



Universidade Federal do Rio Grande do Sul  
Instituto de Matemática e Estatística  
Programa de Pós-Graduação em Estatística

# **Controle Estatístico de Processos por Bateladas: Uma abordagem utilizando o modelo VAR Bayesiano**

Mariana Motta Dias da Silva

Porto Alegre, Dezembro de 2023.



### CIP - Catalogação na Publicação

Silva, Mariana Motta Dias da  
Controle Estatístico de Processos por Bateladas:  
Uma abordagem utilizando o modelo VAR Bayesiano /  
Mariana Motta Dias da Silva. -- 2023.  
42 f.  
Orientador: Danilo Marcondes Filho.

Coorientadora: Gabriela Bettella Cybis.

Dissertação (Mestrado) -- Universidade Federal do  
Rio Grande do Sul, Instituto de Matemática e  
Estatística, Programa de Pós-Graduação em Estatística,  
Porto Alegre, BR-RS, 2023.

1. Processo em Batelada. 2. Controle Estatístico de  
Processo. 3. Priori Minnesota. 4. Fator Bayes. I.  
Filho, Danilo Marcondes, orient. II. Cybis, Gabriela  
Bettella, coorient. III. Título.



Dissertação submetida por Mariana Motta Dias da Silva como requisito parcial para a obtenção do título de Mestre em Estatística pelo Programa de Pós-Graduação em Estatística da Universidade Federal do Rio Grande do Sul.

**Orientador(a):**

Prof. Dr. Danilo Marcondes Filho (PPGEst - UFRGS)

**Co-orientador(a):**

Profa. Dra. Gabriela Bettella Cybis (PPGEst - UFRGS)

**Comissão Examinadora:**

Profa. Dra. Silvana Schneider (PPGEst-UFRGS)

Profa. Dra. Márcia Barbian (PPGEst-UFRGS)

Prof. Dr. Ângelo Márcio Oliveira Sant'Anna (PEI-UFBA)

Data de Apresentação: 20 de Dezembro de 2023



*“Em meio à incerteza, os números falam,  
e na estatística se da voz  
a sabedoria dos dados”*





# AGRADECIMENTOS

*Primeiramente, agradeço de coração aos meus pais, Marcelo e Adriana, pelo apoio, confiança e amor ao longo de toda a minha jornada. Sem o suporte e incentivo de vocês, nada disso seria possível; vocês são minha base e inspiração.*

*À minha irmã, Giovana, por ser minha amiga, companheira incansável e incentivadora. Nos bons e maus momentos, sua amizade foi uma fonte constante de conforto e força. Cada palavra de encorajamento foi um impulso vital que me motivou a persistir nos momentos mais desafiadores. Ter você ao meu lado, sem dúvida, é um presente valioso na minha vida.*

*Ao meu namorado, Daniel, pelo amor, paciência e compreensão demonstrados ao longo desta jornada. Seu apoio foi crucial para manter meu equilíbrio e determinação durante os estudos. Obrigado por estar ao meu lado e por ser uma fonte constante de motivação.*

*Aos colegas de trabalho no Hospital Moinhos de Vento, que proporcionaram um ambiente propício para o meu crescimento, tanto profissional quanto pessoal. A troca de experiências, conhecimentos e vivências foi enriquecedora e contribuiu para o meu amadurecimento ao longo deste período.*

*Ao meu orientador, Danilo, pela orientação, paciência e ensinamentos ao longo deste processo. Seu suporte foi fundamental para o desenvolvimento desta dissertação.*

*Aos professores que participaram da minha formação até o momento, proporcionando aprendizado contínuo e inspiração para buscar a excelência acadêmica. Também aos colegas de mestrado, que sempre estiveram disponíveis para trocar ideias, tirar dúvidas e oferecer sugestões.*

*A todos que de alguma forma estiveram presentes em minha jornada acadêmica, meu sincero agradecimento. Cada contribuição foi valiosa e fez desta realização algo possível.*



## RESUMO

O Controle Estatístico de Processo voltado ao processo batelada é amplamente encontrado na literatura devido à estrutura de dados bastante peculiar, na qual temos diferentes fontes de variabilidade a serem consideradas. Um processo em batelada típico, gera para cada batelada, séries temporais representando medições sucessivas de variáveis do processo. Neste contexto, além da modelagem das correlações seriais e cruzadas, precisamos modelar a variabilidade entre bateladas (ou seja, entre séries temporais). Há ainda um número restrito de abordagens para monitoramento de bateladas com base na teoria de séries temporais, sendo a principal razão o fato de que a teoria de inferência para séries temporais não está direcionada para acomodar replicação de séries temporais, como acontece com dados em batelada. O trabalho recente de [Marcondes and Valk \(2020\)](#) apresenta uma abordagem baseada no Modelo Vetorial Autoregressivo (VAR), em que estimativas dos coeficientes do VAR para um grupo de bateladas de referência sob controle são utilizadas para construir cartas de controle com base nas estatísticas de Hotelling e da Variância Generalizada. Uma nova batelada sob investigação é comparada as de referência em termos de média e variância das estimativas dos coeficientes do VAR. Através de dados simulados, a abordagem proposta pareceu ser muito eficiente na detecção de perturbações; contudo, neste caso, é necessário um número mínimo razoável de amostras de referência. Esta dissertação apresenta uma abordagem alternativa ao trabalho de [Marcondes and Valk \(2020\)](#) que supera esta restrição de lidar com poucas bateladas de referência disponíveis. Com base na combinação do VAR Bayesiano com a *Priori* Minnesota e do Fator Bayes propomos uma regra de decisão para avaliar uma nova batelada através da comparação de verossimilhanças marginais. Realizamos um estudo de simulação para avaliar o desempenho da abordagem proposta em comparação com a abordagem de [Marcondes and Valk \(2020\)](#). Nossa regra de controle parece ser eficiente, mesmo com um número bastante restrito de amostras de referência disponíveis.



## ABSTRACT

The Statistical Process Control aimed at the batch process is widely found in the literature due to the very peculiar data structure in which we have different sources of variability to consider. A typical batch process generates for each batch sample time series representing successive measurements of many variables. Here, beyond the modeling of the serial and the cross-correlations, we need to model the variability among batch samples (i.e., among time series). There are a limited number of approaches for batch monitoring based on time series theory so far, and the main reason is that the time series inference theory is not grounded in dealing with time series replication, as we are in batch data. The recent work of [Marcondes and Valk \(2020\)](#) presents an approach based on the Vector Autoregressive (VAR) model. The VAR coefficient estimates for a group of reference in-control batches are jointly used to build the control charts based on the Hotelling and the Generalized Variance statistics. The new batch under monitoring is compared to the reference batches in terms of the mean and variance of the VAR coefficient estimates. Through the simulation, this approach proposal seemed to be very efficient in detecting disturbances; however, in this case, a reasonable number of reference batches is needed. The proposed work is an alternative approach to that of [Marcondes and Valk \(2020\)](#) that overcomes this constraint to deal with cases in which few reference batches are available. Based on the combination of the Bayesian VAR with Minnesota prior and the Factor Bayes we propose a decision rule to evaluate a new batch in comparison to the reference in-control batches by comparing marginal likelihoods. We conduct a simulated case study to evaluate the performance of the proposed approach compared to the benchmark approach of [Marcondes and Valk \(2020\)](#). Our control rule seems to be efficient, even with very few reference batches available.



---

# ÍNDICE

---

<b>1</b>	<b>Introdução</b>	<b>2</b>
<b>2</b>	<b>Objetivos, Contribuição e Estrutura da Dissertação</b>	<b>4</b>
2.1	Objetivo Geral . . . . .	4
2.2	Objetivos Específicos . . . . .	4
2.3	Contribuição . . . . .	4
2.4	Estrutura do trabalho . . . . .	5
<b>3</b>	<b>CEP para Processos em Bateladas</b>	<b>6</b>
<b>4</b>	<b>Metodologia</b>	<b>10</b>
4.1	Modelo Vetorial Autoregressivo (VAR) . . . . .	10
4.2	VAR Bayesiano (BVAR) . . . . .	11
4.2.1	Priori e Posteriori Minnesota . . . . .	12
4.3	Fator de Bayes . . . . .	14
4.4	Abordagem de monitoramento . . . . .	15
<b>5</b>	<b>Estudo de caso simulado</b>	<b>18</b>
5.1	Abordagem baseada no VAR . . . . .	18
5.2	Definindo os Cenários . . . . .	19
5.3	Apresentação dos resultados . . . . .	21
<b>6</b>	<b>Conclusões e trabalhos futuros</b>	<b>26</b>

---

# CAPÍTULO 1

## INTRODUÇÃO

---

O Controle Estatístico de Processos (CEP) engloba um conjunto de técnicas que objetivam a análise, monitoramento e melhoria da qualidade de processos produtivos [Montgomery \(2020\)](#). Dentre estas, as Cartas de Controle (CCs) desempenham papel importante na etapa de monitoramento de tais processos [Ramos \(2000\)](#).

Em 1924, Walter Andrew Shewhart ([Shewhart, 1931](#)) propôs e desenvolveu as CCs, com o objetivo de identificar discrepâncias em um processo, separando a variabilidade de uma variável em análise devido a causas comuns (variabilidade inerente ao processo) e eventualmente devido a causas específicas (variabilidade devido à algum evento atípico incidindo durante as medições).

As CCs são montadas a partir da distribuição amostral de uma estatística de interesse, sendo a *média*, a *mediana*, o *desvio padrão* e a *amplitude* as de uso mais frequente inicialmente [Montgomery \(2020\)](#). Considerando um histórico de amostras obtidas de um processo sujeito apenas a causas comuns de variação, os limites de controle são determinados utilizando a estatística de interesse e os quantis da distribuição amostral dessa estatística, determinados por uma probabilidade de detecção de alarme falso. A cada nova amostra comparada com esses limites, é feita uma avaliação se o processo está dentro ou fora de controle.

O foco desta dissertação são os processos que ocorrem por bateladas. Este tipo de processo traz uma estrutura de dados bastante peculiar. Durante uma batelada, são disponibilizadas sucessivas amostras que representam medições de um conjunto de variáveis de interesse, configurando assim séries temporais multivariadas. Bateladas históricas trazem então uma estrutura de dados tridimensional ( $I$  bateladas  $\times K$  variáveis  $\times T$  instantes de tempo) bastante desafiadora, visto que nestes casos é preciso considerar e tratar a variabilidade intra bateladas, composta pela correlação serial e a correlação cruzada das variáveis, além da variabilidade entre bateladas (entre séries temporais multivariadas).

Tais características chamam atenção de inúmeros pesquisadores e tornaram o CEP no contexto de processos em bateladas uma área de pesquisa bastante consolidada, o que pode ser evidenciado pela frequência de trabalhos disponíveis em bases de periódicos da área (citamos por exemplo a *web of science*). Abordagens de controle clássicas estão baseadas no trabalho precursor de [Nomikos and MacGregor \(1995b\)](#) e se concentram na modelagem de tal estrutura a partir da utilização de técnicas de redução de dados numa estrutura desdobrada em um arranjo bidimensional, sendo os arranjos  $I \times KT$  e  $IT \times K$  mais frequentes. Técnicas multivariadas tais como Componentes Principais e Mínimos Quadrados Parciais são aplicadas nas colunas das matrizes  $I \times KT$  e  $IT \times K$  e cartas



de controle são propostas para monitorar os escores. O arranjo na primeira matriz está focado em capturar a variação entre bateladas e captura indiretamente a dinâmica temporal dos dados, em que os efeitos da correlação serial e cruzada dos dados estão misturados. O segundo arranjo prioriza a correlação entre as variáveis no conjunto dos instantes de tempo e das bateladas, isto é, a dinâmica temporal acaba sendo deixada de lado [Camacho et al. \(2009\)](#).

Um número relativamente restrito de trabalhos na literatura propõe abordagens que consideram diretamente a natureza temporal dos dados, baseando-se assim na utilização de teoria de séries temporais. Uma razão importante se deve ao fato de que a teoria de séries temporais está no domínio longitudinal, não sendo desenvolvida para um cenário em que existem replicações de séries temporais. Dessa forma, a variabilidade entre séries temporais não está acomodada.

[Marcondes and Valk \(2020\)](#), em um trabalho recente propuseram um conjunto de cartas de controle para tais processos, através da utilização do Modelo Vetorial Autorregressivo (VAR) ([Lütkepohl, 2005](#)). Em linhas gerais, para cada batelada histórica, os coeficientes de um modelo VAR são estimados e armazenados em um vetor de médias e uma matriz de covariâncias. A série temporal da nova batelada sob análise é ajustada via VAR e suas estimativas são comparadas às estimativas das bateladas de referência, em termos de média e variância através das estatísticas de *Hotteling* e da *Variância Generalizada* modificadas. Esta abordagem se mostrou bastante eficiente para capturar a variabilidade intra e entre séries temporais, em um cenário com um número mínimo aceitável de bateladas de referência disponível para estabelecer os limites de controle das cartas.

Esta dissertação apresenta uma versão alternativa a proposição de [Marcondes and Valk \(2020\)](#), baseada em um modelo VAR Bayesiano, utilizando a distribuição *a priori* de Minnesota [Wind \(2015\)](#). Combinado a informação *a priori* com a verossimilhança dada pelo modelo VAR e a distribuição *a posteriori*, propomos uma carta de controle derivada de uma regra de decisão, baseada no Fator de Bayes. Em linhas gerais, uma nova batelada é comparada com as bateladas de referência em termos de verossimilhança marginal, isto é, quão provável esta batelada ter sido gerada do mesmo processo sob controle gerador das bateladas de referência. Um estudo de simulação é apresentado, onde a abordagem proposta apresenta resultados promissores em um cenário onde o número de bateladas de referência é bastante reduzido ( $< 30$ ) e portanto não viável a utilização da abordagem de [Marcondes and Valk \(2020\)](#). Adicionalmente, mesmo em cenários comparáveis a abordagem de [Marcondes and Valk \(2020\)](#), a proposição dessa dissertação apresenta resultados interessantes.

---

## CAPÍTULO 2

# OBJETIVOS, CONTRIBUIÇÃO E ESTRUTURA DA DISSERTAÇÃO

---

### 2.1 *Objetivo Geral*

Apresentar uma nova abordagem, alternativa à de [Marcondes and Valk \(2020\)](#), para monitoramento de processos em bateladas, onde existe um número restrito de amostras de referência, através da utilização combinada do modelo VAR Bayesiano com *a priori* Minnesota e o Fator de Bayes.

### 2.2 *Objetivos Específicos*

- Apresentar uma revisão de literatura sobre abordagens para monitoramento de processos em bateladas.
- Propor uma carta de controle com uma regra de decisão baseada no Fator de Bayes, a partir do VAR Bayesiano apresentado.
- Realizar estudo com dados simulados para avaliar o desempenho da abordagem proposta e comparar com a abordagem clássica de [Marcondes and Valk \(2020\)](#).

### 2.3 *Contribuição*

Desenvolvimento de uma abordagem de monitoramento de processos em bateladas utilizando a inferência bayesiana, através de uma carta de controle. A proposta combina o modelo VAR com distribuição *a priori* de Minnesota e o Fator de Bayes, para criação da estatística de avaliação da qualidade de novas bateladas. O foco de interesse aqui é em um cenário com um número restrito ( $< 30$ ) de amostras históricas de referência disponíveis, cenário esse não acomodado na abordagem de [Marcondes and Valk \(2020\)](#).

## 2.4 *Estrutura do trabalho*

Esta dissertação está estruturada da seguinte forma nos demais capítulos:

Uma revisão da literatura de CEP para processos em bateladas é apresentada no Capítulo 3.

O Capítulo 4 apresenta o detalhamento da abordagem de controle proposta nesta dissertação.

Um estudo de caso simulado será apresentado no Capítulo 5 para ilustrar a performance da abordagem proposta.

O Capítulo 6 apresenta as conclusões do trabalho e perspectivas de extensão da proposição apresentada.

---

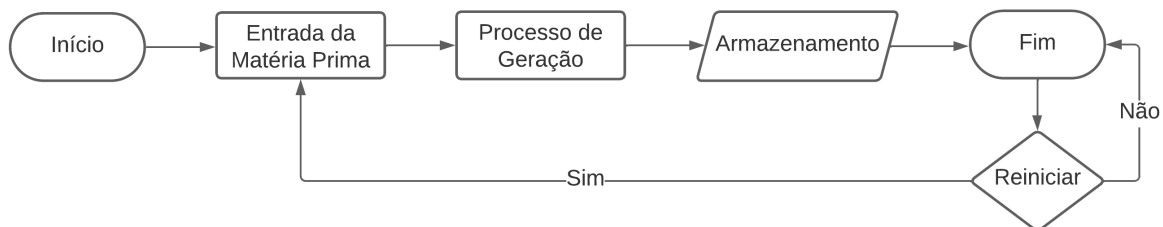
## CAPÍTULO 3

# CEP PARA PROCESSOS EM BATELADAS

---

Este capítulo destina-se à revisão teórica dos sobre Controle Estatístico de Processo para processos em bateladas. Os tópicos a serem apresentados nesta revisão incluem processo em bateladas, abordagens tradicionais de CEP e abordagens baseadas em séries temporais neste contexto.

Os processos em bateladas representam um tipo particular de operação industrial em que a produção ocorre em lotes finitos e cada lote é tratado como uma unidade única e independente. Como ele ocorre de maneira descontínua, a cada batelada é necessário que o sistema seja carregado com as matérias primas que passam pela etapa de produção, onde ocorre a transformação da matéria prima no produto (Tolentino, 2015). Uma vez finalizada a batelada, o produto final é armazenado para posterior distribuição, a Figura 3.1 ilustra o funcionamento.



**Figura 3.1:** Representação do funcionamento do processo em batelada. Fonte: Próprio autor

Esse tipo de processo é comumente encontrado em indústrias farmacêuticas, químicas, de alimentos e de manufatura. Um exemplo prático de processo em batelada é a fabricação de comprimidos, em que após os insumos serem pesados com precisão, eles são colocados em um misturador para melhor distribuição dos ativos, acrescenta-se os demais componentes necessários para formação do medicamento formando os grânulos, que passam em uma secadora industrial. Posteriormente são novamente misturados, passando pelo processo de compressão, onde são submetidos a alta pressão em moldes e assim gerar o comprimido a ser consumido e após a etapa final de embalagem.

Dentre as principais características de um processo em bateladas, tem-se a interrupção ao final de cada lote de produção, possibilitando assim a realização de ajustes, limpeza ou reconfiguração para o próximo lote. Essa característica apresenta desafios para o CEP, já que cada batelada pode apresentar peculiaridades, gerando variabilidade significativa entre as bateladas e, assim, necessidade

de monitoramento rigoroso para garantir a consistência e a qualidade do produto final.

Ainda um outro ponto importante a realização de diversas medições das variáveis durante o procedimento, para cada uma das bateladas (Westerhuis et al., 2000). Quando o processo é finalizado, tem-se disponível as informações das variáveis em questão durante todo o período de tempo da batelada. Desta forma as informações podem ser organizadas em uma matriz tridimensional ( $I \times K \times T$ ), sendo  $I$  o total de bateladas,  $K$  o número de variáveis e  $T$  a quantidade de instantes de tempo. Temos assim para cada batelada dados que descrevem séries temporais multivariadas das variáveis em análise.

A eficácia e a qualidade dos produtos resultantes desses processos são extremamente relevantes, sendo fundamental a realização do monitoramento. O CEP, torna-se uma excelente ferramenta para monitoramento e controle, além de poder proporcionar uma economia e otimização dos insumos, pois através dele é possível identificar e corrigir os problemas existentes, evitando assim um desperdício de matéria prima, energia e tempo. (Torminato, 2004)

Na literatura existem diversas abordagens de CEP para processos contínuos. Este tipo de processo, em que amostras são geradas de modo ininterrupto ao longo do tempo, vem sendo estudado desde o começo do século XX por Walter A. Shewhart e aprimorado por W. Edwards Deming e Joseph M. Juran, e sendo utilizada até os dias de hoje. Destacam-se a carta de controle para a média ( $\bar{X}$ ), utilizado para monitorar a média do processo e a carta de controle para amplitude (R), que avalia a variabilidade do processo. No entanto, a utilização dessas ferramentas se limitam a dados univariados, isto é, uma variável em análise por vez.

Num cenário de processo que geram dados multivariados, como os processos em bateladas, é essencial a utilização de cartas de controle que comportem esse formato. A abordagem mais tradicional para monitoramento de dados provenientes de processos multivariados é a carta de Hotelling (1947). Através da tradicional *estatística de Hotteling*, podemos monitorar simultaneamente um grupo de variáveis utilizando um vetor de médias do processo sob controle, permitindo detectar alterações concomitantes das variáveis. Além disso, por considerar as correlações cruzadas entre as variáveis, minimiza a ocorrência de alertas falsos. Entretanto, esta abordagem tal qual é estruturada é adequada em processos que operam de maneira contínua e que geram amostras não correlacionadas.

Nos processos em bateladas, é fundamental levar em conta a estrutura dos dados de um processo multivariado. Entretanto, sua estrutura de dados multivariados apresenta não apenas a correlação cruzada entre as variáveis (com no caso dos processos contínuos), mas também a correlação serial em cada variável. Além disso, é necessário modelar adicionalmente a variabilidade entre as bateladas. Pelos processos serem dinâmicos, faz com que seja necessário analisar os desvios das trajetórias das variáveis originais, assim como as autocorrelações cruzadas existentes (Kourti, 2003). Inúmeras metodologias foram sugeridas para monitorar esse tipo de processo.

Para avaliar dados multivariados mais extensos as técnicas de redução de dados são largamente empregadas. No contexto do CEP podemos destacar as Cartas de Controle via Análise de Componentes Principais (ACP), introduzidas por Jackson and Mudholkar (1979) e Jackson (1991). A ACP transforma um conjunto de variáveis correlacionadas em um conjunto menor de variáveis latentes não correlacionadas, reduzindo assim a dimensionalidade dos dados e identificando os componentes que explicam a maior parte da variabilidade observada. Ao fazer isso, a ACP facilita a visualização e a compreensão das interações entre as variáveis, facilitando a identificação de alterações nas fontes de

variabilidade no controle de processos multivariados.

Dentre os precursores na utilização de ACP para monitoramento de processos em batelada tem-se [Nomikos and MacGregor \(1994\)](#), que propôs a análise dos dados em uma matriz bidimensional ( $I \times K/T$ ), onde as linhas representam as bateladas e na coluna é realizada a combinação das variáveis e os instantes de tempo. Esta abordagem se denomina Análise de Componentes multidirecionais (ACPM). Entretanto, nesta abordagem as correlações seriais e cruzadas ficam confundidas, dificultando compreender onde ocorreu o descontrole [Ge et al. \(2013\)](#).

Uma extensão da implementação do ACPM, foi proposta por [Nomikos and MacGregor \(1995a\)](#), utilizando Mínimos Quadrados Parciais (MQP). Através desta abordagem, a avaliação não é feita apenas nas informações das variáveis durante a batelada, mas considera as informações das variáveis de avaliação da qualidade das bateladas finalizadas. Entretanto, a MQP é aplicada também a uma matriz de dados ( $I \times K/T$ ), misturando a correlação serial e cruzada das variáveis de processo.

Até hoje, ambas metodologias seguem sendo estudadas, em busca do monitoramento dos desvios das condições de operação, para assim garantir a qualidade dos produtos finais. [Camacho et al. \(2009\)](#) apresentou os principais métodos até então existentes para a realização do monitoramento de processos em batelada online, utilizando as metodologias de controle propondo aplicação de ACP em arranjos alternativos ao proposto por [Nomikos and MacGregor \(1994\)](#). Uma discussão ampliada sobre estes arranjos incluindo formas alternativas para acomodar a estrutura tridirecional ( $I \times K \times T$ ) dos processos em bateladas é encontrada em [Ramos et al. \(2021\)](#).

O monitoramento de processos em bateladas envolvendo a utilização de modelos de séries temporais, embora ainda escasso, vem ganhando atenção na literatura de CEP. As séries temporais representam uma sequência de dados observados ou coletados em intervalos de tempo regulares, e são usadas para analisar padrões temporais, tendências e sazonalidades. Ao integrar séries temporais ao CEP, é possível compreender melhor a dependência temporal dos dados de produção em bateladas.

A principal restrição ao uso da teoria de séries temporais no contexto dos processos em bateladas se deve ao fato de que tais processos geram dados históricos de mais de uma série temporal multivariada (uma para cada batelada). A teoria inferencial de séries não está desenvolvida para tratar com repetições de séries, e sim para construir estimativas para os parâmetros de algum modelo a partir de uma série apenas, isto é, de uma única amostra do processo estocástico gerador dos dados. Este fato trás uma limitação grande na proposição de abordagens de controle, visto que a variabilidade entre bateladas (ou entre séries temporais) não é acomodada em tal teoria.

A utilização de modelos de séries temporais univariados e multivariados é relativamente frequente na literatura de CEP para processos contínuos. Podemos citar o Modelo Autorregressivo de Médias Móveis (ARMA) ([Brockwell et al., 1991](#)) e Modelo Autoregressivo Multivariado (VAR) ([Lütkepohl, 2005](#)). A análise de séries temporais no contexto do CEP é uma estratégia poderosa para otimizar a eficiência e a qualidade dos processos, proporcionando uma boa compreensão para aprimoramentos e melhor gestão da produção industrial.

No contexto de processos em bateladas, [Choi et al. \(2008\)](#) propôs a realização de monitoramento a partir da combinação de ACP com um modelo VAR. Os dados históricos de bateladas de referência são primeiramente ajustados via um modelo VAR e os resíduos da modelagem são analisados via ACP.

Nesta abordagem, os coeficientes do VAR são estimados arranjando os dados com defasagem de lag  $p$  e ajustando uma regressão por MQP.

Wang et al. (2017) propôs o acompanhamento do processo utilizando um modelo Bidimensional Autoregressivo de Médias Móveis (2D-ARMA), com o intuito de monitorar a variabilidade dentro e entre as bateladas. Em cada batelada a regressão *stepwise* é utilizada e o operador LASSO é utilizado para definir o modelo final e os coeficientes estimados. A partir disso índices são criados para monitorar a variabilidade dentro e entre bateladas.

Uma outra metodologia utilizando séries temporais foi apresentada recentemente por Marcondes and Valk (2020) utilizando os Modelos Vetoriais Autoregressivos (VAR), onde realiza-se a estimação dos coeficientes VAR para cada batelada. Estas estimativas por batelada são agrupadas para formar um única estimativa e a estatísticas de Hotelling e da Variância Generalizada são utilizadas para monitorar novas bateladas, comparando os coeficientes estimados dessa batelada com o grupo de estimativas das bateladas de referência.

As cartas de controle utilizando inferência Bayesiana foram originalmente introduzidas por Girshick and Rubin (1952). A utilização de abordagens baseadas em inferência Bayesiana para monitoramento de processos em batelada tem se tornado um objeto de estudo cada vez mais importante.

Wasserman and Sudjianto (1993) apresentou uma CC para monitorar para monitorar um processo que segue um modelo linear dinâmico de segunda ordem, o qual consiste em calcular a média e a variância da distribuição *a posteriori* do processo. Em seguida, ela desenha uma faixa de confiança e identifica qualquer item fora dessa faixa como um sinal de que o processo está fora de controle.

Recentemente, Permin et al. (2018) propôs uma abordagem de monitoramento, em que a partir de um amplo histórico de bateladas, são selecionados algumas amostras de diferentes bateladas. Com os subconjuntos menores são calculadas a semelhança dos dados, assim cada subgrupo, estará refletindo uma abordagem distinta dentro do processo. A probabilidade *a priori* é calculada utilizando as informações das semelhanças obtidas e na sequência são calculadas probabilidades *a posteriori*, através do Fator de Bayes. Com base nessas informações atualizadas, o modelo mais semelhante é selecionado e empregado para conduzir o CEP.

Embora exista uma gama de técnicas eficientes e validadas para realização do monitoramento de processos em batelada, uma característica em comum entre elas é a necessidade de um tamanho mínimo de bateladas de referência, para que a detecção dos desvios do processo seja realmente eficaz. No trabalho mais recente de Marcondes and Valk (2020) é evidenciado a eficiência e o poder de detecção de descontroles na metodologia proposta, entretanto, um número bastante grande de bateladas de referência é empregado para construção dos limites nos gráficos de controle propostos.

Esta dissertação tem como foco o desenvolvimento de uma alternativa para monitoramento de processo em bateladas, com um número reduzido de bateladas de referência disponíveis. Para tanto, propõe-se o desenvolvimento de uma regra de decisão combinando o VAR Bayesiano com a distribuição *a priori* Minnesota (Wind, 2015) e o Fator de Bayes.

---

## CAPÍTULO 4

# METODOLOGIA

---

Neste capítulo será descrito a abordagem proposta neste trabalho, a qual consiste em uma regra de decisão bayesiana para classificar bateladas. Os tópicos que compõem a sua construção são: o Modelo Vetorial Autoregressivo (Lütkepohl, 2005), o VAR Bayesiano focando no BVAR Minnesota (Wind, 2015) e o Fator de Bayes (Jeffreys, 1939).

### 4.1 Modelo Vetorial Autoregressivo (VAR)

Através dos Modelos Vetorial Autorregressivos (VAR), proposto por Sims (1980), é possível expressar um conjunto de variáveis em função das suas observações defasadas no tempo, considerando as interrelações entre as variáveis. O modelo VAR é muito usado no contexto da economia para ajustar um conjunto de séries temporais de variáveis macroeconômicas com a finalidade de analisar suas interações e realizar previsões.

Conforme Lütkepohl (2005) podemos descrever o modelo VAR( $p$ ) de  $K$  variáveis e  $p$  defasagens como

$$\mathbf{z}_t = \mathbf{\Phi}_0 + \mathbf{\Phi}_1 z_{t-1} + \dots + \mathbf{\Phi}_p z_{t-p} + u_t, \quad (4.1)$$

onde  $\mathbf{z}_t$  é um vetor de dimensão  $k$ ,  $\mathbf{\Phi}_0$  é o vetor de constantes  $K \times 1$ ,  $\mathbf{\Phi}_j$  são os coeficientes onde  $j = 1, 2, \dots, p$  sendo matrizes  $K \times K$  e  $u_t$  é um ruído branco através de um vetor  $K \times 1$ . Sendo que  $\mathbb{E}(u_t) = 0$ ,  $\text{Var}[u_t] = \Gamma$  e  $\text{COV}[\mathbf{u}_t, \mathbf{u}_{t+h}] = \mathbf{0}$  para todo  $h = 1, 2, 3, \dots$ . Para o modelo VAR( $p$ ) ser descrito como estável, todos os autovalores de  $\mathbf{\Phi}$ , em módulo, devem ser menor que 1, assim  $\det(\mathbf{I}_k - \sum_{j=1}^p \mathbf{\Phi}_j w^j) \neq 0$  para  $|w| \leq 1$ .

Considerando que  $T$  é o número de instantes de tempo, podemos então reescrever o VAR como

$$\mathbf{Z} = \mathbf{X}\mathbf{\Phi} + \mathbf{U}, \quad (4.2)$$

em que  $\mathbf{Z}$  é uma matriz de variáveis dependentes de dimensão  $T \times K$ ,  $\mathbf{X}$  é a matriz de regressores de dimensão  $T \times l$ ,  $\mathbf{U}$  é a matriz de erros de dimensão  $T \times K$  e  $\mathbf{\Phi}$  é matriz dos parâmetros da regressão  $l \times K$ , com  $l = 1 + pK$  representando o número de parâmetros regressores por equação no VAR.

As matrizes são explicitamente estruturadas do seguinte modo:  $\mathbf{Z} = [\{z_t\}_{t=1}^T]'$ , onde  $z_t$  é o vetor de dimensão  $K \times 1$ , e  $\mathbf{X} = [\{x_t\}_{t=1}^T]$  com  $x_t = [1, \{z'_{t-i}\}_{i=1}^p]'$ . Isto implica que a  $i$ -ésima coluna da



matriz  $\Phi$  traz os parâmetros da regressão representada pela  $i$ -ésima equação do VAR.

Através da regra da vetorização, em que  $\otimes$  é o produto Kronecker e  $\text{vec}(\cdot)$  o vetor operador coluna, conforme descrito em [Lütkepohl \(2005\)](#), permitindo assim que

$$z = (\mathbf{I}_K \otimes X) \phi + u, \quad (4.3)$$

onde o vetor de variáveis dependentes  $z = \text{vec}(Z)$  possui dimensão  $TK \times 1$ , o vetor de parâmetros regressores  $\phi = \text{vec}(\Phi)$  tem dimensão  $IK \times 1$ . O vetor de erros  $u = \text{vec}(U)$  possui dimensão  $TK \times 1$ , sendo  $u \sim N(\mathbf{0}, \Sigma \otimes \mathbf{I}_T)$ .

Os coeficientes do modelo VAR podem ser estimados via mínimos quadrados através de

$$\hat{\phi} = ((X'X)^{-1} \otimes \mathbf{I}_K)z. \quad (4.4)$$

Sob condições de regularidade, conforme demonstrado em [Lütkepohl \(2005\)](#),  $\hat{\phi}$  possui distribuição assintótica Normal Multivariada. Explicitamente temos

$$T^{\frac{1}{2}}(\hat{\phi} - \phi) = T^{\frac{1}{2}}\text{Vec}(\hat{\Phi} - \Phi) \xrightarrow{L} \mathcal{N}(0, \Gamma^{-1} \otimes \Sigma_u) \quad \text{quando } T \rightarrow \infty, \quad (4.5)$$

onde  $\frac{1}{T}XX' \xrightarrow{P} \Gamma$  e  $\Sigma_u = \mathbf{I}_T \otimes \Sigma_u$ . A notação  $\xrightarrow{L}$  implica convergência em distribuição e  $\xrightarrow{P}$  indica convergência em termos de probabilidade.

## 4.2 VAR Bayesiano (BVAR)

O conteúdo desta sessão está baseado principalmente no material de [Wind \(2015\)](#). Na formulação do BVAR definimos inicialmente a função de verossimilhança representada pela distribuição amostral de  $p(z|\phi, \Sigma)$ , vista em função dos parâmetros. Assumindo distribuição Normal Multivariada, condicionada aos parâmetros temos que

$$p(z|\phi, \Sigma) \propto |\Sigma \otimes \mathbf{I}_T|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}u'(|\Sigma \otimes \mathbf{I}_T|)^{-1}u\right) \quad (4.6)$$

$$\propto |\Sigma|^{-\frac{T}{2}} \exp\left(-\frac{1}{2}\text{tr}\left((Z - X\Phi)'(Z - X\Phi)\Sigma^{-1}\right)\right). \quad (4.7)$$

Ao retirar a constante de proporcionalidade, é possível decompor a função de verossimilhança em dois componentes, um onde a distribuição Normal (condicional) de  $\phi$  dado  $\Sigma$  e a outra na distribuição inversa-Wishart (marginal) para  $\Sigma$ . Utilizando a regra de decomposição  $(Z - X\Phi)'(Z - X\Phi) = (Z - X\hat{\Phi})'(Z - X\hat{\Phi}) + (\hat{\Phi} - \Phi)'X'X(\hat{\Phi} - \Phi)$ , onde  $\hat{\Phi} = (X'X)^{-1}X'Z$  é a estimativa de mínimos quadrados ordinários para  $\phi$ , tem-se que

$$p(z|\phi, \Sigma) \propto |\Sigma|^{-\frac{K}{2}} \exp\left(-\frac{1}{2}\text{tr}\left((\hat{\Phi} - \Phi)'X'X(\hat{\Phi} - \Phi)\Sigma^{-1}\right)\right) \quad (4.8)$$

$$\propto |\Sigma|^{-\frac{T-K}{2}} \exp\left(-\frac{1}{2}\text{tr}\left((Z - X\hat{\Phi})'(Z - X\hat{\Phi})\Sigma^{-1}\right)\right). \quad (4.9)$$

A equação (4.8) representa o kernel da distribuição Normal e a (4.9) representa o kernel da distribuição inversa de Wishart. É conveniente escrever no contexto matricial a distribuição em termos da distribuição Normal Multivariada Wind (2015). Assim temos

$$\boldsymbol{\phi}|\boldsymbol{\Sigma}, z \sim N(\widehat{\boldsymbol{\phi}}, \boldsymbol{\Sigma} \otimes (\mathbf{X}'\mathbf{X})^{-1}), \quad (4.10)$$

$$\boldsymbol{\Sigma}|z \sim IW(\widehat{\mathbf{S}}, T - l - K - 1), \quad (4.11)$$

onde  $\widehat{\boldsymbol{\phi}} = \text{vec}(\widehat{\boldsymbol{\Phi}})$  e  $\widehat{\mathbf{S}} = (\mathbf{Z} - \mathbf{X}\widehat{\boldsymbol{\Phi}})'(\mathbf{Z} - \mathbf{X}\widehat{\boldsymbol{\Phi}})$ . Sendo plausível conseguir  $\widehat{\boldsymbol{\phi}} = (\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'z$  com  $\mathbf{x} = (\mathbf{I}_K \otimes \mathbf{X})$  da forma apresentada em (4.3).

De uma maneira geral, assumindo uma distribuição *a priori* padrão, temos uma distribuição condicional Normal para  $\boldsymbol{\phi}$  dado  $\boldsymbol{\Sigma}$  e uma distribuição marginal inversa de Wishart para  $\boldsymbol{\Sigma}$ . As estimativas pontuais *a posteriori* dos parâmetros são dadas por

$$\boldsymbol{\phi}^* = \widehat{\boldsymbol{\phi}} \quad (4.12)$$

e

$$\boldsymbol{\Sigma}^* = \frac{\widehat{\mathbf{S}}}{T}. \quad (4.13)$$

Considerando *a priori* Uniforme imprópria, temos a distribuição *a posteriori* centrada em torno das estimativas de mínimos quadrados ordinários. Ainda sob *a priori* não informativa de Jeffreys, somente os graus de liberdade da distribuição inversa de Wishart aumentam para  $K+1$ .

#### 4.2.1 *Priori e Posteriori Minnesota*

A metodologia BVAR utilizando *a priori* Minnesota proposta por Doan et al. (1984) e Litterman (1986), assume  $\boldsymbol{\Sigma}$  conhecido (ou pré-estimado) e a distribuição *a priori* Normal para os parâmetros  $\boldsymbol{\Phi}$ . Tem-se

$$\boldsymbol{\phi} \sim N(\underline{\boldsymbol{\phi}}, \underline{\mathbf{V}}), \quad (4.14)$$

em que  $\underline{\boldsymbol{\phi}}$  é um vetor de dimensão  $Kl \times 1$  com as médias *a priori* e  $\underline{\mathbf{V}}$  uma matriz de covariância *a priori* de dimensões  $Kl \times K$ .

A ideia subjacente da *priori* Minnesota é encurtar/direcionar o modelo para um passeio aleatório, atenuando os coeficientes das defasagens maiores e da correlação cruzada entre as variáveis, isso implica que

$$\underline{\boldsymbol{\phi}} = \text{vec}([0_{K \times 1} \quad \mathbf{I}_{K \times K} \quad 0_{K \times K(p-1)}]'), \quad (4.15)$$

onde  $\underline{\boldsymbol{\phi}}$  é um vetor de zeros, exceto para os elementos que correspondem a primeira defasagem.

A covariância *a priori* ( $\underline{\mathbf{V}}$ ) é uma matriz diagonal. Sendo  $\underline{v}_i$  um vetor com os elementos diagonais associados aos parâmetros da  $i$ -ésima equação VAR, onde  $\underline{v}_i = \text{diag}(\{v_{ij}\}_{j=1}^K)$  e determinamos que  $v_{iq}$  é o  $q$ -ésimo elemento de  $\underline{v}_i$ . Considerando que  $v_{i1}$  está associado com o termo constante, a medida que  $v_{iq}$ ,  $q \neq 1$  está relacionado com o  $r$ -ésima defasagens da variável  $j$  com  $r = \frac{q-2}{K} + 1$  e

$j = q - 1 - \lfloor \frac{q-2}{K} \rfloor K$ . Normalmente escolhe-se os elementos de  $\underline{v}_i$  para *a priori* Minnesota como Wind (2015), dada por

$$\underline{v}_{iq} = \begin{cases} \frac{\pi_1}{r^2} & \text{se } q \neq 1 \text{ e } j = i \text{ (para coeficientes de defasagem próprias de ordem } r) \\ \frac{\pi_2}{r^2} \frac{\sigma_{ii}}{\sigma_{jj}} & \text{se } q \neq 1 \text{ e } j \neq i \text{ (para coeficientes de defasagem externa de ordem } r) \\ \pi_3 \sigma_{ii} & \text{se } q = 1 \text{ (para coeficientes em termos determinísticos),} \end{cases} \quad (4.16)$$

na qual  $\sigma_{ii}$  é o  $i$ -ésimo elemento da diagonal de  $\Sigma$ .

Os termos  $\sigma_{ii}$  são utilizados para demonstrar as diferenças na escala das variáveis do modelo. As definições acima requerem que a redução seja mais forte em defasagens distantes e entre variáveis em que  $\pi_1 > \pi_2$ . Geralmente  $\pi_3$  é determinado grande, ocasionando em uma diminuição fraca para os termos determinísticos.

Como  $\Sigma$  é conhecido pode-se reduzir a função de verossimilhança, sendo necessário apenas do termo  $\exp(\cdot)$ , conhecido como núcleo da matriz da distribuição Normal, podendo assim reescrever a matriz kernel Normal em função do kernel multivariado Normal, conforme

$$p(\mathbf{z}|\boldsymbol{\phi}) \propto \exp\left(-\frac{1}{2}(\boldsymbol{\phi} - \hat{\boldsymbol{\phi}})' \hat{\mathbf{V}}^{-1}(\boldsymbol{\phi} - \hat{\boldsymbol{\phi}})\right), \quad (4.17)$$

onde

$$\hat{\mathbf{V}} = \Sigma \otimes (\mathbf{X}'\mathbf{X})^{-1}. \quad (4.18)$$

Combinando a função de verossimilhança com a distribuição *a priori*, ao multiplicar o kernel da verossimilhança com o kernel *priori* obtêm-se o kernel posteriori, dado por

$$p(\boldsymbol{\phi}|\mathbf{z}) \propto \exp\left(-\frac{1}{2}(\boldsymbol{\phi} - \hat{\boldsymbol{\phi}})' \hat{\mathbf{V}}^{-1}(\boldsymbol{\phi} - \hat{\boldsymbol{\phi}})\right) \exp\left(-\frac{1}{2}(\boldsymbol{\phi} - \underline{\boldsymbol{\phi}})' \underline{\mathbf{V}}^{-1}(\boldsymbol{\phi} - \underline{\boldsymbol{\phi}})\right). \quad (4.19)$$

Como a soma de duas expressões quadráticas resulta em apenas uma expressão quadrática, então a multiplicação de dois termos  $\exp(\cdot)$  produz apenas um termo  $\exp(\cdot)$ , no caso um kernel Normal multivariado, dados por

$$p(\boldsymbol{\phi}|\mathbf{z}) \propto \exp\left(-\frac{1}{2}(\boldsymbol{\phi} - \bar{\boldsymbol{\phi}})' \bar{\mathbf{V}}^{-1}(\boldsymbol{\phi} - \bar{\boldsymbol{\phi}})\right). \quad (4.20)$$

Ao combinar os coeficiente é possível obter  $\bar{\boldsymbol{\phi}}$  e  $\bar{\mathbf{V}}$ , gerando assim a distribuição *a posteriori* definida como Wind (2015)

$$\boldsymbol{\phi}|\mathbf{z} \sim N(\bar{\boldsymbol{\phi}}, \bar{\mathbf{V}}), \quad (4.21)$$

onde

$$\bar{\boldsymbol{\phi}} = (\hat{\mathbf{V}}^{-1} + \underline{\mathbf{V}}^{-1})^{-1} (\hat{\mathbf{V}}^{-1} \hat{\boldsymbol{\phi}} + \underline{\mathbf{V}}^{-1} \underline{\boldsymbol{\phi}}), \quad (4.22)$$

$$\bar{\mathbf{V}} = (\hat{\mathbf{V}}^{-1} + \underline{\mathbf{V}}^{-1})^{-1}. \quad (4.23)$$

Dessa forma, *a priori* Minnesota gera uma distribuição *a posteriori* Normal para  $\boldsymbol{\phi}$ , com média e variância dados por (4.22) e (4.23), respectivamente.

É possível perceber que a média *a posteriori* de  $\phi$  é uma combinação ponderada das médias *a priori* do estimador  $\underline{\phi}$  com o estimador de máxima verossimilhança  $\widehat{\phi}$ , com pesos  $\bar{V}\underline{V}^{-1}$  e  $\bar{V}\widehat{V}^{-1}$ , respectivamente.

Os pesos são controlados pelo usuário, através de valores dos hiperparâmetros de entrada em  $\underline{V}$ . Entretanto os pesos são endógenos, uma vez que os dados estão inseridos em  $\widehat{V}$ , assim sendo a matriz de covariâncias acomoda a incerteza dos parâmetros.

### 4.3 Fator de Bayes

O Fator de Bayes (FB) é uma metodologia que utiliza a razão entre duas probabilidades para quantificar a comparação entre dois modelos, sendo introduzido por [Jeffreys \(1939\)](#). O FB basicamente compara as verossimilhanças marginais de cada um dos modelos.

Considere duas hipóteses, sendo a hipótese nula ( $H_0$ ) a hipótese que corresponde ao modelo  $M_0$  e a hipótese alternativa ( $H_1$ ), referente ao modelo  $M_1$ . Deseja-se verificar qual modelo é mais verossímil, dado um vetor de observações  $\mathbf{z}$ .

Tem-se uma probabilidade *a priori* para cada um dos modelos considerados, denominados como  $P(M_0)$  e  $P(M_1)$ . Levando em consideração os dados, obtém-se a probabilidade *a posteriori* de ambos os modelos, sendo  $P(M_0|\mathbf{z})$  e  $P(M_1|\mathbf{z})$ .

O Teorema de Bayes para a probabilidade *a posteriori* é descrito por

$$P(M_k|\mathbf{z}) = \frac{P(\mathbf{z}|M_k)P(M_k)}{P(\mathbf{z})}, \quad (4.24)$$

onde  $k = 0, 1$  e  $P(\mathbf{z}) = P(\mathbf{z}|M_0)P(M_0) + P(\mathbf{z}|M_1)P(M_1)$ .

Manipulando a eq. (4.24) podemos obter explicitamente em função da probabilidade dos dados, considerando eles provenientes de  $M_k$ , da forma

$$P(\mathbf{z}|M_k) = \frac{P(\mathbf{z})P(M_k|\mathbf{z})}{P(M_k)}, \quad (4.25)$$

onde  $k = 0, 1$ .

O FB representa a verossimilhança marginal sob o modelo  $M_1$  em comparação à verossimilhança marginal sob o modelo  $M_0$ , desse modo se torna possível compreender qual dos dois modelos explica melhor a origem dos dados. Temos então o FB dado por

$$FB = \frac{P(\mathbf{z}|M_1)}{P(\mathbf{z}|M_0)}. \quad (4.26)$$

[Jeffreys \(1961\)](#) apresentou uma escala de interpretação para o escore FB obtido, classificando os valores em quatro intervalos. Além da interpretação através do valor exato do FB, é possível utilizar o logaritmo dele na análise, com o intuito de simplificar os cálculos e contornar eventuais problemas de instabilidade numérica. A Tabela 4.1 apresenta as faixas de classificação para ambas as escalas.

**Tabela 4.1:** Valores de FB e interpretação

FB	log (FB)	Indicação sobre $M_1$
1 - 3,2	0 - 0,5	Insignificância
3,2 - 10	0,5 - 1	Significativa
10 - 100	1 - 2	Forte
> 100	> 2	Decisiva

Desse modo se o valor obtido está no intervalo da primeira classificação, as evidências referentes a hipótese de que os dados vem do modelo  $M_1$  são mínimas, sendo assim provenientes do modelo  $M_0$ . No segundo intervalo, as evidências referentes a hipótese de que os dados correspondem ao modelo  $M_1$  aumentam consideravelmente. E assim gradativamente até o quarto intervalo que sinaliza forte evidência de que os dados pertencem ao modelo  $M_1$ . É importante destacar que o  $\log(FB)$  assume valores menores do que zero (assim como  $FB < 1$ ) e, neste caso, devemos entender a existência de evidências em favor do modelo  $M_0$ .

#### 4.4 Abordagem de monitoramento

Inicialmente temos que considerar as expressões das verossimilhanças marginais de (4.26). Para tanto devemos considerar explicitamente a dependência dos parâmetros de cada modelo. Note que

$$P(z) = \frac{P(z|\phi)P(\phi)}{P(\phi|z)}, \quad (4.27)$$

onde  $z$  e  $\phi$  são como definidos em (4.3). Sabemos que no VAR bayesiano Minnesota as densidades de probabilidade *a priori* e *a posteriori* seguem distribuição Normal (resultados nas expressões (4.6), (4.14), (4.20) e (4.21)). Com base nestes resultados, podemos escrever explicitamente os elementos de (4.27) como

$$P(z|\phi) = (2\pi)^{\frac{\gamma_y}{2}} |S|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} [z - (I_k \otimes X)\phi]' S^{-1} [z - (I_k \otimes X)\phi] \right\}, \quad (4.28)$$

$$P(\phi) = (2\pi)^{\frac{\gamma_\phi}{2}} |\underline{V}|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (\phi - \underline{\phi})' \underline{V}^{-1} (\phi - \underline{\phi}) \right\}, \quad (4.29)$$

$$P(\phi|z) = (2\pi)^{\frac{\gamma_\phi}{2}} |\bar{V}|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (\phi - \bar{\phi})' \bar{V}^{-1} (\phi - \bar{\phi}) \right\}, \quad (4.30)$$

onde  $\underline{\phi}$ ,  $\underline{V}$ ,  $\hat{V}$ ,  $\bar{\phi}$  e  $\bar{V}$  estão definidos, respectivamente, nas expressões (4.15), (4.16), (4.18), (4.22) e (4.23), com  $S = \Sigma \otimes I_T$  (Kernel da eq.(4.6)).

Note que a verossimilhança marginal em (4.26) resulta de uma integral para todos os valores de  $\phi$  e portanto independem do valor de  $\phi$ . Por outro lado, esta expressão será calculada através da expressão (4.27), onde um valor é necessário. Por tanto, sem perda de generalidade podemos assumir que  $\phi = 0$ . Adicionalmente consideramos em (4.15) da priori Minnesota  $\underline{\phi} = 0$  e sabemos que  $\underline{V}$

(4.16) é uma matriz diagonal. Desta forma, as equações acima podem ser escritas de forma reduzida como se segue

$$P(\mathbf{z}|\boldsymbol{\phi}) = (2\pi)^{\frac{Jy}{2}} |\underline{S}|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \mathbf{z}' \underline{S}^{-1} \mathbf{z} \right\}, \quad (4.31)$$

$$P(\boldsymbol{\phi}) = |\underline{V}|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \boldsymbol{\phi}' \underline{V}^{-1} \boldsymbol{\phi} \right\}, \quad (4.32)$$

$$P(\boldsymbol{\phi}|\mathbf{z}) = |\bar{V}|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (\bar{\boldsymbol{\phi}} \bar{V}^{-1} \bar{\boldsymbol{\phi}}) \right\}. \quad (4.33)$$

Considere um processo em bateladas tal qual definido na seção 2.2, isto é, cada batelada gera uma série temporal de tamanho  $T$  instantes com observações de  $K$  variáveis de processo. Supondo que a dinâmica do processo é bem descrita por um modelo VAR de ordem  $p$  com vetor de parâmetros  $\boldsymbol{\phi} = \text{vec}(\boldsymbol{\Phi})$ , com  $\boldsymbol{\Phi}$  definido na expressão (4.2), onde as observações defasadas na ordem do VAR estão arranjadas em um vetor  $\mathbf{z}$ , onde  $\mathbf{z} = \text{vec}(\underline{Z})$  (como definido em (4.3)), com dimensão  $J \times 1$ , onde  $J = TK$ . Vamos também admitir que dispomos de um número  $I$  de bateladas históricas observadas no processo sob controle e uma nova batelada sob análise.

Seguindo a ideia do FB descrita na seção 2.5 definimos primeiramente o nosso modelo  $M_0$ . Deseja-se determinar a verossimilhança marginal dos dados assumindo que a nova batelada pertence ao processo sob controle, isto é, a nova série temporal gerada não apresenta diferença significativa em relação às características das séries de referência.

Para tanto empilhamos as  $I$  bateladas de referência, isto é, colamos as  $I$  bateladas de tal forma que ao final da última informação da primeira batelada, na sequência estará a primeira informação da segunda batelada e assim sucessivamente até a  $I$ -ésima batelada. Chamamos de  $\mathbf{z}_I$  o vetor de dimensão  $(IJ \times 1)$  composto pela séries temporais coladas contendo observações defasadas na ordem  $p$  do VAR.

A seguir, definimos como  $\mathbf{z}^*$  o vetor de dimensão  $J \times 1$  de observações da nova batelada sob análise e colamos este vetor ao vetor  $\mathbf{z}_I$ . Para  $M_0$  estamos interessados em calcular  $P(\mathbf{z}_0)$ , onde  $\mathbf{z}_0 = [\mathbf{z}_I, \mathbf{z}^*]$ , isto é, quão provável é este vetor assumir uma grande série temporal da variável gerada de um processo em que nenhuma eventual distúrbio ocorreu incidindo na batelada nova. Em outras palavras, aqui supomos que esta grande amostra empilhada foi gerada pelo VAR subjacente com vetor de parâmetros  $\boldsymbol{\phi}$ . Podemos entender que se a nova batelada também é gerada num processo sob controle e considerando a mesma distribuição *a priori* (Minnesota), esperamos que a informação vinda dos dados das  $I + 1$  bateladas formando o vetor  $\mathbf{z}_0$  gerem estimativas próximas aos parâmetros em  $\boldsymbol{\phi}$ .

Posteriormente, definimos o modelo  $M_1$ , onde a ideia é obter separadamente  $P(\mathbf{z}_I)$  e  $P(\mathbf{z}^*)$  e multiplicá-las. Na verdade aqui estamos decompondo  $P(\mathbf{z}_0)$  como  $P(\mathbf{z}_I) \times P(\mathbf{z}^*)$ , isto é, obtendo a verossimilhança do vetor  $\mathbf{z}_0$  através do produto das verossimilhanças de  $\mathbf{z}_I$  e  $\mathbf{z}^*$ , o que denominaremos por  $P(\mathbf{z}_1) = P(\mathbf{z}_I)P(\mathbf{z}^*)$ . Dessa forma, levando em consideração novamente que a nova batelada esteja sob controle e a mesma distribuição *a priori*, esperamos que a informação vinda dos dados considerando separadamente apenas a batelada nova  $\mathbf{z}^*$  e as  $I$  bateladas em  $\mathbf{z}_I$  gerem estimativas próximas entre si e aos parâmetros em  $\boldsymbol{\phi}$ .

O FB irá comparar o modelo  $M_1$  ao modelo  $M_0$  e é explicitamente descrito como

$$FB = \frac{P(z_1)}{P(z_0)} = \frac{P(z_I) * P(z_*)}{P(z_I, z_*)}, \quad (4.34)$$

onde  $P(z_I)$ ,  $P(z_*)$  e  $P(z_0)$  são obtidos através da eq. (4.27), em que cada elemento é obtido de acordo com as eqs.(4.31), (4.32) e (4.33).

Em resumo, sabemos que as  $I$  bateladas históricas foram geradas num processo sob controle e seguem um  $VAR(p)$  com vetor de parâmetros  $\phi = \phi_0$ . Supondo que a nova batelada sob investigação seja gerada por um  $VAR(p)$  com vetor de parâmetros  $\phi_1 \neq \phi_0$ , neste caso, na expressão (4.34), é fácil notar que a verossimilhança marginal  $P(z_1)$  deve ser superior a  $P(z_0)$  e FB tende a ser maior do que 1. Caso contrário, isto é, considerando a nova batelada também sob controle, FB deve ser menor do que 1. Neste caso é importante entender que valores próximos a 1 indicam falta de evidências em favor de ambos os modelos e valores de  $FB$  menores do que 1 indicam maior verossimilhança em favor do modelo  $M_0$ . Entretanto, na nossa regra de classificação  $FB < 1$  nos indica falta de evidências de que a nova batelada sob investigação esteja em desacordo com as históricas de referência, portanto, ela será considerada como sob controle.

A Tabela 4.1 contém os valores de  $FB$  e  $\log(FB)$  considerados significativos em favor de  $M_1$ . Para fins de simplificação das densidades envolvidas e diminuição do tempo computacional, neste trabalho utilizaremos o  $\log(FB)$ . Dessa forma, reescrevemos a expressão (4.34) da seguinte forma

$$\log FB = \log P(z_I) + \log P(z_*) - \log P(z_0). \quad (4.35)$$

No sentido de formalizar a metodologia proposta e operacionalizar sua utilização, definimos o seguinte teste de hipóteses:  $H_0: \phi = \phi_0$  contra  $H_1: \phi \neq \phi_0$ . Utilizando a estatística  $\log(FB)$  no arranjo definido em (4.35) para classificar a nova batelada sob investigação e considerando o valor mínimo de  $\log(FB)$  aceito como significativo na tabela 4.1, definimos a seguinte regra de decisão

Dado uma nova batelada sob análise,

- Se  $\log(FB) \leq 0.5$ , não rejeita  $H_0$ , isto é, não há evidências de que a nova batelada tenha sido gerada em um processo fora de controle.
- Se  $\log(FB) > 0.5$ , rejeita  $H_0$ , isto é, há evidências de que a nova batelada foi obtida em um processo fora de controle.

---

## CAPÍTULO 5

# ESTUDO DE CASO SIMULADO

---

Esta seção apresenta o estudo de caso simulado ilustrativo da metodologia proposta, o qual está baseado no cenário apresentado por [Marcondes and Valk \(2020\)](#). Nos parágrafos abaixo iremos ilustrar resumidamente a abordagem proposta por estes autores (denominaremos *abordagem baseada no VAR*), a qual será utilizada como comparativo da abordagem proposta nesta dissertação. Em seguida, o estudo de caso simulado utilizado será apresentado com detalhes e apontado os resultados obtidos. Ademais, será realizada uma comparação entre a abordagem proposta e a de [Marcondes and Valk \(2020\)](#).

### 5.1 Abordagem baseada no VAR

A abordagem de [Marcondes and Valk \(2020\)](#) busca estimar pra cada batelada de referência os coeficientes de um VAR( $p$ ) e agrupar estas estimativas em um vetor de médias e uma matriz de covariância. A estatística de Hotelling  $\mathcal{T}_\phi^2$  e da Variância Generalizada  $\mathcal{W}_\phi$  são propostas para avaliar as estimativas dos coeficientes de uma nova batelada sob monitoramento.

Considere bateladas que geram séries temporais  $K$  variadas ( $K$  variáveis em análise) e são bem descritas por um modelo VAR de ordem  $p$  (VAR( $p$ )). A cada batelada, obtém-se então o vetor  $\hat{\phi}$  de estimativas dos coeficientes, de comprimento  $K+pK^2$ . As cartas de controle são escritas explicitamente da seguinte forma:

$$\mathcal{T}_\phi^2 = (\hat{\phi}_{new} - \bar{\phi})' \bar{\Sigma}_\phi^{-1} (\hat{\phi}_{new} - \bar{\phi}) \sim \frac{C(I+1)(I-1)}{I(I-C)} F_{C, I-C} \quad (5.1)$$

e

$$\mathcal{W}_\phi = -C(T-1) + C(T-1)\ln(T-1) - (T-1)D + E \sim \chi_{C(C+1)/2}^2, \quad (5.2)$$

onde  $\hat{\phi}_{new}$  é o vetor de estimativas da batelada nova,  $\bar{\phi}$  e  $\bar{\Sigma}_\phi^{-1}$  representam, respectivamente, o vetor média e a matriz de covariância média (de dimensão  $K+pK^2 \times K+pK^2$ ) considerando os coeficientes estimados de cada uma das  $I$  bateladas de referência. As quantidades  $D = \ln(\det(\mathbf{A})/\det(\bar{\Sigma}_\phi))$  e  $E = \text{tr}(\bar{\Sigma}_\phi^{(-1)} \mathbf{A})$ .  $C$  é comprimento de  $\bar{\phi}$  (ou de  $\hat{\phi}_{new}$ ) e  $\mathbf{A} = (T-2)\hat{\Sigma}_{\phi_{new}}$ , onde  $\hat{\Sigma}_{\phi_{new}}$  representa a matriz de covariâncias estimadas de  $\hat{\phi}_{new}$ .

Escores acima do percentil  $\alpha$  em  $\mathcal{T}_\phi^2$  ou  $\mathcal{W}_\phi$  sugerem que a estrutura de correlação cruzada e/ou serial das variáveis considerando a defasagem de primeira ordem está significativamente diferente do



esperado tomando como base as séries temporais oriundas das bateladas de referência.

## 5.2 Definindo os Cenários

Assumindo que as leis físicas de um processo em bateladas seguem um modelo VAR(1) sem intercepto e duas variáveis em análise. Temos assim um processo que gera séries temporais bivariadas, ou seja, medições sucessivas de duas variáveis estão disponíveis em cada batelada. Num processo operando sob controle assumimos bateladas geradas por um VAR(1) bivariado com matriz de coeficientes  $\Phi$  em (4.2) explicitamente escrita como

$$\Phi = \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix} = \begin{bmatrix} -0.3 & 0.4 \\ 0.4 & 0.5 \end{bmatrix}. \quad (5.3)$$

A matriz  $\Phi$  contém na sua diagonal principal informação sobre a correlação serial de defasagem 1, em ambas as variáveis ( $x_1$  e  $x_2$ ). A diagonal secundária guarda uma medida de correlação cruzada da primeira variável com a segunda defasada de 1 instante de tempo e da segunda variável com a primeira defasada de 1 instante de tempo.

Para avaliar a abordagem de controle proposta, novas bateladas ( $I_{new}$ ) deste processo foram geradas apresentando diferentes níveis de descontrole, primeiramente para o parâmetro  $\phi_{11}$  e posteriormente para o parâmetro  $\phi_{12}$ . Adicionalmente, variamos o número de bateladas de referência ( $I$ ) e o tamanho das bateladas, isto é, o comprimento ( $T$ ) das séries temporais geradas. A Tabela 5.1 apresenta a configuração detalhada dos cenários que serão investigados.

**Tabela 5.1:** Configurações das Simulações

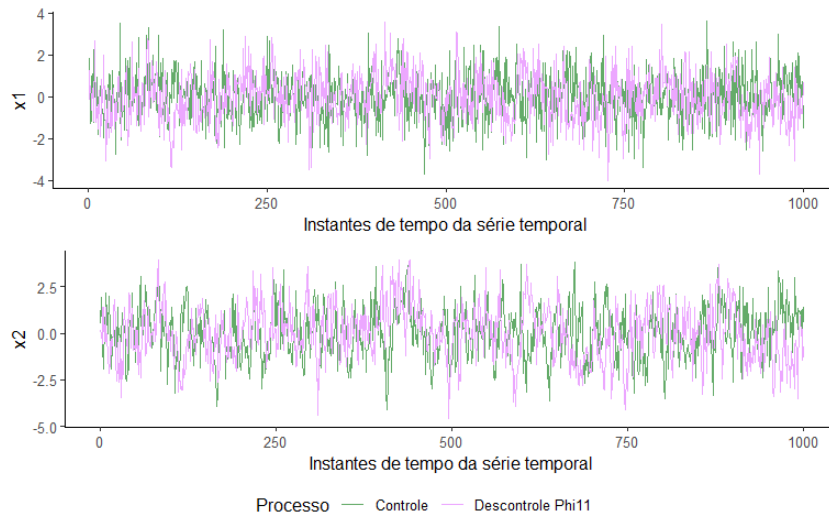
Parâmetro perturbado	Referência	Descontrole	$I$	$I_{new}$	$T$	Replicações
$\phi_{11}$	-0.3	-0.25, -0.2, -0.1, 0, 0.1, 0.3	5,10,30	150	300,500	100
$\phi_{12}$	0.4	0.6, 0.35, 0.3, 0.2,0, -0.2				

As configurações acima apresentadas nos permitem avaliar o poder da abordagem de controle proposta diante de diferentes níveis de correlação serial na primeira variável. Assim como para diferentes níveis de correlação cruzada entre as variáveis.

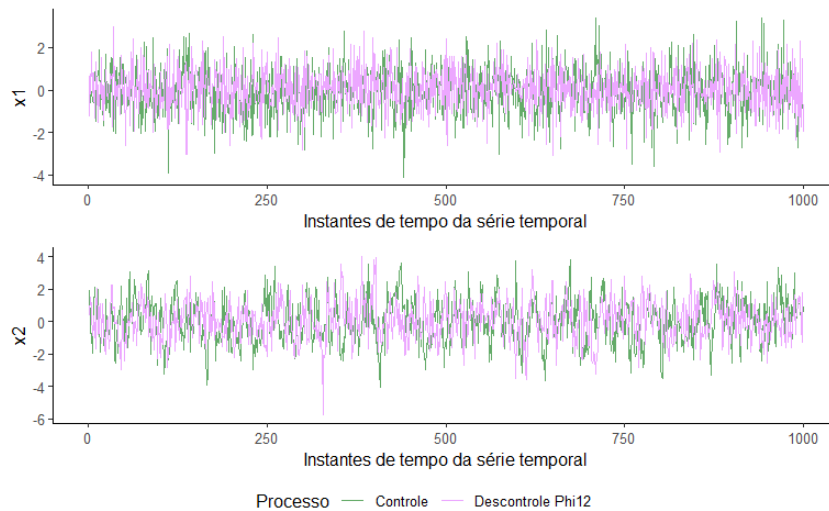
Em cada cenário variamos o número  $I \in (5, 10, 30)$  de bateladas de referência do processo sob controle gerado pelo VAR(1) com  $\Phi$  definido em (5.3). Também variamos o tamanho das séries temporais em 2 níveis,  $T \in (300, 500)$ , e número  $I_{new} = 150$  de novas bateladas com valores de  $\Phi$  com  $\phi_{11} \neq -0.3$  ou  $\phi_{12} \neq 0.4$ . Cada cenário foi repetido 100 vezes. Seguindo a classificação proposta na Tabela 4.1, as novas bateladas em que  $\log(FB) > 0.5$  são classificadas como *fora de controle*. Todas as simulações foram realizadas no software R (R Core Team (2022)). As séries temporais bivariadas seguindo um VAR(1) foram geradas através da função *simoff* do pacote *dvqcc* do R.

Para ilustrar o comportamento das séries temporais simuladas, bateladas de tamanho  $T = 1000$  foram geradas. Duas versões de descontrole foram utilizados, primeiramente perturbando apenas o valor de  $\phi_{11}$  e posteriormente o valor do parâmetro de  $\phi_{12}$ . As Figuras 5.1 e 5.2 ilustram o perfil das

séries temporais geradas durante uma batelada sob controle (linhas verdes) e fora do controle (linhas rosas). É possível visualizar as perturbações impostas nas séries, alterando a estrutura de correlação e correlação cruzada.



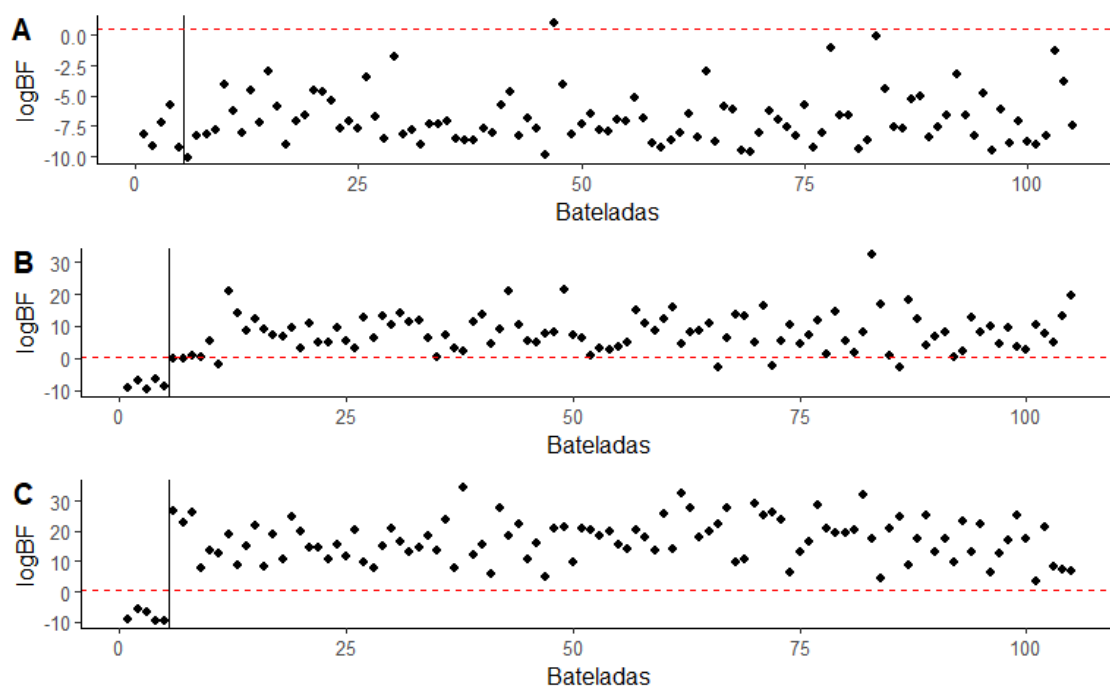
**Figura 5.1:** Séries temporais de  $x_1$  e  $x_2$  para processo sob controle (em verde) e fora de controle (em rosa) em  $\phi_{11}$



**Figura 5.2:** Séries temporais de  $x_1$  e  $x_2$  para processo sob controle (em verde) e fora de controle (em rosa) em  $\phi_{12}$

A Figura 5.3 ilustra graficamente a regra de controle baseada no Fator de Bayes diante de cenários específicos. Três cartas de controle considerando  $I = 5$  bateladas de referência,  $I_{new} = 100$  novas bateladas,  $T = 200$ , e variando as informações consideradas para a geração das novas bateladas. A linha pontilhada em vermelho representa o limite de controle segundo metodologia proposta ( $\log(FB) = 0.5$ ). Os 5 primeiros pontos representam os escores  $\log(FB)$  das bateladas de referência. Após a linha vertical estão os escores  $\log(FB)$  das novas bateladas sob investigação.

Na figura A as novas bateladas são geradas em um processo sob controle, isto é, através de um VAR(1) com  $\Phi$  descrito em (5.3). A figura B apresenta as novas bateladas com descontrole em  $\phi_{11}$  ( $\phi_{11}=0.1$ ) e a figura C apresenta as novas bateladas com descontrole em  $\phi_{12}$  ( $\phi_{12}=0$ ).



**Figura 5.3:** Cartas de controle de processos em batelada com variações nos valores de  $\phi$  para as novas bateladas. **A:**  $\phi_{11} = -0.3$  e  $\phi_{12} = 0.4$ ; **B:**  $\phi_{11} = 0.1$  e  $\phi_{12} = 0.4$ ; **C:**  $\phi_{11} = -0.3$  e  $\phi_{12} = 0.0$

No gráfico A, quase todos os escores estão dentro do limite de controle, como se espera. Importante ressaltar que, como o limite de controle não é estabelecido em função do quantil da calda de uma distribuição de probabilidade, não temos probabilidade teórica de referência para avaliar o número relativo de pontos mal classificados, isto é, o percentual empírico de alarme falso. Note-se, neste cenário um percentual relativamente baixo (1 ponto em 200). Nas análises das simulações nos parágrafos subsequentes voltaremos a discutir sobre esta questão. Na figura B, quase todos os descontroles foram corretamente classificados, enquanto que na figura C todos foram corretamente classificados.

### 5.3 Apresentação dos resultados

As Tabelas 5.2 e 5.3 apresentam os resultados obtidos através das simulações configuradas na Tabela 5.1, com a variação no parâmetro  $\phi_{11}$  e  $\phi_{12}$ , respectivamente. Em ambas as tabelas a linha destacada em cinza ilustra o cenário com as bateladas geradas num processo sob controle. As demais linhas representam perturbações nas séries temporais geradas. Para cada cenário são apresentados o percentual médio  $\hat{\mu}$  e desvio padrão  $\hat{\sigma}$  estimados e o *Average Run Length* ( $ARL(\hat{\mu}) = 1/\hat{\mu}$ ), que representa o número médio de bateladas até que  $\log(FB) > 0.5$ .

**Tabela 5.2:** Percentual médio estimado  $\hat{\mu}$ , desvio padrão  $\hat{\sigma}$  e  $ARL(\hat{\mu})$  de bateladas com  $\log(FB) > 0.5$  para variações em  $\phi_{11}$

$I$	$\phi_{11}$	$T = 300$			$T = 500$		
		$\hat{\mu}$	$\hat{\sigma}$	$ARL(\hat{\mu})$	$\hat{\mu}$	$\hat{\sigma}$	$ARL(\hat{\mu})$
5	-0.30	0.006	0.016	168.539	0.003	0.009	326.087
	-0.25	0.006	0.011	161.290	0.006	0.009	180.723
	-0.20	0.029	0.031	33.937	0.030	0.034	33.482
	-0.10	0.326	0.139	3.066	0.564	0.154	1.772
	0.00	0.867	0.092	1.153	0.988	0.018	1.012
	0.10	0.998	0.005	1.002	1.000	0.000	1.000
	0.30	1.000	0.000	1.000	1.000	0.000	1.000
10	-0.30	0.003	0.010	300.000	0.002	0.005	625.000
	-0.25	0.008	0.013	126.050	0.006	0.010	156.250
	-0.20	0.034	0.043	29.070	0.046	0.044	21.898
	-0.10	0.335	0.112	2.986	0.660	0.101	1.516
	0.00	0.902	0.053	1.108	0.996	0.006	1.004
	0.10	0.998	0.004	1.002	1.000	0.000	1.000
	0.30	1.000	0.000	1.000	1.000	0.000	1.000
30	-0.30	0.002	0.005	576.923	0.002	0.005	576.923
	-0.25	0.007	0.009	147.059	0.006	0.015	172.414
	-0.20	0.032	0.030	31.056	0.047	0.040	21.216
	-0.10	0.358	0.090	2.790	0.686	0.080	1.457
	0.00	0.917	0.040	1.090	0.997	0.004	1.003
	0.10	1.000	0.000	1.000	1.000	0.000	1.000
	0.30	1.000	0.000	1.000	0.997	0.004	1.003

Avaliando primeiramente a Tabela 5.2 quando o processo está sob controle,  $\phi_{11} = -0.30$ , para os diferentes tamanhos de amostra  $I$ , o percentual médio de pontos fora do limite  $\hat{\mu}$  está entre 0.20% e 0.60% (para  $T = 300$ ) e entre 0.20% e 0.30% (para  $T = 500$ ). Isto resulta em  $ARL(\hat{\mu})$  entre 168.539 e 576,923; e entre 326.087 e 625,000, respectivamente. Estas quantidades representam o número médio estimado de bateladas analisadas até um alarme falso, visto que o processo está sob controle (isto é, as bateladas novas permanecem sendo geradas sob um VAR(1) com  $\Phi$  em (5.3)). Comparando com os procedimentos de controle baseados em inferência clássica, onde os limites são definidos em função de quantis obtidos a partir de uma probabilidade teórica de alarme falso  $\alpha = 0.01$  ou  $\alpha = 0.005$ , observa-se que as quantidades  $\hat{\mu}$  na regra proposta estão abaixo, porém próximas a estes valores.

Nas primeiras variações para  $\phi_{11}$ , em que o descontrole é pequeno  $\phi_{11} \in (-0.25, -0.20)$ , através da metodologia proposta é possível identificar parcialmente as bateladas descontroladas, independente do número de bateladas de referência  $I$  e do tamanho das séries  $T$ . Para  $\phi_{11} = -0.20$  são necessárias cerca de 33 bateladas descontroladas nestes nível até que a regra sinalize o descontrole, o que representa um número relativamente alto de amostras.

No momento em que  $\phi_{11}$  assume o valor  $-0.10$ , a abordagem proposta detecta descontrole deste

nível com uma boa frequência. Para  $T = 500$  o descontrole pode ser detectado em média após menos de duas bateladas ( $ARL(\hat{\mu}) = 1.772$ , no máximo). Para descontroles de  $\phi_{11} \in (0, 0.10, 0.30)$  praticamente todas as bateladas são sinalizadas como fora de controle.

Fica evidente o efeito do tamanho da série  $T$  no poder da regra de classificação proposta, principalmente para  $\phi_{11} \in (-0.20, -0.10, 0.00, 0.10, 0.30)$ . Entretanto, não verificamos um efeito tão significativo no poder de discriminação em função do número de bateladas de referência  $I$  testadas.

**Tabela 5.3:** Percentual médio estimado  $\hat{\mu}$ , desvio padrão  $\hat{\sigma}$  e  $ARL(\hat{\mu})$  de bateladas com  $\log(FB) > 0.5$  para variações em  $\phi_{12}$

$I$	$\phi_{12}$	T = 300			T = 500		
		$\hat{\mu}$	$\hat{\sigma}$	$ARL(\hat{\mu})$	$\hat{\mu}$	$\hat{\sigma}$	$ARL(\hat{\mu})$
5	0.60	0.544	0.165	1.838	0.835	0.108	1.198
	0.40	0.004	0.008	263.158	0.003	0.009	375.000
	0.35	0.008	0.015	127.119	0.009	0.017	113.636
	0.30	0.035	0.030	28.517	0.066	0.066	15.106
	0.20	0.440	0.139	2.270	0.789	0.117	1.268
	0.00	1.000	0.001	1.000	1.000	0.000	1.000
	-0.20	1.000	0.000	1.000	1.000	0.000	1.000
10	0.60	0.566	0.125	1.767	0.876	0.060	1.142
	0.40	0.003	0.008	288.462	0.004	0.009	250.000
	0.35	0.007	0.010	147.059	0.006	0.009	170.455
	0.30	0.045	0.035	21.994	0.083	0.080	12.107
	0.20	0.535	0.127	1.870	0.817	0.084	1.223
	0.00	1.000	0.001	1.000	1.000	0.000	1.000
	-0.20	1.000	0.000	1.000	1.000	0.000	1.000
30	0.60	0.627	0.083	1.594	0.905	0.037	1.105
	0.40	0.004	0.008	267.857	0.001	0.003	1000.000
	0.35	0.008	0.011	120.968	0.007	0.009	151.515
	0.30	0.049	0.037	20.325	0.079	0.045	12.723
	0.20	0.543	0.077	1.842	0.864	0.051	1.158
	0.00	1.000	0.000	1.000	1.000	0.000	1.000
	-0.20	1.000	0.000	1.000	1.000	0.000	1.000

A tabela 5.3 ilustra a variação no parâmetro  $\phi_{12}$ . No processo sob controle,  $\phi_{12} = 0.40$ , para os diferentes tamanhos de amostra  $I$ , observamos o percentual médio de pontos fora do limite  $\hat{\mu}$  entre 0.30% e 0.40% (para  $T = 300$ ) e entre 0.10% e 0.30% (para  $T = 500$ ). O número estimado de bateladas até o alarme falso neste caso é de  $ARL(\hat{\mu})$  entre 263.158 e 288,462; e entre 375,000 e 1000,000, respectivamente. Novamente, comparando com os procedimentos baseados em inferência clássica que utilizam probabilidade de alarme falso  $\alpha = 0,01$  ou  $\alpha = 0,005$ , as quantidades  $\hat{\mu}$  estão abaixo mas relativamente próximos a estes níveis.

Assim, como na Tabela 5.2 na vizinhança mais próxima de  $\phi_{12}$  sob controle, isto é, para  $\phi_{12} \in (0.35, 0.30)$ , a regra proposta não identifica boa parte das bateladas perturbadas nestes níveis. O melhor cenário está no  $ARL(\hat{\mu}) = 15.106$  (cerca de 15 bateladas em média até a sinalização da regra). Descontroles de  $\phi_{12} \in (0.20, 0.60)$  são detectados com boa frequência pela regra, com  $1.105 < ARL(\hat{\mu}) < 2.270$ , ou seja, menos do que 3 amostras em média até a sinalização, considerando o pior cenário. A regra de classificação detecta todas as bateladas perturbadas nos níveis  $\phi_{12} \in (0.00, -0.20)$ .

A Tabela 5.3 também evidencia o efeito do tamanho da série  $T$  no poder da regra de classificação proposta. Além disso, podemos aqui notar também um pequeno efeito no número  $I$  de bateladas de referência.

De uma maneira bastante geral, nota-se no estudo apresentado nas duas Tabelas que a medida que aumenta-se o número de instantes de tempo a serem avaliados durante a ocorrência da batelada, bem como o número de bateladas de referência, a tendência é de detecção mais rápida dos descontroles.

A Tabela 5.4 exhibe resultados das simulações conduzidas por Marcondes and Valk (2020), para os descontroles realizados de  $\phi_{11}$  bem como para  $\phi_{12}$ . A simulação apresentada aqui neste trabalho está baseada na dos autores citados, isto é, as leis físicas do processo sob controle são emuladas via um VAR(1) sem intercepto com coeficientes dados em (4.29). Apresentamos aqui os cenários semelhantes aos do nosso estudo para fins de comparação dos resultados. Para determinar o comportamento de cada um dos cenários utilizados foi apresentado a média ( $ARL(\hat{\mu})$ ) e o desvio padrão ( $ARL(\hat{\sigma})$ ) dos valores de ARL, tanto para a estatística de Hotelling ( $\mathcal{T}_\phi^2$ ), expressão (5.1), como da Variância Generalizada ( $\mathcal{W}_\phi$ ), expressão (5.2). A simulações foram realizadas todas com  $I = 1000$  bateladas de referências de tamanho  $T = 300$ . Os valores que estão destacados em cinza são os valores das bateladas sob controle.

**Tabela 5.4:** Média  $ARL(\hat{\mu})$  e desvio padrão  $ARL(\hat{\sigma})$  do ARL nas variações de  $\phi_{11}$  e  $\phi_{12}$  para os gráficos de controle de  $\mathcal{T}_\phi^2$  e  $\mathcal{W}_\phi$

$\phi_{11}$	$\mathcal{T}_\phi^2$		$\mathcal{W}_\phi$		$\phi_{12}$	$\mathcal{T}_\phi^2$		$\mathcal{W}_\phi$	
	$ARL(\hat{\mu})$	$ARL(\hat{\sigma})$	$ARL(\hat{\mu})$	$ARL(\hat{\sigma})$		$ARL(\hat{\mu})$	$ARL(\hat{\sigma})$	$ARL(\hat{\mu})$	$ARL(\hat{\sigma})$
-0.30	101.93	38.36	99.54	30.13	0.65	1.01	0.00	2.30	0.33
-0.25	32.80	6.21	84.90	56.58	0.40	104.62	43.50	115.69	81.73
-0.20	7.26	0.68	48.27	21.21	0.35	24.56	4.41	91.35	58.36
0.00	1.00	0.00	5.01	0.91	0.30	4.97	0.38	51.27	21.85
0.10	1.00	0.00	2.31	0.28	0.05	1.00	0.00	1.97	0.29

Importante ressaltar que na abordagem de Marcondes and Valk (2020) os valores  $ARL(\hat{\mu})$  observados das bateladas sob controle estão próximos ao valor nominal  $\alpha = 0.01$  (o que resulta em  $ARL(\mu) = 100$ ) ao qual as cartas de controle foram planejadas. Comparando com os valores das Tabelas 5.2 e 5.3, reforçamos que o percentual de alarme falso da regra proposta está relativamente próximo (estando abaixo) de  $\hat{\mu} = 0.01$ . Em outras palavras, dado o processo sob controle, nosso procedimento leva mais amostras pra sinalizar um alarme falso, o que é bom sob esta hipótese; entretanto, torna em tese a nossa regra de uma maneira geral menos sensível para detectar descontroles nos parâmetros  $\phi$ . É a mesma ideia da relação dos erros do tipo I (probabilidade de alarme falso  $\alpha$ ) e do tipo II (probabilidade de não detecção de determinado descontrole) no contexto de teste de hipóteses, isto é, quanto menor o  $\alpha$  utilizado para obter os limites de controle, menor a probabilidade de alarme falso e menor o poder de detecção, e vice versa.

Observando os níveis de descontrole  $\phi_{11} \in (-0.25, -0.20)$  e  $\phi_{12} \in (0.35, 0.30)$ , os quais representam alterações bastante sutis nos parâmetros, espera-se que as cartas de controle propostas por Marcondes and Valk (2020), onde os coeficientes foram estimados a partir de  $I = 1000$  bateladas de referência de tamanho  $T = 300$  cada, apresentem desempenho superior nesta vizinhança. Vale

ressaltar que o maior número de bateladas de referência disponíveis é de  $I = 30$  na regra proposta. Para  $\phi_{11} = -0.20$  a menor diferença é  $ARL(\hat{\mu}) = 7.26$  contra  $ARL(\hat{\mu}) = 21.22$  da regra proposta. Em  $\phi_{11} = 0.30$  a menor diferença é  $ARL(\hat{\mu}) = 4.97$  contra  $ARL(\hat{\mu}) = 12.72$  da nossa regra. Para os demais descontroles em  $\phi_{11}$  e  $\phi_{12}$  os resultados são bastante semelhantes com a detecção em quase 100% dos casos com valores  $ARL(\hat{\mu}) \approx 1$ .

Para realizar uma comparação ainda mais abrangente entre as metodologias apresentadas, seria importante executar uma avaliação considerando mais cenários idênticos. Entretanto, os resultados apresentados neste capítulo mostram que a regra proposta baseada no Fator de Bayes aparenta ser bastante promissora. Obtivemos desempenho  $ARL$  bastante interessante mesmo para o número pequeno de  $I = 5$  bateladas de referência. Em outras palavras, construímos uma regra com um bom poder discriminatório a partir de apenas 5 bateladas. Esta configuração nem mesmo é possível na metodologia de [Marcondes and Valk \(2020\)](#). A implementação da abordagem destes autores está no pacote *dvqcc* do R. A função do pacote que gera as cartas de controle necessita de um argumento mínimo de  $I = 30$  bateladas de referência para gerar a análise.

---

## CAPÍTULO 6

# CONCLUSÕES E TRABALHOS FUTUROS

---

Esta dissertação apresentou uma proposta para monitoramento de processos em bateladas. Através da integração do VAR Bayesiano (utilizando *a priori* Minnesota) com os princípios do Fator de Bayes, foi possível criar uma regra de decisão para classificar novas bateladas. Tal abordagem não está limitada a um número suficiente de bateladas disponíveis do processo em controle estatístico.

Nas simulações realizadas assumindo dados temporais gerados através de uma VAR(1) e perturbações nos parâmetros  $\phi_{11}$  e  $\phi_{12}$ , a regra proposta apresentou um bom desempenho de uma maneira geral, identificando com uma boa precisão os descontroles impostos, diante de um número restrito de bateladas de referência ( $I \leq 30$ ). Mesmo para perturbações numa vizinhança próxima aos valores do processo em controle, apesar de menor poder em relação a abordagem clássica de [Marcondes and Valk \(2020\)](#), a qual utiliza  $I = 1000$  bateladas de referência e todos os cenários simulados, a regra apresentou um poder de detecção interessante. Os resultados obtidos através das simulações evidenciam um potencial para aplicabilidade da metodologia desenvolvida.

Para trabalhos futuros, necessita-se da ampliação do estudo para um número maior de cenários de simulação a fim de confirmar e validar a metodologia proposta. Além disso, a metodologia necessita ser testada diante de dados reais de processos em batelada, o que permitirá analisar a viabilidade da abordagem no mundo real. Esperamos que estas extensões tenham uma contribuição relevante na literatura de CEP para processos em batelada.



---

## REFERÊNCIAS BIBLIOGRÁFICAS

---

- Brockwell, P.J., Davis, R.A., Brockwell, P.J., Davis, R.A., 1991. Stationary time series. *Time Series: Theory and Methods* , 1–41.
- Camacho, J., Picó, J., 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, 59–68. doi:<https://doi.org/10.1016/j.aca.2009.02.001>. papers presented at the 11th International Conference on Chemometrics in Analytical Chemistry.
- Choi, S.W., Morris, J., Lee, I.B., 2008. Dynamic model-based batch process monitoring. *Chemical Engineering Science* 63, 622–636. doi:<https://doi.org/10.1016/j.ces.2007.09.046>.
- Doan, T., Litterman, R., Sims, C., 1984. Forecasting and conditional projection using realistic prior distributions. *Econometric Reviews* 3, 1–100. doi:[10.1080/07474938408800053](https://doi.org/10.1080/07474938408800053).
- Ge, Z., Song, Z., Gao, F., 2013. Review of recent research on data-based process monitoring. *Industrial & Engineering Chemistry Research* 52, 3543–3562.
- Girshick, M., Rubin, H., 1952. A bayes approach to a quality control model. *The Annals of Mathematical Statistics* 23, 114–125. doi:[10.1214/aoms/1177729489](https://doi.org/10.1214/aoms/1177729489).
- Hotelling, H., 1947. Multivariate quality control, illustrated by the air testing of sample bombsights. *Techniques of Statistical Analysis* , 111–184.
- Jackson, J.E., 1991. *A user's guide to principal components*. Wiley series in probability and mathematical statistics, Wiley, New York.
- Jackson, J.E., Mudholkar, G.S., 1979. Control procedures for residuals associated with principal component analysis. *Technometrics* 21, 341–349.
- Jeffreys, H., 1939. *Theory of Probability*. First ed., The Clarendon Press, Oxford.
- Jeffreys, H., 1961. *Theory of Probability*. Third ed., Oxford, Oxford.
- Kourti, T., 2003. Multivariate dynamic data modeling for analysis and statistical process control of batch processes, start-ups and grade transitions. *Journal of Chemometrics* 17, 93–109. doi:<https://doi.org/10.1002/cem.778>.
- Litterman, R.B., 1986. Forecasting with bayesian vector autoregressions: Five years of experience. *Journal of Business Economic Statistics* 4, 25–38.
- Lütkepohl, H., 2005. *The New Introduction to Multiple Time Series Analysis*. Springer Science & Business Media. doi:[10.1007/978-3-540-27752-1](https://doi.org/10.1007/978-3-540-27752-1).

- Marcondes, D., Valk, M., 2020. Dynamic var model-based control charts for batch process monitoring. *European Journal of Operational Research* 285, 296 – 305. doi:<https://doi.org/10.1016/j.ejor.2019.12.038>.
- Montgomery, D.C., 2020. *Introduction to Statistical Quality Control*. John Wiley Sons.
- Nomikos, P., MacGregor, J.F., 1994. Monitoring of batch processes using multi-way principal components analysis. *American Institute of Chemical Engineers* 40, 1361–1375.
- Nomikos, P., MacGregor, J.F., 1995a. Multi-way partial least squares in monitoring batch processes. *Chemometrics and Intelligent Laboratory Systems* 30, 97–108. doi:[https://doi.org/10.1016/0169-7439\(95\)00043-7](https://doi.org/10.1016/0169-7439(95)00043-7). inCINC '94 Selected papers from the First International Chemometrics Internet Conference.
- Nomikos, P., MacGregor, J.F., 1995b. Multivariate spc charts for monitoring batch processes. *Technometrics* 37, 41–59.
- Permin, E., Voigtmann, C., Schmitt, R., 2018. A bayesian approach to process model evaluation in short run spc. *International Journal of Engineering Innovation Research* 7, 92–97.
- R Core Team, 2022. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. URL: <https://www.R-project.org/>.
- Ramos, A.W., 2000. *CEP para processos contínuos e em bateladas*. Editora Blücher.
- Ramos, M., Ascencio, J., Hinojosa, M.V., Vera, F., Ruiz, O., Jimenez-Feijoó, M.I., Galindo, P., 2021. Multivariate statistical process control methods for batch production: a review focused on applications. *Production & Manufacturing Research* 9, 33–55.
- Shewhart, W.A., 1931. *Economic Control of Quality of Manufactured Product*. D. Van Nostrand Company.
- Sims, C.A., 1980. Macroeconomics and reality. *Econometrica* 48, 1–48.
- Tolentino, N.M.d.C., 2015. *Processos químicos industriais: Matérias-primas, técnicas de produção e métodos de controle de corrosão*. Saraiva Educação SA.
- Torminato, S.M., 2004. *Análise da utilização da ferramenta CEP: um estudo de caso na manufatura de autopeças*. Ph.D. thesis. Universidade Estadual de Campinas.
- Wang, Y., Sun, J., Lou, T.s., Wang, L., 2017. Stability monitoring of batch processes with iterative learning control. *Advances in Mathematical Physics* 2017, 1–7. doi:[10.1155/2017/5912651](https://doi.org/10.1155/2017/5912651).
- Wasserman, G.S., Sudjianto, A., 1993. Short run spc based upon the second order dynamic linear model for trend detection. *Communications in Statistics - Simulation and Computation* 22, 1011–1036. doi:[10.1080/03610919308813140](https://doi.org/10.1080/03610919308813140).
- Westerhuis, J.A., Gurden, S.P., Smilde, A.K., 2000. Generalized contribution plots in multivariate statistical process monitoring. *Chemometrics and Intelligent Laboratory Systems* 51, 95–114. doi:[https://doi.org/10.1016/S0169-7439\(00\)00062-9](https://doi.org/10.1016/S0169-7439(00)00062-9).
- Wind, D., 2015. *Technical background document for bvar models used at cpb*. CPB Netherlands Bureau for economic policy analysis .