UNIVERSIDADE FEDERAL DE ITAJUBÁ

PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE PRODUÇÃO

Estudo Comparativo entre Redes Neurais Artificiais e *Markov-Switching Model na*modelagem de séries temporais não lineares

Estevão Luiz Romão

UNIVERSIDADE FEDERAL DE ITAJUBÁ

PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE PRODUÇÃO

Estevão Luiz Romão

Estudo Comparativo entre Redes Neurais Artificiais e *Markov-Switching Model* na modelagem de séries temporais não lineares

Dissertação submetida ao Programa de Pós-Graduação em Engenharia de Produção como parte dos requisitos para obtenção do título de **Mestre em Ciências em Engenharia de Produção**.

Área de Concentração: Engenharia de Produção **Orientador:** Prof. Dr. Pedro Paulo Balestrassi

Itajubá, setembro de 2019

UNIVERSIDADE FEDERAL DE ITAJUBÁ

PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE PRODUÇÃO

Estevão Luiz Romão

Estudo Comparativo entre Redes Neurais Artificiais e *Markov-Switching Model* na modelagem de séries temporais não lineares

Dissertação aprovada por banca examinadora em 17 de setembro de 2019, conferindo ao autor o título de **Mestre em** Ciências em Engenharia de Produção.

Banca Examinadora:

Prof.^a Dr.^a Mariângela de Oliveira Abans (LNA)

Prof. Dr. Antonio Fernando Branco Costa (UNIFEI)

Prof. Dr. Pedro Paulo Balestrassi (Orientador)

Itajubá, setembro de 2019

DEDICATÓRIA

A minha mãe por todo apoio e dedicação.

AGRADECIMENTOS

A Deus por mais essa oportunidade!

Ao meu grande amigo e orientador Pedro Paulo Balestrassi por todo apoio prestado no desenvolvimento desse trabalho.

A minha mãe, Orceli e meu pai Juarez que sempre me apoiaram na minha carreira acadêmica e também ao meu padrasto Warlen por sempre me incentivar.

Agradeço imensamente à minha querida noiva Daniela por todo apoio, carinho e companheirismo, estando sempre ao meu lado me dando muita força e não me deixando nunca desanimar.

A todos os meus familiares que sempre estiveram presentes em minha vida, sem eles nunca teria chegado aonde cheguei, em especial a minha tia Mirtes e minha prima Lívia.

Muito obrigado aos meus caros amigos da UNIFEI: Andreza, Eduardo, Renata, Fabrício, Alexandre, Aline, Taynara, João Victor, Vinícius, Clara, Kelly, Afonso, Adriano, Gustavo, Franco, Max, Laila, Natália, Lucas, Marina, e em especial à minha grande amiga Simone que sempre me ajudou e muito colaborou na elaboração e conclusão desse trabalho.

Agradeço também aos meus amigos que moraram comigo durante esse período, Filipe, Rafael, Diogo, João Luís, Athos, Gustavo, Marco Antônio e é claro à dona do Pensionato Rita Helena, o meu muito obrigado!

Aos meus queridos amigos da casa espírita Fé, Esperança e Caridade, a todos o meu muito obrigado!

Agradeço imensamente aos professores que contribuíram imensamente para a minha formação: Pedro Paulo Balestrassi, Anderson Paiva, Rafael Leme, Carlos Henrique Mello, Edson Pamplona, José Henrique, João Turrioni, José Arnaldo e aos membros da banca examinadora, Professora Mariângela e ao Professor Antonio Fernando Branco, pelos seus comentários que muito auxiliaram na melhoria deste trabalho.

Agradeço ao apoio financeiro e estrutural fornecido pela CAPES, CNPq e UNIFEI que tornou possível a elaboração desse trabalho.

EPÍGRAFE

RESUMO

Séries temporais não lineares são amplamente encontradas em situações reais envolvendo diferente cenários tais como crescimento econômico, consumo de energia e mudanças climáticas. Elas são, em muitos casos, não estacionárias e o seu comportamento pode variar ao longo do tempo, o que tem incentivado o uso de técnicas capazes de capturar essas características. Esse trabalho objetiva apresentar um estudo comparativo entre a precisão das Redes Neurais Artificiais (RNA) e o método denominado Markov-Switching Model (MSM) em relação à modelagem de séries temporais não lineares e não estacionárias. Para tal foi desenvolvida e aplicada uma metodologia de apoio à decisão baseada na técnica de Design of Experiments (DOE) para conduzir as análises e auxiliar na elaboração de um banco de dados sintético, representando contextos distintos, o que permitiu generalizar as comparações. Foram considerados os seguintes fatores: tamanho do conjunto de dados, erro aleatório associado a cada série temporal e a correlação existente entre elas. Após modelar cada série usando RNA e MSM, calculou-se o valor do Erro Percentual Absoluto Médio (MAPE) e desenvolveu-se um teste t-pareado que permitiu concluir que o desempenho das RNA é estatisticamente melhor que o do método MSM independente de qual cenário considerado. Além disso, foi desenvolvido um estudo de caso cujo objetivo foi modelar o valor de exportação do Brasil durante os 10 últimos anos como uma função das seguintes séries temporais: taxa de câmbio, importação e PIB, no mesmo período considerado, além da própria série de exportação deslocada de uma unidade no tempo. Assim, pôde-se validar que, de fato, as RNA possuem uma melhor performance em problemas envolvendo séries temporais não lineares e não estacionárias que o MSM.

Palavras-Chaves: Redes Neurais Artificiais, *Markov-Switching Model*, Séries Temporais Não Lineares, *Design of Experiments*

ABSTRACT

Nonlinear time series are widely encountered in real situations involving different scenarios such as economic growth, energy consumption, and climate changes. These time series are usually nonstationary and their behavior can change over time leading us to use sophisticated techniques that are able to capture these characteristics. This work aims to present a comparative study between Artificial Neural Networks (ANN) and the Markov-Switching Model (MSM) performances in modeling nonstationary nonlinear time series. Then, it was developed and it was applied a decision support methodology based on Design of the Experiments (DOE) in order to guide the analysis and the creation of a synthetic dataset, representing distinct contexts, which allows to generalize the comparisons. The design considered the following factors: dataset sizes, random error associated to each time series, and the correlations between them. After modeling each one using ANN and MSM we calculated the Mean Absolute Percentage Error (MAPE) and developed a paired-t test, and this led us to conclude that the performance of ANN is statistically better than MSM's no matter the scenario considered. Furthermore, this study presents a case which aims to model Brazilian exportation during the last 10 years as a function of three time series: exchange rate, importation, and gross domestic product at the same period, besides the exportation series, but delayed one unit in time. Thus, we could validate that ANN has a better performance in problems involving nonstationary nonlinear time series than MSM.

Keywords: Artificial Neural Networks, Markov-Switching Model, Nonlinear Time Series, Design of Experiments

LISTA DE FIGURAS

FIGURA 1. KELATORIO DE PUBLICAÇÕES (A) E CITAÇÕES (B) NA PLATAFORMA WEB OF SCIENCE PARA O TERMO MARKOV M	ODELS 14
FIGURA 2. QUANTIDADE DE PUBLICAÇÕES RELACIONADAS A MODELOS MARKOVIANOS POR ÁREA DE CONHECIMENTO	14
Figura 3. Relatório de publicações (a) e citações (b) na plataforma Web of Science para o termo Artificial N	IEURAL
Network	15
FIGURA 4. QUANTIDADE DE PUBLICAÇÕES RELACIONADAS ÀS REDES NEURAIS ARTIFICIAIS POR ÁREA DE CONHECIMENTO	15
Figura 5. Modelo não linear de um neurônio de índice $m{k}$	21
FIGURA 6. EXEMPLO DA ESTRUTURA DE UMA REDE NEURAL	22
FIGURA 7. PROBABILIDADES ASSOCIADAS AOS REGIMES DOS DADOS MODELADOS POR MSM	24
Figura 8. Metodologia de apoio à decisão na análise e comparação de métodos estatísticos	28
FIGURA 9. DESIGNS DISPONÍVEIS NO SOFTWARE MINITAB E SUAS RESPECTIVAS RESOLUÇÕES	30
FIGURA 10. TELA DO IPS PARA DEFINIR CRITÉRIO DE PARADA E QUANTIDADE DE REDES NEURAIS RETIDAS	35
FIGURA 11. TELA DO IPS PARA DEFINIR O CRITÉRIO DE ESCOLHA DA MELHOR REDE NEURAL	35
FIGURA 12. TELA DO IPS PARA DEFINIÇÃO DOS TIPOS DE REDE TESTADAS	35
Figura 13. Tela do IPS para definição da quantidade de neurônios na camada oculta	36
Figura 14. Tela do IPS para definição da quantidade de termos autoregresivos	36
Figura 15. Tela do IPS para definição do tipo da função de ativação	37
Figura ${f 16}.$ Comportamento da $f{f 1}t$ para valores de $ heta{f 1}$ positivos e negativos	40
Figura $17.$ Comportamento da $f2t$ para valores de $ heta2$ e $ heta3$ positivos e negativos	41
Figura 18. Gráfico de Pareto dos efeitos padronizados para a resposta MAPE MSM e gráfico de distribuiçã	O DE
PROBABILIDADE ASSOCIADO	44
FIGURA 19. GRÁFICO DE PROBABILIDADE PARA OS RESÍDUOS DO MODELO MAPE MSM	46
FIGURA 20. GRÁFICO DE EFEITOS PRINCIPAIS PARA A RESPOSTA MAPE MSM	47
Figura 21. Gráfico de interações para a resposta MAPE MSM	47
Figura 22. Gráfico de Pareto dos efeitos padronizados para a resposta Diferença e gráfico de distribuição i	DE
PROBABILIDADE ASSOCIADO	50
FIGURA 23. GRÁFICO DE PROBABILIDADE NORMAL PARA OS RESÍDUOS DO MODELO PARA DIFERENÇA	51
Figura 24. Gráfico de efeitos principais para a resposta Diferença	52
Figura 25. Gráfico de interações para a resposta Diferença	53
Figura 26. Resultados do algoritmo Desirability	54
Figura 27. Valores de preço de exportação no Brasil de 2008 a 2018	59
Figura 28. Valores da Taxa de Câmbio brasileira no período de 2008 a 2018	60
Figura 29. Valores do PIB brasileiro de 2008 a 2018.	60
Figura 30. Valores de preço de importação no Brasil de 2008 a 2018	61
FIGURA 31. GRÁFICO DE PROBABILIDADE DE RESÍDUOS MSM	62
FIGURA 32. GRÁFICO DE ENSAIOS DE RESÍDUOS MSM	62

Figura 33. Gráfico de probabilidades de Resíduos RNA	63
Figura 34. Gráfico de ensaios de Resíduos RNA	63
Figura 35. Gráfico de autocorrelação para os resíduos do método MSM	64
FIGURA 36. GRÁFICO DE AUTOCORRELAÇÃO PARA OS RESÍDUOS DO MÉTODO RNA	64

LISTA DE QUADROS

QUADRO 1. MÉTODOS COMUMENTE APLICADOS PARA A SOLUÇÃO DE PROBLEMAS ESTATÍSTICOS	29
QUADRO 2. FATORES QUE PODEM SER SIGNIFICATIVOS EM CADA TIPO DE PROBLEMA	30
OUADRO 3. ARGUMENTOS DA FUNCÃO MSMFIT	34

LISTA DE TABELAS

Tabela 1. Fatores considerados no estudo e seus respectivos níveis	38
Tabela 2. Experimentos realizados na pesquisa	39
Tabela 3. Resultados dos experimentos	42
Tabela 4. Análise de variância do modelo MAPE MSM	43
Tabela 5. Análise dos coeficientes do modelo MAPE MSM	43
Tabela 6. Análise de variância do modelo MAPE RNA	48
Tabela 7. Análise dos coeficientes do modelo MAPE RNA	49
Tabela 8. Análise de variância para a resposta diferença	50
Tabela 9. Análise dos coeficientes para o modelo da resposta Diferença	50
Tabela 10. Análise dos resíduos para o método MSM	56
Tabela 11. Análise dos resíduos para o método RNA	57
TABELA 12. ANÁLISE DE AUTOCORRELAÇÃO DOS RESÍDUOS GERADOS PELOS MÉTODOS MSM E RNA	58

LISTA DE ABREVIATURAS E SIGLAS

AR Modelo auto-regressivo

ARMA Modelo auto-regressivo de média móvel

ARCH Modelo de heteroscedasticidade condicional auto-regressiva

BL Modelo bilinear

DOE Design of Experiments

EM Algoritmo Expectation Maximization

EXPAR Modelo auto-regressivo exponencial

GARCH Modelo generalizado de heteroscedasticidade condicional auto-regressiva

IID Independente e Identicamente Distribuída

IPEA Instituto de Pesquisa Econômica Aplicada

IPS Intelligent Problem Solver

MA Modelo de Média Móvel

MAPE Erro percentual absoluto médio

MLP Multilayer Perceptron

MSM Markov-Switching Model

PIB Produto Interno Bruto

RNA Redes Neurais Artificiais

STAR Modelo auto-regressivo com transição suave

TAR Modelo auto-regressivo com limiar

LISTA DE SÍMBOLOS

$ ho_{ij}$	Correlação entre as series temporais i e j
$\sigma_{arepsilon_i}$	Desvio do erro aleatório adicionado à série y_i
σ_{y_i}	Desvio padrão da série y_i
LBQ	Estatística de <i>Ljung-Box</i>
r_k	Estimativa para a autocorrelação de <i>lag k</i>
φ(.)	Função de ativação de um neurônio artificial
$f_i(t)$	i-ésima função não linear utilizada para geração dos dados
$ heta_k$	j-ésimo coeficiente de $f_i(t)$
\bar{y}	Média amostral em um contexto mais genérico
x_n	n-ésima entrada de um neurônio artificial
n_r	Número de réplicas
ψ	Números reais
w_{ij}	Peso atribuído à determinada entrada em um neurônio artificial
b_k	Peso atribuído ao bias de um determinado neurônio artificial
p_{ij}	Probabilidade de a série transitar do regime i para o regime j
n	Quantidade de dados na amostra
α_t	Sequência de variáveis aleatórias IID com distribuição de probabilidade bem
	definida
y_i	Série temporal sintética criada a partir de $f_i(t)$
Σ	Somador de um neurônio artificial
T	Tamanho do conjunto de dados
y_t	t-ésimo valor da série y_i
\hat{y}_t	Valor estimado para o t-ésimo valor da série y_i
μ	Valor médio (constante) de uma série temporal
$\sigma_{ar{y}}^{\;\;2}$	Variância da média amostral
s_t	Variável aleatória não observável que governa a troca de regimes das séries
	temporais no método Markov-switching model

SUMÁRIO

1.	INTRODUÇÃO	12
1.1	Contexto	12
1.2	Relevância	13
1.3	Justificativa e Contribuição	15
1.4	Delimitações do estudo	16
1.5	Objetivos	17
1.5.1	Objetivo geral	17
1.5.2	Objetivos específicos	17
1.6	Classificação da pesquisa científica	17
1.7	Estrutura do trabalho	18
2.	FUNDAMENTAÇÃO TEÓRICA	19
2.1	Séries temporais não lineares	19
2.2	Redes neurais artificiais	20
2.3	Markov-Switching Model	23
2.4	Planejamento de experimentos	25
3.	METODOLOGIA	27
4.	ESTUDO COMPARATIVO	33
4.1	Planejamento pré-experimental	33
4.2	Escolha do design experimental	39
4.3	Realização dos experimentos	39
4.4	Análise estatística dos dados	42
4.4.1	Análise estatística para a resposta MAPE MSM	42
4.4.2	Análise estatística para a resposta MAPE RNA	48
4.4.3	Análise estatística para a resposta Diferença	49
4.4.4	Análises estatísticas complementares	53
5.	ESTUDO DE CASO	59
6.	CONCLUSÕES	65
6.1	Contribuições do trabalho	65
6.2	Sugestão para trabalhos futuros	66
REFI	ERÊNCIAS BIBLIOGRÁFICAS	68
APÊ	NDICE A – GERAÇÃO DOS DADOS SINTÉTICOS	74
APÊ	NDICE B – ANÁLISE DOS RESÍDUOS	99

1. INTRODUÇÃO

1.1 Contexto

Uma série temporal pode ser definida como uma coleção de dados observados sequencialmente ao longo do tempo. Em diversas situações a modelagem desses dados é de extrema importância, visto que com um bom modelo é possível realizar boas previsões sobre o futuro de forma a auxiliar nos estudos de variáveis econômicas e financeiras, previsão de vendas, desempenho de um processo ao longo do tempo, entre outros.

Diversas situações reais podem ser encaradas como séries temporais, entretanto nem sempre elas apresentam comportamentos lineares. Períodos de crescimento e decréscimo em séries relacionadas à economia, ao consumo de energia e às mudanças climáticas, criam um conjunto de dados no qual a média não se mantém constante ao longo do tempo, o que implica na não estacionariedade da série.

A necessidade de se obter melhores modelos tem gerado esforços no desenvolvimento de métodos precisos capazes de melhor representar a realidade. Entretanto, muitos estudos têm sido direcionados à construção de modelos lineares, mesmo quando a suposição sobre linearidade não pode ser satisfeita, levando o pesquisador à elaboração de modelos inadequados e consequentemente a inferências errôneas (CHAUDHURI; KUMAR, 2015).

Segundo Laner, Svoboda e Rupp (2014), modelos como o Auto-regressivo (AR), Média Móvel (MA) e modelos Auto-regressivos de Média Móvel (ARMA) são frequentemente utilizados em análises de flutuações macroeconômicas. Além disso, comportam-se linearmente com parâmetros fixos ao longo do tempo. Por essa razão, ressalta-se a importância de se considerar que as séries temporais nem sempre são estacionárias e que transições de estado dos seus dados devem ser consideradas (TANG, 2011).

Uma alternativa aos métodos anteriormente mencionados é o uso da abordagem proposta por J. D. Hamilton, em 1989, denominada *Markov-Switching Model* (MSM). Esse método é capaz de caracterizar mudanças nos parâmetros do processo autoregressivo ao longo do tempo (HAMILTON, 1989). Muitas séries temporais reais estão submetidas a episódios nos quais o seu comportamento muda drasticamente, por exemplo, em séries temporais macroeconômicas e financeiras quando observadas por um período de tempo suficientemente longo (HAMILTON, 1994).

Outra abordagem para lidar com problemas de séries temporais não lineares são as Redes Neurais Artificiais (RNA) que, segundo Haykin (2009), teve a base de sua teoria em 1943 com o trabalho de McCulloch e Pitts. Em seguida, Rosenblatt (1958) propõe o modelo chamado Perceptron que imita o cérebro humano tanto em sua estrutura como em seu modo de operação. As RNA são orientadas pelos dados, ou seja, são capazes de capturar estruturas não lineares sem a necessidade de possuir previamente algum conhecimento sobre a relação fundamental que rege o comportamento dos dados (TSAY, 2005).

1.2 Relevância

Existem na literatura alguns trabalhos que comparam o desempenho de duas abordagens distintas para a solução de um mesmo problema. Entretanto, esses estudos geralmente consideram um único cenário que é o do problema proposto, impossibilitando a generalização das conclusões para outras situações diferentes da estudada.

Em termos matemáticos, é difícil provar que uma abordagem é superior a outra, por exemplo, por meio de um teorema. Só é possível obter conclusões mais amplas em um estudo comparativo por meio da generalização do problema, avaliando-o em diversos cenários e avaliando o desempenho dos métodos comparados por meio de uma métrica comum previamente selecionada.

O presente estudo aplica a ferramenta de *Design of Experiments* (DOE), que permite generalizar os casos em que ambos os métodos podem ser aplicados. Cada experimento do DOE dará origem a um conjunto de dados criados para representar uma determinada situação específica. Dessa forma, cria-se um banco de dados sintético cuja elaboração será explicada mais adiante em maiores detalhes.

Assim, após realizada uma pesquisa nas bases de dados *Scopus* e *Web of Science* não foi encontrado nenhum trabalho que compara os métodos MSM e RNA. Entretanto, ao buscar por trabalhos que apresentem *Markov Models* no título foram encontrados, na base *Web of Science*, 6.469 artigos nos últimos 10 anos, apresentando crescimento significativo em número de publicações e citações, conforme Figura 1. É importante ressaltar que o termo utilizado envolve *Hidden Markov Models*, *Markov Chains*, além do método, objeto do presente estudo, *Markov-Switching Models*, sendo que todos eles estão associados a modelos que envolvem troca de regime ao longo do tempo.

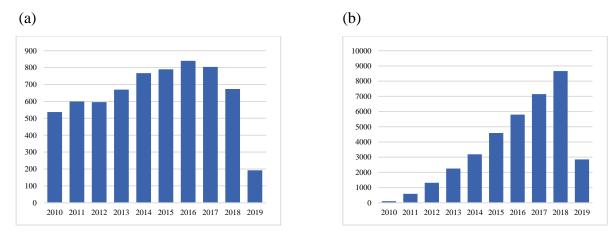


Figura 1. Relatório de publicações (a) e citações (b) na plataforma Web of Science para o termo Markov Models

Observa-se que as citações relacionadas aos modelos Markovianos aumentaram significativamente a partir dos anos 2010. As principais áreas de aplicação desses métodos podem ser observadas na Figura 2.

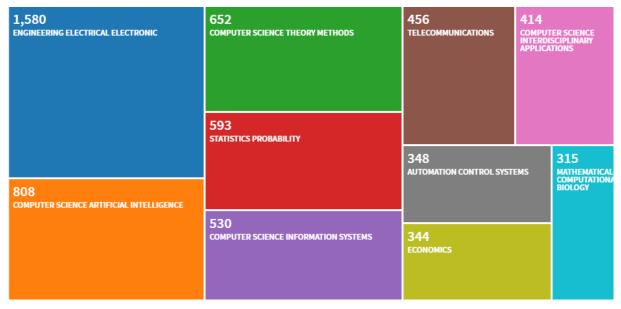


Figura 2. Quantidade de publicações relacionadas a modelos Markovianos por área de conhecimento

A mesma pesquisa foi realizada utilizando *Artificial Neural Networks* como termo de busca no título dos artigos. Foram encontrados 13.059 trabalhos nos últimos 10 anos, entretanto a base *Web of Science*, apenas disponibiliza o relatório de publicações e citações para valores inferiores a 10.000, assim foi necessário refinar a busca considerando apenas os últimos 7 anos. Foram encontrados ao todo 9.713 trabalhos e os relatórios correspondentes podem ser observados na Figura 3.

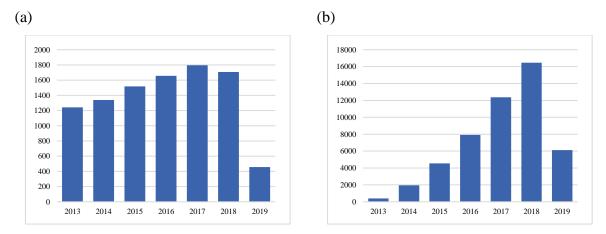


Figura 3. Relatório de publicações (a) e citações (b) na plataforma Web of Science para o termo Artificial
Neural Network

Percebe-se a importância das redes neurais artificiais no cenário atual, sendo que as áreas de aplicação do método podem ser visualizadas na Figura 4.

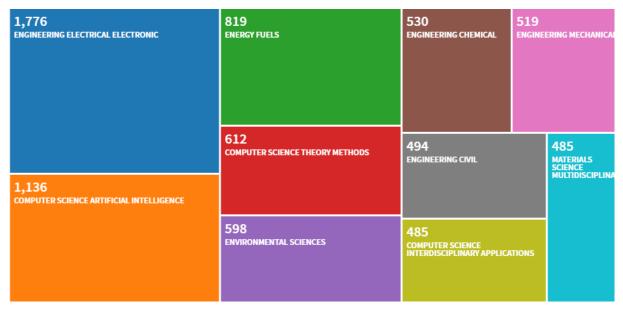


Figura 4. Quantidade de publicações relacionadas às redes neurais artificiais por área de conhecimento

1.3 Justificativa e Contribuição

O surgimento de métodos mais sofisticados, em diversas áreas, relacionados ao estudo de séries temporais tem crescido bastante nas últimas décadas. Dai e Bikdash (2017) propõem métodos para completar os dados faltantes em séries temporais fragmentadas obtidas através de aplicativos de celular relacionados a saúde (monitoramento de peso, sono, etc.). Outro exemplo é o trabalho de Dong, Lian e Liu (2018) no qual é proposto um modelo de rede neural híbrida utilizando o algoritmo de enxame de partículas combinado à função de informação de entropia

para otimizar os valores de peso da rede neural, visando prever características de séries temporais relacionadas a postagens online por intervalo de tempo.

As RNA são amplamente utilizadas com o objetivo de modelar diversos comportamentos não lineares por não precisarem de um modelo inicial para realizar a modelagem dos dados (HAYKIN, 2009). Alguns exemplos recentes de aplicação do método no estudo de séries temporais são: previsão de preço de ações (CAO; LI; LI, 2019), previsão de produção de petróleo (SAGHEER; KOBT, 2019) e preço do petróleo bruto (CEN; WANG, 2019; WU; WU; ZHU, 2019).

O método MSM, é comumente utilizado para modelagem de séries econométricas com comportamentos não lineares e que apresentam diferentes regimes. O trabalho de Fornaciari e Grillenzoni (2017) modela séries financeiras de forma a maximizar a rentabilidade ao lidar com o mercado de ações. Ailliot e Monbet (2012) utilizaram MSM na previsão de dados que descrevem as condições climáticas (vento) de uma região, Mikaeil *et al.* (2014) aplicaram o método na previsão do estado do usuário primário em rádio cognitiva, Lu, Zeng e Chen (2010) usaram MSM para a detecção de surtos de doenças infecciosas.

Diante de dois métodos amplamente utilizados em problemas de grande complexidade, identifica-se uma lacuna na literatura capaz de indicar, por meio de um estudo comparativo, qual deles é mais preciso em diferentes situações. Dessa forma, o presente trabalho busca contribuir no sentido de orientar qual dos métodos possui melhor desempenho na modelagem de séries temporais não lineares e não estacionárias, fornecendo, ao mesmo tempo, informações acerca de quais fatores influenciam a performance de cada um deles.

Outra contribuição desse trabalho é a exemplificação do uso da técnica de *Design of Experiment* (DOE) sendo utilizada para a realização de estudos comparativos. Por meio dela, é possível criar um banco de dados sintético que generaliza os possíveis cenários nos quais os métodos em estudo podem ser aplicados, permitindo assim, alcançar conclusões confiáveis sobre cada um deles.

1.4 Delimitações do estudo

Esse trabalho, conforme anteriormente mencionado, busca realizar um estudo comparativo entre os métodos de RNA e MSM para a modelagem de séries temporais não estacionárias e não lineares.

O *software* Statistica 7.0 foi utilizado para realizar a modelagem das séries temporais por meio das redes neurais através do *Inteligent Problem Solver* (IPS) que testa várias configurações de RNA e seleciona a de menor erro. O tipo de rede testado foi o *three layer*

perceptron e o tempo de treinamento de 1 minuto. Enquanto isso, para a modelagem pelo MSM foi utilizado o pacote MSwM do software R versão 1.4

Assim, esse estudo não se propõe a discutir os detalhes de funcionamento das RNA nem do algoritmo utilizado no método MSM. Os leitores que sentirem necessidade de aprofundarem seus conhecimentos sobre o tema podem buscar os trabalhos de Hamilton (1994) para melhor compreensão do método MSM e Haykin (2009) para as RNA.

1.5 Objetivos

1.5.1 Objetivo geral

Essa pesquisa possui como objetivos gerais:

- Apresentação da metodologia de apoio à decisão desenvolvida e aplicada para a realização estudos comparativos entre métodos estatísticos distintos.
- Realizar um estudo comparativo entre os métodos de RNA e MSM na modelagem de diversas séries temporais sintéticas de forma a possibilitar a identificação de qual dentre essas duas abordagens é a que apresenta menor erro (MAPE) e consequentemente melhor desempenho.

1.5.2 Objetivos específicos

Durante a elaboração do banco de dados sintéticos serão considerados três fatores: tamanho do conjunto de dados, correlação entre as séries temporais e erro aleatório adicionado às séries. Assim, especificamente, objetiva-se definir quais e como os fatores e suas interações influenciam cada um dos métodos, identificar em que circunstâncias eles possuem desempenhos semelhantes, além de validar o estudo comparativo por meio de um estudo de caso.

1.6 Classificação da pesquisa científica

Segundo Bertrand e Fransoo (2002) esta pesquisa pode ser classificada como experimental, pois envolve a realização de experimentos baseados em um conjunto de dados sintéticos elaborado durante o próprio desenvolvimento do estudo. Entretanto também apresenta características de modelagem e simulação, já que a partir de funções não lineares já existentes extraídas do *software* Minitab e com algumas manipulações matemáticas criou-se um conjunto de dados sintéticos seguindo a estrutura do arranjo fatorial completo.

Ainda segundo os autores, a pesquisa também pode ser classificada como axiomática normativa, já que elabora modelos matemáticos para o valor da métrica de erro percentual absoluto médio (MAPE) que indica o desempenho dos métodos. Esses modelos fornecem *feedback* para a literatura acerca de quais fatores e interações influenciam significativamente o desempenho dos métodos e como essa influência ocorre. Além disso, também possibilitam avaliar em que circunstâncias as RNA e o MSM apresentam performances similares.

1.7 Estrutura do trabalho

Esse trabalho está estruturado de forma a, inicialmente, facilitar a compreensão dos temas envolvidos. Assim, na seção 2 está apresentada a fundamentação teórica sendo composta por 4 subtópicos, sendo eles: séries temporais não lineares, redes neurais artificiais, *Markovswitching model* e planejamento de experimentos para simulação.

Na seção 3 é apresentada a metodologia de apoio à decisão desenvolvida para auxiliar na comparação de métodos estatísticos tendo como base a utilização da técnica de DOE. A seção 4 apresenta como se procedeu à aplicação dessa metodologia ao problema proposto neste estudo. Além disso, são apresentados os resultados obtidos e as discussões, elencando-se as vantagens e desvantagens dos métodos, os fatores e como influenciam o desempenho de cada um deles, bem como em qual situação as performances de ambos seria semelhante, cumprindo assim, tanto os objetivos gerais quanto específicos.

Na seção 5 é apresentado um estudo de caso no qual modelou-se por RNA e MSM a série de valores de preço de exportação do Brasil no período de 2008 a 2018 em função dos valores de PIB, preço de importação e taxa de câmbio do país, além da própria série de preço de exportação deslocada de uma unidade no tempo. Esse estudo teve como finalidade verificar os resultados obtidos anteriormente em relação ao método que apresenta menor erro na modelagem de séries temporais não lineares. Finalmente, na seção 6 são apresentadas as principais conclusões do trabalho, juntamente com as contribuições e sugestões para trabalhos futuros.

2. FUNDAMENTAÇÃO TEÓRICA

2.1 Séries temporais não lineares

Modelos de séries temporais lineares têm sido amplamente utilizados nos últimos anos. Segundo Balestrassi *et al.*(2009), uma série temporal puramente estocástica Y_t é dita linear se ela puder ser escrita conforme mostrado na Equação 1.

$$Y_t = \mu + \sum_{i=0}^{\infty} \psi_i \alpha_{t-i} \tag{1}$$

Sendo que μ representa a constante, ψ_i são números reais com $\psi_0 = 1$ e $\{\alpha_t\}$ a sequência de variáveis aleatórias Independente e Identicamente Distribuídas (IID) com uma função de distribuição bem definida. Assume-se que a distribuição de α_t é contínua e $E(\alpha_t) = 0$. Se a série temporal não puder ser bem definida utilizado esse modelo, então ela é dita não linear. Dessa forma, exemplificar o comportamento das séries temporais não lineares não é tão fácil por conta de seu vasto, ou até mesmo ilimitado número de modelos possíveis.

Assim, a preocupação a respeito da não linearidade dos dados têm aumentado consideravelmente, o que resultou no desenvolvimento de modelos não lineares amplamente adotados em pesquisas e nas aplicações de engenharia. O objetivo ao usar tais modelos é capturar uma vasta variedade de comportamentos não lineares de forma a possibilitar maior compreensão do comportamento de sistemas complexos (GAO *et al.*, 2017).

Esses métodos podem ser baseados em redes neurais artificiais (KOCADAGLI; ASIKGIL, 2014); recurrence plot (COSTA et al., 2018); modelos de séries temporais não lineares como: Threshold Autoregressive (TAR), Exponential Autoregressive (EXPAR), Smooth Transition Autoregressive (STAR), e Bilinear Model (BL) (AMIRI,2015); e modelos de troca de regime que podem ser governados por variáveis de estado observáveis (TAR) ou inobserváveis (HAMILTON, 1989).

A modelagem de séries temporais não lineares é aplicada a diversos campos, não se restringindo a apenas uma determinada área de pesquisa. Bonsall *et al.* (2012) caracterizaram a variabilidade no humor de pacientes que sofriam de bipolaridade por meio de uma abordagem não linear baseada em processos autoregressivos. Guan *et al.* (2018) usaram análise de séries temporais não lineares para caracterizar oscilações termo acústicas em motores de combustíveis sólidos de foguetes. O trabalho de Fragkhou, Karakasidis e Nathanail (2018) aplicou análise de séries temporais não lineares para detecção de acidentes no trânsito.

2.2 Redes neurais artificiais

O cérebro humano pode ser entendido como um sistema de processamento de informação altamente complexo, não linear e paralelo. Ele tem a capacidade de reconhecer padrões, por meio da estruturação de seus neurônios, muito mais rápido do que qualquer computador digital existente. Toda essa capacidade se deve ao fato da grande habilidade que o cérebro possui de criar suas próprias regras a partir das inúmeras experiências vividas (HAYKIN, 2009).

Uma das grandes habilidades no processo de aprendizado humano é a regra de aprendizado empírico. Assim como na vida real, as redes neurais artificiais são um tipo de tecnologia computadorizada que aprende a reter conhecimento continuamente por tentativa e erro (CHANG; TSENG, 2017). Dessa forma, assemelham-se ao cérebro já que o conhecimento é adquirido por um processo de aprendizagem a partir do ambiente no qual elas se encontram inseridas e existem forças de conexão entre os neurônios, chamadas de pesos sinápticos, que são utilizadas para reter o conhecimento (HAYKIN, 2009).

Uma rede neural artificial consiste em um arranjo de nós interconectados representados pelos neurônios que são dispostos em diferentes camadas, sendo elas: camada de entrada, camadas escondidas e camada de saída (KUO; FARICHA, 2016). Cada camada é composta por neurônios, sendo que os neurônios em duas camadas adjacentes estão completamente conectados através de pesos respectivos, enquanto que aqueles dispostos em uma mesma camada não possuem nenhuma conexão (BALESTRASSI *et al.*, 2009). Assim, as redes neurais artificiais conseguem mapear uma relação entre os dados de um conjunto de entrada e o de saída.

Um neurônio pode ser entendido como a unidade de processamento de informação para a operação de uma rede neural composta por 3 elementos básicos: sinapse, somador e função de ativação. As sinapses ou elos representam as conexões entre um sinal de entrada e um neurônio, sendo que cada uma é caracterizada por um peso. Já o somador é responsável por realizar a soma dos sinais de entrada multiplicados pelos respectivos pesos sinápticos. Por fim, a função de ativação é responsável por restringir o intervalo de saída de um neurônio, geralmente [0,1] ou [-1,1] (HAYKIN, 2009).

A Figura 5 indica o modelo neuronal anteriormente mencionado, sendo que o *bias* (viés) tem o objetivo de aumentar ou diminuir a entrada líquida da função de ativação.

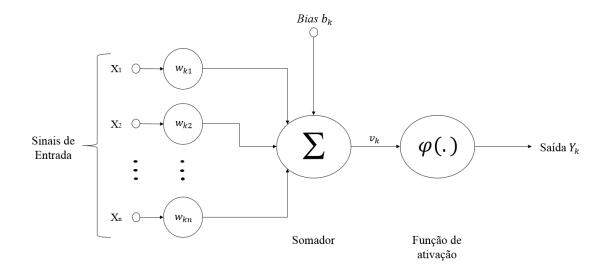


Figura 5. Modelo não linear de um neurônio de índice kFonte: Adaptado de Haykin (2009)

As RNA de acordo com sua arquitetura e interconexão entre seus neurônios podem ser classificadas em 2 classes: redes *feed-forward* ou redes *feed-back*. A rede do tipo *feed-forward*,

também conhecida como estática, consiste em uma função não linear dos valores de entrada e é representada por um conjunto de neurônios interconectados sendo que a informação apenas flui no sentido da camada de entrada para a de saída. Já nas redes do tipo *feed-back*, também conhecidas como dinâmicas, a saída de um neurônio de uma determinada camada pode servir de entrada para o mesmo neurônio ou para neurônios de camadas anteriores (TEALAB;

Conforme mencionado anteriormente, em 1958 se deu o surgimento do modelo Perceptron capaz de imitar a estrutura e a operação do cérebro humano. Posteriormente, na década de 80, um modelo com múltiplos *perceptrons* conectados foi proposto por pesquisadores, doravante chamado de *Multilayer Perceptron* (MLP) (CHANG; TSENG, 2017).

HEFNY; BADR, 2017).

O MLP é um tipo de rede neural artificial com múltiplas camadas de neurônios em uma arquitetura do tipo *feed-forward* e pode ser composto de no mínimo 3 camadas (uma camada de entrada, uma de saída e uma ou mais camadas escondidas). Esse modelo usa função de ativação não linear com cada neurônio em cada camada conectado à camada seguinte com uma certa função peso w_{ij} e um *bias* (viés) b_k (ou intercepto). O MPL utiliza a técnica de aprendizado supervisionado chamada de retropropagação (*backpropagation*) (ZHAI *et al.*, 2016).

Uma rede neural que aprende, a partir de um determinado conjunto de dados de entrada e saída, utilizando a técnica de retropropagação é dita auto adaptativa no sentido de que no

momento do treinamento os valores de w_{ij} são continuamente ajustados de acordo com o feedback obtido a partir do valor real da variável de resposta (MO et al., 2017). Quando essa técnica é escolhida, os pesos e os biases são otimizados baseando-se na minimização da soma de quadrados da diferença entre a variável resposta real e seu valor estimado (BALESTRASSI et al., 2009). O treinamento da rede é realizado até que esta alcance um estado estável, ou seja, quando não existem mais modificações significativas nos valores dos pesos sinápticos (HAYKIN, 2009).

Convém destacar que as RNA, além de apresentarem grande capacidade de aprendizado, também são capazes de generalizar. Isso está relacionado ao fato de que essas redes são capazes de prover boas respostas para valores de entrada que não estavam no grupo de valores utilizados durante o treinamento (HAYKIN, 2009).

A estrutura de uma rede neural artificial do tipo *Multilayer feed-forward* é apresentada na Figura 6.

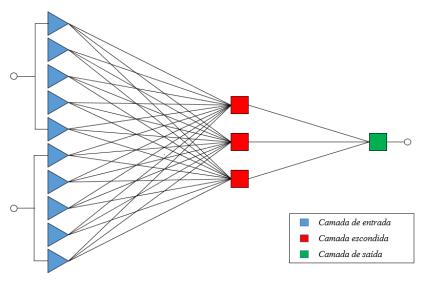


Figura 6. Exemplo da estrutura de uma rede neural

Fonte: O autor

Uma das principais vantagens das redes neurais artificiais é que não é necessário conhecer previamente um modelo matemático que represente o conjunto de dados (CHANG; TSENG, 2017). Assim, são capazes de descrever processos não lineares com uma boa precisão, principal razão para a popularidade desse método já que podem aproximar praticamente qualquer função não linear arbitrariamente (SUN *et al.*, 2017).

Alguns trabalhos científicos (MO et al., 2017; KHANLARI et al., 2012 e MEKANIK et al., 2013) mostram que, em muitos contextos, redes neurais artificiais têm um melhor desempenho quando comparadas com métodos estatísticos convencionais. As RNA também

têm sido utilizadas em processos de previsão, e uma revisão mais específica sobre o tema a partir de uma perspectiva estatística pode ser encontrada em Cheng e Titterington (1994).

2.3 Markov-Switching Model

Muitas variáveis passam por situações em que seu comportamento ao longo do tempo muda drasticamente. Essas mudanças normalmente são originadas de eventos como guerras, crises financeiras ou modificações políticas, e podem ser observadas em quase todas as séries macroeconômicas ou financeiras, desde que, conforme mencionado anteriormente, sejam observadas por um período de tempo suficientemente longo (HAMILTON, 1994).

A modelagem de séries temporais não lineares pelo método conhecido como *Markov-Switching Model*, teve suas bases nos trabalhos de Hamilton (1989), Hamilton (1988) e Hamilton (1990). Inicialmente, o método foi proposto para analisar a taxa de crescimento do Produto Interno Bruto (PIB) americano com um modelo AR alternando entre fases de crescimento e recessão (HAMILTON, 1989).

A mudança no regime, proposta pelo método, não deve ser encarada como uma resposta de um evento determinístico e completamente previsível, mas sim como uma variável aleatória (HAMILTON, 1994). Assim, esse modelo sugere que uma variável resposta num determinado período de tempo t pode ser modelada em k regimes (CAVICCHIOLI, 2014) e a troca entre esses estados é governada por uma matriz de probabilidade de transição. Uma matriz para k=2 é apresentada na Equação 2.

$$P = \begin{bmatrix} P(s_t = 1 \mid s_{t-1} = 1) & P(s_t = 1 \mid s_{t-1} = 2) \\ P(s_t = 2 \mid s_{t-1} = 1) & P(s_t = 2 \mid s_{t-1} = 2) \end{bmatrix} = \begin{bmatrix} p_{11} & p_{21} \\ p_{12} & p_{22} \end{bmatrix}$$
 (2)

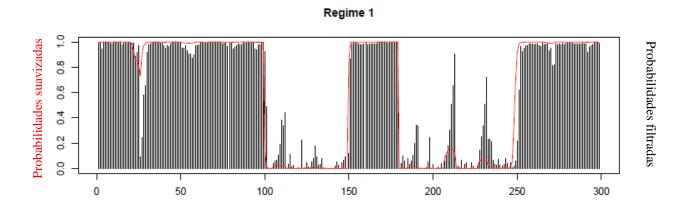
Onde s_t representa a variável de estado inobservável capaz de governar as trocas entre dois regimes seguindo uma cadeia de Markov de primeira ordem. Dessa forma, pode-se generalizar cada elemento dessa matriz pela Equação 3.

$$p_{ij} = P(s_t = j \mid s_{t-1} = i), \forall i, j \in \{1, \dots, k\}$$
(3)

Assim, p_{ij} representa a probabilidade de a série temporal modelada, transitar do regime i para o regime j. Por se tratar de uma probabilidade, os valores possíveis para os elementos dessa matriz sempre variam entre 0 e 1. Dessa forma, a soma de cada coluna deve ser sempre

resultar em 1. Para k=2, o modelo gerado apresentará duas distribuições com duas médias distintas, sendo que uma alta probabilidade na diagonal principal da matriz de probabilidade de transição indica que a série temporal possui baixa tendência em trocar de regime, enquanto que uma alta probabilidade na diagonal secundária relaciona-se a uma forte tendência de alternância entre os regimes.

Um comportamento dinâmico e que é dependente da variável de estado em uma série temporal indica que certas propriedades desta série, tais como variância, média e autocorrelação, são diferentes em regimes distintos (BALESTRASSI *et al.*, 2009). Um exemplo utilizando um conjunto de dados com 300 observações longitudinais obtidas no banco de dados do *software* R é mostrada na Figura 7, representando as probabilidades de uma observação seguir o primeiro ou segundo regime. Assim, o modelo é capaz de identificar e distinguir esses períodos.



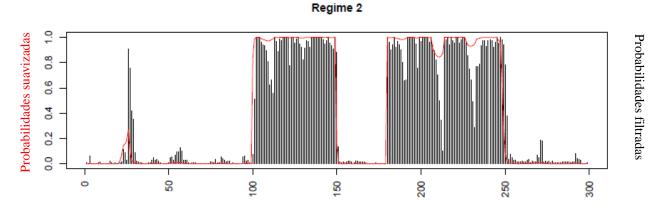


Figura 7. Probabilidades associadas aos regimes dos dados modelados por MSM

O MSM incorpora características da cadeia de Markov com modelos autoregressivos para analisar dados de séries temporais (HAMILTON, 1989). Para estimar os parâmetros do modelo usa-se o algoritmo *Expectation-Maximization* (EM) conforme mostrado em Hamilton

(1990). Esses modelos têm se tornado cada vez mais populares para modelar séries temporais econômicas e financeiras, sendo aplicados à modelagem e previsão de ciclos de negócios, volatilidade em variáveis econômicas e financeiras, taxa de câmbio estrangeira e dinâmicas de inflação (AMIRI, 2015).

2.4 Planejamento de experimentos

O planejamento de experimentos (do inglês *Design of Experiments* – DoE) é uma técnica comumente utilizada quando o pesquisador precisa planejar o menor número de experimentos possível sem que isso cause impactos negativos à sua análise. Envolve uma série de testes nos quais um conjunto de variáveis de entrada ou fatores são variados pelo próprio experimentador de uma maneira controlada. Com isso, podem ser identificadas as causas para as variações sofridas pela resposta analisada (STAICULESCU, 2005).

O DOE permite um profundo entendimento de todos os fatores envolvidos no processo, identificando quais são significativos, como interagem entre si e se as metas almejadas são possíveis dentro das condições da região de estudo. Além disso, o método é bem simples de ser empregado, já que possui um esforço computacional mínimo, sendo normalmente utilizado como uma primeira abordagem para o entendimento de determinado sistema (STAICULESCU, 2005).

Um arranjo fatorial 2^k é muito útil porque fornece a quantidade de experimentos necessária para possibilitar o estudo dos fatores e suas interações em um arranjo completo (MONTGOMERY, 2013; STAICULESCU, 2005). Esse arranjo é o mais simples de todos, sendo que nele cada um dos k fatores possui apenas 2 níveis. Esses níveis, -1 e +1, representam os limites inferior e superior para o intervalo no qual a variável é analisada (STAICULESCU, 2005).

Outra questão que deve ser considerada é o desenvolvimento de um experimento com réplicas, porque a replicação possui duas propriedades importantes. A primeira é que ela permite ao experimentador obter uma estimativa do erro experimental. A outra propriedade é que, se a média amostral (\bar{y}) é utilizada para estimar a média real da população, a réplica permite ao experimentador obter uma estimativa mais precisa desse parâmetros. Por exemplo, considerando σ^2 como a variância de uma observação individual e existem n_r réplicas, então a variância da média amostral será dada pela Equação 4 (MONTGOMERY, 2013).

$$\sigma_{\bar{y}}^2 = \frac{\sigma^2}{n_r} \tag{4}$$

Entre dois experimentos sequenciais apenas um fator irá variar, enquanto os demais serão mantidos fixos. Como consequência, o efeito dessa variável pode ser mensurado pela observação das variações no valor da resposta analisada. É fácil perceber que essa abordagem não é indicada para casos onde se tenha um grande número de variáveis de entrada, já que isso consumirá muito tempo e custo, já que quanto maior o valor de k, maior a quantidade de experimentos necessários (PERRIN $et\ al.$, 2013).

O DOE é uma técnica comumente utilizada para encontrar a solução ótima de determinado processo em estudo (LEE *et al.*, 2007; MCKEIGHEN, 1996; DASCALESCU *et al.*, 2008) ou para descobrir uma solução robusta na qual o objetivo principal é uma boa performance do processo independentemente do cenário ao qual será exposto (WANG *et al.*, 1999; OMEKANDA, 2006; ZHONG; SHI; WU, 2010; QI; OSTERMAN; PECHT, 2009). Contudo, a técnica não é apenas utilizada com esses propósitos, podendo também ser empregada em problemas de simulação (LORSCHEID; HEINE; MEYER, 2012).

Na simulação, o uso do DOE pode ajudar a compreender detalhadamente o comportamento do modelo, levar os pesquisadores a discutirem detalhadamente as implicações das suposições do modelo, além de desafiar ou confirmar expectativas sobre a direção e importância relativa dos efeitos dos fatores. Um arranjo experimental elaborado especificamente para uma determinada situação é muito melhor do que o método de tentativa e erro. Consequentemente, pesquisadores devem compreender que o DOE é parte muito útil e necessária da simulação (BALESTRASSI *et al.*, 2009; LEE *et al.*, 2007).

3. METODOLOGIA

Coleman e Montgomery (1993) apresentam algumas etapas fundamentais para o planejamento de experimentos em simulação, sendo elas:

- a) Estabelecer o problema;
- b) Escolher os fatores e seus respectivos níveis;
- c) Selecionar a variável resposta;
- d) Escolher o tipo de design experimental;
- e) Realizar os experimentos;
- f) Realizar a análise estatística dos dados;
- g) Apresentar as conclusões e recomendações.

Tendo como base essas etapas, desenvolveu-se, para o presente estudo, uma metodologia de apoio à decisão visando a comparação do desempenho de métodos estatísticos distintos, conforme apresentado na Figura 8.

- A1: O planejamento pré-experimental consiste, inicialmente, na identificação do tipo de problema a ser resolvido. As questões mais comuns são categorizadas como classificação, séries temporais, análise de cluster e regressão. Ainda relevante nesta etapa, o pesquisador deve estabelecer as variáveis de entrada e saída para o problema. É necessário destacar que, em um cenário real, a presença de apenas variáveis significativas ajuda na obtenção de melhores modelos. Ao se trabalhar com um conjunto de dados sintético é interessante manter tal característica.
- A2: Este passo consiste na definição de quais métodos serão aplicados. É importante mencionar que, caso pesquisador defina mais de 2 métodos, as comparações deverão ser executadas aos pares. Normalmente, os problemas podem ser resolvidos por mais de um método conforme pode ser observado no Quadro 1.
- A3: Nesta etapa define-se qual o melhor critério para avaliar os desempenhos dos métodos, de acordo com o tipo de problema considerado (univariado ou multivariado). Quando se mede a precisão de modelos univariados podem-se utilizar metas como R^2 , $R^2_{ajustado}$, MAPE, raiz quadrada do erro médio e erro médio quadrático. Para dados multivariados podese utilizar a distância de Mahalanobis e o erro médio quadrado multivariado (do inglês Multivariate Mean Square Error MMSE) proposto por Paiva et al. (2009) como métricas de avaliação. Ressalta-se a importância de escolher uma métrica capaz de avaliar a performance de todos os métodos envolvidos, caso contrário as comparações não poderão ser realizadas.

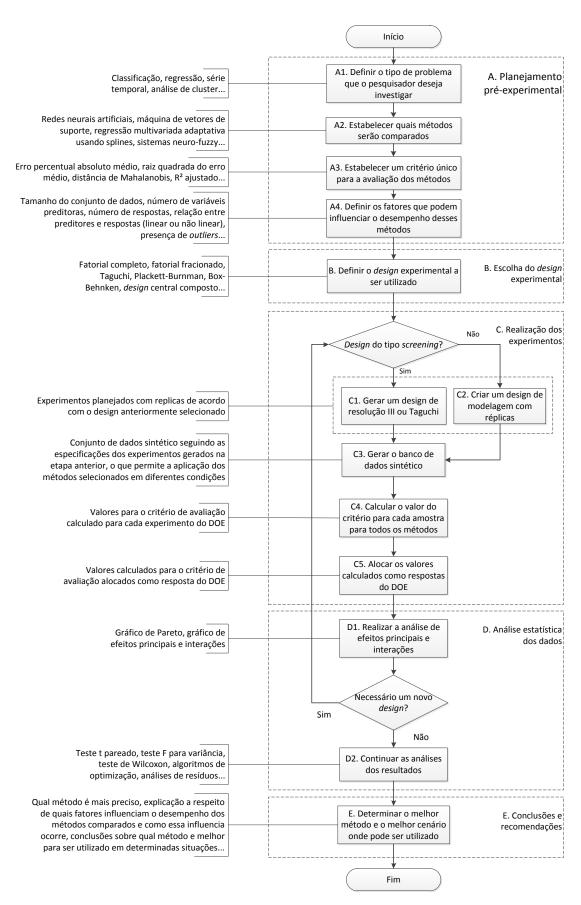


Figura 8. Metodologia de apoio à decisão na análise e comparação de métodos estatísticos

Quadro 1. Métodos comumente aplicados para a solução de problemas estatísticos

Tipo do problema	na Métodos							
	Função de classificação linear							
	Função de classificação quadrática							
	Análise discriminante linear							
	Regressão logística							
Classificação	Análise do vizinho mais próximo							
	Redes neurais artificiais							
	Função densidade de Kernel							
	Máquina de vetores de suporte							
	Árvore de decisão							
	Regressão linear (mínimos quadrados ordinários)							
	Regressão não linear (algoritmo de Gauss-Newton)							
	Regressão usando vetores de suporte							
Regressão	Regressão de Poisson							
	Regressão multivariada adaptativa usando splines							
	Modelos lineares generalizados							
	Redes neurais artificiais							
	Modelo auto-regressivo de média móvel							
	Modelo de média móvel auto-regressivo							
	Modelos de troca de regime Markovianos (<i>Markov-switching models</i>)							
	Modelo de heteroscedasticidade condicional auto-regressiva							
Séries temporais	Modelo generalizado de heteroscedasticidade condicional auto-							
Series temporars	regressiva							
	Redes neurais artificiais							
	Modelo auto-regressivo com limiar							
	Modelo auto-regressivo com transição suave							
	Modelo auto-regressivo exponencial							
	Redes neurais artificiais (Kohonen's Self-Organization Map)							
	Algoritmo K-médias							
	Fuzzy c-médias							
Análise de <i>cluster</i>	Método do centroide							
	Ligação média							
	Ligação única							
	Método Ward							

- A4: Este é provavelmente o passo mais importante de toda a metodologia. O pesquisador, baseado na literatura e em seus próprios conhecimentos, deve determinar quais os fatores possuem certa influência no desempenho dos métodos. O Quadro 2 mostra alguns fatores que podem ser significativos para cada tipo de problema considerado.
- **B:** Após definidos os fatores, é necessário escolher qual o tipo de *design* experimental utilizar. Ao trabalhar com muitos fatores, recomenda-se executar um *design* de Taguchi ou algum com resolução III, de forma a identificar quais deles são significativos, reduzindo, assim, o número de fatores a serem considerados. Ao se trabalhar com um pequeno número de fatores, recomenda-se um fatorial completo com réplicas, ou um fatorial fracionado com resolução

superior a III. Uma síntese dos possíveis *designs* extraída do software Minitab ® pode ser visto na Figura 9.

Quadro 2. Fatores que podem ser significativos em cada tipo de problema

Tipo do problema	Métodos							
	Tamanho do conjunto de dados							
	Distribuição seguida pelas variáveis de entrada							
	Muticolinearidade							
Classificação	Relação (linear ou não linear) entre as variáveis							
	Tipo de distribuição (uni ou multimodal)							
	Natureza do problema (estático ou dinâmico)							
	Proporções amostrais (iguais ou diferentes)							
	Tamanho do conjunto de dados							
	Número de variáveis preditoras e respostas							
	Multicolinearidade							
Regressão	Variância dos dados							
	Relação entre as variáveis (linear ou não linear)							
	Distribuição das variáveis independentes							
	Distribuição das variáveis resposta							
	Tamanho do conjunto de dados							
	Correlação entre as séries							
Séries temporais	Número de variáveis preditoras e nas respostas							
	Autocorrelção nas variáveis preditoras e nas respostas							
	Relação entre as variáveis (linear ou não linear)							
	Número de variáveis							
	Número de clusters							
	Tamanho do conjunto de dados							
Análise de <i>cluster</i>	Quantidade de dados nos <i>clusters</i> (iguais ou diferentes)							
	Nível de sobreposição (overlapping)							
	Presença de <i>outliers</i>							
	Correlação entre as variáveis							

	Número de fatores													
Número de ensaios	2	3	4	5	6	7	8	9	10	11	12	13	14	15
4	Com	Ш												
8		Com	IV	Ш	Ш	Ш								
16			Com	V	IV	IV	IV	III	Ш	Ш	Ш	Ш	Ш	Ш
32				Com	VI	IV	IV	IV	IV	IV	IV	IV	IV	IV
64					Com	VII	V	IV	IV	IV	IV	IV	IV	IV
128						Com	VIII	VI	V	V	IV	IV	IV	IV

Figura 9. Designs disponíveis no software Minitab e suas respectivas resoluções

Uma determinada resolução X indica que um efeito principal de determinado fator poderá estar confundido com uma interação de ordem X-1. Assim, experimentos com resolução III perdem informações relativas a essas interações, já que um efeito principal pode

ser confundido com uma interação de segunda ordem, o que em muitos caso não deve ser negligenciada.

- C1: Considerando que um design escolhido no passo anterior tenha resolução III, o experimentador deve entrar com os fatores selecionados para gerar os experimentos.
- C2: Caso o design escolhido não possua resolução III, então o experimentador deve entrar com os fatores selecionados e gerar os experimentos baseando-se em um design com resolução superior.
- C3: Cada linha do design experimental representa uma amostra diferente que o experimentador deverá gerar sinteticamente seguindo a configuração estabelecida pelos fatores. É recomendável adicionar alguma variação aleatória aos dados de modo a aproximar o conjunto de dados sintéticos de um conjunto real.

A variação depende do tipo de problema, por exemplo, para um problema de regressão, pode-se aleatoriamente gerar valores das variáveis de entrada e então estabelecer uma relação baseada em uma função específica, linear ou não linear, para determinar os valores das respostas. Em seguida, adiciona-se um erro aleatório a esses valores, transformando o problema determinístico em estocástico. O mesmo procedimento poderia ser aplicado a um problema de séries temporais, entretanto neste caso as observações dependeriam do tempo.

- C4: Um dos principais resultados desta metodologia é o conjunto de dados sintético gerado. A utilização da ferramenta DOE permite que esses dados englobem diversas características encontradas em problemas reais, de modo a permitir a generalização dos cenários trabalhados.
- C5: Nesse momento, é necessário resolver o problema utilizando os algoritmos previamente selecionados. O experimentador deve aplicar os métodos escolhidos ao mesmo conjunto de dados. É possível calcular a métrica selecionada e atribuir seu valor como resposta do DoE.
- D1: Nessa etapa, deve-se realizar a análise de efeitos principais para determinar a necessidade de executar um novo design. A resposta será positiva, caso inicialmente tenha-se trabalhado com grande número de fatores e tenha-se realizado um design do tipo screening. Assim, removendo-se os fatores não significativos, retorna-se ao passo C e executa-se novamente o fluxo. Executa-se esse loop até que não seja mais necessária a execução de um screening design.
- D2: Uma vez que se consiga atingir esse cenário, podem-se executar diversas ações como:

- Compara se o desempenho de um método supera o dos demais por meio de um teste t pareado;
- Executar a análise de efeitos principais e das interações;
- Analisar os resíduos para verificar se o modelo gerado com a análise do DOE é satisfatório ou não;
- Analisar os resíduos gerados com a aplicação dos métodos (em problemas e regressão e séries temporais, por exemplo). Assim, pode-se avaliar a normalidade, presença de oscilação, misturas, agrupamentos e tendências;
- Utilizar algoritmos de otimização para identificar em que circunstâncias os métodos possuem desempenhos similares, ou para maximizar o desempenho de um deles.

E: O último passo consiste em extrair as principais informações do passo anterior para determinar o melhor método em diferentes condições. Esses resultados são de extrema importância, já que constituem a base para estudos futuros.

4. ESTUDO COMPARATIVO

De forma a apresentar claramente as etapas do estudo comparativo, essas foram apresentadas de acordo com a metodologia descrita anteriormente.

4.1 Planejamento pré-experimental

A1: Problema de séries temporais

Como o primeiro passo, definiu-se que o problema alvo desse estudo é a modelagem de séries temporais não lineares. Objetiva-se modelar o comportamento de uma série (y_1) em função de seus próprios valores deslocados de uma unidade no tempo (y_{t-1}) e de uma segunda série temporal distinta (y_2) .

A2: Markov-switching model vs Redes neurais artificiais

Em seguida, definiram-se os métodos MSM e RNA como aqueles a serem investigados nesse estudo. Ambos os métodos são adequados para trabalhar com séries temporais não lineares por razões já mencionadas.

Para a modelagem das séries temporais usando MSM optou-se por considerar o número de regimes k=2 já que essa abordagem já se mostrou eficaz em estudos já realizados (MAAZIZ; KHARFOUCHI, 2018; GELATI; MADSEN; ROSBJERG, 2010; PARK; HONG, 2013; HSIEH; CHEN, 2017; BALCILAR *et al.*, 2017). Além disso, definiu-se o termo autoregressivo deslocado de apenas uma unidade no tempo, assim como nas redes neurais, o que possibilita uma comparação justa entre os dois métodos.

Foi utilizado o pacote MSwM do *software* R versão 1.4 que utiliza o algoritmo *Expectation Maximization* (EM) para ajustar um modelo Markoviano aos dados apresentados. O pacote exige que incialmente seja ajustado um modelo de regressão linear aos dados usando a função lm demonstrada na Equação 5. Nele o termo Mod representa o modelo ajustado, o conjunto $Dados_1$ é a resposta, e o conjunto $Dados_2$ a entrada, ou seja, $f_1(t)$ e $f_2(t)$, ambas acrescidas dos valores de erro aleatório.

$$Mod = lm(Dados_1 \sim Dados_2) \tag{5}$$

Em seguida, utilizou-se a função *msmFit* mostrada na Equação 6 para ajustar o modelo de troca de regimes a partir do modelo linear previamente elaborado, sendo que o termo *Swmod* representa o modelo final. Essa função possui ao todo 7 argumentos explicados no Quadro 3.

Quadro 3. Argumentos da função msmFit

Argumento	Significado
objeto	Trata-se de um modelo linear, um modelo linear generalizado, ou uma fórmula. Para o presente estudo, refere-se ao <i>Mod</i> anteriormente gerado, ou seja ao modelo de regressão linear gerado a partir das amostras do banco de dados sintético.
k	Quantidade de regimes que o modelo deverá apresentar.
sw	Um vetor lógico (TRUE ou FALSE) indicando quais os coeficientes serão variáveis.
p	Valor inteiro que indica o número de coeficientes auto-regressivos que o modelo deverá apresentar.
dados	Uma tabela opcional contendo os dados. Caso não seja fornecido esse argumento, os dados serão extraídos do próprio objeto.
família	Um caractere indicando a qual família pertence o modelo. Só é necessário quando o objeto se trata de um modelo linear generalizado.
controle	Lista de parâmetros de controle.

Conforme mencionado anteriormente, essa pesquisa não possui como objetivo aprofundar as demonstrações matemáticas por trás desse método, devendo, aquele que sentir necessidade, recorrer à publicação de Hamilton (1994).

Em relação à modelagem por RNA, utilizou-se o *software* Statistica na sua versão 7.0. Nas Figuras de Figura 10 a Figura 15 é possível observar as configurações do Statistica para a realização dessa modelagem.

Inicialmente o usuário escolhe qual será o critério de parada, podendo ser por meio da quantidade de redes testadas ou da quantidade de horas/minutos que o algoritmo será executado. Também é possível selecionar a quantidade de redes a serem retidas ao final do processo de treinamento. Nesse trabalho optou-se por treinar as redes durante 1 minuto e ao fim reter as 5 melhores, conforme Figura 10.

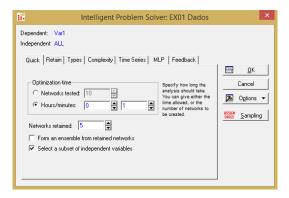


Figura 10. Tela do IPS para definir critério de parada e quantidade de redes neurais retidas

Em seguida, deve-se também escolher qual o critério de seleção das redes neurais. No presente estudo, conforme mostrado na Figura 11, foram escolhidas como melhores redes aquelas que possuíam os menores valores de erro. O *IPS* utiliza como padrão a soma dos quadrados para calcular o erro, ou seja, a soma dos quadrados das diferenças entre as respostas esperadas e as respostas ajustadas.

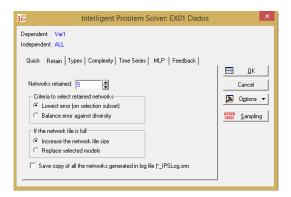


Figura 11. Tela do IPS para definir o critério de escolha da melhor rede neural

Utilizou-se como tipo padrão de rede neural para a modelagem das séries apenas a *Three Layer Perceptron*, Figura 12, por apresentar excelente desempenho ao lidar com conjuntos de dados não lineares e por MLP com retropropagação ser um dos tipos de RNA mais comumente utilizados (OLSON; DELEN; MENG, 2012; AIZENBERG *et al.*, 2016).

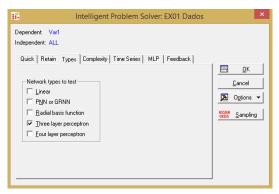


Figura 12. Tela do IPS para definição dos tipos de rede testadas

No próximo passo deve-se determinar o intervalo relacionado à quantidade de neurônios da camada oculta da rede *Three Layer Perceptron*. Quanto maior o número dessas unidades, maior o tempo de processamento para realizar o treinamento da rede e também maior é a probabilidade de ocorrência de sobreajuste (*overfitting*), ou seja, a rede neural começa a decorar as informações e perdendo significativamente sua capacidade de generalização. Nesse caso, optou-se por utilizar o intervalo padrão do software, ou seja, de 1 a 8 unidades, conforme podese observar na Figura 13.

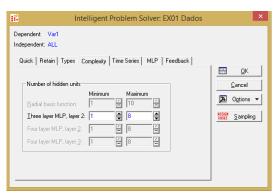


Figura 13. Tela do IPS para definição da quantidade de neurônios na camada oculta

Como o problema deve ser encarado como uma série temporal, deve-se escolher qual o intervalo relacionado ao valor de defasagem da variável de entrada. Conforme pode ser visto na Figura 14, optou-se por usar o intervalo de 1 a 2, ou seja, serão considerados como entradas apenas o valor atual da série em um determinado instante t, e um valor anterior, ou seja, t-1, o que também foi feito no caso do método MSM, conforme mencionado anteriormente.

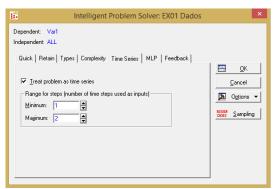


Figura 14. Tela do IPS para definição da quantidade de termos autoregresivos

Por fim, pode-se escolher duas formas para codificação da resposta, linear ou logística, conforme Figura 15. A linear utiliza a função de ativação identidade, enquanto que a logística utiliza a função de ativação logística. Nesse trabalho ambas as abordagens foram selecionadas.

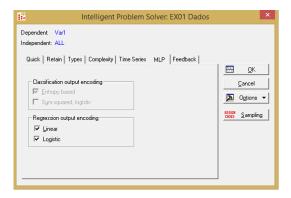


Figura 15. Tela do IPS para definição do tipo da função de ativação

Assim, como para o caso anterior, esse estudo também não irá demonstrar profundamente o funcionamento de uma rede neural artificial, devendo, portanto, os leitores que tiverem necessidade de uma maior compreensão sobre o tema, recorrer a Haykin (2009).

A3: Erro Percentual Absoluto Médio

Escolheu-se a métrica MAPE para avaliar qual dos métodos possui melhor desempenho. O cálculo do MAPE pode ser observado na Equação 7, sendo y_t o valor real da série, \hat{y}_t o valor ajustado pelo modelo e T o valor da quantidade total de dados das séries. Destaca-se que é imprescindível escolher a mesma métrica para avaliação de ambos os métodos, o que irá possibilitar o desenvolvimento das análises futuras. Além disso, a escolha dessa métrica foi importante pelo fato de ela ser relativa, já que erros absolutos poderiam prejudicar a interpretação dos resultado.

$$MAPE(\%) = \frac{1}{T} \sum_{t=1}^{T} \left| \frac{y_t - \widehat{y_t}}{y_t} \right| \times 100$$
 (7)

A4: Fatores possivelmente significativos: desvio padrão do erro aleatório adicionado às séries temporais originais, tamanho do conjunto de dados e correlação entre as séries temporais.

Consideraram-se alguns fatores capazes de influenciar a habilidade de modelagem de séries temporais por RNA e MSM, tais como: desvio padrão do erro aleatório acrescido aos valores de $f_1(t)$ e $f_2(t)$, correlação entre as séries e tamanho do conjunto de dados. Geralmente, apenas dois níveis são necessários para que seja possível avaliar como cada um desses fatores afetam a variável resposta. Dessa forma, foram definidos os níveis para cada um deles variando entre um nível baixo e um alto representados por -1 e +1, respectivamente.

Como as séries serão geradas por meio de funções não lineares conhecidas, o erro aleatório adicionado tem como objetivo transformar o problema determinístico em estocástico, conforme mencionado. Esse erro possui média zero em todos os casos, entretanto o desvio padrão varia ao longo dos experimentos. Esse desvio foi baseado nas considerações de Cohen (1988) a respeito do chamado *Effect Size Index:d.* Assim, segundo o autor, um pequeno efeito é aquele dado por $0.2 \times \sigma_{y_i}$ com i=1 e 2, sendo que, neste estudo, σ_{y_i} representa o desvio padrão da série temporal antes de ser acrescida do erro aleatório, enquanto grande efeito apresenta é dado por $0.8 \times \sigma_{y_i}$. O processo de geração desses erros, bem como o próprio banco de dados sintético podem ser vistos mais detalhadamente no Apêndice A deste trabalho.

Em relação aos valores de correlação entre as séries consideraram-se valores entre 0,20 e 0,40 como uma fraca correlação e valores entre 0,80 e 0,95 como uma forte correlação. É importante destacar que em ambos os casos considerou-se valores de *p-value* significativos de forma a garantir que a correlação seja significativa, entretanto podendo ser forte ou fraca.

Uma vez que se deseja modelar os valores de uma série temporal (y_1) como função de uma outra série temporal (y_2) e de si própria, mas deslocada de uma unidade no tempo (y_{t-1}) é recomendado o uso de no mínimo 20 valores de dados para cada variável de entrada do problema. Assim, nesse caso uma amostra de ao menos 40 valores faz-se necessária, definindose, então, os níveis -1 e +1 para o fator tamanho do conjunto de dados como sendo 40 e 80 dados, respectivamente.

Os fatores considerados neste estudo e os seus respectivos níveis estão sintetizados na Tabela 1, sendo que σ_{y_i} representa o desvio padrão das séries originais, ou seja, sem adição do erro aleatório, com i=1 e 2, ρ_{ij} indica a correlação entre as duas séries temporais consideradas, com i e j=1,2...16, σ_{ε} representa o desvio padrão do erro aleatório adicionado às séries temporais originais e finalmente T representa o tamanho do conjunto de dados.

Tabela 1. Fatores considerados no estudo e seus respectivos níveis

Fator	Nível -1	Nível +1
$\sigma_{\varepsilon}(A)$	$0.2 \times \sigma_{y_i}$	$0.8 \times \sigma_{y_i}$
$ ho_{ij}\left(oldsymbol{B} ight)$	Baixa	Alta
T(C)	40	80

4.2 Escolha do design experimental

B: Fatorial completo

Assim, como se tem um pequeno número de fatores, foi possível utilizar um arranjo fatorial completo 2³ com duas réplicas, resultando em 16 experimentos. Cada um deles gerou uma amostra de dados de acordo com os níveis dos fatores considerados e todas as 16 amostras do banco de dados foi modelada usando os dois métodos RNA e MSM.

4.3 Realização dos experimentos

C2: Experimentos (23) com 2 réplicas

Na Tabela 2 estão apresentados os 16 experimentos, sendo que de acordo com os valores dos fatores em cada um deles gerou-se o banco de dados sintéticos com 32 séries temporais que foram modeladas e cujos resultados estão nas próximas etapas desse trabalho.

Tabela 2. Experimentos realizados na pesquisa

Número do experimento	$\sigma_{arepsilon}$	$ ho_{ij}$	T
1	$0.2 \times \sigma_{y_i}$	Baixa	40
2	$0.8 imes \sigma_{y_i}$	Baixa	40
3	$0.2 \times \sigma_{y_i}$	Alta	40
4	$0.8 imes \sigma_{y_i}$	Alta	40
5	$0.2 imes \sigma_{y_i}$	Baixa	80
6	$0.8 imes \sigma_{y_i}$	Baixa	80
7	$0.2 imes \sigma_{y_i}$	Alta	80
8	$0.8 imes \sigma_{y_i}$	Alta	80
9	$0.2 imes \sigma_{y_i}$	Baixa	40
10	$0.8 imes \sigma_{y_i}$	Baixa	40
11	$0.2 imes \sigma_{y_i}$	Alta	40
12	$0.8 imes \sigma_{y_i}$	Alta	40
13	$0.2 \times \sigma_{y_i}$	Baixa	80
14	$0.8 imes \sigma_{y_i}$	Baixa	80
15	$0.2 imes \sigma_{y_i}$	Alta	80
16	$0.8 \times \sigma_{y_i}$	Alta	80

C3: Banco de dados sintético para o problema de series temporais não lineares

Para a criação do banco de dados sintético, foram selecionadas duas funções não lineares no catálogo de funções para regressão não linear do software Minitab. Essas funções estão indicadas nas Equações 8 e 9 sendo que t representa a variável tempo e θ_k representa os parâmetros das funções.

$$f_1(t) = 1 - \frac{1}{t^{\theta_1}} \tag{8}$$

$$f_2(t) = \theta_2 \exp(\theta_3 t) \tag{9}$$

Essas funções foram escolhidas por assumirem diversos comportamentos de acordo com o valor assumido por θ_k . Para um conjunto com 80 são apresentados na Figura 16 os dois comportamentos distintos para a primeira função, sendo que a série 1 representa a $f_1(t)$ com um valor positivo de θ_1 e a série 2 apresenta o formato da mesma função quando o mesmo parâmetro assume um valor negativo.

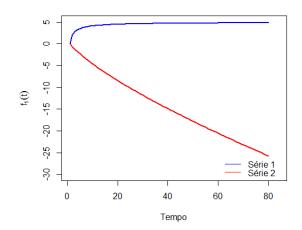


Figura 16. Comportamento da $f_1(t)$ para valores de θ_1 positivos e negativos

A segunda função, por apresentar um maior número de parâmetros, apresenta maior possibilidade de comportamentos conforme mostrado na Figura 17. As séries temporais de 1 a 4 representam o formato de $f_2(t)$ para quando θ_2 e θ_3 são: ambos positivos, ambos negativos, positivo e negativo e negativo e positivo, respectivamente.

Apresentadas as funções, torna-se possível obter os valores de correlação desejados entre as séries. Adotaram-se os valores de 0,30 e 0,90 como sendo os alvos em experimentos nos quais almejava-se uma correlação baixa e alta entre as séries, respectivamente. O objetivo foi reduzir a distância entre a real correlação entre as séries (ρ) e o alvo estabelecido, sendo

que para isso, foi necessário variar os parâmetros θ_k das funções que darão origem às séries temporais.

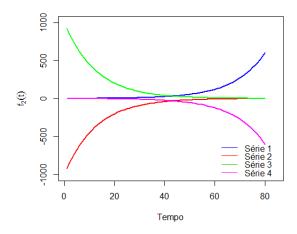


Figura 17. Comportamento da $f_2(t)$ para valores de θ_2 e θ_3 positivos e negativos

Dessa forma, a partir da solução do problema de minimização mostrado na Equação 10, será possível obter os valores dos coeficientes para as funções que fornecem a correlação desejada entre as séries. Acrescenta-se uma restrição de que os valores de θ_k das funções anteriormente mostradas devem estar dentro do intervalo de -10 e 10 para evitar que esses parâmetros assumam valores extremamente altos em módulo para os coeficientes.

$$\min_{\rho \in \Omega} (\rho - alvo)^2$$

$$S. a. -10 \le \theta_i \le 10, i = 1, 2e 3$$
 (10)

C4 e C5: Valor do erro percentual absoluto médio para cada experimento do DOE

Após realizar a modelagem de cada um das séries do banco de dados sintéticos, os valores de MAPE para cada experimento estão elencados na Tabela 3. A resposta diferença é simplesmente o valor da seguinte subtração (MAPE MSM – MAPE RNA), o que possibilitará a comparação entre os métodos, já que quanto maior a diferença melhor a performance da RNA. Esse indicador possui grande importância durante a realização a análise dos dados pelo *Reponse Optimizer* do Minitab que será demonstrada mais adiante.

Tabela 3. Resultados dos experimentos

Número do	NA A DIE NACNA	MADE DNA	DIEEDENICA
experimento	MAPE MSM	MAPE RNA	DIFERENÇA
1	1,8%	0,1%	1,7%
2	6,8%	0,1%	6,7%
3	3,2%	0,1%	3,1%
4	6,7%	0,1%	6,6%
5	2,0%	0,0%	1,9%
6	8,7%	0,0%	8,7%
7	3,7%	0,1%	3,6%
8	6,8%	0,1%	6,7%
9	3,3%	0,2%	3,1%
10	5,7%	0,1%	5,6%
11	3,5%	0,1%	3,4%
12	4,7%	0,2%	4,5%
13	1,7%	0,1%	1,6%
14	7,3%	0,1%	7,2%
15	3,5%	0,1%	3,4%
16	8,1%	0,2%	7,9%

4.4 Análise estatística dos dados

D1: Gráfico de Pareto, efeitos principais e interações

Conforme mencionado anteriormente, para melhor explorar o desempenho dos métodos de modelagem estudados, faz-se necessário identificar quais fatores de fato influenciam na performance de cada um deles e como ocorre essa influência. Para isso utilizou-se o *software* Minitab ® versão 18 para análise do DOE. Nos próximos tópicos estarão descritas as análises realizadas para cada uma das respostas sendo elas MAPE MSM, MAPE RNA e DIFERENÇA, além dos principais resultados da análise dos resíduos para cada um dos métodos em cada um dos 16 experimentos.

4.4.1 Análise estatística para a resposta MAPE MSM

A análise de variância realizada para esse caso pode ser observada na Tabela 4, sendo que os termos A, B e C representam os fatores anteriormente mencionados: $\sigma_{\varepsilon}(A)$, $\rho_{ij}(B)$ e

T (C), respectivamente. Nota-se que não há não há indícios de falta de ajuste no modelo, pois esse termo apresenta p-value maior que 0,05. Nota-se também que os fatores, vistos também como fontes de variação, que possuem um valor de p-value menor que 0,05 exercem influência significativa sobre a resposta MAPE MSM. Assim, pode-se observar que os fatores σ_{ε} e T, a interação entre σ_{ε} e T e a interação entre σ_{ε} e ρ_{ij} exercem influência significativa.

Tabela 4. Análise de variância do modelo MAPE MSM

Fonte	GL	SQ (Aj.)	QM (Aj.)	Valor F	Valor-P
Modelo	5	0,007516	0,001503	26,17	0,000
Linear	3	0,006773	0,002258	39,31	0,000
A	1	0,006491	0,006491	113,02	0,000
В	1	0,000052	0,000052	0,90	0,365
C	1	0,000230	0,000230	4,00	0,073
Interações de 2 fatores	2	0,000743	0,000372	6,47	0,016
A*B	1	0,000342	0,000342	5,96	0,035
A*C	1	0,000401	0,000401	6,98	0,025
Erro	10	0,000574	0,000057		
Falta de ajuste	2	0,000039	0,000020	0,29	0,754
Erro Puro	8	0,000535	0,000067		
Total	15	0,008091			

A partir da análise de variância obtém-se o valor do quadrado médio dos resíduos e dessa forma é possível realizar a próxima análise. Para decidir quais fatores são significativos, é realizada a análise dos coeficientes que irão acompanhar cada um deles no modelo final.

Na Tabela 5 nota-se que os coeficientes associados aos termos A, AB, AC e a constante do modelo possuem *p-value* menor que 0,05, sendo portanto considerados significativos.

Tabela 5. Análise dos coeficientes do modelo MAPE MSM

Termo	Efeito	Coef	EP de Coef	Valor-T	Valor-P	VIF
Constante		0,04843	0,00189	25,56	0,000	
${f A}$	0,04028	0,02014	0,00189	10,63	0,000	1,00
В	0,00360	0,00180	0,00189	0,95	0,365	1,00
C	0,00758	0,00379	0,00189	2,00	0,073	1,00
A*B	-0,00925	-0,00463	0,00189	-2,44	0,035	1,00
A*C	0,01001	0,00501	0,00189	2,64	0,025	1,00

Na Figura 18 pode-se observar o gráfico de Pareto para os fatores que influenciam a resposta MAPE MSM e o gráfico da distribuição de probabilidade associada. Esses gráficos

indicam de forma mais visual quais os fatores que afetam o desempenho do MSM por meio da identificação de quais possuem efeitos maiores que os ruídos.

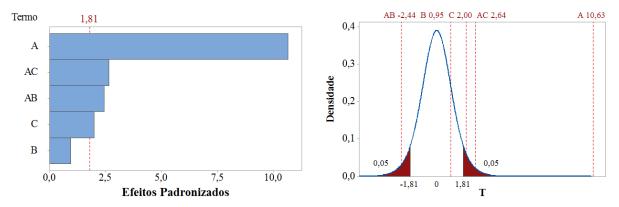


Figura 18. Gráfico de Pareto dos efeitos padronizados para a resposta MAPE MSM e gráfico de distribuição de probabilidade associado

Para a análise do DOE utilizou-se a técnica de eliminação *backward* visando eliminar termos que não contribuem significativamente para o modelo, que quando presentes diminuem o valor do R^2 ajustado, indicador que será detalhado mais adiante. Fixou-se um valor para alfa igual a 0,10, conforme padrão do Minitab, ou seja, caso o *p-value* de um determinado fator ou interação for maior que alfa, ele é eliminado do modelo e seu efeito é considerado como sendo, na verdade, um ruído. Dessa forma, o valor de 1,81 representa o valor crítico associado à distribuição t. Assim, fatores cujos valores de efeitos padronizados são maiores que o t crítico são considerados como significativos. A quantidade de graus de liberdade da distribuição é obtida pela análise de variância mencionada anteriormente, sendo portanto equivalente aos graus de liberdade do erro.

O modelo obtido para o valor do MAPE MSM em unidades codificadas é dado pela Equação 11. Convém destacar que mesmo o coeficiente de B não sendo significativo ele é mantido para que se tenha um modelo hierárquico.

$$MAPE\ MSM = 0.04843 + 0.02014\ A + 0.00180\ B + 0.00379\ C$$

- 0.00463\ A * B + 0.00501\ A * C (11)

Uma das formas de avaliar se o modelo criado realmente representa o conjunto de dados que lhe deu origem é por meio do indicador denominado coeficiente de determinação (R²). Quando maior o valor de R² maior o indício de que o modelo se ajusta bem aos dados históricos indicando o percentual da variação nos dados que é explicado pelo modelo. Seu cálculo pode ser obtido por meio da Equação 12 (MONTGOMERY, 2008).

$$R^{2} = 1 - \frac{\sum_{t=1}^{T} e_{t}^{2}}{\sum_{t=1}^{T} (y_{t} - \bar{y})^{2}}$$
 (12)

Entretanto, como a soma de resíduos reduz significativamente quando novos termos são adicionados ao modelo, então ter esse indicador como única medida para o desempenho pode levar a ocorrência de *overfitting* fazendo com que o modelo perca sua capacidade de generalização. Uma alternativa a esse indicador é o R² ajustado já que seu valor torna-se cada vez menor a medida que mais termos não significativos vão sendo acrescentados ao modelo. O valor do R² ajustado pode ser obtido por meio da Equação 13 sendo *p* o número de termos no modelo (MONTGOMERY, 2008).

$$R^{2} ajustado = 1 - \frac{\frac{\sum_{t=1}^{T} e_{t}^{2}}{(T-p)}}{\frac{\sum_{t=1}^{T} (y_{t} - \bar{y})^{2}}{(T-1)}}$$
(13)

Outro indicador muito importante é o R² de previsão já que indica o quão bem o modelo é capaz de prever observações que não estão contidas no conjunto de dados que o originou. Assim, é muito comum quando o objetivo é comparar dois modelos distintos e suas habilidades em realizar previsão. Seu valor é obtido removendo dados do conjunto fornecido e recalculando a equação que relaciona as variáveis de entrada com as de saída, avaliando assim quão bem o modelo é capaz de prever o ponto faltante. O processo é repetido para todos os pontos do conjunto de dados.

Para o caso do modelo do MAPE MSM, os valores de R², R² ajustado e R² de previsão são, respectivamente, iguais a 92,90%, 89,35% e 81,83%. Como todos esses indicadores apresentaram valores maiores que 80%, pode-se considerar que o modelo para o MAPE MSM produz previsões confiáveis, aumentando, assim, a confiabilidade nos resultados.

Além desses indicadores é importante avaliar a normalidade dos resíduos gerados pelo modelo. Quando os resíduos são normalmente distribuídos, tem-se um forte indicativo de que o modelo é capaz de representar bem os dados. Assim, realizou-se o teste de normalidade dos resíduos, conforme a Figura 19, obtendo-se um valor de *p-value* igual a 0,440, o que indica que a hipótese nula relacionada à normalidade dos resíduos não deve ser rejeitada.

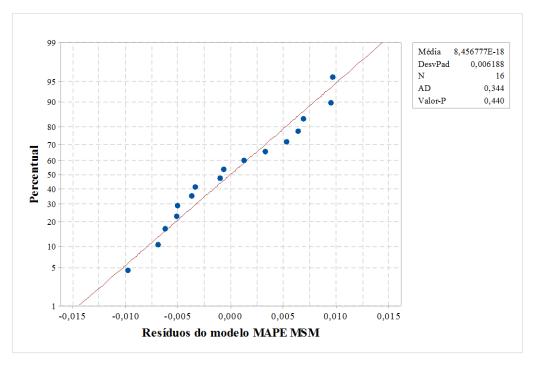


Figura 19. Gráfico de probabilidade para os resíduos do modelo MAPE MSM

No próximo passo, deve-se entender como esses fatores afetam o desempenho do método. Recorreu-se novamente ao *software* Minitab ® versão 18 para a elaboração do gráfico de efeitos principais, conforme mostrado na Figura 20. Nessas análises foram considerados apenas os termos contidos no modelo.

É possível notar que o método MSM possui melhor desempenho quando tem-se um menor valor de desvio padrão associado ao erro aleatório que é adicionado à série temporal, ou seja, quando a série que está sendo modelada é mais comportada ao longo do tempo, sem muitas variações aleatórias. Além disso, também apresenta melhor desempenho quando trabalha com um menor conjunto de dados. A correlação entre as séries, quando considerada individualmente, possui um efeito muito pequeno (apenas 0,00360) na variação da média, ilustrando assim o fato de ter sido considerada não significativa anteriormente e essas variações tidas como aleatórias.

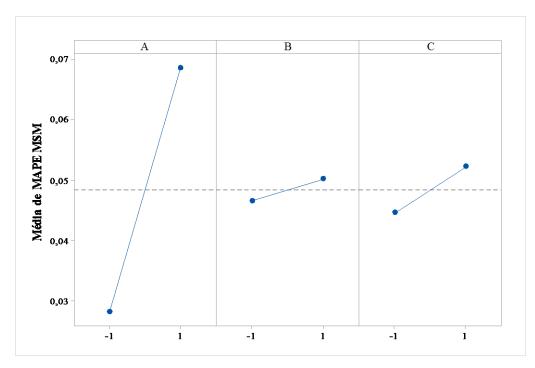


Figura 20. Gráfico de efeitos principais para a resposta MAPE MSM

Entretanto, mesmo que os fatores individualmente não exerçam nenhuma influência sobre o desempenho do método, suas interações podem ser significativas, conforme analisado. Assim a Figura 21 mostra o gráfico de interações entre os fatores do modelo.

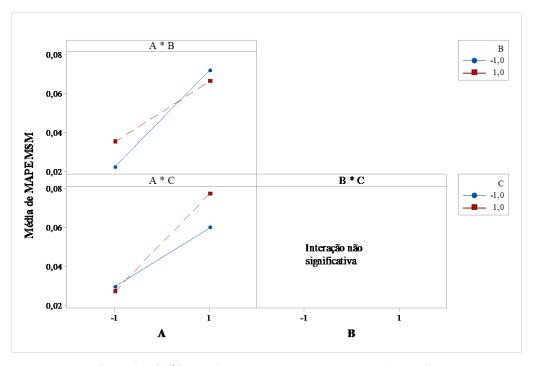


Figura 21. Gráfico de interações para a resposta MAPE MSM

Pode-se observar que quando tem-se um baixo desvio padrão associado ao erro aleatório adicionado à série o efeito da correlação será maior, sendo que uma alta correlação entre as

séries nessas circunstâncias significa um valor mais alto da resposta MAPE MSM. Em relação à interação entre σ_{ε} e T, o efeito do tamanho da amostra será maior quanto maior for o valor do desvio padrão do erro aleatório. A interação entre ρ_{ij} e T não aparece no gráfico, pois não é significativa.

4.4.2 Análise estatística para a resposta MAPE RNA

Em relação ao método RNA, também realizou-se a análise dos valores de MAPE obtidos em cada experimento. Entretanto, ao executar a análise do DOE todos os fatores apresentaram um *p-value* maior que 0,05, podendo concluir, então, que nenhum dos fatores analisados nem suas interações são capazes de influenciar o desempenho da rede neural, portanto todas as variações nos valores do MAPE RNA são atribuídas a causas aleatórias.

Caso seja utilizada a técnica de eliminação *backward* com um alfa igual a 0,10, conforme feito para o caso anterior, todos os termos do modelo são removidos, pois os resultados apresentados pela rede neural apresenta baixa variabilidade que não pode ser atribuída aos fatores, assim não é possível apresentar a análise de variância. Dessa forma, optouse por mostrar aqui a análise de variância e análise dos coeficientes codificados nas Tabela 6 e Tabela 7, respectivamente, sem o uso dessa técnica para que possa ser demonstrado que de fato nenhum dos fatores foi significativo.

Tabela 6. Análise de variância do modelo MAPE RNA

Fonte	GL	SQ(Aj.)	QM(Aj.)	Valor-F	P-Value
Modelo	6	0,000002	0,000000	1,57	0,259
Linear	3	0,000001	0,000000	1,60	0,257
\mathbf{A}	1	0,000000	0,000000	0,68	0,430
В	1	0,000000	0,000000	1,78	0,215
C	1	0,000001	0,000001	2,33	0,161
Interações de 2 fatores	3	0,000001	0,000000	1,55	0,268
A*B	1	0,000001	0,000001	3,30	0,103
A*C	1	0,000000	0,000000	0,63	0,447
B*C	1	0,000000	0,000000	0,73	0,416
Erro	9	0,000002	0,000000		
Falta de ajuste	1	0,000000	0,000000	0,11	0,748
Erro puro	8	0,000002	0,000000		
Total	15	0,000004			

Termo	Efeito	Coef	EP de Coef	Valor-T	Valor-P	VIF
Constante		0,001008	0,000120	8,39	0,000	
\mathbf{A}	0,000198	0,000099	0,000120	0,83	0,430	1,00
В	0,000320	0,000160	0,000120	1,33	0,215	1,00
C	-0,000367	-0,000183	0,000120	-1,53	0,161	1,00
A*B	0,000436	0,000218	0,000120	1,82	0,103	1,00
A*C	0,000191	0,000095	0,000120	0,80	0,447	1,00
B*C	0,000205	0,000102	0,000120	0,85	0,416	1,00

Tabela 7. Análise dos coeficientes do modelo MAPE RNA

Assim, é importante ressaltar que, para esse cenário em que os valores do tamanho da amostra variam em dois níveis iguais a 40 e 80 dados, seu o desempenho não é afetado. As RNA apresentaram um desempenho estatisticamente igual em todos os casos, independentemente de qual fator tenha sido variado apesar de, normalmente, serem aplicadas quando trabalham com maior quantidade de dados. Então, não foi possível criar um modelo que representasse bem o comportamento das RNA na modelagem desse tipo de séries temporais.

4.4.3 Análise estatística para a resposta Diferença

A outra resposta analisada foi a diferença entre o MAPE MSM e o MAPE RNA denominada de diferença na Tabela 3. Inicialmente, para encontrar o modelo mais adequado a essa resposta, realizou-se a análise de variância, conforme mostrado na Tabela 8. Observa-se que não existe falta de ajuste no modelo e que o grau de liberdade para o erro, assim como no caso do MAPE MSM também é igual a 10. O fatores A e C (σ_{ε} e T), a interação AB ($\sigma_{\varepsilon} \times \rho_{ij}$) e a interação AC ($\sigma_{\varepsilon} \times T$) apresentaram p-value menor que 0,05, sendo, portanto, consideradas significativas essas fontes de variação.

Após a análise de variância, faz-se a análise dos coeficientes codificados, conforme Tabela 9 para que seja possível a elaboração do modelo da resposta Diferença. Como esperado, os valores dos coeficientes são muito próximos àqueles anteriormente mostrados na Tabela 5, pois como nenhum fator foi significativo para as redes neurais, assim o modelo para a diferença tende a ser bem próximo ao modelo elaborado para a resposta MAPE MSM.

Para ilustrar os fatores significativos estão demonstrados na Figura 22 os gráficos de Pareto (a esquerda) e de distribuição de probabilidade (a direita). Pode-se observar que, assim como no caso da resposta MAPE MSM, os termos significativos foram os fatores σ_{ε} e T, a interação $\sigma_{\varepsilon} \times \rho_{ij}$ e a interação $\sigma_{\varepsilon} \times T$.

Fonte	GL	SQ (Aj.)	QM (Aj.)	Valor F	Valor-P
Modelo	5	0,007484	0,001497	24,95	0,000
Linear	3	0,006723	0,002241	37,36	0,000
\mathbf{A}	1	0,006427	0,006427	107,15	0,000
В	1	0,000043	0,000043	0,72	0,417
C	1	0,000253	0,000253	4,21	0,067
Interações de 2 fatores	2	0,000761	0,000381	6,35	0,017
$\mathbf{A}^*\mathbf{B}$	1	0,000375	0,000375	6,26	0,031
A*C	1	0,000386	0,000386	6,43	0,030
Erro	10	0,000600	0,000060		
Falta de ajuste	2	0,000034	0,000017	0,24	0,791
Erro Puro	8	0,000566	0,000071		
Total	15	0,008084			

Tabela 9. Análise dos coeficientes para o modelo da resposta Diferença

Termo	Efeito	Coef	EP de Coef	Valor-T	Valor-P	VIF
Constante	-	0,04743	0,00194	24,49	0,000	
${f A}$	0,04009	0,02004	0,00194	10,35	0,000	1,00
В	0,00328	0,00164	0,00194	0,85	0,417	1,00
\mathbf{C}	0,00795	0,00397	0,00194	2,05	0,067	1,00
A*B	-0,00969	-0,00484	0,00194	-2,50	0,031	1,00
A*C	0,00982	0,00491	0,00194	2,54	0,030	1,00

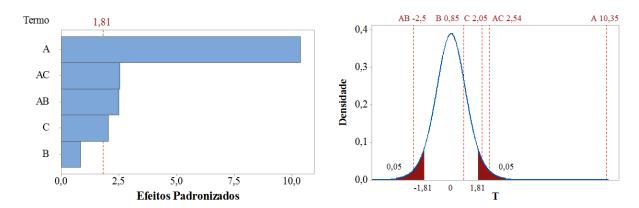


Figura 22. Gráfico de Pareto dos efeitos padronizados para a resposta Diferença e gráfico de distribuição de probabilidade associado

Sendo assim, o modelo obtido para o valor da diferença entre MAPE MSM e MAPE RNA em unidades não codificadas é dado pela Equação 14, sendo que os termos B e C foram mantidos para obter um modelo hierárquico. O modelo apresenta um R² igual a 92,58%, um R² ajustado de 86,79% e um R² de previsão igual a 81,00%. Assim, conforme anteriormente mencionado, o modelo é capaz de explicar o valor da diferença entre os valores de MAPE para os métodos já que todos os indicadores foram superiores a 80%.

$$DIFERENÇA = 0.04743 + 0.02004 A + 0.00164 B + 0.00397 C - 0.00484 A * B + 0.00491 A * C$$
(14)

Avaliando-se os resíduos do modelo obteve-se um *p-value* igual a 0,648, o que significa que os resíduos estão normalmente distribuídos e que o modelo apresentado é capaz de explicar bem os dados. O gráfico de probabilidade para os resíduos pode ser visto na Figura 23.

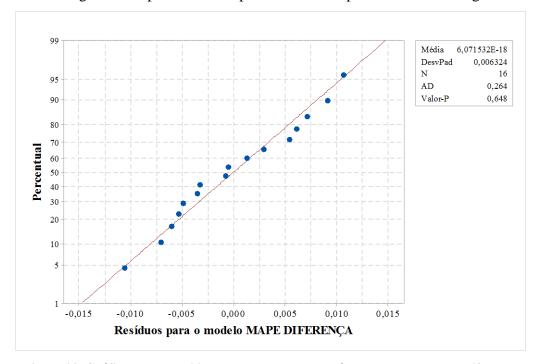


Figura 23. Gráfico de probabilidade normal para os resíduos do modelo para diferença

Por meio do *software* Minitab 18 ® elaborou-se o gráfico de efeitos principais, conforme mostrado na Figura 24, sendo que novamente foram considerados apenas os termos contidos no modelo.

É possível notar que o valor médio da Diferença entre os valores de MAPE para MSM e RNA é maior quando se tem um maior valor de σ_{ε} , o que significa que o desempenho dos dois métodos tende a ser bem diferente. Para baixos valores desse fator serão obtidos baixos valores para a resposta Diferença, o que indica uma maior similaridade entre as performances

entre as RNA e o MSM. Além disso, o valor desta variável tende a ser menor quando menor for o tamanho do conjunto de dados. A variação exercida pela correlação entre as séries foi considerada não significativa.

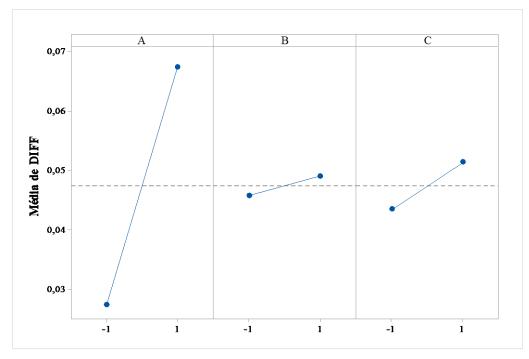


Figura 24. Gráfico de efeitos principais para a resposta Diferença

É válido destacar que nos cenários avaliados, o valor médio da resposta Diferença foi sempre superior a 0, possuindo média em torno de 4,7%, o que significa que nessa região estudada a abordagem por RNA será sempre melhor que pelo método MSM, já que os valores médios para o MAPE dos métodos são de 0,10% e 4,84%, respectivamente. Além disso, observa-se que uma variação nos valores de ρ_{ij} não causa grandes variações na resposta (efeito de apenas 0,00328), o que está de acordo com o fato de esse fator não ter sido considerado significativo anteriormente.

Em seguida realizou-se a análise das interações entre os fatores, conforme mostrado na Figura 25. Pode-se inferir que quando se trabalha com um baixo desvio padrão do erro aleatório o efeito da correlação será maior. Nessas circunstâncias um baixo valor ρ_{ij} indica um pequeno valor médio, mas ainda positivo, para a Diferença, indicando que ambos os métodos possuem performances similares. A interação entre σ_{ε} e T mostra que quanto maior o desvio padrão do erro aleatório, maior será o efeito do tamanho do conjunto de dados. Ressalta-se que o valor médio da diferença é sempre positivo, o que significa dizer o método RNA teve melhor desempenho em todos os cenários estudados.

Assim, é importante ressaltar que como se tem informações suficientes para continuar as análises, não é necessário voltar ao passo C.

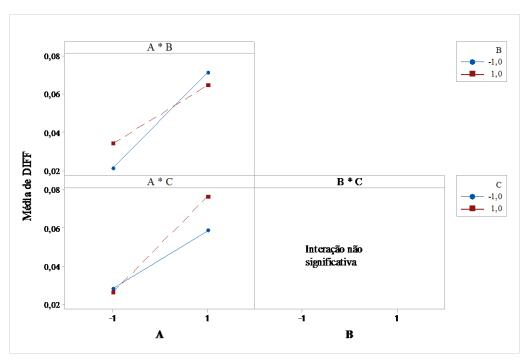


Figura 25. Gráfico de interações para a resposta Diferença

4.4.4 Análises estatísticas complementares

D2: Teste t pareado, desirability e análise de resíduos

Tendo em vista os resultados apresentados, é fácil observar que os valores do MAPE das RNA são, em média, 97,92% menores que os valores do MAPE do MSM. Entretanto, fazse necessário verificar se essa diferença é estatisticamente significativa e por isso realizou-se o chamado teste t pareado. As hipótese de nulidade (H_0) para esse teste indica que as diferenças entre os valores da métrica MAPE para ambos os métodos é estatisticamente igual a zero, enquanto que a hipótese alternativa (H_1) indica que esse valor é estatisticamente diferente de zero, sendo assim um dos métodos possui melhor desempenho que outro.

Para esse teste obteve-se um valor de *p-value* igual a 0,000 e daí pode-se concluir que a hipótese nula deve ser rejeitada. Chega-se, então, ao resultado de que as RNA possuem melhor desempenho que o MSM na modelagem de séries temporais não lineares e não estacionárias.

Buscando entender claramente quais as condições em que os dois métodos aqui estudados apresentam performances similares, resolveu-se um problema de otimização utilizando o algoritmo *desirability* do *software* Minitab 18 ®. Nesse processo, buscou-se obter um cenário no qual a média do valor da resposta Diferença fosse pequena, igual a 0%, ou no máximo 2,5%. Além disso, é desejável um valor de 0% para o valor do MAPE MSM também com um limite superior de 2,5%.

O algoritmo utilizado, requer que sejam atribuídos valores de peso para cada resposta que irão determinar a forma da função *desirability*, o que significa que quanto maior o valor atribuído maior a relevância que o algoritmo dará para o alvo daquela resposta. Um baixo valor de peso atribuído indica que será atribuída menor ênfase ao alvo. Além disso, também é necessário determinar a importância de cada resposta, ou seja, a importância de cada uma das respostas na função *desirability* composta, o que indica qual delas deverá apresentar maior prioridade.

Os valores de peso e importância devem estar compreendidos na faixa de 0,1 a 10. Assim, definiram-se os valores de peso como sendo iguais a 0,1, dessa forma o alvo das respostas deixam de possuir maior ênfase e qualquer valor dentro do intervalo proposto para as respostas analisadas passa a ser aceitável para o algoritmo. Em relação à importância, definiu-se o valor 1 para ambas as respostas, significando que o valores do MAPE MSM e da média da Diferença terão a mesma importância. Os resultados desse problema estão apresentados na Figura 26.

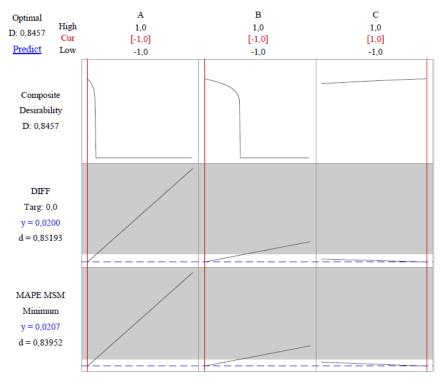


Figura 26. Resultados do algoritmo Desirability

Após analisar os resultados, pode-se concluir que apenas em poucos casos é possível obter uma média para o valor da Diferença que esteja no intervalo de 0% a 2,5%, sendo que as áreas sombreadas indicam as regiões onde o valor do *desirability* é nulo. O algoritmo indica que no máximo 84,57% do que se esperava obter de fato pode ser atingido simultaneamente

considerando as respostas analisadas, para isso é necessário que se trabalhe com dados que possuam baixa variação e correlação, sendo pouco relevante o tamanho do conjunto de dados.

Como forma de avaliar qual dos métodos apresenta os resíduos mais bem comportados, optou-se por realizar uma análise dos resíduos obtidos com a modelagem das séries temporais não lineares. Essas análises foram realizadas por meio do *software* Minitab 18 ® nas quais pôde-se verificar se os resíduos seguem ou não uma distribuição normal, além de verificar a existência de tendência, oscilações, misturas e agrupamentos nesses dados.

As Tabelas 10 e 11 mostram o resumo das informações obtidas com essas análises. A partir delas, é plausível inferir que os resíduos associados aos dois métodos não apresentam nenhum tipo de agrupamento, tendência, mistura ou oscilação, entretanto os resíduos do método RNA não estão distribuídos normalmente.

Assim, pode-se inferir que o resíduos obtidos por meio do MSM apresentam melhor comportamento que os resíduos obtidos por RNA, já que são normalmente distribuídos. Além disso, em 50% dos casos os resíduos das RNA apresentam agrupamentos, o que raramente ocorre para o método MSM.

Por fim, realizou-se a análise de autocorrelação dos resíduos gerados pelos métodos RNA e MSM usando a estatística de qualidade de ajuste de Ljung-Box dada pela Equação 15. Calcularam-se os valores de autocorrelação (r_k) nos resíduos, em que k é igual a T/4, e T o número total de observações na série temporal.

$$Q_{LB} = T(T+2) \sum_{k=1}^{K} \left(\frac{1}{T-k}\right) r_k^2$$
 (15)

Os resultados para o teste podem ser observados na Tabela 12, na qual LBQ representa o valor da estatística de *Ljung-Box*. É importante destacar que essa estatística segue aproximadamente uma distribuição chi-quadrado com k graus de liberdade. Assim, comparase os valores obtidos na tabela com os valores de críticos de $\chi^2_{0,05,k}$ equivalentes a 18,30 e 31,41 para k igual a 10 e 20, respectivamente.

Tabela 10. Análise dos resíduos para o método MSM

Valores de <i>p-value</i> para o método MSM								
Evnavimente	<i>p-value</i> para							
Experimento	Normalidade	Agrupamento	Tendência	Mistura	Oscilação			
1	0.006	0.993	0.818	0.007	0.182			
2	0.032	0.688	0.903	0.312	0.097			
3	0.223	0.073	0.398	0.927	0.602			
4	< 0.005	0.873	0.903	0.127	0.097			
5	0.511	0.154	0.003	0.846	0.997			
6	0.211	0.634	0.964	0.366	0.036			
7	0.158	0.005	0.121	0.995	0.879			
8	0.009	0.214	0.075	0.786	0.925			
9	0.286	0.566	0.552	0.434	0.448			
10	0.030	0.437	0.398	0.563	0.602			
11	0.502	0.008	0.903	0.992	0.097			
12	0.273	0.315	0.077	0.685	0.923			
13	0.021	0.634	0.359	0.366	0.641			
14	0.483	0.894	0.359	0.106	0.641			
15	0.757	0.027	0.264	0.973	0.736			
16	0.043	0.107	0.012	0.893	0.988			

Conclui-se que caso o valor encontrado para a estatística LBQ for menor que o valor crítico, não há nenhuma evidência indicando que as k autocorrelações dos resíduos são diferentes de zero. Uma outra forma de realizar essa análise é por meio do p-value. Quando este é superior a 0,05 não rejeita-se a hipótese nula de que os resíduos são *White noise*.

Conforme pôde-se observar pela Tabela 12, o método MSM gerou resíduos autocorrelacionados em apenas um experimento, enquanto as RNA obtiveram resíduos autocorrelacionados em 3 experimentos. Isso fornece um bom indício de que os métodos são capazes de gerar resíduos que são *White noise* na maior parte dos casos conforme deseja-se. Todo o processo de análise de resíduos, os gráficos de autocorrelação, bem como os gráficos de valores ajustados e reais para ambos os métodos podem ser observados no Apêndice B deste trabalho.

Tabela 11. Análise dos resíduos para o método RNA

Valores de <i>p-value</i> para o método RNA							
Experimento	<i>p-value</i> para	<i>p-value</i> para	<i>p-value</i> para	<i>p-value</i> para	p-value para		
	Normalidade	Agrupamento	Tendência	Mistura	Oscilação		
1	< 0.005	0.037	0.903	0.963	0.097		
2	< 0.005	0.037	0.258	0.963	0.742		
3	0.007	0.210	0.818	0.790	0.182		
4	< 0.005	0.073	0.077	0.927	0.923		
5	< 0.005	0.000	0.674	1.000	0.326		
6	< 0.005	0.715	0.464	0.285	0.536		
7	< 0.005	0.000	0.024	1.000	0.976		
8	< 0.005	0.368	0.121	0.632	0.879		
9	< 0.005	0.018	0.035	0.982	0.965		
10	< 0.005	0.793	0.552	0.207	0.448		
11	0.165	0.000	0.698	1.000	0.302		
12	< 0.005	0.688	0.818	0.312	0.182		
13	< 0.005	0.214	0.839	0.786	0.161		
14	< 0.005	0.286	0.184	0.714	0.816		
15	< 0.005	0.000	0.184	1.000	0.816		
16	< 0.005	0.027	0.359	0.973	0.641		

Tabela 12. Análise de autocorrelação dos resíduos gerados pelos métodos MSM e RNA

Experimento	MSM LBQ	MSM p-value	RNA LBQ	RNA p-value
1	8,411	0,589	0,988	1,000
2	7,251	0,702	6,901	0,734
3	9,677	0,469	17,136	0,071
4	10,582	0,391	4,221	0,937
5	26,596	0,147	30,984	0,055
6	15,521	0,746	4,719	1,000
7	38,004	0,009	54,819	0,000
8	19,172	0,511	15,769	0,731
9	4,181	0,939	0,465	1,000
10	10,156	0,427	1,466	0,999
11	11,415	0,326	103,521	0,000
12	6,205	0,798	6,455	0,776
13	17,501	0,620	4,435	1,000
14	12,076	0,913	20,014	0,457
15	28,131	0,106	17,955	0,590
16	16,704	0,672	52,785	0,000

5. ESTUDO DE CASO

Alguns trabalhos exploram a relação entre variáveis tais quais exportação, importação, taxa de câmbio e PIB (TANG; LAI; OZTURK, 2015; JORDAAN; EITA, 2007). Como essas séries, na maior parte das vezes, apresentam comportamentos não lineares e não estacionários, buscou-se por meio da modelagem destas validar os resultados da presente pesquisa.

O conjunto de dados trabalhado foi obtido no site do Instituto de Pesquisa Econômica Aplicada (IPEA). Assim, modelou-se a série de preços de exportação brasileira durante os anos de 2008 a 2018 em função das séries de preço importação, taxa de câmbio e PIB brasileiros no mesmo período de tempo, além da própria série de preço de exportação deslocada de uma unidade no tempo, conforme feito anteriormente com o banco de dados sintético.

É importante destacar que as séries de preço de exportação e de importação tiveram como base o preço médio do ano de 2006. Assim, adotou-se o valor 100 para o preço médio desse ano e a partir dele definiram-se os valores para os demais anos.

Todas as configurações e softwares utilizados nesse estudo de caso foram os mesmos utilizados para modelar as séries do banco de dados sintético. A série de exportação, variável resposta nesse estudo, pode ser observada na Figura 27. Pode-se perceber que, no período analisado, o Brasil incialmente possuía baixos valores de exportação, passando por um período de alta voltando a ter uma redução a partir do ponto T=80, aproximadamente.

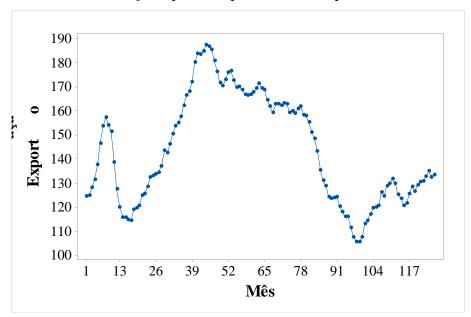


Figura 27. Valores de preço de exportação no Brasil de 2008 a 2018

No período analisado a taxa de cambio brasileira apresentou uma tendência de crescimento ao longo dos anos sofrendo algumas quedas em alguns momentos. Essa série é apresentada na Figura 28.

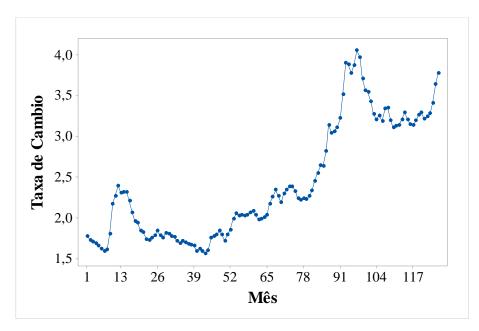


Figura 28. Valores da Taxa de Câmbio brasileira no período de 2008 a 2018

A série do PIB é a única que apresenta tendência de crescimento ao longo dos anos, conforme pode ser observado na Figura 29. Nesse caso ela poderia ser facilmente modelada por regressão linear, entretanto caso fosse abrangido um maior horizonte temporal a série certamente apresentaria um comportamento não linear e com diferentes regimes, já que existem ao longo dos anos, em variáveis macroeconômicas, momentos de crescimento acelerado e momentos em que o crescimento é menos acelerado, bem como crises econômicas e financeiras que caracterizam comportamentos não lineares.

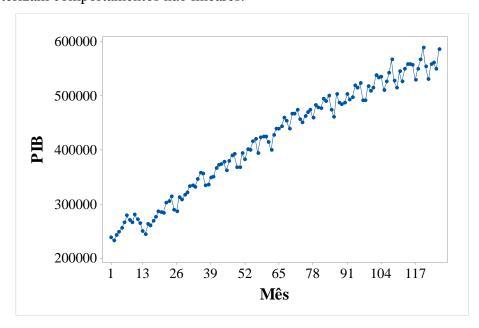


Figura 29. Valores do PIB brasileiro de 2008 a 2018

Similarmente ao que ocorre com os valores de exportação, pode-se observar que os valores de importação brasileira apresentou inicialmente valores mais baixos e posteriormente passa por um momento de grande importação, seguido novamente de uma queda a partir do ponto T=80. O valor da correlação entre essas séries foi de 0,94 e p-value igual a 0,000 de acordo com o método de correlação de Pearson.

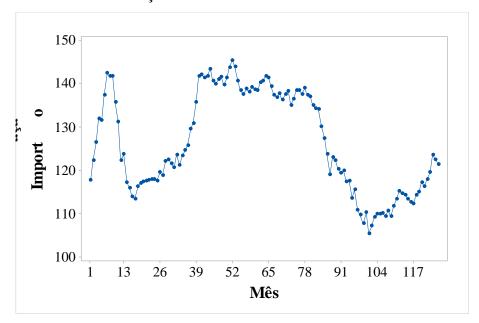


Figura 30. Valores de preço de importação no Brasil de 2008 a 2018

Os valores de MAPE quando se modelou a série exportação utilizando MSM foi de 1,27% contra 0,12% quando utilizado RNA. Assim, por meio desse estudo de caso observa-se que as redes neurais artificiais são mais adequadas para modelagem de séries temporais não lineares que o método *Markov-switching model*. Assim, pode ser entendido como um primeiro passo para a validação da metodologia de apoio à decisão proposta nesse trabalho. Convém ainda destacar que as séries trabalhadas nesse estudo de caso são favoráveis ao método MSM e mesmo assim a abordagem por RNA apresentou melhor desempenho, o que reforça ainda mais robustez das redes neurais.

Nesse momento convém realizar uma análise dos resíduos de cada um dos métodos, assim como foi realizado para os experimentos com banco de dados sintéticos. Utilizou-se o *software* Minitab 18 ® para verificar a normalidade dos resíduos e a presença de qualquer tendência, mistura, agrupamentos ou oscilações. Assim, por meio do teste de normalidade de Anderson-Darling obteve-se o gráfico mostrado na Figura 31 para os resíduos do método MSM. Pode-se observar que os resíduos são normalmente distribuídos, conforme é desejado, já que o *p-value* é maior que 0,05, não rejeitando a hipótese nula de normalidade dos resíduos.

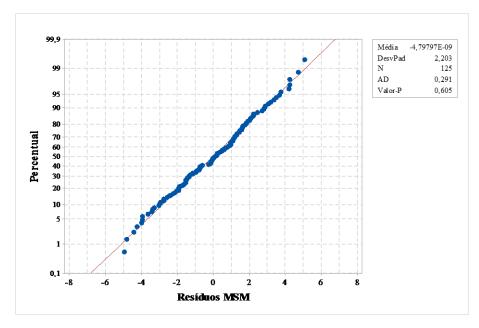


Figura 31. Gráfico de probabilidade de Resíduos MSM

Complementando a análise para os resíduos do método MSM, realizou-se o gráfico de ensaios conforme pode ser observado na Figura 32. Os valores de *p-value* maiores que 0,05 indicam a não existência de mistura e oscilação. Entretanto, os valores de *p-value* menores que 0,05 indicam a presença tendências e agrupamento, o que não é desejado.

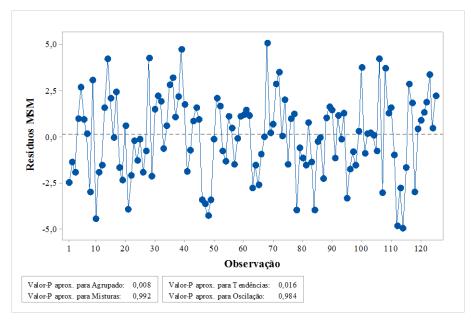


Figura 32. Gráfico de ensaios de Resíduos MSM

A mesma análise foi realizada para os resíduos obtidos com o método RNA. Constatase por meio da Figura 33 que os resíduos não são normalmente distribuídos, mostrando então uma desvantagem das redes neurais artificiais em relação ao *Markov-switching model*.

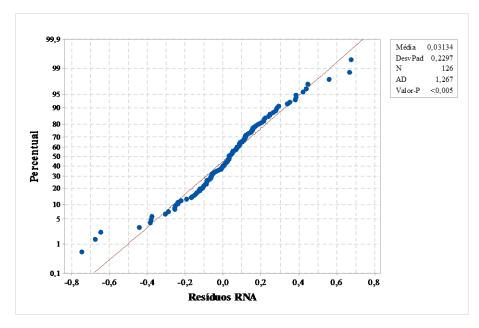


Figura 33. Gráfico de probabilidades de Resíduos RNA

A Figura 34 mostra o gráfico de ensaios realizado para os resíduos do método RNA. As conclusões obtidas por meio dessa figura são as mesmas que se obteve para o método MSM. Os altos valores de *p-value* indicam a ausência de qualquer mistura ou oscilação, enquanto que os baixos valores indicam a presença de agrupamentos e tendências.

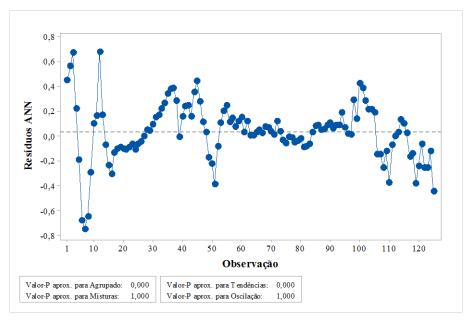


Figura 34. Gráfico de ensaios de Resíduos RNA

Além disso, também avaliou-se a autocorrelação entre os resíduos gerados pelos métodos durante a resolução do problema proposto para o estudo de caso. Para o método MSM obteve-se um valor acumulado para a estatística de *Ljung-Box* igual a 35,784 e para as RNA o valor de 158,701. Uma vez que o valor de LBQ para as RNA foi maior que o valor crítico de

 $\chi^2_{0,05,31}$ = 44,985, rejeita-se a hipótese nula de que os resíduos são *White noise* e conclui-se que são autocorrelacionados com um *p-value* igual a 0,000. Por outro lado, para o método MSM não se rejeita a hipótese nula já que seu valor para a estatística LBQ é menor que o valor crítico, com *p-value* associado igual a 0,254.

Para ambos os casos mencionados acima, tem-se os gráficos da função de autocorrelação mostrados nas Figuras 35 e 36.

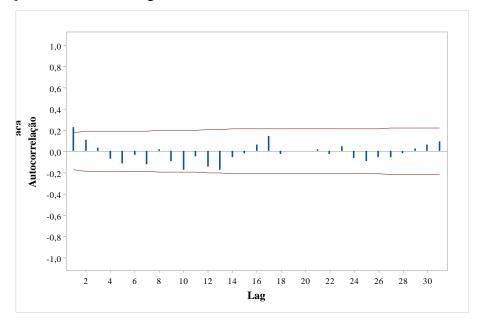


Figura 35. Gráfico de autocorrelação para os resíduos do método MSM

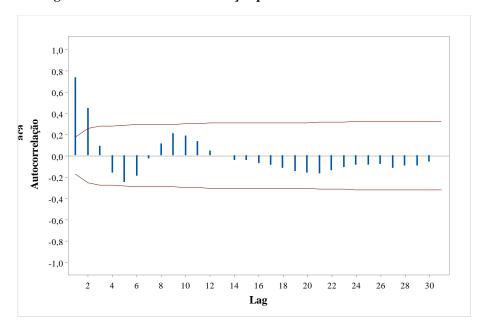


Figura 36. Gráfico de autocorrelação para os resíduos do método RNA

6. CONCLUSÕES

O presente trabalho desenvolveu um estudo comparativo entre os métodos *Markov-switching model* e redes neurais artificiais para a modelagem de séries temporais não lineares. O estudo foi realizado por meio da elaboração de um banco de dados sintéticos criado a partir de um arranjo fatorial completo com 2 réplicas. Consideraram-se 3 fatores que poderiam de alguma forma influenciar no desempenho dos métodos: desvio padrão do erro aleatório das séries temporais (σ_{ε}), tamanho do conjunto de dados (N) e correlação entre as séries (ρ_{ij}).

Modelaram-se os dados originados a partir dos 16 experimentos utilizando ambos os métodos estudados, sendo que o desempenho destes foi mensurado por meio da métrica de erro percentual absoluto médio (MAPE), que indica o erro percentual de cada uma das modelagens. Em todos os casos, o método RNA teve melhor performance que o MSM, apresentando menor valor de MAPE associado. Avaliaram-se as diferenças entre os desempenhos dos métodos por meio do teste t pareado que apresentou *p-value* igual a 0,000 com um nível de significância igual a 95%, confirmando que o MAPE RNA foi significativamente menor que o MAPE MSM.

Pôde-se observar que os fatores analisados, bem como suas interações, de fato influenciam o desempenho do MSM. Por outro lado, o mesmo não acontece para RNA, o que, juntamente com o fato de esse método possuir menores valores de MAPE (em média 97,92% menores que os valores do MAPE MSM), reforça sua robustez, se mostrando aplicável nas diversas situações exploradas.

Assim, pode-se inferir do estudo realizado que as RNA são capazes de compreender maiores variações aleatórias ao longo do tempo, já que seu desempenho não foi influenciado pelo valor do σ_{ε} . Além disso o método também apresenta ótimo resultados independentemente de as séries serem ou não correlacionadas. Em contrapartida, o método MSM é prejudicado quando as séries apresentam alto desvio, mas é pouco influenciado pelo tamanho do conjunto de dados e é indiferente aos valores de correlação entre as séries.

Apesar de um melhor desempenho em relação à métrica MAPE, os resíduos obtidos tanto nos experimentos com dados sintéticos, assim como no estudo de caso, para o método RNA não seguiam distribuição normal.

6.1 Contribuições do trabalho

O presente trabalho apresenta como maior contribuição a conclusão obtida por meio do estudo comparativo de que as redes neurais artificiais apresentam melhor desempenho na modelagem de séries temporais não lineares do que o *Markov-switching model* em termos da

métrica utilizada (MAPE). Entretanto, convém destacar que o método RNA, em relação ao tipo de rede neural utilizada (*Three Layer* Perceptron) não apresentou resíduos que seguem uma distribuição normal, o que é uma desvantagem em relação ao método MSM.

O trabalho também contribui para a comunidade científica em relação à metodologia seguida para realizar a comparação. Ela foi útil em três aspectos principais. Inicialmente em relação à generalização, criando diferentes contextos onde os métodos puderam ser aplicados, já que essa é a única forma de conseguir obter conclusões confiáveis em um estudo comparativo. A metodologia também apoia a execução de uma análise mais profunda visando entender como os fatores e interações significantes influenciam a precisão dos modelos, o que auxilia o pesquisador durante a escolha de qual método aplicar em determinadas circunstâncias a partir das características do conjunto de dados. Finalmente, a metodologia requer o menor número possível de experimentos, já que é baseado na técnica de DOE.

Todas as conclusões que o pesquisador pode obter são estatisticamente confiáveis, a depender de como se procederam a escolha e análises do arranjo experimental. Assim, evitam-se casos nos quais um método é considerado melhor, quando na verdade, a diferença entre os desempenhos não é estatisticamente significativa ao se realizar um teste de hipóteses com baixo nível de significância.

6.2 Sugestão para trabalhos futuros

Visando à modelagem de séries temporais não lineares pelo método RNA, neste trabalho, utilizou-se apenas o tipo de rede neural mais básico conhecida como *Three Layer Perceptron*, objetivando verificar se um dos tipos mais simples de redes neurais apresentaria um desempenho melhor que o método MSM. Assim, um possível estudo a ser realizado é o uso de outras redes como a *Radial Basis Function* (RBF), por exemplo, comparada com métodos que utilizam em sua estrutura os mecanismos de troca do MSM, como o modelo de heteroscedasticidade condicional auto-regressiva (ARCH) e o modelo generalizado de heteroscedasticidade condicional auto-regressiva (GARCH).

Nessas situações seria conveniente explorar outros fatores e outros níveis, diferentes dos que foram considerados neste trabalho. Também seria importante verificar a capacidade dos métodos em generalizar os resultados obtidos para além do conjunto de dados fornecido, ou seja, avaliar capacidade de previsão.

Outra alternativa seria a elaboração de um sistema especializado (*assistant*) para gerar todos os dados automaticamente baseado nas entradas fornecidas pelo pesquisador. Então, o usuário deveria informar o tipo de problema a ser resolvido, quais métodos serão investigados,

quais fatores poderiam influencia-los, e os valores para os níveis desses fatores. O *assistant* sugeriria o *design* experimental mais adequado, além de métodos, fatores e níveis já utilizados por outros estudos anteriormente desenvolvidos. Após a execução do estudo utilizando um *software* especializado no método e no tipo de problema selecionado, o *assistant* demandaria ao pesquisador a entrada das principais conclusões obtidas em seu estudo de forma a criar uma base de dados para apoiar estudos futuros.

REFERÊNCIAS BIBLIOGRÁFICAS

AILLIOT, P.; MONBET, V. Markov-switching autoregressive models for wind time series. **Environmental Modelling and Software**, v. 30, p. 92-101, 2012.

AIZENBERG, I.; SHEREMETOV, L.; VILLA-VARGAS, L.; MARTINEZ-MUÑOZ, J. Multilayer Neural Network with Multi-Valued Neurons in time series forecasting of oil production. **Neurocomputing**, v. 175, p. 980-989, 2016.

AMIRI, E. Forecasting daily river flows using nonlinear time series models. **Journal of Hydrology**, v. 527, p. 1054-1072, 2015.

BALCILAR, M.; EYDEN, R. V.; UWILINGIYE, J.; GUPTA, R. The Impact of Oil Price on South African GDP Growth: A Bayesian Markov Switching-VAR Analysis. **African Development Review**, v. 29, n. 2, p. 319-336, 2017.

BALESTRASSI, P. P.; POPOVA, E.; PAIVA, A. P.; LIMA, J. W. M. Design of experiments on neural network's training for nonlinear time series forecasting. **Neurocomputing**, v. 72, p. 1160-1178, 2009.

BERTRAND, J. W. M.; FRANSOO, J. C. Modelling and Simulation: Operational management research methodologies using quantitative modeling. **International Journal of Operations and Production Management**, v. 22, n. 2, p. 241-264, 2002.

BONSALL, M. B.; WALLACE-HADRILL, S. M. A.; GEDDES, J. R.; GOODWIN, G. M.; HOLMES, E. A. Nonlinear time-series approaches in characterizing mood stability and mood instability in bipolar disorder. **Proceedings of the Royal Society B**, v. 279, p. 916-924, 2012.

CAVICCHIOLI, M. Determining the Number of Regimes in Markov Switching Var and Vma Models. **Journal of Time Series Analysis**, v. 35, p. 173-186, 2014.

CEN, Z.; WANG, J. Crude oil price prediction model with long short term memory deep learning based on prior knowledge data transfer. **Energy**, v. 169, p. 160-171, 2019.

CHANG, J.; TSENG, C. Analysis of Correlation between Secondary PM2.5 and Factory Pollution Sources by Using ANN and the Correlation Coefficient. **IEEE Access**, v. 5, p. 22812-22822, 2017.

CHAUDHURI, K.; KUMAR, A. A Markov-Switching Model for Indian Stock Price and Volume. **Journal of Emerging Market Finance**, v. 14, n. 3, p. 239-257, 2015.

CHENG, B.; TITTERINGTON, D. M. Neural Networks: a review from a statistical perspective. **Statistical Science**, v. 9, n. 1, p. 2-54, 1994.

CIAO, J.; LI, Z.; LI, J. Financial time series forecasting model based on CEEMDAN and LSTM. **Physica A**, v. 519, p. 127-139, 2019.

COHEN, J. **Statistical Power Analysis for the Behavioral Sciences**. 2. ed. New York: Lawrence Erlbaum Associates, 1988.

COLEMAN, D. E.; MONTGOMERY, A. A systematic approach to planning for a designed industrial experiment, **Technometrics**, v. 35, p. 1-27, 1993.

COSTA, D. G. B.; REIS, B. M. F.; ZOU, Y.; QUILES, M. G.; MACAU, E. N. Recurrence Density Enhanced Complex Networks for Nonlinear Time Series Analysis. **International Journal of Bifurcation and Chaos**, v. 28, n. 1, 1850008-1-1850008-12, 2018.

DAI, X.; BIKDASH, M. Trend Analysis of Fragmented Time Series for mHealth Apps: Hypothesis Testing Based Adaptive Spline Filtering Method with Importance Weighting. **IEEE Access**, v. 5, p. 27767-27776, 2017.

DASCALESCU, L.; MEDLES, K.; DAS, S.; YOUNES, M.; CALIAP, L.; MIHALCIOIU, A. Using Design of Experiments and Virtual Instrumentation to Evaluate the Tribocharging of Pulverulent Materials in Compressed-Air Devices. **IEEE Transactions on Industry Applications**, v. 44, n. 1, p. 3-8, 2008.

DONG, X.; LIAN, Y.; LIU, Y. Small and multi-peak nonlinear time series forecasting using a hybrid back propagation neural network. **Information Sciences**, v. 424, p. 39-54, 2018.

FORNACIARI, M.; GRILLENZONI, C. Evaluation of on-line trading systems: Markov-switching vs time-varying parameter models. **Decision Support Systems**, v. 93, p. 51-61, 2017.

FRAGKOU, A. D.; KARAKASIDIS, T. E.; NATHANAIL, E. Detection of traffic incidents using nonlinear time series analysis. **Chaos**, v. 28, p. 063108-1-063108-13, 2018.

GAO, Z.; LI, S.; DANG, W.; YANG, Y.; DO, Y.; GREBOGI, C. Wavelet Multiresolution Complex Network for Analyzing Multivariate Nonlinear Time Series. **International Journal of Bifurcation and Chaos**, v. 27, n. 8, p. 1750123-1-175012-11, 2017.

GELATI, E.; MADSEN, H.; ROBSJERG, D. Markov-switching model for nonstationary runoff conditioned on El Niño information. **Water Resources Research**, v. 46, p. 1-13, 2010.

GUAN, Y.; LIU, P; JIN, B.; GUPTA, V.; LI, L. K. B. Nonlinear time-series analysis of thermoacoustic oscillations in a solid rocket motor. **Experimental Thermal and Fluid Science**, v. 98, p. 217-226, 2018.

HAMILTON, J. D. A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle. **Econometrica Journal of Economic Society**, v. 57, n. 2, p. 357-384, 1989.

HAMILTON, J. D. Analysis of Time Series Subject to Changes in Regime. **Journal of Econometrics**, v. 45, p. 39-70, 1990.

HAMILTON, J. D. Rational-Expectations Econometric Analysis of Changes In Regime an Investigation of the Term Structure of Interest Rates. **Journal of Economic Dynamics and Control**, v. 12, p. 385-423, 1988.

HAMILTON, J. D. **Time Series Analysis**. 1. ed. New Jersey: Princeton University Press, 1994.

HAYKIN, S. **Neural Networks and Learning Machines**. 3. ed. New Jersey: Prentice Hall, 2009.

HSIEH, Y.; CHEN, W. Measuring the value variation of a service system: a Markov-switching model estimation. **Journal of Business and Industrial Marketing**, v. 32, n. 8, p. 1159-1171, 2017.

JORDAAN, A. C.; EITA, J. H. Export and Economic Growth in Namibia: A Granger Causality Analysis. **South African Journal of Economics**, v. 75, p. 540-547, 2007.

KHANLARI, G. R.; HEIDARI, M.; MOMENI, A. A.; ABDILOR, Y. Prediction of shear strength parameters of soils using artificial neural networks and multivariate regression methods. **Engineering Geology**, v. 131-132, p. 11-18, 2012.

KOCADAGLI, O.; ASIKGIL, B. Nonlinear time series forecasting with Bayesian neural networks. **Expert Systems with Applications**, v. 41, p. 6596-6610, 2014.

KUO, H.; FARICHA, A. Artificial Neural Network for Diffraction Based Overlay Measurement. **IEEE Access**, v. 4, p. 7479-7486, 2016.

LANER, M.; SVOBODA, P.; RUPP, M. Parsimonious Network Traffic Modeling By Transformed ARMA Models. **IEEE Access**, v. 2, p. 40-55, 2014.

LEE, J.; CHANG, J.; KANG, D.; KIM, S.; HONG, J. Tooth Shape Optimization for Cogging Torque Reduction of Transverse Flux Rotary Motor Using Design of Experiment and Response Surface Methodology. **IEEE Transactions on Magnetics**, v. 43, n. 4, p. 1817-1820, 2007.

LORSCHEID, I.; HEINE, B.; MEYER, M. Opening the 'black box' of simulations: increased transparency and effective communication through the systematic design of experiments. **Computational and Mathematical Organization Theory**, v. 18, p. 22-62, 2012.

LU, H.; ZENG, D.; CHEN, H. Prospective Infectious Disease Outbreak Detection Using Markov Switching Models. **IEEE Transactions on Knowledge and Data Engineering**, v. 22, n. 4, p. 565-577, 2010.

MAAZIZ, M.; KHARFOUCHI, S. Parameter estimation of Markov switching bilinear model using the (EM) algorithm. **Journal of Statistical Planning and Inference**, v. 192, p. 35-44, 2018.

MCKEIGHEN, R. E. Optimization of Broadband Transducer Designs by Use of Statistical Design of Experiments. **IEEE Transactions on Ultrasonics, Ferroeletrics and Frequency Control**, v. 43, n. 1, p. 63-70, 1996.

MEKANIK, F.; IMTEAZ, M. A.; GATO-TRINIDAD, S. ELMAHDI, A. Multiple regression and Artificial Neural Network for long-term rainfall forecasting using large scale climate modes. **Journal of Hydrology**, v. 503, p. 11-21, 2013.

MIKAEIL, A. M.; GUO, B.; BAI, X.; WANG, Z. Hidden Markov and Markov Switching Model for Primary User Channel State Prediction in Cognitive Radio. **2n** International Conference on Systems and Informatics (ICSAI), United States, 2014.

MO, F.; SHEN, C.; ZHOU, J.; KHONSARI, M. M. Statistical Analysis of the Influence of Imperfect Texture Shape and Dimensional Uncertainty on Surface Texture Performance. **IEEE Access**, v. 5, p. 27023-27035, 2017.

MONTGOMERY, D. C. **Design and Analysis of Experiments**. 8. ed. Danvers: John Wiley and Sons, Inc., 2013.

MONTGOMERY, D. C.; JENNINGS, C. L.; KULAHCI, M. Introduction to Time Series Analysis and Forecasting. 2. ed. New Jersey: Wiley-Interscience, 2008.

OLSON, D. L.; DELEN, D.; MENG, Y. Comparative analysis of data mining methods for bankruptcy prediction. **Decision Support Systems**, v. 52, p. 464-473, 2012.

OMEKANDA, A. M. Robust Torque and Torque-per-Inertia Optimization of a Switched Reluctance Motor Using the Taguchi Methods. **IEEE Transactions on Industry Applications**, v. 42, n. 2, p. 473-478, 2006.

PAIVA, A. P.; PAIVA, E. J.; BALESTRASSI, P. P.; COSTA, S. C. A multivariate mean square error optimization of AISI 52100 hardened steel turning, **International Journal of Advanced Manufacturing Technology**, v. 43, p. 631-643, 2009.

PARK, J.; HONG, T. Analysis of South Korea's economic growth, carbon dioxide emission, and energy consumption using the Markov switching model. **Renewable and Sustainable Energy Reviews**, v. 18, p. 543-551, 2013.

PERRIN, E.; GUIFFAUT, C.; REINEIX, A.; TRISTANT, F. Using a Design-of-Experiment Technique to Consider the Wire Harness Load Impedances in the FDTD Model of an Aircraft Struck by Lightning. **IEEE Transactions on Electromagnetic Compatibility**, v. 55, n. 4, p. 747-753, 2013.

QI, H.; OSTERMAN, M.; PECHT, M. Design of Experiments for Board-Level Solder Joint Reliability of PBGA Package under Various Manufacturing and Multiple Environmental Loading Conditions. **IEEE Transactions on Electronics Packaging Manufacturing**, v. 32, n. 1, p. 32-40, 2009.

ROSENBLATT, F. The perceptron: a probabilistic model for information storage and organization in the brain. **Psychological Review**, v. 65, n. 6, 1958.

SAGHEER, A.; KOTB, M. Time series forecasting of petroleum production using deep LSTM recurrent networks. **Neurocomputing**, v. 323, p. 203-2013, 2019.

STAICULESCU, D.; BUSHYAGER, N.; OBATOYINBO, A.; MARTIN, L. J.; TENTZERIS, M. M. 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, 2005.

SUN, K.; HUANG, S.; WONG, D. S.; JANG, S. Design and Application of a Variable Selection Method for Multilayer Perceptron Neural Network With LASSO. **IEEE Transactions on Neural Networks and Learning Systems**, v. 28, n. 6, p. 1386-1396, 2017.

TANG, C. F.; LAI, Y. W.; OZTURK, I. How stable is the export-led growth hypothesis? Evidence from Asia's Four Little Dragons. **Economic Modelling**, v. 44, p. 229-235, 2015.

TANG, X. Research on Markov-Switching Model. in REVISTA, v. 3, p.15-64, 2011.

TEALAB, A.; HEFNY, H.; BADR, A. Forecasting of nonlinear time series using ANN. **Future Computing and Informatics Journal**, v. 2, p. 39-47, 2017.

TSAY, R. Analysis of Financial Time Series. 2. ed. Wiley-Intersience, 2005.

WANG, H. T.; LIU, Z. J.; CHEN, S. X.; YANG, J. P. Application of Taguchi Method to Robust Design of BLDC Motor Performance. **IEEE Transactions on Magnetics**, v. 35, n. 5, p. 3700-3702, 1999.

WU, Y.; WU, Q.; ZHU, J. Improved EEMD-based crude oil price forecasting using LSTM networks. **Physica A**, v. 516, p. 114-124, 2019.

ZHAI, X.; ALI, A. A. S.; AMIRA, A.; BENSAALI, F. MLP Neural Network Based Gas Classification System on Zynq SoC. **IEEE Access**, v. 4, p. 8138-8146, 2016.

ZHONG, J.; SHI, J.; WU, J. C. F. Design of DOE-Based Automatic Process Controller with Consideration of Model and Observation Uncertainties. **IEEE Transactions on Automation and Engineering**, v. 7, n. 2, p. 266-273, 2010.

APÊNDICE A – GERAÇÃO DOS DADOS SINTÉTICOS

A partir das funções previamente estabelecidas geraram-se os dados determinísticos. Visando obter dados estocásticos foram gerados erros aleatórios no Minitab. De acordo com o DOE foram especificados os valores de desvio dos erros aleatórios a serem acrescidos às séries temporais 1 e 2 respectivamente representados por $\sigma_{\varepsilon 1}$ e $\sigma_{\varepsilon 2}$.

Definiu-se um vetor de médias como sendo um vetor nulo. A matriz de variância e covariância foi especificada tendo na diagonal principal os valores dos desvios do erro aleatório para cada série ao quadrado e demais elementos iguais ao valor da correlação desejada entre as séries multiplicada pelos desvios do erros aleatórios que será adicionado a essas séries. Assim, por meio do software Minitab foi possível gerar os valores dos erros aleatórios conforme mostrado na Figura A. 1.**Erro! Fonte de referência não encontrada.**

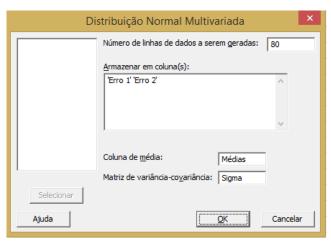


Figura A. 1. Tela do Minitab para geração de dados aleatórios seguindo uma distribuição normal multivariada

As Tabelas de A.1 até A.16 mostram os dados originais gerados pelas funções, os erros aleatórios gerados e o valor final da série sintética representado por $f_i(t) + Erro_i \operatorname{com} i = 1, 2$.

Tabela A. 1. Dados do experimento 1, sendo $\theta_1=0$, 84373; $\theta_2=0$, 10000; $\theta_3=0$, 19812 e σ_{ε_i} =0,2× σ_{y_i}

$f_1(t)$	Erro 1	$f_1(t) + Erro 1$	$f_2(t)$	Erro 2	$f_2(t) + Erro 2$
0,00000	0,00186	0,00186	0,12191	6,32973	6,45164
0,44280	0,01685	0,45965	0,14862	5,55677	5,70539
0,60423	0,00191	0,60614	0,18119	12,94700	13,12819
0,68953	-0,00950	0,68002	0,22088	19,47755	19,69844
0,74281	-0,04621	0,69660	0,26928	2,91805	3,18733
0,77948	0,04267	0,82214	0,32828	-9,99651	-9,66823
0,80637	-0,02758	0,77880	0,40021	-11,44606	-11,04585
0,82700	0,02764	0,85464	0,48789	1,34591	1,83380
0,84337	-0,00219	0,84118	0,59479	10,28630	10,88109
0,85669	-0,00049	0,85620	0,72512	-12,06998	-11,34486
0,86777	-0,03956	0,82821	0,88399	-3,86385	-2,97986
0,87713	0,00500	0,88213	1,07768	3,02497	4,10265
0,88515	-0,00743	0,87772	1,31380	-5,34483	-4,03103
0,89211	-0,04372	0,84840	1,60166	-1,19374	0,40792
0,89821	-0,00697	0,89125	1,95259	3,54421	5,49680
0,90361	-0,04611	0,85750	2,38042	-19,85054	-17,47012
0,90841	-0,00995	0,89847	2,90197	8,11525	11,01722
0,91273	0,02255	0,93528	3,53781	-20,26244	-16,72463
0,91662	-0,06205	0,85456	4,31296	-26,15538	-21,84243
0,92015	-0,01536	0,90479	5,25794	-12,87619	-7,61825
0,92337	0,02788	0,95125	6,40998	-8,69082	-2,28084
0,92632	-0,06935	0,85697	7,81443	-8,52959	-0,71516
0,92903	-0,04467	0,88436	9,52661	-0,33712	9,18949
0,93153	-0,01761	0,91392	11,61393	-11,50374	0,11019
0,93385	0,05332	0,98718	14,15859	29,16700	43,32559
0,93601	0,00838	0,94439	17,26079	19,31841	36,57920
0,93801	-0,01794	0,92007	21,04270	-12,67061	8,37209
0,93988	0,04787	0,98775	25,65324	-4,52750	21,12574
0,94164	0,00844	0,95008	31,27397	9,20017	40,47415
0,94328	-0,04400	0,89928	38,12623	-11,88247	26,24376
0,94483	-0,02350	0,92133	46,47984	-36,11254	10,36730
0,94629	0,04407	0,99036	56,66377	22,53201	79,19578
0,94767	-0,03357	0,91409	69,07903	2,01941	71,09844
0,94897	0,05869	1,00766	84,21453	27,33464	111,54918
0,95020	0,01066	0,96086	102,66628	-9,50788	93,15841
0,95137	0,00948	0,96085	125,16089	3,84339	129,00428
0,95248	0,00376	0,95625	152,58415	-0,39536	152,18879
0,95354	0,01671	0,97025	186,01596	-28,28259	157,73337
0,95455	-0,00014	0,95441	226,77283	-10,56479	216,20804
0,95551	0,00447	0,95998	276,45969	37,59747	314,05716

Tabela A. 2. Dados do experimento 2, sendo $\theta_1=0$, 87956; $\theta_2=0$, 10000; $\theta_3=0$, 20309 e $\sigma_{\varepsilon_i}=0$, $8\times\sigma_{y_i}$.

$f_1(t)$	Erro 1	$f_1(t) + Erro 1$	$f_2(t)$	Erro 2	$f_2(t) + Erro 2$
0,00000	0,01140	0,01140	0,12252	-51,06411	-50,94159
0,45647	0,16140	0,61786	0,15011	65,40845	65,55856
0,61951	-0,03251	0,58700	0,18391	-46,17670	-45,99279
0,70457	0,05680	0,76137	0,22532	57,75553	57,98086
0,75722	-0,07575	0,68146	0,27606	15,54294	15,81901
0,79319	-0,05519	0,73800	0,33823	-60,38744	-60,04921
0,81941	0,02325	0,84267	0,41439	-77,91882	-77,50443
0,83942	0,15877	0,99819	0,50771	121,31936	121,82707
0,85523	-0,15407	0,70116	0,62204	68,34665	68,96868
0,86804	-0,17707	0,69097	0,76211	11,64217	12,40427
0,87865	-0,08736	0,79129	0,93372	10,05242	10,98614
0,88759	-0,13446	0,75313	1,14398	3,03897	4,18295
0,89523	0,16046	1,05569	1,40159	-13,93606	-12,53447
0,90184	-0,00077	0,90107	1,71721	-15,44806	-13,73085
0,90762	-0,11511	0,79252	2,10390	-73,94836	-71,84446
0,91272	0,11071	1,02343	2,57767	-53,67351	-51,09585
0,91725	0,09835	1,01560	3,15812	-93,56288	-90,40476
0,92131	-0,21453	0,70678	3,86928	-131,09581	-127,22654
0,92497	0,20351	1,12848	4,74058	37,13793	41,87851
0,92828	0,23388	1,16215	5,80809	-97,63089	-91,82280
0,93129	0,32526	1,25655	7,11598	83,95042	91,06641
0,93404	-0,00002	0,93402	8,71839	-4,32802	4,39037
0,93657	0,14592	1,08249	10,68165	16,64132	27,32296
0,93890	-0,02674	0,91216	13,08699	-163,98003	-150,89304
0,94106	-0,08350	0,85755	16,03399	38,92661	54,96059
0,94306	-0,05376	0,88930	19,64460	8,24441	27,88901
0,94492	0,05494	0,99986	24,06827	37,03973	61,10799
0,94665	-0,23603	0,71062	29,48808	-27,45216	2,03592
0,94827	0,08098	1,02925	36,12835	99,35600	135,48435
0,94979	0,00919	0,95898	44,26391	-48,09161	-3,82770
0,95122	-0,01292	0,93829	54,23148	14,42437	68,65585
0,95256	-0,11967	0,83289	66,44359	-53,21924	13,22435
0,95383	-0,02504	0,92879	81,40569	99,27000	180,67569
0,95502	-0,17483	0,78020	99,73703	-64,03212	35,70491
0,95616	0,00339	0,95955	122,19631	-20,27954	101,91676
0,95723	0,06942	1,02665	149,71308	88,04832	237,76140
0,95825	0,10933	1,06758	183,42621	23,29479	206,72100
0,95922	0,08361	1,04283	224,73104	76,64854	301,37958
0,96014	0,03627	0,99641	275,33709	-142,60073	132,73637
0,96102	-0,00535	0,95567	337,33887	-0,84813	336,49073

Tabela A. 3. Dados do experimento 3, sendo $\theta_1=0$, 06143; $\theta_2=0$, 10000; $\theta_3=0$, 00011 e $\sigma_{\varepsilon_i}=0$, $2\times\sigma_{y_i}$

$f_1(t)$	Erro 1	$f_1(t) + Erro 1$	$f_2(t)$	Erro 2	$f_2(t) + Erro 2$
0,00000	0,02077	0,02077	0,10001	0,00005	0,10006
0,04169	0,00404	0,04573	0,10002	0,00000	0,10002
0,06526	-0,01050	0,05476	0,10003	-0,00002	0,10001
0,08163	-0,01676	0,06488	0,10004	-0,00003	0,10001
0,09414	-0,00415	0,08999	0,10006	-0,00001	0,10005
0,10423	-0,01957	0,08466	0,10007	-0,00001	0,10005
0,11267	0,01283	0,12550	0,10008	0,00002	0,10010
0,11992	0,00532	0,12523	0,10009	0,00000	0,10009
0,12626	-0,01583	0,11043	0,10010	-0,00006	0,10004
0,13190	0,01042	0,14232	0,10011	0,00003	0,10014
0,13697	0,00934	0,14630	0,10012	0,00001	0,10013
0,14157	-0,00226	0,13930	0,10013	0,00000	0,10013
0,14578	-0,00585	0,13993	0,10014	-0,00001	0,10014
0,14966	-0,00879	0,14087	0,10016	-0,00003	0,10013
0,15325	0,01020	0,16345	0,10017	0,00002	0,10018
0,15660	-0,00510	0,15150	0,10018	-0,00001	0,10016
0,15974	0,01588	0,17561	0,10019	0,00003	0,10021
0,16268	-0,01255	0,15013	0,10020	-0,00004	0,10016
0,16546	0,01253	0,17799	0,10021	0,00004	0,10025
0,16809	0,00077	0,16886	0,10022	-0,00001	0,10021
0,17058	0,00603	0,17661	0,10023	0,00002	0,10026
0,17294	-0,01685	0,15609	0,10024	-0,00004	0,10020
0,17520	0,00652	0,18172	0,10026	0,00003	0,10028
0,17735	0,00453	0,18188	0,10027	0,00000	0,10027
0,17941	-0,01913	0,16028	0,10028	-0,00006	0,10022
0,18139	0,01013	0,19152	0,10029	0,00003	0,10032
0,18328	-0,00432	0,17896	0,10030	0,00000	0,10030
0,18510	-0,00098	0,18412	0,10031	-0,00001	0,10030
0,18686	-0,00618	0,18068	0,10032	0,00000	0,10032
0,18855	0,00035	0,18890	0,10033	0,00000	0,10033
0,19018	-0,01719	0,17300	0,10034	-0,00004	0,10030
0,19176	-0,00049	0,19127	0,10036	0,00002	0,10037
0,19329	0,00987	0,20316	0,10037	0,00002	0,10039
0,19477	0,00334	0,19811	0,10038	0,00001	0,10039
0,19620	-0,01703	0,17917	0,10039	-0,00005	0,10034
0,19759	0,00723	0,20481	0,10040	0,00002	0,10042
0,19894	-0,00793	0,19101	0,10041	-0,00002	0,10039
0,20025	-0,00382	0,19643	0,10042	0,00000	0,10042
0,20152	-0,01356	0,18796	0,10043	-0,00001	0,10043
0,20276	0,00179	0,20456	0,10044	0,00000	0,10044

Tabela A. 4. Dados do experimento 4, sendo $\theta_1=0$, 06143; $\theta_2=0$, 10000; $\theta_3=0$, 00011 e $\sigma_{\varepsilon_i}=0$, $8\times\sigma_{y_i}$

$f_1(t)$	Erro 1	$f_1(t) + Erro 1$	$f_2(t)$	Erro 2	$f_2(t) + Erro 2$
0,00000	0,00601	0,00601	0,10001	0,00002	0,10003
0,04169	-0,01193	0,02975	0,10002	-0,00001	0,10002
0,06526	-0,01727	0,04799	0,10003	-0,00004	0,09999
0,08163	0,00088	0,08251	0,10004	-0,00004	0,10001
0,09414	0,01690	0,11104	0,10006	0,00004	0,10010
0,10423	-0,01526	0,08896	0,10007	-0,00006	0,10001
0,11267	-0,02985	0,08282	0,10008	-0,00007	0,10001
0,11992	0,00709	0,12700	0,10009	0,00001	0,10009
0,12626	0,05720	0,18346	0,10010	0,00015	0,10025
0,13190	-0,01529	0,11660	0,10011	0,00004	0,10015
0,13697	-0,02572	0,11124	0,10012	-0,00007	0,10005
0,14157	-0,11123	0,03033	0,10013	-0,00036	0,09978
0,14578	0,04248	0,18826	0,10014	0,00017	0,10032
0,14966	0,02103	0,17068	0,10016	0,00005	0,10021
0,15325	0,09805	0,25130	0,10017	0,00026	0,10043
0,15660	0,02332	0,17992	0,10018	0,00006	0,10024
0,15974	-0,03468	0,12506	0,10019	0,00000	0,10019
0,16268	0,05981	0,22249	0,10020	0,00016	0,10036
0,16546	-0,01435	0,15111	0,10021	0,00001	0,10023
0,16809	0,04349	0,21158	0,10022	0,00015	0,10037
0,17058	0,04072	0,21130	0,10023	0,00007	0,10031
0,17294	-0,01115	0,16179	0,10024	-0,00005	0,10019
0,17520	-0,02252	0,15267	0,10026	-0,00001	0,10025
0,17735	0,00636	0,18371	0,10027	0,00002	0,10029
0,17941	-0,04602	0,13339	0,10028	-0,00001	0,10027
0,18139	-0,01146	0,16992	0,10029	-0,00003	0,10026
0,18328	0,03108	0,21436	0,10030	0,00010	0,10040
0,18510	0,00536	0,19047	0,10031	-0,00005	0,10026
0,18686	0,03637	0,22323	0,10032	0,00015	0,10048
0,18855	-0,04071	0,14784	0,10033	0,00002	0,10035
0,19018	-0,05971	0,13047	0,10034	-0,00017	0,10018
0,19176	-0,04992	0,14184	0,10036	-0,00012	0,10024
0,19329	0,02839	0,22168	0,10037	0,00005	0,10042
0,19477	-0,05021	0,14455	0,10038	-0,00005	0,10033
0,19620	0,00299	0,19919	0,10039	0,00000	0,10039
0,19759	0,02871	0,22630	0,10040	0,00010	0,10050
0,19894	0,04019	0,23913	0,10041	0,00000	0,10041
0,20025	-0,06462	0,13563	0,10042	-0,00028	0,10015
0,20152	-0,06521	0,13631	0,10043	-0,00010	0,10033
0,20276	-0,02528	0,17748	0,10044	-0,00010	0,10035

Tabela A. 5. Dados do experimento 5, sendo $\theta_1=0$, 75000; $\theta_2=1$, 00000; $\theta_3=0$, 08000 e $\sigma_{\varepsilon_i}=0$, $2\times\sigma_{y_i}$

$f_1(t)$	Erro 1	$f_1(t) + Erro 1$	$f_2(t)$	Erro 2	$f_2(t) + Erro 2$
0,00000	-0,01841	-0,01841	1,08329	45,91393	46,99722
0,40540	0,03000	0,43540	1,17351	-14,05616	-12,88265
0,56131	-0,02559	0,53572	1,27125	-1,43942	-0,16817
0,64645	-0,01531	0,63114	1,37713	-33,65854	-32,28142
0,70093	0,06199	0,76292	1,49182	-5,25465	-3,76282
0,73915	0,01967	0,75882	1,61607	31,92865	33,54473
0,76763	0,02239	0,79002	1,75067	-9,65739	-7,90672
0,78978	0,01469	0,80446	1,89648	-24,35143	-22,45495
0,80755	0,01062	0,81817	2,05443	15,07938	17,13381
0,82217	-0,00448	0,81769	2,22554	19,81817	22,04371
0,83444	0,02071	0,85515	2,41090	-16,37033	-13,95943
0,84490	-0,03038	0,81452	2,61170	6,20537	8,81706
0,85394	0,02151	0,87545	2,82922	28,39252	31,22173
0,86183	-0,04351	0,81833	3,06485	-1,49903	1,56583
0,86880	-0,01356	0,85524	3,32012	50,48205	53,80216
0,87500	-0,04153	0,83347	3,59664	-13,77857	-10,18193
0,88056	-0,04411	0,83644	3,89619	-31,28482	-27,38862
0,88557	0,01870	0,90427	4,22070	16,70855	20,92925
0,89012	-0,02027	0,86984	4,57223	-54,36222	-49,79000
0,89426	0,01063	0,90489	4,95303	22,04110	26,99414
0,89806	0,01221	0,91027	5,36556	-32,06470	-26,69914
0,90156	-0,03067	0,87089	5,81244	-48,30394	-42,49150
0,90479	-0,03784	0,86695	6,29654	-31,73571	-25,43918
0,90778	-0,01026	0,89752	6,82096	-11,79032	-4,96936
0,91056	-0,05297	0,85759	7,38906	-26,91851	-19,52946
0,91315	0,03308	0,94623	8,00447	37,50809	45,51256
0,91557	-0,02740	0,88817	8,67114	4,99526	13,66640
0,91785	0,01031	0,92815	9,39333	-18,65438	-9,26105
0,91998	0,01904	0,93902	10,17567	-7,96924	2,20644
0,92199	-0,00440	0,91759	11,02318	-49,57036	-38,54719
0,92388	-0,01671	0,90717	11,94126	-7,83707	4,10419
0,92567	0,06023	0,98590	12,93582	49,63063	62,56645
0,92737	-0,03478	0,89259	14,01320	28,83451	42,84772
0,92898	0,00779	0,93676	15,18032	0,61487	15,79520
0,93051	-0,00183	0,92867	16,44465	4,06975	20,51439
0,93196	-0,00827	0,92369	17,81427	7,80428	25,61855
0,93334	-0,04039	0,89295	19,29797	-29,72480	-10,42683
0,93466	0,00078	0,93545	20,90524	-31,99597	-11,09072
0,93592	-0,00589	0,93003	22,64638	-57,04749	-34,40111
0,93713	-0,05476	0,88237	24,53253	-20,01417	4,51836
0,93828	-0,04223	0,89606	26,57577	-47,37612	-20,80035
0,93939	0,00963	0,94901	28,78919	-14,76198	14,02721

				Contin	nuação da Tabela A.5
0,94045	0,03935	0,97980	31,18696	7,07428	38,26124
0,94147	-0,00676	0,93471	33,78443	13,71210	47,49653
0,94244	-0,00669	0,93576	36,59823	-33,19113	3,40710
0,94338	0,04948	0,99287	39,64639	61,30852	100,95491
0,94429	-0,00859	0,93570	42,94843	31,90439	74,85281
0,94516	0,00921	0,95437	46,52547	-11,47521	35,05026
0,94601	-0,01833	0,92768	50,40044	-43,54456	6,85589
0,94682	0,03121	0,97803	54,59815	-11,92971	42,66844
0,94760	0,02481	0,97241	59,14547	9,19676	68,34223
0,94836	0,02341	0,97177	64,07152	-12,13619	51,93533
0,94909	-0,05073	0,89836	69,40785	-37,92220	31,48566
0,94980	0,02050	0,97030	75,18863	7,44063	82,62926
0,95049	-0,00741	0,94308	81,45087	4,87282	86,32368
0,95115	-0,01454	0,93661	88,23467	-13,74642	74,48826
0,95179	0,00585	0,95764	95,58348	-11,83165	83,75183
0,95242	0,00328	0,95570	103,54435	-8,79972	94,74463
0,95303	0,02069	0,97371	112,16825	-3,89834	108,26991
0,95361	-0,05656	0,89706	121,51042	-70,61850	50,89192
0,95419	0,00168	0,95587	131,63066	-6,59491	125,03576
0,95474	-0,00946	0,94528	142,59380	-3,18394	139,40985
0,95528	0,00892	0,96420	154,47001	0,37707	154,84709
0,95581	-0,02285	0,93295	167,33537	-7,57798	159,75738
0,95632	0,00146	0,95778	181,27224	6,66060	187,93284
0,95681	-0,01244	0,94437	196,36988	26,30850	222,67837
0,95730	0,01455	0,97185	212,72495	24,21515	236,94009
0,95777	-0,02037	0,93740	230,44218	-36,67579	193,76639
0,95823	-0,01320	0,94503	249,63504	-14,22213	235,41291
0,95868	-0,02625	0,93243	270,42641	-69,72122	200,70519
0,95912	0,00108	0,96020	292,94943	36,30296	329,25239
0,95954	0,01824	0,97778	317,34833	-39,72711	277,62122
0,95996	0,02850	0,98846	343,77934	31,46197	375,24131
0,96037	-0,03988	0,92049	372,41171	23,89461	396,30632
0,96076	-0,02980	0,93096	403,42879	-29,06515	374,36365
0,96115	0,01767	0,97882	437,02919	7,42898	444,45818
0,96153	0,02826	0,98979	473,42807	-12,02837	461,39970
0,96190	0,00177	0,96367	512,85851	-11,67407	501,18444
0,96226	-0,02313	0,93914	555,57299	-23,26168	532,31131
0,96262	0,00113	0,96375	601,84504	-5,68710	596,15794

Tabela A. 6. Dados do experimento 6, sendo $\theta_1=0$, 76836; $\theta_2=1$, 00000; $\theta_3=0$, 08237 e $\sigma_{\varepsilon_i}=0$, $8\times\sigma_{y_i}$

$f_1(t)$	Erro 1	$f_1(t) + Erro 1$	$f_2(t)$	Erro 2	$f_2(t) + Erro 2$
0,00000	-0,04274	-0,04274	1,08586	-122,31680	-121,23095
0,41292	0,10286	0,51578	1,17908	-150,85752	-149,67843
0,57007	-0,06727	0,50281	1,28031	-110,32857	-109,04826
0,65533	-0,13668	0,51866	1,39024	-198,15724	-196,76701
0,70964	-0,07548	0,63416	1,50960	-3,64538	-2,13579
0,74760	0,04373	0,79133	1,63920	86,73273	88,37194
0,77579	-0,03616	0,73963	1,77994	-163,24706	-161,46712
0,79765	-0,01652	0,78113	1,93276	67,00317	68,93592
0,81516	0,13454	0,94970	2,09870	280,57565	282,67435
0,82953	-0,22033	0,60920	2,27888	48,17666	50,45554
0,84157	0,06663	0,90820	2,47454	127,78162	130,25615
0,85182	-0,01430	0,83752	2,68699	-7,31179	-4,62480
0,86066	-0,13049	0,73016	2,91768	37,80841	40,72609
0,86837	0,09504	0,96341	3,16818	-186,74290	-183,57472
0,87517	0,05098	0,92615	3,44019	149,58208	153,02227
0,88121	-0,02556	0,85565	3,73555	-222,05536	-218,31981
0,88661	-0,20629	0,68032	4,05627	-156,97566	-152,91939
0,89148	0,09341	0,98489	4,40452	-51,30799	-46,90347
0,89590	-0,05067	0,84523	4,78267	-85,61371	-80,83104
0,89992	0,01292	0,91285	5,19330	87,61789	92,81118
0,90360	0,02992	0,93352	5,63917	19,38797	25,02714
0,90699	0,14965	1,05663	6,12332	25,79287	31,91619
0,91011	-0,07011	0,84000	6,64905	-148,92339	-142,27434
0,91300	-0,01868	0,89432	7,21991	-56,80053	-49,58062
0,91569	-0,01742	0,89827	7,83978	-56,82664	-48,98686
0,91819	0,12114	1,03934	8,51287	73,63507	82,14794
0,92053	0,16893	1,08946	9,24375	-248,18789	-238,94415
0,92272	-0,09611	0,82661	10,03738	-246,20067	-236,16329
0,92478	0,05079	0,97557	10,89914	1,60303	12,50217
0,92671	0,09705	1,02376	11,83490	-141,01031	-129,17541
0,92854	-0,00748	0,92105	12,85099	-47,47507	-34,62408
0,93026	0,05356	0,98382	13,95432	93,49859	107,45292
0,93189	-0,02925	0,90264	15,15238	109,03544	124,18782
0,93343	-0,06515	0,86829	16,45330	-18,22503	-1,77173
0,93490	-0,07393	0,86096	17,86591	-169,40583	-151,53991
0,93629	-0,14559	0,79070	19,39981	30,12281	49,52262
0,93762	-0,09559	0,84202	21,06539	3,48300	24,54839
0,93888	-0,16827	0,77062	22,87398	-14,23068	8,64330
0,94009	0,13330	1,07339	24,83784	33,90407	58,74191
0,94125	-0,23977	0,70148	26,97031	-345,80987	-318,83956
0,94235	-0,04380	0,89855	29,28587	-88,13592	-58,85005
0,94341	0,01039	0,95380	31,80023	-33,51134	-1,71110

				Contir	nuação da Tabela A.6
0,94442	0,15931	1,10373	34,53047	17,27243	51,80289
0,94540	0,11242	1,05781	37,49511	10,62406	48,11916
0,94633	-0,08783	0,85850	40,71428	-93,32134	-52,60706
0,94723	0,04499	0,99222	44,20984	174,05548	218,26531
0,94809	-0,09499	0,85310	48,00551	-182,76719	-134,76168
0,94893	-0,04973	0,89920	52,12706	-135,15346	-83,02640
0,94973	-0,15690	0,79283	56,60247	42,05794	98,66040
0,95050	-0,11219	0,83831	61,46212	86,14888	147,61100
0,95125	-0,09528	0,85597	66,73899	-18,82038	47,91861
0,95197	-0,04795	0,90403	72,46892	-267,91146	-195,44254
0,95267	0,08376	1,03643	78,69080	122,74630	201,43710
0,95335	-0,08257	0,87077	85,44686	82,26208	167,70894
0,95400	0,04636	1,00036	92,78297	145,04094	237,82391
0,95463	-0,11321	0,84142	100,74892	-100,45112	0,29780
0,95524	0,04961	1,00485	109,39880	-51,12168	58,27712
0,95584	0,10263	1,05847	118,79132	-15,89517	102,89614
0,95641	0,03541	0,99182	128,99024	-198,50111	-69,51087
0,95697	0,18652	1,14349	140,06480	80,71839	220,78319
0,95752	0,11014	1,06765	152,09017	54,83227	206,92244
0,95804	-0,02820	0,92984	165,14799	-92,89730	72,25070
0,95856	-0,01286	0,94570	179,32690	-42,93393	136,39298
0,95906	0,06831	1,02737	194,72316	34,24897	228,97213
0,95954	-0,00405	0,95549	211,44127	32,08309	243,52436
0,96001	-0,10049	0,85952	229,59472	-56,02154	173,57318
0,96047	-0,12552	0,83496	249,30676	-230,36398	18,94278
0,96092	0,09078	1,05170	270,71119	-149,79360	120,91759
0,96135	0,13267	1,09403	293,95331	271,40937	565,36268
0,96178	-0,01014	0,95164	319,19090	87,67537	406,86627
0,96219	-0,03570	0,92650	346,59529	-101,28778	245,30751
0,96260	-0,17095	0,79165	376,35250	106,01251	482,36501
0,96299	0,34861	1,31160	408,66454	-28,51625	380,14829
0,96338	0,14930	1,11268	443,75075	161,66728	605,41803
0,96375	-0,01489	0,94886	481,84932	27,48697	509,33629
0,96412	-0,10128	0,86284	523,21887	-50,34786	472,87102
0,96448	0,10549	1,06997	568,14024	28,52016	596,66041
0,96483	-0,09859	0,86624	616,91837	34,42884	651,34721
0,96517	-0,04936	0,91581	669,88438	176,22985	846,11424
0,96551	-0,16636	0,79915	727,39783	-8,36015	719,03768

Tabela A. 7. Dados do experimento 7, sendo $\theta_1=-0$, 10000; $\theta_2=0$, 10000; $\theta_3=-0$, 10000 e $\sigma_{\varepsilon_l}=0$, $2\times\sigma_{y_l}=0$

$f_1(t)$	Erro 1	$f_1(t) + Erro 1$	$f_2(t)$	Erro 2	$f_2(t) + Erro 2$
0,00000	0,01555	0,01555	0,09048	0,00559	0,09608
-0,07177	-0,01697	-0,08874	0,08187	-0,00277	0,07911
-0,11612	-0,00094	-0,11706	0,07408	-0,00113	0,07296
-0,14870	-0,03091	-0,17961	0,06703	-0,00362	0,06341
-0,17462	0,02085	-0,15377	0,06065	0,00488	0,06553
-0,19623	0,01706	-0,17917	0,05488	0,00189	0,05677
-0,21481	-0,00118	-0,21599	0,04966	0,00056	0,05022
-0,23114	-0,04819	-0,27934	0,04493	-0,00659	0,03834
-0,24573	-0,04196	-0,28770	0,04066	-0,00843	0,03222
-0,25893	0,03320	-0,22573	0,03679	0,00460	0,04139
-0,27098	0,02175	-0,24923	0,03329	0,00383	0,03712
-0,28209	0,00561	-0,27647	0,03012	0,00164	0,03175
-0,29239	-0,00793	-0,30032	0,02725	0,00266	0,02991
-0,30201	-0,00103	-0,30304	0,02466	0,00042	0,02508
-0,31102	0,04700	-0,26402	0,02231	0,00695	0,02926
-0,31951	0,03746	-0,28205	0,02019	0,00572	0,02591
-0,32753	0,02218	-0,30535	0,01827	0,00207	0,02034
-0,33514	0,01382	-0,32132	0,01653	0,00400	0,02053
-0,34238	-0,01927	-0,36165	0,01496	-0,00252	0,01243
-0,34928	0,00340	-0,34588	0,01353	-0,00010	0,01344
-0,35588	0,01202	-0,34386	0,01225	0,00208	0,01432
-0,36220	0,00084	-0,36137	0,01108	-0,00005	0,01103
-0,36827	-0,00991	-0,37818	0,01003	-0,00233	0,00770
-0,37411	0,06175	-0,31236	0,00907	0,00818	0,01726
-0,37973	-0,02837	-0,40810	0,00821	-0,00476	0,00345
-0,38515	0,05656	-0,32860	0,00743	0,00933	0,01676
-0,39039	-0,01515	-0,40554	0,00672	0,00043	0,00715
-0,39545	-0,00439	-0,39985	0,00608	-0,00312	0,00296
-0,40036	0,01073	-0,38963	0,00550	0,00231	0,00781
-0,40512	0,02371	-0,38140	0,00498	0,00503	0,01001
-0,40973	-0,01301	-0,42274	0,00450	-0,00134	0,00317
-0,41421	0,03597	-0,37824	0,00408	0,00656	0,01063
-0,41857	0,02694	-0,39164	0,00369	0,00416	0,00785
-0,42281	0,01605	-0,40676	0,00334	0,00464	0,00798
-0,42694	-0,03020	-0,45715	0,00302	-0,00244	0,00058
-0,43097	-0,01090	-0,44186	0,00273	-0,00339	-0,00066
-0,43490	0,05072	-0,38418	0,00247	0,01078	0,01325
-0,43873	0,02508	-0,41365	0,00224	0,00473	0,00697
-0,44247	-0,03539	-0,47786	0,00202	-0,00358	-0,00155
-0,44613	-0,02315	-0,46927	0,00183	-0,00074	0,00109
-0,44970	-0,03250	-0,48220	0,00166	-0,00423	-0,00258
-0,45320	-0,02890	-0,48210	0,00150	-0,00281	-0,00131

				Conti	nuação da Tabela A.7
-0,45662	-0,02190	-0,47852	0,00136	-0,00197	-0,00061
-0,45997	-0,00607	-0,46604	0,00123	0,00059	0,00182
-0,46326	-0,01316	-0,47642	0,00111	-0,00352	-0,00241
-0,46648	-0,01310	-0,47958	0,00101	-0,00044	0,00056
-0,46964	-0,01324	-0,48287	0,00091	0,00024	0,00115
-0,47273	-0,00875	-0,48149	0,00082	0,00016	0,00098
-0,47577	-0,02785	-0,50362	0,00074	-0,00428	-0,00353
-0,47876	0,01213	-0,46663	0,00067	0,00140	0,00207
-0,48169	0,02114	-0,46055	0,00061	0,00463	0,00524
-0,48457	-0,01487	-0,49944	0,00055	-0,00368	-0,00313
-0,48740	0,01814	-0,46926	0,00050	0,00085	0,00135
-0,49018	-0,00423	-0,49441	0,00045	-0,00428	-0,00383
-0,49292	-0,05782	-0,55074	0,00041	-0,00983	-0,00942
-0,49561	-0,00709	-0,50270	0,00037	0,00049	0,00086
-0,49826	0,03134	-0,46692	0,00033	0,00115	0,00149
-0,50087	-0,01082	-0,51169	0,00030	-0,00115	-0,00084
-0,50344	0,02782	-0,47561	0,00027	0,00261	0,00288
-0,50597	0,03245	-0,47351	0,00025	0,00460	0,00485
-0,50846	0,01106	-0,49740	0,00022	0,00093	0,00115
-0,51091	-0,03687	-0,54778	0,00020	-0,00426	-0,00406
-0,51333	-0,03609	-0,54942	0,00018	-0,00568	-0,00549
-0,51572	-0,01727	-0,53298	0,00017	-0,00259	-0,00243
-0,51807	0,01086	-0,50721	0,00015	0,00133	0,00148
-0,52039	-0,01548	-0,53586	0,00014	-0,00415	-0,00401
-0,52268	0,01398	-0,50870	0,00012	0,00243	0,00255
-0,52493	0,02080	-0,50414	0,00011	0,00074	0,00085
-0,52716	-0,00927	-0,53644	0,00010	-0,00271	-0,00261
-0,52936	0,02798	-0,50138	0,00009	0,00672	0,00681
-0,53153	0,00007	-0,53146	0,00008	0,00055	0,00063
-0,53367	-0,00157	-0,53524	0,00007	-0,00168	-0,00160
-0,53579	-0,02030	-0,55609	0,00007	-0,00173	-0,00166
-0,53788	0,01214	-0,52574	0,00006	0,00276	0,00282
-0,53995	-0,00157	-0,54152	0,00006	0,00201	0,00206
-0,54199	0,01126	-0,53073	0,00005	0,00362	0,00367
-0,54401	0,01320	-0,53081	0,00005	-0,00109	-0,00104
-0,54600	-0,05354	-0,59954	0,00004	-0,01276	-0,01272
-0,54797	-0,01107	-0,55904	0,00004	-0,00495	-0,00491
-0,54992	-0,00657	-0,55649	0,00003	-0,00368	-0,00364

Tabela A. 8. Dados do experimento 8, sendo $\theta_1=$ 0, 10000; $\theta_2=$ 0, 10000; $\theta_3=$ 0, 01000 e $\sigma_{\varepsilon_i}=$ 0, 8 \times σ_{y_i}

$f_1(t)$	Erro 1	$f_1(t) + Erro 1$	$f_2(t)$	Erro 2	$f_2(t) + Erro 2$
0,00000	-0,01162	-0,01162	0,10101	0,00681	0,10782
0,06697	-0,11438	-0,04742	0,10202	-0,04455	0,05747
0,10404	0,12544	0,22948	0,10305	0,05531	0,15835
0,12945	-0,00562	0,12383	0,10408	-0,02357	0,08051
0,14866	-0,01945	0,12921	0,10513	-0,02033	0,08480
0,16404	0,03713	0,20117	0,10618	0,01664	0,12282
0,17683	0,15761	0,33444	0,10725	0,04381	0,15107
0,18775	0,04736	0,23511	0,10833	0,02253	0,13086
0,19726	0,08865	0,28591	0,10942	0,03213	0,14155
0,20567	-0,10352	0,10215	0,11052	-0,05553	0,05499
0,21321	0,03727	0,25047	0,11163	0,02722	0,13885
0,22002	-0,07661	0,14341	0,11275	-0,00859	0,10416
0,22624	0,01937	0,24561	0,11388	-0,00876	0,10513
0,23195	-0,04169	0,19026	0,11503	-0,04568	0,06935
0,23723	-0,02231	0,21492	0,11618	-0,01421	0,10197
0,24214	0,01503	0,25717	0,11735	-0,00230	0,11506
0,24672	0,00465	0,25137	0,11853	0,01840	0,13693
0,25102	0,03907	0,29009	0,11972	0,02436	0,14409
0,25505	0,03954	0,29459	0,12092	0,02265	0,14358
0,25887	0,05667	0,31553	0,12214	0,01907	0,14121
0,26247	0,01091	0,27338	0,12337	0,00291	0,12628
0,26590	0,02848	0,29438	0,12461	0,00075	0,12536
0,26915	0,04129	0,31045	0,12586	0,02767	0,15353
0,27226	0,08694	0,35919	0,12712	0,02584	0,15297
0,27522	0,04084	0,31606	0,12840	0,01493	0,14334
0,27806	-0,06574	0,21232	0,12969	-0,03105	0,09864
0,28078	0,05561	0,33638	0,13100	0,00660	0,13760
0,28339	-0,04575	0,23764	0,13231	-0,02430	0,10801
0,28590	0,03920	0,32510	0,13364	0,00783	0,14147
0,28831	0,00761	0,29593	0,13499	0,00148	0,13647
0,29064	0,04551	0,33615	0,13634	0,00583	0,14217
0,29289	-0,01103	0,28186	0,13771	0,00842	0,14613
0,29507	-0,00180	0,29327	0,13910	-0,00482	0,13428
0,29717	0,01853	0,31570	0,14049	0,00882	0,14932
0,29920	-0,08810	0,21110	0,14191	-0,05180	0,09010
0,30117	-0,03506	0,26612	0,14333	-0,00467	0,13866
0,30308	0,01319	0,31627	0,14477	0,02267	0,16744
0,30494	0,03622	0,34116	0,14623	0,01074	0,15697
0,30674	0,06217	0,36892	0,14770	0,03302	0,18072
0,30850	-0,02875	0,27975	0,14918	-0,02339	0,12579
0,31020	0,01014	0,32035	0,15068	0,00812	0,15881
0,31186	-0,10748	0,20438	0,15220	-0,04463	0,10757

				Conti	nuação da Tabela A.8
0,31348	-0,04699	0,26649	0,15373	-0,00714	0,14659
0,31506	-0,00386	0,31120	0,15527	0,00734	0,16261
0,31659	0,01271	0,32930	0,15683	-0,00441	0,15242
0,31809	-0,08743	0,23067	0,15841	-0,06004	0,09837
0,31956	-0,15823	0,16133	0,16000	-0,07401	0,08599
0,32099	-0,07316	0,24783	0,16161	-0,05439	0,10722
0,32239	0,01490	0,33729	0,16323	-0,00475	0,15848
0,32376	0,00735	0,33111	0,16487	0,00370	0,16858
0,32509	-0,05588	0,26922	0,16653	-0,01848	0,14805
0,32640	0,02538	0,35179	0,16820	0,00280	0,17100
0,32769	-0,03644	0,29125	0,16989	-0,00209	0,16780
0,32894	-0,06036	0,26858	0,17160	-0,01882	0,15278
0,33017	-0,00871	0,32146	0,17333	-0,00245	0,17088
0,33138	0,04009	0,37147	0,17507	0,01454	0,18960
0,33256	0,01349	0,34604	0,17683	0,02526	0,20209
0,33372	-0,02017	0,31354	0,17860	-0,01154	0,16706
0,33486	0,05744	0,39230	0,18040	0,02714	0,20754
0,33597	0,03857	0,37455	0,18221	0,00233	0,18455
0,33707	0,01490	0,35197	0,18404	-0,00049	0,18355
0,33815	0,00007	0,33822	0,18589	-0,02082	0,16507
0,33921	0,03134	0,37055	0,18776	0,00229	0,19005
0,34025	0,11069	0,45094	0,18965	0,04965	0,23930
0,34127	0,03812	0,37939	0,19155	0,00561	0,19716
0,34227	0,01717	0,35944	0,19348	0,00786	0,20134
0,34326	0,00382	0,34708	0,19542	0,00152	0,19694
0,34423	-0,03727	0,30697	0,19739	-0,02231	0,17508
0,34519	0,01297	0,35817	0,19937	0,00131	0,20068
0,34613	0,00644	0,35257	0,20138	0,00051	0,20189
0,34706	-0,10382	0,24324	0,20340	-0,03946	0,16394
0,34797	-0,12401	0,22396	0,20544	-0,04014	0,16531
0,34887	0,03266	0,38153	0,20751	0,01729	0,22479
0,34976	-0,02631	0,32345	0,20959	-0,01641	0,19318
0,35063	0,01362	0,36425	0,21170	0,01627	0,22797
0,35149	0,01343	0,36492	0,21383	0,00624	0,22006
0,35233	0,10412	0,45646	0,21598	0,04693	0,26291
0,35317	0,18009	0,53326	0,21815	0,09460	0,31275
0,35399	-0,01599	0,33800	0,22034	0,00707	0,22741
0,35480	0,09677	0,45158	0,22255	0,04935	0,27191

Tabela A. 9. Dados do experimento 9, sendo $\theta_1=0$, 84373; $\theta_2=0$, 10000; $\theta_3=0$, 19812 e $\sigma_{\varepsilon_i}=0$, $2\times\sigma_{y_i}$

$f_1(t)$	Erro 1	$f_1(t) + Erro 1$	$f_2(t)$	Erro 2	$f_2(t) + Erro 2$
0,00000	0,01741	0,01741	0,12191	0,71890	0,84081
0,44280	-0,11431	0,32849	0,14862	-2,43436	-2,28573
0,60423	0,02733	0,63156	0,18119	3,68511	3,86630
0,68953	0,04522	0,73475	0,22088	6,07217	6,29305
0,74281	0,01111	0,75392	0,26928	-1,88854	-1,61926
0,77948	-0,05008	0,72940	0,32828	-4,88043	-4,55215
0,80637	0,00171	0,80809	0,40021	11,30264	11,70285
0,82700	-0,00393	0,82308	0,48789	-22,96079	-22,47290
0,84337	-0,00580	0,83757	0,59479	2,09849	2,69328
0,85669	0,01794	0,87463	0,72512	-12,52189	-11,79678
0,86777	-0,01098	0,85678	0,88399	19,29394	20,17793
0,87713	0,01481	0,89193	1,07768	-0,07092	1,00676
0,88515	-0,04313	0,84202	1,31380	-9,38135	-8,06754
0,89211	-0,02152	0,87059	1,60166	3,24056	4,84223
0,89821	-0,05060	0,84761	1,95259	-19,86760	-17,91500
0,90361	0,00567	0,90928	2,38042	-13,51438	-11,13396
0,90841	-0,01480	0,89362	2,90197	-19,05773	-16,15575
0,91273	-0,05720	0,85552	3,53781	5,00796	8,54576
0,91662	0,07672	0,99334	4,31296	14,78180	19,09476
0,92015	-0,04648	0,87367	5,25794	-9,78504	-4,52710
0,92337	0,01239	0,93575	6,40998	-13,24705	-6,83707
0,92632	-0,03295	0,89337	7,81443	-7,82126	-0,00682
0,92903	-0,01155	0,91748	9,52661	-5,10004	4,42657
0,93153	-0,01861	0,91292	11,61393	-20,21680	-8,60288
0,93385	0,01551	0,94936	14,15859	8,08407	22,24266
0,93601	-0,01583	0,92018	17,26079	-18,28582	-1,02503
0,93801	-0,02774	0,91027	21,04270	4,91486	25,95756
0,93988	-0,06071	0,87917	25,65324	-11,48637	14,16687
0,94164	0,03429	0,97593	31,27397	0,20578	31,47975
0,94328	0,04653	0,98982	38,12623	-12,80592	25,32031
0,94483	-0,02003	0,92480	46,47984	10,85811	57,33795
0,94629	-0,04674	0,89955	56,66377	-17,40168	39,26209
0,94767	0,00349	0,95116	69,07903	-5,81067	63,26836
0,94897	0,02138	0,97035	84,21453	-16,70941	67,50512
0,95020	0,03065	0,98085	102,66628	-3,39220	99,27408
0,95137	0,00323	0,95460	125,16089	-20,83160	104,32929
0,95248	-0,02642	0,92607	152,58415	-3,08561	149,49854
0,95354	-0,06266	0,89088	186,01596	-11,01315	175,00282
0,95455	0,05008	1,00463	226,77283	2,77722	229,55004
0,95551	-0,00057	0,95494	276,45969	15,39648	291,85616

Tabela A. 10. Dados do experimento 10, sendo $\theta_1=0$, 87956; $\theta_2=0$, 10000; $\theta_3=0$, 20309 e $\sigma_{\varepsilon_l}=0$, $8\times\sigma_{y_l}$

$f_1(t)$	Erro 1	$f_1(t) + Erro 1$	$f_2(t)$	Erro 2	$f_2(t) + Erro 2$
0,00000	-0,20168	-0,20168	0,12252	8,31056	8,43308
0,45647	0,04698	0,50345	0,15011	-2,30681	-2,15670
0,61951	0,28485	0,90436	0,18391	39,81484	39,99875
0,70457	0,09051	0,79508	0,22532	46,92469	47,15001
0,75722	0,19299	0,95021	0,27606	-57,08047	-56,80441
0,79319	-0,09169	0,70150	0,33823	-1,80520	-1,46698
0,81941	-0,00055	0,81886	0,41439	-26,47962	-26,06523
0,83942	0,11843	0,95785	0,50771	-130,43946	-129,93175
0,85523	0,11252	0,96774	0,62204	-71,37935	-70,75732
0,86804	-0,05091	0,81713	0,76211	70,17707	70,93918
0,87865	-0,22302	0,65563	0,93372	-60,77516	-59,84144
0,88759	-0,09974	0,78785	1,14398	43,47601	44,61999
0,89523	-0,06063	0,83460	1,40159	-114,50101	-113,09942
0,90184	0,05853	0,96038	1,71721	-106,30846	-104,59125
0,90762	-0,23725	0,67038	2,10390	-62,51256	-60,40866
0,91272	-0,03739	0,87533	2,57767	-102,33948	-99,76181
0,91725	-0,30815	0,60910	3,15812	-67,27544	-64,11732
0,92131	0,19093	1,11224	3,86928	72,95405	76,82332
0,92497	-0,35171	0,57326	4,74058	-41,35360	-36,61302
0,92828	-0,10751	0,82076	5,80809	41,26106	47,06914
0,93129	-0,23184	0,69945	7,11598	-9,11342	-1,99744
0,93404	0,13592	1,06997	8,71839	69,37016	78,08855
0,93657	-0,18187	0,75470	10,68165	8,40825	19,08990
0,93890	0,14132	1,08023	13,08699	115,11124	128,19823
0,94106	-0,13289	0,80817	16,03399	56,49943	72,53341
0,94306	-0,26098	0,68207	19,64460	-49,75103	-30,10643
0,94492	0,19151	1,13643	24,06827	156,10288	180,17115
0,94665	0,21056	1,15721	29,48808	-67,71454	-38,22646
0,94827	0,01706	0,96533	36,12835	-132,86719	-96,73884
0,94979	-0,05835	0,89144	44,26391	-30,32882	13,93509
0,95122	-0,06693	0,88428	54,23148	77,03296	131,26444
0,95256	-0,07721	0,87535	66,44359	-34,00556	32,43803
0,95383	0,22754	1,18136	81,40569	-4,15348	77,25221
0,95502	0,03361	0,98863	99,73703	92,42898	192,16601
0,95616	-0,04838	0,90777	122,19631	-40,84810	81,34821
0,95723	-0,08510	0,87213	149,71308	47,98159	197,69467
0,95825	-0,10963	0,84862	183,42621	-67,04289	116,38332
0,95922	0,13334	1,09256	224,73104	-49,83975	174,89129
0,96014	-0,00568	0,95446	275,33709	36,86457	312,20166
0,96102	0,06626	1,02728	337,33887	37,99770	375,33656

Tabela A. 11. Dados do experimento 11, sendo $\theta_1 = 0$, 06143; $\theta_2 = 0$, 10000; $\theta_3 = 0$, 00011 e $\sigma_{\varepsilon_i} = 0$, 2 \times σ_{y_i}

$f_1(t)$	Erro 1	$f_1(t) + Erro 1$	$f_2(t)$	Erro 2	$f_2(t) + Erro 2$
0,00000	-0,00595	-0,00595	0,10001	-0,00002	0,10000
0,04169	0,01960	0,06129	0,10002	0,00004	0,10006
0,06526	-0,01357	0,05169	0,10003	-0,00004	0,09999
0,08163	-0,00846	0,07317	0,10004	-0,00005	0,10000
0,09414	-0,00199	0,09215	0,10006	-0,00001	0,10004
0,10423	0,00507	0,10930	0,10007	0,00000	0,10007
0,11267	0,01498	0,12765	0,10008	0,00003	0,10011
0,11992	0,00474	0,12465	0,10009	0,00003	0,10012
0,12626	0,01670	0,14296	0,10010	0,00004	0,10014
0,13190	-0,00883	0,12307	0,10011	-0,00004	0,10007
0,13697	0,00942	0,14639	0,10012	0,00003	0,10015
0,14157	0,01417	0,15574	0,10013	0,00005	0,10018
0,14578	-0,00927	0,13651	0,10014	-0,00003	0,10012
0,14966	0,00890	0,15856	0,10016	0,00005	0,10020
0,15325	0,00816	0,16141	0,10017	0,00002	0,10019
0,15660	-0,00161	0,15499	0,10018	-0,00002	0,10016
0,15974	0,00401	0,16375	0,10019	0,00003	0,10021
0,16268	-0,00126	0,16143	0,10020	-0,00001	0,10019
0,16546	0,02796	0,19342	0,10021	0,00008	0,10029
0,16809	0,02383	0,19191	0,10022	0,00005	0,10027
0,17058	-0,00277	0,16780	0,10023	0,00000	0,10023
0,17294	0,00488	0,17782	0,10024	0,00001	0,10025
0,17520	-0,00266	0,17253	0,10026	0,00001	0,10026
0,17735	-0,01027	0,16708	0,10027	-0,00003	0,10024
0,17941	-0,00495	0,17446	0,10028	0,00000	0,10027
0,18139	-0,00368	0,17771	0,10029	0,00001	0,10030
0,18328	-0,00553	0,17775	0,10030	-0,00002	0,10028
0,18510	-0,00984	0,17527	0,10031	-0,00002	0,10029
0,18686	-0,00076	0,18609	0,10032	-0,00001	0,10031
0,18855	0,00445	0,19300	0,10033	0,00003	0,10036
0,19018	0,01708	0,20727	0,10034	0,00004	0,10038
0,19176	-0,00019	0,19157	0,10036	-0,00001	0,10035
0,19329	0,01389	0,20718	0,10037	0,00001	0,10037
0,19477	0,01719	0,21195	0,10038	0,00003	0,10041
0,19620	0,01276	0,20896	0,10039	0,00005	0,10044
0,19759	-0,00580	0,19178	0,10040	0,00000	0,10040
0,19894	-0,00020	0,19874	0,10041	-0,00003	0,10039
0,20025	-0,01235	0,18790	0,10042	-0,00005	0,10038
0,20152	-0,00545	0,19608	0,10043	-0,00003	0,10040
0,20276	-0,00864	0,19412	0,10044	-0,00002	0,10043

Tabela A. 12. Dados do experimento 12, sendo $\theta_1 = 0$, 06143; $\theta_2 = 0$, 10000; $\theta_3 = 0$, 00011 e $\sigma_{\varepsilon_i} = 0$, $8 \times \sigma_{y_i}$

$f_1(t)$	Erro 1	$f_1(t) + Erro 1$	$f_2(t)$	Erro 2	$f_2(t) + Erro 2$
0,00000	0,00448	0,00448	0,10001	-0,00009	0,09993
0,04169	0,02116	0,06284	0,10002	0,00010	0,10013
0,06526	0,03415	0,09941	0,10003	0,00014	0,10017
0,08163	-0,02863	0,05301	0,10004	-0,00006	0,09998
0,09414	-0,03196	0,06218	0,10006	-0,00010	0,09996
0,10423	0,06626	0,17048	0,10007	0,00014	0,10021
0,11267	0,06187	0,17453	0,10008	0,00020	0,10028
0,11992	0,01989	0,13980	0,10009	0,00001	0,10010
0,12626	0,06617	0,19243	0,10010	0,00015	0,10025
0,13190	0,02011	0,15201	0,10011	0,00005	0,10016
0,13697	0,02521	0,16217	0,10012	0,00009	0,10021
0,14157	0,01660	0,15816	0,10013	0,00004	0,10017
0,14578	0,05025	0,19603	0,10014	0,00012	0,10026
0,14966	0,01691	0,16657	0,10016	0,00011	0,10026
0,15325	-0,02009	0,13316	0,10017	0,00002	0,10019
0,15660	0,05833	0,21493	0,10018	0,00022	0,10040
0,15974	-0,00564	0,15410	0,10019	0,00005	0,10024
0,16268	0,01854	0,18123	0,10020	0,00002	0,10021
0,16546	-0,01674	0,14872	0,10021	-0,00009	0,10012
0,16809	0,01115	0,17923	0,10022	0,00002	0,10024
0,17058	-0,03661	0,13397	0,10023	-0,00016	0,10007
0,17294	0,03383	0,20677	0,10024	0,00005	0,10030
0,17520	0,04269	0,21788	0,10026	0,00008	0,10034
0,17735	-0,01083	0,16652	0,10027	-0,00003	0,10024
0,17941	-0,02506	0,15435	0,10028	0,00002	0,10030
0,18139	-0,02140	0,15999	0,10029	-0,00004	0,10025
0,18328	0,01242	0,19570	0,10030	0,00003	0,10033
0,18510	0,02227	0,20737	0,10031	0,00007	0,10038
0,18686	-0,04915	0,13771	0,10032	-0,00011	0,10021
0,18855	0,04666	0,23522	0,10033	0,00008	0,10041
0,19018	0,01394	0,20413	0,10034	0,00005	0,10040
0,19176	-0,02518	0,16658	0,10036	-0,00001	0,10035
0,19329	-0,01214	0,18115	0,10037	-0,00005	0,10032
0,19477	0,02637	0,22113	0,10038	0,00006	0,10043
0,19620	0,08737	0,28357	0,10039	0,00022	0,10061
0,19759	-0,05112	0,14647	0,10040	-0,00022	0,10018
0,19894	-0,02961	0,16932	0,10041	-0,00004	0,10037
0,20025	0,02397	0,22422	0,10042	-0,00001	0,10041
0,20152	-0,00843	0,19309	0,10043	-0,00006	0,10038
0,20276	0,01392	0,21669	0,10044	0,00011	0,10055

Tabela A. 13. Dados do experimento 13, sendo $\theta_1=0$, 75000; $\theta_2=1$, 00000; $\theta_3=0$, 08000 e $\sigma_{\varepsilon_l}=0$, $2\times\sigma_{y_l}$

$f_1(t)$	Erro 1	$f_1(t) + Erro 1$	$f_2(t)$	Erro 2	$f_2(t) + Erro 2$
0,00000	-0,01326	-0,01326	1,08329	10,43444	11,51772
0,40540	-0,02391	0,38149	1,17351	-6,40093	-5,22742
0,56131	-0,00412	0,55719	1,27125	10,03147	11,30272
0,64645	0,03021	0,67665	1,37713	-24,06893	-22,69181
0,70093	0,00941	0,71034	1,49182	59,61000	61,10182
0,73915	0,00158	0,74074	1,61607	-2,17324	-0,55716
0,76763	0,00637	0,77400	1,75067	-24,78421	-23,03354
0,78978	-0,02395	0,76582	1,89648	-45,16263	-43,26615
0,80755	0,08176	0,88931	2,05443	23,02499	25,07942
0,82217	-0,00093	0,82125	2,22554	-36,09168	-33,86614
0,83444	0,05238	0,88682	2,41090	50,50244	52,91334
0,84490	0,05309	0,89798	2,61170	57,37013	59,98182
0,85394	-0,04439	0,80955	2,82922	-25,94459	-23,11537
0,86183	0,02244	0,88427	3,06485	-20,05525	-16,99040
0,86880	-0,01400	0,85480	3,32012	22,33915	25,65927
0,87500	0,01945	0,89445	3,59664	22,79937	26,39601
0,88056	-0,03709	0,84347	3,89619	8,89501	12,79120
0,88557	0,00436	0,88993	4,22070	-21,88706	-17,66636
0,89012	0,02281	0,91292	4,57223	-3,98143	0,59079
0,89426	0,03227	0,92653	4,95303	25,59673	30,54976
0,89806	0,03159	0,92965	5,36556	14,18898	19,55454
0,90156	0,04834	0,94990	5,81244	-15,23571	-9,42327
0,90479	0,00080	0,90558	6,29654	18,66312	24,95965
0,90778	0,04266	0,95044	6,82096	-34,83302	-28,01206
0,91056	0,00399	0,91455	7,38906	10,41411	17,80317
0,91315	0,05190	0,96505	8,00447	4,74419	12,74865
0,91557	-0,00373	0,91184	8,67114	-21,54729	-12,87615
0,91785	-0,00164	0,91620	9,39333	-0,30519	9,08814
0,91998	-0,04742	0,87256	10,17567	-52,49516	-42,31949
0,92199	0,02166	0,94365	11,02318	-29,51678	-18,49360
0,92388	-0,00195	0,92194	11,94126	-0,70598	11,23528
0,92567	-0,03510	0,89058	12,93582	-24,91193	-11,97611
0,92737	0,01394	0,94131	14,01320	19,00733	33,02053
0,92898	-0,01675	0,91223	15,18032	-10,17404	5,00628
0,93051	0,00701	0,93752	16,44465	5,84069	22,28533
0,93196	0,00531	0,93727	17,81427	23,16974	40,98402
0,93334	-0,03797	0,89537	19,29797	-31,29283	-11,99486
0,93466	0,00699	0,94165	20,90524	-39,18823	-18,28299
0,93592	0,03335	0,96928	22,64638	4,25606	26,90243
0,93713	0,01221	0,94934	24,53253	-18,82724	5,70529
0,93828	0,02685	0,96513	26,57577	-10,13428	16,44149
0,93939	-0,02645	0,91294	28,78919	26,01708	54,80627

				Continu	uação da Tabela A.13
0,94045	0,00033	0,94078	31,18696	8,33439	39,52135
0,94147	-0,00588	0,93559	33,78443	-10,53229	23,25214
0,94244	-0,00310	0,93934	36,59823	22,06271	58,66095
0,94338	-0,03753	0,90586	39,64639	-9,64285	30,00355
0,94429	0,01466	0,95895	42,94843	-9,45030	33,49813
0,94516	-0,01409	0,93107	46,52547	-89,83181	-43,30633
0,94601	0,02941	0,97542	50,40044	-13,35840	37,04205
0,94682	0,01178	0,95860	54,59815	5,12681	59,72496
0,94760	-0,02351	0,92409	59,14547	19,48108	78,62655
0,94836	-0,00860	0,93976	64,07152	22,18487	86,25640
0,94909	-0,04710	0,90199	69,40785	13,46696	82,87481
0,94980	-0,01566	0,93414	75,18863	-10,41369	64,77494
0,95049	0,02036	0,97085	81,45087	18,49078	99,94165
0,95115	-0,00075	0,95040	88,23467	-8,45088	79,78379
0,95179	-0,02241	0,92939	95,58348	12,06721	107,65069
0,95242	-0,05779	0,89463	103,54435	33,67234	137,21668
0,95303	-0,04980	0,90322	112,16825	-8,56988	103,59838
0,95361	-0,01319	0,94042	121,51042	-49,83810	71,67232
0,95419	0,03136	0,98555	131,63066	-17,76187	113,86879
0,95474	0,02865	0,98340	142,59380	-7,49401	135,09979
0,95528	-0,03191	0,92337	154,47001	18,35197	172,82198
0,95581	-0,03639	0,91942	167,33537	-31,35762	135,97775
0,95632	0,00150	0,95781	181,27224	-4,51231	176,75993
0,95681	-0,03116	0,92566	196,36988	18,13689	214,50676
0,95730	0,02050	0,97779	212,72495	25,67722	238,40217
0,95777	0,01438	0,97215	230,44218	31,25086	261,69305
0,95823	0,01330	0,97153	249,63504	10,45064	260,08567
0,95868	0,00616	0,96484	270,42641	-46,93430	223,49210
0,95912	-0,04581	0,91331	292,94943	-1,07426	291,87517
0,95954	-0,01973	0,93981	317,34833	15,50994	332,85827
0,95996	-0,00663	0,95333	343,77934	1,73261	345,51195
0,96037	0,01651	0,97687	372,41171	-19,01624	353,39547
0,96076	-0,03200	0,92876	403,42879	-13,81761	389,61118
0,96115	-0,01318	0,94797	437,02919	24,43181	461,46100
0,96153	0,01111	0,97264	473,42807	-26,59684	446,83124
0,96190	-0,03009	0,93181	512,85851	-16,15628	496,70223
0,96226	-0,00581	0,95645	555,57299	36,60107	592,17406
0,96262	0,02801	0,99063	601,84504	31,69904	633,54408

Tabela A. 14. Dados do experimento 14, sendo $\theta_1=0$, 76836; $\theta_2=1$, 00000; $\theta_3=0$, 08237 e $\sigma_{\varepsilon_i}=0$, $8\times\sigma_{y_i}$

$f_1(t)$	Erro 1	$f_1(t) + Erro 1$	$f_2(t)$	Erro 2	$f_2(t) + Erro 2$
0,00000	-0,25215	-0,25215	1,08586	225,60757	226,69343
0,41292	0,19759	0,61051	1,17908	306,97894	308,15802
0,57007	0,13303	0,70310	1,28031	48,97778	50,25809
0,65533	0,00401	0,65934	1,39024	201,21955	202,60979
0,70964	-0,12473	0,58491	1,50960	90,30766	91,81726
0,74760	-0,00402	0,74358	1,63920	-57,89628	-56,25708
0,77579	-0,04119	0,73460	1,77994	73,72211	75,50205
0,79765	0,12043	0,91808	1,93276	-28,13404	-26,20128
0,81516	-0,16624	0,64892	2,09870	151,94806	154,04676
0,82953	-0,09554	0,73400	2,27888	-113,20890	-110,93002
0,84157	-0,02020	0,82138	2,47454	-188,91425	-186,43971
0,85182	0,12171	0,97352	2,68699	-125,06007	-122,37308
0,86066	-0,02629	0,83436	2,91768	-106,09127	-103,17359
0,86837	-0,04780	0,82057	3,16818	80,32168	83,48986
0,87517	0,18241	1,05758	3,44019	111,19211	114,63230
0,88121	-0,01682	0,86439	3,73555	-84,86727	-81,13172
0,88661	0,13007	1,01668	4,05627	17,04684	21,10311
0,89148	-0,31076	0,58072	4,40452	-169,56469	-165,16017
0,89590	-0,20279	0,69311	4,78267	-183,47243	-178,68975
0,89992	-0,16646	0,73346	5,19330	-283,74101	-278,54771
0,90360	0,09615	0,99976	5,63917	-55,49474	-49,85557
0,90699	0,10065	1,00764	6,12332	115,88955	122,01288
0,91011	0,03668	0,94679	6,64905	-88,81294	-82,16389
0,91300	0,28309	1,19609	7,21991	67,84329	75,06320
0,91569	-0,05991	0,85578	7,83978	-195,21984	-187,38006
0,91819	0,12116	1,03936	8,51287	283,29164	291,80451
0,92053	-0,03194	0,88859	9,24375	-31,28105	-22,03730
0,92272	-0,16320	0,75952	10,03738	-167,71260	-157,67523
0,92478	0,17260	1,09738	10,89914	146,42824	157,32738
0,92671	0,07447	1,00118	11,83490	212,34384	224,17873
0,92854	-0,05949	0,86905	12,85099	72,38964	85,24064
0,93026	-0,00896	0,92130	13,95432	-100,84083	-86,88651
0,93189	0,00183	0,93371	15,15238	113,32803	128,48042
0,93343	-0,09429	0,83914	16,45330	42,01676	58,47007
0,93490	-0,12267	0,81223	17,86591	-189,92120	-172,05529
0,93629	0,21218	1,14847	19,39981	-150,18030	-130,78049
0,93762	0,03353	0,97115	21,06539	-102,85940	-81,79401
0,93888	0,02242	0,96131	22,87398	44,93174	67,80572
0,94009	-0,06253	0,87756	24,83784	-2,67230	22,16555
0,94125	0,01556	0,95680	26,97031	158,32246	185,29278
0,94235	0,06548	1,00783	29,28587	-109,31686	-80,03098
0,94341	-0,02919	0,91422	31,80023	-178,13237	-146,33214

Continu	ıação da Tabela A.14
9063	4,03983

				Continu	iação da Tabela A.14
0,94442	-0,17473	0,76970	34,53047	-30,49063	4,03983
0,94540	-0,03670	0,90869	37,49511	-7,59431	29,90080
0,94633	-0,21995	0,72638	40,71428	-229,33478	-188,62050
0,94723	-0,24196	0,70527	44,20984	-12,54368	31,66615
0,94809	-0,16001	0,78808	48,00551	-81,30804	-33,30253
0,94893	0,01699	0,96592	52,12706	-145,09149	-92,96444
0,94973	-0,18202	0,76770	56,60247	-136,10648	-79,50401
0,95050	-0,04888	0,90162	61,46212	-31,14608	30,31604
0,95125	0,05456	1,00581	66,73899	-175,55177	-108,81278
0,95197	0,05181	1,00379	72,46892	188,33677	260,80569
0,95267	-0,29455	0,65812	78,69080	-104,87233	-26,18153
0,95335	0,07489	1,02824	85,44686	12,96070	98,40756
0,95400	0,01730	0,97130	92,78297	246,64534	339,42831
0,95463	-0,00830	0,94633	100,74892	182,22586	282,97478
0,95524	-0,06162	0,89363	109,39880	87,09208	196,49088
0,95584	0,01170	0,96754	118,79132	-102,81602	15,97530
0,95641	0,00614	0,96255	128,99024	90,16705	219,15729
0,95697	-0,09447	0,86250	140,06480	-171,64580	-31,58100
0,95752	-0,02975	0,92777	152,09017	-140,77158	11,31860
0,95804	0,18241	1,14045	165,14799	236,61380	401,76180
0,95856	0,07102	1,02957	179,32690	278,65666	457,98356
0,95906	0,05649	1,01555	194,72316	-133,33416	61,38899
0,95954	-0,06196	0,89758	211,44127	108,39588	319,83714
0,96001	0,11661	1,07662	229,59472	9,75847	239,35320
0,96047	-0,09438	0,86609	249,30676	-61,04149	188,26527
0,96092	-0,17740	0,78352	270,71119	-101,56269	169,14850
0,96135	-0,01035	0,95101	293,95331	-105,48225	188,47106
0,96178	0,08247	1,04425	319,19090	54,78128	373,97218
0,96219	-0,03726	0,92494	346,59529	52,96742	399,56271
0,96260	0,04754	1,01014	376,35250	-133,70335	242,64914
0,96299	-0,07909	0,88391	408,66454	118,14591	526,81044
0,96338	0,22254	1,18591	443,75075	-35,74651	408,00424
0,96375	-0,00763	0,95612	481,84932	14,71068	496,56000
0,96412	0,09229	1,05641	523,21887	155,15281	678,37169
0,96448	0,12420	1,08868	568,14024	-104,00213	464,13812
0,96483	0,04871	1,01354	616,91837	52,48551	669,40388
0,96517	-0,00091	0,96426	669,88438	-176,79426	493,09012
0,96551	0,22408	1,18959	727,39783	-33,49127	693,90656

Tabela A. 15. Dados do experimento 15, sendo $\theta_1 = -0$, 10000; $\theta_2 = 0$, 1000; $\theta_3 = -0$, 10000 e $\sigma_{\varepsilon_l} = 0$, $2 \times \sigma_{y_l}$

$f_1(t)$	Erro 1	$f_1(t) + Erro 1$	$f_2(t)$	Erro 2	$f_2(t) + Erro 2$
0,00000	0,03987	0,03987	0,09048	0,00353	0,09401
-0,07177	0,03004	-0,04173	0,08187	0,00683	0,08870
-0,11612	0,04126	-0,07486	0,07408	0,00662	0,08070
-0,14870	-0,03826	-0,18695	0,06703	-0,00675	0,06029
-0,17462	-0,00942	-0,18404	0,06065	-0,00379	0,05686
-0,19623	-0,01425	-0,21048	0,05488	-0,00561	0,04927
-0,21481	0,03566	-0,17915	0,04966	0,00767	0,05733
-0,23114	0,00611	-0,22503	0,04493	0,00343	0,04836
-0,24573	0,01257	-0,23316	0,04066	0,00116	0,04181
-0,25893	0,01053	-0,24840	0,03679	0,00214	0,03893
-0,27098	0,03347	-0,23751	0,03329	0,00698	0,04027
-0,28209	-0,00628	-0,28837	0,03012	0,00141	0,03153
-0,29239	-0,01255	-0,30494	0,02725	0,00063	0,02788
-0,30201	0,03698	-0,26503	0,02466	0,00203	0,02669
-0,31102	0,00788	-0,30314	0,02231	0,00048	0,02279
-0,31951	0,02519	-0,29431	0,02019	0,00456	0,02475
-0,32753	-0,02192	-0,34945	0,01827	-0,00335	0,01492
-0,33514	-0,01797	-0,35311	0,01653	-0,00043	0,01610
-0,34238	-0,01281	-0,35519	0,01496	-0,00359	0,01136
-0,34928	-0,00959	-0,35888	0,01353	-0,00097	0,01257
-0,35588	-0,00542	-0,36131	0,01225	-0,00022	0,01203
-0,36220	0,03752	-0,32468	0,01108	0,00679	0,01787
-0,36827	-0,02274	-0,39101	0,01003	-0,00336	0,00666
-0,37411	-0,01581	-0,38992	0,00907	-0,00544	0,00363
-0,37973	0,00379	-0,37594	0,00821	0,00364	0,01184
-0,38515	-0,02506	-0,41021	0,00743	-0,00385	0,00358
-0,39039	0,01661	-0,37378	0,00672	0,00060	0,00732
-0,39545	0,02398	-0,37148	0,00608	0,00383	0,00991
-0,40036	-0,01597	-0,41633	0,00550	-0,00130	0,00421
-0,40512	-0,03743	-0,44255	0,00498	-0,00738	-0,00240
-0,40973	0,03527	-0,37446	0,00450	0,00224	0,00674
-0,41421	-0,02472	-0,43893	0,00408	-0,00217	0,00190
-0,41857	-0,00803	-0,42660	0,00369	-0,00193	0,00175
-0,42281	0,02583	-0,39699	0,00334	0,00569	0,00903
-0,42694	0,01059	-0,41635	0,00302	0,00026	0,00328
-0,43097	0,01774	-0,41323	0,00273	0,00014	0,00287
-0,43490	0,00085	-0,43404	0,00247	0,00105	0,00352
-0,43873	-0,03292	-0,47165	0,00224	-0,00380	-0,00156
-0,44247	-0,03199	-0,47446	0,00202	-0,00443	-0,00240
-0,44613	-0,00548	-0,45161	0,00183	-0,00004	0,00179
-0,44970	0,01329	-0,43641	0,00166	0,00260	0,00425
-0,45320	-0,00365	-0,45685	0,00150	-0,00169	-0,00019

				Continu	uação da Tabela A.15
-0,45662	-0,00942	-0,46604	0,00136	0,00181	0,00317
-0,45997	0,00369	-0,45629	0,00123	0,00071	0,00194
-0,46326	-0,04306	-0,50632	0,00111	-0,00583	-0,00472
-0,46648	-0,02324	-0,48972	0,00101	-0,00240	-0,00140
-0,46964	-0,00017	-0,46981	0,00091	-0,00094	-0,00003
-0,47273	0,02877	-0,44396	0,00082	0,00322	0,00404
-0,47577	0,01076	-0,46501	0,00074	-0,00233	-0,00159
-0,47876	-0,02360	-0,50236	0,00067	-0,00266	-0,00199
-0,48169	-0,01410	-0,49579	0,00061	-0,00468	-0,00407
-0,48457	0,01320	-0,47137	0,00055	0,00308	0,00363
-0,48740	0,00531	-0,48209	0,00050	-0,00038	0,00012
-0,49018	-0,00546	-0,49564	0,00045	-0,00307	-0,00261
-0,49292	0,00034	-0,49258	0,00041	0,00080	0,00121
-0,49561	0,01296	-0,48265	0,00037	0,00048	0,00085
-0,49826	-0,03105	-0,52931	0,00033	-0,00613	-0,00579
-0,50087	0,03678	-0,46409	0,00030	0,00423	0,00453
-0,50344	0,04462	-0,45882	0,00027	0,00666	0,00693
-0,50597	0,01068	-0,49528	0,00025	0,00429	0,00454
-0,50846	0,03265	-0,47581	0,00022	0,00519	0,00541
-0,51091	-0,01543	-0,52635	0,00020	-0,00675	-0,00655
-0,51333	0,02096	-0,49237	0,00018	0,00329	0,00347
-0,51572	0,00251	-0,51321	0,00017	0,00326	0,00343
-0,51807	0,05000	-0,46807	0,00015	0,00388	0,00403
-0,52039	-0,00182	-0,52220	0,00014	-0,00129	-0,00115
-0,52268	-0,05167	-0,57435	0,00012	-0,00559	-0,00546
-0,52493	0,02582	-0,49911	0,00011	0,00386	0,00397
-0,52716	0,04224	-0,48492	0,00010	0,00501	0,00511
-0,52936	-0,00388	-0,53324	0,00009	-0,00012	-0,00003
-0,53153	0,00514	-0,52639	0,00008	0,00843	0,00851
-0,53367	0,00907	-0,52461	0,00007	0,00080	0,00088
-0,53579	0,02243	-0,51336	0,00007	0,00281	0,00288
-0,53788	0,03252	-0,50536	0,00006	0,00351	0,00357
-0,53995	-0,00036	-0,54031	0,00006	0,00018	0,00024
-0,54199	0,00228	-0,53971	0,00005	0,00300	0,00305
-0,54401	0,00316	-0,54085	0,00005	0,00007	0,00012
-0,54600	0,00954	-0,53646	0,00004	0,00061	0,00065
-0,54797	0,00938	-0,53859	0,00004	0,00171	0,00174
-0,54992	0,01665	-0,53327	0,00003	0,00302	0,00305

Tabela A. 16. Dados do experimento 16, sendo $\theta_1=0$, 10000; $\theta_2=0$, 10000; $\theta_3=0$, 01000 e $\sigma_{\varepsilon_l}=0$, $8\times\sigma_{y_l}$

$f_1(t)$	Erro 1	$f_1(t) + Erro 1$	$f_2(t)$	Erro 2	$f_2(t) + Erro 2$
0,00000	0,00549	0,00549	0,10101	-0,00585	0,09516
0,06697	-0,07794	-0,01098	0,10202	-0,03926	0,06276
0,10404	-0,03417	0,06987	0,10305	-0,03273	0,07032
0,12945	-0,11094	0,01851	0,10408	-0,06641	0,03768
0,14866	-0,01815	0,13051	0,10513	-0,01129	0,09384
0,16404	0,02759	0,19163	0,10618	0,01015	0,11634
0,17683	0,02057	0,19740	0,10725	0,01103	0,11828
0,18775	0,07198	0,25973	0,10833	0,03890	0,14723
0,19726	0,04367	0,24093	0,10942	0,01891	0,12833
0,20567	0,04125	0,24692	0,11052	0,02164	0,13215
0,21321	-0,05344	0,15977	0,11163	-0,02228	0,08935
0,22002	-0,09497	0,12505	0,11275	-0,02307	0,08968
0,22624	0,01239	0,23863	0,11388	0,01558	0,12946
0,23195	-0,12799	0,10396	0,11503	-0,06670	0,04833
0,23723	0,00852	0,24575	0,11618	-0,01196	0,10423
0,24214	-0,03506	0,20708	0,11735	0,00407	0,12142
0,24672	0,01873	0,26545	0,11853	-0,00285	0,11568
0,25102	0,07084	0,32185	0,11972	0,03958	0,15930
0,25505	0,00284	0,25790	0,12092	0,00147	0,12240
0,25887	0,04744	0,30630	0,12214	0,01243	0,13457
0,26247	0,02734	0,28981	0,12337	0,01487	0,13823
0,26590	0,00797	0,27386	0,12461	-0,00471	0,11990
0,26915	0,04853	0,31769	0,12586	0,03398	0,15984
0,27226	0,08888	0,36113	0,12712	0,05002	0,17714
0,27522	-0,01384	0,26138	0,12840	-0,01873	0,10967
0,27806	-0,03079	0,24727	0,12969	-0,03374	0,09596
0,28078	-0,03546	0,24531	0,13100	-0,01386	0,11714
0,28339	0,02761	0,31100	0,13231	0,00770	0,14001
0,28590	0,07526	0,36116	0,13364	0,03361	0,16725
0,28831	0,05768	0,34599	0,13499	0,02474	0,15973
0,29064	0,07830	0,36895	0,13634	0,07769	0,21403
0,29289	-0,08845	0,20444	0,13771	-0,04904	0,08867
0,29507	-0,10963	0,18543	0,13910	-0,04299	0,09610
0,29717	-0,03343	0,26374	0,14049	-0,01869	0,12180
0,29920	-0,01504	0,28416	0,14191	0,00228	0,14418
0,30117	-0,00322	0,29795	0,14333	-0,00851	0,13483
0,30308	-0,02216	0,28092	0,14477	0,01729	0,16206
0,30494	0,02479	0,32973	0,14623	-0,00550	0,14073
0,30674	-0,07646	0,23028	0,14770	-0,03916	0,10853
0,30850	-0,12383	0,18466	0,14918	-0,04911	0,10007
0,31020	0,11571	0,42591	0,15068	0,05655	0,20723
0,31186	0,05168	0,36354	0,15220	0,01336	0,16556

				Continu	uação da Tabela A.16
0,31348	-0,07378	0,23970	0,15373	-0,03750	0,11623
0,31506	0,00336	0,31842	0,15527	0,01270	0,16797
0,31659	0,05289	0,36948	0,15683	0,03794	0,19477
0,31809	-0,04556	0,27253	0,15841	-0,01048	0,14793
0,31956	0,02162	0,34117	0,16000	0,00859	0,16859
0,32099	0,00066	0,32165	0,16161	0,00394	0,16555
0,32239	-0,03208	0,29031	0,16323	0,00244	0,16567
0,32376	0,02602	0,34978	0,16487	0,01508	0,17995
0,32509	0,08041	0,40550	0,16653	0,02388	0,19041
0,32640	0,02067	0,34707	0,16820	-0,00207	0,16613
0,32769	0,11325	0,44093	0,16989	0,03179	0,20168
0,32894	-0,06141	0,26753	0,17160	-0,00457	0,16703
0,33017	-0,01362	0,31655	0,17333	0,00559	0,17892
0,33138	0,03389	0,36527	0,17507	0,00103	0,17610
0,33256	-0,06682	0,26574	0,17683	-0,02074	0,15609
0,33372	-0,01777	0,31594	0,17860	-0,02216	0,15644
0,33486	-0,00491	0,32995	0,18040	-0,01114	0,16926
0,33597	-0,02281	0,31317	0,18221	0,00825	0,19046
0,33707	0,00662	0,34369	0,18404	-0,00210	0,18195
0,33815	-0,04306	0,29509	0,18589	-0,01184	0,17406
0,33921	0,07944	0,41865	0,18776	0,03867	0,22643
0,34025	0,04171	0,38196	0,18965	0,00150	0,19115
0,34127	0,00588	0,34715	0,19155	-0,00204	0,18951
0,34227	0,01934	0,36161	0,19348	0,01210	0,20558
0,34326	-0,00152	0,34174	0,19542	0,00158	0,19701
0,34423	-0,15194	0,19229	0,19739	-0,05053	0,14686
0,34519	0,01221	0,35740	0,19937	0,02104	0,22041
0,34613	0,01590	0,36203	0,20138	0,01001	0,21139
0,34706	-0,08104	0,26602	0,20340	-0,02162	0,18178
0,34797	0,01158	0,35955	0,20544	-0,00462	0,20082
0,34887	-0,09102	0,25785	0,20751	-0,05158	0,15592
0,34976	0,05279	0,40255	0,20959	0,03775	0,24734
0,35063	0,02471	0,37534	0,21170	0,00977	0,22147
0,35149	-0,01477	0,33672	0,21383	0,01651	0,23033
0,35233	0,05526	0,40759	0,21598	0,02059	0,23657
0,35317	0,02341	0,37658	0,21815	0,01436	0,23251
0,35399	0,02164	0,37564	0,22034	0,03140	0,25174
0,35480	-0,08568	0,26912	0,22255	-0,04115	0,18141

APÊNDICE B - ANÁLISE DOS RESÍDUOS

Para a análise de resíduos foram avaliadas, inicialmente, questões como como normalidade, presença de agrupamentos (clusters), misturas, oscilações e tendências, sendo que novamente o software Minitab foi utilizado. O teste para normalidade de Anderson-Darling foi realizado nesse trabalho e seu uso pode ser visualizado por meio da Figura B. 1. A hipótese nula para esse teste indica que os dados são normalmente distribuídos, assim é desejável que seja obtido um p-value superior a 0,05 e assim não se rejeite H_0 .

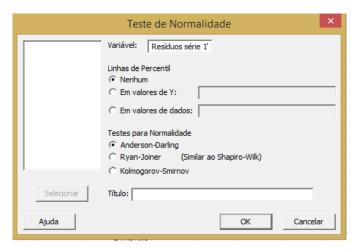


Figura B. 1. Tela do Minitab para a realização do teste de normalidade

Em relação aos demais testes anteriormente mencionados, utilizou-se o recurso de gráficos de ensaios do Minitab, conforme mostrado na Figura B. 2. Esse recurso possibilita analisar a presença de agrupamentos, tendências, oscilações e misturas, sendo que a hipótese nula para os testes indica que não há presença de nenhuma dessas características.

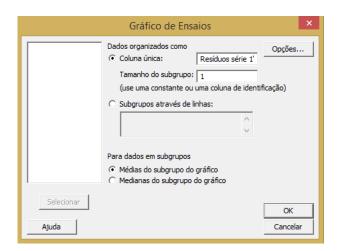


Figura B. 2. Tela do Minitab para realização da análise de resíduos

Além disso, também desenvolveu-se a análise de autocorrelação para os resíduos de cada experimento do DOE. A Figura B. 3 mostra como gerar o gráfico de autocorrelação no Minitab, bem como armazenar as estatísticas de t e Q de Ljung-Box para verificar se os resíduos poderão ser considerados *White noise* ou não. É interessante destacar que foi considerado o número padrão de defasagens, ou seja, um quarto do tamanho da série temporal (T/4), já que segundo Montgomery (2008) trata-se de uma regra comumente utilizada para definir a quantidade de dados a serem utilizados nessas análises.



Figura B. 3. Tela do Minitab para geração dos gráficos de autocorrelação

Nas Figuras de 39 a 70, podem-se observar, em azul, os valores reais das séries temporais sintéticas e em vermelho os valores ajustados pelos métodos MSM e RNA para cada um dos 16 experimentos. Além disso, também estão apresentados os gráficos das funções de autocorrelação para resíduos

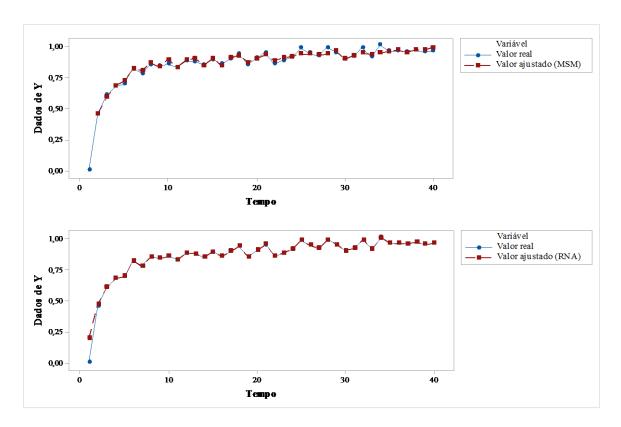


Figura B. 4. Gráfico dos valores reais e ajustados pelos métodos MSM (acima) e RNA (abaixo) ao longo do tempo para o experimento 1

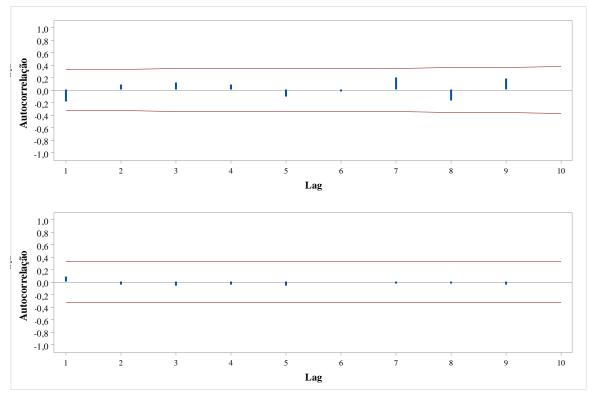


Figura B. 5. Função de autocorrelação para a série de resíduos gerados pelos métodos MSM (acima) e RNA (abaixo) na modelagem do experimento 1

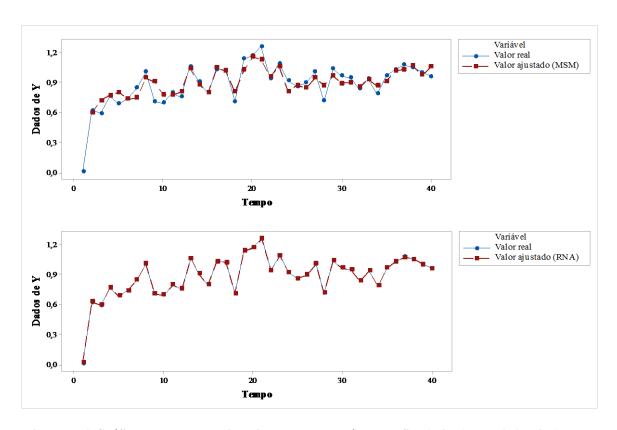


Figura B. 6. Gráfico dos valores reais e ajustados pelos métodos MSM (acima) e RNA (abaixo) ao longo do tempo para o experimento 2

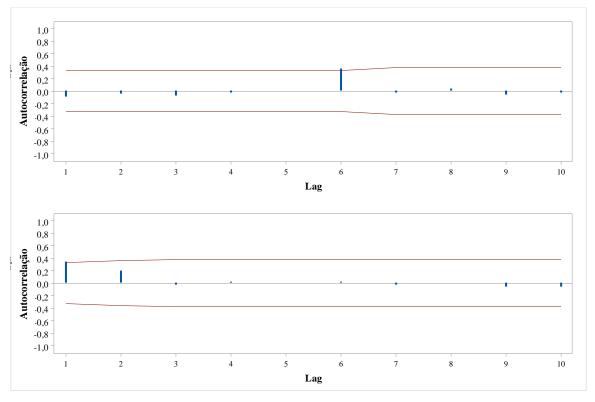


Figura B. 7. Função de autocorrelação para a série de resíduos gerados pelos métodos MSM (acima) e RNA (abaixo) na modelagem do experimento 2

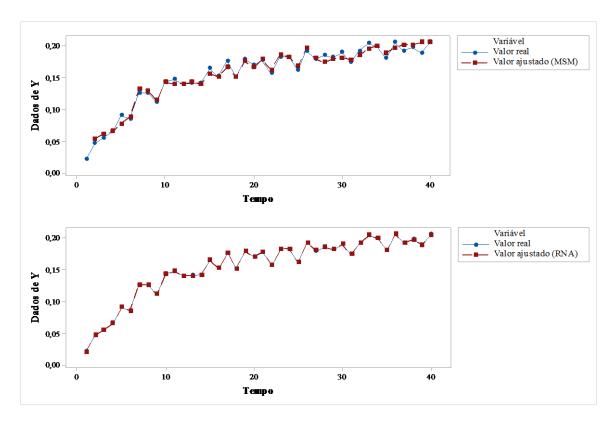


Figura B. 8. Gráfico dos valores reais e ajustados pelos métodos MSM (acima) e RNA (abaixo) ao longo do tempo para o experimento 3

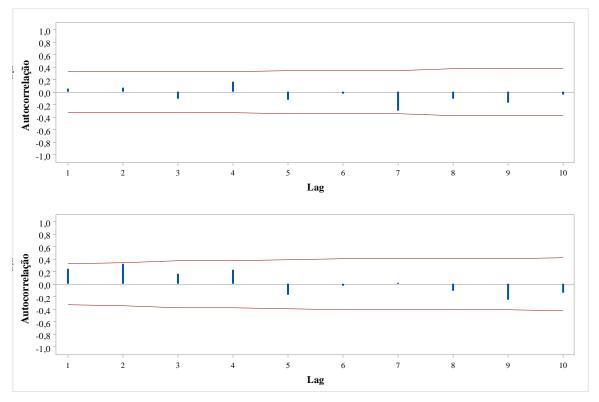


Figura B. 9. Função de autocorrelação para a série de resíduos gerados pelos métodos MSM (acima) e RNA (abaixo) na modelagem do experimento 3

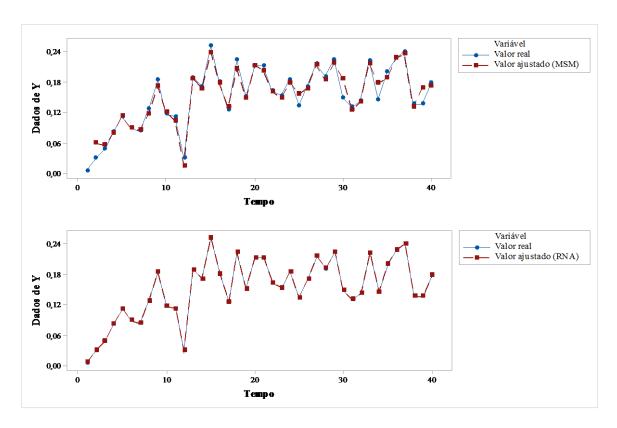


Figura B. 10. Gráfico dos valores reais e ajustados pelos métodos MSM (acima) e RNA (abaixo) ao longo do tempo para o experimento 4

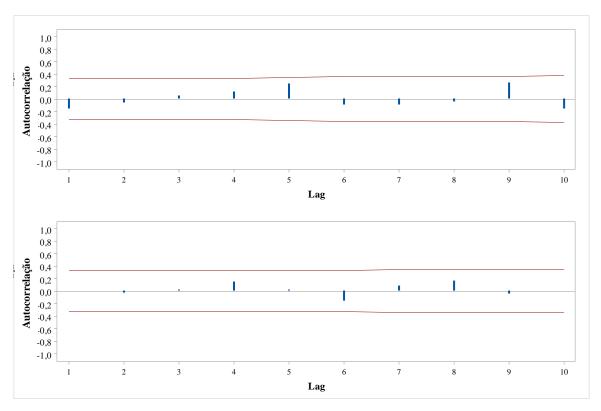


Figura B. 11. Função de autocorrelação para a série de resíduos gerados pelos métodos MSM (acima) e RNA (abaixo) na modelagem do experimento 4

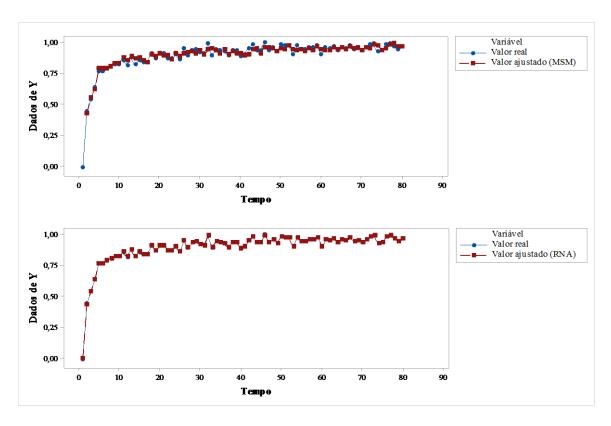


Figura B. 12. Gráfico dos valores reais e ajustados pelos métodos MSM (acima) e RNA (abaixo) ao longo do tempo para o exercício 5

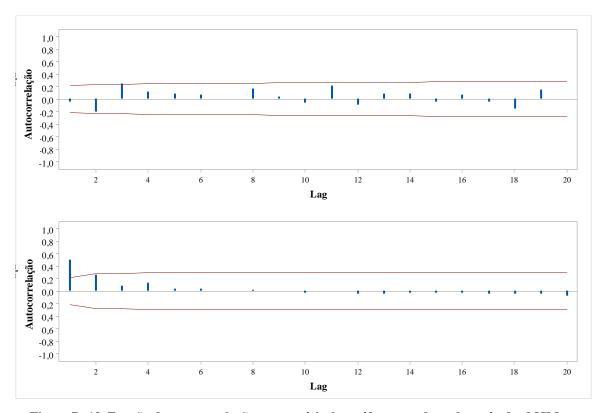


Figura B. 13. Função de autocorrelação para a série de resíduos gerados pelos métodos MSM (acima) e RNA (abaixo) na modelagem do experimento 5

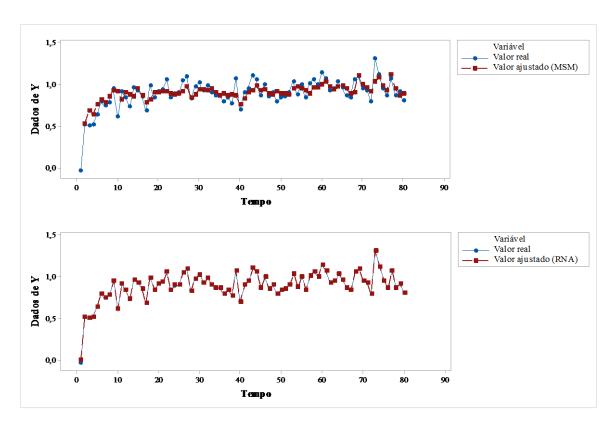


Figura B. 14. Gráfico dos valores reais e ajustados pelos métodos MSM (acima) e RNA (abaixo) ao longo do tempo para o experimento 6

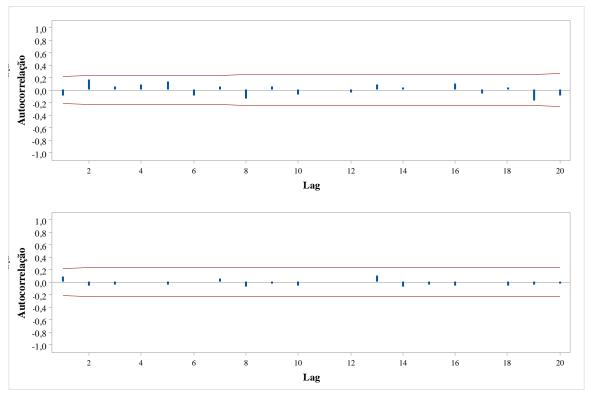


Figura B. 15. Função de autocorrelação para a série de resíduos gerados pelos métodos MSM (acima) e RNA (abaixo) na modelagem do experimento 6

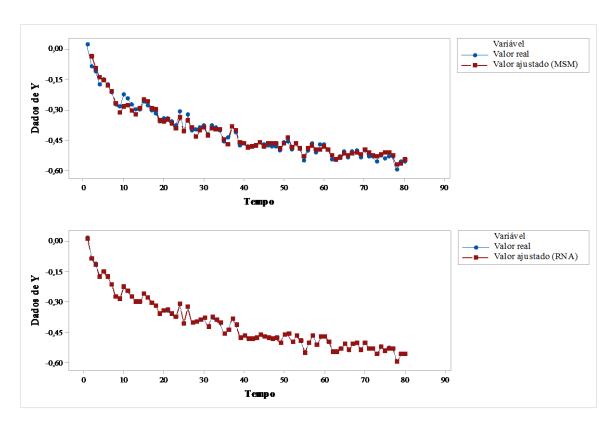


Figura B. 16. Gráfico dos valores reais e ajustados pelos métodos MSM (acima) e RNA (abaixo) ao longo do tempo para o experimento 7

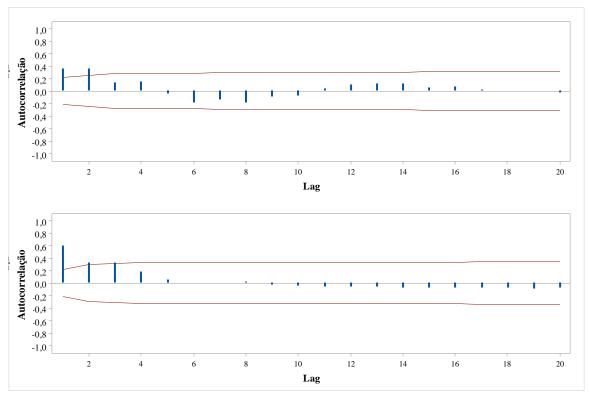


Figura B. 17. Função de autocorrelação para a série de resíduos gerados pelos métodos MSM (acima) e RNA (abaixo) na modelagem do experimento 7

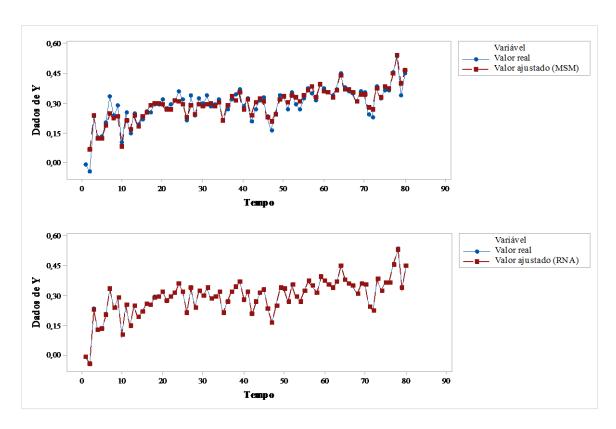


Figura B. 18. Gráfico dos valores reais e ajustados pelos métodos MSM (acima) e RNA (abaixo) ao longo do tempo para o experimento 8

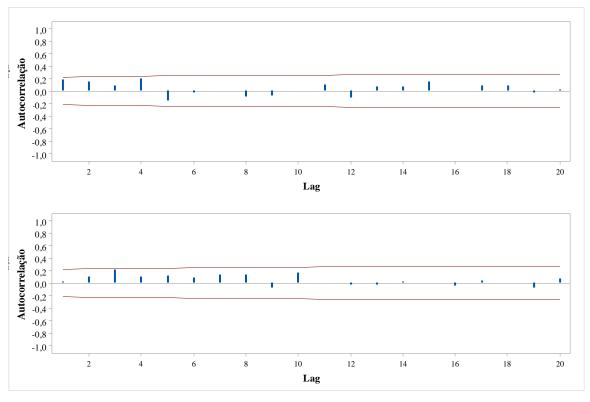


Figura B. 19. Função de autocorrelação para a série de resíduos gerados pelos métodos MSM (acima) e RNA (abaixo) na modelagem do experimento 8

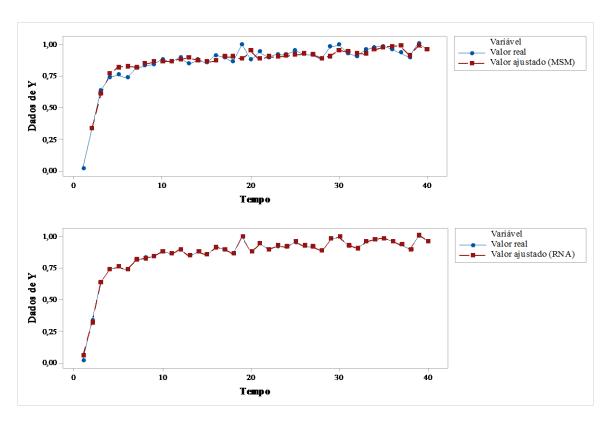


Figura B. 20. Gráfico dos valores reais e ajustados pelos métodos MSM (acima) e RNA (abaixo) ao longo do tempo para o experimento 9

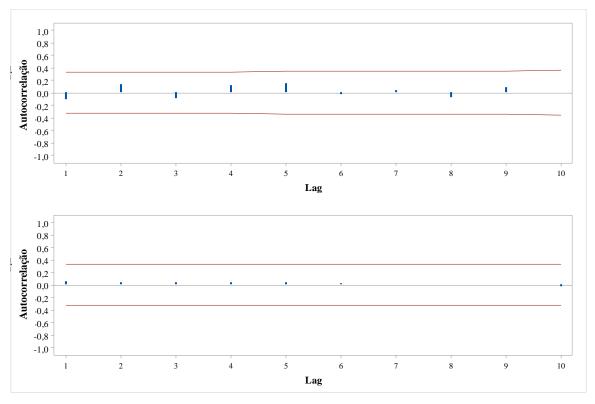


Figura B. 21. Função de autocorrelação para a série de resíduos gerados pelos métodos MSM (acima) e RNA (abaixo) na modelagem do experimento 9

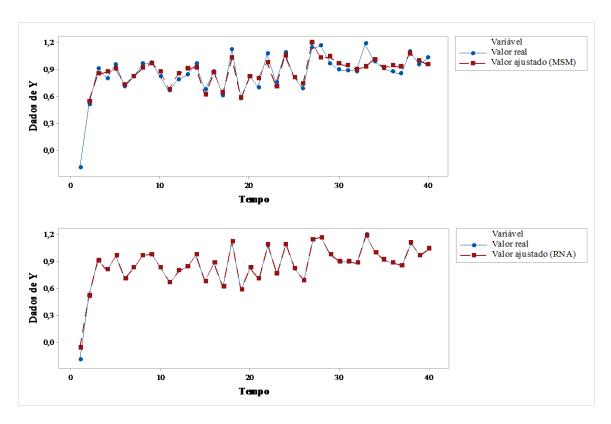


Figura B. 22. Gráfico dos valores reais e ajustados pelos métodos MSM (acima) e RNA (abaixo) ao longo do tempo para o experimento 10

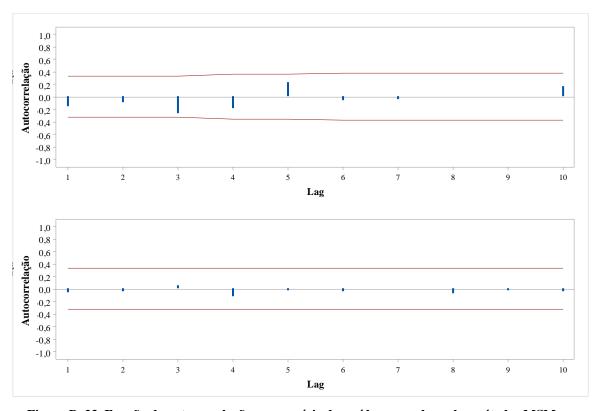


Figura B. 23. Função de autocorrelação para a série de resíduos gerados pelos métodos MSM (acima) e RNA (abaixo) na modelagem do experimento 10

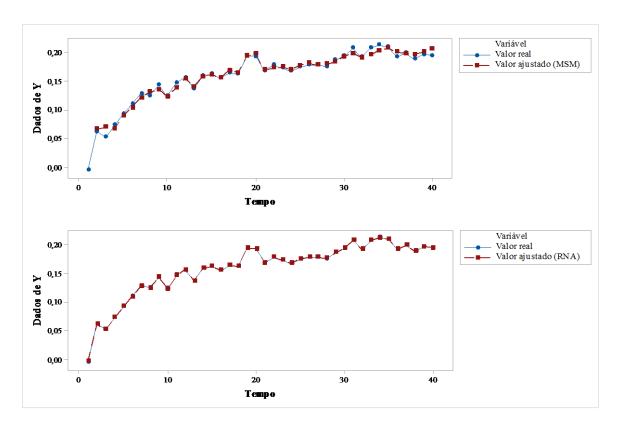


Figura B. 24. Gráfico dos valores reais e ajustados pelos métodos MSM (acima) e RNA (abaixo) ao longo do tempo para o experimento 11

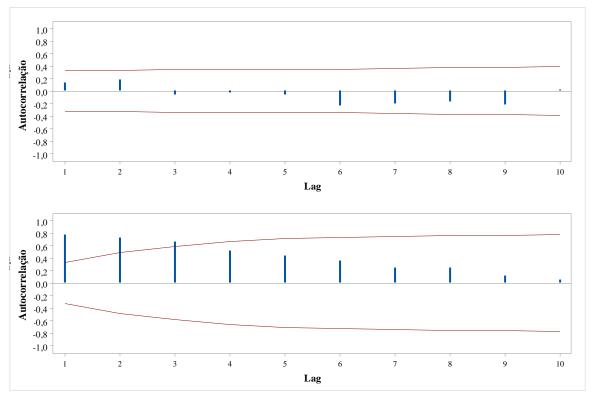


Figura B. 25. Função de autocorrelação para a série de resíduos gerados pelos métodos MSM (acima) e RNA (abaixo) na modelagem do experimento 11

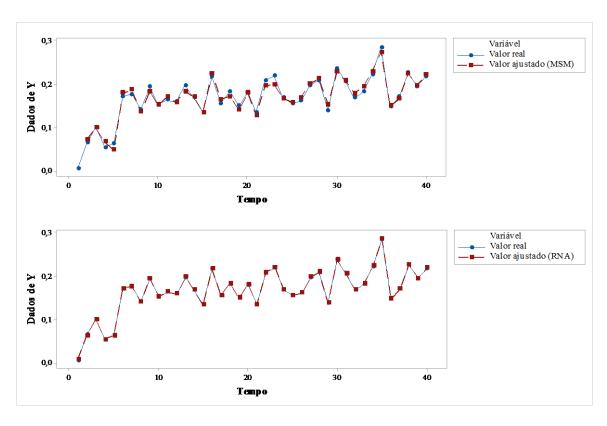


Figura B. 26. Gráfico dos valores reais e ajustados pelos métodos MSM (acima) e RNA (abaixo) ao longo do tempo para o experimento 12

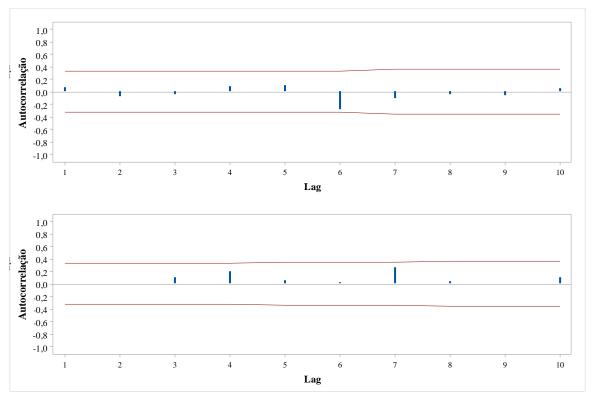


Figura B. 27. Função de autocorrelação para a série de resíduos gerados pelos métodos MSM (acima) e RNA (abaixo) na modelagem do experimento 12

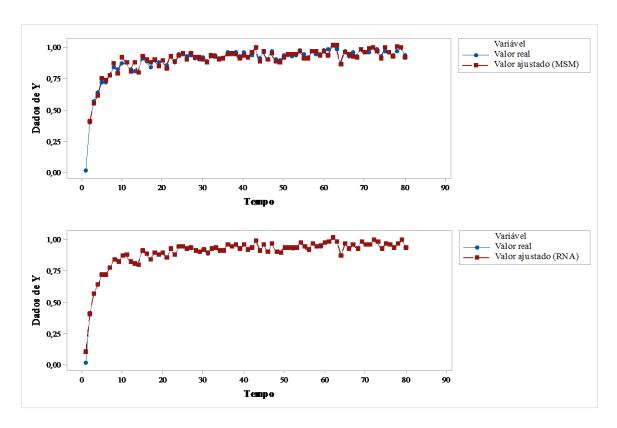


Figura B. 28. Gráfico dos valores reais e ajustados pelos métodos MSM (acima) e RNA (abaixo) ao longo do tempo para o experimento 13

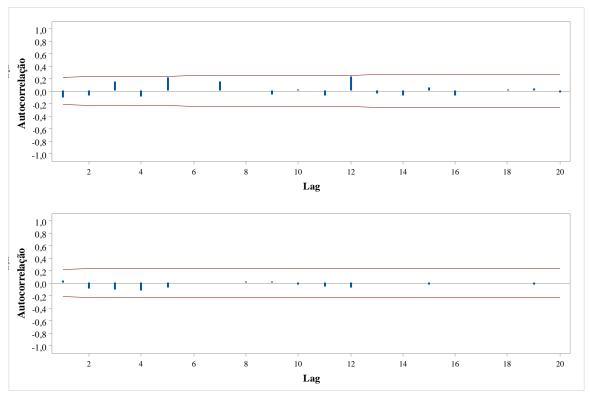


Figura B. 29. Função de autocorrelação para a série de resíduos gerados pelos métodos MSM (acima) e RNA (abaixo) na modelagem do experimento 13

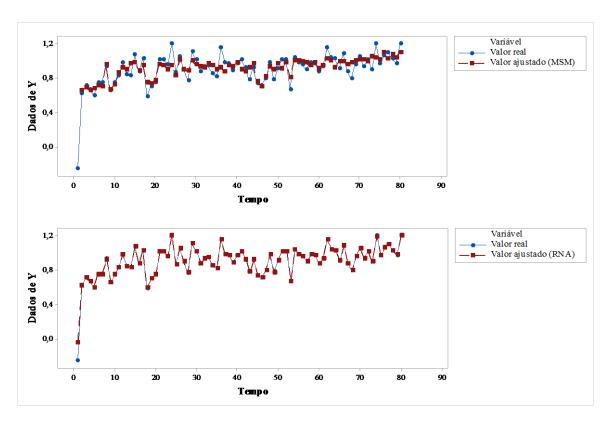


Figura B. 30. Gráfico dos valores reais e ajustados pelos métodos MSM (acima) e RNA (abaixo) ao longo do tempo para o experimento 14

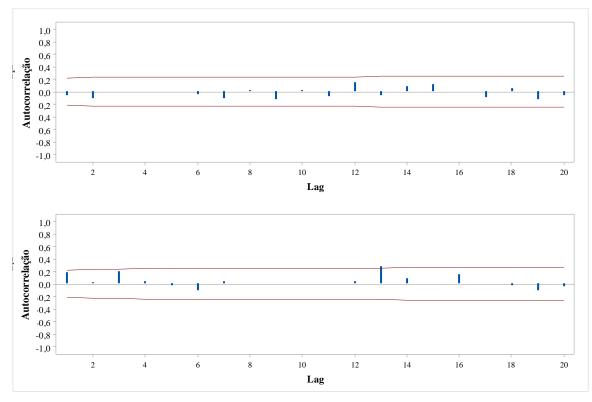


Figura B. 31. Função de autocorrelação para a série de resíduos gerados pelos métodos MSM (acima) e RNA (abaixo) na modelagem do experimento 14

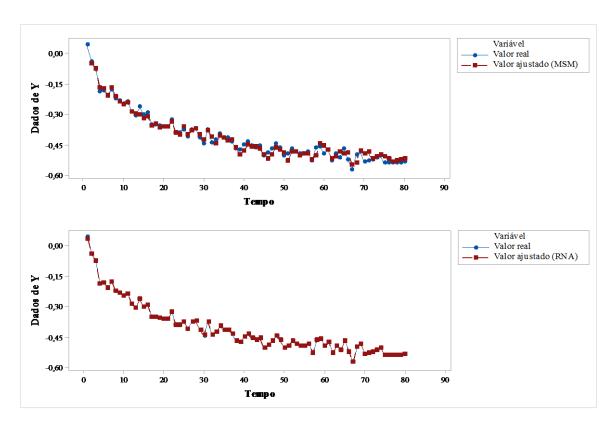


Figura B. 32. Gráfico dos valores reais e ajustados pelos métodos MSM (acima) e RNA (abaixo) ao longo do tempo para o experimento 15

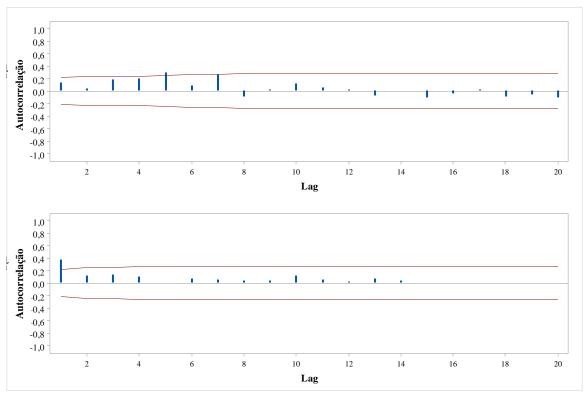


Figura B. 33. Função de autocorrelação para a série de resíduos gerados pelos métodos MSM (acima) e RNA (abaixo) na modelagem do experimento 15

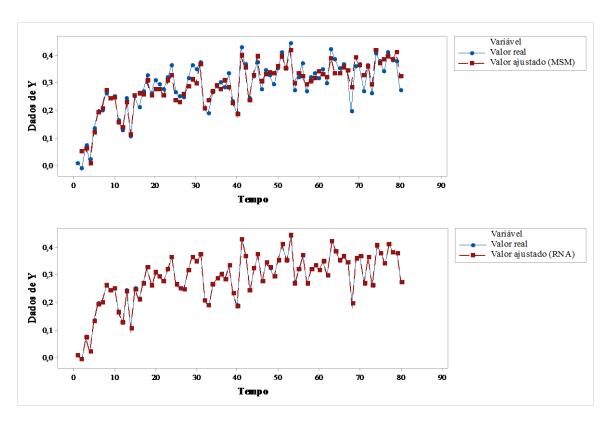


Figura B. 34. Gráfico dos valores reais e ajustados pelos métodos MSM (acima) e RNA (abaixo) ao longo do tempo para o experimento 16

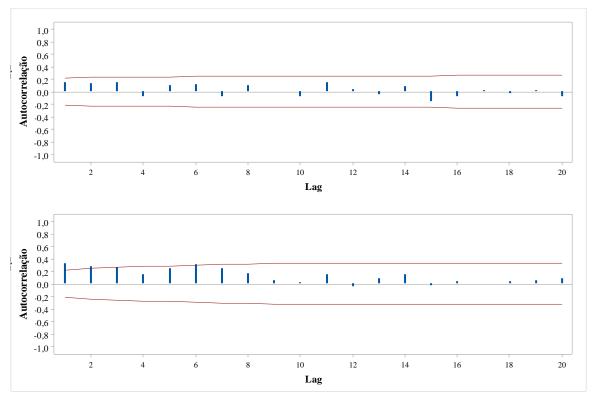


Figura B. 35. Função de autocorrelação para a série de resíduos gerados pelos métodos MSM (acima) e RNA (abaixo) na modelagem do experimento 16