

**UNIVERSIDADE FEDERAL DE ITAJUBÁ**  
**INSTITUTO DE ENGENHARIA DE PRODUÇÃO E GESTÃO**  
**PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE PRODUÇÃO**

**Otimização de modelos SARIMA-DEA com *ensembles* e  
Delineamento de Misturas**

**Gustavo dos Santos Leal**

**ITAJUBÁ**

**2023**

Gustavo dos Santos Leal

## **Otimização de modelos SARIMA-DEA com *ensembles* e Delineamento de Misturas**

Tese submetida ao Programa de Pós-Graduação em Engenharia de Produção da Universidade Federal de Itajubá, como parte dos requisitos para a obtenção do título de Doutor em Ciências em Engenharia de Produção.

**Área de concentração:** Engenharia de Produção

**Orientador:** Prof. Dr. Pedro Paulo Balestrassi

Itajubá

Fevereiro de 2023

UNIVERSIDADE FEDERAL DE ITAJUBÁ  
INSTITUTO DE ENGENHARIA DE PRODUÇÃO E GESTÃO  
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE PRODUÇÃO

Gustavo dos Santos Leal

**Otimização de modelos SARIMA-DEA com *ensembles* e  
Delineamento de Misturas**

Tese \_\_\_\_\_ por banca examinadora no dia 27 de  
fevereiro de 2023,

**Banca examinadora:**

Prof. Dr. Lupércio França Bessegato (UFJF)

Prof. Dr. Wesley Vieira da Silva (UFAL)

Prof. Dr. Carlos Henrique Pereira Mello (UNIFEI)

Prof. Dr<sup>a</sup>. Juliana Helena Daroz Gaudêncio (UNIFEI)

Prof. Dr. Pedro Paulo Balestrassi (Orientador / UNIFEI)

Itajubá

Fevereiro de 2023

## DEDICATÓRIA

*Ao meu eterno pai Chico (in memoriam).*

## **AGRADECIMENTOS**

Agradeço a meus pais, Sonia e Francisco (*in memoriam*) pelo dom da vida e por terem me proporcionado a oportunidade de adquirir conhecimento. Agradeço a meus irmãos, Eduardo e Felipe, por sempre estarem presentes em minha vida.

Agradeço à minha futura esposa, noiva e eternamente amiga Mariana por sempre entender o tempo de ausência e por sempre me apoiar incondicionalmente. À Maria e Rogério por continuarem me dando suporte em tudo. Ao Mateus e Marcos pela companhia prazerosa.

Ao meu orientador, Prof. Dr. Pedro Paulo Balestrassi, por ter tido a confiança de me orientar em tempos de mudanças, e pelas conversas e ensinamentos que tive no convívio diário, ensinando-me mais do que teorias.

Aos professores do Instituto de Engenharia de Produção e Gestão pelos ensinamentos em sala de aula e fora dele. Agradeço em especial ao Prof. Anderson Paiva por ser sempre um suporte aos alunos e de forma especial também ao Prof. Carlos Mello, Prof. João Turrioni, Prof. José Henrique e à Prof<sup>a</sup>. Juliana Gaudêncio por sempre estarem à disposição. Aos membros da banca de qualificação – Prof. Wesley da Silva e Prof<sup>a</sup> Juliana Gaudêncio – pelas contribuições à qualificação do trabalho.

Aos amigos Fernanda, Francisco (Paco) e Prof. Lupércio Bessegato pelo apoio dado. Sem vocês eu não estaria aqui. Aos meus amigos e colegas do PPGEP por todo companheirismo, em especial Anderson, João Victor, Yasmin, Franco, Simone, Aline, Natália, Fabrício, Vinícius e Estevão. Ao Rotary Club de Matias Barbosa Gov. Paulo Brandão.

À UNIFEI pela estrutura cedida. E aos órgãos de fomento à pesquisa CNPq, FAPEMIG e especialmente à CAPES.

## RESUMO

Realizar previsões acuradas é de suma importância para diversas áreas do conhecimento, como Economia, Gestão, Engenharias, Estatística etc. Existem várias abordagens para se realizar previsões: análise de séries temporais, análise de regressão, redes neurais artificiais etc. Contudo, um cuidado que todo pesquisador ou analista deve ter ao aplicar qualquer uma das referidas técnicas é o cuidado com o *overfitting* – que ocorre quando um determinado modelo possui tantos parâmetros que se ajusta bem ao conjunto de treino, mas prevê muito mal o conjunto de teste. Recentemente, técnicas de combinação de modelos estão em voga, uma vez que comprovadamente o *ensemble* de modelos faz com que as métricas de previsão sejam melhores. Entretanto, o problema do *overfitting* ainda pode estar presente nestes casos. Para contornar este problema, esta tese sugere a aplicação de um passo intermediário entre a seleção de modelos e a otimização dos pesos que é a utilização de um modelo de Análise por Envoltória de Dados adequado à presença de variáveis fracionadas para não ferir o pressuposto de convexidade. Para analisar este método, esta tese aplicará modelos de Box & Jenkins. Criam-se, portanto, *Decision Making Units* (DMUs) por meio de um Arranjo Fatorial Completo, modificando os parâmetros computacionais. Aplica-se a análise de supereficiência e retem-se as 4 DMUs com maiores índices de eficiência para posterior combinação por meio de otimização de Superfície de Resposta no contexto de Delineamento de Misturas. Propõe-se também a aplicação de técnicas estatística multivariadas para redução de dimensionalidade, a fim de tornar o problema computacionalmente menor. Para validar o método proposto, foi criado um estudo de simulação, comparando os resultados com o método Naive. A simulação mostrou que o método proposto nesta tese apresenta em média melhores resultados. Por fim, o método foi aplicado em séries de demanda de energia elétrica do Brasil e suas cinco regiões.

**Palavras-chave:** Séries Temporais; Delineamento de Misturas; *Ensemble*; Superfície de Resposta; Otimização; Análise por Envoltória de Dados.

## ABSTRACT

*Accurate forecasting is crucial for several areas of knowledge, such as Economics, Management, Engineering and Statistics. There are several approaches to perform forecasting: time series analysis, regression analysis, artificial neural networks, etc. However, both researchers or analysts must be aware when applying any of the aforementioned techniques because of overfitting – which occurs when a given model has so many parameters that it fits well to the training set, but predicts the test set very poorly. Recently, model combination techniques are widely spread, since the ensemble of models is proven to make the forecast metrics better. However, the overfitting problem may still occur in these cases. To overcome this, this thesis suggests the application of an intermediate step between the selection of models for the ensemble and the optimization of their weights, which is the application of a Data Envelopment Analysis model suitable for the presence of fractional variables so as not to harm the assumption of convexity. To analyze this method, this thesis applies Box & Jenkins models. Therefore, Decision Making Units (DMU) are created through a Complete Factorial Arrangement, modifying the computational parameters. Super-efficiency analysis is applied and the 4 DMUs with the highest efficiency indexes are retained for later combination through Response Surface (RSM) optimization in the context of Mixture Design. It is also proposed the application of multivariate statistical techniques for dimensionality reduction, in order to make the problem computationally smaller. To validate the proposed method, a simulation study was created, comparing the results with the Naive method. The simulation showed that the method proposed in this thesis presents, on average, better results. Finally, the method was applied to series about electricity demand from Brazil and its five geographic regions.*

**Keywords:** *Time Series Analysis; Mixture Design; Ensemble; Response Surface Methodology; Optimization; Data Envelopment Analysis.*

## LISTA DE FIGURAS

Figura 1. Visão geral dos métodos de ensemble. ....	22
Figura 2. Funções de autocorrelação de uma série temporal com tendência e transformações. .....	26
Figura 3. Representação geométrica de desenho fatorial $2^3$ . ....	34
Figura 4. Representação gráfica de design de misturas. ....	35
Figura 5. Interpretação dos coeficientes em um modelo linear de Scheffé. ....	36
Figura 6. Interpretação dos coeficientes no modelo quadrático de Scheffé. ....	37
Figura 7. Desenhos SLD linear e SLD quadrático. ....	37
Figura 8. Fluxograma proposto da metodologia. ....	50
Figura 9. Função de autocorrelação de uma série temporal aleatória com tendência. ....	51
Figura 10. Representação gráfica do Delineamento de Misturas Seleccionado. ....	57
Figura 11. Box-Plots das métricas de performance por método de combinação da Simulação para $n = 100$ . ....	63
Figura 12. Box-Plots das métricas de performance por método de combinação da Simulação para $n = 300$ . ....	63
Figura 13. Box-Plots das métricas de performance por método de combinação da Simulação para $n = 500$ . ....	64
Figura 14. Séries temporais de demanda de energia elétrica do Brasil e de suas regiões (em milhares / MWh). ....	65
Figura 15. Funções de autocorrelação das séries temporais analisadas. ....	67
Figura 16. Funções de autocorrelação da primeira diferença das séries temporais analisadas. ....	68
Figura 17. Padrão de sazonalidade da demanda de energia elétrica da região Norte. ....	69
Figura 18. Padrão de sazonalidade da demanda de energia elétrica da região Nordeste. ....	69
Figura 19. Padrão de sazonalidade da demanda de energia elétrica da região Norte. ....	70
Figura 20. Padrão de sazonalidade da demanda de energia elétrica da região Norte. ....	70
Figura 21. Padrão de sazonalidade da demanda de energia elétrica da região Norte. ....	71
Figura 22. Padrão de sazonalidade da demanda de energia elétrica da região Norte. ....	71
Figura 23. Box-Plot do MAPE do ajuste e da previsão por número de parâmetros dos modelos de previsão para a região Norte. ....	75
Figura 24. Box-Plot do MAPE do ajuste e da previsão por número de parâmetros dos modelos de previsão para a região Norte. ....	75



Figura 25. Box-Plot do MdAE do ajuste e da previsão por número de parâmetros dos modelos de previsão para a região Norte. ....	76
Figura 26. Box-Plot do MAE do ajuste e da previsão por número de parâmetros dos modelos de previsão para a região Norte. ....	76
Figura 27. Box-Plot do MAPE do ajuste e da previsão por número de parâmetros dos modelos de previsão para a região Nordeste. ....	78
Figura 28. Box-Plot do RMSE do ajuste e da previsão por número de parâmetros dos modelos de previsão para a região Nordeste. ....	78
Figura 29. Box-Plot do MdAE do ajuste e da previsão por número de parâmetros dos modelos de previsão para a região Nordeste. ....	79
Figura 30. Box-Plot do MAE do ajuste e da previsão por número de parâmetros dos modelos de previsão para a região Nordeste. ....	79
Figura 31. Box-Plot do MAPE do ajuste e da previsão por número de parâmetros dos modelos de previsão para a região Sudeste. ....	81
Figura 32. Box-Plot do MAPE do ajuste e da previsão por número de parâmetros dos modelos de previsão para a região Sudeste. ....	81
Figura 33. Box-Plot do MdAE do ajuste e da previsão por número de parâmetros dos modelos de previsão para a região Sudeste. ....	82
Figura 34. Box-Plot do MAE do ajuste e da previsão por número de parâmetros dos modelos de previsão para a região Sudeste. ....	82
Figura 35. Box-Plot do MAPE do ajuste e da previsão por número de parâmetros dos modelos de previsão para a região Sul. ....	84
Figura 36. Box-Plot do RMSE do ajuste e da previsão por número de parâmetros dos modelos de previsão para a região Sul. ....	84
Figura 37. Box-Plot do MdAE do ajuste e da previsão por número de parâmetros dos modelos de previsão para a região Sul. ....	85
Figura 38. Box-Plot do MAE do ajuste e da previsão por número de parâmetros dos modelos de previsão para a região Sul. ....	85
Figura 39. Box-Plot do MAPE do ajuste e da previsão por número de parâmetros dos modelos de previsão para a região Centro-Oeste. ....	87
Figura 40. Box-Plot do RMSE do ajuste e da previsão por número de parâmetros dos modelos de previsão para a região Centro-Oeste. ....	87

Figura 41. Box-Plot do MdAE do ajuste e da previsão por número de parâmetros dos modelos de previsão para a região Centro-Oeste.....	88
Figura 42. Box-Plot do MdAE do ajuste e da previsão por número de parâmetros dos modelos de previsão para a região Centro-Oeste.....	88
Figura 43. Box-Plot do MAPE do ajuste e da previsão por número de parâmetros dos modelos de previsão do Brasil. ....	90
Figura 44. Box-Plot do RMSE do ajuste e da previsão por número de parâmetros dos modelos de previsão do Brasil. ....	90
Figura 45. Box-Plot do MdAE do ajuste e da previsão por número de parâmetros dos modelos de previsão do Brasil. ....	91
Figura 46. Box-Plot do MAE do ajuste e da previsão por número de parâmetros dos modelos de previsão do Brasil. ....	91
Figura 47. Análise de resíduos da regressão de superfície de resposta tendo o MAPE como variável resposta. ....	101
Figura 48. Análise de resíduos da regressão de superfície de resposta tendo o RMSE como variável resposta. ....	103
Figura 49. Análise de resíduos da regressão de superfície de resposta tendo o MdAE como variável resposta. ....	104
Figura 50. Análise de resíduos da regressão de superfície de resposta tendo o MAE como variável resposta. ....	106
Figura 51. Gráfico de Traço de Cox para a resposta MAPE da região norte do Brasil.....	107
Figura 52. Gráfico de Traço de Cox para a resposta RMSE da região norte do Brasil.....	108
Figura 53. Gráfico de Traço de Cox para a resposta MdAE da região norte do Brasil.....	108
Figura 54. Gráfico de Traço de Cox para a resposta MAE da região norte do Brasil.....	109
Figura 55. Ajuste do conjunto treino e previsão do conjunto teste para a região Sudeste com o método proposto. ....	111
Figura 56. Gráfico da previsão de demanda de energia elétrica do Sudeste do Brasil utilizando-se o método proposto por este trabalho.....	113

## LISTA DE TABELAS

Tabela 1. DMUs referentes à modelagem ARIMA.....	54
Tabela 2. DMUs referentes à modelagem SARIMA.....	54
Tabela 3. Inputs e Outputs a serem considerados no modelo DEA. ....	56
Tabela 4. Representação geral do delineamento proposto com as variáveis resposta.....	58
Tabela 5. Estatísticas descritivas das métricas no contexto da Simulação realizada. ....	62
Tabela 6. Estatísticas descritivas dos dados de demanda de energia elétrica (em milhares / MWh). ....	65
Tabela 7. Estatísticas descritivas do conjunto treino das séries temporais analisadas. ....	66
Tabela 8. Estatísticas descritivas do conjunto teste das séries temporais analisadas. ....	66
Tabela 9. Teste de raiz unitária de KPSS. ....	67
Tabela 10. Teste de raiz unitária de KPSS sobre a primeira diferença das séries temporais analisadas.....	68
Tabela 11. Teste de Friedman para sazonalidade .....	72
Tabela 12. Métricas de ajuste e quantidade de parâmetros estimados para cada DMU da série de demanda de energia elétrica do Norte do Brasil. ....	73
Tabela 13. Métricas de previsão para cada DMU da série de demanda de energia elétrica do Norte do Brasil. ....	74
Tabela 14. Estatísticas descritivas das métricas de ajuste e previsão para os modelos de demanda de energia elétrica da região Norte.....	74
Tabela 15. Métricas de ajuste e quantidade de parâmetros estimados para cada DMU da série de demanda de energia elétrica do Nordeste do Brasil.....	77
Tabela 16. Métricas de previsão para cada DMU da série de demanda de energia elétrica do Nordeste do Brasil. ....	77
Tabela 17. Estatísticas descritivas das métricas de ajuste e previsão para os modelos de demanda de energia elétrica da região Nordeste. ....	77
Tabela 18. Métricas de ajuste e quantidade de parâmetros estimados para cada DMU da série de demanda de energia elétrica do Sudeste do Brasil.....	80
Tabela 19. Métricas de previsão para cada DMU da série de demanda de energia elétrica do Sudeste do Brasil. ....	80
Tabela 20. Estatísticas descritivas das métricas de ajuste e previsão para os modelos de demanda de energia elétrica da região Nordeste. ....	80

Tabela 21. Métricas de ajuste e quantidade de parâmetros estimados para cada DMU da série de demanda de energia elétrica do Sul do Brasil.....	83
Tabela 22. Métricas de previsão para cada DMU da série de demanda de energia elétrica do Sul do Brasil. ....	83
Tabela 23. Estatísticas descritivas das métricas de ajuste e previsão para os modelos de demanda de energia elétrica da região Sul. ....	83
Tabela 24. Métricas de ajuste e quantidade de parâmetros estimados para cada DMU da série de demanda de energia elétrica do Centro-Oeste do Brasil.....	86
Tabela 25. Métricas de previsão para cada DMU da série de demanda de energia elétrica do Centro-Oeste do Brasil. ....	86
Tabela 26. Estatísticas descritivas das métricas de ajuste e previsão para os modelos de demanda de energia elétrica da região Centro-Oeste. ....	86
Tabela 27. Métricas de ajuste e quantidade de parâmetros estimados para cada DMU da série de demanda de energia elétrica do Brasil. ....	89
Tabela 28. Métricas de previsão para cada DMU da série de demanda de energia elétrica do Brasil.....	89
Tabela 29. Estatísticas descritivas das métricas de ajuste e previsão para os modelos de demanda de energia elétrica do Brasil.....	89
Tabela 30. DMUs eficientes por região analisada e seus respectivos modelos.....	93
Tabela 31. Delineamento de Misturas SLD (4, 5) com as métricas de performance de previsão em relação ao conjunto teste como respostas para a região Norte. ....	94
Tabela 32. Delineamento de Misturas SLD (4, 5) com as métricas de performance de previsão em relação ao conjunto teste como respostas para a região Nordeste. ....	94
Tabela 33. Delineamento de Misturas SLD (4, 5) com as métricas de performance de previsão em relação ao conjunto teste como respostas para a região Sudeste. ....	94
Tabela 34. Delineamento de Misturas SLD (4, 5) com as métricas de performance de previsão em relação ao conjunto teste como respostas para a região Sul. ....	95
Tabela 35. Delineamento de Misturas SLD (4, 5) com as métricas de performance de previsão em relação ao conjunto teste como respostas para a região Centro-Oeste. ....	95
Tabela 36. Delineamento de Misturas SLD (4, 5) com as métricas de performance de previsão em relação ao conjunto teste como respostas para a região o Brasil. ....	95
Tabela 37. Matriz de correlação entre as métricas padronizadas oriundas do delineamento de misturas para a região Norte. ....	96

Tabela 38. Matriz de correlação entre as métricas padronizadas oriundas do delineamento de misturas para a região Nordeste.....	96
Tabela 39. Matriz de correlação entre as métricas padronizadas oriundas do delineamento de misturas para a região Sudeste.....	97
Tabela 40. Matriz de correlação entre as métricas padronizadas oriundas do delineamento de misturas para a região Sul.....	97
Tabela 41. Matriz de correlação entre as métricas padronizadas oriundas do delineamento de misturas para a região Centro-Oeste.....	98
Tabela 42. Matriz de correlação entre as métricas padronizadas oriundas do delineamento de misturas para a região Centro-Oeste.....	98
Tabela 43. Regressão estimada para o MAPE da região Sudeste. ....	100
Tabela 44. Anova para o MAPE da região Sudeste. ....	100
Tabela 45. Regressão estimada para o RMSE da região Sudeste.....	101
Tabela 46. Anova para o RMSE da região Sudeste.....	102
Tabela 47. Regressão estimada para o MdAE da região Sudeste.....	103
Tabela 48. Anova para o MdAE da região Sudeste.....	104
Tabela 49. Regressão estimada para o MAE da região Sudeste.....	105
Tabela 50. Anova para o MdAE da região Sudeste.....	105
Tabela 51. Pesos otimizados via algoritmo de Desirability. ....	107
Tabela 52. Comparação do método proposto SARIMA-DEA-R e o método Naive para previsão de demanda de energia elétrica entre o Brasil e seus 5 regiões. ....	110
Tabela 53. Quinze últimos elementos da previsão do conjunto teste sobre o conjunto teste e intervalos de confiança para a demanda de energia elétrica no Sudeste do Brasil.....	111
Tabela 54. Quinze últimos elementos da previsão da demanda de energia elétrica no Sudeste do Brasil.....	112

# SUMÁRIO

<b>1. INTRODUÇÃO</b> .....	15
1.1 OBJETIVO GERAL .....	17
1.2 OBJETIVOS ESPECÍFICOS .....	17
1.3 CONTRIBUIÇÕES .....	18
1.4 LIMITAÇÕES DA PESQUISA.....	18
1.5 ESTRUTURA .....	19
<b>2. REFERENCIAL TEÓRICO</b> .....	20
2.1 SÉRIES TEMPORAIS .....	20
2.1.1 <i>Ensemble</i> de previsões .....	21
2.1.2 Modelos ARIMA.....	24
2.1.3 Modelos SARIMA .....	26
2.1.4 Métricas para previsão de séries temporais .....	27
2.2 ANÁLISE POR ENVOLTÓRIA DE DADOS .....	28
2.3 MODELOS DE SUPEREFICIÊNCIA.....	30
2.4 PROBLEMAS DE CONVEXIDADE EM ANÁLISE POR ENVOLTÓRIA DE DADOS .	31
2.5 DEA COMO CRITÉRIO DE SELEÇÃO DE MODELOS DE PREVISÃO.....	32
2.6 DESENHO E ANÁLISE DE EXPERIMENTOS .....	33
2.7 DELINEAMENTO DE MISTURAS E METODOLOGIA DE SUPERFÍCIE DE RESPOSTA.....	34
2.8 ESTIMAÇÃO DA SUPERFÍCIE DE RESPOSTA .....	38
2.9 ESCOLHA DE MODELOS DE REGRESSÃO E DE VARIÁVEIS REGRESSORAS .....	39
2.10 OTIMIZAÇÃO DE SUPERFÍCIES DE RESPOSTA .....	42
2.10.1 Otimização Multiobjetiva e Análise de Componentes Principais .....	42
2.10.2 <i>Desirability</i> .....	45
2.11 SIMULAÇÃO DE MONTE CARLO .....	46
2.12 PREVISÃO DE DEMANDA DE ENERGIA ELÉTRICA.....	47
<b>3. ASPECTOS METODOLÓGICOS</b> .....	49
<b>4. RESULTADOS E DISCUSSÕES</b> .....	61
4.1. SIMULAÇÃO .....	61
4.2. APLICAÇÃO EMPÍRICA.....	64
<b>5. CONCLUSÃO</b> .....	114
<b>6. REFERÊNCIAS</b> .....	116

## 1. INTRODUÇÃO

Prever o futuro com exatidão faz parte do âmago do ser humano e é importante para muitas ciências, tais como: Estatística, Economia, Administração, Finanças, Ciências Sociais, Engenharia de Produção etc. Apesar de haver muitos métodos concorrentes para o alcance deste fim, os métodos estatísticos continuam sendo muito utilizados por conseguirem captar o comportamento aleatório dos dados analisados.

A produção de dados cresceu de forma proporcional ao avanço do desenvolvimento tecnológico. Atualmente, dados são produzidos de maneira massiva, como por meio da Internet das Coisas, Indústria 4.0, Cidades Inteligentes e digitalização dos procedimentos médicos. Essa produção em massa de dados traz a oportunidade e a necessidade de se realizar previsões de maneira mais acurada.

Quando falamos de previsão de uma variável aleatória há dois principais recursos estatísticos a serem empregados: regressão múltipla e modelos de séries temporais. Enquanto o primeiro busca criar relações entre uma variável resposta e várias outras variáveis explicativas, o segundo procura criar relações entre a variável de interesse e seus próprios valores passados.

Entre os modelos de séries temporais univariados destacam-se os modelos ARIMA. Esse modelo origina-se da junção de modelos autorregressivos (AR), baseados em Yule (1927) e os modelos de média móveis (MA), propostos por Walker (1931). A junção desses modelos foi proposta por George Box e Gwilym Jenkins em 1970. Inicialmente, a publicação deste trabalho não obteve grande destaque na literatura acadêmica. No entanto, com o advento dos computadores, o modelo ARIMA começou a ser o mais utilizado para modelos de previsão univariados (DE GOOIJER & HYNDMAN, 2006)

Neste contexto, métodos de *ensemble* aparecem como importantes ferramentas para este fim, sendo reconhecidos como uma das formas mais competitivas para resolver problemas de previsão (WICHARD & OGORZALEK, 2007). Técnicas de *ensemble* são consideradas como um “guarda-chuva” que engloba diversos métodos que combinam múltiplas técnicas de aprendizado supervisionadas, dando pesos para diversos modelos com o objetivo de melhorar a acuracidade da previsão (SAGI & ROKACH, 2018).

Entretanto, formas de medir a acuracidade de uma previsão podem gerar controvérsias. Existem diferentes indicadores de previsão que podem ser utilizados, como por exemplo: *Mean Absolute Error* (MAE), *Mean Square Error* (MSE), *Mean Absolute Percentual Error* (MAPE) etc. Desta maneira, determinado método de previsão pode ser escolhido de forma arbitrária dependendo da métrica de avaliação utilizada.

Com o intuito de superar este problema, Emrouznejad, Rostami-Tabar e Petridis (2016) propuseram a aplicação de Análise por Envoltória de Dados (DEA) para aplicar pesos que unifiquem diversas medidas de acuracidade de modo a tratá-las simultaneamente. DEA consiste em um método não paramétrico de múltiplas otimizações lineares que busca criar um *output virtual* e um *input virtual*, permitindo a comparação de diversas unidades (DMUs) entre elas próprias.

Analisando as competições M-Competition e M3-Competition, os autores chegaram à conclusão que a utilização de medidas distintas de acurácia leva a diferentes resultados mesmo analisando-se a mesma série temporal. Os autores argumentam que a utilização de DEA para tratar as diversas métricas simultaneamente pode evitar, inclusive, o problema de *overfitting*<sup>1</sup>.

A aplicação de técnicas relativas à estatística industrial em outros campos que não propriamente a produção em si vem ganhando espaço nos últimos anos. Isaías *et al.* (2021) aplicaram delineamento de misturas para otimizar a seleção de portfólio de projetos de geração de energia fotovoltaica. de Oliveira *et al.* (2011) aplicaram o mesmo método para otimização de agendamento de transações em mercado de eletricidade e Mendes *et al.* (2016) e Leal *et al.* (2022) para otimização de portfólio de ações, mostrando que a aplicação dessas técnicas é relevante para outras áreas do conhecimento.

Assim, este trabalho visa outra aplicação de delineamento de misturas e otimização de superfície de resposta no contexto de *ensemble* de métodos de *Autoregressive Integrated Moving Average* (ARIMA), mais especificamente em séries sazonais que são bem ajustadas por um modelo *Seasonal Autoregressive Integrated Moving Average* (SARIMA).

---

<sup>1</sup> *Overfitting* ocorre quando se gasta tantos graus de liberdade para ajustar um modelo que o mesmo atua bem apenas no conjunto de treino dos dados, não performando de maneira satisfatória no conjunto de teste. O matemático John Von Neumann, ao ser questionado sobre um modelo de *Machine Learning* que classificava bem um conjunto de elefantes, disse: “Dê-me quatro parâmetros que classifico bem os elefantes. Dê-me cinco parâmetros que faço suas trombas se movimentarem” (MAYER, KHAIRY & HOWARD, 2010).



Com o intuito de verificar o comportamento geral do algoritmo, este trabalho aplicará o mesmo em séries aleatórias geradas por meio de um *Data Generating Process* (DGP). A simulação abarcará séries de distintos tamanhos para avaliação do desempenho do algoritmo em diferentes cenários.

Do mesmo modo, para verificar o comportamento específico do algoritmo em um caso real, este trabalho aplicará a metodologia proposta em séries de demanda de energia elétrica do Brasil e de suas cinco regiões: Norte, Nordeste, Sul, Sudeste e Centro-Oeste.

A previsão de demanda de energia é de suma importância para o desenvolvimento sustentável de qualquer país, uma vez que é insumo fundamental para a indústria, agricultura, área de saúde, acesso à água e para a qualidade de vida em si (SUGANTHI & SAMUEL, 2012).

## 1.1 OBJETIVO GERAL

O objetivo geral desta tese é propor um algoritmo para a otimização de pesos em *ensemble* de modelos SARIMA da metodologia de Superfície de Resposta no contexto de Desenho de Misturas, levando em consideração a eficiência dos modelos estimados para gerar previsões acuradas que utilizem o menor número de parâmetros.

## 1.2 OBJETIVOS ESPECÍFICOS

Para alcançar o objetivo geral deste trabalho, podem-se elencar os seguintes objetivos específicos:

- a) Determinar os modelos SARIMA a serem aplicados;
- b) Aplicar todas as configurações de modelos SARIMA, variando os parâmetros computacionalmente de forma planejada;
- c) Determinar métricas de ajuste e de previsão adequadas ao problema;
- d) Determinar modelos de Análise por Envoltória de Dados (DEA) adequados ao problema, uma vez que as métricas de ajuste são variáveis fracionáveis;
- e) Determinar os *inputs* e os *outputs* a serem utilizados no modelo DEA indicado;

- f) Determinar o delineamento de misturas adequado ao problema proposto;
- g) Realizar estudo de Simulação para comparação do algoritmo proposto com o método *Naive* de determinação de pesos de *ensembles*;
- h) Aplicar o algoritmo proposto em séries temporais empíricas para demonstrar a aplicabilidade real do método.

### 1.3 CONTRIBUIÇÕES

Este estudo busca contribuir com analistas e pesquisadores que realizam previsões de forma a propor um algoritmo para determinação de pesos de *ensemble* de modelos, utilizando técnicas de estatística industrial e otimização junto com DEA para a escolha preliminar de modelos eficientes.

Ao se adicionar DEA no algoritmo, espera-se selecionar modelos parcimoniosos, mas que se ajustem e prevejam bem as séries temporais em análise, facilitando a tomada de decisão de quais modelos devem ser considerados em uma análise desse tipo.

O método proposto será aplicado a séries empíricas de demanda de energia elétrica do Brasil e suas respectivas regiões, mostrando a contribuição que o método proposto pode gerar no âmbito prático do problema de previsão de séries temporais.

### 1.4 LIMITAÇÕES DA PESQUISA

Este trabalho propõe a utilização de um modelo de Análise por Envoltória de Dados (DEA) para escolha de modelos de híbridos de séries temporais. A ideia é selecionar modelos parcimoniosos por meio da análise da relação *inputs* e *outputs* de cada modelo.

Como limitação deste estudo, pode-se citar a escolha de apenas um modelo para a análise. Entretanto, ressalta-se que a metodologia proposta é geral, podendo ser aplicada para quaisquer modelos que podem ser considerados, sejam eles baseados em estatística, econometria ou em computação evolutiva.

## 1.5 ESTRUTURA

A partir deste ponto este trabalho se divide em 5 seções: referencial teórico; metodologia; resultados e discussões; e conclusão.

## 2. REFERENCIAL TEÓRICO

### 2.1 SÉRIES TEMPORAIS

Metcalf e Cowpertwait (2009) definem uma série temporal como observações que descrevem o movimento de uma variável ao longo do tempo. Em termos matemáticos, um vetor  $y$  de tamanho  $n$  é uma série temporal se:

$$f(t) = (y_t, y_{t-1}, y_{t-2}, \dots, y_{t-n}) \quad (1)$$

Para:  $y_t$  = variável  $y$  no tempo  $t$ ;  $y_{t-1}$  = variável  $y$  no tempo  $t - 1$ ;  $y_{t-2}$  = variável  $y$  no tempo  $t - 2$ ;  $y_{t-n}$  = variável  $y$  no tempo  $t - n$ .

A partir de uma série temporal o objetivo geralmente é realizar previsões sobre os valores futuros levando em consideração a estrutura dos valores passados da mesma. Segundo Hyndman e Athanasopoulos (2018) as previsões podem ser feitas considerando três horizontes temporais:

- **Previsão de curto prazo**

São necessárias para agendar produção, alocar mão de obra e meios de transporte.

- **Previsões de médio prazo**

São importantes para determinar necessidade de compras e aquisição de matéria prima, contratação ou demissão de mão de obra e investimentos máquinas e equipamentos.

- **Previsões de longo prazo**

Usadas quando se realiza o planejamento estratégico, quando se decide o nível de expansão ou retração do capital físico da organização.

Os mesmos autores elencam o passo a passo para que um modelo de previsão seja selecionado.

#### **Passo 1. Definição do problema**

É possivelmente o passo mais importante do processo, uma vez que a definição incorreta do problema a ser tratado pode levar a aplicações de técnicas inadequadas. Nesta etapa é importante a definição de quais dados coletar e definir qual a finalidade da previsão.

## **Passo 2. Juntar informações**

Nesta etapa duas informações são requeridas: os dados históricos das variáveis selecionadas na primeira etapa e a seleção pessoas capacitadas para realizar a previsão.

## **Passo 3. Análise Exploratória dos Dados**

Extrair estatísticas básicas, como média, mediana, desvio-padrão e variância é a ação necessária nesta etapa. A análise gráfica é de suma importância nesta etapa, uma vez que permite a identificação prévia de padrões na série temporal. Deve-se analisar a presença de tendência, sazonalidade, ciclo e a presença de *outliers* que podem influenciar nos parâmetros estimados.

## **Passo 4. Escolher e ajustar modelos**

Deve-se aplicar, segundo os autores, no mínimo 2 modelos para fins de comparação. Cada modelo deve atender aos pressupostos para que a estimação dos parâmetros seja consistente, eficiente e não enviesada.

## **Passo 5. Utilizar e avaliar o modelo escolhido**

Dado que os parâmetros do modelo selecionado foram estimados, pode-se utilizar o modelo para realizar a previsão para dado período selecionado. Os autores indicam que um dos modos de avaliar a acurácia do modelo é separar o banco de dados em treino (80% dos dados) e teste (20% dos dados).

### **2.1.1 Ensemble de previsões**

De acordo com Polikar (2006), os métodos *ensemble* trabalham para melhorar o desempenho preditivo evitando o *overfitting*. Quando um pequeno número de observações está disponível, um algoritmo de aprendizado tende a se ajustar a um modelo que prediz perfeitamente todas as informações disponíveis, mas falha em prever informações não observadas. Assim, ao reunir

vários modelos diferentes em um único modelo, o risco de cair em um Erro Tipo I é reduzido. Outro fator que faz com que os modelos *ensemble* tenham bons resultados preditivos é o fato desses modelos possuírem vantagens computacionais sobre a previsão simples dos modelos.

Conforme ilustrado na Figura 1, os métodos *ensemble* treinam vários algoritmos de aprendizado para que eles possam resolver o mesmo problema simultaneamente. Os métodos de *ensemble* possuem dois tipos de interação de aprendizado (BARAK & SADEGH, 2016): aprendizado sem interação com os modelos a serem cruzados, que correspondem a interação a posteriori, e aprendizado com interação no processo de aprendizagem (*ensemble a priori*).

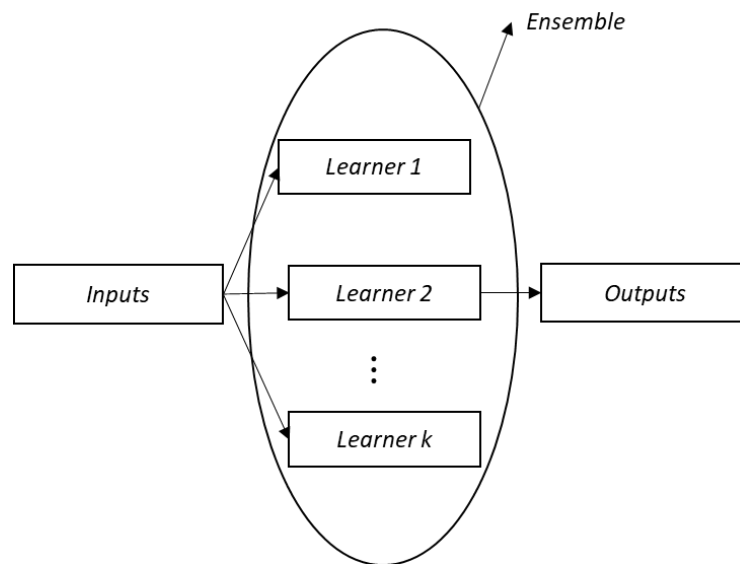


Figura 1. Visão geral dos métodos de *ensemble*.  
Fonte: elaboração própria.

A média é o método mais comumente usado para combinar algoritmos de aprendizado (SAGI & ROKACH, 2018). Considerando um conjunto de  $K$  algoritmos individuais de aprendizagem  $(y_1, y_2, \dots, y_K)$  e um output  $y_i$  para dado  $\mathbf{x}$  é  $y_i(\mathbf{x}) \in \mathbb{R}$ . O objetivo é combinar  $(y_1, y_2, \dots, y_K)$  de modo a se obter uma predição final do valor real da variável em análise,

O cálculo da média simples das predições dos  $k$  modelos considerados é dado por pela Equação 2.

$$Y(x) = \frac{1}{K} \sum_{i=1}^K y_i(x) \quad (2)$$

Considerando um caso em que:

$$y_i(x) = f(x) + u_i(x), i = 1, 2, \dots, T \quad (3)$$

Então o erro quadrático médio é dado por:

$$\int (y_i(x) - f(x))^2 p(x) dx = \int u_i(x)^2 p(x) dx \quad (4)$$

E o erro médio esperado de cada algoritmo de aprendizado é:

$$\bar{u}(y) = \frac{1}{K} \sum_{i=1}^K \int u_i(x)^2 p(x) dx \quad (5)$$

O erro combinado esperado dos modelos de conjunto propostos é:

$$\begin{aligned} \bar{\bar{u}}(Y) &= \int \left( \frac{1}{K} \sum_{i=1}^K y_i(x) - f(x) \right)^2 p(x) d(x) \\ &= \int \left( \frac{1}{K} \sum_{i=1}^K u_i(x) \right)^2 p(x) d(x) \end{aligned} \quad (6)$$

Observa-se, portanto, que  $\bar{\bar{u}}(Y) \leq \bar{u}(y)$ .

(WANG et al., 2010) colocam o problema de *ensemble* em termos de otimização. Seja ainda  $K$  modelos de previsão sendo considerados, então a partir da Equação 3.

$$\hat{Y}_t = \sum_{i=1}^K \hat{\omega}_i \hat{y}_{it}, t = 1, 2, \dots, \quad (7)$$

Então determinar o peso para cada modelo individual é a atividade chave dos modelos de *Ensemble*, como por exemplo dado pela Equação 8.

$$\min_{\omega} \sum_{i=1}^h |y_t - \hat{Y}_t|$$

Sujeito a: (8)

$$\begin{aligned} \sum_{i=1}^h \omega_i &= 1 \\ 0 &\leq \omega_i \leq 1 \\ i &= 1, 2, 3, \dots, h \end{aligned}$$

Como será visto na seção 2.7 a equação acima claramente pode ser tratada no contexto de delineamentos experimentais de misturas.

### 2.1.2 Modelos ARIMA

Existem diversas técnicas que podem ser aplicadas para modelar séries temporais. Dentre as mais simples, tem-se: modelos de extrapolação de tendência linear e quadrática, de crescimento exponencial, de tendência autorregressivo, de tendência autorregressivo nos logaritmos e modelagem por curva logística. Uma outra abordagem para modelar séries temporais é a partir da suavização das mesmas, destacando-se nessa perspectiva os modelos de médias móveis e os de média móveis ponderadas.

Os modelos podem ser construídos de maneira mais formal, utilizando modelos denominados autorregressivos de médias móveis, ou ARIMA( $p, d, q$ ), em que  $p$  é a ordem referente a um modelo autorregressivo,  $d$  é a quantidade de diferenciações necessárias para transformar a série em estacionária e  $q$  é a ordem referente a um modelo de média móveis. Um modelo de médias móveis de ordem  $q$  – MA( $q$ ) é apresentado nas Equações 9 e 10.

$$y_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} \quad (9)$$

$$var(y_t) = \sigma_\varepsilon^2 + (1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2) \quad (10)$$

Para:  $\varepsilon_t$  = termo de perturbação no tempo  $t$ , com distribuição normal e variância  $\sigma_\varepsilon^2$ ;  $\theta_q$  = parâmetros do modelo.

A fim de determinar a ordem de um MA( $q$ ) analisa-se a função de autocorrelação, apresentada na Equação 11.

$$\rho_k = \begin{cases} \frac{-\theta_k + \theta_1 \theta_{k+1} + \dots + \theta_{q-k} \theta_q}{1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2}, & k = 1, \dots, q \\ 0, & k > q \end{cases} \quad (11)$$



Um modelo autorregressivo – AR(p) é formalizado pelas Equações 12 e 13.

$$y_t = \phi_1 y_{t-1} - \phi_2 y_{t-2} - \dots - \phi_p y_{t-p} + \delta + \varepsilon_t \quad (12)$$

$$\gamma_k = \frac{\phi_1^k \sigma_e^2}{1 - \phi_1^2} \quad (13)$$

Para:  $\phi_p$  = parâmetros do modelo;  $\delta$  = termo constante que se relaciona com a média do processo estocástico;  $\gamma_k$  = covariância do processo com deslocamento.

Para determinar a ordem de um modelo autorregressivo, analisa-se a função de autocorrelação parcial, que é dada pela Equação 14.

$$\sigma_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \dots + \phi_p \rho_{k-p} \quad (14)$$

Para:  $\sigma_k$  = autocorrelação parcial;  $\rho_{k-p}$  = correlação no período  $k$  em relação ao período  $p$ .

Muitos processos aleatórios não podem ser modelados apenas como médias móveis ou como puramente autorregressivos, pois possuem características pertinentes a ambos os processos. Desta maneira, utiliza-se um processo misto, denotado por ARMA( $p, q$ ) para modelar essas séries, que é a junção dos modelos MA( $q$ ) e AR( $p$ ). A parte de médias móveis modela a série temporal por meio de uma média móvel de valores de erros que ocorreram no passado (GUJARATI, 2011).

Outro conceito importante a ser destacado é que uma série dita não estacionária homogênea pode ser diferenciada, ou seja, ela pode produzir uma nova série cujos valores são obtidos por meio da diferença entre os valores da série temporal original. A realização deste procedimento uma ou duas vezes já é suficiente para proporcionar uma série temporal estacionária (MONTGOMERY, JENNINGS & KULAHCI, 2015). O número de diferenças ( $d$ ) necessárias para tornar uma série estacionária é chamado de ordem de integração (TSAY, 2005).

Finalmente, um modelo ARIMA irá apresentar os três parâmetros,  $p, d, q$ , explicados anteriormente. Dessa forma, elabora-se um modelo composto de uma parte autorregressiva, outra de médias móveis e uma parte integrada. Se necessário diferenciar os dados para retirar tendência, essa diferenciação pode ser aplicada uma ou mais vezes (TSAY, 2005).

### 2.1.3 Modelos SARIMA

Analisando-se a função de autocorrelação, pode-se chegar à conclusão que a série temporal em análise possui sazonalidade. Quando ocorre este fenômeno há basicamente duas abordagens a serem adotadas: a retirada da sazonalidade, comumente denominado de ajuste sazonal e o tratamento da sazonalidade para fins de previsão.

No contexto de modelos de Box & Jenkins, os parâmetros computacionais a serem selecionados são os mesmos da parte não sazonal do modelo. Ou seja, deve-se selecionar os parâmetros  $P$ ,  $D$  e  $Q$  para fins de modelagem. A Figura 2 apresenta um exemplo de análise de uma série temporal com tendência e suas respectivas transformações via diferenças.

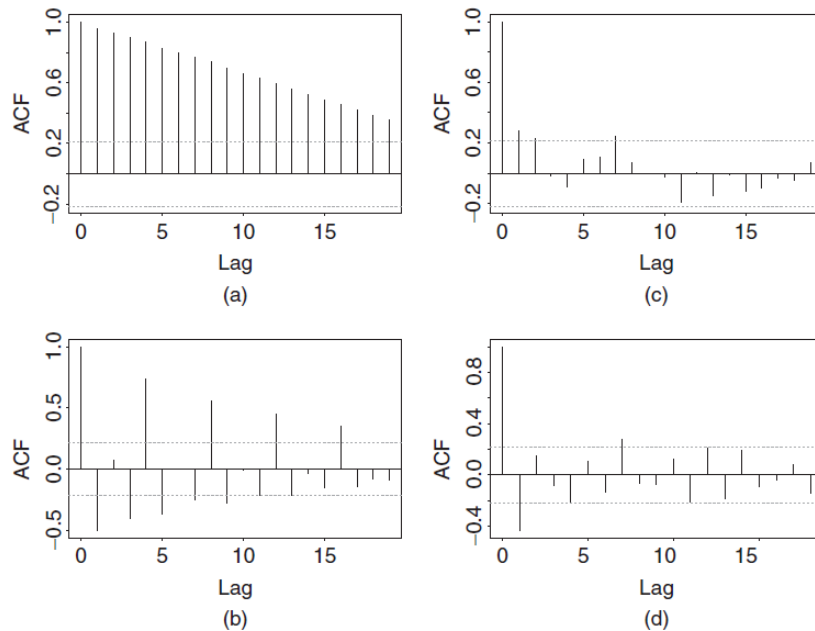


Figura 2. Funções de autocorrelação de uma série temporal com tendência e transformações.  
Fonte: Hyndman e Athanasopoulos (2018).

Em (a) é apresentada uma função de autocorrelação de uma série com tendência. Em (b) verifica-se a primeira diferença da série. Já em (c) aplicou-se a primeira diferença sazonal na série temporal e em (d) observa-se função de autocorrelação com as primeiras diferenças da série e em relação à parte sazonal.

Computacionalmente, um modelo SARIMA é definido por meio dos parâmetros  $p$ ,  $d$ ,  $q$ , que corresponde à série como um todo e aos parâmetros  $P$ ,  $D$ ,  $Q$  que corresponde à parte sazonal da série temporal em análise. Define-se também qual é a frequência ( $F$ ) em que ocorre a

sazonalidade na série. Uma notação comum para um modelo SARIMA é, portanto, SARIMA( $p, d, q$ ) ( $P, D, Q$ )<sub>F</sub>.

#### 2.1.4 Métricas para previsão de séries temporais

Diversas métricas para avaliação têm sido propostas na literatura e há muitas recomendações para o uso de cada uma desses indicadores. No entanto, há um consenso geral de que cada métrica tem suas vantagens e desvantagens, tornando-se um desafio a escolha de qual base utilizar, uma vez que diferentes métricas podem levar a diferentes conclusões referentes a qual modelo é mais adequado para previsão de séries temporais (KOLASSA, 2020).

Além disso, essas métricas têm robustez diferentes. Aquelas que estão relacionados à mediana têm maior robustez a *outliers* em comparação com as métricas que são calculadas com base na média (Hyndman & Athanasopoulos, 2018).

Alguns indicadores que podem ser usados para avaliar o poder preditivo de modelos de séries temporais são: *Mean Absolute Percent Error* (MAPE), *Mean Squared Error* (MSE), *Root Mean Squared Error* (RMSE), *Median Absolute Error* (MAE), *Root Squared Error* (RSE) etc. (HYNDMAN *et al.*, 2020; KOUTSANDREAS *et al.*, 2022).

Considerando uma série temporal com  $t$  observações, o MAPE é a métrica mais utilizada nesse tipo de pesquisa. Uma vez que o MAPE é baseado em valores percentuais, ele aplica uma penalidade diferente para erros positivos e erros negativos, sendo, portanto, assimétrico. (MAKRIDAKIS, 1993) considera este fato uma forte deficiência desta métrica. O autor aponta também que o MAPE é muito instável quando o desvio se encontra em torno de zero.

A métrica MSE, cujo valor da raiz quadrada é chamado de RMSE, e a métrica MAE, têm a mesma base. Enquanto o MSE busca avaliar a qualidade preditiva em relação ao valor médio, o MAE busca avaliar a acurácia em relação à mediana. MSE é a média calculada considerando o quadrado das diferenças entre dois vetores, sendo um composto pelos valores reais e outro composto pelos valores previstos. As equações correspondentes ao MAPE, RMSE, MdAE e MAE são apresentadas nas Equações 15, 16, 17 e 18 respectivamente.

$$MAPE = \frac{\sum_{i=1}^h |y_t^* - y_t|}{\sum_{i=1}^h y_t} \quad (15)$$

$$RMSE = \frac{\sqrt{\sum_{i=1}^h (y_t^* - y_t)^2}}{h} \quad (16)$$

$$MdAE = \text{mediana}(y_t^* - y_t) \quad (17)$$

$$MAE = \frac{\sum_{i=1}^h |y_t^* - y_t|}{h} \quad (18)$$

For:  $y_t^*$  = valor predito;  $y_t$  = valor atual;  $h$  = horizonte de previsão.

## 2.2 ANÁLISE POR ENVOLTÓRIA DE DADOS

A Análise por Envoltória de Dados (DEA) consiste em um método proposto por Charnes, Cooper e Rhodes (1978) com o objetivo de comparar a performance, inicialmente de unidades produtivas (DMUs), quanto à utilização de um conjunto de *inputs* e *outputs*.

A técnica consiste na escolha de uma amostra de insumos e produtos de distintas unidades tomadoras de decisão. Para Thanassoulis (2015) a técnica identifica, por meio de sucessivas programações lineares<sup>2</sup>, pesos que criam *outputs* e *inputs* virtuais, calculando desta forma a eficiência de diversas DMUs de atividades semelhantes, mas que utilizam processos de produção distintos.

A forma objetiva de cálculo dos pesos se deve ao fato de que os mesmos podem ser considerados como subjetivos, principalmente em análises onde carecem aplicações de estatística multivariada (THANASSOULIS, 2015). Esses pesos, através de combinação linear, criam o que se denominam os *inputs* e *outputs* virtuais que serão utilizados para comparar as DMUs.

Existem dois modelos clássicos de DEA: o Charnes, Cooper e Rhodes (CCR) que pressupõe retornos constantes de escala e o modelo de Banker, Charnes e Cooper (BCC) que possui como pressuposto o retorno não constante de escala. Este último expõe uma forte deficiência por considerar que DMUs que apresentem o menor valor de *input* já sejam automaticamente

---

<sup>2</sup> Programação Linear consiste em algoritmos para resolver problemas de otimização com restrições, em que uma função objetivo é linear em relação às variáveis de controle, e o domínio destas variáveis é definido por um sistema de inequações lineares (THANASSOULIS, 2015).

eficientes. Portanto, para aplicar este método de forma mais assertiva possível, deve-se levar em consideração o tamanho das DMUs (THANASSOULIS, 2015).

Considerando  $s$  DMUs em análise,  $m$  outputs e  $r$  inputs, os modelos CCR orientados a *inputs* e a *outputs* são formalizados segundo as Equações 19 e 20, respectivamente:

$$Max\ Eff_0 = \sum_{j=1}^m u_j y_{j0}$$

Sujeito a: (19)

$$\sum_{i=1}^r v_i x_{i0} = 1$$

$$\sum_{j=1}^m u_j y_{jk} - \sum_{i=1}^r v_i x_{ik} \leq 0, \forall k$$

$$v_i; u_j \geq 0, \forall i, j$$

$$Min\ Eff_0 = \sum_{j=1}^r v_j x_{j0}$$

Sujeito a: (20)

$$\sum_{i=1}^m u_i y_{i0} = 1$$

$$-\sum_{i=1}^r v_i x_{ik} + \sum_{j=1}^m u_j y_{jk} \leq 0, \forall k$$

$$u_j; v_i \geq 0, \forall j, i$$

Para:  $Eff_0$  = eficiência da  $DMU_0$ ;  $v_j$  e  $u_j$  = pesos de *inputs*  $i, i=1,2, \dots, r$  e *outputs*  $j, j=1,2, \dots, m$ ;  $x_{i0}$  e  $y_{j0}$  = *inputs*  $i$  e *outputs*  $j$  da  $DMU_0$ ;  $x_{ik}$  e  $y_{jk}$  = *inputs*  $i$  e *outputs*  $j$  das  $DMUs_k$ ;  $k=1,2, \dots, n$ ;  $x_{i0}$  e  $y_{j0}$  = *inputs*  $i$  e *outputs*  $j$  da  $DMU_0$ .

Já os modelos BCC são apresentados nas Equações 21 e 22:

$$Max\ Eff_0 = \sum_{j=1}^m u_j y_{j0} + u_*$$

Sujeito a: (21)

$$\sum_{i=1}^r v_i x_{i0} = 1$$

$$-\sum_{i=1}^r v_i x_{ik} + \sum_{j=1}^s u_j y_{jk} + u_* \leq 0, \forall k$$

$$v_i; u_j \geq 0, u_* \in \mathbb{R}$$

$$\text{Min } Eff_0 = \sum_{i=1}^r v_i x_{i0} + v_*$$

Sujeito a:

$$\begin{aligned} \sum_{i=1}^m u_i y_{i0} &= 1 \\ - \sum_{i=1}^r v_i x_{ik} + \sum_{j=1}^s u_j y_{jk} - v_* &\leq 0, \forall k \\ v_i; u_j &\geq 0, u_* \in \mathbb{R} \end{aligned} \tag{22}$$

Para:  $Eff_0$  = eficiência da  $DMU_0$ ;  $v_j$  e  $u_j$  = pesos de *inputs*  $i, i=1,2, \dots, r$  e *outputs*  $j, j=1,2, \dots, m$ ;  $x_{i0}$  e  $y_{j0}$  = *inputs*  $i$  e *outputs*  $j$  da  $DMU_0$ .  $x_{ik}$  e  $y_{jk}$  = *inputs*  $i$  e *outputs*  $j$  da  $DMU_k$ ;  $u_*$  e  $v_*$  = valores dos fatores de escala;  $k = 1,2, \dots, n$ .

A técnica consiste, portanto, em gerar  $m$  programações lineares com o objetivo de maximizar a produtividade de cada tomadora de decisão, com o intuito de eliminar a subjetividade na determinação dos pesos a serem empregados para a geração dos *outputs* e *inputs* virtuais.

### 2.3 MODELOS DE SUPEREFICIÊNCIA

Os modelos clássicos de DEA não conseguem discriminar as DMUs classificadas como tecnicamente eficientes. Ou seja, as DMUs com escore de eficiência igual a 1 são consideradas tecnicamente empatadas na análise. Andersen e Petersen (1993) propuseram retirar a DMU sob análise das restrições dos modelos CCR e BCC, permitindo que o escore das DMUs eficientes seja irrestrito.

As Equações 23 e 24 apresentam os modelos primais de DEA no contexto de supereficiência, orientados a *inputs* e a *outputs*, respectivamente.

*Min*  $\theta$

Sujeito a:

$$\begin{aligned} \sum_{k=1}^n \lambda_k x_{ik} + s_i^- &= \theta x_{i0}; \forall i, i = 1, 2, \dots, r \\ \sum_{k=1}^n \lambda_k y_{ik} - s_m^+ &= y_{m0}; \forall m, m = 1, 2, \dots, s \\ \lambda_k &\geq 0, \forall k, k \neq 0, \text{ se RCE} \end{aligned} \tag{23}$$

$$\sum_{k=1}^n \lambda_k = 1, \text{ se RVE}$$

$$\sum_{k=1}^n \lambda_k \leq 1, \text{ se RNC}$$

$$\sum_{k=1}^n \lambda_k \leq 1, \text{ se RND}$$

Max  $\phi$

Sujeito a:

$$\sum_{k=1}^n \lambda_k x_{ik} + s_i^- = x_{i0}; \forall i, i = 1, 2, \dots, r$$

$$\sum_{k=1}^n \lambda_k y_{ik} - s_m^+ = \theta y_{m0}; \forall m, m = 1, 2, \dots, s$$

$$\lambda_k \geq 0, \forall k, k \neq 0, \text{ se RCE}$$

$$\sum_{k=1}^n \lambda_k = 1, \text{ se RCE} \tag{24}$$

$$\sum_{k=1}^n \lambda_k \leq 1, \text{ se RNC}$$

$$\sum_{k=1}^n \lambda_k \leq 1, \text{ se RND}$$

Para:  $\lambda_k$  = peso que levará os vetores  $x$  à fronteira;  $s_i^-$  = folga dos *inputs*;  $s_i^+$  = folga dos *outputs*, *RCE* = Retorno Constante de Escala; *RVE* = Retorno Variável de Escala, *RNC* = Retorno Não Crescente de Escala; *RND* = Retorno não Decrescente de Escala.

#### 2.4 PROBLEMAS DE CONVEXIDADE EM ANÁLISE POR ENVOLTÓRIA DE DADOS

Um dos pressupostos para a implementação de um modelo de DEA em um determinado contexto é a convexidade (BOGETOFT & OTTO, 2011). Entretanto, como mostrado por Emrouznejad e Amin (2009) e Olesen, Petersen e Podinovski (2015) quando há variáveis fracionadas no contexto de DEA o pressuposto da convexidade não é atendido.

Um modo de tratar esse problema é utilizar o modelo denominado de *Free Disposal Hull* (FDH), proposto por Deprins, Simar e Tulkens (2006). Outra forma de lidar com variáveis fracionárias em contexto de DEA é usar o volume, colocando o numerador ou o denominador como *inputs* ou *outputs*, como apresentado em Emrouznejad e Amin (2009).

Despić, Despić e Paradi (2007) sugeriram um modelo denominado DEA-R (DEA-*Ratio*) que cria as frações de todas as combinações de *outputs* com todos os *inputs*, transformando os *inputs* em um vetor unitário **I**. Considerando  $n$  DMUs a serem analisadas com  $s$  outputs e  $m$  inputs. A matriz **R** é determinada por  $r_{(i,j)p} = y_{jp}/x_{ip}$ . Portanto, a matriz **R** possui dimensão  $(s \times m) \times n$ , enquanto a matriz **I** possui dimensão  $1 \times n$ . A Equação 25 formaliza o modelo DEA-R matematicamente.

$$Max\ Eff_0 = \sum_{j=1}^m c_{(j,i)} r_{(j,i)0} \tag{25}$$

Sujeito a:

$$\sum_{j=1}^s \sum_{i=1}^m c_{(j,i)} r_{(j,i)k} \leq 1, \forall k$$

$$c_{(j,i)} \geq 0, \forall i, j$$

## 2.5 DEA COMO CRITÉRIO DE SELEÇÃO DE MODELOS DE PREVISÃO

Emrouznejad, Rostami-Tabar e Petridis (2016) propuseram um modelo DEA-Multiplicativo adequado à presença de variáveis fracionárias para medir a eficiência dos modelos mais utilizados pela competição de previsão M-3 (MAKRIDAKIS & HIBON, 2000).

Zheng e Padmanabhan (2007) utilizaram DEA para gerar pesos que combinassem modelos de classificação, utilizando a matriz de confusão como inputs. Os modelos misturados foram: Redes Neurais Artificiais, *Support Vector Machines*, *Naive Bayes*, *Stumps*, Regressão Logística etc.

Narimani e Narimani (2012) propuseram um modelo DEA para escolha dos parâmetros do modelo ARIMA. Modificando os valores dos parâmetros  $p$  e  $q$ , cada uma representa uma DMU. No entanto, os autores criaram um algoritmo que considera os *ensembles* apenas dos modelos simples eficientes. Ao restringir a avaliação inicial apenas para os modelos simples de ARIMA e não considerar todos os *ensembles* possíveis como DMUs, limita-se o poder de discriminação



que o modelo pode fornecer. Além disso, foram utilizados modelos clássicos de DEA que ferem os pressupostos de convexidade ao se utilizar variáveis fracionadas.

## 2.6 DESENHO E ANÁLISE DE EXPERIMENTOS

O desenho de experimentos (DOE) é uma técnica onde o experimentador realiza um certo número de experimentos planejados que permite fazer conclusões estatísticas confiáveis e importantes (LIN, PULIDO & ASPLUND, 2015). O número de experimentos a serem realizados é certamente menor do que àqueles em uma abordagem de tentativa e erro.

Os desenhos fatoriais do tipo  $2^k$ , em que  $k$  representa o número de fatores a serem considerados, são amplamente utilizados em aplicações científicas. Neste tipo de *design*, um modelo de primeira ordem com interações é gerado, ajudando a entender os efeitos principais dos fatores sobre a variável resposta. Além disso, o modelo contribui para entender como os fatores atuam de forma interativa sobre a resposta (MONTGOMERY, 2017).

Em um desenho fatorial  $2^k$ , os níveis dos fatores são usualmente definidos como “baixo” e “alto”. Todas as combinações possíveis desses níveis são investigadas em uma série de testes (YONDO, ANDRÉS & VALERO, 2018). Nestes testes, os níveis dos fatores considerados são alterados pelo pesquisador de forma controlada a fim de observar os efeitos de cada fator e das interações (STAICULESCU *et al.*, 2005).

Normalmente, em um projeto fatorial, os pontos, representando cada execução do projeto estruturado, são definidos nos vértices de um hipercubo, conforme ilustrado na Figura 3 (MONTGOMERY, 2017). Vale ressaltar que (1) indica o ponto em que todos os fatores estão definidos em seus níveis mais baixos e em todos os outros vértices as letras indicam quais fatores estão definidos em seus níveis mais altos. De acordo com Yondo, Andrés e Valero

(2018), os experimentos industriais têm se concentrado principalmente em projetos de 2 e 3 níveis, resultando em  $2^k$  e  $3^k$  experimentos, respectivamente.

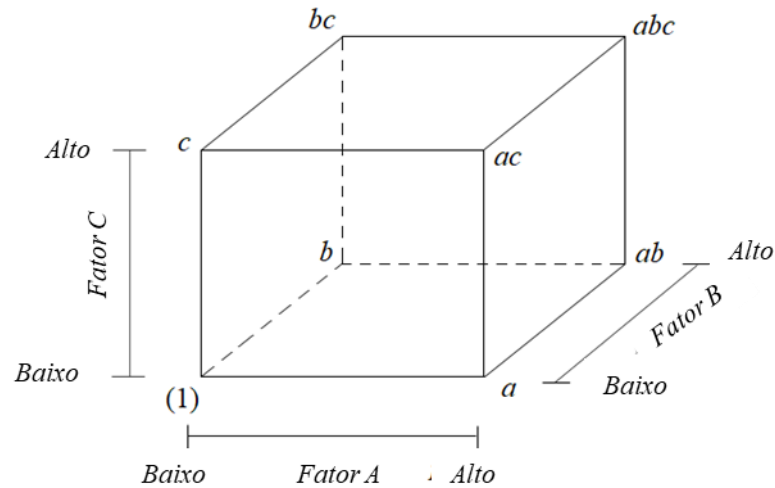


Figura 3. Representação geométrica de desenho fatorial  $2^3$ .  
Fonte: adaptado de Montgomery (2017).

## 2.7 DELINEAMENTO DE MISTURAS E METODOLOGIA DE SUPERFÍCIE DE RESPOSTA

Em muitos casos os fatores possuem como restrição que seus somatórios sejam iguais a um. Ou seja, considerando  $n$  fatores:

$$\sum_{i=1}^n x_i = 1 \quad (26)$$

Nota-se que o problema de dar pesos a modelos de previsão de séries temporais pode ser claramente tratado como um problema de projeto de delineamento de mistura. Em um problema de análise de experimentos, variáveis explicativas (ou fatores) são fixadas em certos níveis para estudar seu comportamento em uma ou mais variáveis de resposta. Em um projeto de mistura, as proporções são modificadas para analisar o comportamento da variável resposta.

Em um problema de mistura com três componentes, por exemplo, a região de soluções factíveis e o sistema de coordenadas podem ser representados por um plano triangular, conforme mostrado nas Figura 4.

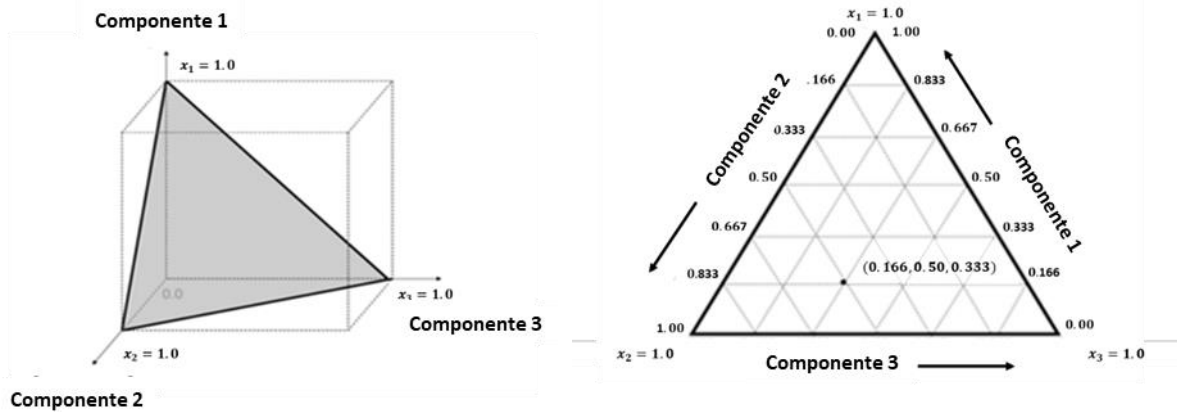


Figura 4. Representação gráfica de design de misturas.  
 Fonte: adaptado de (LAWSON, 2014).

Portanto, pode-se estimar via Mínimos Quadrados Ordinários uma superfície de resposta projetada para cortes paralelos à base triangular (LAWSON, 2014). Em geral, arranja-se uma relação funcional entre as variáveis de resposta e as proporções, de modo que a Equação 27 pode ser escrita de forma que:

$$y_i = \varphi(x_1, x_2, \dots, x_n) \quad (27)$$

Onde  $y_i$  é a variável dependente,  $x_i$  with  $i = 1, 2, \dots, n$ , são as variáveis explanatórias ou proporções, e finalmente,  $\varphi$  é a forma funcional. O problema de mistura consiste então em determinar um modelo que represente a superfície de resposta. Considerando as restrições impostas ao modelo, pode-se escolher entre os seguintes experimentos padrão: rede simplex, projeto centroide simplex ou projeto de vértices extremos.

Seja um problema que apresente  $n$  componentes de mistura. O modelo de regressão linear pode ser representado pela Equação 28.

$$y = \beta_0 + \sum_{i=1}^n \beta_i x_i + \varepsilon \quad (28)$$

Porém, como  $\sum_{i=1}^n x_i = 1$ , ganha-se um grau de liberdade e um dos  $n + 1$  parâmetros da Equação 28 passa a ser redundante. (SCHEFFÉ, 1958) propôs então a reescrita do modelo de regressão conforme a Equação 29.

$$y = \sum_{i=1}^n \beta_i^* x_i + \varepsilon \quad (29)$$

Nesta forma do modelo,  $\beta_0$  na Eq. (18) é igual a 1 e é substituído por  $\sum_{i=1}^n x_i$ . Importante destacar que os coeficientes da Equação 28 não representam os efeitos das variáveis explanatórias consideradas. Além disso, um coeficiente não ser estatisticamente diferente de zero, não significa que não apresente interferência a resposta do modelo. Porém, no modelo de delineamento de misturas, os coeficientes representam a resposta predita no *vertex* da região experimental. A Figura 5 ilustra um caso considerando um modelo de três componentes de misturas,

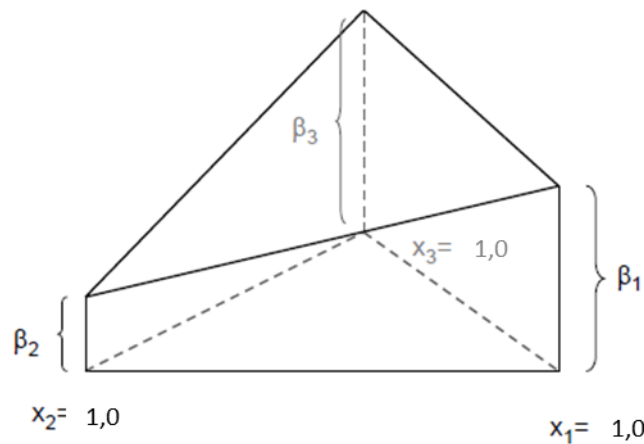


Figura 5. Interpretação dos coeficientes em um modelo linear de Scheffé.  
Fonte: adaptado de (LAWSON, 2014).

Se a superfície de resposta apresentar curvatura, o modelo geral quadrático na Equação 30.

$$y = \sum_{i=1}^n \beta_i x_i + \sum_{i=1}^n \sum_{i < j}^n \beta_{ij} x_i x_j + \varepsilon \quad (30)$$

Desta maneira,  $x_i x_j$  representa a curvatura entre os vértices da região experimental, como apresentado na Figura 5 para uma exemplificação de um modelo com três componentes de mistura.

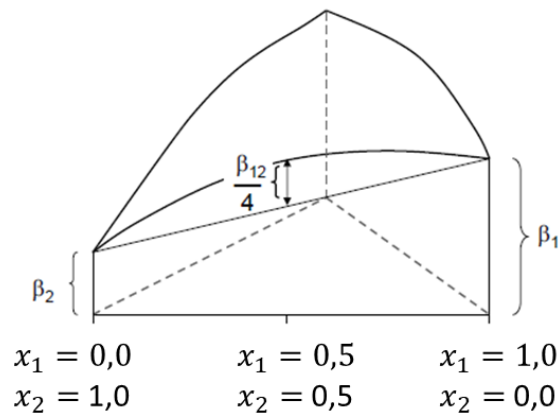


Figura 6. Interpretação dos coeficientes no modelo quadrático de Scheffé.  
 Fonte: adaptado de Lawson (2014).

Baseado na interpretação dos coeficientes dos modelos de desenho de misturas, (SCHEFFÉ, 1958) propôs o desenho simplex-lattice para experimentos com misturas. Este desenho consiste na combinação de todas as proporções cuja  $i$ -ésima proporção  $x_i = 0, \frac{1}{m}, \frac{2}{m}, \dots, 1$ . Desta maneira, um Simplex Lattice Design (SLD) com  $k$  componentes e  $m + 1$  proporções igualmente espaçadas é designado como um desenho SLD ( $k, m$ ).

Como exemplo, a Figura 7 apresenta um SLD (3, 1) e um SLD (3, 2), à esquerda e à direita respectivamente. Importante mencionar que o desenho SLD (3, 1) possui apenas os pontos extremos fazendo com que uma função linear seja suficiente para sua modelagem. Com o intuito de testar a curvatura da função, deve-se adicionar os pontos axiais mostrados no SLD (3, 2).

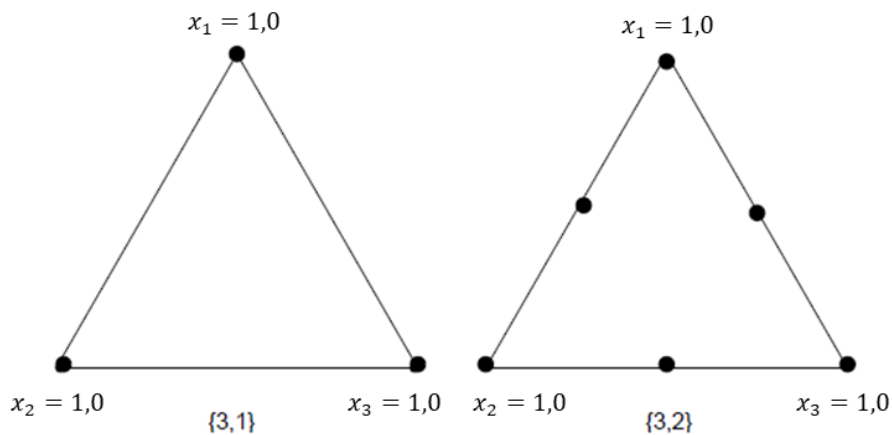


Figura 7. Desenhos SLD linear e SLD quadrático.  
 Fonte: adaptado de Lawson (2014).

Dada a equação de superfície de resposta, se a mesma apresentar curvaturas, pode-se otimizar a mesma por meio de um algoritmo de otimização. Modelos de superfície de resposta correspondem a uma coleção de ferramentas matemáticas e estatísticas que avaliam o comportamento de uma variável resposta, dada a modificação de forma controlada de variáveis explanatórias, chamadas de fatores neste contexto.

## 2.8 ESTIMAÇÃO DA SUPERFÍCIE DE RESPOSTA

Basicamente, uma superfície de resposta corresponde a um modelo de regressão linear aplicado a dados experimentais. Conforme exposto na seção 2.6, os parâmetros do modelo de superfície de resposta podem ser estimados via Mínimos Quadrados Ordinários. Outro método de estimação de parâmetros comumente utilizado é o método de Máximoverossimilhança (MONTGOMERY & RUNGER, 2010)

Sendo  $n$  o número de observações disponíveis, e  $k$  o número de parâmetros a serem estimados em um modelo do tipo da Equação 31, com exceção de  $\beta_0$ , para  $n > k$ . Segundo Montgomery e Runger (2010) a função de mínimos quadrados é dada pela Equação 32.

$$L = \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n (y_i - \beta_0 - \sum_{j=1}^k \beta_j x_{ij})^2 \quad (31)$$

O problema passa a ser minimizar  $L$  com relação a seus parâmetros. Assim:

$$\frac{\partial L}{\partial \beta_j} = -2 \sum_{i=1}^n (y_i - \hat{\beta}_0 - \sum_{j=1}^k \hat{\beta}_j x_{ij}) = 0 \quad (32)$$

Para que os estimadores de Mínimos Quadrados Ordinários para os coeficientes de regressão múltipla sejam não enviesados, eficientes e consistentes o modelo deve atender aos pressupostos de Gauss-Markov (WOOLDRIDGE, 2015), que são listados abaixo:

- 1) A relação entre  $Y$  e  $X$  é linear nos parâmetros;
- 2)  $X$  é uma variável não estocástica cujos valores são fixos;
- 3) O termo de Erro é independente e igualmente distribuído (*i.i.d.*) e tem valor esperado zero,  $E(e|X) = 0$ ;

- 4) O termo de erro tem variância constante para todas as observações (homocedasticidade);
- 5) Não há correlação entre os erros do modelo de regressão,  $cov(e_t, e_{t+s}) = 0$ ;
- 6) O termo de erro tem distribuição normal.

Dado os pressupostos de Gauss-Markov,  $\hat{\beta}_j \sim N(\beta_j, \sigma_{\hat{\beta}_j}^2)$ . Como são os betas amostrais são variáveis aleatórias, eles possuem uma distribuição densidade de probabilidade, que segue uma distribuição normal possuindo média e variância. Pode-se, então, realizar inferência intervalar e testes de hipótese sobre os betas amostrais por intermédio da distribuição *gaussiana*.

Apesar de existirem testes de hipótese para a análise dos pressupostos de Gauss-Markov<sup>3</sup>, Montgomery, Peck e Vining (2021) indicam que em muitos casos a análise gráfica dos resíduos, por meio de *Scatterplots*, Q-Q Plots e Histograma, pode ser bem informativa.

## 2.9 ESCOLHA DE MODELOS DE REGRESSÃO E DE VARIÁVEIS REGRESSORAS

Um dos desafios ao se criar modelos de regressão é selecionar as variáveis que compõem o modelo, fazendo com que o analista ou o pesquisador esteja em um *trade-off* entre ajuste, poder preditivo e o consumo de graus de liberdade do modelo. (GUJARATI, 2011) elenca alguns critérios úteis para seleção de modelos de regressão: o critério do  $R^2$ , do  $R^2$ -ajustado, o critério do  $C_p$  de Mallows, Critério de Informação de Aikake (AIC) e o Critério de Informação de Schwarz (SIC).

### 2.9.1 Critério de $R^2$

O  $R^2$ , muitas vezes chamado de coeficiente de determinação, é um índice que retorna a quantidade percentual da variabilidade da variável resposta que é explicada pelo modelo de regressão proposto. Dentre os critérios mencionados, este é o único que não penaliza a inclusão de novos regressores ao modelo. A Equação 33 apresenta como é realizado o cálculo do coeficiente.

---

<sup>3</sup> Podem-se citar como exemplos de testes para normalidade: Anderson-Darling, Cramer-Von Mises, Komolgorov-Smirnov, Shapiro-Francia e Shapiro-Wilk. Para testes de homocedasticidade dos resíduos, pode-se utilizar o teste de Bresch-Pagan (GUJARATI, 2011).

$$R^2 = 1 - \frac{SQR}{SQT} \quad (33)$$

Onde: SQR = Soma dos Quadrados dos Resíduos; SQT = Soma dos Quadrados Totais.

### 2.9.2 Critério de $R^2$ -Ajustado

Conforme exposto, o  $R^2$  é um índice que não penaliza a inclusão de novas variáveis regressoras. Entretanto, conforme se aumenta o número de regressores, maior quantidade de graus de liberdade são gastos e, conseqüentemente, maiores as chances de ocorrência de *overfitting*.

O  $R^2$ -Ajustado é apresentado pela Equação 34:

$$R^2 = 1 - \frac{SQR/(n - k)}{SQT/(n - 1)} \quad (34)$$

Sendo  $n$  a quantidade de observações e  $k$  o número de parâmetros a serem estimados pelo modelo.

### 2.9.3 Critério de Informação de Akaike (AIC)

O AIC é calculado segundo a Equação 35.

$$\ln AIC = \left(\frac{2k}{n}\right) + \ln \left(\frac{SQR}{n}\right) \quad (35)$$

Como critério de seleção de modelos concorrentes, deve-se selecionar aquele que possui o menor AIC.

### 2.9.4 $C_p$ de Mallows

O  $C_p$  de Mallows compara diversos modelos concorrentes diminuindo o tamanho do modelo em relação ao possível modelo cheio. Considerando um modelo cheio com possíveis  $k$  regressores e que se escolha  $p$  regressores, sendo  $p \leq k$ . O indicador é calculado segundo a Equação 36.



$$C_p = \frac{SQR_p}{\hat{\sigma}^2} - (n - 2p) \quad (36)$$

Para  $SQR_p$  = Soma dos Quadrados dos Resíduos para o modelo  $p$  considerado e  $\hat{\sigma}^2$  = variância do resíduo.

Quando se seleciona modelos por meio do  $C_p$  de Mallows procura-se um modelo com baixo índice que seja próximo a  $p$  e que tenha um bom ajuste.

A seleção de variáveis explanatórias a fazerem parte de um modelo de regressão pode ser uma tarefa desafiadora (SHARPE, DE VEAUX & VELLEMAN, 2021). Segundo os autores, os melhores modelos de regressão, além de atenderem aos pressupostos de Gauss-Markov, devem:

- a) Possuir relativamente poucos previsores para manter o modelo o mais simples possível, mas sem perder tanta informação (critério da Navalha de Okham);
- b) Um  $R^2$  relativamente alto, o que indica que muita da variabilidade de  $y$  é explicada pelo modelo;
- c) Um valor relativamente pequeno do desvio-padrão dos resíduos ( $s_e$ )<sup>4</sup>;
- d) Pequenos p-valores em relação aos testes F do modelo em geral e  $t$  de cada coeficiente, o que indica que o modelo ajustado é melhor para explicar  $y$  do que  $\bar{y}$ ;
- e) Previsores medidos com alta confiabilidade e que não sejam altamente correlacionados entre si.

Dado esses critérios de seleção de modelos, e tendo uma quantidade considerável de possíveis previsores disponíveis, há algoritmos de seleção de modelos que facilitam a escolha do mesmo.

---

<sup>4</sup>  $s_e = \sqrt{\sum_{i=1}^n \varepsilon_i / (n - k)}$ , para  $\varepsilon_i$  = resíduo da  $i$ -ésima observação,  $n$  = número de observações e;  $k$  = número de parâmetros estimados pelo modelo.

O algoritmo dos *Melhores Subconjuntos*, criado por Hocking e Leslie (1967), analisa todos os subconjuntos possíveis de modelos de regressão que podem ser criados e seleciona aquele que retorna o melhor critério de seleção de modelo escolhido pelo pesquisador. O problema deste algoritmo é que ele se torna muito custoso computacionalmente quando o número de variáveis explanatórias elegíveis é grande. Montgomery e Runger (2010) afirmam que se houver  $k$  variáveis candidatas, o número de modelos a serem ajustados é dado por  $2^k$ .

Outra alternativa é utilizar algoritmos que constroem os modelos de regressão em etapas (ou *passo a passo*). Segundo Wang *et al.* (2016) essa técnica evita a inclusão de variáveis explanatórias que explicam o mesmo que outras variáveis ou possuem pouca importância para explicar a variável resposta. Os autores argumentam também que esses algoritmos podem evitar o *overfitting*. Há, basicamente, três algoritmos para criação de modelos de regressão por etapas: *backward selection*, *forward selection* e *stepwise selection* (NORYANI *et al.*, 2019).

O algoritmo de *backward* se inicia com todos os  $k$  regressores candidatos. A variável cujo coeficiente é o menos significativo em relação à estatística  $F$  parcial é removido se a sua remoção faz com que a estatística  $F$  do modelo melhore significativamente.

O algoritmo *forward* começa com uma única variável explanatória (a que possui maior correlação com a resposta, por exemplo) e os demais regressores vão sendo adicionados, um a um, até que não haja mais variáveis explicativas que produzam um aumento significativo na soma dos quadrados explicados da regressão.

Por fim, a seleção de variáveis do tipo *stepwise* é uma junção dos algoritmos de *backward* e *forward*. Adicionando e removendo regressores, o algoritmo constrói de forma iterativa uma série de modelos em cada etapa. O teste  $F$  parcial é mais uma vez utilizado para decidir sobre a exclusão ou a inclusão de determinada variável.

## 2.10 OTIMIZAÇÃO DE SUPERFÍCIES DE RESPOSTA

### 2.10.1 Otimização Multiobjetiva e Análise de Componentes Principais

Como há várias métricas para avaliar o poder preditivo de um modelo de série temporal, pode-se utilizar modelos de otimização adequados para minimizar de forma conjunta essas

métricas. Um modelo de otimização multiobjetivo pode ser resumido de forma geral segundo a Equação 37:

$$\min_{\mathbf{X}} \{f_1(\mathbf{X}), f_2(\mathbf{X}), \dots, f_k(\mathbf{X})\} \quad (37)$$

Sujeito a:  $\mathbf{X} \in \mathcal{S}$

Para  $k > 2$  funções objetivo em  $f_i: \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $\mathbf{X}$  = vetor de variáveis de decisão  $x = (x_1, x_2, \dots, x_n)^T$  no espaço viável não vazio  $S \subset \mathbb{R}^n$ .

Há várias formas de tratar o problema apresentado na Eq. (14), como o método de *e-constraint* (Haimes *et al.*, 1971), que visa otimizar selecionando-se a função mais importante e transferindo as demais funções para o conjunto de restrição do modelo.

Outro método possível proposto por (FISHBURN, 1974) é a Ordem Lexográfica, que classifica as funções segundo suas respectivas importâncias. As funções objetivo consideradas mais importantes são otimizadas por etapas e posteriormente transferidas para as restrições dos problemas subsequentes.

Gass e Saaty (1955) e Zadeh (1963) propuseram resolver o problema de otimização multiobjetiva de forma ponderada, atribuindo pesos  $w_1, w_2, \dots, w_n$  às funções  $f_1(\mathbf{X}), f_2(\mathbf{X}), \dots, f_k(\mathbf{X})$  e transformando o problema em mono-objetivo. Desta maneira, a Equação 37 poderia ser reescrita segundo a Equação 38.

$$\min_{\mathbf{X}} \sum_{i=1}^k w_i f_i(\mathbf{X}) \quad (38)$$

Sujeito a  $\mathbf{X} \in \mathcal{S}$

Conforme pode ser observado, as três soluções exemplificadas apresentam o problema de definir de forma arbitrária a importância (ou pesos) das funções que fazem parte do problema. Com o objetivo de perpassar este problema (GOMES *et al.*, 2012) sugeriram a aplicação de Análise de Componentes Principais (PCA) para unificar as respostas quando as mesmas apresentarem alta correlação.

A PCA baseia-se na escrita de componentes principais dependentes exclusivamente da matriz de covariância das variáveis originais. Embora  $p$  componentes sejam necessários para

explicar a variabilidade total dos dados, geralmente  $k$  componentes são necessários para explicar a variabilidade total dos dados, onde  $k < p$ . (JOHNSON & WICHERN, 2007).

Sendo  $\mathbf{X}^T = [X_1, X_2, \dots, X_p]$  um vetor aleatório de variáveis originalmente consideradas no problema cuja matriz de variância e covariância é dada por  $\Sigma$  com autovalores  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$ , as seguintes combinações lineares mostradas na Equação 39 podem ser obtidas.

$$\begin{aligned} Y_1 &= \mathbf{a}_1^T \mathbf{X} = a_{11}X_1 + a_{12}X_2 + \dots + a_{1p}X_p \\ Y_2 &= \mathbf{a}_2^T \mathbf{X} = a_{21}X_1 + a_{22}X_2 + \dots + a_{2p}X_p \\ &\vdots \\ Y_p &= \mathbf{a}_p^T \mathbf{X} = a_{p1}X_1 + a_{p2}X_2 + \dots + a_{pp}X_p \end{aligned} \quad (39)$$

É conhecido que  $Var(Y_i) = \mathbf{a}_i^T \Sigma \mathbf{a}_i$  e  $Cov(Y_i, Y_k) = \mathbf{a}_i^T \Sigma \mathbf{a}_k$  with  $i, k = 1, 2, \dots, p$ . Assim, o primeiro componente principal é definido como a combinação linear  $\mathbf{a}_1^T \mathbf{X}$  que maximiza o valor de  $Var(\mathbf{a}_1^T \mathbf{X})$  sujeito a  $\mathbf{a}_1^T \mathbf{a}_1 = 1$ . O segundo componente principal é dado pela combinação linear  $\mathbf{a}_2^T \mathbf{X}$  que maximiza o valor de  $Var(\mathbf{a}_2^T \mathbf{X})$  sujeito a  $\mathbf{a}_2^T \mathbf{a}_2 = 1$  e  $Cov(\mathbf{a}_1^T \mathbf{X}, \mathbf{a}_2^T \mathbf{X}) = 0$ . Assim, o  $i$ -ésimo componente principal pode ser obtido por meio da combinação linear  $\mathbf{a}_i^T \mathbf{X}$  que maximiza  $Var(\mathbf{a}_i^T \mathbf{X})$  sujeito a  $\mathbf{a}_i^T \mathbf{a}_i = 1$  e a  $Cov(\mathbf{a}_i^T \mathbf{X}, \mathbf{a}_k^T \mathbf{X}) = 0$  para  $k < i$ .

Schumacker (2015) indica que a adequabilidade para a aplicação das técnicas de redução de dimensionalidade depende de alguns fatores: um número de correlações de Pearson acima ou igual a 0,30 relativamente alta na matriz de correlação, estatística de *Kaiser-Meyer-Olkin* (KMO)<sup>5</sup> no mínimo igual a 0,60 e significância estatística do teste de esfericidade de Bartlett<sup>6</sup>.

Uma questão importante é o número de componentes a serem retidos no estudo. Uma ferramenta útil neste contexto é o *screeplot*. Aqui, os autovalores são ordenados do maior para o menor e geralmente é escolhido manter um número de componentes até o ponto em que os autovalores restantes são relativamente pequenos e de aproximadamente o mesmo tamanho.

<sup>5</sup>  $KMO = \frac{\sum_{l=1}^k \sum_{c=1}^k \rho_{lc}^2}{(\sum_{l=1}^k \sum_{c=1}^k \rho_{lc}^2 + \sum_{l=1}^k \sum_{c=1}^k \varphi_{lc}^2)}$ , para:  $\rho$  = correlação de Pearson;  $\varphi$  = correlação parcial de Pearson.

<sup>6</sup> O teste compara os elementos fora da diagonal principal da matriz de correlação dos dados com uma matriz identidade. Se forem estatisticamente diferentes a um nível de significância  $\alpha$  determinado, considera-se adequada a redução da dimensionalidade.

### 2.10.2 Desirability

O algoritmo selecionado para resolver o problema de otimização proposto foi o *Desirability*, proposto por Derringer e Suich (1980). O método consiste em converter as repostas em funções *desirability* que variam de 0 a 1, ou seja, da menor conveniência para a maior conveniência. Desta maneira, as melhores condições de resposta são obtidas simultaneamente, maximizando ou minimizando as funções objetivo dependendo da situação mais conveniente para cada função (WANG & WAN, 2009).

Dado que cada função pode ser reescrita por meio de uma função  $d_i(Y_i)$ , variando de 0 a 1, a Equação 40 apresenta como pode ser calculado o índice global de desejabilidade.

$$D = (D_1(Y_1) \times D_2(Y_2) \dots \times D_k(Y_k))^{1/k} \quad (40)$$

Desta forma, o ótimo é alcançado pela maximização da função desejabilidade geral, que é função das funções desejabilidades individuais de cada função objetivo. A vantagem da utilização da média geométrica dessas funções, ainda segundo o autor, é que permite atingir de forma balanceada a solução global.

Portanto, a fórmula de transformação  $D_i$  para uma função objetivo que se deseja maximizar é:

$$D_i = \begin{cases} 0, & \hat{Y}_i < L_i \\ \left[ \frac{\hat{Y}_i - L_i}{U_i - L_i} \right]^R, & L_i \leq \hat{Y}_i \leq U_i \\ 1, & \hat{Y}_i > U_i \end{cases} \quad (41)$$

Para:  $U_i$  = limite superior da resposta;  $L_i$  = limite inferior da resposta,  $R$  = peso definindo o quão importante é para  $\hat{Y}_i$  estar próximo do máximo.

Quando se deseja minimizar a função objetivo,  $D_i$  possui a seguinte função:

$$D_i = \begin{cases} 1, & \hat{Y}_i < L_i \\ \left[ \frac{U_i - \hat{Y}_i}{U_i - L_i} \right]^S, & L_i \leq \hat{Y}_i \leq U_i \\ 0, & \hat{Y}_i > U_i \end{cases} \quad (42)$$

Para:  $S$  = peso definindo o quão importante é para  $\hat{Y}_i$  estar próximo do máximo.

Por fim, quando se deseja que o valor de determinada função objetivo se aproxime de um valor alvo  $T_i$ , tem-se:

$$D_i = \begin{cases} 0, & \hat{Y}_i < L_i \\ \left[ \frac{\hat{Y}_i - L_i}{U_i - L_i} \right]^R, & L_i \leq \hat{Y}_i \leq U_i \\ 1, & \hat{Y}_i = T_i \\ \left[ \frac{U_i - \hat{Y}_i}{U_i - L_i} \right]^S, & T_i \leq \hat{Y}_i \leq U_i \\ 0, & \hat{Y}_i > U_i \end{cases} \quad (43)$$

## 2.11 SIMULAÇÃO DE MONTE CARLO

A Simulação de Monte Carlo consiste em uma estratégia popular para comparação de métricas relativas à previsão de séries temporais (BOKDE, YASEEN & ANDERSEN, 2020). A Simulação de Monte Carlo, no contexto de previsão, consiste em gerar conjuntos de dados de forma aleatória a partir de um *Data Generating Process* (DGP) de modo a dar condições de comparar em diversos cenários as métricas de performance de previsão selecionadas.

Segundo Morettin e Bussab (2017) “uma simulação pode ser entendida como uma particular realização de um modelo. Neste sentido, os valores simulados podem ser considerados como uma amostra”. O nome “Monte Carlo” vem da cidade europeia de mesmo nome, famosa por seus cassinos. Os trabalhos pioneiros que utilizaram a técnica foram: Metropolis e Ulam (1949) e Von Neumann (1951), sendo ainda considerados importantes para o desenvolvimento da técnica os trabalhos de Hammersley e Handscomb (1964), Sobol (1976) e Ross (1997).

Robert e Casella (2008) afirmam que a base para toda simulação contínua é um valor inicial  $u_0$  que advém de uma distribuição uniforme  $U(0,1)$  e que é transformado por uma função  $D$ . Então  $u_i = D^i(u_0)$ , tendo o *output* dessa função as mesmas propriedades estatísticas, pois as observações são *i.i.d.*.

Como será visto na seção de *Aspectos Metodológicos*, este trabalho se valerá da referida técnica para comparar o método proposto com outro método de *ensemble* com o intuito de generalizar estatisticamente os resultados.

## 2.12 PREVISÃO DE DEMANDA DE ENERGIA ELÉTRICA

Com o aumento da necessidade de melhor planejamento energético nas últimas décadas a previsão de demanda de energia elétrica se torna essencial para um crescimento econômico sustentável (GHALEHKHONDABI *et al.*, 2017).

Uma previsão acurada de demanda de energia elétrica é importante também para as próprias empresas geradoras e distribuidoras de energia, auxiliando nas decisões de curto e de longo prazo das mesmas (SUGANTHI & SAMUEL, 2012). Pelo prisma das geradoras, a energia não pode ser armazenada. Portanto a sustentabilidade das operações depende claramente de um constante equilíbrio entre a oferta e a demanda. Do ponto de vista das distribuidoras, uma previsão de demanda acurada baliza de maneira primordial as decisões relativas à operação e aos investimentos em capital.

Devido a esta importância, recentemente muitos trabalhos acadêmicos vêm trazendo muitas abordagens distintas a este problema de previsão. Suganthi e Samuel (2012) revisaram modelos baseados em estatística e econometria, como regressão e análise de séries temporais, e também modelos baseados em análise computacional, como Lógica Fuzzy, aplicação de métodos heurísticos (como Algoritmo Genético) e Redes Neurais Artificiais.

Ghosh (2008) aplicou modelos estatísticos para prever a demanda de energia elétrica no Norte da Índia. O autor comparou o modelo SARIMA com Holt-Winters multiplicativo e concluiu que o primeiro performou melhor que o segundo em termos de erros quadráticos e erros absolutos.

Fan, Xiao e Wang (2014) e Hong *et al.* (2011) utilizaram *Support Vector Regression* (SVR), tendo como variáveis explanatórias indicadores socioeconômicos, como população, evolução percentual do Produto Nacional Bruto (PNB) e balança comercial. Em ambos os casos o critério do RMSE foi utilizado para comparação entre os diversos modelos aninhados e o modelo SARIMA.

Tang *et al.* (2012) combinaram modelos SARIMA, Suavização Exponencial Sazonal e *Weighted Support Vector Machines* (WSVM) utilizando o algoritmo *Adaptative Particle Swarm* para otimização. Yanjun e Yuliang (2009) combinaram modelos de teoria cinza com modelos de relação causal de Granger. Em estudo simular, Liu, Cai e WU (2008) apresentaram uma metodologia em que combinam modelos de teoria cinza com um tipo de rede neural chamada Rede de Função de Base Radial.



### **3. ASPECTOS METODOLÓGICOS**

Segundo Malhotra (2019) essa pesquisa pode ser classificada como aplicada, em seu aspecto da natureza, quantitativa quanto à abordagem, explanatória quanto aos objetivos e experimental em relação aos métodos.

Conforme exposto, esta tese possui o objetivo de propor um novo algoritmo de *Machine Learning* de *ensemble* de modelos SARIMA levando em consideração a eficiência determinada por meio de um adequado modelo de Análise por Envoltória de Dados (DEA-R).

A metodologia proposta é apresentada na Figura 8 e segue a explicação de cada passo abaixo.

#### **A. Selecionar as séries temporais**

Inicialmente a série temporal a ser modelada deve ser selecionada pelo pesquisador. Logicamente mais de uma série pode ser selecionada, porém a metodologia proposta não considera uma análise de série temporal multivariada.

Essas séries podem ser séries temporais reais ou séries temporais sintéticas aplicadas para validar a metodologia. Não há restrições quanto ao número de observações sobre as séries ou suas características, como autocorrelação, porém os métodos a serem aplicados devem ser adequados para lidar com as séries selecionadas.

#### **B. Análise Exploratória dos Dados**

Conforme exposto no referencial teórico, a análise exploratória dos dados é um passo muito importante para o ajuste e previsão de séries temporais. Calcular estatísticas de posição e de dispersão, analisar o gráfico de série temporal com o intuito de se ter preliminarmente uma intuição da estrutura da série, como tendência, sazonalidade, ciclo e quebra estrutural da mesma.

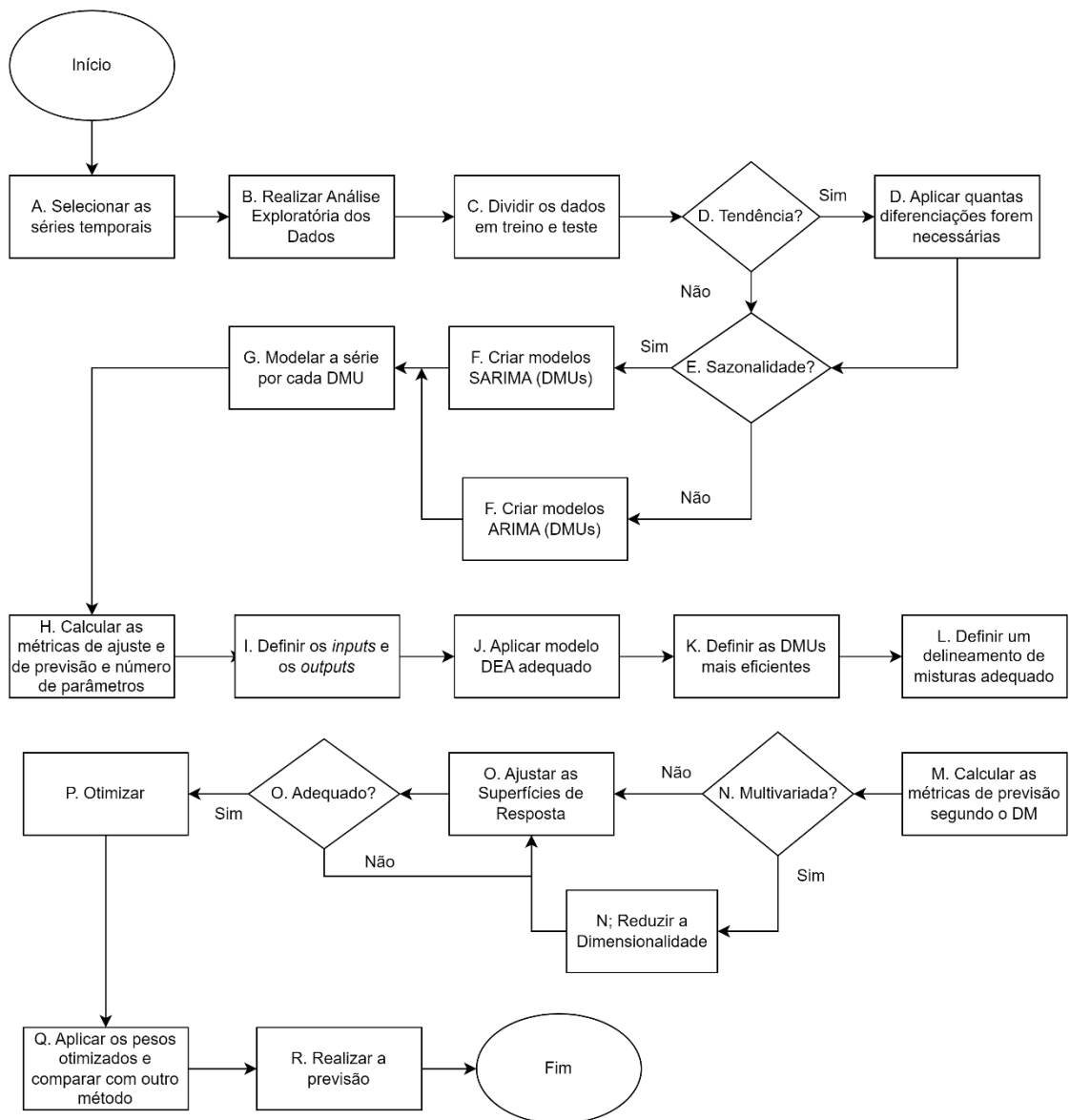


Figura 8. Fluxograma proposto da metodologia.  
Fonte: elaboração própria.

### C. Dividir os dados em treino e teste

Nesta etapa, o conjunto de dados, composto pelas séries temporais selecionadas, deve ser dividido em conjuntos de treinamento e teste. (JAMES *et al.*, 2013) consideram que um número adequado de observações para o conjunto de treino é de 70% a 80% dos dados. Consequentemente, o conjunto de teste terá entre 20% e 30% do número de observações. Contudo, Montgomery, Jennings e Kulahci (2015) recomendam deixar pelo menos 20 a 25 observações para testar o modelo.

#### D. Analisar Tendência

Segundo Fattah *et al.* (2018) um modelo de Box & Jenkins não pode ser aplicado a uma série temporal que apresente tendência. Desta maneira, deve-se analisar se a série possui tendência. Caso positivo, deve-se retirar a tendência por meio de diferenciações da mesma.

Há basicamente dois modos de verificar se uma série temporal não é constante ao longo do tempo. O primeiro por meio do gráfico da função de autocorrelação da série. Se ela decair de forma constante cada vez que o *lag* aumenta, esse é um indício que o tempo cria uma correlação entre as séries defasadas. A Figura 9 apresenta um exemplo de uma função de autocorrelação que dê indícios de que a série em análise possui tendência.

Se a análise da tendência der indícios de que a série temporal em análise possui tendência, a mesma deve ser retirada por meio de diferenciações (TSAY, 2005). A primeira e a segunda diferenças são dadas por:

$$\Delta y_j = y_{j+1} - y_j \quad (44)$$

$$\Delta y_j = y_{j+2} - y_j \quad (45)$$

Para  $j \in \mathbb{N} = [1:T]$ .

Geralmente no máximo duas diferenças são necessárias (MONTGOMERY, JENNINGS & KULAHCI, 2015).

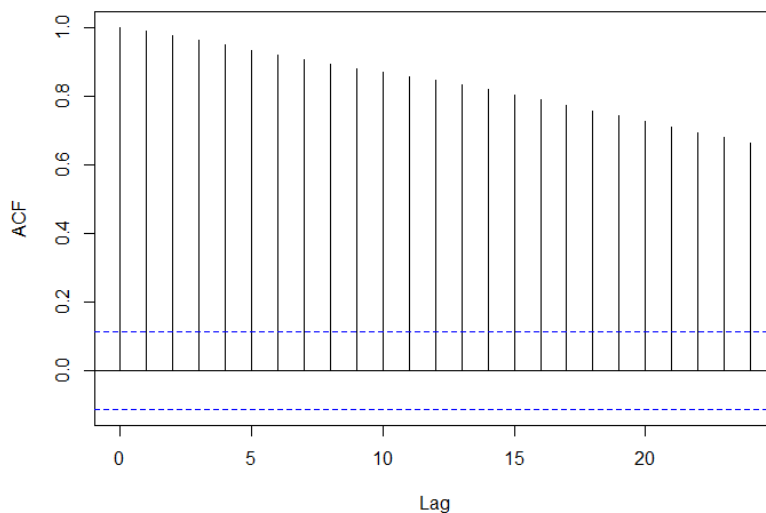


Figura 9. Função de autocorrelação de uma série temporal aleatória com tendência.  
Fonte: elaboração própria.

O segundo modo é aplicar testes de raiz unitária, como de *Phillips-Perron*, *Dickey-Fuller* e *Dickey-Fuller Aumentado*. Formalmente, para testar se uma série temporal  $y$  é um passeio aleatório ou um passeio aleatório com *drift* conceitual, aplicam-se os modelos descritos nas Equações 46 e 47.

$$y_t = \phi_1 y_{t-1} + \varepsilon_t \quad (46)$$

$$y_t = \phi_0 + \phi_1 y_{t-1} + \varepsilon_t \quad (47)$$

Sendo  $\varepsilon_t$  o termo de erro,  $y_t$  a série original e  $y_{t-1}$  a série defasada em  $lag = 1$ . As hipóteses nula e alternativa do teste de raiz unitária de *Dickey-Fuller* são dadas por:  $H_0: \phi_1 = 1$  e  $H_1: \phi_1 < 1$

Considerando a Equação 46, os estimadores de  $\phi_1$  e da variância do termo de erro  $\sigma_\varepsilon^2$  são dados pelas Equações 48 e 49, respectivamente:

$$\hat{\phi}_1 = \frac{\sum_{t=1}^T y_t y_{t-1}}{\sum_{i=1}^T y_{t-1}^2} \quad (48)$$

$$\sigma_\varepsilon^2 = \frac{\sum_{t=1}^T (y_t - \hat{\phi}_1 y_{t-1})^2}{T - 1} \quad (49)$$

Para  $y_0 = 0$  e  $T$  é o tamanho da série temporal.

A estatística razão de  $t$ , que corresponde ao teste de Dickey-Fuller (DF) é dada por:

$$DF \equiv \text{razão de } t = \frac{\hat{\phi}_1 - 1}{\hat{\sigma}_{\hat{\phi}_1}} = \frac{\sum_{t=1}^T \varepsilon_t y_{t-1}}{\hat{\sigma}_\varepsilon \sqrt{\sum_{i=1}^T y_{t-1}^2}} \quad (50)$$

Se  $\varepsilon_t$  é um ruído branco com momentos finitos de ordem maior que 2, então a estatística DF converge para uma função de moção *Browniana* padrão com  $T \rightarrow \infty$ . Se  $\phi_0 = 0$ , a Equação 47 é utilizada mesmo assim e a estatística DF para outra distribuição assimétrica. A Simulação de Monte Carlo é utilizada para determinar os valores críticos para cada um dos casos. Contudo, se  $\phi_0 \neq 0$ , a Equação 46 é utilizada e DF converge assintoticamente para uma distribuição normal.

No entanto, em muitos casos, assim como neste trabalho, modelos ARIMA  $(p, d, q)$  são mais adequados para descrever uma série temporal do que os modelos apresentados pela Equação 46 e Equação 47. Para exemplificar, considerando um modelo AR( $p$ ) e sendo mais uma vez  $y_t$  a série temporal em análise. Uma forma de testar a presença de raiz unitária é ajustar o modelo apresentado pela Equação 51.

$$y_t = c_t + \beta y_{t-1} + \sum_{i=1}^{p-1} \phi_i \Delta y_{t-i} + \varepsilon_t \quad (51)$$

Para  $c_t = \omega_0 + \omega_1 t$  ou  $c_t = 0$  e  $\Delta y_j = y_j - y_{j-1}$  é a  $j$ -ésima diferença da série.

A estatística de teste de Dickey-Fuller Aumentado (DFA) de  $\hat{\beta} - 1$  é dada por:

$$DFA = \frac{\hat{\beta} - 1}{\sigma_{\hat{\beta}}} \quad (52)$$

Sendo:  $\hat{\beta}$  a estimativa de MQO de  $\beta$  e  $\sigma_{\hat{\beta}}$  o erro padrão de  $\hat{\beta}$ .

## E. Análise de Sazonalidade

Esta etapa é necessária para decidir qual modelo de Box & Jenkins será utilizado. O modelo ARIMA ou o modelo SARIMA, que é adequado para séries sazonais. A seção 2.1.3 já tratou dos procedimentos a serem aplicados se a série temporal apresentar sazonalidade.

## F. Criar modelos ARIMA (DMUs) ou modelos SARIMA (DMUs)

Modelos ARIMA  $(p, d, q)$  ou SARIMA  $(p, d, q) (P, D, Q)_F$  são criados nesta etapa considerando defasagem máxima de 3 para os termos autorregressivo ( $p$ ), média móvel ( $q$ ), autorregressivo sazonal ( $P$ ) e média móvel sazonal ( $Q$ ). Já os termos  $d$  e  $D$  são fixados na quantidade de diferenças necessárias para tornar a série estacionária.

Sendo assim, cada parâmetro computacional dos serão números inteiros que podem assumir os valores (0, 1, 2, 3). As DMUs serão geradas por meio de um desenho fatorial completo  $4^k$ , sendo  $k$  o número de fatores a serem considerados. Em ambos os casos a primeira linha experimental ARIMA (0, 0, 0) e SARIMA (0, 0, 0) (0, 0, 0)<sub>F</sub> é retirada. O número de corridas experimentais é 15 (16 – 1) se considerados modelos ARIMA e 255 (256 – 1) se modelos SARIMA forem analisados.

Considerando uma série temporal sem tendência e sem sazonalidade, os modelos ARIMA (2, 0, 3), ARIMA (1, 0, 2), ARIMA (3,0,3) são DMUs selecionadas aleatoriamente da população para exemplificar. Se a série temporal tiver sazonalidade com frequência e 12 meses, pode-se exemplificar as DMUs com os modelos SARIMA (1, 0, 0) (3, 0, 3)<sub>12</sub>, SARIMA (3, 0, 1) (2, 0, 3)<sub>12</sub> e SARIMA (0, 0, 2) (2, 0, 3)<sub>12</sub>.

A Tabela 1 apresenta as DMUs se o modelo ARIMA for considerado, enquanto a Tabela 2 apresenta algumas das DMUs e o modelo SARIMA for utilizado na análise. De modo a diminuir a dimensão da tabela, e como  $d$  e  $D$  possuem os mesmos valores independentemente da corrida experimental, suas respectivas colunas foram suprimidas.

Tabela 1. DMUs referentes à modelagem ARIMA.

DMU	$p$	$q$
1	1	0
2	2	0
3	3	0
4	0	1
5	1	1
6	2	1
7	3	1
8	0	2
9	1	2
10	2	2
11	3	2
12	0	3
13	1	3
14	2	3
15	3	3

Fonte: elaboração própria.

Tabela 2. DMUs referentes à modelagem SARIMA.

DMU	$p$	$q$	$P$	$Q$
1	1	0	0	0
2	2	0	0	0
3	3	0	0	0
4	0	1	0	0
5	1	1	0	0
6	2	1	0	0
7	3	1	0	0
8	0	2	0	0
9	1	2	0	0
10	2	2	0	0
11	3	2	0	0
12	0	3	0	0
13	1	3	0	0
14	2	3	0	0
15	3	3	0	0
⋮	⋮	⋮	⋮	⋮
241	1	0	3	3

242	2	0	3	3
243	3	0	3	3
244	0	1	3	3
245	1	1	3	3
246	2	1	3	3
247	3	1	3	3
248	0	2	3	3
249	1	2	3	3
250	2	2	3	3
251	3	2	3	3
252	0	3	3	3
253	1	3	3	3
254	2	3	3	3
255	3	3	3	3

Fonte: elaboração própria.

### G. Modelar a série para cada DMU

Dado os desenhos fatoriais completos (com exceção dos modelos correspondentes à média), o conjunto treino da série temporal é modelada para cada DMU conforme o desenho de misturas selecionado. Ou seja, se a série temporal não possui sazonalidade, as séries serão modeladas segundo a Tabela 1. Se apresentar sazonalidade, as DMUs serão representadas por cada linha da Tabela 2.

### H. Calcular as métricas de ajuste e de previsão e a quantidade de parâmetros estimados

As métricas apresentadas na seção 2.1.4 são calculadas neste passo para cada DMU, tanto para o ajuste quanto para a previsão. A quantidade de parâmetros estimados ( $k$ ) é dada pela soma dos parâmetros estimados na parte autorregressiva e de médias móveis, tanto na parte sazonal com na parte não sazonal dos modelos ARIMA e SARIMA, respectivamente.

### I. Definir os *inputs* e os *outputs*

Preparando-se a aplicação de DEA, os *inputs* e *outputs* sugeridos no algoritmo deste trabalho são apresentados na Tabela 3. É importante destacar que o objetivo de qualquer modelo de *Machine Learning* para previsão é minimizar as métricas de qualidade de previsão. Este trabalho utilizará um modelo de DEA orientado a *outputs*. Contudo, para manter a coerência entre os modelos, o inverso dessas métricas é utilizado (LEWIS & SEXTON, 2020).

Foram considerados como *inputs* as métricas de ajuste e o número de parâmetros estimado por cada modelo, conforme a discussão apresentada na seção 2.1.

Tabela 3. Inputs e Outputs a serem considerados no modelo DEA.

<i>Inputs</i>	<i>Outputs</i>
$MAPE_{adj}$	$1 / MAPE_{prev}$
$RMSE_{adj}$	$1 / RMSE_{prev}$
$MdAE_{adj}$	$1 / MdAE_{prev}$
$MAE_{adj}$	$1 / MAE_{prev}$
$k$	

Fonte: elaboração própria.

### J. Aplicar um modelo DEA adequado

Como há a presença de variáveis fracionadas (o inverso das métricas de performance de previsão), os modelos clássicos CCR e BCC são inadequados para o problema em questão. Este trabalho utilizará, portanto, o modelo denominado DEA-R orientado a *outputs*, que foi proposto por Despić, Despić e Paradi (2007) e que foi discutido na seção 2.4.

Wang, Smith-Miles e Hyndman (2009) consideram que a utilização de 2 a 4 modelos para fins de comparação é adequado. Portanto, o modelo DEA-R será aplicado no contexto de modelos de super eficiência (seção 2.3) para elencar as quatro DMUs mais eficientes e que serão ponderadas para a realização da previsão.

### K. Definir as DMUs eficientes

A partir do modelo DEA-R, definem-se as DMUs eficientes. Como nos modelos de DEA clássicos, a DMU que apresentar escore de eficiência igual a 1 é considerada eficiente. Contudo, este trabalho reterá as 4 DMUs mais eficientes por meio da aplicação do modelo DEA-R no contexto de super eficiência, em que a *i-ésima* DMU em análise é retirada das restrições correspondentes do modelo DEA. Observa-se pelas Equações 23 e 24 que essa ação permite que as DMUs obtenham eficiência maior que 1.

### L. Definir um Delineamento de Misturas adequado

O objetivo deste trabalho é aplicar a metodologia de Delineamento de Misturas para otimizar o peso de *Ensemble* de modelos eficientes em relação à previsão de séries temporais.

O Desenho de Misturas a ser aplicado é o *Simplex Lattice Design* (SLD) (4, 5). Como a intenção é analisar se a Superfície de Resposta a ser aplicada posteriormente possui curvatura para fins de otimização, optou-se por adicionar 1 ponto central e 4 pontos axiais ao



delineamento selecionado, totalizando 61 corridas (sendo 56 corridas relativas aos pontos do desenho SLD proposto).

A Figura 10 apresenta a representação gráfica do Delineamento de Misturas selecionado. Como há 4 fatores a serem considerados (os 4 modelos selecionados como os mais eficientes na etapa anterior), a figura apresenta o delineamento para 3 componentes.

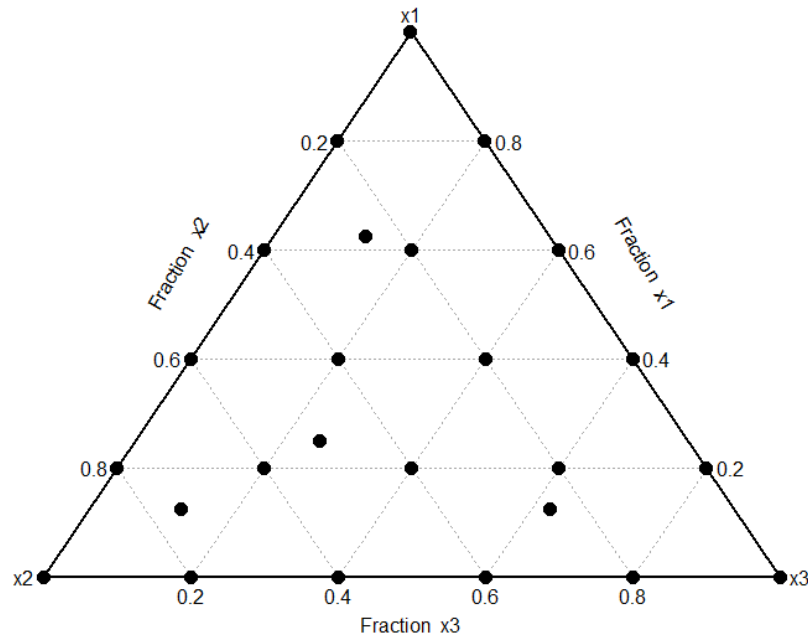


Figura 10. Representação gráfica do Delineamento de Misturas Selecionado.  
Fonte: elaboração própria.

### M. Calcular as métricas de previsão segundo o Delineamento de Misturas

Cada linha experimental do Delineamento de Misturas apresenta os pesos a serem aplicados no *Ensemble* dos modelos selecionados como mais eficientes. As métricas de previsão serão consideradas como as variáveis resposta no modelo de Superfície de Resposta a ser empregado.

Dado o desenho selecionado, a Tabela 4 apresenta as 10 primeiras e últimas linhas experimentais como uma representação geral deste passo. Para fins de conveniência de notação as representações das métricas de previsão na matriz estão representadas por  $y_{ij}$ , em que  $i$  corresponde à linha experimental e  $j$  à métrica em questão. A linha experimental 57 corresponde aos pontos centrais do modelo, enquanto as linhas 58 a 61 correspondem aos pontos axiais adicionados ao modelo.

Tabela 4. Representação geral do delineamento proposto com as variáveis resposta.

Corrida	$x_1$	$x_2$	$x_3$	$x_4$	$MAPE_{prev}$	$RMSE_{prev}$	$MdAE_{prev}$	$MAE_{prev}$
1	1,000	0,000	0,000	0,000	$y_{1_1}$	$y_{1_2}$	$y_{1_3}$	$y_{1_4}$
2	0,800	0,200	0,000	0,000	$y_{2_1}$	$y_{2_2}$	$y_{2_3}$	$y_{2_4}$
3	0,600	0,400	0,000	0,000	$y_{3_1}$	$y_{3_2}$	$y_{3_3}$	$y_{3_4}$
4	0,400	0,600	0,000	0,000	$y_{4_1}$	$y_{4_2}$	$y_{4_3}$	$y_{4_4}$
5	0,200	0,800	0,000	0,000	$y_{5_1}$	$y_{5_2}$	$y_{5_3}$	$y_{5_4}$
6	0,000	1,000	0,000	0,000	$y_{6_1}$	$y_{6_2}$	$y_{6_3}$	$y_{6_4}$
7	0,800	0,000	0,200	0,000	$y_{7_1}$	$y_{7_2}$	$y_{7_3}$	$y_{7_4}$
8	0,600	0,200	0,200	0,000	$y_{8_1}$	$y_{8_2}$	$y_{8_3}$	$y_{8_4}$
9	0,400	0,400	0,200	0,000	$y_{9_1}$	$y_{9_2}$	$y_{9_3}$	$y_{9_4}$
10	0,200	0,600	0,200	0,000	$y_{10_1}$	$y_{10_2}$	$y_{10_3}$	$y_{10_4}$
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
52	0,000	0,000	0,400	0,600	$y_{52_1}$	$y_{52_2}$	$y_{52_3}$	$y_{52_4}$
53	0,200	0,000	0,000	0,800	$y_{53_1}$	$y_{53_2}$	$y_{53_3}$	$y_{53_4}$
54	0,000	0,200	0,000	0,800	$y_{54_1}$	$y_{54_2}$	$y_{54_3}$	$y_{54_4}$
55	0,000	0,000	0,200	0,800	$y_{55_1}$	$y_{55_2}$	$y_{55_3}$	$y_{55_4}$
56	0,000	0,000	0,000	1,000	$y_{56_1}$	$y_{56_2}$	$y_{56_3}$	$y_{56_4}$
57	0,250	0,250	0,250	0,250	$y_{57_1}$	$y_{57_2}$	$y_{57_3}$	$y_{57_4}$
58	0,625	0,125	0,125	0,125	$y_{58_1}$	$y_{58_2}$	$y_{58_3}$	$y_{58_4}$
59	0,125	0,625	0,125	0,125	$y_{59_1}$	$y_{59_2}$	$y_{59_3}$	$y_{59_4}$
60	0,125	0,125	0,625	0,125	$y_{60_1}$	$y_{60_2}$	$y_{60_3}$	$y_{60_4}$
61	0,125	0,125	0,125	0,625	$y_{61_1}$	$y_{61_2}$	$y_{61_3}$	$y_{61_4}$

Fonte: elaboração própria.

## N. Verificar a adequabilidade de redução de dimensionalidade

Se a quantidade de métricas a serem analisadas for relativamente grande, pode ser de interesse a diminuição da dimensionalidade das respostas com o intuito de diminuir o esforço computacional e se trabalhar com otimizações de menor dimensão. A análise da matriz de correlação, da estatística de *Kayser-Maier-Olkin* (KMO) e a aplicação do Teste de Esfericidade de *Bartlett* são comuns para analisar a adequabilidade de aplicação de Análise de Componentes Principais (PCA) ou Análise Fatorial (FA) que são métodos de interesse neste caso.

## O. Ajustar Superfícies de Reposta

Após o cálculo das métricas para cada *ensemble*, cujos pesos da soma ponderada são dados por cada linha experimental do Desenho de Misturas proposto, a metodologia de Superfícies de Resposta é aplicada.

Lawson (2014) indica que uma boa estratégia no caso do Delineamento de Misturas é estimar o modelo de mais alta ordem para posteriormente analisar os testes de hipótese em relação aos coeficientes. Contudo, este trabalho propõe a utilização dos algoritmos de *Stepwise*, para a seleção automática dos modelos.

O estimador de Mínimos Quadrados Ordinários (MQO) foi o selecionado para a estimativa dos modelos, uma vez que é o mais utilizado no contexto de regressão e possui vantagens neste contexto sobre o estimador de máximoverossimilhança (GUJARATI, 2011).

#### **P. Otimização das superfícies de resposta**

Nesta etapa, seleciona-se um algoritmo de otimização multiobjetivo para determinar os pesos que cada modelo receberá na soma ponderada para a criação dos *Ensembles*. Algoritmos evolutivos, *Desirability*, *Normal Boundary Intersection* (NBI) etc. podem ser selecionados neste passo.

Neste passo também pode-se analisar a adequabilidade de redução da dimensionalidade por meio de Análise de Componentes Principais ou Análise Fatorial sobre as respostas do Desenho de Misturas para transformar o problema em uma otimização com apenas um único objetivo. Este passo é importante se o pesquisador ou analista deparar-se com o problema de prever por meio do método proposto dezenas ou centenas de séries temporais de maneira independente.

#### **Q. Aplicar os pesos otimizados e comparar com outro método de *Ensemble***

Com os pesos otimizados, os mesmos são aplicados para criar as previsões por meio de *Ensemble*. Para fins de comparação, outro método de *Ensemble* pode ser selecionado. Como exemplo de métodos de *Ensemble*, como o *Bootstrap Aggregating* (*Bagging*), que corresponde a um meta-algoritmo utilizado para melhorar as previsões e evitar *overfitting*. O método consiste em retirar amostras com repetição da própria série temporal (o que define um *Bootstrap*), treinando modelos para cada uma das amostras. A mediana pode ser utilizada para combinar os modelos (JAMES *et al.*, 2013).

Outro método que pode ser utilizado é o *Random Forest*, que segundo (NIELSEN, 2019) é uma melhoria em relação ao *Bagging*. O *Random Forest* utiliza de um algoritmo de aprendizado de árvore para selecionar em cada etapa do processo um subconjunto de dados candidatos à uma previsão com boa performance.

Por fim, pode-se usar a média ou a média ponderada por algum método, uma vez que é um dos métodos mais utilizados, pela sua simplicidade e relativo custo-benefício positivo em termos computacionais em relação às métricas de previsão (KOURENTZES, BARROW & CRONE, 2014).

## **R. Prever com o melhor**

Dada a comparação das métricas de previsão do passo anterior, o pesquisador ou o analista retem o melhor modelo para previsão  $h$  passos à frente. Se alguma diferenciação for necessária, deve-se retornar a estrutura de tendência sobre o ajuste e a previsão para que preveja os valores reais da série temporal em análise.

## 4. RESULTADOS E DISCUSSÕES

Esta seção apresenta os resultados e discussões da aplicação da metodologia proposta. Com o intuito de generalizar os resultados, foi realizada uma Simulação de Monte Carlo que se encontra na seção 4.1. A seção 4.2. apresenta a aplicação empírica da metodologia em séries temporais relacionadas ao setor de energia elétrica.

### 4.1. SIMULAÇÃO

Com o intuito de generalizar os resultados encontrados nesta tese, foi realizado um estudo de Simulação de Monte Carlo, aplicando-se a metodologia proposta. Foram simuladas 100 séries de forma aleatória de tamanhos  $n_1 = 100$ ,  $n_2 = 300$  e  $n_3 = 500$ , tendo como *Data Generate Process* (DGP) modelos SARIMA. As séries foram simuladas sem tendência e sazonalidade em *lag* 12.

Como há várias séries temporais a serem analisadas por diversos modelos, optou-se por automatizar a análise de adequabilidade das regressões no contexto do desenho de misturas. Adotou-se um nível de significância de 5% para o teste de normalidade de *Anderson-Darling*, teste de homocedasticidade de *Breusch-Pagan*. Para confirmar que a série não possui tendência, o teste de *Dickey-Fuller Aumentado* foi aplicado. Além disso, para confirmar que a série possui sazonalidade, foi aplicado o teste de *Friedman* para sazonalidade. Quando qualquer um dos pressupostos anteriores não foi atendido, substitui-se a série temporal de forma aleatória até que a série e a modelagem atendessem as premissas assumidas.

Optou-se também por automatizar a análise da adequabilidade da aplicação de estatística multivariada para redução da dimensionalidade das respostas. Por conveniência, selecionou-se a Análise de Componentes Principais. Desta maneira, a redução de dimensionalidade foi realizada quando a matriz de correlação apresentasse no mínimo 70% de suas correlações acima de 0,30 e quando o teste de Bartlett e a estatística KMO apresentassem valores que permitissem a continuidade da aplicação do método. A quantidade de componentes foi selecionada a partir do Critério de Kaiser, retendo-se aqueles componentes que apresentam autovalores maiores que 1.

Após a estimação das Superfícies de Resposta e possível redução de dimensionalidade, aplicou-se o método de *Desirability* para otimizar os pesos dos *Ensembles*. Para fins de

comparação, foi utilizado o modelo Naive, que confere pesos iguais aos modelos considerados para serem combinados. As métricas foram calculadas comparando a previsão do conjunto treino sobre o conjunto treino de cada série simulada.

Tabela 5. Estatísticas descritivas das métricas no contexto da Simulação realizada.

		<b>Min.</b>	<b>1° Quartil</b>	<b>Mediana</b>	<b>Média</b>	<b>3° Quartil</b>	<b>Max.</b>
<b>n = 100</b>	MAPE <sub>DEA</sub>	0,644	0,855	0,937	1,190	1,060	10,739
	MAPE <sub>Naive</sub>	0,799	1,147	1,478	2,321	1,881	29,115
	RMSE <sub>DEA</sub>	0,592	0,771	0,869	0,894	0,990	1,357
	RMSE <sub>Naive</sub>	0,727	0,924	1,030	1,063	1,175	1,567
	MdAE <sub>DEA</sub>	0,213	0,430	0,531	0,558	0,673	1,030
	MdAE <sub>Naive</sub>	0,258	0,584	0,722	0,737	0,868	1,269
	MAE <sub>DEA</sub>	0,506	0,609	0,692	0,712	0,816	1,099
	MAE <sub>Naive</sub>	0,572	0,735	0,820	0,857	0,966	1,247
<b>n = 300</b>	MAPE <sub>DEA</sub>	0,838	0,881	0,902	0,947	0,946	1,679
	MAPE <sub>Naive</sub>	0,939	1,022	1,079	1,153	1,148	2,696
	RMSE <sub>DEA</sub>	0,726	0,849	0,907	1,006	1,116	1,637
	RMSE <sub>Naive</sub>	0,809	0,950	1,021	1,135	1,290	1,890
	MdAE <sub>DEA</sub>	0,393	0,551	0,648	0,684	0,749	1,184
	MdAE <sub>Naive</sub>	0,454	0,648	0,751	0,792	0,895	1,386
	MAE <sub>DEA</sub>	0,575	0,676	0,740	0,812	0,909	1,355
	MAE <sub>Naive</sub>	0,647	0,763	0,831	0,915	1,024	1,525
<b>n = 500</b>	MAPE <sub>DEA</sub>	0,838	0,881	0,902	0,947	0,946	1,679
	MAPE <sub>Naive</sub>	0,963	1,025	1,070	1,231	1,168	6,637
	RMSE <sub>DEA</sub>	0,726	0,849	0,908	1,006	1,116	1,637
	RMSE <sub>Naive</sub>	0,836	0,965	1,012	1,055	1,081	1,520
	MdAE <sub>DEA</sub>	0,393	0,551	0,648	0,684	0,749	1,184
	MdAE <sub>Naive</sub>	0,516	0,617	0,689	0,713	0,773	1,128
	MAE <sub>DEA</sub>	0,575	0,676	0,740	0,812	0,909	1,355
	MAE <sub>Naive</sub>	0,654	0,766	0,813	0,850	0,858	1,264

Fonte: elaboração própria.

A Tabela 5 apresenta as estatísticas descritivas das métricas em relação ao método proposto nesta tese e o método Naive. A Figura 11, Figura 12 e Figura 13 apresentam os Box-Plots das métricas, comparando-as entre os métodos propostos, sem os *outliers* a fim de facilitar a visualização dos gráficos.

Observa-se certa simetria entre as distribuições das métricas mostradas por meio dos *Box-Plots*. Quanto às métricas, o MAPE que é retornado pelo método proposto apresenta claramente menor variabilidade e média. Nota-se, ainda, que suas distribuições são assimétricas à direita.

Em relação às demais métricas, nota-se que variabilidade aumenta, mas não é possível afirmar que a média das métricas do modelo proposto são estatisticamente menores que as do método Naive. Para as séries com  $n = 300$  e  $n = 500$  a simulação indica que a variabilidade do método proposto aumenta para o MdAE e MAE.

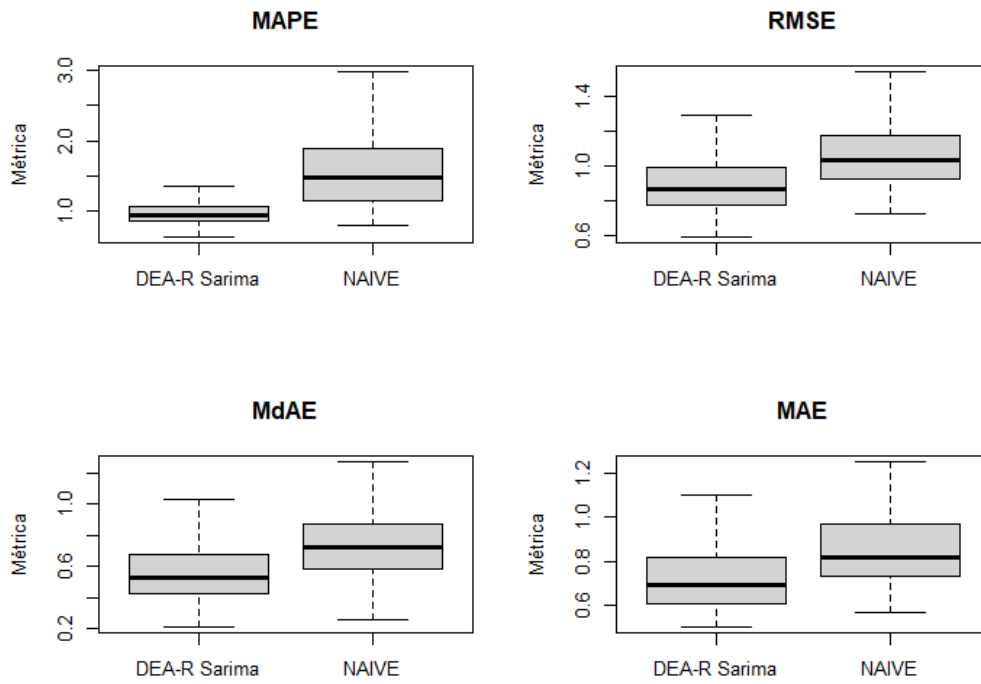


Figura 11. Box-Plots das métricas de performance por método de combinação da Simulação para  $n = 100$ .  
Fonte: elaboração própria.

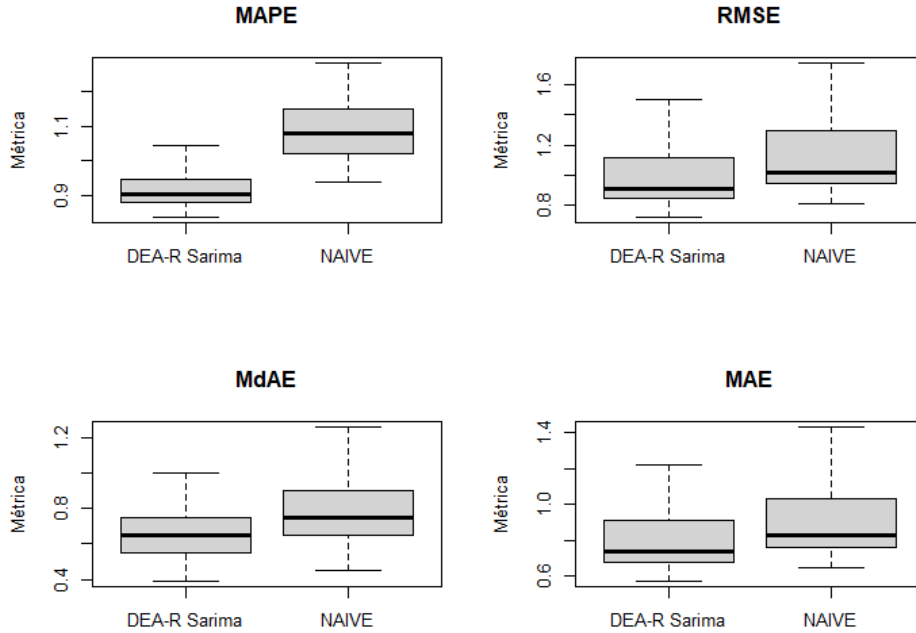


Figura 12. Box-Plots das métricas de performance por método de combinação da Simulação para  $n = 300$ .  
Fonte: elaboração própria.

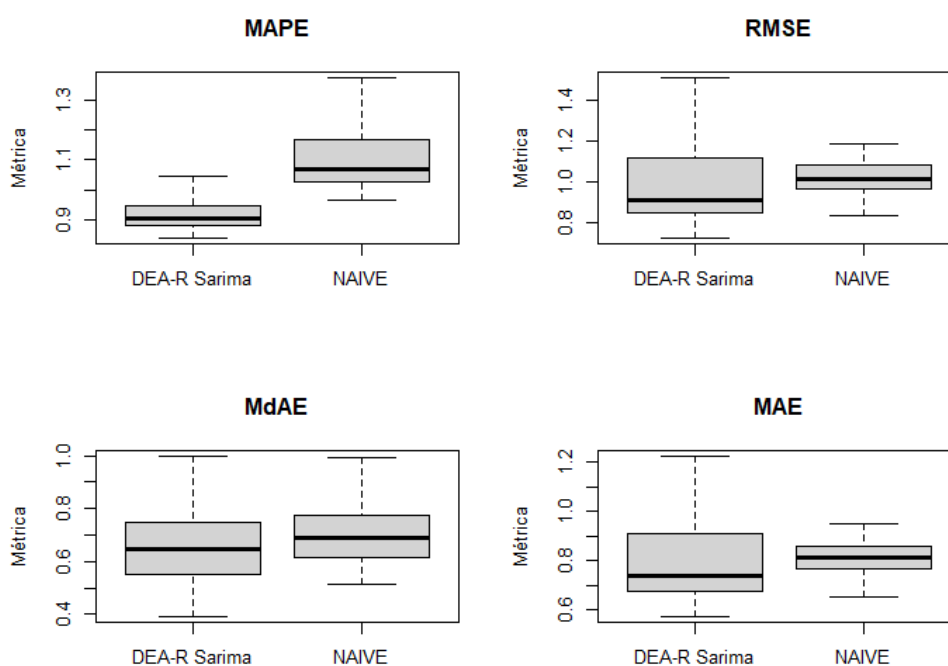


Figura 13. Box-Plots das métricas de performance por método de combinação da Simulação para  $n = 500$ .  
Fonte: elaboração própria.

## 4.2. APLICAÇÃO EMPÍRICA

Para verificar a aplicabilidade empírica da metodologia proposta, nesta seção a mesma será aplicada em séries temporais empíricas.

### A. Selecionar as séries temporais

As séries temporais em que a metodologia proposta será aplicada são séries de demanda de energia elétrica das 5 regiões do Brasil (Centro-Oeste, Nordeste, Norte, Sudeste e Sul), além da demanda de energia elétrica do Brasil. As séries temporais se iniciam em janeiro de 2004 e vão até agosto de 2022, totalizando 224 observações. Os dados foram coletados da Agência Nacional de Energia Elétrica (ANEEL, 2022).

### B. Análise exploratória dos dados



A Tabela 6 apresenta as principais estatísticas descritivas das cinco séries em análise, possibilitando uma visão geral do comportamento em relação às estatísticas de posição e de dispersão dos dados.

Tabela 6. Estatísticas descritivas dos dados de demanda de energia elétrica (em milhares / MWh).

	<b>Min.</b>	<b>1º Quartil</b>	<b>Mediana</b>	<b>Média</b>	<b>3º Quartil</b>	<b>Max.</b>	<b>D.P.</b>
Norte	1.507,45	1.981,67	2.476,55	2.420,11	2.826,52	3.416,76	472,30
Nordeste	4.139,77	5.384,73	6.380,07	6.112,85	6.740,25	7.719,87	851,70
Sudeste	14.714,71	17.293,08	19.246,12	18.663,15	19.920,24	21.963,88	1.723,22
Sul	4.510,11	5.410,73	6.567,12	6.370,46	7.201,71	8.801,05	1.041,73
C. Oeste	1.456,01	1.969,13	2.706,96	2.542,96	2.542,12	3.058,29	593,05
Brasil	26.508,17	32.039,78	37.677,82	36.108,69	39.593,06	44.250,44	4.531,84

Fonte: elaboração própria.

A Figura 14 apresenta as séries temporais que serão analisadas. Preliminarmente, observa-se tendência e sazonalidade em todas as séries, estruturas que serão melhor analisadas na seção análise exploratória dos dados na metodologia proposta.

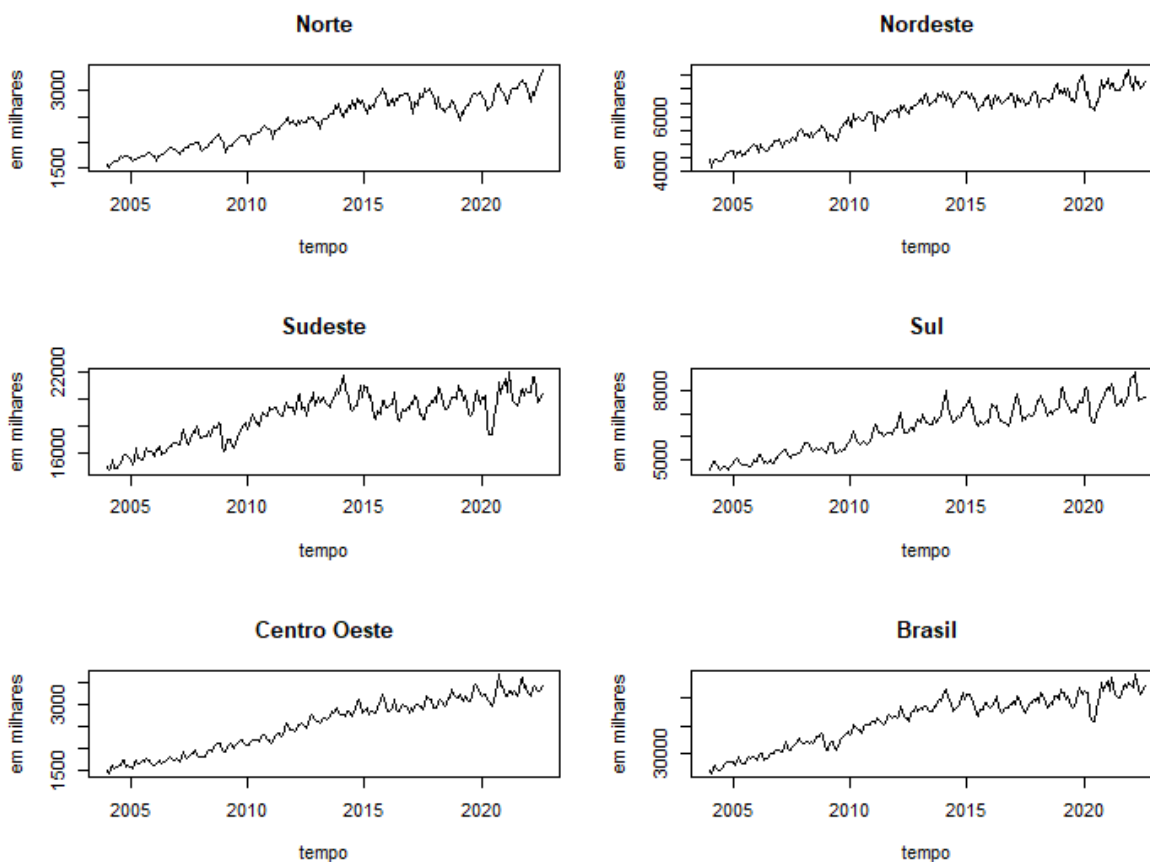


Figura 14. Séries temporais de demanda de energia elétrica do Brasil e de suas regiões (em milhares / MWh).

Fonte: elaboração própria.

### C. Dividir os dados em Treino e Teste

Seguindo a recomendação de Hyndman e Athanasopoulos (2018), 80% dos dados da série temporal foram mantidos como conjunto de dados de treino e 20% foram separados para teste, a fim de comparar o desempenho das DMUs a serem criadas.

As estatísticas descritivas tanto para o conjunto treino, quanto para o conjunto teste são encontradas na Tabela 7 e Tabela 8, respectivamente.

Tabela 7. Estatísticas descritivas do conjunto treino das séries temporais analisadas.

	<b>Min.</b>	<b>1º Quartil</b>	<b>Mediana</b>	<b>Média</b>	<b>3º Quartil</b>	<b>Max.</b>
Norte	1.507,45	1.895,82	2.321,77	2.294,08	2.691,07	3.050,12
Nordeste	4.139,77	5.166,43	6.093,03	5.886,72	6.602,72	7.184,24
Sudeste	14.714,71	16.806,56	18.762,19	18.312,36	19.605,54	21.763,01
Sul	4.510,11	5.272,96	6.147,98	6.058,84	6.801,10	8.019,05
C. Oeste	1.456,01	1.842,46	2.315,72	2.360,48	2.843,22	3.314,45
Brasil	26.508,17	31.247,02	35.956,44	34.912,49	38.618,19	41.653,49

Fonte: elaboração própria.

Tabela 8. Estatísticas descritivas do conjunto teste das séries temporais analisadas.

	<b>Min.</b>	<b>1º Quartil</b>	<b>Mediana</b>	<b>Média</b>	<b>3º Quartil</b>	<b>Max.</b>
Norte	2.421,38	2.770,55	2.938,03	2.921,44	3.053,07	3.416,76
Nordeste	6.234,57	6.820,98	7.035,12	7.012,32	7.239,43	7.719,87
Sudeste	17.315,90	19.706,59	20.187,00	20.058,51	20.642,06	21.963,88
Sul	6.566,36	7.330,46	7.579,90	7.610,03	7.950,41	8.801,05
C. Oeste	2.943,30	3.161,54	3.239,99	3.264,65	3.375,32	3.654,54
Brasil	35.769,05	40.157,52	41.147,99	40.866,93	41.988,44	44.250,44

Fonte: elaboração própria.

### D. Análise de tendência

A Figura 15 apresenta as funções de autocorrelação das séries temporais em análise. Conforme exposto no referencial teórico, o gráfico da função de autocorrelação apresenta as correlações entre a variável analisada e a mesma defasada em *lags*.

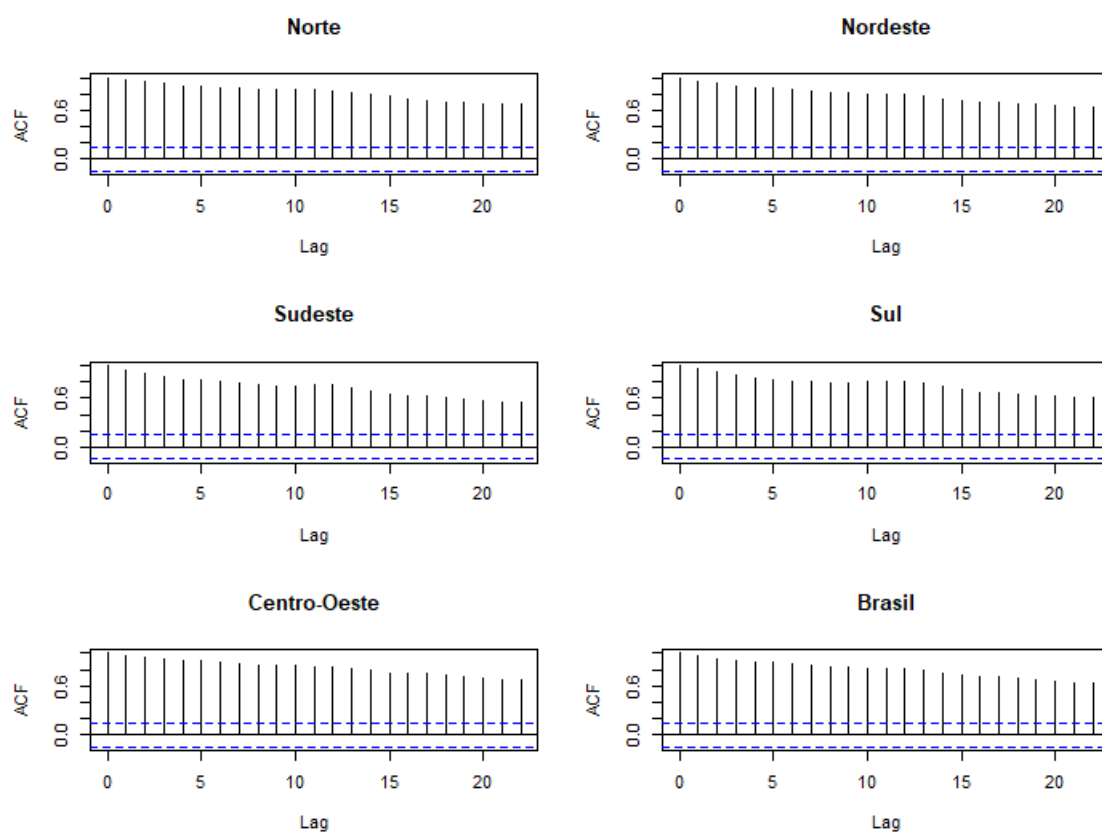


Figura 15. Funções de autocorrelação das séries temporais analisadas.  
 Fonte: elaboração própria.

Pelo formato dos gráficos da função de autocorrelação, observa-se a presença de tendência em todas as séries. Para confirmar a análise da função de autocorrelação, a Tabela 9 apresenta os testes de raiz unitária de KPSS.

Tabela 9. Teste de raiz unitária de KPSS.

	Estatística (KPSS)	Valores Críticos			
		10%	5%	2,5%	1%
Norte	2,8829				
Nordeste	1,6233				
Sudeste	0,5127				
Sul	3,1291	0,347	0,463	0,574	0,739
Centro-Oeste	3,4555				
Brasil	0,9374				

Fonte: elaboração própria.

Os testes indicam que a 95% de nível de confiança há a estrutura de tendência em todas as séries temporais analisadas. Aplicou-se então a primeira diferença, cujos gráficos da função de autocorrelação são apresentados na Figura 16.

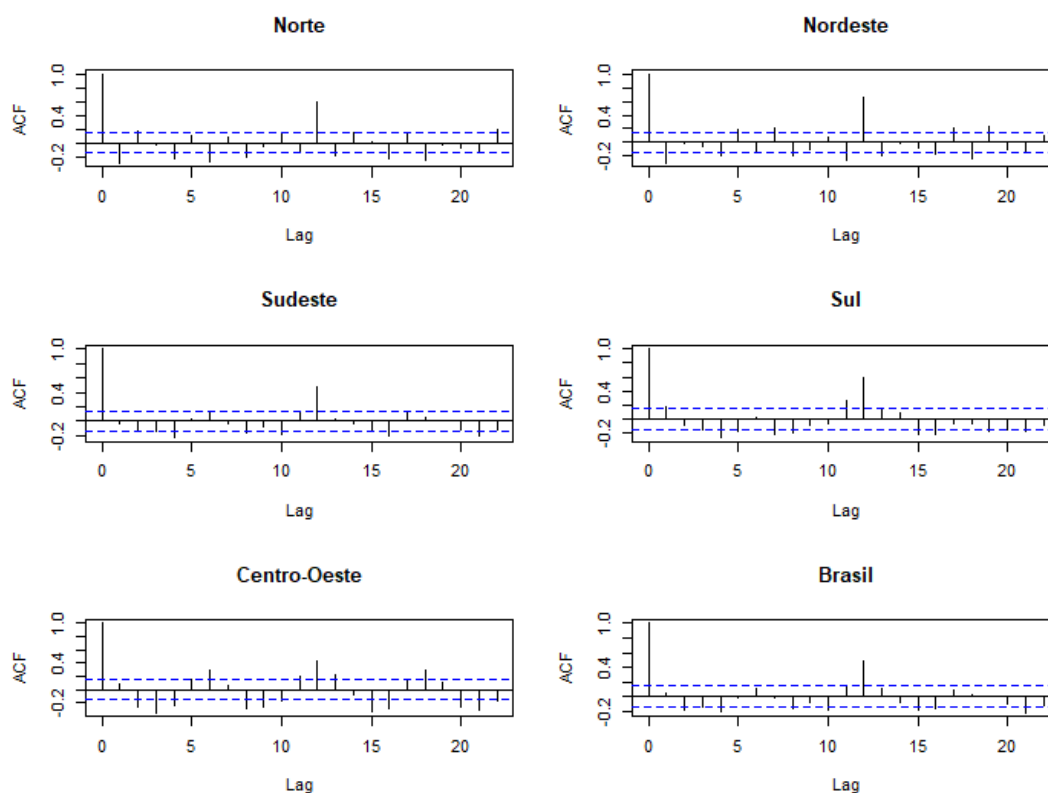


Figura 16. Funções de autocorrelação da primeira diferença das séries temporais analisadas.  
Fonte: elaboração própria.

Analisando-se a Figura 16, observa-se que a primeira diferença foi suficiente para a retirada da tendência, o que é confirmado pela Tabela 10 que apresenta os testes de raiz unitária de KPSS para cada série temporal analisada.

Tabela 10. Teste de raiz unitária de KPSS sobre a primeira diferença das séries temporais analisadas.

	Estatística (KPSS)	Valores Críticos			
		10%	5%	2,5%	1%
Norte	0,2609				
Nordeste	0,1809				
Sudeste	0,0711				
Sul	0,0477	0,347	0,463	0,574	0,739
Centro-Oeste	0,0426				
Brasil	0,1025				

Fonte: elaboração própria.

## E. Análise de sazonalidade

Entre a Figura 17 e Figura 22 são apresentados os gráficos de padrão sazonal de cada série temporal em análise. O gráfico de padrão sazonal plota o gráfico separando os dados por ano como variável categórica.

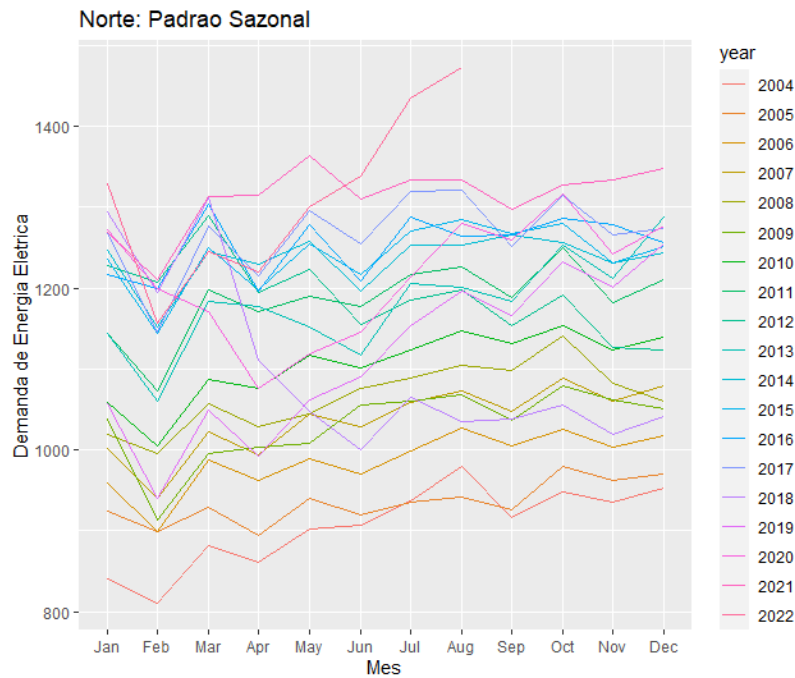


Figura 17. Padrão de sazonalidade da demanda de energia elétrica da região Norte.  
Fonte: elaboração própria.

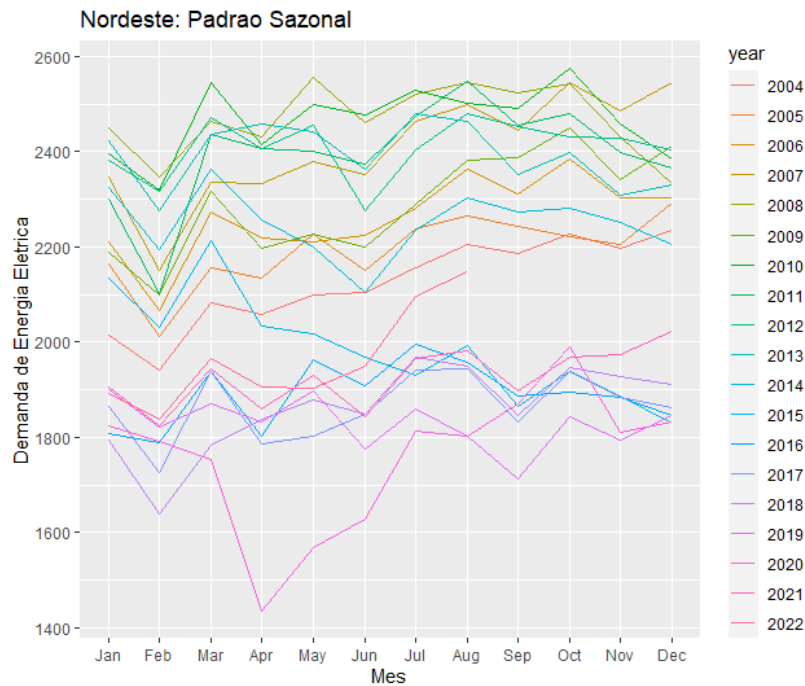


Figura 18. Padrão de sazonalidade da demanda de energia elétrica da região Nordeste.  
Fonte: elaboração própria.

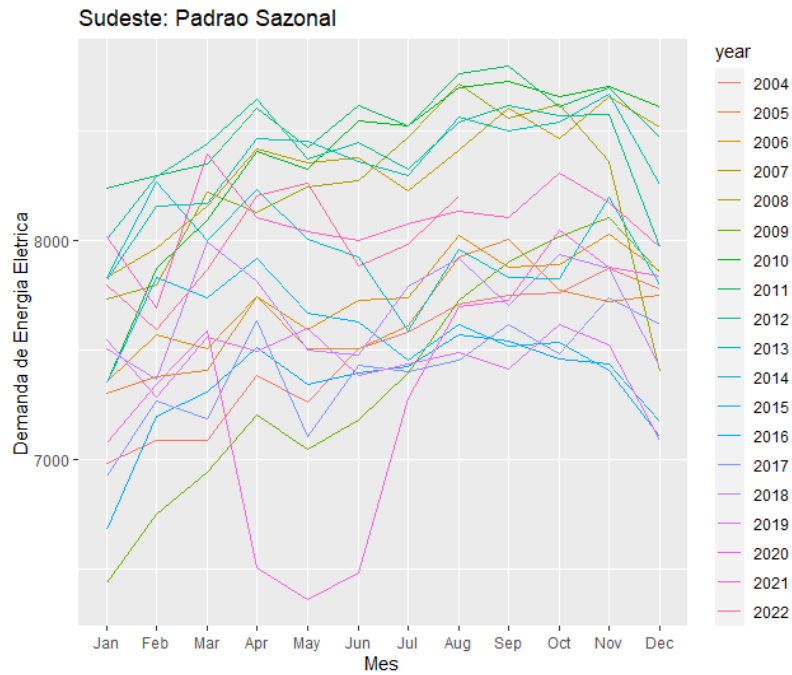


Figura 19. Padrão de sazonalidade da demanda de energia elétrica da região **Norte Sudeste**.  
 Fonte: elaboração própria.

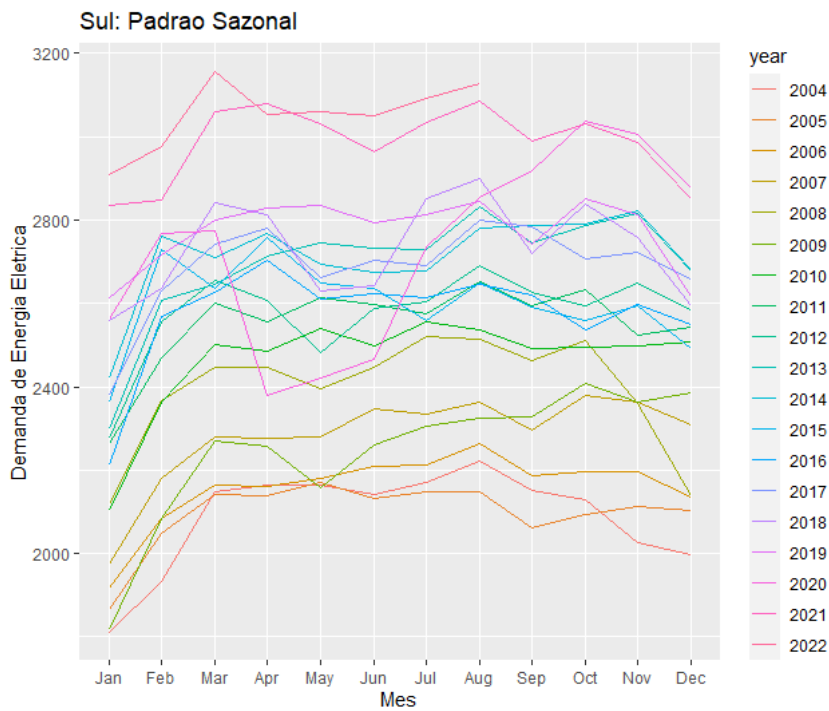


Figura 20. Padrão de sazonalidade da demanda de energia elétrica da região **Norte Sul**.  
 Fonte: elaboração própria.

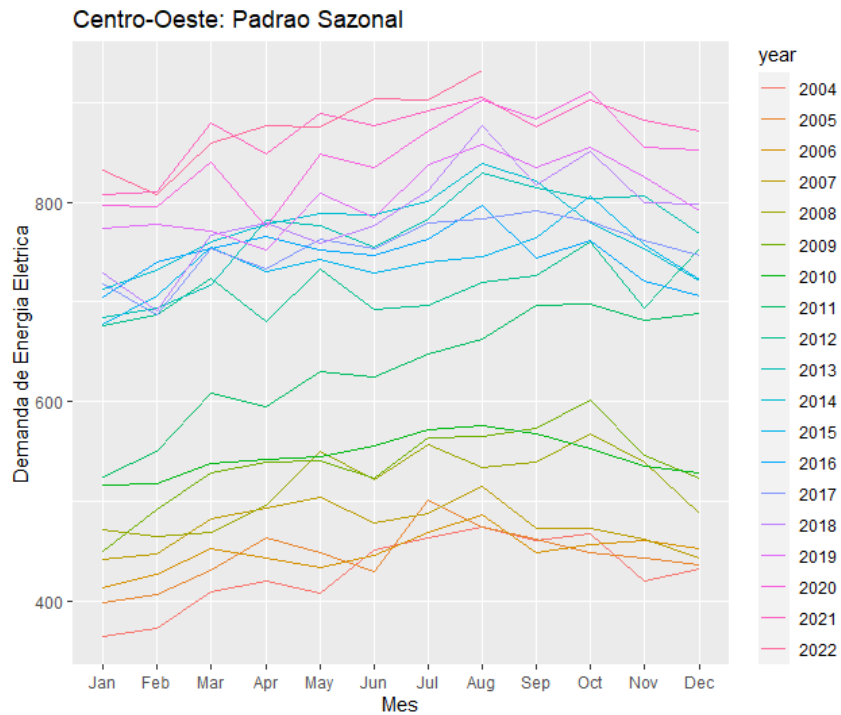


Figura 21. Padrão de sazonalidade da demanda de energia elétrica da região Norte. Centro-Oeste.  
 Fonte: elaboração própria.

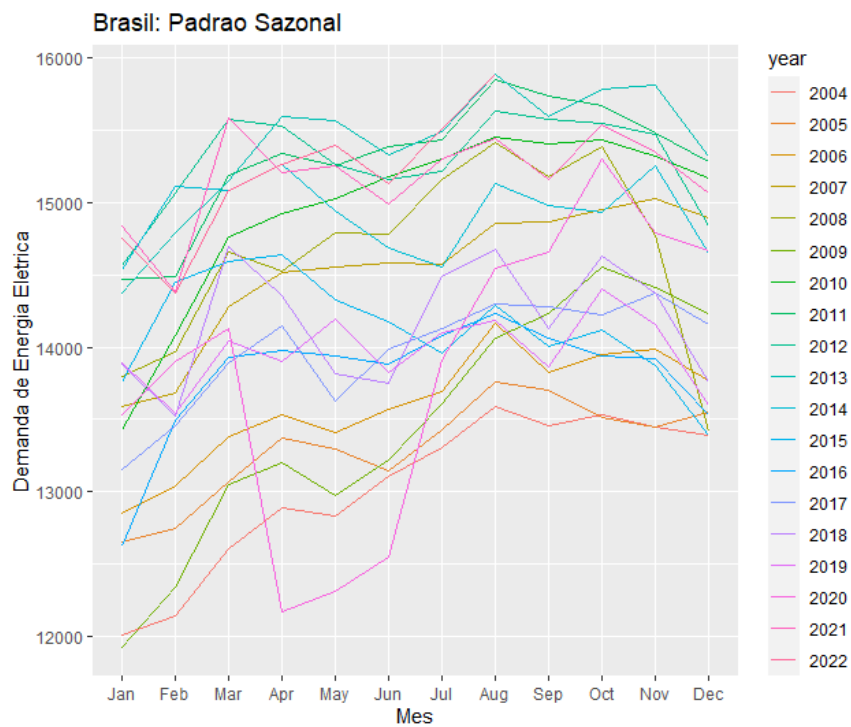


Figura 22. Padrão de sazonalidade da demanda de energia elétrica do Brasil.  
 Fonte: elaboração própria.

Observa-se em todos os gráficos que, independentemente da região analisada, há um aumento na demanda de energia elétrica entre os meses de junho e outubro que corresponde aos

meses referentes ao inverno e início da primavera, dando indícios que há sazonalidade nas séries em análise. Importante mencionar que as funções de autocorrelação apresentadas na Figura 16 já deram fortes indícios de sazonalidade, uma vez que o coeficiente de correlação é significativo em *lags* 12.

A fim de confirmar a análise gráfica, a Tabela 11 apresenta os testes de sazonalidade de Friedman, cuja hipótese nula é a não sazonalidade da série analisada.

Tabela 11. Teste de *Friedman* para sazonalidade

	<b>Estatística (Friedman)</b>	<b>p-valor</b>
Norte	114,95	0,000
Nordeste	118,78	0,000
Sudeste	86,00	0,000
Sul	83,76	0,000
Centro-Oeste	50,60	0,000
Brasil	83,67	0,000

Fonte: elaboração própria.

Como pode ser observado, há fortes evidências contra a hipótese nula de não sazonalidade, uma vez que se pode rejeitar as hipóteses nulas de todos os testes com 99% de nível de confiança.

Esse período é claramente o período mais frio e com duração do dia menos prolongada do ano, fazendo com que as pessoas consumam mais energia em suas residências, uma vez que tomam banhos mais demorados na posição *inverno*, e procuram lazer em casa em detrimento do espaço público.

## **F. Criar modelos SARIMA**

Como as séries temporais apresentaram sazonalidade, todas as séries serão modeladas por meio do modelo SARIMA, conforme a Tabela 4. Importante destacar mais uma vez que cada linha da Tabela 4 será tratada como uma DMU para uma análise de eficiência por meio de Análise por Envoltória de Dados (DEA).

Desta maneira, considerando todas as combinações possíveis e variando os parâmetros computacionais de 0 a 3 há 255 DMUs a serem analisadas em cada uma das 6 séries temporais em estudo nesta seção. Salienta-se que a DMU que corresponde ao modelo SARIMA (0, 0, 0) (0, 0, 0)<sub>12</sub> foi retirada da análise por se tratar de um modelo baseado apenas na média.



## G. Modelar cada série por cada DMU

Aplicou-se para cada série temporal em análise, já com a primeira diferença (ou seja, sem tendência), o modelo SARIMA correspondente à cada linha da Tabela 4, realizando-se a previsão para o tamanho do teste selecionado.

Como há 224 observações em cada série temporal, os modelos retornarão 45 valores de previsão que serão comparados com os dados reais de teste. Já os 179 valores de ajuste serão comparados com os valores reais dos conjuntos de treino.

## H. Calcular as métricas de ajuste e de previsão e a quantidade de parâmetros estimados

Dado os modelos ajustados, calculam-se as métricas MAPE, RMSE, MdAE e MAE para o ajuste e para a previsão para cada DMU. Suas fórmulas de cálculo encontram-se nas Equações 15, 16, 17 e 18. Neste passo calculam-se também a quantidade de parâmetros estimados em cada modelo, somando-se os parâmetros da parte não sazonal dos modelos com a parte sazonal.

Como há 255 modelos a serem estimados (ver Tabela 2) para cada série temporal em análise, as Tabelas 12, 13, 15, 17, 19 e 21 apresentam as métricas de ajuste calculadas e os parâmetros estimados para as 5 primeiras e as 5 últimas DMUs. Já as Tabelas 12, 14, 16, 18, 20, 22 apresentam as métricas de previsão calculadas para as 5 primeiras e as 5 últimas DMUs.

Tabela 12. Métricas de ajuste e quantidade de parâmetros estimados para cada DMU da série de demanda de energia elétrica do Norte do Brasil.

DMU	SARIMA	$MAPE_{adj}$	$RMSE_{adj}$	$MdAE_{adj}$	$MAE_{adj}$	$k$
1	(1, 0, 0) (0, 0, 0) <sub>12</sub>	1,7071	60,2952	31,3376	43,2294	1
2	(2, 0, 0) (0, 0, 0) <sub>12</sub>	1,7474	60,2677	31,7243	43,0602	2
3	(3, 0, 0) (0, 0, 0) <sub>12</sub>	1,7262	60,1576	31,5215	42,7847	3
4	(0, 0, 1) (0, 0, 0) <sub>12</sub>	2,0967	60,6457	30,8732	43,0874	1
5	(1, 0, 1) (0, 0, 0) <sub>12</sub>	1,7498	60,2805	31,6602	43,1541	2
⋮	⋮	⋮	⋮	⋮	⋮	⋮
251	(3, 0, 2) (3, 0, 3) <sub>12</sub>	2,3595	47,0329	18,0140	30,5678	11
252	(0, 0, 3) (3, 0, 3) <sub>12</sub>	3,0672	47,6676	22,2717	31,5241	9
253	(1, 0, 3) (3, 0, 3) <sub>12</sub>	3,4241	47,6236	18,6562	30,7497	10
254	(2, 0, 3) (3, 0, 3) <sub>12</sub>	3,4098	46,1363	18,4415	30,1759	11
255	(3, 0, 3) (3, 0, 3) <sub>12</sub>	3,5085	42,2475	14,7439	27,3453	12

Fonte: elaboração própria.

Tabela 13. Métricas de previsão para cada DMU da série de demanda de energia elétrica do Norte do Brasil.

DMU	SARIMA	$MAPE_{prev}$	$RMSE_{prev}$	$MdAE_{prev}$	$MAE_{prev}$
1	(1, 0, 0) (0, 0, 0) <sub>12</sub>	3,0732	181,4284	110,0924	143,5481
2	(2, 0, 0) (0, 0, 0) <sub>12</sub>	3,0519	181,0247	109,9595	143,1934
3	(3, 0, 0) (0, 0, 0) <sub>12</sub>	3,0484	180,9956	109,9260	143,1792
4	(0, 0, 1) (0, 0, 0) <sub>12</sub>	3,0809	179,8345	109,8348	141,8804
5	(1, 0, 1) (0, 0, 0) <sub>12</sub>	3,0530	181,0358	109,9628	143,1939
⋮		⋮	⋮	⋮	⋮
251	(3, 0, 2) (3, 0, 3) <sub>12</sub>	2,9927	149,8562	97,8016	115,4885
252	(0, 0, 3) (3, 0, 3) <sub>12</sub>	2,5732	142,8052	87,1966	109,5344
253	(1, 0, 3) (3, 0, 3) <sub>12</sub>	2,9042	148,5016	95,7198	114,9762
254	(2, 0, 3) (3, 0, 3) <sub>12</sub>	3,3525	157,3348	102,4588	123,8728
255	(3, 0, 3) (3, 0, 3) <sub>12</sub>	3,2293	156,1232	103,9914	122,7318

Fonte: elaboração própria.

A Tabela 14 apresenta as estatísticas descritivas de cada métrica de ajuste e previsão, calculando-se também as estatísticas por número de parâmetros a fim de analisar o comportamento das métricas a medida que o número de parâmetros estimados entre as DMUs aumenta. Já as Figuras Figura 23, Figura 24, Figura 25 e Figura 26 apresentam os gráficos de Box-Plot comparando as métricas de ajuste e de previsão calculadas pelo número de parâmetros estimados.

Tabela 14. Estatísticas descritivas das métricas de ajuste e previsão para os modelos de demanda de energia elétrica da região Norte.

	Min.	1º Quartil	Mediana	Média	3º Quartil	Max.
$MAPE_{adj}$	1,2907	2,0932	2,3595	2,5438	3,0353	4,5446
$RMSE_{adj}$	2,5732	3,0084	3,0732	3,0911	3,1634	3,6036
$MdAE_{adj}$	42,2475	48,0157	48,9950	49,5824	50,1687	60,6457
$MAE_{adj}$	141,2436	152,7685	156,4895	159,2501	166,3530	181,9334
$MAPE_{prev}$	14,7439	19,7092	21,8272	22,3018	24,5105	32,2155
$RMSE_{prev}$	87,1966	97,1190	98,5812	100,4959	101,2800	120,2243
$MdAE_{prev}$	27,3453	31,3818	32,9006	33,5102	34,9874	43,2509
$MAE_{prev}$	109,5344	118,6015	121,3732	123,8140	129,1852	143,9148

Fonte: elaboração própria.

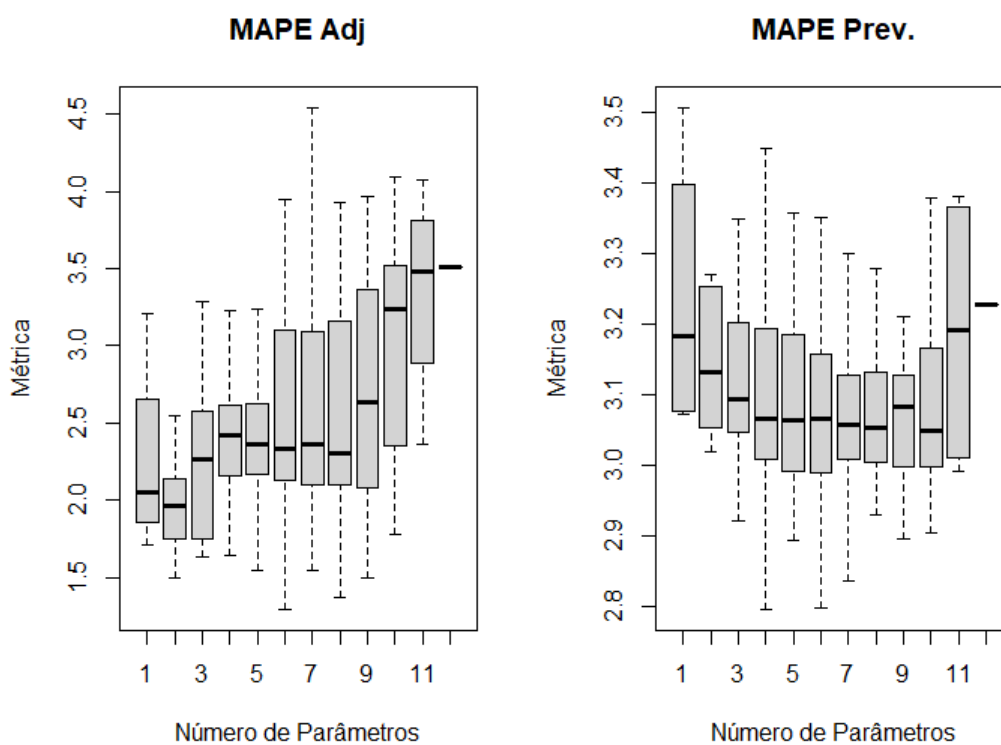


Figura 23. Box-Plot do MAPE do ajuste e da previsão por número de parâmetros dos modelos de previsão para a região Norte.  
 Fonte: elaboração própria.

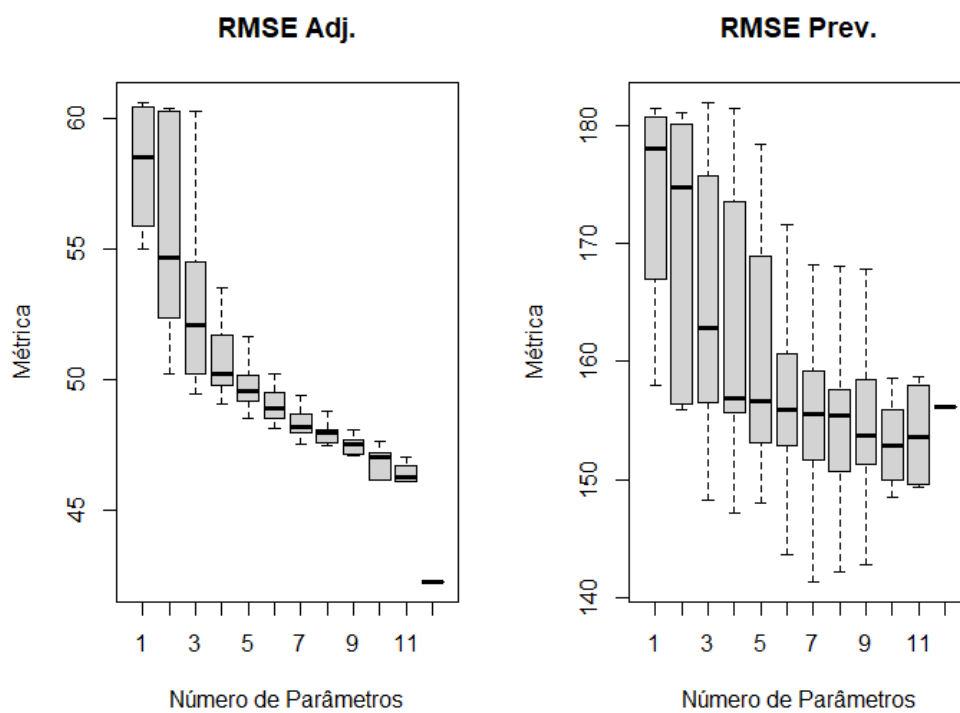


Figura 24. Box-Plot do MAPE do ajuste e da previsão por número de parâmetros dos modelos de previsão para a região Norte.  
 Fonte: elaboração própria.

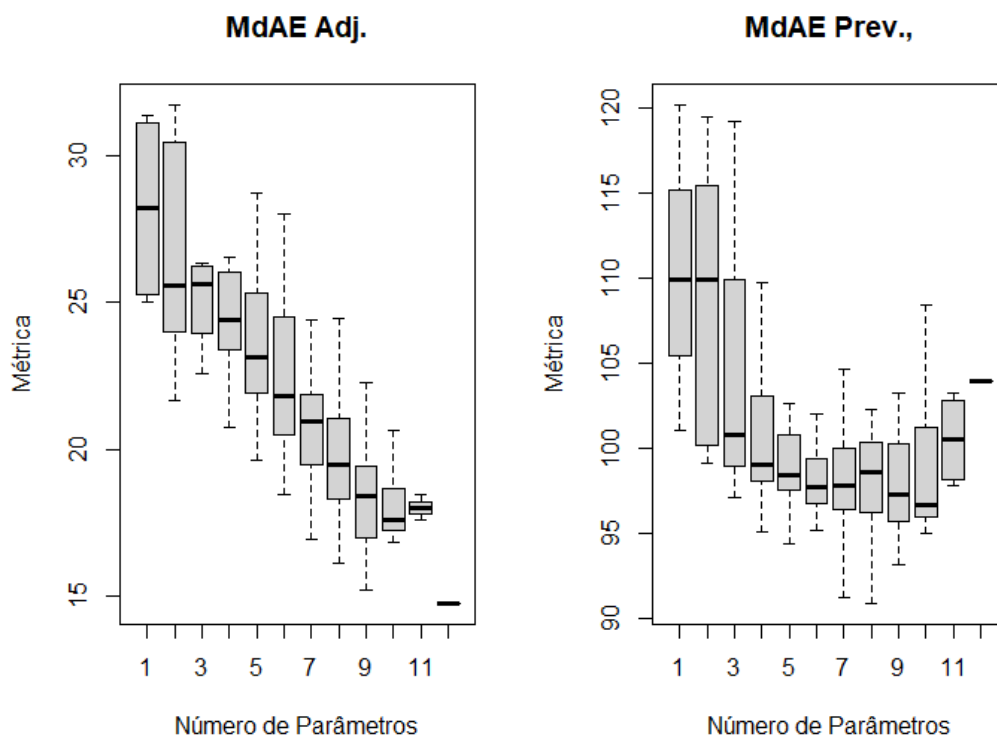


Figura 25. Box-Plot do MdAE do ajuste e da previsão por número de parâmetros dos modelos de previsão para a região Norte.

Fonte: elaboração própria.

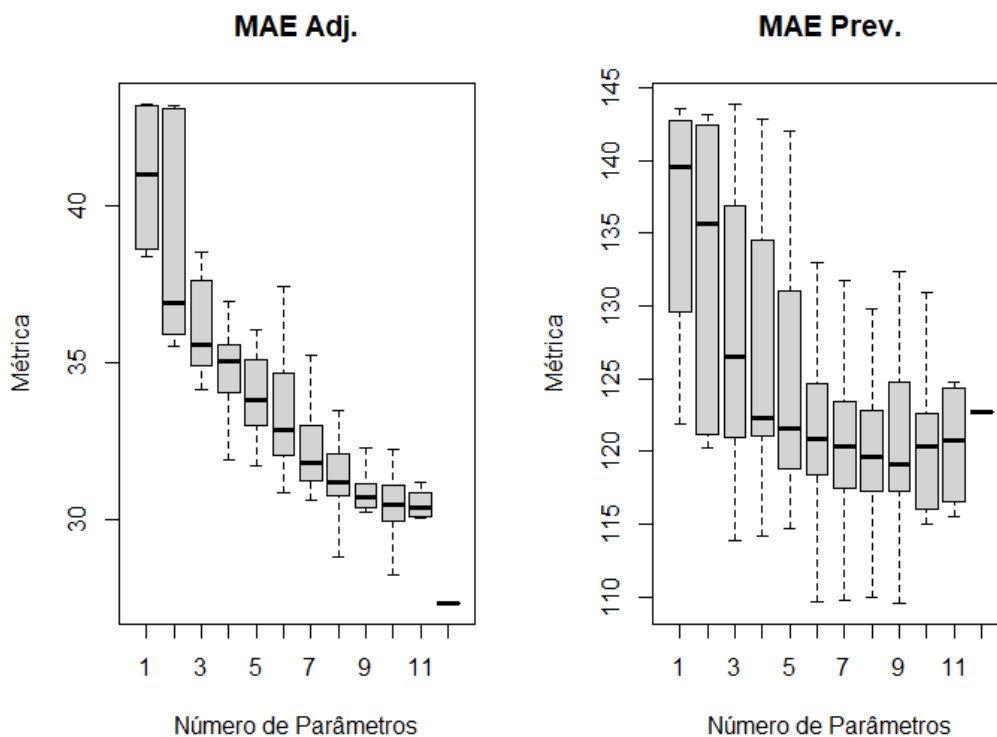


Figura 26. Box-Plot do MAE do ajuste e da previsão por número de parâmetros dos modelos de previsão para a região Norte.

Fonte: elaboração própria.

Observa-se que, com exceção do MAPE, que possui um comportamento instável (ver seção 2.1.4), as métricas diminuem o seu valor central em relação ao ajuste à medida que aumenta o número de parâmetros estimados. Porém, quando se analisa os Box-Plots em relação à previsão, observa-se as métricas caem até determinado nível de parâmetro, aumentando logo em seguida.

Tabela 15. Métricas de ajuste e quantidade de parâmetros estimados para cada DMU da série de demanda de energia elétrica do Nordeste do Brasil.

DMU	SARIMA	$MAPE_{adj}$	$RMSE_{adj}$	$MdAE_{adj}$	$MAE_{adj}$	$k$
1	(1, 0, 0) (0, 0, 0) <sub>12</sub>	8,2655	128,0262	76,2756	96,6329	1
2	(2, 0, 0) (0, 0, 0) <sub>12</sub>	8,9425	126,0937	81,2739	95,3947	2
3	(3, 0, 0) (0, 0, 0) <sub>12</sub>	8,4825	125,5917	77,2279	94,4427	3
4	(0, 0, 1) (0, 0, 0) <sub>12</sub>	6,1474	126,9185	76,0851	95,9334	1
5	(1, 0, 1) (0, 0, 0) <sub>12</sub>	6,0968	126,8933	75,9154	95,7593	2
⋮	⋮	⋮	⋮	⋮	⋮	⋮
251	(3, 0, 2) (3, 0, 3) <sub>12</sub>	8,8033	97,3996	47,1085	65,2434	11
252	(0, 0, 3) (3, 0, 3) <sub>12</sub>	6,7911	99,1988	47,8506	66,3796	9
253	(1, 0, 3) (3, 0, 3) <sub>12</sub>	3,1203	99,4272	47,1799	66,7647	10
254	(2, 0, 3) (3, 0, 3) <sub>12</sub>	5,1875	87,6065	54,5030	61,9241	11
255	(3, 0, 3) (3, 0, 3) <sub>12</sub>	9,0614	94,4445	54,4954	65,2476	12

Fonte: elaboração própria.

Tabela 16. Métricas de previsão para cada DMU da série de demanda de energia elétrica do Nordeste do Brasil.

DMU	SARIMA	$MAPE_{prev}$	$RMSE_{prev}$	$MdAE_{prev}$	$MAE_{prev}$
1	(1, 0, 0) (0, 0, 0) <sub>12</sub>	8,4128	304,9702	218,8257	249,2076
2	(2, 0, 0) (0, 0, 0) <sub>12</sub>	8,4121	296,4987	218,8033	246,4053
3	(3, 0, 0) (0, 0, 0) <sub>12</sub>	8,4253	293,7200	218,7657	244,1333
4	(0, 0, 1) (0, 0, 0) <sub>12</sub>	8,4027	301,6573	218,8259	248,1131
5	(1, 0, 1) (0, 0, 0) <sub>12</sub>	8,4035	301,5669	218,8259	248,2915
⋮	⋮	⋮	⋮	⋮	⋮
251	(3, 0, 2) (3, 0, 3) <sub>12</sub>	6,2451	311,7672	245,5961	256,5106
252	(0, 0, 3) (3, 0, 3) <sub>12</sub>	6,4980	313,4981	249,0493	257,2141
253	(1, 0, 3) (3, 0, 3) <sub>12</sub>	5,9017	313,1556	219,7797	259,0421
254	(2, 0, 3) (3, 0, 3) <sub>12</sub>	8,9140	319,4587	235,1545	262,9654
255	(3, 0, 3) (3, 0, 3) <sub>12</sub>	8,0606	319,3426	231,1107	264,9376

Fonte: elaboração própria.

Tabela 17. Estatísticas descritivas das métricas de ajuste e previsão para os modelos de demanda de energia elétrica da região Nordeste.

	Min.	1º Quartil	Mediana	Média	3º Quartil	Max.
$MAPE_{adj}$	1,4978	4,2284	5,8956	6,8893	9,1738	18,0332
$RMSE_{adj}$	4,2719	6,3459	6,4715	6,9331	7,4097	12,8556
$MdAE_{adj}$	87,6065	101,9839	104,4895	105,6502	107,5573	128,0262
$MAE_{adj}$	293,7200	313,2542	317,1332	316,0957	320,3253	326,3042
$MAPE_{prev}$	36,7221	53,7738	61,4128	60,8009	67,6413	81,2739
$RMSE_{prev}$	198,4997	211,4290	224,2871	225,2028	234,9596	264,7802
$MdAE_{prev}$	61,0516	71,2695	75,5621	76,5825	81,2657	96,6330
$MAE_{prev}$	244,0602	260,2875	263,1989	262,2003	265,4306	272,3996

Fonte: elaboração própria.

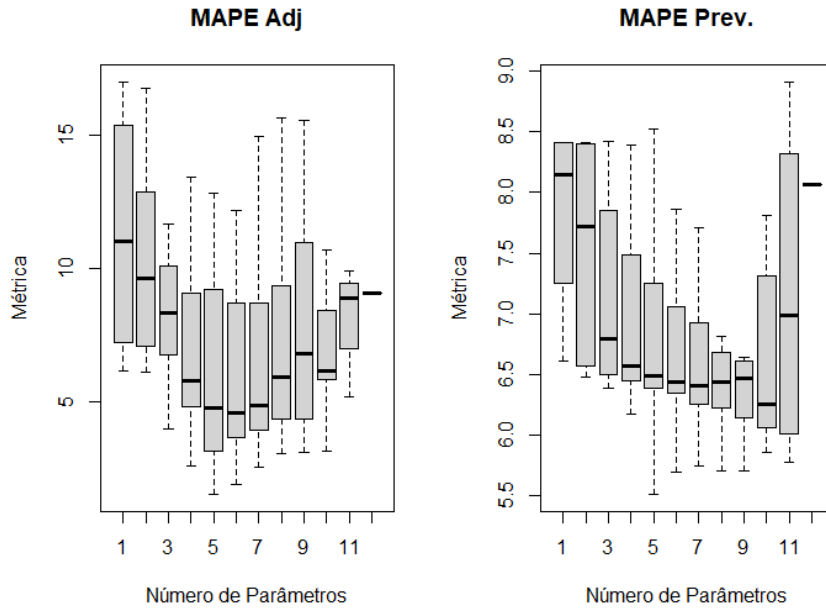


Figura 27. Box-Plot do MAPE do ajuste e da previsão por número de parâmetros dos modelos de previsão para a região Nordeste.  
Fonte: elaboração própria.

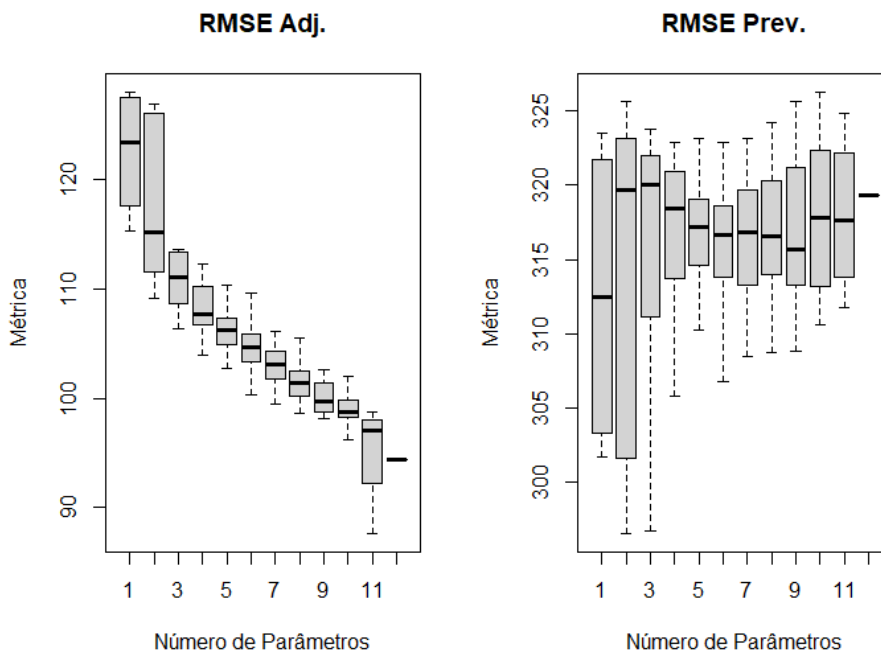


Figura 28. Box-Plot do RMSE do ajuste e da previsão por número de parâmetros dos modelos de previsão para a região Nordeste.  
Fonte: elaboração própria.

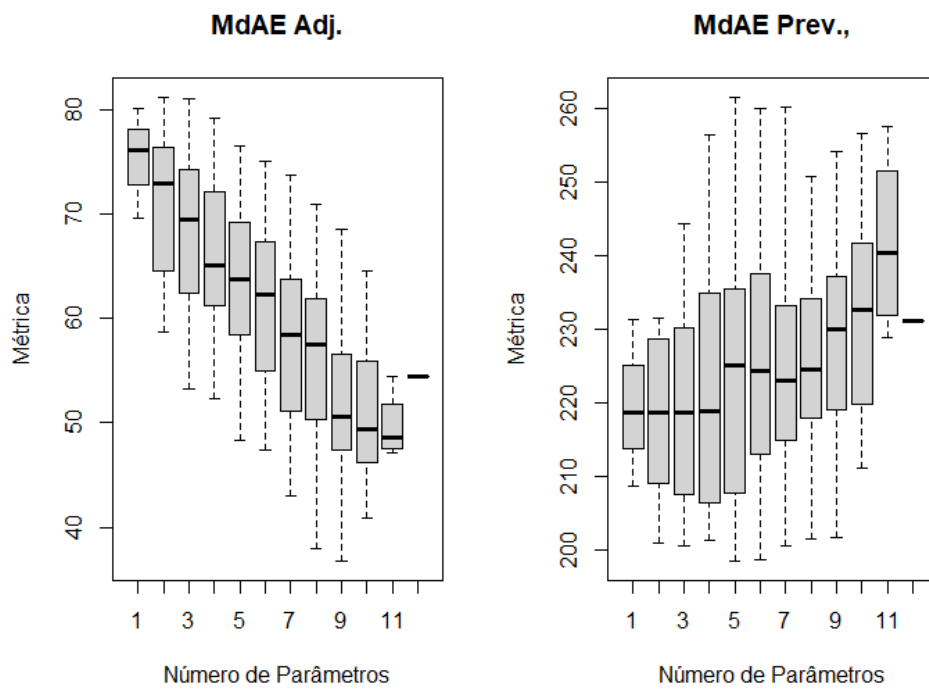


Figura 29. Box-Plot do MdAE do ajuste e da previsão por número de parâmetros dos modelos de previsão para a região Nordeste.  
 Fonte: elaboração própria.

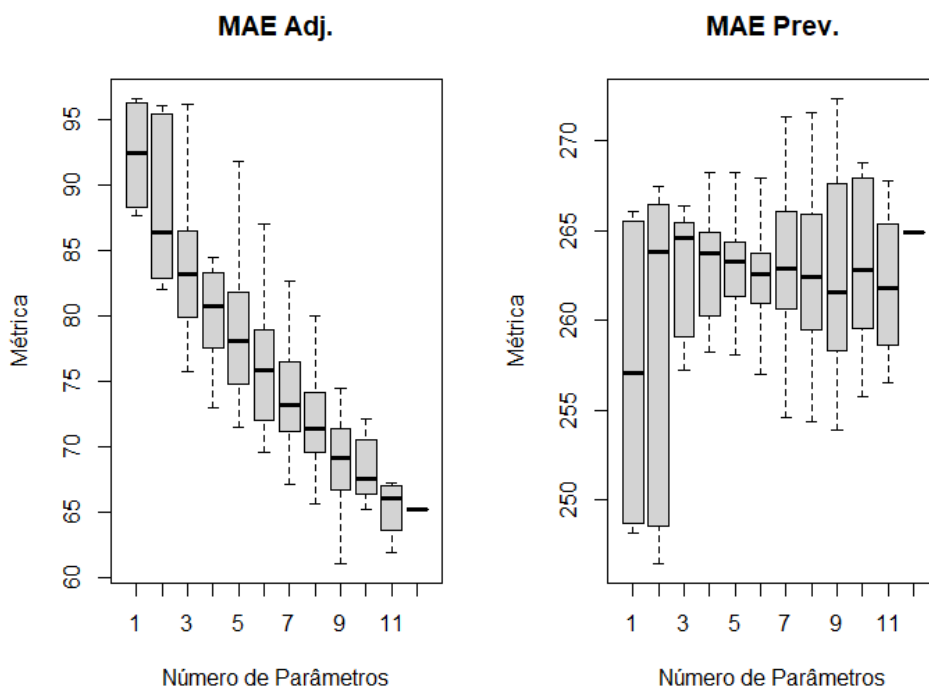


Figura 30. Box-Plot do MAE do ajuste e da previsão por número de parâmetros dos modelos de previsão para a região Nordeste.  
 Fonte: elaboração própria.

Tabela 18. Métricas de ajuste e quantidade de parâmetros estimados para cada DMU da série de demanda de energia elétrica do Sudeste do Brasil.

DMU	SARIMA	$MAPE_{adj}$	$RMSE_{adj}$	$MdAE_{adj}$	$MAE_{adj}$	$k$
1	(1, 0, 0) (0, 0, 0) <sub>12</sub>	1,1671	462,1916	314,8127	360,0597	1
2	(2, 0, 0) (0, 0, 0) <sub>12</sub>	1,2405	458,2455	323,9471	359,0605	2
3	(3, 0, 0) (0, 0, 0) <sub>12</sub>	1,3975	451,1931	294,7964	345,4398	3
4	(0, 0, 1) (0, 0, 0) <sub>12</sub>	1,184	462,4593	312,8026	361,5119	1
5	(1, 0, 1) (0, 0, 0) <sub>12</sub>	1,4919	445,1316	300,9595	347,4701	2
⋮	⋮	⋮	⋮	⋮	⋮	⋮
251	(3, 0, 2) (3, 0, 3) <sub>12</sub>	1,5139	329,4154	151,9639	226,6330	11
252	(0, 0, 3) (3, 0, 3) <sub>12</sub>	1,6746	346,6168	199,3150	248,2707	9
253	(1, 0, 3) (3, 0, 3) <sub>12</sub>	1,5872	328,7909	175,1819	225,8555	10
254	(2, 0, 3) (3, 0, 3) <sub>12</sub>	1,6494	342,6761	183,9287	241,7097	11
255	(3, 0, 3) (3, 0, 3) <sub>12</sub>	1,6173	308,3491	167,0271	215,0820	12

Fonte: elaboração própria.

Tabela 19. Métricas de previsão para cada DMU da série de demanda de energia elétrica do Sudeste do Brasil.

DMU	SARIMA	$MAPE_{prev}$	$RMSE_{prev}$	$MdAE_{prev}$	$MAE_{prev}$
1	(1, 0, 0) (0, 0, 0) <sub>12</sub>	0,9772	771,8741	459,3346	619,3316
2	(2, 0, 0) (0, 0, 0) <sub>12</sub>	0,9784	771,2863	460,1309	618,9171
3	(3, 0, 0) (0, 0, 0) <sub>12</sub>	0,9666	771,2167	462,6927	618,0497
4	(0, 0, 1) (0, 0, 0) <sub>12</sub>	0,9785	771,8363	460,6022	619,4543
5	(1, 0, 1) (0, 0, 0) <sub>12</sub>	0,9761	771,8623	459,6390	619,0674
⋮	⋮	⋮	⋮	⋮	⋮
251	(3, 0, 2) (3, 0, 3) <sub>12</sub>	1,3875	681,1295	520,1896	549,5513
252	(0, 0, 3) (3, 0, 3) <sub>12</sub>	1,4374	679,4799	512,4482	544,9259
253	(1, 0, 3) (3, 0, 3) <sub>12</sub>	1,3540	682,6964	485,9636	549,6393
254	(2, 0, 3) (3, 0, 3) <sub>12</sub>	1,5728	697,0937	501,4557	562,8096
255	(3, 0, 3) (3, 0, 3) <sub>12</sub>	1,5223	681,2694	491,3964	537,4825

Fonte: elaboração própria.

Tabela 20. Estatísticas descritivas das métricas de ajuste e previsão para os modelos de demanda de energia elétrica da região Nordeste.

	Min.	1° Quartil	Mediana	Média	3° Quartil	Max.
$MAPE_{adj}$	1,1671	1,6920	1,8640	1,8673	2,0527	2,2700
$RMSE_{adj}$	0,9576	1,1785	1,3449	1,3340	1,4494	3,2560
$MdAE_{adj}$	308,3491	353,7649	365,5192	370,6883	385,4332	462,4593
$MAE_{adj}$	636,2975	680,6006	733,2024	726,1021	771,2004	1077,8246
$MAPE_{prev}$	146,3487	201,6400	224,2535	226,3896	252,8991	323,9471
$RMSE_{prev}$	412,1434	490,0317	516,5058	512,3636	534,5847	623,7861
$MdAE_{prev}$	215,0820	255,7760	274,9154	276,9047	301,3615	361,5119
$MAE_{prev}$	506,5604	545,5869	594,2837	586,8122	628,3146	817,5988

Fonte: elaboração própria.



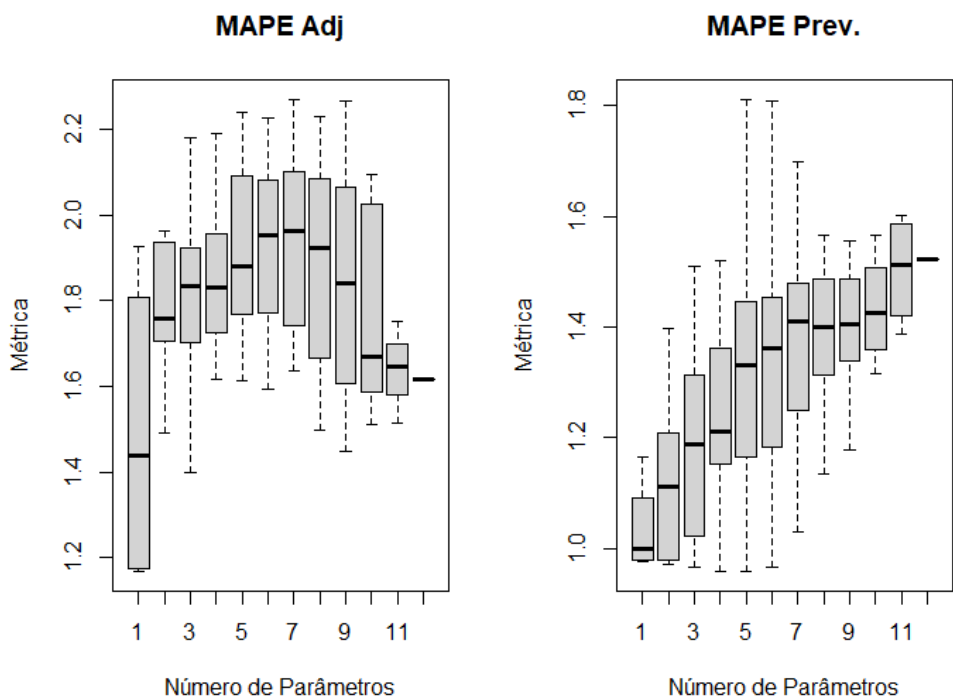


Figura 31. Box-Plot do MAPE do ajuste e da previsão por número de parâmetros dos modelos de previsão para a região Sudeste.  
 Fonte: elaboração própria.

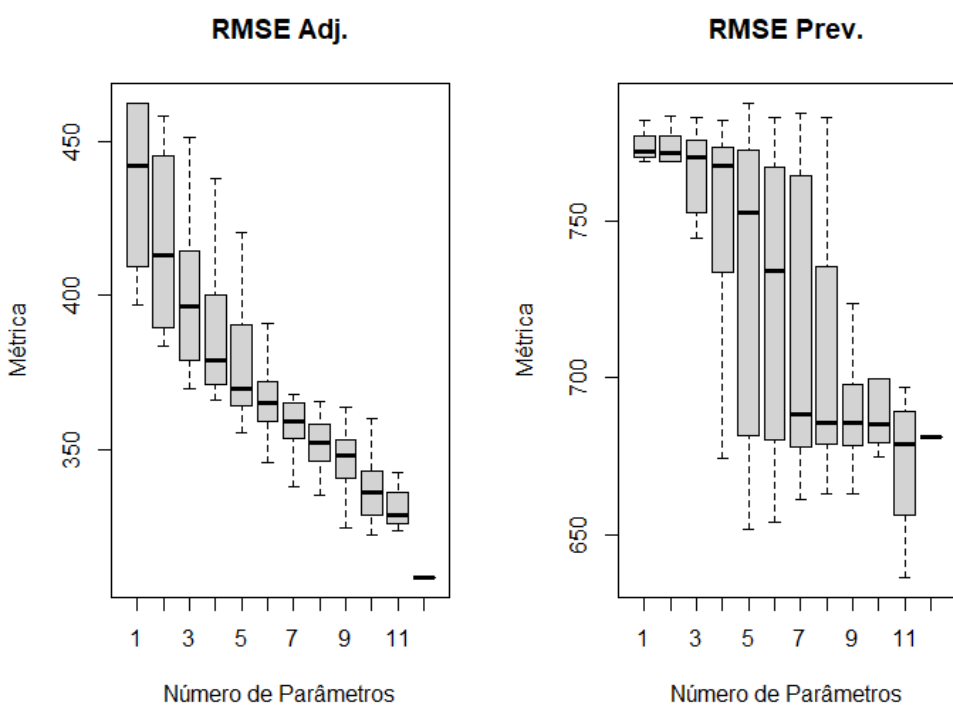


Figura 32. Box-Plot do MAPE do ajuste e da previsão por número de parâmetros dos modelos de previsão para a região Sudeste.  
 Fonte: elaboração própria.

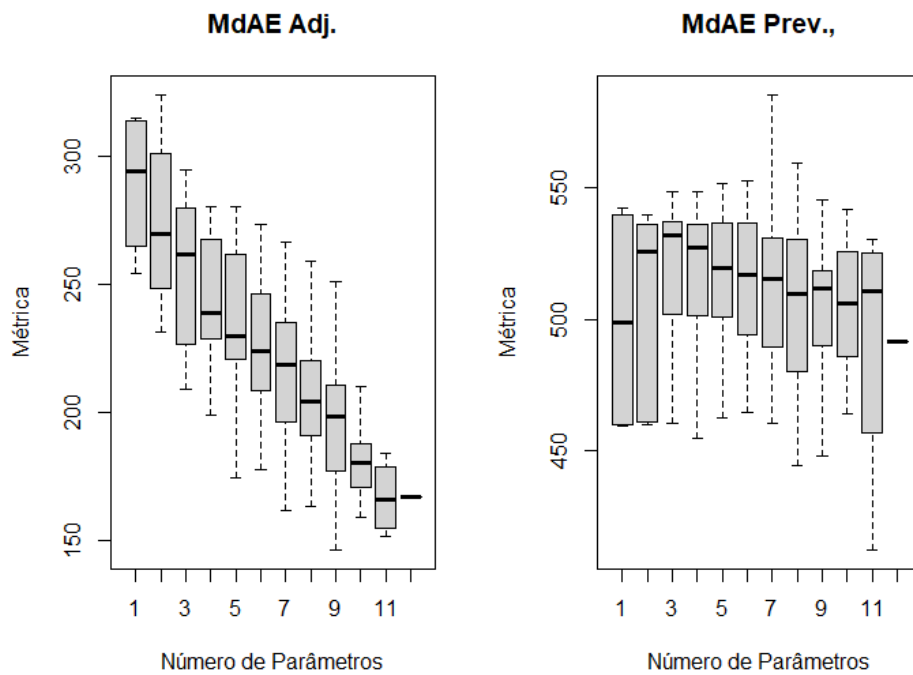


Figura 33. Box-Plot do MdAE do ajuste e da previsão por número de parâmetros dos modelos de previsão para a região Sudeste.  
 Fonte: elaboração própria.

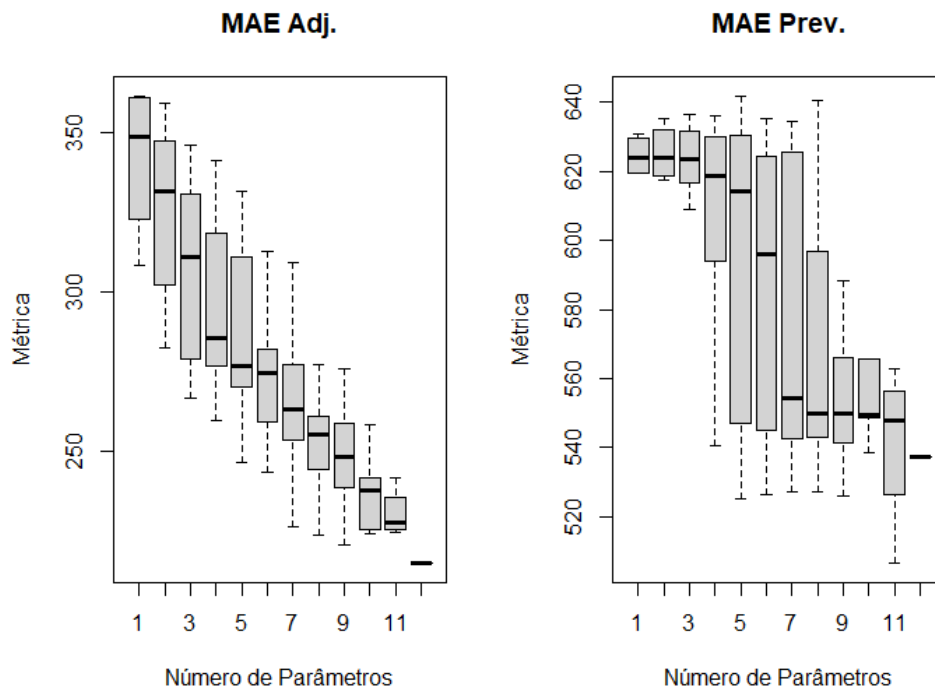


Figura 34. Box-Plot do MAE do ajuste e da previsão por número de parâmetros dos modelos de previsão para a região Sudeste.  
 Fonte: elaboração própria.

Tabela 21. Métricas de ajuste e quantidade de parâmetros estimados para cada DMU da série de demanda de energia elétrica do Sul do Brasil.

DMU	SARIMA	$MAPE_{adj}$	$RMSE_{adj}$	$MdAE_{adj}$	$MAE_{adj}$	$k$
1	(1, 0, 0) (0, 0, 0) <sub>12</sub>	1,4398	201,8629	119,3671	151,2507	1
2	(2, 0, 0) (0, 0, 0) <sub>12</sub>	1,4000	199,9100	115,1724	148,9534	2
3	(3, 0, 0) (0, 0, 0) <sub>12</sub>	1,5764	198,6931	123,6746	148,5392	3
4	(0, 0, 1) (0, 0, 0) <sub>12</sub>	1,4995	201,5301	119,2100	151,6672	1
5	(1, 0, 1) (0, 0, 0) <sub>12</sub>	1,4387	201,2354	118,3566	150,7589	2
⋮	⋮	⋮	⋮	⋮	⋮	⋮
251	(3, 0, 2) (3, 0, 3) <sub>12</sub>	1,2407	123,2196	50,0529	80,5387	11
252	(0, 0, 3) (3, 0, 3) <sub>12</sub>	1,1325	129,1647	61,6498	87,6141	9
253	(1, 0, 3) (3, 0, 3) <sub>12</sub>	1,0724	126,3667	57,2754	84,6657	10
254	(2, 0, 3) (3, 0, 3) <sub>12</sub>	0,9859	127,3406	54,9008	83,5529	11
255	(3, 0, 3) (3, 0, 3) <sub>12</sub>	1,1716	128,5959	50,4083	82,6976	12

Fonte: elaboração própria.

Tabela 22. Métricas de previsão para cada DMU da série de demanda de energia elétrica do Sul do Brasil.

DMU	SARIMA	$MAPE_{prev}$	$RMSE_{prev}$	$MdAE_{prev}$	$MAE_{prev}$
1	(1, 0, 0) (0, 0, 0) <sub>12</sub>	0,9342	324,5135	159,6575	246,8906
2	(2, 0, 0) (0, 0, 0) <sub>12</sub>	0,9411	325,1933	160,6351	247,6614
3	(3, 0, 0) (0, 0, 0) <sub>12</sub>	0,9452	325,8069	160,2299	248,3452
4	(0, 0, 1) (0, 0, 0) <sub>12</sub>	0,9287	324,6628	160,8823	246,6229
5	(1, 0, 1) (0, 0, 0) <sub>12</sub>	0,9307	324,7561	160,3851	246,7787
⋮	⋮	⋮	⋮	⋮	⋮
251	(3, 0, 2) (3, 0, 3) <sub>12</sub>	0,9340	291,2487	160,4311	221,7622
252	(0, 0, 3) (3, 0, 3) <sub>12</sub>	0,9018	298,6603	177,3408	225,0594
253	(1, 0, 3) (3, 0, 3) <sub>12</sub>	7,4817	1313,6460	710,8693	977,8858
254	(2, 0, 3) (3, 0, 3) <sub>12</sub>	1,0753	364,0116	186,4592	275,4128
255	(3, 0, 3) (3, 0, 3) <sub>12</sub>	1,0587	262,2072	163,2742	203,7499

Fonte: elaboração própria.

Tabela 23. Estatísticas descritivas das métricas de ajuste e previsão para os modelos de demanda de energia elétrica da região Sul.

	Min.	1° Quartil	Mediana	Média	3° Quartil	Max.
$MAPE_{adj}$	0,9859	1,1745	1,2604	1,3372	1,3948	2,7057
$RMSE_{adj}$	0,8465	0,9514	1,0234	2,1716	1,3277	142,6225
$MdAE_{adj}$	123,2196	136,0840	141,8330	146,5246	152,1503	201,8629
$MAE_{adj}$	236,0453	274,3138	298,8706	567,0811	324,1151	1313,6460
$MAPE_{prev}$	48,3225	68,4724	75,8461	77,6495	83,8419	123,6746
$RMSE_{prev}$	124,1201	160,7286	175,6078	256,7265	191,2283	857,9652
$MdAE_{prev}$	80,5387	94,7755	100,1142	104,0124	111,3295	151,6672
$MAE_{prev}$	189,4398	209,7979	227,1459	389,0512	246,5960	1356,8800

Fonte: elaboração própria.

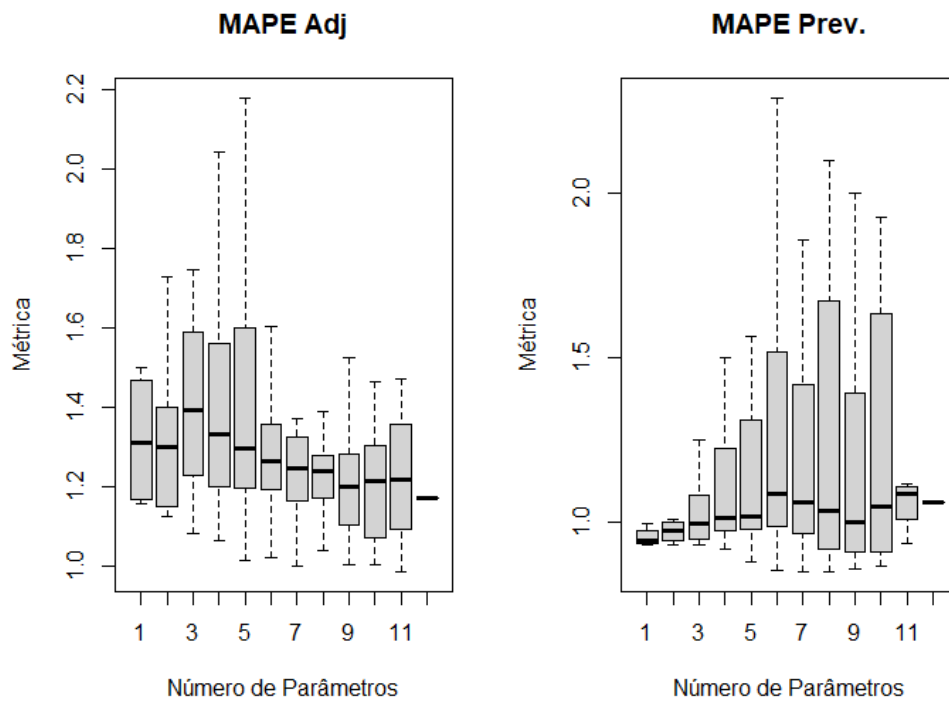


Figura 35. Box-Plot do MAPE do ajuste e da previsão por número de parâmetros dos modelos de previsão para a região Sul.  
 Fonte: elaboração própria.

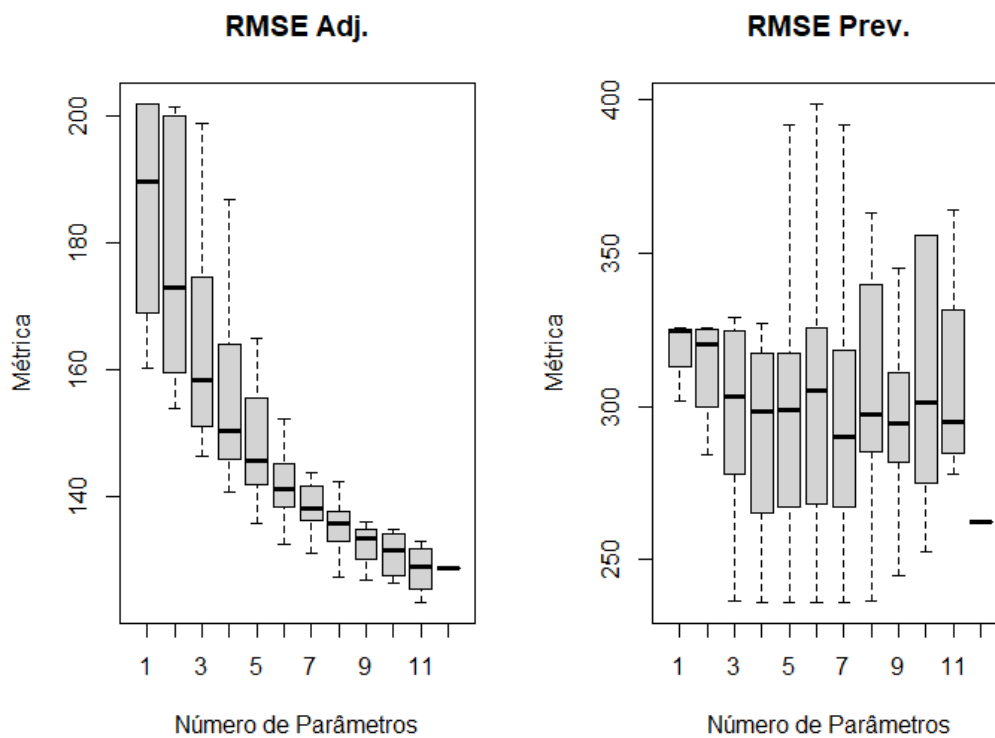


Figura 36. Box-Plot do RMSE do ajuste e da previsão por número de parâmetros dos modelos de previsão para a região Sul.  
 Fonte: elaboração própria.

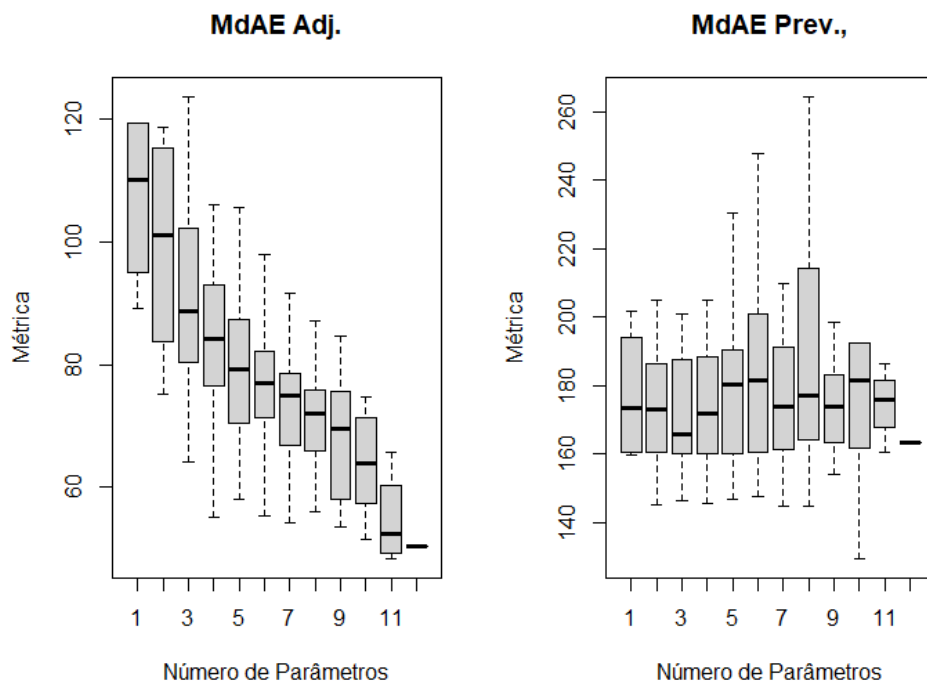


Figura 37. Box-Plot do MdAE do ajuste e da previsão por número de parâmetros dos modelos de previsão para a região Sul.  
 Fonte: elaboração própria.

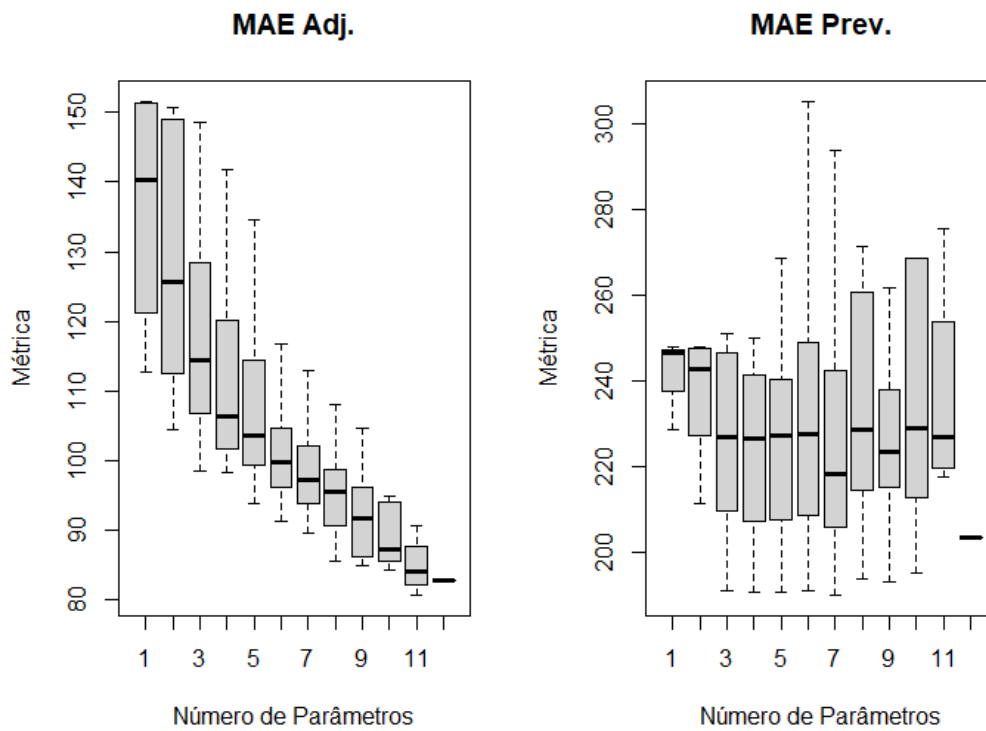


Figura 38. Box-Plot do MAE do ajuste e da previsão por número de parâmetros dos modelos de previsão para a região Sul.  
 Fonte: elaboração própria.

Tabela 24. Métricas de ajuste e quantidade de parâmetros estimados para cada DMU da série de demanda de energia elétrica do Centro-Oeste do Brasil.

DMU	SARIMA	$MAPE_{adj}$	$RMSE_{adj}$	$MdAE_{adj}$	$MAE_{adj}$	$k$
1	(1, 0, 0) (0, 0, 0) <sub>12</sub>	1,1983	77,4716	54,5791	61,1180	1
2	(2, 0, 0) (0, 0, 0) <sub>12</sub>	2,0087	74,5864	47,1198	58,0901	2
3	(3, 0, 0) (0, 0, 0) <sub>12</sub>	2,3396	70,5653	47,5622	54,7664	3
4	(0, 0, 1) (0, 0, 0) <sub>12</sub>	1,2722	77,4084	56,1771	61,1984	1
5	(1, 0, 1) (0, 0, 0) <sub>12</sub>	1,3091	76,8762	51,8879	60,3276	2
⋮	⋮	⋮	⋮	⋮	⋮	⋮
251	(3, 0, 2) (3, 0, 3) <sub>12</sub>	2,1318	52,1546	29,6936	37,6258	11
252	(0, 0, 3) (3, 0, 3) <sub>12</sub>	2,2283	52,1700	29,7974	37,6868	9
253	(1, 0, 3) (3, 0, 3) <sub>12</sub>	2,1963	51,9727	30,4592	37,4582	10
254	(2, 0, 3) (3, 0, 3) <sub>12</sub>	2,1735	51,9335	29,5270	37,3563	11
255	(3, 0, 3) (3, 0, 3) <sub>12</sub>	1,8681	48,0033	23,6639	34,4592	12

Fonte: elaboração própria.

Tabela 25. Métricas de previsão para cada DMU da série de demanda de energia elétrica do Centro-Oeste do Brasil.

DMU	SARIMA	$MAPE_{prev}$	$RMSE_{prev}$	$MdAE_{prev}$	$MAE_{prev}$
1	(1, 0, 0) (0, 0, 0) <sub>12</sub>	0,9798	126,1422	74,4682	97,9607
2	(2, 0, 0) (0, 0, 0) <sub>12</sub>	1,0434	126,4903	74,7084	99,2084
3	(3, 0, 0) (0, 0, 0) <sub>12</sub>	1,0695	126,5767	74,5214	98,4081
4	(0, 0, 1) (0, 0, 0) <sub>12</sub>	0,9808	126,1359	74,0485	98,0073
5	(1, 0, 1) (0, 0, 0) <sub>12</sub>	1,0005	126,2357	74,4054	98,3527
⋮	⋮	⋮	⋮	⋮	⋮
251	(3, 0, 2) (3, 0, 3) <sub>12</sub>	1,8906	124,3358	84,0257	96,8900
252	(0, 0, 3) (3, 0, 3) <sub>12</sub>	2,0117	127,0761	81,8195	99,8623
253	(1, 0, 3) (3, 0, 3) <sub>12</sub>	1,6563	121,5765	79,5954	93,2957
254	(2, 0, 3) (3, 0, 3) <sub>12</sub>	2,6081	142,0319	94,9495	114,2963
255	(3, 0, 3) (3, 0, 3) <sub>12</sub>	1,6157	150,6695	91,6864	121,7084

Fonte: elaboração própria.

Tabela 26. Estatísticas descritivas das métricas de ajuste e previsão para os modelos de demanda de energia elétrica da região Centro-Oeste.

	Min.	1º Quartil	Mediana	Média	3º Quartil	Max.
$MAPE_{adj}$	1,1983	2,2378	2,4617	2,3990	2,5793	3,5369
$RMSE_{adj}$	0,9468	1,2464	1,4700	1,4559	1,6065	2,6554
$MdAE_{adj}$	47,1316	53,6437	55,2163	57,6050	60,6735	77,4716
$MAE_{adj}$	117,4556	119,4546	121,8343	123,2802	126,5496	151,6703
$MAPE_{prev}$	22,0202	31,0279	32,8809	34,9808	38,6308	56,1771
$RMSE_{prev}$	66,4807	74,4368	77,7322	78,4318	82,8855	105,5888
$MdAE_{prev}$	33,8754	39,1300	41,5577	43,1841	46,8465	61,1984
$MAE_{prev}$	90,3266	91,6748	95,6715	95,8323	98,0202	126,4193

Fonte: elaboração própria.

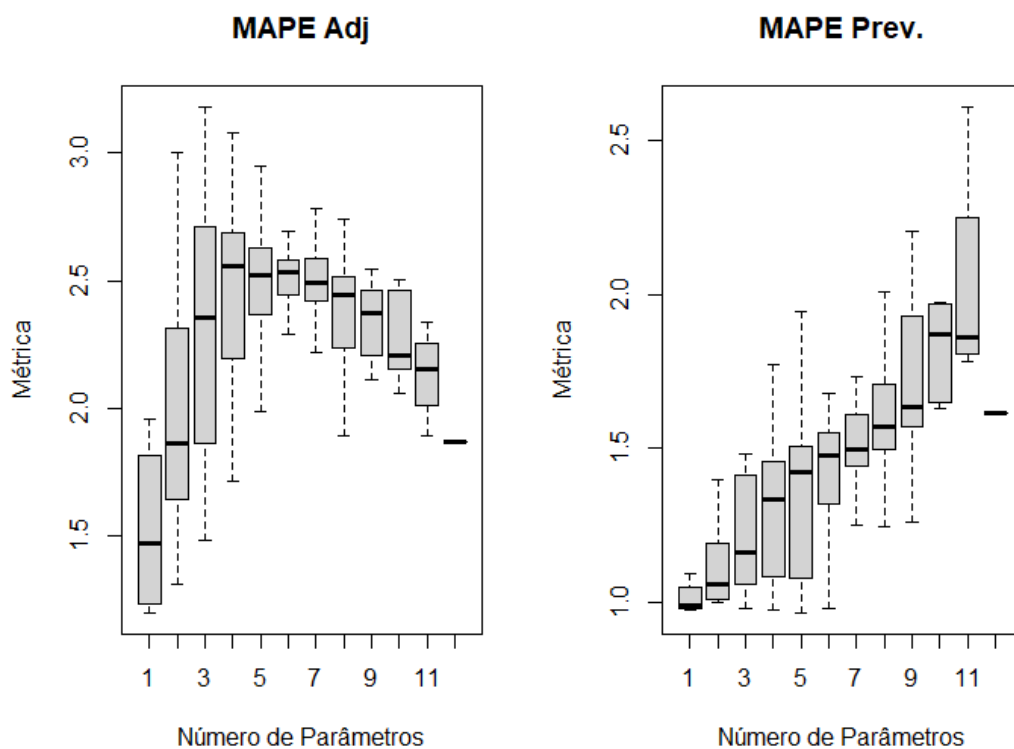


Figura 39. Box-Plot do MAPE do ajuste e da previsão por número de parâmetros dos modelos de previsão para a região Centro-Oeste.  
 Fonte: elaboração própria.

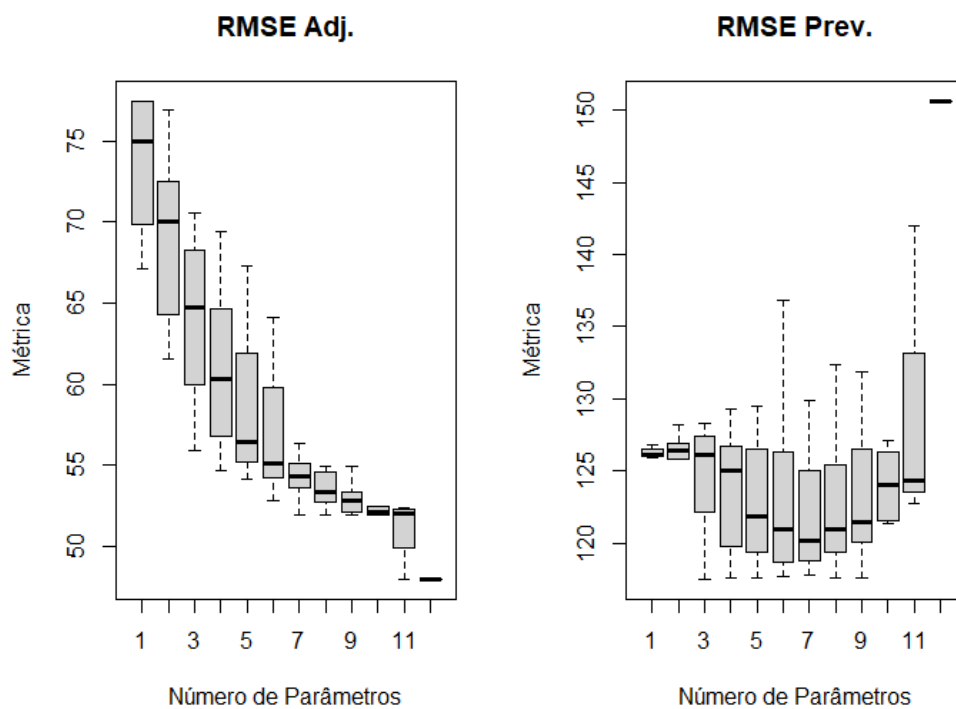


Figura 40. Box-Plot do RMSE do ajuste e da previsão por número de parâmetros dos modelos de previsão para a região Centro-Oeste.  
 Fonte: elaboração própria.

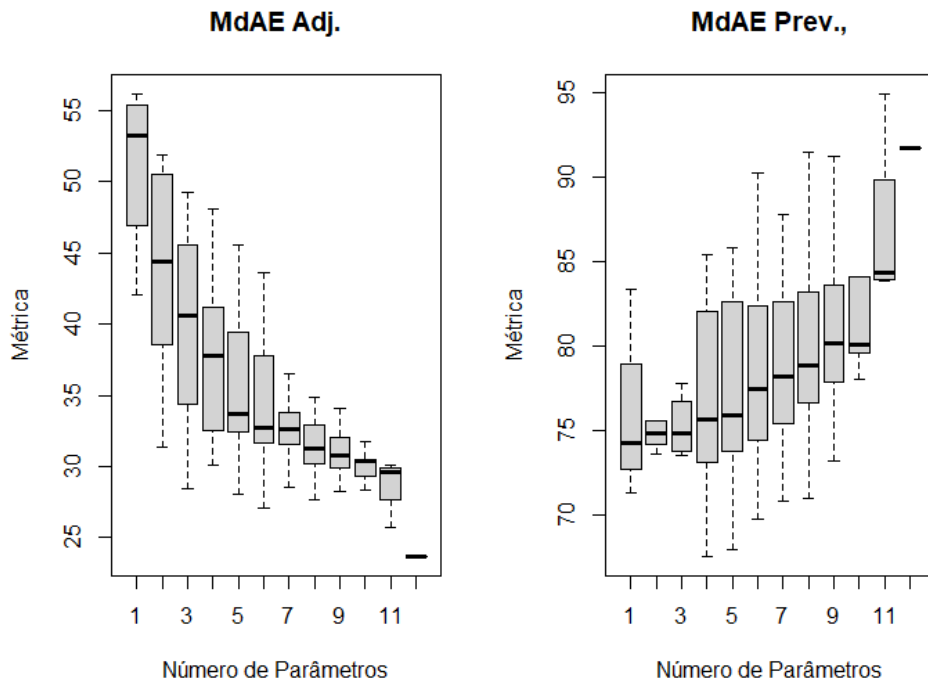


Figura 41. Box-Plot do MdAE do ajuste e da previsão por número de parâmetros dos modelos de previsão para a região Centro-Oeste.  
 Fonte: elaboração própria.

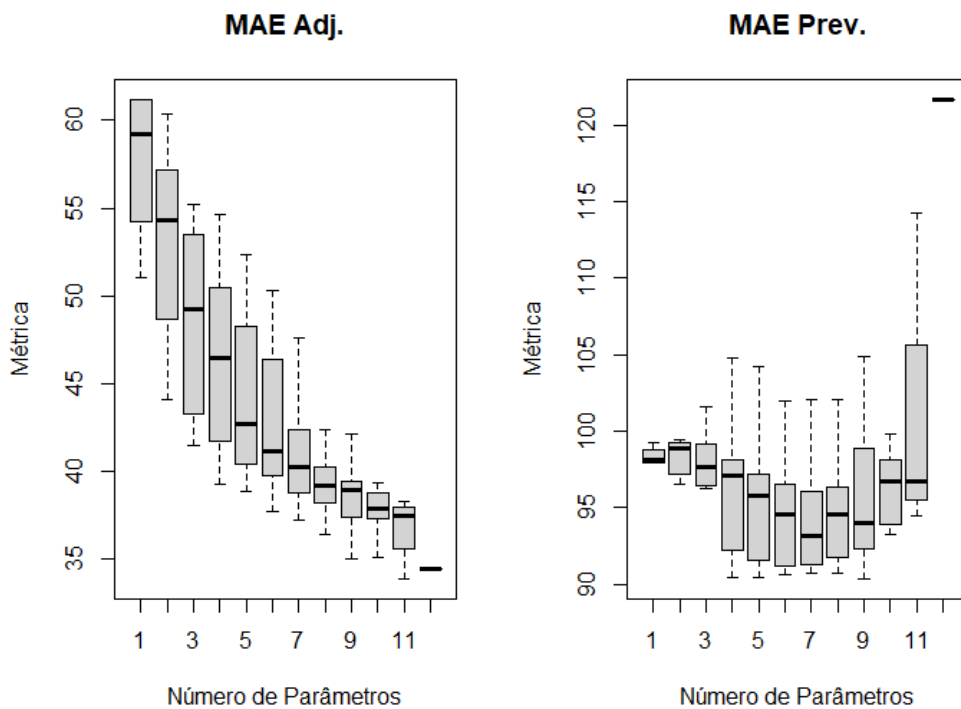


Figura 42. Box-Plot do MdAE do ajuste e da previsão por número de parâmetros dos modelos de previsão para a região Centro-Oeste.  
 Fonte: elaboração própria.



Tabela 27. Métricas de ajuste e quantidade de parâmetros estimados para cada DMU da série de demanda de energia elétrica do Brasil.

DMU	SARIMA	$MAPE_{adj}$	$RMSE_{adj}$	$MdAE_{adj}$	$MAE_{adj}$	$k$
1	(1, 0, 0) (0, 0, 0) <sub>12</sub>	1,4381	703,5226	460,6224	570,6464	1
2	(2, 0, 0) (0, 0, 0) <sub>12</sub>	1,7301	688,6120	488,1998	556,6008	2
3	(3, 0, 0) (0, 0, 0) <sub>12</sub>	1,5504	682,4662	429,1477	538,0519	3
4	(0, 0, 1) (0, 0, 0) <sub>12</sub>	1,4831	704,2698	475,8100	573,8353	1
5	(1, 0, 1) (0, 0, 0) <sub>12</sub>	1,6056	698,5090	465,9786	570,3519	2
⋮	⋮	⋮	⋮	⋮	⋮	⋮
251	(3, 0, 2) (3, 0, 3) <sub>12</sub>	1,7526	493,6236	226,3287	338,1163	11
252	(0, 0, 3) (3, 0, 3) <sub>12</sub>	1,7788	510,0439	279,6171	362,2392	9
253	(1, 0, 3) (3, 0, 3) <sub>12</sub>	1,6045	482,9406	248,3265	339,0706	10
254	(2, 0, 3) (3, 0, 3) <sub>12</sub>	1,4271	479,8770	220,6914	331,6799	11
255	(3, 0, 3) (3, 0, 3) <sub>12</sub>	1,7131	477,1267	219,9473	334,6675	12

Fonte: elaboração própria.

Tabela 28. Métricas de previsão para cada DMU da série de demanda de energia elétrica do Brasil.

DMU	SARIMA	$MAPE_{prev}$	$RMSE_{prev}$	$MdAE_{prev}$	$MAE_{prev}$
1	(1, 0, 0) (0, 0, 0) <sub>12</sub>	1,1222	1.304,7142	848,1811	1.062,0157
2	(2, 0, 0) (0, 0, 0) <sub>12</sub>	1,0865	1.301,4191	851,7315	1.059,1040
3	(3, 0, 0) (0, 0, 0) <sub>12</sub>	1,1162	1.300,9726	854,4623	1.058,7232
4	(0, 0, 1) (0, 0, 0) <sub>12</sub>	1,1159	1.304,4433	851,7025	1.061,5900
5	(1, 0, 1) (0, 0, 0) <sub>12</sub>	1,1187	1.304,7630	851,3435	1.062,4010
⋮	⋮	⋮	⋮	⋮	⋮
251	(3, 0, 2) (3, 0, 3) <sub>12</sub>	5,0147	1.901,7642	1.244,8010	1.543,6872
252	(0, 0, 3) (3, 0, 3) <sub>12</sub>	1,6422	1.541,6144	920,9796	1.195,8654
253	(1, 0, 3) (3, 0, 3) <sub>12</sub>	4,3452	2.291,8183	1.216,4043	1.780,7928
254	(2, 0, 3) (3, 0, 3) <sub>12</sub>	3,8355	1.852,6292	1.118,9397	1.436,7231
255	(3, 0, 3) (3, 0, 3) <sub>12</sub>	2,6205	1.333,7248	881,3352	1.076,6730

Tabela 29. Estatísticas descritivas das métricas de ajuste e previsão para os modelos de demanda de energia elétrica do Brasil.

	Min.	1º Quartil	Mediana	Média	3º Quartil	Max.
$MAPE_{adj}$	1,4271	1,8212	2,3210	2,2392	2,6133	3,2259
$RMSE_{adj}$	1,0865	1,4158	1,5023	1,8112	1,6364	55,0684
$MdAE_{adj}$	466,6483	522,6213	543,7879	553,0188	576,5249	704,2698
$MAE_{adj}$	1.158,4280	1.215,4480	1.298,2690	1.366,0660	1.315,7390	3.765,6880
$MAPE_{prev}$	219,4484	302,4760	331,0562	341,7956	384,6973	503,9837
$RMSE_{prev}$	658,0144	801,9704	863,0888	918,0308	919,0988	9.585,0878
$MdAE_{prev}$	318,5663	380,5532	402,7781	416,0109	452,6146	573,8353
$MAE_{prev}$	913,3215	956,6238	1.058,7228	1101,7271	1.096,9997	3.207,1790

Fonte: elaboração própria.

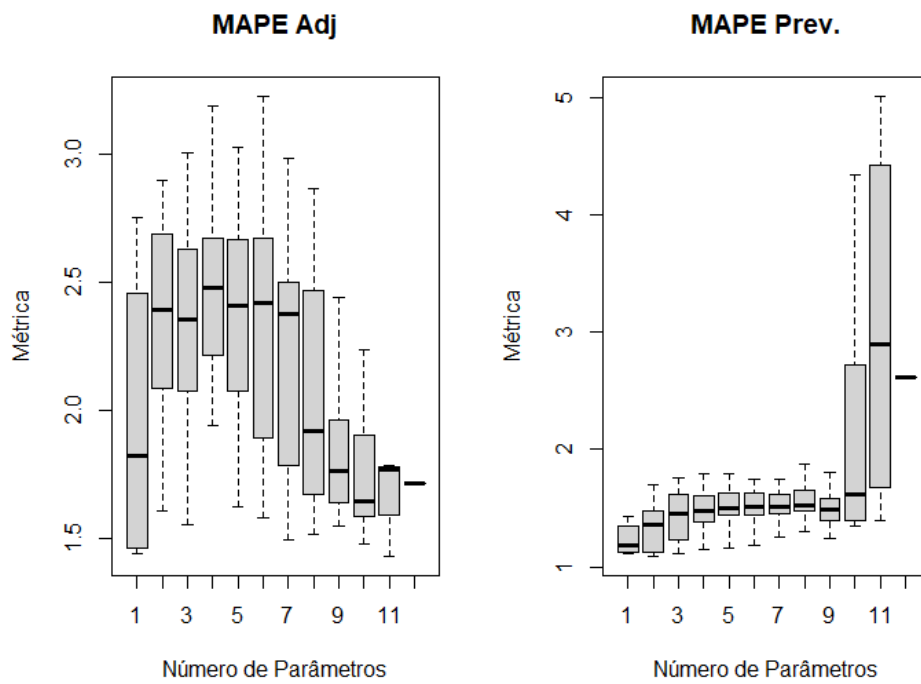


Figura 43. Box-Plot do MAPE do ajuste e da previsão por número de parâmetros dos modelos de previsão do Brasil.  
 Fonte: elaboração própria.

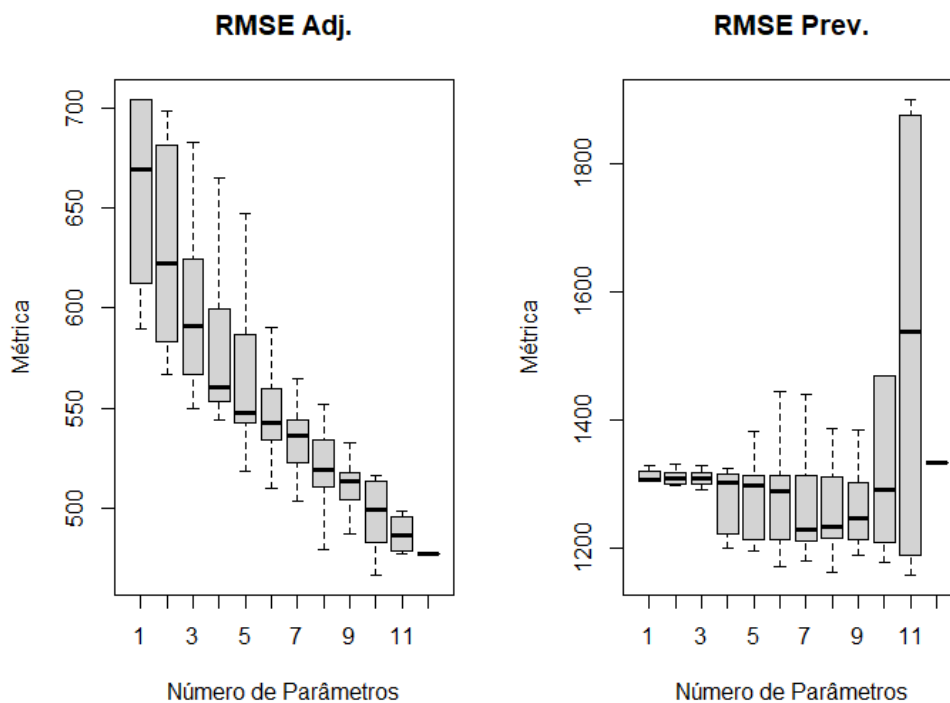


Figura 44. Box-Plot do RMSE do ajuste e da previsão por número de parâmetros dos modelos de previsão do Brasil.  
 Fonte: elaboração própria.

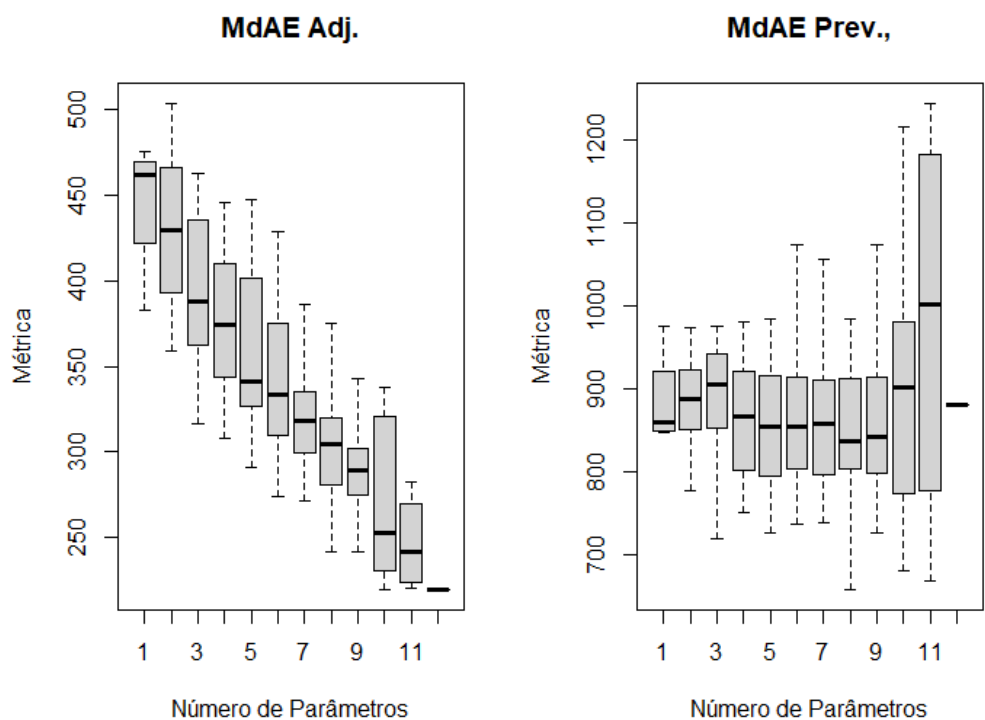


Figura 45. Box-Plot do MdAE do ajuste e da previsão por número de parâmetros dos modelos de previsão do Brasil.  
 Fonte: elaboração própria.

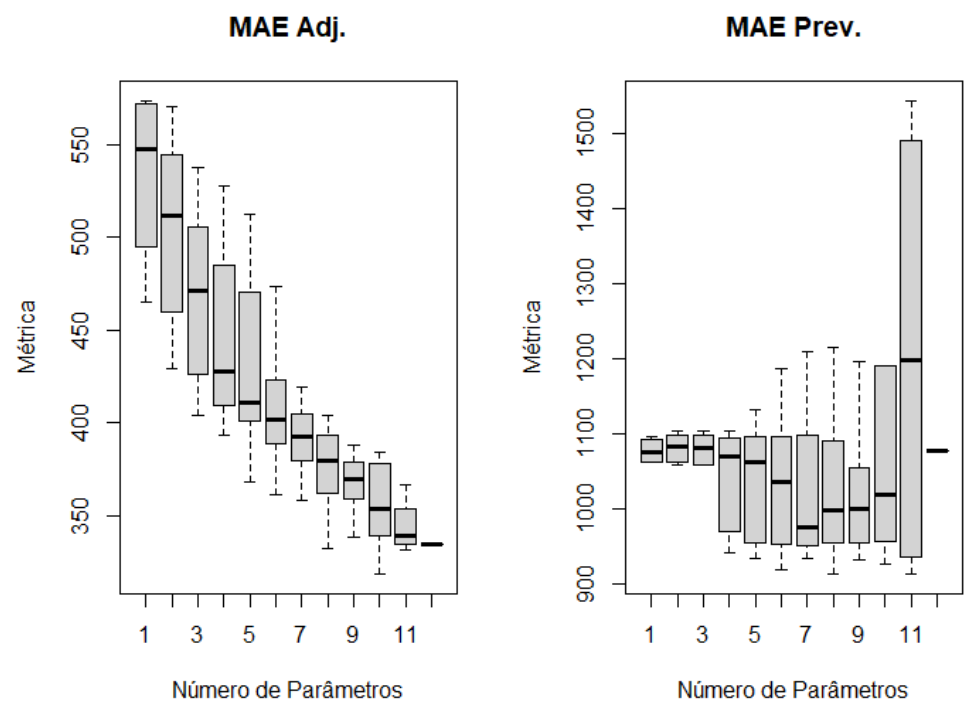


Figura 46. Box-Plot do MAE do ajuste e da previsão por número de parâmetros dos modelos de previsão do Brasil.  
 Fonte: elaboração própria.

## I. Definir os *inputs* e os *outputs*

Os *inputs* e os *outputs* para posterior aplicação do modelo DEA foram selecionados segundo a Tabela 3, isto é, as métricas de acurácia em relação ao ajuste e a quantidade de parâmetros são considerados *inputs* e o inverso das métricas de acurácia em relação à previsão são consideradas como *outputs*.

## J. Aplicar o modelo DEA adequado

Observa-se pelos *inputs* e *outputs* selecionados que há variáveis fracionadas a serem modeladas. Desta maneira, como discutido na seção 2.4, os modelos clássicos CCR e BCC geram problemas de convexidade na fronteira de eficiência.

Assim, este trabalho propõe a utilização de um modelo híbrido que une o modelo DEA-R, discutido na seção 2.4 e o modelo de Supereficiência, selecionando-se as quatro DMUs mais eficientes de cada série temporal analisada (WANG, SMITH-MILES & HYNDMAN, 2009).

O modelo DEA proposto é dado pela Equação 53.

$$\text{Max } Eff_0 = \sum_{j=1}^m c_{(j,i)} r_{(ji)0} \quad (53)$$

Sujeito a:

$$\sum_{j=1}^s \sum_{i=1}^m c_{(j,i)} r_{(ji)k} \leq 1, \forall k, k \neq 0$$

$$c_{(j,i)} \geq 0, \forall i, j$$

Nota-se que a DMU em análise na iteração  $j$  é retirada das restrições, permitindo que seu escore de eficiência seja maior que 1.

## K. Definir as DMUs mais eficientes

As quatro DMUs mais eficientes para cada série analisada e seus respectivos modelos encontram-se na Tabela 30. Observa-se que a DMU125 (modelo SARIMA (1, 0, 3) (3, 0, 1)) se encontra entre as quatro mais eficientes em todas as regiões analisadas. A DMU3 (SARIMA(3, 0, 0) (0, 0, 0)) se apresentou como uma das quatro mais eficientes em 4 dos 6

modelos aplicados nas regiões avaliadas. Importante destacar que entre as 24 DMUs, 18 possuem o componente sazonal.

Tabela 30. DMUs eficientes por região analisada e seus respectivos modelos.

Região	DMUs	SARIMA <sub>12</sub>
Norte	(1) 63	(3, 0, 3) (3, 0, 0)
	(2) 125	(1, 0, 3) (3, 0, 1)
	(3) 3	(3, 0, 0) (0, 0, 0)
	(4) 254	(2, 0, 3) (3, 0, 3)
Nordeste	(1) 61	(1, 0, 3) (3, 0, 0)
	(2) 63	(3, 0, 3) (3, 0, 0)
	(3) 26	(2, 0, 2) (1, 0, 0)
	(4) 125	(1, 0, 3) (3, 0, 1)
Sudeste	(1) 125	(1, 0, 3) (3, 0, 1)
	(2) 26	(2, 0, 2) (1, 0, 0)
	(3) 3	(3, 0, 0) (0, 0, 0)
	(4) 61	(1, 0, 3) (3, 0, 0)
Sul	(1) 182	(2, 0, 1) (3, 0, 2)
	(2) 3	(3, 0, 0) (0, 0, 0)
	(3) 79	(3, 0, 3) (0, 0, 1)
	(4) 125	(1, 0, 3) (3, 0, 1)
Centro-Oeste	(1) 26	(2, 0, 2) (1, 0, 0)
	(2) 15	(3, 0, 3) (0, 0, 0)
	(3) 61	(1, 0, 3) (3, 0, 0)
	(4) 125	(1, 0, 3) (3, 0, 1)
Brasil	(1) 3	(3, 0, 0) (0, 0, 0)
	(2) 125	(1, 0, 3) (3, 0, 1)
	(3) 190	(2, 0, 3) (3, 0, 2)
	(4) 15	(3, 0, 3) (0, 0, 0)

Fonte: elaboração própria.

#### L. Definir um Delineamento de Misturas adequado

O delineamento de misturas selecionado foi o *Simplex Lattice Design* (SLD) de nível (4, 5) que apresenta apenas pontos nos extremos do vértice, conforme discutido na seção 2.7. Foram adicionados 5 pontos adicionais (4 axiais e 1 central) para permitir a análise de presença de curvatura e posterior otimização, totalizando 61 linhas experimentais.

#### M. Calcular as métricas de previsão segundo o Delineamento de Misturas

Com o desenho de misturas criado, cada linha corresponderá ao peso que será dado para combinar as previsões das DMUs mais eficientes em cada região analisada. As Tabelas 31, 32, 33, 34, 35 e 36 apresentam as cinco primeiras corridas experimentais e suas respectivas respostas. Os fatores  $x_1$ ,  $x_2$ ,  $x_3$  e  $x_4$  representam os pesos para as 4 DMUs mais eficientes em relação ao modelo DEA proposto.

Tabela 31. Delineamento de Misturas SLD (4, 5) com as métricas de performance de previsão em relação ao conjunto teste como respostas para a região Norte.

	$x_1$	$x_2$	$x_3$	$x_4$	$MAPE_{ensb}$	$RMSE_{ensb}$	$MdAE_{ensb}$	$MAE_{ensb}$
1	1,000	0,000	0,000	0,000	3,4231	167,8353	114,8672	132,3517
2	0,800	0,200	0,000	0,000	3,3502	165,0178	111,0242	129,4349
3	0,800	0,000	0,200	0,000	3,2789	162,3796	105,4976	126,6702
4	0,800	0,000	0,000	0,200	3,2157	159,9298	102,8383	124,5682
5	0,600	0,400	0,000	0,000	3,1541	157,6771	99,5403	122,5289
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
57	0,250	0,250	0,250	0,250	3,2144	163,5220	103,7553	128,9664
58	0,625	0,125	0,125	0,125	3,3085	165,3921	112,4052	129,6511
59	0,125	0,625	0,125	0,125	3,1485	159,2804	100,1396	124,5709
60	0,125	0,125	0,625	0,125	3,1314	171,2115	98,6174	136,0729
61	0,125	0,125	0,125	0,625	3,2740	160,1745	103,3932	126,0401

Fonte: elaboração própria.

Tabela 32. Delineamento de Misturas SLD (4, 5) com as métricas de performance de previsão em relação ao conjunto teste como respostas para a região Nordeste.

	$x_1$	$x_2$	$x_3$	$x_4$	$MAPE_{ensb}$	$RMSE_{ensb}$	$MdAE_{ensb}$	$MAE_{ensb}$
1	1,000	0,000	0,000	0,000	7,0700	316,6214	260,5756	262,9777
2	0,800	0,200	0,000	0,000	7,9923	316,3088	246,5536	263,9877
3	0,800	0,000	0,200	0,000	8,9673	316,6126	232,5316	265,4180
4	0,800	0,000	0,000	0,200	9,9433	317,5311	218,5096	267,1327
5	0,600	0,400	0,000	0,000	10,9193	319,0588	217,2263	268,8525
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
57	0,250	0,250	0,250	0,250	9,0660	314,5493	224,9996	264,4835
58	0,625	0,125	0,125	0,125	8,0408	315,1423	242,7876	263,4256
59	0,125	0,625	0,125	0,125	10,4794	317,2379	207,7326	267,3674
60	0,125	0,125	0,625	0,125	10,1904	312,0016	239,3280	265,2535
61	0,125	0,125	0,125	0,625	7,5749	316,9143	236,2239	264,1182

Fonte: elaboração própria.

Tabela 33. Delineamento de Misturas SLD (4, 5) com as métricas de performance de previsão em relação ao conjunto teste como respostas para a região Sudeste.

	$x_1$	$x_2$	$x_3$	$x_4$	$MAPE_{ensb}$	$RMSE_{ensb}$	$MdAE_{ensb}$	$MAE_{ensb}$
1	1,000	0,000	0,000	0,000	1,5349	877,1199	480,0261	688,4817
2	0,800	0,200	0,000	0,000	1,4252	848,6699	435,9261	675,0377
3	0,800	0,000	0,200	0,000	1,3231	823,3518	474,7690	661,9726
4	0,800	0,000	0,000	0,200	1,2382	801,4624	526,2070	649,6263
5	0,600	0,400	0,000	0,000	1,1533	783,2895	529,7142	637,2801
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
57	0,250	0,250	0,250	0,250	1,0713	774,9846	555,6230	627,0739
58	0,625	0,125	0,125	0,125	1,2961	820,5548	463,7580	657,4579
59	0,125	0,625	0,125	0,125	1,1046	770,8051	535,3172	627,0914
60	0,125	0,125	0,625	0,125	1,0155	771,6492	546,7430	622,4529
61	0,125	0,125	0,125	0,625	1,1631	750,4528	555,4449	610,6630

Fonte: elaboração própria.

Tabela 34. Delineamento de Misturas SLD (4, 5) com as métricas de performance de previsão em relação ao conjunto teste como respostas para a região Sul.

	$x_1$	$x_2$	$x_3$	$x_4$	$MAPE_{ensb}$	$RMSE_{ensb}$	$MdAE_{ensb}$	$MAE_{ensb}$
1	1,000	0,000	0,000	0,000	0,8465	283,7203	144,5969	213,6225
2	0,800	0,200	0,000	0,000	0,8568	291,3587	158,3323	220,2579
3	0,800	0,000	0,200	0,000	0,8670	299,4186	159,0797	226,8933
4	0,800	0,000	0,000	0,200	0,8810	307,8671	156,6627	233,6719
5	0,600	0,400	0,000	0,000	0,9109	316,6729	161,1987	240,9571
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
57	0,250	0,250	0,250	0,250	0,9171	274,1344	151,8069	209,2886
58	0,625	0,125	0,125	0,125	0,8723	278,4939	148,2019	210,4217
59	0,125	0,625	0,125	0,125	0,8920	297,5595	151,1777	226,3939
60	0,125	0,125	0,625	0,125	0,9094	295,2442	159,9572	225,7054
61	0,125	0,125	0,125	0,625	1,1364	245,2664	173,5392	193,2013

Fonte: elaboração própria.

Tabela 35. Delineamento de Misturas SLD (4, 5) com as métricas de performance de previsão em relação ao conjunto teste como respostas para a região Centro-Oeste.

	$x_1$	$x_2$	$x_3$	$x_4$	$MAPE_{ensb}$	$RMSE_{ensb}$	$MdAE_{ensb}$	$MAE_{ensb}$
1	1,000	0,000	0,000	0,000	1,0990	125,1415	70,4683	97,4441
2	0,800	0,200	0,000	0,000	1,1083	124,0866	77,0983	96,4921
3	0,800	0,000	0,200	0,000	1,1181	123,8772	69,2273	95,5493
4	0,800	0,000	0,000	0,200	1,2035	124,5173	75,3688	95,3871
5	0,600	0,400	0,000	0,000	1,3241	125,9942	79,0070	96,0486
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
57	0,250	0,250	0,250	0,250	1,3217	120,1423	74,2870	93,1182
58	0,625	0,125	0,125	0,125	1,1848	121,1084	70,6293	94,3215
59	0,125	0,625	0,125	0,125	1,3487	123,6649	72,8026	94,9097
60	0,125	0,125	0,625	0,125	1,4293	119,9172	74,9128	94,2933
61	0,125	0,125	0,125	0,625	1,5009	119,4774	80,9492	92,4562

Fonte: elaboração própria.

Tabela 36. Delineamento de Misturas SLD (4, 5) com as métricas de performance de previsão em relação ao conjunto teste como respostas para a região o Brasil.

	$x_1$	$x_2$	$x_3$	$x_4$	$MAPE_{ensb}$	$RMSE_{ensb}$	$MdAE_{ensb}$	$MAE_{ensb}$
1	1,000	0,000	0,000	0,000	1,1162	1300,9715	854,4627	1058,7228
2	0,800	0,200	0,000	0,000	1,0452	1245,4202	864,1199	1017,0472
3	0,800	0,000	0,200	0,000	1,1193	1209,4787	796,8261	977,6944
4	0,800	0,000	0,000	0,200	1,2596	1194,9181	791,5797	946,4802
5	0,600	0,400	0,000	0,000	1,4289	1202,5149	839,4415	944,6559
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
57	0,250	0,250	0,250	0,250	1,3133	1248,3895	816,0222	1023,4727
58	0,625	0,125	0,125	0,125	1,2148	1270,7761	867,8329	1041,0978
59	0,125	0,625	0,125	0,125	1,1840	1195,9534	757,8355	948,3529
60	0,125	0,125	0,625	0,125	1,7916	1327,3668	919,9872	1083,3196
61	0,125	0,125	0,125	0,625	1,2486	1268,7753	832,9851	1039,3913

Fonte: elaboração própria.

## N. Verificar a adequabilidade de redução de dimensionalidade

Para tornar o problema multiobjetivo em um problema com um objetivo apenas e assim tornar o problema computacionalmente mais eficiente, pode-se aplicar técnicas de estatística multivariada para reduzir a dimensionalidade das respostas do problema (NAVES et al., 2017).

Como as respostas se encontram em escalas distintas, optou-se por padronizar as variáveis, segundo a Equação 54.

$$x_{i_{pad}} = \frac{x_i - \bar{X}}{s_{x_i}} \quad (54)$$

Para:  $\bar{X}$  = média amostral de  $X$  e  $s_{x_i}$  = desvio-padrão amostral de  $X$ .

A Tabela 37 apresenta a matriz de covariância entre as variáveis (métricas de performance) padronizadas referentes à região Norte do Brasil.

Tabela 37. Matriz de correlação entre as métricas padronizadas oriundas do delineamento de misturas para a região Norte.

	$MAPE_{ensb}$	$RMSE_{ensb}$	$MdAE_{ensb}$	$MAE_{ensb}$
$MAPE_{ensb}$	1,0000			
$RMSE_{ensb}$	-0,2884	1,0000		
$MdAE_{ensb}$	<b>0,6958</b>	-0,0426	1,0000	
$MAE_{ensb}$	<b>-0,3036</b>	<b>0,9927</b>	-0,0912	1,0000

Fonte: elaboração própria.

Há três correlações maiores que 0,30. (em negrito), indicando que pode haver condições para a aplicação de técnicas de redução de dimensionalidade. O teste de Bartlett apresentou p-valor igual a 0,000. No entanto, a estatística KMO apresentou o valor de 0,45, indicando a não adequabilidade de continuar com a técnica de redução de dimensionalidade.

A matriz de correlação da Tabela (37) exhibe a matriz de correlação das métricas de performance referentes à análise da região Nordeste do Brasil.

Tabela 38. Matriz de correlação entre as métricas padronizadas oriundas do delineamento de misturas para a região Nordeste.

	$MAPE_{ensb}$	$RMSE_{ensb}$	$MdAE_{ensb}$	$MAE_{ensb}$
$MAPE_{ensb}$	1,0000			
$RMSE_{ensb}$	-0,2730	1,0000		
$MdAE_{ensb}$	<b>-0,4472</b>	-0,0394	1,0000	
$MAE_{ensb}$	<b>0,7411</b>	<b>0,3835</b>	<b>-0,3424</b>	1,0000

Fonte: elaboração própria.



Das seis correlações apresentadas, quatro são maiores que 0,30, indicando a possibilidade da aplicação de técnicas de redução de dimensionalidade. O p-valor do teste de esfericidade de Bartlett apresentou valor 0,000, dando fortes evidências de que as correlações são significativamente diferentes de 0. Contudo, a estatística KMO foi de apenas 0,26, sugerindo que técnicas como a Análise de Componentes Principais e a Análise Fatorial não podem ser aplicadas no conjunto de dados analisado.

A Tabela 39 mostra a matriz de correlação para as métricas de previsão no contexto do delineamento de misturas para a região Sudeste do Brasil.

Tabela 39. Matriz de correlação entre as métricas padronizadas oriundas do delineamento de misturas para a região Sudeste.

	$MAPE_{ensb}$	$RMSE_{ensb}$	$MdAE_{ensb}$	$MAE_{ensb}$
$MAPE_{ensb}$	1,0000			
$RMSE_{ensb}$	<b>0,6363</b>	1,0000		
$MdAE_{ensb}$	<b>-0,6231</b>	<b>-0,7000</b>	1,0000	
$MAE_{ensb}$	<b>0,6412</b>	<b>0,9921</b>	<b>-0,6607</b>	1,0000

Fonte: elaboração própria.

Todas as correlações são expressivamente maiores que 0,30. O teste de esfericidade de Bartlett apresentou p-valor igual a 0,000. Porém, a estatística KMO calculada para este caso foi de 0,65, sugerindo uma adequabilidade razoável à aplicação de métodos multivariados.

Já a matriz de correlação das métricas relativas à região Sul do Brasil é apresentada na Tabela 40.

Tabela 40. Matriz de correlação entre as métricas padronizadas oriundas do delineamento de misturas para a região Sul.

	$MAPE_{ensb}$	$RMSE_{ensb}$	$MdAE_{ensb}$	$MAE_{ensb}$
$MAPE_{ensb}$	1,0000			
$RMSE_{ensb}$	<b>-0,6692</b>	1,0000		
$MdAE_{ensb}$	<b>0,6790</b>	-0,2056	1,0000	
$MAE_{ensb}$	<b>-0,5405</b>	<b>0,9860</b>	-0,0991	1,0000

Fonte: elaboração própria.

Mais uma vez, quatro medidas de correlação das seis disponíveis apresentam valores acima de 0,30. O teste de esfericidade de *Bartlett* calculado apresentou valor de 0,000. Porém, a estatística KMO foi de 0,42, indicando que também não se pode aplicar técnicas de redução de dimensionalidade neste caso.

A Tabela 41 exibe matriz de correlação das respostas do delineamento de misturas relativa à região Centro-Oeste.

Tabela 41. Matriz de correlação entre as métricas padronizadas oriundas do delineamento de misturas para a região Centro-Oeste.

	$MAPE_{ensb}$	$RMSE_{ensb}$	$MdAE_{ensb}$	$MAE_{ensb}$
$MAPE_{ensb}$	1,0000			
$RMSE_{ensb}$	-0,2254	1,0000		
$MdAE_{ensb}$	<b>0,6140</b>	0,0425	1,0000	
$MAE_{ensb}$	-0,1881	<b>0,8574</b>	-0,0983	1,0000

Fonte: elaboração própria.

Conforme pode ser observado, apenas duas correlações apresentam valor superior a 0,30. O teste de esfericidade de *Bartlett* apresenta p-valor de 0,0000. Entretanto, a estatística KMO mais uma vez mostra uma pobre aderência dos dados à aplicação de técnicas de estatística multivariada para redução de dimensionalidade.

Por fim, a matriz de correlação das métricas de previsão dado o delineamento de misturas do Brasil é apresentada na Tabela 42.

Tabela 42. Matriz de correlação entre as métricas padronizadas oriundas do delineamento de misturas para a região **Centro-Oeste. Brasil.**

	$MAPE_{ensb}$	$RMSE_{ensb}$	$MdAE_{ensb}$	$MAE_{ensb}$
$MAPE_{ensb}$	1,0000			
$RMSE_{ensb}$	<b>0,7386</b>	1,0000		
$MdAE_{ensb}$	<b>0,3983</b>	<b>0,4218</b>	1,0000	
$MAE_{ensb}$	<b>0,6604</b>	<b>0,9787</b>	<b>0,4566</b>	1,0000

Fonte: elaboração própria.

Como pode-se verificar pela Tabela 42 todas as correlações apresentam valor acima de 0,30. O teste de esfericidade de *Bartlett* indica a pertinência da aplicação de técnicas multivariadas, com o p-valor apresentando valor igual a 0,0000. Porém, a estatística KMO apresentou valor 0,57, indicando uma má adequação global à Análise Fatorial ou à Análise de Componentes Principais.

Como as estatísticas apresentam valores discordantes quanto à possibilidade de redução de dimensionalidade, decidiu-se estimar as superfícies de resposta de maneira separada e tratar os problemas de otimização de maneira multiobjetiva.

## O. Ajustar as superfícies de resposta

Tendo criado os delineamentos de mistura e averiguada a inadequação da aplicação de técnicas de redução de dimensionalidade sobre as respostas dos modelos de delineamento de

misturas propostos, estimam-se os coeficientes das superfícies de resposta de cada métrica, tendo como variáveis explanatórias dos modelos os pesos referentes ao delineamento de misturas. Serão utilizados o método de Mínimos Quadrados Ordinários ou algum método de correção de Heterocedasticidade (Mínimos Quadrados Ponderados ou Correção de White).

Para fins de estimação dos modelos, as variáveis padronizadas foram utilizadas. O algoritmo *stepwise* foi escolhido para seleção das variáveis explanatórias que farão parte de cada modelo. Ressalta-se que o alfa selecionado para definir a entrada ou a saída de variáveis foi de 0,15. Este valor é justificado por Hosmer Jr., Lemeshow e Sturdivant (2013), uma vez que o autor argumenta que quando se analisa um teste de hipótese esperando a não rejeição da hipótese nula o p-valor do teste deve ser aumentado.

Segundo (LAWSON, 2014), foi selecionado um modelo quártico completo, que permite ao algoritmo selecionar o melhor de regressão que se ajusta aos dados, seja de quarta ordem ou de ordem inferior. Um modelo quártico completo pode ser representado de forma geral pela Equação 55.

$$\begin{aligned}
 y = & \sum_{i=1}^q \beta_i x_i + \sum_{i<j}^q \gamma_{ij} x_i x_j (x_i - x_j) + \sum_{i<j}^q \delta_{ij} x_i x_j (x_i x_j)^2 \\
 & + \sum_{i<j<k}^q \beta_{iijk} x_i^2 x_j x_k + \sum_{i<j<k}^q \beta_{ijjk} x_i x_j^2 x_k \\
 & + \sum_{i<j<k}^q \beta_{ijkk} x_i x_j x_k^2 + \sum_{i<j<k<l}^q \beta_{ijkl} x_i x_j x_k x_l
 \end{aligned} \tag{55}$$

Considerando  $x_i$  o peso da  $i$ -ésima DMU mais eficiente em ordem decrescente de escore de eficiência, para  $i = (1, 2, 3, 4)$ . A Tabela 43 apresenta os coeficientes da superfície de resposta referente à análise da demanda de energia elétrica no Sudeste do Brasil, sendo o MAPE a variável resposta, enquanto que a

Tabela 44 apresenta a Análise de Variância (ANOVA) do modelo estimado.

Tabela 43. Regressão estimada para o MAPE da região Sudeste.

	<b>Coef.</b>	<b>Erro Pad.</b>	<b>Valor-T</b>	<b>Valor-P</b>	<b>VIF</b>
$x_1$	3,610	0,408	*	*	4,02
$x_2$	-1,005	0,301	*	*	2,20
$x_3$	0,172	0,302	*	*	2,21
$x_4$	0,593	0,280	*	*	1,90
$x_1x_2$	-6,32	1,82	-3,48	0,001	3,46
$x_1x_3$	-7,85	1,63	-4,82	0,000	2,78
$x_1x_4$	-12,83	1,66	-7,71	0,000	2,90
$x_2x_3(x_2 - x_3)$	8,71	3,39	2,57	0,013	1,32
$x_1^2x_2x_3$	137,8	28,2	4,88	0,000	1,84
$x_1x_2^2x_3$	112,8	27,8	4,05	0,000	1,79
$x_1x_2x_3x_4$	-231,1	82,9	-2,79	0,007	1,33

Fonte: elaboração própria.

Tabela 44. Anova para o MAPE da região Sudeste.

	<b>GL</b>	<b>SQ Seq</b>	<b>SQ (Aj.)</b>	<b>QM (Aj.)</b>	<b>Valor F</b>	<b>Valor-P</b>
Regressão	10	43,9386	43,939	4,3939	13,68	0,000
Linear	3	11,9446	30,209	10,0698	31,35	0,000
Quadrático	3	17,5734	23,527	7,8423	24,41	0,000
$x_1x_2$	1	0,0014	3,884	3,8841	12,09	0,001
$x_1x_3$	1	3,1317	7,464	7,4637	23,23	0,000
$x_1x_4$	1	14,4404	19,080	19,0803	59,40	0,000
Cúbico completo	1	3,1422	2,115	2,1150	6,58	0,013
$x_2x_3(x_2 - x_3)$	1	3,1422	2,115	2,1150	6,58	0,013
Quártico Especial	2	8,7796	10,805	5,4026	16,82	0,000
$x_1^2x_2x_3$	1	4,7754	7,657	7,6574	23,84	0,000
$x_1x_2^2x_3$	1	4,0042	5,274	5,2737	16,42	0,000
Quártico2 completo	1	2,4988	2,499	2,4988	7,78	0,007
$x_1x_2x_3x_4$	1	2,4988	2,499	2,4988	7,78	0,007
Erro de Resíduos	50	16,0614	16,061	0,3212		
Total	60	60,0000				

Fonte: elaboração própria.

O p-valor do teste F indica que o modelo como um todo é significativo para explicar a variabilidade do MAPE. Nota-se que o modelo não possui problemas com multicolinearidade, uma vez que os VIFs se encontram abaixo de 5. A Figura 47 apresenta o diagnóstico gráfico do modelo de regressão proposto.

Observa-se que o modelo não apresenta problemas quanto aos pressupostos de Gauss-Markov. Os resíduos se ajustam bem à curva normal como mostrado pelo Q-Q Plot. O teste de normalidade de Anderson-Darling confirma essa análise, apontando um p-valor de 0,6095.

O gráfico dos valores ajustados contra o resíduo não mostra uma tendência ou a presença de heterocedasticidade. O histograma confirma uma distribuição unimodal e simétrica. A média dos resíduos possui o valor 0, indicando que não há problemas de especificação do modelo.

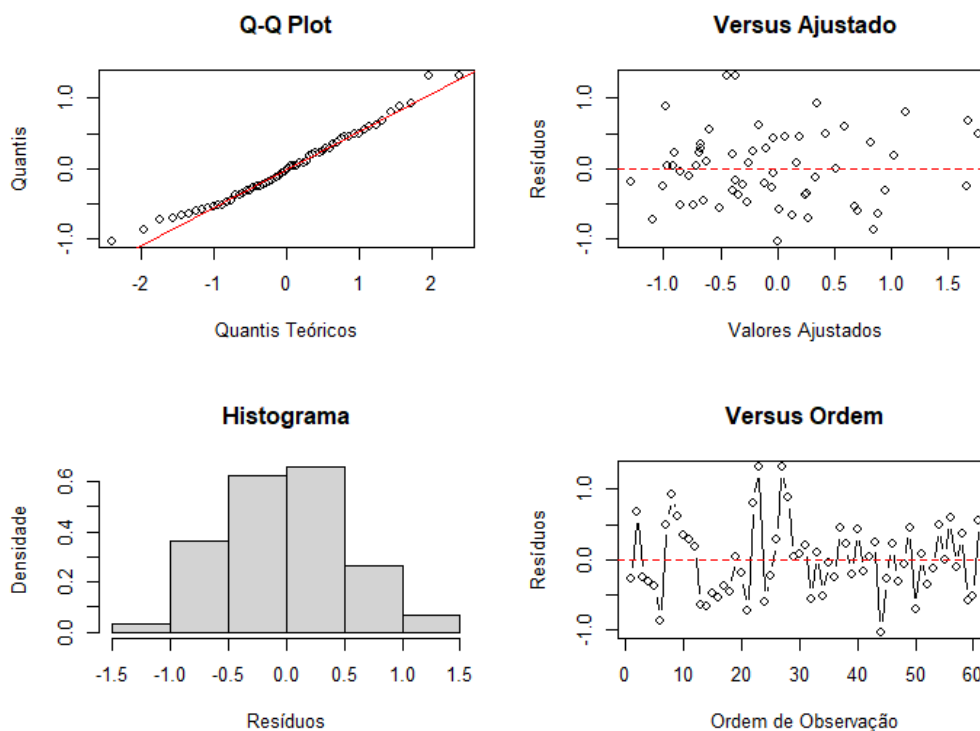


Figura 47. Análise de resíduos da regressão de superfície de resposta tendo o MAPE como variável resposta. Fonte: elaboração própria.

Enquanto a Tabela 45 apresenta a regressão estimada tendo o RMSE como variável resposta sobre o delineamento de mistura da região Sudeste, a Tabela 46 exibe a ANOVA do modelo.

Tabela 45. Regressão estimada para o RMSE da região Sudeste.

	Coef.	Erro Pad.	Valor-T	Valor-P	VIF
$x_1$	3,020	0,320	*	*	3,32
$x_2$	-0,149	0,271	*	*	2,37
$x_3$	-1,170	0,249	*	*	2,01
$x_4$	-1,116	0,248	*	*	2,00
$x_1x_2$	-5,10	1,56	-3,27	0,002	3,40
$x_1x_4$	-4,35	1,46	-2,99	0,004	2,97
$x_1x_3(x_1 - x_3)$	-6,11	2,99	-2,04	0,047	1,37
$x_1^2x_2x_3$	-74,8	31,5	-2,37	0,022	3,07
$x_1^2x_2x_4$	106,0	24,4	4,34	0,000	1,84
$x_1x_2^2x_3$	139,1	31,2	4,45	0,000	3,01
$x_2x_3(x_2 - x_3)^2$	10,26	6,52	1,57	0,122	1,61
$x_1x_2x_3x_4$	-107,3	71,6	-1,50	0,141	1,33

Fonte: elaboração própria.

Tabela 46. Anova para o RMSE da região Sudeste.

	GL	SQ Seq	SQ (Aj.)	QM (Aj.)	Valor F	Valor-P
Regressão	11	48,2278	48,2278	4,3843	18,25	0,000
Linear	3	37,5213	33,4140	11,1380	46,36	0,000
Quadrático	2	0,5088	3,7928	1,8964	7,89	0,001
$x_1x_2$	1	0,2848	2,5727	2,5727	10,71	0,002
$x_1x_4$	1	0,2241	2,1416	2,1416	8,91	0,004
Cúbico completo	1	1,2981	1,0014	1,0014	4,17	0,047
$x_1x_3(x_1 - x_3)$	1	1,2981	1,0014	1,0014	4,17	0,047
Quártico Especial	3	7,7112	8,0248	2,6749	11,13	0,000
$x_1^2x_2x_3$	1	0,0186	1,3529	1,3529	5,63	0,022
$x_1^2x_2x_4$	1	2,9256	4,5290	4,5290	18,85	0,000
$x_1x_2^2x_3$	1	4,7670	4,7613	4,7613	19,82	0,000
Quártico1 completo	1	0,6496	0,5950	0,5950	2,48	0,122
$x_2x_3(x_2 - x_3)^2$	1	0,6496	0,5950	0,5950	2,48	0,122
Quártico2 completo	1	0,5388	0,5388	0,5388	2,24	0,141
$x_1x_2x_3x_4$	1	0,5388	0,5388	0,5388	2,24	0,141
Erro de Resíduos	49	11,7722	11,7722	0,2402		
Total	60	60,0000				

Fonte: elaboração própria.

O modelo como um todo é significativo a 5% de nível de significância, visto que o p-valor da regressão é de 0,000. A Figura 48 apresenta o diagnóstico gráfico para este modelo.

A superfície de resposta estimada não apresenta problemas quanto aos pressupostos de Gauss-Markov. Nota-se que a distribuição dos resíduos se ajusta bem à distribuição teórica normal, como mostra o Q-Q Plot. O gráfico de dispersão entre os valores ajustados e os resíduos não apresenta tendência e nem heterocedasticidade.

O histograma confirma a análise do Q-Q Plot, mostrando uma distribuição unimodal e relativamente simétrica. Os testes de Anderson-Darling e Shapiro-Wilk confirmam a análise. O primeiro apresenta um valor de 0,5953 e o segundo um valor de 0,3852, dando fortes indícios de que não se pode rejeitar a hipótese nula que os resíduos tiveram como uma função geradora uma distribuição normal com média 0.

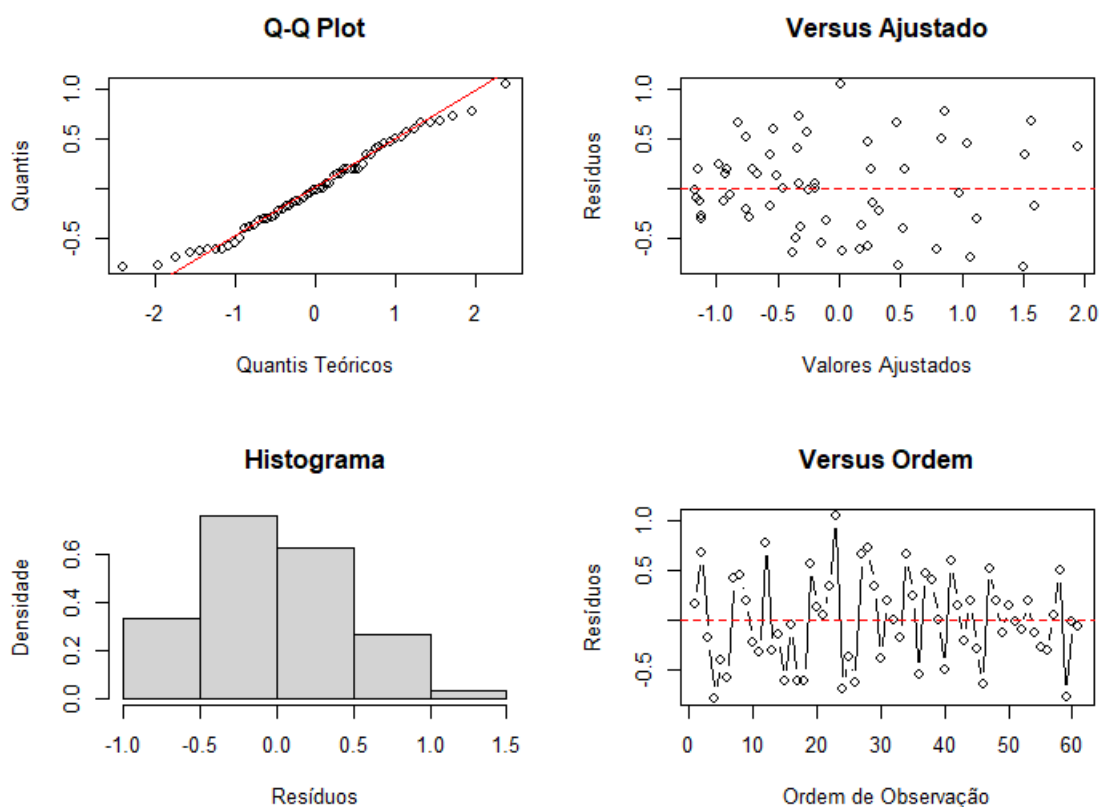


Figura 48. Análise de resíduos da regressão de superfície de resposta tendo o RMSE como variável resposta.  
Fonte: elaboração própria.

Considerando a métrica MdAE como variável resposta, a Tabela 46 e a Tabela 47 apresentam o modelo de regressão e a ANOVA, respectivamente. O modelo não apresenta problema de multicolinearidade avaliado pelos VIFs.

Tabela 47. Regressão estimada para o MdAE da região Sudeste.

	Coef.	Erro. Pad.	Valor-T	Valor-P	VIF
$x_1$	-0,715	0,442	*	*	2,36
$x_2$	0,711	0,370	*	*	1,65
$x_3$	-0,289	0,405	*	*	1,97
$x_4$	-0,323	0,485	*	*	2,83
$x_1x_4$	4,74	2,31	2,05	0,045	2,78
$x_3x_4$	4,59	2,18	2,11	0,040	2,47
$x_1^2x_2x_4$	-153,0	49,9	-3,06	0,003	2,86
$x_1x_2x_4^2$	79,8	50,1	1,60	0,117	2,88
$x_1x_2(x_1 - x_2)^2$	-28,8	10,6	-2,70	0,009	1,60

Fonte: elaboração própria.

Tabela 48. Anova para o MdAE da região Sudeste.

	GL	SQ Seq	SQ (Aj.)	QM (Aj.)	Valor F	Valor-P
Regressão	8	26,433	26,433	3,3042	5,12	0,000
Linear	3	10,577	4,295	1,4316	2,22	0,097
Quadrático	2	3,827	5,119	2,5597	3,97	0,025
$x_1x_4$	1	1,952	2,721	2,7207	4,21	0,045
$x_3x_4$	1	1,875	2,868	2,8680	4,44	0,040
Quártico Especial	2	7,308	6,438	3,2188	4,99	0,010
$x_1^2x_2x_4$	1	5,309	6,064	6,0638	9,39	0,003
$x_1x_2x_4^2$	1	1,999	1,643	1,6425	2,54	0,117
Quártico1 completo	1	4,722	4,722	4,7222	7,32	0,009
$x_1x_2(x_1 - x_2)^2$	1	4,722	4,722	4,7222	7,32	0,009
Erro de Resíduos	52	33,567	33,567	0,6455		
Total	60	60,000				

Fonte: elaboração própria.

Tanto o modelo, quanto a parte linear e quadrática são significativas para modelar a métrica no contexto de delineamento de misturas. A Figura 49 apresenta o diagnóstico da regressão.

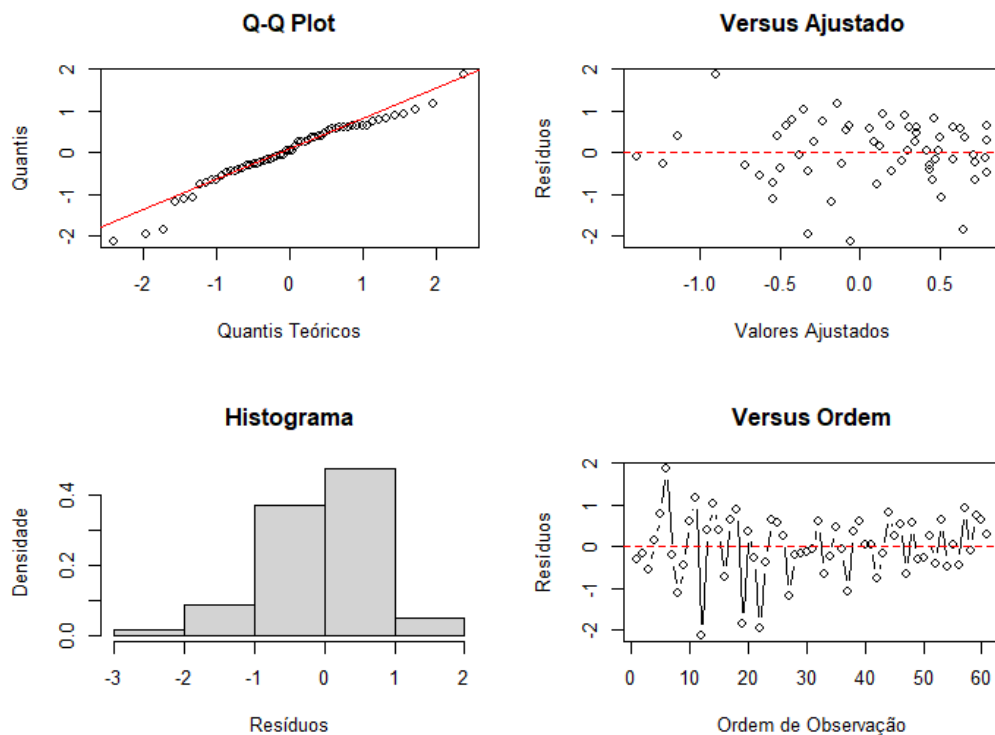


Figura 49. Análise de resíduos da regressão de superfície de resposta tendo o MdAE como variável resposta.  
Fonte: elaboração própria.

Analisando-se o Q-Q Plot e o Histograma da Figura 49, os resíduos vêm de uma distribuição normal, o que é confirmado pelo teste de Anderson-Darling, que apresentou um valor de 0,0574 e o teste de Shapiro-Wilk apresentou um p-valor de 0,0512. Já o gráfico de resíduos *versus* o valor ajustado não apresenta indícios de heterocedasticidade e não possui tendência.



Por fim, a Tabela 49 e a Tabela 50 exibem a estimativa da superfície de resposta tendo a métrica MAE como resposta no contexto do delineamento de misturas e sua respectiva ANOVA.

Tabela 49. Regressão estimada para o MAE da região Sudeste.

	<b>Coef.</b>	<b>Erro Pad.</b>	<b>Valor-T</b>	<b>Valor-P</b>	<b>VIF</b>
$x_1$	3,020	0,294	*	*	3,31
$x_2$	-0,172	0,249	*	*	2,37
$x_3$	-1,199	0,228	*	*	1,99
$x_4$	-1,225	0,228	*	*	1,99
$x_1x_2$	-4,57	1,43	-3,19	0,002	3,40
$x_1x_4$	-4,06	1,34	-3,04	0,004	2,95
$x_1x_4(x_1 - x_4)$	-5,00	2,75	-1,81	0,076	1,37
$x_1^2x_2x_3$	-75,3	28,9	-2,61	0,012	3,04
$x_1^2x_2x_4$	91,4	21,9	4,17	0,000	1,75
$x_1x_2^2x_3$	131,3	28,5	4,60	0,000	2,97
$x_2x_3(x_2x_3)^2$	10,29	5,99	1,72	0,092	1,61

Fonte: elaboração própria.

Tabela 50. Anova para o MdAE da região Sudeste.

	<b>GL</b>	<b>SQ Seq</b>	<b>SQ (Aj.)</b>	<b>QM (Aj.)</b>	<b>Valor F</b>	<b>Valor-P</b>
Regressão	10	49,8274	49,8274	4,9827	24,49	0,000
Linear	3	40,8593	35,1551	11,7184	57,60	0,000
Quadrático	2	0,3584	3,1667	1,5833	7,78	0,001
$x_1x_2$	1	0,1961	2,0678	2,0678	10,16	0,002
$x_1x_4$	1	0,1623	1,8831	1,8831	9,26	0,004
Cúbico completo	1	0,8700	0,6699	0,6699	3,29	0,076
$x_1x_4(x_1 - x_4)$	1	0,8700	0,6699	0,6699	3,29	0,076
Quártico Especial	3	7,1392	6,9291	2,3097	11,35	0,000
$x_1^2x_2x_3$	1	0,0441	1,3834	1,3834	6,80	0,012
$x_1^2x_2x_4$	1	2,4913	3,5351	3,5351	17,38	0,000
$x_1x_2^2x_3$	1	4,6038	4,3028	4,3028	21,15	0,000
Quártico1 completo	1	0,6006	0,6006	0,6006	2,95	0,092
$x_2x_3(x_2x_3)^2$	1	0,6006	0,6006	0,6006	2,95	0,092
Erro de Resíduos	50	10,1726	10,1726	0,2035		
Total	60	60,0000				

Fonte: elaboração própria.

O modelo é significativo como um todo, conforme indicado pela Estatística F do modelo. A parte linear e quadrática também são significativas a 5% de nível de significância. A Figura 50 mostra a análise de resíduos da regressão.

O Q-Q Plot e o Histograma mostram que os resíduos não se distanciam de uma distribuição teórica normal, o que pode ser confirmado pelos testes de normalidade de Anderson-Darling e Shapiro-Willk, que apresentam os p-valores de 0,6929 e 0,5498, respectivamente.

O gráfico de valor ajustado *versus* resíduos mostra aleatoriedade, com os valores se distribuindo em torno de zero, não apresentando tendência.

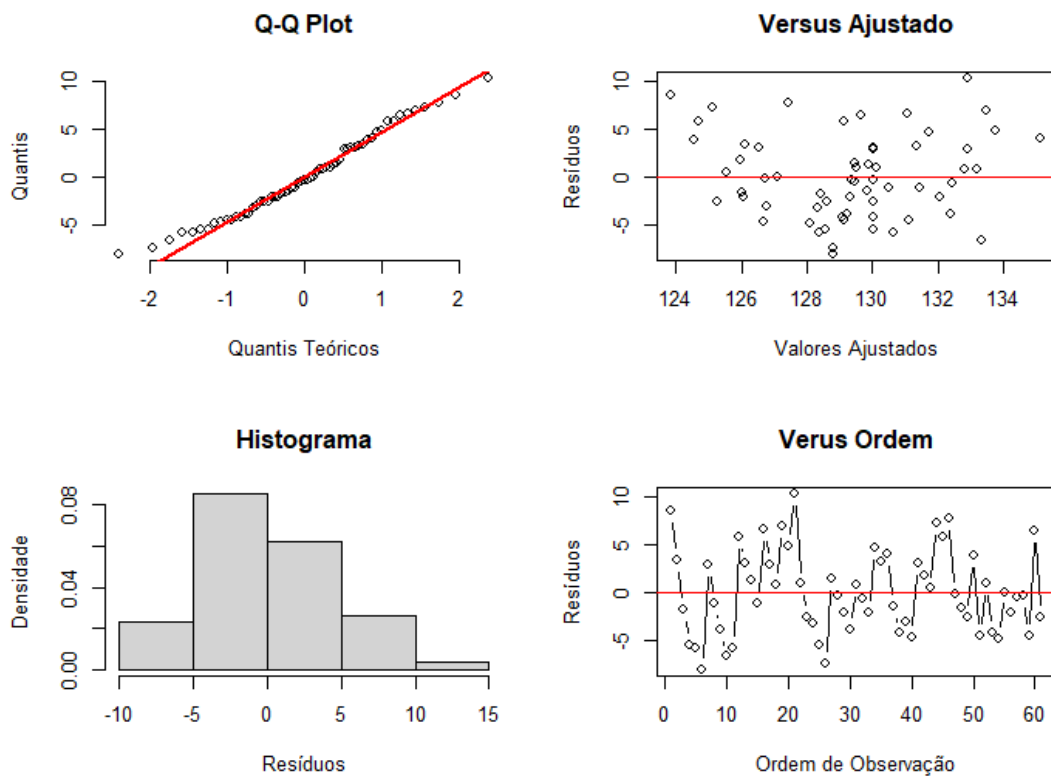


Figura 50. Análise de resíduos da regressão de superfície de resposta tendo o MAE como variável resposta. Fonte: elaboração própria.

As tabelas dos coeficientes estimados, seus respectivos p-valores, ANOVA e diagnóstico das regressões encontram-se no Apêndice A.

## P. Otimização das Superfícies de Resposta

Dado que as superfícies de resposta possuem pertinência quanto aos pressupostos de Gauss-Markov, parte-se para a otimização multiobjetiva das mesmas. Conforme exposto na seção 2.10 optou-se pelo algoritmo denominado *Desirability*.

Os valores ótimos encontrados aplicando-se o algoritmo são apresentados na Tabela 51.

Tabela 51. Pesos otimizados via algoritmo de Desirability.

Região	$x_1$	$x_2$	$x_3$	$x_4$
Norte	0,1010	0,3176	0,3466	0,2349
Nordeste	0,0000	0,1010	0,8990	0,0000
Sudeste	0,1348	0,6465	0,0000	0,2188
Sul	0,9798	0,0023	0,0164	0,0011
Centro-Oeste	0,0000	0,6949	0,2378	0,0674
Brasil	0,0000	0,4949	0,5051	0,0000

$x_i = DMU_{eff_i}$ , para  $i = 1, 2, 3, 4$ .

Fonte: elaboração própria.

Esses pesos são utilizados em cada uma das previsões das DMUs eficientes conforme a Tabela 30.

Abaixo são apresentados os gráficos de Traço de Resposta de Cox para as respostas MAPE, RMSE, MdAE e MAE do modelo referente ao Norte Sudeste do Brasil, sendo  $A = x_1$ ,  $B = x_2$ ,  $C = x_3$  e  $D = x_4$ .

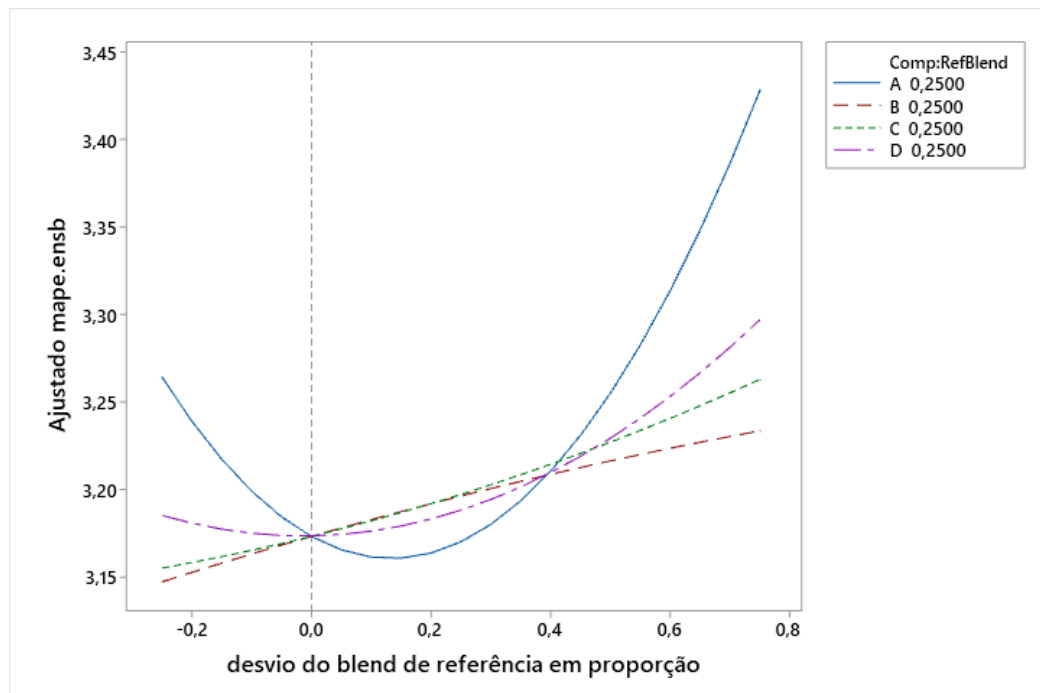


Figura 51. Gráfico de Traço de Cox para a resposta MAPE da região norte sudeste do Brasil.

Fonte: elaboração própria.

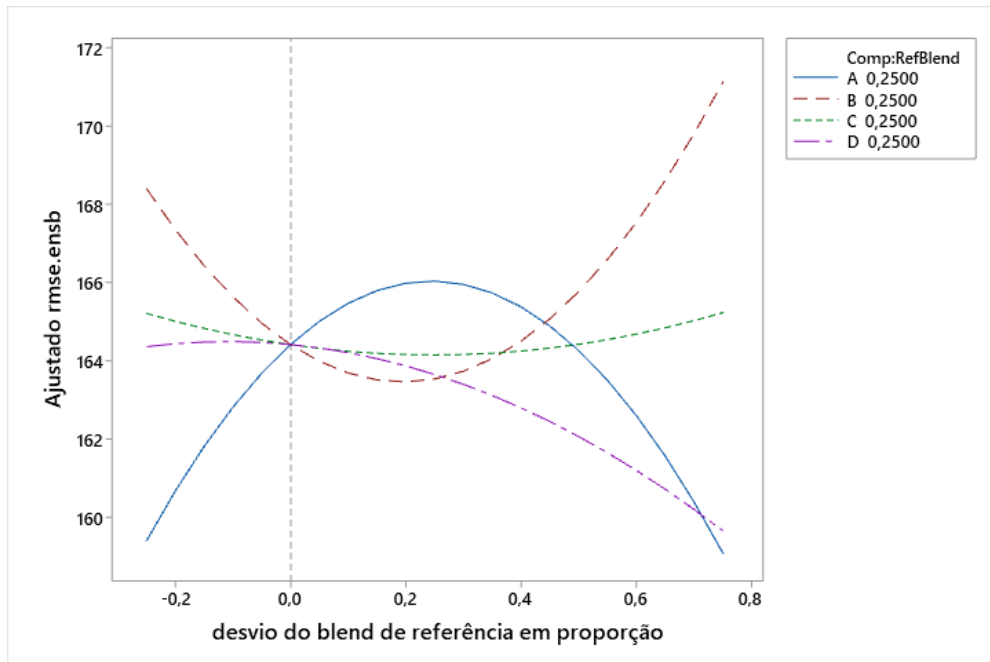


Figura 52. Gráfico de Traço de Cox para a resposta RMSE da região norte sudeste do Brasil.  
Fonte: elaboração própria.

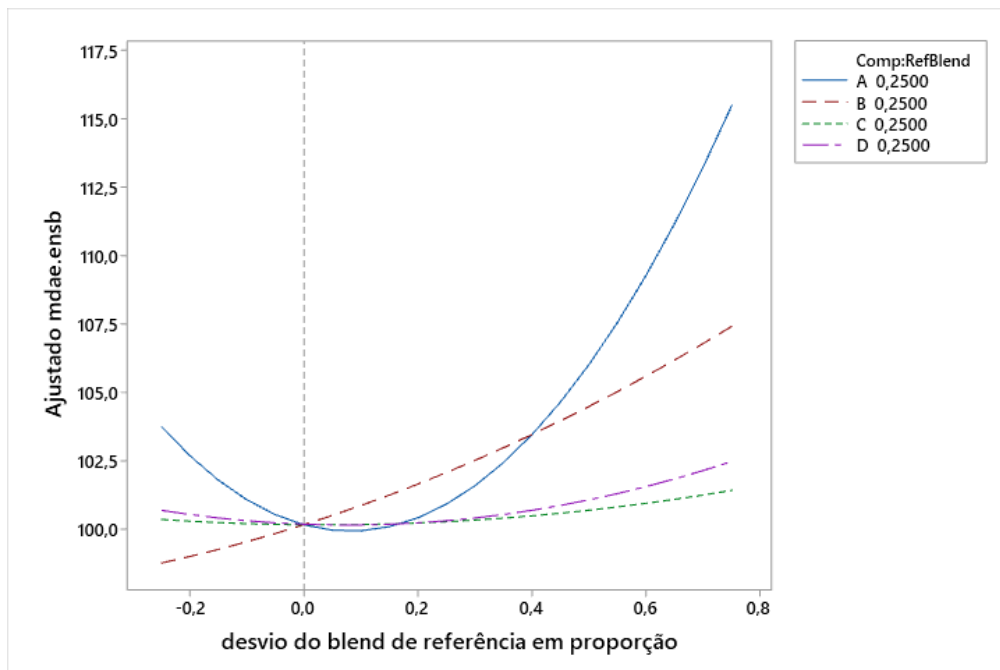


Figura 53. Gráfico de Traço de Cox para a resposta MdAE da região norte sudeste do Brasil.  
Fonte: elaboração própria.

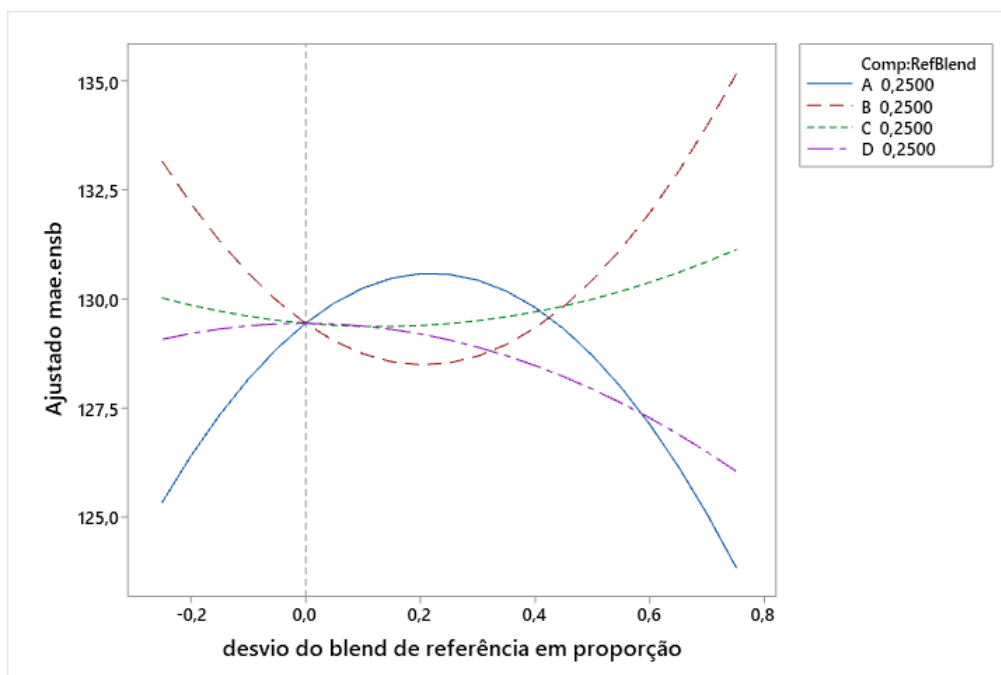


Figura 54. Gráfico de Traço de Cox para a resposta MAE da região norte sudeste do Brasil.  
 Fonte: elaboração própria.

O gráfico de traço de Cox mostra como a mudança das proporções entre os modelos eficientes modifica a resposta. Quanto mais inclinada é a curva, maior o efeito que a DMU em questão possui em relação à resposta analisada. Por exemplo, em relação ao MAPE, aumentar a proporção da DMU mais eficiente faz com que a resposta diminua até determinado ponto e aumente depois. A DMU2 e 3 possuem o mesmo efeito, enquanto a DMU4 possui um efeito quadrático também. Os gráficos de traço de Cox referentes à análise das regiões Nordeste, Sudeste, Sul, Centro-Oeste e Brasil encontram-se no Apêndice A.

#### Q. Aplicar os pesos otimizados e comparar com outro método

Conforme exposto, o método apresentado neste trabalho será comparado com o método *ingênuo* (Naive) de atribuição de pesos para combinação de modelos. Este fato se justifica uma vez que, como discutido na seção 2, o método de dar pesos iguais aos modelos candidatos, além de ser simples e o mais utilizado na prática, é um método que possui uma boa performance (JAGANATHAN; PRAKASH, 2020) dentro das séries temporais da M-4 *Competition*.

As métricas de previsão para o método proposto neste trabalho e o método Naive são apresentadas nas Tabela 52.

Tabela 52. Comparação do método proposto SARIMA-DEA-R e o método Naive para previsão de demanda de energia elétrica entre o Brasil e seus 5 regiões.

REGIÃO	MÉTRICAS	SARIMA-DEA-R	Naive
Norte	MAPE	3,3536	3,2144*
	RMSE	158,1449*	163,522
	MdAE	106,4843	103,7553*
	MAE	124,2826*	128,9664
Nordeste	MAPE	11,1187	9,0660*
	RMSE	311,3229*	314,5493
	MdAE	229,0335	224,9996*
	MAE	266,1205	264,4835*
Sudeste	MAPE	0,9917*	1,0713
	RMSE	766,5302*	774,9846
	MdAE	487,3901*	555,6232
	MAE	616,9574*	627,0743
Sul	MAPE	0,8886*	0,9171
	RMSE	312,3427	274,1344*
	MdAE	155,2597	151,8069*
	MAE	235,7800	209,2886*
Centro-Oeste	MAPE	1,4020	1,3217*
	RMSE	120,1423*	121,0649
	MdAE	72,6328*	74,2870
	MAE	93,11825*	94,4469
Brasil	MAPE	1,0614*	1,3133
	RMSE	1215,0081*	1248,3895
	MdAE	804,9943*	816,0222
	MAE	985,6415*	1023,4727

Fonte: elaboração própria.

Entretanto, o método proposto teve melhor performance de previsão em 15 das 24 comparações analisadas. O método Naive apresentou melhores resultados em relação ao MAPE e ao MdAE quando comparadas as demandas de energia elétrica da região Norte; ao MAPE, MdAE e MAE da região Nordeste, ao RMSE, MdAE e MAE na região Sul e ao MAPE na região Centro-Oeste, confirmando empiricamente o estudo de simulação realizado.

Para exemplificar, a Tabela 53 apresenta as previsões dos últimos 15 elementos em relação ao conjunto treino prevendo o conjunto teste da região Sudeste, utilizando a média ponderada dada pelos pesos otimizados pelo delineamento de misturas. Esta previsão é ilustrada pela Figura 55 que exhibe a previsão pontual e os intervalos de confiança de 80% e 95%, além do ajuste do modelo proposto. Todas as tabelas de previsão podem ser encontradas no Apêndice A.

Tabela 53. Quinze últimos elementos da previsão do conjunto teste e intervalos de confiança para a demanda de energia elétrica no Sudeste do Brasil.

Previsão	80% <sub>inf</sub>	80% <sub>sup</sub>	95% <sub>inf</sub>	95% <sub>sup</sub>
20.171,759	13.558,055	15.847,291	24.496,228	26.785,463
20.244,921	13.409,393	15.775,409	24.714,432	27.080,449
20.351,919	13.301,472	15.741,880	24.961,959	27.402,367
20.593,790	13.334,877	15.847,442	25.340,137	27.852,702
21.087,104	13.625,613	16.208,297	25.965,911	28.548,595
21.146,284	13.487,663	16.138,581	26.153,988	28.804,906
21.086,813	13.007,295	15.803,900	26.369,726	29.166,331
21.400,843	12.988,968	15.900,614	26.901,073	29.812,719
21.048,732	12.326,192	15.345,370	26.752,095	29.771,273
22.165,577	13.152,079	16.271,968	28.059,186	31.179,075
22.053,737	12.751,001	15.971,005	28.136,469	31.356,473
21.054,493	11.471,953	14.788,807	27.320,179	30.637,033
20.508,602	10.654,397	14.065,284	26.951,920	30.362,806
20.586,971	10.469,061	13.971,225	27.202,717	30.704,881
20.694,735	10.319,750	13.910,897	27.478,573	31.069,720

Fonte: elaboração própria.

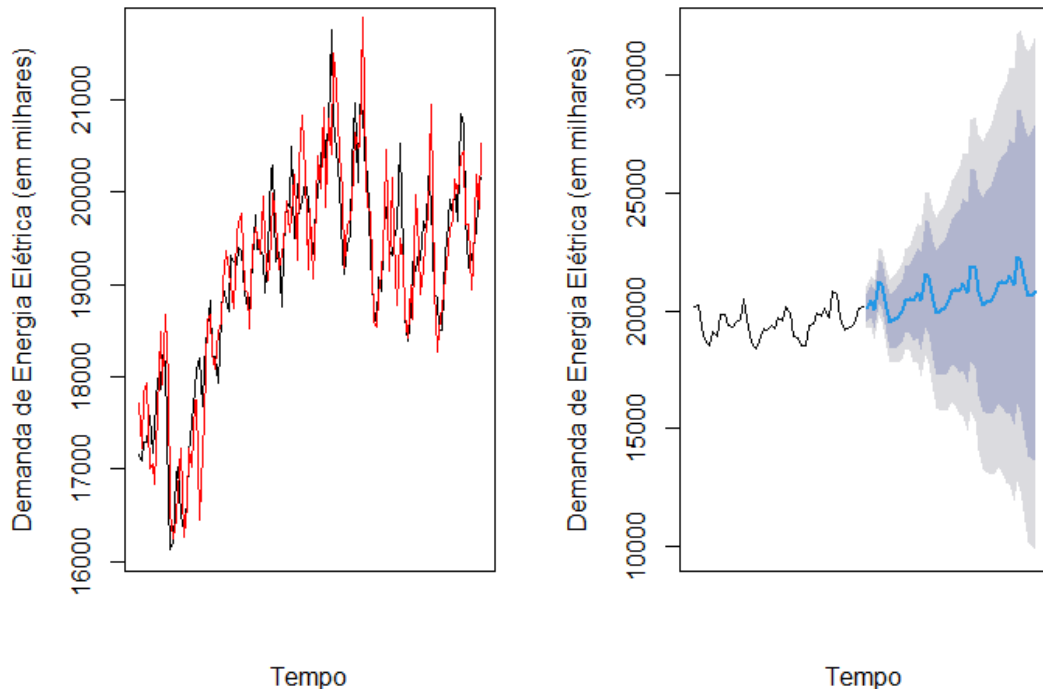


Figura 55. Ajuste do conjunto treino e previsão do conjunto teste para a região Sudeste com o método proposto.  
Fonte: elaboração própria.

## R. Realizar a previsão

O método proposto apresentou melhor performance se comparado com um dos métodos de combinação de modelos mais utilizados. Assim, realizam-se as previsões com o modelo SARIMA-DEA-R. A *Tabela 54* apresenta a previsão e os intervalos de confiança para as 15 últimas previsões da demanda de energia elétrica na região Sudeste do Brasil.

Tabela 54. Quinze últimos elementos da previsão da demanda de energia elétrica no Sudeste do Brasil

<b>Previsão</b>	<b>80%<sub>inf</sub></b>	<b>80%<sub>sup</sub></b>	<b>95%<sub>inf</sub></b>	<b>95%<sub>sup</sub></b>
22.951,449	21.962,051	23.940,847	21.438,797	24.464,101
22.922,320	21.932,922	23.911,718	21.409,668	24.434,972
22.001,626	21.012,228	22.991,024	20.488,974	23.514,278
21.012,700	20.023,302	22.002,098	19.500,048	22.525,352
21.184,407	20.195,009	22.173,805	19.671,755	22.697,059
21.612,190	20.622,792	22.601,588	20.099,538	23.124,842
22.397,318	21.407,920	23.386,716	20.884,666	23.909,970
22.518,941	21.529,543	23.508,339	21.006,289	24.031,593
21.993,593	21.004,195	22.982,991	20.480,941	23.506,245
22.467,236	21.477,838	23.456,634	20.954,584	23.979,888
22.311,449	21.322,051	23.300,847	20.798,797	23.824,101
22.202,066	21.212,668	23.191,464	20.689,414	23.714,718
23.385,596	22.396,198	24.374,994	21.872,944	24.898,248
23.372,906	22.383,508	24.362,304	21.860,254	24.885,558
22.453,693	21.464,295	23.443,091	20.941,041	23.966,345

Fonte: elaboração própria.

A Figura 56 apresenta o gráfico da previsão da demanda de energia elétrica para o Sudeste do Brasil no período considerado.



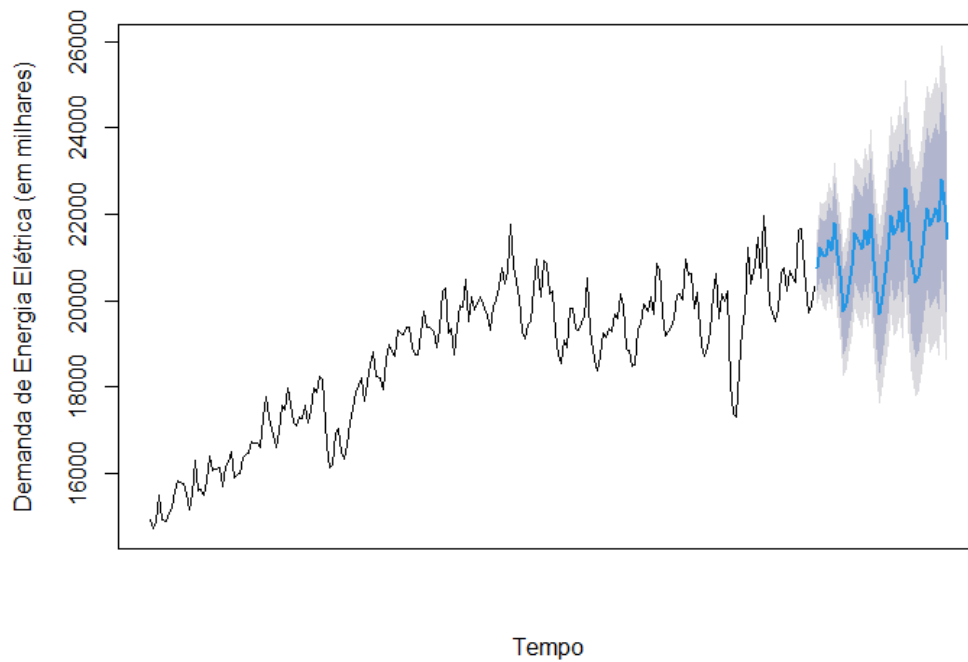


Figura 56. Gráfico da previsão de demanda de energia elétrica do Sudeste do Brasil utilizando-se o método proposto por este trabalho.  
Fonte: elaboração própria.

Considerando o mês de setembro, o método proposto nesta tese apresentou uma previsão de demanda de 21.089,06 milhares de MWh, contra 21.276 milhares de MWh pelo método de *Ensemble Naive*. Segundo dados da ANEEL, o valor real foi de 20.156,56 de milhares de MWh.

## 5. CONCLUSÃO

O presente estudo propõe um novo método de combinação de modelos de séries temporais. O método consiste em criar diversos modelos de Box & Jenkins, modificando os parâmetros computacionalmente seguindo um Arranjo Fatorial Completo. Modela-se a série temporal em análise para cada uma das linhas experimentais e calculam-se métricas de previsão.

Antes do *Ensemble* propriamente dito, este trabalho propõe uma fase intermediária para selecionar preliminarmente quais modelos serão considerados para a combinação. Por meio da utilização de um modelo de Análise por Envoltória de Dados de Superficiência adequado à presença de variáveis fracionadas, selecionam-se aqueles modelos que apresentam um bom desempenho preditivo, considerando o número de parâmetros estimados e as métricas de performance de ajuste como *inputs*.

Propõe-se a retenção de 4 modelos na fase anterior cujos pesos serão dados por meio de otimização de modelos de superfícies de resposta gerados através de Delineamentos Experimentais de Misturas. O arranjo de misturas selecionado foi o Simplex Lattice Design.

As superfícies de resposta foram estimadas por meio de Mínimos Quadrados Ordinários. A seleção das variáveis explanatórias e suas interações foi realizada por meio do algoritmo *Stepwise* sobre um modelo quártico completo inicial. Como algoritmo de otimização, optou-se pelo *Desirability*.

Com a intenção de avaliar o desempenho geral do método, foi realizado um estudo de Simulação, comparando o método proposto por este trabalho com o método Naive, que consiste em dar pesos iguais aos modelos sendo considerados para combinação. O estudo mostrou que o método possui melhor desempenho médio pontual do que o método Naive em todas as métricas consideradas.

Objetivando a diminuição da dimensionalidade das respostas e consequente economia de esforço computacional, analisou-se a adequabilidade de se aplicar ou Análise de Componentes Principais ou Análise Fatorial. Contudo, apesar da análise de correlação e teste de *Bartlett* indicarem a adequação dos mesmos, a estatística KMO não apresentou resultados satisfatórios em nenhum dos casos analisados.

Por fim, aplicou-se o método em seis séries de demanda de energia elétrica correspondentes às cinco regiões do Brasil (Norte, Nordeste, Sudeste, Sul, Centro-Oeste) e ao Brasil. O método proposto apresentou melhor desempenho em 15 das 24 métricas analisadas no estudo, mostrando ser uma alternativa para gerar pesos para combinar séries de forma a evitar o *overfitting*.

Sugere-se como continuidade à esta pesquisa a aplicação e comparação do método proposto com a aplicação de outros métodos de otimização, como o *Normal Boundary Intersection* (NBI), métodos lexográficos e outros algoritmos de otimização. Sugere-se também a aplicação do mesmo método com outros tipos de modelos de séries temporais, como Redes Neurais Artificiais, com o intuito de expandir a aplicabilidade deste trabalho, além de outras situações empíricas.

## 6. REFERÊNCIAS

ANDERSEN, P.; PETERSEN, N. C. A Procedure for Ranking Efficient Units in Data Envelopment Analysis. **Management Science**, v. 39, n. 10, p. 1261–1264, 1993.

BARAK, S.; SADEGH, S. S. Forecasting energy consumption using ensemble ARIMA-ANFIS hybrid algorithm. **International Journal of Electrical Power and Energy Systems**, v. 82, p. 92–104, 2016.

BOGETOFT, P.; OTTO, L. **Benchmarking with DEA, SFA, and R**. [s.l.: s.n.]. v. 157

BOKDE, N. D.; YASEEN, Z. M.; ANDERSEN, G. B. ForecastTB-an R package as a test-bench for time series forecasting-application of wind speed and solar radiation modeling. **Energies**, v. 13, n. 10, 2020.

CHARNES, A.; COOPER, W. W.; RHODES, E. A Data Envelopment Analysis Approach to Evaluation of the Program Follow through Experiment in U.S. Public School Education. **Management Science**, v. 27, n. 6, p. 668–697, 1978.

DE GOOIJER, J. G.; HYNDMAN, R. J. 25 Years of Time Series Forecasting. **International Journal of Forecasting**, v. 22, n. 3, p. 443–473, 2006.

DE OLIVEIRA, F. A. et al. Portfolio optimization using Mixture Design of Experiments: Scheduling trades within electricity markets. **Energy Economics**, v. 33, n. 1, p. 24–32, 2011.

DEPRINS, D.; SIMAR, L.; TULKENS, H. Measuring Labor-Efficiency in Post Offices BT - Public goods, environmental externalities and fiscal competition. In: CHANDER, P. et al. (Eds.). Boston, MA: Springer US, 2006. p. 285–309.

DERRINGER, G.; SUICH, R. Simultaneous Optimization of Several Response Variables. **Journal of Quality Technology**, v. 12, n. 4, p. 214–219, 1980.

DESPIC, O.; DESPIC, M.; PARADI, J. C. DEA-R: Ratio-based comparative efficiency model, its mathematical relation to DEA and its use in applications. **Journal of Productivity Analysis**, v. 28, n. 1–2, p. 33–44, 2007.

DOS SANTOS LEAL, G. et al. A multiobjective portfolio optimization for energy assets using D-Optimal design and mixture design of experiments. **Production**, v. 32, 2022.

EMROUZNEJAD, A.; AMIN, G. R. DEA models for ratio data: Convexity consideration. **Applied Mathematical Modelling**, v. 33, n. 1, p. 486–498, jan. 2009.

EMROUZNEJAD, A.; ROSTAMI-TABAR, B.; PETRIDIS, K. A novel ranking procedure for forecasting approaches using Data Envelopment Analysis. **Technological Forecasting and Social Change**, v. 111, p. 235–243, out. 2016.

FAN, C.; XIAO, F.; WANG, S. Development of prediction models for next-day building energy consumption and peak power demand using data mining techniques. **Applied Energy**, v. 127, p. 1–10, 2014.

FATTAH, J. et al. Forecasting of demand using ARIMA model. **International Journal of Engineering Business Management**, v. 10, p. 1–9, 2018.

FISHBURN, P. C. Lexicographic Orders, Utilities and Decision Rules: A Survey. **Management Science**, v. 20, n. 11, 1974.

GASS, S.; SAATY, T. The computational algorithm for the parametric objective function. **Naval Research Logistics**, v. 2, n. 1, p. 39–45, 1955.

GHALEHKHONDABI, I. et al. An overview of energy demand forecasting methods published in 2005–2015. **Energy Systems**, v. 8, n. 2, p. 411–447, 2017.

GHOSH, S. Univariate time-series forecasting of monthly peak demand of electricity in northern India. **International Journal of Indian Culture and Business Management**, v. 1, n. 4, p. 466, 2008.

GOMES, J. H. F. et al. Global Criterion Method Based on Principal Components to the Optimization of Manufacturing Processes with Multiple Responses. **Journal of Mechanical Engineering**, v. 58, n. 5, 2012.

GUJARATI, D. N. **Econometria básica**. 5. ed. Porto Alegre: [s.n.].

HAMMERSLEY, J. M.; HANDSCOMB, D. C. **Monte Carlo Methods**. [s.l.] Springer US, 1964.

HOCKING, R. R.; LESLIE, R. N. Selection of the Best Subset in Regression Analysis. **Technometrics**, v. 9, n. 4, p. 531–540, 1967.

HONG, W. C. et al. SVR with hybrid chaotic immune algorithm for seasonal load demand forecasting. **Energies**, v. 4, n. 6, p. 960–977, 2011.

HOSMER JR., D. W.; LEMESHOW, S.; STURDIVANT, R. X. **Applied Logistic Regression**. [s.l.] Wiley, 2013.

HYNDMAN, R. J. et al. Time Series Forecasting of Price of Agricultural Products Using Hybrid Methods. **International Journal of Forecasting**, v. 36, n. 1, p. 1–19, 2020.

HYNDMAN, R. J.; ATHANASOPOULOS, G. **Forecasting: principles and practice**. [s.l.] OTexts, 2018.

JAGANATHAN, S.; PRAKASH, P. K. S. A combination-based forecasting method for the M4-competition. **International Journal of Forecasting**, v. 36, n. 1, p. 98–104, 2020.

JAMES, G. et al. **An introduction to statistical learning**. New York: Springer New York, 2013.

JOHNSON, R. A.; WICHERN, D. W. **Applied multivariate statistical analysis**. Sixth ed. Upper Saddle River, New Jersey: Pearson Education, Inc., 2007.

KOLASSA, S. Why the “best” point forecast depends on the error or accuracy measure. **International Journal of Forecasting**, v. 36, n. 1, p. 208–211, jan. 2020.

KOURENTZES, N.; BARROW, D. K.; CRONE, S. F. Neural network ensemble operators for time series forecasting. **Expert Systems with Applications**, v. 41, n. 9, p. 4235–4244, 2014.

KOUTSANDREAS, D. et al. On the selection of forecasting accuracy measures. **Journal of the Operational Research Society**, v. 73, n. 5, p. 937–954, 2022.

LAWSON, J. **Design and Analysis of Experiments with R**. Provo: [s.n.].

LEWIS, H. F.; SEXTON, T. R. Data Envelopment Analysis with Reverse Inputs and Outputs  
Stable URL : <https://www.jstor.org/stable/41770150> Linked references are available on  
JSTOR for this article : Data Envelopment Analysis with Reverse Inputs and Outputs. v. 21,  
n. 2, p. 113–132, 2020.

LIN, J.; PULIDO, J.; ASPLUND, M. Reliability analysis for preventive maintenance based  
on classical and Bayesian semi-parametric degradation approaches using locomotive wheel-  
sets as a case study. **Reliability Engineering and System Safety**, v. 134, p. 143–156, 2015.

LIU, H.; CAI, L.; WU, X. Grey-RBF neural network prediction model for city electricity  
demand forecasting. **2008 International Conference on Wireless Communications,  
Networking and Mobile Computing, WiCOM 2008**, n. 20070228, p. 1–5, 2008.

MAKRIDAKIS, S. Accuracy measures: theoretical and practical concerns. **International  
Journal of Forecasting**, v. 9, n. 4, p. 527–529, 1993.

MAKRIDAKIS, S.; HIBON, M. The M3-competition: Results, conclusions and implications.  
**International Journal of Forecasting**, v. 16, n. 4, p. 451–476, 2000.

MAYER, J.; KHAIRY, K.; HOWARD, J. Drawing an elephant with four complex  
parameters. **American Journal of Physics**, v. 78, n. 6, p. 648–649, 2010.

MENDES, R. R. A. et al. Multiobjective portfolio optimization of ARMA-GARCH time  
series based on experimental designs. **Computers and Operations Research**, v. 66, p. 434–  
444, 2016.

METCALFE, A. V.; COWPERTWAIT, P. S. P. **Introductory Time Series with R**. New  
York, NY: Springer New York, 2009.

METROPOLIS, N.; ULAM, S. The Monte Carlo method. **Journal of the American  
Statistical Association**, v. 44, n. 247, p. 335–341, 1949.

MONTGOMERY, D. C. **Design and Analysis of Experiments**. 9. ed. New York: John  
Wiley & Sons, 2017.

MONTGOMERY, D. C.; JENNINGS, C. L.; KULAHCI, M. Introduction Time Series Analysis and Forecasting. p. 671, 2015.

MONTGOMERY, D. C.; PECK, E. A.; VINING, G. G. **Introduction to linear regression analysis**. [s.l.] John Wiley & Sons, 2021.

MONTGOMERY, D. C.; RUNGER, G. C. **Applied statistics and probability for engineers**. [s.l.] John Wiley & Sons, 2010.

MORETTIN, P. A.; BUSSAB, W. O. **Estatística básica**. [s.l.] Saraiva Educação SA, 2017.

NARIMANI, R.; NARIMANI, A. A new hybrid model for improvement of ARIMA by DEA. **Decision Science Letters**, v. 1, n. 2, p. 59–68, 2012.

NAVES, F. L. et al. Multivariate Normal Boundary Intersection based on rotated factor scores: A multiobjective optimization method for methyl orange treatment. **Journal of Cleaner Production**, v. 143, p. 413–439, 2017.

NIELSEN, A. **Practical time series analysis: prediction with statistics and machine learning**. [s.l.] O'Reilly Media, 2019.

NORYANI, M. et al. Material selection of natural fibre using a stepwise regression model with error analysis. **Journal of Materials Research and Technology**, v. 8, n. 3, p. 2865–2879, 2019.

OLESEN, O. B.; PETERSEN, N. C.; PODINOVSKI, V. V. Efficiency analysis with ratio measures. **European Journal of Operational Research**, v. 245, n. 2, p. 446–462, set. 2015.

POLIKAR, R. Ensemble based systems in decision making. **IEEE Circuits and Systems Magazine**, v. 6, n. 3, p. 21–45, 2006.

ROBERT, C. P.; CASELLA, G. **Introduction to Monte Carlo Methods with R**. New York: Springer New York, 2008.

ROSS, C. A. **Simulation**. New York: Academic Press, 1997.

SAGI, O.; ROKACH, L. Ensemble learning: A survey. **Wiley Interdisciplinary Reviews:**



**Data Mining and Knowledge Discovery**, v. 8, n. 4, p. 1–18, 2018.

SCHEFFÉ, H. Experiments with Mixtures. **Journal of the Royal Statistical Society**, 1958.

SCHUMACKER, R. E. **Using R with multivariate statistics**. [s.l.] Sage Publications, 2015.

SHARPE, N. R.; DE VEAUX, R. D.; VELLEMAN, P. F. **Business statistics**. [s.l.] Pearson Education, Inc., 2021.

SOBOL, I. M. **Método de Monte Carlo**. Moscou: MIR, 1976.

STAICULESCU, D. et al. Design and optimization of 3-D compact stripline and microstrip Bluetooth/WLAN balun architectures using the design of experiments technique. **IEEE Transactions on Antennas and Propagation**, v. 53, n. 5, p. 1805–1812, maio 2005.

SUGANTHI, L.; SAMUEL, A. A. Energy models for demand forecasting - A review. **Renewable and Sustainable Energy Reviews**, v. 16, n. 2, p. 1223–1240, 2012.

TANG, L. et al. A novel hybrid ensemble learning paradigm for nuclear energy consumption forecasting. **Applied Energy**, v. 93, p. 432–443, 2012.

THANASSOULIS, E. **Introduction To the Theory and Application of Data**. [s.l.: s.n.].

TSAY, R. S. **Analysis of financial time series**. [s.l.] John Wiley & Sons, 2005.

VON NEUMANN, J. **Various techniques used in connection with random digits**. **National Bureau of Standards Applied Mathematics Series**, 1951.

WALKER, G. on Periodicity in Series of Related Terms. **Monthly Weather Review**, v. 59, n. 7, p. 277–278, 1931.

WANG, J. et al. Combined modeling for electric load forecasting with adaptive particle swarm optimization. **Energy**, v. 35, n. 4, p. 1671–1678, 2010.

WANG, J.; WAN, W. Experimental design methods for fermentative hydrogen production: A review. **International Journal of Hydrogen Energy**, v. 34, n. 1, p. 235–244, 2009.

WANG, M. et al. A comparison of approaches to stepwise regression on variables sensitivities in building simulation and analysis. **Energy and Buildings**, v. 127, p. 313–326, 2016.

WANG, X.; SMITH-MILES, K.; HYNDMAN, R. Rule induction for forecasting method selection: Meta-learning the characteristics of univariate time series. **Neurocomputing**, v. 72, n. 10–12, p. 2581–2594, jun. 2009.

WICHARD, J. D.; OGORZALEK, M. Time series prediction with ensemble models applied to the CATS benchmark. **Neurocomputing**, v. 70, n. 13–15, p. 2371–2378, 2007.

WOOLDRIDGE, J. M. **Introductory econometrics: A modern approach**. [s.l.] Cengage Learning, 2015.

YANJUN, L.; YULIANG, Z. Energy demand forecast of Henan Province by Using Gray Models GM(1,1). **Henan Sci.**, v. 12, n. 10, 2009.

YONDO, R.; ANDRÉS, E.; VALERO, E. A review on design of experiments and surrogate models in aircraft real-time and many-query aerodynamic analyses. **Progress in Aerospace Sciences**, v. 96, n. March 2017, p. 23–61, 2018.

YULE, G. U. On a Method of Investigating Periodicities in Disturbed Series. **Philosophical Transactions of the Royal Society of London**, v. 226, n. Series A, p. 167–298, 1927.

ZADEH, L. Optimality and non-scalar-valued performance criteria. **IEEE Transactions on Automatic Control**, v. 8, n. 1, p. 59–60, 1963.

ZHENG, Z.; PADMANABHAN, B. Constructing ensembles from data envelopment analysis. **INFORMS Journal on Computing**, v. 19, n. 4, p. 486–496, 2007.

## APÊNDICE A

### A.1. REGIÃO NORTE

Tabela A. 1. Coeficientes de Regressão Estimados para o MAPE considerando a região Norte.

	<b>Coef</b>	<b>EP de Coef</b>	<b>Valor-T</b>	<b>Valor-P</b>	<b>VIF</b>
$x_1$	2,682	0,543	*	*	3,99
$x_2$	-0,014	0,377	*	*	1,92
$x_3$	0,436	0,373	*	*	1,89
$x_4$	0,919	0,374	*	*	1,89
$x_1x_2$	-10,06	2,40	-4,20	0,000	3,37
$x_1x_3$	-8,67	2,16	-4,02	0,000	2,72
$x_1x_4$	-13,41	2,22	-6,04	0,000	2,89
$x_1^2x_2x_4$	109,9	36,8	2,99	0,004	1,75
$x_1x_2^2x_3$	114,9	35,8	3,21	0,002	1,66

Fonte: elaboração própria.

Tabela A. 2. Análise de variância para o MAPE considerando a região Norte.

	<b>GL</b>	<b>SQ Seq</b>	<b>SQ (Aj.)</b>	<b>QM (Aj.)</b>	<b>Valor F</b>	<b>Valor-P</b>
Regressão	8	30,1205	30,121	3,7651	6,55	0,000
Linear	3	0,3186	11,077	3,6925	6,43	0,001
Quadrático	3	20,3407	29,000	9,6665	16,82	0,000
$x_1x_2$	1	1,0914	10,119	10,1185	17,61	0,000
$x_1x_3$	1	3,6913	9,293	9,2934	16,17	0,000
$x_1x_4$	1	15,5580	20,963	20,9631	36,48	0,000
Quártico Especial	2	9,4611	9,461	4,7306	8,23	0,001
$x_1^2x_2x_4$	1	3,5473	5,121	5,1214	8,91	0,004
$x_1x_2^2x_3$	1	5,9139	5,914	5,9139	10,29	0,002
Erro de Resíduos	52	29,8795	29,879	0,5746		
Total	60	60,0000				

Fonte: elaboração própria.

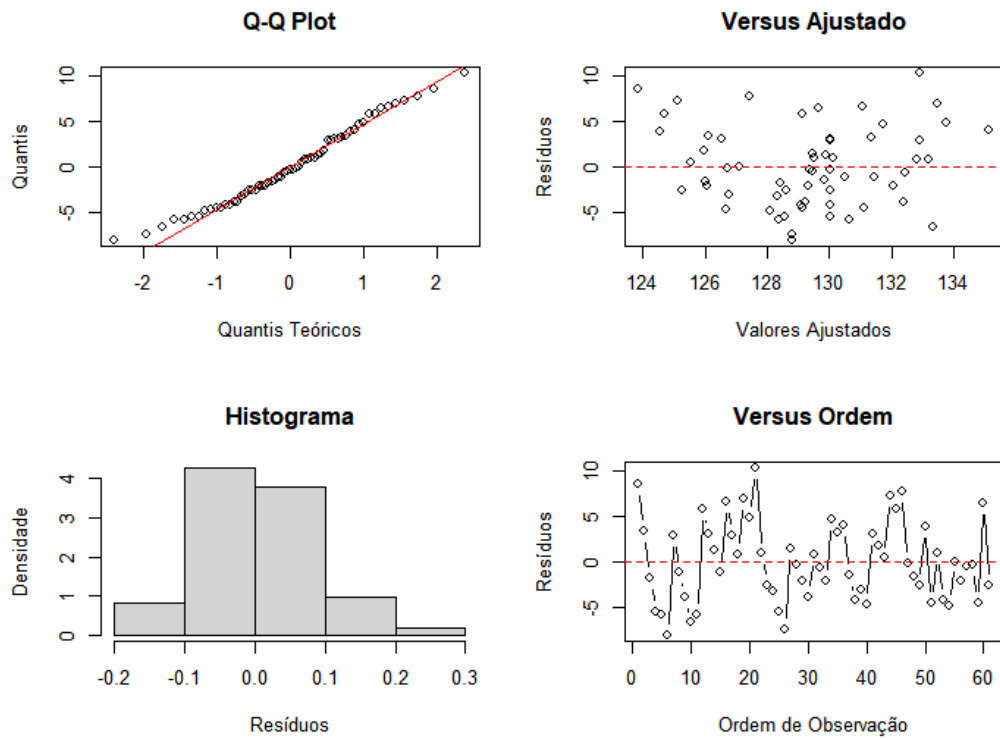


Figura A. 1. Análise de resíduos tendo o MAPE como resposta da região Norte.  
 Fonte: elaboração própria.

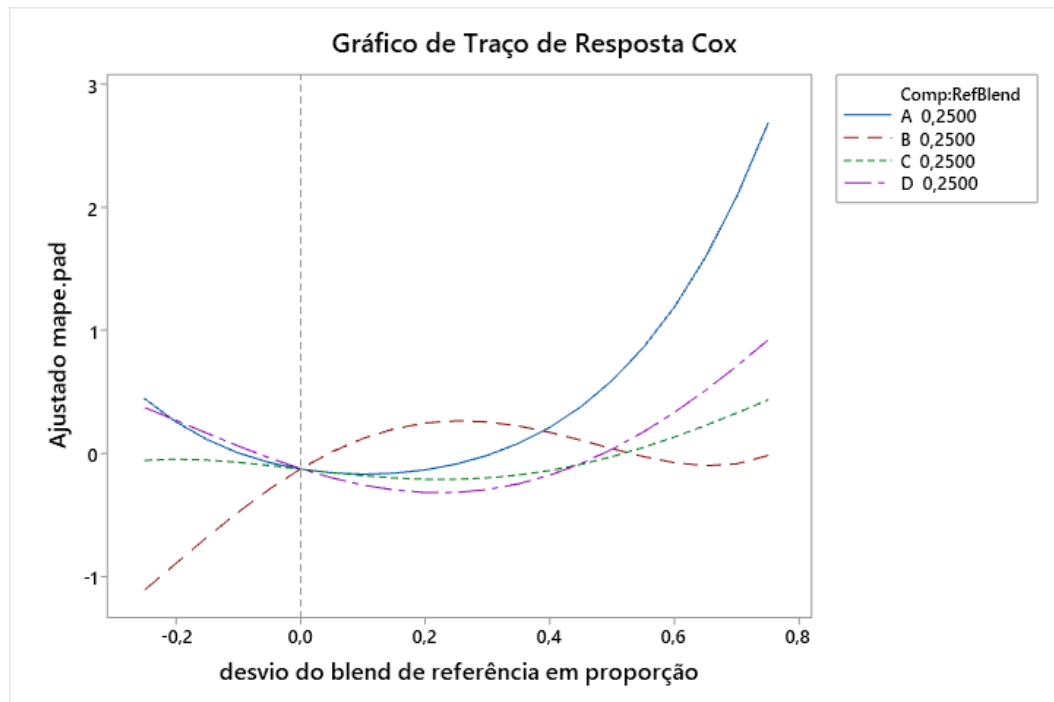


Figura A. 2. Gráfico de Traço de Resposta de Cox para o MAPE como resposta da região Norte.  
 Fonte: elaboração própria.

Tabela A. 3. Coeficientes de Regressão Estimados para o RMSE considerando a região Norte.

	Coef	EP de Coef	Valor-T	Valor-P	VIF
$x_1$	-0,358	0,461	*	*	1,93
$x_2$	0,684	0,461	*	*	1,93
$x_3$	0,733	0,461	*	*	1,93
$x_4$	-1,352	0,461	*	*	1,93
$x_1x_4$	7,18	2,47	2,91	0,005	2,40
$x_2x_3$	-5,39	2,47	-2,18	0,033	2,40

Fonte: elaboração própria.

Tabela A. 4. Análise de variância para o RMSE considerando a região Norte.

	GL	SQ Seq	SQ (Aj.)	QM (Aj.)	Valor F	Valor-P
Regressão	5	13,037	13,037	2,6074	3,05	0,017
Linear	3	2,678	7,870	2,6234	3,07	0,035
Quadrático	2	10,358	10,358	5,1792	6,07	0,004
$x_1x_4$	1	6,293	7,229	7,2292	8,47	0,005
$x_2x_3$	1	4,066	4,066	4,0657	4,76	0,033
Erro de Resíduos	55	46,963	46,963	0,8539		
Total	60	60,000				

Fonte: elaboração própria.

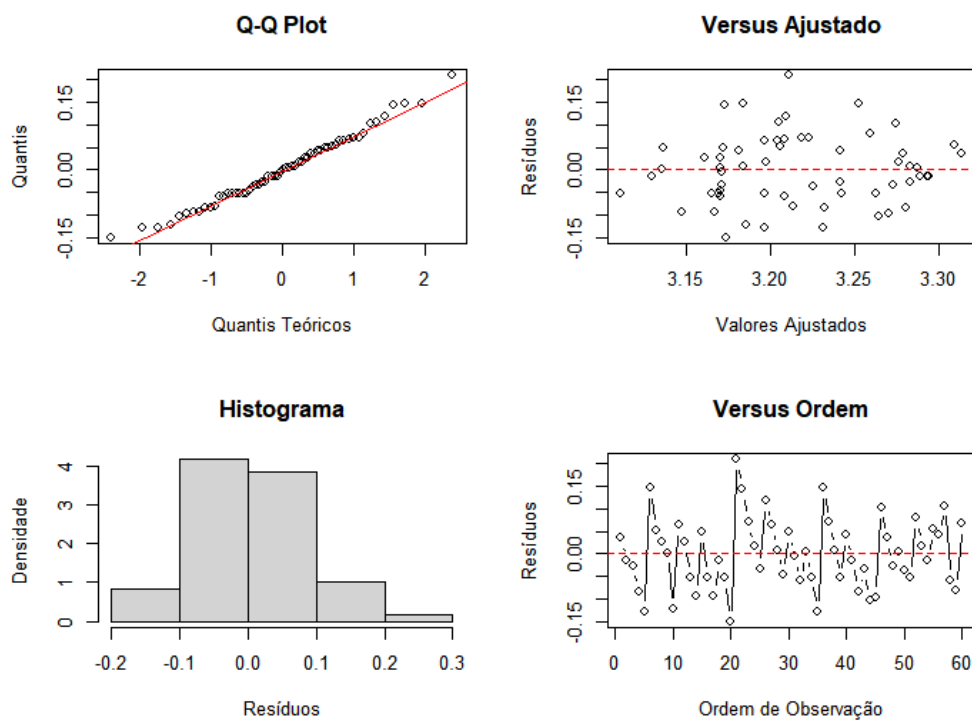


Figura A. 3. Análise de resíduos tendo o RMSE como resposta da região Norte.

Fonte: elaboração própria.

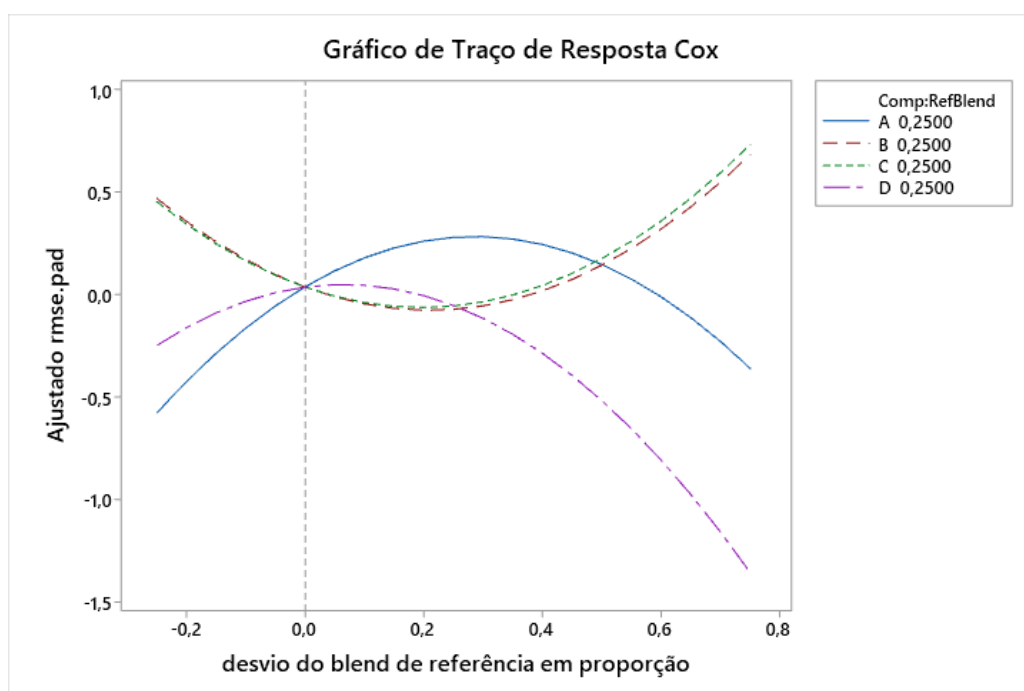


Figura A. 4. Gráfico de Traço de Resposta de Cox para o RMSE como resposta da região Norte.  
Fonte: elaboração própria.

Tabela A. 5. Coeficientes de Regressão Estimados para o MdAE considerando a região Norte.

	Coef	EP de Coef	Valor-T	Valor-P	VIF
$x_1$	2,398	0,623	*	*	3,80
$x_2$	0,916	0,438	*	*	1,88
$x_3$	-0,185	0,438	*	*	1,88
$x_4$	0,005	0,438	*	*	1,88
$x_1x_2$	-6,48	2,42	-2,68	0,010	2,48
$x_1x_3$	-5,81	2,42	-2,41	0,020	2,48
$x_1x_4$	-6,90	2,42	-2,86	0,006	2,48

Fonte: elaboração própria.

Tabela A. 6. Análise de variância para o MdAE considerando a região Norte.

	GL	SQ Seq	SQ (Aj.)	QM (Aj.)	Valor F	Valor-P
Regressão	6	17,185	17,185	2,8642	3,61	0,004
Linear	3	4,271	12,013	4,0042	5,05	0,004
Quadrático	3	12,914	12,914	4,3046	5,43	0,002
$x_1x_2$	1	3,275	5,702	5,7018	7,19	0,010
$x_1x_3$	1	3,164	4,590	4,5895	5,79	0,020
$x_1x_4$	1	6,475	6,475	6,4754	8,17	0,006
Erro de Resíduos	54	42,815	42,815	0,7929		
Total	60	60,000				

Fonte: elaboração própria.

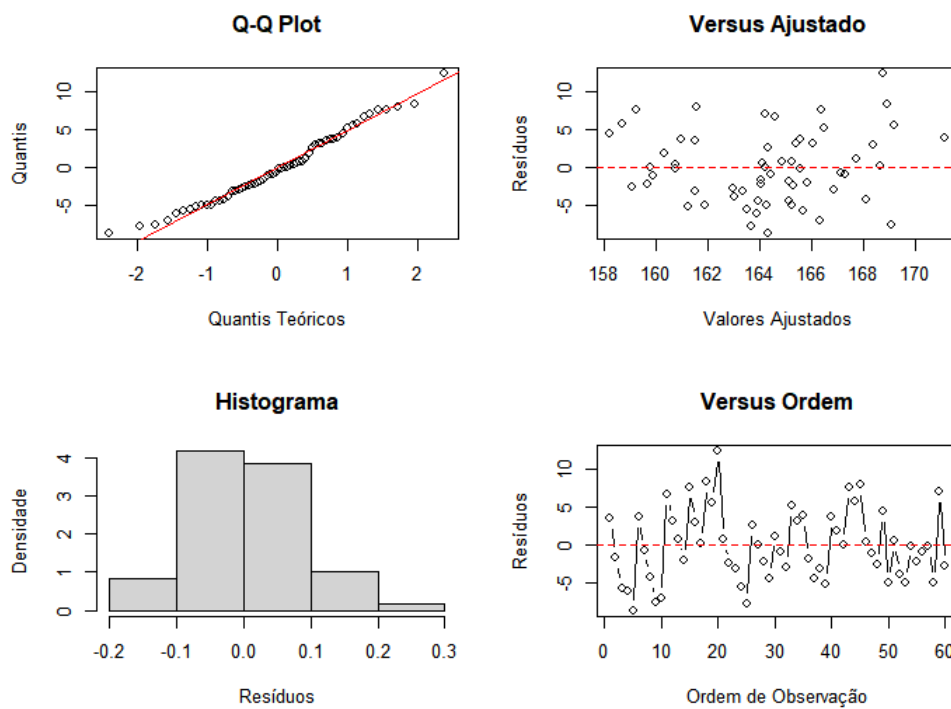


Figura A. 5. Análise de resíduos tendo o MdAE como resposta da região Norte.  
 Fonte: elaboração própria.

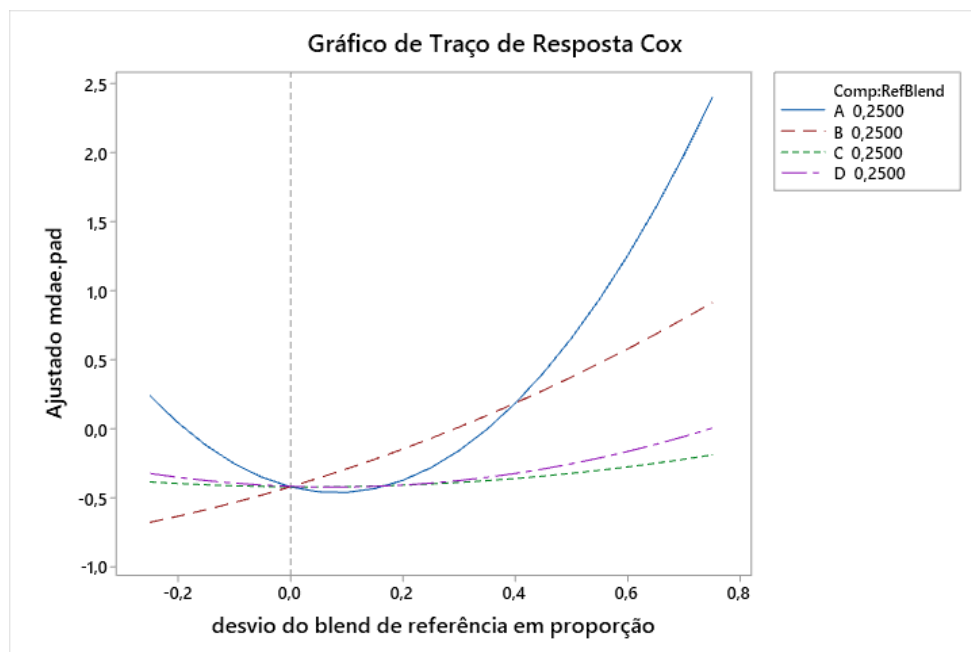


Figura A. 6. Gráfico de Traço de Resposta de Cox para o MdAE como resposta da região Norte.  
 Fonte: elaboração própria.

Tabela A. 7. Coeficientes de Regressão Estimados para o MAE considerando a região Norte.

	Coef	EP de Coef	Valor-T	Valor-P	VIF
$x_1$	-0,559	0,468	*	*	1,93
$x_2$	0,610	0,468	*	*	1,93
$x_3$	0,874	0,468	*	*	1,93
$x_4$	-1,179	0,468	*	*	1,93
$x_1x_4$	7,08	2,51	2,83	0,007	2,40
$x_2x_3$	-5,52	2,51	-2,20	0,032	2,40

Fonte: elaboração própria.

Tabela A. 8 Análise de variância para o MAE considerando a região Norte.

	GL	SQ Seq	SQ (Aj.)	QM (Aj.)	Valor F	Valor-P
Regressão	5	11,580	11,580	2,3161	2,63	0,033
Linear	3	1,224	6,784	2,2614	2,57	0,064
Quadrático	2	10,356	10,356	5,1779	5,88	0,005
$x_1x_4$	1	6,079	7,027	7,0271	7,98	0,007
$x_2x_3$	1	4,277	4,277	4,2768	4,86	0,032
Erro de Resíduos	55	48,420	48,420	0,8804		
Total	60	60,000				

Fonte: elaboração própria.

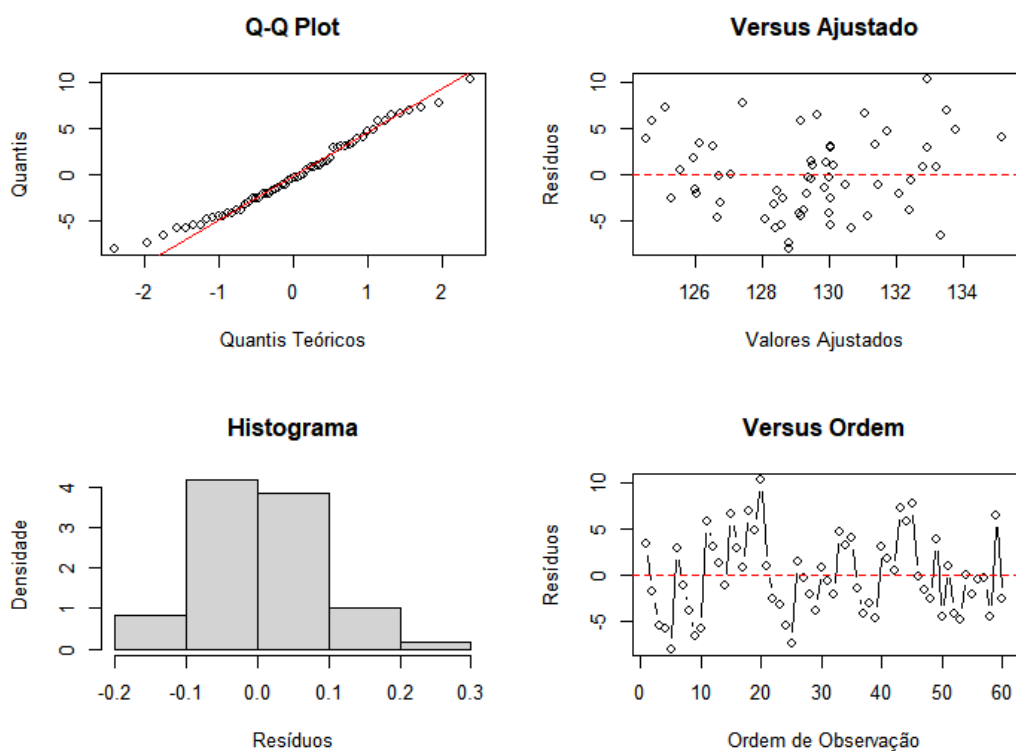


Figura A. 7. Análise de resíduos tendo o MAE como resposta da região Norte.  
Fonte: elaboração própria.



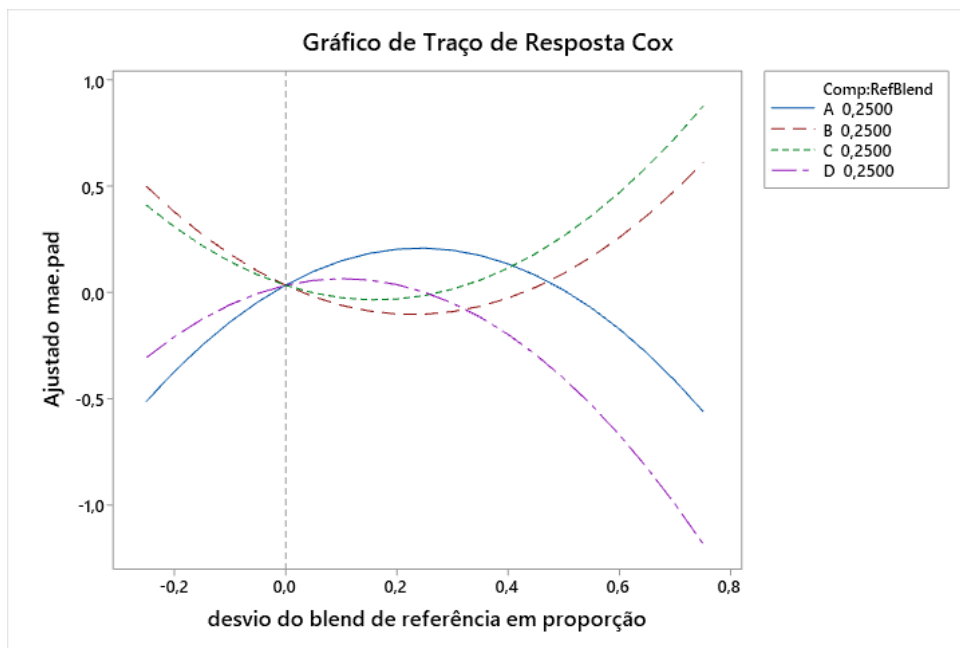


Figura A. 8. Gráfico de Traço de Resposta de Cox para o MAE como resposta da região Norte.  
Fonte: elaboração própria.

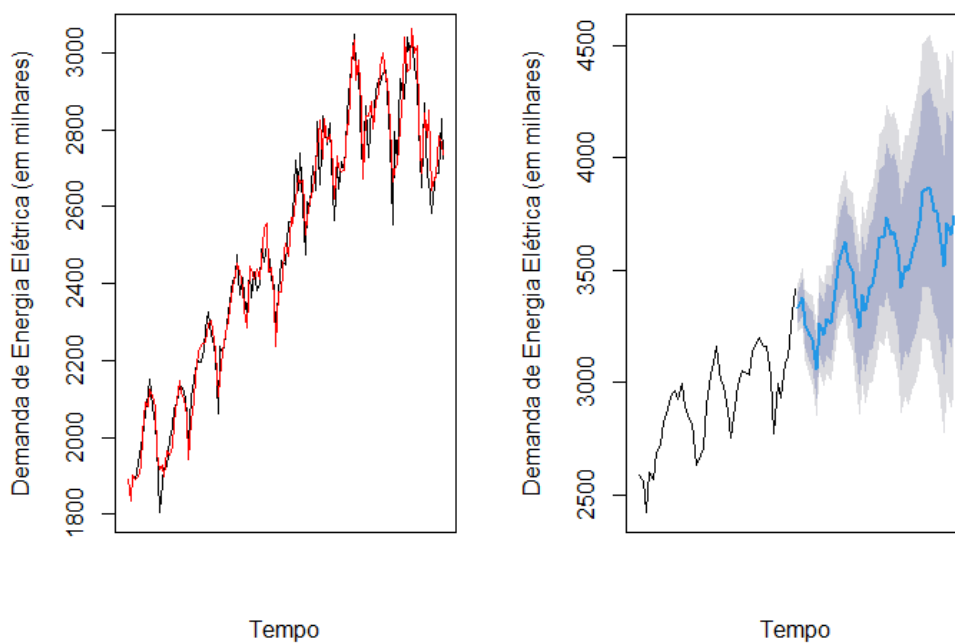


Figura A. 9. Gráfico de ajuste e previsão aplicando a metodologia proposta para a Região Norte.  
Fonte: elaboração própria.

Tabela A. 9. Previsão e intervalos de confiança da demanda de energia elétrica na Região Norte.

<b>Tempo</b>	<b>Previsão</b>	<b>80%<sub>inf</sub></b>	<b>80%<sub>sup</sub></b>	<b>95%<sub>inf</sub></b>	<b>95%<sub>sup</sub></b>
ago/22	3.378,516	3.096,886	3.660,146	3.194,307	3.562,725
set/22	3.452,032	3.170,402	3.733,662	3.267,823	3.636,241
out/22	3.335,115	3.053,485	3.616,745	3.150,906	3.519,324
nov/22	3.274,014	2.992,384	3.555,644	3.089,805	3.458,223
dez/22	3.185,203	2.903,573	3.466,833	3.000,994	3.369,412
jan/23	3.041,329	2.759,699	3.322,959	2.857,120	3.225,538
fev/23	3.235,553	2.953,923	3.517,183	3.051,344	3.419,762
mar/23	3.211,389	2.929,759	3.493,019	3.027,180	3.395,598
abr/23	3.293,809	3.012,179	3.575,439	3.109,600	3.478,018
mai/23	3.274,638	2.993,008	3.556,268	3.090,429	3.458,847
jun/23	3.336,846	3.055,216	3.618,476	3.152,637	3.521,055
jul/23	3.410,618	3.128,988	3.692,248	3.226,409	3.594,827
ago/23	3.434,357	3.152,727	3.715,987	3.250,148	3.618,566
set/23	3.492,108	3.210,478	3.773,738	3.307,899	3.676,317
out/23	3.434,530	3.152,900	3.716,160	3.250,321	3.618,739
nov/23	3.414,947	3.133,317	3.696,577	3.230,738	3.599,156
dez/23	3.315,548	3.033,918	3.597,178	3.131,339	3.499,757
jan/24	3.173,026	2.891,396	3.454,656	2.988,817	3.357,235
fev/24	3.316,554	3.034,924	3.598,184	3.132,345	3.500,763
mar/24	3.225,091	2.943,461	3.506,721	3.040,882	3.409,300
abr/24	3.324,448	3.042,818	3.606,078	3.140,239	3.508,657
mai/24	3.336,217	3.054,587	3.617,847	3.152,008	3.520,426
jun/24	3.455,967	3.174,337	3.737,597	3.271,758	3.640,176
jul/24	3.549,511	3.267,881	3.831,141	3.365,302	3.733,720
ago/24	3.565,839	3.284,209	3.847,469	3.381,630	3.750,048
set/24	3.631,387	3.349,757	3.913,017	3.447,178	3.815,596
out/24	3.543,985	3.262,355	3.825,615	3.359,776	3.728,194
nov/24	3.530,962	3.249,332	3.812,592	3.346,753	3.715,171
dez/24	3.431,259	3.149,629	3.712,889	3.247,050	3.615,468
jan/25	3.316,901	3.035,271	3.598,531	3.132,692	3.501,110
fev/25	3.440,725	3.159,095	3.722,355	3.256,516	3.624,934
mar/25	3.397,956	3.116,326	3.679,586	3.213,747	3.582,165
abr/25	3.465,944	3.184,314	3.747,574	3.281,735	3.650,153
mai/25	3.461,550	3.179,920	3.743,180	3.277,341	3.645,759
jun/25	3.532,252	3.250,622	3.813,882	3.348,043	3.716,461
jul/25	3.612,825	3.331,195	3.894,455	3.428,616	3.797,034
ago/25	3.641,733	3.360,103	3.923,363	3.457,524	3.825,942
set/25	3.693,462	3.411,832	3.975,092	3.509,253	3.877,671
out/25	3.620,403	3.338,773	3.902,033	3.436,194	3.804,612
nov/25	3.608,039	3.326,409	3.889,669	3.423,830	3.792,248
dez/25	3.503,040	3.221,410	3.784,670	3.318,831	3.687,249
jan/26	3.348,432	3.066,802	3.630,062	3.164,223	3.532,641
fev/26	3.509,657	3.228,027	3.791,287	3.325,448	3.693,866
mar/26	3.457,750	3.176,120	3.739,380	3.273,541	3.641,959
abr/26	3.550,857	3.269,227	3.832,487	3.366,648	3.735,066

Fonte: elaboração própria.

## A.2. REGIÃO NORDESTE

Tabela A. 10. Coeficientes de Regressão Estimados para o MAPE considerando a região Nordeste.

	Coef	EP de Coef	Valor-T	Valor-P	VIF
$x_1$	-1,789	0,483	*	*	4,12
$x_2$	-0,164	0,354	*	*	2,21
$x_3$	-0,669	0,351	*	*	2,17
$x_4$	-1,553	0,340	*	*	2,04
$x_1x_2$	9,40	2,16	4,36	0,000	3,55
$x_1x_3$	7,13	1,97	3,62	0,001	2,96
$x_1x_4$	12,92	1,98	6,52	0,000	2,99
$x_1^2x_2x_3$	88,4	42,6	2,08	0,043	3,05
$x_1^2x_2x_4$	-148,4	42,9	-3,46	0,001	3,09
$x_2^2x_3x_4$	100,0	41,2	2,43	0,019	2,86
$x_1x_2^2x_3$	-155,1	41,8	-3,71	0,001	2,95
$x_2x_3^2x_4$	-63,4	41,2	-1,54	0,130	2,86
$x_1x_2x_4^2$	62,2	41,3	1,51	0,139	2,87

Tabela A. 11. Análise de variância para o MAPE considerando a região Nordeste.

	GL	SQ Seq	SQ (Aj.)	QM (Aj.)	Valor F	Valor-P
Regressão	12	38,8507	38,8507	3,2376	7,35	0,000
Linear	3	7,2421	4,8404	1,6135	3,66	0,019
Quadrático	3	19,4180	23,8209	7,9403	18,02	0,000
$x_1x_2$	1	1,2244	8,3818	8,3818	19,02	0,000
$x_1x_3$	1	3,7532	5,7880	5,7880	13,14	0,001
$x_1x_4$	1	14,4404	18,7502	18,7502	42,56	0,000
Quártico Especial	6	12,1906	12,1906	2,0318	4,61	0,001
$x_1^2x_2x_3$	1	0,0296	1,9009	1,9009	4,31	0,043
$x_1^2x_2x_4$	1	3,2585	5,2830	5,2830	11,99	0,001
$x_2^2x_3x_4$	1	1,3039	2,5946	2,5946	5,89	0,019
$x_1x_2^2x_3$	1	5,5816	6,0511	6,0511	13,73	0,001
$x_2x_3^2x_4$	1	1,0185	1,0432	1,0432	2,37	0,130
$x_1x_2x_4^2$	1	0,9984	0,9984	0,9984	2,27	0,139
Erro de Resíduos	48	21,1493	21,1493	0,4406		
Total	60	60,0000				

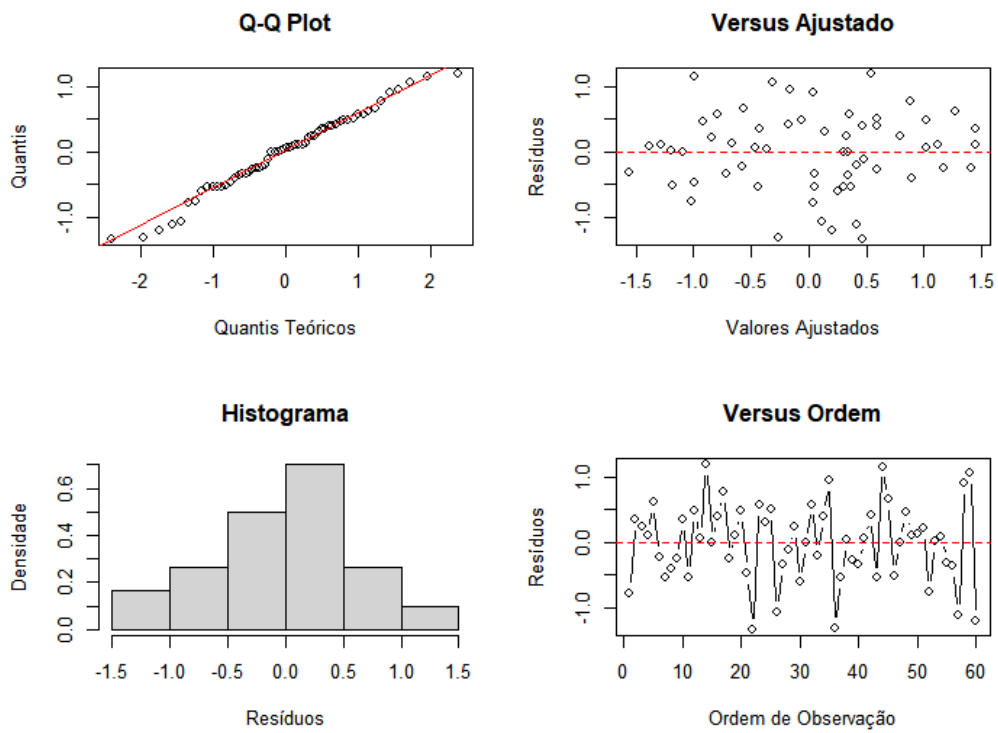


Figura A. 10. Análise de resíduos tendo o MAPE como resposta da região Nordeste.  
 Fonte: elaboração própria.

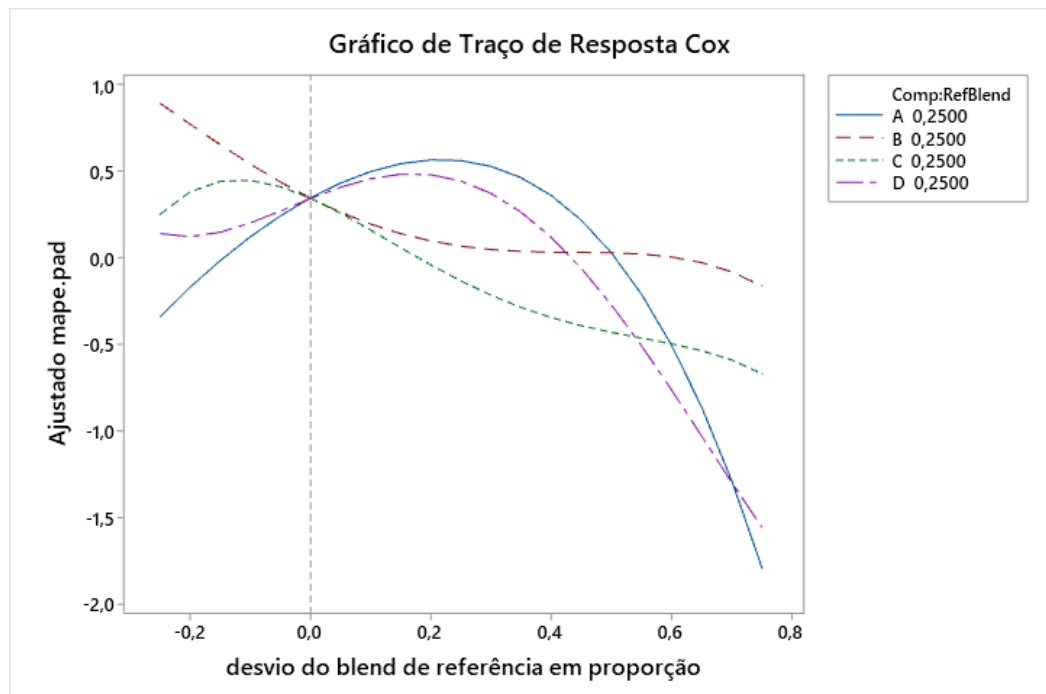


Figura A. 11. Gráfico de Traço de Resposta de Cox para o MAPE como resposta da região Nordeste.  
 Fonte: elaboração própria.

Tabela A. 12. Coeficientes de Regressão Estimados para o RMSE considerando a região Nordeste.

	Coef	EP de Coef	Valor-T	Valor-P	VIF
$x_1$	1,479	0,524	*	*	2,71
$x_2$	-0,649	0,444	*	*	1,95
$x_3$	-0,567	0,524	*	*	2,71
$x_4$	1,332	0,444	*	*	1,95
$x_1x_3$	-4,70	2,40	-1,96	0,055	2,46
$x_1x_4$	-9,26	2,39	-3,88	0,000	2,44
$x_2x_3$	4,18	2,39	1,75	0,086	2,44

Fonte: elaboração própria.

Tabela A. 13. Análise de variância para o RMSE considerando a região Nordeste.

	GL	SQ Seq	SQ (Aj.)	QM (Aj.)	Valor F	Valor-P
Regressão	6	17,448	17,448	2,9081	3,69	0,004
Linear	3	1,656	8,765	2,9216	3,71	0,017
Quadrático	3	15,793	15,793	5,2643	6,68	0,001
$x_1x_3$	1	2,298	3,021	3,0215	3,83	0,055
$x_1x_4$	1	11,078	11,841	11,8406	15,03	0,000
$x_2x_3$	1	2,416	2,416	2,4164	3,07	0,086
Erro de Resíduos	54	42,552	42,552	0,7880		
Total	60	60,000				

Fonte: elaboração própria.

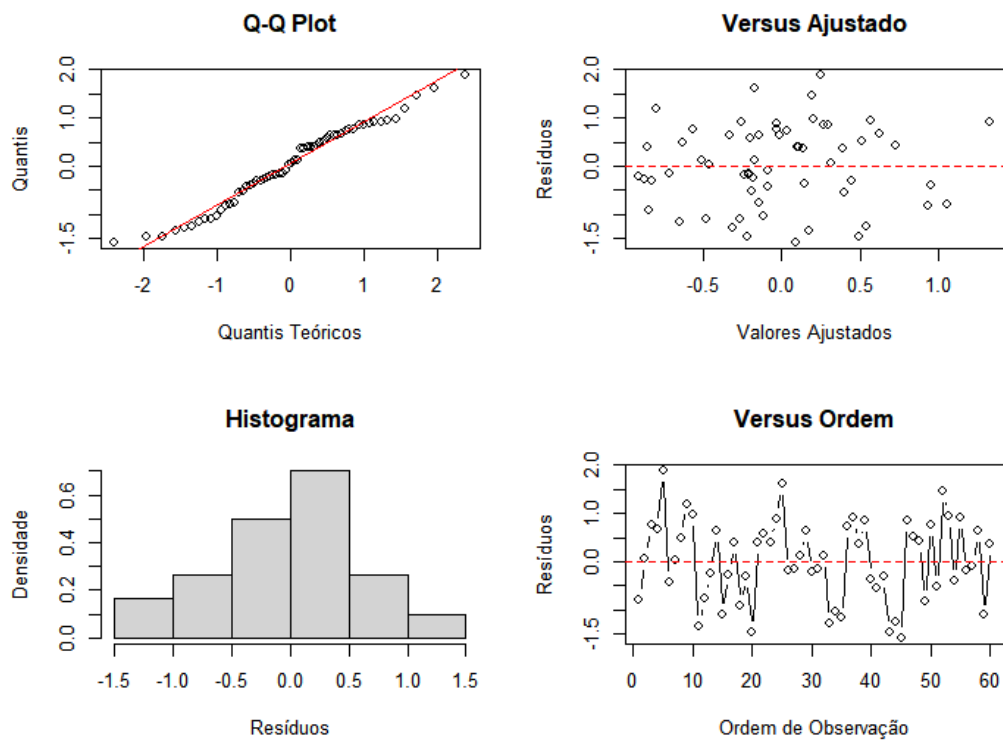


Figura A. 12. Análise de resíduos tendo o RMSE como resposta da região Nordeste.

Fonte: elaboração própria.

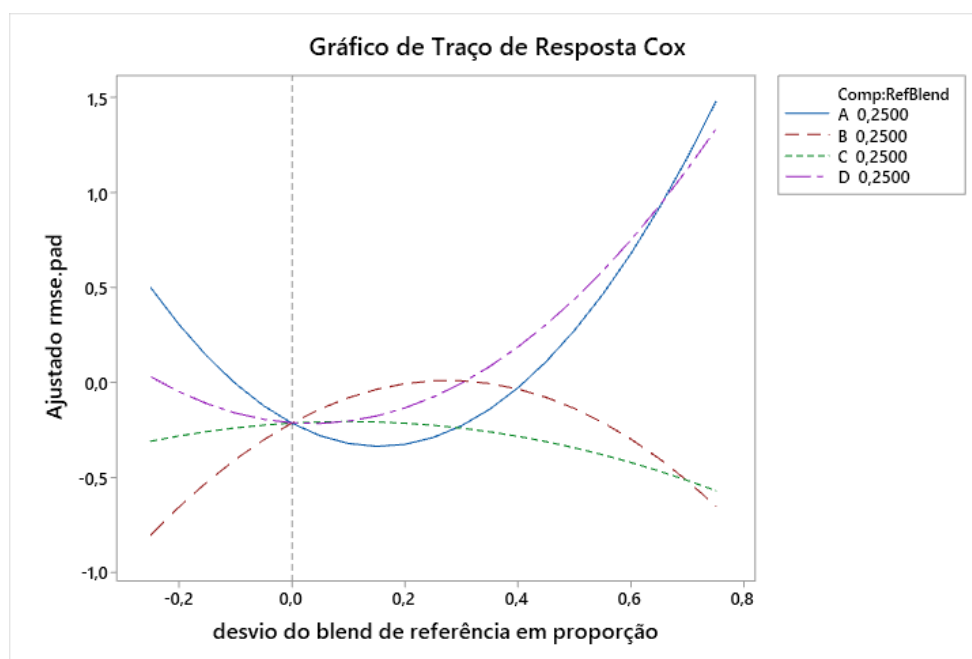


Figura A. 13. Gráfico de Traço de Resposta de Cox para o RMSE como resposta da região Nordeste.  
Fonte: elaboração própria.

Tabela A. 14 Coeficientes de Regressão Estimados para o MdAE considerando a região Nordeste.

	Coef	EP de Coef	Valor-T	Valor-P	VIF
$x_1$	1,366	0,498	*	*	2,84
$x_2$	0,183	0,498	*	*	2,84
$x_3$	-0,303	0,348	*	*	1,38
$x_4$	1,918	0,502	*	*	2,87
$x_1x_2$	-4,08	2,56	-1,60	0,116	3,23
$x_1x_4$	-9,53	2,52	-3,78	0,000	3,14
$x_2x_4$	-8,23	2,52	-3,27	0,002	3,14
$x_1^2x_2x_4$	163,9	53,1	3,09	0,003	3,07
$x_1x_2^2x_4$	-89,7	53,1	-1,69	0,097	3,07

Fonte: elaboração própria.

Tabela A. 15. Análise de variância para o MdAE considerando a região Nordeste.

	GL	SQ Seq	SQ (Aj.)	QM (Aj.)	Valor F	Valor-P
Regressão	8	24,5669	24,567	3,0709	4,51	0,000
Linear	3	6,0673	9,372	3,1240	4,58	0,006
Quadrático	3	11,9878	12,988	4,3293	6,35	0,001
$x_1x_2$	1	0,0135	1,736	1,7365	2,55	0,116
$x_1x_4$	1	5,3874	9,742	9,7418	14,30	0,000
$x_2x_4$	1	6,5869	7,267	7,2665	10,66	0,002
Quártico Especial	2	6,5119	6,512	3,2559	4,78	0,012
$x_1^2x_2x_4$	1	4,5638	6,494	6,4936	9,53	0,003
$x_1x_2^2x_4$	1	1,9481	1,948	1,9481	2,86	0,097
Erro de Resíduos	52	35,4331	35,433	0,6814		
Total	60	60,0000				

Fonte: elaboração própria.

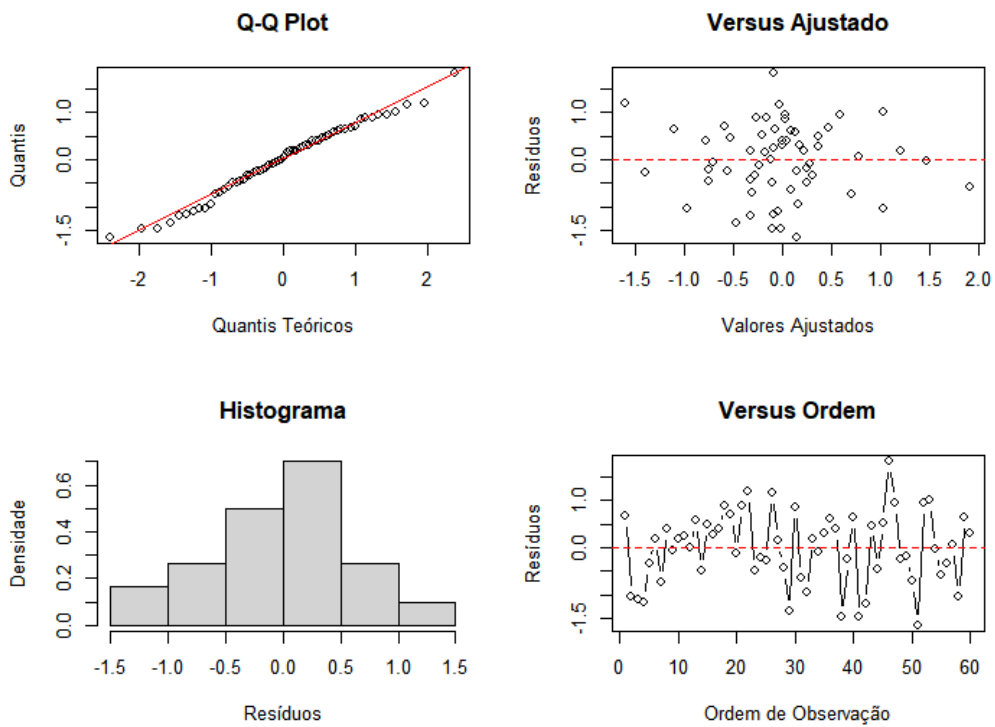


Figura A. 14. Análise de resíduos tendo o MdAE como resposta da região Nordeste.  
Fonte: elaboração própria.

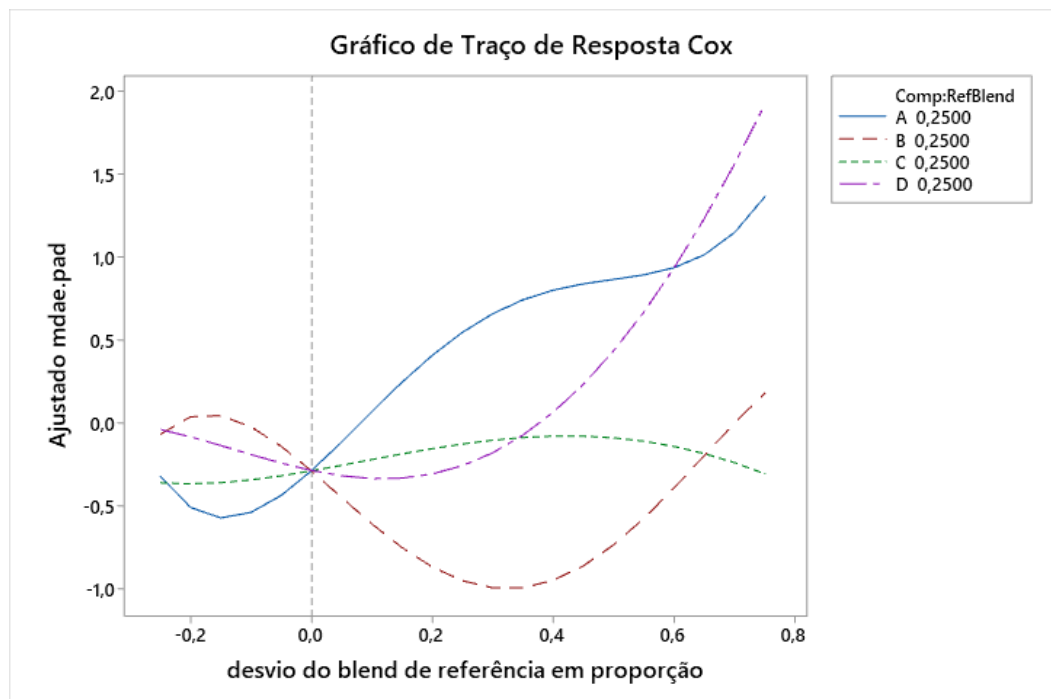


Figura A. 15. Gráfico de Traço de Resposta de Cox para o MdAE como resposta da região Nordeste.  
Fonte: elaboração própria.

Tabela A. 16. Coeficientes de Regressão Estimados para o MAE considerando a região Nordeste.

	<b>Coef</b>	<b>EP de Coef</b>	<b>Valor-T</b>	<b>Valor-P</b>	<b>VIF</b>
$x_1$	-0,231	0,541	*	*	2,79
$x_2$	-0,881	0,472	*	*	2,12
$x_3$	-0,534	0,377	*	*	1,35
$x_4$	-0,370	0,458	*	*	2,00
$x_1x_2$	7,83	2,64	2,96	0,005	2,87
$x_1x_4$	5,00	2,64	1,89	0,064	2,88
$x_1^2x_2x_4$	-163,4	57,7	-2,83	0,007	3,02
$x_2^2x_3x_4$	57,4	38,6	1,49	0,143	1,35
$x_1x_2x_4^2$	91,7	56,0	1,64	0,108	2,85

Fonte: elaboração própria.

Tabela A. 17. Análise de variância para o MAE considerando a região Nordeste.

	<b>GL</b>	<b>SQ Seq</b>	<b>SQ (Aj.)</b>	<b>QM (Aj.)</b>	<b>Valor F</b>	<b>Valor-P</b>
Regressão	8	17,523	17,5231	2,1904	2,68	0,015
Linear	3	4,712	0,8151	0,2717	0,33	0,802
Quadrático	2	4,775	8,2592	4,1296	5,06	0,010
$x_1x_2$	1	3,438	7,1793	7,1793	8,79	0,005
$x_1x_4$	1	1,337	2,9196	2,9196	3,57	0,064
Quártico Especial	3	8,037	8,0366	2,6789	3,28	0,028
$x_1^2x_2x_4$	1	4,193	6,5614	6,5614	8,03	0,007
$x_2^2x_3x_4$	1	1,654	1,8042	1,8042	2,21	0,143
$x_1x_2x_4^2$	1	2,189	2,1895	2,1895	2,68	0,108
Erro de Resíduos	52	42,477	42,4769	0,8169		
Total	60	60,000				

Fonte: elaboração própria.



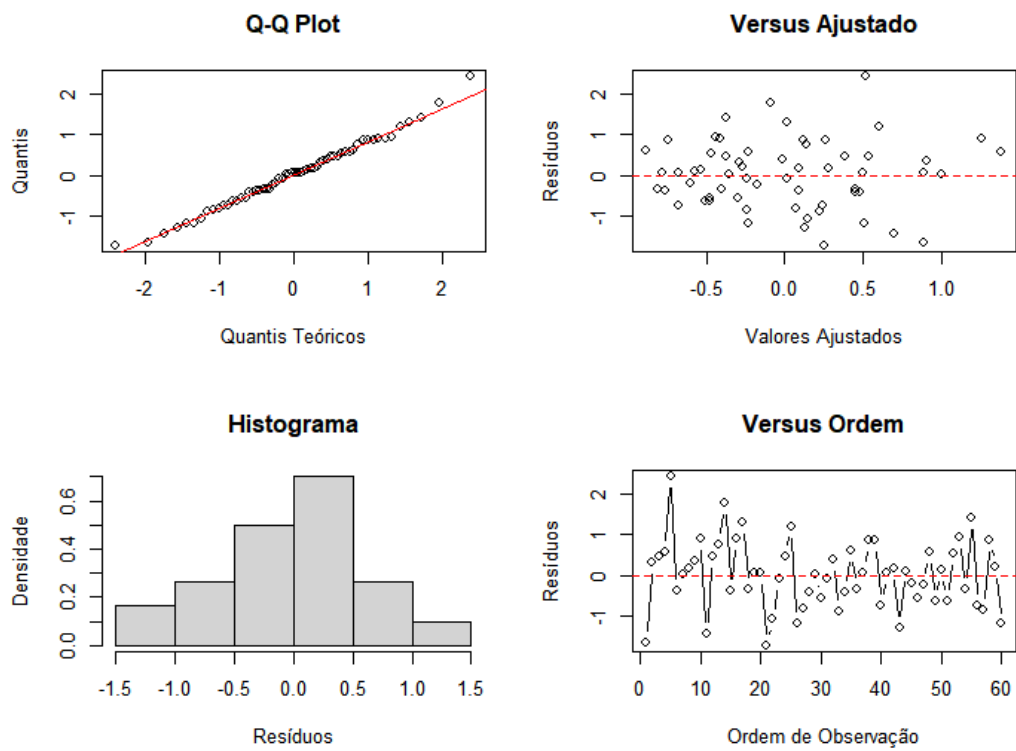


Figura A. 16. Análise de resíduos tendo o MAE como resposta da região Nordeste.  
 Fonte: elaboração própria.

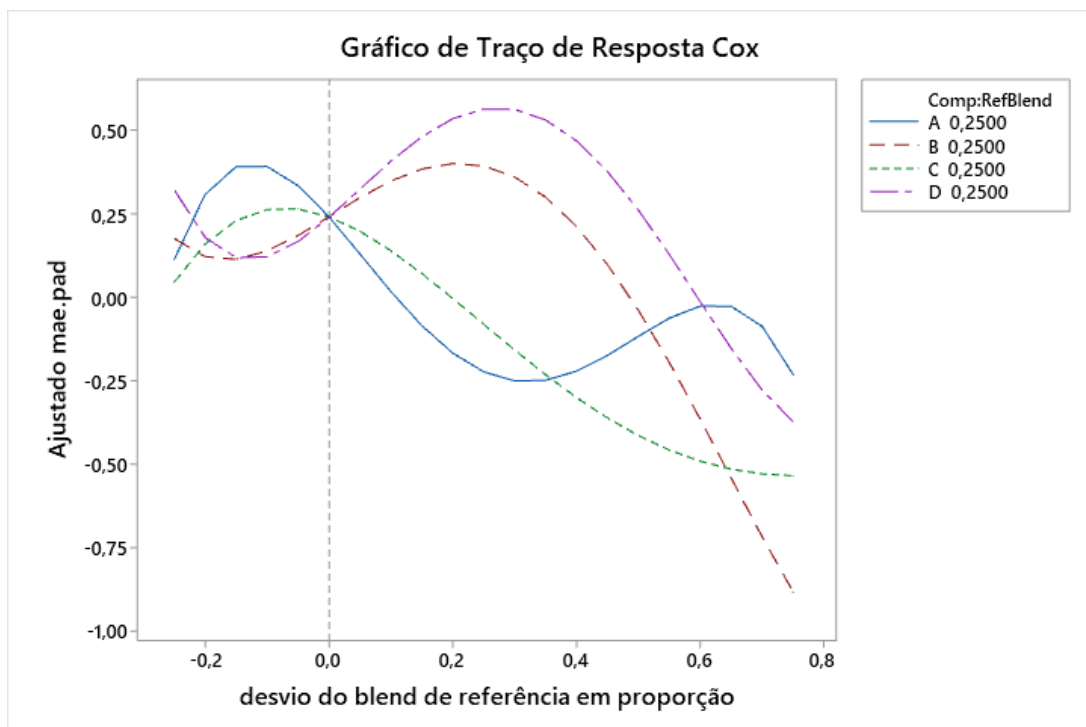


Figura A. 17. Gráfico de Traço de Resposta de Cox para o MAE como resposta da região Nordeste.  
 Fonte: elaboração própria.

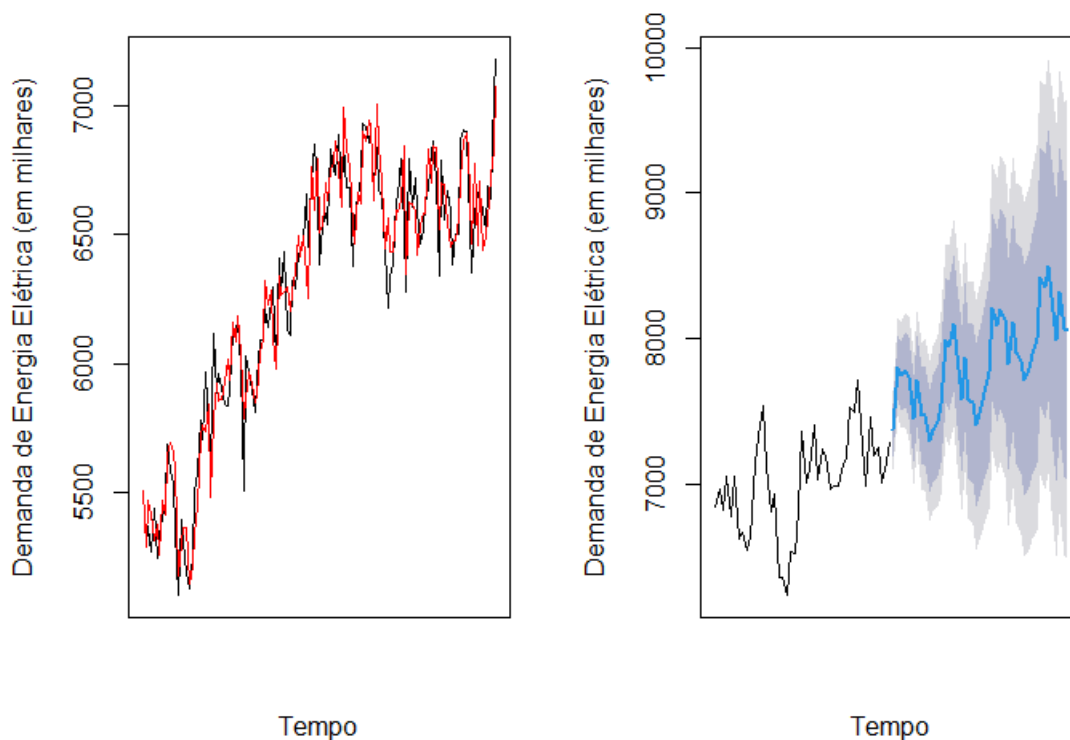


Figura A. 18. Gráfico de ajuste e previsão aplicando a metodologia proposta para a Região Nordeste.  
Fonte: elaboração própria.

Tabela A. 18. Previsão e intervalos de confiança da demanda de energia elétrica na Região Nordeste.

<b>Tempo</b>	<b>Previsão</b>	<b>80%<sub>inf</sub></b>	<b>80%<sub>sup</sub></b>	<b>95%<sub>inf</sub></b>	<b>95%<sub>sup</sub></b>
ago/22	7.479,613	7.055,265	7.903,961	6.830,844	8.128,382
set/22	7.852,847	7.428,499	8.277,195	7.204,078	8.501,616
out/22	7.670,411	7.246,063	8.094,759	7.021,642	8.319,180
nov/22	7.761,225	7.336,877	8.185,573	7.112,456	8.409,994
dez/22	7.772,727	7.348,379	8.197,075	7.123,958	8.421,496
jan/23	7.410,766	6.986,418	7.835,114	6.761,997	8.059,535
fev/23	7.751,686	7.327,338	8.176,034	7.102,917	8.400,455
mar/23	7.585,785	7.161,437	8.010,133	6.937,016	8.234,554
abr/23	7.511,599	7.087,251	7.935,947	6.862,830	8.160,368
mai/23	7.391,078	6.966,730	7.815,426	6.742,309	8.039,847
jun/23	7.443,649	7.019,301	7.867,997	6.794,880	8.092,418
jul/23	7.567,851	7.143,503	7.992,199	6.919,082	8.216,620
ago/23	7.712,100	7.287,752	8.136,448	7.063,331	8.360,869
set/23	8.058,485	7.634,137	8.482,833	7.409,716	8.707,254
out/23	7.976,977	7.552,629	8.401,325	7.328,208	8.625,746
nov/23	8.132,640	7.708,292	8.556,988	7.483,871	8.781,409
dez/23	7.886,333	7.461,985	8.310,681	7.237,564	8.535,102
jan/24	7.567,729	7.143,381	7.992,077	6.918,960	8.216,498
fev/24	7.982,545	7.558,197	8.406,893	7.333,776	8.631,314
mar/24	7.749,302	7.324,954	8.173,650	7.100,533	8.398,071
abr/24	7.758,502	7.334,154	8.182,850	7.109,733	8.407,271
mai/24	7.563,113	7.138,765	7.987,461	6.914,344	8.211,882
jun/24	7.641,056	7.216,708	8.065,404	6.992,287	8.289,825

jul/24	7.795,398	7.371,050	8.219,746	7.146,629	8.444,167
ago/24	7.976,281	7.551,933	8.400,629	7.327,512	8.625,050
set/24	8.336,460	7.912,112	8.760,808	7.687,691	8.985,229
out/24	8.197,062	7.772,714	8.621,410	7.548,293	8.845,831
nov/24	8.312,552	7.888,204	8.736,900	7.663,783	8.961,321
dez/24	8.224,562	7.800,214	8.648,910	7.575,793	8.873,331
jan/25	7.876,016	7.451,668	8.300,364	7.227,247	8.524,785
fev/25	8.249,015	7.824,667	8.673,363	7.600,246	8.897,784
mar/25	8.053,775	7.629,427	8.478,123	7.405,006	8.702,544
abr/25	8.013,211	7.588,863	8.437,559	7.364,442	8.661,980
mai/25	7.862,208	7.437,860	8.286,556	7.213,439	8.510,977
jun/25	7.927,566	7.503,218	8.351,914	7.278,797	8.576,335
jul/25	8.062,926	7.638,578	8.487,274	7.414,157	8.711,695
ago/25	8.224,883	7.800,535	8.649,231	7.576,114	8.873,652
set/25	8.575,620	8.151,272	8.999,968	7.926,851	9.224,389
out/25	8.469,580	8.045,232	8.893,928	7.820,811	9.118,349
nov/25	8.609,178	8.184,830	9.033,526	7.960,409	9.257,947
dez/25	8.427,886	8.003,538	8.852,234	7.779,117	9.076,655
jan/26	8.095,753	7.671,405	8.520,101	7.446,984	8.744,522
fev/26	8.496,126	8.071,778	8.920,474	7.847,357	9.144,895
mar/26	8.278,339	7.853,991	8.702,687	7.629,570	8.927,108
abr/26	8.267,173	7.842,825	8.691,521	7.618,404	8.915,942

Fonte: elaboração própria.

### A.3. REGIÃO SUL

Tabela A. 19. Coeficientes de Regressão Estimados para o MAPE considerando a região Sul.

	Coef	EP de Coef	Valor-T	Valor-P	VIF
$x_1$	-0,980	0,209	*	*	3,64
$x_2$	0,009	0,195	*	*	3,18
$x_3$	0,952	0,232	*	*	4,50
$x_4$	3,968	0,248	*	*	5,13
$x_1x_3$	-2,982	0,941	-3,17	0,003	3,20
$x_1x_4$	-7,637	0,991	-7,71	0,000	3,55
$x_2x_3$	-2,055	0,883	-2,33	0,024	2,82
$x_2x_4$	-7,585	0,881	-8,61	0,000	2,81
$x_1x_3(x_1 - x_3)$	4,62	1,83	2,53	0,015	1,32
$x_1x_4(x_1 - x_4)$	10,52	1,82	5,79	0,000	1,31
$x_2x_4(x_2 - x_4)$	4,03	1,82	2,22	0,031	1,31
$x_3x_4(x_3 - x_4)$	3,68	1,89	1,95	0,058	1,41
$x_1x_3^2x_4$	-58,3	14,9	-3,90	0,000	1,78
$x_1x_4(x_1 - x_4)^2$	-11,45	4,51	-2,54	0,015	1,99
$x_3x_4(x_3 - x_4)^2$	-6,85	4,31	-1,59	0,119	1,82

Fonte: elaboração própria.

Tabela A. 20. Análise de variância para o MAPE considerando a região Sul.

	GL	SQ Seq	SQ (Aj.)	QM (Aj.)	Valor F	Valor-P
Regressão	14	55,7160	55,7160	3,97971	42,73	0,000
Linear	3	32,4863	29,1593	9,71975	104,37	0,000
Quadrático	4	16,9774	12,4427	3,11066	33,40	0,000
$x_1x_3$	1	2,1574	0,9362	0,93621	10,05	0,003
$x_1x_4$	1	10,1151	5,5342	5,53417	59,42	0,000
$x_2x_3$	1	0,1586	0,5042	0,50417	5,41	0,024
$x_2x_4$	1	4,5463	6,8963	6,89631	74,05	0,000
Cúbico completo	4	4,1796	4,6893	1,17232	12,59	0,000
$x_1x_3(x_1 - x_3)$	1	0,9700	0,5942	0,59425	6,38	0,015
$x_1x_4(x_1 - x_4)$	1	2,7562	3,1183	3,11832	33,48	0,000
$x_2x_4(x_2 - x_4)$	1	0,3165	0,4584	0,45838	4,92	0,031
$x_3x_4(x_3 - x_4)$	1	0,1370	0,3527	0,35268	3,79	0,058
Quártico Especial	1	1,2526	1,4171	1,41712	15,22	0,000
$x_1x_3^2x_4$	1	1,2526	1,4171	1,41712	15,22	0,000
$x_1x_4(x_1 - x_4)^2$	2	0,8200	0,8200	0,41001	4,40	0,018
$x_3x_4(x_3 - x_4)^2$	2	0,5847	0,6005	0,60045	6,45	0,015
Erro de Resíduos	46	4,2840	4,2840	0,09313		
Total	60	60,0000				

Fonte: elaboração própria.

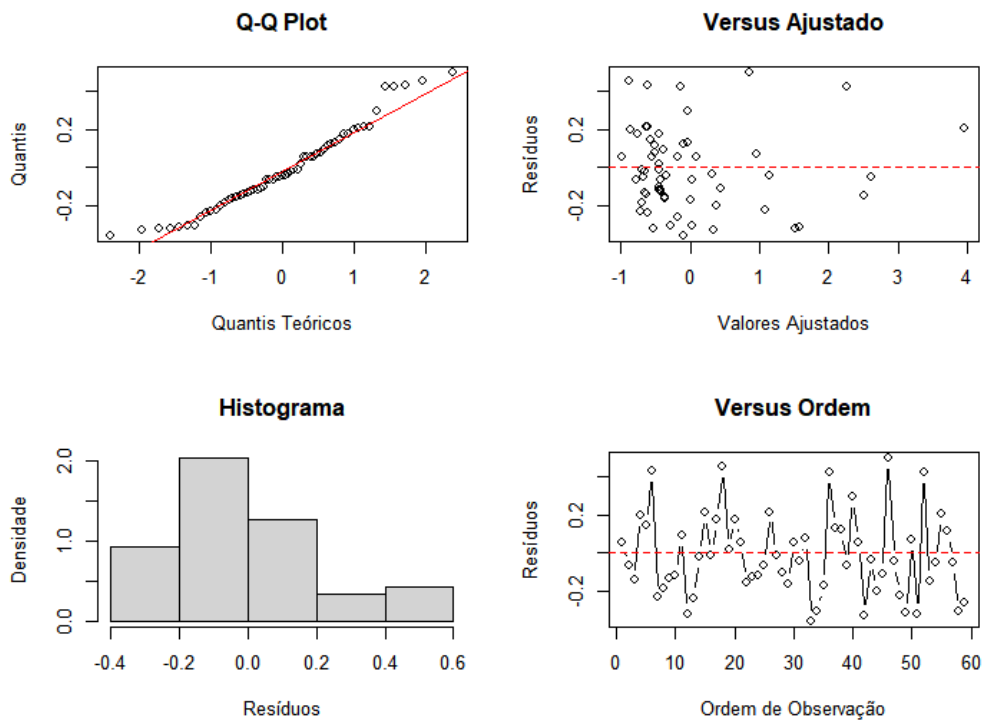


Figura A. 19. Análise de resíduos tendo o MAPE como resposta da região Sul.  
 Fonte: elaboração própria.

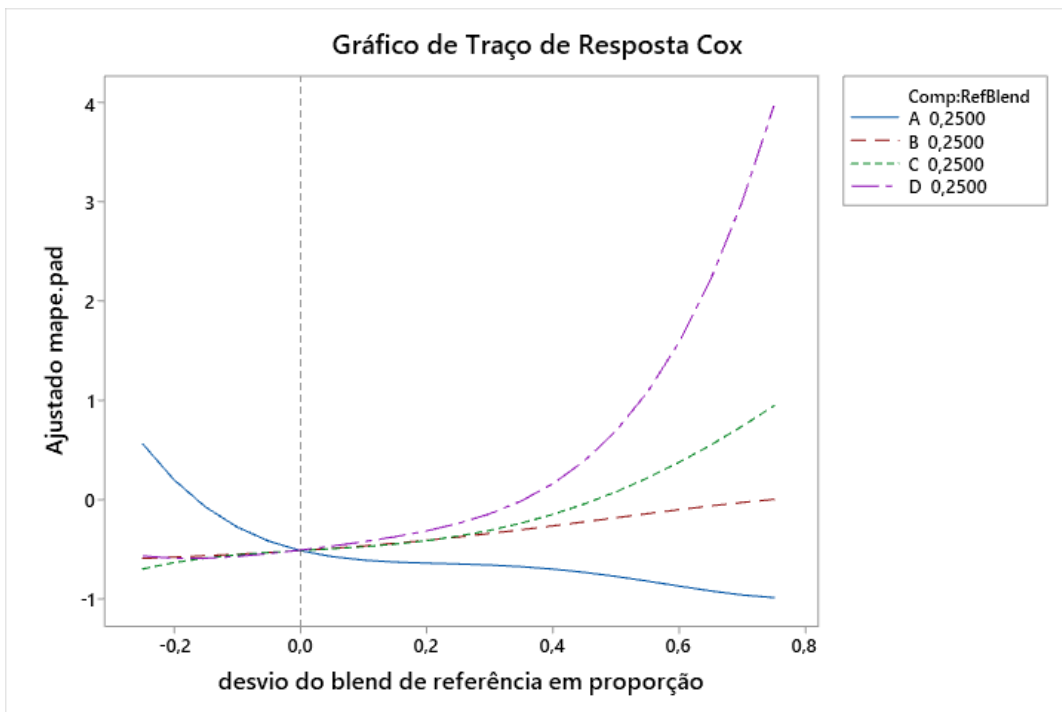


Figura A. 20. Gráfico de Traço de Resposta de Cox para o MAPE como resposta da região Sul.  
 Fonte: elaboração própria.

Tabela A. 21. Coeficientes de Regressão Estimados para o RMSE considerando a região Norte.

	<b>Coef</b>	<b>EP de Coef</b>	<b>Valor-T</b>	<b>Valor-P</b>	<b>VIF</b>
$x_1$	-0,363	0,332	*	*	3,99
$x_2$	-0,323	0,231	*	*	1,92
$x_3$	-1,280	0,228	*	*	1,89
$x_4$	-1,527	0,229	*	*	1,89
$x_1x_2$	7,93	1,47	5,40	0,000	3,37
$x_1x_3$	7,81	1,32	5,91	0,000	2,72
$x_1x_4$	9,63	1,36	7,09	0,000	2,89
$x_1^2x_2x_4$	-62,3	22,5	-2,76	0,008	1,75
$x_1x_2^2x_3$	-56,7	21,9	-2,59	0,013	1,66

Fonte: elaboração própria.

Tabela A. 22. Análise de variância para o RMSE considerando a região Sul.

	<b>GL</b>	<b>SQ Seq</b>	<b>SQ (Aj.)</b>	<b>QM (Aj.)</b>	<b>Valor F</b>	<b>Valor-P</b>
Regressão	8	48,803	48,803	6,1003	28,33	0,000
Linear	3	31,192	3,663	1,2208	5,67	0,002
Quadrático	3	14,966	17,492	5,8306	27,08	0,000
$x_1x_2$	1	1,558	6,284	6,2842	29,18	0,000
$x_1x_3$	1	4,445	7,528	7,5283	34,96	0,000
$x_1x_4$	1	8,962	10,829	10,8287	50,29	0,000
Quártico Especial	2	2,645	2,645	1,3224	6,14	0,004
$x_1^2x_2x_4$	1	1,204	1,644	1,6444	7,64	0,008
$x_1x_2^2x_3$	1	1,441	1,441	1,4407	6,69	0,013
Erro de Resíduos	52	11,197	11,197	0,2153		
Total	60	60,000				

Fonte: elaboração própria.

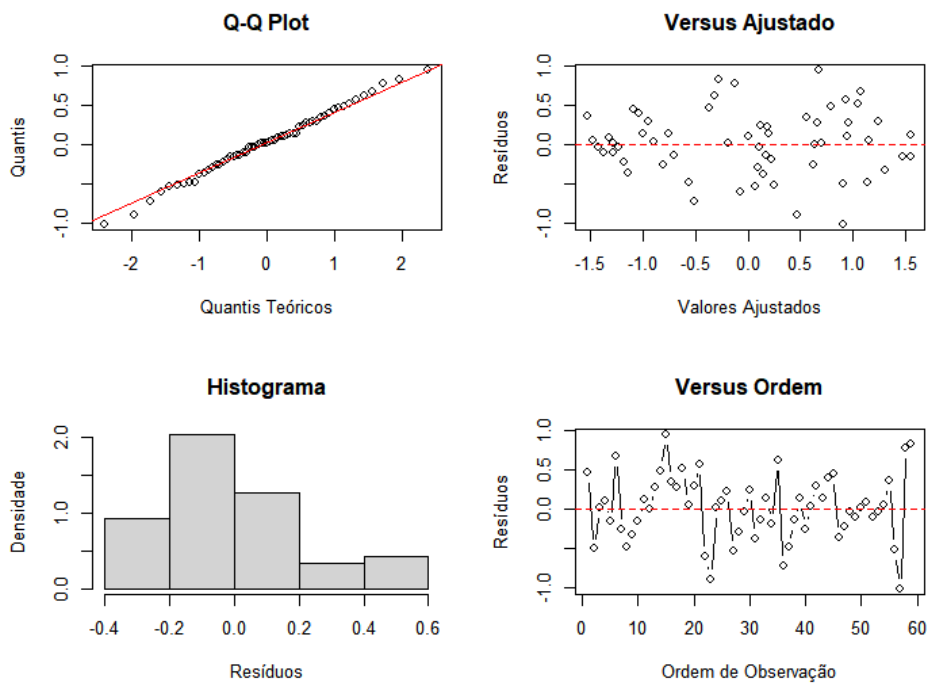


Figura A. 21. Análise de resíduos tendo o RMSE como resposta da região Sul.  
 Fonte: elaboração própria.

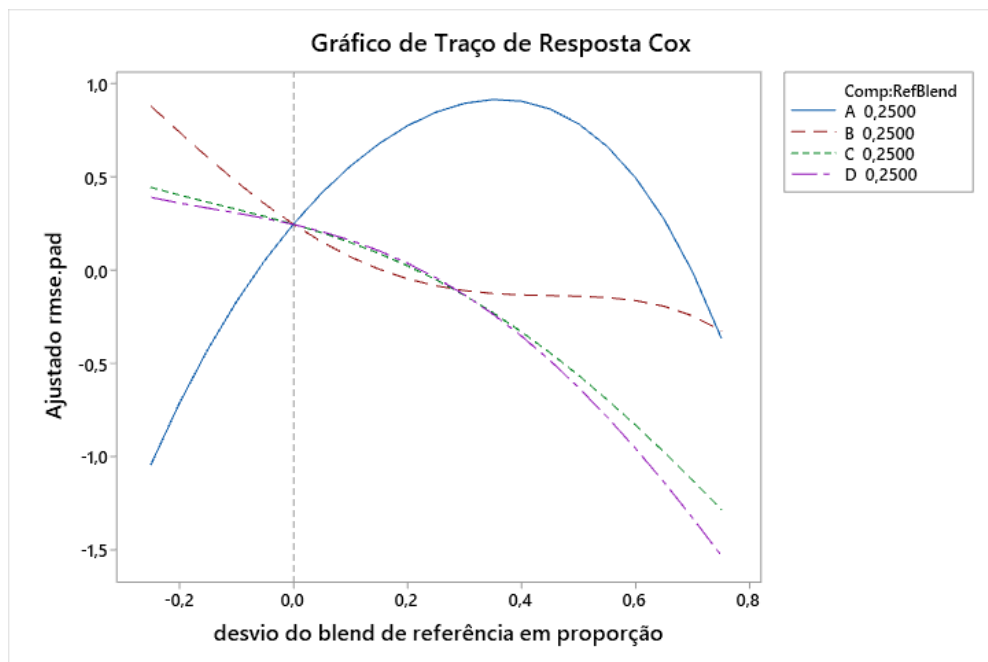


Figura A. 22. Gráfico de Traço de Resposta de Cox para o RMSE como resposta da região Sul.  
 Fonte: elaboração própria.

Tabela A. 23. Coeficientes de Regressão Estimados para o MdAE considerando a região Sul.

	<b>Coef</b>	<b>EP de Coef</b>	<b>Valor-T</b>	<b>Valor-P</b>	<b>VIF</b>
$x_1$	-1,308	0,462	*	*	3,97
$x_2$	-1,173	0,332	*	*	2,05
$x_3$	1,527	0,393	*	*	2,87
$x_4$	2,487	0,392	*	*	2,85
$x_1x_2$	7,27	2,05	3,54	0,001	3,38
$x_3x_4$	-3,92	1,96	-2,00	0,051	3,09
$x_1x_3(x_1 - x_3)$	15,28	3,86	3,96	0,000	1,31
$x_1x_4(x_1 - x_4)$	12,40	3,95	3,14	0,003	1,38
$x_1x_2^2x_1$	-67,9	29,2	-2,33	0,024	1,51
$x_1x_2^2x_4$	-106,6	29,2	-3,65	0,001	1,51
$x_1x_3x_4^2$	-70,7	29,8	-2,38	0,022	1,57
$x_1x_3(x_1 - x_3)^2$	-19,00	8,71	-2,18	0,034	1,66
$x_1x_4(x_1 - x_4)^2$	-23,05	8,71	-2,65	0,011	1,65

Fonte: elaboração própria.

Tabela A. 24. Análise de variância para o MdAE considerando a região Sul.

	<b>GL</b>	<b>SQ Seq</b>	<b>SQ (Aj.)</b>	<b>QM (Aj.)</b>	<b>Valor F</b>	<b>Valor-P</b>
Regressão	12	39,9230	39,923	3,3269	7,95	0,000
Linear	3	10,5550	18,615	6,2048	14,83	0,000
Quadrático	2	1,6621	6,079	3,0396	7,27	0,002
$x_1x_2$	1	1,5122	5,253	5,2530	12,56	0,001
$x_3x_4$	1	0,1499	1,671	1,6710	4,00	0,051
Cúbico completo	2	13,6613	10,355	5,1775	12,38	0,000
$x_1x_3(x_1 - x_3)$	1	7,6207	6,556	6,5558	15,67	0,000
$x_1x_4(x_1 - x_4)$	1	6,0406	4,115	4,1149	9,84	0,003
Quártico Especial	3	9,6781	10,040	3,3466	8,00	0,000
$x_1x_2^2x_1$	1	1,6492	2,264	2,2643	5,41	0,024
$x_1x_2^2x_4$	1	5,6821	5,586	5,5855	13,35	0,001
$x_1x_3x_4^2$	1	2,3468	2,361	2,3608	5,64	0,022
Quártico1 completo	2	4,3665	4,366	2,1832	5,22	0,009
$x_1x_3(x_1 - x_3)^2$	1	1,4365	1,989	1,9886	4,75	0,034
$x_1x_4(x_1 - x_4)^2$	1	2,9300	2,930	2,9300	7,00	0,011
Erro de Resíduos	48	20,0770	20,077	0,4183		
Total	60	60,0000				

Fonte: elaboração própria.



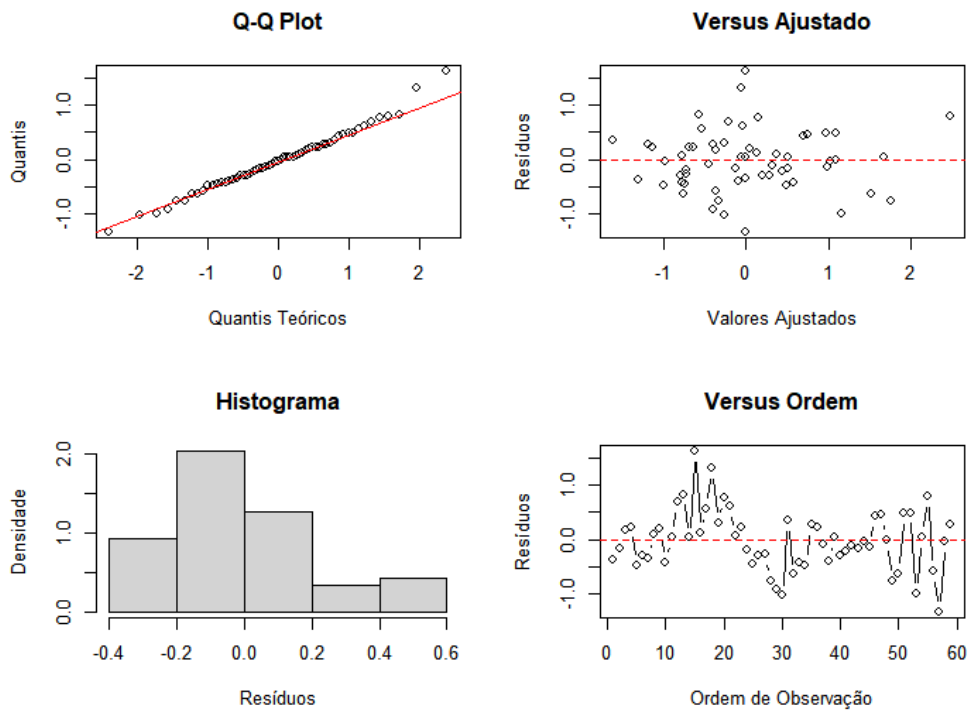


Figura A. 23. Análise de resíduos tendo o MdAE como resposta da região Sul.  
 Fonte: elaboração própria.

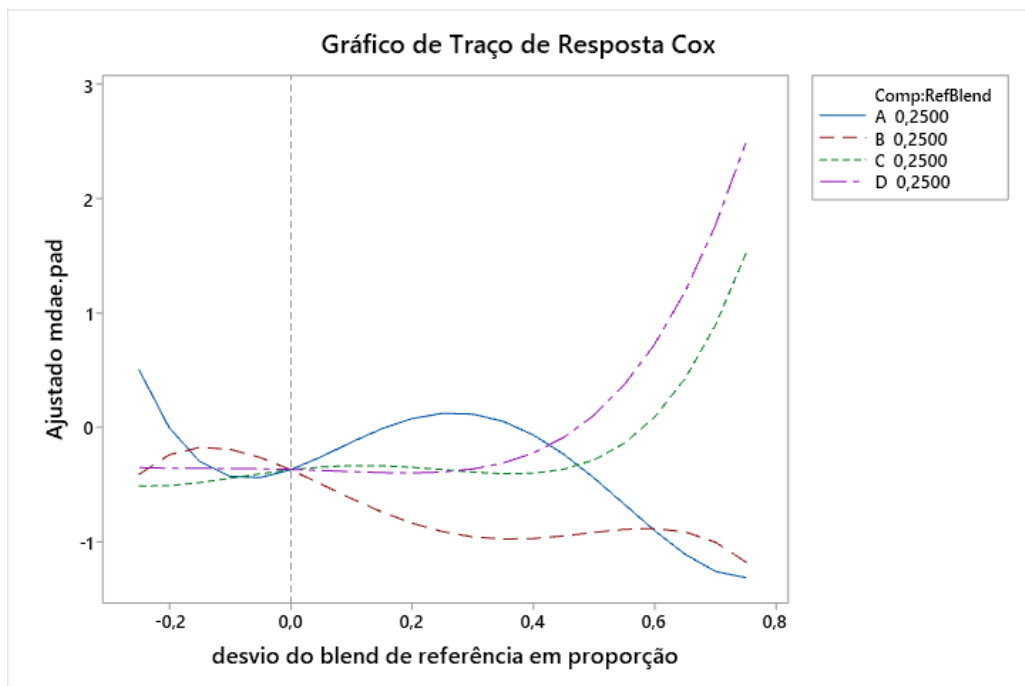


Figura A. 24. Gráfico de Traço de Resposta de Cox para o MdAE como resposta da região Sul.  
 Fonte: elaboração própria.

Tabela A. 25. Coeficientes de Regressão Estimados para o MAE considerando a região Sul.

	<b>Coef</b>	<b>EP de Coef</b>	<b>Valor-T</b>	<b>Valor-P</b>	<b>VIF</b>
$x_1$	-0,480	0,379	*	*	3,99
$x_2$	-0,269	0,292	*	*	2,36
$x_3$	-1,365	0,268	*	*	1,99
$x_4$	-0,974	0,289	*	*	2,32
$x_1x_2$	8,61	1,70	5,07	0,000	3,47
$x_1x_3$	7,93	1,51	5,26	0,000	2,74
$x_1x_4$	8,78	1,58	5,56	0,000	3,00
$x_1^2x_2x_4$	-77,9	25,7	-3,03	0,004	1,76
$x_1x_2^2x_3$	-71,3	25,0	-2,85	0,006	1,66
$x_2x_4(x_2 - x_4)^2$	-13,27	7,13	-1,86	0,068	1,65

Fonte: elaboração própria.

Tabela A. 26. Análise de variância para o MAE considerando a região Sul.

	<b>GL</b>	<b>SQ Seq</b>	<b>SQ (Aj.)</b>	<b>QM (Aj.)</b>	<b>Valor F</b>	<b>Valor-P</b>
Regressão	9	45,7302	45,7302	5,0811	18,16	0,000
Linear	3	26,7061	2,1817	0,7272	2,60	0,062
Quadrático	3	13,8111	16,6103	5,5368	19,79	0,000
$x_1x_2$	1	2,2345	7,2032	7,2032	25,74	0,000
$x_1x_3$	1	4,1542	7,7308	7,7308	27,63	0,000
$x_1x_4$	1	7,4224	8,6567	8,6567	30,94	0,000
Quártico Especial	2	4,2418	4,1528	2,0764	7,42	0,001
$x_1^2x_2x_4$	1	2,0571	2,5604	2,5604	9,15	0,004
$x_1x_2^2x_3$	1	2,1847	2,2749	2,2749	8,13	0,006
Quártico1 completo	1	0,9711	0,9711	0,9711	3,47	0,068
$x_2x_4(x_2 - x_4)^2$	1	0,9711	0,9711	0,9711	3,47	0,068
Erro de Resíduos	51	14,2698	14,2698	0,2798		
Total	60	60,0000				

Fonte: elaboração própria.

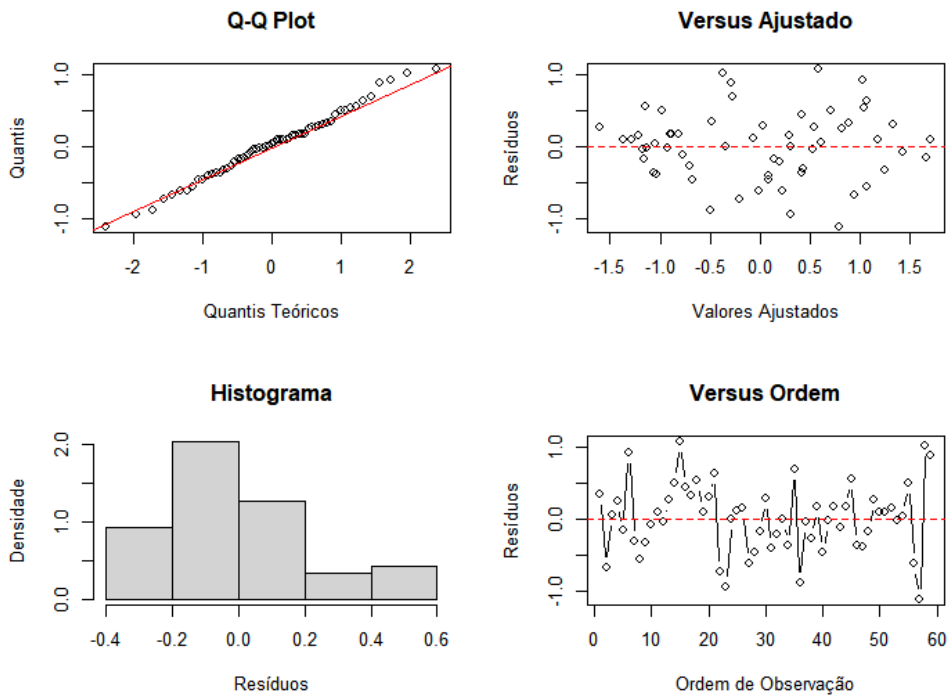


Figura A. 25. Análise de resíduos tendo o MAE como resposta da região Sul.  
 Fonte: elaboração própria.

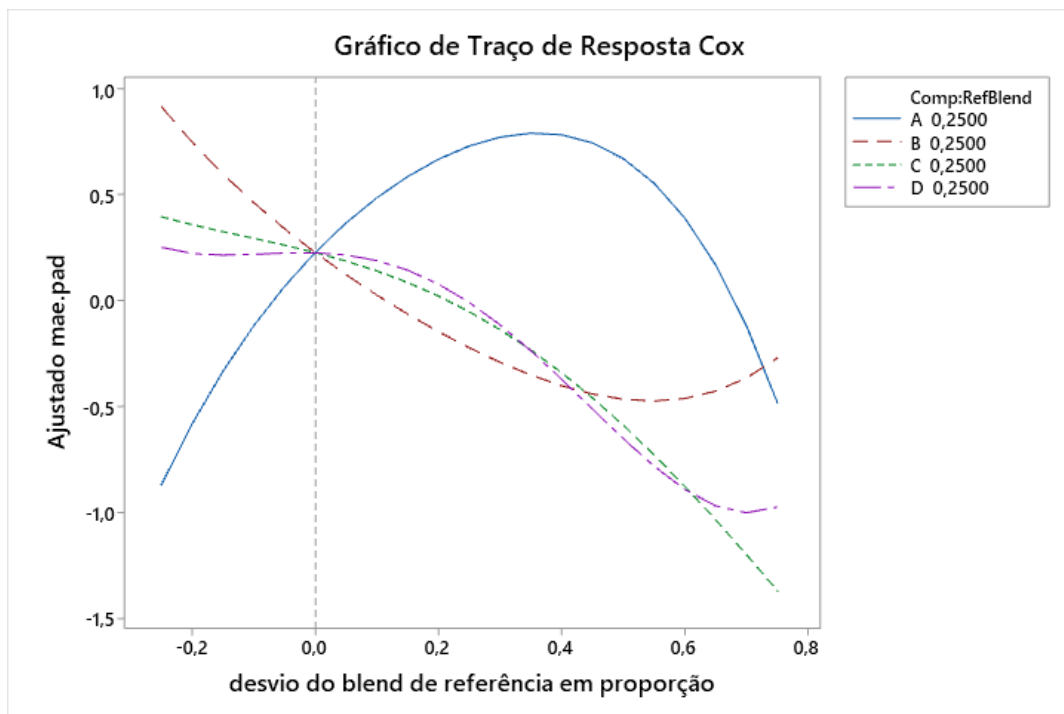


Figura A. 26. Gráfico de Traço de Resposta de Cox para o MAE como resposta da região Sul.  
 Fonte: elaboração própria.

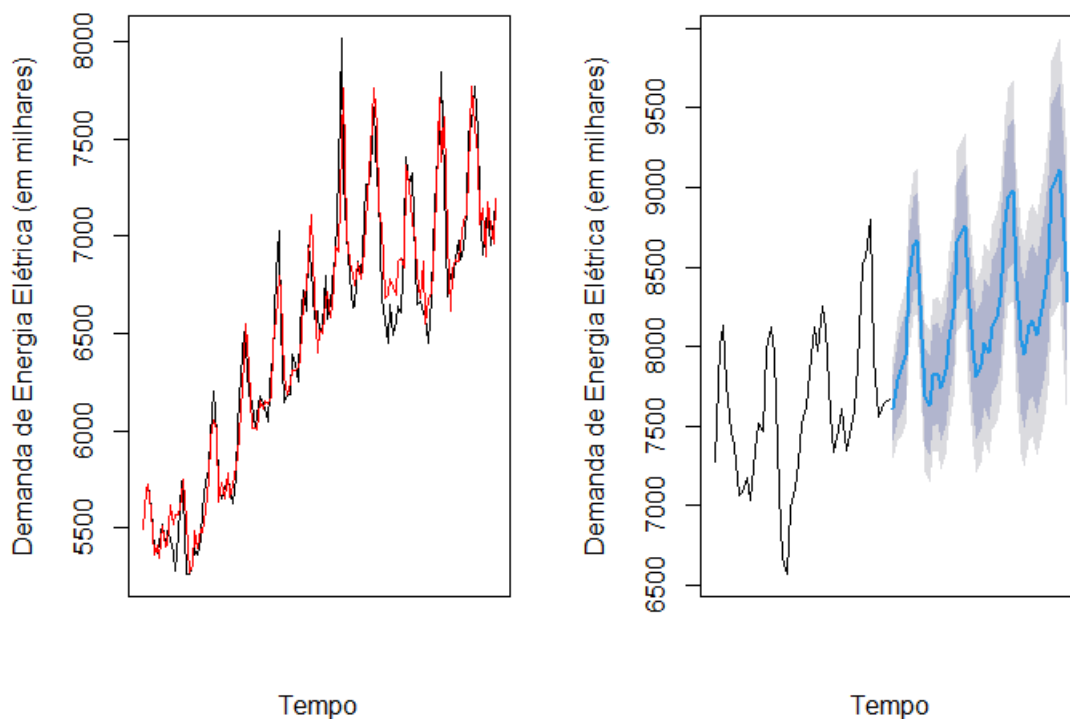


Figura A. 27. Gráfico de ajuste e previsão aplicando a metodologia proposta para a região Sul.  
 Fonte: elaboração própria.

Tabela A. 27. Previsão e intervalos de confiança da demanda de energia elétrica na Região Sul.

<b>Tempo</b>	<b>Previsão</b>	<b>80%<sub>inf</sub></b>	<b>80%<sub>sup</sub></b>	<b>95%<sub>inf</sub></b>	<b>95%<sub>sup</sub></b>
ago/22	7.553,402	6.959,701	8147,103	6.645,715	8.461,089
set/22	7.662,768	7.069,067	8256,469	6.755,081	8.570,455
out/22	7.806,688	7.212,987	8400,389	6.899,001	8.714,375
nov/22	8.055,215	7.461,514	8648,916	7.147,528	8.962,902
dez/22	8.585,322	7.991,621	9179,023	7.677,635	9.493,009
jan/23	8.670,899	8.077,198	9264,600	7.763,212	9.578,586
fev/23	8.840,306	8.246,605	9434,007	7.932,619	9.747,993
mar/23	8.103,755	7.510,054	8697,456	7.196,068	9.011,442
abr/23	7.745,711	7.152,010	8339,412	6.838,024	8.653,398
mai/23	7.727,072	7.133,371	8320,773	6.819,385	8.634,759
jun/23	7.773,751	7.180,050	8367,452	6.866,064	8.681,438
jul/23	7.851,177	7.257,476	8444,878	6.943,490	8.758,864
ago/23	7.695,269	7.101,568	8288,970	6.787,582	8.602,956
set/23	7.802,709	7.209,008	8396,410	6.895,022	8.710,396
out/23	7.969,158	7.375,457	8562,859	7.061,471	8.876,845
nov/23	8.190,150	7.596,449	8783,851	7.282,463	9.097,837
dez/23	8.726,234	8.132,533	9319,935	7.818,547	9.633,921
jan/24	8.824,967	8.231,266	9418,668	7.917,280	9.732,654
fev/24	8.967,590	8.373,889	9561,291	8.059,903	9.875,277
mar/24	8.252,349	7.658,648	8846,050	7.344,662	9.160,036
abr/24	7.901,403	7.307,702	8495,104	6.993,716	8.809,090
mai/24	7.855,958	7.262,257	8449,659	6.948,271	8.763,645
jun/24	7.921,685	7.327,984	8515,386	7.013,998	8.829,372

jul/24	8.001,641	7.407,940	8595,342	7.093,954	8.909,328
ago/24	7.825,555	7.231,854	8419,256	6.917,868	8.733,242
set/24	7.955,240	7.361,539	8548,941	7.047,553	8.862,927
out/24	8.114,218	7.520,517	8707,919	7.206,531	9.021,905
nov/24	8.320,895	7.727,194	8914,596	7.413,208	9.228,582
dez/24	8.880,006	8.286,305	9473,707	7.972,319	9.787,693
jan/25	8.967,937	8.374,236	9561,638	8.060,250	9.875,624
fev/25	9.103,528	8.509,827	9697,229	8.195,841	10.011,215
mar/25	8.405,512	7.811,811	8999,213	7.497,825	9.313,199
abr/25	8.039,088	7.445,387	8632,789	7.131,401	8.946,775
mai/25	7.993,443	7.399,742	8587,144	7.085,756	8.901,130
jun/25	8.076,781	7.483,080	8670,482	7.169,094	8.984,468
jul/25	8.137,731	7.544,030	8731,432	7.230,044	9.045,418
ago/25	7.965,200	7.371,499	8558,901	7.057,513	8.872,887
set/25	8.106,813	7.513,112	8700,514	7.199,126	9.014,500
out/25	8.248,909	7.655,208	8842,610	7.341,222	9.156,596
nov/25	8.465,055	7.871,354	9058,756	7.557,368	9.372,742
dez/25	9.031,491	8.437,790	9625,192	8.123,804	9.939,178
jan/26	9.101,702	8.508,001	9695,403	8.194,015	10.009,389
fev/26	9.249,521	8.655,820	9843,222	8.341,834	10.157,208
mar/26	8.554,714	7.961,013	9148,415	7.647,027	9.462,401
abr/26	8.173,951	7.580,250	8767,652	7.266,264	9.081,638

Fonte: elaboração própria.

#### A.4. REGIÃO CENTRO-OESTE

Tabela A. 28. Coeficientes de Regressão Estimados para o MAPE considerando a região Centro-Oeste.

	<b>Coef</b>	<b>EP de Coef</b>	<b>Valor-T</b>	<b>Valor-P</b>	<b>VIF</b>
$x_1$	-2,318	0,208	*	*	1,55
$x_2$	1,547	0,318	*	*	3,61
$x_3$	0,870	0,243	*	*	2,11
$x_4$	1,950	0,245	*	*	2,14
$x_2x_3$	-5,43	1,49	-3,64	0,001	3,44
$x_2x_4$	-6,13	1,53	-4,00	0,000	3,64
$x_2x_3(x_2 - x_3)$	-11,49	3,22	-3,57	0,001	1,75
$x_2x_4(x_2 - x_4)$	-7,52	2,83	-2,65	0,011	1,36
$x_1^2x_2x_3$	135,8	31,6	4,30	0,000	3,39
$x_2^2x_3x_4$	140,3	33,7	4,16	0,000	3,87
$x_1x_2^2x_3$	-77,1	33,8	-2,28	0,027	3,90
$x_2x_3^2x_4$	-101,8	31,5	-3,23	0,002	3,39
$x_1x_2x_3^2$	-86,3	33,8	-2,55	0,014	3,90
$x_2x_4(x_2 - x_4)^2$	-12,48	6,81	-1,83	0,073	1,94

Fonte: elaboração própria.

Tabela A. 29. Análise de variância para o MAPE considerando a região Centro-Oeste.

	<b>GL</b>	<b>SQ Seq</b>	<b>SQ (Aj.)</b>	<b>QM (Aj.)</b>	<b>Valor F</b>	<b>Valor-P</b>
Regressão	13	49,7744	49,7744	3,8288	17,60	0,000
Linear	3	32,8594	36,2390	12,0797	55,52	0,000
Quadrático	2	5,3961	5,2454	2,6227	12,05	0,000
$x_2x_3$	1	2,3335	2,8879	2,8879	13,27	0,001
$x_2x_4$	1	3,0625	3,4849	3,4849	16,02	0,000
Cúbico completo	2	1,6503	4,0332	2,0166	9,27	0,000
$x_2x_3(x_2 - x_3)$	1	1,0568	2,7705	2,7705	12,73	0,001
$x_2x_4(x_2 - x_4)$	1	0,5935	1,5335	1,5335	7,05	0,011
Quártico Especial	5	9,1381	9,0516	1,8103	8,32	0,000
$x_1^2x_2x_3$	1	0,2337	4,0275	4,0275	18,51	0,000
$x_2^2x_3x_4$	1	2,4453	3,7664	3,7664	17,31	0,000
$x_1x_2^2x_3$	1	3,1416	1,1282	1,1282	5,19	0,027
$x_2x_3^2x_4$	1	1,9107	2,2681	2,2681	10,43	0,002
$x_1x_2x_3^2$	1	1,4068	1,4177	1,4177	6,52	0,014
Quártico1 completo	1	0,7305	0,7305	0,7305	3,36	0,073
$x_2x_4(x_2 - x_4)^2$	1	0,7305	0,7305	0,7305	3,36	0,073
Erro de Resíduos	47	10,2256	10,2256	0,2176		
Total	60	60,0000				

Fonte: elaboração própria.

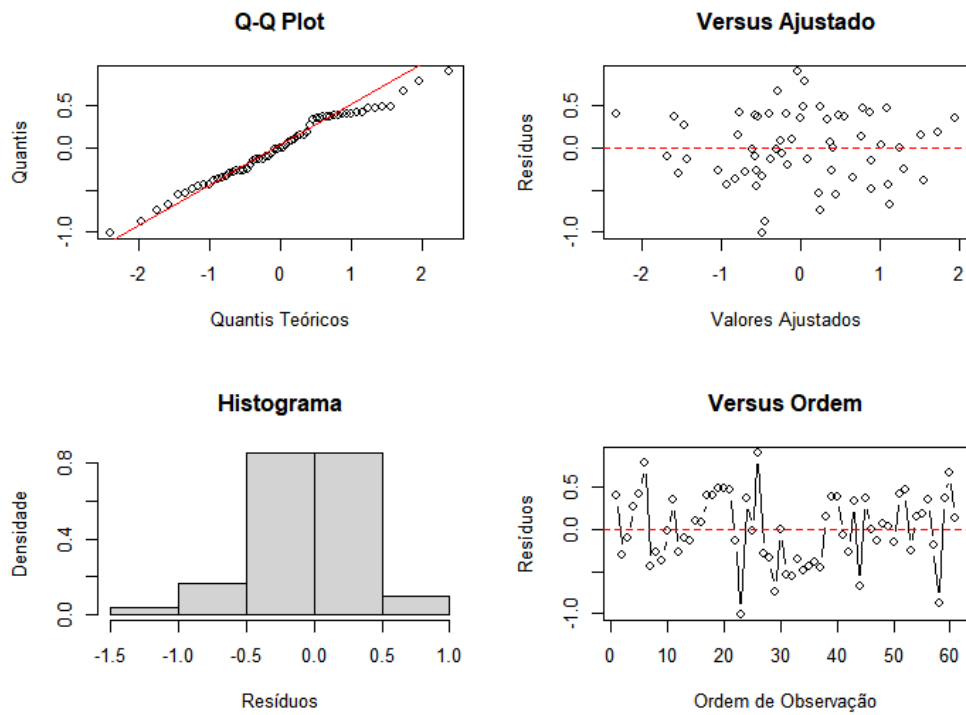


Figura A. 28. Análise de resíduos tendo o MAPE como resposta da região Centro-Oeste.  
Fonte: elaboração própria.

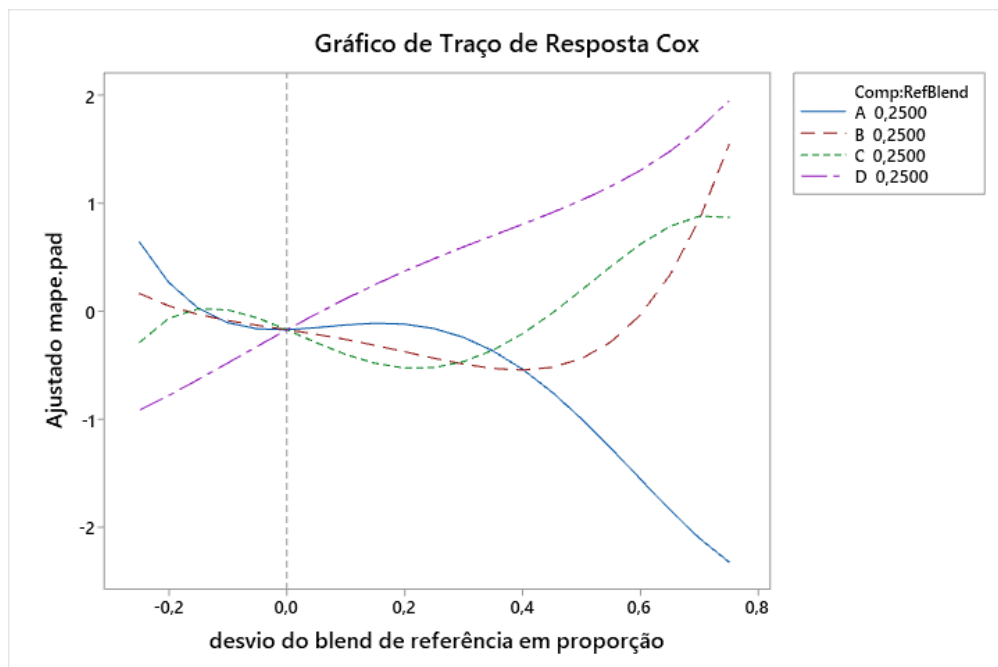


Figura A. 29. Gráfico de Traço de Resposta de Cox para o MAPE como resposta da região Centro-Oeste.  
Fonte: elaboração própria.

Tabela A. 30. Coeficientes de Regressão Estimados para o RMSE considerando a região Centro-Oeste.

	<b>Coef</b>	<b>EP de Coef</b>	<b>Valor-T</b>	<b>Valor-P</b>	<b>VIF</b>
$x_1$	1,416	0,349	*	*	1,89
$x_2$	-0,877	0,372	*	*	2,13
$x_3$	-0,896	0,305	*	*	1,43
$x_4$	-0,647	0,292	*	*	1,31
$x_1x_2$	4,75	1,90	2,49	0,016	2,42
$x_2^2x_3x_4$	116,7	43,8	2,66	0,010	2,83
$x_2x_3^2x_4$	-74,6	43,6	-1,71	0,092	2,79

Fonte: elaboração própria.

Tabela A. 31. Análise de variância para o RMSE considerando a região Centro-Oeste.

	<b>GL</b>	<b>SQ Seq</b>	<b>SQ (Aj.)</b>	<b>QM (Aj.)</b>	<b>Valor F</b>	<b>Valor-P</b>
Regressão	6	32,798	32,798	5,4664	10,85	0,000
Linear	3	26,825	15,951	5,3170	10,56	0,000
Quadrático	1	2,356	3,133	3,1326	6,22	0,016
$x_1x_2$	1	2,356	3,133	3,1326	6,22	0,016
Quártico Especial	2	3,617	3,617	1,8086	3,59	0,034
$x_2^2x_3x_4$	1	2,139	3,568	3,5679	7,08	0,010
$x_2x_3^2x_4$	1	1,479	1,479	1,4786	2,94	0,092
Erro de Resíduos	54	27,202	27,202	0,5037		
Total	60	60,000				

Fonte: elaboração própria.



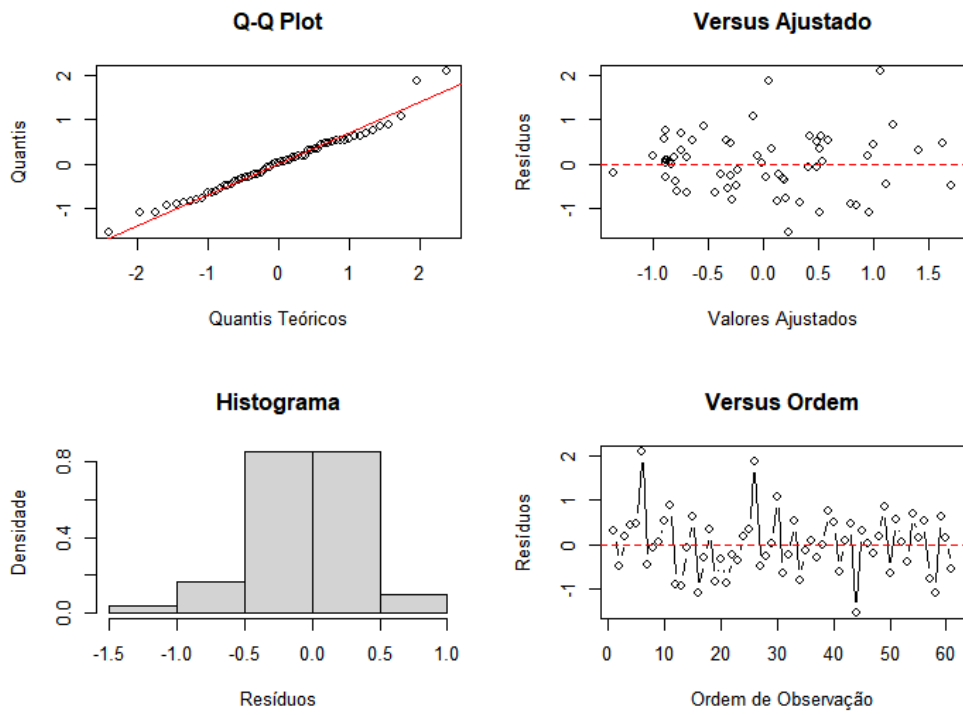


Figura A. 30. Análise de resíduos tendo o RMSE como resposta da região Centro-Oeste.  
Fonte: elaboração própria.

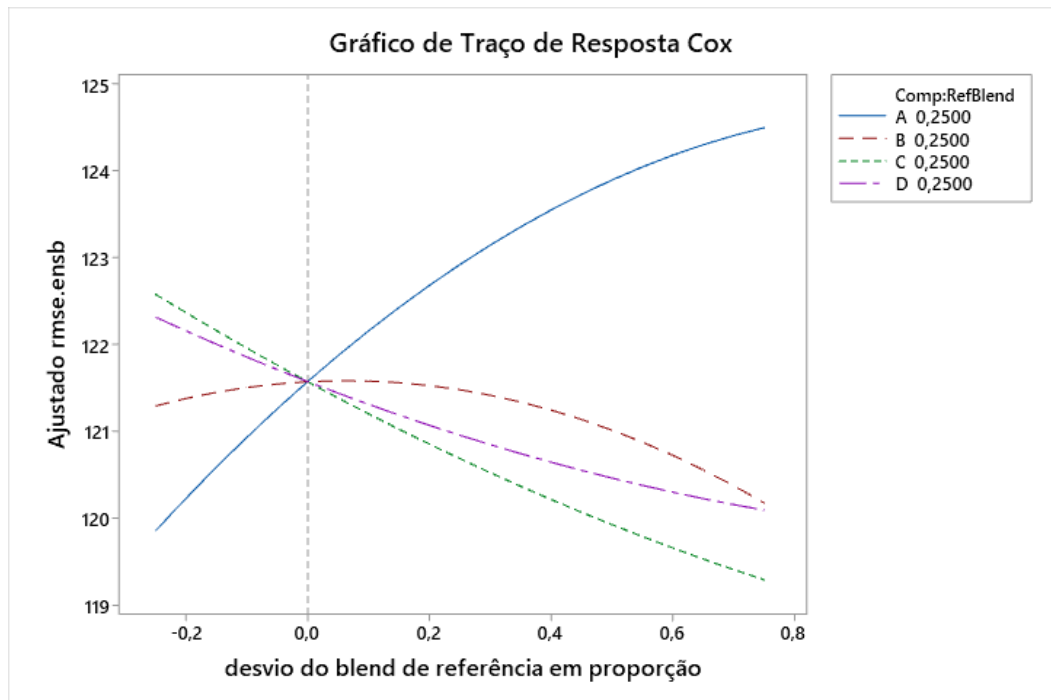


Figura A. 31. Gráfico de Traço de Resposta de Cox para o RMSE como resposta da região Centro-Oeste.  
Fonte: elaboração própria.

Tabela A. 32. Coeficientes de Regressão Estimados para o MdAE considerando a região Centro-Oeste.

	Coef	EP de Coef	Valor-T	Valor-P	VIF
$x_1$	-0,871	0,523	*	*	5,65
$x_2$	-1,204	0,321	*	*	2,12
$x_3$	1,563	0,334	*	*	2,29
$x_4$	1,697	0,343	*	*	2,42
$x_1x_3$	6,86	1,98	3,46	0,001	3,51
$x_1x_3$	-4,09	1,95	-2,10	0,041	3,39
$x_1x_4$	-3,15	2,06	-1,53	0,133	3,79
$x_1x_2(x_1 - x_2)$	6,73	3,77	1,79	0,080	1,39
$x_1x_4(x_1 - x_4)$	11,76	3,76	3,13	0,003	1,38
$x_3x_4(x_3 - x_4)$	-11,59	3,69	-3,15	0,003	1,33
$x_1^2x_2x_4$	-89,4	31,7	-2,82	0,007	1,98
$x_1x_2x_3^2$	-66,3	29,2	-2,27	0,028	1,68
$x_1x_3x_4^2$	-50,6	29,9	-1,69	0,097	1,76

Fonte: elaboração própria.

Tabela A. 33. Análise de variância para o MdAE considerando a região Centro-Oeste.

	GL	SQ Seq	SQ (Aj.)	QM (Aj.)	Valor F	Valor-P
Regressão	12	41,8799	41,8799	3,4900	9,24	0,000
Linear	3	14,6202	19,7558	6,5853	17,44	0,000
Quadrático	3	16,1930	10,4821	3,4940	9,26	0,000
$x_1x_3$	1	2,7474	4,5128	4,5128	11,95	0,001
$x_1x_3$	1	4,9203	1,6595	1,6595	4,40	0,041
$x_1x_4$	1	8,5252	0,8804	0,8804	2,33	0,133
Cúbico completo	3	6,3207	8,4757	2,8252	7,48	0,000
$x_1x_2(x_1 - x_2)$	1	0,3735	1,2039	1,2039	3,19	0,080
$x_1x_4(x_1 - x_4)$	1	2,7318	3,6906	3,6906	9,78	0,003
$x_3x_4(x_3 - x_4)$	1	3,2153	3,7353	3,7353	9,89	0,003
Quártico Especial	3	4,7460	4,7460	1,5820	4,19	0,010
$x_1^2x_2x_4$	1	2,0993	2,9957	2,9957	7,94	0,007
$x_1x_2x_3^2$	1	1,5656	1,9411	1,9411	5,14	0,028
$x_1x_3x_4^2$	1	1,0811	1,0811	1,0811	2,86	0,097
Erro de Resíduos	48	18,1201	18,1201	0,3775		
Total	60	60,0000				

Fonte: elaboração própria.

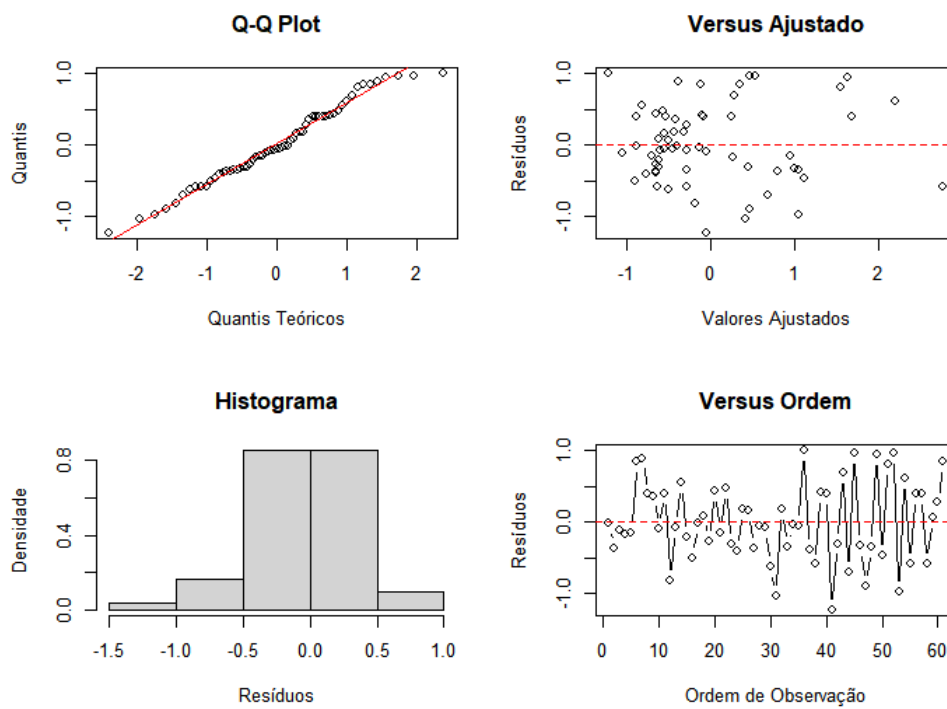


Figura A. 32. Análise de resíduos tendo o MdAE como resposta da região Centro-Oeste.  
 Fonte: elaboração própria.

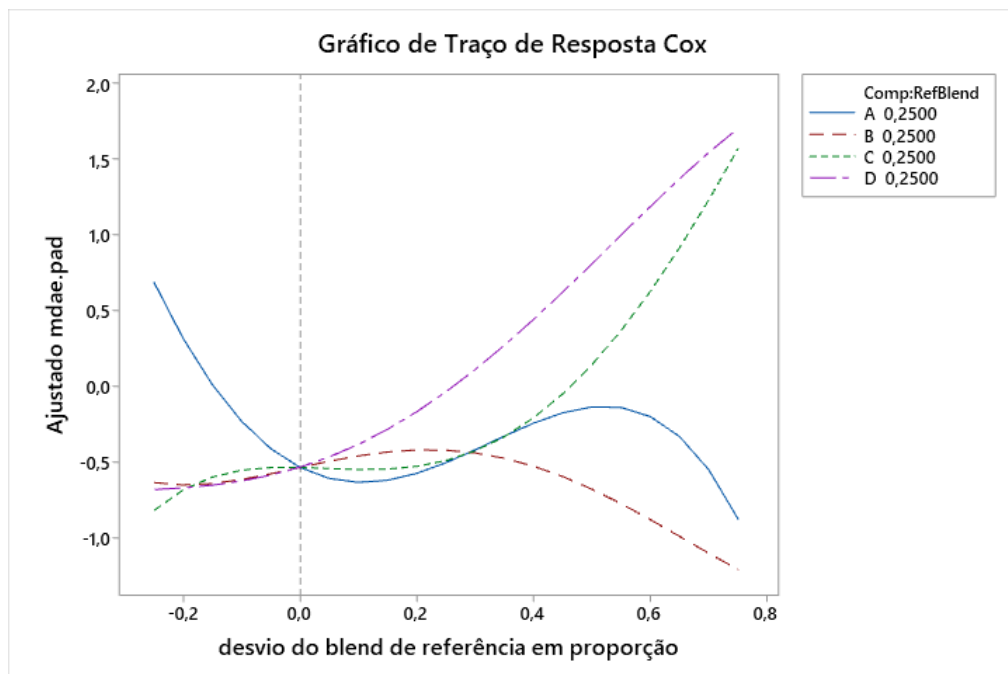


Figura A. 33. Gráfico de Traço de Resposta de Cox para o MdAE como resposta da região Centro-Oeste.  
 Fonte: elaboração própria.

Tabela A. 34. Coeficientes de Regressão Estimados para o MAE considerando a região Centro-Oeste.

	<b>Coef</b>	<b>EP de Coef</b>	<b>Valor-T</b>	<b>Valor-P</b>	<b>VIF</b>
$x_1$	2,039	0,308	*	*	2,04
$x_2$	-0,860	0,307	*	*	2,03
$x_3$	-0,911	0,250	*	*	1,35
$x_4$	-0,756	0,244	*	*	1,28
$x_1x_2$	4,39	1,68	2,61	0,012	2,64
$x_1x_2(x_1 - x_2)$	-5,57	3,51	-1,59	0,119	1,26
$x_1^2x_3x_3$	-42,2	27,0	-1,56	0,124	1,50

Fonte: elaboração própria.

Tabela A. 35. Análise de variância para o MAE considerando a região Centro-Oeste.

	<b>GL</b>	<b>SQ Seq</b>	<b>SQ (Aj.)</b>	<b>QM (Aj.)</b>	<b>Valor F</b>	<b>Valor-P</b>
Regressão	6	40,4908	40,4908	6,7485	18,68	0,000
Linear	3	36,5801	24,5926	8,1975	22,69	0,000
Quadrático	1	1,7986	2,4579	2,4579	6,80	0,012
$x_1x_2$	1	1,7986	2,4579	2,4579	6,80	0,012
Cúbico completo	1	1,2308	0,9077	0,9077	2,51	0,119
$x_1x_2(x_1 - x_2)$	1	1,2308	0,9077	0,9077	2,51	0,119
Quártico Especial	1	0,8813	0,8813	0,8813	2,44	0,124
$x_1^2x_3x_3$	1	0,8813	0,8813	0,8813	2,44	0,124
Erro de Resíduos	54	19,5092	19,5092	0,3613		
Total	60	60,0000				

Fonte: elaboração própria.

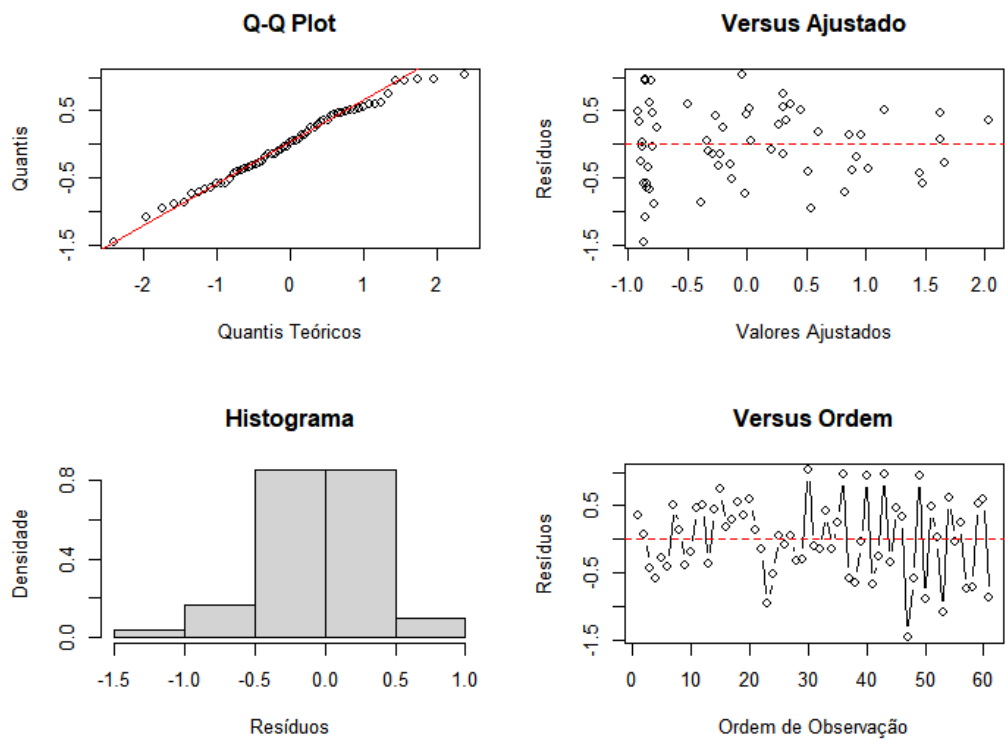


Figura A. 34. Análise de resíduos tendo o MAE como resposta da região Centro-Oeste.  
 Fonte: elaboração própria.

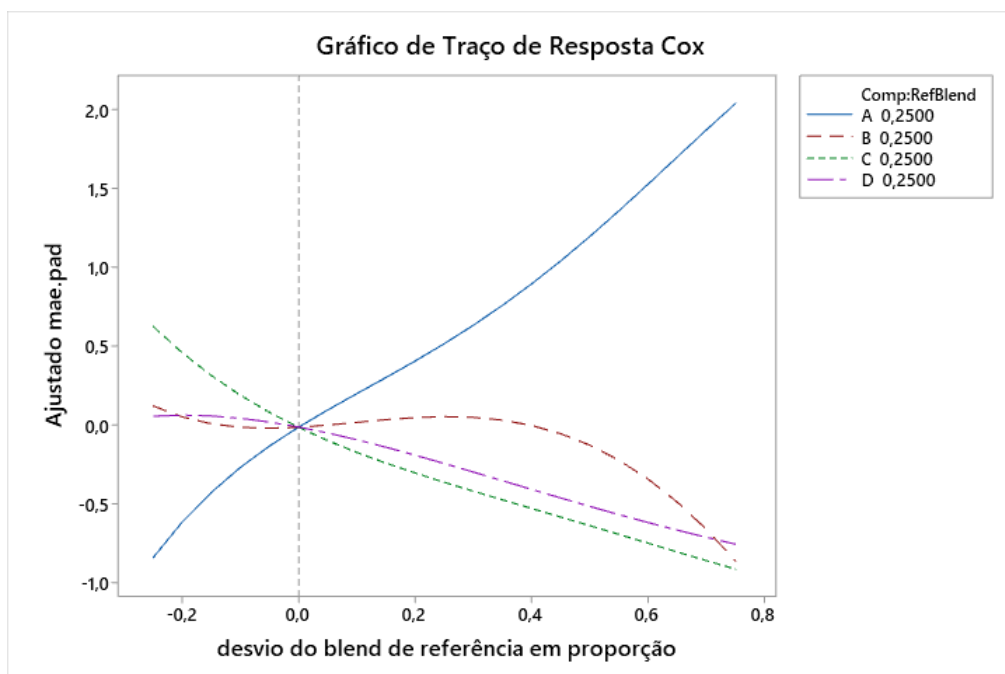


Figura A. 35. Gráfico de Traço de Resposta de Cox para o MAE como resposta da região Centro-Oeste.  
 Fonte: elaboração própria.

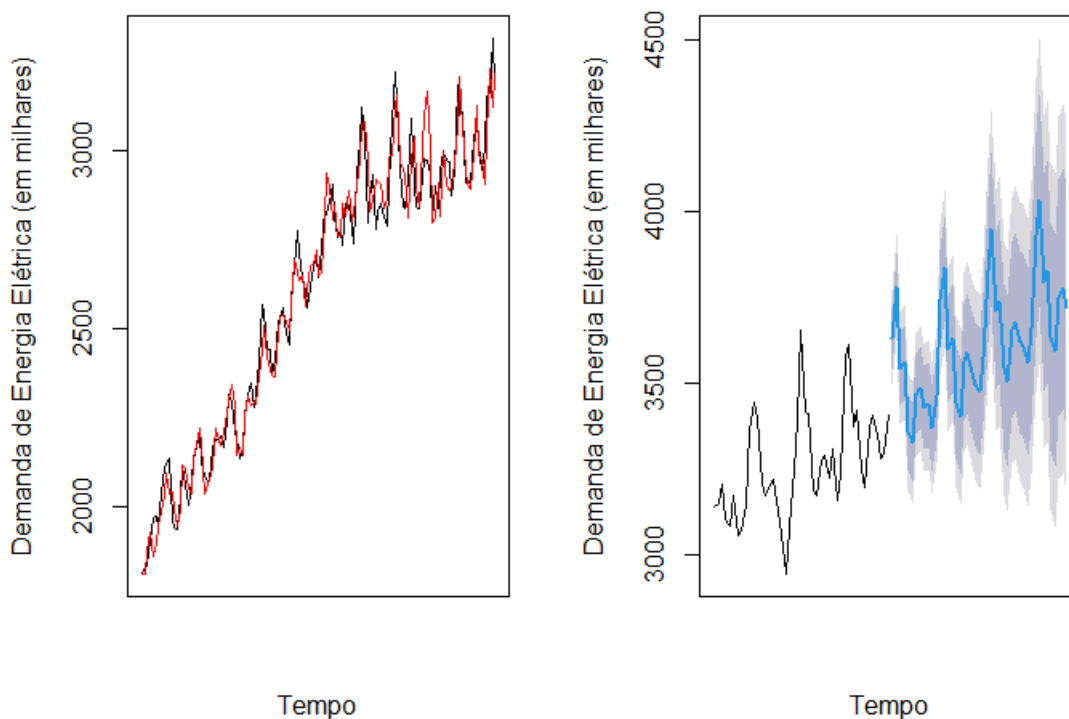


Figura A. 36. Gráfico de ajuste e previsão aplicando a metodologia proposta para a Região Centro-Oeste.  
Fonte: elaboração própria.

Tabela A. 36. Previsão e intervalos de confiança da demanda de energia elétrica na Região Centro-Oeste.

<b>Tempo</b>	<b>Previsão</b>	<b>80%<sub>inf</sub></b>	<b>80%<sub>sup</sub></b>	<b>95%<sub>inf</sub></b>	<b>95%<sub>sup</sub></b>
ago/22	3.614,083	3.427,021	3.801,145	3.328,091	3.900,075
set/22	3.704,813	3.517,751	3.891,875	3.418,821	3.990,805
out/22	3.506,948	3.319,886	3.694,010	3.220,956	3.792,940
nov/22	3.507,556	3.320,494	3.694,618	3.221,564	3.793,548
dez/22	3.378,145	3.191,083	3.565,207	3.092,153	3.664,137
jan/23	3.347,282	3.160,220	3.534,344	3.061,290	3.633,274
fev/23	3.451,901	3.264,839	3.638,963	3.165,909	3.737,893
mar/23	3.480,548	3.293,486	3.667,610	3.194,556	3.766,540
abr/23	3.446,174	3.259,112	3.633,236	3.160,182	3.732,166
mai/23	3.408,359	3.221,297	3.595,421	3.122,367	3.694,351
jun/23	3.379,200	3.192,138	3.566,262	3.093,208	3.665,192
jul/23	3.476,723	3.289,661	3.663,785	3.190,731	3.762,715
ago/23	3.726,825	3.539,763	3.913,887	3.440,833	4.012,817
set/23	3.775,263	3.588,201	3.962,325	3.489,271	4.061,255
out/23	3.611,620	3.424,558	3.798,682	3.325,628	3.897,612
nov/23	3.573,656	3.386,594	3.760,718	3.287,664	3.859,648
dez/23	3.466,265	3.279,203	3.653,327	3.180,273	3.752,257
jan/24	3.432,947	3.245,885	3.620,009	3.146,955	3.718,939
fev/24	3.545,612	3.358,550	3.732,674	3.259,620	3.831,604
mar/24	3.556,142	3.369,080	3.743,204	3.270,150	3.842,134
abr/24	3.503,026	3.315,964	3.690,088	3.217,034	3.789,018
mai/24	3.436,810	3.249,748	3.623,872	3.150,818	3.722,802
jun/24	3.459,280	3.272,218	3.646,342	3.173,288	3.745,272

jul/24	3.580,584	3.393,522	3.767,646	3.294,592	3.866,576
ago/24	3.793,048	3.605,986	3.980,110	3.507,056	4.079,040
set/24	3.891,817	3.704,755	4.078,879	3.605,825	4.177,809
out/24	3.708,879	3.521,817	3.895,941	3.422,887	3.994,871
nov/24	3.683,012	3.495,950	3.870,074	3.397,020	3.969,004
dez/24	3.549,560	3.362,498	3.736,622	3.263,568	3.835,552
jan/25	3.517,221	3.330,159	3.704,283	3.231,229	3.803,213
fev/25	3.627,883	3.440,821	3.814,945	3.341,891	3.913,875
mar/25	3.646,652	3.459,590	3.833,714	3.360,660	3.932,644
abr/25	3.593,372	3.406,310	3.780,434	3.307,380	3.879,364
mai/25	3.570,493	3.383,431	3.757,555	3.284,501	3.856,485
jun/25	3.539,071	3.352,009	3.726,133	3.253,079	3.825,063
jul/25	3.646,149	3.459,087	3.833,211	3.360,157	3.932,141
ago/25	3.902,320	3.715,258	4.089,382	3.616,328	4.188,312
set/25	3.962,674	3.775,612	4.149,736	3.676,682	4.248,666
out/25	3.776,411	3.589,349	3.963,473	3.490,419	4.062,403
nov/25	3.760,030	3.572,968	3.947,092	3.474,038	4.046,022
dez/25	3.628,128	3.441,066	3.815,190	3.342,136	3.914,120
jan/26	3.585,531	3.398,469	3.772,593	3.299,539	3.871,523
fev/26	3.717,240	3.530,178	3.904,302	3.431,248	4.003,232
mar/26	3.738,910	3.551,848	3.925,972	3.452,918	4.024,902
abr/26	3.681,025	3.493,963	3.868,087	3.395,033	3.967,017

Fonte: elaboração própria.

## A.5. BRASIL

Tabela A. 37. Coeficientes de Regressão Estimados para o MAPE considerando o Brasil.

	Coef	EP de Coef	Valor-T	Valor-P	VIF
$x_1$	-1,452	0,473	*	*	3,12
$x_2$	2,453	0,478	*	*	3,18
$x_3$	0,198	0,474	*	*	3,12
$x_4$	-0,943	0,459	*	*	2,93
$x_1x_3$	3,92	2,08	1,88	0,066	2,61
$x_1x_4$	9,46	2,20	4,30	0,000	2,92
$x_2x_3$	-8,63	2,07	-4,16	0,000	2,59
$x_1x_4(x_1 - x_4)$	-10,61	4,44	-2,39	0,021	1,30
$x_2x_3(x_2 - x_3)$	-13,61	4,40	-3,09	0,003	1,28
$x_2x_4(x_2 - x_4)$	-11,20	4,54	-2,47	0,017	1,36
$x_1x_2x_4^2$	-95,9	34,8	-2,76	0,008	1,61
$x_2x_4(x_2 - x_4)^2$	-28,9	10,1	-2,87	0,006	1,65

Fonte: elaboração própria.

Tabela A. 38. Análise de variância para o MAPE considerando o Brasil.

	GL	SQ Seq	SQ (Aj.)	QM (Aj.)	Valor F	Valor-P
Regressão	11	32,6296	32,630	2,9663	5,31	0,000
Linear	3	0,0967	16,617	5,5389	9,92	0,000
Quadrático	3	14,0882	20,074	6,6913	11,98	0,000
$x_1x_3$	1	1,4406	1,982	1,9815	3,55	0,066
$x_1x_4$	1	6,8919	10,322	10,3220	18,48	0,000
$x_2x_3$	1	5,7557	9,659	9,6586	17,29	0,000
Cúbico completo	3	9,1769	12,122	4,0408	7,23	0,000
$x_1x_4(x_1 - x_4)$	1	3,1386	3,191	3,1908	5,71	0,021
$x_2x_3(x_2 - x_3)$	1	3,9953	5,341	5,3405	9,56	0,003
$x_2x_4(x_2 - x_4)$	1	2,0430	3,398	3,3983	6,08	0,017
Quártico Especial	1	4,6654	4,248	4,2484	7,61	0,008
$x_1x_2x_4^2$	1	4,6654	4,248	4,2484	7,61	0,008
Quártico1 completo	1	4,6025	4,602	4,6025	8,24	0,006
$x_2x_4(x_2 - x_4)^2$	1	4,6025	4,602	4,6025	8,24	0,006
Erro de Resíduos	49	27,3704	27,370	0,5586		
Total	60	60,0000				

Fonte: elaboração própria.



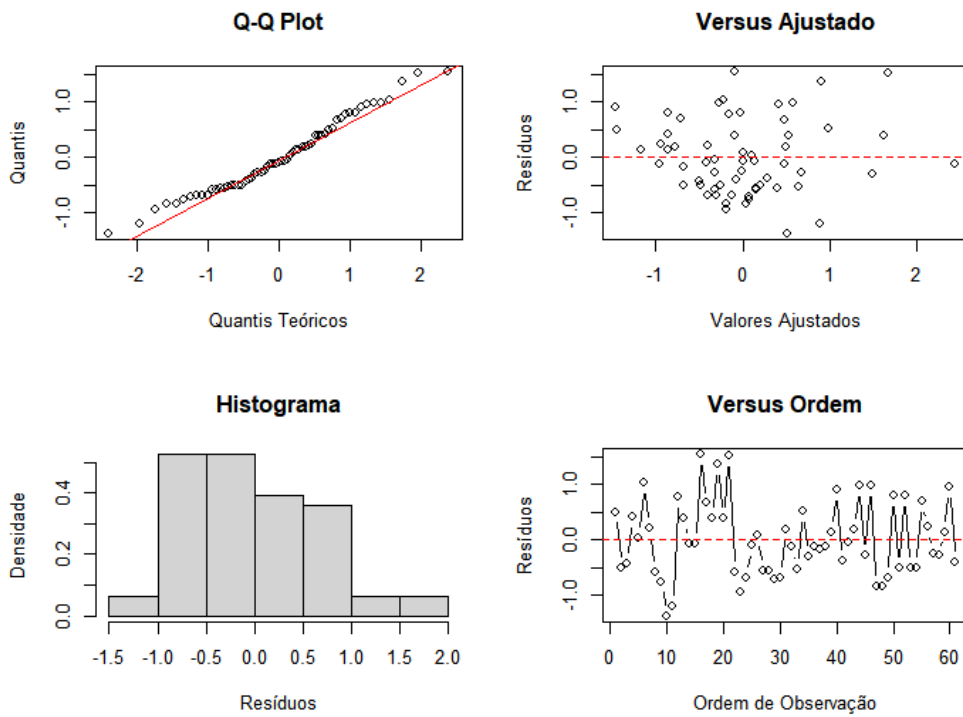


Figura A. 37. Análise de resíduos tendo o MAPE como resposta para o Brasil.  
 Fonte: elaboração própria.

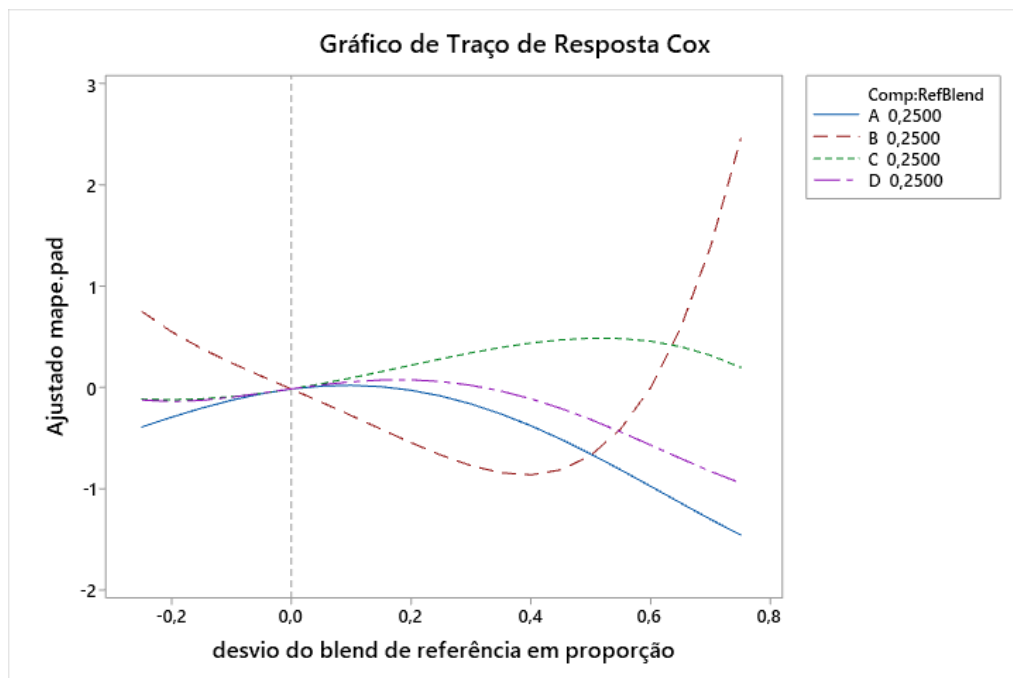


Figura A. 38. Gráfico de Traço de Resposta de Cox para o MAPE como resposta para o Brasil.  
 Fonte: elaboração própria.

Tabela A. 39. Coeficientes de Regressão Estimados para o RMSE considerando o Brasil.

	Coef	EP de Coef	Valor-T	Valor-P	VIF
$x_1$	-0,373	0,390	*	*	1,80
$x_2$	1,208	0,517	*	*	3,17
$x_3$	0,386	0,424	*	*	2,13
$x_4$	0,334	0,395	*	*	1,85
$x_2x_3$	-8,65	2,45	-3,53	0,001	3,07
$x_1x_3(x_1 - x_3)$	-9,79	5,19	-1,89	0,065	1,51
$x_1x_4(x_1 - x_4)$	-11,76	5,13	-2,29	0,026	1,48
$x_2x_3(x_2 - x_3)$	-9,95	4,72	-2,11	0,040	1,25
$x_1^2x_2x_4$	222,2	62,6	3,55	0,001	4,43
$x_1x_2^2x_4$	-173,0	58,5	-2,96	0,005	3,86
$x_1x_2x_4^2$	-103,0	58,4	-1,76	0,084	3,85
$x_2x_3(x_2 - x_3)^2$	27,2	11,6	2,35	0,023	1,86
$x_2x_4(x_2 - x_4)^2$	-18,7	10,9	-1,72	0,092	1,65

Fonte: elaboração própria.

Tabela A. 40. Análise de variância para o RMSE considerando o Brasil.

	GL	SQ Seq	SQ (Aj.)	QM (Aj.)	Valor F	Valor-P
Regressão	12	28,4484	28,448	2,3707	3,61	0,001
Linear	3	1,9105	3,021	1,0070	1,53	0,218
Quadrático	1	4,0226	8,201	8,2014	12,48	0,001
$x_2x_3$	1	4,0226	8,201	8,2014	12,48	0,001
Cúbico completo	3	3,5948	8,292	2,7639	4,20	0,010
$x_1x_3(x_1 - x_3)$	1	0,2403	2,341	2,3413	3,56	0,065
$x_1x_4(x_1 - x_4)$	1	1,1075	3,450	3,4501	5,25	0,026
$x_2x_3(x_2 - x_3)$	1	2,2469	2,917	2,9168	4,44	0,040
Quártico Especial	3	13,3542	11,807	3,9356	5,99	0,001
$x_1^2x_2x_4$	1	0,1024	8,270	8,2703	12,58	0,001
$x_1x_2^2x_4$	1	10,8066	5,753	5,7530	8,75	0,005
$x_1x_2x_4^2$	1	2,4452	2,042	2,0417	3,11	0,084
Quártico1 completo	2	5,5663	5,566	2,7831	4,23	0,020
$x_2x_3(x_2 - x_3)^2$	1	3,6242	3,621	3,6212	5,51	0,023
$x_2x_4(x_2 - x_4)^2$	1	1,9421	1,942	1,9421	2,95	0,092
Erro de Resíduos	48	31,5516	31,552	0,6573		
Total	60	60,0000				

Fonte: elaboração própria.

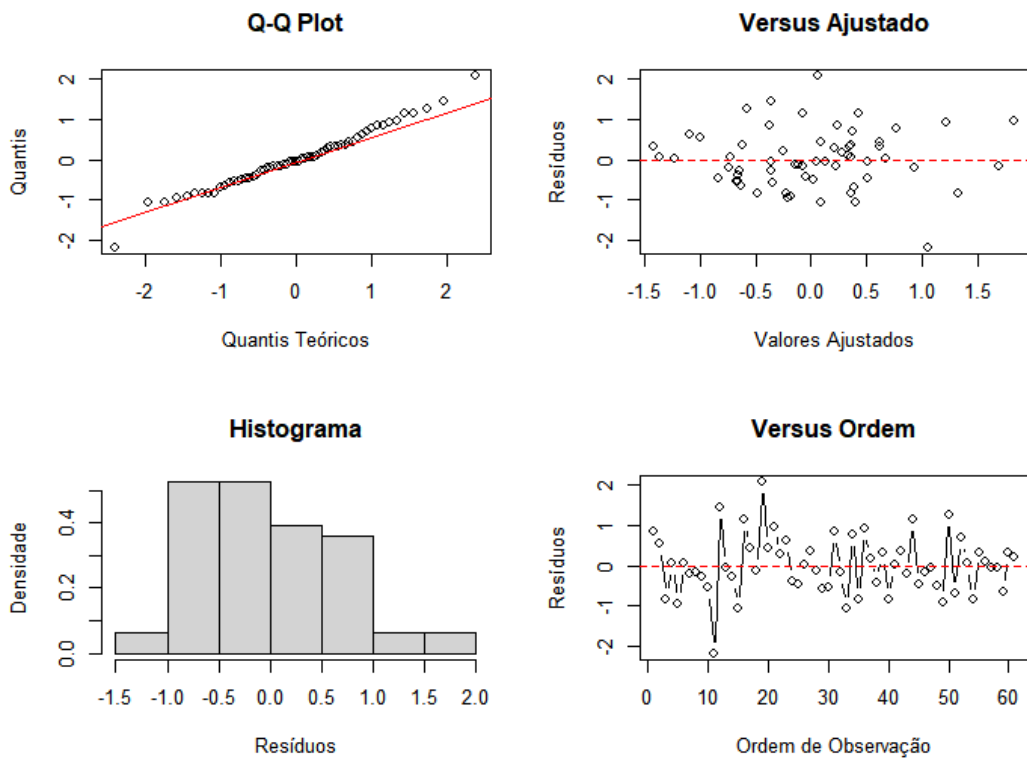


Figura A. 39. Análise de resíduos tendo o RMSE como resposta para o Brasil.  
 Fonte: elaboração própria.

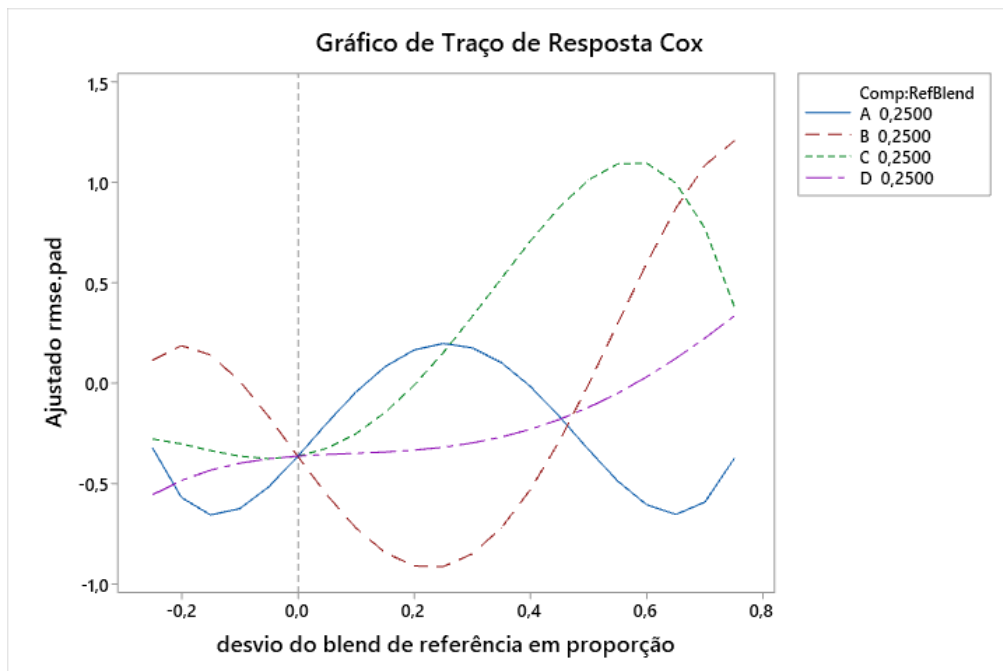


Figura A. 40. Gráfico de Traço de Resposta de Cox para o MAPE como resposta da região Norte.  
 Fonte: elaboração própria.

Tabela A. 41. Coeficientes de Regressão Estimados para o MdAE considerando o Brasil.

	Coef	EP de Coef	Valor-T	Valor-P	VIF
$x_1$	0,059	0,399	*	*	1,60
$x_2$	1,249	0,512	*	*	2,63
$x_3$	0,530	0,455	*	*	2,08
$x_4$	0,029	0,426	*	*	1,82
$x_2x_3$	-6,94	2,38	-2,91	0,005	2,46
$x_1x_4(x_1 - x_4)$	-15,56	5,56	-2,80	0,007	1,47
$x_2x_3(x_2 - x_3)$	-8,00	5,13	-1,56	0,125	1,25
$x_1^2x_2x_4$	106,3	58,6	1,82	0,075	3,28
$x_1x_2x_4^2$	-177,3	58,3	-3,04	0,004	3,25
$x_2x_4(x_2 - x_4)^2$	-19,6	11,7	-1,67	0,100	1,62

Fonte: elaboração própria.

Tabela A. 42. Análise de variância para o MdAE considerando o Brasil.

	GL	SQ Seq	SQ (Aj.)	QM (Aj.)	Valor F	Valor-P
Regressão	9	20,4699	20,470	2,2744	2,93	0,007
Linear	3	0,3048	3,139	1,0463	1,35	0,269
Quadrático	1	5,1581	6,583	6,5830	8,49	0,005
$x_2x_3$	1	5,1581	6,583	6,5830	8,49	0,005
Cúbico completo	2	4,5762	8,014	4,0070	5,17	0,009
$x_1x_4(x_1 - x_4)$	1	3,3163	6,069	6,0686	7,83	0,007
$x_2x_3(x_2 - x_3)$	1	1,2599	1,888	1,8880	2,44	0,125
Quártico Especial	2	8,2570	7,627	3,8135	4,92	0,011
$x_1^2x_2x_4$	1	0,2642	2,554	2,5535	3,29	0,075
$x_1x_2x_4^2$	1	7,9928	7,167	7,1669	9,25	0,004
Quártico1 completo	1	2,1738	2,174	2,1738	2,80	0,100
$x_2x_4(x_2 - x_4)^2$	1	2,1738	2,174	2,1738	2,80	0,100
Erro de Resíduos	51	39,5301	39,530	0,7751		
Total	60	60,0000				

Fonte: elaboração própria.

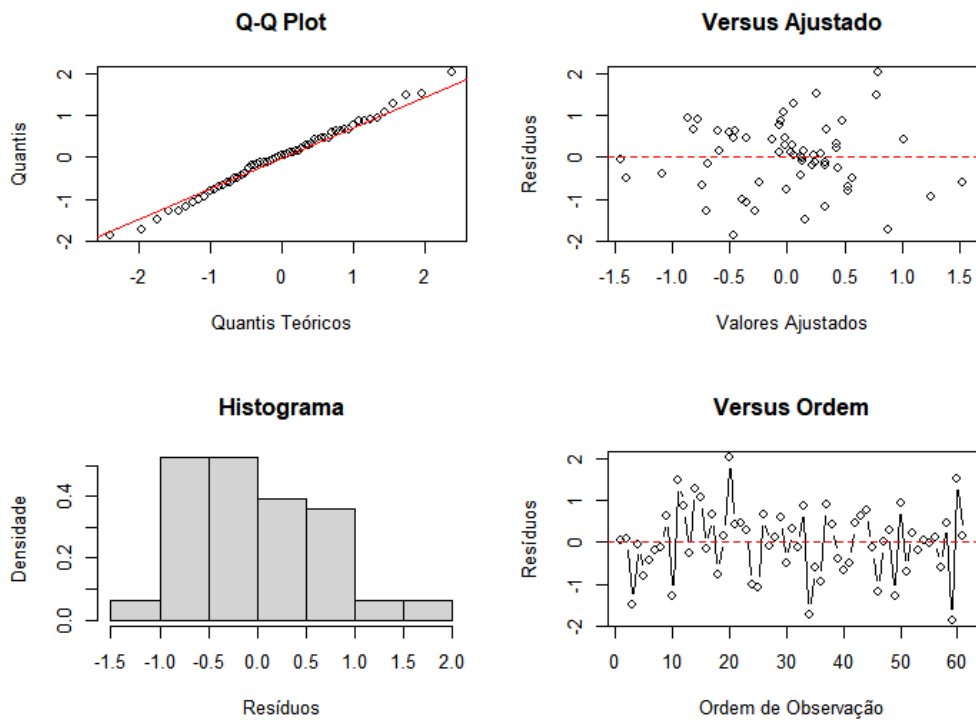


Figura A. 41. Análise de resíduos tendo o MdAE como resposta do Brasil.  
 Fonte: elaboração própria.

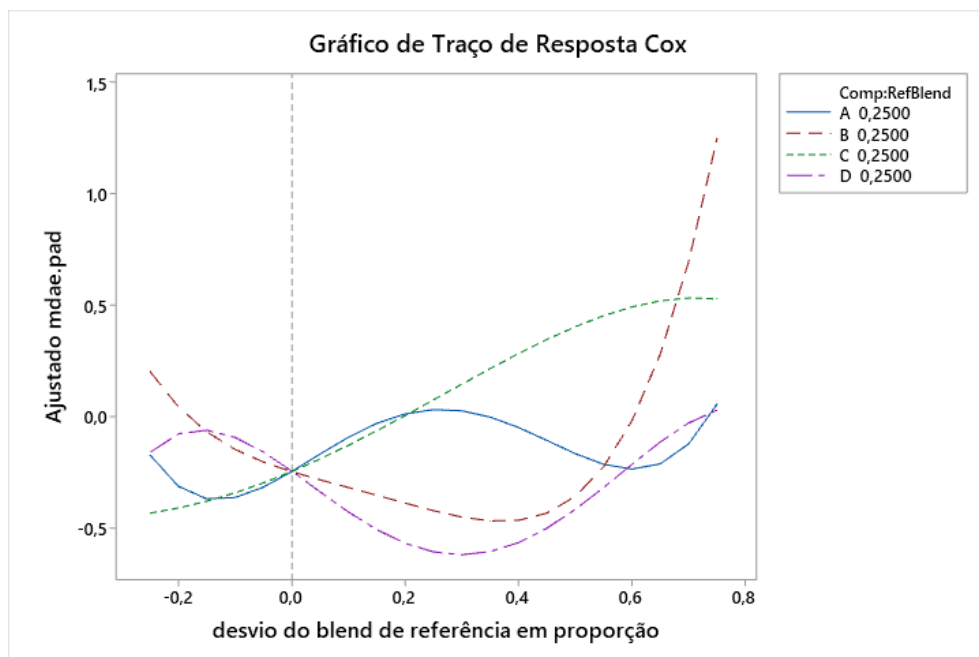


Figura A. 42. Gráfico de Traço de Resposta de Cox para o MdAE como resposta para o Brasil.  
 Fonte: elaboração própria.

Tabela A. 43. Coeficientes de Regressão Estimados para o MAE considerando o Brasil.

	Coef	EP de Coef	Valor-T	Valor-P	VIF
$x_1$	-0,729	0,384	*	*	1,26
$x_2$	0,187	0,403	*	*	1,39
$x_3$	0,651	0,386	*	*	1,28
$x_4$	0,439	0,386	*	*	1,28
$x_2^2 x_3 x_4$	-100,5	40,3	-2,49	0,016	1,32

Fonte: elaboração própria.

Tabela A. 44. Análise de variância para o MAE considerando o Brasil.

	GL	SQ Seq	SQ (Aj.)	QM (Aj.)	Valor F	Valor-P
Regressão	4	9,032	9,032	2,2580	2,48	0,054
Linear	3	3,372	5,462	1,8207	2,00	0,124
Quártico Especial	1	5,660	5,660	5,6599	6,22	0,016
$x_2^2 x_3 x_4$	1	5,660	5,660	5,6599	6,22	0,016
Erro de Resíduos	56	50,968	50,968	0,9101		
Total	60	60,000				

Fonte: elaboração própria.

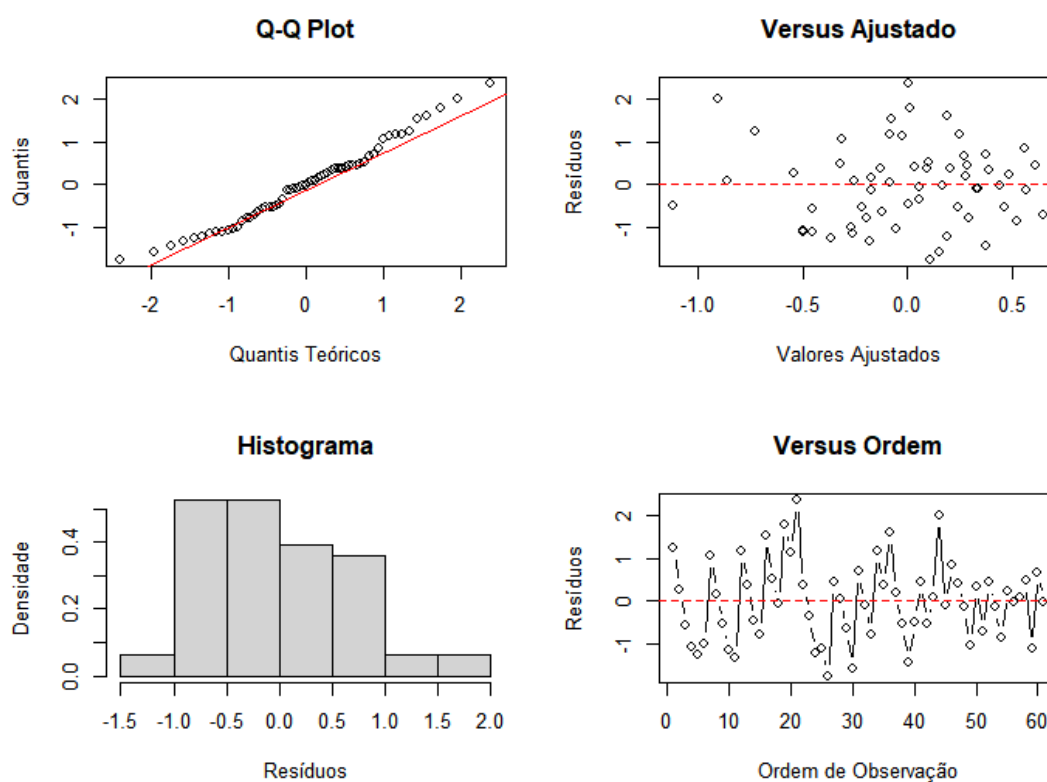


Figura A. 43. Análise de resíduos tendo o MAE como resposta para o Brasil.

Fonte: elaboração própria.

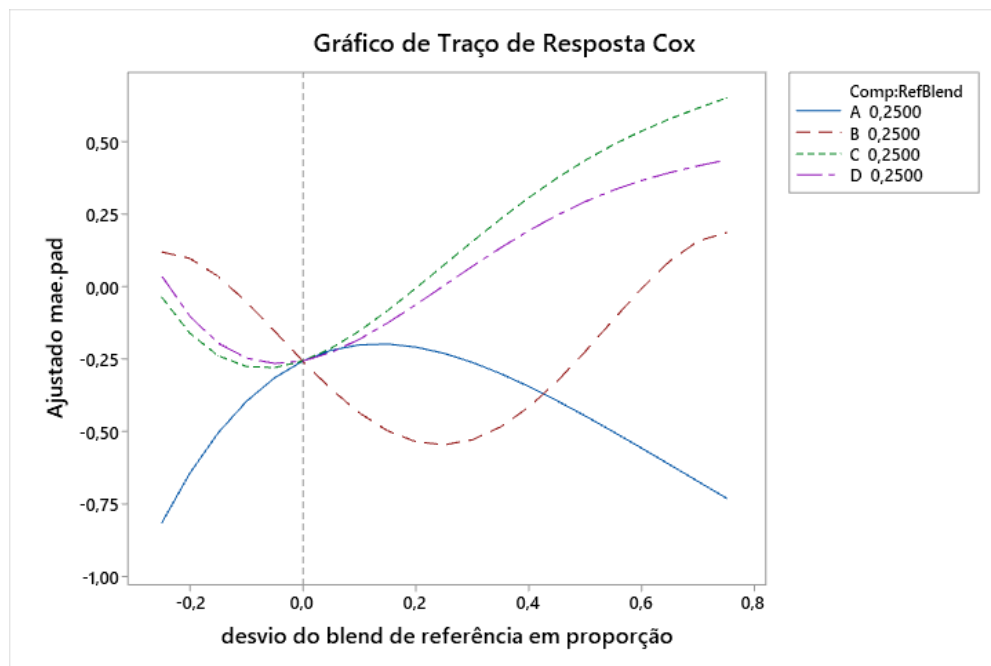


Figura A. 44. Gráfico de Traço de Resposta de Cox para o MAE como resposta para o Brasil.  
 Fonte: elaboração própria.

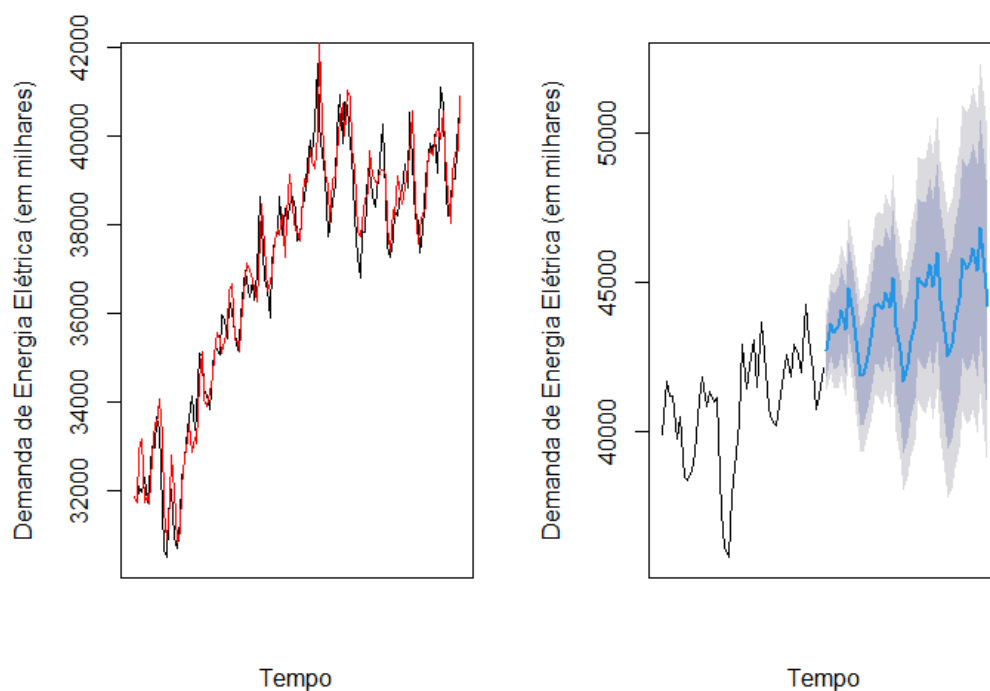


Figura A. 45. Gráfico de ajuste e previsão aplicando a metodologia proposta para o Brasil.  
 Fonte: elaboração própria.

Tabela A. 45. Previsão e intervalos de confiança da demanda de energia elétrica no Brasil.

<b>Tempo</b>	<b>Previsão</b>	<b>80%<sub>inf</sub></b>	<b>80%<sub>sup</sub></b>	<b>95%<sub>inf</sub></b>	<b>95%<sub>sup</sub></b>
ago/22	42.830,077	41.134,504	44.525,650	40.237,781	45.422,373
set/22	43.627,855	41.932,282	45.323,428	41.035,559	46.220,151
out/22	42.978,089	41.282,516	44.673,662	40.385,793	45.570,385
nov/22	43.740,306	42.044,733	45.435,879	41.148,010	46.332,602
dez/22	43.740,919	42.045,346	45.436,492	41.148,623	46.333,215
jan/23	43.044,678	41.349,105	44.740,251	40.452,382	45.636,974
fev/23	45.095,177	43.399,604	46.790,750	42.502,881	47.687,473
mar/23	43.976,299	42.280,726	45.671,872	41.384,003	46.568,595
abr/23	42.957,872	41.262,299	44.653,445	40.365,576	45.550,168
mai/23	41.743,959	40.048,386	43.439,532	39.151,663	44.336,255
jun/23	41.993,797	40.298,224	43.689,370	39.401,501	44.586,093
jul/23	42.875,917	41.180,344	44.571,490	40.283,621	45.468,213
ago/23	43.563,541	41.867,968	45.259,114	40.971,245	46.155,837
set/23	44.416,634	42.721,061	46.112,207	41.824,338	47.008,930
out/23	43.928,744	42.233,171	45.624,317	41.336,448	46.521,040
nov/23	44.513,813	42.818,240	46.209,386	41.921,517	47.106,109
dez/23	44.572,378	42.876,805	46.267,951	41.980,082	47.164,674
jan/24	43.898,877	42.203,304	45.594,450	41.306,581	46.491,173
fev/24	45.803,338	44.107,765	47.498,911	43.211,042	48.395,634
mar/24	44.471,155	42.775,582	46.166,728	41.878,859	47.063,451
abr/24	43.422,934	41.727,361	45.118,507	40.830,638	46.015,230
mai/24	42.293,242	40.597,669	43.988,815	39.700,946	44.885,538
jun/24	42.697,842	41.002,269	44.393,415	40.105,546	45.290,138
jul/24	43.625,792	41.930,219	45.321,365	41.033,496	46.218,088
ago/24	44.369,205	42.673,632	46.064,778	41.776,909	46.961,501
set/24	45.334,676	43.639,103	47.030,249	42.742,380	47.926,972
out/24	44.766,722	43.071,149	46.462,295	42.174,426	47.359,018
nov/24	45.380,155	43.684,582	47.075,728	42.787,859	47.972,451
dez/24	45.480,329	43.784,756	47.175,902	42.888,033	48.072,625
jan/25	44.742,519	43.046,946	46.438,092	42.150,223	47.334,815
fev/25	46.679,008	44.983,435	48.374,581	44.086,712	49.271,304
mar/25	45.385,190	43.689,617	47.080,763	42.792,894	47.977,486
abr/25	44.253,051	42.557,478	45.948,624	41.660,755	46.845,347
mai/25	43.180,323	41.484,750	44.875,896	40.588,027	45.772,619
jun/25	43.552,030	41.856,457	45.247,603	40.959,734	46.144,326
jul/25	44.456,799	42.761,226	46.152,372	41.864,503	47.049,095
ago/25	45.245,970	43.550,397	46.941,543	42.653,674	47.838,266
set/25	46.164,593	44.469,020	47.860,166	43.572,297	48.756,889
out/25	45.540,471	43.844,898	47.236,044	42.948,175	48.132,767
nov/25	46.258,700	44.563,127	47.954,273	43.666,404	48.850,996
dez/25	46.266,824	44.571,251	47.962,397	43.674,528	48.859,120
jan/26	45.537,938	43.842,365	47.233,511	42.945,642	48.130,234
fev/26	47.570,051	45.874,478	49.265,624	44.977,755	50.162,347
mar/26	46.289,523	44.593,950	47.985,096	43.697,227	48.881,819
abr/26	45.149,671	43.454,098	46.845,244	42.557,375	47.741,967

Fonte: elaboração própria.