



**DOUTORADO EM ENGENHARIA DE PRODUÇÃO**  
**UNIVERSIDADE FEDERAL DE ITAJUBÁ**

**Luiz Gustavo de Mello**

**COMBINAÇÃO ÓTIMA DE MÉTODOS DE PREVISÃO SEGUNDO O  
CRITÉRIO PAYOFF-JOLLIFFE FATORIAL: UMA ABORDAGEM  
MULTIVARIADA PARA A ESTIMAÇÃO DE DEMANDA DE GÁS  
NATURAL**

**Itajubá – MG**

**2021**



**DOUTORADO EM ENGENHARIA DE PRODUÇÃO**  
**UNIVERSIDADE FEDERAL DE ITAJUBÁ**

**Luiz Gustavo de Mello**

**COMBINAÇÃO ÓTIMA DE MÉTODOS DE PREVISÃO SEGUNDO O  
CRITÉRIO PAYOFF-JOLLIFFE FATORIAL: UMA ABORDAGEM  
MULTIVARIADA PARA A ESTIMAÇÃO DE DEMANDA DE GÁS  
NATURAL**

Tese submetida ao Programa de Pós-Graduação em Engenharia de Produção como requisito parcial para obtenção do Título de Doutor em Ciências da Engenharia de Produção.

**Área de Concentração: Engenharia de Produção**

**Orientador: Prof. Dr. Anderson Paulo de Paiva.**

**Itajubá – MG**

**2021**



**DOUTORADO EM ENGENHARIA DE PRODUÇÃO**  
**UNIVERSIDADE FEDERAL DE ITAJUBÁ**

**Luiz Gustavo de Mello**

**COMBINAÇÃO ÓTIMA DE MÉTODOS DE PREVISÃO SEGUNDO O  
CRITÉRIO PAYOFF-JOLLIFFE FATORIAL: UMA ABORDAGEM  
MULTIVARIADA PARA A ESTIMAÇÃO DE DEMANDA DE GÁS  
NATURAL**

Tese aprovada por Banca Examinadora em 15 de julho de 2021, conferindo ao autor o Título de Doutor em Ciências da Engenharia de Produção.

**Banca Examinadora:**

Prof. Dr. Anderson Paulo de Paiva (Orientador - UNIFEI)

Profª Dra. Juliana Helena Daroz Gaudêncio (UNIFEI)

Prof. Dr. Pedro Paulo Balestrassi (UNIFEI)

Prof. Dr. Antonio Fernando Branco Costa (UNIFEI)

Prof. Dr. Fabrício José Pontes (EMBRAER)

Prof. Dr. Paulo Rotela Júnior (UFPB)

**Itajubá – MG**

**2021**

## DEDICATÓRIA

Desde sempre, tive minha mãe como minha protetora e grande incentivadora em todos os momentos de dificuldades pelos que passei. Infelizmente, por designos da vida, minha mãe faleceu pouco antes da defesa e do término do trabalho. Assim, esta produção é dedicada a você, minha mãe Nilza Gomes Barbosa de Mello, que tanto me faz falta, mas sei que está me acompanhando. Obrigado por tudo!

## AGRADECIMENTOS

Ensina-me o bom senso e o conhecimento, pois confio em teus mandamentos (Salmos 119:66). Agradeço a Deus ter me proporcionado os caminhos, o conhecimento, a saúde e a motivação para trilhar esse árduo, mas recompensador caminho do Doutorado.

Meus pais, dentro dos limites que a vida os impôs, sempre me deram amor, educação e atenção. A eles, então, agradeço por terem sido esteio sempre em minha vida. Obrigado, meu pai Ivo de Mello e minha mãe Nilza Gomes Barbosa de Mello.

Minha trajetória acadêmica, profissional e, conseqüentemente, pessoal só foi possível graças ao meu Orientador, Professor Doutor Anderson Paulo de Paiva. Desde o Mestrado, seu conhecimento, benevolência, atenção, afeição e profissionalismo serviram como exemplo e incentivaram minha continuidade e conclusão do Doutorado. Agradeço-lhe imensamente, Professor Anderson. Sua atitude alterou toda a minha trajetória de vida para poder me tornar o que sou. Você sempre estará em minhas orações. Não esquecerei toda a ajuda que você me ofertou e sempre estarei à disposição para retribuir e para manter essa corrente de colaboração em contínua atividade.

Agradeço imensamente toda a contribuição recebida dos professores: Prof. Dr. Pedro Paulo Balestrassi, Prof. Dra. Juliana Helena Daroz Gaudêncio, Prof. Dr. Antonio Fernando Branco Costa, Prof. Dr. Fabrício José Pontes, Prof. Dr. Paulo Rotela Júnior. Vocês foram fundamentais durante o período de realização do doutorado e nos processos de Qualificação e de Defesa. Certamente, seus apontamentos contribuíram com o aperfeiçoamento da pesquisa.

Faço um agradecimento especial aos colegas de Doutorado e ao Grupo de Pesquisa em Engenharia, Prof. Dr. Livio Agnew Bacci e Prof. Dr. Fabrício Alves de Almeida, serei grato pela parceria e pela colaboração de ambos. Contem sempre comigo. Não esquecerei nunca da contribuição de vocês.

Agradeço a todos os servidores do Instituto Federal do Sul de Minas Gerais que me apoiaram e me entenderam diante de alguma falha no trabalho, proveniente da dedicação conjunta às atribuições profissionais e da realização do Doutorado. Faço um agradecimento especial ao Diretor-geral do Campus Av. Carmo de Minas, Professor Dr. João Olympio de Araújo Neto, que me acompanhou e prestou todo o apoio necessário nesse processo. Além de colega de trabalho, tornou-se um amigo fraterno.

Nesse caminho, algumas pessoas foram importantes e, por essa razão, faço um agradecimento a elas pelos momentos de apoio ofertados. Minha ex-esposa Eliana Paulino da Silva, que, durante o nosso casamento, acompanhou quase todo o período do Doutorado e me ajudou em

diversos momentos. Meus amigos e professores: Prof. Dr. Livio Agnew Bacci, Prof. MSc. Guido Agnew Bacci, Prof. MSc. Pedro Chaib Bernardes e Prof. MSc. José Celso de Castro. Amigos são imprescindíveis nesse processo, e vocês foram muito importantes.

Destaco, ainda, o Prof. Dr. Renato da Silva Lima. O senhor, dentro do profissionalismo exigido pelo cargo de Coordenador do Programa de Doutorado, compreendeu-me, orientou-me e contribuiu fundamentalmente para que eu pudesse concluir minha tese. Muito obrigado por toda a colaboração, da qual nunca me esquecerei, afinal esse suporte foi essencial nos momentos difíceis por que passei.

Agradeço, também, à Universidade Federal de Itajubá. Essa Instituição mudou minha vida. Tive a oportunidade de concluir 2 Pós-graduações Lato Sensu, o Mestrado e o Doutorado. Essa trajetória acadêmica na Instituição dividiu minha vida em antes e depois do início dos estudos na Instituição. Sem o apoio, ensinamentos, compreensão e suporte em todos os momentos, essa trajetória pessoal e profissional não teria seguido o caminho atual. Meu muito obrigado a todos!

## RESUMO

Este trabalho apresenta um método de otimização multiobjetivo não-linear para definir pesos ótimos para a combinação de métodos de previsões de séries temporais usadas para a estimação de demanda anual de gás natural. A abordagem de alocação de pesos emprega um arranjo de experimentos de misturas para modelar a relação entre diversas métricas de desempenho de previsão e os pesos atribuídos aos resíduos de previsão dos métodos de séries temporais individuais escolhidos para formar a combinação. Nesta pesquisa, foram considerados os métodos de Alisamento Exponencial Duplo (DES) e os de Holt-Winters aditivo (WA) e multiplicativo (WM). As diversas métricas de desempenho relacionadas à localização, à dispersão e à diversidade foram modeladas por polinômios canônicos de misturas e, posteriormente, otimizadas individualmente, formando uma matriz *Payoff* das soluções individuais. Estas foram, então, agrupadas de acordo com a mínima distância entre os pontos de ótimo e os critérios de Jolliffe, definidos segundo uma análise de componentes principais (PCA) e aplicados a cada grupo identificado para uma primeira seleção das métricas não-redundantes (Critérios *Payoff*-Jolliffe). Às métricas remanescentes aplicou-se uma análise fatorial (FA) com extração por componentes principais e rotação varimax, armazenando-se os escores rotacionados dos fatores obtidos. Após a modelagem desses escores pela mesma classe de polinômio canônicos de misturas, aplicou-se o método de otimização NBI (*Normal Boundary Intersection*), modificado pela adição de uma classe de restrições elípticas auxiliares. Para se verificar a razoabilidade da tratativa, confrontou-se o conjunto de resultados obtidos com aqueles proporcionados pelos melhores métodos individuais de previsão, por métodos de combinação tradicionais, pelo método FA-NBI e suas variantes obtidas segundo a aplicação das 3 regras de Jolliffe. Comparando-se, também, os resultados da aplicação de todos os métodos a um conjunto de teste não utilizado nas etapas de modelagem e de otimização (*out-of-sample*), constatou-se uma destacada eficiência do método proposto nesta tese em relação aos demais. Embora os resultados obtidos circunscrevam-se apenas à série estudada, sua adequabilidade pressupõe que quaisquer outros tipos de séries temporais ou combinações de métodos poderiam experimentar melhorias significativas similares quanto à assertividade das previsões geradas.

**Palavras-chave:** Combinação de Métodos de Previsão; Otimização Multiobjetivo; Critérios de Jolliffe; Séries Temporais; Gás Natural; Arranjos de Misturas.

## ABSTRACT

This study presents a nonlinear multi-objective optimization method for defining optimal weights for combining time series forecasting methods used to estimate annual natural gas demands. The weight allocation approach employs mixed experimental arrangements to model the relationship between various predictive performance metrics, and the weights assigned to the prediction residuals of the individual time series methods chosen for the combinations. The Double Exponential Smoothing (DES) method, the Holt-Winters additive (WA) method, and the multiplicative (WM) method were used in this study. Various performance metrics related to location, dispersion, and diversity were modeled using canonical polynomials for mixtures, which were then individually optimized to form a Payoff matrix for the individual solutions. These were then grouped according to the minimum distance between optimal points and the Jolliffe criterion, defined by the Principal Component Analysis (PCA), and applied to each group identified for non-redundant metric first selection (Payoff-Jolliffe Criteria). Factor analysis (FA) was applied to the remaining metrics, via principal component extraction and varimax rotation, storing the rotated factor scores. After modeling these scores with the same canonical polynomial mixture class, the Normal Boundary Intersection (NBI) optimization method was used, modified by adding an auxiliary elliptic constraint class. The set of results was compared with results from the best individual forecasting methods, results from traditional combination methods, results from the FA-NBI method, and its variants according to the 3 applied Jolliffe rules, in order to verify the reasonableness of the data treatment. The results for all methods were compared with a test set not used in the modeling and optimizing stages, i.e., an out-of-sample set, which verified the remarkable efficiency of the method proposed in this paper, relative to the other methods. Although the results are limited to the studied series alone, the adequacy of the methods presupposes that all other types of time series, or combinations of methods, might result in similar significant improvements in forecast assertiveness.

**Keywords:** Combined Forecasting Methods; Multi-objective Optimization; Jolliffe criteria; Time Series; Natural gas; Mixing arrangements.

## LISTA DE FIGURAS

Figura 1-1 - Consumo de Gás Natural por Setor - 1950 a 2019 .....	13
Figura 1-2 - Consumo de Gás Natural pelo Setor Industrial .....	13
Figura 2-1 Agrupamentos das 12 combinações de modelos de previsão propostas por Karabiber e Xydis (2020) considerando a similaridade entre as métricas de performance. ....	36
Figura 2-2 - Representação gráfica do método NBI.....	53
Figura 2-3 - Método FA-NBI .....	65
Figura 3-1 - 95% Elipse de Confiança versus Fronteira de Pareto (NBI). ....	67
Figura 3-2 - Fronteira de Pareto corrigida pela elipse de confiança de 95%.....	69
Figura 3-3 - Dilema das funções objetivo correlacionadas. ....	72
Figura 3-4 - Fluxograma Método Jolliffe Fatorial.....	74
Figura 3-5 - Fluxograma Explodido “Passo 14” do Método proposto .....	75
Figura 3-6 -Restrição elítica para as proporções do arranjo simplex (ou pesos).....	78
Figura 4-1 - Série Original de demanda industrial de gás natural nos EUA (1972-2019). ....	81
Figura 4-2 - Análise estatística da série completa. ....	82
Figura 4-3 - Gráfico de Autocorrelação Parcial e Total .....	82
Figura 4-4 - Ajustes dos melhores métodos em relação à série original no conjunto in-sample. ....	83
Figura 4-5 - Arranjo de Misturas Simplex Lattice {3, 5}.....	84
Figura 4-6 - Dendograma dos agrupamentos de métricas de performance para previsão (MPP) por semelhança entre os ótimos individuais. ....	90
Figura 4-7 - Elipse de confiança de 95% para as métricas MAE e SD. ....	99
Figura 4-8 - Fronteira de Pareto para os resultados do método NBI-FA com filtros de Jolliffe e restrições elípticas.....	103
Figura 4-9 - Simplex Overlaid Plot: superfícies de resposta sobrepostas e peso ótimo.....	103
Figura 4-10 - MDE surface plot para MAE e sMdAPE. ....	104
Figura 4-11 - Resultados das previsões para o conjunto de teste (out-of-sample). ....	105

Figura 4-12 - Resultados das previsões para o conjunto de teste (out-of-sample) (Série seccionada).....	106
---	-----

## LISTA DE TABELAS

Tabela 4-1 - Valores dos parâmetros utilizados nos melhores métodos.....	83
Tabela 4-2 - Métricas de performance de previsão de 25 séries de resíduos ponderadas (Parte D).....	85
Tabela 4-3 - Métricas de performance de previsão de 25 séries de resíduos ponderadas (Parte II). ....	86
Tabela 4-4 - Métricas de performance de previsão de 25 séries de resíduos ponderadas (Parte III). ....	87
Tabela 4-5 - Coeficientes do polinômico cúbico completo para as métricas de performance (Parte I). ....	88
Tabela 4-6 - Coeficientes do polinômico cúbico completo para as métricas de performance (Parte II). ....	89
Tabela 4-7 - Pesos ótimos individuais para cada métrica utilizada para a avaliação da combinação. ....	89
Tabela 4-8 - Análise de Componentes Principais (PCA) para o cluster MSE, RMSE, U1, VAR e SD. ....	91
Tabela 4-9 - Análise Fatorial de MPP's remanescentes (Extração por componentes principais e rotação varimax). ....	92
Tabela 4-10 - Análise de componentes principais das MPP's pertencentes ao Fator 2. ....	92
Tabela 4-11 - Análise fatorial das MPP's mais significativas e não-redundantes. ....	93
Tabela 4-12 - Simplex Lattice Design, MPP's e escores rotacionados de fator. ....	94
Tabela 4-13 - Coeficientes das funções objetivo utilizadas no método NBI.....	95
Tabela 4-14 - Valores utilizados na segunda restrição elíptica acoplada ao método NBI. ....	100
Tabela 4-15 - Coeficientes das funções objetivo utilizadas no método NBI.....	100
Tabela 4-16 - Matriz Payoff para as funções FMSE que serão acopladas ao método NBI....	101
Tabela 4-17 - Resultados do método NBI-FA com filtros de Jolliffe e restrições elípticas...	102
Tabela 4-18 - Comparações entre métodos. ....	106
Tabela 4-19 - Comparação do método proposto com o método NBI-FA modificado pelos critérios de Jolliffe. ....	107

## LISTA DE QUADROS

Quadro Tese 2-1 - Autores e técnicas utilizadas como redução de dimensionalidade dos dados .....	61
---	----

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	<b>12</b>
1.1	OBJETIVO	16
1.1.1	Objetivo principal	16
1.1.2	Objetivos secundários	17
<b>2</b>	<b>REVISÃO DA LITERATURA</b>	<b>18</b>
2.1	O GÁS NATURAL E OS DESAFIOS DE SUA PREVISÃO	18
2.2	MÉTODOS DE PREVISÃO DE SÉRIES TEMPORAIS	23
2.2.1	Métodos Auto-regressivos Integrados de Médias Móveis	24
2.2.2	Alisamento Exponencial Duplo	25
2.2.3	Método de <i>Winters</i>	26
2.3	COMBINAÇÃO DE MÉTODOS DE PREVISÃO	27
2.4	MÉTRICAS DE PERFORMANCE PARA OS MÉTODOS DE PREVISÃO	36
2.4.1	Métricas de Desempenho	37
2.4.2	Métricas de Precisão	42
2.5	PROJETO DE EXPERIMENTOS DE MISTURA	43
2.6	ANÁLISE FATORIAL COM EXTRAÇÃO POR COMPONENTES (PCFA)	48
2.7	MÉTODO DE INTERSECÇÃO NORMAL À FRONTEIRA	50
2.8	MÉTODO DE JOLLIFFE	55
2.9	MÉTODO FA-NBI	63
<b>3</b>	<b>MÉTODO PAYOFF-JOLLIFFE FATORIAL</b>	<b>66</b>
3.1	MÉTODO NBI-FA COM RESTRIÇÕES ELÍPTICAS E FILTROS DE JOLLIFFE	66
3.2	OTIMIZAÇÃO COM FUNÇÕES OBJETIVO CORRELACIONADAS	71
3.3	MÉTODO JOLLIFFE FATORIAL	73
<b>4</b>	<b>APLICAÇÃO, RESULTADOS E DISCUSSÃO</b>	<b>81</b>
4.1	SÉRIE HISTÓRICA DE DEMANDA DE GÁS NATURAL E SUAS CARACTERÍSTICAS	81
4.2	APLICAÇÃO DO MÉTODO PAYOFF-JOLLIFFE FATORIAL COM RESTRIÇÕES ELÍPTICAS	88
4.3	QUALIDADE DA PREVISÃO OBTIDA	105
<b>5</b>	<b>CONCLUSÕES</b>	<b>108</b>
<b>6</b>	<b>REFERÊNCIAS BIBLIOGRÁFICAS</b>	<b>111</b>

## 1 INTRODUÇÃO

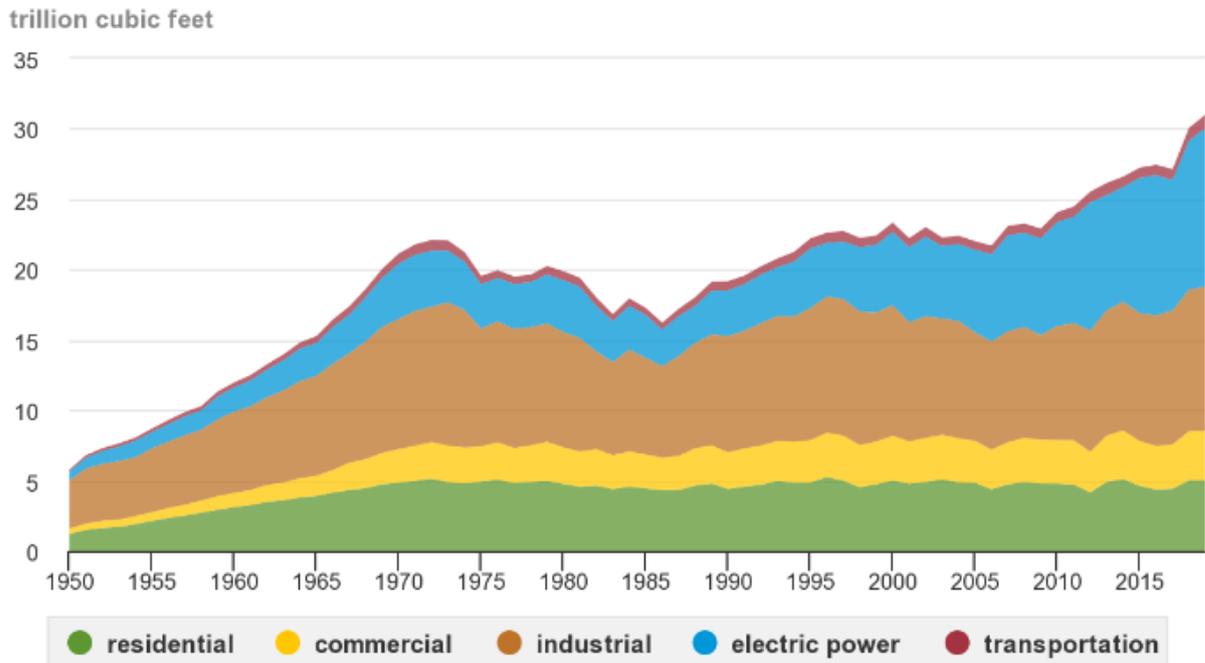
A previsibilidade dos comportamentos fenomenológicos é a condição fundamental para o correto planejamento, controle e acompanhamento dos processos decisórios. A esse respeito, Sanders (1995) destaca que a previsão é a busca pelo processo de estimação de futuros eventos com o propósito de efetivar o planejamento e de se tomar decisões. Contudo, embora seja reconhecidamente importante, a tarefa de se obter previsões adequadas e confiáveis é demasiadamente difícil dado que nem toda dinâmica comportamental de séries temporais é trivial. Os exemplos estendem-se de previsões de ativos econômicos, eventos naturais, comportamentos de processos até a demandas por recursos naturais ou fontes de energia renováveis ou não.

Como objeto principal da previsão, optou-se pelo estudo da demanda por fontes de energia, em específico, o Gás Natural, importante fonte de energia para o setor industrial, principalmente em países como os Estados Unidos da América (EUA). De acordo com a EIA (*U.S. Energy Information Administration*), o gás natural é uma fonte de energia fóssil que se forma nas profundezas da superfície da terra e que, apesar de ser um combustível fóssil, é considerado como fonte de energia limpa e alternativa ao uso de outros tipos de energia que agridem mais o meio ambiente e, conseqüentemente, acarretam mudanças climáticas. Em função dessa característica, a demanda por gás natural é crescente, e sua importância na formação das matrizes energéticas dos países é considerável.

Huntington (2007) cita que uma mudança na composição industrial está transformando o consumo de gás industrial nos EUA. Obviamente, como em toda atividade econômica, a previsão de consumo é uma tarefa primordial. Baldacci *et al.* (2016), por exemplo, citam que a previsão de consumo de gás natural é crítica para muitas tarefas de empresas fornecedoras de gás. Wang *et al.* (2016) pontuam que, na China, o consumo de gás natural como uma fonte de energia limpa e renovável está aumentando para ajudar nas necessidades de energia e nas mudanças climáticas. A introdução do gás natural nas matrizes energéticas, nesse sentido, é um fator preponderante para se diminuir o volume de emissão de  $CO_2$  na atmosfera. O Protocolo de Kyoto, com isso, prevê como meta a redução de emissão de  $CO_2$ , principalmente, para os países que mais emitem.

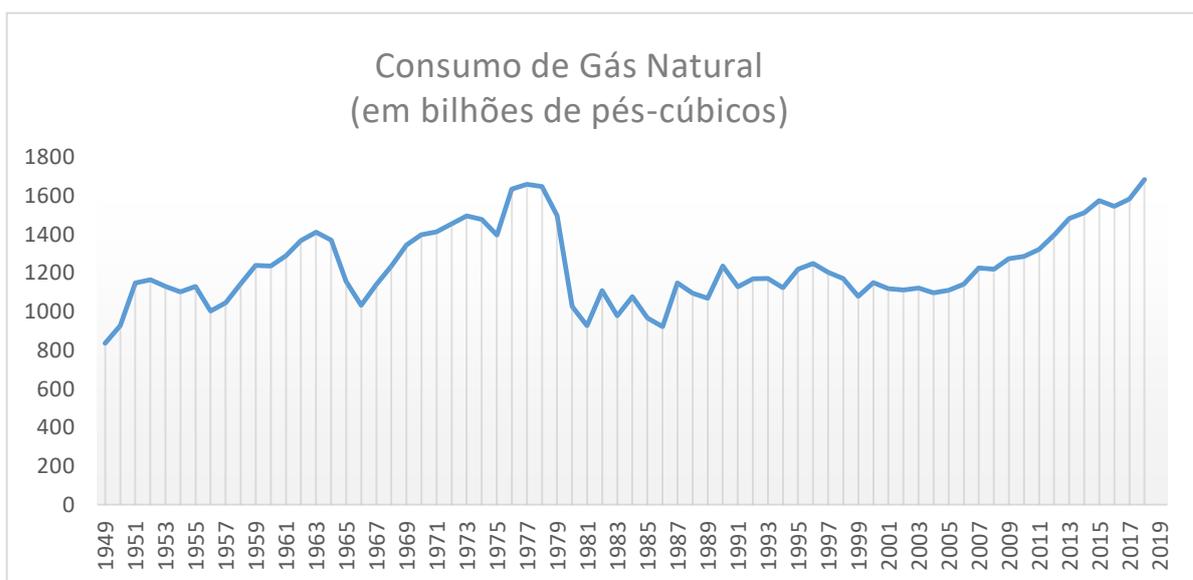
No trabalho de Shahbaz *et al.* (2013), é enfatizada a importância do gás natural como fonte de geração de eletricidade e como uma alternativa de energia limpa e, ao mesmo tempo, atrativa. A EIA fez um importante levantamento sobre o consumo do gás natural (Figura 1.1)

demonstrando que, em todos os setores da economia americana (residencial, industrial, transporte, comercial e energia elétrica), houve um aumento significativo no seu consumo.



**Figura 1-1 - Consumo de Gás Natural por Setor - 1950 a 2019**  
Fonte: EIA

No setor industrial, foco deste trabalho, o aumento foi considerável. As indústrias perceberam uma alternativa segura e eficiente para redução de  $CO_2$  utilizando o gás natural. (HUNTINGTON, 2007; SHAHBAZ *et al.* 2013).



**Figura 1-2 - Consumo de Gás Natural pelo Setor Industrial**  
Fonte: EIA

O consumo de gás natural sofreu redução por volta da década de 80 e depois recuperou seu crescimento ao longo das últimas décadas, justamente quando as exigências em relação à proteção do clima se tornaram mais efetivas. Pela **Figura 1-2**, é possível perceber o aumento do uso de gás natural no setor industrial nos anos entre 2000 e 2019, período em que, em valor absoluto, atingiu 46%. Esse percentual demonstra a importância do gás natural para a composição da matriz energética de um país e, conseqüentemente, para o mundo. Muitos outros países compartilham e experimentam a mesma tendência, tais como China, EUA, Paquistão, Itália, Alemanha, Turquia e Dinamarca. (BIANCO *et al.* 2014; WANG e LIN, 2014; CHENG *et al.* 2018; HUNTINGTON, 2007; BATZIAS *et al.*, 2006; BALDACCI *et al.* 2016; ZENG e LI, 2016; PANAPAKIDIS e DAGOUMAS, 2017; SHAHBAZ *et al.* 2013; WANG *et al.*, 2016, KARABIBER e XYDIS, 2020).

Séries temporais, como a apresentada na **Figura 1-2**, podem ser modeladas e seus comportamentos, previstos com relativo nível de precisão. Para matrizes energéticas cada vez mais dependentes dessas fontes de energia, a previsão de demanda é uma atividade crucial, porém não trivial. Isso porque as dinâmicas das séries são não lineares, às vezes heteroscedásticas, não estacionárias ou pontuadas por sazonalidades. Ademais, cada série é única, nem sempre os métodos de previsão disponíveis são capazes de exibir boas performances de previsão em todas essas condições. Há quase tantos métodos de previsão disponíveis quanto o número de métricas de performance utilizadas. Tal fato contribui para que a importante tarefa de se prever a demanda de uma fonte de energia, por exemplo, torne-se um trabalho árduo e, por vezes, confuso. Que método usar? Em que métrica se deve embasar a análise de performance da previsão? Pensando nessa complexidade, esta tese buscou evidenciar que combinações de métodos apresentam performances melhores do que tratativas individuais e que, nem sempre, a métrica de performance escolhida para se avaliar a qualidade da previsão é a melhor opção. Então, no primeiro caso, como há incerteza quanto ao melhor método, combinam-se os melhores; mas, como a definição de melhor método é baseada na métrica de performance, volta-se à dúvida inicial. Dessa forma, um procedimento mais razoável pressupõe a avaliação de todas as métricas disponíveis com a seleção daquelas não redundantes que mais contribuam para uma performance adequada de previsão. Essa tarefa de eliminação de métricas redundantes pode ser realizada empregando-se métodos multivariados de redução da dimensionalidade dos dados, tal como a análise de componentes principais (PCA). A utilização da redução da dimensionalidade foi feita com as métricas de desempenho e risco, buscando trabalhar com um número reduzido de métricas que preservassem as informações geradas pelo conjunto original de métricas e pudessem melhorar os resultados obtidos. (CUMMING e

WOOF, 2007; LIU *et al.* 2020; HIDALGO *et al.* 2020; KARIMI E FARROKHANIA, 2014; LI e SHU, 2009). No presente trabalho, foram utilizados dois procedimentos para estudar o resultado da redução da dimensionalidade dos dados na combinação de métodos de previsão.

No primeiro procedimento adotado e no escopo principal de estudo da tese, foi desenvolvido o Método NBI com restrições elípticas para escores fatoriais rotacionados de métricas de performance, filtradas pelo critério *Payoff*-Jolliffe. Nessa abordagem, utiliza-se arranjo de experimentos de mistura para modelar a relação entre as métricas de desempenho de previsão e os pesos atribuídos aos resíduos de previsão dos métodos de séries temporais individuais escolhidos para a combinação. Depois de modelar individualmente cada uma das Métricas de Performance para a Previsão (MPP), serão feitas análises (Cluster e Jolliffe) para eliminar as métricas com similaridades entre as soluções ótimas. De posse dos resultados das métricas escolhidas, será aplicada uma análise fatorial com a extração de dois fatores rotacionados, eliminando-se as métricas com menor carregamento dentro deles. Em seguida, aplica-se a 2ª Regra de Jolliffe para eliminar as variáveis redundantes dentro de cada fator, obtendo-se o conjunto final de métricas que será utilizado na análise fatorial. Após o processamento, os escores de cada um dos dois fatores rotacionados de interesse são armazenados, com posterior modelagem segundo o polinômio canônico de interesse. A aplicação do NBI com conjunto de restrições elípticas formará uma fronteira de Pareto com diversas soluções ótimas viáveis. Com base na menor distância de Mahalanobis entre a solução Pareto-ótima e o vetor de ótimos individuais, assim como na máxima diversidade de pesos obtidos pelo índice de entropia de Shannon para o vetor Pareto-ótimo de interesse, é feita a escolha dos pesos ótimos para serem aplicados às séries de resíduos derivados dos métodos de previsão de Holt-Winters aditivo (WA), multiplicativo (WM) e alisamento exponencial duplo (DES).

No segundo procedimento, para buscar a redução da dimensionalidade dos dados, foram utilizados os passos descritos no trabalho de Bacci *et al.* (2019), com a aplicação do FA-NBI na combinação dos métodos de previsão até o momento da realização do PCFA. Nesse ponto, com a intenção de verificar o resultado da redução da dimensionalidade dos dados na utilização das métricas de risco e performance, foi estudada a aplicação dos 3 métodos de Jolliffe para a eliminação de métricas redundantes. Dessa forma, utilizou-se um método de combinação de previsão com uso da redução das métricas de desempenho e performance, mediante a aplicação dos 3 métodos de Jolliffe. O estudo foi feito para que as séries de *Factor Score* utilizadas obtivessem informações somente das métricas selecionadas.

Essa proposta de redução de dimensionalidade dos dados é defendida por Jolliffe em diversos trabalhos publicados e também por diversos autores que utilizaram outros métodos (*Principal Component Analysis (PCA)*”, “*Artificial Neural Networks*”; “*Kernel PCA*”; “*Fuzzy Robust Principal Component Analysis (FRPCA)*”; “*Multi-Criteria Decision Analysis e Machine Learning*” e outros apresentados na revisão bibliográfica), a fim de também obter a redução das variáveis necessárias para realização dos cálculos (JOLLIFFE, 1972; JOLLIFFE, 1973; JOLLIFFE, 2002; ZHONG e ENKE, 2017; LI e SHU, 2009; MARTINS *et al.* 2020; HAN e GE, 2020; FANG e YU, 2020; MEI *et al.*, 2007; KARIMI e FARROKHANIA, 2014; DAI *et al.* 2016; LIU *et al.*, 2020; CUMMING E WOOF, 2007).

Após a aplicação da redução de quantidade das métricas pela aplicação dos 3 métodos de Jolliffe, foram feitos os *Factor Scores* com cada grupo de métrica identificada. Foram realizados os procedimentos referentes à determinação das funções objetivo de cada série de *Factor Score* por meio de um arranjo de misturas (CORNELL, 2002, MYERS *et al.*, 2009; COETZER e HAINES, 2017; MASON *et al.*, 2003). Por fim, foram realizados os cálculos de otimização relativos ao método *Normal Boundary Intersection* (NBI), considerando os objetivos de maximização ou de minimização, dependendo da correlação existente entre as métricas de risco e desempenho, bem como seus respectivos *Factor Score* representativos. Com os pesos derivados de um arranjo de mistura atribuídos a cada método, foram realizados os cálculos de comparação entre o FA-NBI, os 3 métodos de Jolliffe e os de previsão individuais.

## 1.1 OBJETIVO

Diante dos argumentos expostos anteriormente, das lacunas observadas quanto à indefinição de métodos de previsão precisos, da incerteza gerada pela quantidade de métricas de performance disponíveis e das possibilidades de obtenção de soluções razoáveis a partir de uma combinação sistemática de métodos de previsão, de métodos e de planejamento, além de análise de experimentos, de métodos de análise estatística multivariada e de métodos de otimização, propõe-se o seguinte objetivo principal para este trabalho.

### 1.1.1 Objetivo principal

- Desenvolver um procedimento para a combinação de métodos de previsão que seja alicerçado na influência que os pesos atribuídos a cada componente da combinação exercem sobre a minimização dos valores de múltiplas métricas de performance de previsão não-redundantes, com aplicação na previsão de demanda de Gás Natural.

### 1.1.2 Objetivos secundários

- Avaliar se a aplicação dos critérios de redução de dimensionalidade de Jolliffe à matriz de ótimos individuais *Payoff* favorece a obtenção de um conjunto de pesos para as combinações que conduzam aos menores erros de previsão;
- Examinar se a minimização de métricas de performance de previsão correlacionadas conduzem aos mesmos pesos ótimos;
- Estimar qual dos critérios de Jolliffe é o mais adequado para o caso estudado;
- Analisar se o conjunto de restrições elípticas favorece a obtenção de pesos ótimos para o problema, eliminando ótimos individuais extremos (*outliers*) e garantindo que as combinações sejam diversificadas, distanciando-se seguramente dos vértices da região simplex de interesse;
- Comparar os resultados obtidos com os métodos individuais, com combinações obtidas pelo método FA-NBI e com combinações obtidas com FA-NBI modificado pelos critérios de Jolliffe.

## 2 REVISÃO DA LITERATURA

### 2.1 O GÁS NATURAL E OS DESAFIOS DE SUA PREVISÃO

Energia é o fator mais importante para o desenvolvimento sustentável das nações, seja em seu aspecto social, econômico ou ambiental. Dada sua importância em um mundo repleto de transformações, sua demanda cresce ano após ano. Entre as fontes de energia mais importantes disponíveis na atualidade estão as de origem fóssil, na forma de líquidos (como o petróleo), o gás natural e o carvão, que constituem cerca de 80% do consumo global (Hafezi *et al.*, 2021). Entretanto, com a virada do século XXI, a atenção dos países se voltou para as questões ambientais, envolvendo principalmente as causas e os efeitos do aquecimento global.

Dessa forma, esses dois objetivos conflitantes de crescimento do fornecimento de energia e, por outro lado, da preocupação com as questões ambientais e a sustentabilidade, confluíram para o desenvolvimento de modelos capazes de conciliar esses propósitos. Nesse contexto, o gás natural é a escolha mais viável e razoável de suprimento global de energia por duas razões: (a) é uma fonte primária de energia que é muito mais limpa do que o petróleo e o carvão; (b) na maioria dos casos, é uma fonte de energia muito mais barata do que várias outras fontes de energia renováveis (HAFEZI *et al.*, 2021). Consequentemente, o gás natural se apresenta como uma alternativa mais competitiva no curto e no médio prazo para suprir a produção ainda insuficiente das fontes renováveis, com demanda crescendo exponencialmente nos últimos 30 anos (LIU *et al.*, 2021).

O crescente consumo em larga escala não apenas reflete o sucesso das políticas de proteção ambiental e o crescimento econômico, mas também ressalta o quão ineficazes podem ser os processos de gestão do seu fornecimento sem modelos de previsões de demanda confiáveis. Tal mecanismo de auxílio na tomada de decisões conduzirá empresas e países a formular planos de abastecimento mais razoáveis com considerável melhoria da eficiência operacional, fornecendo informações importantes para os setores de produção e de infraestrutura a custos reduzidos.

As pesquisas sobre o desenvolvimento e a aplicação de métodos de previsão às séries de demanda de gás natural se iniciaram nos anos 1950 e, desde então, experimentaram diferentes períodos de desenvolvimento, nos quais os métodos disponíveis evoluíram dos algoritmos convencionais para os baseados em inteligência artificial. Liu *et al.* (2021) descrevem uma linha do tempo que se inicia na década de 1950 com os modelos BAN (*Best asymptotically Normal models*) seguidos pelos modelos convencionais de séries temporais (1970), os primeiros software de previsão (1984), o advento das redes neurais artificiais para

previsão de demanda (1994), chegando aos modelos de aprendizagem de máquina do tipo SVM (*Support Vector Machine*) no ano de 2004.

Em 2006, as pesquisas em previsão de demanda de gás natural começam a ser influenciadas pelo então “novo” conceito de *deep learning*, representado de forma preponderante pelas redes neurais do tipo LSTM (*Long Short-Term Memory*). Em um desses revezes surpreendentes da ciência, entre os anos de 2013 e 2017, os métodos convencionais voltaram a ser amplamente empregados em previsões e têm se mantido como alternativas confiáveis até o presente. Uma característica básica enfatizada por Liu *et al.* (2021) é que, ao longo do tempo, a tendência das pesquisas evoluiu dos modelos de previsão de longo prazo (*Long Term*) para uma preponderância dos modelos para as previsões de curto prazo (*Short Term*), o que se traduziu para dinâmicas seriais mais complexas que modelos tradicionais não eram capazes de detectar e de modelar.

Destaca-se que, de acordo com o estabelecido por esses pesquisadores, denomina-se como “*short-term forecasting*” as previsões feitas em bases **horárias**, **diárias** ou **semanais**; modelos para “*medium-term*” são aqueles que tratam as séries temporais em bases **mensais** ou **quadrimestrais**. Já as previsões do tipo “*Long-term*” são aquelas que consideram a evolução da série temporal em uma base temporal **anual**. Ressalta-se que a escolha do esquema temporal adequado ao problema de previsão dependerá do tipo de planejamento para o qual ela é necessária, ou seja, decisões e ações de curto prazo exigirão previsões de curtíssimo período e assim por diante.

Os primeiros modelos de previsão para gás natural desenvolvidos na década de 1950 eram eminentemente estatísticos e baseavam-se em modelos de regressão que relacionavam a demanda dessa fonte de energia a fatores estruturais, como preços, PIB, nível de produção, volume de comercialização, entre outros. Neste estágio inicial, encontram-se os modelos exponencialmente ponderados (EWMA), o método dos mínimos quadrados ordinários e as regressões logísticas.

Os anos de 1970 apresentam o estágio de evolução dos métodos de previsão denominado de “Convencional”. Esse período é marcado profundamente pelo trabalho seminal de Box e Jenkins publicado no livro *Time Series Analysis: Forecasting and Control*. Nessa fase, explodem as aplicações dos modelos auto-regressivos (AR), os modelos de médias móveis (MA), os modelos autoregressivos e de médias móveis (ARMA) e, conseqüentemente, o modelo auto-regressivo de médias móveis, integrado, popularmente conhecido por ARIMA. O impacto desses desenvolvimentos de Box e Jenkins foi tão grande que, até hoje, são realizados trabalhos com base em tais modelos. Modelos AR são usados por Andersen *et al.* (2013) e

(2014); modelos MA são encontrados no trabalho de Li *et al.* (2010); aplicações de ARMA são relatadas em Pappas *et al.* (2010). Os modelos ARIMA permanecem competitivos até os dias atuais. Karabiber e Xydis (2020), por exemplo, utilizam modelos ARIMA na previsão de demanda de gás natural de “curto prazo” na Dinamarca; apesar de o ARIMA ter uma importância relevante, neste trabalho, o modelo é usado em uma combinação com redes neurais artificiais. Já no trabalho de Akpinar e Yumusak (2016), as previsões usando ARIMA são desenvolvidas para séries anuais de demanda de gás natural, e o método é combinado com os modelos de alisamento exponencial e Holt-Winters.

O método de Holt-Winters é outro exemplo de modelos convencionais que se mantêm competitivos em meio ao turbilhão causado pelo advento da inteligência artificial. O método é encontrado, por exemplo, no trabalho de Li *et al.* (2021) como uma opção para a previsão de “produção” mensal de gás natural. Embora não seja em gás natural, Tratar e Strmčnik (2016) também ressaltam a competitividade e a boa adequação desse método ao apresentar um estudo comparativo da aplicação dos métodos de Holt-Winters aditivo, Holt-Winters multiplicativo e regressão múltipla à previsão de demanda de “curto” e “longo” prazo para fontes de calor como água quente e vapor; a pesquisa aponta para uma boa adequação dos métodos exponenciais para previsões de longo prazo desse tipo de fonte de energia.

Entre os anos de 1992 e 2006, a preponderância na previsão de demanda de gás natural é representada pelos algoritmos de “inteligência artificial”, com destaque para as redes neurais artificiais (RNA) e para a Regressão por vetores de suporte (SVR), como encontrado no trabalho de Azadeh *et al.* (2013). As redes neurais de base radial (RBF) também são encontradas como métodos de I.A. aplicados à previsão de demanda de gás natural, tal como apresentado no trabalho de Azadeh *et al.* (2011).

A previsão de demanda de gás natural em um período mais recente, especificamente entre 2006 e 2017, experimentou uma grande influência do conceito denominado de “*deep learning*”, um termo da inteligência artificial que pode ser operacionalizado pelo uso das redes neurais LSTM (*Long Short-Term Memory*), como citado por Ghasemi *et al.* (2018). Como curiosidade e atualidade, as LSTM’s têm sido usadas desde 2015 pelo *Google* em tarefas de reconhecimento de voz, tal como no *Google Voice*<sup>(TM)</sup> ou no *Facebook*<sup>(TM)</sup>, plataformas nas quais as LSTM’s são empregadas para cerca de 4,5 bilhões de tarefas de transcrição automáticas (atualizações de perfis) diárias. Diferentemente das redes neurais tradicionais cujos algoritmos se baseiam em um sequenciamento do tipo “*feedforward*”, as LSTM’s mantêm conexões de “*feedback*”. Dessa forma, LSTM’s são treinadas de forma “supervisionada” usando algoritmos de otimização, como o “gradiente descendente” (ou “*Steepest Descent*”), combinados com

*backpropagation*, para computar os gradientes necessários para a definição das mudanças de pesos da LSTM proporcionalmente à derivada parcial do erro em relação ao respectivo peso.

No caso das redes neurais tradicionais, o gradiente descendente tende a desaparecer rapidamente ao longo das interações do algoritmo, proporcionalmente ao tamanho do lag entre os eventos (Hochreiter *et al.*, 2001). Na LSTM, entretanto, quando os valores de erro são “retro-propagados” (*back-propagated*) a partir da camada de saída, o erro é mantido em uma unidade celular da LSTM; isso se denomina “*short-term memory*”. Trabalhos como o de Wei *et al.* (2019), por exemplo, mostram o emprego de LSTM’s na previsão de demanda diária de gás natural em diferentes regiões e sua respectiva eficácia em comparação com SVM, além de modelos de regressão.

A literatura evidencia, entretanto, que a grande utilidade de redes LSTM’s são as previsões de curto prazo, principalmente as previsões de base “horária”. Como se percebe na descrição da literatura, parece haver uma relação muito bem definida entre a base temporal da série de dados e o algoritmo capaz de detectar as dinâmicas seriais e de modelar seu comportamento; quanto menor o intervalo de tempo (ou a base temporal), mais complexos se tornam os algoritmos contidos nos métodos de previsão.

Outra característica observada nos trabalhos de previsão de demanda de gás natural, publicados no período 2006-2019, é a presença de estratégias que empregam alguma forma de “combinação” de métodos. Assim como proposto nesta tese, trabalhos como os de Akpinar *et al.* (2017), Cao e Wu (2016) empregaram a estratégia de combinação para superar as deficiências dos melhores métodos individuais de previsão, mesmo com aqueles que se baseiam nos algoritmos de inteligência artificial mais sofisticados: a previsão é sempre melhor com a combinação do que com o melhor método individual. Estas pesquisas enfatizam também que, além da combinação de métodos, a otimização dos parâmetros dos algoritmos utilizados, o pré-processamento dos dados de treinamento e a aplicação de algoritmos, que permitem a extração de características (*feature extraction*) das séries, melhoram consideravelmente o nível de precisão dos modelos de previsão.

A estratégia de otimização dos parâmetros dos algoritmos de redes neurais, por exemplo, pode ser bem exemplificada pelos trabalhos de Balestrassi *et al.* (2009), Pontes *et al.* (2012; 2016); nesses trabalhos, evidencia-se a melhoria obtida na performance de redes MLP e RBF aplicando-se uma análise de parâmetros de redes definidas por arranjos experimentais de DOE.

Segundo Liu *et al.* (2021), embora os métodos de inteligência artificial (destacando-se SVM, SVR, LSTM, RNA) apresentem elevada precisão e exatidão, seu foco é excessivamente

concentrado nos dados, e não nas relações que eles – na forma de variáveis aleatórias - estabelecem entre si, ou seja, os métodos são capazes de fazer boas previsões de valores futuros, mas não são capazes de explicar por que estas previsões foram obtidas nos níveis e nas formas apresentados; não há, portanto, uma justificativa intelegível para os resultados. Esse “desconforto” com a utilização de técnicas massivas centradas nos dados reacendeu o interesse científico pelos métodos estatísticos mais tradicionais, com ênfase nos modelos ARIMA (AKPINAR e YUMUSAK, 2017; KARABIBER e XYDIS, 2020), SARIMA (TASPINAR *et al.*, 2013), MLR (regressão linear múltipla) e MA (OZMEN *et. al.*, 2018), AR, Holt-Winters e alisamento exponencial (AKPINAR e YUMUSAK, 2016), dentre outros. Corroborando esse entendimento, Debnath e Mourshed (2018) enfatizam que os métodos de previsão estatísticos são caracterizados pela capacidade de acumular, de examinar, de elucidar, de apresentar e de associar os dados, atributos que os métodos de processamento “massivo” não possuem.

Apesar de os métodos de inteligência artificial parecerem dominantes no tocante à previsão de demandas por fontes de energia, esse não é bem o cenário real. Debnath e Mourshed (2018) apresentam uma extensa revisão da literatura sobre métodos de previsão utilizados para fins de planejamento de energia. Os pesquisadores ressaltam ainda a imprescindível importância dos métodos de previsão de demanda para a atividade de planejamento. A revisão sistemática de 483 artigos publicada na respeitada revista *Renewable and Sustainable Energy Reviews* (Fator de Impacto 12.11) mostra que, do total de 483 artigos, 297 trabalhos (61% da amostra) utilizaram métodos estatísticos de séries temporais convencionais, o que é uma quantidade bastante expressiva. Observa-se que 49 artigos utilizaram regressões lineares ou não-lineares como método de previsão de demanda de energia; 46 usaram modelos ARIMA; 22 artigos apresentaram aplicações de modelos ARMA; 19 usaram regressão logística; 14 tratam de modelos GARCH; 13 artigos sobre SARIMA, dentre outros. Desses 61%, exatamente 50% mostram métodos de previsão estatísticos aplicados em horizontes de longo prazo.

Esse parece ser um padrão recorrente na literatura: métodos clássicos usados para previsões de longo prazo; métodos de inteligência artificial para curtíssimo prazo. Em contrapartida, os trabalhos remanescentes representam a classe de métodos I.A. e dividem-se em 58 artigos sobre a utilização de SVM (*Support Vector Machine*); 194 trabalhos sobre a aplicação de redes neurais artificiais (RNA); 39 sobre o uso de algoritmos genéticos e 48 sobre a utilização de métodos de otimização como *bee colony*, *ant colony* ou *particle swarm optimization* (PSO). De uma forma geral, Debnath e Mourshed (2018) apresentam ainda as métricas de performance de qualidade de previsão mais recorrentes: o MAPE, MAE, RMSE e MAD. Essas métricas são preponderantes independentemente do tipo de algoritmo usado na

tarefa preditiva. Tal constatação também é observada em Liu *et al.* (2021), trabalho no qual são examinados 72 artigos exclusivamente de previsão de demanda de gás natural; do total de artigos avaliados, constata-se que o MAPE é a principal medida de performance em 44% dos trabalhos, independentemente do período de análise ou do método de previsão utilizado. Além dessa métrica de performance, destacam-se ainda o RMSE, MAE, MSE e SD, dentre outras. Comparando-se o MAPE médio de modelos estatísticos para previsões de longo prazo (1,90%) com o MAPE de modelos de redes neurais (5,3%) e LSTM (2.15%), verifica-se que, nesse horizonte de planejamento, os modelos tradicionais são bem competitivos; o mesmo não ocorre quando o horizonte é de curto prazo (MAPE's respectivamente iguais a 14.2%, 6.52% e 5.94%).

A análise sistemática da literatura no tocante à previsão de demanda de gás natural proposta por Liu *et al.* (2021) é finalizada por um resumo sobre as principais características dos problemas de previsão e sua relação com os métodos disponíveis. No que tange a esta tese de doutorado, verifica-se que os modelos estatísticos de previsão tradicionais serão adequados para a previsão de demanda de gás natural quando: (a) os conjuntos de dados disponíveis apresentarem tamanhos variando de “dezenas” a “centenas” de observações; (b) quando o período da série e suas respectivas previsões forem de longo prazo (dados históricos em bases “anuais”); c) quando o método preditor não envolver fatores (variáveis explicativas externas) em sua composição ou algoritmo e (d), quando o fator “estabilidade” das previsões for mais importante do que “precisão”.

No próximo item, será apresentada uma breve revisão dos métodos de previsão de séries temporais. Alguns desses métodos serão escolhidos (após testes de adequação para previsão) para ilustrar a proposta de alocação ótima de pesos para as combinações em problemas de previsão de demanda de gás natural, com séries históricas organizadas em base anual.

## 2.2 MÉTODOS DE PREVISÃO DE SÉRIES TEMPORAIS

Os métodos de previsão para séries temporais são utilizados para elaborar projeções embasadas em séries históricas, ou seja, valores passados são analisados para fornecer uma previsão de um cenário futuro. Montgomery *et al.* (2008) enfatizam que previsões podem ser classificadas quanto ao período a que se destinam, podendo ser de curto, médio ou longo prazo, além de os modelos matemáticos terem de se adaptar a cada situação de contorno exigida pelos diferentes prazos. Em outras palavras, é difícil que um único modelo seja capaz de capturar dinâmicas e volatilidade de séries em intervalos de amostragem muito diferentes, variando entre milissegundos até anos. Dessa forma, cada tipo de método de previsão está intrinsecamente ligado ao período para o qual se planeja o processo de tomada de decisões.

Makridakis *et al.* (1998) citam que as previsões de médio e longo prazo estão relacionadas à extrapolação dos modelos de decomposição, que estão entre as abordagens mais antigas para análise de séries temporais. Chambers *et al.* (1971) destacam que a extrapolação é mais correta para horizontes de curto prazo, sendo recomendada para períodos de longo prazo se os padrões dos dados forem estáveis. É importante entender, nesse sentido, que métodos de previsão são instrumentos necessários ao processo de tomada de decisão e, portanto, precisam ser precisos e eficientes na tarefa de estimar cenários futuros para o problema de interesse.

Como será apresentado nas próximas seções, a escolha dos métodos de previsão candidatos a participarem de uma combinação é a primeira etapa da abordagem proposta e, para esta etapa, vários métodos “candidatos” podem ser considerados inicialmente, restando apenas aqueles que apresentarem melhor adequação às séries estudadas. Entre os métodos aplicados neste trabalho, destacam-se: **ARIMA**, o modelo de alisamento exponencial duplo (**DES**, do inglês *Double Exponential Smoothing*), o modelo Holt-Winters multiplicativo (**WM**), Holt-Winters aditivo (**WMA**), os modelos de análise de tendência linear, quadrática, exponencial e S-curve. Nos itens seguintes, desenvolve-se uma explicação sobre cada um deles.

### 2.2.1 Métodos Auto-regressivos Integrados de Médias Móveis

O método **ARIMA** (*Autoregressive Integrated Moving Average*) é uma generalização do modelo auto-regressivo de média móvel (ARMA) utilizado para modelar o comportamento de uma série temporal estacionária e para gerar previsões para ela. Zhang (2003) cita que ARIMA é um dos modelos lineares mais populares de previsão de séries temporais das últimas 3 décadas embora tal modelo não seja capaz de capturar e de modelar padrões não lineares das séries de interesse. Büyüksahin e Ertekin (2019) corroboram esse pensamento, citando que os modelos ARIMA fornecem boa precisão na previsão de dados para séries temporais relativamente estacionárias. Em relação aos dados não estacionários, Makridakis *et al.* (1998) ressaltam a importância de se realizar diferenciações para a série original até que se obtenha uma série derivada de comportamento estacionário. As séries estacionárias possuem média, variância e estrutura de autocorrelação estacionárias no decorrer do tempo. Isso significa que essas propriedades se mantêm estáveis ao longo do tempo. De acordo com Makridakis *et al.* (1998), um modelo ARIMA genérico pode ser escrito como:

$$Y_t = \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \dots + \varphi_p Y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} \quad (2.1)$$

Onde:

- $Y_t$  é o valor previsto no período  $t$ ;
- $\varepsilon_t$  é o erro aleatório no período  $t$ , isto é, uma variável aleatória normalmente distribuída, com média zero, variância constante e covariância igual a zero;
- $\varphi_i$  ( $i = 1, 2, 3, \dots, p$ ) denota o  $i$ -ésimo coeficiente dos termos auto-regressivos;
- $\theta_j$  ( $j = 1, 2, 3, \dots, q$ ) denota o  $j$ -ésimo coeficiente dos termos de médias móveis;
- $p$  e  $q$  são números inteiros das ordens do modelo.

De acordo com Liu *et al.* (2021), um modelo ARIMA ( $p, d, q$ ) converte uma série temporal não estacionária em estacionária utilizando uma operação diferencial de ordem  $d$ , além de combinar um modelo auto-regressivo AR de ordem  $p$  com um modelo de médias móveis MA de ordem  $q$ . Utilizando um operador  $B$  de defasagem, estabelece-se um modelo polinomial para os coeficientes de correlação auto-regressivos  $\phi(B) = 1 - \varphi_1 B - \varphi_2 B^2 - \dots - \varphi_p B^p$ ; considerando que  $\nabla^d = (1 - B)^d$  é um operador diferencial de ordem  $d$ ; o modelo polinomial para coeficientes dos termos de médias móveis é dado por  $\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$ ;  $\varepsilon_t, \varepsilon_s$  são os resíduos da série nos instantes  $t$  e  $s$ , respectivamente;  $Var(\varepsilon_t)$  e  $\sigma_{\varepsilon_t}^2$  a variância dos resíduos e o termo de variância do erro residual;  $E(\varepsilon_t \varepsilon_s)$  e  $E(x; \varepsilon_t)$  são ambos iguais a zero e representam, respectivamente, o valor esperado dos resíduos nos instantes  $t$  e  $s$  e o valor esperado da série temporal e seus resíduos.

Embora seja um modelo tradicional de previsão de séries temporais, sua utilização, em modernas aplicações para previsão de demanda de gás natural, é facilmente encontrada (AKPINAR e YUMUSAK, 2017; KARABIBER e XYDIS, 2020; LIU *et al.*, 2021).

## 2.2.2 Alisamento Exponencial Duplo

O modelo de alisamento exponencial duplo (**DES**, do inglês *Double Exponential Smoothing*) é um método de suavização geral que pode ser utilizado com eficácia quando existe uma tendência identificada na série temporal. As estimativas dinâmicas desse modelo são calculadas tanto para nível da média, em um dado intervalo, quanto para a tendência da série. Segundo Chung e Kim (2013), o princípio básico do método **DES** é que, conforme as medidas se tornam mais afastadas de um dado presente, pesos decrescentes exponencialmente são atribuídos a elas; portanto, as medições mais recentes recebem pesos relativamente maiores do que as mais antigas. As equações 2.2, 2.3 e 2.4 apresentam as particularidades desse modelo:

$$L_t = \alpha Y_t + (1 - \alpha)(L_{t-1} + b_{t-1}) \quad (2.2)$$

$$b_t = \beta(L_t - L_{t-1}) + (1 - \beta)b_{t-1} \quad (2.3)$$

$$F_{t+m} = L_t + b_t m \quad (2.4)$$

Na equação (2.2), o parâmetro  $Y_t$  representa a medição original da variável de interesse no instante  $t$ ;  $L_t$ , representa uma medição suavizada da série e  $b_t$  computa o componente de tendência identificado (significando uma mudança nos dados suavizados de um instante para um outro). O parâmetro  $\alpha \in [0, 1]$  é uma constante de suavização, e o parâmetro  $\beta \in [0, 1]$  é uma constante usada para controlar a tendência. O papel da 1ª equação é ajustar  $L_t$  adicionando a tendência do intervalo de tempo anterior;  $b_{t-1}$  representa a suavização do valor suavizado  $L_{t-1}$ . A equação (2.3) atualiza a tendência enquanto a equação (2.4) é utilizada para a realização da previsão da série para  $m$  períodos a frente. Nessa equação, a tendência ( $b_t$ ) é multiplicada por ( $m$ ) e, depois, somada ao valor base  $L_t$ .

Apesar de ser uma técnica clássica desenvolvida na década de 1950, o DES continua sendo uma interessante opção como técnica de previsão em muitos trabalhos atuais. Nguyen *et al.* (2021), por exemplo, utilizaram-no para prever a produção de minerais - como bauxita, ferro, fosfatos, carvão, chumbo e zinco - na indústria mineradora vietnamita. Islam e Huda (2019), por sua vez, empregaram o DES para realizar a previsão de vendas de equipamentos eletroeletrônicos na Austrália, no horizonte 2000-2047, bem como o volume de “lixo eletrônico” (*e-waste*) gerado como consequência dos produtos disponibilizados no mercado e o potencial de recuperação de receita com o reaproveitamento desses materiais. Doorga *et al.* (2019) o aplicaram para previsão da distribuição de irradiação solar. Al-Ghandoor (2013), em um artigo na respeitada “*Renewable and Sustainable Energy Reviews*” utilizou o DES para prever o número de veículos, os preços dos combustíveis, as receitas e as despesas do setor de transportes na Jordânia.

### 2.2.3 Método de Winters

O método de *Winters* utiliza a suavização dos dados mediante a suavização exponencial de *Holt-Winters*, principalmente, para previsões geradas no curto e no médio prazo. O método calcula estimativas dinâmicas de 3 componentes: nível, tendência e sazonalidade. Esse método pode ser utilizado de forma multiplicativa ou aditiva em relação aos componentes tendência e sazonalidade, utilizando, para isso, as equações (2.5), (2.6), (2.7) e (2.8). Dependendo da

variação sazonal dos dados, existem diferentes tipos de sazonalidade: multiplicativa e aditiva (MAKRIDAKIS *et al.*, 1998; DANTAS *et al.*, 2017 e LIU *et al.*, 2020).

$$L_t = \alpha(Y_t - S_{t-s}) + (1 - \alpha)(L_{t-1} + b_{t-1}) \quad (2.5)$$

$$b_t = \beta(L_t - L_{t-1}) + (1 - \beta)b_{t-1} \quad (2.6)$$

$$S_t = \gamma\left(\frac{Y_t}{L_t}\right) + (1 - \gamma)S_{t-s} \quad (2.7)$$

$$F_{t+k} = L_t + Kb_t + S_{t+k-s} \quad (2.8)$$

Onde:  $L_t$  é o nível no tempo  $t$ ;  $Y_t$  é o valor dos dados no tempo  $t$ ;  $b_t$  é a tendência no tempo  $t$ ;  $S_t$  é o componente sazonal no tempo  $t$ ;  $F_{t+k}$  é o valor da previsão no tempo  $t + k$ ;  $s$  é a duração do ciclo sazonal;  $\alpha$  é o fator de suavização ( $0 \leq \alpha \leq 1$ );  $\beta$  é o fator de suavização de tendência ( $0 \leq \beta \leq 1$ ), e  $\gamma$  é o fator de suavização da mudança sazonal ( $0 \leq \gamma \leq 1$ ) (LIU *et al.*, 2020).

Aplicações das técnicas de Holt-Winters em previsão de demanda de gás natural podem ser verificadas nos trabalhos recentes de Akpinar e Yumusak (2016) e Li *et al.* (2021).

### 2.3 COMBINAÇÃO DE MÉTODOS DE PREVISÃO

A combinação de métodos de previsão é uma alternativa altamente eficiente para melhorar a precisão das previsões e para diminuir a variabilidade. Desde o início da década de 1970, muitos trabalhos têm empregado, satisfatoriamente, esta estratégia e, mesmo sendo um assunto bem conhecido na literatura há cinco décadas, ainda hoje desperta o interesse de muitos pesquisadores, a exemplo doo trabalho de Meira *et al.* (2021), publicado no reconhecido *International Journal of Forecasting*, além do trabalho de Song e Fu (2020), os quais apresentam um estudo atual sobre as diversas estratégias de ponderação dos métodos de previsão nas combinações. Na literatura, os trabalhos de Bates e Granger (1969) e Newbold e Granger (1974) destacam-se como os primeiros a combinarem diferentes métodos de previsão empregando-se médias ponderadas. Bates e Granger (1969) propuseram, de início, combinações lineares simples, com pesos percentuais individuais variando no intervalo fechado  $0 \leq w_i \leq 1$  (nota-se que esta é a mesma condição de contorno observada na construção dos arranjos experimentais do tipo “Misturas”). Além dessa recomendação, os autores defendem que uma combinação convexa deve ser utilizada quando se tratar de uma combinação de apenas dois métodos de previsão. As vantagens observadas inicialmente para estratégia de combinação

eram tão proeminentes que os autores clássicos não se continham de entusiasmo. Newbold e Granger (1974), por exemplo, citam, textualmente, que:

*“Parece-nos ... que as previsões dos modelos Box-Jenkins podem ser frequentemente melhoradas quando combinadas como modelos como Holt-Winters ou com previsões de modelos autoregressivos stepwise, e sentimos que nossos resultados indicam que em qualquer situação particular de previsão, a combinação é uma tentativa que vale a pena, embora requeira um pequeno esforço. Melhorias adicionais podem ser frequentemente obtidas considerando-se a combinação destes três tipos de métodos de previsão.”*

Os modelos de Box-Jenkins sobre os quais Newbold e Granger (1974) discorrem eram, na verdade, modelos ARIMA ( $p, d, q$ ) e eram esses modelos que deveriam ser combinados com de alisamento exponencial e os de Holt-Winters (aditivo ou multiplicativo). Makridakis e Winkler (1983) ressaltavam ainda que a precisão das previsões melhora substancialmente à medida que se aumentava o número de métodos que poderiam ser combinados embora se esperasse certo nível de saturação para mais do que quatro ou cinco métodos. Da mesma forma, os pesquisadores também observaram que não só a precisão da previsão melhorava com a combinação, mas também sua variabilidade diminuía consideravelmente. Quanto à forma como se faziam as combinações, Winkler e Makridakis (1983) ressaltam que a média ponderada (supondo pesos diferenciados, obviamente) seria uma técnica muito mais adequada do que a média simples (que corresponderia a uma média ponderada com pesos iguais).

Na literatura atual, a diferença básica entre as duas estratégias é basicamente caracterizada como o nível de “diversificação” dos pesos (Rocha *et al.*, 2020), ou seja, com pesos iguais, a combinação de métodos apresentará o maior nível possível de diversificação, garantindo que as potencialidades individuais de cada método se sobressaiam enquanto suas deficiências se atenuem. Porém, nessa estratégia, perde-se um pouco a vantagem de favorecer adequadamente o método que apresentar as melhores características para previsão. Então, uma alternativa “*Trade-off*” entre a máxima diversificação e a máxima concentração pode ser a melhor solução. Para isso, concorre a soma ponderada – também chamada de “média ponderada” – uma vez que os pesos podem ser relativamente diferentes, com baixa probabilidade de concentração de pesos em apenas um método, mas não tão diversificados quanto o cenário formado por pesos igualmente espaçados.

Os pesos atribuídos aos diferentes métodos de previsão propostos por Winkler e Makridakis (1983) relacionavam-se à matriz de variância-covariância dos erros de previsão ( $\Sigma$ ), de forma que cada peso  $w_i$  era obtido pelo elemento  $\alpha_{ij}$  da inversa da matriz de variância-covariância dos erros de previsão,  $\Sigma^{-1}$ , pela soma de todos esses elementos. Para os casos nos quais a matriz de variância-covariância dos erros de previsão ( $\Sigma$ ) era desconhecida, Newbold e

Granger (1974) propuseram cinco diferentes razões entre os quadrados dos resíduos de previsão como estimativas para os pesos  $w_i$  de cada método de previsão escolhido, das quais, duas estratégias se mostraram muito mais promissoras: (a) pesos proporcionais ao inverso da soma de quadrados dos erros e (b) pesos proporcionais ao inverso da soma de quadrados dos erros alisados exponencialmente.

Para dois modelos competidores e não enviesados de previsão, Diebold (1988) propôs o conceito de “pesos ótimos” para a combinação, utilizando as variâncias  $\sigma_1^2$  e  $\sigma_2^2$  dos erros de previsão dos modelos  $f^1$  e  $f^2$  para um passo à frente (*one-step-ahead forecast*) e sua respectiva covariância  $\sigma_{12}$ . A ideia é bastante intuitiva: suponha que se estabeleça uma combinação linear convexa  $f^c$  entre as previsões dos modelos  $f^1$  e  $f^2$  para um passo à frente, tal que:

$$f^c = \phi_{t-1} f_t^1 + (1 - \phi)_{t-1} f_t^2 \quad (2.9)$$

Então, o erro “*one-step-ahead*” dos modelos combinados será escrito por:

$$e^c = \phi e_t^1 + (1 - \phi) e_t^2 \quad (2.9)$$

Como esta expressão é uma função linear, sua variância será igual a:

$$Var e^c = \phi^2 \sigma_1^2 + (1 - \phi)^2 \sigma_2^2 + 2\phi(1 - \phi)\sigma_{12} \quad (2.9)$$

Desse modo, a minimização dessa expressão com relação ao peso  $\phi$  conduz ao peso ótimo para a combinação dos dois métodos de interesse, tal que:

$$\phi^* = \frac{(\sigma_1^2 - \sigma_{12})}{(\sigma_1^2 + \sigma_2^2 - 2\sigma_{12})} \quad (2.9)$$

As variâncias  $\sigma_1^2$  e  $\sigma_2^2$  dos erros de previsão dos modelos  $f^1$  e  $f^2$  para um passo à frente (*one-step-ahead forecast*) e a covariância  $\sigma_{12}$  são parâmetros populacionais. Em termos práticos, esses termos são substituídos por suas expressões “amostrais” equivalentes,  $\hat{\sigma}_1^2$ ,  $\hat{\sigma}_2^2$  e  $\hat{\sigma}_{12}$ , dadas por:  $\hat{\sigma}_1^2 = (1/T) \sum e_{1t}^2$ ,  $\hat{\sigma}_2^2 = (1/T) \sum e_{2t}^2$  e  $\hat{\sigma}_{12} = (1/T) \sum e_{1t} e_{2t}$ . A métrica de variância é exatamente a mesma usada para se estabelecer a métrica de performance MSE.

Outras literaturas são também constantemente citadas como referências seminais para a utilização de combinações de previsões, tais como Makridakis (1983); Winkler e Makridakis (1983); Armstrong (1986); Lawrence *et al.* (1986); Reeves e Lawrence (1982); Dickinson

(1973) e Granger e Ramanathan (1984). Desses trabalhos originais, depreendem-se algumas constatações interessantes: primeiramente, a noção universal de combinações convexas; segundo, o reconhecimento de que médias ponderadas são melhores do que médias simples; terceiro, que as médias ponderadas de métodos de previsão são mais eficientes quando seus pesos são, de alguma forma, “otimizados”. Embora uma vasta quantidade de pesquisas nesse assunto tenha sido realizada nas últimas cinco décadas, estes três pressupostos continuam ensejando novos desenvolvimentos, como os propostos por Song e Fu (2020). Da mesma forma, os trabalhos originais de combinação também suportam as suposições iniciais propostas nesta tese.

Chambers *et al.* (1971) enfatizam que a abordagem tradicional de previsão pressupõe a escolha do método mais apropriado entre aqueles existentes, porém dificilmente um método único será capaz de suplantar os resultados de combinações de métodos, principalmente quando alternativas sofisticadas são incorporadas para o melhor modelamento das dinâmicas não lineares das séries.

Reeves e Lawrence (1991) foram os pioneiros na utilização de otimização multiobjetivo para formar uma combinação ótima de métodos de previsão. Até essa ocasião, a maioria dos trabalhos empregava otimizações de uma única função objetivo relativa à qualidade da previsão obtida com combinação, função esta geralmente relacionada à precisão da previsão. Dessa forma, o método proposto pelos autores, baseado na estratégia de *goal programming*, permitia a incorporação de múltiplos objetivos na estratégia de tomada de decisão, permitindo um “*trade-off*” entre vários esquemas de ponderação para combinações em relação aos objetivos conflitantes.

Nessa linha, um primeiro grupo de funções objetivo contemplaria as medidas de precisão – descritas em termos de minimização de desvios entre o valor previsto e o valor real – enquanto um segundo grupo representaria a direção da mudança, a qual poderia ser expressa por uma dinâmica de diminuição ou de aumento do valor médio da série temporal. Outras formas de medidas conflitantes de desempenho podem ser exemplificadas pelo critério “precisão-entropia”. Enquanto a precisão é atingida pela minimização das funções características de performance de previsão, a entropia atua como um catalisador de “diversificação” de pesos, os quais, por sua vez, reduzem a variabilidade da previsão. Desse modo, a otimização simultânea dessas funções conduz a valores previstos muito próximos dos valores reais, com intervalos de confiança para previsão os mais estreitos possíveis.

Sanders e Ritzman (1995) apresentam um estudo sobre os benefícios do julgamento de especialistas para a melhoria da precisão das previsões obtidas pela combinação de métodos.

Neste trabalho, a inovação promovida pelo julgamento foi, na verdade, a incorporação de pesos nas proporções de 25-75%, 50-50% e 75-25% para a combinação de dois métodos de previsão baseados em alisamento exponencial (modelo 1) em Holt-Winters (modelo 2). Se os resultados individuais forem incorporados a essas três combinações, obtém-se uma configuração de pesos que é idêntica a um arranjo de misturas do tipo “Simplex Centróide”. Então, de certa forma, o que Sanders e Ritzman (1995) tencionavam inferir era que uma diversificação maior das combinações de pesos no espaço de solução “*simplex*” poderia favorecer a descoberta de condições melhores para as combinações.

Chan *et al.* (1999) apresentam uma abordagem na qual os “pesos ótimos” são obtidos por minimização de uma função erros quadráticos de previsão para os modelos combinados, utilizando programação quadrática com restrições lineares (neste caso, a restrição enfatiza que o somatório dos pesos atribuídos aos modelos individuais deve ser igual a um). Muito similar à abordagem que será proposta nesta tese, aplicou-se o método mínimos quadrados ordinários (OLS) à relação entre o vetor de erros quadráticos e uma matriz de pesos.

O início do século XXI é marcado pelo aumento na utilização de combinações entre modelos tradicionais e as redes neurais artificiais. Zhang (2003) apresenta uma proposta de combinação entre modelos ARIMA e RNA. Nessa estratégia, o modelo ARIMA é utilizado para modelar a parte linear do problema enquanto a rede neural é desenvolvida para capturar a estrutura não linear dos dados que podem ser detectados nos resíduos de previsão do modelo ARIMA.

Hassan *et al.* (2007) propuseram e implementaram um modelo híbrido obtido a partir da combinação de cadeias ocultas de Markov (HMM, do inglês *Hidden Markov Model*), de redes neurais (ANN) e de algoritmos genéticos (GA) para previsões de comportamento do mercado financeiro de ações. Usando as ANN's, os preços diários das ações são transformados em conjuntos independentes de variáveis que alimentam uma HMM que, por sua vez, tem seus parâmetros otimizados por um algoritmo genético. A HMM foi treinada para identificar e para localizar padrões similares na série histórica. Usando os valores previstos pela HMM, as diferenças entre o preço das ações em determinado dia de interesse e o preço no dia subsequente são calculadas. Usando uma média ponderada das diferenças dos preços para padrões similares na série, gera-se um valor de previsão para “*one-day-ahead*”.

Jeong e Kim (2009) propuseram um estudo sobre a adequação das médias simples e ponderada na combinação de métodos individuais de previsão. Diferentemente de outros trabalhos, mas, consonante a ideia inicial de Granger e Ramanathan (1984) e de Granger e Newbold (1977), os pesos são estimados por uma regressão entre  $y_t$  e as previsões não

enviesadas  $F_{(t, i)}$  e, nos casos para os quais as propriedades dos erros de previsão variarem com tempo, os pesos devem ser atualizados de alguma forma. Seguindo o modelo de Bates e Granger (1969), que empregam as variâncias e covariâncias como elementos para o cálculo dos pesos, os modelos de atualização recalculam as variâncias para cada instante  $t+i$  em função dos  $v$  erros de previsão anteriores. Assim, os pesos são atualizados como em uma média móvel.

Poncela *et al.* (2011) apresentaram uma estratégia de combinações de métodos de previsão baseadas em técnicas de redução de dimensionalidade, tais como: PCA, FA e PLS, entre outros. Pela definição teórica de PCA, um escore de componente é uma combinação linear de variáveis  $y_{t+1}$  para a qual os pesos são os autovetores da matriz de variância-covariância, cujo módulo (ou norma) é unitário. Assim, aplicando-se o PCA aos resíduos de previsão de vários métodos, obtêm-se combinações lineares independentes e em diferentes graus de importância ou de explicação, de acordo com a decomposição espectral da matriz de variância-covariância. Nesse mesmo sentido, a pesquisa de Poncela *et al.* (2011) enseja que a utilização da análise fatorial possa ser útil de forma similar: grupos de resíduos de previsão semelhantes são agrupados em escores fatoriais rotacionados diferentes; cada grupo (ou fator), então, pode integrar a combinação usando como pesos os autovalores associados a cada fator. Essa não é a aplicação direta feita pelos autores, mas é uma estratégia que pode ser facilmente derivada da ideia original. Pode-se registrar esta ideia como uma perspectiva para trabalhos futuros. Poncela *et al.* (2011) afirmam ainda que a utilização de escores de componentes principais não só considera a informação obtida com as previsões feitas com múltiplos métodos como só leva para a combinação a informação não redundante que estes dados contêm.

Outra vantagem dos métodos de combinação é o aumento da estabilidade das previsões. Segundo Bordignon *et al.* (2013), quando um único modelo for utilizado e sofrer interferências com oscilações ou outras dinâmicas próprias da série (tendência, sazonalidade, correlação etc.), o modelo de previsão será instável; porém, ao acoplar outros tipos menos sensíveis ou mais adaptáveis a essas dinâmicas entrópicas, as previsões serão mais estáveis.

Publicado no *International Journal of Forecasting*, o trabalho de Conflitti *et al.* (2015) critica a utilização de pesos iguais para as combinações e propõe que o cálculo dos pesos ótimos seja realizado a partir da minimização do erro quadrático médio de previsão (MSE) no caso de previsões pontuais e na utilização da maximização do escore log preditivo derivado do critério de Kullback-Liebler, no caso de previsões para densidades de previsão, ou seja, quando se deseja prever a distribuição de probabilidade de um conjunto de valores futuros previstos.

O conceito de “Fronteira de Pareto” é apresentado por Xiao *et al.* (2017) para a determinação de pesos ótimos para combinações de cinco modelos de redes neurais artificiais

dos tipos BPNN (*Back Propagation Neural Network*), GABPNN (BPNN otimizada por algoritmo genético), WNN (*Wavelet Neural Network*), RBFNN (*Radial Basis Function Neural Network*), GRNN (*Generalized Regression Neural Network*) acoplados a um modelo ARIMA. O objeto da previsão é a demanda de energia elétrica (a cada meia hora) na cidade de Victoria, na Austrália. Um algoritmo de otimização multiobjetivo relaciona a determinação dos pesos ótimos observando que os critérios de escolha devam satisfazer a duas exigências: ( $y_1$ ) a precisão e ( $y_2$ ) a estabilidade da previsão. O primeiro critério dedica-se à redução de viés (ou *Bias*) entre um valor observado (real) e um valor previsto para o mesmo instante  $t$ ; no segundo, minimiza-se a variância do valor previsto. Múltiplas métricas de performance (MAE, MSE e MAPE) são consideradas no estudo.

A combinação de métodos de previsão continua atual mesmo depois de mais de 50 anos desde sua proposição inicial. Entretanto, a tendência observada no último quadriênio (2018-2021) é de uma preocupação mais concentrada na utilização de novos métodos de previsão baseados em inteligência artificial e *machine learning*, tais como SVR (*Support Vector Regression*), LSTM (*Long Short-Term Memory Neural Network*) e modelos ARIMA, com uma preponderante e recursiva utilização de técnicas de processamento digital de sinais capazes de decompor as séries temporais originais e eliminar potenciais ruídos. Como a maioria das séries temporais contém tendência e volatilidade, a decomposição da original em várias subséries de alta frequência e uma de baixa frequência reduzem a dificuldade de aprendizado dos padrões, melhorando a performance de previsão. Para essa tarefa, é destacada a utilização de transformadas de Wavelet (discretas e contínuas) ou Fourier, *Singular Spectral Analysis* (SSA), *Empirical Model Decomposition* (EMD) e *Variational Mode Decomposition* (VMD).

Também se nota a utilização da combinação sendo realizada a partir de diferentes métodos de decomposição. Segundo Su *et al.* (2019), a transformada de Wavelet é capaz de fornecer informação útil sobre a dinâmica das séries tanto no domínio do tempo quanto no da frequência, ambas simultaneamente; seu desempenho é muito bom, principalmente diante de séries não estacionárias. Após a etapa de pré-processamento dos sinais de entrada, a maioria dos artigos desta área dedica-se à otimização das arquiteturas dos algoritmos de aprendizagem. Isto é principalmente verdade quando se deseja otimizar os parâmetros das redes neurais LSTM, por exemplo, empregando-se algoritmos evolutivos (GA, PSO, *Ant Colony*, *Bee Colony* etc.), os quais, quase sempre, também dependem de uma parametrização ótima: estes subsequentes procedimentos são geralmente intercambiados e executados sequencialmente. A combinação se dá com os vetores resultantes da base transformada. Só no final, o problema resolvido no domínio Wavelet é transformado para a unidade original da variável prevista por meio de uma

transformada inversa. É evidente e comum a utilização simultânea de diversas métricas (correlacionadas) de qualidade de previsão, tais como as métricas MSE, MAE, RMSE, MAPE e MASE, entre outras. Raros são os trabalhos que empregam uma única métrica.

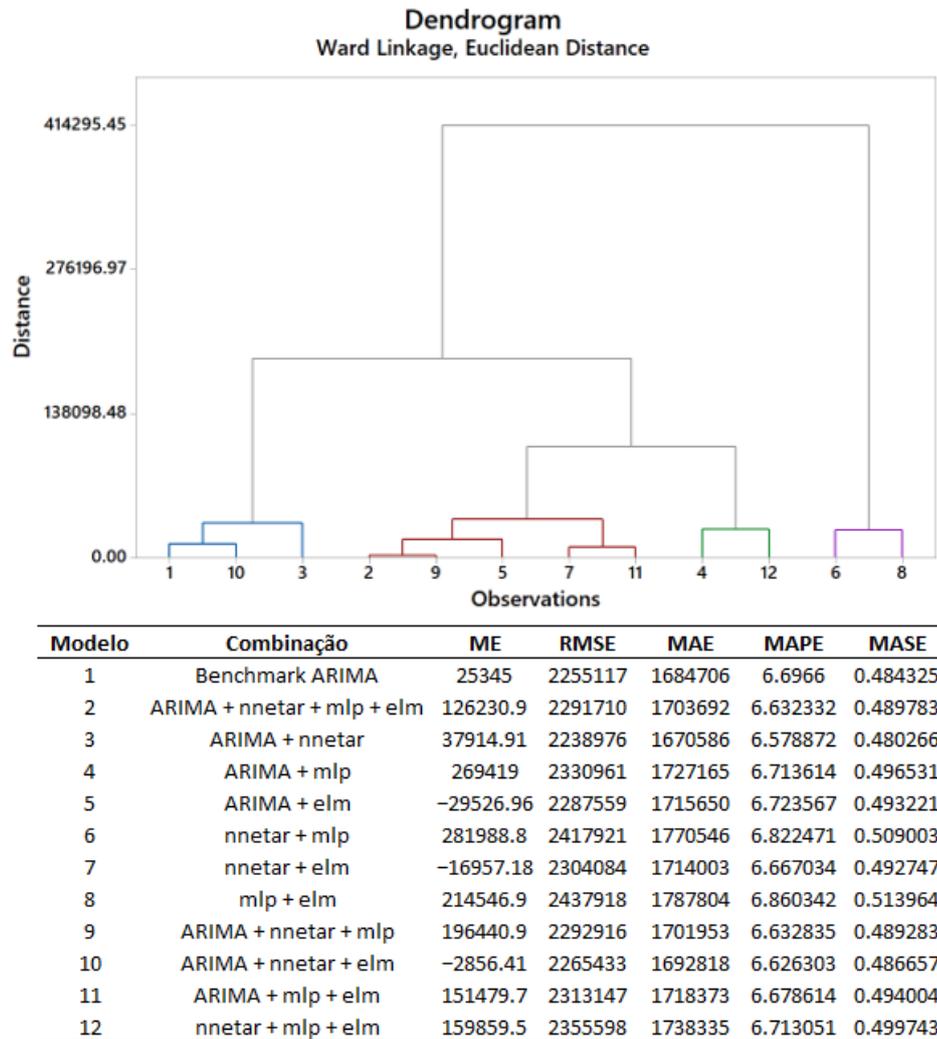
Outra característica recorrente observada nos trabalhos mais recentes é a opção por intervalos de tempo cada vez mais curtos, resultantes da necessidade de previsões de curto-prazo, como as séries de base horárias (SU *et al.*, 2019). Também são comuns regimes em base diária, semanal e quinzenal, sobretudo quando o problema trata de previsões de demanda/consumo de fontes de energia. Há uma quantidade muito grande de trabalhos de previsão para as fontes de energia renováveis ou para a previsão de variáveis relacionadas à exploração econômica dessas fontes, a saber: dinâmicas das propriedades do vento (velocidade e intensidade), radiação solar, gás natural, força das ondas etc.

Bacci *et al.* (2019) desenvolveram um método de combinação de previsões baseado em arranjo de misturas e análise fatorial denominado FA-NBI. Este trabalho, do qual o proponente desta tese é coautor, utilizou o método de Interseção Normal à Fronteira (NBI) para encontrar os pesos ótimos das combinações de previsões para a demanda de café no Brasil. Para tanto, o mapeamento da influência dos pesos sobre quatorze métricas de performance de previsão das combinações foi modelado empregando-se um arranjo de misturas do tipo “Simplex Lattice” {q, m}, com 61 experimentos, os quais descreviam o comportamento de métricas, como MSE, MAPE, MAE, entre outras quando se atribuíam combinações de pesos diferentes aos resíduos de previsão dos métodos DES, WM, ARIMA (1, 1, 1) e ARIMA (2, 2, 3). Para condensar as 14 métricas em apenas dois indicadores de performance multivariados, os autores propuseram a utilização de escores rotacionados de fator. O método NBI foi, então, aplicado às superfícies de resposta dos escores rotacionados de fator, gerando uma Fronteira de Pareto para o problema. O método é finalizado com a escolha do ponto da fronteira que apresentou a máxima razão entre a entropia (S) do vetor de pesos e seu respectivo erro percentual global (EPG). O EPG é uma métrica univariada que considera a distância entre um resultado ótimo multiobjetivo e os ótimos individuais. Uma vantagem da utilização do arranjo de mistura é a possibilidade de se modelar o comportamento de métricas de desempenho do método NBI simultaneamente com o mapeamento dos pesos nas combinações. Embora esta tese tenha sido desenvolvida como continuação do trabalho de Bacci *et al.* (2019), diversas atualizações e adaptações foram adicionadas ao método original, as quais serão especificadas ao longo deste texto.

Em relação ao emprego específico de combinações de métodos para a previsão de demanda de gás natural, pode-se citar o trabalho de Karabiber e Xydis (2020), no qual modelos ARIMA são combinados com três tipos de redes neurais artificiais: uma rede neural do tipo

MLP convencional, uma segunda rede denominada NNETAR (que é uma rede neural com uma única camada oculta) e uma outra rede denominada de ELM (*Extreme learning machining*, que é basicamente também uma SLFN (*Single-hidden layer feedforward neural network*)). Doze tipos de combinações diferentes (empregando pesos iguais) foram utilizados na previsão de curto prazo (“*one-day-ahead*”) de gás natural para quatro regiões da Dinamarca. Como também será evidenciado nesta tese, o trabalho de Karabiber e Xydis (2020) pontua a necessidade de se utilizarem várias métricas de performance de previsão para realizar a seleção das melhores combinações. No trabalho desses autores, as métricas MSE, MAE, RMSE, MAPE e MASE foram escolhidas.

Realizando uma análise estatística dos resultados disponibilizados pelos autores para essas métricas, no caso de uma das regiões estudadas, observam-se altos valores para os coeficientes de correlação entre elas e um elevado nível de semelhança entre grupos de combinações de modelos. Uma análise PCA aplicada a tais métricas de performance revela que a explicação do primeiro componente principal é de 96,6% ( $\lambda=3.86$ ), sendo os demais autovalores menores do que 1 e com um autovalor mínimo nulo, o que caracteriza a presença de variáveis redundantes no conjunto. Nesse caso, as métricas redundantes eram MAE e MASE. Aplicando-se o segundo critério de Jolliffe, verifica-se a possibilidade de eliminação sucessiva das métricas MASE, RMSE e MAPE, restando apenas a métrica MAE. Comparando-se as 12 combinações de ARIMA e RNA's por meio de uma análise hierárquica de cluster, verifica-se que as combinações 2 e 9 produzem resultados idênticos, enquanto que outros pares (1 e 10, 7 e 11, 4 e 12, 6 e 8) apresentam resultados extremamente semelhantes, indicando que muitos modelos de combinação não precisavam ter sido usados. Um resumo dessa análise estatística dos resultados de Karabiber e Xydis (2020) é ilustrado a seguir. Estes resultados mostram que os argumentos estabelecidos inicialmente nesta tese são coerentes e capazes de suportar as hipóteses estabelecidas no preâmbulo deste documento.



**Figura 2-1 Agrupamentos das 12 combinações de modelos de previsão propostas por Karabiber e Xydis (2020) considerando a similaridade entre as métricas de performance.**

**Fonte: Próprio autor**

## 2.4 MÉTRICAS DE PERFORMANCE PARA OS MÉTODOS DE PREVISÃO

Conforme se observa na literatura de combinações de métodos de previsão, a maior parte dos trabalhos avalia a qualidade das previsões com base em múltiplas métricas de performance embora muitas delas sejam redundantes. Apesar de serem bastante semelhantes, as diversas métricas também contêm particularidades que podem ser úteis para uma adequada definição dos pesos ótimos que podem ser atribuídos às combinações. Tais peculiaridades se relacionam com o nível de “Desempenho” – entendido como a mínima diferença entre um valor previsto e o valor real – e com o nível de “Precisão”, o qual está associado ao nível de confiança dos valores previstos representados por variâncias de previsão ou de larguras de intervalos de confiança para valores previstos “*one-step-ahead*”. De acordo com as métricas mais recorrentes

citadas no trabalho de Liu *et al.* (2021) para avaliação de performance de métodos usados para a previsão de demanda de gás natural e em consonância com as métricas sugeridas por Bacci *et al.* (2019), optou-se pela utilização das seguintes métricas de desempenho e de precisão:

### **Métricas de Desempenho**

- *Mean Absolute Error* (MAE);
- *Median Absolute Error* (MdMAE);
- *Mean Square Error* (MSE);
- *Root Mean Square Error* (RMSE);
- *Mean Absolute Scaled Error* (MASE);
- *Root Mean Square Percentage Error* (RMSPE);
- *Mean Absolute Percentage Error* (MAPE);
- *Symmetric Mean Absolute Percentage Error* (sMAPE);
- Coeficiente Estatístico de Theil U1;
- *Median Absolute Percentage Error* (MdAPE);
- *Symmetric Median Absolute Percentage Error* (SMdAPE).

### **Métricas de Precisão (Variabilidade ou estabilidade)**

- Variância (VAR);
- Desvio Padrão (SD);
- Coeficiente Estatístico de Theil U2.

Nos subcapítulos do 2.4.1, cada uma dessas métricas será explicada individualmente.

## **2.4.1 Métricas de Desempenho**

### **2.4.1.1 Mean Absolute Error (MAE)**

O MAE faz a mensuração do erro absoluto. É comumente usado para medir dissimilaridades entre duas séries (FRIAS-PAREDES *et al.* 2017; FRIAS-PAREDES *et al.* 2018). Percebe-se que seu uso é bem difundido entre as pesquisas que utilizam métricas de precisão para demonstrar a capacidade do método na previsão em questão. As métricas de desempenho demonstram a capacidade de um método conseguir uma previsão dos dados reais ou atuais (dados dentro da amostra) quando um modelo de previsão é ajustado a esses dados, ou mesmo para prever dados para períodos futuros (fora da amostra) cujos valores não foram utilizados para desenvolver o modelo de previsão (MAKRIDAKIS, 1993). Wun e Pearn (1991)

citam que, ao lado do MSE, o MAE é a principal medida de desempenho de modelos de previsão, sendo matematicamente descrita tal como a Equação 2.9:

$$MAE = \frac{1}{n} \sum_{t=1}^n |y_t - \hat{y}_t| \quad (2.9)$$

Onde:  $y_t$  é a  $t$ -ésima observação real da série e  $\hat{y}_t$  o valor previsto por um dado modelo de previsão para ela.

#### 2.4.1.2 Median Absolute Error (MdMAE)

Embora seja uma medida bastante difundida, médias aritméticas sofrem uma influência considerável de dados extremos (*outliers*), o que a torna uma medida pouco robusta de tendência central. Dessa forma, os erros absolutos podem ser melhor representados por uma medida não paramétrica, tal como a mediana.

Nesse sentido, Bacci *et al.* (2019) conceituam o MdMAE como uma medida robusta de variabilidade de uma amostra univariada de dados quantitativos descritos pela Equação 2.10:

$$MdMAE = \frac{1}{n} \text{median} |y_t - \hat{y}_t| \quad (2.10)$$

Onde:  $n$  é o número de termos da série,  $y_t$  é a  $t$ -ésima observação real da série  $\hat{y}_t$  o valor previsto por um dado modelo de previsão para ela.

#### 2.4.1.3 Mean Square Error (MSE)

Segundo Thompson (1990), o MSE é uma medida de desempenho que penaliza os grandes erros de previsão mais severamente do que outras estatísticas de precisões comuns. É útil para comparar a precisão da previsão sob perda quadrática. Wang *et al.* (2017) destacam que o MSE é usado para avaliar a eficácia de previsões determinísticas. Há pouco acordo, porém, em como utilizar o MSE para avaliar a precisão geral em muitas séries.

$$MSE = \frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2 \quad (2.11)$$

Onde:  $n$  é o número de termos da série,  $y_t$  é a  $t$ -ésima observação real da série  $\hat{y}_t$  o valor previsto por um dado modelo de previsão para ela.

#### 2.4.1.4 Root Mean Square Error (RMSE)

Como se observa na Equação (2.11), o MSE é uma métrica do tipo “variância” e, como tal, não tem um significado físico ou um fácil entendimento. Além disso, é uma medida que não está no mesmo sistema de unidade da média da série. Então, da mesma forma como se utiliza o desvio padrão para se expressar a variabilidade de um conjunto de dados, emprega-se o RMSE para avaliar a precisão geral dos métodos de previsão. Dessa forma, o RMSE é tão somente a raiz quadrada do MSE (DRAPER *et al.*, 2013; CALASAN *et. al*, 2020; LIU *et al.*, 2021) e pode ser escrito como:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (2.12)$$

Onde:  $n$  é o número de termos da série,  $y_i$  é a  $i$ -ésima observação real da série  $\hat{y}_i$  o valor previsto por um dado modelo de previsão para ela.

Armstrong e Collopy (1992), juntamente a Armstrong (2001), recomendam cautela no uso do RMSE na avaliação de performance de métodos de previsão para séries que contenham *outliers*, pois, como se trata de uma métrica “quadrática” e os *outliers* são dados mais difíceis de serem previstos, o erro aumentaria muito, e seu quadrado conduziria a uma explosão da métrica.

#### 2.4.1.5 Mean Absolute Scaled Error (MASE)

O MASE é uma medida de performance de previsão baseada em “erros escalonados” que apresenta melhor desempenho do que outras métricas baseadas em erros relativos. Embora medidas relativas tenham a capacidade de remover o efeito da escala dos dados sobre os resultados da previsão, tais erros possuem uma distribuição de probabilidade com média indefinida e variância infinita (HYNDMAN e KOEHLER, 2006; KIM e KIM, 2016). Ademais, medidas relativas só podem ser computadas quando existem vários valores previstos para a mesma série, o que as impede de serem usadas para medições “*out-of-sample*” em horizontes de uma única previsão (*one-step-ahead*). Escalonando, portanto, o erro da métrica MAE para uma série “*in-sample*” simples de *random walk* ( $e_t$ ), obtém-se o MASE. Tal medida é menos sensível a *outliers*, mais facilmente interpretado e menos variável em pequenas amostras. O MASE é definido matematicamente como:

$$q_t = \frac{e_t}{\frac{1}{n-1} \sum_{i=2}^n |Y_i - Y_{i-1}|} \quad (2.13)$$

$$MASE = \text{mean}(|q_t|) \quad (2.14)$$

Onde:  $e_t$  é uma série do tipo *random walk*, ou seja, uma constante mais um erro aleatório normalmente distribuído com parâmetros  $N(0, \sigma^2)$ ;  $n$  é o número de termos da série,  $y_i$  é a  $i$ -ésima observação da série,  $y_{i-1}$  é uma observação anterior à  $i$ -ésima.

#### 2.4.1.6 Root Mean Square Percentage Error (RMSPE)

De acordo com Tian *et al.* (2020), o RMSPE é uma métrica mais robusta de performance, sendo menos sensível à presença de *outliers*, além de ser expresso, matematicamente, como uma razão entre o RMSE e a média da série, tal como descrito na Equação (2.15):

$$RMSPE = \frac{\sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}}{\bar{y}} \quad (2.15)$$

Onde:  $n$  é o número de termos da série,  $y_i$  é a  $i$ -ésima observação real da série,  $\hat{y}_i$  representa o valor previsto por um dado modelo de previsão e  $\bar{y}$  representa a média da série temporal.

#### 2.4.1.7 Mean Absolute Percentage Error (MAPE)

Segundo Kim e Kim (2016), o erro médio percentual absoluto é uma das medidas mais difundidas para avaliação da precisão dos métodos de previsão. Os autores destacam, porém, que o MAPE tem a desvantagem de produzir valores infinitos ou indefinidos quando os valores reais são zero ou perto de zero, podendo produzir *outliers*. Matematicamente, o MAPE pode ser escrito da seguinte forma:

$$MAPE = \frac{1}{N} \sum_{t=1}^n \left| \frac{A_t - F_t}{A_t} \right| \quad (2.16)$$

Onde:  $N$  é o número de dados na série temporal;  $A_t$  denota os valores reais da série no instante  $t$  e  $F_t$  os valores da previsão para este mesmo ponto (KIM e KIM,2016; MYTTENAERE et al., 2016).

Lam *et al* (2001) destacam que MAPE pode ser uma boa alternativa para minimizar o erro das combinações de métodos de previsão. Mittenaere *et al.* (2016) enfatizam que o MAPE é frequentemente usado quando a quantidade a prever permanece acima de zero. MacKenzie (2011) pontua que o MAPE é atraente, pois penaliza a sub e a sobre-previsão em relação ao resultado real de forma simétrica. Armstrong e Collopy (1992), por outro lado, citam que o MAPE é relevante apenas para dados dispostos na forma *ratio-scaled*, ou seja, dados com um zero significativo. Outra desvantagem é o fato de que o MAPE penaliza mais severamente os erros de previsão que excedem o valor real do que aqueles que subestimam tal valor.

#### 2.4.1.8 Symmetric Mean Absolute Percentage Error (sMAPE)

MacKenzie (2011) destaca que, para uma determinada previsão, os resultados reais acima e abaixo da previsão são tratados de forma assimétrica, e o MAPE, portanto, penaliza esses desvios de forma diferente. Como estratégia de considerar esses desvios de maneira simétrica, o sMAPE é uma opção bastante adequada. Kim e Kim (2016) explicam que o sMAPE é um MAPE modificado em que o divisor é a metade da soma dos valores reais e previstos, tal como descrito pela Equação (2.17).

$$sMAPE = \frac{1}{n} \sum_{t=1}^n \frac{|\hat{y}_t - y_t|}{(y_t + \hat{y}_t)/2} \quad (2.17)$$

#### 2.4.1.9 Coeficiente Estatístico de Theil - $U_1$

Considere que  $n$  observações de uma variável  $y_t$  sejam dadas, para as quais existe um conjunto de valores previstos associados,  $\hat{y}_t$ . Então, o coeficiente estatístico de Theil  $U_1$  será dado por:

$$U_1 = \frac{\left[ \frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2 \right]^{\frac{1}{2}}}{\left[ \frac{1}{n} \sum_{t=1}^n (y_t)^2 \right]^{\frac{1}{2}} + \left[ \frac{1}{n} \sum_{t=1}^n (\hat{y}_t)^2 \right]^{\frac{1}{2}}} \quad (2.18)$$

A estatística  $U_1$  é limitada ao intervalo fechado  $[0, 1]$ , com valores iguais a zero, representando a previsão perfeita; de forma oposta, valores altos dessa estatística, aproximando-se do limite superior de  $U_1=1$ , são indicadores típicos de inexatidão crescente das previsões (BLIEMEL, 1973).

#### 2.4.1.10 Median Absolute Percentage Error (MdAPE)

Segundo Armstrong e Collopy (1992), usar medianas é uma maneira adequada de representar séries diminuindo a influência de valores extremos superiores e inferiores sobre o valor médio. A mediana APE (MdAPE) reduz o viés em favor de previsões baixas, oferecendo uma vantagem sobre MAPE. Fornece uma regra de corte padrão, e isso ajuda na comparação de estudos que são relatados na literatura. Também pode ser utilizada para modelar o impacto dos *outliers*.

$$MdAPE = \text{median}_{t=1, \dots, n} \left| 100 \times \frac{y_i - \hat{y}_i}{y_i} \right| \quad (2.20)$$

#### 2.4.1.11 Symmetric Median Absolute Percentage Error (sMdAPE)

É uma medida de erro de previsão pouco influenciada por valores extremos de  $y_t$ , além de ser mais robusta do MdAPE. Foi utilizada na M-3 *Competition* realizada por Makridakis e Hibon (2000) em seu artigo “*The M3-competition: Results, conclusions and implications*”. Matematicamente, o sMdAPE pode ser expresso como:

$$sMdAPE = \text{median}_{t=1, \dots, n} \frac{|\hat{y}_t - y_t|}{(y_t - \hat{y}_t)/2} \quad (2.21)$$

### 2.4.2 Métricas de Precisão

#### 2.4.2.1 Variância

Segundo Montgomery (2009), a dispersão ou a variabilidade dos dados de um conjunto é expressa por momento de segunda ordem denominado “variância”. Para o caso de uma série de resíduos  $e_t$ , definidos como a diferença entre  $y_i$  (a  $i$ -ésima observação real de uma série) e o valor previsto  $\hat{y}_i$ , a variância é expressa no quadrado das unidades da variável original.

$$VAR = s^2 = \frac{\sum_{t=1}^n (e_t - \bar{e}_t)^2}{n - 1} \quad (2.22)$$

#### 2.4.2.2 Desvio Padrão

Segundo Montgomery (2009), o desvio padrão é a raiz quadrada da variância e refere-se a uma medida de propagação ou de dispersão expressa na unidade original.

$$SD = \sqrt{\frac{\sum_{t=1}^n (e_t - \bar{e}_t)^2}{n-1}} \quad (2.23)$$

#### 2.4.2.3 Coeficiente Estatístico de Theil - U<sub>2</sub>

Bacci *et al.* (2019) enfatizam que o coeficiente estatístico de Theil U<sub>2</sub> é um coeficiente estatístico de qualidade da previsão, sendo usado para fazer comparações entre a previsão realizada por um determinado método e a previsão realizada pelo método *naïve*, assim considerado como se o valor previsto fosse a própria série (*no-change*). Dessa forma, a estatística U<sub>2</sub> é a razão entre o RMSE do modelo de previsão proposto e o RMSE da própria série, razão esta que pode ser representada por:

$$U_2 = \frac{\left[ \frac{1}{n} \sum_{t=1}^n (\hat{y}_t - y_t)^2 \right]^{\frac{1}{2}}}{\left[ \frac{1}{n} \sum_{t=1}^n (y_t)^2 \right]^{\frac{1}{2}}} \quad (2.19)$$

A estatística U<sub>2</sub> é também limitada ao intervalo [0, 1], mas, enquanto valores iguais a zero para U<sub>2</sub> representam a previsão perfeita, o limite superior de U<sub>2</sub>=1 representa uma precisão que equivale ao método *naïve*.

## 2.5 PROJETO DE EXPERIMENTOS DE MISTURA

As somas ponderadas representam a forma mais tradicional de combinação de métodos de previsão e, considerando que os pesos delas podem ser relativos, sua soma totaliza a unidade. Portanto, qualquer soma ponderada relativa é uma convexa dos produtos entre proporções e valores esperados (OLIVEIRA *et al.*, 2011). Então, variando-se as proporções dentro de uma região convexa (fechada) e restrita, obtém-se uma matriz de “pesos” representativa de uma malha de pontos sobre esse espaço. Este é justamente o conceito de “arranjo de misturas” (MDE, do inglês *Mixture Design of Experiments*). Segundo Mason *et al.* (2003), a esse respeito, o arranjo de misturas é usado quando a resposta de interesse é influenciada não pelas quantidades reais dos fatores (componentes de mistura), mas apenas por suas proporções relativas. Dentro da região simplex, as proporções podem ser uniformemente espalhadas, além de apresentarem alternativas em regiões específicas, como o centro da região ou pontos equidistantes desse centro (MENDES *et al.*, 2016). Myers *et al.* (2009) enfatizam que um experimento de misturas é um tipo especial de superfície de respostas na qual os fatores são os componentes de uma mistura, e a resposta é uma função das proporções de cada componente.

De acordo com Coronado (2015), nesse viés, a técnica presume que, quando as condições do processo são mantidas constantes, as variáveis de resposta são dissuadentes pela proporção dos componentes da mistura. Dessa forma, não é difícil se pensar que os pesos nas somas convexas são equivalentes às proporções no arranjo de misturas. Assim, a matriz simplex pode ser usada como uma malha de opções de pesos para as combinações de métodos também. Essa ideia encontra respaldo em trabalhos anteriores observados na área de porfolio de ações (MENDES *et al.*, 2016), em portfolios de contratos de energia elétrica (OLIVEIRA *et al.*, 2011), em métodos de otimização multiobjetivo (MOP), como o trabalho de Gomes *et al.* (2013), na combinação de métodos de previsão de séries temporais convencionais para a previsão de demanda de café no Brasil (BACCI *et al.*, 2019) e na combinação de diferentes arquiteturas de redes neurais artificiais para a previsão de quantidade de energia fotovoltaica gerada (MOREIRA *et al.*, 2021).

Em todos esses trabalhos, a principal vantagem de se utilizar o arranjo de misturas como matriz de pesos para diferentes tipos de combinações é a capacidade de se mapear e de modelar as características do objeto de estudo. No caso de Mendes *et al.* (2016), as respostas consideradas no MDE eram relativas às propriedades do portfolio de ações, tais como o retorno, o risco, a entropia e a curtose. No trabalho de Oliveira *et al.* (2011), por sua vez, modelou-se a relação entre os pesos do MDE e as características de um portfolio de contratos entre empresas consumidoras e fornecedoras de energia elétrica, caracterizado pelo retorno do portfolio e o seu risco medido pela métrica *CVar* (*Conditional Value-at-Risk*). Na utilização do MDE para a otimização multiobjetivo, tal como descrita no trabalho de Gomes *et al.* (2013), o benefício maior consistiu na modelagem da qualidade da otimização, expressa pela minimização da distância de Mahalanobis de diversas soluções Pareto-ótimas e os respectivos valores ótimos individuais. Frente a isso, depreende-se desse trabalho que, ao se utilizar polinômios canônicos (não lineares) para estabelecer uma função de transferência entre a característica do objeto de estudo (no caso, a distância de Mahalanobis) e a matriz de pesos, pesos ótimos diferentes dos disponibilizados pela matriz podem ser encontrados a partir da otimização dessa função de transferência. Este é um dos principais benefícios desta abordagem híbrida: modelar as funções de interesse em termos dos pesos a fim de que um algoritmo de otimização possa encontrar pesos ainda melhores do que aqueles previamente definidos pelo arranjo. Se apenas os pesos descritos pela matriz fossem utilizados, a abordagem seria meramente discreta e, com isso, limitar-se-ia às possibilidades fornecidas pelo arranjo. Esta estratégia de utilizar o arranjo de misturas para modelar e para explicar as relações entre os pesos das funções em um problema MOP e as características de performance da otimização multiobjetivo foi intensamente utilizada

nos últimos anos em trabalhos como os de Rocha *et al.* (2015, 2017a, 2017b, 2020), Aquila *et al.* (2018) e Gaudêncio *et al.* (2019). Em todos eles, caracterizou-se a performance do método de otimização multiobjetivo por uma relação entre o erro percentual global (medida de afastamento entre os ótimos individuais e global das funções objetivo) e a entropia (medida da diversificação dos pesos dentro da otimização). Em muitos casos, uma subrotina de otimização foi também empregada para se encontrar o melhor vetor de pesos para o problema.

O mesmo raciocínio foi aplicado aos trabalhos mais recentes de Bacci *et al.* (2019) e de Moreira *et al.* (2021). Neles, o MDE foi usado para gerar, além de um rol de proporções diferentes para as combinações, modelos de características de performance e de precisão dos métodos de previsão, como MSE, MAPE, MAE, sMAPE, SdMAPE. Essas métricas podem ser modeladas em função dos pesos, e sua minimização conduz à descoberta de proporções ótimas, provavelmente diferentes daquelas dadas precisamente pela malha de pesos. É nesse contexto que esta tese também propõe a utilização de arranjos de misturas para a definição das proporções ótimas de pesos para a combinação de métodos de previsão, com vistas aos menores valores possíveis dos indicadores de desempenho e de precisão.

Mas o que é na verdade uma mistura? Cornell (2002) define que o problema de mistura é aquele no qual a resposta mensurada depende apenas das proporções dos componentes presentes na mistura, e não da sua quantidade. De acordo com Myers *et al.* (2009), em uma mistura com  $q$  componentes, cada um deles apresenta uma proporção  $x_i$ , tal que:

$$x_i \geq 0, \quad i = 1, 2, \dots, q \quad (2.24)$$

A Equação (2.24) expressa que proporções serão sempre positivas e que sua soma deve ser constante. Uma restrição linear, como descrita na Equação (2.25), garante que a soma das proporções seja unitária, porém, ao mesmo tempo, também torna todas as proporções  $x_i$  dependentes entre si, o que torna as colunas do arranjo de mistura linearmente dependentes. Por decorrência, os coeficientes estimados não serão mais independentes e, por vezes, apresentarão valores significativamente altos de inflação da variância (VIF, do inglês *Variance Inflation Factor*). Essa característica particular do MDE o difere de um arranjo típico de superfície da resposta. Então, pode-se escrever que:

$$\sum_{i=1}^q x_i = x_1 + x_2 + \dots + x_q = 1 \quad (2.25)$$

Segundo Myers *et al.* (2009), o espaço experimental da mistura é uma região fechada, formada por um sistema de coordenadas “triangular” denominado de “Simplex”. Nessa região, os pontos do arranjo se posicionam nos vértices (quando as misturas são “puras” – ou formadas por um só componente), nas bordas ou faces (formando as misturas “binárias”), ou no interior desse simplex, em pontos equidistantes dos vértices (centróide) ou em pontos relativos ao baricentro do triângulo que compreende a região (pontos axiais). O simplex lattice é apenas um conjunto de pontos uniformemente espaçados em um simplex. Assim, um arranjo simplex Lattice  $\{q, m\}$  é aquele projetado para  $q$  componentes e  $n$  experimentos, capazes de estimar os coeficientes de um polinômio canônico de grau  $m$ ; o arranjo consiste em pontos definidos pelas proporções tomadas para cada componente, igualmente espaçados entre 0 e 1. A Equação (2.26) descreve estas proporções.

$$X_i = 0, \frac{1}{m}, \frac{2}{m}, \dots, 1, \quad i = 1, 2, \dots, q \quad (2.26)$$

Dessa forma, o número de pontos em um *simplex lattice design* é igual a:

$$N = \frac{(q + m - 1)!}{m! (q - 1)!} \quad (2.27)$$

Um modelo (ou polinômio canônico) de mistura de primeira ordem é dada por:

$$E(y) = \beta_0 + \sum_{i=1}^q \beta_i x_i \quad (2.28)$$

Porém, como a restrição  $X_1 + X_2 + \dots + X_q = 1$ , os coeficientes  $\beta_0, \beta_1, \dots, \beta_k$  podem ser reescritos, tal que:

$$X_q = 1 - \sum_{i=1}^{q-1} X_i \quad (2.29)$$

Uma outra alternativa é multiplicar alguns dos termos na primeira ordem do modelo por  $X_1 + X_2 + \dots + X_q = 1$  e depois simplificar. Assim, tem-se que:

$$\begin{aligned} E(y) &= \beta_0 (X_1 + X_2 + \dots + X_q) \sum_{i=1}^q \beta_i x_i \\ &= \sum_{i=1}^q \beta_i x_i \quad \text{onde } \beta_i^* = \beta_0 + \beta_i \end{aligned} \quad (2.30)$$

Então, de uma forma artificial, o termo independente da equação desaparece (por ter sido incorporado a todos os coeficientes dos termos lineares). Esse tipo de modelo é chamado de polinômio canônico de *Scheffé* de primeira ordem (*first-order mixture model*). Assim, os possíveis polinômios canônicos de mistura são:

a) Modelo Linear

$$E(y) = \sum_{i=1}^q \beta_i x_i \quad (2.31)$$

b) Modelo Quadrático

$$E(y) = \sum_{i=1}^q \beta_i x_i + \sum_{i<j=2}^q \beta_{ij} X_i X_j \quad (2.32)$$

c) Modelo Cúbico completo (Full Cubic)

$$\begin{aligned} E(y) = & \sum_{i=1}^q \beta_i x_i + \sum_{i<j=2}^q \beta_{ij} X_i X_j + \sum_{i<j=2}^q \delta_{ij} X_i X_j (X_i - X_j) \\ & + \sum_{i<j=3}^q \beta_{ijk} X_i X_j X_k \end{aligned} \quad (2.33)$$

d) Modelo Cúbico Especial (*Special Cubic*)

$$E(y) = \sum_{i=1}^q \beta_i x_i + \sum_{i<j=2}^q \beta_{ij} X_i X_j + \sum_{i<j=3}^q \beta_{ijk} X_i X_j X_k \quad (2.34)$$

Como nos modelos de superfície de resposta e de regressão tradicionais, a estimação dos coeficientes  $\beta_i$  é também realizada pelo algoritmo de mínimos quadrados ordinários (**OLS**, do inglês *Ordinary Least Squares*), de forma que, dada uma matriz experimental de misturas  $\mathbf{X}$  e um vetor de respostas  $\mathbf{y}$ , seja possível escrever que:

$$\frac{\partial \mathbf{L}}{\partial \boldsymbol{\beta}} = -2\mathbf{X}^T \mathbf{y} + 2\mathbf{X}^T \mathbf{X} \hat{\boldsymbol{\beta}} = 0 \therefore \hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} \quad (2.35)$$

Em relação à proposição desta tese, os polinômios canônicos de mistura poderão ser utilizados como uma função de transferência entre as métricas de performance de previsão (MAE, MAPE, RMSE, SdMAPE etc.) e os pesos atribuídos aos métodos de previsão individuais (Holt-Winters Aditivo (WA), Multiplicativo (WM) e Alisamento exponencial duplo (DES) que, neste caso, serão iguais às proporções definidas pelo arranjo de misturas escolhido.

No próximo item, serão descritas as formulações referentes à técnica estatística multivariada denominada “Análise Fatorial” e sua utilidade para reduzir a dimensionalidade do problema multiobjetivo quando os arranjos de mistura são aplicados às diversas métricas de performance, gerando diversos polinômios canônicos.

## 2.6 ANÁLISE FATORIAL COM EXTRAÇÃO POR COMPONENTES (PCFA)

A análise fatorial com extração por componentes principais (**PCFA**, do inglês *Principal component factor analysis*) é uma técnica estatística multivariada que visa à redução de dimensionalidade de um conjunto grande de dados formado por grupos de variáveis aleatórias correlacionadas (JOHNSON e WICHERN, 2007). Essa técnica permite organizar as informações desse conjunto em um número significativamente menor de fatores (ou grupos de variáveis), com perda mínima de informação (HAIR *et al.* 2010). A PCFA também pode ser caracterizada como uma abordagem estatística capaz de analisar as inter-relações entre um grande número de variáveis e de explicar essas variáveis em termos de suas dimensões comuns subjacentes.

Um modelo fatorial postula que um vetor aleatório  $\mathbf{X}$ , com  $p$  componentes, com um respectivo vetor de médias  $\boldsymbol{\mu}$  e matriz de variância-covariância  $\boldsymbol{\Sigma}$ , é linearmente dependente de algumas variáveis aleatórias não  $F_1, F_2, \dots, F_m$ , denominadas *Fatores comuns*, e  $p$  fontes adicionais de variação denominadas erros (ou fatores específicos), tal que:

$$\underset{(p \times 1)}{\mathbf{X}} - \underset{(p \times 1)}{\boldsymbol{\mu}} = \underset{(p \times m)}{\mathbf{L}} \underset{(m \times 1)}{\mathbf{F}} + \underset{(p \times 1)}{\boldsymbol{\varepsilon}} \quad (2.36)$$

Na Eq. 2.36,  $\mathbf{L}$  é a matriz de carregamento dos fatores que podem ser calculados a partir de uma decomposição spectral da matriz de variância-covariância, tal que:

$$\Sigma = \lambda_1 \mathbf{e}_1 \mathbf{e}_1^T + \lambda_2 \mathbf{e}_2 \mathbf{e}_2^T + \dots + \lambda_p \mathbf{e}_p \mathbf{e}_p^T = \begin{bmatrix} \sqrt{\lambda_1} \mathbf{e}_1 & \sqrt{\lambda_2} \mathbf{e}_2 & \dots & \sqrt{\lambda_p} \mathbf{e}_p \end{bmatrix} \begin{bmatrix} \sqrt{\lambda_1} \mathbf{e}_1^T \\ \sqrt{\lambda_2} \mathbf{e}_2^T \\ \vdots \\ \sqrt{\lambda_p} \mathbf{e}_p^T \end{bmatrix} = \mathbf{L} \mathbf{L}^T \quad (2.37)$$

Os carregamentos fatoriais (*factor loadings*) são, portanto, os autovetores da matriz de variância-covariância  $\Sigma$  escalonados pela raiz quadrada dos respectivos autovalores,  $\sqrt{\lambda_j}$ .

Como a covariância pode ser escrita como  $\Sigma = E(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^T$ , então, tem-se que:

$$\Sigma = E[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^T] = E[(\mathbf{L}\mathbf{F} + \boldsymbol{\varepsilon})(\mathbf{L}\mathbf{F} + \boldsymbol{\varepsilon})^T] = E[(\mathbf{L}\mathbf{F} + \boldsymbol{\varepsilon})(\mathbf{L}^T \mathbf{F}^T + \boldsymbol{\varepsilon}^T)] \quad (2.38)$$

Logo:

$$\Sigma = E[\mathbf{L}\mathbf{F}(\mathbf{L}^T \mathbf{F}^T) + \boldsymbol{\varepsilon}(\mathbf{L}^T \mathbf{F}^T) + (\mathbf{L}\mathbf{F})\boldsymbol{\varepsilon}^T + \boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T] \quad (2.40)$$

Portanto:

$$\Sigma = \underbrace{\mathbf{L}E(\mathbf{F}\mathbf{F}^T)\mathbf{L}^T}_{\mathbf{I}} + \underbrace{E(\boldsymbol{\varepsilon}\mathbf{F}^T)\mathbf{L}^T}_0 + \underbrace{\mathbf{L}E(\mathbf{F}\boldsymbol{\varepsilon}^T)}_0 + \underbrace{E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T)}_{\boldsymbol{\Psi}} = \mathbf{L}\mathbf{L}^T + \boldsymbol{\Psi} \quad (2.41)$$

Onde  $\boldsymbol{\Psi}$  é a matriz diagonal formada pelas variâncias específicas  $\psi_i$ , tal que  $\psi_i = \sigma_i^2 - h_i^2$ . O termo  $h_i^2$  denomina-se “comunalidade” e é calculado como a soma de quadrados dos carregamentos fatoriais associados à  $i$ -ésima variável dos  $m$  fatores comuns.

Como na teoria dos Componentes Principais, um conjunto original de dados pode ser também representado por fatores não correlacionados denominados “Escore Fatorial” apenas considerando a matriz de dados originais padronizados  $\mathbf{Z}$  e a matriz de carregamentos  $\mathbf{L}$ , tal que:

$$\mathbf{F} = \mathbf{Z}[\mathbf{L}(\mathbf{L}^T \mathbf{L})^{-1}] \quad (2.42)$$

Os carregamentos  $\mathbf{L}$  podem ainda ser rotacionados em várias direções usando uma transformação ortogonal de modo que uma estrutura de fatores mais simples seja alcançada, isto é, o método de rotação factorial pode conduzir à separação das variáveis originais em grupos mais fáceis de serem interpretados. O carregamento rotacionado  $\mathbf{L}^*$  pode ser calculado como:  $\mathbf{L}^* = \mathbf{L}\mathbf{T}$ , onde  $\mathbf{T}\mathbf{T}^T = \mathbf{T}^T \mathbf{T} = \mathbf{I}$ . Essa transformação não altera as comunalidades e as variâncias específicas.

Existem vários métodos de rotação factorial. O mais efetivo, porém, é o método **Varimax** que utiliza coeficientes rotacionados que são escalonados pela raiz quadrada das comunalidades, tal que  $\tilde{\ell}_{ij}^* = \tilde{\ell}_{ij}^* / \hat{h}_i$ . O objetivo do método Varimax é obter apenas alguns pesos significativos com os demais sendo iguallados a zero. Desse modo, o método pressupõe que a variância total entre os pesos de cada componente seja maximizada, tal que:

$$\text{Max } V = \frac{1}{p} \sum_{j=1}^m \left[ \sum_{i=1}^p \tilde{\ell}_{ij}^{*4} - \left( \sum_{i=1}^p \tilde{\ell}_{ij}^{*2} \right)^2 / p \right] \quad (2.43)$$

A Análise Fatorial com escores rotacionados tem muitas aplicações possíveis. Nesta tese, os escores rotacionados serão utilizados como novas funções objetivo (ou, novos eixos da Superfície ou da Fronteira de Pareto) em um problema de otimização multiobjetivo resolvido pelo algoritmo NBI. Assim, se um problema multiobjetivo original contém  $p$  funções objetivo, depois da redução de dimensionalidade e do agrupamento das funções objetivo originais, o problema terá  $r$  funções objetivo de escores rotacionados. Dessa maneira, a Fronteira (ou Superfície) de Pareto poderá ser construída para grupos de funções. Especificamente, espera-se que, após a eliminação de métricas de performance redundantes, as métricas remanescentes possam ser agrupadas em fatores, de acordo com sua estrutura de correlação. Nesse viés, aplicar-se-á a modelagem por polinômios canônicos de misturas para os escores rotacionados de fator representativos de grupos de métricas de performance. Se o conjunto de métricas puder ser reduzido a apenas dois grupos e, respectivamente, a duas funções objetivo, o problema biobjetivo poderá ser resolvido por uma versão bidimensional do método NBI, cuja apresentação será realizada na próxima seção.

## 2.7 MÉTODO DE INTERSECÇÃO NORMAL À FRONTEIRA

O Método da Interseção Normal à Fronteira (**NBI**, do inglês *Normal Boundary Intersection*) criado por Das e Dennis (1998) é um método de otimização multiobjetivo utilizado para definir um conjunto de soluções Pareto-ótimas e viáveis em um dado problema. Este método, especificamente, permite a construção de uma Fronteira de Pareto com soluções distribuídas de forma contínua, uniforme e equiespaçadas mesmo em regiões não convexas da fronteira (AHMADI *et al.*, 2015).

Para sua consecução, o método NBI depende inicialmente da construção de uma matriz de soluções individuais denominada Matriz *Payoff*,  $\Phi$ , definida tal como descrito pela Equação (2.44).

$$\Phi = \begin{bmatrix} f_1^*(x_1^*) & \cdots & f_1(x_i^*) & \cdots & f_1(x_m^*) \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ f_i(x_1^*) & \cdots & f_i^*(x_i^*) & \cdots & f_i^*(x_m^*) \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ f_m(x_1^*) & \cdots & f_m(x_i^*) & \cdots & f_m^*(x_m^*) \end{bmatrix} \quad (2.44)$$

A diagonal principal da matriz *Payoff* é definida pela combinação de ótimos individuais, de modo que o valor mínimo de  $f_i(x)$ , obtido no ponto  $x_i^*$ , conduza a  $f_i^*(x_i^*)$ . Substituindo os pontos ideais  $x_i^*$  obtidos individualmente nas outras funções objetivo, definem-se demais elementos de cada linha de  $\Phi$ , a saber,  $f_i(x_i^*)$ . Cada linha de  $\Phi$  apresenta um Ponto de Utopia (melhor valor possível para uma dada função),  $f_i^*(x_i^*)$ , e os pontos de Nadir  $f_i^N$  (que é o pior valor para a função  $f_i(\mathbf{x})$  observado na  $i$ -ésima linha da *Payoff*). Estes dois valores são utilizados para realizar o escalonamento (adimensionalização ou normalização) da  $i$ -ésima função objetivo (VAHIDINASAB e JADID 2010; GAUDÊNCIO *et al.*, 2019). O ponto de utopia cai, geralmente, fora da região viável e é representado pelo vetor  $f_i^U = [f_1^*(x_1^*), \dots, f_i^*(x_i^*), \dots, f_m^*(x_m^*)]^T$ . O ponto de Nadir é representado pelo vetor  $f_i^N = [f_1^N, \dots, f_i^N, \dots, f_m^N]^T$ , onde  $f_1^N = \max f_1(x)$ , sujeito a  $s. t. : \{x \in R \mid g(x) \leq 0, h(x) = 0\}$  em um problema de “minimização”, ou seja, obtém-se o ponto de Nadir invertendo-se o tipo de otimização. Em outras palavras, nenhuma solução em um problema de otimização pode ser pior do que aquela obtida por “maximização”. Na prática, porém, o ponto de Nadir é substituído por um ponto denominado de “Pseudo-Nadir”, que não é simplesmente o pior resultado observado na  $i$ -ésima linha da *Payoff*, não sendo necessariamente obtido por uma otimização em sentido oposto ao original (VAHIDINASAB e JADID 2010; ROCHA *et al.*, 2017a; GAUDÊNCIO *et al.*, 2019).

A Fronteira de Pareto é fixada nos pontos ótimos individuais, os quais são denominados de “Pontos de Ancoragem”, porque a fronteira é fixada nos ótimos individuais. Ligando-se estes dois pontos de ancoragem por um segmento de reta, forma-se o que se denomina de “Linha de Utopia”. Durante a execução do NBI, a linha de Utopia é dividida em segmentos equiespaçados relativos aos pesos compreendidos no intervalo  $[0, 1]$ . Assim, uma reta perpendicular (ou Normal) à linha de Utopia move-se entre os pontos de ancoragem, na medida em que se alteram os pesos atribuídos às funções envolvidas na otimização (GAUDÊNCIO *et al.*, 2019).

Ahmadi *et al.* (2015) explicam que, como as múltiplas funções objetivo podem apresentar diferentes magnitudes, elas devem ser primeiramente normalizadas (ou escalonadas) em termos dos pontos de Utopia e Nadir, tal como descrito pela Equação (2.45) a seguir.

$$\bar{f}(x) = \frac{f_i(x) - f_i^U}{f_i^N - f_i^U}, \quad i = 1, \dots, m \quad (2.45)$$

Repetindo-se esse escalonamento para todas as linhas da *Payoff*, obtém-se uma nova matriz denominada de “*Payoff* escalonada”,  $\bar{\Phi}$ , tal como descrito em (2.46).

$$\bar{\Phi} = \begin{bmatrix} \bar{f}_1^*(x_1^*) & \dots & \bar{f}_1^*(x_i^*) & \dots & \bar{f}_1^*(x_m^*) \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \bar{f}_i^*(x_1^*) & \dots & \bar{f}_i^*(x_i^*) & \dots & \bar{f}_i^*(x_m^*) \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \bar{f}_m^*(x_1^*) & \dots & \bar{f}_m^*(x_i^*) & \dots & \bar{f}_m^*(x_m^*) \end{bmatrix} \quad (2.46)$$

Como resultado Simab *et al* (2018) enfatizam que o problema é resolvido em um espaço não dimensional. Gaudêncio *et al.* (2019) explicam, ainda, que as combinações convexas de cada fileira da  $\bar{\Phi}$  escalonada formarão o conjunto de pontos denominados *Convex Hull of Individual Minima* (CHIM), também chamado de linha de utopia para problemas biobjetivos. Em problemas multiobjetivos com mais que duas funções objetivo, os pontos de ancoragem formarão o hiperplano de utopia. Os elementos descritos anteriormente podem ser representados esquematicamente pela figura a seguir.

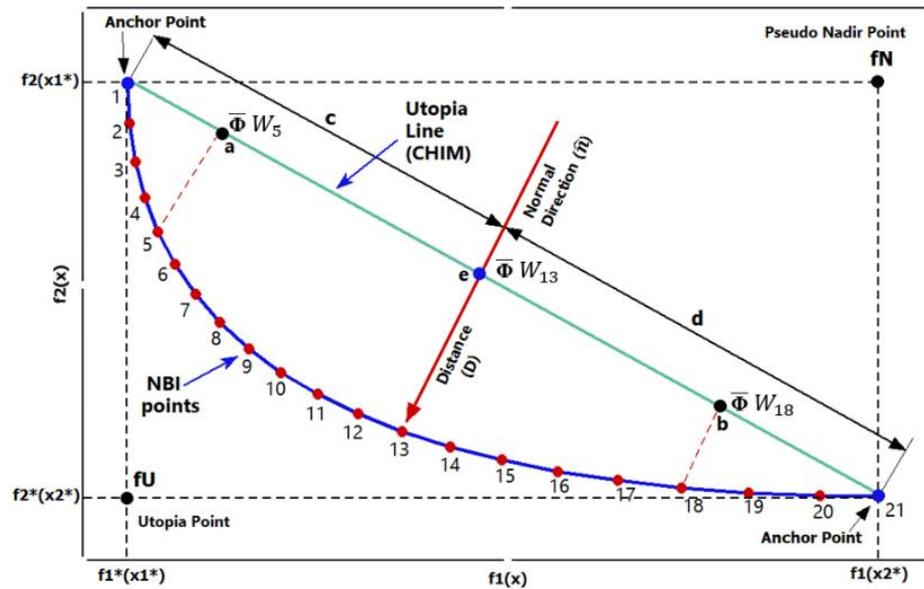


Figura 2-2 - Representação gráfica do método NBI.

Fonte: Próprio autor

Simab *et al.* (2018) explicam que a CHIM é indicada como a combinação dos pontos em cada fileira da  $\Phi$  escalonada. Cada ponto no espaço  $P(\beta_1, \beta_2, \dots, \beta_p)$  pode ser representado por  $P(\beta_1, \beta_2, \dots, \beta_p) = [(\beta_1 \bar{\Phi}_{11} + \dots + \beta_p \bar{\Phi}_{1p}) \dots (\beta_p \bar{\Phi}_{p1} + \dots + \beta_p \bar{\Phi}_{pp})]$ , onde  $\beta_u$  é um valor positivo em acordo com a equação:  $\sum_{i=1}^p \beta_i = 1$ . Consequentemente, o problema multi-objetivo primário é transformado em  $p$  subproblemas de um objetivo único, no qual a função objetivo de cada subproblema,  $D$ , é maximizar a distância entre o CHIM e a fronteira de Pareto. A representação matemática de cada problema acima mencionado seria a seguinte:

$$\text{Max } D \quad (2.47)$$

Sujeito a:

$$\beta + D_{\bar{n}} = \bar{F}(x) \text{ onde: } \{x \in R | g(x) \leq 0, h(x) = 0\} \quad (2.48)$$

Ahmadi *et al.* (2015) elucidam que o ponto  $x$  é mapeado para um ponto na normal, enquanto outras restrições garantem a viabilidade de  $x$  considerando o problema original. Rocha *et al.* (2017b) complementam que o  $D$  é um vetor perpendicular à linha de utopia.  $\beta$  é o vetor de pesos que representa diferentes pesos ao longo da linha de utopia. Simab *et al.* (2018) pontuam que pelo NBI é possível controlar a densidade das soluções ideais de Pareto, determinando os valores de  $\beta$ . Os autores enfatizam, porém, que o maior número de soluções ideais de Pareto gera um custo de maior carga computacional, o que obriga o tomador de decisão a fazer um *trade-off* entre a densidade de soluções ideais de Pareto.

O problema de otimização representado pelo sistema de equações (2.47 e 2.48) pode ser resolvido iterativamente para diferentes valores de  $w$ , o que cria, por conseguinte, uma Fronteira de Pareto igualmente espaçada. Uma escolha comum proposta por Rocha *et al.* (2017) é fazer  $w_n = 1 - \sum_{i=1} w_i$ . Por uma questão de simplificação, o parâmetro conceitual  $\mathbf{D}$  pode ser algebricamente eliminado das Equações (2.47 e 2.48), dado que ele está tanto na função objetivo quanto nas restrições de igualdade. Para o caso bidimensional, esta expressão simplificada pode ser reescrita como (GAUDÊNCIO *et al.*, 2019):

$$\begin{aligned}
 & \text{Min } \bar{f}_1(\mathbf{x}) \\
 & \text{s.t.: } \bar{f}_1(\mathbf{x}) - \bar{f}_2(\mathbf{x}) + 2w - 1 = 0 \\
 & \quad g_j(\mathbf{x}) \geq 0 \\
 & \quad 0 \leq w \leq 1
 \end{aligned} \tag{2.49}$$

Onde  $\bar{f}_1(\mathbf{x})$  e  $\bar{f}_2(\mathbf{x})$  representam duas funções objetivo escalonadas (normalizadas).

Após a resolução de todos os subproblemas propostos no método NBI, um conjunto de soluções Pareto-ótimas e viáveis estará disponível, entretanto, o excesso de soluções ótimas pode ser tão desafiador quanto a sua escassez. Então, a escolha do ponto mais adequado entre os Pareto-ótimos passa pela avaliação da qualidade das soluções em termos de distância em relação aos ótimos individuais e em relação à diversificação dessas soluções. Para auxiliar o tomador de decisão nesta tarefa, inúmeros trabalhos recentes propuseram estratégias de escolha baseadas em uma combinação entre o Índice de Entropia de Shannon ( $\mathbf{S}$ ) (ROCHA *et al.*, 2017a) e o Erro Percentual Global (GPE, do inglês *Global Percentage Error*) proposto inicialmente por Gomes *et al.* (2013).

O índice de Entropia de Shannon ( $S_{\text{Total}}$ ) é usado para se avaliar quão diversificada é uma solução Pareto-ótima. Para vetores de pesos indicativos de soluções individuais, nesse sentido, há uma concentração da solução em uma única função objetivo; portanto, a diversificação é mínima. De outro modo, se os pesos atribuídos às funções forem todos iguais, a diversificação será máxima. Em termos matemáticos, o índice de Entropia de Shannon pode ser dado pela seguinte equação (GAUDÊNCIO *et al.*, 2019):

$$S_{\text{Total}}(x) = - \sum_{i=1}^m w_i \ln(w_i) \tag{2.50}$$

De outra forma, o decisor pode basear-se no critério de máxima proximidade da solução ótima em relação ao vetor de ótimos individuais. Nesta tese, adotar-se-á a  $GPE_{Total}$ , definida de acordo com a Equação (2.51). Cada GPE de uma solução Pareto-ótima é calculado definindo quão longe o ponto analisado ( $y_i^*$ ) está do valor ideal da função objetivo, ou seja, de um determinado alvo ( $T_i$ ) (GOMES *et al.*, 2013; GAUDÊNCIO *et al.*, 2019; ROCHA *et al.*, 2015):

$$GPE_{Total} = \sum_{i=1}^m \left| \frac{y_i^*}{T_i} - 1 \right| \quad (2.51)$$

Na qual:

- $y_i^*$  – Valor da resposta Pareto-ótima;
- $T_i$  – Alvo definido;
- $m$  – Número de objetivos.

A melhor solução pode ser escolhida como aquela que apresenta a maior relação ( $S_{Total} / GPE_{Total}$ ).

## 2.8 MÉTODO DE JOLLIFFE

A redução de dimensionalidade dos dados é uma estratégia bastante eficiente para diminuir a quantidade de processamento necessário preservando a maior parte das informações contidas no conjunto original (CUMMING e WOOLF, 2007). Para muitos conjuntos de dados em grande escala, é necessário reduzir a dimensionalidade ao ponto no qual a exploração e a análise possam ser realizadas de forma eficiente e em tempo apropriado (HIDALGO *et al.*, 2020). Em geral, a redução de dimensionalidade não pressupõe (pelo menos diretamente) a eliminação de variáveis, mas, sim, de dimensões. Nesse caso, a eliminação das variáveis só deve ocorrer se elas forem identificadas como “redundantes”, ou seja, a informação contida nelas já está perfeitamente representada em outras variáveis. Em outras palavras, conjuntos de dados podem ser comprimidos e representados em um número menor de dimensões se possuírem estruturas de correlação significativas. Esta compressão de dimensões pode ser feita de muitas maneiras diferentes, mas, indubitavelmente, a mais apropriada é a compressão por combinações lineares independentes. Esta estratégia permite que as variáveis originais sejam combinadas em novas variáveis não observáveis e, em geral, o número dessas combinações lineares será inferior ao número de variáveis originais. Até certo ponto, esse tipo de estratégia preserva a estrutura e as informações das variáveis originais, desde que o número de combinações retidas na análise não seja, em geral, igual ao número de variáveis. A técnica mais comumente usada para esta tarefa é a Análise de Componentes Principais (PCA). A Análise de

Componentes Principais (PCA, do inglês *Principal Component Analysis*), é uma técnica estatística multivariada criada por Hotelling (1933) e que se dedica à explicação da estrutura de variância-covariância existente em um conjunto de dados, utilizando-se combinações lineares das variáveis originais. Segundo Johnson e Wichern (2002) e Rencher (2002), seus objetivos principais são: (1) a redução de dimensionalidade e (2) a interpretação de dados.

Embora  $p$  componentes sejam necessários para se reproduzir a variabilidade total de um sistema de interesse, geralmente, a maior parte desta variabilidade pode ser representada por um pequeno número  $k$  de componentes principais. Isso quer dizer que existe quase tanta informação em  $k$  componentes principais que nas  $p$  variáveis originais. A ideia geral da ACP é, portanto, que  $k$  componentes principais podem substituir, sem perda considerável de informação, as  $p$  variáveis originais. O conjunto original de dados, consistindo de  $n$  medições das  $p$  variáveis, é reduzido para um conjunto posterior formado por  $n$  medições de  $k$  componentes principais.

De acordo com Rencher (2002), a ACP geralmente revela relacionamentos que não são previamente identificados no conjunto original, o que resulta em uma interpretação mais abrangente do fenômeno. Segundo Johnson e Wichern (2002), a análise de componentes principais serve como um passo intermediário na análise dos dados.

A Análise de Componentes Principais tem uma extensa gama de aplicações. Basicamente, sua utilidade está na redução de dimensionalidade de vetores de entradas ou de saídas em determinados equacionamentos. Nela, os componentes principais formados dependem somente da matriz de variância-covariância  $\Sigma$  ou da matriz de correlação  $\rho$  das variáveis  $X_1, X_2, \dots, X_p$  e seu desenvolvimento não requer o pressuposto de normalidade multivariada. Por outro lado, os componentes principais derivados de uma população normal multivariada conduzem a interpretações úteis em termos de elipsóides de densidade constante. Adicionalmente, inferências podem ser feitas a partir de componentes amostrais quando a população é multivariada normal.

Seja o vetor aleatório  $X^T = [X_1, X_2, \dots, X_p]$ , cuja matriz de variância-covariância  $\Sigma$  possua autovalores  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$ . Para este vetor, sejam consideradas as seguintes combinações lineares:

$$\begin{aligned}
Y_1 &= \ell_1^T \mathbf{X} = \ell_{11}X_1 + \ell_{21}X_2 + \dots + \ell_{p1}X_p \\
Y_2 &= \ell_2^T \mathbf{X} = \ell_{12}X_1 + \ell_{22}X_2 + \dots + \ell_{p2}X_p \\
&\vdots \\
Y_p &= \ell_p^T \mathbf{X} = \ell_{1p}X_1 + \ell_{2p}X_2 + \dots + \ell_{pp}X_p
\end{aligned} \tag{2.52}$$

Os componentes principais serão, portanto, todas as combinações lineares não correlacionadas  $Y_1, Y_2, \dots, Y_p$  cujas variâncias são tão grandes quanto possível.

O primeiro componente principal (PC<sub>1</sub>), segundo a definição de Johnson e Wichern (2002), é a combinação linear que melhor conseguir representar o máximo da estrutura de variância ou, de acordo com o Teorema de Cauchy-Schwarz, aquela combinação que maximizar a variância. Genericamente, o  $i$ -ésimo componente principal será a combinação linear  $\ell_i^T \mathbf{X}$  que resultar do seguinte equacionamento:

$$\text{Maximizar } \text{Var}(\ell_i^T \mathbf{X}) \tag{2.53}$$

$$\text{Sujeito a: } \ell_i^T \ell_i = 1 \tag{2.54}$$

$$\text{Cov}(\ell_i^T \mathbf{X}, \ell_k^T \mathbf{X}) = 0 \text{ para } k < i \tag{2.55}$$

Onde:  $\ell_i^T$  é o autovalor (ortonormal) da matriz de variância-covariância dos dados originais e  $\mathbf{X}$  é o vetor de variáveis aleatórias. A restrição (2.54) é necessária para gerar autovetores unitários.  $\ell_k^T$  é um segundo autovetor perpendicular ao primeiro; nessa condição, as combinações lineares formadas por esses dois vetores gerarão variáveis latentes independentes, tal como descrito pela equação (2.55).

Na maioria das vezes, não se tem conhecimento dos parâmetros populacionais de variância-covariância e correlação, respectivamente,  $\Sigma$  e  $\rho$ . Neste caso, adota-se a matriz de variância-covariância amostral  $\mathbf{S}$  no lugar de  $\Sigma$  e a matriz de correlação amostral  $\mathbf{R}$  no lugar de  $\rho$ . Assim, tem-se que:

$$\mathbf{S} = \begin{bmatrix} \frac{1}{n} \sum_{j=1}^n (x_{1j} - \bar{x}_1)^2 & \dots & \frac{1}{n} \sum_{j=1}^n (x_{1j} - \bar{x}_1)(x_{pj} - \bar{x}_p) \\ \vdots & \ddots & \vdots \\ \frac{1}{n} \sum_{j=1}^n (x_{1j} - \bar{x}_1)(x_{pj} - \bar{x}_p) & \dots & \frac{1}{n} \sum_{j=1}^n (x_{pj} - \bar{x}_p)^2 \end{bmatrix} \tag{2.56}$$

Dessa forma, os componentes principais amostrais são escritos em termos de  $\mathbf{S}$  e  $\mathbf{R}$ , tal que:

$$\sum_{i=1}^p s_{ii} = \hat{\lambda}_1 + \hat{\lambda}_2 + \dots + \hat{\lambda}_p \quad (2.57)$$

$$r_{(\hat{y}_i, x_k)} = \frac{\text{Cov}(x_k, \hat{y}_i)}{\sqrt{\text{Var}(\hat{y}_i)\text{Var}(x_k)}} = \frac{\hat{e}_{ki} \sqrt{\hat{\lambda}_i}}{\sqrt{s_{kk}}}, \quad i, k = 1, 2, \dots, p \quad (2.58)$$

Por vezes, é útil escrever as combinações lineares na forma de escores dos componentes principais. Em muitas aplicações, a matriz de variáveis padronizadas está representada pelas  $p$  colunas das características estudadas, em cada uma das suas  $n$  observações. Assim, na prática, é mais comumente empregada a matriz transposta de  $\mathbf{Z}$ .

Para se encontrar uma expressão adequada a esta realidade que represente a mesma informação que  $Y_i = e_i^T \mathbf{Z}$ ,  $i = 1, 2, \dots, p$ , utiliza-se a entidade estatística denominada de escore de componentes principais ( $PC_k$ ), que pode ser representada como:

$$PC_k = \mathbf{Z}^T \mathbf{e} = \begin{bmatrix} \left( \frac{x_{11} - \bar{x}_1}{\sqrt{s_{11}}} \right) & \left( \frac{x_{21} - \bar{x}_2}{\sqrt{s_{22}}} \right) & \dots & \left( \frac{x_{p1} - \bar{x}_p}{\sqrt{s_{pp}}} \right) \\ \left( \frac{x_{12} - \bar{x}_1}{\sqrt{s_{11}}} \right) & \left( \frac{x_{22} - \bar{x}_2}{\sqrt{s_{22}}} \right) & \dots & \left( \frac{x_{p2} - \bar{x}_p}{\sqrt{s_{pp}}} \right) \\ \vdots & \vdots & \ddots & \vdots \\ \left( \frac{x_{1n} - \bar{x}_1}{\sqrt{s_{11}}} \right) & \left( \frac{x_{2n} - \bar{x}_2}{\sqrt{s_{22}}} \right) & \dots & \left( \frac{x_{pn} - \bar{x}_p}{\sqrt{s_{pp}}} \right) \end{bmatrix} \cdot \begin{bmatrix} e_{11} & e_{12} & \dots & e_{1p} \\ e_{21} & e_{22} & \dots & e_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ e_{1p} & e_{2p} & \dots & e_{pp} \end{bmatrix} \quad (2.59)$$

Geometricamente, a nova combinação linear representa a seleção de um novo sistema de coordenadas obtidas a partir do sistema original. Enquanto os eixos originais representam as variáveis  $X_1, X_2, \dots, X_p$ , o novo eixo representará a direção de máxima variância. Componentes secundários representarão o restante da variância e sempre estarão representados em eixos ortogonais aos primeiros. Os componentes principais são não correlacionados entre si e dependentes apenas na matriz de covariância  $\mathbf{\Sigma}$  (ou na matriz de correlação  $\mathbf{\rho}$ ) das variáveis  $X_1, X_2, \dots, X_p$  e o desenvolvimento não requer que se admita a normalidade multivariada.

Para se saber quantos componentes principais (ou quantas combinações lineares) são adequados para representar a estrutura original, aplicam-se os critérios de Kaiser (JOHNSON & WICHERN, 2002). De acordo com tais exigências, quando a extração se der pela matriz de

correlação, devem ser mantidas apenas as combinações lineares associadas aos autovalores maiores do que a unidade. Além disso, para matriz de covariância  $\Sigma$  ou na matriz de correlação  $\rho$ , a variância acumulada explicada deve ser sempre superior a 80%.

A utilização do PCA como técnica de redução de dimensionalidade é antiga, porém, em constante e atual utilização. Liu *et al.* (2020), por exemplo, utilizaram o PCA para reduzir o número de dimensões relacionadas com um grande número de variáveis de decisão em problemas sintéticos de otimização multiobjetivo. No texto de Liu *et al.* (2020), fica evidenciada a necessidade de redução de dimensionalidade dos dados para obter uma representação menor das variáveis originais, mantendo-se as informações importantes e evitando-se a redundância ou o uso de informações que não agregam. Neste trabalho, especificamente, as variáveis foram primeiramente agrupadas pela técnica de cluster segundo sua (a) diversidade e (b) convergência. Na sequência, o PCA foi aplicado às variáveis desses dois grupos iniciais. Os escores de componentes resultantes alimentaram um algoritmo genético do tipo MOEA (*Multiobjective Optimization Evolutionary Algorithm*) que, por sua vez, realizou as iterações da otimização até a convergência. O que se salienta desse trabalho é o destaque dado pelos autores à diversidade das soluções e à facilidade (maior ou menor) de convergência, pontos importantes e consonantes com a proposta da presente tese.

Khanal *et al.* (2017) utilizaram o PCA para extrair as informações mais relevantes de um grande conjunto de variáveis para a previsão de produção de líquidos ricos em gás de xisto condensado. Além de reduzir a dimensionalidade do conjunto de entrada, o PCA atuou principalmente para a remoção das correlações entre as variáveis preditoras, substituindo-as por variáveis latentes de previsão independentes.

Processos semelhantes de “mineração de dados”, tal como o caso de Khanal *et al.* (2017), tornaram-se importantes nos dias de hoje, em que grandes quantidades de dados precisam de ser tratados para se transformarem em informações auxiliares no processo de tomada de decisões. Dai *et al.* (2016) explicam que conjunto de dados de alta dimensão é encontrado em muitas situações, no entanto, é comum que apenas alguns dos preditores sejam relevantes para a resposta. As técnicas de obtenção de dados proporcionaram um aumento significativo nos conjuntos de dados disponíveis. Em algumas áreas, a exemplo da medicina e da energia, os conjuntos de dados obtidos são extensos – tanto em número de variáveis quanto em número de observações (ou linhas) - gerando uma dificuldade de processamento e de tempo requerido para este. No trabalho apresentado por Karimi e Farrokhnia (2014), é explicitada a dificuldade de trabalhar com elevados dados dimensionais em um conjunto de dados de genes exigindo uma extração de características significativas de alta dimensão de dados. Trabalhos

como esses geram um quantitativo muito grande de dados. A redução da dimensionalidade dos dados visa a otimizar a análise, projetando informações mais validadas e confiáveis.

Em outro estudo, de Mei *et al.* (2007), também foi analisada a dificuldade de entender altas interações em heterogeneidade genética em doenças humanas de alta complexidade. Mais uma vez, destaca-se que o PCA é capaz de capturar “heterogeneidades” (ou “diversidade”) dentro dos conjuntos formados por muitas variáveis “homogêneas” (ou correlatas), aspecto esse que também é muito importante para o desenvolvimento desta tese: identificar as métricas de previsão cuja minimização conduza a vetores de pesos diversificados, o que caracteriza a busca por soluções “homogêneas” dentro de grupos “heterogêneos”.

Martins *et al.* (2020) também apresentam um trabalho no qual destacam a importância da redução de dimensionalidade quando o processo de tomada de decisão se baseia em múltiplos critérios, que representam, na verdade, grupos de subcritérios, com diferentes pesos. Os autores apresentam uma estratégia na qual “fatores” (chamados de critérios principais) englobam “variáveis”, as quais são incorporadas ao processo de tomada de decisão por meio de diferentes pesos atribuíveis. Observa-se que a estratégia proposta pelos autores neste trabalho é muito semelhante à composição matemática observada na análise fatorial, ou seja, grupos de variáveis (cada qual com seu peso – no caso, diferentes *loadings*) agrupadas em “fatores” independentes e com diferentes níveis de explicação (ou diferentes autovalores).

Na literatura, são diversas as técnicas que são utilizadas para realizar a redução da dimensionalidade, entre as quais podem ser citadas:

- “*Principal Component Analysis (PCA)*”;
- “*Artificial Neural Networks*”;
- “*Kernel PCA*” e “*Fuzzy Robust Principal Component Analysis (FRPCA)*”;
- “*Multi-Criteria Decision Analysis e Machine Learning*”;
- “*Stacked Autoencoder (SAE)*” e “*Stacked Restricted Boltzmann Machine (SRBM)*”;
- “*Sliced Inverse Regression (SIR)*”;
- “*Principal Hessian Directions (PHD)*”;
- “*Multifactor Dimensionality Reduction-Phenomics (MDR-Phenomics)*”;
- “*MDR pedigree disequilibrium test (PDT)*”;
- “*MDT com considerações de PCs (MDR-PDT)*” e “*Stratified Phenotype (SP)*”;
- “*Segmented Principal Component Analysis and Regression (SPCAR)*”;
- “*Kohonen Self-organizing Map (KSoMAP)*”;
- “*Linear Surrogate Variable (LSV)*” e “*ALASSO*”;

- “*Multi-objective Evolutionary Algorithms (MOEAs)*”;
- “*Settling Large-scale Multi-objective Problems (MOEA/DVA)*”;
- “*Krzanowski’s KP Method*”;
- “*Jolliffe*”;
- “*McCabes Method*”;
- “*Graphical Gaussian Models (DF)*”;
- “*Multiple Correlation*”; e
- “*Algorithm H e Full Temporal Algorithm (HT)*”.

Zhong e Enke (2017) ponderam, em seu trabalho sobre previsão diária de ativos em mercados de ações, que a redução da dimensionalidade pode ser realizada de duas maneiras diferentes: ou selecionando as variáveis mais relevantes do conjunto de dados original ou gerando um grupo menor de novas variáveis, cada uma sendo uma certa combinação das variáveis de entrada. Neste trabalho, os autores utilizaram o PCA para transformar sessenta características econômicas e financeiras das ações da S&P 500 em um número menor de variáveis de entrada utilizadas em redes neurais artificiais.

É interessante perceber a diversidade de métodos utilizados para se obter a redução da dimensionalidade. O Quadro 2-1 apresenta um resumo dos principais autores desta área e os métodos utilizados para obtenção da redução da dimensionalidade. Observa-se uma preponderância do PCA como estratégia de redução da dimensionalidade em diversas pesquisas.

**Quadro Tese 2-1 - Autores e técnicas utilizadas como redução de dimensionalidade dos dados**

<b>Autores</b>	<b>Técnicas Utilizadas</b>
Zhong e Enke (2017)	PCA, FRPCA e KPCA
Li e Shu (2009)	KPCA
Martins et al. (2020)	MCDA e ML
Han e Ge (2020)	PCA, SAE e SRBM
Fang e Yu (2020)	SIR e PHD
Mei et al. (2007)	MDR-Phenomics, PDT, MDR-PDT e SP
Karimi e Farrokhnia (2014)	PCA, SPCAR e KSoMAP
Dai et al. (2016)	SIR, ALASSO e LSV
Liu et al. (2020)	PCA-MOEA
Cumming e Woof (2007)	Jolliffe's Method, PCA, McCabes Method, H, HT e DF

A crença equivocada de que o PCA é robusto a todo tipo de variável confere a esta técnica estatística um “status” de maleabilidade que ela, na verdade, não apresenta. Em outras palavras, mesmo sendo um algoritmo baseado em otimização, a decomposição em valores singulares (ou PCA) pode ter sua performance substancialmente melhorada se variáveis “redundantes” forem eliminadas do conjunto, de modo prévio, posterior ou recursivo, o que envolve, geralmente, a execução do PCA em sequências sucessivas. Nesse sentido, Jolliffe (1972) discutiu uma série de métodos para selecionar um subconjunto de “ $m$ ” variáveis capazes de preservar a maior parte da variação de “ $x$ ”. A maior parte dos métodos comparados que tiveram um bom desempenho consistia na aplicação de PCA. Jolliffe (1972, 1973) apresentou, basicamente, três tipos de métodos distintos de redução de dimensionalidade e de redundância baseados em PCA que conduzem a resultados diferentes entre si. Esses métodos são descritos a seguir.

**1º Método** - Associar uma variável com cada um dos últimos  $m_1^*(= p - m_1)$  PC's e deletar essa variável  $m_1^*$ . Isso pode ser feito apenas uma vez ou iterativamente rodando um segundo PCA e deletando-se uma outra variável. O procedimento é repetido até que se tenha atingido uma quantidade de variáveis excluídas compatível com o de PCs que apresentaram autovalores maiores que 1 no primeiro PCA.

**2º Método** - Associar um conjunto de  $m^*$  variáveis em bloco com os últimos  $m^*$  PC's para excluir a variável com maior coeficiente (no caso, o coeficiente é elemento do autovetor associado ao autovalor de referência). Nessa situação, a indicação é que, do primeiro PCA, sejam retirados, a partir do último PC, as  $m^*$  variáveis. Em cada  $PC^n$ , será retirada uma das  $m^*$  variáveis, restando um número de  $m^*$  variáveis de acordo, também, com o tamanho dos autovalores maiores que 1 detectados quando da realização do primeiro e, neste caso, único PCA.

**3º Método** - Esse método associa uma variável  $m^*$ , com o coeficiente mais alto, em valor absoluto, a cada um dos primeiros PC's. Essas variáveis são mantidas, e as restantes  $m^* = p - m$  são excluídas (JOLLIFFE, 1972; JOLLIFFE, 1973; JOLLIFFE, 2002). Kaiser (1960) enfatiza que os PC's que obtiverem autovalores maiores que 1 devem ser retidos. Jolliffe

(1972), embasada em estudos de simulação, sugere o valor de 0,7 para ser usado na retenção de PC's.

Nesta tese, estes três métodos de Jolliffe serão testados para se eliminar as métricas de desempenho e de precisão de previsão do conjunto original de métricas que apresentarem níveis altos de redundância. Os dados para aplicação destes testes virão da matriz *Payoff*, cujas linhas representam os valores da otimização individual dos polinômios canônicos de misturas para cada uma das quatorze métricas de desempenho e de precisão de previsão. A ideia é que permaneçam na análise apenas as métricas com pontos de ótimo distintos.

## 2.9 MÉTODO FA-NBI

O método FA-NBI (do inglês *Factorial Analysis Normal Boundary Intersection*), proposto por Bacci *et al.* (2019), é um método de otimização de combinações de métodos de previsão que também utiliza arranjos de misturas e técnicas estatísticas multivariadas para a geração de fronteiras equiespaçadas de Pareto para escores rotacionados de fator de métricas de performance de previsão. Como a presente tese aspira a desenvolver um método semelhante ao FA-NBI, mais aperfeiçoado em diversos aspectos, este método será utilizado para comparação dos resultados. O FA-NBI foi apresentado originalmente com a seguinte estrutura:

- **1º Passo** – coletar as observações da série  $Y_t$
- **2º Passo** – Nesta etapa, são selecionados os  $z$  métodos de previsão individuais ( $F_1, F_2, \dots, F_z$ ) que serão testados. Com esses métodos de previsão, são produzidas as séries de resíduos de cada método ( $R_1, R_2, \dots, R_z$ ).
- **3º Passo** – Através da abordagem de experimentos de mistura (MDE), cria-se a matriz de pesos para as combinações. Um arranjo de misturas do tipo Simplex Lattice para a combinação de três métodos de previsão é gerado para o estudo de caso desta tese, perfazendo o total de 25 combinações diferentes dos produtos “pesos-resíduos” ( $W_{11}R_{11}; W_{12}R_{11}; \dots, W_{1p}R_{11}$ ).
- **4º Passo** – Com as séries formadas por resíduos combinados, obtidas no passo 3, são calculadas as  $j$  métricas de performance e de precisão.
- **5º Passo** – A análise PCFA é aplicada às  $j$  métricas de performance e de precisão para reduzir a dimensionalidade do problema, obtendo-se as séries de fatores (*Factor Scores Series* -  $FS_1; FS_2 \dots FS_n$ ).

- **6º Passo** – Empregando-se o algoritmo OLS, modelam-se as superfícies de respostas para os escores de fatores  $FS_1; FS_2; \dots; FS_n$  representativos das séries de métricas originais. No software MINITAB®, a modelagem utiliza uma parametrização por *mixture components only*; com pesos na forma de pseudocomponentes e com método de ajuste do modelo do tipo *mixture regression*. Nesta etapa, são geradas as respectivas equações polinomiais  $\widehat{FS}_1(\mathbf{W}), \widehat{FS}_2(\mathbf{W}), \widehat{FS}_3(\mathbf{W}), \dots, \widehat{FS}_q(\mathbf{W})$ .
- **7º Passo** – A aplicação do NBI é realizada para resolver um problema de otimização de múltiplos objetivos, no qual as funções objetivo  $\widehat{FS}_1(\mathbf{W}), \widehat{FS}_2(\mathbf{W}), \dots, \widehat{FS}_q(\mathbf{W})$  são simultaneamente otimizadas. Essa otimização conjunta conduz à otimização das  $j$  métricas originais representadas por elas. Após sucessivas otimizações, o NBI forma a Fronteira de Pareto. Para escolher a melhor solução, aplicam-se os critérios  $GPE_{Total}$  e a  $S_{Total}$  para se verificar qual conjunto de pesos apresenta a melhor razão  $S_{Total} / GPE_{Total}$ .
- **8º Passo** – Com base no conjunto de pesos escolhidos, compara-se o desempenho do Método FA-NBI com todos os métodos individuais de previsão, utilizando todas as métricas do trabalho.

A Figura 2.3 apresenta uma visão esquemática desse método.

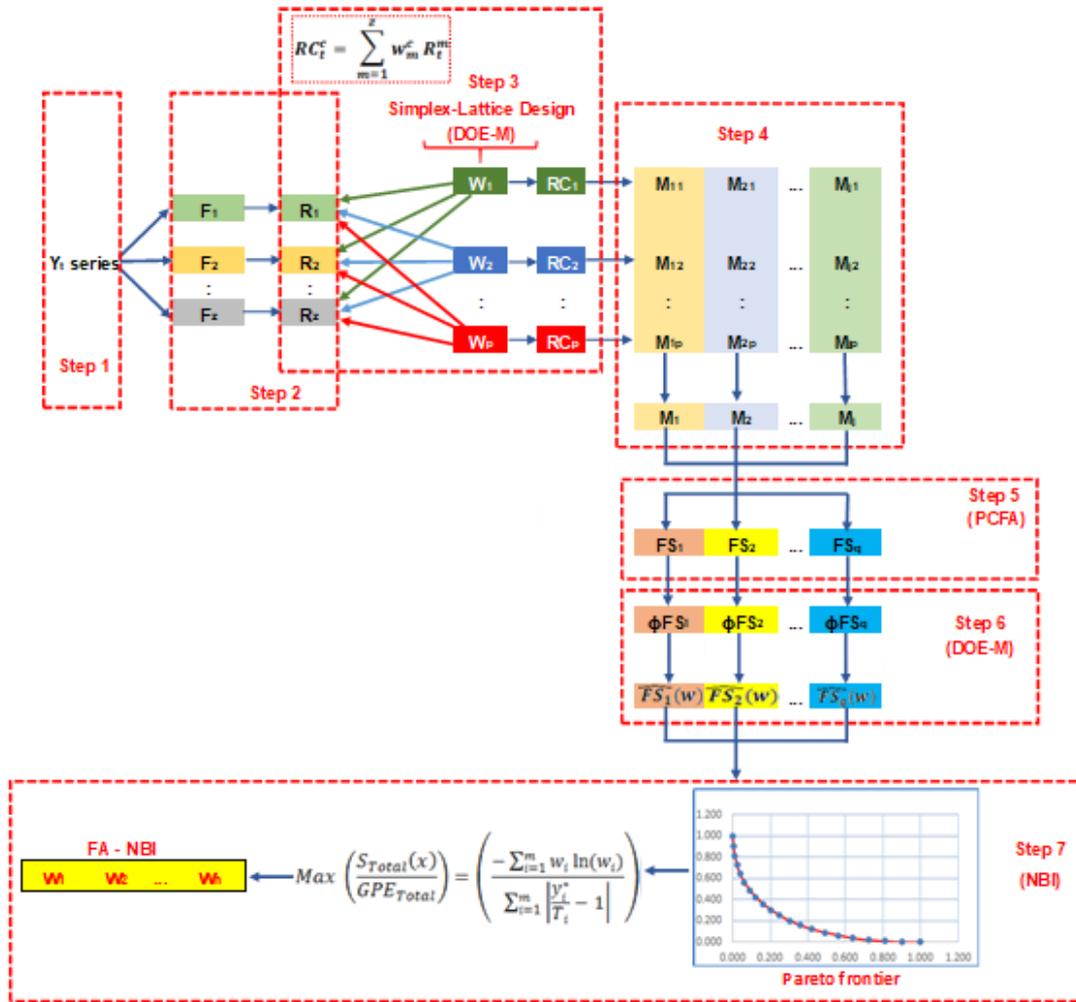


Figura 2-3 - Método FA-NBI

Fonte: Bacci *et al.* (2019)

Apesar de ser um método reconhecidamente robusto para otimização multiobjetivo de combinações de métodos de previsão, o FA-NBI apresenta uma série de lacunas que podem ser preenchidas, aperfeiçoando sua estrutura e a tornando mais flexível e enxuta. A principal modificação reside na utilização de um conjunto de métricas de desempenho não redundantes, escolhidas de acordo com os resultados de otimizações individuais dos polinômios canônicos de misturas. A incorporação de um grupo de restrições elípticas ao método NBI também é uma considerável contribuição para a melhoria da performance de previsão. Essas particularidades serão detalhadas no próximo capítulo.

### 3 MÉTODO PAYOFF-JOLLIFFE FATORIAL

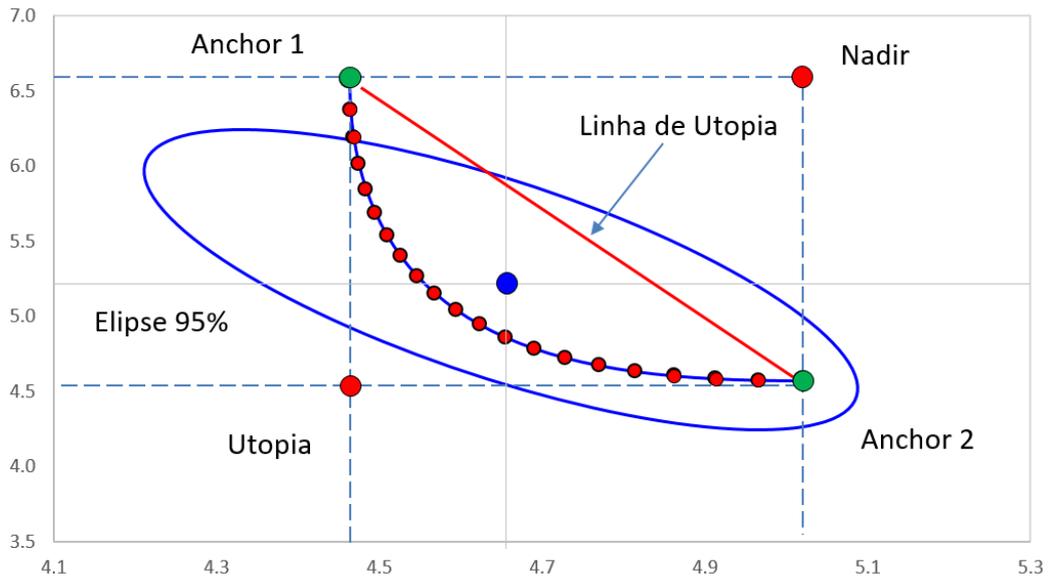
#### 3.1 MÉTODO NBI-FA COM RESTRIÇÕES ELÍPTICAS E FILTROS DE JOLLIFFE

O método NBI foi apresentado no item 2.7 do capítulo 2 desta tese. Sua concepção original pressupunha a utilização de funções objetivo determinísticas e, como tal, não contemplavam as particularidades provenientes dos modelos de superfície de resposta em seu desenvolvimento. Foi apenas no trabalho de Costa *et al.* (2016) que os modelos de superfície de resposta foram primeiramente incorporados ao método NBI na forma de funções objetivo e restrições. A este primeiro modelo adaptado, seguiram-se os trabalhos de Naves *et al.* (2017) e Gaudêncio (2019a, 2019b). Estes trabalhos também foram os precursores na utilização de escores de superfícies de respostas de componentes principais (COSTA *et al.*, 2016) ou de escores rotacionados de fator (NAVES *et al.*, 2017; GAUDÊNCIO *et al.*, 2019a) como funções objetivo ou restrições no método NBI. A ideia comum entre eles era de que, se diferentes características medidas forem correlacionadas, sua dimensionalidade poderá ser reduzida pela substituição de variáveis originais por variáveis latentes, quer sejam componentes (PCA), quer sejam fatores (FA). Entretanto, a translação do método NBI original para sua versão aplicada a superfícies de resposta experimentou muitos percalços e dificuldades, os quais foram contornados ao longo dos últimos anos, tal como pode ser verificado nos trabalhos de Maia (2020) e Maia *et al.* (2021). Os resultados desses estudos mostram uma melhoria de desempenho do método NBI, mas também caracteriza restrições de funcionamento ou condições de contorno para a melhoria de performance do método.

A principal discrepância entre as condições modeladoras do NBI e o espaço de solução de problemas multivariados está relacionada aos “pontos de ancoragem” e à região elíptica que circunscreve os dados das superfícies de resposta. Dados experimentais correlacionados originam regiões de confiança multivariadas e, em função dessa relação de dependência, dificilmente as superfícies de resposta ajustadas a esses dados poderão admitir soluções independentes umas das outras, o que, de acordo com a fundamentação teórica do método NBI, contrapõe-se à fixação de pontos de ancoragem independentes.

Portanto, quando o método NBI é empregado com funções objetivo correlacionadas (positiva ou negativamente) e obtidas experimentalmente por arranjos de superfície de resposta, os pontos de ancoragem recorrentemente saem da elipse de confiança dos dados experimentais, o que envia a linha de utopia e a formação da fronteira e conduz a soluções distantes da realidade. Na Figura 3.1, adaptada de Maia (2020), vê-se claramente que a elipse de confiança – no caso bidimensional – não contém todos as soluções ótimas viáveis descritas pela Fronteira

de Pareto. Logo, um certo desajuste ocorrerá entre as soluções matemáticas e aquelas reais. Esta insuficiência do método NBI pode ser superada se às suas restrições forem acrescentadas funções que delimitem o conjunto “imagem” multivariado. De acordo com a teoria estatística multivariada, tal restrição deve ser escrita na forma de uma equação de elipse (no caso bivariado) ou de uma equação de hiperelipse para problemas com dimensões maiores. Estas formulações serão descritas posteriormente.



**Figura 3-1 - 95% Elipse de Confiança versus Fronteira de Pareto (NBI).**  
**Fonte: Maia (2020).**

A demonstração a seguir consta nos trabalhos de Maia *et al.* (2021), Almeida *et al.* (2020), Maia (2020) e Gaudêncio *et al.* (2019b). Considere, pois, uma equação de elipse escrita em termos de um vetor de variáveis normais e aleatórias  $\mathbf{x}$ , com vetor de média  $\boldsymbol{\mu}$  e matriz de variância-covariância  $\tilde{\boldsymbol{\Sigma}}$ . Se  $c^2$  é uma variável aleatória do tipo Qui-Quadrado, definida para um nível de confiança  $(1-\alpha)$ , tal que  $c^2 = \chi^2_{(p, \alpha)}$ , então, a equação da elipse de confiança será dada por:

$$(\mathbf{x} - \boldsymbol{\mu})^T \tilde{\boldsymbol{\Sigma}}^{-1} (\mathbf{x} - \boldsymbol{\mu}) = c^2 \quad (3.1)$$

Por decomposição espectral,  $\tilde{\boldsymbol{\Sigma}}^{-1}$  pode ser substituída por  $\mathbf{P} \boldsymbol{\Lambda}^{-1} \mathbf{P}^T$ , onde  $\mathbf{P}$  é a matriz de autovetores, e  $\boldsymbol{\Lambda}$  é a matriz diagonal de autovalores. Dado que  $\boldsymbol{\Lambda}^{-1} = \boldsymbol{\Lambda}^{-1/2} \boldsymbol{\Lambda}^{-1/2}$ , então:

$$\sqrt{[\mathbf{P}^T(\mathbf{x} - \boldsymbol{\mu})]^T \boldsymbol{\Lambda}^{-1/2} \boldsymbol{\Lambda}^{-1/2} [\mathbf{P}^T(\mathbf{x} - \boldsymbol{\mu})]} = [\mathbf{P}^T(\mathbf{x} - \boldsymbol{\mu})]^T \boldsymbol{\Lambda}^{-1/2} = \sqrt{\chi_{(p,\alpha)}^2} \quad (3.2)$$

Como  $\mathbf{P}$  é ortogonal, então  $\mathbf{P}^T = \mathbf{P}^{-1}$ , o que simplifica a expressão anterior, tal que:

$$\mathbf{x} = \boldsymbol{\mu} + \mathbf{P} \boldsymbol{\Lambda}^{1/2} \sqrt{\chi_{(p,\alpha)}^2} \quad (3.3)$$

Para se representar graficamente a elipse de confiança, a Equação (3.3) pode ser reescrita, em uma forma matricial-vetorial, de forma que:

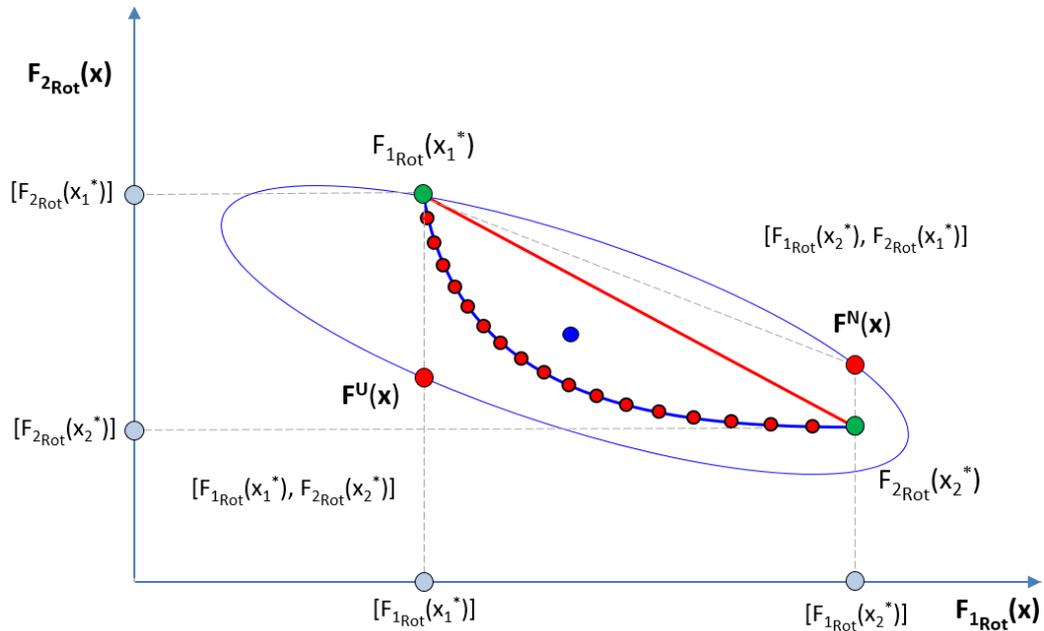
$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} + \sqrt{\chi_{(p,\alpha)}^2} \times \underbrace{\begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix}}_{\mathbf{P}} \underbrace{\begin{bmatrix} \sqrt{\lambda_1} & 0 \\ 0 & \sqrt{\lambda_2} \end{bmatrix}}_{\boldsymbol{\Lambda}^{1/2}} \begin{bmatrix} \cos \alpha \\ \sin \alpha \end{bmatrix} \quad (3.4)$$

No sistema de equações (3.4), o ângulo  $\alpha$  representa a rotação de um ponto  $\mathbf{P}(x, y)$  qualquer na elipse, no domínio  $0 \leq \alpha \leq 2\pi$ , servindo para construir a elipse para todo o espaço de solução.

Pois bem, se a elipse de confiança for considerada como uma restrição adicional àquelas já pré-estabelecidas pelo método NBI, então a nova região de solução cairá dentro da elipse de confiança dos dados experimentais. Desse modo, pode-se ressaltar que a formulação do método NBI fica modificada por dois conjuntos de restrições: uma para o conjunto “Domínio” e outra para o conjunto “Imagem”. Ou seja, restrições de espaço experimental (ou de esfericidade) tal como  $g(\mathbf{x}) = \mathbf{x}^T \mathbf{x} - \rho^2$ , controlam os possíveis valores do vetor de solução  $\mathbf{x}$  (ou o “Domínio”), enquanto restrições elípticas controlam os valores das respostas originais  $\mathbf{y}$  (ou o conjunto “Imagem”). Nessa região, tantos os pontos de ancoragem como todas as demais soluções viáveis representadas pela fronteira tenderão a cair dentro da região de confiança multivariada, preservando a estrutura de variância-covariância dos dados originais do problema. A simples utilização de uma otimização individual – requerida pelo método para firmar os pontos de ancoragem – pode conduzir a soluções fora dessa região de confiança, porque a resposta otimizada individualmente tende a não respeitar as restrições impostas pelas demais funções objetivo, o que provoca a instabilidade dos pontos de fronteira ou a sua incompatibilidade com as soluções reais.

A Figura 3-2 ilustra a formação da Fronteira de Pareto e da elipse de confiança para o caso bidimensional, no qual dois fatores são capazes de representar adequadamente conjuntos

de funções objetivo originais. A fronteira de Pareto que essas funções “multivariadas” descrevem deve estar contida na região de confiança dos dados originais do problema, representada pela elipse de confiança. As superfícies de resposta para os escores de fator rotacionados (FRot) ainda poderão integrar o método na forma simples ou considerando-se um valor de alvo, previamente estabelecido.



**Figura 3-2 - Fronteira de Pareto corrigida pela elipse de confiança de 95%.  
Fonte: Maia (2020).**

Como já demonstrado em Gaudêncio *et al.* (2019b), funções objetivo multivariadas podem ser agrupadas por correlação e, conseqüentemente, representadas por novas funções escritas em termos de escores rotacionados de fatores ( $\mathbf{F}_{Rot}$ ). Além da redução garantida de dimensionalidade do problema multiobjetivo, a utilização de escores rotacionados de fator em substituição aos conjuntos originais de dados evita inversões nos sentidos de correlação de conjuntos de funções cujos sentidos de otimização sejam contrários, apresentando soluções que sejam as mais próximas possíveis da realidade do processo modelado.

O “Escore de Fator” é uma variável adimensional calculada a partir da matriz de dados original padronizada ( $\mathbf{Z}$ ) e de uma matriz de carregamentos (ou *loadings*) ( $\mathbf{L}$ ), tal que:

$$\tilde{\mathbf{F}} = \mathbf{Z}[\mathbf{L}(\mathbf{L}^T\mathbf{L})^{-1}] \quad (3.5)$$

A matriz de carregamentos possui a forma  $\mathbf{L} = [\sqrt{\lambda_1} \mathbf{e}_1 \mid \sqrt{\lambda_2} \mathbf{e}_2 \mid \cdots \mid \sqrt{\lambda_p} \mathbf{e}_p]$ , que basicamente é a matriz  $\mathbf{P}$  de autovetores de  $\Sigma$ , com cada coluna de autovetores, sendo pré-multiplicada pela raiz quadrada do seu respectivo autovalor. É fácil verificar que  $\mathbf{L}\mathbf{L}^T$  representa a decomposição espectral de  $\Sigma$ .

“Fator” é, portanto, uma variável não observável, subjacente, capaz de representar (e substituir) grupos de variáveis originais que se agrupem por suas correlações, de forma que os grupos formados apresentem mínima variância interna e máxima discriminação em relação aos outros grupos. Como o fator é uma variável latente, ele só pode ser extraído por meio de uma decomposição espectral denominada “Análise Fatorial”. Em suma, esta técnica estatística multivariada baseia-se no fato de que, se um vetor aleatório  $\mathbf{X}$ , com  $p$  componentes, apresentar vetor de média  $\boldsymbol{\mu}$  e matriz de variância-covariância  $\Sigma$ , então:

$$\underset{(p \times 1)}{\mathbf{X}} - \underset{(p \times 1)}{\boldsymbol{\mu}} = \underset{(p \times m)}{\mathbf{L}} \underset{(m \times 1)}{\mathbf{F}} + \underset{(p \times 1)}{\boldsymbol{\varepsilon}} \quad (3.6)$$

Quando há algum nível de correlação entre as variáveis originais, a representação de  $\Sigma$  pode ser aproximada por uma combinação linear de carregamentos com um número menor de fatores do que de variáveis; neste caso, a informação descartada é armazenada em uma matriz de erro ( $\Psi$ ), tal que:

$$\Sigma = E[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^T] = E[(\mathbf{L}\mathbf{F} + \boldsymbol{\varepsilon})(\mathbf{L}\mathbf{F} + \boldsymbol{\varepsilon})^T] = E[(\mathbf{L}\mathbf{F} + \boldsymbol{\varepsilon})(\mathbf{L}^T\mathbf{F}^T + \boldsymbol{\varepsilon}^T)] = \mathbf{L}\mathbf{L}^T + \Psi \quad (3.7)$$

Onde  $\Psi$  é matriz diagonal de variâncias específicas  $\psi_i$ , tal que  $\psi_i = \sigma_i^2 - h_i^2$ . O termo  $h_i^2$  é conhecido como “comunalidade” ou quantidade de variância explicada.

A análise fatorial se assemelha muito à análise de componentes principais (PCA), uma vez que ambas objetivam a explicação da informação contida em  $\Sigma$ . A diferença básica é que, em PCA, o componente principal é escrito como uma combinação linear das variáveis originais, enquanto na análise fatorial, a variável original é que deve ser escrita como uma combinação linear dos fatores. Além dessa diferença, a análise fatorial permite também que os autovetores sejam rotacionados, o que melhora sensivelmente a formação e a interpretação dos grupos (ou fatores). A rotação é uma transformação ortogonal, na qual os *loadings* são multiplicados por uma matriz de rotação ortogonal, tal que  $\mathbf{L}^* = \mathbf{L}\mathbf{T}$ . Este tipo de transformação não altera as comunalidades ou as variâncias específicas. Uma escolha comum para rotação no espaço de  $\mathbf{R}^2$  e sentido anti-horário é  $\mathbf{T}_{(2 \times 2)}$ , tal que:

$$\mathbf{T} = \begin{bmatrix} \cos\theta & -\text{sen}\theta \\ \text{sen}\theta & \cos\theta \end{bmatrix} \quad (3.8)$$

Dessa forma, o escore rotacionado de fator é uma combinação de variáveis originais padronizadas ( $\mathbf{Z}$ ), acompanhada por uma matriz de carregamentos ( $\mathbf{L}$ ) e uma matriz de rotação ( $\mathbf{T}$ ), de forma a criar uma nova variável adimensional e representativa de um grupo de variáveis originais calculado como:

$$\tilde{\mathbf{F}}_R = \mathbf{Z} \left\{ (\mathbf{L}\mathbf{T})[(\mathbf{L}\mathbf{T})^T(\mathbf{L}\mathbf{T})]^{-1} \right\} = \mathbf{Z} \left\{ (\tilde{\mathbf{L}})[(\tilde{\mathbf{L}})^T(\tilde{\mathbf{L}})]^{-1} \right\} \quad (3.9)$$

Suponha que as respostas originais medidas ou observadas por meio de um DOE (fatoriais completos, arranjos de superfície de resposta ou misturas) possam ser substituídas pelos escores de fatores dessas mesmas respostas (rotacionados ou não). Após modelagem por mínimos quadrados ordinários (**OLS**) – quando os resíduos forem homoscedásticos - ou por modelos lineares generalizados (**GLM**) – quando os resíduos forem heteroscedásticos - superfícies de respostas serão obtidas de acordo com a ordem do arranjo experimental inicial (cada tipo de arranjo dá origem a um tipo específico de função objetivo não linear, dependendo das interações, efeitos quadráticos ou de ordem superior que são capazes de estimar). Essas superfícies serão as novas funções objetivo representativas de grupos de funções originais parcialmente redundantes ou interdependentes. Suponha, também, que a rotação ou o método de extração dos fatores permita que todas as funções objetivo originais sejam representadas apenas por dois fatores. Então, um problema multiobjetivo consideravelmente grande poderá ter sua dimensão drasticamente reduzida, sem perda significativa de informação e comporá, adicionando-se restrições elípticas, uma versão multivariada para o método NBI em sua versão biobjetivo. Caso se necessite de um terceiro grupo de funções (ou um 3º fator) e ainda se deseje preservar o modelo biobjetivo, este novo fator pode ser considerado como uma restrição adicional.

### 3.2 OTIMIZAÇÃO COM FUNÇÕES OBJETIVO CORRELACIONADAS

A presença de funções objetivo correlacionadas em problemas de otimização é muito mais comum do que se pode imaginar. A presença da correlação entre funções objetivo tem causas diversas: desde a propagação da incerteza de medições para grandezas medidas com um mesmo instrumento de medição até a amplitude dos níveis utilizados em arranjos experimentais de DOE. Entretanto, de que forma a correlação afeta os resultados da otimização e até que ponto é



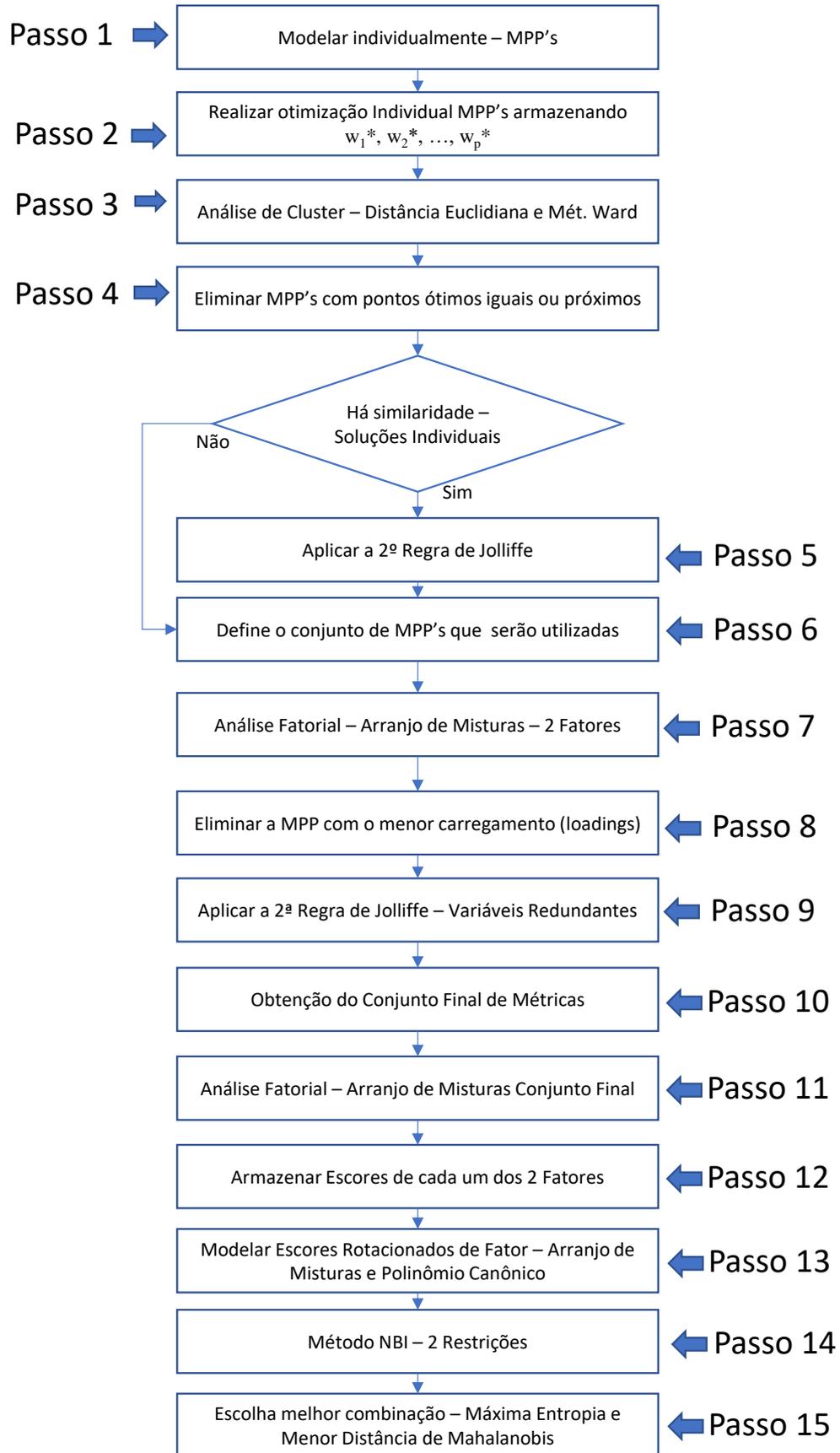
sendo diferentes, existirá uma linha de Utopia de comprimento não nulo unindo estes ótimos individuais, de onde surgirá uma Fronteira de Pareto. Em outras palavras, ainda que as funções sejam correlacionadas, nem sempre as soluções que são boas para uma função objetivo também serão adequadas para as outras.

Logo, é razoável se admitir que algum tipo de filtro pode ser aplicado às múltiplas funções objetivo correlacionadas para identificar se elas devem ou não fazer parte de um problema de otimização multiobjetivo. Como superfícies de resposta (usadas como funções objetivo) podem ter seus dados de origens processados por técnicas multivariadas como PCA (Análise de Componentes Principais) ou por PCFA (Análise Fatorial), então, tais técnicas serão úteis na filtragem e na seleção das funções objetivo. Da mesma forma, se uma matriz de soluções individuais puder ser construída, uma análise de similaridade ou de distância entre as soluções individuais também pode servir como critério de seleção. Assim, soluções idênticas ou muito próximas indicam que apenas uma das funções deve ser selecionada enquanto as demais serão descartadas. Por outro lado, se as soluções são muito diferentes, portanto as funções objetivo devem compor o problema multiobjetivo.

Dentre os possíveis critérios de seleção de respostas correlacionadas, destacam-se os três critérios de Jolliffe já apropriadamente apresentados em seções anteriores. Assim, sugere-se o seguinte método para abordar o problema das múltiplas funções objetivo correlacionadas.

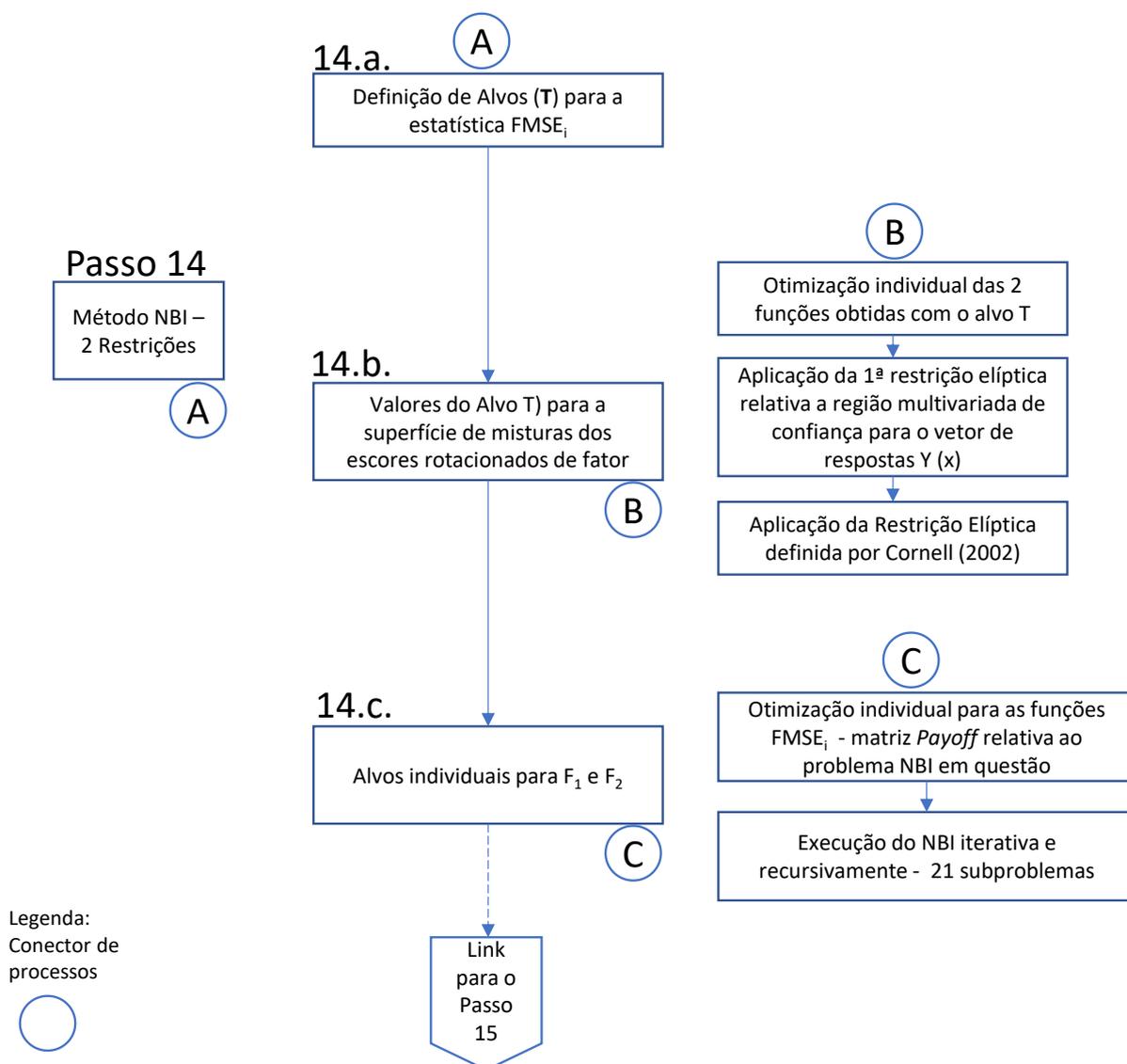
### 3.3 MÉTODO JOLLIFFE FATORIAL

O método proposto nesta tese utiliza a combinação de métodos de previsão, introduzindo a redução da dimensionalidade das métricas de performance de previsão (**MPP's**), para se alcançar um resultado melhor do que aquele obtido pela utilização dos métodos individuais de previsão. Sua estrutura segue o seguinte fluxograma:



**Figura 3-4 - Fluxograma Método Jolliffe Fatorial.**  
**Fonte: Próprio autor**

A parte mais complexa do método se encontra na aplicação do “Passo 14”. Na descrição de cada passo apresentada na seção 4.2, o “Passo 14” foi dividido em 3 (14<sup>a</sup>, 14b e 14c) para melhor entendimento da complexa sequência de cálculos realizados. Abaixo, segue um fluxograma explodido somente do “Passo 14”:



**Figura 3-5 - Fluxograma Explodido “Passo 14” do Método proposto**  
**Fonte: Próprio autor**

O Método Jolliffe Fatorial com restrições elípticas segue os seguintes passos:

- 1) Modelar individualmente cada uma das MPP’s de acordo com o arranjo de misturas escolhido e o polinômio canônico de interesse, por exemplo, um arranjo cúbico completo;

- 2) Realizar a otimização individual de cada MPP, utilizando o grupo de restrições elípticas, além de armazenar as soluções ótimas (pesos)  $w_1^*$ ,  $w_2^*$ , ...,  $w_p^*$  em uma Matriz *Payoff* com cada linha relacionada a uma métrica;
- 3) Aplicar uma Análise Hierárquica de Cluster para estas observações, usando a distância euclideana e o método de Ward para avaliar a similaridade entre as soluções ótimas;
- 4) Eliminar as MPP's com mesmo ponto de ótimo da análise;
- 5) Se a similaridade entre as soluções individuais dentro de um determinado cluster for alta, aplicar a 2ª regra de Jolliffe para escolher qual métrica representará o cluster;
- 6) Definir o conjunto final de métricas MPP;
- 7) Aplicar a Análise Fatorial com rotação do tipo "Varimax" aos dados do arranjo de misturas para as métricas escolhidas, considerando apenas dois fatores;
- 8) Eliminar a MPP com o menor carregamento (*loadings*) dentro de cada um dos dois fatores;
- 9) Aplicar a 2ª regra de Jolliffe para eliminar variáveis redundantes dentro do fator;
- 10) Obter o conjunto final de métricas;
- 11) Aplicar a Análise Fatorial com rotação do tipo "Varimax" aos dados do arranjo de misturas para o conjunto final de métricas, representando a estrutura de variância-covariância do conjunto em dois fatores;
- 12) Armazenar os escores de cada um dos fatores;
- 13) Modelar os escores rotacionados de fator segundo o arranjo de misturas e o polinômio canônico de interesse;
- 14) Aplicar o método NBI com restrições elípticas para formar a Fronteira de Pareto;
- 15) Escolher a melhor combinação de pesos representada na fronteira de acordo com a máxima entropia  $S(w)$  da combinação e a menor distância de Mahalanobis da solução em relação às soluções individuais.

As principais contribuições desta tese estão associadas a quatro pontos principais:

- (a) Obter soluções ótimas individuais diferentes dos pontos de vértice da região Simplex que respeitem o elipsóide de confiança de 95% para todas as métricas. Este ponto obriga que as soluções ótimas sejam obtidas dentro da região Simplex e que os valores das MPP's obedeam à região multivariada de confiança;
- (b) Os valores ótimos individuais ( $\varphi$ ) formam uma Matriz *Payoff* (14 x 14) cujas linhas representam todos os valores de uma determinada MPP<sub>*i*</sub>. O valor "utópico" de uma

MPP é o resultado de sua otimização individual; substituindo-se os ótimos das MPP<sub>j</sub> na função objetivo MPP<sub>i</sub>, obtêm-se os demais valores da linha; o maior valor da MPP<sub>i</sub> na linha será o valor “Nadir”;

- (c) Baseando-se nesta Matriz *Payoff*, são escolhidas as soluções com pontos de ótimo  $\mathbf{w}^*$  e valores ótimos  $f(\mathbf{w}^*)$  mais dissimilares. Esta escolha baseia-se em uma Análise Hierárquica de Cluster;
- (d) Às métricas escolhidas nos pontos (b) e (c) aplica-se a análise fatorial, armazenando-se os escores rotacionados. Depois de modelados individualmente, seus polinômios canônicos  $\hat{f}_i(\mathbf{x})$  são submetidos às restrições elípticas do item (a) para se obter o valor alvo T e, assim, formar a estatística FMSE<sub>i</sub>. A direção de otimização polinômio canônico fatorial  $\hat{f}_i(\mathbf{x})$  depende da correlação entre o fator *i* e as MPP’s que ele representa.

Repetindo-se o procedimento (d) para o escore rotacionado do segundo fator, obtêm-se uma Matriz *Payoff* bidimensional dos polinômios canônicos fatoriais. A partir desse ponto, o método NBI modificado é aplicado. Considerando-se que a variância associada ao fator *i* é seu autovalor  $\lambda$ , a estatística FMSE<sub>i</sub> será dada, segundo Leite (2019), por:

$$FMSE_i(\mathbf{x}) = \left[ \hat{f}_i(\mathbf{x}) - T \right]^2 + \sigma_f^2 \quad (3.10)$$

Há duas restrições elípticas neste problema de otimização. A primeira restrição é dada por:

$$g_1(\mathbf{x}) = \left[ \hat{\mathbf{Y}}(\mathbf{x}) - \boldsymbol{\varphi} \right]^T \hat{\boldsymbol{\Sigma}}^{-1} \left[ \hat{\mathbf{Y}}(\mathbf{x}) - \boldsymbol{\varphi} \right] \leq \chi_{(p,\alpha)}^2 \quad (3.11)$$

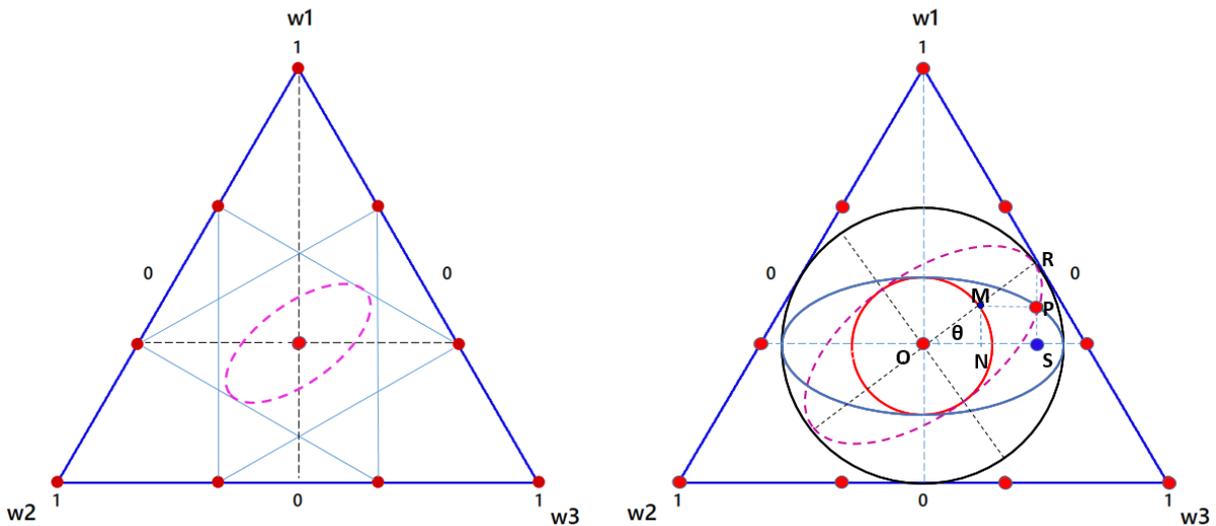
Onde:  $\mathbf{Y}(\mathbf{x})$  é um vetor com todos os polinômios canônicos das funções originais (MPP’s),  $(\boldsymbol{\varphi})$  é um vetor contendo os ótimos individuais dessas funções,  $\hat{\boldsymbol{\Sigma}}^{-1}$  é uma estimativa da inversa da matriz de variância-covariância dos dados experimentais das MPP’s e  $\chi_{(p,\alpha)}^2$  é uma estatística do Qui-quadrado para p funções objetivo e  $\alpha=5\%$ .

A estimativa de  $\hat{\boldsymbol{\Sigma}}^{-1}$  é obtida considerando-se a matriz experimental do arranjo de misturas  $\{q, m\}$ ,  $\mathbf{X}$ , um vetor  $\mathbf{z}^T(\mathbf{x})$  relativo ao modelo descrito pelo polinômio canônico em um dado ponto  $\mathbf{x}$  (pertencente à região Simplex), e a matriz de variância-covariância  $\boldsymbol{\Sigma}$  das

múltiplas respostas de interesse; no caso desta tese, estas respostas serão as métricas MPP's. Com base nestes elementos, Khuri e Conlon (1981) propuseram a restrição elíptica apresentada na equação (3.11) a seguir:

$$g_1(\mathbf{x}) = [\hat{\mathbf{Y}}(\mathbf{x}) - \boldsymbol{\varphi}]^T \left\{ \left[ \mathbf{z}^T(\mathbf{x}) (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{z}(\mathbf{x}) \boldsymbol{\Sigma} \right]^{-1} \right\} [\hat{\mathbf{Y}}(\mathbf{x}) - \boldsymbol{\varphi}] \leq \chi_{(p, \alpha)}^2 \quad (3.12)$$

A segunda restrição elíptica foi proposta por Cornell (2002) e é utilizada para afastar os pesos ótimos dos vértices da região Simplex. Nesta restrição,  $w_i$  é o peso ótimo,  $w_{0i}$  é o centróide ( $1/q$ ) e  $\delta_i$  é a diferença entre limites superior (UB) e inferior (LB) para cada peso  $w_i$ , tal como representado na figura a seguir.



**Figura 3-6 -Restrição elítica para as proporções do arranjo simplex (ou pesos).  
Fonte: Próprio autor**

Dessa forma, a segunda restrição elíptica poderá ser escrita como:

$$g_2(\mathbf{x}) = \sum_{i=1}^q \left( \frac{w_i - w_{0i}}{\delta_i} \right)^2 \leq 1 \quad (3.13)$$

Acrescentando-se a restrição de que a soma dos pesos tem de ser igual à unidade, a otimização do polinômio canônico fatorial sujeito às restrições elípticas poderá ser escrita da seguinte forma:

$$\begin{aligned}
\text{Min}_{\mathbf{x} \in \Omega} \bar{f}_i^{\text{Rot}}(\mathbf{x}) &= \beta_{1i} w_1 + \beta_{2i} w_2 + \beta_{3i} w_3 + \beta_{12i} w_1 w_2 + \beta_{13i} w_1 w_3 + \beta_{23i} w_2 w_3 \\
&\quad + \beta_{12-i} w_1 w_2 (w_1 - w_2) + \beta_{13-i} w_1 w_3 (w_1 - w_3) + \beta_{23-i} w_2 w_3 (w_2 - w_3) \\
\text{S.t.}: g_1(\mathbf{x}) &= [\hat{\mathbf{Y}}(\mathbf{x}) - \boldsymbol{\phi}]^T \left\{ \left[ \mathbf{z}^T(\mathbf{x}) (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{z}(\mathbf{x}) \boldsymbol{\Sigma} \right]^{-1} \right\} [\hat{\mathbf{Y}}(\mathbf{x}) - \boldsymbol{\phi}] \leq \chi_{(p,\alpha)}^2 \\
g_2(\mathbf{x}) &= \sum_{i=1}^q \left( \frac{w_i - w_{0i}}{\delta_i} \right)^2 \leq 1 \\
g_3(\mathbf{x}) &= \sum_{i=1}^q w_i = 1 \\
0 &\leq w_i \leq 1
\end{aligned} \tag{3.14}$$

O sistema de otimização (3.14) descreve a otimização de um polinômio canônico fatorial; entretanto, substituindo-se a função objetivo pelo modelo de uma dada MPP, encontra-se o valor ótimo individual ( $\boldsymbol{\phi}$ ). Denotando-se  $\text{FMSE}_{(g_1)}(\mathbf{x})$  como a superfície de resposta de mistura relativa ao Fator 1, bem como  $\text{FMSE}_{(g_1)}^U(\mathbf{x})$  e  $\text{FMSE}_{(g_1)}^N(\mathbf{x})$  como seus respectivos pontos de Utopia e Nadir, assim como  $\text{FMSE}_{(g_2)}^U(\mathbf{x})$  e  $\text{FMSE}_{(g_2)}^N(\mathbf{x})$ , os respectivos pontos de Utopia e Nadir para a superfície de resposta de mistura relativa ao Fator 2,  $\text{FMSE}_{(g_2)}(\mathbf{x})$ , então, pode-se escrever o método NBI com restrições elípticas da seguinte forma:

$$\begin{aligned}
\text{Min}_{\mathbf{x} \in \Omega} \bar{f}_{(Rot)}(\mathbf{x}) &= \left[ \frac{\text{FMSE}_{(g_1)}(\mathbf{x}) - \text{FMSE}_{(g_1)}^U}{\text{FMSE}_{(g_1)}^N - \text{FMSE}_{(g_1)}^U} \right] \\
\text{S.t.}: \bar{g}_{1(Rot)}(\mathbf{x}) &= \left[ \frac{\text{FMSE}_{(g_1)}(\mathbf{x}) - \text{FMSE}_{(g_1)}^U}{\text{FMSE}_{(g_1)}^N - \text{FMSE}_{(g_1)}^U} \right] - \left[ \frac{\text{FMSE}_{(g_2)}(\mathbf{x}) - \text{FMSE}_{(g_2)}^U}{\text{FMSE}_{(g_2)}^N - \hat{F}_{Rot(g_2)}^U} \right] + 2\beta_i - 1 = 0 \\
g_2(\mathbf{x}) &= [\hat{\mathbf{Y}}(\mathbf{x}) - \boldsymbol{\phi}]^T \left\{ \mathbf{z}^T(\mathbf{x}) (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{z}(\mathbf{x}) \hat{\boldsymbol{\Sigma}} \right\}^{-1} [\hat{\mathbf{Y}}(\mathbf{x}) - \boldsymbol{\phi}] \leq \chi_{(p,\alpha)}^2 \\
g_3(\mathbf{x}) &= \sum_{i=1}^q \left( \frac{w_i - w_{0i}}{\delta_i} \right)^2 \leq 1 \\
g_4(\mathbf{x}) &= \sum_{i=1}^q w_i = 1 \\
\text{FMSE}^i &= [\hat{F}_{Rot(g_1)}(\mathbf{x}) - T_i]^2 + \lambda_i \\
0 &\leq \beta_i \leq 1
\end{aligned} \tag{3.15}$$

No sistema de equações (3.15),  $\mathbf{X}$  representa a matriz experimental do arranjo de misturas  $\{q, m\}$ ,  $\Sigma$  é a matriz de variância-covariância das múltiplas respostas de interesse,  $\mathbf{x}$  é um ponto pertencente à região Simplex e  $\mathbf{z}^T(\mathbf{x})$  é um vetor relativo ao modelo descrito pelo polinômio canônico em um dado ponto, tal que:

$$\mathbf{z}^T(\mathbf{x}) = [w_1 \ w_2 \ w_3 \ w_1w_2 \ w_1w_3 \ w_2w_3 \ w_1w_2(w_1 - w_2) \ w_1w_3(w_1 - w_3) \ w_2w_3(w_2 - w_3)] \quad (3.16)$$

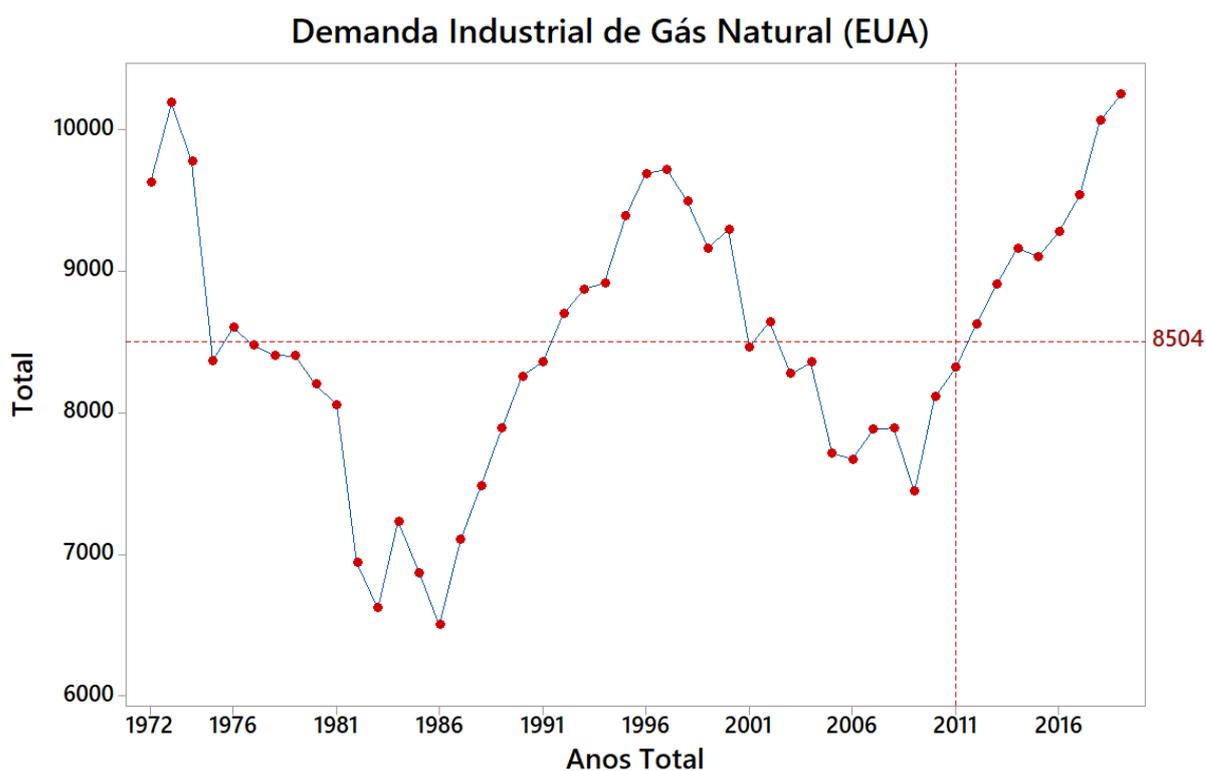
O sistema de equações (3.15) é a principal contribuição desta tese e consiste em uma formulação para o método NBI, em que as funções objetivo são polinômios canônicos dos escores rotacionados de fator de métricas de performance e de precisão de previsões. O método restringe a região de solução por uma elipse inscrita na região Simplex e os valores ótimos das métricas ao elipsóide de confiança de 95%. Destaca-se que as funções  $FMSE_{(i)}(\mathbf{x})$  são polinômios canônicos para escores de fator obtidos apenas com métricas cujos pontos de ótimo não são coincidentes. Esta seleção foi feita com base na Matriz Payoff das soluções individuais, processada por uma Análise Hierárquica de Cluster e reduzida dimensionalmente pela 2ª regra de Jolliffe.

No próximo capítulo, será apresentada uma aplicação desse método ao caso da previsão de demanda anual de gás natural.

## 4 APLICAÇÃO, RESULTADOS E DISCUSSÃO

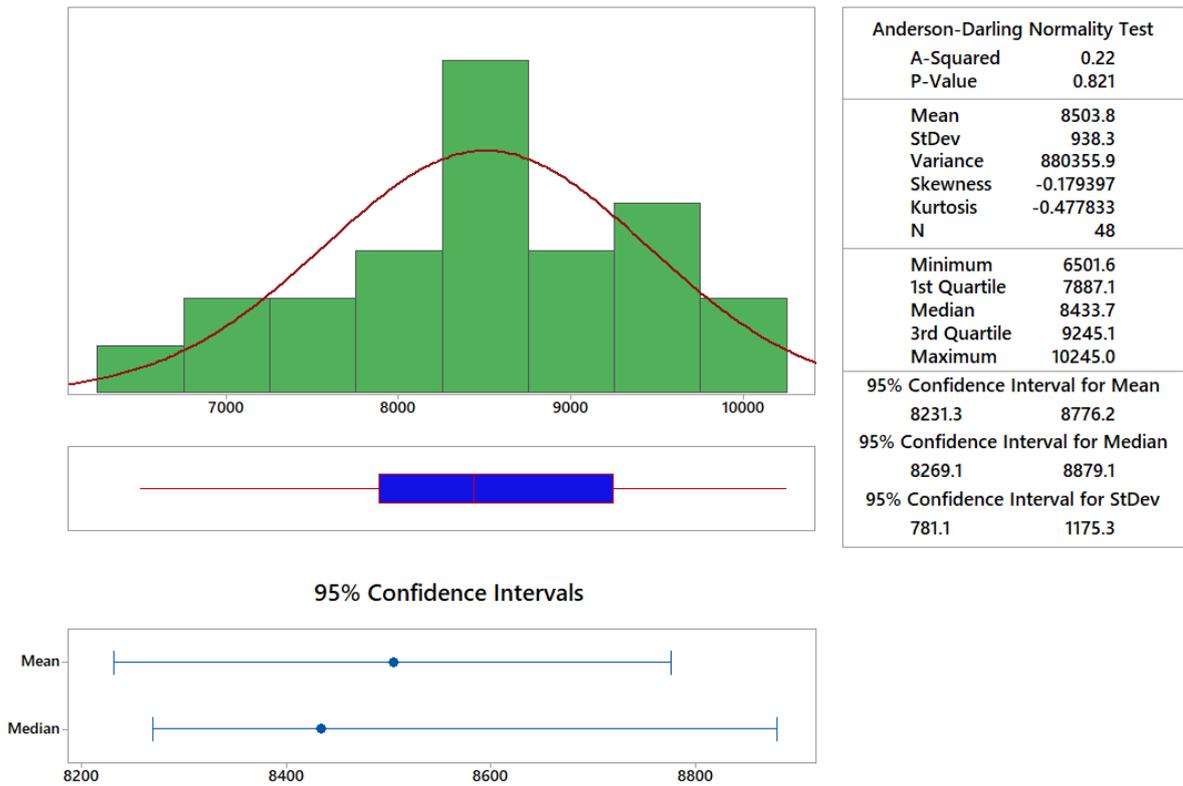
### 4.1 SÉRIE HISTÓRICA DE DEMANDA DE GÁS NATURAL E SUAS CARACTERÍSTICAS

Para testar o método proposto, será utilizada uma série temporal de demanda anual de Gás Natural, de acordo com o volume consumido pelo Setor Industrial dos EUA. Esta série, formada pelo consumo industrial norte-americano durante os últimos 48 anos, foi disponibilizada pela U.S. *Energy Information Administration* em seu site. Além disso, a série foi recortada para formar dois conjuntos: um de treinamento, formado pelas demandas observadas entre os anos de 1972 e 2011 e denominado de período *In-sample*, e um de teste, denominado de período *out-of-sample*, caracterizado por observações entre os anos de 2012 e 2019. A série histórica completa é apresentada na Figura 4.1 a seguir.

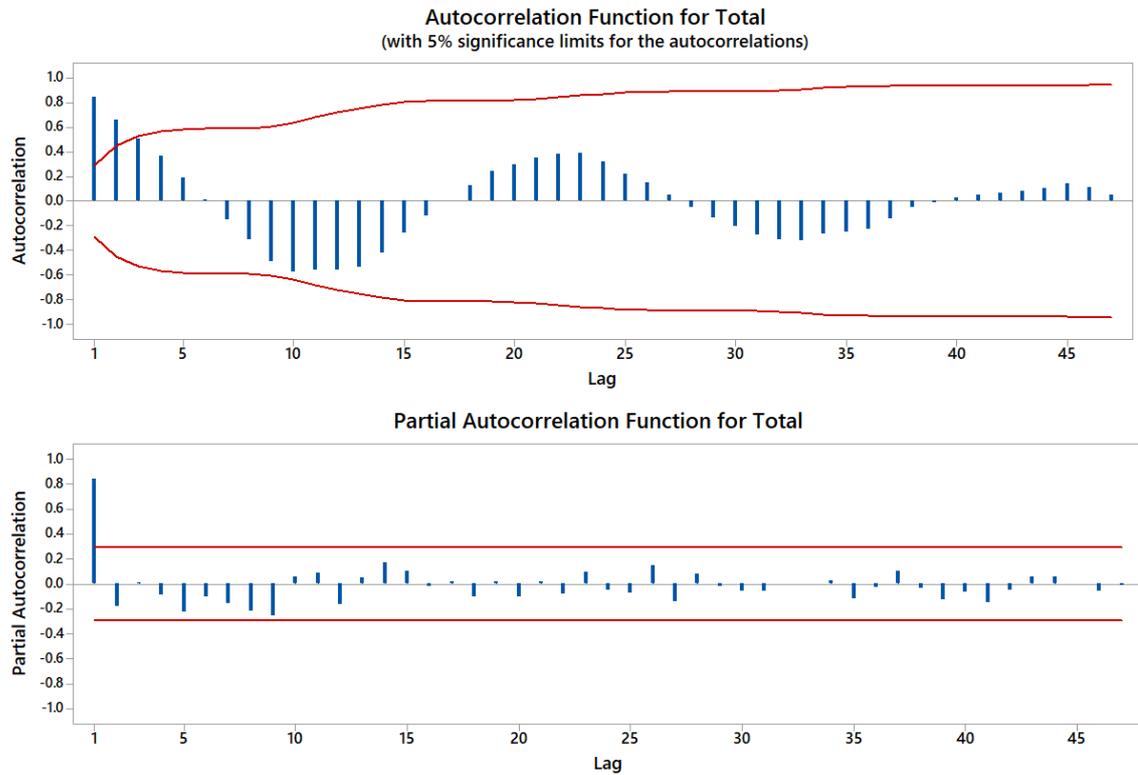


**Figura 4-1 - Série Original de demanda industrial de gás natural nos EUA (1972-2019).**  
**Fonte: Próprio autor**

A Figura 4.2 apresenta uma análise estatística da série completa. Conforme pode ser observado pelo valor do P-Value (0,821), a série segue o modelo de distribuição normal. Para um P-Value > 0,05, aceita-se a hipótese nula de que os dados seguem uma distribuição normal.



**Figura 4-2 - Análise estatística da série completa.  
Fonte: Próprio autor**



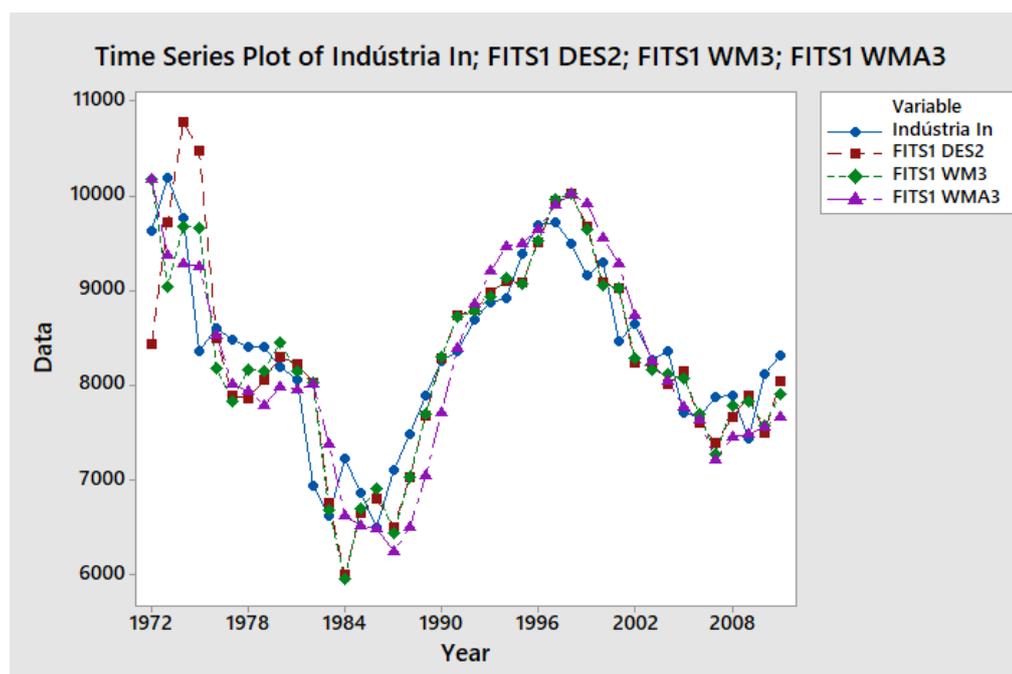
**Figura 4-3 - Gráfico de Autocorrelação Parcial e Total  
Fonte: Próprio autor**

Utilizando-se o conjunto de treinamento, vários métodos de previsão foram avaliados, entre os quais se destacam: o métodos Autoregressive Integrated Moving Average (ARIMA), Double Exponential Smoothing (DES), Winters' Method Multiplicative (WM), Winters' Method Additive (WMA), além dos métodos de análise de tendências linear, quadrática exponencial e logística (S-Curve Pearl-Reed). Alternando-se os parâmetros disponíveis de cada método (Nível, Tendência e Sazonalidade), os melhores ajustes foram alcançados com os parâmetros descritos na Tab. 4.1.

**Tabela 4-1 - Valores dos parâmetros utilizados nos melhores métodos**

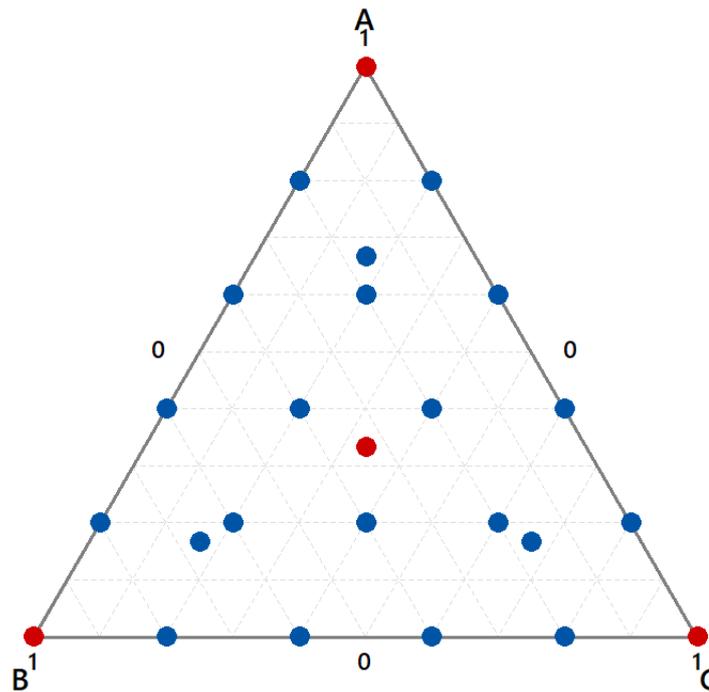
	WMA1	WMA2	WMA3	WMA4	WMA5	WM1	WM2
Level	<b>0.6</b>	<b>0.4</b>	<b>0.3</b>	<b>0.4</b>	<b>0.4</b>	<b>0.3</b>	<b>0.5</b>
Trend	<b>0.8</b>	<b>0.8</b>	<b>0.7</b>	<b>0.5</b>	<b>0.4</b>	<b>0.8</b>	<b>0.8</b>
Seasonal	<b>0.2</b>						
	WM3	WM4	WM5	DES1	DES2	DES 3	DES4
Level	<b>0.6</b>	<b>0.7</b>	<b>0.8</b>	<b>0.4</b>	<b>0.6</b>	<b>0.5</b>	<b>0.5</b>
Trend	<b>0.8</b>	<b>0.8</b>	<b>0.8</b>	<b>0.8</b>	<b>0.8</b>	<b>0.3</b>	<b>0.2</b>
Seasonal	<b>0.2</b>	<b>0.2</b>	<b>0.2</b>	---	---	---	---

Diante desses resultados, os melhores métodos obtidos foram os seguintes: WMA3, WM3 e DES2. Chegou-se a essa conclusão pelos melhores resultados nas análises *in-sample* e *out-of-sample analysis*, bem como seu ajuste em relação à série de gás natural.



**Figura 4-4 - Ajustes dos melhores métodos em relação a série original no conjunto in-sample.**  
**Fonte: Próprio autor**

De cada um desses métodos, foram armazenados os valores ajustados (*fits*), os resíduos e os valores previstos. Considerando os resíduos dos métodos WMA3, WM3 e DES2 e um Arranjo de Misturas do tipo “Simplex Lattice Design” (Figura 4.5), com número de componentes igual a “3” e grau do polinômio (ou *Degree of Lattice*) “5”, incrementado com pontos centrais e axiais com center point e axial points, sem réplicas e sem execuções randomizadas, foram produzidas 25 séries ponderadas de resíduos.



**Figura 4-5 - Arranjo de Misturas Simplex Lattice {3, 5}.**  
Fonte: Próprio autor

Cada ponto representado no *Simplex Lattice* da Figura 4-5 corresponde a um dos 25 vetores distintos de pesos necessários para a combinação das séries de resíduos. A combinação é simples: em face do arranjo feito, os pesos obtidos para cada método são utilizados para a combinação dos resíduos dos 3 métodos ( $W_{11}R_{11}; W_{12}R_{11}; \dots; W_{1p}R_{11}$ ), através da multiplicação de todos os valores da série de resíduos pelo valor atribuído pelo arranjo de misturas. Com a série resultante dessas combinações, foram calculadas as métricas de performance e de precisão de previsão (MPP's) MAE, MSE, RMSE, MASE, RMSPE, MAPE, sMAPE, U1, U2, MdMAE, MdAPE, sMdAPE, VAR e SD, tais como descritas nas tabelas a seguir.

Tabela 4-2 - Métricas de performance de previsão de 25 séries de resíduos ponderadas (Parte I).

WMA3	WM3	DES2	MAE	MSE	RMSE	MASE	RMSPE	MAPE
1.00	0.00	0.00	433.75	283759.59	532.69	1.103	6.686	5.324
0.80	0.20	0.00	412.00	259807.27	509.71	1.048	6.384	5.063
0.80	0.00	0.20	399.80	254483.68	504.46	1.017	6.372	4.948
0.60	0.40	0.00	398.17	245330.65	495.31	1.013	6.190	4.893
0.60	0.20	0.20	386.35	240287.97	490.19	0.982	6.179	4.783
0.60	0.00	0.40	387.96	246711.50	496.70	0.987	6.272	4.812
0.40	0.60	0.00	391.80	240329.74	490.23	0.996	6.115	4.810
0.40	0.40	0.20	379.09	235567.97	485.35	0.964	6.105	4.689
0.40	0.20	0.40	383.23	242272.41	492.21	0.975	6.200	4.742
0.40	0.00	0.60	400.12	260443.06	510.34	1.018	6.396	4.929
0.20	0.80	0.00	391.47	244804.53	494.78	0.996	6.164	4.797
0.20	0.60	0.20	380.85	240323.68	490.23	0.969	6.155	4.696
0.20	0.40	0.40	386.30	247309.03	497.30	0.982	6.250	4.762
0.20	0.20	0.60	402.06	265760.59	515.52	1.022	6.446	4.936
0.20	0.00	0.80	417.82	295678.35	543.76	1.063	6.732	5.109
0.00	1.00	0.00	393.24	258755.03	508.68	1.000	6.333	4.810
0.00	0.80	0.20	386.15	254555.09	504.53	0.982	6.325	4.744
0.00	0.60	0.40	391.19	261821.35	511.68	0.995	6.419	4.805
0.00	0.40	0.60	406.88	280553.82	529.67	1.035	6.611	4.977
0.00	0.20	0.80	422.57	310752.50	557.45	1.075	6.891	5.150
0.00	0.00	1.00	438.25	352417.38	593.65	1.114	7.251	5.322
0.33	0.33	0.33	379.00	239358.38	489.24	0.964	6.160	4.691
0.67	0.17	0.17	392.41	244026.81	493.99	0.998	6.225	4.852
0.17	0.67	0.17	381.66	241980.29	491.91	0.971	6.169	4.701
0.17	0.17	0.67	407.82	276282.26	525.63	1.037	6.551	4.997

**Tabela 4-3 - Métricas de performance de previsão de 25 séries de resíduos ponderadas (Parte II).**

<b>WMA3</b>	<b>WM3</b>	<b>DES2</b>	<b>Smape</b>	<b>U1</b>	<b>U2</b>	<b>MdMAE</b>	<b>MdAPE</b>
1.00	0.00	0.00	0.05365	0.0319	0.0636	467.55	5.49
0.80	0.20	0.00	0.05096	0.0305	0.0608	417.90	4.89
0.80	0.00	0.20	0.04974	0.0302	0.0602	361.59	4.90
0.60	0.40	0.00	0.04921	0.0296	0.0591	322.90	4.01
0.60	0.20	0.20	0.04803	0.0293	0.0585	308.53	3.98
0.60	0.00	0.40	0.04823	0.0297	0.0593	319.43	4.13
0.40	0.60	0.00	0.04835	0.0293	0.0585	274.49	3.66
0.40	0.40	0.20	0.04706	0.0290	0.0579	260.32	3.45
0.40	0.20	0.40	0.04748	0.0294	0.0587	277.47	3.65
0.40	0.00	0.60	0.04926	0.0305	0.0609	329.00	3.96
0.20	0.80	0.00	0.04820	0.0296	0.0590	295.58	3.64
0.20	0.60	0.20	0.04711	0.0293	0.0585	293.18	3.46
0.20	0.40	0.40	0.04766	0.0297	0.0593	285.58	3.53
0.20	0.20	0.60	0.04928	0.0308	0.0615	319.17	4.04
0.20	0.00	0.80	0.05093	0.0325	0.0649	330.58	4.07
0.00	1.00	0.00	0.04834	0.0304	0.0607	330.82	3.69
0.00	0.80	0.20	0.04759	0.0301	0.0602	307.86	3.47
0.00	0.60	0.40	0.04807	0.0306	0.0611	328.34	3.52
0.00	0.40	0.60	0.04967	0.0316	0.0632	336.16	4.19
0.00	0.20	0.80	0.05130	0.0333	0.0665	316.90	4.11
0.00	0.00	1.00	0.05296	0.0354	0.0708	335.23	4.25
0.33	0.33	0.33	0.04700	0.0292	0.0584	286.68	3.60
0.67	0.17	0.17	0.04875	0.0295	0.0589	325.04	4.33
0.17	0.67	0.17	0.04718	0.0294	0.0587	303.91	3.65
0.17	0.17	0.67	0.04986	0.0314	0.0627	314.49	3.96

Tabela 4-4 - Métricas de performance de previsão de 25 séries de resíduos ponderadas (Parte III).

WMA3	WM3	DES2	SMdAPE	VAR	SD
1.00	0.00	0.00	0.05439	285241.67	534.08
0.80	0.20	0.00	0.05011	261756.52	511.62
0.80	0.00	0.20	0.05020	256891.40	506.84
0.60	0.40	0.00	0.04091	247878.45	497.87
0.60	0.20	0.20	0.04064	243234.53	493.19
0.60	0.00	0.40	0.04182	250310.68	500.31
0.40	0.60	0.00	0.03663	243607.47	493.57
0.40	0.40	0.20	0.03451	239184.73	489.07
0.40	0.20	0.40	0.03682	246482.07	496.47
0.40	0.00	0.60	0.04045	265499.49	515.27
0.20	0.80	0.00	0.03669	248943.56	498.94
0.20	0.60	0.20	0.03457	244742.02	494.71
0.20	0.40	0.40	0.03525	252260.55	502.26
0.20	0.20	0.60	0.04050	271499.16	521.06
0.20	0.00	0.80	0.04159	302457.84	549.96
0.00	1.00	0.00	0.03757	263886.74	513.70
0.00	0.80	0.20	0.03534	259906.39	509.81
0.00	0.60	0.40	0.03518	267646.11	517.35
0.00	0.40	0.60	0.04188	287105.92	535.82
0.00	0.20	0.80	0.04094	318285.79	564.17
0.00	0.00	1.00	0.04239	361185.74	600.99
0.33	0.33	0.33	0.03582	243581.53	493.54
0.67	0.17	0.17	0.04430	246691.75	496.68
0.17	0.67	0.17	0.03652	246482.48	496.47
0.17	0.17	0.67	0.03960	282462.75	531.47

Na próxima seção, essas métricas serão utilizadas para a aplicação do Método Payoff-Jolliffe Fatorial com restrições elípticas.

## 4.2 APLICAÇÃO DO MÉTODO PAYOFF-JOLLIFFE FATORIAL COM RESTRIÇÕES ELÍPTICAS

O Método Payoff-Jolliffe Fatorial com restrições elípticas foi genericamente descrito na seção 3.3 em uma sequência de 15 passos. Neste item, será apresentado cada um dos passos por meio de exemplos numéricos derivados dos dados apresentados nas tabelas 4-2, 4-3 e 4-4.

**Passo 1:** Modelar individualmente cada uma das métricas de performance de previsão (MPP's) de acordo com o arranjo de misturas escolhido e o polinômio canônico de interesse, por exemplo, um arranjo cúbico completo. A **Tabela 4-5** a seguir apresenta os resultados obtidos no caso escolhido para a ilustração do funcionamento do método para as métricas MAE, MSE, RMSE, RMSPE, MAPE e sMAPE. Todos os polinômios canônicos de grau 5 (modelo cúbico completo) para estas métricas apresentam ajustes  $R^2_{adj}$  superiores a 99%.

**Tabela 4-5 - Coeficientes do polinômio cúbico completo para as métricas de performance (Parte I).**

Term	MAE	MSE	RMSE	MASE	RMSPE	MAPE	sMAPE
WA	434.31	283759.59	532.91	1.104	6.689	5.330	0.05371
WM	393.75	258755.03	508.75	1.001	6.334	4.815	0.04839
DES	438.06	352417.38	593.84	1.114	7.252	5.320	0.05294
WA*WM	-78.35	-118446.30	-116.56	-0.199	-1.491	-0.907	-0.00932
WA*DES	-172.88	-268796.72	-248.10	-0.440	-2.639	-1.868	-0.01883
WM*DES	-69.88	-143327.57	-127.20	-0.178	-1.156	-0.730	-0.00743
WA*WM*DES	-136.86	0.00	-29.69	-0.348	-0.251	-1.372	-0.01333
WA*WM*(-)	-14.63	0.00	2.30	-0.037	0.035	-0.171	-0.00172
WA*DES*(-)	-90.04	0.00	-14.58	-0.229	-0.119	-0.949	-0.00939
WM*DES*(-)	-60.62	0.00	-9.89	-0.154	-0.079	-0.631	-0.00609

**Tabela 4-6 - Coeficientes do polinômico cúbico completo para as métricas de performance (Parte II).**

Term	U1	U2	MdMAE	MdAPE	sMdAPE	VAR	SD
WA	0.0319	0.0636	476.91	5.61	0.05590	2.85E+05	534.30
WM	0.0304	0.0607	329.76	3.71	0.03792	2.64E+05	513.77
DES	0.0354	0.0709	330.00	4.24	0.04245	3.61E+05	601.18
WA*WM	-0.0070	-0.0139	-389.32	-3.12	-0.02991	-1.20E+05	-117.62
WA*DES	-0.0148	-0.0296	-344.03	-3.36	-0.02938	-2.72E+05	-248.51
WM*DES	-0.0076	-0.0152	-13.16	-0.55	-0.00719	-1.47E+05	-128.34
WA*WM*DES	-0.0018	-0.0035	-417.97	-3.68	-0.05331	1.16E-09	-29.60
WA*WM*(-)	0.0001	0.0003	108.21	-0.12	0.00717	-6.05E-11	1.88
WA*DES*(-)	-0.0009	-0.0017	-348.03	-0.55	-0.00368	-9.39E-11	-15.91
WM*DES*(-)	-0.0006	-0.0012	-8.93	-2.32	-0.02197	-3.66E-10	-10.16

**Passo 2:** Realizar a otimização individual de cada MPP e armazenar as soluções ótimas (pesos)  $w_1^*$ ,  $w_2^*$ , ...,  $w_p^*$ . Neste passo, uma otimização não linear por gradiente reduzido generalizado é aplicada a cada MPP, considerando-se como restrições a soma unitária dos pesos e uma restrição elíptica para definir o espaço de solução dentro da região simplex. As soluções obtidas nesse passo são apresentadas na **Tabela 4-7** a seguir.

**Tabela 4-7 - Pesos ótimos individuais para cada métrica utilizada para a avaliação da combinação.**

Métricas	w1	w2	w3	S(w)
MAE	0.324	0.446	0.230	1.086
MSE	0.400	0.417	0.183	1.060
RMSE	0.399	0.422	0.179	1.056
MASE	0.324	0.446	0.230	1.086
RMSPE	0.378	0.505	0.117	0.965
MAPE	0.282	0.504	0.215	1.044
sMAPE	0.278	0.495	0.228	1.055
U1	0.395	0.423	0.182	1.058
U2	0.399	0.422	0.179	1.056
MdMAE	0.312	0.501	0.187	1.033
MdAPE	0.164	0.652	0.185	0.853
sMdAPE	0.177	0.630	0.193	0.888
VAR	0.415	0.409	0.176	1.053
SD	0.415	0.413	0.172	1.049

Na Tabela 4-7,  $S(w)$  representa a função de entropia, utilizada para se avaliar a diversificação dos pesos na combinação.

**Passo 3:** Aplicar a Análise Hierárquica de Cluster para observações, usando distância euclidiana e o método de Ward para avaliar a similaridade entre as soluções ótimas. Com base nos resultados da Tabela 4-6 (anterior), obtém-se o seguinte dendograma da **Fig. 4-6**:

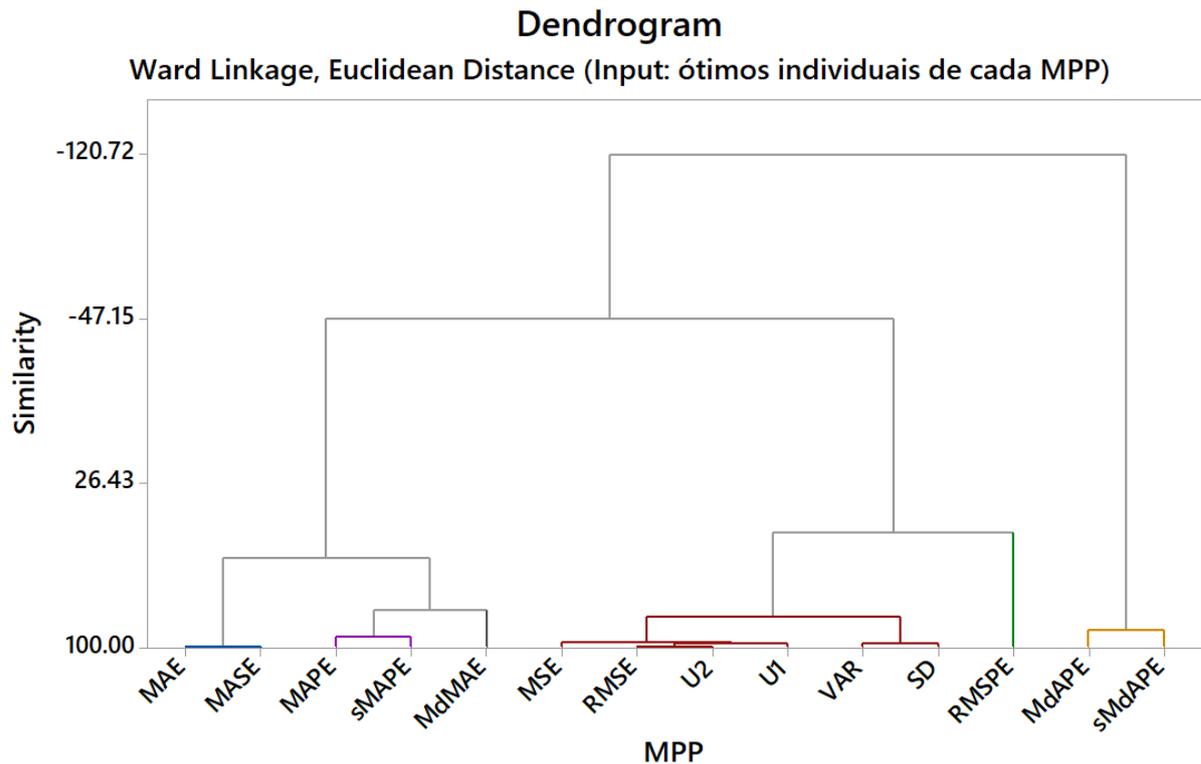


Figura 4-6 - Dendograma dos agrupamentos de métricas de performance para previsão (MPP) por semelhança entre os ótimos individuais.

Fonte: Próprio autor

**Passo 4:** Eliminar as MPP's com mesmo ponto de ótimo da análise, no qual se observa que métricas, como MAE e MASE, ou RMSE e U2, possuem o mesmo ponto de ótimo. Dessa forma, o interessado poderá eliminar uma das variáveis de seu interesse. Por exemplo, retenham-se as métricas MAE e RMSE e eliminem-se as métricas MASE e U2. Contudo, mesmo com essas eliminações, nota-se que alguns agrupamentos apresentam muitas métricas com distâncias (ou similaridades) entre seus pontos de ótimo muito baixas. Para escolher métricas representativas desses agrupamentos, recorre-se ao **Passo 5** do método proposto.

**Passo 5:** Este passo do método define que, se a similaridade entre as soluções individuais dentro de um determinado cluster for alta, aplica-se a 2ª regra de Jolliffe para escolher qual métrica

representará o cluster. Dessa forma, aplica-se a análise de componentes principais (PCA) ao conjunto de métricas do cluster de interesse, eliminando-se as variáveis (MPP's) com maiores autovetores associados aos menores autovalores. Desse modo, tem-se que:

**Tabela 4-8 - Análise de Componentes Principais (PCA) para o cluster MSE, RMSE, U1, VAR e SD.**

<b>Autovalores da matriz de correlação e respectivos percentuais de explicação</b>					
Eigenvalue	4.995	0.005	0.000	0.000	0.000
Proportion	0.999	0.001	0.000	0.000	0.000
Cumulative	0.999	1.000	1.000	1.000	1.000
<b>Autovetores associados a cada PC</b>					
Variable	PC1	PC2	PC3	PC4	PC5
MSE	0.447	0.356	<b>-0.595</b>	-0.273	0.495
RMSE	0.447	0.230	0.433	-0.633	-0.399
U1	0.447	0.486	0.252	<b>0.705</b>	-0.062
VAR	0.447	-0.469	-0.487	0.165	<b>-0.561</b>
SD	0.447	-0.603	0.397	0.036	0.527

Na análise de componentes principais aplicada ao cluster exemplificado, observa-se que três dos cinco autovalores são nulos, o que, segundo a teoria multivariada, indica a presença de variáveis redundantes no conjunto. De acordo com a 2ª regra de Jolliffe, eliminam-se as variáveis nestes autovalores que apresentarem, sucessivamente, os maiores autovetores. Desse modo, eliminam-se as variáveis VAR, U1, MSE. Como os autovetores no primeiro componente são todos iguais, pode-se aplicar a 3ª regra de Jolliffe, o que conduz à eliminação da métrica RMSE e à retenção da variável SD. Análise similar pode ser aplicada ao cluster formado pelas variáveis MAPE, sMAPE e MdMAE. Neste caso, apenas a variável MAPE pode ser eliminada do grupo de acordo com a 2ª regra de Jolliffe.

**Passo 6:** Definir o conjunto final de métricas MPP. Nesse ponto, já se reduziu o conjunto original de 14 métricas para apenas 7, a saber: sMdAPE, MdAPE, MdMAE, SD, RMSPE, MAE e sMAPE.

**Passo 7:** Aplica-se uma análise fatorial com método de rotação “Varimax” para dividir as 7 métricas definidas no **Passo 6** em dois conjuntos ou “fatores rotacionados”. Esta análise fatorial

é realizada com as respostas do arranjo de misturas para as 7 métricas selecionadas, e não com os pontos de ótimo. O resultado dessa análise é apresentado a seguir.

**Tabela 4-9 - Análise Fatorial de MPP's remanescentes (Extração por componentes principais e rotação varimax).**

Métrica	Fator1	Fator2	Comunalidade
sMdAPE	<b>0.965</b>	0.208	0.974
MdAPE	<b>0.959</b>	0.241	0.979
MdMAE	<b>0.914</b>	0.252	0.899
SD	0.110	<b>0.990</b>	0.992
RMSPE	0.210	<b>0.969</b>	0.983
MAE	0.521	<b>0.840</b>	0.977
sMAPE	0.643	<b>0.749</b>	0.975
Variance	3.427	3.351	6.778
% Var	0.490	0.479	0.968

**Passo 8:** Neste passo, pode-se eliminar a MPP com o menor carregamento (*loadings*) dentro de cada um dos dois fatores. No caso do presente exemplo, eliminam-se MdMAE no Fator 1 e sMAPE no Fator 2, respectivamente.

**Passo 9:** Neste passo, aplica-se a 2ª regra de Jolliffe para eliminar variáveis redundantes dentro de cada um dos dois fatores.

**Tabela 4-10 - Análise de componentes principais das MPP's pertencentes ao Fator 2.**

Eigenvalue	2.849	0.145	0.006
Proportion	0.950	0.048	0.002
Cumulative	0.950	0.998	1.000
Variable	PC1	PC2	PC3
SD	0.582	-0.477	-0.659
RMSPE	0.587	-0.315	<b>0.746</b>
MAE	0.563	0.821	-0.097

Observa-se que a eliminação de RMSPE é plausível. Do mesmo modo, sMdAPE e MdAPE podem ser analisadas.

**Passo 10:** Obter o conjunto final de métricas. Considerando os resultados dos passos anteriores, o conjunto final de métricas utilizadas é composto pelas seguintes métricas de performance de previsão: MAE, SD e sMdAPE. Este conjunto define que variáveis de resposta devem ser utilizadas na otimização multiobjetivo.

**Passo 11:** Aplicar a Análise fatorial com rotação do tipo “Varimax” aos dados do arranjo de misturas para o conjunto final de métricas, representando a estrutura de variância-covariância do conjunto em dois fatores. A partir do **Passo 11**, o método proposto passa a tratar do problema de otimização multiobjetivo, e não mais sobre que respostas devem ser consideradas no problema de definição de pesos ótimos em combinações de métodos de previsão. Assim, neste passo, obtêm-se os seguintes resultados para a análise fatorial definida:

**Tabela 4-11 - Análise fatorial das MPP's mais significativas e não redundantes.**

<b>Métrica</b>	<b>Fator 1</b>	<b>Fator 2</b>	<b>Comunalidade</b>
SD	0.988	0.112	0.989
MAE	0.843	0.520	0.981
sMdAPE	0.202	0.977	0.996
Variance	1.7278	1.2383	2.9661
% Var	0.576	0.413	0.989

**Passo 12:** Armazenar os escores de cada um dos fatores. Os resultados desta etapa estão descritos pela **Tabela 4-12**.

**Passo 13:** Modelar os escores rotacionados de fator segundo o arranjo de misturas e o polinômio canônico de interesse. Os resultados desta etapa estão descritos pela **Tabela 4-13**.

**Passo 14:** Aplicar o método NBI com restrições elípticas para formar a Fronteira de Pareto. Os resultados desta etapa estão descritos pela **Tabela 4-14**.

**Tabela 4-12 - Simplex Lattice Design, MPP's e escores rotacionados de fator.**

<b>WA</b>	<b>WM</b>	<b>DES</b>	<b>MAE</b>	<b>sMdAPE</b>	<b>SD</b>	<b>F1</b>	<b>F2</b>
1.000	0.000	0.000	433.7455	0.0544	534.0802	0.602313	2.782344
0.800	0.200	0.000	411.9968	0.0501	511.6215	-0.30924	2.04636
0.800	0.000	0.200	399.8009	0.0502	506.8446	-0.75245	1.999021
0.600	0.400	0.000	398.1666	0.0409	497.8739	-0.49585	0.357312
0.600	0.200	0.200	386.3465	0.0406	493.1881	-0.906	0.244992
0.600	0.000	0.400	387.9594	0.0418	500.3106	-0.73932	0.388988
0.400	0.600	0.000	391.8025	0.0366	493.5661	-0.52325	-0.45686
0.400	0.400	0.200	379.0934	0.0345	489.0652	-0.84276	-0.92725
0.400	0.200	0.400	383.2274	0.0368	496.4696	-0.67141	-0.54721
0.400	0.000	0.600	400.1170	0.0404	515.2664	0.056158	0.062404
0.200	0.800	0.000	391.4673	0.0367	498.9424	-0.38816	-0.51959
0.200	0.600	0.200	380.8484	0.0346	494.7141	-0.64761	-0.9716
0.200	0.400	0.400	386.3011	0.0352	502.2555	-0.34368	-0.88636
0.200	0.200	0.600	402.0590	0.0405	521.0558	0.260277	0.01683
0.200	0.000	0.800	417.8170	0.0416	549.9617	1.383538	0.008269
0.000	1.000	0.000	393.2351	0.0376	513.6991	0.008216	-0.5286
0.000	0.800	0.200	386.1520	0.0353	509.8101	-0.14625	-0.96949
0.000	0.600	0.400	391.1898	0.0352	517.3453	0.19565	-1.04436
0.000	0.400	0.600	406.8776	0.0419	535.8227	0.704743	0.13263
0.000	0.200	0.800	422.5654	0.0409	564.1682	1.929865	-0.25091
0.000	0.000	1.000	438.2532	0.0424	600.9873	3.246912	-0.29697
0.333	0.333	0.333	378.9993	0.0358	493.5398	-0.79924	-0.74153
0.667	0.167	0.167	392.4081	0.0443	496.6807	-0.87167	0.948521
0.167	0.667	0.167	381.6583	0.0365	496.4700	-0.69339	-0.61982
0.167	0.167	0.667	407.8239	0.0396	531.4722	0.742596	-0.22713

**Tabela 4-13 - Coeficientes das funções objetivo utilizadas no método NBI.**

Objetivos	Min	Min	Min	Max	Min	Min
Termos	MAE	SD	sMdAPE	S(w)	F1	F2
WA	433.560	534.395	0.05590	0.074	0.3783	3.177
WM	394.500	513.675	0.03668	0.074	0.0255	-0.848
DES	438.060	601.177	0.04175	0.074	3.4125	-0.065
WA*WM	-78.350	-117.620	-0.02703	2.695	-3.298	-5.27
WA*DES	-171.660	-248.660	-0.02739	2.695	-9.526	-5.44
WM*DES	-71.100	-128.180	0.00000	2.695	-5.234	
WA*WM*DES	-136.900	-29.600	-0.07308	4.356		
WA*DES*(-)	-90.270	-15.880	0.00000	0.000		
WM*DES*(-)	-60.390	-10.190	-0.02213	0.000		

O **Passo 14** compreende a aplicação do método de otimização não linear multiobjetivo NBI, associado às funções objetivo multivariadas (baseadas em escores rotacionados de fator), bem como a um conjunto de restrições elípticas definidas em função dos espaços de solução em  $\mathbf{x}$  (pesos) e em  $\mathbf{Y}$  (métricas de performance de previsão). Sua implementação é realizada a partir de uma sequência de subproblemas de otimização, a saber:

**14.a.** Definição de alvos ( $\mathbf{T}$ ) para a estatística  $FMSE_i$ : a estatística  $FMSE_i$  é uma função composta pelo polinômio canônico associado ao escore de fator  $i$ ,  $f_i(\mathbf{x})$ , seu alvo  $T$  corresponde à otimização individual de  $f_i(\mathbf{x})$ , e sua respectiva variância representada, neste caso, pela variância do fator descrita na tabela de análise fatorial. Os valores dos coeficientes dos polinômios canônicos para  $FMSE_i$  são dados pelas últimas duas colunas da Tabela 4-13. As variâncias associadas a estes polinômios são iguais a  $\sigma^2[f_1(x)] = 1,7278$  para a superfície de misturas do primeiro escore e  $\sigma^2[f_2(x)] = 1,2383$  para o segundo escore rotacionado e são apresentadas na análise fatorial da **Tabela 4-11**.

**14.b.** Com o alvo  $\mathbf{T}$  definido de acordo com o seguinte problema de otimização: os valores de alvo ( $\mathbf{T}$ ) para a superfície de misturas dos escores rotacionados de fator são obtidos pela otimização individual destas funções, tal como descrito pela Equação (3.14) e assumirão a forma do sistema de equações apresentado a seguir. Embora escritas em uma única formulação, as otimizações das duas funções objetivo são feitas individualmente.

$$(a) \underset{\mathbf{x} \in \Omega}{\text{Min}} \bar{f}_1^{\text{Rot}}(\mathbf{x}) = 0.3738w_1 + 0.0255w_2 + 3.4125w_3 - 3.298w_1w_2 - 9.526w_1w_3 - 5.234w_2w_3$$

$$(b) \underset{\mathbf{x} \in \Omega}{\text{Min}} \bar{f}_2^{\text{Rot}}(\mathbf{x}) = 3.177w_1 - 0.848w_2 - 0.065w_3 - 5.270w_1w_2 - 5.440w_1w_3$$

$$S.t.: g_1(\mathbf{x}) = [\hat{\mathbf{Y}}(\mathbf{x}) - \boldsymbol{\varphi}]^T \left\{ \mathbf{z}^T(\mathbf{x})(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{z}^T(\mathbf{x}) \hat{\boldsymbol{\Sigma}} \right\}^{-1} [\hat{\mathbf{Y}}(\mathbf{x}) - \boldsymbol{\varphi}] \leq \chi_{(p, \alpha)}^2 \quad (4.1)$$

$$g_2(\mathbf{x}) = \sum_{i=1}^q \left( \frac{w_i - w_{0i}}{\delta_i} \right)^2 \leq 1$$

$$g_3(\mathbf{x}) = \sum_{i=1}^q w_i = 1 \quad 0 \leq w_i \leq 1$$

Para se resolver o sistema de equações (4.1), é preciso desenvolver a expressão relacionada à primeira restrição elíptica. Nessa restrição,  $\mathbf{X}$  é a matriz de planejamento experimental do arranjo de misturas Simplex-Lattice  $\{3, 5\}$  incrementado. Esta matriz é dada por:

$$\mathbf{X} = \begin{bmatrix} 1.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\ 0.8000 & 0.2000 & 0.0000 & 0.1600 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\ 0.8000 & 0.0000 & 0.2000 & 0.0000 & 0.1600 & 0.0000 & 0.0000 & 0.0960 & 0.0000 \\ 0.6000 & 0.4000 & 0.0000 & 0.2400 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\ 0.6000 & 0.2000 & 0.2000 & 0.1200 & 0.1200 & 0.0400 & 0.0240 & 0.0480 & 0.0000 \\ 0.6000 & 0.0000 & 0.4000 & 0.0000 & 0.2400 & 0.0000 & 0.0000 & 0.0480 & 0.0000 \\ 0.4000 & 0.6000 & 0.0000 & 0.2400 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\ 0.4000 & 0.4000 & 0.2000 & 0.1600 & 0.0800 & 0.0800 & 0.0320 & 0.0160 & 0.0160 \\ 0.4000 & 0.2000 & 0.4000 & 0.0800 & 0.1600 & 0.0800 & 0.0320 & 0.0000 & -0.0160 \\ 0.4000 & 0.0000 & 0.6000 & 0.0000 & 0.2400 & 0.0000 & 0.0000 & -0.0480 & 0.0000 \\ 0.2000 & 0.8000 & 0.0000 & 0.1600 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\ 0.2000 & 0.6000 & 0.2000 & 0.1200 & 0.0400 & 0.1200 & 0.0240 & 0.0000 & 0.0480 \\ 0.2000 & 0.4000 & 0.4000 & 0.0800 & 0.0800 & 0.1600 & 0.0320 & -0.0160 & 0.0000 \\ 0.2000 & 0.2000 & 0.6000 & 0.0400 & 0.1200 & 0.1200 & 0.0240 & -0.0480 & -0.0480 \\ 0.2000 & 0.0000 & 0.8000 & 0.0000 & 0.1600 & 0.0000 & 0.0000 & -0.0960 & 0.0000 \\ 0.0000 & 1.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\ 0.0000 & 0.8000 & 0.2000 & 0.0000 & 0.0000 & 0.1600 & 0.0000 & 0.0000 & 0.0960 \\ 0.0000 & 0.6000 & 0.4000 & 0.0000 & 0.0000 & 0.2400 & 0.0000 & 0.0000 & 0.0480 \\ 0.0000 & 0.4000 & 0.6000 & 0.0000 & 0.0000 & 0.2400 & 0.0000 & 0.0000 & -0.0480 \\ 0.0000 & 0.2000 & 0.8000 & 0.0000 & 0.0000 & 0.1600 & 0.0000 & 0.0000 & -0.0960 \\ 0.0000 & 0.0000 & 1.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\ 0.3333 & 0.3333 & 0.3333 & 0.1111 & 0.1111 & 0.1111 & 0.0370 & 0.0000 & 0.0000 \\ 0.6667 & 0.1667 & 0.1667 & 0.1111 & 0.1111 & 0.0278 & 0.0185 & 0.0556 & 0.0000 \\ 0.1667 & 0.6667 & 0.1667 & 0.1111 & 0.0278 & 0.1111 & 0.0185 & 0.0000 & 0.0556 \\ 0.1667 & 0.1667 & 0.6667 & 0.0278 & 0.1111 & 0.1111 & 0.0185 & -0.0556 & -0.0556 \end{bmatrix}$$

Com esta matriz experimental, pode-se calcular o termo  $(X^T X)^{-1}$ , tal que:

$$(X^T X)^{-1} = \begin{bmatrix} 0.779 & 0.096 & -0.015 & -2.069 & -1.889 & -0.208 & 4.500 & -2.192 & 0.033 \\ 0.096 & 0.779 & -0.015 & -2.069 & -0.208 & -1.889 & 4.500 & 0.033 & -2.192 \\ -0.015 & -0.015 & 0.890 & -0.028 & -2.069 & -2.069 & 4.500 & 2.159 & 2.159 \\ -2.069 & -2.069 & -0.028 & 15.465 & 5.276 & 5.276 & -42.628 & 3.507 & 3.507 \\ -1.889 & -0.208 & -2.069 & 5.276 & 15.173 & 5.568 & -42.628 & 0.054 & -3.561 \\ -0.208 & -1.889 & -2.069 & 5.276 & 5.568 & 15.173 & -42.628 & -3.561 & 0.054 \\ 4.500 & 4.500 & 4.500 & -42.628 & -42.628 & -42.628 & 432.550 & 0.000 & 0.000 \\ -2.192 & 0.033 & 2.159 & 3.507 & 0.054 & -3.561 & 0.000 & 42.097 & -0.638 \\ 0.033 & -2.192 & 2.159 & 3.507 & -3.561 & 0.054 & 0.000 & -0.638 & 42.097 \end{bmatrix}$$

Considerando as variáveis de resposta MAE, sMdAPE, SD e S(w), a matriz de variância-covariância ( $\Sigma$ ) formada e sua inversa serão, respectivamente, iguais a:

$$\Sigma = \begin{bmatrix} 270.12187 & 0.05705 & 379.84071 & -3.55092 \\ 0.05705 & 0.00003 & 0.04311 & -0.00080 \\ 379.84071 & 0.04311 & 694.96304 & -4.89211 \\ -3.55092 & -0.00080 & -4.89211 & 0.09341 \end{bmatrix}$$

$$\Sigma^{-1} = \begin{bmatrix} 0.065 & -85.068 & -0.029 & 0.249 \\ -85.068 & 161359.316 & 37.175 & 97.879 \\ -0.029 & 37.175 & 0.015 & 0.009 \\ 0.249 & 97.879 & 0.009 & 21.488 \end{bmatrix}$$

O vetor  $\mathbf{z}^T(\mathbf{x}) = [w_1 \ w_2 \ w_3 \ w_1 w_2 \ w_1 w_3 \ w_2 w_3 \ w_1 w_2 (w_1 - w_2) \ w_1 w_3 (w_1 - w_3) \ w_2 w_3 (w_2 - w_3)]$  relativo ao modelo descrito pelo polinômio canônico é dado, neste caso, por:

$$\mathbf{z}^T(\mathbf{x}) = \begin{bmatrix} 0.503 & 0.271 & 0.227 & 0.136 & 0.114 & 0.061 & 0.031 & 0.031 & 0.003 \end{bmatrix}$$

Sua multiplicação pelo termo  $(X^T X)^{-1}$  fornece o seguinte resultado:

$$\mathbf{z}^T(\mathbf{x}) \cdot (.) = \begin{bmatrix} -0.025 & -0.032 & 0.036 & 0.229 & -0.010 & -0.228 & 4.568 & 0.983 & 0.082 \end{bmatrix}$$

Multiplicando-se este resultado por  $\mathbf{z}(\mathbf{x})$ , obtém-se o valor 0,175.

Nos passos seguintes, pré-multiplica-se a matriz de variância-covariância ( $\Sigma$ ) formada por este resultado,  $\mathbf{z}^T(\mathbf{x})(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{z}(\mathbf{x})$ , para se obter  $\mathbf{z}^T(\mathbf{x})(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{z}(\mathbf{x})\Sigma$  e sua inversa, tal que:

$\mathbf{z}(\mathbf{x})^T[\mathbf{X}^T\mathbf{X}^{-1}]\mathbf{z}(\mathbf{x})\Sigma$				$[\mathbf{z}(\mathbf{x})^T[\mathbf{X}^T\mathbf{X}^{-1}]\mathbf{z}(\mathbf{x})\Sigma]^{-1}$			
47.235	0.010	66.420	-0.621	0.372	-486.480	-0.163	1.424
0.010	0.000	0.008	0.000	-486.480	922770.663	212.596	559.745
66.420	0.008	121.524	-0.855	-0.163	212.596	0.085	0.052
-0.621	0.000	-0.855	0.016	1.424	559.745	0.052	122.887

Considerando-se que as funções objetivo individuais de MAE, sMdAPE, SD e S(w) podem ser organizadas no vetor  $[\mathbf{Y}(\mathbf{x})]^T$  e seus ótimos individuais componham o vetor  $[\boldsymbol{\varphi}]^T$ , além de as diferenças entre esses dois vetores serem iguais a  $[\mathbf{Y}(\mathbf{x})]^T - [\boldsymbol{\varphi}]^T$ ; então, obtém-se:

	MAE	sMdAPE	SD	S(w)
$[\mathbf{Y}(\mathbf{x})]^T$	382.200	0.038	490.281	1.048
$[\boldsymbol{\varphi}]^T$	397.996	0.040	514.608	0.689
$[\mathbf{Y}(\mathbf{x}) - \boldsymbol{\varphi}]^T$	-15.797	-0.002	-24.328	0.359

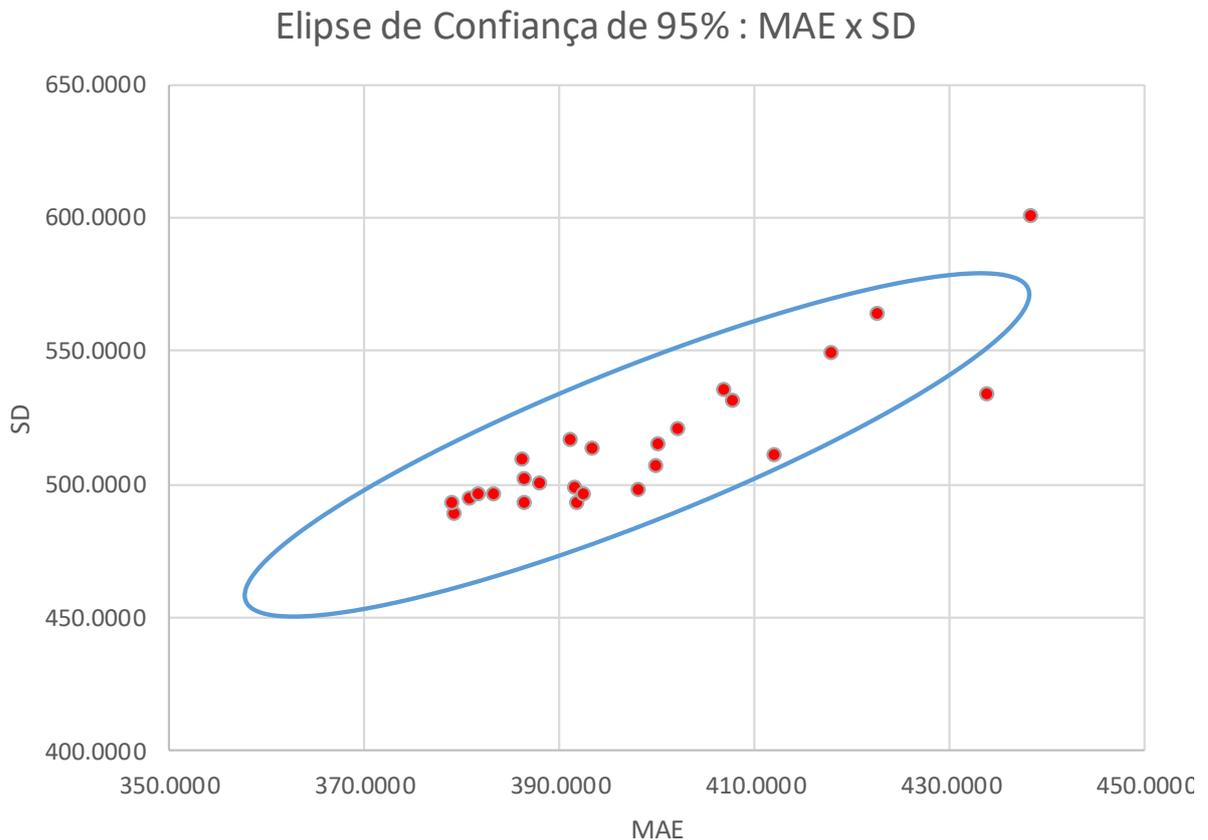
Pré-multiplicando-se  $[\mathbf{z}(\mathbf{x})^T[\mathbf{X}^T\mathbf{X}^{-1}]\mathbf{z}(\mathbf{x})\Sigma]^{-1}$  por  $[\mathbf{Y}(\mathbf{x}) - \boldsymbol{\varphi}]^T$ , obtém-se então:

$$[\mathbf{Y}(\mathbf{x}) - \boldsymbol{\varphi}]^T \{ [\mathbf{z}(\mathbf{x})^T[\mathbf{X}^T\mathbf{X}^{-1}]\mathbf{z}(\mathbf{x})\Sigma]^{-1} \} = \begin{bmatrix} -0.518 & 1045.594 & 0.154 & 19.287 \end{bmatrix}$$

Finalmente, multiplicando-se este resultado por  $[\mathbf{Y}(\mathbf{x}) - \boldsymbol{\varphi}]$ , chega-se ao valor da equação da elipse para um dado ponto x. No presente exemplo, o vetor de solução inclui uma distribuição de pesos ótimos igual a  $\mathbf{x}^T = [0.503, 0.271, 0.227]$  e produz um valor para elipse de 9.466. Este valor deve ser menor do que aquele relativo a uma estatística do Qui-quadrado para um nível de significância de 5% e p=4 (4 respostas/dimensões). Para esses parâmetros, o valor do Qui-quadrado de teste é  $\chi^2_{(0.95;4)} = 9.488$ . Logo, o peso ótimo define valores para MAE, sMdAPE, SD e S(w) dentro do hiperelipsóide quadridimensional definido pela equação (3.12).

O resultado da aplicação de restrição elíptica de  $\mathbf{Y}(\mathbf{x})$  ao problema de otimização multiobjetivo pode ser ilustrado (e entendido) construindo-se uma elipse de confiança para os dados. Neste exemplo, serão utilizadas as métricas MAE e SD, um vetor contendo suas médias, sua matriz de variância-covariância, com uma matriz diagonal ( $\Lambda$ ) contendo os respectivos autovalores e uma matriz ortogonal P, contendo os respectivos autovetores. Compõem, ainda, a equação (3.4) uma estatística de Qui-quadrado  $\chi^2_{(2,0.95)}$  (para duas variáveis e uma região de

confiança de 95%) e um vetor trigonométrico,  $\mathfrak{Z}^T = [\cos\alpha \ \text{sen}\alpha]$ , utilizado para a construção gráfica da função elíptica. No sistema de equações (3.4), o ângulo  $\alpha$  representa a rotação de um ponto  $\mathbf{P}(x, y)$  qualquer na elipse, no domínio  $0 \leq \alpha \leq 2\pi$ , servindo para construí-la para todo o espaço de solução. Dessa forma, obtém-se a elipse de confiança da figura a seguir:



**Figura 4-7 - Elipse de confiança de 95% para as métricas MAE e SD.  
Fonte: Próprio autor**

Observa-se claramente que alguns pontos experimentais saem da região elíptica de confiança. Quando tais pontos são considerados na modelagem dos polinômios canônicos, seus destacados valores podem enviesar os coeficientes dos modelos de MAE e SD. Quando otimizados individualmente, por sua vez, tais otimizações podem conduzir o método NBI a definir pontos de ancoragem fora da região de confiança dos dados. Para se evitar tal efeito e para melhorar a performance do NBI frente às métricas correlacionadas e com valores extremos, é que se justifica a utilização da equação elíptica descrita em detalhe neste item.

Além da restrição elíptica relativa à região multivariada de confiança para o vetor de respostas  $\mathbf{Y}(\mathbf{x})$ , aplicou-se também a restrição elíptica definida por Cornell (2002), descrita pela equação (3.13). Esta restrição é utilizada com a finalidade de afastar as soluções ótimas dos

vértices da região simplex. Para o presente problema-exemplo, os valores dos parâmetros da equação (3.13) são, respectivamente, iguais a:

**Tabela 4-14 - Valores utilizados na segunda restrição elíptica acoplada ao método NBI.**

Item	LB	w <sub>i</sub>	UB	w <sub>0i</sub>	Equação (3.13)	
g <sub>2</sub> (w)	0.070	0.503	0.550	0.333	0.12423	$g_2(\mathbf{x}) = \sum_{i=1}^q \left( \frac{w_i - w_{0i}}{\delta_i} \right)^2 \leq 1$
g <sub>3</sub> (w)	0.070	0.271	0.850	0.333	0.00640	
g <sub>4</sub> (w)	0.050	0.227	0.400	0.333	0.09308	
				1.000	0.22371	

A aplicação de equações semelhantes à equação (4.1) às funções MAE, sMdAPE, SD, S(w) e seus respectivos polinômios canônicos para escores rotacionados de fator, f<sub>1</sub>(x) e f<sub>2</sub>(x), do problema-teste conduz aos resultados descritos na **Tabela 4-15**. A diagonal principal dessa tabela (com valores destacados em azul) representa o vetor de ótimos individuais obtidos com a otimização de cada resposta de interesse, as respectivas soluções ótimas individuais (w<sub>1</sub>, w<sub>2</sub> e w<sub>3</sub>) e sua respectiva métrica de entropia, S(w), mostrando qual é o nível de diversificação das soluções encontradas em cada caso. Por questão de comparação, as três últimas linhas da tabela apresentam os resultados das MPP's obtidas com a utilização dos métodos de previsão individuais.

**Tabela 4-15 - Coeficientes das funções objetivo utilizadas no método NBI.**

MAE	SD	sMdAPE	S(w)	F <sub>1</sub>	F <sub>2</sub>	w <sub>1</sub>	w <sub>2</sub>	w <sub>3</sub>
<b>379.335</b>	491.819	0.03466	1.022	-0.7393	-0.6515	0.267	0.526	0.207
380.840	<b>488.987</b>	0.03644	1.050	-0.8637	-0.3283	0.417	0.410	0.173
380.499	495.753	<b>0.03436</b>	0.881	-0.5903	-0.7616	0.178	0.635	0.187
380.369	489.247	0.03663	<b>1.084</b>	-0.8762	-0.2758	0.424	0.362	0.214
381.583	489.817	0.03782	1.064	<b>-0.8863</b>	-0.0758	0.480	0.298	0.222
381.689	498.639	0.03445	0.773	-0.4891	<b>-0.7895</b>	0.131	0.697	0.172
433.745	534.080	0.05439	0.053	0.6023	2.7823	<b>1.000</b>	0.000	0.000
393.235	513.699	0.03757	0.053	0.0082	-0.5286	0.000	<b>1.000</b>	0.000
438.253	600.987	0.04239	0.053	3.2469	-0.2970	0.000	0.000	<b>1.000</b>

**14.c.** Os resultados das otimizações individuais descritas no item anterior conduzem a alvos individuais para F<sub>1</sub> e F<sub>2</sub>, respectivamente, iguais a -0.8863 e -0.7895 (**Tabela 4-15**), assim como

suas respectivas variâncias são  $\text{Var}_1=1.728$  e  $\text{Var}_2=1.238$  (esses resultados são mostrados na análise fatorial (**Tabela 4-11**). A otimização individual é, então, repetida para as funções  $\text{FMSE}_i$ , e seus resultados formam a matriz *Payoff* relativa ao problema NBI em questão. Esta matriz é descrita a seguir.

**Tabela 4-16 - Matriz Payoff para as funções FMSE que serão acopladas ao método NBI.**

<b>1.72780</b>	1.74930
1.88572	<b>1.23830</b>

Cada linha da matriz *Payoff* está associada a uma função objetivo, no caso, a cada um dos  $\text{FMSE}_i$ . Os valores destacados na diagonal principal da matriz *Payoff* representam os resultados das otimizações individuais das funções  $\text{FMSE}_i$  e são denominados como “Pontos de Utopia”; já os elementos da diagonal secundária, nesse caso, representam os piores valores admitidos para as funções dentro das restrições estabelecidas. Este vetor de soluções ruins denomina-se “Ponto de Nadir”. Essas informações integram a formulação (4.2) a seguir.

$$\begin{aligned}
 \underset{\mathbf{x} \in \Omega}{\text{Min}} \quad & \bar{f}_{(Rot)}(\mathbf{x}) = \left[ \frac{\text{FMSE}_{(g_1)}(\mathbf{x}) - \text{FMSE}_{(g_1)}^U}{\text{FMSE}_{(g_1)}^N - \text{FMSE}_{(g_1)}^U} \right] \\
 \text{S.t.} : \quad & \bar{g}_{1(Rot)}(\mathbf{x}) = \left[ \frac{\text{FMSE}_{(g_1)}(\mathbf{x}) - \text{FMSE}_{(g_1)}^U}{\text{FMSE}_{(g_1)}^N - \text{FMSE}_{(g_1)}^U} \right] - \left[ \frac{\text{FMSE}_{(g_2)}(\mathbf{x}) - \text{FMSE}_{(g_2)}^U}{\text{FMSE}_{(g_2)}^N - \hat{F}_{Rot(g_2)}^U} \right] + 2\beta_i - 1 = 0 \\
 & g_2(\mathbf{x}) = [\hat{\mathbf{Y}}(\mathbf{x}) - \boldsymbol{\varphi}]^T \left\{ \mathbf{z}^T(\mathbf{x}) (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{z}^T(\mathbf{x}) \hat{\boldsymbol{\Sigma}} \right\}^{-1} [\hat{\mathbf{Y}}(\mathbf{x}) - \boldsymbol{\varphi}] \leq \chi^2_{(p,\alpha)} \quad (4.2) \\
 & g_3(\mathbf{x}) = \sum_{i=1}^q \left( \frac{w_i - w_{0i}}{\delta_i} \right)^2 \leq 1 \\
 & g_4(\mathbf{x}) = \sum_{i=1}^q w_i = 1 \\
 & \text{FMSE}^i = \left[ \hat{F}_{Rot(g_1)}(\mathbf{x}) - T_i \right]^2 + \lambda_i \\
 & 0 \leq \beta_i \leq 1
 \end{aligned}$$

No sistema de equações (4.2), denota-se por  $\text{FMSE}_{(g_1)}(\mathbf{x})$  a superfície de resposta de mistura relativa ao Fator 1, bem como  $\text{FMSE}_{(g_1)}^U(\mathbf{x})$  e  $\text{FMSE}_{(g_1)}^N(\mathbf{x})$  como seus respectivos pontos de Utopia e Nadir, assim como  $\text{FMSE}_{(g_2)}^U(\mathbf{x})$  e  $\text{FMSE}_{(g_2)}^N(\mathbf{x})$ , os respectivos pontos de

Utopia e Nadir para a superfície de resposta de mistura relativa ao Fator 2,  $FMSE_{(g_2)}(\mathbf{x})$ . Para um problema de natureza bidimensional, o NBI é executado iterativa e recursivamente, acarretando os resultados de 21 subproblemas descritos na **Tabela 4-17** e na **figura 4-8** a seguir.

**Tabela 4-17 - Resultados do método NBI-FA com filtros de Jolliffe e restrições elípticas.**

<b>W</b>	<b>MAE</b>	<b>SD</b>	<b>sMdAPE</b>	<b>f<sub>1</sub></b>	<b>f<sub>2</sub></b>	<b>w<sub>1</sub></b>	<b>w<sub>2</sub></b>	<b>w<sub>3</sub></b>	<b>S(w)</b>	<b>d</b>
0.00	380.26	491.35	0.03480	1.019	0.019	0.278	0.573	0.149	0.949	947.53
0.05	380.17	491.18	0.03482	0.923	0.023	0.282	0.566	0.152	0.958	882.55
0.10	380.08	491.02	0.03485	0.827	0.027	0.287	0.558	0.155	0.968	815.64
0.15	380.00	490.85	0.03488	0.732	0.032	0.292	0.549	0.159	0.978	746.29
0.20	379.92	490.67	0.03492	0.638	0.038	0.298	0.540	0.162	0.988	674.58
0.25	379.85	490.48	0.03497	0.546	0.046	0.305	0.530	0.166	0.998	600.05
0.30	379.78	490.29	0.03503	0.455	0.055	0.312	0.519	0.169	1.009	523.20
0.35	379.73	490.09	0.03511	0.368	0.068	0.320	0.506	0.174	1.020	443.75
0.40	379.69	489.89	0.03521	0.284	0.084	0.330	0.493	0.178	1.032	362.18
0.45	379.68	489.68	0.03533	0.206	0.106	0.341	0.477	0.183	1.044	280.15
0.50	379.70	489.49	0.03549	0.136	0.136	0.354	0.458	0.188	1.055	200.94
0.55	379.77	489.31	0.03571	0.079	0.179	0.370	0.437	0.193	1.066	131.27
0.60	379.92	489.19	0.03598	0.039	0.239	0.388	0.414	0.199	1.075	82.59
<b>0.65</b>	<b>380.14</b>	<b>489.15</b>	<b>0.03629</b>	<b>0.015</b>	<b>0.315</b>	<b>0.406</b>	<b>0.389</b>	<b>0.204</b>	<b>1.080</b>	<b>65.79</b>
0.70	380.41	489.19	0.03663	0.005	0.405	0.425	0.366	0.209	1.081	83.04
0.75	380.71	489.30	0.03696	0.001	0.501	0.441	0.345	0.213	1.079	127.19
0.80	381.02	489.46	0.03727	0.000	0.600	0.456	0.327	0.217	1.075	189.35
0.85	381.32	489.64	0.03757	0.000	0.700	0.470	0.311	0.220	1.069	262.94
0.90	381.62	489.84	0.03785	0.000	0.800	0.481	0.296	0.222	1.063	344.01
0.95	381.91	490.06	0.03812	0.000	0.900	0.492	0.283	0.225	1.056	429.80
1.00	382.20	490.28	0.03838	0.000	1.000	0.503	0.271	0.227	1.048	519.20

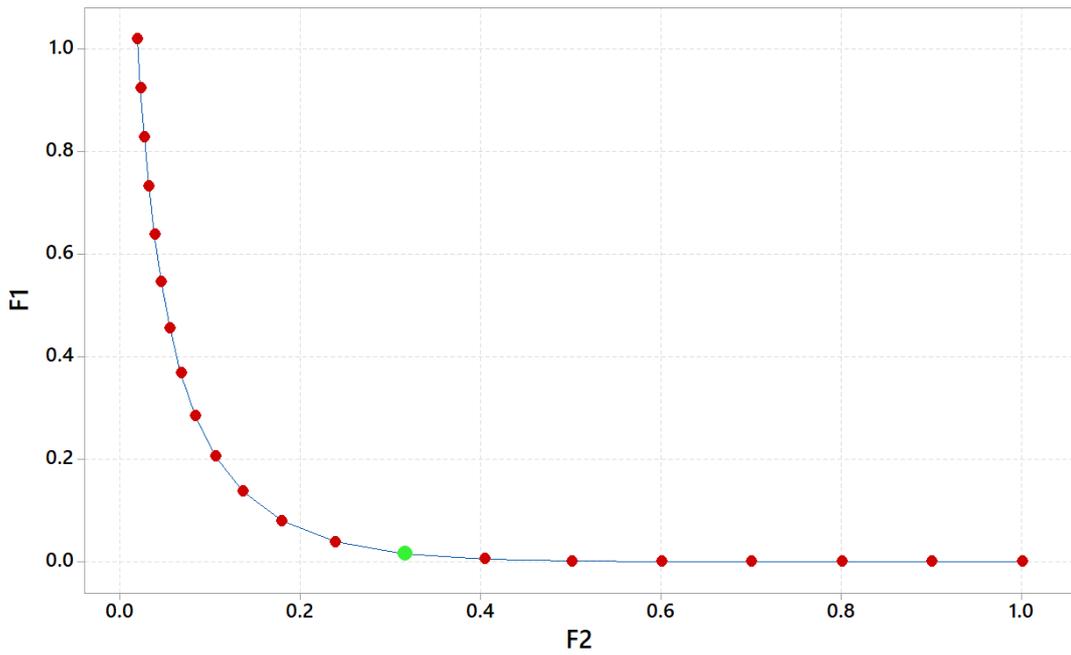


Figura 4-8 - Fronteira de Pareto para os resultados do método NBI-FA com filtros de Jolliffe e restrições elípticas.  
 Fonte: Próprio autor

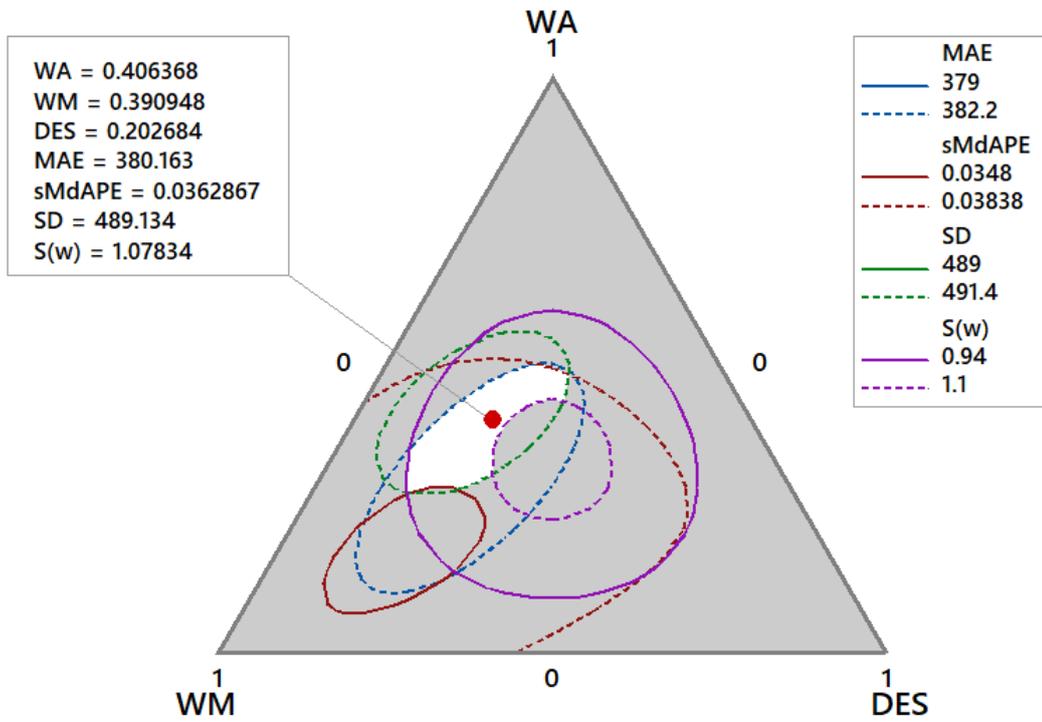
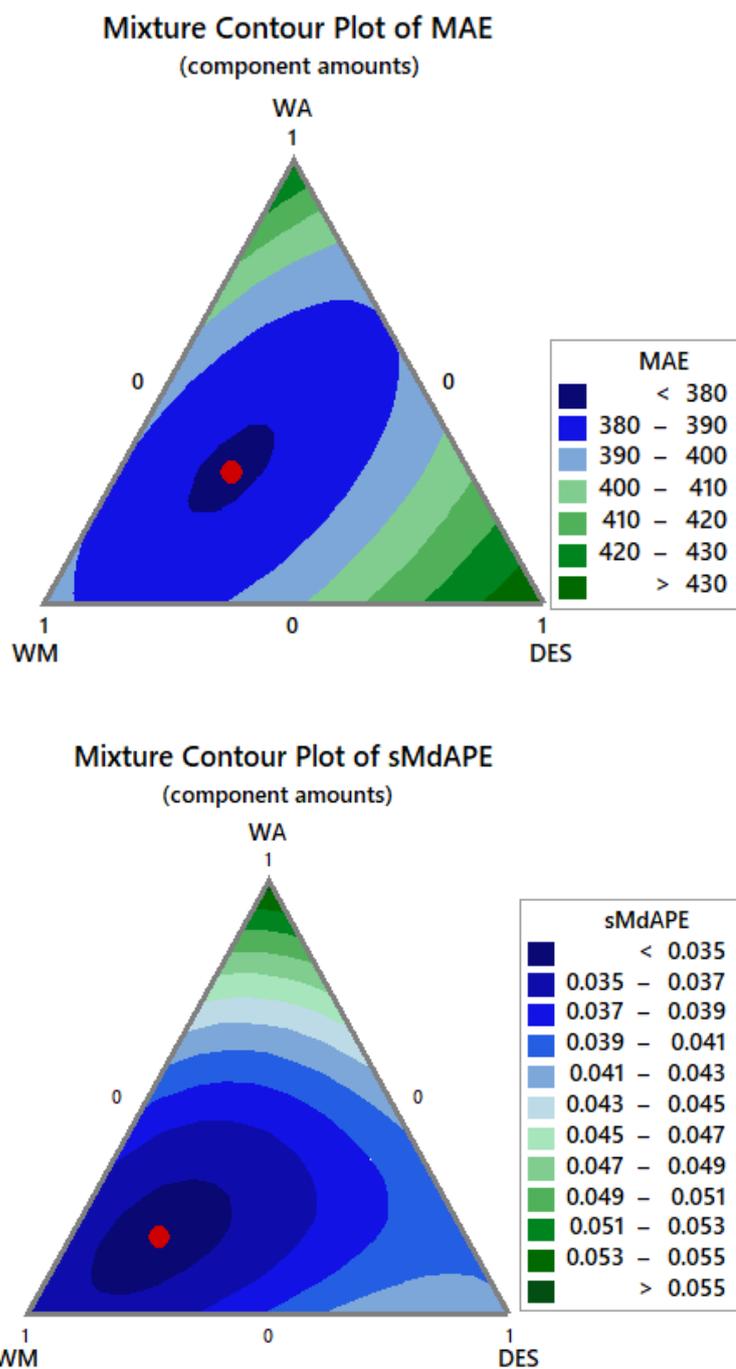


Figura 4-9 - Simplex Overlaid Plot: superfícies de resposta sobrepostas e peso ótimo.  
 Fonte: Próprio autor



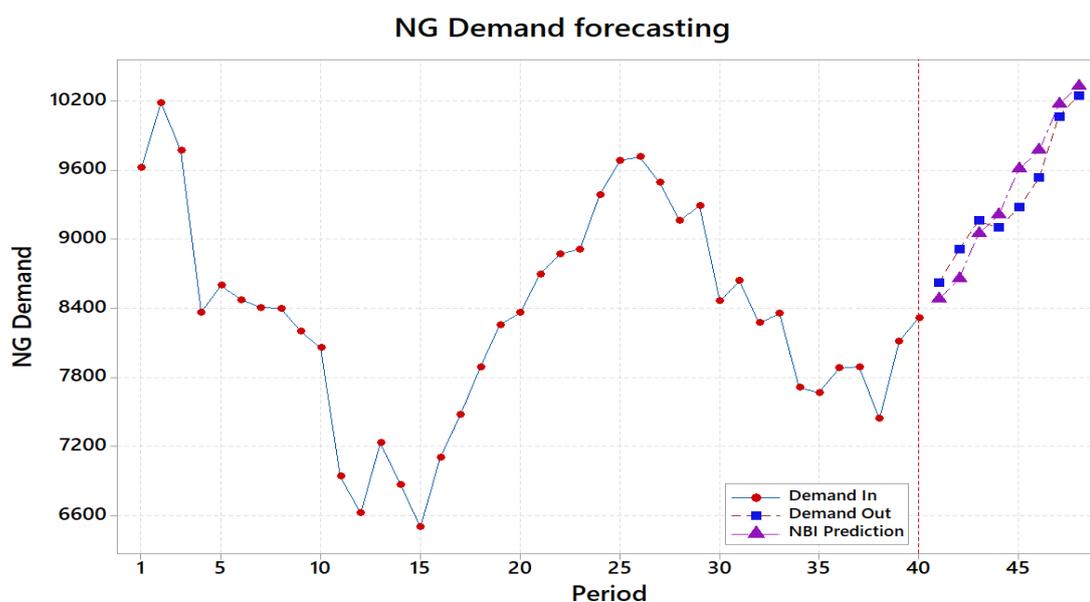
**Figura 4-10 - MDE surface plot para MAE e sMdAPE.**  
Fonte: Próprio autor

**Passo 15:** Este passo consiste na escolha da melhor entre todas as soluções apresentadas pela Fronteira de Pareto. De acordo com a máxima entropia  $S(w)$  da combinação e a menor distância de Mahalanobis da solução em relação às soluções individuais (Tabela 4.17), a melhor solução foi obtida no vetor de pesos  $\mathbf{x}^T = [0.406, 0.389, 0.204]$ , com a maior entropia ( $S=1.080$ ) e a

menor distância de Mahalanobis ( $d=65.79$ ) dentre as 21 soluções obtidas. Essa solução será testada posteriormente com uma amostra *out-of-sample*.

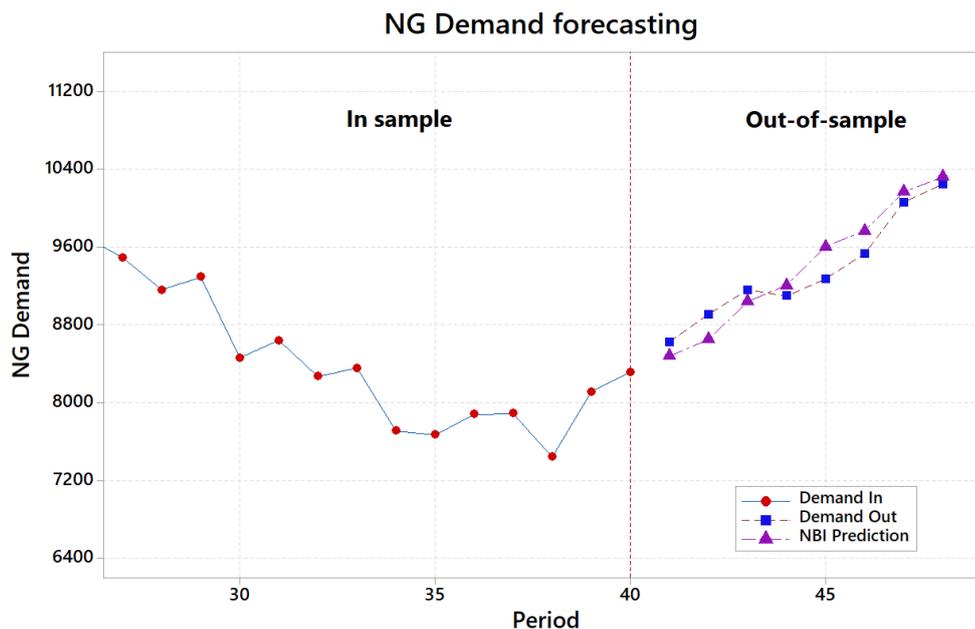
### 4.3 QUALIDADE DA PREVISÃO OBTIDA

Para avaliar a qualidade da solução obtida com os pesos ótimos identificados com o método NBI fatorial com restrições elípticas, aplicou-se o vetor de pesos de pesos ótimos  $w=[0.406, 0.389, 0.204]$  aos resíduos de previsão dos métodos individuais de Winter's aditivo (WA), Winter's multiplicativo (WM) e alisamento exponencial duplo (DES). Com essa recomposição de resíduos, os valores previstos encontrados foram muito próximos aos valores reais de teste (*out-of-sample*), tal como se pode verificar nas **Figuras 4.11 e 4.12** a seguir.



**Figura 4-11 - Resultados das previsões para o conjunto de teste (out-of-sample).**  
**Fonte: Próprio autor**

Comparando-se também todas as métricas de performance de previsão calculadas para o conjunto de teste (*out-of-sample*) com os métodos individuais, com combinações obtidas pelo método FA-NBI, com combinações obtidas com FA-NBI modificado pelos critérios de Jolliffe e o método NBI com restrições elípticas aplicado às métricas selecionadas pelo filtro de Jolliffe Fatorial, observa-se que o método proposto conduz à determinação de um vetor de pesos que supera todas as opções apresentadas. Esses resultados são apresentados nas **Tabelas 4.18 e 4.19** a seguir.



**Figura 4-12 - Resultados das previsões para o conjunto de teste (out-of-sample) (Série seccionada).**  
**Fonte: Próprio autor**

**Tabela 4-18 - Comparações entre métodos.**

	<b>WMA3</b>	<b>WM3</b>	<b>DES2</b>	<b>FA-NBI</b>	<b>NBI Elíptico</b>
MAE	397.75	429.44	274.95	216.26	172.33
MSE	178267.04	259645.01	102305.11	59261.94	36839.05
RMSE	422.22	509.55	319.85	243.44	191.94
MASE	1.44663	1.56191	1.00000	0.78657	0.62679
RMSPE	4.45856	5.29106	3.34970	2.58458	2.08233
MAPE	4.22668	4.47894	2.89246	2.29733	1.86150
sMAPE	0.04329	0.04344	0.02841	0.02273	0.01856
U1	0.02300	0.02656	0.01685	0.01289	0.01021
U2	0.04503	0.05434	0.03411	0.02596	0.02047
VAR	22932.50	98272.15	65115.03	49407.53	39882.89
SD	151.43	313.48	255.18	222.28	199.71
MdAE	392.57	528.61	300.39	206.48	128.12
MdAPE	4.4007	5.4246	3.1564	2.1373	1.4470
sMdAPE	0.04501	0.05279	0.03107	0.02115	0.01458

**Tabela 4-19 - Comparação do método proposto com o método NBI-FA modificado pelos critérios de Jolliffe.**

	<b>1º Regra de Jolliffe</b>	<b>2º Regra de Jolliffe</b>	<b>3º Regra de Jolliffe</b>	<b>NBI Elíptico</b>
MAE	240.16	200.61	256.10	172.33
MSE	78895.14	49404.12	90615.59	36839.05
RMSE	280.88	222.27	301.02	191.94
MASE	0.87349	0.72965	0.93147	0.62679
RMSPE	2.95903	2.37570	3.16102	2.08233
MAPE	2.53339	2.14226	2.69479	1.86150
sMAPE	0.02497	0.02126	0.02651	0.01856
U1	0.01483	0.01179	0.01588	0.01021
U2	0.02995	0.02370	0.03210	0.02047
VAR	55790.08	45737.47	59057.41	39882.89
SD	236.20	213.86	243.02	199.71
MdAE	244.09	171.35	279.68	128.12
MdAPE	2.5179	1.7791	2.8828	1.4470
sMdAPE	0.02487	0.01763	0.02842	0.01458

Como mostram os resultados apresentados anteriormente, o método NBI-FA, incrementado com restrições elípticas e com filtros de Jolliffe para as métricas de performance e de precisão de previsões, superou todos os métodos convencionais de previsão, todos os métodos individuais, o método NBI-FA tradicional e também o método NBI-FA modificado pelas três regras de Jolliffe, o que sugere sua boa adequação como estratégia de definição de pesos ótimos para a combinação de métodos de previsão.

Embora os resultados obtidos sejam relativos, apenas, ao problema-exemplo utilizado nesta tese, pode-se esperar que sua aplicabilidade se estenda a uma diversidade de casos, podendo variar quanto aos inúmeros métodos disponíveis que podem ser combinados ou em relação aos tipos de séries temporais, as quais necessitem de uma previsão.

## 5 CONCLUSÕES

O presente trabalho mostrou o desenvolvimento, a aplicação e o teste do Método de otimização NBI-Payoff-Jolliffe Fatorial com restrições elípticas para a determinação de uma combinação ótima de métodos de previsão para a estimação da demanda de gás natural. Em resumo, a estratégia desenvolvida aperfeiçoa o método de otimização não linear multiobjetivo NBI para ser utilizado com funções objetivo formadas por escores rotacionados de fator. Estes escores foram obtidos com um conjunto reduzido de métricas filtradas a partir de uma matriz original de 14 métricas de performance e de precisão de previsão. Essas métricas foram filtradas de acordo com o segundo critério de Jolliffe, o qual utiliza uma Análise de Componentes Principais (PCA) para identificar e para eliminar variáveis redundantes dentro do conjunto multivariado de interesse.

Para se modelar os escores representativos das métricas não redundantes resultantes da filtragem em função dos pesos utilizados para se combinar os resíduos dos métodos de previsão individuais que devem ser combinados, foram utilizados polinômios canônicos de misturas do tipo {3, 5}. Um grupo de restrições elípticas relacionadas às respostas de interesse e referentes ao espaço de solução simplex foram acopladas ao método NBI. Executando-se o método, 21 soluções Pareto-ótimas, das quais se escolheu a melhor solução pelo critério da máxima entropia (maior diversificação dos pesos) e a menor distância de Mahalanobis – que é uma métrica que considera a distância entre os resultados obtidos com o método multiobjetivo e as soluções individuais (de forma multivariada). O ponto de ótimo escolhido no problema utilizado como exemplo foi  $\mathbf{x}^{T*}=[0.406, 0.389, 0.204]$ , pesos estes que devem ser aplicados, respectivamente, aos resíduos dos métodos Holt-Winters aditivo (WA), Holt-Winters multiplicativo (WM) e ao método de Alisamento Exponencial Duplo (DES).

Comparando-se os resultados obtidos com os resultados apresentados pelos métodos individuais, combinados, de NBI-FA ou NBI-FA modificados pelos três critérios de Jolliffe, observou-se uma excelente performance da combinação proposta, o que sugere que o método é adequado para esse tipo de problema e pode ser, doravante, melhor explorado em outros contextos, com outros tipos de séries históricas e com outros tipos de métodos de previsão para compor as combinações.

Algumas conclusões específicas podem também ser destacadas, a saber:

- Foi evidenciado que, verdadeiramente, há uma combinação ótima de pesos que conduz a uma performance de previsão muito superior àquela permitida por métodos individuais, por melhor desempenho que eles possam apresentar;
- Há uma diversidade de métricas de qualidade de previsão disponíveis; algumas são redundantes, outras são, apenas, correlacionadas. Evidenciou-se que a performance de previsão das combinações melhora quando a escolha dos pesos se baseia em escores rotacionados de métricas não redundantes;
- As melhores métricas para o caso estudado nessa tese foram: MAE, sMdAPE, SD e  $S(w)$ ;
- Métricas correlacionadas podem ser usadas para a tomada de decisão, porém elas devem apresentar pontos de ótimos individuais diferentes para que sejam selecionáveis;
- O segundo critério de Jolliffe se mostrou o mais eficiente para a seleção das métricas mais adequadas de avaliação da qualidade de previsão;
- As restrições elípticas exerceram uma forte influência positiva sobre os resultados da otimização multiobjetivo, promovendo a correta execução do algoritmo de gradiente reduzido generalizado (GRG) em cada uma de suas iterações;
- Os polinômios canônicos de misturas derivados das matrizes experimental e de respostas apresentaram excelentes ajustes e colaboraram para os bons resultados obtidos com o método de otimização multiobjetivo;
- A filtragem de métricas baseada na aplicação da Análise Hierárquica de Cluster aos dados da matriz Payoff mostrou-se muito eficaz para a obtenção dos bons resultados. Recomenda-se a utilização desse procedimento também para outros tipos de problemas abordados por métodos de otimização multiobjetivos;
- Pesos ótimos foram obtidos em segmentos interiores da região Simplex, afastando-se consideravelmente dos vértices. Esta conclusão aponta para a força da combinação em relação aos melhores métodos individuais. Tais soluções também apresentaram um alto nível de diversificação, quantificado pela função de entropia;
- A utilização concomitante da distância de Mahalanobis com a entropia foi muito útil para a adequada escolha da melhor solução disponível no conjunto Pareto-ótimo;
- A Fronteira de Pareto para o método NBI modificado apresentou excelentes características, mantendo-se equiespaçada e convexa em toda a sua extensão;
- O resultado da previsão para o consumo de gás natural no setor industrial dos EUA para uma amostra de teste não utilizada (*out-of-sample*) nas fases de construção e de

avaliação do método proposto mostrou-se excelente, com os menores níveis de métricas em comparação a todos os métodos estudados, combinados ou não.

## 6 REFERÊNCIAS BIBLIOGRÁFICAS

- AHMADI, A.; KAYMANESH, A.; SIANO, P.; JANGHORBANI, M.; NEZHAD, A. E.; SARNO, D. (2015). Evaluating the effectiveness of normal boundary intersection method for short-term environmental/economic hydrothermal self-scheduling: **Electrical Power Systems Research**, 123, 192-204.
- AKPINAR, M.; YUMUSAK, N. (2016). Year ahead demand forecast of city natural gas using seasonal time series method. **Energies**, 9, 727.
- AKTUNC, E.A.; YUKSELTAN, E.; YUCEKAYA, A.; BILGE, A.H.(2020). Managing natural gas demand for free consumers under uncertainty and limited storage capacity. **Journal of Natural Gas Science and Engineering**, 79, 103322.
- AL-GHANDOOR, A. (2013). An approach to energy savings and improved environmental impact through restructuring Jordan's transport sector. **Renewable and Sustainable Energy Reviews**, 18, 31-42.
- ALMEIDA, F. A.; LEITE, R. R.; GOMES, G. F.; GOMES, J. H. F.; PAIVA, A. P. (2019). Multivariate data quality assessment based on rotated factor scores and confidence ellipsoids. **Decision Support Systems**, 129, 113173.
- AQUILA, G.; ROCHA, L. C. S.; PAMPLONA, E. O.; QUEIROZ, A. R.; JÚNIOR, P. R.; BALESTRASSI, P. P.; FONSECA, M. N. (2018). Proposed method for contracting of wind-photovoltaic projects connected to the Brazilian electric system using multiobjective programming, **Renewable and Sustainable Energy Reviews**, 97, 377-389.
- ANDERSEN, S; GREFSRUD, E. S.; HARBOE, T. Effect of increased pCO<sub>2</sub> level on early shell development in great scallop (*Pecten maximus* Lamarck) larvae. **Biogeosciences**, 10 (10), 6161-6184.
- ARMSTRONG, J. S. (1986) The Ombudsman: Research on Forecasting: A Quarter-Century Review, 1960–1984. **Inform Journal on Applied Analytics** 16 (1) 89-109
- ARMSTRONG, J. S.; COLLOPY, F. (1992). Error measures for generalizing about forecasting methods: Empirical comparisons. **International Journal of Forecasting**, 8, 69-80.
- ARMSTRONG, J. (2001). **Principles of Forecasting: a handbook for researches and practitioners**. (1th ed.). New York: Kluwer Academic Publishers.

AZADEH, A. SABERI, M., KAZEM, A., EBRAHIMIPOUR, V., NOURMOHAMMADZADEH, A., SABERI, Z. (2011). A flexible algorithm for fault diagnosis in a centrifugal pump with corrupted data and noise based on ANN and support vector machine with hyper-parameters optimization. **Applied Soft Computing**, 13, Issue 3, 1478-1485,

AZADEH, A.; BABAZADEH, R., ASADZADEH, S.M., (2013). Optimum estimation and forecasting of renewable energy consumption by artificial neural networks, **Renewable and Sustainable Energy Reviews**, 27, 605-612.

AZADEH, A, ASADZADEH S.M., SABERI, M., NADIMI, V., TAJVIDI, A., SHEIKALISHAHI, M., (2011). A Neuro-fuzzy-stochastic frontier analysis approach for long-term natural gas consumption forecasting and behavior analysis: The cases of Bahrain, Saudi Arabia, Syria, and UAE, **Applied Energy**, 88, Issue 11, 3850-3859,

BACCI, L. A.; MELLO, L. G.; INCERTI, T.; PAIVA, A. P.; BALESTRASSI, P. P. (2019). Optimization of combined time series methods to forecast the demand for coffee in Brazil: A new approach using Normal Boundary Intersection coupled with mixture designs of experiments and rotated factor scores. **International Journal of Production Economics**, 212, 186–211.

BALDACCI, L.; GOLFARELLI, M.; LOMBARDI, D.; SAMI, F. (2016). Natural gas consumption forecasting for anomaly detection. **Expert Systems With Applications**, 62, 190–201.

BALESTRASSI, P.P., POPOVA, E., PAIVA, A.P., MARANGON LIMA, J.W., (2009). Design of experiments on neural network's training for nonlinear time series forecasting, *Neurocomputing*, 72, Issues 4–6, 1160-1178,

BATES, J. M.; GRANGER, C. W. J. (1969). The Combination of Forecasts. **Operational Research Society**, 20, 451-468.

BATZIAS F. A.; NIKOLAU, N. P.; KAKOS, A. S. (2006). Modelling the natural gas consumption in a changing environment. **Journal of Computational Methods in Sciences and Engineering**, 6, 339–352.

BEYCA, O. F.; ERVURAL, B. C.; TAGOGLU, E.; OZUYAR, P. G.; ZAIM, S. (2019). Using machine learning tools for forecasting natural gas consumption in the province of Istanbul. **Energy Economics**, 80, 937-949.

BIANCO, V.; SCARPA, F.; TAGLIAFICO, L. A. (2014). Scenario analysis of nonresidential natural gas consumption in Italy. **Applied Energy**, 113, 392–403.

BLIEMEL, F. (1973). Theil's Forecast Accuracy Coefficient: A Clarification. **Journal of Marketing Research**. 1973;10(4):444-446.

BORDIGNON, S.; BUNN, D. W.; LISI, F.; NAN, F. (2013). Combining day-ahead forecasts for British electricity prices, **Energy Economics**, 35, 88-103,

BRITO, T. G.; PAIVA, A. P.; FERREIRA, J. R.; GOMES, J.H.F.; BALESTRASSI, P. P. (2014). A normal boundary intersection approach to multiresponse robust optimization of the surface roughness in end milling process with combined arrays: **Precision Engineering**, 38, 628-638.

BÜYÜKSAHİN, Ü. Ç. ERTEKİN, S. (2019). Improving forecasting accuracy of time series data using a new ARIMA-ANN hybrid method and empirical mode decomposition. **Neurocomputing**, 361, 151–163.

CALASAN, M.; ALEEM, S. H. E.; ZOBAA, A. F. (2020). On the root mean square error (RMSE) calculation for parameter estimation of photovoltaic models: A novel exact analytical solution based on Lambert W function. **Energy Conversion and Management**, 210, 112716.

CALERO, L. A.; FLOREZ, K. G.; RUFFO, J. S.; PALOMIRO, E. R. (2019) Improvement proposal to raise service level in a cosmetics retail company IEEE Xplore. **Congreso Internacional de Innovación y Tendencias en Ingeniería (CONIITI)**, 1-5.

CAO, Y.; NISHIHARA, R.; WU, K.; WANG, M.; OGINO, S.; WILLET, W. C.; SPIEGELMAN, D.; FUCHS, C. S.; GIOVANUCCI, E. L.; CHAN, A. T.; (2016) Population-wide Impact of Long-term Use of Aspirin and the Risk for Cancer. **JAMA Oncol**. 2(6):762–769

CHAMBERS, J.C. MULLICK, S.K. SMITH, D.D. (1971). How do Choose the Right Forecasting Technique. **Harvard Business Review**, 49, 45-57.

CHAN, C. K.; KINGSMAN, B. G.; WONG, H. (1999). The value of combining forecasts in inventory management – a case study in banking. **European Journal of Operational Research**, 117, 199-210.

CHENG, Y.; CHUA, W. S.; KOCH, T. (2018). Forecasting day-ahead high-resolution natural-gas demand and supply in Germany. **Applied Energy**, 228, 1091-1110.

CHIEN, C.; LIN, Y.; LIN, S. (2020) Deep reinforcement learning for selecting demand forecast models to empower Industry 3.5 and an empirical study for a semiconductor component distributor. **International Journal of Production Research**, 58, 2784-2804.

CHUNG, M. G.; KIM, S. (2013). Efficient jitter compensation using double exponential smoothing. **Information Sciences**, 227, 83-89.

COETZER, R.; HAINES, L. M. (2017). The construction of D- and I-optimal designs for mixture experiments with linear constraints on the components. **Chemometrics and Intelligent Laboratory Systems**, 171, 112–124.

CONFLITTI, C.; DE MOL, C.; GIANNONE, D. (2015) Optimal combination of survey forecasts, **International Journal of Forecasting**, 31, 1096-1103.

CORNELL, J. A. (2002). **Experiments with Mixtures: designs, models and the analysis of the mixture data**. (3th. ed.). New York: John Wiley & Sons.

CORONADO, M.; SEGADÃES, A. M.; ANDRÉS, A. (2015). Using mixture design of experiments to assess the environmental impact of clay-based structural ceramics containing foundry wastes. **Journal of Hazardous Materials**, 299, 529–539.

COSTA, D. M. D.; BRITO, T. G.; PAIVA, A. P.; LEME, R. C. BALESTRASSI, P. P. (2016). A normal boundary intersection with multivariate mean square error approach for dry end milling process optimization of the AISI 1045 steel. **Journal of Cleaner Production**, 135, 1658-1672.

CUMMING, J. A.; WOOFF, D. A. (2007). Dimension reduction via principal variables. **Computational Statistics & Data Analysis**, 52, 550 – 565.

DAI, P.; DING, X; WANG, Q. (2016). Dimension reduction based linear surrogate variable approach for model free variable selection. **Journal of Statistical Planning and Inference**, 169, 13–26.

DANTAS, T. M.; OLIVEIRA, F. L.; REPOLHO, M. V. (2017). Air transportation demand forecast through Bagging Holt Winters methods. **Journal of Air Transport Management**, 59, 116-123.

DAS, I.; DENNIS, J. E. (1998). Normal-Boundary Intersection: a new method for generating the Pareto Surface in Nonlinear Multicriteria Optimization Problems. **Society for Industrial and Applied Mathematics**, 8, 631-657.

DEBNATH, K. B.; MOURSHED, M. (2018). Forecasting methods in energy planning models. **Renewable and Sustainable Energy Reviews**, 88, 297-325.

DICKINSON, J. P. (1973). Some Statistical Results in the Combination of Forecasts. **Operational Research Quarterly**. Vol. 24, n° 2, p.p. 253-260.

- DIEBOLD, F. X. (1988). Serial correlation and the combination of forecasts. **Journal of Business and Economic Statistics**, 6, 105-111.
- DOORGA, J. R. S.; DHURMEA, K. R.; RUGHOOPTH, S.; BOOJHAWON, R. (2019). Forecasting mesoscale distribution of surface solar irradiation using a proposed hybrid approach combining satellite remote sensing and time series models. **Renewable and Sustainable Energy Reviews**, 104, 69-85.
- DRAPER, C.; REICHLER, R.; JEU, R.; NAEIMI, V.; PARINUSSA, R.; WAGNER, W. (2013). Estimating root mean square errors in remotely sensed soil moisture over continental scale domains. **Remote Sensing of Environment**, 137, 288–298.
- DUANGNATE, K.; MJELDE, J. W. (2017). Comparison of data-rich and small-scale data time series models generating probabilistic forecasts: an application to U.S. natural gas gross withdrawals. **Energy Economics**, 65, 411-423.
- FANG, F.; YU, Z. (2020). Model averaging assisted sufficient dimension reduction. **Computational Statistics and Data Analysis**, 152, 106993.
- FRIAS-PAREDES, L. F.; MALLOR, F.; GÁSTON-ROMEO, M.; LEÓN, T. (2017). Assessing energy forecasting inaccuracy by simultaneously considering temporal and absolute errors. **Energy Conversion and Management**, 142, 533–546.
- FRIAS-PAREDES, L. F.; MALLOR, F.; GÁSTON-ROMEO, M.; LEÓN, T. (2018). Dynamic mean absolute error as new measure for assessing forecasting errors. **Energy Conversion and Management**, 162, 176-188.
- GAUDÊNCIO, J. H. D.; CORRÊA, J. E. C.; PAES, V. C.; CAMPOS, P. H. S. TURRIONI, J. B.; PAIVA, A. P. (2019). Hybrid multiobjective optimization algorithm based on multivariate mean square error and fuzzy decision maker. **Applied Soft Computing Journal**, 82, 105586.
- GAUDÊNCIO, J. H. D.; ALMEIDA, F. A.; TURRIONI, J. B.; BALESTRASSI, P. P.; PAIVA, A. P. (2019). A multiobjective optimization model for machining quality in the AISI 12L14 steel turning process using fuzzy multivariate mean square error. **Precision Engineering**, 2019, 56, 303–320.
- GHASEMI, Fahimeh et al. (2018). Neural network and deep-learning algorithms used in QSAR studies: merits and drawbacks. **Drug Discov. Today**, v. 23, n. 10, p. 1784-1790.

GOMES, J. H. F.; PAIVA, A. P.; COSTA, S. C.; BALESTRASSI, P. P.; PAIVA, E. J. (2013). Weighted Multivariate Mean Square Error for processes optimization: A case study on flux-cored arc welding for stainless steel claddings. **European Journal of Operational Research**, 226, 522-535.

GRANGER, C.W.J. and NEWBOLD, P. (1977) **Forecasting Economic Time Series**. Academic Press, New York. (1th. ed.). New York: Academic Press

GRANGER, C. W. J.; RAMANATHAN, R. (1984). Improved Methods of Combining Forecasts. **Journal of Forecasting**, 3, 197-204.

HAFEZI, R. AKHAVAN, A. N.; PAKSERESHT, S.; WOODD, D. A. (2021). Global natural gas demand to 2025: A learning scenario development model. **Energy**, 224, 120167.

HAIR Jr., J. F., BLACK, W. C.; BABIN, B. J.; ANDERSON, R. E. (2010). **Multivariate Data Analysis**, (7 th ed.). New York: Prentice Hall.

HANG, J.; GE, Z. (2020). Effect of dimensionality reduction on stock selection with cluster analysis in different market situations. **Expert Systems With Applications**, 147, 113226.

HASSAN, M. R.; NATH, B.; KIRLEY, M. (2007). A fusion model of HMM, ANN and GA for stock market forecasting. **Expert Systems with Applications**, 33, 171-180.

HIDALGO, D. R.; CORTÉS, B. B.; BRAVO, E. C. (2020). Dimensionality reduction of hyperspectral images of vegetation and crops based on self-organized maps. **Information Processing In Agriculture**.

HYNDMAN, R. J.; KOEHLER, A. B.; (2006). Another look at measures of forecast accuracy. **International Journal of Forecasting**, 22, 679– 688.

HOCHREITER, Sepp; MOZER, Michael C. Monaural separation and classification of mixed signals: (2001). A support-vector regression perspective. In: 3rd **International Conference on Independent Component Analysis and Blind Signal Separation**, San Diego, CA. 2001. p. 498-503.

HOTELLING, H. (1933). Analysis of a complex of statistical variables into principal components. **Journal of Educational Psychology**, 24(6), 417–441.

HUNTINGTON, H. G. (2007). Industrial natural gas consumption in the United States: an empirical model for evaluating future trend. **Energy Economics**, 29, 743-749.

ISLAM, M. T.; HUDA, N.; (2019). E-waste in Australia: Generation estimation and untapped material recovery and revenue potential. **Journal of Cleaner Production**, 237, 117787.

- JEONG, D.; KIM, Y. (2009). Combining single-value streamflow forecasts – A review and guidelines for selecting techniques. **Journal of Hydrology**, 377, 284-299.
- JOLLIFFE, I.T. (1972). Discarding variables in a principal component analysis I: artificial data. **Applied Statistics**, 21, 160–173.
- JOLLIFFE, I.T. (1973). Discarding variables in a principal component analysis II: Real data. **Applied Statistics**, 22, 21–31.
- JOLLIFFE, I. T. (2002). **Principal Component Analysis (2th ed.)**. New York, Springer-Verlag.
- JOHNSON, R.; WICHERN, D. W.; (2007). **Statistical Analysis. (6<sup>th</sup> ed.)**. New Jersey: Prentice Hall.
- KAISER, H. F. (1960). The application of electronic computers to factor analysis. **Educational and Psychological Measurement**, 20, 141-151.
- KHANAL, A., KHOSHGHADAM, M., LEE, W.J., NIKOLAOU, M., (2017). New forecasting method for liquid rich shale gas condensate reservoirs with data driven approach using principal component analysis. **J. Nat. Gas Sci. Eng.** 38, 621-637.
- KARABIBER, O. A.; XYDIS, G. (2020). Forecasting day-ahead natural gas demand in Denmark. **Journal of Natural Gas Science and Engineering**, 76, 103-193.
- KARIMI, S.; FARROKHNIYA, M. (2014). Leukemia and small round blue-cell tumor cancer detection using microarray gene expression data set: Combining data dimension reduction and variable selection technique. **Chemometrics and Intelligent Laboratory Systems**, 139, 6–14.
- KIM, S.; KIM, H. (2016). A new metric of absolute percentage error for intermittent demand forecasts. **International Journal of Forecasting**, 32, 669-679.
- KHURI, A. I.; CONLON, M. (1981). Simultaneous Optimization of Multiple Responses Represented by Polynomial Regression Functions. **Technometrics**, 23, 363-375.
- LAM, K. F.; MUI, H. W.; YUEN, H. K. (2001). A note on minimizing absolute percentage error in combined Forecasts. **Computers & Operations Research**, 28, 1141-1147.
- LAWRENCE, M.J. EDMUNDSON, R.H. O'CONNOR, M.J. (1986) The accuracy of combining judgemental and statistical forecasts. **Management Science**, 32 (12), pp. 1521-1532

LEITE, R. R. **Método de interseção normal a fronteira para modelos quadráticos de escores fatoriais rotacionados**. Dissertação (Mestrado em Engenharia de Produção) – Universidade Federal de Itajubá, Itajubá, p. 94, 2019.

LI, J.; HU, C.; XU, D.; XIAO, J.; WANG, H. (2010) Application of time-series autoregressive integrated moving average model in predicting the epidemic situation of Newcastle disease. *Word Automation Congress*, 2010, 141,144

LI, N.; WANG, J.; WU, L. BENTLEY, Y. (2021). Predicting monthly natural gas production in China using a novel grey seasonal model with particle swarm optimization. **Energy**, 215, part. A, 119118.

LI, X.; SHU, L. (2009). Kernel based nonlinear dimensionality reduction for microarray gene expression data analysis. **Expert Systems with Applications**, 36, 7644–7650.

LIU, R.; REN, R.; LIU, J.; LIU, J. (2020). A clustering and dimensionality reduction based evolutionary algorithm for large-scale multi-objective problems. **Applied Soft Computing Journal**, 89, 106-120.

LIU, J.; WANG, S.; WEI, N.; CHEN, X.; XIE, H. WANG, J. (2021). Natural gas consumption forecasting: A discussion on forecasting history and future challenges. **Journal of Natural Gas Science and Engineering**, 22, 103930.

LIU, C.; WU, W.; XIE, W.; ZHANG, T. ZHANG, J. (2021). Forecasting natural gas consumption of China by using a novel fractional grey model with time power term. **Energy Reports**, 7, 788-797.

LIU, M.; DING, L.; BAI, Y. (2021). Application of hybrid model based on empirical mode decomposition, novel recurrent neural networks and the ARIMA to wind speed prediction. **Energy Conversion and Management**, 233, 113917.

MACKENZIE, J. (2011). Mean absolute percentage error and bias in economic forecasting. **Economics Letters**, 113, 259–262.

MAIA, P. R. **Método de interseção normal a fronteira por escores rotacionados em espaço elíptico de solução**. Tese (Doutorado em Engenharia de Produção) – Universidade Federal de Itajubá. Itajubá, p.137, 2020.

MAIA, P. R.; ALMEIDA, F. A.; PAES, V. C.; GOMES, J. H. F.; PAIVA, A. P. (2021). Multivariate steepest ascent method based on latent variables. **Applied Mathematical Modelling**, 90, 30-45.

MAKRIDAKIS, S. (1993). Accuracy measures: theoretic and practical concerns. **International Journal of Forecasting**, 9, 527-529.

MAKRIDAKIS, S.; WINKLER, R. L. (1983). Averages of Forecasts: Some Empirical Results. **Management Science**, 29, 987-996.

MAKRIDAKIS, S.; WHEELWRIGHT, S. C.; HYNDMAN, R. J. (1998). **Forecasting methods and applications**. (3th ed.). New York: John Wiley & Sons.

MAKRIDAKIS, S.; HIBON, M. (2000). The M3-Competition: results, conclusions and implications. **International Journal of Forecasting**, 16, 451-476.

MARTINS, I. D.; BAHIANSE, L.; INFANTE, C. E. D.; ARRUDA, E. F. (2020). Dimensionality reduction for multi-criteria problems: An application to the decommissioning of oil and gas installations. **Expert Systems With Applications**, 148, 113236.

MASON, R. L.; GUNST, R. F.; HESS, J. L. (2003). **Statistical Design and Analysis of Experiments: with applications to Engineering and Science**. (2 th ed.). New Jersey: John Wiley & Sons.

MEI, H.; CUCCARO, M. L.; MARTIN, E. R. (2007). Multifactor Dimensionality Reduction–Phenomics: a novel method to capture genetic heterogeneity with use of phenotypic variables. **The American Journal of Human Genetics**, 81, 1251-1261.

MEIRA, E.; OLIVEIRA, F. L. C.; JEON, J. (2021). Treating and Pruning: New approaches to forecasting model selection and combination using prediction intervals. **International Journal of Forecasting**, 37, 547-768.

MENDES, R. R. A.; PAIVA, A. P.; PERUCHI, R. S.; BALESTRASSI, P. P.; LEME, R. C.; SILVA, M. B. (2016). Multiobjective portfolio optimization of ARMA-GARCH time series based on experimental designs. **Computers and Operations Research**, 66, 434-444.

MYTTENAERE, A. GOLDEN, B. GRAND, B.; ROSSI, F. (2016) Mean Absolute Percentage Error for regression models, Mean Absolute Percentage Error for regression models. **Neurocomputing**, 192, 38-48.

MYERS, R. H.; MONTGOMERY, D. C.; ANDERSON-COOK, C. M; (2009). **Response Surface Methodology: process and product optimization using designed experiments**. (3th ed.). New Jersey: John Wiley & Sons.

MINITAB 17 Statistical Software. (2017). Versão 17. Minitab Inc.

MOREIRA, M. O. BALESTRASSI, P. P.; PAIVA, A. P.; RIBEIRO, P. F.; BONATTO, B. D. (2021). Design of experiments using artificial neural network ensemble for photovoltaic generation forecasting. **Renewable and Sustainable Energy Reviews**, 135, 110450.

MONTGOMERY, D. C.; JENNINGS, C.L.; KULAHCI, M. (2008). **Introduction to Times Series Analysis and Forecasting**. New York: John Wiley & Sons.

MONTGOMERY, D. C.; (2009). **Introduction to Statistical Quality Control**. (6<sup>th</sup> ed.). New York: John Wiley & Sons.

MYERS, R. H.; MONTGOMERY, D. C.; ANDERSON-COOK, C. M. (2009). **Response Surface Methodology: process and product optimization using Design Experiments**. (3th. ed.). New York: John Wiley & Sons.

MYTTENAERE, A.; GOLDEN, B.; LE GRAND, B.; ROSSI, F. (2016) Mean Absolute Percentage Error for regression models. **Neurocomputing**, 192, 8–48

NAVES, F. L.; PAULA, T. I.; BALESTRASSI, P. P.; BRAGA, W. L. M.; SAWHNEY, S. S.; PAIVA, A. P. (2017). Multivariate Normal Boundary Intersection based on rotated factor scores: A multiobjective optimization method for methyl orange treatment. **Journal of Cleaner Production**, 143, 413-439.

NEWBOLD, P. GRANGER, W. J. (1974). The combination of forecasts. **Journal of the Royal Statistical Society**, 146, 150-157.

NGUYEN-LE, V., KIM, M., SHIN, H., LITTLE, E., (2021). Multivariate approach to the gas production forecast using early production data for Barnett shale reservoir. **J. Nat. Gas Sci. Eng.** 87, 103776.

NGUYEN, B. N.; BORUFF, B.; TONTS, M. (2021). Looking through a crystal ball: Understanding the future of Vietnam's minerals and mining industry. **The Extractive Industries and Society**, 100907.

OLIVEIRA, L. G. ; OLIVEIRA, C. H. BRITO, T. G.; PAIVA, E, J. ; PAIVA, A. P.; FERREIRA, J. R. (2020). Nonlinear optimization strategy based on multivariate prediction capability ratios: Analytical schemes and model validation for duplex stainless steel end milling. **Precision Engineering**, 66, 229-254.

OLIVEIRA, F. A.; PAIVA, A. P.; LIMA, J. W. M.; BALESTRASSI, P. P.; MENDES, R. R. A. (2011). Portfolio optimization using Mixture Design of Experiments: Scheduling trades within electricity markets. **Energy Economics**, 33, 24-32.

ÖZMEN, Ayşe; YILMAZ, Yavuz; WEBER, Gerhard-Wilhelm. Natural gas consumption forecast with MARS and CMARS models for residential users. **Energy Economics**, v. 70, p. 357-381.

PANAPAKIDIS, I. P.; DAGOUMAS, A. (2017). Day-ahead natural gas demand forecasting based on the combination of wavelet transform and ANFIS/genetic algorithm/neural network model. **Energy**, 118, 231-245.

PAPPAS, S. S.; EKONOMOU, L.; KARAMPELAS, P.; KARAMOUSANTAS, D. C.; KATSIKAS, S. K.; CHATZARAKIS, G. E.; SKAFIDAS, P. D. (2010). Electricity demand load forecasting of the Hellenic Power System using an Arma Model. *Electric Power Systems Research*, 80, 256-264

PONCELA, P.; RODRIGUEZ, J.; SANCHES-MANGAS, R.; SENRA, E. (2011). Forecast combination through dimension reduction techniques. **International Journal of Forecasting**, 27, 224-237.

PONTES, Fabrício José et al. Optimization of Radial Basis Function neural network employed for prediction of surface roughness in hard turning process using Taguchi's orthogonal arrays. **Expert Systems with Applications**, v. 39, n. 9, p. 7776-7787, 2012.

PONTES, Fabrício José et al. Design of experiments and focused grid search for neural network parameter optimization. **Neurocomputing**, v. 186, p. 22-34, 2016.

REEVES, G. R.; LAWRENCE, K. D.; LAWRENCE. (1982). Combining Multiple Forecasts given Multiple Objectives. **Journal of Forecast**, 1, 271-279.

REEVES, G. R.; LAWRENCE, K. D.; LAWRENCE, S. M.; GUERARD JR., J. B. (1988). Combining Earnings Forecasts Using Multiple Objective Linear Programming. **Computers & Operations Research**, 15, 551-559.

REEVES, G. R.; LAWRENCE, K. D. (1991). Combining Forecasts Given Different Types of Objectives. **European Journal of Operational Research**, 51, 65-72.

RENCHER, A. C. **Methods of Multivariate Analysis**, Second Edition. New Jersey: JOHN WILEY & SONS, INC., 2009.

ROCHA, L. C. S.; PAIVA, A. P.; BALESTRASSI, P. P.; SEVERINO, G. JÚNIOR, P. R. (2015) Entropy-based weighting for multiobjective optimization: an application on vertical turning. **Mathematical Problems in Engineering**, 1-11.

ROCHA, L. C. S.; PAIVA, A. P.; JÚNIOR, P. R., BALESTRASSI, P. P.; CAMPOS, P. H. S. (2017a). Robust multiple criteria decision making applied to optimization of AISI H13 hardened steel turning with PCBN wiper tool. **International Journal of Advanced Manufacturing Technology**, 89, 2251-2268.

ROCHA, L. C. S.; PAIVA, A. P.; JÚNIOR, P. R., BALESTRASSI, P. P.; CAMPOS, P. H. S.; DAVIM, J. P. (2017b) Robust weighting applied to optimization of AISI H13 hardened-steel turning process with ceramic wiper tool: A diversity-based approach. **Precision Engineering**, 50, 235-247.

ROCHA, C. S.; JÚNIOR, P. R.; AQUILA, G.; PAIVA, A. P.; BALESTRASSI, P. P. (2020). Toward a robust optimal point selection: a multiple-criteria decision-making process applied to multi-objective optimization using response surface methodology. **Engineering with Computers**, <https://doi.org/10.1007/s00366-020-00973-5>

SANDERS, N. R.; RITZMAN, L. P. (1995). Bringing judgment into combination forecasts. **Journal of Operations Management**, 13, 311-321.

SIMAB, M.; JAVADI, M. S.; NEZHAD, A. E.; (2018). Multi-objective programming of pumped-hydro-thermal scheduling problem using normal boundary intersection and VIKOR. **Energy**, 143, 854-866.

SHAHBAZ, M.; LEAN, H. H.; FAROOQ, A. (2013). Natural gas consumption and economic growth in Pakistan. **Renewable and Sustainable Energy Reviews**, 18, 87–94.

SOLDO, B. (2012). Forecasting natural gas consumption. **Applied Energy**, 92, 26–37.

SONG, C. FU, X. (2020). Research on different weight combination in air quality forecasting models. **Journal of Cleaner Production**, 261, 121169.

SU, H.; ZIO, E.; ZHANG, J.; XU, M.; LI, X.; ZHANG, Z. (2019). A hybrid hourly natural gas demand forecasting method based on the integration of wavelet transform and enhanced Deep-RNN model. **Energy**, 178, 224-237.

TAK, N. (2021). Forecast combination with meta possibilistic fuzzy functions. **Information Sciences**, 560, 168-182.

TAŞPINAR, F.; CELEBI, N.; TUTKUN, N. (2013). Forecasting of daily natural gas consumption on regional basis in Turkey using various computational methods. **Energy and Buildings**, v. 56, p. 23-31.

TIAN, Y.; YU, J.; ZHAO, A. (2020). Predictive model of energy consumption for office building by using improved GWO-BP. **Energy Reports**, 6, 620-627.

THOMPSON, P. A. (1990). An MSE statistic for comparing forecast accuracy across series. **International Journal of Forecasting**, 6, 219-227.

TRATAR, F. L.; STRMČNIK, E. (2016). The comparison of Holt–Winters method and Multiple regression method: A qcase study. **Energy, Elsevier**, vol. 109(C), 266-276.

VAHIDINASAB, V.; JADID, S. (2010). Normal boundary intersection method for suppliers' strategic bidding in electricity markets: An environmental/economic approach. **Energy Conversion and Management**, 51, 1111-1119.

WANG, T.; LIN, B. (2014). China's naturalgas consumption and subsidies — From a sector perspective. **Energy Policy**, 65, 541–551.

WANG, J.; JIANG, H.; ZHOU, Q.; WU, J.; QIN, S. (2016). China's natural gas production and consumption analysis based on the multicycle Hubbert model and rolling Grey model. **Renewable and Sustainable Energy Reviews**, 53, 1149–1167

WANG, J.; ZHANG, X.; GUO, Z.; LU, H. (2017). Developing an early-warning system for air quality prediction and assessment of cities in China. **Expert Systems With Applications**, 84, 102–116.

WEI, W.; WU, H.; MA, H.. An autoencoder and LSTM-based traffic flow prediction method. **Sensors**, v. 19, n. 13, p. 2946,

WINKLER, R. L.; MAKRIDAKIS, S. (1983). Research on different weight combination in air quality forecasting models. **Journal of Cleaner Production**, 261, 121169.

WUN, L.; PEARN, W. L. (1991). Assessing the statistical characteristics of the mean absolute error or forecasting. **International Journal of Forecasting**, 7, 335-337.

YUKSELTAN, E.; YUCEKAYA, A.; BILGE, A. H. AKTUNC, E. A. (2021). Forecasting models for daily natural gas consumption considering periodic variations and demand segregation. **Socio-Economic Planning Sciences**, 74, 100937.

XIAO, L.; SHAO, W.; YU, M.; MA, J.; JIN, C. (2017) Research and application of a hybrid wavelet neural network model with the improved cuckoo search algorithm for electrical power system forecasting. *Applied Energy*, 198, 203-222

ZENG, B.; LI, C. (2016). Forecasting the natural gas demand in China using a self-adapting intelligent grey model. **Energy**, 112, 810-825.

ZHANG, G. P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. **Neurocomputing**, 50, 159-175

ZHONG, X.; ENKE, D. (2017). Forecasting daily stock market return using dimensionality reduction. **Expert Systems With Applications**, 67, 126–139.