

**UNIVERSIDADE FEDERAL DE SANTA MARIA
CENTRO DE TECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA**

**UMA METODOLOGIA PARA A IDENTIFICAÇÃO DE
PERDAS NÃO TÉCNICAS EM GRANDES
CONSUMIDORES RURAIS**

DISSERTAÇÃO DE MESTRADO

Maicon Coelho Evaldt

Santa Maria, RS, Brasil.

2014

UMA METODOLOGIA PARA A IDENTIFICAÇÃO DE PERDAS NÃO TÉCNICAS EM GRANDES CONSUMIDORES RURAIS

Maicon Coelho Evaldt

Dissertação apresentada ao Curso de Mestrado do Programa de Pós-Graduação em Engenharia Elétrica Área de Concentração em Processamento de Energia, Centro de Estudos em Energia e Sistemas de Potência (CEESP), da Universidade Federal de Santa Maria (UFSM, RS), como requisito parcial para obtenção do grau de **Mestre em Engenharia Elétrica.**

**Orientadora: Prof^a. Dr^a. Alzenira da Rosa Abaide
Coorientador: Prof. Dr. Daniel Pinheiro Bernardon**

Santa Maria, RS, Brasil.

2014

Ficha catalográfica elaborada através do Programa de Geração Automática da Biblioteca Central da UFSM, com os dados fornecidos pelo(a) autor(a).

Evaldt, Maicon Coelho

Uma metodologia para a identificação de perdas não técnicas em grandes consumidores rurais / Maicon Coelho Evaldt.-2014.

122 p.; 30cm

Orientadora: Alzenira da Rosa Abaide

Coorientador: Daniel Pinheiro Bernardon

Dissertação (mestrado) - Universidade Federal de Santa Maria, Centro de Tecnologia, Programa de Pós-Graduação em Engenharia Elétrica, RS, 2014

1. Rede neural artificial 2. Perdas não técnicas 3. Consumidores rurais I. Abaide, Alzenira da Rosa II. Bernardon, Daniel Pinheiro III. Título.

**Universidade Federal de Santa Maria
Centro de Tecnologia
Programa de Pós-Graduação em Engenharia Elétrica**

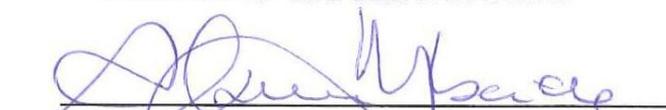
A Comissão Examinadora, abaixo assinada,
aprova a Dissertação de Mestrado

**UMA METODOLOGIA PARA A IDENTIFICAÇÃO DE
PERDAS NÃO TÉCNICAS EM GRANDES
CONSUMIDORES RURAIS**

elaborada por
Maicon Coelho Evaldt

como requisito parcial para obtenção do grau de
Mestre em Engenharia Elétrica

COMISSÃO EXAMINADORA:

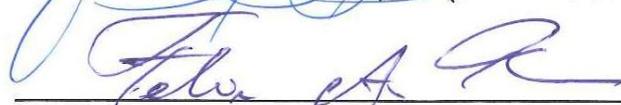


Alzenira da Rosa Abaide, Dr^a. (UFSM/RS)
(Presidente/Orientadora)

Daniel Pinheiro Bernardon, Dr. (UFSM/RS)
(Coorientador)



José Vicente Canto dos Santos, Dr. (UNISINOS/RS)



Felix Alberto Farret, Dr. (UFSM/RS)

Santa Maria, 26 de Setembro de 2014.

"O Sucesso na vida não se mede pelo que você conquistou, mas sim pelas dificuldades que superou no caminho."

Abraham Lincoln

AGRADECIMENTOS

Na caminhada da vida se agrega muito do que as pessoas lhe oferecem; o conhecimento é só um deles e este é apenas uma consequência da permissão divina e do amor das pessoas que estão de alguma maneira presente nos momentos mais importantes da vida. Algumas destas cito aqui:

A meu querido pai, Sr. Antônio José Evaldt (em memória), minha mãe Sra. Maria Eponina Coelho Evaldt, meu irmão Márcio Coelho Evaldt, minha tia Sra. Maria das Graças Porto Coelho e demais familiares. Obrigado por estarem comigo e entenderem muitos momentos de ausência em prol desta causa.

À minha esposa e meu amor Daiane Calheiro Evaldt, pelo companheirismo e pela nossa maior conquista, Alice Calheiro Evaldt, nossa amada filha de um mês de idade no momento em que apresento essa Dissertação.

À professora Dr^a. Alzenira da Rosa Abaide, orientadora desta Dissertação, pelas inúmeras orientações, incentivo e motivação. Ao professor Daniel Pinheiro Bernardon, coorientador, pelas intervenções cirúrgicas nos rumos do trabalho.

Ao professor Luciano Lopes Pfitscher, grande amigo e “pai” acadêmico, com quem aprendi muito para chegar até aqui.

Aos amigos da Unisinos pela compreensão e apoio, estes representados nos nomes de Guilherme Luís Roehe Vaccaro, Paulo Ricardo da Silva Pereira, Rodrigo Iván Goytia Mejía e em especial ao professor Rodrigo Marques de Figueiredo, pelo apoio técnico nas RNA,s e no *Software Python*, bem como, compreensão como gestor.

Aos colegas do CEESP pelo companheirismo e carinho de forma acentuada ao amigo Enoque Dutra Garcia por sua colaboração.

Em fim, a todos com participações e intensidades diferentes me ajudaram nesta trajetória que me levou a esta conquista.

RESUMO

Dissertação de Mestrado
Programa de Pós-Graduação em Engenharia Elétrica
Universidade Federal de Santa Maria

UMA METODOLOGIA PARA A IDENTIFICAÇÃO DE PERDAS NÃO TÉCNICAS EM GRANDES CONSUMIDORES RURAIS

AUTOR: MAICON COELHO EVALDT
ORIENTADORA: ALZENIRA DA ROSA ABAIDE
CORIENTADOR: DANIEL PINHEIRO BERNARDON
Data e Local da Defesa: Santa Maria, 22 de Agosto de 2014.

A irrigação de grandes áreas agrícolas representa uma porção significativa do consumo de energia elétrica em países produtores de alimentos. A potência das bombas de irrigação em lavouras de arroz, por exemplo, pode ser superior a 800 CV, sendo que esses sistemas são utilizados continuamente no período da safra. No Brasil, as perdas não técnicas neste tipo de consumidor geralmente são devidas a fraudes, erros em medidores de energia e ligações clandestinas. Esse tipo de problema é de difícil fiscalização devido à grande extensão das linhas rurais, além da dificuldade de acesso, em muitos casos. Este trabalho apresenta uma proposta para a identificação de perdas não técnicas em alimentadores rurais contendo sistemas de bombeamento para irrigação de lavouras de arroz. A metodologia proposta é baseada na correlação dos padrões de consumo de energia elétrica, das características da área irrigada e das condições climáticas do período de irrigação. A metodologia emprega a técnica de Rede Neural Artificial, e tem como entrada um conjunto de dados de precipitação pluviométrica, temperatura, incidência solar, umidade do ar, carga instalada e área de solo irrigado característico do cultivo de arroz. O resultado final das análises indica o risco percentual de cada conjunto de dados e inconsistências que possam implicar em perdas não técnicas. Os resultados do trabalho foram obtidos e validados a partir de uma base de dados reais de safras do período entre 2009 e 2014, de lavouras do Estado do Rio do Grande do Sul.

Palavras-chave: Rede neural artificial. Perdas não técnicas. Consumidores rurais.

ABSTRACT

Master Course Dissertation
Post Graduation Program in Electric Engineering
Universidade Federal de Santa Maria

A METHODOLOGY FOR IDENTIFYING NON-TECHNICAL LOSSES IN LARGE RURAL CONSUMERS

AUTHOR: MAICON COELHO EVALDT
ADVISER: ALZENIRA DA ROSA ABAIDE
CO ADVISER: DANIEL PINHEIRO BERNARDON
Defense Place and Date: Santa Maria, August 22, 2014.

Irrigation of large agricultural areas represents a significant portion of energy consumption in food producing countries. The electric power of irrigation pumps in rice crops, for example, may exceed 800 HP, while these systems are continually used during the harvest. In Brazil, non-technical losses in this type of consumer are generally due to fraud or error in power meters, or illegal connections. This type of problem is difficult to control because of the large length of rural feeder, in addition to access difficulties in many cases. This work presents a proposal for identifying non-technical losses in rural feeders containing pumping systems for irrigation of rice crops. The proposed methodology is based on the correlation of patterns of energy consumption, characteristics of the irrigated area and climatic conditions of the irrigation period. The developed system employs Artificial Neural Network technique, and it has as input a dataset of rainfall, temperature, solar irradiation, humidity, installed power and irrigated area of rice cultivation. The final result of the analysis indicates the percentage risk of each set of data and inconsistencies that can result in non-technical losses. The results of the developed methodology were obtained and validated from a real data base of crops of the period between 2009 and 2014, in the State of Rio Grande do Sul, Brazil.

Index Terms: Artificial neural network. Non-technical losses. Rural Consumers.

LISTA DE FIGURAS

Figura 2.1 – Percentual das perdas elétricas globais de concessionárias de energia em 2012.....	28
Figura 2.2 – Percentuais de perdas técnicas e não técnicas de 64 concessionárias de energia no período de 2000 a 2012 sobre o total gerado.....	28
Figura 2.3 – Esquema de uma rede elétrica com perdas técnicas e não Técnicas	32
Figura 2.4 – Representação de uma Rede Neural Artificial.....	36
Figura 2.5 – Rede Neural Artificial – Estrutura de Kohonen	41
Figura 2.6 – Rede Neural Artificial – Estrutura Multilayer Perceptron.....	41
Figura 2.7 – Bombas de irrigação para lavouras de arroz	45
Figura 3.1 – Exemplo de rede neural artificial simples.	55
Figura 3.2 – Representação de um neurônio	57
Figura 3.3 – Exemplos de funções de ativação	58
Figura 3.4 – Exemplos de estruturas de RNAs.....	59
Figura 3.5 – Estrutura de Aprendizado Não-Supervisionado.....	60
Figura 3.6 – Estrutura de Aprendizado Supervisionado	61
Figura 3.7 – Gráfico de Fluxo de Sinal de uma MLP com retropropagação de erro	64
Figura 3.8 – Divisão do conjunto de dados para validação do modelo da RNA	65
Figura 3.9 – Técnica de Validação Cruzada <i>K-fold</i>	66
Figura 3.10 – Exemplo de análise por Correlação Gráfica	69
Figura 3.11 – Exemplo de análise por Sobreposição de Curvas	69
Figura 4.1 – Fluxograma geral da metodologia proposta	72
Figura 4.2 – Dados de Entrada e Saída do Sistema Proposto.	73
Figura 4.3 – Unidade consumidora UC-1.	76
Figura 4.4 - Unidade consumidora UC-2.....	76
Figura 4.5 – Unidade consumidora UC-3.	76
Figura 4.6 – Consumo de energia elétrica em um sistema de irrigação obtido com o programa Elo50	77
Figura 4.7 – Resultado da Regressão <i>Stepwise</i> para UC-1/ Safra 2013-2014.78..	82

Figura 4.8 – Rede Neural Artificial proposta	83
Figura 4.9 – Relação entre número de neurônios na camada intermediária e o erro de treinamento da RNA.	84
Figura 4.10 – Validação Cruzada com K = 10.....	85
Figura 4.11 – Resultado da RNA com Validação Cruzada (1º conjunto de dados)	86
Figura 4.12 – Resultado da RNA com Validação Cruzada (2º conjunto de dados)	86
Figura 4.13 – Resultado da RNA com Validação Cruzada (3º conjunto de dados)	87
Figura 4.14 – Três redes treinadas com Validação Direta	88
Figura 4.15 – Resultado da análise de UC-1 com RNA e Validação Direta.....	89
Figura 4.16 – Detalhe do resultado da análise de um mês de UC-1 com RNA e Validação Direta.....	89
Figura 4.17 – Resultado da análise de UC-2 com RNA e Validação Direta.....	90
Figura 4.18 – Detalhe do resultado da análise de um mês de UC-2 com RNA e Validação Direta.....	90
Figura 4.19 – Resultado da análise de UC-3 com RNA e Validação Direta.....	91
Figura 4.20 – Detalhe do resultado da análise de um mês de UC-3 com RNA e Validação Direta.....	91
Figura 4.21 – Resultado da simulação de falha na medição de UC-3.	93
Figura 4.22 – Detalhe do resultado da simulação de falha na UC-3, Safra 2009-2010.....	94

LISTA DE TABELAS

Tabela 2.1 – Percentual de perdas não técnicas das concessionárias de energia brasileiras em 2011	30
Tabela 2.2 – Perdas técnicas típicas em redes elétricas no Brasil	33
Tabela 3.1 – Resultados hipotéticos para exemplo de análise da Regressão <i>Stepwise</i>	53
Tabela 4.1 – Índice de possibilidade de perdas não técnicas	74
Tabela 4.2 – Dados gerais das lavouras	77
Tabela 4.3 – Análise de correlação pelo Coeficiente de Pearson / UC-1 2009/2010	80
Tabela 4.4 – Análise de correlação do consumo de energia com clima para UC-1 ..	80
Tabela 4.5 – Resumo da análise de UC-1 modelada com RNA e Validação Direta.....	92
Tabela 4.6 – Resumo da análise de UC-2 modelada com RNA e Validação Direta.....	92
Tabela 4.7 – Resumo da análise de UC-3 modelada com RNA e Validação Direta.....	92
Tabela 4.8 – Resumo da análise de UC-3 com simulação de falha de medição.....	94

LISTA DE ABREVIATURAS E SIGLAS

ANEEL	Agência Nacional de Energia Elétrica
BT	Baixa Tensão
DNU	Dia Não Útil
DU	Dia Útil
I	Corrente
MAE	<i>Mean Absolute Error</i>
MAPE	<i>Mean Absolute Percentage Error</i>
P	Potência ativa
p.u.	Por unidade
PRODIST	Procedimento de Distribuição do Setor Elétrico Nacional
RNA	Redes Neurais Artificiais
RSME	<i>Root Mean Square Error</i>
SEP	Sistema Elétrico de Potência
T	Temperatura

LISTA DE APÊNDICES

APÊNDICE A –RESULTADOS DE CORRELAÇÃO ENTRE CONSUMO DE ENERGIA E VARIÁVEIS CLIMÁTICAS	107
APÊNDICE B –REDE NEURAL ARTIFICIAL DESENVOLVIDO EM PYTHON®..	111
APÊNDICE C –REDE NEURAL ARTIFICIAL DESENVOLVIDO EM MATLAB	113

SUMÁRIO

1 INTRODUÇÃO	23
1.1 Justificativa.....	24
1.2 Metodologia	24
1.3 Objetivos	25
1.4 Estrutura do trabalho	26
2 REVISÃO BIBLIOGRÁFICA	27
2.1 Perdas não técnicas em Sistemas de Potência.....	27
2.2 Detecção de perdas não técnicas	31
2.2.1 Estimação de perdas não técnicas a partir das perdas globais e técnicas.....	32
2.2.2 Estimação de perdas a partir de Estimadores de Estado das Redes de Distribuição.....	35
2.2.3 Detecção de perdas não técnicas a partir dos perfis de consumidores	37
2.2.4 Monitoramento de perdas com medidores eletrônicos	42
2.3 Particularidades do consumo de energia em áreas rurais	44
2.4 Considerações finais	46
3. FUNDAMENTAÇÃO TEÓRICA DOS MÉTODOS APLICADOS	49
3.1 Análise Estatística – Seleção e Regressão Stepwise	49
3.1.1 Algoritmo de Regressão Stepwise	50
3.1.2 Exemplo numérico.....	52
3.2 Redes Neurais Artificiais	55
3.2.1 Representação das RNAs	56
3.2.2 Estruturas de RNAs.....	59
3.2.3 Aprendizagem das RNAs	60
3.2.4 Redes <i>Perceptrons</i> de Múltiplas Camadas	62
3.2.4.1 Algoritmo de Retropropagação de Erro	62
3.2.4.2 Validação Cruzada	65
3.3 Avaliação de desempenho	66
3.3.1 Análise Estatística	67
3.3.2 Análise Gráfica	68
3.4 Considerações Finais	70
4. IDENTIFICAÇÃO DE PERDAS NÃO TÉCNICAS NO SUPRIMENTO DE ENERGIA A CONSUMIDORES RURAIS - METODOLOGIA PROPOSTA E RESULTADOS	71
4.1 Fluxograma geral proposto para identificação de perdas não técnicas	71
4.2 Aquisição e tratamento de dados	72
4.3 Ferramentas de modelagem, análise e validação	74
4.4 Indicativo de perdas elétricas não técnicas.....	74

4.5 Descrição do conjunto de dados para teste e validação	75
4.6 Análise Estatística e Regressão Stepwise	78
4.7 Correlação do consumo de energia elétrica com o clima	78
4.8 Análise por Regressão Stepwise.....	81
4.9 Implementação e Análise por Redes Neurais Artificiais	82
4.10 Treinamento com Validação Cruzada K-Fold.....	85
4.11 Treinamento com Validação Direta	87
4.12 Resultados de simulação de uma situação de perda não técnica	93
CONCLUSÕES	95
REFERÊNCIAS BIBLIOGRÁFICAS	99

1 INTRODUÇÃO

O consumo de energia elétrica pode ser considerado um indicador do desenvolvimento econômico de um país ou região. A ampliação da produção industrial e o crescimento do poder aquisitivo da população impactam e são diretamente impactados pela capacidade de produção de energia elétrica. Nesse sentido, o aumento da demanda de energia deve ser acompanhado de estudos e planejamento de expansão do Sistema Elétrico de Potência.

Por outro lado, quando a energia elétrica produzida não é devidamente aproveitada, entende-se que o sistema é ineficiente do ponto de vista energético, e que os investimentos em ampliação da capacidade de produção de energia não são plenamente usufruídos. É o que acontece, por exemplo, em um sistema com grandes quantidades de perdas elétricas.

De modo geral, as perdas elétricas nos Sistemas de Potência podem ser divididas em dois grandes grupos: as perdas inerentes aos sistemas elétricos, devidas aos fenômenos físicos envolvidos na geração, transmissão e distribuição de energia elétrica, e que são usualmente referidas como *perdas técnicas*, e as perdas relacionadas ao processo de faturamento das concessionárias distribuidoras de energia, que devido ao caráter estreitamente relacionado a fatores econômicos, eram habitualmente denominadas perdas comerciais. Atualmente, é mais usual empregar o termo “*perdas não técnicas*” para designar esse tipo de perdas. Entre os exemplos mais comuns de perdas não técnicas, destacam-se o furto de energia e fraudes no sistema de medição.

No Brasil, as perdas não técnicas implicam em prejuízos na faixa de bilhões de Reais, no contexto global das Concessionárias de Energia. Para ilustrar um exemplo, a CELPA (Centrais Elétricas do Pará, PA) registrou em 2011 quase um quarto (1/4) de perdas não técnicas dentre a sua produção e distribuição anual de energia.

Além da questão econômica, as perdas não técnicas impactam no próprio planejamento e operação da rede, uma vez que investimentos em expansão e equipamentos podem ser necessários para suprir um consumo não previsto e /ou

não faturado, e que consumidores ilegais podem interferir na segurança e qualidade de energia da rede.

1.1 Justificativa

Diante do contexto apresentado, essa dissertação aborda o problema das perdas não técnicas em sistemas de distribuição de energia. O escopo do trabalho são alimentadores com grandes consumidores rurais, caracterizados pela utilização de sistemas de irrigação em lavouras de arroz.

A definição desse escopo levou em consideração a forte característica de produção agrícola de arroz irrigado do Estado do Rio Grande do Sul e a carência de estudos sobre perdas não técnicas nesse tipo de consumidores de energia elétrica.

No Rio Grande do Sul, e mesmo em outros Estados do Brasil, os sistemas de irrigação de grandes lavouras de arroz possuem potência instalada na faixa de algumas centenas de kVA, o que ressalta a relevância desse tipo de consumidor na rede elétrica. Além disso, devido a grandes distâncias envolvidas e dificuldade de acesso em algumas situações, a fiscalização e identificação de fraudes acaba sendo limitada, por parte das Concessionárias de Energia, o que torna o sistema mais susceptível ao problema das perdas não técnicas.

1.2 Metodologia

A proposta dessa dissertação é desenvolver um sistema que auxilie na identificação de possíveis situações de perdas não técnicas, baseado em históricos e padrões de consumo considerando características próprias das lavouras e registros climatológicos.

A metodologia proposta emprega Redes Neurais Artificiais (RNAs), treinadas com um conjunto de dados reais das lavouras de arroz irrigado do Rio Grande do Sul. O sistema desenvolvido emprega a RNA *Perceptron* de múltiplas camadas, que tem como entrada dados como dimensões da lavoura, potência instalada,

temperatura e precipitação pluvial. A saída da RNA é uma indicação do consumo de energia esperado. O desenvolvimento das RNAs foi realizado em linguagem de programação Python®, que possui bibliotecas específicas para essa aplicação.

Adicionalmente, também foi realizado um estudo sobre a correlação entre os dados utilizados no treinamento da RNA e o consumo de energia da lavoura, através de ferramentas estatísticas. A análise estatística básica foi realizada através do programa Minitab®. Especificamente, para análise do método estatístico de Regressão *Stepwise*, foi utilizado o programa Matlab®.

A partir do consumo de energia determinado pela RNA, e do consumo de energia observado na lavoura, foi definido um índice de possibilidade de ocorrência de perdas não técnicas em cada lavoura analisada.

1.3 Objetivos

O objetivo geral desse trabalho é:

- Desenvolver uma ferramenta que proporcione um indicativo da possibilidade de perdas não técnicas no suprimento de energia de lavouras de arroz irrigado.

Para atender ao objetivo geral, o mesmo foi desmembrado nos seguintes objetivos específicos:

- Analisar a relação entre consumo de energia elétrica, características das lavouras, e informações climatológicas.
- Desenvolver uma Rede Neural Artificial para determinação do consumo de energia esperado em lavouras específicas, empregando um processo de seleção de dados para treinamento e validação.
- Desenvolver uma ferramenta computacional que avalie os resultados de análise e indique a possibilidade de perda não técnica em cada caso.

1.4 Estrutura do trabalho

Esse trabalho foi dividido em cinco capítulos, conforme segue:

- Capítulo 1: nesse capítulo, introdutório, é feita a contextualização do tema abordado na dissertação, e são definidos os objetivos do trabalho.
- Capítulo 2: apresenta a revisão bibliográfica do trabalho, dividida em duas partes principais: caracterização e detecção de perdas não técnicas, e caracterização dos sistemas de irrigação de lavouras de arroz.
- Capítulo 3: apresenta o embasamento teórico das ferramentas estatísticas e das Redes Neurais Artificiais empregadas no trabalho.
- Capítulo 4: nesse capítulo, é apresentada a metodologia desenvolvida para detecção de perdas não técnicas nas lavouras de arroz irrigado. Nele também são apresentados os resultados da metodologia empregada, com análise de dados reais de safras de lavouras do Rio Grande do Sul.
- Capítulo 5: nesse capítulo são feitas as considerações finais sobre a metodologia empregada e os resultados obtidos, e são apresentadas sugestões de continuidade do trabalho.

2 REVISÃO BIBLIOGRÁFICA

Neste capítulo é apresentada uma síntese da revisão bibliográfica realizada sobre perdas não técnicas e ferramentas para modelagem e análise relacionadas a esse tema. A abordagem foi dividida em duas partes principais: caracterização e detecção de perdas em Sistemas de Potência. Ao final do capítulo, são feitas considerações que embasam a metodologia proposta nessa Dissertação.

2.1 Perdas não técnicas em Sistemas de Potência

As perdas elétricas em Sistemas de Potência podem ser classificadas em três tipos: i) perdas globais, que envolvem todos os tipos de perdas de energia no sistema; ii) perdas técnicas, que são referentes principalmente a perdas de dissipação de potência nos condutores e equipamentos, desde a geração até o ponto final de entrega de energia e; iii) perdas não técnicas, antes chamadas de comerciais, que são a soma de fraudes, desvios e diversas falhas no processo de faturamento. Desenvolver métodos precisos para identificação e mitigação dessas perdas contribui com a qualidade e a segurança da distribuição de energia elétrica, uma vez que elas impactam diretamente na operação do sistema, causando problemas como sobrecarga e níveis de tensão abaixo dos limites aceitáveis.

Para ilustrar um panorama da dimensão das perdas nos Sistemas de Potência em redes elétricas brasileiras, as Figuras 2.1 e 2.2 apresentam percentuais de perdas globais, técnicas e não técnicas de diversas concessionárias de energia, a partir de um estudo desenvolvido pela Associação Brasileira de Distribuidoras de Energia Elétrica (ABRADEE, 2012). Na Figura 2.1 observam-se as perdas globais individualizadas por concessionária, e na Figura 2.2 observa-se um histórico de perdas técnicas e não técnicas (no gráfico, apresentadas como comerciais) no período de 10 anos, para um conjunto de 64 concessionárias. Ressalta-se que os índices apresentados referem-se ao total da energia elétrica injetada na rede pelas concessionárias. Esses elevados índices implicam na preocupação das

concessionárias no combate às perdas e cumprimento de metas definidas pela ANEEL, uma vez que o montante de energia não faturada equivale a diversas usinas de porte considerável, operando apenas para suprir perdas (ANEEL, 2011).

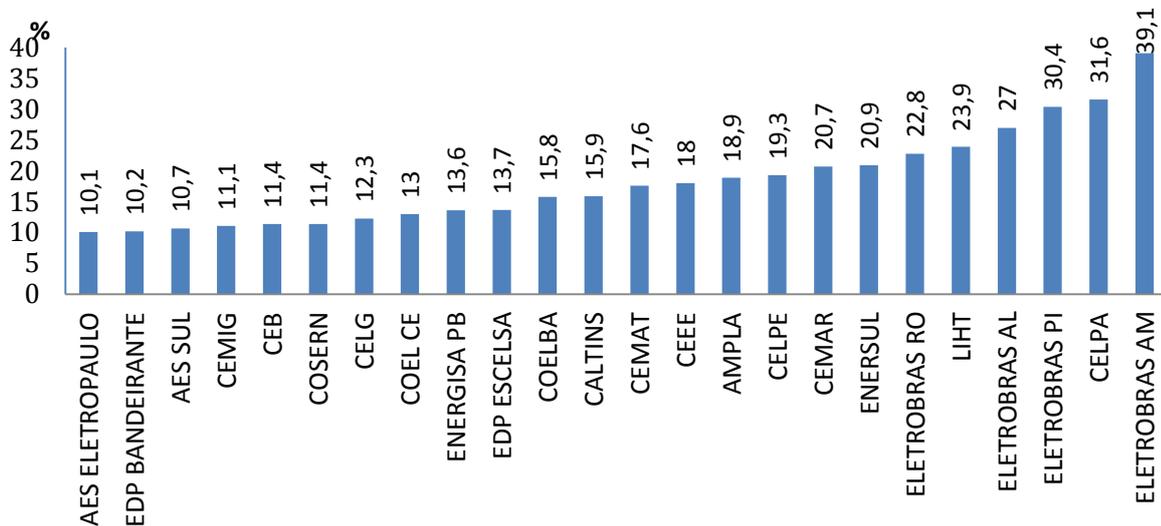


Figura 2.1 – Percentual das perdas elétricas globais de concessionárias de energia em 2012.

Fonte: Sistema de Informação para Gestão, ABRADDEE.

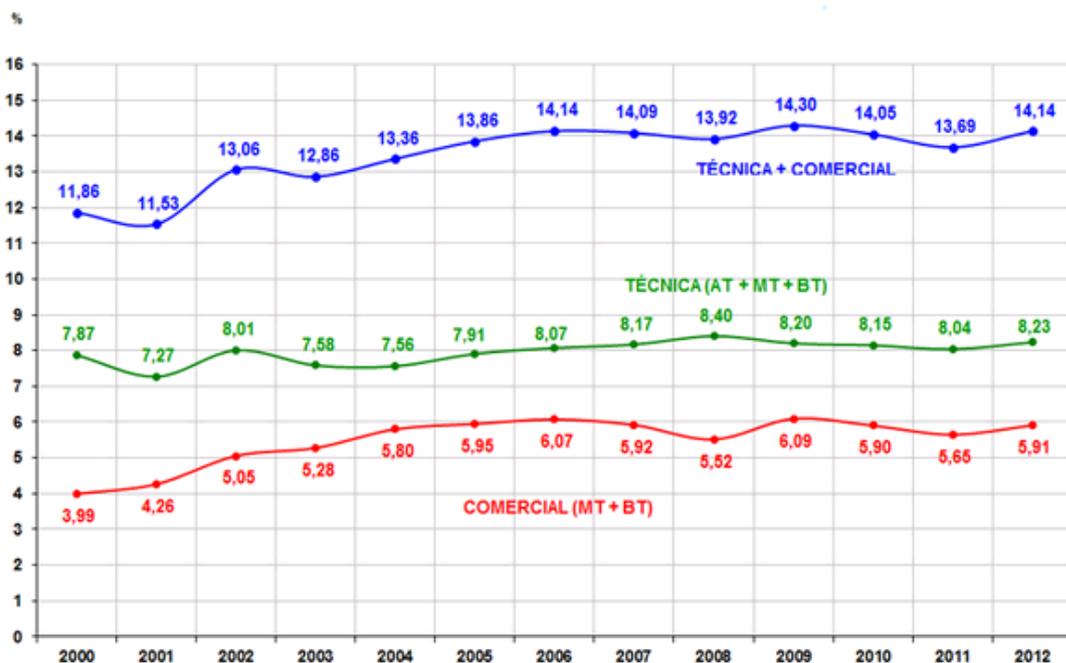


Figura 2.2 – Percentuais de perdas técnicas e não técnicas de 64 concessionárias de energia no período de 2000 a 2012 sobre o total gerado.

Fonte: Sistema de Informação para Gestão, ABRADDEE.

As perdas não técnicas, objeto de estudo nessa Dissertação, são geralmente associadas a fatores sociais e econômicos. No Brasil, a meta de perdas não técnicas a ser alcançada pelas concessionárias de energia é definida pela ANEEL a partir de modelos econométricos, que medem a complexidade socioeconômica de cada região de concessão.

Observa-se, no entanto, que a questão socioeconômica é relevante também em outros países, em especial naqueles com alta população e grande extensão territorial, como Índia, Malásia e Estados Unidos.

Chauhan e Rajvanshi (2013) apresentam uma revisão sobre perdas não técnicas em Sistemas de Potência, e relacionam as suas principais causas no sistema de distribuição brasileiro:

- Adulteração de medidores de energia;
- Conexão ilegal à rede (no Brasil, popularmente conhecida como “gato”);
- Erros ou falsas leituras de consumo, com eventuais subornos;
- Suprimento de energia não medido, devido a falhas no registro de consumidores ou no processo de faturamento.

De modo geral, as perdas não técnicas de energia elétrica são relacionadas ao processo de comercialização, e incluem erros de medição, fraudes, e principalmente, furto da energia elétrica (QUEIROS, 2010). A maioria dos furtos de energia ocorre através de desvios ou ligações clandestinas, como perfuração de eletroduto ou ganchos de rede para ligação *by-pass* de energia elétrica, desvios de energia com instalação de acionamento remoto, entre outros (REIS, 2005).

Fraudes significativas ocorrem também nos equipamentos do sistema de distribuição, em especial nos medidores de energia. Muitas vezes os defeitos nesses equipamentos podem parecer naturais e são registrados como perdas técnicas, mas na verdade foram provocados de forma voluntária, com objetivo de reduzir o consumo medido (CALILI, 2005).

Segundo a ANEEL (2010), no Brasil as perdas não técnicas são responsáveis por aproximadamente R\$ 8,1 bilhões de prejuízos por ano. A região com maior índice de consumo irregular é o Norte, com 20% da energia distribuída, seguida do Sudeste, com 10% e do Nordeste, com 9%. No Centro-Oeste, o percentual é de 5%, e no Sul, de 3%. A Tabela 2.1 apresenta um *ranking* de concessionárias mais atingidas pelas perdas não técnicas, de acordo com a ANEEL (2011).

Tabela 2.1 – Percentual de perdas não técnicas das concessionárias de energia brasileiras em 2011.

Posição	Concessionária	Perdas não técnicas
1	CELPA	24,4 %
2	LIGHT	24,2 %
3	CERON	22,0 %
4	CEMAR	17,8 %
5	AMPLA	17,1 %
6	CEAL	17,0 %
7	AMAZONAS ENERGIA	16,8 %
8	ELETROACRE	15,9 %
9	CEPISA	15,8 %
10	ENERGISA PARAÍBA	11,2 %
11	ELETROPAULO	10,8 %
12	CEEE	10,5 %
13	BANDEIRANTE	10,1 %
14	EXCELSA	10,0 %
15	BOA VISTA	10,0 %

Fonte: ANEEL (2011)

Este tipo de perda, bastante comum no Brasil, ocorre também em grande parte do mundo, incluindo os países mais desenvolvidos. Nos Estados Unidos, por exemplo, o valor com o furto de energia é em torno de U\$ 1,6 bilhões por ano, enquanto o Canadá tem perda anual menor, mas ainda significativa, de aproximadamente U\$ 100 milhões por ano (DEPURU, 2012).

Apesar de não ser um problema apenas de países pobres, nestes locais a ocorrência é mais evidenciada. A Malásia, por exemplo, atingiu um recorde de perda por furto de energia elétrica em 2004. Este recorde negativo custou à companhia Tenaga Nasional Berhad cerca de U\$ 229 Milhões. A Índia se compara ao valor de perdas não técnicas do Brasil, atingindo um valor na faixa de U\$ 4,5 bilhões por ano (DEPURU, 2012).

A redução das perdas não técnicas tem um incentivo natural relacionado ao aumento de receita, por parte das concessionárias. No entanto, o impacto das perdas no sistema elétrico pode ser ainda mais relevante. Para o sistema elétrico, as perdas não técnicas implicam na imprudência no consumo de energia, uma vez que o consumidor que não paga a energia que consome não tem compromisso com a racionalidade do seu uso. Indiretamente, o uso irracional leva a uma necessidade cada vez mais crescente da disponibilidade do recurso, com a construção de novas usinas e a ampliação da capacidade das redes. Além disso, existe o impacto na qualidade de energia, tendo em vista que o consumo irregular pode ocasionar sobrecarga nos transformadores e linhas de distribuição, interrupções frequentes nos circuitos dos alimentadores, e queda de tensão para os consumidores adimplentes.

2.2 Detecção de perdas não técnicas

Embora a pesquisa relacionada a perdas não técnicas não seja recente, ela tem ganhado maior enfoque com o aumento da disponibilidade de recursos tecnológicos e computacionais, que permitem a detecção das perdas.

O procedimento mais elementar para detecção de perdas não técnicas é a inspeção local. No entanto, há uma série de fatores limitantes a essa prática. Destaca-se o custo com equipes de inspeção, o tempo necessário para inspecionar grandes áreas e a própria dificuldade de inspeção, uma vez que as situações irregulares nem sempre são perceptíveis com uma simples inspeção visual.

Devido a essas dificuldades, diversos métodos são propostos, para auxiliar equipes de inspeção, limitando as regiões de buscas e, em alguns casos, indicando consumidores suspeitos de fraudes. A seguir, são apresentadas algumas técnicas utilizadas para diagnóstico das perdas não técnicas encontradas na literatura.

2.2.1 Estimação de perdas não técnicas a partir das perdas globais e técnicas

Devido ao fato das perdas não técnicas não serem medidas diretamente nos locais em que ocorrem, o seu montante deve ser estimado. Uma possibilidade é a estimação a partir de uma análise dos alimentadores de distribuição, conhecendo-se seu modelo elétrico (impedâncias da rede e cargas de consumidores regulares) e os totais de energia suprida e faturada.

Da diferença entre a energia suprida e faturada, obtém-se uma indicação das perdas globais, e a partir do modelo elétrico da rede, pode-se obter uma indicação aproximada das perdas técnicas. As perdas não técnicas podem, então, ser estimadas pela diferença entre as perdas globais e as perdas técnicas. A Figura 2.3 ilustra esse conceito. A comparação entre energia suprida e faturada é usualmente conhecida como balanço de energia.

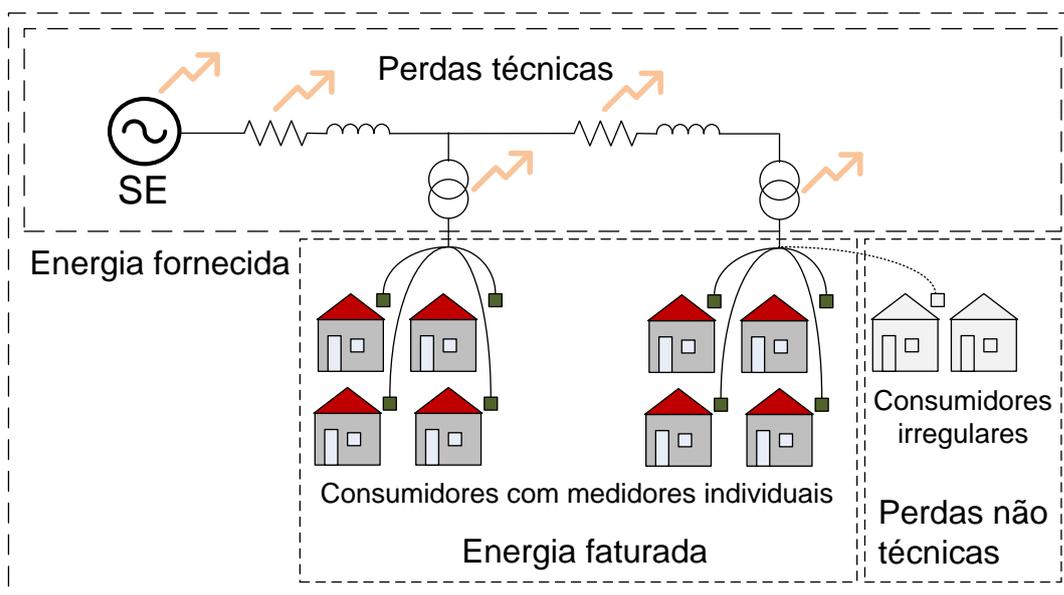


Figura 2.3 – Esquema de uma rede elétrica com perdas técnicas e não técnicas

Fonte: Elaborado pelo autor

Essa metodologia necessita que as perdas técnicas da rede sejam conhecidas. Essas perdas são intrínsecas ao Sistema de Distribuição e devidas principalmente à resistência de condutores elétricos, que transforma a energia

elétrica em calor quando a corrente flui através deles. A perda de energia decorrente da dissipação de calor em redes elétricas (linhas, cabos, transformadores e outros elementos) é normalmente referida como física ou ôhmica.

As perdas técnicas variam de acordo com o nível de utilização da capacidade e qualidade da rede, ou seja, a quantidade de energia a ser transmitida e distribuída. Em particular, elas são proporcionais ao quadrado da corrente. As redes de distribuição em tensões mais baixas estão sujeitas a um nível mais elevado de perdas, pois a corrente é mais elevada nesses trechos. Além disso, as perdas técnicas também são dependentes do comprimento e da seção transversal da linha de rede (ERGEG, 2008).

Oliveira et al. (2001) apresentam uma metodologia para cálculo de perdas técnicas nos sistemas de distribuição, que consiste na separação dos sistemas em segmentos, cada qual com um procedimento específico para determinação de perdas. Por exemplo, para perdas em medidores de energia, foi adotado um valor fixo, de acordo com o tipo de medidor; para o cálculo de perdas na rede de baixa tensão, foi utilizado o cálculo do fluxo de potência para diferentes patamares de carga da curva diária de demanda e a determinação das perdas nos trechos de rede, de acordo com a resistência dos condutores. A partir do balanço de energia, sugere-se a possibilidade de se determinar perdas não técnicas. A Tabela 2.2 apresenta valores típicos de perdas técnicas por segmentos em redes elétricas brasileiras, de acordo com os autores.

Tabela 2.2 – Perdas técnicas típicas em redes elétricas no Brasil

Segmento	Perda técnica típica
Subtransmissão	2,0 – 3,0 %
Subestação de Distribuição	0,5 – 2,0 %
Rede primária	0,5 – 2,5 %
Transformadores de Distribuição	1,0 – 2,0 %
Rede secundária	0,5 – 2,0 %
Conexões dos consumidores	0,05 – 0,15 %
Medidores de energia	0,2 – 0,4 %
Outros	0,2 – 0,8 %

Fonte: Oliveira et al. (2001)

A metodologia proposta por Méffe (2002) considera a utilização de dados de medições na rede para reajustar uniformemente as curvas de cargas típicas dos alimentadores, de forma a calcular perdas técnicas com maior precisão.

Celik (1999) destaca a importância de se considerar as características reais dos alimentadores, como potências ativa e reativa, tensão e corrente. Com isto, o resultado do cálculo do fluxo de potência se torna mais confiável. Esse sistema também permite a obtenção de curvas de perdas ao longo de um determinado período e para um determinado componente, aumentando a precisão da obtenção de índices de perdas técnicas.

Para que seja possível estimar as perdas técnicas de maneira satisfatória, é importante que a base de dados seja detalhada e confiável, o que nem sempre é possível. Com isso, se faz necessário buscar métodos que necessitem de uma quantidade menor de dados. Segundo Dortolina (2005), existem basicamente três tipos de métodos de análise:

- *Bottom-Up*: utilizado em situações que se conhece o sistema de maneira completa, possuindo informações detalhadas da rede. Neste método existe uma maior precisão nos cálculos, através de simulação do fluxo de potência.

- *Top-Down*: utilizado quando não se possui muitos dados, realizando a estimativa das perdas através de comparação com outros sistemas. Tende a produzir estimativas mais rápidas, porém com um grau mais elevado de incertezas.

- Sistema Híbrido: como sugere o nome, é uma mescla dos dois métodos anteriores. Parte do pressuposto que existam dados suficientes para algumas regiões do sistema, enquanto que para outros não. A principal vantagem desta abordagem é que ela leva menos tempo do que a abordagem *bottom-up* e é mais precisa do que a abordagem *top-down*. Tem como desvantagem requerer informações mais detalhadas do que os métodos *top-down*, porém produzindo um resultado menos preciso que os métodos *bottom-up*.(DORTOLINA, 2005).

A estimação das perdas não técnicas a partir da diferença entre perdas globais e técnicas é uma maneira clássica de se obter um indicativo da ocorrência do problema na rede. Ela não evita, contudo, o alto custo com equipes para percorrer os trechos de rede, fiscalizando e identificando os locais irregulares. Além disso, há que se considerar a dificuldade do cálculo preciso das perdas técnicas na rede, incluindo, por exemplo, perdas em reguladores de tensão, isoladores,

conexões da rede, bancos de capacitores, etc. Algumas técnicas são empregadas para melhorar os resultados dessa análise primária, destacando-se métodos baseados em Estimação de Estados, em perfis típicos de consumo, e no investimento em equipamentos de monitoramento, conforme apresentado a seguir.

2.2.2 Estimação de perdas a partir de Estimadores de Estado das Redes de Distribuição

Nos trabalhos de detecção de perdas em sistemas de potência, a estimação de estados é utilizada como parte de um processo que visa obter um modelo da rede mais aproximado da realidade. Na forma mais básica, calcula-se o fluxo de potência na rede e ajusta-se linearmente as demandas com base em valores medidos.

Méffe (2002) propõe uma metodologia que inclui as perdas técnicas como parte do fator de correção da demanda em um estimador de estados.

Cruz, Quintero e Pérez (2006) apresentam uma metodologia baseada na estimação de estados para detecção das perdas não técnicas. A estimação é feita em tempo real, para cada ponto com transformador considerado relevante na rede primária. A análise parte de um macro balanço de energia, para detecção de regiões afetadas por perdas não técnicas. A potência estimada fornecida em cada transformador é calculada a partir de um modelo matemático, que inclui consumo faturado, perdas técnicas, e um fator de comportamento, de acordo com o tipo de consumidor, entre outras variáveis. A metodologia empregada pelos autores é baseada na modelagem de carga para estimação de estados apresentada por Ghosh, Lubkeman e Jones (1997).

Donadel et al. (2009) utilizam a estimação de perdas não técnicas para o refinamento do cálculo de perdas técnicas. As perdas não técnicas são calculadas utilizando três procedimentos: inspeções por amostragem, seleção de grupos de consumidores para ajustes de consumo e aplicação de um fator de correção para ajuste de consumo. Os três procedimentos combinados, produzem uma curva de demanda ajustada para os grupos de consumidores, e então é aplicado o método apresentado por Oliveira et al. (2001), descrito na seção anterior.

Métodos estatísticos são aplicados para refinar os resultados dos estimadores, entre os quais se destaca o método dos Mínimos Quadrados

Ponderados, cuja solução é buscada de forma recursiva e através do algoritmo iterativo de Gauss-Newton. A base teórica desse método aplicada a Sistemas de Potência é apresentada nos trabalhos de Schweppe (1970 a,b,c). Ferreira Neto (2011) emprega essa metodologia e introduz pseudo medidas a partir de dados de faturamento dos consumidores, para estimar o estado da rede em um sistema de detecção de perdas não técnicas. O autor divide o estimador de estados em três submodelos, sendo um para cada fase da rede trifásica.

Outras técnicas estatísticas apresentadas por ZARCO P. e EXPÓSITO A.G., 2001, incluem o estimador de estado de Mínimo Erro Absoluto Ponderado e da Mínima Mediana Ponderada de Quadrado dos Resíduos. Esses métodos são propostos para sistemas em que ocorrem erros grosseiros de medições.

Métodos heurísticos e de inteligência artificial também são empregados para estimação de estados, destacando-se os sistemas Fuzzy e as Redes Neurais. Os sistemas Fuzzy caracterizam-se pela capacidade de representação numérica da informação descrita em uma linguagem naturalmente imprecisa ou incerta (SARIC e CIRIC, 2003). As Redes Neurais serão tratadas com maior profundidade no Capítulo 4 dessa Dissertação. Elas são representadas por sistemas de neurônios, que recebem e processam informação, com capacidade de aprendizagem, de maneira a simular um cérebro (HAYKIN, 2001), como mostra a Figura 2.4.

O Quadro 2.1 apresenta, de uma forma resumida, uma relação de trabalhos que abordam a estimação de estados em Sistemas de Potência.

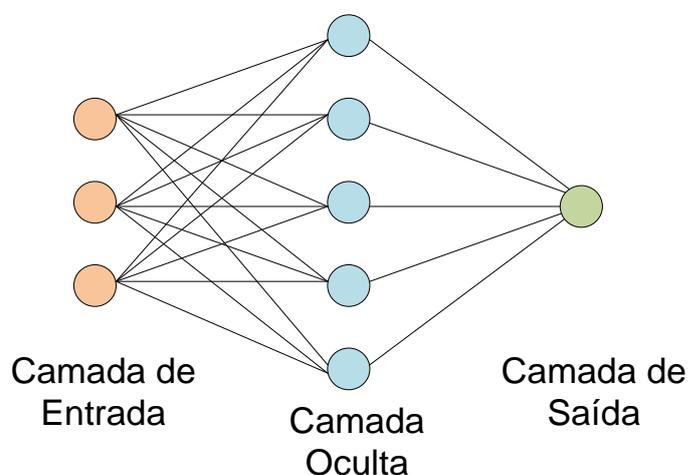


Figura 2.4 – Representação de uma Rede Neural Artificial

Fonte: Haykin, 2001

Método / Técnica	Referência
Estatístico – Básico	Deng, He e Zangh (2002); Méffe (2002)
Estatístico – Mínimos Quadrados Ponderados	Monticelli (2000); Ferreira Neto (2011); Lo, Huang e Lu (2012, 2013)
Estatístico – Mínimo Erro Absoluto Ponderado	Kotyuga e Vidyasagar (1982); Falcão e Assis (1988); Ramírez e Barocio (2000)
Estatístico – Mínima Mediana Ponderada de Quadrado dos Resíduos	Nanni (2009)
Redes Neurais	Abbasy e Hassawy (1996); Antonio (2001); Shafiu (2005); Rakpenthai (2005)
Fuzzy	Zimmerman (1984); Saric e Siric (2003);
Híbridos e Outros	Kennedy e Eberhart (1995); Naka (2001); Naka, Genji e Fukuyama (2003);

Quadro 2.1 – Relação de trabalhos que empregam Estimação de Estados no Sistema de Distribuição.

Fonte: Elaborado pelo autor

2.2.3 Detecção de perdas não técnicas a partir dos perfis de consumidores

A partir de um determinado padrão de consumo de energia, é possível identificar situações discrepantes, que podem levar à identificação de fraudes ou erros de medições. Nessa metodologia, é bastante usual o emprego de técnicas estatísticas e de inteligência artificial, que são utilizadas como parte do processo de identificação de perdas, para diferentes finalidades; destacam-se, entre essas, a mineração de dados e a classificação.

A determinação de um perfil padrão de consumo deve levar em conta diversos aspectos, sendo as características da curva de demanda o principal deles.

No entanto, os hábitos de consumo estão sujeitos a variações, dependendo do dia da semana, período do ano, clima, região, entre outros. Por isso, são necessárias técnicas de seleção dos dados que influenciam no consumo, seleção de clientes e classificação dos perfis das curvas de demanda. Em um conjunto grande de dados, por exemplo, a seleção de clientes pode ser interpretada como um filtro, que pode ter como objetivos excluir clientes repetidos, sem consumo ou que foram cadastrados após o início das análises.

Nizar, Zhao e Dong (2006) apresentam uma metodologia para detectar e estimar perdas não técnicas baseadas no desvio de comportamento dos consumidores e a partir das curvas de cargas típicas. Para determinação do perfil de carga e caracterização dos consumidores, os autores avaliam diferentes ferramentas de seleção e tratamento de dados.

As técnicas de seleção de recursos têm como objetivo eliminar variáveis ou atributos irrelevantes de maneira a melhorar o desempenho dos algoritmos de modelagem ou aprendizagem. Entre as técnicas, pode-se citar: busca completa, busca heurística, algoritmos genéticos, busca randômica e busca gradual (CELIK e LIU, (1999). A caracterização dos consumidores consiste em agrupar perfis típicos de carga de acordo com índices pré-determinados (*a priori*, como tipo de atividade econômica e demanda contratada) ou de campo (geralmente a partir das curvas de carga diárias). Para Nizar, Zhao e Dong (2006), uma curva de carga diária está sujeita a variações de acordo com diferentes condições. Os autores ainda consideram outras características para caracterização dos consumidores, como fator de carga e período do dia (manhã, tarde e noite).

A partir da classificação, os autores (id. 2009) analisaram diferentes consumidores, e propuseram a identificação de perdas não técnicas a partir de anormalidades nas curvas de demanda, estabelecendo limites máximos e mínimos para as diferentes situações apresentadas no Quadro 2.2.

Condição a ser considerada	Classe
Tipo de consumidor	Residencial Comercial Industrial
Local	Urbana Rural
Nível de tensão	Baixa Média/Alta
Clima	Temperatura Precipitação pluviométrica Umidade Pressão atmosférica Velocidade do vento Visibilidade
Tipo de dia	Dia útil Sábado Domingo / Feriado

Quadro 2.2 – Fatores que podem influenciar na curva de carga de um consumidor

Fonte: Nizar, Zhao e Dong (2006)

Nagi et al. (2008) utilizam uma técnica baseada em inteligência artificial denominada Máquina Vetorial de Apoio (do inglês, *Support Vector Machine (SVM)*). Os padrões de consumo são obtidos a partir de mineração de dados e métodos estatísticos. A técnica SVM é baseada na teoria de aprendizagem estatística, e tem aplicação em ferramentas de classificação. O algoritmo utilizado pelos autores classifica os consumidores em uma de quatro classes: suspeito de fraude confirmado, suspeito de fraude não confirmado, suspeito de regular confirmado e suspeito de regular não confirmado. O sistema desenvolvido foi testado em redes reais de uma concessionária de energia da Malásia, e os resultados mostram que o programa alcançou uma taxa de acerto de 53%.

Méffe e Oliveira (2002) propõem a utilização de medições para correção de cálculos de perdas técnicas. Eles analisam a origem de discrepâncias entre valores reais e estimados pela metodologia proposta. Algumas situações podem impactar

significativamente os resultados, destacando-se: a reconfiguração da rede, erros na classificação dos consumidores (comercial, residencial ou industrial), e diferenças entre períodos de leitura para faturamento. Os autores apresentam um fator de correção para essas situações.

Ramos et al. (2013) introduzem a técnica OPF – *Optimum-path forest*, para identificação de perdas não técnicas. Essa técnica é baseada em grafos e é utilizada para reconhecimento de padrões. Os autores comparam o desempenho do método proposto com outras técnicas (SVM e RNA) e demonstram precisão e rapidez de processamento dos resultados usando OPF.

Monedero et al (2006) apresentam um projeto, denominado MIDAS, no qual são empregadas Redes Neurais Artificiais e também métodos estatísticos para detecção de perdas não técnicas. As RNAs e os métodos estatísticos são empregados para mineração de dados. Foram utilizadas Redes Kohonen, que permitem o agrupamento de dados em mapas auto organizáveis, como mostra a Figura 2.5. As variáveis de entrada utilizadas no projeto foram:

- Valores máximos e mínimos da conta de energia do consumidor, no ano anterior;
- Consumo médio de energia, a partir das tarifas, no ano anterior;
- Diferença entre as médias de consumo de energia, entre os consumidores analisados;
- Diferença entre o consumo de um mês, para um consumidor, e o consumo médio e máximo do mês dos consumidores analisados;
- Relação entre consumo e demanda contratada.

Os métodos estatísticos usados no projeto têm como objetivo identificar discrepâncias e fornecer uma lista de consumidores atípicos, usando como referência um consumidor geral padrão.

Os dois métodos empregados pelos autores (id. 2006) foram testados em uma rede real, e alcançaram uma taxa de acerto de 50%.

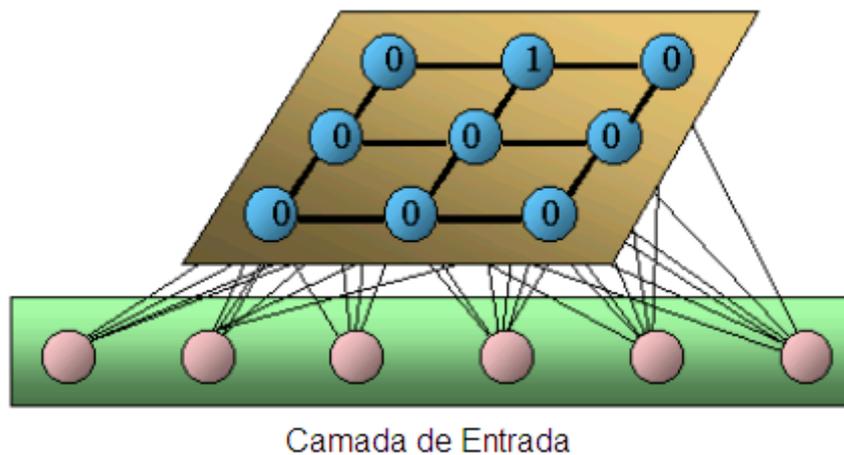


Figura 2.5 – Rede Neural Artificial – Estrutura de Kohonen

Fonte: Monedero et al (2006)

Pereira et al (2013) utilizam Redes Neurais Multilayer Perceptron, visto na figura 2.6, para classificação, combinadas com uma técnica denominada Busca por Sistema de Cargas (CSS – *Charged System Search*), que é baseada no conceito de interação entre partículas eletricamente carregadas.

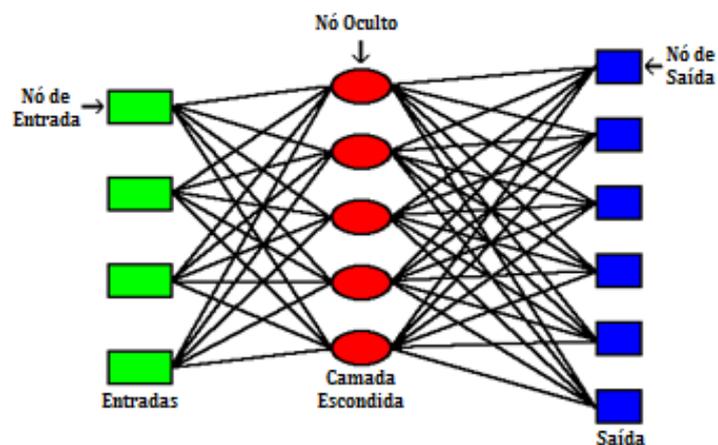


Figura 2.6 – Rede Neural Artificial – Estrutura Multilayer Perceptron

Fonte: Pereira et al (2013)

Guerrero et al (2010) empregam RNA e métodos estatísticos para identificar perfis de consumos considerados anormais, em um grupo de mais de 50.000

consumidores com demanda contratada superior a 15 kW, em Andaluzia (Espanha). O processo de aprendizagem é alimentado com informações utilizando processamento de linguagem natural (NPL – *Natural Language Processing*), que inclui dados geográficos, datas e comentários de inspetores e de equipes de medição.

Alguns trabalhos relacionam o consumo de energia elétrica com as condições climáticas, embora não seja dada ênfase ao aspecto das perdas de energia. Altran (2005) demonstra em seu trabalho a influência de fatores climáticos como temperatura, nebulosidade, velocidade do vento, entre outros, para a modelagem da previsão de carga. O autor menciona que a influência de tais fatores é fundamental na representação do comportamento da carga, utilizando esses dados como entrada dos algoritmos de previsão, que empregam RNAs.

Fatores climáticos, como temperatura, umidade relativa do ar, velocidade do vento e nível de luminosidade possuem influência nas previsões de curto prazo, sendo a temperatura o principal fator. A distribuidora de energia elétrica deve conhecer cada região da sua concessão para então avaliar as influências da temperatura sobre a demanda solicitada ao seu sistema.

O sucesso da aplicação dos métodos baseados nos perfis de consumo depende da correta modelagem das curvas de demanda, considerando a diversidade de influências a que essas curvas estão sujeitas. No nível da curva de demanda de um alimentador, um consumidor irregular pode passar despercebido; mas no nível individual, considerando, por exemplo, os históricos das faturas de energia elétrica, os métodos baseados no perfil de consumo podem indicar pontualmente consumidores fora dos padrões esperados, o que pode ser um indicativo de ocorrência de perdas não técnicas.

2.2.4 Monitoramento de perdas com medidores eletrônicos

Com o avanço das Redes Elétricas Inteligentes, a utilização de medidores eletrônicos inteligentes e estruturas avançadas de medição (AMI – *Advanced Metering Infrastructure*) tende a ser mais comum. Esses equipamentos e

tecnologias podem ser utilizados para a detecção de perdas técnicas e não técnicas nas redes de distribuição.

No trabalho de Alves et al. (2006), a solução proposta para a redução das perdas não técnicas é baseada na modernização dos sistemas de medição e gerenciamento das faturas de energia elétrica. Os autores destacam que, em muitos casos, as perdas não técnicas são devidas a equipamentos obsoletos e à ineficiência dos serviços relacionados ao faturamento.

Paruchuri e Dubey (2012) propõem a utilização de medição inteligente (do inglês, *smart metering*) no nível residencial e em equipamentos da rede, como relés e transformadores. Com a medição, é possível um balanço contínuo de energia entre pontos da rede e consumidores. Um sinal de verificação com frequência específica é injetado na rede, entre os equipamentos com medição, e deve ser anulado por medidores regulares. No caso de furto de energia ou mau funcionamento de medidores, é possível identificar o ponto de ocorrência, pelo sinal de verificação.

A instalação de medidores de totalização (macro medidores ou medidores coletivos) é mencionada por Agüero (2012), como uma alternativa para a identificação e redução de perdas não técnicas. Com os medidores, é possível fazer a comparação entre energia fornecida e faturada (balanço de energia) em trechos menores dos alimentadores, diminuindo, dessa forma, a região de busca por fraudes. Esse recurso é utilizado no trabalho apresentado por Gonzalez-Urdaneta (2010), que mostra ações feitas por uma concessionária de energia da Venezuela para identificação e redução de perdas não técnicas.

Lo, Huang e Lu (2012, 2013) combinam a técnica de Estimadores de Estados com Mínimos Quadrados Ponderados, apresentada na Seção 2.2.2, com a utilização de AMIs para ajuste do estimador da rede.

No trabalho de Faria et al. (2013), a identificação de perdas não técnicas é feita através de dispositivos de Resposta da Demanda (DR – *Demand Response*), um recurso que permite o monitoramento local de cargas dos consumidores, e que também está incluído nos conceitos das Redes Elétricas Inteligentes.

A implantação de medidores eletrônicos e estruturas de medição avançadas implica na necessidade de investimentos para a modernização dos sistemas de medição e faturamento das concessionárias de energia. O planejamento desses investimentos deve ser cuidadosamente elaborado, para obter a melhor relação

custo-benefício. Ribeiro et al. (2012) apresentam uma metodologia para otimização dos investimentos em medidores totalizadores e AMIs em uma região atendida pela Light (Rio de Janeiro). Eles consideram restrições como número mínimo de consumidores por transformador e número máximo de inspeções locais permitidas pelo orçamento disponível.

2.3 Particularidades do consumo de energia em áreas rurais

As redes elétricas que atendem consumidores em áreas rurais têm características bastante diferenciadas das redes localizadas em centros urbanos. De modo geral, as principais diferenças a serem destacadas, considerando o sistema brasileiro, são:

- Maiores extensões de rede, podendo chegar a dezenas de quilômetros, com menor densidade populacional por quilômetro de rede;
- Sazonalidade das curvas de cargas, impactadas especialmente em regiões com intensa atividade agrícola e agroindustrial; isso reflete em um baixo fator de carga dos transformadores. A Eletrobrás utiliza 0,20 como fator de carga de áreas rurais, considerando inclusive as grandes propriedades rurais que possuem grande consumo de energia;
- Dificuldade de manutenção de padrões de qualidade de energia, sendo necessário, em muitos casos, a utilização de equipamentos para compensação das quedas de tensão nos longos trechos de rede. Ressalta-se, também, que as redes rurais são mais susceptíveis a falhas, devido a uma série de fatores, como dificuldade de monitoramento das condições físicas da rede, limitação de investimentos em equipamentos, maior extensão de rede, entre outros. Além disso, algumas soluções implementadas em redes urbanas são impraticáveis nas redes rurais, como a transferência de cargas para isolar uma determinada falha.

Kersting e Phillips (1992) e Freeman et al. (2009) apresentam algumas considerações sobre redes rurais nos Estados Unidos, mencionando alguns aspectos e problemas similares encontrados no Brasil.

Em áreas rurais, existe uma grande dispersão de consumidores, variando de 0,1 a 5,0 consumidores por quilômetro de rede (RIBEIRO et al, 2012). Segundo a COELBA 2012, de 120.166 domicílios atendidos pelo programa Luz no Campo, 61,8% estavam situados no raio de um quilômetro da rede existente, enquanto 25,7% estavam localizados entre um e cinco quilômetros da rede de energia.

No nível residencial, a utilização da energia elétrica nas áreas rurais é primordialmente para iluminação, embora o uso de eletrodomésticos nesse meio seja crescente. No nível agroindustrial, a eletricidade é utilizada na produção rural em sistemas de bombeamento de água para irrigação, equipamentos utilizados para tratamento após a colheita, como ventiladores para aeração em secadores, por exemplo, ferramentas elétricas, elevadores de carga, esteiras transportadoras, compressores para refrigeração, entre outros.

O escopo dessa Dissertação abrange consumidores rurais com sistemas de irrigação para lavouras de arroz. Um típico sistema é ilustrado na Figura 2.7, sendo que a principal carga a ser considerada é o motor da bomba de irrigação. Esse tipo de sistema é conhecido como levante, no qual a água é bombeada de reservatórios de água, rios ou lagos, e distribuída para a lavoura através de canais. O trabalho de Behrens (2013) apresenta uma descrição de técnicas de sistematização do solo e irrigação das lavouras, com ênfase na automação e eficiência energética do sistema elétrico de alimentação das bombas.



Figura 2.7 – Bombas de irrigação para lavouras de arroz

Fonte: AES SUL (2006)

Os motores elétricos utilizados nesses sistemas de irrigação são motores de indução trifásicos, em instalações com potências que variam tipicamente entre 100 e 300 CV (AES SUL, 2006). O regime de funcionamento é praticamente contínuo no período de irrigação. Isso significa que, em condições normais, não existe um controle automático da irrigação, e as bombas permanecem ligadas por aproximadamente 21 horas por dia, sendo desligadas apenas nos horários de ponta; isso representa um grande desperdício de energia elétrica e água. Behrens (2013) propõe uma metodologia de monitoramento das condições climáticas para acionar as bombas com rotação variável, de acordo com índices pluviométricos e previsão de clima. Mesmo que haja expectativa de redução de consumo, esse tipo de sistema não é útil para lavouras de arroz no Brasil.

Cerca de 50% da produção nacional de arroz é proveniente do Rio Grande do Sul, segundo dados do IBGE 2013 – Instituto Brasileiro de Geografia e Estatística. As regiões com maior quantidade de área cultivada situam-se nas áreas de concessão das empresas AES SUL (Fronteira Oeste, Campanha e Centro) e CEEE (Litoral). O período de irrigação nesse Estado ocorre tipicamente entre os meses de setembro e janeiro.

2.4 Considerações finais

Na revisão bibliográfica apresentada nesse capítulo, foram relacionados métodos de identificação de perdas não técnicas nos sistemas de distribuição de energia elétrica. De modo geral, os métodos podem ser divididos em quatro categorias, com as seguintes características:

1) Inspeção Local: maior custo com recursos humanos e dificuldade de execução;

2) Baseados em Estimação de Estados: fortemente dependentes da correta modelagem da rede; as perdas não técnicas são estimadas indiretamente, a partir das perdas técnicas; não há indicação pontual de consumidores irregulares, mas sim de regiões com ocorrência de problemas;

3) Baseados no Perfil de Consumo: sujeitos a fatores que afetam o comportamento da curva de demanda, como local, clima e sazonalidade; permitem a identificação pontual de consumidores atípicos;

4) Baseados na Instalação de Medidores: prática que tende a se tornar mais comum, com o avanço da tecnologia e das redes inteligentes; necessidade de investimentos e modernização de equipamentos.

Ressalta-se que é possível uma solução híbrida, que utilize elementos de dois ou mais métodos simultaneamente.

Na revisão bibliográfica, constata-se que a utilização de Redes Neurais Artificiais e métodos estatísticos têm grande aplicação como parte de processos das metodologias de detecção de perdas baseadas em estimadores de estado e em perfis de consumo. Entre as finalidades das RNAs e métodos estatísticos, destacam-se a mineração de dados, a seleção de variáveis e a classificação.

Nessa Dissertação, é proposto um método para identificação de perdas não técnicas em consumidores rurais. Esse tipo de consumidor possui particularidades, apresentadas na Seção 2.3, especialmente em instalações com sistemas de irrigação, que são o foco desse trabalho.

No sistema proposto, serão testados métodos estatísticos e RNAs, baseados na caracterização do perfil de consumo de energia, tendo em vista que um dos objetivos do trabalho é a identificação pontual de perdas não técnicas. Considerando as características dos sistemas de irrigação, serão utilizados dados de consumo, área cultivada e dados climatológicos.

Durante a pesquisa na bibliografia publicada, não foram encontrados trabalhos com o enfoque proposto nessa Dissertação, o que demonstra o caráter pioneiro desse trabalho.

3 FUNDAMENTAÇÃO TEÓRICA DOS MÉTODOS APLICADOS

Com base na Revisão Bibliográfica apresentada no Capítulo 2, dentre as técnicas escolhidas para o desenvolvimento da metodologia proposta nessa Dissertação, adotou-se a Regressão *Stepwise* e o emprego de Redes Neurais Artificiais. Neste Capítulo é apresentada uma revisão teórica dos principais conceitos e avaliação técnica, a serem empregadas nesse trabalho.

3.1 Análise Estatística – Seleção e Regressão Stepwise

Um dos ramos mais importantes e práticos da Estatística é a Análise de Regressão. Em síntese, o objetivo fundamental de uma análise de regressão consiste em estabelecer relações entre variáveis de entrada e saída, de maneira que se possa explicar ou fazer previsões sobre o comportamento da saída a partir de observações das entradas. Esse conceito reforça a hipótese de utilizar um modelo de regressão nesse trabalho, no qual se deseja investigar a relação de variáveis climatológicas e de características de lavouras com o consumo de energia elétrica.

De modo geral, um modelo de regressão linear pode ser caracterizado por uma variável dependente Y e suas relações com um conjunto de variáveis independentes X , também chamadas de variáveis explicativas ou regressores. As relações, na maioria das aplicações, podem ser descritas por uma relação linear – ou ainda serem linearizadas – de acordo com forma apresentada na Equação 3.1:

$$Y(x_1, x_2, \dots, x_K) = \beta_0 + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + \dots + \beta_K \cdot x_K \quad (3.1)$$

Os coeficientes constantes β relacionam as entradas com a saída e são denominados coeficientes de regressão. O desenvolvimento matemático e estatístico para determinação desses coeficientes não está no escopo desse trabalho, mas pode ser encontrado em livros de Estatística e Análises de Regressão Linear ((WALPOLE e MYERS, 2009); (MONTGOMERY e RUNGER, 2009)).

Nos sistemas em que a análise de regressão múltipla envolve um conjunto grande de variáveis, pode ser necessária a utilização de um método para seleção das variáveis independentes que deverão ser utilizadas no modelo. A seleção de variáveis reduz a necessidade computacional dos cálculos de regressão, bem como a necessidade de aquisição de novos dados para atualização ou manutenção do modelo. Além disso, pode-se demonstrar que, em alguns casos, modelos mais próximos da realidade podem ser obtidos excluindo-se variáveis de entrada do processo de análise (SEBER e LEE, 2003).

A experiência prévia de especialistas é um fator que pode ajudar no processo de seleção, mas, de modo geral, quando há um grande número de variáveis e uma grande quantidade de dados a serem analisados, as técnicas computacionais podem ser mais efetivas.

De acordo com Seber e Lee (2003), à medida que o número de variáveis a serem analisadas aumenta, os métodos baseados na análise “passo a passo” podem ser mais apropriados, reduzindo os esforços computacionais. Dentre esses métodos, destaca-se o método “*Stepwise*”, discutido a seguir. Como esse método de seleção está diretamente relacionado ao processo de regressão, também é denominado de Regressão *Stepwise*.

3.1.1 Algoritmo de Regressão Stepwise

A Regressão *Stepwise* foi inicialmente descrita por Efroymsom (1960). Ela é baseada na aplicação conjunta de duas técnicas de seleção: a inclusão (*Forward Selection*) e a exclusão (*Backward Elimination*) de variáveis, passo a passo, no modelo. Dessa forma, ela requer que dois níveis de significância sejam estabelecidos, sendo um para cada técnica.

Assume-se que a função de regressão seja descrita pela Equação 3.2:

$$Y_i = \beta_i \cdot X_i + \varepsilon_i \quad (3.2)$$

na qual:

$i = 1 \dots n$, sendo n o número de análises computadas;

Y = variável de saída (função prevista);

β = coeficientes de regressão;

X = variável de entrada;

ε = erro independente.

A solução utiliza a função estatística F apresentada na Equação 3.3:

$$F = \frac{RSS_p - RSS_{p+1}}{RSS_{p+1}} \cdot (n - p - 1) \quad (3.3)$$

Sendo $p = K$, o número de variáveis totais do sistema, e RSS (*residual sum of squares*) o valor da soma dos erros quadráticos. RSS pode ser determinada computacionalmente pelo somatório dos quadrados das diferenças entre o valor real da função a ser prevista e o valor calculado através de (3.1) (SEBER e LEE, 2003, pg. 38). De modo alternativo, outros parâmetros podem ser utilizados para testar se uma variável é importante na definição do modelo, citando-se o coeficiente de determinação (R^2) e a análise de variância.

- Procedimento para *Forward Selection* (FS)

Nessa técnica, o modelo inicia sem nenhuma variável. Calcula-se o valor de F (Equação 3.3) com $p = 1$ para todas as variáveis, e guarda-se (mantém no modelo) a que apresentar o maior valor. Repete-se o procedimento para $p = 2$, selecionando novamente a variável que resultou no maior incremento de F . Assim, sucessivamente ($p = 3, 4\dots$), vão sendo adicionadas no modelo as variáveis que resultam no maior valor de F , até que não seja mais percebida uma variação significativa nessa função. Define-se, previamente, um nível de significância, ou valor de referência, $F_{\text{Ref-FS}}$, acima do qual a função F calculada deve estar para que a variável seja incluída no modelo. Ressalta-se que, na aplicação da técnica, uma vez inserida no modelo, a variável não é mais excluída.

- Procedimento para *Backward Elimination* (BE)

Ao contrário da técnica FS, na técnica BE todas as variáveis são inicialmente incluídas no modelo. A função F é calculada para $p = K$, sendo K o número total de variáveis, e p vai diminuindo a cada análise. Em cada análise, exclui-se do modelo a variável que resultou no menor valor de F , considerando ainda que F esteja abaixo de um nível de significância mínimo de referência ($F_{\text{ref-BE}}$). Uma vez excluída do modelo, a variável não é mais reinserida.

- Procedimento para Regressão *Stepwise*

Na Regressão *Stepwise*, o procedimento FS é aplicado inicialmente para selecionar uma variável que é inserida no modelo, e é sucedido pelo procedimento BE, para selecionar uma variável que será excluída do modelo. Os procedimentos FS e BE são aplicados alternadamente, considerando-se que para ser incluída no modelo, a variável deve ter $F > F_{\text{Ref-FS}}$, e para ser excluída, a variável deve ter $F < F_{\text{Ref-BE}}$. O procedimento da Regressão *Stepwise* é encerrado quando a adição de variáveis no modelo, definida pelo procedimento FS, for encerrada. Para a correta aplicação do método, deve-se garantir que $F_{\text{Ref-FS}} \geq F_{\text{Ref-BE}}$.

3.1.2 Exemplo numérico

Com o objetivo de exemplificar a aplicação passo a passo da Regressão *Stepwise*, considere o sistema hipotético descrito a seguir, no qual os dados apresentados nos resultados são apenas para ilustração do método.

Sistema Exemplo:

- Variável de Saída: Y
- Variáveis de Entrada: x_1, x_2, x_3 e x_4
- $F_{\text{Ref-FS}} = F_{\text{Ref-BE}} = 0,50$

A sequência de resultados e análise pode ser acompanhada na Tabela 3.1.

Tabela 3.1 – Resultados hipotéticos para exemplo de análise da Regressão *Stepwise*

1º Passo: analisar $Y(x_1)$, $Y(x_2)$, $Y(x_3)$ e $Y(x_4)$.						
	x_1	x_2	x_3	x_4		
β	1,55	- 1,83	0,75	0,56		
F	0,72	0,65	0,82	0,60		
2º Passo: Adicionar x_3 ao modelo, e analisar $Y(x_3x_1)$, $Y(x_3x_2)$, $Y(x_3x_4)$.						
	x_3	x_1	x_3	x_2	x_3	x_4
β	0,71	1,33	0,65	- 1,50	0,80	0,47
F	0,85	0,80	0,41	0,32	0,85	0,78
3º Passo: Adicionar x_1 ao modelo, e analisar $Y(x_3x_1x_2)$ e $Y(x_3x_1x_4)$.						
	x_3	x_1	x_2	x_3	x_1	x_4
β	0,35	1,32	- 1,34	0,72	0,99	0,22
F	0,47	0,69	0,73	0,65	0,67	0,70
4º Passo: Adicionar x_2 ao modelo e testar modelo prévio sem x_3 (analisar $Y(x_1x_2)$).						
	x_1	x_2				
β	1,36	-1,77				
F	0,97	0,98				
5º Passo: Remover x_3 do modelo, e analisar $Y(x_1x_2x_3)^*$ e $Y(x_1x_2x_4)$.						
	x_1	x_2	x_3	x_1	x_2	x_4
β	1,32	-1,34	0,35	1,47	-1,68	0,19
F	0,69	0,73	0,47	0,90	0,85	0,39
6º Passo: Resultado final**.						
	x_1	x_2				
β	1,36	-1,77				
F	0,97	0,98				

* Como x_3 fora removido no passo imediatamente anterior, esse teste serve apenas para ilustrar que variáveis que foram excluídas previamente devem ser testadas novamente.

** Para simplificar a tabela, não foram ilustrados termos que não dependem das variáveis X (ou seja, termos que são constantes) da expressão Y.

Sequência de análise da Tabela 3.1:

1º Passo: Calcular a regressão de Y para cada variável, individualmente. Selecionar a variável com maior F para adicionar ao modelo. No exemplo, x_3 apresentou maior F, e satisfaz $F > 0,5$ ($F > F_{\text{Ref-FS}}$).

2º Passo: Calcular a regressão de Y considerando x_3 combinado com as demais variáveis. Selecionar a variável (ainda não adicionada) com maior F para adicionar ao modelo. No exemplo, x_1 apresentou maior F, e satisfaz $F > 0,5$.

3º Passo: Calcular a regressão de Y considerando x_3 e x_1 combinados com as demais variáveis. Selecionar a variável (ainda não adicionada) com maior F para adicionar ao modelo. No exemplo, x_2 apresentou maior F, e satisfaz $F > 0,5$. Selecionar a variável com menor F para remover do modelo. No exemplo, x_3 apresentou menor F, e satisfaz $F < 0,5$ ($F < F_{\text{Ref-BE}}$).

4º Passo: Calcular a regressão de Y com as variáveis restantes do modelo. Excluir a variável selecionada no passo anterior, caso $F > 0,5$. No exemplo, apenas x_1 e x_2 continuam no modelo, e x_3 é excluída.

5º Passo: Calcular a regressão de Y considerando x_1 e x_2 combinados com as demais variáveis. Selecionar a variável (ainda não adicionada) com maior F para adicionar ao modelo. No exemplo, nenhuma das variáveis satisfaz $F > 0,5$, logo o algoritmo é encerrado.

6º Passo: Quando não há mais variáveis a serem inseridas no modelo, a solução final foi encontrada. Para o exemplo, a função final seria: $Y = \beta_0 + 1,36 \cdot x_1 - 1,77 \cdot x_2$.

O uso de recursos computacionais para a aplicação do algoritmo de Regressão *Stepwise* é fundamental, devido à quantidade de dados a serem processados. Deve-se ter em mente que a determinação dos parâmetros β e o cálculo da soma dos quadrados dos resíduos - cujos cálculos não são apresentados nessa dissertação – envolve uma grande quantidade de operações matemáticas e estatísticas, o que torna a análise inviável sem o auxílio computacional. Destacam-se alguns programas para análise estatística que podem ser utilizados: Action (Portal Action), NCSS (NCSS LLC), e módulos específicos do MATLAB, Python e Excel.

3.2 Redes Neurais Artificiais

Uma Rede Neural Artificial (RNA) tem seu funcionamento baseado em conceitos que são observados no cérebro humano, reproduzindo características e funções de neurônios e caminhos sinápticos do cérebro. A RNA funciona como um processador paralelo distribuído, constituído por unidades de processamento simples, denominadas neurônios (ou nós computacionais), com capacidade de armazenar conhecimento e torná-lo disponível para uso. Assim como ocorre no cérebro, o conhecimento adquirido pela rede a partir de seu ambiente é realizado através de um processo de aprendizagem (HAYKIN, 2001).

As RNAs constituem um ramo importante da Inteligência Artificial (IA) e são amplamente utilizadas devido à possibilidade de se implementar computacionalmente versões simplificadas dos neurônios biológicos. As RNAs são capazes de organizar dados, reconhecer padrões, aprender, tolerar falhas e tomar decisões com base no aprendizado, extraindo informações não apresentadas de forma explícita, entre outras habilidades (ZOUCCAS, 2009).

A capacidade de aprender de uma RNA é obtida por um algoritmo de aprendizagem. O mesmo altera os pesos das ligações sinápticas entre os neurônios, distribuindo-as de maneira ordenada para a rede alcançar seus objetivos. Este processo possui as características que serão apresentadas nas próximas seções.

Na estrutura mais simples, os neurônios de uma RNA são organizados em camadas, sendo estas divididas em três tipos: Entrada, Oculta (Intermediária ou Escondida), e Saída. A Figura 3.1 ilustra um exemplo de RNA simples com três neurônios na camada de entrada, 4 neurônios na camada oculta e 3 neurônios na camada de saída.

O número de neurônios das camadas de entrada e saída deve ser definido previamente, de acordo com o número de entradas e saídas do processo a ser modelado. Os conjuntos, ou vetores, de entrada e saída de uma RNA também são chamados de *padrões*. Por sua vez, a camada oculta deve ser definida em função da busca do melhor resultado oriundo de treinamento, podendo ser acrescentados ou retirados neurônios, e refazendo o treinamento até que se consiga uma configuração satisfatória (AGUIAR et al., 2007).

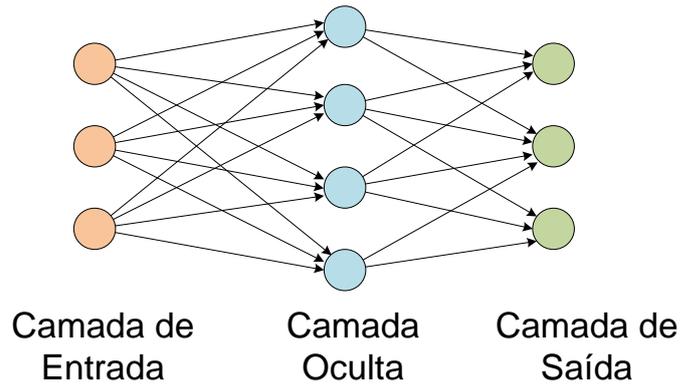


Figura 3.1 – Exemplo de rede neural artificial simples.

Fonte: Elaborado pelo autor

3.2.1 Representação das RNAs

A unidade básica de processamento de informação de uma RNA é o neurônio. Esse pode ser dividido em três elementos principais, como mostra a Figura 3.2:

- **Sinapses:** responsáveis por ponderar as entradas do neurônio, multiplicando-as por um peso;
- **Somador:** responsável pela soma das entradas multiplicadas pelos respectivos pesos;
- **Função de ativação:** responsável por limitar a amplitude de saída do neurônio tipicamente para valores normalizados nos intervalos $[0,1]$ ou $[-1,1]$.

Nesse modelo, está representado também um sinal b_k , usualmente denominado *bias*, que tem a função de deslocar o resultado do somador para cima ou para baixo, antes de sua entrada na função de ativação.

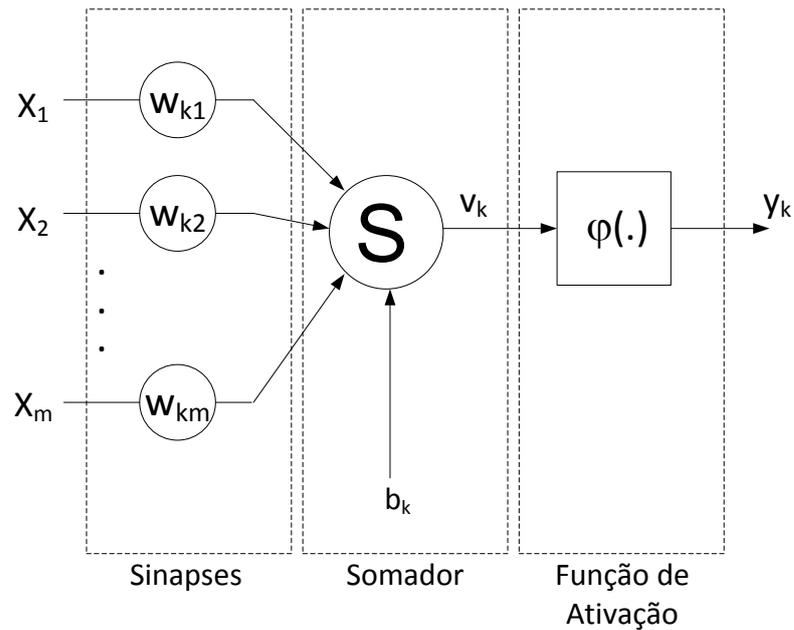


Figura 3.2 – Representação de um neurônio

Fonte: Adaptado de Haykin (2001)

A saída y_k do neurônio pode ser expressa em função das entradas x_k pela Equação 3.4:

$$y_k = \varphi \left(b_k + \sum_{j=1}^m w_{kj} \cdot x_j \right) \quad (3.4)$$

sendo:

φ = operador da função de ativação;

k = índice do neurônio;

j = índice da sinapse;

w = peso sináptico;

b = bias;

m = número de sinapses.

A função de ativação pode ter diferentes formas, sendo as mais usuais apresentadas na Figura 3.3. Além dessas funções, existem também funções que restringem a saída ao intervalo que inclui valores negativos, $[-1, 1]$.

Nessa Dissertação, foi utilizada a função denominada Sigmóide (em forma de S), que pode ser representada pela Equação 3.5:

$$\varphi(v) = \frac{1}{1 + e^{-av}} \quad (3.5)$$

na qual:

v = saída do somador (também denominado campo local induzido);

a = inclinação da curva.

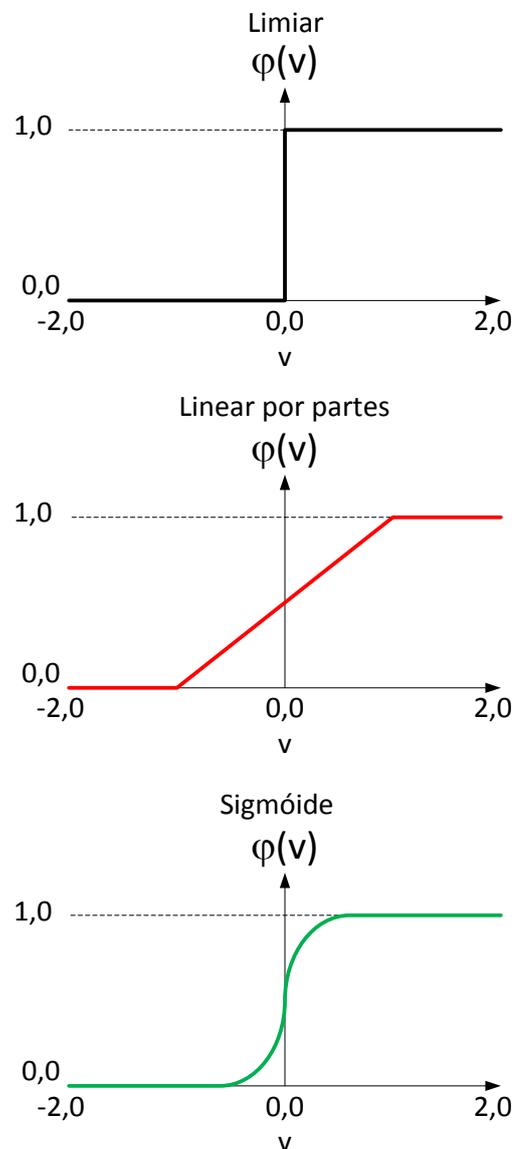


Figura 3.3 – Exemplos de funções de ativação

Fonte: Elaborado pelo autor

3.2.2 Estruturas de RNAs

De modo geral, existem três estruturas distintas relacionadas à forma como os neurônios estão organizados em uma RNA:

- **Redes alimentadas adiante com camada única:** nas quais existe apenas uma camada de neurônios onde ocorre processamento (Figura 3.4.a); o termo “adiante” se refere ao fato da entrada se propagar para frente, através da rede.
- **Redes alimentadas adiante com múltiplas camadas:** nas quais existem camadas de neurônios ocultas (Figura 3.4.b);
- **Redes recorrentes:** nas quais existe pelo menos um laço de realimentação, caracterizado pelo retorno da informação de saída de um neurônio para camadas anteriores da rede que influenciam em sua entrada (Figura 3.4.c). O operador do ramo de realimentação é representado por z^{-1} , e consiste na aplicação de um atraso temporal unitário em sua saída em relação à sua entrada.

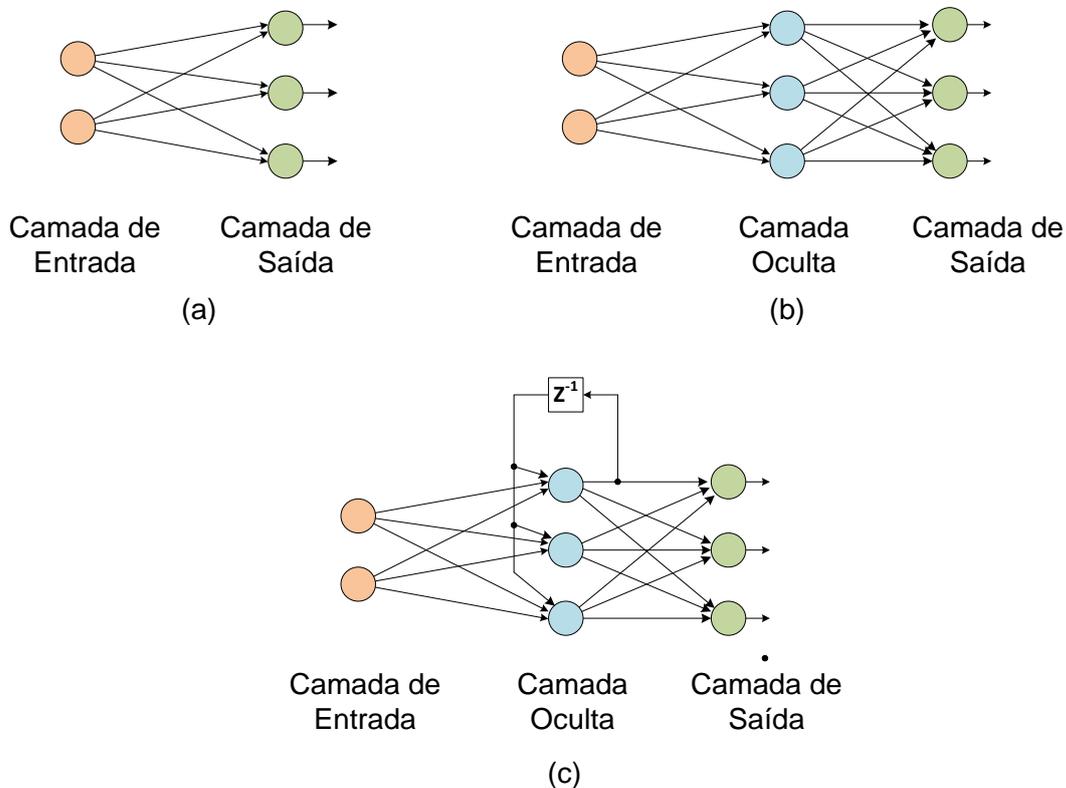


Figura 3.4 – Exemplos de estruturas de RNAs.

Fonte: Elaborado pelo autor

Quando todos os neurônios de camadas consecutivas da RNA estão conectados entre si através de sinapses, a rede é dita *totalmente conectada*; caso contrário, a rede é dita *parcialmente conectada*.

3.2.3 Aprendizagem das RNAs

O conceito de aprendizagem de uma RNA está relacionado com mecanismos de adaptação de seus parâmetros através de estímulos provenientes do ambiente no qual a rede está inserida (HAYKIN, 2011). Em termos de aprendizagem, as RNAs podem ser divididas basicamente em duas classes: Redes Não-Supervisionadas e Redes Supervisionadas.

As RNAs Não-Supervisionadas obtêm seu aprendizado a partir somente dos dados de entrada e descobrindo, estaticamente, padrões relevantes destes dados. Neste treinamento, não há a associação entre os dados de entrada com os valores desejados de saída e, dessa forma, o sistema desenvolve seu próprio aprendizado (PEREIRA, 2013). A Figura 3.5 ilustra o conceito de uma RNA com aprendizado Não-Supervisionado.

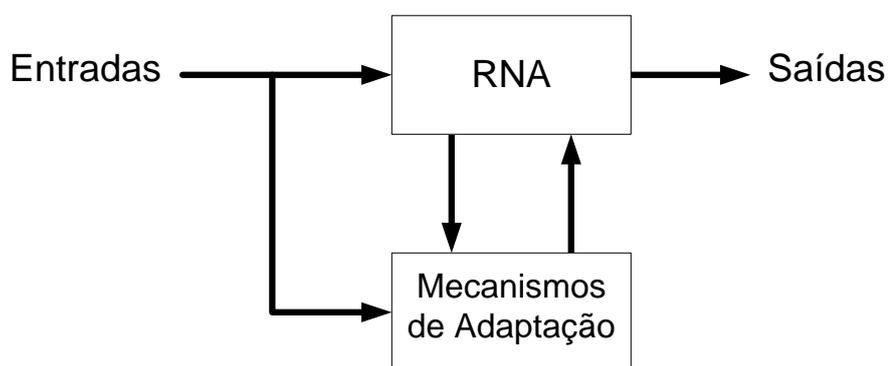


Figura 3.5 – Estrutura de Aprendizado Não-Supervisionado.

Fonte: Adaptado de Altran (2005)

Nas RNAs Supervisionadas, existe uma associação entre os dados de entrada com valores almejados na saída, que podem ser, por exemplo, dados reais de um sistema e que serão utilizados como referência no treinamento da rede. Sempre que houver divergência entre a saída da RNA e a saída desejada, os valores dos pesos sinápticos da rede são ajustados, até que o erro entre a saída e a referência atinja um valor aceitável (PASQUOTTO, 2010). Esse processo de aprendizagem é denominado *aprendizagem por correção de erro*, e essa estrutura emprega o paradigma conhecido como aprendizagem com professor.

Na Figura 3.6 está ilustrado o conceito de aprendizado supervisionado. O bloco “Professor” recebe os mesmos dados de estímulo do ambiente que a rede, porém, ele detém o conhecimento prévio sobre o ambiente, a partir de exemplos. Dessa forma, para um determinado conjunto de entradas, o professor é capaz de produzir uma saída desejada, ou referência, que é comparada com a saída produzida pela RNA e utilizada para produzir o sinal de erro, que por fim é utilizado para ajustar os parâmetros da rede.

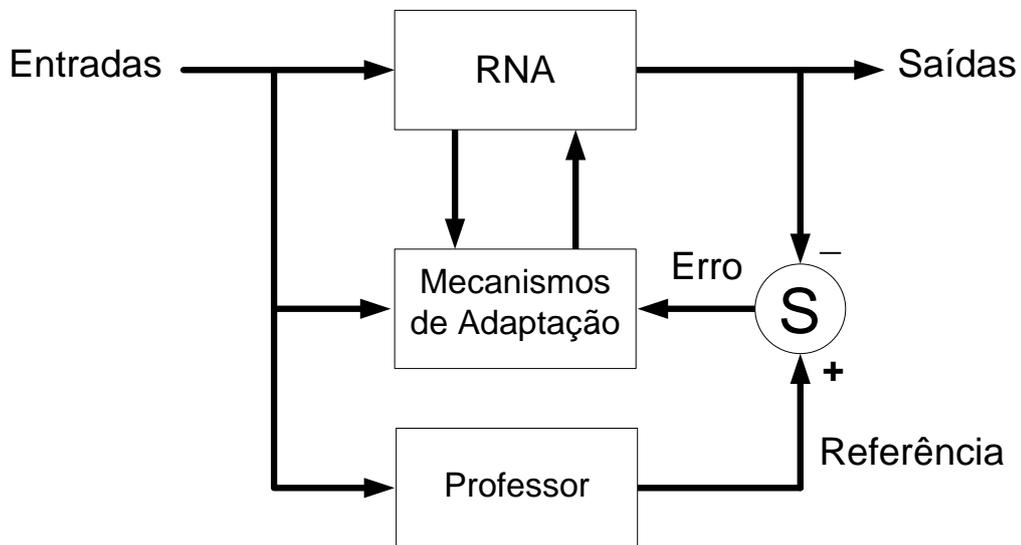


Figura 3.6 – Estrutura de Aprendizado Supervisionado.

Fonte: Adaptado de Altran (2005)

As redes supervisionadas são as mais comumente utilizadas por permitirem um controle maior de sua resposta através da interpretação e atuação sobre o erro

apresentado pela RNA. Elas foram originalmente propostas por Rosenblatt (1958), que introduziu o conceito do *perceptron*, que será abordado a seguir.

Nessa Dissertação é empregado o processo de aprendizagem supervisionada, com múltiplas camadas alimentadas adiante, ou rede *Perceptron* de Múltiplas Camadas. As principais características dessa rede são apresentadas na próxima seção.

3.2.4 Redes *Perceptrons* de Múltiplas Camadas

Um *perceptron* consiste em um neurônio que computa uma única saída a partir de uma combinação linear de múltiplas entradas com pesos sinápticos ajustáveis e um valor de bias. Ele é considerado a forma mais simples de neurônio, e segue o modelo apresentado previamente na Figura 3.2. Tipicamente, ele é utilizado para classificação de padrões.

As redes *Perceptrons* de Múltiplas Camadas (MLP – *Multilayer Perceptron*) são redes com pelo menos uma camada oculta e uma camada de saída de nós computacionais, alimentadas adiante, cujo treinamento se dá de forma supervisionada.

O algoritmo de aprendizagem mais conhecido para esse tipo de rede é o Algoritmo de Retropropagação de Erro (*error back-propagation*), que emprega a regra de aprendizagem por correção de erro.

3.2.4.1 Algoritmo de Retropropagação de Erro

De modo resumido, o algoritmo de retropropagação de erro consiste no ajuste dos pesos sinápticos da rede em função de um sinal de erro, e pode ser representado pelo conjunto formado pelas Equações 3.6 a 3.9:

Erro $e_j(n)$: Diferença entre o valor da saída real ($y_j(n)$) de um neurônio e o valor esperado ($d_j(n)$), ou de referência. O índice j , nesse caso, denota um neurônio de

saída e “n” indica uma iteração que representa o n-ésimo padrão de treinamento, ou exemplo, apresentado à rede.

$$e_j(n) = d_j(n) - y_j(n) \quad (3.6)$$

Ajuste do peso sináptico: Alteração feita ao peso sináptico de uma conexão, que consiste na soma do peso atual ($w_{ji}(n)$) a um valor de correção ($\Delta w_{ji}(n)$). O índice ji indica a ligação do neurônio da camada j a um neurônio da camada i , à sua esquerda.

$$w_{ji}(n + 1) = w_{ji}(n) + \Delta w_{ji}(n) \quad (3.7)$$

Correção $\Delta w_{ji}(n)$: também referida como regra delta, dada por:

$$\Delta w_{ji}(n) = \eta \cdot \delta_j(n) \cdot y_j(n) \quad (3.8)$$

Sendo η definida como a taxa de aprendizagem do algoritmo, que implica em o quanto maior ou menor será a variação do peso sináptico de uma iteração para outra, e $\delta_j(n)$ definido como o gradiente local, que aponta para as modificações necessárias no peso sináptico, em função do erro.

Gradiente local $\delta_j(n)$: é o produto entre o erro $e_j(n)$ e a derivada da função de ativação $\varphi(v_n)$, definida anteriormente.

$$\delta_j(n) = e_j(n) \cdot \varphi'_j(v_j(n)) \quad (3.9)$$

O treinamento utilizando retropropagação consiste em dois passos básicos:

1. Passo adiante (propagação): Um padrão é apresentado às unidades da camada de entrada e, a partir desta camada, as unidades de processamento calculam a resposta que é produzida na camada de saída; a partir dessa resposta, o erro é calculado.

2. Passo atrás (retropropagação): O erro é propagado de volta a partir da camada de saída até a camada de entrada; nas unidades das camadas internas os pesos das conexões vão sendo modificados utilizando a regra delta.

Uma apresentação completa do conjunto de treinamento que está sendo processado, no qual os pesos são processados padrão por padrão, é denominada *época*.

A Figura 3.7 ilustra um exemplo de RNA *Perceptron* com duas camadas de processamento ((1) e (2), cujos parâmetros estão em cores diferentes), e que contém dois neurônios em cada camada. A RNA está representada através de um gráfico de fluxo de sinal.

Na parte superior da figura, ocorre o primeiro passo do algoritmo, no qual as setas de fluxo apontam da entrada para a saída. Na parte inferior, as setas estão em sentido contrário, indicando o segundo passo do algoritmo (retropropagação). Essa figura também ilustra que a Equação 3.9 assume formas diferentes, dependendo se o neurônio está na camada de saída ou na camada oculta (HAYKIN, 2001, pg. 202).

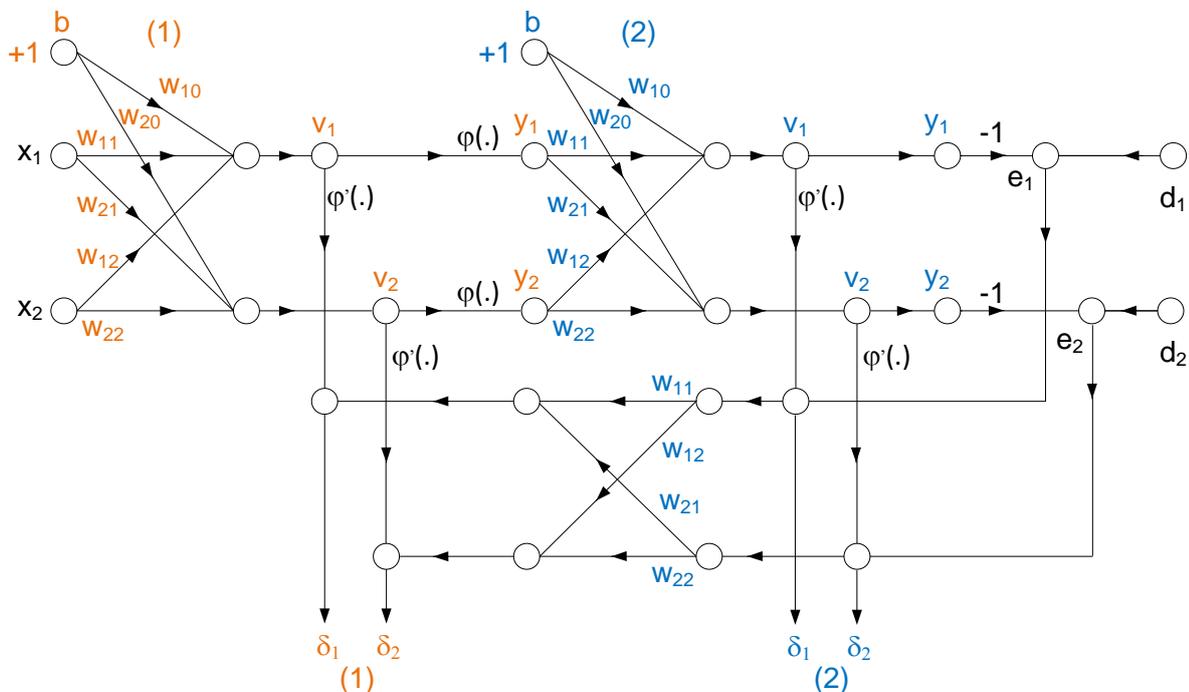


Figura 3.7 – Gráfico de Fluxo de Sinal de uma MLP com retropropagação de erro.

Fonte: Elaborado pelo autor.

3.2.4.2 Validação Cruzada

Conforme visto na seção anterior, a aprendizagem por retropropagação está fortemente relacionada com a aproximação da saída real da RNA a um valor esperado. Dessa forma, é necessário haver um conjunto de dados de exemplo, a partir do qual será codificado o mapeamento das entradas e saídas nos pesos sinápticos do *perceptron*. A validação do modelo deve ser feita com o conjunto de dados restantes (Figura 3.8), sendo que esse tipo de validação é normalmente conhecido como *validação direta*, na qual não existe nenhuma manipulação dos dados.

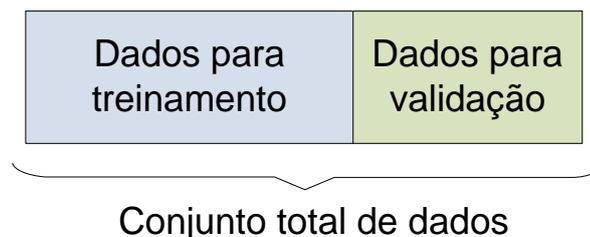


Figura 3.8 – Divisão do conjunto de dados para validação do modelo da RNA.

Fonte: Elaborado pelo autor.

A validação cruzada (*cross-validation*) é uma das ferramentas mais conhecidas para validar um modelo com um conjunto de dados diferente do que foi usado para o treinamento da rede. Segundo FIGUEIREDO, R. M. ; SANTOS, J. V. C.; CECHIN, A. L. este método consiste na divisão do conjunto de dados disponíveis em grupos e no cruzamento desses grupos para formar um conjunto de treinamento e outro para validação. Entre as técnicas mais utilizadas que empregam esse conceito destaca-se a *Validação Cruzada K-fold*, que foi empregada nessa Dissertação.

Na Validação Cruzada *K-fold*, o conjunto total de dados é dividido em K subconjuntos cruzados de mesmo tamanho, sendo que desses, “K – 1” são utilizados para treinamento da rede, e o subconjunto restante é usado para

validação. O processo é repetido K vezes, alternando-se o subgrupo de validação, até que todos os subgrupos tenham sido utilizados para essa finalidade. A Figura 3.9 ilustra o conceito dessa validação, para $K = 3$. O valor usual de K em aplicações de RNAs é 10.

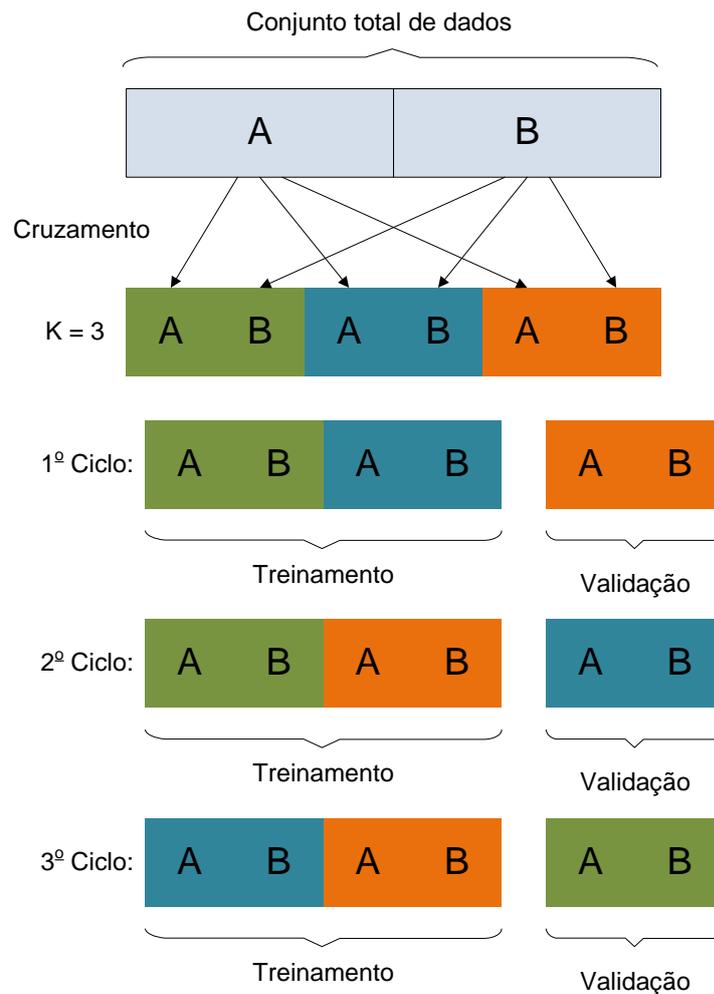


Figura 3.9 – Técnica de Validação Cruzada K -fold

Fonte: Elaborado pelo autor

3.3 Avaliação de desempenho

Uma vez determinados os modelos que serão empregados para atender aos objetivos do projeto – seja utilizando Regressão *Stepwise*, Redes Neurais Artificiais

ou outro método – é necessário fazer uma avaliação do seu desempenho, a fim de verificar o quanto sua resposta se aproxima dos valores esperados. Algumas ferramentas estatísticas ou gráficas podem ser utilizadas para essa finalidade. No primeiro caso, obtém-se uma avaliação quantitativa, onde normalmente se quantifica o erro do previsor utilizado no modelo. No segundo caso, obtém-se uma análise qualitativa, na qual visualmente se verifica a aderência dos resultados.

3.3.1 Análise Estatística

Na maioria dos métodos de análise, os dados são previamente normalizados, de acordo com a Equação 3.10:

$$x_i^* = \frac{x_i}{x_{max}} \quad (3.10)$$

Sendo:

x_i^* = valor normalizado de x_i ;

x_{max} = valor máximo da variável x , no conjunto total de dados.

Dentre as medidas mais utilizadas na análise de desempenho dos modelos estão o Erro Médio Absoluto (MAE – *Mean Absolute Error*) e a Raiz do Erro Médio Quadrático (RMSE – *Root Mean Square Error*), que foi empregada nessa Dissertação.

A análise a partir do erro médio absoluto é feita de forma linear, confrontando-se os valores estimados com os valores reais, um a um, e calculando-se a média dos resultados. A Equação 3.11 mostra como a métrica é determinada.

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| = \frac{1}{N} \sum_{i=1}^N e_i \quad (3.11)$$

Sendo:

N = número total de dados;

y_i = saída prevista pelo modelo;

\hat{y}_i = saída esperada;

e_i = erro absoluto.

A raiz do erro médio quadrático consiste, basicamente, no desvio padrão da diferença entre os resultados calculados pelo modelo e os resultados previstos, ou seja, o desvio padrão do erro. Ela pode ser determinada a partir da Equação 3.12.

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (3.12)$$

O termo sob a raiz quadrada é definido como o erro médio quadrático, que é uma medida da variância do erro.

3.3.2 Análise Gráfica

Para uma análise qualitativa do desempenho de um modelo, as ferramentas mais simples são a Correlação Gráfica e a Sobreposição de Curvas.

A Correlação Gráfica consiste em correlacionar os dados de saída produzidos pelo modelo com os dados reais esperados, utilizando eixos diferentes do mesmo gráfico. A aderência perfeita consiste em obter os pontos dos dados estimados sobre a reta diretriz, que representa o lugar geométrico na qual as saídas estimadas são iguais às saídas esperadas. Quanto maior for a distância dos pontos estimados a essa reta, maior é o erro associado aos instantes de previsão daqueles pontos. A Figura 3.10 ilustra um exemplo de Correlação Gráfica para um conjunto hipotético de dados.

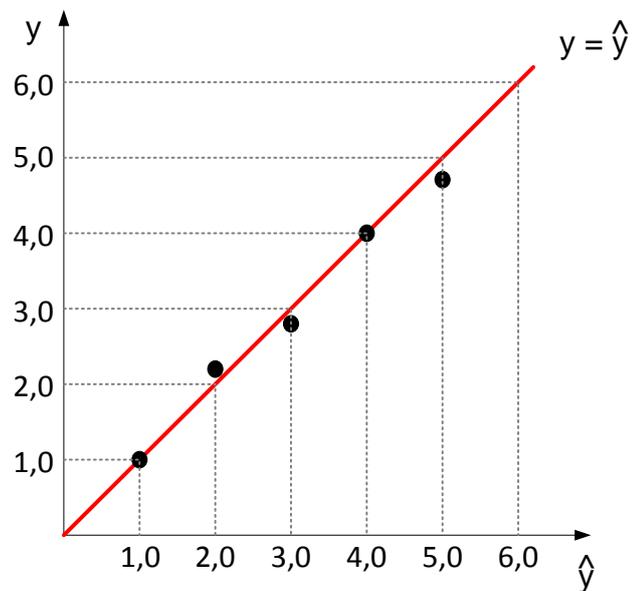


Figura 3.10 – Exemplo de análise por Correlação Gráfica

Fonte: Elaborado pelo autor

A Sobreposição de Curvas consiste no desenho simultâneo das séries de resultados produzidas pelo modelo e das séries de dados esperados, para cada resultado produzido, em um intervalo de tempo. Esse método oferece uma maneira simples e direta de visualizar a aderência dos resultados e o comportamento do modelo. A Figura 3.11 mostra um exemplo de aplicação da Sobreposição de Curvas.

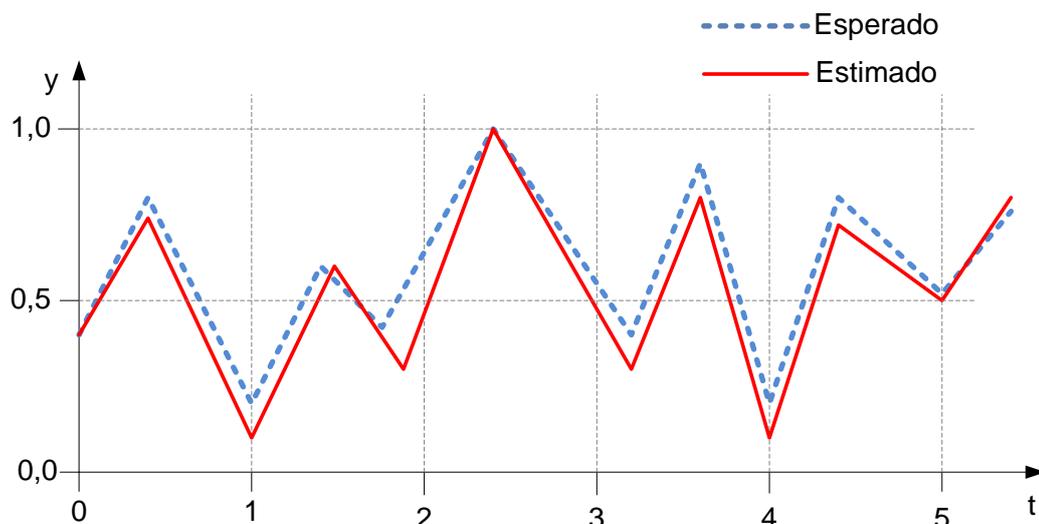


Figura 3.11 – Exemplo de análise por Sobreposição de Curvas

Fonte: Elaborado pelo autor

3.4 Considerações Finais

Esse capítulo apresentou os principais conceitos teóricos relacionados à análise estatística por Regressão *Stepwise* e às Redes Neurais Artificiais, e que serão empregados como parte da metodologia proposta nessa Dissertação.

A Regressão *Stepwise*, como outros métodos de regressão, permite criar um modelo de previsão de uma saída com base na análise estatística de um conjunto de dados de entrada disponíveis. Ela possui a vantagem de incluir uma etapa de seleção de variáveis no processo de criação do modelo, o que pode ser apropriado quando se trabalha com um conjunto grande de variáveis, como é o caso desse trabalho.

As Redes Neurais Artificiais possuem diferentes aplicações, como reconhecimento de padrões, otimização, e também a criação de modelos de previsão, entre outras. Nessa Dissertação, foi desenvolvida uma Rede *Perceptron* de Múltiplas Camadas, que emprega o conceito de aprendizagem supervisionada com retropropagação de erro. Trata-se de um modelo de rede bastante consolidado na literatura, de fácil implementação prática, e robusto, conforme apresentado em exemplos na Revisão Bibliográfica.

Os métodos destacados nesse Capítulo foram empregados para a criação de um modelo previsor, que tem a finalidade de estimar o consumo de energia em uma lavoura de arroz irrigado, com base em informações das características da lavoura, como área irrigada e potência das bombas de irrigação, e em dados climatológicos. Essa estimativa faz parte da metodologia proposta para detecção de perdas não-técnicas nesse tipo de consumidor, apresentada no próximo Capítulo.

4 IDENTIFICAÇÃO DE PERDAS NÃO TÉCNICAS NO SUPRIMENTO DE ENERGIA A CONSUMIDORES RURAIS - METODOLOGIA PROPOSTA E RESULTADOS

Nesse Capítulo é apresentada a metodologia proposta para identificar a possibilidade de ocorrência de perdas não técnicas no suprimento de energia de consumidores rurais, caracterizados pelo uso de levantes para irrigação de lavouras de arroz. O Capítulo está dividido em três seções principais, sendo que na primeira seção são apresentadas considerações gerais sobre a proposta, na segunda seção é apresentada uma análise baseada no método estatístico de Regressão Stepwise, e na terceira seção é apresentada uma metodologia baseada em Redes Neurais Artificiais. Em cada caso, são apresentados os resultados obtidos a partir da análise de dados reais de lavouras de arroz irrigado do Rio Grande do Sul.

4.1 Fluxograma geral proposto para identificação de perdas não técnicas

A metodologia geral proposta para identificação de perdas não técnicas em consumidores rurais segue o fluxograma geral apresentado na Figura 4.1. Esse fluxograma se aplica tanto ao método de Regressão *Stepwise* quanto ao método de Redes Neurais artificiais. A proposta consiste em determinar o consumo esperado de energia elétrica em uma lavoura (calculado a partir das ferramentas estatísticas ou RNA) e compará-lo com o consumo real. O desenvolvimento de um modelo que represente satisfatoriamente o consumo da lavoura é o principal desafio dessa proposta. Com os valores de consumo reais e estimados, é feita uma comparação entre os padrões observados em diferentes lavouras e safras, e é definido um índice que representa a possibilidade de perdas não técnicas nas lavouras analisadas.

As principais etapas do fluxograma, e a descrição do conjunto de dados utilizados para testes nessa Dissertação são detalhadas a seguir.

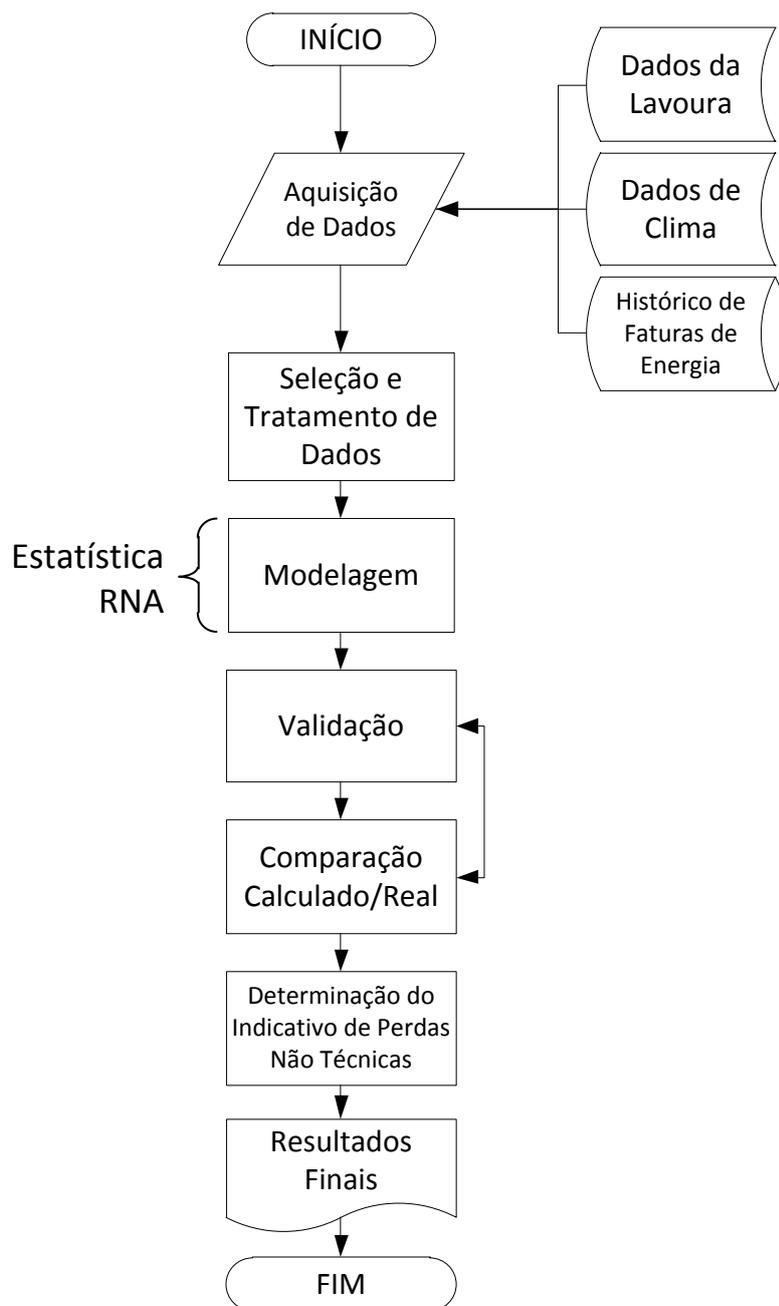


Figura 4.1 – Fluxograma geral da metodologia proposta

Fonte: Elaborado pelo autor

4.2 Aquisição e tratamento de dados

A primeira etapa da metodologia consiste em adquirir um conjunto de dados de diferentes lavouras e safras, que servirão para modelagem do sistema. Esse

conjunto é provido por bases distintas de dados, como históricos e medições de consumo de energia, obtidos a partir de concessionárias de energia, e dados climatológicos, obtidos a partir de estações meteorológicas.

Normalmente, a base de tempo dos registros dessas fontes não é a mesma, sendo necessária uma seleção e tratamento prévio de dados. Essa seleção também serve para excluir registros inconsistentes, que caracterizam erros nas medições.

A Figura 4.2 esquematiza o conjunto de dados definidos como entrada e saída nesse trabalho. A potência instalada, o consumo de energia elétrica e a área irrigada foram obtidos a partir de registros de uma distribuidora de energia local. Os dados climatológicos foram obtidos de uma estação meteorológica próxima às lavouras.

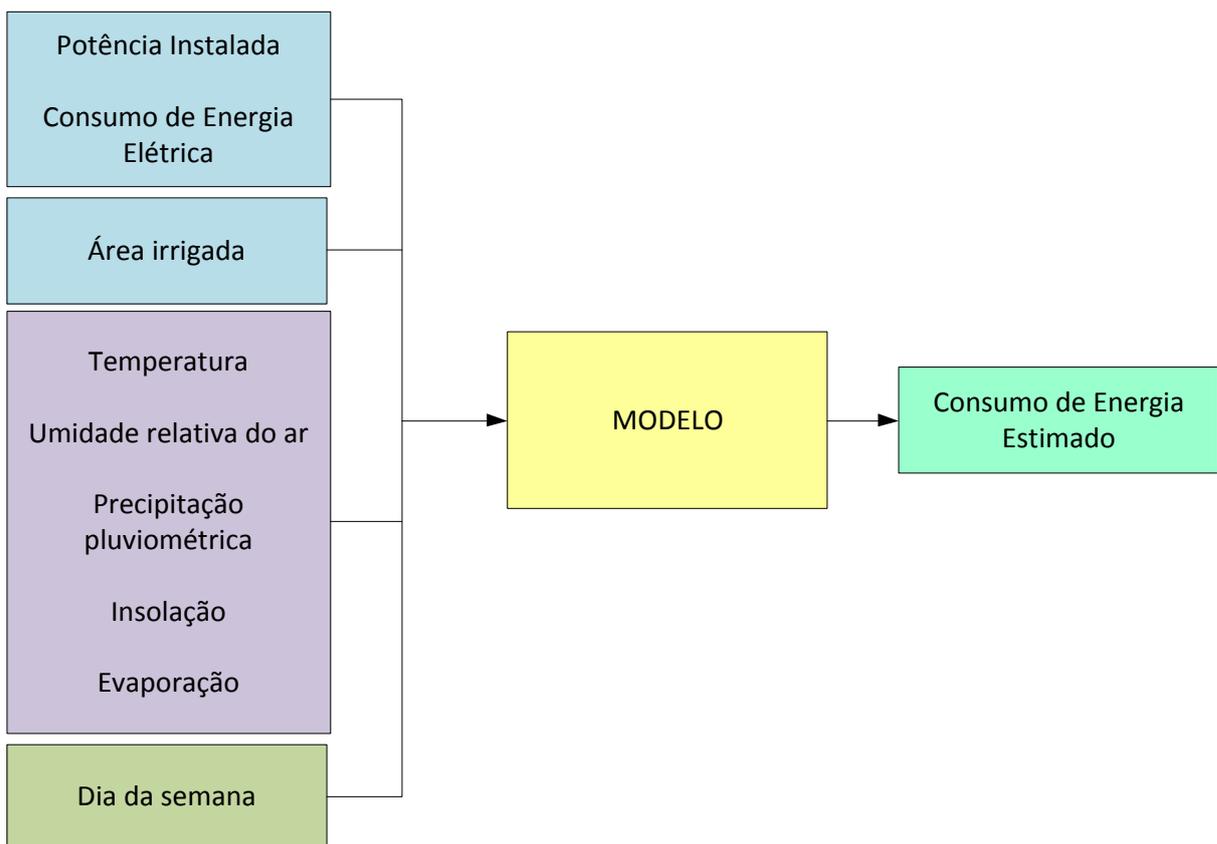


Figura 4.2 – Dados de Entrada e Saída do Sistema Proposto.

Fonte: Elaborado pelo autor

Todos os dados foram normalizados em relação aos respectivos valores

máximos de cada variável, conforme Equação 3.10. Além disso, foi realizada a média diária dos dados climáticos e de consumo de energia elétrica.

4.3 Ferramentas de modelagem, análise e validação

A base da metodologia proposta são as ferramentas de modelagem por Regressão *Stepwise*, e por Redes Neurais Artificiais. A teoria envolvida na implementação dessas ferramentas foi apresentada no Capítulo 3, sendo que no presente capítulo serão discutidos detalhes específicos de sua aplicação para os propósitos desse trabalho, com os respectivos resultados, nas Seções 4.8, 4.10 e 4.11.

4.4 Indicativo de perdas elétricas não técnicas

A possibilidade de ocorrência de perdas elétricas não técnicas nos consumidores analisados é determinada a partir da comparação entre os valores de consumo de energia calculados pelo modelo desenvolvido e os valores reais de consumo, obtidos a partir das medições disponíveis. O resultado é fornecido de forma qualitativa, considerando o erro médio percentual dado pela Equação 3.12 e a Tabela 4.1, proposta a seguir.

Tabela 4.1 – Índice de possibilidade de perdas não técnicas

Faixa de erro	Possibilidade de perdas não técnicas
$\text{Erro} \leq 15\%$	Baixa
$15 < \text{Erro} \leq 30\%$	Média
$\text{Erro} > 30\%$	Alta

Fonte: Elaborado pelo autor

Essa análise parte do pressuposto de que as lavouras têm padrões similares de consumo de energia elétrica, considerando a área irrigada e as condições

climáticas. Entende-se, no entanto, que não há uma formulação matemática precisa que relacione todas as variáveis, uma vez que o consumo de energia pode ser afetado por fatores como variações na cultura do arroz irrigado, práticas de irrigação diferenciadas, contratos entre consumidores e distribuidoras de energia, qualidade do fornecimento de energia, manutenção e automação de equipamentos, etc. Dessa forma, como uma proposta inicial sobre o assunto, optou-se por utilizar faixas de erros de 15% para o indicativo da possibilidade de ocorrência de perdas não técnicas nas lavouras analisadas.

O resultado final fornecido pelo sistema proposto servirá de apoio às equipes de inspeção, que poderão dedicar mais atenção aos casos com alta possibilidade de perdas não técnicas, e planejar vistorias *in loco*.

4.5 Descrição do conjunto de dados para teste e validação

Nas próximas seções, são apresentados a modelagem e os resultados dos métodos empregados nessa Dissertação, que foram desenvolvidos com base no conjunto de dados apresentados a seguir.

Os dados empregados nesse trabalho compreendem um conjunto de cinco safras de três lavouras distintas, formando um total de quinze blocos de dados. As lavouras possuem características semelhantes entre si em relação às técnicas de irrigação, e são localizadas na mesma região do Estado do Rio Grande do Sul. As unidades consumidoras são identificadas pelas siglas UC-1, UC-2 e UC-3. As Figuras 4.3 a 4.5 apresentam imagens de satélite, obtidas através do Google Earth®, que mostram as três lavouras utilizadas nos testes.

O Período de irrigação (levante) nas culturas de arroz é geralmente compreendido entre os meses de Setembro a Março. Para este estudo utilizou-se como parâmetro de o início de registro dos medidores dedicados, conforme segue:

- Set 2009 - Mar 2010 (Safrá 2009/2010)
- Set 2010 - Mar 2011 (Safrá 2010/2011)
- Set 2011 - Mar 2012 (Safrá 2011/2012)
- Set 2012 - Mar 2013 (Safrá 2012/2013)
- Set 2013 - Mar 2014 (Safrá 2013/2014)



Figura 4.3 – Unidade consumidora UC-1.
Fonte: Adaptado de Google Earth® (2014)

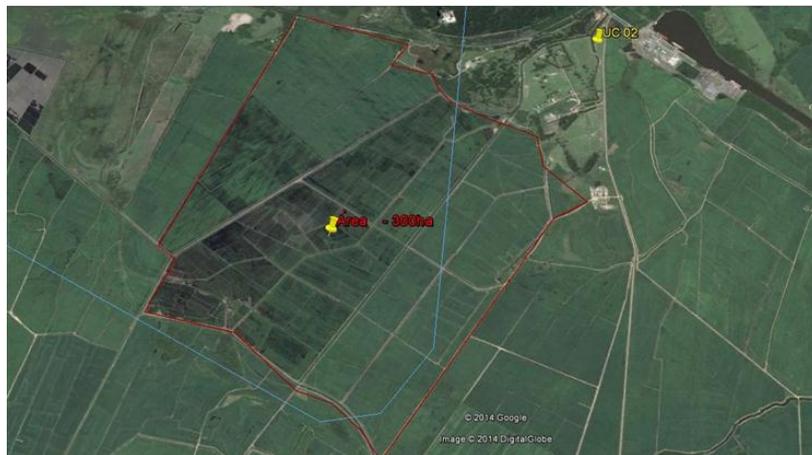


Figura 4.4 – Unidade consumidora UC-2.
Fonte: Adaptado de Google Earth® (2014)



Figura 4.5 – Unidade consumidora UC-3.
Fonte: Adaptado de Google Earth® (2014)

Dados gerais das lavouras

A Tabela 4.2 apresenta um sumário das informações gerais das lavouras, utilizadas no trabalho.

Tabela 4.2 – Dados gerais das lavouras

Unidade Consumidora	Área irrigada aproximada	Potência instalada
UC-1	90 ha	112,5 kVA
UC-2	300 ha	300 kVA
UC-3	50 ha	75 kVA

Fonte: Elaborado pelo autor

Consumo de energia elétrica

As lavouras possuem medidores eletrônicos exclusivos para irrigação, instalados em média tensão. Os dados adquiridos pelos medidores são registrados a cada 30 minutos, e posteriormente são extraídos através do programa ELO50. A Figura 4.6 ilustra um exemplo de gráfico extraído do programa.

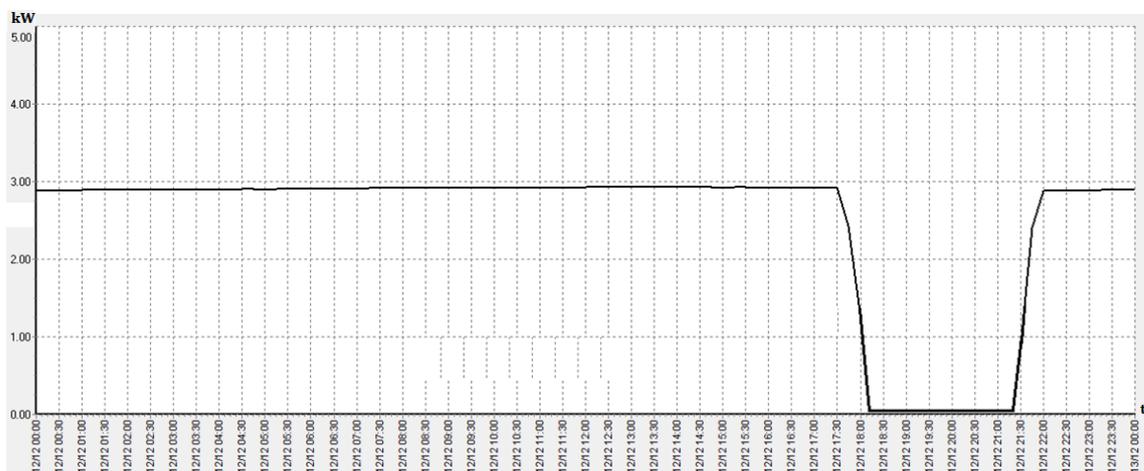


Figura 4.6 – Consumo de energia elétrica em um sistema de irrigação obtido com o programa Elo50.

Fonte: Elaborado pelo autor

Dados climatológicos

Os dados climatológicos foram obtidos a partir de uma estação meteorológica do Centro de Previsão de Tempo e Estudos Climáticos (CPTEC/ INPE), localizada aproximadamente a 40 km do local das lavouras. As informações utilizadas são as que constam na Figura 4.2: precipitação pluviométrica, temperatura, umidade relativa do ar, insolação e evaporação pelo método de Piche.

A maioria dos dados é disponibilizada com intervalos de 15 minutos. No entanto, algumas medições, como a precipitação pluviométrica, são integralizadas durante um dia. Desta forma, considerando esta característica da coleta de dados climatológicos, optou-se por trabalhar com dados diários nessa Dissertação.

4.6 Análise Estatística e Regressão Stepwise

Conforme apresentado no Capítulo 3, as ferramentas estatísticas podem ser úteis para a modelagem de um sistema quando se dispõe de um conjunto de variáveis de entrada e se busca encontrar uma relação de seu efeito em uma variável de saída.

Nesse trabalho, o consumo de energia elétrica das lavouras de arroz foi avaliado considerando as informações sobre as lavouras e a influência do clima. A ideia inicial é identificar como as variáveis de entrada se relacionam com a saída, e se é possível utilizar um modelo de regressão para representar o sistema.

4.7 Correlação do consumo de energia elétrica com o clima

Em uma primeira análise, foi realizado um estudo de correlação entre as variáveis, aplicando a estatística do Coeficiente de Pearson. Este coeficiente de correlação tem valores entre -1 e $+1$, sendo negativo quando uma variável diminui

com o aumento da outra variável, e positivo quando uma variável aumenta com o aumento da outra. O coeficiente de correlação de Pearson, r , é determinado pela

Equação 4.1.

$$r = \frac{\sum(x - \bar{x}) \cdot (y - \bar{y})}{(n - 1) \cdot \sigma_x \cdot \sigma_y} \quad (4.1)$$

onde:

x, y : variáveis a serem correlacionadas;

\bar{x}, \bar{y} : média do conjunto de cada variável;

σ_x, σ_y : desvio padrão do conjunto de cada variável;

n : número de dados do conjunto analisado.

Na estatística clássica, a significância da correlação é testada através do valor- p para as hipóteses nula e alternativa:

H_0 : $r = 0$, não há correlação

H_1 : $r \neq 0$, há correlação.

Sendo que a rejeição da hipótese nula pode ser determinada de acordo com o seguinte critério:

$0,5 < p \leq 1,0$: H_0 é fracamente rejeitada;

$0,1 < p \leq 0,5$: H_0 é moderadamente rejeitada;

$p \leq 0,1$: H_0 é fortemente rejeitada.

Dessa forma, além do coeficiente de correlação, pode-se obter uma indicação do grau de correlação de uma variável com outra.

A técnica foi aplicada ao um conjunto de dados da UC-1, e os resultados da Safra 2009/2010 são apresentados na Tabela 4.3. Para obtenção desses resultados foi utilizado o programa MiniTab®, que possui recursos específicos para análise estatística. A Tabela 4.4 apresenta os resultados de correlação das variáveis com o consumo de energia elétrica, das cinco safras de UC-1. Os resultados completos fornecidos pelo programa são apresentados no Apêndice A.

Tabela 4.3 – Análise de correlação pelo Coeficiente de Pearson / UC-1 2009/2010

		Consumo Energia	Precipit. Pluviom.	Insolação	Evapor.	Temperatura	Umidade Relativa
Precipit.	r	0,063	-	*	*	*	*
Pluviom.	p	0,381					
Insolação	r	0,077	-0,386	-	*	*	*
	p	0,286	0,000				
Evaporação	r	-0,060	-0,357	0,276	-	*	*
	p	0,402	0,000	0,000			
Temperatura	r	0,026	-0,325	0,247	0,405	-	*
	p	0,720	0,000	0,000	0,000		
Umidade Relativa	r	-0,036	0,344	-0,780	-0,265	-0,179	-
	p	0,617	0,000	0,000	0,000	0,012	

r: Coeficiente de Pearson; p: Valor-p

* Os valores das células assinaladas com "*" repetem-se em relação ao mesmo par de variáveis indicado nas demais células.

Fonte: Elaborado pelo autor

Tabela 4.4 – Análise de correlação do consumo de energia com clima para UC-1

		2009-2010	2010-2011	2011-2012	2012-2013	2013-2014
Precipitação	r	0,063	-0,037	-0,147	-0,122	-0,155
Pluviom.	p	0,381	0,607	0,155	0,161	0,119
Insolação	r	0,077	0,092	0,072	0,136	0,247
	p	0,286	0,196	0,491	0,117	0,012
Evaporação	r	-0,060	0,142	0,101	0,123	0,355
	p	0,402	0,045	0,331	0,158	0,000
Temperatura	r	0,026	0,003	0,059	0,092	0,383
	p	0,720	0,970	0,573	0,288	0,000
Umidade Relativa	r	-0,036	-0,112	-0,155	-0,101	-0,371
	p	0,617	0,115	0,133	0,244	0,000

Fonte: Elaborado pelo autor

Nas Tabelas 4.3 e 4.4 pode-se observar que existe alguma correlação entre as variáveis climáticas, porém a correlação dessas com o consumo de energia elétrica é fraca. A hipótese H0 é moderadamente ou fracamente rejeitada na maior

parte dos dados da Tabela 4.4, e os coeficientes de correlação são, em sua maioria, menores que 0,3.

Apesar dos valores obtidos nesse teste indicarem relações fracas entre as variáveis, especialmente em relação ao consumo de energia elétrica, que é a variável de interesse nesse trabalho, a análise estatística foi aprofundada, com a implementação do método de Regressão *Stepwise*.

4.8 Análise por Regressão Stepwise

O método de Regressão Stepwise foi aplicado em diferentes conjuntos de dados, considerando a disponibilidade de quinze safras para análise. O programa utilizado foi o Matlab®, que possui a função específica (*stepwise*), para essa aplicação.

Os resultados apresentaram grandes discrepâncias entre os valores de consumo de energia previstos pelo método e os valores de consumo reais, reforçando a conclusão, previamente indicada pela análise de correlação, de que não existe uma relação linear direta entre as variáveis de entrada e a saída.

Para exemplificar, é apresentado o resultado da análise de UC-1 / Safra 2013-2014.

A equação determinada para o conjunto de dados dessa safra é dada por:

$$Y = -3,6196 - 0,03843x_1 + 0,07323x_2 + 0,3215x_3 + 0,0009x_4 - 0,0048x_5 - 0,0067x_6 \quad (4.2)$$

sendo: Y – consumo de energia elétrica previsto; x_1 - precipitação pluviométrica, x_2 - evaporação, x_3 - umidade relativa do ar, $x_4 = x_1^2$, $x_5 = x_2^2$ e $x_6 = x_3^2$.

A Figura 4.7 ilustra as curvas de consumo da safra, evidenciando que os valores reais e previstos diferem significativamente, tanto em escala quanto em forma.

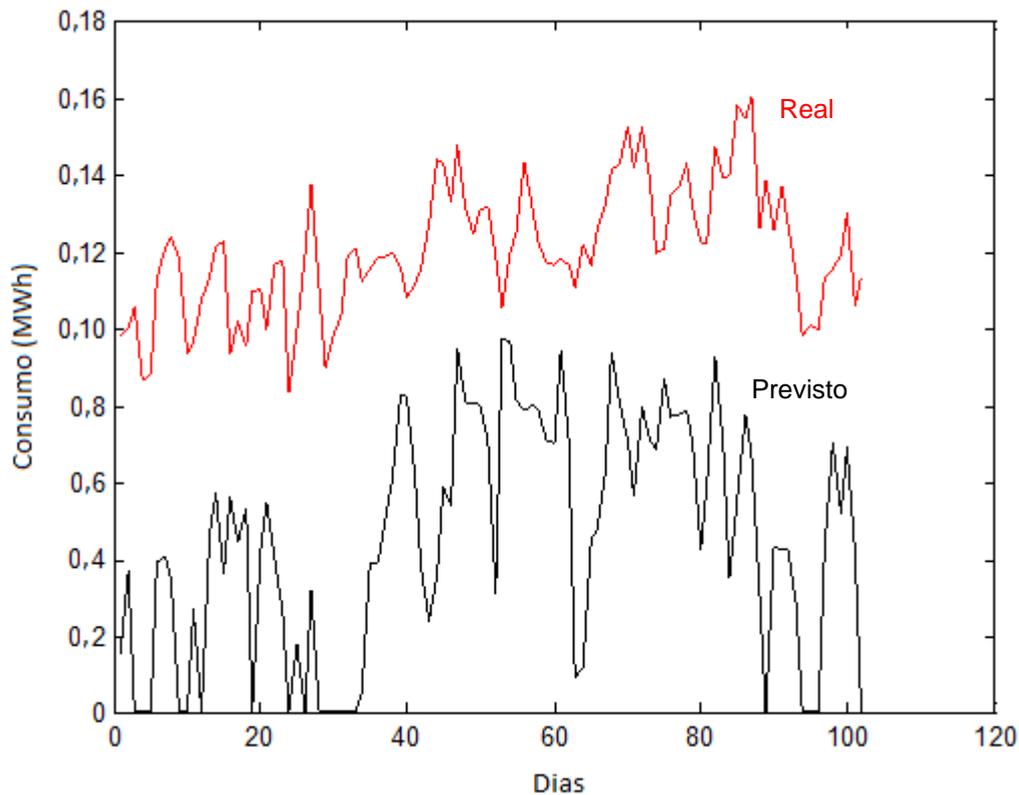


Figura 4.7 – Resultado da Regressão *Stepwise* para UC-3/ Safra 2013-2014.

Fonte: Elaborado pelo autor.

Os resultados das demais safras foram igualmente inconsistentes e podem ser observados no Apêndice C.

4.9 Implementação e Análise por Redes Neurais Artificiais

Como alternativa ao método estatístico apresentado na seção anterior, foi desenvolvido um modelo para estimar o consumo de energia elétrica nas lavouras utilizando Redes Neurais Artificiais. Para isso, foi implementada uma rede com as características apresentadas na Figura 4.8.

O programa utilizado foi desenvolvido em linguagem Python® com a biblioteca de Redes Neurais chamada *PyBrain*. Esta biblioteca é fechada, só havendo interação com o usuário no momento do treinamento e no momento da validação, sob a forma de entrega dos padrões de entrada e saída esperados. Neste

pacote não há a possibilidade de extração dos pesos dados para cada conexão entre os neurônios da RNA por ele produzida. Isto pode ser feito *a posteriori* com o uso de redes *bayesianas*.

Para automatizar e sistematizar o treinamento e manipulação dos dados foram criadas funções, também em Python, para extração dos dados dos arquivos em formato MS Excel®, normalização, automatização dos testes e verificação dos resultados por meio de testes de erro.

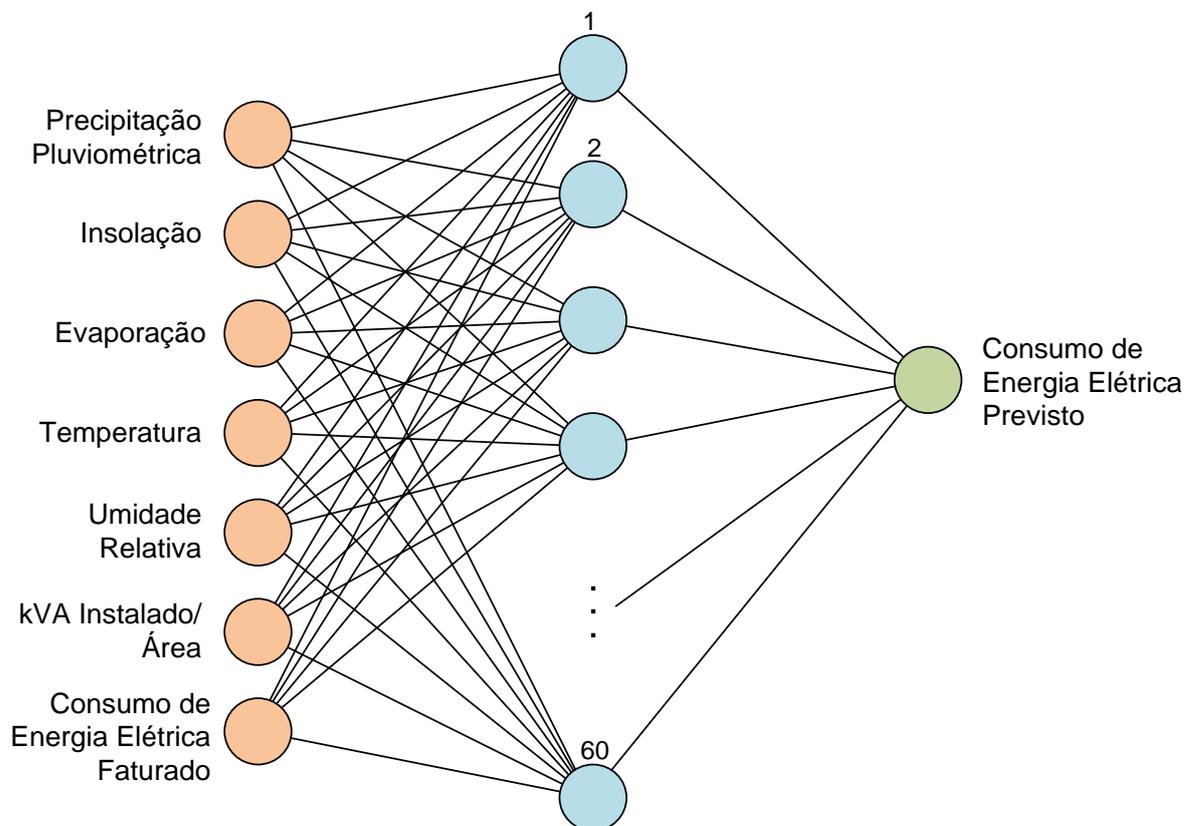


Figura 4.8 – Rede Neural Artificial proposta

Fonte: Elaborado pelo autor

A automatização do treinamento e da validação foi realizada para que os resultados pudessem ser avaliados e a RNA reconfigurada (em número de neurônios na camada escondida e número de épocas de treinamento) até se atingir valores pré-definidos. No caso desta Dissertação, o erro máximo aceitável foi

definido como 10%, em termos absolutos. A avaliação do erro é realizada por um módulo também desenvolvido para este projeto, de acordo com a Equação 3.12. O código do programa está apresentado no Apêndice B.

Em uma análise inicial, utilizando a metodologia de validação apresentada posteriormente, na Seção 4.3.1, foi verificado o número de neurônios na camada oculta que apresentasse o menor erro de treinamento (RMSE), com 1000 épocas. O resultado está ilustrado na Figura 4.9. O erro de treinamento com 60 neurônios foi o mais baixo entre as alternativas testadas, sendo inferior a 1,5%.

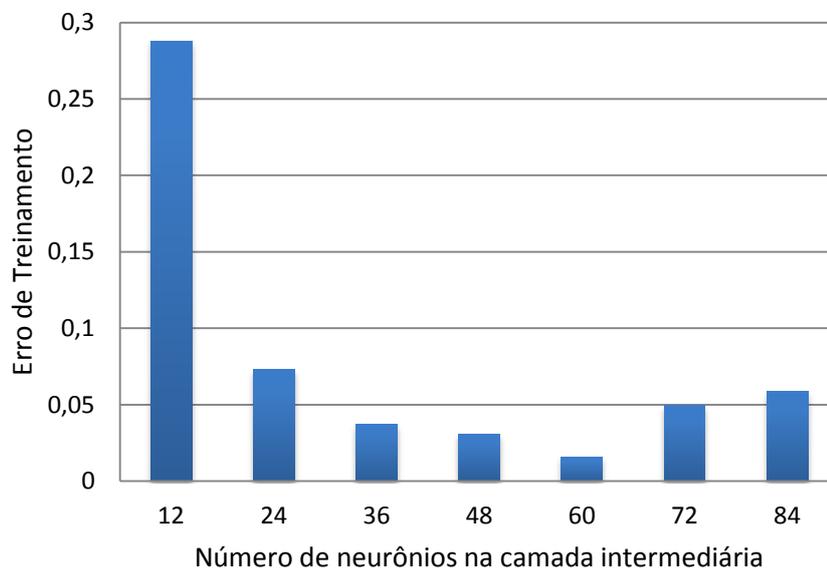


Figura 4.9 – Relação entre número de neurônios na camada intermediária e o erro de treinamento da RNA.

Fonte: Elaborado pelo autor

Os dados, para serem utilizados em uma RNA, devem ser normalizados, para que todos tenham o mesmo peso matemático dentro do cálculo dos pesos da rede. Para aplicação em RNAs, as normalizações podem ser dadas de -1 a 1 ou de 0 a 1, dependendo da função de ativação a ser utilizada. No caso da função de ativação utilizada nesta Dissertação (função sigmoide - Figura 3.3) a normalização definida foi com intervalo de 0 a 1, permitindo que os processos matemáticos ocorram dentro da faixa de operação da função de ativação.

Os treinamentos, validações e análises da RNA proposta foram realizados de duas maneiras, que são distintas pelo método de validação. Inicialmente, foram realizados testes com a Validação *K-Fold*, e posteriormente foram realizados testes com Validação Direta. Os procedimentos e resultados obtidos são apresentados nas duas seções que seguem.

4.10 Treinamento com Validação Cruzada K-Fold

Na Validação Cruzada *K-Fold*, a massa de dados é dividida em K partes e um processo iterativo é realizado K vezes, sendo que em cada vez, uma fração da massa de dados é removida para validação, e o restante é usado para treinamento da rede.

Nessa Dissertação, foi utilizado $K = 10$, que é um valor usual em RNAs e que designa o método conhecido como *Ten-Fold Cross Validation*. A Figura 4.10 ilustra o procedimento em uma das etapas do ciclo de treinamento. Os blocos de dados representados na figura correspondem a toda a massa de dados disponíveis de cada lavoura, nas cinco safras analisadas.

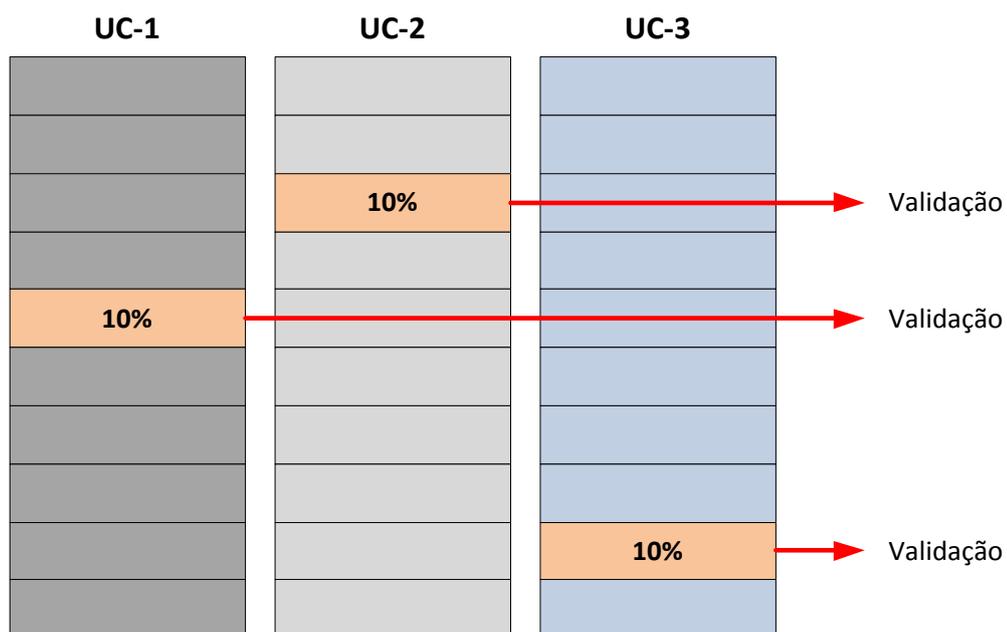


Figura 4.10 – Validação Cruzada com $K = 10$

Fonte: Elaborado pelo autor

Quando se utiliza o *Ten-Fold Cross Validation* é eliminada a influência temporal da série de dados, pois os dados estão apresentados de maneira aleatória para a RNA.

Para facilitar a visualização dos resultados na forma gráfica, foram selecionados aleatoriamente três períodos do conjunto total de dados, que estão apresentados nas Figuras 4.11 a 4.13.

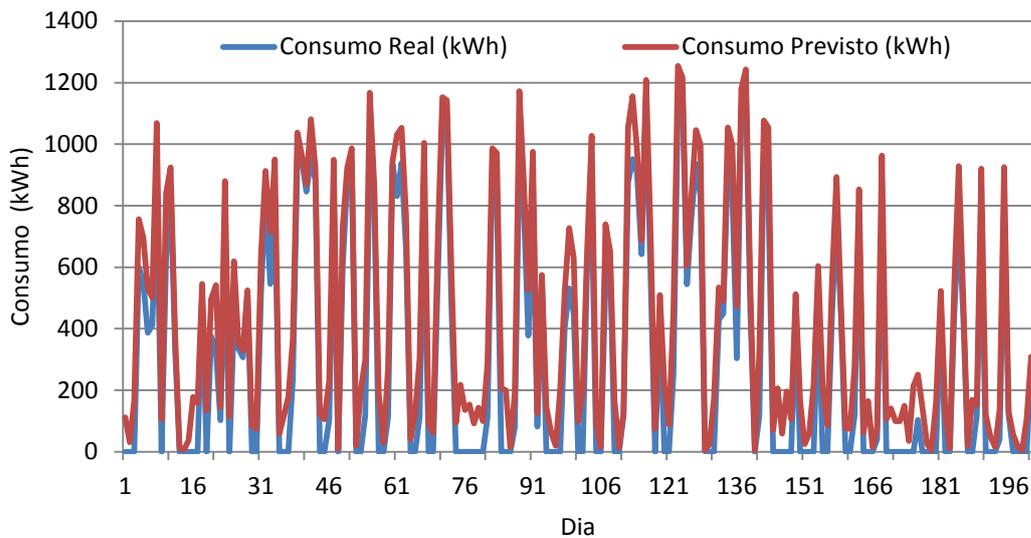


Figura 4.11 – Resultado da RNA com Validação Cruzada (1º conjunto de dados)

Fonte: Elaborado pelo autor

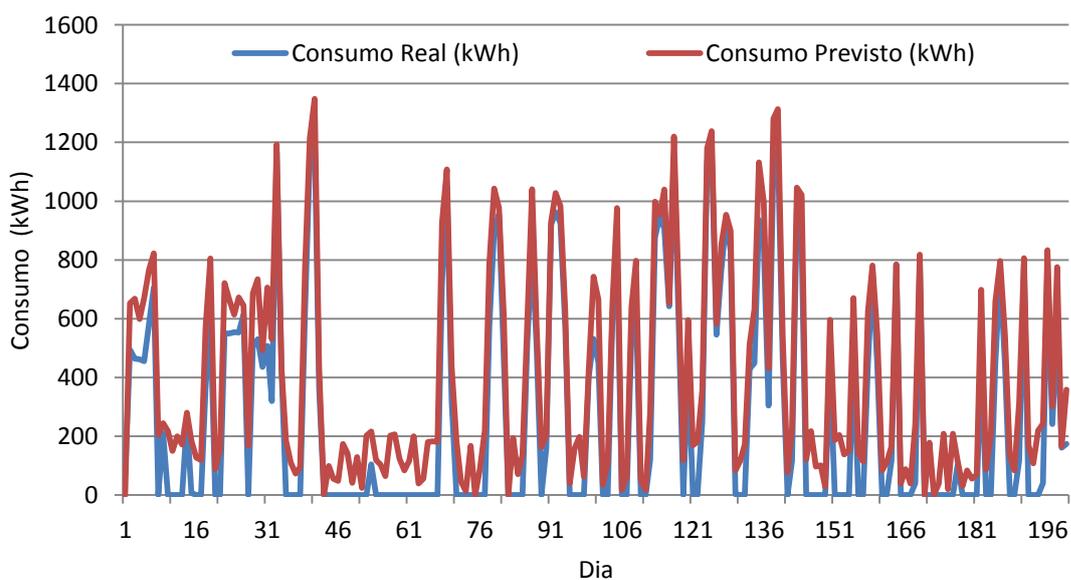


Figura 4.12 – Resultado da RNA com Validação Cruzada (2º conjunto de dados)

Fonte: Elaborado pelo autor

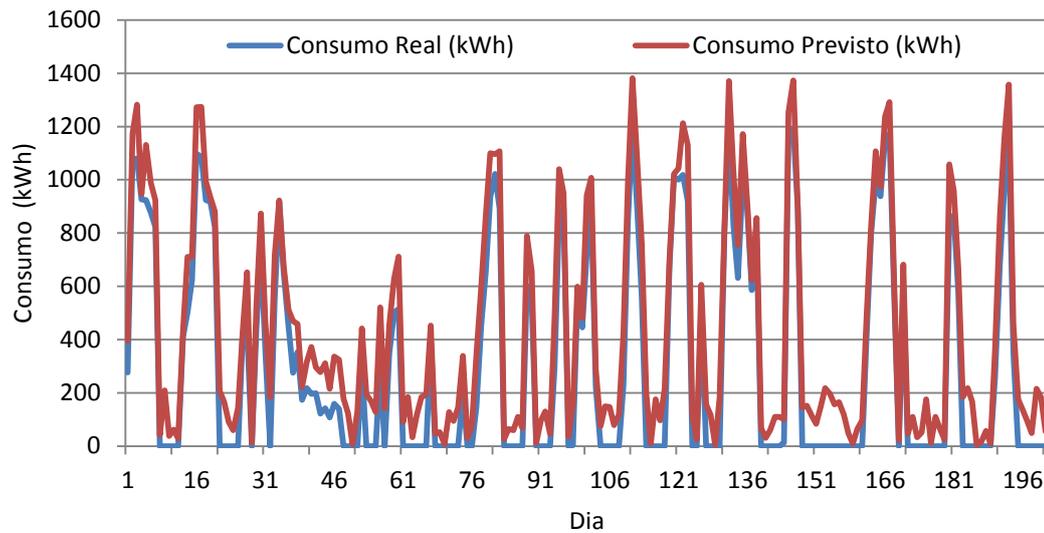


Figura 4.13 – Resultado da RNA com Validação Cruzada (3º conjunto de dados)

Fonte: Elaborado pelo autor

Através da análise gráfica, pode-se verificar que o consumo previsto com a RNA treinada possui boa aproximação do consumo real, apresentando pequenas discrepâncias, principalmente nos pontos de consumo abaixo de 100 kWh. Essas discrepâncias resultam em diferenças percentuais elevadas nesses pontos, superiores a 100% e, para efeito de análise, podem ser consideradas como erros grosseiros – por estarem muito acima do padrão dos demais resultados.

Os erros médios percentuais nos resultados finais de cada conjunto selecionado e apresentado nas Figuras 4.11 a 4.13 foram, respectivamente: 24.86%, 24.15% e 23.45%. O erro total de todo o conjunto de dados testado foi 19,35%.

4.11 Treinamento com Validação Direta

No treinamento com Validação Direta, o conjunto total de dados disponíveis é simplesmente dividido em duas partes, sendo uma para treinamento e outra para validação, sem cruzamento de dados.

Nessa Dissertação, foram treinadas três redes diferentes, sendo que cada uma foi desenvolvida para fornecer os resultados de uma das lavouras analisadas na pesquisa. Nesse caso, foram utilizados os dados de duas lavouras para treinamento, e os dados da terceira lavoura para validação e resultados. Esse processo está ilustrado na Figura 4.14.

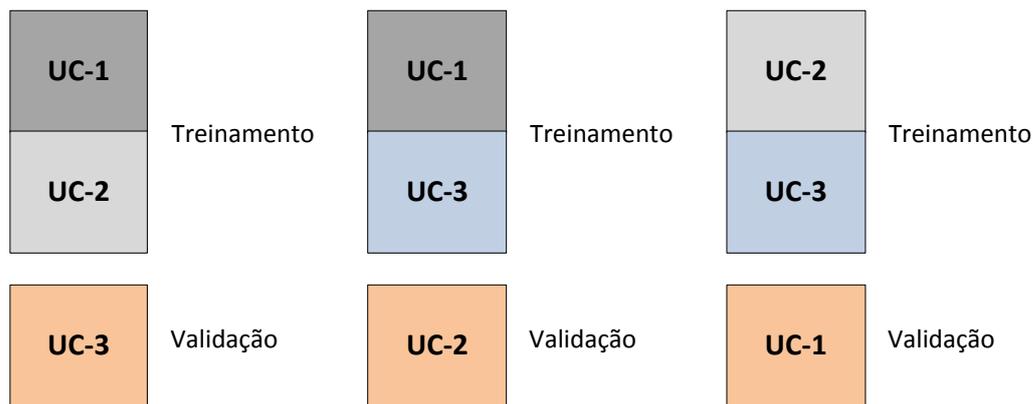


Figura 4.14 – Três redes treinadas com Validação Direta

Fonte: Elaborado pelo autor

A questão temporal é considerada pelo uso dos dados consecutivos apresentados à RNA, em seu treinamento e validação de maneira direta. Os dados são apresentados com uma sequencialidade temporal que permite à RNA reconhecer as tendências sazonais da série de dados.

Os resultados gráficos da análise das RNAs com Validação Direta são apresentados nas Figuras 4.15 a 4.18.

As Tabelas 4.5 a 4.7 resumem os resultados globais de cada unidade consumidora e de cada safra analisada. Pode-se constatar, nessas tabelas, que a possibilidade de ocorrência de perdas não técnicas é baixa, em todas as situações simuladas.

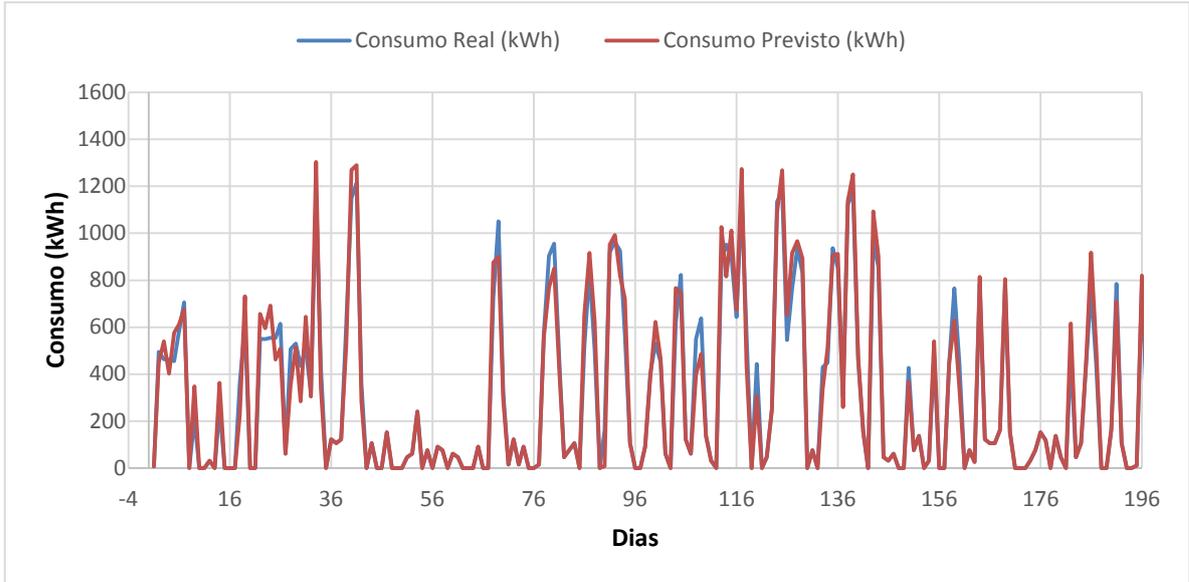


Figura 4.15 – Resultado da análise de UC-1 com RNA e Validação Direta

Fonte: Elaborado pelo autor

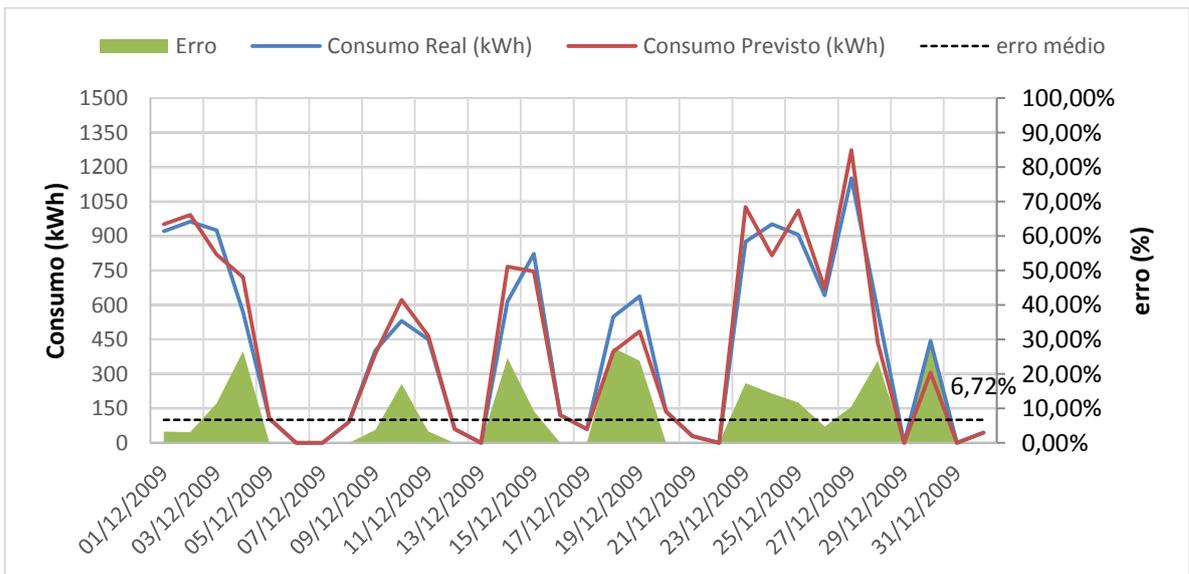


Figura 4.16 – Detalhe do resultado da análise de um mês de UC-1 com RNA e Validação Direta

Fonte: Elaborado pelo autor

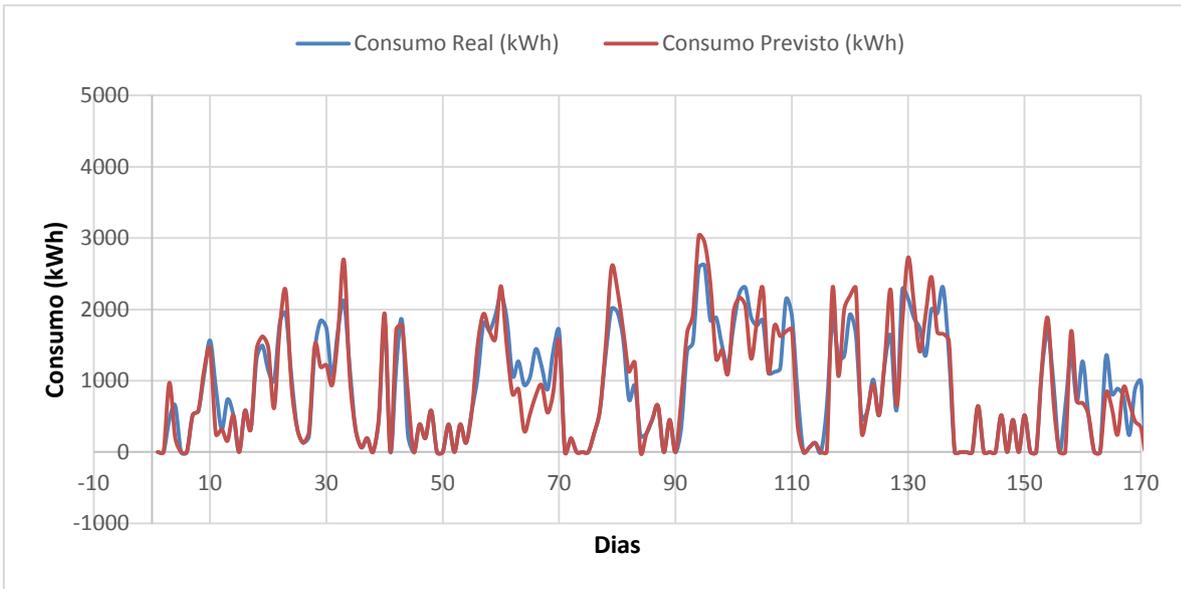


Figura 4.17 – Resultado da análise de UC-2 com RNA e Validação Direta
 Fonte: Elaborado pelo autor

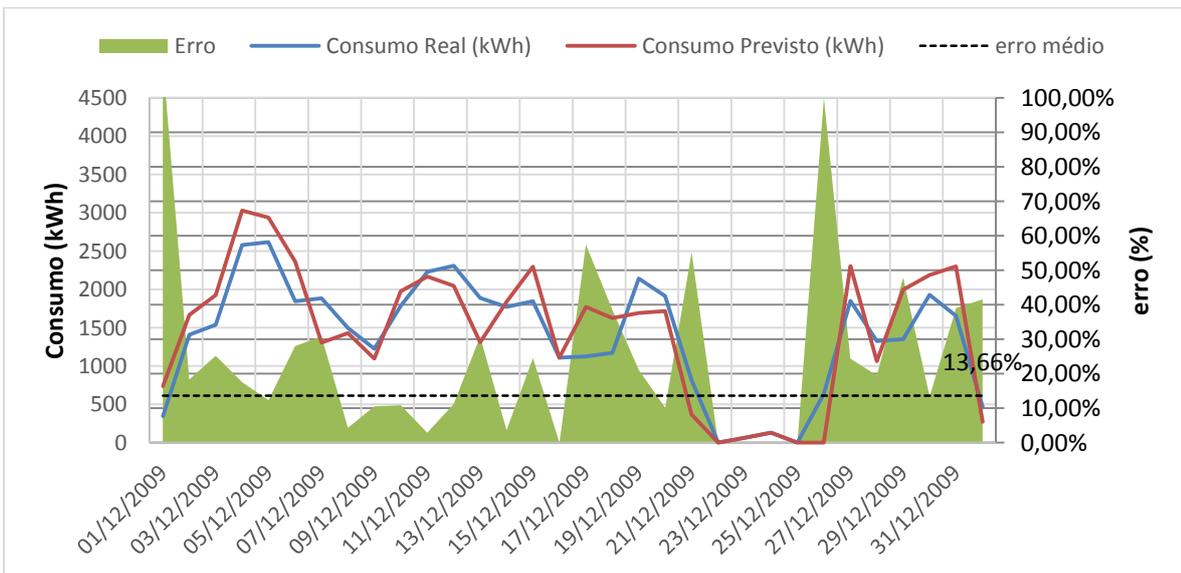


Figura 4.18 – Detalhe do resultado da análise de um mês de UC-2 com RNA e Validação Direta
 Fonte: Elaborado pelo autor

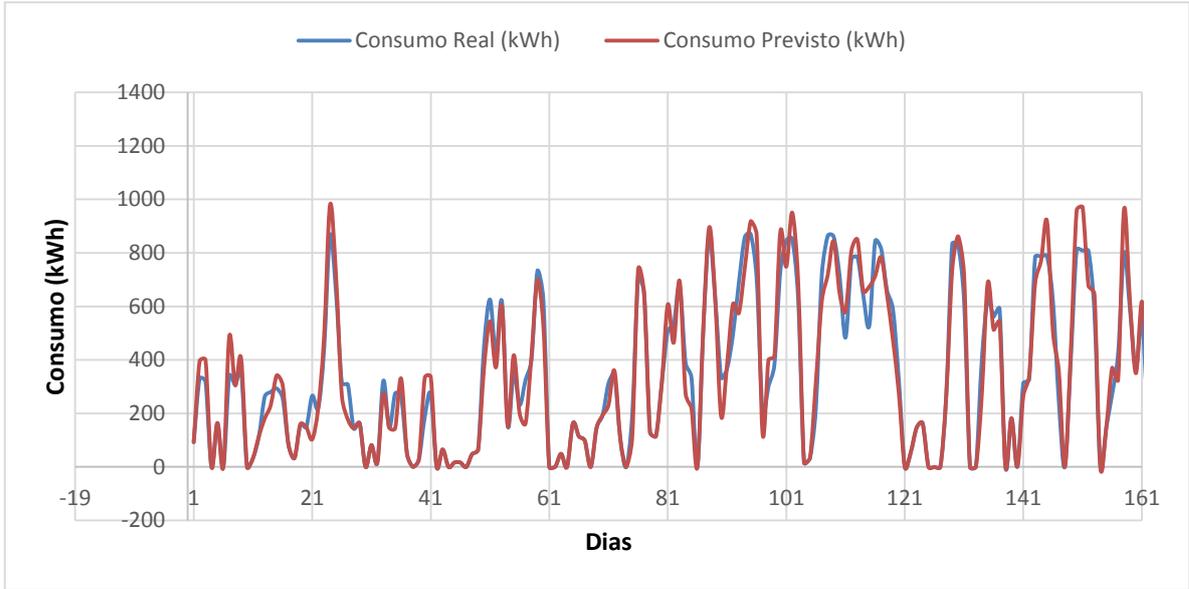


Figura 4.19 – Resultado da análise de UC-3 com RNA e Validação Direta
 Fonte: Elaborado pelo autor

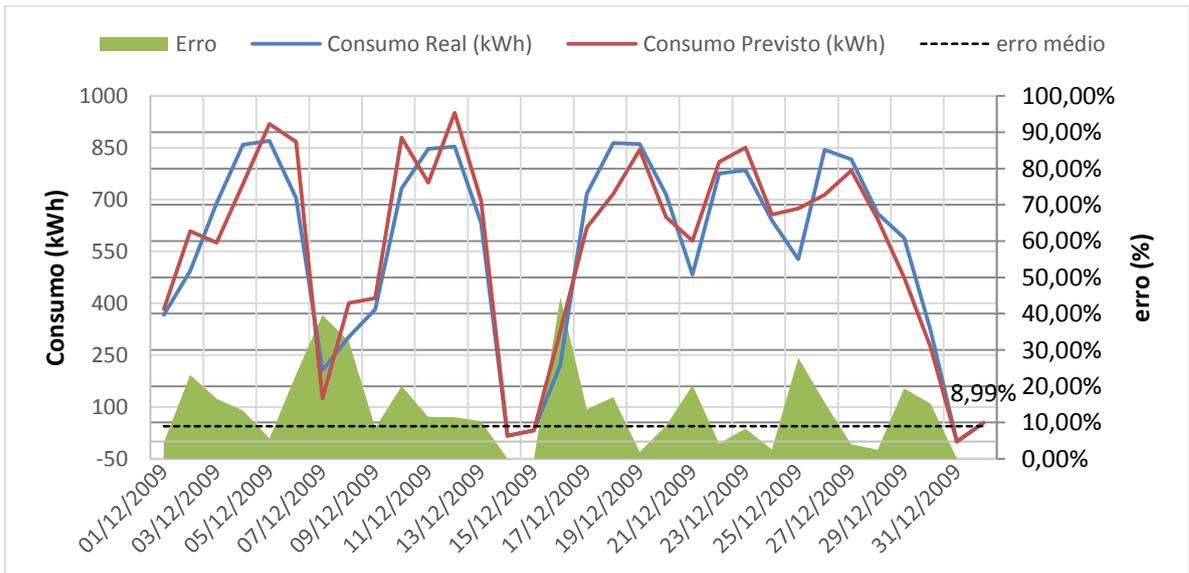


Figura 4.20 – Detalhe do resultado da análise de um mês de UC-3 com RNA e Validação Direta
 Fonte: Elaborado pelo autor

Tabela 4.5 – Resumo da análise de UC-1 modelada com RNA e Validação Direta.

Safra	Consumo médio diário real (kWh)	Consumo médio diário previsto (kWh)	Erro %	Erro Médio Absoluto %
2009-2010	309,25	313,66	1,43 %	6,75%
2010-2011	353,10	354,09	0,28 %	6,68%
2011-2012	306,66	302,19	-1,46 %	6,04%
2012-2013	318,70	320,74	0,64 %	7,92%
2013-2014	418,99	427,10	1,93 %	5,79%

Fonte: Elaborado pelo autor

Tabela 4.6 – Resumo da análise de UC-2 modelada com RNA e Validação Direta.

Safra	Consumo médio diário real (kWh)	Consumo médio diário previsto (kWh)	Erro	Erro Médio Absoluto %
2009-2010	928,06	905,61	-2,42 %	20,03%
2010-2011	550,78	544,14	-1,21 %	3,31%
2011-2012	728,37	709,23	-2,63 %	12,91%
2012-2013	1.372,90	1.365,87	-0,51 %	21,70%
2013-2014	1.296,44	1.287,79	-0,67 %	11,29%

Fonte: Elaborado pelo autor

Tabela 4.7 – Resumo da análise de UC-3 modelada com RNA e Validação Direta.

Safra	Consumo médio diário real (kWh)	Consumo médio diário previsto (kWh)	Erro	Erro Médio Absoluto %
2009-2010	315,11	311,78	-1,06%	11,02%
2010-2011	201,22	201,88	0,33%	6,70%
2011-2012	312,58	321,68	2,91%	11,27%
2012-2013	338,07	337,95	-0,04%	8,79%
2013-2014	307,90	310,76	1,01%	10,44%

Fonte: Elaborado pelo autor

4.12 Resultados de simulação de uma situação de perda não técnica

Nessa seção, são apresentados os resultados obtidos com o modelo de RNA desenvolvido na Seção 4.3.1, para uma situação em que foi aplicada uma falha intencional no sistema de medição de energia de uma das lavouras.

O conjunto de dados apresentado à RNA desconsidera a potência de uma das fases da rede, simulando, por exemplo, uma falha ou fraude em um transformador de corrente do sistema de medição. Nesse caso, a energia faturada pela Concessionária de Energia seria menor que a energia real fornecida.

As Figuras 4.21 e 4.22 ilustram os resultados obtidos através da RNA com Validação Direta, após a aplicação da falha intencional nos dados de consumo da lavoura UC-3.

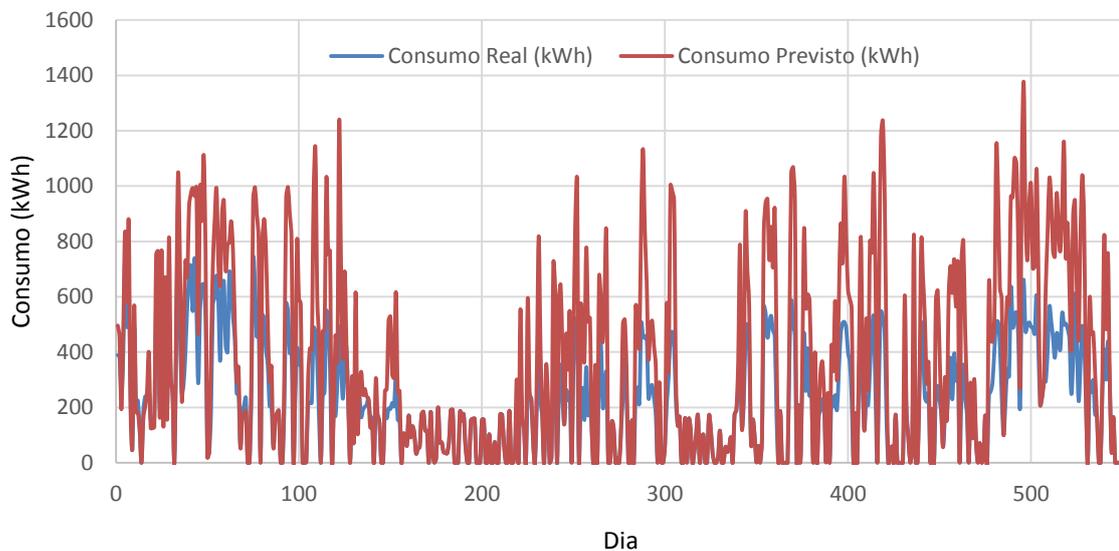


Figura 4.21 – Resultado da simulação de falha na medição de UC-3.

Fonte: Elaborado pelo autor

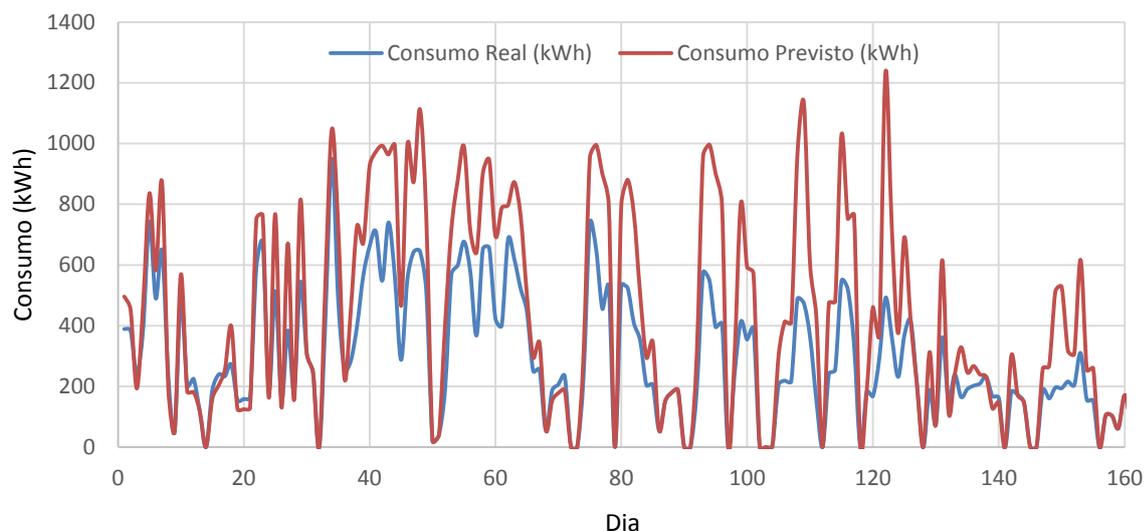


Figura 4.22 – Detalhe do resultado da simulação de falha na UC-3, Safra 2009-2010.

Fonte: Elaborado pelo autor

Os resultados numéricos das análises de simulação de falha estão resumidos na Tabela 4.8.

Tabela 4.8 – Resumo da análise de UC-3 com simulação de falha de medição

Safra	Consumo médio diário real (kWh)	Consumo médio diário previsto (kWh)	Erro
2009-2010	263,52	378,04	43,46 %
2010-2011	172,57	285,59	65,49 %
2011-2012	229,00	413,23	80,45 %
2012-2013	153,55	270,66	76,26 %
2013-2014	276,98	397,56	32,83 %

Fonte: Elaborado pelo autor

Como pode ser verificado na Tabela 4.8, e considerando os critérios estabelecidos na Tabela 4.1, os resultados da análise indicam forte possibilidade de ocorrência de perdas não técnicas nas cinco safras analisadas, refletindo o efeito da simulação de falha intencional aplicada aos dados de consumo dessa lavoura.

CONCLUSÕES

Essa Dissertação abordou o problema das perdas não técnicas nos sistemas de distribuição de energia, tendo como escopo a análise de consumidores rurais que utilizam sistemas de irrigação para lavouras de arroz. O trabalho foi motivado pela relevância desse tipo de consumidor em alimentadores de concessionárias de energia elétrica do Rio Grande do Sul, pela dificuldade de fiscalização e identificação de fraudes nas linhas de distribuição rurais e pela carência de estudos que abordam esse problema.

O objetivo geral do trabalho foi desenvolver um sistema que permitisse identificar possíveis situações de perdas não técnicas, baseado em padrões de consumo e características próprias das lavouras, associados a registros climatológicos.

Para atingir esse objetivo, foi desenvolvida uma metodologia que emprega Redes Neurais Artificiais (RNAs) *Perceptron* de múltiplas camadas, treinadas com um conjunto de dados reais de lavouras de arroz irrigado do Rio Grande do Sul. O conjunto de entradas utilizado compreende: potência instalada por área irrigada, consumo faturado de energia elétrica, precipitação pluviométrica, temperatura, umidade relativa do ar, insolação e evaporação. A saída da RNA é uma indicação do consumo de energia previsto. A validação foi realizada de duas maneiras: utilizando a técnica de Validação Cruzada *Ten-Fold*, e utilizando a técnica de Validação Direta.

Também foi realizado um estudo da correlação entre os dados utilizados no treinamento da RNA e o consumo de energia da lavoura, através de ferramentas estatísticas e do método de Regressão *Stepwise*. Os resultados foram obtidos a partir da análise de três lavouras e das informações de cinco safras de cada uma, totalizando quinze blocos de dados. A análise estatística revelou que existe uma fraca correlação entre os dados de consumo de energia elétrica e de clima analisados. Essa verificação se deu a partir de uma análise básica de correlação entre os dados disponíveis, e também pela aplicação da Regressão *Stepwise*. Os resultados obtidos com o modelo estatístico apresentaram grandes discrepâncias em relação aos valores esperados, e o modelo pode ser considerado como não adequado para os propósitos desse trabalho.

Deve-se ressaltar, no entanto, que as ferramentas estatísticas utilizadas fazem uma relação *linear* da saída com as entradas. Uma análise estatística mais aprofundada poderia levar a um modelo de regressão *não linear* que apresentasse melhores resultados. Aqui também cabe ressaltar que a influência do clima no consumo de energia elétrica dos sistemas de irrigação não é investigada com a mesma intensidade feita para casos de alimentadores urbanos. Nesse sentido, esse trabalho representa uma pequena contribuição a esse estudo.

Com o emprego das Redes Neurais Artificiais foi possível obter um modelo que apresentou uma boa aproximação entre os valores de consumo de energia elétrica faturados e os valores previstos. O teste da RNA pelo método de *k-fold cross validation* permite que se tenha uma visão da sua capacidade de adquirir conhecimento do sistema, separando a componente temporal (sequencialidade) do sistema, deixando o mesmo atemporal onde verifica-se apenas seu comportamento, mesmo este sendo estocástico. Uma vez que a RNA comprovou-se apta, o método de teste direto, mais comumente chamado de validação direta, faz a comprovação do uso da RNA em ambiente de produção, ou seja, como ela será utilizada em sua implementação final. Nesta condição a RNA proposta mostrou um excelente resultado, porém deve-se tomar algumas ressalvas, pois os dados utilizados são de uma mesma região e em lavouras com características parecidas. Quando houver uma diversidade maior de lavouras a massa de dados deve aumentar de maneira substancial e visitas *in loco* devem ser feitas para validar de maneira categórica para que a RNA seja sintonizada a contento.

A partir de uma simulação de falha no sistema de medição de uma das lavouras analisadas - na qual a potência em uma das fases da rede não foi computada no cálculo de consumo de energia - pôde-se demonstrar que a RNA desenvolvida foi capaz de identificar uma situação anômala, indicando forte possibilidade de ocorrência de perdas não técnicas.

O sistema desenvolvido com RNAs pode ser utilizado como mecanismo de apoio à identificação de possíveis perdas não técnicas, por parte das concessionárias de energia elétrica. Diferentemente de métodos baseados em balanços de energia, ele apresenta a característica de identificar pontualmente as unidades consumidoras suspeitas. Também apresenta a vantagem de não utilizar equipamentos e recursos tecnológicos exclusivos para a identificação de perdas. Por

outro lado, é necessária uma grande quantidade de dados para a modelagem da RNA, e os dados - no caso desse trabalho - são oriundos de fontes diferentes.

Embora não tenha sido dado ênfase nessa Dissertação, a metodologia desenvolvida com as RNAs pode ser empregada com o viés da eficiência energética, nos casos em que o erro percentual obtido seja negativo e relevante. Nesses casos, a interpretação que pode ser feita é de que o consumidor está consumindo mais energia elétrica do que o padrão esperado para o seu tipo de lavoura. Os encaminhamentos, nessa situação, podem levar a uma revisão no dimensionamento das bombas de irrigação, obsolescência de equipamentos, e rendimento dos sistemas de bombeamento e tubulações.

Destaca-se, por fim, que essa Dissertação apresenta pioneirismo no estudo de perdas não técnicas em consumidores com sistemas de irrigação de grande porte. Isso é constatado pela ausência de publicações técnicas sobre o tema. Nesse sentido, a metodologia aqui proposta pode ser considerada promissora, e com possibilidade de ser melhorada, considerando as sugestões de continuidade apresentadas a seguir.

Sugestões de Continuidade

Com o objetivo de melhorar o desempenho do sistema de identificação de perdas não técnicas proposto nessa Dissertação, sugere-se as seguintes possibilidades de continuidade do trabalho:

- Aprofundar a análise estabelecendo uma relação entre consumo real esperado (resultado obtido conceitualmente levado em conta às características de lavoura e da rede elétrica) utilizando técnicas de RNA e Algoritmos Genéticos.
- Ampliar o conjunto de variáveis de entrada da RNA, considerando, por exemplo: tipo de relevo e solo, absorção de água do solo, velocidade do vento, vazão média de água dos levantes, altura do levante, entre outros. Esse tipo de informação poderia ser obtido através de parcerias com os produtores, utilizando os resultados do projeto com a finalidade de melhorar a eficiência energética nos consumidores parceiros.

- Ampliar a base de dados de treinamento da RNA, considerando lavouras de diferentes regiões e com diferentes práticas de irrigação.
- Integrar a metodologia proposta com tecnologias de inspeção remotas como, por exemplo, a utilização de imagens de satélite para verificação da área irrigada.

REFERÊNCIAS BIBLIOGRÁFICAS

ABBASY, N. H. **Neural Network Aided Design for Metering System of Power System State Estimation**, Stellenbosch: Proceedings of 4th IEEE AFRICON, vol. 2, pp. 607 – 610, sept. 1996. Abbasy e Hassawy (1996).

ABBASY, N. H.; EL-HASSAWY, W. **Power system State Estimation: ANN-Application to bad data detection and identification**. Stellenbosch: IEEE, 1996, pp. 611-615.

ABRADEE – Associação Brasileira de Distribuidores de Energia Elétrica. **SIG – Sistemas de Informação para a Gestão**. Brasília, DF, 2012.

AES-SUL – **“Experiência da AES Sul nos Projetos de Eficiência Energética em Sistemas de Bombeamento”** Seminário: Diretrizes para comissões de conservação de água e energia, ABAE: Associação Brasileira de Água e Energia. Porto Alegre, 2006.

AGÜERO, J. R. **“Improving the efficiency of power distribution systems through technical and non-technical losses reduction,”** IEEE Transmission and Distribution Conference and Exposition (T&D), Orlando FL, USA, May 2012.

AGUIAR, H.; OLIVEIRA JUNIOR (Coord.). **Inteligência Computacional: Aplicada à Administração, Economia e Engenharia em MatLab®**. São Paulo: Thomson, 2007, 320 p.

ALTRAN, A. B. **Aplicação de Funções de Base Radial em Problemas de Previsão de Cargas Elétricas Via Redes Neurais Artificiais**. 2005. 112 f. Dissertação (Mestrado em Engenharia Elétrica)–Universidade Estadual Paulista, Ilha Solteira, 2005.

ALVES, P. CASANOVA, E. QUIROGAS, O. RAVELO, AND W. GIMENEZ, **“Reduction of non-technical losses by modernization and updating of measurement systems,”** in Proc. IEEE/PES Transmission and Distribution Conf. Expo.: Latin America, 2006, pp. 1–5.

ANEEL Agência Nacional de Energia Elétrica. **Projeto Estratégico: “Atlas de Energia Elétrica do Brasil - Aneel”** Brasília - DF, (2011) Acesso em maio 2014, disponível em: <http://www.aneel.gov.br>.

ANEEL Agência Nacional de Energia Elétrica. **Tratamento regulatório de perdas Não técnicas - Aneel**” Rio de Janeiro - RJ, (2010) Acesso em agosto 2014, disponível em: <http://www.aneel.gov.br>.

ANTONIO, A. B.; TORREÃO, J. R. A.; DO COUTTO FILHO, M. B. **Meter Placement for Power System State Estimation Using Simulated Annealing**. Portugal: IEEE Proceedings of Porto Power Tech. Conference, vol. 3, 5 pp., sept. 2001.

CALILI, R. F. **Desenvolvimento de um sistema de detecção de perda comerciais no sistema de distribuição**. Dissertação de Mestrado – Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2005.

CELIK, M. K.; LIU, W. H. E. **A Practical Distribution State Calculation Algorithm**. Power Engineering Society 1999 Winter Meeting, IEEE. Vol. 1, pp. 442 – 447, jan. 1999. INSPEC Accession Number: 6209741.

CELIK, M.; LIU, W. H. E. **A Practical Distribution State Estimation Algorithm**. Proceedings of PES Winter Meeting, 1999.

CHAUHAN, A.; RAJVANSHI, S. **Non-Technical Losses in power system: A review**. Proceedings of 2013 International Conference on Power, Energy and Control, ICPEC 2013.

CRUZ, R. V.; QUINTERO, C. V.; PEREZ, F. **Detecting Non-Technical Losses in Radial Distribution System Transformation Point through the Real Time State Estimation Method**. Transmission & Distribution Conference and Exposition: Latin America, 2006.

CRUZ, R. V.; QUINTERO, C. V.; PEREZ, F. Detecting Non-Technical Losses in Radial Distribution System Transformation Point through the Real Time State Estimation Method. **Transmission & Distribution Conference and Exposition: Latin America, 2006. TDC '06. IEEE/PES. 2006.**

DENG, Y.; HE, Y.; ZHANG, B. **A Branch-Estimation-Based State Estimation Method for Radial Distribution Systems**. Transactions on Power Systems Delivery, Vol. 17, n. 4, oct. 2002. ISSN : 0272-1724.

DEPURU, S. S. S. R. **Modeling, Detection, and Prevention of Electricity Theft for Enhanced Performance and Security of Power Grid**. Master Thesis, The University of Toledo, Toledo , 2012.

DONADEL, C.; ANICIO, J.; FREDES, M.; VAREJÃO, F.; COMARELA, G.; PERIM, G. **A methodology to refine the technical losses calculation from estimates of non-technical losses.** In: INTERNATIONAL CONFERENCE ON ELECTRICITY DISTRIBUTION, 20., 2009, Praga. Anais... Praga: CIRED, 2009. 4 p.

DORTOLINA, C. A.; NADIRA, R. **The loss that is unknown is no loss at all: a topdown/bottom-up approach for estimating distribution losses.** IEEE Transactions PowerSystems, New York, v. 20, n. 2, p. 1119-1125, 2005.

EFROYMSON, M. A. **"Multiple regression analysis"**. Ralston HS Wilf (Ed. 1.), New York, 1960.

ERGEG (European Regulators Group for Electricity and Gas), **Treatment of Losses by Network Operators.** 2008.

FALCÃO, D.; ASSIS, S. **Linear Programming State Estimation: Error Analysis and Gross Error Identification.** IEEE Transactions on Power Systems, Vol. 3, n. 3, aug. 1988. ISSN : 0885-8950.

FARIA, P.; VALE, Z.; ANTUNES, P. and SOUZA, A. **"Using baseline methods to identify non-technical losses in the context of smart grids,"** in Innovative Smart Grid Technologies Latin America (ISGT LA), 2013 IEEE PES Conference On, 2013, pp. 1–7.

FERREIRA NETO, L. H. T. **Estimador de estado inteligente para análise de perdas em sistemas de distribuição de energia.** 2011. 99 f. Dissertação (Mestrado) - Escola Politécnica da Universidade de São Paulo, 2011.

FIGUEIREDO, R. M.; SANTOS, J. V. C.; CECHIN, A. L. **Um Novo Sistema Neural para Previsão de Carga em Sistemas de Energia Elétrica.** In: VIII CBQEE - Conferência Brasileira sobre Qualidade de Energia Elétrica, 2009, Blumenau. Anais da VIII CBQEE - Blumenau - SC, 2009.

FREEMAN, L.; MARSHALL, M.; WALLING, R. and WEISS, J. **"Loss characteristics of local-delivery distribution systems,"** in Electricity Distribution - Part 1, 2009. CIRED 2009. 20th International Conference and Exhibition on, 2009, pp. 1-4.

GHOSH, A. K.; LUBKEMAN, D. L.; JONES, R. H. **Load modeling for distribution circuit state estimation.** Power Delivery, IEEE Transactions on. 1997.

GONZALEZ-URDANETA, G. **“A Venezuelan experience in the reduction of non-technical power losses”** China International Conference on Electricity Distribution, China, 2010.

HAYKIN, S. **Redes Neurais: Princípios e Prática**; trad. Engel. P. M. - 2. ed. - Porto Alegre, Bookman, 2001. 902 p.

HUANG, S.; LO, Y.; and LU, C. **“Non-technical Loss Detection Using State Estimation and Analysis of Variance.”** IEEE Trans. Power Syst., pp. 2959-2966, 2013.

HUANG, S. J.; SHIH, K. R.; **“Short-Term Load Forecasting Via ARMA Model Identification Including Non-Gaussian Process Considerations”**, In: IEEE Transactions on Power Systems, v. 18, n. 2, pp. 673-679, May., 2003. Disponível em: <<http://www.ieee.org.br>>. Acesso em: 26 fev., 2013.

GUERRERO, J. I.; LEÓN, C.; BISCARRI, F.; MONEDERO, I.; BISCARRI, J.; and MILLÁN, R. **“Increasing the efficiency in non-technical losses detection in utility companies”**. MELECON 2010, 15th IEEE Mediterranean Electromechanical Conference. Pp. 136-141. Valleta, Malta. 25-28 April, 2010.

KENNEDY, J.; EBERHART, R. **Particle swarm optimization**. Perth, Australia: Proceedings IEEE Int. Conf. Neural Networks, pp. 1942–1948, vol. 4, 1995.

KERSTING, W. H. and PHILLIPS, W. H. **“Modeling and analysis of rural electric distribution feeders.”** IEEE Transactions on Industry Applications. vol. 28(4), pp. 767.

KOTIUGA, W. W.; VIDYASAGAR, M. **Bad Data Rejection Properties of Weighted Least Absolute Value Techniques Applied to Static State Estimation**. IEEE Transactions on Power Apparatus and Systems, v. PAS-101, n. 4, p. 844-853, 1982. ISSN: 0272-1724.

MÉFFE, A. **Cálculo de Perdas Técnicas em Sistemas de Distribuição – Modelos Adequáveis às Características do Sistema e à Disponibilidade de Informações**. São Paulo, Tese de Doutorado, EPUSP, 2007.

MÉFFE, A. **Metodologia para Cálculo de Perdas Técnicas por Segmento do Sistema de Distribuição**, Dissertação de Mestrado, Departamento de Engenharia de Energia e Automação Elétricas da Escola Politécnica da Universidade de São Paulo, São Paulo, 2001.

MÉFFE, A.; OLIVEIRA, C. C. B.; KAGAN, N.; JONATHAN, S.; CAPARROZ, S.; CAVARETTI, J. L. **Technical and economic analysis for the reduction of losses in distribution systems**. In: Transmission and Distribution Conference and Expositio: Latin America, 2002, São Paulo. IEEE/PES, 2002. 6p.

MONEDERO, F.; BISCARRI, C.; LE & OACUTE, N, J. E R. BISCARRI MILL & AACUTE; **"MIDAS: Detecção de perdas não-técnicas em consumo de energia elétrica utilizando redes neurais e técnicas estatísticas"**, Proc. Int. Conf. Ciência e Aplicações Computacional , vol. 3984, 2006.

MONTGOMERY, D. C. and RUNGER, G. C. **"Engineering Statistics"**, 3rd ed. Hoboken, NJ: Wiley, 2009.

MONTICELLI, A. **State Estimation in Electric Power Systems: A Generalized Approach**. Massachusetts: Kluwer Academic Publishers, 2000.

NAKA, S. et al. **Practical Distribution State Estimation Using Hybrid Particle Swarm Optimization**. Columbus, Ohio: IEEE Power Engineering Society Winter Meeting, pp. 815-820, 2001.

NAKA, S.; GENJI, T.; FUKUYAMA, Y. **A hybrid particle swarm optimization for distribution state estimation**. IEEE Transaction on Power Systems, Vol. 18, n. 1, feb. 2003. INSPEC Accession Number: 7992583.

NANNI, M. **Estimador de Estado Robusto Baseado no Método da Mínima Mediana**. São Carlos: Tese (Mestrado em Engenharia Elétrica). Escola de Engenharia de São Carlos da Universidade de São Paulo, 145 f., abr. 2009.

Nizar, A. H.; Zhao, J. H; Dong Z. Y. **Customer Information System Data Pre-Processing with Feature Selection Techniques for Non-Technical Losses Prediction in an Electricity Market**. Power System Technology, 2006. PowerCon 2006 International Conference on DOI: 10.1109/ICPST.2006.321964.

OLIVEIRA, C. C. B.; MÉFFE, A.; CAMPOS, M. B. M. **Implantação em área piloto e treinamento para utilização do software, Relatório Técnico 04, Projeto de P&D: Metodologia e Determinação do Nível Ótimo de Perdas em um Sistema de Distribuição de Energia**, São Paulo, 2005.

OLIVEIRA, M. O. et al. **Projeção de Demanda de Energia Elétrica a Curtíssimo Prazo Considerando o Índice de Desconforto**, In: SBSE – SIMPÓSIO BRASILEIRO DE SISTEMAS ELÉTRICOS, 3, Belém, 2001. **Anais...** Disponível em: < <http://www.researchgate.net/>>. Acesso em: mar. 2014.

PARUCHURI, V and DUBEY, S. **“An Approach to Determine Non-Technical Energy Losses in India”**, Proc. of 14th Int'l Conf. on Advaned Comm. Tech., 2012, pp. 111-115.

PASQUOTTO, J. L. D. **Previsão de Séries Temporais no Varejo Brasileiro: Uma Investigação Comparativa da Aplicação de Redes Neurais Recorrentes de Elman**. 2010. 200 f. Dissertação (Graduação em Administração)–Universidade de São Paulo, São Paulo, 2010.

PEREIRA, L. A. M.; AFONSO, L. C. S.; PAPA, J. P.; VALE, Z. A.; RAMOS, C. C. O.; GASTALDELLO, D. S.; SOUZA, A. N. **Multilayer perceptron neural networks training through charged system search and its Application for non-technical losses detection**. Innovative Smart Grid Technologies Latin America (ISGT LA), 2013 IEEE PES Conference On. 2013.

QUEIROS, L. M. O. **Estimação e Análise das Perdas Técnicas na Distribuição de Energia Elétrica**. Tese de Doutorado – Universidade Federal de Alfenas – universidade estadual de Campinas, Campinas, 2010.

RAKPENTHAI, C.; PREMRUDEEPREECHACHARN, S.; UATRONGJIT, S.; WATSON, N. R. **Measurement Placement for Power System State Estimation using Decomposition Technique**. Electric Power Systems Research, Elsevier, vol. 75, n. 1, pp. 41 – 49, jul. 2005, INSPEC Accession Number: 8332124.

RAMÍREZ J. M.; BAROCIO. E. E. **Solving State Estimation in Power Systems by an Interior Point method**. Vancouver: ELSEVIER- International Journal of Electrical Power & Energy Systems, n. 22, pp. 355-365, jun. 2000.

REIS, C. Z. **Eficácia de Solução Tecnológica para Redução de Furtos de Energia Elétrica em Empresas Distribuidoras: Estudo de Casos**. Dissertação de Mestrado – Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2005.

RIBEIRO, S. S.; CAZES, T.; MANO, R. F.; MAIA, D. **“Investment optimization methodology applied to investments on non-technical losses reduction actions.”** Computers and Communications (ISCC), 2012 IEEE Symposium on DOI: 10.1109/ISCC.2012.6249321.

ROSENBLATT, F. **“The perceptron: A probabilistic model for information storage and organization in the brain”**. Psychological Review, Vol 65(6), Nov 1958.

SARIC, A. T.; CIRIC, R. M. **Integrated Fuzzy State Estimation and Load Flow Analysis in Distribution Networks**. IEEE Transaction on Power Delivery, Vol. 18, n. 2, apr. 2003. ISSN: 0885-8977.

SCHWEPPE, F. C. **Power System Static-State, Part II: Approximate Model**. IEEE transactions on Power Apparatus and Systems, V.PAS-89, n. 1, p. 120-130, 1970, ISSN: 0018-9510.

SEBER, G. A. F. and LEE, A. J. **“Linear regression analysis”** (2nded.). Hoboken, NJ: Wiley, 2003.

SHAFIU, A.; JENKINS, N.; STRBAC, G. **Measurement Location for State Estimation of Distribution Networks with Generation**. IEE Proceedings on Generation, Transmission and Distribution,, vol. 152, n. 2, pp. 240–246, march 2005. ISSN: 1350-2360.

SIMÕES, M. G.; SHAW, I. S. **Controle e Modelagem Fuzzy**. Editora: EDGARD BLUCHER. 200 p. São Paulo, 2007.

WALPOLE, R. and MYERS, R. **“Probability and Statistics for Engineers and Scientists”**. New York: Macmillan, 2009.

ZARCO, P.; EXPÓSITO, A. G. **Power System Parameter Estimation: A Survey**. IEEE Transactions on Power Systems, vol. 15, n. 1: 216-222, 2000.

ZIMMERMAN, H. J. **Fuzzy Set Theory – and Its Application**. Boston: Kluwer-Nijhoff Publishing, 3rd Edition, pp. 281, 1984.

ZOUCAS, F. A. M. **Aplicação de Redes Neurais para Previsão de Séries Temporais em Segmentos Industriais**, 2009. 170 f. Dissertação (Mestrado em Engenharia Mecânica)–Centro Universitário da FEI, São Bernardo do Campo, 2009.

APÊNDICE A – RESULTADOS DE CORRELAÇÃO ENTRE CONSUMO DE ENERGIA E VARIÁVEIS CLIMÁTICAS

Nesse Apêndice são apresentados os resultados extraídos do programa Minitab®, para a correlação entre as variáveis de entrada da pesquisa, da Unidade Consumidora 1 (UC-1). Os dados estão apresentados no formato disponibilizado pelo programa, sendo que em cada linha de variável se observa o Coeficiente de Correlação de Pearson e o Valor p .

UC1 Safra 2009/2010

Welcome to Minitab, press F1 for help.

Retrieving project from file: 'D:\MAICON\1UFISM\DROPOBOX\DISSERTAÇÃO\TESTES MINITAB\TESTE 1.MPJ'

Results for: Worksheet 4

Correlations: Cosumo real ; Precipitacao; Insolacao; Evaporacao P; ...

	Cosumo real (kWh)	Precipitacao (mm)	Insolacao
Precipitacao (mm)	0,063 0,381		
Insolacao	0,077 0,286	-0,386 0,000	
Evaporacao Piche	-0,060 0,402	-0,357 0,000	0,276 0,000
Temp Media	0,026 0,720	-0,325 0,000	0,247 0,000
Umidade Relativa	-0,036 0,617	0,344 0,000	-0,780 0,000

	Evaporacao Piche	Temp Media	Umidade Relativa
Temp Media	0,405 0,000		
Umidade Relativa	-0,365 0,000	-0,179 0,012	

Cell Contents: Pearson correlation
P-Value

* NOTE * All values in column are identical.

UC1 Safra 2010/2011

	Consumo Real	Precipitacao	Insolacao
Precipitacao	-0,037 0,607		
Insolacao	0,092 0,196	-0,291 0,000	
Evaporacao Piche	0,142 0,045	-0,254 0,000	0,254 0,000
Temp Media	0,003 0,970	-0,022 0,755	0,186 0,009
Umidade Relativa	-0,112 0,115	0,322 0,000	-0,722 0,000

	Evaporacao Piche	Temp Media	Umidade Relativa
Temp Media	0,199 0,005		
Umidade Relativa	-0,337 0,000	0,076 0,289	

UC1 Safra 2011/2012

	Consumo Real	Precipitacao	Insolacao
Precipitacao	-0,147 0,155		
Insolacao	0,072 0,491	-0,283 0,006	
Evaporacao Piche	0,101 0,331	-0,485 0,000	0,129 0,213
Temp Media	0,059 0,573	-0,202 0,050	0,203 0,049
Umidade Relativa	-0,155 0,133	0,365 0,000	-0,706 0,000

	Evaporacao Piche	Temp Media	Umidade Relativa
Temp Media	0,369 0,000		
Umidade Relativa	-0,189 0,067	-0,084 0,418	

UC1 Safra 2012/2013

Precipitacao	Consumo Real	Precipitacao	Insolacao
	-0,122		
	0,161		
Insolacao	0,136	-0,291	
	0,117	0,001	
Evaporacao Piche	0,123	-0,280	0,331
	0,158	0,001	0,000
Temp Media	0,092	-0,197	0,351
	0,288	0,022	0,000
Umidade Relativa	-0,101	0,385	-0,800
	0,244	0,000	0,000
Temp Media	Evaporacao Piche	Temp Media	Umidade Relativa
	0,482		
	0,000		
Umidade Relativa	-0,515	-0,413	
	0,000	0,000	

UC1 Safra 2013/2014

Precipitacao	Consumo Real	Precipitacao	Insolacao
	-0,155		
	0,119		
Insolacao	0,247	-0,256	
	0,012	0,009	
Evaporacao Piche	0,355	-0,304	0,179
	0,000	0,002	0,070
Temp Media	0,383	-0,238	0,287
	0,000	0,015	0,003
Umidade Relativa	-0,371	0,242	-0,740
	0,000	0,014	0,000
Temp Media	Evaporacao Piche	Temp Media	Umidade Relativa
	0,463		
	0,000		
Umidade Relativa	-0,470	-0,484	
	0,000	0,000	

APÊNDICE B – REDE NEURAL ARTIFICIAL DESENVOLVIDA EM PYTHON®

Nesse Apêndice é apresentado o código de treinamento e validação da Rede Neural Artificial proposta nessa Dissertação. O código foi desenvolvido em Linguagem de Programação Python® e está apresentado na Figura Apêndice B-1.

```
Python 2.7.8: teste (4).py - C:\Users\MAICON\Downloads\teste (4).py
File Edit Format Run Options Windows Help

# Chamada das bibliotecas
import numpy as np
import neurolab as nl
import misc
import ann_handling as ah
import ann_plot as ap
import ann_error_measure as em

# Leitura do arquivo para as variáveis
carga, precip, insol, evap, temp, umidade, cargaprev = misc.read_data("c:/trafo.txt")

# Normalização das variáveis de 0 a 1
carga_n = misc.normalize(carga, 1, 0)
precip_n = misc.normalize(precip, 1, 0)
insol_n = misc.normalize(insol, 1, 0)
evap_n = misc.normalize(evap, 1, 0)
temp_n = misc.normalize(temp, 1, 0)
umidade_n = misc.normalize(umidade, 1, 0)
cargaprev_n = misc.normalize(cargaprev, 1, 0)

# Configuração dos dados para leitura da RNA
carga_in = misc.conv_list2array(carga_n)
precip_in = misc.conv_list2array(precip_n)
insol_in = misc.conv_list2array(insol_n)
evap_in = misc.conv_list2array(evap_n)
temp_in = misc.conv_list2array(temp_n)
umidade_in = misc.conv_list2array(umidade_n)
cargaprev_in = misc.conv_list2array(cargaprev_n)

in_data = misc.matrix_concat(carga_in, precip_in, insol_in, evap_in, temp_in, umidade_in)
t_data = cargaprev_in

# Separação dos dados para treinamento
data_in = in_data[0:200] # Dados de entrada da RNA
data_t = t_data[0:200] # Dados alvo da RNA

# Separação dos dados para validação
data_test = in_data[201:len(in_data)] # Dados de entrada da RNA
data_val = t_data[201:len(t_data)] # Dados alvos da RNA

# Criação da RNA
```

(Continua)

(Conclusão)

```
ann = nl.net.newff([[0,1],[0,1],[0,1],[0,1],[0,1],[0,1]], [16, 1])

# Treinamento da RNA
err = ann.train(data_in, data_t, epochs=10000, show=100, goal=0.02)

# Simulação (Validação) da RNA
res = ah.ann_sim(ann,data_test)

# Cálculo do RMSE da RNA (verificação quantitativa)
print("RMSE: ")
print(em.rmse(res, data_val))

# Plot do gráfico da resposta da RNA (verificação qualitativa)
ap.plot_data(data_val, res, "", "")
```

Figura B-1 - Código de programação da RNA

Fonte: Desenvolvedor

APÊNDICE C – CÓDIGO PARA ANÁLISE DE REGRESSÃO *STEPWISE* E RESULTADOS PARA UC-3

Nesse Apêndice são apresentados o código de programação para implementação da metodologia de Regressão Stepwise no Matlab, e os resultados obtidos a partir da análise da lavoura da unidade consumidora UC-3. A base do código foi desenvolvida pelo Professor Rodrigo Mejia (@) e adaptada para este trabalho.

O Quadro C.1 apresenta o código utilizado para a análise da UC-3, Safra 2013/2014.

Regressão Stepwise – UC3 – Safra 2013/2014

Parte 1 - Calculo dos parâmetros para determinar o indicador de consumo

```
% Autor:
% Prof. Rodrigo Iván Goytia Mejía
% rmejia@unisin.br
% All rights reserved, 2014
%
% Observations:
% . Sample developed for Maicon Evaldt
% . Using the code do not forget to cite the author.
% -----

clear all
close all
clc

% 1. Carregando os dados
%-----

% 1.1. Definindo o caminho dos arquivos em EXCEL que contem os dados
%-----
pathfile = 'C:\09_ACADEMIC\Organizar\Projetos Pesquisa\Docs Stepwise
Regression\Projeto Maicon\02_Calculo
Indicador\Cliente_3_UC20399\Cliente_3_UC20399.xls';

% 1.2. Fazendo a leitura dos dados da matriz de dados X
%-----
pagina      = 'Safra_2013_2014';
range       = strcat('E',num2str(2),':','I',num2str(103));
[Xvars aux] = xlsread(pathfile,pagina,range);

% 1.2. Fazendo a leitura dos dados do vetor de dados Y
%-----
pagina      = 'Safra_2013_2014';
range       = strcat('D',num2str(2),':','D',num2str(103));
```

(Continua)

```
[Yopts aux] = xlsread(pathfile,pagina,range);

save('XY_Safra_2013_2014', 'Xvars', 'Yopts')

% 2. Normalização Z-score
%-----
Xvars = [Xvars Xvars.^2];

filcol = size(Xvars);
fil = filcol(1);
Xnormal = (Xvars-repmat(mean(Xvars),fil,1))./repmat(std(Xvars),fil,1);
Xnormal(isnan(Xnormal))=0;

Yoptnor = (Yopts-mean(Yopts))/std(Yopts);

% 3. Calculando os parametros do modelo utilizando StepWise Regression
%-----

var(1) = 1; % Precipitacao
var(2) = 1; % Evaporacao Piche
var(3) = 1; % Temp Comp Media
var(4) = 1; % Umidade Relativa Media
var(5) = 1; % Area Plantada
var(6) = 0; % Carga Instalada
var(7) = 0; %
var(8) = 1; %
var(9) = 1; %
var(10) = 1; %
% var(11) = 1; %
% var(12) = 1; %
% var(13) = 0; %
% var(14) = 0; %

initvar = find(var==1);

stepwise(Xnormal,Yoptnor, initvar)

% Logo de determinar os Xi que formarão parte do modelo, exportar os
% coeficientes e Intercept e modificar o arquivo IConsumo.m logo é só
% validar os resultados com o arquivo teste_IConsumoTE.
```

Parte 2 - Determina indicador

```
Intercept = 2.56631e-16;

beta = [
    0
    -0.74458
    0
    0.57858
    0
    0
    0.6385
    0
    0
    0];

% beta = [
%
% -38849.1184075767
% 178164.311643218
% -55683.160567067
% 0
% 0
```

(Continuação)

```

%           -9256.12362990054
%           60433.0532804545
%           22038.639223978
%           -45109.0120245811
%           0
%           0
%           0];
b0      = Intercept;
b(1)    = beta(1);
b(2)    = beta(2);
b(3)    = beta(3);
b(4)    = beta(4);
b(5)    = beta(5);
b(6)    = beta(6);
b(7)    = beta(7);
b(8)    = beta(8);
b(9)    = beta(9);
b(10)   = beta(10);
% b(11) = beta(11);
% b(12) = beta(12);
% b(13) = beta(13);
% b(14) = beta(14);
Idxval = b0+span+b(1)*X(1)...
         +b(2)*X(2)...
         +b(3)*X(3)...
         +b(4)*X(4)...
         +b(5)*X(5)...
         +b(6)*X(6)...
         +b(7)*X(7)...
         +b(8)*X(8)...
         +b(9)*X(9)...
         +b(10)*X(10);
%         +b(11)*X(11)...
%         +b(10)*X(12)...
%         +b(11)*X(13)...
%         +b(12)*X(14);

End

```

Parte 3 – Validação do indicador de consumo

```

clear all
close all
clc
% 1. Determinando os parametros de entrada
% -----
load XY_Safra_2013_2014
% load XY_Safra_2012_2013
% load XY_Safra_2011_2012
% load XY_Safra_2010_2011
tag      = 50;           % Variável a ser analisada
span     = 0;           % Parametro de ajuste pelo operador

% Isto é so para testar se o modelo é adequado (A seguinte linha tem que ser
eliminada)

% 2. Normalização Z-score
% -----
Xvars = [Xvars Xvars.^2];
filcol = size(Xvars);
fil    = filcol(1);

```

(Continua)

```

Xnormal = (Xvars-repmat(mean(Xvars),fil,1))./repmat(std(Xvars),fil,1);
Xnormal(isnan(Xnormal))=0;

% X = [Xvars(tag,:) Xvars(tag,:).^2]; % <--- Isto é usando dados com os que foi
gerado o modelo

% Aqui tem que ser inserido os valores dos dados medidos para validação
% X(1) = 0; % Precipitacao
% X(2) = 2.3; % Evaporacao Piche
% X(3) = 11.06; % Temp Comp Media
% X(4) = 75.25; % Umidade Relativa Media
% X(5) = 140; % Area Plantada
% X(6) = 220; % Carga Instalada
% X(7) = 0; %
% X(8) = 5.29; %
% X(9) = 122.3236; %
% X(10) = 5662.5625; %
% X(11) = 19600; %
% X(12) = 48400; %

% 2. Cálculo do indicador consolidado
% -----
Idxval = IConsumo_Cliente_3_UC20399_Safra_2013_2014(Xnormal,span);

for k=1:length(Xvars)
    Index(k) = IConsumo_Cliente_3_UC20399_Safra_2013_2014(Xnormal(k,:),span);
end

Idxval = Index(tag);

% 3. Impressao dos resultados
% -----
disp('-----')
disp(' Cálculo do indicador consolidado de consumo')
disp('-----')

% Normalizando os valores
% -----
filcol = size(Xvars);
fil = filcol(1);
Xn = (Xvars-repmat(mean(Xvars),fil,1))./repmat(std(Xvars),fil,1);
Xn(isnan(Xn))=0;

Yn = (Yopts-mean(Yopts))./std(Yopts);
figure (1)
t = (1:length(Xvars))';
plot(t,Xn(:,1),'k','linewidth',1), hold on
plot(t,Xn(:,2),'-k','linewidth',1), hold on
plot(t,Xn(:,3),'--k','linewidth',2), hold on
plot(t,Xn(:,4),'-k','linewidth',1), hold on
plot(t,Xn(:,5),':k','linewidth',1), hold on
% plot(t,Xn(:,6),'b','linewidth',1), hold on
% plot(t,Xn(:,7),'--b','linewidth',1), hold on
plot(t,Yn,'r','linewidth',2)
legend('Precipitacao (mm)', 'Insolacao', 'Evaporacao Piche', 'Temp Media', 'Umidade
Relativa Media', 'Área (há)', 'Carga Instalada (kw)')

figure (2)
subplot(7,1,1), plot(t,Yn,'k','linewidth',2), title('Consumo (kwh)')
subplot(7,1,1), plot(t,Xn(:,1),'k','linewidth',2), title('Precipitacao (mm)')
subplot(7,1,2), plot(t,Xn(:,2),'k','linewidth',2), title('Insolacao')
subplot(7,1,3), plot(t,Xn(:,3),'k','linewidth',2), title('Evaporacao Piche')
subplot(7,1,4), plot(t,Xn(:,4),'k','linewidth',2), title('Temp Comp Media')
subplot(7,1,5), plot(t,Xn(:,5),'k','linewidth',2), title('Umidade Relativa Media')

```

(Conclusão)

```

% subplot(7,1,6), plot(t,Xn(:,6),'k','linewidth',2), title('Area Plantada (há)')

% subplot(7,1,7), plot(t,Xn(:,7),'k','linewidth',2), title('Carga Instalada (kw)')
xlabel('Tempo, dias')

figure (3)
plot(Yn,'Color',[0.7294 0.8314 0.9569]), ylabel('IConsumo'), hold on
plot(tag,Yn(tag),'ko','linewidth',1)
plot(tag,Idxval,'ro','linewidth',1)
legend('Yopt', 'Yopts(tag)', 'Idxval')

figure (4)
plot(Yn,'k'), ylabel('IConsumo'), hold on
plot(Index,'r')
legend('Yopt', 'Index')

end

```

Quadro C.1 – Código Matlab utilizado para análise de Regressão *Stepwise*
 Fonte: Desenvolvedor

As Figuras C.1 a C.10 apresentam os resultados obtidos para a análise de Regressão *Stepwise* dos dados das cinco safras da UC-3.

Nas figuras da interface gráfica (C.1, C.3, C.5, C.7 e C.9), no canto superior esquerdo, apresentam-se as estimativas dos coeficientes de todos os termos potenciais, com barras horizontais, indicando os intervalos de confiança. A cor vermelha indica que, inicialmente, os termos não estão no modelo. Os valores mostrados na tabela ao lado do gráfico são os que resultariam se os termos fossem adicionados ao modelo.

As figuras C.2, C.4, C.6, C.8 e C.10 apresentam um comparativo do consumo real (*Yopt* - linha preta) com o consumo estimado (*Index* - linha vermelha). Observa-se que em todas as safras analisadas, há uma grande discrepância entre as duas curvas.

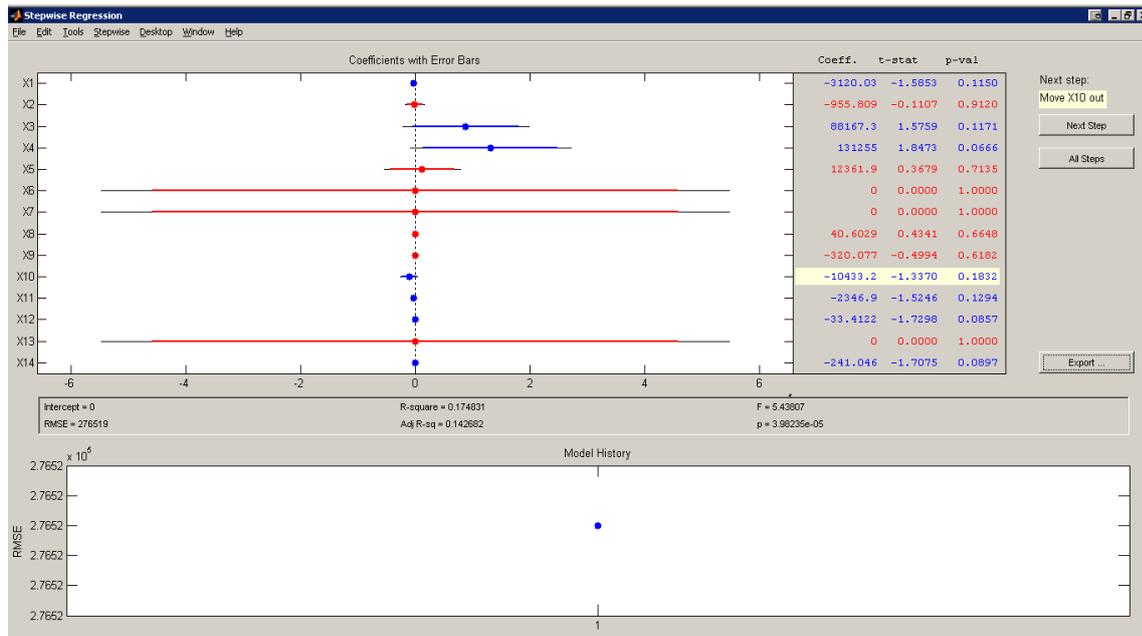


Figura C.1 – Interface gráfica da análise de Regressão *Stepwise*, UC-3, Safra 2009/2010.

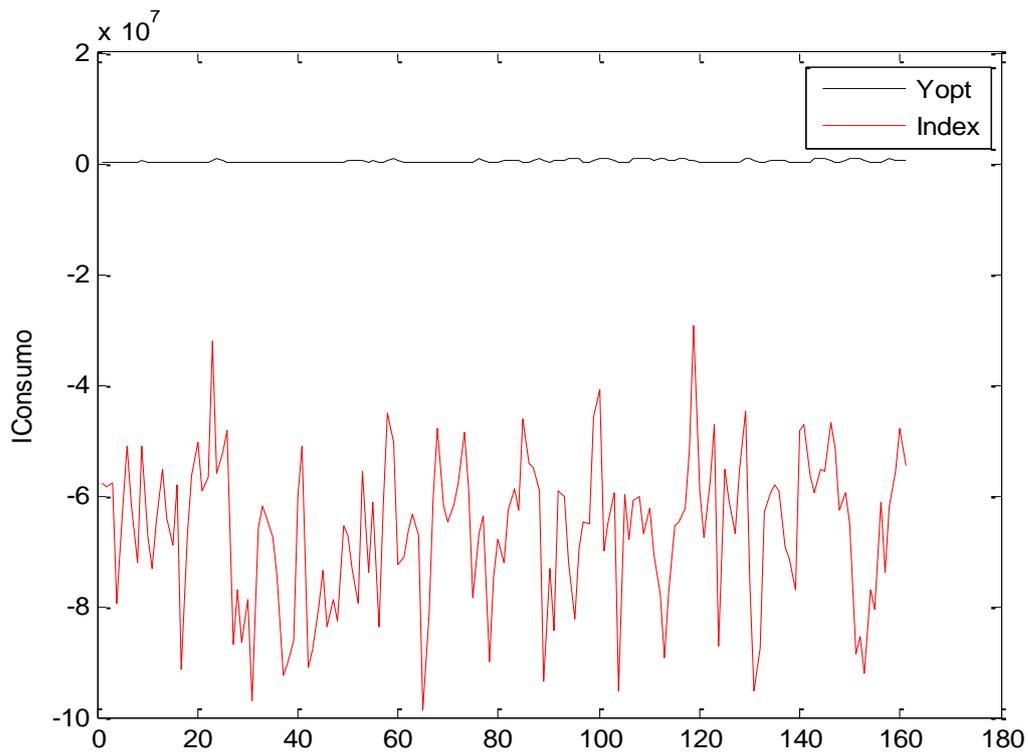


Figura C.2 – Comparativo do consumo real (Y_{opt}) com o consumo estimado ($Index$) por Regressão *Stepwise*, UC-3, Safra 2009/2010.

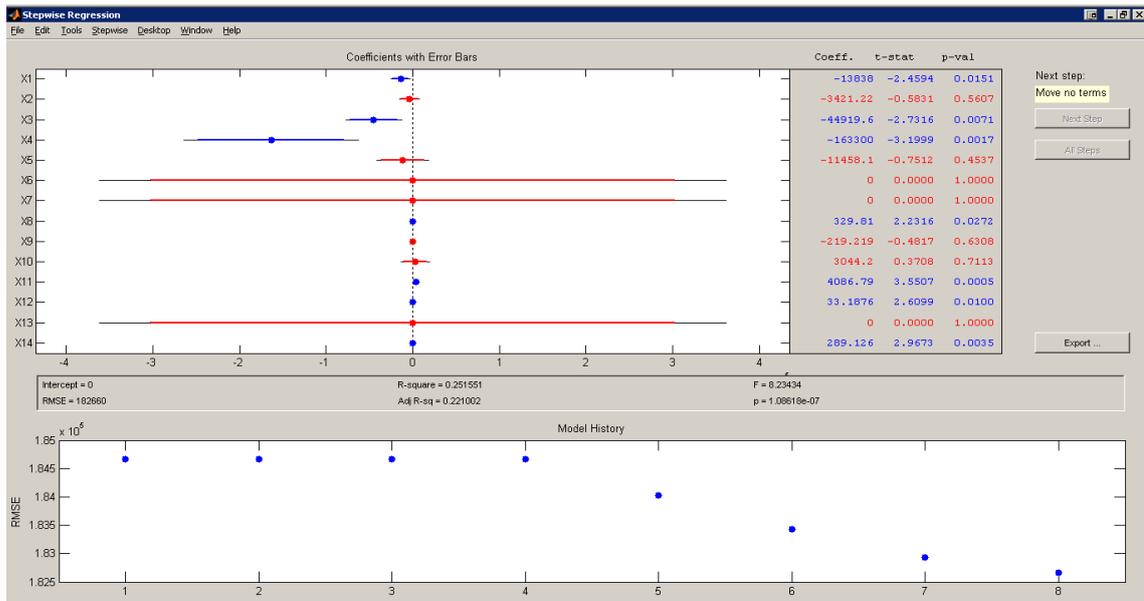


Figura C.3 – Interface gráfica da análise de Regressão *Stepwise*, UC-3, Safra 2010/2011.

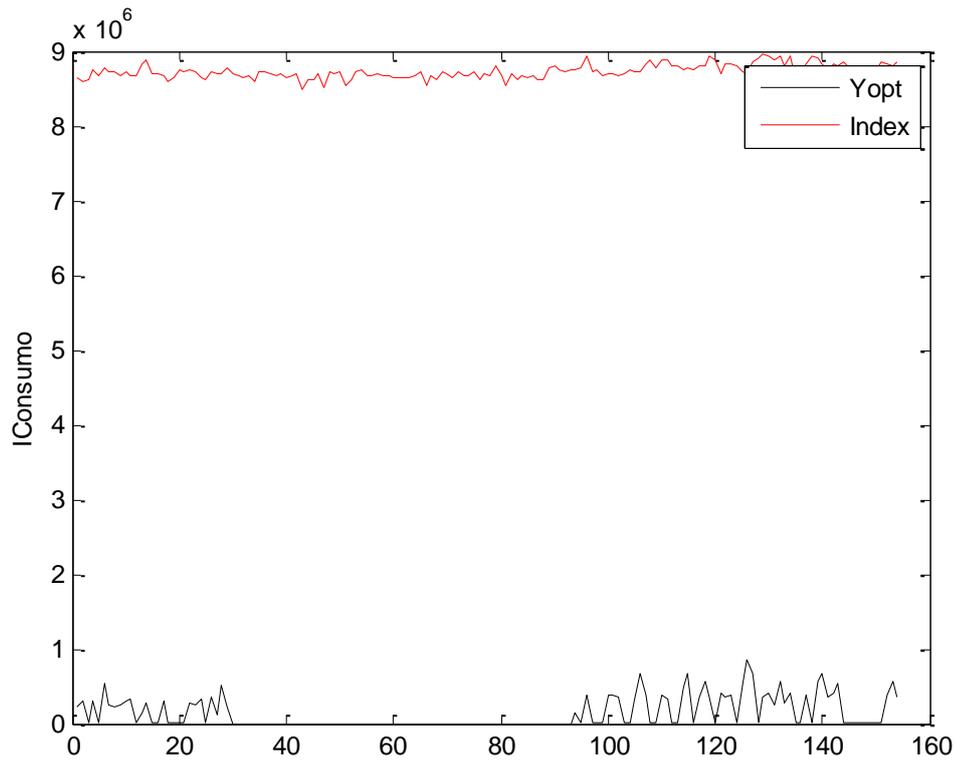


Figura C.4 – Comparativo do consumo real (Y_{opt}) com o consumo estimado ($Index$) por Regressão *Stepwise*, UC-3, Safra 2010/2011.

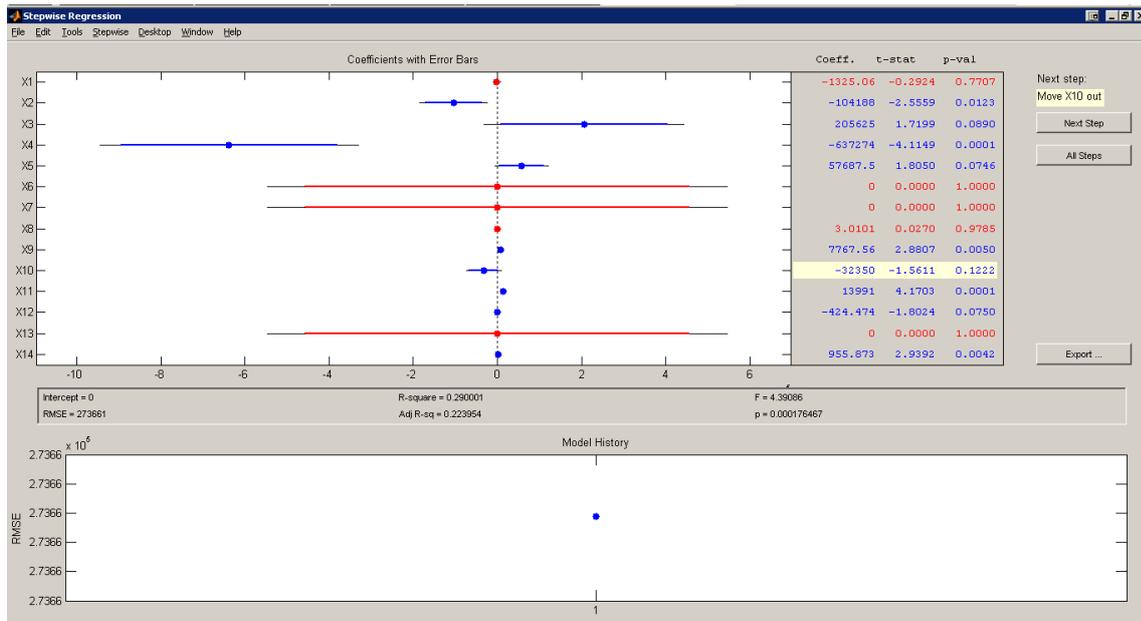


Figura C.5 – Interface gráfica da análise de Regressão *Stepwise*, UC-3, Safra 2011/2012.

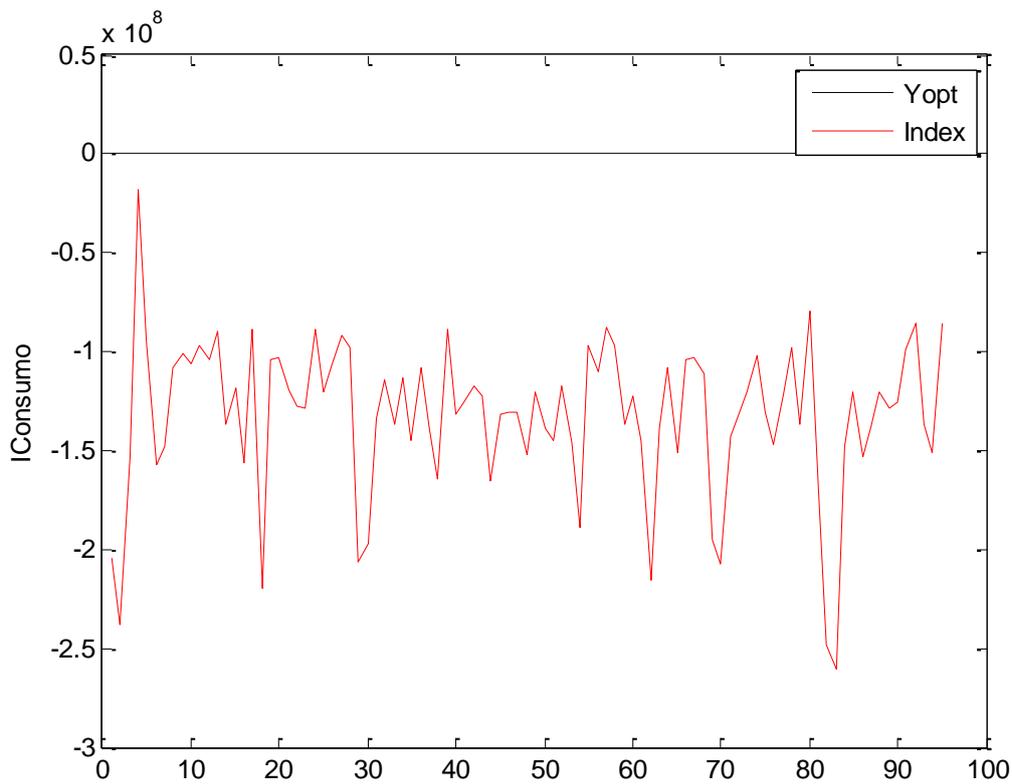


Figura C.6 – Comparativo do consumo real (Y_{opt}) com o consumo estimado ($Index$) por Regressão *Stepwise*, UC-3, Safra 2011/2012.

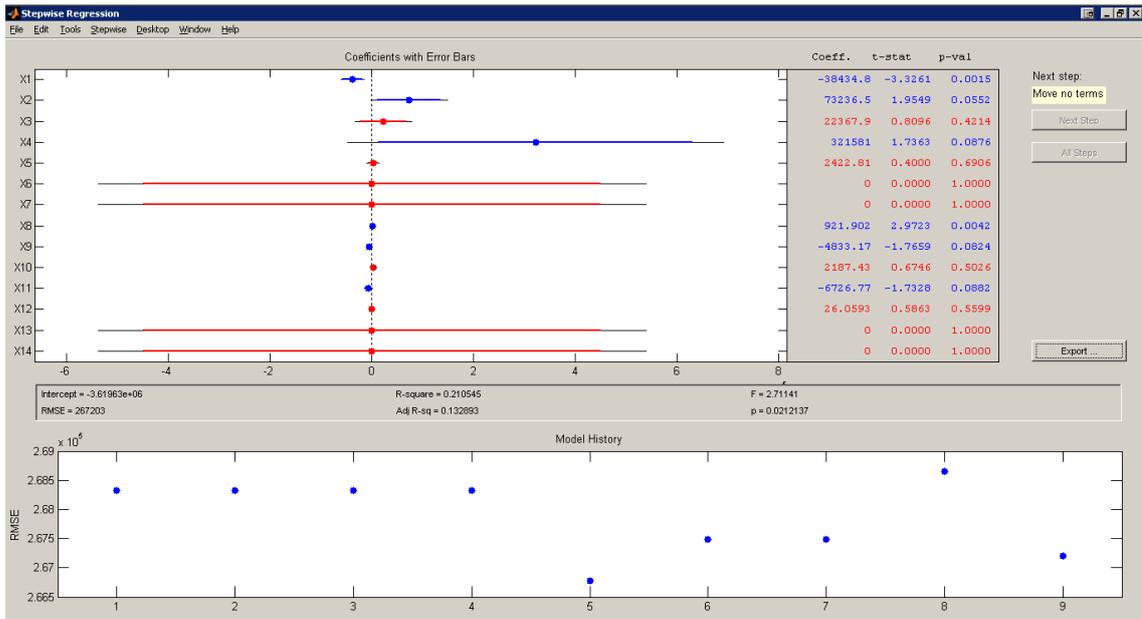


Figura C.7 – Interface gráfica da análise de Regressão Stepwise, UC-3, Safra 2012/2013.

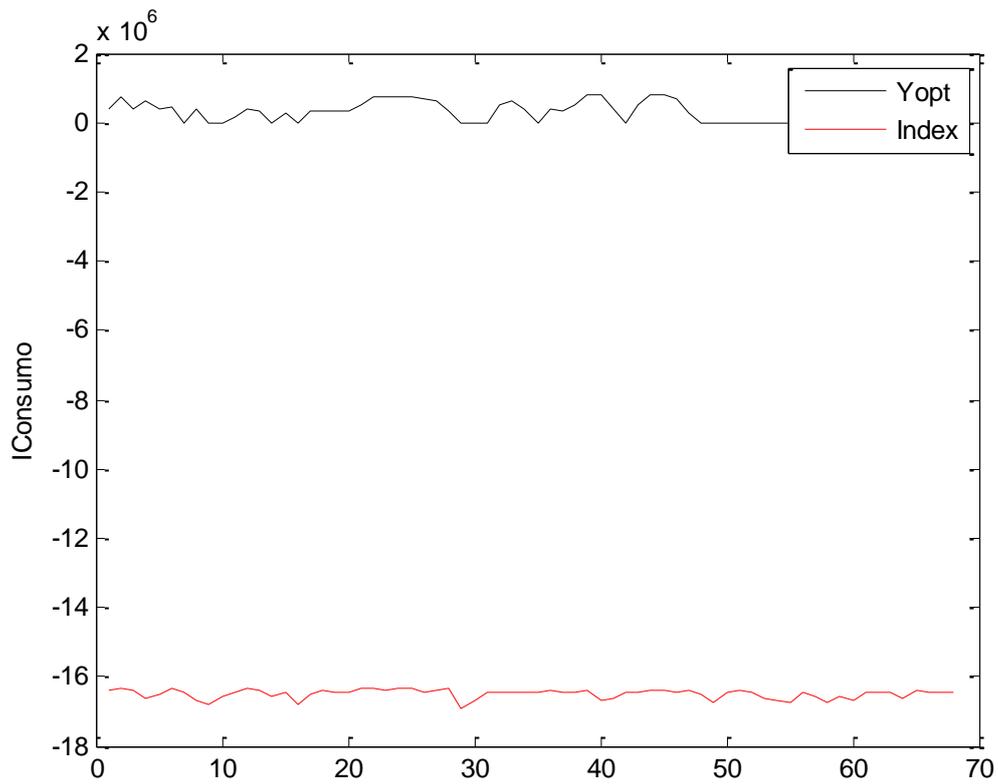


Figura C.8 – Comparativo do consumo real (Y_{opt}) com o consumo estimado ($Index$) por Regressão Stepwise, UC-3, Safra 2012/2013.

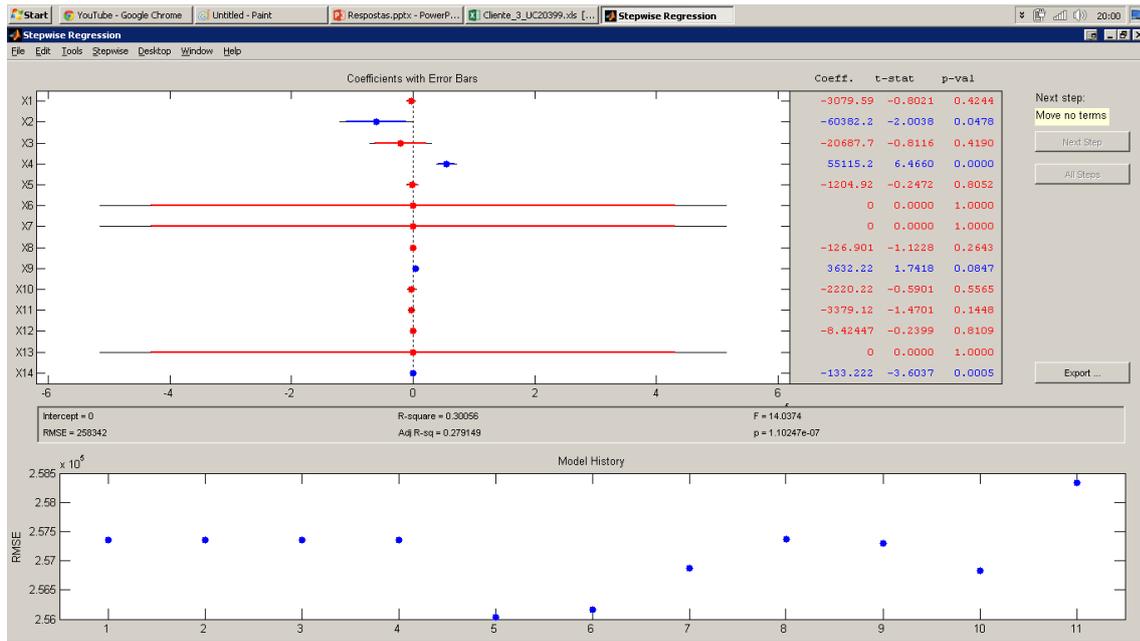


Figura C.9 – Interface gráfica da análise de Regressão *Stepwise*, UC-3, Safra 2013/2014.

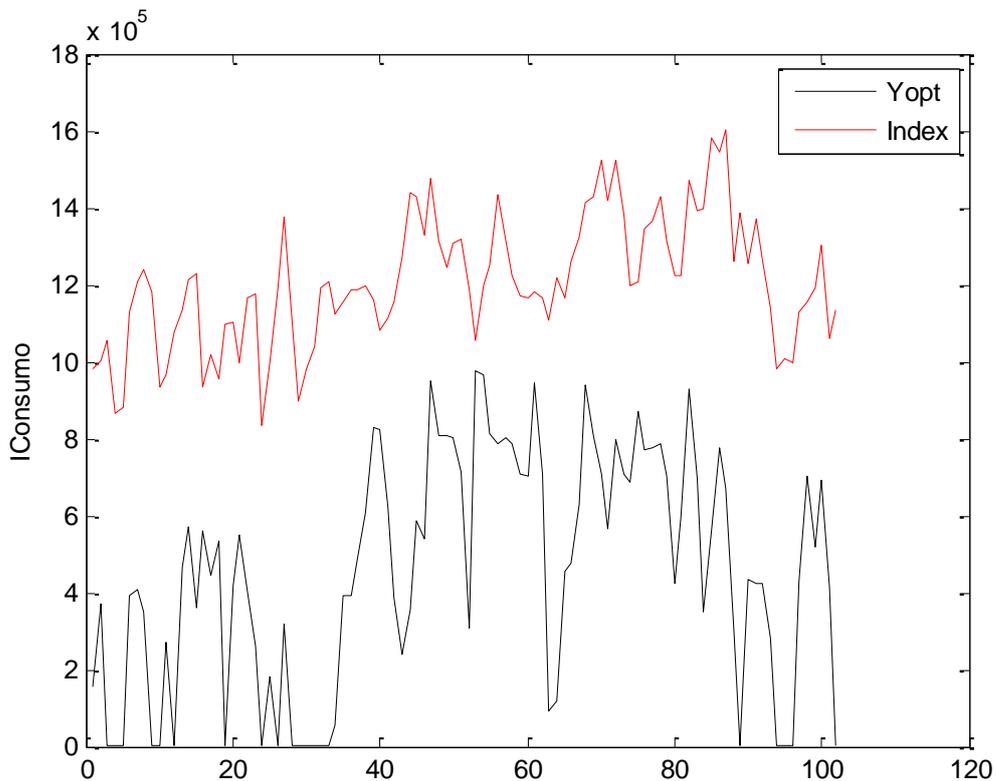


Figura C.10 – Comparativo do consumo real (Y_{opt}) com o consumo estimado ($Index$) por Regressão *Stepwise*, UC-3, Safra 2013/2014.