



UFBA

UNIVERSIDADE FEDERAL DA BAHIA
ESCOLA POLITÉCNICA
PROGRAMA DE PÓS GRADUAÇÃO EM
ENGENHARIA INDUSTRIAL - PEI

MESTRADO EM ENGENHARIA INDUSTRIAL

ANA ROSA CARIBÉ DE GÓES MASSA

ANALISADOR VIRTUAL PARA A DETERMINAÇÃO DO TEOR DOS
CONTAMINANTES MAPD EM UM REATOR TRICKLE BED



SALVADOR
2017



UNIVERSIDADE FEDERAL DA BAHIA
ESCOLA POLITÉCNICA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA INDUSTRIAL

ANA ROSA CARIBÉ DE GÓES MASSA

**ANALISADOR VIRTUAL PARA A DETERMINAÇÃO DO TEOR DOS
CONTAMINANTES MAPD EM UM REATOR TRICKLE BED**

Salvador
2017

ANA ROSA CARIBÉ DE GÓES MASSA

**ANALISADOR VIRTUAL PARA A DETERMINAÇÃO DO TEOR DOS
CONTAMINANTES MAPD EM UM REATOR TRICKLE BED**

Dissertação de Mestrado Profissional apresentada ao Programa de Pós-graduação em Engenharia Industrial, da Universidade Federal da Bahia, como parte dos requisitos necessários à obtenção do título de Mestre em Engenharia Industrial.

Orientadora: Karla Patrícia Oliveira Esquerre
Co-orientador: Adonias Magdiel Silva Ferreira

Salvador
2017

Modelo de ficha catalográfica fornecido pelo Sistema Universitário de Bibliotecas da UFBA para ser confeccionada pelo autor

Massa, Ana Rosa Caribé de Góes Massa
Analisador virtual para a determinação do teor dos
contaminantes mapd em um reator tricklebed. / Ana Rosa Caribé
de Góes Massa Massa. -- Salvador, 2017.
71 f. : il

Orientador: Karla Patrícia Oliveira Esquerre Esquerre.
Coorientador: Adonias Magdiel Silva Ferreira Ferreira.
Dissertação (Mestrado - Programa de Pós-graduação em
Engenharia Industrial.) -- Universidade Federal da Bahia,
Escola Politécnica, Universidade Federal da Bahia., 2017.

1. Analisador virtual. 2. Reator tricklebed. I. Esquerre,
Karla Patrícia Oliveira Esquerre. II. Ferreira, Adonias Magdiel
Silva Ferreira. III. Título.

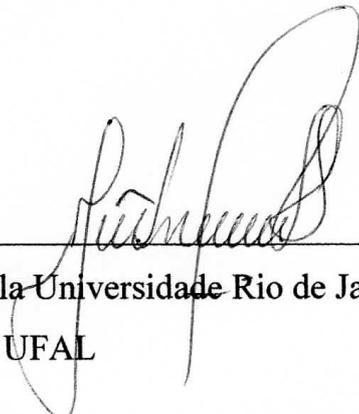
ANALISADOR VIRTUAL PARA A DETERMINAÇÃO DO TEOR DOS CONTAMINANTES MAPD EM UM REATOR TRICKLE-BED

Ana Rosa Caribé de Góes Massa

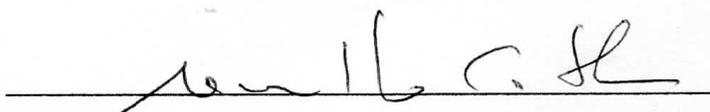
Dissertação submetida ao corpo docente do Programa de Pós Graduação em Engenharia Industrial da Escola Politécnica da Universidade Federal da Bahia – UFBA, como parte dos requisitos necessários para obtenção do grau de mestre em Engenharia Industrial.

Examinada por:

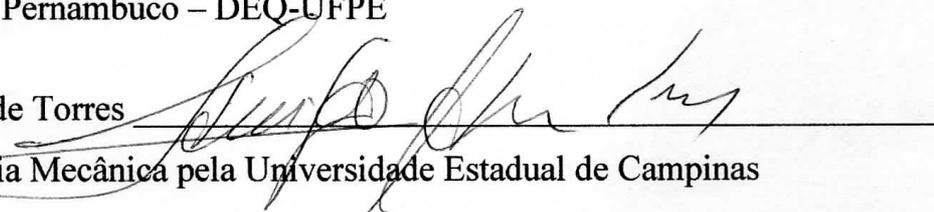
Prof. Dr. João Inácio Soleti


Doutor em Engenharia Química pela Universidade Rio de Janeiro
Universidade Federal de Alagoas - UFAL

Prof. Dr. Neirivaldo Cavalcante da Silva


Doutor em Química pela Universidade Federal de Pernambuco
Universidade Federal de Pernambuco – DEQ-UFPE

Prof. Dr. Ednildo Andrade Torres


Doutorado em Engenharia Mecânica pela Universidade Estadual de Campinas
Universidade Federal da Bahia - UFBA –Escola Politécnica

AGRADECIMENTOS

Primeiramente, a Deus, quem *permitted* que este momento fosse vivido por mim; quero dedicar-Lhe este trabalho. Tenho plena convicção que sem Ele não poderia ter realizado este sonho.

Agradeço de forma especial à minha mãe Áurea e ao meu pai (*In Memoriam*), por não medirem esforços para que eu pudesse levar meus estudos adiante, e me ensinar à importância da paciência e humildade. Aos meus principais incentivadores que sempre acreditam em mim. Vocês me fazem ter certeza de que o amor existe e é possível ser vivido.

Agradeço imensamente ao meu marido Antônio Massa pelo amor incondicional, não medindo esforços para que eu chegasse até esta etapa de minha vida, MEU MUITO OBRIGADO! TE AMO.

Ao presente de Deus na minha vida meu filho Miguel, meu anjo! À minha eterna companheira, amiga Dalva, meu braço direito.

Sou grata à Dra. Karla Esquerre, que me recebeu tão prontamente como sua orientanda, sempre confiando na minha capacidade de realizar esse trabalho. E por fim, por ser uma excelente professora e profissional, a qual me espelho.

Aos líderes Dra. Daniela Domingos e M.^a Márcio Borges, por todo apoio e por acreditarem no meu trabalho e à toda equipe do Físico-Químico e a todos os colegas e amigos do Laboratório pela amizade, paciência, ternura e convivência destes 25 anos, que serão infindáveis. E entre os que merecem especialmente os meus agradecimentos o meu amigo e mentor Dr. Márcio Rebouças. Amo todos vocês e os carrego sempre em nossos corações!

Ao meu co-orientador Dr. Adonias Magdiel por todas as contribuições, pelas discussões e apoio, que foram fundamentais. Agradeço ao Dr. Neirivaldo Cavalcante pela paciência, boa vontade em ajudar e por estar sempre me motivando durante nossas conversas acadêmicas.

Estaria também em falta se não reconhecesse os sacrifícios feitos pelos meus queridos amigos da Universidade Federal da Bahia: Pâmella Mendes, Vicente Barbosa, Adelmo Aguiar e, Luciane Brito por todo esforço em ajudar a resolver problemas e pela crítica certa e construtiva sobre o meu trabalho. Sou grata à Dra. Maria Fernanda Pimentel, pela confiança, incentivo e colaborações.

À Braskem, pelo apoio para realização desse trabalho e tornado possível a realização deste projeto.

*“Seja Deus gracioso para conosco, e nos
abençoe, e faça resplandecer sobre nós o teu
rosto, para que se conheça na terra o teu
caminho.” (Salmo 67:1)*

MASSA, Ana Rosa Caribé de Góes. **Analisador virtual para a determinação do teor dos contaminantes mapd em um reator trickle bed**. 2018. 71 f. il. Dissertação (Mestrado em Engenharia Industrial) – Programa de Pós-graduação em Engenharia Industrial, Escola Politécnica, Universidade Federal da Bahia, Salvador, 2018.

RESUMO

Analisadores em linha fornecem uma resposta rápida de composição em comparação às análises laboratoriais. Porém, esses estão sujeitos a frequentes interferências e contaminações devido às substâncias utilizadas nas correntes reais, que agredem, contaminam e comprometem o funcionamento do equipamento, exigindo manutenções rotineiras. Durante estas, há perda de informações vitais que podem culminar numa parada da produção, a não ser que outro equipamento permita estimar tais informações de maneira confiável. Desta forma, o presente trabalho tem por objetivo desenvolver um analisador virtual para estimar a concentração dos contaminantes metilacetileno e propadieno (MAPD) em um reator *trickle bed* em uma indústria de propileno no Brasil. A partir da coleta de dados de processos de uma campanha catalítica, coletados por cromatógrafos a gás, termopares e medidores de vazão, foram desenvolvidos modelos de calibração multivariada utilizando a técnica de Regressão por Mínimos Quadrados Parciais (PLS), para dois leitos catalíticos, A e B. Dados de treze variáveis de processo, monitoradas a cada 10 minutos durante uma campanha catalítica para cada leito, foram utilizados. Os modelos PLS foram desenvolvidos e validados e foram capazes de fornecer valores preditos confiáveis, com R^2 de 0,84 para o leito A e 0,92 para o leito B. Normalidade e homocedacidade dos resíduos foram observadas em ambos modelos. Além disso, foi realizada uma seleção de variáveis utilizando o gráfico de escores VIP (*Variable Importance in Projection*) obtido durante o desenvolvimento dos modelos PLS. As variáveis mais importantes foram selecionadas e os modelos PLS construídos apenas com essas variáveis mantiveram a capacidade de predição em ambos os leitos, com valores de R^2 de 0,82 para o leito A e 0,87 para o leito B. Normalidade e homocedacidade dos resíduos foram mantidas, e um teste F não gerou evidência que indicasse diferença significativa entre os modelos desenvolvidos antes e após a seleção de variáveis. Dessa forma, os modelos PLS-VIP fornecem uma estimativa confiável do teor de MAPD no reator *trickle bed* na planta de propeno estudada. Esses resultados mostraram que os métodos desenvolvidos possuem um alto potencial de aplicação nos reatores estudados caso haja necessidade, evitando assim uma parada da planta e subsequente perda de capital investido.

Palavras-Chave: Reator *Trickle bed*. Teor de MAPD. Analisadores Virtuais. Quimiometria. Regressão PLS.

MASSA, Ana Rosa Caribé de Góes. **Analisador virtual para a determinação do teor dos contaminantes mapd em um reator trickle bed**. 2018. 71 f. il. Dissertação (Mestrado em Engenharia Industrial) – Programa de Pós-graduação em Engenharia Industrial, Escola Politécnica, Universidade Federal da Bahia, Salvador, 2018.

ABSTRACT

Online analysers grant a faster answer on the composition of products when compared with laboratory analysis. However, the former is often affected by substances in the stream line which harm and compromise its normal working state, calling for frequent maintenance. During those, the loss of vital information could lead to a halt in production, unless another device allows for such information to be carefully estimated. As such, this paper aims at developing a Virtual Analyser that can estimate the concentration of methylacetylene and propadiene (MAPD) contaminants in a trickle bed reactor at a propene industrial plant in Brazil. Process data collected in the reactor by gas chromatographers, temperature probes and flowmeters were employed to build multivariate calibration models by using the Partial Least Square Regression (PLS) technique, for two bed reactors, A and B. Data from thirteen process variables, monitored every 10 minutes during one catalytic campaign for each bed, of about three months each, were used. The developed PLS models for both beds have shown a great prediction capacity and remarkable performances, with R^2 of 0.84 for bed A and 0.92 for bed B. Residual normality and homoscedasticity were observed for both models. In addition, a variable selection approach was carried out using the VIP (Variable Importance in Projection) score plot obtained during the developed PLS models. The most important variables (process variables) were selected and the PLS models built with only these variables were still able to keep a remarkable prediction ability for both beds, with a R^2 of 0.82 for bed A and 0.87 for bed B. Residual normality and homoscedasticity were kept, and an F test did not provide evidence for significant difference between the models developed before and after the variable selection. Therefore, the PLS-VIP models provided a reliable estimate of the MAPD content in the trickle bed reactors at the studied propene plant. These outcomes showed that the developed methods present a high potential for application in the studied reactors, if necessary, in order to prevent a halt in production and its subsequent loss of invested capital.

Keywords: *Trickle bed* Reactor. MAPD content. Virtual Analyzer. Chemometrics. PLS Regression.

LISTA DE ILUSTRAÇÕES

Figura 1	Representação do sistema de hidrogenação catalítica do MAPD	18
Figura 2	Variáveis de processo (PVs)	19
Figura 3	Configurações do reator tipo trickle bed convencional. Correntes concomitantes (a), e contra corrente (b)	22
Figura 4	Estrutura básica de um cromatógrafo a gás. Componentes V1 - válvula reguladora de pressão; V2 - válvula de injeção; V3 - válvula de comutação da coluna; C1 - coluna; C2 - coluna; D detector	22
Figura 5	Representação de um analisador virtual desenvolvido com entradas diretas de medidores físicos, com analisadores virtuais (AV 1 e AV 2)	24
Figura 6	Esquema simplificado da técnica PLS	28
Figura 7	Metodologia utilizada para a construção e avaliação dos modelos	38
Figura 8	Correlação entre as Variáveis de Processo	41
Figura 9	Séries Temporais: Vazões de Carga Fresca, de Reciclo e de Hidrogênio	42
Figura 10	Séries Temporais: Temperaturas no Interior do Leito Catalítico	43
Figura 11	Séries Temporais: Teor de MAPD na Entrada e na Saída	44
Figura 12	Escores das Primeiras Duas Componentes - Leito A	45
Figura 13	Escores das Primeiras Duas Componentes - Leito B	45
Figura 14	Escores da PC1 em Função do Tempo - Leitos A e B	47
Figura 15	Teor de MAPD: Predito vs. Observado, em mg mg Kg ⁻¹ - Leito A	47
Figura 16	Cargas Fatoriais das PVs nas LVs do Modelo PLS - Leito A	48
Figura 17	Resíduos do modelo em Função das PVs-Leito A	49
Figura 18	Teor de MAPD: Predito vs. Observado, em mg Kg ⁻¹ - Leito B	50
Figura 19	Cargas Fatoriais das PVs nas LVs do Modelo PLS - Leito B	51
Figura 20	Resíduos do modelo em Função das PVs- Leito B	51

Figura 21	Variáveis Seleccionadas pelo VIP - Leito A	52
Figura 22	Teor de MAPD: Predito vs. Observado, em mg Kg ⁻¹ - Leito A – VIP	53
Figura 23	Resíduos do modelo em Função das PVs - Leito A – VIP	53
Figura 24	Variáveis Seleccionadas pelo VIP - Leito B	54
Figura 25	Teor de MAPD: Predito vs. Observado, em mg Kg ⁻¹ - Leito B – VIP	55
Figura 26	Resíduos do modelo em Função das PVs - Leito B – VIP	56
Figura 27	RMSEP em função do número de LVs - Leito A	69
Figura 28	RMSEP em função do número de LVs - Leito B	69
Figura 29	Gráfico Integrado de Análise Exploratória Inicial - Leito A	70
Figura 30	Gráfico Integrado de Análise Exploratória Inicial - Leito B	71

LISTA DE TABELAS

Tabela 1	Faixa de composição que alimenta o reator em estudo	17
Tabela 2	Descrição das Variáveis de processo (PVs)	20
Tabela 3	Métodos utilizados para desenvolvimento de analisadores virtuais desenvolvidos em trabalhos publicados nos últimos cinco anos	25
Tabela 4	Aplicação da PLS em processos industriais	36
Tabela 5	Cargas fatoriais das cinco primeiras componentes - Leito B	46
Tabela 6	Figuras de mérito dos modelos antes e depois da seleção de variáveis	56
Tabela 7	p-valores correlações entre PVs - Leito A	68
Tabela 8	p-valores correlações entre PVs - Leito B	68

LISTA DE ABREVIATURAS E SIGLAS

ITC	Intermediate Thermal Detector
JK	Jack- Knife
LV	Variável Latente
MA	Metilacetileno
MAPD	Metilacetileno e Propadieno
MLR	Regressão Linear Múltipla
NNR	Regressão por Redes Neurais
PCA	Análise de Componentes Principais
PCR	Regressão por Componente Principais
PD	Propadieno
PLS	Mínimos Quadrados Parciais
PLSR	Regressão por Mínimos Quadrados Parciais
PV	Variável de Processo
RMSEC	Raiz Quadrada do Erro Médio Quadrático de Calibração
RMSECV	Raiz Quadrada do Erro Médio Quadrático de Cross – Validation
RMSEP	Raiz Quadrada do Erro Médio Quadrático de Previsão
SNV	Standard Normal Variate
SDV	Standard deviation of validation
SPXY	Sample Set Partitioning based on Joint X-Y Distances
SSI	Identificação Subespacial
SV	Sample Valve
TCD	Thermal Conductivity Detector
VIP	Projeção de influência da variância

SUMÁRIO

1	INTRODUÇÃO	14
2	OBJETIVOS	16
2.1	OBJETIVO GERAL	16
2.2	OBJETIVOS ESPECÍFICOS	16
3	FUNDAMENTAÇÃO TEÓRICA	17
3.1	SISTEMA DE HIDROGENAÇÃO DO MAPD	17
3.2	REATOR DO TIPO <i>TRICKLE BED</i>	18
3.3	ANALISADOR EM LINHA	21
3.4	ANALISADORES VIRTUAIS	23
3.5	PRÉ-PROCESSAMENTO DE DADOS	25
3.6	ANÁLISE POR COMPONENTES PRINCIPAIS – PCA	26
3.6.1	Descrição da PCA	26
3.7	REGRESSÃO POR MÍNIMOS QUADRADOS PARCIAIS – PLS	27
3.7.1	Descrição do método PLS	28
3.7.2	Seleção de amostras	29
3.8	AVALIAÇÃO DOS MODELOS DE CALIBRAÇÃO MULTIVARIADA ...	30
3.9	SELEÇÃO DE VARIÁVEIS EMPREGANDO VIP	32
3.10	APLICAÇÃO DA TÉCNICA PLS EM PROCESSOS INDUSTRIAIS	35
4	METODOLOGIA	38
4.1	AQUISIÇÃO DE DADOS	38
4.2	ANÁLISE EXPLORATÓRIA E DETECÇÃO DE OUTLIERS	39
4.3	CONSTRUÇÃO DE MODELOS	39
4.4	AVALIAÇÃO DE MODELOS	40

4.5	SOFTWARE R-STUDIO	40
5	RESULTADOS E DISCUSSÃO	41
5.1	ANÁLISE EXPLORATÓRIA INICIAL	41
5.1.1	Correlação entre as Variáveis de Processo e Séries Temporais	41
5.1.2	Análise de Componentes Principais	44
5.2	MODELO DE REGRESSÃO PLS	47
5.3	MODELOS DE REGRESSÃO PLS-VIP	52
6	CONCLUSÕES	58
7	SUGESTÕES PARA TRABALHOS FUTUROS	60
8	PRODUTOS GERADOS NESTA PESQUISA	61
	REFERÊNCIAS	62
	APÊNDICE A – MATRIZ DOS P-VALORES DAS CORRELAÇÕES ENTRE VARIÁVEIS DE PROCESSO	68
	APÊNDICE B – GRÁFICO – RMSEP – LEITO A E LEITO B	69
	APÊNDICE C – ANÁLISE EXPLORATÓRIA INICIAL – LEITO A ...	70

1 INTRODUÇÃO

A indústria petroquímica é considerada o maior segmento da indústria de transformação brasileira, desde a década de 1960 (SEIDL et al., 2012). Os produtos derivados desse ramo industrial estão presentes em nosso dia-a-dia, como por exemplo: os plásticos, detergentes, solventes, pesticidas, tintas, fibras, cosméticos, borrachas sintéticas, explosivos etc. (GREEN; WITTCOFF, 2003; LAUZON, 2013). Dentre as petroquímicas, destaca-se a brasileira Braskem S.A., a maior produtora de resinas termoplásticas das Américas, pioneira na integração de operações da cadeia produtiva do plástico, onde o propileno é uma das suas principais matérias-primas.

O propileno ou propeno é o segundo produto de maior produção mundial ficando atrás apenas do eteno (ANTUNES, 2007); e um dos produtos do craqueamento da nafta, obtido através da hidrogenação catalítica dos isômeros metilacetileno (MA) e propadieno (PD). Ele é submetido a diversas etapas de destilação, para que compostos leves e pesados sejam retirados, e um produto de alto grau de pureza seja obtido (MOHUNDRO, 2003). As reações de hidrogenação desses isômeros, denominados MAPD, configuram uma etapa crucial do processo de obtenção do propileno, uma vez que esses compostos, mesmo em pequenas quantidades, são contaminantes para os catalisadores das indústrias de polímeros de terceira geração. Caso estejam presentes, retiram a especificação do propileno como produto final para a venda.

A reação de hidrogenação do MAPD é dependente da atividade e da seletividade do catalisador. Neste caso, alguns contaminantes presentes na nafta podem desativar os catalisadores utilizados, sendo prejudiciais à obtenção do propileno. Há relatos na literatura acerca dos efeitos da desativação catalítica em reatores de hidrogenação do MAPD. Wang e Froment (2005) analisaram os efeitos de desativação catalítica em um modelo de reator adiabático. Já Wu e colaboradores (2011) utilizaram dados reais de um sistema de hidrogenação com reatores tipo *trickle bed* para a criação de um modelo cinético incorporando a desativação catalítica, validado por uma simulação dinâmica. Por fim, Wu e Li (2011) realizaram uma otimização multiobjetivo baseada em Fuzzy nesse mesmo sistema, visando encontrar uma razão ótima de operação entre o H₂ e o MAPD em cada reator.

Os reatores de hidrogenação do tipo *trickle bed* são bastante utilizados na reação de hidrogenação catalítica do MAPD. Neste processo, duas correntes, uma líquida e outra gasosa, são inseridas simultaneamente no topo do leito e escoam pela superfície dos

catalisadores onde se processa a reação de hidrogenação. As vantagens da utilização deste tipo de catalisador são: a operação em condições próximas ao *plugflow*, redução do risco de transbordamento; e altas conversões por volume. A modelagem de reatores do tipo *trickle bed* tem sido relatada na literatura; a maioria dos modelos encontrados assume a operação isotérmica, sem vaporização e desconsideram a desativação catalítica (AL-DAHMAN et al., 1997; DUDOKOVIC; LARACHI; MILLS, 1999; GUO; JIANG; AL-DAHMAN, 2008; HUNGER, 2002; KUZELJEVIC; DUDUKOVIC, 2012; SAMIMI et al., 2015).

A capacidade mundial de produção de processos via reatores do tipo *trickle bed* é de aproximadamente 1,6 bilhão de toneladas por ano. O valor dos produtos processados via estes reatores é da ordem média de 300 bilhões de dólares por ano (SIE; KRISHNA, 1998). Tamanha utilização é devido à variedade de reações que ocorrem neste tipo de reator, além das já citadas, em que podem ser conduzidas reações de hidrogenação, oxidação, alquilação e cloração com maior eficiência nos reatores *trickle bed*.

A relevância desse trabalho está na construção de um Analisador Virtual, a partir de dados reais, que consiga prever com relativa precisão a conversão de MAPD no reator. O modelo utilizado na predição do analisador será construído a partir de uma regressão multivariada, especificamente a PLS (Regressão por Mínimos Quadrados Parciais, da sigla em inglês), que tem desfrutado de alta popularidade em diversas áreas de pesquisa, dentre elas a Engenharia Química (MAITRA; YAN, 2008). Isto se deve principalmente pelo fato de tais modelos serem construídos de forma a enfrentar situações onde há alta correlação entre as variáveis preditoras, o que afeta de forma significativa o desempenho de regressões lineares multivariadas. Tal cenário é observado no caso a ser estudado, e fundamenta a escolha do modelo PLS para construção do analisador.

Há diversos relatos na literatura onde o PLS foi utilizado em modelos de predição do teor de substâncias em processos químicos (KAMRUZZAMAN et al., 2012; WU et al., 2011). Não houve, porém, a localização de relatos em reatores *trickle bed*, o que não inviabiliza a relevância deste trabalho, de maneira contrária, o engrandece ao incorporar originalidade ao caso estudado, pioneiro na questão de modelagem empírica de um reator *trickle bed*.

Espera-se que o uso deste analisador viabilize o planejamento de melhores estratégias para a produção e troca dos reatores devido à desativação. Adicionalmente, espera-se que o controle do processo passe a não depender exclusivamente dos analisadores *online*, os quais podem estar sujeitos a interferências e contaminações, implicando na sua desativação e, conseqüentemente, na perda de informações sobre o processo.

2 OBJETIVOS

2.1 OBJETIVO GERAL

Desenvolver um analisador virtual para estimativa das principais variáveis de operação de dois reatores *trickle bed* utilizados no sistema de reação de hidrogenação de MAPD de uma planta de propileno e etileno, presentes em uma grande indústria do Polo Petroquímico de Camaçari.

2.2 OBJETIVOS ESPECÍFICOS

- Observar o comportamento dos dados, tendências de comportamento dos dados, tendências entre as amostras e variáveis;
- Identificar situações atípicas;
- Identificar variáveis que afetam o desempenho do reator;
- Modelar o teor de MAPD no reator *trickle bed*;
- Testar e Validar os modelos construídos.

3 FUNDAMENTAÇÃO TEÓRICA

3.1.SISTEMA DE HIDROGENAÇÃO DO MAPD

A Tabela 1 apresenta a composição típica da corrente de corte C₃ do topo da coluna depropanizadora que alimenta o reator *trickle bed* na planta de etileno.

Tabela 1 – Faixa de composição que alimenta o reator em estudo

Componente	Composição / (%p.p) [†]
Etano + Butadieno	0,25
Propadieno	1,5 a 3,0
Metilacetileno	1,5 a 4,0
Propeno + Propano	90 a 95

[†] porcentagem em peso

Fonte: Elaborado pelo autor (2017).

A reação de hidrogenação de propeno ocorre em um reator heterogêneo, diferenciado dos demais reatores encontrados na indústria. O leito do reator em estudo está dividido em duas partes, sendo que a primeira opera como um reator *trickle bed* (Leito A e Leito B) que é usado como reator primário de forma alternada, e a segunda como um reator de leito fixo (Leito C), podendo ser substituído pelo conversor leito A, ocasionalmente. A Figura 1 representa de forma esquemática o sistema de hidrogenação.

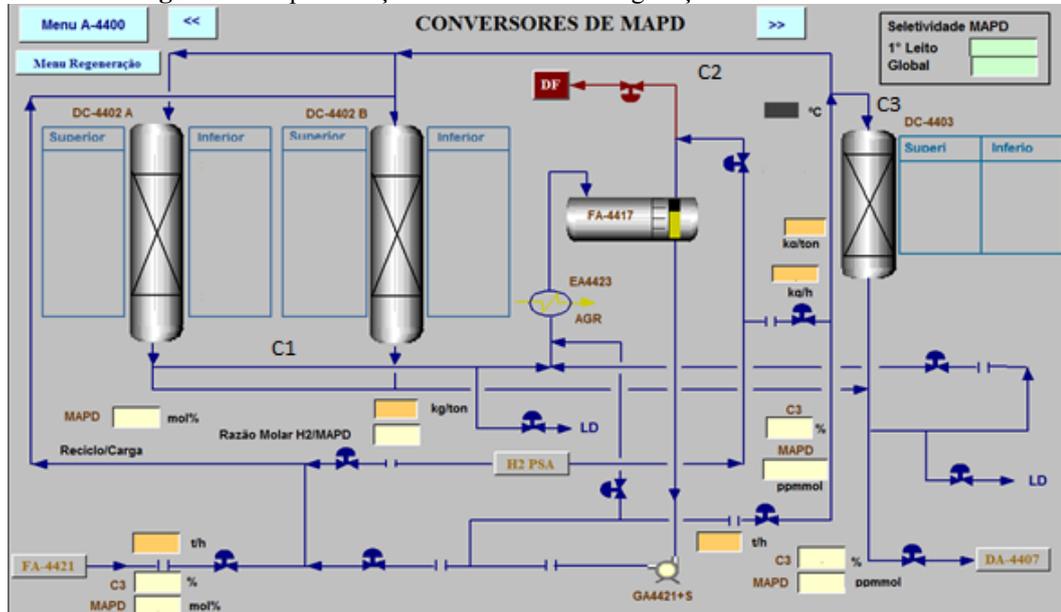
A corrente efluente do conversor primário (C1) é acumulada no vaso de separação, após passagem por um resfriador. Ela é então bombeada e separada em duas correntes (C2 e C3): uma forma o reciclo do conversor primário e a outra serve como carga para o reator secundário, ao qual é adicionada uma segunda injeção de hidrogênio.

O reator catalítico heterogêneo *trickle bed* possui uma carga de propeno que vem com alta concentração de MAPD (metilacetileno e propadieno). Para diminuir a concentração deste contaminante, é necessário a hidrogenação dos metilacetilenos e propadienos e convertê-los a propeno.

Deve-se garantir que na saída do reator a concentração de MAPD seja inferior a 0,35% p.p. para aumentar o tempo de vida do catalisador de paládio, que é limitado pela formação de *greenoil*, e diminuir a quantidade da carga de hidrogênio no reator, pois a hidrogenação é uma

reação exotérmica podendo ocorrer reações de polimerização. Os calores de reação são apresentados nas Equações ((1) e(2).

Figura 1 – Representação do sistema de hidrogenação catalítica do MAPD



Fonte: Elaborado pelo próprio autor (2017).



No reator, onde ocorre a maior parte da conversão, há uma grande liberação de calor. A Equação (1) corresponde à Hidrogenação do Metilacetileno e Equação (2) à Hidrogenação do Propadieno. Mesmo assim, consegue-se manter uma temperatura saída baixa, usando-se uma pressão tal que haja uma vaporização parcial da mistura reacional. As reações de polimerização são muito favorecidas por temperaturas elevadas. Por isso, deve-se operar nas temperaturas mais baixas.

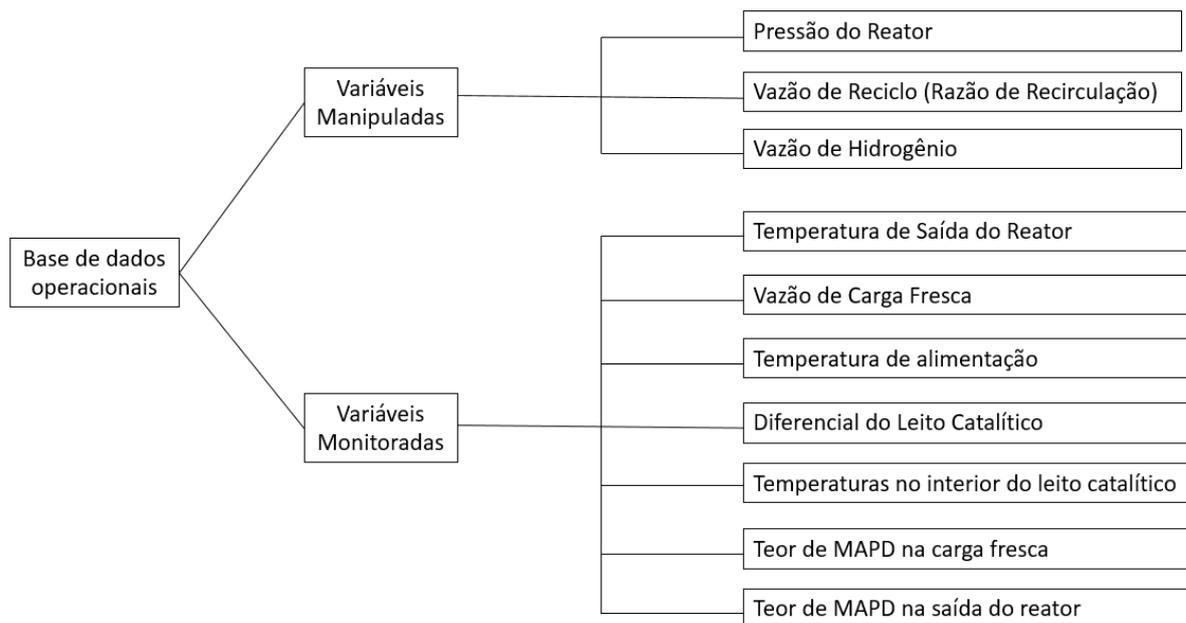
A razão molar H_2 /MAPD é próxima de 1, para que haja uma forte competição entre as reações de hidrogenação do MAPD e do propeno. Assim, converte-se a maior parte do MAPD em propeno, reduzindo a hidrogenação deste último, obtendo-se assim aumento no ganho de propeno.

As variáveis de processo monitoradas e controladas do reator *trickle bed* (descrito na Figura 1 e detalhado na Seção 3.2) estão detalhadas na Figura 2 e na Tabela 2.

3.2 REATOR DO TIPO *TRICKLE BED*

Reatores do tipo *trickle bed* são caracterizados por possuírem as três fases da matéria em seu leito de reação. Por tal característica, eles são utilizados em diversos campos da indústria química, tal como o setor de petróleo, petroquímico, de química fina e de produção bioquímica industrial.

Figura 2 – Variáveis de processo (PVs)



Fonte: Elaborado pelo próprio autor (2017).

Seu grande uso na indústria química e de petróleo, mais especificamente, é devido às vantagens únicas que podem ser obtidas a partir de uso em pressões elevadas e possibilidade de uma maior gama de catalisadores que podem ser usados em seu leito, se comparado com demais reatores que fazem uso de três fases.

A família de reatores *trickle bed* compreende aqueles nos quais a fase vapor e líquida fluem na direção da gravidade passando por um leito sólido de catalisador. A fase gasosa pode ter seu sentido compreendido junto com o líquido, para baixo (Figura 3a), ou pode ter sentido contrário (Figura 3b), dependendo da aplicação em questão. A fase líquida obrigatoriamente deve seguir o sentido da gravidade passando pela fase sólida de catalisador.

A performance geral dos reatores *trickle bed* depende dos fatores do leito catalítico, os quais compreendem: configuração de empacotamento do sólido, porosidade, tamanho da partícula e tamanho da cobertura formada por catalisador. Também influenciam a distribuição da vazão de líquidos e vapores: as condições das partículas do catalisador (envenenamento,

umidificação), a geração de pontos quentes no reator (incluindo catalisador e paredes do reator), as taxas de transferência de calor e massa, além da variação da cinética de reação ao longo do reator (RANADE; CHAUDHARI; GUNJAL, 2011).

Tabela 2 – Descrição das Variáveis de processo (PVs)

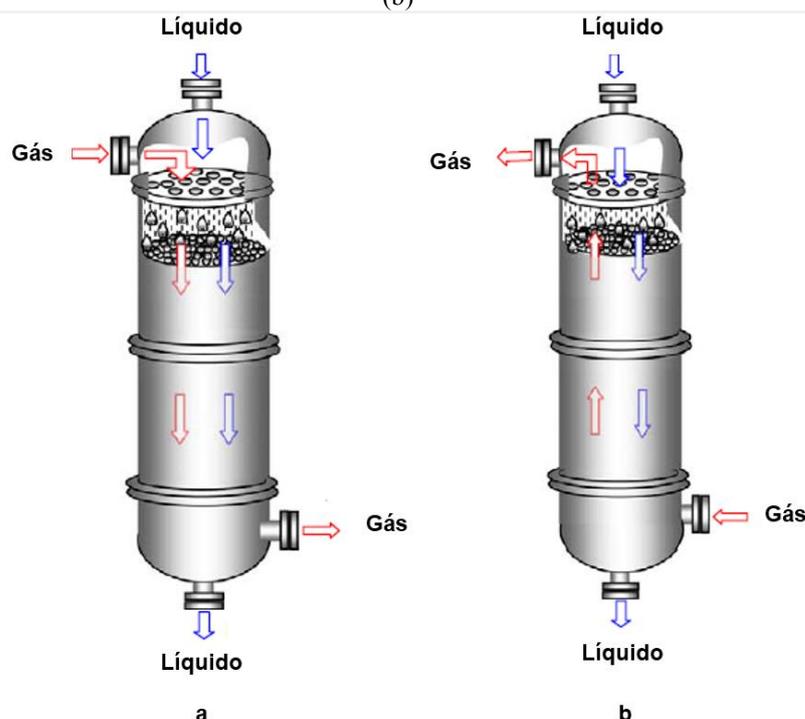
Variável	Símbolo	Descrição
Vazão de Carga Fresca	V1	Carga bruta vinda do processo, que contém a corrente de propileno e contaminantes.
Vazão de Reciclo	V2	Visa garantir simultaneamente que 1) a entrada no reator seja ajustada a um valor de aproximadamente 4,0 % em peso e 2) a velocidade do líquido no reator esteja acima do valor crítico, senão haverá formação de caminhos preferenciais.
Vazão de Hidrogênio Alimentada	V3	Controla a extensão da conversão. O ajuste desta variável deverá ser cuidadoso, a fim de evitar, insuficiente conversão aliado a perdas excessivas de propileno.
Pressão do Leito	V4	Variável real é, teoricamente, a pressão parcial do hidrogênio (ppH_2). Um aumento na ppH_2 resulta em um aumento das taxas de todas as reações e numa baixa seletividade. Isto é, causa uma diminuição no rendimento do propeno.
Temperatura Alimentação	V5	Impacta a cinética das reações envolvidas, porém, para um dado fluxo de alimentação/razão de reciclo, a temperatura de entrada é fixa e não há condição de ser alterada.
Temperaturas no interior do leito catalítico	V6 a V11	Impactam as reações envolvidas, principalmente no que tange a cinética. Esta diz que a taxa de reação aumenta com a temperatura. Tais variáveis são monitoradas em seis diferentes locais ao longo do reator.
Temperatura de Saída	V12	
Teor de MAPD na carga fresca	V13	Teor de contaminante proveniente de seção anterior do processo. Valor máximo permitido de 7% em peso.
Teor de MAPD na saída do reator	Y	Teor de contaminante na saída do reator <i>trickle bed</i> . Medido através de um analisador <i>online</i> (ver seção 3.3).

Fonte: Elaborado pelo autor (2017).

3.3 ANALISADOR EM LINHA

A competitividade na produção dos insumos exige um aparato de tecnologias complexas que garantam a especificação dos produtos em conjunto com um retorno do capital investido. Neste contexto, os analisadores em linha fornecem uma resposta mais rápida da composição dos produtos em comparação às análises laboratoriais. Porém, estes estão sujeitos a frequentes interferências e contaminações devido às substâncias utilizadas nas correntes reais, que agredem, contaminam e comprometem o funcionamento do equipamento, exigindo manutenções mais frequentes, período em que informações importantes deixam de ser coletadas.

Figura 3 – Configurações do reator tipo *trickle bed* convencional. Correntes concomitantes (a), e contra corrente (b)



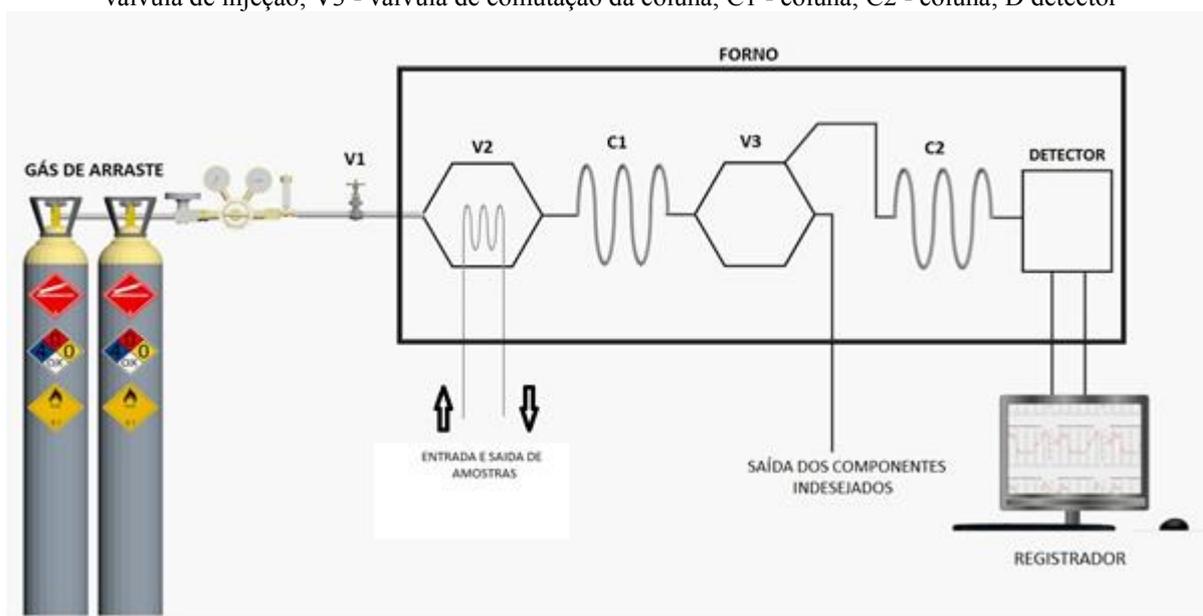
Fonte: Adaptado de Ranade, Chaudhari e Gunjal (2011).

Analisadores em linha são instrumentos utilizados para medições de variáveis do processo. Entre suas principais vantagens estão a disponibilização de resultados em tempo real, o que permite a atuação/intervenção rápida no processo, e a minimização de procedimentos de coletas de amostras em campo. Exemplos de analisadores incluem medidores de pH, condutívimetros e cromatógrafos (KROL; ZABIEGALA; NAMIESNIK, 2010).

Neste contexto, destacam-se os analisadores por cromatografia gasosa, equipamentos que permitem a determinação de vários componentes de uma amostra, desde que suas naturezas sejam conhecidas e programadas previamente. O método de medição consiste em separar e detectar cada componente, relacionando-os com o volume total da amostra. A análise é descontínua, ou seja, se inicia com a injeção da amostra e demora alguns minutos para ser completada. A separação acontece devido às diferentes velocidades com que os componentes da amostra se deslocam pela coluna cromatográfica (COHN, 2006).

Os cromatógrafos a gás são importantes, pois ajudam a controlar e reduzir a variabilidade dos processos (uma vez que representam elementos primários de controle, essenciais à malha de controle do processo), e na obtenção de análise quantitativa de contaminantes, evitando assim liberação de produtos fora de especificações (THOMAS E HAIDER, 2013). Na Figura 4, pode-se observar a estrutura básica de um cromatógrafo a gás.

Figura 4 – Estrutura básica de um cromatógrafo a gás. Componentes V1 - válvula reguladora de pressão; V2 - válvula de injeção; V3 - válvula de comutação da coluna; C1 - coluna; C2 - coluna; D detector



Fonte: Elaborado pelo próprio autor (2017).

O circuito analítico de um cromatógrafo, geralmente, é instalado em um forno e trabalha a uma temperatura controlada. Ele é composto de: válvulas reguladoras de pressão (V1); válvula de injeção da amostra ou *Sample Valve* – SV (V2); colunas (C1, C2); válvula de comutação de colunas ou *Column Valve* – CV (V3). Ele contém ainda um detector térmico intermediário ou *Intermediate Thermal Detector* – ITC (D), o qual permite a visualização das atuações das válvulas e um detector principal, o qual é responsável pela quantificação e visualização do componente.

Um analisador em linha funciona 24 horas por dia, reinjetando a amostra a cada final do ciclo analítico. Todo o sistema é controlado por um microprocessador, que gerencia uma programação automática. A cada injeção, o gás de arraste transporta um determinado volume de amostra, através de válvulas e colunas, até chegar ao detector. Em seguida, um sinal proporcional à concentração dos componentes é gerado e pode ser registrado numa carta gráfica, sob a forma de picos, que é denominada cromatograma. Além disso, os resultados são transmitidos para o painel da operação, para que ajustes possam ser feitos no processo (ANNINO, 1995).

O gás de arraste que conduz a amostra pela coluna cromatográfica, na maioria das vezes, fica armazenado em cilindros, dispostos em série, sob alta pressão. Quando um cilindro acaba, o outro assume automaticamente a função de fornecedor, devido ao diferencial de pressão. O detector por condutividade térmica ou *Thermal Conductivity Detector* – TCD tem seu funcionamento baseado no fato de que a velocidade de perda de calor de um corpo quente para um corpo mais frio é proporcional, dentre outros fatores, à condutividade térmica do gás que separa esses corpos. O TCD geralmente utiliza hélio (He) ou hidrogênio (H₂) como gás de arraste, pois estes gases possuem altíssimas condutividades térmicas, diminuindo as possíveis interferências (AUGUSTO, 2000).

3.4 ANALISADORES VIRTUAIS

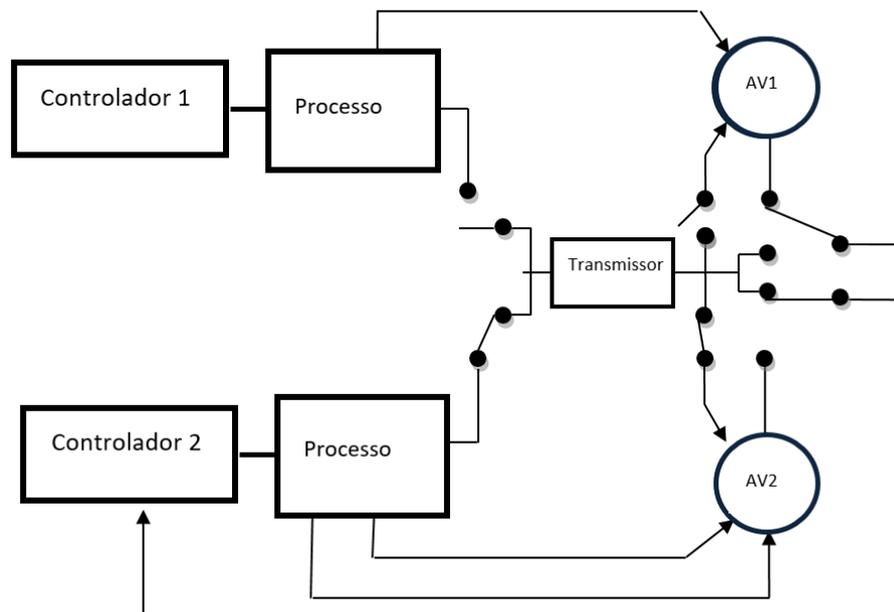
Os analisadores virtuais podem ser definidos como modelos matemáticos desenvolvidos para estimar variáveis de interesse, e podem ser uma alternativa viável aos analisadores *on-line* e às análises laboratoriais, que demandam alto custo devido à aquisição e manutenção dos equipamentos, bem como a utilização de mão-de-obra qualificada, e maior tempo de análise, a depender da técnica empregada (NAGAI; ARRUDA, 2004).

A expressão “analisador virtual” tem origem na substituição de sensores físicos por um *software*. Eles são algoritmos matemáticos desenvolvidos a partir de dados de um processo industrial, capazes de encontrar uma relação matemática entre as variáveis de entrada (dados obtidos através de análises laboratoriais ou provenientes de sensores físicos instalados na planta). São obtidas medidas continuamente, sendo retomadas pelo algoritmo como saídas após serem calculadas através de um modelo matemático determinado de forma empírica ou fenomenológica (LOTUFO; GARCIA, 2008).

Um analisador virtual que utiliza dados de uma planta industrial, medidos por sensores físicos (como medidores de vazão, pressão, temperatura, dentre outros), é representado na Figura 5.

Em um analisador virtual, a modelagem empírica objetiva construir modelos matemáticos de processos dinâmicos a partir de dados experimentais. Estes modelos são de têm grande aplicabilidade, pois apresentam a vantagem de não demandar conhecimento fenomenológico do processo; interessando-se apenas nas relações entre as entradas e saídas do mesmo. Dentre as técnicas de modelagem, empregadas para a construção do analisador virtual estão o MLR (Regressão Linear Múltipla), o PCR (Regressão por Análise de Componentes Principais), o PLS (Regressão por Mínimos Quadrados Parciais), o ICA (Análise Canônica Independente) e a SSI (Identificação Subespacial). A Tabela 3 ilustra a quantidade de trabalhos publicados nos últimos 5 anos, em inglês, e que estão disponíveis na plataforma *Science Direct*, onde analisadores virtuais foram desenvolvidos utilizando os métodos citados.

Figura 5 – Representação de um analisador virtual desenvolvido com entradas diretas de medidores físicos, com analisadores virtuais (AV 1 e AV 2)



Fonte: Lotufo e Garcia (2008).

Dentre os métodos descritos na Tabela 3, o PLS e o PCR se distinguem dos demais devido à capacidade destes de enfrentar situações onde há alta correlação entre variáveis preditoras, o que afeta de forma significativa o desempenho do modelo. Entretanto, o PLS

demonstra melhor desempenho quando há alta correlação entre uma das variáveis preditoras e a variável predita quando comparado a PCR (Maitira & Yan, 2008), o que é observado no caso estudado.

Tabela 3 – Métodos utilizados para desenvolvimento de analisadores virtuais desenvolvidos em trabalhos publicados nos últimos cinco anos

Método	Trabalhos publicados (em inglês)
MLR	31
PCR	346
PLS	74
ICA	45
SSI	35

Fonte: Science Direct (2017).

3.5 PRÉ-PROCESSAMENTO DE DADOS

Ao longo das décadas os estudos, acerca de distintos assuntos da química sejam instrumentais e/ou analíticos, aprimoraram-se e com isso houve um crescimento exponencial das tecnologias modernas de análise (FERREIRA et al, 1999). Essas tecnologias que tem como principal característica a possibilidade de detectar inúmeras variáveis, com suas respectivas unidades, gerando uma vasta quantidade de dados. Assim, devido à quantidade de dados possíveis de ser agrupados, tornou-se necessário o uso de ferramentas de análises para selecionar os dados relevantes, cujo conjunto destas é denominado Quimiometria.

A Quimiometria pode ser definida como uma área destinada à análise de dados de natureza multivariada. Essa área envolve métodos como o ajuste de curva, a otimização e validação de metodologias analíticas e a calibração multivariada. Devido à versatilidade da calibração multivariada, houve um aumento acentuado na sua aplicação a partir da década de 90 (SOUZA; POPPI, 2012).

Uma etapa importante da calibração multivariada se antecede à mesma, sendo o pré-processamento de dados, o qual visa eliminar interferentes e informações superpostas, antes do uso da calibração multivariada (FERREIRA et al., 1999). Especificamente, espera-se uma melhora do conceito de calibração através de manipulações matemáticas comumente utilizadas para reduzir variações aleatórias ou sistemáticas que não estão relacionadas com o problema estudado e que podem acarretar em resultados insatisfatórios ao final da análise em questão. Segundo Beebe, Pell e Seasholtz (1998), essas manipulações podem ser aplicadas

tanto ao conjunto de amostras como às variáveis e são de suma importância para o desempenho dos modelos de calibração multivariada.

No pré-tratamento de dados é possível determinar a melhor estratégia ou método a ser utilizado. Entre os métodos mais comuns, é possível citar o escalonamento, a linearização, a seleção de variáveis, a centralização dos dados na média, derivatização e a *standard normal variate* (SNV).

O escalonamento tem por objetivo fazer com que cada variável se torne unitária e passe a ser expressa em unidades de desvios padrão, colocando de certa forma todas as unidades com a mesma influência em relação à análise como um todo. Já a linearização, utiliza-se de um filtro com média móvel através das transformadas de Fourier ou Wavelet para minimizar o ruído dos dados. A seleção de variáveis gera outra matriz apenas com dados relevantes através da eliminação dos termos sobressalentes. Por fim, a centralização dos dados na média consiste na subtração dos elementos de cada linha da matriz pela média da respectiva coluna em que o elemento se encontra. Uma descrição detalhada destes métodos e aplicações pode ser encontrada em Silva (2017).

3.6 ANÁLISE POR COMPONENTES PRINCIPAIS – PCA

A análise de componentes se preocupa em explicar a estrutura de variância-covariância entre um conjunto de dados através de combinações lineares das variáveis originais. Apesar de p componentes serem necessários para reproduzir a variabilidade total do sistema, onde p representa o número total de variáveis, uma grande porção da variabilidade pode ser explicada pelas k primeiras componentes, denominadas de componentes principais (JOHNSON; WICHERN, 1988).

3.6.1 Descrição da PCA

As componentes principais são combinações lineares específicas das p variáveis originais $X_1, X_2, X_3, \dots, X_p$ de um conjunto de dados. Geometricamente, essas combinações lineares definem um novo sistema de coordenadas, obtido através da rotação do sistema original, onde $X_1, X_2, X_3, \dots, X_p$ são os novos eixos coordenados. Estes representam as direções de máxima variabilidade e fornecem uma descrição mais simples e parcimoniosa da estrutura de covariância (JOHNSON; WICHERN, 2007).

Sendo a matriz de dados iniciais $\mathbf{X} = [X_1 X_2 X_3 \dots X_p]$ com matriz de covariância Σ e autovalores $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$. Considere as combinações lineares:

$$\begin{aligned} Y_1 &= \mathbf{a}'_1 \mathbf{X} = a_{11}X_1 + a_{12}X_2 + \dots + a_{1p}X_p \\ Y_2 &= \mathbf{a}'_2 \mathbf{X} = a_{21}X_1 + a_{22}X_2 + \dots + a_{2p}X_p \\ &\vdots \\ Y_p &= \mathbf{a}'_p \mathbf{X} = a_{p1}X_1 + a_{p2}X_2 + \dots + a_{pp}X_p \end{aligned} \quad (3)$$

Desta forma, temos que:

$$Var(Y_i) = \mathbf{a}'_i \sum \mathbf{a}_i \quad (4)$$

$$Cov(Y_i, Y_k) = \mathbf{a}'_i \sum \mathbf{a}_k \quad (5)$$

As componentes principais são essas combinações lineares ortogonais entre si. Desse modo, a primeira componente principal representa a direção de maior variância no conjunto de dados, a segunda componente principal representa a direção de segunda maior variância ortogonal a primeira componente e assim por diante. Para evitar que $Var(Y_1)$ seja maximizada pela simples multiplicação de a_i por uma constante, os vetores coeficientes são limitados à unidade. Desta forma, tem-se:

- Primeira componente principal = combinação linear $\mathbf{a}'_1 \mathbf{X}$ que maximiza $Var(\mathbf{a}'_1 \mathbf{X})$, sujeito a $\mathbf{a}'_1 \mathbf{a}_1 = 1$;
- Segunda componente principal = combinação linear $\mathbf{a}'_2 \mathbf{X}$ que maximiza $Var(\mathbf{a}'_2 \mathbf{X})$, sujeito a $\mathbf{a}'_2 \mathbf{a}_2 = 1$ e $Cov(\mathbf{a}'_1 \mathbf{X}, \mathbf{a}'_2 \mathbf{X}) = 0$;
- i componente principal = combinação linear $\mathbf{a}'_i \mathbf{X}$ que maximiza $Var(\mathbf{a}'_i \mathbf{X})$, sujeito a $\mathbf{a}'_i \mathbf{a}_i = 1$ e $Cov(\mathbf{a}'_i \mathbf{X}, \mathbf{a}'_k \mathbf{X}) = 0$ para todo $k < i$.

Se a maior parte (90%, por exemplo) da variância total da amostra pode ser explicada pelas primeiras componentes, tidas como “*principais*”, então estas componentes podem substituir as variáveis originais sem perda significativa de informação (JOHNSON; WICHERN, 1998).

3.7 REGRESSÃO POR MÍNIMOS QUADRADOS PARCIAIS - PLS

A regressão por mínimos quadrados parciais, do inglês *Partial Least Squares* (PLS), foi desenvolvida nos anos 60 por Wold para aplicação no ramo da economia (GELADI; KOWLASKI, 1986). Ela consiste em um método linear de calibração multivariada que relaciona uma matriz de dados X com outra matriz Y com o intuito de encontrar fontes de variabilidade em X que são relevantes para explicar Y. Dessa forma, a regressão PLS busca um conjunto de variáveis latentes no processo de decomposição simultânea de X e Y, restringindo-os à explicação da covariância entre X e Y (ABDI, 2010).

Segundo Morellato (2010), as principais vantagens da PLS são: habilidade para modelar regressões com múltiplas respostas, não sendo afetado pela multicolinearidade; e produzir fatores com alto poder de predição, uma vez que tem grandes covariâncias com as variáveis repostas.

A utilização da técnica PLS pode ser interessante na redução da dimensionalidade dos dados e fornece fatores que podem ser usados na construção de cartas de controle (MACGREGOR; KOURTI, 1995).

3.7.1 Descrição do método PLS

Considerando o esquema mostrado na Figura 6, em que X é a matriz de dados de dimensão $(m \times n)$, Y é a matriz de respostas da propriedade de interesse de dimensão $(m \times k)$. Ambas as matrizes são decompostas em uma soma de h variáveis latentes (BRERETON, 2007).

Figura 6 – Esquema simplificado da técnica PLS

$$\begin{array}{c}
 \begin{array}{c} \xrightarrow{n} \\ \boxed{X} \\ \downarrow m \end{array} = \begin{array}{c} \xrightarrow{h} \\ \boxed{T} \\ \downarrow m \end{array} \cdot \begin{array}{c} \xrightarrow{n} \\ \boxed{P^t} \\ \downarrow h \end{array} + \begin{array}{c} \xrightarrow{n} \\ \boxed{E} \\ \downarrow m \end{array} \\
 \\
 \begin{array}{c} \xrightarrow{k} \\ \boxed{Y} \\ \downarrow m \end{array} = \begin{array}{c} \xrightarrow{h} \\ \boxed{U} \\ \downarrow m \end{array} \cdot \begin{array}{c} \xrightarrow{k} \\ \boxed{Q^t} \\ \downarrow h \end{array} + \begin{array}{c} \xrightarrow{k} \\ \boxed{F} \\ \downarrow m \end{array}
 \end{array}$$

Fonte: Brereton (2007).

A decomposição das matrizes gera as Equações (6) e (7), apresentadas abaixo (FILGUEIRAS, 2011):

$$X_{(m \times h)} = T_{(m \times h)} \times P_{(h \times n)}^t + E_{(m \times n)} \quad (6)$$

$$Y_{(m \times k)} = U_{(m \times h)} \times Q_{(h \times k)}^t + F_{(m \times k)} \quad (7)$$

em que:

- T e U - matrizes de escores de X e Y , respectivamente;
- P e Q - matrizes de pesos de X e Y , respectivamente; e
- E e F - resíduos.

A relação linear obtida pelos coeficientes de regressão é a representação da correlação entre os dois blocos \mathbf{X} e \mathbf{Y} , conforme a Equação (8).

$$U_{(m \times h)} = T_{(m \times h)} \times B_{(h \times h)} \quad (8)$$

em que B é definida como a pseudo-inversa da matriz T .

Assim, Y pode ser reescrita utilizando as Equações (7) e (8), originando a Equação(9).

$$Y_{(m \times k)} = T_{(m \times h)} \times B_{(h \times h)} \times Q_{(h \times k)}^t + F_{(m \times k)} \quad (9)$$

De tal modo, que a matriz \hat{Y} para novas amostras (g) pode ser estimada a partir de novos dados, X^* da matriz de previsão na Equação (10).

$$\hat{Y}_{(g \times k)} = X_{(g \times n)}^* \times P_{(n \times h)} \times B_{(h \times h)} \times Q_{(h \times k)}^t \quad (10)$$

3.7.2 Seleção de amostras

A seleção de amostras é uma parte da teoria estatística que define os procedimentos para os planejamentos amostrais e as técnicas utilizadas para o tratamento dos dados. Essas ações devem ser realizadas para que os dados colhidos sejam representativos em relação ao todo. Assim, a seleção de amostras contribui para a construção de modelos de calibração

simples e robustos, além de minimizar erros de predição. Devido à dificuldade de se obter conjuntos representativos ao trabalhar com matrizes complexas, percebe-se a importância da seleção do método adequado, e por isso os algoritmos são utilizados. Dentre os algoritmos destaca-se o Kennard-Stone.

O algoritmo Kennard-Stone em 1969, foi criado visando selecionar amostras de calibração em um agrupamento de dados utilizando para isso a distância entre as amostras para a uniformidade da seleção (KENNARD; STONE, 1969). Por necessitar de aprimoramento, criou-se o algoritmo de partição do conjunto de amostras baseado nas distâncias conjuntas X-y.

O algoritmo inicia a seleção encontrando as duas observações que possuem maior distância geométrica. Dentre as observações restantes, o algoritmo seleciona então a que possui maior separação geométrica da mais próxima observação já selecionada (KENNARD; STONE, 1969). Esse processo é repetido até que o subconjunto de amostras selecionadas possua k observações, que geralmente representa entre 20% e 30% da massa inicial de dados.

3.8 AVALIAÇÃO DOS MODELOS DE CALIBRAÇÃO MULTIVARIADA

Após construção do modelo PLS, parâmetros analíticos de desempenho são utilizados para avaliar a capacidade preditiva do modelo. Esses parâmetros são conhecidos como figuras de mérito, e podem ser considerados como indicadores quantitativos da capacidade preditiva dos modelos multivariados. Tendo em vista a melhor eficiência do modelo, as figuras de mérito devem descrever a melhor situação real levando em conta o maior número possível de variações.

As figuras de mérito utilizadas para a avaliação dos modelos são: **RMSEC**, **RMSECV**, **RMSEP**, **BIAS** e R^2 (NAES et al., 2002), descritas abaixo.

O RMSEC (*Root Mean Square Error of Calibration*) ou raiz quadrada do erro médio quadrático de calibração, fornece informações sobre o ajuste do modelo em relação aos dados, e é calculado através da Equação (11) (VALDERRAMA; BRAGA; POPPI, 2009).

$$RMSEC = \sqrt{\frac{\sum_{i=1}^{N_{cal}} (\hat{y}_i - y_i)^2}{(N_{cal} - A - 1)}} \quad (11)$$

em que:

- \hat{y}_i é o valor predito para valor de referência (y);
- N_{cal} é o número de amostras de calibração e;
- A é o número de variáveis latentes.

Da mesma forma, o ajuste do modelo nas etapas de validação cruzada e previsão é avaliado através do erro quadrático médio da validação cruzada e de previsão, RMSECV e RMSEP, calculados conforme Equações (12) e (13).

$$RMSECV = \sqrt{\frac{\sum_{i=1}^{N_{cal}} (\hat{y}_{CV,i} - y_i)^2}{N_{cal}}} \quad (12)$$

em que:

- $\hat{y}_{CV,i}$ é o valor previsto na etapa de validação cruzada;
- y_i é o valor de referência para a amostra i ;
- N_{cal} é o número de amostras do conjunto.

$$RMSEP = \sqrt{\frac{\sum_{i=1}^{N_{pred}} (\hat{y}_i - y_i)^2}{N_{pred}}} \quad (13)$$

em que:

- \hat{y}_i é o valor predito;
- y_i é o valor de referência para cada amostra i ;
- N_{pred} é o número de amostras do conjunto de predição.

O BIAS (erro sistemático) indica a existência de tendências significativas na estimativa dos valores preditos. O BIAS é calculado pela Equação (14).

$$BIAS = \frac{\sum_{i=1}^{N_{pred}} (\hat{y}_i - y_i)}{N_{pred}} \quad (14)$$

O gráfico dos resíduos para as amostras de calibração e validação pode indicar se os dados seguem um comportamento linear caso a distribuição destes resíduos seja aleatória, observando o valor de R^2 . Já que a linearidade não é uma figura de mérito que melhor descreva a calibração multivariada utilizando PLS, uma vez que as variáveis são decompostas pelos componentes principais.

3.9 SELEÇÃO DE VARIÁVEIS EMPREGANDO VIP

A presença de inúmeras variáveis de processo tem incentivado os pesquisadores na busca por modelos com um número reduzido de variáveis, devido aos custos de coletas de dados, aumento da precisão das informações geradas e maior aplicabilidade dos modelos (MONTGOMERY; RUNGER; HUBELE, 2009). Dentre as abordagens para seleção de variáveis aplicadas no contexto de regressões lineares múltiplas, estudos indicam o método PLS-VIP é o mais adequado para provisões onde há alta colinearidade (CHONGC; JUN, 2005).

Proposto inicialmente por Wold, Sjöström e Eriksson (2001), o gráfico dos escores VIP (*Variable Importance in Projection*), obtidos através do método PLS, dá origem a índices de importância das variáveis de processo, definidos como a soma ponderada dos quadrados dos pesos (*weights*) do modelo PLS Equação (15), sendo um resumo da importância de uma variável em relação aos dados de X e Y (GUYON; ELISSEEFF, 2003).

$$VIP_j = \sqrt{\frac{\sum_{f=1}^F w_{if}^2 \cdot SSY_{total} \cdot J}{SSY_{total} \cdot F}} \quad (15)$$

em que:

- w_{if} são os pesos do modelo PLS;
- SSY_f é a soma dos quadrados da variância explicada por cada componente;
- J é o número de variáveis;
- SSY_{total} é a soma dos quadrados da variância explicada total e;
- F é o número total de componentes.

A inclusão dos pesos na estimativa do cálculo dos VIP permite tanto a descrição das variáveis dependentes como as das variáveis independentes, ambas são parcelas importantes para o modelo. Por exemplo, variáveis com \overline{VIP} menores que 1 são menos importantes e podem ser excluídas do modelo, enquanto que aquelas com valores próximo ou superior a 1 são consideradas importantes. Isto implica na eliminação de variáveis, e escolhas das variáveis que são relevantes em relação à resposta.

Para comparar o desempenho dos diferentes modelos PLS são realizados testes estatísticos de Fischer e t de *Student*. (BRUNS, SCARMINIO e NETO, 2006; INMETRO, 2010). O teste estatístico de Fischer (Teste F) pode ser empregado para testar a hipótese da relação entre as variâncias populacionais, utilizando a Equação (16).

$$\frac{S_A^2}{S_B^2} \cong \frac{\sigma_A^2}{\sigma_B^2} F_{cal} \quad (16)$$

em que, S_A^2 e S_B^2 são as variâncias amostrais σ_A^2 e σ_B^2 as variâncias populacionais.

Em alguns casos, a possibilidade de duas variâncias populacionais serem idênticas pode ser testada, ou seja, casos em que a razão $\frac{\sigma_A^2}{\sigma_B^2}$ é igual a 1. Assim, faz necessário testar a hipótese nula em que as variâncias amostrais S_A^2 e S_B^2 são estimativas da mesma variância populacional. Dessa forma, o teste F pode ser calculado considerando a razão entre duas variâncias a fim obter o valor de F_{cal} , conforme a Equação(17).

$$F_{cal} = \frac{\sigma_A^2}{\sigma_B^2} \quad (17)$$

O $F_{crítico}$ é tabelado, obtido conhecendo graus de liberdade (ν) e o nível de confiança desejado (geralmente 95 % de confiança). O grau de liberdade é dado por $\nu = n - 1$, em que n é o número de elementos da amostra.

Ao comparar o F_{cal} e $F_{crítico}$, caso o F_{cal} seja maior, a hipótese nula deve ser rejeitada.

O teste t de *student* avalia a significância estatística do *bias* dos modelos PLS, de acordo coma a norma ASTM E 1655-05. Assim, o cálculo do t envolve o desvio padrão dos erros de validação (SDV - *standard deviation of validation*). O SDV e o valor de t são calculados pelas Equações (18) e (19).

$$SDV = \sqrt{RMSEP^2 - BIAS^2} \quad (18)$$

$$t_{cal} = \frac{|BIAS\sqrt{N_{pred}}|}{SDV} \quad (19)$$

em que N_{pred} é o número de amostras de previsão do modelo.

O valor do $t_{crítico}$ também é tabelado, obtido através do valor do nível de confiança e dos graus de liberdade. Ao comparar o t_{cal} e $t_{crítico}$, caso o calculado seja maior, indica que o *bias* do modelo PLS é estatisticamente significativo.

É possível também verificar se os valores de *bias* de dois modelos PLS diferentes podem ser estatisticamente iguais ou diferentes. Para isso, usando o teste F já descrito, é certificado se as variâncias dos modelos são significativamente diferentes. Caso os desvios padrões não sejam diferentes estatisticamente, é preciso uma estimativa conjunta do desvio padrão s_c , calculado da através da Equação (20).

$$s_c = \sqrt{\frac{(N_{pred1} - 1)SDV_1^2 + (N_{pred2} - 1)SDV_2^2}{N_{pred1} + N_{pred2} - 2}} \quad (20)$$

Assim, para avaliar se os dois *bias* são significativamente diferentes o valor de t_{cal} é calculado através da Equação (21).

$$t_{calc} = \frac{Bias_1 - Bias_2}{s_c \sqrt{\frac{1}{N_{pred1}} + \frac{1}{N_{pred2}}}} \quad (21)$$

No entanto, se os dois conjuntos de amostras possuírem desvios padrões diferentes, a estimativa conjunta do desvio padrão não é indicada. Assim, o teste t deve ser realizado utilizando as Equações, considerando o t_{calc} ev:

$$t_{calc} = \frac{Bias_1 - Bias_2}{\sqrt{\frac{SDV_1^2}{N_{pred1}} + \frac{SDV_2^2}{N_{pred2}}}} \quad (22)$$

$$v = \frac{\left(\frac{SDV_1^2}{N_{pred1}} + \frac{SDV_2^2}{N_{pred2}} \right)^2}{\frac{\left(\frac{SDV_1^2}{N_{pred1}} \right)^2}{N_{pred1+1}} + \frac{\left(\frac{SDV_2^2}{N_{pred2}} \right)^2}{N_{pred2+1}}} \quad (23)$$

Para efeito de obtenção do valor de t crítico tabelado, considera-se o valor de graus de liberdade arredondado para um número inteiro.

3.10 APLICAÇÃO DA TÉCNICA PLS EM PROCESSOS INDUSTRIAIS

Trabalhos recentemente publicados na literatura demonstram que a regressão por mínimos quadrados parciais tem sido aplicada nos mais diversos campos industriais. Na Tabela 4 – Aplicação da PLS em processos industriais, é apresentado um resumo dos trabalhos levantados e, logo em seguida, uma descrição mais detalhada sobre cada um.

Tabela 4 – Aplicação da PLS em processos industriais

Autor(es) (Ano)	Objetivos	Variáveis Avaliadas	Métodos	Principais Resultados obtidos
Pasquini e Bueno (2007)	Prever o ponto de ebulição (TBP) e estimar a gravidade da API para caracterizar o petróleo processado em refinarias	Espectro infravermelho	PLS	Os modelos construídos usando o algoritmo PLS mostraram desempenho superior ao construído usando ANN, o que deu bons resultados durante a calibração, mas falhou na validação externa
Anzanello, Albin e Chaovalitwoegse (2009)	Reduzir a ocorrência de overfitting no processo de seleção de variáveis	Variáveis reais de um processo de produção de látex.	PLS	O método proposto apresentou significativa redução no percentual de variáveis retidas e na sua variância, e manteve a acurácia de classificação em patamares satisfatórios
Lira <i>et al.</i> (2010)	Prever parâmetros de qualidade de misturas de diesel / biodiesel	Densidade, teor de enxofre e temperaturas de destilação	PLS	Os modelos globais PLS que empregam regiões NIR e MIR podem ser usados para prever temperaturas de destilação e teor de enxofre em misturas de biodiesel / diesel
Reiss, Wojsznis e Wojewodka (2010)	Propor métodos para modelagem PLS e cálculos de previsão on-line para um processo em lote	Variáveis de entrada e saída de cada lote	PLS	Análise PLS é mais informativa e robusta e uma melhor compreensão e percepção do intervalo de confiança por parte do pessoal de operação em lote, bem como uma melhor interpretação e ação tomadas durante a evolução do lote.

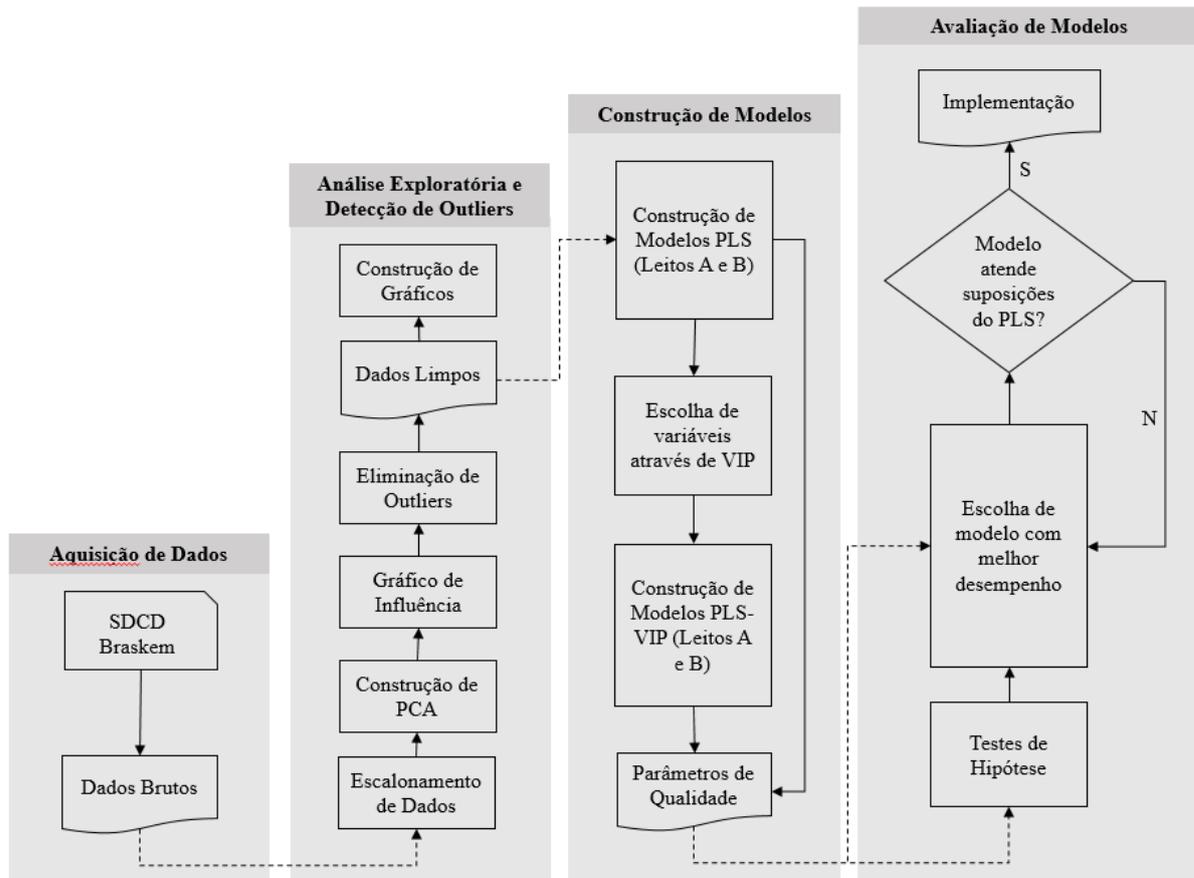
Autor(es) (Ano)	Objetivos	Variáveis Avaliadas	Métodos	Principais Resultados obtidos
Anzanello (2013)	Seleção de variáveis para classificação de bateladas produtivas com base em múltiplos critérios.	Temperaturas, pressões e concentrações de reagentes químicos, enquanto a variável de resposta denota uma característica do produto, como viscosidade ou teor de pureza.	PLS	O método reteve, em média, 12% das variáveis originais. As variáveis selecionadas elevaram a sensibilidade de classificação da porção de teste em 9%, de 0,78 para 0,85, enquanto que a especificidade da mesma porção aumentou 20%, de 0,64 para 0,77.
Afanador <i>et al.</i> (2014)	Formulação de fator de correção de auto-correlação para PLS e aplicação ao método de seleção de variável de Correlação Multivariada Significativa (SMC)	Fluxo, pressão, temperatura e mistura de composições para reações com catalisadores de níquel. E a variável rendimento para produção de vacina	PLS e SMC	Os resultados demonstram que a formulação do fator de correção apresentada neste trabalho tem o efeito desejado de diminuir a taxa de falso positivo quando aplicado ao SMC.
Harrou, Nounoue and Madakyaru (2015)	Combinar as vantagens do EWMA univariado e PLS para aprimorar seus desempenhos e ampliar sua aplicabilidade na prática	Dados da coluna de destilação	PLS	Os resultados da simulação mostram claramente a eficácia do método proposto sobre o PLS convencional, especialmente na presença de falhas com pequenas magnitudes.
Hug <i>et al.</i> (2015)	Identificar compostos que co-variam com a atividade mutagênica de amostras ambientais complexas	Compostos que covariam com atividade mutagênica	PLS	A covariância entre compostos provenientes de águas residuais industriais e atividade mutagênica apoiou a aplicação de EDA "virtual" como ferramenta estatística para separar compostos toxicologicamente relevantes de menos relevantes.

Fonte: Elaborado pelo autor (2017).

4 METODOLOGIA

Na Figura 7 é apresentado um desenho esquemático da metodologia desenvolvida e utilizada nesta pesquisa envolvendo todas as etapas necessárias para a construção modelos e validação dos modelos de predição do MAPD.

Figura 7 – Metodologia utilizada para a construção e avaliação dos modelos



Fonte: Elaborado pelo próprio autor (2017).

4.1 AQUISIÇÃO DE DADOS

Os dados foram adquiridos em duas campanhas catalíticas: uma campanha para o leito A e uma campanha para o leito B. A periodicidade de aquisição das variáveis do processo foram de 10 minutos em 10 minutos, tanto para as variáveis do reator quanto para o analisador em linha. No entanto, os dados relativos à saída do reator e ao analisador possuem um atraso de três e sete minutos, respectivamente, em relação às variáveis de entrada do sistema, em função do tempo de residência do reator e do ciclo do analisador. As variáveis de entrada foram organizadas em uma matriz \mathbf{X} e a variável de saída em um vetor \mathbf{y} . A matriz \mathbf{X} contém

13 colunas, correspondentes às variáveis independentes (vazão de carga fresca, vazão de reciclo, vazão de hidrogênio, temperatura de alimentação combinada, pressão, temperaturas dos seis sensores dentro do reator, temperatura de saída e o teor de MAPD na carga fresca). Foram coletadas observações durante uma campanha catalítica, num período aproximado de quatro meses (11445 / 13540 observações para os leitos A e B, respectivamente). O vetor \mathbf{y} possui os valores de MAPD determinados pelo cromatógrafo *online* durante o processo.

4.2 ANÁLISE EXPLORATÓRIA E DETECÇÃO DE *OUTLIERS*

O analisador de condutividade térmica (TCD), que fornece a resposta de MAPD na saída, possui um limite de detecção para ambos MA e PD que, somados, resultam em 200 mg Kg^{-1} . No início da campanha, quando o catalisador ainda está reativo, a conversão da reação, a depender da vazão de H_2 , pode gerar valores tão baixos de contaminante na saída, que o valor fornecido pelo TCD não possui significância alguma. A única informação de um valor abaixo do limite de detecção é dada por comparação: sabe-se que, naquele momento, o teor de contaminante está abaixo de 200 mg Kg^{-1} . O TCD pode, inclusive, gerar valores negativos, que não possuem nenhum significado real. Desta forma, todas as observações abaixo de 200 mg Kg^{-1} foram desconsideradas no modelo.

Como as variáveis possuem diferentes dimensões, é necessário um escalonamento, que também tem por objetivo fazer com que cada variável possua média nula e desvio unitário. Após o escalonamento, a técnica PCA foi utilizada para realizar uma análise exploratória dos dados do conjunto de dados de ambos os leitos, com o intuito de observar o comportamento dos dados, tendências entre as amostras e variáveis, e identificar possíveis valores aberrantes (*outliers*). A partir de um modelo PCA, obteve-se as estatísticas Q-resíduo e T^2 de Hotelling, as quais representam, respectivamente, o resíduo deixado pelo modelo ao representar determinada amostra e a influência que cada amostra exerce sobre o ajuste do modelo. (RASMUS; AGE, 2014).

Observações com alto valor numérico para ambas estatísticas se distanciam da massa de dados e podem ser considerados *Outliers*. Estas observações podem, portanto, ser removidas do conjunto de dados. Os dados, após a remoção de *outliers*, foram utilizados para construção dos gráficos de séries temporais das variáveis independentes e dependente.

4.3 CONSTRUÇÃO DE MODELOS

Os modelos PLS foram desenvolvidos conforme a seção 3.7 deste trabalho. Para isso, os dados foram divididos em subgrupos de treinamento (80%) e teste (20%). A separação dos blocos de calibração e validação foi realizada de forma que cada um dos dois conjuntos tivesse representações de diferentes condições operacionais utilizadas durante as campanhas, conforme descrito na seção 3.9.

Os modelos PLS construídos foram utilizados para realizar uma seleção das variáveis mais importantes e construção de novos modelos, conforme descrito na seção 3.7. Para avaliar a capacidade preditiva dos modelos, foram utilizados os critérios de qualidade de modelo: ***RMSEC***, ***RMSECV***, ***RMSEP***, ***BIAS*** e R^2 , descritas na seção 3.8.

4.4 AVALIAÇÃO DE MODELOS

Os testes F e t descritos na seção 3.9 foram utilizados para decidir se há diferença significativa entre os modelos PLS e PLS-VIP. Após avaliação dos critérios de qualidade de modelo e capacidade preditiva, o modelo com melhor desempenho foi escolhido para implementação no controle de processo. Suposições inerentes ao modelo PLS foram verificadas no modelo escolhido antes da implementação.

4.5 SOFTWARE R-STUDIO

As análises estatísticas, bem como as análises gráficas, apresentadas neste trabalho foram realizadas por meio do ambiente de desenvolvimento integrado (do inglês, IDE) do RStudio®. O RStudio IDE constitui uma plataforma moderna que integra programação, geração de imagens e importação de dados em um único ambiente. O RStudio IDE opera através da linguagem de programação R, uma linguagem aberta, livre e com destaque na área de análise de dados e modelagem. Por operar através de bibliotecas de funções, muitas desenvolvidas e validadas pela comunidade de programadores de R, a linguagem R possui aplicações em diferentes áreas do conhecimento, desde a estatística clássica, psicologia, biologia até tópicos modernos como o *big data*. Além de contribuir para uma pesquisa com foco na reprodutibilidade dos resultados, o desenvolvimento de pesquisas quantitativas aliadas a uma linguagem de programação torna possível a implementação de soluções não disponíveis em softwares comerciais. Foram empregados, em todas as etapas metodológicas do presente trabalho os pacotes de funções *tidyverse*, *pls* e *mixOmics* (LILAND, 2016; LIQUET, 2017; WICKHAM, 2017).

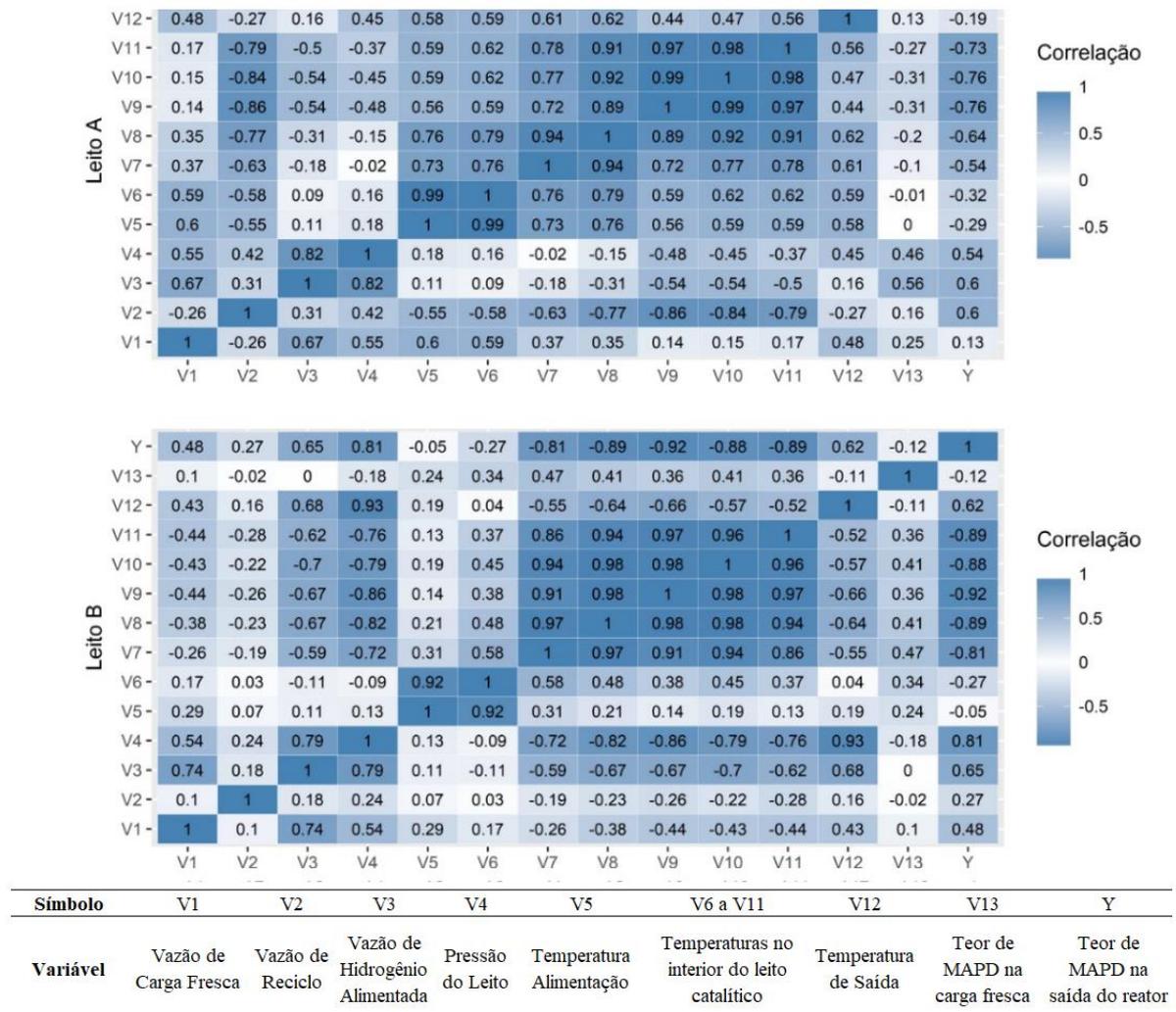
5 RESULTADOS E DISCUSSÃO

5.1 ANÁLISE EXPLORATÓRIA INICIAL

5.1.1 Correlação entre as Variáveis de Processo e Séries Temporais

As correlações entre as variáveis de processo (PVs) descritas na Tabela 2 podem ser observadas na Figura 8. Regiões mais escuras representam valores numéricos de correlação mais elevados em termos absolutos, enquanto que regiões mais claras representam valores mais baixos. Todas as figuras apresentadas nesta seção não incluem a presença de *outliers*.

Figura 8 – Correlação entre as Variáveis de Processo



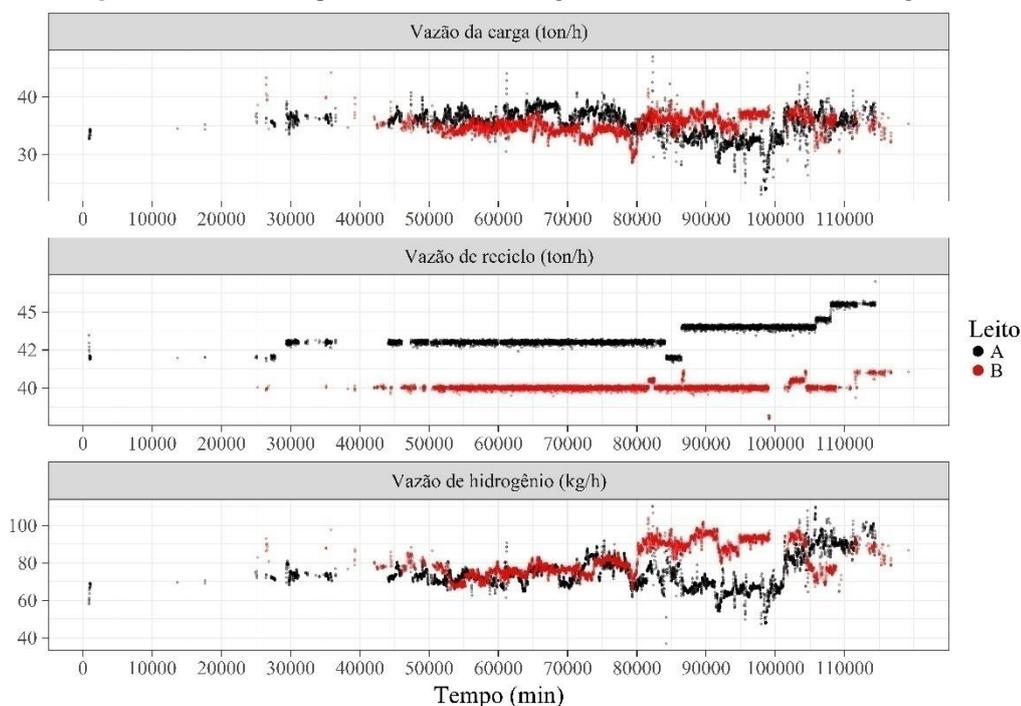
Fonte: Elaborado pelo próprio autor (2017).

A matriz de p-valores das correlações na Figura 8 é apresentada no Apêndice A. A 95% de confiança, todas as correlações são significativas, com exceção das correlações entre V2 e V13, e V3 e V13, em ambos os leitos. Observa-se na Figura 8 que as variáveis de temperatura interior catalítico (V6 a V11) apresentam correlações moderadas/altas entre si, o que é esperado, já que o aumento de temperatura em um ponto do leito influencia o aumento em outro ponto por transporte de calor. Também é possível encontrar altas correlações entre a vazão de hidrogênio (V3), e as temperaturas dos leitos, já que um aumento na vazão de reciclo (V2) torna mais significativa a reação de conversão. Entretanto, é importante notar que sistemas reais e complexos como este estão sujeitos a interferências diversas que podem resultar em correlações espúrias.

As regiões mais escuras entre as variáveis preditoras justificam a aplicação da Análise de Componentes Principais. Entretanto, é a coluna relacionada ao MAPD de saída (Y) que fundamenta a utilização de um modelo PLS e não de uma PCR, como detalhado na seção 3.7.

A Figura 9 apresenta as séries temporais das variáveis com cargas de entrada. É importante observar que os dados de início de campanha estão majoritariamente omitidos por apresentarem teor de MAPD na saída menores que o limite de detecção, sendo desconsiderados na análise dos dados/construção do modelo, conforme descrito na seção 4.1.

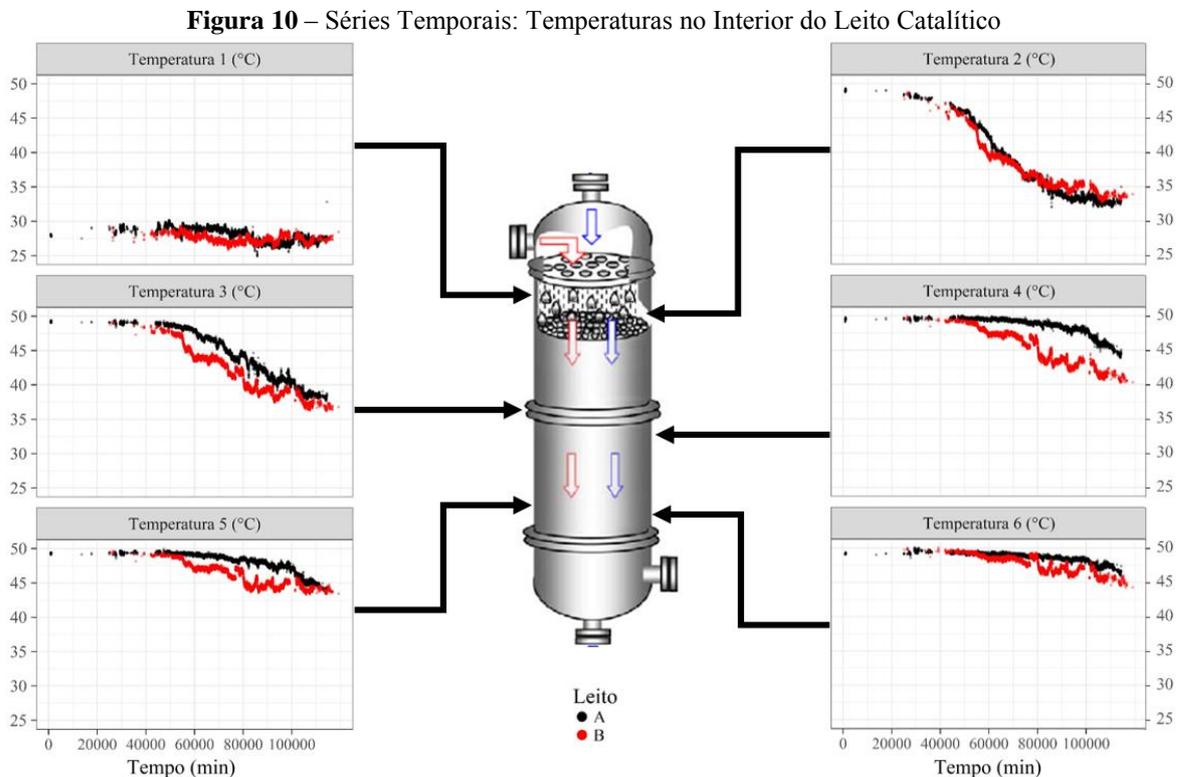
Figura 9 – Séries Temporais: Vazões de Carga Fresca, de Reciclo e de Hidrogênio



Fonte: Elaborado pelo próprio autor (2017).

É possível observar na Figura 9 que as variáveis apresentam comportamento similar em ambos os leitos. A vazão de reciclo, uma variável controlada (ver Figura 2), do leito A possui valor numérico maior que B, mas o comportamento geral é semelhante, com saltos somente no final da campanha. Observa-se, entretanto, uma diferença gráfica na vazão de hidrogênio entres os leitos. Em aproximadamente 80000 minutos, há um degrau na carga de hidrogênio no leito B. Como V2 é controlada, esse salto foi determinado pelo operador, que deve ter observado uma queda de conversão, que precisou ser contrabalanceada por um aumento na vazão de H₂. Esse salto pode causar a presença de grupos diferentes na massa de dados do leito B que devem ser considerados na construção dos modelos.

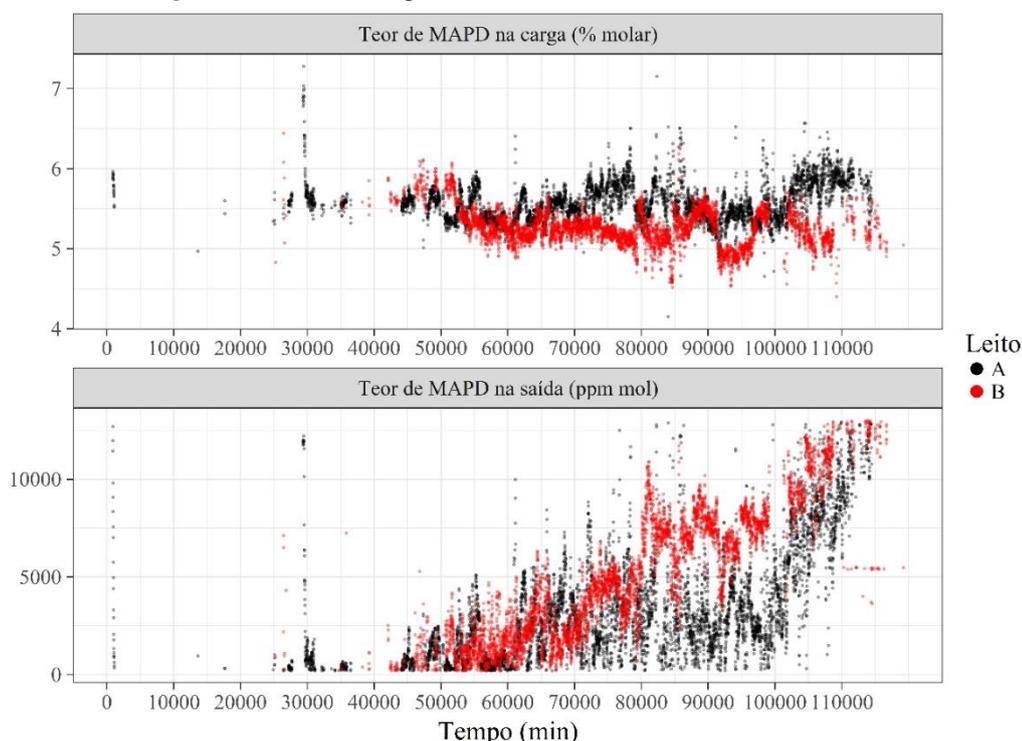
Já na Figura 10, é possível observar a natureza exotérmica da reação. A Temperatura 1, mais próxima a entrada do reator, varia no intervalo de 25°C a 30°C, enquanto que a Temperatura 6, mais ao fundo, varia entre 45°C e 50°C. A queda observada nas temperaturas que seguem a primeira, principalmente nas intermediárias (3 e 4), se deve à queda na reatividade do catalisador e consequente queda de conversão. Dada a natureza exotérmica da reação, uma queda de conversão implica numa menor geração de calor, o que reduz a temperatura observada.



Fonte: Elaborado pelo próprio autor (2017).

A Figura 11 indica que a carga fresca possui níveis constantes de teor do contaminante, com picos esporádicos. É importante notar que a carga fresca é proveniente de uma seção anterior do processo, a qual impede que a porcentagem molar de MAPD não ultrapasse 7%. A omissão de dados no início do da campanha é resultado do limite de detecção do TCD, como descrito anteriormente. Há uma tendência clara de aumento no teor de MAPD na saída do reator ao longo do tempo, evidenciando mais uma vez a perda de reatividade do catalisador. Observa-se também um degrau no teor de MAPD na saída no leito B ao redor de 80000 minutos. Este aumento repentino deve ter causado o operado da planta a aumentar drasticamente a vazão de H_2 , causando o salto observado na Figura 9.

Figura 11 – Séries Temporais: Teor de MAPD na Entrada e na Saída



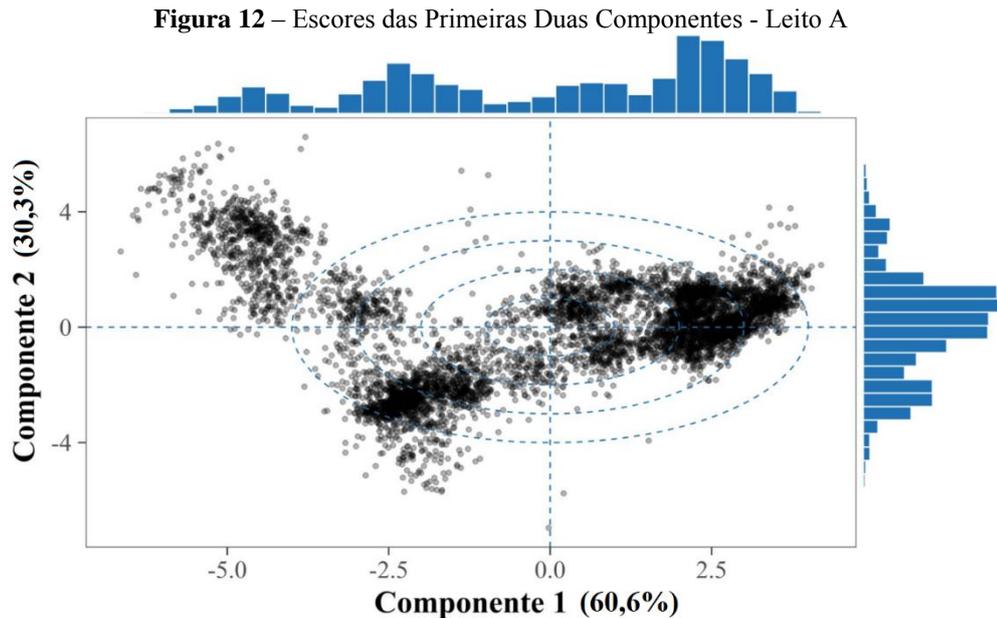
Fonte: Elaborado pelo próprio autor (2017).

5.1.2 Análise de Componentes Principais

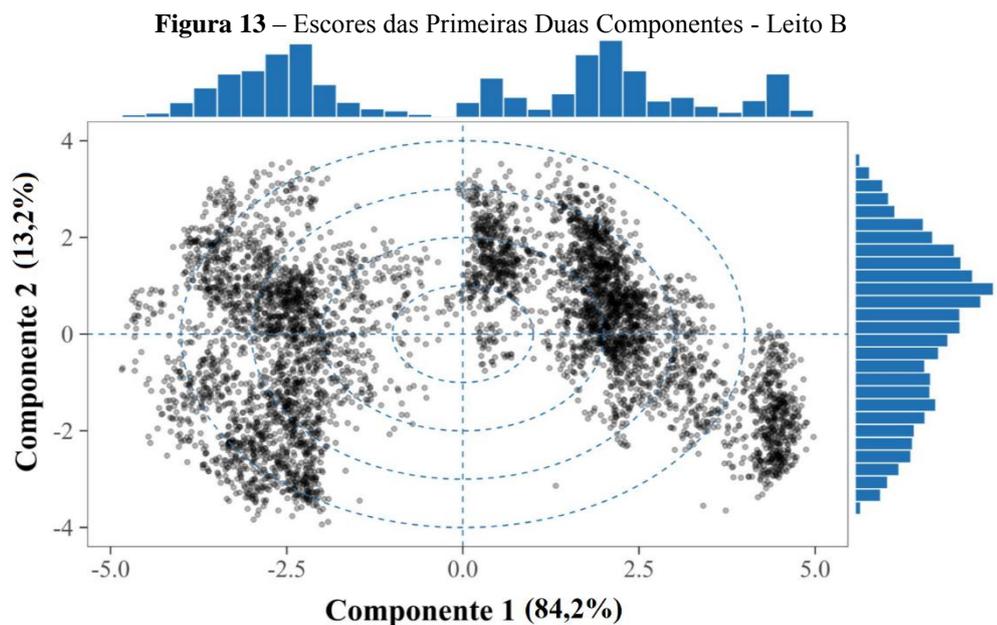
Os escores das primeiras duas componentes para os dados do leito A são apresentados na Figura 12. Graficamente, não há a presença de grupos claramente definidos, apesar de se verificar um agrupamento de dados mais à direita (valores positivos da PC 1).

De maneira análoga, há indicação clara de dois grupos distintos quando os escores das duas primeiras componentes para o leito B são analisados na Figura 13. É possível observar que a tendência de agrupamento é dada pelo valor numérico da primeira componente. O grupo

mais à direita possui valores positivos, enquanto que o grupo mais esquerda possui valores negativos. A presença de dois grupos distintos foi causada pelo salto na carga de H_2 , discutido anteriormente, e confirmado pelo gráfico de *loadings* apresentado na Figura 14. De fato, o salto em H_2 impacta em quase todas as outras variáveis, o que gera condições diferentes de processo e o conseqüentemente agrupamento de dados.



Fonte: Elaborado pelo próprio autor (2017).



Fonte: Elaborado pelo próprio autor (2017).

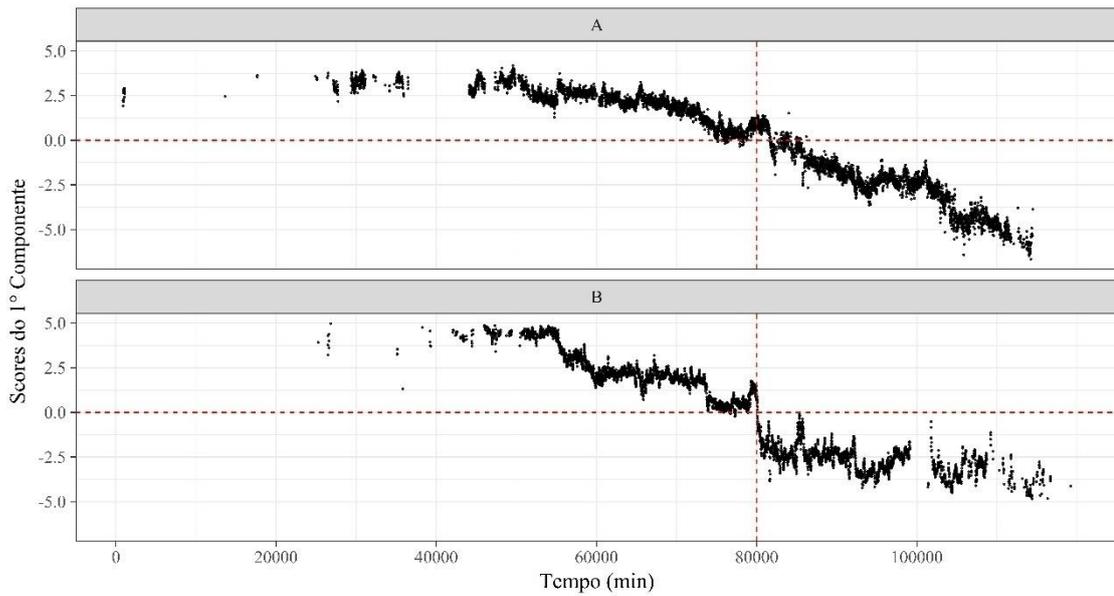
Porém, não há nenhuma variável com carga fatorial na PC1 elevada o suficiente para explicar isoladamente a separação dos grupos, como é possível ver na Tabela 5. Isto dito, as cargas fatoriais da vazão de hidrogênio (V2) e de temperatura (V7-V11) possuem valores superiores quando comparadas às demais, o que indica que a tendência de separação pode estar relacionada a estas variáveis. Como se sabe que a queda de reatividade do catalisador, com o conseqüente salto na vazão de H₂, é função do tempo, procurou-se investigar a relação entre os escores da PC1 e o tempo.

Tabela 5 – Cargas fatoriais das cinco primeiras componentes - Leito B

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5
V1	-0,190	-0,354	0,325	-0,176	0,574
V2	-0,100	-0,089	-0,776	-0,565	0,019
V3	-0,290	-0,239	0,275	-0,197	0,074
V4	-0,334	-0,205	0,016	0,068	-0,332
V5	0,054	-0,549	-0,219	0,317	0,123
V6	0,150	-0,527	-0,213	0,250	0,058
V7	0,346	-0,171	0,027	-0,064	0,028
V8	0,366	-0,082	0,034	-0,038	0,015
V9	0,366	-0,023	0,058	-0,020	-0,018
V10	0,364	-0,067	0,008	-0,023	-0,141
V11	0,348	-0,037	0,099	0,010	-0,221
V12	-0,270	-0,255	0,047	0,149	-0,617
V13	0,139	-0,284	0,325	-0,643	-0,293

Fonte: Elaborado pelo autor (2017).

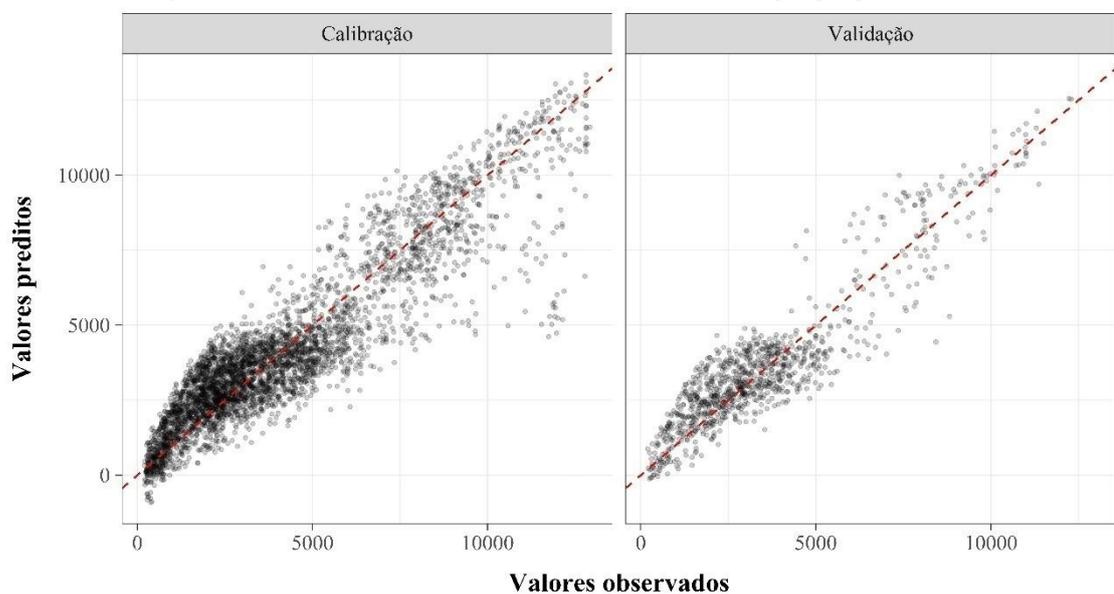
A Figura 14 evidencia de maneira clara que o agrupamento de dados no leito B através do valor numérico da PC1 é consequência do salto na vazão de H₂ em 80000 minutos, como predito anteriormente. Desta forma, a não ser que um modelo único consiga representar ambas condições de operação, é necessária a construção de dois modelos distintos para o leito B. Apesar de comportamento similar, queda ao longo do tempo, não há fortes evidências na Figura 14 que indiquem a presença do grupos no leito A, o que já havia sido observado na Figura 12.

Figura 14 – Escores da PC1 em Função do Tempo - Leitos A e B

Fonte: Elaborado pelo próprio autor (2017).

5.2 MODELO DE REGRESSÃO PLS

Os valores numéricos do teor de MAPD, preditos através do modelo PLS do leito A, são apresentados na Figura 15 em função dos valores observados, medidos pelo TCD. É possível observar que os pontos se espalham ao redor da reta de suporte de maneira adequada, o que indica a capacidade de predição do modelo. O modelo possui R^2 de 0,84, RMSEP de 1125 mg Kg^{-1} e BIAS de $22,7 \text{ mg Kg}^{-1}$. O modelo foi construído com cinco variáveis latentes (ver Apêndice B).

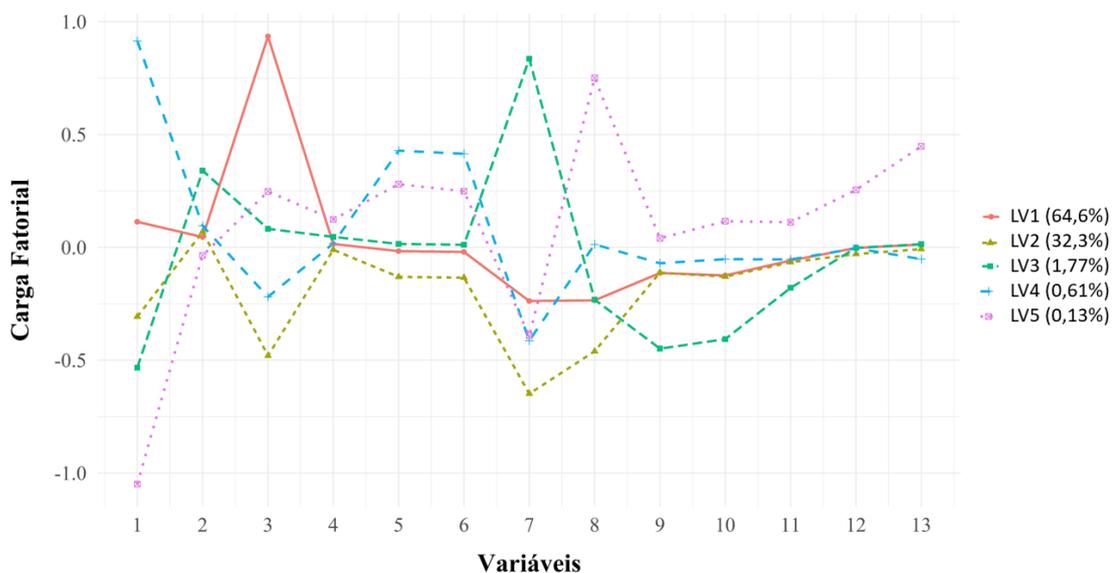
Figura 15 – Teor de MAPD: Predito vs. Observado, em mg mg Kg^{-1} - Leito A

Fonte: Elaborado pelo próprio autor (2017).

Também é possível observar na Figura 15 que o conjunto de dados de calibração e de validação se comportam de maneira semelhante, o que indica que o método de amostragem, descrito na seção 3.7.2, é eficiente na representatividade das amostras.

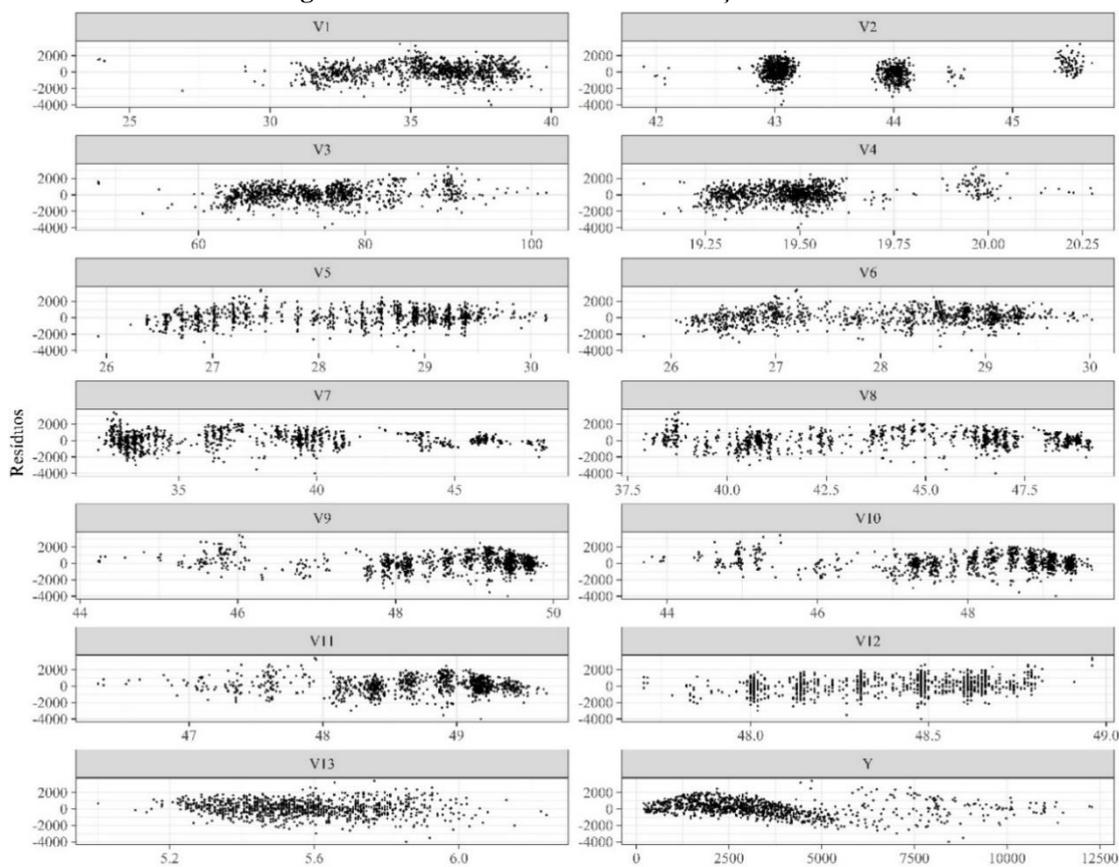
Na Figura 16 estão representadas as cargas fatoriais de cada variável de processo nas cinco LVs do modelo PLS para o Leito A, bem como a variância explicada para cada LV. É possível observar o alto peso de V3 (vazão de hidrogênio) na primeira LV, a que explica a maior variabilidade observada nos dados. Tal variável é de extrema importância ao processo já que ela impacta fenomenologicamente em diversas outras variáveis, incluindo de temperatura, pressão e saída do MAPD. Assim sendo, o modelo consegue ilustrar de forma matemática a importância de V3, extraíndo características inerentes ao processo, tornando-se mais relevante.

Figura 16 – Cargas Fatoriais das PVs nas LVs do Modelo PLS - Leito A



Fonte: Elaborado pelo próprio autor (2017).

Figura 17 – Resíduos do modelo em Função das PVs-Leito A



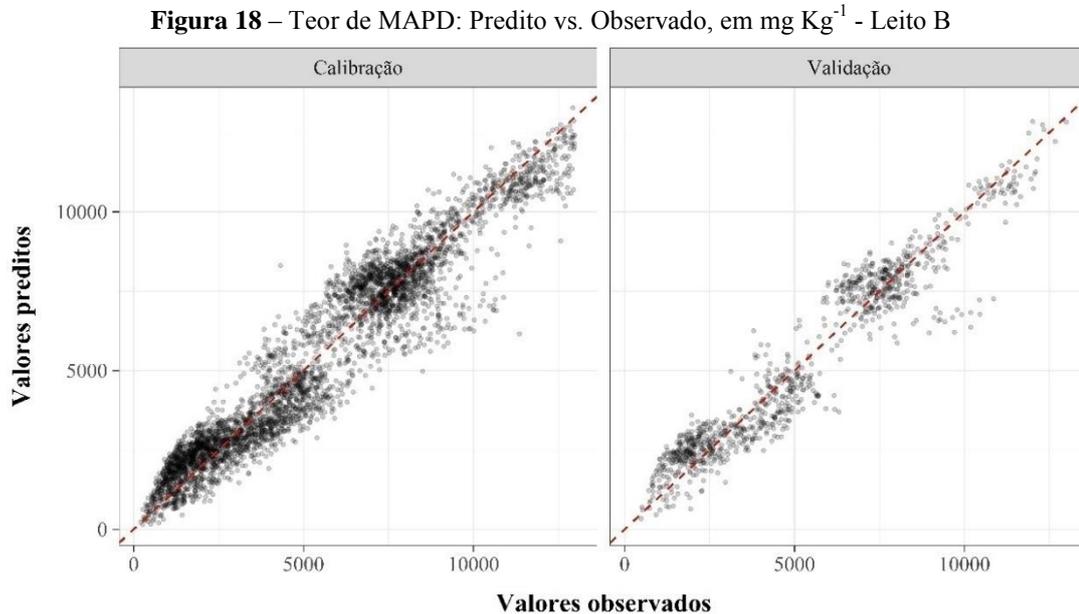
Fonte: Elaborado pelo próprio autor (2017).

Na Figura 17 é verificada a hipótese de homocedasticidade dos resíduos contra todas as variáveis de processo. Não há uma indicação clara de tendências na Figura 17, e a presença de agrupamento acontece em variáveis controladas onde não há a presença de valores intermediários. Assim, não há indicação gráfica da falta de homocedasticidade dos mesmos.

De maneira análoga, os valores numéricos do teor de MAPD, preditos através do modelo PLS do leito B, são apresentados na Figura 18 em função dos valores observados, medidos pelo TCD. Novamente, é possível observar que os pontos se espalham ao redor de reta de suporte de maneira adequada, o que indica uma boa capacidade de predição do modelo. Devido à existência de dois grupos com comportamentos distintos na massa de dados do leito B, discutida na seção 5.1.2, haveria necessidade de construção de dois modelos, um para cada condição do leito.

Entretanto, a Figura 18 indica que um modelo PLS único consegue representar de maneira adequada ambas as situações. Desta forma, decidiu-se pela simplicidade de um modelo único, com R^2 de 0,92, RMSEP de 878 mg Kg^{-1} e *bias* de $1,8 \text{ mg Kg}^{-1}$. Assim como

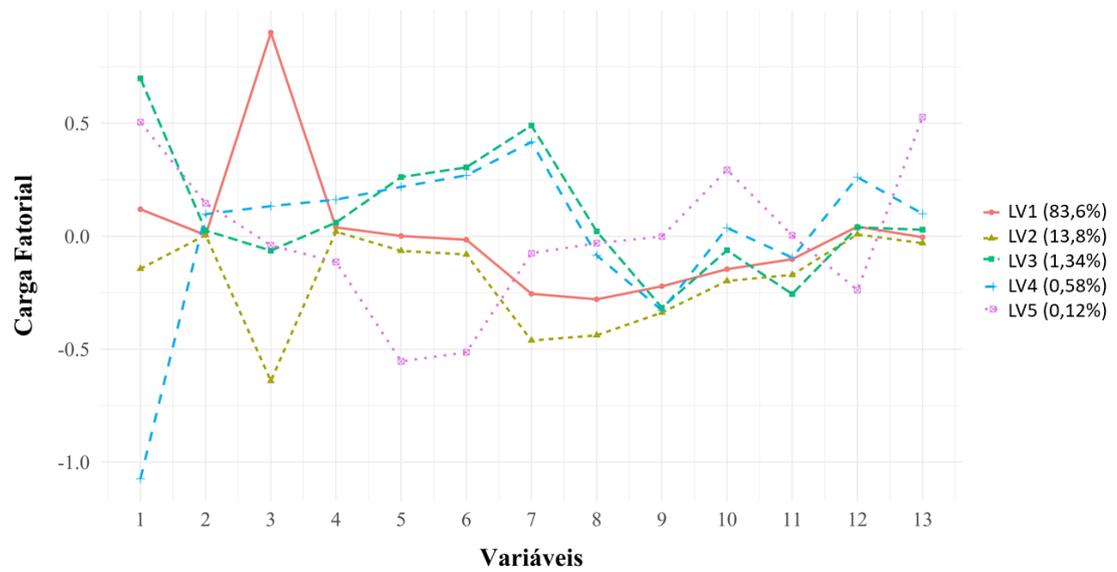
no leito A, o modelo foi construído com cinco LVs, havendo notável semelhança entre os comportamentos das amostras de calibração e validação.



Fonte: Elaborado pelo próprio autor (2017).

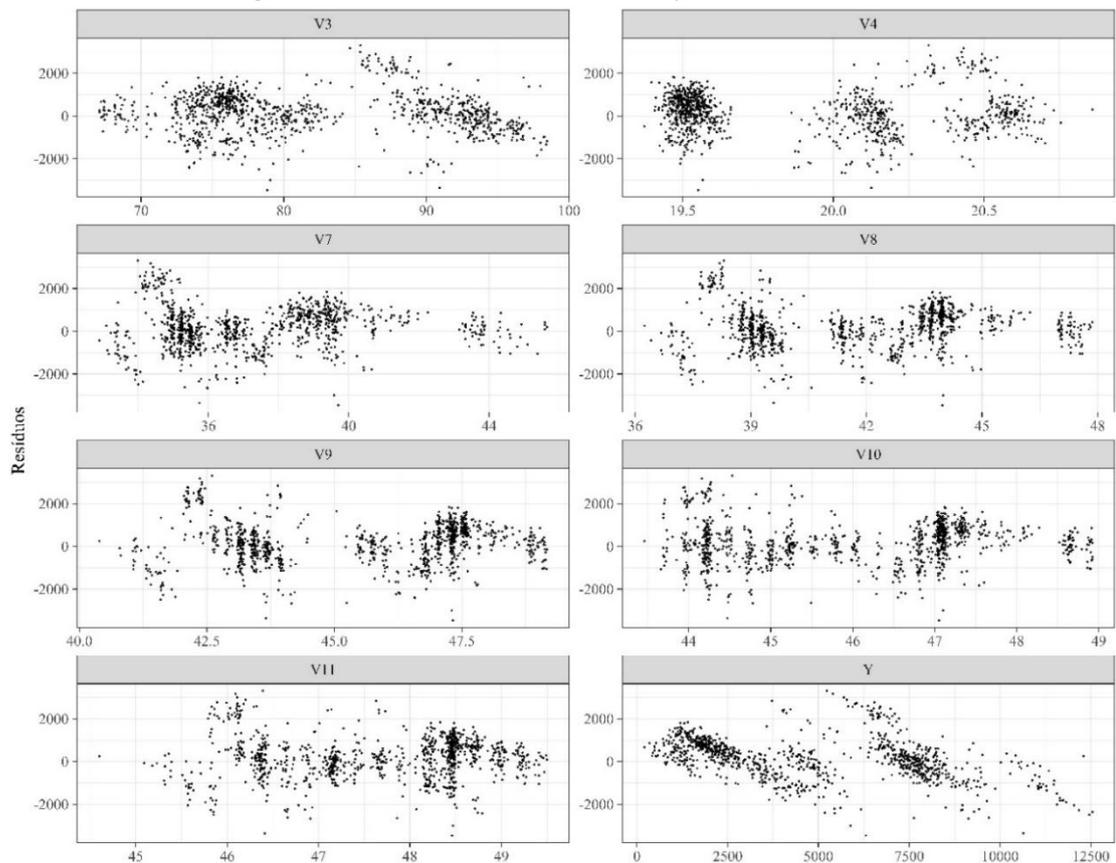
A Figura 19 ilustra as cargas fatoriais das PVs nas cinco LVs do modelo PLS para o leito B. Novamente, é possível observar o alto peso de V3 (vazão de hidrogênio) na primeira LV, característica semelhante ao modelo PLS do leito A. É importante notar que um possível teste de significância das cargas fatoriais das PVs no modelo seria sensível a enorme quantidade de dados utilizada na geração do modelo, o que geraria uma conclusão distorcida. De fato, a extensão da massa de dados anula a maioria dos testes de hipóteses que assume normalidade já que, de acordo com o teorema do limite central, um número imenso de observações enviesaria o teste. Dessa forma, a significância das PVs será testada através da técnica de VIP, que, de certa forma, indica, através de um índice, se a carga fatorial é importante ou não ao modelo.

Figura 19 – Cargas Fatoriais das PVs nas LVs do Modelo PLS - Leito B



Fonte: Elaborado pelo próprio autor (2017).

Figura 20 – Resíduos do modelo em Função das PVs- Leito B



Fonte: Elaborado pelo próprio autor (2017).

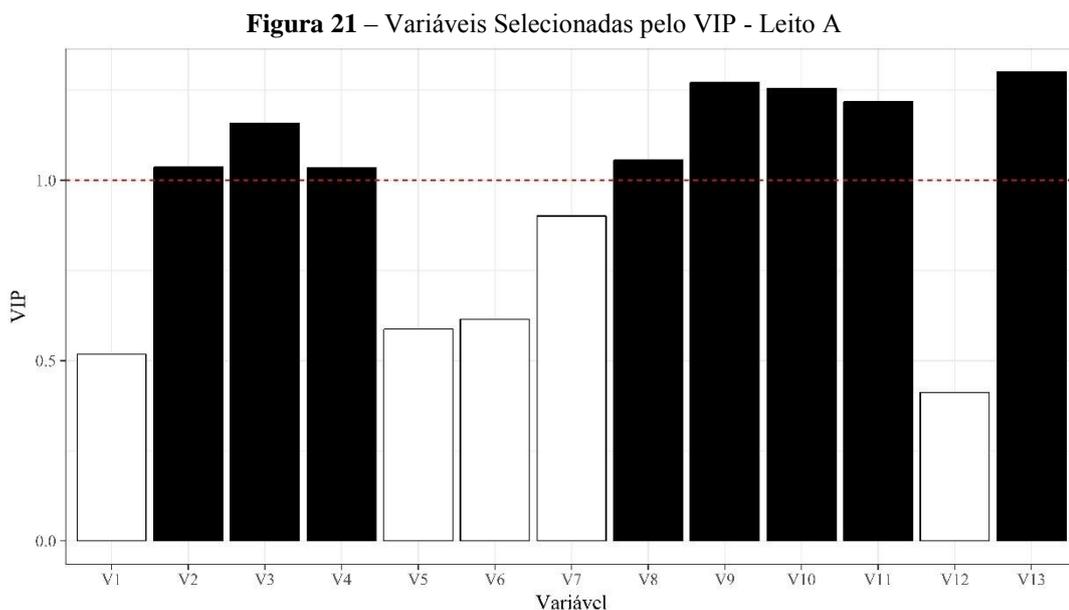
De maneira análoga à Figura 17, na Figura 20 é verificada a hipótese de homocedasticidade dos resíduos contra todas as variáveis de processo, e, novamente, não há

indicação gráfica da falta de homocedasticidade dos mesmos. Além disso, os resíduos também se mostraram aleatórios frente a variáveis preditas e observadas.

5.3 MODELO DE REGRESSÃO PLS-VIP

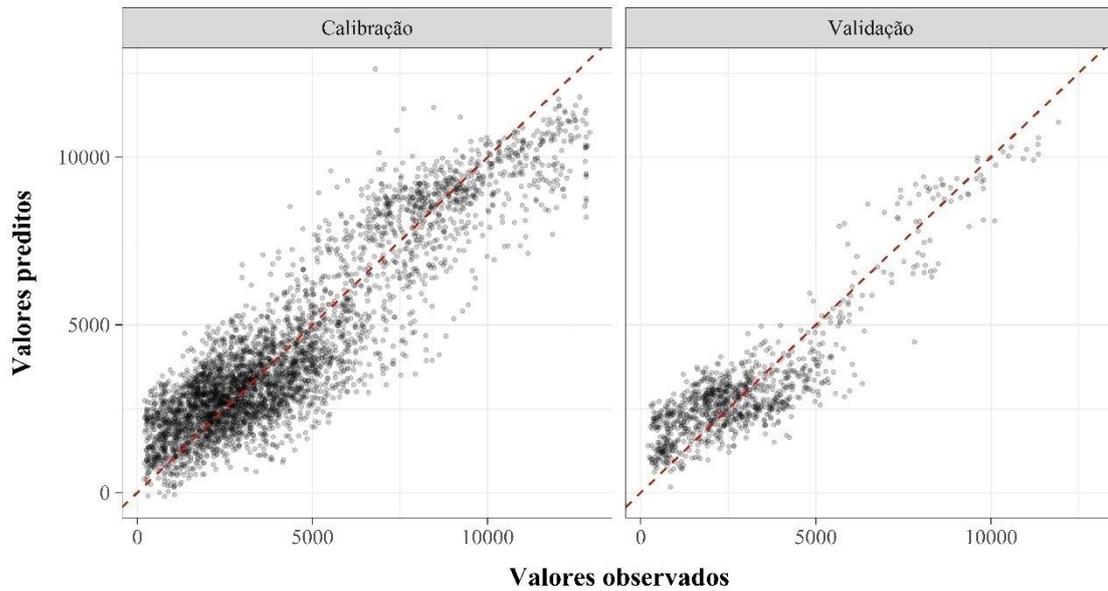
Os escores VIP das PVs do modelo PLS construído para o leito A são apresentados na Figura 21. As variáveis que apresentaram valor numérico maior que a unidade foram V2 (vazão de reciclo), V3 (vazão de hidrogênio), V4 (pressão de leito), V8 a V11 (temperaturas no interior do leito catalítico) e V13 (teor de MAPD na carga fresca). Um novo modelo PLS foi construído somente com as variáveis selecionadas, e os novos valores preditos do teor de MAPD para o leito A utilizando este modelo são apresentados na Figura 22.

É possível observar que, quando comparados à Figura 15, os pontos na Figura 22 se espalham ao redor de reta de suporte numa região maior, porém ainda de maneira adequada. Esse maior espalhamento é refletido no aumento de RMSEP, que passa de 1125 mg Kg^{-1} para 1207 mg Kg^{-1} , um aumento de aproximadamente 7%. O novo modelo possui R^2 de 0,81 e *bias* de 28 mg Kg^{-1} . A partir das equações (22) e (23), na seção 3.9, chega-se a um t_{cal} de -0,231 e ν de 23138. Considerando 95% de confiança, o $t_{critical}$ com esse grau de liberdade é de 1,645, bem próximo à distribuição normal, devido ao alto grau de liberdade. Como $|t_{cal}| < |t_{critical}|$, não há evidência significativa para rejeitar a hipótese nula de que os modelos são estatisticamente iguais.



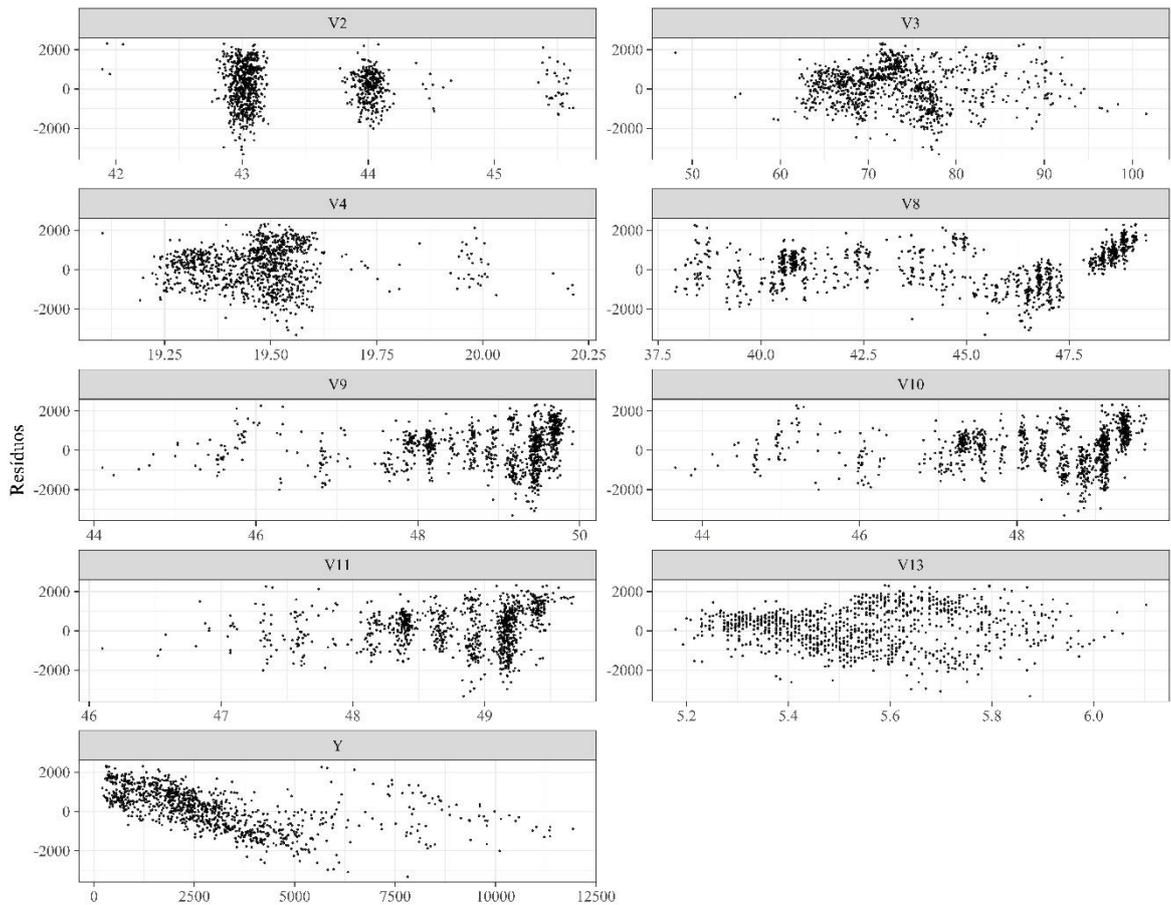
Fonte: Elaborado pelo próprio autor (2017).

Figura 22 – Teor de MAPD: Predito vs. Observado, em mg Kg^{-1} - Leito A - VIP



Fonte: Elaborado pelo próprio autor (2017).

Figura 23 – Resíduos do modelo em Função das PVs - Leito A - VIP



Fonte: Elaborado pelo próprio autor (2017).

Por fim, na Figura 23 é verificada graficamente a hipótese de homocedasticidade e aleatoriedade dos resíduos contra todas as variáveis de processo selecionadas. Assim, sendo

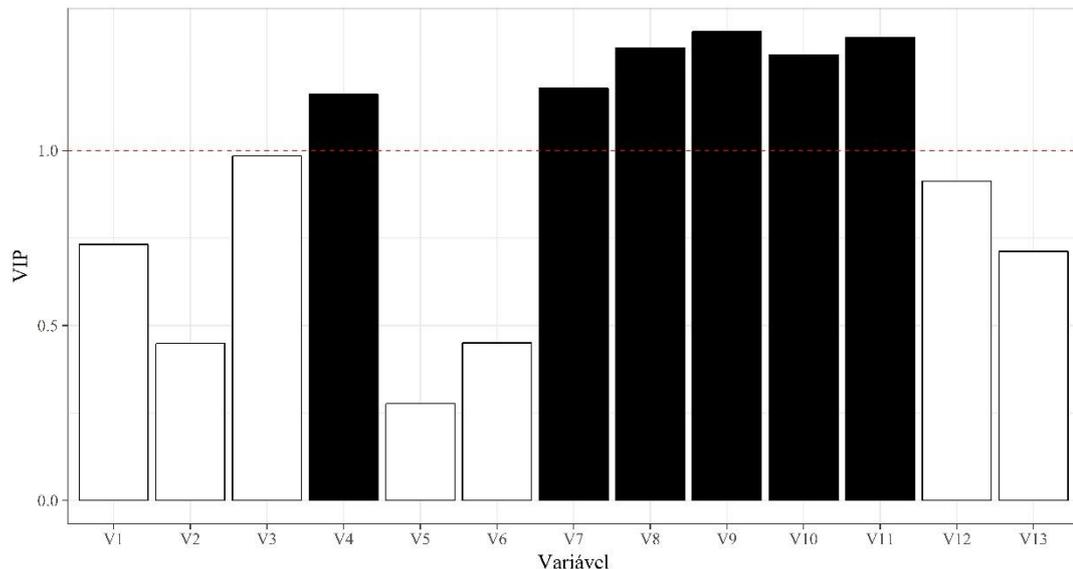
os modelos estatisticamente iguais, deve-se usar o modelo mais simples, PLS-VIP, que exige o monitoramento de um número menor de variáveis.

Assim, a equação (24) apresenta o modelo PLS-VIP para o modelo A

$$Y = -26,8 V2 - 13,4 V3 + 786,2 V4 + 61,4 V8 - 1023,9 V9 - 256,2 V10 - 304,1 V11 + 5523,9 V13 \quad (24)$$

De maneira análoga, os índices VIP das PVs do modelo PLS construído para o leito B são apresentados na Figura 24. As variáveis que apresentaram valor numérico maior que a unidade foram V4 (pressão de leito) e V7 a V11 (temperaturas no interior do leito catalítico). É interessante notar que V3 (vazão de hidrogênio) obteve valor bem próximo da unidade (0,985), porém ainda assim abaixo do valor mínimo para seleção da variável. Entretanto, a importância fenomenológica da vazão de hidrogênio no processo associada à proximidade do índice obtido pela variável sugere que a mesma deve ser mantida na construção do modelo PLS-VIP para o leito B. Além disso, no leito anterior também houve seleção de V13 (teor de MAPD na carga fresca), descartada no leito B. De fato, as únicas variáveis selecionadas pelo VIP para o leito B, com exceção de V3, são as variáveis de estado, relacionadas à pressão e à temperatura.

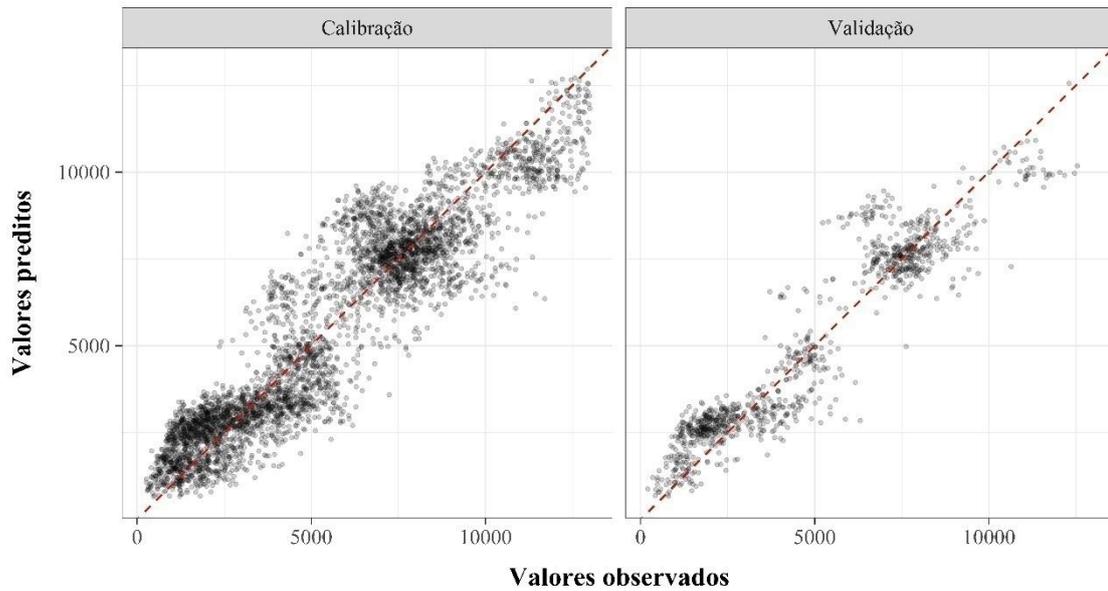
Figura 24 – Variáveis Seleccionadas pelo VIP - Leito B



Fonte: Elaborado pelo próprio autor (2017).

Os valores preditos do teor de MAPD para o leito B a partir do novo modelo PLS construído somente com as variáveis selecionadas pelo VIP são apresentados na Figura 25.

Figura 25 – Teor de MAPD: Predito vs. Observado, em mg Kg^{-1} - Leito B - VIP



Fonte: Elaborado pelo próprio autor (2017).

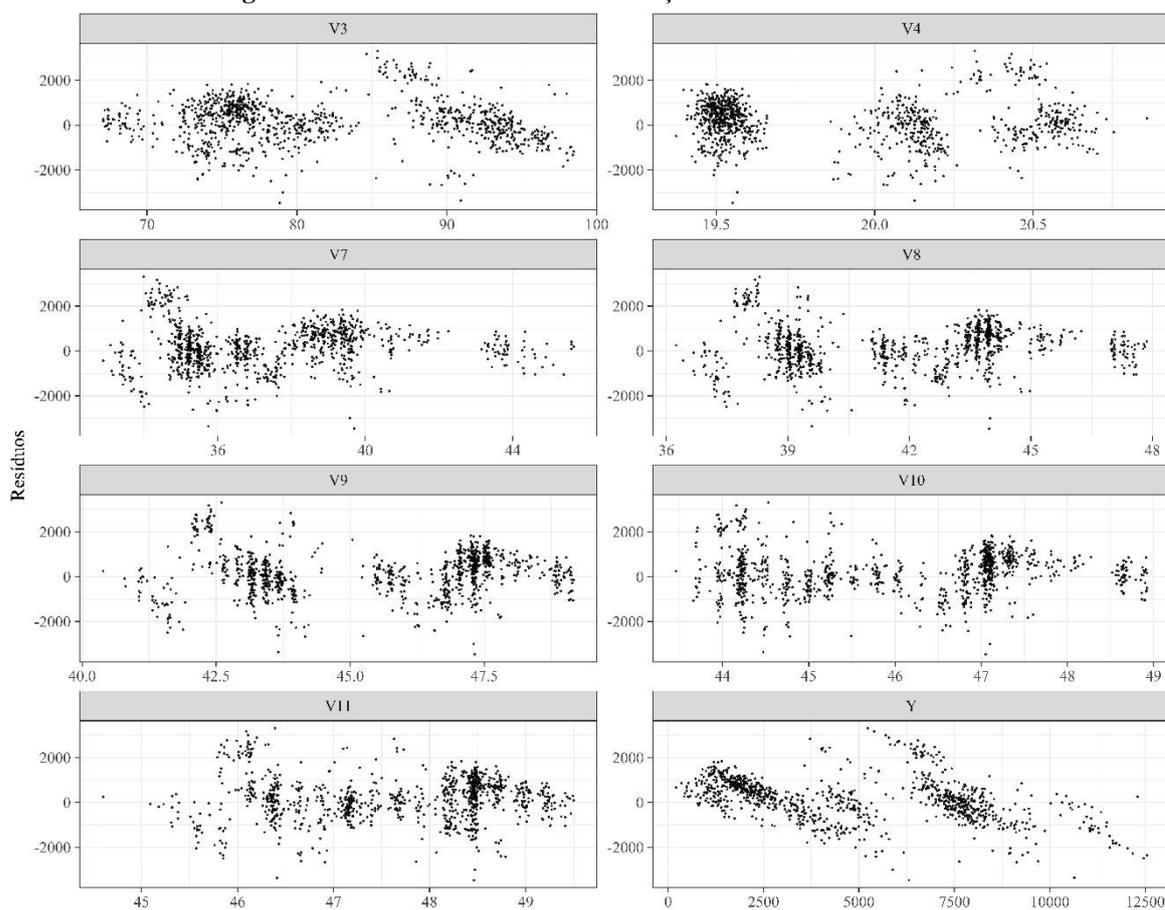
Novamente, é possível observar que, quando comparados ao modelo PLS anterior ao VIP, ilustrado na Figura 18, os pontos na Figura 25 se espalham ao redor de reta de suporte numa região maior, porém ainda de maneira adequada. Como antes, esse maior espalhamento é refletido no aumento de RMSEP, que passa de 878 mg Kg^{-1} para 1132 mg Kg^{-1} , um aumento de aproximadamente 29%, visivelmente superior ao aumento no modelo do leito A. O novo modelo possui R^2 de 0,87 e *bias* de 32 mg Kg^{-1} . A partir das equações (22) e (23), na seção 3.9, chega-se a um t_{cal} de -1,505 e v de 35585. Considerando 95% de confiança, o $t_{critical}$ com esse grau de liberdade é de 1,645, bem próximo à distribuição normal, devido ao alto grau de liberdade. Assim como no modelo PLS-VIP para o leito A, $|t_{cal}| < |t_{critical}|$, não há evidência suficiente para rejeitar a hipótese nula de que os modelos são estatisticamente iguais.

Por fim, na Figura 26 é verificada graficamente a hipótese de homocedasticidade e aleatoriedade dos resíduos contra todas as variáveis de processo selecionadas. Assim sendo, os modelos estatisticamente iguais, deve-se usar o modelo mais simples, PLS-VIP, que exige o monitoramento de um número menor de variáveis.

Os resultados apresentados na Tabela 6 são utilizados para avaliar o efeito da seleção de variáveis nos modelos, baseada nas figuras de mérito descritas na seção 3.8. É possível observar que nos modelos PLS-VIP de ambos os leitões houve um aumento no valor do RMSEP e de BIAS. Este último, porém, ainda ilustra a presença de um baixo erro sistemático considerando a ordem de grandeza dos valores do teor de MAPD, particularmente ao final da

campanha. Entretanto, em ambos os casos, não houve diferença significativa entre os modelos PLS e PLS-VIP à um nível de confiança de 95%.

Figura 26 – Resíduos do modelo em Função das PVs - Leito B - VIP



Fonte: Elaborado pelo próprio autor (2017).

Tabela 6 – Figuras de mérito dos modelos antes e depois da seleção de variáveis

Leito	Antes da seleção de Variáveis				Depois da seleção de Variáveis			
	Nº PVs	R ²	RMSEP / mg Kg ⁻¹	BIAS / mg Kg ⁻¹	Nº PVs	R ²	BIAS / mg Kg ⁻¹	RMSEP / mg Kg ⁻¹
A	13	0.84	1125	23	9	0.81	28	1207
B	13	0.92	878	1.8	8	0.87	32	1131

Fonte: Elaborado pelo autor (2017).

Assim, a equação (25) apresenta o modelo PLS-VIP para o modelo B.

$$Y = 59,8 V3 - 585,0 V4 + 278,7 V7 - 884,3 V9 - 1734,0 V10 - 1260,9 V11 \quad (25)$$

Testes F foram realizados para determinar se as variâncias dos modelos PLS-VIP são significativamente diferentes das variâncias dos modelos PLS. A um nível de confiança de 95%, com os graus de liberdade descritos anteriormente, o F crítico para ambos os leitos é de 1,30. Para o leito A, o F_{cal} foi de 1,07, e para o leito B o F_{cal} foi de 1,29. Como no teste F a hipótese nula é de as variâncias são iguais, e $F_{cal} < F_{crítico}$, não há evidência suficiente para rejeitar a hipótese nula. Desta forma, as variâncias (RMSEP) dos modelos são estaticamente iguais.

Dada a natureza industrial dos dados, os coeficientes de determinação (R^2) para ambos os modelos, de ambos os leitos, demonstram-se satisfatórios. Desta forma, foi possível alcançar um desempenho adequado para os modelos PLS mesmo quando foram empregadas um menor número de variáveis de processo. A nível de comparação, todos os modelos foram construídos com cinco variáveis latentes, correspondente ao número de LVs para os modelos sem VIP (ver Apêndice 2).

É importante notar que a implementação do modelo PLS-VIP não implica que as variáveis excluídas do modelo deixarão de ser monitoradas pelo operador da planta. Entretanto, caso a variável deixe de ser medida, por falha do equipamento de medição, o modelo PLS-VIP consegue estimar o teor de MAPD, enquanto que o modelo PLS não, já que não há informação sobre uma das variáveis de entrada, o que impede uma saída do modelo.

6 CONCLUSÕES

Na indústria química brasileira, não é comum a aplicação de analisadores virtuais como substituinte de analisadores físicos. Possíveis razões incluem o desconhecimento dos métodos de regressão multivariada associada ao *know-how* das técnicas analíticas já estabelecidas. Os modelos PLS desenvolvidos neste trabalho demonstram que é possível realizar, de forma satisfatória, a previsão do teor de MAPD em reatores *trickle bed* a partir da informação numérica das variáveis de processo.

Este trabalho avaliou não somente a capacidade preditiva dos modelos construídos, mas também o impacto da seleção de variáveis no desempenho dos mesmos. Para o leito A, foi possível, a partir do VIP, reduzir o número de variáveis de entrada do modelo PLS de 13 para 9 variáveis de processo. Essa redução provocou um leve aumento no RMSEP e na *bias* do modelo, porém não houve diferença significativa entre os dois modelos à um nível de significância de 5%. Ao final, o modelo PLS-VIP para o leito A apresentou R^2 de 0,81 e RMSEP de 1207 mg Kg⁻¹. Para o leito B a redução do número de variáveis de entrada foi ainda maior, saindo de 13 para 8, e mantendo somente variáveis de estado (pressão e temperatura). O impacto no RMSEP e no *bias* foi mais visível, porém, assim como em A, não houve diferença significativa entre os modelos a um teste de significância de 5%. O modelo PLS-VIP para o leito B apresentou R^2 de 0,87 e RMSEP de 1131 mg Kg⁻¹. Assim, foi possível manter modelos com boa capacidade preditiva a partir do monitoramento de um número menor de variáveis de processo, o que, além de tornar o modelo mais simples, pode gerar uma redução nos custos de manutenção dessas variáveis, caso essas não sejam críticas à segurança do processo.

Além da construção dos modelos, a análise exploratória dos dados elucidou fatores presentes no processo que são essenciais ao entendimento da resposta final do analisador virtual. Foi possível detectar a influência da vazão de hidrogênio sobre as demais variáveis de processo, e como uma mudança abrupta naquela impacta o comportamento de toda a massa de dados seguinte. Não somente isso, houve indícios de excesso na vazão de hidrogênio, que, junto à reatividade do catalisador no início da campanha, torna os valores do teor de MAPD abaixo do limite de detecção do analisador real, o TCD. Isto impede que o analisador virtual trabalhe em faixas menores de concentração, já que não há significado real, mesmo que haja valores numéricos, para respostas do TCD abaixo deste limite. Como os parâmetros do modelo PLS são estimados levando em consideração não somente as variáveis de entrada,

como também a de saída, a utilização das respostas abaixo do limite de detecção comprometeria a capacidade preditiva do modelo.

Desta forma, a metodologia desenvolvida neste trabalho, bem como seu produto final, os modelos PLS-VIP para os leitos A e B, apresentam alto potencial para implementação num processo de conversão de contaminantes MAPD num reator *trickle bed*. Além do impedimento de uma possível parada na produção devido a queda de analisadores reais, que fornecem informações essenciais à continuidade do processo, uma vez desenvolvido o modelo matemático, a implementação do analisador virtual nos sistemas de controle presentes na indústria não demanda grandes esforços ou custos. Assim sendo, o analisador desenvolvido neste trabalho pode ser prontamente implementado na indústria, como também a metodologia de construção pode ser replicada em outros processos, desde que devidamente seguida.

7 SUGESTÕES PARA TRABALHOS FUTUROS

Como visto neste trabalho, o desenvolvimento de um analisador virtual confiável para medir o teor de MAPD num reator *trickle bed* não é trivial. Porém, uma vez construído, como todas as implicações estatísticas necessárias testadas, sua implementação ao sistema de controle de uma planta industrial não demanda muitos esforços. Portanto, para garantir a exatidão da resposta do modelo, é necessário assegurar que os dados de entradas estarão disponíveis e serão confiáveis. Isso implica num esforço maior para garantir a manutenibilidade dos equipamentos de medição que irão fornecer os dados de entrada do sistema. Dessa forma, existe um *trade-off* entre o custo de manutenção do modelo PLS para o analisador virtual e a precisão da resposta do modelo. A importância de um trabalho futuro que analise essa relação está no fato de adicionar um fator econômico ao quase que puramente estatístico desenvolvimento do modelo. A posse dessa informação econômica do modelo contribuirá para uma decisão mais orientada por dados em processos industriais complexos.

Além disso, o analisador construído neste trabalho estima somente o teor de contaminantes MAPD no reator *trickle bed*. Porém, existem outros índices importantes ao bom funcionamento do processo. Como o produto final é propeno químico, uma vazão excessiva de gás hidrogênio, para compensar a falta de reatividade do catalisador, pode converter propeno em propano, reduzindo o rendimento global do processo. A partir de uma metodologia semelhante, é possível construir um analisador virtual que, além de fornecer o teor de MAPD, também estime o teor de propano no produto final. Esse modelo pode ser utilizado, num trabalho futuro, para otimizar a vazão de hidrogênio no reator, já que existe um *trade off* entre a eliminação do contaminante MAPD e a geração de propeno para altas vazões de H₂. Estimativas iniciais indicam que a redução de 1 ponto percentual na concentração do propano na saída do reator significa uma captura anual de aproximadamente 9 milhões de reais, um ganho significativo aos rendimentos de uma empresa petroquímica.

8 PRODUTOS GERADOS NESTA PESQUISA

Esta dissertação é produto de uma pesquisa que gerou outros trabalhos, publicados em eventos na área de engenharia industrial, cujas referências são apresentadas abaixo.

MASSA, A. R. C. G. et al. Modelagem empírica de um reator trickle bed com desativação catalítica através de redes neuronais artificiais. In: CONGRESSO BRASILEIRO DE ENGENHARIA QUÍMICA, 21., 2016. **Anais...** [S.l.: s.n.], 2016.

MASSA, A. R. C. G. et al. Desenvolvimento de um analisador virtual para determinação do teor dos contaminantes MAPD em um reator *trickle bed*, utilizando a regressão por mínimos quadrados parciais. In: CONGRESSO BRASILEIRO DE ENGENHARIA QUÍMICA, 21., 2016. **Anais...** [S.l.: s.n.], 2016.

MASSA, A. R. C. G. et al. Development of a virtual analyser through partial least square regression and variance influence projection to estimate the content of MAPD contaminants in a trickle bed reactor using R. In: Seminário Internacional de Estatística com R, 2., 2017, Niterói. [S.l.: s.n.], 2017.

Além destes, dois resumos foram aceitos na conferência de sistemas de Processos de Engenharia (PSE), que acontecerá em 2018 em San Diego, EUA, com os títulos:

- *Estimating the content of MAPD contaminants in a trickle bed reactor by developing a virtual analyser;*
- *Enabling online virtual analysers in industrial processes: a trade-off between implementation costs and output reliability.*

Por fim, os modelos desenvolvidos para ambos os leitos foram configurados no sistema do PIMS (*Plant Information Management System*) da planta industrial de origem dos dados. Desta forma, o analisador virtual detalhado neste trabalho poderá ser acessado a qualquer momento pelas equipes que necessitarem verificar o comportamento do teor de MAPD na saída do reator. Também será possível analisar a diferença entre o teor de MAPD estimado pelo analisador virtual e o fornecido pelo analisador *online*. Esta é a primeira fase da implementação completa do analisador virtual ao controle do processo.

REFERÊNCIAS

- ABDI, H. Partial least squares regression and projection on latent structure regression (PLS Regression). **Wiley Interdisciplinary Reviews: computational statistics**, v. 2, n. 1, p. 97-106, 2010. Disponível em: <<https://www.utdallas.edu/~herve/abdi-wireCS-PLS2010.pdf>>.
- AFANADOR, N. L. et al. Variable importance in PLS in the presence of autocorrelated data – case studies in manufacturing processes. **Chemometrics and Intelligent Laboratory Systems**, v. 139, p. 139-145, 2014.
- AL-DAHMAN, M. H. et al. High-pressure trickle bed reactors: a review. **Industrial & engineering chemistry research**, v. 36, n. 8, p. 3292-3314, 1997.
- ANNINO, R. Chromatographs: gas. In: LIPTÁK, B. G. **Instrument engineer's handbook: Process measurement and analysis**. Danvers: Library of Congress, 1995.
- ANTUNES, A. M. de S. **Setores da Indústria Química Orgânica**. Rio de Janeiro: E-papers, 2007.
- ANZANELLO, M. J. Seleção de variáveis para classificação de bateladas produtivas com base em múltiplos critérios. **Produção**, Porto Alegre, v. 23, n. 4, p. 858-865, 2013. Disponível em: <http://www.scielo.br/pdf/prod/v23n4/aop_t6_0006_0674.pdf>. Acesso em: 24 maio 2017.
- ANZANELLO, M. J.; ALBIN, S. L.; CHAOVALITWONGSE, W. A. Identificação das variáveis mais relevantes para categorização de bateladas de produção: reduzindo a variância do percentual de variáveis retidas. **Produto & produção**, Porto Alegre, v. 10, n. 3, p. 19-27, out. 2009. Disponível em: <<http://www.seer.ufrgs.br/ProdutoProducao/article/viewFile/7916/6529>>. Acesso em: 17 jan. 2017.
- ASTM Internacional. **ASTM E 1655-05** – Standard Practices for infrared Multivariate Quantitative Analysis. [S.l.: s.n.], 2012.
- AUGUSTO, F. **Cromatografia a gás: curso em dispositivos**. Campinas: Unicamp, 2000.
- BEEBE, K. R.; PELL, R. J.; SEASHOLTZ, M. B. **Chemometrics a Practical Guide**. Hoboken: John Wiley & Sons, 1998.
- BRANDÃO, D. S. **Aplicação da Espectroscopia Nir, associada à técnicas de calibração multivariada, na determinação da composição química do solvente utilizado na produção de isopreno**. 2012.105 f. Dissertação (Mestrado em Química) – Programa de Pós-

graduação em Química, Departamento de Química Analítica, Instituto de Química, Universidade Federal da Bahia, Salvador, 2012. Disponível em: <https://repositorio.ufba.br/ri/bitstream/ri/13093/1/Apresenta%C3%A7%C3%A3o_Denise_Vers%C3%A3oFinal.pdf>. Acesso em: 16 ago. 2017.

BRERETON, R. G. **Applied Chemometrics for Scientist**. Hoboken: John Wiley & Sons, 2007.

BRO, R.; SMILDE, A. K. Principal Component Analysis. **Analytical Methods**, v. 6, p. 2812-2831, 2014.

BRUNS, R. E.; SCARMINIO, I. S.; BARROS NETO, B. **Statistical design – Chemometrics**. Amsterdam: Elsevier, 2006.

CHONG, I-G.; JUN, C-H. Performance of some variable selection methods when multicollinearity is present. **Chemometrics and Intelligent Laboratory Systems**, v. 78, p. 103-112, 2005. Disponível em: <http://automatica.dei.unipd.it/public/Schenato/PSC/2010_2011/gruppo4-Building_termo_identification/IdentificazioneTermodinamica20072008/Biblio/Biblio-PLS/Performance%20of%20some%20variable%20selection%20methods%20when%20multicollinearity.pdf>. Acesso em: 12 abr. 2017.

COHN, P. E. **Analisadores Industriais**: no processo, na área de utilidades, na supervisão da emissão de poluentes e na segurança. Rio de Janeiro: Interciência; IBP, 2006.

DUDUKOVIC, M. P.; LARACHI, F.; MILLS, P. L. Multiphase reactors – revisited. **Chemical Engineering Science**, v. 54, n. 13-14, p. 1975-1995, 1999.

FERREIRA, M. M. et al. Quimiometria I: calibração multivariada, um tutorial. **Química Nova**, São Paulo, v. 22, n. 5, p. 724-731, 1999. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-40421999000500016>. Acesso em: 10 jul. 2017.

GELADI, P.; KOWLASKI B. Partial least square regression: a tutorial, **Analytica Chimica Acta**, Amsterdam, n. 185, p. 1-117, 1986. Disponível em: <<http://home.mit.bme.hu/~horvath/IDA/1-s2.0-0003267086800289-main.pdf>>. Acesso em: 15 maio 2017.

GREEN, M. M.; WITTCOFF, H. A. **Organic Chemistry Principles and Industrial Practice**. Weinheim: Wiley-VCH, 2003.

GUYON, I.; ELISSEEFF, A. An introduction to variable and feature selection. **The Journal of Machine Learning Research**, v. 3, p. 1157-1182, 2003. Disponível em: <<http://www.jmlr.org/papers/volume3/guyon03a/guyon03a.pdf>>. Acesso em: 6 jun. 2017.

GUO, J.; JIANG, Y.; AL-DAHMAN, M. H. Modeling of trickle bed reactors with exothermic reactions using cell network approach. **Chemical Engineering Science**, v. 63, n. 3, p. 751-764, 2008.

HARROU, F. et al. PLS-based EWMA fault detection strategy for process monitoring. **Journal of Loss Prevention in the Process Industries**, v. 36, p. 108-119, jul. 2015.

HUG, C. et al. Linking mutagenic activity to micropollutant concentrations in wastewater samples by partial least square regression and subsequent identification of variables. **Chemosphere**, v. 138, p. 176-182, nov. 2015.

HUNGER, R. Modelagem e Simulação de um Reator trickle bed para hidrogenação de propeno. 2002. 117 f. Dissertação (Mestrado em Engenharia Química) – Programa de Pós-graduação em Engenharia Química, Departamento de Engenharia Química, Escola de Engenharia, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2002. Disponível em: <<https://www.lume.ufrgs.br/bitstream/handle/10183/3719/000342230.pdf?sequence=1>>. Acesso em: 7 mar. 2017.

Instituto Nacional de Metrologia, Qualidade e Tecnologia – INMETRO. **Orientação sobre validação de métodos de ensaios químicos** – QOQCGCRE 088. Revisão 03 – fevereiro/2010. Rio de Janeiro, 2010.

JOHNSON, R. A.; WICHERN, D. W. **Applied Multivariate Statistical Analysis**. 2. ed. New Jersey: Prentice Hall, 1988.

KAMRUZZAMAN, M. et al. Non-destructive prediction and visualization of chemical composition in lamb meat using NIR hyperspectral imaging and multivariate regression. **Innovative Food Science & Emerging Technologies**, v. 16, p. 218-226, 2012.

KENNARD, R.W.; STONE, L. A. Computer-aided design of experiments, **Technometrics**, v. 11, p. 137-148, 1969.

KRÓL, S.; ZABIEGALA, B.; NAMIESNIK, J. Monitoring VOCs in atmospheric air I. On-line gas analyzers. **Trends in Analytical Chemistry**, Amsterdam, v. 29, n. 9, p. 1092-1100, out. 2010.

KUZELJEVIC, Z. V.; DUDUKOVIC, M. P. Computational modeling of trickle bed reactors. **Industrial & Engineering Chemistry Research**, Washington, v. 51, n. 4, p. 1663-1671, 2012.

LIRA, L. D. F. B de et al. Prediction of properties of diesel/biodiesel blends by infrared spectroscopy and multivariate calibration. **Fuel**, v. 89, n. 2, p. 405-409, feb. 2010.

LOTUFO F.A.; GARCIA C. Sensores virtuais ou soft sensors: uma introdução. In: BRAZILIAN CONFERENCE ON DYNAMICS, CONTROL AND APPLICATIONS -

DINCON, 7., Presidente Prudente, 2008. **Anais eletrônicos...** [S.l.: s.n.], 2008. Disponível em: <https://www.researchgate.net/publication/228413086_Sensores_Virtuais_ou_Soft_Sensors_Uma_introducao>. Acesso em: 27 abr. 2017.

MACGREGOR, J. F.; KOURTI, T. Statistical process control of multivariate processes, **Control Engineering Practice**, v. 3, n. 5, 403-414, 1995. Disponível em: <http://files.isec.pt/DOCUMENTOS/SERVICOS/BIBLIO/Documentos%20de%20acesso%20remoto/Statistical_process_control_of_multivariate_processes.pdf>. Acesso em: 2 maio 2017.

MAITRA, S.; YAN, J. Principle component analysis and partial least squares: Two-dimension reduction techniques for regression. **Casualty Actuarial Society**, Arlington, v. 79, p. 79-90, 2008. Disponível em: <<https://www.casact.org/pubs/dpp/dpp08/08dpp76.pdf>>. Acesso em: 8 jun. 2017.

MEVIK, B-H.; WEHRENS, R.; LILAND, K. H. **PLS: Partial Least Squares and Principal Component Regression – R package version 2.6-0**. [S.l.: s.n.], 2016. Disponível em: <<https://CRAN.R-project.org/package=pls>>. Acesso em: 16 ago. 2017.

MOHUNDRO, E. L. Overview on C2 and C3 Selective Hydrogenation in Ethylene Plants. In: ETHYLENE PRODUCERS CONFERENCE, 15., 2003, New Orleans. **Electronic annals...** New Orleans: American Institute of Chemical Engineers, 2003. Disponível em: <<http://www.klmtechgroup.com/PDF/Articles/articles/edm64a.pdf>>. Acesso em: 8 ago. 2017.

MONTGOMERY, D. C.; RUNGER, G. C.; HUBELE, N. F. **Engineering Statistics**. Hoboken: John Wiley & Sons, 2009.

MORELLATO, S. A. **Modelos de Regressão PLS com Erros Heteroscedásticos**. 2010. 49 f. Dissertação (Mestrado em Estatística) – Departamento de Estatística, Universidade Federal de São Carlos, São Carlos, 2010. Disponível em: <<http://www.ime.unicamp.br/sinape/sites/default/files/DissSAM.pdf>>. Acesso em: 12 jun. 2017.

NAES, T. et al. **A user-friendly guide to multivariate calibration and classification**. Chichester: NIR, 2002.

NAGAI, E. Y.; ARRUDA, L. V. R. de. Identificação de modelos inferenciais fuzzy para a estimação de medidas de concentração em processos de refino. In: CONGRESSO BRASILEIRO DE PETRÓLEO E GÁS, 3., 2004, Salvador. **Anais eletrônicos...** [S.l.: s.n.], 2004. Disponível em: <<http://docplayer.com.br/10335683-Identificacao-de-modelos-inferenciais-fuzzy-para-a-estimacao-de-medidas-de-concentracao-em-processos-de-refino.html>>. Acesso em: 27 jul. 2017.

PASQUINI, C.; BUENO, A. F. Characterization of petroleum using near-infrared spectroscopy: quantitative modeling for the true boiling point curve and specific gravity. **Fuel**, v. 86, p. 1927-1934, 2007.

PINTO, A. et al. Monitoramento da produção de etanol celulósico por espectroscopia no infravermelho próximo. In: CONGRESSO BRASILEIRO DE ENGENHARIA QUÍMICA, 20., 2015, Florianópolis. **Anais eletrônicos...** Florianópolis: [s.n.], 2014.

RANADE, V. V.; CHAUDHARI, R. V.; GUNJAL, P. R. **Trickle Bed Reactors**. [S.l.: Elsevier, 2011.

REISS, R.; WOJSZNIS, W.; WOJEWODKA, R. Partial least squares confidence interval calculation for industrial end-of-batch quality prediction. **Chemometrics and Intelligent Laboratory Systems**, v. 100, n. 2, p. 75-82, 2010.

SAMIMI, F. et al. Application of response surface methodology for optimization of an industrial methylacetylene and propadiene hydrogenation reactor. **Journal of the Taiwan Institute of Chemical Engineers**, Taiwan, v. 46, p. 51- 64, 2015.

SEIDL, P. R. et al. **Integração e Refino- petroquímica: tendências e impactos**. Rio de Janeiro: Publit, 2012.

SIE, S. T.; KRISHNA, R. Process development and scale up: III. Scale-up and scale-down of trickle bed processes. **Reviews in Chemical Engineering**, v. 14, n. 3, p. 203-252, 1998.

SILVA, N. C. D. **Uso da espectroscopia NIR no desenvolvimento de um simulador para gasolina e na transferência de calibração entre instrumentos de bancada e portátil**. 2017. Tese (Doutorado em Quimioterapia) – Programa de Pós-Graduação em Química, Universidade Federal de Pernambuco, Recife, 2017.

SOUZA, A. M. D.; POPPI, R. J. Experimento didático de quimiometria para análise exploratória de óleos vegetais comestíveis por espectroscopia no infravermelho médio e análise de componentes principais: um tutorial, parte I. **Química Nova**, São Paulo, v. 35, n. 1, p. 223-229, 2012. Disponível em:
<http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-40422012000100039&lng=en&nrm=iso&tlng=pt>. Acesso em: 4 jul. 2017.

THOMAS, S.; HAIDER, N. S. A Study on Basics of a Gas Analyzer. **International Journal of Advanced Research In Electrical, Electronics And Instrumentation Engineering**, Chennai, v. 2, n. 12, p. 6016-6025, dez. 2013. Disponível em:
<https://www.ijareeie.com/upload/2013/december/19H_A%20Study.pdf>. Acesso em: 15 abr. 2017.

VALDERRAMA, P.; BRAGA, J. W. B.; POPPI, R. J. Estado da arte de figuras de mérito em calibração multivariada. **Química Nova**, São Paulo, v. 32, n. 5, 2009. Disponível em:

<http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-40422009000500034>. Acesso em: 23 jan. 2017.

WANG, B.; FROMENT, G. F. Kinetic Modeling and Simulation of the Selective Hydrogenation of the C₃-Cut of a Thermal Cracking Unit. **Industrial & Engineering Chemistry Research**, v. 44, p. 9860- 9867, 2005.

WEHRENS, R.; MEVIK, B-H. The PLS package: principal component and partial least squares regression in R. *Journal of Statistical of Software*, Innsbruck, v. 18, n. 2, 2007.

WICKHAM, H.. **Tidyverse**: Easily Install and Load “Tidyverse” Packages – R package version 1.1.1. [S.l.: s.n.], 2017.

WOLD, S.; ESBENSEN, K.; GELADI, P. Principal component analysis. **Chemometrics Intelligent Laboratory Systems**, v. 2, n. 1-3, p. 37-52, 1987.

WOLD, S.; SJÖSTRÖM, M.; ERIKSSON, L. PLS-regression: a basic tool of chemometrics. **Chemometrics Intelligent Laboratory Systems**, v. 58, p. 109-130, 2001. Disponível em: <<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.603.4176&rep=rep1&type=pdf>>. Acesso em: 7 mar. 2017.

WU, W. et al. Kinetic Studies and Operating Strategies for an Industrial Selective Hydrogenation Process. **Industrial & Engineering Chemistry Research**, v. 50, n. 3, p.1264-1271, 2011.

WU, W.; LI, Y-L. Selective Hydrogenation of Methylacetylene and Propadiene in an Industrial Process: A Multiobjective Optimization Approach.. **Industrial & Engineering Chemistry Research**, v. 50, n. 3, p. 1453-1459, 2011.

APÊNDICE B – GRÁFICO – RMSEP – LEITO A E LEITO B

É possível observar o RMSEP em função do número de variáveis latentes do modelo PLS nas Figuras 27 e 28. É possível observar que a adição de uma LV ao modelo sempre implica numa queda de RMSEP. Começando da primeira, adiciona-se LVs ao modelo até que a nova adição de uma variável latente implique numa queda de RMSEP menor que 5%. Em ambos os casos, o número de LVs escolhidas é de 5.

Figura 27 - RMSEP em função do número de LVs - Leito A

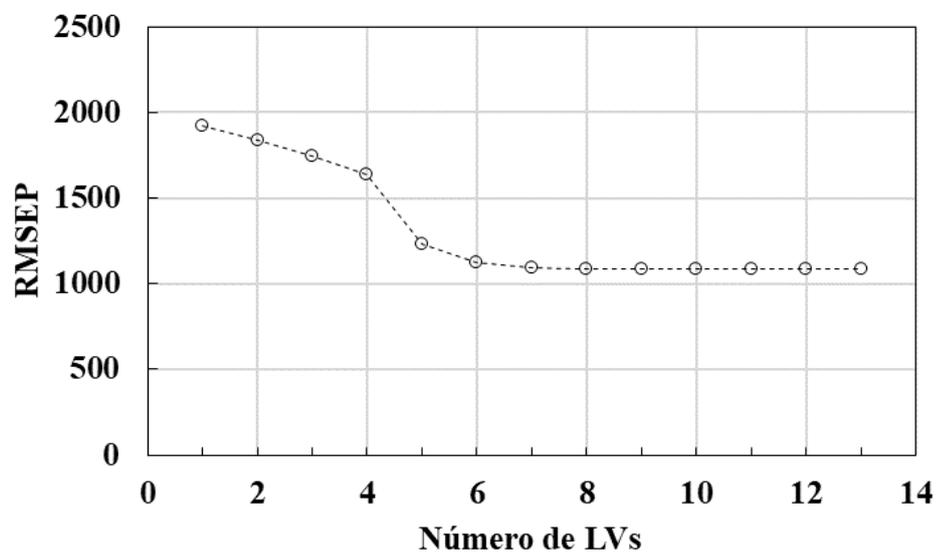
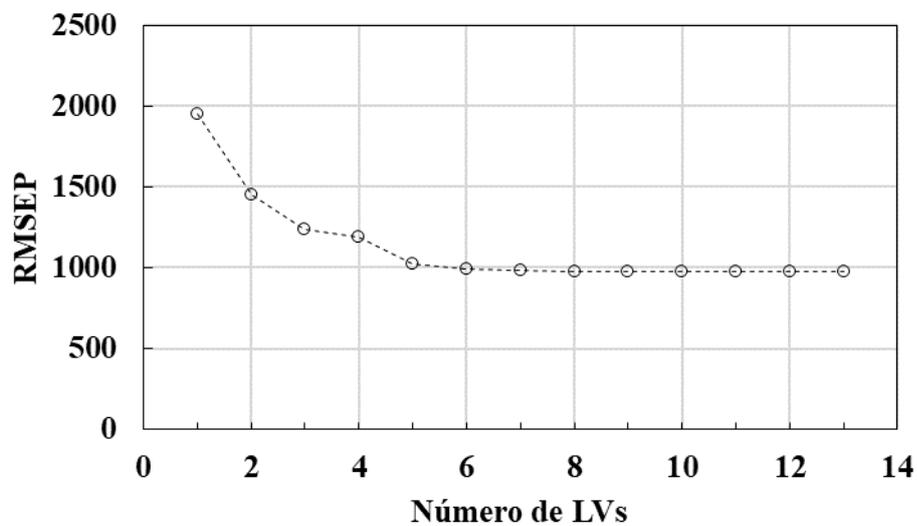


Figura 28 - RMSEP em função do número de LVs - Leito B



UFBA
UNIVERSIDADE FEDERAL DA BAHIA
ESCOLA POLITÉCNICA

PROGRAMA DE PÓS GRADUAÇÃO EM ENGENHARIA INDUSTRIAL - PEI

Rua Aristides Novis, 02, 6º andar, Federação, Salvador BA

CEP: 40.210-630

Telefone: (71) 3283-9800

E-mail: pei@ufba.br

Home page: <http://www.pei.ufba.br>

