

UNIVERSIDADE FEDERAL DE MINAS GERAIS

ESCOLA DE ENGENHARIA

DEPARTAMENTO DE ENGENHARIA QUÍMICA

CURSO DE PÓS-GRADUAÇÃO EM ENGENHARIA QUÍMICA

LEONARDO CATÃO DE OLIVEIRA

DESENVOLVIMENTO DE ANALISADORES
VIRTUAIS E SUA APLICAÇÃO NA PREDIÇÃO DO
PONTO FINAL DE EBULIÇÃO DA NAFTA DE
CRAQUEAMENTO

BELO HORIZONTE, MINAS GERAIS

OUTUBRO DE 2010

LEONARDO CATÃO DE OLIVEIRA

DESENVOLVIMENTO DE ANALISADORES
VIRTUAIS E SUA APLICAÇÃO NA PREDIÇÃO DO
PONTO FINAL DE EBULIÇÃO DA NAFTA DE
CRAQUEAMENTO

Dissertação de mestrado apresentada ao departamento de engenharia química como parte dos requisitos exigidos para obtenção do título de MESTRE EM ENGENHARIA QUÍMICA.

ORIENTADOR:

PROF. DR. MARCELO CARDOSO

BELO HORIZONTE, MINAS GERAIS

OUTUBRO DE 2010

AGRADECIMENTOS

Agradeço a Deus, por tornar tudo possível.

Agradeço aos meus pais pelo carinho, exemplo de vida e pelo incentivo constante aos meus estudos.

Agradeço à minha esposa pelo amor, apoio e compreensão nas horas mais turbulentas.

Aos meus colegas de trabalho pelo incentivo, discussões e orientações constantes.

Ao amigo Marcelo Cardoso pela enorme paciência e boa vontade.

Conta-me o teu passado e saberei o teu futuro

Confúcio

RESUMO

Analisadores virtuais são empregados com sucesso em aplicações de controle e otimização. Representam uma alternativa eficiente e de baixo custo para o monitoramento online de indicadores de qualidade de correntes de processo. Este trabalho buscou desenvolver um conjunto de rotinas computacionais contendo técnicas matemáticas e estatísticas para facilitar a criação destes analisadores. O programa desenvolvido apresenta uma estrutura modular, o que permite ao usuário habilitar independentemente as ferramentas de pré-tratamento de dados, seleção de variáveis e modelagem matemática de sistemas. A ferramenta foi utilizada para a predição do ponto final de ebulição (PFE) da nafta 'leve' de uma unidade de craqueamento catalítico fluidizado da Petrobrás, adotando-se a premissa de obtenção do modelo mais simples, contendo o menor número de regressores possível. O sistema em questão foi especialmente escolhido para o estudo em função da suspeita de problemas hidráulicos na torre fracionadora principal, o que levou à hipótese de impossibilidade de obtenção de um modelo que capture o PFE nestas condições. Pelo amplo espectro de faixa de operação, o modelo obtido não apresentou variabilidade inferior à reprodutibilidade do ensaio de laboratório. No entanto, o modelo consegue sinalizar condições favoráveis à perda de especificação do PFE. Foi possível verificar, ainda, que o PFE é influenciado por três fatores distintos: temperatura do topo da torre fracionadora principal, cuja influência se dá no equilíbrio termodinâmico desta seção; temperatura de reação, cuja influência se dá no maior ou menor rendimento de hidrocarbonetos mais 'leves'; diferencial de pressão no topo da torre fracionadora, cujo comportamento pode sinalizar condições hidráulicas desfavoráveis para o controle do PFE. Dentre os modelos obtidos, destacam-se as redes neurais tipo *Perceptron* de múltiplas camadas, com função de performance dada por regularização Bayesiana, em função de seu maior poder de generalização.

ABSTRACT

Soft sensors can be successfully used in control and optimization applications. They can be seen as an efficient low cost alternative for online monitoring of quality indicators of process streams. The goal of this work is to develop a set of computational routines based on mathematical and statistical techniques in order to improve the building of these soft sensors. The application here developed presents a modular structure which allows the user to access independent tools for data pre processing, variable selection and mathematical system modeling. The application was used to predict final boiling point (FBP) of light naphtha in a fluid catalytic cracking unit of Petrobrás refinery. We looked for the simplest model with the smallest number of variables. This case study was designed to test the difficulty of prediction of the FBP in case of hydraulic problems in the main fractional distillation column. Because of the large range of modeled operations performed by the soft sensor it was not possible to reach values close to the reproducibility obtained in laboratory tests. On the other hand the model is capable of signaling the presence of favorable conditions for the loss of specification of FBP. Also, it was possible to perceive that the FBP is sensible to three variables: main column top temperature, which influences the thermodynamic equilibrium of this section; raiser top temperature, which influences the amount of light hydrocarbons production; and the differential pressure in the top section of the main fractional distillation column, which signals hydraulically unfavorable conditions for controlling the FBP. Amongst the models that were obtained the multilayer *Perceptron* with Bayesian regulation stands out due to its more powerful generalization capabilities.

LISTA DE FIGURAS

Figura 2.1: Interação da unidade de craqueamento catalítico com outros processos de refino (adaptado de Mattos (2004, p.4)).	5
Figura 2.2: Esquema simplificado de uma unidade de craqueamento catalítico em leito fluidizado (adaptado de Mattos (2004, p.8)).	6
Figura 2.3: Diagrama esquemático do conversor e soprador de uma unidade de craqueamento catalítico (adaptado de Mattos (2004, p.9)).	8
Figura 2.4: Fluxograma esquemático das principais correntes de processo referentes à seção de fracionamento de uma unidade de craqueamento (adaptado de Mattos (2004, p.11)).	10
Figura 2.5: Fluxograma esquemático exibindo os principais equipamentos da seção de recuperação de gases (adaptado de Mattos (2004, p.13)).	11
Figura 2.6: Aparato manual para ensaio de destilação (adaptado da norma ASTM D86).	16
Figura 2.7: Principais tipos de pratos utilizados em torres de separação: a) prato com borbulhador; b) prato perfurado; c) prato valvulado (Site da Koch-Glitsch).	20
Figura 2.8: Principais tipos de recheios utilizados em torres de separação: a) recheio randômico tipo IMTP [®] ; b) recheio estruturado de alta eficiência INTALOX [®] (Site da Koch-Glitsch).	21
Figura 2.9: Distribuidor de líquido tipo chapa-perfurada e chaminé (Site da Koch-Glitsch).	21
Figura 2.10: Representação esquemática dos fluidos presentes em uma torre de pratos.	22
Figura 2.11: Condições hidráulicas de operação para uma torre de pratos (adaptado de Caldas et al. (2007, p.76)).	23
Figura 2.12: Diagrama esquemático do processo de compressão no PI: a) pontos reportados por exceção; b) paralelogramo gerado com o último ponto reportado por exceção; c) penúltimo ponto armazenado.	29
Figura 2.13: Dendograma exemplificando o histórico de agrupamentos em um algoritmo hierárquico	38
Figura 2.14: Interpretação geométrica dos <i>scores</i> .	46
Figura 2.15: Variáveis de processo correlacionadas (adaptado de Feital (2009, p.19)).	47

Figura 2.16: Variáveis de processo representadas nos eixos originais e nos novos eixos de coordenadas (adaptado de Feital (2009, p.19)).	48
Figura 2.17: Reconstrução de um valor a partir da sua projeção na componente mais significativa (adaptado de Feital (2009, p.21)).	48
Figura 2.18: Interpretação geométrica da técnica de PLS (adaptado de Wold et al., (2004, p.44)).	51
Figura 2.19: Etapas sucessivas do algoritmo NIPALS.	53
Figura 2.20: Neurônio MCP (BRAGA; CARVALHO; LUDERMIR, 2007, p.8).	55
Figura 2.21: Principais funções de ativação utilizadas. a) função degrau; b) função logaritmo sigmoide; c) função linear; d) função gaussiana (adaptado de Braga; Carvalho; Ludermir, (2007, p.10)).	58
Figura 2.22: Diagrama esquemático de aprendizado supervisionado (BRAGA; CARVALHO; LUDERMIR, 2007, p.13).	59
Figura 2.23: Diagrama esquemático de aprendizado não supervisionado (BRAGA; CARVALHO; LUDERMIR 2007, p.17).	59
Figura 2.24: Rede feedforward com uma camada oculta.	60
Figura 2.25: Rede feedforward com uma camada oculta, apresentando recorrência entre a camada de saída e a camada oculta.	60
Figura 2.26: Rede neural autoassociativa.	61
Figura 2.27: Diagrama esquemático do algoritmo back-propagation (adaptado de Braga; Carvalho; Ludermir, 2007, p.76).	64
Figura 3.1: Interface gráfica para utilização do INFERATOR.	76
Figura 3.2: Fluxograma esquematizando as etapas de cálculo do programa INFERATOR.	77
Figura 3.3: Diagrama esquemático da localização da instrumentação do conversor.	80
Figura 3.4: Diagrama esquemático da localização da instrumentação da fracionadora.	80
Figura 3.5: Variável considerada adequada para utilização do identificador de Hampel.	82
Figura 3.6: Variável inadequada para estabelecimento de <i>outliers</i> pelo identificador de Hampel.	83
Figura 3.7: Gráfico de correlação cruzada do diferencial de pressão entre a panela de LCO e o topo da fracionadora (variável 23) em relação ao PFE (variável 35).	85
Figura 3.8: Representação gráfica da classificação obtida pelo método das k-médias.	89

Figura 3.9: Histograma de distribuição ilustrativo característico da curva W para a variável 23.	92
Figura 3.10: Fluxograma esquematizando as principais operações realizadas pelo programa na etapa de modelagem.	99
Figura 4.1: Limites de corte obtidos para as variáveis normalizadas selecionadas para aplicação do identificador de Hampel; a) pressão de topo da fracionadora principal; b) temperatura na base do riser; c) temperatura da panela de retirada de nafta ‘pesada’; d) temperatura do refluxo circulante de HCO; e) diferencial de pressão entre as panelas de LCO e HCO.	105
Figura 4.2: Gráfico de correlação cruzada entre o PFE e a temperatura de topo da fracionadora principal.	107
Figura 4.3: Gráfico de correlação cruzada entre o PFE e a temperatura do refluxo de topo da fracionadora principal.	107
Figura 4.4: Gráfico de correlação cruzada entre o PFE e a vazão de nafta ‘leve’ produzida.	108
Figura 4.5: Gráfico de correlação cruzada entre o PFE e a vazão de carga A de gasóleo para a unidade.	109
Figura 4.6: Gráfico de correlação cruzada entre o PFE e a vazão de carga B de gasóleo para a unidade.	109
Figura 4.7: Gráfico de correlação cruzada entre o PFE e a temperatura de reação no topo do riser.	110
Figura 4.8: Gráfico de correlação cruzada entre o PFE e o diferencial de pressão topo/panela de LCO.	111
Figura 4.9: Gráfico de correlação cruzada entre o PFE e o diferencial de pressão panela de LCO/panela de HCO.	111
Figura 4.10: Gráfico de correlação cruzada entre o PFE e o diferencial de pressão topo/fundo.	112
Figura 4.11: Gráfico de correlação cruzada entre o PFE e a retirada de nafta ‘pesada’.	112
Figura 4.12: Variáveis selecionadas pela regressão linear por etapas, para nível de significância de 0,05 e 0,01.	115
Figura 4.13: Variáveis selecionadas pela regressão linear por etapas, para nível de significância de 0,01.	116

Figura 4.14: Ordem preferencial de seleção das variáveis pela rotina de regressão linear por etapas.....	117
Figura 4.15: Clusters formados considerando-se as variáveis PFE e diferencial de pressão topo/panela de LCO no vetor das k-médias.....	119
Figura 4.16: Clusters formados considerando-se as variáveis PFE, diferencial de pressão topo/panela de LCO e temperatura do refluxo de topo no vetor das k-médias.	120
Figura 4.17: Clusters formados considerando-se as variáveis PFE, diferencial de pressão topo/panela de LCO, temperatura do refluxo de topo e temperatura de reação no vetor das k-médias.	121
Figura 4.18: Clusters formados considerando-se as variáveis PFE, diferencial de pressão topo/panela de LCO, temperatura do refluxo de topo, temperatura de reação e a temperatura de topo no vetor das k-médias.	122
Figura 4.19: Clusters formados considerando-se as variáveis PFE e diferencial de pressão topo/panela de LCO no vetor das k-médias.....	123
Figura 4.20: Clusters formados considerando-se as variáveis PFE, diferencial de pressão topo/panela de LCO e temperatura de reação no vetor das k-médias.	124
Figura 4.21: Clusters formados considerando-se as variáveis PFE, diferencial de pressão topo/panela de LCO, temperatura de reação e temperatura de topo no vetor das k-médias.....	125
Figura 4.22: Clusters formados considerando-se as variáveis PFE, diferencial de pressão topo/panela de LCO, temperatura de reação, temperatura de topo e temperatura do refluxo de topo no vetor das k-médias.....	125
Figura 4.23: Histograma utilizado para o cálculo do <i>LSC</i> da variável temperatura de topo da torre fracionadora principal (variável 3).....	129
Figura 4.24: Histograma utilizado para o cálculo do <i>LSC</i> da variável temperatura do refluxo de topo para a torre fracionadora principal (variável 4).....	129
Figura 4.25: Histograma utilizado para o cálculo do <i>LSC</i> da variável temperatura de reação (variável 21).	130
Figura 4.26: Histograma utilizado para o cálculo do <i>LSC</i> da variável diferencial de pressão topo/panela de LCO (variável 23).	130
Figura 4.27: Efeito do acréscimo de componentes principais na variável ‘incerteza’.	131
Figura 4.28: Efeito do acréscimo de componentes principais no coeficiente de correlação entre os dados reais e os dados preditos.....	131

Figura 4.29: Efeito do acréscimo de componentes principais na soma quadrática dos erros.....	132
Figura 4.30: Efeito do acréscimo de componentes principais na soma quadrática da derivada do erro em relação ao tempo.....	132
Figura 4.31: Efeito do acréscimo de variáveis latentes na variável ‘incerteza’.....	133
Figura 4.32: Efeito do acréscimo de variáveis latentes no coeficiente de correlação entre os dados reais e os dados preditos.....	133
Figura 4.33: Efeito do acréscimo de variáveis latentes na soma quadrática dos erros.....	134
Figura 4.34: Efeito do acréscimo de variáveis latentes na soma quadrática da derivada do erro em relação ao tempo.....	134
Figura 4.35: Efeito do acréscimo de neurônios à camada oculta na variável ‘incerteza’ para o caso de função de performance dada por regularização Bayesiana.....	135
Figura 4.36: Efeito do acréscimo de neurônios à camada oculta no coeficiente de correlação entre os dados reais e os dados preditos para o caso de função de performance dada por regularização Bayesiana.....	135
Figura 4.37: Efeito do acréscimo de neurônios à camada oculta na soma quadrática dos erros para o caso de função de performance dada por regularização Bayesiana.....	136
Figura 4.38: Efeito do acréscimo de neurônios à camada oculta na soma quadrática da derivada do erro em relação ao tempo para o caso de função de performance dada por regularização Bayesiana.....	136
Figura 4.39: Efeito do acréscimo de neurônios à camada oculta na variável número de parâmetros efetivos.....	137
Figura 4.40: Efeito do acréscimo de neurônios à camada oculta na variável ‘incerteza’, utilizando-se um máximo de vinte nós para o caso de função de performance dada por regularização Bayesiana.....	137
Figura 4.41: Efeito do acréscimo de neurônios à camada oculta na variável ‘incerteza’ para o caso do algoritmo de treinamento back-propagation resiliente.....	138
Figura 4.42: Efeito do acréscimo de neurônios à camada oculta no coeficiente de correlação para o caso do algoritmo de treinamento back-propagation resiliente.....	138
Figura 4.43: Efeito do acréscimo de neurônios à camada oculta na soma quadrática dos erros para o caso do algoritmo de treinamento back-propagation resiliente.....	139
Figura 4.44: Efeito do acréscimo de neurônios à camada oculta na soma quadrática da derivada do erro em relação ao tempo para o caso do algoritmo back-propagation resiliente.....	139

Figura 4.45: Efeito do acréscimo de neurônios à segunda camada, conforme legenda, e à primeira camada, conforme abscissa, na variável ‘incerteza’, para dados caracterizados como ‘transientes’	140
Figura 4.46: Efeito do acréscimo de neurônios à segunda camada, conforme legenda, e à primeira camada, conforme abscissa, na variável coeficiente de correlação, para dados caracterizados como ‘transientes’	141
Figura 4.47: Efeito do acréscimo de neurônios à segunda camada, conforme legenda, e à primeira camada, conforme abscissa, na variável média da soma quadrática dos erros, para dados caracterizados como ‘transientes’	141
Figura 4.48: Efeito do acréscimo de neurônios à segunda camada, conforme legenda, e à primeira camada, conforme abscissa, na variável média da soma quadrática da derivada do erro em relação ao tempo, para dados caracterizados como ‘transientes’	142
Figura 4.49: Efeito do acréscimo de neurônios à segunda camada, conforme legenda, e à primeira camada, conforme abscissa, na variável ‘incerteza’, para dados caracterizados como ‘estacionários’	142
Figura 4.50: Efeito do acréscimo de neurônios à segunda camada, conforme legenda, e à primeira camada, conforme abscissa, na variável coeficiente de correlação, para dados caracterizados como ‘estacionários’	143
Figura 4.51: Efeito do acréscimo de neurônios à segunda camada, conforme legenda, e à primeira camada, conforme abscissa, na variável média da soma quadrática dos erros, para dados caracterizados como ‘estacionários’	143
Figura 4.52: Efeito do acréscimo de neurônios à segunda camada, conforme legenda, e à primeira camada, conforme abscissa, na variável média da soma quadrática da derivada do erro em relação ao tempo, para dados caracterizados como ‘estacionários’	144
Figura 4.53: Efeito da modificação do parâmetro de espalhamento, conforme legenda, e do incremento do número de nós na camada oculta, conforme abscissa, na variável ‘incerteza’, para dados caracterizados como ‘transientes’	145
Figura 4.54: Efeito da modificação do parâmetro de espalhamento, conforme legenda, e do incremento do número de nós na camada oculta, conforme abscissa, na variável coeficiente de correlação, para dados caracterizados como ‘transientes’	145
Figura 4.55: Efeito da modificação do parâmetro de espalhamento, conforme legenda, e do incremento do número de nós na camada oculta, conforme abscissa, na variável média da soma quadrática dos erros, para dados caracterizados como ‘transientes’	146

Figura 4.56: Efeito da modificação do parâmetro de espalhamento, conforme legenda, e do incremento do número de nós na camada oculta, conforme abscissa, na variável média da soma quadrática da derivada do erro em relação ao tempo, para dados caracterizados como ‘transientes’.....	146
Figura 4.57: Efeito da modificação do parâmetro de espalhamento, conforme legenda, e do incremento do número de nós na camada oculta, conforme abscissa, na variável ‘incerteza’, para dados caracterizados como ‘estacionários’.....	147
Figura 4.58: Efeito da modificação do parâmetro de espalhamento, conforme legenda, e do incremento do número de nós na camada oculta, conforme abscissa, na variável coeficiente de correlação, para dados caracterizados como ‘estacionários’.....	147
Figura 4.59: Efeito da modificação do parâmetro de espalhamento, conforme legenda, e do incremento do número de nós na camada oculta, conforme abscissa, na variável média da soma quadrática dos erros, para dados caracterizados como ‘estacionários’.	148
Figura 4.60: Efeito da modificação do parâmetro de espalhamento, conforme legenda, e do incremento do número de nós na camada oculta, conforme abscissa, na variável média da soma quadrática da derivada do erro em relação ao tempo, para dados caracterizados como ‘estacionários’.....	148
Figura 4.61: Comparação da variável ‘incerteza’ para as diferentes técnicas de modelagem.....	149
Figura 4.62: Comparação da variável ‘coeficiente de correlação’ para as diferentes técnicas de modelagem.....	149
Figura 4.63: Comparação da variável ‘soma quadrática dos erros’ para as diferentes técnicas de modelagem.....	150
Figura 4.64: Comparação da variável ‘soma quadrática da derivada do erro em relação ao tempo’ para as diferentes técnicas de modelagem.....	150
Figura 4.65: Modelo linear (dados estacionários) com correlação igual a 0.64547 e variável ‘incerteza’ igual a 22.0°C.....	151
Figura 4.66: Modelo de PCR (dados estacionários) com variabilidade explicada de 100%, com correlação igual a 0.67251 e ‘incerteza’ de 20.9°C.....	151
Figura 4.67: Modelo de PLS (dados estacionários) com 2 componentes, com correlação igual a 0.64647 e ‘incerteza’ de 21.7°C.....	152
Figura 4.68: RNA regularização Bayesiana (dados estacionários), uma camada oculta, com 2 neurônios, com correlação igual a 0.81534 e ‘incerteza’ de 15.8°C.....	152

Figura 4.69: RNA back-propagation (dados estacionários), uma camada oculta, com 4 neurônios, com correlação igual a 0.71374 e ‘incerteza’ de 20.0°C.....	153
Figura 4.70: RNA back-propagation (dados estacionários), duas camadas ocultas, com 4 neurônios na 1ª, 2 neurônios na 2ª, com correlação igual a 0.75329 e ‘incerteza’ de 18.0°C.	153
Figura 4.71: RBF (dados estacionários), camada de saída linear, com 6 neurônios, com correlação igual a 0.80755, parâmetro de espalhamento 10 e ‘incerteza’ de 16.8 °C .	154
Figura 4.72: Modelo linear (dados estacionários) com correlação igual a 0.67593 e ‘incerteza’ de 23.3°C.	154
Figura 4.73: Modelo de PCR (dados estacionários) com variabilidade explicada de 100%, com correlação igual a 0.68094 e ‘incerteza’ de 23.1 °C.	155
Figura 4.74: Modelo de PLS (dados estacionários) com 2 componentes, com correlação igual a 0.63902 e ‘incerteza’ de 24.3°C.....	155
Figura 4.75: RNA regularização Bayesiana (dados estacionários), uma camada oculta, com 2 neurônios, com correlação igual a 0.78088 e ‘incerteza’ de 19.7°C.....	156
Figura 4.76: RNA back-propagation (dados estacionários), uma camada oculta, com 4 neurônios, com correlação igual a 0.58154 e ‘incerteza’ de 26.1°C.....	156
Figura 4.77: RNA back-propagation (dados estacionários), duas camadas ocultas, com 4 neurônios na 1ª, 2 neurônios na 2ª, com correlação igual a 0.69942 e ‘incerteza’ de 18.5°C.	157
Figura 4.78: RBF (dados estacionários), camada de saída linear, com 6 neurônios, com correlação igual a 0.78631, parâmetro de espalhamento 10 e ‘incerteza’ de 20°C.	157
Figura 4.79: Gráfico de probabilidade normal dos resíduos, referente ao modelo de redes neurais MLP (regularização Bayesiana), considerando-se a métrica da soma quadrática dos erros.	159
Figura 4.80: Gráfico de probabilidade normal dos resíduos, referente ao modelo de redes neurais MLP (regularização Bayesiana), considerando-se a métrica da soma quadrática da derivada do erro em relação ao tempo.	159

LISTA DE TABELAS

Tabela 2.1: Fatores para a construção dos gráficos de controle para amplitude.	35
Tabela 3.1: Relação das variáveis coletadas inicialmente.	78
Tabela 3.2: Variáveis selecionadas para aplicação do identificador de Hampel.	83
Tabela 4.1: Disponibilidade de dados após operação de filtragem para perda de sinal.	101
Tabela 4.2: Número de falhas de sinal identificadas para os medidores das variáveis coletadas.	102
Tabela 4.3: Disponibilidade de dados após aplicação do identificador de Hampel.	105
Tabela 4.4: Estimativa para o atraso de transporte.	106
Tabela 4.5: Redução do número de dados disponíveis ocasionada pela montagem da matriz das variáveis candidatas à regressoras.	114
Tabela 4.6: Dados numéricos referentes ao arquivo de configuração para regressão linear por etapas.	114
Tabela 4.7: Variáveis selecionadas para novo procedimento de regressão linear por etapas.	116
Tabela 4.8: Variáveis selecionadas para os procedimentos posteriores de análise de clusters e modelagem matemática de sistemas.	117
Tabela 4.9: Primeira etapa de inclusão de variáveis no vetor de k-médias, para dois clusters a priori.	118
Tabela 4.10: Segunda etapa de inclusão de variáveis no vetor de k-médias, para dois clusters a priori.	119
Tabela 4.11: Terceira etapa de inclusão de variáveis no vetor de k-médias, para dois clusters a priori.	120
Tabela 4.12: Quarta etapa de inclusão de variáveis no vetor de k-médias, para dois clusters a priori.	121
Tabela 4.13: Centroides dos clusters referentes à Figura 4.15.	122
Tabela 4.14: Primeira etapa de inclusão de variáveis no vetor de k-médias, para três clusters a priori.	123
Tabela 4.15: Segunda etapa de inclusão de variáveis no vetor de k-médias, para três clusters a priori.	123
Tabela 4.16: Terceira etapa de inclusão de variáveis no vetor de k-médias, para três clusters a priori.	124

Tabela 4.17: Quarta etapa de inclusão de variáveis no vetor de k-médias, para três clusters a priori.	124
Tabela 4.18: Centroides dos clusters referentes à Figura 4.21.	126
Tabela 4.19: Valores obtidos para a média d_2 da distribuição W	127
Tabela 4.20: Resumo dos fatores para a construção dos gráficos de controle para amplitude.	127
Tabela 4.21: Resumo dos fatores para a determinação do desvio padrão d_3	128

LISTA DE SÍMBOLOS

Variável	Descrição
A_{pp}	Matriz de autovalores obtida pela decomposição espectral da matriz de covariâncias;
ANP	Agência Nacional do Petróleo, Gás Natural e Biocombustíveis;
C_i	<i>Cluster</i> de dados i ;
\bar{C}_i	Centroide do i -ésimo <i>cluster</i> ;
c_{ik}	<i>Load</i> ou direção da k -ésima componente principal;
\mathbf{c}_k	Autovetor k obtido pela decomposição espectral da matriz de covariâncias;
C_{pp}	Matriz de autovetores obtida pela decomposição espectral da matriz de covariâncias;
CCF	Craqueamento catalítico em leito fluidizado;
\mathbf{D}^*	Matriz diagonal advinda da decomposição em valores singulares;
\mathbf{D}^{**}	Matriz diagonal simplificada pela consideração de k componentes principais;
\mathbf{D}	Conjunto dos dados de entrada e saída;
\mathcal{D}	Operador diferencial linear;
\mathbf{D}_{PLS}	Matriz diagonal do modelo de variáveis latentes;
d_2	Média da distribuição W ;
d_3	Desvio padrão da distribuição W ;
$d(C_i, C_j)$	Métrica de distância entre o i -ésimo <i>cluster</i> e o j -ésimo <i>cluster</i> ;
$d(\mathbf{x}_i, \mathbf{x}_k)$	Distância entre o vetor coluna da i -ésima variável regressora em relação ao vetor coluna da k -ésima variável regressora;
d_{kk}	Elemento k da matriz diagonal;
D_t	Vazão de produto de topo (m ³ /d, kmol/h, kg/h);
E	Função erro quadrático;
$E(X)$	Esperança matemática da variável X ;
f	Limiar de aceitação para teste parcial de Fischer;
F	Estatística do teste parcial de Fischer;

$F(\mathbf{x}_j, \mathbf{w})$	Função que descreve o sistema físico em estudo;
$F_r(x)$	Função de distribuição de probabilidade para a r -ésima amostra ordenada;
$f(u)$	Função de ativação para modelo de redes neurais;
g^*	Número de <i>clusters</i> existentes na amostra;
GLP	Gás Liquefeito de Petróleo;
\mathbf{H}	Matriz hessiana;
\mathbf{H}_p	Matriz de ponderação para o cálculo de distâncias;
h_{ii}	<i>Leverage</i> referente à amostra de dados i ;
h_i	Resultado do cálculo da função de ativação da camada escondida para o neurônio i ;
\mathbf{I}	Matriz identidade;
k	Número de variáveis na matriz de dados (i.e número de colunas);
L	Somatório do quadrado dos desvios para modelo linear;
LCO	Óleo leve de reciclo;
LIC	Limite inferior de controle para detecção de <i>outliers</i> ;
LSC	Limite superior de controle para detecção de <i>outliers</i> ;
\mathcal{M}^B	Modelo específico de arquitetura de rede neural;
MAD_k	Mediana dos dados gerados pelo desvio absoluto entre os valores amostrados e a mediana dos valores amostrados;
MCP	Modelo de neurônio McCulloch e Pitts;
MLP	Redes neurais tipo <i>Perceptron</i> de múltiplas camadas;
MLR	Técnica de modelagem que utiliza regressão linear multivariada;
MPC	Controlador preditivo baseado em modelos;
MQ_E	Média quadrática do erro;
MQ_R	Média quadrática da regressão;
n	Número de amostras na matriz de dados (i.e número de linhas);
N_B	Número total de parâmetros livres da rede;
n_B	Produto do número de neurônios na camada de saída pelo número de amostras no conjunto de dados;
n_i	Número de amostras presentes no i -ésimo <i>cluster</i> ;

P	Matriz de <i>loads</i> para o modelo de variáveis latentes;
\mathbf{P}_{pp}^*	Matriz contendo vetores de norma unitária, advinda da decomposição em valores singulares;
\mathbf{P}_{pk}^{**}	Matriz contendo vetores de norma unitária, reduzida pela utilização de k componentes, para modelo de regressão de componentes principais;
$P(x y)$	Probabilidade de ocorrência do evento x dada ocorrência do evento y ;
PCR	Técnica de modelagem que utiliza regressão de componentes principais;
PFE	Ponto final de ebulição;
PI	Conjunto de módulos de software servidor/cliente para monitoramento e análise de plantas de processo;
PI-UDS	Servidor de dados PI Universal Data Server;
PLS	Técnica de modelagem que utiliza mínimos quadrados parciais;
Q	Matriz de <i>loads</i> para o modelo de variáveis latentes;
q	Vetor de <i>loads</i> do algoritmo <i>NIPALS</i> ;
R	Amplitude, diferença entre o maior e o menor valor amostrado para uma dada variável aleatória;
\mathcal{R}^B	Termo de regularização específico;
r	Parâmetro de espalhamento da função gaussiana;
R^2	Coefficiente de determinação múltipla;
R_f	Razão de refluxo (adimensional);
\bar{R}	Amplitude média observável da variável regressora, considerando-se m amostras;
R_t	Vazão de refluxo de topo (m^3/d , kmol/h , kg/h);
$\hat{r}_{xy}(k)$	Valor da função de correlação cruzada tendo em vista k deslocamentos no tempo;
$r(\mathbf{z}, \mathbf{w})$	Coefficiente de correlação de Pearson entre os vetores genéricos \mathbf{z} e \mathbf{w} ;
RBF	Redes neurais de base radial;
RNA	Redes neurais artificiais;

RNA1 BP	Redes neurais tipo <i>Perceptron</i> de múltiplas camadas, com uma camada oculta e algoritmo de resolução tipo <i>back-propagation</i> resiliente;
RNA1 RB	Redes neurais tipo <i>Perceptron</i> de múltiplas camadas, com uma camada oculta, modelagem por regularização bayesiana e algoritmo de resolução tipo Levenberg-Marquardt;
RNA2 BP	Redes neurais tipo <i>Perceptron</i> de múltiplas camadas, com duas camadas ocultas e algoritmo de resolução tipo <i>back-propagation</i> resiliente;
s_{ij}	Covariância entre a i -ésima variável regressora e a j -ésima variável regressora;
\mathbf{S}_{pp}	Matriz de covariância para um total de p variáveis regressoras;
SQ_E	Soma quadrática dos erros;
SQ_R	Soma quadrática da regressão;
SQ_T	Soma quadrática total corrigida;
S_r	Variável “espaçamento”, diferença entre duas amostras de uma variável aleatória ordenada;
t	Critério de corte para detecção de <i>outliers</i> (adimensional);
t_a	Instante de coleta das amostras de nafta para análise de laboratório;
\mathbf{T}	Matriz de <i>scores</i> ;
\mathbf{T}^*	Matriz de <i>scores</i> obtida pela decomposição em valores singulares da matriz de dados;
\mathbf{T}^{**}	Matriz de <i>scores</i> reduzida a partir da utilização de k componentes para modelo de regressão de componentes principais;
\mathbf{t}_i	Vetor coluna de <i>scores</i> referentes a i -ésima componente principal;
t_{ik}	<i>Score</i> (projeção) da k -ésima componente considerando-se o i -ésimo instante de amostragem;
\mathbf{U}	Matriz de <i>scores</i> do modelo de variáveis latentes;
\mathbf{U}_{nm}^*	Matriz com norma unitária advinda da decomposição em valores singulares para modelo de regressão de componentes principais;
\mathbf{U}_{nk}^{**}	Matriz reduzida pela utilização de k componentes para modelo de regressão de componentes principais;

u	Valor aplicado na função de ativação para cálculo por redes neurais;
V	Volume de torres e vasos;
\mathbf{W}	Matriz de pesos do algoritmo <i>NIPALS</i> ;
\mathcal{W}	Distribuição obtida pela divisão das amplitudes das amostras pelo desvio padrão populacional;
\mathbf{w}	Vetor de pesos para cálculo de redes neurais;
$\overline{\mathbf{w}}$	Média dos dados do vetor coluna genérico \mathbf{w} empregado na análise de correlação;
w_i	Peso para ponderação das variáveis no cálculo de distância Minkowsky entre vetores;
w_{ij}	Pesos aplicados no cálculo da saída u utilizada nas funções de ativação;
\mathbf{X}	Matriz de dados das variáveis regressoras;
\mathbf{X}_{MLR}	Matriz de dados das variáveis regressoras acrescida de uma coluna unitária;
X_n	n -ésima amostra de uma variável aleatória genérica;
$X_{(n)}$	n -ésima amostra ordenada de uma variável aleatória genérica;
x_{ik}	Valor da k -ésima variável no i -ésimo instante de tempo;
\mathbf{x}_k	Vetor coluna contendo os dados da k -ésima variável regressora;
$\overline{\mathbf{x}}_k$	Média amostral dos dados coletados para a k -ésima variável;
$\left(\overline{\mathbf{x}}_k\right)_m$	Média amostral dos dados da k -ésima variável coletados em janela temporal entre os instantes m e $(l+1)$;
$\left(\overline{\mathbf{x}}\right)^T$	Vetor linha obtido pela média dos valores dos vetores coluna das k variáveis regressoras;
\mathbf{x}_i^T	Vetor linha contendo os valores das k -ésimas variáveis regressoras amostrados no instante i ;
\mathbf{y}	Vetor coluna da variável resposta;
$\hat{\mathbf{y}}$	Vetor coluna estimado para a variável resposta a partir da aplicação dos dados de entrada em um modelo matemático;
\mathbf{Y}	Matriz de dados das variáveis de saída;
y_i	Valor da variável resposta referente ao instante i ;

$y_j^D(n)$	Saída desejada obtida do conjunto de dados de treinamento;
$\bar{\mathbf{z}}$	Média dos dados do vetor coluna genérico \mathbf{z} empregado na análise de correlação

Letras gregas

Variável	Descrição
α	Termo <i>momentum</i> para modificação de convergência;
α^B	Constante da modelagem por regularização bayesiana que multiplica o termo de regularização;
β^B	Constante da modelagem por regularização bayesiana que multiplica o termo do erro quadrático;
β_i	Coefficientes do modelo de regressão linear;
$\hat{\boldsymbol{\beta}}_{MLR}$	Vetor coluna de coeficientes do modelo de regressão linear;
$\hat{\boldsymbol{\beta}}_{PCR}$	Vetor de coeficientes ajustado pelo modelo de regressão de componentes principais;
$\hat{\boldsymbol{\beta}}_{PLS}$	Vetor de coeficientes ajustado pelo modelo de mínimos quadrados parciais;
β	Parâmetro da função de ativação logaritmo sigmoide;
δ_i	Gradiente local referente ao neurônio i ;
Δ_{ji}	Incremento do vetor de pesos para algoritmo <i>back-propagation</i> resiliente;
δ_D	Intervalo temporal desejado para a captura das médias de processo de cada variável (min).
δ_J	Janela móvel de tempo para captura das medições históricas da variável X_i entre os instantes t_a e $t_a - \delta_J$ (min);
δ_P	Deslocamentos de tempo referentes aos picos de correlação de cada uma das variáveis (min);
δ_S	Passo temporal para recuos sucessivos no tempo, de modo a capturar os dados entre os instantes $t_a - n\delta_S$ e $t_a - \delta_J - n\delta_S$, para n inteiro positivo (min);
δ_{SS}	Janela temporal para captura das médias de processo (min);

$(\sigma_k)_m$	Desvio padrão amostral dos dados da k -ésima variável coletados em janela temporal entre os instantes m e $(l+1)$;
$\boldsymbol{\varepsilon}$	Vetor coluna de resíduos;
ε_E	Somatório do erro quadrático presente na obtenção do modelo por regularização bayesiana;
ε_F	Função objetivo a ser minimizada na obtenção do modelo por regularização bayesiana;
ε_R	Somatório do quadrado dos pesos presente na obtenção do modelo por regularização bayesiana;
$\mathcal{E}_{derivada}$	Soma para o quadrado da diferença da derivada em relação ao tempo da variável resposta;
ε_{erro}	Soma para o erro quadrático;
γ	Número efetivo de parâmetros utilizados pela rede para redução da função erro;
λ	Parâmetro do equacionamento de distância <i>Minkowsky</i> ;
λ_G	Parâmetro da distribuição de probabilidade <i>Gamma</i> ;
λ_k	Autovalor obtido pela decomposição espectral da matriz de covariâncias;
λ_R	Termo de regularização;
μ	Média da função de ativação gaussiana;
μ_{LM}	Coefficiente do algoritmo Levenberg-Marquardt;
$\boldsymbol{\mu}$	Vetor centroide para cálculo de redes de base radial;
η	Taxa de aprendizagem de redes neurais;
ν_i	Ruído branco nos dados de saída;
\dot{v}	Vazão volumétrica (m^3/h , m^3/d);
σ_{ERRO}	Desvio padrão dos dados do vetor obtido pela diferença entre os valores reais da variável resposta e os valores estimados pelo modelo;
σ_k	Desvio padrão amostral para os dados da k -ésima variável;
σ_v^2	Variabilidade relacionada ao ruído branco de medição;
σ_w^2	Variabilidade relacionada à probabilidade de obtenção do vetor de

	pesos;
σ_k^2	Variância amostral para os dados da k -ésima variável;
$\sigma_{Medidor}^2$	Variabilidade observável do dispositivo de medição;
$\sigma_{Repetibilidade}^2$	Variabilidade advinda da repetibilidade do ensaio;
$\sigma_{Reprodutibilidade}^2$	Variabilidade advinda da reprodutibilidade do ensaio;
$\sigma_{Sistema}^2$	Variabilidade inerente ao sistema;
σ_{TOTAL}^2	Variabilidade total observável do sistema;
τ	Tempo de residência do inventário (s, min, h);
τ_{SS}	Tempo referente à quantidade de dias subsequentes de operação da planta selecionados para análise (min);
θ	Limiar de corte para função de ativação degrau;
$(v_k)_m$	Variância amostral dos dados da k -ésima variável coletados em janela temporal entre os instantes m e $(l+1)$;
χ	Domínio dos valores de entrada;
ξ_{ERRO}	Variável “incerteza”;
Ψ	Domínio dos valores de saída;

SUMÁRIO

1	INTRODUÇÃO.....	1
1.1	Importância do craqueamento catalítico no refino de petróleo	1
1.2	Utilização de analisadores virtuais para fins de otimização e controle de processos.....	1
1.3	Contextualização do problema e objetivos do trabalho.....	2
2	REVISÃO BIBLIOGRÁFICA.....	4
2.1	Processo de craqueamento catalítico fluidizado (CCF).....	4
2.1.1	Seção de conversão.....	6
2.1.2	Seção de fracionamento.....	8
2.1.3	Seção de recuperação de gases	10
2.1.3.1	Compressão dos gases	11
2.1.3.2	Absorção.....	12
2.1.3.3	Retificação.....	12
2.1.3.4	Debutanização	13
2.1.4	Seção de tratamento.....	13
2.1.5	Considerações finais sobre a descrição do processo	15
2.2	Ensaio de destilação ASTM D86.....	15
2.2.1	Montagem de laboratório.....	15
2.2.2	Reprodutibilidade e repetibilidade do ensaio de destilação ASTM D86.....	17
2.2.3	Considerações finais sobre o ensaio de destilação	18
2.3	Principais problemas hidráulicos em torres de fracionamento.....	19
2.4	Modelagem matemática de sistemas	24
2.4.1	Pré-tratamento dos dados de entrada.....	25
2.4.1.1	Coleta de dados.....	25
2.4.1.1.1	Seleção da frequência de amostragem dos dados.....	25
2.4.1.1.2	Sistema PI para aquisição de dados.....	26
2.4.1.2	Filtragem dos dados.....	30
2.4.1.2.1	Exclusão de outliers (abordagem monovariada)	30
2.4.1.2.2	Seleção de dados característicos de estado estacionário	32
2.4.2	Análise de <i>cluster</i>	35
2.4.2.1	Técnicas hierárquicas	38
2.4.2.2	Técnicas não hierárquicas.....	38
2.4.3	Regressão linear multivariada (MLR).....	40
2.4.3.1	Estimação de parâmetros.....	40
2.4.3.2	Cálculo da adequação do modelo de regressão	41
2.4.3.3	Coeficiente de determinação múltipla	41
2.4.4	Análise de componentes principais	43
2.4.4.1	Regressão de componentes principais (PCR).....	49
2.4.4.2	Mínimos quadrados parciais (PLS).....	50
2.4.5	Redes neurais artificiais (RNA).....	54
2.4.5.1	Modelo de neurônio McCulloch e Pitts (MCP).....	54
2.4.5.2	Funções de ativação.....	55
2.4.5.3	Aprendizado.....	57
2.4.5.4	Arquitetura de redes neurais artificiais (RNA's).....	59
2.4.5.4.1	Número de camadas.....	61
2.4.5.4.2	Número de neurônios	61

2.4.5.5	Abordagens para o treinamento de redes de <i>Perceptron</i> de múltiplas camadas (MLP)	63
2.4.5.5.1	Algoritmo back-propagation	63
2.4.5.5.2	Algoritmo back-propagation resiliente (Rprop)	66
2.4.5.5.3	Regularização Bayesiana combinada com algoritmo de treinamento Levenberg-Marquardt.....	67
2.4.6	Considerações finais sobre as técnicas de modelagem.....	72
2.5	Técnicas para seleção de variáveis	72
2.5.1	Análise de correlação.....	72
2.5.2	Regressão linear por etapas (<i>stepwise regression</i>)	73
2.5.3	Considerações finais sobre as técnicas de seleção de variáveis	74
3	METODOLOGIA.....	75
3.1	Descrição do INFERATOR – Rotinas para ajuste de inferências.....	75
3.2	Coleta e pré-tratamento dos dados.....	76
3.2.1	Aquisição dos dados do PI	76
3.2.2	Exclusão dos dados com perda de sinal.....	81
3.2.3	Exclusão de <i>outliers</i>	81
3.3	Seleção das variáveis regressoras	84
3.3.1	Obtenção dos gráficos de correlação cruzada.....	84
3.3.2	Regressão linear por etapas	86
3.3.3	Análise de <i>clusters</i>	88
3.4	Modelagem matemática de sistemas	91
3.4.1	Seleção de dados em regime estacionário	91
3.4.2	Aplicação das técnicas de modelagem	94
3.4.3	Seleção da arquitetura dos modelos	100
4	RESULTADOS E DISCUSSÃO	101
4.1	Coleta e pré-tratamento dos dados.....	101
4.1.1	Exclusão dos dados com perda de sinal.....	101
4.1.2	Exclusão de <i>outliers</i>	103
4.2	Seleção das variáveis regressoras	105
4.2.1	Obtenção dos gráficos de correlação cruzada.....	105
4.2.2	Regressão linear por etapas	113
4.2.3	Análise de clusters	118
4.3	Modelagem matemática de sistemas	126
4.3.1	Seleção de dados em regime estacionário	127
4.3.2	Seleção da arquitetura dos modelos	130
4.3.2.1	Regressão de componentes principais (PCR).....	131
4.3.2.2	Mínimos quadrados parciais (PLS)	133
4.3.2.3	Regularização Bayesiana combinada com algoritmo de treinamento Levenberg-Marquardt.....	134
4.3.2.4	Algoritmo back-propagation resiliente (Rprop)	138
4.3.2.4.1	Utilização de uma camada oculta	138
4.3.2.4.2	Utilização de duas camadas ocultas	140
4.3.2.5	Redes neurais de base radial.....	144
4.3.3	Comparação das técnicas de modelagem	149
4.3.4	Considerações finais sobre a etapa de modelagem.....	157
5	CONCLUSÃO.....	161
5.1	Sugestões para trabalhos futuros	162
6	REFERÊNCIAS BIBLIOGRÁFICAS	164
	ANEXO A	170

ANEXO B	172
ANEXO C	174

1 INTRODUÇÃO

1.1 Importância do craqueamento catalítico no refino de petróleo

Segundo Abadie (2008, p.1), o objetivo principal da unidade de craqueamento catalítico fluidizado é o aumento da produção de gasolina e gás liquefeito de petróleo (GLP) de uma refinaria. Algumas correntes de processo de maior densidade e menor valor econômico agregado, ditas ‘correntes pesadas’, tais como os gasóleos e resíduos de destilação, podem ser convertidas quimicamente em frações menos densas e com maior valor econômico, ditas correntes ‘leves’. No passado, o crescimento da indústria automobilística dos Estados Unidos impulsionou o desenvolvimento dessa tecnologia, face à crescente demanda por gasolina. Até o ano de 1913, a gasolina era obtida diretamente por destilação, sendo que sua qualidade e a quantidade obtida dependiam das características do petróleo processado. A partir da década de 1920, o craqueamento térmico começou a ser utilizado, possibilitando um aumento no rendimento de gasolina nas refinarias. Essa tecnologia evoluiu para a versão catalítica em leito fixo, posteriormente em leito móvel e, finalmente, em leito fluidizado.

Secchi et al. (2001, p.851) ressaltam que em função do grande impacto econômico da unidade de craqueamento, a implementação de estratégias avançadas de controle e otimização nesta unidade apresenta grandes ganhos. Além disso, o craqueamento catalítico confere flexibilidade operacional à refinaria, possibilitando adequar o perfil de produção para melhor atendimento ao mercado local. Mattos apud Gilbert (2004, p.3) afirma que 70% das correntes de processo que compõe o *pool* da gasolina utilizada no Brasil são oriundas do processo de craqueamento catalítico fluidizado (CCF) e, portanto, o esquema de refino adotado torna o país bastante dependente dessa tecnologia.

1.2 Utilização de analisadores virtuais para fins de otimização e controle de processos

Sharmin et al. (2006, p.6373) afirmam que, com o desenvolvimento das tecnologias de aquisição e armazenamento de histórico de dados, é grande a disponibilidade de medições de variáveis de processo. A utilização eficiente destes dados pode conduzir a

melhorias através da construção de analisadores virtuais (inferências) e técnicas de detecção de falhas. Segundo Zhao (2005, p.233), uma medida precisa de indicadores de qualidade para correntes de processo é fundamental para fins de monitoramento e controle. Em muitos casos, existem dificuldades para a medição online destas variáveis, tais como custo e confiabilidade dos medidores, presença de atrasos de transporte na unidade, necessidade de análises laboratório mais detalhadas, dentre outras. Essas dificuldades de medição podem ocasionar perda de especificação de produtos, maior consumo energético, geração de subprodutos ou mesmo condições inseguras para o processo.

Sendo assim, a estimativa das variáveis de processo através do uso de analisadores virtuais se apresenta como alternativa interessante e de baixo custo que possibilita melhorias nas estratégias de otimização e controle da unidade. Como exemplo, cita-se a possibilidade do uso de uma inferência corretamente calibrada em ferramentas como os controladores preditivos baseados em modelos (MPC's).

1.3 Contextualização do problema e objetivos do trabalho

Este trabalho teve como objetivo o desenvolvimento de uma ferramenta computacional que auxilie na criação de analisadores virtuais, dada a importância destes últimos para o controle e otimização de processos. Esta ferramenta computacional agrupou um conjunto de técnicas matemáticas e estatísticas aplicáveis às etapas de pré-tratamento dos dados coletados, seleção das variáveis candidatas à regressoras e modelagem matemática do sistema. Buscou-se a aplicação desta ferramenta na predição do ponto final de ebulição (PFE) da nafta de craqueamento catalítico de uma refinaria do sistema Petrobrás, dado que o controle desta propriedade permite uma melhor adequação do modo de operação às condições do mercado local. Com este controle é possível tanto a maior incorporação de produtos mais 'pesados' à gasolina, em detrimento à produção de correntes para o *pool* de óleo diesel, quanto a maximização da incorporação de produtos 'leves' ao diesel, em detrimento à produção de gasolina.

A predição desta propriedade foi especialmente escolhida em função da suspeita de limitações hidráulicas na torre de fracionamento desta unidade, as quais ocorrem em condições operacionais específicas e que serão descritas ao longo deste trabalho. Segundo relatos dos operadores da produção, estas limitações hidráulicas levam à

elevação excessiva do PFE da nafta craqueada. Desta forma, os limites desta propriedade na especificação da gasolina, estabelecidos pela Agência Nacional do Petróleo, Gás Natural e Biocombustíveis (ANP), podem ser excedidos. Uma predição satisfatória para esta propriedade permitiria, além da sua utilização no estabelecimento dos objetivos de otimização, a detecção de períodos de anormalidade em que o controle do PFE foi perdido, o que reduziria os riscos de perda de especificação de produto final e conseqüente necessidade de correção de tanques, gerando atraso de vendas.

Os principais objetivos desta dissertação são resumidos a seguir:

- desenvolvimento de uma ferramenta computacional aplicável a qualquer cálculo de inferências, sendo possível analisar um número genérico de variáveis candidatas à regressoras, desde que observada a restrição de apenas uma variável de saída fornecida pelo programa;
- comparação do desempenho de técnicas lineares e não lineares na modelagem matemática do problema, bem como verificação do desempenho de ferramentas estatísticas nas etapas prévias de pré-tratamento dos dados e seleção das variáveis regressoras;
- obtenção de um modelo estático capaz de prever o valor do PFE da nafta da unidade de craqueamento.

A revisão bibliográfica a seguir tem como objetivo fornecer ao leitor, inicialmente, uma visão ampla do processo de craqueamento catalítico, de modo a descrever seus principais fluxos de massa, em especial a nafta de craqueamento e as operações unitárias que a envolvem. O ensaio laboratorial para determinação do PFE é brevemente descrito, com objetivo de destacar a influência de sua repetibilidade e reprodutibilidade na precisão do analisador virtual desenvolvido. Em seguida, serão descritos os principais problemas hidráulicos que ocorrem nas torres de fracionamento, dado que os relatos obtidos dos operadores indicam a ocorrência da perda de especificação do PFE em função de um destes problemas. Na sequência, são abordadas as principais técnicas matemáticas e estatísticas utilizadas nas etapas de pré-tratamento dos dados de processo, seleção de variáveis regressoras e modelagem matemática.

2 REVISÃO BIBLIOGRÁFICA

Conforme citado no item 1.3, inicia-se a revisão bibliográfica pela descrição do processo de craqueamento catalítico fluidizado.

2.1 Processo de craqueamento catalítico fluidizado (CCF)

Segundo Mattos (2004, p.7), uma unidade de craqueamento catalítico fluidizado (CCF) emprega um catalisador à base de sílica e alumina, a altas temperaturas, capaz de promover a quebra das moléculas ‘pesadas’ presentes na carga. Os principais produtos destas unidades são a nafta craqueada, importante componente do ‘*pool*’ de gasolina, e o GLP. Conforme Farah (2008, p.139), o GLP é composto fundamentalmente por hidrocarbonetos que contêm entre três e quatro átomos de carbono. Na faixa de destilação da nafta craqueada, encontram-se compostos que possuem, fundamentalmente, entre cinco a doze átomos de carbono.

São gerados também, no processo de CCF, hidrocarbonetos ‘leves’, contendo entre um e dois átomos de carbono, os quais podem ser enviados para o sistema de gás combustível das refinarias. Outras correntes oriundas do processo de CCF são o óleo ‘leve’ de reciclo (LCO), caracterizado como componente do diesel de craqueamento, e o óleo decantado. O LCO pode ser incorporado ao *pool* de óleo diesel mediante hidrotratamento. Já o óleo decantado pode ser utilizado como diluente de óleo combustível. A carga tipicamente processada é o gasóleo oriundo da destilação a vácuo. Outras correntes podem ser incorporadas à carga, tais como o gasóleo de coque e a nafta de coque. A Figura 2.1 mostra a inserção da unidade de craqueamento catalítico em um esquema de refino, indicando a interação desta unidade com outros processos de refino, tais como destilação e coqueamento retardado:

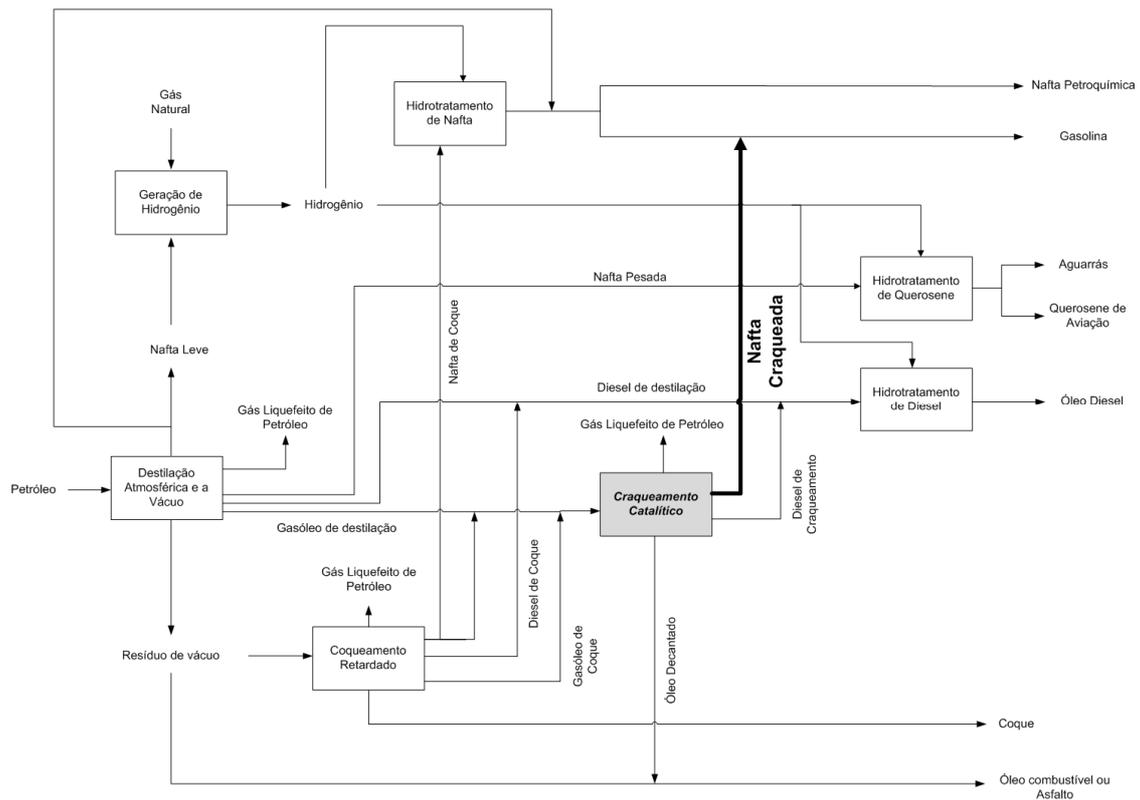


Figura 2.1: Interação da unidade de craqueamento catalítico com outros processos de refino (adaptado de Mattos (2004, p.4)).

A unidade de craqueamento catalítico pode ser subdividida nas seguintes seções, representadas na Figura 2.2:

- Bateria de pré-aquecimento;
- Seção de conversão;
- Soprador;
- Seção de recuperação de calor;
- Seção de fracionamento;
- Seção de recuperação de gases;
- Seção de tratamentos.

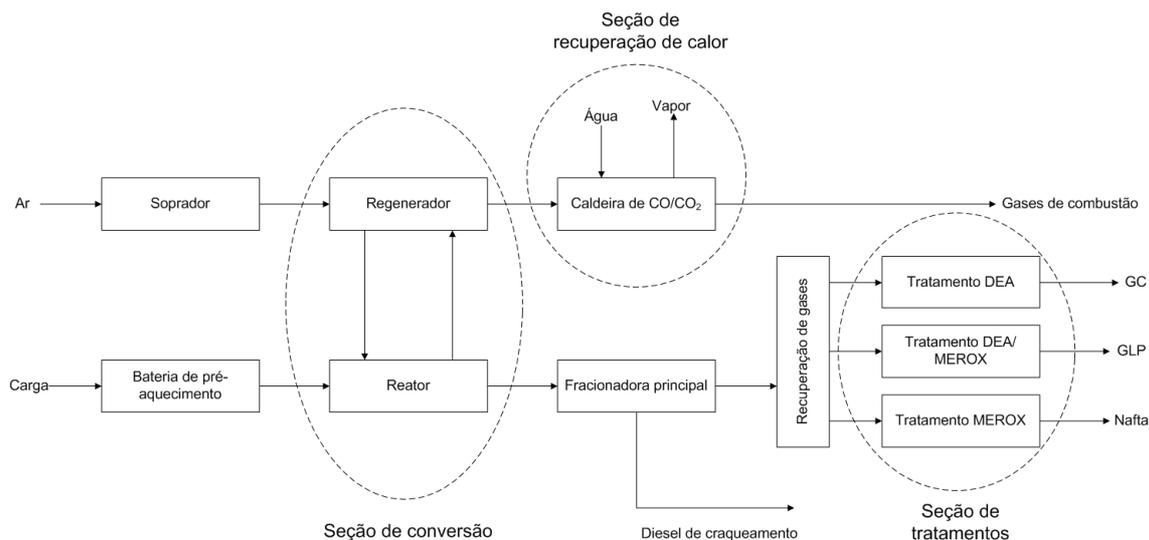


Figura 2.2: Esquema simplificado de uma unidade de craqueamento catalítico em leito fluidizado (adaptado de Mattos (2004, p.8)).

Serão descritas, a seguir, as principais seções de uma unidade de craqueamento catalítico fluidizado.

2.1.1 Seção de conversão

Antes do envio da carga para o conversor, esta passa por uma bateria de trocadores de calor, onde é pré-aquecida pelos produtos quentes oriundos do fracionamento. Após o aquecimento, a carga segue para o conversor, entrando inicialmente na base do ‘riser’ de reação. Trata-se de uma tubulação vertical, fluidizada pela injeção de vapor d’água de elevação (*vapor lift*), na qual a carga é vaporizada pelo contato com o catalisador quente oriundo da etapa de regeneração. Além de possibilitar a movimentação ascendente da mistura catalisador/carga, a injeção de vapor possibilita a redução da pressão parcial dos hidrocarbonetos no *riser* de reação.

As moléculas vaporizadas se difundem nos poros das partículas de catalisador, ocorrendo, então, as reações primárias de craqueamento (ABADIE, 2008, p.13), as quais são endotérmicas. Progressivamente, os componentes mais ‘pesados’ originam coque via reações secundárias, tais como as reações de transferência de hidrogênio e condensação (i.e formação de estruturas polinucleares), sendo que este coque vai se depositando na superfície do catalisador, reduzindo sua atividade.

No final do *riser*, a mistura escoar através de ciclones que se encontram dentro do vaso separador, os quais auxiliam na recuperação dos finos de catalisador. Este sistema reduz o arraste de hidrocarbonetos para o *stripper*. Os vapores de hidrocarbonetos saem dos ciclones praticamente sem finos de catalisador, sendo então enviados para a seção de fracionamento e, posteriormente, para a unidade de recuperação de gases (MATTOS, 2004, p.10).

Os finos de catalisador são recuperados pelos ciclones e retornam para o *stripper*. O catalisador gasto, contendo coque acumulado oriundo das reações secundárias do processo de craqueamento, desce por gravidade, entrando em contato com vapor para retificação dos hidrocarbonetos voláteis ainda aderidos. Um sistema de chicanas, localizadas ao longo do *stripper* favorece um contato mais eficiente entre o catalisador que desce e o vapor ascendente, auxiliando na remoção dos voláteis.

O catalisador gasto forma uma fase densa que escoar por gravidade para o regenerador. No regenerador, ocorre a combustão do coque depositado sobre o catalisador, o que possibilita restabelecer sua atividade. A fluidização do sistema, especificamente neste equipamento, é obtida pela injeção do ar utilizado para combustão, sendo o mesmo fornecido por um compressor de alta capacidade. Os gases de combustão ascendentes arrastam os finos de catalisador e, por isso, passam por ciclones para separação das fases. Os finos de catalisador recuperados pelos ciclones retornam para o leito de catalisador e os gases de combustão escoam para as etapas posteriores de aproveitamento do seu conteúdo entálpico. O retorno do catalisador regenerado para a base do *riser* de reação fecha o circuito do catalisador. A Figura 2.3 sintetiza a explicação anterior:

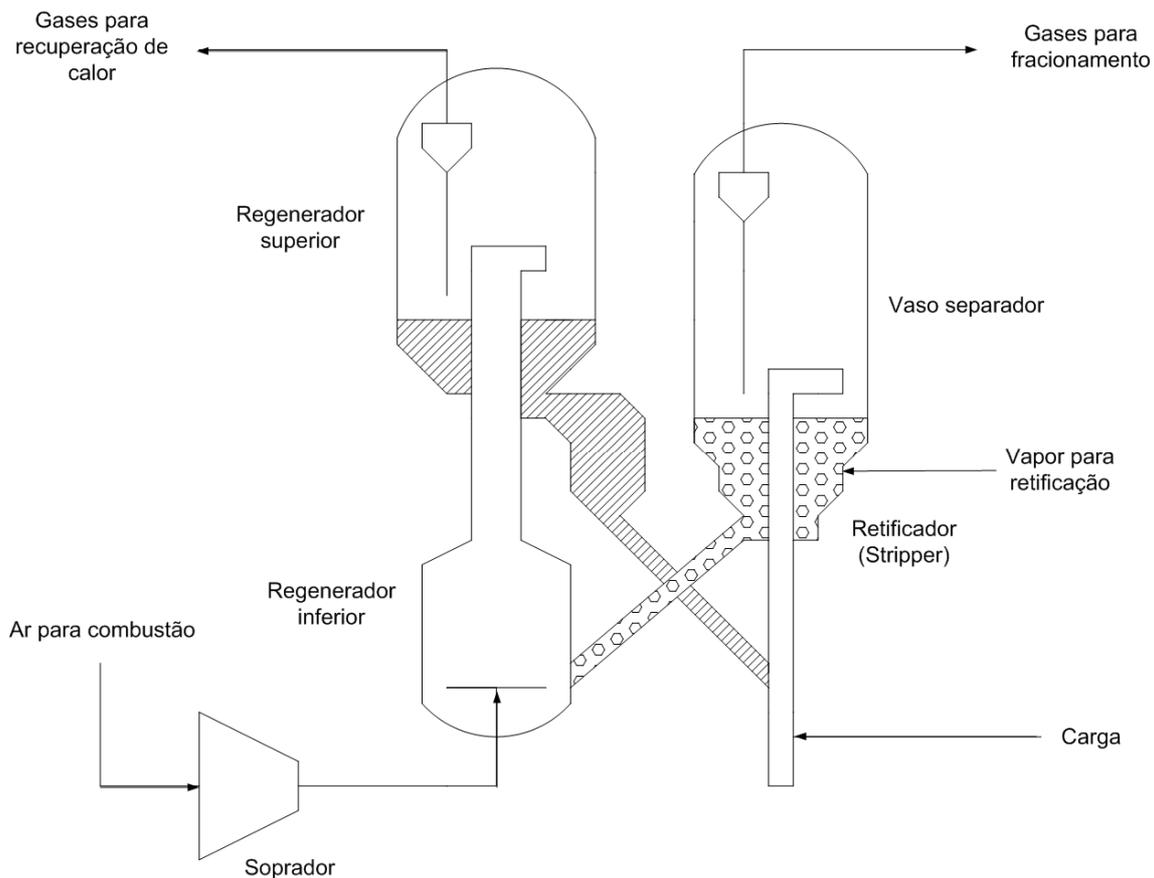


Figura 2.3: Diagrama esquemático do conversor e soprador de uma unidade de craqueamento catalítico (adaptado de Mattos (2004, p.9)).

Algumas unidades podem operar a etapa de queima no regenerador tanto por combustão parcial, dando origem aos compostos CO e CO₂, quanto por queima total, originando CO₂. Para unidades de refino que operam com queima parcial, existe uma caldeira a jusante do processo, que converte o CO produzido a CO₂, aproveitando a energia liberada (calor latente e calor sensível) para a produção de vapor. No caso de queima total, a caldeira opera apenas removendo o calor sensível dos gases, gerando vapor. Existe a possibilidade de utilização dos gases de combustão, em função de seu elevado conteúdo entálpico, para acionamento de um turbogerador, produzindo energia elétrica.

2.1.2 Seção de fracionamento

A seção de fracionamento conta com uma torre de destilação, conhecida como fracionadora principal. No entanto, existem diferenças conceituais significativas entre a fracionadora de uma unidade de CCF e uma coluna de uma unidade de destilação de petróleo, tais como:

- A produção de compostos ‘leves’ no CCF é significativa;
- Existe um acoplamento energético entre a seção de fracionamento e a seção de recuperação de gases da unidade de CCF, sendo significativa a remoção de calor pelos refluxos circulantes da unidade;
- O controle de pressão do vaso de topo da torre fracionadora do CCF influencia não só as correntes obtidas por destilação, como também a operação do conversor.

As retiradas obtidas na torre fracionadora são consequências direta da operação do conversor, tendo em vista a severidade adotada. Esta severidade é função da temperatura do riser de reação, relação catalisador/óleo e atividade/especificidade do catalisador empregado.

Os gases efluentes do conversor entram a uma temperatura incompatível com o processo de fracionamento, que precisa ser reduzida (ABADIE, 2008, p.10). O óleo clarificado, ao entrar em contato com os gases oriundos do conversor no fundo da fracionadora, possibilita um resfriamento inicial. Ao ceder calor externamente à fracionadora, esta corrente auxilia no controle do perfil térmico da região de fundo. O óleo clarificado é a retirada mais ‘pesada’ da coluna de fracionamento e pode ser incorporado ao óleo combustível para acerto de sua viscosidade.

As retiradas laterais denominadas óleo ‘leve’ (LCO) e óleo ‘pesado’ (HCO) participam da integração energética com a seção de recuperação de gases. Parte do óleo ‘leve’ é utilizada na etapa de absorção dos componentes da faixa do GLP presentes no gás combustível. A retirada de óleo ‘leve’ pode ser encaminhada para o *pool* de óleo diesel ou para a diluição do óleo combustível e conseqüente acerto de viscosidade. A retirada lateral de nafta ‘pesada’ também pode auxiliar na absorção de GLP e nafta do gás combustível, podendo seguir para tratamento e incorporação a produtos finais.

O sistema de topo tem por finalidade promover o resfriamento final da fracionadora principal, mantendo a especificação do ponto final de ebulição (PFE) da nafta craqueada (MATTOS, 2004, p.12). Auxilia também na condensação do vapor utilizado no conversor para fluidização do sistema e retificação dos hidrocarbonetos, separando-o da nafta produto. A corrente gasosa é transferida à absorvedora primária por meio de um compressor. As frações de GLP e nafta presentes no gás serão absorvidas por fluxo de

nafta craqueada instabilizada. O condensado, formado devido à injeção de vapor, é recolhido no vaso e segue para tratamento nas unidades de águas ácidas.

A Figura 2.4 exhibe os fluxos mássicos descritos anteriormente:

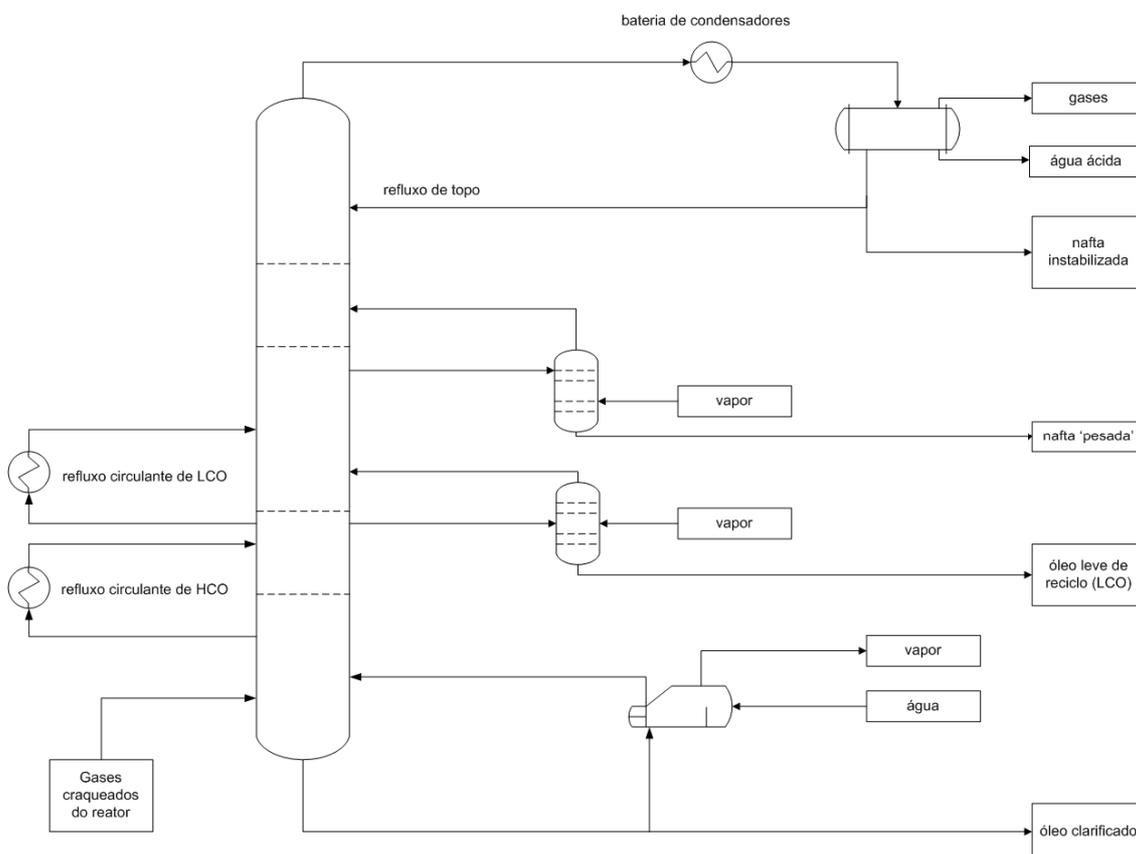


Figura 2.4: Fluxograma esquemático das principais correntes de processo referentes à seção de fracionamento de uma unidade de craqueamento (adaptado de Mattos (2004, p.11)).

2.1.3 Seção de recuperação de gases

O objetivo desta seção é recuperar do gás combustível o máximo de frações na faixa do GLP e nafta, bem como enviar para a torre debutanizadora uma nafta com a menor porcentagem possível de compostos 'leves' na faixa do gás combustível. Além disso, pode-se considerar um objetivo secundário a remoção do H_2S presente na fase líquida através da etapa de retificação. Essa etapa permite a remoção deste composto juntamente com o gás combustível, o que possibilita sua separação posterior na etapa de tratamento do gás com aminas, além de favorecer o tratamento cáustico regenerativo da nafta pelo menor teor de H_2S presente na fase líquida. A carga da unidade de recuperação consiste de uma corrente líquida, formada pela nafta 'leve' de topo e por

uma corrente gasosa, formada pelo gás combustível e o GLP. A Figura 2.5 exibe os principais fluxos mássicos presentes na seção de recuperação de gases:

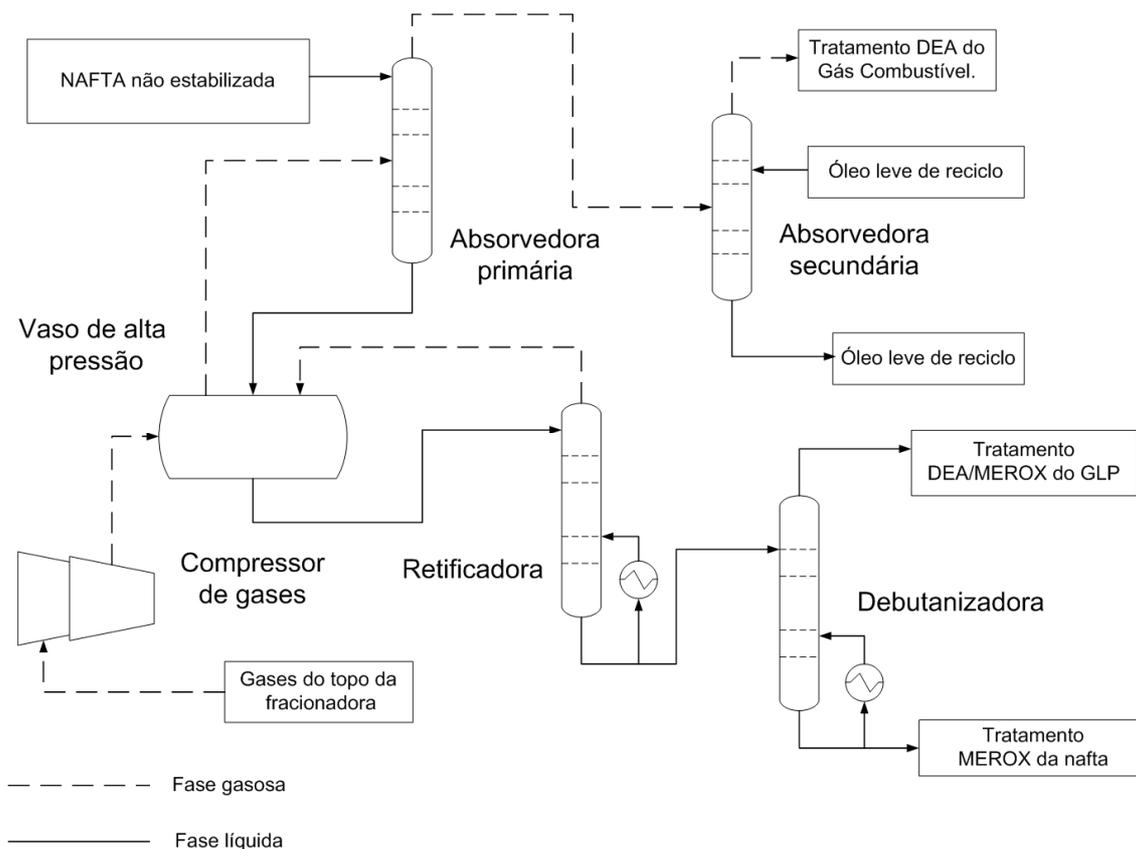


Figura 2.5: Fluxograma esquemático exibindo os principais equipamentos da seção de recuperação de gases (adaptado de Mattos (2004, p.13)).

2.1.3.1 Compressão dos gases

Os gases de topo da fracionadora são enviados à sucção do compressor. O aumento da pressão favorece a condensação dos hidrocarbonetos na faixa do GLP. A pressão de descarga do compressor corresponde à pressão de operação da absorvedora de H_2S do gás combustível, acrescida da perda de carga do sistema. Portanto, a pressão de descarga é determinada pelas condições de operação do sistema de tratamento com aminas. A composição dos gases craqueados oriundos do conversor afeta o sistema de compressão, pois um aumento na produção de compostos ‘leves’ tenderá a sobrecarregar o equipamento, em função do aumento da massa de gás a ser comprimida. Os gases são comprimidos e enviados ao vaso de alta pressão.

2.1.3.2 Absorção

A função da torre absorvedora primária é reter as frações mais ‘pesadas’ não condensadas no vaso de alta pressão. Esta seção é a principal responsável pela recuperação dos componentes na faixa do GLP, presentes no gás combustível. Entretanto, uma grande recuperação torna-se antieconômica devido à necessidade do aumento da circulação do óleo de absorção, o que resulta na sobrecarga das torres retificadora e debutanizadora. Neste caso, há necessidade do aumento do fornecimento de carga térmica da seção de fracionamento para a seção de recuperação, o que interfere no balanço de energia da torre fracionadora.

Para que ocorra a absorção, que configura uma mudança de estado físico das frações mais ‘pesadas’ do gás para a fase líquida, o calor de vaporização deve ser recebido como calor sensível pelo líquido de absorção. Este fato justifica o aumento da temperatura do óleo de absorção durante o processo. Favorecem o processo de absorção:

- A menor a temperatura do líquido absorvente;
- Pressões mais elevadas no sistema;
- Um maior contato entre o líquido e o gás;
- Um menor índice de saturação do líquido absorvente.

Além disso, quanto menor a diferença da massa molar do líquido absorvente em relação à massa molar do gás a ser absorvido, maior a capacidade de absorção. Os gases de topo são enviados para a absorvedora secundária, com a função de recuperar frações da nafta e GLP que, por ventura, foram arrastadas da primeira torre de absorção.

2.1.3.3 Retificação

A nafta oriunda do fundo do vaso de alta pressão vai para o topo da torre retificadora, sendo previamente pré-aquecida para favorecer a liberação dos compostos mais ‘leves’ para a fase gasosa (MATTOS, 2004, p.13). Pelo topo da torre saem os gases retificados que retornam ao vaso de alta pressão, incorporando-se à corrente gasosa que segue para a absorção. A eficiência da retificação depende, sobretudo, da relação entre o vapor do refulvador e o refluxo interno provido pela carga admitida no topo da torre.

2.1.3.4 Debutanização

A carga da torre debutanizadora é a nafta retificada. Nesta etapa realiza-se o acerto do teor de compostos ‘pesados’ do GLP (intemperismo), bem como a pressão de vapor da nafta debutanizada (ÍNDIO DO BRASIL, 2008, p.103). O grau de fracionamento numa torre debutanizadora é determinado pela razão de refluxo (TREYBAL, 1980, p.373), definida como:

$$R_f = \frac{R_t}{D_t} \quad (2.1)$$

Define-se R_f como a razão de refluxo, R_t como a vazão (volumétrica ou molar) de refluxo para o topo da torre e D_t como a vazão de retirada do destilado de topo (volumétrica ou molar). Quanto maior a razão de refluxo, melhor o fracionamento. A pressão da torre somente deve ser manipulada quando não se conseguir os resultados desejados através da utilização das variáveis usuais: temperatura e razão de refluxo. Ela deve ser mantida constante para que se tenha uma operação estável (ÍNDIO DO BRASIL, 2008, p.97). Mantendo-se constantes as temperaturas, o efeito do aumento de pressão na torre vai acarretar um pior fracionamento, devido à diminuição da volatilidade relativa dos componentes. Além disso, o produto de fundo incorpora uma maior quantidade de ‘leves’, diminuindo seu ponto inicial de ebulição, bem como diminuindo o ponto final de ebulição do produto de topo. Sendo assim, as retiradas de topo e fundo se tornam mais ‘leves’.

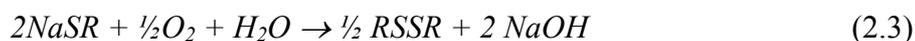
A nafta circula pelos refeedores da debutanizadora, sendo retirada do fundo da torre para tratamento. Parte do GLP produto constitui o refluxo de topo; a outra parte também segue para tratamento.

2.1.4 Seção de tratamento

Apenas o tratamento cáustico regenerativo da nafta será apresentado, uma vez que os tratamentos do GLP não serão significativos para posterior desenvolvimento do trabalho.

O tratamento cáustico tem a finalidade de remoção de compostos indesejáveis, tais como: H_2S , mercaptans, enxofre total, fenóis, etc. Define-se por ‘adoçamento’ da nafta a conversão de compostos de enxofre tóxicos e corrosivos em outros não agressivos, que permanecem no produto na forma de dissulfetos. No caso do ‘adoçamento’, utiliza-se o tratamento cáustico regenerativo, cujo catalisador é conhecido como MEROX (MERcaptan OXidation). Os contaminantes reagem com a soda/merox e são oxidados a dissulfetos ($RSSR$)¹ devido ao ar injetado no fluxo de nafta. A soda cáustica é regenerada, o que contribui para uma menor reposição (*make-up*) deste reagente.

As reações químicas envolvidas na oxidação de mercaptans e na posterior regeneração da soda cáustica são as seguintes:



Por ser um processo de ‘adoçamento’, os compostos de enxofre não são eliminados do produto final. Neste caso, o dissulfeto é incorporado à corrente de nafta, saindo com o produto tratado. Entretanto, este composto é muito menos prejudicial do que os mercaptans (RSH).

Após a oxidação dos mercaptans e remoção dos tiofenóis, a nafta craqueada está em condições de receber a adição de antioxidantes, que retardam a oxidação de hidrocarbonetos e evitam a formação de goma na gasolina (MATTOS, 2004, p.15).

O processo tem início com a adição da mistura soda/catalisador à corrente de nafta craqueada oriunda da seção de recuperação de gases. O ar requerido para oxidação dos mercaptans é injetado na corrente e a mistura é enviada para a torre de lavagem cáustica. A solução cáustica que sai pelo topo é enviada ao vaso decantador, no qual a soda decantada é bombeada novamente para a entrada da torre. Temperaturas mais baixas facilitam a extração dos mercaptans, porém dificultam a regeneração da soda. Por outro lado, temperaturas elevadas aumentam a solubilidade da soda na nafta, o que implica em uma maior necessidade de reposição deste insumo (MATTOS, 2004, p.34). Após a decantação da soda, a nafta sai pelo topo e segue para o leito coalescedor de soda residual. A nafta tratada e filtrada pode receber injeção de inibidor de oxidação.

¹ A abreviação “R” representa uma cadeia carbônica genérica.

2.1.5 Considerações finais sobre a descrição do processo

Embora não estejam explícitas as dimensões dos equipamentos tais como torres, vasos ou permutadores, é possível afirmar que o processo apresenta um tempo de residência significativo entre a retirada de nafta craqueada no topo da fracionadora e a sua obtenção na saída da seção de tratamento. Uma estimativa deste tempo de residência deve ser obtida, para que as condições da fracionadora que deram origem à nafta amostrada sejam corretamente capturadas no processo de modelagem. Fisicamente, as condições termodinâmicas e hidráulicas do topo da fracionadora principal devem determinar a faixa de destilação da corrente de nafta ‘leve’ craqueada produzida.

A seguir, descreve-se brevemente o ensaio de destilação ASTM D86, bem como a influência da sua repetibilidade e reprodutibilidade na obtenção de um analisador virtual.

2.2 Ensaio de destilação ASTM D86

2.2.1 Montagem de laboratório

O ensaio de destilação de produtos de petróleo à pressão atmosférica segue a norma ASTM D-86 “Standard Test Method for Distillation of Petroleum Products at Atmospheric Pressure”. Este ensaio é realizado em batelada e possibilita determinar quantitativamente a faixa de ebulição de destilados ‘leves’ e médios, como naftas, querosene de aviação, diesel, dentre outros. A curva de destilação do derivado é representada por um conjunto de pontos de ebulição, que correspondem às temperaturas de equilíbrio necessárias para a vaporização de quantidades definidas da mistura. (FARAH, 2008, p. 90)

A partir das expectativas iniciais para a composição, pressão de vapor, ponto inicial e final de ebulição da amostra, a mesma é classificada inicialmente em um dentre quatro grupos possíveis. O arranjo experimental do ensaio é determinado a partir dessa classificação, que define a temperatura na qual será mantido o condensador. Um volume padrão de 100 mL da amostra é destilado de acordo com condições padronizadas para o grupo em questão.

O ensaio descrito pela norma ASTM D86 caracteriza um procedimento de destilação diferencial sem refluxo e com apenas um estágio de equilíbrio. Na destilação diferencial o vapor formado é continuamente retirado do contato com o líquido residual, sendo totalmente condensado. O aquecimento segue sempre com a condensação do vapor formado, obtendo-se ao longo do ensaio dois líquidos de composição diferentes: um condensado, mais rico em compostos voláteis, e um líquido residual no balão de destilação (ÍNDIO DO BRASIL, 2008, p. 52). São realizadas leituras frequentes para a temperatura e os volumes de condensado obtidos ao longo do ensaio. Comparando-se o volume inicial de 100 mL com a soma do volume de resíduo restante no balão de destilação mais o condensado recuperado, é possível estimar as perdas por evaporação de compostos incondensáveis. As temperaturas obtidas no teste são corrigidas em função da pressão atmosférica local.

Os componentes principais do aparato de medição são: o balão de destilação, o condensador imerso em banho para resfriamento, uma fonte de calor, compartimento para o balão de destilação, dispositivo de medição de temperatura e cilindro graduado para coleta do condensado. Um dispositivo manual para o ensaio de destilação é exibido na Figura 2.6:

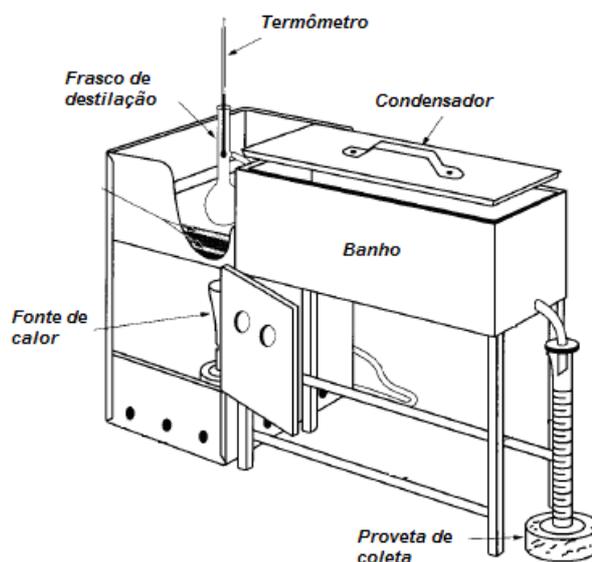


Figura 2.6: Aparato manual para ensaio de destilação (adaptado da norma ASTM D86).

De acordo com a resolução da ANP nº38, de 09 de Dezembro de 2009, o PFE máximo previsto para as gasolinas Comum e Premium é de 215°C. Farah (2008, p.164) ressalta

que o PFE da gasolina tem influência na eficiência de operação do motor em plena carga. Neste caso, deseja-se produzir a máxima potência com o menor consumo, sendo que os hidrocarbonetos da faixa final de destilação da gasolina contribuem para esse fato. No entanto, um PFE muito elevado acarreta maior formação de resíduo de carbono e conseqüente formação de depósitos no motor.

Descreve-se, no item 2.2.2, as incertezas estatísticas referentes ao ensaio padrão para determinação do PFE, bem como sua influência na construção dos analisadores virtuais.

2.2.2 Reprodutibilidade e repetibilidade do ensaio de destilação ASTM D86

Em qualquer atividade que envolva medidas, parte da variabilidade observada será atribuída ao sistema estudado; outra parte será atribuída à variabilidade dos dispositivos e procedimentos analíticos empregados na medição, conforme a equação (2.4) (MONTGOMERY, 2005, p. 357). Neste sentido, a calibração dos instrumentos de medida é fundamental para o acompanhamento de indicadores de qualidade obtidos por análises laboratoriais, tais como o PFE da nafta craqueada.

$$\sigma_{TOTAL}^2 = \sigma_{Sistema}^2 + \sigma_{Medidor}^2 \quad (2.4)$$

Duas características inerentes ao processo de medição são a precisão e a acurácia. Um instrumento com elevada acurácia é capaz de estimar corretamente a média do sistema em estudo, embora a dispersão entre as medidas experimentais para uma mesma amostra possa ser elevada em torno deste valor médio. Por outro lado, um instrumento preciso apresenta uma baixa variabilidade dentre as medidas realizadas considerando-se uma mesma amostra, embora o valor médio obtido possa apresentar um desvio em relação ao valor médio real. A avaliação da acurácia de um sistema de medição sempre requer um padrão para o qual o valor mais provável da variável medida seja conhecido, sendo possível melhorar a acurácia do instrumento através da sua regulagem, construção de curva de calibração, dentre outros procedimentos.

A variabilidade dos sistemas de medição pode ser decomposta em duas parcelas, conforme equação (2.5). Estas parcelas são denominadas repetibilidade e reprodutibilidade.

$$\sigma_{\text{Medidor}}^2 = \sigma_{\text{Reprodutibilidade}}^2 + \sigma_{\text{Repetibilidade}}^2 \quad (2.5)$$

A reprodutibilidade, segundo a norma ASTM D86, é definida como a diferença máxima entre dois testes independentes, realizados por indivíduos diferentes em aparatos de medição distintos, considerando-se a mesma amostra. Para um número grande de corridas experimentais, a norma ASTM D86 determina que a reprodutibilidade para a determinação do PFE de hidrocarbonetos na faixa de destilação da nafta craqueada não deva exceder 6,78°C em mais de uma dentre vinte corridas experimentais.

A repetibilidade é definida como a diferença máxima aceitável entre pelo menos dois resultados consecutivos, obtidos a partir de corridas experimentais realizadas em uma mesma amostra, pelo mesmo indivíduo, mantendo-se as condições de medição constantes, ou seja, uso do mesmo aparato de medição. Analogamente, para um número grande de corridas experimentais, a norma ASTM D86 determina que a repetibilidade para a determinação do PFE de hidrocarbonetos na faixa de destilação da nafta craqueada não deva exceder 3,33°C em mais de uma dentre vinte corridas experimentais.

2.2.3 Considerações finais sobre o ensaio de destilação

Conclui-se, portanto, que o melhor analisador virtual construído a partir dos dados de laboratório não pode possuir precisão superior à da repetibilidade do ensaio. Se o procedimento analítico atende a norma ASTM D86, a incerteza máxima do ensaio de laboratório é de 6,66°C, dada pela repetibilidade do ensaio, sendo $\pm 3,33^\circ\text{C}$ em relação ao valor obtido. Normalmente, os laboratórios atendem à norma com uma incerteza inferior à 6,66°C. No entanto, está implícito neste caso que as condições de amostragem são perfeitas, isto é, o produto analisado no laboratório é representativo em relação àquele contido no inventário do sistema. Para um correto procedimento de amostragem, é necessário drenar adequadamente a linha de coleta, eliminando-se o produto antigo acumulado, não representativo das condições atuais do processo.

Além das incertezas referentes ao procedimento de amostragem, a utilização do histórico de dados para a construção do analisador virtual apresenta outro problema: a imprecisão do horário de coleta da amostra. Normalmente, o horário de amostragem é informado ao banco de dados pelo analista do laboratório, após o término da análise,

horas depois. Porém, esta informação diz respeito a um horário típico em que, teoricamente, a amostra foi realizada. Esta regularidade, na prática, pode não ocorrer. A estimativa mais precisa do tempo de residência médio do inventário de nafta necessita desta maior regularidade em relação aos horários de coleta, para que o intervalo de tempo referente ao atraso de transporte não seja mascarado pela irregularidade do horário de amostragem.

Conclui-se esta seção ressaltando-se as limitações intrínsecas referentes à construção de analisadores virtuais baseados em histórico de dados de laboratório, dada as limitações do procedimento de amostragem e mesmo a precisão do método analítico empregado.

A seguir, são discutidos os principais problemas hidráulicos de torres de fracionamento, em função da suspeita dos operadores e engenheiros de acompanhamento quanto à inundação da seção de topo da fracionadora principal.

2.3 Principais problemas hidráulicos em torres de fracionamento

Os principais tipos de internos utilizados em torres de separação são os pratos e os recheios (CALDAS, 2007, p.29). Estes dispositivos devem permitir um contato eficiente entre o vapor ascendente na torre e o fluxo líquido descendente, proporcionando os fenômenos de transferência de massa e calor ao longo do equipamento.

Dentre os principais tipos de pratos, citam-se:

- Pratos com borbulhadores, que foram inicialmente utilizados nos antigos dispositivos de separação. Seu uso vem sendo reduzido em função da maior perda de carga da fase gasosa com este dispositivo;
- Pratos perfurados, que apresentam fabricação simplificada e baixa perda de carga. Contudo eles operam com eficiência de separação cerca de 15% inferior quando comparados aos pratos com borbulhadores (CALDAS, 2007, p.36), além de apresentarem limitação de operação em vazões baixas, em função da possibilidade de gotejamento excessivo;

- Pratos valvulados, que apresentam maior flexibilidade dentre os modelos citados, com perda de carga inferior ao prato com borbulhador e faixa de operação mais ampla quando comparados aos pratos perfurados.

A Figura 2.7 ilustra os modelos citados:

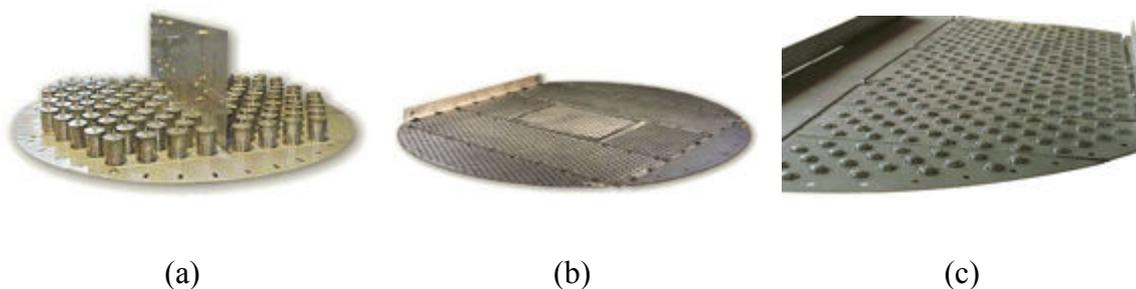


Figura 2.7: Principais tipos de pratos utilizados em torres de separação: a) prato com borbulhador; b) prato perfurado; c) prato valvulado (Site da Koch-Glitsch).

Dentre os tipos de recheios mais utilizados, citam-se:

- Classe de recheios randômicos, tais como os anéis de Raschig, a sela de Berl, os anéis de Pall e o IMTP[®];
- Classe de recheios estruturados, subdividida em ‘tradicionais’ e os ‘de alta eficiência’, como o INTALOX[®].

Os recheios randômicos são colocados no interior de torres sem a preocupação com sua disposição no leito (ÍNDIO DO BRASIL, 2008, p.23). Já os recheios estruturados são instalados com disposição ordenada, para maximizar o contato entre líquido e vapor. Os recheios estruturados de alta eficiência são caracterizados pela elevada transferência de massa e baixa perda de carga da fase gasosa. A Figura 2.8 exhibe alguns tipos de recheios:

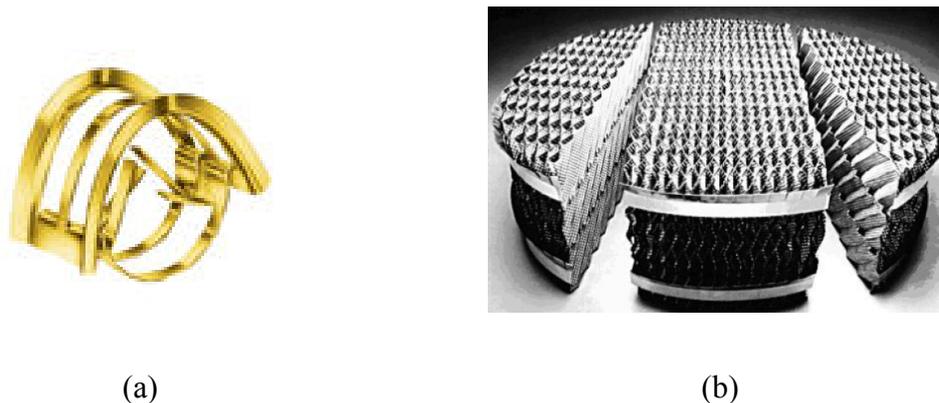


Figura 2.8: Principais tipos de recheios utilizados em torres de separação: a) recheio randômico tipo IMTP[®]; b) recheio estruturado de alta eficiência INTALOX[®] (Site da Koch-Glitsch).

O emprego de recheios em torres de separação pressupõe a utilização de um distribuidor de líquido sobre o leito recheado, de forma a garantir seu correto molhamento. A Figura 2.9 exibe um distribuidor tipo chapa-perfurada e chaminé. O líquido verte pelos pequenos furos da chapa, de forma a garantir o maior molhamento possível. O gás ascende aos estágios superiores passando pelas chaminés.

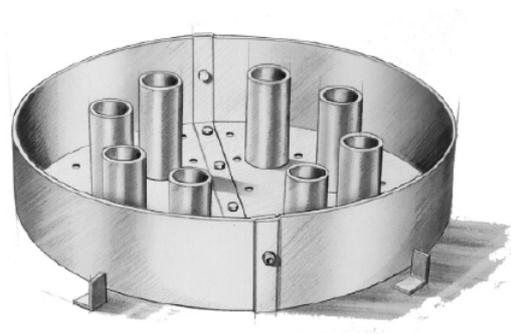


Figura 2.9: Distribuidor de líquido tipo chapa-perfurada e chaminé (Site da Koch-Glitsch).

No caso dos pratos, os dispositivos como borbulhadores, furos e válvulas são destinados à passagem do vapor, que deve entrar em contato com o nível de líquido, possibilitando a transferência de massa e calor. O líquido deve escoar em direção ao *downcomer*², em função de um gradiente hidráulico presente no prato, encontrando neste percurso o vapor que é admitido pelos dispositivos citados. Considerando-se o uso de pratos, é possível identificar em operação estável três regiões distintas formadas pelo tráfego de líquido e vapor, a saber:

² “O downcomer é a parte da torre que liga dois pratos conduzindo o fluxo de líquido entre os pratos.” (CALDAS, 2007, p.501)

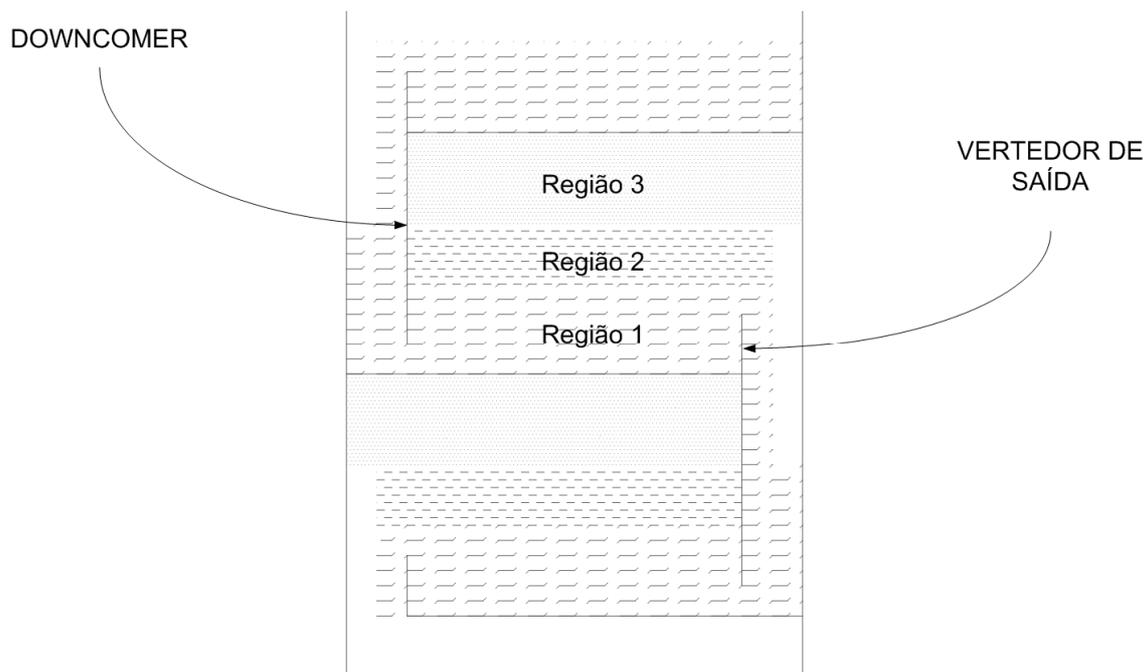


Figura 2.10: Representação esquemática dos fluidos presentes em uma torre de pratos.

- a região 1 constitui um líquido com espuma e representa a maior parte do volume que escoa pelo *downcomer*;
- a região 2 constitui grandes gotas, que coalescem e retornam ao prato;
- a região 3 constitui pequenas gotas, que podem ser arrastadas para o prato superior pelo fluxo de vapor, interferindo na eficiência de separação.

Os dispositivos de contato líquido-vapor, tanto pratos quanto recheios, apresentam limites de capacidade máxima e mínima para as vazões de líquido e vapor. Caso estes limites sejam ultrapassados, surgem problemas hidráulicos (ÍNDIO DO BRASIL, 2008, p.110) que comprometem a operação estável do equipamento de separação. No caso do uso de pratos, citam-se dentre os principais problemas:

- Gotejamento (*weeping*), que constitui a passagem do líquido através dos furos, válvulas e borbulhadores dos pratos;
- Gotejamento excessivo (*dumping*), ocasionado pela baixa vazão de vapor pelas válvulas, furos ou borbulhadores, o que possibilita a passagem do líquido por estes dispositivos, ocasionando perda da eficiência de separação;

- Inundação por *downcomer*, em que a altura do líquido aerado neste dispositivo atinge o prato superior. Pode ser ocasionado pelo baixo tempo de residência do líquido no *downcomer* (vazões de líquido elevadas), o que dificulta sua desaeração, ou ainda pela elevada perda de carga na entrada do prato inferior;
- Arraste mecânico excessivo de gotas para o prato superior, ocasionada pelo alto fluxo de vapor, resultando em perda na eficiência de separação e possibilidade de inundação das seções superiores;
- *Blowing*, semelhante ao arraste mecânico, porém mais brando. Neste caso, as gotículas arrastadas são muito pequenas para que ocorra a sua coalescência. Este fenômeno ocorre em baixas vazões de líquido combinadas com altas vazões de vapor;
- Inundação (*flooding*), em que vazões de líquido excessivas dificultam o escoamento de prato a prato (ÍNDIO DO BRASIL, 2008, p.112). O líquido no *downcomer* atinge o prato superior, impedindo o escoamento e aumentando o nível deste prato. Caso este nível suba excessivamente, ocorrerá arraste de líquido pela vazão de gás para as bandejas superiores, possibilitando a inundação de seções inteiras da torre.

A Figura 2.11 sintetiza as condições hidráulicas anteriores, de acordo com as vazões de líquido e gás em uma torre de pratos:

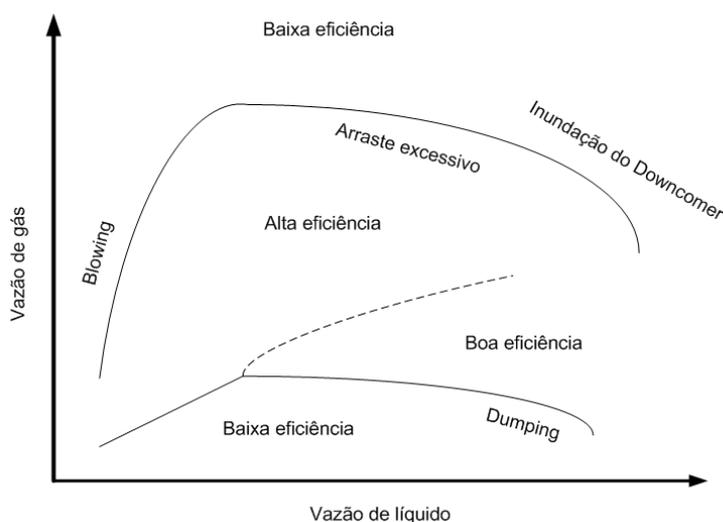


Figura 2.11: Condições hidráulicas de operação para uma torre de pratos (adaptado de Caldas et al. (2007, p.76)).

Vazões elevadas de refluxo circulante, baixa retirada de produtos, formação excessiva de espuma e alta velocidade dos vapores são as causas comuns de inundação em torres de destilação. Este fenômeno pode ser percebido pela elevação do diferencial de pressão da fase vapor na seção inundada, pois é necessária uma maior perda de carga para vencer a coluna de líquido sobre os pratos. Além disso, o gradiente térmico na seção inundada diminui, em função da grande presença de líquido que esfria a região. Nesta situação, o risco de arraste de compostos ‘pesados’ e a perda de especificação do ponto final de ebulição das retiradas acima da seção inundada são elevados.

No caso da fracionadora objeto deste estudo, sabe-se que a redução da vazão volumétrica de retirada da nafta ‘pesada’ da torre ocasiona indícios de inundação na seção de topo, pois se verifica elevação do diferencial de pressão entre a região de retirada de óleo ‘leve’ e o topo da torre. Aparentemente, o excesso de líquido que deixou de ser retirado apresenta dificuldades de escoamento nesta seção, em função do volume excessivo que verte por gravidade, causando inundação. Neste caso, o fluxo de nafta ‘pesada’ não retirada é incorporado à vazão de saída imediatamente inferior, neste caso o óleo ‘leve’ (LCO), conforme expresso pela Figura 2.4 exibida na seção 2.1.2.

Encerra-se, portanto, a revisão bibliográfica referente à parte de processo. Ressalta-se que o conhecimento do processo é fundamental na etapa de modelagem do sistema, em especial na etapa de seleção das variáveis regressoras. Os resultados das técnicas matemáticas e estatísticas descritas a seguir devem ser validados pelo conhecimento prévio do processo e, portanto, não devem ser avaliados isoladamente ou de forma absoluta.

2.4 Modelagem matemática de sistemas

Segundo Aguirre (2007, p.51), entende-se por modelagem matemática o desenvolvimento e implementação de modelos matemáticos referentes à sistemas reais. Quando o pesquisador conhece bem o sistema em questão, é possível partir para uma modelagem dita ‘caixa branca’, baseada nas leis físicas que regem o processo. Esse tipo de modelagem é denominado também modelagem fenomenológica ou conceitual. Mesmo uma modelagem ‘caixa branca’ rigorosa é apenas uma representação aproximada de um sistema real.

As técnicas de identificação de sistemas permitem ao pesquisador uma abordagem dita modelagem ‘caixa preta’, em que pouco conhecimento prévio do sistema é necessário para modelá-lo matematicamente. No entanto, neste caso, as relações físicas fundamentais permanecem ocultas ao pesquisador, o que não se torna um empecilho para o uso do modelo desenvolvido, desde que as características de interesse do sistema sejam bem representadas pelo mesmo. Para analisadores virtuais, por exemplo, o modelo deve ser adequado no que diz respeito à sua capacidade de predição. Não é necessário o fechamento rigoroso das equações de balanço de massa, energia e quantidade de movimento do sistema. Apenas os dados de entrada e saída são necessários para a obtenção do modelo dito ‘caixa preta’.

Uma abordagem intermediária, conhecida como modelagem ‘caixa cinza’, utiliza tanto os dados de entrada e saída como um conhecimento prévio do sistema, expresso nas equações fenomenológicas básicas que regem o modelo. Neste caso, busca-se estimar os parâmetros destas leis físicas fundamentais a partir dos dados de entrada e saída do sistema.

Facchin, Trierwieler e Conz (2005, p.1) ressaltam que um analisador virtual eficiente apresenta como pilares um modelo matemático adequado do processo e a correta seleção de variáveis. Lin et al. (2007, p.418) tratam da importância do pré-processamento dos dados para obtenção de um modelo robusto e com boa capacidade de generalização. A qualidade da inferência é comprometida ao se utilizar medições que não representam adequadamente o sistema, tais como as falhas de instrumento ou períodos transientes significativos no processo. Sendo assim, as técnicas de pré-processamento dos dados são tratadas a seguir.

2.4.1 Pré-tratamento dos dados de entrada

2.4.1.1 Coleta de dados

2.4.1.1.1 Seleção da frequência de amostragem dos dados

Para a construção de um modelo dinâmico, é necessário capturar a evolução temporal de um sistema em uma frequência de amostragem entre 5 a 10 vezes maior do que a maior frequência de interesse contida nos dados (AGUIRRE, 2007, p.463). Desta forma, é possível ter informação sobre as dinâmicas mais rápidas do sistema. Na prática, o desconhecimento *a priori* dessas frequências força o pesquisador a realizar uma coleta

de dados sobreamostrada, realizando posteriormente uma decimação do sinal, de modo a evitar o mau condicionamento na etapa de estimação de parâmetros. Uma frequência muito baixa de amostragem da variável de interesse a ser modelada impede a obtenção de um modelo dinâmico, pois a evolução temporal do processo não pode ser capturada. Nesta situação é possível apenas construir um modelo estático do sistema em questão.

Modelos estáticos, que são normalmente definidos por equações algébricas, relacionam variáveis sem quantificar sua dependência temporal (AGUIRRE, 2007, p.56). Um sistema real é satisfatoriamente descrito por um modelo estático se a sua variação temporal (dinâmica do processo) não é relevante, o que pode caracterizar um sistema com dinâmica muito rápida, a ponto de se tornar não observável, ou dinâmica muito lenta, quando comparada à janela temporal de interesse. Sistemas cuja dinâmica sejam de interesse do pesquisador são normalmente modelados por equações diferenciais.

O histórico das variáveis regressoras utilizadas na construção dos analisadores virtuais é obtido de um banco de dados de processo. No entanto, em função da capacidade deste banco de dados, existe um *trade-off* entre a frequência de armazenamento e o horizonte de tempo armazenado. Sendo assim, a taxa de aquisição de dados configurada pode ser baixa para observação das dinâmicas de interesse. Além disso, o pré-processamento dos dados pode gerar perda de informação. Este tópicos são abordados a seguir.

2.4.1.1.2 Sistema PI para aquisição de dados

Segundo Scheuer (2004, p.18), o sistema PI (*Plant Information*) corresponde a um conjunto de módulos de software servidor/cliente para monitoramento e análise de plantas de processo. O núcleo deste sistema é o servidor de dados PI Universal Data Server (PI-UDS), baseado em arquitetura Microsoft Windows, que permite acesso à base de dados armazenada no PI-Archive. As funções da biblioteca PI-API, as quais permitem ler e escrever valores no servidor, podem ser utilizadas a partir de linguagens de programação como Visual Basic, Delphi, C e C++, possibilitando a criação de estruturas distribuídas para coleta de dados. A aplicação cliente que permite a integração do servidor PI com o Excel é um *Add-in* denominado PI-DataLink. Esta aplicação permite disponibilizar os recursos da planilha para a análise dos dados armazenados.

Cada ponto no PI possui um conjunto de atributos que definem como o dado será coletado e armazenado. Os dados armazenados caracterizam ‘eventos’. Estes dados possuem um indicador de tempo, associado ao momento da ocorrência do ‘evento’, um indicador de *status* e um valor associado à variável medida. ‘Eventos’ significativos devem ser armazenados para possibilitar a recuperação dos dados originais, pois as rotinas de exceção e compressão dos dados, necessárias para um armazenamento eficiente, geram perda das informações de processo. Define-se como um armazenamento eficiente aquele que atende satisfatoriamente a dois requisitos conflitantes: o primeiro diz respeito ao espaço em disco utilizado para o armazenamento; o segundo diz respeito à frequência e acuidade da informação armazenada. Configurada corretamente, a aquisição de dados possibilitará a obtenção de um histórico suficientemente longo com a precisão necessária às análises de processo. Existem dois testes que determinam se o valor lido será armazenado no PI-Archive: o teste de exceção e o teste de compressão. A etapa de exceção é realizada inicialmente no nó de coleta de dados, enquanto que a compressão é realizada posteriormente pelo servidor PI-UDS.

O teste de exceção consiste em um filtro de dados com banda centrada no último valor reportado por exceção, sendo este filtro configurado no ponto de coleta (SCHEUER, 2004, p.26). O valor atual da variável é comparado ao último valor reportado ao servidor PI-UDS por exceção. Caso a diferença dos dois seja superior à metade da banda, gera-se um ‘evento’, sendo então o valor lido reportado ao servidor. A limitação na frequência de envio de dados no ponto de coleta já acarreta uma perda intrínseca de informação para a análise de processo, especialmente se a dinâmica de interesse for rápida, da mesma ordem de grandeza do período de amostragem.

O teste seguinte, referente à etapa de compressão (SCHEUER, 2004, p.26), é conhecido como ‘*swinging door compression*’. Trata-se da construção gradativa de um paralelogramo baseado nos pontos recebidos pelo servidor, em que se conecta o último ponto armazenado no PI-Archive ao último ponto recebido por exceção. Este método descarta do armazenamento os valores que estão dentro deste paralelogramo, cuja espessura (banda) é definida pelo critério de compressão. Caso o último valor recebido por exceção gere um paralelogramo que não compreenda todos os pontos previamente

recebidos por exceção desde o último ponto arquivado, então o penúltimo valor reportado é armazenado. A Figura 2.12 exhibe graficamente estas etapas.

Conforme explicado, o processo de compressão é similar ao teste de exceção, comportando-se também como um filtro. A diferença é que as especificações de exceção determinam quais ‘eventos’ devem ser enviados à PI-UDS, enquanto que a compressão determina quais ‘eventos’ serão arquivados. A requisição pelo usuário de um histórico de dados coletado em intervalos de tempos regulares implica em ‘reverter’ o algoritmo de compressão, ou seja, interpolar entre os dados contidos no PI-Archive para estimar um valor entre os dois pontos armazenados. Sendo assim, ressalta-se que o dado recuperado na frequência desejada não é real, e sim aproximado por um algoritmo de interpolação.

Após a aquisição dos dados, um pré-tratamento para a exclusão de *outliers* deve ser realizado. Do contrário, a etapa de modelagem buscará descrever eventos atípicos, que não caracterizam o processo, ocasionando perda da capacidade de generalização do modelo nas situações ditas ‘normais’. O tópico a seguir aborda estas técnicas.

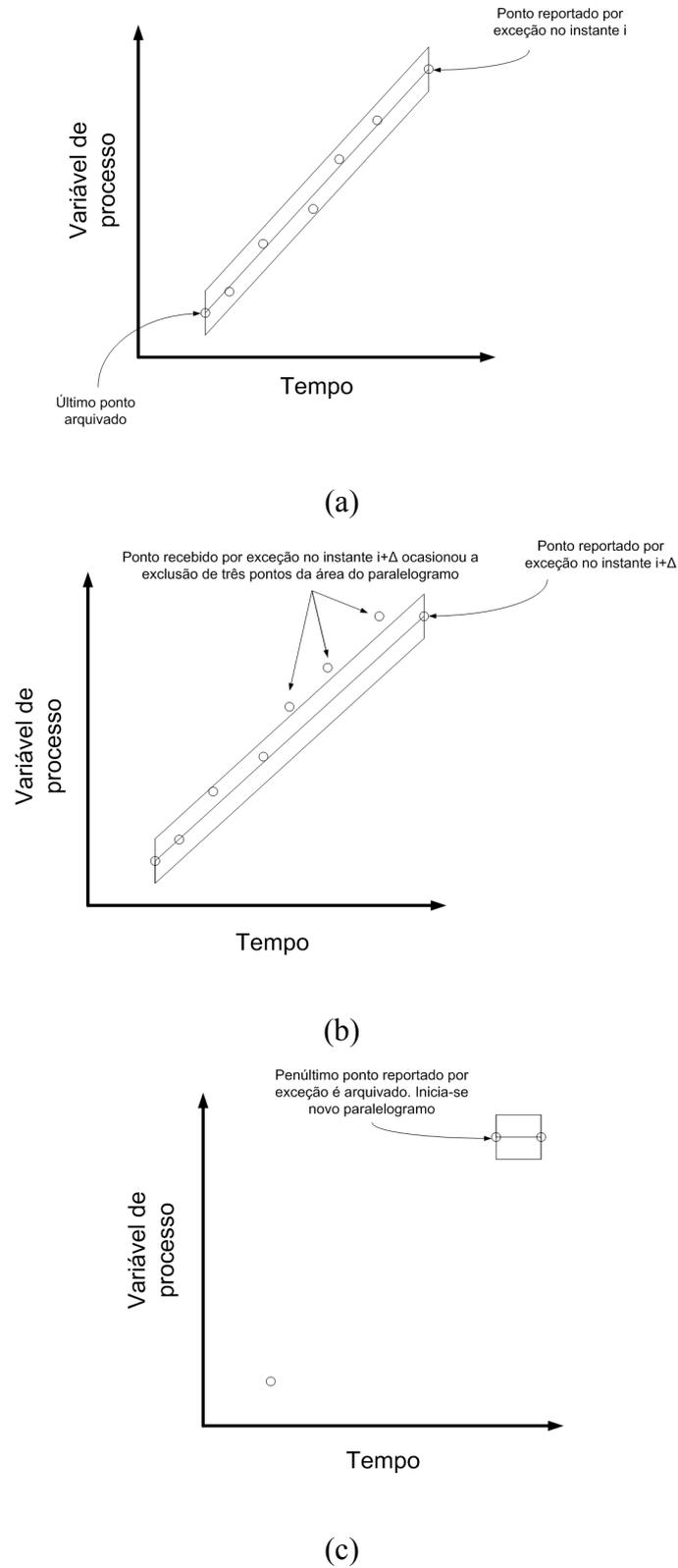


Figura 2.12: Diagrama esquemático do processo de compressão no PI: a) pontos reportados por exceção; b) paralelogramo gerado com o último ponto reportado por exceção; c) penúltimo ponto armazenado.

2.4.1.2 Filtragem dos dados

2.4.1.2.1 Exclusão de outliers (abordagem monovariada)

Segundo Lin et al. (2007, p.420), *outliers* podem ser definidos como observações não consistentes com o conjunto de dados, podendo ser consequência de falhas na aquisição da informação ou condições de processo atípicas. Os analisadores virtuais têm seu desempenho prejudicado quando *outliers* estão presentes nos dados utilizados para modelagem, devendo-se remover os mesmos. Para a detecção destes pontos, podem ser utilizadas técnicas monovariadas e/ou multivariadas.

Uma abordagem monovariada bastante utilizada (LIN, 2007, p.420) é a regra 3σ , expressa pela equação (2.6):

$$|x_{ik} - \bar{\mathbf{x}}_k| > t\sigma_k \quad (2.6)$$

Nesta inequação $\bar{\mathbf{x}}_k$ representa a média amostral dos dados da k -ésima coluna da matriz \mathbf{X} , referentes à k -ésima variável regressora. Já x_{ik} representa o i -ésimo dado amostral referente à k -ésima variável da matriz \mathbf{X} . O parâmetro t diz respeito ao critério de corte adotado e σ_k se refere ao desvio padrão amostral dos dados presentes na k -ésima coluna.

O módulo da diferença que exceder $t\sigma_k$ é classificado como *outlier*, devendo ser excluído. O valor $t = 3$ é um critério de corte comumente utilizado. No entanto, este método falha frequentemente pois a presença de *outliers* no conjunto de dados tende a ‘inflar’ a estimativa do desvio padrão σ_k , o que ocasiona um aumento do intervalo de aceitação e, conseqüentemente, reduz a sensibilidade para a detecção de *outliers*.

O identificador de Hampel substitui os estimadores média amostral e desvio padrão amostral, os quais são mais sensíveis à presença de *outliers*, respectivamente pelos estimadores mediana e MAD³. Considere a matriz de dados genérica expressa pela equação (2.7), contendo n amostras e p variáveis:

³ Trata-se da abreviação em inglês para *median absolute deviation from the median*, ou seja, a mediana dos dados gerados pelo desvio absoluto entre os valores amostrados e a mediana dos valores amostrados.

$$\mathbf{X} = [\mathbf{x}_1 \quad \mathbf{x}_2 \quad \cdots \quad \mathbf{x}_p] = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} \quad (2.7)$$

O identificador de Hampel pode ser expresso conforme equações (2.8) e (2.9):

$$MAD_k = 1,4826(\text{mediana}(|x_{ik} - \text{mediana}(\mathbf{x}_k)|)) \quad (2.8)$$

Neste caso, o MAD da k -ésima variável é obtido variando-se o índice i de 1 até n . O fator multiplicativo de 1,4826 iguala o MAD_k ao desvio padrão σ_k para uma população com distribuição normal. Neste caso, o critério de teste para a detecção de *outliers* é dado, à semelhança da equação (2.6), pela equação (2.9):

$$|x_{ik} - \text{mediana}(\mathbf{x}_k)| > t \text{ MAD}_k \quad (2.9)$$

Técnicas multivariadas para exclusão de *outliers* incluem a análise estatística do 'leverage', expresso matematicamente por h_{ii} . Este parâmetro pode ser interpretado geometricamente como a distância de uma amostra \mathbf{x}_i^T ao centroide $(\bar{\mathbf{x}})^T$ do conjunto de dados, considerando-se um total de n amostras (FERREIRA et al., 1999, p.728). Neste caso, \mathbf{x}_i^T é expresso pela equação (2.10) e $(\bar{\mathbf{x}})^T$ é expresso pelas equações (2.11) e (2.12):

$$\mathbf{x}_i^T = [x_{i1} \quad x_{i2} \quad \cdots \quad x_{ip}] \quad (2.10)$$

$$(\bar{\mathbf{x}})^T = [\bar{\mathbf{x}}_1 \quad \bar{\mathbf{x}}_2 \quad \cdots \quad \bar{\mathbf{x}}_p] \quad (2.11)$$

$$\bar{\mathbf{x}}_k = \frac{1}{n} \sum_{i=1}^n x_{ik} \quad (2.12)$$

O 'leverage' é calculado pela equação (2.13):

$$h_{ii} = \frac{1}{n} + (\mathbf{x}_i^T - (\bar{\mathbf{x}})^T)(\mathbf{X}^T \mathbf{X})^{-1}(\mathbf{x}_i^T - (\bar{\mathbf{x}})^T)^T \quad (2.13)$$

Esta grandeza é utilizada conjuntamente com a análise dos resíduos de *Student*. Uma amostra com alto resíduo e pequeno ‘leverage’ exige uma inspeção mais detalhada para aceitação no conjunto de dados. Já uma amostra com alto ‘leverage’ e alto resíduo de *Student* deve ser excluída do conjunto de dados.

2.4.1.2.2 Seleção de dados característicos de estado estacionário

As variáveis de um processo industrial apresentam uma dispersão inerente causada pelo efeito cumulativo de pequenas causas, as quais, na maioria das vezes, são de difícil identificação (MONTGOMERY, 2003, p.361). Estas causas, ditas ‘casuais’, produzem um ruído branco de fundo, que é observado ao se coletarem dados históricos referentes a essas variáveis. Caso o ruído presente seja pequeno em relação aos níveis de tolerância estabelecidos para as mesmas, o processo pode ser caracterizado como em ‘estado estacionário’.

Eventos transientes como as perturbações podem modificar a média e a dispersão das variáveis de processo, considerando-se a janela de tempo de interesse do analista. Kim et al. (2008, p.791) propuseram a utilização de uma janela móvel para a média e variância. Esta janela móvel é aplicada na detecção dos dados correspondentes ao estado estacionário. Considere novamente a matriz de dados \mathbf{X} dada pela equação (2.7). Para a variável regressora k define-se uma janela móvel m contendo l dados, tal que:

$$\left(\overline{\mathbf{x}}_k\right)_m = \frac{1}{l} \sum_{i=m-l+1}^m x_{ik} \quad (2.14)$$

A partir da definição matemática da janela móvel de dados (equação (2.14)), é possível realizar a seguinte simplificação:

$$\left(\overline{\mathbf{x}}_k\right)_m - \left(\overline{\mathbf{x}}_k\right)_{m-1} = \frac{1}{l} \left(\sum_{i=m-l+1}^m x_{ik} - \sum_{i=m-l}^{m-1} x_{ik} \right) = \frac{1}{l} (x_{mk} - x_{(m-l)k}) \quad (2.15)$$

A equação (2.15) resulta na seguinte definição:

$$\left(\overline{\mathbf{x}}_k\right)_m = \left(\overline{\mathbf{x}}_k\right)_{m-1} + \frac{1}{l} (x_{mk} - x_{(m-l)k}) \quad (2.16)$$

A variância móvel da variável k , calculada na janela de dados m , pode ser definida matematicamente por:

$$(\nu_k)_m = \frac{1}{l} \sum_{i=m-l+1}^m (x_{ik} - (\bar{\mathbf{x}}_k)_m)^2 = \frac{1}{l} \left[\left(\sum_{i=m-l+1}^m x_{ik}^2 \right) - \left((\bar{\mathbf{x}}_k)_m \right)^2 \right] \quad (2.17)$$

Analogamente à equação (2.15), é possível definir matematicamente a diferença da variância móvel da variável k considerando-se dois instantes subsequentes capturados pela janela móvel, conforme equação (2.18):

$$(\nu_k)_m - (\nu_k)_{m-1} = \frac{1}{l} \left[\left(\sum_{i=m-l+1}^m x_{ik}^2 \right) - \left((\bar{\mathbf{x}}_k)_m \right)^2 \right] - \left[\left(\sum_{i=m-l}^{m-1} x_{ik}^2 \right) - \left((\bar{\mathbf{x}}_k)_{m-1} \right)^2 \right] \quad (2.18)$$

Simplificando-se a equação (2.18), obtêm-se a equação (2.19), que expressa a variância móvel da variável k , calculada em cada janela temporal m :

$$(\nu_k)_m = (\nu_k)_{m-1} + \frac{1}{l} \left[(x_{mk})^2 - (x_{(m-l)k})^2 \right] - \left[\left((\bar{\mathbf{x}}_k)_m \right)^2 - \left((\bar{\mathbf{x}}_k)_{m-1} \right)^2 \right] \quad (2.19)$$

O desvio padrão da variável k , calculado na janela temporal m , é dado por:

$$(\sigma_k)_m = \sqrt{(\nu_k)_m} \quad (2.20)$$

Percebe-se que o identificador proposto por Kim infere o desvio padrão esperado para o conjunto de dados subsequentes, os quais caracterizam a janela temporal m , a partir da variância do conjunto de dados coletados até então, caracterizados pela janela temporal $(m-1)$. O identificador caracteriza como estado estacionário os dados em que os valores de desvio padrão amostral da variável k $(\sigma_k)_S$ sejam inferiores aos valores limites de desvio padrão $(\sigma_k)_m$ previstos utilizando-se a janela temporal anterior $(m-1)$.

As cartas de controle para a média, desvio padrão e amplitude (i.e diferença entre o valor máximo e o valor mínimo do intervalo de dados) são ferramentas estatísticas tradicionalmente empregadas no controle de qualidade (MONTGOMERY, 2003, p.367). São construídas através da amostragem periódica da variável de interesse, que

serve de base para o estabelecimento dos limites inferior e superior que caracterizam a condição de controle estatístico. Esta abordagem é semelhante a um teste de hipótese bilateral. As amostras que extrapolam esses limites são consideradas como fora de controle estatístico.

Para uma população com distribuição normal, a razão entre as amplitudes R de amostras subsequentes, cada uma contendo n dados, e o desvio padrão σ da população original geram valores que compõem a distribuição denominada W , tal que:

- A média da distribuição W é denominada d_2 ;
- O desvio padrão da distribuição W é denominado d_3 .

A distribuição W é conhecida como integral primeira de Tippett. Os detalhes matemáticos referentes à distribuição W são exibidos no ANEXO A (BARBOSA, GNERI e MENEGUETTI, 2010). A equação (2.21) define a amplitude média observável \bar{R} dentre todas as m amostras coletadas, cada uma delas contendo n dados:

$$\bar{R} = \frac{1}{m} \sum_{i=1}^m R_i \quad (2.21)$$

Os limites inferior (LIC) e superior de controle (LSC) para as amplitudes das amostras coletadas são definidos pelas equações (2.22) e (2.23):

$$LIC = \left(1 - \frac{3d_3}{d_2}\right) \bar{R} \quad (2.22)$$

$$LSC = \left(1 + \frac{3d_3}{d_2}\right) \bar{R} \quad (2.23)$$

As amostras que violam estes limites podem ser interpretadas como características de instantes de processo em que não se verificam fortes indícios para as condições de controle estatístico ou, analogamente, estado estacionário. A Tabela 2.1 exhibe valores característicos da distribuição W tabelados para amostras de tamanho n variável:

Tabela 2.1: Fatores para a construção dos gráficos de controle para amplitude.

n	d_2	$\left(1 - \frac{3d_3}{d_2}\right)$	$\left(1 + \frac{3d_3}{d_2}\right)$
2	1,1280	0,0000	3,2670
3	1,6930	0,0000	2,5750
4	2,0590	0,0000	2,2820
5	2,3260	0,0000	2,1150
6	2,5340	0,0000	2,0040
7	2,7040	0,0760	1,9240
8	2,8470	0,1360	1,8640
9	2,9700	0,1840	1,8160
10	3,0780	0,2230	1,7770

(MONTGOMERY, 2003, p.425)

2.4.2 Análise de *cluster*

A análise de *cluster* tem como objetivo dividir os elementos de uma amostra em grupos de características semelhantes (MINGOTI, 2005, p.155). As características dizem respeito a um conjunto de variáveis medidas e, neste sentido, são utilizadas métricas matemáticas que possibilitam a comparação entre vetores contendo p -variáveis de interesse. Várias métricas para a distância entre vetores podem ser utilizadas na construção dos *clusters*, sendo que os elementos que apresentam as menores distâncias são agrupados.

Considere os vetores \mathbf{x}_l^T e \mathbf{x}_k^T definidos pelas equações abaixo:

$$\mathbf{x}_l^T = [x_{l1} \quad x_{l2} \quad \cdots \quad x_{lp}] \quad (2.24)$$

$$\mathbf{x}_k^T = [x_{k1} \quad x_{k2} \quad \cdots \quad x_{kp}] \quad (2.25)$$

A equação (2.26), denominada distância generalizada ou ponderada, é expressa por:

$$d(\mathbf{x}_l^T, \mathbf{x}_k^T) = \sqrt{(\mathbf{x}_l^T - \mathbf{x}_k^T) \mathbf{H}_p (\mathbf{x}_l^T - \mathbf{x}_k^T)^T} \quad (2.26)$$

A matriz \mathbf{H}_p é uma matriz de ponderação, positiva definida. Tem-se que:

$$\mathbf{H}_p = \mathbf{I} \quad \text{Distância Euclidiana} \quad (2.27)$$

$$\mathbf{H}_p = \mathbf{S}_{pp}^{-1} \quad \text{Distância de Mahalanobis} \quad (2.28)$$

$$\mathbf{H}_p = \text{diag}\left(\frac{1}{p}\right) \quad \text{Distância Euclidiana média}^4 \quad (2.29)$$

A matriz \mathbf{S}_{pp} é definida pelas equações (2.12), (2.30) e (2.31):

$$\mathbf{S}_{pp} = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1p} \\ s_{21} & s_{22} & \cdots & s_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ s_{p1} & s_{p2} & \cdots & s_{pp} \end{bmatrix} \quad (2.30)$$

$$s_{ij} = \frac{\sum_{l=1}^n (x_{li} - \bar{x}_i)(x_{lj} - \bar{x}_j)}{n-1} \quad (2.31)$$

A distância de Minkowsky, métrica menos afetada pela presença de valores discrepantes do que a distância Euclidiana (MINGOTI, 2005, p.158), é definida por:

$$d(\mathbf{x}_l^T, \mathbf{x}_k^T) = \left(\sum_{i=1}^p w_i |x_{li} - x_{ki}|^\lambda \right)^{1/\lambda} \quad (2.32)$$

$$w_i \quad \text{Pesos de ponderação das variáveis} \quad (2.33)$$

$$\text{Para } \lambda = 1 \quad \text{Distância city-block} \quad (2.34)$$

$$\text{Para } \lambda = 2 \quad \text{Distância Euclidiana} \quad (2.35)$$

Utilizam-se duas técnicas principais para a construção de *clusters*: as hierárquicas e as não hierárquicas. As técnicas hierárquicas são subdivididas ainda em aglomerativas e divisivas (MINGOTI, 2005, p.164).

⁴ O valor p diz respeito ao número de variáveis incluídas na análise.

Além de variadas métricas de distâncias, existem também diferentes critérios de agrupamento de *clusters*, dentre eles:

- Método de ligação simples: a distância d entre dois *clusters* (C_i, C_j) é definida pelos vizinhos mais próximos, ou seja, por amostras presentes em *clusters* distintos que se encontram na menor distância dentre todas as amostras do conjunto de dados;

$$d(C_i, C_j) = \min \left[d(\mathbf{x}_l^T, \mathbf{x}_k^T, l \neq k, l \in C_i, k \in C_j) \right] \quad (2.36)$$

- Método de ligação completa: analisam-se, dentre todos os *clusters* candidatos, os elementos amostrais que apresentam a maior distância, sendo que cada um dos elementos é pertencente a um *cluster* distinto. Utiliza-se, para o agrupamento, a ‘menor’ dentre as máximas distâncias computadas;

$$d(C_i, C_j) = \min \left[\max \left[d(\mathbf{x}_l^T, \mathbf{x}_k^T, l \neq k, l \in C_i, k \in C_j) \right] \right] \quad (2.37)$$

- Método da média das distâncias: a distância entre dois *clusters* é definida como a média das distâncias entre todos os pares de elementos que podem ser formados entre dois *clusters* distintos, considerando-se n_i elementos presentes no *cluster* C_i e n_j elementos presentes no *cluster* C_j ;

$$d(C_i, C_j) = \min \left[\sum_{l \in C_i} \sum_{k \in C_j} \left[\left(\frac{1}{n_i n_j} \right) d(\mathbf{x}_l^T, \mathbf{x}_k^T) \right] \right] \quad (2.38)$$

- Método do centroide: a distância entre dois *clusters* é definida pela distância entre os vetores de média, denominados ‘centroides’ dos conglomerados. Trata-se do quadrado da distância Euclidiana entre os vetores de média;

$$d(C_i, C_j) = \min \left[(\bar{C}_i - \bar{C}_j)(\bar{C}_i - \bar{C}_j)^T \right] \quad (2.39)$$

$$\bar{C}_i = \frac{1}{n_i} \sum_{l \in C_i} \mathbf{x}_l^T \quad (2.40)$$

$$\bar{C}_j = \frac{1}{n_j} \sum_{k \in C_j} \mathbf{x}_k^T \quad (2.41)$$

2.4.2.1 Técnicas hierárquicas

As técnicas hierárquicas são utilizadas para a determinação do número aproximado de *clusters* presentes nos dados. No caso das técnicas aglomerativas, cada amostragem constitui, inicialmente, um *cluster*. Em cada passo do algoritmo, apenas um novo conglomerado é formado pelo agrupamento do par mais semelhante, os quais se mantêm unidos até o término dos cálculos. Sendo assim, é possível construir uma representação gráfica do histórico de agrupamentos, denominada ‘dendograma’ (Figura 2.13). Busca-se encontrar uma divisão de grupos que represente uma ‘partição natural’ dos elementos que estão sendo comparados.

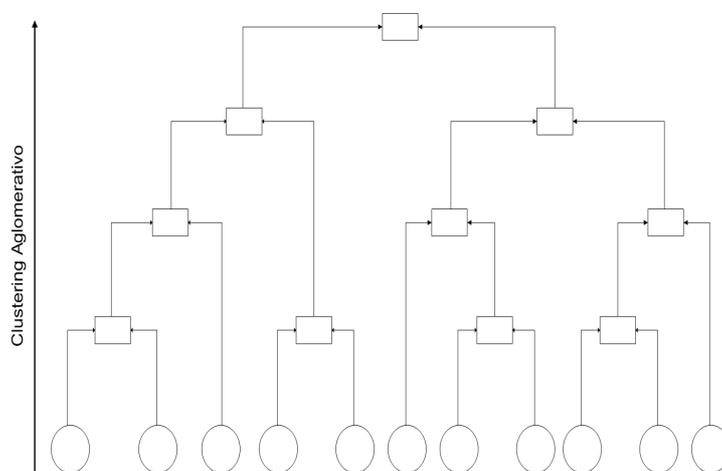


Figura 2.13: Dendograma exemplificando o histórico de agrupamentos em um algoritmo hierárquico

Existem diversos testes estatísticos propostos por autores para verificar o número mais provável de *clusters* dentro de um conjunto de dados. Para maiores detalhes recomenda-se Mingoti (2005, p.180):

2.4.2.2 Técnicas não hierárquicas

As técnicas não hierárquicas requerem que o analista especifique previamente o número g^* de *clusters* existentes na amostra. A partição proposta pelo algoritmo busca atender a dois critérios básicos: semelhança interna entre os elementos de cada *cluster* e o isolamento entre os *clusters* distintos formados (MINGOTI, 2005, p.192). Uma diferença fundamental nestes algoritmos é que, no caso das técnicas hierárquicas, a união de dois *clusters* em passos intermediários do algoritmo implica na união de ambos

ao término da rotina. No caso das técnicas não hierárquicas, novos *clusters* podem ser formados a partir da divisão de outros formados anteriormente, através de cálculos iterativos.

O método conhecido como k-médias utiliza a medida de distância do elemento amostral aos centroides dos *clusters* para alocá-lo em um grupo específico. O centroide de menor distância em relação ao elemento amostral determina sua alocação. O algoritmo tem início com a escolha prévia de ‘sementes’ para os centroides, que iniciam o processo de divisão. Todos os elementos amostrais são alocados de acordo com sua distância em relação a esses centroides iniciais. Normalmente, utiliza-se a distância Euclidiana padrão. Em seguida, inicia-se o passo dois com um novo cálculo dos centroides a partir dos elementos amostrais alocados no passo anterior. Procede-se então um novo ciclo de alocação dos elementos, que continua até que nenhuma realocação seja necessária.

O agrupamento final obtido pelo método das k-médias é influenciado pelas ‘sementes’ utilizadas. Dentre outras alternativas possíveis, elas podem ser escolhidas de acordo com os seguintes critérios:

- Utilizando-se g^* amostras aleatórias do conjunto de dados;
- Utilizando-se os centroides obtidos por estudos prévios através de técnicas hierárquicas;
- Utilizando-se valores pré-fixados pelo analista;
- Utilizando-se as g^* primeiras amostras do banco de dados, padrão utilizado por diversos softwares estatísticos.

O programa computacional desenvolvido disponibiliza rotinas para a análise de *clusters*. Cabe ressaltar, porém, que esta técnica não é empregada diretamente para obtenção dos modelos. Sua principal finalidade é a orientação do usuário na escolha das variáveis regressoras, as quais deverão ser utilizadas nas etapas de modelagem. No entanto, os aspectos principais da técnica, bem como suas limitações devem ser conhecidos, para permitir melhor julgamento de seus resultados.

2.4.3 Regressão linear multivariada (MLR)

Um modelo de regressão linear multivariada descreve uma relação do tipo (MONTGOMERY, 2003, p.230):

$$\mathbf{y} = \beta_0 + \beta_1 \mathbf{x}_1 + \beta_2 \mathbf{x}_2 + \cdots + \beta_k \mathbf{x}_k + \boldsymbol{\varepsilon} \quad (2.42)$$

Trata-se de um hiperplano presente em um espaço contendo $k+1$ dimensões, sendo k o número de variáveis regressoras no modelo. Ele é dito linear nos parâmetros, pois as variáveis regressoras são multiplicadas diretamente por coeficientes de regressão β_j , sendo que a variável \mathbf{x}_j pode apresentar uma função não linear em relação aos dados de entrada coletados. Os coeficientes representam a mudança esperada na variável resposta \mathbf{y} por unidade de variação da regressora \mathbf{x}_j , desde que todas as outras variáveis sejam mantidas constantes.

Esta técnica se mostra útil para a aproximação de funções não lineares em torno de um ponto de operação. Sendo assim, a verdadeira relação entre a variável resposta \mathbf{y} e as variáveis regressoras \mathbf{x}_j pode ser desconhecida e, ainda assim, satisfatoriamente aproximada por um hiperplano em condições específicas de processo. Quanto maior for a não linearidade da função original e quanto mais distante se encontrar o processo do ponto de operação considerado na modelagem, pior será a descrição do fenômeno físico pelo modelo linear.

2.4.3.1 Estimação de parâmetros

Para ajustar os coeficientes do modelo de regressão, utiliza-se uma abordagem matricial (MONTGOMERY, 2003, p.234). Supondo a existência de $n > k$ observações, dado a inclusão de k variáveis no modelo, tem-se que:

$$\mathbf{y} = \mathbf{X}_{MLR} \boldsymbol{\beta}_{MLR} + \boldsymbol{\varepsilon} \quad (2.43)$$

$$\mathbf{X}_{MLR} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} \quad (2.44)$$

A matriz \mathbf{X}_{MLR} representa os dados de entrada \mathbf{X} acrescidos de uma coluna unitária para ajuste do termo constante β_0 . O vetor \mathbf{y} representa as observações da variável dita dependente. O vetor $\boldsymbol{\varepsilon}$ representa um resíduo estocástico.

O vetor $\hat{\boldsymbol{\beta}}_{MLR}$ pode ser calculado pela solução de mínimos quadrados, dada por:

$$\hat{\boldsymbol{\beta}}_{MLR} = (\mathbf{X}_{MLR}^T \mathbf{X}_{MLR})^{-1} \mathbf{X}_{MLR}^T \mathbf{y} \quad (2.45)$$

Os detalhes desta solução podem ser encontrados no ANEXO B.

2.4.3.2 Cálculo da adequação do modelo de regressão

Para um modelo linear, algumas suposições devem ser respeitadas. Os resíduos $\boldsymbol{\varepsilon}$ do modelo linear ajustado devem apresentar média nula e variância constante, além de comportamento estocástico. Algumas vezes, a hipótese de que o modelo se comporte de maneira linear não é válida para as faixas de operação de interesse, sendo necessária a constante avaliação do analista, com embasamento nos conhecimentos de processo, para verificação da adequabilidade do modelo.

Um gráfico de probabilidade normal dos resíduos (*normal plot*) auxilia na verificação da hipótese de distribuição normal. Um gráfico da distribuição temporal dos resíduos pode indicar uma mudança da variância com o tempo. Da mesma forma, a verificação da distribuição dos resíduos em função dos níveis das variáveis regressoras pode auxiliar na identificação de mudanças na variabilidade do sistema para diferentes pontos de operação.

Modelos não lineares podem, eventualmente, ser linearizados através da aplicação de operadores matemáticos. Este é o caso dos modelos ditos *intrinsecamente lineares*, tais como a função exponencial.

2.4.3.3 Coeficiente de determinação múltipla

Determina-se coeficiente de determinação múltipla (R^2) pela equação (2.46):

$$R^2 = \frac{SQ_R}{SQ_T} = 1 - \frac{SQ_E}{SQ_T} \quad (2.46)$$

O coeficiente de determinação pode ser compreendido como a fração da variabilidade total dos dados explicada pelo modelo de regressão. A variância total de um sistema modelado linearmente pode ser dividida em duas parcelas principais, segundo a equação (2.47) (MONTGOMERY, 2003, p.214):

$$SQ_T = SQ_R + SQ_E \quad (2.47)$$

$$SQ_R = \beta_{MLR}^T \mathbf{X}^T \mathbf{y} - \frac{\left(\sum_{i=1}^n y_i \right)^2}{n} \quad (2.48)$$

$$SQ_E = \mathbf{y}^T \mathbf{y} - \beta_{MLR}^T \mathbf{X}^T \mathbf{y} \quad (2.49)$$

O primeiro termo do lado direito da equação (2.47) é denominado soma quadrática da regressão (SQ_R) e diz respeito à variabilidade em y_i devido à linha de regressão. O segundo termo é denominado soma quadrática dos erros (SQ_E) e diz respeito à variação residual deixada sem explicação pela linha de regressão.

Deve-se estar atento, durante a etapa de modelagem, para o fato de que a adição de termos ao modelo implica no aumento de R^2 , o que não significa que o novo modelo é superior ao antigo.

A menos que a soma quadrática dos erros no novo modelo seja reduzida por uma quantidade igual à média quadrática dos erros do modelo original, o novo modelo terá uma média quadrática dos erros maior do que o modelo antigo, por causa da perda de grau de liberdade no erro. (MONTGOMERY, 2003, p.221)

Define-se a média quadrática do erro MQ_E , bem como a média quadrática da regressão MQ_R , como:

$$MQ_E = \frac{SQ_E}{(n - (k + 1))} \quad (2.50)$$

$$MQ_R = \frac{SQ_R}{k} \quad (2.51)$$

2.4.4 Análise de componentes principais

A análise de componentes principais tem como objetivo explicar a estrutura de variância e covariância de um conjunto de dados composto de p -variáveis aleatórias, através de combinações lineares das variáveis originais. Ela foi introduzida por Pearson em 1901, sendo aperfeiçoada pelos trabalhos de Hotelling em 1933 (MINGOTI, 2005, p. 59). Estas combinações lineares são chamadas de componentes principais, sendo as mesmas ortogonais entre si e, portanto, não correlacionadas. Essa técnica é bastante utilizada para fins exploratórios de simplificação da estrutura de variabilidade dos dados (JOHNSON, WICHERN, 1992, p. 356).

Utiliza-se a decomposição espectral da matriz de correlação ou covariância de um sistema para a determinação das componentes principais. Para a exemplificação desta técnica, considera-se uma matriz de dados de processo \mathbf{X} que possua vetores colunas \mathbf{x}_k referentes à p -variáveis aleatórias, sendo que cada elemento do vetor se refere à amostragem em um instante i , considerando-se um total de n amostragens:

$$\mathbf{X} = [\mathbf{x}_1 \quad \mathbf{x}_2 \quad \cdots \quad \mathbf{x}_p] = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} \quad (2.7)$$

Estima-se o vetor de médias $(\bar{\mathbf{x}})^T$ através dos dados amostrais:

$$\bar{\mathbf{x}}_k = \frac{1}{n} \sum_{i=1}^n x_{ik} \quad (2.52)$$

$$(\bar{\mathbf{x}})^T = [\bar{\mathbf{x}}_1 \quad \bar{\mathbf{x}}_2 \quad \cdots \quad \bar{\mathbf{x}}_p] \quad (2.53)$$

A covariância mede o grau de associação linear entre variáveis aleatórias (MONTGOMERY, 2003, p. 116), ou seja, quantifica se o acréscimo em uma variável implica no acréscimo ou decréscimo da outra variável, conforme já expresso anteriormente pela equação (2.31).

Destaca-se que, se a relação entre as variáveis aleatórias não for linear, a covariância pode não ser sensível à relação. Além disso, embora a covariância traga alguma

informação sobre o relacionamento linear entre duas variáveis, é difícil quantificar se essa relação é forte ou fraca, em função da inexistência de valores máximos e mínimos de referência. A análise da matriz de correlação, ao invés da matriz de covariância, soluciona este problema (MINGOTI, 2005, p. 29). A matriz de covariância é definida pelas equações (2.30) e (2.31).

A definição de covariância resulta que a matriz \mathbf{S}_{pp} é simétrica, pois $s_{ij} = s_{ji}$ para $i \neq j$. Segundo o teorema da decomposição espectral, a matriz \mathbf{S}_{pp} garante a existência de uma matriz ortogonal \mathbf{C}_{pp} , ou seja, $\mathbf{C}_{pp}^T \mathbf{C}_{pp} = \mathbf{C}_{pp} \mathbf{C}_{pp}^T = \mathbf{I}_p$ (MINGOTI, 2005, p. 35), tal que:

$$\mathbf{C}_{pp}^T \mathbf{S}_{pp} \mathbf{C}_{pp} = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_p \end{bmatrix} = \mathbf{A}_{pp} \quad (2.54)$$

Classifica-se a matriz \mathbf{A}_{pp} como similar à matriz \mathbf{S}_{pp} , pois a variância total e a variância generalizada de ambas são iguais:

- Variância total:

$$\text{traço}(\mathbf{S}_{pp}) = \sigma_1^2 + \sigma_2^2 + \dots + \sigma_p^2 = \lambda_1 + \lambda_2 + \dots + \lambda_p = \text{traço}(\mathbf{A}_{pp}) \quad (2.55)$$

- Variância generalizada:

$$\det(\mathbf{S}_{pp}) = \prod_{i=1}^p \lambda_i = \det(\mathbf{A}_{pp}) \quad (2.56)$$

Considerando-se o teorema da decomposição espectral, expresso pela equação (2.54), a seguinte igualdade é válida:

$$\mathbf{S}_{pp} = \mathbf{C}_{pp} \mathbf{A}_{pp} \mathbf{C}_{pp}^T \quad (2.57)$$

A matriz \mathbf{C}_{pp} , ou matriz de autovetores, é composta de vetores coluna de norma unitária, ortogonais entre si. Então:

$$\mathbf{C}_{pp} = [\mathbf{c}_1 \quad \mathbf{c}_2 \quad \cdots \quad \mathbf{c}_p] = \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1p} \\ c_{21} & c_{22} & \cdots & c_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ c_{p1} & c_{p2} & \cdots & c_{pp} \end{bmatrix} \quad (2.58)$$

A ortogonalidade dos vetores coluna implica que:

$$\mathbf{c}_i^T \mathbf{c}_i = 1 \quad (2.59)$$

$$\mathbf{c}_i^T \mathbf{c}_j = 0 \quad (2.60)$$

Estes vetores constituem as direções das componentes principais. Os detalhes algébricos de obtenção dos autovetores e matriz de autovalores \mathbf{A}_{pp} podem ser encontrados em Johnson e Wichern (1992, p. 48). Sendo assim, a existência de p -variáveis no conjunto de dados originais garante obtenção de p -componentes, todas elas ortogonais entre si. Os *scores* t_{ik} das componentes obtidas são definidos, em termos algébricos, por:

$$\begin{aligned} t_{i1} &= x_{i1}c_{11} + x_{i2}c_{21} + \cdots + x_{ip}c_{p1} \\ t_{i2} &= x_{i1}c_{12} + x_{i2}c_{22} + \cdots + x_{ip}c_{p2} \\ &\vdots \\ t_{ip} &= x_{i1}c_{1p} + x_{i2}c_{2p} + \cdots + x_{ip}c_{pp} \end{aligned} \quad (2.61)$$

Trata-se, portanto, de combinações lineares das variáveis originais. As novas variáveis t_{ij} 's são não correlacionadas, em função da ortogonalidade das direções das componentes. Os coeficientes c_{ij} , direções das componentes principais, são denominados *loads*, e representam, geometricamente, o cosseno formado entre os eixos originais e o novo eixo definido pela componente em questão. Os valores de t_{ij} , também denominados *scores*, representam a projeção de um ponto original, p -dimensional, no novo eixo definido pela componente, conforme Figura 2.14:

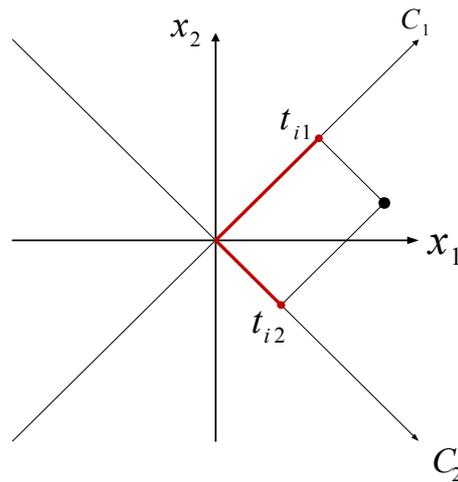


Figura 2.14: Interpretação geométrica dos scores.

A matriz de *scores* \mathbf{T} é obtida pelo produto matricial:

$$\mathbf{T} = \mathbf{X}\mathbf{C}_{pp} = [\mathbf{t}_1 \quad \mathbf{t}_2 \quad \cdots \quad \mathbf{t}_p] = \begin{bmatrix} t_{11} & t_{12} & \cdots & t_{1p} \\ t_{21} & t_{22} & \cdots & t_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ t_{n1} & t_{n2} & \cdots & t_{np} \end{bmatrix} \quad (2.62)$$

Embora p -componentes sejam requeridas para reproduzir a variabilidade total do sistema, normalmente grande parte dessa variabilidade pode ser explicada por um número inferior de componentes. Sendo assim, essa informação pode ser reproduzida com qualidade utilizando-se k componentes, tal que $k < p$. As matrizes \mathbf{T} e \mathbf{X} possuem a mesma variância total e a mesma variância generalizada. Porém, em função da ortogonalidade das bases, a variabilidade pode ser atribuída em separado a cada uma das componentes, sendo que a porcentagem da variabilidade dos dados originais explicada em cada direção é dada por:

$$\% \text{ var}_i = \frac{\lambda_i}{\lambda_1 + \lambda_2 + \dots + \lambda_p} 100\% \quad (2.63)$$

Os autovalores da matriz \mathbf{A}_{pp} são apresentados em ordem decrescente na diagonal principal. Portanto, a primeira componente principal explica a maior porcentagem da variabilidade. Conceitualmente, ela é a combinação linear que maximiza a variância do

primeiro vetor coluna de \mathbf{T} , respeitando-se a restrição de ortogonalidade da componente (JOHNSON, WICHERN, 1992, p. 357). Matematicamente:

$$\max(\text{Var}(\mathbf{t}_1)) = \mathbf{c}_1^T \mathbf{S}_{pp} \mathbf{c}_1 \quad (2.64)$$

$$\mathbf{c}_1^T \mathbf{c}_1 = 1 \quad (2.65)$$

Analogamente, a segunda componente principal é aquela que maximiza a variância do segundo vetor coluna de \mathbf{T} , respeitando-se as restrições de ortogonalidade, e assim sucessivamente:

$$\max(\text{Var}(\mathbf{t}_2)) = \mathbf{c}_2^T \mathbf{S}_{pp} \mathbf{c}_2 \quad (2.66)$$

$$\mathbf{c}_2^T \mathbf{c}_2 = 1 \quad (2.67)$$

$$\mathbf{c}_1^T \mathbf{c}_2 = 0 \quad (2.68)$$

A Figura 2.15 representa a medição de três variáveis de processo, sendo duas delas classificadas como regressoras (x_1, x_2) e uma classificada como dependente (y). Percebe-se que a maior parte da variabilidade dos dados se distribui ao longo da direção da componente principal:

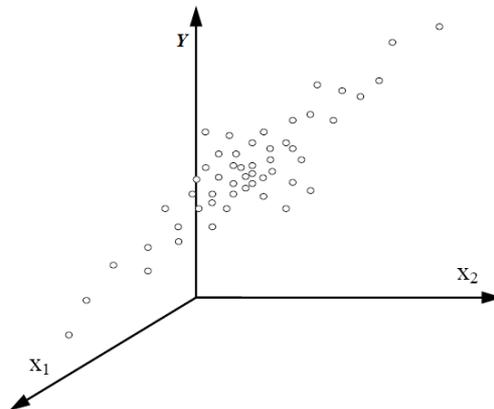


Figura 2.15: Variáveis de processo correlacionadas (adaptado de Feital (2009, p.19)).

A aplicação da decomposição espectral da matriz de correlação dos dados originais possibilita a obtenção de novas bases ortogonais (Figura 2.16), as quais definem um novo sistema de coordenadas, sendo que a medição original define também os *scores* (t_1, t_2, t_3).

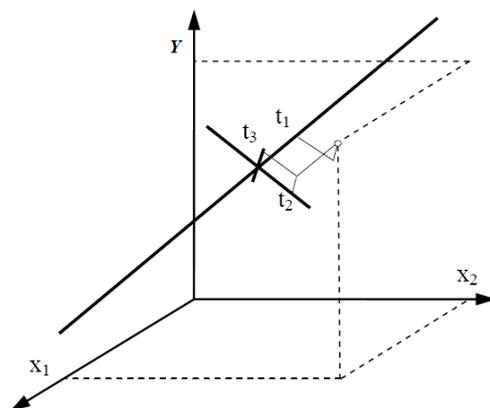


Figura 2.16: Variáveis de processo representadas nos eixos originais e nos novos eixos de coordenadas (adaptado de Feital (2009, p.19)).

Sendo assim, seria possível utilizar apenas uma das projeções (t_1) para ‘reconstruir’ o ponto original no sistema de coordenadas inicial $(\hat{x}_1, \hat{x}_2, \hat{y})$. A projeção escolhida, conforme Figura 2.17, seria aquela cuja componente explica a maior parte da variabilidade dos dados:

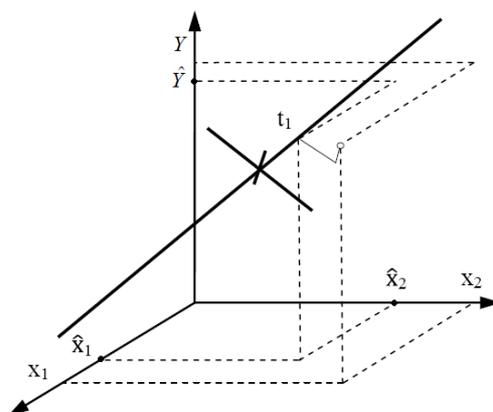


Figura 2.17: Reconstrução de um valor a partir da sua projeção na componente mais significativa (adaptado de Feital (2009, p.21)).

Obtendo-se uma estimativa dos *scores* t_{ij} , os mesmos podem ser utilizados, de acordo com a Figura 2.17, para calcular um vetor estimado \hat{y} . Selecionando-se as componentes mais significativas, é possível realizar uma redução da dimensionalidade do problema, desde que o número k de componentes principais selecionadas para descrição da variabilidade original do sistema seja inferior ao número n de amostras presentes no conjunto de dados.

2.4.4.1 Regressão de componentes principais (PCR)

A regressão de componentes principais (PCR) é baseada na decomposição em valores singulares (FERREIRA et al., 1999, p.727). Uma vantagem desta técnica em relação à regressão linear é a possibilidade de tratar amostras multicolineares. A colinearidade adiciona redundância ao modelo, o que causa instabilidade numérica na estimativa dos coeficientes de regressão.

A decomposição em valores singulares da matriz de regressores \mathbf{X} é expressa por:

$$\mathbf{X} = \mathbf{T}^* \mathbf{P}_{pp}^{*T} \quad (2.69)$$

$$\mathbf{T}^* = \mathbf{U}_{nm}^* \mathbf{D}^* \quad (2.70)$$

A matriz \mathbf{D}^* não é quadrada e, portanto, não pode ser invertida diretamente. Apenas os elementos da diagonal d_{ij}^* , para $i = j$, apresentam valores não nulos. Analogamente à equação (2.63), estes elementos estão relacionados à porcentagem de variabilidade explicada por cada componente, tal que:

$$\% \text{ var}_i = \frac{d_{ii}^2}{d_{11}^2 + d_{22}^2 + \dots + d_{pp}^2} 100\% \quad (2.71)$$

A matriz \mathbf{D}^* pode ser simplificada, ao se considerar as k componentes principais que explicam a maior parte da variabilidade dos regressores, obtendo-se:

$$\mathbf{D}^{**} = \begin{bmatrix} d_{11} & 0 & \dots & 0 \\ 0 & d_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & d_{kk} \end{bmatrix} \quad (2.72)$$

De maneira análoga, a consideração de k componentes principais implica na simplificação das matrizes de *loads* \mathbf{U}_{nm}^* e \mathbf{P}_{pp}^* e da matriz de *scores* \mathbf{T}^* , tal que:

$$\mathbf{U}_{nk}^{**} = \begin{bmatrix} u_{11} & u_{12} & \cdots & u_{1k} \\ u_{21} & u_{22} & \cdots & u_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ u_{n1} & u_{n2} & \cdots & u_{nk} \end{bmatrix} \quad (2.73)$$

$$\mathbf{T}^{**} = \mathbf{U}_{nk}^{**} \mathbf{D}^{**} \quad (2.74)$$

$$\mathbf{P}_{pk}^{**} = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1k} \\ p_{21} & p_{22} & \cdots & p_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ p_{p1} & p_{p2} & \cdots & p_{pk} \end{bmatrix} \quad (2.75)$$

A regressão de componentes principais é dada pelas equações (2.76) e (2.77):

$$\mathbf{y} = \mathbf{X} \boldsymbol{\beta}_{PCR} + \boldsymbol{\varepsilon} \quad (2.76)$$

$$\hat{\boldsymbol{\beta}}_{PCR} = \mathbf{P}_{pk}^{**} \left((\mathbf{T}^{**})^T \mathbf{T}^{**} \right)^{-1} (\mathbf{T}^{**})^T \mathbf{y} \quad (2.77)$$

A predição do modelo de PCR é dada pela equação (2.78):

$$\hat{\mathbf{y}} = \mathbf{X} \hat{\boldsymbol{\beta}}_{PCR} \quad (2.78)$$

Um aspecto característico importante deste método é a construção das componentes principais utilizando-se unicamente as informações presentes nos regressores de entrada, conforme equação (2.69), sem levar em consideração a informação contida na variável resposta durante a etapa de obtenção destas componentes. Isto constitui uma fragilidade do método, dado que informações preditivas úteis podem ser perdidas. Deste modo, pode ser necessária uma maior adição de componentes para melhor descrição da variável resposta.

2.4.4.2 Mínimos quadrados parciais (PLS)

A técnica de regressão de mínimos quadrados parciais (*Partial Least Square*), também denominada projeção em variáveis latentes (*Projections to Latent Structures*), desenvolvida por H. Wold e colaboradores (ROSIPAL; KRÄMER, 2006, p. 34), é muito utilizada na quimiometria para fins de calibração multivariada. Consiste na relação entre dois conjuntos de variáveis, $(\mathbf{X})_{n \times p}$ e $(\mathbf{Y})_{q \times u}$, através de um modelo linear.

Em linhas gerais, a técnica de PLS cria vetores de *scores*, denominados variáveis latentes ou componentes, pela maximização da covariância entre diferentes conjuntos de variáveis. A regressão linear multivariada pode ser compreendida como um caso particular dos mínimos quadrados parciais, em que o número de variáveis latentes se iguala ao número de regressores originais, presentes no conjunto de dados.

Esta técnica apresenta grande robustez para tratar dados ruidosos e com elevada colinearidade. Ao contrário da técnica de PCR, ela utiliza informações presentes na variável resposta para a obtenção das direções das variáveis latentes T e U . Geometricamente, a técnica de PLS pode ser compreendida como a projeção da matriz de dados originais \mathbf{X} , contendo n amostragens e p variáveis regressoras, em um hiperplano de k dimensões, tal que $k < p$, conforme Figura 2.18:

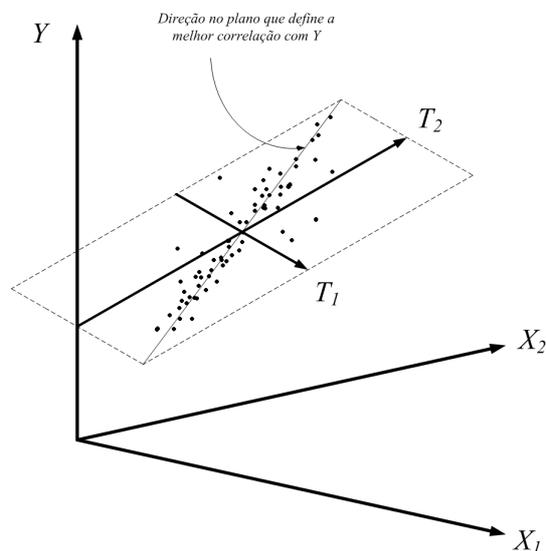


Figura 2.18: Interpretação geométrica da técnica de PLS (adaptado de Wold et al., (2004, p.32)).

As projeções \mathbf{T} no plano são correlacionadas com a resposta \mathbf{Y} . Matematicamente, este procedimento pode ser expresso por (ROSIPAL; KRÄMER, 2006, p. 39):

$$\mathbf{X} = \mathbf{T} \mathbf{P}^T + \boldsymbol{\varepsilon}_X \quad (2.79)$$

$$\mathbf{Y} = \mathbf{U} \mathbf{Q}^T + \boldsymbol{\varepsilon}_Y \quad (2.80)$$

$$\mathbf{U} = \mathbf{T} \mathbf{D}_{PLS} + \boldsymbol{\varepsilon}_H \quad (2.81)$$

Substituindo-se a equação (2.81) em (2.80), obtêm-se:

$$\mathbf{Y} = \mathbf{T} \mathbf{D}_{PLS} \mathbf{Q}^T + (\boldsymbol{\varepsilon}_H \mathbf{Q}^T + \boldsymbol{\varepsilon}_Y) \quad (2.82)$$

$$\mathbf{Y} = \mathbf{T} \mathbf{D}_{PLS} \mathbf{Q}^T + \boldsymbol{\varepsilon}_F \quad (2.83)$$

Rosipal e Krämer (2006, p. 40), demonstram matematicamente que, considerando-se apenas uma variável de saída:

$$\mathbf{y} = \mathbf{X} \hat{\boldsymbol{\beta}}_{PLS} + \boldsymbol{\varepsilon}_F \quad (2.84)$$

$$\hat{\boldsymbol{\beta}}_{PLS} = \mathbf{W}(\mathbf{P}^T \mathbf{W})^{-1} \mathbf{q} \quad (2.85)$$

Os dados $\hat{\mathbf{y}}$ previstos pelo modelo são expressos por:

$$\hat{\mathbf{y}} = \mathbf{X} \hat{\boldsymbol{\beta}}_{PLS} \quad (2.86)$$

A técnica de PLS se baseia no algoritmo NIPALS (*Nonlinear Iterative Partial Least Square*) para obtenção das matrizes \mathbf{W} , \mathbf{P} e \mathbf{q} . O algoritmo apresenta os seguintes passos (Figura 2.19):

NIPALS

Dados de entrada : matriz de dados X , vetor resposta y , número máximo de componentes (variáveis latentes) c e a tolerância ε ;

$$\mathbf{X}_1 \leftarrow \mathbf{X}$$

$$\mathbf{w}_1 \leftarrow \frac{\mathbf{X}_1^T \mathbf{y}}{\|\mathbf{X}_1^T \mathbf{y}\|}$$

$$\mathbf{T}_1 \leftarrow \mathbf{X}_1 \mathbf{w}_1$$

para $k = 1$ até c

$$t_k \leftarrow \mathbf{T}_k^T \mathbf{T}_k$$

$$\mathbf{T}_k \leftarrow \left(\frac{1}{t_k} \right) \mathbf{T}_k$$

$$\mathbf{p}_k \leftarrow \mathbf{X}_k^T \mathbf{T}_k$$

$$q_k \leftarrow \mathbf{y}^T \mathbf{T}_k$$

se $|q_k| < \varepsilon$, então

$$c \leftarrow k$$

para e sai do loop

do contrário

se $k < c$, então

$$\mathbf{X}_{k+1} \leftarrow \mathbf{X}_k - t_k (\mathbf{T}_k \mathbf{p}_k^T)$$

$$\mathbf{w}_{k+1} \leftarrow \mathbf{X}_{k+1}^T \mathbf{y}$$

$$\mathbf{T}_{k+1} \leftarrow \mathbf{X}_{k+1} \mathbf{w}_{k+1}$$

do contrário

para e sai do loop

fim se

fim se

fim para

$$\mathbf{W} \leftarrow [\mathbf{w}_1 \quad \mathbf{w}_2 \quad \cdots \quad \mathbf{w}_c]$$

$$\mathbf{P} \leftarrow [\mathbf{p}_1 \quad \mathbf{p}_2 \quad \cdots \quad \mathbf{p}_c]$$

$$\mathbf{q} \leftarrow \begin{bmatrix} q_1 \\ q_2 \\ \vdots \\ q_c \end{bmatrix}$$

Figura 2.19: Etapas sucessivas do algoritmo NIPALS.

2.4.5 Redes neurais artificiais (RNA)

Segundo Braga, Carvalho e Ludermir (2007, p.3), as redes neurais artificiais (RNA's) constituem sistemas paralelos distribuídos, compostos por unidades de processamento simples, denominadas neurônios. Estas unidades calculam funções matemáticas, normalmente não lineares, e são dispostas em camadas interligadas. A cada neurônio está associado um conjunto de dados de entrada com seus respectivos pesos. O ajuste desses pesos possibilita um mapeamento não linear entre os dados de entrada e a saída desejada.

A capacidade de aprender por meio de exemplos é uma das características mais atrativas das redes neurais, permitindo uma grande aplicabilidade em sistemas industriais. As RNA's são capazes de extrair informações não apresentadas de forma explícita através dos exemplos, podendo atuar como mapeadores universais de funções multivariáveis. Esquerre (2003, p.45) afirma que as redes neurais fornecem um poderoso conjunto de ferramentas para a solução de problemas de reconhecimento de padrões, processamento de dados, modelagem e controle não linear.

2.4.5.1 Modelo de neurônio McCulloch e Pitts (MCP)

O neurônio biológico é formado por três partes fundamentais: corpo celular, dendritos e axônios (BRAGA; CARVALHO; LUDERMIR, 2007, p.6). Ele recebe impulsos nervosos de outros neurônios através dos dendritos (receptores), sendo a informação processada no corpo celular. Impulsos nervosos gerados neste neurônio passam através do axônio e são transmitidos aos dendritos de outros neurônios da rede biológica. As conexões entre os neurônios são denominadas conexões sinápticas. Quando um impulso nervoso chega às conexões sinápticas, neurotransmissores modificam o potencial elétrico das membranas, inibindo ou promovendo a passagem do impulso (ESQUERRE, 2003, p.48). A RNA é um instrumento computacional que procura imitar o sistema de aprendizagem do cérebro humano. A essência de sistemas neurais biológicos está na capacidade de modificar respostas como um resultado de exposição a sinais externos. Este princípio é válido para o neurônio artificial proposto por McCulloch e Pitts (neurônio MCP) em 1943.

A descrição matemática deste tipo de neurônio resultou em um modelo composto por n terminais de entrada, os quais representam o análogo matemático dos dendritos. A cada terminal está associado um peso w_i e a cada neurônio MCP está associada apenas uma saída, conforme Figura 2.20:

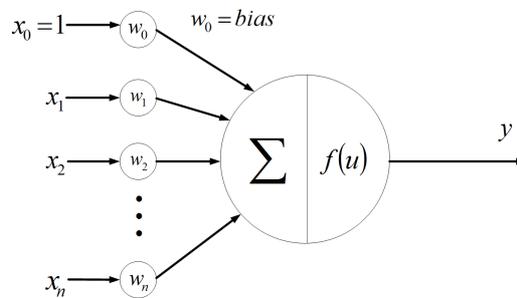


Figura 2.20: Neurônio MCP (BRAGA; CARVALHO; LUDERMIR, 2007, p.8)

Pesos positivos correspondem a sinais de excitação; já os pesos negativos exercem função de inibição do neurônio. Os pesos determinam a proporção em que um neurônio deve considerar sinais de excitação ou inibição oriundos de uma conexão específica. O produto da entrada x_i pelo peso w_i é o análogo matemático de uma sinapse nervosa.

McCulloch e Pitts simplificaram o modelo do neurônio artificial ao considerar uma camada com disparo síncrono. Sistemas biológicos apresentam uma complexidade superior, pois os disparos não são necessariamente síncronos, como também não ocorrem apenas em tempos discretos. Além disso, os disparos dos neurônios biológicos são dependentes também das ativações em estados anteriores.

2.4.5.2 Funções de ativação

A função de ativação processa as sinapses produzindo uma saída. No caso de neurônios MCP, as entradas são multiplicadas pelos pesos e então somadas (Figura 2.20). Este sinal é comparado a um limiar de ativação θ , de acordo com a equação:

$$f(u) = \begin{cases} 1 & \sum_{i=1}^n x_i w_i \geq \theta \\ 0 & \sum_{i=1}^n x_i w_i < \theta \end{cases} \quad (2.87)$$

Trata-se, portanto, de uma função de ativação do tipo degrau, transladada por um *bias* de θ . Considerando-se um sinal de entrada constante, $x_0 = 1$, multiplicado por um peso $w_0 = -\theta$, a equação (2.87) pode ser matematicamente expressa por:

$$f(u) = \begin{cases} 1 & \sum_{i=0}^n x_i w_i \geq 0 \\ 0 & \sum_{i=0}^n x_i w_i < 0 \end{cases} \quad (2.88)$$

A função degrau apresenta uma descontinuidade em sua função derivada, o que pode representar um problema na utilização de algoritmos de aprendizagem que dependam do cálculo do gradiente. Uma aproximação para a função degrau, diferenciável em todos os pontos, é a função logaritmo sigmoide:

$$f(u) = \frac{1}{1 + e^{-\beta(u)}} \quad (2.89)$$

A variável u é definida por:

$$u = \sum_{i=0}^n x_i w_i \quad (2.90)$$

A função logaritmo sigmoide apresenta uma região semilinear que auxilia na aproximação de funções contínuas (BRAGA; CARVALHO; LUDERMIR, 2007, p.9). Funções de ativação lineares são utilizadas no modelo *Adaptive Linear* (ADALINE) (cf. BRAGA; CARVALHO; LUDERMIR, 2007, p.39):

$$f(u) = u \quad (2.91)$$

O neurônio MCP é a unidade de processamento das redes de *Perceptron* de múltiplas camadas (MLP). Nestas redes, o argumento das funções de ativação é definido como o produto interno dos vetores de entrada pelos vetores de pesos (BRAGA; CARVALHO; LUDERMIR, 2007, p.170). No caso das redes de base radial (RBF), as quais utilizam funções de ativação radiais, o argumento de entrada nestas funções é a distância euclidiana entre os vetores de entrada e o vetor de pesos, conforme equação (2.92):

$$y_j = \sum_{i=1}^k w_{ji} \left(f \left(\left\| \mathbf{x}_j^T - \boldsymbol{\mu}^T \right\| \right) \right) + w_{j0} \quad (2.92)$$

Estas redes separam o espaço de padrões de entrada pela definição de hiperelipsoides, definindo aproximadores locais. Já redes MLP separam o espaço por hiperplanos e, portanto, se constituem em mapeadores globais com maior capacidade de generalização. No entanto a maior capacidade de generalização implica em um maior risco de classificação errônea dos dados de entrada (problema dos ‘falsos padrões’). A função gaussiana é um exemplo de função de ativação radial:

$$f(u) = \exp\left(\frac{-(u - \mu)^2}{r^2}\right) \quad (2.93)$$

As funções mencionadas são exibidas graficamente na Figura 2.21.

2.4.5.3 Aprendizado

Uma das características principais das redes neurais é a sua capacidade de aprendizagem a partir do ajuste das intensidades das conexões entre neurônios. Segundo Haykin apud Mendel e McClaren (2001, p.75) define-se o processo de aprendizagem dentro do contexto de redes neurais por:

... processo pelo qual os parâmetros livres de uma rede neural são adaptados através de um processo de estimulação pelo ambiente no qual a rede está inserida. O tipo de aprendizagem é determinado pela maneira pela qual a modificação dos parâmetros ocorre.

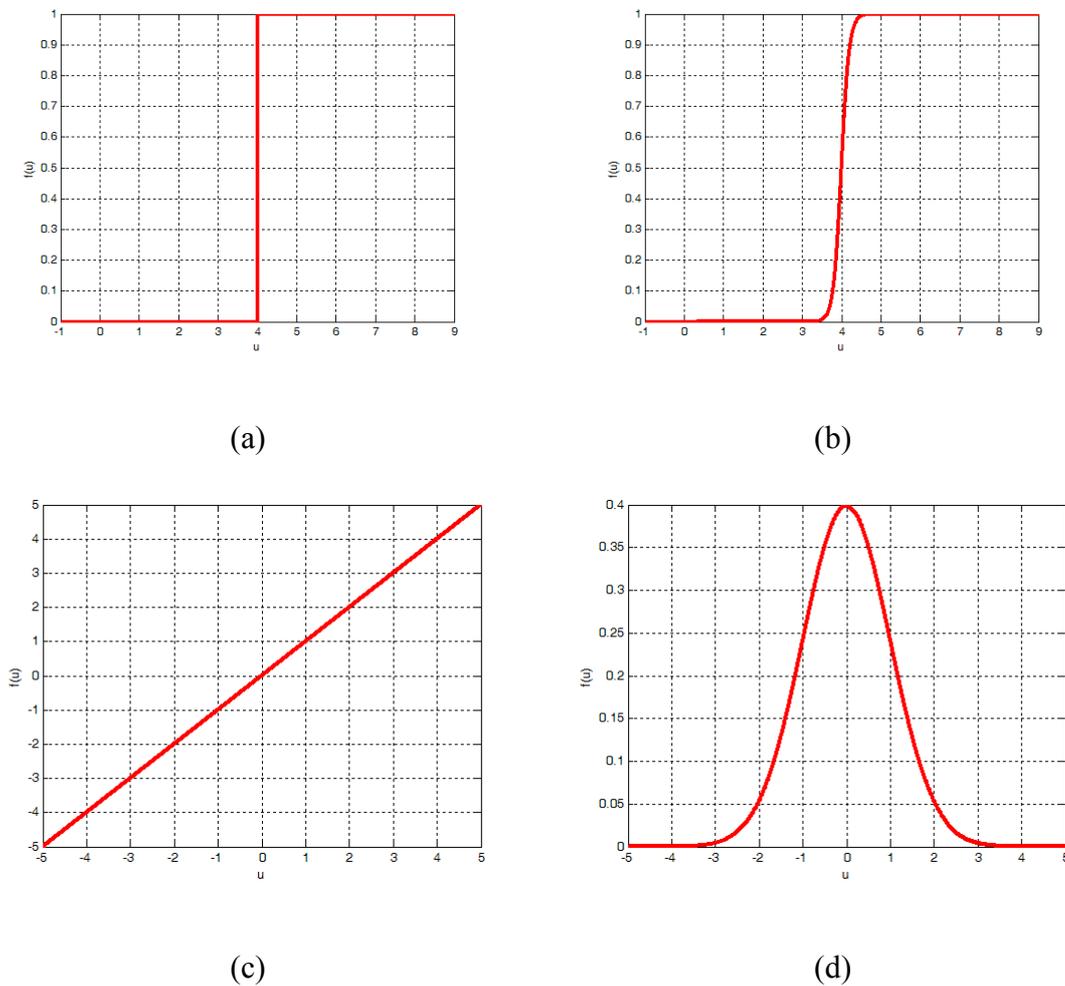


Figura 2.21: Principais funções de ativação utilizadas. a) função degrau; b) função logaritmo sigmoide; c) função linear; d) função gaussiana (adaptado de Braga; Carvalho; Ludemir, (2007, p.10)).

Braga, Carvalho e Ludemir (2007, p.12) ressaltam que o conceito de aprendizado se relaciona à melhoria de algum critério de performance estabelecido, tal como o erro quadrático. De uma forma geral, o ajuste dos pesos pode ser matematicamente expresso por:

$$\mathbf{w}(t+1) = \mathbf{w}(t) + \Delta\mathbf{w}(t) \quad (2.94)$$

Os algoritmos de aprendizagem diferem fundamentalmente na forma como $\Delta\mathbf{w}(t)$ é calculado.

Na forma de aprendizado dito supervisionado existe a figura de um professor, que apresenta para a rede exemplos de entrada e saída (i.e dados utilizados na etapa de treinamento). A resposta da rede é obtida tendo em vista o vetor de pesos $\mathbf{w}(t)$. O erro calculado é retroalimentado para a obtenção do ajuste $\Delta\mathbf{w}(t)$, conforme Figura 2.22:

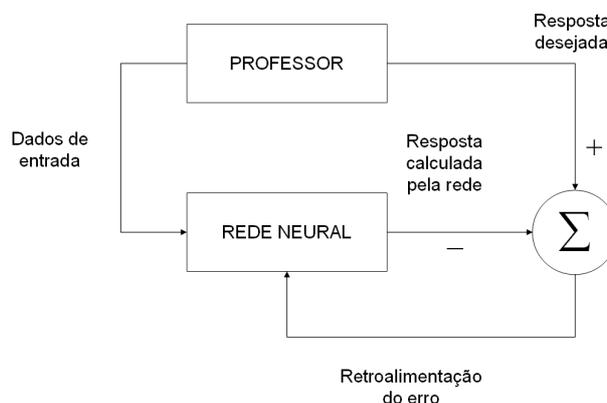


Figura 2.22: Diagrama esquemático de aprendizado supervisionado (BRAGA; CARVALHO; LUDERMIR, 2007, p.13)

No caso do aprendizado não supervisionado não existe a figura de um professor, conforme expresso pela Figura 2.23. Neste caso, a existência de regularidade e redundância nos padrões de entrada torna possível o ajuste dos pesos sinápticos independentemente de um critério de desempenho da resposta da rede. Essa metodologia é empregada em problemas de categorização, no qual as classes não estão definidas de antemão (BRAGA; CARVALHO; LUDERMIR, 2007, p.18).



Figura 2.23: Diagrama esquemático de aprendizado não supervisionado (BRAGA; CARVALHO; LUDERMIR 2007, p.17)

2.4.5.4 Arquitetura de redes neurais artificiais (RNA's)

A definição da arquitetura da rede neural implica na determinação do número de camadas da rede, número de nós em cada camada e o tipo de conexão entre os nós (MATTOS, 2004, p.46). A Figura 2.24 exhibe uma rede *feedforward*, que é caracterizada

pela existência de apenas um sentido de estímulo dos neurônios. Além disso, ela pode ser classificada também como estática, porque as saídas y_i , para um dado instante, dependem apenas das entradas x_i atuais (BRAGA; CARVALHO; LUDERMIR, 2007, p.12).

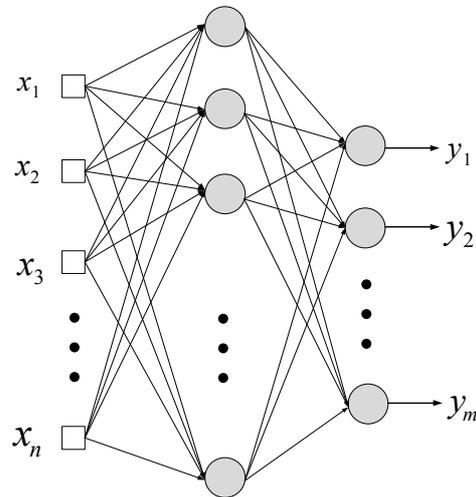


Figura 2.24: Rede feedforward com uma camada oculta.

A Figura 2.25 exibe uma rede com recorrência entre as saídas e a camada oculta. A Figura 2.26 exibe uma rede com recorrência autoassociativa, em que não se verificam entradas externas, estrutura típica de uma rede de Hopfield.

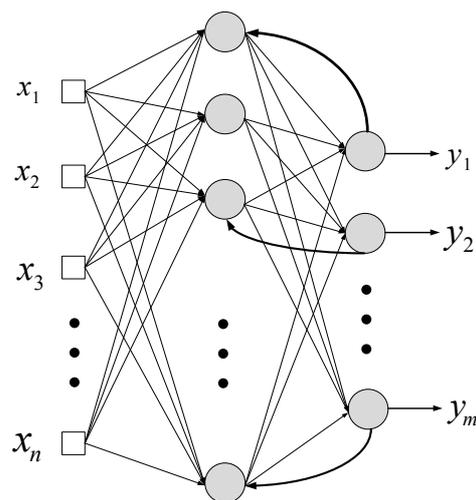


Figura 2.25: Rede feedforward com uma camada oculta, apresentando recorrência entre a camada de saída e a camada oculta.

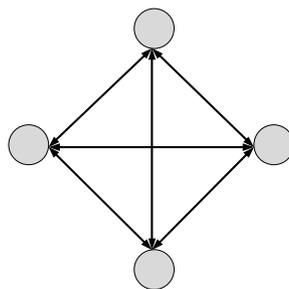


Figura 2.26: Rede neural autoassociativa.

2.4.5.4.1 Número de camadas

A conexão de neurônios artificiais apresenta grande capacidade de computação. Segundo Cybenko (1989, p.303), redes com uma camada intermediária conseguem aproximar qualquer função contínua. Utilizam-se duas camadas intermediárias quando a função a ser aproximada apresenta descontinuidades (BRAGA; CARVALHO; LUDERMIR, 2007, p.70). Em geral, problemas práticos necessitam de apenas uma camada intermediária. No caso das redes MLP com duas camadas intermediárias, cada neurônio da primeira contribui com retas para a formação da superfície do espaço de entrada. Na segunda camada essas retas são combinadas, formando regiões convexas. A camada linear de saída combina essas regiões convexas, definindo diversos formatos no espaço (BRAGA; CARVALHO; LUDERMIR, 2007, p.69).

2.4.5.4.2 Número de neurônios

A capacidade de generalização da rede é influenciada por três fatores fundamentais (HAYKIN, 2001, p.233):

1. Tamanho e representatividade do conjunto de treinamento;
2. Escolha da arquitetura da rede neural;
3. Complexidade do fenômeno modelado.

O número de neurônios utilizados nas camadas intermediárias influencia na qualidade da aproximação obtida. O uso excessivo de neurônios pode causar *'overfitting'*, o que significa que a rede neural superestima a complexidade do problema. O conceito de *'overfitting'* se refere ao ajuste excessivo de exemplos específicos apresentados à rede. Neste caso, o ruído presente nos dados também é modelado, contribuindo para a perda

da capacidade de generalização. Esse fato pode ser verificado pelo erro excessivo obtido entre os dados de teste (i.e dados avaliados pela rede, que são diferentes daqueles utilizados na etapa de treinamento) e as previsões do modelo. Além do uso excessivo de neurônios, o *'overtraining'*, ou seja, o uso de muitas iterações no procedimento de treinamento da rede, também pode causar *'overfitting'* (AMARI et al., 1997, p.985).

Braga, Carvalho e Ludemir. (2007, p.71) ressaltam que, embora haja na literatura inúmeras abordagens que buscam determinar o tamanho ótimo da rede, até então não foi encontrada uma resposta formal para esta questão. Segundo Rafiq, Bugmann e Easterbrook (2001, p.1545), alguns autores sugerem que o número máximo de neurônios na camada oculta não deve exceder em mais que uma unidade o dobro de variáveis regressoras, mas que esta regra não garante a capacidade de generalização da rede.

Um dos métodos utilizados frequentemente para a determinação do número ótimo de neurônios é a validação cruzada. Neste método, os dados são divididos em dois conjuntos distintos: dados de treinamento e dados de teste (HAYKIN, 2001, p.240). Na validação cruzada *k-fold*, o conjunto de dados originais é subdividido em k conjuntos com aproximadamente o mesmo tamanho. A rede é treinada k vezes, sendo que a cada ciclo um dos conjuntos é deixado de fora. Este conjunto é utilizado para verificação do erro de teste. Um critério de performance, tal como a soma quadrática dos erros, pode ser avaliado, considerando-se todos os k testes realizados. Repete-se este procedimento acrescentando-se neurônios, gradativamente, às camadas ocultas. A arquitetura que fornecer o melhor critério de performance, tendo em vista k testes realizados em cada uma dessas etapas, será escolhida.

No método de validação cruzada, se o valor k iguala o número de amostras presentes nos dados de treinamento, o procedimento é dito *'leave-one-out'*. O procedimento *'leave-v-out'* é mais elaborado e com maior custo computacional, envolvendo todas as possibilidades de subconjuntos contendo v amostras de treinamento.

A validação cruzada difere do procedimento *'split-sample'*, utilizado normalmente nos métodos de parada antecipada para treinamento de redes. No caso da parada antecipada, o conjunto de treinamento é, ainda, subdividido em dados de estimação e dados de validação, sendo que apenas um único conjunto de dados de validação é utilizado no

procedimento; diferentemente da validação cruzada k-fold, em que k conjuntos são utilizados na avaliação de um critério de performance.

Na parada antecipada, os dados de estimação são utilizados para a obtenção dos pesos da rede, considerando-se uma arquitetura fixa (i.e número de neurônios fixos para as camadas ocultas). Os dados de validação são utilizados para aceitar um determinado conjunto de pesos, tendo em vista que os algoritmos de treinamento são normalmente iterativos e realizam ajustes sequenciais dos mesmos. A performance da aproximação da função é verificada após cada época. Caracteriza-se por época a apresentação completa de todos os elementos do conjunto de treinamento, seguida dos ajustes de pesos correspondentes (BRAGA; CARVALHO; LUDERMIR 2007, p.42). O treinamento é interrompido após cada época e os dados de validação são, então, avaliados. No momento em que o erro da validação começa a aumentar com o acréscimo do número de épocas, é possível que a rede esteja começando a modelar o ruído presente nos dados, caracterizando '*overtraining*'.

A técnica computacional denominada *bootstrapping* (EFRON e TIBSHIRANI, 1995, p.1) apresenta, frequentemente, melhor desempenho quando comparada à validação cruzada. Na forma mais simples do *bootstrapping*, ao invés de se analisar subgrupos de dados, amostras aleatórias são sorteadas, com reposição repetidas vezes, para estimativa do erro de teste em cada tipo de arquitetura. Não apenas o erro de teste pode ser estimado, como também intervalos de confiança para a saída da rede.

2.4.5.5 Abordagens para o treinamento de redes de *Perceptron* de múltiplas camadas (MLP)

Duas abordagens distintas comumente utilizadas para o treinamento de redes MLP são as técnicas de retropropagação de erros (algoritmo *back-propagation*) e as técnicas de regularização, em especial a regularização Bayesiana. Ambas serão brevemente comentadas a seguir.

2.4.5.5.1 Algoritmo *back-propagation*

Trata-se de um algoritmo de aprendizado supervisionado que opera em duas fases. A primeira fase é denominada *forward* e diz respeito ao sentido de fluxo de processamento

dos sinais de entrada pela rede. A segunda fase é denominada *backward* e diz respeito ao fluxo de erro calculado a partir da saída \hat{y}_j , conforme Figura 2.27.

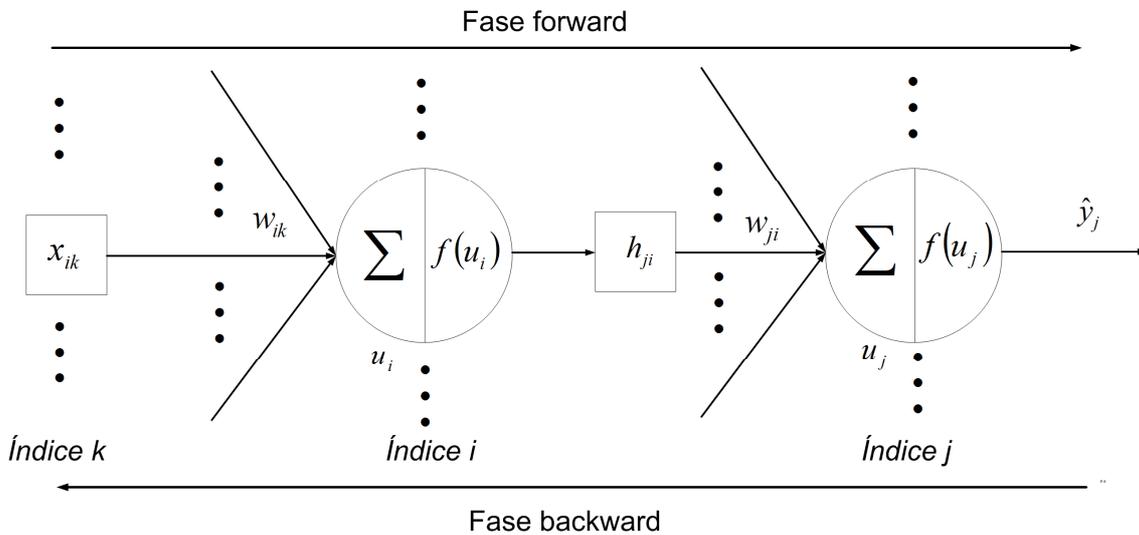


Figura 2.27: Diagrama esquemático do algoritmo back-propagation (adaptado de Braga; Carvalho; Ludermir, 2007, p.76).

Busca-se identificar o mínimo global da superfície criada pelo erro quadrático em função dos pesos da rede, conforme equação (2.95):

$$\varepsilon(n) = \frac{1}{2} \sum_j (y_j(n) - f(u_j(n)))^2 \quad (2.95)$$

As equações descritas a seguir se aplicam ao caso de redes MLP com uma camada oculta. Para detalhes das deduções das equações de ajuste, cf. Braga, Carvalho e Ludermir (2007, p.75).

Na fase *forward*, os dados de treinamento são avaliados pela rede para obtenção da saída \hat{y}_j . A cada iteração n , avalia-se uma amostra do conjunto de dados de treinamento $\mathbf{X}_{n \times p}$, considerando-se os pesos e *bias* da rede ajustados na iteração $n-1$. Caso esta seja a primeira iteração do algoritmo, consideram-se os pesos e *bias* de inicialização. A saída calculada $\hat{y}_j(n)$ é comparada, a cada n -ésima iteração, com a saída desejada $y_j^D(n)$, obtida do conjunto de dados de treinamento. O erro $e(n)$ é, então, calculado.

Na fase *backward*, o erro $e(n)$ calculado é utilizado para ajustar o vetor de pesos da camada de saída em sentido oposto ao gradiente do erro quadrático. Considerando-se uma rede MLP com uma camada oculta, a equação de ajuste para os pesos w_{ji} da camada de saída pode ser expressa segundo a seguinte regra (HAYKIN, 2001, p.193):

$$\left(\begin{array}{c} \text{Correção de peso} \\ \Delta w_{ji} \end{array} \right) = \left[\begin{array}{c} \text{Taxa de} \\ \text{aprendizagem} \\ (\eta) \end{array} \right] \left[\begin{array}{c} \text{Gradiente local} \\ \delta_j \end{array} \right] \left[\begin{array}{c} \text{Sinal de entrada} \\ \text{oriundo do neurônio } i \\ \text{para o neurônio } j \\ (h_{ji}) \end{array} \right]$$

Esta expressão apresenta analogia com a equação de ajuste dos pesos de redes *Adaptive Linear* (ADALINE), conhecida com regra Delta (BRAGA; CARVALHO; LUDERMIR, 2007, p.79). Matematicamente, a regra anterior é expressa pela equação (2.96):

$$\Delta w_{ji} = [\eta] \left[e_j(n) \left(\frac{d}{du_j} f(u_j(n)) \right) \right] [h_i(n)] \quad (2.96)$$

De maneira semelhante, o ajuste dos pesos da camada oculta é expresso pela mesma regra anterior, conforme demonstra equação (2.97):

$$\left(\begin{array}{c} \text{Correção de peso} \\ \Delta w_{ik} \end{array} \right) = \left[\begin{array}{c} \text{Taxa de} \\ \text{aprendizagem} \\ (\eta) \end{array} \right] \left[\begin{array}{c} \text{Gradiente local} \\ \delta_i \end{array} \right] \left[\begin{array}{c} \text{Sinal oriundo} \\ \text{da entrada } k \\ \text{para o neurônio } i \\ (x_{ik}) \end{array} \right]$$

$$\Delta w_{ik} = [\eta] \left[\left(\frac{d}{du_i} f(h_i(u_i(n))) \right) \sum_j \left(e_j(n) \left(\frac{d}{du_j} f(u_j(n)) \right) w_{ji}(n) \right) \right] [x_{ik}(n)] \quad (2.97)$$

No entanto, o vetor gradiente local δ_i é calculado a partir de uma soma ponderada dos erros dos neurônios j , da camada de saída, que se conectam ao neurônio i . Desta forma ocorre uma retropropagação dos erros.

O algoritmo *back-propagation* pode apresentar variações que objetivam acelerar a convergência ou mesmo reduzir sua sensibilidade a mínimos locais. Como exemplo,

cita-se a adição de um termo *momentum* (α), expresso pela equação (2.98), ao ajuste dos pesos, o que resulta na aceleração da convergência do algoritmo em regiões de maior declínio da superfície do erro quadrático. No entanto, a adição deste termo pode também ocasionar instabilidade ao algoritmo (BRAGA; CARVALHO; LUDERMIR, 2007, p.80).

$$w_{ik}(t+1) = w_{ik}(t) + \eta \delta_i(t) x_{ik}(t) + \alpha (w_{ik}(t) - w_{ik}(t-1)) \quad (2.98)$$

Outras estratégias incluem utilização de taxa de aprendizado variável. Dentre as variações do algoritmo *back-propagation*, citam-se os algoritmos Quickprop e Rprop (BRAGA; CARVALHO; LUDERMIR, 2007, p.81). O algoritmo Rprop será brevemente descrito na sequência.

2.4.5.5.2 Algoritmo *back-propagation* resiliente (Rprop)

O algoritmo *back-propagation* resiliente (Rprop) consiste de uma modificação do *back-propagation* tradicional para eliminar a influência negativa do valor da derivada parcial na etapa de ajuste dos pesos (BRAGA; CARVALHO; LUDERMIR, 2007, p.82). Nos casos em que a saída do neurônio encontra-se próxima à saturação, ou seja, próxima à 0 ou 1, a derivada parcial será próxima de 0, resultando em um ajuste mínimo para o vetor de peso. Este problema é minimizado pelo algoritmo ao se utilizar apenas o sinal da derivada ao invés do seu valor, de acordo com a equação (2.99)

$$\Delta w_{ji}(t) = \begin{cases} -\Delta_{ji}(t), & \text{se } \frac{\partial E(t)}{\partial w_{ji}} > 0 \\ +\Delta_{ji}(t), & \text{se } \frac{\partial E(t)}{\partial w_{ji}} < 0 \\ 0, & \text{se } \frac{\partial E(t)}{\partial w_{ji}} = 0 \end{cases} \quad (2.99)$$

O valor Δ_{ji} é definido por um processo adaptativo, conforme equação (2.100)

$$\Delta_{ji}(t) = \begin{cases} \eta^+ \Delta_{ji}(t-1), & \text{se } \frac{\partial E(t-1)}{\partial w_{ji}} \frac{\partial E(t)}{\partial w_{ji}} > 0 \\ \eta^- \Delta_{ji}(t-1), & \text{se } \frac{\partial E(t-1)}{\partial w_{ji}} \frac{\partial E(t)}{\partial w_{ji}} < 0 \\ \Delta_{ji}(t-1), & \frac{\partial E(t)}{\partial w_{ji}} = 0 \end{cases} \quad (2.100)$$

em que $0 < \eta^- < 1 < \eta^+$.

2.4.5.5.3 Regularização Bayesiana combinada com algoritmo de treinamento Levenberg-Marquardt

A aprendizagem pode ser interpretada como “... um problema de reconstrução de uma hipersuperfície, dado um conjunto de pontos de dados que podem ser esparsos.” (HAYKIN, 2001, p.293). Este problema é dito bem formulado ou ‘bem posto’, no sentido de Hadamard, caso três condições sejam satisfeitas simultaneamente:

1. Existência: para cada vetor de entrada $\mathbf{x} \in \chi$, existe uma saída $y = f(\mathbf{x})$ onde $y \in \psi$, onde χ e ψ são, respectivamente, os domínios de entrada e saída;
2. Unicidade: para qualquer par de vetores de entrada $\mathbf{x}, \mathbf{t} \in \chi$, temos $f(\mathbf{x}) = f(\mathbf{t})$ se e somente se $\mathbf{x} = \mathbf{t}$;
3. Continuidade ou estabilidade: para qualquer $\varepsilon > 0$ existe $\delta = \delta(\varepsilon)$ tal que a distância $d(\mathbf{x}, \mathbf{t}) < \delta$ (i.e distância entre os argumentos \mathbf{x} e \mathbf{t} no domínio χ) implica em $d(f(\mathbf{x}), f(\mathbf{t})) < \varepsilon$ (i.e distância do mapeamento de \mathbf{x} e \mathbf{t} no domínio ψ).

O fenômeno físico responsável pela geração dos dados atende às condições de um problema bem formulado. No entanto, a reconstrução da hipersuperfície a partir dos dados, o que caracteriza o problema inverso, pode ser mal formulada. Dentre as causas pode-se citar a insuficiência de informação nos dados, o que pode prejudicar a reconstrução do mapeamento unívoco. Haykin apud Lanczos (2001, p.294) ressalta que “... falta de informação não pode ser remediada por nenhum truque matemático.”. Além disso, um nível de ruído elevado nos dados pode resultar em um mapeamento do vetor \mathbf{x} fora do espaço de soluções ψ , violando o critério de

continuidade. Este último problema, em especial, pode ser contornado caso alguma informação prévia sobre o mapeamento de entrada e saída esteja disponível.

Cabe ressaltar que a modelagem de processos industriais baseada em técnicas de identificação tipo ‘caixa preta’ pode resultar em um problema mal formulado. Variáveis importantes do ponto de vista físico-químico, e que sabidamente influenciam na variável resposta, podem não ser suficientemente sensibilizadas durante a etapa de coleta de dados. Sendo assim, informações importantes não são incorporadas ao modelo. Este fato resulta, já na etapa de seleção de variáveis, na exclusão de variáveis significativas, dado que as mesmas não apresentam conteúdo de informação suficiente.

O método de regularização, proposto por Tikhonov em 1963, propõe estabilizar a solução (i.e garantir continuidade) por meio de um funcional que incorpore informação prévia sobre a mesma. A forma mais comum de informação prévia incorporada consiste na suposição de suavidade, ou seja, entradas similares devem produzir saídas similares.

A teoria de regularização de Tikhonov envolve dois termos:

1. Termo do erro padrão, dado pela equação (2.101);

$$\varepsilon_E = \frac{1}{2} \sum_j (y_j - F(\mathbf{x}_j, \mathbf{w}))^2 \quad (2.101)$$

2. Termo de regularização, dado pela equação (2.102):

$$\varepsilon_R = \frac{1}{2} \|\mathcal{D}(F)\|^2 \quad (2.102)$$

O termo de regularização apresenta um operador diferencial linear \mathcal{D} aplicado à função $F(\mathbf{x}, \mathbf{w})$, responsável por garantir continuidade à solução. A função a ser minimizada, neste caso, é dada pela equação (2.103):

$$\varepsilon_F = \varepsilon_E + \lambda_R \varepsilon_R \quad (2.103)$$

Percebe-se que o termo de regularização ε_R atua com uma “navalha de Occam”, pois os modelos mais complexos, que possuem maior quantidade de parâmetros livres w_i , são punidos na função objetivo ε_F . Sjöberg e Ljung (1992, p.6) ressaltam que determinados parâmetros podem contribuir mais com a variabilidade do erro do que com a efetividade do ajuste, sendo caracterizados como ‘supérfluos’, devendo os mesmos ser removidos do modelo.

O parâmetro de regularização λ_R caracteriza a suficiência de informação nos dados para determinação dos parâmetros da função $F(\mathbf{x}, \mathbf{w})$. Quando λ_R tende a zero, os exemplos determinam a solução do problema de forma irrestrita. Já quando λ_R tende a infinito, a restrição de suavidade é, por si só, suficiente para determinar a solução de $F(\mathbf{x}, \mathbf{w})$, o que significa que os dados coletados não são confiáveis. Para detalhes da obtenção da estimativa ótima para o parâmetro de regularização λ_R , cf. Haykin (2001, p.314).

Mackay (1992a, p.418) aplicou a técnica de regularização Bayesiana nas etapas de modelagem e seleção de arquitetura de redes. A função objetivo a ser minimizada é expressa pela equação (2.104), que resulta matematicamente na expressão (2.105) (MACKAY, 1992b, p.449).

$$\varepsilon_F = \beta^B \varepsilon_E + \alpha^B \varepsilon_R \quad (2.104)$$

$$\varepsilon_F = \frac{\beta^B}{2} \sum_j [(y_j - f(\mathbf{x}_j, \mathbf{w}))^2] + \frac{\alpha^B}{2} \sum_i w_i^2 \quad (2.105)$$

Para a variável resposta observada y_i , a abordagem Bayesiana considera a presença de um ruído branco v aditivo na saída, com média nula e variância σ_v^2 , conforme equação (2.106):

$$y_i = f(\mathbf{x}_i, \mathbf{w}) + v_i \quad (2.106)$$

Considera-se também tanto o vetor de parâmetros \mathbf{w} da rede quanto o ruído v observado na variável resposta \mathbf{y} como variáveis aleatórias com distribuição normal. Sendo assim, é possível expressar matematicamente as respectivas distribuições de probabilidade pelas equações (2.107) e (2.108):

$$P(\mathbf{w} | \alpha^B; \mathcal{M}^B, \mathcal{R}^B) = \frac{\exp(-\alpha^B \varepsilon_R)}{\left(\frac{\pi}{\alpha^B}\right)^{\frac{N_B}{2}}} \quad (2.107)$$

$$P(\mathbf{D} | \mathbf{w}; \beta^B; \mathcal{M}^B) = \frac{\exp(-\beta^B \varepsilon_E)}{\left(\frac{\pi}{\beta^B}\right)^{\frac{n_B}{2}}} \quad (2.108)$$

A expressão (2.107) representa a probabilidade de se obter os pesos \mathbf{w} dado o parâmetro de ponderação α^B , a arquitetura de rede \mathcal{M}^B e o funcional de regularização \mathcal{R}^B . Já a expressão (2.108) representa a probabilidade de se obter os dados \mathbf{D} tendo em vista o vetor de pesos \mathbf{w} , o parâmetro β^B e a arquitetura \mathcal{M}^B . A variável N_B representa o número total de parâmetros livres da rede, ou seja, a dimensão do vetor \mathbf{w} . Já a variável n_B diz respeito ao produto do número de neurônios na camada de saída pelo número de amostras no conjunto de dados, ou seja, caracteriza o número de graus de liberdade do problema para determinação do vetor \mathbf{w} (MACKAY, 1992b, p.454).

Tendo em vista a distribuição normal, verifica-se que os parâmetros α^B e β^B estão relacionados às variâncias das distribuições de probabilidade anteriores pelas equações (2.109) e (2.110):

$$\sigma_w^2 = \frac{1}{\alpha^B} \quad (2.109)$$

$$\sigma_v^2 = \frac{1}{\beta^B} \quad (2.110)$$

A abordagem Bayesiana para a etapa de modelagem implica em maximizar a probabilidade de se obter o vetor de parâmetros \mathbf{w} , dados os parâmetros α^B , β^B , o conjunto de dados \mathbf{D} , a arquitetura de rede \mathcal{M}^B e o funcional de regularização \mathcal{R}^B . A equação (2.111) ilustra essa abordagem:

$$P(\mathbf{w} | \mathbf{D}; \alpha^B; \beta^B; \mathcal{M}^B) = \frac{P(\mathbf{D} | \mathbf{w}; \beta^B; \mathcal{M}^B)P(\mathbf{w} | \alpha^B; \mathcal{M}^B; \mathcal{R}^B)}{P(\mathbf{D} | \alpha^B; \beta^B; \mathcal{M}^B; \mathcal{R}^B)} \quad (2.111)$$

É necessário, no entanto, determinar os valores ótimos de α^B e β^B , baseando-se nos dados coletados do problema. Este procedimento de determinação está resumidamente descrito no ANEXO C.

Demonstra-se matematicamente no ANEXO C que os parâmetros de maior probabilidade α^B e β^B são expressos por:

$$\gamma = N_B - 2\alpha^B(\text{Traço}(\mathbf{H}^{-1})) \quad (2.112)$$

$$\alpha^B = \frac{\gamma}{2\varepsilon_R} \quad (2.113)$$

$$\beta^B = \frac{n_B - \gamma}{2\varepsilon_E} \quad (2.114)$$

A matriz \mathbf{H} representa a hessiana da função objetivo ε_F . A variável γ representa o número efetivo de parâmetros utilizados pela rede para redução da função erro (FORESSE; HAGAN, 1997, p. 1931). Após o treinamento, o número de parâmetros efetivos γ deve ser verificado. Caso este esteja muito próximo do número de parâmetros da rede, provavelmente a mesma necessitará de um número maior de neurônios para representar o problema. A partir do momento em que a adição de neurônios não modificar o cálculo do número de parâmetros efetivos, presume-se que a rede alcançou seu tamanho ótimo.

O algoritmo Levenberg-Marquardt combina os métodos de Newton e do gradiente descendente em uma rotina de otimização, de modo a evitar dois problemas: a falha do método de Newton no caso em que a matriz Hessiana seja singular; a convergência lenta do método do gradiente descendente. Os pesos da rede são ajustados segundo a equação (2.115) (BRAGA; CARVALHO; LUDERMIR, 2007, p.83):

$$\Delta w_{ji}(t) = -[\nabla^2 E(w_{ji}(t)) + \mu_{LM} \mathbf{I}]^{-1} \nabla E(w_{ji}(t)) \quad (2.115)$$

A matriz Hessiana é expressa por $\nabla^2 E(w_{ji}(t))$; o vetor gradiente descendente é expresso por $-\nabla E(w_{ji}(t))$. Um valor elevado para μ_{LM} implica na utilização do método do gradiente descendente preferencialmente ao método de Newton, ambos implícitos no algoritmo Levenberg-Marquardt.

Foresse e Hagan (1997, p.1932) utilizaram a aproximação da matriz Hessiana disponível no algoritmo Levenberg-Marquardt para a determinação dos parâmetros ótimos α^B e β^B . Para detalhes do algoritmo, cf. Foresse e Hagan (1997, p.1932)

2.4.6 Considerações finais sobre as técnicas de modelagem

Encerra-se, portanto, a base teórica referente às técnicas de modelagem utilizadas no programa desenvolvido. Esta fundamentação é importante para a posterior discussão dos resultados obtidos, referentes à determinação do PFE da nafta de craqueamento.

2.5 Técnicas para seleção de variáveis

2.5.1 Análise de correlação

O coeficiente de correlação r (MAGALHÃES; LIMA, 2002, p.142) é definido matematicamente para os vetores coluna \mathbf{z} e \mathbf{w} pela equação (2.116):

$$r(\mathbf{z}, \mathbf{w}) = \frac{\sum_{i=1}^n z_i w_i - n(\bar{\mathbf{z}})(\bar{\mathbf{w}})}{\sqrt{\left(\sum_{j=1}^n z_j^2 - n(\bar{\mathbf{z}})^2\right)\left(\sum_{j=1}^n w_j^2 - n(\bar{\mathbf{w}})^2\right)}} \quad (2.116)$$

Este último é conhecido como coeficiente de correlação de Pearson (MATTOS apud NETO, 2004, p.38). Percebe-se que é possível construir um gráfico o qual expressa o valor da correlação cruzada entre o sinal da variável regressora x_j e o sinal da variável dependente y , de acordo com a equação (2.116), deslocando-se uma dessas variáveis por k instantes de amostragem a cada cálculo de correlação. Este gráfico é denominado função de correlação cruzada. Outra abordagem utilizada para a construção deste gráfico é dada pela equação (2.117), de uso frequente nos estudos de séries temporais de duração finita (AGUIRRE, 2007, p.186):

$$\hat{r}_{xy}(k) = \begin{cases} \frac{1}{N} \sum_{i=1}^{N-k} x(i)y(i+k), & k = 0, 1, \dots, N-1 \\ \frac{1}{N} \sum_{i=1}^{N+k} y(i)x(i-k), & k = 0, -1, \dots, -N+1 \end{cases} \quad (2.117)$$

Caso as variáveis x e y sejam independentes no tempo, sua função de correlação cruzada não será estatisticamente significativa para diversos deslocamentos k . Caso contrário, o gráfico de correlação cruzada apresentará regiões nas quais se verifica algum indício de correlação linear entre ambas. Aguirre (2007, p.197) define como intervalo de confiança para funções de correlação cruzada, ao nível de significância de 95%:

$$\frac{-1,96}{\sqrt{n}} < \hat{r}_{xy} < \frac{+1,96}{\sqrt{n}} \quad (2.118)$$

em que n representa o número total de elementos dos vetores coluna x e y .

Este procedimento pode ser utilizado para estimar atrasos de transporte em processos químicos. Neste caso, é possível observar um pico de correlação, seja ele positivo ou negativo, após k instantes de amostragem, que sugere uma ordem de grandeza para o tempo morto (AGUIRRE, 2007, p.190).

2.5.2 Regressão linear por etapas (*stepwise regression*)

A regressão linear por etapas é uma técnica bastante utilizada para a seleção de variáveis. Consiste na adição e remoção iterativa de regressores x_j , a partir de um teste estatístico parcial F , considerando-se valores limites para aceitação ou eliminação da variável (MONTGOMERY, 2003, p.259). O procedimento é iniciado com a inclusão do regressor de maior correlação x_1 com a variável resposta y , dado que o mesmo apresenta a maior estatística parcial F dentre os candidatos (F_{I-MAX}). Em seguida, as variáveis candidatas restantes são testadas no modelo e aquela que apresentar um valor máximo da estatística parcial F , tal que a mesma seja superior ao critério de aceitação, será adicionada ao modelo. Então:

$$F_{2-MAX} = \frac{SQ_R(\beta_2 | \beta_1, \beta_0)}{MQ_E(\mathbf{x}_2, \mathbf{x}_1)} > f_{in} \quad (2.119)$$

No próximo passo, o algoritmo verifica se o regressor x_1 adicionado anteriormente deve ser mantido no modelo. Realiza-se o seguinte teste parcial F :

$$F_1 = \frac{SQ_R(\beta_1 | \beta_2, \beta_0)}{MQ_E(\mathbf{x}_1, \mathbf{x}_2)} > f_{out} \quad (2.120)$$

Se a estatística F_1 for superior a f_{out} , o regressor é mantido no modelo. O algoritmo busca, então, a inserção de um novo regressor. Este procedimento é finalizado quando não for possível, dados os critérios de aceitação/eliminação, incluir ou retirar variáveis do modelo.

Um procedimento simplificado da regressão por etapas consiste na adição progressiva de variáveis ao modelo, considerando-se os mesmos critérios de aceitação expressos na equação (2.119). Sendo assim, variáveis são inseridas até que não haja mais candidatos a regressor que produzam aumento significativo na soma quadrática da regressão. No entanto, a etapa de remoção de variáveis é omitida. Desta forma, o efeito que a adição de uma nova variável exerce nos regressores já incluídos permanece indeterminada. Ainda assim, esta técnica possibilita uma análise rápida e abrangente das variáveis regressoras mais prováveis de serem incluídas nos modelos.

Segundo Montgomery (2003, p.212) a regressão linear deve ser utilizada com cuidado no que diz respeito à seleção de variáveis regressoras, pois é possível desenvolver relações estatísticas entre variáveis que apresentem correlação, porém não apresentem causalidade. Além disso, ressalta-se que as relações de regressão são válidas apenas para valores dos regressores dentro da faixa dos dados originais, devendo-se evitar extrapolações do modelo.

2.5.3 Considerações finais sobre as técnicas de seleção de variáveis

Encerram-se os aspectos teóricos relevantes referentes às técnicas de seleção de variáveis, as quais estão presentes na ferramenta computacional desenvolvida. A seguir, demonstra-se a aplicação prática das técnicas tratadas na presente revisão bibliográfica.

3 METODOLOGIA

3.1 Descrição do *INFERATOR* – Rotinas para ajuste de inferências.

Com o objetivo de disponibilizar uma ferramenta que acelere a construção de analisadores virtuais, baseados em modelos tipo ‘caixa-preta’, buscou-se agrupar um conjunto de técnicas matemáticas e estatísticas aplicadas a este propósito em um programa computacional. A aplicação desenvolvida, denominada *INFERATOR*, foi testada em um problema prático de forma a possibilitar sua validação e constante aperfeiçoamento.

O programa possibilita ao usuário comparar o desempenho das técnicas disponíveis, tendo em vista a capacidade de generalização e robustez de cada uma delas. Optou-se pelo desenvolvimento deste programa no aplicativo MATLAB[®] R2008b em função das facilidades já inclusas no software, tais como: toolbox de redes neurais, toolbox de funções matemáticas e estatísticas para ajuste de modelos, disponibilidade de algoritmos para decomposição em valores singulares, mínimos quadrados parciais, regressão linear, dentre outras.

O programa foi concebido com uma estrutura modular, de tal forma que o usuário possa habilitar exclusiva ou simultaneamente as ferramentas de pré-tratamento dos dados, seleção de variáveis e modelagem matemática de sistemas. Encontram-se disponíveis para o usuário cinco técnicas tradicionais para ajuste de inferências, a saber: regressão linear multivariada (MLR), regressão em componentes principais (PCR), mínimos quadrados parciais (PLS), redes neurais tipo *Perceptron* de múltiplas camadas e redes neurais de base radial (RBF). As redes tipo *Perceptron* de múltiplas camadas são subdivididas nas categorias: redes com uma (RNA1 BP) ou duas (RNA2 BP) camadas ocultas e treinamento baseado no algoritmo *back-propagation* resiliente; redes com uma camada oculta, com a função de custo estabelecida por regularização Bayesiana e treinamento realizado pelo algoritmo Levenberg-Marquardt.

Buscou-se, neste programa, a construção de uma interface gráfica mínima (Figura 3.1) que facilite a utilização das principais ferramentas:

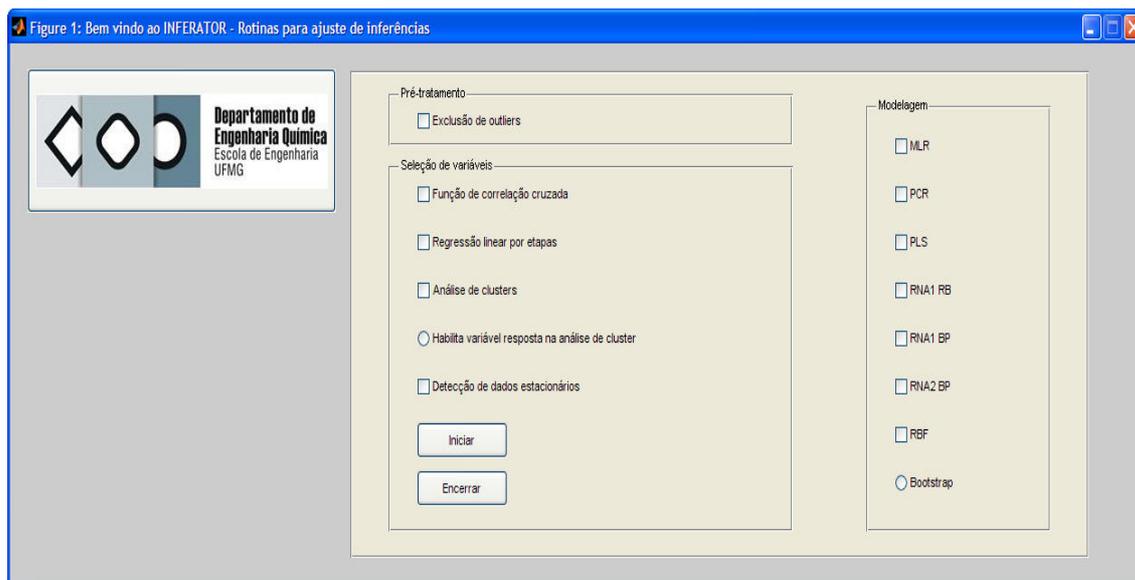


Figura 3.1: Interface gráfica para utilização do INFERATOR.

A Figura 3.2 descreve, em linhas gerais, as etapas de coleta e tratamento dos dados, bem como os cálculos executados pelo programa, resumindo a metodologia adotada ao longo deste trabalho. Cada etapa será detalhada nas seções seguintes.

3.2 Coleta e pré-tratamento dos dados

3.2.1 Aquisição dos dados do PI

Inicialmente, foram obtidos dados históricos referentes à operação de uma unidade de craqueamento do sistema Petrobrás, entre os dias 28/11/2008 e 28/08/2010, com o objetivo de desenvolvimento de um modelo estático para o PFE da nafta ‘leve’. Os dados foram coletados a cada 5 minutos, sendo que esta escolha foi orientada pela imprecisão intrínseca das rotinas de exceção e compressão do PI as quais limitam a acuidade da informação em taxas de amostragem elevadas. Foi possível verificar, através de uma análise gráfica preliminar dos dados coletados, que essa taxa de amostragem seria suficiente para a captura das médias das variáveis do sistema em regime estacionário ao longo do horizonte de tempo estudado. Ressalta-se que, como o modelo desenvolvido é estático, as dinâmicas mais rápidas não são de grande interesse na análise, sendo mais importante a correta observação das médias das variáveis de processo.

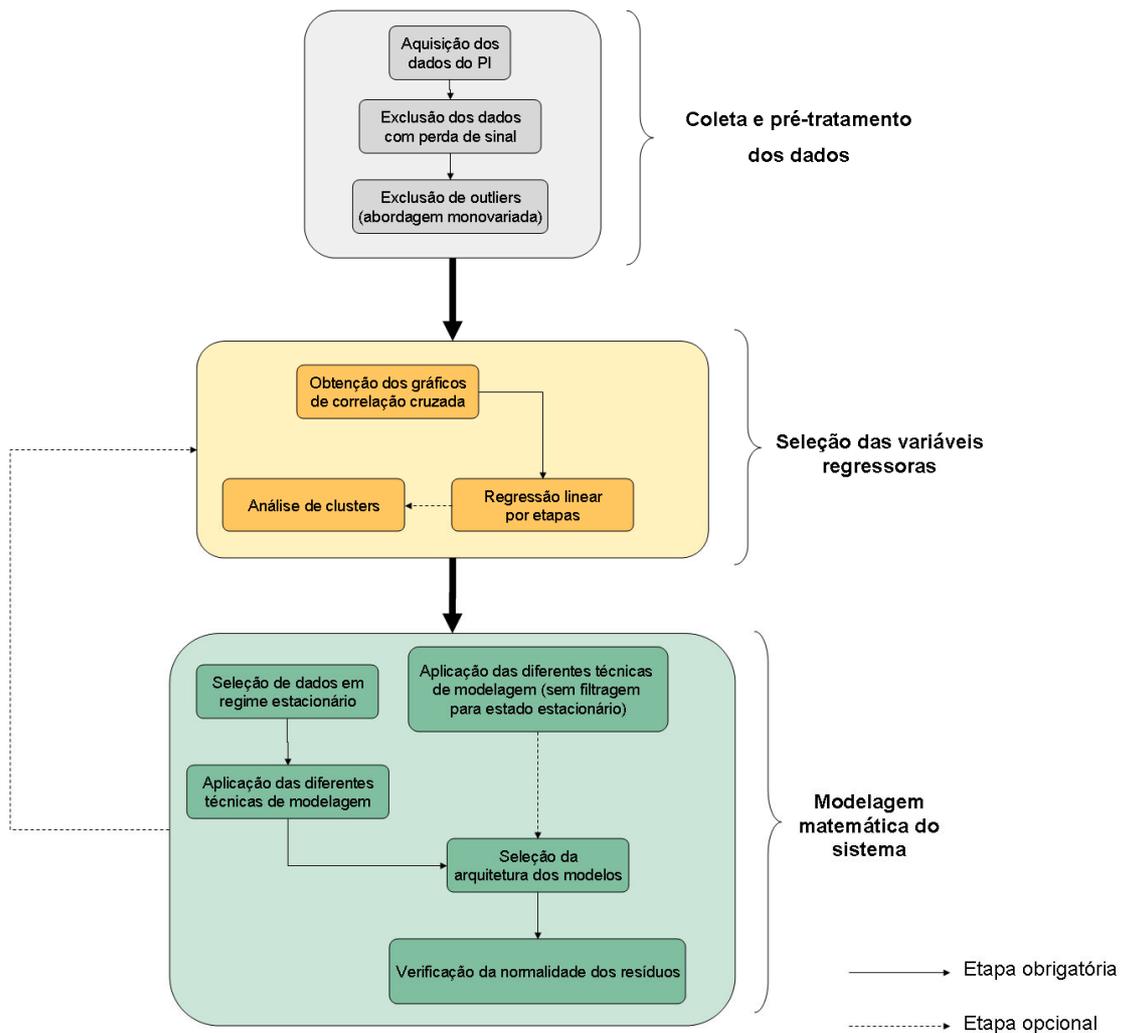


Figura 3.2: Fluxograma esquematizando as etapas de cálculo do programa INFERATOR.

A Tabela 3.1 exhibe as trinta e quatro variáveis escolhidas a partir de uma análise prévia. Estas possíveis variáveis regressoras foram selecionadas a partir de relações de causa ou efeito historicamente conhecidas, baseadas no conhecimento de processo e nos fenômenos físicos de provável ocorrência no sistema. As variáveis são classificadas, na Tabela 3.1, como ‘causa’, ‘correlacionada’ ou ‘consequência’. As variáveis classificadas como ‘causa’ são aquelas cujos estados determinam o valor obtido para a variável inferida. Já aquelas classificadas como ‘correlacionadas’ não determinam o valor da propriedade inferida; no entanto, são influenciadas por uma ‘causa comum’ à variável inferida. Conhecendo-se os efeitos desta variável ‘causa comum’ em uma variável ‘correlacionada’ é possível determinar seus efeitos na variável inferida. Outras variáveis podem, ainda, ser ‘consequência’ direta dos estados da variável a ser inferida.

Tabela 3.1: Relação das variáveis coletadas inicialmente.

Variável	Descrição	Suposta relação com a variável inferida	Fenômeno físico relacionado
1	Tempo (min)	Não se aplica	Não se aplica
2	Pressão de topo da fracionadora principal	Causa	Influência no equilíbrio termodinâmico da região de topo
3	Temperatura de topo da fracionadora principal	Causa	Influência no equilíbrio termodinâmico da região de topo
4	Temperatura do refluxo de topo para torre	Correlacionada	Influência no equilíbrio termodinâmico da região de topo
5	Vazão do refluxo de topo para a fracionadora	Causa	Influência no refluxo interno e no fracionamento
6	Vazão de retirada de nafta 'leve' craqueada	Consequência	Incorporação de compostos 'pesados' à corrente
7	Pressão do vapor de retificação no vaso separador	Causa	Pressão parcial dos hidrocarbonetos na fracionadora
8	Vazão A de vapor para retificação no vaso separador	Causa	Pressão parcial dos hidrocarbonetos na fracionadora
9	Vazão B de vapor para retificação no vaso separador	Causa	Pressão parcial dos hidrocarbonetos na fracionadora
10	Vazão C de vapor para retificação no vaso separador	Causa	Pressão parcial dos hidrocarbonetos na fracionadora
11	Vazão D de vapor para retificação no vaso separador	Causa	Pressão parcial dos hidrocarbonetos na fracionadora
12	Vazão A de vapor para o riser	Causa	Pressão parcial dos hidrocarbonetos na fracionadora
13	Vazão B de vapor para o riser	Causa	Pressão parcial dos hidrocarbonetos na fracionadora
14	Vazão C de vapor para o riser	Causa	Pressão parcial dos hidrocarbonetos na fracionadora
15	Vazão D de vapor para o riser	Causa	Pressão parcial dos hidrocarbonetos na fracionadora
16	Vazão E de vapor para o riser	Causa	Pressão parcial dos hidrocarbonetos na fracionadora
17	Vazão A de gásóleo para riser	Causa	Rendimentos dos produtos e hidráulica da torre
18	Vazão B de gásóleo para riser	Causa	Rendimentos dos produtos e hidráulica da torre
19	Vazão de nafta 'leve' de coque	Causa	Influência nos rendimentos dos produtos
20	Vazão de gásóleo de coque	Causa	Influência nos rendimentos dos produtos

Tabela 3.1 (continuação): Relação das variáveis coletadas inicialmente.

Variável	Descrição	Suposta relação com a variável inferida	Fenômeno físico relacionado
21	Temperatura no topo do riser	Causa	Influência nos rendimentos dos produtos
22	Temperatura na base do riser	Causa	Influência nos rendimentos dos produtos
23	Diferencial de pressão entre o topo e a panela de LCO	Correlacionada	Condição hidráulica da torre
24	Temperatura da panela de retirada de nafta 'pesada'	Causa	Influência no equilíbrio termodinâmico da região de topo
25	Vazão de refluxo circulante de nafta 'pesada'	Causa	Influência no perfil térmico da torre
26	Vazão de retirada de nafta 'pesada' para tratamento	Causa	Condição hidráulica da torre
27	Vazão de LCO na saída da fracionadora principal	Correlacionada	Influência no refluxo interno da torre
28	Retirada de LCO para tanque	Correlacionada	Influência no refluxo interno da torre
29	Vazão de LCO para integração energética	Correlacionada	Influência no perfil térmico da torre
30	Temperatura do refluxo de HCO para integração energética	Correlacionada	Influência no perfil térmico da torre
31	Vazão de HCO para integração energética	Correlacionada	Influência no perfil térmico da torre
32	Diferencial de pressão entre a panela de LCO e a panela de HCO	Correlacionada	Condição hidráulica da torre
33	Diferencial de pressão entre a panela de HCO e o fundo	Correlacionada	Condição hidráulica da torre
34	Diferencial de pressão entre o topo e o fundo	Correlacionada	Condição hidráulica da torre
35	PFE da nafta craqueada	Não se aplica	Não se aplica

A Figura 3.3 e a Figura 3.4 exibem um diagrama esquemático da localização da instrumentação analisada no presente trabalho. Adotou-se a seguinte notação:

- P – pressão;
- T – temperatura;
- V – vazão;
- DP – diferencial de pressão;
- A – análise de laboratório;

O número posterior à identificação do tipo de medição se relaciona ao índice da variável expressa na Tabela 3.1 (ex: DP – 23; diferencial de pressão, variável 23).

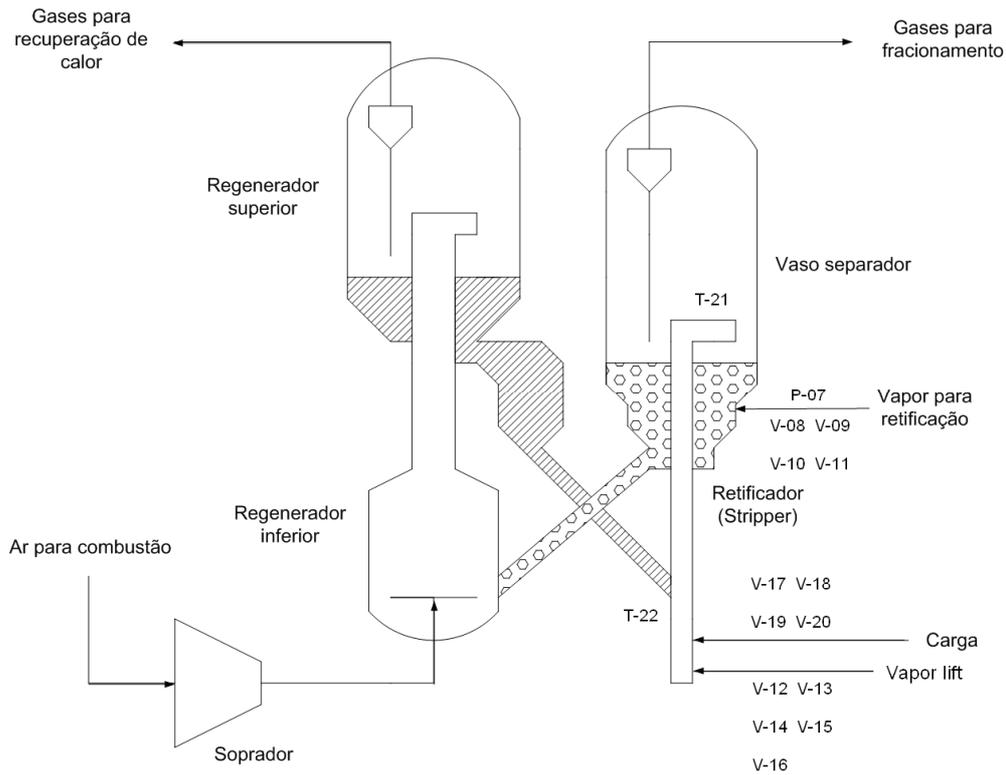


Figura 3.3: Diagrama esquemático da localização da instrumentação do conversor.

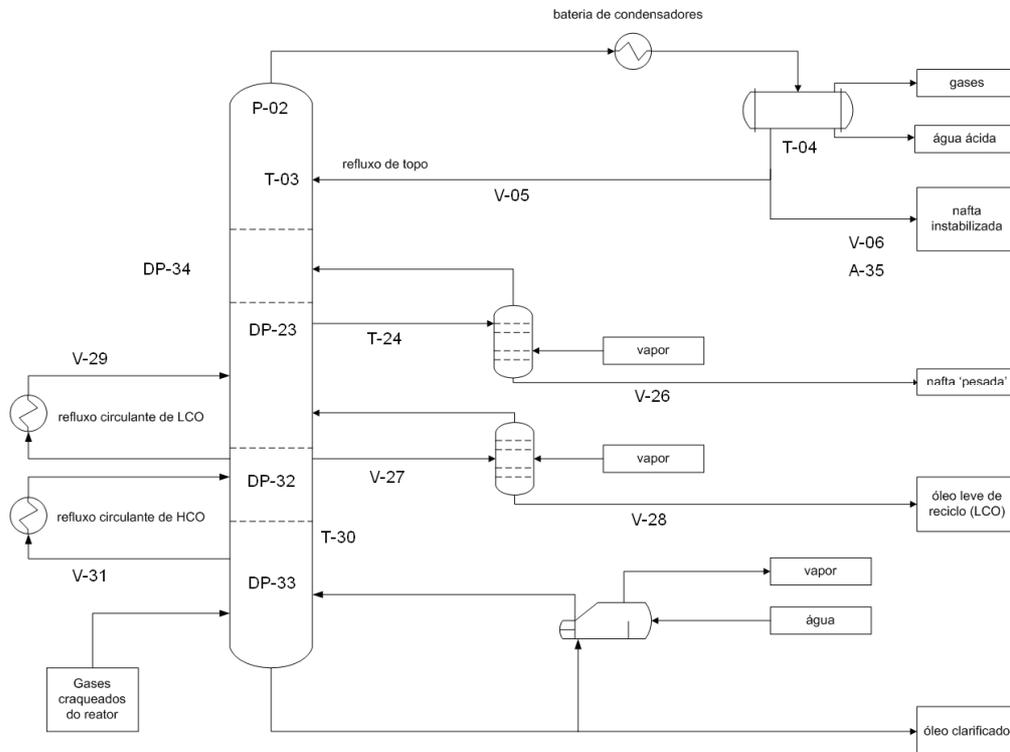


Figura 3.4: Diagrama esquemático da localização da instrumentação da fracionadora.

3.2.2 Exclusão dos dados com perda de sinal

Inicialmente, foram eliminadas as amostras que apresentavam perda de sinal, tais como indicações fora do *range* de medição, falhas de comunicação com o servidor PI-UDS, dentre outras causas possíveis. Em situações operacionais atípicas, tais como paradas repentinas de unidades, alguns medidores apresentam indicações fora do *range* de medição, ocasionando perda temporária de sinal. Sendo assim, estes eventos são eliminados do conjunto de dados por essa filtragem prévia.

Para esta etapa de filtragem foi desenvolvida uma rotina em Visual Basic que identifica se uma determinada linha do arquivo de dados em Excel apresenta pelo menos um elemento tipo texto. Em caso afirmativo, todos os elementos da linha são excluídos e o conjunto de dados é automaticamente concatenado. Com intuito de verificar se alguma variável em especial estava operando sistematicamente fora do range de medição, o que acarretaria a exclusão sistemática de várias informações de interesse, buscou-se registrar na planilha de filtragem o número de dados antes e após a exclusão.

Destaca-se que todos os cálculos de média conduzidos pelo programa e descritos nas seções posteriores consideram a possibilidade de existência de descontinuidades temporais no conjunto de dados. Estas descontinuidades são consequência da exclusão de dados, quer seja em função da perda de sinal quer seja pela presença de *outliers*. Para determinação da média, o programa sempre considera o instante de tempo de aquisição do dado de processo, normalmente obtido por interpolação pelo método ‘*swinging door compression*’, descrito na seção 2.4.1.1.2. O programa utiliza uma janela temporal, estabelecida pelo usuário, para calcular a média. Desta forma, o número de dados presentes no cálculo da média pode ser variável, dependendo do número de dados previamente excluídos dentro da janela temporal de interesse. Este procedimento será exemplificado nas seções 3.3.1 e 3.3.2.

3.2.3 Exclusão de *outliers*

Após a exclusão dos dados com perda de sinal, os dados restantes são normalizados e, em seguida, é aplicado o módulo de “Exclusão de outliers” (Figura 3.1). Este módulo permite ao usuário escolher as variáveis mais adequadas para aplicação do identificador de Hampel. Para o cálculo da mediana, *MAD* e dos limites de corte caracterizados pelas

equações (2.8) e (2.9), foi utilizada uma janela temporal contemplando todo o conjunto de dados. Através de uma rotina auxiliar, o programa exibe gráficos característicos dos limites de corte para todas as variáveis, ficando a cargo do usuário a escolha daquelas sobre as quais será aplicado o identificador de Hampel. Esta decisão deve ser baseada no conhecimento prévio de processo. Como exemplo, cita-se a temperatura de reação no topo do riser. Em função de seu histórico, verifica-se que estando esta temperatura abaixo de um determinado valor, é possível presumir que a unidade de craqueamento se encontrava parada ou em procedimento de partida. Outra hipótese possível é a ocorrência de falha de medição. Em ambas situações o dado deve ser expurgado.

A Figura 3.5 exemplifica uma variável caracterizada como adequada para detecção de *outliers*, pois nela não se verificam descontinuidades nos patamares de operação. Os valores que fogem da normalidade são evidentes, e bem capturados pelos limites de corte. Nela se verificam condições atípicas de processo tanto acima quanto abaixo das linhas de corte tracejadas em vermelho.

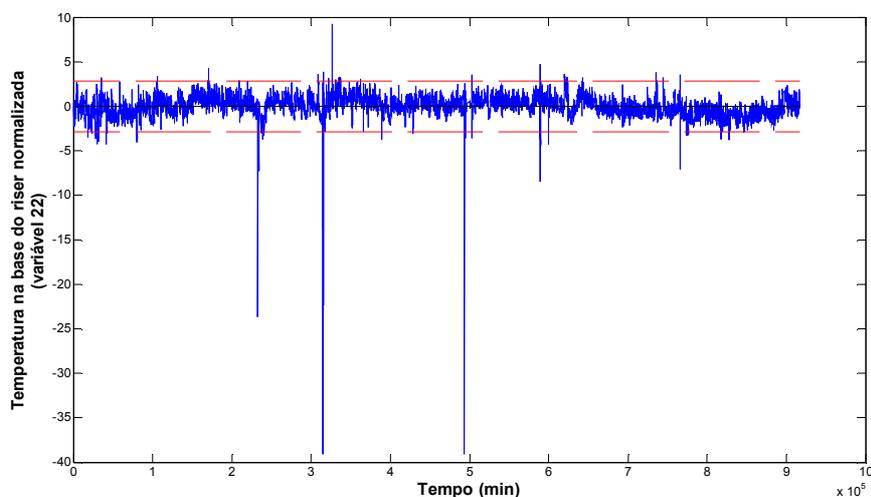


Figura 3.5: Variável considerada adequada para utilização do identificador de Hampel.

Já a Figura 3.6 exibe uma variável em que, dada a diversidade de patamares de operação da planta ao longo do tempo de campanha, a utilização do identificador Hampel, considerando-se toda a janela temporal dos dados adquiridos, resultaria em um grande número de prováveis *outliers*. Portanto, esta variável é caracterizada como inadequada para aplicação do filtro. Já com a utilização de uma janela temporal móvel para o

identificador de Hampel seria possível escolher esta variável para o procedimento de exclusão de *outliers*.

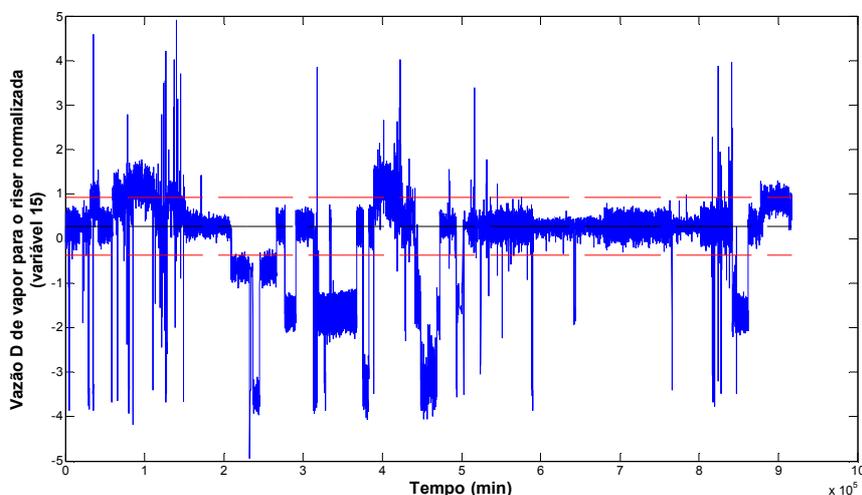


Figura 3.6: Variável inadequada para estabelecimento de *outliers* pelo identificador de Hampel.

Baseando-se nesta discussão prévia, foram escolhidas as seguintes variáveis para identificação dos *outliers*, sendo que seus respectivos gráficos serão exibidos posteriormente na seção referente aos resultados:

Tabela 3.2: Variáveis selecionadas para aplicação do identificador de Hampel.

Variável	Descrição
2	Pressão de topo da fracionadora principal
22	Temperatura na base do riser de reação
24	Temperatura da panela de retirada de nafta 'pesada'
30	Temperatura do refluxo circulante de HCO para integração energética
32	Diferencial de pressão entre a panela de LCO e a panela de HCO

De maneira similar à filtragem para exclusão de dados com perda de sinal, a identificação de um único *outlier* em uma determinada amostra implica na exclusão de todos os dados daquela amostra. Após o término do procedimento de exclusão, a porcentagem de dados restantes é registrada pelo programa, de modo a fornecer uma estimativa da quantidade de informação perdida.

3.3 Seleção das variáveis regressoras

3.3.1 Obtenção dos gráficos de correlação cruzada

A técnica de análise de correlação é empregada neste ponto para estabelecer quais variáveis, dentre as variáveis candidatas, apresentam alguma correlação linear estatisticamente significativa com a variável a ser inferida. Este módulo é denominado “Função de correlação cruzada” (Figura 3.1). Como o valor do PFE só é obtido em instantes discretos correspondentes aos momentos da amostragem para análise laboratorial, a seguinte metodologia foi adotada para estimar a existência de correlação linear:

1. Os valores de PFE referentes aos instantes t_a de coleta das amostras de nafta para análise foram obtidos ao longo de todo período analisado (28/11/2008 até 28/08/2010), constituindo o vetor \mathbf{y} ;
2. Médias sucessivas referentes à variável candidata X_i foram obtidas, considerando-se uma janela móvel de tempo de δ_j minutos, a qual se inicia a cada instante t_a de amostragem de campo. Sendo assim, capturam-se as medições históricas da variável X_i entre os instantes t_a e $t_a - \delta_j$, sendo efetuada uma média destes valores. Esta sequência de médias irá constituir o vetor $(\mathbf{x}_i)_0$;
3. Calcula-se a correlação $r(\mathbf{y}; (\mathbf{x}_i)_0)$ pela equação (2.116), sendo este valor associado ao instante t_a ;
4. Médias sucessivas referentes à variável candidata X_i são novamente calculadas, considerando-se uma janela móvel de tempo de δ_j minutos, que se inicia a cada instante $t_a - \delta_s$, sendo δ_s o passo temporal sucessivo de recuo no tempo. Assim, capturam-se os dados entre os instantes $t_a - \delta_s$ e $t_a - \delta_j - \delta_s$, sendo efetuada uma média destes valores. Esta sequência de médias irá constituir o vetor $(\mathbf{x}_i)_1$;
5. Calcula-se a correlação $r(\mathbf{y}; (\mathbf{x}_i)_1)$ pela equação (2.116), sendo associado este valor ao instante $t_a - \delta_s$;

6. Retorna-se ao passo 4, considerando-se deslocamentos temporais sucessivos dados por $n\delta_s$, onde n é um número inteiro.

A Figura 3.7 exibe um gráfico gerado com os valores de correlação obtidos a partir da metodologia descrita. É possível perceber que os dados da variável candidata apresentam um pico de correlação cerca de 180 minutos antes da amostragem, o que sugere, ainda, a existência de um atraso de transporte dessa magnitude no sistema.

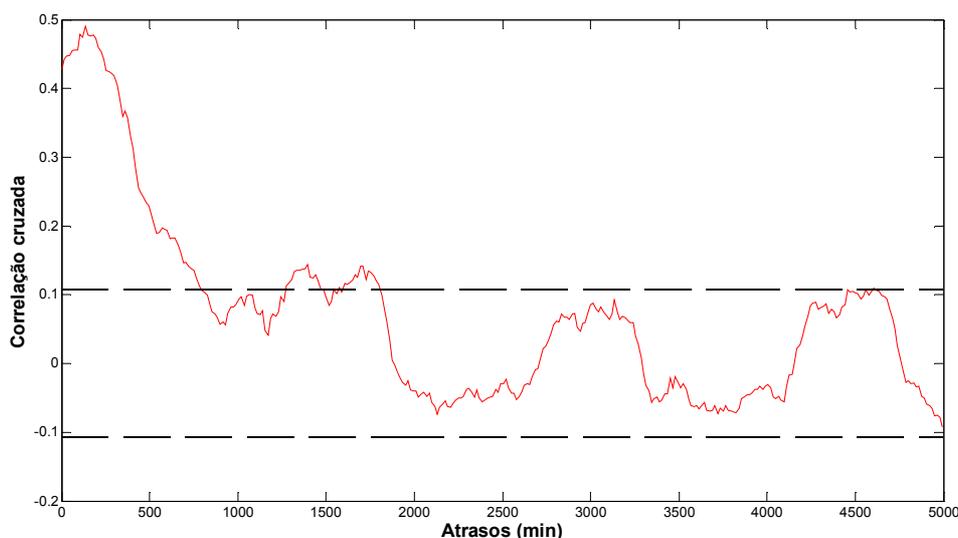


Figura 3.7: Gráfico de correlação cruzada do diferencial de pressão entre a panela de LCO e o topo da fracionadora (variável 23) em relação ao PFE (variável 35).

A metodologia adotada consistiu na análise dos gráficos de correlação, selecionando-se para as próximas etapas as variáveis que apresentaram alguma correlação estatística linear significativa. Estas foram capturadas próximas aos seus respectivos picos de correlação, quando perceptíveis e desde que coerentes com a estimativa do atraso de transporte do sistema, para posterior geração do conjunto de dados utilizados nas etapas de modelagem e validação. A janela temporal δ_j adotada para o cálculo das médias foi de 10 minutos, ou seja, inclusão de três pontos adjacentes do conjunto de dados para o cálculo. A janela temporal δ_s adotada foi de 15 minutos a cada iteração.

Ressalta-se que as variáveis a serem escolhidas nesta etapa devem apresentar indícios de correlações lineares estatísticas compatíveis com a dinâmica do processo, ou seja, que estejam dentro de um intervalo de tempo possível para o atraso de transporte, pois cabe

lembrar que o PFE é determinado por condições no topo da torre fracionadora, porém o mesmo só pode ser observado na saída da seção de tratamento. Sendo assim, uma estimativa grosseira para o tempo de residência do inventário entre o topo da torre fracionadora e a saída da seção de tratamentos deve ser obtida, para conferir ordem de grandeza para o atraso de transporte. Este procedimento foi realizado através do levantamento dos volumes dos principais vasos da unidade. O inventário contido nas bandejas e recheios no interior de torres foi desprezado, considerando-se somente o volume do fundo destes equipamentos. O inventário das linhas também foi desconsiderado. A estimativa do tempo de residência é dada por:

$$\tau = \frac{V}{\dot{v}} \quad (3.1)$$

em que τ é o tempo de residência do inventário, V é o volume total de inventário na unidade e \dot{v} é a produção média de nafta de craqueamento.

O programa INFERATOR identifica automaticamente o instante de amostragem de campo no conjunto de dados ao constatar uma mudança no valor do PFE. A existência de descontinuidade no conjunto de dados, resultantes dos processos prévios de filtragem, implica que alguns cálculos de médias serão efetuados com um número inferior de pontos disponíveis na janela temporal considerada.

Todas as capturas de médias em janelas temporais móveis são realizadas baseadas no valor de tempo associado às variáveis. Esta diretriz é válida não apenas para a função de correlação cruzada, mas qualquer rotina que utilize janelas móveis para a captura de médias, tais como a montagem da matriz de dados, a ser mencionada no item 3.3.2. Desta forma, duas linhas adjacentes na matriz de dados, referentes a instantes de tempo os quais não permitem que ambas sejam enquadradas na mesma janela temporal, não serão indevidamente utilizadas no cálculo das médias. Apenas as amostras cujos instantes de tempo se enquadram na janela temporal determinada pelo usuário é que serão utilizadas nos cálculos de médias.

3.3.2 Regressão linear por etapas

Após a análise dos gráficos de correlação cruzada, as variáveis candidatas selecionadas foram capturadas em instantes de tempo próximos àqueles referentes a seus respectivos

picos de correlação, desde que coerentes com a estimativa de atraso de transporte. A metodologia adotada incluiu as seguintes etapas:

1. Criação de um arquivo de configuração em Excel, contendo três informações relevantes: índices das colunas das variáveis selecionadas na etapa de análise de correlação; deslocamentos de tempo δ_p (min) referentes aos picos de correlação de cada uma dessas variáveis; intervalo temporal δ_D (min) desejado para a captura das médias de processo de cada variável. Este arquivo é um argumento de entrada para a rotina de regressão linear por etapas;
2. O programa identifica automaticamente os instantes de amostragem de campo e, então, utilizando o arquivo de configuração, recua no tempo δ_p minutos para capturar as médias de processo próximo ao pico de correlação identificado, considerando uma janela temporal captura de δ_D minutos. Sendo assim, capturam-se as medições históricas da variável X_i entre os instantes $t_a - \delta_p$ e $t_a - \delta_p - \delta_D$, sendo efetuada uma média destes valores. Esta sequência de médias irá constituir o vetor coluna \mathbf{x}_i ;
3. Repete-se o passo 2 para todas as variáveis selecionadas no arquivo de configuração;

Desta forma, obtêm-se a matriz dos dados de entrada \mathbf{X} a ser utilizada posteriormente nas etapas de modelagem e validação.

Padronizou-se a janela temporal δ_D em 30 minutos para todas as variáveis. Este período foi escolhido em função dos relatos dos operadores quanto à incerteza média do horário de amostragem. Ressalta-se ainda a grande diminuição do número de dados disponíveis para o ajuste dos modelos. Como a frequência de análise laboratorial é, aproximadamente, diária, o conjunto de dados para a modelagem é bastante reduzido em relação aos dados originalmente coletados.

Os dados são então normalizados e divididos em dois grupos: dados destinados à modelagem e dados destinados à validação. Aplica-se aos dados destinados à modelagem a regressão linear por etapas, utilizando-se a função *stepwisefit* do MATLAB[®]. O usuário fornece ao programa INFERATOR o número de vezes que

deseja repetir este procedimento, bem como a proporção dos dados destinada à modelagem. Como a separação do conjunto de dados é feita de maneira aleatória, é possível verificar quais variáveis foram selecionadas para os diferentes dados de modelagem sorteados em cada repetição do algoritmo. Considerou-se, neste trabalho, um total de 100 repetições por nível de significância analisado. Foram analisados os níveis de significância de 5% e 1% para o procedimento de regressão linear por etapas.

Vários níveis de significância para seleção de variáveis no modelo linear podem ser avaliados a critério do usuário. O programa INFERATOR guarda todo o histórico de seleção das variáveis por este procedimento, assim como a prioridade de seleção, porque, dentre as variáveis escolhidas, aquelas que apresentam as maiores correlações lineares com a resposta são selecionadas primeiro pelo algoritmo de regressão linear por etapas.

A avaliação de relações não lineares é possível desde que as variáveis de interesse sejam linearizadas previamente através da aplicação de uma função, sendo então a variável transformada fornecida à matriz de dados.

O procedimento é repetido até que todas as variáveis informadas no arquivo de configuração sejam selecionadas como regressoras no modelo. Isso quer dizer que apenas as variáveis que foram selecionadas em todos os procedimentos de regressão linear realizados serão utilizadas no equacionamento. Deste modo, a premissa de obtenção de um modelo com o menor número de variáveis regressoras possível pode ser alcançada.

3.3.3 Análise de *clusters*

Utilizando-se o algoritmo das k-médias presente no MATLAB[®], buscou-se verificar, dentre as variáveis escolhidas pelo algoritmo de regressão linear por etapas, se um número ainda menor destas possibilitaria uma partição natural do conjunto de dados segundo níveis pré-estabelecidos para o PFE, a saber: *clusters* com PFE alto, *clusters* com PFE médio e *clusters* com PFE baixo. O programa INFERATOR adiciona gradativamente as variáveis selecionadas na etapa de regressão linear à matriz de dados **D** utilizada no algoritmo das k-médias, seguindo nesta adição a mesma ordem verificada no procedimento de regressão linear por etapas. O MATLAB[®] utiliza a

distância Euclidiana para construção dos *clusters*, os quais são inicializados a partir de amostras aleatórias sorteadas do conjunto de dados \mathbf{D} .

A cada nova variável adicionada, o programa exibe graficamente a distribuição dos *clusters*. Neste gráfico a abscissa representa o tempo em que em que foi coletada a variável resposta e a ordenada representa o valor normalizado da variável resposta. Cada *cluster* é indicado por um símbolo específico diferente. A Figura 3.8 exibe o resultado da classificação do método, considerando-se dois *clusters*. Neste exemplo ilustrativo, as três primeiras variáveis selecionadas pela regressão linear por etapas compõem o vetor das k-médias. Nesta classificação não fica clara a segregação do PFE em função dos níveis alto e baixo, embora uma segregação temporal dos *clusters* seja evidente.

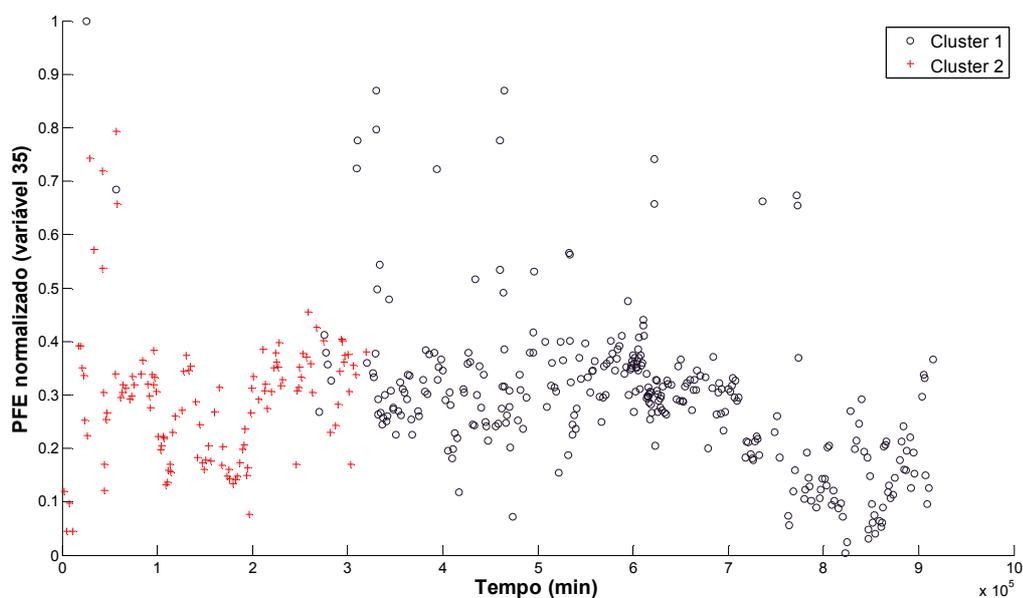


Figura 3.8: Representação gráfica da classificação obtida pelo método das k-médias.

Para ilustrar como funciona o módulo de análise de *clusters*, supõe-se que as variáveis 23, 3 e 21 (ver seção 3.2.1) sejam selecionadas, nesta ordem de prioridade, pelo algoritmo de regressão linear por etapas. O módulo de análise de *clusters* constrói, a cada ciclo de execução, uma matriz de dados \mathbf{D} , tal que:

$$\mathbf{D}_1 = [\mathbf{x}_{23} \quad \mathbf{y}] = \begin{bmatrix} x_{1-23} & y_1 \\ x_{2-23} & y_2 \\ x_{3-23} & y_3 \\ \vdots & \vdots \\ x_{n-23} & y_n \end{bmatrix} \quad (3.2)$$

$$\mathbf{D}_2 = [\mathbf{x}_3 \quad \mathbf{x}_{23} \quad \mathbf{y}] = \begin{bmatrix} x_{1-3} & x_{1-23} & y_1 \\ x_{2-3} & x_{2-23} & y_2 \\ x_{3-3} & x_{3-23} & y_3 \\ \vdots & \vdots & \vdots \\ x_{n-3} & x_{n-23} & y_n \end{bmatrix} \quad (3.3)$$

$$\mathbf{D}_3 = [\mathbf{x}_3 \quad \mathbf{x}_{21} \quad \mathbf{x}_{23} \quad \mathbf{y}] = \begin{bmatrix} x_{1-3} & x_{1-21} & x_{1-23} & y_1 \\ x_{2-3} & x_{2-21} & x_{2-23} & y_2 \\ x_{3-3} & x_{3-21} & x_{3-23} & y_3 \\ \vdots & \vdots & \vdots & \vdots \\ x_{n-3} & x_{n-21} & x_{n-23} & y_n \end{bmatrix} \quad (3.4)$$

As amostras da matriz de dados \mathbf{D} (i.e linhas da matriz \mathbf{D}) são segregadas de acordo com o número de *clusters* estabelecidos pelo usuário. A Figura 3.8 ilustra uma análise em que foram escolhidos, a priori, dois *clusters*. A rotina computacional retorna para o usuário o valor dos centroides obtidos em cada *cluster*. Ressalta-se que o programa INFERATOR permite que o usuário estabeleça quantos *clusters* desejar, o que torna a rotina computacional genérica, adaptada para conduzir outras análises além do presente estudo.

Os centroides destes clusters foram analisados para estabelecer como os níveis das variáveis regressoras estão correlacionados com os níveis observados para o PFE. Dado que a técnica de classificação das k-médias é não hierárquica, este estudo foi realizado segundo duas suposições distintas e exclusivas entre si:

- Existência de dois *clusters* apenas, classificando-se os dados em níveis de PFE alto e PFE baixo;
- Existência de três *clusters*, classificando-se os dados em níveis de PFE alto, PFE médio e PFE baixo.

Após a análise de *clusters*, as variáveis que indicaram a possibilidade de fornecer uma partição natural para os níveis do PFE foram selecionadas para análise complementar na etapa de modelagem.

3.4 Modelagem matemática de sistemas

3.4.1 Seleção de dados em regime estacionário

O método da carta de controle para a variável amplitude foi escolhido como modo de detecção dos dados em estado estacionário em função da sua maior facilidade para observação dos limites *LIC* e *LSC* de cada variável, os quais constituem o limiar de seleção. O método do desvio padrão móvel apresenta um limiar de seleção σ_k variável a cada janela de inspeção k , conforme equação (2.20), o que dificulta a sua avaliação.

Para a realização da seleção dos dados estacionários, são observadas três etapas. A primeira etapa utiliza o conjunto de dados referentes à saída dos métodos de exclusão de *outliers* para a determinação dos limites *LIC* e *LSC*, conforme as equações (2.22) e (2.23). Nesta etapa, verificam-se os seguintes eventos:

- O usuário determina um intervalo de tempo τ_{SS} referente aos dias subsequentes de operação da planta que serão analisados. Com o intuito de não tornar tendenciosa a escolha desta janela, apenas o número de dias é recebido pelo programa INFERATOR, que se encarrega de escolher aleatoriamente o dia inicial do intervalo para a análise. Neste estudo, foram analisados 240 dias subsequentes;
- Uma janela temporal em minutos δ_{SS} é estabelecida pelo usuário, sendo que neste trabalho foi considerado o tempo de 30 minutos, o mesmo escolhido para a captura das médias de processo. Esta janela se moverá ao longo do intervalo de dias selecionados, procedendo, a cada iteração, o cálculo da amplitude e do desvio padrão do intervalo, para cada uma das variáveis listadas no arquivo de configuração. São gerados, então, os histogramas para a variável W , definida como a razão entre a amplitude R_i da i -ésima janela analisada e o desvio padrão populacional σ . O desvio padrão populacional de cada variável analisada é aproximado pela média dos desvios padrões calculados nas i janelas

temporais móveis, ao longo de todo o intervalo de tempo τ_{SS} . Um histograma obtido por este procedimento é exemplificado na Figura 3.9:

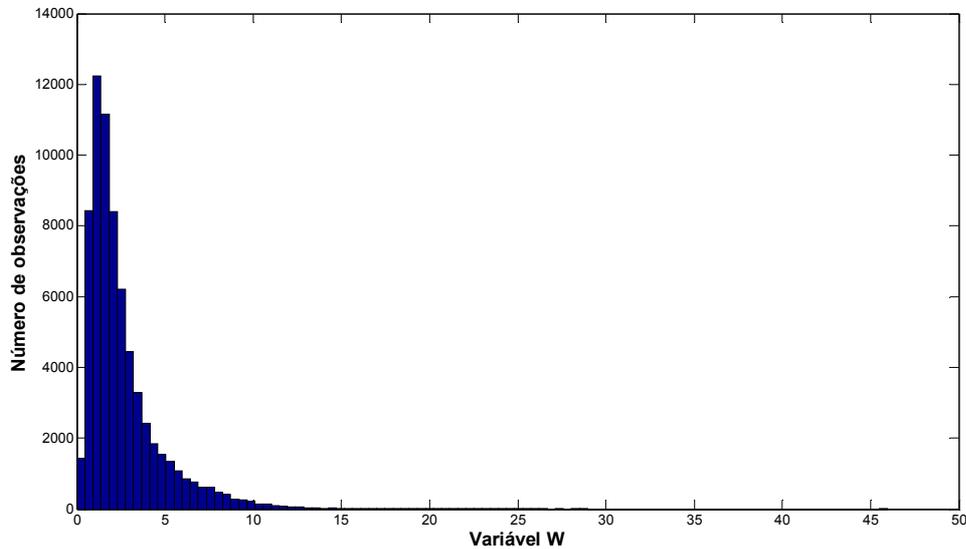


Figura 3.9: Histograma de distribuição ilustrativo característico da curva W para a variável 23.

- O histograma da variável W é aproximado pela distribuição de probabilidade Gamma, sendo que esta consideração foi realizada com base na análise visual dos gráficos obtidos. A distribuição de probabilidade Gamma é expressa matematicamente pela equação (3.5) (MONTGOMERY, 2005, p.94):

$$f(x) = \frac{\lambda_G^r x^{r-1} e^{-\lambda_G x}}{\int_0^{\infty} x^{r-1} e^{-x} dx} \quad \therefore r > 0 \quad (3.5)$$

- Os parâmetros r e λ da distribuição Gamma foram calculados, para cada variável, pela função *gamfit* do MATLAB[®]. A partir destes parâmetros, é possível determinar, juntamente com o uso da função *gamstat*, a média d_2 e o desvio padrão d_3 para a distribuição W , considerando-se cada uma das variáveis de processo analisadas;
- As equações (2.21), (2.22) e (2.23) são utilizadas para a determinação dos limites inferior e superior para a amplitude. Desta forma, são estabelecidos

os limites de corte, em que os valores que os excedem não são classificados como característicos de estado estacionário.

A segunda etapa consiste na avaliação dos resultados da rotina de seleção de variáveis. Os dados de entrada fornecidos às rotinas de seleção de variáveis, os quais foram obtidos pelos passos 2 e 3 expressos no item 3.3.2, não foram tratados para a exclusão dos períodos transientes. Mesmo assim, eles foram utilizados nas etapas posteriores de modelagem em função da possibilidade do fenômeno de inundação hidráulica do topo da torre caracterizar períodos transientes. A exclusão destes dados poderia acarretar perda de informações importantes para a predição do PFE.

Para a detecção do estado estacionário, utilizou-se apenas o *LSC* das variáveis selecionadas pela regressão linear por etapas. As outras variáveis fornecidas no arquivo de configuração não foram consideradas no teste. Ao se utilizar uma variável menos relevante para a determinação do estado estacionário, verifica-se a possibilidade de perda desnecessária de informação, dado que as variáveis relevantes podem se encontrar em condição estacionária no referido momento. Novamente, pelo critério de exclusão adotado, a identificação de apenas um dado transiente implica na exclusão de todos os dados da i -ésima amostra da matriz \mathbf{X} (i.e exclusão de uma linha da matriz de dados). O uso exclusivo do limite superior como critério de corte é justificável, pois é esperado que o aumento da amplitude das amostras se dê em condições de mudança de patamar operacional ou presença de distúrbios no sistema, o que conseqüentemente eleva o numerador da variável W , resultando em valores mais susceptíveis a exceder o limite superior de controle (*LSC*).

Após a determinação dos limites de corte (primeira etapa) e seleção das variáveis para aplicação destes limites (segunda etapa), tem início a terceira etapa. Com intuito de maximizar a aleatoriedade na etapa de modelagem, um novo sorteio dos dados é realizado, considerando-se como fonte original àqueles obtidos após as etapas de exclusão de *outliers*. No entanto, apenas as variáveis selecionadas na análise de regressão linear irão compor os dados de entrada. Aplica-se, então, os limites de corte. Os dados restantes são novamente enviados às rotinas de regressão linear por etapas, sendo que este procedimento está representado pela seta tracejada ligando os blocos 'Modelagem matemática do sistema' a 'Seleção das variáveis regressoras' na Figura

3.2. Desta forma, é verificada a hipótese da exclusão dos dados transientes possibilitar ou não a eliminação de mais alguma variável candidata a regressora do conjunto de dados de entrada, o que favoravelmente restringe ainda mais a possibilidade de inclusão de variáveis não significativas nos modelos.

Tem início, então, a etapa de modelagem matemática. Dois conjuntos de dados serão modelados: o conjunto contendo dados ‘transientes’ (i.e sem filtragem para o estado estacionário) e o conjunto de dados estacionários filtrados.

3.4.2 Aplicação das técnicas de modelagem

O usuário estabelece inicialmente o número de repetições da etapa de modelagem, considerando que em cada uma delas um novo conjunto de dados será sorteado, as principais variáveis serão selecionadas pela regressão linear por etapas e, caso o módulo de ‘Detecção de dados estacionários’ esteja ativado, um conjunto de dados filtrados adicional será obtido a cada iteração. O programa INFERATOR calcula o erro entre o variável resposta real y_i e a variável resposta estimada $f(\mathbf{x}_i, \mathbf{w})$, para cada uma das técnicas habilitadas no menu ‘Modelagem’, exibido na Figura 3.1. Os principais resultados de saída referentes à cada uma delas são armazenados em arquivos *.xls específicos. Outras informações, as quais podem ser de interesse do analista, são armazenadas em arquivos de extensão *.txt e *.mat.

Todas as soluções obtidas, considerando-se cada uma das técnicas habilitadas, são armazenadas em estruturas salvas em arquivos *.mat. Dentre as diversas soluções encontradas, o programa INFERATOR representa graficamente a que melhor atendeu à métrica estabelecida, tendo em vista cada uma das técnicas separadamente. As métricas possíveis são os seguintes:

- Seleciona-se a solução que apresentou a menor soma para o erro quadrático;

$$\varepsilon_{\text{erro}}(n) = \sum_i (y_i - f(\mathbf{x}_i, \mathbf{w}))^2 \quad (3.6)$$

- Seleciona-se a solução que apresentou a menor soma para o quadrado da diferença da derivada em relação ao tempo da variável resposta;

$$\varepsilon_{derivada}(n) = \sum_i \left(\frac{dy_i}{dt} - \frac{d[f(\mathbf{x}_i, \mathbf{w})]}{dt} \right)^2 \quad (3.7)$$

O segundo critério busca selecionar o modelo que melhor representa a tendência dos dados. Esta métrica é especialmente importante quando a inferência ajustada é utilizada em aplicações como controladores preditivos multivariáveis (MPC's), em que o sentido para o qual se dirige a variável (i.e aumento ou diminuição) deve ser conhecido para que uma correta ação de controle seja tomada.

A variável 'incerteza' foi definida para estimar a aproximação do modelo de predição, com intuito de verificar sua aplicabilidade prática para o controle do PFE. Define-se a variável 'incerteza' ξ (°C) como:

$$\xi = 2 (1,96 \sigma_{ERRO}) = 3,92 \sigma_{ERRO} \quad (3.8)$$

Esta variável foi calculada em cada iteração, considerando-se cada uma das técnicas separadamente. Outra variável calculada para verificação da qualidade da predição é o coeficiente de correlação linear entre o PFE real \mathbf{y} e o PFE estimado pelo modelo $\hat{\mathbf{y}}$, dado pela equação (2.116). Considerando-se o melhor modelo selecionado para cada uma das técnicas, foi verificada a hipótese de distribuição normal dos resíduos que, caso concretizada, permite estimar que o valor predito difere do valor real, em módulo, $1,96 \sigma_{ERRO}$ graus Celsius em cada uma dentre vinte amostras analisadas em laboratório.

A matriz de regressores \mathbf{X} e o vetor de dados de saída \mathbf{y} são argumentos de entrada comuns a todas as técnicas. Analogamente, a variável incerteza ξ (°C) e o valor numérico obtido para o coeficiente de correlação linear são também saídas fornecidas por cada uma delas. Apresentam-se, a seguir, as variáveis de entrada e saída particulares de cada técnica, bem como os parâmetros de modelagem:

- Regressão linear multivariada (MLR)
 - Variável de saída:
 1. Vetor de coeficientes para os modelos MLR ajustados;
- Regressão de componentes principais (PCR)
 - Variáveis de entrada:

1. Incremento da porcentagem de variabilidade explicada pelas componentes;
 2. Máxima porcentagem de variabilidade explicada pelas componentes;
- Variáveis de saída:
 1. Vetor de coeficientes para os modelos de PCR encontrados;
 2. Variabilidade explicada pelas componentes em cada simulação;
- Mínimos quadrados parciais (PLS)
 - Variáveis de entrada:
 1. Número máximo de variáveis latentes permitidas no modelo;
 - Variáveis de saída:
 1. Vetor de coeficientes para os modelos de PLS encontrados;
 2. Número de variáveis latentes utilizadas em cada simulação;
 - Redes neurais *Perceptron* de múltiplas camadas, função de performance baseada em regularização Bayesiana, dada pela equação (2.104), algoritmo de treinamento Levenberg-Marquardt, uma camada oculta com função de ativação logaritmo sigmoide e camada de saída linear (RNA1 RB)
 - Variáveis de entrada:
 1. Número máximo de neurônios para realização de testes subsequentes com diferentes arquiteturas de redes,

$$n_{NEURON(RNA1RB)} = 10;$$
 - Parâmetros de treinamento de redes:
 1. Número máximo de épocas para convergência,

$$n_{EPOCHS(RNA1RB)} = 1000;$$
 - Variáveis de saída:
 1. Pesos e bias para os modelos RNA1 RB encontrados;
 2. Número de parâmetros efetivos da rede, γ , em cada simulação;
 - Redes neurais *Perceptron* de múltiplas camadas, algoritmo *back-propagation* resiliente, uma camada oculta com função de ativação logaritmo sigmoide e camada de saída linear (RNA1 BP)
 - Variáveis de entrada:

1. Número máximo de neurônios para realização de testes subsequentes com diferentes arquiteturas de redes,

$$n_{NEURON(RNA1BP)} = 10;$$
 - Parâmetros de treinamento de redes:
 1. Número máximo de épocas para convergência,

$$n_{EPOCHS(RNA1BP)} = 1000;$$
 2. Taxa de aprendizagem $\eta_{RNA1BP} = 0,005;$
 - Variáveis de saída:
 1. Pesos e bias para os modelos RNA1 BP encontrados;
- Redes neurais *Perceptron* de múltiplas camadas, algoritmo *back-propagation* resiliente, duas camadas ocultas, função de ativação tipo logaritmo sigmoide na primeira e segunda camada, camada de saída linear (RNA2 BP)
 - Variáveis de entrada:
 1. Número máximo de neurônios para realização de testes subsequentes com diferentes arquiteturas de redes, considerando-se $n_{NEURON1(RNA2BP)} = 4$ na primeira camada e $n_{NEURON2(RNA2BP)} = 4$ na segunda camada;
 - Parâmetros de treinamento de redes:
 1. Número máximo de épocas para convergência,

$$n_{EPOCHS(RNA2BP)} = 1000;$$
 2. Taxa de aprendizagem $\eta_{RNA2BP} = 0,005;$
 - Variáveis de saída:
 1. Pesos e bias para os modelos RNA2 BP encontrados.
- Redes neurais de base radial, com função gaussiana na camada oculta (equação (2.93)) e camada de saída linear (RBF)
 - Variáveis de entrada:
 1. Número máximo de neurônios para realização de testes subsequentes com diferentes arquiteturas de redes, considerando-se $n_{NEURON1(RBF)} = 10;$

2. Parâmetro de espalhamento da curva gaussiana, sendo avaliados a cada iteração três níveis distintos, dados por

$$r_{RBF} = [0,1 \quad 1 \quad 10];$$

- Variáveis de saída:

1. Pesos e bias para os modelos RBF encontrados.

As etapas seguidas pelo programa são exibidas na Figura 3.10. A cada iteração, um conjunto de dados é sorteado para modelagem, sendo que os dados restantes são armazenados para etapa de validação. São modeladas diferentes arquiteturas em cada um dos diferentes tipos de técnica (i.e variação do número de neurônios, mudança do número de variáveis latentes e modificação da porcentagem de variabilidade explicada pelas componentes). Exemplificando o caso específico de redes neurais, o programa INFERATOR acrescenta neurônios ao modelo gradativamente, treinando e validando a rede de maneira subsequente até que o limite máximo de neurônios especificado seja atingido. O mesmo conjunto de dados de modelagem e validação é utilizado para avaliar as diferentes arquiteturas, permitindo comparar a eficiência de generalização de cada uma delas.

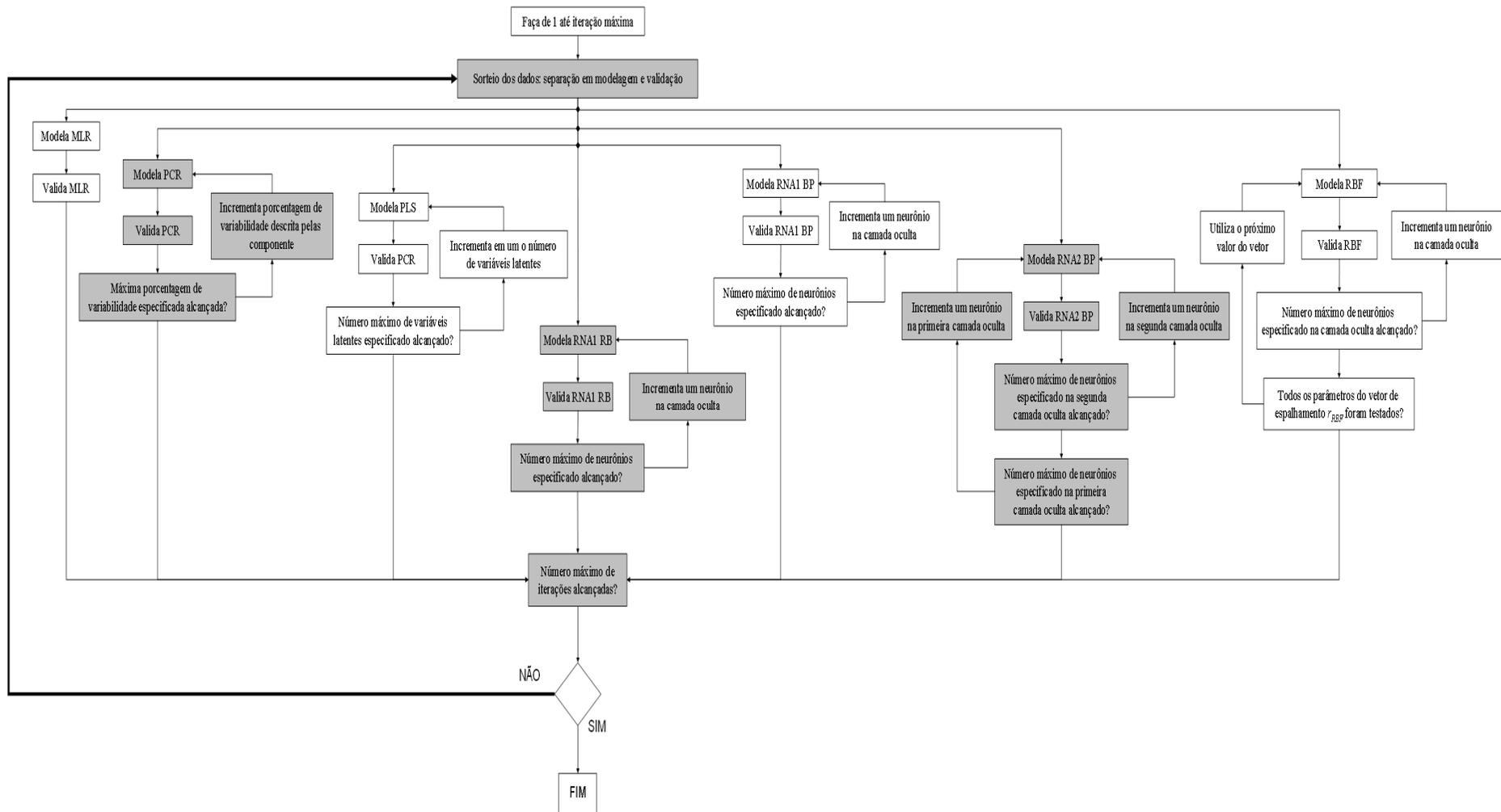


Figura 3.10: Fluxograma esquematizando as principais operações realizadas pelo programa na etapa de modelagem.

3.4.3 Seleção da arquitetura dos modelos

Para selecionar a arquitetura ótima para as diferentes técnicas, foram verificados os seguintes aspectos:

1. Qual arquitetura forneceu a menor ‘incerteza’, expressa pela equação (3.8);
2. Qual arquitetura forneceu o maior coeficiente de correlação entre os dados reais de PFE e o valor inferido pelo modelo;
3. Qual arquitetura forneceu a menor soma quadrática dos erros, expressa pela equação (3.6);
4. Qual arquitetura forneceu a menor soma quadrática para a derivada do erro em relação ao tempo, expressa pela equação (3.7);

Sendo assim, foram realizadas 30 simulações subseqüentes, cujas repetições são expressas pelo laço externo da Figura 3.10. Verificou-se, finalmente, a hipótese de normalidade dos resíduos, com intuito de estabelecer uma estimativa para a ‘incerteza’, expressa pela equação (3.8).

4 RESULTADOS E DISCUSSÃO

4.1 Coleta e pré-tratamento dos dados

4.1.1 Exclusão dos dados com perda de sinal

A Tabela 4.1 exibe o número de dados brutos coletados em cada mês. Estes foram filtrados para eliminação das amostras com perda de sinal.

Tabela 4.1: Disponibilidade de dados após operação de filtragem para perda de sinal.

Mês cujos dados foram coletados	Número de dados brutos coletados do PI	Dados restantes após filtragem por perda de sinal	Dados excluídos	Porcentagem de dados excluídos no período (%)
12/2008	8641	7577	1064	0,59
01/2009	8641	8589	52	0,03
02/2009	8641	8620	21	0,01
03/2009	8065	8018	47	0,03
04/2009	8641	8630	11	0,01
05/2009	8641	5451	3190	1,77
06/2009	8641	7457	1184	0,66
07/2009	8641	8031	610	0,34
08/2009	8641	8121	520	0,29
09/2009	8641	8413	228	0,13
10/2009	8629	8268	361	0,20
11/2009	8641	5918	2723	1,51
12/2009	8641	8640	1	0,00
01/2010	8641	8589	52	0,03
02/2010	8641	8600	41	0,02
03/2010	8065	8063	2	0,00
04/2010	8641	8641	0	0,00
05/2010	8641	8590	51	0,03
06/2010	8641	8641	0	0,00
07/2010	8641	8641	0	0,00
08/2010	8641	8634	7	0,00
TOTAL	180297	170132	10165	5,64

Frente ao número total de dados coletados, a quantidade de amostras restantes foi considerada satisfatória. Buscou-se identificar os medidores com maior frequência de perda de sinal. Eles estão listados na Tabela 4.2:

Tabela 4.2: Número de falhas de sinal identificadas para os medidores das variáveis coletadas.

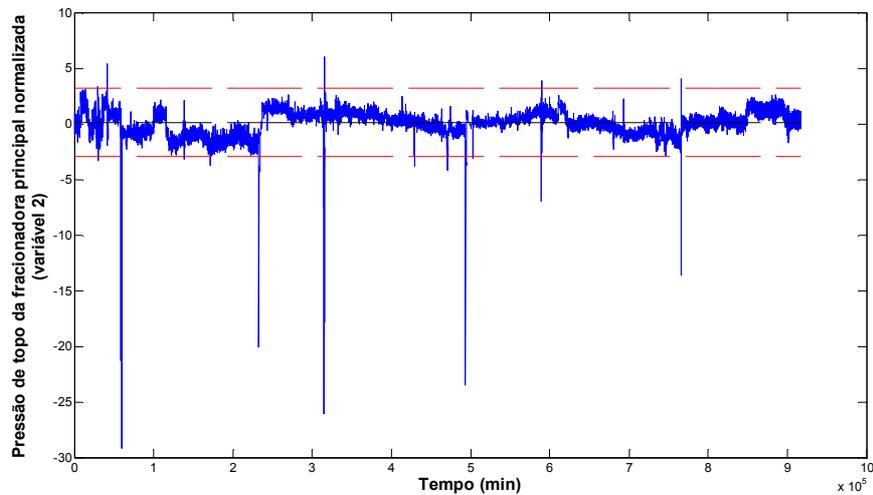
Variável	Descrição	Falhas de sinal
33	Diferencial de pressão entre a panela de HCO e o fundo	6994
3	Temperatura de topo da fracionadora principal	2698
5	Vazão do refluxo de topo para a fracionadora	799
32	Diferencial de pressão entre as panelas de LCO e HCO	417
23	Diferencial de pressão entre o topo e a panela de LCO	258
12	Vazão A de vapor para o riser	146
21	Temperatura no topo do riser	137
34	Diferencial de pressão entre o topo e o fundo	99
2	Pressão de topo da fracionadora principal	90
16	Vazão E de vapor para o riser	87
24	Temperatura da panela de retirada de nafta 'pesada'	84
4	Temperatura do refluxo de topo para torre	78
18	Vazão B de gasóleo para riser	78
27	Vazão de LCO na saída da fracionadora principal	76
17	Vazão A de gasóleo para riser	74
28	Retirada de LCO para tanque	74
10	Vazão C de vapor para retificação no vaso separador	73
22	Temperatura na base do riser	73
31	Vazão de HCO para integração energética	72
29	Vazão de LCO para integração energética	69
13	Vazão B de vapor para o riser	66
14	Vazão C de vapor para o riser	66
19	Vazão de nafta 'leve' de coque	65
8	Vazão A de vapor para retificação no vaso separador	62
6	Vazão de retirada de nafta 'leve' craqueada	61
7	Pressão do vapor de retificação no vaso separador	61
11	Vazão D de vapor para retificação no vaso separador	61
9	Vazão B de vapor para retificação no vaso separador	60
15	Vazão D de vapor para o riser	56
26	Vazão de retirada de nafta 'pesada' para tratamento	54
20	Vazão de gasóleo de coque	53
25	Vazão de refluxo circulante de nafta 'pesada'	50
30	Temperatura do refluxo de HCO para integração energética	41
35	PFE da nafta craqueada	0

As medições das variáveis 3 e 33 apresentaram maior recorrência de falha. Ressalta-se que a temperatura de topo é, em teoria, importante variável para descrição do PFE, em função do equilíbrio termodinâmico da região do topo da fracionadora. No entanto, em

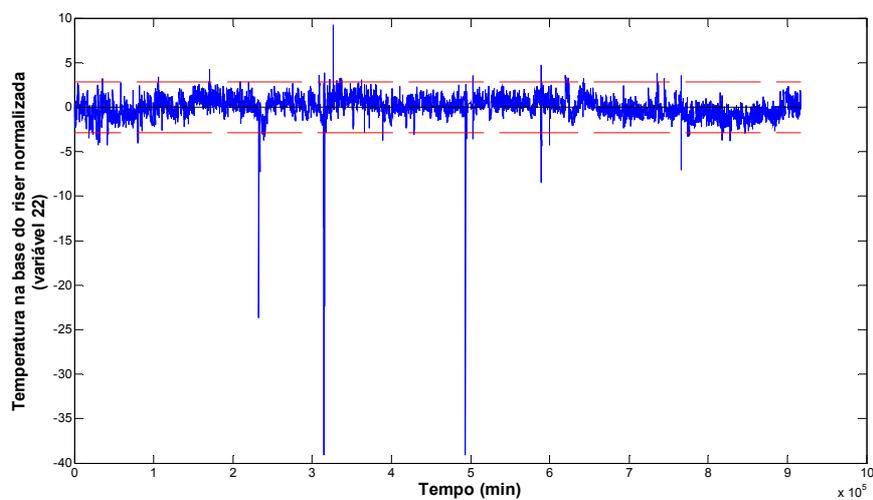
função da pequena quantidade de dados expurgados, essa exclusão não acarretou grandes perdas de informações.

4.1.2 Exclusão de *outliers*

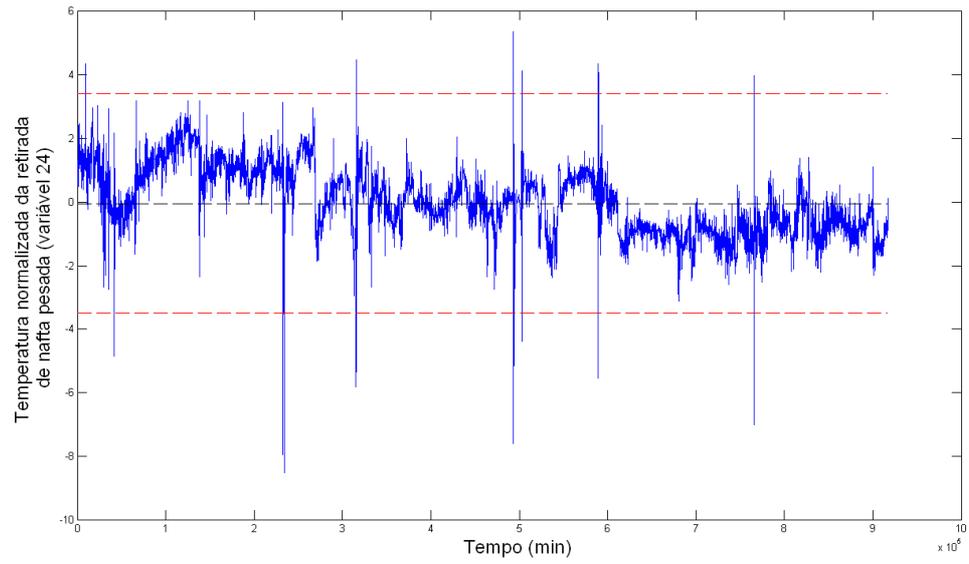
As variáveis selecionadas para aplicação do identificador de Hampel e exclusão de *outliers* estão relacionadas na Tabela 3.2. Os gráficos obtidos para estas variáveis, com seus respectivos limites de corte, são exibidos a seguir:



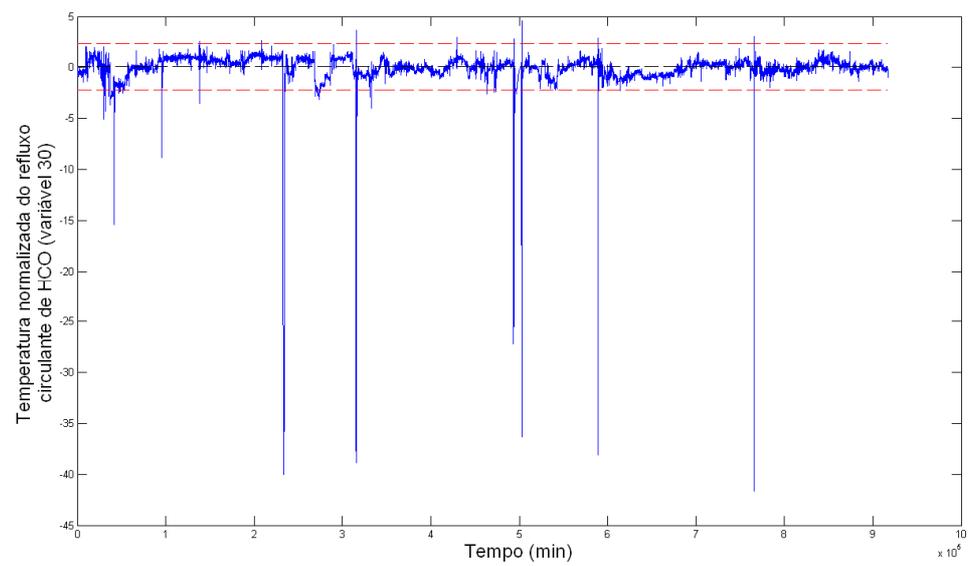
(a)



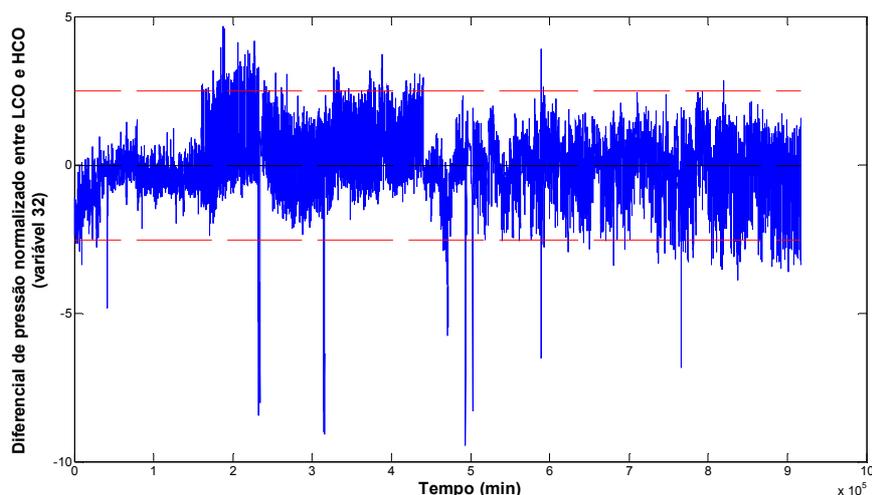
(b)



(c)



(d)



(e)

Figura 4.1: Limites de corte obtidos para as variáveis normalizadas selecionadas para aplicação do identificador de Hampel; a) pressão de topo da fracionadora principal; b) temperatura na base do riser; c) temperatura da panela de retirada de nafta ‘pesada’; d) temperatura do refluxo circulante de HCO; e) diferencial de pressão entre as panelas de LCO e HCO.

A Tabela 4.3 exibe a disponibilidade de dados após filtragem. Percebe-se que mais do que 5% dos dados foram excluídos. Caso o desvio padrão e a média amostral fossem utilizados para obtenção das linhas de corte, seria esperado que cerca de 5% dos dados fossem expurgados. O identificador de Hampel demonstra maior rigor na detecção de *outliers*.

Tabela 4.3: Disponibilidade de dados após aplicação do identificador de Hampel.

Número de amostras originais	Número de amostras após exclusão de outliers	Porcentagem excluída (%)
170132	157481	7,43

4.2 Seleção das variáveis regressoras

4.2.1 Obtenção dos gráficos de correlação cruzada

Com a utilização do módulo de ‘Função de correlação cruzada’ exibido no menu da Figura 3.1, aplicado às variáveis candidatas a regressoras e à variável resposta PFE, foram obtidos os gráficos exibidos nesta seção. Foi verificada, então, a existência de correlação linear significativamente estatística em períodos compatíveis com o atraso de transporte do sistema.

Para uma estimativa grosseira do tempo morto, segundo a equação (3.1), os volumes dos principais vasos das seções de fracionamento, seção de recuperação de gases e seção de tratamentos foram coletados. O tempo de residência foi estimado considerando-se todos estes vasos em 20% do nível máximo, 50% do nível máximo 80% do nível máximo. Esta é uma condição sabidamente irreal no processo (i.e manter todos os vasos com a mesma porcentagem do nível máximo), porém auxilia na estimativa da ordem de grandeza para o tempo morto. A Tabela 4.4 exhibe estes resultados:

Tabela 4.4: Estimativa para o atraso de transporte.

Porcentagem do nível dos vasos (%)	Tempo de residência (h)
20	2,1
50	5,1
80	8,2

Sendo assim, buscou-se selecionar as variáveis cujos picos de correlação ocorram entre 1,5 horas até 8,5 horas. Por experiência prévia, vários operadores relatam que o tempo de renovação do inventário é de aproximadamente 4 horas.

Verificou-se a coerência dos gráficos de correlação cruzada obtidos. Eles indicam se um aumento da variável analisada em questão implica no aumento do PFE (i.e valores positivos para a correlação cruzada) ou diminuição do PFE (i.e valores negativos para a correlação cruzada). A provável relação de causa-efeito responsável pelo fenômeno físico-químico observado, considerando-se os valores de correlação obtidos entre 90 minutos e 510 minutos (i.e coerentes com a ordem de grandeza estimada para o atraso de transporte), é brevemente sugerida a seguir:

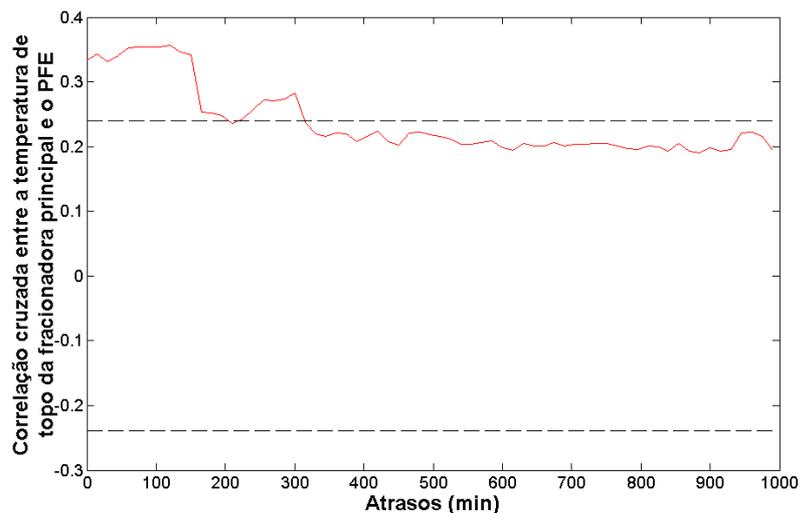


Figura 4.2: Gráfico de correlação cruzada entre o PFE e a temperatura de topo da fracionadora principal.

Espera-se que o aumento da temperatura de topo implique no aumento do PFE (Figura 4.2), em função da maior evaporação de compostos mais densos referentes ao final da faixa de destilação da nafta ‘leve’ e início da faixa de destilação da retirada de nafta ‘pesada’.

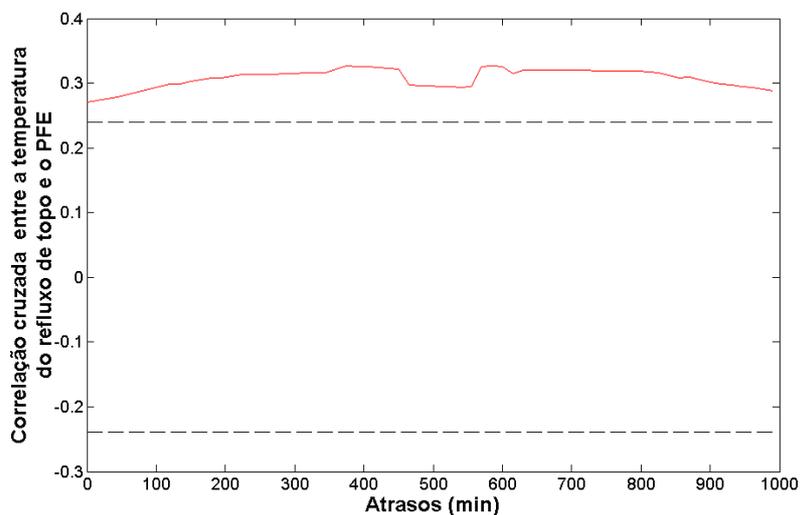


Figura 4.3: Gráfico de correlação cruzada entre o PFE e a temperatura do refluxo de topo da fracionadora principal.

É possível que o pequeno aumento da temperatura do refluxo de topo esteja correlacionado ao aumento do PFE. Uma maior temperatura do refluxo de topo é

consequência direta da maior temperatura do inventário contido no vaso de topo, o que termodinamicamente indica maior fração de compostos ‘pesados’, de maior temperatura de ebulição, presentes no equilíbrio. (Figura 4.3).

Contudo, sabe-se que o fator que mais influencia a temperatura do inventário do vaso de topo diz respeito à capacidade de sub-resfriamento dos condensadores de topo. Esta capacidade está diretamente ligada à temperatura da água de refrigeração: quanto mais quente se encontrar a água de refrigeração, menor será o diferencial de temperatura para a transferência de calor, ocasionando um menor sub-resfriamento da nafta no vaso de topo.

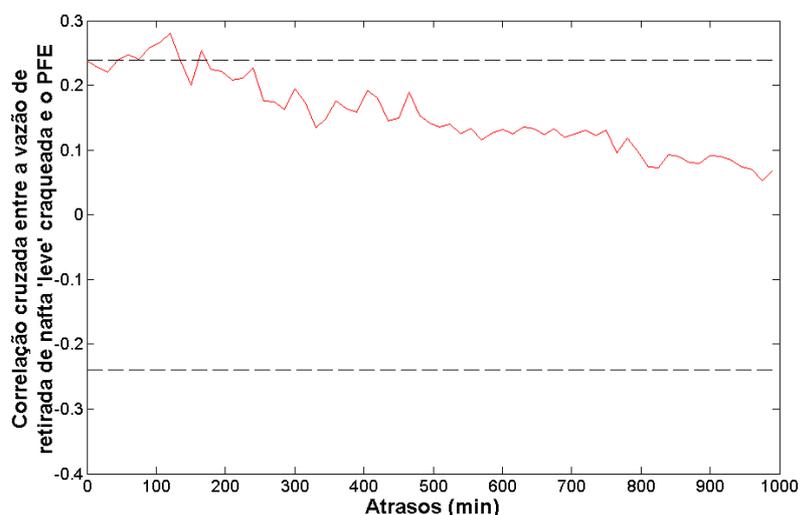


Figura 4.4: Gráfico de correlação cruzada entre o PFE e a vazão de nafta ‘leve’ produzida.

Uma maior produção de nafta ‘leve’ craqueada indica incorporação do corte de destilação inicial da nafta ‘pesada’ a esta corrente, o que resulta na elevação de seu PFE (Figura 4.4).

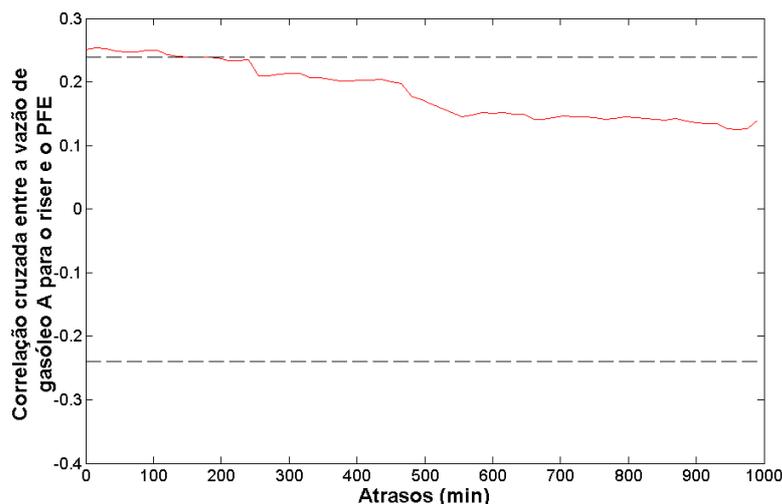


Figura 4.5: Gráfico de correlação cruzada entre o PFE e a vazão de carga A de gásóleo para a unidade.

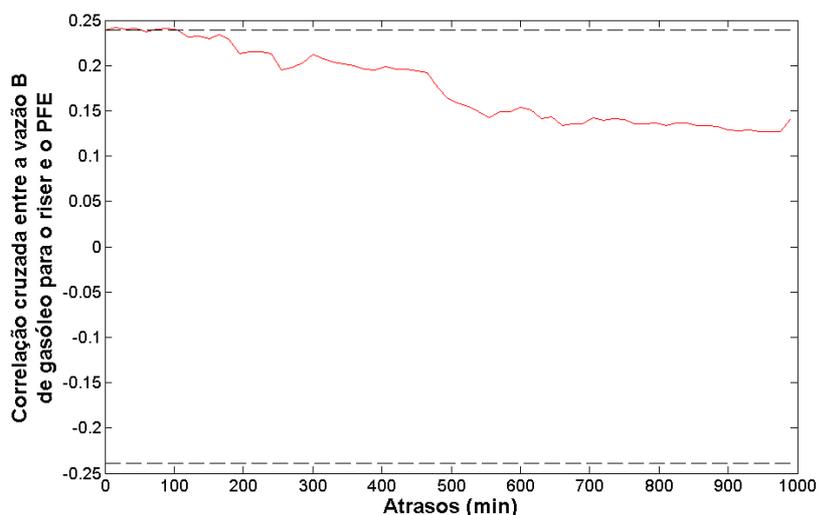


Figura 4.6: Gráfico de correlação cruzada entre o PFE e a vazão de carga B de gásóleo para a unidade.

Um aumento na vazão de carga pode implicar, indiretamente, no aumento do PFE, em especial nas condições de inundação hidráulica da seção de topo. A redução desta vazão dificulta o alcance das condições de inundação na seção de topo, em função do menor refluxo interno de gases e líquidos ao longo da torre. Além disso, uma menor vazão de carga facilita a remoção do calor nas seções intermediárias da torre, aumentando a condensação de hidrocarbonetos abaixo da seção de topo. Desta forma, evita-se o

arraste excessivo de compostos mais ‘pesados’ para o topo e sua consequente incorporação à nafta ‘leve’ (Figura 4.6);

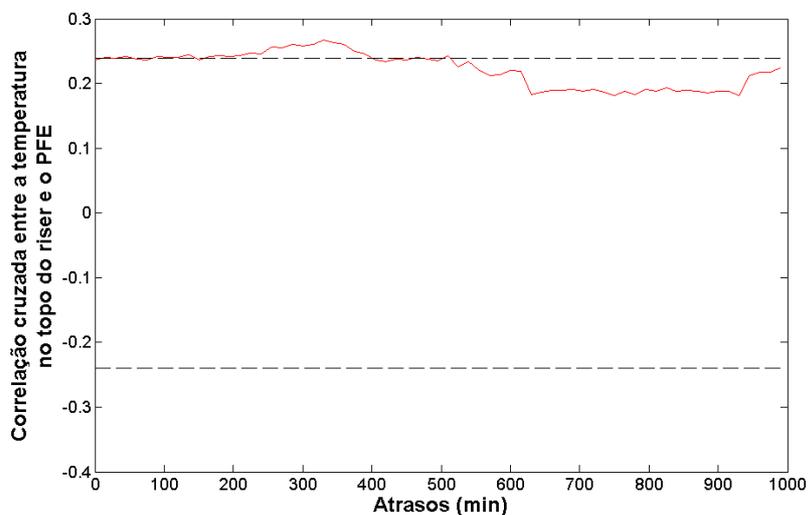


Figura 4.7: Gráfico de correlação cruzada entre o PFE e a temperatura de reação no topo do riser.

O aumento da temperatura de reação ao longo do *riser* favorece uma operação mais severa, com uma maior taxa de conversão dos hidrocarbonetos ‘pesados’ presentes na carga em frações mais ‘leves’. Favorece também a produção de coque, oriundo das reações secundárias, o qual se deposita sobre o catalisador. Este aumento do rendimento de produtos ‘leves’ possibilita uma maior incorporação destes às retiradas superiores da torre. Além disso, a elevação da temperatura de reação favorece a elevação da temperatura da carga da torre, o que contribui para maior evaporação de hidrocarbonetos para as seções superiores. Sugere-se que estes fatores contribuam para o aumento do PFE com o aumento da temperatura de reação (Figura 4.7).

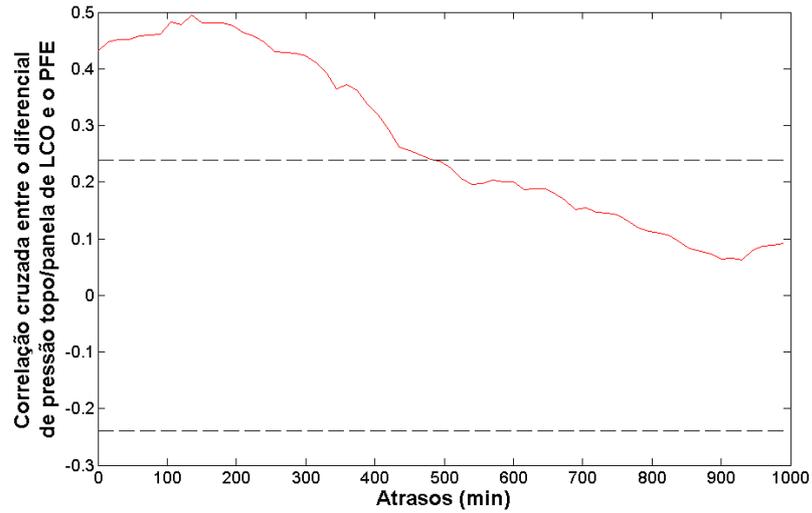


Figura 4.8: Gráfico de correlação cruzada entre o PFE e o diferencial de pressão topo/panela de LCO.

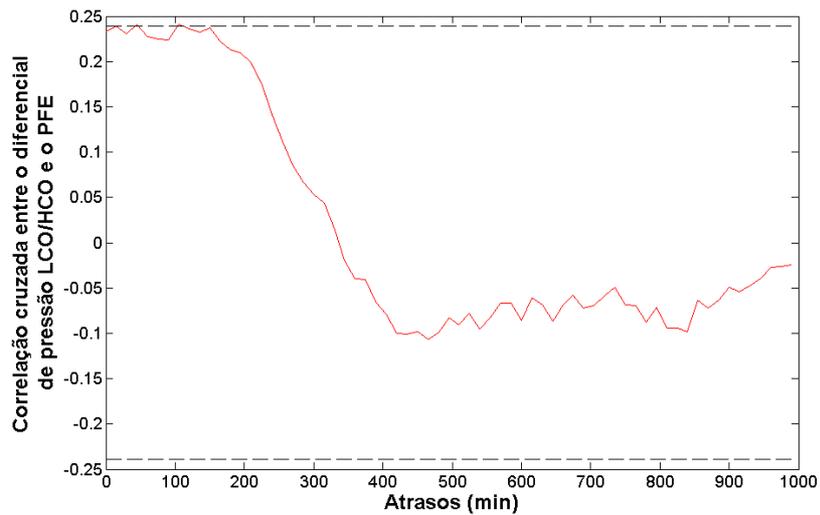


Figura 4.9: Gráfico de correlação cruzada entre o PFE e o diferencial de pressão panela de LCO/panela de HCO.

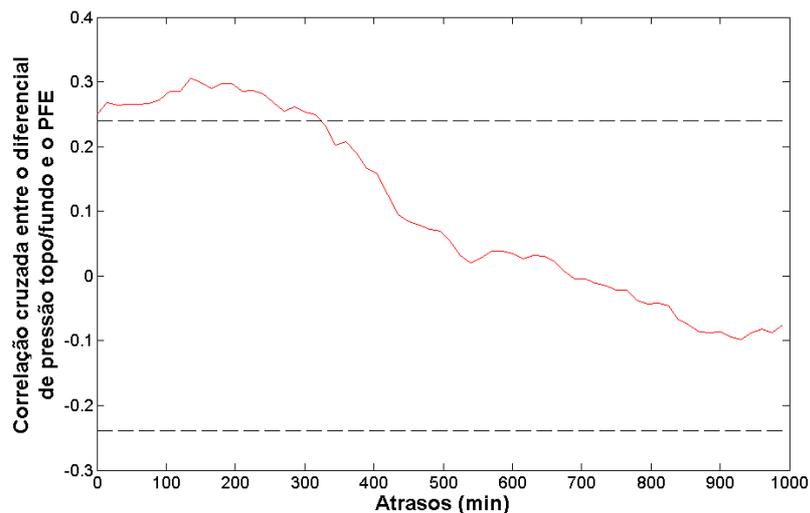


Figura 4.10: Gráfico de correlação cruzada entre o PFE e o diferencial de pressão topo/fundo.

A elevação de medições de diferencial de pressão ao longo da torre é um indício de inundação em alguma de suas seções. No caso, é sabido que o corte da vazão de retirada de nafta ‘pesada’ faz com que o diferencial de pressão entre o topo e a panela de LCO se eleve de maneira significativa, o que indica que a inundação ocorre fundamentalmente nesta seção. Este fenômeno pode ocasionar arraste excessivo de gotas de líquido para os estágios superiores, com conseqüente perda da eficiência de fracionamento e elevação do PFE da nafta ‘leve’ craqueada (Figura 4.8). O mesmo fenômeno é identificado, porém com menor sensibilidade, pelas outras duas medições de diferencial de pressão (Figura 4.9 e Figura 4.10).

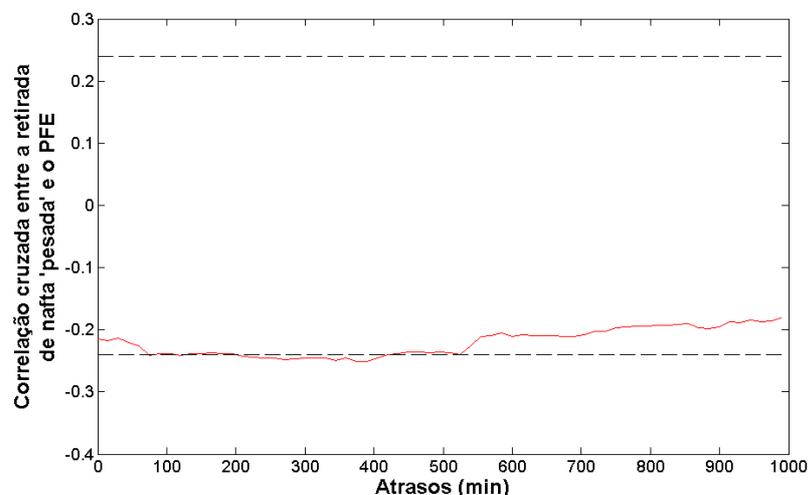


Figura 4.11: Gráfico de correlação cruzada entre o PFE e a retirada de nafta ‘pesada’.

A redução da vazão de nafta ‘pesada’ para tratamento favorece a inundação da seção do topo da torre. Nestas circunstâncias, a retenção do fluxo ascendente de gás, em função do maior nível de líquido nos internos, faz com que a pressão a montante da região inundada se eleve. Ao vencer essa coluna, o gás provoca um arraste excessivo de líquido com faixa de destilação superior à da nafta ‘leve’ de craqueamento, contribuindo para elevação do seu PFE (Figura 4.11);

Os gráficos de correlação cruzada possibilitaram estimar uma ordem de grandeza para o atraso de transporte. Neste caso, o diferencial de pressão entre o topo e a panela de LCO (variável 23) foi identificado como a variável de maior correlação linear com a resposta, sendo seguida pela temperatura de topo da torre (variável 3). No caso do diferencial de pressão, verifica-se que o gráfico de correlação tem a forma de um pico, o qual se mantém elevado até cerca de 180 minutos antes do momento da amostragem (Figura 4.8). A temperatura de topo também apresentou um formato de pico de correlação, com patamar elevado até cerca de 150 minutos antes do período de amostragem (Figura 4.2). Padronizou-se, portanto, a escolha do tempo de 180 minutos para o atraso de transporte das variáveis, ainda que os seus picos de correlação ocorressem em períodos distintos, como é o caso da temperatura de reação (Figura 4.7). Esta medida foi adotada tendo em vista a implementação posterior da inferência ajustada, a qual utilizará dados de processo online, todos eles referentes ao instante temporal presente, para fornecer uma estimativa baseada em modelo estático para o PFE. Sugere-se que as diferenças nos picos de correlação sejam função não apenas do atraso de transporte, mas também das diferentes dinâmicas de resposta do PFE às modificações das variáveis regressoras.

4.2.2 Regressão linear por etapas

A matriz de dados X das variáveis candidatas é construída conforme procedimento relatado na seção 3.3.2. Como a frequência de amostragem da nafta ‘leve’ para análise do PFE é aproximadamente diária, verifica-se uma grande redução na disponibilidade de dados para ajuste dos modelos:

Tabela 4.5: Redução do número de dados disponíveis ocasionada pela montagem da matriz das variáveis candidatas à regressoras.

Número de amostras após exclusão de outliers	Número de amostras após montagem da matriz de dados
157481	619

A Tabela 4.6 exhibe as informações presentes no arquivo de configuração para a regressão linear por etapas:

Tabela 4.6: Dados numéricos referentes ao arquivo de configuração para regressão linear por etapas.

Variável	Identificador da coluna na matriz de dados	Atraso referente ao pico de correlação (min)	Janela para captura das médias (min)
Temperatura de topo da fracionadora principal	3	180	30
Temperatura do refluxo de topo	4	180	30
Vazão de retirada de nafta 'leve' craqueada	6	180	30
Vazão A de gasóleo para riser	17	180	30
Vazão B de gasóleo para riser	18	180	30
Temperatura no topo do riser	21	180	30
Diferencial de pressão entre topo/panela LCO	23	180	30
Vazão de retirada de nafta 'pesada'	26	180	30
Diferencial de pressão LCO/HCO	32	180	30
Diferencial de pressão topo/fundo	34	180	30

Executando-se este módulo do programa INFERATOR, considerando-se os níveis de significância de 5% e 1% na regressão linear por etapas, para um total de 100 repetições deste procedimento, foi possível verificar a escolha do método para as seguintes variáveis:

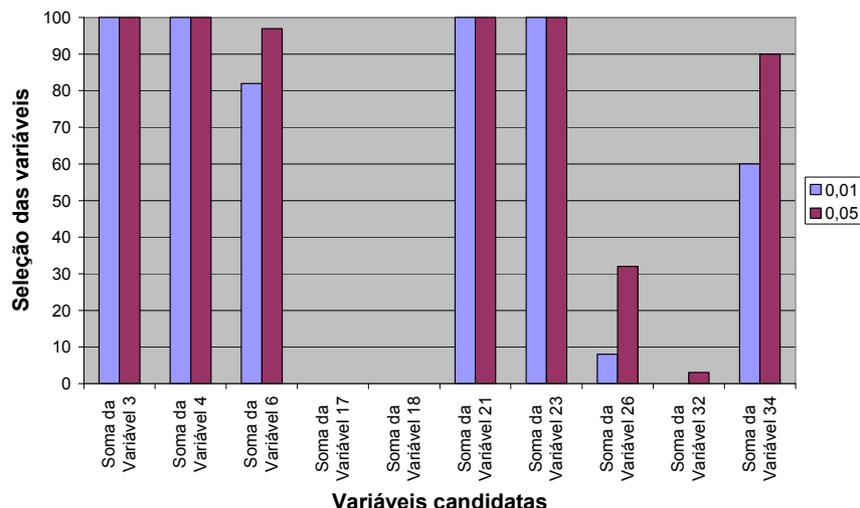


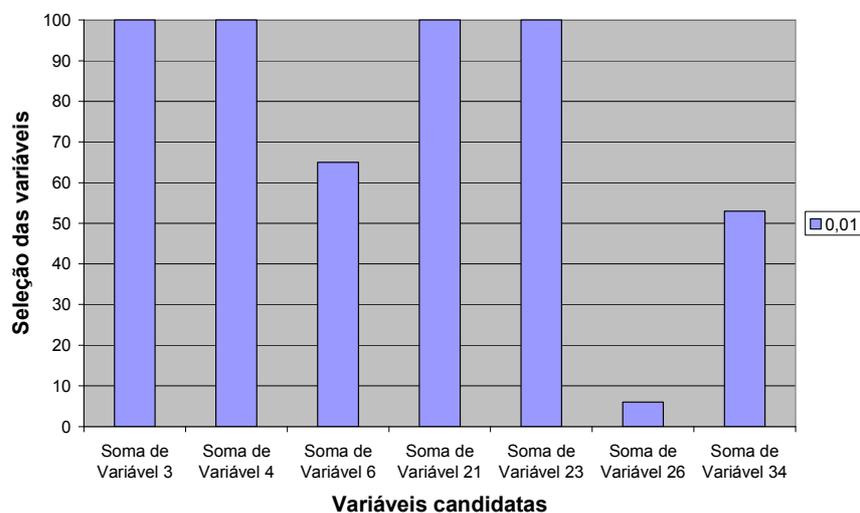
Figura 4.12: Variáveis selecionadas pela regressão linear por etapas, para nível de significância de 0,05 e 0,01.

Este gráfico expressa quantas vezes, em um total de 100 repetições do procedimento de regressão linear por etapas, as variáveis foram selecionadas. É possível perceber que as variáveis 3, 4, 21 e 23 foram selecionadas em todas as 100 repetições do procedimento, considerando-se amostragens aleatórias dos dados de modelagem. Percebe-se, ainda, que a variação do nível de significância não resultou em respostas muito diferentes no que diz respeito às variáveis regressoras selecionadas (Figura 4.12). Além disso, três variáveis candidatas foram descartadas após essa análise, a saber: as vazões A e B de carga de gásóleo e o diferencial de pressão entre a panela de LCO e a panela de HCO.

A Tabela 4.7 exhibe as variáveis selecionadas. A priori, as variáveis temperatura de topo (variável 3), temperatura do refluxo de topo (variável 4), temperatura de reação (variável 21) e diferencial de pressão topo/panela de LCO (variável 23) devem fazer parte do modelo, pois as mesmas foram selecionadas em todos os procedimentos de regressão linear por etapas realizados (Figura 4.12). Realiza-se novo procedimento de regressão linear para determinação da permanência ou exclusão das variáveis vazão de produção de nafta craqueada, (variável 6), vazão de retirada de nafta ‘pesada’ (variável 26) e diferencial de pressão entre topo e fundo (variável 34):

Tabela 4.7: Variáveis selecionadas para novo procedimento de regressão linear por etapas.

Variável	Identificador da coluna na matriz de dados	Atraso referente ao pico de correlação (min)	Janela para captura das médias (min)
Temperatura de topo da fracionadora principal	3	180	30
Temperatura do refluxo de topo	4	180	30
Vazão de retirada de nafta 'leve' craqueada	6	180	30
Temperatura no topo do riser	21	180	30
Diferencial de pressão entre topo/panela LCO	23	180	30
Vazão de retirada de nafta 'pesada'	26	180	30
Diferencial de pressão topo/fundo	34	180	30

**Figura 4.13: Variáveis selecionadas pela regressão linear por etapas, para nível de significância de 0,01.**

Analisando-se a Figura 4.13, optou-se pela escolha do número mínimo de regressores para o modelo, o que inclui as variáveis expressas na Tabela 4.8, as quais foram selecionadas em todas as 100 repetições realizadas para o procedimento de regressão linear por etapas, considerando-se que os dados de modelagem foram sorteados aleatoriamente.

Tabela 4.8: Variáveis selecionadas para os procedimentos posteriores de análise de clusters e modelagem matemática de sistemas.

Variável	Identificador da coluna na matriz de dados	Atraso referente ao pico de correlação (min)	Janela para captura das médias (min)
Temperatura de topo da fracionadora principal	3	180	30
Temperatura do refluxo de topo	4	180	30
Temperatura no topo do riser	21	180	30
Diferencial de pressão entre topo/panela LCO	23	180	30

A prioridade de seleção das variáveis candidatas também é fornecida pelo programa INFERATOR. Conforme relatado no item 2.5.2, as variáveis de maior correlação linear são selecionadas nos passos iniciais do algoritmo.

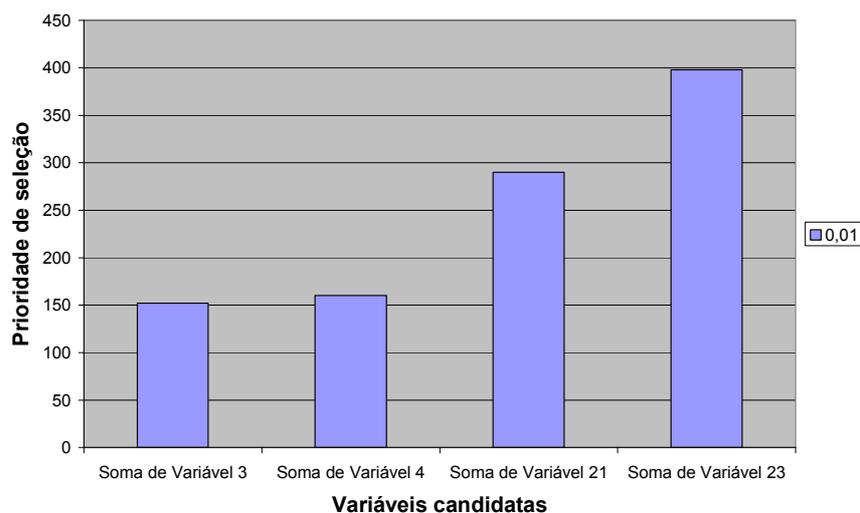


Figura 4.14: Ordem preferencial de seleção das variáveis pela rotina de regressão linear por etapas.

A variável de maior correlação linear com a resposta (i.e diferencial de pressão topo/panela de LCO) foi, de fato, selecionada preferencialmente. Porém, a variável temperatura de topo era esperada como a segunda na hierarquia de prioridade, tendo em vista os valores presentes em seu gráfico de correlação cruzada (Figura 4.2). No entanto, ela foi preterida em relação à temperatura no topo do *riser* e a temperatura do refluxo de topo no que diz respeito à prioridade de seleção, conforme Figura 4.14, o que significa que a adição destes regressores produziu uma estatística F mais elevada do a adição da

temperatura de topo. De fato, o teste parcial F considera, a partir da primeira variável incluída, o aumento da soma quadrática de regressão com a inclusão do novo regressor; o valor de correlação individual desta variável candidata com a variável resposta não é considerado diretamente.

4.2.3 Análise de clusters

Uma amostra aleatória para modelagem foi sorteada para aplicação da análise de clusters, utilizando-se 80% dos dados disponíveis no conjunto original. Iniciou-se a avaliação considerando-se a possibilidade de existência de dois clusters a priori: um para PFE alto e outro para PFE baixo. A inserção gradativa das variáveis selecionadas pela regressão linear por etapas no vetor do algoritmo das k -médias inclui gradativamente uma dimensão adicional aos centroides, modificando os elementos classificados em cada um dos grupos estabelecidos. A seguir, exibe-se o efeito da adição gradativa das variáveis:

Tabela 4.9: Primeira etapa de inclusão de variáveis no vetor de k -médias, para dois clusters a priori.

Variável	Inclusão
Temperatura de topo da fracionadora principal	NÃO
Temperatura do refluxo de topo	NÃO
Temperatura no topo do riser	NÃO
Diferencial de pressão entre topo/panela LCO	SIM
PFE	SIM

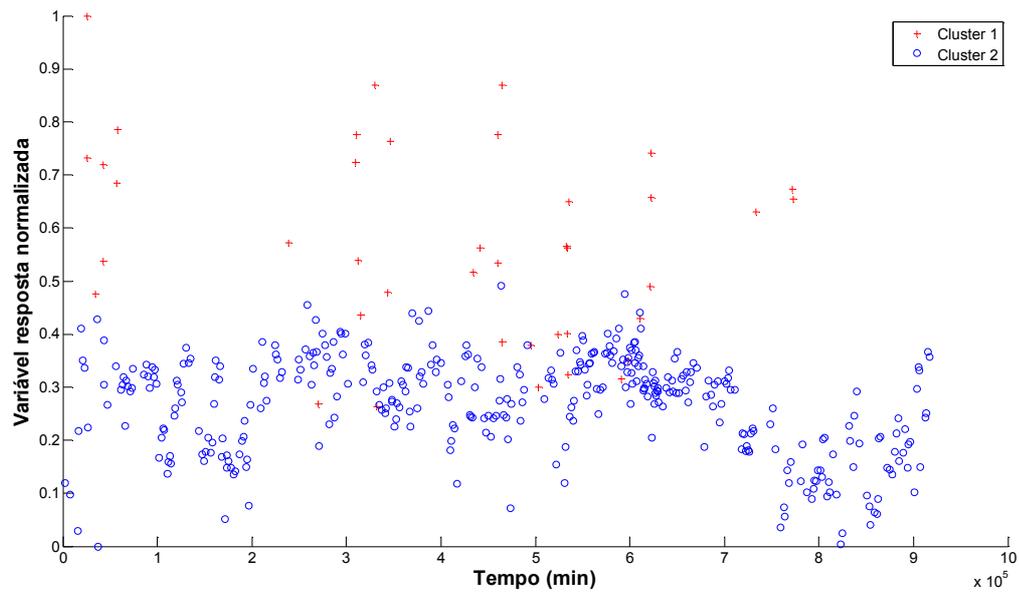


Figura 4.15: Clusters formados considerando-se as variáveis PFE e diferencial de pressão topo/panela de LCO no vetor das k-médias.

Tabela 4.10: Segunda etapa de inclusão de variáveis no vetor de k-médias, para dois clusters a priori.

Variável	Inclusão
Temperatura de topo da fracionadora principal	NÃO
Temperatura do refluxo de topo	SIM
Temperatura no topo do riser	NÃO
Diferencial de pressão entre topo/panela LCO	SIM
PFE	SIM

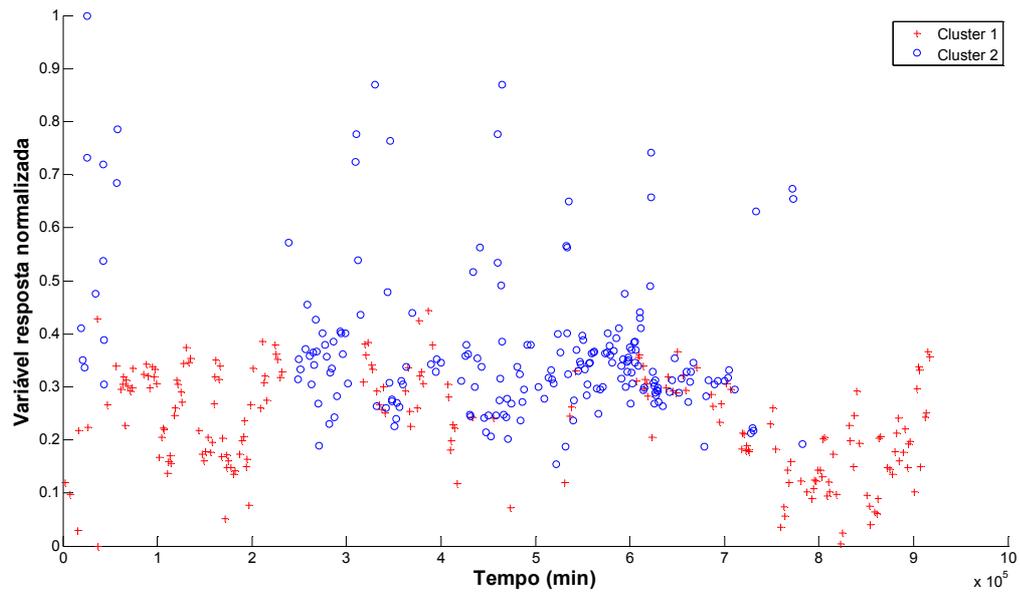


Figura 4.16: Clusters formados considerando-se as variáveis PFE, diferencial de pressão topo/panela de LCO e temperatura do refluxo de topo no vetor das k-médias.

Tabela 4.11: Terceira etapa de inclusão de variáveis no vetor de k-médias, para dois clusters a priori.

Variável	Inclusão
Temperatura de topo da fracionadora principal	NÃO
Temperatura do refluxo de topo	SIM
Temperatura no topo do riser	SIM
Diferencial de pressão entre topo/panela LCO	SIM
PFE	SIM

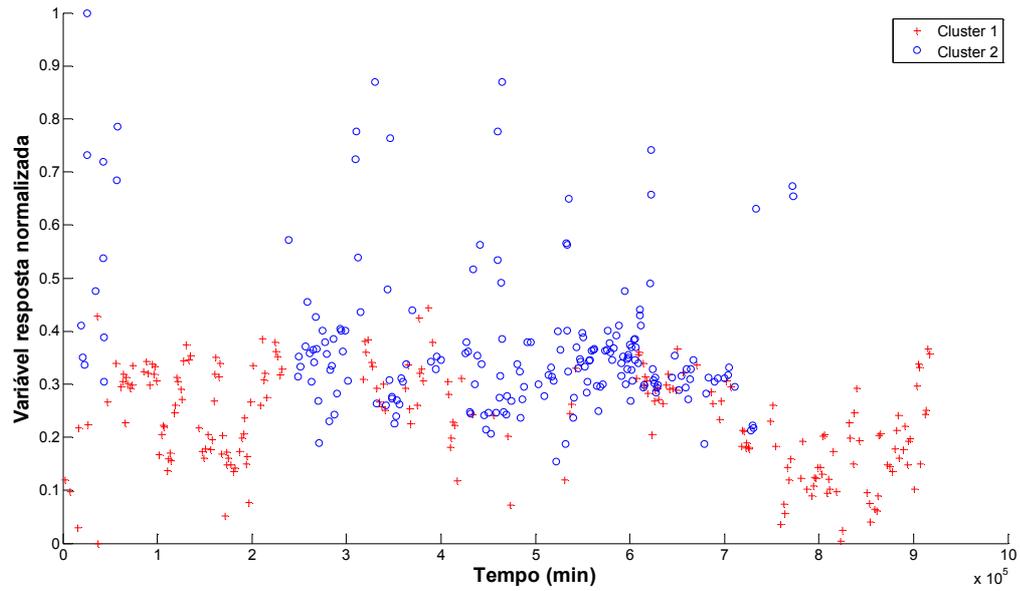


Figura 4.17: Clusters formados considerando-se as variáveis PFE, diferencial de pressão topo/panela de LCO, temperatura do refluxo de topo e temperatura de reação no vetor das k-médias.

Tabela 4.12: Quarta etapa de inclusão de variáveis no vetor de k-médias, para dois clusters a priori.

Variável	Inclusão
Temperatura de topo da fracionadora principal	SIM
Temperatura do refluxo de topo	SIM
Temperatura no topo do riser	SIM
Diferencial de pressão entre topo/panela LCO	SIM
PFE	SIM

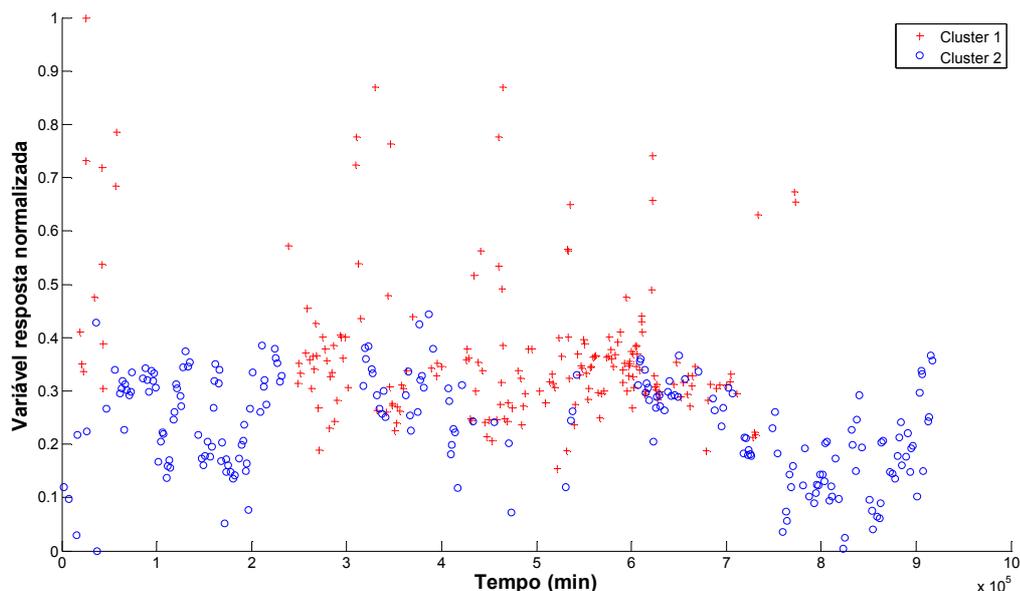


Figura 4.18: Clusters formados considerando-se as variáveis PFE, diferencial de pressão topo/panela de LCO, temperatura do refluxo de topo, temperatura de reação e a temperatura de topo no vetor das k-médias.

Percebe-se que a melhor partição natural dos clusters formados ocorre na Figura 4.15, em que é possível verificar um conjunto de PFE alto (cluster 1) e outro de PFE baixo (cluster 2). Considerando-se tanto a variável resposta PFE quanto o regressor diferencial de pressão topo/panela de LCO, ambos normalizados entre 0 e 1, os seguintes centroides são obtidos:

Tabela 4.13: Centroides dos clusters referentes à Figura 4.15.

Variável	Centroide	
	Cluster 1	Cluster 2
Diferencial de pressão entre topo/panela LCO	0,7646	0,3226
PFE	0,5697	0,2709

O centroide do cluster 1 apresenta uma média mais elevada para o diferencial de pressão e para o PFE, quando comparado ao centroide do cluster 2. Isto indica que valores mais elevados do diferencial de pressão topo/panela de LCO favorecem valores mais elevados para o PFE. Este fenômeno é explicado pelo efeito de arraste mecânico verificado nas condições de inundação.

Analisa-se, então, a possibilidade de existência de três clusters a priori: PFE baixo, PFE médio e PFE elevado. A inserção gradativa das variáveis candidatas à regressoras na ordem de seleção do procedimento de regressão linear por etapas resultou em:

Tabela 4.14: Primeira etapa de inclusão de variáveis no vetor de k-médias, para três clusters a priori.

Variável	Inclusão
Temperatura de topo da fracionadora principal	NÃO
Temperatura do refluxo de topo	NÃO
Temperatura no topo do riser	NÃO
Diferencial de pressão entre topo/panela LCO	SIM
PFE	SIM

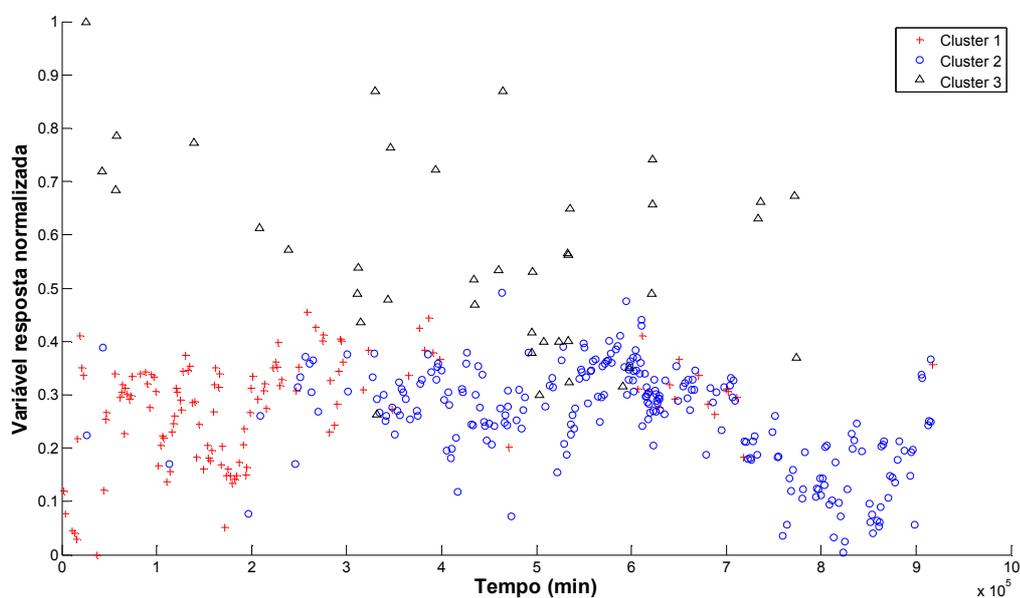


Figura 4.19: Clusters formados considerando-se as variáveis PFE e diferencial de pressão topo/panela de LCO no vetor das k-médias.

Tabela 4.15: Segunda etapa de inclusão de variáveis no vetor de k-médias, para três clusters a priori.

Variável	Inclusão
Temperatura de topo da fracionadora principal	NÃO
Temperatura do refluxo de topo	NÃO
Temperatura no topo do riser	SIM
Diferencial de pressão entre topo/panela LCO	SIM
PFE	SIM

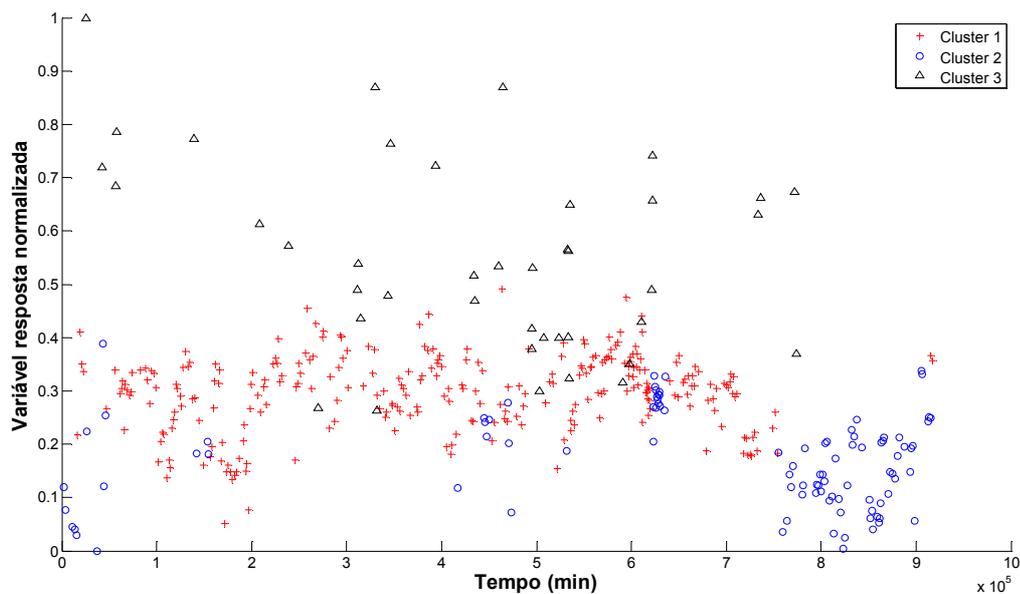


Figura 4.20: Clusters formados considerando-se as variáveis PFE, diferencial de pressão topo/panela de LCO e temperatura de reação no vetor das k-médias.

Tabela 4.16: Terceira etapa de inclusão de variáveis no vetor de k-médias, para três clusters a priori.

Variável	Inclusão
Temperatura de topo da fracionadora principal	SIM
Temperatura do refluxo de topo	NÃO
Temperatura no topo do riser	SIM
Diferencial de pressão entre topo/panela LCO	SIM
PFE	SIM

Tabela 4.17: Quarta etapa de inclusão de variáveis no vetor de k-médias, para três clusters a priori.

Variável	Inclusão
Temperatura de topo da fracionadora principal	SIM
Temperatura do refluxo de topo	SIM
Temperatura no topo do riser	SIM
Diferencial de pressão entre topo/panela LCO	SIM
PFE	SIM

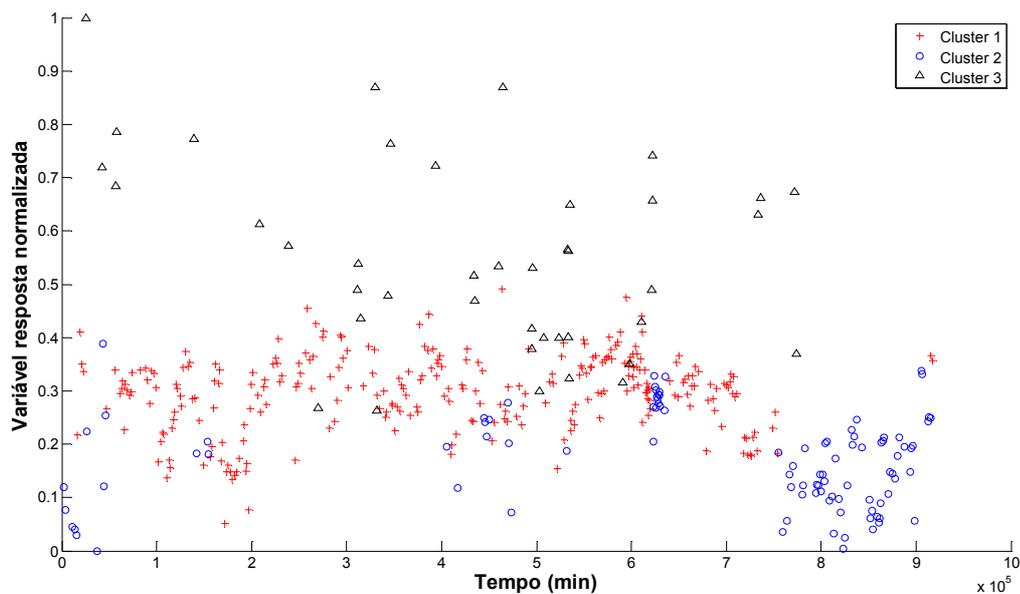


Figura 4.21: Clusters formados considerando-se as variáveis PFE, diferencial de pressão topo/panela de LCO, temperatura de reação e temperatura de topo no vetor das k-médias.

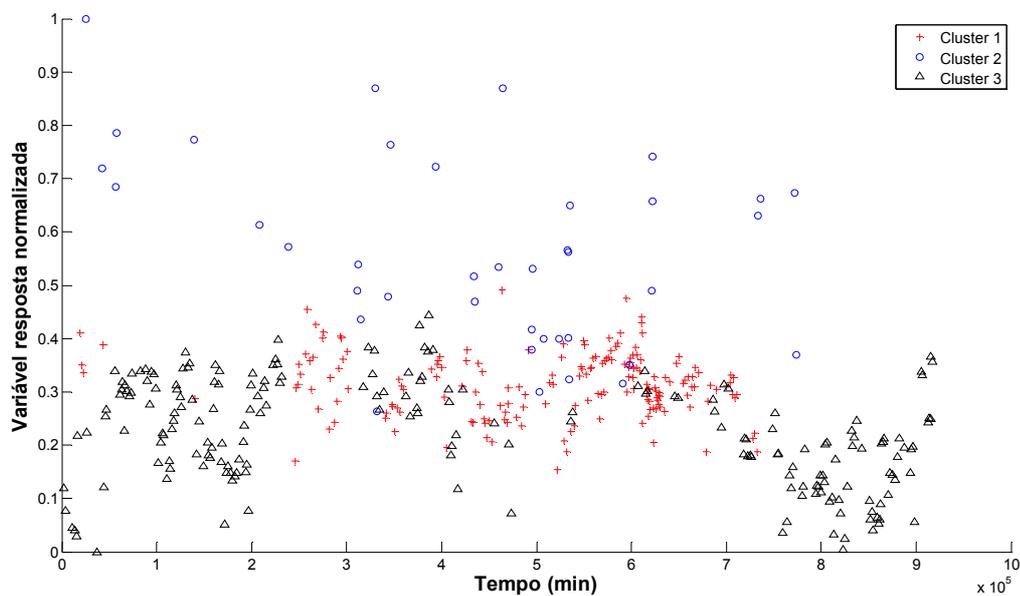


Figura 4.22: Clusters formados considerando-se as variáveis PFE, diferencial de pressão topo/panela de LCO, temperatura de reação, temperatura de topo e temperatura do refluxo de topo no vetor das k-médias.

Neste caso, a melhor partição natural dos dados encontrada é exibida na Figura 4.21, embora a mesma não seja suficientemente clara. Ela resulta na obtenção dos centroides:

Tabela 4.18: Centroides dos clusters referentes à Figura 4.21.

Variável	Centroide		
	Cluster 1	Cluster 2	Cluster 3
Temperatura de topo da fracionadora principal	0,8973	0,8726	0,8766
Temperatura no topo do riser	0,8426	0,6371	0,7793
Diferencial de pressão entre topo/panela LCO	0,3023	0,3682	0,7743
PFE	0,2981	0,1715	0,5515

O cluster 1 caracteriza uma operação com maximização do PFE da nafta ‘leve’ craqueada, considerando-se as especificações de qualidade vigentes. Neste caso, o PFE apresenta um valor intermediário quando comparado aos outros dois centroides. Tanto a temperatura de topo da torre fracionadora quanto a temperatura de reação são mantidas em um patamar um pouco mais elevado, o que caracteriza maior produção de nafta ‘leve’ craqueada.

O cluster 2 caracteriza também uma condição de operação normal, em que a severidade do conversor é um pouco reduzida para favorecer a produção de médios (i.e nafta ‘pesada’ e LCO). Neste caso, o PFE da nafta ‘leve’ craqueada é um pouco reduzido, porém mantendo-se especificado dentro dos requisitos de qualidade.

Já no caso do cluster 3, a elevação do diferencial de pressão parece caracterizar a elevação do PFE, em função dos fenômenos de inundação anteriormente descritos. Este fenômeno favorece a perda de especificação do PFE.

A análise de cluster permite detectar três influências principais na especificação do PFE. A condição de operação do conversor (i.e temperatura de reação) influencia nos rendimentos das retiradas da torre e, conseqüentemente, no PFE da retirada superior. A condição termodinâmica do topo da torre (i.e temperatura no topo da torre) influencia no equilíbrio de condensação de hidrocarbonetos. Já o diferencial de pressão traduz as condições hidráulicas de escoamento de líquidos e gases no topo da torre, permitindo mapear a possibilidade de perda de especificação do PFE.

4.3 Modelagem matemática de sistemas

Após a análise de clusters, optou-se pela manutenção das variáveis temperatura de topo (variável 3), temperatura do refluxo de topo (variável 4), temperatura do topo do *riser*

(variável 21) e diferencial de pressão topo/panela de LCO (variável 23) como variáveis regressoras, dado que as mesmas são sempre selecionadas pelo método de regressão linear por etapas, independentemente dos dados sorteados. A forte correlação e, portanto, redundância de informação entre as temperaturas de topo e do refluxo de topo não tornaria necessária a manutenção de ambas no modelo. Contudo, com intuito de verificar o poder de filtragem e tratamento de colinearidade das técnicas lineares de PCR e PLS frente à técnica de MLR, ambas as variáveis foram mantidas.

4.3.1 Seleção de dados em regime estacionário

A aproximação do histograma da distribuição W pela distribuição Gamma demonstrou coerência quanto aos valores das médias d_2 obtidas, conforme exibido na Tabela 4.19.

Tabela 4.19: Valores obtidos para a média d_2 da distribuição W .

Variável	d_2
Temperatura de topo da fracionadora principal	2,5529
Temperatura do refluxo de topo	2,6115
Temperatura no topo do riser	2,6219
Diferencial de pressão entre topo/panela LCO	2,6000
MÉDIA	2,5966

Tabela 4.20: Resumo dos fatores para a construção dos gráficos de controle para amplitude.

n	d_2	$\left(1 - \frac{3d_3}{d_2}\right)$	$\left(1 + \frac{3d_3}{d_2}\right)$
6	2,5340	0,0000	2,0040
7	2,7040	0,0760	1,9240

(MONTGOMERY, 2005, p.425)

Como o intervalo de tempo estabelecido para a captura das médias foi de 30 minutos, com frequência de coleta a cada 5 minutos, cada janela temporal apresentou no máximo sete pontos de amostragem. A utilização de filtros para exclusão de *outliers* ou mesmo exclusão de dados com perda de sinal ocasiona menor disponibilidade de informação em algumas janelas de captura. Como o valor médio obtido para d_2 foi de 2,5966, intermediário entre os valores referentes à seis e sete pontos de amostragem, acredita-se

que a aproximação pela função Gamma é válida. No entanto, a estimativa do desvio padrão d_3 da distribuição Gamma diferiu dos valores recomendados, dado que:

Tabela 4.21: Resumo dos fatores para a determinação do desvio padrão d_3 .

Variável	$\left(1 + \frac{3d_3}{d_2}\right)$
Temperatura de topo da fracionadora principal	7,6454
Temperatura do refluxo de topo	9,2698
Temperatura no topo do riser	2,7293
Diferencial de pressão entre topo/panela LCO	3,1616
MÉDIA	5,7015

De acordo com a Tabela 4.20, o valor esperado para o termo $\left(1 + \frac{3d_3}{d_2}\right)$ estaria entre 1,924 e 2,004. No entanto, a estimativa do desvio padrão d_3 da distribuição W foi superior ao referido na literatura. Desta forma, o limite superior de controle (LSC) obtido pelo programa para classificação dos dados apresentou um ponto de corte superior. Com isso, dados os quais poderiam ser identificados como característicos de períodos ‘transientes’ pelo valor referido em literatura (i.e janela de dados com variância superior àquela verificada comumente na massa de dados) foram incorporados ao conjunto de dados supostamente ‘estacionários’ (i.e janela de dados com variância equivalente àquela comumente verificada nos dados). Este fato pode estar ligado à qualidade do ajuste dos parâmetros da função Gamma pelo software, em especial para valores de W próximos de zero (i.e amplitude nula no intervalo). Além disso, cabe ressaltar a perda de sensibilidade para as pequenas variações dos dados de processo, em função dos procedimentos de exceção e compressão no PI, o que poderia comprometer a precisão do dado fornecido ao ajuste dos parâmetros.

No entanto, o programa INFERATOR possibilita que o usuário ajuste o critério de classificação dos dados de regime estacionário, variando-se o parâmetro k da equação (4.1), tornando o limite de corte mais ou menos restritivo:

$$LSC = \left(1 + \frac{kd_3}{d_2}\right) \bar{R} \quad (4.1)$$

Neste trabalho, o parâmetro k adotado foi tal que $k = 3$, para permitir comparação com os dados de literatura presentes na Tabela 4.20. Apesar da maior nota de corte obtida para o *LSC*, a filtragem dos dados estacionários eliminou 6,46% das amostras disponíveis na matriz de dados \mathbf{X} . Os histogramas obtidos para o cálculo do *LSC* para cada uma das variáveis são exibidos a seguir:

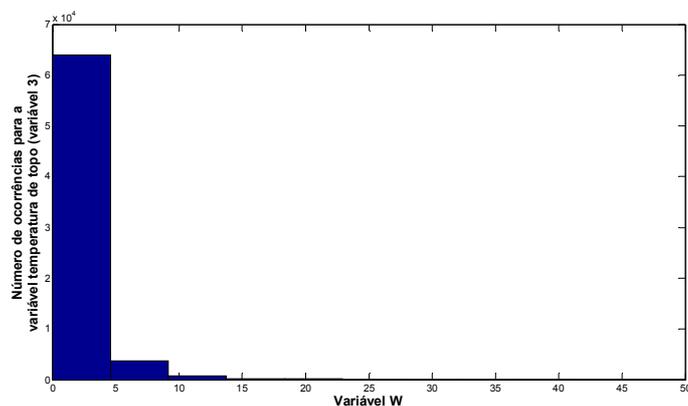


Figura 4.23: Histograma utilizado para o cálculo do *LSC* da variável temperatura de topo da torre fracionadora principal (variável 3).

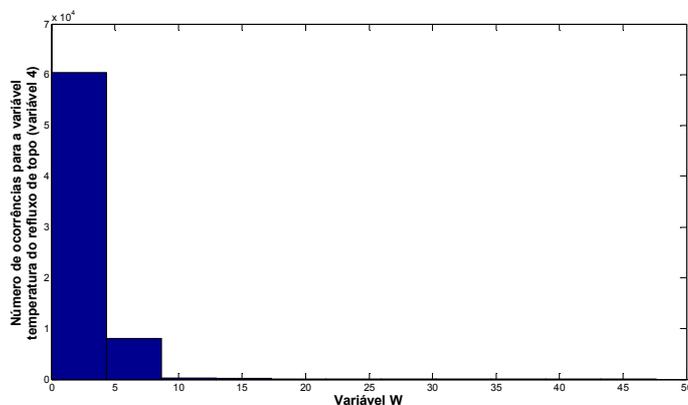


Figura 4.24: Histograma utilizado para o cálculo do *LSC* da variável temperatura do refluxo de topo para a torre fracionadora principal (variável 4).

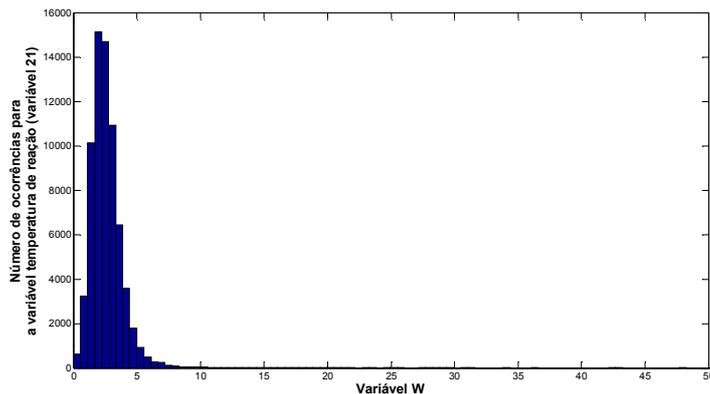


Figura 4.25: Histograma utilizado para o cálculo do *LSC* da variável temperatura de reação (variável 21).

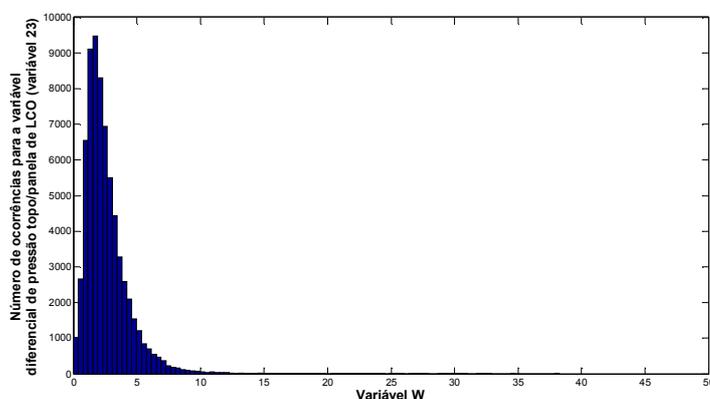


Figura 4.26: Histograma utilizado para o cálculo do *LSC* da variável diferencial de pressão topo/panela de LCO (variável 23).

4.3.2 Seleção da arquitetura dos modelos

As técnicas de modelagem foram aplicadas de maneira exploratória, para a verificação da arquitetura ótima em cada uma delas, considerando-se dois conjuntos de dados distintos, sorteados aleatoriamente: dados ditos ‘transientes’ e ‘estacionários’. Os dados ditos ‘transientes’ não foram filtrados para obtenção exclusiva dos dados característicos de regime estacionário. As simulações foram repetidas trinta vezes para cada caso estudado (i.e tipo de arquitetura testada), conforme Figura 3.10.

Foram obtidas as médias das variáveis ‘incerteza’ e correlação, bem como o somatório do erro quadrático e da derivada do erro quadrático em relação ao tempo. Estes

parâmetros foram considerados como parâmetros de desempenho para escolha da arquitetura dos modelos.

Ressalta-se que todas as análises expressas nos gráficos a seguir foram conduzidas considerando-se os dados de teste.

4.3.2.1 Regressão de componentes principais (PCR)

Buscou-se, em cada simulação, aumentar a variabilidade explicada pelas componentes principais, de modo a verificar a influência do acréscimo de componentes na sua capacidade de predição:

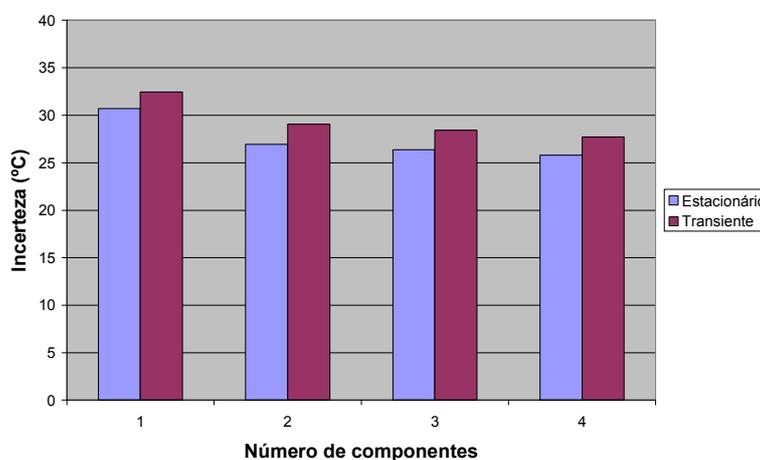


Figura 4.27: Efeito do acréscimo de componentes principais na variável 'incerteza'.

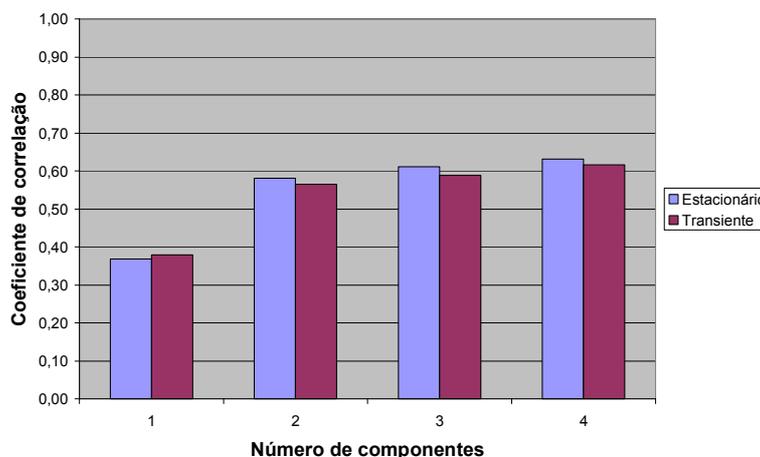


Figura 4.28: Efeito do acréscimo de componentes principais no coeficiente de correlação entre os dados reais e os dados preditos.

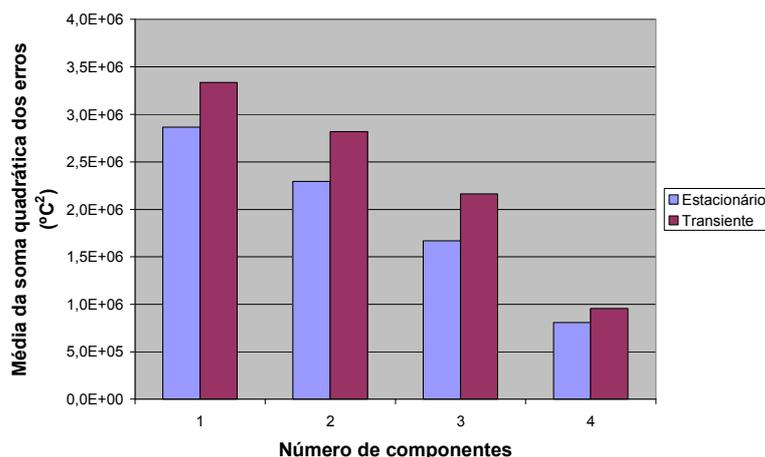


Figura 4.29: Efeito do acréscimo de componentes principais na soma quadrática dos erros.

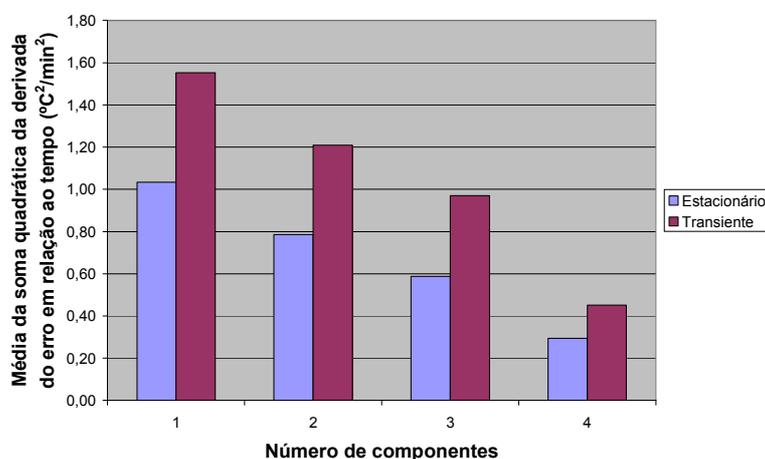


Figura 4.30: Efeito do acréscimo de componentes principais na soma quadrática da derivada do erro em relação ao tempo.

Pode-se perceber que a inclusão de quatro componentes principais proporciona os melhores resultados: a menor incerteza, a menor soma quadrática dos erros e a menor soma quadrática da derivada do erro em relação ao tempo. Este último indicador favorece a melhor adequação da tendência do modelo. Além disso, observa-se também o maior coeficiente de correlação no caso do uso de quatro componentes principais. No entanto, a escolha de todas as componentes reduz a potencialidade de filtragem de ruído, tornando o desempenho do método similar à regressão linear multivariada. Mesmo assim, optou-se pelo uso de quatro componentes principais.

4.3.2.2 Mínimos quadrados parciais (PLS)

Verificou-se o efeito do aumento do número de variáveis latentes na capacidade de predição do modelo:

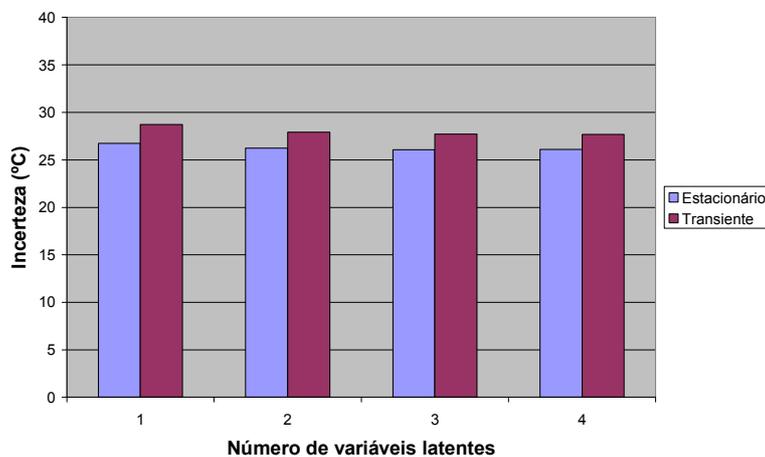


Figura 4.31: Efeito do acréscimo de variáveis latentes na variável ‘incerteza’.

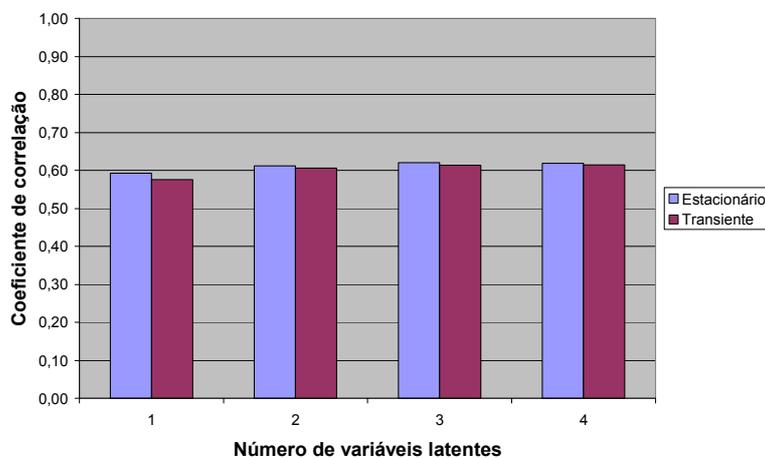


Figura 4.32: Efeito do acréscimo de variáveis latentes no coeficiente de correlação entre os dados reais e os dados preditos.

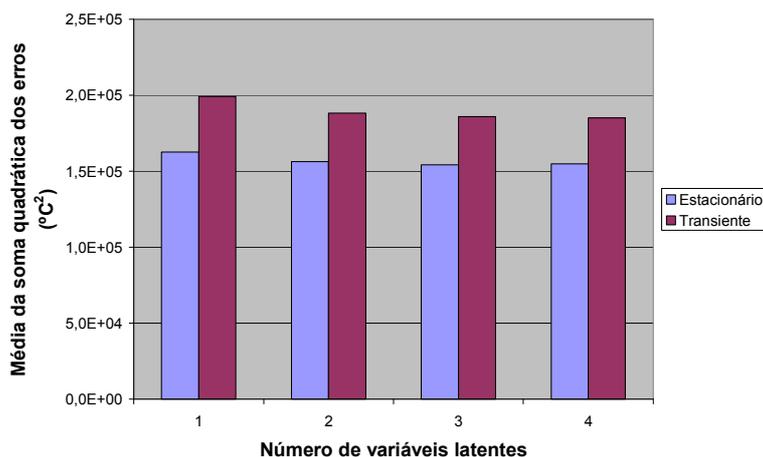


Figura 4.33: Efeito do acréscimo de variáveis latentes na soma quadrática dos erros.

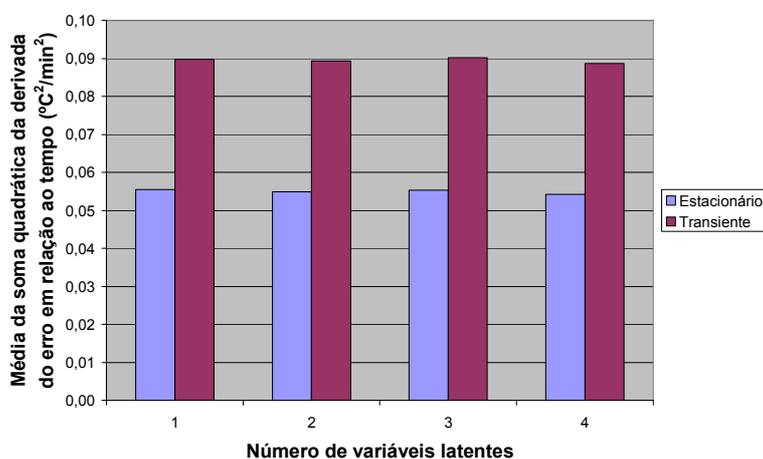


Figura 4.34: Efeito do acréscimo de variáveis latentes na soma quadrática da derivada do erro em relação ao tempo.

Dentre os parâmetros destacados para avaliação, as respostas se mostraram semelhantes. Apenas a variável ‘Média da soma quadrática dos erros’ apresentou diminuição sensível com o uso de duas variáveis latentes. Sendo assim, adotou-se a utilização de duas variáveis latentes.

4.3.2.3 Regularização Bayesiana combinada com algoritmo de treinamento Levenberg-Marquardt

Verificou-se o efeito do acréscimo de neurônios na camada oculta e sua influência na capacidade de predição do modelo:

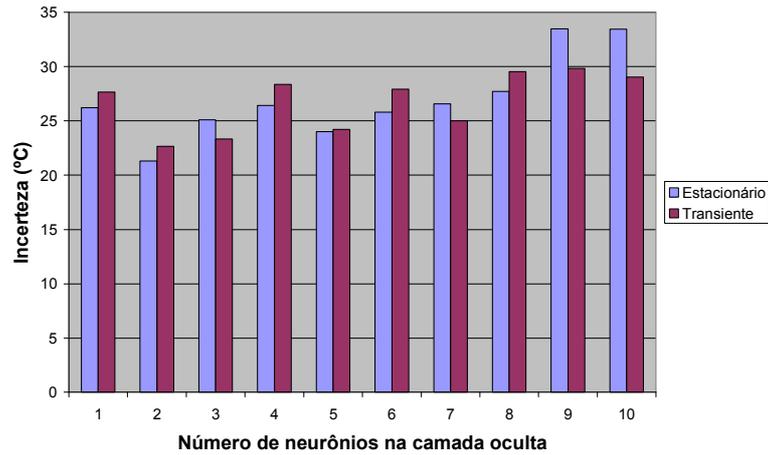


Figura 4.35: Efeito do acréscimo de neurônios à camada oculta na variável ‘incerteza’ para o caso de função de performance dada por regularização Bayesiana.

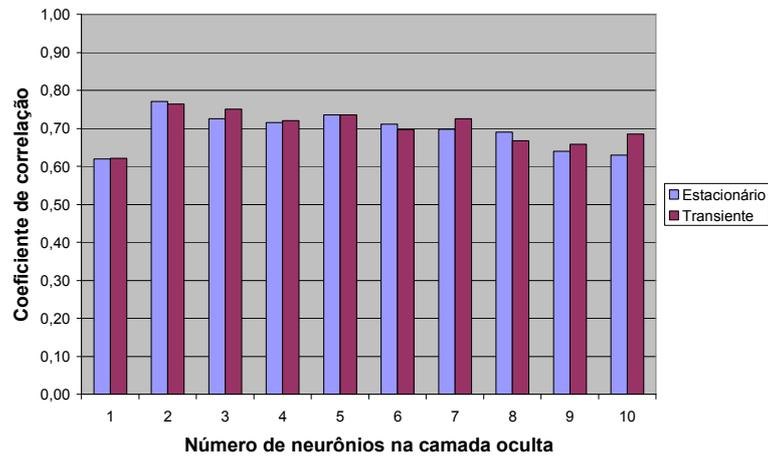


Figura 4.36: Efeito do acréscimo de neurônios à camada oculta no coeficiente de correlação entre os dados reais e os dados preditos para o caso de função de performance dada por regularização Bayesiana.

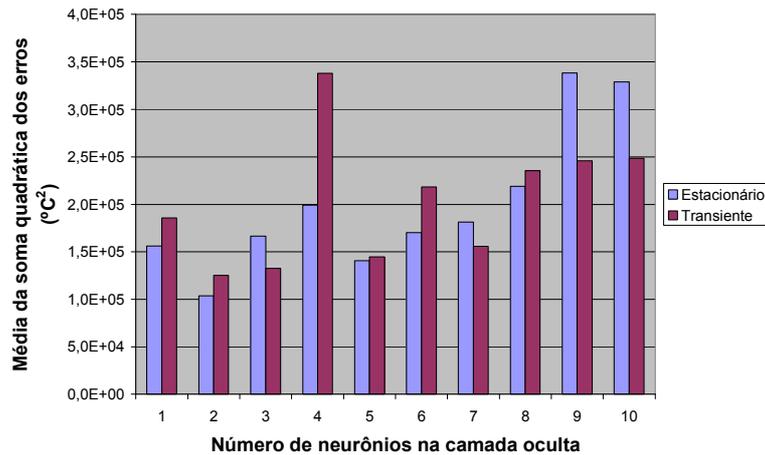


Figura 4.37: Efeito do acréscimo de neurônios à camada oculta na soma quadrática dos erros para o caso de função de performance dada por regularização Bayesiana.

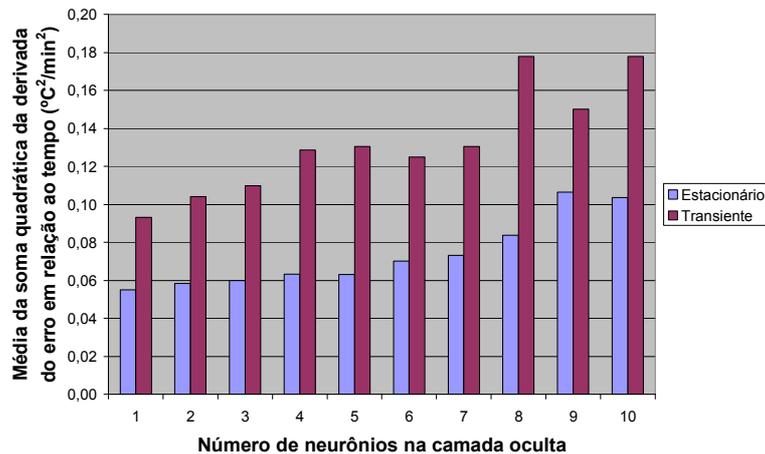


Figura 4.38: Efeito do acréscimo de neurônios à camada oculta na soma quadrática da derivada do erro em relação ao tempo para o caso de função de performance dada por regularização Bayesiana.

Os parâmetros de adequação indicam que o uso de apenas dois neurônios é suficiente para minimizar a soma quadrática de erros, bem como minimizar a incerteza e maximizar o coeficiente de correlação. Com o intuito de verificar o comportamento do número de parâmetros efetivos γ , foi realizada uma simulação adicional até o limite máximo de vinte neurônios. O comportamento da variável “Número de parâmetros efetivos” é exibida a seguir:

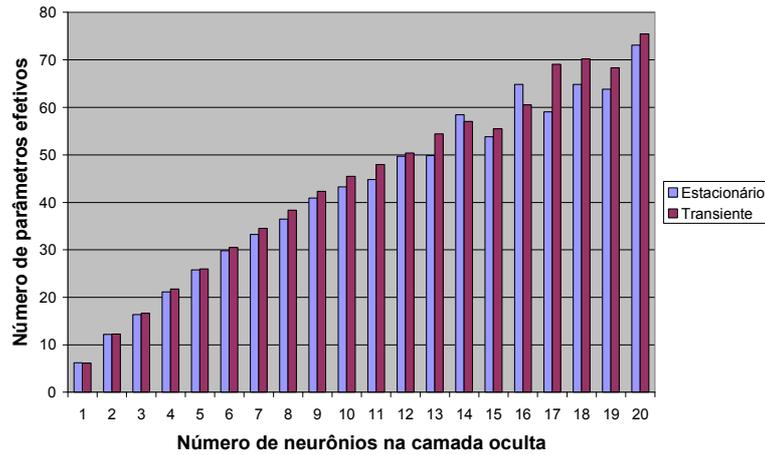


Figura 4.39: Efeito do acréscimo de neurônios à camada oculta na variável número de parâmetros efetivos.

Como pode ser observado na Figura 4.39, não foi verificado um patamar para o número de parâmetros efetivos. Este seria um indicativo de que a adição de novos neurônios não implicaria mais na redução do termo do erro quadrático. No entanto, verifica-se uma pequena tendência de redução do número de parâmetros efetivos para mais de dezesseis neurônios. O gráfico obtido para a variável ‘incerteza’, no caso da simulação com vinte neurônios também sugere, tal como o gráfico da Figura 4.35, o uso de apenas dois nós:

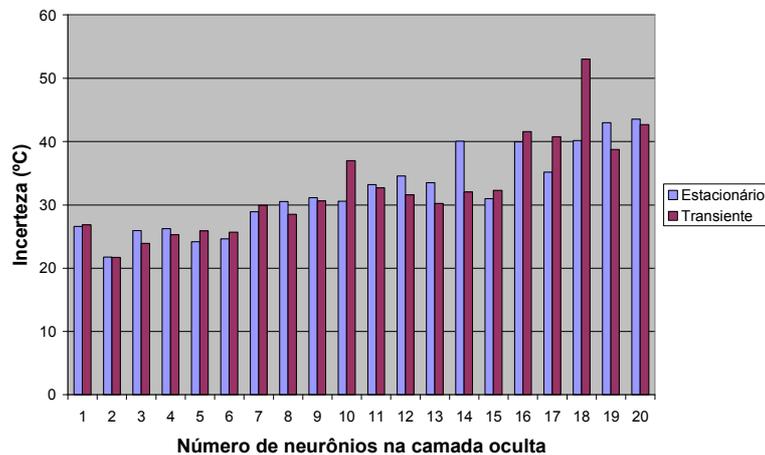


Figura 4.40: Efeito do acréscimo de neurônios à camada oculta na variável ‘incerteza’, utilizando-se um máximo de vinte nós para o caso de função de performance dada por regularização Bayesiana.

4.3.2.4 Algoritmo *back-propagation* resiliente (Rprop)

4.3.2.4.1 Utilização de uma camada oculta

Novamente, foram acrescentados neurônios à camada oculta, de modo a verificar o seu efeito na capacidade de predição do modelo:

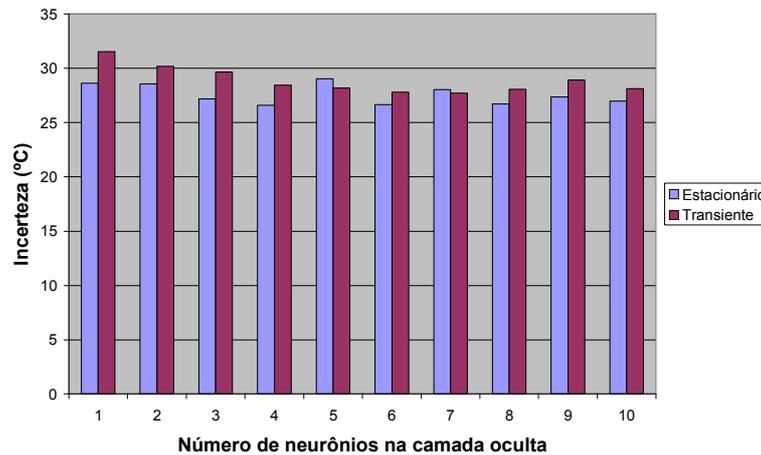


Figura 4.41: Efeito do acréscimo de neurônios à camada oculta na variável ‘incerteza’ para o caso do algoritmo de treinamento *back-propagation* resiliente.

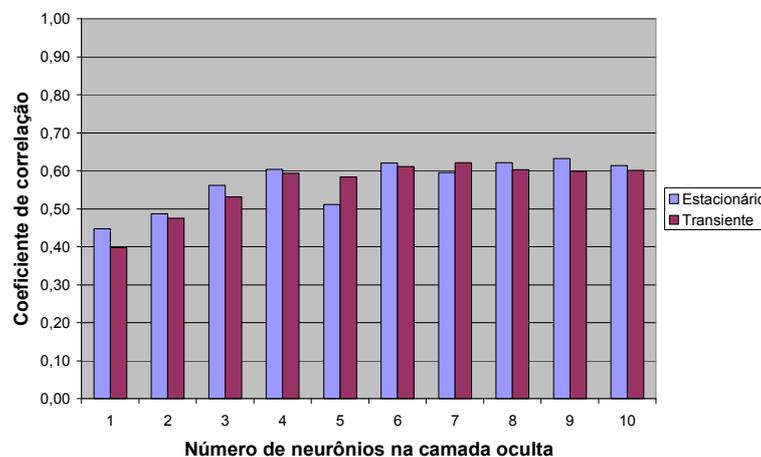


Figura 4.42: Efeito do acréscimo de neurônios à camada oculta no coeficiente de correlação para o caso do algoritmo de treinamento *back-propagation* resiliente.

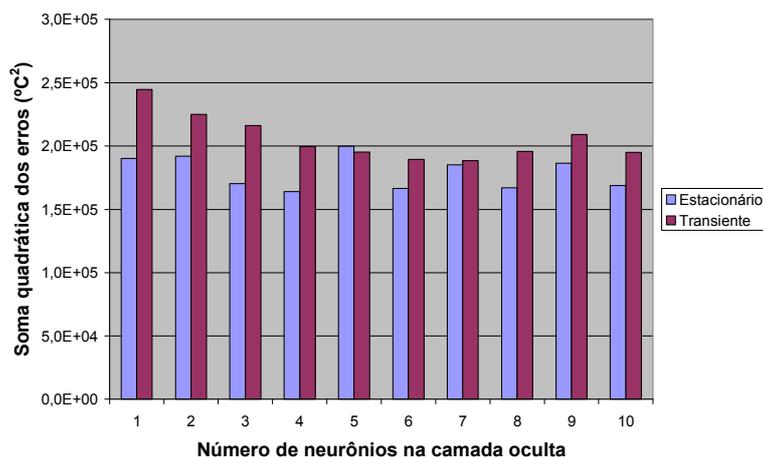


Figura 4.43: Efeito do acréscimo de neurônios à camada oculta na soma quadrática dos erros para o caso do algoritmo de treinamento back-propagation resiliente.

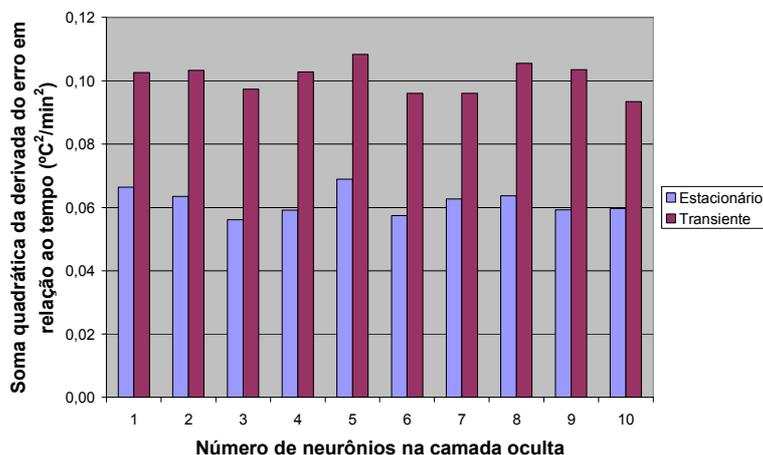


Figura 4.44: Efeito do acréscimo de neurônios à camada oculta na soma quadrática da derivada do erro em relação ao tempo para o caso do algoritmo back-propagation resiliente.

A utilização de quatro neurônios na camada oculta implica na obtenção da menor incerteza, maior correlação e menor soma quadrática dos erros. Além disso, o parâmetro indicativo da tendência (i.e soma quadrática da derivada do erro em relação ao tempo) é satisfatória para esta arquitetura.

4.3.2.4.2 Utilização de duas camadas ocultas

Neste caso, variou-se o número de neurônios na segunda camada oculta, para um valor fixo de neurônios na primeira camada. Após alcance do número máximo de neurônios na segunda camada, foi incrementado um neurônio na primeira camada. Este procedimento é exemplificado pela Figura 3.10. Utilizou-se um máximo de quatro neurônios em cada camada. Os resultados são exibidos a seguir, segregados por dados referentes a estado ‘transiente’ e ‘estacionário’. Para os dados ditos ‘transientes’, os seguintes resultados foram obtidos:

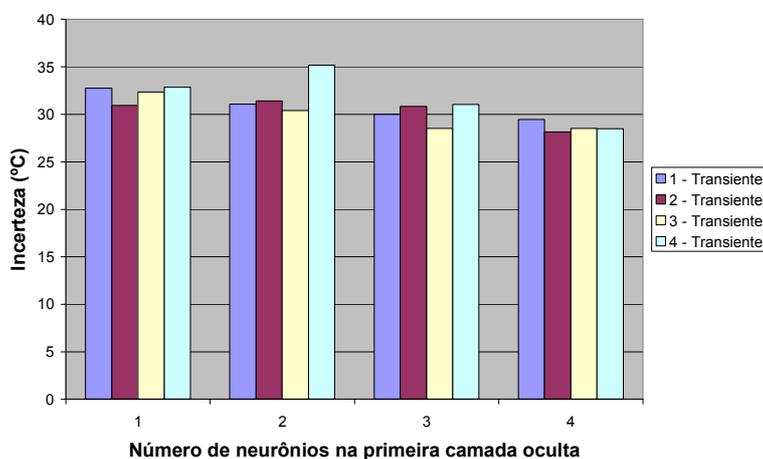


Figura 4.45: Efeito do acréscimo de neurônios à segunda camada, conforme legenda, e à primeira camada, conforme abscissa, na variável ‘incerteza’, para dados caracterizados como ‘transientes’.

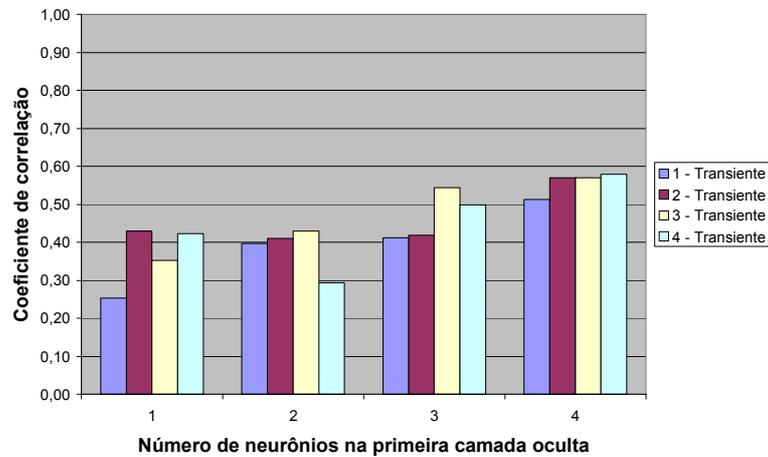


Figura 4.46: Efeito do acréscimo de neurônios à segunda camada, conforme legenda, e à primeira camada, conforme abscissa, na variável coeficiente de correlação, para dados caracterizados como ‘transientes’.

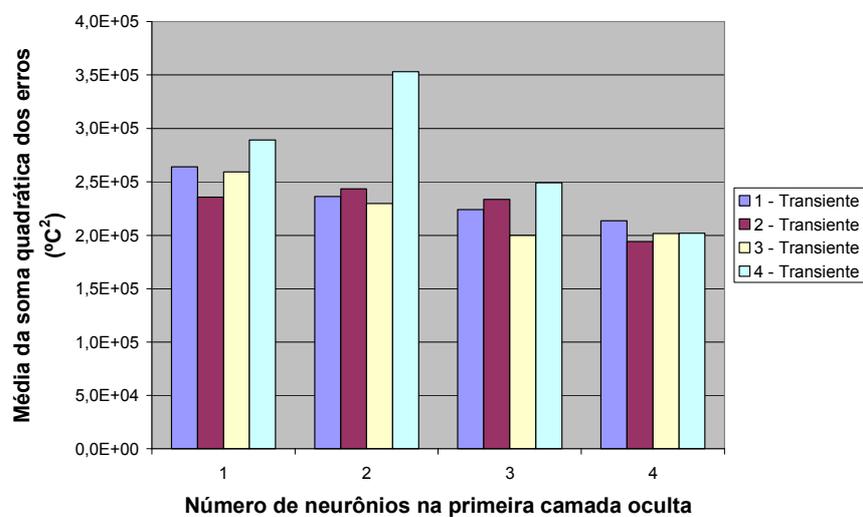


Figura 4.47: Efeito do acréscimo de neurônios à segunda camada, conforme legenda, e à primeira camada, conforme abscissa, na variável média da soma quadrática dos erros, para dados caracterizados como ‘transientes’.

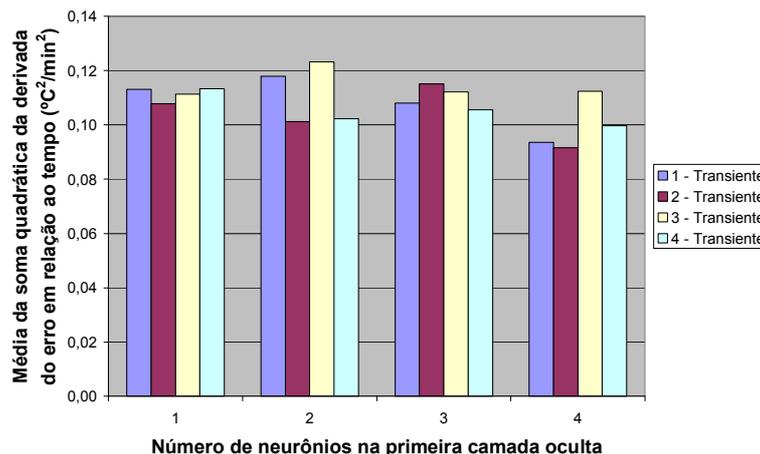


Figura 4.48: Efeito do acréscimo de neurônios à segunda camada, conforme legenda, e à primeira camada, conforme abscissa, na variável média da soma quadrática da derivada do erro em relação ao tempo, para dados caracterizados como ‘transientes’.

A análise dos dados ‘transientes’ sugere a utilização de quatro neurônios na primeira camada e dois neurônios na segunda camada. Esta configuração fornece uma menor média para a ‘incerteza’ e, conseqüentemente, a menor soma quadrática dos erros. Além disso, fornece também o maior coeficiente de correlação.

Resultados semelhantes são observados para o caso dos dados ditos ‘estacionários’:

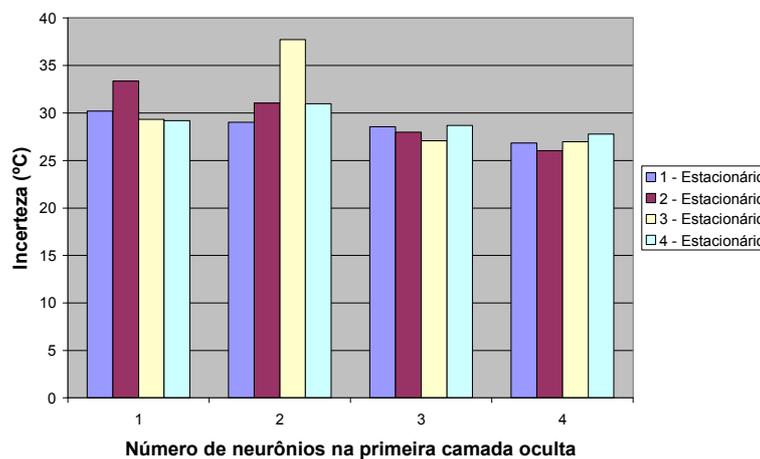


Figura 4.49: Efeito do acréscimo de neurônios à segunda camada, conforme legenda, e à primeira camada, conforme abscissa, na variável ‘incerteza’, para dados caracterizados como ‘estacionários’.

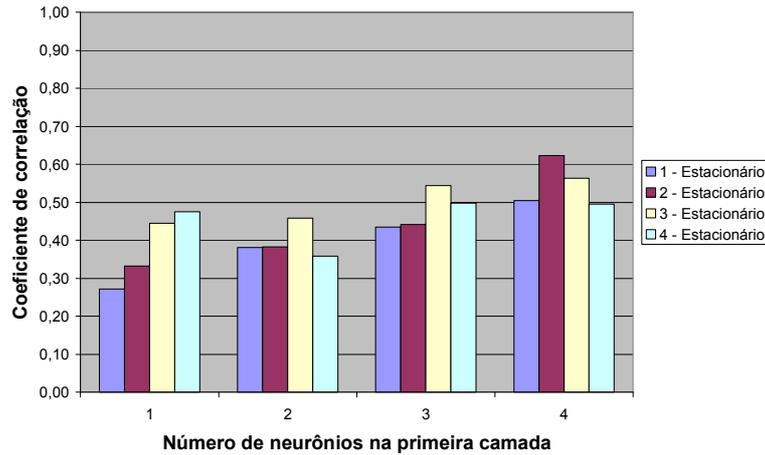


Figura 4.50: Efeito do acréscimo de neurônios à segunda camada, conforme legenda, e à primeira camada, conforme abscissa, na variável coeficiente de correlação, para dados caracterizados como ‘estacionários’.

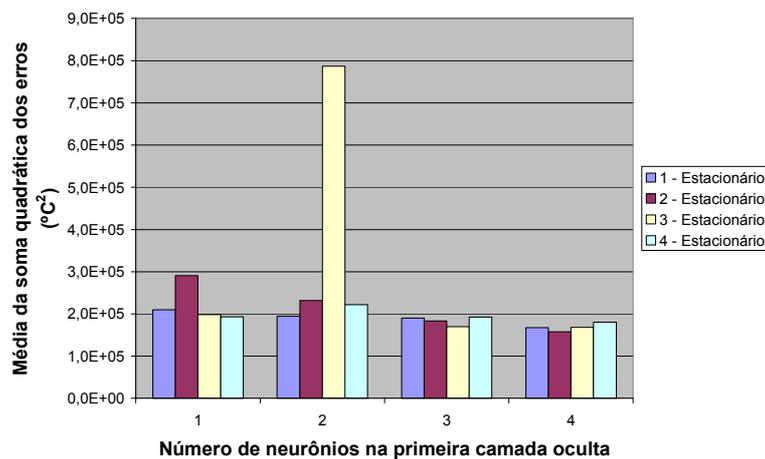


Figura 4.51: Efeito do acréscimo de neurônios à segunda camada, conforme legenda, e à primeira camada, conforme abscissa, na variável média da soma quadrática dos erros, para dados caracterizados como ‘estacionários’.

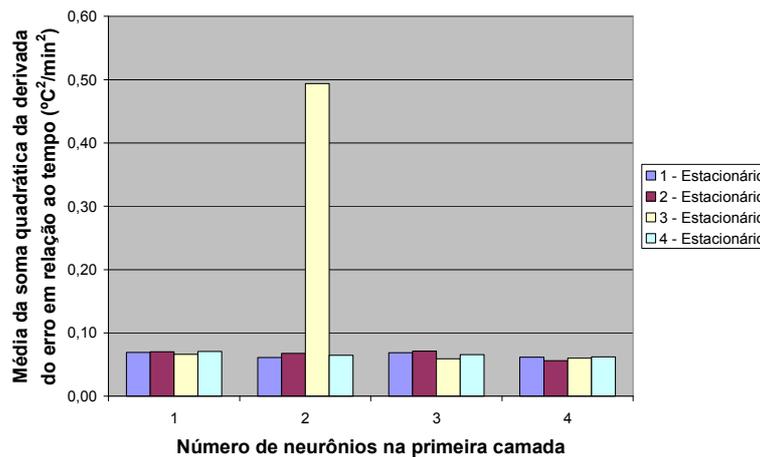


Figura 4.52: Efeito do acréscimo de neurônios à segunda camada, conforme legenda, e à primeira camada, conforme abscissa, na variável média da soma quadrática da derivada do erro em relação ao tempo, para dados caracterizados como ‘estacionários’.

4.3.2.5 Redes neurais de base radial

Neste caso, variou-se gradativamente o número de neurônios na camada oculta, para um valor fixo do parâmetro de espalhamento. Após o alcance do número máximo de neurônios, foi testado um novo valor para o parâmetro de espalhamento, dentre um total de três possibilidades. Novamente, este procedimento é exemplificado pela Figura 3.10. Utilizou-se um máximo de dez neurônios na camada oculta. Os resultados são exibidos a seguir, divididos entre dados referentes a estado ‘transiente’ e ‘estacionário’. Para os dados ditos ‘transientes’, os seguintes resultados foram obtidos:

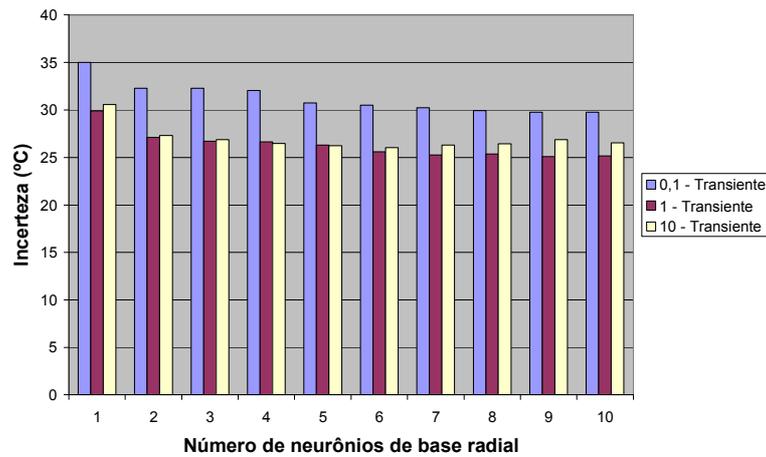


Figura 4.53: Efeito da modificação do parâmetro de espalhamento, conforme legenda, e do incremento do número de nós na camada oculta, conforme abscissa, na variável ‘incerteza’, para dados caracterizados como ‘transientes’.

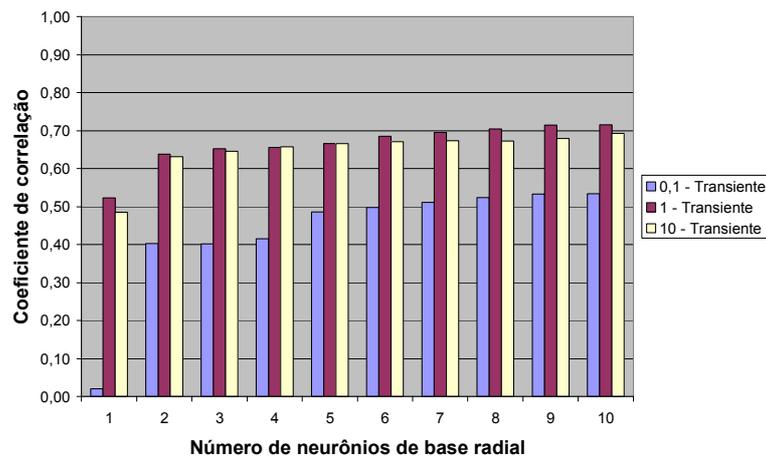


Figura 4.54: Efeito da modificação do parâmetro de espalhamento, conforme legenda, e do incremento do número de nós na camada oculta, conforme abscissa, na variável coeficiente de correlação, para dados caracterizados como ‘transientes’.

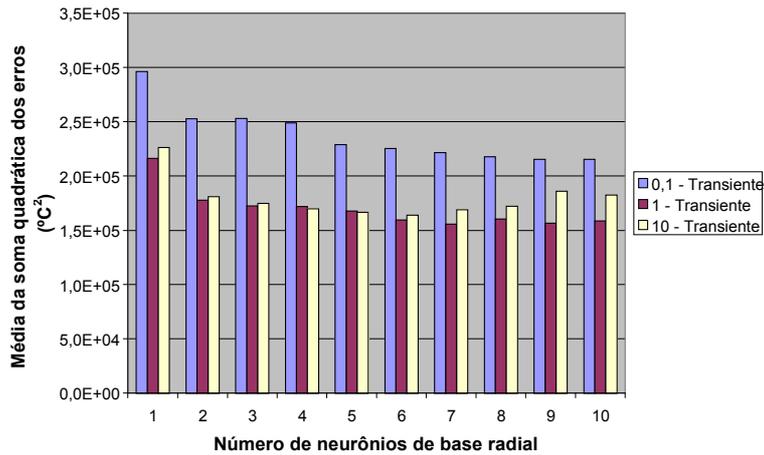


Figura 4.55: Efeito da modificação do parâmetro de espalhamento, conforme legenda, e do incremento do número de nós na camada oculta, conforme abscissa, na variável média da soma quadrática dos erros, para dados caracterizados como ‘transientes’.

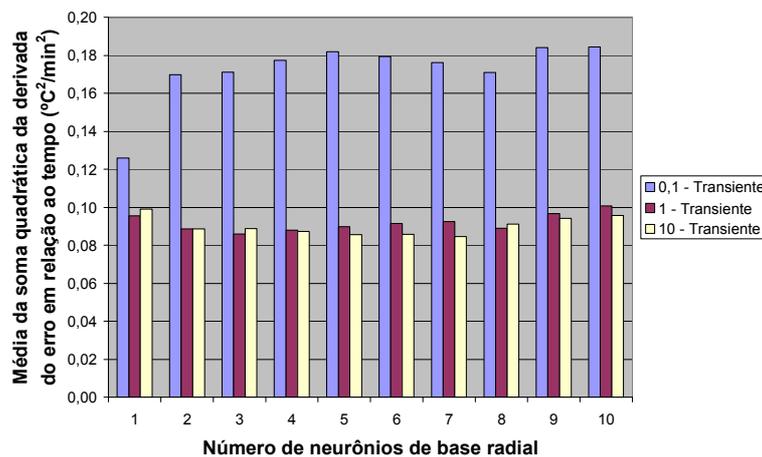


Figura 4.56: Efeito da modificação do parâmetro de espalhamento, conforme legenda, e do incremento do número de nós na camada oculta, conforme abscissa, na variável média da soma quadrática da derivada do erro em relação ao tempo, para dados caracterizados como ‘transientes’.

Resultados semelhantes foram obtidos para o caso dos dados caracterizados como ‘estacionários’, sendo exibidos a seguir:

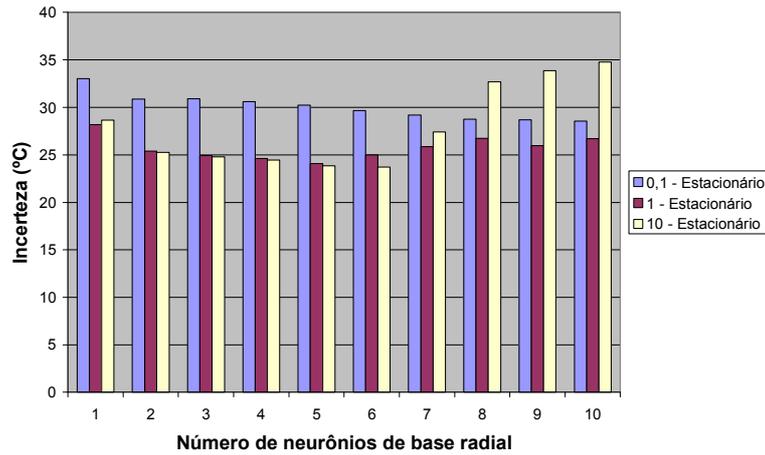


Figura 4.57: Efeito da modificação do parâmetro de espalhamento, conforme legenda, e do incremento do número de nós na camada oculta, conforme abscissa, na variável ‘incerteza’, para dados caracterizados como ‘estacionários’.

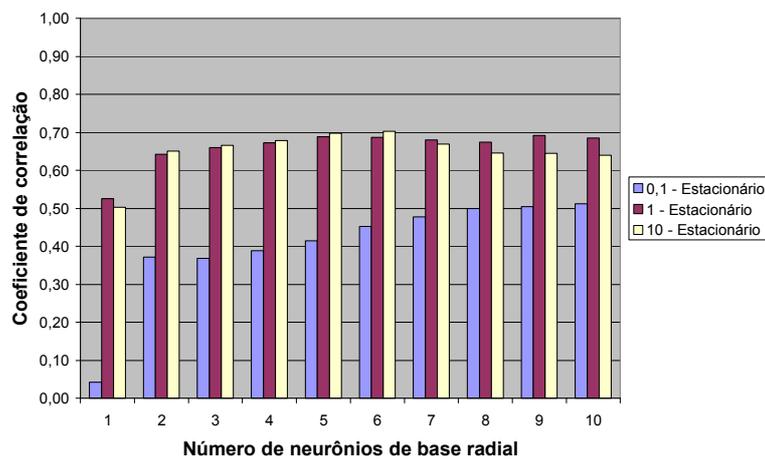


Figura 4.58: Efeito da modificação do parâmetro de espalhamento, conforme legenda, e do incremento do número de nós na camada oculta, conforme abscissa, na variável coeficiente de correlação, para dados caracterizados como ‘estacionários’.

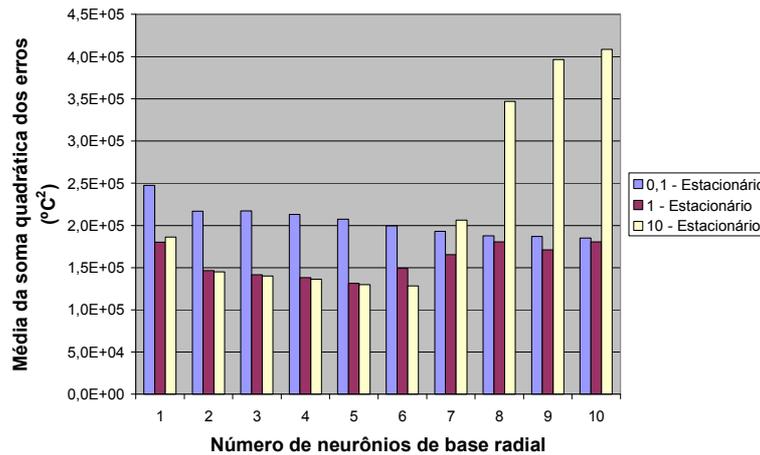


Figura 4.59: Efeito da modificação do parâmetro de espalhamento, conforme legenda, e do incremento do número de nós na camada oculta, conforme abscissa, na variável média da soma quadrática dos erros, para dados caracterizados como ‘estacionários’.

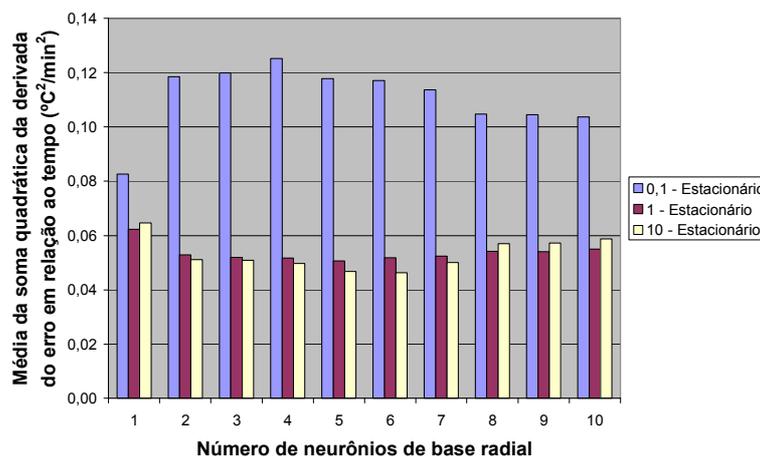


Figura 4.60: Efeito da modificação do parâmetro de espalhamento, conforme legenda, e do incremento do número de nós na camada oculta, conforme abscissa, na variável média da soma quadrática da derivada do erro em relação ao tempo, para dados caracterizados como ‘estacionários’.

Adotou-se a utilização de seis neurônios de base radial, com parâmetro de espalhamento igual a dez. Esta arquitetura forneceu a menor média da soma quadrática do erro e, conseqüentemente, a menor ‘incerteza’, além do maior coeficiente de correlação.

4.3.3 Comparação das técnicas de modelagem

As arquiteturas escolhidas com base nas análises exploratórias anteriores foram novamente simuladas, com o intuito de se comparar o desempenho de cada uma das técnicas em sua melhor configuração. Novamente, os parâmetros utilizados para avaliação de desempenho foram: variável ‘incerteza’, coeficiente de correlação, soma quadrática dos erros e soma quadrática da derivada dos erros em relação ao tempo. Os gráficos a seguir resumem os principais resultados:

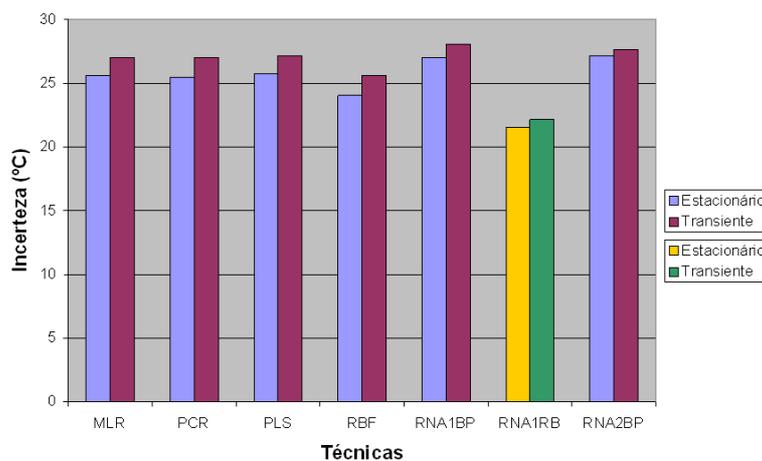


Figura 4.61: Comparação da variável ‘incerteza’ para as diferentes técnicas de modelagem.

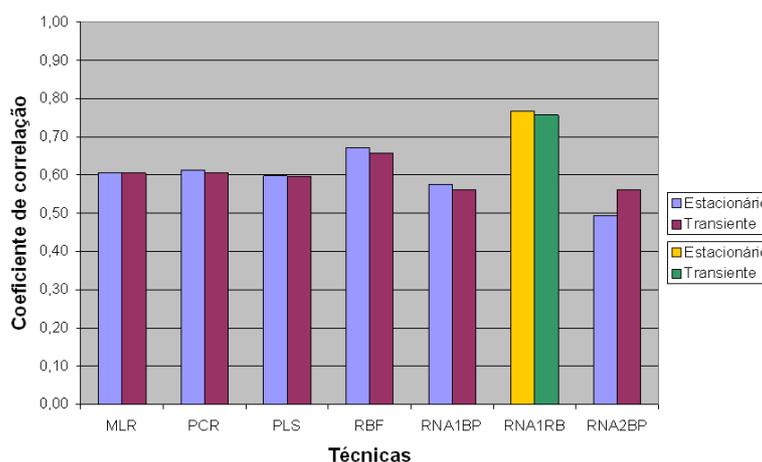


Figura 4.62: Comparação da variável ‘coeficiente de correlação’ para as diferentes técnicas de modelagem.

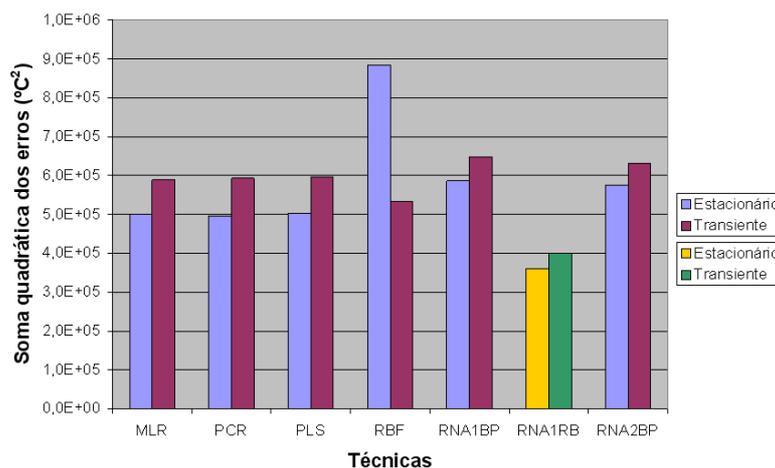


Figura 4.63: Comparação da variável ‘soma quadrática dos erros’ para as diferentes técnicas de modelagem.

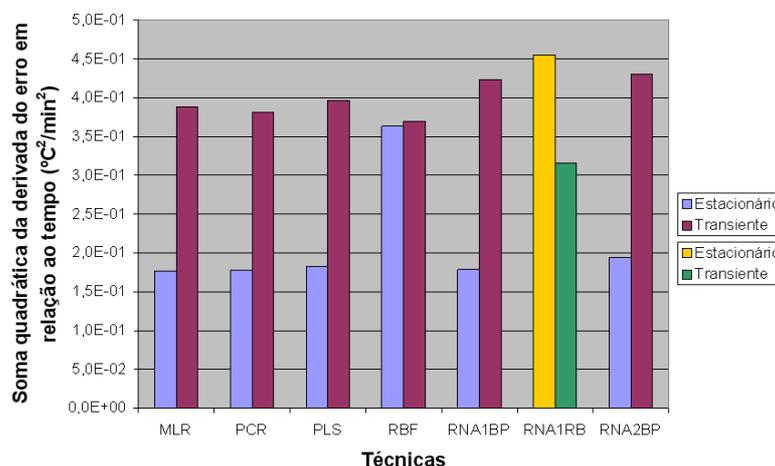


Figura 4.64: Comparação da variável ‘soma quadrática da derivada do erro em relação ao tempo’ para as diferentes técnicas de modelagem.

Os gráficos exibidos a seguir foram gerados considerando-se cada uma das técnicas e se referem ao modelo obtido que melhor atendeu as métricas de desempenho propostas: mínima soma do erro quadrático ou mínima soma da derivada do erro quadrático em relação ao tempo. Ressalta-se que os modelos encontrados não são necessariamente os melhores possíveis para cada arquitetura, e sim os melhores modelos encontrados tendo em vista os dados sorteados e a premissa do menor uso de parâmetros.

Os gráficos seguintes se referem aos melhores modelos obtidos tendo em vista a minimização do erro quadrático:

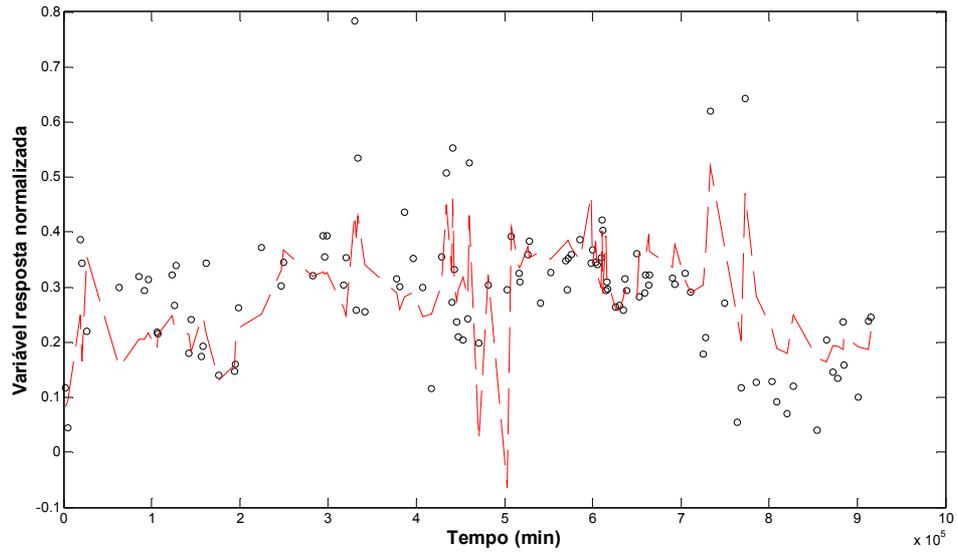


Figura 4.65: Modelo linear (dados estacionários) com correlação igual a 0.64547 e variável ‘incerteza’ igual a 22.0°C.

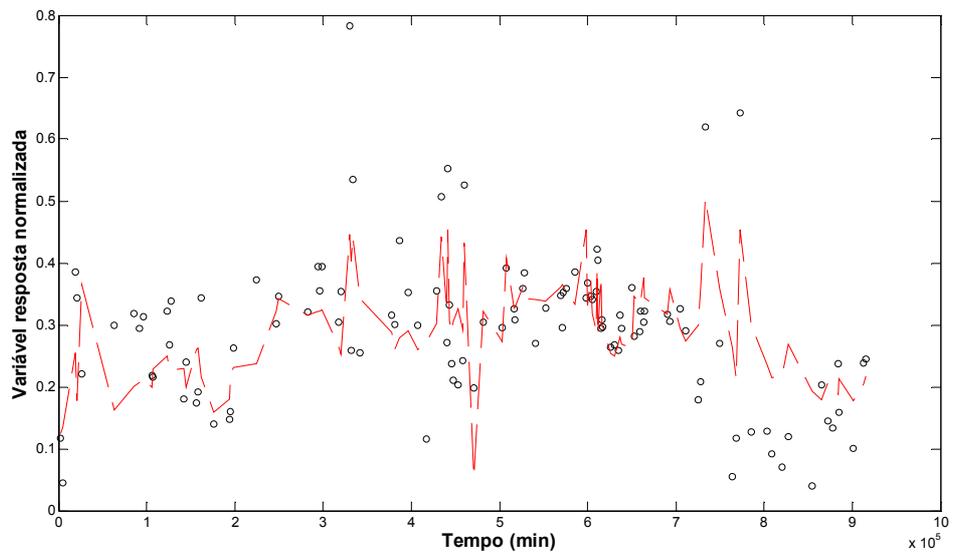


Figura 4.66: Modelo de PCR (dados estacionários) com variabilidade explicada de 100%, com correlação igual a 0.67251 e ‘incerteza’ de 20.9°C.

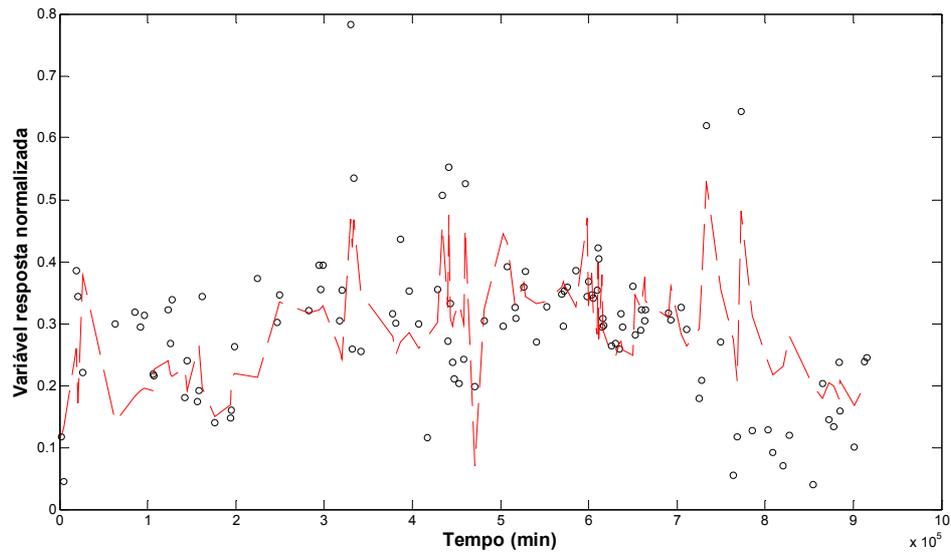


Figura 4.67: Modelo de PLS (dados estacionários) com 2 componentes, com correlação igual a 0.64647 e ‘incerteza’ de 21.7°C.

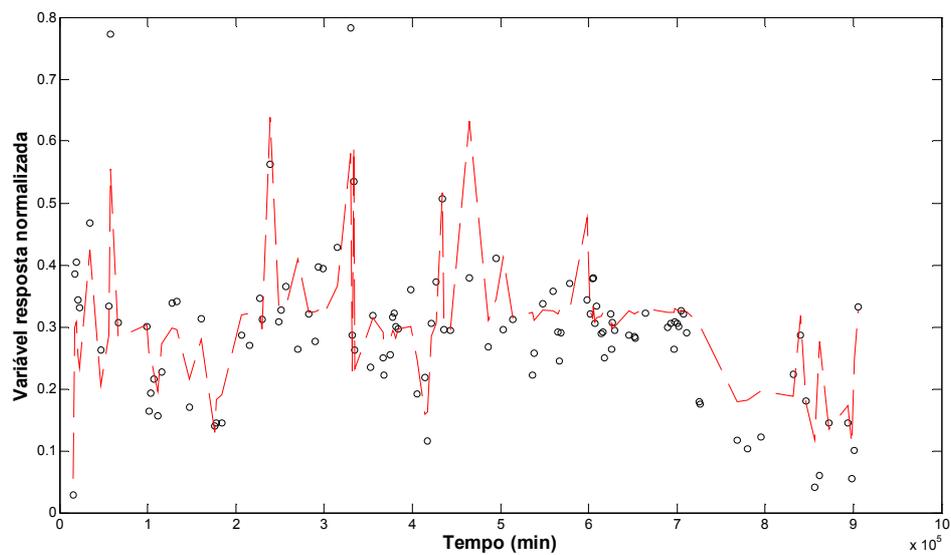


Figura 4.68: RNA regularização Bayesiana (dados estacionários), uma camada oculta, com 2 neurônios, com correlação igual a 0.81534 e ‘incerteza’ de 15.8°C.

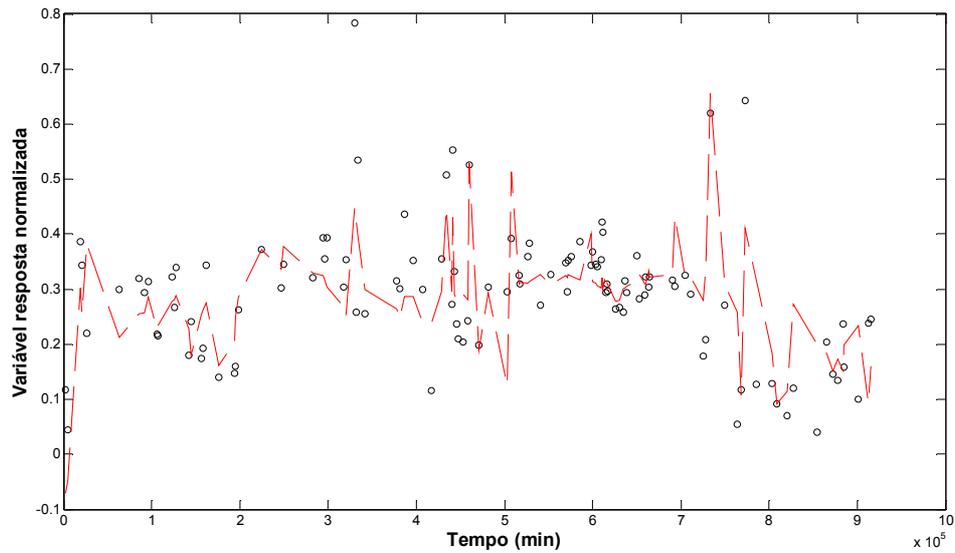


Figura 4.69: RNA back-propagation (dados estacionários), uma camada oculta, com 4 neurônios, com correlação igual a 0.71374 e ‘incerteza’ de 20.0°C.

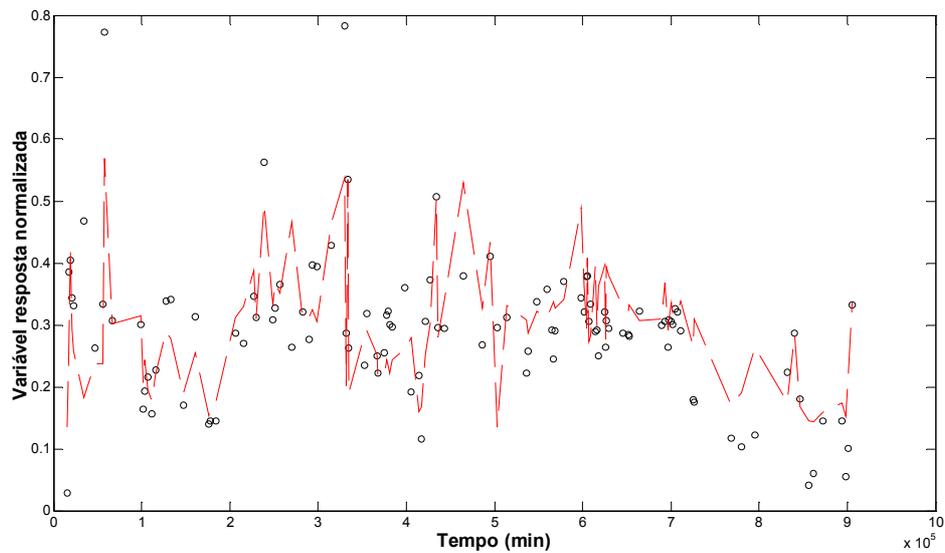


Figura 4.70: RNA back-propagation (dados estacionários), duas camadas ocultas, com 4 neurônios na 1ª, 2 neurônios na 2ª, com correlação igual a 0.75329 e ‘incerteza’ de 18.0°C.

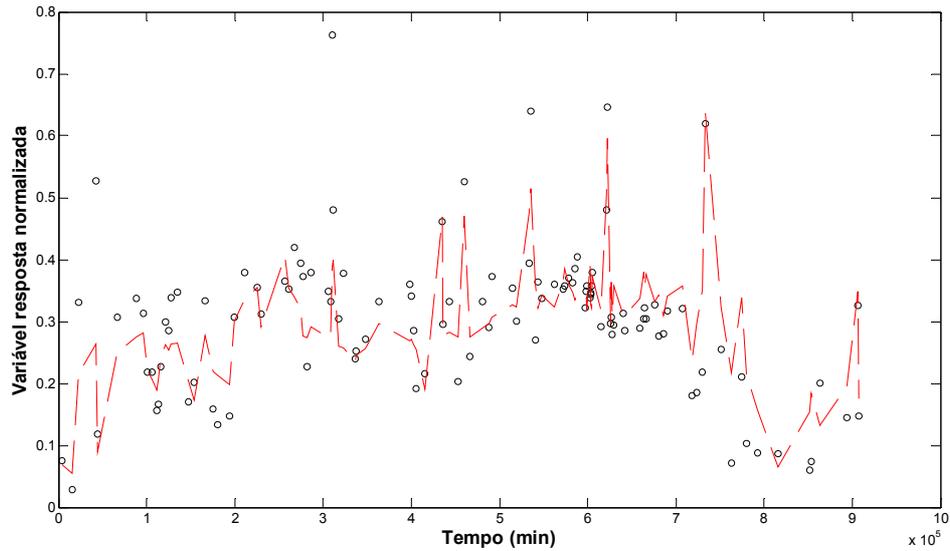


Figura 4.71: RBF (dados estacionários), camada de saída linear, com 6 neurônios, com correlação igual a 0.80755, parâmetro de espalhamento 10 e ‘incerteza’ de 16.8 °C .

Utilizando-se como métrica para seleção de modelos a soma quadrática da derivada do erro em relação ao tempo, que tem como objetivo verificar se a tendência da resposta é condizente com os dados, são obtidas as seguintes curvas:

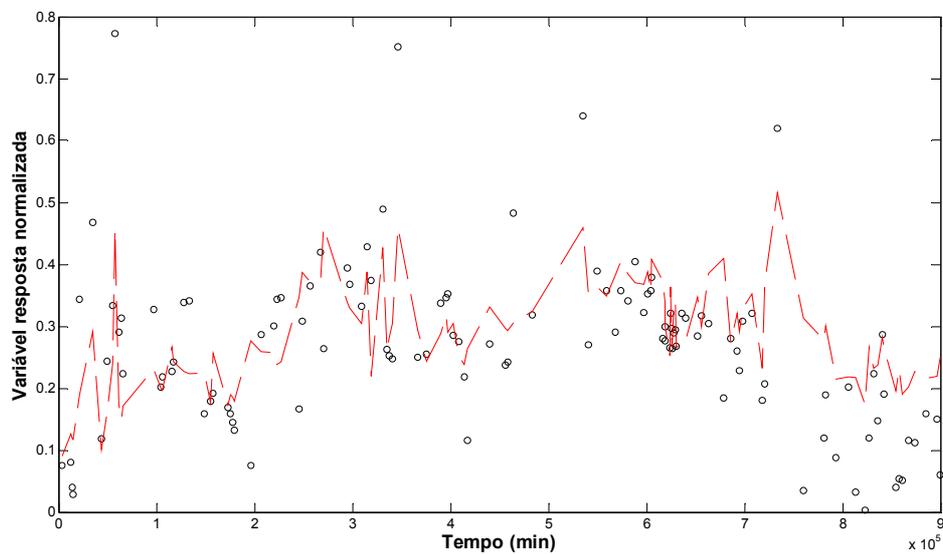


Figura 4.72: Modelo linear (dados estacionários) com correlação igual a 0.67593 e ‘incerteza’ de 23.3°C.

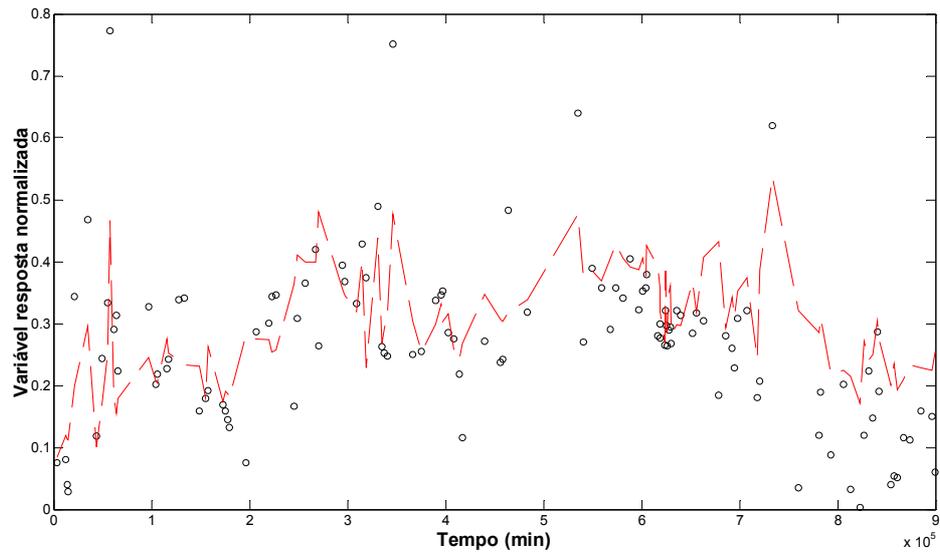


Figura 4.73: Modelo de PCR (dados estacionários) com variabilidade explicada de 100%, com correlação igual a 0.68094 e ‘incerteza’ de 23.1 °C.

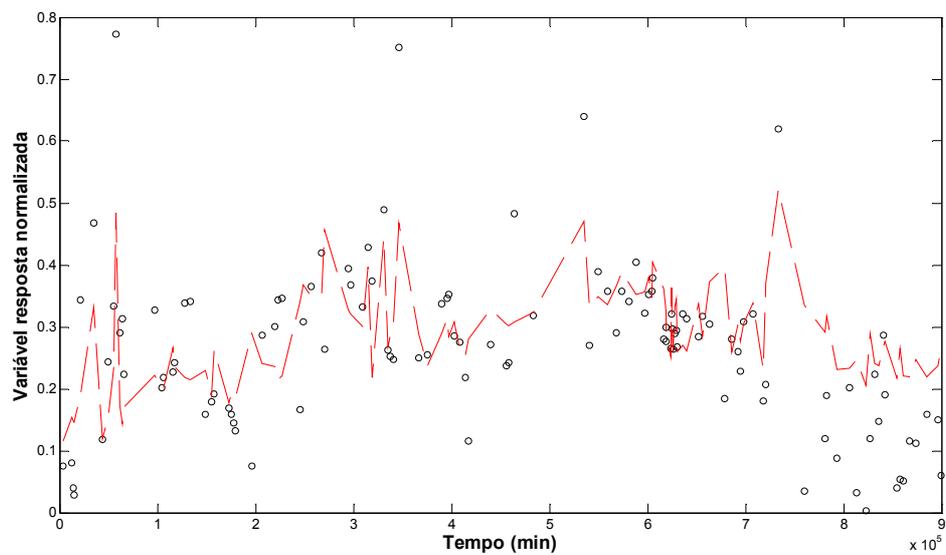


Figura 4.74: Modelo de PLS (dados estacionários) com 2 componentes, com correlação igual a 0.63902 e ‘incerteza’ de 24.3°C.

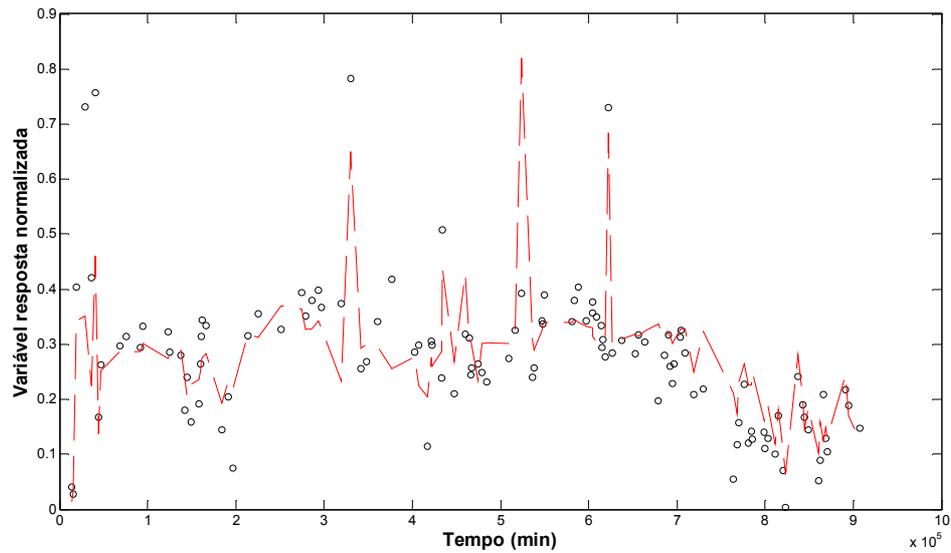


Figura 4.75: RNA regularização Bayesiana (dados estacionários), uma camada oculta, com 2 neurônios, com correlação igual a 0.78088 e ‘incerteza’ de 19.7°C.

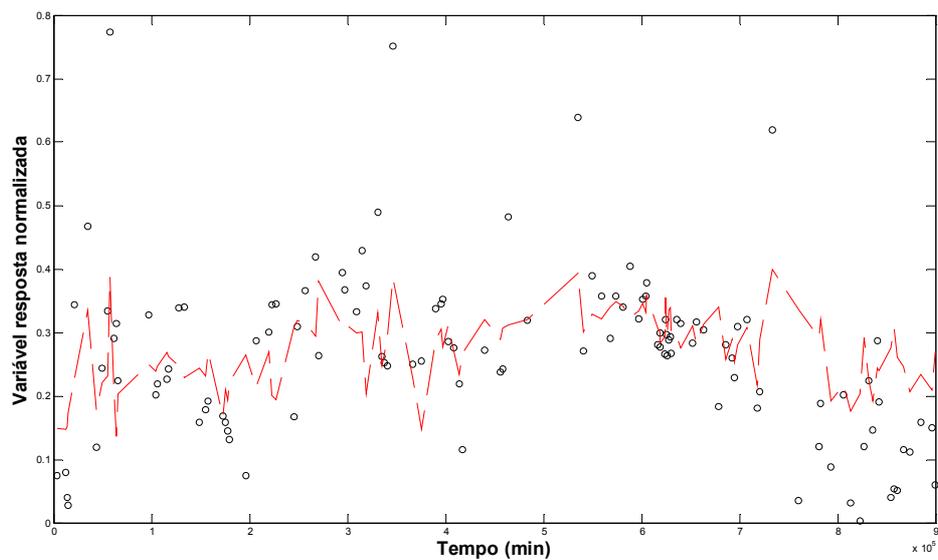


Figura 4.76: RNA back-propagation (dados estacionários), uma camada oculta, com 4 neurônios, com correlação igual a 0.58154 e ‘incerteza’ de 26.1°C.

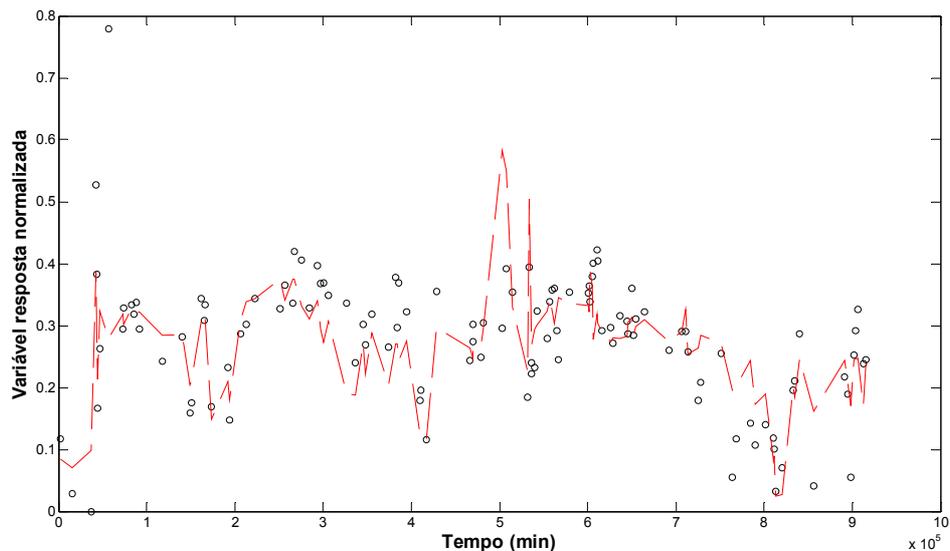


Figura 4.77: RNA back-propagation (dados estacionários), duas camadas ocultas, com 4 neurônios na 1ª, 2 neurônios na 2ª, com correlação igual a 0.69942 e ‘incerteza’ de 18.5°C.

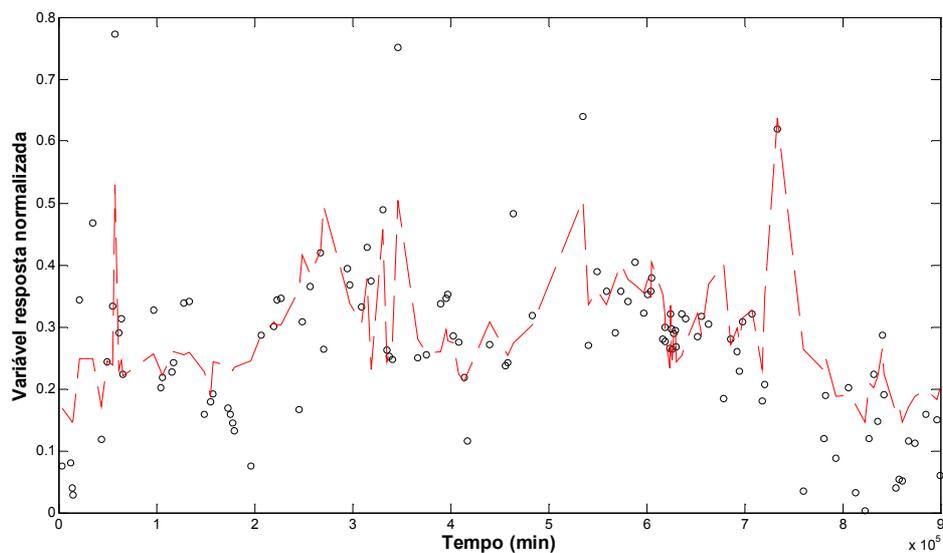


Figura 4.78: RBF (dados estacionários), camada de saída linear, com 6 neurônios, com correlação igual a 0.78631, parâmetro de espalhamento 10 e ‘incerteza’ de 20°C.

4.3.4 Considerações finais sobre a etapa de modelagem

A técnica que apresentou o melhor desempenho foi a rede neural tipo MLP com regularização Bayesiana (RNA1 RB), considerando-se uma camada oculta com dois neurônios. Conforme previsto, o termo de regularização introduz maior suavidade ao

modelo, o que pode ser observado comparando-se a saída da rede com aquela gerada pelo algoritmo *back-propagation* resiliente.

Comparando-se as redes MLP obtidas pelo algoritmo *back-propagation* com as redes tipo RBF, percebe-se que, apesar do melhor poder de generalização das redes MLP em função da utilização de hiperplanos para aproximação, o uso de seis neurônios geram hiperelipsoides com capacidade de predição satisfatória, superior à das redes MLP considerando-se o algoritmo *back-propagation* e quatro neurônios na camada oculta. As redes com duas camadas ocultas, sendo quatro neurônios na primeira camada e dois neurônios na segunda camada, não apresentaram desempenho sensivelmente superior às redes com uma camada oculta e quatro neurônios.

As técnicas lineares testadas apresentaram resultados semelhantes. A utilização de todas as componentes principais no método PCR equipara o mesmo à técnica de MLR, além de reduzir a sua capacidade de filtrar o ruído. Apesar da existência de variáveis correlacionadas entre os regressores (i.e temperatura de topo da fracionadora principal e temperatura do refluxo de topo), essa redundância de informação não se tornou problemática para estimativa dos coeficientes do modelo de MLR, considerando-se que o algoritmo de resolução dos mínimos quadrados pressupõe inversão de matriz. Caso fosse adicionada imprecisão em função da colinearidade, o modelo de PCR seria superior, pois mesmo com a utilização de informação de todas as componentes, a regressão é feita a partir de variáveis ortogonais.

Conforme mencionado na revisão bibliográfica, a técnica de PCR não utiliza o conteúdo de informação presente na variável resposta para a obtenção das componentes. No caso do PLS, o algoritmo NIPALS calcula os *loads* e *scores* das componentes incorporando informações da variável resposta e posteriormente correlaciona estes *scores* com a variável resposta, no caso unidimensional. Neste caso, a quantidade de informação adicional incorporada no cálculo das componentes por esta técnica não foi suficiente para torná-la superior às outras duas técnicas lineares.

Os gráficos de probabilidade normal dos resíduos para os dois modelos de regularização obtidos, baseados nas métricas utilizadas, são exibidos a seguir:

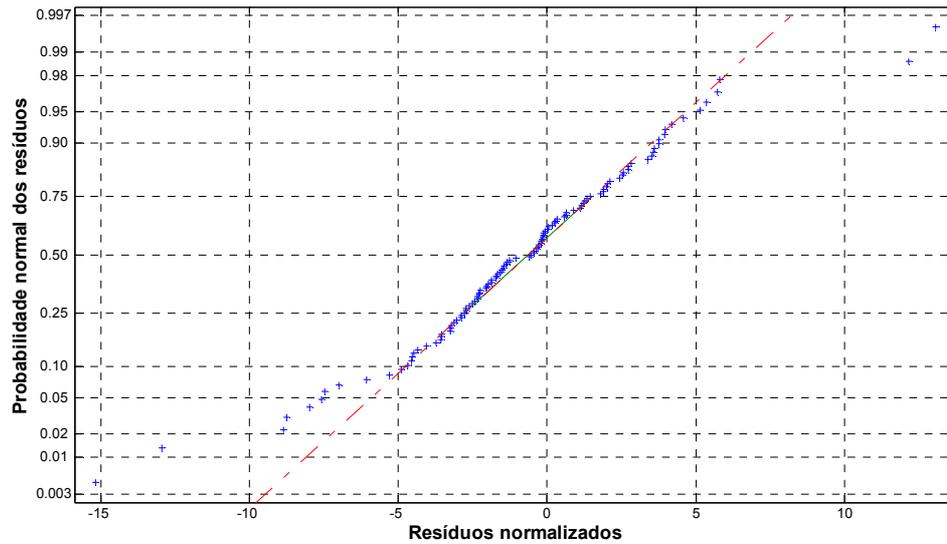


Figura 4.79: Gráfico de probabilidade normal dos resíduos, referente ao modelo de redes neurais MLP (regularização Bayesiana), considerando-se a métrica da soma quadrática dos erros.

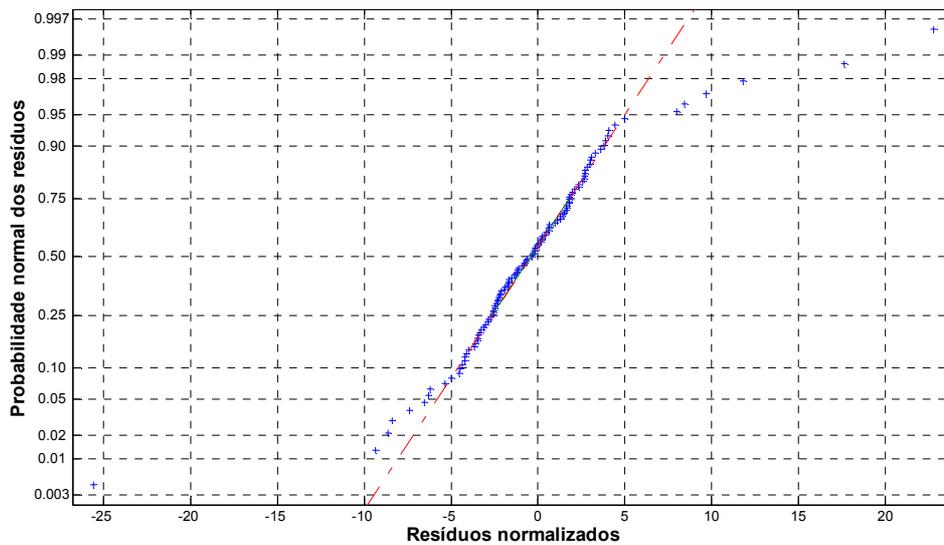


Figura 4.80: Gráfico de probabilidade normal dos resíduos, referente ao modelo de redes neurais MLP (regularização Bayesiana), considerando-se a métrica da soma quadrática da derivada do erro em relação ao tempo.

É possível perceber acentuados desvios da normalidade para valores mais baixos e mais elevados da variável resposta, significando que a variável ‘incerteza’ pode exceder o valor médio determinado. Desta forma, a afirmativa de que apenas uma a cada vinte amostras excederia o valor previsto ao nível de significância de 5% não pode ser feita.

No entanto, para o caso de PFE elevado, o modelo não necessita de prever com exatidão o valor desta propriedade. Basta o mesmo sinalizar que os limites superiores estabelecidos pela operação estão sendo excedidos. No caso de PFE baixo, a maior imprecisão não acarreta perda de especificação da corrente; porém dificulta a otimização da produção de nafta.

Como a modelagem por regularização Bayesiana prevê um ruído normal na saída (equação (2.106)), o gráfico da distribuição normal dos resíduos torna-se especialmente interessante para a verificação da aderência do modelo obtido. Desvios acentuados da normalidade podem sugerir a necessidade de formação de *clusters* de dados, com diferentes funções de aproximação, modeladas para cada conjunto.

5 CONCLUSÃO

Foi desenvolvida uma ferramenta computacional para auxiliar o desenvolvimento de inferências, sendo a mesma aplicável a sistemas contendo um número genérico de variáveis de entrada, porém com a restrição de apenas uma variável de saída. Esta ferramenta pôde ser especialmente aplicada nos casos em que a variável resposta apresenta uma baixa frequência de atualização de informação, tais como os ensaios de laboratório, quando os mesmos são comparados às medições *online* da planta. Esta ferramenta foi testada na obtenção de um modelo de predição para o PFE da nafta leve de craqueamento. A qualidade da inferência obtida pode ainda ser melhorada; contudo, as técnicas disponibilizadas pela aplicação permitem agilizar estudos mais detalhados deste sistema.

O identificador Hampel se mostrou um dispositivo simples para a exclusão de *outliers*. Ressalta-se que esta etapa de filtragem pode ser melhorada através da introdução de técnicas multivariadas, tais como o PCA, para verificação de pontos que não sejam aderentes à distribuição normal multivariada.

A análise dos gráficos de correlação cruzada possibilitou uma estimativa do tempo de residência do inventário no sistema. O atraso de transporte identificado por esta técnica, correspondente a três horas, foi inferior ao período de referência utilizado pelos operadores para amostragens da corrente de nafta após modificações no processo. O período de referência adotado é de quatro horas o que é coerente, ou seja, um período superior ao estimado pela correlação cruzada. Este período é adotado pelos operadores por experiência prévia porque garante a observação da modificação realizada. Contudo, ressaltam-se os cuidados necessários na eliminação de variáveis ao se basear exclusivamente nos limites de corte que caracterizam correlação estatística significativa. Estes limites dependem do número de dados utilizados no vetor para o cálculo de correlação, sendo que o intervalo para rejeição dos dados com correlação estatística significativa será maior quanto menor for o número de dados presentes neste vetor.

As variáveis selecionadas pela regressão linear por etapas são bastante influenciadas pelos períodos de captura de médias configurados. Sendo assim, capturando-se uma variável próxima ao seu pico de correlação cruzada com a variável resposta implica em maior chance de aceitação da mesma no modelo linear. O conjunto de dados deve ser

cuidadosamente explorado por esta ferramenta para garantir a aceitação das melhores variáveis candidatas à regressoras nos modelos. O tempo de residência esperado deve ser coerente com o pico de correlação identificado. Caso contrário, este pico de correlação pode ser resultado de alguma ação indeterminada no processo, sendo que a mesma atua como ‘causa comum’ tanto para a variável resposta quanto para a variável candidata à regressora. A técnica de análise de *clusters* permitiu verificar as variáveis que possibilitaram uma partição mais natural dos níveis do PFE. O vetor de centroides sintetiza informações de diferentes condições de processo, o que auxilia na análise de causalidade para as variáveis regressoras. Ressalta-se que a análise de processo deve nortear as escolhas das variáveis; relações que não fazem sentido fisicamente devem ser investigadas de modo a impedir que as variáveis sejam escolhidas apenas por sua correlação com a resposta.

A caracterização do estado estacionário pelo método de carta de controle pode ser melhorada em função de uma estimativa mais precisa do desvio padrão da variável W . A análise exploratória das arquiteturas para as técnicas de modelagem deve considerar o máximo de repetições possíveis no que diz respeito ao procedimento de validação, garantindo maior exploração do conteúdo de informação dos dados coletados destinados à modelagem.

Conforme esperado, as técnicas não lineares representaram melhor o sistema em estudo. Isso sugere a presença de não linearidades no sistema, o que prejudica o poder de predição das técnicas lineares ao se considerar uma faixa de operação muito ampla.

Conclui-se o trabalho enfatizando o desenvolvimento da ferramenta, a qual pode ser aplicada no dia a dia de engenheiros e técnicos de operação para a obtenção de analisadores virtuais, o que permite a sua utilização em dispositivos de otimização e controle de processos.

5.1 Sugestões para trabalhos futuros

A técnica de redes neurais baseada em regularização Bayesiana apresentou desempenho superior. Contudo, o número de parâmetros efetivos não alcançou o patamar esperado na arquitetura que minimiza o somatório do erro quadrático dos dados de validação.

Sugere-se a investigação deste fenômeno por meio de um novo estudo de caso que utilize esta técnica.

Um recurso interessante a ser incorporado no programa INFERATOR seria a utilização das respostas do módulo de análise de *clusters* para balanceamento das entradas das redes neurais, permitindo melhor aderência do modelo obtido nos casos em que determinados *clusters* não apresentem um volume de dados suficiente para sensibilizar o treinamento.

6 REFERÊNCIAS BIBLIOGRÁFICAS

ABADIE, E. **Craqueamento catalítico**. Rio de Janeiro: Universidade Petrobrás, 2008.

AGUIRRE, L. A. **Introdução à identificação de sistemas**: Técnicas lineares e não-lineares aplicadas a sistemas reais. 3. ed. Belo Horizonte: Editora UFMG, 2007.

AMARI, S. Asymptotic Statistical Theory of Overtraining and Cross-Validation. **IEEE Transactions on Neural Networks**. v.8, n.5, p.985-996, 1997.

AMERICAN SOCIETY FOR TESTING AND MATERIALS. **D-86**: standard test method for distillation of petroleum products at atmospheric pressure. West Conshohocken, 2009.

BARBOSA, E. P.; GNERI, M. A.; MENEGUETTI, A. Range Control Charts Revisited: Simpler Tippett-like Formulae, Its Practical Implementation and the Study of False Alarm. In: SINAPE - SIMPÓSIO NACIONAL DE PROBABILIDADE E ESTATÍSTICA, 19, 2010, São Pedro, SP.

BRAGA, A. P.; CARVALHO, A. C. P. L. F.; LUDEMIR, T. B. **Redes neurais artificiais**: Teoria e aplicações. 2. ed. Rio de Janeiro: LTC, 2007.

CALDAS, J. N. et al. **Internos de torres**: pratos & recheios. Rio de Janeiro: Editora Interciência, 2007.

CYBENKO, G. Approximation by Superpositions of a Sigmoidal Function. **Math. Control Signals Systems**. v.2, p. 303-314, 1989.

EFRON, B; TIBSHIRANI, R. **Cross-validation and the bootstrap: estimating the error rate of a prediction rule.** Technical Report n° 176, Stanford, California, 1995.

ESQUERRE, K. P. S. O. R. **Aplicação de técnicas estatísticas multivariadas e de redes neurais na modelagem de um sistema de tratamento de efluentes industriais.** 2003. 150f. Tese (Doutorado em Engenharia Química) - Universidade Estadual de Campinas, Campinas, 2003.

FACCHIN, S; TRIERWIELER, J. O; CONZ, V. **Soft Sensor Design: A New Approach for Variable Selection.** In: Mercosur Congress on Chemical Engineering, n 2.; 2005. Rio de Janeiro. p.1-10.

FARAH, M. A. **Caracterização do petróleo e seus derivados.** Rio de Janeiro: Universidade Petrobrás, 2008.

FEITAL, T. S. **Monitoramento da Condição de Processos Químicos Industriais.** 2009. Exame de Qualificação ao Doutorado - Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2009.

FERREIRA, M. M. C. et al. **Quimiometria I: Calibração Multivariada, um Tutorial. Química Nova.** v. 22 (5), p. 724- 731, 1999.

FORESEE, F. D; HAGAN, M. T. **Gauss-Newton approximation to Bayesian learning.** In: INTERNATIONAL CONFERENCE ON NEURAL NETWORKS, 1997, Houston, IEEE Xplore, v.3, p. 1930-1935

HAYKIN, S. **Redes neurais: Princípios e prática.** 2. ed. Porto Alegre: Bookman, 2001.

INDIO DO BRASIL, N. **Destilação de petróleo**. Rio de Janeiro: Universidade Petrobrás, 2008.

JOHNSON, R. A.; WICHERN, D. W. **Applied multivariate statistical analysis**. 3rd ed. Upper Saddle River, New Jersey: Prentice Hall, 1992.

KIM, M. et al. Design of a steady-state detector for fault detection and diagnosis of a residential air conditioner. **International Journal of Refrigeration**, v.31, n.5, p. 790-799, 2008.

KOCH-GLITSCH. **Intalox packed tower system: IMTP[®] high performance packing**. United States of America, 2003.

Disponível em: < <http://www.koch-glitsch.com/Document%20Library/KGIMTP.pdf> >.
Acesso em: 25 set. 2010.

KOCH-GLITSCH. **Intalox packed tower system: packed tower internals**. United States of America, 2003.

Disponível em: <http://www.kochglitsch.com/koch/product_brochures/KGMTIG.pdf>.
Acesso em: 25 set. 2010.

LIN, B. et al. A systematic approach for soft sensor development. **Computers and Chemical Engineering**. v.31, p. 419–425, 2007.

MACKAY, D. J. C. Bayesian interpolation. **Neural Computation**. v.4, p.415-447, 1992a.

MACKAY, D. J. C. A Practical Bayesian Framework for Backpropagation Networks. **Neural Computation**. v.4, p.448-472, 1992b.

MAGALHÃES, M. N.; LIMA, A. C. P. **Noções de probabilidade e estatística**. 4. ed. São Paulo: Edusp, 2002.

MANNE, R. Analysis of two partial-least-squares algorithms for multivariate calibration. **Chemometrics and Intelligent Laboratory Systems**.v.2, p. 187-197, 1987.

MATTOS, L. G. Efeito das variáveis operacionais na formação de goma na nafta craqueada. 2004. 132 f. Dissertação (Mestrado em Engenharia Química) – Escola de Engenharia, Universidade Federal de Minas Gerais, Belo Horizonte, 2004.

MINGOTI, S. A. **Análise de dados através de métodos de estatística multivariada: uma abordagem aplicada**. Belo Horizonte: Editora UFMG, 2005.

MONTGOMERY, D. **Introduction to Statistical Quality Control**. 5. ed. United States of America: John Wiley & Sons, 2005.

MONTGOMERY, D; RUNGER, G. C. **Estatística aplicada e probabilidade para engenheiros**. 2. ed. Rio de Janeiro: LTC, 2003.

RAFIC, M. Y.; BUGMANN, G; EASTERBROOK, D. J. Neural Network Design for Engineering Applications. **Computers and Structures**. v.79, p.1541-1552, 10 Abr 2001.

RESOLUÇÃO ANP Nº 38, 09 DEZ 2009 DOU 10.12.2009

ROSIPAL, R; KRÄMER, N. Overview and Recent Advances in Partial Least Squares. **LNCS**. v.3940, p. 34–51, 2006.

SCHEUER, A. **Instalação e Administração do Sistema PI na Unidade Multipropósito de FCC**. 2004. Monografia de Projeto de Fim de Curso - Universidade Federal de Santa Catarina, Florianópolis, 2004.

SECCHI, A. R. et al. A dynamic model for a FCC UOP stacked converter unit. **Computers and Chemical Engineering** .v. 25, p.851-858, 2001.

SHARMIN, R. et al. Inferential sensors for estimation of polymer quality parameters: Industrial application of a PLS-based soft sensor for a LDPE plant. **Chemical Engineering Science**. v. 61, p. 6372-6384, 2006.

SJÖBERG, J; LJUNG, L. Overtraining, Regularization, and Searching of Minimum in Neural Networks. Linköping University, p. 1-16, 1992.

TREYBAL, R. E. **Mass-Transfer operations**. 3rd ed. Singapore: McGraw-Hill Book Co, 1980.

WOLD, S. et al. The PLS method - partial least squares projections to latent structures - and its applications in industrial RDP (research, development, and production). p. 1-44, 2004.

ZHAO,Y. A Soft Sensor Based on Orthogonal Nonlinear Principal Component Analysis. **Dev. Chem. Eng. Mineral Process.**v.13, p.233-242, 2005.

ANEXO A

Considere amostras aleatórias da variável X , tal que:

$$X_1, X_2, \dots, X_n \quad (\text{A.1})$$

Estas amostras são independentes e identicamente distribuídas. Considere ainda que a variável X possua função de distribuição de probabilidade $F(x)$.

Ordenando-se de forma crescente as amostras anteriores, obtêm-se:

$$X_{(1)}, X_{(2)}, \dots, X_{(n)} \quad (\text{A.2})$$

A variável “espaçamento” S_r é definida como:

$$S_r = X_{(r+1)} - X_{(r)} \quad r = 1, 2, \dots, n-1 \quad (\text{A.3})$$

Define-se a amplitude R como:

$$R = \sum_{r=1}^{n-1} S_r = X_{(n)} - X_{(1)} \quad (\text{A.4})$$

A função de distribuição para a r -ésima amostra ordenada e dada pelo binômio

$$F_r(x) = \sum_{i=r}^n \binom{n}{i} F^i(x) (1-F(x))^{n-i} \quad (\text{A.5})$$

Esta expressão significa a probabilidade de r ou mais valores de X_i serem inferiores a um valor de corte x . Define-se a esperança da variável X como:

$$E(X) = \int_0^{\infty} (1-F(X)) dx - \int_{-\infty}^0 F(X) dx \quad (\text{A.6})$$

A esperança da variável espaçamento pode ser obtida aplicando-se a definição (A.3) em (A.6):

$$E(S_r) = E(X_{(r+1)} - X_{(r)}) = \int_{-\infty}^{\infty} (F_r(x) - F_{r+1}(x)) dx \quad (\text{A.7})$$

Aplicando-se a definição (A.5) em (A.7), obtêm-se:

$$E(S_r) = \int_{-\infty}^{\infty} \binom{n}{r} F^r(x) [1 - F(x)]^{n-r} dx \quad (\text{A.8})$$

A definição (A.4) pode ser aplicada na equação (A.8), o que resulta a expressão da média d_2 da distribuição W , denominada primeira integral de Tippett:

$$W = \frac{R}{\sigma} \quad (\text{A.9})$$

$$E(R) = E\left(\sum_{r=1}^{n-1} S_r\right) = \sum_{r=1}^{n-1} E(S_r) = \int_{-\infty}^{\infty} \sum_{r=1}^{n-1} \binom{n}{r} F^r(x) [1 - F(x)]^{n-r} dx \quad (\text{A.10})$$

$$d_2 = E(W) = \frac{1}{\sigma} \int_{-\infty}^{\infty} \sum_{r=1}^{n-1} \binom{n}{r} F^r(x) [1 - F(x)]^{n-r} dx \quad (\text{A.11})$$

Pode-se demonstrar que o desvio padrão d_3 da distribuição W pode ser calculado aplicando-se:

$$d_3 = \sqrt{E(W^2) - (d_2)^2} \quad (\text{A.12})$$

$$E(W^2) = 2 \int_{-\infty}^{\infty} \int_{-\infty}^v \left\{ 1 - [1 - F(u)]^n - F^n(v) + [F(v) - F(u)]^n \right\} dudv \quad (\text{A.13})$$

ANEXO B

Para ajustar os coeficientes do modelo de regressão, utiliza-se uma abordagem matricial (MONTGOMERY, 2003, p.234). Supondo a existência de $n > k$ observações, dado a inclusão de k variáveis no modelo, tem-se que:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (\text{B.1})$$

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \quad (\text{B.2})$$

$$\mathbf{X} = \begin{bmatrix} 1 & x_{12} & x_{13} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix} \quad (\text{B.3})$$

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} \quad (\text{B.4})$$

O valor L , expresso na equação (B.5), representa o somatório do quadrado da distância entre o valor estimado pelo modelo e o valor real obtido do processo:

$$L = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \quad (\text{B.5})$$

$$L = \mathbf{y}^T \mathbf{y} - \mathbf{y}^T \mathbf{X}\boldsymbol{\beta} - \boldsymbol{\beta}^T \mathbf{X}^T \mathbf{y} + \boldsymbol{\beta}^T \mathbf{X}^T \mathbf{X}\boldsymbol{\beta} \quad (\text{B.6})$$

Deseja-se obter o vetor $\boldsymbol{\beta}$ que minimize este somatório, expresso pela equação (B.6). Sendo assim (AGUIRRE, 2007, p.225), deriva-se a função L em relação ao vetor $\boldsymbol{\beta}$, tal que:

$$\frac{\partial L}{\partial \boldsymbol{\beta}} = 0 \quad (\text{B.7})$$

Utilizando-se as identidades (B.8), (B.9) e (B.10), obtêm-se a equação (B.12):

$$\frac{\partial(\mathbf{z}^T \mathbf{w})}{\partial \mathbf{w}} = \mathbf{z} \quad (\text{B.8})$$

$$\frac{\partial(\mathbf{w}^T \mathbf{z})}{\partial \mathbf{w}} = \mathbf{z} \quad (\text{B.9})$$

$$\frac{\partial(\mathbf{z}^T \mathbf{A} \mathbf{z})}{\partial \mathbf{z}} = (\mathbf{A} + \mathbf{A}^T) \mathbf{z} \quad (\text{B.10})$$

$$\frac{\partial L}{\partial \boldsymbol{\beta}} = -(\mathbf{y}^T \mathbf{X})^T - \mathbf{X}^T \mathbf{y} + (\mathbf{X}^T \mathbf{X} + \mathbf{X}^T \mathbf{X}) \boldsymbol{\beta} = 0 \quad (\text{B.11})$$

$$\boldsymbol{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} \quad (\text{B.12})$$

ANEXO C

A abordagem Bayesiana para a etapa de modelagem implica em maximizar a probabilidade de se obter o vetor de parâmetros \mathbf{w} , dados α^B, β^B , o conjunto de dados \mathbf{D} , a arquitetura de rede \mathcal{M}^B e o termo de regularização \mathcal{R}^B . A equação (C.1) ilustra essa abordagem:

$$P(\mathbf{w} | \mathbf{D}; \alpha^B; \beta^B; \mathcal{M}^B) = \frac{P(\mathbf{D} | \mathbf{w}; \beta^B; \mathcal{M}^B) P(\mathbf{w} | \alpha^B; \mathcal{M}^B; \mathcal{R}^B)}{P(\mathbf{D} | \alpha^B; \beta^B; \mathcal{M}^B; \mathcal{R}^B)} \quad (\text{C.1})$$

Para a variável resposta observada y_i , a abordagem Bayesiana considera a presença de um ruído branco aditivo na saída ν , com média nula e variância σ_ν^2 , conforme equação (C.2):

$$y_i = f(\mathbf{x}_i, \mathbf{w}) + \nu_i \quad (\text{C.2})$$

Mackay (1992b, p.449) aplicou a técnica de regularização Bayesiana nas etapas de modelagem e seleção de arquitetura de redes. A função objetivo a ser minimizada é expressa pela equação (C.3), que resulta matematicamente na expressão (C.4) (MACKAY, 1992b, p.450).

$$\varepsilon_F = \beta^B \varepsilon_E + \alpha^B \varepsilon_R \quad (\text{C.3})$$

$$\varepsilon_F = \frac{\beta^B}{2} \sum_j [(y_j - f(\mathbf{x}_j, \mathbf{w}))^2] + \frac{\alpha^B}{2} \sum_i w_i^2 \quad (\text{C.4})$$

A abordagem Bayesiana considera os pesos \mathbf{w} da rede como variáveis aleatórias com distribuição normal:

$$P(\mathbf{w} | \alpha^B; \mathcal{M}^B, \mathcal{R}^B) = \frac{\exp(-\alpha^B \varepsilon_R)}{\int d^{N_B} \mathbf{w} \exp(-\alpha^B \varepsilon_R)} \quad (\text{C.5})$$

Da mesma forma, a abordagem Bayesiana considera como normal a distribuição de probabilidade do ruído presente nos dados \mathbf{D} de treinamento:

$$P(\mathbf{D} | \mathbf{w}; \beta^B; \mathcal{M}^B) = \frac{\exp(-\beta^B \varepsilon_E)}{\int d \mathbf{D} \exp(-\beta^B \varepsilon_E)} \quad (\text{C.6})$$

A partir das hipóteses anteriores de normalidade, as seguintes simplificações matemáticas das equações (C.5) e (C.6) podem ser obtidas:

$$P(\mathbf{w} | \alpha^B; \mathcal{M}^B, \mathcal{R}^B) = \frac{\exp(-\alpha^B \varepsilon_R)}{\left(\frac{\pi}{\alpha^B}\right)^{\frac{N_B}{2}}} \quad (\text{C.7})$$

$$P(\mathbf{D} | \mathbf{w}; \beta^B; \mathcal{M}^B) = \frac{\exp(-\beta^B \varepsilon_E)}{\left(\frac{\pi}{\beta^B}\right)^{\frac{n_B}{2}}} \quad (\text{C.8})$$

A variável N_B representa o número total de parâmetros livres da rede, ou seja, a dimensão do vetor \mathbf{w} . Já a variável n_B diz respeito ao produto do número de neurônios na camada de saída pelo número de amostras no conjunto de dados, ou seja, caracteriza o número de graus de liberdade do problema para determinação do vetor \mathbf{w} (MACKAY, 1992b, p.454). A expressão (C.7) representa a probabilidade de se obter os pesos \mathbf{w} dado o parâmetro de ponderação α^B , a arquitetura de rede \mathcal{M}^B e o funcional de regularização \mathcal{R}^B . Já a expressão (C.8) representa a probabilidade de se obter os dados \mathbf{D} tendo em vista o vetor de pesos \mathbf{w} , o parâmetro β^B e a arquitetura \mathcal{M}^B .

Tendo em vista as hipóteses de normalidade, verifica-se que os parâmetros α^B e β^B estão relacionados às variâncias das distribuições de probabilidade anteriores pelas equações (C.9) e (C.10):

$$\sigma_w^2 = \frac{1}{\alpha^B} \quad (\text{C.9})$$

$$\sigma_v^2 = \frac{1}{\beta^B} \quad (\text{C.10})$$

Aplicando-se as equações (C.7) e (C.8) na equação (C.1), obtêm-se as seguintes igualdades:

$$P(\mathbf{w} | \mathbf{D}; \alpha^B; \beta^B; \mathcal{M}^B) = \frac{\exp(-\varepsilon_F)}{Z_F(\alpha^B; \beta^B)} \quad (\text{C.11})$$

$$\exp(-\varepsilon_F) = \exp(-\alpha^B \varepsilon_R - \beta^B \varepsilon_E) \quad (\text{C.12})$$

$$Z_F(\alpha^B; \beta^B) = \int d^N \mathbf{w} \exp(-\varepsilon_F) \quad (\text{C.13})$$

$$Z_F(\alpha^B; \beta^B) = \left(\frac{\pi}{\beta^B} \right)^{\frac{n_B}{2}} \left(\frac{\pi}{\alpha^B} \right)^{\frac{N_B}{2}} P(\mathbf{D} | \alpha^B; \beta^B; \mathcal{M}^B; \mathcal{R}^B) \quad (\text{C.14})$$

A abordagem Bayesiana consiste em maximizar $P(\mathbf{w} | \mathbf{D}; \alpha^B; \beta^B; \mathcal{M}^B)$, ou seja, minimizar a função objetivo ε_F . Os pesos obtidos pela minimização desta função são caracterizados como os “mais prováveis”, ou seja, \mathbf{w}^{MP} . Contudo, os valores ótimos dos parâmetros de regularização α^B e β^B precisam ser inicialmente obtidos à luz dos dados \mathbf{D} coletados para modelagem.

A determinação dos parâmetros ótimos α^B e β^B implica na maximização da seguinte função de probabilidade:

$$P(\alpha^B; \beta^B | \mathbf{D}; \mathcal{M}^B, \mathcal{R}^B) = \frac{P(\mathbf{D} | \alpha^B; \beta^B; \mathcal{M}^B, \mathcal{R}^B) P(\alpha^B; \beta^B)}{P(\mathbf{D} | \mathcal{M}^B, \mathcal{R}^B)} \quad (\text{C.15})$$

Assumindo-se uma distribuição uniforme de probabilidade $P(\alpha^B; \beta^B)$ para os parâmetros de regularização α^B e β^B , a maximização da função de distribuição $P(\alpha^B; \beta^B | \mathbf{D}; \mathcal{M}^B, \mathcal{R}^B)$ implica na maximização do termo $P(\mathbf{D} | \alpha^B; \beta^B; \mathcal{M}^B, \mathcal{R}^B)$, que nada mais é do que o fator de normalização da equação (C.1). Conhecendo-se a distribuição de probabilidade dos termos de informação a priori, evidência e verossimilhança da equação (C.1) (lado direito da equação), a distribuição do termo posterior pode ser inferida (lado esquerdo da equação).

Dito isso, uma aproximação do termo $Z_F(\alpha^B; \beta^B)$ precisa ser obtida. Assumindo-se que a função objetivo apresenta formato quadrático em uma área próxima ao mínimo global, considerando-se a aplicação dos pesos “mais prováveis” na função objetivo, uma aproximação por série de Taylor resulta nas seguintes equações:

$$\varepsilon_F = \varepsilon_F^{MP} + \frac{1}{2}(\mathbf{w} - \mathbf{w}^{MP})^T \mathbf{H}(\mathbf{w} - \mathbf{w}^{MP}) \quad (\text{C.16})$$

$$\mathbf{H} = \alpha^B (\nabla^2 \varepsilon_R) + \beta^B (\nabla^2 \varepsilon_E) \quad (\text{C.17})$$

As equações (C.16) e (C.17) combinadas com a equação (C.13), por sua vez, resultam na aproximação:

$$Z_F(\alpha^B; \beta^B) = (2\pi)^{N_B/2} (\det(\mathbf{H}))^{-1/2} \exp(-\varepsilon_F) \quad (\text{C.18})$$

Reorganizando-se os termos da equação (C.1) é possível substituir a aproximação obtida na equação (C.18), o que resulta na equação (C.19):

$$P(\mathbf{D} | \alpha^B; \beta^B; \mathcal{M}^B; \mathcal{R}^B) = \frac{(2\pi)^{N_B/2} (\det(\mathbf{H}))^{-1/2} \exp(-\varepsilon_F)}{\left(\frac{\pi}{\beta^B}\right)^{\frac{n_B}{2}} \left(\frac{\pi}{\alpha^B}\right)^{\frac{N_B}{2}}} \quad (\text{C.19})$$

A resolução para os valores ótimos de α^B e β^B implica em se obter as derivadas do logaritmo da equação (C.19). Antes, porém, Mackay apud Gull (MACKAY, 1992a, p.432) realiza uma transformação geométrica de modo que o hessiano da função ε_R coincida com a matriz identidade, ou seja:

$$\nabla^2 \varepsilon_R = \mathbf{I} \quad (\text{C.20})$$

Aplica-se, em seguida, a função logaritmo em ambos os lados da equação (C.19), de modo a se obter:

$$\begin{aligned} \ln(P(\mathbf{D} | \alpha^B; \beta^B; \mathcal{M}^B; \mathcal{R}^B)) &= -\alpha^B \nabla^2 \varepsilon_R - \beta^B \nabla^2 \varepsilon_E - \frac{1}{2} \ln(\det(\mathbf{H})) + \\ &\frac{n_B}{2} \ln \beta^B + \frac{N_B}{2} \ln \alpha^B + \frac{N_B}{2} \ln(2\pi) \end{aligned} \quad (\text{C.21})$$

A diferenciação da equação (C.21) em relação à α^B necessita da avaliação do termo

$\frac{d}{d\alpha^B} \ln(\det(\mathbf{H}))$. Utilizando-se a transformação geométrica referenciada, tem-se que:

$$\mathbf{H} = \alpha^B \mathbf{I} + \beta^B \nabla^2 \varepsilon_E \quad (\text{C.22})$$

$$\frac{d}{d\alpha^B} \ln(\det(\mathbf{H})) = \text{Traço}\left(\mathbf{H}^{-1} \frac{d\mathbf{H}}{d\alpha^B}\right) = \text{Traço}(\mathbf{H}^{-1}\mathbf{I}) = \text{Traço}(\mathbf{H}^{-1}) \quad (\text{C.23})$$

Diferenciando-se a equação (C.21) em relação à α^B e β^B , igualando-se à zero, obtêm-se os seguintes resultados:

$$\gamma = N_B - 2\alpha^B (\text{Traço}(\mathbf{H}^{-1})) \quad (\text{C.24})$$

$$\alpha^B = \frac{\gamma}{2\varepsilon_R} \quad (\text{C.25})$$

$$\beta^B = \frac{n_B - \gamma}{2\varepsilon_E} \quad (\text{C.26})$$

A variável γ representa o número efetivo de parâmetros utilizados pela rede para redução da função erro (FORESSE; HAGAN, 1997, p. 1931).