



Trabalho de Conclusão de Curso

**Otimização do método de monitoramento de  
processos em bateladas baseado na dinâmica do  
modelo VAR**

Tobias dos Santos Gomes

18 de julho de 2019

Tobias dos Santos Gomes

**Otimização do método de monitoramento de processos em  
bateladas baseado na dinâmica do modelo VAR**

Trabalho de Conclusão apresentado à comissão de Graduação do Departamento de Estatística da Universidade Federal do Rio Grande do Sul, como parte dos requisitos para obtenção do título de Bacharel em Estatística.

Orientador(a): Profa. Dr. Marcio Valk

Porto Alegre  
Junho de 2019

Tobias dos Santos Gomes

**Otimização do método de monitoramento de processos em  
bateladas baseado na dinâmica do modelo VAR**

Este Trabalho foi julgado adequado para obtenção dos créditos da disciplina Trabalho de Conclusão de Curso em Estatística e aprovado em sua forma final pela Orientador(a) e pela Banca Examinadora.

Orientador(a): \_\_\_\_\_  
Profa. Dr. Marcio Valk, UNICAMP  
Doutor(a) pela Universidade Estadual de Campinas - Campinas, SP

Banca Examinadora:

Prof. Dr. Danilo Marcondes Filho, UFRGS  
Doutor pela Universidade Federal do Rio Grande do Sul - Porto Alegre, RS

Porto Alegre  
Junho de 2019

*“Muitas coisas não ousamos empreender por parecerem difíceis; entretanto, são difíceis porque não ousamos empreendê-las.”. (Sêneca)*

# Agradecimentos

Agradeço a Deus, pela vida e as pessoas que me rodeiam. A Adevar Bueno (in memorian), Guaracy Bomfim Viana (in memorian), Francisco Carlos de Oliveira (in memorian), Manuel Pedro da Conceição (in memorian), pelos exemplos que foram em vida.

Agradeço a meu pai (Valfrei Rodrigues Gomes), que me ensinou a integridade e o respeito ao próximo. A minha mãe, por seu enorme carinho e sua paciência.

Ao meu amor (Janaína Vianna da Conceição) por estar sempre ao meu lado, fazendo-me sempre mais feliz e me apoiando neste processo de tentar sempre melhorar como pessoa.

Ao meu orientador (Marcio Valk), pela parceria, ajuda e paciência. Ao curso de Estatística, por todos ensinamentos que me proporcionou nestes anos.

# Resumo

Este trabalho tem como objetivo estudar e otimizar o método de monitoramento de processos em batelada baseado na dinâmica do modelo VAR proposto por [Marcondes e Valk \(2019\)](#). Nessa abordagem, os autores propõem um método que depende de uma matriz acoplada que é utilizada na estimação das cartas de controle para o processo por bateladas *on-line*. Esse processo de acoplagem consiste em juntar duas partes de séries temporais distintas em que cada parte a ser acoplada tem um tamanho. Nesse estudo, buscamos otimizar os tamanhos das partes que formam a matriz acoplada. Partimos de valores de referência propostos no trabalho original e investigamos outros possíveis valores para os tamanhos da acoplagem e também os tamanhos de cada uma das partes. O desempenho dessa abordagem é comparada com uma abordagem de referência, através do  $ARL_0$  e do  $ARTL$  das cartas  $\mathcal{T}^2$  de Hotelling e  $\mathcal{W}$  de Variância Generalizada. Resultados de um estudo de simulação mostram que, em geral, essa abordagem supera a de referência e que os parâmetros de acoplagem influenciam tanto na velocidade de detecção do descontrole do processo quanto no efeito de comparação.

**Palavras-Chave:** Cartas de Controle, Séries Temporais, Modelo VAR.

# Abstract

This work aims to study and optimize the method of monitoring batch process dynamics based on the VAR model proposed by [Marcondes e Valk \(2019\)](#). In their approach the authors propose a method depending on a coupled array that is used in the estimation of control charts for the online batch process. This coupling method consists of joining two parts of distinct time series in which each part to be coupled has its own size. In this study, we seek to optimize the sizes of the parts that form the coupled array. We start from reference values proposed in the original work and investigate other possible values for the coupling sizes and also the sizes of each part. The performance of this approach is compared to a reference approach, using  $ARL_0$  and  $ARTL$  from  $\mathcal{T}^2$ -Hotelling and  $\mathcal{W}$ -Generalized Variance charts. Results from a simulation study show that, in general, the performance of this approach exceeds the reference and that the coupling parameters influence the rate of detection of out-of-control process and the comparison between the two methodologies can also change.

**Keywords:** Chart Control, Time Series, VAR Models.

# Sumário

<b>1</b>	<b>Introdução</b>	<b>11</b>
<b>2</b>	<b>Cartas de Controle</b>	<b>14</b>
2.1	A Carta de Controle de Hotelling . . . . .	14
2.2	Carta da Variância Generalizada . . . . .	15
2.3	Modelando as Bateladas por Séries Temporais . . . . .	15
<b>3</b>	<b>Metodologia</b>	<b>18</b>
3.1	Modelo VAR . . . . .	18
3.2	Cartas de Controle Baseadas no Monitoramento da Dinâmica do Modelo VAR . . . . .	19
<b>4</b>	<b>Simulação</b>	<b>23</b>
4.1	Resultados . . . . .	24
<b>5</b>	<b>Conclusões</b>	<b>34</b>
	<b>Referências Bibliográficas</b>	<b>34</b>



# Lista de Figuras

Figura 1.1: Representação de uma carta de controle padrão. Os limites são, respectivamente, $\sigma$ , $2\sigma$ e $3\sigma$ . . . . .	12
Figura 1.2: Três bateladas de um processo com duas variáveis observadas em 300 períodos de tempo . . . . .	13
Figura 2.1: Representação da redução de dimensionalidade. Esquema tree-way para two-way. . . . .	16
Figura 3.1: Processos fora de controle. Os tempos em que o processo começa a sair de controle são, respectivamente, 5,50 e 200. . . . .	20
Figura 3.2: Fluxograma da metodologia online e off-line . . . . .	21

## Lista de Tabelas

Tabela 4.1: Resultados do <i>ARTL</i> e <i>ARL</i> <sub>0</sub> para $L_c = 15$ e $L_f=6$ . . . . .	25
Tabela 4.2: Resultados do <i>ARTL</i> e <i>ARL</i> <sub>0</sub> para $L_c = 20$ e $L_f=8$ . . . . .	26
Tabela 4.3: Resultados do <i>ARTL</i> e <i>ARL</i> <sub>0</sub> para $L_c = 25$ e $L_f=10$ . . . . .	27
Tabela 4.4: Resultados do <i>ARTL</i> e <i>ARL</i> <sub>0</sub> para $L_c = 30$ e $L_f=18$ . . . . .	28
Tabela 4.5: Resultados do <i>ARTL</i> e <i>ARL</i> <sub>0</sub> para $L_c = 40$ e $L_f=24$ . . . . .	29
Tabela 4.6: Resultados do <i>ARTL</i> e <i>ARL</i> <sub>0</sub> para $L_c = 30$ e $L_f=12$ . . . . .	30
Tabela 4.7: Resultados do <i>ARTL</i> e <i>ARL</i> <sub>0</sub> para $L_c = 40$ e $L_f=16$ . . . . .	31
Tabela 4.8: Resultados do <i>ARTL</i> e <i>ARL</i> <sub>0</sub> para $L_c = 60$ e $L_f=24$ . . . . .	32
Tabela 4.9: Resultados do <i>ARTL</i> e <i>ARL</i> <sub>0</sub> para $L_c = 70$ e $L_f=28$ . . . . .	33

# 1 Introdução

Processos por bateladas são amplamente usados nas indústrias química, bioquímica, farmacêutica e alimentícia. Este tipo de processo envolve processar os materiais em condições controladas por muitas variáveis. Essas variáveis, em geral contínuas, são observadas periodicamente durante o tempo de duração de uma batelada. Uma batelada consiste num procedimento completo, desde a mistura dos componentes, as reações, até o produto final. Para cada batelada, temos um conjunto de variáveis. Se, ao final, o produto resultante desta batelada for considerado adequado, a trajetória descrita pelas variáveis é considerada uma *trajetória padrão*.

Para monitorar trajetórias, utilizamos as cartas de controle. As cartas de controle foram criadas por [Shewhart \(1925, 1931\)](#) para monitorar e medir a qualidade de um processo produtivo, isto é, sua adequação aos critérios estabelecidos. Com elas, podemos verificar se o processo está sob controle ou não e entender quais são as variáveis decisivas neste processo, o que pode auxiliar na melhoria da qualidade e na redução da variabilidade. Caso os valores da carta de controle, calculada em pontos amostrais de um novo processo, fiquem dentro dos limites de controle, o processo é tido como *sob controle*. Em alguns casos, é necessário a suposição de normalidade da variável medida, permitindo construir limites com propriedades estatísticas conhecidas. Para que esta normalidade seja efetiva, é necessário que a variabilidade externa ao processo em si tenha sido controlada, ou seja, nos termos de controle de qualidade, que as causas especiais estejam controladas, tornando o processo estável e previsível. Assim, pode-se agir sob as causas comuns, reduzindo a variabilidade do processo. Definidos os limites de controle, uma quantidade de valores fora dos limites pode indicar que o processo está fora de controle ou ainda pode ser o que chamamos *alarme falso*. Um alarme falso é quando o valor da carta está fora dos limites de controle, mas o processo permanece sob controle. A Figura 1.1 apresenta um exemplo de uma carta de controle em que os limites de controle são definidos por  $\sigma$ ,  $2\sigma$  e  $3\sigma$ .

Existem diferentes características das variáveis que podem ser monitoradas através de uma carta de controle. [Shewhart \(1925, 1931\)](#) propuseram um conceito de monitoramento visual com limites de controle para detectar mudanças na média do processo ao longo do tempo. Essa abordagem deu início ao monitoramento da média e da variância de um processo. Desde então, contribuições significativas foram feitas e novas metodologias com melhor desempenho foram propostas. No contexto de processos em bateladas é preciso considerar que as variáveis são observadas em cada instante de tempo e em diferentes bateladas. Logo, é necessário monitorar a evolução no tempo, mas também existe uma variabilidade entre bateladas que deve ser

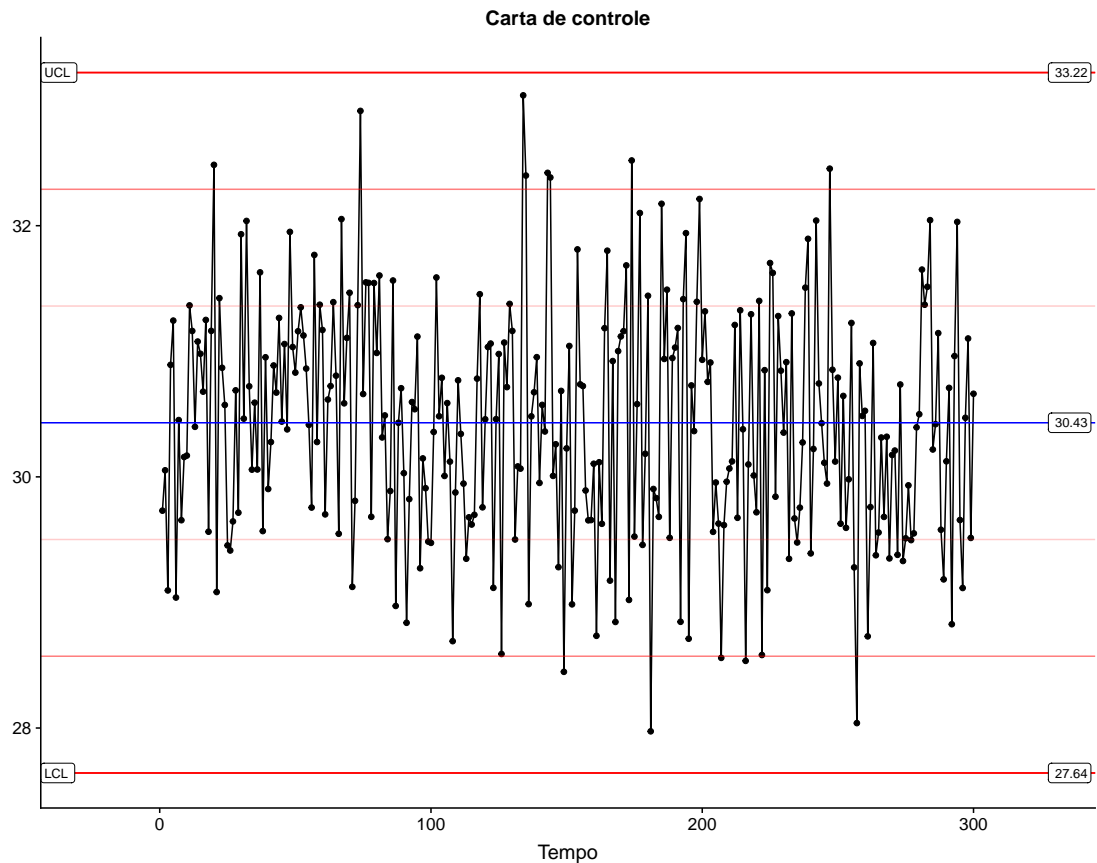


Figura 1.1: Representação de uma carta de controle padrão. Os limites são, respectivamente,  $\sigma$ ,  $2\sigma$  e  $3\sigma$

Fonte: autor

considerada. Trabalhos mais recentes propõem o uso de modelos de séries temporais no contexto do Controle Estatístico Multivariado de Processo, mais especificamente, o Modelo de Vetores Autoregressivos (VAR, do inglês Vector Autoregressive Model). No entanto, todos eles dependem de algum tipo de redução de dimensionalidade. Um exemplo em que 2 variáveis são observadas em 300 períodos de tempo, em 3 repetições do processo (3 bateladas) é apresentado na Figura 1.2.

Uma abordagem inovadora é apresentada no trabalho de [Marcondes e Valk \(2019\)](#). A proposta é monitorar os coeficientes do modelo VAR no lugar das tradicionais cartas baseadas em monitoramento de resíduos. O procedimento é dividido em duas etapas. A primeira, consiste em usar as bateladas históricas sob controle para construir a carta. Em um segundo momento, pode-se monitorar um processo assim que as observações são coletadas (monitoramento *on-line*). A medida em que as observações chegam no tempo, pode-se estimar os coeficientes do modelo VAR usando as informações que estiverem disponíveis até aquele tempo e verificar se o processo encontra-se sob controle. Além disso, também é possível verificar se uma batelada completa está adequada aos padrões históricos observados (monitoramento *off-line*). A carta de Hotelling e a carta da Variância Generalizada são construídas para monitoramento *off-line* e *on-line* de novos lotes.

Estudos de simulação mostram que essa abordagem baseada no monitoramento dos coeficientes do modelo VAR supera a abordagem baseada em resíduos. Ainda sobre o procedimento *on-line*, quando temos poucas observações disponíveis, como

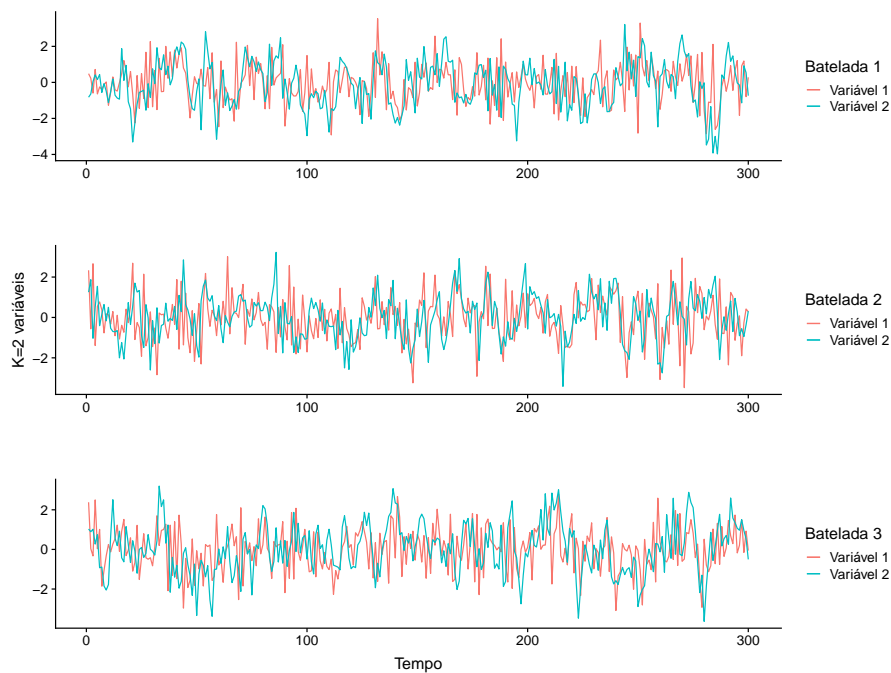


Figura 1.2: Três bateladas de um processo com duas variáveis observadas em 300 períodos de tempo

Fonte: Autor

por exemplo para os tempos  $t=1,2,3,\dots$ , não é razoável estimar os coeficientes do modelo VAR. Marcondes e Valk (2019) propõe o uso de acoplamento de séries temporais. Significa juntar as informações disponíveis até o tempo  $t$  desse processo sob observação com um “pedaço” de uma série que sabidamente está sobre controle. Como o objetivo é observar a dinâmica desse processo, é preciso respeitar a ordem dessas observações. Além disso, é necessário identificar o quanto de informação do processo em observação devemos usar e quantas observações da série de acoplamento devemos utilizar. A teoria sobre as propriedades dos estimadores nessa situação de processos acoplados foi obtida em Marcondes e Valk (2019).

Este trabalho propõe explorar abordagem apresentada no trabalho de Marcondes e Valk (2019) com o objetivo de otimizar as quantidades de cada parcela da acoplagem. Para isto, simularemos diferentes tamanhos da matriz acoplada e diferentes tamanhos para as partes de acoplagem. As perturbações ou descontroles serão inseridas em diferentes tempos do processo. Usaremos medidas de sensibilidade para verificar tanto a quantidade de alarmes falsos quanto a velocidade de detecção dos descontroles causadas pelas perturbações nos dados simulados.

Além desta introdução, este trabalho está organizado da seguinte forma: no Capítulo 2, apresentamos as cartas tradicionais de Hotelling e da Variância Generalizada e detalhamos como é feita a abordagem para séries temporais com e sem redução de dimensionalidade. No Capítulo 3, apresentamos a abordagem baseada em modelos VAR que utiliza acoplamento de séries temporais para construir as cartas de controle. No Capítulo 4, um estudo de simulação é proposto para investigar o desempenho do *método de monitoramento de processos em batelada baseado na dinâmica do modelo VAR*. Por fim, no Capítulo 5, reportamos as conclusões obtidas a partir do estudo de simulação.

## 2 Cartas de Controle

### 2.1 A Carta de Controle de Hotelling

No processo por bateladas temos mais de uma variável a ser controlada. Logo, temos uma distribuição multivariada. Nessa situação multivariada, as cartas mais utilizadas são a  $\mathcal{T}^2$  de Hotelling para a matriz de médias amostrais e  $\mathcal{W}$  para a matriz de covariância, [Johnson e Wichern \(2007\)](#). Neste caso, não basta observarmos cada variável separadamente e verificar se a mesma se encontra dentro dos limites especificados, mas observar as variáveis simultaneamente, dado que a quantidade de variáveis age sob a probabilidade de erro tipo I ( $\alpha$ ).

O procedimento para a construção de uma carta segue o seguinte raciocínio. Dada a distribuição normal univariada de média  $\mu$  e variância  $\sigma^2$

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \quad -\infty < x < \infty, \quad (2.1)$$

o expoente, conhecido como distância padronizada, expressa a distância da média em unidades de desvio padrão, isto é, define as caudas da distribuição. Esta distância pode ser definida como

$$(x - \mu)(\sigma^2)^{-1}(x - \mu). \quad (2.2)$$

No caso multivariado,  $x$  é substituído pelo vetor das variáveis  $\mathbf{x}$ ,  $\mu$  pelo vetor das médias  $\boldsymbol{\mu}$  e  $\sigma^2$  pela matriz de covariância  $\Sigma$ . Assim, a distância padronizada dada em (2.2) pode ser reescrita como

$$(\mathbf{x} - \boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu}). \quad (2.3)$$

A carta  $\mathcal{T}^2$  de Hotelling é a carta de controle mais usada para monitorar o vetor média do processo. Seja o vetor média amostral  $\bar{\mathbf{x}}$ , e a matriz de covariância amostral  $\mathbf{S}$ , estimadores não viesados da média e da covariância, e dada uma amostra de tamanho  $n$ , a estatística  $\mathcal{T}^2$  é a seguinte:

$$(\mathbf{x} - \bar{\mathbf{x}})'\mathbf{S}^{-1}(\mathbf{x} - \bar{\mathbf{x}}) \sim \frac{K(n+1)(n-1)}{n(n-K)} F_{K, n-K}, \quad (2.4)$$

onde

$$\bar{\mathbf{x}} = \frac{1}{n} \sum_{i=1}^K \mathbf{x}_i \quad (2.5)$$

e

$$\mathbf{S} = \frac{1}{n-1} \sum_{i=1}^K (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})'. \quad (2.6)$$

O procedimento para encontrar uma distribuição referência da carta  $\mathcal{T}^2$  consiste em usar um conjunto de dados sob controle (fase 1) para obter os limites de controle. Na fase 2, para uma amostra de tamanho um, a estatística (2.4) é calculada. Escores acima do percentil  $\alpha$  para essa nova observação indicam que a média  $\boldsymbol{\mu}$  mudou.

No caso desta pesquisa, na fase 1, a média e a matriz de covariância são estimadas de amostras de referência do processo sob controle sem nenhum problema na estrutura de covariância causados por eventos atípicos.

## 2.2 Carta da Variância Generalizada

A estrutura de covariância do processo é monitorada pela estatística Variância Generalizada ( $\mathcal{W}$ ). A covariância do processo resumida pela matriz quadrada  $\Sigma$ , na qual os elementos da diagonal principal são as variâncias das variáveis do processo, e o restante são as covariâncias. A estatística utilizada na carta de controle para monitoramento do processo é

$$\mathcal{W} = -Kn + Kn \ln(n) - n \ln\left(\frac{\det(\mathbf{A})}{\det(\mathbf{S})}\right) + \text{tr}(\mathbf{S}^{-1}\mathbf{A}) \sim \chi_{K(K+1)/2}^2 \quad (2.7)$$

onde  $\mathcal{W}$  tem distribuição  $\chi^2$  e  $\mathbf{S}$  vem das  $n$  amostras de referência. Aqui,  $\mathbf{A} = (m-1)\mathbf{S}_i$ , onde  $\mathbf{S}_i$  é a matriz de covariância do novo grupo de observações de tamanho  $m$ . Para o novo grupo de  $m$  amostras, o escore  $\mathcal{W}$  é obtido por 2.7 e os escores acima do percentil  $\alpha$  implicam que o novo grupo de observações foi obtido em um processo com  $\boldsymbol{\Sigma} \neq \boldsymbol{\Sigma}_0$ . Uma ampla discussão sobre a carta  $\mathcal{W}$  pode ser encontrada em Montgomery (2007) e Johnson e Wichern (2007).

## 2.3 Modelando as Bateladas por Séries Temporais

Os dados dos processos por bateladas podem ser vistos como uma matriz tri-dimensional ( $I \times K \times T$ ), onde  $I$  representa o número de bateladas,  $K$  o número de variáveis e  $T$  o número de instantes no tempo. Mais especificamente, para cada batelada, uma série temporal  $K$ -dimensional de tamanho  $T$  é gerada tendo dados de forte característica dinâmica.

Em muitas abordagens, o dado é rearranjado dando origem a uma matriz bidimensional e sendo analisado por técnicas multivariadas como Componentes Principais (PCA), Mínimos Quadrados Parciais, Máquinas de Vetores de Suporte, Redes Neurais, etc. Estas estruturas dimensionais têm duas formas clássicas. A abordagem de Nomikos e MacGregor (1995) usa a matriz de dados ( $I \times KT$ ), assim o comportamento dinâmico das variáveis do processo é priorizado (análise temporal), por exemplo, focando em correlações seriais. Wold et al. (1998) propôs uma abordagem alternativa, usando o arranjo ( $IT \times K$ ), no qual uma análise das variáveis é priorizada. Essa abordagem foca nas correlações cruzadas entre variáveis e permite o diagnóstico de distúrbios no processo. Camacho et al. (2009) apresenta uma ampla discussão sobre as vantagens e desvantagens das duas abordagens. Contudo, não há consenso sobre a melhor estratégia e há perdas e ganhos na escolha entre modelar correlação cruzada e correlação serial. A Figura 2.1 apresenta os dois esquemas de

redução de dimensionalidade. Na letra (a), temos a representação da análise temporal proposta por [Nomikos e MacGregor \(1995\)](#), enquanto que (b) representa a análise das variáveis de [Wold et al. \(1998\)](#).

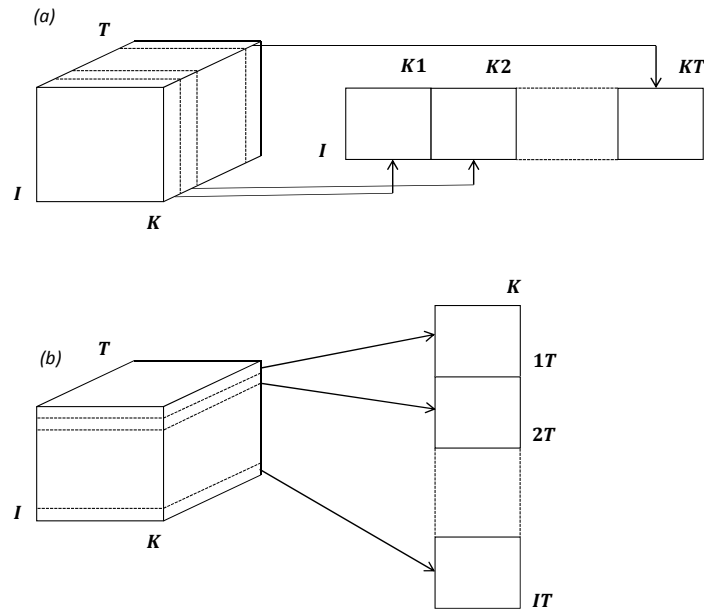


Figura 2.1: Representação da redução de dimensionalidade. Esquema tree-way para two-way.

Fonte: [Marcondes e Valk \(2019\)](#)

Alguns modelos de séries temporais têm sido usados para lidar com controle de processo estatístico multivariado, mais especificamente os Modelos Vetoriais Autor-regressivos (VAR). Todavia, muitas abordagens se restringem a monitorar processos contínuos, isto é, processos que tem uma estrutura intrinsecamente dupla ( $I \times K$ ). Logo, não havendo replicações das medidas em cada amostra, não há dimensão temporal. Essas abordagens usam o modelo VAR para eliminar a correlação entre as  $I$  amostras, o que é condição necessária para a aplicação das técnicas multivariadas clássicas. Basicamente, em um estágio preliminar, propõe-se um modelo VAR para estimar as correlação seriais. Posteriormente, as clássicas cartas de controle são usadas nos resíduos do modelo para monitorar o comportamento das futuras amostras.

[Choi et al. \(2008\)](#) propôs uma abordagem para lidar com o monitoramento de um processo por bateladas usando modelos VAR. Ele considera a estrutura de dados da batelada ( $I \times K \times T$ ) como uma  $I$  série temporal,  $K$ -dimensional de tamanho  $T$ . Resumindo, o autor sugere um conjunto de cartas baseadas na estatística de Hotelling para modelar os resíduos do modelo VAR ajustado nas  $I$  amostras históricas. Antes de ajustar o VAR, um método de redução de dimensionalidade PCA é aplicado para reduzir o número de variáveis  $K$ . Tendo como referência um processo



por bateladas, ele mostra que a abordagem baseada no controle dos resíduos supera as abordagens clássicas concorrentes, como [Nomikos e MacGregor \(1995\)](#) e [Wold et al. \(1998\)](#).

[Marcondes e Valk \(2019\)](#) criaram uma abordagem para lidar com os processos em bateladas usando modelos VAR com foco nos coeficientes do modelo ao invés dos resíduos e sem reduzir a dimensionalidade. As cartas de controle são baseadas nos coeficientes estimados do modelo VAR ajustado. Em resumo, são estimados os coeficientes VAR para cada amostra de batelada no tempo e guardadas as informações sobre as correlações cruzada e a serial padrão do processo sob controle. Após, são usadas as cartas de controle  $\mathcal{T}^2$  de Hotelling e  $\mathcal{W}$  da Variância Generalizada para monitorar as bateladas on-line e off-line. Em um processo de bateladas simulado, [Marcondes e Valk \(2019\)](#) demonstram que as cartas de controle baseadas nos coeficientes do VAR superam a abordagem com foco nos resíduos.

Dada esta metodologia, o objetivo deste trabalho é encontrar os tamanhos dos grupos que compõem a matriz acoplada de forma que tenhamos a maior sensibilidade na detecção de processos fora de controle. Aqui coloca-se a questão de termos os menores grupos possíveis, o que reduziria o tempo de processamento, tendo-se o cuidado para não perdermos as propriedades do modelo.

## 3 Metodologia

A metodologia utilizada neste trabalho passa pela modelagem do conjunto de dados por modelo VAR e análise de coeficientes advindos da dinâmica deste modelo para a construção de uma carta de controle para cada instante de tempo. A análise dos coeficientes tem resposta mais rápida ao descontrole do processo do que análise dos resíduos do VAR, conforme [Marcondes e Valk \(2019\)](#). Para que o processo possa ser analisado desde seu primeiro momento, faremos uso de acoplagem com dados do processo sob controle. Simularemos perturbações no processo, para verificar a velocidade de detecção. Neste momento, alteraremos o tamanho das partes que compõem a matriz acoplada.

### 3.1 Modelo VAR

O modelo VAR é amplamente usado na modelagem de séries temporais multivariadas. Proposto inicialmente por [Sims \(1980\)](#), é especialmente útil na previsão e conjuntamente descrevendo o comportamento dinâmico de uma série econômica financeira e permitindo a expressão da evolução de um conjunto de  $K$  variáveis como uma função linear dos seus valores anteriores. A forma básica do modelo VAR( $p$ ) é

$$\mathbf{z}_t = \boldsymbol{\Phi}_0 + \sum_{j=1}^p \boldsymbol{\Phi}_j \mathbf{z}_{t-j} + \mathbf{u}_t, \quad t \in \mathbb{Z} \quad (3.1)$$

onde  $\boldsymbol{\Phi}_0$  é um vetor de  $K \times 1$  constantes,  $\boldsymbol{\Phi}_j$  para  $j = 1, \dots, p$  é uma matriz  $K \times K$  de coeficientes do modelo e  $\mathbf{u}_t$  é o vetor de  $K \times 1$  erros satisfazendo  $\mathbb{E}[\mathbf{u}_t] = \mathbf{0}$ ,  $\text{Var}[\mathbf{u}_t] = \boldsymbol{\Gamma}$  e  $\text{Cov}[\mathbf{u}_t, \mathbf{u}_{t+h}] = \mathbf{0}$  para todo  $h = 1, 2, 3, \dots$ . Seja  $z = (z_1, \dots, z_T)$  uma série temporal de dimensão  $K \times T$ . Denotemos por

$$\boldsymbol{\Phi} = (\boldsymbol{\Phi}_0, \boldsymbol{\Phi}_1, \dots, \boldsymbol{\Phi}_p) \quad (3.2)$$

uma matriz de coeficientes  $K \times (Kp + 1)$ . Seja  $\mathbf{w}_t = (1, \mathbf{z}_t, \dots, \mathbf{z}_{t-p+1})'$  um vetor de comprimento  $Kp + 1$  e definamos  $\mathbf{W} = (\mathbf{w}_0, \dots, \mathbf{w}_{T-1})$  uma matriz de dimensões  $(Kp + 1) \times T$ . Por fim, definamos a matriz  $\mathbf{U} = (\mathbf{u}_1, \dots, \mathbf{u}_T)$  de dimensão  $K \times T$ . Com isto, podemos reescrever a equação [3.1](#), utilizando a notação vetorial como

$$\mathbf{v} = (\mathbf{W}' \otimes \mathbf{I}_K) \boldsymbol{\phi} + \mathbf{u} \quad (3.3)$$

onde  $\boldsymbol{\phi} = \text{Vec}(\boldsymbol{\Phi})$  e  $\mathbf{u} = \text{Vec}(\mathbf{U})$ ,  $\otimes$  é o produto de Kronecker e  $\text{Vec}$  é o operador que transforma a matriz em vetor.

Os estimadores de Mínimos Quadrados para o modelo VAR tem sua demonstração em Lütkepohl (2005), onde são detalhadas suas propriedades.

Por exemplo, se  $\mathbf{y}_t = [y_{1,t}, y_{2,t}]'$  é um VAR(1) bivariado, então  $\mathbf{y}_t$  pode ser escrito como

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} \phi_{0;1} \\ \phi_{0;2} \end{bmatrix} + \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} u_{1,t} \\ u_{2,t} \end{bmatrix}, \quad (3.4)$$

onde

$$\Sigma = \mathbb{E}(\mathbf{u}_t \mathbf{u}_t') = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}.$$

A expressão (3.4) pode ser escrita individualmente como

$$\begin{aligned} y_{1,t} &= \phi_{0;1} + \phi_{11}y_{1,t-1} + \phi_{12}y_{2,t-1} + u_{1,t}; \\ y_{2,t} &= \phi_{0;2} + \phi_{21}y_{1,t-1} + \phi_{22}y_{2,t-1} + u_{2,t}. \end{aligned} \quad (3.5)$$

Perceba que a expressão acima deixa claro que as variáveis se influenciam mutuamente, pois temos valores de  $y_2$  na equação de  $y_1$  e vice-versa.

## 3.2 Cartas de Controle Baseadas no Monitoramento da Dinâmica do Modelo VAR

Considere um conjunto de dados históricos contendo as realizações de  $I$  bateladas de produtos dentro das especificações. O dado representa  $I$  bateladas bem-sucedidas, cada uma contendo a trajetória no tempo de  $K$  variáveis do processo, medidas em instantes de tempo iguais  $t$ . Logo, para cada batelada temos uma matriz ( $K \times T$ ). Assumindo que essas  $I$  amostras venham de processo sob controle, isto é, essas amostras representam a trajetória de  $K$  variáveis em um processo normal como apresentado na Figura 1.2. Estatisticamente, isso significa que o comportamento destas amostras é relacionado à dinâmica do mesmo processo estocástico. Logo, é razoável assumirmos que esta dinâmica é bem caracterizada por um modelo VAR(p). O objetivo é identificar se um processo está mudando a dinâmica por algum motivo. Um exemplo de processo que muda o comportamento dinâmico a partir de um certo período de tempo é apresentado na Figura 3.1

Para cada batelada histórica, modelamos um VAR(p) e armazenamos o vetor das estimativas dos parâmetros para cada batelada. No passo seguinte, calculamos a estimativa da média e da variância dos parâmetros ao combinar as estimativas individuais. Essa estimativa contém informação relevante sobre a estrutura de correlação, incluindo correlação cruzada e serial. Como  $\bar{\phi}$  tem distribuição assintoticamente normal, propomos cartas de controle baseadas nas estatísticas  $\mathcal{T}^2$  de Hotelling e na Variância Generalizada  $\mathcal{W}$ . Na fase II, a nova batelada é monitorada, sendo que valores acima do percentil  $\alpha$  implicam que os valores são distintos dos valores esperados em um processo sob controle.

Considere novamente a matriz contendo  $I$  amostras  $K$ -dimensionais séries temporais de tamanho  $T$ , obtida de processo por bateladas sob controle. Para propormos uma abordagem de controle on-line, como no estágio preliminar, precisamos

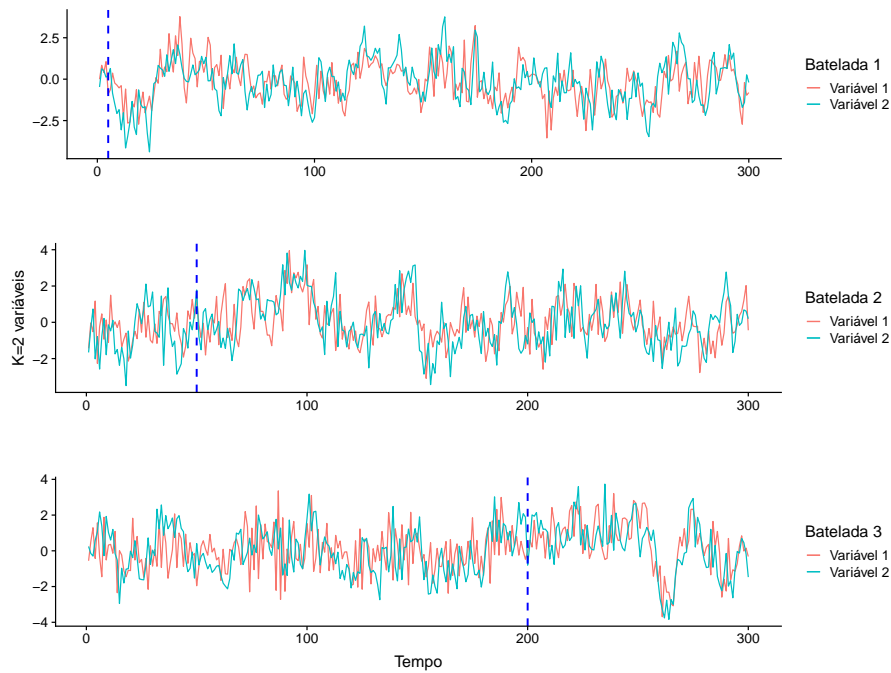


Figura 3.1: Processos fora de controle. Os tempos em que o processo começa a sair de controle são, respectivamente, 5, 50 e 200.

Fonte: Autor

construir uma distribuição de referência on-line para o processo sob controle utilizando as bateladas históricas. Num primeiro momento, aleatoriamente separamos as amostras em dois grupos de mesmo tamanho, o grupo aleatório e o grupo ajustado.

Para cada elemento do grupo ajustado, em cada instante do tempo  $t$ , definimos a matriz acoplada onde  $L_r$  e  $L_f$  são o número máximo de observações dos vetores aleatórios e ajustados. Em cada instante, uma nova série temporal é escolhida (de acordo com distribuição uniforme discreta) para que parte dela componha a matriz acoplada.

A matriz acoplada é composta de  $L_r$  sucessivas observações de um elemento do grupo aleatório e  $L_f$  sucessivas observações do elemento do grupo ajustado. Logo, para cada  $t$ , temos tantas matrizes acopladas quantos os elementos ajustados na distribuição de referência online.

No próximo passo, para cada matriz acoplada, obtemos as estimativas de Mínimos Quadrados para os coeficientes de um VAR(p). Logo, para um instante de tempo fixo, temos a metade do número de bateladas de matrizes contendo informações das correlações serial e cruzada em períodos 1 a  $p$  entre variáveis de um processo sob controle. A matriz acoplada é uma série temporal não-estacionária e as propriedades dos estimadores de Mínimos Quadrados são as mesmas do modelo VAR(p) clássico. Essa abordagem objetiva remover a correlação entre os  $\phi$ 's estimados para um tempo  $t$  fixo e reduzir a correlação serial entre  $\phi$ 's para cada batelada  $i$ . Ela permite monitorar a nova batelada desde o primeiro momento.

O procedimento consiste na aplicação da abordagem de controle proposta para cada instante no tempo. Na fase I, temos a metade do total de bateladas na distribuição de referência. Para cada instante no tempo, estimamos a média e a variância dos parâmetros. Na fase II, a nova batelada em curso no instante de tempo  $t$  é

monitorada, substituindo  $T$  por  $L_c$ . Então, pode-se monitorar a correlação cruzada e serial por meio dos escores  $\mathcal{T}^2$  e  $\mathcal{W}$ . Um fluxograma representando a metodologia online e off-line é apresentado na figura 3.2.

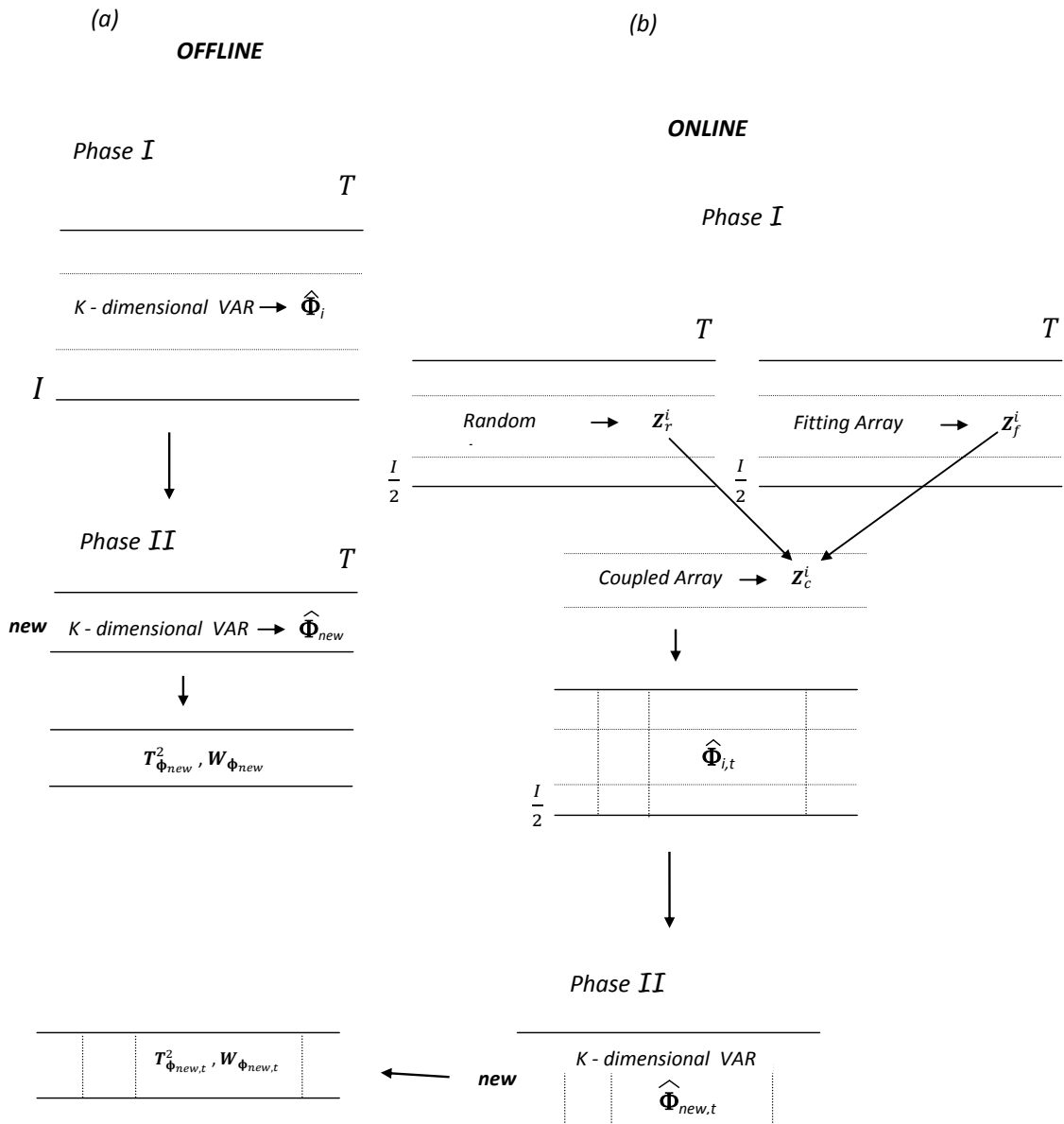


Figura 3.2: Fluxograma da metodologia online e off-line  
Fonte: Marcondes e Valk (2019)

Resumindo:

- (i) Utilizamos uma abordagem baseada em modelos VAR no contexto de processos por batelada. Dado processo gera, para cada batelada, séries temporais  $K$ -dimensionais de tamanho  $T$ . Pela metodologia baseada no VAR, podemos

modelar e monitorar as correlações cruzadas e seriais simultaneamente.

- (ii) Monitoramos os coeficientes do modelo VAR. Os resíduos captam mudanças na dinâmica das variáveis, incluindo as correlações cruzada e serial. Porém, monitorando os parâmetros estimados pelo VAR, podemos construir uma abordagem de controle com maior sensibilidade e que consiga detectar mudanças na dinâmica das variáveis mais rapidamente. Uma das contribuições mais significativas da metodologia proposta em [Marcondes e Valk \(2019\)](#), segundo os próprios autores, é permitir o monitoramento do processo em andamento desde o primeiro instante.
- (iii) Por fim, dividimos os dados nos grupos aleatório e ajustado na fase I e verificamos como o tamanho de cada porção impacta na performance. [Marcondes e Valk \(2019\)](#) estão conscientes de que a abordagem é sensível à escolha dos valores de  $L_c$ ,  $L_f$  e  $L_r$  e ponderam que se  $L_r$  for grande a performance da carta de controle baixa, pois a maior parte da informação vem do processo sob controle. Por outro lado,  $L_f$  grande aumenta a correlação serial entre os parâmetros, decaindo a performance das cartas de controle. Para um  $L_c$  grande, temos muita informação de processos sob controle, diminuindo a performance e aumentando o tempo computacional. Alternativamente,  $L_c$  pequeno causa estimação pobre dos coeficientes VAR, restringindo o uso da teoria assintótica e, conseqüentemente, causando instabilidade nos limites das cartas de controle. Adicionalmente, a porção aleatória na matriz acoplada garante a independência entre bateladas para cada instante de tempo  $t$  fixado, o que é um pressuposto para as estatísticas  $\mathcal{T}^2$  e  $\mathcal{W}$ .

## 4 Simulação

Considerando um processo por bateladas com 2 variáveis no qual o comportamento da dinâmica segue um modelo VAR(1), com  $p = 1$  e  $K = 2$ , conforme 3.4, porém com o vetor de constantes  $\phi_0 = \mathbf{0}$ . Na fase I, assumimos que, sob condições normais de operação, as bateladas são geradas por um processo com coeficientes dados pela matriz  $\Phi$

$$\Phi = \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix} = \begin{bmatrix} -0.3 & 0.4 \\ 0.4 & 0.5 \end{bmatrix}. \quad (4.1)$$

A matriz  $\Phi$  contém a informação da correlação serial defasada no tempo 1 na diagonal principal e da correlação cruzada na antidiagonal defasada no tempo 1.

Adotamos o índice *Average Run Length* (*ARL*), mais precisamente uma variação dele, que será descrito posteriormente. O *ARL* é usado para avaliar a performance das cartas de controle. Esse índice é definido como  $ARL = 1/r$ , onde  $r$  é a proporção de bateladas além dos limites de controle. Consideramos  $ARL_0$  como o número médio de bateladas até o alarme falso (para  $\alpha = 0.01$ , é esperado um  $ARL_0 = 100$ ), isto é, pontos fora dos limites de controle em um processo sem perturbações (processo sob controle).

Seja o processo sob controle, um processo VAR(1) com os parâmetros dados em (4.1). Na fase II, consideramos o cenário para perturbações com  $\phi_{11}$  tomando valores no conjunto  $\{-0.4, -0.3, -0.2, 0, 0.2, 0.4\}$ . Para cada cenário na fase I, geramos 2000 bateladas de  $T = 300$  instantes no tempo, que são igualmente divididas entre conjunto aleatório e ajustado. Na fase II, geramos 100 bateladas com 100 replicações de cada perturbação. Para os processos fora de controle, a perturbação foi feita em três instantes de tempo  $t^*$  diferentes, tomando os valores de  $\{5, 50, 200\}$ . Para a matriz acoplada, configuramos diferentes valores para  $L_c$ ,  $L_r$  e  $L_f$  que é o objetivo do nosso estudo. Para o tamanho da matriz acoplada  $L_c$  foram escolhidos os valores de  $\{15, 20, 25, 30, 40, 60, 70\}$ . Como o  $L_c = L_f + L_r$ , escolhemos também diferentes valores de  $L_f$ . Nesse capítulo reportamos a Tabela 4.1, por exemplo, onde consideramos poucas observações para a estimação de  $\Phi$ . Escolhemos  $L_c = 15$ , com  $L_f = 6$ , o que representa um valor bem baixo para usarmos teoria assintótica e propriedades de estimadores. No outro extremo, na Tabela 4.9, por exemplo, usamos  $L_c = 70$  com  $L_f = 28$ . Neste caso, temos um número razoável de observações para a questão da estimação da matriz  $\Phi$  do modelo VAR. No caso do  $L_c = 30$ , reportado nas Tabelas 4.4 e 4.6, testamos dois valores de  $L_f$ ,  $\{18, 12\}$ . Nas Tabelas 4.5 e 4.7, reportamos os resultados para um  $L_c$  de 40 com valores de  $L_f$  de  $\{24, 16\}$ .

Usamos o índice *Average Run Time Length* (*ARTL*) para avaliar a performance das cartas. O *ARTL* grava o número de instantes de tempo, a partir de  $t^*$ , até

que a carta sinalize a condição fora de controle, dado o número máximo  $(T - t^*)$  de instantes no tempo disponíveis. A probabilidade de alarme falso foi configurada para  $\alpha = 0.01$ .

A abordagem de referência é baseada no trabalho de [Choi et al. \(2008\)](#). Os resíduos para modelo VAR ajustado foram organizados em uma matriz  $I(p-1) \times K$ . No método [Choi et al. \(2008\)](#), não podemos aplicar totalmente nossa abordagem, pois as estimativas da Covariância Residual em cada instante do tempo não estão disponíveis. Assim, não podemos usar a carta baseada em  $\mathcal{W}$  para monitorar o processo em andamento. Por esta razão, o estudo comparativo é feito usando a carta  $\mathcal{T}^2$ .

Para cada uma das 1000 bateladas do grupo de ajuste, ajustamos 300 modelos VAR(1). Isto é, temos 300 matrizes de coeficientes estimadas em cada batelada. Esse número de  $T = 300$  foi fixo nesse estudo. Além disso, todas as simulações foram feitas considerando uma única matriz de coeficientes  $\Phi$ . A perturbação foi somente analisada no coeficiente  $\phi_{11}$ , ou seja, somente avaliamos o efeito de uma perturbação na correlação serial.

## 4.1 Resultados

Neste capítulo, apresentamos os resultados do nosso estudo de simulação. As Tabelas 4.1 a 4.9 resumem os resultados das abordagens de [Marcondes e Valk \(2019\)](#) e [Choi et al. \(2008\)](#) em cada cenário. Reportamos a média e o desvio padrão dos valores de  $ARTL$  para cada perturbação. Na linha cinza, destacamos os valores de  $ARL_0$ .

Para todos os casos considerados, observamos que os valores de  $ARL_0$  estão acima, mas relativamente próximos, do valor nominal de 100, dado que as cartas foram construídas com a probabilidade de alarme falso  $\alpha = 0.01$ . Nesse ponto, podemos ressaltar que os alarmes falsos estão controlados, dentro dos limites aceitáveis em todos os casos considerados.

Note que o valor de  $ARTL$ , para cada caso  $t^*$ , é associado a um espaço de probabilidade distinto, dado que o tempo remanescente  $(T - t^*)$  não tem o mesmo tamanho. Logo, não podemos comparar o  $ARTL$  para  $t^*$  diferentes.



Tabela 4.1: Resultados do  $ARTL$  e  $ARL_0$  para  $L_c = 15$  e  $L_f=6$ 

t*	Perturbação	$\mathcal{T}_\phi^2$		$\mathcal{W}_\phi$		$\mathcal{T}_u^2$	
		Média	Desvio	Média	Desvio	Média	Desvio
5	0.4	0.123	0.011	0.197	0.024	0.088	0.01
5	0.2	0.195	0.019	0.25	0.024	0.177	0.017
5	0	0.242	0.024	0.28	0.026	0.269	0.025
5	-0.2	0.277	0.026	0.303	0.028	0.318	0.028
5	-0.3	114.096	7.818	132.245	9.315	121.689	9.237
5	-0.4	0.283	0.025	0.294	0.027	0.321	0.027
50	0.4	0.147	0.016	0.227	0.025	0.107	0.011
50	0.2	0.217	0.023	0.273	0.027	0.207	0.021
50	0	0.267	0.025	0.305	0.026	0.29	0.027
50	-0.2	0.295	0.025	0.319	0.028	0.34	0.028
50	-0.3	111.77	7.234	131.083	8.219	121.201	7.456
50	-0.4	0.303	0.029	0.316	0.025	0.34	0.026
200	0.4	0.289	0.03	0.334	0.042	0.241	0.022
200	0.2	0.335	0.034	0.332	0.045	0.318	0.035
200	0	0.333	0.045	0.336	0.042	0.33	0.045
200	-0.2	0.344	0.049	0.331	0.048	0.316	0.053
200	-0.3	113.261	7.553	132.755	8.997	120.748	8.613
200	-0.4	0.338	0.048	0.329	0.057	0.315	0.057

Fonte: Autor.

Tabela 4.2: Resultados do  $ARTL$  e  $ARL_0$  para  $L_c = 20$  e  $L_f=8$ .

$t^*$	Perturbação	$\mathcal{T}_\phi^2$		$\mathcal{W}_\phi$		$\mathcal{T}_u^2$	
		Média	Desvio	Média	Desvio	Média	Desvio
5	0.4	0.096	0.01	0.181	0.022	0.091	0.01
5	0.2	0.159	0.018	0.247	0.026	0.181	0.018
5	0	0.226	0.023	0.28	0.03	0.269	0.023
5	-0.2	0.274	0.025	0.302	0.024	0.323	0.03
5	-0.3	114.086	8.292	133.606	8.089	120.617	9.73
5	-0.4	0.288	0.025	0.302	0.026	0.323	0.026
50	0.4	0.113	0.011	0.212	0.023	0.105	0.011
50	0.2	0.189	0.018	0.278	0.026	0.206	0.019
50	0	0.254	0.023	0.307	0.025	0.297	0.025
50	-0.2	0.295	0.024	0.319	0.026	0.338	0.027
50	-0.3	114.684	8.077	131.05	9.277	121.516	8.208
50	-0.4	0.308	0.028	0.326	0.028	0.339	0.03
200	0.4	0.261	0.025	0.333	0.035	0.244	0.022
200	0.2	0.326	0.031	0.344	0.037	0.32	0.039
200	0	0.348	0.039	0.334	0.05	0.317	0.048
200	-0.2	0.332	0.043	0.325	0.052	0.314	0.052
200	-0.3	114.499	7.095	131.783	8.735	121.019	8.898
200	-0.4	0.322	0.052	0.328	0.051	0.313	0.056

Fonte: Autor.

Tabela 4.3: Resultados do  $ARTL$  e  $ARL_0$  para  $L_c = 25$  e  $L_f=10$  .

t*	Perturbação	$\mathcal{T}_\phi^2$		$\mathcal{W}_\phi$		$\mathcal{T}_u^2$	
		Média	Desvio	Média	Desvio	Média	Desvio
5	0.4	0.08	0.007	0.177	0.017	0.09	0.009
5	0.2	0.142	0.016	0.248	0.025	0.177	0.017
5	0	0.21	0.021	0.28	0.03	0.272	0.027
5	-0.2	0.273	0.027	0.301	0.026	0.324	0.026
5	-0.3	114.06	7.722	132.893	7.921	120.897	9.388
5	-0.4	0.291	0.026	0.31	0.029	0.324	0.029
50	0.4	0.095	0.009	0.201	0.021	0.108	0.011
50	0.2	0.17	0.018	0.275	0.024	0.209	0.02
50	0	0.246	0.023	0.303	0.028	0.291	0.027
50	-0.2	0.292	0.026	0.323	0.028	0.344	0.029
50	-0.3	114.905	8.348	132.182	8.191	121.76	7.511
50	-0.4	0.314	0.025	0.321	0.026	0.336	0.027
200	0.4	0.229	0.02	0.331	0.032	0.237	0.026
200	0.2	0.31	0.029	0.337	0.046	0.32	0.037
200	0	0.334	0.04	0.33	0.05	0.329	0.057
200	-0.2	0.334	0.043	0.317	0.058	0.313	0.051
200	-0.3	112.558	8.48	131.518	7.945	121.098	9.329
200	-0.4	0.33	0.052	0.323	0.051	0.308	0.053

Fonte: Autor.

Tabela 4.4: Resultados do  $ARTL$  e  $ARL_0$  para  $L_c = 30$  e  $L_f=18$ .

t*	Perturbação	$\mathcal{T}_\phi^2$		$\mathcal{W}_\phi$		$\mathcal{T}_u^2$	
		Média	Desvio	Média	Desvio	Média	Desvio
5	0.4	0.063	0.004	0.167	0.014	0.09	0.01
5	0.2	0.108	0.009	0.26	0.026	0.175	0.018
5	0	0.192	0.019	0.315	0.027	0.274	0.026
5	-0.2	0.302	0.028	0.348	0.029	0.323	0.025
5	-0.3	120.849	9.15	131.301	9.133	125.957	9.978
5	-0.4	0.341	0.03	0.363	0.031	0.324	0.026
50	0.4	0.075	0.005	0.196	0.019	0.104	0.011
50	0.2	0.13	0.011	0.287	0.026	0.206	0.019
50	0	0.225	0.019	0.334	0.03	0.292	0.024
50	-0.2	0.319	0.027	0.368	0.03	0.338	0.029
50	-0.3	119.984	8.38	131.679	7.758	129.207	10.576
50	-0.4	0.35	0.029	0.376	0.031	0.338	0.027
200	0.4	0.187	0.012	0.322	0.038	0.236	0.024
200	0.2	0.284	0.022	0.316	0.049	0.323	0.036
200	0	0.334	0.04	0.274	0.058	0.321	0.058
200	-0.2	0.301	0.055	0.255	0.069	0.309	0.049
200	-0.3	120.215	9.762	131.567	7.735	127.091	9.792
200	-0.4	0.262	0.066	0.246	0.066	0.314	0.053

Fonte: Autor.

Tabela 4.5: Resultados do  $ARTL$  e  $ARL_0$  para  $L_c = 40$  e  $L_f=24$  .

t*	Perturbação	$\overline{\mathcal{T}}_\phi^2$		$\mathcal{W}_\phi$		$\overline{\mathcal{T}}_u^2$	
		Média	Desvio	Média	Desvio	Média	Desvio
5	0.4	0.061	0.004	0.162	0.016	0.091	0.01
5	0.2	0.097	0.007	0.26	0.027	0.177	0.016
5	0	0.181	0.017	0.323	0.033	0.267	0.026
5	-0.2	0.307	0.027	0.374	0.031	0.323	0.025
5	-0.3	123.316	8.822	130.85	7.916	129.044	10.167
5	-0.4	0.356	0.034	0.375	0.034	0.32	0.027
50	0.4	0.074	0.004	0.197	0.019	0.106	0.01
50	0.2	0.119	0.01	0.288	0.03	0.204	0.019
50	0	0.216	0.02	0.353	0.028	0.296	0.025
50	-0.2	0.338	0.024	0.386	0.032	0.342	0.026
50	-0.3	122.738	10.074	132.89	9.167	129.497	9.666
50	-0.4	0.371	0.031	0.396	0.032	0.341	0.027
200	0.4	0.188	0.011	0.326	0.036	0.237	0.025
200	0.2	0.284	0.023	0.307	0.055	0.317	0.037
200	0	0.337	0.04	0.251	0.068	0.328	0.044
200	-0.2	0.263	0.056	0.201	0.079	0.322	0.055
200	-0.3	121.418	9.336	130.74	8.094	129.699	10.997
200	-0.4	0.222	0.067	0.204	0.074	0.301	0.06

Fonte: Autor.

Tabela 4.6: Resultados do  $ARTL$  e  $ARL_0$  para  $L_c = 30$  e  $L_f=12$ 

t*	Perturbação	$\mathcal{T}_\phi^2$		$\mathcal{W}_\phi$		$\mathcal{T}_u^2$	
		Média	Desvio	Média	Desvio	Média	Desvio
5	0.4	0.074	0.007	0.17	0.018	0.091	0.009
5	0.2	0.129	0.016	0.241	0.026	0.179	0.016
5	0	0.203	0.019	0.279	0.026	0.268	0.022
5	-0.2	0.269	0.026	0.305	0.025	0.323	0.025
5	-0.3	116.054	7.587	133.096	8.781	121.632	8.416
5	-0.4	0.296	0.027	0.308	0.027	0.32	0.023
50	0.4	0.086	0.008	0.192	0.018	0.107	0.011
50	0.2	0.153	0.017	0.271	0.026	0.207	0.02
50	0	0.234	0.024	0.305	0.026	0.297	0.026
50	-0.2	0.296	0.026	0.325	0.026	0.342	0.027
50	-0.3	115.03	8.677	131.586	9.032	121.41	8.293
50	-0.4	0.312	0.028	0.324	0.027	0.342	0.024
200	0.4	0.213	0.018	0.327	0.033	0.236	0.021
200	0.2	0.306	0.027	0.332	0.046	0.317	0.036
200	0	0.337	0.039	0.323	0.049	0.329	0.046
200	-0.2	0.322	0.049	0.322	0.048	0.308	0.053
200	-0.3	112.575	8.59	132.877	9.593	120.474	8.51
200	-0.4	0.324	0.048	0.308	0.057	0.316	0.062

Fonte: Autor.

Tabela 4.7: Resultados do  $ARTL$  e  $ARL_0$  para  $L_c = 40$  e  $L_f=16$ 

t*	Perturbação	$\mathcal{T}_\phi^2$		$\mathcal{W}_\phi$		$\mathcal{T}_u^2$	
		Média	Desvio	Média	Desvio	Média	Desvio
5	0.4	0.066	0.006	0.162	0.017	0.091	0.009
5	0.2	0.113	0.011	0.242	0.026	0.177	0.017
5	0	0.191	0.02	0.284	0.026	0.267	0.026
5	-0.2	0.271	0.025	0.311	0.028	0.325	0.027
5	-0.3	116.403	9.084	131.768	8.744	123.011	7.766
5	-0.4	0.306	0.029	0.317	0.027	0.318	0.028
50	0.4	0.078	0.006	0.191	0.017	0.104	0.011
50	0.2	0.132	0.015	0.271	0.025	0.206	0.02
50	0	0.22	0.023	0.311	0.027	0.294	0.024
50	-0.2	0.291	0.028	0.335	0.028	0.34	0.028
50	-0.3	116.424	7.761	130.851	8.747	121.51	9.012
50	-0.4	0.326	0.027	0.339	0.027	0.343	0.026
200	0.4	0.191	0.015	0.331	0.035	0.236	0.024
200	0.2	0.294	0.02	0.331	0.038	0.32	0.032
200	0	0.335	0.039	0.318	0.054	0.33	0.054
200	-0.2	0.317	0.053	0.308	0.057	0.309	0.057
200	-0.3	116.892	7.587	132.264	8.909	123.206	8.534
200	-0.4	0.308	0.06	0.297	0.057	0.312	0.056

Fonte: Autor.

Tabela 4.8: Resultados do  $ARTL$  e  $ARL_0$  para  $L_c = 60$  e  $L_f=24$ 

t*	Perturbação	$\mathcal{T}_\phi^2$		$\mathcal{W}_\phi$		$\mathcal{T}_u^2$	
		Média	Desvio	Média	Desvio	Média	Desvio
5	0.4	0.064	0.004	0.16	0.016	0.092	0.009
5	0.2	0.099	0.008	0.242	0.023	0.175	0.016
5	0	0.174	0.018	0.296	0.032	0.27	0.025
5	-0.2	0.275	0.029	0.329	0.034	0.325	0.029
5	-0.3	117.907	7.42	131.751	7.861	123.907	10.502
5	-0.4	0.31	0.029	0.33	0.027	0.319	0.029
50	0.4	0.076	0.005	0.188	0.016	0.108	0.011
50	0.2	0.118	0.01	0.272	0.025	0.205	0.02
50	0	0.206	0.019	0.32	0.027	0.291	0.024
50	-0.2	0.302	0.024	0.351	0.031	0.341	0.028
50	-0.3	116.852	8.997	132.191	8.854	123.525	9.172
50	-0.4	0.334	0.026	0.354	0.029	0.34	0.027
200	0.4	0.19	0.013	0.332	0.037	0.233	0.024
200	0.2	0.281	0.021	0.333	0.043	0.321	0.03
200	0	0.337	0.035	0.307	0.057	0.33	0.05
200	-0.2	0.319	0.058	0.297	0.059	0.319	0.056
200	-0.3	117.295	8.555	131.085	8.061	123.521	9.035
200	-0.4	0.295	0.055	0.289	0.059	0.307	0.056

Fonte: Autor.



Tabela 4.9: Resultados do  $ARTL$  e  $ARL_0$  para  $L_c = 70$  e  $L_f=28$ 

$t^*$	Perturbação	$\mathcal{T}_\phi^2$		$\mathcal{W}_\phi$		$\mathcal{T}_u^2$	
		Média	Desvio	Média	Desvio	Média	Desvio
5	0.4	0.064	0.004	0.156	0.014	0.091	0.01
5	0.2	0.098	0.007	0.24	0.026	0.176	0.015
5	0	0.165	0.016	0.295	0.032	0.265	0.024
5	-0.2	0.277	0.029	0.33	0.03	0.322	0.027
5	-0.3	118.463	8.45	130.902	8.825	125.575	9.073
5	-0.4	0.316	0.028	0.341	0.032	0.317	0.028
50	0.4	0.077	0.004	0.188	0.018	0.107	0.01
50	0.2	0.117	0.008	0.271	0.024	0.204	0.019
50	0	0.197	0.019	0.324	0.027	0.294	0.024
50	-0.2	0.308	0.029	0.347	0.029	0.34	0.029
50	-0.3	118.75	8.493	131.487	8.318	126.07	9.165
50	-0.4	0.338	0.032	0.343	0.026	0.341	0.028
200	0.4	0.192	0.01	0.335	0.032	0.234	0.025
200	0.2	0.279	0.021	0.332	0.046	0.32	0.035
200	0	0.351	0.032	0.306	0.048	0.324	0.055
200	-0.2	0.313	0.054	0.276	0.062	0.308	0.055
200	-0.3	117.646	9.228	133.513	8.995	125.072	8.552
200	-0.4	0.29	0.054	0.284	0.063	0.311	0.056

Fonte: Autor.

No artigo de [Marcondes e Valk \(2019\)](#) o valor de referência é  $L_c = 50$ , com  $L_f = 20$ . Partindo dos resultados obtidos no artigo original, é natural começar a análise pelas tabelas com valores de  $L_c$  e  $L_f$  próximos aos padrões. Na Tabela 4.6, para a média, observamos que:

- (i) A detecção de descontrole é mais rápida na perturbação 0.4 e mais devagar em -0.4. Isso se explica por estes valores do coeficiente de perturbação estarem mais próximos de -0.3, que é a situação em que o processo está sob controle ou, sob  $H_0$ .
- (ii) Os resultados da análise dos coeficientes foram superiores à dos resíduos, sendo a exceção quando  $t^*$  foi 200. Neste caso, os  $ARTL$  dos resíduos foram menores na maioria dos casos (perturbações 0, 0.2, -0.2 e -0.4).
- (iii) Podemos observar uma diminuição no valor do  $ARL_0$ .
- (iv) No tempo  $t^* = 5$ , temos os maiores  $ARTL$  para as perturbações 0.4 (0.074), 0.2 (0.129), 0 (0.203) e os menores para -0.2 (0.269) e -0.4 (0.296).

Após a simulação reportada nas Tabelas 4.9, 4.8, 4.7 e 4.6, optamos por alterar os valores de  $L_f$  e  $L_r$  para verificar se a diferença nos valores de  $ARL_0$  e  $ARTL$  se deve a alguma dessas medidas. Notamos que a medida em que diminuimos o  $L_c$ , aumentamos consideravelmente o tempo computacional. Notamos que existe uma leve tendência em detectar mais rápido o descontrole para pequenas perturbações. No entanto, podemos notar que quando diminuimos o  $L_c$ , a metodologia de [Marcondes e Valk \(2019\)](#) perde em mais casos para a metodologia de [Choi et al. \(2008\)](#).

## 5 Conclusões

Neste trabalho, objetivamos encontrar tamanhos ótimos para as partes  $L_f$ ,  $L_r$  e  $L_c$  do método de acoplamento proposto por [Marcondes e Valk \(2019\)](#). Realizamos um estudo de simulação com parâmetros baseados naqueles propostos no artigo original. Fizemos pequenas variações de alguns parâmetros para avaliar o comportamento da metodologia baseada em acoplamento de séries temporais. Basicamente, exploramos somente o tamanho das matrizes acopladas e as quantidades de cada uma das partes que acoplamos.

Observamos que o aumento de  $L_f$  reduz os valores de  $ARTL$  para os valores mais distantes de  $H_0$ , ou seja para perturbações maiores. Porém, nota-se um aumento dos valores de  $ARL_0$ , mostrando-se mais conservador nesses casos. No entanto, ainda, nota-se um aumento nos valores de  $ARTL$  para valores próximos de  $H_0$   $\{0, -0.2$  e  $-0.4\}$ .

Como conclusão final, baseada nesse estudo, podemos dizer que se o objetivo for detectar com maior rapidez, devemos diminuir o  $L_c$ . Mas isso torna a metodologia nem sempre superior a metodologia de [Choi et al. \(2008\)](#). Se o objetivo for ser consistentemente melhor do que a abordagem de [Choi et al. \(2008\)](#), devemos aumentar o  $L_c$ . Para esses casos, observamos que o  $L_f$  se comporta de forma semelhante. Aumentando o  $L_f$ , melhora a metodologia comparativamente à de [Choi et al. \(2008\)](#).

## Referências Bibliográficas

- Camacho, J., Picó, J., e Ferrer, A. (2009). The best approaches in the on-line monitoring of batch processes based on pca: Does the modelling structure matter? *Analytica chimica acta*, 642(1-2):59–68.
- Choi, S. W., Morris, J., e Lee, I.-B. (2008). Dynamic model-based batch process monitoring. *Chemical Engineering Science*, 63(3):622–636.
- Johnson, R. A. e Wichern, D. W. (2007). Applied multivariate statistical analysis.
- Lütkepohl, H. (2005). *New introduction to multiple time series analysis*. Springer Science & Business Media.
- Marcondes, D. F. e Valk, M. (2019). Dynamic var model-based control charts for batch process monitoring. *Submitted do EJOR*.
- Montgomery, D. C. (2007). *Introduction to statistical quality control*. John Wiley & Sons.
- Nomikos, P. e MacGregor, J. F. (1995). Multivariate spc charts for monitoring batch processes. *Technometrics*, 37(1):41–59.
- Shewhart, W. (1931). Statistical method from an engineering viewpoint. *Journal of the American Statistical Association*, 26(175):262–269.
- Shewhart, W. A. (1925). The application of statistics as an aid in maintaining quality of a manufactured product. *Journal of the American Statistical Association*, 20(152):546–548.
- Sims, C. A. (1980). Macroeconomics and reality. *Econometrica: journal of the Econometric Society*, pages 1–48.
- Wold, S., Kettaneh, N., Fridén, H., e Holmberg, A. (1998). Modelling and diagnostics of batch processes and analogous kinetic experiments. *Chemometrics and intelligent laboratory systems*, 44(1-2):331–340.