterais e as respectivas versões com uma formulação para monitorar variações positivas.

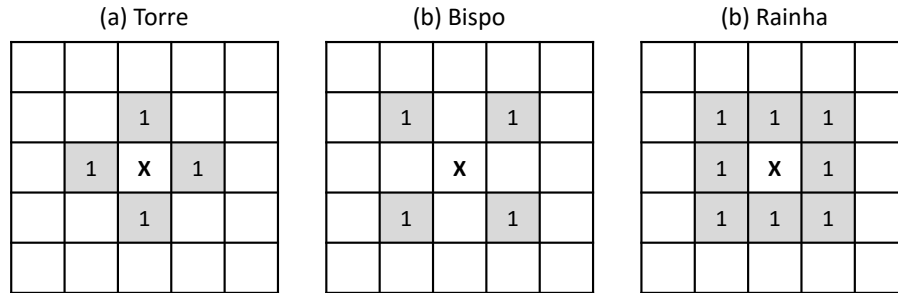
Para incorporar um termo que descreva a relação de vizinhança na modelagem espacial é utilizada uma matriz \mathbf{W} padronizada com dimensão $N \times N$, ou seja, as linhas devem estar padronizadas para que a soma seja igual a 1. Isso traz uma interpretação natural que é a noção de média da vizinhança. Os elementos w_{ij} indicam a dependência na vizinhança entre o i -ésimo e j -ésimo local.

$$\sum_{j=1}^N w_{ij} = 1, i = 1, \dots, N$$

Outra característica desta matriz é que os termos da diagonal são iguais zeros, para não existir uma relação de vizinhança com o próprio elemento em análise.

Existem construções dessa matriz com alusão ao jogo de xadrez, conforme Figura 2. A vizinhança formada pelo movimento da rainha é todo o espaço adjacente ao elemento, o formado pela torre corresponde ao movimento horizontal e vertical e o do bispo é formado pela diagonal do elemento (GETIS; ALDSTADT, 2004).

Figura 2 – Vizinhança formada pelos movimentos do jogo de xadrez



Pode-se ainda utilizar a relação de distância ao invés de adjacência. Nessa proposta atribui-se 1 para as regiões que estão até uma determinada distância e zero para aquelas que ultrapassarem esse limite. Essa denominação é chamada de matriz de distância de ordem 1. A matriz de ordem 2 parte dos elementos cuja distância é superior à de ordem anterior até um próximo limite de interesse. As matrizes de ordens superiores seguem a mesma lógica.

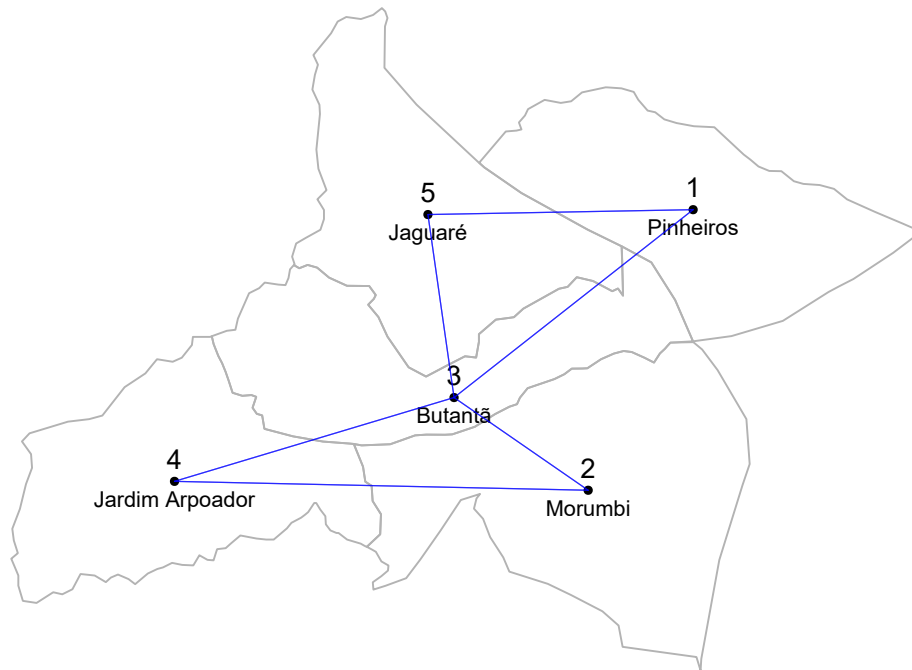
Uma outra proposta é considerar que a relação de vizinhança se forme pelo inverso da distância entre os elementos. Se estão próximos essa relação é perto de 1, mas conforme a distância aumenta então essa relação tende a zero. Deste método surgem outros como o inverso da distância ao quadrado ou outras potências que sejam adotadas. A finalidade é que as regiões mais próximas tenham uma influência maior que aquelas consideradas distantes.

Por exemplo, a Figura 3 apresenta uma estrutura na qual duas regiões são denominadas vizinhas se houver uma fronteira em comum para ambas, cuja representação matricial é expressa em $\mathbf{W}_{exemplo}$.

$$\mathbf{W}_{exemplo} = \begin{pmatrix} 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{4} & \frac{1}{4} & 0 & \frac{1}{4} & \frac{1}{4} \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \end{pmatrix}$$

Essa relação de vizinhos também pode ser descrita como sendo de ordem 1, ordem 2 ou superior. A notação da matriz de ordem k será representada por \mathbf{W}_k . Seja uma lista de distâncias $(0, d_1, d_2, \dots, d_k)$, com $0 < d_1 < d_2 < \dots < d_k$, então \mathbf{W}_1 é obtida pelos locais com distância entre 0 e d_1 e \mathbf{W}_k é obtida pelos locais entre d_{k-1} e d_k . Será utilizada a notação $\mathbf{W} = \mathbf{W}_1$ quando a ordem não for relevante para o contexto.

Figura 3 – Exemplo de vizinhança segundo mapa de influência das delegacias policiais da cidade de São Paulo



Neste trabalho a variável de interesse poderá ser representada como $Z_{i,t}$ no contexto espaço-temporal e indicará o valor da variável de interesse no i -ésimo local e instante t . Na forma vetorial, o instante t será indicado por $\mathbf{Z}_t = (Z_{1,t}, Z_{2,t}, \dots, Z_{N,t})$. Quando houver ausência de informação temporal a variável de interesse será indicada por $\mathbf{Z} = (Z_1, Z_2, \dots, Z_N)$, nesse contexto o escalar Z_i indicará a taxa de roubo de veículos na i -ésima localidade.

Nos estudos com análise espacial recorre-se frequentemente ao Variograma, que reflete a dependência fixado o instante de tempo, para identificar se há relação da

variável de interesse em função das distâncias espaciais. Para sua construção supõe-se que a variância não depende da posição no espaço, mas sim da distância entre os pontos. Seja s uma posição no espaço e $s+h$ uma nova posição cuja distância a partir de s resulta em h . Cressie (1993) indica duas condições de estacionariedade a partir da primeira diferença de $Z_{s+h} - Z_s$, com $E(Z_s) = \mu$ para qualquer posição espacial s , segue que

$$E(Z_{s+h} - Z_s) = 0$$

e

$$Var(Z_{s+h} - Z_s) = 2\gamma(h)$$

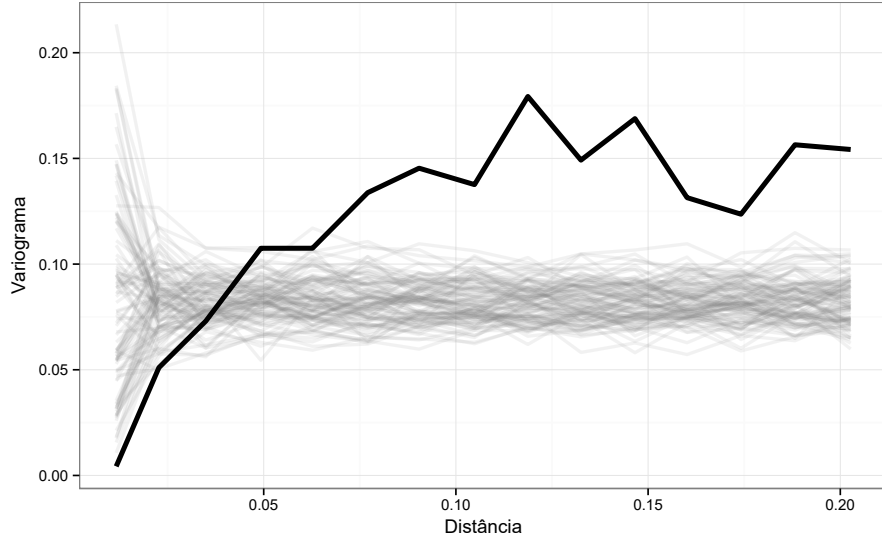
O valor de $2\gamma(h)$ é chamado variograma e $\gamma(h)$ é o semivariograma. Na prática consiste em calcular a matriz de distância dos pontos, em seguida são separadas faixas de interesse do estudo e, por fim, é obtida a correlação entre os pontos de cada faixa estabelecida previamente. No total existem N localidades, no entanto existem N_h localidades quando são agrupadas em faixas de distâncias até h . O estimador mais usual da variância é:

$$2\hat{\gamma}(h) = \frac{1}{N_h} \sum_{s=1}^{N_h} (Z_s - Z_{s+h})^2$$

Quando há indícios de correlação o gráfico tem o comportamento de crescimento e quando não há dependência tem a forma de uma linha paralela ao eixo da distância. Na Figura 4 foram geradas 100 amostras com dados independentes, representados pelas linhas cinzas e a linha preta mostra o que seria o comportamento de dados com dependência espacial.

O teste mais comum para medir a correlação espacial é o \mathcal{I} de Moran (MORAN, 1950). Nesse teste considera-se o instante de tempo fixo e o cálculo é realizado a partir da expressão (3.1), onde w_{ij} é um elemento da matriz de vizinhança \mathbf{W} e o

Figura 4 – Exemplo de Variograma. Linha preta indica correlação e linhas cinzas indicam independência. Fonte: Próprio autor.



escalar $e_i = Z_i - \bar{Z}$ representa o desvio do valor observado da i -ésima localidade em relação à média \bar{Z} , sua forma vetorial assume $\mathbf{e}' = (e_1, e_2, \dots, e_N)$.

$$\hat{\mathcal{I}} = \frac{\sum_{i=1}^N \sum_{j=1}^N w_{ij} e_i e_j}{\sum_i e_i^2} = \frac{\mathbf{e}' \mathbf{W} \mathbf{e}}{\mathbf{e}' \mathbf{e}} \quad (3.1)$$

O valor de $\hat{\mathcal{I}}$ varia de -1 a +1 e na ausência de correlação espacial assume $E(\hat{\mathcal{I}}) = -1/(N-1)$, que tende a zero quando N tende ao infinito. A variância de $\hat{\mathcal{I}}$ é:

$$Var(\hat{\mathcal{I}}) = \frac{N^3 + 3}{N^2(N-1)(N+1)} - \frac{1}{(N-1)^2}$$

Quando o valor de $\hat{\mathcal{I}}$ é maior que $-1/(N-1)$ há indícios de correlação positiva, caso contrário há indícios de correlação negativa. A hipótese nula é $\mathcal{I}_0 = -1/(n-1)$ e a estatística de teste com distribuição Normal é apresentada em (3.2).

$$\frac{\hat{\mathcal{I}} - \mathcal{I}_0}{\sqrt{Var(\hat{\mathcal{I}})}} \sim N(0, 1) \quad (3.2)$$

A estatística de teste (3.2) converge para uma distribuição Normal se o N for grande e se os dados forem normais, do contrário é necessário construir a distribuição de \mathcal{I} por meio de permutações. Em cada iteração os valores Z_i são trocados entre

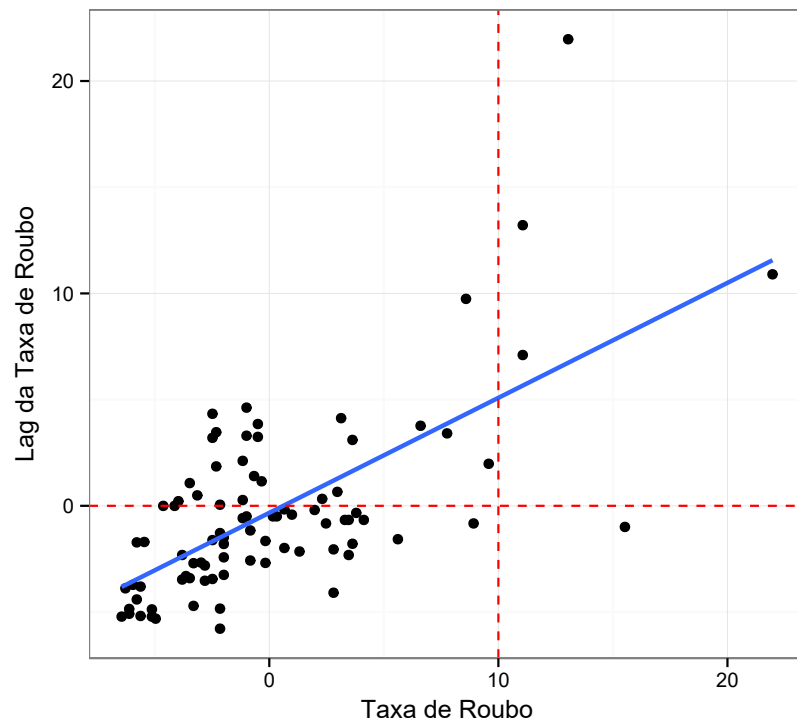
si e uma nova estatística $\hat{\mathcal{I}}_{\text{permutado}}$ é obtida. Após muitas iterações o p-valor do teste unilateral é calculado a partir da contagem do número de vezes que o valor observado de $\hat{\mathcal{I}}$ foi maior que as estatísticas obtidas pelas permutações.

Uma outra forma de interpretar o valor da estatística $\hat{\mathcal{I}}$ de Moran é utilizando um gráfico de dispersão, onde no eixo das abscissas são colocados os valores de e_i para cada uma das localidades e no eixo das ordenadas são colocados os lags espaciais de e_i . Para obter o lag espacial da variável utiliza-se a matriz de vizinhança W e calcula-se a média segundo $\sum_{j=1}^N w_{ij}e_i$. O valor de $\hat{\mathcal{I}}$ corresponde ao coeficiente angular da reta que ajusta os dados.

Este tipo de gráfico é uma forma natural de visualizar a dependência espacial, já que será possível notar descritivamente o grau de influência da medida espacializada. Se a reta de regressão tiver inclinação positiva então valores altos em e_i também estão associados a valores altos na i -ésima vizinhança. Se a inclinação for negativa a interpretação diz que valores altos em e_i estão associados a valores baixos na i -ésima vizinhança.

No exemplo da Figura 5 foram utilizados os dados da Taxa de Roubo de agosto/2014 dos distritos policiais da cidade de São Paulo (ver Figura 32) com uma matriz que indica a vizinhança se as regiões estão a menos de 4km, cada ponto é um distrito e o coeficiente angular da reta ajustada equivale à estatística $\hat{\mathcal{I}}$. Para o evento de roubo a inclinação positiva da reta tem o mesmo impacto de uma epidemia, se um local tem baixa incidência é provável que a vizinhança também seja semelhante. Agora, o descontrole de uma região pode afetar facilmente uma vizinhança que não esteja preparada.

Figura 5 – Exemplo do Gráfico de Dispersão para Obtenção da Estatística \mathcal{I} . Fonte: Próprio autor.



3.1 Modelos para Dados com Correlação Espacial

Neste Capítulo são apresentados os modelos espaciais mais frequentes encontradas na literatura como o SEM (*Spatial Error Model*), SAR (*Spatial Autoregressive Model*) e a extensão multivariada do SAR que é o MSAR. Este tipo de modelagem é indicado quando há apenas interesse em analisar a dependência espacial em um instante de tempo específico, por exemplo, o preço de residências, o potencial de consumo de um produto ou serviço.

O uso da matriz de vizinhança W sempre será na forma padronizada, ou seja, as linhas somam 1. Isso faz com que a medida de um local esteja associada à média dos valores da sua vizinhança. A notação utilizada será X para a matriz $N \times N$ de variáveis exógenas e $Z = (Z_1, \dots, Z_i, \dots, Z_N)'$ para as respostas de N regiões.

3.1.1 SEM - Spatial Error Model

Apresentado em Ord (1975), esse modelo incorpora a correlação espacial pela componente de erro com distribuição Normal Multivariada, dada por.

$$\mathbf{Z} = \mathbf{X}\beta + \mathbf{u} \quad (3.3)$$

Com

$$\mathbf{u} = \phi \mathbf{W}\mathbf{u} + \varepsilon$$

A componente \mathbf{u} pode ser isolada como (3.4).

$$\begin{aligned} (\mathbf{I} - \phi \mathbf{W})\mathbf{u} &= \varepsilon \\ \mathbf{u} &= (\mathbf{I} - \phi \mathbf{W})^{-1} \varepsilon \end{aligned} \quad (3.4)$$

Combinando as equações (3.3) e (3.4), o modelo SEM assume uma forma mais simples:

$$\mathbf{Z} = \mathbf{X}\beta + (\mathbf{I} - \phi \mathbf{W})^{-1} \varepsilon \quad (3.5)$$

O processo é estacionário, ou $(\mathbf{I} - \phi \mathbf{W})$ é não-singular, se o parâmetro espacial está no intervalo $\omega_{min}^{-1} < \phi < \omega_{max}^{-1}$, com ω sendo os autovalores de \mathbf{W} . Como a matriz \mathbf{W} é padronizada, então o autovalor máximo é igual a $\omega_{max} = 1$, mas $\omega_{min} > -1$, implicando em $\phi < 1$.

3.1.2 SAR - Spatial Lag ou Spatial Autoregressive Model

O modelo de lag espacial, muito conhecido por SAR, foi estudado por Ord (1975) e Anselin (1988), que expressa por meio do lag a correlação de regiões vizinhas, dado por

$$\mathbf{Z} = \phi \mathbf{W}\mathbf{Z} + \mathbf{X}\beta + \varepsilon \quad (3.6)$$

sendo o escalar ϕ o coeficiente autoregressivo espacial, ε um vetor de erros com distribuição Normal Multivariada, \mathbf{W} a matriz de vizinhança padronizada e \mathbf{X} a matriz de variáveis exógenas.

Ao isolar a expressão (3.6) para \mathbf{Z} o modelo tem a forma:

$$\mathbf{Z} = (\mathbf{I} - \phi \mathbf{W})^{-1} \mathbf{X} \boldsymbol{\beta} + (\mathbf{I} - \phi \mathbf{W})^{-1} \boldsymbol{\varepsilon}$$

Da mesma forma que o modelo SEM, $(\mathbf{I} - \phi \mathbf{W})$ é não-singular se o parâmetro espacial $\phi < 1$.

Kazar e Celik (2012) apresentam diversas formas de estimação dos parâmetros do modelo, inclusive em uma perspectiva recente para volumes maiores de dados utilizando algoritmos para processamento em paralelo.

3.1.3 MRSAR - High-Order Mixed-Regressive Spatial Autoregressive Model

Uma expansão do SAR é considerar lags espaciais maiores que 1 na modelagem, (LEE; LIU, 2009), com $r \geq 1$ ordens espaciais e ε o vetor de erros com distribuição Normal Multivariada.

$$\mathbf{Z} = \sum_{j=1}^r \phi_j \mathbf{W}_j \mathbf{Z} + \mathbf{X} \boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

A forma reduzida do modelo em função apenas de \mathbf{Z} é:

$$\mathbf{Z} = \left(\mathbf{I} - \sum_{j=1}^r \phi_j \mathbf{W}_j \right)^{-1} (\mathbf{X} \boldsymbol{\beta} + \boldsymbol{\varepsilon})$$

Se fosse necessário um modelo com lag espacial de ordem r , então o MRSAR(r) seria uma escrito por:

$$\mathbf{Z} = (\mathbf{I} - \phi_1 \mathbf{W}_1 - \phi_2 \mathbf{W}_2 \cdots - \phi_r \mathbf{W}_r)^{-1} (\mathbf{X} \boldsymbol{\beta} + \boldsymbol{\varepsilon})$$

Da mesma forma que os modelos anteriores, para atender a condição de não singularidade é necessário que $\sum_{j=1}^r |\phi_j| < 1$.

3.2 Modelos para Dados com Correlação Espaço-Temporal

Neste Capítulo serão apresentados modelos cuja finalidade é capturar correlações que existam no tempo e no espaço. É possível pensar em exemplos cuja natureza envolve dependência espaço-temporal como a evolução da distribuição de renda, a demanda de produtos ou serviços, a demanda por transporte público, contágio de doenças, criminalidade etc.

Em econometria são frequentemente chamados de modelos em painel (ANSELIN, 2003), porém será dada atenção especial ao modelo STARMA, que é uma expansão dos tradicionais modelos ARMA, com apresentação do método de estimação e diagnósticos.

3.2.1 Modelo em Painel

Anselin (2003) discutiu a formação de quatro modelos quando há informação no espaço e no tempo, porém com apenas um lag temporal ($t - 1$), são eles:

1. *Espaço-Recursivo Puro*, com dependência espacial nas informações do instante $t - 1$.

$$\mathbf{Z}_t = \phi \mathbf{W} \mathbf{Z}_{t-1} + \mathbf{X} \boldsymbol{\beta} + \boldsymbol{\varepsilon}_t \quad (3.7)$$

2. *Tempo-Espaço Recursivo*, com dependência espacial também em $t - 1$ e com mais um parâmetro autoregressivo temporal ψ .

$$\mathbf{Z}_t = \psi \mathbf{Z}_{t-1} + \phi \mathbf{W} \mathbf{Z}_{t-1} + \mathbf{X} \boldsymbol{\beta} + \boldsymbol{\varepsilon}_t \quad (3.8)$$

3. *Tempo-Espaço Simultâneo*, com um parâmetro autoregressivo temporal ψ e o parâmetro espacial ϕ sendo aplicado no próprio instante de tempo t .

$$\mathbf{Z}_t = \psi \mathbf{Z}_{t-1} + \phi \mathbf{W} \mathbf{Z}_t + \mathbf{X} \boldsymbol{\beta} + \boldsymbol{\varepsilon}_t \quad (3.9)$$

4. *Tempo-Espaço Dinâmico*, com ϕ_1 o parâmetro espacial aplicado em \mathbf{Z}_t e ϕ_2 o parâmetro espacial aplicado em \mathbf{Z}_{t-1}

$$\mathbf{Z}_t = \psi \mathbf{Z}_{t-1} + \phi_1 \mathbf{W} \mathbf{Z}_t + \phi_2 \mathbf{W} \mathbf{Z}_{t-1} + \mathbf{X} \boldsymbol{\beta} + \boldsymbol{\varepsilon}_{it} \quad (3.10)$$

Os modelos (3.7) e (3.8) são chamados recursivos porque dependem de \mathbf{Z}_{t-1} , já os modelos (3.9) e (3.10) são denominados simultâneo e dinâmico, pois apresentam o termo \mathbf{Z}_t no lado direito da equação.

3.2.2 Modelo STARMA

Cliff e Ord (1975) e Pfeifer e Deutsch (1980) estudaram o modelo com lag temporal e espacial chamado *Spatio-Time Autoregressive Moving Average* (STARMA) cuja estrutura foi baseada nos modelos ARMA.

$$\mathbf{Z}_t = \sum_{k=1}^p \sum_{j=0}^{v_k} \phi_{k,j} \mathbf{W}_j \mathbf{Z}_{t-k} - \sum_{k=1}^q \sum_{j=0}^{u_k} \theta_{k,j} \mathbf{W}_j \boldsymbol{\varepsilon}_{t-k} + \boldsymbol{\varepsilon}_t \quad (3.11)$$

O vetor \mathbf{Z}_t contém as respostas das N localidades no instante t , a matriz \mathbf{W}_j é a estrutura de vizinhança de j -ésima ordem com os pesos padronizados, v_k representa a quantidade de ordens espaciais existentes dentro do k -ésimo lag temporal autoregressivo, u_k é a quantidade de ordens espaciais existentes dentro do k -ésimo lag temporal de média móvel. O erro, i.i.d., é indicado no tempo e no espaço, embora a notação seja representada na forma reduzida, $\boldsymbol{\varepsilon}_t = \{\varepsilon_{1,t}, \dots, \varepsilon_{N,t}\}$. Com $\phi_{k,j}$ os parâmetros dos termos autoregressivos e $\theta_{k,j}$ os parâmetros dos termos de médias móveis. Para $j = 0$ a matriz de vizinhança torna-se neutra, $\mathbf{W}_0 = \mathbf{I}$, por isso os termos $\phi_{k,0}$ e $\theta_{k,0}$ serão puramente temporais.

A notação quando há apenas componentes autoregressivos é $\text{STAR}(p, v_p)$, quando há apenas médias móveis é $\text{STMA}(q, u_q)$ e quando tem ambos termos é $\text{STARMA}(p,$

v_p, q, u_q). Por exemplo, o modelo STAR(1,1) é expandido como:

$$\begin{aligned} \mathbf{Z}_t &= \phi_{1,0} \mathbf{W}_0 \mathbf{Z}_{t-1} + \phi_{1,1} \mathbf{W}_1 \mathbf{Z}_{t-1} + \varepsilon_t \\ &= \phi_{1,0} \mathbf{Z}_{t-1} + \phi_{1,1} \mathbf{W}_1 \mathbf{Z}_{t-1} + \varepsilon_t \end{aligned}$$

Outro exemplo é um STARMA(1,1,1,1).

$$\begin{aligned} \mathbf{Z}_t &= \phi_{1,0} \mathbf{W}_0 \mathbf{Z}_{t-1} + \phi_{1,1} \mathbf{W}_1 \mathbf{Z}_{t-1} - \theta_{1,0} \mathbf{W}_0 \varepsilon_{t-1} - \theta_{1,1} \mathbf{W}_1 \varepsilon_{t-1} + \varepsilon_t \\ &= \phi_{1,0} \mathbf{Z}_{t-1} + \phi_{1,1} \mathbf{W}_1 \mathbf{Z}_{t-1} - \theta_{1,0} \varepsilon_{t-1} - \theta_{1,1} \mathbf{W}_1 \varepsilon_{t-1} + \varepsilon_t \end{aligned}$$

É importante notar que os lags espaciais estão em função dos lags temporais, isso significa que para o lag temporal 1 poderiam existir, por exemplo, duas ordens espaciais. Para o lag temporal 2, poderia existir apenas uma ordem espacial e para um lag temporal 3 poderiam existir 4 ordens espaciais.

O modelo (3.11) pode ser expandido para incorporar uma tendência e também regressores exógenos, o STARIMAR.

$$\Delta \mathbf{Z}_t = \sum_{k=1}^p \sum_{j=0}^{v_k} \phi_{k,j} \mathbf{W}_j \Delta \mathbf{Z}_{t-k} - \sum_{k=1}^q \sum_{j=0}^{u_k} \theta_{k,j} \mathbf{W}_j \Delta \varepsilon_{t-k} + \mathbf{X} \beta + \varepsilon_t$$

No modelo de séries temporais são utilizadas as funções de autocorrelação ACF (*autocorrelation function*) e autocorrelação parcial PACF (*partial autocorrelation function*). Para os modelos espaço-temporais existem a função de autocorrelação espaço-temporal STACF (*spatio-time autocorrelation function*) e a função de autocorrelação parcial espaço-temporal STPACF (*spatio-time partial autocorrelation function*). Em Morettin e Toloi (2006), autocorrelação é a correlação entre os valores de uma série temporal em diferentes momentos no tempo. E a autocorrelação parcial de uma série de tempo, indicada por $\phi_{k,k}$, é definida como a correlação restante entre Z_t e Z_{t+k} após a retirada da influência das variáveis intermediárias $Z_{t+1}, \dots, Z_{t+k-1}$. Mais

detalhes sobre essas funções e também sobre a estacionariedade são encontrados no APÊNDICE C.

Para ajustar o modelo STARMA, na formulação de Pfeifer e Deutsch (1980), o erro ε assume distribuição normal multivariada com média zero e matriz de variância-covariância $\mathbf{G} = \sigma^2 \mathbf{I}_N$, onde \mathbf{I}_N é a matriz identidade com dimensão $N \times N$ e T o total de unidades de tempo.

$$\varepsilon = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_T)'$$

Agora, sejam $\Phi = (\phi_{1,1}, \dots, \phi_{p,v_p})'$ o vetor de parâmetros autoregressivos e $\Theta = (\theta_{1,1}, \dots, \theta_{q,u_q})'$ o vetor de parâmetros de médias móveis. A função densidade, com $S(\Phi, \Theta) = \varepsilon' \varepsilon = \sum_{i=1}^N \sum_{t=1}^T \varepsilon_{i,t}^2$, é escrita em (3.12).

$$\begin{aligned} f(\varepsilon | \Phi, \Theta, \sigma^2) &= (2\pi)^{-TN/2} |\sigma^2 \mathbf{I}_{NT}|^{-1/2} \exp \left\{ -\frac{\varepsilon' \mathbf{I} \varepsilon}{2\sigma^2} \right\} \\ &= (2\pi\sigma^2)^{-TN/2} \exp \left\{ \frac{-S(\Phi, \Theta)}{2\sigma^2} \right\} \end{aligned} \quad (3.12)$$

Como apenas os valores de $\mathbf{Z}(t)$ são conhecidos, então a obtenção dos ε são obtidos recursivamente.

$$\varepsilon_t = \mathbf{Z}_t - \sum_{k=1}^p \sum_{j=0}^{v_k} \phi_{k,j} \mathbf{W}_j \mathbf{Z}_{t-k} + \sum_{k=0}^q \sum_{j=0}^{u_k} \theta_{k,j} \mathbf{W}_j \varepsilon_{t-k} \quad (3.13)$$

A função de verossimilhança condicional assume a forma expressa em (3.14), com os vetores \mathbf{Z}_t e ε_t iguais a zero para $t < 1$:

$$L(\Phi, \Theta, \sigma^2 | \varepsilon) = (2\pi\sigma^2)^{-TN/2} \exp \left\{ \frac{-S(\Phi, \Theta)}{2\sigma^2} \right\} \quad (3.14)$$

Portanto, a estimação dos parâmetros por máxima verossimilhança, condicionada a σ^2 , é obtida minimizando:

$$S(\Phi, \Theta) = \varepsilon' \varepsilon \quad (3.15)$$

Os valores estimados para ε são calculados com a equação (3.13). Se o modelo for um STAR, então é utilizada a mesma teoria de mínimos quadrados uma vez que os resíduos condicionais estão vinculados às médias móveis e não aos termos auto-regressivos, como pode ser visto na equação (3.16).

$$\varepsilon_t = \mathbf{Z}_t - \sum_{k=1}^p \sum_{j=0}^{v_k} \phi_{k,j} \mathbf{W}_j \mathbf{Z}_{t-k} \quad (3.16)$$

A estimativa de máxima verossimilhança para $\hat{\sigma}^2$ é:

$$\hat{\sigma}^2 = \frac{\hat{S}(\hat{\Phi}, \hat{\Theta})}{NT}$$

Agora, seja $\delta = (\Phi', \Theta')'$. Para testar a hipótese nula $H_0 : \Phi = \Theta = 0$ sobre os parâmetros do modelo STARMA é necessário repetir o processo de otimização de $\hat{S}(\delta)$ com e sem o parâmetro de interesse. Seja $\hat{S}(\hat{\delta})$ a soma dos resíduos ao quadrado do modelo completo e $\hat{S}(\hat{\delta}^*)$ a soma dos resíduos ao quadrado do modelo reduzido. A estatística para o teste assume distribuição F sob a hipótese nula, com K sendo o número de parâmetros do modelo completo.

$$(TN - K) \frac{[\hat{S}(\hat{\delta}^*) - \hat{S}(\hat{\delta})]}{\hat{S}(\hat{\delta})} \sim F_{(1, TN-K)}$$

De forma análoga ao diagnóstico residual dos modelos ARMA, as estruturas ST-PACF e STACF serão utilizadas repetidamente no processo de modelagem até que não exista mais correlação residual nas componentes de espaço e tempo. Deutsch e Pfeifer (1981) apresenta a Tabela 1 para identificação dos modelos de acordo com o comportamento das correlações.

Para tornar a matriz de variância-covariância menos rígida, Deutsch e Pfeifer (1981) sugere adotar a própria matriz de variância-covariância dos resíduos para G. Dessa forma cada localidade terá a sua variância.

$$\mathbf{G} = \frac{1}{T} \sum_{t=1}^T \hat{\varepsilon}_t \hat{\varepsilon}_t'$$

Tabela 1 – Identificação dos modelos STARMA, com $G = \sigma^2 I$

Modelo	STACF	STPACF
STAR(p)	Decai Gradualmente	Zero após lag p
STMA(q)	Zero após lag q	Decai Gradualmente
STARMA(p,q)	Decai Gradualmente	Decai Gradualmente

Essa flexibilidade também resultará em algumas modificações nas funções que foram descritas anteriormente. A função de autocorrelação no lag espacial r e lag temporal k terá a incorporação da matriz \hat{G} , estimada por:

$$\hat{\rho}_{r,0}^k = \frac{T}{T-k} \frac{\sum_{t=1}^{T-k} (\mathbf{W}_r \mathbf{Z}_t)' \hat{G}^{-1} \mathbf{Z}_{t+k}}{\left(\sum_{t=1}^{T-k} (\mathbf{W}_r \mathbf{Z}_t)' \hat{G}^{-1} (\mathbf{W}_r \mathbf{Z}_t) \sum_{t=1}^{T-k} \mathbf{Z}_t' \hat{G}^{-1} \mathbf{Z}_t \right)^{-1/2}}$$

E a variância da autocorrelação também tem uma forma modificada.

$$\hat{Var}(\hat{\rho}_{r,0}^k) \approx \frac{T}{T-k} \left[\frac{tr(\mathbf{W}_r' \hat{G} \mathbf{W}_r \hat{G})}{tr(\hat{G}) tr(\mathbf{W}_r' \mathbf{W}_r \hat{G})} \right]$$

Além disso, ao invés dos parâmetros do modelo STARMA serem obtidos pela minimização dos quadrados dos resíduos conforme (3.15), agora na presença de correlação residual entre as localidades a proposta é minimizá-la buscando o menor determinante de G .

$$\hat{\delta} = \underset{\delta}{\operatorname{argmin}} \{ \det[G(\hat{\delta})] \}.$$

3.3 Gráficos de Controle

Neste Capítulo serão apresentados os procedimentos de controle de qualidade mais utilizados. Será feita uma breve descrição do gráfico de controle de Shewhart, pois o foco será nos gráficos de controle univariados e multivariados de Médias Móveis e da Soma Acumulada. Para os casos introduzidos o monitoramento será apenas

para detecção de mudanças na média considerando a variância constante, outras abordagens podem ser encontradas em Montgomery (2009) bem como a descrição a seguir sobre a difusão do conceito de controle de qualidade a partir dos anos 1900 nas grandes empresas.

Segundo Montgomery (2009), Henry Ford aprimorou a linha de produção de automóveis para reduzir problemas e aumentar a produtividade. Em 1907 a AT&T sistematizou processos de inspeção; entre 1920-30 AT&T e Bell Labs criaram um departamento de controle de qualidade. Na mesma década a General Electric começou a utilizar procedimentos estatísticos para o controle de qualidade na produção de lâmpadas.

Em 1924, W. A. Shewhart introduziu o conceito de gráfico de controle no Bell Labs. Em 1940, o Departamento de Guerra Norte-Americano publicou um guia para uso dos gráficos de controle na análise de dados. Ao fim da segunda guerra mundial, em 1946, Deming iniciou um ciclo de palestras para divulgar o controle de qualidade nas indústrias japonesas.

E em 1954 E. S. Page apresentou o gráfico de controle chamado CUSUM (Cumulative Sum). Em 1959 S. Roberts introduziu o gráfico de controle EWMA (Exponentially Weighted Moving Average).

Em 1960 os cursos de controle de qualidade estatístico foram disseminados na área de engenharia industrial e na mesma década a indústria norte-americana criou o programa *Zero-Defects* (ZD). Já em 1987 a Motorola começa a utilizar os processos seis-sigmas, mas foi em 1997 que ganhou importância nas outras indústrias.

3.3.1 Medida de Desempenho

Para avaliar se um processo está ou não sob controle pode-se pensar em um teste de hipótese no qual o processo está sob controle quando não temos rejeição da hipótese nula $H_0 : \mu = \mu_0$, ou que o processo está fora de controle quando a hipótese alternativa prevalece com $H_1 : \mu \neq \mu_0$. Com isso, o erro tipo I significa que o processo é dito fora de controle quando na verdade ele está sob controle e o erro tipo II significa que o processo é dito sob-controle quando na verdade ele está fora de controle. Neste trabalho a hipótese de processo sob-controle é satisfeita quando todos os distritos estão sob controle, $H_0 : \mu = \mu_0$ onde $\mu_0 = (\mu_{0,1}, \dots, \mu_{0,93})$, contra a hipótese alternativa que existe pelo um dos 93 distritos fora de controle, $H_1 : \exists \mu_i \in (\mu_1, \dots, \mu_{93}) : \mu_i > \mu_{0,i}$.

Um conceito importante em controle de qualidade e que é comum aos métodos é o Comprimento Médio da Sequência, ou *Average Run Length* (ARL). Com origem na distribuição Geométrica ele mede o número de eventos até ocorrer o primeiro sinal, ou seja, o número de pontos dentro da região de controle até que ocorra o primeiro ponto fora da região de controle. Seja S_t uma estatística de controle e h o limite estabelecido para que o processo esteja sob controle, então $ARL = \min\{t : S_t \geq h\}$. Agora, seja α a probabilidade de um ponto ultrapassar os limites estabelecidos, como o número de tentativas independentes até a primeira falha segue uma distribuição geométrica com média $1/\alpha$, então a quantidade média de eventos (ARL) pode ser calculada por:

$$ARL = \frac{1}{\alpha} \quad (3.17)$$

Em um exemplo em que a probabilidade da amostra exceder os limites de controle é de 5% então o $ARL = 20$. Assim, são necessárias em média 20 tentativas para encontrar um ponto fora dos limites de controle. Montgomery (2009) cita exemplos em

que a medida pode ser alvo de críticas por ter um grande desvio padrão.

O ARL também é descrito sob hipóteses, quando o processo estiver sob-controle teremos o ARL sob H_0 , que é chamado de ARL_0 . O outro conceito importante é o ARL_1 que indica o tempo médio até a detecção de um processo que está fora de controle, e portanto sob H_1 . Lembrando que $P(\text{aceitar } H_0 \mid H_1 \text{ é verdade}) = \beta(\mu_1)$, onde o valor de β é obtido após a fixação de um valor para a média do processo μ_1 . Utilizando a probabilidade complementar, portanto $P(\text{rejeitar } H_1 \mid H_1 \text{ é verdade}) = 1 - \beta(\mu_1)$, temos nesse caso que o número de eventos até ocorrer a sinalização quando H_1 é verdade segue uma distribuição geométrica com média $1/[1 - \beta(\mu_1)]$. Portanto, o ARL_1 é descrito por:

$$ARL_1 = \frac{1}{1 - \beta(\mu_1)}$$

3.3.2 Gráfico de Controle Shewhart

O gráfico de controle Shewhart foi proposto por Walter Shewhart, em 1931, como forma de medir oscilações que excedam um padrão histórico. Isso é feito a partir de dois gráficos, o da média (\bar{X}) e o da amplitude (R). Essas medidas são calculadas em instantes de tempo definidos a priori e são compostas por amostras geralmente de 5 unidades amostrais. A amplitude corresponde à diferença entre o máximo e o mínimo valor observado em cada amostra. O gráfico de controle da média seria composto por três linhas: o valor médio real, os limites superior e inferior com 3 erros padrões com origem na média. Desta forma, as médias amostrais fora dos limites indicam que o processo está fora de controle.

O gráfico de controle de Shewhart tem bom desempenho em detectar grandes desvios de um valor de referência, no entanto essa eficiência não se repete na detecção

de pequenas mudanças na média (MONTGOMERY, 2009; HAWKINS; WU, 2014), que foi melhorada com as novas abordagens como o *EWMA* e o *CUSUM*.

3.3.3 Exponentially Weighted Moving Average - EWMA

Esse gráfico de controle foi proposto por Roberts(1959), ver também Montgomery (2009), com intuito de detectar alterações no padrão de dados coletados no tempo. A estatística utilizada no monitoramento é definida como:

$$y_t = \lambda z_t + (1 - \lambda)y_{t-1}$$

onde, o parâmetro $0 < \lambda \leq 1$ é uma constante e z_0 pode ser inicializado em valor fixo μ_0 ou no valor da média amostral \bar{z} .

Se as variáveis z_t forem independentes com variância $\sigma_{z_t}^2$, então a variância de y_t é:

$$\sigma_{y_t}^2 = \sigma_{z_t}^2 \left(\frac{\lambda}{2 - \lambda} \right) [1 - (1 - \lambda)^{2t}]$$

O gráfico de controle contém os valores de y_t no tempo, a linha central em μ_0 e os limites de controle:

$$\begin{aligned} UCL &= \mu_0 + L\sigma_{z_t} \sqrt{\frac{\lambda}{2 - \lambda} [1 - (1 - \lambda)^{2t}]} \\ LCL &= \mu_0 - L\sigma_{z_t} \sqrt{\frac{\lambda}{2 - \lambda} [1 - (1 - \lambda)^{2t}]} \end{aligned}$$

Montgomery (2009) também comenta que é comum a adoção de $0.05 \leq \lambda \leq 0.25$ e $L = 3$.

Lowry et al. (1992) expandiu o EWMA para a versão multivariada e Joner et al. (2008) apresentou uma versão para detectar mudanças em apenas uma direção, cuja ponderação das observações segue por:

$$\mathbf{Y}_t = \max [0, \lambda(\mathbf{Z}_t - \boldsymbol{\mu}) + (1 - \lambda)\mathbf{Y}_{t-1}] \quad (3.18)$$

A ponderação dos valores ocorre por meio de $0 < \lambda < 1$. Para $t = 0$ inicia-se o processo com $Y_0 = 0$. A matriz de variância-covariância do processo é expressa por:

$$\Sigma_{Y_t} = \frac{\lambda[1 - (1 - \lambda)^{2t}]}{2 - \lambda} \Sigma_Z \quad (3.19)$$

O processo emite o sinal quando o escalar $E_t = Y_t' \Sigma_{Y_t}^{-1} Y_t > b$ estabelecido previamente para atender algumas medidas de desempenho.

3.3.4 Cumulative Sum - CUSUM

O processo univariado utilizando o *CUSUM* foi amplamente estudado por diversos autores como Ewan e Kemp (1960), Van Dobben de Bruyn (1968), Bissell (1969), Brook e Evans (1972), Lucas e Crosier (1982), Woodall (1983), Woodall (1984).

Esta estatística, assim como o EWMA, mantém uma memória dos valores observados anteriormente o que torna mais eficaz na detecção de mudanças pequenas e repetidas (HAWKINS; OLWELL, 1998). Enquanto diversos pontos consecutivos acima da média seriam ditos sob controle segundo Shewhart, na versão do *CUSUM* este comportamento pode indicar uma anomalia uma vez que o excesso acima da média será acumulado e conseqüentemente pode ultrapassar o limite estabelecido. A estatística é inicializada com valor zero, $C_0 = 0$, e o indicador segue acumulando recursivamente como:

$$C_t = C_{t-1} + (X_t - \mu)$$

Também é possível utilizar a versão padronizada, com a vantagem de ter os resultados em uma escala menor sem preocupação com a unidade de medida. A versão recursiva da fórmula também é útil computacionalmente, pois é neste formato que a estatística deve ser implementada para evitar somas desnecessárias sempre que uma

nova informação é coletada.

$$C_t = C_{t-1} + \frac{(X_t - \mu)}{\sigma}$$

Seja x_t a n -ésima observação, com distribuição normal, média μ e desvio padrão σ . Os limites superior e inferior são descritos por:

$$C_t^+ = \max(0, C_{t-1}^+ + X_t - \mu - k)$$

$$C_t^- = \min(0, C_{t-1}^- + X_t - \mu + k)$$

O valor de k é chamado de valor de referência, serve como uma variável de folga ou de excesso nos cálculos de C_t^+ e C_t^- . A inicialização em $t = 0$ é $C_0^- = C_0^+ = 0$, no entanto Hawkins e Wu (2014) sugere que essa prática pode tornar o procedimento pouco sensível a mudanças logo no início da medição, nesse caso seria indicado adotar um valor c diferente de zero e as novas inicializações seriam $C_0^- = -c$ e $C_0^+ = c$. A sinalização ocorre quando $C_t^- < -h$ ou $C_t^+ > h$, onde h é um valor limite e é obtido de forma a calibrar o ARL_0 .

Quando havia a necessidade de monitorar dois ou mais processos a prática comum indicava o uso de múltiplos *CUSUM*, no entanto isso deprezava possíveis correlações. Também foi aplicado o uso de componentes principais para eliminar a correlação dos processos (JACKSON; MORRIS, 1957; JACKSON, 1959; JACKSON; MUDHOLKAR, 1979). O uso de uma teoria que contemplasse processos multivariados foi introduzido por Woodall e Ncube (1985).

A estatística proposta por Healy (1987) e posteriormente por Crossier (1988), também foi estudada por Pignatiello e Runger (1990) que introduziram dois novos esquemas do MCUSUM denominados *MC1* e *MC2*, por Ngai e Zhang (2001) e por Bodnar e Schmid (2006), Bodnar e Schmid (2007). Nessas versões o controle é feito apenas

para média, enquanto que Smiley e Keoagile (2005) introduziram um novo esquema do MCUSUM para detectar mudanças tanto na média quanto na matriz de covariância.

No esquema do MCUSUM de Crossier (1988), seja $\mathbf{Z}_t = \{Z_{1,t}, \dots, Z_{n,t}\}$ um vetor cujos valores serão acumulados a partir da estatística S_t .

$$S_t = (S_{t-1} + \mathbf{Z}_t - \boldsymbol{\mu}) \left(1 - \frac{k}{d_t}\right) \quad (3.21)$$

Onde $\boldsymbol{\mu}$ é o vetor de médias de \mathbf{Z}_t , k é obtido pela solução de $k^2 = \mathbf{k}' \boldsymbol{\Sigma}^{-1} \mathbf{k}$, e o valor de d_t é calculado por (3.22). Se $d_t \leq k$, então o valor de S_t é reiniciado em zero.

$$d_t = \sqrt{(\mathbf{S}_{t-1} + \mathbf{Z}_t - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{S}_{t-1} + \mathbf{Z}_t - \boldsymbol{\mu})} \quad (3.22)$$

Por fim, a estatística final é dada por $C_t = \sqrt{\mathbf{S}_t' \boldsymbol{\Sigma}^{-1} \mathbf{S}_t}$ e a sinalização de mudança ocorre quando $C_t > h$.

Pignatiello e Runger (1990) propuseram dois novos esquemas para o *CUSUM Multivariado* chamados *MC1* e *MC2*. Seja $S_{t-n_t,t} = \sum_{i=t-n_t+1}^t (\mathbf{Z}_i - \boldsymbol{\mu}_0)$, a soma acumulada pode ser reescrita por

$$\frac{1}{n_t} S_{t-n_t,t} = \left(\frac{1}{n_t} \sum_{i=t-n_t+1}^t \mathbf{Z}_i \right) - \boldsymbol{\mu}_0.$$

Seja também $\|S_{t-n_t,t}\|_{\boldsymbol{\Sigma}} = \sqrt{\mathbf{S}_{t-n_t,t}' \boldsymbol{\Sigma}^{-1} \mathbf{S}_{t-n_t,t}}$, com $\boldsymbol{\mu}_0$ a média sob H_0 , $n_t = n_{t-1} + 1$ se $MC1_{t-1} > 0$ ou $n_t = 1$ se $MC1_{t-1} = 0$. Se a estatística $MC1_t = \max \{ \|S_{t-n_t,t}\|_{\boldsymbol{\Sigma}} - kn_t, 0 \} > h$ então há indícios de uma possível mudança no padrão da média de \mathbf{Z}_t . Embora não seja possível derivar o valor do parâmetro k a partir do esquema apresentado os autores sugerem utilizar $k = \|(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)\| / 2$, com $\boldsymbol{\mu}_1$ a média sob H_1 . Pignatiello e Runger (1990) concluíram que o *MC1* tem desempenho similar ao proposto por Crossier (1988).

A estatística *MC2* difere de *MC1* na ordem em que as distâncias entre \mathbf{Z}_t e $\boldsymbol{\mu}_0$ são obtidas, no *MC1* os valores são acumulados e depois calcula-se o desvio em relação a

μ_0 , já no *MC2* o desvio é calculado a cada instante t . Seja $D_t^2 = (\mathbf{Z}_t - \mu_0)' \Sigma^{-1} (\mathbf{Z}_t - \mu_0)$, a estatística é calculada recursivamente por $MC2_t = \max\{0, MC2_{t-1} + D_t^2 - k\}$, com $MC2_0 = 0$, e a sinalização ocorre quando $MC2_t > h$.

Ao utilizar o MCUSUM existe a possibilidade de monitorar de forma direcional apenas os aumentos de uma variável. Este método foi proposto por Fricker, Knitt e Hu (2008) de forma que os aumentos são acumulados, mas ao ocorrer uma queda a soma volta a ser igual a zero. Para fazer esse controle houve uma modificação na estrutura de S_t apresentada por:

$$S_{t,j} = \max \left[0, (S_{t-1,j} + Z_{t,j} - \mu_j) \left(1 - \frac{k}{d_t} \right) \right] \quad (3.23)$$

O processo segue com o cálculo de $C_t = \sqrt{\mathbf{S}_t' \Sigma^{-1} \mathbf{S}_t}$ e há uma sinalização quando $C_t > h$.

4 Modelagem e Resultados

A proposta deste trabalho é a aplicação de gráficos de controle quando os dados possuem correlação espaço temporal, como é o caso do roubo de veículos na cidade de São Paulo. Nesta primeira parte serão utilizadas as técnicas discutidas anteriormente para avaliar a presença dessa dependência como a construção da matriz de vizinhança, o uso do variograma e do \mathcal{I} de Moran.

A aplicação de gráficos de controle requer que os dados sejam independentes e uma forma de adequar esse ponto é a estimação de um modelo que acomode essa dependência e no caso do STARMA restará a única variável aleatória e independente que é o erro (ALWAN; ROBERTS, 1988). A escolha do STARMA reside na simplificação e redução do tempo de modelagem, o processo de escolha dos parâmetros de um STARMA via análise de correlação para modelar 93 distritos é praticamente equivalente ao de um modelo ARMA. A outra vantagem de um modelo que capture a correlação é que seria possível fazer simulações do impacto da migração da criminalidade quando aplica-se um aumento ou redução em um distrito policial.

Após essa etapa inicia-se a construção dos gráficos de controle MEWMA e o MCU-SUM, ambos na versão direcional, pois a ideia é ter alertas apenas no caso de aumentos no vetor de taxas médias e ignorar as quedas do indicador de taxa de roubo de veículos.

Para que as estatísticas sinalizem um alerta de forma eficaz é necessário determinar seus parâmetros para que respeitem uma probabilidade de falso alarme e que é traduzida na medida de desempenho apresentada anteriormente, o ARL_0 , ou seja, o ARL sob a hipótese nula. A avaliação de desempenho do gráfico de controle é calculada na presença de mudanças intencionais na média do processo e com isso será

avaliado o tempo até a emissão de um alerta, o ARL_1 , ou seja, o ARL sob a hipótese alternativa.

Nos Capítulos 4.1 e 4.2 a modelagem será feita com dados de 2002 até 2013, enquanto que no Capítulo 4.3 os dados de 2014 até 2016 serão utilizados na aplicação do gráfico de controle.

4.1 Ajuste do Modelo STARMA

Os dados foram disponibilizados após o preenchimento de uma solicitação feita no site Governo Aberto, do Estado de São Paulo, com os números de roubos de veículos por região da atuação das Delegacias. A partir de 2001, período que o dado foi coletado, os registros de roubos são vinculados ao local de ocorrência, o que é importante para não existir um viés que pode ocorrer ao deixar a escolha da delegacia para o cidadão que sofreu o crime. A Figura 6 traz a forma que o banco de dados foi organizado inicialmente para criação de mapas e filtros na seleção dos meses para modelagem do STARMA.

Figura 6 – Estrutura do banco de dados

15,996 observations of 10 variables										
	ano	mes	anomes	id	delegacia	Roubo	descricao	data	frota	taxa.roubo
1	2002	2	200202	59	059 DP - Jardim Noemia	34	modelagem	fev-02	3676180	9.2487310
2	2002	2	200202	27	027 DP - Campo Belo	65	modelagem	fev-02	3676180	17.6813975
3	2002	2	200202	23	023 DP - Perdizes	44	modelagem	fev-02	3676180	11.9689460
4	2002	2	200202	21	021 DP - Vila Matilde	93	modelagem	fev-02	3676180	25.2979996

E a Figura 7 traz a forma do banco de dados transposto contendo 93 linhas que representam os distritos e os 172 meses nas colunas. Os 4 primeiros números no nome das colunas indicam o ano e os 2 últimos números representam o mês, assim o nome da primeira coluna 200201 indica janeiro de 2002.

Como o número de veículos aumentou durante todo o período de análise também

Figura 7 – Estrutura do banco de dados para modelagem

	200201	200202	200203	200204	200205	200206	200207	200208	200209	200210
1	0.8160645	3.5362795	1.904151	3.536280	2.7202150	1.0880860	3.8083010	3.264258	3.536280	3.536280
2	1.0880860	1.3601075	2.720215	1.088086	2.1761720	0.8160645	1.0880860	2.992237	1.904151	3.264258
3	1.3601075	0.8160645	3.536280	1.904151	2.4481935	1.9041505	3.2642580	1.088086	1.360108	2.176172
4	1.0880860	3.2642580	3.264258	7.616602	1.9041505	2.1761720	1.9041505	2.176172	1.088086	1.632129
5	3.8083010	2.7202150	2.176172	2.992237	2.4481935	1.0880860	2.7202150	2.720215	2.992237	4.080323

seria esperado existir uma quantidade maior de roubos devido ao tamanho do mercado de peças. Por conta disso foi feita a primeira transformação e a variável resposta passou a ser chamada de índice de roubo por um milhão de veículos. Dessa forma a quantidade de veículos roubados do i -ésimo distrito no mês t foi dividida pelo tamanho da frota de veículos da cidade de São Paulo do mês t e o resultado foi multiplicado por 1 milhão.

$$Z_{i,t}^* = \frac{\#Roubos_{i,t}}{\#Frota_t} \times 1.000.000 \quad (4.1)$$

A Figura 8a indica que as regiões do Jabaquara, na zona Sul, e São Mateus, na zona leste, possuem os maiores valores médios e a Figura 8b mostra que essas localidades tem um baixo coeficiente de variação (desvio-padrão/média) indicando que o índice de roubo é consistentemente alto no período de 2002 à 2013. Já a região central da cidade tem a menor média, mas o índice sofre mais oscilação. Os mapas mensais de todo o período estão entre as Figuras 20-33.

A série agregada mensalmente é exibida na Figura 9, nela o índice de roubo cai de 2002 até 2008, quando entrou em um período de estabilidade até 2011 e disparou fortemente até 2014 e então surgiu uma nova queda.

Como o variograma indica a apenas a correlação espacial, os dados foram avaliados também por mês para demonstrar se há indícios de dependência. O variograma na Figura 10 apresenta maiores valores nos meses em que houve maior variabilidade

espacial nas quantidade de veículos roubados. O formato de curva de crescimento indica correlação espacial.

Para o problema de roubo de veículos foram declarados como vizinhas aquelas regiões que estão abaixo de uma determinada distância. Com isso, cada elemento da matriz recebe 1 se for vizinho e zero caso contrário. As opções de matrizes com vizinhança por contiguidade não foram consideradas, pois um veículo será alvo de roubo dependendo da distância do local de captura e entrega.

No cálculo da distância entre dois distritos policiais foi utilizada a distância Haversine, que contempla o raio de curvatura da Terra. Sejam dois pontos $P_i = (P_{i,x}, P_{i,y})$

Figura 8 – Indicadores dos distritos policiais entre 2002 e 2013

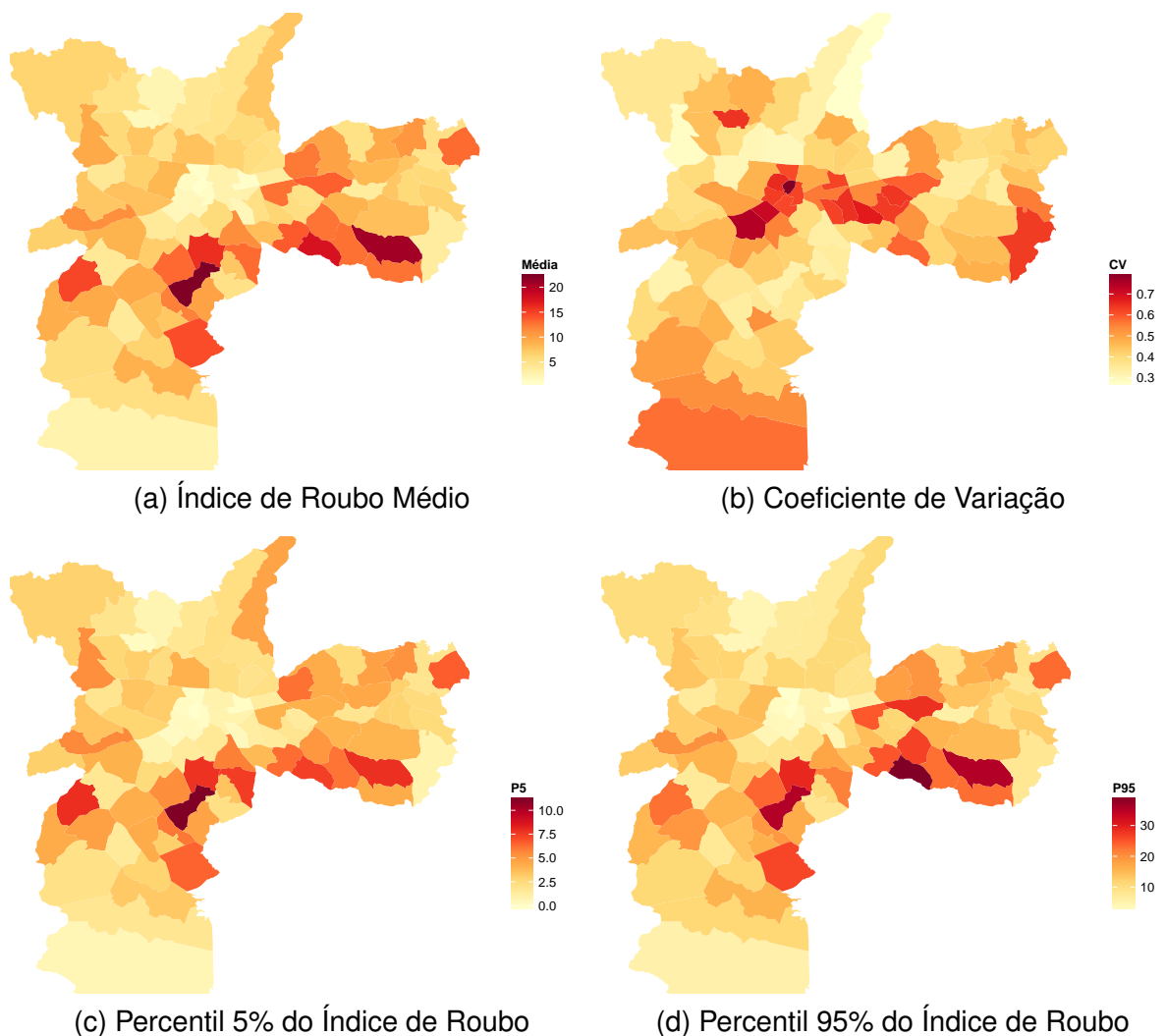
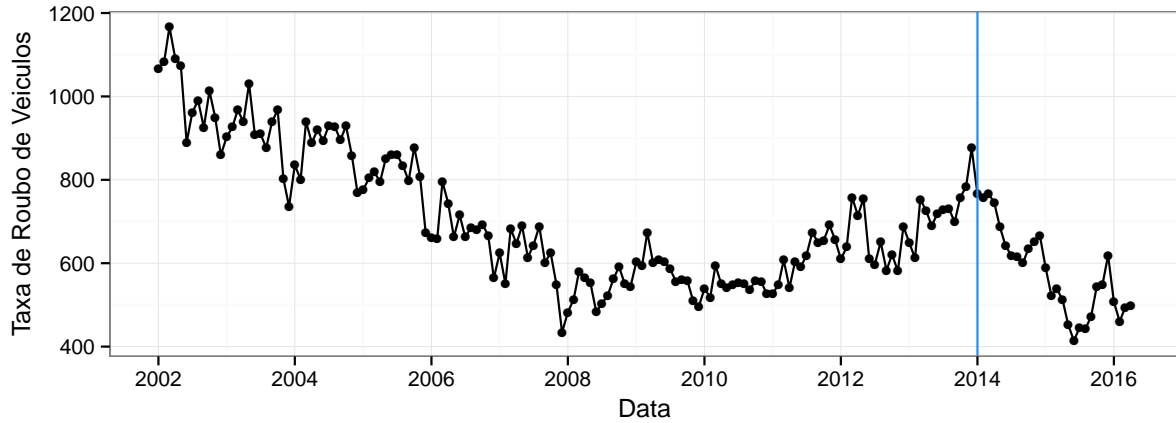


Figura 9 – Taxa de Roubo para cada 1 milhão de veículos

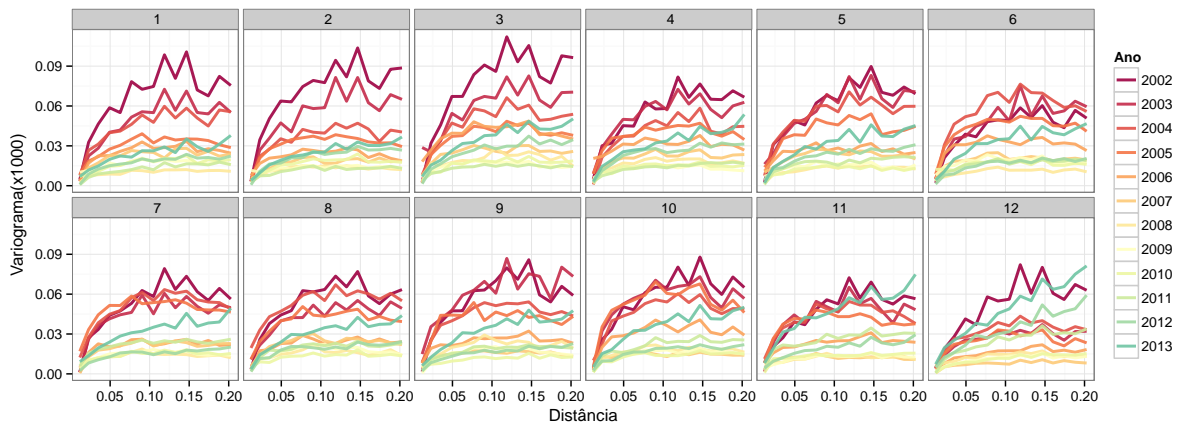


e $P_j = (P_{j,x}, P_{j,y})$ em graus, por exemplo, a localização georeferenciada da USP fica em $P_0 = (-46.73079, -23.56140)$. Sejam o raio da Terra $R \approx 6372.79548km$, $d_y = (P_{j,y} - P_{i,y})/2$ e $d_x = (P_{j,x} - P_{i,x})/2$. Portanto, a distância d , em quilômetros, entre P_i e P_j é calculada a partir de (4.2).

$$\frac{d(P_i, P_j)}{2R} = \arcsen \sqrt{\sen^2 \left(\frac{d_y \pi}{180} \right) + \cos \left(\frac{P_{i,y} \pi}{180} \right) \cos \left(\frac{P_{j,y} \pi}{180} \right) \sen^2 \left(\frac{d_x \pi}{180} \right)} \quad (4.2)$$

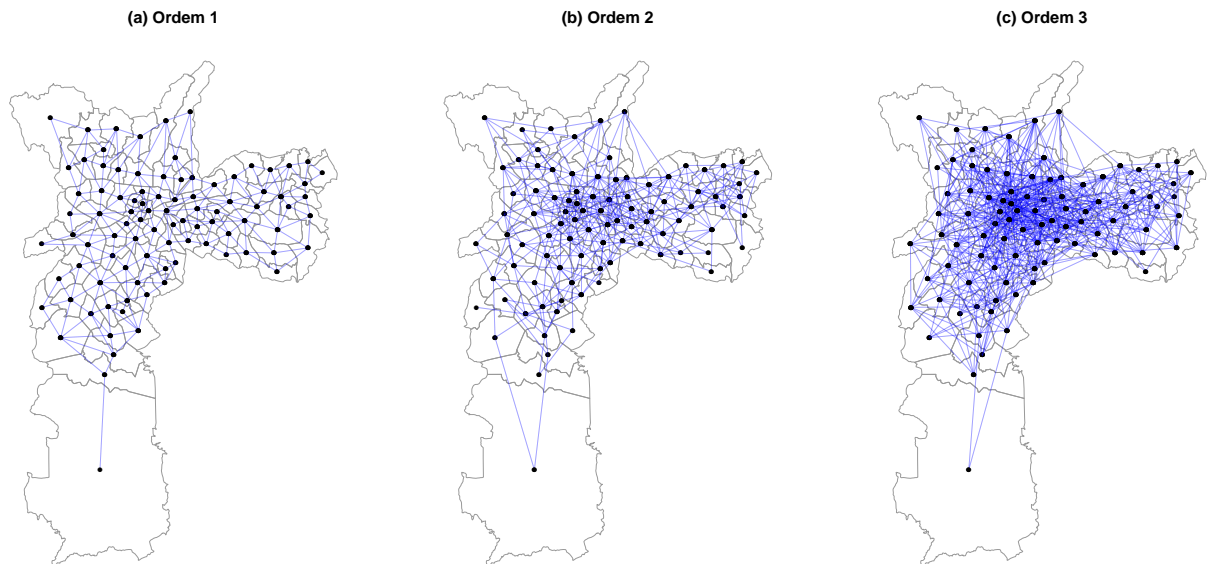
A estrutura de vizinhança adotada consta na Figura 11, do item (a) ao (c) representam as ordens espaciais de 1 a 3. Foi utilizada a distância Haversine entre todos os pontos que formam os polígonos das DP e o menor deles foi definido como a dis-

Figura 10 – Variograma da taxa de roubo para com os meses de jan(1) até dez(12).



tância oficial, esse método evita que regiões com grandes áreas não sejam vizinhas por causa da distância entre seus centróides.

Figura 11 – Representação gráfica das matrizes de vizinhanças W_1 , W_2 e W_3

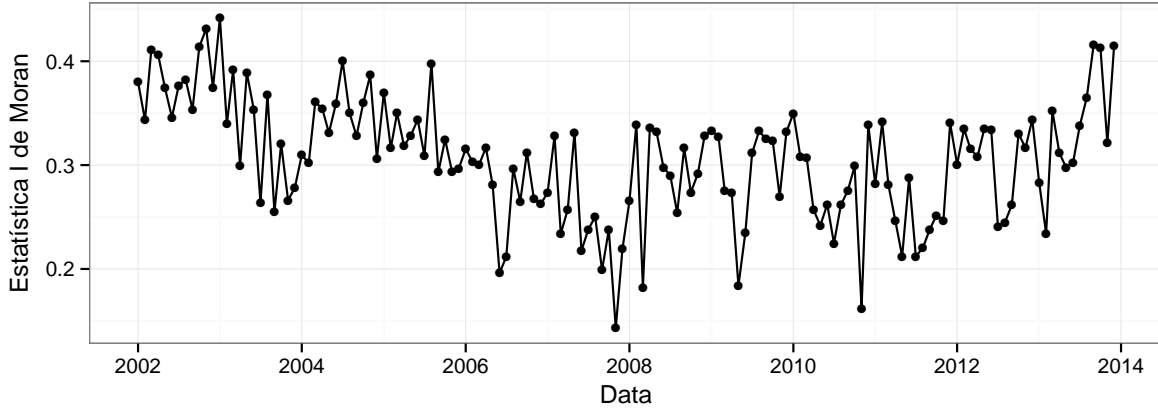


A matriz (a) W_1 tem como vizinhos regiões até 0,5km; a matriz (b) W_2 é formada por regiões não adjacentes até 3km(inclusive) e a matriz; (c) W_3 é formada é pelas regiões entre 3km e 6km(inclusive). Esses intervalos foram obtidos a partir da simulação de diversos cenários com duas restrições, a primeira era que todos tivessem ao menos um vizinho e a segunda restrição é que a correlação espacial obtida por Moran fosse igual ou decrescente conforme o avanço da ordem. A medida final foi maximizar o valor da mediana da estatística de Moran entre os anos de 2002 até o final de 2013.

Com uso da matriz de vizinhança W_1 foram obtidos os valores da estatística \mathcal{I} de Moran em cada um dos meses, Figura 12. Note que quando a criminalidade diminui o índice diminui junto e quando a criminalidade aumenta o índice também acompanha esse movimento. A explicação para isso é que para existir dependência espacial é necessário ter heterogeneidade. No entanto, quando os índices atingem os menores patamares entre 2007 e 2012 as regiões ficam mais homogêneas e por isso a

dependência espacial também diminui.

Figura 12 – Estatística \mathcal{I} de Moran para cada mês.



No modelo STARMA há suposição de resíduos normais e isso é alcançado de forma mais simples se os próprios dados tiverem uma estrutura simétrica, no entanto o item (a) da Figura 13 mostra justamente a assimetria da variável taxa de roubo. A primeira tentativa de modelagem do STARMA indicou que essa assimetria não poderia ser ignorada e por isso foram proposta duas transformações. A primeira foi a raiz quadrada e a segunda foi a *Deviance Residual* (ALENCAR; HO; ALBARRACÍN, 2015) que é expressa em (4.3). Seja $Z_{i,t}^*$ a taxa de roubo de veículos, então a variável transformada indicada por $Z_{i,t}^{DR}$ fica:

$$Z_{i,t}^{DR} = \text{sign}(Z_{i,t}^* - \mu_i) \sqrt{d_{i,t}^2} \quad (4.3)$$

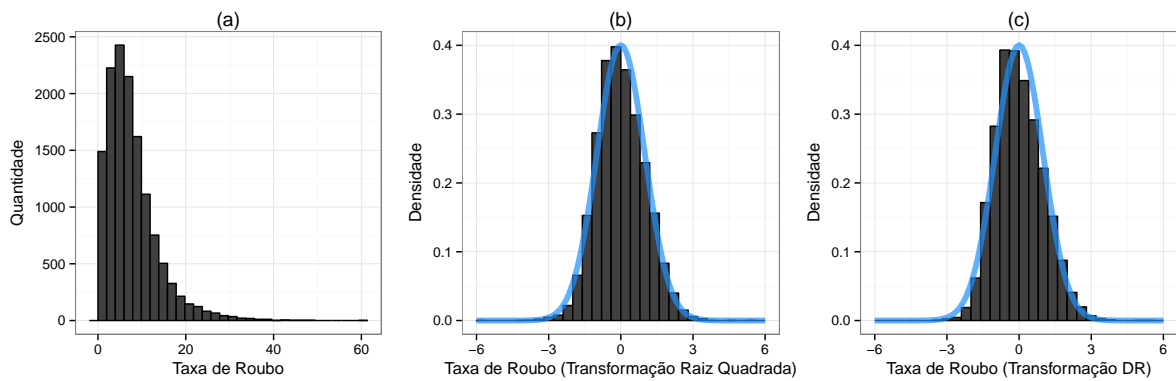
Onde $d_{i,t}^2$ é calculado por:

$$d_{i,t}^2 = \begin{cases} 2\tau_i \ln(1 + \mu_i/\tau_i) & \text{se } Z_{i,t}^* = 0 \\ 2Z_{i,t}^* \ln(Z_{i,t}^*/\mu_i) - 2\tau_i(1 + Z_{i,t}^*/\tau_i) \ln\left(\frac{1+Z_{i,t}^*/\tau_i}{1+\mu_i/\tau_i}\right) & \text{se } Z_{i,t}^* > 0 \end{cases} \quad (4.4)$$

O valor de μ_i é a média dos dados na i -ésima localidade e o parâmetro τ_i é estimado previamente observando $\text{Var}(Z_{i,t}^*) = \tau_i \pi_Z / (1 - \pi_Z)^2$, com $\pi_Z = \mu_i / (\mu_i + \tau_i)$.

O item (a) da Figura 13 mostra o histograma dos dados originais, o item (b) apresenta a transformação raiz quadrada e o item (c) mostra a transformada DR. Em ambas transformações os dados foram posteriormente padronizados pela média e variância. O modelo STARMA foi ajustado na série original e nas duas transformações citadas. Para a Raiz Quadrada os resíduos ainda indicaram muitos pontos fora da banda de confiança do gráfico QQ-Plot e foi apenas a transformação (c) que apresentou um ajuste satisfatório segundo a aderência dos resíduos.

Figura 13 – Transformações para tornar os dados simétricos



Com o intuito de não carregar a notação, a partir de agora a variável resposta Z_t^{DR} será escrita apenas como Z_t .

O modelo STARMA foi escolhido por ter uma estrutura que capture a correlação espaço-temporal dos dados e também pela possibilidade de identificação de suas componentes por meio das autocorrelações. O Algoritmo 1 indica o fluxo desse processo de modelagem que continua até que não exista mais correlação residual e ao final também espera-se que esses resíduos tenham uma aderência adequada à distribuição Normal.

De acordo com a Figura 34 as autocorrelações espaço-temporais apresentam decaimento lento para as 4 ordens espaciais. Pfeifer e Deutsch (1980) sugere um modelo STAR quando a STACF tem queda lenta e a STPACF tem uma forte queda nos lags

temporais (Tabela 1).

Algoritmo 1: Etapas de Estimação do STARMA

- 1 Calcular STACF e STPACF dos dados;
 - 2 Identificação do modelo inicial;
 - 3 **repeat**
 - 4 Estimação do modelo;
 - 5 Análise do P-Valor dos parâmetros;
 - 6 ACF e PACF dos resíduos;
 - 7 **until** *Ausência de correlação residual*;
-

Nas Figuras 34 e 35 o lag temporal 1 é candidato na ordem espacial zero. Já na ordem espacial 1 há um candidato autoregressivo de lag 1. Por isso, o modelo inicial é um STAR de ordem espacial 1 e lag temporal 1 expressa em (4.5).

$$\begin{aligned} \mathbf{Z}_t &= \phi_{1,0} \mathbf{W}_0 \mathbf{Z}_{t-1} + \phi_{1,1} \mathbf{W}_1 \mathbf{Z}_{t-1} \\ &= \phi_{1,0} \mathbf{Z}_{t-1} + \phi_{1,1} \mathbf{W}_1 \mathbf{Z}_{t-1} \end{aligned} \quad (4.5)$$

Após o ajuste do modelo STAR(1,1) os resíduos ainda apresentam correlação conforme as Figuras 36 e 37. Como o lag temporal 1 ainda é alto (Figura 34) será introduzido um termo de média móvel. Esse procedimento segue até reduzir ao máximo a correlação existente nos resíduos, que ocorreu no modelo STARMA(12,2,1,0) incompleto conforme descrito em (4.6). Esse modelo foi escolhido pois apresentou correlação residual próxima de zero (veja Figuras 38 e 39).

$$\mathbf{Z}_t = \phi_{1,0} \mathbf{Z}_{t-1} + \phi_{12,0} \mathbf{Z}_{t-12} + \phi_{1,1} \mathbf{W}_1 \mathbf{Z}_{t-1} + \phi_{2,1} \mathbf{W}_1 \mathbf{Z}_{t-2} - \theta_{1,0} \boldsymbol{\varepsilon}_{t-1} + \boldsymbol{\varepsilon}_t \quad (4.6)$$

O modelo STARMA apresentou significância em todos os parâmetros utilizados, conforme Tabela 2. Os lags puramente temporais são os autoregressivos 1 e 12, bem

como o termo de média móvel no lag 1. Na ordem espacial 1, vizinhança adjacente, os lags temporais foram de ordem 1 e 2.

Tabela 2 – Parâmetros do Modelo STARMA(12,2,1,0) Final

Categoria	Parâmetro	Estimativa	Erro Padrão	P-Valor
AR Temporal Lag 1	$\phi_{1,0}$	0,80409	0,02172	< 0,0001
AR Temporal Lag 12	$\phi_{12,0}$	0,04687	0,00605	< 0,0001
MA Temporal Lag 1	$\theta_{1,0}$	0,51406	0,02115	< 0,0001
AR Spatio-Temporal Lag 1	$\phi_{1,1}$	0,22849	0,01645	< 0,0001
AR Spatio-Temporal Lag 2	$\phi_{2,1}$	-0,10124	0,01848	< 0,0001

Com base no modelo estimado (4.7), nota-se que no lag temporal puro o valor de \hat{Z}_t corresponde a 80% do índice do mês anterior e a 5% do mesmo mês do ano anterior. Já nas ordens espaciais, a taxa de roubo é composta por 23% da vizinhança do mês anterior e uma correlação negativa no lag temporal 2 com uma correção de -10%.

$$\begin{aligned}
 \hat{Z}_t = & + 0,80409Z_{t-1} + 0,04687Z_{t-12} \\
 & + 0,22849W_1Z_{t-1} - 0,10124W_1Z_{t-2} \\
 & - 0,51406\varepsilon_{t-1} + \varepsilon_t
 \end{aligned} \tag{4.7}$$

E, como $\hat{\varepsilon}_{t-1} = Z_{t-1} - \hat{Z}_{t-1}$, a projeção para \hat{Z}_t é corrigida em aproximadamente 51% do resíduo no lag anterior.

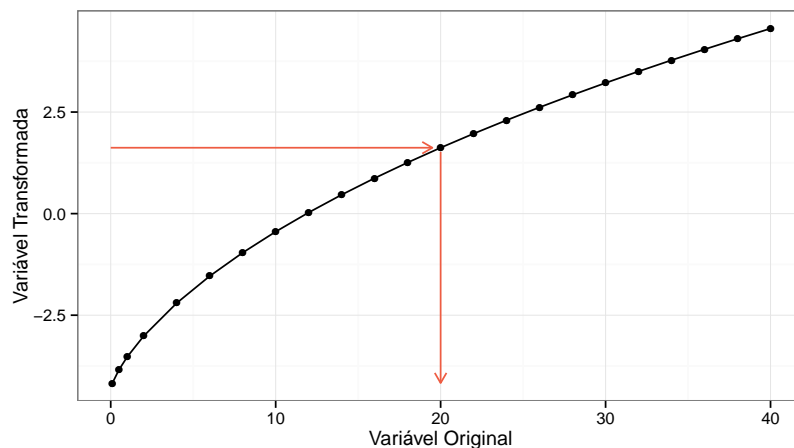
Vale lembrar que os lags espaço-temporais WZ_t são obtidos pelo valor que forma a média da vizinhança por causa da matriz W padronizada. Nesta aplicação, o valor de Z_t é influenciado positivamente pelo que ocorreu no lag espaço-temporal $t - 1$. No entanto, as ações policiais na vizinhança terão efeito negativo após 2 meses, o que indicaria a migração do crime para as regiões vizinhas. Isso quer dizer que uma região

poderá experimentar um aumento de roubo caso tenha ocorrido uma ação policial ostensiva na vizinhança em $t - 2$. Essa indicação do modelo pode expor que as ações tem sido de curto prazo e locais, o que permite o espalhamento nas regiões vizinhas, da mesma forma que um vírus se espalha quando não há vigilância permanente.

O modelo STARMA foi construído a partir da transformação *Deviance Residual* para cada um dos distritos e uma outra forma de visualizar o resultado estimado seria o consolidado dessas 93 unidades e na unidade de taxa de roubo. Uma forma de fazer isso seria calcular a inversa da função (4.3), no entanto não é possível fazer isso de forma direta já que trata-se de uma função não-linear. Por isso, o retorno para variável original foi obtido por um algoritmo de busca indicado pela Figura 14.

Este algoritmo consiste em gerar uma grande sequência numérica da variável original x_i que representa a taxa de roubo entre um intervalo mínimo e máximo estabelecido e com isso obter a respectiva transformação *Deviance Residual* para formar os pares $(x_i, f_{DR}(x_i))$. Por fim, comparou-se o valor estimado pelo STARMA representado por Z_{dr} com cada um dos valores de $f_{DR}(x_i)$, desta forma foi usado o x_i que minimizava a distância $|Z_{dr} - f_{DR}(x_i)|$. O algoritmo foi repetido para cada um dos distritos utilizando suas respectivas médias e variâncias.

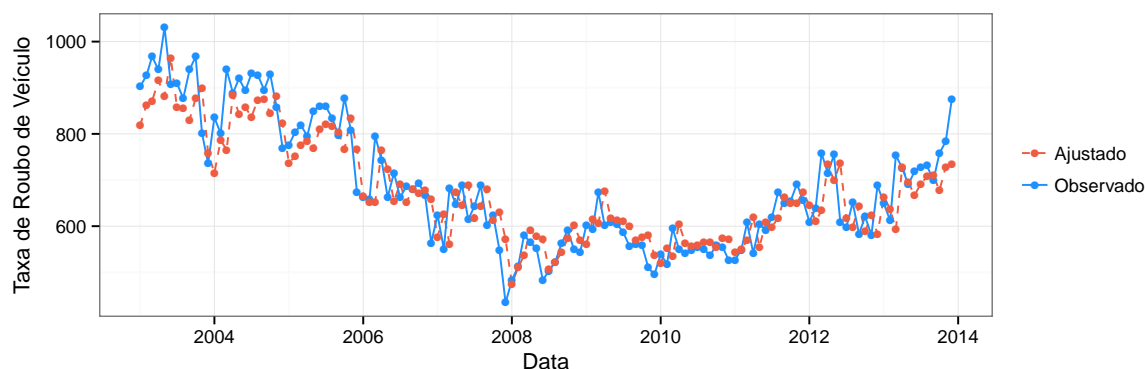
Figura 14 – Transformação *Deviance Residual*



Os resultados desta aplicação são apresentados nas Figuras 44-48 com os valores reais e os estimados em cada um dos 93 distritos e convertidos para taxa de roubo por milhão de veículos. Os resultados indicam que os valores estimados acompanham bem os movimentos reais, com exceção do distrito policial Jardim Miriam, cujo aumento da taxa de roubo foi a maior em comparação a todos os outros distritos.

A Figura 15 mostra a série da taxa de roubo e os valores estimados pelo modelo STARMA em um nível agregado para o município de São Paulo. Na análise de erro entre o valor observado e ajustado nota-se que 80% dos dados estão no intervalo entre -9% e $+8\%$. Ainda que as séries individuais apresentem um ajuste satisfatório é possível observar um processo de subestimação que ocorreu entre os anos de 2002 e 2006 ao trazer do nível das delegacias policiais para o nível do município.

Figura 15 – Ajuste do modelo STARMA convertido para taxa de roubo



4.2 Gráficos de Controle do MCUSUM e MEWMA

Neste Capítulo será apresentada a determinação dos parâmetros de controle do MCUSUM e do MEWMA. A primeira etapa é definir o ARL que será utilizado para controlar o erro tipo I. O valor mais usual encontrado na literatura é de 370 dias, isso seria aproximadamente 1 alarme falso por ano. Esse número tem como origem a probabilidade de uma variável exceder 3 desvios-padrões (MONTGOMERY, 2009), ou seja, se Y é uma variável aleatória com distribuição Normal Padrão então $ARL = 1/P(|Y| > 3) \approx 1/0,0027 \approx 370$. Na ausência de literatura sobre este tema na criminalidade para dados mensais adotou-se um erro de 5% para a probabilidade de falso alarme, o que leva a um ARL_0 de 20 meses.

Quando esse erro conjunto é aplicado para cada um dos 93 distritos o valor individual é ainda menor. Considere que as probabilidades de erro sejam independentes para o propósito de uma estimativa individual e que pode ser obtida pela expressão $\alpha_{global} = 1 - (1 - \alpha_i)^{93}$, com α_i a probabilidade de falso alarme no i -ésimo distrito. Ao resolver a equação chega-se com $\alpha_i = 1 - (1 - \alpha_{global})^{1/93}$, portanto a probabilidade individual de falso alarme em cada distrito policial é aproximadamente 0,055%.

Com o valor do ARL_0 fixado a segunda etapa consiste em um procedimento de otimização a fim de determinar os parâmetros (k, h) do MCUSUM:

$$\arg \min_{(k,h)} |ARL_{(k,h)} - ARL_0|$$

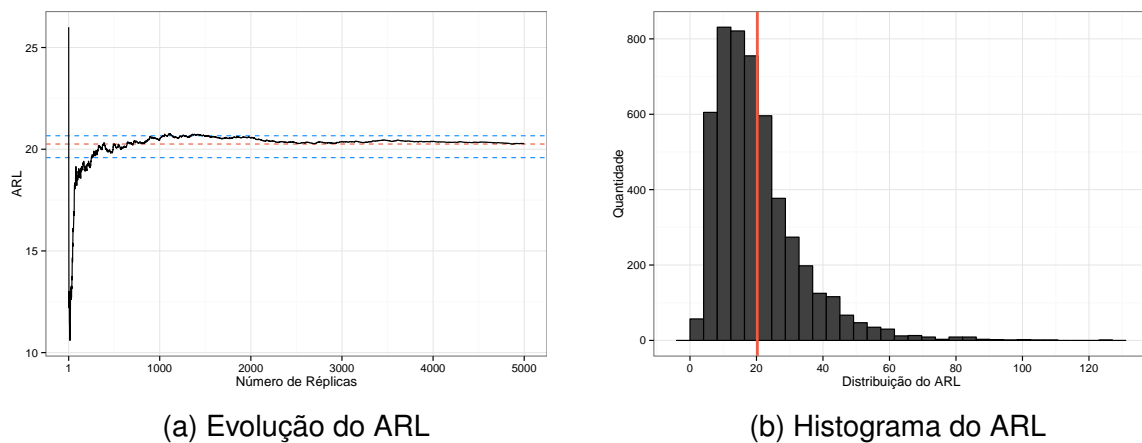
E para o MEWMA em função de (λ, b) é:

$$\arg \min_{(\lambda,b)} |ARL_{(\lambda,b)} - ARL_0|$$

Para o MCUSUM os valores de k foram avaliados no intervalo de 0,10 até 1,00 em intervalos de 0,10 e os valores de h foram avaliados entre 0,10 e 100 com espaça-

mento não uniforme para fazer uma varredura inicial rápida. Nessa busca se o valor do ARL simulado alcançava uma média de 30 nas primeiras mil simulações o processo era interrompido por ter informação suficiente nos pontos anteriores. Inclusive, de acordo com as Figuras 16a e 16b, nota-se que a convergência do ARL estabiliza em torno da média entre 4 mil e 5 mil réplicas. Como é necessário avaliar uma quantidade grande de parâmetros na busca pelo ARL_0 e para não tornar a simulação mais onerosa foi adotado 5 mil réplicas de *run length* para o obter o MCUSUM e o MEWMA.

Figura 16 – Avaliação do números de réplicas para obter a distribuição do ARL



Na segunda etapa os valores de h e do respectivo ARL , fixado o valor de k , foram usados para obter um spline que pudesse estimar o melhor valor de h dado um $ARL = 20$. Essa estimativa pontual não terá precisão de 100% se for utilizada uma nova simulação, por isso foi gerado um novo intervalo de variação para h com espaçamentos de 0,005 partindo de $0,98h$ até $1,02h$.

Para acelerar o processamento dos 5 métodos de gráficos de controle o software *RStudio* foi instalado nos servidores de nível gratuito da *Amazon Web Services* (SCHMIDBERGER, 2015).

O Algoritmo 2 apresenta a estrutura para otimizar os parâmetros k e h do MCUSUM, mas ele pode ser replicado para o MEWMA fazendo a troca de k por λ , h por b

Algoritmo 2: Cálculo do ARL_0 para o MCSUM.

```

1  Inicializa a faixa de variação de  $k$  e de  $h$ ;
2  forall  $k_i$  do
3      forall  $h_j$  do
4          Inicializa a estatística  $C_t$ ;
5          for  $r = 1$  to 5.000 do
6              repeat
7                  Simula erros normais com média  $\mu_0$  para obter  $Z_t$ ;
8                  until  $C_t > h_i$ ;
9                   $RL[r] \leftarrow t$ ;
10             end
11              $ARL \leftarrow Média(RL)$ ;
12         end
13         Ajusta-se um spline para os pares  $(ARL_{h_j}, h_j)$  tal que  $h_j = f_{spline}(ARL_{h_j})$ ;
14         Obtém o novo  $h$  utilizando o spline em  $ARL = 20$ ,  $h_{spline} = f_{spline}(20)$ ;
15         forall  $v \in \{0,98; 0,9805; \dots; 1,015; 1,02\}$  do
16             for réplicas = 1 to 5.000 do
17                 repeat
18                     Simula erros normais com média  $\mu_0$  para obter  $Z_t$ ;
19                     until  $C_t > v h_{spline}$ ;
20                      $RL[r] \leftarrow t$ ;
21                 end
22                  $ARL \leftarrow Média(RL)$ ;
23                 Armazena os novos ARLs obtidos dos pares  $(k_i, v h_{spline})$ ;
24             end
25         end
26 Seleciona os parâmetros  $\{(k_i, h_i) \mid -0,2 < ARL - 20 < 0,2\}$ ;

```

e de C_t por E_t .

Após esta simulação apenas os pares (k, h) cuja diferença do ARL_0 era inferior a 0,2 foram usados para estimar o ARL_1 . Nesta simulação do ARL_1 foram considerados

shifts com múltiplos do desvio padrão sob a média do modelo STARMA, a variância continua inalterada. Lembrando que $\varepsilon_t \sim N(\mu_0, \sigma^2 \mathbf{I}_N)$, então para cada shift de ordem δ a nova média é deslocada de forma que $\mu_1 = \mu_0 + \delta(\sigma \mathbf{I}_N)$, com $\delta \geq 0$.

Algoritmo 3: Cálculo do ARL_1

```

1 forall pares  $(k_i, h_i)$  da simulação anterior do
2   forall  $\delta_u$  do
3     for  $r = 1$  to 5.000 do
4       repeat
5         Simula erros normais com média  $\mu_1 = \mu_0 + \delta_u(\sigma \mathbf{I}_N)$  para obter  $Z_t$ ;
6       until  $C_t > h_i$ ;
7        $RL[r] \leftarrow t$ ;
8     end
9     Calcula o ARL;
10  end
11 end

```

A otimização do MEWMA ocorreu para $0,05 \leq \lambda \leq 0,20$ em intervalos de 0,05 e $25 \leq h \leq 500$ em intervalos de 25. No entanto, os valores do ARL_0 ficou muito distante de 20, por isso foi ajustado um spline no intervalo de h que continha o ARL_0 de 20 e em seguida iniciou-se uma nova busca varrendo este novo intervalo.

Ainda na estrutura do MEWMA foi proposta uma terceira simulação com uso da variância assintótica fazendo $t \rightarrow \infty$, o que leva $\sigma_{yt}^2 = \sigma^2 [\lambda/(2 - \lambda)]$, com finalidade de medir os efeitos nos parâmetros e também nos ARLs.

A Tabela 3 traz um resumo dos três métodos segundo os parâmetros que melhor aproximou o ARL_0 em 20 meses, para o MCUSUM foram $h = 2,16$ e $k = 0,7$ e nas duas versões do MEWMA foram $b = 213,06$ e $\lambda = 0,2$. Quando $\delta = 0$ encontra-se o valor do ARL_0 e as demais linhas correspondem ao ARL_1 para um *shift* na

média e a coluna DP indica o desvio padrão do ARL . Nota-se que o desvio padrão no ARL_0 foi menor no MCUSUM, já os desvios para $\delta > 0$ alternaram-se entre os métodos. Como destaque, o MCUSUM apresentou capacidade de sinalizar um alerta em 1 passo quando $\delta \geq 1$. Os resultados detalhados das simulações com seleção dos melhores parâmetros após aplicação do Algoritmo 2 encontram-se nas Tabelas 5, 6 e 7.

Tabela 3 – ARL para MCUSUM e MEWMA - segundo o modelo (4.7)

δ	MEWMA				MCUSUM	
	Exato		Assintótico			
	ARL	DP	ARL	DP	ARL	DP
0	20,10	17,18	20,10	17,18	19,88	12,36
0,25	4,11	0,64	4,67	0,59	3,96	1,03
0,50	2,76	0,44	3,24	0,43	2,22	0,48
0,75	2,01	0,11	2,94	0,23	1,57	0,50
1,00	1,92	0,28	2,06	0,24	1,04	0,19
1,25	1,22	0,42	2,00	$< 10^{-2}$	1,00	$< 10^{-2}$
1,50	1,00	0,04	2,00	$< 10^{-2}$	1,00	$< 10^{-2}$
1,75	1,00	$< 10^{-2}$	2,00	$< 10^{-2}$	1,00	$< 10^{-2}$
2,00	1,00	$< 10^{-2}$	2,00	0,03	1,00	$< 10^{-2}$
b ou h	213,06		213,06		2,16	
λ ou k	0,2		0,2		0,70	

Na próxima simulação foi avaliado o efeito no ARL quando as componentes espaciais do modelo STARMA são omitidas, pois o mais usual é considerar modelos puramente temporais como os modelos ARMA. Nesta avaliação os parâmetros do MCUSUM e do MEWMA foram obtidos fixando o $ARL_0 = 20$ com o STARMA $\hat{Z}_t = 0,80409\mathbf{Z}_{t-1} + 0,04687\mathbf{Z}_{t-12} - 0,51406\epsilon_{t-1} + \epsilon_t$. Mas, o ARL_1 foi calculado utilizando o modelo STARMA com a presença das componentes espaciais. Nota-se, de

acordo com a Tabela 4, que os ARLs do MCUSUM e do MEWMA ficaram abaixo de 20 meses para $\delta = 0$, esse cenário indica a sinalização de falso alarme 7 meses antes do esperado para o MCUSUM e de 2,3 meses antes para o MEWMA. Em termos probabilísticos o erro tipo I que antes foi fixado em 5% agora é de 7,6% para o MCUSUM e 5,7% para o MEWMA. As Tabelas 8 e 9 possuem os valores de ARLs para os parâmetros que foram selecionados como potenciais.

Tabela 4 – *ARL* para MCUSUM e MEWMA - STARMA sem componentes espaciais.

δ	MCUSUM		MWMA	
	ARL	DP	ARL	DP
0,00	13,09	7,50	17,68	14,68
0,25	3,57	0,97	3,94	0,62
0,50	2,03	0,48	2,64	0,48
0,75	1,35	0,48	2,00	0,07
1,00	1,01	0,11	1,84	0,37
1,25	1,00	$< 10^{-2}$	1,13	0,33
1,50	1,00	$< 10^{-2}$	1,00	0,03
1,75	1,00	$< 10^{-2}$	1,00	$< 10^{-2}$
2,00	1,00	$< 10^{-2}$	1,00	$< 10^{-2}$
b ou h	1,72		202,33	
λ ou k	0,70		0,20	

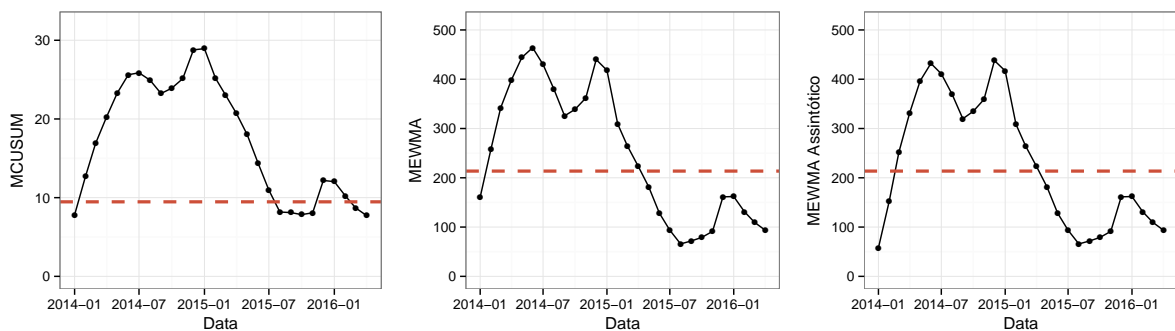
Quando esses procedimentos de controle de qualidade são utilizados na indústria e após uma amostragem detecta-se uma medida fora de controle pode-se optar por interromper o processo de produção para investigar em qual ponto o problema ocorreu. Agora, nesta aplicação com a criminalidade não há o que ser interrompido, os crimes continuarão acontecendo. Uma decisão imediata seria rever a logística de patrulhamento a fim de reforçar as atividades policiais nos distritos em que a sinalização ocorreu e isso deve ser realizado com cautela dependendo da magnitude da sinali-

zação da estatística de controle. Lembrando do efeito vírus que foi comentado no Capítulo 4.1 de estimação do modelo STARMA, a ação em uma região pode provocar a migração para regiões vizinhas, o cenário requer mais cautela quando ocorre um falso alarme, pois isso pode provocar um aumento real da criminalidade logo após a mudança logística na região que sofreu algum esvaziamento policial.

4.3 Aplicação do Gráfico de Controle

A aplicação do gráfico de controle será feita no período que vai de janeiro de 2014 até abril de 2016. O *MCUSUM*, o *MEWMA* padrão e o assintótico são apresentados na Figura 17.

Figura 17 – Gráficos de Controle *MCUSUM* e *MEWMA* entre jan/14 e abr/16.



O *MCUSUM* e o *MEWMA* já indicam um alerta no segundo mês, enquanto que o *MEWMA* Assintótico demorou um mês a mais. Após 9 meses o *MEWMA* padrão e o assintótico convergem para os mesmos resultados, o que seria esperado para t grande. No entanto, apenas o *MCUSUM* indicou novos alertas em dezembro de 2015.

Para identificar quais os distritos tiveram maiores aumentos foram construídos dois mapas, Figuras 18 e 19, para o *MCUSUM* foram utilizados os valores das componentes do vetor S_t (Eq. 3.23) e para o *MEWMA* foram utilizadas as componentes do vetor Y_t (Eq. 3.18). Ambos os mapas apontam que os distritos da periferia foram os que mais influenciaram os resultados do *MCUSUM* e do *MEWMA* com destaque para Jardim Arpoador na zona oeste, Jardim Miriam na zona sul, Parque do Carmo, Parque São Rafael e Cidade Tiradentes localizados na zona leste e Parada de Taipas e Vila Pereira Barreto na zona norte.

Essa sinalização de aumento parece contraditória quando analise-se a taxa geral de roubo da Figura 9 que apresenta uma queda, no entanto isso decorre do método

que padroniza todos os distritos e os deixam com a mesma importância.

Figura 18 – Resultado da aplicação do *MCUSUM*

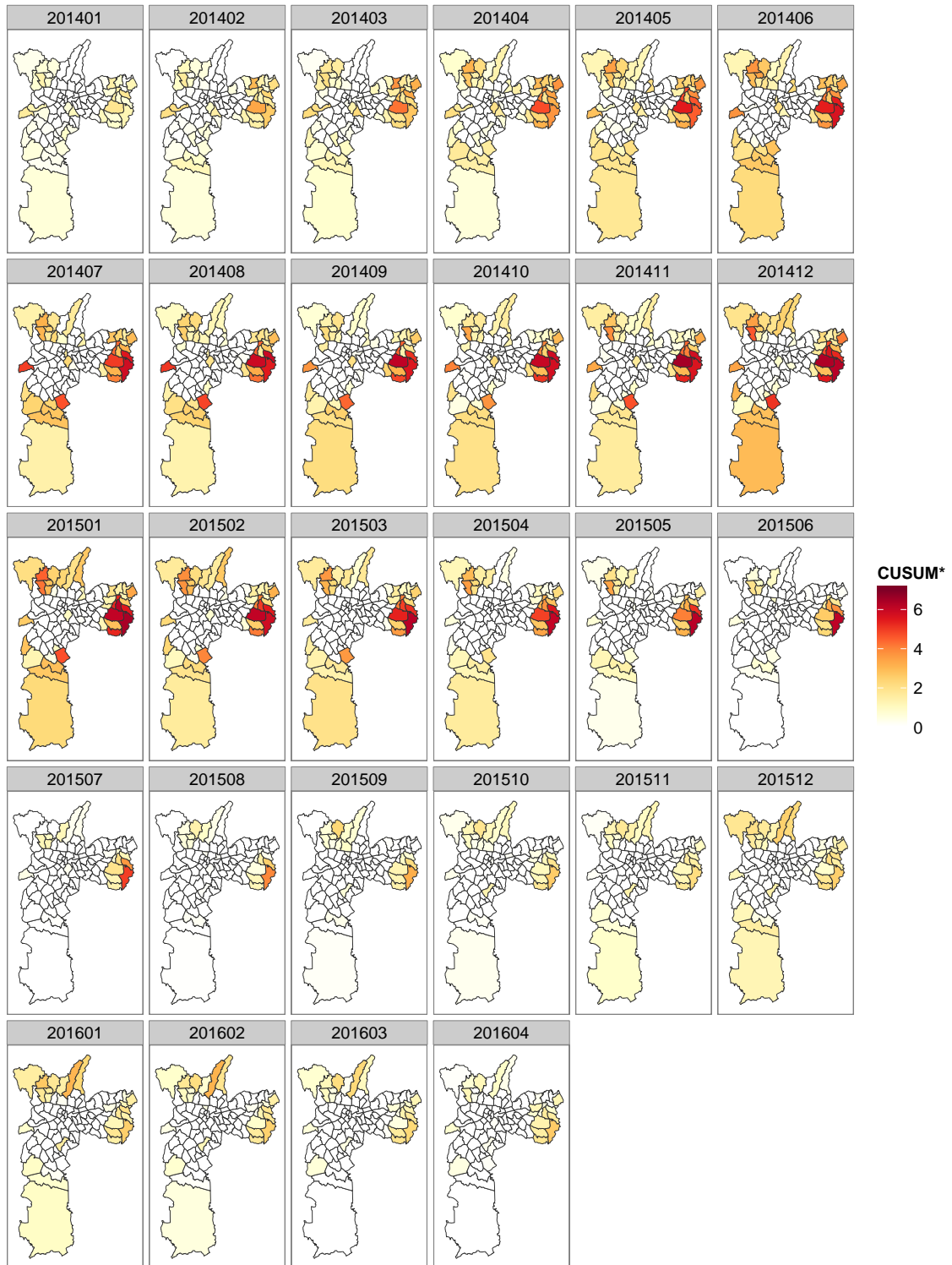
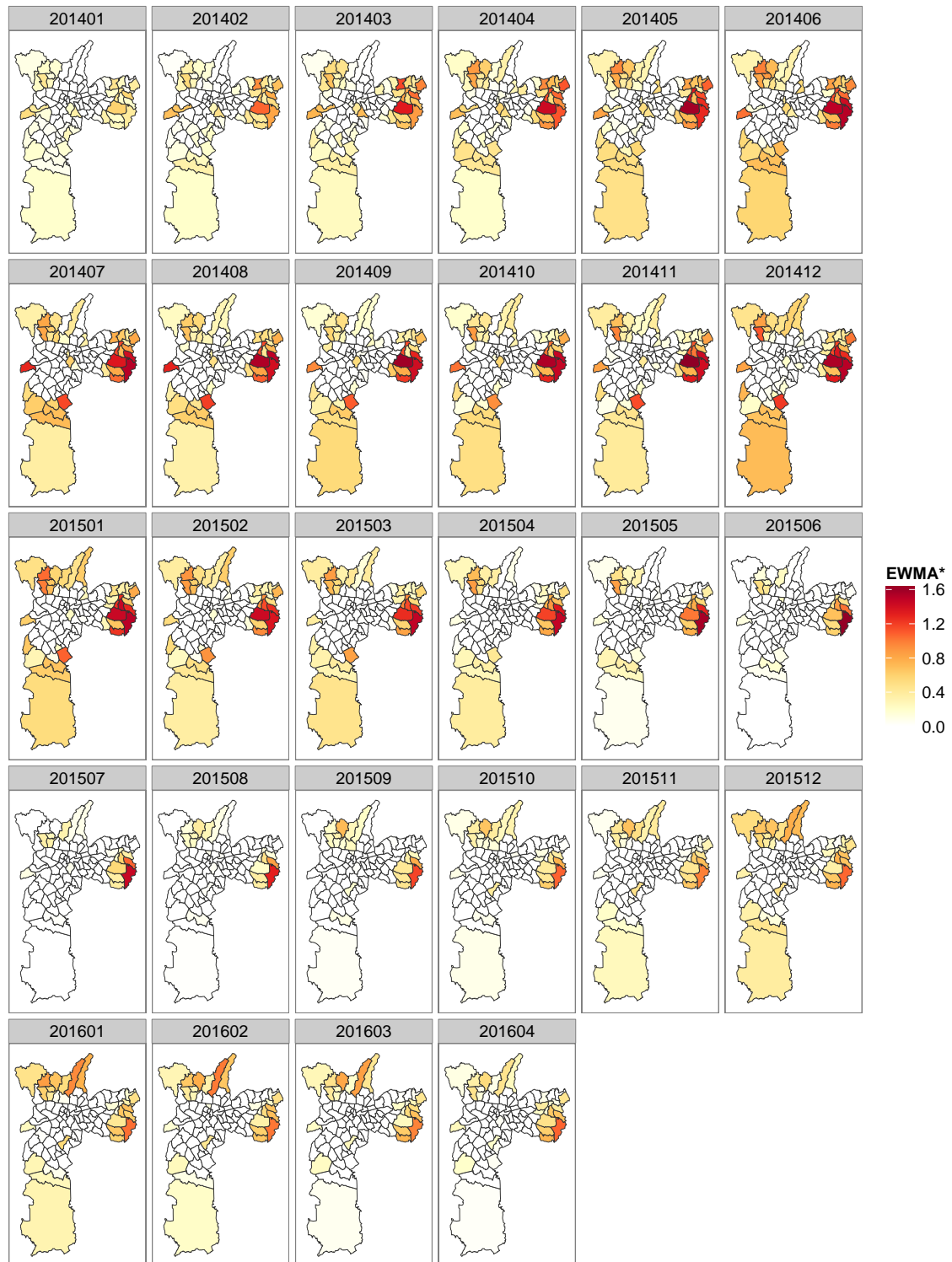


Figura 19 – Resultado da aplicação do *MEWMA*



5 Conclusões

O ajuste do modelo STARMA, por sua generalidade, apresentou um bom ajuste na maioria das localidades. Isso porque o modelo tem apenas 6 parâmetros e foi capaz de explicar grande parte da variabilidade nas 93 regiões. O modelo final incorporou sazonalidades anuais, além de forte dependência com o mês anterior. A vantagem do STARMA é o uso das autocorrelações e autocorrelações parciais para identificar o modelo mais apropriado com base na correlação residual.

No problema em questão o uso dos indicadores direcionais poupam tempo ao trabalhar apenas com aumento da taxa de roubo. No período de modelagem o *MCUSUM* foi a estatística que sinalizou alarmes mais cedo, embora no período de validação ambos tenham emitido alerta já no segundo mês foi apenas o *MCUSUM* que voltou a emitir alertas no final do ano de 2015.

Um destaque importante foi a redução do crime nas áreas mais nobres da cidade, que envolve praticamente a mesma área da zona de rodízio de veículo, o aumento simultâneo do roubo nas zonas periféricas e o aumento substancial no bairro Jardim Miriam, que fica localizado na zona sul da cidade e que faz divisa com a cidade de Diadema. Também destaca-se o aumento no bairro Cidade Tiradentes localizado no extremo leste da cidade. O modelo STARMA já havia capturado esse tipo de comportamento viral do crime, em que ações localizadas de repressão poderiam causar aumentos de roubos em regiões vizinhas sem o mesmo tipo de combate.

O processo de otimização dos parâmetros é o ponto computacional mais oneroso, na tentativa de reduzir o espaço de busca optou-se uma varredura inicial mais esparsa e depois por interpolação dos pontos foi estabelecida uma região mais detalhada. O desenvolvimento desse algoritmo ajudou ao refinar a região de busca e a convergência

do alvo de 20 meses definido a priori foi alcançado.

Ainda sob a ótica computacional o uso de servidores da Amazon Web Services para executar as simulações em paralelo ajudou a acelerar o processo, mas uma abordagem diferente ter surtido um efeito promissor nos problemas de Inteligência Artificial que é a utilização de GPUs (*Graphics Processing Units*). Elas tem sido usadas para acelerar a estimação dos parâmetros de redes neurais devido à grande capacidade de processamento em paralelo e o mesmo poderia ser aplicado na busca dos parâmetros de controle. Outra forma de aumentar o paralelismo é na simulação dos modelos STARMA, como a distribuição do ARL tem uma convergência lenta são necessárias muitas réplicas e o uso dessa abordagem ajudaria a aumentar o número de réplicas e reduzir o tempo total da simulação.

Outro ponto de melhoria é a expansão da região de vizinhança para toda a zona metropolitana de São Paulo. A inclusão do IDH ou de informações de renda que possam denotar locais de vulnerabilidade social também ajudariam a reduzir o erro do modelo STARMA.

Uma sugestão que pode direcionar ainda mais a logística policial seria ter a informação georeferenciada de cada um dos crimes em nível diário ou semanal ao invés de mensal.

A maior dificuldade encontrada neste trabalho foi a adequação dos dados, que originalmente vem de uma contagem, para um modelo com distribuição Normal nos erros. Um avanço seria expandir o modelo da classe *GARMA* (*Generalized Autoregressive Moving Average*), (BENJAMIN; RIGBY; STASINOPOULOS, 2003), para tratar dados de contagem com dependência espaço-temporal e utilizar o tamanho da frota como *off-set* (PAULA, 2013).

REFERÊNCIAS

- ALENCAR, A. P.; HO, L. L.; ALBARRACÍN, O. Y. E. Cusum control charts to monitor series of negative binomial count data. *Statistical Methods in Medical Research*, 2015.
- ALWAN, L. C.; ROBERTS, H. V. Time-series modeling for statistical process control. *Journal of Business and Economic Statistics*, v. 6, n. 1, p. 87, 1988.
- ANDERSON, E. A.; DIAZ, J. Using process control chart techniques to analyse crime rates in houston, texas. *Journal of the Operational Research Society*, JSTOR, p. 871–881, 1996.
- ANSELIN, L. *Spatial Econometrics: Methods and Models*. [S.l.]: Springer Netherlands, 1988.
- ANSELIN, L. Spatial econometrics. *A Companion to Theoretical Econometrics*, p. 310–330, 2003.
- BENJAMIN, M. A.; RIGBY, R. A.; STASINOPOULOS, D. M. Generalized autoregressive moving average models. *Journal of the American Statistical Association*, v. 98, n. 461, p. 214–223, 2003.
- BISSELL, A. F. Cusum techniques for quality control (with discussion). *Applied Statistics*, v. 18, p. 1–30, 1969.
- BODNAR, O.; SCHMID, W. Cusum control schemes for multivariate time series. In: _____. *Frontiers in Statistical Quality Control 8*. [S.l.]: Springer, 2006. p. 55–73.
- BODNAR, O.; SCHMID, W. Surveillance of the mean behavior of multivariate time series. *Statistica Neerlandica*, Blackwell Publishing Ltd, v. 61, n. 4, p. 383–406, 2007.
- BROOK, D.; EVANS, D. A. An approach to the probability distribution of cusum run length. *Biometrika*, [Oxford University Press, Biometrika Trust], v. 59, n. 3, p. 539–549, 1972.
- CERQUEIRA, D. et al. Análise dos custos e consequências da violência no brasil. *IPEA, Instituto de Pesquisa Econômica Aplicada*, 2007.
- CLIFF, A. D.; ORD, J. K. Space-time modelling with an application to regional forecasting. *Transactions of the Institute of British Geographers*, The Royal Geographical Society (with the Institute of British Geographers), n. 64, p. pp. 119–128, 1975.
- CONSTANTIN, S.; DOXARAN, D.; CONSTANTINESCU, S. Estimation of water turbidity and analysis of its spatio-temporal variability in the danube river plume (black sea) using modis satellite data. *Continental Shelf Research*, Elsevier, v. 112, p. 14–30, 2016.
- CRESSIE, N. A. C. *Statistics for spatial data*. [S.l.]: Wiley, 1993. 449-552 p.
- CROSSIER, R. B. Multivariate generalizations of cumulative sum quality-control schemes. *Technometrics*, v. 30, n. 3, p. 291, 1988.

DEUTSCH, S. J.; PFEIFER, P. E. Space-time arma modeling with contemporaneously correlated innovations. *Technometrics*, v. 23, n. 4, p. pp. 401–409, 1981.

DUBÉ, J.; LEGROS, D. A spatio-temporal measure of spatial dependence: An example using real estate data. *Papers in Regional Science*, Wiley Online Library, v. 92, n. 1, p. 19–30, 2013.

EWAN, W. D.; KEMP, K. W. Sampling inspection of continuous processes with no autocorrelation between successive results. *Biometrika*, [Oxford University Press, Biometrika Trust], v. 47, n. 3/4, p. 363–380, 1960.

FRICKER, R. D.; KNITT, M. C.; HU, C. X. Comparing directionally sensitive mcusum and mewma procedures with application to biosurveillance. *Quality Engineering*, v. 20, n. 4, p. 478–494, 2008.

GETIS, A.; ALDSTADT, J. Constructing the spatial weights matrix using a local statistic. *Geographical Analysis*, Blackwell Publishing Ltd, v. 36, n. 2, p. 90–104, 2004.

HAWKINS, D. M.; OLWELL, D. H. *Cumulative sum charts and charting for quality improvement*. [S.I.]: Springer, 1998.

HAWKINS, D. M.; WU, Q. The cusum and the ewma head-to-head. *Quality Engineering*, v. 26, n. 2, p. 215–222, 2014.

HEALY, J. D. A note on multivariate cusum procedures. *Technometrics*, v. 29, n. 4, p. 409, 1987.

HOLLY, S.; PESARAN, M. H.; YAMAGATA, T. A spatio-temporal model of house prices in the usa. *Journal of Econometrics*, v. 158, n. 1, p. 160–173, 2010.

JACKSON, J. E. Quality control methods for several related variables. *Technometrics*, [Taylor & Francis, Ltd., American Statistical Association, American Society for Quality], v. 1, n. 4, p. 359–377, 1959.

JACKSON, J. E.; MORRIS, R. H. An application of multivariate quality control to photographic processing. *Journal of the American Statistical Association*, [American Statistical Association, Taylor & Francis, Ltd.], v. 52, n. 278, p. 186–199, 1957.

JACKSON, J. E.; MUDHOLKAR, G. S. Control procedures for residuals associated with principal component analysis. *Technometrics*, [Taylor & Francis, Ltd., American Statistical Association, American Society for Quality], v. 21, n. 3, p. 341–349, 1979.

JONER, M. D. et al. A one-sided mewma chart for health surveillance. *Qual. Reliab. Engng. Int.*, v. 24, n. 5, p. 503–518, 2008.

KANTAKUMAR, L. N.; KUMAR, S.; SCHNEIDER, K. Spatio temporal urban expansion in pune metropolis, india using remote sensing. *Habitat International*, Elsevier, v. 51, p. 11–22, 2016.

KAZAR, B. M.; CELIK, M. *Spatial AutoRegression (SAR) model*. [S.I.]: Springer, 2012.

KERRY, R. et al. Applying geostatistical analysis to crime data: Car-related thefts in the baltic states. *Geographical analysis*, Wiley Online Library, v. 42, n. 1, p. 53–77, 2010.

- LEE, L.; LIU, X. Efficient gmm estimation of high order spatial autoregressive models with autoregressive disturbances. *Econom. Theory*, v. 26, n. 01, p. 187, 2009.
- LOWRY, C. A. et al. A multivariate exponentially weighted moving average control chart. *Technometrics*, v. 34, n. 1, p. 46, 1992.
- LUCAS, J. M.; CROSIER, R. B. Fast initial response for cusum quality-control schemes: Give your cusum a head start. *Technometrics*, [Taylor & Francis, Ltd., American Statistical Association, American Society for Quality], v. 24, n. 3, p. 199–205, 1982.
- MAIUMUN, N. H. A. et al. Spatiotemporal autoregressive modelling of residential property prices. *International Journal of Real Estate Studies*, v. 7, n. 1, 2012.
- MONTGOMERY, D. C. *Introduction to statistical quality control*. [S.l.]: Wiley, 2009.
- MORAN, P. A. P. Notes on continuous stochastic phenomena. *Biometrika*, JSTOR, p. 17–23, 1950.
- MORETTIN, P. A.; TOLOI, C. M. C. *Análise de Séries Temporais*. [S.l.]: Edgard Blucher, 2006.
- NGAI, H. M.; ZHANG, J. Multivariate cumulative sum control charts based on projection pursuit. *Statistica Sinica*, Institute of Statistical Science, Academia Sinica, v. 11, n. 3, p. 747–766, 2001.
- ORD, K. Estimation methods for models of spatial interaction. *Journal of the American Statistical Association*, v. 70, n. 349, p. pp. 120–126, 1975.
- PAULA, G. A. *Modelos de Regressão com Apoio Computacional*. 2013. Disponível em: <https://www.ime.usp.br/~giapaula/texto_2013.pdf>.
- PFEIFER, P. E.; DEUTSCH, S. J. A starima model-building procedure with application to description and regional forecasting. *Transactions of the Institute of British Geographers*, v. 5, n. 3, p. pp. 330–349, 1980.
- PIGNATIELLO, J. J. J.; RUNGER, G. C. Comparison of multivariate cusum charts. *Journal of Quality Technology*, v. 22, n. 3, p. 173–186, 1990.
- PIROUTEK, A.; ASSUNÇÃO, R.; PAIVA, T. Space-time prospective surveillance based on knox local statistics. *Statistics in Medicine*, v. 33, n. 16, p. 2758–2773, 2014.
- SCHEEL, I. et al. A bayesian hierarchical model with spatial variable selection: the effect of weather on insurance claims. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, Blackwell Publishing Ltd, v. 62, n. 1, p. 85–100, 2013.
- SCHMIDBERGER, M. *Running R on AWS*. 2015. Disponível em: <<https://blogs.aws.amazon.com/bigdata/post/Tx3IJSB6BMHWZE5/Running-R-on-AWS>>.
- SMILEY, W. C.; KEOAGILE, T. Multivariate max-cusum chart. *Quality Technology & Quantitative Management*, v. 2, n. 2, p. 221–235, 2005.
- SUN, B.; ZHOU, S.; ZHAO, Q. Evaluation of spatial and temporal changes of soil quality based on geostatistical analysis in the hill region of subtropical china. *Geoderma*, Elsevier, v. 115, n. 1, p. 85–99, 2003.

SUSEP. *Autoseg*. 2015. Disponível em: <<http://www2.susep.gov.br/menuestatistica/Autoseg/principal.aspx>>.

Van Dobben de Bruyn, C. S. *Cumulative sum tests: theory and practice*. London: Griffin, 1968. (Griffin 's Stat. Mono. Courses).

WOODALL, W. H. The distribution of the run length of one-sided cusum procedures for continuous random variables. *Technometrics*, [Taylor & Francis, Ltd., American Statistical Association, American Society for Quality], v. 25, n. 3, p. 295–301, 1983.

WOODALL, W. H. On the markov chain approach to the two-sided cusum procedure. *Technometrics*, [Taylor & Francis, Ltd., American Statistical Association, American Society for Quality], v. 26, n. 1, p. 41–46, 1984.

WOODALL, W. H.; NCUBE, M. M. Multivariate cusum quality-control procedures. *Technometrics*, [Taylor & Francis, Ltd., American Statistical Association, American Society for Quality], v. 27, n. 3, p. 285–292, 1985.

WOODARD, J. D. et al. A spatial econometric analysis of loss experience in the u.s. crop insurance program. *Journal of Risk and Insurance*, Blackwell Publishing Inc, v. 79, n. 1, p. 261–286, 2012.

APÊNDICE A – Tabelas e Figuras

Os mapas entre as Figuras 20-33 apresentam a taxa de roubo por 1 milhão de veículos nos anos de 2002 até 2015. Em cada página são exibidos os 12 meses de cada ano no formato de AAAAMM. Ou seja, os quatro primeiros dígitos indicam o ano e os dois últimos indicam o mês.

Figura 20 – Taxa de Roubo de Veículos na Cidade de São Paulo em 2002

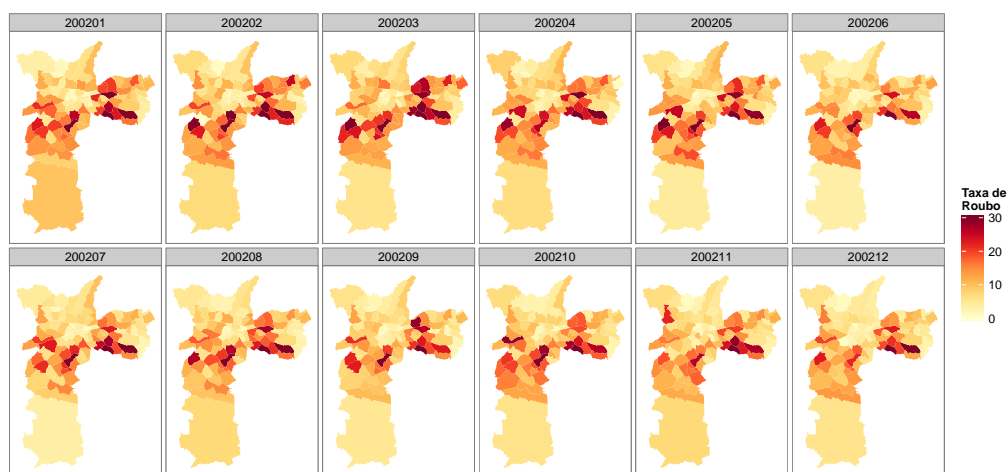


Figura 21 – Taxa de Roubo de Veículos na Cidade de São Paulo em 2003

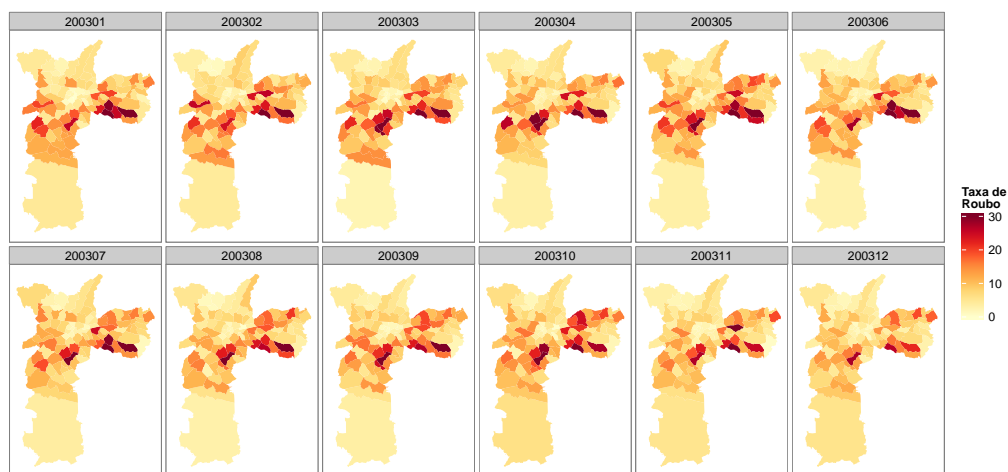


Figura 22 – Taxa de Roubo de Veículos na Cidade de São Paulo em 2004

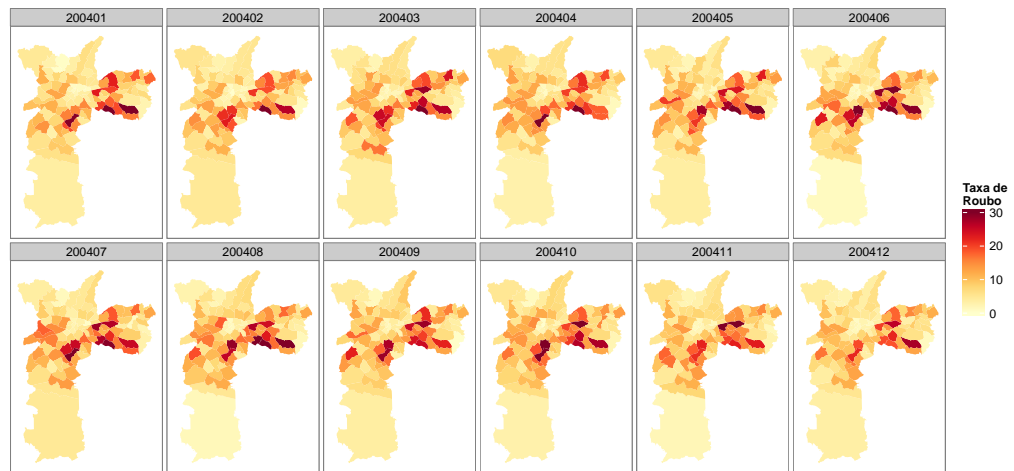


Figura 23 – Taxa de Roubo de Veículos na Cidade de São Paulo em 2005



Figura 24 – Taxa de Roubo de Veículos na Cidade de São Paulo em 2006

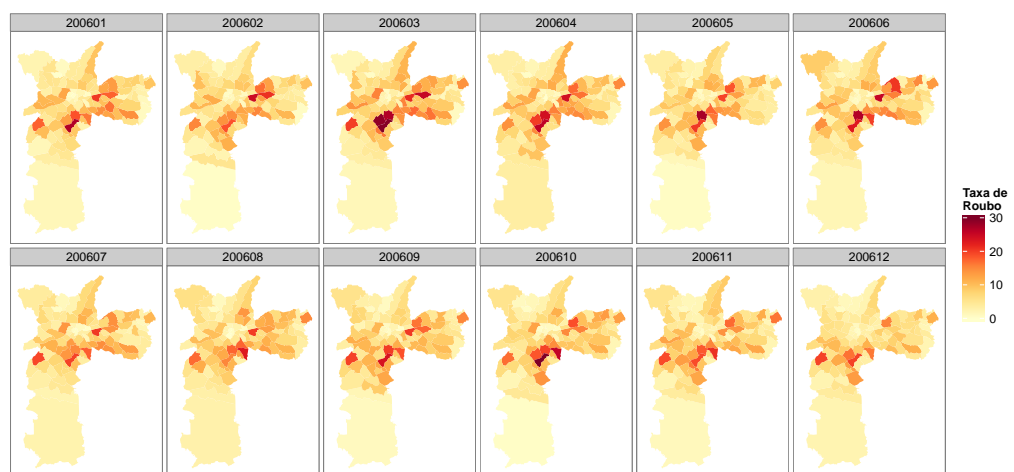


Figura 25 – Taxa de Roubo de Veículos na Cidade de São Paulo em 2007

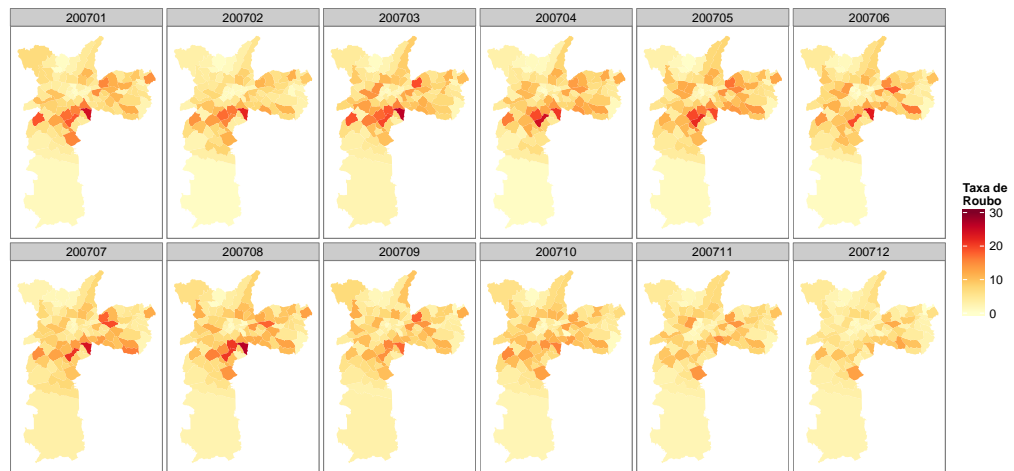


Figura 26 – Taxa de Roubo de Veículos na Cidade de São Paulo em 2008



Figura 27 – Taxa de Roubo de Veículos na Cidade de São Paulo em 2009

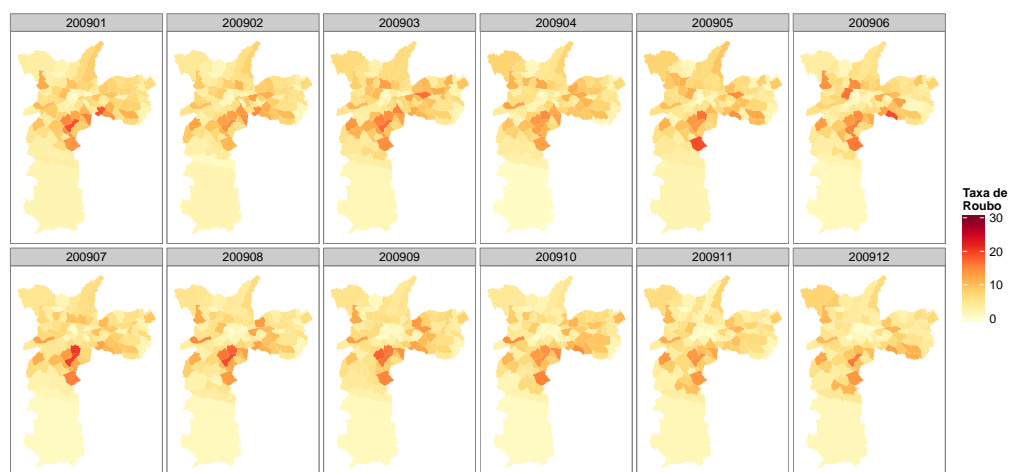


Figura 28 – Taxa de Roubo de Veículos na Cidade de São Paulo em 2010

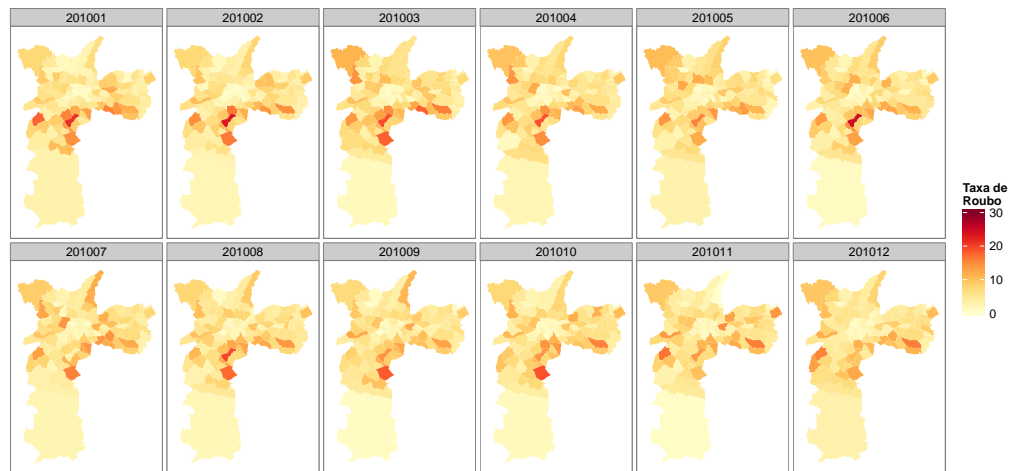


Figura 29 – Taxa de Roubo de Veículos na Cidade de São Paulo em 2011

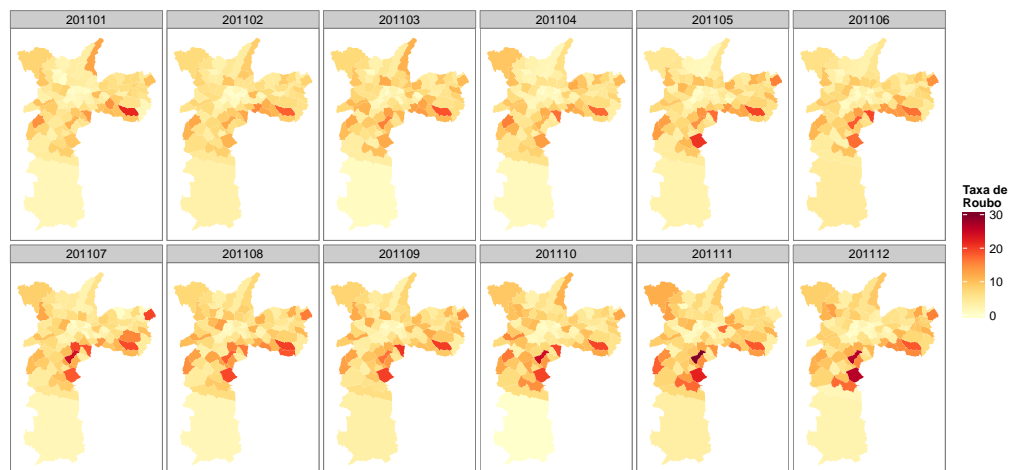


Figura 30 – Taxa de Roubo de Veículos na Cidade de São Paulo em 2012

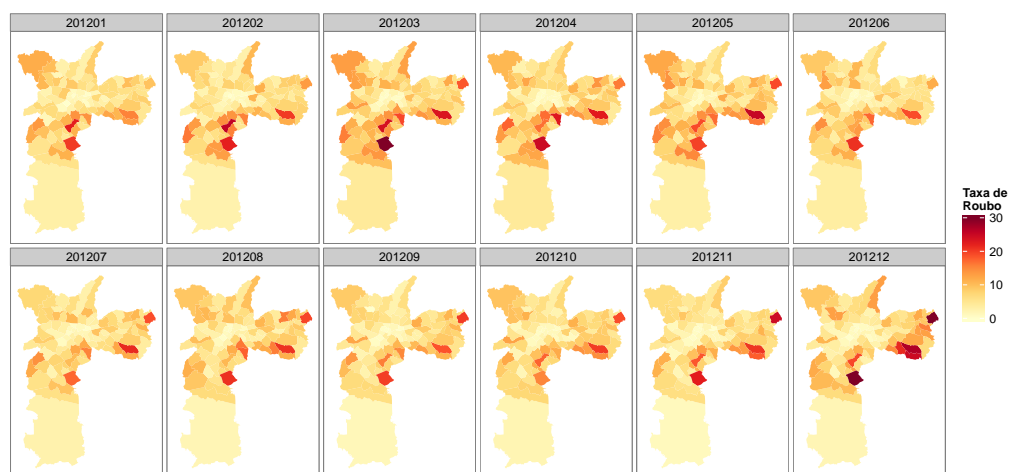


Figura 31 – Taxa de Roubo de Veículos na Cidade de São Paulo em 2013

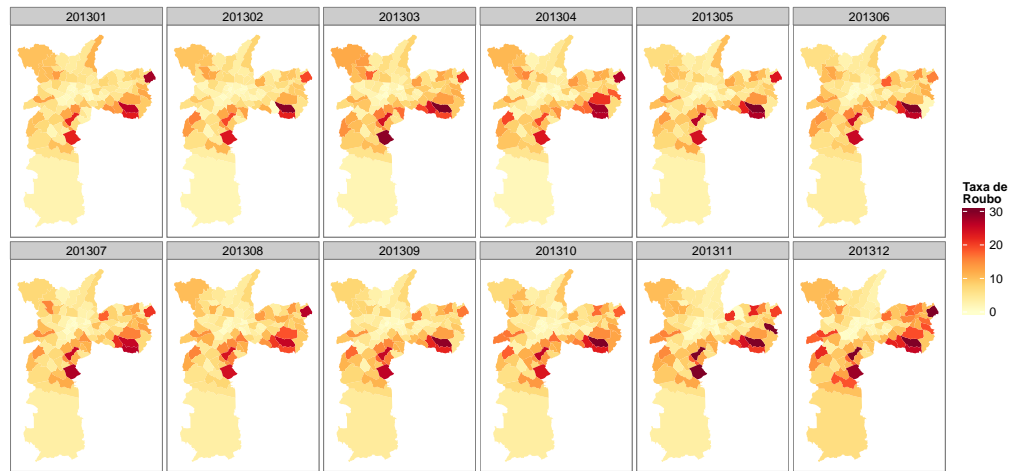


Figura 32 – Taxa de Roubo de Veículos na Cidade de São Paulo em 2014

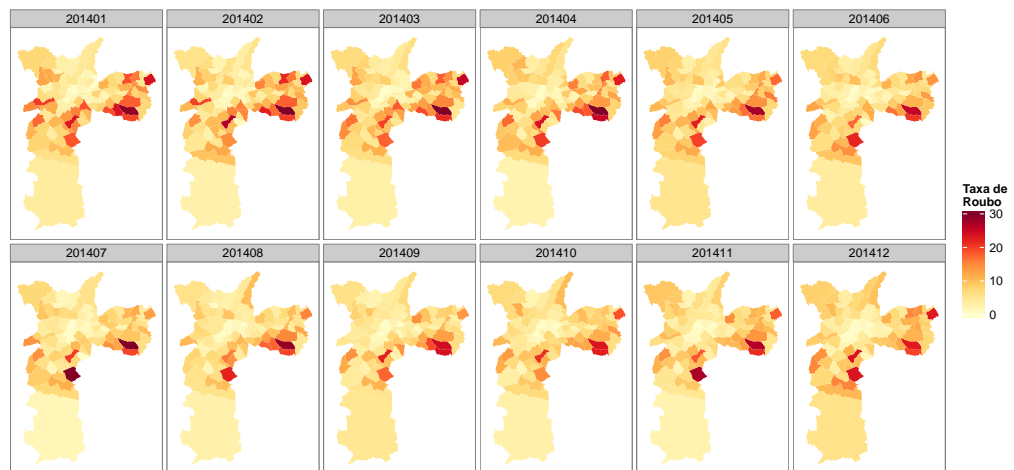


Figura 33 – Taxa de Roubo de Veículos na Cidade de São Paulo em 2015

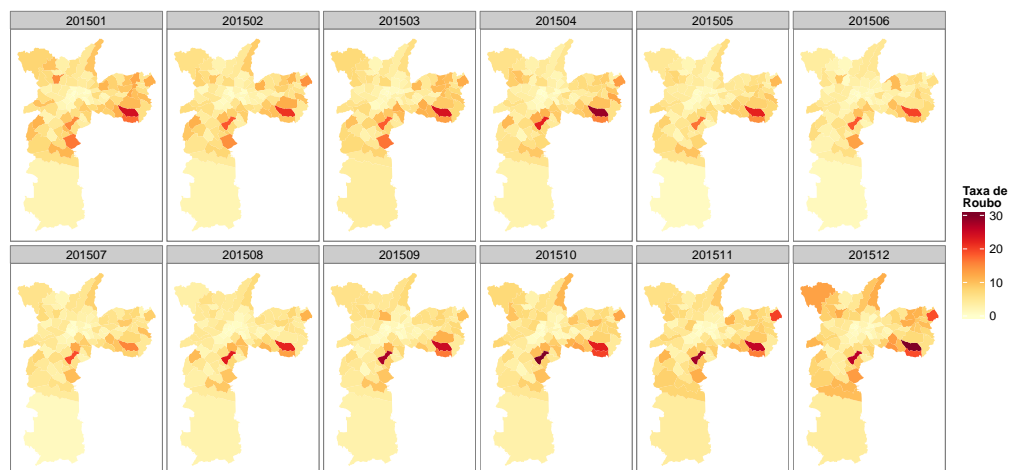


Figura 34 – STACF de Ordem Espacial 0, 1, 2 e 3

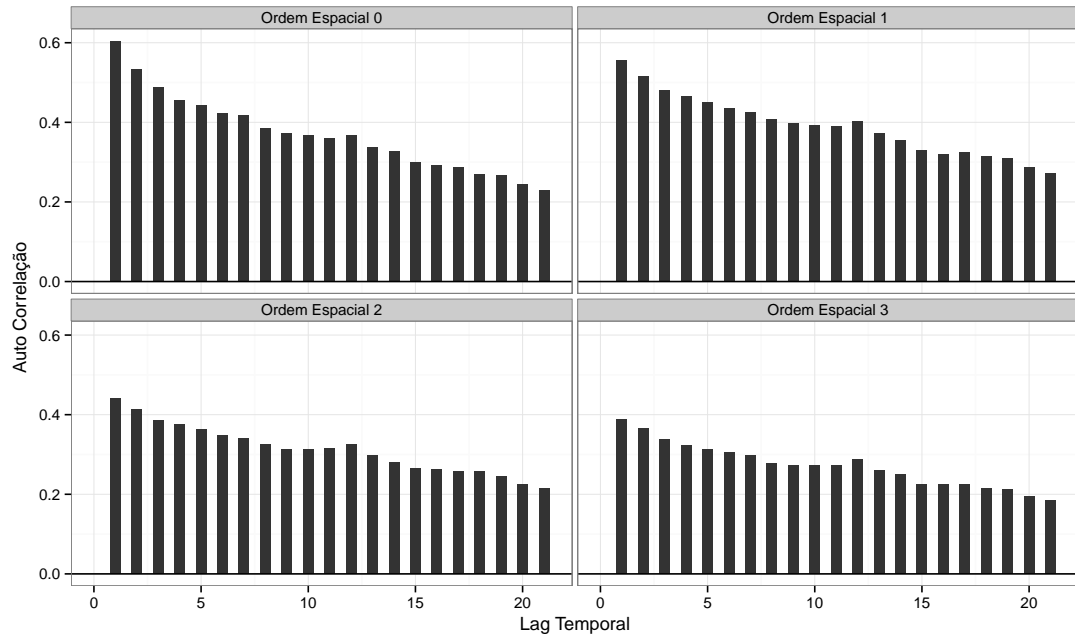


Figura 35 – STPACF de Ordem Espacial 0, 1, 2 e 3

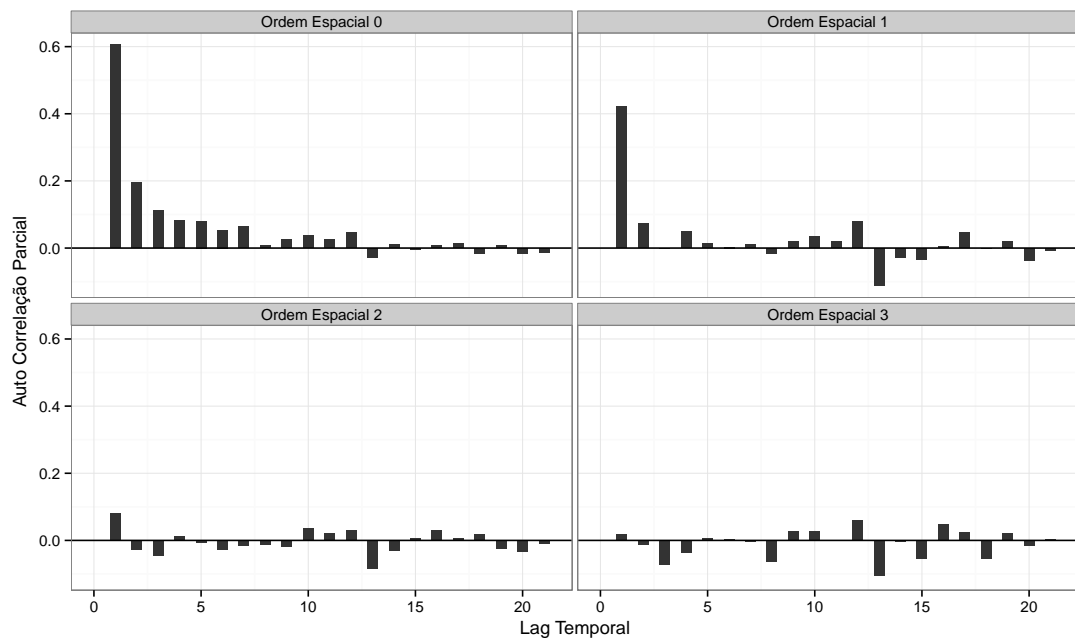


Figura 36 – STACF dos Resíduos do modelo STAR(1,1), nas ordens espaciais 0, 1, 2 e 3

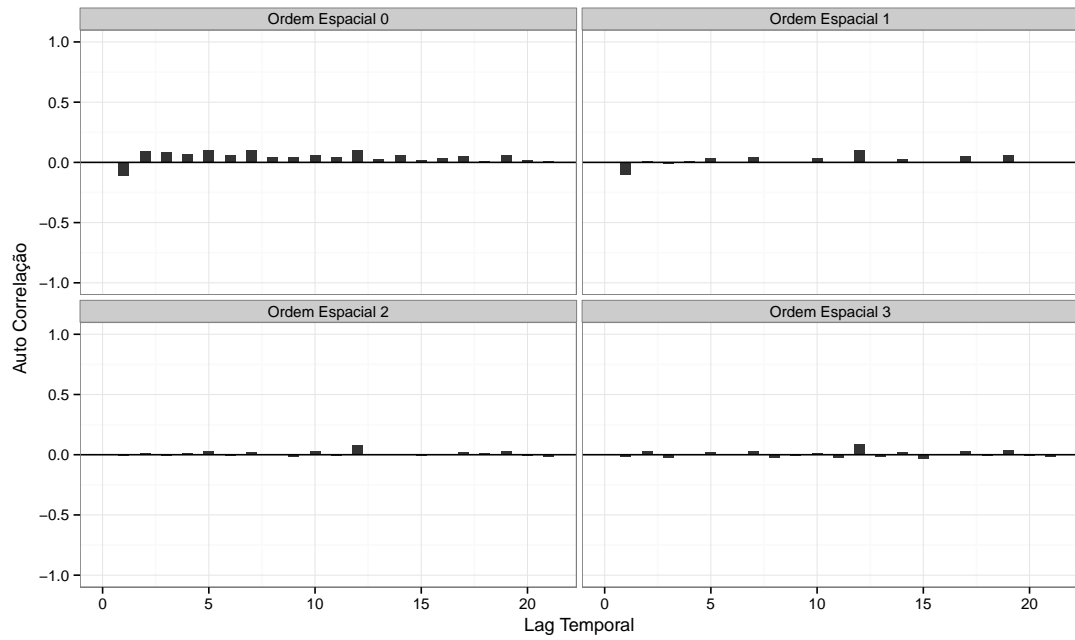


Figura 37 – STPACF dos Resíduos do modelo STAR(1,1), nas ordens espaciais 0, 1, 2 e 3

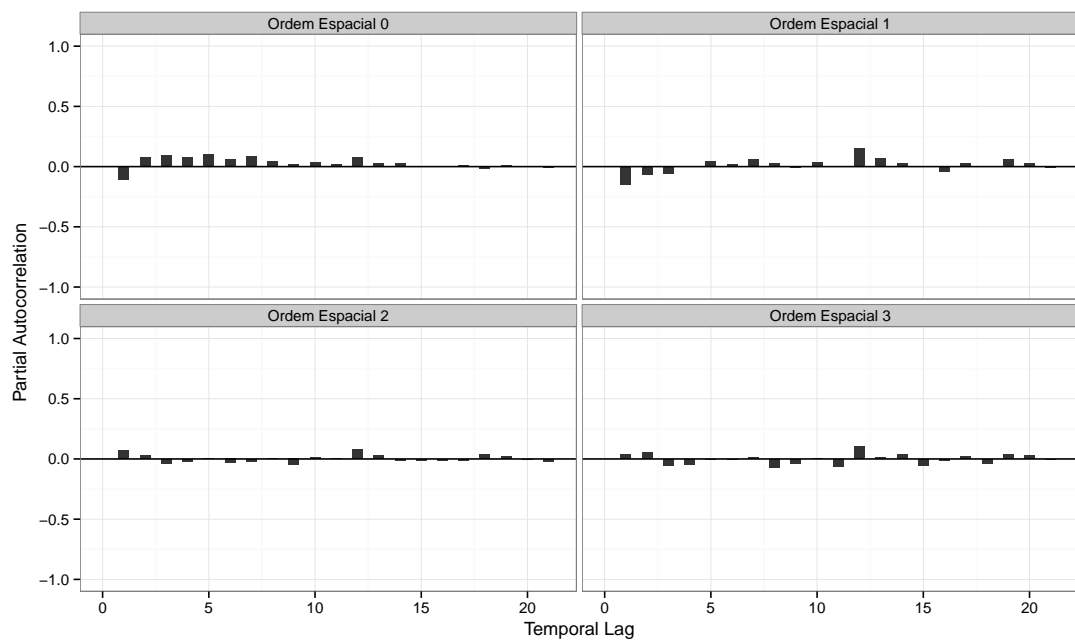


Figura 38 – STACF dos Resíduos do modelo STARMA(12,2,1,0) incompleto, nas ordens espaciais 0, 1, 2 e 3

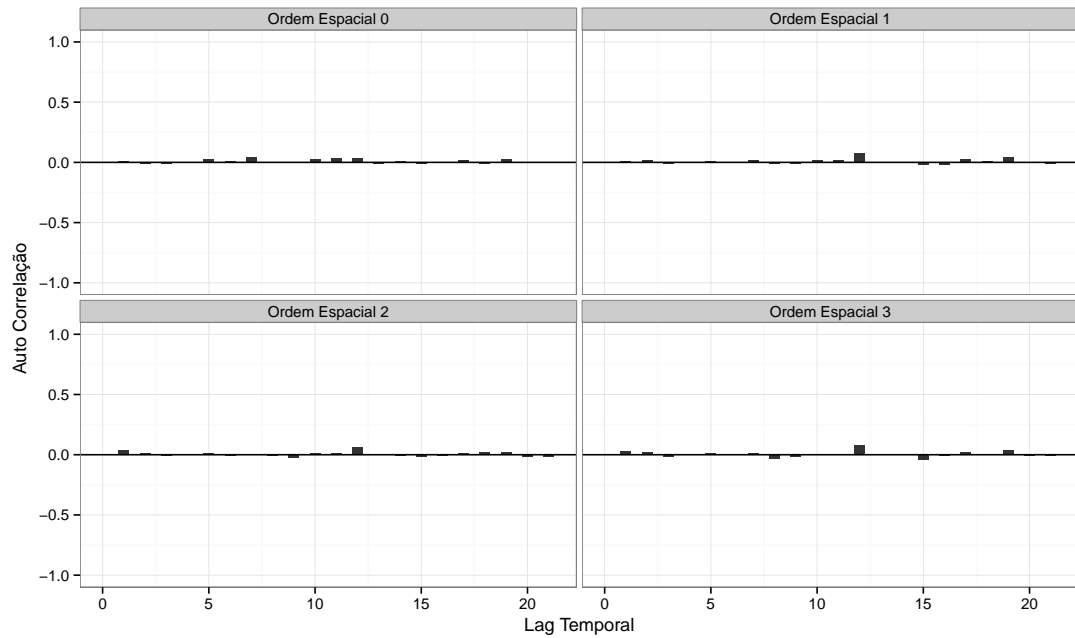


Figura 39 – STPACF dos Resíduos do modelo STARMA(12,2,1,0) incompleto, nas ordens espaciais 0, 1, 2 e 3

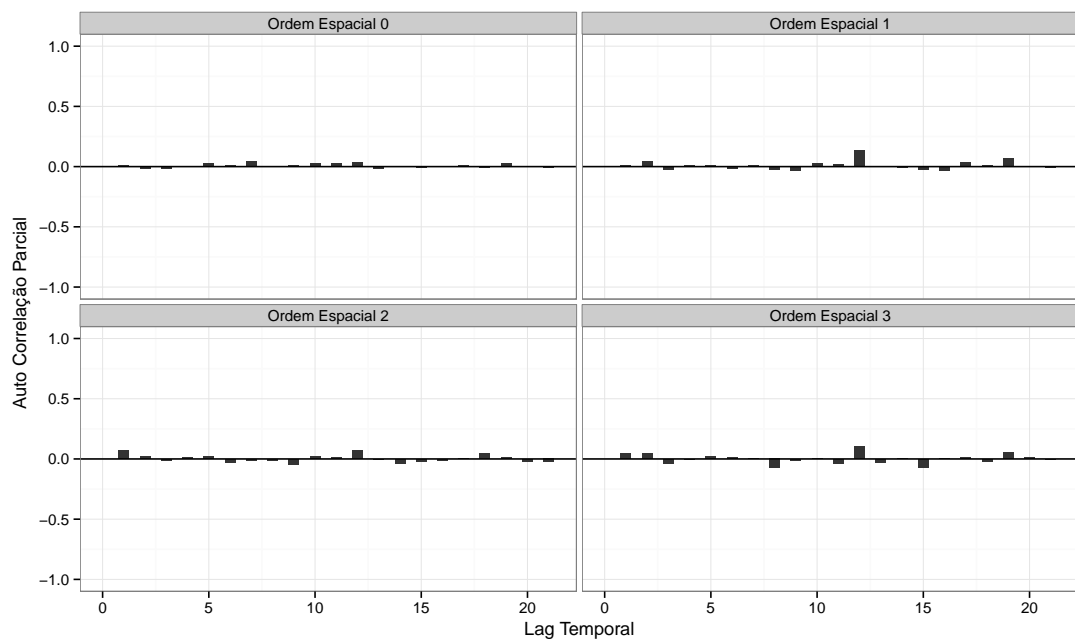


Figura 40 – QQ-Normal dos Resíduos do Modelo STARMA Final - Parte 1/4

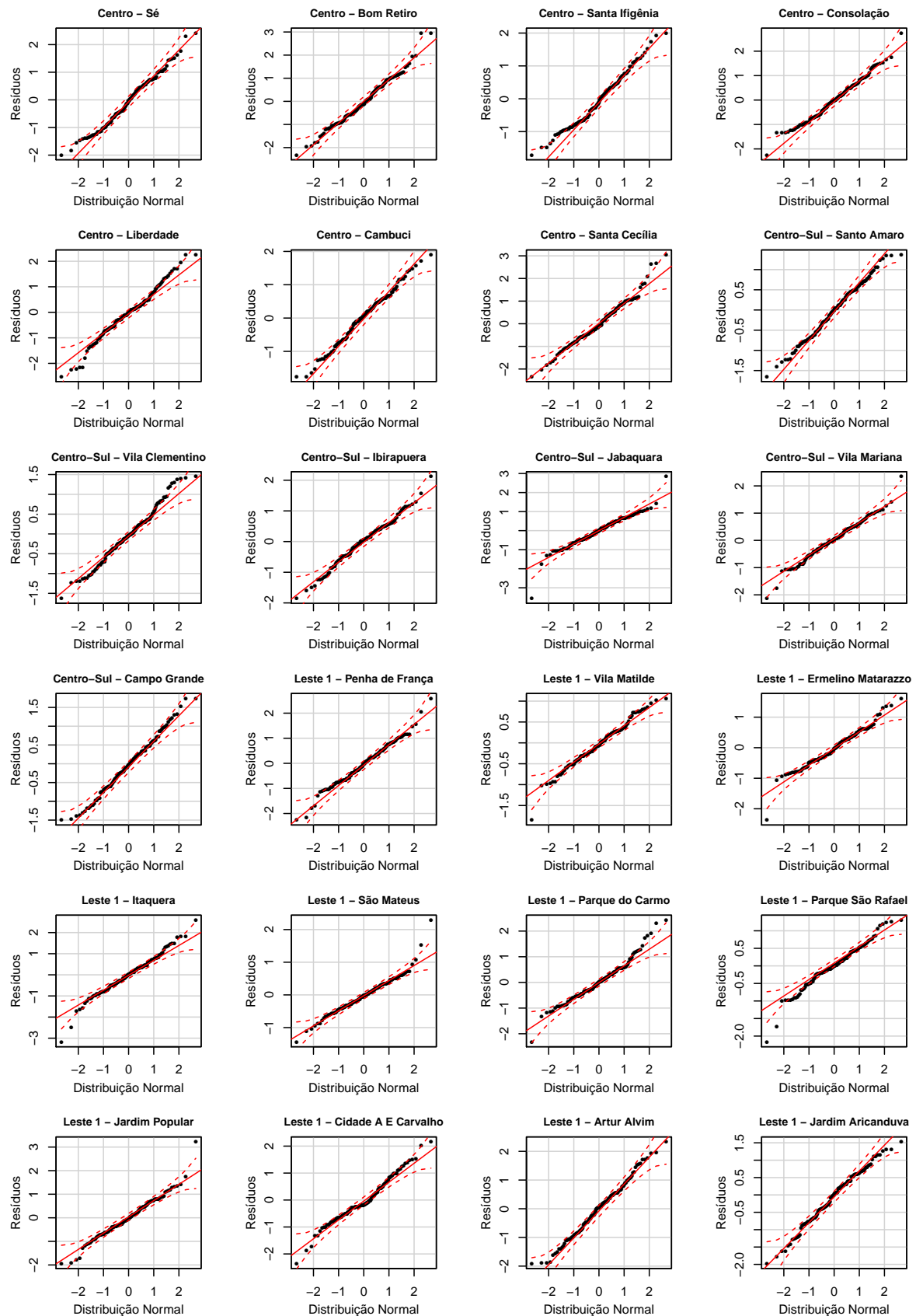


Figura 41 – QQ-Normal dos Resíduos do Modelo STARMA Final - Parte 2/4

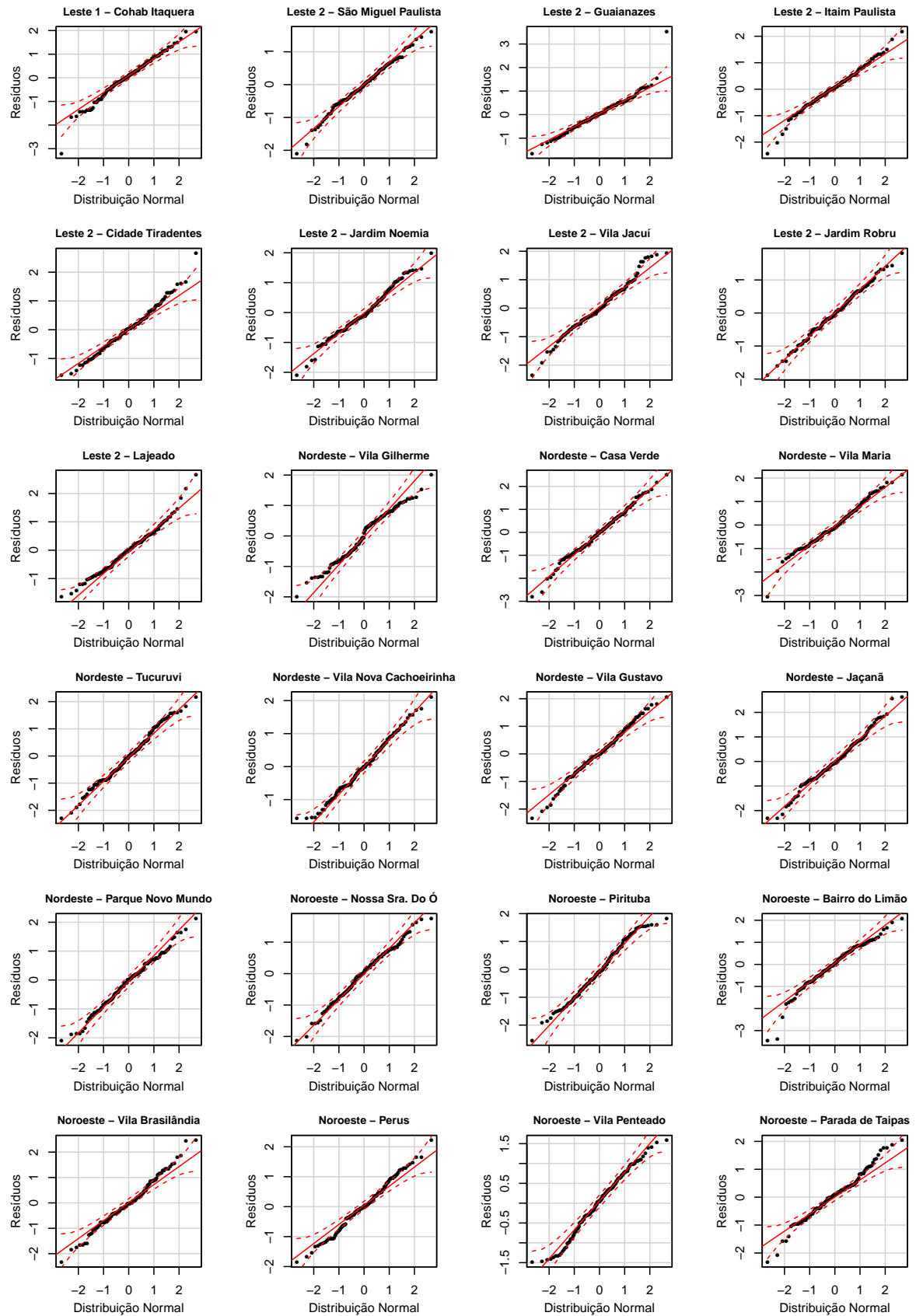


Figura 42 – QQ-Normal dos Resíduos do Modelo STARMA Final - Parte 3/4

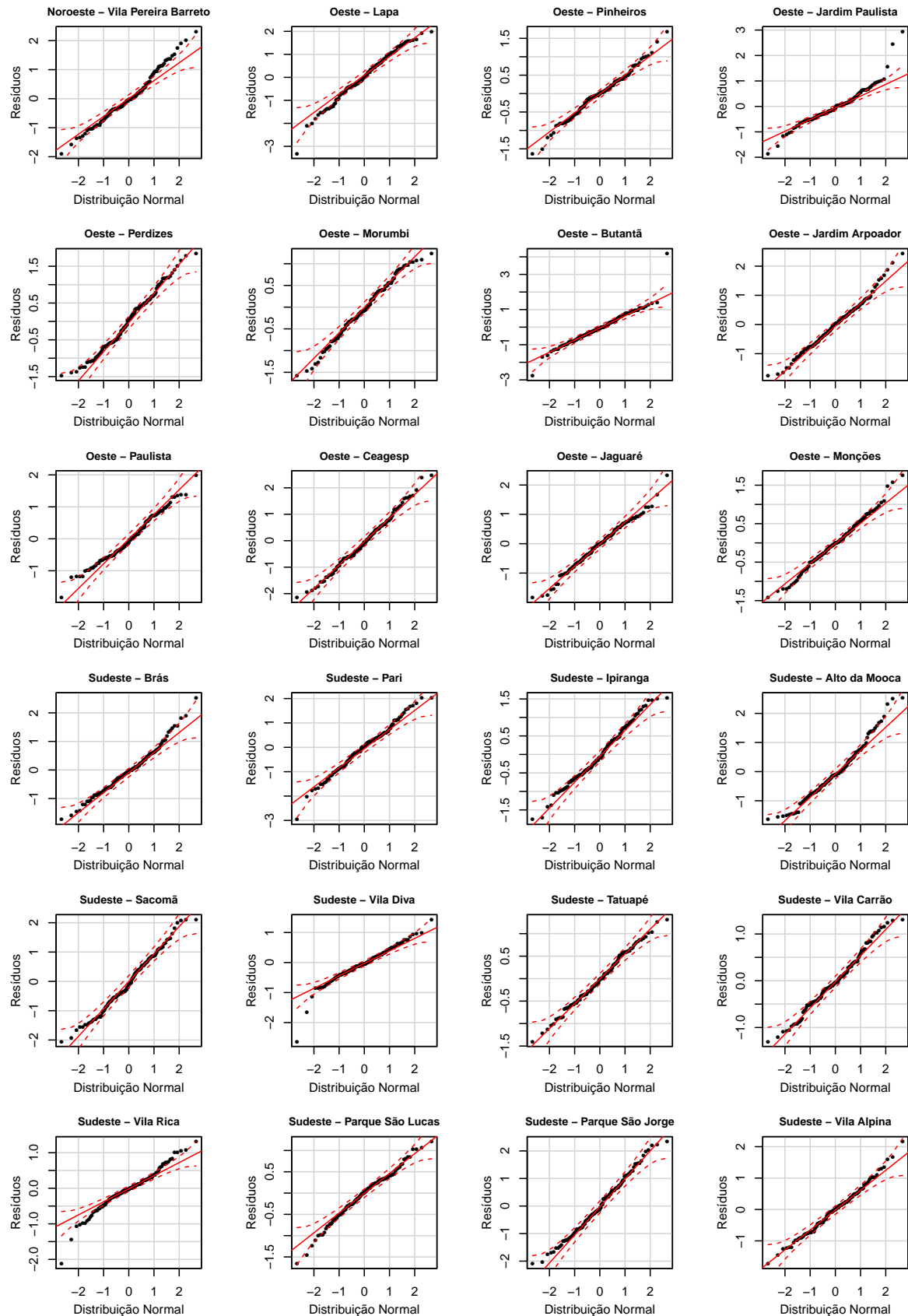


Figura 43 – QQ-Normal dos Resíduos do Modelo STARMA Final - Parte 4/4

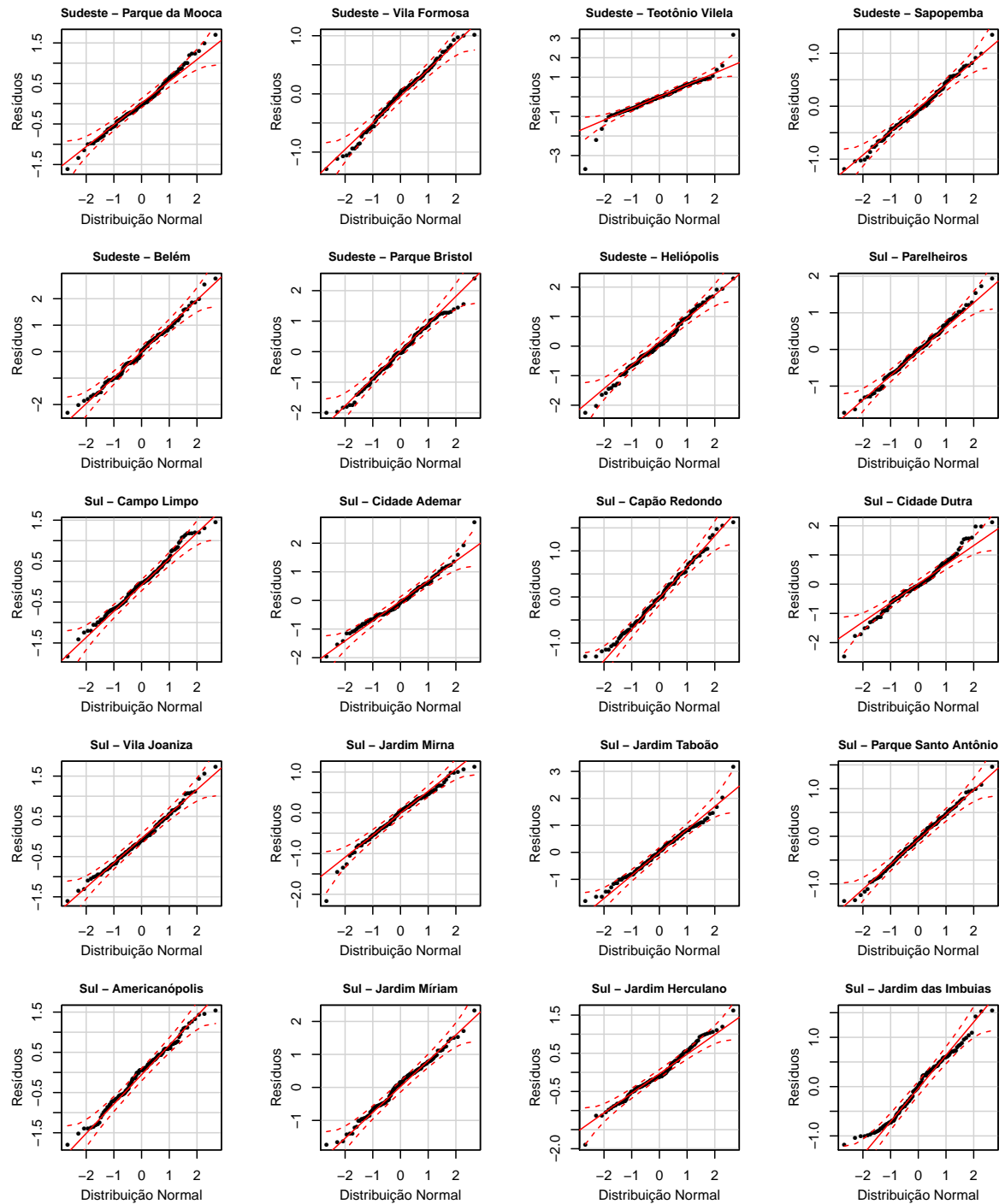


Figura 44 – Comparação entre Observado e Estimado pelo Modelo STARMA - Parte 1/5.

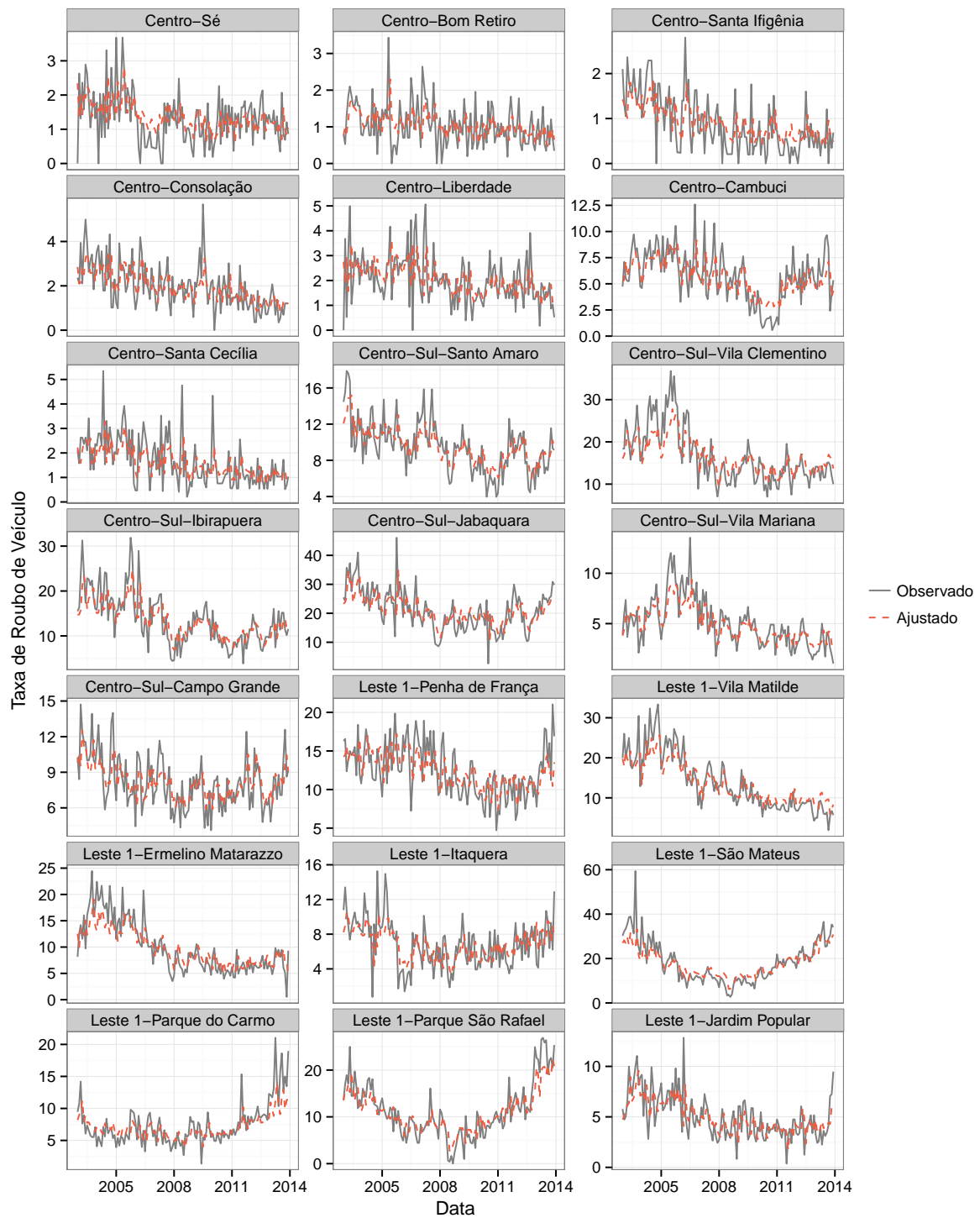


Figura 45 – Comparação entre Observado e Estimado pelo Modelo STARMA - Parte 2/5.

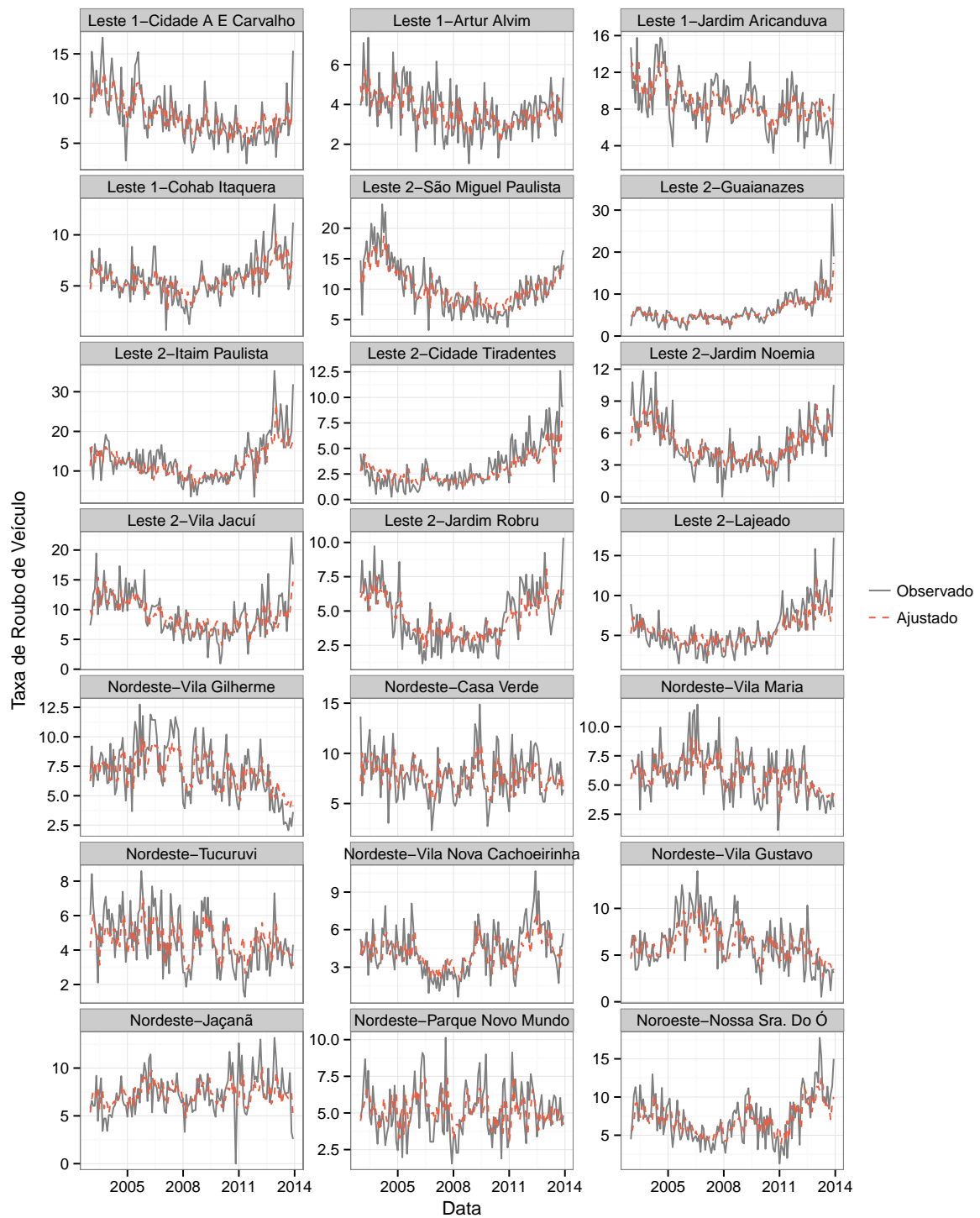


Figura 46 – Comparação entre Observado e Estimado pelo Modelo STARMA - Parte 3/5.

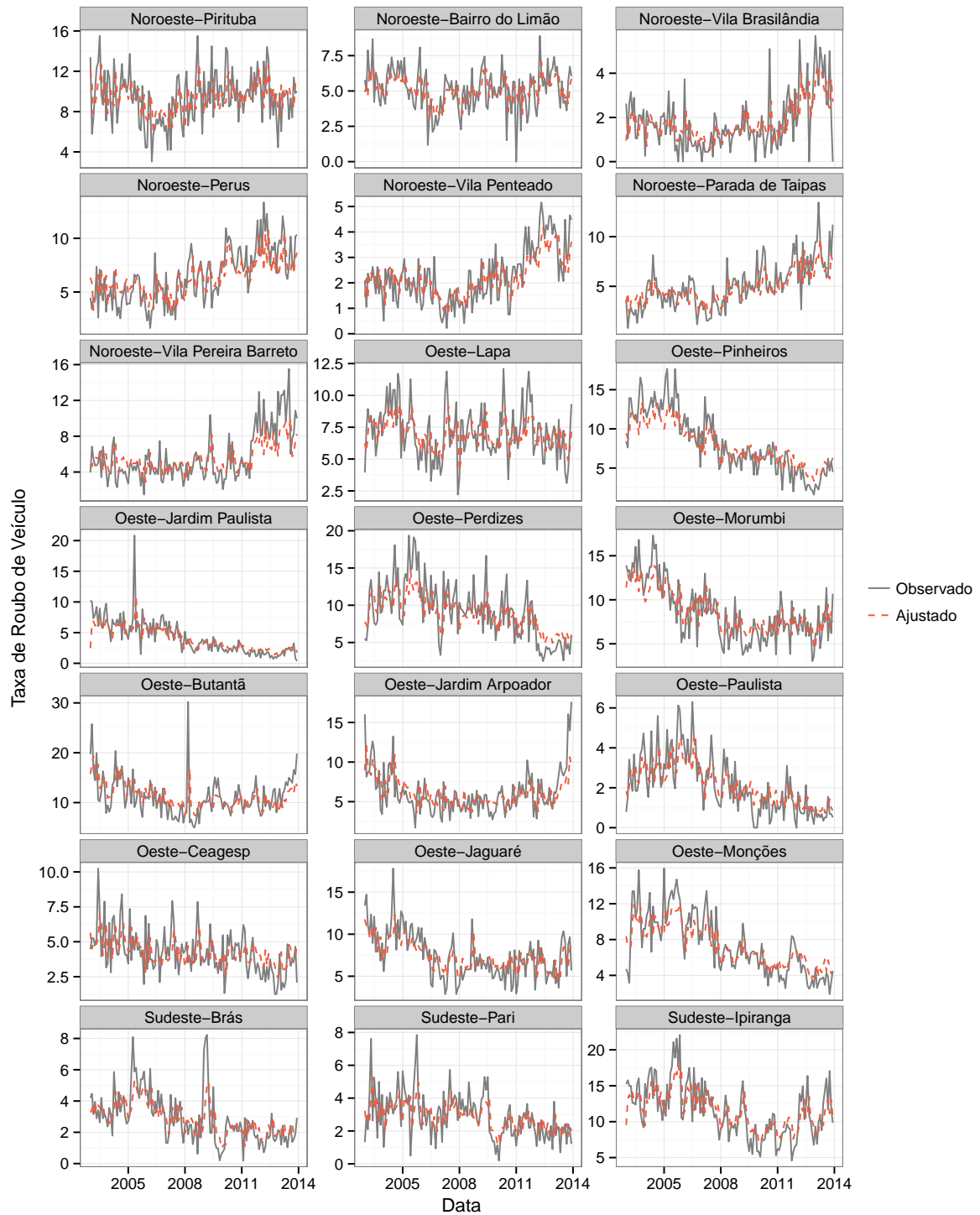


Figura 47 – Comparação entre Observado e Estimado pelo Modelo STARMA - Parte 4/5.

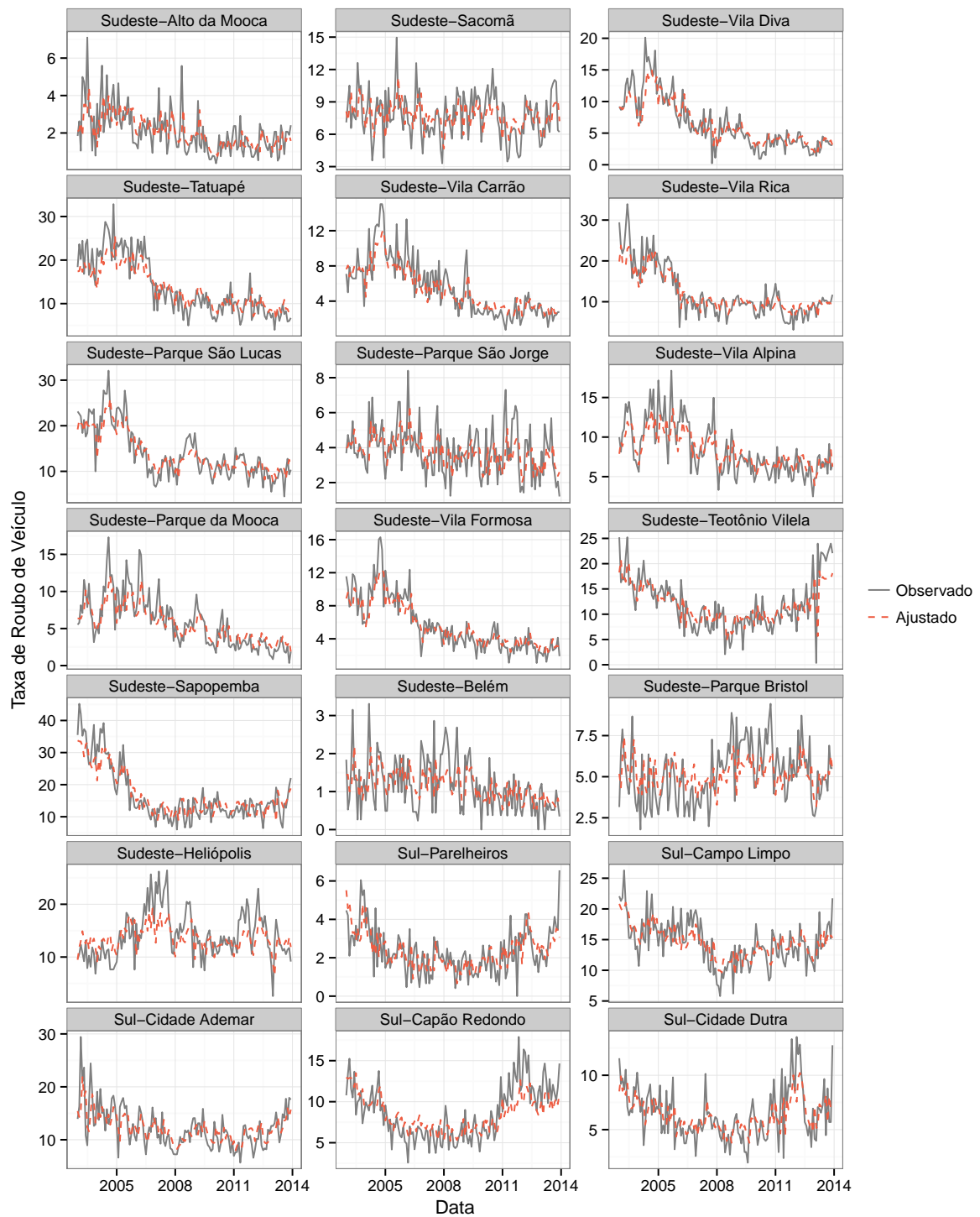


Figura 48 – Comparação entre Observado e Estimado pelo Modelo STARMA - Parte 5/5.

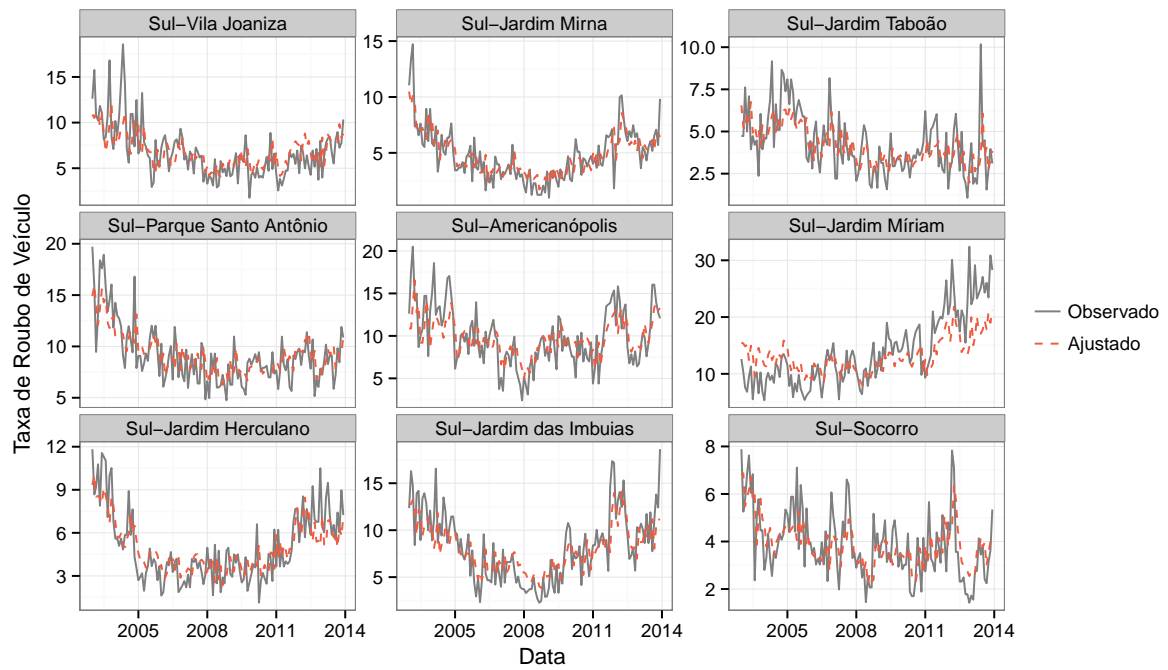


Tabela 5 – ARL e desvio padrão(DP) para o MCUSUM.

k	h	δ								
		0	0,25	0,50	0,75	1,00	1,25	1,50	1,75	2,00
0,10	54,88	20,04	8,05	5,85	4,83	4,00	3,60	3,00	3,00	3,00
		(6,08)	(0,68)	(0,39)	(0,37)	(0,02)	(0,49)	(0,01)	(0,01)	(0,01)
0,15	45,08	19,92	7,42	5,22	4,10	3,85	3,00	3,00	2,99	2,15
		(8,14)	(0,65)	(0,42)	(0,30)	(0,36)	(0,04)	(0,01)	(0,11)	(0,36)
0,20	35,56	19,88	6,65	4,80	3,96	3,02	3,00	2,78	2,01	2,00
		(10,21)	(0,65)	(0,41)	(0,19)	(0,13)	(0,01)	(0,41)	(0,08)	(0,01)
0,25	27,91	20,00	6,02	4,15	3,24	3,00	2,59	2,00	2,00	2,00
		(12,46)	(0,65)	(0,36)	(0,43)	(0,01)	(0,49)	(0,04)	(0,01)	(0,01)
0,30	21,03	19,88	5,30	3,79	3,00	2,50	2,00	2,00	2,00	2,00
		(14,12)	(0,63)	(0,42)	(0,04)	(0,50)	(0,03)	(0,01)	(0,01)	(0,01)
0,35	16,19	20,04	4,80	3,23	2,78	2,01	2,00	2,00	1,97	1,35
		(14,70)	(0,65)	(0,42)	(0,41)	(0,11)	(0,01)	(0,01)	(0,18)	(0,48)
0,40	12,40	20,11	4,39	3,02	2,18	2,00	2,00	1,79	1,08	1,00
		(15,54)	(0,64)	(0,25)	(0,38)	(0,01)	(0,04)	(0,41)	(0,27)	(0,01)
0,45	9,47	20,01	4,06	2,80	2,02	2,00	1,72	1,06	1,00	1,00
		(16,01)	(0,65)	(0,41)	(0,12)	(0,06)	(0,45)	(0,23)	(0,01)	(0,01)
0,70	2,16	19,88	3,96	2,22	1,57	1,04	1,00	1,00	1,00	1,00
		(12,36)	(1,03)	(0,48)	(0,50)	(0,19)	(0,01)	(0,01)	(0,01)	(0,01)
0,75	1,38	20,00	4,22	2,26	1,51	1,03	1,00	1,00	1,00	1,00
		(11,73)	(1,12)	(0,53)	(0,50)	(0,18)	(0,01)	(0,01)	(0,01)	(0,01)
0,80	0,75	20,08	4,47	2,32	1,54	1,04	1,00	1,00	1,00	1,00
		(11,63)	(1,28)	(0,57)	(0,50)	(0,20)	(0,01)	(0,01)	(0,01)	(0,01)

Tabela 6 – ARL e desvio padrão para o MEWMA.

λ	h	δ								
		0	0,25	0,5	0,75	1	1,25	1,5	1,75	2
0,05	282,74	19,87	5,10	3,23	2,41	2,00	1,95	1,31	1,00	1,00
		(12,53)	(0,72)	(0,43)	(0,49)	(0,01)	(0,22)	(0,46)	(0,06)	(0,01)
0,05	284,18	19,94	5,09	3,25	2,43	2,00	1,96	1,31	1,01	1,00
		(12,47)	(0,70)	(0,44)	(0,50)	(0,01)	(0,20)	(0,46)	(0,07)	(0,01)
0,05	287,05	20,16	5,16	3,26	2,47	2,00	1,97	1,35	1,00	1,00
		(12,42)	(0,73)	(0,44)	(0,50)	(0,03)	(0,18)	(0,48)	(0,06)	(0,01)
0,10	257,91	19,81	4,73	3,06	2,18	2,00	1,81	1,09	1,00	1,00
		(14,56)	(0,68)	(0,30)	(0,38)	(0,03)	(0,40)	(0,29)	(0,01)	(0,01)
0,10	261,88	20,10	4,78	3,09	2,21	2,00	1,85	1,12	1,00	1,00
		(14,82)	(0,70)	(0,33)	(0,41)	(0,02)	(0,36)	(0,32)	(0,02)	(0,01)
0,10	263,20	20,04	4,79	3,10	2,21	2,00	1,85	1,13	1,00	1,00
		(14,20)	(0,69)	(0,33)	(0,41)	(0,04)	(0,36)	(0,34)	(0,01)	(0,01)
0,15	236,32	20,18	4,41	2,94	2,05	1,99	1,55	1,02	1,00	1,00
		(15,76)	(0,65)	(0,32)	(0,23)	(0,12)	(0,50)	(0,13)	(0,01)	(0,01)
0,15	237,49	19,95	4,42	2,96	2,06	1,99	1,56	1,02	1,00	1,00
		(15,57)	(0,67)	(0,31)	(0,24)	(0,11)	(0,50)	(0,13)	(0,01)	(0,01)
0,20	213,06	20,10	4,11	2,76	2,01	1,92	1,22	1,00	1,00	1,00
		(17,18)	(0,64)	(0,44)	(0,11)	(0,28)	(0,42)	(0,04)	(0,01)	(0,01)
0,20	214,12	19,88	4,13	2,77	2,01	1,92	1,25	1,00	1,00	1,00
		(16,68)	(0,65)	(0,44)	(0,11)	(0,27)	(0,43)	(0,06)	(0,01)	(0,01)
0,20	215,18	19,80	4,14	2,79	2,01	1,93	1,25	1,00	1,00	1,00
		(16,45)	(0,65)	(0,42)	(0,10)	(0,25)	(0,44)	(0,05)	(0,01)	(0,01)

Tabela 7 – ARL e desvio padrão para o MEWMA assintótico.

λ	h	δ								
		0	0,25	0,5	0,75	1	1,25	1,5	1,75	2
0,05	282,74	19,87	7,81	5,62	4,52	4,00	3,19	3,00	3,00	2,99
		(12,53)	(0,69)	(0,49)	(0,50)	(0,02)	(0,39)	(0,01)	(0,01)	(0,11)
0,05	284,18	19,94	7,82	5,62	4,55	4,00	3,21	3,00	3,00	2,99
		(12,47)	(0,66)	(0,49)	(0,50)	(0,02)	(0,41)	(0,01)	(0,01)	(0,10)
0,05	287,05	20,16	7,88	5,65	4,59	4,00	3,24	3,00	3,00	3,00
		(12,42)	(0,69)	(0,48)	(0,49)	(0,02)	(0,43)	(0,01)	(0,01)	(0,06)
0,10	257,91	19,81	6,10	4,32	3,66	3,00	3,00	2,22	2,00	2,00
		(14,56)	(0,60)	(0,47)	(0,48)	(0,01)	(0,07)	(0,42)	(0,01)	(0,01)
0,10	261,88	20,10	6,16	4,36	3,69	3,00	3,00	2,28	2,00	2,00
		(14,82)	(0,61)	(0,48)	(0,46)	(0,01)	(0,06)	(0,45)	(0,02)	(0,01)
0,10	263,20	20,04	6,15	4,38	3,71	3,00	3,00	2,30	2,00	2,00
		(14,20)	(0,60)	(0,49)	(0,46)	(0,02)	(0,06)	(0,46)	(0,01)	(0,01)
0,15	236,32	20,18	5,25	3,89	3,01	2,90	2,02	2,00	2,00	2,00
		(15,76)	(0,57)	(0,33)	(0,07)	(0,30)	(0,15)	(0,01)	(0,01)	(0,01)
0,15	237,49	19,95	5,26	3,89	3,01	2,92	2,03	2,00	2,00	2,00
		(15,57)	(0,58)	(0,32)	(0,07)	(0,27)	(0,18)	(0,01)	(0,01)	(0,01)
0,20	213,06	20,10	4,67	3,24	2,94	2,06	2,00	2,00	2,00	2,00
		(17,18)	(0,59)	(0,43)	(0,23)	(0,24)	(0,01)	(0,01)	(0,01)	(0,03)
0,20	214,12	19,88	4,66	3,25	2,95	2,07	2,00	2,00	2,00	2,00
		(16,68)	(0,59)	(0,44)	(0,22)	(0,26)	(0,01)	(0,01)	(0,01)	(0,04)
0,20	215,18	19,80	4,67	3,25	2,96	2,07	2,00	2,00	2,00	2,00
		(16,45)	(0,59)	(0,44)	(0,21)	(0,26)	(0,01)	(0,01)	(0,01)	(0,03)

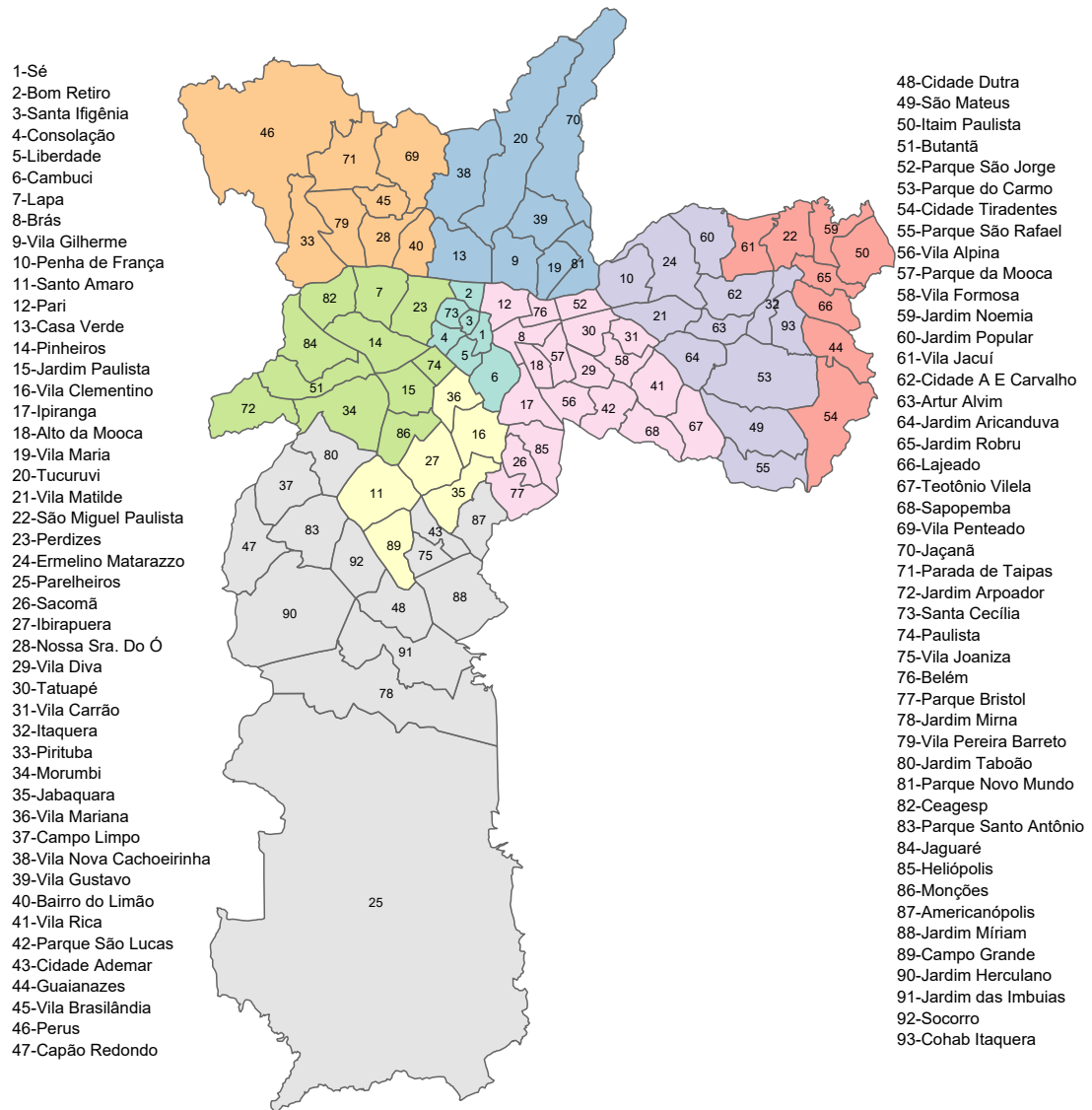
Tabela 8 – ARL e desvio padrão para o MCUSUM com STARMA Incorreto.

k	h	δ								
		0	0,25	0,5	0,75	1	1,25	1,5	1,75	2
0,10	51,80	18,45	7,73	5,57	4,49	4,00	3,17	3,00	3,00	2,98
		(5,83)	(0,67)	(0,50)	(0,50)	(0,04)	(0,38)	(0,01)	(0,01)	(0,14)
0,15	42,96	18,73	7,15	5,08	4,02	3,56	3,00	3,00	2,85	2,01
		(7,37)	(0,65)	(0,30)	(0,15)	(0,50)	(0,01)	(0,01)	(0,36)	(0,10)
0,20	34,18	18,41	6,47	4,62	3,88	3,00	3,00	2,47	2,00	2,00
		(8,73)	(0,64)	(0,49)	(0,32)	(0,05)	(0,02)	(0,50)	(0,02)	(0,01)
0,25	26,31	17,99	5,76	4,05	3,07	2,99	2,23	2,00	2,00	2,00
		(10,44)	(0,63)	(0,25)	(0,26)	(0,07)	(0,42)	(0,01)	(0,01)	(0,01)
0,30	20,06	17,61	5,13	3,62	2,99	2,26	2,00	2,00	2,00	2,00
		(11,74)	(0,62)	(0,49)	(0,07)	(0,44)	(0,01)	(0,01)	(0,01)	(0,03)
0,35	15,17	17,81	4,56	3,10	2,54	2,00	2,00	2,00	1,78	1,06
		(13,53)	(0,61)	(0,31)	(0,50)	(0,03)	(0,01)	(0,05)	(0,41)	(0,24)
0,40	11,27	16,69	4,09	2,90	2,05	2,00	1,96	1,33	1,00	1,00
		(13,28)	(0,61)	(0,33)	(0,22)	(0,01)	(0,19)	(0,47)	(0,05)	(0,01)
0,45	8,53	15,82	3,76	2,53	2,00	1,95	1,31	1,00	1,00	1,00
		(12,22)	(0,64)	(0,50)	(0,06)	(0,22)	(0,46)	(0,06)	(0,01)	(0,01)
0,55	4,87	14,81	3,38	2,14	1,84	1,17	1,00	1,00	1,00	1,00
		(11,17)	(0,71)	(0,36)	(0,37)	(0,37)	(0,05)	(0,01)	(0,01)	(0,01)
0,60	3,62	14,27	3,37	2,08	1,62	1,05	1,00	1,00	1,00	1,00
		(9,90)	(0,79)	(0,33)	(0,49)	(0,22)	(0,01)	(0,01)	(0,01)	(0,01)
0,70	1,72	13,09	3,57	2,03	1,35	1,01	1,00	1,00	1,00	1,00
		(7,50)	(0,97)	(0,48)	(0,48)	(0,11)	(0,01)	(0,01)	(0,01)	(0,01)
0,80	0,45	13,08	4,01	2,13	1,38	1,02	1,00	1,00	1,00	1,00
		(7,28)	(1,21)	(0,56)	(0,49)	(0,13)	(0,01)	(0,01)	(0,01)	(0,01)

Tabela 9 – ARL e desvio padrão para o MEWMA com STARMA incorreto.

λ	h	δ								
		0	0,25	0,5	0,75	1	1,25	1,5	1,75	2
0,05	266,40	18,01	4,89	3,12	2,24	2,00	1,88	1,15	1,00	1,00
		(11,13)	(0,72)	(0,36)	(0,43)	(0,02)	(0,33)	(0,35)	(0,03)	(0,01)
0,05	267,75	18,26	4,89	3,14	2,27	2,00	1,88	1,15	1,00	1,00
		(10,98)	(0,70)	(0,36)	(0,44)	(0,01)	(0,32)	(0,36)	(0,02)	(0,01)
0,10	246,02	17,65	4,58	3,00	2,10	1,99	1,69	1,04	1,00	1,00
		(11,90)	(0,67)	(0,30)	(0,30)	(0,08)	(0,46)	(0,19)	(0,01)	(0,01)
0,10	247,25	17,83	4,60	3,02	2,10	2,00	1,68	1,04	1,00	1,00
		(12,07)	(0,66)	(0,30)	(0,30)	(0,05)	(0,47)	(0,20)	(0,01)	(0,01)
0,15	223,86	17,41	4,26	2,86	2,03	1,97	1,37	1,01	1,00	1,00
		(13,00)	(0,65)	(0,38)	(0,17)	(0,18)	(0,48)	(0,07)	(0,01)	(0,01)
0,20	202,33	17,68	3,94	2,64	2,00	1,84	1,13	1,00	1,00	1,00
		(14,68)	(0,62)	(0,48)	(0,07)	(0,37)	(0,33)	(0,03)	(0,01)	(0,01)
0,20	203,34	17,77	3,97	2,66	2,00	1,85	1,14	1,00	1,00	1,00
		(15,03)	(0,64)	(0,48)	(0,07)	(0,36)	(0,35)	(0,01)	(0,01)	(0,01)
0,20	204,35	17,69	3,99	2,67	2,01	1,86	1,14	1,00	1,00	1,00
		(14,42)	(0,63)	(0,48)	(0,08)	(0,35)	(0,34)	(0,01)	(0,01)	(0,01)

Figura 49 – Mapa Numerado



APÊNDICE B – STARMA

Estacionariedade do STARMA

Uma condição fundamental é que o processo seja estacionário, ou seja, a estrutura de covariância de $\mathbf{Z}(t)$ não se altera com o tempo. Segundo Pfeifer e Deutsch (1980), há estacionariedade se, e somente se, as raízes da equação (0.1) resolvida em y estiverem dentro do círculo unitário, ou seja, $|y| \leq 1$. Esse resultado serve para determinar os valores para os quais $\phi_{k,j}$ torna o processo estacionário e com solução única. E, se não houver estacionariedade é necessário diferenciá-la com lags simples ou sazonais a fim de torná-la.

$$\det \left[\mathbf{I} - \sum_{k=1}^p \sum_{j=0}^{v_k} \phi_{k,j} \mathbf{W}_j y^{-k} \right] = 0 \quad (0.1)$$

E, o modelo STARMA será inversível se, e somente se, as raízes da equação (0.2) resolvidas em y estiverem dentro do círculo unitário, ou seja, $|y| \leq 1$.

$$\det \left[\mathbf{I} - \sum_{k=1}^q \sum_{j=1}^{u_k} \theta_{k,j} \mathbf{W}_j y^{-k} \right] = 0 \quad (0.2)$$

Autocorrelação do STARMA

Pfeifer e Deutsch (1980) apresenta a expressão (0.3) para obter a covariância entre os lags espaciais r e s no lag temporal k .

$$\gamma_{r,s}^k = E \left\{ [\mathbf{W}_r \mathbf{Z}_t]' [\mathbf{W}_s \mathbf{Z}_{t+k}] \right\} \quad (0.3)$$

O cálculo ainda pode ser expresso de forma simplificada. Seja $\Gamma(k) = E[\mathbf{Z}_t \mathbf{Z}_{t+k}']$, então a covariância assume:

$$\gamma_{r,s}^k = tr \left\{ \mathbf{W}_r' \mathbf{W}_s \Gamma(k) \right\} \quad (0.4)$$

A estimativa amostral para $\hat{\gamma}_{r,s}^k$ é obtida trocando-se $\Gamma(k)$ por $\hat{\Gamma}(k)$.

$$\hat{\Gamma}(k) = \sum_{t=1}^{T-k} \frac{(\mathbf{Z}_t - \bar{\mathbf{Z}})(\mathbf{Z}_{t+k} - \bar{\mathbf{Z}})'}{T-k} \quad (0.5)$$

Sendo $\bar{\mathbf{Z}}$ um vetor de tamanho $N \times 1$ e cada elemento é igual à média sobre todas as localidades e tempo. Com isso, autocorrelação fica expressa como:

$$\rho_{r,s}^k = \frac{\gamma_{r,s}^k}{\sqrt{\gamma_{r,r}^0 \gamma_{s,s}^0}} \quad (0.6)$$

E a variância da autocorrelação é:

$$\hat{Var}(\hat{\rho}_{r,0}^k) \approx \frac{1}{N(T-k)} \quad (0.7)$$

Segundo Pfeifer e Deutsch (1980), a autocorrelação parcial é obtida do modelo STAR.

$$\mathbf{Z}_t = \sum_{k=1}^p \sum_{j=0}^{v_k} \phi_{k,j} \mathbf{W}_j \mathbf{Z}_{t-k} + \varepsilon_t \quad (0.8)$$

Ele inicia multiplicando toda a equação (0.8) por $[\mathbf{W}_h \mathbf{Z}_{t-u}]'$, com u sendo um lag temporal e h uma ordem espacial:

$$\mathbf{Z}_{t-u}' \mathbf{W}_h' \mathbf{Z}_t = \sum_{k=1}^p \sum_{j=0}^{v_k} \phi_{k,j} \mathbf{Z}_{t-u}' \mathbf{W}_h \mathbf{W}_j' \mathbf{Z}_{t-k} + \mathbf{Z}_{t-u}' \mathbf{W}_h' \varepsilon_t \quad (0.9)$$

As autocovariâncias são derivadas da expressão (0.9) ao calcular a esperança.

$$E \left(\mathbf{Z}_{t-u}' \mathbf{W}_h' \mathbf{Z}_t \right) = \sum_{k=1}^p \sum_{j=0}^{v_k} \phi_{k,j} E \left(\mathbf{Z}_{t-u}' \mathbf{W}_h \mathbf{W}_j' \mathbf{Z}_{t-k} \right) + E \left(\mathbf{Z}_{t-u}' \mathbf{W}_h' \varepsilon_t \right) \quad (0.10)$$

Recuperando a expressão (0.3) e fazendo as equivalências em (0.10) temos:

$$E \left(\mathbf{Z}_{t-u}' \mathbf{W}_h' \mathbf{Z}_t \right) = \gamma_{h0}^u \quad (0.11)$$

$$E \left(\mathbf{Z}_{t-u}' \mathbf{W}_h \mathbf{W}_j' \mathbf{Z}_{t-k} \right) = \gamma_{h,j}^{u-k} \quad (0.12)$$

$$E \left(\mathbf{Z}_{t-u}' \mathbf{W}_h' \varepsilon_t \right) = 0 \quad (0.13)$$

Agora, a equação (0.10) pode ser reescrita como:

$$\gamma_{h,0}^u = \sum_{k=1}^p \sum_{j=0}^{v_k} \phi_{k,j} \gamma_{h,j}^{u-k} \quad (0.14)$$

As equações de (0.14) compõem o sistema de Yule-Walker, utilizado para estimar os parâmetros $\phi_{k,j}$ que formam a autocorrelação parcial.

APÊNDICE C – Códigos R

A seguir serão apresentados os códigos em R que foram usados neste trabalho.

```
1 library(reshape2)
2 library(starma) #utilizada apenas para obter a STACF e STPACF
3 library(car) #faz o QQ-Plot Normal
4 library(ggplot2)
5 library(gridExtra)
6 library(RColorBrewer) #paleta de cores
7 library(scales)
8 library(rgdal)
9 library(dplyr)
10 library(plyr)
11 library(zoo) #Formato de data Ano/Mes
```

Código C.1 – "Bibliotecas Utilizadas"

```
1 #Polígonos
2 dat = read.table("clipboard", dec = ",", header=T, sep="\t")
3
4 #Dados dos crimes
5 dat.rou.frota.atual = read.table("clipboard",
6                                 dec = ",", header=T, sep="\t")
7
8 datapoly.atual <- merge(dat,
9                        dat.rou.frota.atual[,c("id",
10                                               "anomes", "taxa.roubo", "ano")],
11                        by=c("id"), sort=F)
12 datapoly.atual = arrange(datapoly.atual, anomes, id, ordem)
13 head(datapoly.atual)
14
15 lista.anomes = sort(unique(datapoly.atual$ano))[-1]
16 for (i in lista.anomes) {
17   tmp <- subset(datapoly.atual, ano == i)
18
19   m = ggplot(tmp, aes(x=lng, y=lat)) +
20     geom_polygon(aes(fill=taxa.roubo, group=id),
21                 size=0.0001, colour=NA)+
22     scale_fill_gradientn(colours=rev(brewer.pal(11,
23                                                "Spectral"))),
24                        limits=c(0,30),
25                        na.value = rev(brewer.pal(11,
26                                                "Spectral"))[11],
27                        name="Taxa de\nRoubo")+
28     facet_wrap(~anomes, ncol=6)+
29     ggtitle(paste0("Taxa de Roubo de Veículos em ", i,
30                   " (por milhão)"))+
31     theme_bw() +
```

```

32     theme(plot.title = element_text(vjust = 1.5),
33           panel.grid.major = element_blank(),
34           panel.grid.minor = element_blank(),
35           axis.title.x = element_blank(),
36           axis.ticks.x = element_blank(),
37           axis.text.x = element_blank(),
38           axis.title.y = element_blank(),
39           axis.ticks.y = element_blank(),
40           axis.text.y = element_blank())
41   ggsave(paste0("MapasRouboTaxa__", i, ".pdf"),
42         m, width=12, height=6)
43 }

```

Código C.2 – "Mapa do Roubo"

```

1 tmp.id.t = as.matrix((dcast(dat.rou.frota[, c("id",
2                                           "anomes",
3                                           "taxa.roubo")],
4                               id ~ anomes,
5                               value.var="taxa.roubo")[, -1]))
6 #Retira amostra para validação posterior
7 tmp.id.t = tmp.id.t[, 1:(ncol(tmp.id.t)-12-7)]
8
9 medias.sqrt <- sapply(1:93, function(x) {
10   mean(sqrt(tmp.id.t[x,]))
11 })
12 var.sqrt <- sapply(1:93, function(x) {
13   var(sqrt(tmp.id.t[x,]))
14 })
15
16 dim(as.matrix(tmp.id.t))
17 z.sqrt.pad <- matrix(0, 93, 144)
18 for(i in 1:93) {
19   z.sqrt.pad[i, ] <- (sqrt(as.numeric(as.matrix(
20     tmp.id.t)[i,])) - medias.sqrt[i]) /
21     sqrt(var.sqrt[i])
22 }
23
24 lista.cenarios <- t(combn(x=c(0.05, 0.1, 0.2, 0.3, 0.4, 0.5,
25                               1, 2, 3, 4, 5, 6, 7, 8, 9, 10),
26                             m=4))
27 nrow(lista.cenarios)
28
29 i.moran.median <- vector("list")
30 i.moran.q05 <- vector("list")
31 i.moran.q95 <- vector("list")
32 i.moran.sd <- vector("list")
33 i.moran.lista <- vector("list")
34
35 i.moran.lista[[1]] <- rep(0, ncol(tmp.id.t))
36 i.moran.lista[[2]] <- rep(0, ncol(tmp.id.t))
37 i.moran.lista[[3]] <- rep(0, ncol(tmp.id.t))

```

```

38 i.moran.lista[[4]] <- rep(0, ncol(tmp.id.t))
39
40 for(k in lista.cenarios.ativos) {
41
42   W1 <- ifelse(mat.dist <= lista.cenarios[k,1], 1, 0)
43   diag(W1) <- 0
44   W1std = W1 / rowSums(W1)
45   W1std[!is.finite(W1std)] <- 0
46
47   mat.dist.bin <- ifelse(mat.dist > lista.cenarios[k,1] &
48                         mat.dist <= lista.cenarios[k,2],
49                         1, 0)
50   diag(mat.dist.bin) <- 0
51   W2std = mat.dist.bin / rowSums(mat.dist.bin)
52   W2std[!is.finite(W2std)] <- 0
53
54   mat.dist.bin <- ifelse(mat.dist > lista.cenarios[k,2] &
55                         mat.dist <= lista.cenarios[k,3],
56                         1, 0)
57   diag(mat.dist.bin) <- 0
58   W3std = mat.dist.bin / rowSums(mat.dist.bin)
59   W3std[!is.finite(W3std)] <- 0
60
61   mat.dist.bin <- ifelse(mat.dist > lista.cenarios[k,3] &
62                         mat.dist <= lista.cenarios[k,4],
63                         1, 0)
64   diag(mat.dist.bin) <- 0
65   W4std = mat.dist.bin / rowSums(mat.dist.bin)
66   W4std[!is.finite(W4std)] <- 0
67
68   Wstd <- list(W1std, W2std, W3std, W4std)
69
70   for (j in 1:ncol(tmp.id.t)) {
71     y.moran <- c()
72     ordem.lag <- j
73     z.bar.moran <- mean(z.dr[,ordem.lag])
74     x.moran <- z.dr[,ordem.lag] - z.bar.moran
75
76     for(q in 1:4) {
77       for(i in 1:93){
78         indices.destino = which((Wstd[[q]][i,] > 0) == T)
79         y.moran[i] <- mean(z.dr[indices.destino,ordem.lag]) -
80                       z.bar.moran
81       }
82       i.moran.lista[[q]][j] <- as.numeric(
83         lm(y.moran~0+x.moran)$coefficients['x.moran'])
84     }
85   }
86
87   i.moran.median[[k]] <- sapply(1:4,
88                                 function(x) {

```

```

89         quantile(i.moran.lista[[x]],
90                  probs=0.50,
91                  na.rm=T)})
92 i.moran.q05[[k]] <- sapply(1:4,
93                            function(x) {
94                                quantile(i.moran.lista[[x]],
95                                         probs=0.05,
96                                         na.rm=T)})
97 i.moran.q95[[k]] <- sapply(1:4,
98                             function(x) {
99                                 quantile(i.moran.lista[[x]],
100                                         probs=0.95,
101                                         na.rm=T)})
102 i.moran.sd[[k]] <- sapply(1:4,
103                           function(x) {
104                               sd(i.moran.lista[[x]],
105                                  na.rm=T)})
106 cat(paste0("Cenário: ", k, "\n"))
107 }
108
109 sim.cenarios.moran <-
110 rbind(
111 cbind(rep(1, length(lista.cenarios.ativos)), seq(1, length(
112     lista.cenarios.ativos)),
113       (sapply(lista.cenarios.ativos, function(x) {
114         i.moran.median[[x]][1]
115       }))),
116 cbind(rep(2, length(lista.cenarios.ativos)), seq(1, length(
117     lista.cenarios.ativos)),
118       (sapply(lista.cenarios.ativos, function(x) {
119         i.moran.median[[x]][2]
120       }))),
121 cbind(rep(3, length(lista.cenarios.ativos)), seq(1, length(
122     lista.cenarios.ativos)),
123       (sapply(lista.cenarios.ativos, function(x) {
124         i.moran.median[[x]][3]
125       }))),
126 cbind(rep(4, length(lista.cenarios.ativos)), seq(1, length(
127     lista.cenarios.ativos)),
128       (sapply(lista.cenarios.ativos, function(x) {
129         i.moran.median[[x]][4]
130       })))
131 )
132 sim.cenarios.moran <- as.data.frame(sim.cenarios.moran)
133 names(sim.cenarios.moran)[1] <- "Orden"
134 names(sim.cenarios.moran)[2] <- "Seq"
135 names(sim.cenarios.moran)[3] <- "Moran"

```

```

136
137 compara.cenarios =
138   data.frame(
139     Seq=subset(sim.cenarios.moran, Ordem==1)$Seq,
140     Ordem1=abs(subset(sim.cenarios.moran, Ordem==1)$Moran),
141     Ordem2=abs(subset(sim.cenarios.moran, Ordem==2)$Moran),
142     Ordem3=abs(subset(sim.cenarios.moran, Ordem==3)$Moran),
143     Ordem4=abs(subset(sim.cenarios.moran, Ordem==4)$Moran))
144
145 compara.cenarios$soma2 = apply(compara.cenarios[2:3], 1, sum)
146 compara.cenarios$soma3 = apply(compara.cenarios[2:4], 1, sum)
147 compara.cenarios$soma4 = apply(compara.cenarios[2:5], 1, sum)
148
149 head(lista.cenarios.ativos)
150 ano.mes.lab <- sort(
151   unique(dat.rou.frota[, "anomes"])
152   )[1:(ncol(tmp.id.t))]
153
154 qplot(x=sim.cenarios.moran$Seq,
155       y=sim.cenarios.moran$Moran,
156       colour=factor(sim.cenarios.moran$Ordem)) +
157   geom_line() +
158   theme_bw() +
159   theme(plot.title = element_blank(),
160         axis.title.x = element_text(vjust = 0.2),
161         axis.title.y = element_text(vjust = 1.2)) +
162   xlab("Data") + ylab("Estatística I de Moran")
163
164 W1 <- ifelse(mat.dist <= 0.5, 1, 0)
165 diag(W1) <- 0
166 W1std = W1 / rowSums(W1)
167 W1std[!is.finite(W1std)] <- 0
168
169 W2 <- ifelse(mat.dist <= 3 & mat.dist > 0.5, 1, 0)
170 diag(W2) <- 0
171 W2std = W2 / rowSums(W2)
172 W2std[!is.finite(W2std)] <- 0
173
174 W3 <- ifelse(mat.dist <= 6 & mat.dist > 3, 1, 0)
175 diag(W3) <- 0
176 W3std = W3 / rowSums(W3)
177 W3std[!is.finite(W3std)] <- 0
178 Wstd <- list(W1std, W2std, W3std)
179
180 pdf("VizinhancaModelo.pdf", width=11, height=5)
181 par(mar=c(0,0,3,0), mfrow=c(1,3))
182 item=c("(a) Ordem 1", "(b) Ordem 2", "(c) Ordem 3")
183
184 for (j in 1:length(Wstd)){
185   plot(SP2, border="gray60", lwd=0.3, main=item[j])
186   Wviz <- Wstd[[j]]

```



```

187 for(i in 1:93){
188   indices.destino = which((Wviz[i,] > 0) == T)
189   indice.origem = rep(i, length(indices.destino))
190   for(j in indices.destino){
191     lines(rbind(coords[i, ],
192                 coords[j, ]),
193           col=alpha("blue",0.2), lwd=0.3)
194   }
195   points(coords[indices.destino, ], cex=0.8, pch=16)
196 }
197 }
198 }
199 dev.off()

```

Código C.3 – "Seleção da Melhor Matriz W"

```

1 #0 resíduo até o lag 12 vale 0 por causa dos termos
  autoregressivos
2 vlr.est.dr <- (as.matrix(z.dr)) - (t(fit.residuals))
3 dim(vlr.est.dr)
4
5 #obtem a série antes da padronização
6 vlr.est.dr.despad <- matrix(0,
7                             nrow=dim(vlr.est.dr)[1],
8                             ncol=dim(vlr.est.dr)[2])
9 for (i in 1:nrow(vlr.est.dr)) {
10   vlr.est.dr.despad[i,] = vlr.est.dr[i,]*
11                           sqrt(variancia.z.dr[i]) +
12                           medias.z.dr[i]
13 }
14
15 #Recalcula a função DR para obter o valor original de x
16 tmp <- seq(0, 100, 0.001)
17 vlr.est.dr.despad[1,]
18
19 vlr.est.orig <- matrix(0,
20                       nrow=dim(vlr.est.dr)[1],
21                       ncol=dim(vlr.est.dr)[2])
22
23 for(i in 1:nrow(vlr.est.dr.despad)) {
24   tmp.y <- transf.DR.param(tmp, vlr.media[i], vlr.est.phi[i])
25
26   tmp.z <- sapply(1:length(vlr.est.dr.despad[i,]),
27                 function(x) {
28                   tmp[which.min(abs(vlr.est.dr.despad[i,x] - tmp.y))]
29                 })
30
31   vlr.est.orig[i,] <- tmp.z
32 }
33
34 indiv.df <- data.frame(id=numeric(),
35                        data=numeric(),

```

```

36         Taxa=numeric(),
37         Tipo=character(),
38         Bairro=character())
39 for (j in 1:93){
40   indiv.df <- rbind(indiv.df, data.frame(
41     id=j,
42     data = c(eixo.x[13:144], eixo.x[13:144]),
43     Taxa=c(as.numeric(tmp.id.t.atual[j,13:144]),
44           as.numeric(vlr.est.orig[j, 13:144])),
45     Tipo=c(rep("Observado", 144-13+1),
46           rep("Ajustado", 144-13+1)),
47     Bairro=as.character(paste0(dat.zon$regiao[j], "-",
48                               dat.id$Bairro[j]))
49   ))
50 }
51
52
53 dat.zon.bairro <- dat.zon
54 dat.zon.bairro$Bairro <- as.character(
55   paste0(dat.zon$regiao,
56         "-", dat.id$Bairro))
57
58 nova.ordem <- arrange(dat.zon.bairro, regiao, id)$Bairro
59
60 indiv.df$Bairro <- factor(indiv.df$Bairro, levels=nova.ordem)
61
62 table(dat.zon$regiao)
63 #selecao.ids = subset(dat.zon, regiao=="Centro-Sul")$id
64
65 zonas.ordenadas = sort.int(as.numeric(dat.zon$regiao),
66                             index.return=T)$ix
67 lista.qqplot=c(seq(1, 94, 7*3), 94)
68
69 selecao.ids = zonas.ordenadas[lista.qqplot[5]:
70                               (lista.qqplot[5+1]-1)]
71 tmp=subset(indiv.df, id %in% selecao.ids)
72 table(tmp$Bairro)
73
74 for (i in 1:(length(lista.qqplot)-1)) {
75   selecao.ids = zonas.ordenadas[lista.qqplot[i]:
76                                   (lista.qqplot[i+1]-1)]
77   ggplot(data=subset(indiv.df, id %in% selecao.ids),
78         aes(x=data, y=Taxa, colour=factor(Tipo))) +
79   geom_line(aes(linetype=factor(Tipo))) +
80   geom_point()+
81   theme_bw()+
82   facet_wrap(~Bairro, ncol=3, scales="free_y")+
83   theme(plot.title = element_blank(),
84         axis.title.x = element_text(vjust = 0.2),
85         axis.title.y = element_text(vjust = 1.2),
86         legend.title=element_blank())+

```

```

87 xlab("Data")+
88 ylab("Taxa de Roubo de Veículo")+
89 scale_colour_manual(values=c("dodgerblue", "gray50"))+
90 scale_linetype_manual(values=c("dashed", "solid"))+
91 scale_x_date(labels = date_format("%Y"),
92             breaks = date_breaks("3 years"),
93             limits=c(as.Date("2003-01-01"),
94                     as.Date("2013-12-01")))
95 #Ajuste para encaixar gráfico no tamanho correto da página
96 if (length(selecao.ids) == 21){
97   ggsave(filename=paste0("EstimadoSTARMA_", i, ".pdf"),
98         width=9, height=11)
99 } else {
100   ggsave(filename=paste0("EstimadoSTARMA_", i, ".pdf"),
101         width=9, height=11*3/7*1.1)
102 }
103 }
104
105 #Gráfico do erro entre observado e estimado
106 ggplot(data=agg.err, aes(x=data, y=erro, colour=erro)) +
107   geom_line() + geom_point() +
108   xlab("Data") +
109   ylab("Erro [(Observado/Ajustado-1) x 100]") +
110   scale_colour_gradientn(colours=c(rev(
111     brewer.pal(9, "YlOrRd")[-c(1,2)]),
112     brewer.pal(9, "YlOrRd")[-c(1,2)]),
113     name="Erro") +
114   scale_y_continuous(breaks=seq(-30, 30, 10),
115                     limits=c(-30, 30))+
116   theme_bw()
117 ggsave("SerieAgregadaAjustadaErroTemporal.pdf",
118       width=7, height=4)

```

Código C.4 – "GGplot"

```

1 #Função que calcula o resíduo de cada ponto da série
2 fun.res=function(DELTA, PARAM_LIST,y){
3
4   #Exemplo de como chamar a função
5   #Os parâmetros serão devolvidos na ordem AR, MA, STAR e STMA
6   #PARAM_LIST = list(
7     #           PARAM_AR=c(1,12),
8     #           PARAM_MA=c(0),
9     #           PARAM_STAR_1=c(1,2,12),
10    #           PARAM_STMA_1=c(1)
11    #)
12
13   count_AR    = sum(PARAM_LIST$PARAM_AR>0)
14   count_MA    = count_AR    + sum(PARAM_LIST$PARAM_MA>0)
15   count_STAR  = count_MA    + sum(PARAM_LIST$PARAM_STAR>0)
16   count_STMA  = count_STAR  + sum(PARAM_LIST$PARAM_STMA>0)
17

```

```

18 index_AR =if (PARAM_LIST$PARAM_AR[1] !=0) {
19     c(1, count_AR) }
20 index_MA =if (PARAM_LIST$PARAM_MA[1] !=0) {
21     c(count_AR+1, count_MA) }
22 index_STAR=if (PARAM_LIST$PARAM_STAR[1] !=0) {
23     c(count_MA+1, count_STAR) }
24 index_STMA=if (PARAM_LIST$PARAM_STMA[1] !=0) {
25     c(count_STAR+1, count_STMA) }
26
27 #inicializa valores dos parâmetros
28 if (PARAM_LIST$PARAM_AR[1] !=0) {
29     eval(parse(text=paste0("AR_",PARAM_LIST$PARAM_AR,
30                             "=DELTA[",
31                             seq(from=index_AR[1],
32                                 to=index_AR[2]),"]"))))
33 }
34
35 if (PARAM_LIST$PARAM_MA[1] !=0) {
36     eval(parse(text=paste0("MA_",PARAM_LIST$PARAM_MA,
37                             "=DELTA[",
38                             seq(from=index_MA[1],
39                                 to=index_MA[2]),"]"))))
40 }
41
42 if (PARAM_LIST$PARAM_STAR[1] !=0) {
43     eval(parse(text=paste0("STAR_",PARAM_LIST$PARAM_STAR,
44                             "=DELTA[",
45                             seq(from=index_STAR[1],
46                                 to=index_STAR[2]),"]"))))
47 }
48
49 if (PARAM_LIST$PARAM_STMA[1] !=0) {
50     eval(parse(text=paste0("STMA_",PARAM_LIST$PARAM_STMA,
51                             "=DELTA[",
52                             seq(from=index_STMA[1],
53                                 to=index_STMA[2]),"]"))))
54 }
55 eps=matrix(0,nrow=nrow(y),ncol=ncol(y))
56
57 for(i in 13:ncol(y)){
58     yhat = paste0(
59         if (PARAM_LIST$PARAM_AR[1] !=0)
60         paste0(paste0("+AR_",PARAM_LIST$PARAM_AR,
61                     paste0("*y[,i-]",PARAM_LIST$PARAM_AR,"]")),
62             collapse=""),
63
64         if (PARAM_LIST$PARAM_MA[1] !=0)
65         paste0(paste0("-MA_",PARAM_LIST$PARAM_MA,
66                     paste0("*eps[,i-]",PARAM_LIST$PARAM_MA,"]")),
67             collapse=""),
68

```

```

69     if (PARAM_LIST$PARAM_STAR[1] != 0)
70       paste0(paste0("+STAR_", PARAM_LIST$PARAM_STAR,
71         paste0("*WW[[2]]%*%
72           y[,i-]", PARAM_LIST$PARAM_STAR, "]")),
73         collapse=""),
74
75     if (PARAM_LIST$PARAM_STMA[1] != 0)
76       paste0(paste0("-STMA_", PARAM_LIST$PARAM_STMA,
77         paste0("*WW[[2]]%*%
78           eps[,i-]", PARAM_LIST$PARAM_STMA, "]")),
79         collapse="")
80   )
81   eps[,i] = y[,i] - (eval(parse(text=yhat)))
82 }
83 return(eps)
84 }
85
86 #Função para obter o resíduo quadrático total do modelo
87 fun.est=function(Delta,PARAM_LIST,y) {
88   return(sum(fun.res(Delta, PARAM_LIST,y)^2))
89 }
90
91 #Otimiza o modelo completo
92 system.time(
93 fit1 <-optim(rep(0,5),
94   fun.est,
95   method="BFGS",
96   y=as.matrix(z.dr),
97   PARAM_LIST=list(PARAM_AR=c(1,12),
98     PARAM_MA=c(1),
99     PARAM_STAR_1=c(1,2),
100     PARAM_STMA_1=c(0)))
101
102 #Retira cada um dos parâmetros do modelo completo para obter
103 #o p-value
104 system.time(
105 fit1_less_MA_1 <-optim(rep(0,4),
106   fun.est,
107   method="BFGS",
108   y=as.matrix(z.dr),
109   PARAM_LIST=list(PARAM_AR=c(1,12),
110     PARAM_MA=c(0),
111     PARAM_STAR_1=c(1,2),
112     PARAM_STMA_1=c(0)))
113
114 system.time(
115 fit_less_STAR_2 <-optim(rep(0,4),
116   fun.est,
117   method="BFGS",
118   y=as.matrix(z.dr),
119   PARAM_LIST=list(PARAM_AR=c(1,12),

```

```

120                                     PARAM_MA=c(1),
121                                     PARAM_STAR_1=c(1),
122                                     PARAM_STMA_1=c(0)))
123
124 system.time(
125 fit_less_STAR_1 <-optim(rep(0,4),
126                         fun.est,
127                         method="BFGS",
128                         y=as.matrix(z.dr),
129                         PARAM_LIST=list(PARAM_AR=c(1,12),
130                                         PARAM_MA=c(1),
131                                         PARAM_STAR_1=c(2),
132                                         PARAM_STMA_1=c(0)))
133
134 system.time(
135 fit_less_AR_12 <-optim(rep(0,4),
136                       fun.est,
137                       method="BFGS",
138                       y=as.matrix(z.dr),
139                       PARAM_LIST=list(PARAM_AR=c(1),
140                                       PARAM_MA=c(1),
141                                       PARAM_STAR_1=c(1,2),
142                                       PARAM_STMA_1=c(0)))
143
144 system.time(
145 fit_less_AR_1 <-optim(rep(0,4),
146                      fun.est,
147                      method="BFGS",
148                      y=as.matrix(z.dr),
149                      PARAM_LIST=list(PARAM_AR=c(12),
150                                      PARAM_MA=c(1),
151                                      PARAM_STAR_1=c(1,2),
152                                      PARAM_STMA_1=c(0)))
153
154 #p-value do MA(1)
155 x=(163*93-5)*(fit1_less_MA_1$value-fit1$value)/fit1_less_MA_1$
    value;x
156 pf(x, df1=1, df2=163*93-5, lower.tail=F)
157
158 #p-value do AR(1)
159 x=(163*93-5)*(fit_less_AR_1$value-fit1$value)/fit_less_AR_1$
    value;x
160 pf(x, df1=1, df2=163*93-5, lower.tail=F)
161
162 #p-value do AR(12)
163 x=(163*93-5)*(fit_less_AR_12$value-fit1$value)/fit_less_AR_12$
    value;x
164 pf(x, df1=1, df2=163*93-5, lower.tail=F)
165
166 #p-value do AR Espacial(1)
167 x=(163*93-5)*(fit_less_STAR_1$value-fit1$value)/fit_less_STAR_1

```

```

    $value;x
168 pf(x, df1=1, df2=163*93-5, lower.tail=F)
169
170 #p-value do AR Espacial(2)
171 x=(163*93-5)*(fit_less_STAR_2$value-fit1$value)/fit_less_STAR_2
    $value;x
172 pf(x, df1=1, df2=163*93-5, lower.tail=F)
173
174
175 #Cálculo da variância
176 fit.residuals <- fun.res(fit1$par,
177                           PARAM_LIST=list(PARAM_AR=c(1,12),
178                                             PARAM_MA=c(1),
179                                             PARAM_STAR_1=c(1,2),
180                                             PARAM_STMA_1=c(0)),
181                           y=as.matrix(z.dr))
182 sum(fit.residuals^2)
183
184 dim(fit.residuals)
185
186 fit.residuals <- t(fit.residuals)
187
188 tmp.stacf <- stacf(fit.residuals, WW, plot=F)
189 tmp.stpacf <- stpacf(fit.residuals, WW, plot=F)
190
191 xparcial.melt = melt(tmp.stpacf)
192 xparcial.melt$Var1=as.numeric(xparcial.melt$Var1)
193 xparcial.melt$Var2=as.numeric(xparcial.melt$Var2)-1
194 xparcial.melt = subset(xparcial.melt, Var2 <= 4)
195 texto.lag.espacial <- c("Ordem Espacial 0",
196                          "Ordem Espacial 1",
197                          "Ordem Espacial 2",
198                          "Ordem Espacial 3")
199
200 xparcial.melt$texto <- texto.lag.espacial[xparcial.melt$Var2+1]
201
202 ggplot(data=xparcial.melt, aes(x=Var1, y=value))+
203   geom_bar(stat = "identity",
204            position = "identity", width=0.5)+
205   geom_hline(aes(xintercept = 0))+
206   facet_wrap(~texto)+
207   ylim(-1, 1)+
208   theme_bw()+
209   theme(plot.title = element_blank(),
210         axis.title.x = element_text(vjust = 0.2),
211         axis.title.y = element_text(vjust = 1.2))+
212   xlab("Lag Temporal")+ylab("Autocorrelação Parcial")
213 ggsave("STPACF_Res_STAR_1_1.pdf", width=10, height=6)
214
215 xacf.melt = melt(tmp.stacf)
216 xacf.melt$Var1=as.numeric(xacf.melt$Var1)

```

```

217 xacf.melt$Var2=as.numeric(xacf.melt$Var2)-1
218 xacf.melt = subset(xacf.melt, Var2 < 4)
219 xacf.melt$texto <-texto.lag.espacial[xacf.melt$Var2+1]
220
221 ggplot(data=xacf.melt, aes(x=Var1, y=value))+
222   geom_bar(stat = "identity",
223           position = "identity", width=0.5)+
224   geom_hline(aes(xintercept = 0))+
225   facet_wrap(~texto)+
226   ylim(-1, 1)+
227   theme_bw()+
228   theme(plot.title = element_blank(),
229         axis.title.x = element_text(vjust = 0.2),
230         axis.title.y = element_text(vjust = 1.2))+
231   xlab("Lag Temporal")+ylab("Autocorrelação")
232 ggsave("STACF_Res_STAR_1_1.pdf", width=10, height=6)
233
234 zonas.ordenadas = sort.int(as.numeric(dat.zon$regiao),
235                             index.return=T)$ix
236 lista.qqplot=c(seq(1, 93, 6*4), 94)
237 contador = 1
238 for (j in 1:(length(lista.qqplot)-1)){
239   pdf(paste0("STARMA_Final_ZDR", j, ".pdf"),
240       width=8.27, height=11.69)
241   par(mfrow=c(6,4), mar=c(4,4,2,2), mgp=c(2.2,1,0))
242   for (i in lista.qqplot[j]:(lista.qqplot[j+1]-1)){
243     qqPlot(fit.residuals[-(1:12),zonas.ordenadas[i]],
244           pch=16, lwd=1,
245           ylab="Resíduos",
246           xlab="Distribuição Normal")
247     title(main=paste0(dat.id$Bairro[zonas.ordenadas[i]]),
248           cex.main=1)
249     contador = contador+1
250   }
251   dev.off()
252 }

```

Código C.5 – "Estimação do STARMA"

```

1 #Recalcula o valor original da serie (taxa de roubo) a
2 #partir da serie transformada e modelada
3
4 #Exemplos da inversão
5 range(as.matrix(tmp.id.t.atual))
6 dim(as.matrix(tmp.id.t.atual))
7
8 x.seq.ex <- c(c(0.1, 0.5, 1), seq(2, 40, 2))
9 y.seq.ex <- transf.DR.param(x.seq.ex,
10                             vlr.media[51],
11                             vlr.est.phi[51])
12 plot(y.seq.ex, x.seq.ex)
13 plot(x.seq.ex, y.seq.ex)

```



```

14 plot(as.numeric(tmp.id.t.atual[1,]), type="l")
15
16 qplot(x.seq.ex, y.seq.ex) +
17   geom_point() + geom_line() +
18   ylab("Variável Transformada") +
19   xlab("Variável Original") +
20   theme_bw() +
21   geom_segment(aes(x = 0, y = 1.621576,
22                     xend = 20-0.5, yend = 1.621576),
23                 colour = "tomato2",
24                 arrow=arrow(length = unit(0.03, "npc")))) +
25   geom_segment(aes(x = 20, y = 1.621576-0.1,
26                     xend = 20, yend = -4.17),
27                 colour = "tomato2",
28                 arrow=arrow(length = unit(0.03, "npc")))
29 ggsave("FuncaoInversaDR.pdf", width=7, height=4)
30
31 vlr.est.dr <- (as.matrix(z.dr)) - (t(fit.residuals))
32
33 #obtem a série antes da padronização
34 vlr.est.dr.despad<-matrix(0,
35                            nrow=dim(vlr.est.dr)[1],
36                            ncol=dim(vlr.est.dr)[2])
37
38 for (i in 1:nrow(vlr.est.dr)) {
39   vlr.est.dr.despad[i,] = vlr.est.dr[i,] *
40                           sqrt(variancia.z.dr[i]) +
41                           medias.z.dr[i]
42 }
43
44 #recalcula a função DR para obter o valor original de x
45 tmp <- seq(0, 100, 0.001)
46 vlr.est.orig <- matrix(0,
47                        nrow=dim(vlr.est.dr)[1],
48                        ncol=dim(vlr.est.dr)[2])
49 for(i in 1:nrow(vlr.est.dr.despad)) {
50   tmp.y <- transf.DR.param(tmp, vlr.media[i], vlr.est.phi[i])
51
52   tmp.z <- sapply(1:length(vlr.est.dr.despad[i,]),
53                   function(x){
54                     tmp[which.min(abs(vlr.est.dr.despad[i,x] - tmp.y))]
55                   })
56   vlr.est.orig[i,] <- tmp.z
57 }
58
59 indiv.df <- data.frame(id=numeric(),
60                        data=numeric(),
61                        Taxa=numeric(),
62                        Tipo=character(),
63                        Bairro=character())
64 for (j in 1:93){

```

```

65 indiv.df <- rbind(indiv.df, data.frame(
66   id=j,
67   data = c(eixo.x[13:144], eixo.x[13:144]),
68   Taxa = c(as.numeric(tmp.id.t.atual[j,13:144]),
69            as.numeric(vlr.est.orig[j, 13:144])),
70   Tipo=c(rep("Observado", 144-13+1),
71          rep("Ajustado", 144-13+1)),
72   Bairro=as.character(paste0(dat.zon$regiao[j], "-",
73                               dat.id$Bairro[j]))
74 ))
75 }
76
77 dat.zon.bairro <- dat.zon
78 dat.zon.bairro$Bairro <- as.character(paste0(
79   dat.zon$regiao, "-", dat.id$Bairro))
80 nova.ordem <- arrange(dat.zon.bairro, regiao, id)$Bairro
81
82 indiv.df$Bairro <- factor(indiv.df$Bairro, levels=nova.ordem)
83
84 table(dat.zon$regiao)
85
86 zonas.ordenadas = sort.int(as.numeric(dat.zon$regiao),
87                             index.return=T)$ix
88 lista.qqplot=c(seq(1, 94, 7*3), 94)
89
90 selecao.ids = zonas.ordenadas[lista.qqplot[5]:
91                               (lista.qqplot[5+1]-1)]
92 tmp=subset(indiv.df, id %in% selecao.ids)
93 table(tmp$Bairro)
94
95 indiv.df$Tipo <- factor(indiv.df$Tipo,
96                         levels=c("Observado", "Ajustado"))
97 head(indiv.df)
98
99 for (i in 1:(length(lista.qqplot)-1)) {
100   selecao.ids = zonas.ordenadas[lista.qqplot[i]:
101                                 (lista.qqplot[i+1]-1)]
102   m<-ggplot(data=subset(indiv.df, id %in% selecao.ids),
103             aes(x=data, y=Taxa, colour=factor(Tipo))) +
104     geom_line(aes(linetype=factor(Tipo)), size=0.5) +
105     theme_bw()+
106     facet_wrap(~Bairro, ncol=3, scales="free_y")+
107     theme(plot.title = element_blank(),
108           axis.title.x = element_text(vjust = 0.2),
109           axis.title.y = element_text(vjust = 1.2),
110           legend.title=element_blank(),
111           legend.key = element_blank())+
112     xlab("Data")+
113     ylab("Taxa de Roubo de Veículo")+
114     scale_colour_manual(values=c("gray50", "tomato2"))+
115     scale_linetype_manual(values=c("solid", "dashed"))+

```

```

116     scale_x_date(labels = date_format("%Y"),
117                 breaks = date_breaks("3 years"),
118                 limits=c(as.Date("2003-01-01"),
119                         as.Date("2013-12-01")))
120   if (length(selecao.ids) == 21){
121     ggsave(filename=paste0("EstimadoSTARMA_", i, ".pdf"),
122           width=9, height=11)
123   } else {
124     ggsave(filename=paste0("EstimadoSTARMA_", i, ".pdf"),
125           width=9, height=11*3/7*1.1)
126   }
127 }
128
129 mean(as.matrix(tmp.id.t.atual)[1:144])
130 sd(as.matrix(tmp.id.t.atual)[1:144])
131
132 mean(vlr.est.orig)
133
134 #Agrega as séries e faz o grafico
135 agg.obs <- apply(as.matrix(tmp.id.t.atual), 2, sum)[13:144]
136 mean(agg.obs)
137 sd(agg.obs)
138 agg.obs <- apply(as.matrix(tmp.id.t), 2, sum)[13:144]
139 mean(agg.obs)
140 sd(agg.obs)
141 agg.est <- apply(vlr.est.orig, 2, sum)[13:144]
142 mean(agg.est)
143 plot(agg.obs, type="l", ylab="Taxa de Roubo")
144 lines(agg.est, col="dodgerblue")
145
146 plot(agg.obs, agg.est)
147 abline(0,1)
148
149 agg.df <- data.frame(
150     data = c(eixo.x[13:144], eixo.x[13:144]),
151     Taxa=c(as.numeric(agg.obs), as.numeric(agg.est)),
152     Tipo=c(rep("Observado", length(agg.obs)),
153           rep("Ajustado", length(agg.obs)))
154 )
155
156 ggplot(data=agg.df, aes(x=data, y=Taxa, colour=factor(Tipo))) +
157   geom_line(aes(linetype=factor(Tipo))) + geom_point()+
158   theme_bw()+
159   theme(plot.title = element_blank(),
160         axis.title.x = element_text(vjust = 0.2),
161         axis.title.y = element_text(vjust = 1.2),
162         legend.title=element_blank(),
163         legend.key = element_blank())+
164   xlab("Data")+
165   ylab("Taxa de Roubo de Veículo")+
166   scale_colour_manual(values=c("tomato2", "dodgerblue"))+

```

```

166 scale_linetype_manual(values=c("dashed", "solid"))
167
168 ggsave("SerieAgregadaAjustada.pdf", width=9, height=3)
169
170
171 #Calcula o erro de projeção e faz histograma
172 agg.err <- data.frame(data=eixo.x[13:144],
173                       erro=as.numeric((agg.obs/agg.est-1)*100))
174
175 ggplot(agg.err, aes(erro)) +
176   geom_histogram(aes(y=100*(..count..)/sum(..count..)),
177                 binwidth=5) +
178   scale_x_continuous(breaks=seq(-25, 30, 5), limit=c(-25,30))+
179   scale_y_continuous(breaks=seq(0, 30, 5), limit=c(0,30))+
180   xlab("Erro [(Observado/Ajustado-1) x 100]") +
181   ylab("Percentual (%)") +
182   theme_bw()
183 ggsave(SerieAgregadaAjustadaHistErro.pdf, width=7, height=4)
184
185 quantile(agg.err$erro, probs=seq(0, 1, 0.1))
186 quantile(agg.err$erro, probs=c(0.025, 0.05, 0.95, 0.975))
187 quantile(agg.err$erro, probs=c(0.25, 0.75))

```

Código C.6 – "Recupera Valores Originais após Transformação Deviance Residual"

```

1 fit1=c()
2 fit1$value = 6228.062
3 fit1$par = c(0.80409012, 0.04687098, 0.51406362,
4             0.22849816, -0.10123630)
5
6 #####
7 ### CUSUM DIRECIONAL - ARL0
8 sigmafit <- matrix(0,93,93)
9 zsim <- matrix(0, nrow=93, ncol=144)
10 sigma2 = (1/(144*93))*fit1$value
11 diag(sigmafit) <- sigma2
12 sigmafit.inv <- sigmafit
13 diag(sigmafit.inv) <- 1/diag(sigmafit)
14
15 hseq = c(0.1, 0.25, 0.5, 0.75, 1, 1.5, 2, 3, 4, 5, 7.5, 10,
16         15, 20, 25, 30, 40, 50, 60, 70, 80, 90, 100)
17 hseq.n <- length(hseq)
18 k.mv.seq <- seq(0.1, 1, 0.05)
19 list.sim.cenarios.arl <- list()
20
21 for(i in 1:length(k.mv.seq)) {
22   list.sim.cenarios.arl[[i]] <- list(
23     k.mv=k.mv.seq[i],
24     k=NA,
25     h=hseq,
26     arl=rep(NA,hseq.n),
27     arl.sd=rep(NA,hseq.n),

```



```

79         zsim[,t]))))
80
81     cusum.tmp <- ((lista.sim.cusum[,t-1] + zsim[,t]) *
82                 (1-k/dt)) * (dt > k)
83
84     lista.sim.cusum[,t] <- ifelse(cusum.tmp > 0,
85                                 cusum.tmp, 0)
86
87     Ct <- sqrt(t(lista.sim.cusum[,t]) %*%
88                sigmafit.inv %*% lista.sim.cusum[,t])
89     if (Ct > h) {
90         arl.sim[r] <- t-12
91         break
92     }
93 }
94
95 if (r == 1000) {
96     if (sum(is.na(arl.sim)) > 500) {
97         break
98     } else if (mean(arl.sim, na.rm=T) < 5 ||
99               mean(arl.sim, na.rm=T) > 40) {
100         break
101     }
102 }
103 }
104
105 list.sim.cenarios.arl[[i]]$arl[j] <-
106     mean(arl.sim, na.rm=T)
107
108 list.sim.cenarios.arl[[i]]$arl.sd[j] <-
109     sd(arl.sim, na.rm=T)
110
111 list.sim.cenarios.arl[[i]]$arl.p05[j] <-
112     as.numeric(quantile(arl.sim, probs=0.05, na.rm=T))
113
114 list.sim.cenarios.arl[[i]]$arl.p50[j] <-
115     as.numeric(quantile(arl.sim, probs=0.50, na.rm=T))
116
117 list.sim.cenarios.arl[[i]]$arl.p95[j] <-
118     as.numeric(quantile(arl.sim, probs=0.95, na.rm=T))
119
120 list.sim.cenarios.arl[[i]]$arl.miss[j] <-
121     sum(is.na(arl.sim))
122
123 cat((paste0("Sim: (", i, ", ", j, ")", " - Com ARL0: ",
124             round(list.sim.cenarios.arl[[i]]$arl[j], 2),
125             " - ", Sys.time(), "\n")))
126
127 if(sum(!is.na(list.sim.cenarios.arl[[i]]$arl[1:j])) > 1) {
128     if(max(list.sim.cenarios.arl[[i]]$arl[1:j], na.rm=T) > 30) {
129         cat((paste0("Interrupção da busca nesta lista - ",

```

```

130         Sys.time(), "\n"))
131     break
132 }
133 }
134
135 if (!is.na(list.sim.cenarios.arl[[i]]$arl[j])){
136     if (list.sim.cenarios.arl[[i]]$arl[j] < 5) {
137         cat((paste0("Avança para o próximo item da lista - ",
138             Sys.time(), "\n")))
139         next
140     }
141 }
142 }
143 }
144 fim <- Sys.time()
145 fim - inicio
146
147 #####Inicia segunda fase com ajuste do Spline
148 list.sim.cenarios.arl.poly <- list()
149
150 retorna.indice <- function(x){
151     a = list.sim.cenarios.arl[[x]]$arl
152     a = a[!is.na(a)]
153     b = sign(a - 20)
154     indice <- which.min(b[1:(length(b)-1)] - b[2:length(b)])
155     return(indice)
156 }
157
158 for (cen in 1:length(list.sim.cenarios.arl)){
159
160     ri <- retorna.indice(cen)
161     va=max(ri-2, 1)
162     vb=min(ri+2, length(list.sim.cenarios.arl[[cen]]$h))
163
164     lim.spline <- list.sim.cenarios.arl[[cen]]$h[ri:(ri+1)]
165
166     x = list.sim.cenarios.arl[[cen]]$h[va:vb]
167     y = list.sim.cenarios.arl[[cen]]$arl[va:vb]
168
169     x = x[!is.na(y)]
170     y = y[!is.na(y)]
171
172     if (length(x) > 3) {
173
174         fit.spline <- smooth.spline(x, y)
175         xlin = seq(range(x)[1], range(x)[2], 0.001)
176
177         r=which.min(abs(predict(fit.spline, xlin)$y-20))
178
179         u = xlin[r]*seq(from=0.98, to = 1.02, by = 0.005)
180

```

```

181 h.tmp = u[u > lim.spline[1] & u < lim.spline[2]]
182 h.tmp.n = length(h.tmp)
183
184 list.sim.cenarios.arl.poly[[cen]] <- list(
185   k.mv=list.sim.cenarios.arl[[cen]]$k.mv,
186   h= h.tmp,
187   arl=rep(NA, h.tmp.n),
188   arl.sd=rep(NA, h.tmp.n),
189   arl.p05=rep(NA, h.tmp.n),
190   arl.p50=rep(NA, h.tmp.n),
191   arl.p95=rep(NA, h.tmp.n),
192   arl.miss=rep(NA, h.tmp.n))
193 }
194 }
195
196 set.seed(2016)
197 inicio <- Sys.time()
198 replicas=5000
199 m1 = length(list.sim.cenarios.arl.poly)
200 for (i in seq(1,m1)){
201
202   k.mv = list.sim.cenarios.arl.poly[[i]]$k.mv
203
204   k = sqrt(t(as.matrix(rep(k.mv, 93))) %*%
205             sigmafit.inv %*%
206             as.matrix(rep(k.mv, 93)))
207
208   list.sim.cenarios.arl.poly[[i]]$k <- k
209
210   m2 = length(list.sim.cenarios.arl.poly[[i]]$h)
211   for (j in seq(1,m2)){
212
213     h=list.sim.cenarios.arl.poly[[i]]$h[j]
214     arl.sim <- rep(NA, replicas)
215
216     for(r in seq(1,replicas)){
217
218       #Inicializa o modelo STARMA
219       eps <- matrix(mvrnorm(n=144,
220                             mu=rep(0, 93),
221                             Sigma=sigmafit),
222                     ncol=144)
223
224       eps[,1:12] <- 0
225       zsim[,1:12]<- 0
226
227       #Inicializa o CUSUM
228       lista.sim.cusum <- matrix(0, nrow=93, ncol=144)
229       Ct <- 0
230
231       for(t in seq(13,144)) {
232         #Simula Z(t) por Monte Carlo

```



```

232     zsim[,t] <- fit1$par[1] * zsim[,t-1] +
233               fit1$par[2] * zsim[,t-12] +
234               fit1$par[4] * WW[[2]] %*% zsim[,t-1] +
235               fit1$par[5] * WW[[2]] %*% zsim[,t-2] -
236               fit1$par[3] * eps[,t-1] + eps[,t]
237
238     dt <- sqrt(t(as.matrix((lista.sim.cusum[,t-1] +
239                           zsim[,t]))) %*%
240               sigmafit.inv %*%
241               as.matrix((lista.sim.cusum[,t-1] +
242                           zsim[,t])))
243
244     cusum.tmp <- ((lista.sim.cusum[,t-1] + zsim[,t]) *
245                 (1-k/dt)) * (dt > k)
246
247     lista.sim.cusum[,t] <- ifelse(cusum.tmp > 0,
248                                   cusum.tmp, 0)
249
250     Ct <- sqrt(t(lista.sim.cusum[,t]) %*%
251               sigmafit.inv %*%
252               lista.sim.cusum[,t])
253
254     if (Ct > h) {
255       arl.sim[r] <- t-12
256       break
257     }
258   }
259
260   if (r == 2500) {
261     if (sum(is.na(arl.sim)) > 500) {
262       break
263     } else if (mean(arl.sim, na.rm=T) < 19 ||
264               mean(arl.sim, na.rm=T) > 21) {
265       break
266     }
267   }
268 }
269
270 list.sim.cenarios.arl.poly[[i]]$arl[j] <-
271   mean(arl.sim, na.rm=T)
272
273 list.sim.cenarios.arl.poly[[i]]$arl.sd[j] <-
274   sd(arl.sim, na.rm=T)
275
276 list.sim.cenarios.arl.poly[[i]]$arl.p05[j] <-
277   as.numeric(quantile(arl.sim, probs=0.05, na.rm=T))
278
279 list.sim.cenarios.arl.poly[[i]]$arl.p50[j] <-
280   as.numeric(quantile(arl.sim, probs=0.50, na.rm=T))
281
282 list.sim.cenarios.arl.poly[[i]]$arl.p95[j] <-

```

```

283         as.numeric(quantile(arl.sim, probs=0.95, na.rm=T))
284
285     list.sim.cenarios.arl.poly[[i]]$arl.miss[j]<-
286         sum(is.na(arl.sim))
287
288     cat((paste0("Sim: (", i, ",", j, ")", " - Com ARL0: ",
289         round(list.sim.cenarios.arl.poly[[i]]$arl[j],2),
290         " - ", Sys.time(), "\n")))
291
292     if(sum(!is.na(list.sim.cenarios.arl.poly[[i]]$arl[1:j]))>1)
293     {
294         #dois pontos acima de 20, então interrompe
295         if(min(list.sim.cenarios.arl.poly[[i]]$arl[(j-1):j],
296             na.rm=T) > 20) {
297             cat((paste0("Interrupção da busca nesta lista,
298             já ultrapassou o ARL0 - ", Sys.time(), "\n")))
299             break
300         }
301
302         #diferença pequena, então acelera
303         if(abs(diff(list.sim.cenarios.arl.poly[[i]]$arl[(j-1):j],
304             na.rm=T) < 0.1)
305             & list.sim.cenarios.arl.poly[[i]]$arl[j] < 19.5) {
306             cat((paste0("Avança para o próximo nível - ",
307             Sys.time(), "\n")))
308             next()
309         }
310     }
311
312     if (!is.na(list.sim.cenarios.arl.poly[[i]]$arl[j])){
313         if (list.sim.cenarios.arl.poly[[i]]$arl[j] < 5) {
314             cat((paste0("Avança para o próximo item da lista - ",
315             Sys.time(), "\n")))
316             next
317         }
318     }
319
320
321 }
322 }
323 fim <- Sys.time()
324 fim - inicio
325
326 ##### Simula ARL1
327 #Compila os resultados em uma tabela
328 resultados.cusum.poly = rbind.fill(
329     lapply(1:length(list.sim.cenarios.arl.poly), function(x) {
330         data.frame(
331             k.mv = list.sim.cenarios.arl.poly[[x]]$k.mv,
332             h = list.sim.cenarios.arl.poly[[x]]$h,

```

```

333     arl = list.sim.cenarios.arl.poly[[x]]$arl,
334     arl.sd = list.sim.cenarios.arl.poly[[x]]$arl.sd,
335     arl.p05 = list.sim.cenarios.arl.poly[[x]]$arl.p05,
336     arl.p50 = list.sim.cenarios.arl.poly[[x]]$arl.p50,
337     arl.p95 = list.sim.cenarios.arl.poly[[x]]$arl.p95)
338   )))
339
340 resultados.cusum.poly.subset <- subset(resultados.cusum.poly,
341                                       abs(arl-20)<0.20)
342
343 list.sim.cenarios.arl.1 <- list()
344
345 nshift <- length(seq(0.25,2,0.25))
346
347 for(i in 1:nrow(resultados.cusum.poly.subset)) {
348   list.sim.cenarios.arl.1[[i]] <- list(
349     k.mv=resultados.cusum.poly.subset$k.mv[i],
350     h=resultados.cusum.poly.subset$h[i],
351     arl0=resultados.cusum.poly.subset$arl[i],
352     shift=seq(0.25,2,0.25),
353     arl=rep(NA, nshift),
354     arl.sd=rep(NA, nshift),
355     arl.p05=rep(NA, nshift),
356     arl.p50=rep(NA, nshift),
357     arl.p95=rep(NA, nshift),
358     arl.miss=rep(NA, nshift))
359 }
360
361 set.seed(2016)
362 inicio <- Sys.time()
363 replicas=5000
364 m1=length(list.sim.cenarios.arl.1)
365 for (i in seq(1,m1)){
366
367   k.mv = list.sim.cenarios.arl.1[[i]]$k.mv
368   k = sqrt(t(as.matrix(rep(k.mv, 93))) %*%
369           sigmafit.inv %*%
370           as.matrix(rep(k.mv, 93)))
371
372   list.sim.cenarios.arl.1[[i]]$k <- k
373
374   h=list.sim.cenarios.arl.1[[i]]$h
375   m2=length(list.sim.cenarios.arl.1[[i]]$shift)
376
377   for (j in seq(1,m2)){
378
379     arl.sim <- rep(NA, replicas)
380
381     for(r in seq(1,replicas)){
382
383       #Inicializa STARMA

```

```

384     eps <- matrix(mvrnorm(n=144,
385                        mu=rep(0, 93),
386                        Sigma=sigmafit),
387                  ncol=144) +
388      list.sim.cenarios.arl.1[[i]]$shift[j]*
389      sqrt(sigma2)
390
391     eps[,1:12] <- 0
392     zsim[,1:12]<- 0
393
394     #Inicializa o MCUSUM
395     lista.sim.cusum <- matrix(0, nrow=93, ncol=144)
396     Ct <- 0
397
398     for(t in seq(13,144)) {
399       #Simula Z(t) por Monte Carlo
400       zsim[,t] <- fit1$par[1] * zsim[,t-1] +
401         fit1$par[2] * zsim[,t-12] +
402         fit1$par[4] * WW[[2]] %*% zsim[,t-1] +
403         fit1$par[5] * WW[[2]] %*% zsim[,t-2] -
404         fit1$par[3] * eps[,t-1] + eps[,t]
405
406       dt <- sqrt(t(as.matrix((lista.sim.cusum[,t-1] +
407                             zsim[,t]))) %*%
408                  sigmafit.inv %*%
409                  as.matrix((lista.sim.cusum[,t-1] +
410                             zsim[,t])))
411
412       cusum.tmp <- ((lista.sim.cusum[,t-1] + zsim[,t])*
413                    (1-k/dt))*(dt > k)
414
415       lista.sim.cusum[,t] <- ifelse(cusum.tmp > 0,
416                                    cusum.tmp, 0)
417
418       Ct <- sqrt(t(lista.sim.cusum[,t]) %*%
419                  sigmafit.inv %*%
420                  lista.sim.cusum[,t])
421
422       if (Ct > h) {
423         arl.sim[r] <- t-12
424         break
425       }
426
427     }
428   }
429
430   list.sim.cenarios.arl.1[[i]]$arl[j] <-
431     mean(arl.sim, na.rm=T)
432
433   list.sim.cenarios.arl.1[[i]]$arl.sd[j] <-
434     sd(arl.sim, na.rm=T)

```

```

435
436   list.sim.cenarios.arl.1[[i]]$arl.p05[j] <-
437       as.numeric(quantile(arl.sim, probs=0.05, na.rm=T))
438
439   list.sim.cenarios.arl.1[[i]]$arl.p50[j] <-
440       as.numeric(quantile(arl.sim, probs=0.50, na.rm=T))
441
442   list.sim.cenarios.arl.1[[i]]$arl.p95[j] <-
443       as.numeric(quantile(arl.sim, probs=0.95, na.rm=T))
444
445   list.sim.cenarios.arl.1[[i]]$arl.miss[j] <-
446       sum(is.na(arl.sim))
447
448   cat((paste0("MCUSUM ARL1 Sim: (", i, ", ", j, ") ",
449               " - ", Sys.time(), "\n")))
450 }
451 }
452 fim <- Sys.time()
453 fim - inicio
454
455 #Compila os resultados em uma tabela
456 cusum.resultados.arl1 = rbind.fill(
457   lapply(1:length(list.sim.cenarios.arl.1),
458     function(x) {
459       data.frame(
460         k.mv = list.sim.cenarios.arl.1[[x]]$k.mv,
461         h = list.sim.cenarios.arl.1[[x]]$h,
462         shift = list.sim.cenarios.arl.1[[x]]$shift,
463         arl0 = list.sim.cenarios.arl.1[[x]]$arl0,
464         arl = list.sim.cenarios.arl.1[[x]]$arl,
465         arl.sd = list.sim.cenarios.arl.1[[x]]$arl.sd,
466         arl.p05 = list.sim.cenarios.arl.1[[x]]$arl.p05,
467         arl.p50 = list.sim.cenarios.arl.1[[x]]$arl.p50,
468         arl.p95 = list.sim.cenarios.arl.1[[x]]$arl.p95)
469     })
470
471 tmp=subset(cusum.resultados.arl1)
472 ggplot(data=tmp, aes(x=h, y=arl, colour=factor(shift))) +
473   geom_line() + geom_point()
474
475 #Lista os menores ARLs dentro de cada shift
476 resultados.ewma.arl1.menores.por.shift = rbind.fill(
477   lapply(unique(cusum.resultados.arl1[[1]]$shift),
478     function(x) {
479       tmp = subset(cusum.resultados.arl1, shift==x & lambda==0.3)
480       return (tmp[which.min(tmp$arl),])
481     })
482
483 ggplot(resultados.ewma.arl1.menores.por.shift,
484   aes(y=arl, x=shift)) +
485   geom_point() +

```

```

486 geom_errorbar(aes(ymin=arl-arl.sd, ymax=arl+arl.sd),
487               width=0.2)+
488 xlab("Shift") + ylab(expression(ARL[1]))+
489 theme_bw()

```

Código C.7 – "Otimização dos parâmetros do MCUSUM"

```

1 sigmafit <- matrix(0,93,93)
2 zsim <- matrix(0, nrow=93, ncol=144)
3 sigma2 = (1/(144*93))*fit1$value
4 diag(sigmafit) <- sigma2
5 sigmafit.inv <- sigmafit
6 diag(sigmafit.inv) <- 1/diag(sigmafit)
7
8 hseq = seq(25,500,25)
9 hseq.n <- length(hseq)
10
11 ewma.list.sim.cenarios.arl <- list()
12
13 m = 1
14 for(i in seq(0.05, 0.5, 0.05)) {
15   ewma.list.sim.cenarios.arl[[m]] <- list(
16     lambda=i,
17     h=hseq,
18     arl=rep(NA,hseq.n),
19     arl.sd=rep(NA,hseq.n),
20     arl.p05=rep(NA,hseq.n),
21     arl.p50=rep(NA,hseq.n),
22     arl.p95=rep(NA,hseq.n),
23     arl.miss=rep(NA,hseq.n))
24   m = m+1
25 }
26
27 #Quanto maior o k.mv, menor tem que ser o h. E quanto maior o k
   .mv, menor tem que ser o h.
28
29 set.seed(2016)
30
31 inicio <- Sys.time()
32 replicas=1000
33 m1 = length(ewma.list.sim.cenarios.arl)
34 for (i in seq(1,m1)){
35
36   m2 = length(ewma.list.sim.cenarios.arl[[i]]$h)
37   for (j in seq(1,m2)){
38
39     h      = ewma.list.sim.cenarios.arl[[i]]$h[j]
40     lambda = ewma.list.sim.cenarios.arl[[i]]$lambda
41
42     arl.sim <- rep(NA, replicas)
43
44     for(r in seq(1,replicas)){

```

```

45
46 #Inicializa o modelo de séries
47 eps <- matrix(mvrnorm(n=144, mu=rep(0, 93),
48                               Sigma=sigmafit),
49                               ncol=144)
50
51 eps[,1:12] <- 0
52 zsim[,1:12]<- 0
53
54 #Inicializa o MEWMA
55 Z.sim.ewma <- matrix(0, nrow=93, ncol=144)
56 Et <- 0
57
58 for(t in seq(13,144)) {
59   #Simula Z(t)
60   zsim[,t] <- fit1$par[1] * zsim[,t-1] +
61               fit1$par[2] * zsim[,t-12] +
62               fit1$par[4] * WW[[2]] %*% zsim[,t-1] +
63               fit1$par[5] * WW[[2]] %*% zsim[,t-2] -
64               fit1$par[3] * eps[,t-1] + eps[,t]
65
66   y <-      lambda * zsim[,t] +
67           (1 - lambda) * Z.sim.ewma[,t-1]
68
69   Z.sim.ewma[,t] <- ifelse(y > 0, y, 0)
70
71   ewma.sigma.z <- (lambda*(1-(1-lambda)^(2*(t-12)))/
72                   (2-lambda)) * sigmafit
73   ewma.sigma.z.inv <- ewma.sigma.z
74   diag(ewma.sigma.z.inv) <- 1/diag(ewma.sigma.z.inv)
75
76   Et <- t(Z.sim.ewma[,t]) %*%
77         ewma.sigma.z.inv %*%
78         Z.sim.ewma[,t]
79
80   if (Et > h) {
81     arl.sim[r] <- t-12
82     break
83   }
84 }
85
86
87 ewma.list.sim.cenarios.arl[[i]]$arl[j] <-
88   mean(arl.sim, na.rm=T)
89
90 ewma.list.sim.cenarios.arl[[i]]$arl.sd[j] <-
91   sd(arl.sim, na.rm=T)
92
93 ewma.list.sim.cenarios.arl[[i]]$arl.p05[j] <-
94   as.numeric(quantile(arl.sim, probs=0.05, na.rm=T))
95

```

```

96 ewma.list.sim.cenarios.arl[[i]]$arl.p50[j] <-
97   as.numeric(quantile(arl.sim, probs=0.50, na.rm=T))
98
99 ewma.list.sim.cenarios.arl[[i]]$arl.p95[j] <-
100   as.numeric(quantile(arl.sim, probs=0.95, na.rm=T))
101
102 ewma.list.sim.cenarios.arl[[i]]$arl.miss[j] <-
103   sum(is.na(arl.sim))
104
105   cat((paste0("MEWMA Sim: (", i, ", ", j, ")",
106               " - ", Sys.time(), "\n")))
107 }
108 }
109 fim <- Sys.time()
110 fim - inicio
111
112 ##Busca refinada pelo spline
113 ewma.list.sim.cenarios.arl.poly <- list()
114
115 retorna.indice <- function(x){
116   a = ewma.list.sim.cenarios.arl[[x]]$arl
117   a = a[!is.na(a)]
118   b = sign(a - 20)
119   indice <- which.min(b[1:(length(b)-1)] - b[2:length(b)])
120   return(indice)
121 }
122
123 for (cen in 1:length(ewma.list.sim.cenarios.arl)){
124
125   ri <- retorna.indice(cen)
126
127   va=max(ri-2, 1)
128   vb=min(ri+2, length(ewma.list.sim.cenarios.arl[[cen]]$h))
129
130   lim.spline <- ewma.list.sim.cenarios.arl[[cen]]$
131                 h[ri:(ri+1)]
132
133   x = ewma.list.sim.cenarios.arl[[cen]]$h[va:vb]
134   y = ewma.list.sim.cenarios.arl[[cen]]$arl[va:vb]
135
136   fit.spline <- smooth.spline(x, y)
137   xlin = seq(range(x)[1], range(x)[2], 0.001)
138
139   r = which.min(abs(predict(fit.spline, xlin)$y-20))
140
141   u = xlin[r]*seq(from=0.95, to = 1.05, by = 0.005)
142
143   h.tmp = u[u > lim.spline[1] & u < lim.spline[2]]
144
145   h.tmp.n = length(h.tmp)
146

```



```

147 ewma.list.sim.cenarios.arl.poly[[cen]] <- list(
148   lambda = ewma.list.sim.cenarios.arl[[cen]]$lambda,
149   h= h.tmp,
150   arl=rep(NA, h.tmp.n),
151   arl.sd=rep(NA, h.tmp.n),
152   arl.p05=rep(NA, h.tmp.n),
153   arl.p50=rep(NA, h.tmp.n),
154   arl.p95=rep(NA, h.tmp.n),
155   arl.miss=rep(NA, h.tmp.n))
156 }
157
158 set.seed(2016)
159
160 inicio <- Sys.time()
161 replicas=5000
162 for (i in seq(1,length(ewma.list.sim.cenarios.arl.poly))) {
163
164   h.n = length(ewma.list.sim.cenarios.arl.poly[[i]]$h)
165
166   lambda = ewma.list.sim.cenarios.arl.poly[[i]]$lambda
167
168   for(j in seq(1,h.n)) {
169
170     h <- ewma.list.sim.cenarios.arl.poly[[i]]$h[j]
171
172     arl.sim <- rep(NA, replicas)
173
174     for(r in seq(1,replicas)) {
175
176       #Inicializa o modelo de séries por Monte Carlo
177       eps <- matrix(mvrnorm(n=144, mu=rep(0, 93),
178                                     Sigma=sigmafit),
179                                     ncol=144)
180       eps[,1:12] <- 0
181       zsim[,1:12] <- 0
182
183       #Inicializa o CUSUM
184       Z.sim.ewma <- matrix(0, nrow=93, ncol=144)
185       Et <- 0
186
187       for(t in seq(13,144)) {
188         #Simula Z(t)
189         zsim[,t] <- fit1$par[1] * zsim[,t-1] +
190                   fit1$par[2] * zsim[,t-12] +
191                   fit1$par[4] * WW[[2]] %*% zsim[,t-1] +
192                   fit1$par[5] * WW[[2]] %*% zsim[,t-2] -
193                   fit1$par[3] * eps[,t-1] + eps[,t]
194
195         y <-          lambda * zsim[,t] +
196              (1 - lambda) * Z.sim.ewma[,t-1]
197

```

```

198     Z.sim.ewma[,t] <- ifelse(y > 0, y, 0)
199
200     ewma.sigma.z <- (lambda*
201                     (1-(1-lambda)^(2*(t-12)))/
202                     (2-lambda)) * sigmafit
203
204     ewma.sigma.z.inv <- ewma.sigma.z
205     diag(ewma.sigma.z.inv) <- 1/diag(ewma.sigma.z.inv)
206
207     Et <- t(Z.sim.ewma[,t]) %*%
208           ewma.sigma.z.inv %*%
209           Z.sim.ewma[,t]
210
211     if (Et > h) {
212         arl.sim[r] <- t-12
213         break
214     }
215 }
216 }
217
218 ewma.list.sim.cenarios.arl.poly[[i]]$arl[j] <-
219     mean(arl.sim, na.rm=T)
220
221 ewma.list.sim.cenarios.arl.poly[[i]]$arl.sd[j] <-
222     sd(arl.sim, na.rm=T)
223
224 ewma.list.sim.cenarios.arl.poly[[i]]$arl.p05[j] <-
225     as.numeric(quantile(arl.sim, probs=0.05, na.rm=T))
226
227 ewma.list.sim.cenarios.arl.poly[[i]]$arl.p50[j] <-
228     as.numeric(quantile(arl.sim, probs=0.50, na.rm=T))
229
230 ewma.list.sim.cenarios.arl.poly[[i]]$arl.p95[j] <-
231     as.numeric(quantile(arl.sim, probs=0.95, na.rm=T))
232
233 ewma.list.sim.cenarios.arl.poly[[i]]$arl.miss[j] <-
234     sum(is.na(arl.sim))
235
236 cat((paste0("MEWMA Poly Sim: (",
237             i,",", j,") com ARL0 = ",
238             ewma.list.sim.cenarios.arl.poly[[i]]$arl[j],
239             " - ", Sys.time(), "\n"))
240     )
241
242 }
243 fim <- Sys.time()
244 fim - inicio
245
246 #####3
247 #####3
248 # Simula ARL1

```

```

249
250
251 #Compila os resultados em uma tabela
252 resultados.ewma.poly = rbind.fill(
253   lapply(1:length(ewma.list.sim.cenarios.arl.poly), function(x)
254     {
255       data.frame(
256         lambda = ewma.list.sim.cenarios.arl.poly[[x]]$lambda,
257         h = ewma.list.sim.cenarios.arl.poly[[x]]$h,
258         arl = ewma.list.sim.cenarios.arl.poly[[x]]$arl,
259         arl.sd = ewma.list.sim.cenarios.arl.poly[[x]]$arl.sd,
260         arl.p05 = ewma.list.sim.cenarios.arl.poly[[x]]$arl.p05,
261         arl.p50 = ewma.list.sim.cenarios.arl.poly[[x]]$arl.p50,
262         arl.p95 = ewma.list.sim.cenarios.arl.poly[[x]]$arl.p95)
263     })
264 resultados.ewma.poly.subset <- subset(resultados.ewma.poly,
265                                       abs(arl-20)<0.20)
266
267 ewma.list.sim.cenarios.arl.1 <- list()
268
269 nshift <- length(seq(0.25,2,0.25))
270
271 for(i in 1:nrow(resultados.ewma.poly.subset)) {
272   ewma.list.sim.cenarios.arl.1[[i]] <- list(
273     lambda=resultados.ewma.poly.subset$lambda[i],
274     h=resultados.ewma.poly.subset$h[i],
275     arl0=resultados.ewma.poly.subset$arl[i],
276     shift=seq(0.25,2,0.25),
277     arl=rep(NA, nshift),
278     arl.sd=rep(NA, nshift),
279     arl.p05=rep(NA, nshift),
280     arl.p50=rep(NA, nshift),
281     arl.p95=rep(NA, nshift),
282     arl.miss=rep(NA, nshift))
283 }
284
285 set.seed(2016)
286
287 inicio <- Sys.time()
288 replicas <- 5000
289 m1 <- length(ewma.list.sim.cenarios.arl.1)
290 for (i in seq(1, m1)){
291
292   m2 <- length(ewma.list.sim.cenarios.arl.1[[i]]$shift)
293   for (j in seq(1, m2)){
294
295     h = ewma.list.sim.cenarios.arl.1[[i]]$h
296     lambda = ewma.list.sim.cenarios.arl.1[[i]]$lambda
297
298     arl.sim <- rep(NA, replicas)

```

```

299
300   for(r in seq(1,replicas)){
301
302       #Inicializa o modelo de séries por Monte Carlo
303       eps <- matrix(mvrnorm(n=144, mu=rep(0, 93),
304                               Sigma=sigmafit), ncol=144) +
305           ewma.list.sim.cenarios.arl.1[[i]]$shift[j]*
306           sqrt(sigma2)
307
308       eps[,1:12] <- 0
309       zsim[,1:12] <- 0
310
311       #Inicializa o CUSUM
312       Z.sim.ewma <- matrix(0, nrow=93, ncol=144)
313       Et <- 0
314
315       for(t in seq(13,144)) {
316           #Simula Z(t)
317           zsim[,t] <- fit1$par[1] * zsim[,t-1] +
318               fit1$par[2] * zsim[,t-12] +
319               fit1$par[4] * WW[[2]] %*% zsim[,t-1] +
320               fit1$par[5] * WW[[2]] %*% zsim[,t-2] -
321               fit1$par[3] * eps[,t-1] + eps[,t]
322
323           y <-          lambda * zsim[,t] +
324               (1 - lambda) * Z.sim.ewma[,t-1]
325
326           Z.sim.ewma[,t] <- ifelse(y > 0, y, 0)
327
328           ewma.sigma.z <- (lambda/(2-lambda)) * sigmafit
329           ewma.sigma.z.inv <- ewma.sigma.z
330           diag(ewma.sigma.z.inv) <- 1/diag(ewma.sigma.z.inv)
331
332           Et <- t(Z.sim.ewma[,t]) %*%
333               ewma.sigma.z.inv %*%
334               Z.sim.ewma[,t]
335           if (Et > h) {
336               arl.sim[r] <- t-12
337               break
338           }
339       }
340   }
341
342   ewma.list.sim.cenarios.arl.1[[i]]$arl[j] <-
343       mean(arl.sim, na.rm=T)
344   ewma.list.sim.cenarios.arl.1[[i]]$arl.sd[j] <-
345       sd(arl.sim, na.rm=T)
346   ewma.list.sim.cenarios.arl.1[[i]]$arl.p05[j] <-
347       as.numeric(quantile(arl.sim, probs=0.05, na.rm=T))
348   ewma.list.sim.cenarios.arl.1[[i]]$arl.p50[j] <-
349       as.numeric(quantile(arl.sim, probs=0.50, na.rm=T))

```

```

350 ewma.list.sim.cenarios.arl.1[[i]]$arl.p95[j] <-
351     as.numeric(quantile(arl.sim, probs=0.95, na.rm=T))
352 ewma.list.sim.cenarios.arl.1[[i]]$arl.miss[j] <-
353     sum(is.na(arl.sim))
354
355     cat((paste0("MEWMA ARL1 Sim: (", i, ",", j, ")",
356               " - ", Sys.time(), "\n")))
357 }
358 }
359 fim <- Sys.time()
360 fim - inicio
361
362 #Compila os resultados em uma tabela
363 ewma.resultados.arl1 = rbind.fill(
364     lapply(1:length(ewma.list.sim.cenarios.arl.1), function(x) {
365         data.frame(
366             lambda = ewma.list.sim.cenarios.arl.1[[x]]$lambda,
367             h = ewma.list.sim.cenarios.arl.1[[x]]$h,
368             shift = ewma.list.sim.cenarios.arl.1[[x]]$shift,
369             arl0 = ewma.list.sim.cenarios.arl.1[[x]]$arl0,
370
371             arl = ewma.list.sim.cenarios.arl.1[[x]]$arl,
372             arl.sd = ewma.list.sim.cenarios.arl.1[[x]]$arl.sd,
373             arl.p05 = ewma.list.sim.cenarios.arl.1[[x]]$arl.p05,
374             arl.p50 = ewma.list.sim.cenarios.arl.1[[x]]$arl.p50,
375             arl.p95 = ewma.list.sim.cenarios.arl.1[[x]]$arl.p95)
376         })
377
378 subset(ewma.resultados.arl1, lambda==0.2 & shift == 1)
379
380 write.table(ewma.resultados.arl1,
381             "ewma_resultados_arl1.txt",
382             sep="|", dec=",", row.names = F)
383
384 tmp=subset(ewma.resultados.arl1)
385 ggplot(data=tmp, aes(x=lambda, y=arl, colour=factor(shift))) +
386     geom_line() + geom_point()
387
388 #Lista os menores ARLs dentro de cada shift
389 resultados.ewma.arl1.menores.por.shift = rbind.fill(
390     lapply(unique(ewma.list.sim.cenarios.arl.1[[1]]$shift),
391             function(x) {
392
393                 tmp = subset(ewma.resultados.arl1, shift==x & lambda==0.3)
394                 return (tmp[which.min(tmp$ar1),])
395             })
396
397 ggplot(resultados.ewma.arl1.menores.por.shift,
398         aes(y=arl, x=shift)) +
399     geom_point() +
400     geom_errorbar(aes(ymin=arl-arl.sd, ymax=arl+arl.sd),

```

```
400         width=0.2)+  
401     xlab("Shift") + ylab(expression(ARL[1])) +  
402     theme_bw()
```

Código C.8 – "Otimização dos parâmetros do MEWMA"