

UNIVERSIDADE FEDERAL DO PARÁ
INSTITUTO DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

PREVISÃO MULTI-PASSOS À FRENTE DO PREÇO DE ENERGIA ELÉTRICA DE
CURTO PRAZO NO MERCADO BRASILEIRO

JOSÉ CARLOS RESTON FILHO

TD: 10/2014

UFPA/ ITEC/ PPGEE
Campus Universitário do Guamá
Belém – Pará – Brasil
2014

UNIVERSIDADE FEDERAL DO PARÁ
INSTITUTO DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

JOSÉ CARLOS RESTON FILHO

PREVISÃO MULTI-PASSOS A FRENTE DO PREÇO DE ENERGIA ELÉTRICA DE
CURTO PRAZO NO MERCADO BRASILEIRO

TD: 10/2014

UFPA/ ITEC/ PPGEE
Campus Universitário do Guamá
Belém – Pará – Brasil
2014

UNIVERSIDADE FEDERAL DO PARÁ
INSTITUTO DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

JOSÉ CARLOS RESTON FILHO

PREVISÃO MULTI-PASSOS A FRENTE DO PREÇO DE ENERGIA ELÉTRICA DE
CURTO PRAZO NO MERCADO BRASILEIRO

Tese submetida à Banca Examinadora do Programa de Pós-Graduação em Engenharia Elétrica da UFPA para a obtenção do Grau de Doutor em Engenharia Elétrica na área de computação aplicada.

UFPA/ ITEC/ PPGEE
Campus Universitário do Guamá
Belém – Pará – Brasil
2014

Dados Internacionais de Catalogação-na-Publicação (CIP)
Sistema de Bibliotecas da UFPA

Reston Filho, José Carlos, 1974-

Previsão multi-passos a frente do preço de energia
de curto prazo no mercado brasileiro / José Carlos
Reston Filho. - 2014.

Orientadora: Carolina de Mattos Affonso;

Coorientador: Roberto Célio Limão de

Oliveira.

Tese (Doutorado) - Universidade Federal do
Pará, Instituto de Tecnologia, Programa de
Pós-Graduação em Engenharia Elétrica, Belém,
2014.

1. Energia elétrica - comercialização. 2.
Redes neurais (computação). 3. Análise de séries
temporais. I. Título.

CDD 22. ed. 621.31

UNIVERSIDADE FEDERAL DO PARÁ
INSTITUTO DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

PREVISÃO MULTI-PASSOS A FRENTE DO PREÇO DE ENERGIA ELÉTRICA DE
CURTO PRAZO NO MERCADO BRASILEIRO

AUTOR: JOSÉ CARLOS RESTON FILHO

TESE DE DOUTORADO SUBMETIDA À AVALIAÇÃO DA BANCA EXAMINADORA
APROVADA PELO COLEGIADO DO PROGRAMA DE PÓS-GRADUAÇÃO EM
ENGENHARIA ELÉTRICA DA UNIVERSIDADE FEDERAL DO PARÁ E JULGADA
ADEQUADA PARA OBTENÇÃO DO GRAU DE DOUTOR EM ENGENHARIA
ELÉTRICA NA ÁREA DE COMPUTAÇÃO APLICADA.

APROVADA EM: ____/____/____

BANCA EXAMINADORA:

Prof^a. Dr^a. Carolina de Mattos Affonso
(ORIENTADORA – UFPA)

Prof. Dr. Roberto Célio Limão de Oliveira
(CO-ORIENTADOR – UFPA)

Prof. Dr. Takashi Yoneyama
(AVALIADOR EXTERNO – ITA)

Prof^a. Dr^a. Elizete Maria Lourenço
(AVALIADOR EXTERNO – UFPR)

Prof. Dr. ÁDAMO LIMA DE SANTANA
(AVALIADOR INTERNO – UFPA)

VISTO:

Prof. Dr. EVALDO GONÇALVES PELAES
(COORDENADOR DO PPGEE / ITEC / UFPA)

Dedicatória

Dedico este trabalho aos meus pais, José Carlos Reston e M^a Leonarda da C. Reston, como forma de registrar a minha gratidão, pela oportunidade de estudar, da infância até a universidade, em condições privilegiadas.

Dedico também a minha esposa, Luciana Botelho Reston, amizade raríssima, meu porto seguro, meu amor.

AGRADECIMENTOS

A Professora Dra. Carolina de Mattos Affonso e ao Professor Dr. Roberto Célio Limão de Oliveira, pela confiança depositada, pela orientação competente, segura e lúcida que recebi. Registro aqui o meu reconhecimento pela oportunidade, pelo incentivo, pela paciência e pela disponibilidade ao longo da elaboração deste trabalho.

Aos amigos João Evangelista, Carlos Vasquez Seisdedos, Hiran de Melo, José Ruben Sicchar e a todos os colegas de curso, pelas contribuições, pela palavra amiga e pelas mensagens de ânimo e encorajamento sempre presentes na trajetória deste doutoramento.

Aos sócios, Vicente Fernandes Tino e Daniel Diniz Fregapani, pelo apoio recebido sempre que precisei me ausentar da empresa para dedicação mais ostensiva ao doutoramento. Sem o suporte de vocês não teria sido possível a conclusão exitosa deste projeto.

A todos os Professores do Programa de Pós-Graduação em Engenharia Elétrica - PPGEE, sem exceção, pelos ensinamentos fundamentais a minha formação.

Este trabalho foi desenvolvido com o apoio do Governo do Estado do Amazonas por meio Fundação de Amparo à Pesquisa do Estado do Amazonas, com a concessão de bolsa de estudo.

“Se queres prever o futuro, estuda o passado”

Confúncio

LISTA DE FIGURAS

Figura 2.1	Instituições do setor elétrico brasileiro	11
Figura 2.2	Diagrama esquemático dos 4 submercados brasileiros	12
Figura 2.3	Capacidade instalada no SIN	13
Figura 2.4	Processo de decisão para um sistema hidrotérmico	14
Figura 2.5	Curvas de custo do uso da água	15
Figura 2.6	Comportamento do PLD para o submercado SE/CO	18
Figura 2.7	Processo de comercialização de energia no curto-prazo	19
Figura 2.8	Ambientes de contratação	19
Figura 3.1	Classificação de metodologias de previsão de preços	23
Figura 3.2	Histograma de energia armazenada	30
Figura 3.3	Metodologia Box-Jenkins para a modelagem de séries temporais	31
Figura 3.4	Modelo do neurônio artificial	36
Figura 3.5	Elementos de uma árvore de decisão	44
Figura 3.6	Hiper-retângulos da árvore de decisão	45
Figura 4.1	Fragmento do IPDO do dia 07/06/2003 com dados da ONS	50
Figura 4.2	Séries temporais do banco de dados para região nordeste	51
Figura 4.3	Diagrama de blocos do modelo híbrido preditor	53
Figura 4.4	Diagrama de blocos da solução de classificação	56
Figura 5.1	Predição PLD até t+12 com modelo híbrido para a região S	63
Figura 5.2	Predição PLD até t+12 com modelo híbrido para a região SE/CO	63
Figura 5.3	Predição PLD até t+12 com modelo híbrido para a região NE	64
Figura 5.4	Predição PLD até t+12 com modelo híbrido para a região N	64
Figura 5.5	Indicadores de erro para cada modelo nos 4 submercados	65
Figura 5.6	Patamares dos valores do PLD	68
Figura 5.7	Variáveis explanatórias selecionadas	69
Figura 5.8	Regras para valores muito altos de PLD para a região N	72
Figura 5.9	Regras para valores altos de PLD para a região N	72

LISTA DE TABELAS

Tabela 3.1	Matriz de confusão para um problema de duas classes	42
Tabela 4.1	Atributos componentes do banco de dados	50
Tabela 4.2	Medidas de posição e dispersão da região SE/CO	52
Tabela 4.3	Medidas de posição e dispersão da região S	52
Tabela 4.4	Medidas de posição e dispersão da região NE	52
Tabela 4.5	Medidas de posição e dispersão da região N	52
Tabela 5.1	Teste f ANOVA das variáveis explanatórias dos 4 submercados	60
Tabela 5.2	Modelos ARIMA obtidos para os 4 submercados	61
Tabela 5.3a	Topologias das RN aplicadas ao submercado S	62
Tabela 5.3b	Topologias das RN aplicadas ao submercado N	62
Tabela 5.3c	Topologias das RN aplicadas ao submercado NE	62
Tabela 5.3d	Topologias das RN aplicadas ao submercado SE/CO	62
Tabela 5.4	Correlação linear (CL) e Desvio Padrão (DP) do preço PLD	66
Tabela 5.5	Resultados do teste de Kupiec para previsões t+12	66
Tabela 5.6	Critério EMPA horizontes superiores de predição futura	67
Tabela 5.7	Indicadores EMPC e APC para 3 modelos classificadores	70
Tabela 5.8	EMPC para até 5 passos a frente de classificadores da região N	70
Tabela 5.9	Resultado modelos combinados 2 a 2 para a região N	71
Tabela 5.10	Matriz de confusão do modelo C50 + CART para região N	71

LISTA DE ABREVIATURAS

% M.L.T.	Média de longo termo
ACF	Função de autocorrelação
ACL	Ambiente de contratação livre
ACR	Ambiente de contratação regulada
AIDS	Síndrome da imunodeficiência adquirida
ANEEL	Agência nacional de energia elétrica
ANOVA	<i>Analysis of variance</i>
APC	Acurácia percentual de classificação
AR	<i>Autoregressive</i>
ARIMA	<i>Autoregressive integrated moving average</i>
ARIMAX	<i>ARIMA with explanatory variables</i>
ARMA	<i>Autoregressive moving average</i>
ARMAX	<i>ARMA with explanatory variables</i>
ARX	<i>AR with explanatory variables</i>
AWNN	<i>Adaptive wavelet neural network</i>
C5.0	Modelo classificador baseado em regras
CARGA	Carga de energia
CART	<i>Classification and regression tree</i>
CCEAR	Contratos de comercialização de energia no ambiente regulado
CCEE	Câmara de comercialização de energia elétrica
CL	Correlação linear
CMO	Custos marginais de operação
CMSE	Comitê de monitoramento do setor elétrico
CNPE	Conselho nacional de pesquisa energética
DP	Desvio padrão
EARM	Energia armazenada
EMA	Erro médio absoluto
EMPA	Erro médio percentual absoluto
EMPC	Erro médio percentual de classificação
ENA	Energia natural afluyente
EPE	Empresa de pesquisa energética
ET	Erro total
EUA	Estados Unidos da América

f ANOVA	Teste que avalia a relação linear entre variáveis de saída e entrada
FCF	Função de custo futuro
FCI	Função de custo imediato
GARCH	<i>Generalized autoregressive conditional heteroskedasticity process</i>
GerHid	Geração hidráulica
GerTer	Geração térmica
IPDO	Informativo preliminar diario da operação
LRPOF	<i>Likelihood ratio for Probability of failure</i>
LSSVM	Least squares support vector machine
MA	<i>Moving average</i>
MAE	Mercado atacadista de energia elétrica
MDG	Média dentro do grupo
MEG	Média entre grupos
MME	Ministério de minas e energia
N	Norte
NE	Nordeste
N_{err}	Número de amostras classificadas erradas
N_{tot}	Número total de amostras
ONS	Operador nacional do sistema
PACF	Função de autocorrelação parcial
PCA	<i>Principal components analysis</i>
PCM	<i>Production cost model</i>
PIB	Produto interno bruto
P_i^{Prev}	Valor predito do PLD
P_i^{real}	Valor real do PLD
PJM	<i>Pennsylvania-New Jersey-Maryland</i>
PLD	Preço de liquidação de diferenças
PROINFA	Programa de incentivo às fontes alternativas de energia elétrica
REMQ	Raiz do erro médio quadrático
RNA	Redes neurais
S	Sul
SARIMA	<i>Seasonal autoregressive integrated moving average</i>
SARIMAX	<i>SARIMA with explanatory variables</i>
SE/CO	Sudeste/Centro-oeste

SIN	Sistema interligado nacional
SPCM	<i>Strategic production cost model</i>
SQDG	Soma quadrada dentro do grupo
SQEG	Soma quadrada entre grupos
URL	<i>Uniform Resource Locator</i>

RESUMO

A predição do preço da energia elétrica é uma questão importante para todos os participantes do mercado, para que decidam as estratégias mais adequadas e estabeleçam os contratos bilaterais que maximizem seus lucros e minimizem os seus riscos. O preço da energia tipicamente exhibe sazonalidade, alta volatilidade e picos. Além disso, o preço da energia é influenciado por muitos fatores, tais como: demanda de energia, clima e preço de combustíveis. Este trabalho propõe uma nova abordagem híbrida para a predição de preços de energia no mercado de curto prazo. Tal abordagem combina os filtros autorregressivos integrados de médias móveis (ARIMA) e modelos de Redes Neurais (RNA) numa estrutura em cascata e utiliza variáveis explanatórias. Um processo em dois passos é aplicado. Na primeira etapa, as variáveis explanatórias são preditas. Na segunda etapa, os preços de energia são preditos usando os valores futuros das variáveis exploratórias. O modelo proposto considera uma predição de 12 passos (semanas) a frente e é aplicada ao mercado brasileiro, que possui características únicas de comportamento e adota o despacho centralizado baseado em custo. Os resultados mostram uma boa capacidade de predição de picos de preço e uma exatidão satisfatória de acordo com as medidas de erro e testes de perda de cauda quando comparado com técnicas tradicionais. Em caráter complementar, é proposto um modelo classificador composto de árvores de decisão e RNA, com objetivo de explicitar as regras de formação de preços e, em conjunto com o modelo preditor, atuar como uma ferramenta atrativa para mitigar os riscos da comercialização de energia.

PALAVRAS CHAVE: predição de preços de energia, multi-passos à frente, mercado de curto-prazo, ARIMA, redes neurais artificiais, seleção de variáveis explanatórias.

ABSTRACT

Electricity price forecasting is an important issue to all Market participants in order to decide bidding strategies and to establish bilateral contracts, maximizing their profits and minimizing their risks. Energy price typically exhibits seasonality, high volatility and spikes. Also, energy price is influenced by many factors such as power demand, weather, and fuel price. This work proposes a new hybrid approach for short-term energy price prediction. This approach combines auto-regressive integrated moving average (ARIMA) and neural network (ANN) models in a cascaded structure and uses explanatory variables. A two step procedure is applied. In the first step, the selected explanatory variables are predicted. In the second one, the energy prices are forecasted by using the explanatory variables prediction. The proposed model considers a multi-step ahead price prediction (12 weeks-ahead) and is applied to Brazilian market, which adopts a cost-based centralized dispatch with unique characteristics of price behavior. The results show good ability to predict spikes and satisfactory accuracy according to error measures and tail loss test when compared with traditional techniques. Additionally, is proposed a classifier model consisting of ANN and decision trees in order to explain the rules of price formation and, together with the predictor model, acting as an attractive tool to mitigate the risks of energy trading.

KEYWORDS: electricity price forecasting, multi-step ahead, short-term market, ARIMA, neural networks, explanatory variables selection.

SUMÁRIO

	Capítulo I – Introdução	1
1.1	Motivação	1
1.2	Objetivos	3
1.3	Revisão da literatura	4
1.4	Contribuições do trabalho	6
1.5	Estrutura da tese	7
1.6	Considerações finais	8
	Capítulo II - Panorama da comercialização de energia no Brasil	9
2.1	Visão histórica	9
2.2	O novo marco regulatório	10
2.3	O Sistema Interligado Nacional	12
2.4	O modelo <i>tight pool</i>	13
2.5	O cálculo do Preço de Liquidação de Diferenças - PLD	16
2.6	Processo de comercialização de energia	18
2.7	Ambientes de contratação	19
2.8	Considerações finais	21
	Capítulo III - Métodos para a predição e classificação de séries temporais	23
3.1	Introdução	23
3.2	Pré-processamento	25
3.2.1	Seleção de variáveis exploratórias	25
3.2.2	Redução de dimensionalidade	27
3.2.3	Balanceamento de amostras	29
3.3	Metodologia Box-Jenkins	30
3.3.1	Modelos autorregressivos (AR)	32
3.3.2	Modelos de médias móveis (MA)	33
3.3.3	Modelos autorregressivos de médias móveis (ARMA)	33
3.3.4	Modelos autorregressivos integrados de médias móveis (ARIMA)	33

3.4	Redes Neurais Artificiais	35
3.4.1	O algoritmo <i>Backpropagation</i>	38
3.4.2	Sobre-treinamento	39
3.5	Árvores de decisão	40
3.5.1	Indução de árvores de decisão	44
3.5.2	Algoritmos de indução de árvores de decisão	46
3.6	Considerações finais	46
	Capítulo IV - Metodologia	48
4.1	Representação da solução	48
4.2	Banco de dados	49
4.3	Modelo híbrido proposto	53
4.4	Modelo de classificação proposto	55
4.5	Considerações finais	58
	Capítulo V – Resultados, validação experimental e análise	59
5.1	Introdução	59
5.2	Resultados obtidos utilizando o modelo híbrido	59
5.2.1	Seleção de variáveis explanatórias	59
5.2.2	Modelo ARIMA	60
5.2.3	Topologia das redes neurais	61
5.2.4	Resultados para predição do PLD	62
5.3	Resultados obtidos utilizando o modelo de classificação	67
5.3.1	Definição de classes do PLD	67
5.3.2	Seleção das variáveis explanatórias	68
5.3.3	Resultados para classificação do PLD	69
5.4	Considerações finais	72
	Capítulo VI - Conclusões e trabalhos futuros	74
	Conclusões e recomendações para trabalhos futuros	74
	Referências	77

CAPÍTULO I

INTRODUÇÃO

1.1 Motivação

Antes da década de 1980, a energia elétrica era vista como um monopólio natural de propriedade pública na Europa e de propriedade privada nos Estados Unidos, na maioria dos casos. No Brasil, o setor elétrico era totalmente estatal. Assim, empresas estatais tinham o direito exclusivo de suprir energia elétrica para os consumidores de varejo residencial, comercial e industrial, dentro de uma estabelecida área geográfica, sendo responsáveis pela geração, transmissão, distribuição e comercialização da energia elétrica.

Este quadro começou a mudar com a reforma no setor elétrico dos países desenvolvidos, com a separação das atividades de geração, transmissão, distribuição e comercialização de energia (descentralização do setor), introduzindo competição nas atividades, visando torná-la mais eficiente, mais transparente e com custos menores de energia. No Brasil este processo iniciou-se principalmente pela necessidade do governo de privatizar seus ativos para pagamento de suas dívidas e pelas pressões de grupos estrangeiros interessados no mercado brasileiro, com taxas de crescimento superiores ao PIB do país.

Embora o preço médio da energia elétrica em vários países tenha caído significativamente na fase inicial das reformas, tem-se observado desde então um aumento da volatilidade nos preços de energia no mercado à vista ou *spot* em diversos países como, por exemplo, EUA (Califórnia especialmente), Austrália e Brasil. O mercado de energia possui características físicas singulares que fazem com que a ocorrência de picos súbitos e

temporários (*price spikes*) do preço de energia seja maior que em quaisquer outros mercados de *commodities* (Mayo, 2009). As razões que causam a volatilidade dos preços da energia são:

- A impossibilidade de armazenamento da energia para venda futura;
- A incerteza da demanda;
- A inelasticidade de preço da demanda;
- A hidrologia;
- O perfil íngreme ascendente da função de custo da geração;
- O longo tempo de resposta (*lead-time*) necessário para a expansão da geração a fim de atender o crescimento da demanda.

Segundo Melo et al (2011), durante os anos 1990, o modelo regulatório brasileiro de energia elétrica começou a ser reestruturado com o objetivo de promover a concorrência e atrair investimentos para o setor. Esta primeira fase da reforma do setor foi marcada pela implementação do Projeto de Reestruturação do Setor Elétrico Brasileiro. Os objetivos básicos da reestruturação foram a introdução da concorrência nas atividades de geração e comercialização e a construção de uma regulação eficiente nas atividades de transmissão e distribuição.

A reforma no mercado de energia elétrica trouxe a elevação dos riscos associados à contratação de energia. O ‘quanto’ e ‘quando’ comprar no mercado *spot* são decisões que passam a ser cruciais para as empresas do setor elétrico. Como tais decisões dependem especialmente do preço da energia elétrica, um julgamento mal feito ou uma estratégia de investimentos mal traçada hoje poderá, no futuro, acarretar um prejuízo de milhares a milhões de dólares.

Segundo Freire et al (2012), as taxas de adesão aos contratos de curto prazo no mercado brasileiro de energia apresentam uma expressiva variabilidade de mês a mês, enquanto as taxas de adesão aos contratos de longo prazo aumentam de forma suave e discreta ao longo do tempo.

Por isso, desenvolver modelos de previsão de preços no mercado de curto prazo para reduzir possíveis riscos associados, são grandes desafios que o setor energia elétrica está enfrentando. A previsão de valores futuros do preço da energia e de seus possíveis picos é de

grande a importância para o processo de tomada de decisão e elaboração de estratégias de comercialização de energia.

Para Catalão et al (2011), as ferramentas de previsão de preço são essenciais para todos os participantes do mercado para a sua sobrevivência em ambiente competitivo. No curto prazo, o produtor precisa de previsões de preços para organizar o seu calendário de produção e para derivar a sua estratégia de lance no mercado de energia .

Para Liu et al (2014) entre os possíveis fatores de risco no ambiente do mercado de curto prazo de eletricidade, a volatilidade do preço é o que mais contribui para o risco financeiro em uma carteira. Assim, antever os momentos de maior volatilidade pode auxiliar bastante na tomada de decisão e na gestão de riscos da carteira.

1.2 Objetivos

O principal objetivo desta tese de doutorado é propor um modelo de previsão futura do preço de energia no mercado de curto prazo Brasileiro, capaz de estimar o preço da energia até 12 passos (semanas) a frente. A metodologia proposta baseia-se em um sistema híbrido composto pelos modelo autorregressivo integrado de médias móveis (ARIMA) e Redes Neurais (RNA), utilizando variáveis explanatórias.

Além disso, esta tese de doutorado propõe um modelo de classificação do preço de energia futura utilizando árvores de decisão, numa abordagem personalizada para prever o comportamento dos preços futuros em 5 patamares de preço previamente especificados. Para tal são utilizados o C5.0 e CART de modo combinando. Com a árvore de decisão pode-se obter a explicitação do conhecimento, evidenciando a combinação de fatores que geram maior volatilidade no preço de energia no Brasil, e assim criar uma estratégia de gerenciamento de risco eficaz para participantes do mercado de curto prazo de energia elétrica no Brasil.

Assim, esta tese de doutorado tem como objetivos específicos:

- Propor um modelo híbrido para previsão do preço futuro de energia 12 passos a frente composto pelos modelos ARIMA e Redes Neurais utilizando variáveis explanatórias;
- Propor um modelo de classificação do preço de energia futura utilizando árvores de decisão, com finalidade de previsão, combinando os algoritmos C5.0 e CART;
- Analisar e discutir os fatores que mais impactam a formação de preços de energia no mercado Brasileiro em patamares de preço considerados alto;

1.3 Revisão da literatura

Diversas pesquisas tem sido feitas com o objetivo de desenvolver modelos eficientes capazes de prever o valor futuro da energia. Para Aggarwal et al (2009a) e Gonzales et al (2005) os modelos de previsão dos preços de eletricidade podem ser classificados principalmente em três grupos: teoria dos jogos, modelos fundamentais e modelos de séries temporais.

O primeiro grupo é baseado na teoria dos jogos, que analisa o comportamento estratégico dos agentes e seu impacto nos preços da eletricidade, considerando-se como chave os modelos de equilíbrio do mercado de eletricidade (como equilíbrio de Nash). Yang et al (2013) apresenta um modelo que usa a teoria dos jogos na análise da demanda estimada versus o consumo real. No trabalho de Molina *et al.*(2010) é desenvolvido um modelo baseado no equilíbrio de Nash-Cournot, específico para o mercado chileno de energia. Assim também ocorre no trabalho de Villar e Rudnick (2003). Já Carpio e Pereira Jr. (2007) desenvolveram um modelo semelhante, também baseado na teoria dos jogos, mas voltado para o mercado brasileiro.

O segundo grupo é baseado em modelos fundamentais que simulam sistemas de potência capazes de satisfazer as demandas a um custo mínimo, modelando o impacto dos fatores físicos e econômicos que afetam o preço da eletricidade. Esta abordagem expressa os preços da eletricidade com base nos custos marginais de geração, considerando congestionamento de transmissão, perdas, e outras solicitações de serviços auxiliares em mercados de energia. Pode-se mencionar o trabalho de Battle e Barquín (2005), denominado *Strategic Production Cost Model* – SPCM, que é uma evolução do modelo *Production Cost Model* – PCM, capaz de fazer previsões a partir do custo de produção e bem adaptado ao mercado atacadista de energia. Dois outros trabalhos sugeridos para análise são Vehviläinen e Pyykkönen (2005) e Munoz *et al.* (2009). Em ambos, a abordagem para a modelagem de preços se dá em cima das relações físicas e econômicas presentes na produção e comercialização de eletricidade.

Finalmente, o terceiro grupo é baseado em modelos de séries temporais, que inclui modelos baseados em regressão, tais como ARIMA, Redes Neurais, *Generalized Autoregressive Conditional Heteroskedasticity Process* (GARCH), lógica fuzzy e outros. Estes modelos estão focados principalmente na análise de dados, sem focar os processos

físicos subjacentes em detalhe. Eles usam séries históricas de preços, por vezes, outros fatores explicativos, como temperatura, hora do dia, e a demanda de carga. Modelos quantitativos que caracterizam as propriedades estatísticas do preço da eletricidade ao longo do tempo, com o objetivo de avaliação de derivativos e gestão de riscos. Os trabalhos desenvolvidos por Weron e Misiorek (2008), Garcia-Martos *et al.* (2007), Aggarwal *et al.* (2009) e Contreras *et al.* (2003) e Bowden e Payne (2008) abordam de maneira quantitativa o problema da predição futura dos preços de eletricidade, a partir da replicação das propriedades estatísticas dos preços de energia elétrica. Nestes trabalhos são construídos modelos estocásticos ou econométricos, utilizando as metodologias de predição de séries temporais capazes de auxiliar na avaliação de derivativos. Alguns trabalhos ampliam os modelos de predição de carga de energia para a predição de preços de energia, também a partir de métodos estatísticos. Nesta categoria se enquadram os trabalhos de Burguer *et al.* (2004), Davison *et al.* (2002) e Anderson (2004). Já as técnicas baseadas na inteligência artificial conseguem lidar bem com a complexidade e as não linearidades. Há muitos trabalhos com o uso de modelagem a partir de redes neurais, lógica difusa, mineração de dados e uso combinado de técnicas, chamados modelos híbridos. Nesta área destacam-se os seguintes trabalhos: Pindoriya *et al.* (2008), Yun *et al.* (2008), Catalão *et al.* (2011), Gareta *et al.* (2006), Mandal *et al.* (2006), Pao (2007), Amjady (2006), Amjady e Hemmati (2009), Lora *et al.* (2007), Meng *et al.* (2008), Fan *et al.* (2007) e Zareipour *et al.* (2006).

Recentemente, métodos de previsão híbridos com resultados promissores têm sido propostos na literatura. A maior vantagem dos modelos híbridos é combinar os pontos fortes de diferentes técnicas que estabelecem um quadro de modelagem robusta. Por exemplo, Conejo *et al.* (2005) propuseram um método híbrido baseado em transformada wavelet e modelos ARIMA para a previsão dos preços da energia para o dia seguinte no mercado de Espanha. Os autores concluem que o uso da transformada wavelet como um pré-processador de dados de previsão melhora o comportamento de previsão de modelos ARIMA. Já Yan e Showdhury (2013) propuseram máquina de vetor de suporte de mínimos quadrados (LSSVM) combinada com um filtro autorregressivo de média móvel com variáveis exógenas (ARMAX) para prever o preço de hora em hora por um mês inteiro no mercado PJM (Pennsylvania-New Jersey-Maryland). O LSSVM é utilizado para prever os preços iniciais da eletricidade. Em seguida, o método ARMAX é utilizado para melhorar os resultados da previsão com o LSSVM. O modelo proposto melhorou a previsão de precisão em comparação com um modelo de previsão usando um único LSSVM. No entanto, o método ainda tem baixa

acurácia na previsão de preços de pico. Em Wu e Shahidehpour (2010) um modelo híbrido é proposto que combina métodos de séries temporais (ARMAX + GARCH) e um modelo adaptativo rede neural wavelet (AWNN) para previsão de preço para o dia seguinte no mercado PJM. Neste método, a carga de energia é utilizada como variável explicativa. O modelo ARMAX é usado para capturar a relação linear entre preço e carga. O modelo GARCH é usado para mostrar a heteroscedasticidade dos resíduos e AWNN é usada para apresentar o impacto não-linear e não-estacionário da carga nos preços da eletricidade. Amjady e Keynia (2008) propuseram uma combinação de transformada wavelet, RNA e um algoritmo evolutivo para a previsão do preço do dia seguinte no mercado PJM. Os resultados obtidos foram comparados com outras técnicas tradicionais e apresentaram indicadores de erros menores.

Ao contrário dos métodos de previsão de carga que têm erros de previsão inferiores a 3%, as técnicas de previsão de preço de energia ainda estão em seus estágios iniciais de maturidade e os erros relatados geralmente variam de 5% a 36% (Zareipur et al, 2011), e variam de acordo com a técnica utilizada e do mercado analisado. Uma revisão dos diferentes métodos de previsão dos preços da eletricidade pode ser encontrada em Aggarwal et al (2009) e a maioria dos métodos são aplicados aos mercados PJM (*Pennsylvania-New Jersey-Maryland Interconnection*), Nova Inglaterra e Espanha. É importante mencionar que apenas alguns estudos foram desenvolvidos na área de previsão do preço da energia aplicada ao mercado brasileiro, como é o caso dos trabalhos de Reston Filho et al (2009) e Queiroz et al (2007).

No que tange a classificação de preços destacam-se os trabalhos de Zareipour et al (2011), Zhao et al (2007) e Yan e Chowdhury (2014). A classificação de preços de curto prazo é uma alternativa para a previsão numérica de preço. Na classificação de preços, as previsões são feitas com respeito a se o preço está acima ou abaixo dos limites de preços pré-especificadas e definidos por usuários com base em seus objetivos de operação e planejamento. A classificação de preços é especificamente útil quando o valor exato de preços futuros não é criticamente importante.

1.4 Contribuições do trabalho

As principais contribuições deste trabalho são o próprio modelo de previsão e sua aplicação para o mercado brasileiro, que adota um modelo baseado em custos com

características únicas do comportamento dos preços de energia. Alguns aspectos importantes dos métodos de previsão de preço propostos são:

- Combina os modelos ARIMA e RNA, e não assume que a relação entre os componentes lineares e não-lineares é aditiva (pode ser multiplicativo por exemplo), o que pode degradar o desempenho do preditor;
- Utiliza a técnica de seleção de variáveis explanatórias;
- O modelo proposto é simples, visto que é composto de blocos constitutivos bastante explorados na literatura específica, utilizados nas configurações mais usuais;
- Adota um horizonte temporal de previsão de 12-passos à frente (semanas);
- O modelo proposto é aplicado ao mercado brasileiro.

O modelo híbrido proposto foi aplicado para o mercado brasileiro e os resultados foram comparados com outras técnicas disponíveis na literatura aplicada a outros mercados. Os resultados mostram medidas de erro semelhantes ou melhores do que outros métodos.

1.5 Estrutura da Tese

Este trabalho está dividido em seis capítulos assim constituídos:

O Capítulo 1 apresenta os aspectos introdutórios sobre a problemática do assunto, contendo a justificativa, objetivos a serem alcançados, revisão bibliográfica e a descrição e organização dos capítulos.

O Capítulo 2 apresenta a fundamentação teórica sobre a comercialização de energia elétrica no Brasil, com o detalhamento do mercado de curto prazo, a metodologia de formação do preço de liquidação de diferenças, as modalidades de contratação e os atores envolvidos nos ambientes de contratação, abordando as especificidades do mercado Brasileiro.

O Capítulo 3 versa sobre as técnicas de predição de séries temporais, dando especial ênfase aos métodos utilizados: ARIMA, Redes Neurais e árvores de decisão. Ainda neste capítulo é abordada a redução da dimensão do espaço de entrada pelo uso da técnica de Análise dos Componentes Principais (PCA) e a estratégia de balanceamento de amostras de treinamento utilizadas nas redes neurais.

O Capítulo 4 mostra os procedimentos metodológicos desenvolvidos no trabalho, destacando a representação da solução, a montagem do banco de dados, o pré-processamento dos dados, a solução de modelagem híbrida utilizando as RNA associadas ao ARIMA e a explicitação de conhecimento a partir das árvores de classificação e regressão. No Capítulo 5 são apresentados os resultados obtidos com os modelos de previsão de preço de energia propostos, além de realizar a avaliação da precisão obtida por meio destes resultados.

O Capítulo 6 apresenta as principais conclusões obtidas com esta pesquisa além de sugestões para trabalhos futuros

1.6 Considerações Finais

Este capítulo apresentou os principais objetivos desta tese de doutorado, além da contribuição e relevância deste estudo. O próximo capítulo irá apresentar os principais aspectos relacionados ao mercado brasileiro de energia, do modelo utilizado e o método do cálculo do preço de energia no mercado de curto prazo.

CAPÍTULO II

PANORAMA DA COMERCIALIZAÇÃO DE ENERGIA NO BRASIL

2.1 Visão histórica

A reforma do setor elétrico brasileiro iniciou no início dos anos 90, a partir de um processo de privatizações das empresas federais (Mayo, 2009). Dois eram os objetivos fundamentais neste processo:

- ❖ Assegurar os investimentos necessários para expansão da oferta de energia, uma vez que havia uma percepção de esgotamento da capacidade do estado de investir em infraestrutura para atender o aumento da demanda;
- ❖ Assegurar que o setor fosse economicamente eficiente, utilizando os recursos disponíveis para garantir um suprimento confiável de energia elétrica ao menor custo possível.

Para Silva (2001), os pilares da reforma do segmento de energia elétrica se apoiam na introdução da competição e na imposição de que os preços reflitam os custos reais, coerentemente com o uso. A competição surge a partir da desverticalização, onde há a separação das atividades de geração, transmissão e distribuição, surgindo o segmento da comercialização de energia. As consequências do livre acesso à transmissão por todos os agentes de mercado seriam:

- ❖ Redução do preço da energia elétrica;
- ❖ Ganho de eficiência da indústria de energia elétrica;
- ❖ Atração do capital privado.

De acordo com Souza (2008), no momento posterior a desverticalização, surgiu a necessidade da criação de um órgão regulador (a Agência Nacional de Energia Elétrica - ANEEL), de um operador para o sistema elétrico nacional (Operador Nacional do Sistema Elétrico - ONS) e de um ambiente para a realização das transações de compra e venda de energia elétrica (o Mercado Atacadista de Energia Elétrica – MAE).

No entanto, o modelo regulatório não foi totalmente implementado. Contribuíram para este fato as deficiências do modelo em não contemplar as especificidades técnicas e institucionais do setor elétrico e a insuficiência do aparato regulatório. Após o racionamento de energia elétrica de 2001 foi instituído o Comitê de Revitalização do Modelo do Setor Elétrico, cujo trabalho resultou em um conjunto de propostas de alterações no setor elétrico brasileiro.

2.2 O novo marco regulatório

Durante os anos de 2003 e 2004 o Governo Federal lançou as bases de um novo modelo para o Setor Elétrico Brasileiro, sustentado pelas Leis nº 10.847 e 10.848, de 15 de março de 2004; e pelo Decreto nº 5.163, de 30 de julho de 2004.

O modelo contém um conjunto de medidas a serem cumpridas pelos agentes, tais como a obrigatoriedade da contratação de totalidade da demanda por parte das distribuidoras e dos consumidores livres, um novo método de cálculo do lastro para venda de geração, contratação de usinas hidrelétricas e termelétricas em proporções que assegurem melhor equilíbrio entre garantia e custo de suprimento, bem como o monitoramento permanente da continuidade e da segurança de suprimento, visando detectar desequilíbrios conjunturais entre oferta e demanda.

Em termos de modicidade tarifária, o modelo prevê a compra de energia elétrica pelas distribuidoras no ambiente regulado por meio de leilões – observado o critério de menor tarifa, objetivando a redução do custo de aquisição da energia elétrica a ser repassada para a tarifa dos consumidores cativos.

A inserção social busca promover a universalização do acesso e do uso do serviço de energia elétrica, criando condições para que os benefícios da eletricidade sejam disponibilizados aos cidadãos que ainda não contam com esse serviço, e garantir subsídio para os consumidores de

baixa renda, de tal forma que estes possam arcar com os custos de seu consumo de energia elétrica.

O novo marco regulatório do setor definiu a criação da(o):

- Empresa de Pesquisa Energética (EPE): entidade responsável pelo planejamento do setor elétrico a longo prazo;
- Comitê de Monitoramento do Setor Elétrico (CMSE): instituição com a função de avaliar permanentemente a segurança do suprimento de energia elétrica;
- Câmara de Comercialização de Energia Elétrica (CCEE): instituição para comercialização de energia elétrica no Sistema Interligado Nacional Brasileiro;

Além disso, houve a definição do exercício do Poder Concedente ao Ministério de Minas e Energia (MME) e a ampliação da autonomia do ONS. A Figura 2.1 apresenta as instituições do setor elétrico brasileiro.

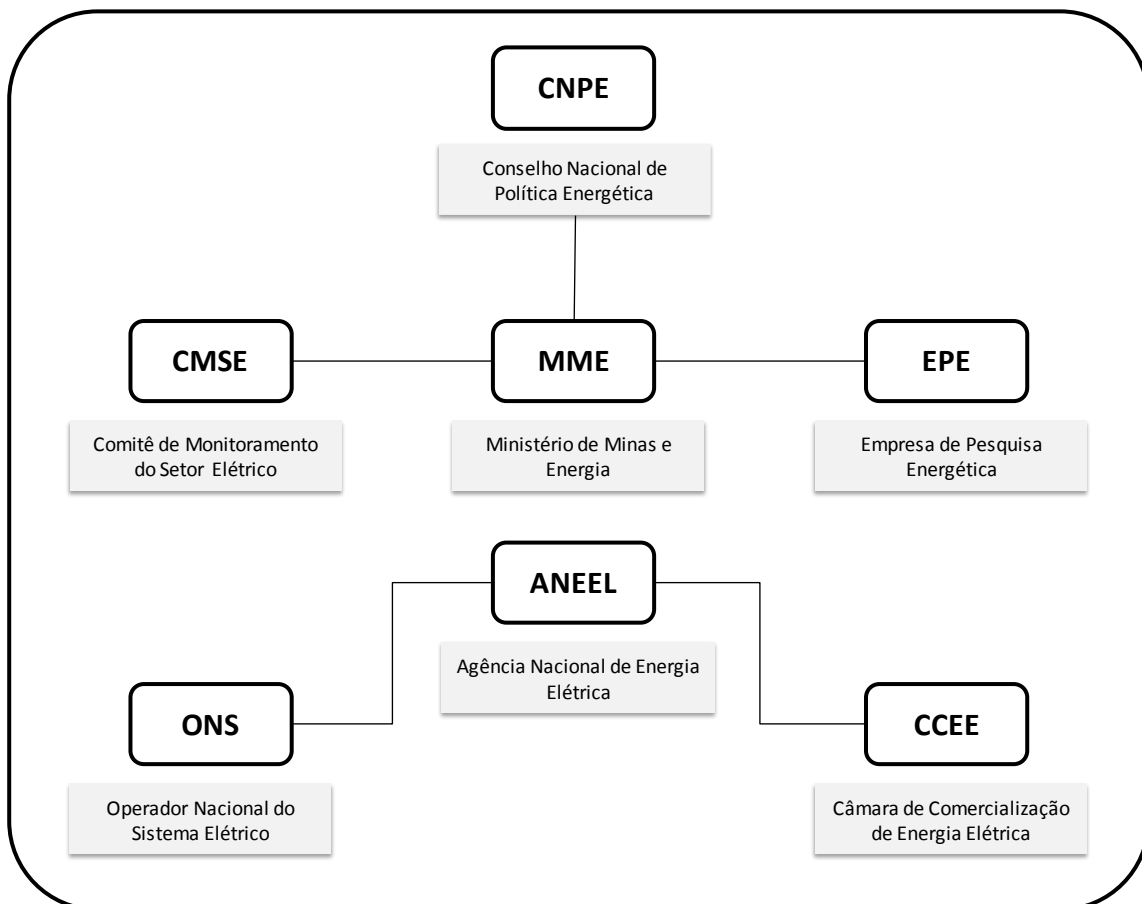


Figura 2.1 - Instituições do setor elétrico brasileiro. Fonte: site da CCEE – www.ccee.org.br

2.3 O Sistema Interligado Nacional

O Sistema Interligado Nacional (SIN) é formado pelas regiões Sul, Sudeste, Centro-Oeste, Nordeste e parte da região Norte. Apenas 3,4% da capacidade de produção de eletricidade do país está fora do SIN, em pequenos sistemas isolados localizados principalmente na região amazônica. A malha de transmissão existente transfere energia entre as regiões geo-elétricas do país aproveitando-se da diversidade hidrológica entre as mesmas e reduzindo os custos operativos do sistema. Porém, devido às restrições de transmissão entre as regiões geo-elétricas, o mercado de eletricidade brasileiro foi dividido em 4 (quatro) submercados como mostra a Figura 2.2.

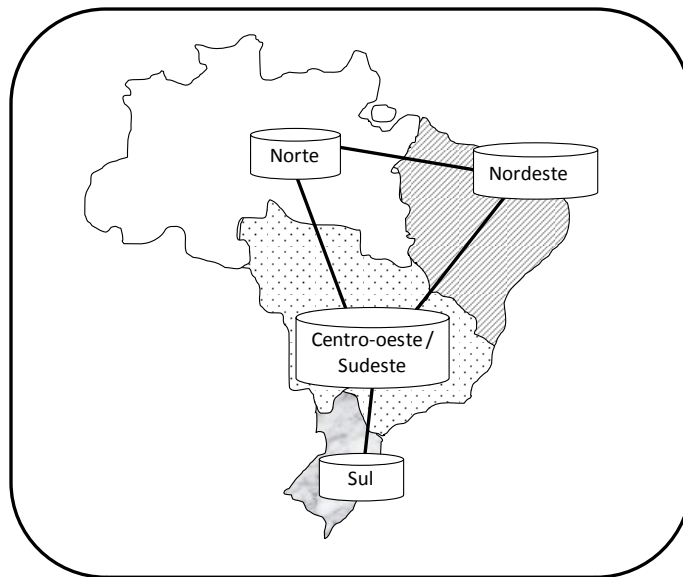


Figura 2.2 - Diagrama esquemático dos 4 submercados brasileiros. Fonte: site da CCEE – www.ccee.org.br

O sistema de geração de energia elétrica Brasileiro é um sistema hidrotérmico de grande porte, com forte predominância de usinas hidrelétricas e com múltiplos proprietários. Isto pode ser constatado pela figura 2.3, que ilustra a estrutura da capacidade instalada do SIN por tipo de fonte de energia, referente a 20/12/2013, de acordo com o ONS. Em 2013, o sistema apresentou 79,1 GW relativos a usinas hidráulicas, sendo 7 GW de Itaipu. As usinas térmicas respondem por 32,3 GW, fonte nuclear representam 2 GW, e geração eólica representam 1,63 GW.

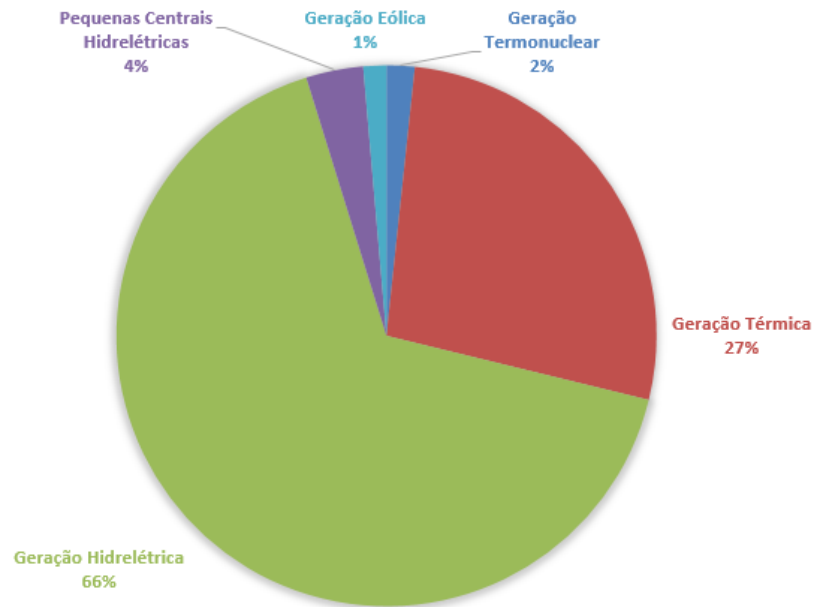


Figura 2.3 – Capacidade instalada no SIN em 20/12/2013. Fonte: site do ONS – www.ons.org.br

2.4 O modelo *tight pool*

Devido à predominância hidráulica do parque gerador brasileiro, foi adotado o modelo de despacho centralizado (*tight pool*), em que o ONS, com base em uma cadeia de modelos de otimização do uso da água estocada nos reservatórios, define o montante a ser despachado por cada usina integrante do sistema interligado (Souza, 2008). O objetivo deste modelo é definir uma política de operação de modo a suprir a carga e minimizar o custo de operação do sistema.

Para Silva (2001), diferentemente dos sistemas térmicos, qualquer decisão tomada em uma etapa do planejamento em sistemas hidrotérmicos, com base hidráulica, traz embutida uma consequência futura e um risco associado, pois a afluência futura é uma variável aleatória e existe um limite na capacidade de armazenamento dos reservatórios. A Figura 2.4 ilustra este processo.

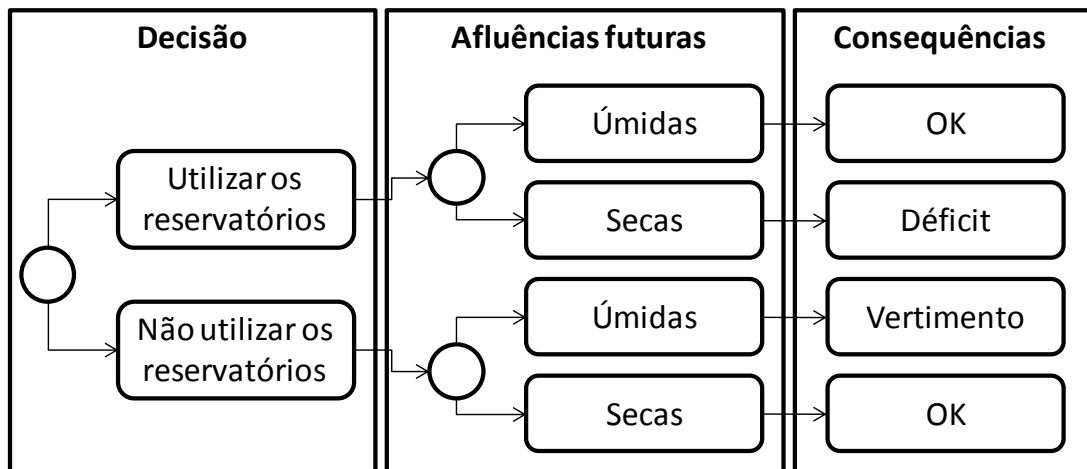


Figura 2.4 - Processo de decisão para um sistema hidrotérmico. Fonte: Silva (2001).

Nota-se que a decisão de utilizar os reservatórios do sistema leva fatalmente a duas possíveis consequências operativas.

- Se as afluências futuras forem altas (período úmido), tem-se uma operação econômica;
- Se as afluências futuras forem baixas (período seca), tem-se um déficit, com possíveis cortes de carga ou medidas de racionamento de energia.

Por outro lado, a decisão de não utilizar os reservatórios e sim a geração térmica presente leva também a duas possíveis consequências operativas futuras:

- Se as afluências futuras forem baixas (período seca), tem-se uma operação econômica
- Se as afluências futuras forem altas (período úmido), tem-se vertimento, que não chega a ser tão danoso quanto um déficit, mas indica um desperdício de energia.

Assim, a operação do sistema hidrotérmico deve considerar não apenas os custos de geração térmica em um estágio t (normalmente mês ou semana), mas também o custo esperado dessa geração e do déficit no final desse estágio t (início de $t+1$). Matematicamente, esta lógica pode ser expressa pela “Função de Custo Imediato” (FCI), representando o benefício do uso imediato da água, e pela “Função de Custo Futuro” (FCF), representando o benefício de armazená-la hoje para seu uso no futuro, conforme ilustra a Figura 2.5 (Silva, 2001).

A Função de Custo Imediato (FCI) mede os custos de geração térmica no momento t . O custo imediato aumenta à medida que diminui a utilização de recursos hidráulicos. Já a Função de Custo Futuro (FCF) está relacionada ao custo esperado de geração térmica e déficit do momento t até o final do período sob análise. A FCF diminui à medida que aumenta o volume armazenado final.

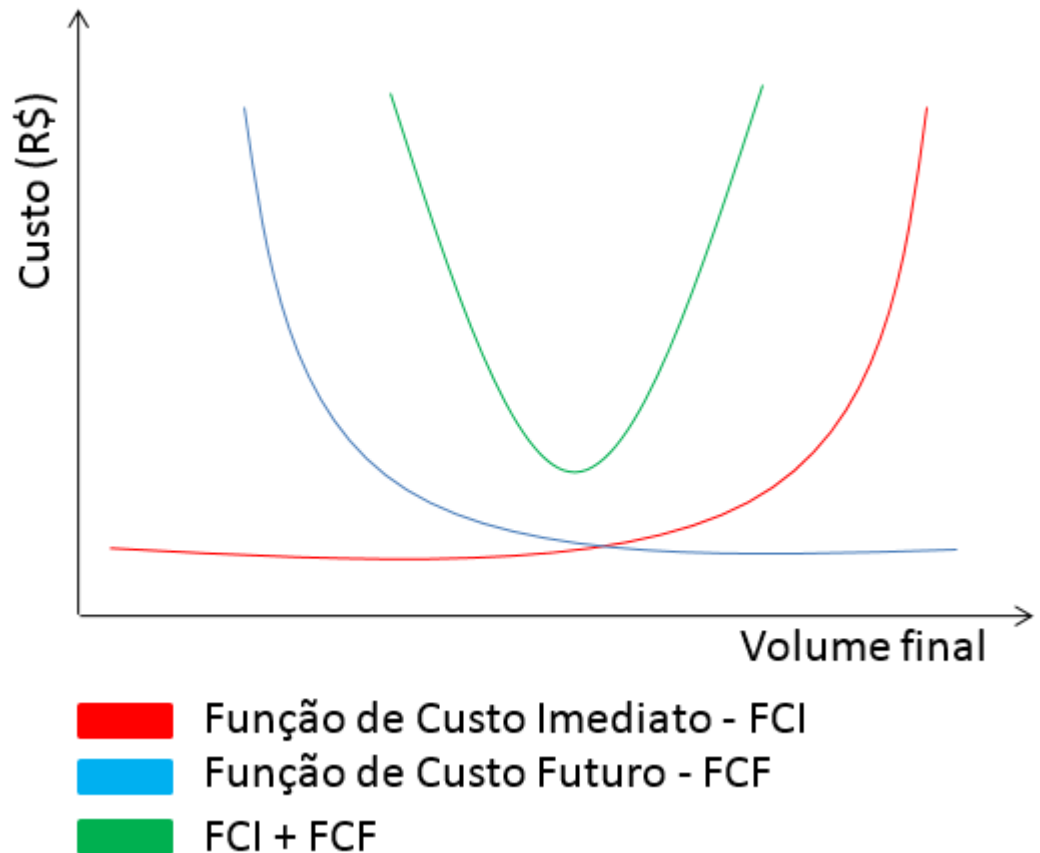


Figura 2.5 - Curvas de custo do uso da água. Fonte Silva (2001).

A curva de FCF pode ser estimada através de simulações da operação do sistema para cada nível de armazenamento no final do momento t . A modelagem da função de custo futuro é desenvolvida por um procedimento recursivo denominado programação dinâmica dual estocástica.

A otimização do uso da água se dá no ponto que minimiza a soma dos custos imediato e futuro (FCI+FCF). Diferentemente dos sistemas termelétricos, onde a produção de energia possui um custo operativo direto, a produção das hidrelétricas têm um valor indireto, relacionado à oportunidade de economizar combustível deslocando uma térmica hoje ou no futuro.

No modelo de despacho brasileiro há uma incorporação apropriada do valor implícito da água armazenada no momento do despacho, possibilitando uma otimização dos recursos do sistema, isto é, despacha-se o sistema ao mínimo custo. O preço da energia é determinado então pelo recurso de maior custo que foi utilizado e, considerando-se a forte predominância hidráulica do sistema brasileiro, o preço da energia é normalmente determinado pelo valor da água e eventualmente pelo preço de uma termelétrica ou pelos custos de déficit.

2.5. O cálculo do Preço de Liquidação de Diferenças – PLD

Segundo Silva (2001), os modelos matemáticos usados para calcular o despacho ótimo baseiam-se nas previsões hidrológicas, na demanda futura, nos preços dos combustíveis, no custo do déficit, na entrada em operação de novas plantas e na disponibilidade dos sistemas de geração e transmissão. Tem-se então o despacho ótimo para horizonte estudado para ambos os tipos de geração, hidráulica e térmica, e para cada submercado (norte, sul, nordeste e centro-oeste/sudeste). Como resultados são obtidos os Custos Marginais de Operação (CMO) para o período sob análise, para cada patamar de carga e para cada submercado. O valor do preço de liquidação de diferenças (PLD) é originado do custo marginal de operação e atribuído em bases semanais, porém limitado por um valor mínimo e máximo definido pela ANEEL.

O mesmo processamento realizado pelo ONS é utilizado pela CCEE, porém com as restrições internas aos submercados suprimidas, de modo que a energia comercializada seja considerada igualmente disponível em todos os pontos, e conseqüentemente seu preço uniforme em cada submercado (as restrições entre submercados são mantidas). O despacho econômico tem propósito distinto do despacho elétrico, mas ambos são gerados pelo ONS e fornecem os montantes que cada usina deveria gerar (com e sem restrições), sendo possível então que o despacho real seja diferente daquele previsto. Essas diferenças, por não estarem contempladas no cálculo do PLD pela CCEE, são pagas às usinas através dos Encargos de Serviços do Sistema.

Para Melo *et al* (2011), sob a abordagem de despacho centralizado, o mecanismo de preços do mercado de curto prazo considera os seguintes aspectos:

- ❖ Hidroelétricas e termoelétricas apresentam dados técnicos sobre a sua instalação, incluindo níveis de água nos reservatórios, taxa de entrada, a eficiência térmica e de combustível, dados de custo operacional e da disponibilidade técnica de suas turbinas;
- ❖ Agentes representantes dos consumidores enviam os dados de demanda;
- ❖ Com base nesses dados, juntamente com a previsão da entrada de água nos reservatórios e com base em cenários derivados de dados de oferta e demanda e projeções históricas, o ONS faz o envio em tempo real, baseado em custo centralizado usando os modelos de otimização (NEWAVE / DECOMP). O modelo NEWAVE calcula o despacho hidrotérmico em uma base mensal, enquanto DECOMP estabelece o despacho hidrotérmico em uma base semanal;
- ❖ Além da programação de envio para as plantas, este modelo estabelece os custos marginais (implícitos) para cada submercado, por período de programação;
- ❖ A CCEE calcula o preço de curto prazo (PLD) com base no mesmo modelo de otimização utilizado pelo ONS. Este preço é calculado e divulgado semanalmente pela CCEE, tendo por base o custo marginal operacional do sistema com limites inferior e superior de preços, em três níveis de carga diferentes ("alto", "médio", "baixo") e para quatro submercados diferentes (Sul, Sudeste, Norte, Nordeste).

Os sistemas com geração de predominância hidráulica com grandes reservatórios, como é o caso do Brasil, são projetados para assegurar o fornecimento de energia em condições hidrológicas adversas, que não ocorrem com frequência. Com isso, na maioria do tempo há energia armazenada em excesso (água armazenada nos reservatórios), resultando em PLDs muito baixos. Por outro lado, quando um período de seca ocorre, os PLDs se tornam bastante elevados. A figura 2.6 ilustra esta situação para o mercado Sudeste/Centro-Oeste.

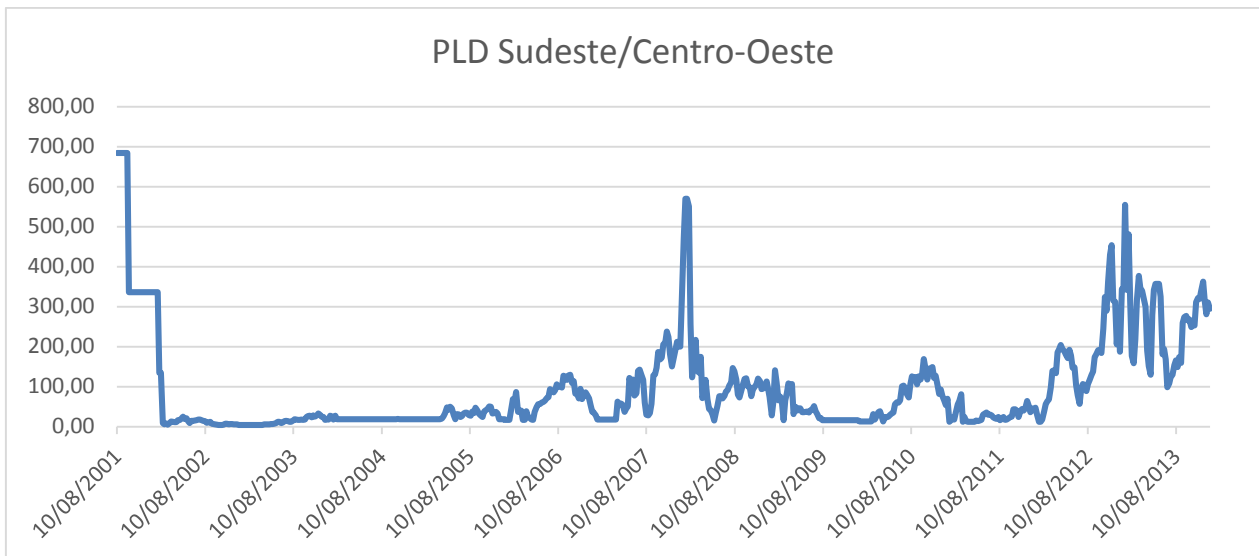


Figura 2.6 – Comportamento do PLD para o submercado SE/CO. Fonte: site da CCEE – www.ccee.org.br

2.6. Processo de comercialização de energia

Segundo Souza (2008), no processo de comercialização de energia, a CCEE contabiliza as diferenças entre o que foi produzido ou consumido e o que foi contratado. As diferenças positivas ou negativas são liquidadas no mercado de curto prazo e valoradas ao PLD. Para Da Silva et al (2012), como o Brasil não possui ainda um mercado pool, de curto prazo, no qual os preços spot sejam calculados hora-a-hora, o cálculo do PLD é semanal, para três patamares de carga, utilizando os softwares NEWAVE e DECOMP para cada submercado (Norte, Nordeste, Sudeste/Centro-Oeste e Sul), tendo como base o custo marginal de operação do sistema, sendo este limitado por um preço mínimo e por um preço máximo estipulado pela ANEEL.

Tanto os Contratos de Comercialização de Energia Elétrica no Ambiente Regulado (CCEAR) quanto os contratos bilaterais do ambiente de comercialização livre são registrados na CCEE e servem de base para a contabilização e liquidação de diferenças no mercado de curto prazo. Dessa forma, pode-se dizer que o mercado de curto prazo é o mercado das diferenças entre montantes contratados e montantes medidos, conforme ilustra a Figura 2.7.

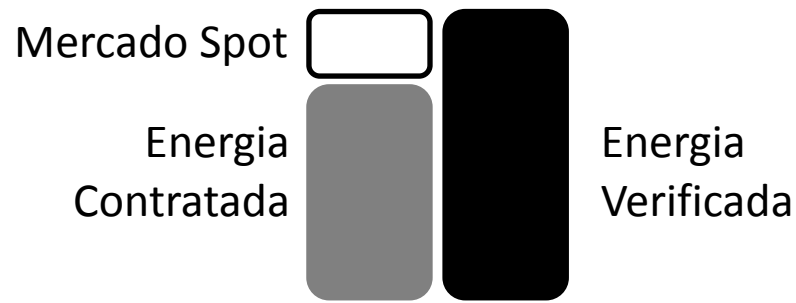


Figura 2.7 – Processo de comercialização de energia no curto prazo. Fonte: site da CCEE – www.ccee.org.br

2.7. Ambientes de contratação

A comercialização de energia elétrica é realizada em dois ambientes de mercado, o Ambiente de Contratação Regulada (ACR) e o Ambiente de Contratação Livre (ACL), do qual participam agentes de geração, comercializadores, importadores e exportadores de energia e consumidores livres, conforme ilustra Figura 2.8.

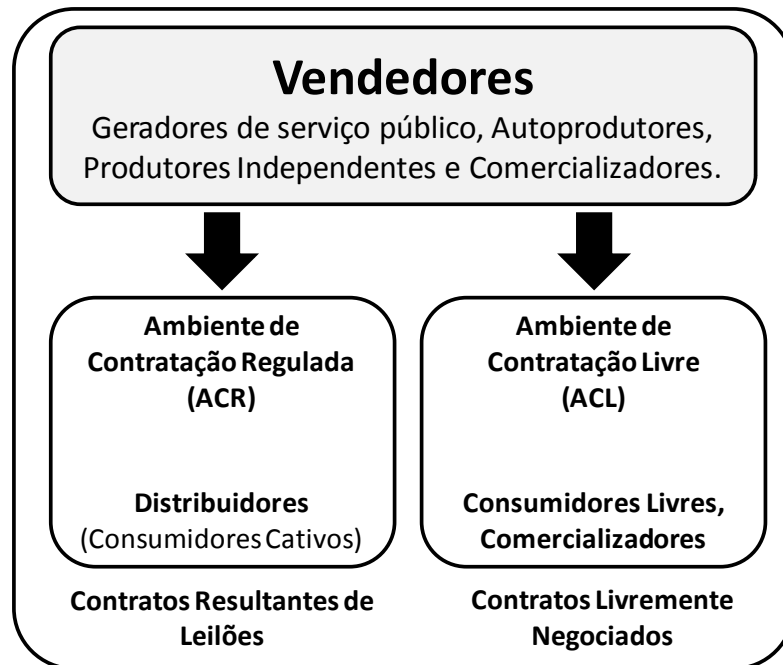


Figura 2.8 - Ambientes de contratação. Fonte: site da CCEE – www.ccee.org.br

A contratação no ACR é formalizada através de contratos bilaterais regulados, denominados Contratos de Comercialização de Energia Elétrica no Ambiente Regulado (CCEAR), celebrados entre Agentes Vendedores (comercializadores, geradores, produtores independentes ou

autoprodutores) e Compradores (distribuidores) que participam dos leilões de compra e venda de energia elétrica.

Segundo Barroso et al (2011), embora os contratos sejam instrumentos financeiros, estes devem ser ancorados por uma capacidade física de geração de energia. Os contratos de energia são mecanismos essenciais para a viabilidade comercial de geradores visto que os preços spot são utilizados para resolver os desequilíbrios energéticos (diferenças entre os volumes produzidos e contratados) e são muito voláteis.

Os Agentes de Geração, sejam concessionários de serviço público de Geração, produtores Independentes de energia ou auto-produtores, assim como os comercializadores, podem vender energia elétrica nos dois ambientes, mantendo o caráter competitivo da geração, e todos os contratos, sejam do ACR ou do ACL, são registrados na CCEE e servem de base para a contabilização e liquidação das diferenças no mercado de curto prazo.

Participam do ACR os agentes vendedores e agentes de distribuição de energia elétrica. Para garantir o atendimento aos seus mercados, os agentes de distribuição podem adquirir energia das seguintes formas, de acordo com o art. 13 do Decreto nº 5.163/2004:

- ❖ Leilões de compra de energia elétrica proveniente de empreendimentos de geração existentes e de novos empreendimentos de geração;
- ❖ Geração distribuída, desde que a contratação seja precedida de chamada pública realizada pelo próprio Agente de Distribuição e com montante limitado a 10% do mercado do distribuidor;
- ❖ Usinas que produzem energia elétrica a partir de fontes eólicas, pequenas centrais hidrelétricas e biomassa, contratadas na primeira etapa do Programa de Incentivo às Fontes Alternativas de Energia Elétrica – PROINFA;
- ❖ Itaipu Binacional.

Além disso, conforme descrito no mesmo artigo do Decreto nº 5.163/2004, os contratos firmados pelos Agentes de Distribuição até 16/03/2004 também são considerados como energia contratada para atendimento à totalidade de seus respectivos mercados.

No Ambiente de Contratação Livre participam agentes de geração, comercializadores, importadores e exportadores de energia elétrica e consumidores livres. Nesse ambiente há liberdade para se estabelecer volumes de compra e venda de energia e seus respectivos preços, sendo as transações pactuadas através de contratos bilaterais.

Segundo Torres (2006), em um portfólio de contratos de energia elétrica administrado por um comercializador, normalmente os contratos bilaterais de compra possuem características diferentes dos contratos bilaterais de venda. Isto envolve o montante de energia contratada, o período de suprimento, o preço em R\$/MWh, o submercado de entrega, e algumas flexibilidades que podem ser incorporadas, como por exemplo, a sazonalização e modulação.

Por sazonalização, entende-se a distribuição da energia ao longo do ano, que é, na prática, a distribuição mensal do montante anual contratado. Já modulação é a distribuição horária dos montantes mensais definidos. Tais flexibilidades permitem atender a um maior consumo em um determinado período do ano ou em determinadas horas do dia.

Ainda segundo Torres (2006), outra flexibilidade que pode ser incorporada aos contratos bilaterais de compra é a possibilidade de aumento ou redução dos montantes contratos. Esta flexibilidade equivale a uma opção embutida no contrato, onde o aumento do montante contratado equivale a uma opção de compra e a redução do montante contratado equivale a uma opção de venda. O preço de exercício desta opção, e o próprio valor do MWh é estabelecido em contrato.

Os contratos bilaterais de venda negociados diretamente com os consumidores livres possuem características semelhantes aos contratos de compra, mas com a incorporação de alguns detalhes próprios do consumo industrial, como é o caso das paradas programadas e a variação do montante contratado dentro de uma faixa, entre um limite superior e um limite inferior. Esta variação atrelada ao consumo é denominada *Take or Pay*.

2.8 Considerações Finais

Neste capítulo foi apresentado um panorama geral da comercialização de energia no mercado brasileiro, abordando o processo de desverticalização da indústria de energia e o novo marco regulatório brasileiro. Também foi apresentado o Sistema Interligado Nacional – SIN e o

modelo de despacho centralizado *tight pool* utilizado no Brasil, além do processo de comercialização de energia e o modelo de formação do Preço de Liquidação de Diferenças.

No capítulo a seguir serão apresentadas as técnicas de predição de séries temporais com detalhamento dos métodos que serão utilizadas para a estimação futura do PLD e explicitação do conhecimento do processo para a formação do PLD.

CAPÍTULO III

MÉTODOS PARA A PREDIÇÃO E CLASSIFICAÇÃO DE SÉRIES TEMPORAIS

3.1 Introdução

Séries temporais são estudadas a bastante tempo pelo meio científico (Levy, 1969; Nelson, 1973; Granger e Newbold, 1977; White, 1988; Lo e Mackinley, 1988; Kimoto *et al*, 1990; Refenes *et al*, 1997) visto que as mesmas representam os mais variados fenômenos, tais como, preços de ações do mercado, valor trimestral do PIB, valor do dólar, números de casos mensais de AIDS, temperatura média de uma região, etc.

Na análise de séries temporais um dos procedimentos básicos é a sua predição. A realização da predição de séries temporais é feita por diversas técnicas, metodologias, algoritmos computacionais ou simplesmente modelos. Aggarwal et al (2009b) propõe, conforme a Figura 3.1 uma classificação para as metodologias de predição de preços.

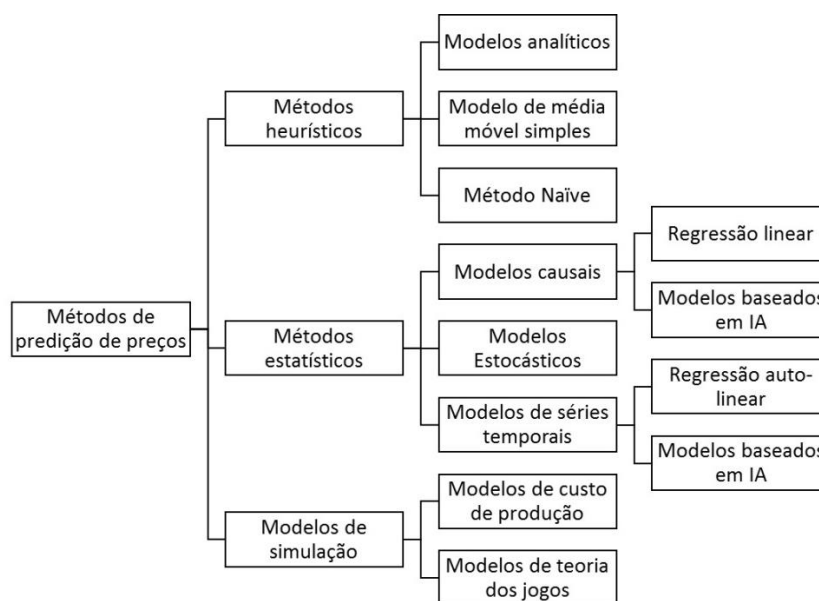


Figura 3.1 - Classificação de metodologias de predição de preços

Os métodos heurísticos são simples de implementar e pela rapidez com que são modelados, se tornam excelentes parâmetros de comparação para avaliar desempenho e exatidão de qualquer outro modelo de previsão. Os modelos analíticos baseiam-se na relação matemática exata entre um valor desconhecido, normalmente o preço, e outro valor de variável conhecido, comumente a carga de energia, em algum cenário. Pressupõe-se aqui que a variação de preço é linearmente relacionada com a demanda de energia. Os modelos de média móvel (MA) simples estimam o preço com base na média simples dos valores dos preços no passado. Os modelos Naïve circunscrevem o valor do preço a algum valor passado, mantendo a periodicidade das séries temporais de preços. Um modelo analítico é apresentado em Weron (2006) e em Su e Kirschen (2009). São exemplos de modelos Naïve os trabalhos de Zhao et al (2007a) e Lu et al (2005).

Os métodos de previsão de preços por simulação, imitam a expedição real do sistema com seus requisitos operacionais e restrições. Os *Production Cost Models* - PCM são projetados para calcular os custos de produção de um sistema de geração considerando o padrão de carga esperado, as curvas de taxa de calor dos diferentes geradores, os custos de combustível, o despacho econômico e as agendas de compromissos de cada unidade. Os modelos baseados na teoria dos jogos possuem os refinamentos dos PCM e acrescentam a complexidade dos fatores econômicos, as estratégias dos participantes do mercado e a elasticidade da demanda. São exemplos de modelos PCM os contidos nos trabalhos de Maenhoudt e Deconinck (2014) e Maenhoudt e Deconinck (2012). São exemplos de modelos baseados em teoria dos jogos os trabalhos apresentados por Molina et al (2010) e Yang et al (2013).

Os modelos estatísticos são modelos de função de aproximação na qual a variável prevista é definida em termos de um conjunto de equações comportamentais que envolvem algumas variáveis observadas e um termo de erro. Em modelos de séries temporais, o preço é modelado como uma função de seus valores observados no passado. Em modelos causais, o preço é modelado como uma função de alguns fatores exógenos, especialmente da carga, dos preços dos combustíveis e outros. Os modelos estocásticos são derivados dos modelos de precificação de ativos financeiros, que são usados para a previsão do preço de ações e adaptados à dinâmica dos preços da eletricidade. Modelos lineares estatísticos baseados em filtros ARIMA e GARCH são propostos em Contreras et al (2003), Garcia et al (2005) e Nogales e Conejo (2006). Modelos baseados em inteligência artificial são apresentados em Chen et al (2012) e Zhao et al (2007b).

Entre as diferentes abordagens dos sistemas de classificação, podemos citar as árvores de decisão, redes neurais, modelos Naïve, classificadores bayesianos, K-Nearest Neighbors (KNN) e máquinas de suporte vetorial (SVM). Apenas alguns trabalhos têm proposto o uso de técnicas de classificação para os preços da energia elétrica. Em Zhao et al (2007), é proposto um classificador para prever picos de preços no mercado da Austrália usando a máquina de vetor de suporte. Já Zareipour et al (2011) propõe uma classificação dos preços futuros do mercado de energia elétrica através de máquinas de vetores de suporte com dados de sistemas de energia elétrica de Ontário e Alberta. Em Anbazhagan e Kumarappan (2012) é proposta uma abordagem de rede neural para classificar os preços da eletricidade para os mercados da Espanha continental e Nova Iorque. O trabalho de Huang et al (2012) realiza a classificação de preços e investiga a efetividade de três técnicas: árvores de decisão, redes neurais MLP, classificador Naïve, e KNN. Os resultados são obtidos usando dados de Nova York, Ontário e mercados de Alberta na Austrália

Neste trabalho, o foco está limitado aos modelos preditores de séries temporais e modelos causais e ao modelo classificador de preços futuros de energia elétrica, em especial com o uso de filtros ARIMA, redes neurais e árvores de decisão.

3.2. Pré-processamento

A etapa de pré-processamento na predição ou classificação de séries temporais, compreende a aplicação de várias técnicas para captação, organização, tratamento e a preparação dos dados. É uma etapa que possui fundamental relevância porque vai desde a correção de dados errados até o ajuste da formatação dos dados para os algoritmos de predição e classificação que serão utilizados.

3.2.1. Seleção das variáveis explanatórias

O procedimento de seleção de variáveis explanatórias é utilizado para reduzir a dimensionalidade do espaço de entrada antes do processo de predição. O uso de um menor número de preditores gera economia de tempo e recursos computacionais.

Para Jeong et al (2012), o processo de seleção de variáveis explanatórias é um importante método de processamento de dados, usado para remover entradas irrelevantes ou redundantes, que não fornecem informações adicionais para o modelo em desenvolvimento.

Existem dois tipos de técnicas de seleção de variáveis explanatórias: ranqueamento de variáveis e subconjunto de variáveis. O ranqueamento de variáveis avalia o poder preditivo de

uma variável individualmente com base em uma função critério. Neste caso não é necessário um algoritmo de busca de um subconjunto de variáveis, pois o objetivo se limita a determinar um ranque das variáveis individualmente. Um exemplo muito comum de função critério é a correlação, como apresentado em Guyon e Elisseeff (2003). Em contraste, o subconjunto de variáveis é um método que exige tanto uma função critério quanto um algoritmo de busca. Existem três categorias de técnicas recurso subconjunto de variáveis: Wrappers, filtros e métodos embarcados. Maiores detalhes sobre os métodos de subconjunto de variáveis são encontrados em Kohavi e John (1997).

O método utilizado neste trabalho é de ranqueamento de variáveis explanatórias a partir do teste F ANOVA (Análise de Variância). Segundo Wilcox (2003), o método ANOVA é um teste estatístico para comparar as médias de duas ou mais variáveis aleatórias independentes que possuam distribuição normal, e pode ser utilizado como técnica de seleção de variáveis explanatórias. O teste F determina se as diferenças entre dois ou mais conjuntos de dados são estatisticamente significativos.

O algoritmo ANOVA considera uma variável de entrada por vez e observa de que maneira uma entrada é capaz de prever o valor da saída de maneira isolada. A importância de cada preditor é calculada como sendo $(1-p)$, onde p é o valor do teste apropriado entre a variável candidata a preditor e a variável de saída.

O método ANOVA separa a variação do conjunto de dados em duas partes: entre-grupos e dentro do grupo. A proporção F é calculada dividindo a variação entre grupos pela variação dentro do grupo.

Para se obter o valor-p, este deve ser testado contra uma distribuição F de uma variável aleatória com graus de liberdade associados ao numerador e denominador da proporção F. Valores grandes de F geram pequenos valores de p. Um valor-p estatisticamente significativo (usualmente usa-se $p < 0.05$) sugere que pelo menos a média de um grupo é significativamente diferente das demais.

O objetivo do método ANOVA é testar a hipótese das médias iguais:

$$H_0: \mu_1 = \mu_2 \quad (3.1)$$

Etapas do método:

- A) Calcule as médias \bar{X}_1 e \bar{X}_2 e as variâncias S_1^2 e S_2^2
- B) Calcule a grande média

$$\bar{X}_G = \frac{1}{2}(\bar{X}_1 + \bar{X}_2) \quad (3.2)$$

- C) Calcule as somas quadradas.

C1) Entre grupos:

$$SQEG = n \sum_{j=1}^2 (\bar{X}_j - \bar{X}_G)^2 \quad (3.3)$$

C2) Variação dentro do grupo:

$$SQDG = (n - 1) \sum_{j=1}^2 s_j^2 \quad (3.4)$$

D) Compute as médias

D1) Entre grupos

$$MEG = \frac{SQEG}{2-1} = SQEG \quad (3.5)$$

D2) Dentro do grupo

$$MDG = \frac{SQDG}{2n-2} \quad (3.6)$$

E) Compute a estatística F

$$F = \frac{MEG}{MDG} \quad (3.7)$$

Regra de decisão: Rejeitar a H_0 se $F \geq f$, onde f é o quantil de uma distribuição F com $v_1 = 2-1 = 1$ e $v_2 = 2n-2$ graus de liberdade.

3.2.2 Redução de dimensionalidade

Segundo Tan *et al.* (2009), conjuntos de amostras podem ter um grande número de características. Muitos algoritmos funcionam melhor se a dimensionalidade (o número de atributos nos dados) for menor. Isto ocorre em parte porque a redução da dimensionalidade pode eliminar características irrelevantes e reduzir o ruído. A redução da dimensionalidade pode levar a um modelo mais compreensível, por envolver menos atributos, e onde a quantidade de tempo e memória requerida pelo algoritmo de mineração de dados é reduzida.

Neste trabalho optou-se por utilizar a Análise dos Componentes Principais (PCA – *Principal Component Analysis*). Da perspectiva de reconhecimento estatístico de padrões, o valor prático do PCA é que ela fornece uma técnica efetiva para redução de dimensionalidade. Em particular, pode-se reduzir o número de características necessárias para a representação efetiva de dados, descartando aquelas combinações lineares que têm variâncias pequenas e retendo apenas aqueles termos que têm variâncias grandes (MANLY, 2008) e (HAYKIN, 2007).

A PCA é uma técnica da álgebra linear para atributos contínuos que encontra novos atributos (componentes principais) que: (1) sejam combinações lineares dos atributos

originais, (2) sejam ortogonais (perpendiculares) entre si e (3) capturem a quantidade máxima de variações nos dados.

O PCA usa um conjunto de dados representado por uma matriz de n registros por k atributos, que podem estar correlacionados, e sumariza esse conjunto por eixos não correlacionados (componentes principais) que são uma combinação linear das k variáveis originais. Objetos são representados por uma nuvem de n pontos em um espaço multidimensional, com um eixo para cada uma dos k atributos. O centróide dos pontos é definido pela média de cada atributo e a variância de cada atributo é média dos quadrados da diferença dos n pontos com relação a média de cada atributo. A equação 3.8 apresenta como são calculadas as variâncias de cada atributo:

$$V_i = \frac{1}{n-1} \sum_{m=1}^n (X_{im} - \bar{X}_i)^2 \quad (3.8)$$

O grau com que cada variável é linearmente correlacionado é representado pela sua covariância. A equação 3.9 apresenta como é calculada a covariância entre duas variáveis.

$$C_{ij} = \frac{1}{n-1} \sum_{m=1}^n (X_{im} - \bar{X}_i)(X_{jm} - \bar{X}_j) \quad (3.9)$$

O objetivo da PCA é rotacionar rigidamente os eixos desse espaço p -dimensional para novas posições (eixos principais) que tem a seguinte propriedade:

- ❖ Ordenado de tal maneira que o eixo principal 1 tem a maior variância, o eixo 2 tem a próxima maior variância, até o último eixo de menor variância.
- ❖ Covariância entre cada par de eixos é zero (os eixos principais não são correlacionados).

As componentes principais são variáveis aleatórias que não podem ser medidas diretamente, mas observadas apenas a partir da informação do vetor aleatório x . É comum utilizar os escores das componentes para condução de análise estatística de dados ou para a simples ordenação (ranking) dos elementos amostrais observados, com o intuito de identificar aqueles que estão com maiores, ou menores, valores globais das componentes.

Uma aplicação útil do PCA é como pré-processador para uma rede neural supervisionada (HAYKIN, 2007). Busca-se com o pré-processamento a convergência do processo de aprendizagem retirando a correlação dos dados de entrada. Um processo de convergência baseado no algoritmo de retropropagação do erro é tipicamente lento devido a efeitos interativos de pesos sinápticos de uma rede neural sobre o sinal de erro, mesmo com o

uso de procedimentos simples de aceleração local como o momento e taxas de aprendizagens adaptativas para pesos individuais. Se, entretanto, as entradas da rede neural consistirem de componentes não-correlacionadas, acelera-se o processo de convergência da rede neural.

3.2.3. Balanceamento de amostras

Para Haibo He e Garcia (2009), considera-se uma base de dados desbalanceada, aquela onde há distribuição desigual, significativa, entre as classes de valores. Ainda segundo os autores, duas são as estratégias de reamostragem utilizadas para balanceamento de dados: supra-amostragem de valores majoritários e super-amostragem de valores minoritários.

Neste trabalho é utilizada a estratégia de super-amostragem de valores minoritários. Nesta estratégia, um conjunto de dados tendencioso tem o balanceamento realizado através da duplicação e, em seguida, descarte de registros com base nas condições especificadas.

Para Haibo He et al (2008), usa-se o balanceamento para corrigir os desequilíbrios em conjuntos de dados. Por exemplo, suponha-se que um conjunto de dados tem apenas dois valores de saída (baixo ou elevado) e que 90% dos casos são baixos, enquanto apenas 10% dos casos são elevados. Muitas técnicas de modelagem têm problemas com esses dados tendenciosos, porque os modelos tendem a acertar bastante nos casos onde a saída for baixo e errar muito quando a saída for de valor elevado. Se os dados estão bem equilibrados com aproximadamente o mesmo número de resultados de baixo e elevado, os modelos terão uma melhor chance de encontrar padrões que distinguem os dois grupos. Neste caso, é útil uma estratégia de balanceamento que reduz os casos com resultado baixo.

Como exemplo apresenta-se na Figura 3.2 que ilustra o histograma de amostras da energia armazenada. Analisando o diagrama, há valores que se repetem 24 vezes no espaço amostral, mas há outros valores que só ocorrem duas vezes. A estratégia de balanceamento se baseia em dividir as amostras de maneira proporcional nas partições de treino e teste. Assim, a rede é treinada para uma maior gama de valores e não apenas para um percentual de amostras iniciais, correspondentes a partição de treino, que nem sempre contemplam toda a variedade de valores.

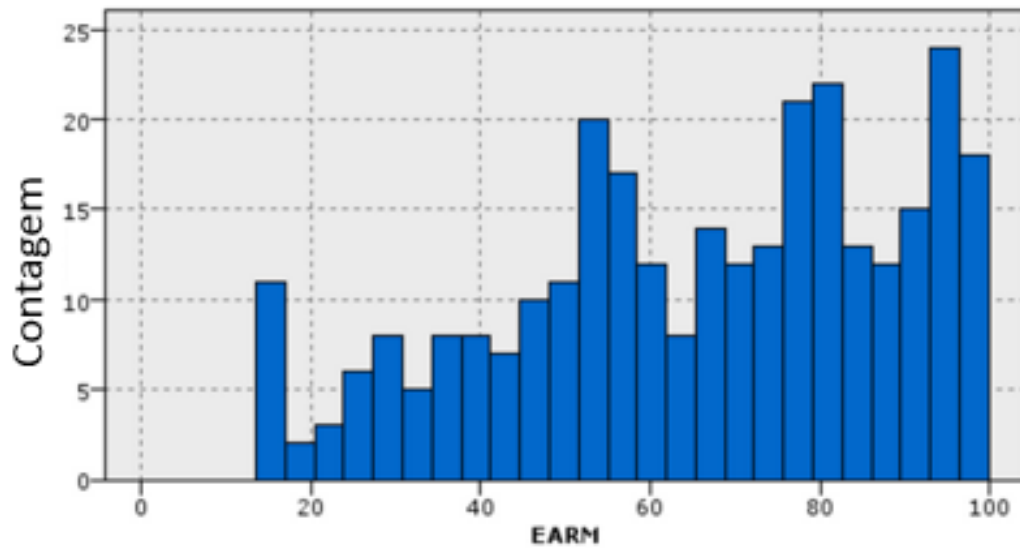


Figura 3.2- Histograma de energia armazenada

3.3. Metodologia Box-Jenkins

Segundo Makridakis *et al.* (1998), o modelo autoregressivo integrado de médias móveis (*Autoregressive Integrated Moving Average – ARIMA*) tem sido estudado extensivamente. Os modelos foram popularizados por George Box e Gwilym Jenkins no início dos anos 70 e os nomes dos autores têm sido usados como sinônimo dos modelos ARIMA e utilizados na análise e previsão de séries temporais. Box e Jenkins descreveram de maneira compreensiva as relevantes informações necessárias para compreender e utilizar os modelos ARIMA.

A base da abordagem de Box-Jenkins é composta de três fases: I) identificação, II) estimação/testes e III) aplicação. O algoritmo é sintetizado na figura 3.3.

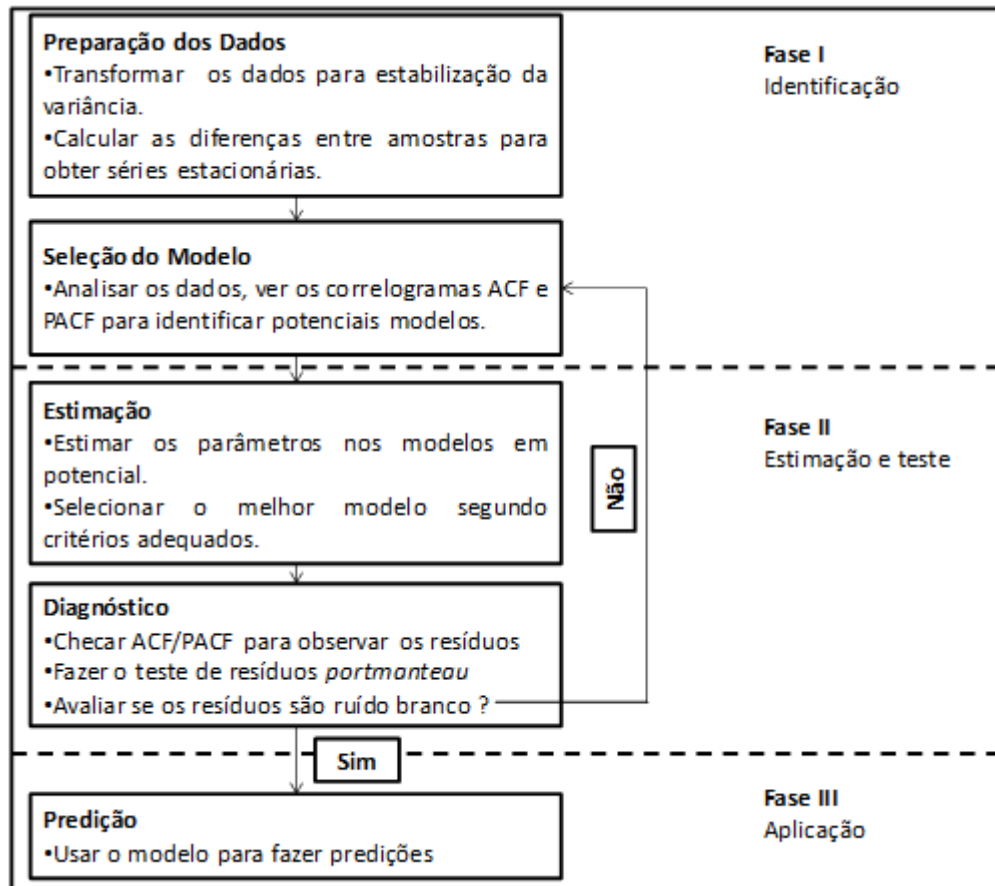


Figura 3.3 - Metodologia Box-Jenkins para a modelagem de séries temporais. Fonte Makridakis *et al.* (1998)

A primeira etapa da metodologia é a fase de identificação, onde são pré-processados os dados e, a partir das funções de autocorrelação (ACF) e autocorrelação parcial (PACF) define-se uma classe geral de modelos para análise. A seguir é realizada a etapa de estimação dos parâmetros do modelo identificado. Através de uma série de testes, sendo o principal a análise dos resíduos (erros de predição), ajusta-se o modelo. Se o modelo não for satisfatório, o ciclo é repetido, voltando-se à fase de identificação. A primeira fase é a mais crítica, visto que é possível que vários modelos diferentes se adaptem bem a uma determinada série temporal. Nestes casos, deve-se utilizar o modelo mais simples, com menos parâmetros.

A metodologia de Box-Jenkins é baseada na ACF e a PACF. A ACF é obtida através da correlação linear de cada valor $x(n)$ da série de dados com outros valores em atrasos distintos, como $x(n-1)$, $x(n-2)$ e assim por diante. Já a PACF é a correlação entre, por exemplo, $x(n)$ e $x(n-3)$, mas sem levar em consideração os efeitos causados por $x(n-1)$ e $x(n-2)$. Na ACF, tais efeitos são considerados. Maiores detalhes sobre a influência das funções ACF e PACF na determinação do modelo a ser utilizado podem ser obtidas a partir da leitura

de (Makridakis, Wheelwright e Hyndman 1998), (Brockwell e Davis 2002) e (Morettin e Toloi 2006).

Para Morettin e Toloi (2006), dentre os processos estocásticos importantes da metodologia de Box-Jenkins, destaca-se os modelos autoregressivos (AR), médias móveis (MA) e combinações destes, tais como os modelos ARMA, ARIMA e SARIMA. Todos eles são paramétricos, ou seja, possuem um número finito de parâmetros cujos valores são estimados a partir do sinal ou série temporal sob análise.

3.3.1 Modelos Autoregressivos (AR)

Os modelos autoregressivos de ordem p , $AR(p)$, são os modelos de Box-Jenkins mais simples, em que se escreve o valor atual da variável aleatória $x(n)$ como uma soma ponderada de seus valores passados mais o ruído branco gaussiano.

$$\begin{aligned} x(n) &= \phi_0 + \phi_1 x(n-1) + \phi_2 x(n-2) + \dots + \phi_p x(n-p) + a(n) \quad (3.10) \\ &= \phi_0 + \sum_{i=1}^p \phi_i x(n-i) + a(n) \end{aligned}$$

Em que $\phi_i, i = 0, \dots, p$, são os coeficientes do modelo, que juntamente com a ordem da memória p , constituem os parâmetros do modelo. Na equação 3.11 a sequência $\{a(n), n \geq 0\}$ de ruído branco gaussiano tem média nula e variância $\sigma_a^2 \neq 0$.

Na forma preditiva, o modelo AR pode ser escrito da seguinte maneira:

$$x(n+1) = \phi_0 + \phi_1 x(n) + \phi_2 x(n-1) + \dots + \phi_p x(n-p+1) + a(n) \quad (3.11)$$

Na equação 3.11 valem todas as definições da equação 3.10.

Para Aguirre (2000), independentemente da formulação escolhida, existem várias técnicas para calcular os coeficientes de um modelo AR, sendo a mais comum a dos Mínimos Quadrados (MQ), que é equivalente ao método de estimação por máxima verossimilhança (*maximum likelihood*) quando o ruído é gaussiano.

3.3.2 Modelos de médias móveis (MA)

O modelo $MA(q)$, é definido pelo comportamento dos erros no passado e é representado pela equação 3.12. Segundo Morettin e Toloí (2006), os modelos $MA(q)$ são sempre estacionários. Modelos de médias móveis de ordem q , denotados $MA(q)$, são descritos como uma combinação linear finita de q valores passados da seqüência de ruído branco:

$$x(n) = a(n) + \theta_1 a(n-1) + \theta_2 a(n-2) + \dots + \theta_q a(n-q) \quad (3.12)$$

em que θ_i são os coeficientes do modelo, que juntamente com sua ordem q , constituem os parâmetros do modelo. Estes modelos são mais difíceis de aplicar que modelos $AR(p)$ e o cálculo de seus coeficientes, a partir dos dados observados, é geralmente feito através do método de máxima verossimilhança. Em geral, modelos $MA(q)$ são usados em conjunção com modelos $AR(p)$, afim de reduzir o número de parâmetros deste último.

3.3.3 Modelos autorregressivos de médias móveis (ARMA)

Muitas séries temporais podem ser descritas com um número menor de parâmetros através de um modelo $ARMA(p,q)$, quando comparados ao uso de um modelo $AR(p)$ ajustado à mesma série. Segundo Morettin e Toloí (2006), o uso combinado de termos autoregressivos e de médias móveis é a solução adequada. A equação 3.13 apresenta o modelo.

$$x(n) = \phi_0 + \phi_1 x(n-1) + \phi_2 x(n-2) + \dots + \phi_p x(n-p) + a(n) \quad (3.13) \\ + \theta_1 a(n-1) + \theta_2 a(n-2) + \dots + \theta_q a(n-q)$$

em que ϕ_i e θ_i são os coeficientes autoregressivos e de médias móveis do modelo, que juntamente com as ordens p e q , constituem os parâmetros do mesmo. O modelo é especialmente útil em séries já estacionárias.

3.3.4 Modelos autorregressivos integrados de médias móveis (ARIMA)

O processo estocástico ARIMA é a combinação de modelos $AR(p)$ e modelos $MA(q)$, onde são modeladas as partes autorregressiva e de médias móveis. Além disso, é integrada a combinação com d diferenças, para que a série temporal se torne estacionária.

A estacionariedade de uma série temporal pode ser obtida através de transformações atuando sobre a série temporal original. Uma forma simples de tornar séries não-estacionárias em séries estacionárias é através de diferenças entre seus valores consecutivos.

Como exemplo, parte-se de uma dada série $\{x(n)\}_{n=1}^N$ não estacionária para uma nova série $\{w(n)\}_{n=1}^{N-1}$ obtida por meio da seguinte operação:

$$w(n) = \Delta x(n) = x(n) - x(n - 1) \quad (3.14)$$

Se mesmo após o procedimento da equação 3.14 não se obtiver uma série estacionária, o mesmo procedimento pode ser repetido sobre as amostras $w(n)$ até que uma série estocástica seja estacionária o suficiente para permitir que um modelo linear de Box-Jenkins possa ser ajustado a ela.

Para Morettin e Tolo (2006) uma série temporal $\{x(n)\}_{n=1}^N$ que, tomando-se um número finito de diferenças entre amostras sucessivas torna-se estacionária, é chamada não-estacionária homogênea. Como o processo é reversível, a série não-estacionária original $\{x(n)\}_{n=1}^N$ pode ser obtida a partir da série estacionária omitida pela soma (ou integração) de amostras sucessivas, daí este modelo ser chamado de Auto-regressivo Integrado de Médias Móveis de ordens p , d e q , ou simplesmente $ARIMA(p,d,q)$. Destaca-se também a existência de uma variação do processo estocástico ARIMA capaz de captar a sazonalidade de uma série temporal, denominada SARIMA.

Muitas séries temporais contêm uma componente periódica sazonal que se repete a cada s observações ($s > 1$). Por exemplo, com dados mensais $s = 12$, dados semanais $s = 52$ e dados diários $s = 365$. Para tais séries temporais, o uso do modelo $SARIMA(p,d,q)(P,D,Q)_s$ é o mais adequado. Esses modelos contêm uma parte não sazonal com parâmetros (p,d,q) , e uma outra sazonal com parâmetros (P,D,Q) . O valor de s se refere ao número de observações que descreve o padrão de sazonalidade. Quando se está modelando uma série que apresenta sazonalidade, o objetivo é estimar a componente sazonal e, na sequência, subtrair tal parte do modelo.

Há também os casos onde as saídas do modelo ARIMA não dependem apenas das amostras passadas da série temporal sob análise. Nestes casos usam-se os valores atuais e atrasados no tempo de séries temporais pertinentes. Estes modelos não representam uma nova classe de processos estocásticos, mas a generalização dos modelos já conhecidos. As contrapartes dos modelos AR, ARMA, ARIMA e SARIMA que utilizam variáveis exógenas são

ARX, ARMAX, ARIMAX e SARIMAX. Um maior detalhamento dos modelos que captam sazonalidade e usam variáveis exógenas pode ser obtido em Weron (2006).

3.4. Redes Neurais Artificiais

Uma RNA é um conjunto de elementos processadores (neurônios) interligados por um grande número de conexões, os quais são capazes de aprender (ajuste de pesos sinápticos), armazenar conhecimento e se adaptar a novas situações. A capacidade de aprender através de exemplos (padrões), e de generalizar (fornecer respostas adequadas para dados que não estão no conjunto de exemplos) é o principal atrativo das RNA (BRAGA *et al*, 2000). Uma de suas propriedades é a capacidade de aprender a partir de seu ambiente e de aperfeiçoar seu desempenho (‘performance’) através do aprendizado (HAYKIN, 2007).

Para Anbazhagan e Kumarappan (2013), as redes neurais têm sido usadas para resolver problemas de previsão de carga ou de preços, diagnóstico de componentes e falhas de sistema, avaliação de segurança entre outros. A RNA é uma ferramenta simples poderosa e flexível para a previsão, proporcionando uma melhor solução para modelar relações não-lineares complexas.

A principal limitação das redes neurais se dá pelo fato de que os dados entram na rede e uma previsão sai, mas o tipo de relacionamento entre as variáveis não é revelado (Francis, 2001), assim como detalhes de como são processados os dados para se chegar aos mapeamentos de resposta.

Embora existam inúmeras arquiteturas de redes neurais, a arquitetura das redes neurais de múltiplas camadas é a mais frequentemente encontrada na literatura (Skapura, 1996). Entre as razões para a sua popularidade cita-se a capacidade de aproximação de funções e a flexibilidade para formar soluções de qualidade para uma ampla classe de problemas, a partir do algoritmo de retropropagação do erro. O *backpropagation* foi um dos principais responsáveis pelo ressurgimento do interesse da comunidade científica em redes neurais, após um período de grande ceticismo em relação às suas potencialidades.

Para Patrikar (2013), uma das críticas a arquitetura das redes neurais de múltiplas camadas é que sua operação é como uma caixa preta, e os seus pesos não estão sujeitos a qualquer tipo de interpretação.

As arquiteturas do tipo múltiplas camadas, tipicamente, consistem em um conjunto de unidades sensoriais que formam uma camada de entrada, uma ou mais camadas intermediárias (ou escondidas) de unidades computacionais e uma camada de saída. Os sinais de entrada são propagados camada a camada pela rede em uma direção positiva, ou seja, da entrada para a saída. As camadas intermediárias funcionam como extratoras de características, seus pesos são uma codificação de características apresentadas nos padrões de excitações de entrada e permitem que a rede crie sua própria representação, mais rica e complexa.

O elemento básico de cada camada é o neurônio artificial. No modelo matemático proposto para o neurônio artificial, o corpo faz a soma ponderada do produto dos pesos e sinais de entrada e uma função é aplicada sobre a função de ativação para gerar a saída, como visto na Figura 3.4. Os pesos são as intensidades da força sináptica e podem ser fixos ou treináveis.

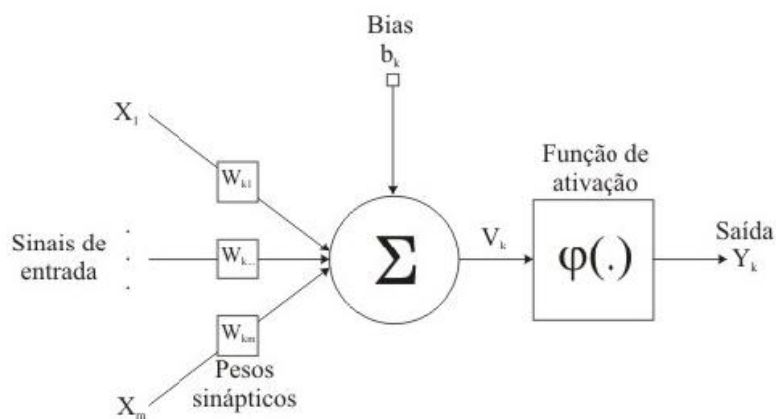


Figura 3.4 - Modelo de um neurônio artificial. Fonte: (Haykin, 2007).

As funções ideais para função de ativação dos neurônios são as consideradas quase lineares. Tais funções introduzem uma não linearidade sem no entanto alterar de forma radical a resposta da rede e possibilitam o cálculo da derivada parcial do erro em relação aos pesos de uma forma elegante e generalizável para todas as camadas. As equações 3.20 e 3.21 apresentam as funções tangente hiperbólica e sigmoide, que são comumente utilizadas como função de ativação.

$$\varphi(c) = \tanh(c) = \frac{e^c - e^{-c}}{e^c + e^{-c}} \quad (3.20)$$

$$\varphi(c) = \text{sigmoide}(c) = \frac{1}{1 + \exp(-c)} \quad (3.21)$$

As redes neurais múltiplas camadas levam o foco para um importante problema na metodologia científica, denominado de indução. A indução e a dedução são os únicos procedimentos sistemáticos conhecidos para construção do conhecimento científico. A dedução aplica princípios gerais a situações específicas. A dedução é bastante conhecida e teve enorme impacto na matemática, engenharia, ciência da computação, e na ciência em geral. Do outro lado, a indução é pouco compreendida e menos aplicada do que a dedução. A indução é o princípio de se abstrair regras gerais a partir de casos específicos. A indução é mais difícil de ser aplicada com validade do que a dedução. Nem sempre se consegue generalizar a partir de um conjunto pequeno de casos. A indução é também matematicamente menos formalizada do que a dedução (PRINCIPE et al., 2000).

A capacidade de generalização de uma RNA está intrinsecamente relacionada com fatores como a boa escolha de uma arquitetura de rede e o tipo de algoritmo de treinamento utilizado. Quando se escolhe de maneira satisfatória a arquitetura de rede e o algoritmo de treinamento consegue-se otimizar a resposta obtida pela RNA. Não há uma expressão analítica que defina uma arquitetura adequada para cada problema específico, apesar de existirem vários estudos nesta direção. Estas definições acabam por ser feitas de forma empírica, através do teste de vários modelos com diferentes números de camadas e neurônios e escolhendo aquele que proporcione o melhor desempenho.

Um dos desafios de aplicar as RNA na predição de séries temporais é incorporar a relação temporal existente entre as amostras de entrada e saída. O meio mais simples é utilizando diferentes variáveis de entrada defasadas ao longo do tempo. Por exemplo, para um vetor alvo de saída $s(t)$ da RNA, a série $\{s(t - 1), s(t - 2), \dots, s(t - \tau)\}$ pode ser utilizada como conjunto de variáveis de entrada.

A proposta deste trabalho não é de implementar uma rede neural puramente recorrente como são os casos das redes de Jordan ou redes de Elman (LIN et al, 2013). Nas redes puramente recorrentes, há neurônios adicionais, retroalimentados, atrasadores e mecanismos de reforço na propagação de parte dos sinais internos a rede. As recorrências serão implementadas através de memórias, com a introdução de elementos de atraso na estrutura da RNA, conforme proposto por Haykin (2007). Mantém-se assim uma rede RNA do tipo *Multilayer Perceptron* e topologia direta e com uso do algoritmo de treinamento de retropropagação do erro.

3.4.1. O Algoritmo *Backpropagation*

O algoritmo *backpropagation* ou algoritmo de retropropagação do erro, é o algoritmo mais utilizado para o treinamento das redes neurais múltiplas camadas. Foi desenvolvido por Rumelhart, Hinton e Williams em 1986 (Rumelhart et al., 1986) e consiste numa retropropagação de erro através das diversas camadas de uma RNA.

O algoritmo *backpropagation*, em sua maneira mais simples, é um algoritmo de gradiente descendente, assim como a regra de aprendizagem de Widrow-Hoff. Para o caso específico de predição de séries temporais, uma RNA é treinada com pares de vetores de entrada e saída. Durante o treinamento há o ajuste dos pesos sinápticos.

Basicamente, o processo de retropropagação do erro é constituído de duas fases: uma fase de propagação do sinal funcional (direta) e uma de retropropagação do erro (inversa). Na fase positiva, os vetores de dados são aplicados às unidades de entrada, e seu efeito se propaga pela rede, camada a camada. Finalmente, um conjunto de saídas é produzido como resposta da rede. Durante a fase positiva, os pesos das conexões são mantidos fixos. Na retropropagação do erro, por outro lado, os pesos são ajustados adaptativamente de acordo com uma regra de correção do erro.

Especificamente, a resposta da rede em um instante de tempo $s(t)$ é subtraída da saída desejada $sd(t)$ para produzir um sinal de erro. Este sinal de erro é propagado da saída para a entrada, camada a camada, originando o nome “retropropagação do erro”. Os pesos são ajustados de forma que a “distância” entre a resposta da rede e a resposta desejada seja reduzida.

Para Shi e Gao (2013), o algoritmo de retropropagação do erro procura o mínimo da função erro no espaço dos pesos sinápticos usando o método de gradiente descendente. A combinação de pesos que minimiza a função de erro é considerada uma solução de o problema de aprendizagem. Como este método requer o cálculo do gradiente da função de erro em cada passo de iteração, temos de garantir a continuidade e diferenciabilidade da função erro.

Pode-se iniciar o algoritmo atribuindo valores aleatórios para os pesos sinápticos. Quando o vetor de dados de entrada $x(t)$ é apresentado para a rede, a RN produz uma saída $\hat{y}(t)$, que em geral é diferente do vetor alvo $y(t)$. O que se quer é fazer com que $\hat{y}(t)$ e $y(t)$ sejam idênticos para $i = 1, \dots, P$ através do uso de um algoritmo de aprendizagem. Mais precisamente, queremos minimizar a função erro da rede, definido como:

$$E = \frac{1}{2} \sum_{i=1}^p \|\hat{y}(t) - y(t)\|^2 \quad (3.22)$$

A rede, na fase direta, calcula o erro de saída de cada neurônio. O valor do erro é propagado na rede, de modo que os neurônios das camadas escondidas e das camadas de saídas recebem os erros das camadas anteriores. Cada neurônio soma os valores do erro dos neurônios de camadas anteriores ligados a ele através de sinapses e, ao final, nos neurônios de saída da rede, tem-se o valor do erro total.

A esta altura, tem-se uma rede capaz de calcular o erro total para uma dado conjunto de treinamento. Os pesos da rede são os únicos parâmetros que podem ser modificados para tornar o valor da função erro o mais baixo possível. Assim, a função erro E é uma função contínua e diferenciável dos pesos w_1, w_2, \dots, w_l da rede. Podemos, assim, minimizar E usando um processo iterativo do gradiente descendente, processo de descida do gradiente:

$$\nabla E = \left(\frac{\partial E}{\partial w_1} + \frac{\partial E}{\partial w_2} + \dots + \frac{\partial E}{\partial w_l} \right) \quad (3.23)$$

Cada peso é ajustado utilizando-se o seguinte incremento:

$$\nabla w_i = -\gamma \frac{\partial E}{\partial w_i} \text{ para } i = 1, 2, \dots, l \quad (3.24)$$

Onde γ representa uma constante de aprendizagem, ou seja, um parâmetro de proporcionalidade que define o comprimento do passo de cada iteração na direção negativa do gradiente. Uma vez que se tem um método para calcular o gradiente, pode-se ajustar os pesos da rede de forma iterativa de maneira a encontrar o valor mínimo da função erro.

3.4.2. Sobre-treinamento

Para Wang et al (2008), o sobre-treinamento (*overfitting*) é um problema de generalização conhecido pelos pesquisadores das redes neurais. Ele ocorre porque conjuntos de treinamento finitos reduzem consideravelmente a capacidade de generalização da RN em aplicações práticas. Visando contornar o problema, um grande número de pesquisas têm sido conduzidas para que a capacidade de generalização das RN seja melhorado.

Conforme escreve Shao et al (2011), os aspectos chave nas RN são a otimização dos protocolos de treinamento, que incluem a arquitetura de rede, as taxas de aprendizagem, e o critério para parada do treinamento.

Quanto a escolha do critério de parada, 3 são as abordagens comumente encontradas na literatura: parar o treinamento de acordo com um número predefinido de iterações; parar de

treinar quando a taxa de erro pré-definido para o conjunto de treinamento é atingido; parar de treinar quando uma taxa de erro mínima seja alcançada por um conjunto de validação.

Nesse trabalho, o conjunto de dados de treino é de 70% da base de dados e o conjunto de teste de 30%. O critério de parada é o momento onde ocorre a taxa de erro mínima no conjunto de validação. O conjunto de validação neste caso está inserido dentro do conjunto de treino.

3.5 Árvores de decisão

Para Hullermeier e Vanderlooy (2009), as árvores de decisão são um dos métodos mais estudados em aprendizado de máquina e mineração de dados. Vários fatores contribuem para sua popularidade:

- As árvores de decisão são simples de compreender e de interpretar;
- Lidam bem com atributos numéricos e categóricos;
- Trabalham bem com bases de dados com muitos atributos e instâncias;
- Possuem um mecanismo intrínseco de seleção de atributos.
- Possui performance de classificação/predição igual ou superior aos métodos de destaque no estado-da-arte, especialmente quando se usam estratégias de *bagging/boosting* ou de hibridização.

Para Manwani e Sastry (2012), em uma árvore de decisão, cada nó não-folha está associado com um teste de decisão. Cada nó folha está associada com uma etiqueta de classe. Para classificar um conjunto de dados, em cada nó não-folha é feito um teste que ramifica o registro com base no valor assumido pela regra de divisão do respectivo nó.

Para Tan et al. (2009) as árvores de decisão são modelos estatísticos que utilizam um treinamento supervisionado para a classificação e previsão de dados. O treinamento supervisionado é aquele onde é utilizado um conjunto de treinamento formado por entradas e saídas. Assim, os valores da classe ou atributo da classe são conhecidos. As saídas podem ter domínios contínuos (numéricos) ou discretos (categóricos). Se o domínio do atributo classe é discreto, tem-se um problema de classificação. Se numérico, tem-se um problema de regressão ou aproximação de funções.

Surge daí os dois propósitos da classificação: a modelagem descritiva e a modelagem preditiva. A modelagem descritiva é útil quando se quer descrever o motivo pelo qual um determinado exemplo pertence a uma determinada classe. Na modelagem preditiva, um modelo é utilizado para classificar exemplos cujas classes são desconhecidas.

A classe é uma variável dependente e seu valor é determinado a partir dos valores das respectivas variáveis independentes.

Ainda segundo Tan et al. (2009), os problemas de classificação ocorrem onde a tarefa de aprender uma função alvo f que mapeia cada conjunto de atributos x , ou seja, cada exemplo das y classes pré-definidas. Nestes casos, os dados de entrada são definidos pelo par ordenado (x,y) , em que x é um vetor que representa os atributos preditivos, $x = (x_1, x_2, x_3, \dots, x_n)$, e y é o rótulo da classe à qual este exemplo pertence. Alguns pares ordenados podem ser tomados como conjunto de treinamento, e a partir deste conjunto de dados, faz-se a inferência indutiva, que pode criar hipóteses verdadeiras ou não. O modelo criado pode ser aplicado ao conjunto de teste, que consiste de registros com rótulos de classe desconhecidos.

Um grande desafio para os algoritmos de aprendizado é construir modelos que possuam boa capacidade de generalização, ou seja, consigam prever com alta taxa de acerto, rótulos de classe para exemplos que não foram utilizados na construção do modelo.

Outra característica dos classificadores a ser observada diz respeito a interpretabilidade do conhecimento adquirido. As redes neurais criam classificadores do tipo *Black-box* onde o conhecimento obtido estará representado pelo peso das entradas dos neurônios. Já o conhecimento obtido por árvores de decisão é explicitado a partir de regras do tipo SE (condição) ENTÃO (classe). Estas regras são expressas em linguagem natural, facilitando o entendimento por parte das pessoas.

A árvore de decisão é construída usando um conjunto de amostras de treino e então é aplicada para classificar um conjunto de amostras novas, e ainda não vistas, chamadas de amostras de teste. A árvore de decisão é moldada usando uma estratégia de desenvolvimento do tipo top-down. Começa a partir de um nó raiz e as amostras são então classificadas, colocando uma série de testes lógicos de características associadas com os dados. Um nó é dividido em dois sub ramos de acordo com as possíveis respostas para a pergunta. Para responder a cada pergunta, o algoritmo usa um conjunto de critérios estatísticos para a classificação de dados. As duas medidas comumente utilizadas são entropia e o índice Gini.

Para um conjunto de dados típico de duas classes, com as classes alvo negativo e positivo, a entropia dos dados é definida como se segue:

$$\Delta I(p, n) = -D_p \log_2(D_p) - D_n \log_2(D_n) \quad (3.23)$$

Onde D_p é a precisão e D_n a taxa de erro

A matriz de confusão é uma técnica empregada para se analisar o desempenho de árvores de decisão. Trata-se de uma matriz é formada por um arranjo quadrado de números dispostos em linhas e colunas que expressam o número de unidades de amostras de uma categoria particular relativa – inferida por um classificador (ou regra de decisão), comparado com a categoria atual verificada no campo (CONGALTON, 1991). A diagonal principal da matriz exibe o número de acertos para as classes analisadas, enquanto os elementos fora da diagonal representam o número de erros. A matriz gera indicadores de exatidão e de taxa de erros. Congalton e Green (1999) descrevem os cálculos associados com estas medidas.

Tabela 3.1 - Matriz de confusão para um problema de duas classes

		Classe prevista	
		Classe = 1	Classe = 0
Classe Real	Classe = 1	f_{11}	f_{10}
	Classe = 0	f_{01}	f_{00}

Cada entrada f_{ii} na tabela 3.1 denota o número de registros da classe 0 previstos incorretamente como classe 1. Baseado nas entradas da matriz de confusão, o número total de previsões corretas feita pelo modelo é $(f_{11} + f_{00})$ e o número total de previsões incorretas é $(f_{10} + f_{01})$.

Embora a matriz de confusão forneça as informações necessárias para determinar o quão bem um modelo de classificação é executado, resumir estas informações com um único número tornaria mais conveniente comparar o desempenho de diferentes modelos. Isto pode ser feito usando uma métrica de desempenho como a precisão que pode ser definida conforme a equação 3.24:

$$\text{Precisão} = \frac{\text{Número de previsões corretas}}{\text{Número total de previsões}} = \frac{f_{11} + f_{00}}{f_{11} + f_{10} + f_{01} + f_{00}} \quad (3.24)$$

De forma equivalente, o desempenho de um modelo pode ser expresso em termos da sua taxa de erro, que é dada pela equação 3.25:

$$\text{Taxa de erro} = \frac{\text{Número de previsões erradas}}{\text{Número total de previsões}} = \frac{f_{10} + f_{01}}{f_{11} + f_{10} + f_{01} + f_{00}} \quad (3.25)$$

A maioria dos algoritmos de classificação procura modelos que atinjam a maior precisão ou, equivalentemente, a menor taxa de erro quando aplicados a um conjunto de testes.

Para Amraee & Ranjbar (2013), a entropia é menor quando todas as amostras têm a mesma classe de destino, enquanto o valor máximo ocorre quando o número de amostras classificado como positivo é igual ao número de dados classificado como negativo. O índice de Gini em um nó, é definido pela equação 3.26:

$$G(e) = 1 - \sum_i P\langle i|e \rangle^2 \quad (3.26)$$

onde $P(i|e)$ é a probabilidade condicional de categoria i no nó e . Define-se esta probabilidade através das equações 3.27, 3.28 e 3.29:

$$P(i|e) = \frac{P(i,e)}{P(e)} \quad (3.27)$$

$$P(i, e) = \frac{(\pi_i N_i(e))}{N_i} \quad (3.28)$$

$$P(e) = \sum_i P(i, e) \quad (3.29)$$

onde π_i é o valor de probabilidade prévia para a categoria i , $N_i(e)$ é o número de registros da classe i no nó e , e N_i é o número de registros da categoria i no nó raiz. O índice de Gini é zero quando todas as amostras em conjunto têm a mesma classe de destino. A árvore de decisão se desenvolve adicionando testes lógicos sobre amostras de treinamento restantes. Um bom teste vai dividir uma grande coleção de dados com diferentes classes de destino em subconjuntos com quase os mesmos tipos de rótulos, de modo que há pouca variação em cada ramo.

A diferença entre a entropia da distribuição das classes no nó raiz e a média ponderada da entropia de um nó interno é chamado o ganho de informação. O ganho de informação no nó é expressa como:

$$\text{Ganho de informação } (s, e) = I(p, n) - E(s) \quad (3.30)$$

onde s é o conjunto de dados divididos no nó e . O ganho de informação tem sempre um valor positivo, porque $I(p, n)$ é a entropia de todas as amostras em um nó e E é a entropia de várias classes do nó. Portanto $I(p, n)$ é sempre maior do que E . O ganho de informação é calculado em cada nó não-folha e a característica com o maior ganho de informação será selecionado como recurso de entrada para esse nó. Este processo será repetido recursivamente para classificar os dados em subconjuntos menores no espaço de árvore até que todos os dados sejam classificados.

O crescimento integral das árvores de decisão leva ao over-fitting. Portanto, as árvores de decisão devem ser podadas para evitar o excesso de ajuste dos dados de treinamento.

Várias técnicas de poda foram introduzidas na literatura. Alguns deles são: *cost complexity pruning*, *reduced error pruning*, *pessimistic pruning*, *error-based pruning*, *penalty pruning* e *MDLpruning*. Os detalhes destes métodos podem ser encontrados em Esposito et al (1997).

3.5.1 Indução de árvores de decisão

Para ilustrar como as árvores de decisão são induzidas, vamos utilizar a figura 3.5 que representa uma árvore de decisão onde cada nó de decisão contém um teste para algum atributo, cada ramo descendente corresponde a um possível valor deste atributo, cada folha está associada a uma classe e, cada percurso da árvore, da raiz à folha corresponde uma regra de classificação.

No exemplo a raiz da árvore é o atributo carga de energia. Trata-se da raiz porque não há ramos chegando e há um ou mais ramos saindo. Os nós internos são Geração Hidráulica e Geração térmica, visto que tais nós possuem ramos chegando e saindo. As folhas ou nós terminais são PLD1, PLD2, PLD3, PLD4 e PLD 5.

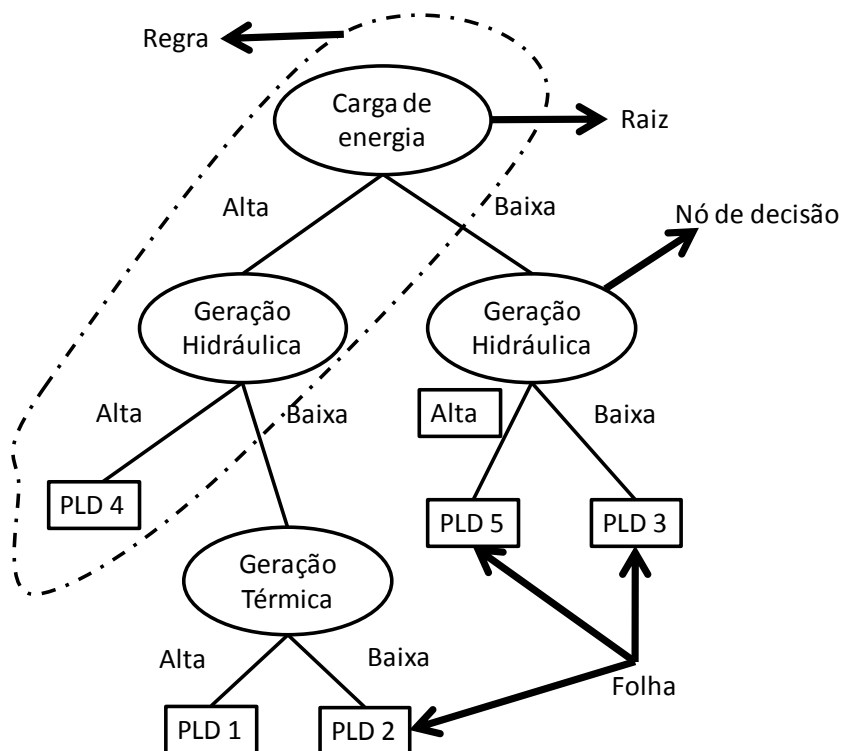


Figura 3.5 - Elementos de uma árvore de decisão

No espaço definido pelos atributos, cada folha corresponde a um hiper-retângulo onde a interseção destes é vazia e a união é todo o espaço, conforme a Figura 3.6.

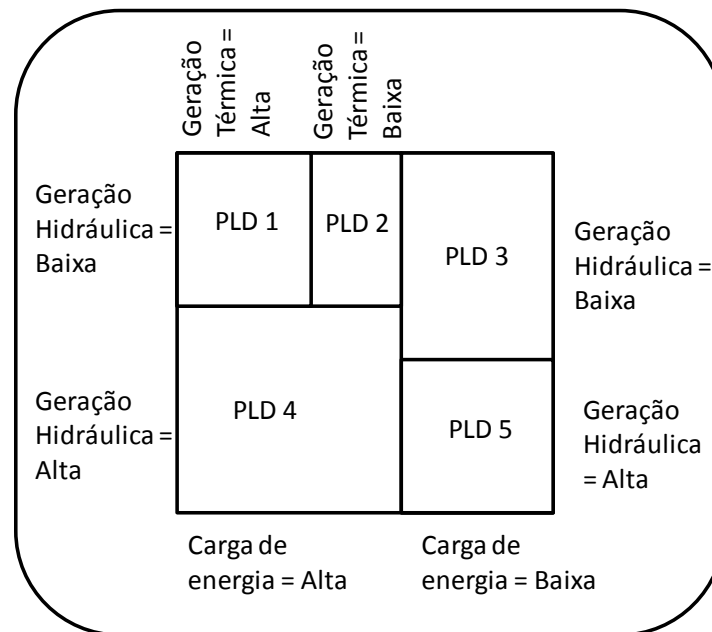


Figura 3.6 - Hiper-retângulos da árvore de decisão

Na fase de construção da árvore de decisão, uma árvore é gerada pelo particionamento recursivo dos dados de treinamento. O conjunto de treinamento é separado em duas ou mais partições usando restrições sobre os conjuntos de valores de cada atributo. O processo é repetido recursivamente até que todos ou a maioria dos exemplos em cada partição pertençam a uma classe. A árvore gerada abrange todo o conjunto de treinamento e é construída por meio da estratégia de busca em largura. Assim, todos os nós em uma determinada altura da árvore devem ser processados antes do início do processo do nível subsequente.

Há duas operações principais durante o processo de construção de uma árvore: a avaliação dos pontos de separação de cada nó interno da árvore com a identificação de qual é o melhor ponto de separação e a criação de partições usando o melhor ponto de separação identificado para os casos pertencentes a cada nó. Uma vez determinado o melhor ponto de separação de cada nó, as partições podem ser criadas pela simples aplicação do critério de separação identificado. Para a o tipo de árvore sob análise, CART, o índice utilizado é o índice Gini.

3.5.2 Algoritmos de indução de árvores de decisão

O algoritmo C5.0 é uma evolução do C4.5 (Quinlan, 1993) e ID3 (Quinlan, 1986). O C5.0 lida tanto com atributos categóricos como contínuos. No caso dos atributos contínuos, o algoritmo define um limiar e divide os registros de forma binária: aqueles cujo valor do atributo são maiores que o limiar e aqueles cujo valor do atributo são menores ou iguais ao limiar. Ele trata valores desconhecidos de forma especial, não utilizando-os nos cálculos de ganho e de entropia. Para Bagaslupp (2010), o algoritmo utiliza a medida chamada *gain ratio* para selecionar o atributo que melhor dividem os exemplos. Essa medida se mostrou superior ao ganho de informação, gerando árvores mais precisas e menos complexas. Apresenta um método de pós-poda das árvores geradas. O algoritmo C5.0 faz uma busca na árvore, de baixo para cima, e transforma em nodos folha aqueles ramos que não apresentam nenhum ganho significativo. Esta família de algoritmos é a mais utilizada na literatura, justamente por ter mostrado ótimos resultados em problemas de classificação (WITTEN & FRANK, 2011).

O algoritmo CART (*Classification and Regression Trees*) foi proposto em Breiman et al. (1984) e consiste de uma técnica não paramétrica com grande capacidade de pesquisa de relações entre os dados, mesmo quando elas não são evidentes, bem como a produção de resultados sob a forma de árvores de decisão de simplicidade e legibilidade. Para Bagaslupp (2010), o CART produz árvores binárias que podem ser percorridas da sua raiz até as folhas respondendo apenas a questões simples do tipo “sim” ou “não”. Os nodos que correspondem a atributos categóricos são representados por agrupamento de valores em dois conjuntos. O algoritmo utiliza técnica de pesquisa exaustiva para definir os limiares a serem utilizados nos nodos para dividir os atributos contínuos e permite a utilização de combinações lineares entre atributos. Do ponto de vista da poda, o CART expande a árvore exaustivamente, realizando pós-poda por meio da redução do fator custo-complexidade (Breiman et al., 1984). Segundo os autores, a técnica de poda utilizada é extremamente eficiente e produz árvores mais simples, precisas e com ótima capacidade de generalização.

3.6 Considerações Finais

Este capítulo apresentou os fundamentos dos métodos utilizados para predição e classificação do preço de energia no mercado brasileiro. Dois modelos são propostos: um modelo híbrido que utiliza ARIMA e RN e um modelo que combina árvores de decisão e RN.

No capítulo seguinte será apresentada a metodologia completa do modelo proposto para a predição do preço de energia no mercado Brasileiro.

CAPÍTULO IV

METODOLOGIA

4.1. Introdução

Esta tese de doutorado tem por objetivo desenvolver modelos de previsão futura do preço de energia aplicado ao mercado Brasileiro de curto prazo. Para tal dois modelos são propostos. O primeiro modelo é híbrido e combina as técnicas ARIMA e RN, sendo capaz de estimar o preço da energia até 12 passos (semanas) a frente. O segundo modelo, propõe o uso de árvores de decisão e RN para a classificação do preço futuro de energia elétrica, e assim criar uma estratégia de gerenciamento de risco eficaz para participantes do mercado de curto prazo de energia elétrica no Brasil. Ambas as propostas serão apresentadas a seguir.

O software utilizado nesta tese de doutorado é o IBM SPSS *Modeler*. Trata-se de um ambiente de trabalho de mineração de dados, versátil e poderoso, projetado para construção de modelos preditivos sem programação, de forma rápida e intuitiva. Este software que possui uma grande variedade de técnicas analíticas para acessar diversas fontes de dados, tais como data *warehouses*, bancos de dados, *Hadoop* ou *flat files*. A vasta gama de técnicas de modelagem, oportuniza que em uma única execução sejam utilizadas várias técnicas de modelagem e comparados os resultados de diferentes algoritmos de mineração de dados. O SPSS *Modeler* possui todos os algoritmos utilizados neste projeto, a saber: ANOVA, PCA, ARIMA, RN, CART, C5.0 e seleção de variáveis explanatórias.

4.2. Banco de Dados

Esta seção apresenta a metodologia de montagem do banco de dados utilizado nesta tese de doutorado. Foram usadas duas fontes de informação na montagem do banco de dados:

- ❖ A série histórica de preços PLD obtida junto a Coordenação de preços da Câmara comercialização de energia elétrica (CCEE). A série obtida vai de janeiro de 2001 até dezembro de 2013 e apresenta os valores semanais de PLD para os 4 submercados (Norte, Nordeste, Sul e Centro-oeste/Sudeste) em 3 patamares de carga: pesado, médio e leve.

- ❖ O Informativo preliminar diário da operação (IPDO) fornecido pelo operador nacional do sistema (ONS) através da URL <http://www.ons.org.br/publicacao/ipdo/>. Foram feitos os downloads de todos os informativos de janeiro de 2001 até dezembro de 2013. De cada informativo foram extraídas as informações de carga de energia, a energia natural afluyente, a energia armazenada, a geração hidráulica e a geração térmica, para os 4 submercados brasileiros.

Foi digitalizado um total de 4522 informativos, totalizando 18.031 páginas. A partir dos IPDO's, os dados diários foram registrados em tabelas mensais e anuais. Devido ao fato de os valores do PLD estarem disponíveis em base semanal e os dados da ONS em base diária, houve o esforço de padronização das tabelas de dados oriundos do IPDO. Desta forma, as medidas de carga de energia, energia natural afluyente, energia armazenada, geração hidráulica e geração térmica tiveram seus valores ajustados para medidas semanais. A figura 4.1 apresenta a seção do IPDO de onde os dados são obtidos.

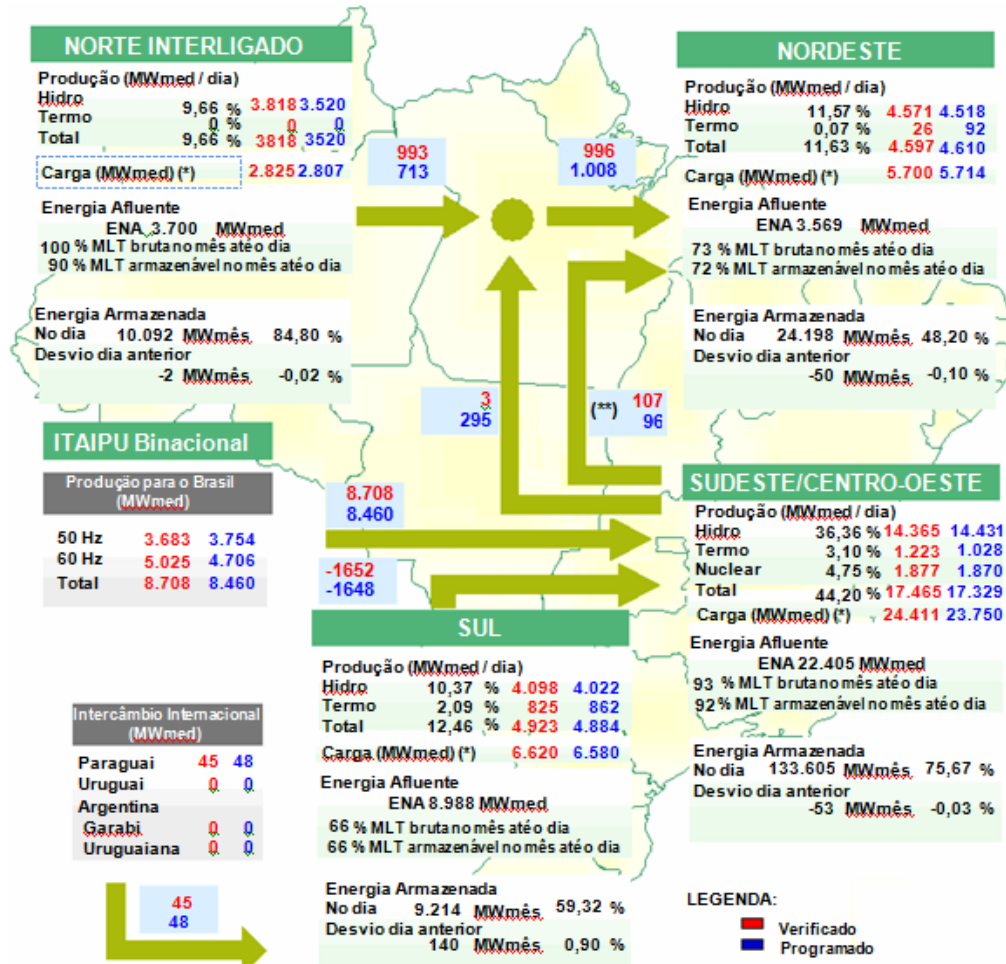


Figura 4.1 - Fragmento do IPDO do dia 07/06/2003 com a disponibilização dos dados da ONS

A Tabela 4.1 sintetiza a descrição, a unidade de medida e a fonte de informação de cada uma das variáveis que compõem o banco de dados utilizado. As variáveis são: PLD, geração hidráulica, geração térmica, carga de energia, energia armazenada e energia natural afluente.

Tabela 4.1 - Atributos componentes do banco de dados

Sigla	Descrição	Unidade	Orgão
PLD	Preço de Liquidação de Diferenças - PLD no patamar médio	R\$	C.C.E.E
GerHid	Geração Hidráulica	MWmed	O.N.S
GerTer	Geração térmica	MWmed	O.N.S
Carga	Carga de energia	MWmed	O.N.S
EARM	Quantidade de água armazenada em um reservatório, transformada em energia	% MLT	O.N.S
ENA	Quantidade de água que entra em um reservatório, transformada em energia	% MLT	O.N.S

A Figura 4.2 apresenta o comportamento das 6 séries temporais componentes do banco de dados para o submercado nordeste. A análise estatística dos dados das 6 séries temporais, apresentando as medidas de posição e de dispersão, são apresentadas nas Tabelas 4.2, 4.3, 4.4 e 4.5. As séries não apresentam um comportamento gaussiano e apresentam valores muito elevados de desvio padrão e variância.

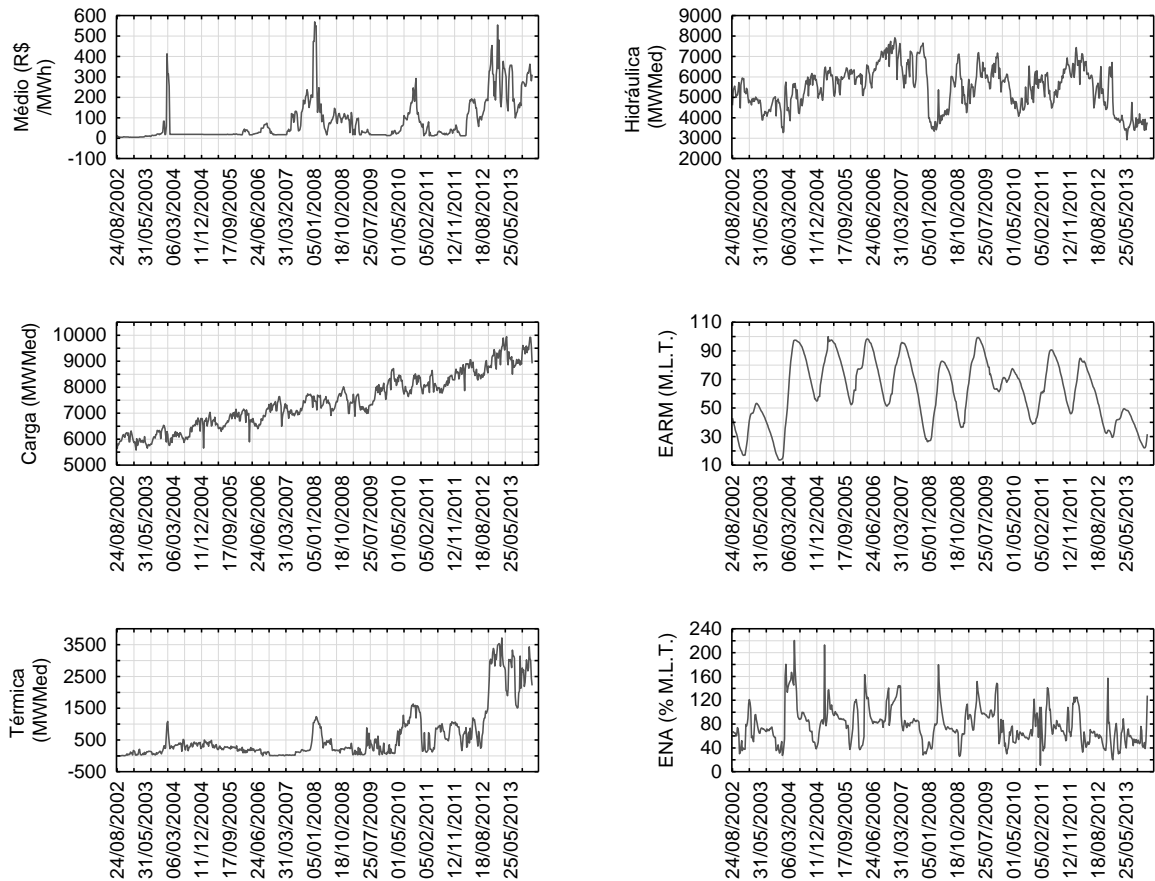


Figura 4.2 - Séries temporais do banco de dados para região nordeste

Tabela 4.2 - Medidas de posição e dispersão da região SE/CO

Região	Sudeste					
	PLD	Hidraulica	Térmica	Carga	EARM	ENA
Média	74,65	17.635,02	1.112,57	28.048,88	65,30	103,87
Mín.	4,00	8.854,14	192,86	19.295,57	20,71	47,86
Máx.	684,00	23.378,14	3.258,86	34.668,00	87,64	182,00
Amplitude	680,00	14.524,00	3.066,00	15.372,43	66,93	134,14
Variância	13.812,64	10.587.485,82	379.375,16	9.913.769,09	278,87	594,54
Desvio padrão	117,53	3.253,84	615,93	3.148,61	16,70	24,38
Mediana	27,72	18.217,22	982,93	28.365,79	68,96	101,50
Moda	18,59	18.426,57	245,00	31.425,86	73,67	83,71

Tabela 4.3 - Medidas de posição e dispersão da região S

Região	Sul					
	PLD	Hidraulica	Térmica	Carga	EARM	ENA
Média	51,82	5.606,83	975,67	7.709,98	68,07	98,02
Mín.	4,00	422,57	156,57	5.893,29	27,83	15,14
Máx.	569,59	16.149,57	1.861,14	9.432,71	97,96	366,43
Amplitude	565,59	15.727,00	1.704,57	3.539,42	70,13	351,29
Variância	4.545,69	4.855.836,54	84.607,87	444,82	368,05	4.134,85
Desvio padrão	67,42	2.203,60	290,87	666,95	19,19	64,30
Mediana	24,17	5.402,57	997,71	7.666,14	68,56	78,36
Moda	18,59	6.607,29	499,29	6.081,00	36,93	51,00

Tabela 4.4 - Medidas de posição e dispersão da região NE

Região	Nordeste					
	PLD	Hidraulica	Térmica	Carga	EARM	ENA
Média	54,35	5.648,83	244,93	6.872,12	66,43	86,60
Mín.	4,00	3.273,57	0,86	5.580,71	13,53	25,71
Máx.	569,59	7.918,86	1.234,57	8.156,00	99,92	220,43
Amplitude	565,59	4.645,29	1.233,71	2.575,29	86,39	194,72
Variância	6.657,46	1.097.009,16	50.574,46	367.025,80	520,03	1.032,12
Desvio padrão	81,59	1.047,38	224,89	605,83	22,80	32,13
Mediana	18,59	5.765,29	198,00	6.932,57	69,54	84,43
Moda	18,59	3.552,29	5,43	6.175,43	13,83	72,71

Tabela 4.5 - Medidas de posição e dispersão da região N

Região	Norte					
	PLD	Hidraulica	Térmica	Carga	EARM	ENA
Média	66,84	3.633,04	0,00	3.115,77	62,65	93,17
Mín.	4,00	539,00	0,00	2.041,57	8,73	31,86
Máx.	684,00	7.279,71	0,00	3.780,00	99,67	264,43
Amplitude	680,00	6.740,71	0,00	1.738,43	90,94	232,57
Variância	13.172,80	1.495.703,98	0,00	194,64	671,69	1.120,67
Desvio padrão	114,77	1.222,99	0,00	441,18	25,92	33,48
Mediana	18,59	3.384,57	0,00	3.165,57	65,75	87,29
Moda	18,59	2.527,00	0,00	2.811,14	19,66	78,00

4.3. Modelo híbrido proposto

O modelo híbrido proposto combina a capacidade do modelo ARIMA com uma Rede Neural para a predição do Preço de Liquidação de Diferenças (PLD) em até 12 passos (semanas) à frente, utilizando variáveis explanatórias. A metodologia proposta segue duas etapas. Em uma primeira etapa, os valores futuros das variáveis explanatórias selecionadas são preditos. Em uma segunda etapa, o preço de energia é predito utilizando a previsão das variáveis explanatórias.

No caso do ARIMA, utiliza-se uma metodologia consolidada, que facilita a identificação da configuração mais adequada modelo mais adequado para modelar as séries temporais pertinentes à formação de preços de energia elétrica. Já para as RN, aproveita-se a capacidade flexível de modelagem não-linear. Assim, a abordagem híbrida permite que um modelo complemente o outro na captura de padrões e das dependências internas das séries temporais (Sterba e Hilovska, 2010). Partindo do princípio que as series temporais não são completamente lineares ou não-lineares, o uso de um modelo composto permite a apreensão tanto da parte linear dos dados quanto das irregularidades e eventos pouco frequentes.

A Figura 4.3 abaixo apresenta a metodologia proposta para predição futura do PLD, a qual pode ser resumida em 4 etapas.

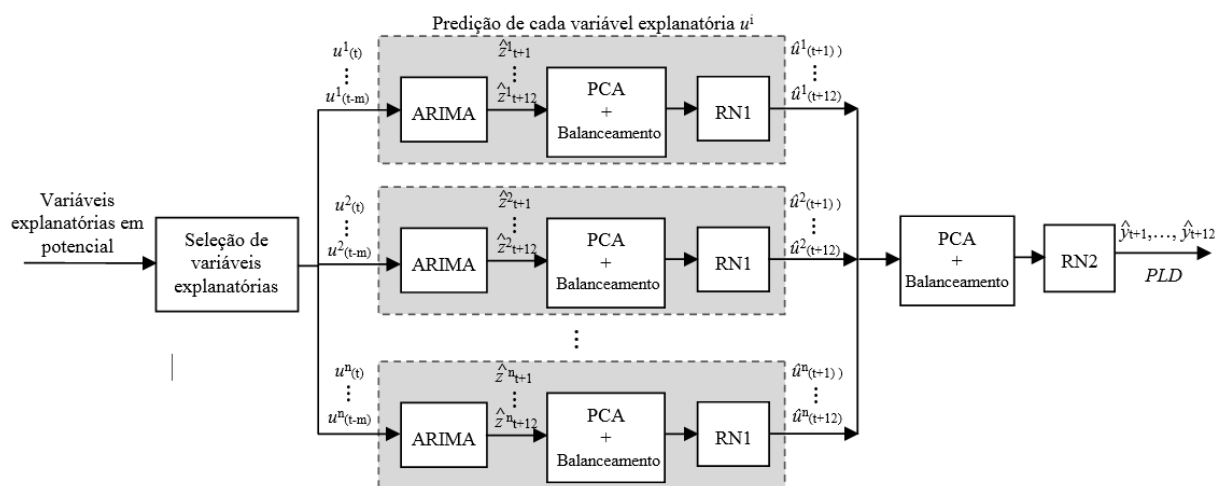


Figura 4.3 - Diagrama de blocos do modelo híbrido preditor

- ❖ Etapa 1: cria-se um grande banco de dados que inclui dados históricos do PLD e variáveis explanatórias em potencial que afetam o preço da energia de curto prazo. Considera-se

que a série temporal do PLD y_t é função de n variáveis explanatórias u_t^i , $i = 1 \dots n$. Cada variável explanatória u_t^i é composta de uma estrutura linear e uma componente não-linear, $u_t^i = f(L_t, N_t)$, onde L_t denota a componente linear e N_t denota a componente não-linear;

- ❖ Etapa 2: aplica-se uma técnica de seleção de variáveis explanatórias para identificar o melhor conjunto de variáveis explanatórias (u_t^i) para cada submercado. Neste trabalho utilizou-se o método teste f ANOVA.
- ❖ Etapa 3: faz-se a predição de cada variável explanatória (\hat{u}_{t+12}^i) 12 passos (semanas) a frente. Primeiramente, um filtro ARIMA é utilizado para modelar as relações lineares da série temporal. Depois de ajustado o modelo ARIMA, ele passa a ser utilizado para fazer a predição de valores futuros das variáveis explanatórias, $\hat{z}_{t+12}^i = L^{ARIMA}(u_t^i)$. São encontrados valores até 12 passos à frente, utilizando uma estratégia recursiva, na qual valores preditos pelo modelo são novamente injetados na entrada dele. Estes valores entram na primeira Rede Neural (RN1), que é utilizada para mapear as relações não-lineares das séries de variáveis explanatórias. De modo a garantir capacidade de generalização, é importante reduzir a dimensionalidade dos vetores de entrada e adotar a melhor configuração de aprendizagem antes de treinar a rede. Isso é feito na etapa de preparação de dados com a aplicação da Análise de Componentes Principais (PCA) e um procedimento de balanceamento das amostras de treinamento. A RN1 tem apenas um neurônio de saída, que é utilizado para fazer a predição futura do valor da variável explanatória um passo à frente. Um processo de iteração recursivo é novamente usado para garantir a predição futura 12 passos à frente, $\hat{u}_{t+12}^i = N^{RN1}(\hat{z}_{t+12}^i)$.
- ❖ Etapa 4: faz-se a predição do PLD 12 semanas à frente (\hat{y}_{t+12}), utilizando uma segunda Rede Neural (RN2) precedido do processo de preparação de dados novamente. As

entradas desta Rede Neural são as variáveis explanatórias preditas 12 semanas à frente. A RN2 tem apenas um neurônio de saída e prevê as 12 semanas à frente utilizando técnica de previsão iterativa, $\hat{y}_{t+12} = N^{RN2}[\hat{u}_{t+12}^1, \hat{u}_{t+12}^2, \dots, \hat{u}_{t+12}^n]$.

Antes do fluxo de dados chegar na RNA, é necessário usar uma estratégia de redução de dimensionalidade e balanceamento das amostras de treinamento.

Segundo Haykin (2007), as redes neurais apresentam grande sensibilidade à dimensão do espaço amostral dos vetores de entrada. Através do uso da PCA, são reduzidos os atributos de entrada para 5 atributos com variância superior aos atributos originais, o que melhora a capacidade de treinamento da rede.

Para Park e Gosh (2014), as distribuições desequilibradas de amostras são encontradas com maior frequência em problemas de classificação do mundo real, para os casos de detecção de fraudes, gerenciamento de riscos, classificação texto, diagnóstico médico, e muitos outros domínios. Bases de dados desbalanceadas, possuem assimetria das distribuições de classes. Há pouquíssimos trabalhos abordando esta questão.

Para Haibo He e Garcia (2009), considera-se uma base de dados desbalanceada, aquela onde há distribuição desigual, significativa, entre as classes de valores. Ainda segundo os autores, duas são as estratégias de reamostragem utilizadas para balanceamento de dados: supra-amostragem de valores majoritários e super-amostragem de valores minoritários.

Neste trabalho usa-se o método em nível de dados de super-amostragem, onde os pontos de dados das classes minoritárias são replicadas aleatoriamente usando os membros da classe minoritária existentes. Atuando no nível de dados em uma etapa de pré-processamento, reequilibra-se a distribuição de classes redefinindo o espaço de dados, sem modificar o algoritmo de aprendizagem e de forma independente da modelo de predição utilizado. Maiores detalhes sobre esta solução podem ser vistos em Galar et al (2012).

4.4. Modelo de classificação proposto

Além do modelo híbrido de predição para determinar o valor exato do PLD, este trabalho propõe um modelo de classificação de preço da energia de curto prazo utilizando árvores de decisão. Observa-se na literatura existente que os modelos tradicionais de previsão de preços tradicionais tentam prever o valor exato de preços no futuro. No entanto, nem todos os participantes do mercado precisam saber o valor exato de preços futuros em seu processo

de tomada de decisão. Em alguns casos, é suficiente saber se o preço vai ser "muito caro" para tomar uma decisão. O modelo proposto não prevê o valor exato do preço futuro de energia elétrica, mas a classe a que pertence, estabelecida em relação a valores pré-especificados, dentro da faixa de valores da ANEEL cujo piso é R\$ 15,62 e o teto R\$ 822,83.

Para o modelo classificador, são investigados os algoritmos de árvore de decisão são investigados CART e C5.0. Os resultados obtidos são comparados a um classificador baseado em Redes Neurais e com modelos combinando CART, C5.0 e RN dois a dois. A técnica de seleção de variáveis explanatórias é novamente utilizada. O horizonte de tempo de previsão é de uma semana à frente.

O C5.0 é um avanço do algoritmo C4.5, o mais utilizado na literatura pelos ótimos resultados obtidos em problemas de classificação. Ele lida bem tanto com atributos categóricos e contínuos. Uma das vantagens do C5.0 é que os valores desconhecidos não impactam nas medidas de avaliação de desempenho do modelo. Além disso, usa a medida *gain ratio* para dividir os exemplos, o que gera árvores mais precisas e menos complexas. Apresenta um método de pós-poda das árvores geradas.

O algoritmo CART é uma técnica não paramétrica com grande capacidade de produção da relação entre dados sob a forma de árvores de decisão de grande simplicidade, legibilidade e ótima capacidade de generalização. A exemplo do C5.0, também funciona para dados contínuos e categóricos. O CART produz árvores binárias com testes cujas respostas são sim ou não. Do ponto de vista da poda, o CART expande a árvore exaustivamente, realizando pós-poda por meio da redução do fator custo-complexidade.

A figura 4.4 apresenta o diagrama de blocos da metodologia proposta para classificação do preço de energia no mercado Brasileiro.

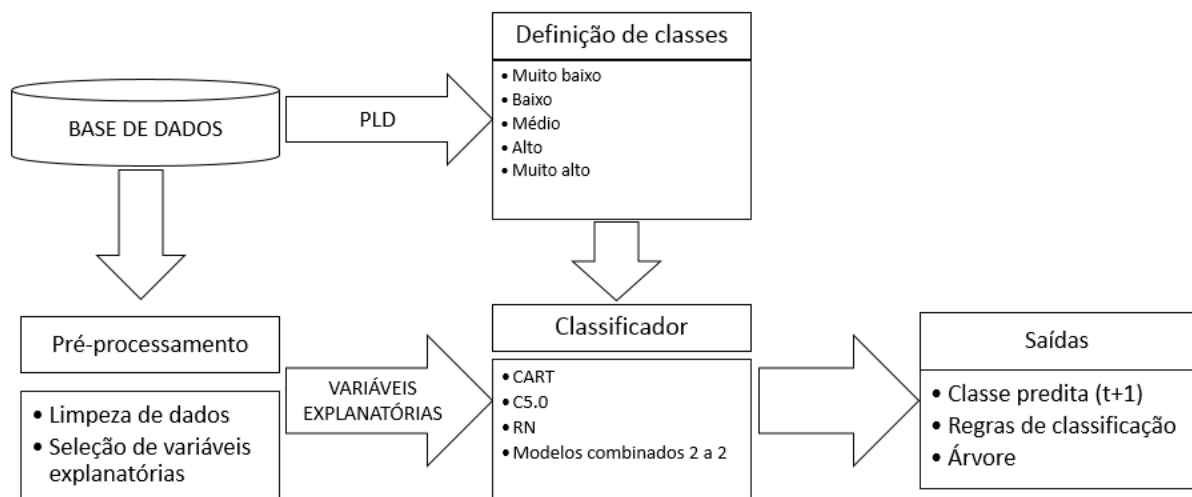


Figure 4.4 – Diagrama de blocos de solução de classificação

A metodologia de classificação proposta pode ser resumida nas seguintes etapas:

Etapa 1: Criação da base de dados

Esta etapa prevê a criação de um grande banco de dados que inclui dados históricos do PLD e um conjunto inicial de potenciais variáveis explanatórias que possam explicar as variações dos preços de energia. Tais variáveis são: carga de energia, a energia armazenada, geração hidráulica, geração térmica e energia natural afluenta.

Etapa 2: Pré-processamento

Nesta etapa aplica-se o pré-processamento de dados, que inclui limpeza, normalização e seleção de variáveis explanatórias, a fim de eliminar entradas irrelevantes e redundantes. O método utilizado foi o teste f ANOVA.

Etapa 3: Definição das classes de preços

Nesta etapa definiu-se 5 classes para o preço de energia. Os intervalos de classe foram manualmente definidos com base em informações de empresas de energia elétrica e os intervenientes no mercado, de acordo com o seu valor e significado para operações no mercado de energia. Essas classes são rotuladas de acordo com as faixas de preço: muito baixo, baixo, médio, alto e muito alto. Foram adotados quatro limiares de preços por megawatt-hora: T1 = R\$ 50,00; T2 = R\$ 90,00; R\$ 150,00 e T4 = R \$ 450,00. Então, cinco classes foram criadas com base nesses limites:

- Classe 1 (preço muito baixo): abaixo de T1;
- Classe 2 (preço baixo): entre T1 e T2;
- Classe 3 (preço médio): entre T2 e T3;
- Classe 4 (preço alto): entre T3 e T4;
- Classe 5 (preço muito alto): acima de T4 até o teto definido pela ANEEL;

Etapa 4: Classificação

Nesta etapa faz-se a previsão das classes de preços de energia elétrica de curto prazo 1 semana à frente. O classificador é treinado para determinar a "classe" de preços futuros. Em outras palavras, o classificador determina em que faixa de preços futuros cairão sem estimar o valor exato do preço. Uma árvore de decisão é obtida.

Para potencializar os resultados dos modelos CART e C5.0 são utilizadas as técnicas de *bagging* e *boosting*. Para Banfield et al (2007) *bagging* é uma das mais simples, conhecidas e antigas técnicas de amostragem. Nesta técnica cria-se um conjunto de classificadores, por amostragem, com a substituição do conjunto de dados original de treinamento por novos conjuntos de treinamento chamados de *bags*. A quantidade de registros

de cada *bag* é o mesmo do conjunto original de treinamento. Um classificador separado é formado a partir de cada saco. Já Freund e Schapire (1996) propuseram um algoritmo de *boosting* que cria classificadores utilizando um conjunto de treino com pesos atribuídos a cada registro. Aos exemplos que são incorretamente classificados por um classificador é dado um incremento no peso para a iteração seguinte.

Para combinar os modelos 2 a 2 foi utilizada a estratégia de votação. Para Littlestone e Warmuth (1994), a *Weighted-Majority Voting* (WMV) é um procedimento de meta-aprendizagem. Conforme proposto por Fangming Ye et al (2013), o primeiro tipo de votação por maioria refere-se à decisão quando todos os classificadores concordam com a mesma saída (unanimidade). O segundo tipo de votação por maioria refere-se à decisão quando a metade mais um dos classificadores concordam com a mesma saída (maioria simples). A terceira abordagem de votação é chamada de votação ponderada de maioria. Se há evidências que certos classificadores são mais qualificados do que outros, ponderando as decisões desses modelos de acordo com a exatidão e pode-se melhorar ainda mais o desempenho global do sistema.

Neste trabalho serão usados os processos de *boosting* para as árvores de decisão, votação ponderada de maioria para combinação de modelos, validação cruzada do tipo 10-fold para árvores de decisão e partição de dados na proporção 70% (treino) e 30% (teste) para evitar o sobre-treinamento das Redes Neurais.

No total, seis casos são analisados nos modelos de classificação, cujos resultados serão apresentados na próxima seção:

- CART;
- C5.0;
- Redes Neurais;
- RN + CART;
- RN + C5.0;
- C5.0 + CART.

4.5 Considerações Finais

Descrito os detalhes da metodologia empregada para a predição e classificação do PLD, o próximo capítulo apresentará os resultados obtidos com o uso da metodologia híbrida de predição de valores até t+12 passos à frente para o mercado brasileiro, assim como o modelo classificador que identifica a qual classe de valores pré-definidas estará o preço da energia elétrica em t+1.

CAPÍTULO V

RESULTADOS, VALIDAÇÃO EXPERIMENTAL E ANÁLISE

5.1 Introdução

Este capítulo apresenta os resultados obtidos com a aplicação dos métodos propostos para predição do PLD nas bases de dados dos 4 submercados brasileiros. Com o objetivo de fazer uma análise comparativa, os resultados são comparados com aqueles alcançados através de outras técnicas comumente empregadas.

Para análise dos dados foi utilizado o *software* IBM SPSS MODELER versão 15, já descrito em detalhes no capítulo anterior. A ferramenta possibilita a descoberta rápida de padrões e tendências em dados numéricos estruturados de maneira mais fácil, usando uma única interface visual apoiada por análises avançadas.

5.2 Resultados obtidos utilizando o modelo híbrido

Esta seção apresenta os resultados obtidos para previsão do PLD utilizando o modelo híbrido proposto, composto pelos modelos ARIMA e Redes Neurais, utilizando variáveis explanatórias. O modelo tem por objetivo prever o valor exato do PLD 12 semanas a frente.

5.2.1 Seleção das variáveis explanatórias

Para seleção das variáveis explanatórias, foram utilizados o teste f ANOVA que indica a significância e importância de cada variável. Detalhes do teste f ANOVA já foram apresentados no Capítulo 3. Os resultados de significância e importância para cada submercado é apresentado na tabela 5.1. A análise de significância mostra que a geração hidráulica deve ser eliminada dos mercados do Norte e Nordeste. O mesmo teste de hipóteses

indica que a Energia Natural Afluente (ENA) deve ser removida do submercado Centro-Oeste/Sudeste e do submercado Sul.

Tabela 5.1 – Teste f ANOVA das variáveis explanatórias dos 4 submercados

	Região Norte		Região Nordeste		Região SE/CO		Região Sul	
	p-value	Imp	p-value	Imp	p-value	Imp	p-value	Imp
EARM	0.001	0,62	0.000	0,18	0.000	0,53	0.000	0,2
ENA	0.050	0,23	0.000	0,09	0.002	---	0.024	---
GerHid	0.263	---	0.116	---	0.000	0,05	0.000	0,09
GerTer	---	---	0.000	0,61	0.000	0,3	0.000	0,25
Carga	0.023	0,15	0.000	0,12	0.000	0,12	0.000	0,46

Os conjuntos de variáveis explanatórias selecionadas para cada submercado são:

- Norte: EARM, ENA e Carga;
- Nordeste: EARM, ENA, GerTer e Carga;
- Centro-Oeste/Sudeste: EARM, GerHid, GerTer, Carga;
- Sul: EARM, GerHid, GerTer e Carga.

Os parâmetros de cada algoritmo serão apresentados junto com os resultados, devido à grande variação dos mesmos para obtenção da maior correlação linear e menor erro médio absoluto.

5.2.2. Modelo ARIMA

O modelo ARIMA (Auto Regressivo Integrado de Média Móvel) resulta da combinação de três filtros: o componente Auto-regressivo (AR), o filtro de Integração (I) e o componente de Médias Móveis (MA). Este modelo é comumente representado pela notação ARIMA(p,d,q), sendo (p,d,q) a representação da ordem do modelo. Uma representação ARIMA(1,2,0), indica um modelo de ordem 1 para o componente AR (Auto-Regressivo), ordem 2 para o componente I (Integração ou diferenciação) e o último 0 para o componente MA (Média Móvel). Já os fatores grafados em letras maiúsculas se referem a:

- P = número de termos auto-regressivos sazonais;

- D = Número de diferenças sazonais;
- Q = Número de médias móveis sazonais.

Os modelos ARIMA (p,d,q)(P,D,Q) obtidos para cada uma das séries temporais são apresentados na Tabela 5.2

Tabela 5.2 - Modelos ARIMA obtidos para os 4 submercados

Série Temporal	Região			
	N	S	NE	SE/CO
GerHid	--	(0,1,1) (0,0,0)	--	(1,1,1) (0,0,0)
GerTer	--	(0,1,1) (0,0,1)	(0,1,1) (0,0,0)	(0,1,1) (1,0,0)
Carga	(0,1,1) (1,1,0)	(0,1,3) (1,1,0)	(0,1,2) (1,1,0)	(0,1,2) (1,1,1)
ENA	(1,1,1) (1,0,1)	--	(0,1,1) (1,0,1)	(1,1,1) (1,0,1)
EARM	(1,1,0) (0,1,0)	(1,1,0) (0,1,0)	(1,1,0) (0,1,0)	--
PLD	(0,1,1) (1,0,0)	(0,1,2) (1,1,0)	(0,1,1) (0,0,1)	(0,1,1) (1,1,0)

5.2.3 Topologia das Redes Neurais

De acordo com o modelo proposto e apresentado na Figura 4.3, a primeira Rede Neural - RN1 é utilizada para prever as variáveis explanatórias. Já a segunda Rede Neural – RN2 é utilizada para prever o PLD. A topologia de ambas as Redes Neurais foi escolhida de acordo com o melhor desempenho e melhor capacidade de generalização.

As Tabelas 5.3a, 5.3b, 5.3c e 5.3d apresentam as topologias das RNA utilizadas. Nas tabelas constam o número de neurônios nas camadas de entrada, camadas intermediárias e camadas de saída. Em todos os casos foi utilizado o algoritmo de treinamento de retropropagação do erro, de modo a ajustar os pesos da rede com objetivo de minimizar o erro entre as saídas reais e saídas previstas. O particionamento de dados foi de 70% para treino e 30% para teste. O critério de parada é o momento onde ocorre a taxa de erro mínima no conjunto de validação. A rede foi treinada com função de ativação do tipo tangente sigmoidal para todos os neurônios.

As tabelas 5.3a, 5.3b, 5.3c e 5.3d quantificam os neurônios em cada camada da RNA.

Tabela 5.3a - Topologias das RNA aplicadas ao submercado S

Submercado sul		Variável Preditada				
		EARM	GerTer	GerHid	Carga	PLD
Camadas	Entrada	5	5	5	5	5
	Intermediária 1	8	9	6	12	8
	Intermediária 2	5	5	3	8	5
	Saída	1	1	1	1	1

Tabela 5.3b - Topologias da RNA aplicadas ao submercado N

Submercado norte		Variável Preditada			
		EARM	ENA	Carga	PLD
Camadas	Entrada	5	5	5	5
	Intermediária 1	8	9	8	10
	Intermediária 2	5	4	3	5
	Saída	1	1	1	1

Tabela 5.3c - Topologias da RNA aplicadas ao submercado NE

Submercado nordeste		Variável Preditada				
		EARM	ENA	GerTer	Carga	PLD
Camadas	Entrada	5	5	5	5	5
	Intermediária 1	8	12	10	8	10
	Intermediária 2	3	4	4	3	3
	Saída	1	1	1	1	1

Tabela 5.3d - Topologias da RNA aplicadas ao submercado SE/CO

Submercado Sudeste/Centro-oeste		Variável Preditada				
		ENA	GerTer	GerHid	Carga	PLD
Camadas	Entrada	5	5	5	5	5
	Intermediária 1	11	10	12	13	11
	Intermediária 2	10	8	10	8	9
	Saída	1	1	1	1	1

5.2.4 Resultados para predição do PLD

O modelo híbrido proposto foi aplicado para predição do valor do preço de liquidação de diferenças – PLD a partir dos valores futuros preditos das variáveis explanatórias selecionadas para cada submercado conforme a análise de importância e significância.

A aplicação da Análise de Componentes Principais – PCA objetivou diminuir o número de entradas da Rede Neural. Assim, com o uso do PCA, as 5 variáveis explanatórias,

cada um com 5 amostras atrasadas no tempo, totalizando 25 sinais de entrada, são condensadas em apenas 5 sinais de entrada.

De modo a melhorar o desempenho do treinamento, utiliza-se o balanceamento de amostras de entrada, conforme descrito no capítulo 4. As figuras 5.1, 5.2, 5.3 e 5.4 apresentam a comparação entre os valores reais do PLD (preto) e os valores preditos pelo sistema híbrido (cinza).

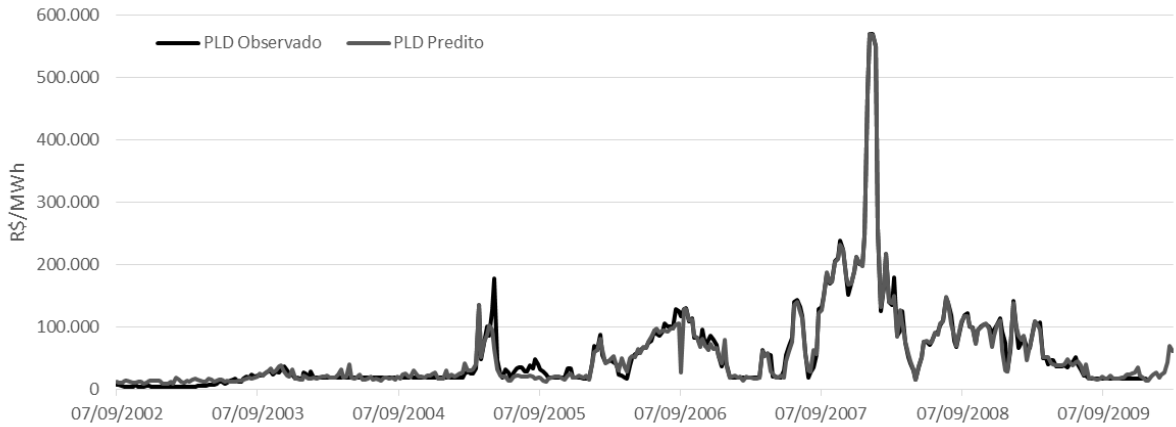


Figura 5.1 - Predição do PLD até $t+12$ com o modelo híbrido para a região S

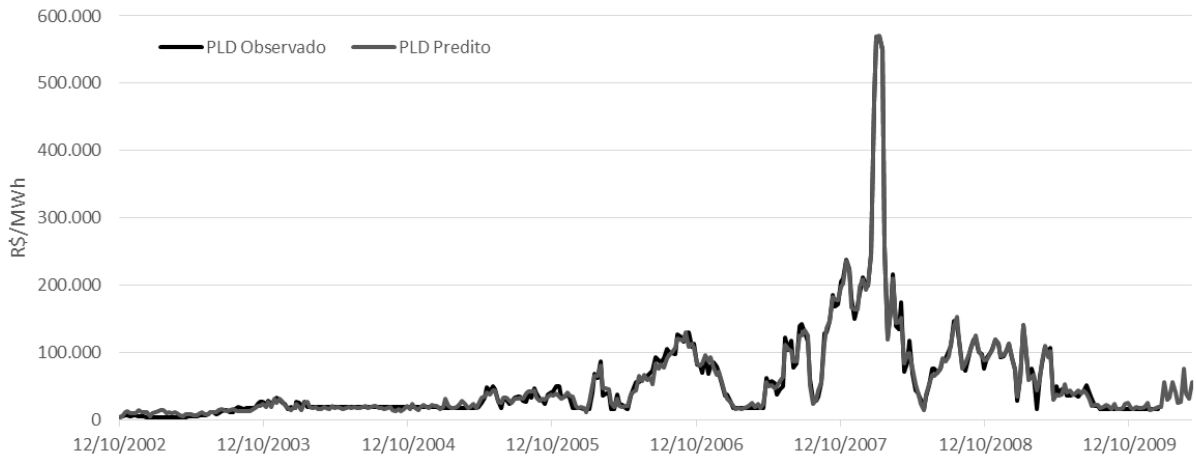


Figura 5.2 - Predição do PLD até $t+12$ com o modelo híbrido para SE/CO



Figura 5.3 - Predição do PLD até t+12 com o modelo híbrido para a região NE

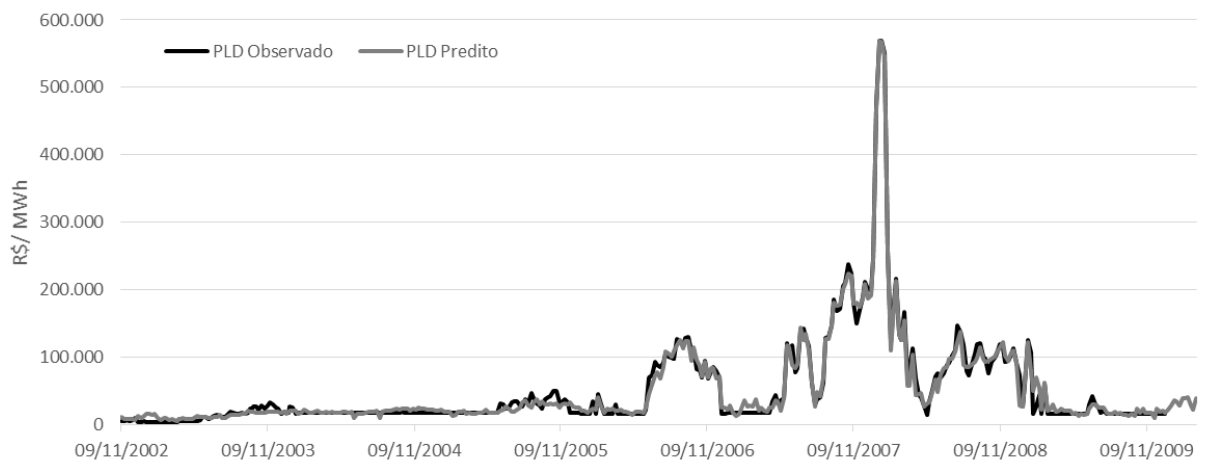


Figura 5.4 - Predição do PLD até t+12 com o modelo híbrido para a região N

Alguns critérios comumente utilizados para avaliar a precisão da previsão do PLD são empregados neste trabalho: Raiz do Erro Médio Quadrático (REMQ), Erro Médio Absoluto (EMA) e Erro Médio Percentual Absoluto (EMPA). Maiores detalhes podem ser vistos em Wu e Shahidehpour (2010):

$$REMQ = \sqrt{\frac{1}{N} \sum_{i=1}^N (P_i^{real} - P_i^{prev})^2} \quad (5.1)$$

$$EMA = \frac{1}{N} \sum_{i=1}^N |P_i^{real} - P_i^{prev}| \quad (5.2)$$

$$EMPA = \frac{1}{N} \sum_{i=1}^N \frac{|P_i^{real} - P_i^{prev}|}{P_{MED}} \times 100\% \quad (5.3)$$

$$PMED = \frac{1}{N} \sum_{i=1}^N P_i^{real} \quad (5.4)$$

Onde, N é o número de amostras, P_i^{real} é o valor real do PLD e P_i^{prev} é o preço predito do PLD.

A figura 5.5 apresenta os resultados obtidos de REMQ, EMA e EMPA com a predição do PLD 12 semanas à frente para as regiões Norte, Nordeste, Sul, Centro-Oeste/Sudeste utilizando o modelo híbrido proposto, bem como outras técnicas conhecidas como ARIMA, suavização exponencial, GARCH e RNA.

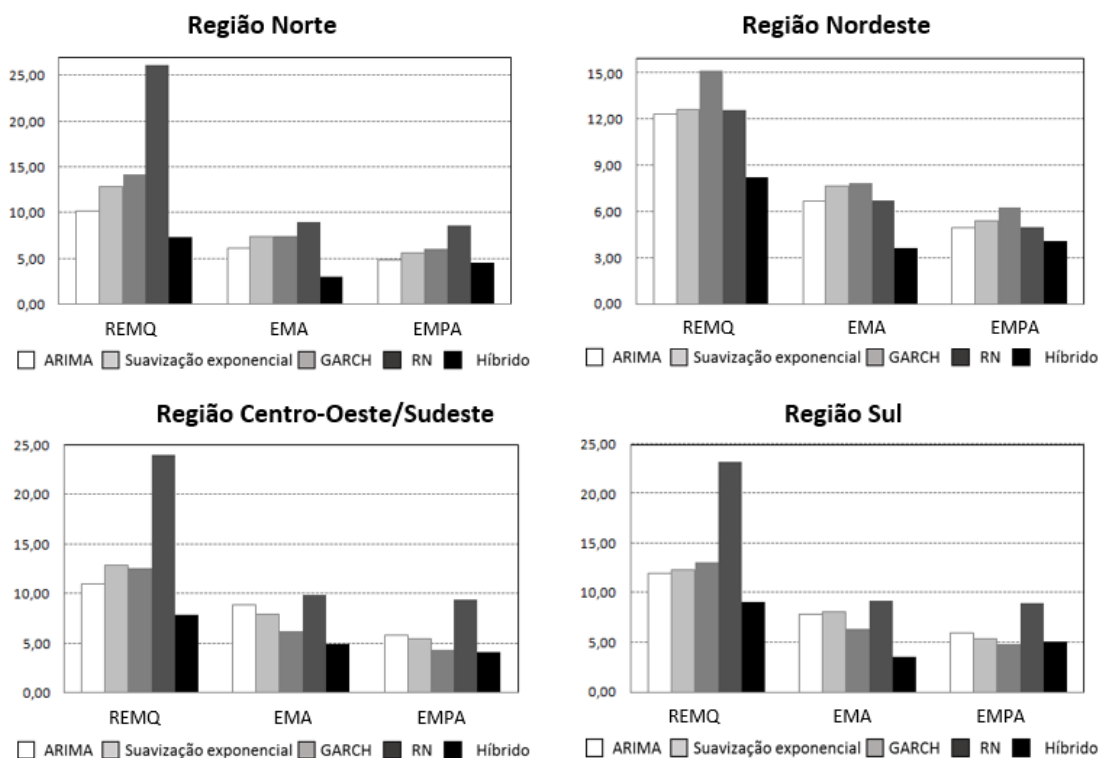


Figura 5.5 - Indicadores de erro para cada modelo nos 4 submercados

Na figura 5.5 observa-se que para todas as regiões o modelo híbrido apresenta os melhores indicadores dos erros REMQ, EMA e EMPA. As demais técnicas apresentam bons

resultados isoladamente, mas em nenhum caso se repete a constância de bons resultados do modelo híbrido.

A tabela 5.4 apresenta o quadro comparativo entre os métodos ARIMA, suavização exponencial, GARCH, RNA e híbrido, com os resultados de correlação linear (CL) e desvio padrão (DP) comparado entre séries reais e preditas dos Preços de Liquidação de Diferenças (PLD) dos 4 submercados brasileiros.

Tabela 5.4 - Correlação linear (CL) e Desvio Padrão (DP) do preço PLD

	N		NE		SE/CO		S	
	CL	DP	CL	DP	CL	DP	CL	DP
ARIMA	0,96	33,63	0,90	34,35	0,96	32,08	0,92	25,36
SUAV. EXP.	0,95	37,00	0,89	37,18	0,96	33,94	0,92	28,50
GARCH	0,75	22,25	0,93	23,67	0,88	18,63	0,70	19,80
RNA	0,77	42,50	0,85	22,00	0,81	34,17	0,69	32,05
HÍBRIDO	0,99	14,72	0,97	19,75	0,99	11,30	0,98	14,75

De modo a fazer um *backtesting* no modelo híbrido, foi escolhido o teste de *Kupiec* proposto em Kupiec (1995). *Backtesting* é um jargão usado em indústrias financeiras para se referir a testar um modelo de previsão com base em dados históricos existentes. Trata-se de um tipo especial de validação cruzada aplicada a dados de séries temporais. A tabela 5.5 apresenta o quadro comparativo entre os métodos ARIMA, suavização exponencial, GARCH, RNA e híbrido, com os resultados da estatística LRPOF dos 4 submercados brasileiros.

Tabela 5.5 - Resultados do teste de Kupiec para previsões t+12

Região	Estatística LRPOF				
	ARIMA	GARCH	RNA	Suaviza. Exp.	Híbrido
Norte	15.48	18.39	5.82	7.85	0.298
Nordeste	12.74	24.69	4.78	10.19	1.108
SE/CO	18.39	28.05	3.83	7.85	1.108
Sul	12.74	15.48	2.97	18.39	0.298

Os resultados apresentados na Tabela 5.6 do teste de *Kupiec* (LRPOF) mostram que a técnica híbrida proposta é o melhor modelo para todas as regiões uma vez que os valores obtidos estão dentro do número de violações aceitos para este tipo de *backtesting*. Os piores

modelos são o GARCH para as regiões norte, nordeste e centro-oeste/sudeste. Já para a região sul, o pior modelo é o de suavização exponencial.

O comportamento do modelo híbrido proposto também foi testado em horizontes temporais superiores. A Tabela 5.6 mostra os valores de EMPA para 12, 24 e 36 semanas à frente para o modelo híbrido proposto, o modelo ARIMA e RNA para efeito de comparação. Os resultados mostram que o modelo híbrido apresenta menor erro para todas as regiões. Além disso, o desempenho dos modelos comparados se deteriora conforme o horizonte temporal de previsão aumenta. É importante mencionar que os melhores resultados foram obtidos considerando um horizonte temporal de previsão inferior a 12 semanas à frente. No entanto, o horizonte de tempo de 12 semanas está à frente mais apropriado para arriscar práticas de gestão no mercado brasileiro.

Tabela 5.6 - Critério EMPA horizontes superiores de predição futura

Região	12-semanas			24-semanas			36-semanas		
	ARIMA	RNA	Híbrido	ARIMA	RNA	Híbrido	ARIMA	RNA	Híbrido
Norte	5,879	8,201	4,481	17,839	20,000	9,353	20,431	23,115	16,387
Nordeste	11,851	10,759	4,122	14,215	19,813	9,148	23,930	27,717	15,885
SE/CO	10,666	11,838	4,066	17,363	17,597	9,253	28,237	22,694	15,274
Sul	10,600	11,770	5,111	19,032	20,591	11,922	23,400	24,235	17,838

5.3 Resultados obtidos utilizando o modelo de classificação

Esta seção apresenta os resultados obtidos para previsão do PLD utilizando um modelo de classificação, onde tem-se por objetivo prever a qual classe o PLD pertence, exemplo: muito baixa, baixa, media, alta ou muito alta. Neste caso, o valor exato do PLD não é o objetivo do método.

5.3.1 Definição das classes do PLD

A definição dos valores de PLD em 5 classes foi feita a partir da avaliação do preços do MWh para consumidores cativos, assim como observação dos preços praticados em outros mercados de energia de porte similar ao Brasil. A Figura 5.6 apresenta o histograma dos valores do PLD separados pelas 5 classes de valores.

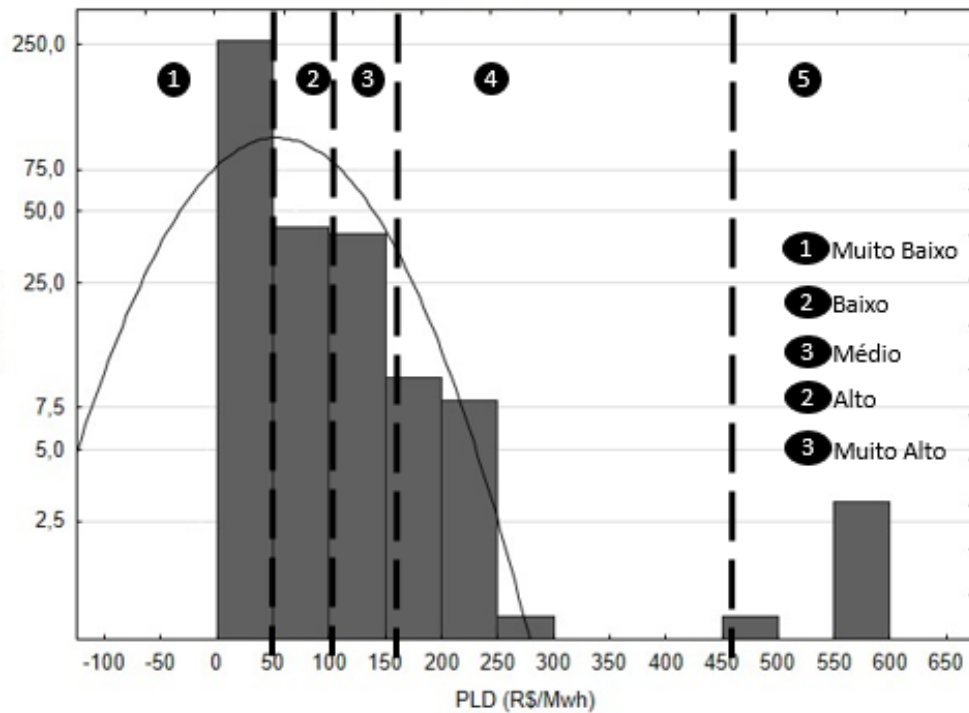


Figura 5.6 - Patamares dos valores do PLD

Essas classes são rotuladas de acordo com as faixas de preço: muito baixo, baixo, médio, alto e muito alto. Foram adotados quatro limiares de preços por megawatt-hora: T1 = R\$ 50,00; T2 = R\$ 90,00; R\$ 150,00 e T4 = R \$ 450,00. Então, cinco classes foram criadas com base nesses limites:

- Classe 1 (preço muito baixo): abaixo de T1;
- Classe 2 (preço baixo): entre T1 e T2;
- Classe 3 (preço médio): entre T2 e T3;
- Classe 4 (preço alto): entre T3 e T4;
- Classe 5 (preço muito alto): acima de T4;

5.3.2 Seleção das Variáveis Explanatórias

No pré-processamento de dados do modelo classificador, são novamente selecionadas as variáveis explanatórias que mais contribuem com a classificação do valor futuro da energia elétrica.

Na região Norte, por exemplo, trabalha-se com as 5 variáveis explanatórias em potencial atrasadas no tempo em até 52 semanas (1 ano). Como resultado, tem-se que para a classificação futura da região norte o valor mais relevante com 19% de importância é a carga

de energia de 28 semanas antes da classificação do PLD, seguido da energia armazenada da semana anterior a classificação (18%), da carga de energia de 36 semanas atrás e de uma semana atrás e da energia natural afluyente de 5 semanas atrás e da semana imediatamente anterior a classificação. A figura 5.7 ilustra as variáveis explanatórias selecionadas para a região Norte:

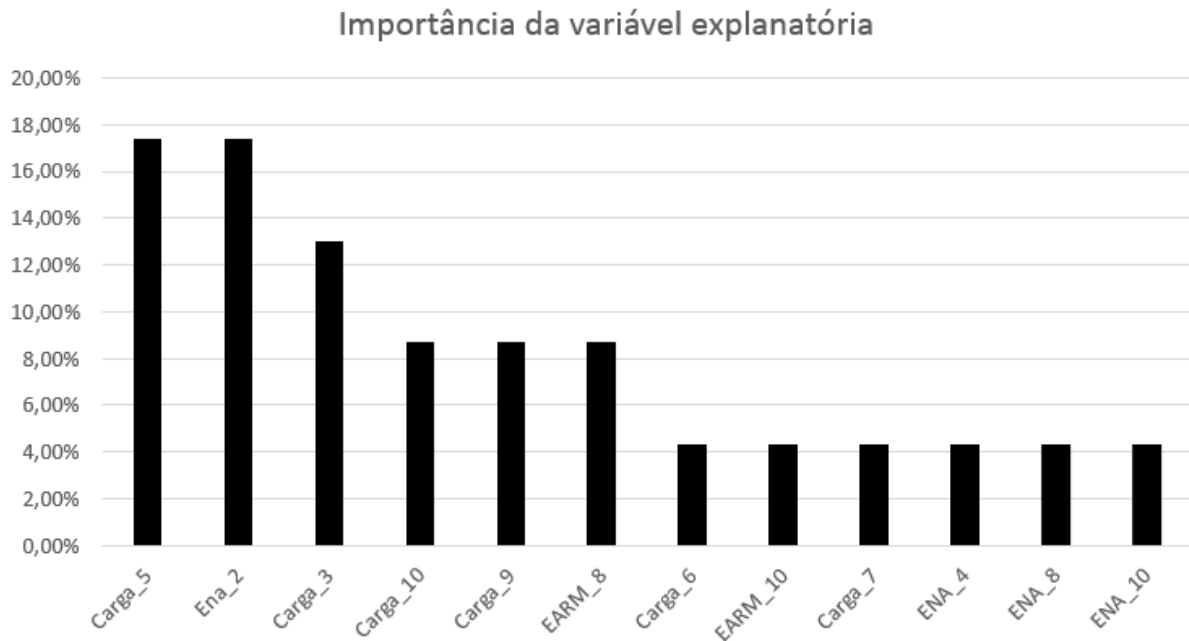


Figura 5.7 - Variáveis explanatórias selecionadas

5.3.3 Resultados para classificação do PLD

Para os 4 submercados foram criados modelos classificadores do PLD futuro. Os algoritmos utilizados para indução da árvore de decisão foram o CART e C5.0. Os resultados obtidos utilizando estes algoritmos foram comparados com resultados obtidos utilizando a Rede Neural.

De acordo com o trabalho de TAN et al (2009), optou-se para a avaliação do desempenho dos algoritmos de classificação dos critérios de erro médio percentual de classificação (EMPC) e acurácia percentual de classificação (APC), dados por:

$$EMPC = \frac{N_{err}}{N_{tot}} \times 100 \quad (5.5)$$

$$APC = \frac{N_{tot} - N_{err}}{N_{tot}} \times 100 \quad (5.6)$$

Onde N_{tot} se refere ao total de amostras e N_{err} se refere a quantidade de amostras que foram classificadas de forma errada.

A tabela 5.7 apresenta os resultados obtidos com os modelos de classificação para os 4 submercados brasileiros. Nota-se um melhor desempenho com o modelo C5.0.

Tabela 5.7 - Indicadores EMPC e APC para 3 modelos classificadores

Região	Norte		Nordeste		Sul		Centro-Oeste/Sudeste	
	EMPC(%)	APC(%)	EMPC(%)	APC(%)	EMPC(%)	APC(%)	EMPC(%)	APC(%)
RN	14,11%	85,88%	14,36%	84,56%	16,91%	82,73%	15,70%	84,36%
CART	15,83%	84,16%	17,84%	82,18%	18,49%	81,48%	15,92%	84,11%
C5.0	1,20%	98,79%	1,26%	97,85%	1,97%	97,25%	1,46%	98,55%

De modo a auxiliar na avaliação da capacidade de classificação para 1, 2, 3, 4 e 5 semanas a frente, a tabela 5.8 apresenta os resultados dos modelos classificadores quanto ao critério do erro médio percentual de classificação (EMPC) para a região Norte. Novamente, nota-se um melhor desempenho com o modelo C5.0.

Tabela 5.8 - EMPC para até 5 passos a frente de classificadores da região N

Região Norte	EMPC				
	1 passo	2 passos	3 passos	4 passos	5 passos
RNA	14,11 %	16,17 %	19,79 %	21,34 %	21,51 %
CART	15,83 %	17,38 %	18,24 %	18,37 %	19,11 %
C5.0	1,20 %	2,06 %	2,34 %	2,94 %	3,27 %

Com o intuito de obter melhores resultados, foi realizada a combinação dos modelos RNA, CART e C5.0 dois a dois, utilizando a estratégia de votação *Weighted Majority Voting* (WMV). Assim, foram analisados os seguintes casos:

- RNA + CART;
- RNA + C5.0;
- C5.0 + CART.

A Tabela 5.9 apresenta os resultados obtidos com a aplicação destes modelos para a região norte. Nota-se que o modelo com melhor desempenho para todas as regiões foi o que combina os algoritmos C5.0 e CART.

Tabela 5.9- Resultado modelos combinados 2 a 2 para a região N

Região	Norte		Nordeste		Sul		Centro-Oeste/Sudeste	
Modelo	EMPC	APC	EMPC	APC	EMPC	APC	EMPC	APC
RN + CART	8,43%	91,56%	8,83%	91,22%	8,64%	91,41%	9,59%	91,04%
RN + C5.0	1,03%	98,96%	1,48%	98,02%	1,32%	98,18%	1,65%	97,93%
C5.0 + CART	0,68%	99,31%	0,72%	99,12%	0,65%	99,27%	0,81%	99,06%

A matriz de confusão para a região norte do melhor modelo combinado (C5.0 + CART) é apresentada na Tabela 5.10. Pode-se observar pela Tabela 5.11 o excelente poder classificador do método para todas as classes, mesmo para as classes com poucas amostras.

Tabela 5.10 - Matriz de confusão do modelo C50 + CART para região N

	Muito Baixo	Baixo	Médio	Alto	Muito Alto
Muito Baixo	342	0	0	2	0
Baixo	1	50	0	0	0
Médio	1	1	76	5	0
Alto	1	0	0	95	0
Muito alto	0	0	0	0	7

Enquanto explicitação do conhecimento, as regras que merecem maior destaque são as que se referem aos valores **Alto** e **Muito Alto**. Isso porque nem todos os participantes do mercado precisam saber o valor exato de preços futuros em seu processo de tomada de decisão. Em alguns casos passa a ser mais interessante a utilização de limiares de preços específicos, como base para a tomada de decisões de escalonamento de curto prazo. Os limiares que determinam os preços **Alto** e **Muito Alto** são os mais significativos neste caso, porque denotam de forma clara os cenários onde não vale a pena comprar energia no mercado de curto prazo. A figura 5.8 apresenta as regras para os valores muito altos da região norte. Já a figura 5.9 apresenta as regras para valores altos da mesma região.

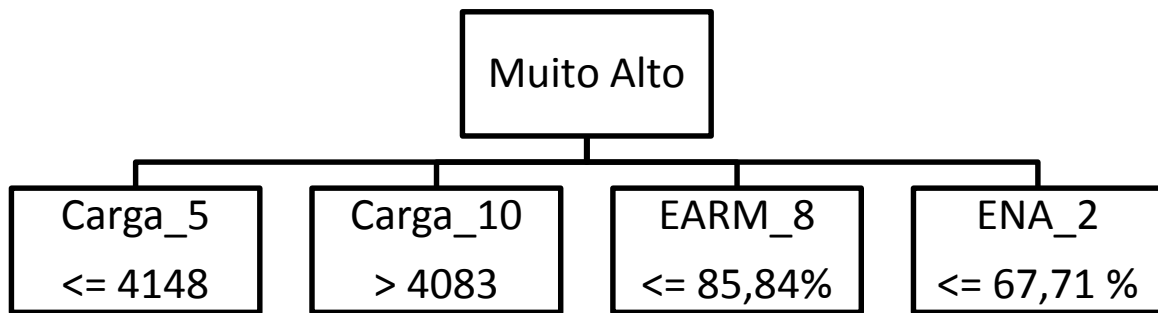


Figura 5.8 - Regras para valores muito altos de PLD para a região N

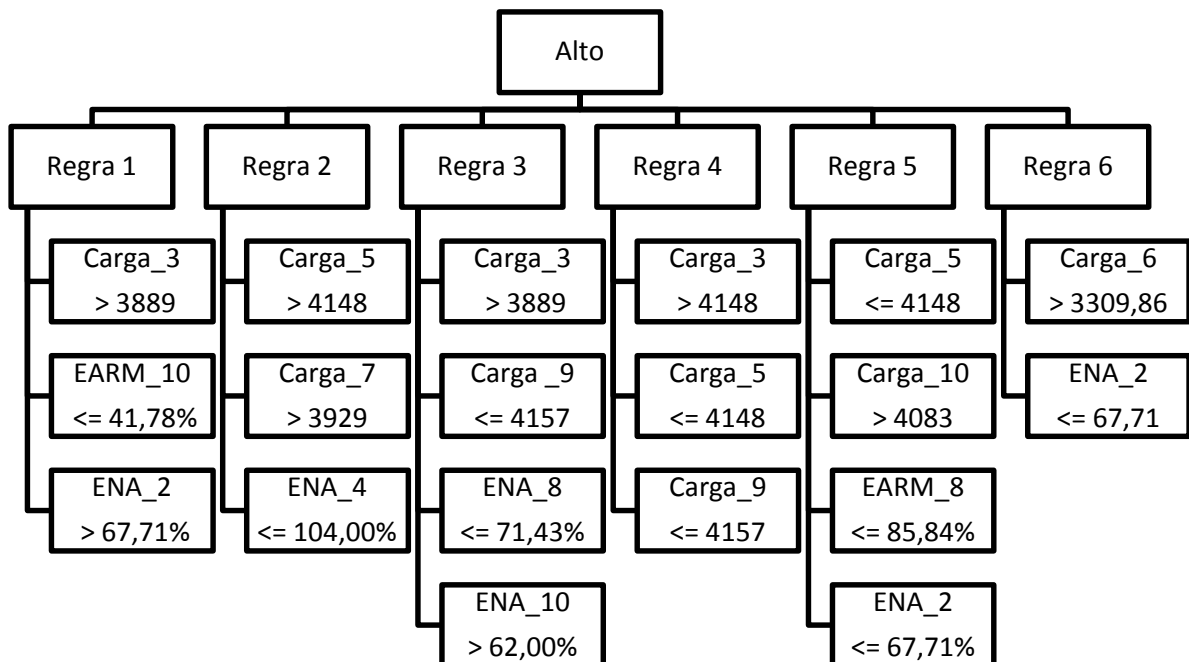


Figura 5.9 - Regras para valores altos de PLD para a região N

5.4 Considerações finais

Este capítulo apresentou uma análise experimental dos modelos preditor híbrido ARIMA + RNA e classificador futuro. Os resultados obtidos com o modelo híbrido foram comparados com o uso das RNA, ARIMA, GARCH, Suavização Exponencial com o objetivo de predição até 12 passos à frente. O uso da predição de 12 passos à frente se deve a necessidade de se criar cenários do preço PLD em até 3 meses à frente, permitindo a elaboração de uma estratégia de compra/venda de energia no mercado de curto prazo. Os

resultados superiores do modelo híbrido se dão por haver em cada sinal modelado uma parte linear e outra não-linear, que o uso conjunto do ARIMA e das RNA consegue mapear, o que não é obtido com o uso isolado dos dois métodos de predição futura.

Os resultados do modelo classificador são muito precisos, principalmente para 1 passo à frente. A análise dos resultados demonstra que os modelos combinados classificadores são alternativas mais eficientes que os classificadores atuando de maneira isolada. As árvores de decisão apresentam possíveis regras de formação de preços do PLD, em especial para valores altos e muito altos.

CAPÍTULO VI

CONCLUSÕES E TRABALHOS FUTUROS

6.1. Conclusões e recomendações para trabalhos futuros

Os processos de compra e venda de energia se dão em função de várias condições que vão desde o mapeamento dos riscos inerentes ao mercado de eletricidade, até o cálculo de indicadores que orientam o processo de tomada de decisão. Neste trabalho foram preparadas as condições para criar cenários e para medir quais as combinações de fatores que geram maior volatilidade no Preço de Liqueidação de Diferenças - PLD, e assim propor uma estratégia de gerenciamento de risco eficaz para participantes do mercado de curto prazo de energia elétrica no Brasil.

Da revisão bibliográfica do estado da arte acerca da formação de preços no mercado de curto prazo de energia elétrica no Brasil, constatou-se que existem poucos estudos sobre o mercado brasileiro, que abranjam as peculiaridades da formação de preços e estratégias de mitigação de riscos existentes no modelo da Câmara de Comercialização de Energia Elétrica – CCEE. Além disso, apesar do grande número de modelos utilizados para predição do preço de energia elétrica no curto prazo nos mercados internacionais, não se tem publicado nenhum trabalho comparativo que avalie as variáveis exploratórias e a explicitação do conhecimento da formação de regras de determinação dos preços. O entendimento das regras de formação de preços em conjunto com um modelo de predição futura do preço PLD pode auxiliar na elaboração de estratégias de mitigação de riscos para os agentes da CCEE.

Nesta tese, uma combinação híbrida de filtro ARIMA e RNA é proposta para predição do Preço de Liquidação de Diferenças – PLD usando como variáveis explanatórias a geração hidráulica, geração térmica, carga de energia, energia natural afluyente e energia armazenada para os 4 submercados brasileiros, a saber: norte, sul/centro-oeste, nordeste e norte. Em uma primeira a fase, o modelo prediz os valores futuros das variáveis explanatórias que afetam o valor do preço do mercado de curto prazo de energia. O modelo considera uma predição multi-passos à frente (12 semanas) e é aplicado ao mercado brasileiro de eletricidade. Os resultados obtidos com a metodologia proposta são comparados com técnicas tradicionais, a exemplo do ARIMA, RN, GARCH e suavização exponencial, e o modelo híbrido responde melhor que estas técnicas. Além disso, os resultados obtidos com a metodologia híbrida aplicada ao mercado brasileiro apresentam exatidão suficientemente boas quando comparadas com outros métodos de predição reportados na literatura existente e aplicados a outros mercados.

Uma vez que os critérios de erro não captam as perdas de cauda (*tail loss*), esta tese também avalia a performance da metodologia proposta usando o teste de cobertura incondicional de perda de cauda desenvolvido por *Kupiec*. Os resultados apontam que a metodologia híbrida é o melhor modelo diante deste critério. Os resultados considerando horizontes de tempo superiores também foram analisados e a técnica proposta apresenta performance superior a técnicas clássicas como a RN e ARIMA para 24 e 36 passos a frente. Desse modo, esta técnica pode ser uma ferramenta importante para auxiliar os participantes do mercado brasileiro para reduzir riscos na compra e venda de energia.

Em caráter complementar, foi proposto um modelo classificador dos preços futuros da energia elétrica, através do uso combinado de algoritmos de árvores de decisão e de redes neurais. Estratégias de *bagging* e *boosting* foram utilizadas para potencializar o poder classificador das árvores de decisão, assim como técnicas de combinação de algoritmos através de estratégias de votação. O modelo classificador se mostrou bastante efetivo na classificação futura de valores considerados altos e muito altos e possibilita a compreensão das condições que criam tais valores a partir das variáveis explanatórias. Com isso explicita-se o conhecimento das regras que levam os valores do PLD a se tornarem elevados, a ponto de não valer a pena a compra de energia no mercado de curto prazo. Analisou-se o uso da técnica para horizontes de tempo de 1 a 5 semanas com resultados bastante significativos. A técnica classificadora, a exemplo do modelo preditivo, pode ajudar na gestão de riscos dos *players* do mercado de curto prazo.

É importante mencionar que as metodologias propostas nesta tese, foram desenvolvidas com base nas características únicas do mercado brasileiro de eletricidade, e os preços preditos/classificados são específicos para os limites deste mercado. Até este momento, não foi testado o nível de adaptação dos algoritmos a predição ou classificação do preço de energia do mercado de eletricidade de outros países. Em trabalhos futuros, pretende-se investigar a performance destas metodologias em outros mercados de energia de modo a criar uma abordagem com maior capacidade de generalização.

REFERÊNCIAS

- AGGARWAL, S. K.; SAINI, L.M.; KUMAR, A. Short term price forecasting in deregulated electricity markets. *International journal of energy sector management*. v. 3, p. 333-358, 2009a.
- AGGARWAL, S.K.; SAINI, L.M.; KUMAR, A. Electricity price forecasting in deregulated markets: A review and evaluation. *International Journal of Electrical Power & Energy Systems*, v.31, p.13-22, 2009b.
- AGUIRRE, L. A. Introdução à identificação de sistemas: técnicas lineares e não-lineares aplicadas a sistemas reais. Belo Horizonte: UFMG, 2000. 735p.
- AMJADY, N. Day-ahead price forecasting of electricity markets by a new fuzzy neural network. *IEEE transactions on power systems*, v.21, p. 867-896, 2006.
- AMJADY, N., HEMMATI, M. Day-ahead price forecasting of electricity markets by a hybrid intelligent system. *IEEE transactions on power systems*, v.19, p.89-102, 2009.
- AMJADY, N.; KEYNIA, F. Day ahead price forecasting of electricity markets by a mixed data model and hybrid forecast method, *International Journal of Electrical Power & Energy Systems*, v. 30, p. 533–546, 2008.
- AMRAEE, T.; RANJBAR, S. Transient Instability prediction using decision tree technique. *IEEE Transactions on power systems*, v.28, p. 3028-3037, 2013.
- ANBAZHAGAN, S.; KUMARAPPAN, N. Day-Ahead Deregulated Electricity Market Price Forecasting Using Recurrent Neural Network. *IEEE systems journal*, v. 7, p. 866-872, 2013.
- ANBAZHAGAN, S.; KUMARAPPAN, N. Day-ahead deregulated electricity market price classification using neural network input featured by DCT. *Electrical Power and Energy Systems*, v.37, p. 103–109, 2012.
- ANDERSON, C.L. A hybrid model for electricity spot price. 2004. 153 f. Tese (Doutorado em Engenharia Elétrica) - University of Western Ontario, Ontário, 2004.
- BANFIELD, E.; HALL, L.O.; BOWYER, K. W.; KEGELMEYER, W.P. A comparison of decision tree ensemble creation techniques. *IEEE transactions on pattern analysis and machine intelligence*, v. 29, p.173-180, 2007.
- BARROSO, L.A.; STREET, A.; GRANVILLE, S.; PEREIRA, M.V. Offering strategies and simulation of multi-item iterative auctions of energy contracts. *IEEE transactions on power systems*, v. 26, p.1917-1928, 2011.

- BASGALUPP, M. P. LEGAL-Tree: um algoritmo genético multi-objetivo lexicográfico para a indução de árvores de decisão. 2010. 116 f. Tese (Doutorado em Matemática) – Instituto de Ciências Matemáticas e Computação. São Carlos: USP, 2010.
- BATLLE, C.; BARQUÍN, E. J. A strategic cost production model for electricity market price analysis. *IEEE Transactions on Power Systems*, v. 20, p. 657-74, 2005.
- BOWDEN, N.; PAYNE, J. Short term forecasting of electricity prices for MISO hubs: evidence from ARIMA-EGARCH models. *Energy Economics*, v.30, p.3186-3197, 2008.
- BRAGA, A. DE P.; LUDEMIR, T. B.; CARVALHO, A. T. L. F. Redes Neurais. Artificiais: teoria e aplicações. São Paulo: LTC, 2000. 262 p.
- BREIMAN, L.; FRIEDMAN, J. H.; OLSHEN, R. A.; STONE, C. J. Classification and Regression Trees. New York: Chapman and hall, 1984. 368 p.
- BROCKWELL, P. J.; DAVIS, R. A. Introduction to times series and forecasting. New York: Springer, 2002. 437 p.
- BURGER, M.; KLAR, B.; MÜLLER, A.; SCHINDLMAYR, E. G. A spot market model for pricing derivatives in electricity markets. *Quantitative finance*, v.4, p. 109-122, 2004.
- CARPIO, L. G. T.; PEREIRA JR., A. O. Economical efficiency of coordinating the generation by subsystems with the capacity of transmission in the brazilian market of electricity. *Energy Economics*, v. 29, p. 454-466, 2007.
- CATALÃO, J.P.S.; POUSINHO, H.M.I.; MENDES, V.M.F. Hybrid Wavelet-PSO-ANFIS Approach for Short-Term Electricity Prices Forecasting. *IEEE Transactions on Power Systems*, v.26 , p. 137-144, 2011.
- CHEN, X.; DONG, Z.Y.; MENG, K.; XU, Y.; WONG, K.P., NGAN, H.W. Electricity price forecasting with extreme learning machine and bootstrapping. *IEEE transactions on power systems*, v. 27, p. 2055-2062, 2012.
- CONEJO, A.J.; PLAZAS, M.A.; ESPÍNOLA, R.; MOLINA, A.B. Day-ahead electricity price forecasting using the wavelet transform and ARIMA models. *IEEE Transactions on Power Systems*, v.20, p. 1035-1042 ,2005.
- CONGALTON, R. G. A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment*, v. 49, p. 1671-1678, 1991.
- CONGALTON, R. G., GREEN, K. Assessing the Accuracy of Remotely Sensed Data: Principles and Practices. New York: Lewis Publishers, 1999. 137p.
- CONTRERAS, J.; ESPÍNOLA, R.; NOGALES, F.J.; CONEJO, A.J. ARIMA Models to Predict Next-Day Electricity Prices. *IEEE Transactions on Power Systems*, v.18, p. 1014-1020, 2003.

- DA SILVA, A.L.; DA SILVA, P.S.; BREDA, J.C.; NEPOMUCENO, J. Combined pool/bilateral short-term hydrothermal scheduling model for day-ahead energy markets. IEEE latin america transactions, v. 10, p. 2094-2104, 2012.
- DAVISON, M.; ANDERSON, C.L.; MARCUS, B.; ANDERSON, E. K. Development of a hybrid model for electric power spot prices. IEEE Transactions on Power Systems, v.17, p.257-264, 2002.
- ESPOSITO, F.; MALERBA, D.; SEMERARO, G.; KAY, J. A comparative analysis of methods for pruning decision trees. IEEE transactions on pattern analysis and machine intelligence, v. 19, p. 476-491, 1997.
- FAN, S.; MAO, C.; CHEN, L. Next-day electricity price forecasting using a hybrid network. IET Generation, Transmission & Distribution, v.1, p. 176-182, 2007.
- FANGMING YE; ZHAOBO ZHANG, Z.; CHAKRABARTY, K. XINLI GU. Board-Level Functional Fault Diagnosis Using Artificial Neural Networks, Support-Vector Machines and Weighted-Majority Voting. IEEE Computer-Aided Design of Integrated Circuits and Systems. v. 32, p. 723-736, 2013.
- FRANCIS, L. The basics os neural network demystified. American Academy of Actuaries Contingencies, p.56-61, novembro de 2001.
- FREIRE, L.M. ; NEVES, E.M.A. ; TSUNECHIRO, L.I. ; CABRAL, R. ; SOUZA, Z. Liquidity in the Brazilian electricity Market. 9th International Conference on the European Energy Market – EEM, Florença, p. 1 – 8, 2012.
- FREUND, Y.; SCHAPIRE, R. Experiments with a New Boosting Algorithm. Proceedings of 13th Nat’l Conf. Machine Learning. p.148-156, 1996.
- GALAR, M.; FERNÁNDEZ, A.; BARRENCHEA, E.; BUSTINCE, H.; HERRERA, F. A Review on Ensembles for the Class Imbalance Problem: Bagging, Boosting, and Hybrid-Based Approaches. IEEE transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, v. 42, p. 463-484, 2012.
- GARCIA, R.C.; CONTRERAS, J.; VAN AKKEREN, M.; GARCIA, J.B.C. GARCH forecasting model to predict day-ahead electricity prices. IEEE transactions on power systems, v. 2, p. 867-874, 2005.
- GARCÍA-MARTOS, C.; RODRÍGUEZ, J.; SÁNCHEZ, M.J. Mixed models for short-run forecasting of electricity prices: application for the Spanish market. IEEE Transactions on Power Systems, v. 22, p. 544-552, 2007.
- GARETA, R.; ROMEO, L. M.; GIL, E A. Forecasting of electricity prices with neural networks. Energy conversion and management, v.47, p. 1770-1778, 2006 .

- GONZALES, A.M.; SAN ROQUE, A.M; GARCIA-GONZALES, J. Modeling and forecasting electricity prices with input/output hidden Markov models. *IEEE Transactions on power systems*, v.20, p.13-24, 2005.
- GRANGER, C. W. J.; NEWBOLD, P. *Forecasting Economic Time Series*. New York : Academic Press, 1977. 262 p.
- GUYON, I.; ELISSEEFF, A. An introduction to variable and feature selection. *Journal of machine learning research*, v. 3, p. 1157–1182, 2003
- HAIBO HE; GARCIA, E.A. Learning from imbalanced data. *IEEE transactions on knowledge and data engineering*, v.21, p. 1263-1284, 2009.
- HAIBO HE; YANG BAU; GARCIA, E.A. SHUTAO LI. ADASYN : Adaptative synthetic sampling approach for imbalanced learning. *IEEE joint conference on computational intelligence*, p. 1322-1328, 2008.
- HAYKIN, S. *Redes neurais artificiais: princípios e prática*. Porto Alegre: Bookman, 2007. 900 p.
- HUANG, D.; ZAREIPOUR, W.; ROSEHART, D.; AMJADY, N. Data Mining for Electricity Price Classification and the Application to Demand-Side Management. *IEEE Transactions on Smart Grid*, v.3, 2012.
- HULLERMEIER, E., VANDERLOOY, S. Why decision trees are good rankers. *IEEE Transactions on Fuzzy Systems*, v.17, p. 1233-1244, 2009.
- JEONG, Y. S.; KANG, I.H.; JEONG, M.K.; KONG, D. A new feature selection method for one-class classification problems. *IEEE transactions on systems, man and cybernetics – part C: applications and reviews*, v. 42, p. 1500-1509, 2012.
- KIMOTO, T.; ASAWAKA, K.; YODA, M; TAKEOKA, M. Stock Market Prediction System with Modular Neural Networks. *Proceedings of the IEEE International Joint Conference on Neural Networks*, v. 1, p. 1-6, 1990.
- KOHAVI, R.; JOHN, G. Wrappers for feature selection. *Artificial intelligence*, v. 97, p. 273-324, 1997.
- KUPIEC, P. Techniques for verifying the accuracy of risk measurement models. *Journal of derivatives*, v. 3, p. 73-84, 1995.
- LEVY, R. A. *The Relative Strength Concept of Common Stock Price Forecasting*. New York: Investors Intelligence Inc, 1969. 318 p.
- LIN, W.M.; HONG, C.M.; HUANG, C.H. OU, T. C. Hybrid Control of a Wind Induction Generator Based on Grey–Elman Neural Network. *IEEE Transactions on Control Systems Technology*, v.21, p. 2367-2373, 2013.

- LITTLESTONE, N.; WARMUTH, M. The weighted majority algorithm. *Journal Information and computation*, v. 108, p. 212-261, 1994.
- LIU, W.; CHUNG, C. Y.; WEN, F. Multifractal based return interval approach for short-term electricity price volatility risk estimation. *IET Generation, Transmission & Distribution*, v.8, p. 1550-1560, 2014.
- LO, A.; MACKINLEY, A. C. Stock Prices do not follow Random Walks : Evidence from a simple specification test. *The Review of Financial Studies*, v.1, p. 41-66, 1988.
- LORA, A. T.; SANTOS, J. M. R.; EXPÓSITO, A.G.; RAMOS, J.L.M.; SANTOS, E.J.C.R. Electricity market price forecasting based on weighted nearest neighbors technique. *IEEE Transactions on Power Systems*, v.22, p.1294-1301, 2007.
- LU, X., DONG, Z.Y., LI, X. Electricity market price spike forecast with data mining techniques. *Electric power systems research*. v. 1, p. 19-29, 2005.
- MAENHOUDT, M.; DECONINCK, G. Detailed modelling of thermal units from a price taker's perspective, *Proceedings 9th International conference of european energy markets* , p. 1-8, 2012.
- MAENHOUDT, M.; DECONINCK, G. Strategic offering to maximize day-ahead profit by hedging against an infeasible Market clearing result. *IEEE transactions on power systems*, v. 29, p. 854- 862, 2014.
- MAKRIDAKIS, S.; WHEELWRIGHT, S. C.; HYNDMAN, R.J. *Forecasting methods and applications*. New Jersey: John Wiley & Sons, 1998. 656 p.
- MANDAL, P.; SENJYU, T.; FUNABASHI, T. Neural networks approach to forecast several hour ahead electricity prices and loads in deregulated market. *Energy conversion and management* 47, p. 2128-2142, 2006 .
- MANLY, B.F.J. *Métodos estatísticos multivariados: uma introdução*. Porto Alegre: Bookman, 2008. 3ª ed. 230 p.
- MANWANI, N.; SASTRY, P.S. Geometric Decision Tree. *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, v.42, p. 181-192, 2012.
- MAYO, R. *Derivativos de eletricidade e gerenciamento de risco*. Rio de Janeiro: Synergia, 2009. 124 p.
- MELO, E.; NEVES, E.M.A.; PAZZINI, L.H.A. The Brazilian Electricity Model: An Overview of the Current Structure and Market Design. *IEEE 8th International Conference on the European Energy Market (EEM)*, Zagreb, p. 592-597, 2011.

- MENG, K.; DONG, Z.Y.; WONG, K.P. Self-adaptative radial basis function neural network for short-term electricity price forecasting. IET Generation, Transmission & Distribution, v.3, p. 325-335, 2008.
- MOLINA, J.P.; ZOLEZZI, J.M; CONTRERAS, J.; RUDNICK, H.; REVECO, M.J. Nash-Cournot Equilibria in Hydrothermal Electricity Markets. IEEE Transactions on Power Systems. v. 26, p. 1089-1101, 2010 .
- MORETTIN, P.A.; TOLOI, C.M.C. Análise de séries temporais. São Paulo: Edgard Blücher, 2006. 544 p.
- MUNOZ, M.; CORCHERO, C.; HEREDIA, F. Improving electricity market price scenarios by means of forecasting factor models. Barcelona: Universitat Politècnica de Catalunya, 2009.
- NELSON, C. R. Applied Time Series Analysis for Managerial Forecasting. San Francisco : Holden-Day, 1973. 231 p.
- NOGALES, F.J.; CONEJO, A.J. Electricity price forecasting through transfer function models. Journal of operational research society, v. 4, p.350-358, 2006.
- PAO, H. T. Forecasting electricity market pricing using artificial neural networks . Energy conversion and management, v. 48, p. 907-912, 2007.
- PARK, Y.; GOSH, J. Ensembles of α -Trees for Imbalanced Classification Problems. IEEE Transactions on Knowledge and Data Engineering, v. 26, p. 131-143, 2014.
- PATRIKAR, A.M. Approximating Gaussian Mixture Model or Radial Basis function Network With Multilayer Perceptron. IEEE Transactions on Neural Networks and Learning Systems, v. 24, p.1161 - 1166, 2013.
- PINDORIYA, N. M.; SINGH, S. N.; SINGH, S. K. An adaptive wavelet neural network-based energy price forecasting in electricity markets. IEEE Transactions on Power Systems, v. 23, p. 1423-1432, 2008 .
- PRÍNCIPE, J.C., EULIANO, N.R.; LEFEBVRE, W.C. Neural and adaptive systems: fundamentals through simulations. New Jersey: John Wiley & sons, 2000. 627 p.
- QUEIROZ, A.R.; OLIVEIRA, F.A.; LIMA, J.W.M.; BALESTRASSI, P.P. Simulating electricity spot prices in Brazil using neural network and design of experiments. IEEE PowerTech Conf. Proc., Lausanne, P. 2029-2034, 2007.
- QUINLAN, J. R. C4.5: programs for machine learning. San Francisco: Morgan Kaufmann, 1993. 312 p.
- QUINLAN, J. R. Induction of decision trees. Machine Learning, v.1, p.1–106, 1986.

- REFENES, A.N.; BURGESS, A.N.; BENTZ, Y. Neural networks in financial engineering: a study in methodology. *IEEE transactions on Neural Networks*, v. 8, p. 1222-1267, 1997.
- RESTON FILHO, J.C.; AFFONSO, C.M.; OLIVEIRA, R.C.L. Pricing analysis in the Brazilian energy market: a decision tree approach. *IEEE PowerTech Conf. Proc.*, Bucharest, p. 1-6, 2009.
- RUMELHART, D.; HINTON, G.E.; WILLIAMS, R.J. Learning representations by back-propagation errors. *Nature*, v. 323, p. 533-536, 1986.
- SHAO, Y.; TAFF, G.N.; WALSH, S.J. Comparison of Early Stopping Criteria for Neural-Network-Based Subpixel Classification. *IEEE Geoscience and Remote Sensing Letters*, v.8, p.113-117, 2011.
- SHI, D.; GAO, Y. A New Method for Identifying Electromagnetic Radiation Sources Using Backpropagation Neural Network. *IEEE transactions on electromagnetic compatibility*, v. 55, p. 842-848, 2013.
- SILVA, E. L. Formação de preços em mercados de energia elétrica. Porto Alegre: Sagra Luzzato, 2001. 184 p.
- SKAPURA, D.M. Building neural networks. New York: Addison-Wesley, 1996. 304 p.
- SOUZA, F.C. Dinâmica da gestão de riscos no ambiente de contratação regulada do setor elétrico brasileiro. Tese (Doutorado em Planejamento Energético). 2008. 154 f. Rio de Janeiro: Universidade Federal do Rio de Janeiro, 2008.
- STERBA, J.; HILOVSKA, K. The implementation of hybrid ARIMA neural network prediction model for aggregate water consumption prediction. *Aplimat—Journal of Applied Mathematics*, v.3, p.123–131,2010.
- SU, C.L.; KIRSCHEN, D. Quantifying the effect of demand response on electricity markets. *IEEE Transactions on power systems*. v. 24, p. 1199–1207, 2009.
- TAN, P. N.; STEINBACH, M.; KUMAR, V. Introdução ao Data Mining – Mineração de dados. Rio de Janeiro: Ciência Moderna, 2009. 928 p.
- TORRES, R. C. Avaliação de Portfolios de Contratos de Compra e Venda de Energia Elétrica: uma abordagem pela teoria das opções. 2006. 103 f. Dissertação (Mestrado em Engenharia Industrial). Rio de Janeiro: PUC-Rio.
- VEHVILÄILEN, I.; PYYKKÖNEN, E T. Stochastic factormodel for electricity spot price - the case of the nordic market. *Energy Economics*, v.27, p. 351-367, 2005 .
- VILLAR, J.; RUDNICK, H. Hydrothermal Market simulator using game theory: assessment of Market power. *IEEE Transactions on Power Systems*, v.18, p.91-98, 2003.

- WANG, K.; YANG, J.; SHI, G.; WANG, Q. An Expanded Training Set Based Validation Method to Avoid Overfitting for Neural Network Classifier. Fourth international conference on natural computation - ICNC '08, v. 3, p. 83-87, 2008.
- WERON, R. Modeling and forecasting electricity loads and price: a statistical approach. Chichester: John Wiley & sons, 2006.
- WERON, R.; MISIOREK, E.A. Forecasting Spot Electricity Prices: A Comparison of Parametric and Semiparametric Time Series Models. International journal of forecasting, v.24, p. 744-763, 2008.
- WHITE, H. Economic Prediction using Neural Networks : The case of the IBM Daily Stock Returns. Proceedings of the IEEE International Joint Conference on Neural Networks, v.2, p. 451-458, 1988.
- WILCOX, R. Applying contemporary statistical techniques. Academic Press, 2003. 608 p.
- WITTEN, I.H.; FRANK, E.; HALL, M.A. Data mining: practical machine learning tools and techniques. San Francisco: Morgan Kaufmann, 2011. 629 p.
- WU, L. ; SHAHIDEHPOUR, M. A hybrid model for day-ahead price forecasting, IEEE Transactions on power systems, v. 25, p. 1519–1530, 2010.
- YAN, X.; CHOWDHURY, N.A. Mid-term electricity market clearing price forecasting: a hybrid LSSVM and ARMAX approach. International Journal of Electrical Power & Energy Systems, v. 53, p. 20-26, 2013.
- YAN, X.; CHOWDHURY, N.A. Mid-term electricity market clearing price forecasting using multiple least squares support vector machines. IET - Generation, Transmission & Distribution, v.8, p. 1572-1582, 2014.
- YANG, P.; TANG, G.; NEHORAI, A. A game-theoretic approach for optimal time-of-use electricity pricing. IEEE Transactions on Power Systems, v. 28, p. 884-892, 2013.
- YUN, Z.; QUAN, Z.; CAIXIN, S.; SHAOLAN, L.; YUMING, L.; YANG, E.S. RBF neural network and ANFIS based short term load forecasting approach in real time price environment. IEEE Transactions on Power Systems, v.23, p. 853-858 , 2008.
- ZAREIPOUR, H.; CANIZARES, C.; BHATTACHARYA, K.; THOMSOM, J. Application of public-domain Market information to forecast Ontario wholesale electricity prices. IEEE Transactions on Power Systems, v. 21, p. 1707-1717, 2006.
- ZAREIPOUR, H.; JANJANI, A.; LEUNG, H.; MOTAMEDI, A.; SCHELLENBERG, A. Classification of future electricity Market prices. IEEE Transactions on Power Systems, v.26, p. 165-173, 2011

- ZHAO, J. H.; DONG, Z.Y.; LI, X. Electricity market price spike forecasting and decision making. IET Generation, Transmission & Distribution, v. 4, p. 647-654, 2007.
- ZHAO, J. H.; DONG, Z.Y.; LI, X.; WONG, K.P. A framework for electricity price spike analysis with advanced data-mining methods. IEEE transactions on Power Systems, v. 22, p. 376-385, 2007.