

UNIVERSIDADE FEDERAL DE ITAJUBÁ
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA
ELÉTRICA

Previsão de curto prazo de cargas elétricas usando combinação de
redes neurais artificiais e delineamento de experimentos

Franco Bassi Rocha

Itajubá
Dezembro de 2020

UNIVERSIDADE FEDERAL DE ITAJUBÁ
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA
ELÉTRICA

Franco Bassi Rocha

Previsão de curto prazo de cargas elétricas usando combinação de
redes neurais artificiais e delineamento de experimentos

Tese submetida ao Programa de Pós-Graduação em Engenharia Elétrica como parte dos requisitos para obtenção do título de Doutor em Ciências em Engenharia Elétrica.

Banca Examinadora:

Prof. Dr. Wesley Vieira da Silva

Prof. Dr. José Paulo Carvalho dos Santos

Prof. Dr. Paulo Fernando Ribeiro

Prof. Dr. Antonio Carlos Zambroni de Souza

Prof. Dr. Pedro Paulo Balestrassi (Orientador)

Prof. Dr. Anderson Paulo de Paiva (Coorientador)

Itajubá
Dezembro de 2020

*Dedico este trabalho especialmente ao meu pai Indalécio Rocha Júnior (in memoriam)
por ser o melhor professor que já tive na minha vida.*

AGRADECIMENTOS

Acredito que essa parte do trabalho seja a mais complexa de se escrever, pois não é pelo p-valor que atestamos a significância das pessoas em nossa vida.

Inicialmente agradeço a Deus, por me guiar principalmente nos momentos em que a visão ficou turva, não permitindo que eu desistisse em meio as diversas dificuldades.

Meu agradecimento mais profundo só poderia ser dedicado a duas pessoas: minha esposa Fernanda e minha filha Lara. À minha esposa por ter sido tão guerreira e suportado todo o peso da minha ausência nestes 4 anos. Assumi toda a responsabilidade, se doando a minha filha e a minha família. Seu amor, sua forma positiva e serena de encarar as dificuldades foram determinantes para que eu chegasse até aqui. À minha filha Lara, ainda pequenina não entende o significado dessas palavras, mas eu te agradeço filha pelo simples fato de existir na minha vida. Por fim, faço uma promessa a vocês duas: jamais sairei de perto de vocês novamente!

Agradeço a minha mãe Maria Olenca e ao meu pai Indalécio Rocha (*in memoriam*) por me ensinarem durante toda a vida a ser merecedor das coisas. O amor, a retidão e o caráter de vocês foram fundamentais para que este dia chegasse.

À minha sogra Elisa e meu sogro Paulo Fernando (*in memoriam*) por todo apoio dedicado a minha esposa e a minha filha neste período de ausência. Vocês foram fundamentais!

Ao meu irmão Bruno, que sempre estive de prontidão, me representando em situações de emergência e dando apoio a nossa mãe.

Ao professor, educador e orientador Pedro Paulo Balestrassi, por me acolher em um momento que eu estava totalmente desacreditado. Gratidão pela oportunidade de conviver e aprender contigo. Seu exemplo vai muito além de um trabalho de doutorado. Meu muito obrigado!

Ao também educador e coorientador Anderson Paulo de Paiva, pelos inúmeros conselhos e pelas inúmeras conversas quando eu estava vivendo momentos de desespero. Sua atuação foi fundamental e me proporcionou boas lições de como ser um docente de verdade.

Aos amigos do GEPE qualidade, que mesmo eu não fazendo parte do grupo, me acolheram com muito respeito e carinho. Fica aqui um agradecimento especial aos meus amigos Alexandre Fonseca Torres e Fabrício Alves de Almeida por terem sido tão importantes no processo de doutoramento.

Ao meu amigo e companheiro de Doutorado em Engenharia Elétrica Max Moreira pelas inúmeras reflexões, debates e pensamentos compartilhados.

Quando não estamos com as pessoas que amamos, nos resta a recorrer aos nossos amigos. E aqui registro meu agradecimento ao meu irmão de república César Veras Casañas, pela força que me deu nos momentos em que eu pensei em desistir de tudo. Valeu irmão!

Aos professores do departamento de matemática da Universidade Federal de Alfenas, em especial aos professores José Paulo, José Carlos e Evandro por me incentivarem nos momentos difíceis.

À Universidade Federal de Alfenas, pela concessão do afastamento para qualificação.

Ao coordenador do programa de pós graduação em engenharia elétrica da Unifei, professor Tales Cleber Pimenta pela paciência e prontidão nos vários questionamentos.

“Se não puder voar, corra. Se não puder correr, ande. Se não puder andar, rasteje, mas continue em frente de qualquer jeito.”
(Martin Luther King)

RESUMO

É proposto neste trabalho uma nova abordagem para previsão de demanda de energia elétrica de curto prazo para subestações de distribuição de energia utilizando o ensemble de redes neurais artificiais. Nesse sentido, o principal objetivo desta abordagem é fazer previsões de uma mesma série temporal através de diferentes ferramentas que sozinhas são competentes para este tipo de problema e em seguida combinar as soluções, obtendo uma solução melhor em comparação a essas ferramentas utilizadas de forma individual. Para a construção do ensemble a metodologia de planejamento experimental (DOE) foi empregada inicialmente para identificar a influência de 6 fatores relacionados a parametrização da RNA e a partir do método de otimização desirability obter uma parametrização para determinar a arquitetura das redes neurais que formaram o ensemble. Na sequência, o método de otimização interseção normal a fronteira (NBI) aliado com a técnica de análise fatorial exploratória (baseado em planejamento com experimentos de mistura) foi utilizado para estabelecer um conjunto de soluções Pareto ótimas para a combinação das saídas produzidas pelas redes neurais, formando a saída do ensemble. Como critério de escolha, a razão máxima entre entropia de Shannon e erro percentual global foi utilizada e com base nos 72 dados deixados fora da amostra o ensemble de redes neurais artificiais apresentou melhores resultados comparados a cada rede individual.

Palavras-chave: Previsão de carga elétrica de curto prazo; Ensemble de Redes Neurais Artificiais; Planejamento de experimentos; Interseção normal a fronteira; Análise fatorial.

ABSTRACT

This work proposes a new approach to short-term load forecast for power distribution substations using the artificial neural network ensemble. In this sense, the main objective of this approach is to make predictions for the same time series using different tools that are competent for this type of problem and then combine the solutions, obtaining a better solution compared to those tools used individually. For the construction of the ensemble, the experimental planning methodology (DOE) was used initially to identify the influence of 6 factors related the parameterization of the ANN and from the desirability optimization method to obtain a parameterization to determine the architecture of the neural networks that formed the ensemble . Then, the normal intersection optimization method (NBI) combined with the technique of exploratory factor analysis (based on mixture design of experiments) was used to establish a set of optimal Pareto solutions for combining the outputs produced by neural networks, forming the output of the ensemble. As a criterion of choice, the maximum ratio between Shannon's entropy and global percentage error was used and based on the 72 out of sample data the ensemble of artificial neural networks presented better results compared to each individual network.

Keywords: Short-term load forecasting; Artificial neural network ensemble; Design of experiments; Normal boundary intersection; Factor analysis.

LISTA DE ILUSTRAÇÕES

Figura 1.1 – Classificação da pesquisa científica	3
Figura 2.1 – Classificação e aplicação da previsão de carga elétrica	5
Figura 2.2 – Neurônio biológico	7
Figura 2.3 – Estrutura de uma RNA	8
Figura 2.4 – Funções de ativação mais utilizadas em uma RNA	8
Figura 2.5 – Ensemble de redes neurais artificiais	10
Figura 2.6 – Fatores de um processo	12
Figura 2.7 – Passos para a realização de um experimento	16
Figura 2.8 – Superfície de resposta para y	18
Figura 2.9 – Região experimental para experimentos de misturas com duas (a) e três (b) componentes	20
Figura 2.10–Arranjos simplex lattice (a) e centróide (b)	21
Figura 2.11–Arranjos simplex incorporado	22
Figura 2.12–Região experimental definida por restrições limitadas inferiormente	23
Figura 2.13–Região experimental definida por restrições limitadas superiormente: (a) simplex invertido; (b) não simplex	24
Figura 2.14–Otimidade e dominância de Pareto	26
Figura 2.15–Pontos de ancoragem, utopia, nadir e pseudo nadir para o caso bi-objetivo	27
Figura 2.16–Definição do alvo para o desirability	29
Figura 2.17–Formato da função de maximização para o desirability	30
Figura 2.18–Relação de importância entre os alvos e os limites	31
Figura 2.19–Etapas para execução do método NBI	32
Figura 2.20–Árvore de pesos para um problema objetivo com três funções objetivos e espaçamento de 0,2 entre os pesos.	34
Figura 2.21–Arranjo simplex lattice {3,5}.	34
Figura 2.22–Representação gráfica do NBI para o caso bi-objetivo	35
Figura 2.23–Entropia	37
Figura 3.1 – Etapas para a obtenção da parametrização das redes neurais	41
Figura 3.2 – Séries temporais de energia elétrica das 4 subestações	43
Figura 3.3 – Análise espectral	47
Figura 3.4 – Gráfico de Pareto para a resposta Média MAPE	51
Figura 3.5 – Gráfico de Probabilidade para os resíduos para o modelo Média MAPE	53
Figura 3.6 – Gráfico de efeitos principais para a resposta Média MAPE	53
Figura 3.7 – Gráfico de interações para a resposta Média MAPE	54
Figura 3.8 – Gráfico de Pareto para a resposta DP MAPE	56

Figura 3.9 – Gráfico de Probabilidade para os resíduos para o modelo DP MAPE	56
Figura 3.10–Gráfico de efeitos principais para a resposta DP MAPE	57
Figura 3.11–Gráfico de interações para a resposta DP MAPE	57
Figura 3.12–Parametrização pelo método desirability	59
Figura 3.13–Arquitetura das redes neurais usadas no ensemble	60
Figura 4.1 – Metodologia para combinação das saídas das redes neurais	61
Figura 4.2 – Dados utilizados no ajuste e validação dos modelos	62
Figura 4.3 – Arranjo experimental simplex lattice $\{3, 5\}$	63
Figura 4.4 – Gráfico de contorno para y_1, y_2, y_3 e y_4	67
Figura 4.5 – Gráfico de probabilidade de resíduos para y_1 e y_2	67
Figura 4.6 – Gráfico de probabilidade de resíduos para y_3 e y_4	68
Figura 4.7 – <i>Scree Plot</i>	69
Figura 4.8 – Gráfico de carregamentos das quatro respostas	70
Figura 4.9 – Dendograma de similaridade entre as quatro resposta e seus respectivos fatores.	70
Figura 4.10–Superfície de respostas para os fatores F_1 e F_2	72
Figura 4.11–Gráfico de contorno para os fatores F_1 e F_2	72
Figura 4.12–Gráfico de traços para F_1	74
Figura 4.13–Gráfico de traços para F_2	74
Figura 4.14–Relação entre os fatores e as respostas	76
Figura 4.15–Relação entre os pesos do método NBI com os pesos relacionados as RNAs	77
Figura 4.16–Relação entre os pesos do método NBI com os fatores	78
Figura 4.17–Entropia, EPG e H/EPG em cada solução Pareto-ótima	78
Figura 4.18–Fronteira de Pareto $F_1 \times F_2$ nos espaços de solução normalizado e original	79
Figura 4.19–Variação das respostas entre cada ponto da Fronteira de Pareto	80
Figura 4.20–Nuvem de pontos com soluções viáveis e Fronteira de Pareto	80
Figura 4.21–Gráfico de contorno sobreposto	81
Figura 4.22–Previsão de demanda de energia elétrica para as subestações	82

LISTA DE TABELAS

Tabela 2.1 – Arranjo fatorial 2^4	13
Tabela 2.2 – Arranjo Fatorial Fracionado 2^{4-1}	15
Tabela 3.1 – Entradas das RNAs	43
Tabela 3.2 – Arranjo experimental 2_{IV}^{6-2}	49
Tabela 3.3 – MAPE subestações 1 e 2	49
Tabela 3.4 – MAPE subestações 3 e 4	50
Tabela 3.5 – Respostas de interesse em cada experimento	50
Tabela 3.6 – ANOVA para a resposta Média MAPE	52
Tabela 3.7 – ANOVA para a resposta DP MAPE	55
Tabela 3.8 – Valores utilizados no algoritmo desirability	58
Tabela 4.1 – Séries de resíduos produzidas pelas redes neurais para as subestações 1 e 2	64
Tabela 4.2 – Séries de resíduos produzidas pelas redes neurais para as subestações 3 e 4	65
Tabela 4.3 – Métricas de performance das redes neurais para as subestações	65
Tabela 4.4 – MAPE das séries de resíduos combinados	66
Tabela 4.5 – Estrutura de correlação entre as respostas y_1, y_2, y_3 e y_4	68
Tabela 4.6 – Componentes principais	69
Tabela 4.7 – Análise fatorial baseada em componentes principais e rotação <i>varimax</i>	69
Tabela 4.8 – Correlações entre as respostas e os fatores	71
Tabela 4.9 – Série de escores dos fatores	71
Tabela 4.10–ANOVA para F_1	73
Tabela 4.11–ANOVA para F_2	73
Tabela 4.12–Resultados das otimizações do método NBI	76
Tabela 4.13–Escolha do solução ótima segundo critério da máxima razão $\frac{H}{EPG}$	79
Tabela 4.14–MAPE obtidos com base nos dados fora da amostra para as subestações 1 e 2	81
Tabela 4.15–MAPE obtidos com base nos dados fora da amostra para as subestações 3 e 4	82

LISTA DE QUADROS

Quadro 3.1 – Projetos experimentais e resoluções disponíveis no software Minitab .	48
--	----

LISTA DE ABREVIATURAS E SIGLAS

RNA	Redes Neurais Artificiais
DOE	Design of Experiments - Planejamento de experimentos
NBI	Normal Boundary Intersection - Interseção Normal a Fronteira
PSO	Particle Swarm Optimization - Otimização por enxame de partículas
MLP	Multi Layer Perceptron - Perceptron multicamada
RBF	Radial Basis Function - Função de base radial
BEM	Basic Ensemble Method - Método do ensemble básico
MDE	Mixture Design of Experiments - Projeto de Experimentos de Misturas
RSM	Response Surface Methodology - Metodologia de superfície de resposta
OLS	Ordinary Least Squares - Mínimos quadrados ordinários
CHIM	Convex hull of individual minima - Envoltória convexa de mínimos individuais
EPG	Erro percentual global
IPS	Intelligent Problem Solver - Solucionador inteligente de problemas
PNN	Rede neural probabilística
MAPE	Erro percentual absoluto médio
PD	Pré-processamento de dados
LI	Transformação Linear
EV	Variáveis explicativas
CS	Critério de seleção das redes a serem retidas
IV	Seleção de subconjunto de variáveis independentes
SS	Seleção de subconjuntos para cada rede treinada
R	Número de etapas de tempo usadas como entradas
DP	Desvio Padrão

SUMÁRIO

Dedicatória	iii
Agradecimentos	v
Epígrafe	vi
Resumo	vii
Abstract	viii
Lista Ilustrações	viii
Lista de Tabelas	x
Lista de Quadros	xi
Lista de Abreviaturas e Siglas	xiii
1 INTRODUÇÃO	1
1.1 Justificativa e Relevância	1
1.2 Objetivos	1
1.2.1 Objetivo Geral	1
1.2.2 Objetivos Específicos	2
1.3 Método de Pesquisa	2
1.4 Organização do Texto	3
2 FUNDAMENTAÇÃO TEÓRICA	5
2.1 Previsão de carga elétrica	5
2.2 Redes Neurais artificiais (RNAs) para previsão de carga elétrica	6
2.3 Ensemble de redes neurais artificiais	9
2.4 Planejamento e análise de experimentos (DOE)	10
2.4.1 Arranjos fatoriais	12
2.4.2 Arranjos Fatoriais Fracionados	14
2.4.3 Resolução de um experimento	14
2.4.4 Etapas para a execução do experimento	15
2.4.5 Planejamento de experimentos em simulação	16
2.4.6 Metodologia de superfície de resposta	18
2.4.7 Experimentos por arranjos de misturas	19

2.5	Otimização Multiobjetivo	25
2.5.1	Formulação matemática de problemas de otimização multiobjetivo	25
2.5.2	Somas Ponderadas	27
2.5.3	Método de otimização Desirability	28
2.5.4	Método de interseção normal a fronteira	31
2.6	Erro percentual global e entropia	36
2.7	Análise Fatorial	38
2.8	Considerações finais sobre o capítulo	40
3	PARAMETRIZAÇÃO DAS REDES NEURAIIS	41
3.1	Estratégia utilizada	41
3.2	Séries temporais de energia elétrica das subestações	42
3.2.1	Planejamento Pré-Experimental	44
3.2.2	Definição do arranjo experimental e execução dos experimentos	47
3.2.3	Análise dos resultados experimentais para a resposta média MAPE	50
3.2.3.1	Análise dos resultados experimentais para a resposta DP MAPE	54
3.3	Parametrização de redes usando o método desirability	58
3.4	Considerações finais sobre o capítulo	60
4	PREVISÃO DE CARGA ELÉTRICA USANDO O ENSEMBLE DE REDES NEURAIIS ARTIFICIAIS	61
4.1	Ponderação utilizada no ensemble de redes neurais	61
4.2	Resultados e discussão	64
5	CONCLUSÕES	84
	REFERÊNCIAS	86

1 INTRODUÇÃO

1.1 Justificativa e Relevância

Em sistemas de energia, a previsão de carga elétrica de curto prazo é importante para mitigar os desequilíbrios do sistema e reduzir os custos de operação por meio do planejamento [1]. As séries temporais de carga elétrica têm particularidades que tornam suas previsões desafiadoras, como a não linearidade de seus eventos e para a previsão deste tipo de série, técnicas baseadas em inteligência artificial (especialmente redes neurais artificiais-RNAs) têm sido amplamente utilizadas [2, 3]. Devido à necessidade de previsões cada vez mais precisas, uma combinação de previsões usando mais de uma rede neural (denominada de ensemble) é explorada neste trabalho apresentando performance superior quando comparadas às previsões obtidas pelas redes neurais individuais. No entanto, ao se combinar modelos de previsão utilizando redes neurais alguns cuidados devem ser tomados em dois aspectos: o primeiro diz respeito à parametrização da RNA, dificuldade frequentemente encontrada pelos pesquisadores e normalmente abordada por meio de ajustes manuais, ou tentativa e erro [4, 5]. O outro aspecto está relacionado à forma como as saídas produzidas pelas redes neurais serão combinadas para formar a saída do ensemble. Nesse sentido, a técnica de planejamento de experimentos (DOE - Design of Experiments) auxilia na resposta a esses questionamentos, pois através dessa ferramenta poderão ser detectados se há parâmetros e interações entre esses parâmetros que afetam o desempenho da rede e também como é a participação de cada saída da rede na formação da saída do ensemble. Esses dois aspectos serão discutidos para que o modelo de ensemble para previsão de demanda de energia elétrica seja elaborado.

1.2 Objetivos

De acordo com as necessidades e relevâncias expostas nos itens anteriores, esta pesquisa foi elaborada com o intuito de cumprir com o objetivo geral e específicos a seguir.

1.2.1 Objetivo Geral

Este trabalho propõe o ensemble de RNA's para realizar previsão de carga elétrica de curto prazo de quatro subestações de distribuição de energia elétrica.

1.2.2 Objetivos Específicos

Diante do objeto de estudo escolhido os objetivos específicos deste trabalho incluem:

- Estudo de como a parametrização da rede impacta na sua medida desempenho através do planejamento de experimentos;
- Baseado em um problema de otimização multiobjetivo, determinar a parametrização das redes;
- Através do planejamento de misturas, otimização multiobjetivo e análise fatorial exploratória, determinar um conjunto de soluções para determinar a saída do ensemble de redes neurais;
- Comparar os resultados do ensemble com cada uma das RNA's que o compõem.

1.3 Método de Pesquisa

O presente trabalho segue a linha de pesquisa quantitativa, pois usa métodos estatísticos fazendo a análise de dados de maneira quantitativa utilizando as respostas dos diferentes cenários gerando, na fase final, um ponto ótimo das variáveis de entrada para a variável de saída, que nesse trabalho, será a medida de desempenho do ensemble de redes neurais.

A classificação quantitativa é dada as pesquisas onde modelos de relações causais entre variáveis de controle e variáveis de saída são desenvolvidas, analisadas e alteradas sobre um domínio específico [6]. Por isso, este trabalho é classificado como uma pesquisa baseada em modelagem quantitativa, uma vez que se utiliza de métodos estatísticos para realizar as otimizações.

Para uma visualização geral da classificação da pesquisa científica tem-se a figura 1.1 que, de acordo com as definições de Turrioni e Mello [7], classifica a pesquisa científica segundo alguns critérios, como natureza, objetivos, abordagem e métodos utilizados.

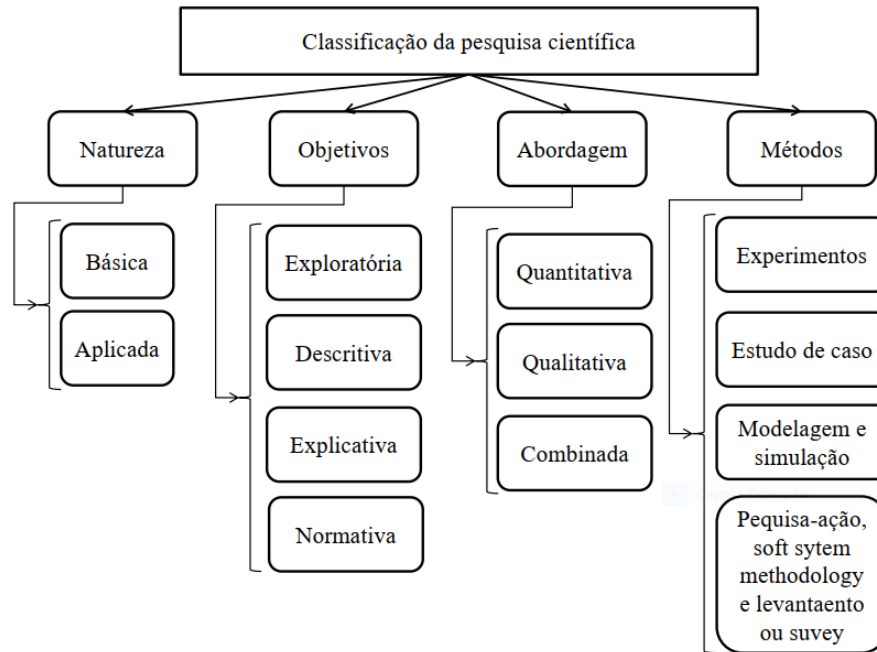


Figura 1.1 – Classificação da pesquisa científica

Fonte: adaptado de [7]

Segundo a figura acima, este trabalho está classificado da seguinte forma:

- Quanto à natureza: tem característica **aplicada**, com interesse prático na utilização em problemas reais.
- Quanto aos objetivos: é caracterizada como **explicativa**, pois visa explicar, analisar e avaliar fatores que contribuem para a ocorrência de fenômenos.
- Quanto à abordagem do problema: classificada como **quantitativa**, traduzindo em números os resultados da pesquisa.
- Quanto ao método de pesquisa: **experimental** pois ele é empregado quando se determina um objeto de estudo (previsão de carga elétrica através do ensemble de redes neurais). Neste método, selecionam-se as variáveis que seriam capazes de influenciá-los, definem-se as formas de controle e a observação dos efeitos que a variável produz no objeto de estudo.

1.4 Organização do Texto

Este documento está estruturado da seguinte forma:

- **Capítulo 2:** é apresentada a fundamentação teórica e revisão de literatura.
- **Capítulo 3:** destaca os fatores e níveis escolhidos para definição do arranjo experimental, bem como a parametrização definida via otimização multiobjetivo.

- **Capítulo 4:** apresenta o planejamento experimental de mistura para obtenção de um conjunto de soluções (via método de otimização) para ponderação das saídas produzidas pelas redes neurais artificiais para formar a saída do ensemble.
- **Capítulo 5:** traz as conclusões.

2 FUNDAMENTAÇÃO TEÓRICA

Neste capítulo serão apresentados as ferramentas utilizadas no trabalho, evidenciando desde os conceitos básicos sobre previsões de demanda de energia elétrica, ensemble de redes neurais artificiais, encerrando com as ferramentas de planejamento de experimentos e otimização.

2.1 Previsão de carga elétrica

A previsão de carga de elétrica tornou-se cada vez mais importante em sistemas elétricos. É relevante no planejamento e operação de sistemas de energia, infraestrutura, expansão da rede de transmissão de energia, realocação de carga e programação de manutenção, entre outras aplicações [8, 9, 10, 11]. Segundo Hong e Wang [12], a previsão de carga elétrica é classificada em quatro classes diferentes, também chamadas de horizontes de previsão, conforme figura 2.1.

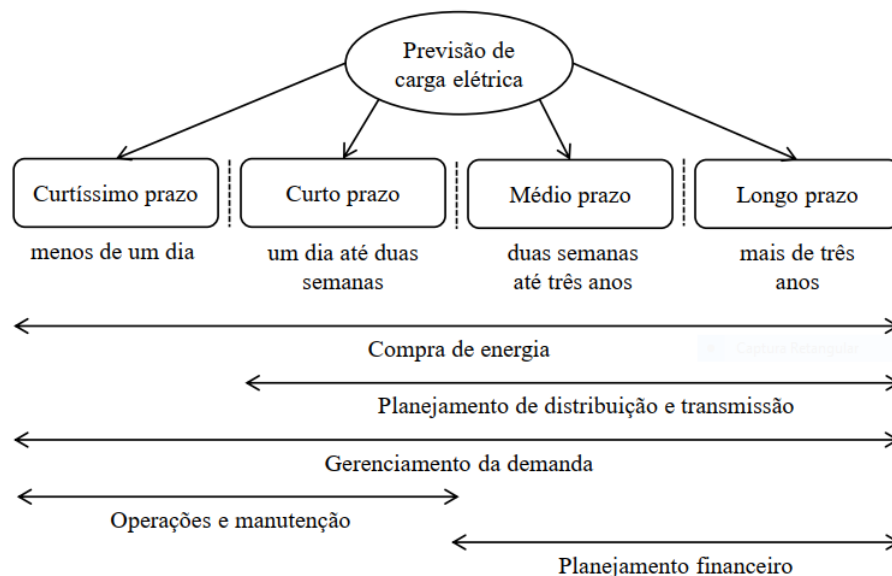


Figura 2.1 – Classificação e aplicação da previsão de carga elétrica
Fonte: adaptado de [13].

De acordo com a figura 2.1, A previsão de carga de elétrica é subdividida em quatro classes: curtíssimo prazo para previsões de menos de um dia, curto prazo (escolhido neste trabalho) para previsão de um dia a duas semanas, médio prazo para previsões no horizonte de duas semanas a três anos e a longo prazo, para previsões além de três anos. Todas essas subdivisões são importantes para o planejamento de compras de eletricidade e gerenciamento de demanda. A previsão de curto prazo busca estimar a carga que

será demandada, auxiliando no planejamento da operação, transferência de energia e gerenciamento da demanda. As previsões de curto, médio e longo prazo são importantes para o planejamento do sistema de transmissão e distribuição, e o planejamento financeiro de um sistema elétrico pode se beneficiar de informações de previsões de médio e longo prazo.

A previsão de energia elétrica tem sido um tópico de interesse significativo na comunidade científica, e várias abordagens foram usadas na literatura. Métodos analíticos como média móvel integrada autoregressiva (ARIMA) [14, 15], filtro de Kalman [16], decomposição em valores singulares [17, 18] e suavização exponencial [19] foram explorados para esse tipo de previsão. Verdejo et al. [20] revisaram os principais métodos paramétricos lineares estatísticos e os implementaram para analisar os dados de medidas reais em sistemas chilenos. Devido à presença de fenômenos não lineares no processo de previsão de carga estocástica, as técnicas baseadas em inteligência artificial foram aplicadas a problemas de previsão de carga elétrica de curto prazo e apresentaram desempenho superior aos métodos de previsão convencionais.

Chen et al. [21] recomendaram um modelo de máquina de vetores de suporte em uma competição organizada pela EUNITE para prever uma carga elétrica de médio prazo (prevendo uma carga diária máxima para os 31 dias seguintes). Eles concluíram que as informações de temperatura (ou outros tipos de informações meteorológicas) podem não ser úteis em um problema de previsão de carga a médio prazo, e que a introdução de um conceito de série temporal pode melhorar a previsão. Niu et al. [22] criaram um sistema para prever a carga de energia usando uma máquina de vetores de suporte e otimização de colônias de formigas. Eles o compararam com uma rede neural de retropropagação e uma única máquina de vetores de suporte para previsão de curto prazo demonstrando desempenho aprimorado. Yang et al. [23] desenvolveram um modelo para prever carga elétrica combinando a função de autocorrelação e uma máquina de vetores de suporte de mínimos quadrados (LSSVM). Os resultados mostraram que o método é eficaz para melhorar significativamente o desempenho das previsões em comparação com os modelos de referência. Métodos preditivos que usam transformada Wavelet [24, 25], bem como modelos que usam lógica fuzzy [26, 27, 28] também são encontrados na literatura. As redes neurais serão abordadas de forma separada na próxima seção por ser o método adotado neste trabalho.

2.2 Redes Neurais artificiais (RNAs) para previsão de carga elétrica

As redes neurais artificiais são sistemas paralelos distribuídos compostos por unidades de processamento, chamados de neurônios artificiais. As RNAs buscam implementar modelos matemáticos que assemelham-se aos neurônios biológicos (veja figura 2.2).

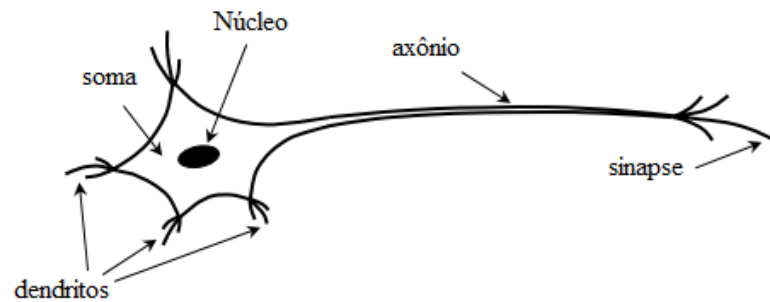


Figura 2.2 – Neurônio biológico
Fonte: [29].

O neurônio biológico é uma célula estimulada por uma ou mais entradas, gerando uma saída que é encaminhada a outro neurônio. A saída depende da força de cada uma das entradas e da magnitude da conexão associada a entrada (sinapse). O funcionamento das RNAs é inspirado em uma simplificação do sistema neural biológico, com capacidade de aprendizado, generalização, associação e abstração.

A idéia de redes neurais como máquinas computacionais teve origem pelo trabalho de McCulloch e Pitts [30], em que um neurônio com pesos fixos foi proposto. As redes neurais, pioneiras no campo da ciência e da engenharia cognitiva, são aproximações de funções universais e altamente flexíveis, podendo capturar estruturas de dados não lineares sem conhecimento prévio de um problema específico [31]. Assim, quando aplicada a uma série temporal caracterizada por relações não lineares, a rede neural irá detectá-las e fornecer um ajuste superior em comparação aos modelos lineares.

Uma rede neural é classificada de acordo com o padrão de conexão entre os neurônios (também chamado arquitetura), os métodos de pesagem de conexão (chamados treinamento, aprendizado ou algoritmo) e as funções de ativação [32]. Existem muitos tipos de arquitetura de rede neural, incluindo rede feedforward, rede recorrente, base radial, etc. Como mostrado na figura 2.3, em uma rede neural artificial as informações recebidas das camadas de entrada passam sequencialmente pelas camadas da rede até atingir a camada de saída. As camadas intermediárias que estão entre as camadas de entrada e saída são chamadas de camadas ocultas.

Em um neurônio, a saída é calculada por uma função da combinação das suas entradas, ou seja, para cada neurônio na camada de saída oculta, a seguinte transformação de entrada e saída é usada:

$$y = f \left(\sum_{j=1}^K w_j u_j + w_o \right) \quad (2.1)$$

em que y é a saída, K o número total de neurônios na camada anterior, u_j é a saída do j -ésimo neurônio na camada anterior, w_j são os pesos de conexão correspondentes e w_o é o bias. As funções de ativação mais utilizadas estão na figura 2.4.

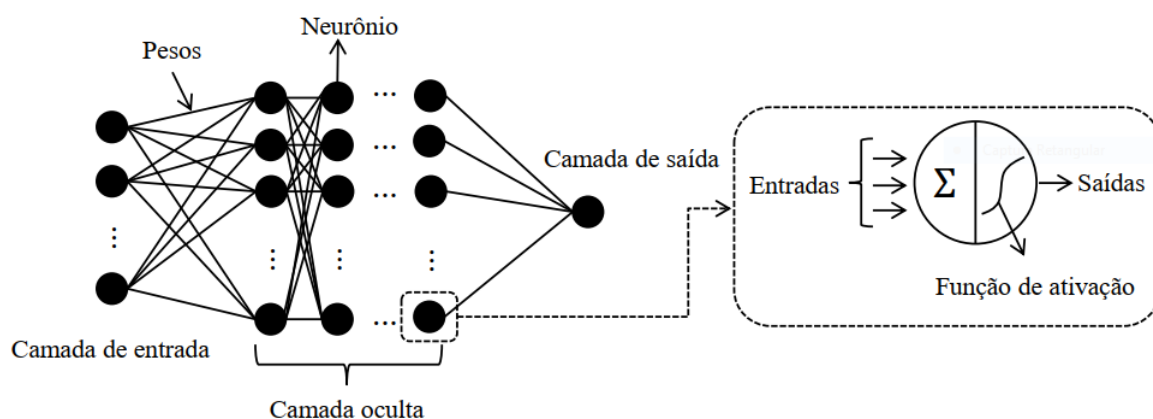


Figura 2.3 – Estrutura de uma RNA

Fonte: adaptado de [4].

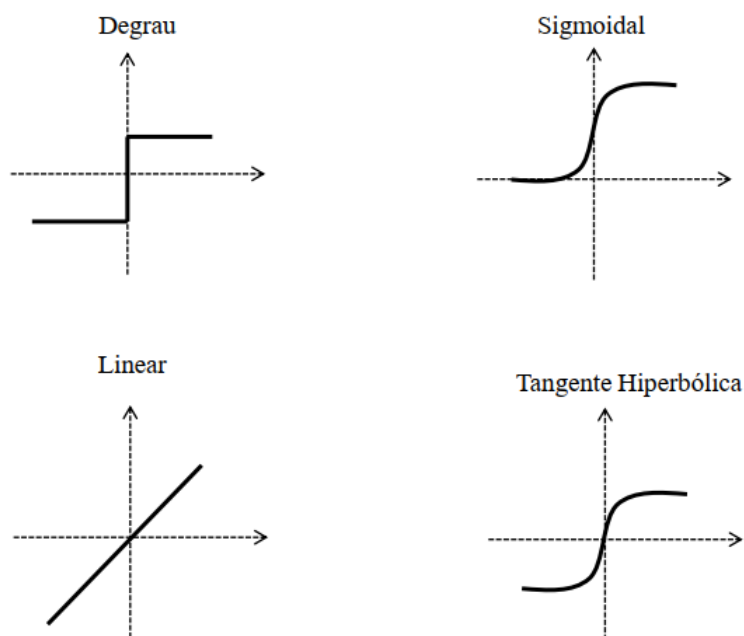


Figura 2.4 – Funções de ativação mais utilizadas em uma RNA

Fonte: autor.

Em uma pesquisa recente na base de dados do Web of Science disponível no portal periódico da CAPES, uma busca feita combinando a palavra chave "short-term load forecasting" com "neural network" encontrou 129 referências no período de 2000 até 2019. Alguns estudos importantes consideraram as redes neurais como uma metodologia significativa em problemas de previsão de carga elétrica. Muralitharan et al. [33] sugeriram uma nova abordagem de otimização baseada em rede neural para prever demandas de energia. Neste trabalho, uma RNA baseada em um algoritmo genético e otimização de enxame de partículas (PSO) (onde os pesos da rede são ajustados automaticamente) foi usada para determinar a previsão de demanda de energia para o consumidor final. Com base em simulações, o método proposto obteve um desempenho melhor do que uma rede neural

convencional. Redes neurais desenvolvidas por algoritmo genético para previsões de carga de curto prazo foram apresentadas em [34] usando dados de carga real e previsões meteorológicas, revelando boa precisão. Kong et al. [35] propuseram uma estrutura baseada em redes neurais recorrentes com memória longa para previsão de carga de curto prazo para famílias. Dudek [36] apresentou e comparou várias abordagens univariadas para previsão de carga elétrica de curto prazo usando várias arquiteturas de redes neurais, alcançando desempenho superior em comparação com métodos de previsão populares, como ARIMA e suavização exponencial. Uma previsão aprimorada de curto prazo baseada em um ensemble de redes neurais foi relatada em [37], reduzindo erros de estimativa em comparação com várias técnicas existentes, enquanto Quan et al. [38] implementaram um método baseado em rede neural para construir intervalos de previsão.

Recentemente, foi dada ênfase ao desenvolvimento de métodos híbridos combinando várias técnicas de inteligência computacional. As redes neurais já foram combinadas com wavelets [39, 40, 41], lógica fuzzy [42, 43, 44] e teoria do caos [45, 46]. Redes neurais de diferentes tipos (perceptron multicamada (MLP), função de base radial (RBF), redes recorrentes) também foram combinadas com vários algoritmos de inspiração biológica, com o objetivo de otimizar as matrizes de pesos. Exemplos de tais algoritmos são o algoritmo de colônia de abelhas [47], enxame de partículas (*particle swarm*) [48, 49]. A próxima subseção abordará o ensemble de redes neurais para previsão.

2.3 Ensemble de redes neurais artificiais

Hansen e Salamon [50] foram os primeiros a demonstrar que a capacidade de generalização de um sistema neural poderia ser melhorada com a utilização de várias redes neurais, ou seja, treinando individualmente cada rede neural e posteriormente combinando seus resultados de alguma forma. No caso de RNA's, o ensemble é definido como uma maneira de combinar saídas obtidas de outras redes neurais. O ensemble de redes neurais ocorre em dois estágios: no primeiro, um determinado número de redes neurais é treinada, e no segundo as saídas dessas redes neurais são combinadas. A figura 2.5 ilustra o funcionamento de um ensemble de redes neurais.

Como se pode observar pela figura, cada rede neural RNA_t , $t = 1, 2, \dots, k$, recebe os dados de entrada x_{tu} com u representando a quantidade de entradas de cada rede (os dados podem não ser os mesmos para as redes) produzindo a saída y_t . Essas saídas são ponderadas para compor a solução do ensemble.

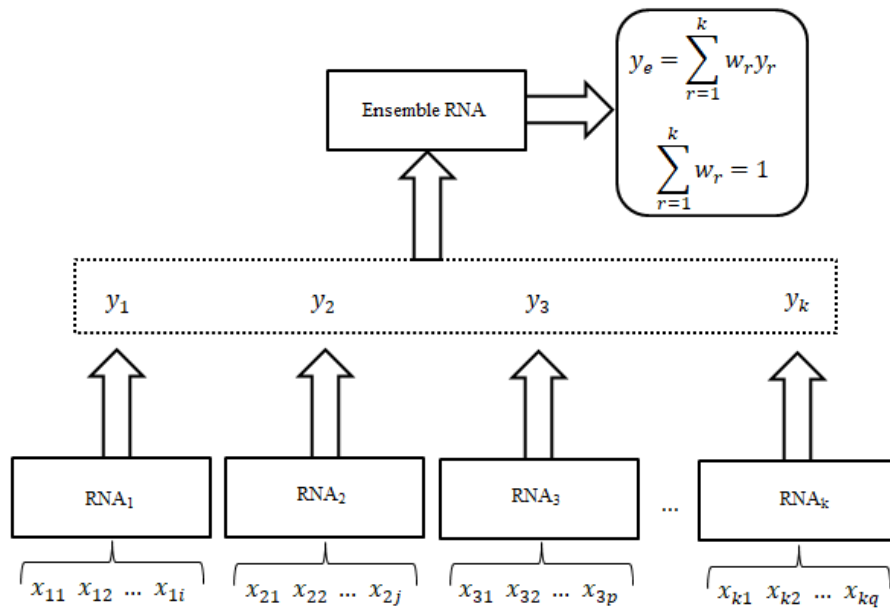


Figura 2.5 – Ensemble de redes neurais artificiais
Fonte: adaptado de [51].

A literatura inclui estudos sobre o uso do ensemble de redes neurais para fazer previsão. Jovanović et al. [52] concluíram que a precisão de previsão por meio do ensemble de três redes neurais (feedforward com backpropagation, rede RBF e sistema de inferência neuro-fuzzy adaptável) na previsão do consumo de energia solar de um campus universitário alcança melhores resultados do que redes individuais. Saviozzi et al. [53] apresentaram o design e a implementação de duas funcionalidades em um sistema de distribuição real: previsão de carga e modelagem de carga. Os algoritmos foram baseados em um ensemble de redes neurais, proporcionando bom desempenho em que a previsão de carga foi realizada usando o método do ensemble básico (BEM), que define a saída do ensemble como uma média aritmética simples das saídas das redes neurais individuais. Uma metodologia baseada em redes neurais foi apresentada em [54] para gerar previsões para três usinas de energia solar na Itália e indicou que a solução combinada produz os melhores resultados. Jetcheva et al. [55] demonstraram um modelo de ensemble de redes neurais para previsão de eletricidade no dia seguinte. Neste trabalho também é descrito um método automatizado para selecionar os parâmetros do modelo, facilitando sua implementação em problemas reais.

2.4 Planejamento e análise de experimentos (DOE)

A metodologia de planejamento de experimentos é antiga, sendo desenvolvida por Ronald A. Fisher na década de 20 e aprimorada por diversos pesquisadores, dentre eles Taguchi. Segundo Montgomery [56] um experimento é um teste ou uma série de testes

nas quais mudanças propositais são realizados nas variáveis de entrada de um processo ou sistema, visando observar e identificar mudanças correspondentes na resposta de interesse. Uma grande vantagem desta metodologia é estabelecer uma relação entre as saídas e os fatores de entrada e comparado ao método "um fator de cada vez" um importante ganho que esta metodologia proporciona é que os fatores de entrada podem ser modificados simultaneamente, ao contrário do outro método em que os fatores são alterados de forma individual, mantendo os demais constantes.

O DOE é uma metodologia de otimização estatística que objetiva encontrar através dessa análise simultânea os melhores níveis dos fatores para um conjunto de parâmetros que influenciam um processo. Ainda de acordo Montgomery [56] os principais termos para o entendimento desta metodologia são:

- Respostas: é a medição de desempenho de um sistema para uma determinada configuração de fatores. Essas variáveis descrevem o comportamento de um sistema para estas configurações. São as variáveis dependentes y do experimento.
- Fatores: são classificados em controláveis e incontroláveis. Os fatores controláveis são variáveis de entrada que sofrerão alterações durante a execução dos experimentos com o objetivo de provocar estímulos nas variáveis de respostas de interesse [57]. Os fatores incontroláveis são conhecidos como ruídos, pois tem interferência nos processos. Segundo Kelton [58] procura-se trabalhar com os fatores controláveis nos experimentos de simulação, pois eles são mais relevantes para decisões que devem ser feitas sobre implementação de sistemas.
- Níveis de fatores: são as variações de um mesmo fator, isto é, quais valores ou qual a configuração que o fator poderá assumir no arranjo experimental.
- Matriz experimental: é uma tabela que apresenta as combinações dos fatores analisados e seus respectivos níveis assumidos. A dimensão desta tabela depende do tipo de experimento a ser executado (completo, fracionado, com dois ou mais níveis, etc.)
- Tratamento: É a combinação dos fatores controláveis e seus respectivos níveis assumidos para a realização de cada experimento segundo a matriz experimental.
- Efeito principal: são os efeitos observados na resposta quando se muda os níveis dos fatores de controle.
- Efeito interação: acontecem quando o efeito combinado de dois ou mais fatores interferem na resposta. Pode ocorrer a situação em que um efeito principal não é observado, mas quando combinado com outros fatores o efeito principal pode ser manifestado.

- Aleatorização: É a maneira de se definir a ordem dos tratamentos da matriz experimental. Geralmente utilizado para diluir os efeitos produzidos por fatores de ruído.
- Modelo: relação entre os fatores estudados e as saídas do processo.

A relação entre as entradas, saídas, fatores controláveis e incontroláveis são exibidos na figura 2.6

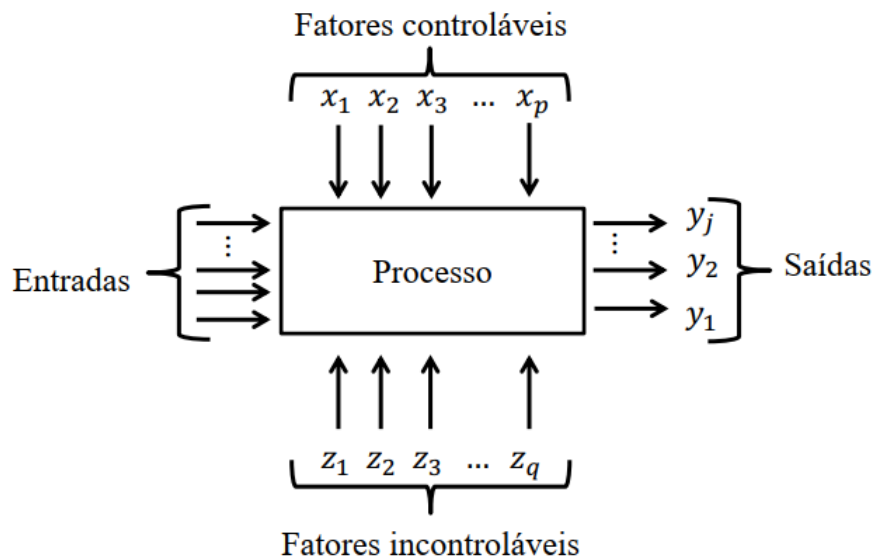


Figura 2.6 – Fatores de um processo
Fonte: adaptado de [56].

2.4.1 Arranjos fatoriais

Os arranjos fatoriais são indicados em situações nas quais há a existência de diversos fatores e se deseja saber quais são os de maior influência sobre uma determinada resposta. Os arranjos fatoriais são construídos com base em regras e usualmente contém os dois valores extremos de cada fator do problema (podendo ser qualitativos ou quantitativos e estar codificados ou não). Arranjos com mais de dois valores para os seus níveis também são encontrados [59].

O arranjo fatorial mais comum é o fatorial completo, onde o número de experimentos N depende da quantidade de fatores em estudo e dos níveis assumidos por esses fatores. Para o caso de fatores com dois níveis, N obedece a expressão

$$N = 2^k \quad (2.2)$$

onde k é o número de fatores envolvidos.

Devor et al. [60] detalha o procedimento que pode ser utilizado para determinar a matriz experimental em um experimento fatorial 2^k . Considere então um delineamento

fatorial 2^k sendo os k fatores representados por $x_1, x_2, x_3, \dots, x_k$ e os seus níveis superior e inferior representado por $+$ e por $-$ respectivamente. A matriz experimental é determinada da seguinte forma:

- Para a coluna que representa o fator x_1 os níveis mudam de sinal a cada mudança de linha, ou seja, o sinal dessa colua se altera a cada $2^0 = 1$ experimento;
- Para a coluna que representa o fator x_2 , os níveis mudam de sinal a cada duas mudanças de linhas, ou seja, o sinal dessa coluna se modifica a cada $2^1 = 2$ experimentos;
- Para a coluna que representa o fator x_3 , os níveis mudam de sinal a cada duas mudanças de linhas, ou seja, o sinal dessa coluna se modifica a cada $2^2 = 2$ experimentos; a coluna que representa o fator x_k , os níveis mudam de sinal a cada 2^{k-1} experimentos.

Com este planejamento garante-se que todas as colunas da matriz experimental sejam ortogonais entre si. De forma a exemplificar o procedimento exposto acima, a tabela 2.1 ilustra um delineamento fatorial 2^4 .

Tabela 2.1 – Arranjo fatorial 2^4

Tratamento	Fatores			
	x_1	x_2	x_3	x_4
1	-	-	-	-
2	+	-	-	-
3	-	+	-	-
4	+	+	-	-
5	-	-	+	-
6	+	-	+	-
7	-	+	+	-
8	+	+	+	-
9	-	-	-	+
10	+	-	-	+
11	-	+	-	+
12	+	+	-	+
13	-	-	+	+
14	+	-	+	+
15	-	+	+	+
16	+	+	+	+

Observe que os sinais da primeira, segunda, terceira e quartas colunas se modificam a cada 1, 2, 4 e 8 experimentos respectivamente.

Este tipo de arranjo contempla todo o espaço experimental. No entanto uma quantidade elevada de fatores implicará em um número elevado de experimentos e como alternativa para suprir este problema, surgem os arranjos fatoriais fracionados.

2.4.2 Arranjos Fatoriais Fracionados

Quanto maior o número de fatores (k) em estudo em um experimento, maior será a quantidade de experimentos a realizar. Considere, por exemplo, um experimento fatorial onde o número de fatores em estudo é $k = 6$, resultando em 2^6 experimentos. Neste caso, 6 graus de liberdade são destinados aos efeitos principais e 15 graus de liberdade são destinados às interações de segunda ordem. Dos 62 graus de liberdade totais, 42 graus de liberdade são usados para estimar as interações de ordem mais altas, isto é, ordens maiores ou iguais a três. O que ocorre é que muitas vezes existe pouco interesse nessas interações mais altas, principalmente quando se estuda um processo ou sistema. Caso o experimentador possa admitir que as interações de ordem mais alta possam ser descartadas, as informações sobre os efeitos principais e as interações de ordem baixa podem ser alcançadas com a realização de uma fração do experimento fatorial completo. De forma geral, um experimento fatorial completo 2^k pode ser estudado em uma fração de $\frac{1}{2^p}$ do experimento completo (p é a quantidade de geradores de confundimento independentes). Assim teremos um experimento fatorial fracionado 2^{k-p} . Quando $p = 2$ temos $\frac{1}{2^2}$ de fração, resultando em 2^{k-2} experimentos.

De forma semelhante ao que acontece com o delineamento fatorial completo, a construção da matriz experimental no arranjo fatorial fracionado segue o mesmo raciocínio, porém a quantidade de colunas Q da matriz que receberão mudança dos níveis dos fatores obedece a regra

$$Q = k - p. \quad (2.3)$$

Para maiores detalhes sobre como é originado cada gerador de confundimento veja [56]. No nosso exemplo de um fatorial 2^4 , o correspondente fatorial fracionado de meia fração, isto é 2^{4-1} apresenta a matriz experimental conforme a tabela 2.2.

Veja que segundo 2.3, $Q = 3$, ou seja, somente as três primeiras colunas recebem mudanças de sinal de níveis de fatores a cada um, dois e três experimentos. A quarta coluna é definida através de um gerador de confundimento (na tabela acima o nível do fator D é gerado a partir dos níveis dos fatores A, B e C).

2.4.3 Resolução de um experimento

Um detalhe importante que deve ser considerado nos experimentos fatoriais fracionados é a resolução do experimento que está sendo utilizada. Segundo [56], a resolução de um delineamento fatorial fracionado está associada ao nível de confundimento entre

Tabela 2.2 – Arranjo Fatorial Fracionado 2^{4-1}

Tratamento	Fatores			
	x_1	x_2	x_3	x_4
1	-	-	-	-
2	+	-	-	+
3	-	+	-	+
4	+	+	-	-
5	-	-	+	+
6	+	-	+	-
7	-	+	+	-
8	+	+	+	+

os efeitos dos fatores. Segundo o autor citado anteriormente, recomenda-se que os experimentos tenham maior resolução possível, pois através da alta resolução é possível desprezar os efeitos de interações de ordem mais alta. Definiremos abaixo as resoluções III, IV e V de um delineamento fatorial fracionado, que são consideradas como as mais importantes e utilizadas [56, 61].

1. Resolução III: nenhum efeito principal é confundido com outro efeito principal, mas efeitos principais podem ser confundidos com interações de segunda ordem, podendo haver confundimento de interações de segunda ordem entre si.
2. Resolução IV: nenhum efeito principal é confundido com outro efeito principal ou qualquer interação de dois fatores, mas interações de dois fatores podem ser confundidas entre si.
3. Resolução V: nenhum efeito principal é confundido com outro efeito principal ou interação de dois fatores, mas interações de dois fatores podem ser confundidas com interações de três fatores.

2.4.4 Etapas para a execução do experimento

Coleman e Montgomery [62] indicam alguns passos para a realização dos experimentos (figura 2.7) visando garantir maior confiabilidade nas conclusões obtidas. Essas etapas são:

- (a) Definição do problema: refere-se a definição do problema a ser atacado, bem como os objetivos a serem alcançados.
- (b) Escolha dos fatores e níveis: relacionada a coleta de informações do produto ou processo. Devem ser listados todos os fatores de controle, ruídos e variáveis de resposta. É necessário planejar muito bem quais são os fatores que serão analisados

no experimento, de maneira a não excluir fatores relevantes e também não testar um número muito grande de fatores, uma vez que o número de experimentos a serem realizados é função direta da quantidade de fatores.

- (c) Seleção das respostas: definição das medidas de desempenho que serão observadas durante o processo experimental.
- (d) Escolha do delineamento experimental: determinação de qual arranjo experimental é indicado para o problema (fatorial completo, fatorial fracionado, etc.)
- (e) Realização dos experimentos: execução de cada tratamento segundo a matriz experimental definida no item anterior.
- (f) Análise estatística: Nesta etapa os conceitos estatísticos são aplicados para descrever qual o efeito provocado na variável de resposta (fator dependente), pelas variáveis de controle (fatores independentes). A análise estatística ainda permite tomar decisões quanto a aceitação ou rejeição das hipóteses formuladas nas etapas iniciais do experimento. Atualmente, estas análises são auxiliadas por softwares estatísticos como MINITAB, STATISTICA, EXCEL, etc., que realizam cálculos estatísticos, geram relatórios e gráficos para a realização das análises.
- (g) Conclusões e recomendações: relacionada a extração de informações das análises estatísticas para a recomendação de ações.

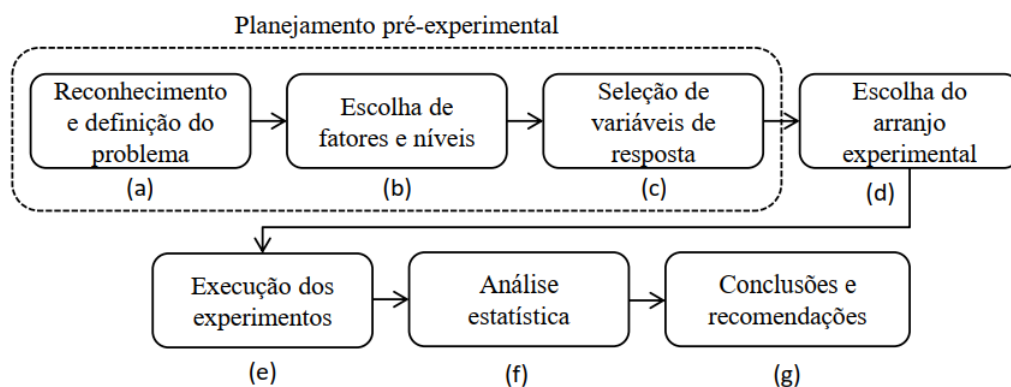


Figura 2.7 – Passos para a realização de um experimento

Fonte: adaptado de [62].

2.4.5 Planejamento de experimentos em simulação

O uso do delineamento de experimentos é amplamente utilizado em manufatura, sendo que sua utilização fora do campo do desenvolvimento e aperfeiçoamento de produtos e serviços demoraram para ser descobertas, firmando-se como estratégia vantajosa [57].

Segundo Balestrassi et al. [4], o treinamento de uma RNA (que envolve a experimentação com diversos parâmetros de entrada em algoritmos computacionais inteligentes) pode ser encarado como um estudo simulatório do problema de RNAs. Como afirmado anteriormente, as RNAs são candidatas à previsão de carga de eletricidade devido às características não lineares de uma série temporal de carga elétrica. Entretanto, configurar uma rede neural não é uma tarefa simples e a configuração escolhida pode afetar o desempenho. Técnicas de tentativa e erro, bem como indicações bibliográficas são comumente usadas para parametrizar uma rede e o estudo de cada parâmetro individualmente pode levar a conclusões duvidosas. Para evitar esses problemas, a metodologia estatística do DOE leva a resultados mais confiáveis, pois além de estudar vários fatores simultaneamente, as interações entre esses fatores também podem ser detectadas.

Em uma pesquisa recente na base de dados do Web of Science disponível no portal periódicos da CAPES, uma busca feita combinando a palavra chave "artificial neural network" com "design of experiments" encontrou 42 referências no período de 2000 até 2019, o que revela ser um indicativo de que a ferramenta de planejamento experimental é pouco utilizada em assuntos envolvidos com redes neurais. Khaw et al. [63] descreveram uma aplicação inovadora do método Taguchi para a determinação de parâmetros de projeto incluindo os aspectos microestruturais e macroestruturais de uma rede neural. Os resultados obtidos nesta pesquisa mostraram que o método Taguchi fornece um meio eficaz para melhorar o desempenho da rede neural em termos de velocidade e precisão de aprendizado. Um método foi proposto em [64] para otimizar a arquitetura de uma rede neural *perceptron* multicamadas aplicada para a previsão rugosidade média de superfície. O método foi expresso no formato de um algoritmo que emprega duas técnicas da metodologia DOE: fatoriais completos e operações evolucionárias, levando a uma redução significativa de erros de previsão de rugosidade nas operações de usinagem. Balestrassi et al. [4] usaram a técnica DOE para treinar redes neurais em problemas de previsão de séries temporais não lineares. Os autores utilizaram um projeto experimental utilizando os fatores de classificação da caixa de ferramentas de redes neurais do software Statistica [65] no modo de RNA customizado e apresentaram um estudo de caso de previsão de carga elétrica de curto prazo de consumidores industriais de uma empresa no Brasil. Seguindo a mesma estratégia, este trabalho fará um estudo de fatores significativos na parametrização de uma rede neural utilizando o modo solucionador de problemas inteligentes no referido software. Saviozzi et al. [66] propuseram um ensemble de RNA para previsão de carga em sistemas de gerenciamento de distribuição. As saídas do ensemble foram obtidas através da média das saídas das redes neurais que formaram o ensemble.

A partir do que foi feito no trabalho de Saviozzi et al. [66], este trabalho propõe a utilização da técnica DOE não apenas na parametrização da rede. Também é proposto um projeto de experimentos com mistura (MDE) para a determinação da ponderação a ser utilizada no ensemble. Mais especificamente, esta técnica, que será tratada adiante,

auxilia através de uma malha de pesos para determinar uma superfície de resposta onde a resposta de interesse é modelada como uma função dos pesos. Essa superfície é útil no processo de otimização a fim de se determinar a ponderação que será usada para encontrar a série residual de previsões do ensemble.

2.4.6 Metodologia de superfície de resposta

Segundo Montgomery [56] a metodologia de superfície de resposta (RSM) (Response Surface Methodology) é uma coleção de técnicas matemáticas e estatísticas úteis para a modelagem e análise de problemas nos quais uma resposta de interesse no qual deseja-se otimizar é influenciada por várias variáveis. Considerando que as relações entre as respostas e as variáveis independentes são desconhecidas, objetiva-se então encontrar uma boa aproximação para representar as respostas de interesse como função destas variáveis. Suponha que uma determinada resposta y é determinada em função de duas variáveis controláveis x_1 e x_2 e que objetiva-se determinar um valor ótimo para y . A resposta y pode ser modelada em função dessas variáveis, isto é,

$$y = f(x_1, x_2) + \epsilon \quad (2.4)$$

onde ϵ denota o erro observado na resposta. Se denotarmos por $\zeta = E(y)$ então a superfície representada por $\zeta = f(x_1, x_2)$ é chamada de superfície de resposta conforme representado na figura 2.8.

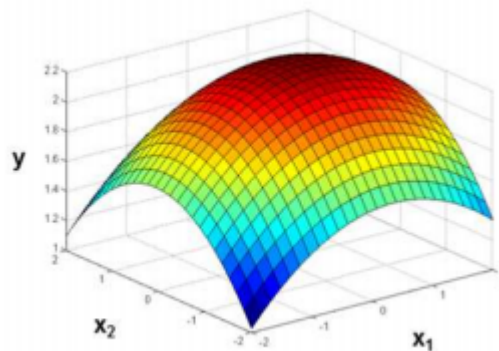


Figura 2.8 – Superfície de resposta para y
Fonte: [67]

O primeiro passo dentro da RSM é encontrar uma aproximação adequada para o verdadeiro relacionamento entre y e o conjunto de variáveis. Usualmente, um polinômio de baixa ordem para qualquer região de interesse é utilizado. Se a resposta é bem modelada por uma função linear das variáveis independentes, essa relação pode ser aproximada pelo modelo de primeira ordem

$$y = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \dots + \alpha_k x_k + \epsilon \quad (2.5)$$

em que y denota a resposta de interesse, x_i ($i = 1, 2, \dots, k$) as variáveis independentes, α_i os coeficientes do modelo a serem estimados e ϵ o erro experimental. Entretanto, caso exista curvatura na resposta y , então a função de aproximação mais indicada é um polinômio de ordem superior, como o modelo de segunda ordem descrito pela equação 2.6.

$$y = \alpha_o + \sum_{i=1}^k \alpha_i x_i + \sum_{i=1}^k \alpha_{ii} x_i^2 + \sum_{i,j} \sum_{i<j} \alpha_{ij} x_i x_j + \epsilon. \quad (2.6)$$

De acordo com Box e Draper [68] os dois modelos descritos acima conseguem representar quase todos os problemas relacionados à superfície de respostas. No entanto, como um dos objetivos da RSM consiste na otimização das respostas é recomendável (quando possível) representá-las por meio dos modelos de segunda ordem, já que a curvatura apresentada por estes define a localização de um ponto estacionário. Com relação à capacidade de representação das respostas, é improvável que o modelo polinomial se comporte como uma aproximação adequada para todo o espaço experimental coberto pelas variáveis independentes. Entretanto, para uma região específica, tais modelos tem se mostrado eficientes [56].

A determinação dos coeficientes dos modelos das equações 2.5 e 2.6 é realizada através do emprego do método dos Mínimos Quadrados Ordinários (Ordinary Least Squares - OLS). Segundo [69] algumas razões para a utilização da metodologia de superfície de resposta podem ser destacadas abaixo:

- a função de resposta é caracterizada em uma região de interesse;
- níveis de fatores podem ser determinados para que a resposta seja otimizada (maximizada, ou minimizada, por exemplo);
- níveis de fatores podem ser determinados a fim de otimizar diversas respostas simultaneamente.

2.4.7 Experimentos por arranjos de misturas

Segundo Montgomery [56] os experimentos de misturas são uma classe especial de experimentos do tipo superfície de resposta nos quais as variáveis de entrada são componentes ou ingredientes de uma mistura e as respostas são funções das proporções de cada componente. Nesse tipo de experimento, os níveis dos parâmetros não são independentes, isto é, se w_i , $i = 1, \dots, k$ representam as proporções de k componentes de uma mistura, em cada experimento os valores de w_i devem satisfazer uma restrição de totalidade, ou seja

$$\sum_{i=1}^k w_i = 1 \quad (2.7)$$

onde $w_i \geq 0, i = 1, \dots, k$. Geometricamente esta condição pode ser visualizada para o caso com duas e três componentes conforme figura 2.9.

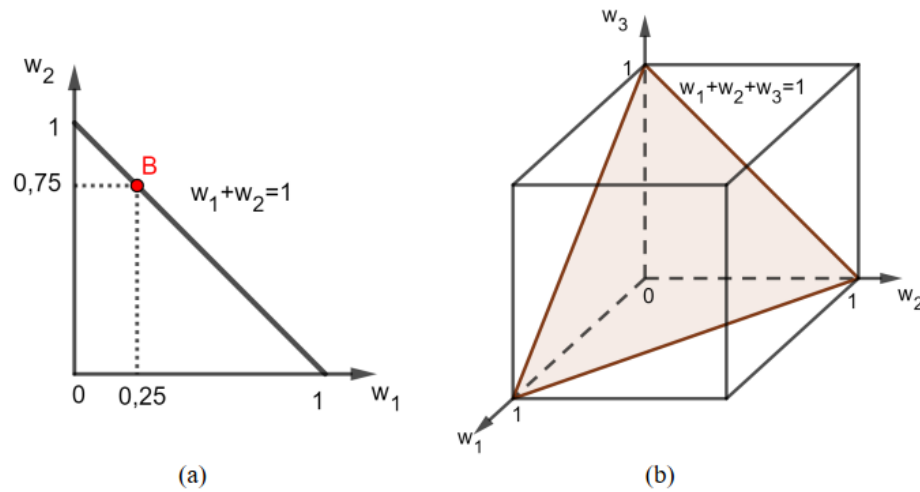


Figura 2.9 – Região experimental para experimentos de misturas com duas (a) e três (b) componentes

Fonte: adaptado de [56].

Para o caso $k = 2$ esta restrição descreve uma reta no plano $w_1 \times w_2$ limitado ao primeiro quadrante $w_1, w_2 \geq 0$ e para $k = 3$ a restrição é determinada pelo plano $w_1 + w_2 + w_3 = 1$ limitada pelo primeiro octante. Para o caso de mistura com duas componentes, as intersecções da reta com os eixos w_1 e w_2 representam as misturas puras e ao longo da reta as misturas binárias (o ponto B na figura 2.9a representa a mistura de dois componentes com as proporções de 0,25 para w_1 e 0,75 para w_2). Já no caso $k = 3$ as misturas puras localizam-se nos vértices do triângulo equilátero da figura 2.9b. As misturas binárias estão presentes nos lados deste triângulo e as misturas ternárias estão presentes no interior do triângulo. A existência dessas características torna então necessário que os experimentos de misturas sejam planejados e conduzidos através de arranjos específicos e, nesse contexto, os arranjos simplex têm sido os mais utilizados [70, 56].

Os arranjos simplex são definidos como uma configuração triangular na qual os vértices do triângulo representam as proporções máximas das variáveis de entrada e os pontos interiores a este triângulo descrevem as possíveis combinações destas variáveis. Podem ser de dois tipos principais: o arranjo simplex lattice e o arranjo simplex centroide. No simplex lattice, as k variáveis de entrada definem pontos cujas proporções assumidas levam em consideração $m + 1$ valores igualmente espaçados entre 0 e 1, de tal forma que:

$$w_i = 0, \frac{1}{m}, \frac{2}{m}, \dots, 1 \quad i = 1, \dots, k \quad (2.8)$$

e neste arranjo, todas as combinações possíveis das proporções são definidas pela equação 2.8. Sendo m o grau lattice do arranjo podemos estabelecer uma relação que determina

a quantidade de experimentos em um arranjo simplex lattice em função do seu grau m e da quantidade k de componentes:

$$S_L = \frac{(k + m - 1)!}{m!(k - 1)!}. \quad (2.9)$$

A figura 2.10a mostra um arranjo simplex lattice com três componentes e grau $m = 2$. Em relação ao simplex centróide as k componentes de entrada são configuradas em $2^k - 1$ pontos, obtidos a partir do seguinte critério:

$$\binom{k}{j} \text{ permutações de } \left(\frac{1}{j}, \frac{1}{j}, \dots, \frac{1}{j}, 0, 0, \dots, 0 \right)$$

$j = 1, 2, \dots, k$ A figura 2.10b ilustra um simplex centróide com 3 componentes. Veja que neste caso temos:

- 3 permutações de $(1, 0, 0)$ representando os vértices do triângulo (misturas puras);
- 3 permutações de $\left(\frac{1}{2}, \frac{1}{2}, 0\right)$ representando os pontos médios dos lados do triângulo (misturas binárias)
- o centróide $\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$ representando o ponto no interior do triângulo (mistura de três componentes)

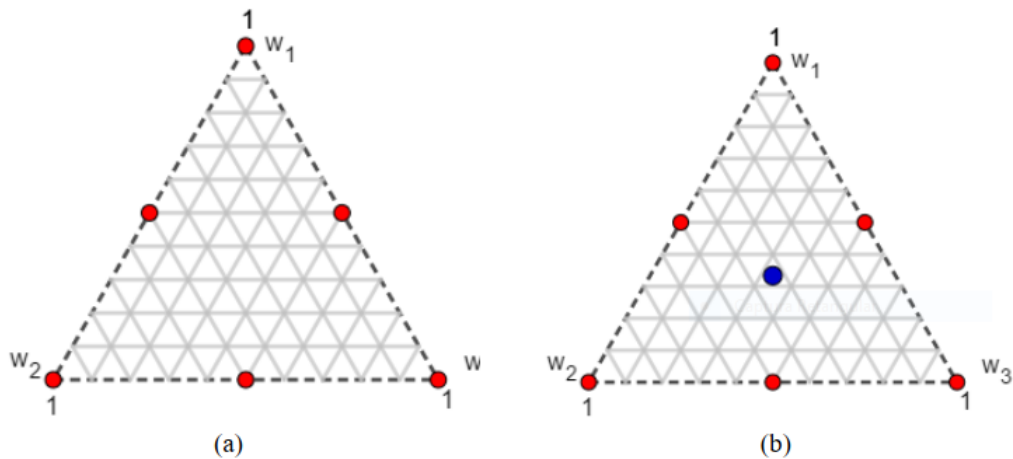


Figura 2.10 – Arranjos simplex lattice (a) e centróide (b)

Fonte: adaptado de [67].

Uma desvantagem dos arranjos simplex diz respeito ao fato de que a maioria dos experimentos ocorre nas fronteiras do arranjo, o que faz com que poucos pontos da parte interna sejam testados. Assim, recomenda-se, sempre que possível, aumentar o número de experimentos através da adição de pontos internos aos arranjos, como os pontos centrais e também os pontos axiais. Os pontos axiais podem ser definidos como todas as permutações de

$$\left(\frac{k+1}{2k}, \frac{1}{2k}, \dots, \frac{1}{2k} \right).$$

No caso dos arranjos de misturas os pontos centrais correspondem ao próprio centroide. A Figura 2.11 apresenta os arranjos incorporados com a adição destes pontos.

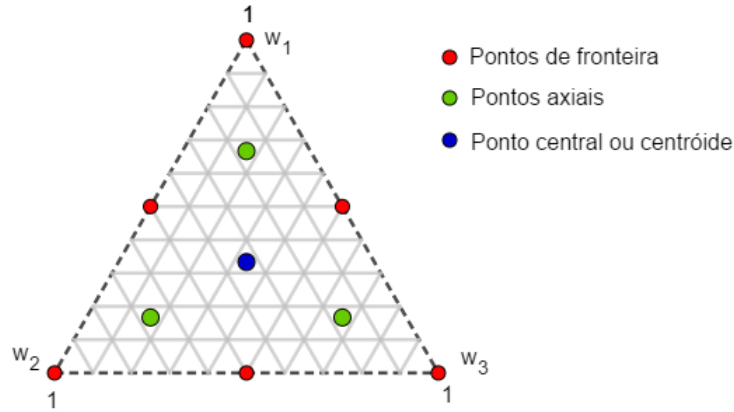


Figura 2.11 – Arranjos simplex incorporado
Fonte: autor.

Quanto aos modelos matemáticos utilizados para a representação das respostas, verifica-se que os modelos de misturas apresentam algumas diferenças em relação aos polinômios empregados na metodologia de superfície de resposta, principalmente devido a existência da restrição $w_1 + w_2 + \dots + w_k = 1$. Dependendo do comportamento da resposta analisada, os seguintes modelos podem ser utilizados:

1. Modelo linear:

$$y(\mathbf{w}) = \sum_{i=1}^k \gamma_i w_i \quad (2.10)$$

2. Modelo quadrático:

$$y(\mathbf{w}) = \sum_{i=1}^k \gamma_i w_i + \sum_{i < j}^k \sum_{i < j}^k \gamma_{ij} w_i w_j \quad (2.11)$$

3. Modelo cúbico completo:

$$y(\mathbf{w}) = \sum_{i=1}^k \gamma_i w_i + \sum_{i < j}^k \sum_{i < j}^k \gamma_{ij} w_i w_j + \sum_{i < j}^k \sum_{i < j}^k \delta_{ij} w_i w_j (w_i - w_j) + \sum_{i < j < l}^k \sum_{i < j < l}^k \gamma_{ijl} w_i w_j w_l \quad (2.12)$$

Modelo cúbico especial:

$$y(\mathbf{w}) = \sum_{i=1}^k \gamma_i w_i + \sum_{i < j}^k \sum_{i < j}^k \gamma_{ij} w_i w_j + \sum_{i < j < l}^k \sum_{i < j < l}^k \gamma_{ijl} w_i w_j w_l \quad (2.13)$$

As equações 2.10, 2.11, 2.12 e 2.13 são denominadas polinômios canônicos de misturas ou polinômios de Scheffé [70].

Até o momento, o que expomos aqui levou em consideração a possibilidade de explorar toda a região experimental em um arranjo de mistura. No entanto, pode ocorrer restrições existentes nas proporções dos componentes, impedindo que toda região experimental seja analisada. As restrições relacionadas aos componentes se mostram na forma de intervalos, ou seja:

$$\tau_i \leq w_i \leq \varphi_i \quad i = 1, 2, \dots, k \quad (2.14)$$

em que τ_i e φ_i são respectivamente os limites inferiores e superiores para o i -ésimo componente. Essas intervalos fazem com que a região experimental agora seja um subconjunto da região experimental 2.7. Se $\varphi = 1$ no intervalo 2.14 verifica-se que o arranjo simplex não perde o formato, ou seja, a região experimental para os experimentos de mistura com essas restrições será um simplex contido dentro da região experimental 2.7 (figura 2.12).

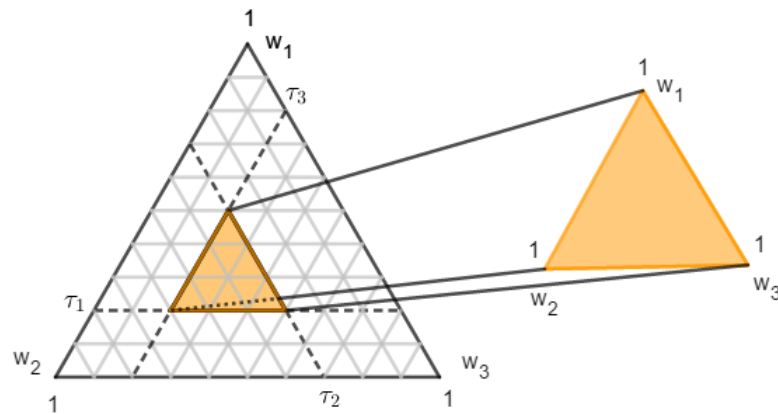


Figura 2.12 – Região experimental definida por restrições limitadas inferiormente
 Fonte: adaptado de [67].

Por outro lado, quando $\tau_i = 0$, ou seja, $0 \leq w_i \leq \varphi_i$ a região experimental terá um formato que irá depender dos valores exigidos, de forma que:

- Se $\sum_{i=1}^k \varphi_i - \min_{1 \leq i \leq k} \{\varphi_i\} \leq 1$, a região experimental será um simplex invertido contido dentro da região experimental original;
- Se $\sum_{i=1}^k \varphi_i - \min_{1 \leq i \leq k} \{\varphi_i\} \geq 1$ a região não será um simplex.

A figura ilustra essas afirmações para o caso de três componentes:

Vale ressaltar que o mesmo ocorre para os casos em que os limites inferiores τ_i e superiores φ_i aparecem simultaneamente, ou seja, a região experimental pode ser ou não um simplex, sendo o formato desta região determinada pelos valores dos limites inferiores e superiores. Myers e Montgomery [71] recomenda que quando a região experimental não apresentar o formato de um simplex indica-se a utilização de arranjos grados através de computador, como o arranjo de vértices extremos (*extreme vertice design*).

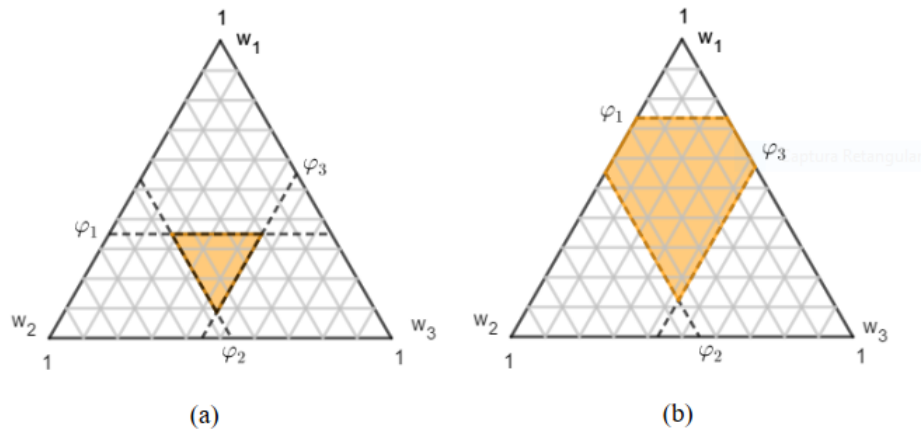


Figura 2.13 – Região experimental definida por restrições limitadas superiormente: (a) simplex invertido; (b) não simplex
 Fonte: adaptado de [67].

Para os casos em que não há perda de formato do arranjo simplex, é conveniente a definição de um novo conjunto de componentes que irá assumir os valores extremos 0 e 1 na região. Esta transformação é importante na modelagem e análise dos polinômios de misturas, uma vez que tende a reduzir os erros na estimação dos coeficientes dos modelos. Com esta transformação, um novo conjunto de variáveis caracteriza os pseudocomponentes (W_i). Para arranjos definidos por restrições inferiores, essas pseudocomponentes são calculados através da relação 2.15

$$W_i = \frac{w_i - \tau_i}{1 - \sum_{i=1}^k \tau_i} \tag{2.15}$$

em que $\sum_{i=1}^k \tau_i < 1$. Já para as restrições superiores, as pseudocomponentes são dadas por

$$W_i = \frac{\varphi_i - w_i}{\sum_{i=1}^k \varphi_i - 1} \tag{2.16}$$

onde $\sum_{i=1}^k \varphi_i > 1$.

O Cálculo dos pseudocomponentes é considerada uma tarefa relativamente simples. No entanto, softwares estatísticos realizam estas transformações de forma automática, fornecendo opções aos usuários para a visualização dos arranjos de mistura tanto na forma dos pseudocomponentes quanto na quantidade real das misturas.

2.5 Otimização Multiobjetivo

A otimização multiobjetivo consiste no processo de otimização de mais de uma resposta simultaneamente com objetivo de se encontrar as condições que possam satisfazer ambas as respostas consideradas. Entretanto, segundo Paiva [72] devido à característica dos problemas em engenharia e a competição entre as múltiplas respostas, é muito difícil determinar uma condição que otimize todas as respostas analisadas, fazendo com que a solução não seja única, mas sim um conjunto de soluções ótimas.

De forma prática, não existe uma única solução que otimize todas as respostas, tornando-se relevante a identificação de soluções que proporcione uma relação de perda e ganho (conhecida como *trade-off*) para as diferentes respostas, permitindo opções de escolha no processo de tomada de decisão. De acordo com Berezkin e Lotov [73] a solução de um problema de otimização multiobjetivo é encontrada por meio da construção da fronteira de Pareto. Existem inúmeros métodos na literatura para a solução de um problema de otimização multiobjetivo e conseqüentemente a determinação da fronteira. Tais estratégias de solução podem ser divididas em duas classes: os métodos de priorização e os métodos de agrupamento. Nos métodos de priorização, define-se uma função objetivo como sendo a mais importante e as demais são utilizadas como restrições. Como exemplos dessa abordagem podemos citar a programação lexicográfica [74] e o método restritivo ϵ [75]. Para o método de agrupamento a idéia é combinar as funções objetivos em uma única função, tornando-se o objetivo global do problema. A otimização da função global garante a otimização das funções individuais. O método de somas ponderadas e o método desirability são exemplos dessa abordagem de otimização.

2.5.1 Formulação matemática de problemas de otimização multiobjetivo

Matematicamente, quando-se possuem j funções a serem otimizadas ($j \geq 2$) o problema de otimização multi-objetivo pode ser formulado segundo a equação 2.17

$$\begin{aligned} \text{Min } \mathbf{F}(x) &= [f_1(\mathbf{x}), f_2(\mathbf{x}), \dots, f_j(\mathbf{x})] & (2.17) \\ \text{sujeito a: } \mathbf{g}_r(\mathbf{x}) &\leq 0 \quad r = 1, 2, \dots, k \\ \mathbf{h}_l(\mathbf{x}) &= 0 \quad l = 1, 2, \dots, m \end{aligned}$$

em que $\mathbf{x} \in \mathbb{R}^n$ são as variáveis de decisão, r e l o número de restrições de desigualdade e igualdade respectivamente que juntas determinam o espaço viável \mathbb{V}_v de variáveis de decisão. O espaço viável das funções objetivos é o conjunto $\mathbf{F}(\mathbb{V}_v)$. Em relação aos problemas de otimização multiobjetivo se faz necessário a definição de alguns conceitos:

- (1) Otimalidade de Pareto: Um vetor $\mathbf{x}^* \in \mathbb{V}_v$ é dito Pareto-ótimo se não existe qualquer outra solução $x \in \mathbb{V}_v$ tal que $f_i(x) \leq f_i(x^*)$ pra $i = 1, 2, \dots, j$ com $f_p(x) < f_p(x^*)$ em pelo menos uma função objetivo p . De forma mais simples Gomes [67] afirma que \mathbf{x}^* é Pareto-ótimo se não existe outra solução \mathbf{x} que acarrete redução em alguma função objetivo sem causar aumento simultâneo em pelo menos uma das outras funções. O conjunto de todos os pontos ótimos é denominado de soluções Pareto-ótimas (\mathbb{P})
- (2) Pontos dominados e não-dominados: Um ponto $\mathbf{F}(\mathbf{x}^*) \in \mathbf{F}(\mathbb{V}_v)$ é dito ser não-dominado se não existir $\mathbf{F}(\mathbf{x}) \in \mathbf{F}(\mathbb{V}_v)$ tal que $f_i(x) \leq f_i(x^*)$ para $i = 1, 2, \dots, j$ e $f_i(\mathbf{x}) < f_i(\mathbf{x}^*)$ para no mínimo uma função objetivo. Caso contrário o ponto é dito ser dominado.
- (3) Fronteira de Pareto (∂F): conjunto das soluções Pareto-ótimas.

Podemos ver através da figura 2.14 as três definições para o caso bi-objetivo.

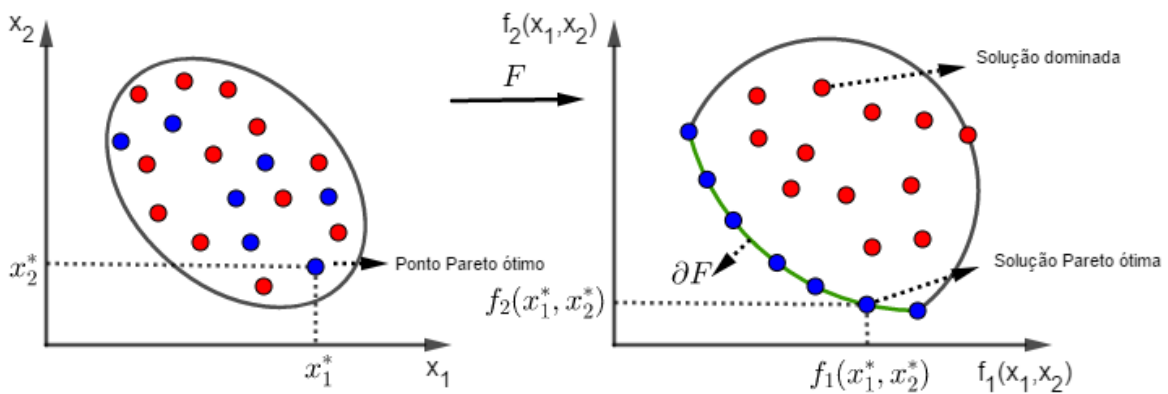


Figura 2.14 – Otimalidade e dominância de Pareto
 Fonte: adaptado de [76].

- (4) Pontos de âncora: Seja $\mathbf{x}^{i*} \in \mathbb{R}^n$ que minimiza o valor de $f_i(\mathbf{x})$. O Ponto de âncora F^{i*} é definido como:

$$\mathbf{F}^{i*} = (f_1(\mathbf{x}^{i*}), f_2(\mathbf{x}^{i*}), \dots, f_i(\mathbf{x}^{i*}), \dots, f_j(\mathbf{x}^{i*})) \quad (2.18)$$

Vale ressaltar que o ponto \mathbf{x}^{i*} é encontrado através da minimização da função $f_i(\mathbf{x})$ sujeitas as mesmas restrições do problema de otimização 2.17. No caso em que existem duas funções objetivo, teremos dois pontos de âncora. A linha que une esses dois pontos recebe o nome de linha de utopia e para os casos em que existam mais de duas funções objetivos temos um hiperplano de utopia.

- (5) Ponto de Utopia: O ponto

$$F^U = (f_1(\mathbf{x}^{1*}), f_2(\mathbf{x}^{2*}), \dots, f_j(\mathbf{x}^{j*})) \quad (2.19)$$

é chamado de ponto de utopia. Denotaremos daqui em diante $f_j(\mathbf{x}^{j*}) = f_j^U$.

(6) Pontos de nadir e pseudo nadir: Seja \mathbf{z}^{i*} o vetor solução do problema

$$\begin{aligned} \text{Max } f_i(\mathbf{x}) & & (2.20) \\ \text{sujeito a: } \mathbf{g}_r(\mathbf{x}) & \leq 0 \quad r = 1, 2, \dots, k \\ \mathbf{h}_l(\mathbf{x}) & = 0 \quad l = 1, 2, \dots, m. \end{aligned}$$

O Ponto

$$F^N = (f_1(\mathbf{z}^{1*}), f_2(\mathbf{z}^{2*}), \dots, f_j(\mathbf{z}^{j*})) \quad (2.21)$$

é chamado de ponto de nadir. Para simplificar a notação faremos $f_i(\mathbf{z}^{i*}) = f_i^N$. O ponto de pseudo nadir é definido como

$$F^{PN} = (f_1^{PN}, f_2^{PN}, \dots, f_j^{PN}) \quad (2.22)$$

em que $f_i^{PN} = \max \{f_i(\mathbf{x}^{1*}), f_i(\mathbf{x}^{2*}), \dots, f_i(\mathbf{x}^{j*})\}$.

A figura 2.15 ilustra essas definições para o caso bi-objetivo.

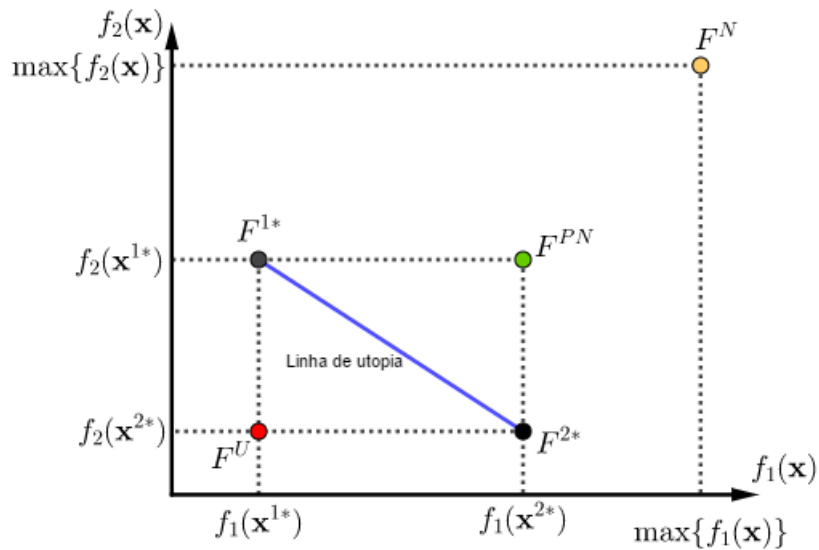


Figura 2.15 – Pontos de ancoragem, utopia, nadir e pseudo nadir para o caso bi-objetivo
Fonte: adaptado de [76].

2.5.2 Somas Ponderadas

A formulação matemática do método de somas ponderadas é:

$$\begin{aligned} \text{Min } \mathbf{F}(x) & = \sum_{i=1}^j w_i f_i(\mathbf{x}) & (2.23) \\ \text{Sujeito a: } \mathbf{g}_r(\mathbf{x}) & \leq 0 \quad r = 1, 2, \dots, k \\ \mathbf{h}_l(\mathbf{x}) & = 0 \quad l = 1, 2, \dots, m. \end{aligned}$$

Segundo Gomes [67] este método é caracterizado como uma sequência de problemas onde a função objetivo global $F(\mathbf{x})$ é definida por uma combinação linear das funções objetivos individuais $f_i(\mathbf{x})$ e seus respectivos graus de importância (w_i). Para a determinação do conjunto de soluções Pareto-ótimas, o problema 2.23 é resolvido diversas vezes utilizando diferentes combinações de pesos entre as funções objetivo [75].

O método de somas ponderadas é amplamente utilizado em problemas de otimização multiobjetivo. Entretanto, o método de somas ponderadas não é capaz de gerar uma fronteira uniformemente espaçada, mesmo quando a distribuição dos pesos w_i seja uniforme. Além disso, outro ponto importante a se destacar é que se o conjunto de soluções de Pareto for não convexo, a fronteira passa a ser não convexa e descontínua, formando clusters de soluções Pareto-ótimas em regiões de grande curvatura, porém, descontínuas no espaço de solução, o que é típico de problemas mal condicionados [77].

2.5.3 Método de otimização Desirability

De forma geral, procedimentos de otimização são construídos sobre uma abordagem utilizando programação linear que em síntese considera uma única resposta que está condicionada a múltiplas restrições. Em contraste a esses métodos, Derringer e Suich [78] aprimoraram um algoritmo denominado Desirability, proposto originalmente por Harrington [79] para otimização de respostas simultâneas que trata um conjunto de variáveis respostas sujeitas a restrições relacionadas a diferentes objetivos (maximização, minimização ou um alvo específico). De acordo com Gyseghem et al. [80] o Desirability é um método capaz de avaliar um conjunto de respostas simultaneamente, permitindo a determinação do conjunto de condições mais desejável para as respostas estudadas.

O fundamento é bastante simples: inicialmente é estabelecido um relacionamento entre as respostas e as variáveis independentes. Em geral isso pode ser feito através de RSM e de OLS. Fazendo o uso da formulação de Harrington (unilateral ou bilateral), cada uma das respostas do conjunto original é transformada, tal que d_i pertença ao intervalo $[0, 1]$. O valor de d_i aumenta quando a i -ésima resposta se aproxima dos limites estabelecidos.

O índice global D é obtida através da combinação das respostas transformadas por uma média geométrica simples, conforme a equação 2.24

$$D = \left(\prod_{i=1}^k d_i \right)^{\frac{1}{k}} \quad (2.24)$$

ou através da média geométrica ponderada

$$D = \left(\prod_{i=1}^k d_i^{z_i} \right)^{\frac{1}{\sum_{i=1}^k z_i}} \quad (2.25)$$

Vale destacar que se qualquer $d_i = 0$, isto é, quando uma resposta não consegue atingir o valor ótimo, $D = 0$, significando que não foi encontrado uma configuração de parâmetros de entrada que otimizasse a resposta. Há funções de desejabilidade particulares para cada tipo de resposta. Dependendo da aplicação, há três tipos de respostas desejáveis: quanto maior, melhor (maximização); quanto menor, melhor (minimização) e quanto mais próximo do alvo, melhor (alvo). Baseado nesta idéia e nos limites superior, inferior e no alvo de cada função, Derringer e Suich [78] criaram funções específicas para cada tipo de resposta desejada.

Considere a situação em que se deseja que cada resposta y_i esteja dentro de um intervalo $[a_i, c_i]$ de forma que a resposta é inaceitável se $y_i < a_i$ ou se $y_i > c_i$. Além disso, existe um ponto $b_i \in [a_i, c_i]$ tal que a desirability aumenta em $[a_i, b_i]$ e diminui em $[b_i, c_i]$. Assim, pode-se definir então o mínimo e o máximo valor aceitável para \hat{y}_i , que são respectivamente a_i e c_i . Através da escolha dos parâmetros r e s , podem ser atribuídos vários níveis da desirability para \hat{y}_i , conforme figura 2.16 [81].

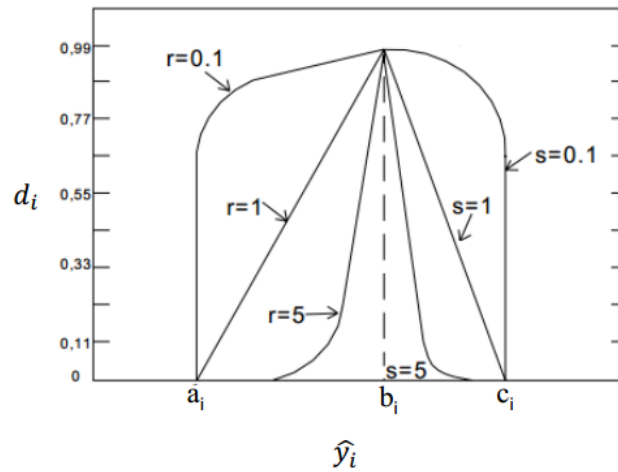


Figura 2.16 – Definição do alvo para o desirability
 Fonte: [70]

Note que para grandes valores de r e s (~ 10), o valor \hat{y}_i fica próximo de b_i . Por outro lado, para qualquer valor de \hat{y}_i acima de a_i e baixo de c_i aceitável, pequenos valores de r e s devem ser escolhidos. Isso é útil no caso em que se deseja que uma variável resposta y_i atinja um valor alvo. Com o objetivo de se obter uma resposta no alvo, a função desirability é calculada através da equação

$$d_i = \begin{cases} \left(\frac{\hat{y}_i - a_i}{b_i - a_i} \right)^r, & \text{se } a_i \leq \hat{y}_i \leq b_i \\ \left(\frac{\hat{y}_i - c_i}{b_i - c_i} \right)^s, & \text{se } b_i \leq \hat{y}_i \leq c_i \\ 0, & \text{se } \hat{y}_i < a_i \text{ ou } \hat{y}_i > c_i \end{cases} \quad (2.26)$$

Quando se deseja a maximização de uma resposta, a fórmula de transformação empregada é

$$d_i = \begin{cases} 0, & \text{se } \hat{y}_i < L_i \\ \left(\frac{\hat{y}_i - L_i}{U_i - L_i} \right)^t, & \text{se } L_i \leq \hat{y}_i \leq U_i \\ 1, & \text{se } \hat{y}_i > U_i \end{cases} \quad (2.27)$$

onde L_i e U_i são respectivamente o menor e o maior valor aceitável para a resposta. Observe que se $\hat{y}_i < L_i$ então a desirability assume o valor nulo, e que quando $L_i \leq \hat{y}_i \leq U_i$ assume um valor entre 0 e 1, sendo o formato da função dependente do valor assumido por r , conforme figura 2.17.

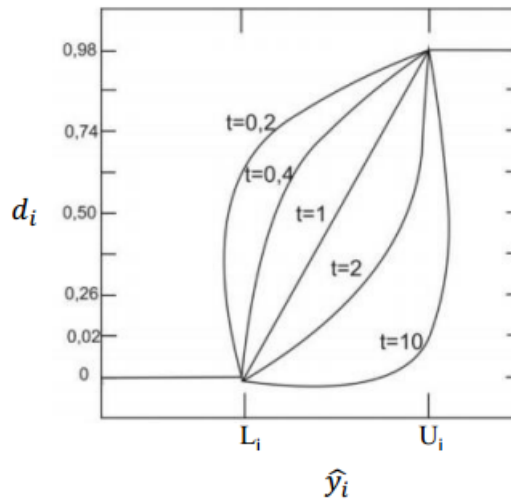


Figura 2.17 – Formato da função de maximização para o desirability
Fonte: [70]

No caso em que se deseja a minimização a utiliza-se

$$d_i = \begin{cases} 0, & \text{se } \hat{y}_i > H_i \\ \left(\frac{H_i - \hat{y}_i}{H_i - T_i} \right)^r, & \text{se } L_i \leq \hat{y}_i \leq T_i \\ 1, & \text{se } \hat{y}_i < T_i \end{cases} \quad (2.28)$$

em que H_i é valor aceitável mais alto para a i -ésima resposta.

Em se tratando da curvatura apresentada nas funções exponenciais especificando o peso, elas podem ser manipuladas de forma a alcançar com maior rapidez o alvo de interesse. Paiva [72] apresenta um esquema envolvendo alguns casos em que esses pesos são alterados conforme a figura 2.18.

Assim, com a determinação da desirability total, o problema de otimização multivariado fica reduzido a um problema univariado.

O processo de solução baseia-se na maximização de D . Pelo que já foi exposto, D é função de d_i , que por sua vez é função da variável predita y_i . As variáveis preditas

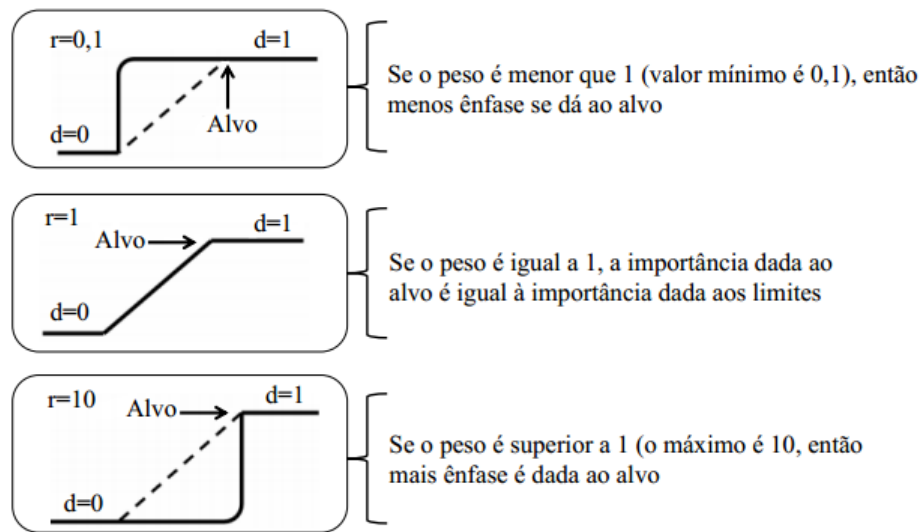


Figura 2.18 – Relação de importância entre os alvos e os limites

Fonte: [72]

dependem das variáveis de entrada do processo, logo D também é dependente destas variáveis. O problema reside no fato de se encontrar um cenário para estas variáveis independentes que maximizam D .

É evidente que o método desirability é flexível, devido ao fato de direcionar as respostas de acordo com seus respectivos objetivos individuais, maximizando algumas, minimizando outras ou até mesmo buscar um valor alvo (tudo isso simultaneamente). Entretanto, o método desirability possui pontos negativos, não levando em consideração a variância e nem a correlação entre as respostas.

Outro ponto importante no método desirability é a escolha subjetiva das funções individuais, onde a inexperiência do pesquisador pode conduzir a resultados imprecisos. Adicionalmente, negligenciar a correlação entre as respostas altera a estrutura da função D , colocando em risco a determinação das condições ótimas.

Para contornar os problemas do método desirability e somas ponderadas, Das e Dennis [82] desenvolveram o método Interseção Normal à Fronteira (NBI - *Normal Boundary Intersection*), mostrando ser possível a construção de fronteiras contínuas e uniformemente espaçadas. Este método será tratado de forma isolada na próxima subseção.

2.5.4 Método de interseção normal a fronteira

Como mencionado anteriormente, o método NBI é considerado eficiente para construção de fronteiras de Pareto equiespaçadas, devendo-se ao fato de que o método é independente das escalas relativas das funções otimizadas [83]. Existem alguns passos para a correta execução do método. A figura 2.19 apresenta de forma sucinta o método.

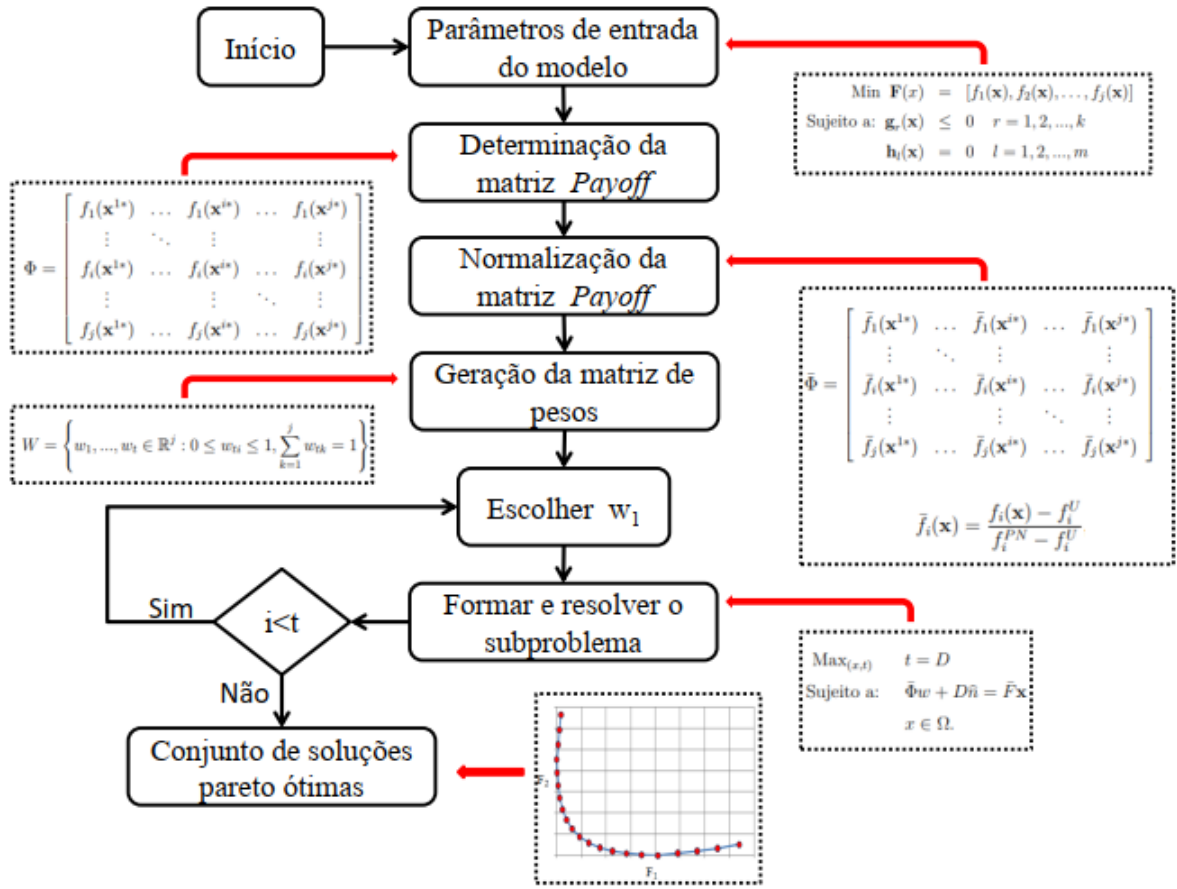


Figura 2.19 – Etapas para execução do método NBI

Fonte: adaptado de [84].

O ponto de partida para a execução do método NBI envolve o cálculo dos elementos da matriz payoff

$$\Phi = \begin{bmatrix} f_1(\mathbf{x}^{1*}) & \dots & f_1(\mathbf{x}^{i*}) & \dots & f_1(\mathbf{x}^{j*}) \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ f_i(\mathbf{x}^{1*}) & \dots & f_i(\mathbf{x}^{i*}) & \dots & f_i(\mathbf{x}^{j*}) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ f_j(\mathbf{x}^{1*}) & \dots & f_j(\mathbf{x}^{i*}) & \dots & f_j(\mathbf{x}^{j*}) \end{bmatrix}. \quad (2.29)$$

Note que a diagonal principal da matriz Φ é formada por todos os pontos de utopia (f_j^U) das funções individuais. Vizando diminuir esforços computacionais quando as funções objetivos possuem escalas muito discrepantes, uma normalização é feita nas funções objetivo segundo a expressão:

$$\bar{f}_i(\mathbf{x}) = \frac{f_i(\mathbf{x}) - f_i^U}{f_i^{PN} - f_i^U}, \quad i = 1, 2, \dots, j. \quad (2.30)$$

A partir disso chega-se a matriz payoff normalizada

$$\bar{\Phi} = \begin{bmatrix} \bar{f}_1(\mathbf{x}^{1*}) & \dots & \bar{f}_1(\mathbf{x}^{i*}) & \dots & \bar{f}_1(\mathbf{x}^{j*}) \\ \vdots & \ddots & \vdots & & \vdots \\ \bar{f}_i(\mathbf{x}^{1*}) & \dots & \bar{f}_i(\mathbf{x}^{i*}) & \dots & \bar{f}_i(\mathbf{x}^{j*}) \\ \vdots & & \vdots & \ddots & \vdots \\ \bar{f}_j(\mathbf{x}^{1*}) & \dots & \bar{f}_j(\mathbf{x}^{i*}) & \dots & \bar{f}_j(\mathbf{x}^{j*}) \end{bmatrix} \quad (2.31)$$

onde a diagonal principal da matriz $\bar{\Phi}$ passa ser nula, uma vez que $\bar{f}_i(\mathbf{x}^{i*}) = 0$.

Segundo [77] as combinações convexas de cada linha da matriz $\bar{\Phi}$ formam a "Envoltória convexa de mínimos individuais" (CHIM - Convex hull of individual minima), também conhecida como linha de utopia ou seja, sendo $w_i = (w_{i1}, \dots, w_{ij}) \in \mathbb{R}^j$ e

$$W = \left\{ w_1, \dots, w_t \in \mathbb{R}^j : 0 \leq w_{ti} \leq 1, \sum_{k=1}^j w_{tk} = 1 \right\} \quad (2.32)$$

em que t é o número de subproblemas de otimização a serem resolvidos, o conjunto dos pontos da envoltória convexa é

$$\bar{\Phi}_w = \{ \bar{\Phi}w_1, \bar{\Phi}w_2, \dots, \bar{\Phi}w_t \}. \quad (2.33)$$

Sendo \hat{n} um vetor normal à linha de utopia nos pontos de $\bar{\Phi}_w$ na direção da origem tem se que $\bar{\Phi}w + D\hat{n}$, $D \in \mathbb{R}$ é a equação vetorial da reta normal a linha de utopia passando pelos pontos de $\bar{\Phi}_w$. O ponto em que $\bar{\Phi}w + D\hat{n}$ intercepta fronteira da região viáveis mais próxima da origem corresponderá a maximização da distância (D) entre a linha de utopia e a fronteira de Pareto, marcando desta forma um ponto de solução ideal nessa fronteira [85, 86]. Logo o método NBI pode ser escrito como um problema de programação não-linear

$$\begin{aligned} \text{Max}_{(x,t)} \quad & t = D \\ \text{Sujeito a:} \quad & \bar{\Phi}w + D\hat{n} = \bar{F}\mathbf{x} \\ & x \in \Omega \end{aligned} \quad (2.34)$$

e para a determinação de ∂F o problema de otimização 2.34 deve ser solucionado para diferentes vetores de peso w , ou seja, para cada vetor de pesos w teremos um subproblema a ser resolvido. De acordo com Das e Dennis [82], quando se possuem mais do que duas funções objetivo a atribuição de pesos segue o diagrama de árvore. Para ilustrar tal diagrama, considere um problema multiobjetivo com três funções objetivos e um espaçamento uniforme entre os pesos de 0,2. Por meio do diagrama de árvore ilustrado na figura

2.20 podemos ver que neste tipo de problema multiobjetivo teríamos 21 subproblemas a serem resolvidos.

Ainda segundo Das e Dennis [82] o número de subproblemas N_s a serem resolvidos em um problema mutiobjetivo obedece a relação

$$N_s = \binom{n+p-1}{p} = \frac{(n+p-1)!}{p!(n-1)!} \tag{2.35}$$

onde n é o número de funções objetivo e $p = \frac{1}{\delta}$, sendo δ o espaçamento uniforme entre dois pesos consecutivos.

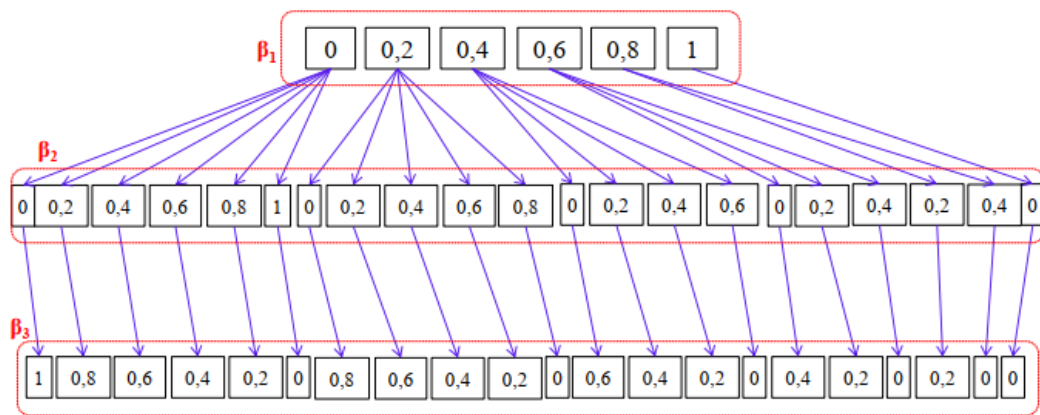


Figura 2.20 – Árvore de pesos para um problema objetivo com três funções objetivo e espaçamento de 0,2 entre os pesos.

Fonte: adaptado de [82].

Comparando as relações 2.35 e 2.9 chegamos a conclusão de que a árvore proposta por Das e Dennis [82] está diretamente relacionado a um arranjo de misturas de n componentes e grau polinomial p (veja equação 2.9). Assim, no caso da árvore representada na figura, a mesma combinação de pesos pode ser obtida através do arranjo simplex lattice $\{3,5\}$, como exibido na figura 2.21. Para os casos em que há duas funções onjetivo, Jia e Iarapetritou

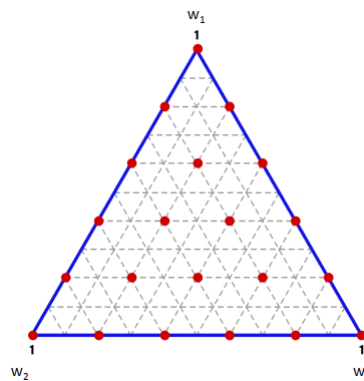


Figura 2.21 – Arranjo simplex lattice $\{3,5\}$.

Fonte: autor.

[86] propõe a utilização da relação $w_j = 1 - \sum_{i=1}^n w_i$ para determinar W . Assim, o problema 2.34 pode ser escrito na forma:

$$\begin{aligned} \text{Min} \quad & \bar{f}_1(\mathbf{x}) \\ \text{Sujeito a:} \quad & \bar{f}_1(\mathbf{x}) - \bar{f}_2(\mathbf{x}) + 2w - 1 = 0 \\ & \mathbf{g}_r(\mathbf{x}) \leq 0 \quad , r = 1, 2, \dots, k \\ & 0 \leq w \leq 1 \end{aligned} \tag{2.36}$$

A figura 2.22 ilustra o método NBI no espaço normalizado para duas funções objetivo.

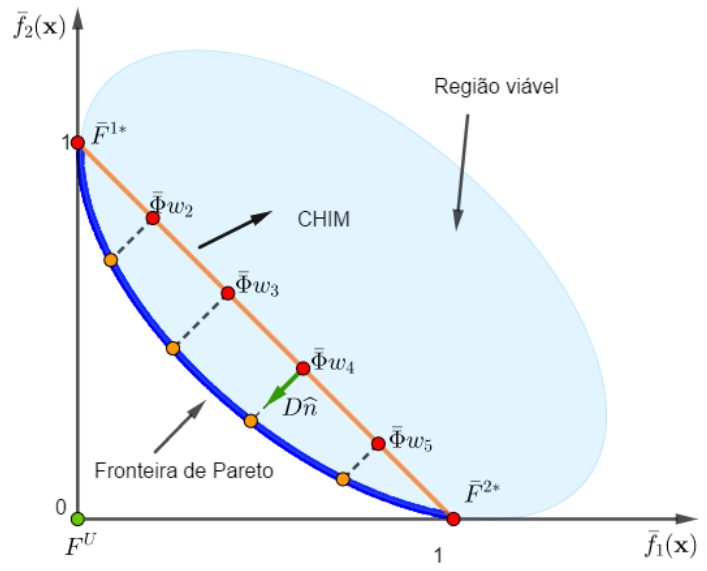


Figura 2.22 – Representação gráfica do NBI para o caso bi-objetivo
 Fonte: adaptado de [76].

Muitas pesquisas tem utilizado o método NBI em problemas de otimização multi-objetivo [77, 87, 88]. Este método tem sido utilizado para determinar uma curva *trade-off* em problemas de otimização multiobjetivo para geração da Fronteira de Pareto e apesar da sua versatilidade, o método NBI pode apresentar falhas quando as respostas possuem correlação forte e positiva, produzindo resultados não condizentes com o processo modelado [89].

2.6 Erro percentual global e entropia

Conforme dito anteriormente, em um processo de otimização multiobjetivo, um grande número de soluções podem ser geradas formando a fronteira de Pareto.

De acordo com Taboada et al. [90] o conjunto de soluções Pareto ótimas inclui todas as decisões racionais, sendo o tomador de decisão responsável pela escolha de uma solução dentre as soluções existentes. No entanto, devido complexidade da atribuição dos pesos para as funções objetivo, os tomadores de decisão tem dificuldade em eleger uma solução. Seguindo neste sentido, diversas técnicas de tomada de decisão multi-objetivo podem ser utilizadas. Dentre elas, podemos destacar a métrica desenvolvida por Rocha et al. [91] que combina o índice de erro percentual global e o índice de entropia de Shanon [92].

A palavra "entropia" surgiu por volta de 1865 pelo alemão Rudolf Clausius para dar um novo nome a perda de calor. A entropia esteve associada ao domínio da Física, até que em 1948 Claude Shannon usou o termo para representar uma medida de informação. A partir daí, esse conceito passou a estar presente em variadas disciplinas como mecânica, estatística, finanças, transporte, planejamento urbano, teoria das filas, teoria da informação e programação linear e não linear [93].

Segundo Fang et al. [94] a entropia está intimamente ligada ao conceito de incerteza e pode ser considerada como uma medida de incerteza probabilística, que assume o máximo valor quando os resultados em uma distribuição de probabilidade são igualmente prováveis. Com isso, a entropia pode ser utilizada como um critério para se avaliar a qualidade das soluções Pareto-ótimas obtidas por um processo de otimização, a partir da máxima diversificação do vetor de pesos e variáveis de decisão.

Para distribuições contínuas de probabilidade, a função densidade é avaliada em todos os valores do argumento. Assim, dada uma distribuição contínua de probabilidade (p) a entropia é definida por:

$$H = - \int_{-\infty}^{\infty} p(x) \ln(x) dx \quad (2.37)$$

onde $\int_{-\infty}^{\infty} p(x) = 1$ e $p(x) \geq 0$. Como os pesos utilizados na ponderação de funções em otimização multiobjetivo são proporções, $p(x)$ segue uma distribuição discreta de probabilidade. Assim, segundo Shannon [92], a entropia passa a ser representada como:

$$H = - \sum_{i=1}^m x_i \ln(x_i) \quad (2.38)$$

sendo que x_i são os pesos atribuídos aos objetivos a serem otimizados.

Entre as muitas propriedades desejáveis do índice de entropia de Shannon, pode se enfatizar:

- (1) A entropia de uma distribuição de probabilidade representando um resultado completamente certo é 0 e a entropia de qualquer distribuição de probabilidade representando resultado incertos é positiva;
- (2) A entropia é côncava.

A utilidade da propriedade (1) reside no fato de garantir a existência de soluções não nulas e a propriedade (2) facilita no processo de otimização, visto que é mais fácil maximizar uma função côncava do que uma não côncava. Podemos visualizar essas duas propriedades por meio da figura 2.23 para o caso de duas possibilidades.

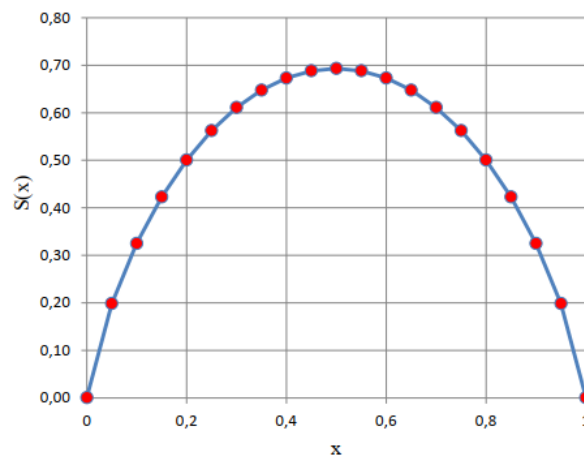


Figura 2.23 – Entropia
Fonte: autor.

Outra estratégia para determinar a ponderação é o Erro Percentual Global (EPG). O EPG calcula o somatório das diferenças das soluções na fronteira de Pareto em relação aos seus alvos. O EPG é calculado segundo a expressão

$$EPG = \sum_{i=1}^j \left| \frac{y_i^* - T_i}{T_i} \right| \quad (2.39)$$

em que y_i^* representa os valores das respostas Pareto-ótimas, T_i são os alvos definidos e j o número de objetivos.

Neste trabalho, para apontar uma solução dentre as soluções Pareto-ótimas, uma medida que utiliza tanto a entropia de Shannon quanto o erro percentual global será utilizada, mais especificamente o critério será o da máxima razão entre entropia e erro percentual global

$$\zeta = \frac{H}{EPG}. \quad (2.40)$$

O valor máximo de ζ corresponderá a maior entropia (e isso acarretará na diversificação dos pesos) e um somatório das diferenças entre as soluções Pareto-ótimas e seus alvos menor.

2.7 Análise Fatorial

Devido a incapacidade do método de otimização desirability e somas ponderadas incorporar as correlações e variâncias entre as respostas, técnicas estatísticas de redução de dimensionalidade do problema baseadas na matriz de variância e covariância das respostas originais tem sido utilizadas. Uma delas é a análise fatorial, técnica que permite reduzir a dimensionalidade do problema, condensando as informações contidas em um conjunto de variáveis originais em um número menor de variáveis (fatores), sempre visando a mínima perda de informações.

Em um processo onde se estuda inúmeras variáveis de respostas pode existir correlação entre essas variáveis, e isso gera a necessidade da utilização de técnicas estatísticas capazes de gerenciar a estruturas de correlação, obtendo uma nova medida que represente essas variáveis correlacionadas. Segundo Jhonson e Wichern [95] a análise fatorial é uma técnica estatística multivariada capaz de descrever variáveis correlacionadas em grupos de fatores ou seja, é uma técnica que descreve a estrutura de covariância entre as variáveis de resposta y_1, y_2, \dots, y_q em termos de poucos fatores F_1, F_2, \dots, F_s ($s < q$). Com essa técnica se reduz a repetição de informação entre as variáveis através de um número menor de variáveis latentes.

O modelo de análise fatorial é descrito segundo a equação

$$\mathbf{X} - \mu = \mathbf{L}\mathbf{F} + \epsilon \quad (2.41)$$

em que $X = [X_1, X_2, \dots, X_q]^T$ é um vetor aleatório observável com q componentes com vetor de médias $\mu = [\mu_1, \mu_2, \dots, \mu_q]^T$, $F = [F_1, F_2, \dots, F_s]^T$ é um vetor aleatório de variáveis não observáveis ($s < q$), L é uma matriz de dimensão ($q \times s$) de coeficientes de cargas fatoriais ou carregamentos. O vetor $\epsilon = [\epsilon_1, \epsilon_2, \dots, \epsilon_q]$ é um vetor de termos de erros conhecidos como fatores específicos. Em termos de sistemas a equação 2.41 fica:

$$\begin{cases} X_1 - \mu_1 = l_{11}F_1 + l_{12}F_2 + \dots + l_{1s}F_s + \epsilon_1 \\ X_2 - \mu_2 = l_{21}F_1 + l_{22}F_2 + \dots + l_{2s}F_s + \epsilon_2 \\ \vdots \\ X_q - \mu_q = l_{q1}F_1 + l_{q2}F_2 + \dots + l_{qs}F_s + \epsilon_q \end{cases} \quad (2.42)$$

É importante destacar que a matriz L pode ser calculada a partir de uma decomposição espectral da matriz Σ

$$\Sigma = \lambda_1 e_1 e_1^T + \lambda_2 e_2 e_2^T + \dots + \lambda_q e_q e_q^T = \begin{bmatrix} \sqrt{\lambda_1} e_1 & \sqrt{\lambda_2} e_2 & \dots & \sqrt{\lambda_q} e_q \end{bmatrix} \begin{bmatrix} \sqrt{\lambda_1} e_1^T \\ \sqrt{\lambda_2} e_2^T \\ \vdots \\ \sqrt{\lambda_q} e_q^T \end{bmatrix} = LL^T. \quad (2.43)$$

Observe que as cargas fatoriais são, portanto, os autovetores da matriz de variância-covariância Σ escalonados por $\sqrt{\lambda_q}$. Dado que $Cov(X) = E[(x - \mu)(x - \mu)^T]$ a equação 2.43 fica:

$$\begin{aligned} \Sigma &= Cov(X) = E[(x - \mu)(x - \mu)^T] = E[(LF + \epsilon)(LF + \epsilon)^T] & (2.44) \\ &= E[LF(L^T F^T) + \epsilon(L^T F^T) + (LF)\epsilon^T + \epsilon\epsilon^T] \\ &= \underbrace{E[FF^T]}_I L^T + \underbrace{E[\epsilon F^T]}_0 L^T + L \underbrace{E[F\epsilon^T]}_0 + \underbrace{E[\epsilon\epsilon^T]}_\Psi \\ &= LL^T + \Psi \end{aligned}$$

onde $\Psi_{q \times q}$ é uma matriz diagonal formada pelas variâncias específicas ψ_i (proporção da variância de X_i explicada pelo fator específico ϵ_i). Logo,

$$\sigma_{ii} = Var(X_i) = (l_{i1}^2 + l_{i2}^2 + \dots + l_{iq}^2) + \psi_i = h_i^2 + \psi_i \quad (2.45)$$

onde h_i^2 recebe o nome de comunalidade (parte da variância de X_i explicada pelos fatores comuns). Sendo $Cov(\epsilon, F) = 0$ segue que

$$Cov(X, F) = E[(X - \mu)F^T] = E[(LF + \epsilon)F^T] = E[LFF^T] + E[\epsilon^T] = L \quad (2.46)$$

e conseqüentemente que $Cov(X_i, X_k) = l_{i1}l_{k1} + \dots + l_{im}l_{km}$ e $Cov(X_i, F_j) = l_{ij}$.

Como se pode notar, escrevemos as variáveis X_q em função das variáveis latentes F_s ($s < q$). Uma questão chave dentro da análise fatorial é justamente determinar quantas serão as variáveis, ou seja, determinar s . Jolliffe [96] sugeriu o uso da regra de Kaiser [97] para determinar s . Esta regra seleciona apenas as componentes principais que possuem autovalores maiores do que um (utilizando a matriz de correlação), e uma variância cumulativa de no mínimo 80%.

Considerando uma matriz de dados patronizados Z de X e a matriz de carregamentos L , a matriz F dos escores dos fatores (não correlacionados) é dada por

$$F = Z [L(L^T L)^{-1}]. \quad (2.47)$$

Uma vantagem apresentada por esta técnica comparada à análise de componentes principais diz respeito a rotação da matriz L , o que pode acarretar na separação das variáveis em grupos onde a interpretação fica facilitada. Essa rotação pode ser dada em várias direções por meio de um transformação ortogonal, preservando a comunalidade e variâncias específicas. A matriz rotacionada de carregamentos pode ser calculada como $\hat{L} = LB$, onde B é uma matriz ortogonal ($BT^T = B^T B = I$) [95].

Vários são os métodos para rotacionar a matriz L . Estes métodos podem ser ortogonais e oblíquos. Os ortogonais produzem fatores não correlacionados e os oblíquos fatores correlacionados. Para os métodos ortogonais damos atenção especial ao método varimax,

que maximiza a soma das variâncias das cargas fatoriais da matriz de carregamentos fatoriais.

Caso o conjunto de funções objetivo apresente respostas correlacionadas, a análise fatorial com escores extraídos por componentes principais pode ser utilizada, aplicando uma rotação varimax para obter eixos independentes obtendo escores fatoriais para substituir as respostas originais sem perda de informações e sem eliminar as respostas correlacionadas.

Dessa forma, neste trabalho, os escores de fator serão utilizados para substituir as funções objetivos em um problema de otimização multiobjetivo resolvido pelo método NBI. Assim, se o problema original possui j funções objetivos, após a redução de dimensionalidade e agrupamento das funções objetivo originais o problema possuirá l funções objetivos de escores rotacionados de fator e a Fronteira (ou superfície) de Pareto poderá ser obtida para este grupo de funções.

2.8 Considerações finais sobre o capítulo

Este capítulo teve o objetivo de apresentar o suporte teórico necessário para o desenvolvimento do presente trabalho. A partir da discussão de conceitos relacionados ao ensemble de redes neurais, chegou-se a conclusão que a utilização de redes neurais artificiais são amplamente utilizadas em problemas de previsão de demanda de energia elétrica de curto prazo, principalmente pelo caráter não linear das séries temporais de carga elétrica.

Quanto às ferramentas de análise e otimização, foram apresentados os conceitos de um DOE, caracterizando esta como uma técnica importante no estudo científico dos problemas de engenharia. Entre seus projetos experimentais, discutiu-se os experimentos com misturas, que é uma ferramenta importante no estudo de processos envolvendo combinação de componentes e será utilizado para a determinar como será realizado será obtido a saída do ensemble de redes neurais. Em seguida, apresentou-se algumas ferramentas de otimização de múltiplas respostas, bem como as vantagens e desvantagens. Como critério de escolha de uma solução, foi apresentado a máxima razão entre entropia e erro percentual global. Por fim, de grande importância neste trabalho, a técnica de Análise fatorial também foi abordada e será importante neste trabalho para reduzir a dimensionalidade do problema, ou seja, por meio desta técnica o número de funções objetivo a serem otimizadas será reduzido.

3 PARAMETRIZAÇÃO DAS REDES NEURAIS

Este capítulo tem o objetivo de descrever o planejamento experimental utilizado neste trabalho para obter uma parametrização para as redes neurais que irão compor o ensemble. A definição do método de pesquisa a ser seguido é importante, pois assim o processo estudado é tratado de forma sistemática, contribuindo para obtenção de informações que sejam objetivas e de relevância. Desta forma, o problema de pesquisa é exposto novamente para que em seguida o método experimental seja definido. Com isso, o processo experimental começa a ser descrito através da definição dos parâmetros, fixação dos níveis de trabalho, escolha das respostas de interesse, definição da matriz experimental e a explicação de como foi feito o registro das repostas.

3.1 Estratégia utilizada

Este trabalho utilizou a experimentação para obter uma parametrização das redes neurais. Seguindo as etapas descritas por Coleman e Montgomery [62] na figura 2.7 na seção 2.4.4 a experimentação realizada pode ser descrita segundo a figura 3.1.

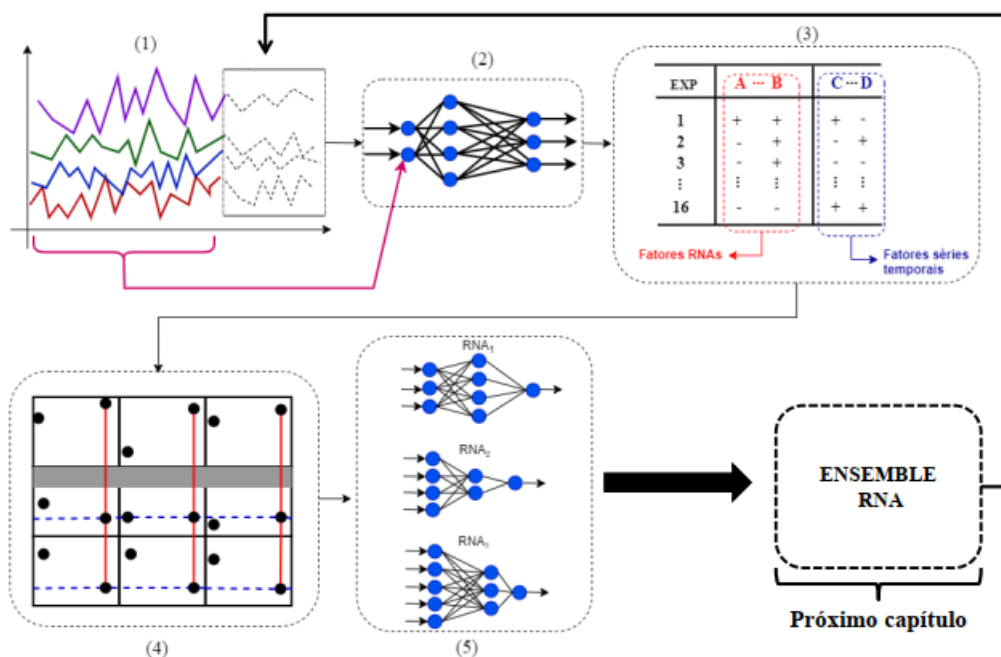


Figura 3.1 – Etapas para a obtenção da parametrização das redes neurais
Fonte: autor.

Portanto podemos dividir o planejamento experimental em algumas etapas, discutidas de forma separada:

- (a) Os passos (1) e (2) na figura representam a a seleção dos fatores e a fixação dos níveis escolhidos no trabalho. Esses fatores estão relacionados à parametrização das redes neurais e incluem fatores das séries temporais de energia elétrica das quatro subestações e fatores que são exclusivos do software utilizado para obter as redes neurais;
- (b) A etapa (3) está relacionado a definição do arranjo experimental utilizado para registro das respostas de interesse;
- (c) A etapa (4) refere-se à análise dos resultados obtidos através da execução dos experimentos segundo a matriz experimental definida na etapa (3). Neste momento, através do método desirability uma parametrização é determinada para a seleção das redes neurais que irão compor o ensemble (etapa 5).

A próxima etapa, que será abordada no capítulo seguinte, refere-se ao ensemble de redes neurais artificiais que será proposto com as redes neurais (5) obtidas após a parametrização.

Discutiremos a seguir de forma detalhada toda a execução do experimento que será utilizado para obter uma parametrização para as redes neurais.

3.2 Séries temporais de energia elétrica das subestações

As séries temporais de energia elétrica das quatro subestações de distribuição de energia elétrica que serão utilizadas como entrada das redes neurais estão dispostos na figura 3.2. Esses dados representam as medições de energia elétrica em (KWh) para as quatro subestações coletadas durante o período de um mês, observadas a cada hora.

Associado a essas medições, temos o dia da semana e o período do dia em que a medida foi observada. Os dias da semana foram codificados de 1=domingo a 7=sábado e o período do dia foi codificado de 1 a 4, correspondendo manhã, tarde, noite e madrugada, respectivamente. Essa classificação se deu de acordo com o seguinte critério:

- Manhã: medições de 06:00 até 11:59;
- Tarde: medições de 12:00 até 18:59;
- Noite: medições de 19:00 até 23:59;
- Madrugada: medições de 00:00 até 05:59.

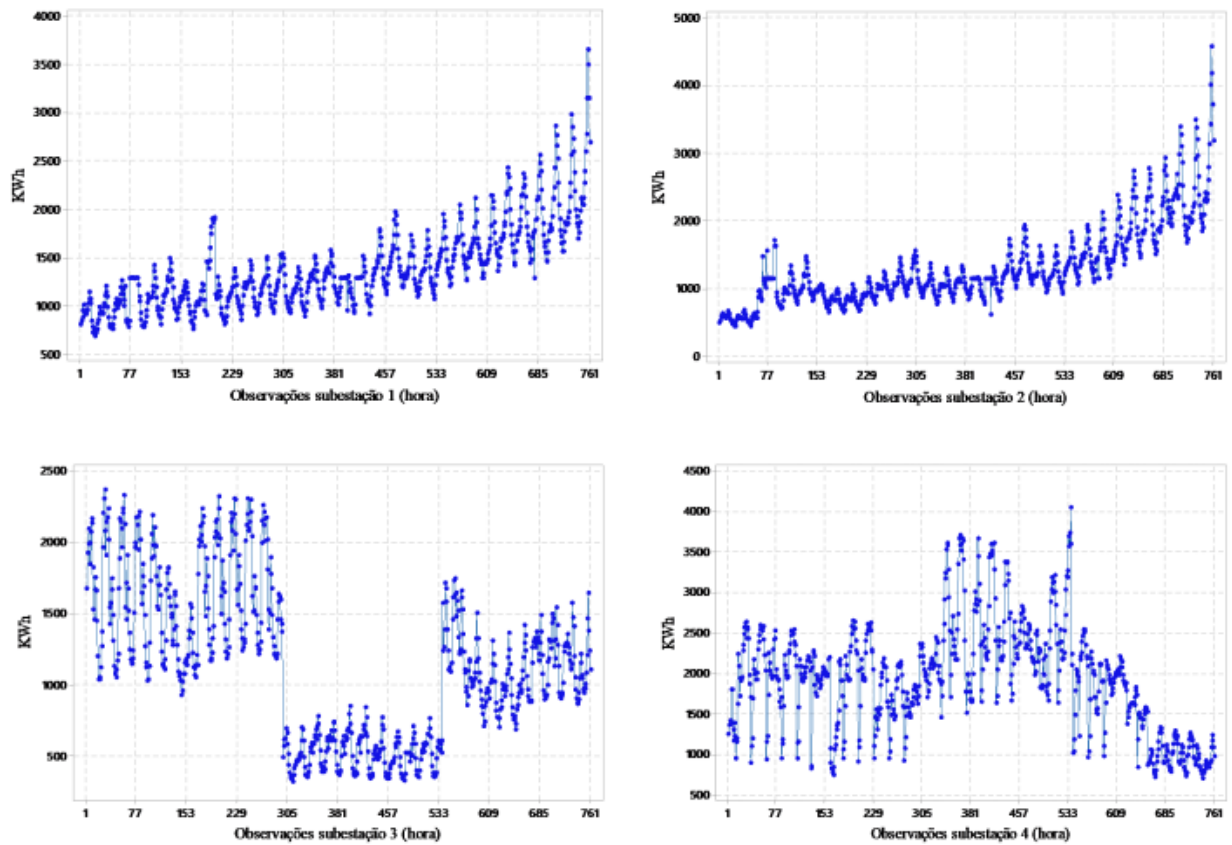


Figura 3.2 – Séries temporais de energia elétrica das 4 subestações
 Fonte: autor

A tabela 3.1 apresenta de forma resumida os dados que serão utilizados como entradas para treinamento das redes neurais.

Tabela 3.1 – Entradas das RNAs

Medição	Dia	Período	Subestação 1 (KWh)	Subestação 2 (KWh)	Subestação 3 (KWh)	Subestação 4 (KWh)
1	2	2	795,74	490,25	1678,56	1244,58
2	2	2	831,67	523,27	1927,98	1352,83
3	2	2	875,62	564,67	1987,34	1360,31
⋮	⋮	⋮	⋮	⋮	⋮	⋮
760	5	4	3151,13	3717,79	1236,6	1081,96
761	5	4	2683,05	3170,38	1102,42	972,22

A próxima subseção discutirá o planejamento pré-experimental, onde discutiremos os fatores e níveis que foram escolhidos neste estudo.

3.2.1 Planejamento Pré-Experimental

O planejamento pré-experimental começa com a definição do tipo de problema a ser resolvido. Reconhecer e definir o problema corretamente é fundamental para alcançar os objetivos desejados. Neste estágio do trabalho, o interesse é encontrar uma parametrização para obter as redes neurais que irão compor o ensemble para que posteriormente a previsão de demanda de carga elétrica das subestações possa ser realizada.

Definido o problema a ser atacado, é necessário determinar o método que será usado para resolver o problema. Pelas várias razões já listadas, usaremos o ensemble de redes neurais para previsão. Neste trabalho foi utilizado o solucionador inteligente de problemas (Intelligent Problem Solver - IPS) do software Statistica em sua versão 7.0 para executar o treinamento da RNA automaticamente, indicando o tipo de rede neural que melhor se ajusta ao problema. O IPS é uma ferramenta que procura, de forma automática, a arquitetura de rede ideal.

Inicialmente o usuário determina qual será a regra de parada, podendo ser por meio da quantidade de redes testadas ou pelo tempo (em horas/minutos) que o algoritmo será executado. Também é possível selecionar a quantidade de redes a serem retidas ao final do treinamento. Neste trabalho, optou-se por reter três redes e treiná-las durante 3 minutos em cada experimento.

Em relação aos tipos de redes para o teste, optou-se por utilizar todos os tipos disponíveis no software: linear, PNN, RBF e a MLP (perceptrons de 3 e 4 camadas). Com a definição dos tipos de RNAs, o próximo passo é determinar o intervalo relacionado a quantidade de neurônios da camada oculta das redes RBF, MLP de três e quatro camadas. Optou-se por utilizar o intervalo padrão do software, que estabelece que o número de neurônios na camada oculta da rede RBF é no mínimo 1 e no máximo 190, e das redes MLP (de três e quatro camadas) é no mínimo 1 e no máximo 10. Por fim, existem duas formas de codificação da resposta no caso das redes perceptrons multicamadas: a linear, que utiliza a função de ativação identidade, enquanto que a logística utiliza a função de ativação logística. Ambas as abordagens foram selecionadas. É importante ressaltar que outras opções de configuração das redes neurais foram analisadas neste trabalho, sendo consideradas como fatores para determinação da matriz experimental. Essas opções serão descritas de forma mais detalhada a seguir.

Após o método ser previamente determinado, definiu-se uma métrica de desempenho para avaliar o método. Várias medidas de erro, como o erro médio quadrático (MSE) e o erro percentual absoluto médio (MAPE) podem ser usadas como medidas de desempenho para a previsão de séries temporais. Para um estudo específico, informações adicionais podem ser encontradas em [98]. Neste estudo, o MAPE foi escolhido como uma

métrica de desempenho, sendo definido como:

$$MAPE(\%) = \frac{1}{T} \sum_{i=1}^T \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100 \quad (3.1)$$

em que y_i é a observação atual no instante i , \hat{y}_i é o valor previsto e T é o número de previsões.

Para finalizar o planejamento pré-experimental, a próxima etapa está relacionada à escolha dos fatores e seus níveis que irão compor a matriz experimental. Foram escolhidos 6 fatores nas quais dois fatores estão relacionados às séries temporais de energia exibidas na figura 3.2 e os demais são inerentes à configuração das redes neurais. Os fatores são:

- *Pré-processamento de dados (PD)*. O PD está relacionado à análise e transformação das medidas de energia horária das quatro subestações mostradas na tabela 3.1, afim de se detectar relacionamentos e tendências importantes e para auxiliar as redes neurais no aprendizado de importantes padrões. Para esse fator, foram utilizadas duas formas de pré-processamento: transformação linear (LI) em que os dados são incluídos no intervalo $[0, 1]$ e normalização estatística (Z), definidas respectivamente por

$$l_i = \frac{y_i - y_{min}}{y_{max} - y_{min}} \quad (3.2)$$

e

$$z_i = \frac{y_i - \bar{y}}{s} \quad (3.3)$$

onde y_{max} , y_{min} , \bar{y} e s denotam respectivamente os valores máximo, mínimo, médio e desvio padrão das observações.

- *Variáveis Explicativas (EV)*. De acordo com a tabela 3.1, o dia e o período de medição são classificados como variáveis explicativas. Para fins experimentais, foram escolhidos dois níveis para esse fator: uso (Y) e não uso (N) dessas variáveis, ou seja, usar ou não usar essas variáveis como entradas da RNAs.
- *Critério de seleção das redes a serem retidas (CS)*. Neste estudo, temos interesse em combinar três redes neurais para a previsão de energia elétrica de curto prazo. Em cada execução experimental três redes neurais foram retidas e a retenção é baseada em dois critérios: menor erro (LE) (no subconjunto da seleção) e erro equilibrado (BE) em relação à diversidade. No LE, as redes são retidas baseadas no menor erro de seleção. Para o caso BE, as redes são retidas equilibrando o erro com tipos e diversidade. A escolha dessa opção preserva um conjunto de redes com uma variedade de tipos de desempenho e complexidade.
- *Seleção de um subconjunto de variáveis independentes (IV)*. Esta opção determina que o IPS selecione quais variáveis devem ser usadas como entradas. As variáveis

são escolhidas dentre as variáveis independentes. Cada rede testada pode usar uma combinação diferente de entradas. Esse fator foi utilizado de duas maneiras: permitir que seja feita a seleção (Y) e não permitir (N).

- *Seleção de subconjuntos para cada rede treinada (SS)*. Utilizamos duas opções para esse fator: fixo (F) e reamostragem (RS) (atribuídas aleatoriamente a cada rede). Ao selecionar a primeira opção, os subconjuntos de treinamento, seleção e teste são atribuídos aleatoriamente no início da execução aos números especificados nos campos de treinamento, seleção e teste. Para a segunda opção, os subconjuntos são atribuídos aleatoriamente aos números especificados nos campos de treinamento, seleção e teste cada vez que uma rede é criada e testada. Por padrão, a proporção usada para os subconjuntos de treinamento, seleção e teste é 2:1:1.
- *Número de etapas de tempo usadas como entradas (R)*. Neste estudo, as redes neurais são utilizadas para análise de séries temporais. Como tal, as informações atrasadas para o processo autoregressivo são preservadas. Dessa forma, a entrada será uma amostra da série temporal $(y_{t-1}, y_{t-2}, \dots, y_{t-k})$ considerando um atraso de k , e a resposta será a observação um passo à frente (y_t). Para esse fator, o software exige a fixação de um número mínimo e um máximo.

Usamos dois métodos para definir o número de etapas de tempo usadas como entradas. O primeiro é o padrão (D), onde o software fixa os valores mínimo e máximo como sendo 1 e 10 respectivamente. Para este caso, um valor é selecionado entre esses números e então esse valor é determinado como o número de etapas de tempo usadas como entrada. A segunda maneira é usar a Transformada rápida de Fourier (FFT) (figura 3.3) para determinar a sazonalidade da série temporal de carga elétrica e conseqüentemente a definição dos limites mínimo e máximo. A figura 3.3 mostra claramente a sazonalidade presente nas séries temporais de carga elétrica das subestações. Notavelmente, há um pico proeminente na densidade espectral no período 24 em todas as subestações, indicando que há um período sazonal de 24 horas. Através desta análise definiremos o mínimo e máximo na janela de seleção igual a 24. Sendo assim, as duas situações consideradas para análise deste fator são: padrão (mínimo e máximo sendo 1 e 10 respectivamente) e FFT (mínimo e máximo iguais a 24).

Vale lembrar que os fatores *PD* e *EV* são relacionados às séries temporais e os demais relacionados à parametrização das redes neurais no software

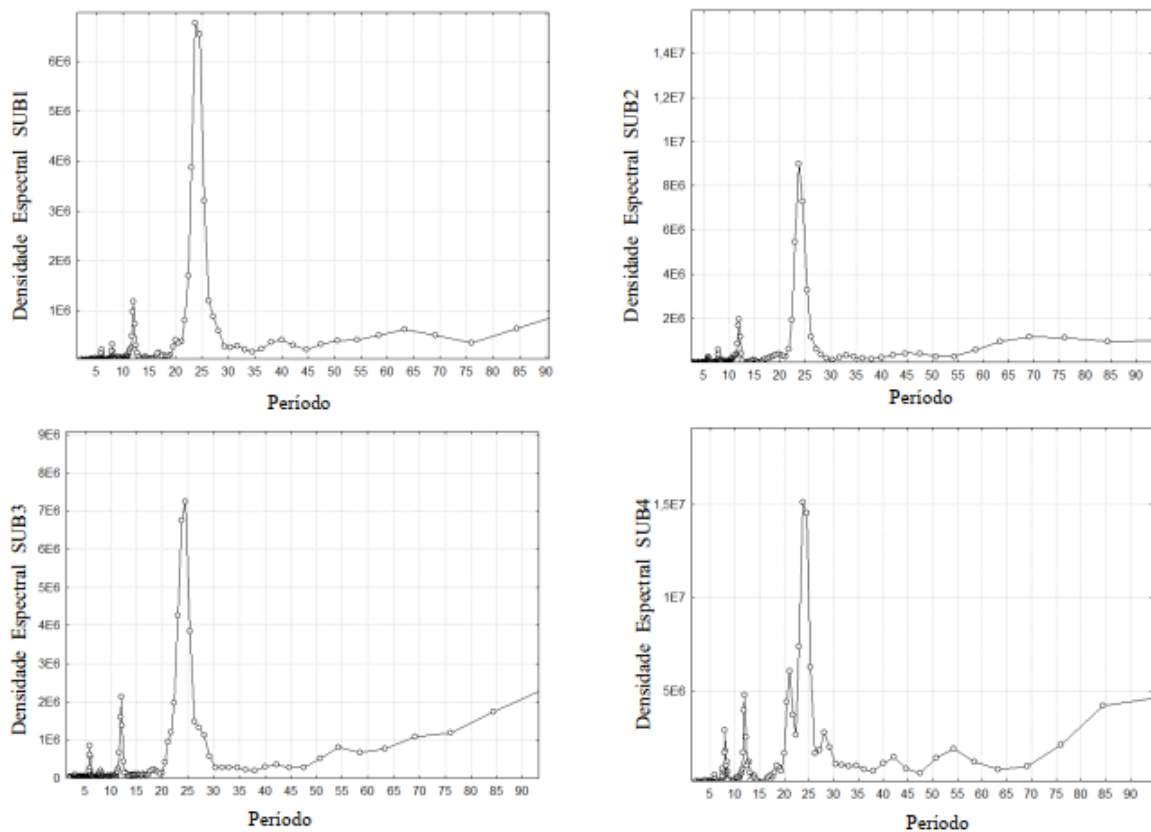


Figura 3.3 – Análise espectral

Fonte: autor.

Uma vez definidos os fatores e níveis escolhidos, a próxima etapa consiste em definir o arranjo experimental. A próxima subseção abordará a escolha do arranjo.

3.2.2 Definição do arranjo experimental e execução dos experimentos

Após definir os fatores e níveis escolhidos, precisamos definir o arranjo experimental. Aqui, o termo arranjo indica uma matriz na qual as colunas representam fatores de entrada e as linhas representam combinações desses fatores. O arranjo experimental geralmente é escolhido com base no número de fatores e níveis, estrutura e resolução do experimento, quantidade de tempo e recursos para realizar o experimento, etc. Existem programas (Minitab, Statistica, SPSS, JMP, Matlab) que oferecem uma biblioteca de arranjos clássicos predeterminados.

Neste estudo, usamos um planejamento fatorial fracionário que, de acordo com [56], é usado quando é necessário estudar os efeitos de vários fatores de entrada simultaneamente. Tal situação pode impossibilitar o uso do fatorial completo, pois quanto maior o número de fatores analisados, maior o número de experimentos necessários. Segundo o mesmo autor, é possível, realizando apenas parte dos experimentos, analisar os efeitos em uma resposta de interesse. Conforme exposto anteriormente, seis fatores com dois níveis cada foram definidos neste estudo, o que resultaria em um total de $2^6 = 64$ experimentos.

Então é razoável supor que informações sobre os principais efeitos e interações de baixa ordem podem ser obtidas considerando-se apenas uma fração do experimento completo. A principal vantagem de usar uma fração do total de experimentos é que informações importantes (como quais fatores são importantes e suas interações com outros fatores) podem ser extraídas usando um pequeno número de experimentos. Um resumo de alguns projetos extraídos do software Minitab é apresentado no quadro 3.1.

Uma determinada resolução Z aponta que um efeito principal de um fator poderá estar confundido com uma interação de ordem $Z - 1$. Desta forma, os experimentos com resolução III (destacados em vermelho no Quadro 3.1) perdem informações relativas a essas interações, já que um efeito principal pode ser confundido com uma interação de baixa ordem, o que em muitos casos não deve ser omitida. Isso não ocorre nos experimentos de resolução igual o superior a IV (destacados em amarelo e verde), tornando estes experimentos com menor risco. Neste trabalho, o arranjo experimental escolhido foi o arranjo fracionário de 6 fatores (com dois níveis cada) de resolução IV, resultando em um total de $2^{6-2} = 16$ experimentos. Com isso, temos a matriz experimental exibida na tabela 3.2.

Após definir a matriz experimental, o próximo passo é realizar os experimentos de acordo com as entradas da RNA e a matriz experimental, seguindo rigorosamente os níveis dos fatores determinados em cada experimento. As tabelas 3.3 e 3.4 exibem os resultados em termos de MAPE das três redes retidas para os experimentos realizados com os fatores e níveis escolhidos segundo a tabela 3.2.

Quadro 3.1 – Projetos experimentais e resoluções disponíveis no software Minitab

	Fatores													
N	2	3	4	5	6	7	8	9	10	11	12	13	14	15
4	Full	III												
8		Full	IV	III	III	III								
16			Full	V	IV	IV	IV	III	III	III	III	III	III	III
32				Full	VI	IV	IV	IV	IV	IV	IV	IV	IV	IV
64					Full	VII	V	IV	IV	IV	IV	IV	IV	IV
128						Full	VIII	VI	V	V	IV	IV	IV	IV

Tabela 3.2 – Arranjo experimental 2_{IV}^{6-2}

Exp.	Fatores					
	PD	EV	CS	IV	SS	R
1	LI	N	LE	N	F	D
2	Z	N	LE	N	RS	D
3	LI	Y	LE	N	RS	FFT
4	Z	Y	LE	N	F	FFT
5	LI	N	BE	N	RS	FFT
6	Z	N	BE	N	F	FFT
7	LI	Y	BE	N	F	D
8	Z	Y	BE	N	RS	D
9	LI	N	LE	Y	F	FFT
10	Z	N	LE	Y	RS	FFT
11	LI	Y	LE	Y	RS	D
12	Z	Y	LE	Y	F	D
13	LI	N	BE	Y	RS	D
14	Z	N	BE	Y	F	D
15	LI	Y	BE	Y	F	FFT
16	Z	Y	BE	Y	RS	FFT

Tabela 3.3 – MAPE subestações 1 e 2

Exp.	Subestação 1			Subestação 2		
	RNA ₁	RNA ₂	RNA ₃	RNA ₁	RNA ₂	RNA ₃
1	0,044	0,044	0,04	0,085	0,076	0,058
2	0,051	0,051	0,052	0,086	0,085	0,088
3	0,051	0,044	0,045	0,076	0,077	0,066
4	0,043	0,053	0,053	0,054	0,084	0,075
5	0,033	0,041	0,061	0,047	0,052	0,086
6	0,039	0,041	0,061	0,051	0,048	0,093
7	0,041	0,038	0,051	0,055	0,051	0,079
8	0,041	0,038	0,043	0,055	0,051	0,061
9	0,039	0,05	0,048	0,047	0,082	0,077
10	0,039	0,055	0,059	0,051	0,084	0,086
11	0,055	0,04	0,054	0,086	0,052	0,086
12	0,043	0,048	0,046	0,057	0,083	0,081
13	0,053	0,037	0,056	0,074	0,05	0,085
14	0,054	0,039	0,044	0,07	0,054	0,076
15	0,065	0,036	0,06	0,081	0,047	0,074
16	0,046	0,056	0,034	0,059	0,077	0,043

Tabela 3.4 – MAPE subestações 3 e 4

Exp.	Subestação 3			Subestação 4		
	RNA ₁	RNA ₂	RNA ₃	RNA ₁	RNA ₂	RNA ₃
1	0,089	0,083	0,072	0,097	0,088	0,088
2	0,083	0,085	0,081	0,096	0,089	0,094
3	0,062	0,07	0,064	0,065	0,082	0,074
4	0,064	0,06	0,058	0,075	0,068	0,071
5	0,06	0,076	0,093	0,066	0,092	0,098
6	0,054	0,073	0,089	0,07	0,086	0,088
7	0,073	0,077	0,073	0,09	0,087	0,082
8	0,069	0,078	0,074	0,076	0,092	0,084
9	0,072	0,068	0,067	0,086	0,083	0,083
10	0,077	0,091	0,09	0,092	0,092	0,092
11	0,069	0,08	0,067	0,106	0,054	0,034
12	0,065	0,07	0,065	0,082	0,086	0,085
13	0,075	0,075	0,081	0,086	0,089	0,088
14	0,078	0,076	0,081	0,096	0,085	0,088
15	0,064	0,068	0,061	0,103	0,078	0,068
16	0,068	0,063	0,071	0,077	0,075	0,077

Múltiplas respostas são comuns em experimentos projetados e de acordo com Balestrassi et al. [4] média e desvio padrão (DP), são frequentemente usados como variáveis de resposta e a otimização simultânea de várias variáveis de resposta envolve funções de desejabilidade. Utilizamos a média e o DP da medida de performance MAPE como respostas de interesse neste estudo, sendo exibidas na tabela 3.5.

Tabela 3.5 – Respostas de interesse em cada experimento

Exp.	Média MAPE	DP MAPE	Exp.	Média MAPE	DP MAPE
1	0,072	0,02	9	0,067	0,017
2	0,078	0,017	10	0,076	0,019
3	0,065	0,013	11	0,065	0,021
4	0,063	0,012	12	0,068	0,016
5	0,067	0,022	13	0,069	0,018
6	0,066	0,02	14	0,069	0,019
7	0,066	0,018	15	0,067	0,018
8	0,063	0,018	16	0,062	0,015

A próxima subseção apresenta uma análise dos resultados obtidos nos experimentos.

3.2.3 Análise dos resultados experimentais para a resposta média MAPE

Diante do que já foi exposto, com o objetivo de se determinar uma parametrização da rede neural a partir dos fatores e níveis definidos, é necessário identificar quais fatores de

fato influenciam nas duas respostas de interesse definidas anteriormente. Para realizar esta análise, utilizou-se o software Minitab® versão 18 para análise do experimento planejado.

A análise de variância realizada para a resposta média MAPE pode ser observada na tabela 3.6. Por meio desta análise podemos observar que os fatores que possuem *p*-valor inferior a 0,05 exercem influência significativa sobre a resposta. Assim, podemos observar que as variáveis explicativas (EV), a forma de retenção das redes (CS) e o número de etapas de tempo usadas como entradas na rede são significativos, bem como as interações entre a forma de padronizar os dados (PD) e as variáveis explicativas, o critério de retenção das redes e a seleção de subconjuntos para cada rede treinada exercem influência sobre a resposta. Como o fator PD e SS não são significativos, mas interações em que eles participam são, podemos dizer que estes dois fatores também influenciam a resposta analisada. Através do gráfico de Pareto exibido na figura 3.4, podemos indicar de forma mais visual os fatores e as interações impactam na resposta, sendo aqueles que ultrapassam a linha vertical pontilhada. Para a análise do experimento utilizou-se a técnica de eliminação *backward* visando eliminar fatores que não contribuem significativamente para o modelo, com alfa fixado em 0,05, isto é, caso o *p*-valor de um determinado fator ou interação for superior a 0,05, ele é eliminado do modelo. Assim, o valor de 2,365 representa o valor crítico associado à distribuição *t*. Com isso, fatores cujos efeitos padronizados são maiores que este valor são considerados significativos.

Vale ressaltar que devido a existência de interações significativas com participação de fatores não significativos o princípio da hierarquia inclui as interações de ordem inferior no modelo, ou seja, devido a existência de uma interação de segunda ordem significativa as interações de ordem inferior associadas aos fatores deverão constar no modelo. Esse fato justifica a presença de fatores não significativos (PD e SS) no modelo que representa a resposta (equação 3.4).

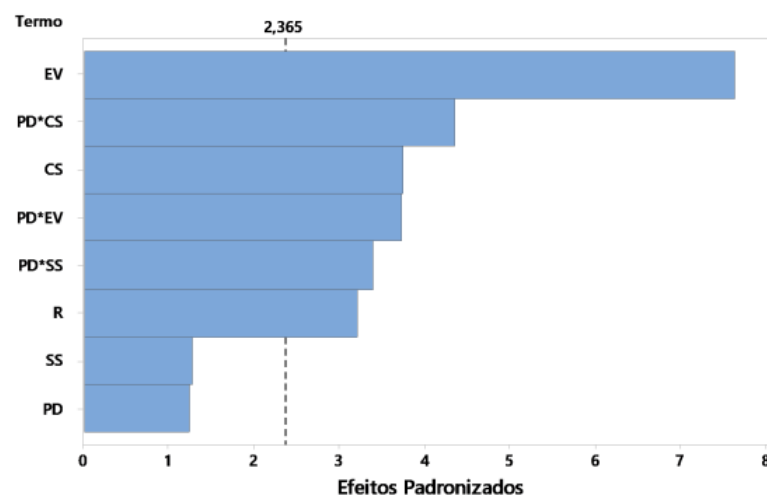


Figura 3.4 – Gráfico de Pareto para a resposta Média MAPE
Fonte: autor.

Tabela 3.6 – ANOVA para a resposta Média MAPE

Fonte	GL	SQ (A.j)	QM (A.j)	Valor F	p-valor
Modelo	8	0,000278	0,000035	16,3	0,001
Linear	5	0,000183	0,000037	17,19	0,001
PD	1	0,000003	0,000003	1,55	0,253
EV	1	0,000124	0,000124	58,41	0,000
CS	1	0,00003	0,00003	14	0,007
SS	1	0,000003	0,000003	1,63	0,242
R	1	0,000022	0,000022	10,34	0,015
Interações de 2 fatores	3	0,000095	0,000032	14,83	0,002
PD*EV	1	0,00003	0,00003	13,98	0,007
PD*CS	1	0,00004	0,00004	18,99	0,003
PD*SS	1	0,000025	0,000025	11,52	0,012
Error	7	0,000015	0,000002		
Total	15	0,000293			

$$\begin{aligned}
\text{Média MAPE} &= 0,067797 + 0,000454(PD) - 0,002789(EV) - 0,001365(CS) \\
&+ 0,000466(SS) - 0,001174R - 0,001364(PD * EV) \\
&- 0,001590(PD * CS) + 0,001239(PD * SS).
\end{aligned} \tag{3.4}$$

Uma das formas de avaliar se o modelo criado realmente representa o conjunto de dados que lhe deu origem é por meio do indicador denominado coeficiente de determinação (R^2). Quando maior o valor de R^2 maior o indício de que o modelo se ajusta bem aos dados indicando o percentual da variação nos dados que é explicado pelo modelo. Seu cálculo pode ser obtido por meio da equação 3.5 [56]

$$R^2 = 1 - \frac{\sum_{t=1}^T e_t^2}{\sum_{t=1}^T (y_t - \bar{y})^2}. \tag{3.5}$$

Entretanto, como a soma de resíduos reduz significativamente quando novos termos são adicionados ao modelo, então ter esse indicador como única medida para o desempenho pode levar a ocorrência de *overfitting* fazendo com que o modelo perca sua capacidade de generalização. Uma alternativa a esse indicador é o R^2 ajustado já que seu valor torna-se cada vez menor a medida que mais termos não significativos vão sendo acrescentados ao modelo. O valor do R^2 ajustado pode ser obtido por meio da equação 3.6 sendo p o número de termos do modelo

$$R^2 \text{ ajustado} = 1 - \frac{\frac{\sum_{t=1}^T e_t^2}{T - p}}{\frac{\sum_{t=1}^T (y_t - \bar{y})^2}{T - 1}}. \tag{3.6}$$

Para a resposta média MAPE esses valores foram de 98,30% e 98,41% respectivamente. Além desses indicadores é importante avaliar a normalidade dos resíduos gerados pelo

modelo, o que é um indicativo de que o modelo é capaz de representar bem os dados. Assim, realizou-se o teste de normalidade dos resíduos (figura 3.5) onde podemos constatar por meio do p-valor (0,989) que a hipótese nula relacionada à normalidade dos resíduos não deve ser rejeitada.

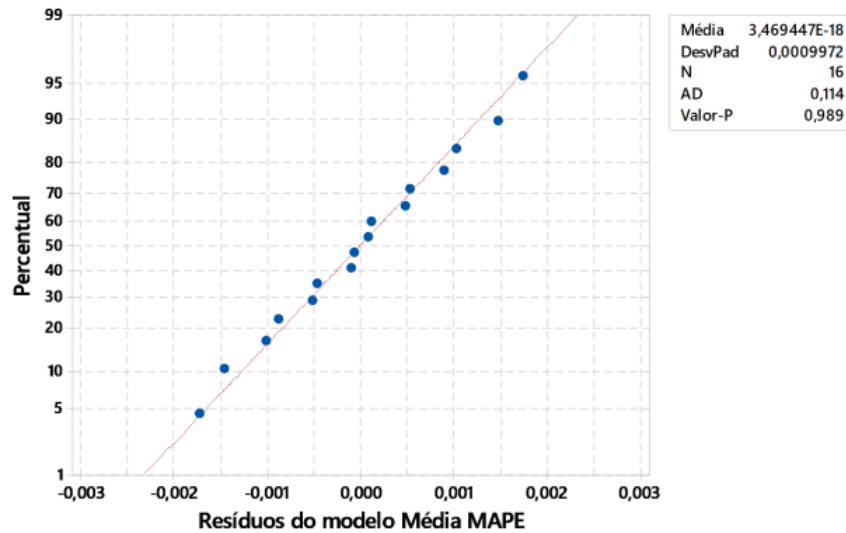


Figura 3.5 – Gráfico de Probabilidade para os resíduos para o modelo Média MAPE
 Fonte: autor.

Além de determinação de quais fatores e interações afetam a resposta é necessário entender como ocorre a influência. O gráfico de efeitos principais exibido na figura 3.6 auxilia nesse entendimento. Observe que dentre os efeitos considerados significativos pela

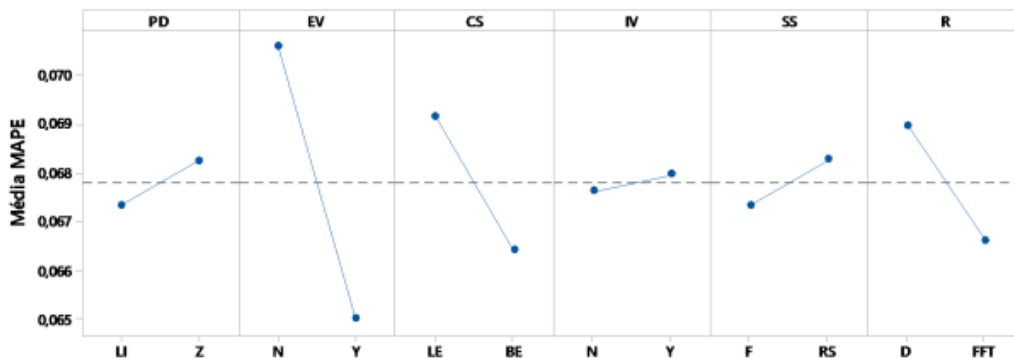


Figura 3.6 – Gráfico de efeitos principais para a resposta Média MAPE
 Fonte: autor.

ANOVA na tabela 3.6 o uso de variáveis explicativas proporciona melhor desempenho. Além disso, a rede também apresenta melhor desempenho quando-se usa o erro equilibrado como forma de seleção de redes a serem retidas. Utilizar a transformada rápida da Fourier para determinação da sazonalidade da série e consequentemente o número de etapas de tempo usadas como entrada também foi importante para a melhoria da resposta. Os

demais fatores, quando considerados de forma individual, possuem efeito muito pequeno na variação da média, o que ilustra o fato de não serem significativas. Entretanto, mesmo que os fatores individualmente não exerçam nenhuma influência sobre o desempenho do método, suas interações podem ser significativas; é o que mostra o gráfico de interações, exibido na figura 3.7.

Observe que o fator PD não impacta a resposta diretamente, mas quando analisado em conjunto com os fatores EV, CS e SS ocorre influência na resposta. Nota-se que a resposta apresenta melhor resultado quando a forma de pré-processamento de dados é utilizada com o uso de variáveis explicativas, ou seja, o uso de variáveis explicativas atrelado à forma de como os dados serão utilizados como entradas na rede melhora a resposta. Com relação ao critério de seleção de redes a serem retidas, quando se utiliza o erro equilibrado como forma de seleção a performance também é melhorada. O mesmo acontece quando a seleção de subconjuntos para cada rede treinada é fixada (F).

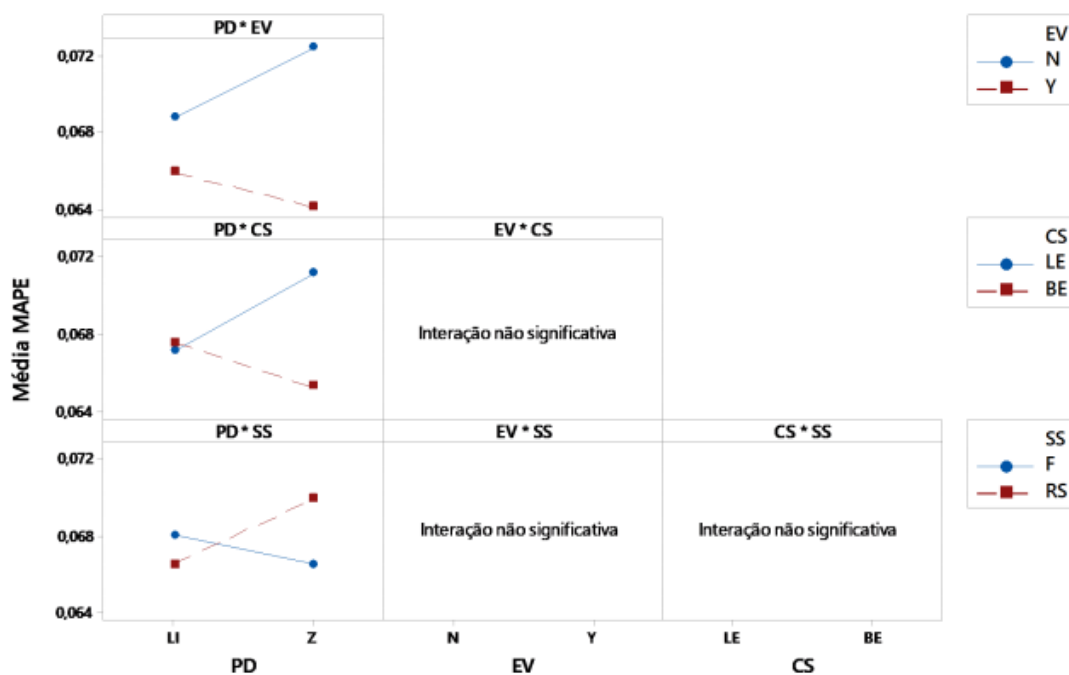


Figura 3.7 – Gráfico de interações para a resposta Média MAPE
Fonte: autor.

Seguindo a mesma lógica, a próxima subseção faz as mesmas análises para a resposta DP MAPE.

3.2.3.1 Análise dos resultados experimentais para a resposta DP MAPE

Os fatores e interações significativas podem ser detectadas por meio da análise de variância descrita na tabela 3.7.

Tabela 3.7 – ANOVA para a resposta DP MAPE

Fonte	GL	SQ. (Aj.)	QM. (Aj.)	Valor F	p-valor
Model	10	0,000117	0,000012	28,96	0,001
Linear	5	0,000056	0,000011	27,59	0,001
PD	1	0,000007	0,000007	18,21	0,008
EV	1	0,000028	0,000028	70,28	0,000
CS	1	0,00001	0,00001	23,65	0,005
IV	1	0,000000	0,000000	1,07	0,348
R	1	0,00001	0,00001	24,73	0,004
Interações de 2 fatores	4	0,000045	0,000011	27,94	0,001
PD*EV	1	0,000004	0,000004	8,92	0,031
PD*IV	1	0,000001	0,000001	1,25	0,314
EV*IV	1	0,000015	0,000015	37,27	0,002
EV*R	1	0,000026	0,000026	64,32	0,000
Interação de 3 fatores	1	0,000016	0,000016	39,86	0,001
PD*EV*IV	1	0,000016	0,000016	39,86	0,001
Error	5	0,000002	0,000000		
Total	15	0,000119			

Observa-se que a forma de padronizar os dados, as variáveis explicativas, o critério de retenção das redes e o número de etapas de tempo impactam na resposta, Interação entre a forma de padronizar os dados e as variáveis explicativas e a seleção de subconjuntos de variáveis independentes são detectadas. Também há interações entre as variáveis explicativas com a seleção de subconjunto de variáveis independentes e a quantidade de entradas da rede influenciam de forma significativa a resposta DP MAPE (p-valor inferior a 0,05, utilizando o método de eliminação backward). O gráfico de pareto na figura 3.8 ilustra o resultado obtido pela ANOVA e o modelo para esta resposta está exibida na equação 3.7.

$$\begin{aligned}
\text{DP MAPE} = & 0,017600 - 0,000678(PD) - 0,001333(EV) + 0,000773(CS) \\
& + 0,000165(IV) - 0,000791(R) - 0,000475(PD * EV) \\
& + 0,000178(PD * IV) + 0,000971(EV * IV) \\
& - 0,001275(EV * R) - 0,001004(PD * EV * IV)
\end{aligned} \tag{3.7}$$

O modelo para a resposta DP MAPE também apresentou um ajuste satisfatório com R^2 e R^2 ajustado de respectivamente 98,30% e 94,91%, com os resíduos gerados por esse modelo sendo normalmente distribuídos segundo o teste Anderson-Darling (p-valor igual a 0,818), conforme a figura 3.9.

A figura 3.10 exhibe os efeitos principais dos fatores sobre a resposta DP MAPE, possibilitando o entendimento do impacto de cada um dos fatores na resposta.

Ao contrário do que aconteceu com a resposta Média MAPE, a forma de padronização dos dados (PD) afeta a resposta DP MAPE. Veja que a rede possui melhor

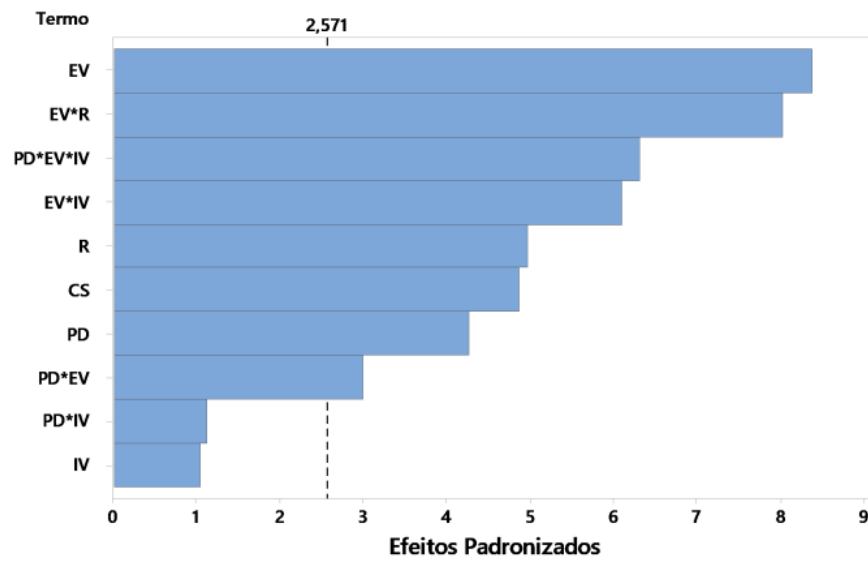


Figura 3.8 – Gráfico de Pareto para a resposta DP MAPE
 Fonte: autor.

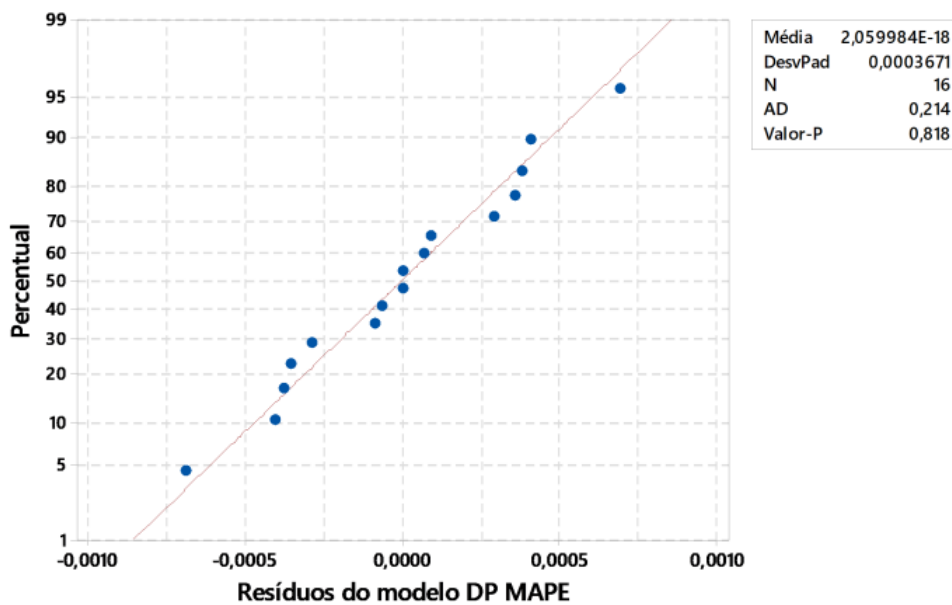


Figura 3.9 – Gráfico de Probabilidade para os resíduos para o modelo DP MAPE
 Fonte: autor.

performance para esta resposta quando os dados são padronizados (Z). O uso de variáveis explicativas como entrada da rede neural também auxilia na melhoria da performance para esta resposta. Já em relação ao critério de seleção de redes retidas, para a resposta DP MAPE o melhor desempenho acontece quando se usa o critério do menor erro (LE). Observe que uso da Transformada rápida de Fourier para determinação da quantidade de unidades de tempo usadas como entrada também melhora o resultado da rede. Assim como aconteceu com a outra resposta analisada, devido a existência de uma interação tripla significativa (PD*EV*IV) as interações de ordem inferiores estão presentes no modelo.

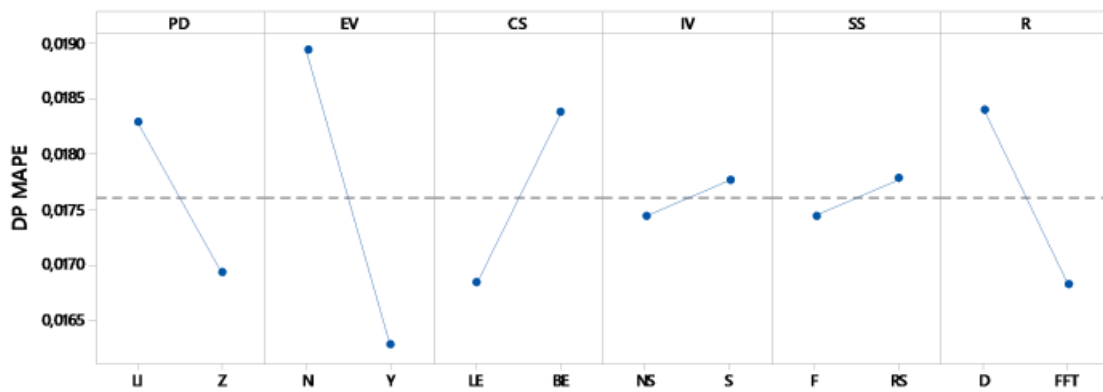


Figura 3.10 – Gráfico de efeitos principais para a resposta DP MAPE
Fonte: autor.

A figura 3.11 exibe as interações. Nota-se que assim como ocorreu com a resposta Média MAPE, quando se utiliza a padronização estatística associada ao uso de variáveis explicativas como entrada da rede a performance é melhorada.

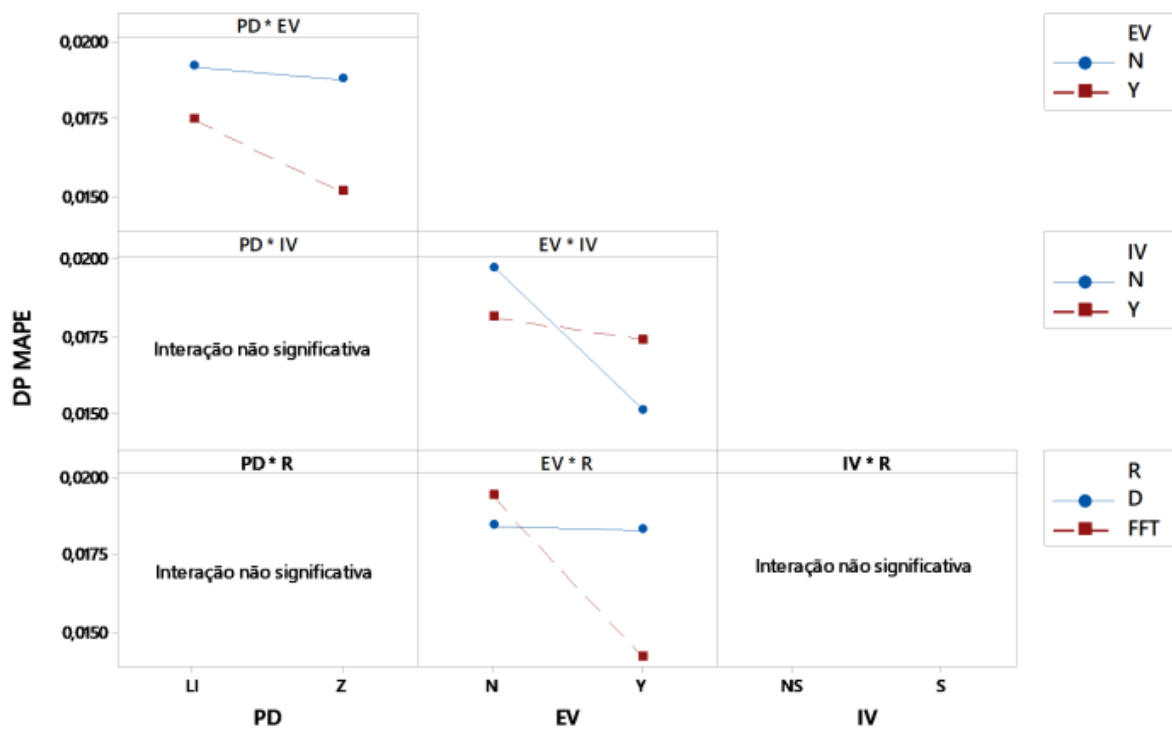


Figura 3.11 – Gráfico de interações para a resposta DP MAPE
Fonte: autor.

Uma melhor resposta também é observada quando existe a utilização das variáveis explicativas (EV) associado a não seleção de subconjuntos de variáveis independentes, ou seja, quando o software não faz a seleção de quais variáveis devem ser usadas como entrada no período de teste da rede associado ao uso de variáveis explicativas a rede apresenta melhor resultado. Por fim, uma atenção especial pode ser dada à interação entre os fatores EV e R, isto é, associado ao uso da Transformada de Fourier para determinação do número de

etapas de tempo usadas como entradas as variáveis explicativas auxiliam na melhoria da resposta DP MAPE.

Após o entendimento de como os fatores e suas respectivas interações impactam nas respostas analisadas prossegue-se para a obtenção de uma parametrização para as redes neurais que irão compor o ensemble.

3.3 Parametrização de redes usando o método desirability

Baseado nos resultados apresentados anteriormente, próximo passo é de extrema importância para os objetivos deste estudo. Neste momento, necessita-se determinar uma configuração para realizar o treinamento da rede e, conseqüentemente, formar o ensemble de redes neurais. Para obter uma configuração que resulte em uma média e um desvio padrão mais baixos para o MAPE, resolvemos um problema de otimização usando o algoritmo desirability do software Minitab®. Como o intuito é minimizar as duas respostas simultaneamente usando a função de conveniência, é necessário definir os valores para o alvo T_i e para o limite superior H_i conforme equação 2.28. Como estratégia para se determinar esses valores alguns indicadores estatísticos como média, valor máximo, valor mínimo e os quartis são utilizados. Além de T_i e H_i é necessário a atribuição dos valores de peso para as duas respostas, determinando assim a forma da função desirability. Por fim, é necessário definir também a importância de cada resposta, ou seja, a importância de cada uma das respostas na função desirability composta, o que indica qual delas deverá apresentar maior prioridade. A tabela 3.8 resume quais foram os parâmetros utilizados no algoritmo e os resultados do problema de otimização são exibidos na figura 3.12

Tabela 3.8 – Valores utilizados no algoritmo desirability

Resposta	Objetivo	Alvo (T_i)	Limite superior (H_i)	Peso	Importância
Média MAPE	Minimizar	0,0582	0,0668 (mediana)	0,1	1
DP MAPE	Minimizar	0,0119	0,0178 (mediana)	0,1	1

As linhas verticais vermelhas representam a configuração otimizada encontrada pelo método (os níveis escolhidos para cada fator são destacados em vermelho na parte superior da figura 3.12) e os valores em azul representam os valores de resposta encontrados. As regiões cinzas indicam onde a resposta correspondente tem conveniência nula, ou seja $d = 0$.

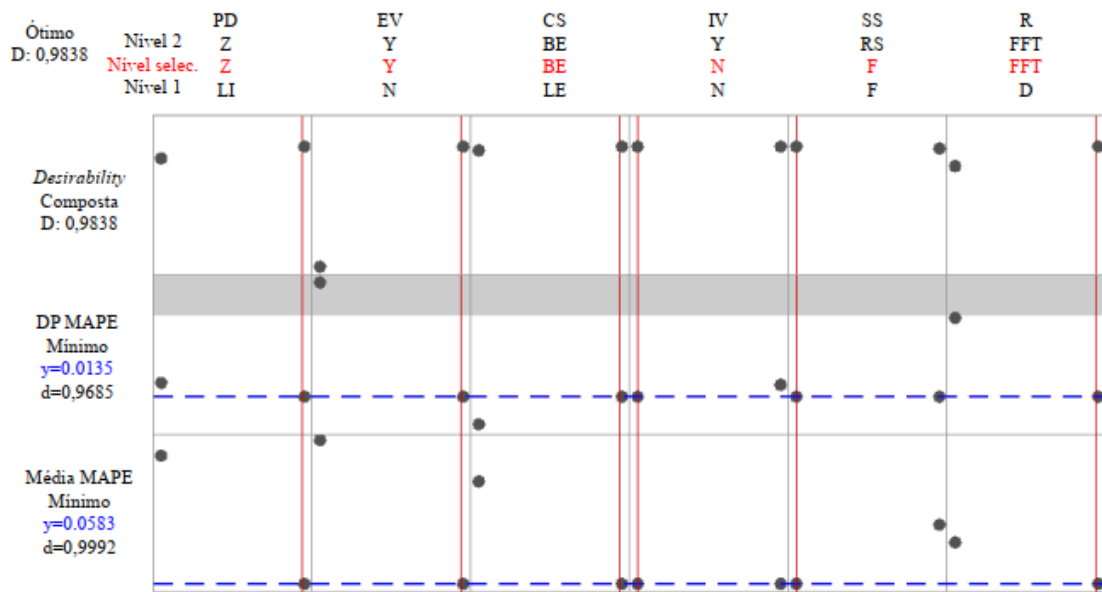


Figura 3.12 – Parametrização pelo método desirability

Fonte: autor.

Para as duas respostas de interesse neste estudo, a conveniência composta é de 0,9838 e os níveis escolhidos para as variáveis aqui estudadas são os seguintes:

- Pré-processamento de dados: padronização estatística (Z);
- Variáveis explicativas: usar as variáveis explicativas como entrada da rede neural (Y);
- Critério de seleção de redes retidas: erro equilibrado (BE);
- Seleção de subconjunto de variáveis independentes: não fazer a seleção (N);
- Seleção de subconjuntos para cada rede treinada: fixo (F);
- Número de etapas de tempo usadas como entradas: utilização da transformada rápida de Fourier (FFT), ou seja, serão utilizadas 24 etapas de tempo como entradas.

A partir da parametrização determinada pelo algoritmo, foram selecionadas três redes neurais por meio do software. A figura 3.13 exibe a arquitetura das redes selecionadas.

Pode-se observar por meio da figura que todas as redes selecionadas possuem 144 neurônios na cada de entrada cuja função de ativação é a função identidade. Esses neurônios não executam o processamento e simplesmente inserem os 144 valores de entrada que representam as seis variáveis de entrada utilizadas: os dados de carga elétrica pré-processados das quatro subestações de distribuição de energia elétrica e as variáveis explicativas (dia e período da medição).

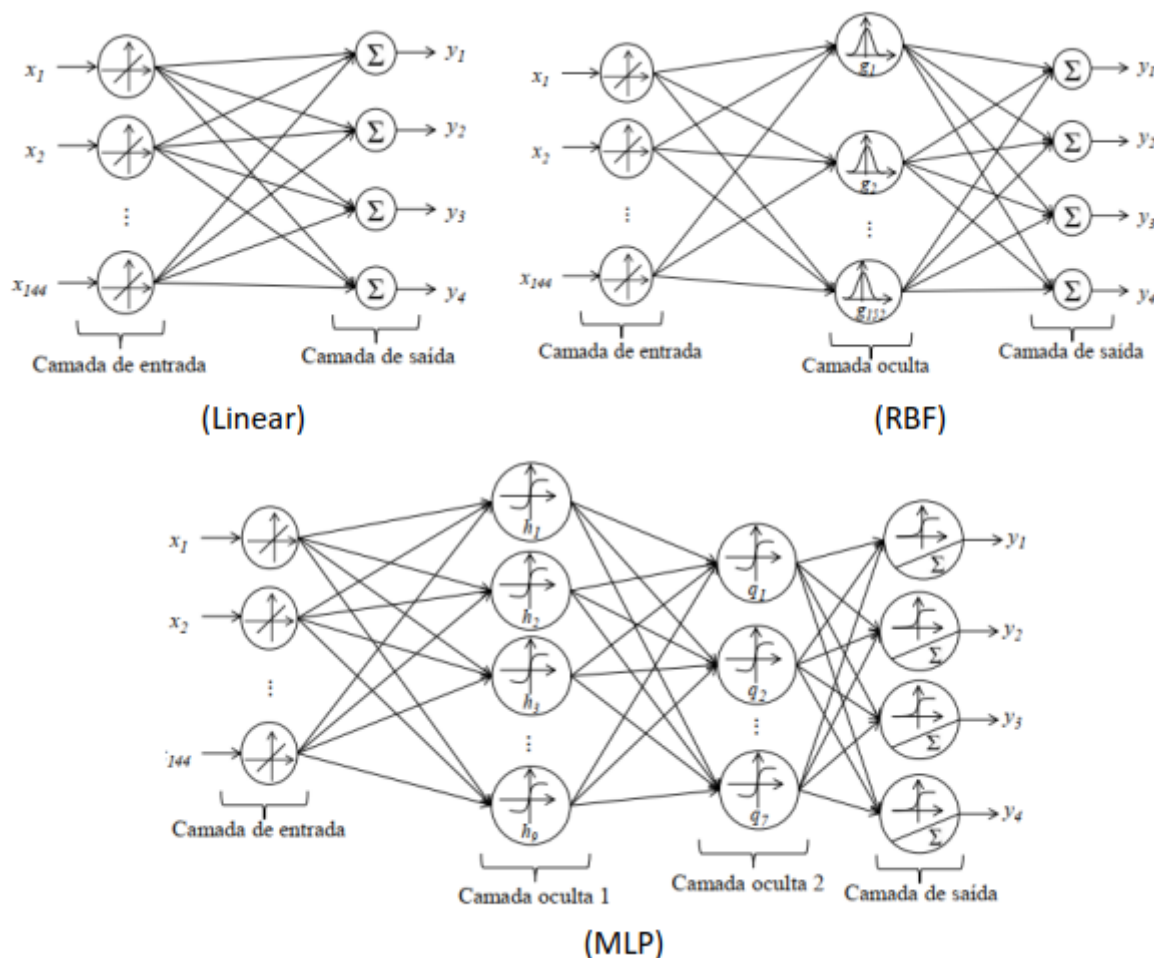


Figura 3.13 – Arquitetura das redes neurais usadas no ensemble
 Fonte: autor.

A rede MLP possui duas camadas ocultas contendo cada uma 9 e 7 neurônios tendo como função de ativação a função hiperbólica e a função sigmoideal como função de ativação na camada de saída. Já as redes RBF (com 152 neurônios na camada oculta e função de ativação gaussiana) e linear a função de ativação na camada de saída é a função identidade.

3.4 Considerações finais sobre o capítulo

Este capítulo teve o intuito de descrever todo o planejamento experimental para a obtenção das redes neurais que serão utilizadas no ensemble. Considerando que o objeto de estudo foi considerado como um problema de otimização complexo por envolver múltiplas opções de parametrização da redes neurais e mais de uma variável de resposta o delineamento experimental foi dividido em duas etapas: a primeira referindo-se a execução dos experimentos e análise de resultados através de gráficos de efeitos principais e interações e a segunda que esteve relacionada com a obtenção de uma parametrização através de uma ferramenta de otimização.

4 PREVISÃO DE CARGA ELÉTRICA USANDO O ENSEMBLE DE REDES NEURAIS ARTIFICIAIS

Este capítulo apresenta a segunda etapa para obtenção do ensemble de redes neurais artificiais para previsão de demanda de energia elétrica de subestações de distribuição de energia elétrica. Baseado em uma ferramenta de otimização multiobjetivo (método de otimização interseção normal a fronteira - NBI) associado a técnica de análise fatorial exploratória a ponderação utilizada para combinar as saídas das redes para formar a saída do ensemble foi obtida através do critério da máxima razão entre entropia e erro percentual global. Todo esse ferramental foi aplicado baseado em um experimento de mistura, mais especificamente um arranjo simplex lattice.

4.1 Ponderação utilizada no ensemble de redes neurais

Para obter a ponderação utilizadas para combinar as saídas das redes neurais foram aplicadas as etapas ilustradas na figura 4.1, detalhadas a seguir.

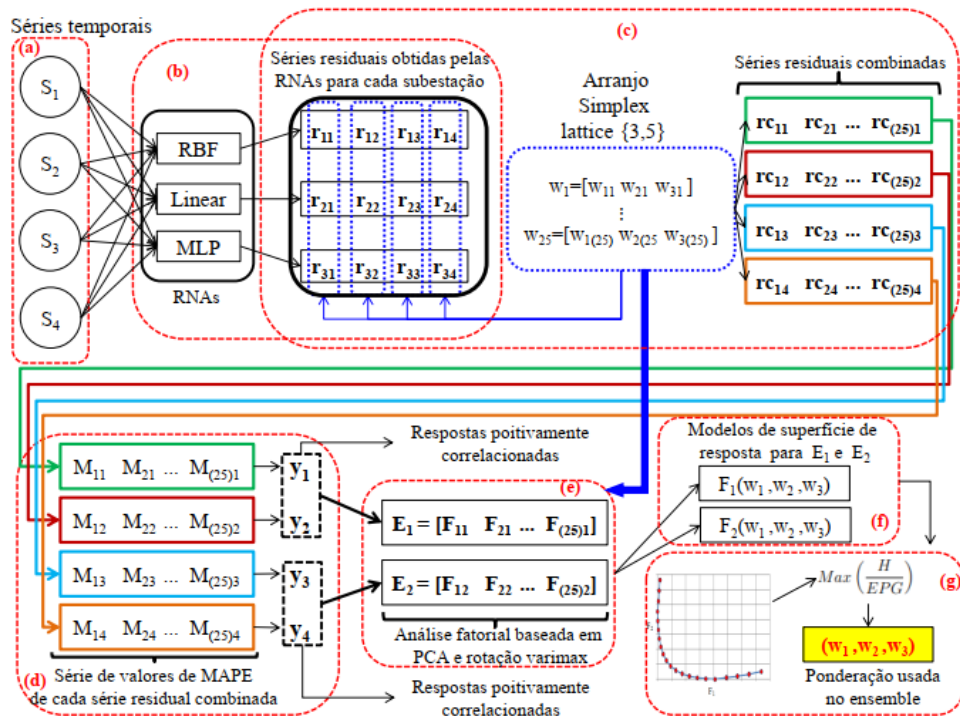


Figura 4.1 – Metodologia para combinação das saídas das redes neurais

Fonte: autor.

- (a) Coleta das observações das séries temporais de demanda de energia elétrica das 4 subestações (S_1 , S_2 , S_3 e S_4) a serem usadas para a obtenção dos pesos utilizados no ensemble (figura 4.2). Das 761 observações pertencentes as séries temporais de

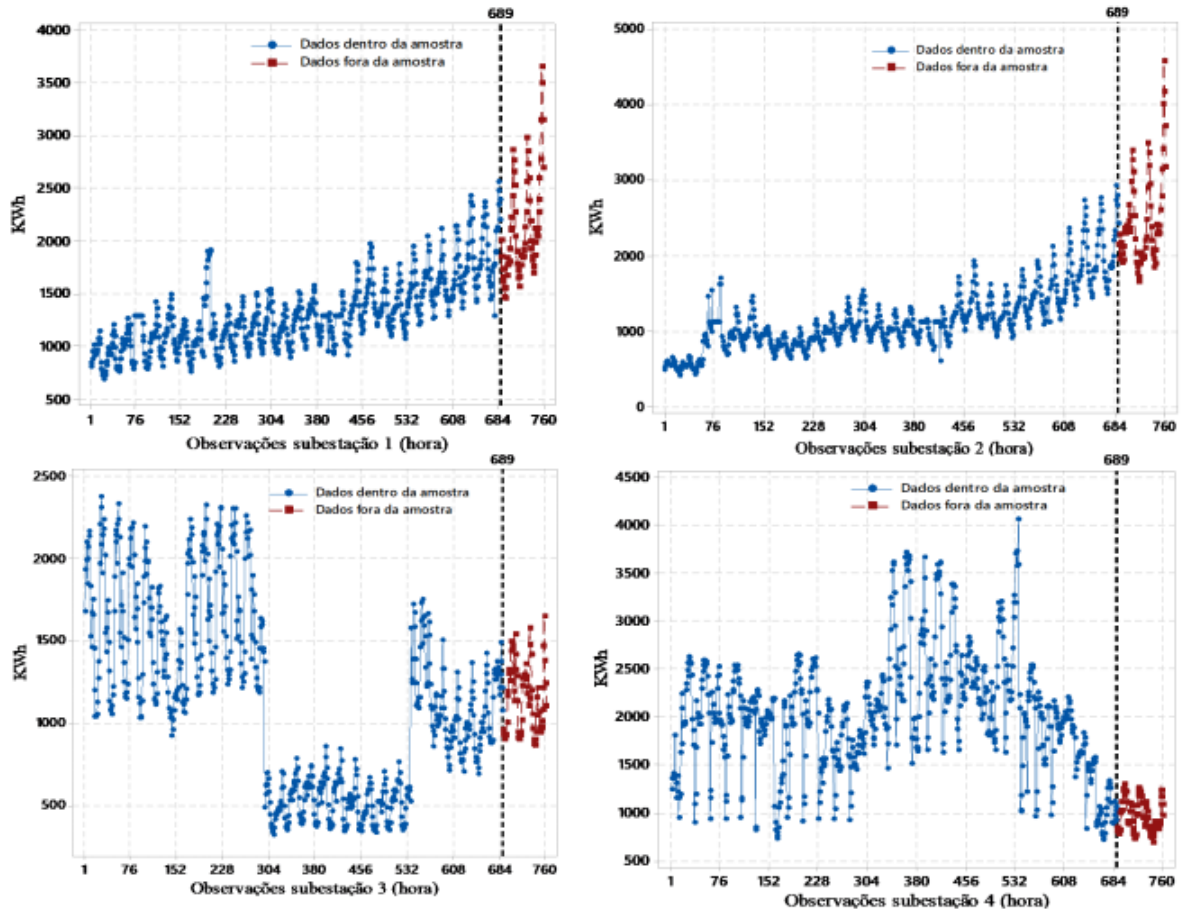


Figura 4.2 – Dados utilizados no ajuste e validação dos modelos

Fonte: autor.

demanda de energia elétrica das subestações, 72 observações foram deixadas de fora para validar os métodos de previsão. Assim, os métodos testados foram ajustados considerando as 689 observações restantes.

- (b) De acordo com a parametrização determinada no capítulo anterior, três redes neurais foram escolhidas para participar do ensemble: uma rede Linear, uma rede RBF e uma rede MLP conforme figura 3.13. Cada rede produziu quatro séries residuais r_{ij} (em que $i = 1, 2, 3$ representam as redes RBF, Linear e MLP respectivamente e $j = 1, \dots, 4$ as subestações, nesta ordem) totalizando 12 séries de resíduos com 665 observações.
- (c) Usando a abordagem de projeto de experimentos de misturas, foram definidos 25 vetores de pesos $w_k = [w_{1k} \ w_{2k} \ w_{3k}]$, $k = 1, \dots, 25$ segundo um arranjo simplex lattice $\{3, 5\}$ com um ponto central e 3 pontos axiais incorporados, conforme figura

4.3. Para cada vetor w_k e para cada série residual obtida na etapa (b) foram calculadas as séries de resíduos combinados \mathbf{rc}_{kj} , de acordo com a equação

$$\mathbf{rc}_{kj} = \sum_{t=1}^3 w_{tk} \mathbf{r}_{tj} \quad k = 1, \dots, 25, j = 1, \dots, 4 \quad (4.1)$$

cada uma também com 665 observações.

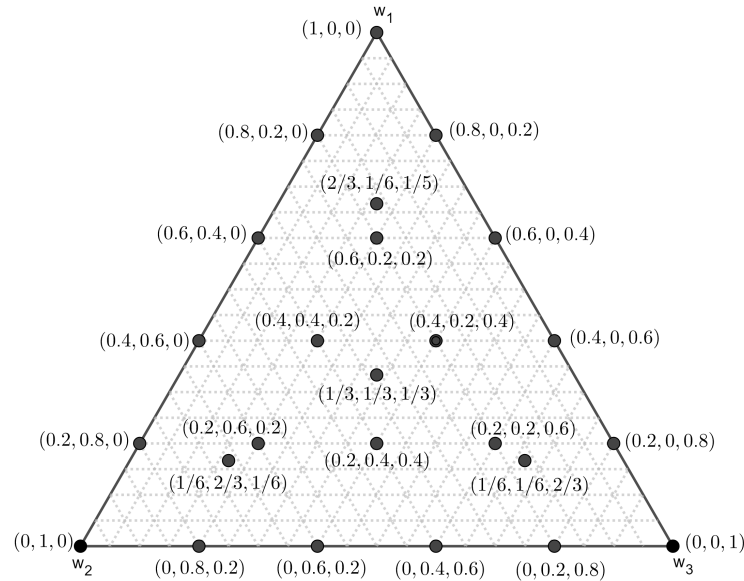


Figura 4.3 – Arranjo experimental simplex lattice {3, 5}
Fonte: autor.

- (d) Para cada série de resíduos combinados \mathbf{rc}_{kj} foram calculados a medida MAPE (veja equação 3.1) associada dando origem as medidas M_{kj} , $k = 1, \dots, 25$, $j = 1, \dots, 4$, que foram modeladas matematicamente em função dos pesos w_1 , w_2 e w_3 , dando origem as funções y_1 , y_2 , y_3 e y_4 que representam o valor da medida MAPE para as subestações 1, 2, 3 e 4 respectivamente.
- (e) Após a análise e constatação de existência de correlação entre as respostas y_1, y_2, y_3 e y_4 procede-se com a aplicação da análise fatorial para cálculo dos escores de fatores. É importante ressaltar que apesar de ser possível extrair um número de fatores igual ao número de respostas, geralmente um número menor de fatores já garante a explicação da variâncias das respostas reduzindo a dimensionalidade do problema. Assim, aplicando a análise fatorial com extração por componentes principais e método de rotação varimax, determinou-se a retenção de 2 fatores aplicando-se a regra de kaiser [97]. Cada uma das observações das 2 séries de escores de fatores $E_1 = [F_{11} \dots F_{(25)1}]$ e $E_2 = [F_{21} \dots F_{(25)2}]$, (com 25 observações) representou uma resposta relacionada a um determinado conjunto de resíduos combinados e, conseqüentemente, a um determinado conjunto de pesos selecionado na etapa (c).

- (f) As 2 séries de escores de fatores E_1 e E_2 representativas das 4 séries de respostas M_{k1} , M_{k2} , M_{k3} e M_{k4} , $k = 1, \dots, 25$ foram modeladas em função de w_1 , w_2 e w_3 . Cada uma dessas 25 observações das 2 séries de escores de fatores representou uma resposta relacionada a um dado conjunto de resíduos combinados e , conseqüentemente a um dado conjunto de pesos no simplex lattice $\{3, 5\}$ selecionado na etapa (c). Esses conjuntos de pesos definidos na etapa (c), juntamente com as respostas produzidas na etapa (e) produziram o desenho experimental de misturas para cada série de escores de fatores F_1 e F_2 , e conseqüentemente suas superfícies de respostas ϕE_1 e ϕE_2 gerando as respectivas funções para $F_1(w_1, w_2, w_3)$ (representando y_1 e y_2) e $F_2(w_1, w_2, w_3)$ (representando y_3 e y_4) sobre a região simplex lattice $\{3, 5\}$.
- (g) Obtidas as funções $F_1(w_1, w_2, w_3)$ e $F_2(w_1, w_2, w_3)$, estas foram otimizadas simultaneamente através do método NBI, sendo que a otimização conjunta dessas duas funções representou, na realidade, a otimização conjunta das quatro respostas originais y_1 , y_2 , y_3 e y_4 . Nesse processo de otimização conjunta, cada uma das soluções Pareto-ótimas encontradas e conseqüentemente cada um dos conjuntos Pareto ótimos encontrados representaram um ponto na fronteira de Pareto. Para cada uma das soluções Pareto ótimas foram calculados EPG e entropia segundo as equações 2.38 e 2.39 respectivamente. A solução foi escolhida a que apresentou maior relação $\zeta = \frac{H}{EPG}$.
- (h) Após a determinação da ponderação, os resultados da combinação dos resíduos com a ponderação obtida pelo método NBI-FA foi comparado com os resultados produzidos por cada uma das redes neurais nos dados deixados fora da amostra.

4.2 Resultados e discussão

Cada rede neural produziu uma série de resíduos com 665 observações para cada subestação. Esses resíduos estavam padronizados segundo a equação 3.3 pois as redes neurais receberam as entradas conforme esta padronização. As tabelas 4.1 e 4.2 exibem de forma resumida as séries residuais.

Tabela 4.1 – Séries de resíduos produzidas pelas redes neurais para as subestações 1 e 2

Obs.	Subestação 1				Subestação 2			
	Real (KWh)	RBF	LIN.	MLP	Real (KWh)	RBF	LIN.	MLP
1	752,58	250,75	52,32	153,26	436,64	160,49	148,09	237,42
2	812,09	376,18	23,21	232,68	498,24	-8,46	-106,93	230,64
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
665	2199,90	-70,75	26,23	-24,86	2427,40	29,15	51,16	80,03

Tabela 4.2 – Séries de resíduos produzidas pelas redes neurais para as subestações 3 e 4

Obs.	Subestação 3				Subestação 4			
	Real (KWh)	RBF	LIN.	MLP	Real (KWh)	RBF	LIN.	MLP
1	1964,89	-203,70	-357,13	-85,79	2475,20	-79,18	66,40	-62,10
2	2207,82	-332,46	-108,53	-93,37	2557,00	-223,39	263,78	88,31
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
665	1070,49	23,68	-110,30	42,25	917,85	-62,18	-3,40	58,73

Denotando por \mathbf{r}_{ij} os vetores de resíduos acima (de dimensão 665×1) em que $i = 1, 2, 3$ representa a rede neural RBF, Linear e MLP respectivamente e $j = 1, \dots, 4$ corresponde a subestações, podemos representar as tabelas 4.1 e 4.2 em uma matriz, afim de simplificar a notação

$$A = \begin{bmatrix} \mathbf{r}_{11} & \mathbf{r}_{12} & \mathbf{r}_{13} & \mathbf{r}_{14} \\ \mathbf{r}_{21} & \mathbf{r}_{22} & \mathbf{r}_{23} & \mathbf{r}_{24} \\ \mathbf{r}_{31} & \mathbf{r}_{32} & \mathbf{r}_{33} & \mathbf{r}_{34} \end{bmatrix}. \quad (4.2)$$

A cada série de resíduos da matriz A temos associada a métrica de performance MAPE coforme euquação 3.1, exibidas na tabela 4.3.

Tabela 4.3 – Métricas de performance das redes neurais para as subestações

RNA	Subestação 1	Subestação 2	Subestação 3	Subestação 4
RBF	0,0449	0,0589	0,0676	0,0724
Linear	0,0322	0,0515	0,0745	0,0707
MLP	0,0571	0,0831	0,0594	0,0586

Após a obtenção desta matriz residual, um delineamento simplex lattice $\{3, 5\}$ foi criado, com um ponto central e 3 pontos axiais incorporados, resultando em 25 pontos no arranjo, conforme pode ser visto na figura 4.3. Estes 25 conjuntos de pesos foram utilizados para combinar as 685 observações de cada uma das 12 séries residuais conforme a equação 4.1 produzindo no total 100 séries de resíduos combinados, conforme mostra a matriz RC

$$RC = \begin{bmatrix} \mathbf{rc}_{11} & \mathbf{rc}_{12} & \mathbf{rc}_{13} & \mathbf{rc}_{14} \\ \mathbf{rc}_{21} & \mathbf{rc}_{22} & \mathbf{rc}_{23} & \mathbf{rc}_{24} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{rc}_{(25)1} & \mathbf{rc}_{(25)2} & \mathbf{rc}_{(25)3} & \mathbf{rc}_{(25)4} \end{bmatrix} \quad (4.3)$$

em que cada elemento \mathbf{rc}_{ij} , $i = 1, \dots, 25$, $j = 1, \dots, 4$ da matriz RC é uma série contendo 665 observações. A partir daí, para cada série de resíduos \mathbf{rc}_{ij} foram calculadas a métrica MAPE (conforme equação 3.1) resultando então na tabela de métricas 4.4.

Tabela 4.4 – MAPE das séries de resíduos combinados

k	w_1	w_2	w_3	Subestação 1	Subestação 2	Subestação 3	Subestação 4
1	1,0	0	0	0,0449	0,0589	0,0676	0,0724
2	0,8	0,2	0	0,0377	0,0501	0,0594	0,0649
3	0,8	0	0,2	0,0429	0,0545	0,0602	0,0633
4	0,6	0,4	0	0,0321	0,0436	0,0562	0,0611
5	0,6	0,2	0,2	0,0363	0,0463	0,0530	0,0582
6	0,6	0	0,4	0,0438	0,0543	0,0554	0,0570
7	0,4	0,6	0	0,0286	0,0404	0,0580	0,0610
8	0,4	0,4	0,2	0,0314	0,0413	0,0520	0,0569
9	0,4	0,2	0,4	0,0381	0,0478	0,0500	0,0543
10	0,4	0	0,6	0,0468	0,0583	0,0538	0,0535
11	0,2	0,8	0	0,0290	0,0418	0,0646	0,0645
12	0,2	0,6	0,2	0,0293	0,0399	0,0559	0,0595
13	0,2	0,4	0,4	0,0343	0,0446	0,0510	0,0560
14	0,2	0,2	0,6	0,0422	0,0536	0,0506	0,0539
15	0,2	0	0,8	0,0513	0,0660	0,0553	0,0542
16	0	1,0	0	0,0322	0,0474	0,0745	0,0707
17	0	0,8	0,2	0,0308	0,0433	0,0643	0,0654
18	0	0,6	0,4	0,0333	0,0453	0,0567	0,0613
19	0	0,4	0,6	0,0396	0,0525	0,0532	0,0585
20	0	0,2	0,8	0,0478	0,0634	0,0542	0,0577
21	0	0	1,0	0,0571	0,0765	0,0594	0,0586
22	1/3	1/3	1/3	0,0341	0,0439	0,0502	0,0553
23	2/3	1/6	1/6	0,0373	0,0478	0,0548	0,0597
24	1/6	2/3	1/6	0,0291	0,0400	0,0583	0,0611
25	1/6	1/6	2/3	0,0444	0,0569	0,0515	0,0542

Baseado nos valores da tabela 4.4 determinou-se então os modelos matemáticos de superfícies para as respostas y_i , $i = 1, \dots, 4$ (y_i representa o valor de MAPE da série residual combinada da subestação i) exibidas nas equações 4.4 a 4.7.

$$y_1 = 0,04506w_1 + 0,03144w_2 + 0,05815w_3 - 0,03378w_1w_2 - 0,02634w_1w_3 - 0,03512w_2w_3 \quad (4.4)$$

$$y_2 = 0,05925w_1 + 0,04653w_2 + 0,07776w_3 - 0,04671w_1w_2 - 0,05112w_1w_3 - 0,05625w_2w_3 \quad (4.5)$$

$$y_3 = 0,06767w_1 + 0,07546w_2 + 0,05953w_3 - 0,06093w_1w_2 - 0,03892w_1w_3 - 0,0527w_2w_3 \quad (4.6)$$

$$y_4 = 0,07264w_1 + 0,0711246w_2 + 0,05830w_3 - 0,04499w_1w_2 - 0,0429w_1w_3 - 0,02036w_2w_3 \quad (4.7)$$

cujos gráficos de contorno e probabilidade de resíduos são exibidos nas figuras 4.4, 4.5 e

4.6 respectivamente.

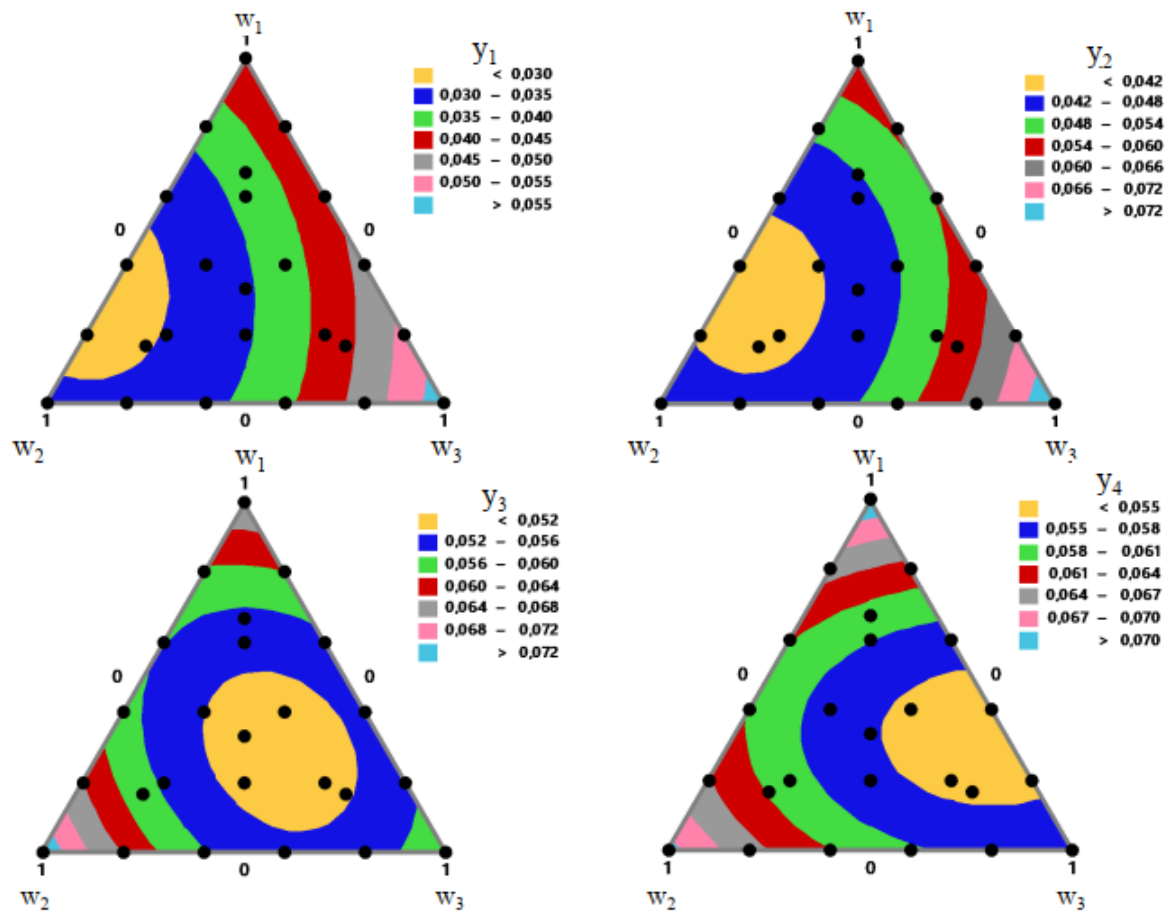


Figura 4.4 – Gráfico de contorno para y_1 , y_2 , y_3 e y_4
 Fonte: autor.

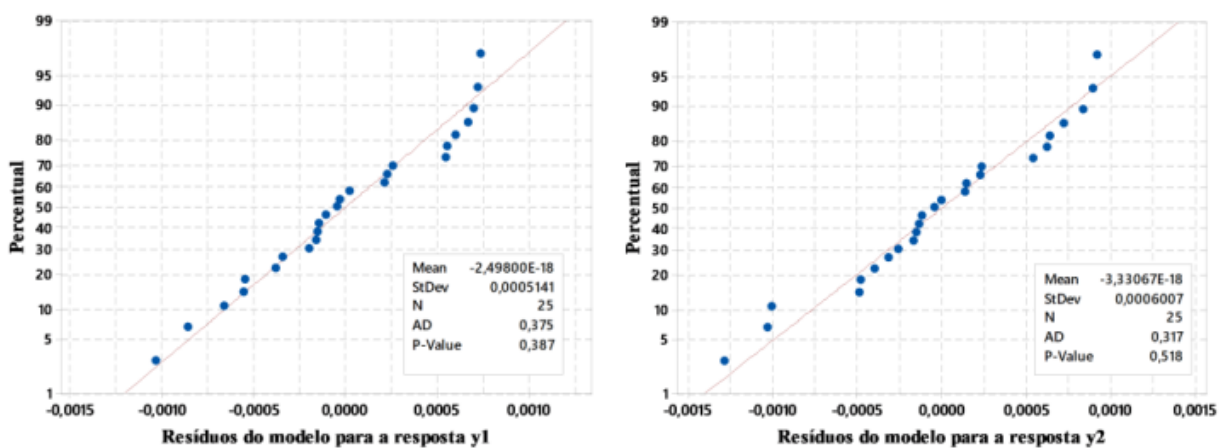


Figura 4.5 – Gráfico de probabilidade de resíduos para y_1 e y_2
 Fonte: autor.

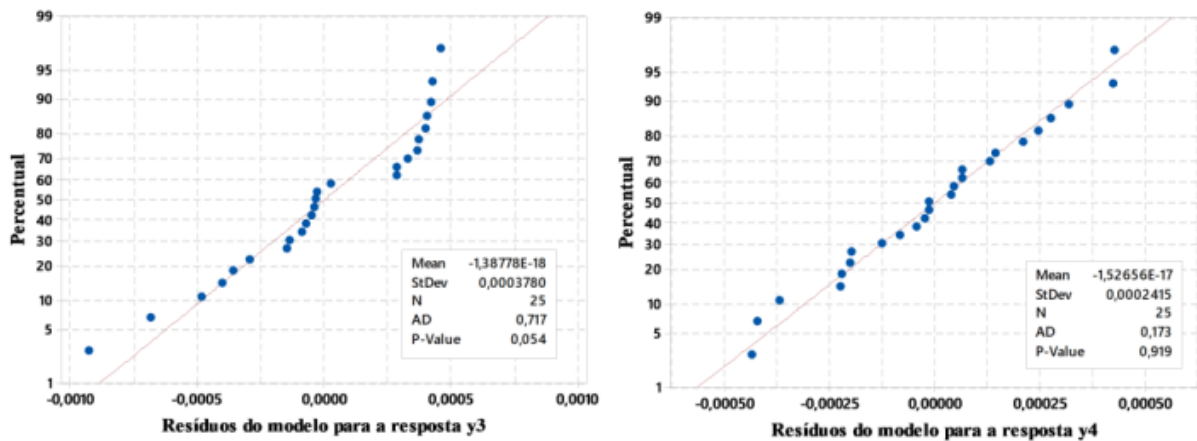


Figura 4.6 – Gráfico de probabilidade de resíduos para y_3 e y_4 .
Fonte: autor.

Como se pode observar pelas figuras 4.5 e 4.6, todos os modelos apresentaram resíduos normalmente distribuídos pelo teste de Anderson-Darling e os valores de R^2 e R^2_{adj} foram superiores a 99%, revelando que os modelos representam o conjunto de dados.

Por meio da tabela 4.5 podemos detectar correlações significativas (p-valor inferior a 0,05) e positivas entre os pares de respostas y_1 e y_2 , y_3 e y_4 , sendo portanto justificada a utilização da análise fatorial visando a redução de dimensionalidade.

Tabela 4.5 – Estrutura de correlação entre as respostas y_1 , y_2 , y_3 e y_4

	y_1	y_2	y_3
y_2	0,98*		
	0,000**		
y_3	-0,143	-0,001	
	0,494	0,998	
y_4	-0,297	-0,177	0,911
	0,149	0,397	0,000

*Correlação de Pearson

**p-valor

Para aplicar a técnica de análise fatorial com extração por componentes principais é necessário uma análise prévia das componentes principais. Como pode ser observado na tabela 4.6 as componentes principais PC_1 e PC_2 cujos autovalores são maiores do que 1 conseguem explicar, acumuladamente, 97,9% da variância das respostas y_1 , y_2 , y_3 e y_4 . Logo, pelo critério de Kaiser é suficiente utilizar somente as componentes principais PC_1 e PC_2 . De forma mais visual, o scree plot na figura 4.7 mostra que os autovalores a partir da terceira componente ficam próximos de zero, indicando que extrair somente as duas primeiras componentes é suficiente neste caso.

Tabela 4.6 – Componentes principais

	PC ₁	PC ₂	PC ₃	PC ₄
Autovalores	2,2666	1,6482	0,0759	0,0092
Variância (%)	56,7%	41,2%	1,9%	0,2%
Variância acumulada (%)	56,7%	97,9%	99,8%	100%

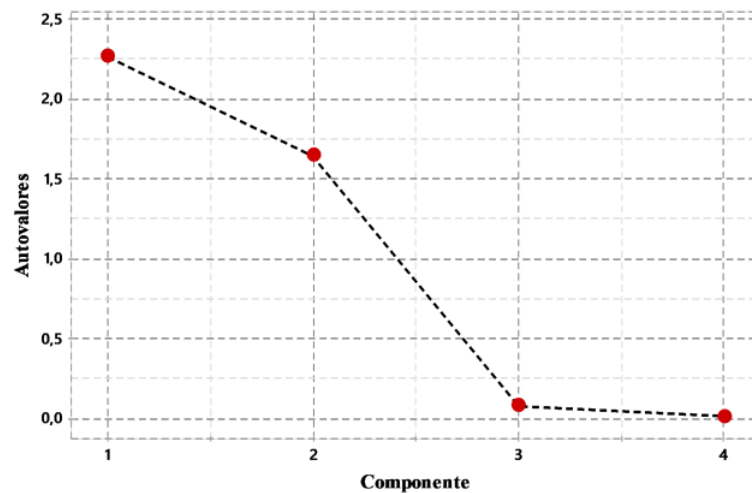


Figura 4.7 – Scree Plot

Fonte: autor.

A análise fatorial foi então aplicada (tabela 4.7) para a extração dos escores rotacionados dos dois primeiros fatores. Foi utilizado a extração por componentes principais utilizando rotação *varimax*.

Tabela 4.7 – Análise fatorial baseada em componentes principais e rotação *varimax*

Variáveis	Fator 1	Fator 2	Comunalidade
y_1	0,986	-0,147	0,993
y_2	0,998	-0,01	0,995
y_3	0,009	0,982	0,965
y_4	-0,166	-0,966	0,961
Variância	1,9948	1,9200	3,9149
% Variância	0,499%	0,48%	0,979%

Através da análise fatorial juntamente com o gráfico de carregamentos (figura 4.8) pode se observar que as respostas y_1 e y_2 definiram o primeiro fator dado que elas apresentaram cargas elevadas neste fator e pequenas no segundo fator. De forma análoga, o segundo fator foi determinado pelas variáveis y_3 e y_4 . Observa-se também a existência de elevados valores de comunalidades na tabela 4.7, indicando que os dois fatores explicam uma alta porcentagem de variabilidade de cada resposta e explicando 97,9% da variância total dessas quatro respostas.

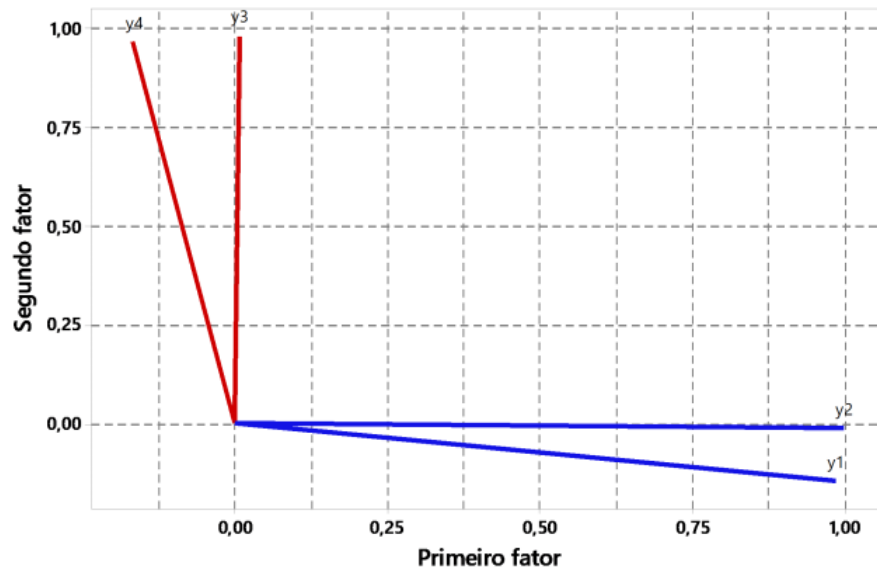


Figura 4.8 – Gráfico de carregamentos das quatro respostas
 Fonte: autor.

O dendograma na figura 4.9 mostra que de fato os fatores 1 e 2 representam os pares de respostas y_1 e y_2 , y_3 e y_4 , respectivamente. Veja que as respostas e seus fatores pertencem aos mesmo grupo e possuem alto índice de similaridade.

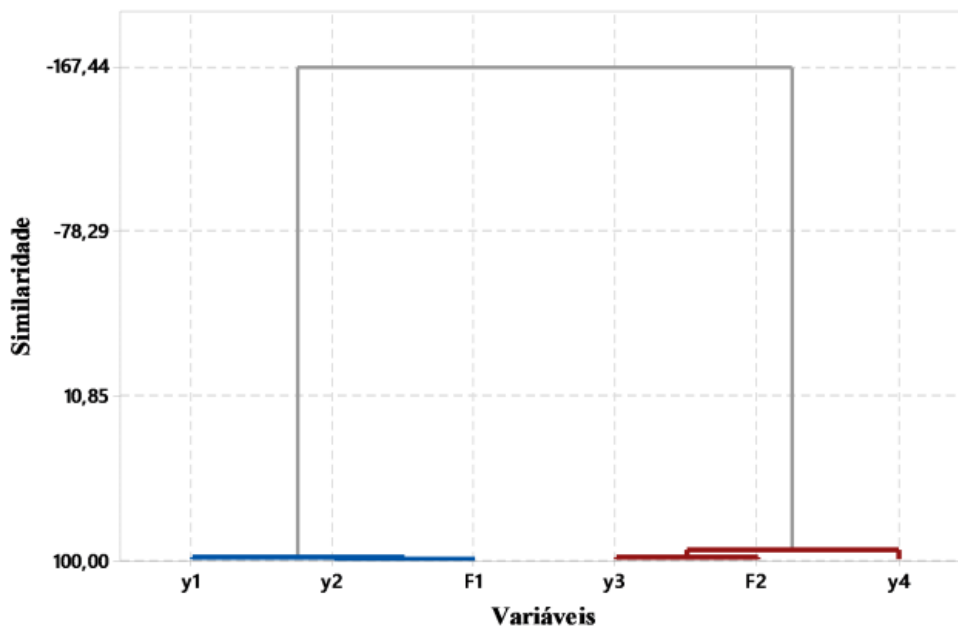


Figura 4.9 – Dendograma de similaridade entre as quatro resposta e seus respectivos fatores.
 Fonte: autor.

Podemos observar também através da tabela 4.8 que os fatores 1 e 2 estão positivamente correlacionado com as resposta eles representam. Devido a este fato, no probema de otimização bi-objetivo, as funções $F_1(w_1, w_2, w_3)$ e $F_2(w_1, w_2, w_3)$ deverão ser minimi-

zadas simultaneamente, acarretando na minimização das funções originais y_1 , y_2 , y_3 e y_4 .

Tabela 4.8 – Correlações entre as respostas e os fatores

	y_1	y_2	y_3	y_4
F_1	0,986*	0,998	0,009	-0,166
	0,000**	0,000	0,968	0,428
F_2	-0,147	-0,010	0,982	0,966
	0,483	0,962	0,000	0,000

*Coeficiente de correlação de Pearson

**p-valor

Duas séries de escores rotacionados $E_1 = [F_{11}, \dots, F_{(25)1}]$ e $E_2 = [F_{12}, \dots, F_{(25)2}]$ (tabela 4.9) foram produzidas, e foram utilizadas para modelar matematicamente os fatores 1 e 2. Os conjuntos de pesos conforme a tabela 4.4 combinados com as séries de escores E_1 e E_2 formaram o desenho experimental e conseqüentemente as superfícies de respostas ϕE_1 e ϕE_2 exibidas na figura 4.10.

Tabela 4.9 – Série de escores dos fatores

k	Escore F_1	Escore F_2	k	Escore F_1	Escore F_2
1	1,046272	2,303223	14	0,354586	-1,10681
2	-0,01647	0,749276	15	1,703799	-0,54689
3	0,577842	0,698419	16	-0,30473	2,640046
4	-0,77856	0,024934	17	-0,7628	1,169596
5	-0,39866	-0,52669	18	-0,59646	0,101105
6	0,565389	-0,35607	19	0,160847	-0,41897
7	-1,15402	0,146976	20	1,311555	-0,3213
8	-1,00447	-0,78711	21	2,710336	0,346353
9	-0,23552	-1,1709	22	-0,70577	-1,08203
10	0,963331	-0,81826	23	-0,22837	-0,20527
11	-0,95767	1,08917	24	-1,13721	0,172747
12	-1,16131	-0,19252	25	0,69509	-0,96661
13	-0,64705	-0,94241			

A exploração dessas superfícies na região simplex proporcionou a modelagem dos fatores 1 e 2, cujos gráficos de contorno com os intervalos assumidos por F_1 e F_2 são exibidos na figura 4.11

$$F_1(w_1, w_2, w_3) = 1,0789w_1 - 0,3888w_2 + 2,8494w_3 - 5,5898w_1w_2 - 5,0441w_1w_3 - 6,116w_2w_3 \quad (4.8)$$

$$F_2(w_1, w_2, w_3) = 2,33w_1 + 2,7581w_2 + 0,3334w_3 - 10,297w_1w_2 - 8,172w_1w_3 - 7,184w_2w_3. \quad (4.9)$$

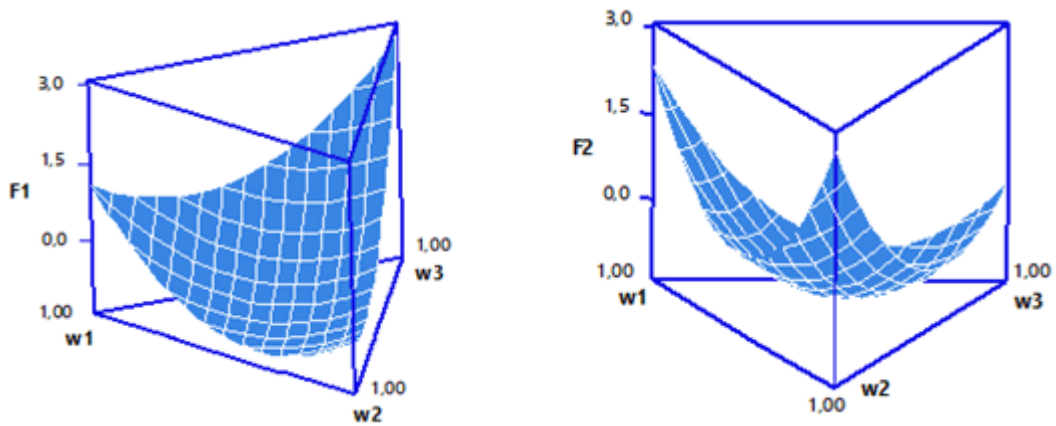


Figura 4.10 – Superfície de respostas para os fatores F_1 e F_2
 Fonte: autor.

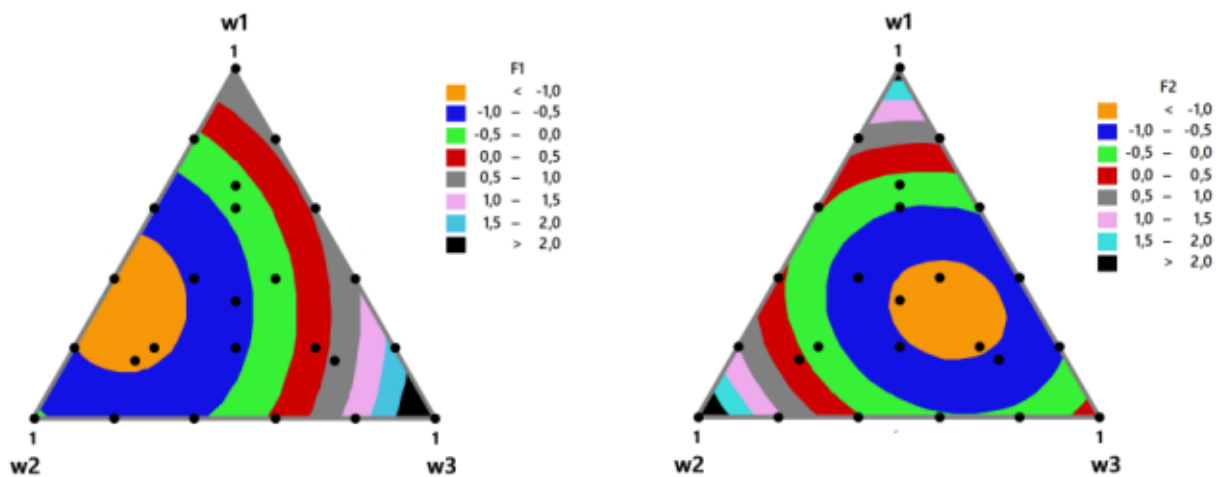


Figura 4.11 – Gráfico de contorno para os fatores F_1 e F_2
 Fonte: autor.

Nesta figura são exibidos os intervalos de valores assumidos pelas funções $F_1(w_1, w_2, w_3)$ e $F_2(w_1, w_2, w_3)$ onde os pontos destacados em negrito representam as proporções de acordo com o arranjo simplex lattice $\{3,5\}$. Como pode ser visto, as regiões onde os valores para w_1, w_2 e w_3 representam menores valores de F_1 e F_2 não se cruzam, indicando que no problema de otimização conjunta de F_1 e F_2 será impossível encontrar um conjunto de valores para w_1, w_2 e w_3 que minimiza simultaneamente as duas funções (e consequentemente as funções originais y_1, y_2, y_3 e y_4).

A modelagem foi realizada através do software Minitab, partindo-se de um modelo quadrático e utilizando-se o método backward elimination, que remove os termos não significativos do modelo. Em relação ao ajuste dos modelos, para a função F_1 os coeficientes R^2 e R^2 ajustado foram de 99,59% e 99,48% respectivamente. Para a função F_2 esses valores foram de 99,75% e 99,69%. Pelo teste de Anderson-Darling os resíduos dos dois modelos são normalmente distribuídos (p-valores de 0,342 e 0,556). Essas informações

revelam que os modelos representam bem os fatores.

As tabelas 4.10 e 4.11 exibem a análise de variância (ANOVA) para os modelos quadráticos F_1 e F_2 . Pode se observar através do p-valor que os termos de ambos os modelos foram estatisticamente significativos ao nível de 5%.

Tabela 4.10 – ANOVA para F_1

Fonte	DF	Adj. SS	Adj. MS	F	p-valor
Regressão	5	23,9015	4,78031	922,46	0
Linear	2	9,2645	4,63227	893,9	0
Quadrático	3	7,0333	2,34442	452,41	0
$w_1 \times w_2$	1	2,927	2,92699	564,83	0
$w_1 \times w_3$	1	2,3838	2,38382	460,01	0
$w_2 \times w_3$	1	3,5059	3,50595	676,55	0
Erro residual	19	0,0985	0,00518		
Total	24	24			

Tabela 4.11 – ANOVA para F_2

Fonte	DF	Adj. SS	Adj. MS	F	p-valor
Regressão	5	23,941	4,78819	1540,77	0
Linear	2	5,9016	2,95081	949,53	0
Quadrático	3	16,8993	5,63312	1812,66	0
$w_1 \times w_2$	1	9,9363	9,93634	3197,38	0
$w_1 \times w_3$	1	6,2588	6,25884	2014,01	0
$w_2 \times w_3$	1	4,8371	4,8371	1556,51	0
Erro residual	19	0,059	0,00311		
Total	24	24			

As curvas de traço de cada componente da mistura (w_1 , w_2 e w_3) mostradas nas figuras 4.12 e 4.13 indicam como as mudanças nas proporções de cada uma das componentes impactam as respostas estimadas pelas funções F_1 e F_2 . A interseção das três curvas em cada gráfico representam a mistura de referência, isto é, 1/3 para cada componente. Ao deslocar-se do ponto de mistura de referência 1/3 para a direita, a indicação é de que a proporção relativa do componente aumenta na mistura. Ao contrário, deslocar-se para a esquerda indica que a proporção relativa do componente diminui. É importante ressaltar que aumentos ou diminuições nas respostas de F_1 e F_2 representam aumentos ou diminuições nos valores das respostas que eles representam, ou seja, y_1 e y_2 , y_3 e y_4 .

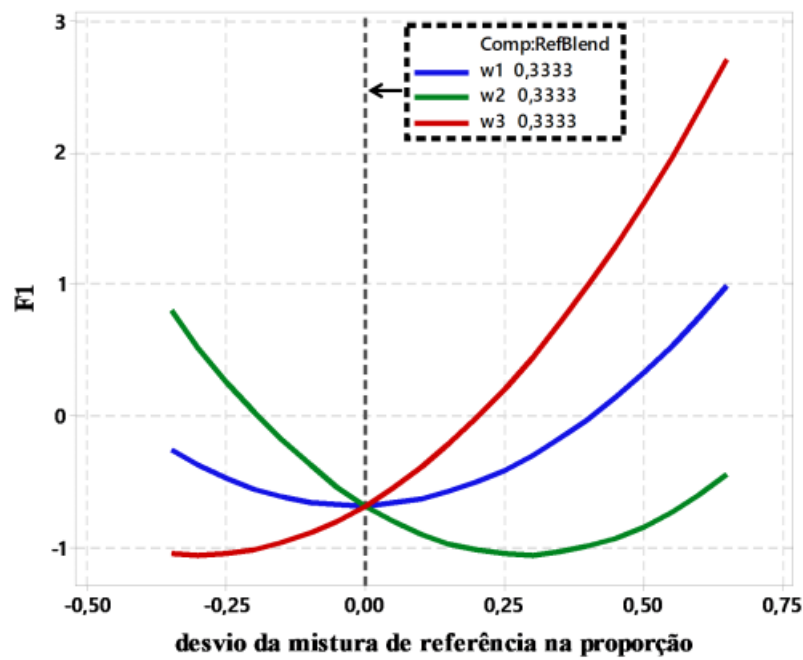


Figura 4.12 – Gráfico de traços para F_1
Fonte: autor.

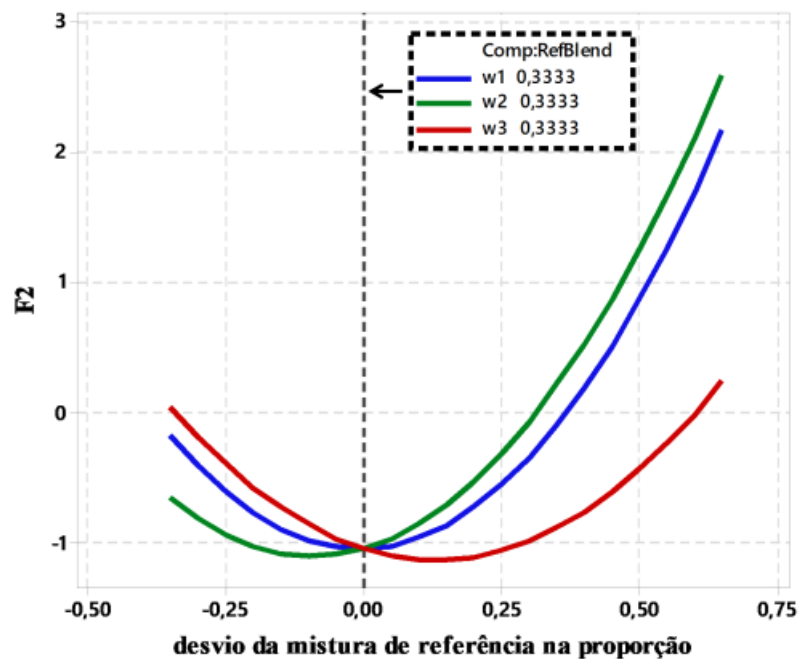


Figura 4.13 – Gráfico de traços para F_2
Fonte: autor.

Veja por exemplo que quando a proporção de w_1 (peso que representa a rede neural RBF) aumenta, o valor para a função F_1 também aumenta. O mesmo ocorre para w_3 (que representa a rede MLP), porém rapidamente. Já para o peso w_2 (que representa a rede linear), a sua adição na combinação proporciona diminuição no valor previsto da função F_1 até um determinado momento, depois passa a contribuir para o crescimento da

resposta. No caso da resposta F_2 , o aumento na proporção de w_1 ou w_2 provoca aumento na resposta. O aumento do peso w_3 também proporciona o decréscimo de F_2 até certo ponto, passando a contribuir para o acréscimo de F_2 , mas com uma velocidade menor em comparação aos demais pesos.

Após a modelagem das funções F_1 e F_2 o método de otimização NBI foi aplicado. A matriz Payoff foi calculada com base na equação 2.29, resultando na matriz

$$\Phi = \begin{bmatrix} -1,1593 & -0,1698 \\ -0,0027 & -1,15 \end{bmatrix} \quad (4.10)$$

em que a diagonal principal representam as otimizações individuais (mínimo de F_1 e F_2 respectivamente), ou seja, os valores de utopia das duas funções. A diagonal secundária são os pontos de pseudo-Nadir. As funções objetivo foram escalonadas conforme a equação 2.30, obtendo assim

$$\bar{F}(w) = \begin{bmatrix} \bar{F}_1(w_1, w_2, w_3) = \frac{F_1(w_1, w_2, w_3) - (-1,1593)}{-0,1698 - (-1,1593)} \\ \bar{F}_2(w_1, w_2, w_3) = \frac{F_2(w_1, w_2, w_3) - (-1,15)}{-0,0027 - (-1,15)} \end{bmatrix}. \quad (4.11)$$

A seguir, o problema de otimização bi-objetivo

$$\begin{aligned} \text{Min} \quad & \bar{F}_1(w_1, w_2, w_3) \\ \text{Sujeito a:} \quad & \bar{F}_1(w_1, w_2, w_3) - \bar{F}_2(w_1, w_2, w_3) + 2\beta - 1 = 0 \\ & w_1 + w_2 + w_3 = 1 \\ & 0 \leq w_i \leq 1, \quad i = 1, 2, 3 \end{aligned} \quad (4.12)$$

foi resolvido empregando o algoritmo Gradiente reduzido generalizado (GRG) e usando incrementos de 5% para a distribuição dos pesos de ponderação das funções objetivo. As restrições $w_1 + w_2 + w_3 = 1$, com $0 \leq w_i \leq 1$ para $i = 1, 2, 3$ são inerentes ao delineamento de mistura utilizado. A tabela 4.12 e exibe os resultados da orimização usando o método NBI.

Vale lembrar que devido a correlação positiva de F_1 e F_2 com as respostas que representam, diminuições nos valores de F_1 e F_2 acarretam diminuições nos valores das respostas y_1 e y_2 , y_3 e y_4 , conforme mostra a figura 4.14.

Tabela 4.12 – Resultados das otimizações do método NBI

s	β_i	$1-\beta_i$	F_1	F_1	F_1	F_2	w_1	w_2	w_3
1	0	1	1,0000	0,0000	-0,1699	-1,1500	0,3090	0,2225	0,4686
2	0,05	0,95	0,9025	0,0025	-0,2664	-1,1471	0,3118	0,2412	0,4470
3	0,1	0,9	0,8100	0,0100	-0,3579	-1,1386	0,3147	0,2599	0,4254
4	0,15	0,85	0,7224	0,0224	-0,4446	-1,1243	0,3172	0,2788	0,4040
5	0,2	0,8	0,6399	0,0399	-0,5262	-1,1043	0,3198	0,2977	0,3825
6	0,25	0,75	0,5623	0,0623	-0,6030	-1,0785	0,3219	0,3168	0,3613
7	0,3	0,7	0,4897	0,0897	-0,6748	-1,0471	0,3246	0,3356	0,3398
8	0,35	0,65	0,4222	0,1222	-0,7416	-1,0098	0,3268	0,3547	0,3185
9	0,4	0,6	0,3596	0,1596	-0,8035	-0,9669	0,3289	0,3738	0,2973
10	0,45	0,55	0,3021	0,2021	-0,8604	-0,9181	0,3308	0,3931	0,2761
11	0,5	0,5	0,2496	0,2496	-0,9124	-0,8637	0,3326	0,4124	0,2550
12	0,55	0,45	0,2021	0,3021	-0,9594	-0,8034	0,3342	0,4317	0,2341
13	0,6	0,4	0,1596	0,3596	-1,0014	-0,7374	0,3356	0,4512	0,2132
14	0,65	0,35	0,1221	0,4221	-1,0385	-0,6657	0,3369	0,4707	0,1924
15	0,7	0,3	0,0897	0,4897	-1,0706	-0,5882	0,3379	0,4904	0,1717
16	0,75	0,25	0,0622	0,5622	-1,0978	-0,5050	0,3387	0,5101	0,1512
17	0,8	0,2	0,0398	0,6398	-1,1199	-0,4160	0,3392	0,5300	0,1308
18	0,85	0,15	0,0224	0,7224	-1,1372	-0,3213	0,3394	0,5500	0,1106
19	0,9	0,1	0,0099	0,8099	-1,1495	-0,2208	0,3394	0,5701	0,0905
20	0,95	0,05	0,0025	0,9025	-1,1569	-0,1146	0,3390	0,5903	0,0707
21	1	0	0,0000	1,0000	-1,1593	-0,0027	0,3383	0,6107	0,0511

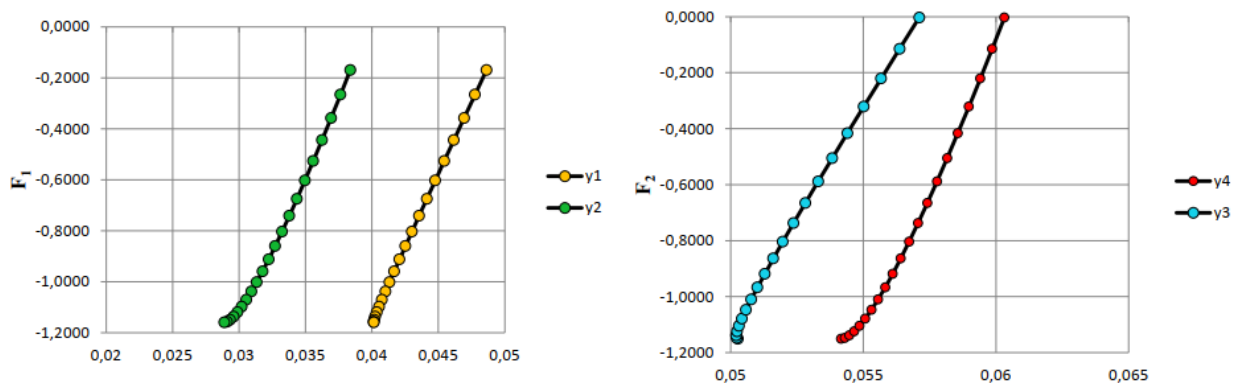


Figura 4.14 – Relação entre os fatores e as respostas

Fonte: autor.

Podemos observar na primeira linha da tabela 4.12 que foi conferido peso nulo para a função F_1 ($\beta_i = 0$) e peso unitário para F_2 ($1 - \beta_i = 1$) significando que para esta otimização somente F_2 foi considerada. Logo, neste caso foi dado maior valor para o peso que apresenta a rede MLP (peso w_3) por ser o método que mais contribui para minimizar o valor de F_2 , quando ele é adicionado na combinação (veja figura 4.13). Do mesmo modo, quando se atribui peso unitário para a função F_1 , somente F_1 participou da otimização, logo, a rede Linear terá um valor de peso (w_2) maior na combinação por ser o método que mais contribui para a minimização de F_1 . Veja que neste caso a rede MLP teve um

peso muito pequeno por ser a rede que mais contribui para o aumento da resposta F_1 . Partindo-se da solução Pareto-ótima $s = 1$ em direção as outras soluções é notório que a medida que mais peso se dá a função F_1 (e conseqüentemente menos peso para a função F_2) o peso referente a rede MLP vai ficando menor e o peso referente a rede Linear fica maior, como mostra a figura 4.15. Isto se deve ao fato de que a rede MLP é a rede que mais contribui para a diminuição de F_1 e a rede Linear contribui para a minimização de F_1 .

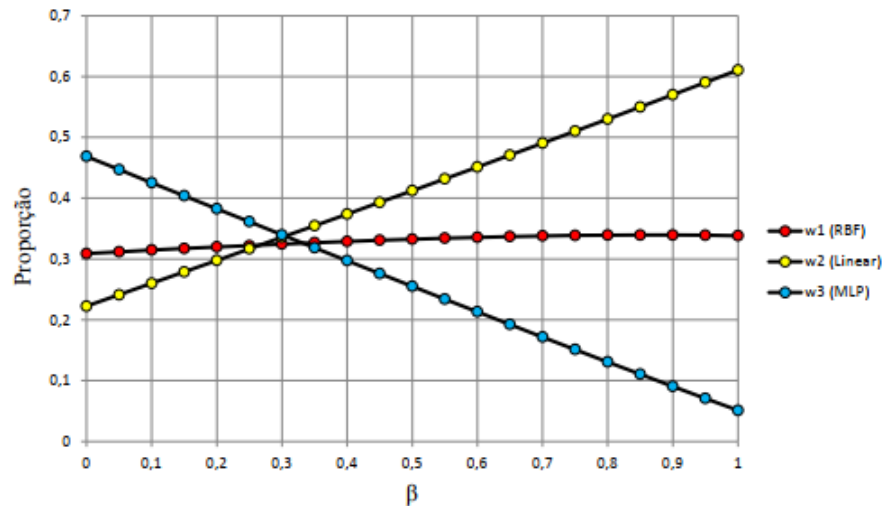


Figura 4.15 – Relação entre os pesos do método NBI com os pesos relacionados as RNAs
Fonte: autor.

Pelo gráfico 4.16 podemos observar que quanto maior o peso no processo de otimização é conferido a função F_1 , mais as respostas de F_1 se aproxima do seu ponto de utopia. Entretanto, isso faz com que as respostas de F_2 fiquem mais distantes da sua utopia e isso permite a conclusão de que é impossível atingir os ótimos individuais simultaneamente.

Para encontrar a melhor solução dentre as 21 soluções Pareto-ótimas obtidas pelo método de otimização NBI foi calculado para cada conjunto de pesos $(\beta_i, 1 - \beta_i)$ sua respectiva entropia H total conforme equação 2.38. Também foram calculados para cada par de solução (F_1, F_2) seus respectivos EPG (equação 2.39), considerando o alvo como sendo o ponto de utopia $F_U = (F_1^U; F_2^U) = (-1, 159; -1, 15)$ e por fim calculamos o índice $\frac{H}{EPG}$ para cada solução Pareto-ótima do método. Podemos identificar pela tabela 4.13, e de forma mais visual pela figura 4.17, que a solução $s = 11$ apresentou maior razão $\frac{H}{EPG}$, sendo portanto a solução escolhida.

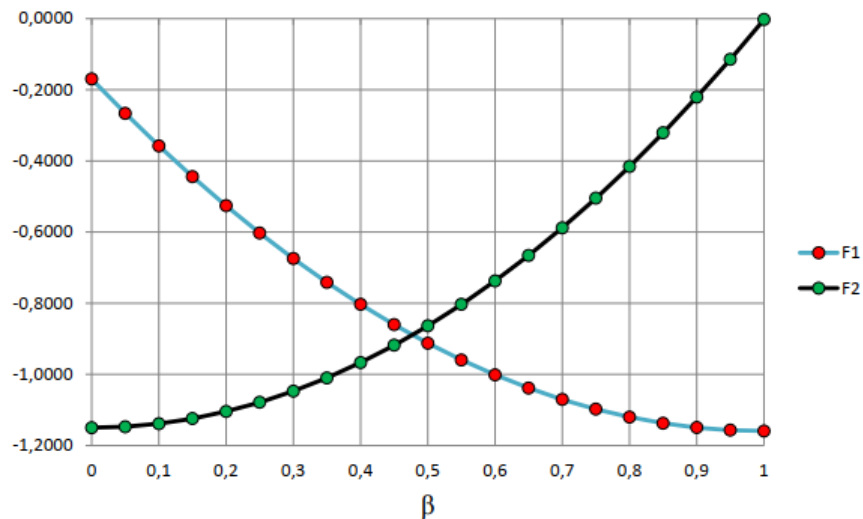


Figura 4.16 – Relação entre os pesos do método NBI com os fatores
Fonte: autor.

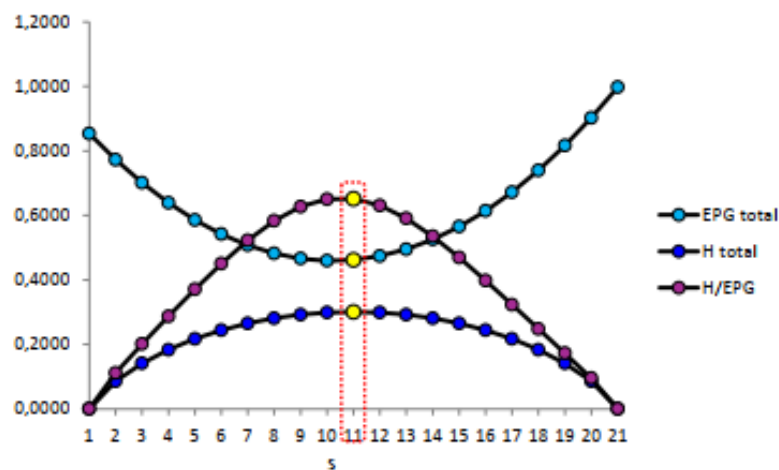


Figura 4.17 – Entropia, EPG e H/EPG em cada solução Pareto-ótima
Fonte: autor.

A figura 4.18 mostra as soluções pareto ótimas nos espaços de solução normalizado e original inerentes as funções objetivos utilizadas no problema de otimização, onde a solução em destaque corresponde a solução escolhida. Para a solução escolhida ($s = 11$) segundo o critério da máxima razão $\frac{H}{EPG}$, o conjunto de pesos associado para combinar as séries resíduais das redes RBF, Linear e MLP serão $w_1 = 0,3326$, $w_2 = 0,4124$ e $w_3 = 0,255$ respectivamente (veja tabela 4.12). Usando o critério da máxima razão $\frac{H}{EPG}$, a melhor solução dentre as soluções do foi a solução $s = 11$. É evidente que o vetor de soluções viáveis ($s = 11$) é pareto ótimo, uma vez que não existe outro ponto viável que possa reduzir qualquer uma das funções objetivo sem causar aumento simultâneo de pelo menos outra função objetivo. Observe que não é possível reduzir o valor de y_1 ou y_2 sem aumentar o valor de y_3 ou y_4 (figura 4.19).

Tabela 4.13 – Escolha do solução ótima segundo critério da máxima razão $\frac{H}{EPG}$

s	β_i	$1-\beta_i$	F_1	F_2	H	EPG	H/EPG
1	0	1	-0,1699	-1,1500	0,0004	0,8535	0,0005
2	0,05	0,95	-0,2664	-1,1471	0,0862	0,7727	0,1116
3	0,1	0,9	-0,3579	-1,1386	0,1412	0,7012	0,2013
4	0,15	0,85	-0,4446	-1,1243	0,1836	0,6389	0,2873
5	0,2	0,8	-0,5262	-1,1043	0,2173	0,5858	0,3710
6	0,25	0,75	-0,6030	-1,0785	0,2442	0,5420	0,4506
7	0,3	0,7	-0,6748	-1,0471	0,2653	0,5075	0,5228
8	0,35	0,65	-0,7416	-1,0098	0,2812	0,4822	0,5831
9	0,4	0,6	-0,8035	-0,9669	0,2923	0,4662	0,6270
10	0,45	0,55	-0,8604	-0,9181	0,2989	0,4595	0,6504
11	0,5	0,5	-0,9124	-0,8637	0,3010	0,4620	0,6516
12	0,55	0,45	-0,9594	-0,8034	0,2989	0,4738	0,6307
13	0,6	0,4	-1,0014	-0,7374	0,2923	0,4950	0,5905
14	0,65	0,35	-1,0385	-0,6657	0,2812	0,5254	0,5352
15	0,7	0,3	-1,0706	-0,5882	0,2653	0,5650	0,4695
16	0,75	0,25	-1,0978	-0,5050	0,2442	0,6140	0,3977
17	0,8	0,2	-1,1199	-0,4160	0,2173	0,6722	0,3233
18	0,85	0,15	-1,1372	-0,3213	0,1836	0,7397	0,2482
19	0,9	0,1	-1,1495	-0,2208	0,1412	0,8165	0,1729
20	0,95	0,05	-1,1569	-0,1146	0,0862	0,9024	0,0955
21	1	0	-1,1593	-0,0027	0,0004	0,9976	0,0004

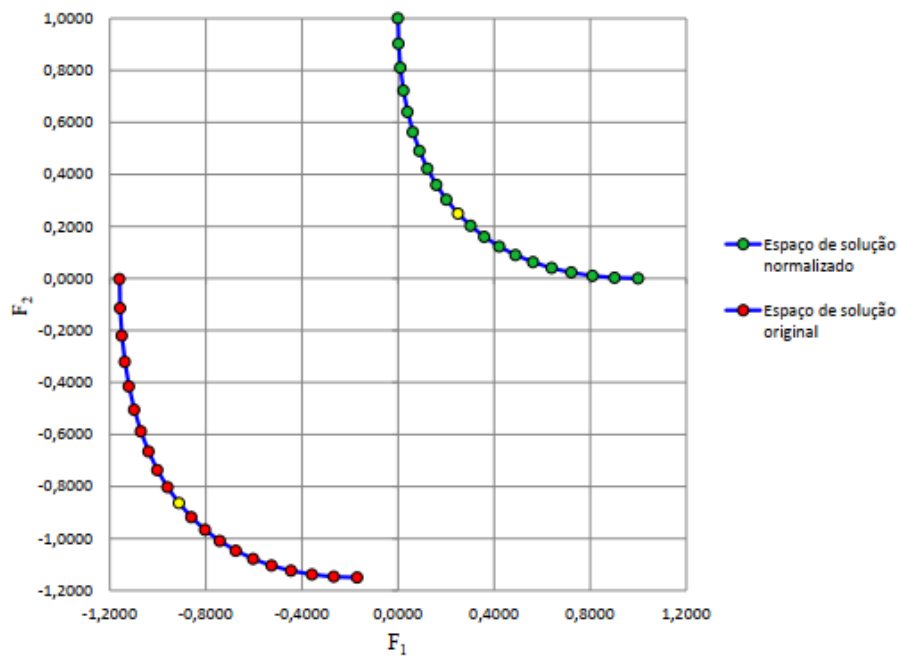


Figura 4.18 – Fronteira de Pareto $F_1 \times F_2$ nos espaços de solução normalizado e original
 Fonte: autor.

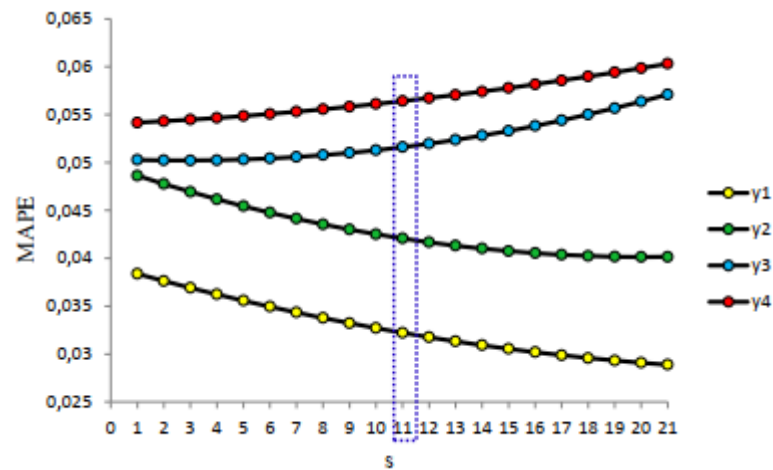


Figura 4.19 – Variação das respostas entre cada ponto da Fronteira de Pareto
 Fonte: autor.

A figura 4.20 mostra a dominância dos pontos da fronteira em relação a outras soluções viáveis dentro do espaço de solução.

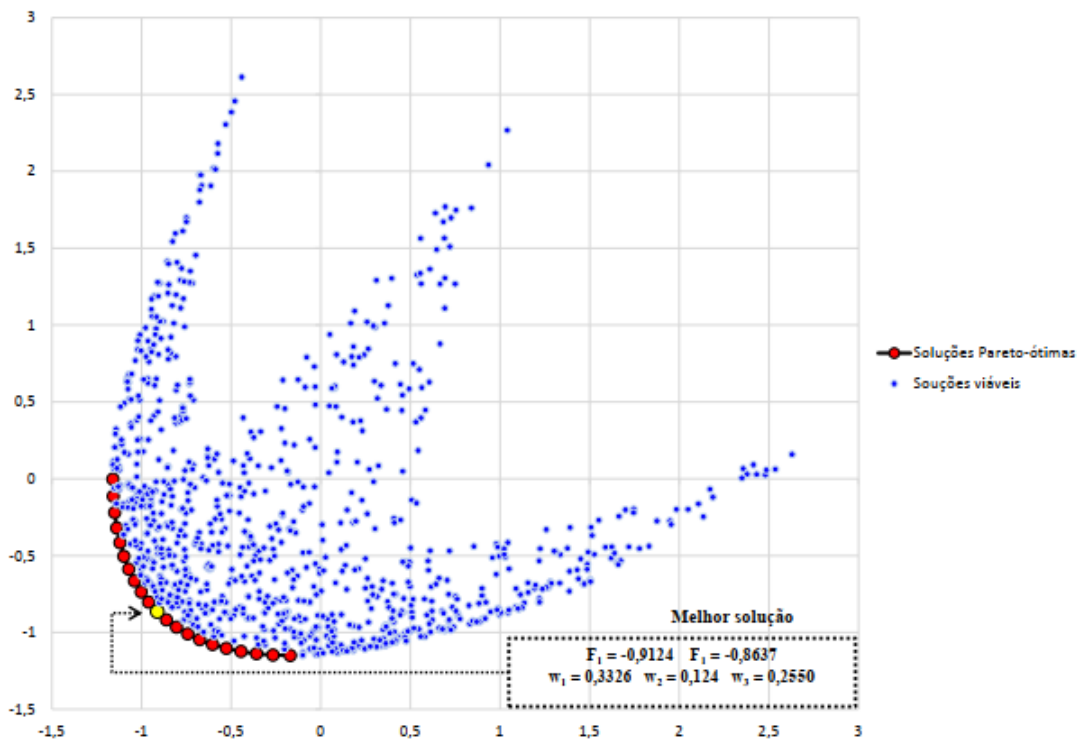


Figura 4.20 – Nuvem de pontos com soluções viáveis e Fronteira de Pareto
 Fonte: autor.

Podemos observar a localização das soluções Pareto-ótimas no gráfico de contorno sobreposto na figura 4.21. Os pontos das extremidades (destacados em preto) representam os valores das proporções de w_1 , w_2 e w_3 para os quais as funções F_1 e F_2 são mínimas. A medida que uma função vai tendo sua prioridade diminuída na otimização, as soluções vão "caminhando" de um ponto mínimo a outro. O ponto destacado em vermelho revela

a solução escolhida segundo o critério da máxima razão $\frac{H}{EPG}$, pertencendo a uma região viável. Por meio da equação

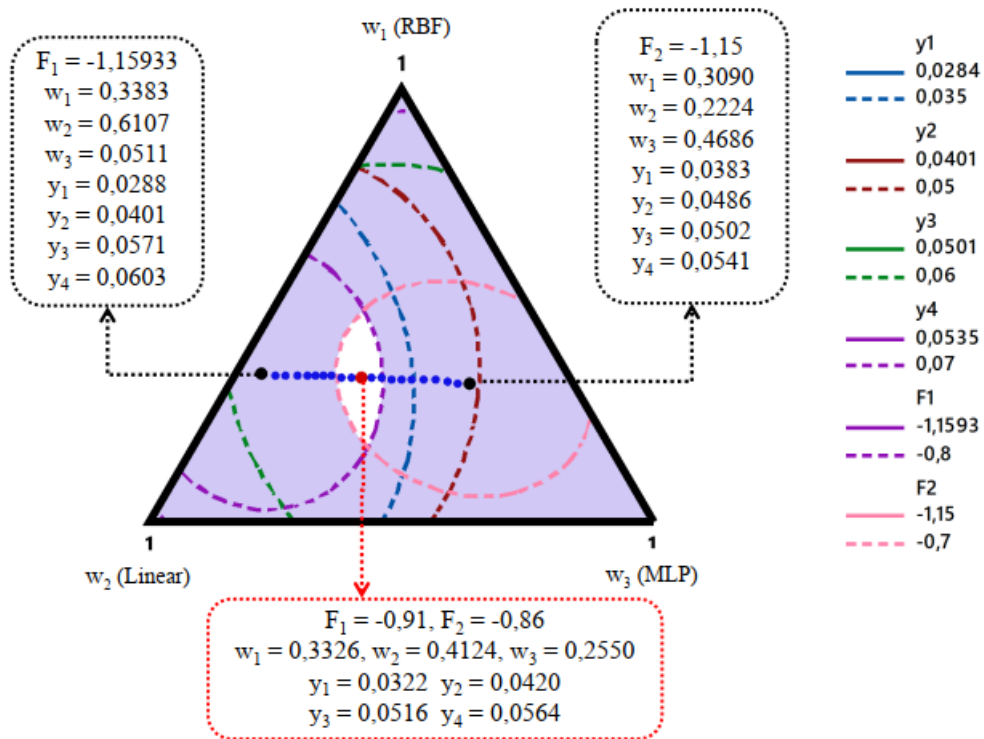


Figura 4.21 – Gráfico de contorno sobreposto
 Fonte: autor.

$$\mathbf{r}_{e_j} = 0,3326\mathbf{r}_{1j} + 0,4124\mathbf{r}_{2j} + 0,255\mathbf{r}_{3j} \quad (4.13)$$

$j = 1, \dots, 4$ a série residual fora da amostra para o ensemble foi calculada e os resultados em termos de MAPE foram calculados para previsões de 24, 48 e 72 pontos a frente, mostrados nas tabelas 4.14 e 4.15. Na figura 4.22 é exibida a previsão para 72 pontos a frente.

Tabela 4.14 – MAPE obtidos com base nos dados fora da amostra para as subestações 1 e 2

Pontos a frente	Subestação 1			Subestação 2		
	24	48	72	24	48	72
RBF	0,0426	0,0439	0,0390	0,0820	0,0791	0,0655
Linear	0,0329	0,0305	0,0299	0,0640	0,0572	0,0519
MLP	0,0435	0,0416	0,0357	0,1078	0,0805	0,0682
Ensemble	0,0270	0,0255	0,0228	0,0618	0,0560	0,0472

Tabela 4.15 – MAPE obtidos com base nos dados fora da amostra para as subestações 3 e 4

	Subestação 3			Subestação 4		
Pontos a frente	24	48	72	24	48	72
RBF	0,0457	0,0477	0,0558	0,0761	0,0925	0,1046
Linear	0,0542	0,0601	0,0663	0,1376	0,1332	0,1310
MLP	0,0425	0,0446	0,0544	0,0753	0,0853	0,1005
Ensemble	0,0399	0,0416	0,0398	0,0739	0,0791	0,0864

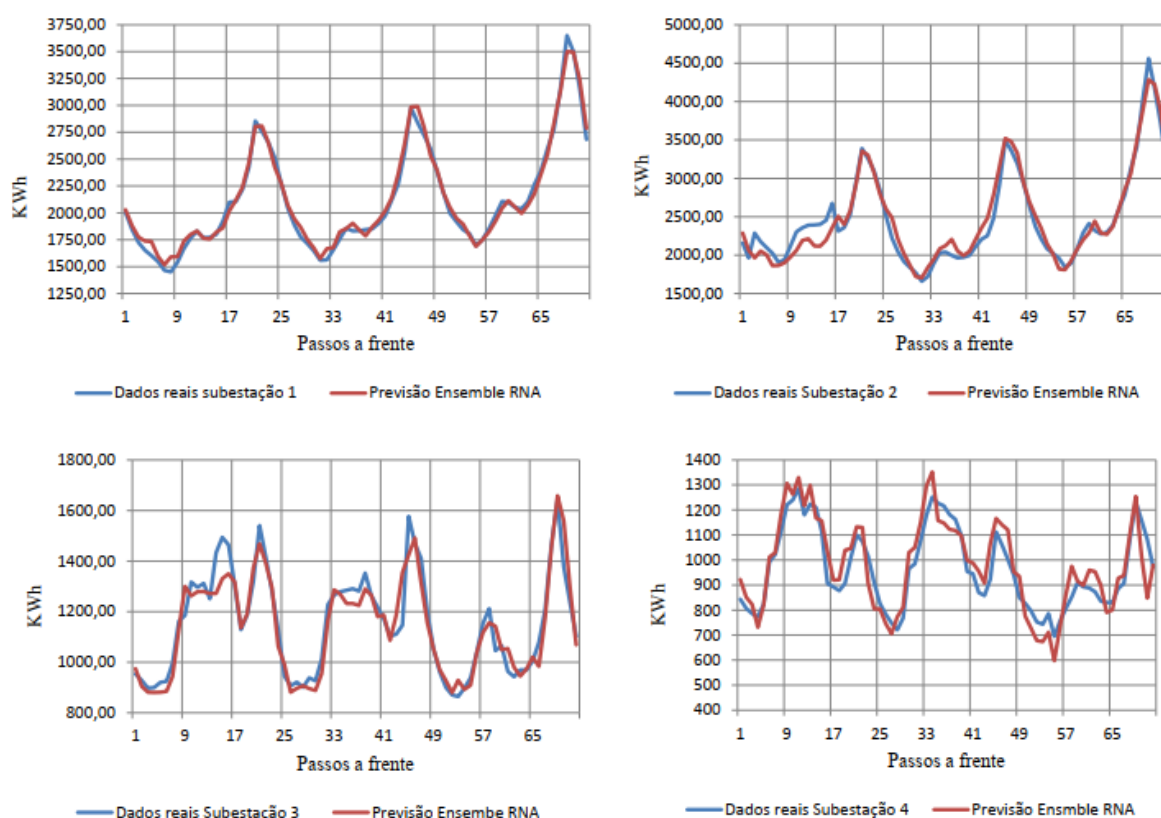


Figura 4.22 – Previsão de demanda de energia elétrica para as subestações
Fonte: autor.

Por meio da tabela 4.14 podemos notar que a combinação superou as redes individuais para todos estes horizontes de previsão. Pela análise que fizemos até aqui, vale destacar que individualmente a rede linear apresentou os melhores valores de MAPE para as subestações 1 e 2, pois como mostra a figura 4.12 a rede linear é a rede que mais contribuiu para o decréscimo da resposta F_1 (e portanto dos valores de MAPE das subestações 1 e 2). De forma contrária, para as subestações 3 e 4, a rede linear apresentou os piores resultados em termos de MAPE para todos os horizontes de previsão testados, pois a resposta do ensemble foi obtida através de uma combinação onde a resposta referente a rede linear recebeu o maior peso, acarretando desta forma para o acréscimo da função F_2 . Um raciocínio análogo para as subestações 3 e 4 podem ser aplicados, esclarecendo

os motivos pelos quais a rede MLP apresentou melhores resultados, em comparação com as demais redes.

5 CONCLUSÕES

Este trabalho abordou o problema de previsão de demanda de energia elétrica por meio do ensemble de redes neurais artificiais sob uma perspectiva totalmente voltada para o planejamento experimental. Em função da dificuldade de parametrizar uma rede neural, este estudo empregou a metodologia de planejamento experimental para melhorar o desempenho de redes neurais em uma previsão de carga elétrica de curto prazo para subestações de distribuição de energia elétrica. Por meio de um arranjo fatorial fracionário baseado em fatores relacionados a série temporal de demanda de energia elétrica e fatores relacionados à parametrização das redes neurais foi possível identificar como a forma de parametrizar a rede neural pode impactar na sua performance. Seis fatores de parametrização foram estudados neste trabalho: a forma de padronizar os dados das séries temporais de demanda de energia elétrica das subestações, utilizar ou não as variáveis explicativas (dia da semana e o período do dia em que a observação da série temporal foi registrada), critério de seleção de redes neurais, seleção de subconjuntos de variáveis independentes, seleção de subconjuntos para cada rede treinada e o número de etapas de tempo usadas como entradas na rede (esses 4 últimos são características avaliadas durante a configuração das redes neurais no software Statistica, modo IPS).

Com base nos resultados dos experimentos pode constatar-se, de forma individual, que a utilização das variáveis explicativas como entradas nas redes e a utilização da Transformada rápida de Fourier para a detecção da sazonalidade da série para determinar o número de etapas de tempo usadas como informações que melhoraram o desempenho das respostas analisadas (média e desvio padrão do erro percentual absoluto médio - MAPE). Algumas características inerentes a parametrização não foram significativas, mas participaram da parametrização da rede em função de suas interações com outros fatores serem significativos. Foi o caso do fator seleção de subconjuntos de variáveis independentes, que mesmo não sendo significativo para nenhuma das respostas, participou da parametrização em função da sua interação com a utilização das variáveis explicativas auxiliarem na redução do desvio padrão do MAPE. A partir desta análise, a ferramenta de otimização multiobjetivo desirability foi empregada para obter uma parametrização que buscasse um menor valor de média e desvio padrão para a medida de desempenho adotada neste trabalho, obtendo três redes neurais para fazerem parte do ensemble: uma rede Linear, uma rede Perceptron multicamadas e uma rede neural de função de base radial.

Um planejamento experimental também foi definido para estudar a participação de cada uma das redes neurais na saída do ensemble, formada pela combinação das saídas produzidas por cada uma das redes. Mais especificamente, um arranjo de mistura simplex

lattice de 3 componentes e grau 5, com três pontos axiais e um ponto central foi definido para combinar as séries residuais produzidas por cada uma das redes para cada uma das subestações, em que a resposta analisada nesta etapa foi a medida MAPE. Com base nesta medida, observou-se que a rede Linear obteve um rendimento melhor para duas subestações, enquanto a rede MLP obteve rendimento melhor para as outras duas. Em função da existência de correlação positiva e significativa entre as respostas MAPE para pares de subestações, para determinar um vetor de ponderação para combinar as séries residuais produzidas pelas redes neurais o método de otimização Interseção Normal a Fronteira (NBI) associado a técnica de análise fatorial exploratória foi aplicado. Por meio desta ferramenta, foi possível reduzir a dimensionalidade dos problema, onde dois fatores foram suficientes para explicar 97,9% da variância das respostas das quatro subestações reduzindo o número de funções objetivos de quatro para duas sem perda de informação, obtendo funções não correlacionadas e mantendo a estrutura de covariância das respostas originais. O processo de otimização bi-objetivo foi então aplicado a essas duas funções de escores rotacionados de fator, obtendo uma fronteira de Pareto contínua e equiespaçada com 21 soluções Pareto-ótimas, na qual uma solução foi escolhida usando o critério da máxima razão entre entropia e erro percentual global. Por este critério, a solução com máxima diversificação dos pesos relacionados as funções objetivo e mínima distância em relação ao pontos de mínimo das funções de escores de fator foi escolhida. Esta solução estava associada a um vetor de pesos que foram utilizados para combinar as séries residuais deixada fora da amostra, formando a saída do ensemble, que revelou ser superior às saídas produzidas pelas redes neurais para os horizontes de previsão de 24, 48 e 72 pontos a frente.

Como se pode observar, em problemas de previsão de séries temporais não lineares por meio do ensemble de redes neurais artificiais o planejamento experimental configura-se como uma boa estratégia para parametrizar as redes, bem como para determinar como as saídas serão combinadas, sendo mais confiável do que a famosa técnica da tentativa e erro.

Ao longo deste trabalho, os planejamentos experimentais para parametrização das redes neurais e para determinação da forma de combinação das saídas foram separados. Entretanto, é possível que esses dois planejamentos experimentais sejam acoplados. Isso possibilitará o estudo de possíveis interações entre os parâmetros das redes neurais com a ponderação utilizadas para a combinação das saídas, sendo portanto uma continuação deste trabalho.

REFERÊNCIAS

- 1 SHORT-TERM load forecasting based on artificial neural networks parallel implementation. *Electric Power Systems Research*, v. 63, n. 3, p. 185 – 196, 2002. Citado na página 1.
- 2 PARK, D. C. et al. Electric load forecasting using an artificial neural network. *IEEE Transactions on Power Systems*, v. 6, n. 2, p. 442–449, 1991. Citado na página 1.
- 3 HSU, Y.-Y. et al. Two-Stage Artificial Neural Network Model for Short-Term Load Forecasting. *IFAC-PapersOnLine*, Elsevier, v. 51, n. 28, p. 678–683, jan 2018. Citado na página 1.
- 4 BALESTRASSI, P. et al. Design of experiments on neural network’s training for nonlinear time series forecasting. *Neurocomputing*, Elsevier, v. 72, n. 4-6, p. 1160–1178, jan 2009. Citado 4 vezes nas páginas 1, 8, 17 e 50.
- 5 CHEN, M.-Y. et al. Design of experiments on neural network’s parameters optimization for time series forecasting in stock markets. *Neural Network World*, v. 23, p. 369–393, jan 2013. Citado na página 1.
- 6 J., W. M. B. Operations management research methodologies using quantitative modeling. *International Journal of Operations & Production Management*, MCB UP Ltd, v. 22, n. 2, p. 241–264, jan 2002. Citado na página 2.
- 7 TURRIONI J. B.; MELLP, C. H. P. *Metodologia de pesquisa em engenharia de produção*. [S.l.]: Universidade Federal de Itajubá, 2012. Citado 2 vezes nas páginas 2 e 3.
- 8 WANG, S. et al. Bi-directional long short-term memory method based on attention mechanism and rolling update for short-term load forecasting. *International Journal of Electrical Power & Energy Systems*, v. 109, p. 470–479, 2019. Citado na página 5.
- 9 BARINGO, L.; CONEJO, A. J. Transmission and Wind Power Investment. *IEEE Transactions on Power Systems*, v. 27, n. 2, p. 885–893, 2012. Citado na página 5.
- 10 ATWA, Y. M. et al. Optimal Renewable Resources Mix for Distribution System Energy Loss Minimization. *IEEE Transactions on Power Systems*, v. 25, n. 1, p. 360–370, 2010. Citado na página 5.
- 11 WANG, P.; LIU, B.; HONG, T. Electric load forecasting with recency effect: A big data approach. *International Journal of Forecasting*, Elsevier, v. 32, n. 3, p. 585–597, jul 2016. Citado na página 5.
- 12 HONG, T.; WANG, P. Fuzzy interaction regression for short term load forecasting. *Fuzzy Optimization and Decision Making*, v. 13, n. 1, p. 91–103, mar 2014. Citado na página 5.
- 13 Zor, K.; Timur, O.; Teke, A. A state-of-the-art review of artificial intelligence techniques for short-term electric load forecasting. p. 1–7, 2017. Citado na página 5.

- 14 AL-MUSAYLH, M. S. et al. Short-term electricity demand forecasting with MARS, SVR and ARIMA models using aggregated demand data in Queensland, Australia. *Advanced Engineering Informatics*, v. 35, p. 1–16, 2018. Citado na página 6.
- 15 CANCELO, J. R.; ESPASA, A.; GRAFE, R. Forecasting the electricity load from one day to one week ahead for the Spanish system operator. *International Journal of Forecasting*, v. 24, n. 4, p. 588–602, 2008. Citado na página 6.
- 16 TAKEDA, H.; TAMURA, Y.; SATO, S. Using the ensemble Kalman filter for electricity load forecasting and analysis. *Energy*, v. 104, p. 184–198, 2016. Citado na página 6.
- 17 LI, F. A. N. Singular value decomposition expansion for electrical demand analysis. *IMA Journal of Management Mathematics*, n. 1, p. 37–48, 2000. Citado na página 6.
- 18 ABU-SHIKHAH, N.; ELKARMI, F. Medium-term electric load forecasting using singular value decomposition. *Energy*, v. 36, n. 7, p. 4259–4271, 2011. Citado na página 6.
- 19 RENDON-SANCHEZ, J. F.; MENEZES JOURNAL = European Journal of Operational Research, n. . . p. . . t. . S. v. . . y. . . de. Citado na página 6.
- 20 VERDEJO, H. et al. Statistic linear parametric techniques for residential electric energy demand forecasting. A review and an implementation to Chile. *Renewable and Sustainable Energy Reviews*, v. 74, p. 512–521, 2017. Citado na página 6.
- 21 Chen, Bo-Juen; Chang, Ming-Wei; Lin, C.-J. Load forecasting using support vector Machines: a study on EUNITE competition 2001. *IEEE Transactions on Power Systems*, v. 19, n. 4, p. 1821–1830, nov 2004. Citado na página 6.
- 22 NIU, D.; WANG, Y.; WU, D. D. Power load forecasting using support vector machine and ant colony optimization. *Expert Systems with Applications*, v. 37, n. 3, p. 2531–2539, 2010. Citado na página 6.
- 23 YANG, A.; LI, W.; YANG, X. Short-term electricity load forecasting based on feature selection and Least Squares Support Vector Machines. *Knowledge-Based Systems*, v. 163, p. 159–173, 2019. Citado na página 6.
- 24 BAHRAMI, S.; HOOSHMAND, R.-A.; PARASTEGARI, M. Short term electric load forecasting by wavelet transform and grey model improved by PSO (particle swarm optimization) algorithm. *Energy*, v. 72, p. 434–442, 2014. Citado na página 6.
- 25 ZHENG, T.; GIRGIS, A. A.; MAKRAM, E. B. A hybrid wavelet-Kalman filter method for load forecasting. *Electric Power Systems Research*, v. 54, n. 1, p. 11–17, 2000. Citado na página 6.
- 26 RANAWEERA, D. K.; HUBELE, N. F.; KARADY, G. G. Fuzzy logic for short term load forecasting. *International Journal of Electrical Power & Energy Systems*, v. 18, n. 4, p. 215–222, 1996. Citado na página 6.
- 27 MAMLOOK, R.; BADRAN, O.; ABDULHADI, E. A fuzzy inference model for short-term load forecasting. *Energy Policy*, v. 37, n. 4, p. 1239–1248, 2009. Citado na página 6.

- 28 PAPADAKIS, S. E.; THEOCHARIS, J. B.; BAKIRTZIS, A. G. A load curve based fuzzy modeling technique for short-term load forecasting. *Fuzzy Sets and Systems*, v. 135, n. 2, p. 279–303, 2003. Citado na página 6.
- 29 DINIZ, C. A. *Otimização Estrutural de Elementos em Compósitos usando Redes Neurais Artificiais*. Dissertação (Mestrado) — Universidade Federal de Itajubá, Fevereiro 2017. Citado na página 7.
- 30 MCCULLOCH, W. S.; PITTS, W. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, v. 5, n. 4, p. 115–133, Dec 1943. Citado na página 7.
- 31 TSAY, R. S. *Analysis of financial time series*. 2. ed. ed. Hoboken, NJ: Wiley-Interscience, 2010. (Wiley series in probability and statistics). Citado na página 7.
- 32 FAUSETT, L. (Ed.). *Fundamentals of Neural Networks: Architectures, Algorithms, and Applications*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1994. Citado na página 7.
- 33 MURALITHARAN, K.; SAKTHIVEL, R.; VISHNUVARTHAN, R. Neural network based optimization approach for energy demand prediction in smart grid. *Neurocomputing*, v. 273, p. 199–208, 2018. Citado na página 8.
- 34 SRINIVASAN, D. Evolving artificial neural networks for short term load forecasting. *Neurocomputing*, v. 23, n. 1, p. 265–276, 1998. Citado na página 9.
- 35 KONG, W. et al. Short-Term Residential Load Forecasting Based on LSTM Recurrent Neural Network. *IEEE Transactions on Smart Grid*, v. 10, n. 1, p. 841–851, jan 2019. Citado na página 9.
- 36 DUDEK, G. Neural networks for pattern-based short-term load forecasting: A comparative study. *Neurocomputing*, v. 205, p. 64–74, 2016. Citado na página 9.
- 37 KHWAJA, A. S. et al. Improved short-term load forecasting using bagged neural networks. *Electric Power Systems Research*, v. 125, p. 109–115, 2015. Citado na página 9.
- 38 QUAN, H.; SRINIVASAN, D.; KHOSRAVI, A. Short-Term Load and Wind Power Forecasting Using Neural Network-Based Prediction Intervals. *IEEE Transactions on Neural Networks and Learning Systems*, v. 25, n. 2, p. 303–315, 2014. Citado na página 9.
- 39 YAO, S. J. et al. Wavelet transform and neural networks for short-term electrical load forecasting. *Energy Conversion and Management*, v. 41, n. 18, p. 1975–1988, 2000. Citado na página 9.
- 40 GHAYEKHLOO, M.; MENHAJ, M. B.; GHOFRANI, M. A hybrid short-term load forecasting with a new data preprocessing framework. *Electric Power Systems Research*, v. 119, p. 138–148, 2015. Citado na página 9.
- 41 LIU, Z.; LI, W.; SUN, W. A novel method of short-term load forecasting based on multiwavelet transform and multiple neural networks. *Neural Computing and Applications*, v. 22, n. 2, p. 271–277, 2013. Citado na página 9.

- 42 ÇEVİK, H. H.; ÇUNKACS, M. Short-term load forecasting using fuzzy logic and ANFIS. *Neural Computing and Applications*, v. 26, n. 6, p. 1355–1367, 2015. Citado na página 9.
- 43 KHOTANZAD, A.; and H. Elragal. A neuro-fuzzy approach to short-term load forecasting in a price-sensitive environment. *IEEE Transactions on Power Systems*, v. 17, n. 4, p. 1273–1282, nov 2002. Citado na página 9.
- 44 ZAHEDI, G. et al. Electricity demand estimation using an adaptive neuro-fuzzy network: A case study from the Ontario province – Canada. *Energy*, Pergamon, v. 49, p. 323–328, jan 2013. Citado na página 9.
- 45 LIU, Y. et al. A multivariate forecasting method for short-term load using chaotic features and RBF neural network. *European Transactions on Electrical Power*, v. 21, n. 3, p. 1376–1391, 2011. Citado na página 9.
- 46 KOUHI, S.; KEYNIA, F.; RAVADANEGH, S. N. A new short-term load forecast method based on neuro-evolutionary algorithm and chaotic feature selection. *International Journal of Electrical Power & Energy Systems*, v. 62, p. 862–867, 2014. Citado na página 9.
- 47 AWAN, S. M. et al. An efficient model based on artificial bee colony optimization algorithm with Neural Networks for electric load forecasting. *Neural Computing and Applications*, v. 25, n. 7, p. 1967–1978, 2014. Citado na página 9.
- 48 QUAN, H.; SRINIVASAN, D.; KHOSRAVI, A. Uncertainty handling using neural network-based prediction intervals for electrical load forecasting. *Energy*, v. 73, p. 916–925, 2014. Citado na página 9.
- 49 EL-TELBANY, M.; EL-KARMI, F. Short-term forecasting of Jordanian electricity demand using particle swarm optimization. *Electric Power Systems Research*, v. 78, n. 3, p. 425–433, 2008. Citado na página 9.
- 50 HANSEN, L. K.; SALAMON, P. Neural network ensembles. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 12, n. 10, p. 993–1001, 1990. Citado na página 9.
- 51 FILHO, M. de A. L. *Previsão de carga de curto prazo usando ensembles de previsores selecionados e evoluídos por algoritmos genéticos*. Dissertação (Mestrado) — Universidade Estadual de Campinas, Janeiro 2006. Citado na página 10.
- 52 JOVANOVIĆ, R. Ž.; SRETENOVIĆ, A. A.; ŽIVKOVIĆ, B. D. Ensemble of various neural networks for prediction of heating energy consumption. *Energy and Buildings*, Elsevier, v. 94, p. 189–199, may 2015. Citado na página 10.
- 53 SAVIOZZI, M.; MASSUCCO, S.; SILVESTRO, F. Implementation of advanced functionalities for Distribution Management Systems: Load forecasting and modeling through Artificial Neural Networks ensembles. *Electric Power Systems Research*, v. 167, p. 230–239, 2019. Citado na página 10.
- 54 CERVONE, G. et al. Short-term photovoltaic power forecasting using Artificial Neural Networks and an Analog Ensemble. *Renewable Energy*, v. 108, p. 274–286, 2017. Citado na página 10.

- 55 JETCHEVA, J. G.; MAJIDPOUR, M.; CHEN, W.-P. Neural network model ensembles for building-level electricity load forecasts. *Energy and Buildings*, v. 84, p. 214–223, 2014. Citado na página 10.
- 56 MONTGOMERY, D. C. *Design and Analysis of Experiments*. [S.l.]: John Wiley & Sons, 2017. ISBN 9780470128664. Citado 10 vezes nas páginas 10, 11, 12, 14, 15, 18, 19, 20, 47 e 52.
- 57 KLEIJNEN, J. P. C. *Experimental Design for Sensitivity Analysis, Optimization, and Validation of Simulation Models*. [S.l.]: John Wiley Sons, 2007. 173-223 p. Citado 2 vezes nas páginas 11 e 16.
- 58 KELTON, W. D. *Simulation with Arena (McGraw-Hill Series in Industrial Engineering and Management)*. [S.l.]: McGraw-Hill Higher Education, 2007. Citado na página 11.
- 59 BOX, G. E. P.; BEHNKEN, D. W. Some new three level designs for the study of quantitative variables. *Technometrics*, Taylor Francis, v. 2, n. 4, p. 455–475, 1960. Citado na página 12.
- 60 DEVOR, R.; TH, C.; SUTHERLAND, J. [S.l.]: New Jersey, Prentice Hall, Inc. Citado na página 12.
- 61 STEINBERG, D. M.; HUNTER, W. G. Experimental design: Review and comment. *Technometrics*, Taylor Francis, v. 26, n. 2, p. 71–97, 1984. Citado na página 15.
- 62 COLEMAN, D. E.; MONTGOMERY, D. C. A Systematic Approach to Planning for a Designed Industrial Experiment. *Technometrics*, Taylor & Francis, v. 35, n. 1, p. 1–12, 1993. Citado 3 vezes nas páginas 15, 16 e 41.
- 63 KHAW, J. F. C.; LIM, B. S.; LIM, L. E. N. Optimal design of neural networks using the Taguchi method. *Neurocomputing*, v. 7, n. 3, p. 225–245, 1995. Citado na página 17.
- 64 PONTES, F. et al. Design of experiments and focused grid search for neural network parameter optimization. *Neurocomputing*, Elsevier, v. 186, p. 22–34, apr 2016. Citado na página 17.
- 65 STATSOFT Inc., Statistica (data analysis software system), version 7.1, 2005. 2005. Citado na página 17.
- 66 SAVIOZZI, M.; MASSUCCO, S.; SILVESTRO, F. Implementation of advanced functionalities for Distribution Management Systems: Load forecasting and modeling through Artificial Neural Networks ensembles. *Electric Power Systems Research*, Elsevier, v. 167, p. 230–239, feb 2019. Citado na página 17.
- 67 GOMES, J. H. *Método dos polinômios canônicos de misturas para otimização multi-objetivo*. Tese (Doutorado) — Universidade Federal de Itajubá, 2013. Citado 6 vezes nas páginas 18, 21, 23, 24, 26 e 28.
- 68 BOX, G. E. P.; DRAPER, N. R. *Empirical Model-Building and Response Surfaces*. USA: Wiley, 1987. Citado na página 19.

- 69 AMIRI, A.; SABOUR, M. R. Multi-response optimization of fenton process for applicability assessment in landfill leachate treatment. *Waste Management*, v. 34, n. 12, p. 2528 – 2536, 2014. Citado na página 19.
- 70 CORNELL, J. *Experiments with Mixtures*. third. [S.l.]: John Wiley & Sons, 2002. Citado 4 vezes nas páginas 20, 22, 29 e 30.
- 71 MYERS R. H. E MONTGOMERY, D. C. *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*. fourth. [S.l.]: John Wiley & Sons, 2016. Citado na página 23.
- 72 PAIVA, A. *Metodologia de superfície de resposta e análise de componentes principais em otimização de processos de manufatura com múltiplas respostas correlacionadas*. Tese (Doutorado) — Universidade Federal de Itajubá, 2006. Citado 3 vezes nas páginas 25, 30 e 31.
- 73 BEREZKIN, V. E.; LOTOV, A. V. Comparison of two Pareto frontier approximations. *Computational Mathematics and Mathematical Physics*, v. 54, n. 9, p. 1402–1410, 2014. Citado na página 25.
- 74 RAO, S. S. *Engineering optimization: theory and practice*. 4. ed. [S.l.]: John Wiley & Sons, 2009. Citado na página 25.
- 75 LI, D.; YANG, J.-B.; BISWAL, M. Quantitative parametric connections between methods for generating noninferior solutions in multiobjective optimization. *European Journal of Operational Research*, v. 117, n. 1, p. 84 – 99, 1999. Citado 2 vezes nas páginas 25 e 28.
- 76 LEITE, R. R. *Método de interseção normal à fronteira para modelos quadráticos de escores fatoriais rotacionados*. Dissertação (Mestrado) — Universidade Federal de Itajubá, Fevereiro 2019. Citado 3 vezes nas páginas 26, 27 e 35.
- 77 VAHIDINASAB, V.; JADID, S. Normal boundary intersection method for suppliers' strategic bidding in electricity markets: An environmental/economic approach. *Energy Conversion and Management*, v. 51, n. 6, p. 1111 – 1119, 2010. Citado 3 vezes nas páginas 28, 33 e 35.
- 78 DERRINGER, G.; SUICH, R. Simultaneous optimization of several response variables. *Journal of Quality Technology*, Taylor Francis, v. 12, n. 4, p. 214–219, 1980. Citado 2 vezes nas páginas 28 e 29.
- 79 HARRINGTON, E. The desirability function. *Industrial Quality Control*, n. 21, p. 494–498, 1965. Citado na página 28.
- 80 GYSEGHEM, E. V. et al. Selection of reversed-phase liquid chromatographic columns with diverse selectivity towards the potential separation of impurities in drugs. *Journal of Chromatography A*, v. 1042, n. 1, p. 69 – 80, 2004. Citado na página 28.
- 81 OLIVEIRA, F. *Estratégia de comercialização de energia elétrica através da otimização de portfólios de contratos utilizando projetos de experimentos de misturas*. Tese (Doutorado) — Universidade Federal de Itajubá, 2009. Citado na página 29.

- 82 DAS, I.; DENNIS, J. Normal-boundary intersection: A new method for generating the pareto surface in nonlinear multicriteria optimization problems. *SIAM Journal on Optimization*, v. 8, 07 2000. Citado 3 vezes nas páginas 31, 33 e 34.
- 83 CHARWAND, M. et al. Benders decomposition and normal boundary intersection method for multiobjective decision making framework for an electricity retailer in energy markets. *IEEE Systems Journal*, v. 9, n. 4, p. 1475–1484, 2015. Citado na página 31.
- 84 PAULA, T. I. de. *Método da interseção normal à fronteira baseado em análise fatorial para otimização de problemas multivariados utilizando-se algoritmo genético*. Tese (Doutorado) — Universidade Federal de Itajubá, 2019. Citado na página 32.
- 85 SHUKLA, P. K.; DEB, K. On finding multiple pareto-optimal solutions using classical and evolutionary generating methods. *European Journal of Operational Research*, v. 181, n. 3, p. 1630 – 1652, 2007. Citado na página 33.
- 86 JIA, Z.; IERAPETRITOU, M. G. Generate pareto optimal solutions of scheduling problems using normal boundary intersection technique. *Computers Chemical Engineering*, v. 31, n. 4, p. 268 – 280, 2007. Citado 2 vezes nas páginas 33 e 35.
- 87 GANESAN, T.; VASANT, P.; ELAMVAZUTHI, I. Normal-boundary intersection based parametric multi-objective optimization of green sand mould system. *Journal of Manufacturing Systems*, v. 32, n. 1, p. 197 – 205, 2013. Citado na página 35.
- 88 PAIXAO, K. et al. Normal boundary intersection applied in the scale-up for the treatment process of eriochrome black t through the uv/tio2/o3 system. *Journal of Environmental Chemical Engineering*, v. 7, n. 1, p. 102801, 2019. Citado na página 35.
- 89 NAVES, F. L. *Método de interseção normal de fronteira por análise fatorial aplicado ao tratamento de efluentes*. Tese (Doutorado) — Universidade Federal de Itajubá, 2017. Citado na página 35.
- 90 TABOADA, H. A. et al. Practical solutions for multi-objective optimization: An application to system reliability design problems. *Reliability Engineering System Safety*, v. 92, n. 3, p. 314 – 322, 2007. Selected Papers Presented at the Fourth International Conference on Quality and Reliability. Citado na página 36.
- 91 ROCHA, L. C. S. et al. Entropy-Based Weighting for Multiobjective Optimization: An Application on Vertical Turning. *Mathematical Problems in Engineering*, Hindawi Publishing Corporation, p. 608325, 2015. Citado na página 36.
- 92 SHANNON, C. E. A mathematical theory of communication. *Bell Syst. Tech. J.*, v. 27, n. 3, p. 379–423, 1948. Citado na página 36.
- 93 ROCHA, L. C. S. *Método da ponderação entrópica para otimização multiobjetivo robusta de modelos de superfície de resposta*. Tese (Doutorado) — Universidade Federal de Itajubá, 2017. Citado na página 36.
- 94 FANG S.; RAJASEKERA, J. R. H.-S. J. *Entropy optimization and mathematical programming*. [S.l.]: Kluwer Academic Publishers, 1997. Citado na página 36.
- 95 JOHNSON R. A.; WICHERN, D. W. *Applied multivariate statistical analysis*. 6. ed. [S.l.]: Prentice Hall, 2007. Citado 2 vezes nas páginas 38 e 39.

-
- 96 JOLLIFFE, L. *Principal component analysis*. 2. ed. [S.l.]: New York: Springer series in statistics, 2002. Citado na página 39.
- 97 KAISER, H. F. The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, v. 20, n. 1, p. 141–151, 1960. Citado 2 vezes nas páginas 39 e 63.
- 98 ARMSTRONG, J.; COLLOPY, F. Error measures for generalizing about forecasting methods: Empirical comparisons. *International Journal of Forecasting*, Elsevier, v. 8, n. 1, p. 69–80, jun 1992. Citado na página 44.