PONTIFÍCIA UNIVERSIDADE CATÓLICA DO PARANÁ ESCOLA POLITÉCNICA PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE PRODUÇÃO E SISTEMAS

SIMONE MASSULINI ACOSTA

CARTAS DE CONTROLE BASEADAS EM MÁQUINAS DE VETORES DE RELEVÂNCIA, MÁQUINAS DE VETORES DE SUPORTE E REDES NEURAIS ARTIFICIAIS PARA MONITORAMENTO DE PROCESSOS INDUSTRIAIS

> CURITIBA 2018

SIMONE MASSULINI ACOSTA

CARTAS DE CONTROLE BASEADAS EM MÁQUINAS DE VETORES DE RELEVÂNCIA, MÁQUINAS DE VETORES DE SUPORTE E REDES NEURAIS ARTIFICIAIS PARA MONITORAMENTO DE PROCESSOS INDUSTRIAIS

Tese apresentada ao Programa de Pós-Graduação em Engenharia de Produção e Sistemas da Pontifícia Universidade Católica do Paraná, como requisito à obtenção do título de Doutor em Engenharia de Produção e Sistemas.

Área de concentração: Gerência de Produção e Logística.

Orientador: Prof. Dr. Ângelo Márcio Oliveira Sant'Anna

Dados da Catalogação na Publicação Pontifícia Universidade Católica do Paraná Sistema Integrado de Bibliotecas – SIBI/PUCPR Biblioteca Central Edilene de Oliveira dos Santos CRB/9 -1636

Acosta, Simone Massulini A185c Cartas de controle baseadas em máquinas de vetores de relevância, máquinas de vetores de suporte e redes neurais artificiais para monitoramento 2018 de processos industriais / Simone Massulini Acosta ; orientador, Ângelo Márcio Oliveira Sant'Anna. --2018 153 f. : il. ; 30 cm Tese (doutorado) - Pontifícia Universidade Católica do Paraná, Curitiba, 2018. Bibliografia: f. 136-153 1. Administração da produção. 2. Engenharia da produção. 3. Controle de Processos - Métodos estatísticos. 4. Redes neurais (Computação). I. SantAnna, Ângelo Márcio Oliveira II. Pontifícia Universidade Católica do Paraná. Programa de Pós-Graduação em Engenharia de Produção e Sistemas. III. Título. CDD 20. ed. - 658.51



Pontifícia Universidade Católica do Paraná Escola Politécnica Programa de Pós Graduação em Engenharia de Produção e Sistemas - PPGEPS

TERMO DE APROVAÇÃO

Simone Massulini Acosta

CARTAS DE CONTROLE BASEADAS EM MÁQUINAS DE VETORES DE RELEVÂNCIA, MÁQUINAS DE VETORES DE SUPORTE E REDES NEURAIS ARTIFICIAIS PARA MONITORAMENTO DE PROCESSOS INDUSTRIAIS.

Tese aprovada como requisito parcial para obtenção do grau de Doutor no Curso de Doutorado em Engenharia de Produção e Sistemas, Programa de Pós-Graduação em Engenharia de Produção e Sistemas, da Escola Politécnica da Pontifícia Universidade Católica do Paraná, pela seguinte banca examinadora:

Presidenté da Banca Prof. Dr. Ângelo Márcio Oliveira Sant'Anna (Orientador)

Prof. Dr. Leandro dos Santos Coelho (Membro Interno)

> Prof. Dr. Osiris Canciglieri Junior (Membro Interno)

Docco Profa. Dra. Viviana Cocco Mariani (Membro Externo)

Prof. Dr. Danilo Marcondes Filho (Membro Externo)

Curitiba, 14 de novembro de 2018.

Rua Imaculada Conceição, 1155 - Prado Velho - CEP 80215-901 - Curitiba - Paraná - Brasil Tel.: (41) 3271-2579 - www.pucpr.br/ppgeps

Aos meus pais, Pedro e Glaci

Ao meu marido, Anderson

AGRADECIMENTOS

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), pela concessão da bolsa de estudos.

Ao Programa de Pós-Graduação em Engenharia de Produção e Sistemas da Pontifícia Universidade Católica do Paraná, pelas condições disponibilizadas para a realização desta tese.

Ao meu orientador, professor Ângelo Márcio Oliveira Sant'Anna, por sua orientação, incentivo e amizade durante a realização do doutorado.

Aos membros das bancas examinadoras de qualificação e de defesa, professores Danilo Marcondes Filho, Leandro dos Santos Coelho, Osiris Canciglieri Junior, Roberto Zanetti Freire e professora Viviana Cocco Mariani, pelas contribuições para a conclusão desta tese.

Aos meus colegas de trabalho do Grupo de Automação do Departamento de Eletrônica, pelo auxílio durante a realização do doutorado, e à Universidade Tecnológica Federal do Paraná, que condições disponibilizadas para a realização do doutorado.

Aos meus pais, Pedro e Glaci, pelo apoio aos meus estudos em toda a minha formação educacional e pelo amor em todos os momentos.

À João Alexandre, Ana Patrícia, Júlia, Carmen, Álvaro, Rodrigo, Rafaela, Paulo, Rose, Ana Carolina, Pedro Paulo, Tânia, Rodrigo, Zenaide, Andressa, Daniel e Heloísa, pelas alegrias de todos os dias.

Ao meu marido Anderson, por seu amor, apoio, companheirismo e incentivo em todos estes anos de convivência.

À Deus por guiar meus passos.

À todas as pessoas que contribuíram de alguma forma para a realização do doutorado.

RESUMO

O Controle Estatístico de Processos (CEP) é uma das estratégias da engenharia da qualidade mais utilizadas na indústria para o monitoramento do desempenho de processos industriais. A principal ferramenta do CEP é a carta de controle, sendo utilizada para monitorar o desempenho das variáveis do processo. Cartas de controle baseadas em modelos de regressão são apropriadas para o monitoramento de características da qualidade que variam em função de uma ou mais variáveis de controle. Nos últimos anos, as técnicas de aprendizado de máguina, tais como Redes Neurais Artificiais (RNAs) e Máquinas de Vetores de Suporte (SVMs), estão sendo utilizadas para modelagem e monitoramento de processos, e a técnica de Máguinas de Vetores de Relevância (RVMs) apresenta aplicações em modelagem de processos. As RNAs, SVMs e RVMs podem ser utilizadas para problemas de classificação e de regressão, sendo que não é necessário assumir uma distribuição de probabilidade específica para a implementação dos modelos. As RNAs do tipo Perceptron Multicamadas (MLP) possuem a capacidade de aproximação universal de mapeamentos não lineares contínuos. As SVMs e RVMs são métodos baseados em kernel que mapeiam dados não lineares do espaço de características original de entrada para um espaço de características de kernel de maior dimensionalidade e, então, resolvem um problema linear nesse espaço. Neste sentido, esta tese apresenta o desenvolvimento de cartas de controle baseadas em modelos de regressão utilizando Regressão por Vetores de Relevância (RVR), Regressão por Vetores de Suporte (SVR) e RNAs do tipo MLP para o monitoramento de variáveis não conforme às especificações em processos industriais. A escolha dos parâmetros (constante de regularização, função de perda insensível e parâmetros do kernel) apropriados é essencial para a obtenção de modelos RVR e SVR bem ajustados, sendo utilizado um algoritmo de evolução diferencial auto-adaptativo, que é um algoritmo evolutivo para otimização global. Os modelos de regressão e as cartas de controle propostas foram aplicadas no processo produtivo de uma empresa siderúrgica para a modelagem e o monitoramento da proporção da concentração de fósforo nas ligas metálicas produzidas pela empresa. Os modelos de regressão foram avaliados por medidas de diagnósticos e comparados com modelos existentes na literatura. As cartas de controle dos resíduos dos modelos ajustados de acordo com as observações históricas do processo foram utilizadas para o monitoramento do processo. As cartas de controle foram avaliadas por meio de simulação utilizando o método de Monte Carlo para geração das observações necessárias. Os resultados mostraram que os modelos de regressão utilizando as três técnicas de aprendizado de máguina apresentaram desempenho melhor do que os métodos estatísticos analisados e as cartas de controle detectaram as diferentes alterações induzidas. Desta forma, os modelos de regressão e as cartas de controle baseadas em RVR, SVR e RNAs do tipo MLP são adequados para a modelagem e o monitoramento da característica da qualidade do processo produtivo analisado. Além disso, as cartas de controle desenvolvidas podem ser aplicadas no monitoramento de processos industriais em diferentes áreas de aplicação.

Palavras-chave: Controle Estatístico de Processos. Cartas de Controle. Máquinas de Vetores de Relevância. Máquinas de Vetores de Suporte. Redes Neurais Artificiais.

ABSTRACT

Statistical process control (SPC) is one of the most used quality engineering strategies in the industry to monitor the performance of industrial processes. The main tool of SPC is the control chart that is used to monitor the performance of the process variables. Control charts based on regression models are appropriate for monitoring quality characteristics as a function of one or more control variables. In recent years, machine learning techniques, such as artificial neural networks (ANN) and support vector machines (SVM), are being used for process modeling and monitoring, and relevance vector machines (RVM) is used for process modeling. ANN, SVM, and RVM can be used for classification and regression problems, and do not need to assume a specific probability distribution for models implementation. Multilayer perceptron neural networks (MLP) has the capability of universal approximation of continuous nonlinear mappings. SVM and RVM are kernel-based methods that mapping nonlinear data from the original input feature space to a kernel feature space of higher dimensionality and then solving a linear problem in that space. This thesis presents the development of control charts based on regression models using relevance vector regression (RVR), support vector regression (SVR) and MLP for monitoring nonconforming fraction data in industrial processes. The choice of appropriate learning parameters (regularization parameter, insensitive loss function, and kernel parameters) is a crucial step in obtaining well-tuned RVR and SVR models, and in this thesis has been applied a selfadaptive differential evolution algorithm, an evolutionary algorithm for global optimization. The regression models and the proposed control charts were applied in the steelmaking production process for modeling and monitoring phosphorus concentration levels in the metal alloys produced by the company. The regression models were evaluated by diagnostic criteria and compared with models reported in the literature. The residuals control chart from the fitted model according to historical process observations were used for process monitoring. The performance analysis of the control charts was developed using the Monte Carlo simulation method to generate the required observations. The comparative analysis indicates that the regression models using the three machine learning techniques achieved better performance than statistical methods analyzed and the control charts detected the different applied shifts. Thus, regression models and control charts based on RVR, SVR, and MLP are adequate tools for modeling and monitoring the process quality characteristic analyzed. Furthermore, control charts developed can also be useful in the process monitoring in different areas of application.

Keywords: Statistical Process Control. Control Charts. Relevance Vector Machines. Support Vector Machines. Artificial Neural Networks.

LISTA DE FIGURAS

Figura 1 - Sequência metodológica da pesquisa	27
Figura 2 - Carta de controle para o número de não conformes	32
Figura 3 - Carta p para a fração não conforme de um processo fora de controle	36
Figura 4 - Função de perda ε -insensível	52
Figura 5 - Influência do valor de ε no ajuste do modelo SVR	55
Figura 6 - RVR e SVR aplicadas a um problema de regressão	61
Figura 7 - Publicações de RNAs em CEP por ano	65
Figura 8 - Publicações de RNAs em CEP por abordagem	65
Figura 9 - Publicações de SVMs e SVR em CEP por ano	70
Figura 10 - Publicações de SVMs e SVR em CEP por abordagem	70
Figura 11 - Publicações de SVDD em CEP por ano	73
Figura 12 - Fluxograma das etapas de implementação da Fase I e da Fase II das	
cartas de controle baseadas em modelos de regressão	83
Figura 13 - Fluxograma para implementação dos modelos RVR e SVR otimizados	3
pelo algoritmo ED	88
Figura 14 - Box-percentile plot e boxplot para uma distribuição Normal	93
Figura 15 - Histogramas com as distribuições de probabilidade Normal	
sobrepostas para as variáveis de controle	100
Figura 16 - Proporção da concentração de fósforo no final do processo de refino .	101
Figura 17 - (a) Histograma com distribuição Beta sobreposta, (b) gráfico de	
probabilidade normal, (c) gráfico de autocorrelação e (d) gráfico de	
autocorrelação parcial da proporção da concentração de fósforo	102
Figura 18 - Gráficos de probabilidade normal para os resíduos dos modelos para	
o grupo de treinamento	108
Figura 19 - Gráficos de probabilidade normal para os resíduos dos modelos para	
o grupo de teste	109
Figura 20 - Gráficos dos resíduos dos modelos versus a ordem de coleta dos	
dados	110
Figura 21 - Gráficos da função de autocorrelação para os resíduos obtidos com	
o grupo de treinamento	112
Figura 22 - Gráficos da função de autocorrelação para os resíduos obtidos com	
o grupo de teste	113

Figura 23 - Gráficos dos resíduos versus valores estimados pelos modelos114
Figura 24 - Gráficos dos resíduos padronizados versus valores estimados pelos
modelos115
Figura 25 - Gráficos dos valores observados e valores estimados pelos modelos.117
Figura 26 - Gráficos dos valores estimados versus valores observados no
processo118
Figura 27 - Box-percentile plots dos resíduos obtidos com os modelos119
Figura 28 - Box-percentile plots dos resíduos dos modelos RVR, SVR, RNA,
Beta e Modelo 1 à Modelo 8 para os dados do grupo de teste121
Figura 29 - Gráficos dos valores estimados versus observados para os modelos
RVR, SVR, RNA, Beta e Modelo 8122
Figura 30 - Gráficos da variação na proporção da concentração de fósforo em
função da variação da quantidade inicial de fósforo (P^*)123
Figura 31 - Cartas de controle para os resíduos padronizados da Fase I124
Figura 32 - Cartas de controle para o monitoramento do processo na Fase II 126
Figura 33 - Gráficos da função distribuição cumulativa (CDF) para alterações
induzidas na média da característica da qualidade sob controle130
Figura 34 - Box-percentile plots do número de amostras para o processo sob
controle132
Figura 35 - Box-percentile plots do número de amostras para diferentes
alterações induzidas na média da característica da qualidade sob
controle133

LISTA DE TABELAS

Tabela 1 - Principais periódicos com publicações de RNAs em CEP 66
Tabela 2 - Principais periódicos com publicações de técnicas baseadas em
vetores de suporte em CEP73
Tabela 3 - Informações estatísticas das variáveis selecionadas
Tabela 4 - Resultados das simulações para as RNAs103
Tabela 5 - Melhores parâmetros dos modelos SVR 104
Tabela 6 - Melhores parâmetros dos modelos RVR106
Tabela 7 - Informações estatísticas da proporção da concentração de fósforo
estimadas pelos modelos115
Tabela 8 - Resultados dos erros e do coeficiente de determinação para o grupo
de treinamento116
Tabela 9 - Resultados dos erros e do coeficiente de determinação para o grupo
de teste116
Tabela 10 - Modelos de regressão linear múltipla120
Tabela 11 - Resultados dos erros calculados para o grupo de teste120
Tabela 12 - NMA (desvio padrão), SRL e percentis da distribuição do número de
amostras para as cartas de controle dos resíduos padronizados129

LISTA DE ABREVIATURAS E SIGLAS

AICc Coeficiente de Informação de Akaike corrigido (do inglês, correc			
	Akaike Information Criterion)		
AG	Algoritmo Genético		
ART	Teoria da Ressonância Adaptativa (do inglês, Adaptive Resonance		
	Theory)		
CDF	Função Distribuição Cumulativa (do inglês, Cumulative Distribution		
	Function)		
CEP	Controle Estatístico de Processos		
CEQ	Controle Estatístico da Qualidade		
ED	Evolução Diferencial		
ELM	Aprendizado Extremo de Máquina (do inglês, Extreme Machime		
	Learning)		
ERM	Minimização do Risco Empírico (do inglês, Empirical Risk Minimization)		
EWMA	Média Móvel Exponencialmente Ponderada (do inglês, Exponentially		
	Weighted Moving Average)		
GRNN	Rede Neural de Regressão Generalizada (do inglês, General Regression		
	Neural Network)		
i.i.d	independente e identicamente distribuída		
LC	Linha Central		
LIC	Limite Inferior de Controle		
LM	Linha Média		
LSC	Limite Superior de Controle		
LS-SVM	Máquina de Vetor de Suporte por Mínimos Quadrados (do inglês, Least		
	Squares Support Vector Machine)		
LS-SVR	Regressão por Vetor de Suporte por Mínimos Quadrados (do inglês,		
	Least Squares Support Vector Regression)		
LVQ	Quantização Vetorial por Aprendizagem (do inglês, Learning Vector		
	Quantization)		
MAE	Erro Médio Absoluto (do inglês, Mean Absolute Error)		
MAPE	Erro Percentual Absoluto Médio (do inglês, Absolute Percentage Error		
	Medium)		

MEWMA Multivariada de Média Móvel Exponencialmente Ponderada (do ing			
Multivariate Exponentially Weighted Moving Average)			
MLG	Modelo Linear Generalizado		
MLP	Perceptron Multicamadas (do inglês, Multilayer Perceptron)		
MLR	Regressão Linear Múltipla (do inglês, Multiple Linear Regression)		
MQV	Modelo de Quase-Verossimilhança		
MRB	Modelo de Regressão Beta		
MRL	Número mediano de amostras até a presença de um sinal (do inglês,		
	Median Run Length)		
MSE	Erro Médio Quadrático (do inglês, Mean Squared Error)		
NMA	Número médio de amostras até a presença de um sinal		
NMA ₀	Número médio de amostras até a presença de um alarme falso		
NMA ₁	Número médio de amostras até a presença de causas especiais		
PNN	Rede Neural Probabilística (do inglês, Probabilistic Neural Network)		
PSO	Otimização por Enxame de Partículas (do inglês, Particle Swarm		
	Optimization)		
RBF	Função de Base Radial (do inglês, Radial Basis Function)		
RMSE	Raiz do Erro Médio Quadrático (do inglês, Root Mean Squared Error)		
RF	Random Forest		
RNA	Rede Neural Artificial		
RProp	Resilient Propagation		
RS	Rough Set		
RV	Vetor de Relevância (do inglês, <i>Relevance Vector</i>)		
RVM	Máquina de Vetor de Relevância (do inglês, Relevance Vector Machine)		
RVR	Regressão por Vetor de Relevância (do inglês, Relevance Vector		
	Regression)		
SCG	Gradiente Conjugado Escalonado (do inglês, Scaled Conjugate		
	Gradiente)		
SOM	Rede Mapa Auto Organizável (do inglês, Self-Organized Map)		
SQE	Soma dos Quadrados dos Resíduos da Regressão		
SQR	Soma dos Quadrados da Regressão		
SQT	Soma dos Quadrados Total		
SRM	Minimização do Risco Estrutural (do inglês, Structural Risk Minimization)		

SRL	Desvio padrão do número de amostras até a presença de um sinal (do		
	inglês, Standard Error of the Run Length)		
SV	Vetor de Suporte (do inglês, Support Vector)		
SVDD	Descrição de Dados com Vetor de Suporte (do inglês, Support Vector		
	Data Description)		
SVC	Classificação por Vetor de Suporte (do inglês, Support Vector		
	Classification)		
SVM	Máquina de Vetor de Suporte (do inglês, Support Vector Machine)		
SVR	Regressão por Vetor de Suporte (do inglês, Support Vector Regression)		
VC	Vapnik-Chervonenkis		

LISTA DE SÍMBOLOS

b	limiar (<i>bias</i>)
$oldsymbol{b}_L$	limite inferior dos vetores de parâmetros x na ED
$oldsymbol{b}_U$	limite superior dos vetores de parâmetros x na ED
С	constante de regularização na SVR
C_r	taxa de cruzamento na ED
d	grau do kernel polinomial para a SVR e RVR
D	número de parâmetros na ED
$d_{\scriptscriptstyle L}$	limite inferior obtido da Tabela de Durbin-Watson
$d_{_U}$	limite superior obtido da Tabela de Durbin-Watson
d_{w}	estatística de Durbin-Watson
$e_{_{pt}}$	resíduo padronizado do modelo
$-\mathbf{e}_{\rho t}$	média dos resíduos padronizados do modelo
\boldsymbol{e}_t	resíduo do modelo
$\overline{\mathbf{e}}_t$	média dos resíduos do modelo
F	fator de escala na ED
$f(\mathbf{x})$	valor previsto na SVR
g	geração na ED
$k(x_i, x_j)$	função kernel
h	número de variáveis de controle do processo
j	número de neurônios na camada de saída na RNA
J	vetor a ser mutado na ED
k	neurônio na RNA
K	quantidade de vetores diferença usados na ED
1	número de neurônios na camada oculta na RNA
L	mecanismo de cruzamento na ED
т	número de entradas na RNA
n	número de observações
Na	tamanho da amostra
N _{nc}	número de unidades não conformes da amostra
n_{p}	número de parâmetros do modelo

N_{p}	número de vetores da população na ED		
p	fração não conforme		
$\overline{\rho}$	fração média de não conformes		
Р	população na ED		
$p^{^{th}}$	valor do percentil		
p_{o}	média da característica da qualidade de um processo sob controle		
p_1	média da caracterísitica da qualidade de um processo fora de controle		
R	coeficiente de correlação		
R^2	coeficiente de determinação		
$RL_{p^{th}}$	valor do p^m percentil do número de amostras		
t	vetor dos sinais de saída (alvo) na RVR		
t _{min}	valor mínimo do intervalo de normalização		
t_{max}	valor máximo do intervalo de normalização		
$t_{ ho}$	valor normalizado da variável		
t _r	sinal de saída (alvo) na RVR		
u	parâmetro para os kernels polinomial e sigmoidal para a SVR e RVR		
Uk	saída do combinador linear devido aos sinas de entrada na RNA		
v	vetor mutante na ED		
V_d	variável dependente		
V_i	variável independente		
w	vetor de pesos no modelo RVR		
x	sinal de entrada		
\overline{x}	média aritmética do sinal de entrada		
x	indivíduo da população na ED		
X _{min}	valor mínimo dos dados do sinal de entrada		
X _{min} X _{max}	valor mínimo dos dados do sinal de entrada valor máximo dos dados do sinal de entrada		
X _{min} X _{max} X _r	valor mínimo dos dados do sinal de entrada valor máximo dos dados do sinal de entrada vetor dos sinais de entrada na RVR		
X _{min} X _{max} X _r Y	valor mínimo dos dados do sinal de entrada valor máximo dos dados do sinal de entrada vetor dos sinais de entrada na RVR sinal de saída		
X _{min} X _{max} X _r Y y y	valor mínimo dos dados do sinal de entrada valor máximo dos dados do sinal de entrada vetor dos sinais de entrada na RVR sinal de saída média aritmética do sinal de saída		
$\begin{array}{l} \mathbf{x}_{min} \\ \mathbf{x}_{max} \\ \mathbf{x}_{r} \\ \mathbf{y} \\ \overline{\mathbf{y}} \\ \overline{\mathbf{y}} \\ \mathbf{y}_{i} \end{array}$	valor mínimo dos dados do sinal de entrada valor máximo dos dados do sinal de entrada vetor dos sinais de entrada na RVR sinal de saída média aritmética do sinal de saída valor observado no processo		
X _{min} X _{max} X _r Y <u>Y</u> Y _i Y _k	valor mínimo dos dados do sinal de entrada valor máximo dos dados do sinal de entrada vetor dos sinais de entrada na RVR sinal de saída média aritmética do sinal de saída valor observado no processo sinal de saída do neurônio <i>k</i> na RNA		
x_{min} x_{max} x_{r} y \overline{y} \overline{y} y_{i} y_{k} y_{t}	valor mínimo dos dados do sinal de entrada valor máximo dos dados do sinal de entrada vetor dos sinais de entrada na RVR sinal de saída média aritmética do sinal de saída valor observado no processo sinal de saída do neurônio <i>k</i> na RNA valor estimado pelo modelo		

α	probabilidade de ocorrer um erro do tipo l		
α_i	variável de Lagrange		
α_r	parâmetro associado a cada peso ω_r na RVR		
α	vetor de parâmetros α_r na RVR		
β	probabilidade de ocorrer um erro do tipo II		
Е	função de perda (<i>loss function</i>) na SVR		
ϵ_{r}	ruído gaussiano com média zero e variância $\sigma_{ m r}^2$		
$\phi(\mathbf{x}_r)$	função de base na RVR		
Φ	matriz das funções de base na RVR		
γ	controla a flexibilidade da função dos kernels na SVR e RVR		
$\varphi(.)$	função de ativação na RNA		
λ	largura dos limites de controle da carta de controle		
μ	média		
μ_r	<i>r-</i> ésimo componente do vetor média na RVR		
μ	vetor média na RVR		
$ au_1$	probabilidades de ajustar o fator <i>F</i> na ED		
$ au_2$	probabilidades de ajustar o fator <i>Cr</i> na ED		
σ	desvio padrão		
$\sigma_{\scriptscriptstyle e_t}$	desvio padrão dos resíduos		
$\sigma_{\scriptscriptstyle e_{\scriptscriptstyle pt}}$	desvio padrão dos resíduos padronizados		
σ^{2}	variância		
$\hat{\sigma}^2$	variância estimada		
σ_r^2	variância do ruído na RVR		
ν	largura dos kernels na SVR e RVR		
ω	peso no modelo SVR		
ω_{kj}	peso do neurônio <i>k</i> na RNA		
ω_r	peso no modelo RVR		
ξi	variável de folga na SVR		
х	espaço de entrada na SVR		
Σ	matriz de covariância na RVR		

SUMÁRIO

1	INTRODUÇÃO	.18
1.1	MOTIVAÇÃO	.22
1.2	OBJETIVOS	.23
1.2.1	Objetivo geral	.23
1.2.2	Objetivos específicos	.24
1.2.3	Contribuições da tese	.24
1.3	JUSTIFICATIVA	.25
1.4	METODOLOGIA DA PESQUISA	.26
1.5	ESTRUTURA DO DOCUMENTO	.28
2	CONTROLE ESTATÍSTICO DA QUALIDADE	.29
2.1	CONTROLE ESTATÍSTICO DE PROCESSOS	.29
2.2	CARTAS DE CONTROLE	.30
2.2.1	Cartas de controle para atributos	.33
2.2.2	Cartas de controle baseadas em modelos de regressão	.37
3	APRENDIZADO DE MÁQUINA	.40
3.1	TEORIA DO APRENDIZADO SUPERVISIONADO	.41
3.2	REDES NEURAIS ARTIFICIAIS	.43
3.3	MÁQUINAS DE VETORES DE SUPORTE	.48
3.3.1	Regressão por vetores de suporte	.49
3.3.2	Funções kernel	.53
3.3.3	Parâmetros para a regressão por vetores de suporte	.54
3.4	MÁQUINAS DE VETORES DE RELEVÂNCIA	.57
3.5	REDES NEURAIS ARTIFICIAIS E MÁQUINAS DE VETORES DE	
	SUPORTE EM CONTROLE ESTATÍSTICO DE PROCESSOS	.62
3.5.1	Aplicações de redes neurais artificiais em CEP	.63
3.5.2	Aplicações de técnicas baseadas em vetores de suporte em CEP	.68
3.6	SELEÇÃO DOS PARÂMETROS PARA REGRESSÃO POR VETORES DE	
	SUPORTE E REGRESSÃO POR VETORES DE RELEVÂNCIA	.74
3.6.1	Métodos para seleção dos parâmetros para SVR e RVR	.74
3.6.2	Evolução diferencial	.75
4	MÉTODO PROPOSTO	.81

4.1	FASE I – ANÁLISE RETROSPECTIVA	82	
4.1.1	Coleta, análise e preparação dos dados82		
4.1.2	Desenvolvimento dos modelos de regressão	85	
4.1.3	Medidas de diagnóstico dos modelos de regressão	87	
4.1.4	Desenvolvimento das cartas de controle	94	
4.2	FASE II – MONITORAMENTO DO PROCESSO	95	
4.3	ANÁLISE DE SENSIBILIDADE DAS CARTAS DE CONTROLE	95	
5	ESTUDO EXPERIMENTAL	98	
5.1	MODELOS BASEADOS EM REDES NEURAIS ARTIFICIAIS	102	
5.2	MODELOS BASEADOS EM REGRESSÃO POR VETORES DE		
	SUPORTE	103	
5.3	MODELOS BASEADOS EM REGRESSÃO POR VETORES DE		
	RELEVÂNCIA	105	
5.4	ANÁLISE DOS MODELOS DE REGRESSÃO	107	
5.5	IMPLEMENTAÇÃO DAS CARTAS DE CONTROLE NA FASE I	123	
5.6	IMPLEMENTAÇÃO DAS CARTAS DE CONTROLE NA FASE II	125	
5.7	ANÁLISE DE SENSIBILIDADE DAS CARTAS DE CONTROLE	126	
6	CONCLUSÃO	134	
	REFERÊNCIAS	136	

1 INTRODUÇÃO

O Controle Estatístico da Qualidade (CEQ) diz respeito ao uso de métodos estatísticos no monitoramento e na manutenção da qualidade nos processos e produtos. O CEQ inclui, principalmente, as áreas de controle estatístico de processos, planejamento de experimentos, análise de capacidade e amostragem de aceitação (SAMOHYL, 2009).

O Controle Estatístico de Processos (CEP) é utilizado para monitorar a qualidade dos produtos em diversos processos industriais, devido principalmente a sua capacidade em distinguir entre causas naturais e causas especiais de variação. As causas naturais são variações aleatórias inerentes ao processo enquanto as causas especiais são aquelas resultantes de um modo de operação anormal do processo (SAMOHYL, 2009).

Segundo Montgomery (2004) a engenharia da qualidade é o conjunto de atividades operacionais, de gerenciamento e de engenharia que uma empresa utiliza para garantir que as características da qualidade de um produto estejam nos níveis nominais ou exigidos. Werkema (1995) define um processo como uma combinação de elementos, equipamentos, insumos, métodos ou procedimentos, condições ambientais, pessoas e informações do processo ou medidas, tendo como objetivo a fabricação de um bem ou o fornecimento de um serviço.

A definição de Taguchi expressa que um produto ou serviço de qualidade é aquele que atende perfeitamente às especificações, atingindo o valor desejado com a menor variância possível em torno dele. Os elementos que, em conjunto, descrevem a adequação de um produto ao uso são denominados características da qualidade ou indicadores de desempenho (RIBEIRO; CATEN, 2012).

A carta ou gráfico de controle é a principal ferramenta do CEP e possui o objetivo de monitorar o processo e identificar a ocorrência de causas especiais permitindo que ações corretivas sejam realizadas antes que muitas unidades defeituosas sejam produzidas. As cartas de controle também fornecem informações para que o operador faça um diagnóstico sobre o processo, podendo implantar mudanças que melhorem seu desempenho (SAMOHYL, 2009).

De acordo com Montgomery (2004) existem cinco razões principais para a popularidade das cartas de controle: contribuem para aumentar a produtividade, pois

reduzem desperdícios e diminuem custos; são efetivas na prevenção de defeitos; previnem ajustes desnecessários no processo, pois distinguem as causas de variabilidade naturais das especiais; fornecem informações para o diagnóstico sobre o desempenho do processo e quanto à capacidade do processo.

As cartas de controle tradicionais de Shewhart requerem que os dados monitorados sejam independentes e identicamente distribuídos (i.i.d.) em torno de um modelo de referência com média constante (MONTGOMERY, 2004). Entretanto, é frequente observar processos onde as características da qualidade variam em função de alterações do ajuste das variáveis de controle do processo, gerando assim um modelo de referência distinto para cada novo ajuste. Uma alternativa proposta na literatura são as cartas de controle baseadas em modelos de regressão (PSARAKIS; PAPALEONIDA, 2007).

A análise de regressão consiste na modelagem e investigação da relação entre as características da qualidade e as variáveis de controle do processo (MONTGOMERY; RUNGER, 2003). As cartas de controle baseadas em modelos de regressão monitoram os resíduos entre as características da qualidade observadas e as características da qualidade previstas pelo modelo de regressão (MANDEL, 1969). Os resíduos são definidos como sendo a diferença entre os valores observados e os valores previstos pelo modelo.

O monitoramento das características da qualidade utilizando cartas de controle baseadas em modelos apresenta algumas vantagens: a operacionalização da carta é relativamente simples, identifica a influência das variáveis de controle, o controle estatístico é baseado nos resíduos do modelo de regressão que usualmente são não correlacionados no tempo, e permite monitorar alterações na média da característica da qualidade e nos coeficientes das variáveis de controle (SANT'ANNA, 2009).

O ajuste inadequado do modelo de regressão pode afetar o desempenho das cartas de controle no monitoramento do processo (JEARKPAPORN *et al.*, 2005) e, se os resíduos obtidos com o modelo de regressão não apresentam distribuição aproximadamente normal e variâncias iguais, precisam ser utilizados métodos de transformação dos dados das características da qualidade como, por exemplo, a transformação de Box-Cox (BOX; COX, 1964).

As cartas de controle podem ser classificadas em dois tipos: cartas de controle para variáveis, que monitoram características da qualidade em escala numérica, e cartas de controle para atributos, que monitoram características da qualidade por comparação com um padrão de referência, que pode ser o número de não conformes (fração defeituosa ou fração não conforme) ou o número de não conformidades (defeitos) existentes na amostra, pressupondo distribuição Binomial e Poisson, respectivamente. As características da qualidade do tipo fração não conforme são observações expressas no intervalo [0,1].

A maioria dos métodos aplicados em CEP é paramétrica na medida em que fazem suposições sobre as propriedades de distribuição e estrutura de autocorrelação dos parâmetros do processo sob controle. Gujarati (2006) apresenta que a autocorrelação pode ser definida como a correlação entre integrantes de séries de observações ordenadas no tempo ou no espaço.

Segundo Sant'Anna (2009), as cartas de controle baseadas em modelos de regressão linear não são adequadas quando se monitoram características da qualidade do tipo fração de produtos não conformes, pois pressupõem que a característica da qualidade possui distribuição Normal e possibilitam a previsão de valores fora do limite do intervalo [0,1]. Conforme Sant'Anna (2009), os modelos pertencentes à classe dos Modelos Lineares Generalizados (MLGs) e o Modelo de Regressão Beta (MRB) são alternativas para a modelagem de variáveis com distribuição não Normal.

Nos Modelos de Quase-Verossimilhança (MQVs), pertencentes à classe dos MLGs, as variáveis são consideradas independentes sem ser necessário especificar a distribuição de probabilidade para a característica da qualidade (CORDEIRO; DEMÉTRIO, 2008) e no MRB supõe-se que a fração não conforme segue a distribuição Beta, que assume valores restritos ao intervalo [0,1] (FERRARI; CRIBARI-NETO, 2004). Sant'Anna e Caten (2012) propuseram uma carta de controle baseada em MRB para monitorar processos com características da qualidade do tipo fração de produtos não conformes.

Cartas de controle não paramétricas baseadas em técnicas de aprendizado de máquina têm sido propostas na literatura, conforme apresentado na seção 3.4 desta tese. A carta de controle é não paramétrica na medida em que não é necessário assumir uma distribuição de probabilidade específica para a sua implementação (CAMCI *et al.*, 2008). O aprendizado de máquina trata do desenvolvimento de técnicas computacionais sobre o aprendizado e a construção de sistemas capazes de aprender

e melhorar seu desempenho baseado em experiências acumuladas através da solução de problemas anteriores (MITCHELL, 1997).

As técnicas de aprendizado de máquina utilizadas nesta tese são as Máquinas de Vetores de Relevância (*Relevance Vector Machine*, RVM), as Máquinas de Vetores de Suporte (*Support Vector Machine*, SVM) e as Redes Neurais Artificiais (RNAs).

As RNAs são modelos de processamento de dados que são inspirados no que se conhece das redes de neurônios biológicos. Segundo Haykin (2009), a principal propriedade de uma RNA é a sua habilidade de aprender a partir de seu ambiente e de melhorar o seu desempenho através do aprendizado. As RNAs podem ser utilizadas para problemas de classificação de padrões e para problemas de regressão. Das técnicas de aprendizado de máquina, a RNA do tipo Perceptron Multicamadas (*Multilayer Perceptron*, MLP) é uma das mais aplicadas em CEP (ATASHGAR, 2015).

As SVMs são fundamentadas na Teoria de Aprendizado Estatístico ou teoria VC (Vapnik-Chervonenkis), desenvolvida por Vapnik em 1995 visando a proposição de técnicas de aprendizado de máquina que buscam maximizar a capacidade de generalização (VAPNIK, 1998). A aplicação de SVMs foi originalmente desenvolvida para problemas de classificação de padrões, sendo posteriormente estendida para o tratamento de problemas de regressão e denominada Regressão por Vetores de Suporte (*Support Vector Regression*, SVR).

As RVMs são modelos probabilísticos baseados em métodos kernel com forma funcional similar as SVMs (TIPPING, 2000). As RVMs possuem a propriedade de soluções esparsas, proporcionando um desempenho mais rápido sob uma mesma base de testes e mantendo o poder de generalização dos dados comparável às SVMs. As RVMs podem ser utilizadas para problemas de classificação de padrões e para problemas de regressão, sendo denominada de Regressão por Vetores de Relevância (*Relevance Vector Regression*, RVR).

As RNAs, SVMs e RVMs não necessitam do conhecimento sobre a distribuição de probabilidade dos dados para a implementação dos modelos. Estas três técnicas de aprendizado de máquina foram selecionadas nesta tese pois, em nossas pesquisas nas bases de dados: (i) não foram encontrados trabalhos empregando RVR no desenvolvimento de cartas de controle, sendo inédita a aplicação de RVR nesta aplicação de CEP; (ii) não foram encontrados trabalhos empregando SVR no monitoramento de processos que mensuram variáveis não conformes às

especificações, sendo inédita a aplicação de RVR e SVR nestes processos; e (iii) as RNAs são uma das mais utilizadas técnicas de aprendizado de máquina em cartas de controle e estão disponíveis na literatura algumas aplicações de RNAs no monitoramento de processos com características da qualidade em atributos.

Em geral, a escolha dos parâmetros apropriados é um passo crucial na utilização de RVMs e SVMs para classificação de padrões e para regressão. Na RVR deve-se selecionar o tipo de função kernel e seus parâmetros. A capacidade de generalização da SVR depende da seleção dos parâmetros constante de regularização (C) e função de perda insensível (\mathcal{E}), além da seleção do tipo de função kernel e seus parâmetros.

Nas últimas décadas, algoritmos evolutivos foram propostos para aplicações de otimização. Evolução diferencial (ED) é uma variante dos algoritmos evolutivos proposta por Storn e Price (1997). O potencial da ED inclui sua estrutura simples, a facilidade de uso, a propriedade de convergência e, também, a robustez (SANTOS *et al.*, 2012). Nesta tese, um algoritmo de ED foi utilizado para a seleção dos parâmetros apropriados da RVR e da SVR.

1.1 MOTIVAÇÃO

As RNAs são aplicadas em CEP desde o final da década de 1980. De acordo com Psarakis (2011), a principal razão para a aplicação de RNAs na detecção de alterações da média e da variância em processos é automatizar a interpretação das cartas de controle. É essencial para as indústrias detectarem a ocorrência dessas alterações rapidamente, para que suas causas possam ser encontradas e as ações corretivas necessárias sejam tomadas antes que uma grande quantidade de produtos não conformes seja produzida.

As SVMs são aplicadas em CEP desde os anos 2000. Segundo Camci *et al.* (2008) existem muitas razões para propor cartas de controle baseadas em princípios SVM: (i) é um sistema de aprendizado de máquina eficiente no espaço de características induzidas kernel, (ii) controlam com sucesso a flexibilidade do espaço de características induzidas kernel através da teoria da generalização, e (iii) exploram

a teoria da otimização, além de possuírem uma importante característica que é utilizar o modelo dual.

Em nossas pesquisas nas bases de dados foi verificado que, nos últimos anos, estão sendo aplicadas abordagens empregando RNAs, SVMs e SVR para o reconhecimento de padrões, o monitoramento de processos industriais para a detecção de alterações da média e da variância e para o desenvolvimento de cartas de controle porém, não foram encontrados trabalhos empregando RVMs e RVR nestas aplicações de CEP. Algumas aplicações de RNAs no monitoramento de processos que mensuram variáveis não conformes às especificações estão disponíveis na literatura porém, não foram encontrados trabalhos utilizando SVR no monitoramento de processos que mensuram variáveis não conformes às especificações.

Nesta pesquisa são aplicadas RVR, SVR e RNAs para a modelagem e o desenvolvimento de cartas de controle destinadas ao monitoramento de processos que mensuram variáveis não conformes às especificações.

1.2 OBJETIVOS

Nesta seção são apresentados o objetivo geral da tese, os objetivos específicos e as contribuições da tese.

1.2.1 Objetivo geral

Desenvolver cartas de controle baseadas em modelos de regressão utilizando Regressão por Vetores de Relevância, Regressão por Vetores de Suporte e Redes Neurais Artificiais do tipo Perceptron Multicamadas para o monitoramento de processos que mensuram variáveis não conformes às especificações do tipo proporção.

1.2.2 Objetivos específicos

Os objetivos específicos da tese são os seguintes:

- Descrever as aplicações de Máquinas de Vetores de Suporte e Redes Neurais Artificiais em Controle Estatístico de Processos.
- Aplicar um algoritmo de otimização denominado Evolução Diferencial para a seleção dos parâmetros dos modelos baseados em Regressão por Vetores de Relevância e Regressão por Vetores de Suporte.
- Desenvolver a modelagem aplicando Regressão por Vetores de Relevância, Regressão por Vetores de Suporte e Redes Neurais Artificiais do tipo Perceptron Multicamadas em processos que mensuram variáveis não conformes às especificações do tipo proporção.
- Desenvolver as cartas de controle baseadas em modelos de Regressão por Vetores de Relevância, Regressão por Vetores de Suporte e Redes Neurais Artificiais do tipo Perceptron Multicamadas para o monitoramento de processos que mensuram variáveis não conformes às especificações do tipo proporção.

1.2.3 Contribuições da tese

Alinhado com os objetivos da tese, levantou-se a seguinte hipótese: A utilização de RVR, SVR e RNAs do tipo MLP é capaz de fornecer resultados eficientes na modelagem e no monitoramento utilizando cartas de controle em processos que mensuram variáveis não conformes às especificações?

Como principais contribuições desta tese destacam-se a aplicação inédita de RVR no desenvolvimento de cartas de controle baseadas em modelos de regressão e a aplicação inédita de RVR e SVR no monitoramento de processos que mensuram variáveis não conformes às especificações utilizando cartas de controle.

1.3 JUSTIFICATIVA

As cartas de controle tradicionais usadas para monitorar processos e determinar quando o processo está ou não sob controle, em muitos casos, não conseguem detectar e classificar as causas especiais existentes. Nestes casos, normalmente são identificadas as mudanças nas medidas utilizadas para a observação do processo, principalmente a média e a variância.

Quando um processo está sob controle estatístico, as amostras em uma carta de controle seguem um padrão aleatório ou natural, onde as amostras possuem um comportamento aleatório, seguem uma distribuição de probabilidade específica, nenhuma amostra está fora dos limites de controle da carta e as amostras não apresentam ciclos ou tendências. Se as amostras em uma carta de controle possuem um padrão não aleatório ou anormal, onde as amostras seguem tendências, ciclos ou configurações dos pontos como as descritas em Western Electric (1958), representam que o processo pode possuir causas especiais mesmo se as amostras estiverem dentro dos limites de controle.

As cartas de controle tradicionais assumem que os dados são independentes e, desta forma, apresentam alarmes falsos quando os processos possuem dados autocorrelacionados. Estes problemas levaram ao desenvolvimento de abordagens para o monitoramento de processos univariados e multivariados empregando técnicas de aprendizado de máquina. Destaca-se ainda a possibilidade de integração entre as técnicas estatísticas e de aprendizado de máquina.

A utilização de técnicas de aprendizado de máquina se apresenta como uma opção para a modelagem e o desenvolvimento de cartas de controle pois, em casos apresentados na literatura, essas técnicas apresentam desempenho superior as técnicas estatísticas utilizadas para a mesma finalidade (PSARAKIS, 2011; ATASHGAR, 2015).

A escolha do tema também se justifica por sua relevância prática. Com a crescente informatização dos processos industriais, tem-se verificado um aumento na quantidade de informações disponíveis sobre as variáveis de processos. Em muitos casos, as informações descrevem a variação, não normalidade e correlação dessas variáveis. A análise das variáveis de processo por técnicas de aprendizado de

máquina permite a caracterização do fenômeno em estudo visando a melhoria efetiva na qualidade dos produtos e o aumento na competitividade das empresas.

1.4 METODOLOGIA DA PESQUISA

O método de pesquisa adotado caracteriza-se como sendo de natureza aplicada, possuindo uma abordagem quantitativa.

Segundo Silva e Menezes (2005) e Appolinário (2012) um trabalho pode ser classificado como sendo uma pesquisa básica ou aplicada. Uma pesquisa aplicada objetiva gerar conhecimentos dirigidos à solução de problemas específicos. Uma pesquisa com abordagem quantitativa é focada em análises numéricas e utiliza analises estatísticas e computacionais (GIL, 2002).

De acordo com seus objetivos, pode ser classificada como pesquisa exploratória, onde se deseja obter melhor conhecimento e aprofundamento da situação a ser pesquisada (SILVA e MENEZES, 2005). Segundo Oliveira (1999) esse tipo de pesquisa desenvolve estudos que permitem uma visão geral do fato ou fenômeno estudado, através da delimitação do estudo, levantamento bibliográfico, leitura e análise de documentos.

As etapas desenvolvidas na pesquisa foram:

- (i) consolidação do levantamento bibliográfico sobre os temas abordados;
- (ii) desenvolvimento das modelagens e das cartas de controle propostas;
- (iii) validação das modelagens e das cartas de controle desenvolvidas por meio de aplicações reais e de simulações;
- (iv) comparação dos modelos e das cartas de controle propostas com técnicas apresentadas na literatura.

A Figura 1 apresenta a sequência metodológica desta pesquisa, cujo objetivo principal foi a aplicação de RVR, SVR e RNAs na modelagem e no monitoramento de processos que mensuram variáveis não conformes às especificações.

A etapa inicial foi a revisão da literatura sobre Controle Estatístico da Qualidade, Controle Estatístico de Processos, Cartas de Controle, Redes Neurais Artificiais, Máquinas de Vetores de Suporte, Máquinas de Vetores de Relevância e Evolução Diferencial.

Após esta revisão, foi realizado o desenvolvimento da modelagem e das cartas de controle utilizando RVR, SVR e RNAs e sua aplicação por meio de estudo de casos e simulações foi realizada na sequência. A seguir foi realizada a avaliação dos resultados dos métodos propostos e a comparação com técnicas disponíveis na literatura e, finalizando a tese, foi apresentada a discussão dos resultados obtidos e a identificação de novas oportunidades para a pesquisa.





Fonte: a autora, 2016.

1.5 ESTRUTURA DO DOCUMENTO

Este texto está organizado em capítulos, sendo que o Capítulo 1 consistiu na apresentação da introdução, objetivos, contribuições da tese, justificativas e metodologia da pesquisa. No Capítulo 2 foram apresentados os conceitos sobre Controle Estatístico da Qualidade, Controle Estatístico de Processos e Cartas de Controle. No Capítulo 3 foram apresentados os conceitos sobre Redes Neurais Artificiais, Máquinas de Vetores de Suporte, Máquinas de Vetores de Relevância e Evolução Diferencial, além das aplicações de Redes Neurais Artificiais e Máquinas de Vetores de Suporte em Controle Estatístico de Processos. Após, o Capítulo 4 apresentou a metodologia de desenvolvimento da modelagem e das cartas de controle baseadas em modelos de regressão utilizando RVR, SVR e RNAs. O Capítulo 5 apresentou a aplicação desta metodologia em um processo industrial, a comparação com técnicas apresentadas na literatura, as simulações e a análise dos resultados. O Capítulo 6 apresentou as considerações finais e algumas das novas oportunidades para a pesquisa.

2 CONTROLE ESTATÍSTICO DA QUALIDADE

Este capítulo apresenta a revisão da literatura necessária para o desenvolvimento da tese englobando Controle Estatístico da Qualidade, Controle Estatístico de Processos e Cartas de Controle.

O conceito de qualidade foi definido de diferentes maneiras na literatura. Mitra (1998) define que qualidade de um produto ou serviço é a adequação do produto ou serviço em atender ou exceder as necessidades de uso solicitadas pelo consumidor. Para Juran e Godfrey (1999) qualidade significa adequação ao uso. Montgomery (2004) prefere a definição de que qualidade é inversamente proporcional à variabilidade e que melhoria da qualidade é a redução da variabilidade nos processos e produtos. Para Samohyl (2009) o conceito da qualidade se apoia essencialmente na sua operacionalidade e sua definição requer a mensuração de características concretas em termos quantitativos.

Os métodos estatísticos e sua aplicação na melhoria da qualidade começaram a ser utilizados por Walter Andrew Shewhart, na década de 1920, ao desenvolver e aplicar as cartas ou gráficos de controle no monitoramento de processos nos *Bell Telephone Laboratories*. Este é considerado o começo formal do Controle Estatístico de Processos (SHEWHART, 1926; COSTA *et al.*, 2010).

2.1 CONTROLE ESTATÍSTICO DE PROCESSOS

O Controle Estatístico de Processos (CEP) pode ser definido como um conjunto de ferramentas de resolução de problemas utilizadas para planejar, monitorar e aprimorar um processo produtivo, sendo úteis na obtenção da estabilidade e na melhoria da capacidade do processo através da redução da variabilidade nas características da qualidade de interesse (RIBEIRO; CATEN, 2012).

As sete principais ferramentas utilizadas no CEP são: o histograma ou diagrama de ramos e folhas, a folha de controle, o gráfico de Pareto, o diagrama de causa e efeito, o diagrama de concentração de defeito, o diagrama de dispersão e a carta ou gráfico de controle (MONTGOMERY, 2004; RAMOS *et al.*, 2013).

Os processos produtivos estão sujeitos a variabilidade, que está relacionada com as diferenças existentes entre as unidades produzidas. Segundo Costa *et al.* (2010), conforme o valor da variabilidade do processo, as diferenças entre as unidades produzidas serão fáceis ou difíceis de observar. Para o gerenciamento do processo e redução da variabilidade é importante analisar os motivos da variabilidade no processo. A variabilidade pode ser ocasionada por causas naturais e por causas especiais.

As variações aleatórias inerentes ao processo são denominadas de causas naturais ou causas aleatórias e são resultantes de uma série de perturbações que não podem ser eliminadas. Essa variabilidade representa o padrão natural do processo. Quando o processo apresenta apenas a variabilidade natural atuando diz-se que está no modo de controle estatístico ou sob controle (LOUZADA *et al.*, 2013).

As causas especiais, também denominadas causas assinaláveis ou causas atribuíveis, são aquelas resultantes de um modo de operação anormal do processo e que podem prejudicar a qualidade do produto manufaturado. Essas causas fazem com que o processo saia fora de seu padrão natural de operação e possuem o efeito de deslocar a distribuição da variável aleatória e/ou aumentar sua dispersão. As causas especiais devem ser identificadas e podem ser corrigidas ou eliminadas. Quando, além das causas naturais de variabilidade, as causas especiais estiverem presentes diz-se que o processo está fora de controle estatístico (OAKLAND, 2003; MONTGOMERY, 2004; COSTA *et al.*, 2010).

A distinção entre as duas causas de variabilidade do processo é fundamental visto que as causas especiais não fazem parte do processo e devem ser investigadas e eliminadas. As cartas de controle são ferramentas gráficas do CEP que permitem distinguir esses dois tipos de variabilidade, uma vez que estabelecem limites para as variações devido as causas naturais que possam ocorrer em um processo (LOUZADA *et al.*, 2013).

2.2 CARTAS DE CONTROLE

Segundo Tôrres (2015), a carta de controle é empregada sistematicamente para o monitoramento de uma ou mais variáveis (características da qualidade) que

estão direta ou indiretamente relacionadas com o processo de produção. Na construção de uma carta de controle devem ser consideradas a estatística de interesse e os limites de controle. A carta de controle é uma representação gráfica de valores, que ilustram a evolução da estatística de interesse ao longo do tempo, referente a uma determinada característica da qualidade. Os limites de controle são geralmente calculados com base na distribuição de probabilidade desta estatística e sobre as taxas de alarmes falsos definidos pelo usuário.

As cartas de controle tradicionais podem ser classificadas em duas categorias (MITRA, 1998; COSTA *et al.*, 2010):

- cartas de controle para variáveis, que monitoram características da qualidade em escala numérica, pressupondo distribuição Normal; e
- cartas de controle para atributos, que monitoram características da qualidade por comparação com um padrão de referência: número de não conformes (defeituosos) e número de não conformidades (defeitos), pressupondo distribuição Binomial e Poisson, respectivamente.

Os métodos estatísticos para analisar variáveis estão dispostos em dois grupos: um que trata da estatística para as variáveis de maneira isolada, a estatística univariada, e outro que trata as variáveis de forma conjunta, a estatística multivariada. Desta forma, com base no número de variáveis envolvidas, as cartas de controle podem ser classificadas como univariadas ou multivariadas. As cartas de controle univariadas são utilizadas para monitorar uma única característica da qualidade, enquanto as cartas multivariadas permitem monitorar conjuntamente uma série de características da qualidade, que podem estar correlacionadas umas com a outras (VICINI, 2005; TÔRRES, 2015).

Para a aplicação das cartas de controle tradicionais de Shewhart as características da qualidade devem ser independentes e identicamente distribuídas (i.i.d.), ou seja, as variáveis não podem ser autocorrelacionadas (COSTA *et al.*, 2010). Gujarati (2006) apresenta que a autocorrelação pode ser definida como a correlação entre integrantes de séries de observações ordenadas no tempo ou no espaço. Quando existe autocorrelação nos dados, o uso das cartas de controle não exibe bom desempenho podendo fornecer alarmes falsos, ou seja, indicar que o processo está fora de controle quando na verdade continua estável (RAMOS *et al.*, 2013).

Além disso, as cartas de controle tradicionais devem possuir distribuição de probabilidade Normal. Em algumas ocasiões, a suposição de normalidade pode ser moderadamente violada sem afetar os resultados (RAMOS *et al.*, 2013).

Uma carta de controle possui três linhas paralelas ao eixo das abcissas que representam a média ou alvo do processo, denominada Linha Central (LC) ou Linha Média (LM), e os limites de controle, denominados Limite Superior de Controle (LSC) e Limite Inferior de Controle (LIC). Os limites de controle da carta definem a sua região de ação e são baseados em estatísticas como média amostral, mediana, desvio-padrão, entre outras. A Figura 2 apresenta um exemplo de carta de controle para o número de carcaças não conformes observadas em um processo de fabricação de carcaças de tratores (MONTGOMERY, 2004; COSTA *et al.*, 2010).





A verificação se o processo está ou não sob controle é feito pela coleta periódica das observações. Para um processo sob controle os pontos amostrais devem estar entre os limites de controle e devem ser distribuídos segundo padrões aleatórios, como o apresentado na Figura 2. A indicação básica de falta de controle estatístico é a ocorrência de pontos fora dos limites de controle, indicando que provavelmente o processo foi afetado por alguma causa especial (MONTGOMERY, 2004; RAMOS *et al.*, 2013; LOUZADA *et al.*, 2013).

No entanto, mesmo com todos os pontos entre os limites de controle, a presença de tendências, ciclos ou alguma outra configuração típica dos pontos pode

revelar falta de controle estatístico, ou seja, indicar perda da aleatoriedade do processo e a presença de causas especiais (JURAN; GODFREY, 1999; MONTGOMERY, 2004). Existem diversas regras que tentam identificar padrões formados em cartas de controle, sendo uma das mais mencionadas e aplicadas a regra de Western Electric (1958), onde são listados um conjunto de 15 padrões não aleatórios.

Em um processo sob controle em que a característica da qualidade do conjunto de dados possua uma distribuição Normal, o intervalo da distribuição corresponde a 99,7% quando usado $\mu \pm 3\sigma$, onde μ representa a média e σ o desvio padrão. A faixa de variabilidade normal é a do intervalo $\mu - 3\sigma$ e $\mu + 3\sigma$. A escolha dos limites de controle de $\pm 3\sigma$ leva em conta aspectos econômicos e o erro do tipo I e o erro do tipo II (SIQUEIRA, 1997).

O erro do tipo I ocorre quando se procura por uma causa especial de variação que, na realidade, não está presente no processo. Esse erro representa um alarme falso, ou seja, são pontos que a carta assinala como estando fora dos limites de controle, quando na verdade não deveriam estar. A probabilidade de ocorrer um erro do tipo I é denominado de (risco) α (WERKEMA, 1995; SIQUEIRA, 1997).

O erro do tipo II acontece quando se assume que uma causa natural está presente quando, na realidade, existe uma causa especial de variação no processo. Esse erro ocorre quando o processo está fora de controle e a carta de controle não sinaliza esta situação, ou seja, os pontos continuam sendo assinalados dentro dos limites de controle quando deveriam estar além deles. A probabilidade de cometer um erro do tipo II é denominada de (risco) β (WERKEMA, 1995; SIQUEIRA, 1997).

2.2.1 Cartas de controle para atributos

Em um processo industrial pode ser definido um conjunto de causas ou fatores que produzem determinado efeito sobre uma ou mais características da qualidade do processo ou produto. Certas causas são inerentes ao processo e, desta forma, não é possível controlar todas as causas de variação do processo. As causas de variação que podem ser controladas podem resultar na produção de produtos com características da qualidade não conformes às especificações desejadas (MONTGOMERY, 2004).

Os atributos são características da qualidade que são comparadas com um certo padrão (especificações) e podem assumir valores discretos, como a classificação como conforme ou não-conforme, ou uma certa contagem de defeitos (RIBEIRO; CATEN, 2012; SAMOHYL, 2009).

Nas cartas de controle para atributos existem duas categorias (OAKLAND, 2003; RAMOS *et al.*, 2013):

- a) as que classificam os itens em conformes ou não conformes, como é o caso da carta da fração defeituosa ou fração não conforme (carta *p*) e da carta do número de unidades não conformes (carta *np*). Estas cartas seguem a distribuição Binomial.
- b) as que consideram o número de defeitos ou não conformidades existentes na amostra, como a carta do número de não conformidades (carta *c*) e a carta para o número de não conformidades por unidade (carta *u*). Estas cartas seguem a distribuição Poisson.

As características da qualidade do tipo fração são observações expressas no intervalo [0,1] e podem ser definidas em duas categorias, conforme Quadro 1 apresentado em Sant'Anna (2009).

Termo	Categoria	Forma	Exemplo
Fração	Percentual	<u>discreto</u>	<u>nº de peças defeituosas</u>
		discreto	nº total de peças
(ou Razao)	Proporção	<u>contínuo</u> contínuo	volume sob pesquisa volume total

Quadro	1	- Características	da	qualidade	do	tipo	fração
--------	---	-------------------	----	-----------	----	------	--------

Fonte: Sant'Anna, 2009.

A primeira categoria, apresentada no Quadro 1, é denominada de percentual e compreende a fração ou razão entre dois números discretos como, por exemplo, o número de peças defeituosas dividido pelo número total de peças do lote. A fração do tipo percentual tem sido comumente modelada como distribuição de probabilidade Binomial, com parâmetros n_a e p, onde n_a representa o tamanho da amostra e p a fração não conforme (SANT'ANNA, 2009; COSTA *et al.*, 2010).
Segundo Montgomery (2004), a aproximação da distribuição Binomial pela Normal é satisfatória quando $n_a p > 10$ e p está no intervalo entre ($0,1 \le p \le 0,9$). Em muitos estudos a carta p é utilizada em situações onde o parâmetro p é considerado pequeno (por exemplo, p = 0,001; 0,01; 0,05; 0,1; ...) e nestes casos a distribuição Binomial é assimétrica e a aproximação pela distribuição Normal não é satisfatória, permitindo estimar valores negativos ou maiores do que 1 (um) (SANT'ANNA, 2009).

A segunda categoria, apresentada no Quadro 1, é denominada de proporção e compreende a razão entre dois números contínuos como, por exemplo, o volume de óleo bruto convertido em gasolina dividido pelo volume total. A fração do tipo proporção pode ser modelada como distribuição de probabilidade Beta, por ser uma distribuição contínua e seus parâmetros não dependem do tamanho da amostra (n_a) (SANT'ANNA, 2009).

A fração não conforme (p) é definida como a razão entre o número de unidades não conformes da amostra (N_{nc}) e o tamanho da amostra (n_a). Para a construção de uma carta p o tamanho do subgrupo deve ser escolhido em função da fração não conforme do processo para permitir a construção de uma carta eficiente na detecção de alterações no processo. Mitra (1998) sugere $n_a \overline{p} \ge 5$, ou seja, o tamanho da amostra (n_a) deve ser $n_a \ge 5/\overline{p}$, onde \overline{p} representa a fração média de não conformes.

Na carta tradicional de Shewhart para monitorar a fração não conforme, a linha média (LM) é igual a fração média de não conformes (\overline{p}) e o limite superior de controle (LSC) e o limite inferior de controle (LIC) são calculados assumindo que a distribuição Binomial possa ser aproximada pela distribuição Normal, conforme apresentado na equação (1)

$$LSC = \overline{p} + \lambda \sqrt{\frac{\overline{p}(1-\overline{p})}{n_a}} \qquad LM = \overline{p} \qquad LIC = \overline{p} - \lambda \sqrt{\frac{\overline{p}(1-\overline{p})}{n_a}} \qquad (1)$$

onde \overline{p} representa a fração média de não conformes, n_a representa o tamanho da amostra e λ é uma constante que define a largura dos limites de controle correspondente a uma região de controle $(1 - \alpha)$ e um número médio desejado de amostras até a presença de um sinal (NMA). O valor λ igual a 3 (três) é comumente utilizado devido a aproximação pela distribuição Normal, correspondendo a uma região 0,9973 de controle e um nível de NMA de 370,4 (MONTGOMERY, 2004; SANT'ANNA, 2009).

A Figura 3 apresenta um exemplo de carta *p* para a fração não conforme de um processo de fabricação de parafusos. Verifica-se que a 12^a amostra está acima da linha superior de controle (LSC), ou seja, com mais do que três desvios padrões em relação a linha média, necessitando que seja diagnosticada a causa especial pois o processo está fora de controle estatístico (MONTGOMERY, 2004).



Uma das medidas de desempenho mais utilizadas em cartas de controle é o número médio de amostras até a presença de um sinal (NMA). O NMA visa determinar a taxa de alarmes falsos da carta de controle investigada e a sensibilidade desta carta em detectar alterações pré-definidas no processo (MONTGOMERY, 2004). O número médio de observações analisadas até a primeira ocorrência de alarme falso é igual a $1/\alpha$. Isto significa que, com limites de $\mu \pm 3\sigma$, um alame falso ocorrerá em média a cada 370,4 pontos plotados (COSTA *et al.*, 2010).

2.2.2 Cartas de controle baseadas em modelos de regressão

A análise de regressão consiste de uma técnica estatística para modelar e investigar a relação entre uma ou mais características da qualidade (variáveis dependentes, V_d), com os fatores controláveis que podem afetá-las (variáveis independentes, V_i), ou seja, $V_d = f(V_i)$ (MONTGOMERY; RUNGER, 2003; CHARNET *et al.*, 2008).

Nos processos em que a característica da qualidade é diretamente dependente de algumas variáveis de controle do processo, e estas podem assumir diferentes valores em função de alterações frequentes de ajuste do processo modificando o modelo de referência da característica da qualidade, podem ser utilizadas as cartas de controle baseadas em modelos de regressão.

Nestas cartas de controle inicialmente é realizada a previsão do valor das características da qualidade por um modelo de regressão em função do seu relacionamento com as variáveis de controle do processo. Ao invés de monitorar a média das características da qualidade, como as cartas de controle tradicionais de Shewhart, as cartas de controle baseadas em modelos de regressão monitoram os resíduos entre as características da qualidade observadas e as características da qualidade previstas pelo modelo (MANDEL, 1969). Os resíduos são definidos como sendo a diferença entre o valor observado (dados medidos) e o valor estimado (previsto) pelo modelo.

Hawkins (1991) apresenta que, conforme o modelo de regressão é ajustado, o desempenho das cartas de controle é aperfeiçoado, pois é possível prever a variável dependente em função do ajuste das variáveis independentes. Segundo Jearkpaporn *et al.* (2005), o ajuste inadequado do modelo de regressão pode afetar o desempenho das cartas de controle no monitoramento do processo.

Para Psarakis e Papaleonida (2007), caso o modelo represente corretamente os dados, espera-se que os resíduos sejam normais, independentes e identicamente distribuídos, condição para a aplicação das cartas de controle convencionais. Se for detectada a variação da média ou da variância dos resíduos, conclui-se que a média ou a variância do processo em si foram alteradas. Assim, o monitoramento dos resíduos em uma carta de controle fornece um mecanismo para detectar uma alteração no processo. As cartas de controle baseadas em modelos de regressão linear para a previsão das características da qualidade requerem que as mesmas sejam normalmente distribuídas e homoscedásticas (variâncias iguais). Se após a realização da análise dos resíduos verificar-se que não foi possível satisfazer as suposições de normalidade e homoscedasticidade para o modelo de regressão linear, é possível que uma transformação dos dados das características da qualidade possa produzir a homoscedasticidade e a distribuição aproximadamente Normal dos resíduos. Um dos métodos que pode ser utilizado é a transformação de Box-Cox (BOX; COX, 1964).

As suposições de normalidade e homoscedasticidade muitas vezes não são satisfeitas em processos industriais quando se monitoram as características da qualidade do tipo fração de produtos não conformes, pois pressupõe-se que a característica da qualidade possui distribuição Normal e possibilita a previsão de valores fora dos limites do intervalo [0,1] (SANT'ANNA, 2009).

Alternativas para a modelagem de variáveis com distribuição não Normal são os modelos pertencentes à classe dos Modelos Lineares Generalizados (MLGs). Nos Modelos de Quase-Verossimilhança (MQV), pertencentes à classe dos MLGs, as variáveis são consideradas independentes sem ser necessário especificar qualquer distribuição de probabilidade para a variável resposta (CORDEIRO; DEMÉTRIO, 2008).

Outra alternativa para a modelagem da característica da qualidade do tipo fração é o Modelo de Regressão Beta (MRB), que se baseia na suposição de que a fração segue a distribuição Beta, que assume valores restritos ao intervalo [0,1] (FERRARI; CRIBARI-NETO, 2004). Sant'Anna e Caten (2012) propuseram uma carta de controle baseada em modelos Beta para monitorar processos com características da qualidade do tipo fração. Os resíduos gerados pelo método são plotados na carta de controle para monitorar o comportamento do processo e detectar mudanças devido as causas especiais.

Uma etapa essencial na análise de ajuste dos modelos de regressão é a verificação da adequação dos modelos aos dados, sendo constituída por um conjunto de critérios de adequação e ferramentas gráficas. As medidas de diagnóstico fornecem subsídios para detectar: possível violação de alguma das suposições feitas para o modelo, aleatoriedade dos dados, presença de pontos extremos ou com valores atípicos (do inglês, *outliers*), adequação da distribuição de probabilidade

proposta para a variável dependente e observação de pontos influentes (WERKEMA; AGUIAR, 1996; MONTGOMERY, 2004; CHARNET *et al.*, 2008; SANT'ANNA, 2009). Para a análise da adequação dos modelos de regressão podem ser utilizados cálculos de erros, coeficiente de determinação, análise dos resíduos e outros critérios, conforme cada caso.

3 APRENDIZADO DE MÁQUINA

Este capítulo apresenta a revisão da literatura necessária para o desenvolvimento da tese englobando Redes Neurais Artificiais, Máquinas de Vetores de Suporte, Máquinas de Vetores de Relevância e Evolução Diferencial, além das aplicações de Redes Neurais Artificiais e Máquinas de Vetores de Suporte no Controle Estatístico de Processos.

Para Mitchell (1997), o aprendizado de máquina trata do desenvolvimento de algoritmos e técnicas que permitem que computadores aprendam e melhorem o desempenho na realização de alguma tarefa por meio da experiência.

Conforme Faceli *et al.* (2011) as tarefas de aprendizado preditivas possuem a meta de encontrar uma função, hipótese ou modelo a partir dos dados de treinamento que possa ser utilizada para prever um rótulo ou valor que caracterize um novo exemplo, com base nos valores de seus atributos de entrada. Os algoritmos utilizados seguem o paradigma de aprendizado supervisionado, onde se conhece a saída desejada para cada conjunto de valores para os dados de entrada. Desta forma, pode ser avaliada a capacidade de hipótese induzida de predizer o valor de saída para novos exemplos. As tarefas supervisionadas dependem do tipo de rótulo dos dados: discreto, no caso de classificação, e contínuo, no caso de regressão.

Algumas das dificuldades na utilização do aprendizado de máquina se referem a ineficiência do algoritmo no caso da existência de mínimos locais e a complexidade da descrição da função de saída. Outra dificuldade ocorre quando está disponível um número limitado de exemplos de treinamento e validação e enriquecer a classe de hipóteses levará a problemas de sobre-treinamento (do inglês, *overfitting*) e a uma generalização ruim (CRISTIANINI; SHAWE-TAYLOR, 2000).

Deve-se levar em consideração, também, que o algoritmo de aprendizado normalmente é controlado por um grande número de parâmetros cuja escolha é baseada em heurísticas. Uma heurística pode ser definida como um algoritmo capaz de apresentar uma solução de boa qualidade, sem necessariamente garantir a otimalidade, para um determinado problema em tempo computacional compatível com a necessidade de rapidez presente no problema.

3.1 TEORIA DO APRENDIZADO SUPERVISIONADO

A teoria da otimização tem como objetivo determinar soluções para certa classe de problemas e desenvolver algoritmos eficazes que permitam encontrá-las (CRISTIANINI; SHAWE-TAYLOR, 2000).

Em problemas de otimização busca-se maximizar ou minimizar uma função objetivo, que depende de um número finito de variáveis de entrada. Estas variáveis podem ser independentes uma das outras ou podem estar relacionadas por meio de restrições. Dependendo da natureza da função objetivo, diz-se que um problema é de minimização ou de maximização (BELTRAMI, 2009).

Todo problema de maximização pode ser transformado em um problema de minimização equivalente. As soluções globais e locais de ambos os problemas possuem os mesmos valores ótimos, porém com sinais opostos. Os resultados obtidos para uma, podem ser estendidos para a outra, sem grandes dificuldades (IZMAILOV; SOLODOV, 2005). Cristianini e Shawe-Taylor (2000) e Izmailov e Solodov (2005) apresentam informações detalhadas sobre a teoria de otimização.

Normalmente, utiliza-se o princípio da indução para inferir uma função que minimize o erro sobre os dados de treinamento conhecidos e espera-se que esse procedimento leve também a um menor erro sobre os dados de teste. Na teoria do aprendizado estatístico, o erro de treinamento é chamado de risco empírico. O risco empírico mede o desempenho nos dados de treinamento por meio dos erros obtidos no conjunto de treinamento. Esse processo de indução constitui o princípio de minimização do risco empírico (*Empirical Risk Minimization,* ERM) (GUNN, 1998; VAPNIK, 1998).

O princípio de minimização do risco estrutural (*Structural Risk Minimization,* SRM) envolve a minimização de um limite superior sobre o erro de generalização, enquanto que o princípio ERM envolve a minimização do erro sobre os dados de treinamento. Os modelos de aprendizado de máquina baseados no princípio SRM tendem a apresentar uma maior habilidade para generalizar bem frente a dados de teste não observados (GUNN, 1998; VAPNIK, 1998; DU, SWAMY, 2013).

As técnicas de aprendizado de máquina selecionadas nesta pesquisa são as Redes Neurais Artificiais, as Máquinas de Vetores de Suporte e as Máquinas de Vetores de Relevância. As Redes Neurais Artificiais (RNAs) podem ser definidas como sistemas de processamento de informação baseados no comportamento do sistema nervoso humano. Segundo Haykin (2009), algumas propriedades das RNAs são: (i) a capacidade de aproximar o comportamento não linear de fenômenos físicos, não exigindo conhecimento estatístico dos dados modelados, (ii) a capacidade de aprender mapeamentos de entrada/saída de natureza contínua, e (iii) a habilidade de generalização do conhecimento adquirido após o processo de treinamento da rede e da possibilidade de estimar soluções para dados não conhecidos.

As RNAs utilizam o princípio ERM. O projeto e o treinamento de RNAs geralmente resultam em uma tarefa complexa na qual vários parâmetros devem ser ajustados, a quantidade necessária de dados de treinamento é geralmente grande e as RNAs tendem a causar sobre-treinamento (*overfitting*) (GUNN, 1998; VAPNIK, 1998).

Nos últimos anos, as Máquinas de Vetores de Suporte (SVMs) foram introduzidas como uma das várias técnicas baseadas em kernel disponíveis no campo do aprendizado de máquina para classificação, previsão e outras tarefas de aprendizado. Os métodos baseados em kernel utilizam o mapeamento de dados do espaço de características original de entrada para um espaço de características de kernel de maior dimensionalidade e, em seguida, na solução de um problema linear nesse espaço (SCHÖLKOPF; SMOLA, 2002).

Conforme Wang *et al.* (2003) comparadas com as RNAs tradicionais, as SVMs possuem algumas vantagens: (i) possuem alta capacidade de generalização e podem evitar mínimos locais, (ii) sempre possuem uma solução, (iii) não precisam determinar antecipadamente a topologia da rede, que pode ser obtida automaticamente quando o processo de treinamento termina, e (iv) possuem interpretação geométrica simples e resultam em uma solução esparsa.

As SVMs adotam o princípio SRM e Vapnik *et al.* (1997) mostraram que este princípio é superior ao princípio ERM, empregado nas RNAs, o que torna as SVMs com excelente capacidade de generalização. Vapnik introduziu as SVMs primeiro para resolver problemas de classificação de padrões, mas a técnica de vetores de suporte foi aplicada para o tratamento de problemas de regressão e denominada Regressão por Vetores de Suporte (*Support Vector Regression*, SVR) (VAPNIK, 1998). Tipping (2000) introduziu as Máquinas de Vetores de Relevância (RVMs), um tratamento Bayesiano de um modelo linear generalizado de forma funcional idêntica às SVMs, sendo utilizado para problemas de classificação e de regressão. Algumas vantagens das RVMs sobre as SVMs são: (i) podem produzir uma saída probabilística; (ii) não existe a necessidade de definir os parâmetros constante de regularização (*C*) e função de perda insensível (*ε*) existentes na SVR; e (iii) a capacidade de usar funções kernel que não seguem os princípios do Teorema de Mercer. A característica mais atraente das RVMs é que, embora possua capacidade de generalização comparável a uma SVM equivalente, a RVM normalmente utiliza menos funções kernel. A máquina de vetor de relevância para regressão (*Relevance Vector Regression*, RVR), constitui uma aproximação que pode ser usada para resolver modelos de regressão não linear.

3.2 REDES NEURAIS ARTIFICIAIS

Segundo Braga *et al.* (2012), as RNAs são sistemas paralelos distribuídos compostos por unidades de processamento simples (neurônios artificiais), dispostas em uma ou mais camadas e interligadas por um grande número de conexões, que calculam determinadas funções matemáticas (normalmente não lineares). Na maioria dos modelos as conexões estão associadas a pesos, que armazenam o conhecimento adquirido pelo modelo e servem para ponderar a entrada recebida por cada neurônio da rede.

O procedimento usual na solução de problemas utilizando RNAs passa inicialmente por uma fase de aprendizado, em que um conjunto de exemplos é apresentado para a rede, a qual extrai as características necessárias para representar a informação fornecida. Essas características são utilizadas para gerar respostas para o problema (BRAGA *et al.*, 2012).

As tarefas principais nas quais as RNAs se aplicam são: classificação, aproximação de funções, predição ou previsão, categorização (agrupamento ou *clustering*) e otimização (REZENDE, 2005).

Em termos matemáticos um neurônio artificial pode ser representado pelas equações (2), (3) e (4) (HAYKIN, 2009), tal que

$$\boldsymbol{u}_{k} = \sum_{i=1}^{m} \boldsymbol{\omega}_{ki} \boldsymbol{x}_{i}$$
⁽²⁾

$$v_k = u_k + b_k \tag{3}$$

$$\boldsymbol{y}_{k} = \boldsymbol{\varphi} \left(\boldsymbol{v}_{k} \right) = \boldsymbol{\varphi} \left(\boldsymbol{u}_{k} + \boldsymbol{b}_{k} \right) \tag{4}$$

onde u_k é a saída do combinador linear devido aos *m* sinas de entrada; ω_{ki} , com i = 1, ..., m, são os pesos sinápticos do neurônio *k*; x_i , com i = 1, ..., m, são os sinais de entrada; b_k é o termo de limiar (*bias*) do neurônio *k*; $\varphi(.)$ é a função de ativação e y_k é o sinal de saída do neurônio *k*.

Existem várias funções de ativação que podem ser utilizadas para o projeto de RNAs. Algumas destas funções são: degrau, linear, funções de base radial (por exemplo, gaussiana) e sigmoidais (por exemplo, logística e tangente hiperbólica). As três últimas funções são totalmente diferenciáveis e facilitam o processo de aprendizado da rede. As funções do tipo sigmoidal são as mais utilizadas em RNAs (ENGELBRECHT, 2007; HAYKIN, 2009; SILVA *et al.*, 2010).

De acordo com De Castro (2006), as RNAs podem ser caracterizadas por três aspectos principais: (i) um grupo de neurônios artificiais, denominados nós, unidades ou simplesmente neurônios; (ii) um padrão de conectividade entre os neurônios, denominado de arquitetura ou estrutura; e (iii) um método para determinar os valores dos pesos, denominado de algoritmo de treinamento ou de aprendizado.

A arquitetura de uma RNA define a forma como seus neurônios estão arranjados, uns em relação aos outros, através do direcionamento das conexões sinápticas destes neurônios. A topologia de uma RNA, considerando determinada arquitetura, pode ser definida como sendo as diferentes formas de composições estruturais que esta poderá assumir (SILVA *et al.*, 2010).

São basicamente três camadas em uma RNA (SILVA et al., 2010):

 a) Camada de entrada: é responsável pelo recebimento de informações do meio externo;

- b) Camadas escondidas, intermediárias ou ocultas: são compostas de neurônios que extraem as características associadas ao processo a ser inferido; e
- c) Camada de saída: é responsável pela produção e apresentação dos resultados finais da rede, obtidos dos processamentos efetuados pelos neurônios das camadas anteriores.

De acordo com De Castro (2006), em geral é possível distringuir três tipos principais de estruturas de RNAs: redes *feedforward* (alimentação à frente ou adiante) de camada única, redes *feedforward* de camadas múltiplas e redes recorrentes. Nas redes *feedforward* não existe conexão entre os neurônios da mesma camada e não existe realimentação entre as camadas (DU; SWAMY, 2013).

Nas redes *feedforward* de camada única tem-se a camada de entrada e uma única camada de neurônios, que é a própria camada de saída. O sinal é propagado na rede sempre da camada de entrada para a camada de saída (unidirecional) e nunca em caminho oposto (*backward*). Uma limitação das redes de uma camada é que elas conseguem classificar apenas objetos que são linearmente separáveis (ENGELBRECHT, 2007; FACELI *et al.*, 2011).

As redes *feedforward* de camadas múltiplas possuem uma ou mais camadas escondidas de neurônios. Pela adição das camadas escondidas é aumentado o processamento computacional e a capacidade de armazenamento da rede. A saída de cada camada da rede é usada como entrada da camada seguinte (DE CASTRO, 2006). O papel das múltiplas camadas é transformar, sucessivamente, o problema descrito pelo conjunto de dados no espaço de entrada em uma representação que pode ser tratada pela camada de saída da rede (BRAGA *et al.*, 2012).

Para resolver problemas não linearmente separáveis utilizando RNAs, a alternativa mais utilizada são as redes *feedforward* de camadas múltiplas. As camadas escondidas utilizam funções de ativação não lineares, como as funções sigmoidais (FACELI *et al.*, 2011).

Segundo Cybenko (1989), o teorema de Kolmogorov garante que uma rede *feedforward* com uma camada escondida, com um número suficiente de neurônios, pode aproximar qualquer função contínua com qualquer grau de precisão. Baseado no teorema de Kolmogorov, o teorema de Hecht-Nielsen apresenta que qualquer função contínua pode ser aproximada por uma rede *feedforward* com *m* nós de

entrada, 2*m*+1 neurônios escondidos e *j* neurônios de saída. A utilização de duas camadas escondidas permite a aproximação de qualquer função (SCARCELLI; TSOI, 1998; FACELI *et al.*, 2011).

Braga *et al.* (2012) apresentam que a maioria dos problemas práticos necessita de somente uma camada escondida e que a utilização de um número maior de camadas do que o necessário pode levar a problemas de convergência da rede. O número de neurônios em cada uma das camadas da rede é de extrema importância para o seu desempenho, principalmente no que se refere a capacidade de generalização. Quanto maior o número de neurônios nas camadas escondidas, mais complexas são as funções mapeadas com a RNA (REZENDE, 2005).

Entre os principais tipos de redes com arquitetura *feedforward* de camadas múltiplas estão o Perceptron Multicamadas (*Multilayer Perceptron*, MLP) e as Redes de Base Radial (*Radial Basis Function*, RBF).

Segundo Haykin (2009), a propriedade de importância primordial para uma RNA é a sua habilidade de aprender a partir de seu ambiente e de melhorar o seu desempenho através do aprendizagdo.

Uma RNA aprende por meio de um processo iterativo de ajustes dos pesos sinápticos e níveis de *bias*, tornando-se mais instruída sobre seu ambiente após cada iteração do processo de aprendizado. Um conjunto preestabelecido de regras para a solução de um problema de aprendizado é denominado de algoritmo de aprendizado, tendo-se como objetivo final a generalização de soluções a serem produzidas pelas saídas (HAYKIN, 2009; SILVA *et al.*, 2010).

Normalmente, o conjunto total de amostras disponíveis é dividido em dois subconjuntos, o de treinamento e o de teste. O subconjunto de treinamento é usado no processo de aprendizado da rede, sendo composto por cerca de 60 a 90% das amostras do conjunto total. O subconjunto de teste possui o restante das amostras do conjunto total e é utilizado para verificar se os aspectos referentes a generalização de soluções por parte da rede estão em patamares aceitáveis, possibilitando a validação da topologia assumida (BRAGA *et al.*, 2012).

Os algoritmos de aprendizado diferem entre si na maneira de obter o ajuste do peso sináptico de um neurônio, podendo ser agrupados em aprendizado supervisionado, aprendizado não supervisionado e aprendizado por reforço (DE CASTRO, 2006; BRAGA *et al.*, 2012).

O aprendizado supervisionado envolve a existência de um supervisor, que possui o conhecimento sobre o ambiente em que a rede está operando. Este conhecimento é representado na forma de um grupo de amostras ou padrões de entrada-saída. Os parâmetros livres da RNA são ajustados, por exemplo, através da combinação do sinal de entrada e do sinal de erro, sendo o sinal de erro obtido pela diferença entre a saída desejada e a saída atual calculada da rede (DE CASTRO, 2006). Como a resposta da rede é função dos valores atuais do seu conjunto de pesos, estes são ajustados de forma a aproximar à saída atual calculada da rede à saída desejada, ou seja, procura-se minimizar o sinal de erro (BRAGA *et al.*, 2012).

No treinamento de redes *feedforward* de camada única por meio do aprendizado supervisionado e correção de erros, o erro é obtido diretamente através da diferença entre a saída desejada e a saída atual da rede. No entanto, nas redes *feedforward* de camadas múltiplas esse procedimento pode ser aplicado somente para a camada de saída, pois não existem saídas desejadas definidas para as camadas escondidas (REZENDE, 2005).

Conforme Braga *et al.* (2012), para as redes MLPs, que se caracterizam por utilizarem funções de ativação sigmoidais nas camadas escondidas e sigmoidais ou lineares na saída, é utilizado o algoritmo de retropropagação de erros (do inglês, *error backpropagation*). Este algoritmo utiliza a descida de encosta e estima o erro das camadas escondidas por meio de uma estimativa do efeito que estas causam no erro da camada de saída.

Existem várias modificações do algoritmo de retropropagação de erros que visam acelerar seu tempo de treinamento e melhorar seu desempenho, sendo algumas destas modificações: a Levenberg-Marquardt, *backpropagation* com *momentum*, *quick propagation* (*Quickprop*), *resilient propagation* (RProp), gradiente conjugado escalonado (*Scaled Conjugate Gradiente*, SCG) e Gauss-Newton (BRAGA *et al.*, 2012; DU; SWAMY, 2013).

Segundo Haykin (2009), quando uma RNA aprende um número excessivo de exemplos de entrada-saída, a rede pode acabar memorizando os dados de treinamento e não conseguir generalizar os padrões de entrada-saída similares. Isto é conhecido como sobre-treinamento (do inglês, *overfitting*).

Uma topologia de RNA com um número reduzido de neurônios pode ser insuficiente para a extração de características que permitam à rede implementar as

hipóteses a respeito do comportamento do processo. Isto é conhecido como sub-treinamento (do inglês, *underfitting*) (SILVA *et al.*, 2010).

Uma das dificuldades na utilização das RNAs é a definição de sua arquitetura, que engloba a escolha das funções de ativação e da topologia da rede. A escolha da arquitetura mais adequada para um conjunto de dados normalmente é realizada por um processo de tentativa e erro, quando diferentes configurações são treinadas e avaliadas de acordo com uma precisão estabelecida (FACELI *et al.*, 2011). Outra abordagem é a utilização de heurísticas para a escolha dos melhores parâmetros para a RNA, tais como Algoritmos Genéticos (AGs), Evolução Diferencial (ED) e Otimização por Enxame de Partículas (PSO).

Além das arquiteturas citadas nesta seção, existem diversas estruturas de RNAs, sendo algumas destas: Redes Neurais de Regressão Generalizada (*General Regression Neural Networks*, GRNNs), Redes Neurais Probabilísticas (*Probabilistic Neural Networks*, PNNs), rede de Quantização Vetorial por Aprendizagem (*Learning Vector Quantization*, LVQ), Redes Auto Organizáveis de Kohonen, redes com Teoria da Ressonância Adaptativa (*Adaptive Resonance Theory*, ART), rede Mapa Auto Organizável (*Self-Organized Map*, SOM), Redes de Elman e Redes de Jordan.

3.3 MÁQUINAS DE VETORES DE SUPORTE

A estratégia de aprendizado baseada em vetores de suporte é fundamentada na Teoria de Aprendizado Estatístico ou teoria VC (Vapnik-Chervonenkis), desenvolvida por Vapnik em 1995 visando a proposição de técnicas de aprendizado de máquinas que buscam maximizar a capacidade de generalização (VAPNIK, 1998).

A aplicação de SVMs foi originalmente desenvolvida para problemas de classificação de padrões, sendo posteriormente estendida para o tratamento de problemas de regressão por Drucker *et al.* (1996) e denominada Regressão por Vetores de Suporte (*Support Vector Regression*, SVR). Vapnik *et al.* (1997) demonstraram que a técnica de vetores de suporte é uma ferramente útil para regressão (BURGES, 1998; SCHÖLKOPF; SMOLA, 2002).

Não existe um consenso na literatura para a aplicação da terminologia SVM, sendo esta muitas vezes utilizada para descrever classificação por vetores de suporte,

que também pode ser apresentada como SVC (*Support Vector Classification*). A terminologia SVR (*Support Vector Regression*) é utilizada para descrever Regressão por Vetores de Suporte. Nesta tese foi utilizada SVM de maneira geral para descrever problemas de classificação e regressão, SVM para classificação por vetores de suporte e SVR para regressão por vetores de suporte.

Nesta pesquisa, a Regressão por Vetores de Suporte (SVR) é aplicada para o desenvolvimento das cartas de controle baseadas em modelos de regressão. Para informações sobre as máquinas de vetores de suporte para classificação recomenda-se consultar Vapnik (1998), Kecman (2001), Schölkopf e Smola (2002), Smola e Schölkopf (2004) e Cherkassky e Mulier (2007). Beltrami (2009) apresenta um resumo das principais definições e teoremas utilizados em SVMs.

3.3.1 Regressão por vetores de suporte

A SVM foi inicialmente desenvolvida para resolver problemas de classificação e depois foi estendida para resolver problemas de regressão e denominada de Regressão por Vetores de Suporte (*Support Vector Regression*, SVR), mantendo as principais propriedades que caracterizam o algoritmo de máxima margem, como a dualidade, a esparsidade (conjunto esparso de vetores de suporte), a utilização de função kernel e a convexidade (BISHOP, 2006).

Basak *et al.* (2007) apresentam uma revisão da teoria, métodos e desenvolvimentos utilizando SVR. Eles apresentam que a idéia da SVR é baseada no cálculo de uma função de regressão linear em um espaço de características de alta dimensão onde os dados de entrada são mapeados por meio de uma função linear ou não linear.

Segundo Verdério (2015), a SVR difere da técnica de classificação no sentido de que, enquanto a segunda busca dividir os dados em diferentes classes e classificar os dados futuros, a primeira busca encontrar um preditor que aproxime bem os dados da amostra.

O algoritmo da SVR utiliza uma função de perda ε -insensivel (*loss function*) que ignora erros que estão além de uma certa distância dos valores considerados válidos, ou seja, erros são permitidos somente se forem menores do que ε . A função de perda ε -insensivel também garante a esparsidade da variável dual. Assim, na SVR tem-se dois conceitos importantes, que são o ε -tubo e a função de perda ε -insensível (KECMAN, 2001; ALVES, 2012).

A SVR foi desenvolvida para estimar funções de regressão e para resolver problemas não lineares. A construção da SVR utiliza a função de perda ε -insensível proposta por Vapnik (1998) e, na equação (5), *y* é o valor medido (alvo) e f(x) é o valor previsto, tal que

$$|y - f(x)|_{\varepsilon} = \begin{cases} 0 & \text{se } |y - f(x)| \le \varepsilon \\ |y - f(x)| - \varepsilon & \text{caso contrário} \end{cases}$$
(5)

De acordo com Smola e Schölkopf (2004), o objetivo do ε -SVR é encontrar uma função f(x) que produza saídas contínuas para todos os dados de treinamento que desviem no máximo de ε dos valores alvo y_n desejados e, ao mesmo tempo, permite que o ε -tubo seja o mais delgado possível. Essa função também deve ser o mais uniforme e regular possível (KECMAN, 2001).

A função de perda ε -insensivel da equação (5) define um tubo com raio ε ajustado aos dados, chamado ε -tubo. Se o valor previsto estiver dentro do tubo, a perda (erro) é zero. Para todos os outros pontos previstos fora do tubo, a perda é igual à magnitude da diferença entre o valor previsto e o raio ε do tubo (KECMAN, 2001).

Considerando os dados de treinamento $\{(x_1, y_1), ..., (x_n, y_n)\} \subset \aleph x \mathfrak{R}$, onde \aleph representa o espaço de entrada, o caso da função linear f(x) pode ser descrito na forma da equação (6) (SMOLA; SCHÖLKOPF, 2004)

$$f(\mathbf{x}) = \langle \omega, \mathbf{x} \rangle + b \quad \text{com} \quad \omega \in \mathfrak{R}, \ b \in \mathfrak{R}$$
(6)

onde $\langle .,. \rangle$ representa o produto escalar em \aleph , ω representa o vetor de pesos, x é o vetor de entrada e *b* o termo de limiar (*bias*).

Na equação (6) busca-se uma função com pequeno ω , o que pode ser conseguido pela minimização da norma Euclidiana, ou seja, $\|\omega\|^2 = \langle \omega, \omega \rangle$. O algoritmo

resolve, então, o problema de otimização apresentado na inequação (7) com as restrições apresentadas em (8) (SMOLA; SCHÖLKOPF, 2004; FACELI *et al.*, 2011)

$$\operatorname{Minimizar}_{\omega,b} \frac{1}{2} \left\| \omega \right\|^2 \tag{7}$$

com as restrições
$$\begin{cases} y_i - \omega . x_i - b \le \varepsilon_i \\ \omega . x_i + b - y_i \le \varepsilon_i \end{cases}$$
 (8)

Procura-se então a função linear que aproxime os pares (x_i, y_i) de treinamento com uma precisão de ε . A minimização da inequação (7) corresponde à obtenção do ε -tubo o mais delgado possível.

O problema de minimização da inequação (7) pode ser relaxado com a introdução de variáveis de folga (ξ_i), permitindo assim lidar com ruídos e *outliers* nos objetos e que alguns exemplos fiquem fora da região entre - ε e + ε . A função de perda ε -insensível introduz variáveis de folga não negativas que consideram os pontos situados fora da margem $|f(x_i) - y_i| \le \varepsilon$ com uma certa penalidade. A função ε -insensível é apresentada na equação (9) (SCHÖLKOPF; SMOLA, 2002)

$$\left|\xi\right|_{\varepsilon} = \begin{cases} 0 & \text{se } \left|\xi\right| \le \varepsilon \\ \left|\xi\right| - \varepsilon & \text{caso contrário} \end{cases}$$
(9)

A Figura 4 apresenta graficamente a situação apresentada pela equação (9), sendo a área sombreada correspondente ao ε -tubo (SCHÖLKOPF; SMOLA, 2002).

A variável de folga ξ_i está associada com os dados localizados abaixo da margem inferior e ξ_i^* com os localizados acima da margem superior. A inequação (10) apresenta a introdução das variáveis de folga e as restrições são apresentadas em (11) (SMOLA; SCHÖLKOPF, 2004; FACELI *et al.*, 2011)

$$\operatorname{Minimizar}_{\omega,b,\xi,\bar{\xi}} \frac{1}{2} \left\| \omega \right\|^2 + C \left(\sum_{i=1}^n \left(\xi_i + \xi_i^* \right) \right)$$
(10)

com as restrições
$$\begin{cases} y_i - \omega . x_i - b \le \varepsilon_i + \xi \\ \omega . x_i + b - y_i \le \varepsilon_i + \xi^* \\ \xi_i, \xi^* \ge 0 \end{cases}$$
 (11)





Fonte: Schölkopf e Smola, 2002.

Na inequação (10), a constante *C* é a constante de regularização e determina a quantidade de erros de treinamento que são permitidos e o custo para obtenção da região delgada do ε -tubo. A constante *C* > *0* é escolhida pelo usuário. Somente as amostras fora do ε -tubo na Figura 4 contribuem para o funcional da inequação (10), isto é, as variações que excedam ε são penalizadas (SMOLA; SCHÖLKOPF, 2004).

Os pontos fora do *ɛ*-tubo da Figura 4 são chamados de vetores de suporte (SVs), porque estabelecem os fundamentos para a função de regressão estimada. Isso significa que todos os outros pontos não são incluídos no modelo e podem ser removidos após a construção do modelo SVR. Assim, normalmente, muito menos objetos de treinamento constituem o modelo de regressão e, por esse motivo, a solução é definida como sendo esparsa. Tal solução é inerente ao algoritmo SVM, inicialmente desenvolvido para resolução de problemas de classificação de padrões (SCHÖLKOPF; SMOLA, 2002; ALVES, 2012).

Quando um modelo não linear é necessário para modelar adequadamente os dados deve-se estender o ε -SVR linear para a regressão não linear. A utilização de uma função kernel possibilita que os dados de entrada originais sejam mapeados no espaço de características de elevada dimensão, onde uma regressão linear pode ser utilizada.

Utilizando o método do multiplicador de Lagrange pode ser obtido o problema dual apresentado na inequação (12), descrito em termos de produtos internos entre os objetos, sendo as restrições apresentadas em (13). A função objetivo é apresentada na equação (14) (SMOLA; SCHÖLKOPF, 2004)

$$\operatorname{Maximizar}_{\alpha,\overline{\alpha}} - \frac{1}{2} \sum_{i,j=1}^{n} (\alpha_{i} - \alpha_{i}^{*}) (\alpha_{j} - \alpha_{j}^{*}) k(\mathbf{x}_{i}, \mathbf{x}_{j}) - \varepsilon \sum_{i=1}^{n} (\alpha_{i} + \alpha_{i}^{*}) + \sum_{i=1}^{n} \mathbf{y}_{i} (\alpha_{i} - \alpha_{i}^{*})$$
(12)

com as restrições
$$\begin{cases} \sum_{i=1}^{n} (\alpha_{i} - \alpha_{i}^{*}) = 0\\ \alpha_{i}, \alpha_{i}^{*} \in [0, C] \end{cases}$$
 (13)

$$f(\mathbf{x}) = \sum_{i=1}^{n} (\alpha_i - \alpha_i^*) k(\mathbf{x}_i, \mathbf{x}_j) + b$$
(14)

onde os termos α_i e α_i^* representam as variáveis de Lagrange, $k(x_i, x_j)$ é a função kernel e *b* o termo de limiar (*bias*). As variáveis de Lagrange associadas a todos os exemplos que se encontram dentro da margem entre - ε e + ε são nulas. As amostras fora do ε -tubo correspondem aos vetores de suporte (SVs).

A formulação dual da SVR permite que se trabalhe em um espaço de alta dimensionalidade. Assim, utilizando as funções kernel pode-se realizar um mapeamento não linear dos dados de entrada para um espaço de dimensão maior, onde a regressão linear torna-se possível. As funções kernel mais utilizadas são apresentadas no Quadro 2.

3.3.2 Funções kernel

Um kernel $k(x_i, x_j)$ é uma função de recebe dois pontos (x_i e x_j) no espaço de entrada original e calcula o produto escalar desses objetos no espaço de características de elevada dimensão. Dessa forma, o kernel pode transformar um conjunto de dados não linearmente separáveis em um conjunto de dados linearmente separáveis (FACELI *et al.*, 2011).

Na SVR são utilizadas funções kernel que seguem os princípios do Teorema de Mercer, em que uma função é definida como sendo uma função kernel se a matriz k é positivamente definida, onde $k = k_{ij} = \kappa (x_i, x_j)$. Uma matriz é positivamente definida se seus autovalores são maiores do que zero (CRISTIANINI; SHAWE-TAYLOR, 2000).

O Quadro 2 apresenta as funções kernel mais utilizadas, que são a linear, polinomial, laplaciano, função de base radial (*Radial Basis Function*, RBF) e sigmoidal, onde *u* é o parâmetro necessário para os kernels polinomial e sigmoidal, *d* é o grau do kernel polinomial e $\gamma = 1/2\nu^2$, sendo $\nu > 0$ o parâmetro que define a largura do kernel. Os parâmetros do kernel devem ser adequadamente determinados pelo usuário durante a etapa de treinamento para a obtenção de uma boa performance de generalização (FACELI *et al.*, 2011).

Kernel	Equação	Parâmetros do kernel
Linear	$\boldsymbol{k}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \boldsymbol{x}_i^T \boldsymbol{x}_j$	-
Polinomial	$k(\mathbf{x}_i, \mathbf{x}_j) = \left(\frac{1}{2\nu^2} \mathbf{x}_i^T \mathbf{x}_j + \mathbf{u}\right)^d = \left(\gamma \mathbf{x}_i^T \mathbf{x}_j + \mathbf{u}\right)^d$	γ, u, d
Laplaciano	$\boldsymbol{k}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \boldsymbol{exp}\left(-\frac{1}{2\nu^2} \ \boldsymbol{x}_i - \boldsymbol{x}_j\ \right) = \boldsymbol{exp}\left(-\gamma \ \boldsymbol{x}_i - \boldsymbol{x}_j\ \right)$	γ
Função de base radial (RBF)	$k(\boldsymbol{x}_{i}, \boldsymbol{x}_{j}) = exp\left(-\frac{1}{2\nu^{2}} \ \boldsymbol{x}_{i} - \boldsymbol{x}_{j}\ ^{2}\right) = exp\left(-\gamma \ \boldsymbol{x}_{i} - \boldsymbol{x}_{j}\ ^{2}\right)$	γ
Sigmoidal	$k(x_i, x_j) = tanh\left(\frac{1}{2\nu^2}x_i^Tx_j + u\right) = tanh(\gamma x_i^Tx_j + u)$	γ, u

Quadro 2 - Funções kerne	
--------------------------	--

Fonte: a autora, 2018.

Segundo Hsu *et al.* (2010), em geral o kernel RBF é uma primeira escolha razoável, pois esse kernel mapeia as amostras não linearmente em um espaço dimensional maior, possui menos hiperparâmetros e possui menos dificuldades numéricas.

3.3.3 Parâmetros para a regressão por vetores de suporte

A capacidade de generalização do modelo SVR depende da correta seleção dos parâmetros C e ε e do tipo de função kernel e seus parâmetros.

O parâmetro ε é responsável por definir o raio do tubo ε ao redor da função de regressão e, desta forma, o número de vetores de suporte que são selecionados para a construção da função de regressão, proporcionando uma solução esparsa.

A Figura 5 apresenta a influência do valor de ε no ajuste do modelo SVR para a função sinc(x) = sin(x) / x (SMOLA; SCHÖLKOPF, 2004). A Figura 5 (a) ilustra o uso de $\varepsilon = 0,1$, onde a linha central é a aproximação da função, o que torma o ε -tubo mais delgado e permite que mais pontos estejam externos ao ε -tubo, resultando em mais vetores de suporte. A Figura 5 (b) apresenta a aproximação de uma função com $\varepsilon = 0,5$, o que resulta em poucos vetores de suporte (mais pontos são ajustados no interior do ε -tubo) e uma função de regressão menos robusta. O valor de ε está relacionado com a amplitude do ruído presente no conjunto de treinamento e, sendo esta informação normalmente desconhecida, ε pode variar em um intervalo proposto pelo usuário (SMOLA; SCHÖLKOPF, 2004; ALVES, 2012).





Fonte: adaptado de Smola e Schölkopf, 2004.

O parâmetro *C* determina a penalização para os casos em que a função de regressão aceita objetos com desvios maiores que o valor de ε . Dessa forma, a robustez do modelo de regressão depende da escolha do valor de *C*, que pode assumir valores entre 0 e ∞ . Quando *C* é muito alto, o algoritmo é mais flexível, mas há uma probabilidade alta de uma adequação ao conjunto de treinamento (*overfitting*). Isto significa que o algoritmo terá baixo desempenho na generalização. No entanto,

um baixo valor de *C* pode fazer com que o algoritmo faça predições pobres devido à falta de flexibilidade (*underfitting*) (KUHN; JOHNSON, 2013).

Segundo Cristianini e Shawe-Taylor (2000), o parâmetro γ controla a flexibilidade da função kernel. Valores pequenos de γ podem levar ao sobre-treinamento (*overfitting*). Valores grandes de γ reduzem o kernel para uma função constante, tornando difícil o processo de aprendizado.

Inicialmente, Vapnik (1998) sugeriu a escolha manual dos parâmetros $C \in \varepsilon$ pelo especialista baseado no conhecimento *a priori* do conjunto de dados a ser avaliado. Porém, esse processo de seleção de parâmetros não é adequado para usuários que não são especialistas em SVR. Lima (2004) cita algumas propostas para seleção ótima dos parâmetros da SVR.

De acordo com Cherkassky e Mulier (2007), existem três abordagens para a seleção dos parâmetros da SVR: (i) busca em grade (*grid search*) em muitas combinações de valores de parâmetros em conjunto com a validação cruzada (*cross validation*) em cada conjunto candidato, (ii) busca eficiente no espaço de parâmetros usando limites analíticos VC, e (iii) seleção analítica dos parâmetros da SVR adaptados a um determinado tipo de problema de aprendizado.

Cherkassky e Ma (2004) resumiram muitas abordagens práticas para definir os valores dos parâmetros $C \in \varepsilon$ da SVR. Eles propuseram a seguinte abordagem para a seleção dos parâmetros utilizando o kernel RBF: seleção analítica do parâmetro C diretamente dos dados de treinamento, seleção analítica do parâmetro ε com base no nível de ruído (conhecido ou estimado) dos dados de treinamento e no número de amostras de treinamento, e seleção da largura do kernel RBF para refletir o intervalo de entrada dos dados de treinamento/teste.

Das diversas propostas existentes na literatura, quando não se tem o conhecimento prévio do conjunto de dados, a busca em grade e a validação cruzada são as mais utilizadas, apesar de custosas computacionalmente (HSU *et al.*, 2010).

A busca em grade se refere ao processo de busca exaustiva em um espaço formado pela combinação de alguns parâmetros de interesse e resolvendo o modelo para cada combinação de parâmetros. O objetivo da busca em grade é encontrar pontos nos quais a acurácia global seja a maior possível. O elevado tempo de processamento e a delimitação do espaço a ser investigado quando não se conhece previamente os dados são algumas desvantagens da busca em grade (MOORE *et al.*, 2011; HSU *et al.*, 2010).

A validação cruzada é um método utilizado para estimar diretamente o erro de predição esperado. Na *k-fold cross-validation* os dados são divididos em *k* conjuntos de tamanhos aproximadamente iguais, onde *k*-1 conjuntos são usados para treinar o modelo e o conjunto restante é utilizado para validar o modelo. Este procedimento é realizado *k* vezes até que cada conjunto seja utilizado para a validação do modelo. Este método é amplamente utilizado para avaliar a precisão de algoritmos (MOORE *et al.*, 2011; HASTIE *et al.*, 2013).

Outra abordagem é a utilização de heurísticas para escolha dos melhores parâmetros para a SVR, tais como, a Evolução Diferencial (ED) utilizada em Santos *et al.* (2012), os Algoritmos Genéticos (AGs) utilizados em Ebrahimzadeh *et al.* (2013) e a Otimização por Enxame de Partículas (PSO) utilizada em Yang (2016).

3.4 MÁQUINAS DE VETORES DE RELEVÂNCIA

As Máquinas de Vetores de Relevância (RVM) são modelos probabilísticos baseados em métodos kernel esparso que possuem algumas das vantajosas características das SVMs e evitam as suas principais limitações (TIPPING, 2000). A RVM possui a propriedade de soluções esparsas, proporcionando um desempenho mais rápido sob uma mesma base de testes e mantendo o poder de generalização dos dados comparável à SVM. As RVMs podem ser utilizadas para classificação e regressão, sendo denominada de Regressão por Vetores de Relevância (*Relevance Vector Regression*, RVR).

Nesta pesquisa, a RVR é aplicada para o desenvolvimento das cartas de controle baseadas em modelos de regressão. Para informações sobre as máquinas de vetores de relevância para classificação recomenda-se consultar Tipping (2000, 2001), Schölkopf e Smola (2002), Tipping e Faul (2003) e Bishop (2006).

A abordagem para a RVR considera uma função de um tipo correspondente àquelas implementadas pela SVR, equação (14). Dado um conjunto de pares de dados de entrada e saída (alvo) $\{\mathbf{x}_r, t_r\}_{r=1}^n$, considerando somente valores escalares para as funções de saída (alvo), segue-se a formulação probabilística padrão e se assume que as saídas (alvos) são observações de um modelo com ruído aditivo, conforme equação (15) (TIPPING, 2001)

$$t_r = \sum_{r=1}^{n} \omega_r k(\mathbf{x}_i, \mathbf{x}_j) + b + \epsilon_r$$
(15)

onde ω_r são os pesos do modelo, $k(\mathbf{x}_i, \mathbf{x}_j)$ é uma função kernel, b é o termo de limiar (*bias*) e ϵ_r são amostras independentes de ruído consideradas como ruído Gaussiano com média zero e variância σ_r^2 . Então, $p(t_r | \mathbf{x}) = \mathcal{N}(t_r | \mathbf{y}(\mathbf{x}_r), \sigma_r^2)$, onde a notação especifica uma distribuição Gaussiana sobre t_r com média $\mathbf{y}(\mathbf{x}_r)$ e variância σ_r^2 .

A RVR adota uma estrutura totalmente probabilística e utiliza a probabilidade *a priori* sobre os pesos do modelo regidos por um conjunto de hiperparâmetros, um associado a cada peso, cujos valores mais prováveis são estimados de forma iterativa a partir dos dados de treinamento.

A RVR é o tratamento Bayesiano da equação (15). Devido a consideração de independência de t_r , a estimativa de verossimilhança do conjunto de dados pode ser escrita pela equação (16) (TIPPING, 2001)

$$p(\mathbf{t}|\mathbf{w},\sigma_r^2) = \left(2\pi\sigma_r^2\right)^{-n/2} \exp\left\{-\frac{1}{2\sigma_r^2}\|\mathbf{t}-\mathbf{\Phi}\mathbf{w}\|^2\right\}$$
(16)

onde $\mathbf{t} = (t_1, ..., t_n)^T$, $\mathbf{w} = (\omega_0, ..., \omega_n)^T$ e Φ é a matriz com dimensão n x(n+1) sendo $\Phi = [\phi(x_1), \phi(x_2), ..., \phi(x_n)]^T$, em que $\phi(x_r) = [1, k(x_r, x_1), k(x_r, x_2), ..., k(x_r, x_n)]^T$.

De acordo com Tipping (2000), a estimativa de máxima verossimilhança de **w** e σ_r^2 da equação (16) levará ao sobre-treinamento (*overfitting*) severo. Na abordagem, a preferência é por funções mais suaves (menos complexas) sendo introduzida uma distribuição *a priori* sobre o vetor de pesos **w**, sendo esta uma função gaussiana com média zero. A introdução de um hiperparâmetro α_r separado para cada um dos parâmetros de peso ω_r é a característica principal da RVR e que a diferencia da SVR, que possui um único hiperparâmetro compartilhado. Assim, a distribuição a priori dos parâmetros de peso possui a forma dada pela equação (17)

$$\boldsymbol{\rho}(\mathbf{w}|\boldsymbol{\alpha}) = \prod_{r=0}^{n} \mathcal{N}(\omega_r | \mathbf{0}, \alpha_r^{-1})$$
(17)

onde α é um vetor de (n+1) hiperparâmetros e α_r representa a precisão do parâmetro correspondente ω_r (BISHOP, 2006).

Com a distribuição *a priori*, utiliza-se a regra de Bayes para determinar a distribuição *a posteriori* sobre os pesos, dada pela equação (18) (TIPPING, 2001)

$$p(\mathbf{w}|\mathbf{t}, \mathbf{\alpha}, \sigma_r^2) = (2\pi)^{-(n+1)/2} |\mathbf{\Sigma}|^{-1/2} \exp\left\{-\frac{1}{2}(\mathbf{w}-\mathbf{\mu})^{\mathsf{T}} \mathbf{\Sigma}^{-1}(\mathbf{w}-\mathbf{\mu})\right\}$$
(18)

onde a matriz de covariância (Σ) e o vetor média (μ) *a posteriori* são estimados, respectivamente, pelas equações (19) e (20)

$$\boldsymbol{\Sigma} = \left(\boldsymbol{\sigma}_r^{-2} \boldsymbol{\Phi}^{\mathsf{T}} \boldsymbol{\Phi} + \boldsymbol{\mathsf{A}}\right)^{-1}$$
(19)

$$\boldsymbol{\mu} = \sigma_r^{-2} \boldsymbol{\Sigma} \boldsymbol{\Phi}^{\mathsf{T}} \mathbf{t}$$
 (20)

onde $\mathbf{A} = diag(\alpha_{t_0}, \alpha_{t_1}, ..., \alpha_{t_n})$. A verossimilhança marginalizada dos hiperparâmetros é obtida pela integração dos pesos, conforme equação (21)

$$p(\mathbf{t}|\boldsymbol{\alpha},\sigma_{r}^{2}) = (2\pi)^{-n/2} \left|\sigma_{r}^{2}\mathbf{I} + \boldsymbol{\Phi}\mathbf{A}^{-1}\boldsymbol{\Phi}^{\mathsf{T}}\right|^{-1/2} exp\left\{-\frac{1}{2}\mathbf{t}^{\mathsf{T}} \left(\sigma_{r}^{2}\mathbf{I} + \boldsymbol{\Phi}\mathbf{A}^{-1}\boldsymbol{\Phi}^{\mathsf{T}}\right)^{-1}\mathbf{t}\right\}$$
(21)

Note que σ_r^2 é estimado como um hiperparâmetro do modelo, já que este pode ser estimado a partir dos dados. Os valores de α e σ_r^2 são determinados usando o método-II de maximização da verossimilhança (do inglês, *type-II maximum likelihood*), em que a função de verossimilhança marginalizada é maximizada pela integração dos parâmetros de peso. Para α , realizando a diferenciação da equação (21), igualando a zero e reorganizando a expressão, obtem-se a equação (22) (TIPPING, 2001)

$$\alpha_r^{novo} = \frac{\delta_r}{\mu_r^2}$$
(22)

onde μ_r é o *r*-ésimo componente do vetor média do peso *a posteriori* da equação (20), e a quantidade δ_r é definida por $\delta_r = 1 - \alpha_r \sum_{rr}$, onde \sum_{rr} é o *r*-ésimo elemento da diagonal da matriz de covariância do peso *a posteriori* da equação (19), calculados com os valores atuais de α e σ_r^2 . Cada $\delta_r \in [0,1]$ pode ser interpretado como uma medida de quanto "bem determinado" está sendo cada parâmetro ω_r pelos dados.

Para a variância do ruído σ_r^2 , a diferenciação da equação (21) leva à reestimativa apresentada pela equação (23), onde *n* no denominador se refere ao número de observações e não ao número de funções de base (TIPPING, 2001)

$$\left(\sigma_r^2\right)^{novo} = \frac{\left\|\mathbf{t} - \mathbf{\Phi}\boldsymbol{\mu}\right\|^2}{n - \sum_r \delta_r}$$
(23)

O algoritmo de aprendizado prossegue pela aplicação repetida das equações (22) e (23), concomitante à atualização da estatística *a posteriori* Σ e μ das equações (19) e (20), até que alguns critérios de convergência adequados tenham sido satisfeitos, como um limiar pré-determinado ou um certo número de iterações tenha sido atingido.

Como resultado do processo de otimização, a proporção dos hiperparâmetros α_r é conduzida para valores elevados (em princípios infinitos) e, portanto os parâmetros de peso ω_r correspondentes a estes hiperparâmetros possuem distribuições *a posteriori* com média e variância nulas (BISHOP, 2006). Assim, os parâmetros infinitos e nulos resultante da maximização são removidos do modelo e as funções de base $\phi(x_r)$ correspondentes podem ser podadas e a esparsividade é obtida. No caso dos modelos com a forma da equação (15), as entradas x_n

correspondentes aos pesos remanescentes não nulos são denominados vetores de relevância (RVs) e são análogos aos vetores de suporte (SVs) de uma SVR.

Tipping (2004) apresenta um exemplo de comparação entre a RVR e a SVR utilizando o kernel RBF, que demonstra algumas das vantagens da RVR. As técnicas foram aplicadas a um problema de regressão para a função sinc(x) = sin(x)/x com ruído uniforme adicionado aos dados, conforme Figura 6. A RVR treinada utilizou 7 (sete) vetores de relevância (RVs) e obteve erro máximo de 0,0664 e raiz do erro médio quadrático (*Root Mean Squared Error*, RMSE) de 0,0322, enquanto a SVR utilizou 29 vetores de suporte (SVs) e obteve erro máximo de 0,0896 e RMSE de 0,0420. Os parâmetros *C* e ε da SVR foram ajustados através de um procedimento de validação cruzada anterior a etapa de aprendizado, enquanto o algoritmo RVR estima automaticamente seus parâmetros (α_r 's e σ_r^2) durante o procedimento de aprendizado.



Figura 6 - RVR e SVR aplicadas a um problema de regressão

Uma característica competitiva da RVR é que, embora com capacidade de generalização comparável a uma SVR equivalente, o número de RVs é, na maioria dos casos, significativamente menor que o número de SVs usados por uma SVR para resolver o mesmo problema. A principal desvantagem da RVR está na complexidade da fase de treinamento porque envolve a otimização de uma função não convexa. Para grandes conjuntos de dados, os tempos de treinamento podem ser maiores do

que para uma SVR equivalente. Mais significativamente, na RVR os parâmetros que controlam a complexidade e a variância do ruído (α_r 's e σ_r^2) são automaticamente estimados pelo procedimento de aprendizado, enquanto na SVR é necessário ajustar os parâmetros *C* e ε , geralmente utilizando validação cruzada que envolve várias execuções de treinamento. No entanto, alguns autores desenvolveram procedimentos para treinamento da RVR que melhora significativamente a velocidade de treinamento (TIPPING; FAUL, 2003; BISHOP, 2006).

3.5 REDES NEURAIS ARTIFICIAIS E MÁQUINAS DE VETORES DE SUPORTE EM CONTROLE ESTATÍSTICO DE PROCESSOS

Esta seção tem como finalidade apresentar alguns dos primeiros trabalhos e o que tem sido desenvolvido nos últimos anos relativo à aplicação de Redes Neurais Artificiais, Máquina de Vetores de Suporte e Máquinas de Vetores de Relevância em Controle Estatístico de Processos. Não estão incluídas nesta seção as dissertações de mestrado, as teses de doutorado, os livros e os artigos publicados em conferências e congressos nacionais e internacionais.

Para realizar este levantamento foi utilizado o portal de periódicos da CAPES, com a busca dos artigos nos periódicos entre os anos de 2008 e 2018. As bases de dados consultadas foram: ACS *Publications*, *Emerald Insight*, *Engineering Village*, IEEE *Xplore*, *Proquest*, *ScienceDirect*, *Springer Link*, Taylor & Francis *Online*, *Wiley Online Library* e *World Scientific*.

Para realizar a busca dos artigos somente na área de CEP foram utilizadas as palavras-chaves "*neural network*", "*relevance vector*" e "*support vector*" em conjunto com as palavras-chaves "*control chart*", "*statistical process control*", "*pattern recognition*" e "*process monitoring*".

Entre 2008 e 2018, verificou-se que foram publicados artigos empregando RNAs, SVMs e SVR para o reconhecimento de padrões em cartas de controle, para o monitoramento de processos industriais e para o desenvolvimento de cartas de controle. Porém, não foram encontrados artigos empregando RVMs e RVR nestas aplicações de CEP. As aplicações de RVMs e RVR se limitaram a classificação de padrões, modelagem e predição em diferentes processos industriais. Para o reconhecimento de padrões em cartas de controle são utilizadas as propriedades de classificação das RNAs e SVMs, no monitoramento de processos industriais para a identificação de alterações nas características utilizadas para o acompanhamento dos processos (principalmente a média e a variância) podem ser utilizadas as propriedades de classificação ou de regressão das RNAs, SVMs e SVR e, para o monitoramento de processos por meio das cartas de controle, a maioria das publicações pesquisadas utilizaram as propriedades de regressão das RNAs e SVR.

3.5.1 Aplicações de redes neurais artificiais em CEP

Conforme Psarakis (2011), as RNAs começaram a ser utilizadas em CEP na década de 1980. Os primeiros artigos publicados envolvendo RNAs e CEP são de Pugh (1989), que propôs uma RNA do tipo MLP com três camadas com algoritmo de treinamento de retropropagação do erro para determinar variações na média do processo, e Pugh (1991), que propôs uma RNA do tipo MLP com quatro camadas também para determinar variações na média do processo, ambas com função de ativação sigmoidal.

Smith (1994), além de apresentar uma RNA do tipo MLP com algoritmo de treinamento de retropropagação do erro para determinar as variações na média e na variância do processo, utilizou a capacidade de predição das RNAs e propôs que essas redes podem ser usadas para predizer os valores das amostras futuras com base no padrão identificado das amostras medidas no processo.

Com relação ao reconhecimento de padrões, um dos primeiros artigos é de Hwarng e Hubele (1993), que propuseram uma RNA do tipo MLP com três camadas com algoritmo de treinamento de retropropagação do erro para reconhecimento de oito padrões anormais em cartas de controle.

Para a identificação de mudanças nos parâmetros de processos com dados autocorrelacionados, um dos primeiros artigos é de Cook e Chiu (1998), que usam RNA do tipo RBF em dois processos industriais.

Zorriassatine e Tannock (1998) apresentam uma revisão da literatura das aplicações de RNAs em cartas de controle até 1997. Paliwal e Kumar (2009) realizaram uma revisão de artigos que envolvem um estudo comparativo de redes neurais *feedforward* e técnicas estatísticas usadas para problemas de predição e classificação em várias áreas de aplicações.

Masood e Hassan (2010) apresentam uma revisão da literatura sobre o reconhecimento de padrões em cartas de controle com RNAs, abordando os dados de entrada e padrões do processo, projeto e treinamento da rede, monitoramento e diagnóstico de processos multivariados. Basu *et al.* (2010) apresentam uma revisão da utilização de RNAs no reconhecimento de padrões em diversas áreas.

Psarakis (2011) apresenta uma revisão da literatura sobre a utilização de RNAs em cartas de controle na detecção e identificação de variações na média e na variância de processos univariados e multivariados, no reconhecimento de padrões específicos nas cartas de controle e para dados de processos autocorrelacionados.

Hachicha e Ghorbel (2012) apresentam uma revisão da literatura de reconhecimento de padrões em cartas de controle de 1991 a 2010. Nesta pesquisa, a maioria dos artigos analisados (61,47%) usaram a abordagem de RNAs para o reconhecimento de padrões, sendo mais de 80% de RNAs supervisionadas, e apenas cinco artigos aplicam os métodos propostos com dados de processos reais.

Atashgar (2015) apresenta uma revisão da literatura sobre o monitoramento de processos multivariados utilizando RNAs. Ele analisou as vantagens e desvantagens dos esquemas propostos e comparou suas capacidades e propriedades. A RNA é o mais utilizado dos métodos de aprendizado de máquina para o projeto de sistemas de reconhecimento de padrões em cartas de controle. Nestas aplicações são utilizadas diferentes arquiteturas de RNAs, tais como MLP, RBF, LVQ e PNN.

A Figura 7 apresenta o número de publicações entre os anos de 2008 e 2018 nos periódicos pesquisados, totalizando 93 publicações. É possível observar que o número de publicações de abordagens de RNAs em CEP é significativo, mas está diminuindo a partir de 2016.

A Figura 8 apresenta os artigos agrupados por abordagem de utilização: reconhecimento de padrões em cartas de controle e monitoramento de processos (identificação de alterações na média e na variância dos processos) e o desenvolvimento de cartas de controle. Para o reconhecimento de padrões em cartas de controle são utilizadas as propriedades de classificação das RNAs, no monitoramento de processos são utilizadas as propriedades de classificação ou de regressão das RNAs e, para o desenvolvimento de cartas de controle, a maioria das

publicações pesquisadas utilizou as propriedades de regressão das RNAs. Como pode ser observado na Figura 8, as aplicações em maior número envolvendo as RNAs são relativas ao monitoramento de processos.



Figura 7 - Publicações de RNAs em CEP por ano





Figura 8 - Publicações de RNAs em CEP por abordagem

Fonte: a autora, 2018.

A Tabela 1 apresenta os principais periódicos responsáveis pelas publicações de RNAs em CEP, sendo detalhados os que possuem a partir de três publicações sobre a temática. Da Tabela 1 pode-se verificar que dez periódicos são responsáveis por 66,67% das publicações entre 2008 e 2018.

Periódico	Frequência absoluta	Frequência Percentual
International Journal of Advanced Manufacturing Technology	14	15,05%
International Journal of Production Research	10	10,75%
Computers & Industrial Engineering	9	9,68%
Expert Systems with Applications	8	8,60%
Quality and Reliability Engineering International	4	4,30%
Journal of Intelligent Manufacturing	4	4,30%
ISA Transactions	4	4,30%
Applied Soft Computing	3	3,23%
Journal of Industrial Engineering International	3	3,23%
Neurocomputing	3	3,23%
Outros	31	33,33%
Total	93	100,00%

Tabela 1 - Principais periódicos com publicações de RNAs em CEP

Fonte: a autora, 2018.

Para o reconhecimento de padrões em cartas de controle as RNAs são utilizadas, na maioria dos artigos, para identificar e classificar os padrões anormais nos processos e muitos pesquisadores utilizaram a rede MLP e a rede RBF. Alguns artigos nesta abordagem são os de Gauri e Chakraborty (2008), Gauri e Chakraborty (2009), Jiang *et al.* (2009), Ebrahimzadeh e Ranaee (2010), El-Midany *et al.* (2010), Guh (2010), Ghomi *et al.* (2011), Ebrahimzadeh *et al.* (2012), Ebrahimzadeh *et al.* (2013), Ghiasabadi *et al.* (2013), Addeh *et al.* (2014), Barghash (2015), Cheng *et al.* (2015), Yang e Zhou (2015), Naeini e Bayatib (2017) e Addeh *et al.* (2018). Awadalla e Sadek (2012) utilizaram a arquitetura de redes neurais pulsadas para o reconhecimento de padrões em cartas de controle.

Em função do bom desempenho no reconhecimento de padrões anormais, apresentado nos trabalhos pesquisados, as RNAs são uma boa opção para a automatização do reconhecimento de padrões nas cartas de controle.

Nas aplicações mais recentes de RNAs na detecção e identificação de alterações na média e na variância em processos univariados e multivariados, muitos pesquisadores utilizaram a rede MLP e a rede RBF. Na detecção e identificação de alterações na média podem ser citados Guh e Shiue (2008), Niaki e Davoodi (2009), Ahmadzadeh (2009), Noorossana *et al.* (2011), Atashgar e Noorossana (2011), Ahmadzadeh *et al.* (2013), Huwang *et al.* (2014) e Maleki e Amiri (2015).

Na detecção e identificação de alterações na variância tem-se Du *et al.* (2010), Cheng e Cheng (2011) e Zaman *et al.* (2015). Niaki e Abbasi (2008) utilizaram RNAs para monitorar sinais fora de controle em processos.

Niaki e Abassi (2008), Niaki e Nafar (2008), Hwarng e Wang (2010) e Wu e Yu (2010) utilizaram RNAs em processos com dados autocorrelacionados. Yu *et al.* (2009) e Du *et al.* (2010) utilizaram ensemble neural para monitorar e diagnosticar sinais fora de controle em processos bivariados. Pacella e Semeraro (2011) e Ghomi *et al.* (2011) utilizaram redes com aprendizado não supervisionado para o monitoramento de processos. Niaki e Nasaji (2011) modificaram uma RNA de Elman para classificar alterações na média de processos autocorrelacionados.

Yu e Xi (2009) desenvolveram uma carta de controle para o monitoramento e diagnóstico de sinais fora de controle utilizando ensemble neural. Garjani *et al.* (2010) desenvolveram uma carta de controle para detecção de alterações na média de processos. Kim *et al.* (2012) desenvolveram cartas de controle para os resíduos de modelos utilizando RNAs para o monitoramento da média de processos.

Amiri *et al.* (2015) propuseram um procedimento baseado em redes PNNs para detectar alterações na variância e Yang *et al.* (2015) aplicaram redes neurais auto associativas para detecção de alterações em cartas de controle. Rezki *et al.* (2016) propuseram um sistema inteligente para detecção, diagnóstico, identificação e reconfiguração de sistemas complexos que combina cartas de controle multivariadas, RNAs, redes bayesianas e sistemas especialistas.

Shao e Chiu (2016) utilizaram as técnicas de Aprendizado Extremo de Máquina (*Extreme Learning Machine*, ELM) e *Random Forest* (RF), além de RNAs e *Rough Set* (RS) para o reconhecimento de padrões. Wang *et al.* (2017) utilizaram redes neurais profundas (*Deep Neural Networks*) para identificação de falhas em turbinas eólicas e monitoraram as alterações utilizando cartas de Média Móvel Exponencialmente Ponderada (*Exponentially Weighted Moving Average*, EWMA). Guo *et al.* (2018) utilizaram redes neurais convolucionais e redes MLP para o monitoramento das condições de operação e diagnósticos de falhas em turbinas eólicas.

Os trabalhos de Niaki e Nafar (2008), Niaki e Abbasi (2008), Niaki e Nasaji (2011) e Maleki e Amiri (2015) utilizaram processos com características da qualidade em atributos, proporção de não conformes e número de defeitos no processo.

Em muitos dos trabalhos citados os autores apresentaram que as RNAs, quando comparadas às cartas de controle tradicionais, apresentaram um desempenho similar ou superior na identificação de pequenas alterações nos valores da média ou da variância. Desta forma, as RNAs se apresentam como alternativas às cartas de controle tradicionais para o monitoramento de processos. Apesar das RNAs apresentarem bons resultados no monitoramento de processos, estas ainda são mais complexas que as técnicas estatísticas convencionais.

Uma das dificuldades na utilização de RNAs é a definição de sua arquitetura, ou seja, a escolha das funções de ativação e da topologia da rede. Outra dificuldade é a escolha do melhor conjunto de parâmetros para a arquitetuta da rede, pois o desempenho de uma RNA depende muito dos parâmetros.

Normalmente, a escolha da arquitetura de uma RNA mais adequada para um conjunto de dados é realizada por um processo de tentativa e erro, quando diferentes configurações são treinadas e avaliadas de acordo com uma precisão estabelecida. Este método é considerado exaustivo e tendencioso a erros (SILVA *et al.*, 2010).

Muitos autores apresentados nesta seção utlizaram redes MLP com algoritmo de treinamento de retropropagação do erro e processo de análise dos erros para a obtenção dos melhores parâmetros para as RNAs.

Diferentes heurísticas têm sido utilizadas para a obtenção dos melhores parâmetros para a arquitetuta de uma RNA. Dos artigos apresentados nesta seção, Niaki e Nasaji (2008) utilizaram *Simulated Annealing*, o algoritmo PSO foi utilizado por Yu e Xi (2009) e Yu *et al.* (2009), enquanto Du *et al.* (2010) utilizaram PSO e *Simulated Annealing*. Addeh *et al.* (2014) utilizaram o algoritmo inspirado no comportamento de cucos (*Cuckoo search algorithm*).

3.5.2 Aplicações de técnicas baseadas em vetores de suporte em CEP

As primeiras aplicações de SVMs em controle estatístico de processos foram apresentadas por Chinnam e Kumar (2001), que utilizaram SVMs para reconhecer alterações na média em processos autocorrelacionados, e Chinnam (2002), que realizou o reconhecimento de alterações em processos correlacionados e não correlacionados utilizando SVMs para minimização de erros. Nesta mesma década, podem-se citar os trabalhos de Sun e Tsung (2003) e Camci *et al.* (2008), que propuseram novas cartas de controle usando princípios de vetores de suporte, e Cheng e Cheng (2008), que utilizaram RNAs e SVMs para a identificação de alterações na variância em processos multivariados.

Rivas-Perea *et al.* (2013) revisaram os conceitos gerais de SVMs, os métodos de treinamento de SVMs, as formulações mais utilizadas para SVR e apresentam formulações de SVR baseadas em programação linear. Salcedo-Sanz *et al.* (2014) fornecem uma visão geral de SVMs e SVR e apresentam resultados bem sucedidos de suas aplicações em problemas reais em diferentes campos da engenharia.

Weese *et al.* (2016) apresentam uma visão geral e o estado atual dos métodos do aprendizado estatístico e técnicas de redução da dimensão aplicadas em cartas de controle para monitoramento de processos multivariados. Neste artigo são apresentados métodos de aprendizado supervisionado (regressão logística, RNAs, SVMs, *Support Vector Data Description* (SVDD) e árvores de decisão, entre outros) e não supervisionado (análise de componentes principais, análise de agrupamentos, entre outros).

Cuentas *et al.* (2017) apresentam uma revisão de artigos sobre a aplicação de SVMs para o reconhecimento de padrões em cartas de controle utilizando as propriedades de classificação das SVMs.

A Figura 9 apresenta o número de publicações de SVMs e SVR em CEP entre os anos de 2008 e 2018 nos periódicos pesquisados, totalizando 76 publicações. É possível observar que o número de publicações é significativo entre 2013 e 2016, ocorrendo uma redução no número de publicações em 2017 e 2018.

A Figura 10 apresenta os artigos agrupados por abordagem de utilização: reconhecimento de padrões em cartas de controle, monitoramento de processos (identificação de alterações na média e na variância) e desenvolvimento de cartas de controle. Para o reconhecimento de padrões em cartas de controle são utilizadas as propriedades de classificação das SVMs, para a detecção de alterações nos processos podem ser utilizadas as propriedades de classificação das SVMs ou de regressão da SVR e, para o desenvolvimento de cartas de controle, a maioria das publicações pesquisadas utilizou as propriedades de regressão da SVR. Como pode ser observado na Figura 10, as aplicações em maior número são relativas ao reconhecimento de padrões em cartas de controle.



Figura 9 - Publicações de SVMs e SVR em CEP por ano

Fonte: a autora, 2018.





Fonte: a autora, 2018.

Segundo Han *et al.* (2011), as SVMs tendem a ser mais precisas e menos propensas a sobre-treinamento (*overfitting*) do que outros métodos de classificação. Ranaee e Ebrahimzadeh (2011) mostraram que o kernel RBF possui o melhor desempenho para aplicações de reconhecimento de padrões em cartas de controle.

Alguns dos pesquisadores que utilizaram SVMs para o reconhecimento de padrões em cartas de controle são: Ranaee *et al.* (2010), Lin *et al.* (2011), Lu *et al.* (2011), Ranaee e Ebrahimzadeh (2011), Salehi *et al.* (2012), Du *et al.* (2013), Xie *et*
al. (2013), Zhang *et al.* (2015), Zhang e Cheng (2015), Wu *et al.* (2015), Kao *et al.* (2016), Kazemi *et al.* (2016) e Zhang *et al.* (2018).

Das e Banerjee (2011), Ebrahimzadeh *et al.* (2013), Chiñas *et al.* (2015) e Gutierrez e Pham (2016) utilizaram RNAs e SVMs para reconhecimento de padrões em cartas de controle.

O reconhecimento de padrões em cartas de controle usando Máquina de Vetor de Suporte por Mínimos Quadrados (*Least Squares Support Vector Machine*, LS-SVM) foi realizado por Zhiqiang e Yizhong (2010) e Cheng *et al.* (2011). Xanthopoulos e Razzaghi (2014), Zhou *et al.* (2015) e Khormali e Addeh (2016) apresentaram o reconhecimento de padrões baseado em *fuzzy* SVM.

Alguns dos pesquisadores que aplicaram SVMs para a identificação de alterações na média e na variância em processos univariados e multivariados são: Hsu *et al.* (2010a), Hsu *et al.* (2010b), Cheng *et al.* (2011), Salehi *et al.* (2011), Cheng e Lee (2012), Du *et al.* (2012), Hu *et al.* (2016), Wang *et al.* (2017) e Onel *et al.* (2018).

Cheng e Cheng (2008) utilizaram RNAs e SVMs para a identificação da alterações na variância em processos multivariados. Kim *et al.* (2012) desenvolveram cartas de controle para os resíduos de modelos utilizando RNAs e SVR para o monitoramento da média de processos.

Zhao *et al.* (2013) apresentaram um método de detecção de falhas utilizando SVR e cartas de controle EWMA. Zhang *et al.* (2015) desenvolveram um sistema de monitoramento para detectar variações em processos utilizando SVMs e cartas de controle Multivariada de Média Móvel Exponencialmente Ponderada (*Multivariate Exponentially Weighted Moving Average*, MEWMA).

Fan *et al.* (2016) propuseram um sistema para detectar alarmes falsos utilizando Regressão por Vetor de Suporte por Mínimos Quadrados (*Least Squares Support Vector Regression*, LS-SVR) e traçaram a carta de controle dos resíduos. Yang (2016) apresentou um modelo para monitoramento *on-line* de alterações na média e na variância de processos bivariados utilizando PSO discreto, redes neurais RBF, SVMs e SVR. Yang *et al.* (2018) utilizaram SVR para o monitoramento de sistemas de turbinas eólicas e traçaram a carta de controle dos resíduos.

Issam e Mohamed (2008) e Khediri *et al.* (2010) propuseram a utilização de SVR no desenvolvimento de cartas de controle para monitoramento de processos multivariados autocorrelacionados. Gani *et al.* (2010) utilizaram SVR para o

desenvolvimento de cartas de controle para resíduos em processos multivariados, denominadas SVR-chart. Chongfuangprinya *et al.* (2011) desenvolveram uma carta de controle utilizando SVMs e probabilidade de classes (*Probability of Class*, PoC).

Du e Lv (2013) desenvolveram uma carta de controle baseada em SVR e mínima distância Euclidiana para monitoramento de alterações na média de processos autocorrelacionados. Yampikulsakul *et al.* (2014) utilizaram LS-SVR para o desenvolvimento de cartas de controle. He *et al.* (2016) propuseram uma carta de controle para processos multivariados utilizando SVMs, denominada de D-SVM *chart.* Khusna *et al.* (2018) desenvolveram uma carta de controle MEWMA baseada em LS-SVR para monitoramento de processos multivariados com dados autocorrelacionados.

A Descrição de Dados com Vetor de Suporte (*Support Vector Data Description*, SVDD) é um método kernel inspirado nas propriedades de classificação das SVMs, sendo utilizada em tarefas de classificação de uma classe (*one class classification*), detecção de anomalias, detecção de novidades, classificação binária e multi-classe, entre outras. A SVDD é caracterizada pela habilidade em lidar com dados de elevadas dimensões e limites de controle flexíveis (TAX *et al.*, 1999, TAX; DUIN, 2004).

Sun e Tsung (2003) propuseram uma carta de controle baseada em distância kernel, denominada *k-chart*, que utiliza os princípios da SVDD. Nos últimos anos alguns pesquisadores desenvolveram cartas de controle baseadas nos princípios SVDD e utilizaram estas cartas de controle em aplicações industriais.

As publicações de aplicações de SVDD em CEP não foram incluídas nos dados das Figuras 9 e 10, por se tratar de uma técnica baseada em vetores de suporte inspirado nas SVMs. A Figura 11 apresenta o número de publicações utilizando princípios da SVDD em CEP entre os anos de 2008 e 2018 nos periódicos pesquisados, totalizando 33 publicações.

Alguns dos pesquisadores que aplicaram SVDD em controle estatístico de processos são: Camci *et al.* (2008), Sukchotrat *et al.* (2010), Gani e Limam (2011), Khediri *et al.* (2012), Gani e Limam (2013a), Gani e Limam (2013b), Ning e Tsung (2013), Yao *et al.* (2014), Huang e Yan (2016), Lv e Yan (2016), Lv *et al.* (2017), Lee e Kim (2018) e Wang *et al.* (2018).

A Tabela 2 apresenta os principais periódicos responsáveis pelas publicações de técnicas baseadas em vetores de suporte (SVMs, SVR e SVDD) em CEP, sendo

detalhados os que possuem a partir de três publicações sobre a temática. No total são 109 publicações, sendo 76 publicações sobre SVMs e SVR e 33 publicações de abordagens de SVDD. Da Tabela 2 pode-se verificar que doze periódicos são responsáveis por 56,88% das publicações entre 2008 e 2018 e oito dos periódicos apresentados na Tabela 2 são os mesmos apresentados na Tabela 1.





Fonte: a autora, 2018.

Tabela 2 -	Principais pe	eriódicos con	n publicações	de técnicas	baseadas em	vetores de	suporte
	em CEP						

Periódico	Frequência absoluta	Frequência Percentual
Computers & Industrial Engineering	9	8,26%
Expert Systems with Applications	9	8,26%
Journal of Process Control	8	7,34%
International Journal of Production Research	7	6,42%
Quality and Reliability Engineering International	5	4,59%
Applied Soft Computing	5	4,59%
Journal of Intelligent Manufacturing	4	3,67%
IIE Transactions	3	2,75%
Communications in Statistics - Simulation and Computation	3	2,75%
ISA Transactions	3	2,75%
Neurocomputing	3	2,75%
Mathematical Problems in Engineering	3	2,75%
Outros	47	43,12%
Total	109	100,00%

Fonte: a autora, 2018.

Nenhuma das publicações apresentadas nesta seção utilizou processos com características da qualidade em atributos, proporção de não conformes e número de defeitos no processo.

Em muitos dos trabalhos apresentados os autores relataram que os sistemas utilizando as técnicas baseadas em vetores de suporte apresentaram desempenho similar ou superior, quando comparados com as cartas de controle tradicionais e os sistemas utilizando RNAs. Desta forma, as técnicas baseadas em vetores de suporte se apresentam como alternativas para o monitoramento de processos e são uma boa opção para a automatização do reconhecimento de padrões nas cartas de controle.

3.6 SELEÇÃO DOS PARÂMETROS PARA REGRESSÃO POR VETORES DE SUPORTE E REGRESSÃO POR VETORES DE RELEVÂNCIA

A SVR é utilizada para encontrar um preditor que aproxime bem os dados da amostra. A capacidade de generalização do modelo SVR depende da correta seleção dos parâmetros $C \in \varepsilon$ e do tipo de função kernel e seus parâmetros. Na RVR a capacidade de generalização do modelo depende da correta seleção do tipo de função kernel e seus parâmetros.

3.6.1 Métodos para seleção dos parâmetros para SVR e RVR

Conforme Hsu *et al.* (2010), para a determinação dos valores ótimos para os parâmetros da SVR os métodos mais utilizados são a busca em grade (*grid search*) e a validação cruzada (*cross validation*).

Em aplicações reais, muitos profissionais selecionam os parâmetros da SVR empiricamente por tentativa e erro ou utilizando a busca em grade em conjunto com a validação cruzada, devido a sua facilidade de implementação. Além de consumir muito tempo, esses procedimentos para selecionar os parâmetros da SVR podem não obter o melhor desempenho. Vários trabalhos apresentados na seção 3.5.2 utilizaram busca em grade e validação cruzada para a determinação dos melhores valores para os parâmetros das SVMs e SVR: Issam e Mohamed (2008), Cheng *et al.* (2011), Wang *et al.* (2009), Kheridi *et al.* (2010), Lu *et al.* (2011), Cheng e Lee (2012), Kao *et al.* (2016), Hwang (2016) e Kazemi *et al.* (2016). Como exemplos de aplicações das SVMs em outras áreas pode ser citado Alves (2012).

Outra abordagem utilizada em trabalhos apresentados na seção 3.5.2 é a aplicação de heurísticas para a escolha dos melhores valores para os parâmetros das SVMs e SVR.

Como exemplos de aplicação de AGs podem ser citados os trabalhos de Ranaee e Ebrahimzadeh (2011), Ebrahimzadeh *et al.* (2013), Zhou *et al.* (2015) e Zhang e Cheng (2015). O algoritmo de PSO foi utilizado por Ranaee *et al.* (2010), Du *et al.* (2012), Du *et al.* (2013), Du e Lv (2013), Zhiqiang e Yizhong (2010), Ebrahimzadeh *et al.* (2013), Zhang *et al.* (2015) e Yang (2016).

O trabalho de Du *et al.* (2012) utilizou *Simulated Annealing*. O algoritmo inspirado no comportamento de cucos (*Cuckoo search algorithm*) foi aplicado por Khormali e Addeh (2016). Colônia de abelhas foi aplicado por Ebrahimzadeh *et al.* (2013) e Gutierrez e Pham (2016).

Em aplicações reais de regressão, os parâmetros da RVR normalmente são selecionados empiricamente por tentativa e erro ou utilizando a busca em grade em conjunto com a validação cruzada, devido a sua facilidade de implementação. O procedimento de tentativa e erro foi utilizado por Verma *et al.* (2017) e Imani *et al.* (2018).

Outra abordagem utilizada para a otimização dos parâmetros da RVR é a aplicação de heurísticas. O algoritmo de PSO foi utilizado por He *et al.* (2017) e a ED foi utilizada por Zhang *et al.* (2015), He *et al.* (2017) e Liu (2017).

3.6.2 Evolução diferencial

O algoritmo de Evolução Diferencial (ED) é um algoritmo evolucionário (EA) para otimização global proposto por Storn e Price (1997). Conforme Das e Suganthan (2011), o algoritmo de ED possui as seguintes características, quando comparado

com outros algoritmos evolutivos: é mais fácil de aplicar, apresenta excelente performance e o número de parâmetros ajustaveis é relativamente menor.

A ED clássica ocorre em quatro passos principais, que entram em ciclo: a inicialização dos vetores, a mutação, o cruzamento e a seleção. De acordo com Chen e Chiang (2015), um indivíduo codifica uma solução para o problema de otimização e uma população é composta de muitos indivíduos. Em uma geração, novos indivíduos são produzidos através dos operadores de reprodução (cruzamento e mutação), e indivíduos bons sobrevivem para a próxima geração através do operador de seleção. A população evolui repetindo a reprodução e a seleção. Espera-se que os indivíduos se tornem cada vez melhores, o que também significa que as soluções se aproximam do ideal.

A versão original do algoritmo ED pode ser definida pelos seguintes elementos (STORN, 2008):

 A população: A população P de uma geração g possui Np vetores D-dimensionais (indivíduos x), cada um representando uma solução candidata para o problema de otimização, equação (24) (PRICE *et al.*, 2006)

$$P_{x,g} = (\mathbf{x}_{i,g}), \quad i = 0, 1, \dots, N_p - 1, \quad g = 0, 1, \dots, g_{max}$$

$$\mathbf{x}_{i,g} = (x_{j,i,g}), \quad j = 0, 1, \dots, D - 1$$
(24)

onde *Np* representa o número de vetores da população, *g* define o contador de gerações, e *D* a dimensionalidade, ou seja, o número de parâmetros.

2) A população inicial da ED é gerada aleatoriamente, conforme equação (25)

$$x_{j,i,g} = rand_{j} [0,1).(b_{j,U} - b_{j,L}) + b_{j,L}$$
(25)

Os vetores de inicialização *D*-dimensionais, $\mathbf{b}_{L} \in \mathbf{b}_{U}$ indicam os limites inferior e superior dos vetores de parâmetros $\mathbf{x}_{i,j}$. O gerador de números aleatórios $rand_{j}[0,1)$ retorna um número aleatório com distribuição uniforme dentro do intervalo [0,1]. O subscrito, *j*, indica que um novo valor aleatório é gerado para cada parâmetro.

3) Uma vez inicializado, o algoritmo de ED realiza a mutação e a recombinação da população para produzir uma população de vetores de tentativas *Np*. Para cada

indivíduo da população corrente ($\mathbf{x}_{i,g}$), um vetor mutante \mathbf{v} é gerado através da seleção aleatória de três outros vetores da geração atual (x_{r1}, x_{r2}, x_{r3}), equação (26)

$$\mathbf{v}_{i,g} = \mathbf{X}_{r1,g} + F.(\mathbf{X}_{r2,g} - \mathbf{X}_{r3,g})$$
(26)

De acordo com Storn e Price (1997), o fator de escala F é um fator com valor real e constante no intervalo [0, 2], que controla a amplificação da variação da diferença entre os vetores diferença.

4) O cruzamento (*crossover*) é introduzido para aumentar a diversidade dos indivíduos que sofreram a mutação. A variante clássica do aumento da diversidade é o cruzamento que mistura parâmetros do vetor de mutação v_{i,g}, e o vetor alvo x_{i,g} para gerar o vetor experimental u_{i,g}. A forma mais comum de cruzamento é definida pela equação (27)

$$\mathbf{v}_{i,g} = \mathbf{x}_{r1,g} + F.\left(\mathbf{x}_{r2,g} - \mathbf{x}_{r3,g}\right)$$
$$\mathbf{u}_{i,g} = \mathbf{u}_{j,i,g} = \begin{cases} \mathbf{v}_{j,i,g} & \text{se } \left(rand_{j}\left[0,1\right] \le Cr \right) \\ \mathbf{x}_{j,i,g} & \text{caso contrário} \end{cases}$$
(27)

A taxa de cruzamento $Cr \in [0,1]$ é definida pelo usuário e representa a probabilidade do vetor experimental herdar os valores das variáveis do vetor mutante (PRICE *et al.*, 2006).

5) Seleção: Para decidir se o vetor experimental $\boldsymbol{u}_{i,g}$ será membro da geração (g+1)ele é comparado ao vetor alvo $\boldsymbol{x}_{i,g}$. Caso o vetor $\boldsymbol{u}_{i,g}$ seja melhor que $\boldsymbol{x}_{i,g}$, a solução deste substitui a solução atual, caso contrário, o antigo $\boldsymbol{x}_{i,g}$ é mantido na população por pelo menos mais uma geração, conforme equação (28)

$$\mathbf{x}_{i,g+1} = \begin{cases} \mathbf{u}_{i,g} & \text{se } f\left(\mathbf{u}_{i,g}\right) \le f\left(\mathbf{x}_{i,g}\right) \\ \mathbf{x}_{i,g} & \text{caso contrário} \end{cases}$$
(28)

Conforme Price *et al.* (2006), ao comparar cada vetor experimental com o vetor alvo do qual ele herda parâmetros, o algoritmo ED integra mais fortemente a recombinação e seleção do que outros EAs.

De acordo com Santos *et al.* (2012), as estratégias da evolução diferencial podem variar de acordo com o tipo de indivíduo a ser modificado na formação do vetor mutante, o número de indivíduos considerados para a perturbação e o tipo de cruzamento a ser usado. A notação é definida por *DE/J/K/L*, onde:

- J representa o vetor a ser mutado, pode ser rand (um vetor da população escolhido aleatoriamente) ou best (o vetor do melhor indivíduo da população);
- *K* representa a quantidade de vetores diferença usados;
- *L* define o mecanismo de cruzamento utilizado, por exemplo *exp* para exponencial e *bin* para binomial.

A configuração dos parâmetros de controle do algoritmo ED é crucial para o desempenho do algoritmo e estes são, geralmente, os principais fatores que afetam a convergência da ED (PRICE *at el.*, 2006). Das *et al.* (2016) resumiram e organizaram as informações sobre os desenvolvimentos atuais da ED e apresentaram as recentes propostas de adaptação de parâmetros do algoritmo ED.

Segundo Storn e Price (1997), a ED é muito mais sensível à escolha do parâmetro *F* do que à escolha do parâmetro *Cr*. Os parâmetros de controle são, frequentemente, definidos arbitrariamente dentro de alguns intervalos pré-definidos ou obtidos por experimentos empíricos que determinam os melhores valores para esses parâmetros, processo este que demanda tempo e esforço do usuário.

Segundo Eiben *et al.* (2007), existem duas formas principais de definir os valores dos parâmetros da ED: ajuste dos parâmetros e controle dos parâmetros. O ajuste dos parâmetros é a abordagem normalmente praticada que tenta encontrar bons valores para os parâmetros antes de executar o algoritmo e, em seguida, executa o algoritmo usando esses valores, que permanecem fixos durante a execução. No controle dos parâmetros, os valores dos parâmetros são alterados durante a execução. Os métodos para alterar o valor de um parâmetro podem ser classificados em uma das três categorias:

- O controle dos parâmetros de maneira determinística ocorre quando o valor de um parâmetro é alterado utilizando alguma regra determinística. Essa regra modifica o parâmetro de maneira fixa e predeterminada.
- 2) O controle dos parâmetros de maneira adaptativa ocorre quando existe alguma forma de realimentação da busca que serve como entrada para um mecanismo usado para determinar a direção ou a magnitude da alteração no parâmetro.
- 3) O controle dos parâmetros de maneira auto-adaptativa ocorre quando os parâmetros a serem adaptados são codificados nos cromossomos e sofrem mutação e recombinação. Os melhores valores desses parâmetros codificados levam à melhores indivíduos, que por sua vez são mais propensos a sobreviver e produzir descendentes e, portanto, propagar esses melhores valores de parâmetros.

O algoritmo proposto por Brest *et al.* (2006), o algoritmo *jDE*, emprega um esquema auto-adaptativo para realizar o ajuste automático dos parâmetros de controle, fator de escala *F* e taxa de cruzamento *Cr*. O parâmetro de controle tamanho da população *Np* não é alterado durante a execução, porque este parâmetro está fortemente ligado a complexidade do problema em questão. Brest *et al.* (2006) implementaram a estratégia *DE/rand/1/bin*, que é uma das mais usadas na prática.

Segundo Brest *et al.* (2006), os parâmetros de controle *F* e *Cr* são ajustados usando a evolução e ambos são aplicados no nível individual. Os melhores valores desses parâmetros de controle (codificados) levam a melhores indivíduos que, por sua vez, são mais propensos a sobreviver e produzir descendentes e, assim, propagar esses melhores valores dos parâmetros. A solução é representada por um vetor $\mathbf{x}_{i,g}$ *D*-dimensional, onde *i* = 1,..., *Np*. Novos parâmetros ou fatores de controle $F_{i,g+1}$ e $Cr_{i,g+1}$ são calculados por meio da equação (29)

$$F_{i,g+1} = \begin{cases} F_i + rand_1 * F_u, & \text{se } rand_2 < \tau_1 \\ F_{i,g}, & \text{caso contrário} \\ Cr_{i,g} = \begin{cases} rand_3, & \text{se } rand_4 < \tau_2 \\ Cr_{i,g}, & \text{caso contrário} \end{cases}$$
(29)

Os fatores calculados na equação (29) produzem fatores $F \in Cr \text{ em um novo}$ vetor pai, onde *rand*, com $j \in \{1,2,3,4\}$, são valores uniformes e aleatórios pertencentes ao intervalo [0,1], $\tau_1 \in \tau_2$ representam as probabilidades de ajustar os fatores *F* e *Cr*, respectivamente. Neste método, o algoritmo pode alterar os parâmetros de controle com algumas probabilidades ($\tau_1 \in \tau_2$) e, depois disso, melhores parâmetros de controle são usados nas próximas gerações.

Nesta tese foi utilizado o algoritmo *jDE* proposto por Brest *et al.* (2006) e implementado por Conceição e Mächler (2015) no pacote *DEoptimR* do programa R[®]. A implementação de Conceição e Mächler (2015) difere do algoritmo ED proposto por Brest *et al.* (2006) principalmente no uso da estratégia de mutação *DE/rand/1/either-or* (PRICE *et al.*, 2006), combinação de *jitter* com *dither* (STORN, 2008) e por substituir imediatamente cada pai na população atual por sua melhor ou igual recém-gerada geração (BABU; ANGIRA, 2006), ao invés de atualizar a população atual com todas as novas soluções ao mesmo tempo como na ED clássica.

4 MÉTODO PROPOSTO

Esta pesquisa teve como objetivo geral desenvolver estratégias para a modelagem e o monitoramento de processos que medem variáveis não conformes às especificações aplicando RVR, SVR e RNAs. Os resíduos padronizados obtidos com os modelos de regressão do processo sob controle são traçados em cartas de controle para a detecção de alterações no processo, sendo analisada a sensibilidade das cartas de controle utilizando o NMA. Os modelos e as cartas de controle desenvolvidas são comparadas com os obtidos a partir de outras técnicas.

Um método de implantação de cartas de controle pode ser dividido em duas fases: a Fase I de análise retrospectiva e a Fase II de monitoramento do processo. Jones-Farmer *et al.* (2014) apresentam uma revisão sobre a análise de dados na Fase I e as perspectivas de utilização na melhoria de processos e no desenvolvimento de cartas de controle.

Na Fase I, utilizando os dados históricos do processo sob controle, é estimado o modelo de regressão que relacione a característica da qualidade com as variáveis de controle do processo, são obtidos os resíduos padronizados, são calculados os limites de controle da carta de controle e é verificada a estabilidade do processo no período em que os dados históricos foram coletados.

Na Fase II considera-se que os novos dados coletados possuem a mesma distribuição de probabilidade que os dados do processo sob controle e que podem ser ajustados pelo mesmo modelo de regressão obtido na Fase I. Nesta fase, novos dados são coletados e utilizados com o modelo obtido na Fase I, os resíduos padronizados são calculados e plotados na carta de controle com os limites de controle calculados na Fase I (WOODALL; MONTGOMERY, 1999).

A implementação da Fase I e da Fase II das cartas de controle dos resíduos baseadas em modelos utilizando RVR, SVR e RNAs segue as seguintes etapas: Fase I – Fase da análise retrospectiva:

- (a) Coleta dos dados do processo ou geração dos dados sem alterações nas características da qualidade.
- (b) Análise e preparação dos dados, compreendendo seleção das variáveis, normalização e divisão dos dados.
- (c) Desenvolvimento dos modelos de regressão utilizando RVR, SVR e RNAs.

- (d) Cálculo dos resíduos padronizados obtidos com os modelos.
- (e) Obtenção dos limites de controle das cartas de controle.
- (f) Desenvolvimento das cartas de controle dos resíduos padronizados. Se o processo não estiver sob controle devem ser analisadas as causas especiais. Se as causas especiais forem identificadas deve-se ajustar o modelo de regressão e, se não forem identificadas, devem ser coletados novos dados.

Fase II – Fase do monitoramento do processo:

- (a) Coleta de novos dados do processo.
- (b) Cálculo dos resíduos padronizados dos novos dados através dos modelos de regressão obtidos na Fase I.
- (c) Desenvolvimento das cartas de controle dos resíduos padronizados com os limites de controle calculados na Fase I. Se o processo não estiver sob controle devem ser identificadas e analisadas as causas especiais.

A Figura 12 apresenta o fluxograma simplificado das estapas de implementação da Fase I e da Fase II das cartas de controle baseadas em modelos de regressão.

4.1 FASE I – ANÁLISE RETROSPECTIVA

Na Fase I são coletados os dados, desenvolvidos os modelos de regressão e construídas as cartas de controle dos resíduos padronizados.

4.1.1 Coleta, análise e preparação dos dados

Conforme apresentado na Figura 12, a primeira etapa da Fase I é a coleta dos dados do processo ou a geração de dados das características da qualidade. Para os dados simulados, conforme a distribuição de probabilidade definida, deve ser gerado um conjunto de dados para as características da qualidade e as variáveis de controle simulando o comportamento de um processo e o tipo de correlação existente.



Figura 12 - Fluxograma das etapas de implementação da Fase I e da Fase II das cartas de controle baseadas em modelos de regressão

Fonte: a autora, 2018.

A etapa seguinte compreende a análise e preparação dos dados. Nesta etapa é realizada uma análise preliminar dos dados (cálculo da média, desvio padrão, entre outros), verificados os efeitos das variáveis de controle sobre as características da qualidade, selecionadas as variáveis de controle a serem utilizadas para o desenvolvimento do modelo de regressão e realizadas as transformações necessárias nas variáveis. Caso necessário, são realizadas medidas corretivas nos dados coletados dos processos. Para a seleção das variáveis de controle que apresentam relação com a característica da qualidade é realizado o teste de correlação de Pearson, sendo que são selecionadas as variáveis de controle com maiores coeficientes de correlação (R) com a característica da qualidade. O teste de correlação de Pearson é uma medida estatística que indica a relação entre duas variáveis (x e y), conforme apresentado na equação (30) (MONTGOMERY; RUNGER, 2003)

$$R = \frac{\sum_{i=1}^{n} (x_i - \overline{x}) (y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \overline{x})^2 \sum_{i=1}^{n} (y_i - \overline{y})^2}} \quad \text{onde} \quad \overline{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \quad \text{e} \quad \overline{y} = \frac{1}{n} \sum_{i=1}^{n} y_i \quad (30)$$

onde x_i e y_i , com i=1,...,n, são os valores medidos das variáveis, \overline{x} e \overline{y} são as médias aritméticas das variáveis e n é o número de pares de observações (x, y). O coeficiente de correlação (R) é um número que varia de -1 a +1, ou seja, $-1 \le R \le 1$.

Caso as variáveis selecionadas para a construção do modelo possuam valores mínimos e máximos muito diferentes, para evitar que uma variável predomine sobre a outra e para obter um bom desempenho do modelo, deve ser realizada a normalização dos dados no intervalo [0,1] ou [-1,1], conforme apresentado na equação (31)

$$t_{p} = \frac{X_{i} - X_{min}}{X_{max} - X_{min}} (t_{max} - t_{min}) + t_{min}$$
(31)

onde t_p é o valor normalizado, x_i é o valor observado, x_{min} é o valor mínimo e x_{max} é o valor máximo dos dados a serem normalizados, t_{min} é o valor mínimo e t_{max} é o valor máximo do intervalo de normalização.

Após a normalização dos dados, o conjunto total de dados do processo é dividido aleatoriamente em dois subconjuntos: o grupo de treinamento e o grupo de teste. Os dados de treinamento são utilizados para estimar os modelos de regressão utilizando RVR, SVR e RNAs e os dados de teste são utilizados para avaliar o poder preditivo dos modelos de regressão.

4.1.2 Desenvolvimento dos modelos de regressão

Após a coleta dos dados históricos do processo, da seleção das variáveis, da normalização dos dados e da divisão aleatória do conjunto de dados, são desenvolvidos os modelos de regressão utilizando RVR, SVR e RNAs. Os modelos de regressão desenvolvidos são avaliados por medidas de diagnósticos, selecionadas conforme cada caso. Caso os modelos gerados não sejam válidos, são realizados ajustes nos modelos de regressão.

Utilizando as RNAs para obtenção dos modelos de regressão devem ser seguidas as etapas:

- (a) Definição da arquitetura;
- (b) Definição do método de seleção dos parâmetros;
- (c) Seleção dos melhores parâmetros;
- (d) Simulações de treinamento e de teste;
- (e) Análise do desempenho utilizando medidas de diagnósticos;
- (f) Validação do modelo.

A arquitetura de RNAs utilizada nesta pesquisa é a MLP com diferentes algoritmos de treinamento e a obtenção dos melhores parâmetros para a RNA é realizada pelo processo de análise dos erros.

Para os modelos de regressão utilizando RVR e SVR devem ser seguidas as etapas:

- (a) Definição do método de seleção dos parâmetros;
- (b) Escolha das funções kernel;
- (c) Seleção dos melhores parâmetros;
- (d) Simulações de treinamento e teste;
- (e) Análise do desempenho utilizando medidas de diagnósticos;
- (f) Validação do modelo.

Para a seleção dos melhores parâmetros do modelo RVR (escolha do kernel e obtenção de seus parâmetros) e do modelo SVR (obtenção dos parâmetros $C \in \varepsilon$ e escolha do kernel e seus parâmetros) é utilizada a busca em grade em conjunto com a validação cruzada, sendo que os melhores parâmetros são os que resultam no menor erro médio quadrático (*Mean Squared Error*, MSE) da função com os dados de treinamento.

Um algoritmo de ED auto-adaptativo também é utilizado para a obtenção dos melhores parâmetros do modelo RVR e do modelo SVR. Nenhum dos trabalhos apresentados na seção 3.5.2 utilizou ED para a determinação dos melhores valores para os parâmetros do modelo SVR.

Nesta tese foi utilizado o algoritmo de ED auto-adaptativo denominado *jDE* do pacote *DEoptimR* do programa R[®] (BREST *et al.*, 2006, CONCEIÇÃO; MÄCHLER, 2015). O algoritmo *jDE* emprega um esquema auto-adaptativo para realizar o ajuste automático dos parâmetros de controle *F* e *Cr* da ED. O tamanho da população (*Np*) não é alterado durante a execução do algoritmo *jDE*, sendo que este foi definido como $Np = 10 \times n_p$, onde n_p é o número de parâmetros do modelo RVR ou SVR. Os critérios de parada utilizados foram o número máximo de iterações igual a $200 \times n_p$ e a tolerância igual a 1×10^{-7} .

Para o desenvolvimento dos modelos de regressão baseados em SVR é utilizada a biblioteca para Máquinas de Vetores de Suporte denominada LIBSVM, desenvolvida por Chang e Lin (2011). O LIBSVM é um programa integrado que pode ser utilizado para problemas de classificação e de regressão com vetores de suporte e possui interfaces para diversas ferramentas, tais como Matlab[®] e R[®].

A Figura 13 apresenta o fluxograma simplificado para o desenvolvimento dos modelos RVR e SVR otimizados pelo algoritmo ED, incluindo as etapas de coleta dos dados históricos do processo, seleção das variáveis, normalização dos dados e divisão aleatória do conjunto de dados.

Os passos sugeridos para a obtenção dos modelos RVR e SVR com parâmetros otimizados pelo algoritmo ED são os seguintes:

- Passo 1: Coletar os dados do processo, selecionar as variáveis, normalizar os dados e dividir aleatoriamente o conjunto de dados em duas partes (dados de treinamento e dados de teste).
- Passo 2: Selecionar os dados de treinamento. Definir os parâmetros iniciais do kernel do modelo RVR ou os parâmetros livres (C, ε e parâmetros do kernel) do modelo SVR.
- Passo 3: Definir o tamanho da população *Np*, sendo *Np* = $10 \times n_p$ onde n_p é o número de parâmetros do modelo RVR ou SVR, e definir o critério de parada: número máximo de iterações ($200 \times n_p$) e tolerância (1×10^{-7}).

Passo 4: Treinar o modelo RVR ou SVR e calcular o valor da função objetivo. A função objetivo é definida como a raiz do erro médio quadrático (*Root Mean Squared Error*, RMSE), equação (32), e o objetivo é minimizar RMSE

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - y_i)^2}$$
(32)

onde y_i é o valor medido no processo, y_t é o valor predito pelo modelo de regressão e *n* é o número de observações usados na obtenção do modelo.

- Passo 5: Se o número máximo de iterações ou a tolerância definidos no critério de parada for atingido, o menor RMSE é selecionado e são obtidos os melhores parâmetros estimados para o modelo RVR ou SVR.
- Passo 6: Treinar o modelo RVR ou SVR com os melhores parâmetros estimados e obter o modelo otimizado.
- Passo 7: Obter os valores preditos com os dados de treinamento e de teste usando os modelos RVR ou SVR otimizados.
- Passo 8: Utilizando os valores preditos e os valores dos resíduos, realizar a análise de desempenho dos modelos de regressão por medidas de diagnósticos.
- Passo 9: Se o modelo for válido, este pode ser utilizado para predizer a característica da qualidade do processo analisado.

Os melhores parâmetros selecionados são utilizados para o treinamento dos modelos e, após o treinamento, é realizada a predição com os dados de teste usando os modelos RVR e SVR otimizados. Os modelos RVR e SVR são usados para predizer os valores futuros da característica da qualidade.

4.1.3 Medidas de diagnóstico dos modelos de regressão

Para a análise da adequação dos modelos de regressão desenvolvidos foram utilizados métodos gráficos, testes estatísticos, cálculos de erros e cálculo do coeficiente de determinação. Nesta tese são utilizados os cálculos de erros e o cálculo do coeficiente de determinação para a realização de uma análise abrangente da adequação dos modelos de regressão desenvolvidos, pois muitas publicações da seção 3.5 utilizam somente um destes critérios de análise.



Figura 13 - Fluxograma para implementação dos modelos RVR e SVR otimizados pelo algoritmo ED

Fonte: a autora, 2018.

Modelos de regressão com bom desempenho estatístico apresentam pequena discrepância entre os dados reais do processo e seus respectivos valores preditos. Para a situação do processo sob controle os erros de predição, ou resíduos, obtidos com o modelo de regressão devem ser normal e independentemente distribuídos com média zero e variância aproximadamente constante. Para a situação fora de controle, a alteração do valor da média do processo é refletida na alteração do valor da média dos resíduos.

O resíduo (e_t) é a diferença entre o valor real (y_i) e o valor estimado pelo modelo (y_t), conforme apresentado na equação (33). O resíduo padronizado (e_{pt}) utiliza a variância estimada do resíduo ($\hat{\sigma}^2$), conforme apresentado na equação (34), onde *n* representa o número de observações e *h* o número de variáveis de controle

$$\boldsymbol{e}_t = \boldsymbol{y}_i - \boldsymbol{y}_t \tag{33}$$

$$e_{pt} = \frac{(y_i - y_t)}{\sqrt{\hat{\sigma}^2}} = \frac{(y_i - y_t)}{\sqrt{\frac{\sum_{t=1}^n (y_i - y_t)^2}{n - h - 1}}} = \frac{e_i}{\sqrt{\frac{\sum_{t=1}^n e_i^2}{n - h - 1}}}$$
(34)

Os resíduos do modelo de regressão devem ser analisados com relação a normalidade, independência e variância. A verificação da normalidade dos resíduos pode ser verificada por meio do gráfico de probabilidade normal e pelo teste de normalidade de Shapiro-Wilk, Anderson-Darling, dentre outros (MONTGOMERY; RUNGER, 2003).

A validade da suposição de que os resíduos não são correlacionados pode ser verificada por meio de gráficos de autocorrelação dos resíduos do modelo e do gráfico de resíduos contra a ordem de coleta dos dados, sendo que a presença de padrões específicos pode indicar que os resíduos são autocorrelacionados (WERKEMA; AGUIAR, 1996).

A estatística de Durbin-Watson pode ser utilizada para verificação da hipótese de que os resíduos são independentes, conforme equação (35)

$$d_{w} = \frac{\sum_{i=2}^{n} (e_{i} - e_{i-1})^{2}}{\sum_{i=1}^{n} e_{i}^{2}}$$
(35)

onde *e*_i é o resíduo associado à *i*-ésima observação e *n* é o número de onservações.

As decisões são feitas comparando o valor de d_w ($0 \le d_w \le 4$) com os valores do limite superior (d_u) e do limite inferior (d_L) da Tabela de Durbin-Watson (WERKEMA; AGUIAR, 1996). Assim:

- se 0 ≤ d_w ≤ d_L autocorrelação positiva dos resíduos (dependência);
- se $d_L \le d_w \le d_U$ o teste é inconclusivo;
- se $d_U \le d_w \le 4 d_U$ os resíduos não são autocorrelacionados (independência);
- se $4 d_U \le d_w \le 4 d_L o$ teste é inconclusivo;
- se $4 d_L \le d_w \le 4$ autocorrelação negativa dos resíduos (dependência).

A verificação da homoscedasticidade, ou seja, se os resíduos possuem variância aproximadamente constante, pode ser analisada pelo teste de Levene e pelo gráfico dos resíduos versus os valores estimados. Os resíduos devem ser aleatórios e não devem seguir padrões específicos, conforme os apresentados em Montgomery e Runger (2003) e Werkema e Aguiar (1996).

O grau de concordância entre os resíduos e os valores preditos pode ser estimado pela média (\overline{e}_t) e pelo desvio padrão (σ_{e_t}) dos resíduos, sendo esperado que em torno de 95% dos resíduos estejam compreendidos em $\overline{e}_t \pm 2\sigma_{e_t}$. Na sequência, deve-se verificar a existência de possíveis *outliers*, sendo que um *outlier* é um resíduo maior do que $\overline{e}_t \pm 3\sigma_{e_t}$.

Conforme Werkema e Aguiar (1996), pode-se utilizar o gráfico dos resíduos padronizados versus os valores estimados e, se os resíduos padronizados são i.i.d., em torno de 95% dos resíduos devem estar no intervalo [-2,2] e 99% no intervalo [-3;+3]. Os resíduos fora destes intervalos (*outliers*) devem ser analisados para verificação da presença de causas especiais e, após esta análise, mantidos ou retirados. Após a análise dos *outliers* realiza-se a estimação de um novo modelo e, quando não houver mais *outliers*, o modelo é considerado válido (WERKEMA; AGUIAR, 1996).

Estratégias de minimização dos resíduos podem ser utilizadas para avaliar a capacidade de generalização dos modelos desenvolvidos. A equação (36) apresenta o erro médio absoluto (*Mean Absolute Error*, MAE), a equação (37) o erro médio quadrático (*Mean Squared Error*, MSE), a equação (38) a raiz do erro médio quadrático (*Root Mean Squared Error*, RMSE) e a equação (39) apresenta o erro percentual absoluto médio (*Mean Absolute Percentage Error*, MAPE), onde *n* representa o número de observações

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |y_i - y_i| = \frac{1}{n} \sum_{i=1}^{n} |e_i|$$
(36)

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (y_i - y_i)^2 = \frac{1}{n} \sum_{i=1}^{n} (e_i)^2$$
(37)

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - y_i)^2} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (e_i)^2}$$
(38)

$$MAPE = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{y_i - y_t}{y_t} \right| = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{e_t}{y_t} \right|$$
(39)

O coeficiente de determinação (R^2) é uma medida global de qualidade do ajuste e pode ser interpretado como a proporção da variabilidade presente nas observações da característica da qualidade que é explicada pelas variáveis de controle no modelo de regressão ajustado dos dados. Werkema e Aguiar (1996) apresentam que o coeficiente de determinação deve ser empregado com cuidado pois o mesmo aumenta com a inclusão de novas variáveis independentes.

O coeficiente de determinação (R^2) pondera matematicamente a separação da característica da qualidade (y_i) nas suas duas partes distintas: a parte representada pela estimação (y_i) e a parte advinda do resíduo (e_i). O coeficiente de determinação (R^2) é um número que varia de 0 a 1, ou seja, $0 \le R^2 \le 1$ (SAMOHYL, 2009).

O coeficiente de determinação (R^2) pode ser determinado utilizando-se a soma dos quadrados total (SQT), a soma dos quadrados dos resíduos da regressão (SQE) e a soma dos quadrados da regressão (SQR), conforme equação (40)

$$R^{2} = \frac{SQR}{SQT} = 1 - \frac{SQE}{SQT} = 1 - \frac{\sum_{i=1}^{n} (y_{i} - y_{i})^{2}}{\sum_{i=1}^{n} (y_{i} - \overline{y_{i}})^{2}}$$
(40)

O Coeficiente de Informação de Akaike corrigido (*corrected Akaike Information Criterion*, AICc) pode ser utilizado para verificar a adequação do modelo (SANT'ANNA, 2009). A decisão quanto ao melhor modelo ajustado é realizado escolhendo o menor valor de AICc, obtido por meio da equação (41)

$$AICc = In\left(\frac{\sum_{t=1}^{n} e_{t}^{2}}{n}\right) + \frac{2n(n_{p}+1)}{n - n_{p}-2}$$
(41)

onde *n* representa o número de observações e n_p representa o número de parâmetros do modelo. Para as RNAs o número de parâmetros do modelo corresponde ao número de pesos e de *bias* a serem ajustados, sendo $n_p = I(m+1) + j(I+1)$, onde *m* é o número de entradas, *I* é o número de neurônios na camada oculta e *j* é o número de neurônios na camada de saída (YU *et al.*, 2006).

O *box-percentile plot* foi utilizado para comparar os modelos de regressão desenvolvidos e na etapa de análise de sensibilidade das cartas de controle. O *box-percentile plot* foi desenvolvido por Esty e Banfield (2003) e combina as virtudes dos *boxplots* (facilidade de interpretação e capacidade de comparar vários conjuntos de dados simultaneamente) com as dos gráficos de percentil (exibir todos os dados e sem escolha arbitrária na construção).

A ideia *box-percentile plot* é usar a largura (como no *boxplot*) para enfatizar o meio dos dados e continuar a usar a largura (mas não de forma arbitrária) para dar menos ênfase aos dados mais extremos (como são dadas menos ênfase as caudas e *outliers* nos *boxplots*). No *boxplot* o comprimento das linhas fora do retângulo (chamadas de *whisquers*) informam sobre a cauda da distribuição.

Assim, as "caixas" do *box-percentile plot* são largas no meio (como as "caixas" do *boxplot*), estreitas a partir do meio (como são as caudas) e muito estreitas nos extremos. Em qualquer altura, a largura da "caixa" irregular é proporcional ao percentil

dessa altura até o 50º percentil e, acima do 50º percentil, a largura é proporcional a 100 menos o percentil. Assim, a largura em qualquer altura é proporcional à porcentagem das observações que são mais extremas nessa direção (ESTY; BANFIELD, 2003).

Como nos *boxplots*, a mediana (50° percentil, 2° quartil - Q₂), o 25° percentil (1° quartil - Q₁) e o 75° percentil (3° quartil - Q3) são marcados com segmentos de linha (ESTY; BANFIELD, 2003, MONTEGOMERY; RUNGER, 2003). Na Figura 14, a mediana, o 1° quartil (Q₁) e o 3° quartil (Q3) foram destacados com segmentos de linha nas cores vermelha e azul, respectivamente.

No *boxplot* um ponto é considerado um valor extremo ou valor atípico (*outlier*) quando estiver fora do intervalo compreendido ente o limite inferior $(Q_1 - 1, 5(Q_3 - Q_1))$ e o limite superior $(Q_3 + 1, 5(Q_3 - Q_1))$, conforme apresentado no exemplo da Figura 14 (MONTEGOMERY; RUNGER, 2003).

A Figura 14 (a) apresenta o *box-percentile plot* e a Figura 14 (b) apresenta o *boxplot* para a mesma distribuição Normal com $\mu = 0$, $\sigma = 1$ e 500 observações, onde verifica-se que no *box-percentile plot* a largura em qualquer altura é proporcional à porcentagem das observações.



Figura 14 - Box-percentile plot e boxplot para uma distribuição Normal

4.1.4 Desenvolvimento das cartas de controle

O modelo de regressão validado é utilizado para o desenvolvimento da carta de controle dos resíduos padronizados, com a definição dos limites de controle e a verificação se o processo está sob controle. Caso o processo não esteja sob controle, devem ser analisadas as causas especiais e tomadas as medidas necessárias para a correção destas. Uma vez identificadas e eliminadas as causas especiais, o processo se torna estável e deve ser desenvolvido novo modelo de regressão com os novos dados e construída a nova carta de controle. Na sequência, realiza-se o monitoramento do processo.

As estimativas de média (\overline{e}_{pt}) e do desvio padrão $(\sigma_{e_{pt}})$ dos resíduos padronizados dos dados do processo sob controle são utilizadas para calcular o limite superior de controle (LSC), o limite inferior de controle (LIC) e a linha central (LM) da carta de controle univariada de Shewhart, conforme equações (42), (43) e (44) (SANT´ANNA, 2009)

$$LSC = e_{\rho t} + \lambda \sigma_{e_{\sigma t}} \tag{42}$$

$$LM = \overline{e}_{\rho t} \tag{43}$$

$$LIC = \overline{e}_{\rho t} - \lambda \sigma_{e_{\alpha t}} \tag{44}$$

onde λ é a constante que define a largura dos limites de controle correspondente a uma região de controle (1- α) e um número médio desejado de amostras até um alarme falso (NMA₀). Habitualmente, utiliza-se o valor de λ igual a 3 (três) devido à aproximação pela distribuição Normal, correspondendo a NMA₀ = 370,4.

Dentre os testes de estabilidade apresentados por Western Electric (1958), nesta pesquisa são utilizados os seguintes testes:

- Teste 1: um ponto acima do LSC ou abaixo do LIC;
- Teste 2: dois em três pontos consecutivos se localizam além dos linites de alerta de dois desvios padrão;
- Teste 3: uma seqüência de oito pontos localizados do mesmo lado da linha.

Caso ocorra um destes três tipos de causas especiais, o processo é considerado como fora de controle estatístico e, desta forma, as causas especiais que alteraram o estado do processo devem ser investigadas. Se as causas especiais forem detectadas, os pontos fora de controle devem ser desconsiderados, um novo modelo de regressão deve ser obtido e os limites de controle devem ser calculados novamente. Caso não existam novos pontos fora de controle, os modelos de regressão e os limites de controle são utilizados na Fase II do método proposto.

4.2 FASE II – MONITORAMENTO DO PROCESSO

A Fase II é o monitoramento propriamente dito do processo, sendo assumido que o modelo de regressão estimado na Fase I está correto e que os novos dados a serem monitorados seguem a mesma distribuição de probabilidade que os dados da Fase I.

Após o cálculo dos resíduos padronizados dos novos dados através dos modelos de regressão obtidos na Fase I, são construídas as cartas de controle dos resíduos com os limites de controle calculados na Fase I. Se após a realização dos testes de estabilidade o processo estiver fora de controle estatístico, devem ser identificadas e analisadas as causas especiais.

4.3 ANÁLISE DE SENSIBILIDADE DAS CARTAS DE CONTROLE

Na próxima etapa é realizada a análise de sensibilidade das cartas de controle através de um estudo de simulação de Monte Carlo para a obtenção do número médio de amostras (NMA) até a detecção de causas especiais de cada carta de controle, para um processo sob controle e um processo fora de controle, onde são induzidas diferentes alterações na média da característica da qualidade.

Segundo Costa *et al.* (2010), uma medida de sensibilidade de uma carta de controle pode ser obtida por meio do NMA, que visa determinar a taxa de alarmes falsos da carta de controle e a sensibilidade da carta em detectar alterações pré-definidas no processo.

Quando o processo está sob controle estatístico, o NMA₀ indica o número médio de pontos necessários para a ocorrência do primeiro alarme falso na carta de controle. Quando o processo está fora de controle estatístico, o NMA₁ indica o número médio de amostras até a presença de causas especiais (MONTGOMERY, 2004; COSTA *et al.*, 2010).

Um maior valor de NMA₀ indica uma menor probabilidade de alarmes falsos no processo, ou seja, um valor alto de NMA₀ é um indício de que serão necessárias poucas interrupções no processo quando este está sob controle estatístico. Um menor valor de NMA₁ representa um menor número médio de amostras coletadas até a detecção da alteração induzida, ou seja, um baixo valor de NMA₁ está diretamente relacionado ao tempo necessário para a detecção de uma causa especial (MONTGOMERY, 2004; COSTA *et al.*, 2010).

O NMA₀ pode ser escrito em função da probabilidade de erro do tipo I (α), que é a probabilidade da carta de controle detectar uma alteração no modelo de referência monitorado quando o processo está sob controle estatístico (MONTGOMERY, 2004), conforme equação (45)

$$NMA_{o} = \frac{1}{\alpha} \quad e \quad \alpha = 1 - P[LIC < x < LSC | p = p_{o}]$$
(45)

onde p_o é a média da característica da qualidade de um processo sob controle, LSC é o limite superior de controle e LIC é o limite inferior de controle.

O NMA₁ pode ser escrito em função da probabilidade de erro do tipo II (β), que é a probabilidade da carta de controle não detectar uma alteração no modelo de referência monitorado quando o processo está fora de controle estatístico (MONTGOMERY, 2004), conforme equação (46)

$$NMA_{1} = \frac{1}{(1-\beta)} \quad e \quad \beta = P[LIC < x < LSC | p = p_{1}], \text{ com } p_{1} = p_{0} + \delta$$
(46)

onde δ é a alteração induzida no processo, p_0 é a média da característica da qualidade de um processo sob controle, p_1 é a média da caracterísitica da qualidade

de um processo fora de controle, LSC é o limite superior de controle e LIC é o limite inferior de controle.

Dyer (2016) apresenta que o NMA tem sido tipicamente usado para quantificar o desempenho da carta de controle. Um critério de desempenho alternativo é a função distribuição cumulativa (*Cumulative Distribution Function*, CDF) do número de amostras coletadas até a presença de um sinal e o número mediano de amostras (*Median Run Length*, MRL). O CDF mede a proporção cumulativa dada pelo *i*-ésimo período após a alteração e indica a probabilidade de uma variável ter um valor menor ou igual a um valor selecionado. O CDF caracteriza completamente a distribuição do número de amostras, enquanto o NMA é apenas a média. O MRL é definido como o número mediano (50º percentil) de amostras coletadas até a presença de um sinal.

Para uma distribuição assimétrica, tal como a distribuição do número de amostras que segue uma distribuição geométrica, a média é maior do que a mediana. Portanto, a média não é uma representação adequada do centro da distribuição e os percentis da distribuição do número de amostras precisam ser calculados para complementar o NMA (TEOH *et al.*, 2016). Assim, o MRL pode ser usado em conjunto com o NMA e o CDF.

O desvio padrão do número de amostras (*Standard Error of the Run Length*, SRL) e o MRL podem ser escritos como uma função da probabilidade de erro do tipo I (α) (DYER, 2016). O 100° percentil ($0 < p^{th} < 1$) é definido como o menor inteiro *t* de modo que o CDF de *t* é pelo menos igual a p^{th} (CHAKRABORTI, 2007). O valor do p^{th} percentil do número de amostras ($RL_{p^{th}}$) pode ser escrito como uma função do α e p^{th} percentil. O $RL_{p^{th}}$ calculado é aproximado para o inteiro mais próximo com um valor mínimo de um. O 50° percentil ($p^{th} = 0,5$) é o MRL. A equação (47) apresenta as equações para o SRL, MRL e $RL_{p^{th}}$

$$SRL = \frac{\sqrt{1-\alpha}}{\alpha} \qquad MRL = \frac{ln(0,50)}{ln(1-\alpha)} \qquad RL_{p^{th}} = \frac{ln(1-p^{th} percentil)}{ln(1-\alpha)}$$
(47)

5 ESTUDO EXPERIMENTAL

A implementação da modelagem e do monitorameno utilizando cartas de controle dos resíduos baseadas em modelos RVR, SVR e RNAs foi aplicada ao estudo de caso experimental da proporção da concentração de fósforo nas ligas metálicas produzidas por uma empresa siderúrgica. A Figura 12 apresenta o fluxograma simplificado das estapas de implementação da Fase I e da Fase II das cartas de controle baseadas em modelos de regressão.

A qualidade de produtos e processos é um desafio constante nas organizações industriais. Um dos principais fatores que afetam a qualidade dos produtos nas empresas siderúrgicas é a existência de contaminantes nas ligas de aço.

O processo de refino adotado pela empresa utiliza o oxigênio de alta pureza para realizar a redução dos níveis de carbono do Ferro Manganês Alto Carbono (FeMnAC), produzindo o Ferro Manganês Médio Carbono (FeMnMC), que é mais valorizado pelo mercado. Durante o processo de refino, ocorre a alteração dos teores dos outros elementos, dentre estes, o fósforo.

Uma das principais impurezas encontradas no processo siderúrgico é a presença de fósforo (*phosphorus*, P). As ligas de ferromanganês são uma das principais fontes de contaminação de fósforo na fabricação de aço, o que demanda o uso limitado deste tipo de liga ao longo do processo (UM *et al.*, 2014). Altos níveis de fósforo causam danos às características físicas das ligas de aço, impactando na qualidade deste produto. Portanto, é necessário o uso de um método que permita a análise e o controle do processo, para gerar alta confiabilidade do processo e produtos de alta qualidade (BARELLA *et al.*, 2017).

Conforme Chaudhary *et al.* (2001), os principais efeitos do fósforo sobre o aço são: (i) aumento da dureza; (ii) diminuição da ductibilidade; (iii) fragilidade a frio; (iv) propicia o surgimento de *ghost lines* em aços com teores mais elevados de carbono; (v) em altas concentrações o fósforo aumenta a fragilidade a quente.

Pedrini e Caten (2010) discutem que esse processo consiste na reação de descarburação entre o metal líquido com a injeção de oxigênio no banho metálico. Para realizar o processo de desfosforação é dissolvido cal (CaO) durante a descarburação na intenção de reduzir a proporção de fósforo no produto final.

O estudo do processo de refino das ligas metálicas de FeMnMC consiste em 21 variáveis de controle definidas como relevantes para a modelagem do processo de desfosforação, as quais foram agrupadas como: composição das ligas metálicas, composição da escória, composição das cargas solidas e níveis de basicidade: binária, quaternaria e ótica, conforme apresentado no Quadro 3. A característica da qualidade analisada é a proporção da concentração de fósforo nas ligas metálicas.

Grupo	Variáveis	Unid
Composição da liga	Si*, P*, Fe*, Ti*, C*, O ₂	%
Composição da escória	MgO, MnO, CaO, SiO ₂ , Al ₂ O ₃ , BaO, K ₂ O, TiO ₂ , FeO	wt-%
Composição das cargas sólidas	Initial, Liquid, Dross	Nm ³
Níveis de basicidade	BB, BQ, BO	Nm ³
Característica da qualidade	Р	%

Quadro 3 - Variáveis iniciais do processo em estudo

Fonte: a autora, 2018.

Foram analisados dados de 257 ligas metálicas do processo. Para a seleção das variáveis relevantes para a construção do modelo de regressão foi realizado o teste de correlação de Pearson, conforme equação (30).

Após o teste de correlação de Pearson foram selecionadas as quatro variáveis de entrada (controle) que apresentaram coeficiente de correlação com a proporção da concentração de fósforo no final do processo igual ou superior à 0,2: quantidade inicial de fósforo (P^*), teor inicial de Carbono (C^*), teor de Óxido de Manganês (MnO) e basicidade líquida (*Liquid*). As demais variáveis de entrada do processo apresentaram coeficiente de correlação inferior à 0,2 e não foram utilizadas para modelar o processo. A variável de saída (característica da qualidade) do processo é a proporção da concentração de fósforo (P) nas ligas metálicas.

A Figura 15 apresenta os histogramas com as distribuições de probabilidade Normal sobrepostas para as quatro variáveis de controle selecionadas: (a) quantidade inicial de fósforo (P^*), (b) teor inicial de Carbono (C^*), (c) teor de Óxido de Manganês (MnO), (d) basicidade líquida (*Liquid*).

A Tabela 3 apresenta informações estatísticas das variáveis de controle selecionadas e da característica da qualidade. A Figura 16 apresenta o gráfico da proporção da concentração de fósforo no final do processo de refino das 257 ligas metálicas.





Tome. a autora, 2010.

Tabela 3 - Informa	ções estatísticas	das variáveis	selecionadas
--------------------	-------------------	---------------	--------------

		Variável	Média	Desvio padrão	Máximo	Mínimo	Correlação
	X 1	P*	0,216	0,026	0,280	0,155	0,707
Variáveis de	X 2	C*	6,752	0,147	7,070	6,077	-0,228
controle	X 3	MnO	39,062	5,971	51,993	24,202	0,259
	X 4	Liquid	9,866	1,547	12,700	5,510	-0,301
Característica da qualidade	у	Р	0,260	0,026	0,331	0,204	1,000

Fonte: a autora, 2018.

O teste de Shapiro-Wilk com 95% de nível de confiança foi utilizado para a verificação da normalidade da proporção da concentração de fósforo e a hipótese de normalidade foi rejeitada (*p-value* < 0,05). Por meio da análise da distribuição dos dados verificou-se que a distribuição de probabilidade Beta é a que melhor representa a distribuição de proporção da concentração de fósforo.

A autocorrelação da proporção da concentração de fósforo foi examinada pelo gráfico da função de autocorrelação e pela estatística de Durbin-Watson (d_w) , calculada pela equação (35). Os valores do limite superior (d_u) e limite inferior (d_L) da Tabela de Durbin-Watson, para 257 observações e nível de significância de 5% são: $d_L = 1,73$ e $d_U = 1,78$. O valor obtido pela estatística de Durbin-Watson foi

 $d_w = 1,64$. Desta forma, $0 \le d_w \le d_L$ e os dados são considerados com autocorrelação positiva.





A Figura 17 (a) apresenta o histograma com a distribuição de probabilidade Beta sobreposta, a Figura 17 (b) apresenta o gráfico de probabilidade normal, a Figura 17 (c) apresenta o gráfico de autocorrelação e a Figura 17 (d) apresenta o gráfico de autocorrelação parcial da proporção da concentração de fósforo.

Conforme apresentado na Tabela 3, as variáveis de controle selecionadas possuem valores mínimos e máximos muito diferentes como, por exemplo, o valor máximo de *MnO* é quase duzentas vezes maior do que o valor máximo de P^* . Para evitar que uma variável predomine sobre a outra e para obter um bom desempenho dos modelos foi realizada a normalização dos dados no intervalo [0,1], utilizando a equação (31).

Após a fase de pré-processamento dos dados, o conjunto total de dados do processo foi dividido aleatoriamente em dois subconjuntos: 205 (80%) dados para o grupo de treinamento e 52 (20%) para o grupo de teste.



Figura 17- (a) Histograma com distribuição Beta sobreposta, (b) gráfico de probabilidade normal, (c) gráfico de autocorrelação e (d) gráfico de autocorrelação parcial da proporção da concentração de fósforo

Fonte: a autora, 2018.

5.1 MODELOS BASEADOS EM REDES NEURAIS ARTIFICIAIS

Para o desenvolvimento dos modelos baseados em RNAs foi utilizado o programa R[®]. A arquitetura selecionada foi a Perceptron Multicamadas (*Multilayer Perceptron*, MLP), que é uma rede totalmente conectada.

As simulações foram realizadas com a rede MLP com quatro variáveis de entrada, uma camada oculta e uma variável de saída. As variáveis de entrada do

modelo são: quantidade inicial de fósforo (P^*), teor inicial de Carbono (C^*), teor de Óxido de Manganês (MnO) e basicidade líquida (*Liquid*). A variável de saída do modelo é a proporção da concentração de fósforo (P) nas ligas metálicas. O conjunto total de dados do processo foi dividido aleatoriamente em dois subconjuntos: 205 (80%) dados para o grupo de treinamento e 52 (20%) para o grupo de teste.

Para as simulações, as funções de ativação utilizadas na camada oculta e na camada de saída foram: linear, gaussiana, logística e tangente hiperbólica. Os algoritmos de treinamento utilizados foram: *backpropagation*, *backpropagation* com *momentum*, Levenberg-Marquardt e gradiente conjugado escalonado (*Scaled Conjugate Gradiente*, SCG).

Após as simulações, a RNA do tipo MLP selecionada utiliza: algoritmo de treinamento Levenberg-Marquardt, função de ativação logística na camada oculta, função de ativação linear na camada de saída, taxa de aprendizado de 0,01, taxa de momentum de 0,1 e número de épocas de 500.

O número de neurônios da camada oculta foi obtido analisando-se o MSE, o RMSE, o MAPE e o AICc. A Tabela 4 apresenta os resultados de algumas simulações, sendo que foram selecionados 11 neurônios para a camada oculta.

Após a seleção da arquitetura e dos parâmetros, a RNA foi teinada e, após a fase de treinamento, o modelo obtido foi utilizado para predizer a característica da qualidade utilizando os dados do grupo de treinamento e de teste.

Índice de	Número de neurônios							
desempenho	10	11	12	13	14			
MSE	0,000239	0,000235	0,000266	0,000266	0,000282			
RMSE	0,0155	0,0153	0,0163	0,0163	0,0168			
MAPE (%)	4,64	4,72	4,97	4,82	4,98			
AICc	166	192	221	252	287			

Tabela 4 -	Resultados	das	simulações	para a	s RNAs
------------	------------	-----	------------	--------	--------

Fonte: a autora, 2018.

5.2 MODELOS BASEADOS EM REGRESSÃO POR VETORES DE SUPORTE

Para o desenvolvimento dos modelos baseados em SVR foi utilizado o programa R[®] e foram utilizadas as funções kernel do Quadro 2. O algoritmo ED é

utilizado para a minimização da função objetivo, que é definida como o RMSE dos dados de treinamento, dado pela equação (32).

As variáveis de entrada do modelo são: quantidade inicial de fósforo (P^*), teor inicial de Carbono (C^*), teor de Óxido de Manganês (MnO) e basicidade líquida (*Liquid*). A variável de saída do modelo é a proporção da concentração de fósforo (P) nas ligas metálicas. O conjunto total de dados do processo foi dividido aleatoriamente em dois subconjuntos: 205 (80%) dados para o grupo de treinamento e 52 (20%) para o grupo de teste.

A Figura 13 apresenta o fluxograma simplificado para o desenvolvimento do modelo SVR otimizado pelo algoritmo ED. Nesta tese foi utilizado o algoritmo de ED auto-adaptativo *jDE* do pacote *DEoptimR* do programa R[®]. O tamanho da população (*Np*) foi definido como $Np = 10 \times n_p$, onde n_p é o número de parâmetros do modelo SVR. Os critérios de parada utilizados foram o número máximo de iterações igual a $200 \times n_p$ e a tolerância igual a 1×10^{-7} .

Para a seleção dos parâmetros livres do modelo SVR com o kernel função de base radial (*Radial Basis Function*, RBF) (parâmetros *C*, ε e γ), primeiro foi utilizado o procedimento proposto por Cherkassky e Ma (2004) e identificado o melhor espaço de busca. O espaço de busca para os parâmetros de controle do modelo SVR foi: $C \in [1;50]$, $\varepsilon \in [0,001;1]$, $\gamma \in [0,001;1]$, $u \in [0;10]$ e $d \in [1;5]$.

Os parâmetros do modelo SVR também foram obtidos utilizando a busca em grade em conjunto com a validação cruzada (10-*fold*) nos dados de treinamento.

Os melhores valores obtidos para os parâmetros do modelo SVR são apresentados na Tabela 5, onde ED representa a otimização pelo algoritmo ED auto-adaptativo e BG representa o método de busca em grade em conjunto com a validação cruzada (10-*fold*).

	Parâmetr						
Modelo	γ	d	u	С	3	SVs(%)	RMSE
ED Linear	-	-	-	1,6382	0,2537	135 (65,8)	0,0151627
ED Polinomial	0,02227	1	3,2336	44,6386	0,2529	136 (66,3)	0,0151633
ED Laplaciano	0,01854	-	-	6,7758	0,2656	138 (67,3)	0,0142609
ED RBF	0,04175	-	-	3,3651	0,3449	135 (65,8)	0,0140385
ED Sigmoidal	0,00551	-	0,0014	15,8328	0,3422	114 (55,6)	0,0151339
BG Laplaciano	0,00503	-	-	7,7224	0,2445	144 (70,2)	0,0148802
BG RBF	0,01050	-	-	4,0025	0,0300	138 (67,3)	0,0147601

Tabela 5 - Melhores parâmetros dos modelos SVR

Fonte: a autora, 2018.

Da Tabela 5 verifica-se que o modelo SVR que obteve o menor valor de RMSE foi o obtido com o kernel RBF e otimizado por ED, cujos parâmetros são: C = 3,3651, $\varepsilon = 0,3449$ e $\gamma = 0,04175$, número de SVs = 135 e RMSE dos dados de treinamento igual a 0,0140385. O número de vetores de suporte (SVs) deste modelo ficou em 65,8% da quantidade de dados do grupo de treinamento, o que pode ser considerado uma indicação de bom ajuste do modelo, pois considera-se que quanto menor o número de vetores de suporte utilizados, menor é a possibilidade de existir sobreajuste do modelo.

Os parâmetros do modelo SVR obtidos com o kernel RBF e otimizado por ED foram utilizados para o treinamento e, após a fase de treinamento, o modelo obtido foi utilizado para predizer a característica da qualidade utilizando os dados do grupo de treinamento e de teste.

5.3 MODELOS BASEADOS EM REGRESSÃO POR VETORES DE RELEVÂNCIA

Para o desenvolvimento dos modelos baseados em RVR foi utilizado o programa R[®] e foram utilizadas as funções kernel do Quadro 2. O algoritmo ED é utilizado para a minimização da função objetivo, que é definida como o RMSE dos dados de treinamento, dado pela equação (32).

As variáveis de entrada do modelo são: quantidade inicial de fósforo (P^*), teor inicial de Carbono (C^*), teor de Óxido de Manganês (MnO) e basicidade líquida (*Liquid*). A variável de saída do modelo é a proporção da concentração de fósforo (P) nas ligas metálicas. O conjunto total de dados do processo foi dividido aleatoriamente em dois subconjuntos: 205 (80%) dados para o grupo de treinamento e 52 (20%) para o grupo de teste.

A Figura 13 apresenta o fluxograma simplificado para o desenvolvimento do modelo RVR otimizado pelo algoritmo ED. Nesta tese foi utilizado o algoritmo de ED auto-adaptativo *jDE* do pacote *DEoptimR* do programa R[®]. O tamanho da população (*Np*) foi definido como $Np = 10 \times n_p$, onde n_p é o número de parâmetros do modelo RVR. Os critérios de parada utilizados foram o número máximo de iterações igual a $200 \times n_p$ e a tolerância igual a 1×10^{-7} .

O espaço de busca para os parâmetros do modelo RVR foi: $\gamma \in [0,001;1]$, $u \in [0;10]$ e $d \in [1;5]$. Os parâmetros do modelo RVR também foram obtidos utilizando a busca em grade em conjunto com a validação cruzada (10-fold) nos dados de treinamento.

Os melhores valores obtidos para os parâmetros do modelo RVR são apresentados na Tabela 6, onde ED representa a otimização pelo algoritmo ED auto-adaptativo e BG representa o método de busca em grade em conjunto com a validação cruzada (10-fold).

Parâmetros do kernel									
Modelo	γ	d	u	RVs(%)	RMSE				
ED Linear	-	-	-	7 (3,4)	0,0292070				
ED Polinomial	0,019960	1	5,6122	3 (1,5)	0,0144930				
ED Laplaciano	0,002224	-	-	10 (4,9)	0,0137368				
ED RBF	0,003713	-	-	7 (3,4)	0,0142890				
ED Sigmoidal	0,045197	-	0,4746	4 (1,9)	0,0150930				
BG Laplaciano	0,002824	-	-	9 (4,4)	0,0143510				

-

Tabela 6 - Melhores parâmetros dos modelos RVR

0,002510

Fonte: a autora, 2018.

BG RBF

Da Tabela 6 verifica-se que o modelo RVR que obteve o menor valor de RMSE foi o obtido com o kernel Laplaciano e otimizado por ED, cujos parâmetros são: γ = 0,002224, número de RVs = 10 e RMSE dos dados de treinamento igual a 0,0137368. O número de vetores de relevância (RVs) deste modelo ficou em 4,9% da quantidade de dados do grupo de treinamento, o que pode ser considerado uma indicação de bom ajuste do modelo.

-

Analisando os resultados da Tabela 5 e da Tabela 6 verifica-se que o DE-RVR Laplaciano apresenta desempenho um pouco melhor do que o DE-SVR RBF, mas o DE-RVR Laplaciano apresenta um número menor de RVs (10) comparado ao número de SVs do DE-SVR RBF (135). Desta forma, o número de SVs do modelo SVR obtido com o kernel RBF e otimizado por ED é aproximadamente treze vezes maior do que o número de RVs do modelo RVR obtido com o kernel Laplaciano e otimizado por ED. Isto significa que o modelo RVR é mais esparso do que o modelo SVR.

Da análise da Tabela 5 e da Tabela 6 observa-se que: (i) DE-RVR Laplaciano e DE-RVR RBF obtiveram valores menores de RMSE do que o BG-RVR Laplaciano e BG-RVR RBF; (ii) DE-SVR Laplaciano e DE-SVR RBF possuem valores menores

0,0144207

19 (9,2)
de RMSE do que BG-SVR Laplaciano and BG-SVR RBF; (iii) BG-RVR Laplaciano e BG-RVR RBF apresentam desempenho um pouco melhor do que BG-SVR Laplaciano e BG-SVR RBF; (iv) RVR Linear apresentou o maior valor de RMSE entre todos os modelos RVR e SVR.

Os parâmetros do modelo RVR obtido com o kernel Laplaciano e otimizado por ED foram utilizados para o treinamento e, após a fase de treinamento, o modelo obtido foi utilizado para predizer a característica da qualidade utilizando os dados do grupo de treinamento e de teste.

5.4 ANÁLISE DOS MODELOS DE REGRESSÃO

Para realizar a comparação com os modelos desenvolvidos com RVR, SVR e RNAs foi utilizado o modelo de regressão Beta para modelar a proporção de concentração de fósforo utilizando os dados de treinamento (205 observações).

A equação (48) apresenta o modelo de regressão Beta estimado para a proporção de concentração de fósforo no processo siderúrgico contendo as variáveis de controle que apresentaram significância estatística para explicar a característica da qualidade, ao nível de significância de 5%, baseado no teste estatístico de *Wald*. O modelo estimado apresenta distribuição Beta, função de ligação *logit* e método de estimação por máxima verossimilhança. Os dados de teste (52 observações) são usados com o modelo de regressão Beta para a predição de valores futuros da proporção de concentração de fósforo.

$$P = \mathsf{BETA}\left[\exp\left(-1.2985 + 0.4607 P^* - 0.0547 C^* + 0.1704 MnO - 0.0489 Liquid\right)\right] (48)$$

Para a análise da adequação dos modelos de regressão desenvolvidos utilizando RVR, SVR e RNAs e regressão Beta são utilizados métodos gráficos, testes estatísticos, cálculos de erros e cálculo do coeficiente de determinação.

A verificação da normalidade dos resíduos dos quatro modelos foi realizada utilizando o teste de normalidade de Shapiro-Wilk com 95% nível de confiança. Os valores de *p*-value obtidos nos modelos para os dados do grupo de treinamento foram: modelo RVR: *p*-value = 0,39; modelo SVR: *p*-value = 0,25; modelo RNA:

p-value = 0,53; modelo Beta: p-value = 0,65. Baseados nestes resultados, pode-se considerar que os resíduos seguem a distribuição normal, pois p-value > 0,05. Foram obtidos, também, os valores de p-value para os dados do grupo de teste: modelo RVR: p-value = 0,27; modelo SVR: p-value = 0,12; modelo RNA: p-value = 0,20; modelo Beta: p-value = 0,17. Baseados nestes resultados, pode-se considerar que os resíduos seguem a distribuição normal, pois p-value > 0,05.

A Figura 18 apresenta os gráficos de probabilidade normal dos resíduos dos modelos para os dados do grupo de treinamento e a Figura 19, para os dados do grupo de teste, onde os resíduos são obtidos por meio da equação (33).



Figura 18 - Gráficos de probabilidade normal para os resíduos dos modelos para o grupo de treinamento: (a) modelo RVR, (b) modelo SVR, (c) modelo RNA, (d) modelo Beta

Fonte: a autora, 2018.



Figura 19 - Gráficos de probabilidade normal para os resíduos dos modelos para o grupo de teste: (a) modelo RVR, (b) modelo SVR, (c) modelo RNA, (d) modelo Beta

A Figura 20 apresenta o gráfico dos resíduos dos modelos versus a ordem de coleta dos dados para os grupos de treinamento e de teste, não sendo detectada a presença de configurações especiais que possam indicar que os resíduos são autocorrelacionados. Na Figura 20, as 52 observações do grupo de teste são apresentados após as 205 observações do grupo de treinamento para facilitar a visualização dos dados dos dois grupos nos gráficos.



Figura 20 - Gráficos dos resíduos dos modelos versus a ordem de coleta dos dados: (a) modelo RVR, (b) modelo SVR, (c) modelo RNA, (d) modelo Beta

Fonte: a autora, 2018.

Outro teste realizado para a validação do modelo de regressão estimado é a estatística de Durbin-Watson (d_w), calculada pela equação (35), para verificar a hipótese de que os resíduos são independentes.

Para o grupo de treinamento, os valores do limite superior (d_U) e limite inferior (d_L) da Tabela de Durbin-Watson, para 205 observações e nível de significância de 5% são: $d_L = 1,69$ e $d_U = 1,75$. Os valores obtidos com os dados do grupo de treinamento foram: modelo RVR: $d_w = 1,72$; modelo SVR: $d_w = 1,71$; modelo RNA:

 $d_w = 1,71$; modelo Beta: $d_w = 1,70$. Desta forma, para todos os modelos, os valores calculados de d_w estão entre $d_L \le d_w \le d_U$ e os testes são inconclusivos.

Para o grupo de teste, os valores do limite superior (d_U) e limite inferior (d_L) da Tabela de Durbin-Watson, para 52 observações e nível de significância de 5% são: $d_L = 1,35$ e $d_U = 1,59$. Os valores obtidos com os dados do grupo de teste foram: modelo RVR: $d_w = 1,62$; modelo SVR: $d_w = 1,72$; modelo RNA: $d_w = 1,62$; modelo Beta: $d_w = 1,70$. Desta forma, para todos os modelos, os valores calculados de d_w estão entre $d_U \le d_w \le 4 - d_U$ indicando que os resíduos são independentes.

A Figura 21 apresenta os gráficos da função de autocorrelação para os resíduos obtidos com os dados do grupo de treinamento e a Figura 22, para os dados do grupo de teste.

Para a verificação se os resíduos possuem variância aproximadamente constante (homoscedasticidade) foi realizado o teste de Levene e construído o gráfico dos resíduos versus os valores estimados pelos modelos. Os valores de *p-value* obtidos nos modelos com os dados do grupo de treinamento foram: modelo RVR: *p-value* = 0,11; modelo SVR: *p-value* = 0,58; modelo RNA: *p-value* = 0,12; modelo Beta: *p-value* = 0,15. Os valores de *p-value* obtidos nos modelos com os dados do grupo de teste foram: modelo RVR: *p-value* = 0,081; modelo SVR: *p-value* = 0,28; modelo RNA: *p-value* = 0,079; modelo Beta: *p-value* = 0,091. Como *p-value* > 0,05 em todos os modelos, com os dados do grupo de treinamento e de teste, não se rejeita a hipótese de homoscedasticidade dos resísuos.

Na Figura 23 são apresentados os gráficos dos resíduos versus os valores estimados pelos modelos. Verifica-se que os resíduos obtidos com os dados dos grupos de treinamento e de teste estão dispersos aleatoriamente em torno do eixo horizontal e não estão seguindo padrões específicos. Desta forma, pode-se considerar que os resíduos possuem variância aproximadamente constante.

Na sequência, foi verificado o grau de concordância entre os resíduos e os valores estimados utilizando a média (\overline{e}_t) e o desvio padrão (σ_{e_t}) dos resíduos, sendo esperado que a maioria dos resíduos estejam compreendidos em $\overline{e}_t \pm 2\sigma_{e_t}$. Na Figura 23 verifica-se que todos os modelos apresentam mais de 94% dos resíduos dentro deste intervalo, representando um bom ajuste entre os dados do processo e os dados estimados pelos modelos. A Figura 23 foi utilizada, também, para a verificação dos

outliers, sendo que um *outlier* é um resíduo maior do que $\bar{e}_t \pm 3\sigma_{e_t}$. Pode-se verificar que nenhum dos modelos possui resíduos fora destes limites.



Figura 21 - Gráficos da função de autocorrelação para os resíduos obtidos com o grupo de treinamento: (a) modelo RVR, (b) modelo SVR, (c) modelo RNA, (d) modelo Beta

Para a verificação do grau de concordância entre os resíduos e os valores estimados e da presença de *outliers* pode ser utilizado o gráfico dos resíduos padronizados versus os valores estimados e, se os resíduos padronizados são i.i.d., em torno de 95% devem estar no intervalo [-2,2] e 99% no intervalo [-3;+3] (WERKEMA; AGUIAR, 1996). Os resíduos padronizados são obtidos por meio da equação (34). A Figura 24 apresenta os gráficos dos resíduos padronizados versus

Fonte: a autora, 2018.

os valores estimados pelos modelos para os dados dos grupos de treinamento e de teste. Conforme apresentado na Figura 24, em todos os modelos mais de 94% dos resíduos estão no intervalo [-2,2] e 100% no intervalo [-3;+3], comprovando o bom ajuste dos modelos aos dados do processo.





Fonte: a autora, 2018.



Figura 23 - Gráficos dos resíduos versus valores estimados pelos modelos: (a) modelo RVR, (b) modelo SVR, (c) modelo RNA, (d) modelo Beta

Após estes testes, verifica-se que os resíduos obtidos com os modelos RVR, SVR, RNA e Beta podem ser considerados normais, independentes e identicamente distribuídos com variância aproximadamente constante, sendo evidências de que os modelos são adequados para os dados do processo. Desta forma, os modelos podem ser utilizados para predizer a proporção da concentração de fósforo no processo.

A Tabela 7 apresenta informações estatísticas da proporção da concentração de fósforo estimada pelos quatro modelos, utilizando os dados dos grupos de treinamento e de teste. Verifica-se que os resultados obtidos pelos modelos são similares às informações obtidas com os dados experimentais.



Figura 24 - Gráficos dos resíduos padronizados versus valores estimados pelos modelos: (a) modelo RVR, (b) modelo SVR, (c) modelo RNA, (d) modelo Beta

Tabela 7 - Informações estatísticas da proporção da concentração de fósforo estimada pelos modelos

	Proporção da concentração de fósforo					
	Experimental	RVR	SVR	RNA	Beta	
Máximo	0,3240	0,3044	0,3045	0,3057	0,3136	
Mínimo	0,2040	0,2237	0,2186	0,2219	0,2198	
Média	0,2599	0,2597	0,2596	0,2596	0,2597	
Mediana	0,2590	0,2554	0,2582	0,2591	0,2581	
Desvio padrão	0,02546	0,02045	0,02031	0,01858	0,02024	
Variância	0,0006483	0,0004185	0,0004125	0,0003453	0,0004098	

Fonte: a autora, 2018.

A Tabela 8 apresenta os valores dos erros, calculados pelas equações (36) a (39), e do coeficiente de determinação, calculado pela equação (40), para os dados do grupo de treinamento e a Tabela 9, para os dados do grupo de teste. Verifica-se que todos os modelos representam adequadamente a proporção da concentração de fósforo no final do processo de refino. Os valores dos erros obtidos pelos modelos para os grupos de treinamento e de teste não diferem significativamente, indicando não haver sobreajuste dos modelos. O modelo RVR apresenta um desempenho um pouco melhor do que o modelo SVR. Porém, o modelo RVR apresenta um número menor de RVs (10) comparado ao número de SVs (135) do modelo SVR e, desta forma, o modelo RVR é mais esparso do que o modelo SVR.

Tabela 8 - Resultados dos erros e do coeficiente de determinação para o grupo de treinamento

Grupo de treinamento								
Modelo	MSE	RMSE	MAE	MAPE(%)	SVs/RVs(%)	R ²		
RVR	0,0001887	0,0137368	0,01096	4,219	10 (4,9)	0,792		
SVR	0,0001970	0,0140385	0,01108	4,266	135 (65,8)	0,782		
RNA	0,0002310	0,0151986	0,01190	4,605	-	0,728		
Beta	0,0002304	0,0151804	0,01185	4,554	-	0,734		

Fonte: a autora, 2018.

labela 9 - Resultados dos erros e do	coeficiente de	determinação pa	ra o grupo de teste
--------------------------------------	----------------	-----------------	---------------------

Grupo de teste							
Modelo	MSE	RMSE	MAE	MAPE(%)	R ²		
RVR	0,0002225	0,0149175	0,01203	4,602	0,800		
SVR	0,0002243	0,0149787	0,01204	4,603	0,798		
RNA	0,0002289	0,0151294	0,01231	4,722	0,786		
Beta	0,0002468	0,0157098	0,01278	4,824	0,784		

Fonte: a autora, 2018.

Analisando os resultados da Tabela 9 verifica-se que: (i) o modelo RVR apresenta os menores valores de erros entre os modelos; (ii) os modelos RVR e SVR apresentam desempenho um pouco melhor do que os modelos RNA e Beta; (iii) as três técnicas de aprendizado de máquina apresentam desempenho melhor do que o método estatístico. Desta forma, os modelos RVR, SVR e RNA apresentam resultados comparáveis com o modelo Beta sem a necessidade de assumir uma distribuição de probabilidade específica para a implementação dos modelos.

A Figura 25 apresenta os valores observados do processo e os valores estimados pelos quatro modelos, para os grupos de treinamento e de teste. A Figura 26 apresenta os valores observados versus os valores estimados obtidos com os modelos, para os grupos de treinamento e de teste. As Figuras 25 e 26 confirmam o bom desempenho preditivo dos modelos, pois os valores estimados se aproximam adequadamente dos valores experimentais do processo.



Figura 25 - Gráficos dos valores observados e valores estimados pelos modelos: (a) modelo RVR, (b) modelo SVR, (c) modelo RNA, (d) modelo Beta

O *box-percentile plot* foi utilizado para a comparação dos resíduos dos modelos obtidos com os grupos de treinamento e de teste. Conforme apresentado na seção 4.1.3 e na Figura 14, no *box-percentile plot* a largura em qualquer altura é proporcional à porcentagem das observações e a mediana, o 25º percentil (1º quartil) e o 75º percentil (3º quartil) são marcados com segmentos de linha.



Figura 26 - Gráficos dos valores estimados versus valores observados no processo: (a) modelo RVR, (b) modelo SVR, (c) modelo RNA, (d) modelo Beta

A Figura 27 apresenta os *box-percentile plots* dos resíduos obtidos com os modelos para os grupos de treinamento e de teste, sendo que a mediana, o 25^o percentil (1º quartil) e o 75º percentil (3º quartil) foram destacados com segmentos de linha nas cores vermelha e azul, respectivamente.

Na Figura 27 verifica-se que a mediana dos resíduos dos modelos RVR, SVR, RNA e Beta do grupo de treinamento é aproximadamente igual, assim como a mediana dos resíduos dos modelos do grupo de teste. No grupo de treinamento, a diferença entre o valor máximo e o valor mínimo dos resíduos é maior no modelo Beta e menor no modelo RVR. No grupo de teste, a diferença entre o valor máximo e o valor mínimo dos resíduos é maior no modelo Beta e menor no modelo RNA.

Com base nos erros calculados e na análise dos resíduos, o modelo RVR apresenta desempenho ligeiramente superior aos modelos SVR, RNA e Beta.





Para evidenciar o desempenho dos modelos desenvolvidos com RVR, SVR e RNA foi realizada a comparação com os modelos de regressão linear múltipla (*Multiple Linear Regression*, MLR) desenvolvidos por Pedrini e Caten (2010) para prever a concentração de fósforo no mesmo processo produtivo. Com a aplicação da MLR foram obtidos sete modelos diferentes, sendo utilizadas transformações lineares da característica da qualidade, como o logaritmo natural da concentração de fósforo e a diferença entre a concentração de fósforo inicial e final da liga. A Tabela 10 apresenta os sete modelos desenvolvidos por Pedrini e Caten (2010).

Para realizar a comparação entre os modelos RVR, SVR, RNA e Beta e os modelos de MLR, os dados do grupo de teste foram usados com os modelos de regressão apresentados na Tabela 10 para prever valores futuros da proporção da concentração de fósforo.

A Tabela 11 apresenta os resultados dos erros calculados por meio das equações (36) à (39). Analisando os resultados obtidos com os dados do grupo de teste apresentados na Tabela 11, verifica-se que os modelos RVR, SVR, RNA e Beta apresentam erros menores do que os modelos de MLR da Tabela 10.

Modelo	Equação
1	$P = 0,112 + 0,748 P^{*} - 0,00211 Fe^{*} + 0,000015 (MnO)^{2} - 0,0121 BaO$
2	$P = 0,103 + 0,752P^* - 0,00212Fe^* + 0,000017(MnO)^2$
4	$P = 0,155 + 0,750 P^{*} - 0,0309 \ln(Fe^{*}) + 0,000017 (MnO)^{2}$
5	$ln(P) = 0,0983 + 0,751P^* + 0,00000026(MnO)^3 - 0,000003(Fe^*)^3 - 0,0128BaO$
6	$ln(P) = 0,501 + 0,163 \ln(P^*) + 0,0000003(MnO)^3 - 0,000003(Fe^*)^3 - 0,00381 \ln(BaO)$
7	$ln(P-P^{*}) = 0,0453 + 0,0000003(MnO)^{3} - 0,000004(Fe^{*})^{3} - 0,017(BaO)^{3}$
8	$ln(P-P^{*}) = -0,804 ln(Fe^{*}) + 0,371 ln(MnO) - 0,656 ln(CaO)$

Tabela 10 - Modelos de regressão linear múltipla

Fonte: adaptado de Pedrini e Caten, 2010.

Tabela 11 - Resultados dos erros calculados para o grupo de teste

			• •	
Modelo	MSE	RMSE	MAE	MAPE(%)
Modelo 1	0,000251168	0,0158482	0,0129624	4,963
Modelo 2	0,000259737	0,0161164	0,0130344	5,014
Modelo 4	0,000258168	0,0160676	0,0130044	5,004
Modelo 5	0,000264255	0,0162559	0,0129054	4,916
Modelo 6	0,000270178	0,0164371	0,0130135	4,985
Modelo 7	0,000300424	0,0173327	0,0131207	4,989
Modelo 8	0,000247714	0,0157389	0,0127920	4,851
RVR	0,000222532	0,0149175	0,0120329	4,602
SVR	0,000224363	0,0149787	0,0120491	4,603
RNA	0,000228898	0,0151294	0,0123100	4,722
Beta	0,000246797	0,0157098	0,0127802	4,824

Fonte: a autora, 2018.

De acordo com Pedrini e Caten (2010), o Modelo 8 foi o melhor modelo para prever a concentração de fósforo, o que pode ser comprovado pela análise dos resultados da Tabela 11, pois o Modelo 8 apresenta os menores erros quando comparado aos Modelos 1 à 7.

A Figura 28 apresenta os *box-percentile plots* dos modelos RVR, SVR, RNA e Beta e dos Modelos 1 à 8 para a comparação dos resíduos obtidos com os dados do grupo de teste. Conforme apresentado na seção 4.1.3 e na Figura 14, no *box-percentile plot* a largura em qualquer altura é proporcional à porcentagem das observações e a mediana, o 25º percentil (1º quartil) e o 75º percentil (3º quartil) são marcados com segmentos de linha. Na Figura 28, observa-se que a diferença entre o valor máximo e o valor mínimo dos resíduos é maior nos Modelos 1 à 8 e menor nos modelos RVR, SVR, RNA e Beta.





A Figura 29 apresenta os valores estimados versus os valores observados no processo, obtidos com os dados do grupo de teste, para os modelos RVR, SVR, ANN e Beta e o Modelo 8. A Figura 29 confirma que os modelos RVR, SVR, RNA e Beta apresentam melhor desempenho do que o Modelo 8.

Um modelo aceitável deve ter a capacidade de representar adequadamente a variação na característica da qualidade em relação às variáveis de controle, além de ser capaz de prever a característica da qualidade (proporção da concentração de fósforo).

Analisando as equações dos modelos de MLR da Tabela 10 e o modelo de regressão Beta da equação (48), verifica-se que a proporção da concentração de fósforo aumenta com o aumento na variável de controle quantidade inicial de fósforo (P^*) na composição da liga.

Foi analisado o efeito das variações na quantidade inicial de fósforo (P^*) na composição da liga na proporção da concentração de fósforo obtida com os modelos RVR, SVR e RNA. No entanto, análises semelhantes podem ser realizadas para outras variáveis de controle do processo. As variáveis de controle teor inicial de Carbono (C^*), teor de Óxido de Manganês (*MnO*) e basicidade líquida (*Liquid*) são

mantidas constantes nos valores de 6,7 wt-%, 39 wt-% e 9,8 Nm³, respectivamente. A quantidade inicial de fósforo (P^*) foi variada de 0,17 a 0,25%.





A variação da proporção da concentração de fósforo obtida nos modelos RVR, SVR, RNA e Beta para a variação da quantidade inicial de fósforo (P^*) é apresentada na Figura 30. Uma linha de ajuste é usada para descrever a tendência geral, sendo que verifica-se que a proporção da concentração de fósforo aumenta com o aumento da quantidade inicial de fósforo (P^*) para os modelos RVR, SVR e RNA, apresentando tendência similar à dos modelos de MLR (Tabela 10) e do modelo de regressão Beta (equação (48) e Figura 30).

Baseado em todos os testes e análises realizadas, os modelos RVR, SVR, RNA e Beta podem ser considerados válidos para prever a proporção da concentração de fósforo no processo siderúrgico analisado.



Figura 30 - Gráficos da variação na proporção da concentração de fósforo em função da variação da quantidade inicial de fósforo (*P**): (a) modelo RVR, (b) modelo SVR, (c) modelo RNA, (d) modelo Beta

5.5 IMPLEMENTAÇÃO DAS CARTAS DE CONTROLE NA FASE I

Como os modelos desenvolvidos foram validados, foi possível utilizá-los para estimar a característica da qualidade, dados os valores das variáveis de controle do processo. Para a implementação das cartas de controle na Fase I são calculados os resíduos padronizados obtidos com os modelos RVR, SVR, RNA e Beta, por meio da equação (34).

Para a comparação entre as cartas de controle desenvolvidas considera-se o mesmo valor de NMA em controle (NMA₀) para todas as cartas de controle. Os limites de controle das cartas de controle são calculados pelas equações (42), (43) e (44) considerando a probabilidade α = 0,0027 de alarmes falsos no monitoramento do processo, correspondente a um NMA₀ de 370,4 para as quatro cartas de controle. A adoção de valores da constante λ referentes a mesma probabilidade de alarmes falsos das cartas de controle permite o estudo de comparação gráfica entre as cartas de controle analisadas.

A Figura 31 apresenta o monitoramento dos resíduos padronizados utilizando os limites de controle calculados a partir dos resíduos dos modelos de regressão desenvolvidos. Verifica-se que as cartas de controle apresentam similaridade no monitoramento da variabilidade da característica da qualidade do processo.



Figura 31 - Cartas de controle para os resíduos padronizados da Fase I: (a) carta RVR, (b) carta

Fonte: a autora, 2018.

Como nas quatro cartas de controle todos os resíduos padronizados encontram-se dentro dos limites de controle e não apresentaram resultados positivos para os testes de estabilidade realizados, nenhuma evidência de presença de causas especiais foi detectada e o processo pode ser considerado sob controle estatístico. Desta forma, todos os resíduos padronizados são bem classificados pelas cartas de controle, uma vez que as observações foram de fato obtidas a partir do processo sob controle.

Desta forma, os modelos de regressão RVR, SVR, RNA e Beta podem ser utilizados para estimar a característica da qualidade e monitorar a proporção da concentração de fósforo do processo. Os limites de controle calculados a partir dos resíduos padronizados dos modelos podem ser utilizados para o monitoramento do processo na Fase II.

5.6 IMPLEMENTAÇÃO DAS CARTAS DE CONTROLE NA FASE II

Após a Fase I, iniciou-se a Fase II do método proposto, que consiste na coleta de novos dados provenientes do processo e no monitoramento do processo. Os dados coletados devem conter o valor da característica da qualidade e os valores das variáveis de controle, no momento da coleta dos dados.

Para a aplicação da Fase II do método proposto, foram utilizadas 60 novas observações provenientes do processo, sendo 40 observações obtidas com o processo sob controle e as últimas 20 observações resultaram na proporção da concentração de fósforo acima dos resultados esperados. A análise dessas observações fora de controle indica que estas foram causadas por valores incomuns das variáveis de controle que alteraram a média da característica da qualidade em dois desvios padrão.

Na sequência, foram obtidos os resíduos padronizados das novas observações utilizando os modelos desenvolvidos na Fase I e esses resíduos padronizados foram monitorados nas cartas de controle com os limites de controle calculados na Fase I.

A Figura 32 apresenta o monitoramento dos resíduos padronizados nas cartas de controle. Novamente, verifica-se que todos os resíduos foram bem classificados

desde os 40 primeiros, que estão dentro dos limites de controle, até os 20 últimos, que estão deslocados em dois desvios padrão da média da característica da qualidade.





5.7 ANÁLISE DE SENSIBILIDADE DAS CARTAS DE CONTROLE

Um estudo de simulação foi realizado para analisar a sensibilidade das cartas de controle dos resíduos baseadas em modelos de regressão RVR, SVR, RNA e Beta desenvolvidas na Fase I, considerando um processo sob controle e outro fora de controle. O método utilizado para a análise de sensibilidade foi a simulação de Monte Carlo usando o programa R[®].

Na análise de sensibilidade foram utilizados o número médio de amostras (NMA), o número mediano de amostras (*Median Run Length*, MRL), o desvio padrão do número de amostras (*Standard Error of the Run Length*, SRL) e a função distribuição cumulativa (*Cumulative Distribution Function*, CDF).

Para o processo sob controle foi calculado o número médio de amostras até a presença de um alarme falso (NMA₀), calculado pela equação (45), e para o processo fora de controle, o número médio de amostras até a presença de causas especiais (NMA₁), calculado pela equação (46). É desejável que uma carta de controle apresente um NMA₀ grande a fim de evitar alarmes falsos e um NMA₁ pequeno a fim de detectar rapidamente alterações no processo. O MRL e o SRL foram calculados pela equação (47).

Para o processo sob controle, a simulação consistiu na geração de 10.000 observações, assumindo-se que as quatro variáveis de controle simuladas e a característica da qualidade simulada possuem as distribuições de probabilidade empíricas das observações históricas do processo em estudo. Esses dados foram utilizados para estimar os valores preditos a partir dos modelos de regressão desenvolvidos na Fase I e calcular os respectivos resíduos padronizados. A simulação do processo sob controle foi realizada 5.000 vezes, sendo computada a média e o erro padrão dos percentuais de pontos fora dos limites de controle estatístico.

Para o processo fora de controle foram induzidas alterações (δ) aditivas na média (μ_y) da característica da qualidade sob controle. Isto significa que a média foi alterada de (μ_y) para ($\mu_y + \delta \sigma_y$), onde σ_y é o desvio padrão da característica da qualidade e δ = 0,5; 1; 1,5; 2; 2,5; 3; 3,5 e 4. Para cada alteração induzida foram geradas 10.000 observações e cada simulação foi realizada 5.000 vezes.

Para realizar a análise de sensibilidade das cartas de controle, inicialmente foi utilizada a simulação de Monte Carlo para obter o valor da constante λ que equaliza os valores de NMA₀ entre as cartas de controle desenvolvidas. Os valores da constante λ foram ajustados para NMA₀ igual a 370,4 nas quatro cartas de controle. A adoção de valores da constante λ referentes a mesma probabilidade de alarmes falsos das cartas de controle permite o estudo de comparação numérica e gráfica entre as cartas de controle desenvolvidas.

A Tabela 12 apresenta os valores do NMA (desvio padrão entre parênteses), SRL e dos percentis da distribuição do número de amostras para as cartas de controle dos resíduos padronizados para os processos sob controle e fora de controle simulados, onde δ representa a alteração na média da característica da qualidade sob controle e δ_r representa a modificação resultante nos resíduos padronizados obtidos com os modelos. As alterações aditivas induzidas na média da característica da qualidade sob controle foram $\delta = 0.5$; 1; 1.5; 2; 2.5; 3; 3.5 e 4.

A Figura 33 apresenta os gráficos da função distribuição cumulativa (CDF) para o processo sob controle e para os processos com as diferentes alterações aditivas induzidas na média da característica da qualidade sob controle. A CDF mede a proporção cumulativa dada pelo *i*-ésimo período após a alteração e indica a probabilidade de uma variável ter um valor menor ou igual a um valor selecionado. A distribuição completa do número de amostras fornece informações úteis sobre o desempenho das quatro cartas de controle.

A partir da Tabela 12 e da Figura 33, observa-se que todas as cartas de controle dos resíduos padronizados obtiveram bom desempenho e boa sensibilidade para as diferentes alterações induzidas na média da característica da qualidade sob controle.

Introduzindo uma alteração $\delta = 0.5$ nos dados principais para as cartas de controle RVR, SVR, RNA e Beta, obteve-se os valores de NMA₁ de 163,71; 167,37; 163,83 e 175,62, respectivamente, e MRL de 114, 117, 114 e 122, respectivamente. Observa-se que as três técnicas de aprendizado de máquina apresentaram desempenho um pouco melhor que o método estatístico.

As cartas de controle dos resíduos utilizando os modelos SVR e RNA apresentaram um desempenho um pouco melhor para pequenas alterações, pois geraram os menores valores de NMA₁ para os valores menores de alterações induzidas de descontrole, quando comparadas às outras cartas de controle. Para alterações induzidas maiores, as quatro cartas de controle apresentaram capacidade de detecção equivalentes. Pode-se notar que as quatro cartas de controle apresentaram praticamente o mesmo desempenho para alterações na média do processo a partir de 1,5, pois necessitam praticamente da mesma quantidade de amostras para detectar um MRL fora de controle.

					Percentis da distribuição do número de amostras				
	δ	$\boldsymbol{\delta}_{r}$	NMA (desvio padrão)	SRL	5°	25°	50° (MRL)	75°	95°
	0,0	0,0	370,19 (0,073)	369,39	19	107	257	513	1108
	0,5	0,4	168,37 (0,034)	167,37	9	49	117	234	503
	1,0	0,8	59,68 (0,012)	59,18	3	17	41	82	178
	1,5	1,2	23,56 (0,0045)	23,05	2	7	16	32	69
RVR	2,0	1,6	10,67 (0,0020)	10,16	1	3	7	14	31
	2,5	2,0	5,53 (0,0010)	5,00	1	2	4	7	15
	3,0	2,4	3,27 (0,00053)	2,72	1	1	2	4	8
	3,5	2,8	2,18 (0,00033)	1,60	1	1	1	3	5
	4,0	3,2	1,62 (0,00020)	1,00	1	1	1	2	3
	0,0	0,0	370,78 (0,073)	370,28	19	107	257	514	1110
	0,5	0,4	164,22 (0,033)	163,71	9	47	114	227	491
	1,0	0,8	57,25 (0,011)	56,75	3	17	40	79	170
	1,5	1,2	22,39 (0,0043)	21,88	2	7	16	30	66
SVR	2,0	1,6	10,10 (0,0019)	9,58	1	3	7	14	29
	2,5	2,0	5,23 (0,00094)	4,71	1	2	4	7	14
	3,0	2,5	3,11 (0,00050)	2,55	1	1	2	4	8
	3,5	2,9	2,09 (0,00031)	1,51	1	1	1	3	5
	4,0	3,3	1,56 (0,00019)	0,93	1	1	1	2	3
	0,0	0,0	370,51 (0,074)	370,01	19	107	257	513	1109
	0,5	0,4	164,36 (0,033)	163,83	9	48	114	228	491
	1,0	0,8	58,52 (0,012)	58,02	3	17	41	81	174
	1,5	1,2	23,23 (0,0045)	22,72	2	7	16	31	69
RNA	2,0	1,6	10,56 (0,0020)	10,05	1	3	7	14	30
	2,5	2,0	5,48 (0,0010)	4,95	1	2	4	7	15
	3,0	2,4	3,24 (0,00053)	2,69	1	1	2	4	8
	3,5	2,8	2,16 (0,00031)	1,58	1	1	1	3	5
	4,0	3,3	1,60 (0,00019)	0,98	1	1	1	2	3
	0,0	0,0	369,98 (0,072)	369,48	19	107	257	513	1107
	0,5	0,4	175,62 (0,034)	175,12	9	51	122	243	525
	1,0	0,8	61,79 (0,012)	61,29	3	18	43	85	184
	1,5	1,2	24,08 (0,0047)	23,57	2	7	17	33	71
Beta	2,0	1,6	10,80 (0,0020)	10,29	1	3	7	15	31
	2,5	2,0	5,56 (0,0010)	5,04	1	2	4	7	15
	3,0	2,4	3,27 (0,00053)	2,72	1	1	2	4	9
	3,5	2,8	2,18 (0,00032)	1,60	1	1	1	3	5
	4,0	3,2	1,62 (0,00020)	1,00	1	1	1	2	3

Tabela 12 - NMA (desvio padrão), SRL e percentis da distribuição do número de amostras para as cartas de controle dos resíduos padronizados

Fonte: a autora, 2018.

Nota: δ é a alteração na média da característica da qualidade sob controle e δ_r é a modificação resultante nos resíduos padronizados obtidos com os modelos.



Figura 33 – Gráficos da função distribuição cumulativa (CDF) para alterações induzidas na média da característica da qualidade sob controle: (a) carta RVR, (b) carta SVR, (c) carta RNA, (d) carta Beta

Conforme a Tabela 12, para todas as cartas de controle na situação sob controle o NMA₀ é próximo de 370,4; MRL sob controle é 257, 5º percentil é 19 e o 25º percentil é 107. Para a carta SVR o 75º percentil é 514 e o 95º percentil é 1.110. O NMA₀ indica que um alarme falso é observado, em média, a cada 370,4 amostras. O 50º percentil (MRL₀) indica que existe 50% de probabilidade de que o primeiro alarme falso ocorra em menos do que 257 amostras, ou seja, a carta sinaliza pela primeira vez na 257ª amostra ou pelo menos em 50% do tempo (CHAKRABORTI, 2007). O 25º percentil indica que existe a probabilidade de 0,25 de que o número de

amostras das cartas seja menor do que 107 e o 75º percentil indica que 25% do tempo o número de amostras é maior que 514.

Teoh *et al.* (2016) considera que, em muitos processos industriais, os primeiros alarmes falsos são uma consideração significativa a ser analisada. Os menores percentis da distribuição do número de amostras, como os 5º, 10º e 25º percentis quando o processo está sob controle, fornecem informações sobre os primeiros alarmes falsos. Valores pequenos para esses percentis caracterizam alarmes falsos mais frequentes, resultando em repetidas paradas dos processos e inicialização dos processos em intervalos de tempo mais curtos.

Para a carta SVR, o 5º percentil indica que um alarme falso no processo sob controle não ocorre dentro das primeiras 19 amostras, com uma probabilidade de pelo menos 95% (CHAKRABORTI, 2007). A assimetria da distribuição do número de amostras pode ser observada a partir da diferença entre os 5º e 95º percentis. O valor do 5º percentil é um valor baixo (19), e o 95º percentil possui um valor significativamente maior (1.110), e isso indica que existe mais variação na distribuição do número de amostras que possui uma cauda mais longa (TEOH *et al.*, 2016).

A Figura 34 compara os *box-percentile plots* do número de amostras para o processo sob controle e a Figura 35 compara os *box-percentile plots* do número de amostras para o processo fora de controle para alterações na média da característica da qualidade do processo sob controle, para as quatro cartas de controle. Conforme apresentado na seção 4.1.3 e na Figura 14, no *box-percentile plot* a largura em qualquer altura é proporcional à porcentagem das observações e a mediana, o 25º percentil (1º quartil) e o 75º percentil (3º quartil) são marcados com segmentos de linha.

Nos *box-percentile plots* das Figura 34 e 35, a largura contém informações precisas sobre a distribuição do número de amostras e os *box-percentile plots* apresentam todas as informações do gráfico CDF do número de amostras. A Figura 34 e a Figura 35 ilustram o padrão típico de uma distribuição assimétrica, tal como é a distribuição do número de amostras.

A Tabela 12 apresenta o 95º percentil do número de amostras, e isso indica que 5% do tempo o número de amostra é maior do que, aproximadamente, 1.110 para todos os processos sob controle. A Figura 34 e a Figura 35 contêm as informações até o 100º percentil do número de amostras.

Conforme apresentado nas Figuras 34 e 35, observa-se que as quatro cartas de controle apresentam praticamente os mesmos valores de mediana, 25° e 75° percentis do número de amostras (ver Tabela 12). No entanto, a diferença entre os valores máximo e mínimo do número de amostras é maior para as cartas SVR e RNA em comparação com as cartas RVR e Beta para o processo sob controle.



Figura 34 - Box-percentile plots do número de amostras para o processo sob controle

Fonte: a autora, 2018.



Figura 35 - *Box-percentile plots* do número de amostras para diferentes alterações induzidas na média da característica da qualidade sob controle: (a) carta RVR, (b) carta SVR, (c) carta RNA, (d) carta Beta

Fonte: a autora, 2018.

6 CONCLUSÃO

Esta pesquisa teve como objetivo desenvolver cartas de controle baseadas em modelos de regressão utilizando Regressão por Vetores de Relevância (RVR), Regressão por Vetores de Suporte (SVR) e Redes Neurais Artificiais (RNA) do tipo Perceptron Multicamadas para o monitoramento de processos que mensuram variáveis não conformes às especificações do tipo proporção.

Nos últimos anos foram aplicadas abordagens empregando RNAs e SVMs para o reconhecimento de padrões, o monitoramento de processos industriais e o desenvolvimento de cartas de controle porém, não foram encontradas publicações empregando RVR nestas aplicações de CEP. Algumas publicações de RNAs no monitoramento de processos que mensuram variáveis não conformes às especificações estão disponíveis porém, não foram encontradas publicações aplicando SVR no monitoramento de processos com este tipo de variáveis.

A escolha dos parâmetros apropriados é essencial para a obtenção de modelos RVR e SVR bem ajustados. Nesta pesquisa foi utilizado um algoritmo de evolução diferencial auto-adaptativo para a obtenção dos parâmetros dos modelos RVR e SVR com diferentes funções kernel.

Os modelos de regressão e as cartas de controle dos resíduos propostos foram aplicados no estudo experimental do processo produtivo de uma empresa siderúrgica para a modelagem e o monitoramento da proporção da concentração de fósforo nas ligas metálicas produzidas pela empresa.

Os modelos RVR, SVR e RNA foram avaliados por medidas de diagnósticos e comparados com o modelo de regressão Beta e com modelos de regressão linear múltipla existentes na literatura. Os resultados indicaram que os modelos RVR, SVR e RNA apresentaram resultados comparáveis com o modelo de regressão Beta sem a necessidade de assumir uma distribuição de probabilidade específica para a implementação dos modelos. O modelo RVR apresentou desempenho similar ao desempenho do modelo SVR porém, o número de RVs do modelo RVR foi aproximadamente treze vezes menor do que o número de SVs do modelo SVR, comprovando a esparsividade do modelo RVR.

A sensibilidade das cartas de controle foi avaliada por meio de simulação de Monte Carlo, em processos sob controle e fora de controle. As simulações mostraram que as cartas de controle propostas apresentaram sensibilidade adequada na detecção das diferentes alterações induzidas na média da característica da qualidade do processo sob controle, com sensibilidade um pouco melhor para as cartas de controle SVR e RNA para pequenas alterações induzidas.

Em função dos resultados obtidos verificou-se que os modelos de regressão e as cartas de controle dos resíduos baseadas em modelos RVR, SVR e RNA são adequadas para a modelagem e para o monitoramento da característica da qualidade do processo produtivo analisado.

Desta forma, os objetivos específicos propostos na tese foram abordados com sucesso e, como principais contribuições desta tese, destacam-se a aplicação inédita de RVR no desenvolvimento de cartas de controle e a aplicação inédita de RVR e SVR no monitoramento de processos que mensuram variáveis não conformes às especificações utilizando cartas de controle.

Para trabalhos futuros sugere-se a utilização da carta de controle baseada em modelo RVR em outros processos industriais e a aplicação de comitê de máquinas, *support vector data description* (SVDD), *weighted support vector machines* (WSVM), máquinas de aprendizado extremo, redes bayesianas e sistemas *fuzzy* na modelagem e no monitoramento de processos que mensuram variáveis não conformes às especificações utilizando cartas de controle.

REFERÊNCIAS

ADDEH, Jalil; EBRAHIMZADEH, Ata; AZARBAD, Milad; RANAEE, Vahid. Statistical process control using optimized neural networks: a case study. **ISA Transactions**, v. 53, p. 1489-1499, 2014.

ADDEH, Abdoljalil; KHORMALI, Aminollah; GOLILARZ, Noorbakhsh A. Control chart pattern recognition using RBF neural network with new training algorithm and practical features. **ISA Transactions**, v. 79, p. 202-2016, 2017.

AHMADZADEH, Farzaneh. Change point detection with multivariate control charts by artificial neural network. **International Journal of Advanced Manufacturing Technology**, p. 1-12, 2009. doi: 10.1007/s00170-009-2193-6

AHMADZADEH, Farzaneh; LUNDBERG, Jan; STRÖMBERG, Thomas. Multivariate process parameter change identification by neural network. **The International Journal of Advanced Manufacturing Technology**, v. 69, p. 2261-2268, 2013.

ALVES, Júlio C. L. Máquina de vetores de suporte aplicada a dados de espectroscopia NIR de combustíveis e lubrificantes para o desenvolvimento de modelos de regressão e classificação. 2012. 271 f. Tese (Doutorado em Química) – Instituto de Química, Universidade Estadual de Campinas, Campinas, 2012.

AMIRI, Amirhossein; NIAKI, Seyed T. A.; MOGHADAM, Alireza T. A probabilistic artificial neural network-based procedure for variance change point estimation. **Soft Computing**, v. 19, p. 691-700, 2015.

APPOLINÁRIO, Fábio. **Metodologia da ciência: filosofia e prática da pesquisa**. 2. ed. São Paulo: Cengage Learning, 2012.

ATASHGAR, Karim; NOOROSSANA, R. An integrating approach to root cause analysis of a bivariate mean vector with a linear trend disturbance. **International Journal of Advanced Manufacturing Technology**, v. 52, p. 407-420, 2011.

ATASHGAR, K. Monitoring multivariate environments using artificial neural network approach: an overview. **Scientia Iranica**, v. 22, 2527-2547, 2015.

AWADALLA, Medhat H. A.; SADEK, M. Abdellatif. Spiking neural network-based control chart pattern recognition. **Alexandria Engineering Journal**, v. 51, p. 27-35, 2012.

BABU, B. V., ANGIRA, R. Modified differential evolution (MDE) for optimization of nonlinear chemical processes. **Computers and Chemical Engineering**, v. 30, p. 989-1002, 2006.

BARELLA, S., MAPELLI, C., MOMBELLI, D., GRUTTADAURIA, A., LAGHI, E., ANCONA, V., VALENTINO, G.. Model for the final decarburization of the steel bath through a self-bubbling effect. **Ironmaking & Steelmaking**, p. 1-4, 2017. doi: 10.1080/03019233.2017.1405179

BARGHASH, Mahmoud A. An effective and novel neural network ensemble for shift pattern detection in control charts. **Computational Intelligence and Neuroscience**, vol. 2015, p. 1-9, 2015. doi: 10.1155/2015/939248

BASAK, Debasish; PAL, Srimanta; PATRANABIS, Dipak C. Support vector regression. **Neural Information Processing – Letters and Reviews**, v. 11, n. 10, p. 203-224, 2007.

BASU, Jayanta K.; BHATTACHARYYA, Debnath; KIM, Tai-hoon. Use of artificial neural network in pattern recognition. **International Journal of Software Engineering and Its Applications**, v. 4, n. 2, p. 23-34, 2010.

BELTRAMI, Mônica. **Precificação de opções sobre ações por modelos de support vector regression**. 2009. 124 f. Dissertação (Mestrado em Métodos Numéricos em Engenharia) – Pós-Graduação em Métodos Numéricos em Engenharia, Universidade Federal do Paraná, Curitiba, 2009.

BISHOP, Christopher M. **Pattern recognition and machine learning**. New York: Springer, 2006.

BOX, G. E. P.; COX, D. R. An analysis of transformation. **Journal of the Royal Statistical Society**, v. 26, n. 2, p. 211-252, 1964.

BRAGA, Antônio de P.; CARVALHO, André P. de L. F.; LUDERMIR, Teresa B. **Redes neurais artificiais: teoria e aplicações**. 2. ed. Rio de Janeiro: LTC, 2012.

BREST, J.; GREINER, S.; BOSKOVIC, B.; MERNIK, M.; ZUMER, V. Self-adapting control parameters in differential evolution: a comparative study on numerical benchmark problems. **IEEE Transactions on Evolutionary Computation**, v.10, p. 646-657, 2006.

BURGES, Christopher J.C. A tutorial on support vector machines for pattern recognition. **Data Mining and Knowledge Discovery**, v. 2, p. 121-167, 1998.

CAMCI, Fatih; CHINNAM, Ratna B.; ELLIS, Richard D. A robust kernel-distance multivariate control chart using support vector principles. **International Journal of Production Research**, v. 46, p. 5075-5095, 2008.

CHAUDHARY, P. N.; GOEL, R. P.; ROY, G. G.. Dephosphorisation of high carbon ferromanganese using BaCO3 based fluxes. **Ironmaking & Steelmaking**, v.28, n. 5, p. 396-403, 2001.

CHAKRABORTI, S. Run length distribution and percentiles: the Shewhart chart with unknown parameters. **Quality Engineering**, v. 19, n. 2, p. 119-127, 2007.

CHANG, Chih-Chung; LIN, Chih-Jen. LIBSVM: a library for support vector machines. **ACM Transactions on Intelligent Systems and Technology**, v. 2, n. 27, p. 1-27, 2011. Disponível em: https://www.csie.ntu.edu.tw/~cjlin/libsvm/. Acesso em: 10 out. 2017.

CHARNET, Reinaldo; FREIRE, Clarice A. L.; CHARNET, Eugênia M. Reginato; BONVINO, Heloísa. **Análide de modelos de regressão linear com aplicações**. Campinas: Editora da Unicamp, 2008.

CHEN, C.; CHIANG, T. Adaptive differential evolution: A visual comparison. **IEEE Congress on Evolutionary Computation (CEC)**, Sendai, Japan, p. 401-408, 2015.

CHENG, Chuen-Sheng; CHENG, Hui-Ping. Identifying the source of variance shifts in the multivariate process using neural networks and support vector machines. **Expert Systems with Applications**, v. 35, n. 1-2, p. 198-206, 2008.

CHENG, Chuen-Sheng; CHEN, Pei-Wen; HUANG, Kuo-ko. Estimating the shift size in the process mean with support vector regression and neural networks. **Expert Systems with Applications**, v. 38, p. 10624-10630, 2011.

CHENG, Chuen-Sheng; CHENG, Hui-Ping. Using neural networks to detect the bivariate process variance shifts pattern. **Computers & Industrial Engineering**, v. 60, n. 2, p.269-278, 2011.

CHENG, Chuen-Sheng; LEE, Hung-Ting. Identifying the out-of-control variables of multivariate control chart using ensemble SVM classifiers. **Journal of the Chinese Institute of Industrial Engineers**, v. 29, n. 5, p. 314-323, 2012.

CHENG, Chuen-Sheng; HUANG, Kuo-Ko; CHEN, Pei-Wen. Recognition of control chart patterns using a neural network-based pattern recognizer with features extracted from correlation analysis. **Pattern Analysis and Applications**, v. 18, p. 75-86, 2015.

CHENG, Zhi-Qiang; MA, Yi-Zhong; BU, And J. Variance shifts identification model of bivariate process based on LS-SVM pattern recognizer. **Communications in Statistics - Simulation and Computation**, v. 40, p. 274-284, 2011.

CHERKASSKY, Vladimir; MA, Yunqian. Practical selection of SVM parameters and noise estimation for SVM regression. **Neural Networks**, v. 17, p. 113-126, 2004.

CHERKASSKY, V., MULIER, F. Learning from data: concepts, theory, and methods. 2nd ed. New Jersey: John Wiley & Sons, 2007.

CHIÑAS, Pamela; LOPEZ-JUAREZ, I.; VAZQUEZ, José A.; OSORIO, Roman; LEFRANC, Gaston. SVM and ANN application to multivariate pattern recognition using scatter data. **IEEE Latin America Transactions**, v. 13, n. 5, p. 1633-1639, 2015.

CHINNAM, Ratna B.; KUMAR, Vinay S. Using support vector machines for recognizing shifts in correlated manufacturing processes. **International Joint Conference on Neural Networks** (IJCNN), v. 3, p. 2276-2280, 2001.

CHINNAM, Ratna B. Support vector machines for recognizing shifts in correlated and other manufacturing processes. **International Journal of Production Research**, v. 40, n. 17, p. 4449-4466, 2002.

CHONGFUANGPRINYA, Panitarn; KIM, Seoung B.; PARK, Sun-Kyoung; SUKCHOTRAT, Thuntee. Integration of support vector machines and control charts for multivariate process monitoring. **Journal of Statistical Computation and Simulation**, v. 81, p. 1157-1173, 2011.

CONCEIÇÃO, E.; MÄCHLER, M. DEoptimR: differential evolution optimization in pure R. R package version 1.0-8, 2015. Disponível em: https://cran.r-project.org/web/packages/DEoptimR/index.html. Acesso em: 10 fev. 2018.

COOK, Deborah F.; CHIU, Chih-Chou. Using radial basis function neural networks to recognize shifts in correlated manufacturing process parameters. **IIE Transactions**, v. 30, p. 227-234, 1998.

CORDEIRO, Gauss M.; DEMÉTRIO, Clarice G. B. **Modelos lineares** generalizados. São Paulo: Editora da USP, 2008.

COSTA, Antônio F. B.; EPPRECHT, Eugênio K.; CARPINETTI, Luiz C. R. **Controle** estatístico de qualidade. 2. ed. São Paulo: Atlas, 2010.

CRISTIANINI, Nello; SHAWE-TAYLOR, John. **An introduction to support vector machines**: and other kernel-based learning methods. Cambridge: Cambridge University Press, 2000.

CUENTAS, Sandra; PEÑABAENA-NIEBLES, Rita; GARCIA, Ethel. Support vector machine in statistical process monitoring: a methodological and analytical review. **International Journal of Advanced Manufacturing Technology**, v. 91, p. 485-500, 2017. doi: 10.1007/s00170-016-9693-y

CYBENKO, G. Approximations by superpositions of sigmoidal functions. **Mathematics of Control, Signals, and Systems**, v. 2, n. 4, p. 303-314, 1989.

DAS, P.; BANERJEE, I. An hybrid detection system of control chart patterns usingcascaded SVM and neural network-based detector. **Neural Computing & Applications**, v. 20, p. 287-296, 2011.

DAS, Swagatam; SUGANTHAN, Ponnuthurai N. Differential evolution: a survey of the state-of-the-art. **IEEE Transactions on Evolutionary Computation**, v. 15, n. 1, p. 4-31, 2011.

DAS, S.; MULLICK, S. S.; SUGANTHAN, P. N. Recent advances in differential evolution – An updated survey. **Swarm and Evolutionary Computation**, v. 27, p. 1-30, 2016.

DE CASTRO, Leandro N. Fundamentals of natural computing: basic concepts, algorithms, and applications. London: Chapman & Hall/CRC, 2006.

DRUCKER, Harris; BURGES, Chris J. C.; KAUFMAN, Linda; SMOLA, Alex; VAPNIK, Vladimir. Support vector regression machines. **Advances in Neural Information Processing Systems**, v. 9, p. 155-161, 1996.

DU, Ke-Lin; SWAMY, M.N.S. **Neural networks and statistical learning**. Heidelberg: Springer, 2013.

DU, Shichang; LV, Jun; XI, Lifeng. An integrated system for on-line intelligent monitoring and identifying process variability and its application. **International Journal of Computer Integrated Manufacturing**, v. 23, n. 6, p. 529-542, 2010.

DU, Shichang; XI, Lifeng; YU, Jianbo; SUN, Jiwen. Online intelligent monitoring and diagnosis of aircraft horizontal stabilizer assemble processes. **International Journal of Advanced Manufacturing Technology**, v. 50, p. 377-389, 2010.

DU, Shichang; LV, Jun; XI, Lifeng. On-line classifying process mean shifts in multivariate control charts based on multiclass support vector machines. **International Journal of Production Research**, v. 50, p. 6288-6310, 2012.

DU, Shichang; HUANG, Delin; LV, Jun. Recognition of concurrent control chart patterns using wavelet transform decomposition and multiclass support vector machines. **Computers & Industrial Engineering**, v. 66, p. 683-695, 2013.

DU, Shichang; LV, Jun. Minimal euclidean distance chart based on support vector regression for monitoring mean shifts of auto-correlated processes. **International Journal of Production Economics**, v. 141, p. 377-387, 2013.

DYER, J.N. Monte Carlo simulation design for evaluating normal-based control chart properties. **Journal of Modern Applied Statistical Methods**, v. 15, p. 580-626, 2016.

EBRAHIMZADEH, A.; RANAEE, V. Control chart pattern recognition using an optimized neural network and efficient features. **ISA Transactions**, v. 49, p.387-393, 2010.

EBRAHIMZADEH, A.; ADDEH, J.; RAHMANI, Z. Control chart pattern recognition using K-MICA clustering and neural networks. **ISA Transactions**, v.51, p.111-119, 2012.

EBRAHIMZADEH, A., ADDEH, J., RANAEE, V. Recognition of control chart patterns using an intelligent technique. **Applied Soft Computing**, v. 13, n. 5, p. 2970-2980, 2013.

EIBEN, A.; MICHALEWICZ, Z.; SCHOENAUER, M.; SMITH, J. Parameter control in evolutionary algorithms. In: LOBO, F. G.; LIMA, C. F.; MICHALEWICZ, Z. (Eds), **Parameter Setting in Evolutionary Algorithms**, v. 54, n. 54, Springer Verlag, p. 19-46, 2007.

EL-MIDANY, T. T.; EL-BAZ, M. A.; ABD-ELWAHED, M. S. A proposed framework for control chart pattern recognition in multivariate process using artificial neural networks. **Expert Systems with Applications**, v. 37, p.1035-1042, 2010.

ENGELBRECHT, Andries P. **Computational intelligence**: **an introduction**. 2nd ed. West Sussex: John Wiley & Sons, 2007.

ESTY, W. W.; BANFIELD, J. D. The box-percentile plot. Journal of Statistical Software, v. 8, n. 17, p. 1-14, 2003.

FACELI, Katti; LORENA, Ana C.; GAMA, João; CARVALHO, André C. P. L. F. **Inteligência artificial: uma abordagem de aprendizado de máquina**. Rio de Janeiro: LTC, 2011.

FAN, Bi; LI, Han-Xiong; HU, Yong. An intelligent decision system for intraoperative somatosensory evoked potential monitoring. **IEEE Transactions on Neural Systems and Rehabilitation Engineering**, v. 24, n. 2, 2016.

FERRARI, S.; CRIBARI-NETO, F. Beta regression for modelling rates and proportions. **Journal of Applied Statistics**, v. 31, n. 7, p. 799-815, 2004.

GANI, Walid; TALEB, Hassen; LIMAM, Mohamed. Support vector regression based residual control charts. **Journal of Applied Statistics**, v. 37, n. 2, p. 309-324, 2010.

GANI, Walid; TALEB, Hassen; LIMAM, Mohamed. An assessment of the kerneldistance-based multivariate control chart through an industrial application. **Quality and Reliability Engineering International**, v. 27, n. 3, p. 391-401, 2011.

GANI, Walid; LIMAM, Mohamed. Performance evaluation of one-class classificationbased control charts through an industrial application. **Quality and Reliability Engineering International**, v. 29, p. 841-854, 2013a.

GANI, Walid; LIMAM, Mohamed. On the use of the K-chart for phase II monitoring of simple linear profiles. **Journal of Quality and Reliability Engineering**. v. 2013, p.1-8, 2013b. doi: 10.1155/2013/705450

GARJANI, M; NOOOROSSANA, R.; SAGHAEI, A. A neural network-based control scheme for monitoring start-up processes and short runs. **The International Journal of Advanced Manufacturing Technology**, v. 52, p. 1023-1032, 2010.

GAURI, S. K.; CHAKRABORTY, S. Improved recognition of control chart patterns using artificial neural networks. **International Journal of Advanced Manufacturing Technology**, v. 36, n. 11-12, p. 1191-1201, 2008.

GAURI, S. K.; CHAKRABORTY, S. Recognition of control chart patterns using improved selection of features. **Computers & Industrial Engineering**, v. 56, n. 4, p. 1577-1588, 2009.

GHIASABADI, A.; NOOROSSANA, R.; SAGHAEI, A. Identifying change point of a non-random pattern on control chart using artificial neural networks. **The International Journal of Advanced Manufacturing Technology**, v. 67, p. 1623-1630, 2013. GHOMI, S. M. T. F.; LESANY, S. A.; KOOCHAKZADEH, A. Recognition of unnatural patterns in process control charts through combining two types of neural networks. **Applied Soft Computing**, v. 11, p. 5444-5456, 2011.

GIL, Antônio C. **Como elaborar projetos de pesquisa**. 4. ed. São Paulo: Atlas, 2002.

GUH, R. S.; SHIUE, Y. R. An effective application of decision tree learning for on-line detection of mean shifts in multivariate control charts. **Computers & Industrial Engineering**, v. 55, n. 2, p. 475-493, 2008.

GUH, R. S. Simultaneous process mean and variance monitoring using artificial neural networks. **Computers & Industrial Engineering**, v. 58, n. 4, p. 739-753, 2010.

GUJARATI, Damodar N. Econometria básica. São Paulo: Elsevier Campus, 2006.

GUO, Peng; FU, Jian; YANG, XiYun. Condition monitoring and fault diagnosis of wind turbines gearbox bearing temperature based on Kolmogorov-Smirnov test and convolutional neural network model. **Energies**, v. 11, p. 1-16, 2018. doi: 10.3390/en11092248

GUNN, Steve. R. **Support vector machines for classification and regression**. Technical Report. University of Southampton, 1998. Disponível em: <http://www.dec.usc.es/persoal/cernadas/tc03/mc/SVM.pdf>. Acesso em: 20 mar. 2017.

GUTIERREZ, H. T.; PHAM, D. T. Estimation and generation of training patterns for control chart pattern recognition. **Computers & Industrial Engineering**, v. 95, p. 72-82, 2016.

HACHICHA, Wafik; GHORBEL, Ahmed. A survey of control-chart pattern-recognition literature (1991–2010) based on a new conceptual classification scheme. **Computers & Industrial Engineering**, v. 63, p. 204-222, 2012.

HAN, J.; KAMBER, M.; PEI, J. **Data mining: concepts and techniques**. 3rd ed. San Francisco: Morgan Kaufmann Publishers, 2011.

HASTIE, Trevor; TIBSHIRANI, Robert; FRIEDMAN, Jerome. **The elements of statistical learning: data mining, inference, and prediction**. 2nd ed., New York: Springer, 2013.

HAYKIN, Simon. **Neural networks and learning machines**. 3rd ed. New York: Prentice Hall, 2009.

HAWKINS, Douglas M. Multivariate quality control based on regression-adjusted variables. **Technometrics**, v. 33, n. 1, p. 61-75, 1991.
HE, Shuguang; JIANG, Wei; DENG, Houtao. A distance-based control chart for monitoring multivariate processes using support vector machines. **Annals of Operations Research**, v. 263, n. 1, p. 191-207, 2016.

HE, Shiming; XIAO, Long; WANG, Yalin, LIU, Xinggao; YANG, Chunhua; LU, Jiangang; GUI, Weihua; SUN, Youxian. A novel fault diagnosis method. Based on optimal relevance vector machine. **Neurocomputing**, p. 1-13, 2017.

HSU, Chih-Wei; CHANG, Chih-Chung; LIN, Chih-Jen. A practical guide to support vector classication. National Taiwan University. 2010. Disponível em: ">http://www.csie.ntu.edu.tw/~cjlin>. Acesso em: 22 mar. 2017.

HSU, Chun-Chin; CHEN, Mu-Chen; CHEN, Long-Sheng. Integrating independent component analysis and support vector machine for multivariate process monitoring. **Computers & Industrial Engineering**, v. 59, p. 145-156, 2010a.

HSU, Chun-Chin; CHEN, Mu-Chen; CHEN, Long-Sheng. Intelligent ICA–SVM fault detector for non-Gaussian multivariate process monitoring. **Expert Systems with Applications**, v. 37, p. 3264-3273, 2010b.

HU, Sheng; ZHAO, Liping; YAO, Yiyong; DOU, Rushan. A variance change point estimation method based on intelligent ensemble model for quality fluctuation analysis. **International Journal of Production Research**, v. 54, n. 19, p. 5783-5797, 2016. doi: 10.1080/00207543.2016.1178862

HUANG, Jian; YAN, Xuefeng. Related and independent variable fault detection based on KPCA and SVDD. **Journal of Process Control**, v. 39, p. 88–99, 2016.

HUWANG, Longcheen; WANG, Yi-Hua T.; XUE, Shuhan; ZOU, Changliang. Monitoring general linear profiles using simultaneous confidence sets schemes. **Computers & Industrial Engineering**, v. 68, p. 1-12, 2014.

HWARNG, H. Brian; HUBELE, Norma F. Back-propagation pattern recognizers for Xbar control charts: methodology and performance. **Computers & Industrial Engineering**, v. 24, n. 2, p. 219-235, 1993.

HWARNG, H. B.; WANG, Y. Shift detection and source identification in multivariate autocorrelated processes. **International Journal of Production Research**, v. 48, n. 3, p. 835-859, 2010.

IMANI, Moslem; KAO, Huan-Chin; LAN, Wen-Hau; KUO, Chung-Yen. Daily sea level prediction at Chiayi coast, Taiwan using extreme learning machine e relevance vector machine. **Global and Planetary Change**, v. 161, p. 211-221, 2018.

ISSAM, B. K., MOHAMED, L. Support vector regression based residual MCUSUM control chart for autocorrelated process. **Applied Mathematics and Computation**, v. 201, p. 565-574, 2008.

IZMAILOV, Alexey; SOLODOV, Mikhail. Otimização: condições de otimalidade, elementos de análise convexa e de dualidade. v. 1. Rio de Janeiro: IMPA, 2005.

JAMES, Gareth; WITTEN, Daniela; HASTIE, Trevor; TIBSHIRANI, Robert. **An introduction to statistical learning with applications in R**. New York: Springer, 2014.

JEARKPAPORN, D., MONTGOMERY, D. C., RUNGER, G. C., BORROR, C. M. Model-based process monitoring using robust generalized linear models. International Journal of Production Research, v. 43, n. 7, p. 1337-1354, 2005.

JIANG, P.; LIU, D.; ZENG, Z. Recognizing control chart patterns with neural network and numerical fitting. **Journal of Intelligent Manufacturing**, v. 20, n. 6, p. 625-635, 2009.

JONES-FARMER, L. A.; WOODALL, William H.; STEINER, Stefan H.; CHAMP, Charles W. An overview of phase I analysis for process improvement and monitoring. **Journal of Quality Technology**, v. 46, n. 3, p. 265-280, 2014.

JURAN, Joseph M.; GODFREY, A. Blanton. **Juran's quality handbook**. 5th ed. New York: McGraw-Hill, 1999.

KAO, Ling-Jing; LEE, Tian-Shyug; LU, Chi-Jie. A multi-stage control chart pattern recognition scheme based on independent component analysis and support vector machine. **Journal of Intelligent Manufacturing**, v. 27, p. 653-664, 2016.

KAZEMI, M. S.; KAZEMI, K.; YAGHOOBI, M. A.; BAZARGAN, H. A hybrid method for estimating the process change point usingsupport vector machine and fuzzy statistical clustering. **Applied Soft Computing**, v. 40, p. 507-516, 2016.

KECMAN, Vojislav. Learning and soft computing: support vector machines, neural networks and fuzzy logic models. London: MIT Press, 2001.

KHEDIRI, Issam B.; WEIHS, Claus; LIMAM, Mohamed. Support vector regression control charts for multivariate nonlinear autocorrelated processes. **Chemometrics and Intelligent Laboratory Systems**, v. 103, p. 76-81, 2010.

KHEDIRI, Issam B.; WEIHS, Claus; LIMAM, Mohamed. Kernel k-means clustering based local support vector domain description fault detection of multimodal processes. **Expert Systems with Applications**, v. 39, p. 2166-2171, 2012.

KHUSNA, H.; MASHURI, M.; SUHARTONO, S.; PRASTYO, D. D.; AHSAN, M. Multioutput least square SVR-based multivariate EWMA control chart: the performance evaluation and application. **Cogent Engineering**, v. 5, n. 1, p. 1-14, 2018.

KIM, S. B.; JITPITAKLERT, W.; PARK, S. K.; HWANG, S. J. Data mining modelbased control charts for multivariate and autocorrelated processes. **Expert Systems with Applications**, v. 39, p. 2073-2081, 2012. KHORMALI, Aminollah; ADDEH, Jalil. A novel approach for recognition of control chart patterns: type-2 fuzzy clustering optimized support vector machines. **ISA Transactions**, v. 63, p. 256-264, 2016.

KUHN, Max; JOHNSON, Kjell. **Applied predictive modeling**. New York: Springer, 2013.

LEE, Seulki; KIM, Seoung B. Time-adaptive support vector data description for nonstationary process monitoring. **Engineering Applications of Artificial Intelligence**, v. 68, p. 18-31, 2018.

LIMA, Clodoaldo A. de M. **Comitê de máquinas: uma abordagem unificada empregando máquinas de vetores-suporte**. 2004. 377 f. Tese (Doutorado em Engenharia Elétrica) – Pós-graduação da Faculdade de Engenharia Elétrica e de Computação, Universidade Estadual de Campinas, Campinas, 2004.

LIN, S. Y; GUH, R. S.; SHIUE, R. Effective recognition of control chart patterns inautocorrelated data using a support vector machine based approach. **Computers & Industrial Engineering**, v. 61, p. 1123-1134, 2011.

LIU, Yiqi. Adaptive just-in-time and relevant vector machine based soft-sensorswith adaptive differential evolution algorithms for parameter optimization. **Chemical Engineering Science**, v. 172, p. 571-584, 2017.

LOUZADA, Francisco; DINIZ, Carlos A.R.; FERREIRA, Paulo H.; FERREIRA, Edil L. **Controle estatístico de processos: uma abordagem prática para cursos de engenharia e administração.** Rio de Janeiro: LTC, 2013.

LU, C. J.; SHAO, Y. E.; LI, P. H. Mixture control chart patterns recognition using independent component analysis and support vector machine. **Neurocomputing**, v. 74, p. 1908-1914, 2011.

LV, Zhaomin; YAN, Xuefeng. Hierarchical support vector data description for batch process monitoring. **Industrial & Engineering Chemistry Research**, v. 55, n. 34, p. 9205-9214, 2016. doi: 10.1021/acs.iecr.6b00901

LV, Zhaomin; YAN, Xuefeng; JIANG, Qingchao. Batch process monitoring based on self-adaptive subspace support vector data description. **Chemometrics and Intelligent Laboratory Systems**, v. 170, p. 25-31, 2017.

MALEKI, M. R., AMIRI, A. Simultaneous monitoring of multivariate-attribute process mean and variability using artificial neural networks. **Journal of Quality Engineering and Production Optimization**, v. 1, n. 1, p. 43-54, 2015.

MANDEL, B. J. The regression control chart. **Journal of Quality Technology**, v. 1, n. 1, p. 1-9, 1969.

MASOOD, Ibrahim; HASSAN Adnan. Issues in development of artificial neural network-based control chart pattern recognition schemes. **European Journal of Scientific Research**, v. 39, n. 3, p. 336-355, 2010.

MITRA, Amitava. **Fundamentals of quality control and improvement**. 2nd ed., New York: Prentice Hall, 1998.

MITCHELL, Tom M. Machine learning. New York: McGraw-Hill, 1997.

MONTGOMERY, D. C.; VINING, G. G.; PECK, E. A. Introduction to Linear Regression Analysis. 3a. ed., New York: John Wiley & Sons, 2001.

MONTGOMERY, Douglas C. Introdução ao controle estatístico da qualidade. 4a ed. Rio de Janeiro: LTC, 2004.

MONTEGOMERY, Douglas C.; RUNGER, George C. **Estatística aplicada e probabilidade para engenheiros**. Rio de Janeiro: LTC, 2003.

MOORE, Gregory; BERGERON, Charles; BENNETT, Kristin P. Model selection for primal SVM. **Machine Learning**, v. 85, p. 175-208, 2011.

NAEINI, M. K.; BAYATI, N. Pattern recognition in control chart using neural network based on a new statistical feature. **International Journal of Engineering**, v. 30, n. 9, p. 1372-1380, 2017.

NIAKI, S. T. A.; ABBASI, B. Detection and classification mean shifts in multi-attribute processes by artificial neural networks. **International Journal of Production Research**, v. 46, p. 2945-2963, 2008.

NIAKI, S. T. A., NAFAR, M. An Artificial Neural Network approach to monitor and diagnose multi-attribute quality control processes. **Journal of Industrial Engineering International**, v. 4, n. 7, p. 10-24, 2008.

NIAKI, S. T. A.; DAVOODI, M. Designing a multivariate multistage quality control system using artificial neural networks. **International Journal of Production Research**, v. 47, p. 251-271, 2009.

NIAKI, Seyed T. A.; NASAJI, Shirin A. A hybrid method of artificial neural networks and simulated annealing in monitoring auto-correlated multi-attribute processes. **International Journal of Advanced Manufacturing Technology**, v. 56. p. 777-788, 2011.

NING, Xianghi; TSUNG Fugee. Improved design of kernel-distance-based charts using support vector methods. **IIE Transactions**, v. 45, p. 464-476, 2013.

NOOROSSANA, R.; ATASHGAR, K.; SAGHAEI, A. An integrated supervised learning solution for monitoring process mean vector. **International Journal of Advanced Manufacturing Technology**, v. 56, p. 755-765, 2011.

OAKLAND, John S. **Statistical process control**. 5th ed. Oxford: Butterworth-Heinemann, 2003. OLIVEIRA, Silvio L. Tratado de metodologia científica: projetos de pesquisas, TGI, TCC, monografias, disertações e teses. 2. ed. São Paulo: Pioneira, 1999.

ONEL, Melis; KIESLICH, Chris A.; GUZMAN Yannis A.; FLOUDAS, Christodoulos A.; PISTIKOPOULOS, Efstratios N. Big data approach to batch process monitoring: simultaneous fault detection and diagnosis using nonlinear support vector machine-based feature selection. **Computers and Chemical Engineering**, v. 115, p. 46-63, 2018.

PALIWAL, M., KUMAR U.A. Neural networks and statistical techniques: a review of applications. **Expert System and Applications**, v. 36, p. 2-17, 2009.

PACELLA, M.; SEMERARO, Q. Monitoring roundness profiles based on an unsupervised neural network algorithm. **Computers & Industrial Engineering**, v. 60, n. 4, p. 677-689, 2011.

PEDRINI, D. C., CATEN, C.S. Modelagem estatística para a previsão do teor de fósforo em ligas de ferromanganês. **Revista Ingepro**, v. 2, p. 14-25, 2010.

PRICE, K.; STORN, R.; LAMPINEN, J. **Differential evolution: a practical approach to global optimization.** New York: Springer-Verlag, 2006.

PSARAKIS, S.; PAPALEONIDA, G. E. A. SPC procedures for monitoring autocorrelated processes. **Quality Technology & Quantitative Management**, v. 4, p. 501-540, 2007.

PSARAKIS, Stelios. The use of neural networks in statistical process control charts. **Quality and Reliability Engineering International**, v. 27, p. 641-650, 2011.

PUGH, G. Allen. Synthetic neural networks for process control. **Computers & Industrial Engineering**, v. 17, p. 24-26, 1989.

PUGH, G. Allen. A comparison of neural networks to SPC charts. **Computers & Industrial Engineering**, v. 21, p. 253-255, 1991.

RAMOS, Edson M. L. S.; ALMEIDA, Sílvia dos S.; ARAÚJO, Adrilayne dos R. **Controle estatístico da qualidade**. Porto Alegre: Bookman, 2013.

RANAEE, V.; EBRAHIMZADEH, A. Control chart pattern recognition using a novel hybrid intelligent method. **Applied Soft Computing**, v. 11, p. 2676-2686, 2011.

RANAEE, V.; EBRAHIMZADEH, A.; GHADERI, R. Application of the PSO-SVM model for recognition of control chart patterns. **ISA Transactions**, v. 49, p. 577-586, 2010.

REZENDE, Solange O. **Sistemas inteligentes: fundamentos e aplicações**. São Paulo: Manole, 2005.

REZKI, Nafissa; KAZAR, Okba; MOUSS, Leila H.; KAHLOUL, Laid; REZKI, Djamil. On the use of multi-agent systems for the monitoring of industrial systems. **Journal of Industrial Engineering International**, v. 12, p. 111-118, 2016.

RIBEIRO, José L. D.; CATEN, Carla S. **Controle estatístico do processo**. Série Monográfica Qualidade. Porto Alegre: Editora da UFRGS, 2012.

RIVAS-PEREA, Pablo; COTA-RUIZ, Juan; CHAPARRO, David G.; VENZOR, Jorge A. P.; CARREÓN, Abel Q.; ROSILES, Jose G. Support vector machines for regression: a succinct review of large-scale and linear programming formulations. **International Journal of Intelligence Science**, v. 3, p. 5-14, 2013. doi: 10.4236/ijis.2013.31002

SALCEDO-SANZ, S.; ROJO-ÁLVAREZ, J. L.; MARTÍNEZ-RAMÓN, M.; CAMPS-VALLS, G. Support vector machines in engineering: an overview. **WIREs Data Mining and Knowledge Discovery**, v. 4, p. 234-267, 2014. doi: 10.1002/widm.1125

SALEHI, M.; BAHREININEJAD, A.; NAKHAI, I. On-line analysis of out-of-control signals in multivariate manufacturing processes using a hybrid learning-based model. **Neurocomputing**, v. 74, p. 2083-2095, 2011.

SALEHI, M.; KAZEMZADEH, R. B.; SALMASNIA, A. On line detection of mean and vari-ance shift using neural networks and support vector machine in multivariateprocesses. **Applied Soft Computing**, v. 12, p. 2973-2984, 2012.

SAMOHYL, Robert W. **Controle estatístico de qualidade**. Rio de Janeiro, RJ: Elsevier, 2009.

SANT'ANNA, Ângelo. M. O. Ferramentas para modelagem e monitoramento de características de qualidade do tipo fração. 2009. 150 f. Tese (Doutorado em Engenharia de Produção) – Escola de Engenharia, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2009.

SANT'ANNA, Ângelo M. O.; CATEN, Carla S. Beta control charts for monitoring fraction data. **Expert Systems with Applications**, v. 39, p. 10236-10243, 2012.

SANTOS, G. S.; LUVIZOTTO, L. G. J; MARIANI, V. C.; COELHO, L. S. Least squares support vector machines with tuning based on chaotic differential evolution approach applied to the identification of a thermal process. **Expert Systems with Applications**, v. 39, p. 4805-4812, 2012.

SCARSELLI, Franco; TSOI, Ah C. Universal approximation using feed-forward neural networks: a survey of some existing methods and some new results. **Neural Networks**, v. 11, p. 15-37, 1998.

SCHÖLKOPF, Bernhard; SMOLA, Alexander J. **Support vector machines and kernel algorithms**, 2002. Disponível em: http://alex.smola.org/papers/2002/SchSmo02b.pdf>. Acesso em: 20 mar. 2016.

SHAO, Yuehjen E.; CHIU, Chih-Chou. Applying emerging soft computing approaches to control chart pattern recognition for an SPC–EPC process. **Neurocomputing**, v. 201, p. 19-28, 2016.

SHEWHART, W. A. Quality Control. **Bell System Technical Journal**, p.722-735, 1926.

SILVA, Edna L.; MENEZES, Estera M. **Metodologia da pesquisa e elaboração de dissertação**. 4a ed. Florianópolis: Laboratório de Ensino a Distância da UFSC, 2005.

SILVA, Ivan N. da; SPATTI, Danilo H.; FLAUZINO, Rogério A. **Redes neurais** artificiais: para engenharia e ciências aplicadas. São Paulo: Artliber: 2010.

SIQUEIRA, Luiz G. P. **Controle estatístico do processo**. São Paulo: Pioneira, 1997.

SMITH, Alice E. X-Bar and R control charts interpretation using neural networks. **International Journal of Production Research**, v. 32, p. 309-320, 1994.

SMOLA, Alex J.; SCHÖLKOPF, Bernhard. A tutorial on support vector regression. **Statistics and Computing**, v.14. n. 3, p. 199-222, 2004.

STORN, Rainer; PRICE, Kenneth. Differential evolution – a simple and efficient heuristic for global optimization over continuous spaces. **Journal of Global Optimization**, v. 11, p. 341-359, 1997.

STORN, R. Differential evolution research - trends and open questions. In: Chakraborty, U. K. (Ed.), **Advances in differential evolution. Studies in Computational Intelligence**, v. 143, p. 11-12, 2008.

SUKCHOTRAT, T.; KIM, S. B.; TSUNG, F. One-class classification-based control charts for multivariate process monitoring. **IIE Transactions**, v. 42, p. 107-120, 2010.

SUN, Ruixiang; TSUNG, Fugee. A kernel-distance-based multivariate control chart using support vector methods. **International Journal of Production Research**, v. 41, n. 13, p. 2975-2989, 2003.

TAX, D.M. J.; YPMA, A.; DUIN, R. P. W. Support vector data description applied to machine vibration analysis. **Proceedings of the Fifth Annual Conference of the Advanced School for Computing and Imaging**, Heijen, Netherlands, p. 398-405, 1999.

TAX, David M. J.; DUIN, Robert P.W. Support vector data description. **Machine** Learning, v. 54, p. 45-66, 2004.

TEOH, W. L.;FUN, M. S.; TEH, S. Y.; KHOO, M. B. C.; YEONG, W. C. Exact run length distribution of the double sampling X chart with estimated process parameters. **South African Journal of Industrial Engineering**, v. 27, n. 1, p. 20-31, 2016.

TIPPING, M. The relevance vector machine. In: SOLLA, S. A.; LEEN, T. K.; MÜLLER, K. R. (Eds.), **Advances in neural information processing systems**, v. 12, p. 652-658, MIT Press, 2000.

TIPPING, M. E. Sparse Bayesian learning and the relevance vector machine. **Journal of Machine Learning Research**, v. 1, p. 211–244, 2001.

TIPPING, M. E., FAUL, A. C. Fast marginal likelihood maximization for sparse Bayesian models. In: BISHOP, C. M.; FREY, B.J. (Eds.), **Proceedings of the Ninth International Workshop on Artificial Intelligence and Statistics**, Key West, Florida, p. 3-6, 2003.

TIPPING, M. E. Bayesian inference: an introduction to principles and practice in machine learning. In: BOUSQUET, O; VON LUXBURG, U.; RÄATSCH, G. (Eds.), Advanced Lectures on Machine Learning, Springer, p. 41-62, 2004.

TÔRRES, Adamastor R. Cartas de controle multivariadas aplicadas na revisão periódica de produtos e no estudo de estabilidade em uma indústria farmacêutica nacional. 2015. 97 f. Tese (Doutorado em Química) – Pós-Graduação em Química, Universidade Federal da Paraíba, João Pessoa, 2015.

UM, H.; LEE, K.; KIM, K-Y.; SHIN, G.; CHUNG, Y. Effect of carbon content of ferromanganese alloy on corrosion behaviour of MgO-C refractory. **Ironmaking & Steelmaking**, v. 41, p. 31-27, 2014.

VAPNIK, Vladimir; GOLOWICH, Steven E.; SMOLA, Alex. Support vector method for function approximation, regression estimation, and signal processing. **Advances in Neural Information Processing Systems**, v. 9, p. 281-287, 1997.

VAPNIK, Vladimir N. **Statistical learning theory**. New York: John Wiley & Sons, 1998.

VERDÉRIO, Adriano. Sobre o uso de regressão por vetores suporte para a construção de modelos em um método de região de confiança sem derivadas. 2015. 115 f. Tese (Doutorado em Matemática) – Pós-graduação em Matemática, Universidade Federal do Paraná, 2015.

VERMA, M.; THIRUMALAISELVI, A.; RAJASANKAR, J. Kernel-based models for prediction of cement compressive strength. **Neural Computing & Applications**, v. 28, n. 1, p. 1083-1100, 2017.

VICINI, Lorena. **Análise multivariada da teoria à prática**. Santa Maria, RS: UFSM, 2005.

WANG, W.; XU, Z.; LU, W.; ZHANG X. Determination of the spread parameter in the Gaussian kernel for classification and regression. **Neurocomputing**, v. 55, p. 643-663, 2003.

WANG, Long; ZHANG, Zijun; LONG, Huan. Wind turbine gearbox failure identification with deep neural networks. **IEEE Transactions on Industrial Informatics**, v. 13, n. 3, p. 1360-1368, 2017.

WANG, Jianlin; LIU, Weimin; QIU, Kepeng; YU, Tao; ZHAO, Liqiang. Dynamic hypersphere based support vector data description for batch process monitoring. **Chemometrics and Intelligent Laboratory Systems**, v. 172, p. 17-32, 2018.

WEESE, Maria; MARTINEZ, Waldyn; MEGAHED, Fadel M.; JONES-FARMER, L. Alisson. Statistical learning methods applied to process monitoring: an overview and perspective. **Journal of Quality Technology**, v. 48, n. 1, p. 4-27, 2016.

WERKEMA, Maria C. C. As ferramentas da qualidade no gerenciamento de processos. Belo Horizonte: Universidade Federal de Minas Gerais, 1995.

WERKEMA, Maria C. C.; AGUIAR, Sílvio. Análise de regressão: como entender o relacionamento entre as variáveis de um processo. Belo Horizonte: Fundação Christiano Ottoni, 1996.

WESTERN ELECTRIC. Western Electric Company. **Statistical quality control handbook**. Easton: Mack Printing Company, 1958.

WOODALL, W.H., MONTGOMERY, D.G. Research issues and ideas in statistical process control. **Journal of Quality Technology**, v. 31, p. 376-385, 1999.

WU, B.; YU, J. A. neural network ensemble model for on-line monitoring of process mean and variance shifts in correlated processes. **Expert Systems with Applications**, v. 37, p. 4058-4065, 2010.

WU, Cang; LIU, Fei; ZHU, Bo. Control chart pattern recognition using an integrated model based on binary-tree support vector machine. **International Journal of Production Research**, v. 53, n. 7, p. 2026-2040, 2015.

XANTHOPOULOS, Petros; RAZZAGHI, Talayeh. A weighted support vector machine method for control chart pattern recognition. **Computers & Industrial Engineering**, v. 70, p. 134-149, 2014.

XIE, L.; GU, N.; LI, D.; CAO, Z.; TAN, M.; NAHAVANDI, S. Concurrent control chart patterns recognition with singular spectrum analysis and support vector machine. **Computers & Industrial Engineering**, v. 64, n. 1, p. 280-289, 2013.

YAMPIKULSAKUL, Nattavut; BYON, Eunshin; HUANG, Shuai; SHENG, Shuangwen; YOU, Mingdi. Condition monitoring of wind power system with nonparametric regression analysis. **IEEE Transactions on Energy Conversion**, v. 29, n. 2, p. 288-299, 2014.

YANG, Wen-An; ZHOU, Wei. Autoregressive coefficient-invariant control chart pattern recognition in autocorrelated manufacturing processes using neural network ensemble. **Journal of Intelligent Manufacturing**, v. 26, p. 1161-1180, 2015.

YANG, Wen-An; ZHOU, Wei; LIAO, Wenhe; GUO, Yu. Identification and quantification of concurrent control chart patterns using extreme-point symmetric mode decomposition and extreme learning machines. **Neurocomputing**, v. 147, p. 260-270, 2015.

YANG, Wen-An. Simultaneous monitoring of mean vector and covariance matrix shifts in bivariate manufacturing processes using hybrid ensemble learning-based model. **Journal of Intelligent Manufacturing**, v. 27, p. 845-874, 2016.

YANG, Cong; QIAN, Zheng; PEI, Yan; WEI, Lu. A data-driven approach for condition monitoring of wind turbine pitch systems. **Energies**, v. 11, p. 1-17, 2018. doi: 10.3390/en11082142

YAO, Ma; WANG, Huangang; XU, Wenli. Batch process monitoring based on functional data analysis and support vector data description. **Journal of Process Control**, v. 24, p. 1085-1097, 2014.

YU, Lean; WANG, Shouyang; LAI, K. K. An integrated data preparation scheme for neural network data analysis. **IEEE Transactions on Knowledge and Data Engineering**, v. 18, n. 2, p. 1-14, 2006.

YU, Jianbo; XI, Lifeng. A Neural network ensemble-based model for on-line monitoring and diagnosis of out-of-control signals in multivariate manufacturing processes. **Expert Systems with Applications**, v. 36, p. 909-921, 2009.

YU, J.; XI, L.; ZHOU, X. Identifying source(s) of out-of- control signals in multivariate manufacturing processes using selective neural network ensemble. **Engineering Applications of Artificial Intelligence**, v. 22, p. 141-152, 2009.

ZAMAN, Babar; RIAZ, Muhammad; AHMAD, Shabbir; ABBASI, Saddam A. On artificial neural networking-based process monitoring under bootstrapping using runs rules schemes. **International Journal of Advanced Manufacturing Technology**, v. 76, p. 311-327, 2015.

ZHANG, Min; CHENG, Wenming. Recognition of mixture control chart pattern using multiclass support vector machine and genetic algorithm based on statistical and shape features. **Mathematical Problems in Engineering**, v. 2015, p. 1-10, 2015. doi: 10.1155/2015/382395

ZHANG, Min; CHENG, Wenming; GUO, Peng. Intelligent recognition of mixture control chart pattern based on quadratic feature extraction and SVM with AMPSO. **Journal of Coastal Research**, v. 73, p. 304-309, 2015.

ZHANG, Chaolong; HE, Yigang; YUAN, Lifeng; XIANG, Sheng; WANG, Jinping. Prognostics of lithium-ion batteries based on wavelet denoising and DE-RVM. **Computational Intelligence and Neuroscience**, v. 2015, p. 1-8, 2015. doi: 10.1155/2015/918305 ZHANG, Chi; TSUNG, Fugee; ZHOU, Changliang. A general framework for monitoring complex processes with both in-control and out-of-control information. **Computers & Industrial Engineering**, v. 85, p. 157-168, 2015.

ZHANG, Min; YUAN, Yi; WANG, Ruiqi; CHENG, Wenming. Recognition of mixture control chart patterns based on fusion feature reduction and fireworks algorithm-optimized MSVM. **Pattern Analysis and Applications**, p. 1-12, 2018. doi: 10.1007/s10044-018-0748-6

ZHAO, Yang; WANG, Shengwei; XIAO, Fu. A statistical fault detection and diagnosis method for centrifugal chillers based on exponentially-weighted moving average control charts and support vector regression. **Applied Thermal Engineering**, v. 51, p. 560-572, 2013.

ZHIQIANG, Cheng; YIZHONG, Ma. Control chart pattern recognition based on PSO and LS-SVM. **The Hong Kong Institution of Engineers Transactions**, v. 17, n. 1, p. 7-10, 2010.

ZHOU, Xueliang; JIANG, Pingyu; WANG, Xianxiang. Recognition of control chart patterns using fuzzy SVM with a hybrid kernel function. **Journal of Intelligent Manufacturing**, v. 29, p. 51-67, 2015.

ZORRIASSATINE, F; TANNOCK, J. D. T. A review of neural networks for statistical process control. **Journal of Intelligent Manufacturing**, v. 9, p. 209-224, 1998.