

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

**COMBINAÇÃO DAS PREVISÕES DOS MODELOS DE
BOX-JENKINS E MLP/RNA PARA A PREVISÃO DE
DEMANDA NO PLANEJAMENTO DA PRODUÇÃO**

DISSERTAÇÃO DE MESTRADO

William Jacobs

Santa Maria, RS, Brasil

2014

COMBINAÇÃO DAS PREVISÕES DOS MODELOS DE BOX-JENKINS E MLP/RNA PARA A PREVISÃO DE DEMANDA NO PLANEJAMENTO DA PRODUÇÃO

William Jacobs

Dissertação apresentada ao Curso de Mestrado do Programa de Pós-Graduação em Engenharia de Produção, Área de Concentração Gerência de Produção, da Universidade Federal de Santa Maria (UFSM, RS), como requisito parcial para obtenção do grau de
Mestre em Engenharia de Produção

Orientador: Profa. Dra. Roselaine Ruviaro Zanini

Santa Maria, RS, Brasil

2014

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

Jacobs, William

Combinação das previsões dos modelos de Box-Jenkins e MLP/RNA para a previsão de demanda no planejamento da produção / William Jacobs.-2014.

101 p.; 30cm

Orientadora: Roselaine Ruviaro Zanini

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

1. Previsão de demanda 2. Combinação de previsões 3. Planejamento e controle da produção I. Ruviaro Zanini, Roselaine II. Título.

© 2014

Todos os direitos autorais reservados a William Jacobs. A reprodução de partes ou do todo deste trabalho só poderá ser feita mediante a citação da fonte.

Endereço: Rua 15 de Novembro, 1335, Teutônia, RS, 95.890-000

End. Eletr: eng.williamjacobs@gmail.com

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

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

**COMBINAÇÃO DAS PREVISÕES DOS MODELOS DE BOX-JENKINS
E MLP/RNA PARA A PREVISÃO DE DEMANDA NO
PLANEJAMENTO DA PRODUÇÃO**

elaborada por
William Jacobs

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

COMISSÃO EXAMINADORA:

Roselaine Ruviaro Zanini, Dra. (UFSM)
(Presidente/Orientador)

Adriano Mendonça Souza, Dr. (UFSM)

Liane Werner, Dra. (UFRGS)

Santa Maria, 26 de junho de 2014.

AGRADECIMENTOS

Primeiramente, agradeço à Deus, por ele me dar saúde e motivação para realizar um curso de mestrado, que é algo que eu sempre quis muito em minha vida.

Sou muito grato pela imensa ajuda que a minha orientadora dispensou ao longo destes dois anos de mestrado. Tive a sorte de ter tido como orientadora uma pessoa muito amiga, compreensiva e motivadora! Muito obrigado Prof. Roselaine!

Agradeço à minha família, especialmente aos meus pais, pois sempre investiram em minha educação e me incentivaram ao estudo.

Agradeço à minha noiva Adriani, que ao longo destes 13 anos de convívio, sempre esteve ao meu lado, me motivando a seguir em frente naquilo que realmente é importante para as nossas vidas.

Sou grato à CAPES, pela bolsa de estudos concedida para que eu pudesse realizar o mestrado.

Agradeço à todos os professores e funcionários do PPGEP/UFSM que fizeram parte desses dois anos de minha vida.

Aos professores integrantes da banca examinadora, pelas preciosas contribuições que tornaram este trabalho mais interessante.

RESUMO

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

COMBINAÇÃO DAS PREVISÕES DOS MODELOS DE BOX-JENKINS E MLP/RNA PARA A PREVISÃO DE DEMANDA NO PLANEJAMENTO DA PRODUÇÃO

AUTOR: WILLIAM JACOBS

ORIENTADORA: ROSELAINÉ RUVIARO ZANINI

Data e Local da Defesa da Dissertação: Santa Maria, 26 de junho de 2014.

A previsão da demanda futura dos produtos é a principal variável a ser considerada no planejamento e controle da produção nas organizações. As técnicas de previsão de demanda são fundamentais no planejamento da produção de nível tático e operacional, especialmente as séries temporais, pois não requerem do planejador, uma investigação mais aprofundada acerca dos fatores que influenciam a demanda. Dois métodos de previsão de séries temporais frequentemente utilizados na literatura são os modelos ARIMA e os modelos MLP/RNA. Uma prática que surgiu em 1969 e já consolidada para obter maior acurácia é a combinação das previsões individuais de dois ou mais modelos. Considerando a necessidade das organizações por técnicas preditivas que gerem melhores resultados, este estudo tem como objetivo prever os valores futuros de uma série temporal da demanda de leite UHT em uma indústria de lácteos, por meio da combinação dos modelos ARIMA e MLP/RNA, e comparar os resultados obtidos pelas combinações em relação aos modelos individuais, exemplificando a obtenção da previsão combinada no planejamento da produção. As medidas de acurácia para mensurar os resultados obtidos e selecionar o melhor modelo, foram o RMSE e o MAPE de previsão. Os resultados mostraram que a combinação dos modelos SARIMA(3,0,1)(1,1,0)₁₂ e DMLP pelo método *inverse mean square* forneceu um desempenho na previsão para 6 meses adiante, de até 66,5% superior em relação aos modelos individuais utilizados, onde a combinação das previsões obteve um RMSE de 1,43 e um MAPE de 2,16. Na previsão para 12 meses adiante, o desempenho da combinação foi de até 56,5% superior em relação aos modelos individuais, caso em que obteve um RMSE de 2,86 e um MAPE de 3,70%. A combinação de modelos de séries temporais possibilitou um aumento significativo no desempenho de previsão dos modelos, mas para que se obtenham resultados absolutos satisfatórios, devem-se utilizar modelos previsores que complementem mutuamente a capacidade preditiva.

Palavras-chave: Previsão de demanda. Combinação de previsões. Modelos ARIMA. Redes neurais artificiais. Planejamento e controle da produção.

ABSTRACT

Master Degree Thesis
Program of Post-Graduation on Production Engineering
Federal University of Santa Maria

COMBINATION OF BOX-JENKINS AND MLP/RNA MODELS FOR FORECASTING IN PRODUCTION PLANNING

AUTHOR: WILLIAM JACOBS

ADVISOR: ROSELAINÉ RUVIARO ZANINI

Date and Local of Defense of Thesis: Santa Maria, June 26th 2014.

A forecast of future demand for the products is the main variable to be considered in the planning and in production control in organizations. Two methods of time series forecasting often used in the literature are the ARIMA and MLP/RNA models. A practice that began in 1969 and has consolidated for greater accuracy is the combination of individual forecasts from two or more models. Considering the need for organizations by predictive techniques that generate better results, this study aims to predict the future values of a time series of the demand for UHT milk in a dairy industry, through the combination of ARIMA and MLP/RNA models, and to compare the results obtained by the combinations compared to individual models, exemplifying the achievement of combined forecasting in production planning. Accuracy measures to measure the results and to select the best model were the RMSE and MAPE for forecasting. The results showed that the combination of models SARIMA(3,0,1)(1,1,0)₁₂ and DMLP the inverse mean square method provided a performance forecast for the six months ahead, up to 66.5% higher than individual models used, where the combination of the predictions obtained a RMSE of 1.43, and a MAPE of 2.16. In the 12 month ahead prediction for the performance of the combination was up to 56.5% higher compared to individual models, in which case obtained a RMSE of 2.86 and 3.70% MAPE. The combination of time series models enabled a significant increase in performance prediction models, but in order to produce satisfactory absolute results should be used to complement their predictive abilities mutually.

Keywords: Forecasting. Forecasting combination. ARIMA models. Artificial neural networks. Production planning and controlling.

LISTA DE FIGURAS

Figura 1 – Planejamento hierárquico da produção	20
Figura 2 – Estratégias para o atendimento à demanda.....	21
Figura 3 – Lista de materiais do produto 1.....	23
Figura 4 – Decomposição de uma série temporal.....	26
Figura 5 – Correlograma da ACF de uma série temporal não estacionária simulada por meio do R.....	29
Figura 6 – Correlograma da ACF de um modelo AR(1) e correlograma da PACF de um modelo AR(1).....	29
Figura 7 – Correlograma da ACF de um modelo MA(1) e correlograma da PACF de um modelo MA(1).....	30
Figura 8 – Correlograma da ACF de um modelo ARMA(2,2) e correlograma da PACF de um modelo ARMA(2,2).....	30
Figura 9 – Correlograma da ACF e correlograma da PACF da série temporal de vazões de afluência da hidrelétrica Itaipu Binacional.....	35
Figura 10 – Correlograma da ACF e correlograma da PACF da série temporal de vazões de afluência integrada de ordem $s=12$	35
Figura 11 – Metodologia de Box-Jenkins para a modelagem e previsão de séries temporais .	36
Figura 12 – Modelo de um neurônio artificial.....	43
Figura 13 – Arquitetura de rede alimentada adiante de camada única	45
Figura 14 – Arquitetura de rede alimentada adiante com múltiplas camadas.....	45
Figura 15 – Arquitetura de rede recorrente e arquitetura de rede recorrente com camada oculta e laços de auto-realimentação.....	46
Figura 16 – A propagação do sinal (fase <i>feedforward</i>) e a retropropagação do erro em uma rede MLP (fase <i>backward</i>).....	48
Figura 17 – Obtenção da combinação das previsões individuais	53
Figura 18 – Fluxograma da metodologia utilizada na previsão e comparação dos modelos e técnicas de combinação.....	58
Figura 19 – Fluxograma da metodologia para a obtenção do modelo ARIMA para a previsão da série temporal.....	59
Figura 20 – Fluxograma da metodologia utilizada para a obtenção do modelo MLP e do modelo DMLP para a previsão de valores futuros da série temporal.....	60
Figura 21 – Metodologia para a combinação das previsões obtidas.....	63
Figura 22 – Série temporal da demanda agregada de leite do tipo UHT	69
Figura 23 – Correlograma da ACF e correlograma da PACF da série temporal estudada	70
Figura 24 – Correlograma da ACF e correlograma da PACF da série temporal estudada diferenciada uma vez na componente sazonal.....	71
Figura 25 – Resíduos padronizados do modelo SARIMA(3,0,1)(1,1,0) ₁₂ em função do período da observação	72
Figura 26 – Correlograma da ACF dos resíduos do modelo SARIMA(3,0,1)(1,1,0) ₁₂	73
Figura 27 – Representação do modelo MLP com 6 e 12 neurônios na camada de entrada.....	74
Figura 28 – Gráfico de dispersão dos resultados do processo de treinamento do modelo MLP e gráfico de dispersão dos resultados do processo de teste do modelo.....	75
Figura 29 – Representação do modelo DMLP com 6 e com 12 neurônios na camada de entrada.....	76
Figura 30 – Gráficos das componentes dessazonalizadas: tendência, sazonalidade e componente irregular da série temporal utilizadas como entrada nos modelos DMLP.....	77
Figura 31 – Gráfico de dispersão dos resultados do processo de treinamento do modelo DMLP e gráfico de dispersão dos resultados do processo de validação do modelo	78

Figura 32 – Gráfico dos valores previstos pela combinação AR_DMLP_IM para 6 meses adiante e dos valores previstos pela combinação para 12 meses adiante.....	82
Figura 33 – Gráfico da resposta média de cada grupo, para o RMSE de previsão de 6 e de 12 meses e da resposta média de cada grupo para o MAPE de previsão de 6 e 12 meses	87
Figura 34 – <i>Box-plot</i> do desempenho de cada grupo de combinação, conforme MAPE e RMSE de previsão de 6 meses e para a previsão de 12 meses.....	88
Figura 35 – Gráfico comparativo dos valores previstos pelos modelos do grupo 1 e da combinação AR_DMLP_IM para o horizonte de 6 meses adiante	90
Figura 36 – Gráfico comparativo dos valores previstos pelos modelos do grupo 1 e da combinação AR_DMLP_IM para o horizonte de 12 meses adiante	90

LISTA DE QUADROS

Quadro 1 – Exemplos de modelos pertencentes a cada uma das classes	24
Quadro 2 – Resumo para a aplicação do teste de raiz unitária ADF	38
Quadro 3 – Padrões teóricos do correlograma da ACF e da PACF a serem comparados com os correlogramas da série	39
Quadro 4 – Categorização dos métodos de combinação de previsões.....	52
Quadro 5 – Grupos de resultados de previsões para a comparação dos modelos e combinações e suas composições	66

LISTA DE TABELAS

Tabela 1 – Combinação dos erros de previsão de dois métodos na série temporal de milhas voadas por passageiro, em 1953	52
Tabela 2 – Teste ADF para a verificação de estacionariedade da série temporal da demanda agregada de leite tipo UHT	71
Tabela 3 – Modelos SARIMA selecionados para a estimação dos parâmetros	72
Tabela 4 – Estimação dos parâmetros dos modelos selecionados	73
Tabela 5 – Resultados dos modelos MLP/RNA utilizados para a modelagem e previsão da série temporal da demanda agregada de leite UHT	76
Tabela 6 – Resultados dos modelos DMLP utilizados para a modelagem e previsão da série temporal da demanda agregada de leite UHT	79
Tabela 7 – Estimação das medidas e definição do método para o cálculo da combinação da variância mínima.....	80
Tabela 8 – Estimação das medidas para a combinação das previsões por meio do método IM, exemplificação para a previsão de um período adiante	83
Tabela 9 – Resultados obtidos nas combinações das previsões dos modelos ARIMA, MLP e DMLP.....	84
Tabela 10 – Testes de Kruskal-Wallis para a comparação do RMSE e do MAPE dos grupos nos horizontes de previsão	85
Tabela 11 – Resultados do teste de Shapiro-Wilk para os conjuntos de valores em que se obtiveram o RMSE e o MAPE, nos respectivos horizontes de previsão	86
Tabela 12 – Resultados do teste de Kruskal-Wallis e do teste de comparação múltipla entre os grupos de modelos e de combinações das previsões	87
Tabela 13 – Desempenho das combinações em relação aos modelos SARIMA, MLP e DMLP, considerando o RMSE e o MAPE, para os respectivos horizontes de previsão	90
Tabela 14 – Pesos da combinação AR_DMLP_IM para os horizontes de previsão de 6 e de 12 meses adiante	92

LISTA DE ABREVIATURAS E SIGLAS

ACF	Autocorrelation Function / Função de Autocorrelação
ADF	Augmented Dickey-Fuller / Teste aumentado de Dickey-Fuller
AIC	Critério de Akaike
ARCH	Autoregressive Conditional Heterocedasticity
ARIMA	Autoregressive Integrated Moving Average / Autorregressivo Integrado de Médias Móveis
AR(p)	Autorregressivo de ordem p
ARMA(p,q)	Autoregressive Moving Average / Autorregressivo de ordem p e de Médias Móveis de ordem q
BIC	Bayesian Criterion / Critério Bayesiano ou de Schwarz
DF	Dickey-Fuller Test / Teste de Dickey-Fuller
DMLP	Modelo MLP dessazonalizado
IM	Método de combinação inverse-MSE
KW	Teste de Kruskal-Wallis
MA(q)	Moving Average / Médias Móveis de ordem q
MAPE	Mean Absolute Percentage Error / Erro Médio Percentual Absoluto
MLP	Multilayer Perceptron / Perceptron de Multicamadas
MD	Método de combinação da média aritmética
MG	Método de combinação da média geométrica
MGP	Método de combinação da média geométrica ponderada
MH	Método de combinação da média harmônica
MHP	Método de combinação da média harmônica ponderada
MQ	Método de combinação de mínimos quadrados (ou regressão)
MRP	Material requirement planning / Planejamento de recursos materiais
MSE	Mean Square Error / Erro Médio Quadrático
PACF	Partial Autocorrelation Function / Função de Autocorrelação Parcial
PAP	Planejamento agregado de produção
PCP	Planejamento e Controle da Produção
PMP	Planejamento mestre de produção
PNB	Produto Nacional Bruto
RMSE	Root mean square error
RNA	Redes Neurais Artificiais
SARIMA	Seasonal Autoregressive Integrated Moving Average / Sazonal Autorregressivo Integrado de Médias Móveis
SARIMABP	Técnica de combinação dos modelos SARIMA e MLP
STAR	Smooth transition autoregressive
SW	Teste de Shapiro-Wilk
TAR	Transition autoregressive
UHT	Ultra high temperature
VAR	Método de combinação da variância mínima

SUMÁRIO

1 INTRODUÇÃO	14
1.1 Justificativa	15
1.2 Objetivos.....	17
1.3 Delimitação do trabalho	17
1.4 Estrutura do trabalho	18
2 PREVISÃO DE DEMANDA NO PLANEJAMENTO E CONTROLE DA PRODUÇÃO.....	19
2.1 Previsão de demanda.....	19
2.2 Previsão de demanda no planejamento e controle da produção	20
2.2.1 Previsão de demanda por meio de métodos qualitativos	24
2.2.2 Previsão de demanda por meio de métodos quantitativos	24
3 PREVISÃO DE SÉRIES TEMPORAIS POR MEIO DOS MODELOS DE BOX-JENKINS, MLP/RNA E SUAS COMBINAÇÕES.....	28
3.1 Modelos ARIMA de Box-Jenkins	28
3.1.1 A importância da ACF e da PACF na metodologia de Box-Jenkins.....	28
3.1.2 Modelos ARIMA estacionários e não estacionários	31
3.1.3 Modelos ARIMA sazonal estacionário e não estacionário	34
3.1.4 Metodologia de Box-Jenkins para a modelagem e previsão de séries temporais.....	36
3.2 Modelos de Redes Neurais Artificiais.....	41
3.2.1 Definições iniciais acerca das Redes Neurais Artificiais.....	42
3.2.2 Arquiteturas de RNA	44
3.2.3 Aprendizagem e treinamento de uma RNA	46
3.2.4 Modelo <i>Multilayer Perceptron</i> para a previsão de séries temporais	49
3.3 Combinação de previsões individuais	51
3.3.1 Motivos para se combinar duas ou mais previsões.....	51
3.3.2 Técnicas de combinação das previsões.....	53
4 METODOLOGIA	58
4.1 Etapa 1 – Modelagem e previsão por meio da metodologia de Box-Jenkins.....	58
4.2 Etapa 2 – Modelagem e previsão por meio dos modelos MLP/RNA.....	60
4.3 Etapa 3 – Combinação das previsões dos modelos de Box-Jenkins e MLP/RNA	63
4.4 Etapa 4 – Comparação dos resultados de previsão	64
5 COMBINAÇÃO DAS PREVISÕES DOS MODELOS ARIMA E MLP/RNAs NO PLANEJAMENTO DA PRODUÇÃO.....	68
5.1 Modelagem, previsão, combinação das previsões e comparação dos resultados	68
5.1.1 Série temporal da demanda agregada de leite tipo UHT	68
5.1.2 Etapa 1 – Modelagem e previsão por meio da metodologia de Box-Jenkins	69
5.1.3 Etapa 2 – Modelagem e previsão por meio dos modelos de MLP/RNA.....	73
5.1.4 Etapa 3 – Combinação das previsões dos modelos de Box-Jenkins e MLP/RNA.....	79
5.1.5 Etapa 4 – Comparação dos resultados de previsão	83
5.2 Discussão dos resultados obtidos.....	90
5.2.1 Desempenho das combinações das previsões	90
5.2.2 Vantagens e dificuldades na utilização das combinações dos modelos.....	93
6 CONSIDERAÇÕES FINAIS.....	94
REFERÊNCIAS.....	96

1 INTRODUÇÃO

A globalização, o desenvolvimento tecnológico, o aumento do poder de compra e outros fatores, influenciam para que a competitividade entre as empresas aumente. Neste contexto, para manter-se no mercado, torna-se fundamental que as empresas busquem constantemente o aprimoramento dos seus métodos de planejamento e de gestão. Dentre estes métodos, a previsão de demanda é essencial, pois é o adequado atendimento à demanda que viabiliza a permanência no mercado e o crescimento da organização.

A previsão de demanda é um processo que envolve uma metodologia apoiada em métodos quantitativos e/ou qualitativos, para determinar e analisar os valores futuros da demanda por um ou mais produto(s) e/ou serviço(s), porque a demanda futura é o ponto de partida de todos os planejamentos dos níveis estratégico, tático e operacional (GAITHER; FRAZIER, 2002). As informações sobre a demanda são necessárias desde a concepção da empresa até o gerenciamento de sua rotina, pois é a principal entrada nos respectivos planejamentos envolvidos, que resultam nos seguintes planos: de negócio, estratégico, de marketing, de finanças, de produção, de distribuição, entre outros.

Na área de produção, conhecer a demanda futura dos produtos é fundamental para os três níveis de planejamento, que são: o estratégico, o agregado e o mestre. No primeiro nível, o conhecimento sobre a demanda é essencial para a tomada de decisão a respeito dos aspectos estruturais e infraestruturais necessários para o seu atendimento. No segundo, torna-se fundamental para o delineamento das políticas de atendimento à demanda e os respectivos recursos humanos e tecnológicos necessários no médio/longo prazo. No operacional, é necessário conhecer a demanda individual de cada produto para que se possa planejar os níveis de estoque necessários e o momento em que este será produzido (SLACK et al., 2009).

O desdobramento do nível tático para o operacional, no gerenciamento da produção, é realizado por meio da desagregação do plano agregado para o plano mestre de produção. No primeiro, as demandas são agregadas em famílias de produtos e no plano mestre, são obtidas as previsões individuais de cada produto (MESQUITA, 2008). A previsão de demanda em cada nível da função produção, também pode ser obtida pela agregação ou pela desagregação. No primeiro caso, as demandas são geradas individualmente para cada produto e em seguida, agregadas nos respectivos níveis superiores. No segundo caso, a demanda é obtida de forma agregada em níveis superiores e então desagregada até chegar à previsão individual de cada produto (SLACK et al., 2009). A previsão a partir dos produtos individuais possui a vantagem

de permitir maior conhecimento a respeito do comportamento da demanda individual, o que gera maior acurácia quando estas são agregadas. A desvantagem é que o processo torna-se trabalhoso na medida em que o número de produtos cresce.

A previsão de demanda para produtos individuais geralmente é obtida por meio de métodos de séries temporais, área na qual existem vários modelos que se baseiam em diversos pressupostos quanto ao comportamento da série de dados (FLORES, 2009), desde modelos simples até os que apresentam alta complexidade. Um modelo há muito tempo utilizado é o *Autoregressive Integrated Moving Average* (ARIMA), que é implementado por uma metodologia popularizada por George Box e Gwilym Jenkins, a metodologia de Box-Jenkins (MAKRIDAKIS et al., 1998). Outro modelo que tem mostrado acurácia é o de Redes Neurais Artificiais (RNA), onde diversas arquiteturas ou topologias já são aplicadas à previsão. A topologia de RNA comumente utilizada na previsão é a *Multilayer Perceptron* (MLP) e tem obtido sucesso, principalmente, na previsão de séries temporais com padrões não lineares, caso em que os modelos de Box-Jenkins não são eficazes (LEE; TONG, 2011).

Ressaltam Box et al. (2008) que grande parte dos processos empíricos, como é o caso da demanda, exibem, no todo ou pelo menos em parte, alguma forma de comportamento não linear. Para capturar melhor os padrões que influenciam na série temporal e otimizar a capacidade de previsão dos modelos, muitos pesquisadores da área argumentam que a melhoria da capacidade preditiva nas séries temporais, se dá pela combinação de modelos de previsão (TSENG et al., 2002; ZHANG, 2003; ABURTO; WEBER, 2007; FARUK, 2010; KHASHEI; BIJARI, 2010; CHOI et al., 2011; LEE; TONG, 2011; STOCK; WATSON, 2004; ANDRAWIS et al., 2011).

Considerando o contexto apresentado, este estudo propõe-se a descrever e exemplificar a utilização da combinação das previsões obtidas por meio dos modelos ARIMA e dos modelos MLP/RNA, na previsão de demanda de um produto. Para isso, foi realizada a modelagem e a previsão da série temporal da demanda agregada de leite produzido pelo processo *Ultra High Temperature*, ou leite tipo UHT, de uma indústria de lácteos da região do Vale do Taquari, interior do Rio Grande do Sul. Esta família de produtos foi escolhida em função de apresentar a maior demanda valorizada dentre as produzidas pela empresa.

1.1 Justificativa

A previsão de demanda no planejamento da produção, em geral, é realizada por métodos quantitativos, pois se deseja estimar a venda a ser realizada de um determinado

produto num horizonte de curto a médio prazo. Dentre os diversos modelos quantitativos existentes para a previsão de demanda, destacam-se os modelos de séries temporais, pois para a utilização desses, o pesquisador ou analista não precisa conhecer, *a priori*, os fatores que influenciam o processo gerador da série. Esta é uma vantagem fundamental no planejamento de produção de curto e médio prazo, tendo em vista o tempo necessário para se estudar os fatores de maior influência.

Uma forma de se melhorar as previsões de séries temporais é por meio da combinação das previsões de modelos, instituídas após o trabalho seminal de Bates e Granger (1969). Diversos estudos publicados, como Makridakis e Hibon (2000), Stock e Watson (2004), Andrawis et al. (2011), Martins e Werner (2012), entre outros, demonstraram que a combinação das previsões obtém um desempenho superior em comparação à utilização individual de cada modelo. A combinação pode ser realizada por diferentes métodos dispostos na literatura, nas diversas áreas de aplicação, mas é importante que se utilize de bons modelos previsores, com capacidade de capturar os padrões existentes na série de dados (ANDRAWIS et al., 2011).

Neste estudo, são utilizados os modelos de Box-Jenkins (ARIMA) e os modelos MLP/RNA como previsores utilizados nas combinações. Os modelos ARIMA, por mais de meio século têm sido utilizado e obtido sucesso em diversas áreas para a realização de previsões por meio de séries temporais (HO et al., 2002). A desvantagem destes, é que apresenta acurácia para a previsão de séries temporais com padrões predominantemente lineares (LEE; TONG, 2011; ZHANG, 2003). Porém, conforme Box et al. (2008), grande parte das séries temporais são não lineares ou no todo ou em parte e, frequentemente, torna-se difícil obter acurácia por meio de somente um modelo. Por esses motivos, foram utilizados os modelos MLP/RNA combinado com as previsões obtidas pelo modelo ARIMA, pois os modelos MLP/RNA têm mostrado capacidade de capturar padrões não lineares presentes na série (LEE; TONG, 2011).

Para que se otimize a previsão de demanda em empresas com um grande número de produtos, uma alternativa proposta por Pellegrini e Fogliatto (2001), é de se utilizar uma metodologia que considere o impacto do produto no faturamento e, então, alocar mais recursos para a previsão e acompanhamento da demanda dos produtos classificados como de maior impacto. Assim, a maior preocupação na obtenção de acurácia na previsão deve ser nos produtos que mais impactam nas receitas da empresa, que geralmente são poucos, onde se recomenda utilizar métodos de previsão que obtenham melhor desempenho. Para os demais

produtos, pode-se utilizar métodos simples e rápidos para se obter a previsão, pois o impacto destes nos resultados não é significativo.

O estudo e obtenção de modelos de previsão de demanda, neste caso por meio de séries temporais, que resultem em maior capacidade de previsão, influenciam de forma direta nos resultados da empresa. Com a obtenção de maior capacidade de previsão de demanda dos produtos, torna-se possível para a organização obter um melhor balanceamento no fluxo de insumos dos fornecedores, no nível de estoques ao longo do processo e na entrega do produto aos clientes. Desta forma é que o presente estudo é justificado, pois este pode ser utilizado como um modelo para a combinação da previsão de demanda no planejamento da produção e na obtenção de melhores resultados neste processo.

1.2 Objetivos

Este estudo tem por objetivo prever os valores futuros da demanda de um produto por meio da combinação dos modelos ARIMA e MLP/RNA, comparar e analisar os resultados das previsões combinadas em relação aos modelos individuais, identificando o melhor modelo na previsão de demanda para o planejamento agregado da produção.

1.3 Delimitação do trabalho

O estudo visa obter a previsão de valores futuros para a série temporal da demanda do produto. O foco não está no processo de modelagem. Assim, a análise dos resultados será realizada com base nos valores das previsões das combinações dos modelos e dos modelos na forma individual.

Outra delimitação do trabalho se relaciona aos modelos individuais utilizados na previsão, que foram dois: o ARIMA e o MLP/RNA. A verificação da acurácia de previsão de cada modelo, também foi delimitada a duas medidas de acuracidade: o RMSE e o MAPE.

Os modelos MLP/RNA, por serem inicializados com valores aleatórios entre os pesos de suas conexões, podem gerar valores diferentes dos obtidos, em caso de replicação dos modelos de previsão de demanda utilizados nesta dissertação.

1.4 Estrutura do trabalho

Para atender aos propósitos estabelecidos, esta dissertação foi estruturada em seis capítulos, que são discriminados a seguir.

O primeiro capítulo aborda uma introdução ao tema estudado, apresentando o problema, as justificativas para a sua realização, bem como os objetivos e a estrutura do trabalho.

O capítulo 2 é constituído por uma revisão bibliográfica sobre previsão de demanda e sobre o planejamento da produção. Os assuntos tratados neste capítulo são: previsão de demanda no contexto do planejamento e controle da produção, os níveis de planejamentos na função produção e a importância da previsão nestes.

O capítulo 3 apresenta os modelos e as técnicas de combinação para a previsão da série temporal. Primeiramente, são apresentados os modelos da metodologia de Box-Jenkins, onde se descreve o processo para a obtenção do melhor modelo e posterior previsão dos valores futuros. Após, são apresentados os modelos MLP/RNA para a previsão de séries temporais. Por fim, descrevem-se as técnicas de combinação das previsões a serem utilizadas no estudo.

O capítulo 4 apresenta a metodologia utilizada para a modelagem, previsão e comparação dos resultados das combinações e dos modelos individuais.

O capítulo 5 apresenta um estudo de caso utilizando a combinação dos modelos. O capítulo segue a apresentação dos modelos dada no capítulo 3, onde primeiramente são apresentados os resultados obtidos pelo modelo ARIMA, seguido dos resultados do modelo MLP/RNA e então da combinação dos modelos. Em seguida é apresentado um estudo comparativo entre os resultados de previsão obtidos, e por último, uma discussão destes resultados.

O capítulo 6 é constituído pelas considerações finais, onde são apresentadas as conclusões deste estudo e sugestões para trabalhos futuros.

2 PREVISÃO DE DEMANDA NO PLANEJAMENTO E CONTROLE DA PRODUÇÃO

Este capítulo tem como objetivo apresentar a maneira como a previsão de demanda se insere no Planejamento e Controle da Produção (PCP) e como a obtenção de resultados mais acurados neste processo influenciam o resultado global da área.

2.1 Previsão de demanda

A previsão, no contexto organizacional, é uma análise de eventos futuros utilizada para estabelecer objetivos e planejar como atingi-los. As previsões são necessárias por diversos motivos: (i) para a análise e tomada de decisão frente às ações dos concorrentes, (ii) para o gerenciamento do ciclo de vida de produtos, (iii) na definição de políticas de atendimento à demanda, (iv) no planejamento da capacidade de produção, (v) no planejamento de recursos humanos e materiais, (vi) na programação dos recursos, entre outros (RUSSOMANO, 2000).

No contexto da administração da produção, a importância das previsões é devido, principalmente, a dois fatores, que são: (i) a quantidade de recursos físicos necessários e (ii) a inércia inerente ao processo decisório (CORRÊA; CORREA, 2004). Para a produção de bens e serviços, são necessários: matéria-prima, máquinas, equipamentos, pessoas e demais insumos; e, esses recursos precisam estar disponíveis em momento específico para que a produção possa ocorrer (RIGGS et al., 1981). Para se otimizar a tomada de decisão, considerando o tempo para que as decisões tomadas tenham efeito, é necessário periodicamente realizar e revisar as previsões da demanda futura (FERNANDES; GODINHO FILHO, 2010).

Conforme Kotler e Keller (2012), demanda é o desejo que um consumidor possui por um produto específico, considerando que este tenha poder de compra. A demanda pode ser influenciada por uma série de fatores, desde as condições macroeconômicas até as questões operacionais. Alguns exemplos destes fatores são os seguintes: inflação, crescimento da economia, disponibilidade de crédito, preço da concorrência, promoção, capacidade de produção, capacidade de entrega, entre outros (MESQUITA, 2008).

A definição ampla de demanda dada por Kotler e Keller (2012) evidencia o grau de complexidade para a realização de previsão. Em termos práticos, seria muito oneroso para as

empresas conhecer todas as variáveis que a influenciam, medi-las, controlá-las e analisá-las, com o propósito de realizar previsão, principalmente para horizontes de planejamento de curto e médio prazo. Estes horizontes necessitam de informações menos estruturadas e de constante revisão dos valores previstos (FERNANDES; GODINHO FILHO, 2010).

Este é o caso do PCP, onde as decisões a serem tomadas tem impacto especialmente no curto e no médio prazo. Nesta área, são utilizadas diversas técnicas e modelos de previsão, onde alguns são descritos na seção 2.2.

2.2 Previsão de demanda no planejamento e controle da produção

O Planejamento e Controle da Produção é a área da administração da produção que tem como objetivo gerenciar com eficiência o fluxo de materiais, a utilização de pessoas e equipamentos e responder às necessidades dos clientes utilizando a capacidade dos fornecedores e da estrutura interna (VOLLMANN et al., 2006). A tomada de decisão ocorre no nível tático e operacional, e objetiva responder às seguintes questões: (i) o que produzir e comprar, (ii) quanto produzir e comprar, (iii) quando produzir e comprar e, (iv) com que recursos produzir (CORRÊA et al., 2011).

O planejamento da produção é realizado a partir da demanda e hierarquizado conforme os horizontes de tempo, classificados como: longo, médio e curto prazo. Os períodos inclusos em cada horizonte de planejamento variam de uma empresa para outra. Geralmente o curto prazo é de uma semana a um mês, o médio prazo de um mês a um ano e o de longo prazo de um a três anos (VOLLMANN et al., 2006). Para Lustosa e Nanci (2008), os níveis de planejamento no PCP são aqueles mostrados na Figura 1.

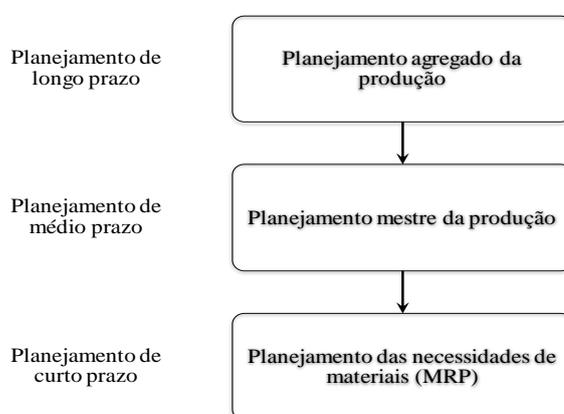


Figura 1 – Planejamento hierárquico da produção. Fonte: adaptado de Lustosa e Nanci (2008)

O planejamento agregado de produção conecta as metas estratégicas da empresa com a produção. A sua principal entrada é a previsão da demanda agregada, que obtém-se pela soma das demandas individuais dos produtos constantes na agregação. A elaboração é coordenada com a demanda prevista, com as disponibilidades de recursos e com o orçamento financeiro. O resultado do planejamento agregado é o plano agregado de produção (PAP), que permite à empresa um melhor gerenciamento de sua capacidade, de forma a obter um custo mínimo. O horizonte de planejamento do PAP é na ordem de seis meses a dois anos e planejado em períodos espaçados mensalmente (VOLLMANN et al., 2006).

As principais políticas estratégicas consideradas na elaboração do PAP são: de atendimento à demanda, de mão de obra e de estoques (FERNANDES; GODINHO FILHO, 2010). A política a ser considerada pela empresa no atendimento à sua demanda influencia diretamente na escolha das outras duas. Esta é definida a partir da estratégia de atendimento à demanda, que pode ser: (i) estratégia nivelada, (ii) estratégia seguidora e (iii) estratégia mista, conforme pode-se visualizar na Figura 2 (LUSTOSA; NANJI, 2008).

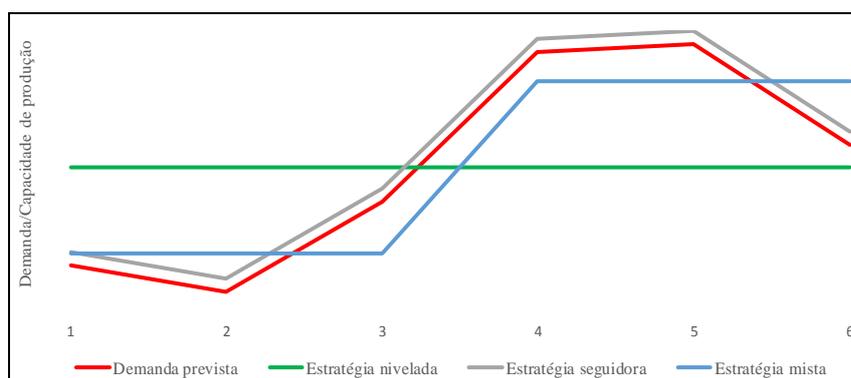


Figura 2 – Estratégias para o atendimento à demanda. Fonte: adaptado de Lustosa e Nanci (2008)

Na estratégia nivelada, a empresa busca manter constante a utilização de sua capacidade ao longo do tempo (RUSSOMANO, 2000). A capacidade de produção é definida a partir da demanda média a ser atendida em determinado horizonte. A adoção desta estratégia impacta diretamente nas políticas de mão de obra e de estoques, pois o objetivo é atender a demanda de forma a minimizar a flutuação destas variáveis (RIGGS et al., 1981). Os estoques são formados em períodos anteriores à demanda prevista, de forma a aproveitar a ociosidade dos recursos. Esta estratégia necessariamente é adotada em organizações onde a mão de obra é especializada, pois dessa forma são evitados os custos de demissão e contratação (LUSTOSA; NANJI, 2008).

Na estratégia seguidora, a empresa produz conforme a previsão de demanda para o horizonte de planejamento (JACOBS et al., 2009). A busca é por minimizar os custos de estoques, de forma a mantê-lo em um nível mínimo, pois são estes que mais impactam no resultado da área operacional. Esta estratégia é operacionalizada por meio da variação na capacidade de mão de obra e/ou da variação na capacidade das máquinas (RIGGS et al., 1981).

A forma mista engloba as duas estratégias descritas e tem como objetivo suavizar a variação na capacidade para o atendimento à demanda, tendo em vista que a variação de capacidade no curto prazo pode apresentar custos muito elevados devidos, principalmente, à demissão ou contratação de mão de obra (JACOBS et al., 2009).

Segundo Ritzman e Krajewski (2004), a diferença entre o PAP e o plano mestre de produção (PMP) está na agregação da demanda, no horizonte de tempo do planejamento e nos objetivos. O planejamento mestre da produção desdobra a previsão de demanda do PAP, realizada em períodos mensais para as famílias de produtos, em previsões individuais dos produtos e em períodos semanais. O horizonte de planejamento geralmente adotado no PMP varia de 4 a 8 semanas. O objetivo é utilizar eficientemente a capacidade planejada no PAP, projetar os estoques e o momento da disponibilidade do produto (CORRÊA et al., 2001).

Para a elaboração do PMP são necessárias as seguintes entradas: a previsão de demanda individual, os pedidos firmes de clientes e o estoque inicial do período anterior (LUSTOSA; NANJI, 2008). Esses dados geram as saídas do PMP, que são: o estoque projetado, os recebimentos do PMP e a quantidade de produtos disponíveis para prometer.

A utilização do PMP gera maior integração entre as áreas de produção e vendas nos horizontes de curto e médio prazo (CORRÊA et al., 2001). A área de vendas pode consultar instantaneamente a quantidade disponível de cada produto que pode ser comercializada. Essa é a principal informação de interesse desta área, pois as quantidades disponíveis para promessa são colocadas como metas a serem atingidas pelas equipes, otimizando os resultados da empresa (RIGGS et al., 1981).

O planejamento de recursos materiais é aquele utilizado para gerenciar o curto prazo, de nível operacional (FERNANDES; GODINHO FILHO, 2010). Este planejamento surgiu na década de 1960 e até então é conhecido como MRP, que é um acrônimo de *Materials Requirement Planning*. O MRP é obtido a partir das seguintes entradas: do PMP, da lista de materiais do produto, dos prazos de entrega e dos estoques disponíveis (RIGGS et al., 1981). A partir destas entradas, são realizados os procedimentos de cálculo do MRP e então obtidas as saídas, que são: as ordens de produção e as ordens de compra, que indicam quando e

quanto será fabricado e/ou comprado em cada período, respectivamente (BURBIDGE; CRUZ, 1981).

A lista de materiais apresenta a estrutura de composição do produto. Os itens que compõem o produto são agrupados em componentes e subcomponentes. Esses itens são organizados em diferentes níveis, do nível 0 até o nível n , onde n é o número de níveis do produto. O principal fundamento na utilização desta técnica é a desagregação da demanda do produto para seus componentes e subcomponentes. Ao nível 0 está associado uma demanda independente e aos demais níveis, demandas dependentes (CORRÊA et al., 2011). Para exemplificação, na Figura 3 se apresenta uma lista de materiais do produto 1, composta de 2 níveis.

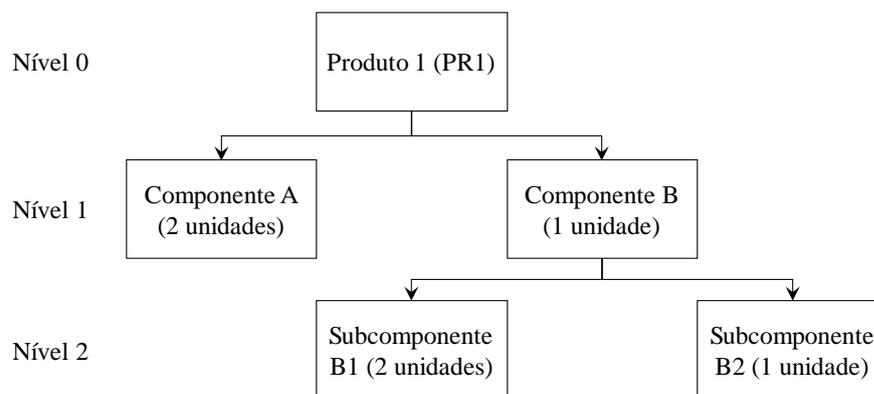


Figura 3 – Lista de materiais do produto 1. Fonte: elaborado a partir de Lustosa e Nanci (2008)

Na lógica do MRP, a demanda independente está associada aos produtos, que é obtida a partir do PMP, realizado para o horizonte de médio prazo. A demanda dependente é aquela gerada para cada componente ou subcomponente, dado a demanda independente do produto e a quantidade de cada componente para cada unidade demandada (RUSSOMANO, 2000). Na Figura 3, por exemplo, cada unidade demandada do produto 1, gera uma demanda dependente do componente A em duas unidades.

Existem diversas técnicas disponíveis para se realizar previsão de demanda dos itens independentes, que são divididas em duas classes: (i) métodos qualitativos e (ii) métodos quantitativos. A escolha da técnica mais apropriada a ser utilizada no PCP, dependerá da análise de alguns fatores, que são principalmente os seguintes: (i) o objetivo para o qual se utilizará a previsão, (ii) o horizonte de tempo a ser considerado, (iii) a precisão desejada e (iv) os dados disponíveis para se realizar a previsão. A partir dessa análise, é possível identificar o(s) modelo(s) mais apropriado(s) ao(s) objetivo(s) do estudo (DAVIS et al., 2001).

2.2.1 Previsão de demanda por meio de métodos qualitativos

Os métodos qualitativos possuem uma natureza subjetiva e são utilizados por meio de uma abordagem estruturada. Utilizam-se estes na formulação de estratégias, desenvolvimento de novos produtos e tecnologias com aceitação ainda incerta pelo mercado e para o planejamento estratégico de longo prazo das organizações (MAKRIDAKIS et al., 1998). Segundo Sanders e Manrodt (1994), tais modelos incorporam uma grande quantidade de conhecimento empírico e adicionam consenso ao processo de previsão. Alguns exemplos destes métodos são: consenso do comitê executivo, método Delphi, pesquisa da equipe de vendas, pesquisa de clientes, analogia histórica e pesquisa de mercado (SLACK et al., 2009).

2.2.2 Previsão de demanda por meio de métodos quantitativos

As técnicas de previsão quantitativa utilizam modelos matemáticos para estabelecer os valores previstos. Estas técnicas buscam descrever o comportamento da demanda por meio do ajuste de um modelo que gere o menor erro possível e são divididas em dois tipos: (i) nos modelos causais ou de correlação e (ii) nos modelos de séries temporais (TUBINO, 2000; MOREIRA, 2008; MESQUITA, 2008). Existem diversos modelos para cada classe descrita, alguns exemplos são apresentados no Quadro 1.

Modelos causais ou de correlação	Modelos de séries temporais
Regressão linear simples	Médias móveis
Regressão exponencial simples	Suavização exponencial
Regressão quadrática simples	Decomposição
Regressão linear múltipla	Box-Jenkins
Regressão quadrática múltipla	Redes neurais artificiais

Quadro 1 – Exemplos de modelos pertencentes a cada uma das classes. Fonte: elaborado pelo autor

Os modelos causais ou de correlação, são utilizados quando se deseja prever a demanda a partir de uma ou mais variáveis explanatórias ou causais, que podem ser internas ou externas à empresa. Alguns exemplos de variáveis causais são os seguintes: a variação populacional, o Produto Nacional Bruto (PNB), o número de alvarás expedidos para a construção, o consumo de produtos correlacionados à demanda do produto que se analisa, entre outras. O que determina a escolha de uma particular variável, é a existência de

correlação significativa entre as variáveis independente e dependente (MAKRIDAKIS et al., 1998).

Nos modelos de séries temporais, as observações passadas da mesma variável são coletadas e analisadas para a obtenção de um modelo que descreva a relação existente entre a variável e o tempo (ZOU et al., 2007). Estes modelos baseiam-se na hipótese de que os fatores que influenciaram a ocorrência da demanda ao longo do tempo, continuará acontecendo no futuro (RITZMAN; KRAJEWSKI, 2004). A área de séries temporais tem sido extensivamente estudada como ferramenta de apoio em muitos sistemas de suporte à decisão, principalmente na previsão de demanda (CHOI et al., 2011).

Estes modelos são utilizados quando não existem variáveis explicativas para aquela que se deseja realizar a previsão ou quando a obtenção de um modelo de regressão resulta em erros-padrão tão elevados, que tornam a maioria dos coeficientes estimados pouco significativos (PINDYCK; RUBINFELD, 2004). Na previsão de demanda para o PCP, a principal vantagem na utilização de séries temporais é devido à redução no tempo necessário para a obtenção de um modelo, tendo em vista que não há maior necessidade de investigação a respeito dos fatores que influenciam a demanda (TUBINO, 2000).

Na literatura da área, existem diversos modelos de previsão por meio de séries temporais (MORETTIN; TOLOI, 2006). Alguns dos modelos que com frequência são observados, são os seguintes: médias móveis, suavização exponencial, decomposição, Box-Jenkins e redes neurais artificiais. Os três primeiros modelos são apresentados a seguir nesta seção e, os dois últimos, são apresentados em maiores detalhes no Capítulo 3, visto que são alvo desta pesquisa.

O modelo de previsão de séries temporais mais simples é o de médias móveis, que é útil para alguns tipos de série, especialmente as que apresentam características estacionárias e não sazonais (PINDYCK; RUBINFELD, 2004). A previsão por meio deste modelo é obtida calculando-se a média aritmética dos n valores reais da demanda imediatamente passados. A escolha do tamanho de n geralmente varia entre 3 a 5 períodos (MORETTIN; TOLOI, 2006).

Os modelos de suavização exponencial são assim denominados por aplicarem um conjunto de pesos desiguais aos valores passados da série, que decaem exponencialmente da mais recente a mais distante observação (MAKRIDAKIS et al., 1998). Os modelos de suavização exponencial mais conhecidos são os de Holt e de Holt-Winters, assim denominados devido ao nome dos dois pesquisadores que o popularizaram.

O modelo de Holt é utilizado em séries que apresentam tendência de crescimento linear. O modelo é descrito por três equações e dois parâmetros de suavização, α e β . Estes

parâmetros podem variar no intervalo de 0 a 1 e são obtidos por meio da minimização da soma dos erros quadrados (MORETTIN; TOLOI, 2006).

Os modelos de Holt-Winters são utilizados em séries temporais que apresentam as componentes tendência linear e sazonalidade e dividem-se em dois outros tipos: (i) no de Holt-Winters aditivo e (ii) no de Holt-Winters multiplicativo. O modelo aditivo é adequado para a previsão de séries temporais em que a amplitude da variação sazonal permaneça constante ao longo do tempo. O multiplicativo, do contrário, em séries que a amplitude da variação sazonal não permaneça constante ao longo do tempo. Os modelos de Holt-Winters são função de três parâmetros de suavização, α , β e γ . Tais parâmetros são obtidos pelo mesmo processo descrito nos modelos de Holt (MORETTIN; TOLOI, 2006).

A decomposição é um método que visa promover melhor entendimento a respeito do comportamento de séries temporais sazonais. Trata-se de um dos mais antigos métodos de análise de séries temporais. A utilização desta técnica assume que as séries temporais possam apresentar as seguintes componentes: (i) tendência, (ii) sazonalidade e (iii) aleatória ou irregular; conforme podem ser observadas na Figura 4. As principais vantagens advindas na decomposição, é a separação da componente irregular ou aleatória e a utilização das outras componentes para realizar a previsão de valores futuros (MAKRIDAKIS et al., 1998).

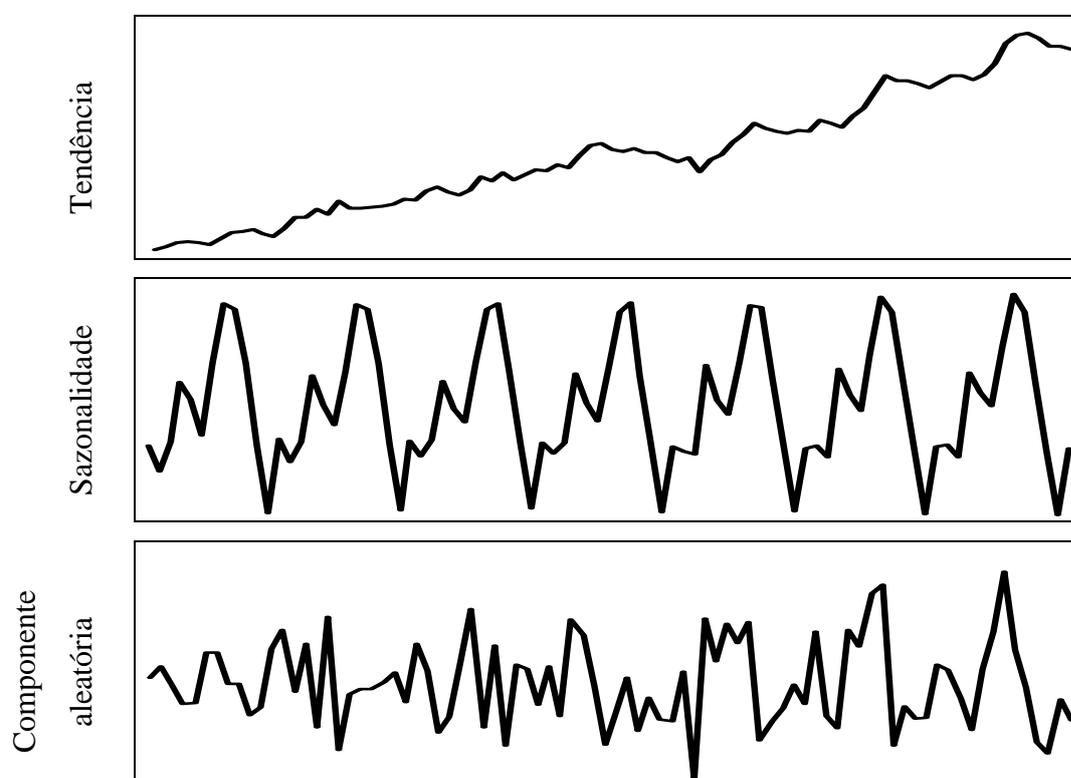


Figura 4 – Decomposição de uma série temporal. Fonte: adaptado a partir de Makridakis et al. (1998)

Os modelos de decomposição podem ser do tipo aditivo ou do tipo multiplicativo. No modelo aditivo, como é sugerido, a série é composta pela soma de seus componentes. E no multiplicativo, pela multiplicação de seus componentes. Nas equações 1 e 2 são dadas as respectivas formas de composição das séries temporais (MORETTIN; TOLOI, 2006).

$$y_t = T_t + S_t + C_t \quad (1)$$

$$y_t = T_t \cdot S_t \cdot C_t \quad (2)$$

onde y_t é o valor da série no tempo t , T_t é o valor da tendência no tempo t , S_t é o valor da sazonalidade no tempo t e C_t é o valor da componente irregular no tempo t .

3 PREVISÃO DE SÉRIES TEMPORAIS POR MEIO DOS MODELOS DE BOX-JENKINS, MLP/RNA E SUAS COMBINAÇÕES

Este capítulo tem como objetivo apresentar os modelos e as técnicas de combinação para a previsão da série temporal considerada nesta dissertação. Os modelos de previsão utilizados foram o ARIMA e o MLP/RNA, e são apresentados nas seções 3.1 e 3.2, respectivamente. A combinação das previsões dos modelos foram realizadas por meio dos oito métodos de combinação apresentados na seção 3.3.

3.1 Modelos ARIMA de Box-Jenkins

Os modelos ARIMA, utilizados por meio da metodologia de Box-Jenkins, vêm sendo utilizados por pesquisadores e profissionais desde a década de 70 na área de previsão de séries temporais (MADDALA, 2003; ZHANG, 2003). Esta metodologia já foi aplicada de forma ampla em diversos problemas, onde se podem citar alguns: aplicações na área médica, ambiental, financeira, de engenharia, na modelagem de séries temporais para a previsão da qualidade do ar, na modelagem da demanda de turistas que chegam à Austrália proveniente de três países da Ásia (WERNER; RIBEIRO, 2003). Além destes, na previsão de falhas em sistemas mecânicos (HO; XIE, 1998), na previsão da taxa de inflação futura na Finlândia (JUNTILA, 2001), na previsão da demanda de máquinas e de bebidas em Taiwan (TSENG; TZENG, 2002), na previsão da demanda de eletricidade na Turquia (ERDOGDU, 2007), entre outras diversas aplicações.

3.1.1 A importância da ACF e da PACF na metodologia de Box-Jenkins

Para a aplicação da metodologia de Box e Jenkins (1976), duas importantes funções a serem consideradas são a função de autocorrelação (ACF) e a função de autocorrelação parcial (PACF), devido a isso, optou-se por abordar esse assunto primeiramente, tendo em vista que os subsequentes frequentemente farão menção a tais medidas (MAKRIDAKIS et al., 1998). A ACF é uma ferramenta para a investigação das propriedades da série, como: identificação da condição de estacionariedade e identificação de potenciais modelos (GUJARATI, 2000; MORETTIN, 2008). Conforme Makridakis et al. (1998, p. 342), a

“PACF é utilizada conjuntamente com a ACF para a identificação dos modelos a serem considerados na modelagem e previsão da série”.

A primeira utilidade citada para a ACF é a análise de estacionariedade. As séries temporais não estacionárias apresentam forte correlação serial, que pode ser visualizada por meio do correlograma da ACF, sendo o declínio da ACF lento na medida em que aumenta o número de *lags* (FAVA, 2000). Na Figura 5 é apresentado o correlograma da ACF de uma série temporal não estacionária simulada por meio do R (R CORE TEAM, 2014), onde se pode observar o lento decrescimento da ACF.

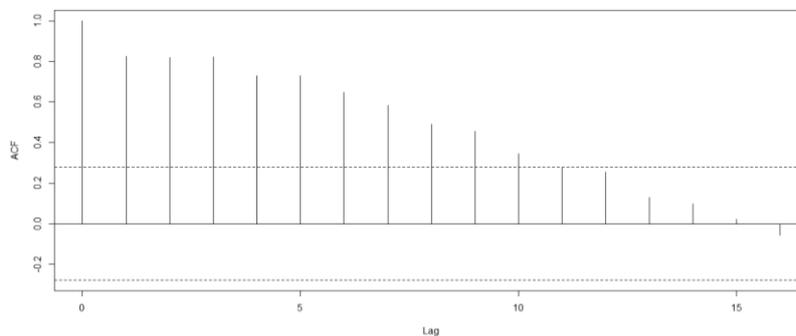


Figura 5 – Correlograma da ACF de uma série temporal não estacionária simulada por meio do R. Fonte: elaborado pelo autor por meio do R (R CORE TEAM, 2014)

A segunda utilidade citada para a ACF, em conjunto com a PACF é a identificação de potenciais modelos para a série temporal. Segundo Box e Jenkins (1976) e Makridakis et al. (1998), uma série temporal que pode ser descrita por um modelo AR(1), apresenta em sua PACF somente o primeiro *lag* como estatisticamente significativo e, a ACF, apresenta um padrão de decaimento exponencial (ou senoidal) após o primeiro *lag*. Nas Figuras 6(a) e 6(b) são apresentadas a ACF e a PACF, respectivamente, de uma série temporal simulada, exemplificando os conceitos abordados (R DEVELOPMENT CORE TEAM, 2014).

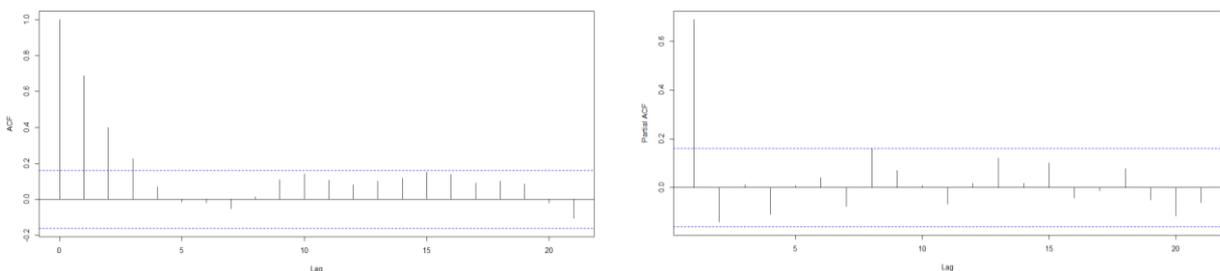


Figura 6 – (a) Correlograma da ACF de um modelo AR(1) e (b) correlograma da PACF de um modelo AR(1). Fonte: elaborado pelo autor por meio do R (R CORE TEAM, 2014)

O exemplo anterior apresenta o padrão teórico da ACF e da PACF de séries temporais passíveis de serem modeladas por um modelo com um componente autorregressivo, ou AR(1). Os modelos aqui citados são descritos em maiores detalhes na seção seguinte, onde também se descreve o processo de identificação do modelo para a série a partir da ACF e da PACF. Para exemplificar o padrão dos correlogramas quando da presença de somente uma componente de médias móveis na série, ou um modelo MA(1), nas Figuras 7(a) e 7(b) são apresentadas as respectivas ACF e PACF.

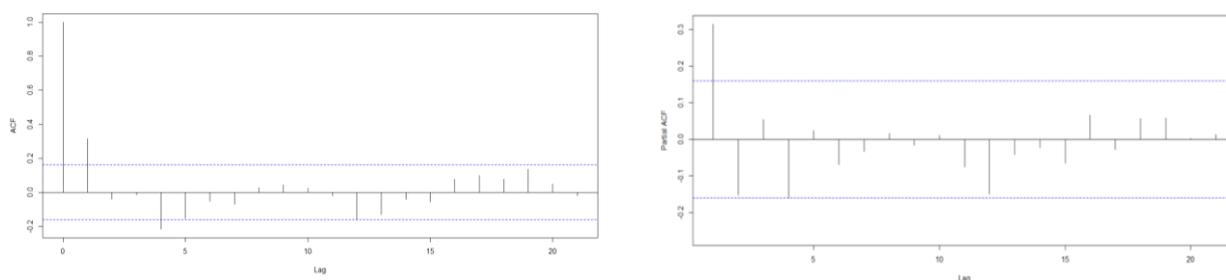


Figura 7 – (a) Correlograma da ACF de um modelo MA(1) e (b) correlograma da PACF de um modelo MA(1). Fonte: elaborado pelo autor por meio do R (R CORE TEAM, 2014)

Até então, foram citados os modelos teóricos para séries temporais com componentes puramente autorregressivo e puramente de médias móveis, respectivamente. Porém, segundo Makridakis et al. (1998), grande parte das séries temporais reais apresentam tanto componentes autorregressivos, quanto componentes de médias móveis. Para tanto, nas Figuras 8(a) e 8(b) são apresentados os correlogramas da ACF e da PACF, respectivamente, para um modelo teórico ARMA(2,2) simulado por meio do R (R CORE TEAM, 2014).

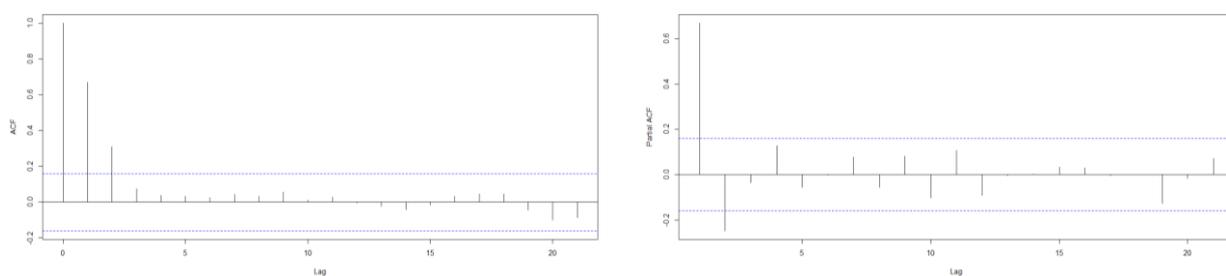


Figura 8 – (a) Correlograma da ACF de um modelo ARMA(2,2) e (b) correlograma da PACF de um modelo ARMA(2,2). Fonte: elaborado pelo autor por meio do R (R CORE TEAM, 2014)

O coeficiente de autocorrelação utilizado na modelagem é o coeficiente de autocorrelação amostral, tendo em vista a impossibilidade de obtenção do coeficiente populacional. O coeficiente de autocorrelação amostral descreve a correlação entre dois

valores de determinada variável, em diferentes períodos de tempo. O conjunto de tais coeficientes é denominado ACF e, a sua representação gráfica é denominada correlograma da ACF. Por meio da equação 3, pode-se calcular o coeficiente de autocorrelação amostral de *lag* (ou defasagem) k (MATOS, 2000; MAKRIDAKIS et al., 1998).

$$r_k = \frac{\sum_{t=k+1}^n (y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=1}^n (y_t - \bar{y})^2} \text{ para } k = 0, 1, 2, \dots, n. \quad (3)$$

onde r_k é o coeficiente de autocorrelação amostral, y_t representa o valor da série no tempo t , \bar{y} é o valor médio da série e n é o número de amostras contidos na série.

Já o coeficiente de autocorrelação parcial mede a correlação entre o valor de determinada série no tempo t e seu valor defasado em $t-k$ depois da influência de $y_{t-1}, y_{t-2}, \dots, y_{t-k+1}$ sobre y_t ser descontada (FAVA, 2000; ENDERS, 2004). Semelhantemente a ACF, o conjunto de valores de tal coeficiente é denominado PACF e a sua representação gráfica é dada por meio do correlograma da PACF. Fava (2000) e Morettin e Tolo (2006) citam duas formas para o cálculo dos coeficientes e da PACF, conforme as equações 4 a 6 onde a sequência de pares (k, ϕ_{kk}) constitui a PACF.

$$y_t = \beta_{11}y_{t-1} + \varepsilon_t \rightarrow \phi_{11} = \beta_{11} \quad (4)$$

$$y_t = \beta_{11}y_{t-1} + \beta_{22}y_{t-2} + \varepsilon_t \rightarrow \phi_{22} = \beta_{22} \quad (5)$$

$$y_t = \beta_{k1}y_{t-1} + \beta_{k2}y_{t-2} + \dots + \beta_{kk}y_{t-k} + \varepsilon_t \rightarrow \phi_{kk} = \beta_{kk} \quad (6)$$

onde β_{kk} são os k coeficientes do modelo de regressão donde se obtém os valores da função e ε_t é o resíduo do modelo de regressão para a estimativa dos coeficientes.

3.1.2 Modelos ARIMA estacionários e não estacionários

Esta seção tem como objetivo descrever os quatro grupos genéricos que compõem a classe dos modelos ARIMA estacionários e não estacionários.

O primeiro modelo ARIMA é o $AR(p)$, apropriado quando os dados de determinada série temporal, em determinados pontos no tempo, estão fortemente correlacionados com os valores que os antecedem e os que o sucedem (LEVINE et al., 2005). São modelos classificados como univariados, onde p é o número de defasagens de y_t (MATOS, 2000). O sentido prático do modelo é que a observação y no período t é gerada por uma média

ponderada de observações passadas, sendo p períodos anteriores inclusos na série, além do erro do período corrente. Este modelo é útil na representação de um grande número de séries temporais (PELLEGRINI, 2000). Segundo Fava (2000), um modelo genérico AR(p) pode ser representado conforme é dado na equação 7.

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t \quad (7)$$

Para simplificar a representação do modelo, utiliza-se o operador de defasagem B . Diversos autores da área adotam essa prática e argumentam que esse processo auxilia na modelagem algébrica da série temporal (MAKRIDAKIS et al., 1998; ENDERS, 2004; PINDYCK; RUBINFELD, 2004). Dessa forma, o modelo genérico AR(p) dado na equação 34 pode ser representado pelas equações 8 a 10 (FAVA, 2000; MORETTIN; TOLOI, 2006).

$$y_t = \phi_1 B y_t + \phi_2 B^2 y_t + \dots + \phi_p B^p y_t + \varepsilon_t \quad (8)$$

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) y_t = \varepsilon_t \quad (9)$$

$$\phi(B) y_t = \varepsilon_t \quad (10)$$

Os coeficientes autorregressivos das equações 8-10 (ϕ_p) são parâmetros que descrevem como um valor corrente relaciona-se com os seus respectivos valores passados e podem ser estimados através de mínimos quadrados ou máxima verossimilhança (PELLEGRINI, 2000; MATOS, 2000). Um dos pressupostos do modelo, é que o termo ε_t comporte-se como um ruído branco, ou seja, que o resíduo do modelo seja não autocorrelacionado, com média zero e variância constante (GUJARATI, 2000). O outro pressuposto do modelo, conforme Box et al. (2008), é a condição de estacionariedade, na qual é necessário que o operador ($\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$) tenha todas as suas raízes $\phi(B) = 0$ maiores que 1, ou seja, todas as raízes devem estar fora do círculo unitário.

O segundo modelo é o de médias móveis, denotado por MA(q). Tais modelos são gerados por uma média ponderada dos erros aleatórios com q defasagens (PINDYCK; RUBINFELD, 1981), diferentemente dos modelos AR(p) que são gerados por uma média ponderada dos valores defasados da série. Neste caso, o número de defasagens é determinado com base na ACF (MATOS, 2000). A equação 11 apresenta o modelo genérico MA(q) (FAVA, 2000).

$$y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} \quad (11)$$

Os parâmetros (θ_q) do modelo tanto podem ser positivos quanto negativos (PINDYCK; RUBINFELD, 1981) e, assim como para o modelo AR(p), utiliza-se o operador

de defasagem B (MAKRIDAKIS et al., 1998; ENDERS, 2004). Assim, com o operador de defasagem, o modelo $MA(q)$ apresentado da equação 11 pode ser representado pelo modelo dado na equação 12 (FAVA, 2000; MORETIN; TOLOI, 2006).

$$y_t = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) \varepsilon_t = \theta(B) \varepsilon_t \quad (12)$$

Os modelos $MA(q)$ são descritos por $q+2$ parâmetros, que são: (i) a média, (ii) a variância do erro aleatório e (iii) os parâmetros que determinam os pesos (PINDYCK; RUBINFELD, 1981). Para tanto, um dos pressupostos é que os erros aleatórios do modelo $MA(q)$ sejam gerados por um ruído branco, assim como nos modelos $AR(p)$ (FAVA, 2000).

Porém, pode ocorrer nas séries temporais estacionárias a presença das componentes $AR(p)$ e $MA(q)$ (MAKRIDAKIS et al., 1998), formando assim o modelo misto $ARMA(p,q)$. Descritivamente, o valor de y em determinado período de tempo t , portanto, é dado pela ponderação de seus valores passados (p períodos defasados), pela ponderação dos resíduos passados (q períodos defasados) e pelo erro aleatório.

No modelo $ARMA(p,q)$, utiliza-se um filtro ou componente autorregressivo de ordem p e um filtro ou componente de médias móveis de ordem q . Para que para as condições de estacionariedade e de invertibilidade de um modelo $ARMA(p,q)$ sejam satisfeitas, requer-se, respectivamente, que todas as p raízes de $\phi(B)$ iguais a zero e que todas as q raízes de $\theta(B)$ iguais a zero, caiam fora do círculo unitário (FAVA, 2000). Na equação 13 é apresentado o modelo genérico e nas equações 14 e 15, sua forma simplificada por meio do operador de defasagem B (MAKRIDAKIS et al., 1998).

$$y_t = \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q} \quad (13)$$

$$(1 - \phi_1 B - \dots - \phi_p B^p) y_t = (1 - \theta_1 B - \dots - \theta_q B^q) \varepsilon_t \quad (14)$$

$$\phi(B) y_t = \theta(B) \varepsilon_t \quad (15)$$

Porém, o fato é que muitas séries temporais são não estacionárias, pois as características do processo estocástico subjacente mudam ao longo do tempo (PINDYCK; RUBINFELD, 2004; GUJARATI, 2000). Caso a série apresente tal comportamento, realizam-se transformações denominadas diferenças e, a ordem de integração, é o número de diferenças necessárias para tornar a série não estacionária em estacionária (FAVA, 2000). Portanto, os modelos $ARIMA(p,d,q)$ são formados pelos componentes $AR(p)$ e $MA(q)$, diferenciados d vezes na série original (MAKRIDAKIS et al., 1998). O modelo genérico, após a série tornar-se estacionária, é dado na equação 16 e o modelo reduzido por meio do operador de

defasagem, é dado na equação 20, obtido por meio das combinações dos modelos dados nas equações 17 a 19 (FAVA, 2000).

$$w_t = \phi_1 w_{t-1} + \dots + \phi_p w_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}, \quad (16)$$

$$\text{Onde } w_t = \Delta^d y_t \quad (17)$$

$$(1 - \phi_1 B - \dots - \phi_p B^p) w_t = (1 - \theta_1 B - \dots - \theta_q B^q) \varepsilon_t \quad (18)$$

$$w_t = (1 - B)^d y_t, \text{ ou} \quad (19)$$

$$(1 - B)^d \phi(B) y_t = \theta(B) \varepsilon_t \quad (20)$$

onde w_t é o valor da série diferenciada, B^p é o operador de defasagem de autocorrelação e B^q é o operador de defasagem de médias móveis.

3.1.3 Modelos ARIMA sazonal estacionário e não estacionário

A última complexidade a ser adicionada aos modelos ARIMA é a sazonalidade (MAKRIDAKIS et al., 1998). A sazonalidade, segundo Ballou (2001), refere-se aos picos e vales regulares que se repetem anualmente na série. A notação do modelo ARIMA sazonal ou SARIMA, é a seguinte: SARIMA(p, d, q)(P, D, Q) $_s$, onde (p, d, q) representa a parte não sazonal, (P, D, Q) representa a parte sazonal do modelo e s representa o número de períodos da sazonalidade (TSENG et al., 2002). O modelo SARIMA genérico é dado na equação 21.

$$(1 - \phi_p B^p) (1 - \Phi_P B^s) (1 - B^d) (1 - B^D) y_t = (1 - \theta_q B^q) (1 - \Theta_Q B^s) \varepsilon_t \quad (21)$$

Na equação 21, a primeira parte da igualdade representa, respectivamente: os p componentes autorregressivos, os P componentes sazonais autorregressivos, as d diferenças e as D diferenças sazonais consideradas na modelagem da série temporal de período igual a s . A segunda parte da igualdade representa, respectivamente: as q componentes de médias móveis e as Q componentes sazonais de médias móveis de s períodos (MAKRIDAKIS et al., 1998).

A identificação de um modelo sazonal, assim como os não sazonais, é realizada por meio da ACF e da PACF (TSENG et al., 2002). Por exemplo, um modelo com uma componente sazonal de médias móveis na sazonalidade de $s=12$ períodos, ou SARIMA(0,0,0)(0,0,1) $_{12}$, espera-se, teoricamente, que o correlograma da ACF mostrará um pico significativo no *lag* 12. A PACF apresentará decaimento sendo os *lags* 12, 24, 36 e assim por diante, significativos. Similarmente, o modelo SARIMA(0,0,0)(1,0,0) $_{12}$ apresentará um decaimento exponencial no *lag* sazonal da ACF e um único pico significativo no *lag* 12 da PACF (MAKRIDAKIS et al., 1998; GHEYAS; SMITH, 2011).

Para exemplificar o processo de identificação de um modelo SARIMA, utilizou-se a série temporal de vazões de afluência (de janeiro de 1986 a dezembro de 2011) da hidrelétrica Itaipu Binacional (ONS, 2014) e então obtiveram-se os correlogramas da ACF e da PACF da série. Nas Figuras 9(a) e 9(b), são mostrados os respectivos correlogramas seguidos após, pela análise de seu comportamento.

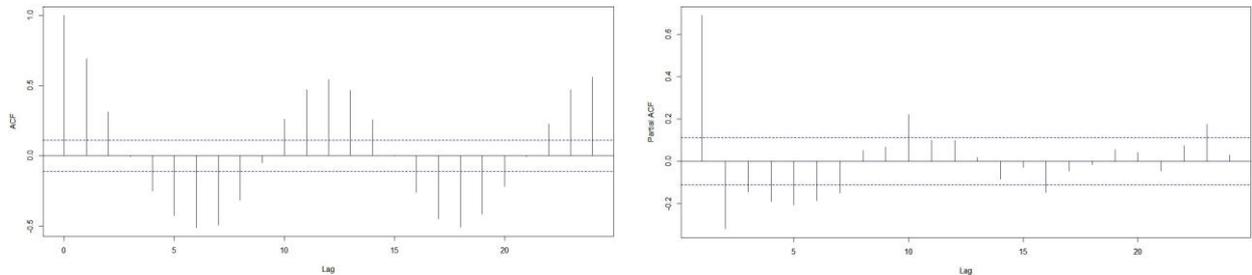


Figura 9 – (a) Correlograma da ACF e (b) correlograma da PACF da série temporal de vazões de afluência da hidrelétrica Itaipu Binacional. Fonte: ONS (2014)

Nas Figuras 9(a) e 9(b), são mostrados os correlogramas da ACF e da PACF para a série temporal obtida. Observa-se na ACF, que a série temporal apresenta um comportamento sazonal. À primeira vista, torna-se difícil identificar qualquer padrão existente nesta, pois observando-se o correlograma da ACF, visualiza-se uma persistência na autocorrelação sazonal (onde $s=6$) da série. Antes das demais análises, realizou-se um teste de raiz unitária, ADF (teste que será apresentado na Seção 3.1.4.1) e ao nível de 5% de confiança, foi rejeitada a hipótese de que a série seja não estacionária. Então, realizou-se uma diferença sazonal, ou de *lag* 12, para observar os padrões remanescentes na ACF e na PACF e assim, possivelmente, facilitar o processo de identificação do modelo. Nas Figuras 10(a) e 10(b), são mostrados os correlogramas da ACF e da PACF para a série temporal integrada de ordem $s=12$.

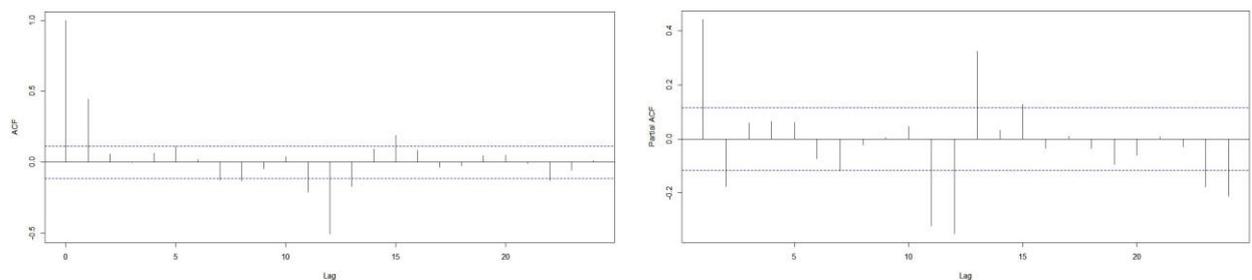


Figura 10 – (a) Correlograma da ACF e (b) correlograma da PACF da série temporal de vazões de afluência integrada de ordem $s=12$. Fonte: ONS (2014)

Conforme delineado por Makridakis et al. (1998) e Gheyas e Smith (2011), observa-se na Figura 10(a), que a ACF apresenta um pico significativo no *lag* 12, seguido de um decaimento senoidal, indicando uma componente sazonal de médias móveis. Além desta, verifica-se também um componente não sazonal de médias móveis, devido ao pico significativo no *lag* 1 da ACF. Na Figura 10(b), que mostra a PACF da série integrada, observa-se que os dois primeiros *lags* são significativos e seguidos de um decaimento senoidal. Além desses, os *lags*: 11, 12, 13, 15, 23 e 24 também são estatisticamente significativos. Como se torna difícil obter um modelo que capture todos os padrões descritos conjuntamente, um modelo candidato neste caso é o SARIMA(2,0,1)(2,1,1)₁₂, como provável candidato à modelagem da série.

Conforme exposto no parágrafo anterior, um único modelo SARIMA não teria a capacidade de capturar todos os padrões identificados na série. Para isso, utiliza-se a metodologia desenvolvida por Box-Jenkins (1976) para melhor identificação do modelo mais adequado, que é apresentada na seção 3.1.4.

3.1.4 Metodologia de Box-Jenkins para a modelagem e previsão de séries temporais

A metodologia de Box e Jenkins (1976) é desenvolvida conforme os seguintes passos: (i) obtenção dos dados históricos da variável de interesse; (ii) análise e teste de estacionariedade da série; (iii) obtenção da ACF e da PACF e, com base nessas, identificação do modelo ARIMA(p,d,q) mais apropriado; (iv) estimação do(s) parâmetro(s) do modelo identificado; e, (v) verificação da adequação do modelo, por meio do atendimento da condição de ruído branco (WANG, 2011; ZHANG, 2003). Os passos são realizados tantas vezes quantas forem necessárias, para obtenção de um modelo satisfatório que descreva adequadamente o processo gerador da série (ZHANG, 2003; MORETTIN; TOLOI, 2006). Os passos descritos são realizados por meio das três etapas da metodologia, que pode ser visualizada na Figura 11.

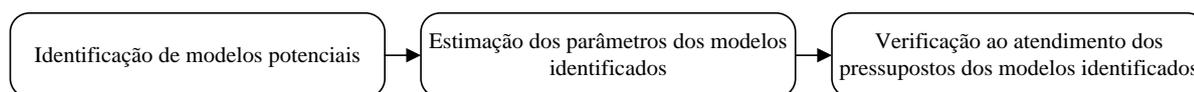


Figura 11 – Metodologia de Box-Jenkins para a modelagem e previsão de séries temporais. Fonte: Elaborado pelo autor com base em Makridakis et al. (1998)

Nas três seções a seguir, são descritos os passos e as análises realizadas em cada uma das etapas mostradas na Figura 11, para a aplicação da metodologia de Box-Jenkins.

3.1.4.1 Identificação de modelos potenciais

O primeiro passo a ser realizado na busca de um modelo que descreva a série, é analisar ou testar sua condição de estacionariedade (KUMAR; JAIN, 1999; ZHANG, 2003). Os testes para a verificação de estacionariedade são denominados testes de raiz unitária e objetivam diagnosticar se há a presença ou não de raiz unitária na série temporal. Caso diagnosticada como não estacionária, possuindo, portanto, raiz unitária, utilizam-se diferenciações na série até torná-la estacionária (MATOS, 2000; KUMAR; JAIN, 1999). Existem vários testes de raiz unitária, dentre eles: ADF e DF de Dickey e Fuller, Ljung-Box, Box-Pierce, KPSS, Phillips-Perron, entre outros (FAVA, 2000). Nesta dissertação, são abordados apenas os três primeiros.

Os testes de Box-Pierce e de Ljung-Box, são testes de estacionariedade baseados nos valores da ACF, e a diferença entre ambos é que o primeiro é apropriado para grandes amostras, enquanto o segundo para pequenas. As estatísticas dos testes citados, são dadas pelas equações 22 e 23, respectivamente, e seguem uma distribuição qui-quadrado, com m graus de liberdade (MAKRIDAKIS et al., 1998). A aceitação da hipótese nula conclui, ao nível de significância estabelecido, pela ausência de raiz unitária. Caso a hipótese nula seja rejeitada ao nível de significância estabelecido, conclui-se pela presença de raiz unitária, ou seja, a série é considerada então como não estacionária. A decisão por uma das hipóteses é dada por meio do cálculo da respectiva estatística e da comparação com a de qui-quadrado, se $Q(m)$ ou $L(m) > \chi_{\alpha}^2$, rejeita-se a hipótese nula e aceita-se a alternativa (MADDALA, 2003).

$$Q(m) = n \sum_{k=1}^m r_k^2 \quad (22)$$

$$L(m) = n(n+2) \sum \left[\frac{r_k^2}{(n-k)} \right] \quad (23)$$

Um teste de raiz unitária utilizado com frequência na literatura é o *Augmented Dickey-Fuller* ou ADF. O teste ADF é utilizado caso a série temporal apresente a possibilidade do erro aleatório ser autocorrelacionado (ENDERS, 2004) e, a raiz unitária é testada da mesma forma que no teste de Dickey-Fuller – para maiores detalhes sobre o teste de Dickey-Fuller,

consultar Box et al. (2008). Nas equações 24 a 26 são dados os modelos de regressão e os respectivos parâmetros a serem testados para a realização do teste ADF.

$$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \sum_{i=1}^{p-1} \delta_i \Delta y_{t-i} + \varepsilon_t \quad (24)$$

$$\Delta y_t = \alpha + \gamma y_{t-1} + \sum_{i=1}^{p-1} \delta_i \Delta y_{t-i} + \varepsilon_t \quad (25)$$

$$\Delta y_t = \gamma y_{t-1} + \sum_{i=1}^{p-1} \delta_i \Delta y_{t-i} + \varepsilon_t \quad (26)$$

Antes da aplicação do teste ADF, deve-se observar o comportamento em termos das componentes presentes na série temporal. O modelo de regressão para o teste de raiz unitária dado na equação 24, é utilizado em séries que apresentam nível (ou intercepto) e tendência; o modelo dado na equação 25, é utilizado em séries que apresentam somente nível; e, o modelo dado na equação 26, utilizado em séries que não possuam nenhum desses componentes. Para o teste, a aceitação da hipótese nula implica na rejeição de que a série temporal seja não estacionária, sendo o γ (conforme dado nas equações 24 a 26) o parâmetro testado. No Quadro 2, é mostrado um resumo da aplicação do teste, suas estatísticas e regras de decisão (FAVA, 2000). A estatística calculada para o teste ADF é comparada ao valor crítico da distribuição de Dickey-Fuller, dado o nível de significância estabelecido (GUJARATI, 2000).

Modelo estimado	H_0	Estatística do teste	Regra de decisão
Com nível e tendência (equação 24)	$\gamma = 0$	$\tau_\tau =$ estatística t de $\hat{\gamma}$	$\tau_\tau >$ valor crítico $\rightarrow H_0$ não é rejeitada
Com nível (equação 25)	$\gamma = 0$	$\tau_\mu =$ estatística t de $\hat{\gamma}$	$\tau_\mu >$ valor crítico $\rightarrow H_0$ não é rejeitada
Sem as componentes (equação 26)	$\gamma = 0$	$\tau =$ estatística t de $\hat{\gamma}$	$\tau >$ valor crítico $\rightarrow H_0$ não é rejeitada

Quadro 2 – Resumo para a aplicação do teste de raiz unitária ADF. Fonte: adaptado pelo autor com base em Fava (2000, p. 247)

Então, o resultado do teste indicará a necessidade ou não de se realizarem transformações na série temporal. Caso rejeitada a hipótese nula, é necessária a realização de uma diferença na série e, neste resultado, aplicar novamente o teste, finalizando esse ciclo somente após encontrar a ordem de integração. Porém, ressalta-se que a maior parte das séries temporais não estacionárias tornam-se estacionárias com, no máximo, duas diferenciações (BOX et al., 2008). O procedimento de diferenciação é realizado conforme dado nas equações

27 e 28 para a diferença de primeira ordem e para a diferença de segunda ordem, respectivamente (MAKRIDAKIS et al., 1998; GHEYAS; SMITH, 2011).

$$y_t' = y_t - y_{t-1} \quad (27)$$

$$y_t'' = y_t' - y_{t-1}' = (y_t - y_{t-1}) - (y_{t-1} - y_{t-2}) = y_t - 2y_{t-1} + y_{t-2} \quad (28)$$

O próximo passo é a identificação do modelo, segundo critérios previamente apresentados na Seção 3.1.4 e aqueles descritos no Quadro 3. Segundo Gujarati (2000), o padrão da ACF e da PACF da série temporal, quando comparado ao padrão teórico da ACF e da PACF, dá diretrizes ao pesquisador/analista para a escolha do(s) modelo(s) que possa(m) descrever o processo gerador da série temporal. Para tanto, pode-se observar no Quadro 3, as diretrizes para a identificação do modelo.

Tipo de modelo	Padrão teórico típico de FAC	Padrão teórico típico de FACP
AR (p)	Declina exponencialmente ou com padrão de onda senoidal amortecida, ou ambos	Picos significativos através das defasagens p . Truncada em $k = p$
MA (q)	Picos significativos através das defasagens q . Truncada em $k = p$	Declina exponencialmente
ARMA (p, q)	Declínio exponencial. Declinante a partir de $k = p$	Declínio exponencial

Quadro 3 – Padrões teóricos do correlograma da ACF e da PACF a serem comparados com os correlogramas da série. Fonte: adaptado pelo autor de Gujarati (2000), Fava (2000) e Morettin (2008)

Assim, a comparação dos correlogramas das ACF e PACF da série temporal com os padrões teóricos descritos no Quadro 3, possibilita a identificação de modelos e o seu respectivo número de parâmetros. Uma regra empírica desta metodologia, é que deve-se ter o cuidado de que o modelo identificado seja o mais parcimonioso possível (BOX; JENKINS, 1976). É natural que o aumento do número de parâmetros do modelo reduza a soma quadrada dos erros de ajustamento, porém, esse aumento provoca a diminuição do grau de liberdade da série. Segundo Maddala (2003) e Enders (2004), o número ótimo de parâmetros a serem introduzidos no modelo pode analisado por meio do critério informacional de Akaike (AIC) e do critério informacional Bayesiano (BIC). O número ótimo de parâmetros a serem inclusos, é aquele que minimiza os valores obtidos por meio dos critérios, dados nas equações 29 e 30.

$$AIC = n \log \left[\frac{SQR}{(n-p)} \right] + 2p \quad (29)$$

$$BIC = n \log \left[\frac{SQR}{(n-p)} \right] + p \log n \quad (30)$$

3.1.4.2 Estimação dos parâmetros do(s) modelo(s) identificado(s)

No próximo passo, são realizadas as estimações dos valores dos parâmetros do modelo identificado, ou seja, estimados os valores de p e q e/ou P e Q . A estimação dos parâmetros é realizada por meio do método dos mínimos quadrados ou pela máxima verossimilhança. Caso o modelo identificado tenha um componente de médias móveis, então se utiliza o método de mínimos quadrados não lineares (FAVA, 2000; MADDALA, 2003; MAKRIDAKIS et al., 1998). Para mais detalhe sobre tais métodos, consultar a obra de Box et al. (2008). Os modelos para a estimação dos parâmetros por meio do método de mínimos quadrados lineares, são dados nas equações 31 e 32. Na equação 33 é dado o método de estimação de máxima verossimilhança (MORETTIN, 2008).

$$S(\hat{\phi}_1, \dots, \hat{\phi}_p, \hat{\theta}_1, \dots, \hat{\theta}_q) = \sum_{t=1}^n \hat{\varepsilon}_t^2, \text{ onde} \quad (31)$$

$$\hat{\varepsilon}_t = \theta^{-1}(B)\phi(B)w_t \quad (32)$$

$$L(\xi|Y, y_0) = (2\pi\sigma^2)^{-\frac{T}{2}} \exp \left\{ \frac{-1}{2\sigma^2} \sum_{t=1}^T (y_t - \phi_0 - \phi_1 y_{t-1})^2 \right\} \quad (33)$$

3.1.4.3 Verificação ao atendimento dos pressupostos do(s) modelo(s) identificado(s)

O último passo da metodologia para a obtenção do modelo, é o de verificação. Este consiste em analisar os resíduos e se o modelo é parcimonioso, para constatar se este é ou não adequado para prever os valores futuros da série temporal. Caso não seja o adequado, analisam-se outras especificações para modelar a série (MORETTIN, 2008). A análise é realizada em duas etapas, que são descritas a seguir: (i) na análise do comportamento dos resíduos, ou, se estes apresentam ou não comportamento de ruído branco e (ii) na análise do critério de parcimônia do modelo obtido (FAVA, 2000).

Para que os resíduos do modelo ARIMA/SARIMA sejam considerados ruído branco, este deve apresentar média zero, variância constante e serem independentes ou não autocorrelacionados. Uma forma comum utilizada na literatura para se analisar o comportamento dos resíduos, é por meio de um gráfico da série temporal e do correlograma

da ACF dos resíduos (GUJARATI, 2000). Pelo gráfico, pode-se observar caso a média seja constante e igual a zero e se a variância dos resíduos permanece constante ao longo do tempo. Pelo correlograma, pode-se observar se as correlações entre os valores defasados do resíduo são estatisticamente significativas, caso este em que o resíduo é autocorrelacionado e o modelo, então, impróprio para realizar a previsão dos valores futuros da série.

Na análise da parcimônia do modelo, é verificado se este apresenta o número adequado de parâmetros e, o resultado dessa análise, pode ser classificado em três situações: (i) o modelo possui um número adequado de parâmetros, atendendo ao critério de parcimônia; (ii) o modelo está super especificado; ou, (iii) o modelo está sub especificado (FAVA, 2000). No caso de ocorrência das situações descritas em (ii) ou em (iii), as ações a serem tomadas em cada uma são descritas nos dois parágrafos subsequentes, respectivamente.

A super especificação acontece quando algum dos coeficientes estimados para o modelo, seja pequeno em relação ao erro-padrão proveniente de sua estimação. Tal situação pode ser verificada analisando a estatística t , ou o p -valor do parâmetro, ou até mesmo comparando o erro-padrão obtido pela estimação ao valor estimado do parâmetro. No caso de ocorrência da super especificação, deve-se retirar do modelo o respectivo parâmetro não significativo e, então, estimar os parâmetros do novo modelo, observando se o resultado gerado tenha então atendido a tal critério. A ocorrência desta situação é prejudicial, pois, embora o modelo obtenha um resultado melhor no ajustamento da série, a capacidade preditiva do modelo é reduzida (FAVA, 2000; MATOS, 2000).

Caso ocorra a situação oposta, deve-se introduzir novos parâmetros ao modelo e, então, semelhantemente ao processo descrito no parágrafo anterior, deve-se analisar a sua significância estatística. Os parâmetros a serem inclusos no modelo, neste caso, podem ser analisados por meio da ACF e da PACF e/ou por meio de uma análise das estatísticas descritas no parágrafo anterior, no que tange à análise de significância de cada parâmetro. A utilização de um modelo sub especificado, assim como no caso da situação oposta, descrita anteriormente, é prejudicial para a previsão de valores futuros da série temporal, pois gera maiores erros nas previsões (FAVA, 2000).

3.2 Modelos de Redes Neurais Artificiais

O segundo modelo considerado nesta dissertação são os modelos de Redes Neurais Artificiais (RNA). Esta seção tem como objetivo apresentar a utilização das RNA para a

previsão de demanda, especificamente no que tange ao modelo ou topologia *Multilayer Perceptron*, aqui denominado de MLP/RNA.

As RNA são sistemas computacionais de processamento de informações, inspiradas na arquitetura das redes neurais biológicas do cérebro humano, que por meio de algoritmos de treinamento, apresentam a habilidade de aprender padrões complexos de informação e generalizar a informação aprendida (ZHANG et al., 2001). As RNA são aplicadas a diversos problemas, dentre eles: *clusterização* ou análise de agrupamento, previsão de séries temporais e ao reconhecimento de padrões (ALLEMÃO, 2004). Uma das principais vantagens na utilização de RNA para a previsão de séries temporais, é devido a sua capacidade em modelagem não-linear, situação não verificada nos modelos de Box-Jenkins, por exemplo, apresentados na seção anterior (LEE; TONG, 2011; ZHANG et al., 2001).

3.2.1 Definições iniciais acerca das Redes Neurais Artificiais

As RNA começaram a ser estudadas na literatura da área, após a publicação do trabalho de Warren McCulloch e Walter Pitts, denominado “*A Logical Calculus of the Ideas Immanent in Nervous Activity*”, em 1943. A ideia inicial do estudo, foi em demonstrar a capacidade computacional de um neurônio biológico. McCulloch era pesquisador da área de psicologia e neuro-fisiologia e Pitts, que era recém graduado em matemática, juntou-se a McCulloch para colaborar na pesquisa. Este foi o trabalho pioneiro na área de RNA, pois abriu caminho para um novo ramo da ciência.

O trabalho de McCulloch e Pitts, por ser o trabalho seminal da área e devido ao seu foco, deixou uma importante questão aberta – como seria realizado o treinamento ou aprendizagem do neurônio artificial, tendo em vista que este busca simular o comportamento de um neurônio biológico? Para responder a essa questão, Donald Hebb, em 1949, publicou a obra intitulada “*The Organization of Behavior*”, onde o autor desenvolveu um modo de proporcionar capacidade de aprendizado às RNA, baseado na tentativa de explicar o aprendizado das redes neurais biológicas (BRAGA et al., 2011).

Após o trabalho de Hebb, já ao final da década de 50, o pesquisador Frank Rosenblatt aprimorou a representação do modelo computacional das redes neurais biológicas e desenvolveu uma rede de múltiplos neurônios, que ficou conhecida como *perceptron*, sendo então criada a primeira topologia de RNA. Além de desenvolver a topologia, Rosenblatt também propôs um algoritmo para o treinamento da RNA, para que essa pudesse realizar determinados tipos de funções. Concomitantemente ao trabalho de Rosenblatt, Bernard

Widrow desenvolveu um princípio de treinamento que ficou conhecido como *regra delta* e, este princípio, deu origem ao algoritmo de treinamento *backpropagation*, que é o mais conhecido e utilizado dos algoritmos de treinamentos de RNA (KOVÁCS, 2006).

Após o desenvolvimento do modelo computacional de um neurônio e dos algoritmos para treinamento, foram desenvolvidas as arquiteturas ou topologias de RNA. Para a compreensão destas, faz-se necessária a compreensão do funcionamento de um neurônio artificial. Um neurônio é uma unidade de processamento de informação que é fundamental para a operação de uma rede neural. Na Figura 12, é mostrado um neurônio artificial e os seus respectivos componentes, que são descritos na sequência (HAYKIN, 1999).

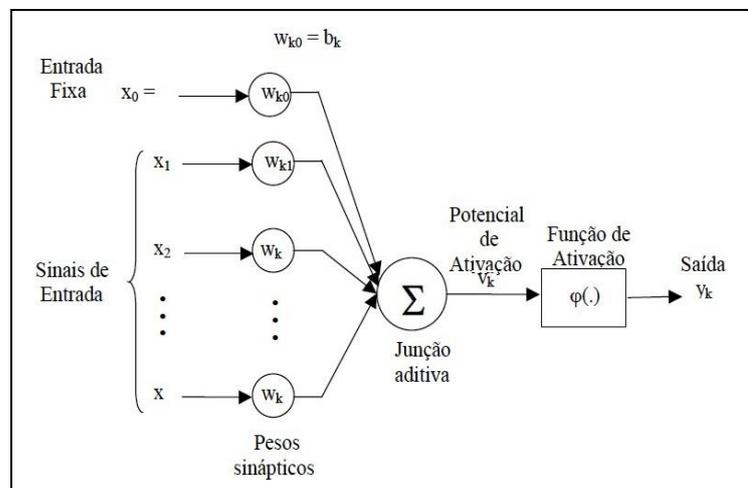


Figura 12 – Modelo de um neurônio artificial. Fonte: adaptado de Haykin (1999)

Na Figura 12, pode-se observar os elementos que descrevem um neurônio artificial, que são: (i) os sinais de entrada, (ii) a entrada fixa ou *bias*, (iii) os pesos sinápticos entre as conexões, (iv) a junção aditiva e potencial de ativação, (v) a função de ativação e (vi) a saída. Cada um destes elementos, é descrito nos parágrafos a seguir.

As conexões entre os sinais de entrada e o neurônio, é dada por meio das sinapses. Cada sinapse possui um peso atribuído, que na primeira iteração do algoritmo de treinamento, são valores aleatórios. Especificamente, um sinal x_j na entrada da sinapse j conectada ao neurônio k é multiplicado pelo peso sináptico w_{kj} . O primeiro índice se refere ao neurônio em questão e o segundo se refere ao terminal de entrada da sinapse à qual o peso se refere.

Na junção aditiva ou somador, são realizadas a soma dos sinais de entrada que são ponderados pelas respectivas sinapses do neurônio, que constitui-se num combinador linear.

A função de ativação tem o objetivo de restringir a amplitude da saída de um neurônio. A família de funções mais utilizada são as funções sigmóides e, dentre estas, uma das mais conhecidas é a função logística, dada na equação 34 (HAYKIN, 1999):

$$\varphi(v) = \frac{1}{1 + \exp - av} \quad (34)$$

onde a é o parâmetro de inclinação da função.

O *bias* de uma RNA tem o objetivo de aumentar ou diminuir a entrada líquida da função de ativação, evitando o seu enviesamento (HAYKIN, 1999; FLORES, 2009). Assim, pode-se definir um neurônio segundo as equações 35 e 36.

$$u_k = \sum_{j=1}^m w_{kj} x_j \quad (35)$$

$$y_k = \varphi(u_k + b_k) \quad (36)$$

onde x_1, x_2, \dots, x_m são os sinais de entrada; $w_{k1}, w_{k2}, \dots, w_{km}$ são os pesos sinápticos do neurônio k ; u_k é a saída do combinador linear devido aos sinais de entrada; b_k é o *bias*; $\varphi(\cdot)$ a função de ativação e y_k é sinal de saída do neurônio.

3.2.2 Arquiteturas de RNA

A arquitetura de uma RNA, diz respeito a maneira pelo qual os neurônios de uma rede neural estão estruturados e está diretamente ligada com o algoritmo de aprendizagem usado para treinar a rede (HAYKIN, 1999). A arquitetura, ou topologia de uma RNA, pode ser dividida em 3 classes gerais, que diferem em seu funcionamento, que são: (i) redes alimentadas adiante com camada única; (ii) redes alimentadas adiante com múltiplas camadas e (iii) redes recorrentes (KOVÁCS, 2006).

3.2.2.1 Redes alimentadas adiante com camada única

Trata-se da forma mais simples de uma RNA em camadas. Nesta topologia, tem-se uma camada de entrada de nós de fonte que se projeta sobre uma camada de saída de neurônios, mas não vice-versa. A designação “camada única”, se refere à camada de saída, porque é apenas nesta que é realizada algum tipo de computação (HAYKIN, 1999). Na Figura 13, pode-se observar a arquitetura de uma rede alimentada adiante de camada única.

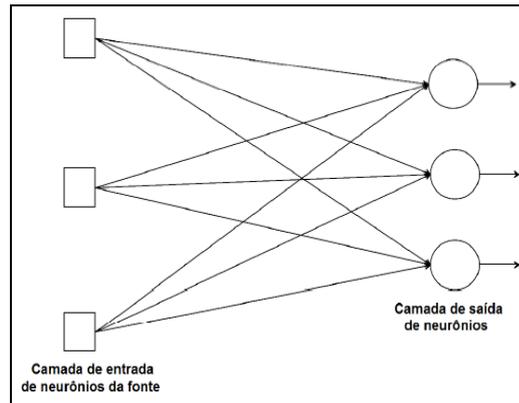


Figura 13 – Arquitetura de rede alimentada adiante de camada única. Fonte: Flores (2009)

3.2.2.2 Redes alimentadas diretamente com múltiplas camadas

Segundo Qi e Zhang (2001), esta é a arquitetura de RNA mais conhecida e mais utilizada também. Essa classe distingue-se pela presença de uma ou mais camadas ocultas, cujos neurônios presentes nessas, também são chamados de neurônios ocultos. A função desses neurônios, é intervir entre a entrada externa e a saída da rede de uma maneira útil, tornando a rede capaz de extrair informações de ordem elevada a partir dos dados de entrada (HAYKIN, 1999; BRAGA et al., 2011).

Os nós da camada de entrada fornecem os elementos do padrão de ativação, dado pelo vetor de entrada, que constituem os sinais de entrada aplicados aos neurônios na segunda camada, ou primeira camada oculta. Os sinais de saída da segunda camada são utilizados como entradas para a terceira camada, se houver, até o resto da rede (HAYKIN, 1999; KOVÁCS, 2006). Na Figura 14, mostra-se a arquitetura de uma rede alimentada diretamente com múltiplas camadas (FLORES, 2009).

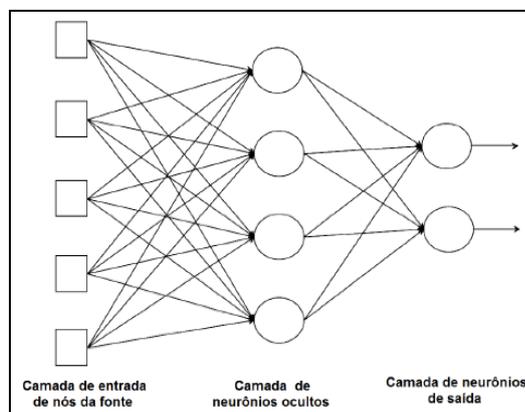


Figura 14 – Arquitetura de rede alimentada adiante com múltiplas camadas. Fonte: Flores (2009)

A arquitetura de rede alimentada adiante com múltiplas camadas seguem a seguinte nomenclatura: $X - Y - Z$, onde X é o número de neurônios na camada de entrada, Y é o número de neurônios na camada oculta (ou camada escondida) e Z é o número de neurônios na camada de saída. Caso a rede apresente mais de uma camada oculta, introduz-se a respectiva quantidade de neurônios na sequência. Na Figura 14, pode-se verificar que esta arquitetura corresponde ao modelo 5-4-2, com 5 neurônios na camada de entrada, 4 neurônios na camada oculta e 2 neurônios na camada de saída (HAYKIN, 1999; FLORES, 2009).

3.2.2.3 Redes recorrentes

São redes que possuem a capacidade de um comportamento dinâmico e não linear, devido ao uso de ramos particulares compostos de elementos de atraso unitário (representados por Z^{-1} nas Figuras 15(a) e 15(b)). Esta rede distingue-se de uma rede alimentada adiante por possuir, pelo menos, um laço de realimentação, que é quando a saída da rede alimenta a entrada de mais de um neurônio. O laço citado também pode ser de auto realimentação, que acontece quando a saída de um neurônio é realimentada para a sua própria entrada. Nas Figuras 15(a) e 15(b), pode-se observar as características das arquiteturas de rede recorrente sem e com auto realimentação, respectivamente (HAYKIN, 1999; FLORES, 2009).

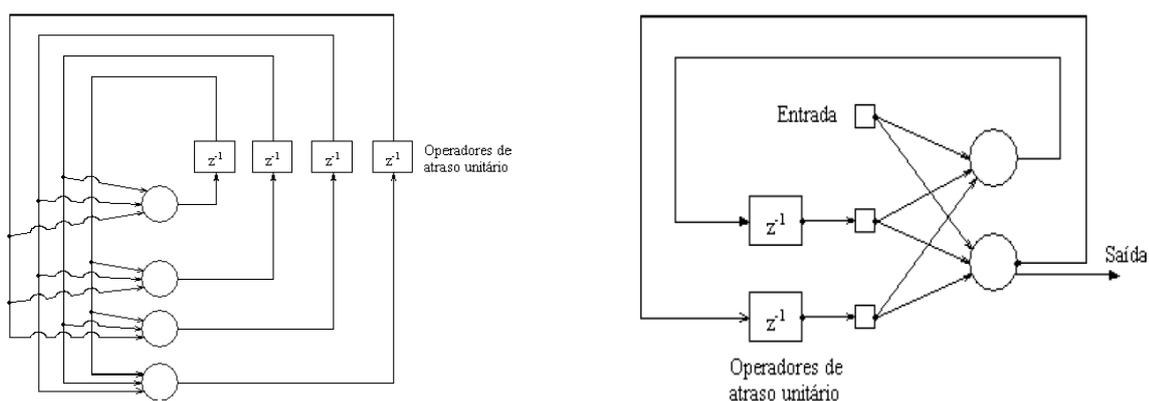


Figura 15 – (a) Arquitetura de rede recorrente e (b) arquitetura de rede recorrente com camada oculta e laços de auto-realimentação. Fonte: Flores (2009)

3.2.3 Aprendizagem e treinamento de uma RNA

A principal propriedade de uma RNA é a sua habilidade de aprender a partir de seu ambiente e de melhorar o desempenho através da aprendizagem. Simplificadamente, a aprendizagem de uma RNA é o processo de ajuste dos pesos sinápticos a cada iteração do

algoritmo de aprendizagem. O treinamento é parte integrante deste processo, para que a RNA obtenha a capacidade de generalização, que diz respeito ao quão bem a rede responde a um estímulo em sua camada de entrada para os dados que não foram apresentados à RNA no referido processo. A aprendizagem divide-se em três etapas: (i) tarefa de aprendizagem, (ii) método de aprendizagem e (iii) aplicação do método (HAYKIN, 1999; BRAGA et al., 2011).

O objetivo da aprendizagem de uma RNA, ou tarefa de aprendizagem, é dividida nas seguintes finalidades: (i) associação de padrões, (ii) reconhecimento de padrões, (iii) controle, (iv) aproximação de funções e (v) filtragem (BRAGA et al., 2011). Conforme Zhang e Qi (2001), na previsão de séries temporais, a tarefa de aprendizagem realizada pela RNA é a de aproximação de funções e essa constitui-se em uma das suas principais aplicações. Um vetor contendo as entradas de uma função desconhecida e um vetor contendo os resultados desta função são apresentados à RNA. Após o treinamento da rede, esta é capaz de reproduzir a função do respectivo conjunto de dados, inclusive para valores da função que não foram apresentados à RNA na fase de treinamento (FLORES, 2009).

O método de aprendizagem diz respeito a como a rede será habilitada a cumprir com a sua tarefa previamente definida (FLORES, 2009; HAYKIN, 1999). Na previsão de séries temporais, o método de aprendizado utilizado é o do tipo supervisionado por correção de erros (QI; ZHANG, 2001). Neste método, procura-se minimizar o erro da resposta da rede em relação à saída desejada, que é dado na equação 38. A aplicação deste método é conhecida como do tipo supervisionado, pois para a aproximação da função no treinamento da RNA, além dos valores de entrada da função, deve-se apresentar à rede os valores obtidos pela função para o respectivo conjunto de entrada (KOVÁCS, 2006; BRAGA et al., 2011).

$$e_k(n) = d_k(n) - y_k(n) \quad (38)$$

onde n é a iteração dos ajustes dos pesos sinápticos, k é o neurônio de saída, $y_k(n)$ é a resposta estimada do neurônio k no passo n , $d_k(n)$ é a resposta desejada do neurônio k , conforme os valores obtidos pela função para o conjunto de entrada e $e_k(n)$ é o erro do neurônio k no passo n (HAYKIN, 1999; FLORES, 2009).

Este método de aprendizagem é conhecido como *backpropagation*, pois a correção dos erros por meio dos pesos sinápticos, é realizada em duas fases: (i) na propagação ou fase *feedforward* e (ii) na retropropagação ou fase *backward* ou *error propagation*, conforme é mostrado na Figura 16 (FARUK, 2010).

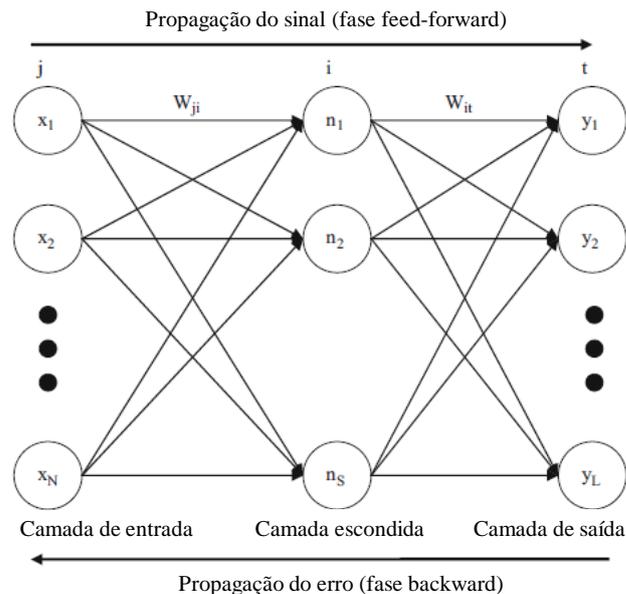


Figura 16 – A propagação do sinal (fase *feed-forward*) e a retropropagação do erro em uma rede MLP (fase *backward*). Fonte: Faruk (2010)

Na Figura 16 é ilustrado o processo de aprendizagem por meio do algoritmo *backpropagation* em uma RNA com três camadas: j , i e t . Na primeira fase, da propagação, um vetor de entrada é aplicado aos nós sensoriais da rede e seu efeito se propaga através desta, camada por camada. Durante esse passo, os pesos sinápticos são todos fixos. Na retropropagação do erro, é realizada a correção dos erros, ajustando os pesos sinápticos internos da rede, de acordo com a regra de correção do erro do gradiente descendente (BRAGA et al., 2011). Conforme mostrado na equação 38, a resposta real da rede é subtraída de um resposta desejada (alvo) para produzir um sinal de erro. Este sinal de erro é então propagado para trás através da rede, contra a direção das conexões sinápticas (HAYKIN, 1999). O mecanismo de correção do erro é dado na equação 40. Para isso, utiliza-se uma função de custo, segundo a equação 39 (FLORES, 2009):

$$E(n) = \frac{1}{2} e_k^2(n) \quad (39)$$

$$\Delta w_{kj}(n) = -\eta e_k(n) x_j(n) \quad (40)$$

onde $w_{kj}(n)$ é o peso sináptico do neurônio k excitado por um elemento $x_j(n)$ do vetor $x(n)$ na iteração n e η é uma constante positiva para a taxa de aprendizado para cada nova iteração. A taxa de aprendizado é determinada pelo modelador e é importante, pois influencia diretamente no desempenho da aprendizagem da RNA (HAYKIN, 1999; FLORES, 2009).

Por fim, a RNA é submetida ao processo de treinamento, que consiste no ajuste dos pesos sinápticos aos exemplos apresentados à rede (ZHANG; QI, 2001; BRAGA et al., 2011). Flores (2009) destaca que o treinamento é a parte aplicada da aprendizagem, sendo essa última, um conceito mais amplo, conforme abordado nos parágrafos anteriores nesta seção.

No treinamento de RNA para a realização de previsão de séries temporais, a série temporal é separada em dois conjuntos: no conjunto de teste e no conjunto de treinamento (ZOU et al., 2007; FLORES, 2009). O conjunto de teste é utilizado para verificar a capacidade de generalização da RNA aos valores da série que não foram apresentados à rede na fase de treinamento (ZHANG, 2003). O conjunto de treinamento é utilizado para o ajuste dos pesos sinápticos da RNA, de forma a minimizar a função dada pela equação 39. Esses conjuntos são comparados após o processo de treinamento, para verificar se houve a ocorrência de super-treinamento, que acontece quando a rede apresenta ótimos resultados para o conjunto de treinamento e um baixo ajuste ao conjunto de teste, demonstrando que esta não possui a capacidade de generalização (ZOU et al., 2007; BRAGA et al., 2011).

3.2.4 Modelo *Multilayer Perceptron* para a previsão de séries temporais

A área de previsão de séries temporais tem sido por muitas décadas, dominada por métodos estatísticos lineares. A aproximação por meio de modelos lineares a séries temporais com padrões não lineares, geralmente não apresenta resultados satisfatórios (ZHANG et al., 2001). No início dos anos 80, Makridakis et al. (1982) desenvolveram um estudo de previsão de séries temporais em larga escala, onde grande parte dos modelos lineares comumente utilizados foram testados utilizando 1001 séries temporais reais. Os resultados demonstraram que não há um único modelo globalmente melhor e, segundo Zhang et al. (2001), a maior razão para esse resultado, é em função do variado grau de não linearidade presente nas séries.

Nas últimas três décadas, diversos modelos não lineares para a previsão de séries temporais têm sido desenvolvidos, por exemplo: o modelo bilinear (GRANGER; ANDERSON, 1978); o modelo *threshold autoregressive* (TAR) (TONG; LIM, 1980); o modelo *smoothing transition autoregressive* (STAR) (CHAN; TONG, 1986); e, o modelo *autoregressive conditional heteroscedastic* (ARCH) (ENGLE, 1982). Conforme Zhang et al. (2001), tais modelos são úteis para determinados problemas e dados, porém, não podem ser generalizados para grande parte das aplicações. Isso porque, nas séries temporais podem haver diversos padrões não lineares quanto diversos padrões lineares presentes, fato esse que dificulta a parametrização deste tipo de modelo.

Em contrapartida, as RNA são modelos não paramétricos que podem capturar estruturas não lineares sem qualquer tipo de definição em relação ao processo gerador da série temporal (ZHANG et al., 2001). Já em relação à captura de padrões lineares na série de dados, Faruk (2010) afirma que os modelos de RNA não são tão capazes de capturar tais padrões quanto são os modelos ARIMA de Box-Jenkins.

Existem diversas topologias de RNAs aplicáveis à previsão de séries temporais (ZHANG et al., 2001; ABURTO; WEBER, 2007). Porém, a *multilayer perceptron* (MLP) é uma das mais utilizadas para as mais variadas aplicações (GURESEN et al., 2011). As RNA da topologia MLP são geralmente compostas por uma ou diversas camadas escondidas, além das camadas de entrada e de saída (ZHANG et al., 2001).

A previsão de séries temporais por meio de RNA geralmente é realizada subdividindo o conjunto ou amostra de dados em dois subconjuntos: no conjunto de treinamento e no conjunto de teste. O conjunto de treinamento é usado para a construção ou treinamento da RNA, conhecido na literatura da área como *in-sample data*. O conjunto de teste ou validação, por sua vez, é usado para medir a capacidade preditiva do modelo treinado, e é conhecido na literatura como *out-of-sample* ou *hold-out sample*, em virtude de não ter sido utilizado para a modelagem e treinamento da RNA (ABURTO; WEBER, 2007).

Para a previsão de séries temporais univariadas por meio de RNAs, as entradas da rede são os valores passados ou defasados e, a saída é o valor predito. As RNAs da topologia MLP realizam a modelagem dos valores da camada de entrada para a camada de saída conforme a equação 41 (ZHANG et al., 2001).

$$y_t = f(y_{t-1}, y_{t-2}, \dots, y_{t-p}) \quad (41)$$

onde y_t é a observação no tempo t , p é a dimensão do vetor da camada de entrada ou número de observações passadas usadas para prever o comportamento futuro da série e f geralmente é uma função não linear determinada pela estrutura da rede MLP. É possível observar que, neste caso, a rede MLP pode ser vista como um modelo geral autorregressivo não linear (ZHANG et al., 2001).

Para Su et al. (1997) e Tseng et al. (2002), o procedimento para a utilização de um modelo MLP para a previsão de séries temporais pode ser realizado da seguinte forma: (i) normalizar o conjunto de treinamento, (ii) decidir a respeito da arquitetura e seus parâmetros, (iii) inicializar todos os pesos da rede aleatoriamente, (iv) treinar a rede, até o momento em que atinja algum critério de parada, (v) escolher a rede MLP que gerou o menor erro e (vi) realizar previsão com a rede selecionada.

3.3 Combinação de previsões individuais

A última seção do Capítulo 3 trata da combinação das previsões. Para melhor organização, a seção descreve inicialmente os motivos de utilização e como se utiliza a combinação das previsões individuais em séries temporais. Em seguida, descreve algumas das técnicas existentes de combinação e que serão utilizadas no presente estudo.

3.3.1 Motivos para se combinar duas ou mais previsões

A combinação de previsões é uma técnica que surgiu com o trabalho de Bates e Granger (1969) e consiste, após obtida a previsão por meio de dois ou mais métodos, em utilizar técnicas para agregar as duas previsões em uma. Conforme Makridakis e Hibon (2000) e Costantini e Pappalardo (2010), a utilização das técnicas de combinação geralmente apresenta maior acurácia em relação aos modelos individuais que compõem a previsão. A obtenção de maior acurácia é explicada por Bates e Granger (1969), por meio de um exemplo utilizando a média dos erros de previsão, conforme mostrado na Tabela 1.

Tabela 1 – Combinação dos erros de previsão de dois métodos na série temporal de milhas voadas por passageiro, em 1953. Fonte: adaptado de Bates e Granger (1969)

Mês	Erros de previsão pelo método 1	Erros de previsão pelo método 2	Combinação dos erros de previsão
Janeiro	1	-3	-1
Fevereiro	6	-10	-2
Março	18	24	21
Abril	18	22	20
Maiο	3	-9	-3
Junho	-17	-22	-19,5
Julho	-24	10	-7
Agosto	-16	2	-7
Setembro	-12	-11	-11,5
Outubro	-9	-10	-9,5
Novembro	-12	-12	-12
Dezembro	-13	-7	-10
Variância dos erros	196	188	150

Os métodos de previsão utilizados na série de milhas voadas por passageiros em 1953, foram o de Brown, método 1, e o de Box e Jenkins, método 2. Observa-se na Tabela 1, que a variância dos erros de previsão é reduzida quando são combinados os métodos 1 e 2 por meio da média aritmética, evidenciando que a combinação pode melhorar o resultado de previsão.

Segundo Bates e Granger (1969), quando se realiza previsão de eventos futuros geralmente utiliza-se mais de um modelo de previsão e, então, compara-se os resultados obtidos e descarta-se o modelo que obteve o pior. Porém, o objetivo da previsão é obter maior acurácia e o modelo descartado, pode conter informações importantes a respeito do processo gerador da série. Assim, a combinação é um meio para se aproveitar as informações obtidas nas previsões e o tempo despendido na modelagem dos modelos de previsão descartados.

Os métodos de combinação atualmente existentes vão desde a simples realização da média das previsões, até métodos teoricamente mais complexos, capazes de modelar a não estacionariedade na combinação dos pesos (MENEZES et al., 2000). Donaldson e Kamstra (1996), por exemplo, estimaram os pesos da combinação utilizando RNAs. Em geral, conforme Menezes et al. (2000), os principais métodos de combinação de previsões são subdivididos nas categorias descritas no Quadro 4.

Categoria	Descrição
Média simples	Conforme Clemen (1989), esta categoria tem sido a principal escolha entre os métodos, devido a sua simplicidade e robustez.
<i>Outperformance</i>	Nesta categoria, o peso da combinação é obtido por métodos Bayesianos e são adaptativos para cada passo da previsão. Cada peso é interpretado como a probabilidade de que a previsão obterá o melhor desempenho.
Ótimo	A categoria de métodos de otimização visa minimizar a variância do erro de combinação. Os métodos de Bates e Granger (1969), por exemplo, enquadram-se nesta categoria.
Ótimo assumindo independência	Nestes métodos de otimização, os pesos são adaptativos conforme o passo de previsão e os resíduos de previsão dos modelos devem ser independentes.
Ótimo com restrições nos pesos	Aqui, além das restrições descritas na categoria anterior, é acrescentada a restrição de que os pesos devem estar inclusos no intervalo [0,1].
Regressão	Neste método, as previsões obtidas pelos modelos são utilizadas em uma regressão por meio de mínimos quadrados.
Regressão com restrições nos pesos	O método é semelhante ao anterior, porém é adicionada a restrição de que a soma dos pesos deve ser igual a 1.

Quadro 4 – Categorização dos métodos de combinação de previsões. Fonte: adaptado de Menezes et al. (2000)

A combinação de previsões é realizada baseando-se em um conjunto de dados históricos e informações contextuais (SILVA, 2014). Assim, diferentes métodos de previsão podem ser utilizados na combinação, dependendo do interesse do analista em determinados métodos e seus benefícios. Na Figura 17, é mostrada a obtenção da previsão final utilizando-se da combinação dos métodos de previsão.

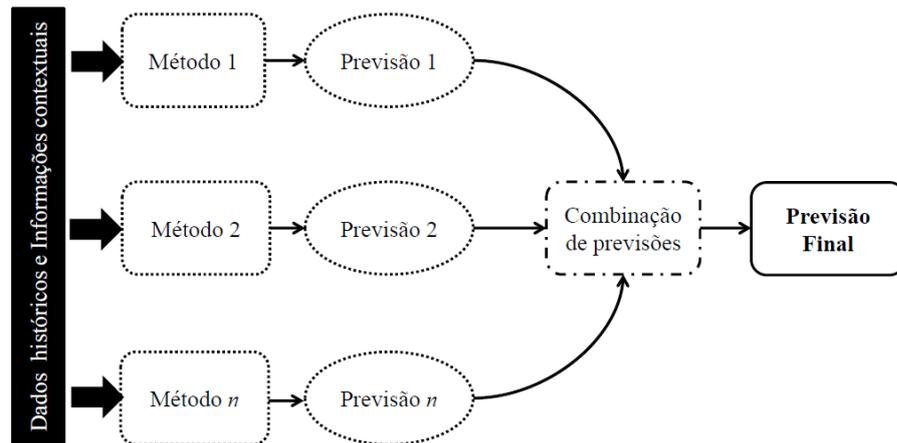


Figura 17 – Obtenção da combinação das previsões individuais. Fonte: Silva (2014)

A combinação das previsões mostrada na Figura 17, pode ser expressa matematicamente como a soma ponderada das previsões individuais obtidas pelos métodos utilizados, conforme é dado na equação 42. O(s) peso(s) w_n , dados na equação 42, é(são) obtido(s) pela técnica de combinação utilizada.

$$\hat{z}_t = w_1 \hat{x}_t^{(1)} + w_2 \hat{x}_t^{(2)} + \dots + w_n \hat{x}_t^{(n)} \quad (42)$$

onde \hat{z}_t é a combinação das previsões no período t obtidas por $\hat{x}_t^{(1)}$, $\hat{x}_t^{(2)}$ e $\hat{x}_t^{(n)}$, que são as previsões pelo método 1, pelo método 2 e pelo método n , respectivamente, e w_n é o peso de cada n previsão na combinação.

Dentre as diversas técnicas de combinação existente, o presente estudo utilizará oito, que são descritas na seção a seguir.

3.3.2 Técnicas de combinação das previsões

O método de combinação inicialmente proposto por Bates e Granger (1969), consiste na obtenção do peso ótimo e combinação linear de duas previsões. A obtenção do peso ótimo é baseada nas variâncias dos erros das previsões individuais e na correlação entre os erros de

previsão das previsões individuais. Portanto, antes da aplicação do método, deve ser analisado se a correlação entre os erros das previsões individuais é significativa. Caso seja, utiliza-se a equação 43 para a obtenção do peso ótimo e, caso contrário, utiliza-se a equação 44.

$$w = \frac{\sigma_2^2 - \rho\sigma_1\sigma_2}{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2} \quad (43)$$

$$w = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} \quad (44)$$

onde w é o peso do modelo 1 na combinação, σ_1^2 é a variância dos erros de previsão do primeiro modelo, σ_2^2 é a variância dos erros de previsão do segundo modelo, ρ é a correlação entre os erros de previsão do primeiro e do segundo modelo, σ_1 e σ_2 são os desvios-padrão dos erros de previsão dos modelos 1 e 2, respectivamente.

Este método de combinação é conhecido como método da variância mínima (VAR) e é utilizado na combinação de duas previsões. Após a obtenção do peso ótimo da combinação, realiza-se a combinação das previsões por meio da equação 45.

$$\hat{z}_t = w\hat{x}_t^{(1)} + (1-w)\hat{x}_t^{(2)} \quad (45)$$

onde \hat{z}_t é a combinação das previsões no período t obtidas por $\hat{x}_t^{(1)}$ e $\hat{x}_t^{(2)}$, a previsão pelo método 1 e pelo método 2, respectivamente, e w é o peso ótimo da combinação.

A média aritmética (MD) é uma técnica de combinação de fácil utilização e que pode apresentar desempenho superior em relação a técnicas sofisticadas. Diversos estudos têm comparado o desempenho da combinação da média aritmética e concluído que esta é uma boa alternativa quando se deseja a utilização de um método simples (WINKLER; MAKRIDAKIS, 1983; STOCK; WATSON, 2004). A combinação das previsões de dois modelos por meio da média aritmética é dada na equação 46.

$$\hat{z}_t = 0,5\hat{x}_t^{(1)} + 0,5\hat{x}_t^{(2)} \quad (46)$$

Além da média aritmética, alguns autores como Andrawis et al. (2011), Patton e Scheppard (2009) e Chen et al. (2007), testaram e obtiveram resultados na utilização das médias geométrica (MG) e harmônica (MH), tendo em vista que estas são mais sensíveis a valores de previsão menores. Nas equações 47 e 48 são dadas as combinações das previsões por meio dos métodos citados, respectivamente.

$$\hat{z}_t = \sqrt{\hat{x}_t^{(1)}\hat{x}_t^{(2)}} \quad (47)$$

$$\hat{z}_t = \frac{2\hat{x}_t^{(1)}\hat{x}_t^{(2)}}{\hat{x}_t^{(1)} + \hat{x}_t^{(2)}} \quad (48)$$

As médias geométrica e harmônica dadas anteriormente, respectivamente, foram adaptadas por Andrawis et al. (2011) de forma que os pesos associados aos valores das previsões individuais fossem obtidos através de uma busca unidimensional no intervalo [0,1] que minimiza o MSE, dado na equação 54. No estudo realizado pelos autores, as médias ponderadas obtiveram um resultado significativamente superior na comparação com as médias simples e foi uma das técnicas que apresentaram o melhor resultado na previsão da demanda de turistas no Egito. Nas equações 49 e 50 são dadas a médias geométrica ponderada (MGP) e média harmônica ponderada (MHP), respectivamente.

$$\hat{z}_t = [\hat{x}_t^{(1)}]^w [\hat{x}_t^{(2)}]^{1-w} \quad (49)$$

$$\hat{z}_t = \frac{\hat{x}_t^{(1)} \hat{x}_t^{(2)}}{(1-w)\hat{x}_t^{(1)} + w\hat{x}_t^{(2)}} \quad (50)$$

O penúltimo método utilizado foi da classe dos modelos de regressão, conforme estudado por Granger e Ramanathan (1984). Segundo os autores, a principal vantagem deste tipo de método na combinação das previsões é que mesmo previsões enviesadas geram combinações não enviesadas, caracterizando-se como uma alternativa aos métodos de otimização, principalmente. Na equação 51 é dada a técnica de combinação do tipo regressão (MQ) abordada neste estudo.

$$\hat{z}_t = w_0 + w_1 \hat{x}_t^{(1)} + w_2 \hat{x}_t^{(2)} \quad (51)$$

onde w_0 é o intercepto vertical da regressão e w_1 e w_2 são os coeficientes da equação de regressão obtidos por meio de uma busca unidimensional, sendo que a soma dos coeficientes deve ser igual a 1.

O último método de combinação considerado no estudo foi o denominado *inverse mean square* (IM), desenvolvido por Stock e Watson (1999). Trata-se de um método com pesos adaptativos para cada horizonte de previsão, enquadrando-se na classe dos métodos ótimos, semelhante ao proposto por Bates e Granger (1969). Para a utilização do MSE como medida de *performance* na obtenção dos pesos, foi proposto por Andrawis et al. (2011) a utilização de uma média móvel de três períodos para o cálculo. Nas equações 52 e 53 são dados os pesos para a combinação dos dois modelos nos respectivos horizontes de previsão. A posterior combinação das previsões é realizada através da equação 42.

$$w_1^h = \frac{\sum_{j=-k}^k MSE_{h+j}^{(2)}}{\sum_{j=-k}^k MSE_{h+j}^{(1)} + \sum_{j=-k}^k MSE_{h+j}^{(2)}} \quad (52)$$

$$w_2^h = \frac{\sum_{j=-k}^k MSE_{h+j}^{(1)}}{\sum_{j=-k}^k MSE_{h+j}^{(1)} + \sum_{j=-k}^k MSE_{h+j}^{(2)}} \quad (53)$$

onde w_1^h e w_2^h são os pesos dos modelos 1 e 2 para cada horizonte h de previsão; $MSE_{h+j}^{(1)}$ é a média móvel da soma dos erros quadráticos do modelo 1 e $MSE_{h+j}^{(2)}$ é a média móvel da soma dos erros quadráticos do modelo 2.

Apresentadas as técnicas de combinação de previsões a serem utilizadas, a seção seguinte descreve um tema importante surgido no início deste século na área de combinação, que é a combinação de modelos.

3.3.2.1 Técnicas de combinação de modelos

Outras técnicas de combinação que têm sido utilizada na literatura e têm obtido resultado é a hibridização dos modelos ARIMA e RNAs. Neste tipo de técnica, combinam-se os modelos e não as previsões. Em geral, primeiramente obtém-se o modelo ARIMA para uma determinada série temporal e em seguida, modela-se uma RNA do tipo MLP conforme os parâmetros e resíduos obtidos no modelo ARIMA. Diversos estudos, tanto teóricos quanto empíricos, abordaram a hibridização de modelos de previsão e obtiveram resultados superiores em relação aos obtidos com os modelos individuais (TSENG et al., 2002; ZHANG, 2003; ABURTO; WEBER, 2007; FARUK, 2010; KASHEI; BIJARI, 2010; WONG; GUO, 2010; YADAV; SRINIVASAN, 2011; LEE; TONG, 2011). A seguir são descritos alguns dos resultados obtidos com a utilização deste tipo de combinação.

Tseng, Yu e Tzeng (2002), combinaram os modelos SARIMA e RNAs, o qual denominaram SARIMABP. As duas etapas da implementação são as seguintes: (i) modelagem da série utilizando o modelo SARIMA e (ii) modelagem das RNAs utilizando como *input* os erros e os valores previstos pelo modelo SARIMA.

Em Zhang (2003) é apresentada a combinação de modelos utilizando a metodologia de Box e Jenkins (1976) e RNAs, ou hibridização de modelos de previsão. O objetivo deste método de combinação é melhorar a capacidade preditiva quando comparada à utilização individual de cada modelo e capturar linearidades e não linearidades na série temporal. No estudo, para a comparação das *performances* dos modelos, o autor utilizou três séries temporais amplamente estudadas na literatura, que são: a série de manchas solares, o número

de lincos capturados por ano no rio Mackenzie, no Canadá, e a taxa de câmbio libra (Britânica) por dólar (Estados Unidos).

Para a série de manchas solares, a percentagem de melhoria da combinação sobre os modelos ARIMA e RNAs, no horizonte de 35 períodos adiante, foi de 16,13% e 9,89%, respectivamente. No horizonte de 67 períodos adiante, a percentagem de melhoria foi de 12,45% e 25,35%, respectivamente. Na série de lincos capturados por ano, prevista somente no horizonte de 14 períodos adiante, a combinação obteve a melhoria de 18,87% em relação ao modelo ARIMA e 18,76% em relação ao modelo MLP. Para a série taxa de câmbio libra por dólar, nos três diferentes horizontes de tempo utilizados na previsão, um mês, seis meses e doze meses, constatou-se que a combinação obteve resultados superiores de até 23% em relação aos modelos ARIMA e de 19% em relação aos modelos de RNAs, considerando todos os horizontes de previsão.

Khashei e Bijari (2010) combinaram os modelos ARIMA e RNAs, os quais denominaram *artificial neural network (p,d,q)*. O modelo foi implementado através das seguintes etapas: (i) modelagem da série utilizando o modelo ARIMA e (ii) modelagem da série por meio de RNAs utilizando como *input* os valores originais (considerando p defasagens) e os valores dos resíduos do modelo ARIMA (considerando q defasagens). Foram então modeladas as séries e comparados os resultados com aqueles obtidos por Zhang (2003). Para o horizonte de previsão de 35 períodos adiante da série de manchas solares, a combinação dos modelos pelo método proposto apresentou a melhoria de 48,50% e para o horizonte de previsão de 67 períodos adiante, a melhoria de 19,62%.

Não foram encontrados na literatura, estudos que definissem as nuances existentes entre as técnicas clássicas de combinação de previsões e as técnicas de combinação de modelos. Pesquisas com este objetivo e com o de comparar os desempenhos obtidos com cada tipo de combinação, forneceriam importantes informações acerca de qual tipo seria mais adequado para cada situação.

4 METODOLOGIA

Neste capítulo, aborda-se a descrição da metodologia adotada na modelagem, previsão e comparação dos resultados gerados pelos modelos utilizados e as suas respectivas combinações, a partir da série temporal estudada. A metodologia deste estudo de caso foi desenvolvida com base em 4 etapas, que são mostradas na Figura 18.

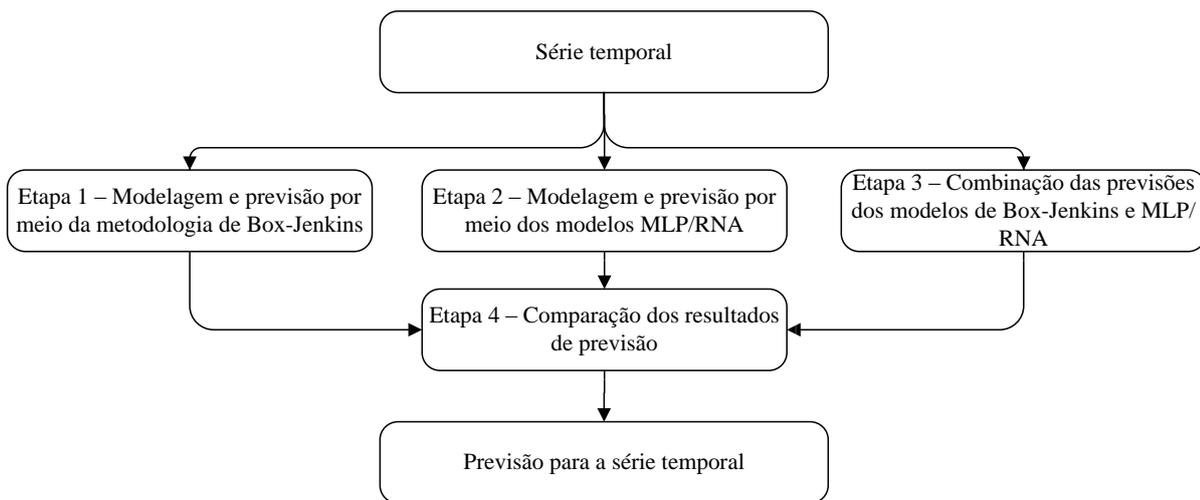


Figura 18 – Fluxograma da metodologia utilizada na previsão e comparação dos modelos e técnicas de combinação

As seções subsequentes descrevem os passos para a modelagem, previsão e comparação dos resultados, conforme mostrado na Figura 18.

4.1 Etapa 1 – Modelagem e previsão por meio da metodologia de Box-Jenkins

A primeira etapa foi a aplicação da metodologia de Box-Jenkins na obtenção do modelo ARIMA para a previsão dos valores futuros da série temporal. Os passos previstos na aplicação desta metodologia – a identificação dos modelos, a estimação dos parâmetros dos modelos identificados e a verificação ou validação do modelo obtido, foram incluídos no estudo na forma descrita na sequência. Na Figura 19, são mostrados os passos dados na obtenção do modelo ARIMA para a previsão dos valores futuros da série temporal.

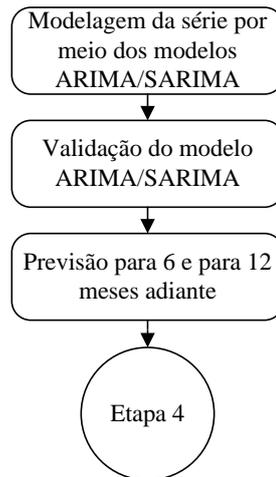


Figura 19 – Fluxograma da metodologia para a obtenção do modelo ARIMA para a previsão da série temporal

O primeiro passo foi realizar uma análise prévia do comportamento da ACF e da PACF, onde o objetivo foi verificar a estacionariedade ou não da série, tanto em tendência, quanto em sazonalidade. Em seguida, foi aplicado o teste estatístico de raiz unitária ADF, para confirmar a hipótese inicialmente estabelecida acerca da estacionariedade da série quanto à tendência. O teste foi aplicado ao nível de significância de 1%, 5% e 10%. Para a verificação da estacionariedade na sazonalidade, foi observado o comportamento da ACF e da PACF nos períodos sazonais s , ou seja, se houve persistência de picos significativos ao longo dos $lags$.

Após, foram identificados, por meio da ACF e da PACF e das diretrizes descritas no Quadro 3, os potenciais modelos ARIMA para a série. Os modelos sazonais, conforme Makridakis et al. (1998), foram identificados seguindo a mesma lógica. Os parâmetros sazonais autorregressivos foram identificados por meio do correlograma da ACF e os parâmetros sazonais de médias móveis foram identificados por meio do correlograma da PACF, em ambos os casos, observando a significância estatística dos coeficientes ao longo do número de defasagens da série.

Na validação do modelo ARIMA foram realizados os seguintes passos: (i) estimação dos parâmetros dos modelos identificados; (ii) seleção do modelo pelos critérios AIC e BIC; e, (iii) validação do modelo pela análise do comportamento dos resíduos obtidos. A validação foi realizada por meio da obtenção e análise do gráfico dos resíduos padronizados e pela obtenção e análise da ACF dos resíduos obtidos pelo modelo. Se o modelo obteve resíduos com média zero, variância constante e ausência de autocorrelação, o modelo foi validado para ser utilizado na previsão dos valores futuros da série temporal.

O último passo da primeira etapa foi realizar a previsão dos valores futuros por meio do modelo ARIMA, para 6 e para 12 meses adiante. Os dois critérios utilizados para analisar os resultados de previsão do modelo selecionado, para ambos os horizontes de previsão, foram o *Root Mean Square Error* (RMSE) e o *Mean Absolute Percentage Error* (MAPE), dado nas equações 58 e 59, respectivamente.

4.2 Etapa 2 – Modelagem e previsão por meio dos modelos MLP/RNA

Na segunda etapa, foram realizadas a modelagem e a previsão dos valores futuros da série por meio do modelo MLP/RNA. Foram utilizados dois tipos de modelos, classificados quanto ao número de neurônios considerado na camada de entrada: (i) um tipo com 6 e com 12 valores defasados da série – denominado a partir de então como MLP; e (ii) um tipo com, além desses valores, valores defasados de uma decomposição clássica realizada na série – o modelo DMLP.

Na Figura 20 é mostrada a metodologia considerada para a obtenção do modelo MLP e do DMLP. Cada modelo obtido nesta etapa foi utilizado na combinação realizada posteriormente, por meio dos métodos considerados dados na seção 3.3.

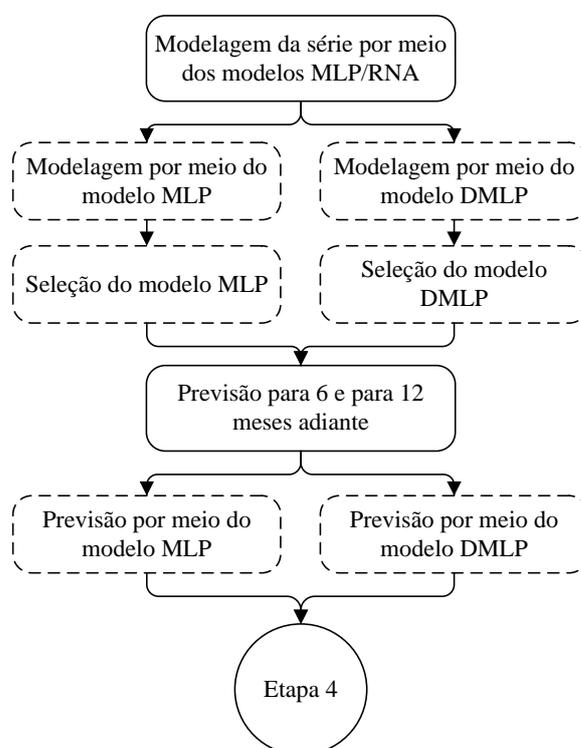


Figura 20 – Fluxograma da metodologia utilizada para a obtenção do modelo MLP e do modelo DMLP para a previsão de valores futuros da série temporal

Ao contrário da modelagem ARIMA, não há uma metodologia estabelecida para a utilização do modelo MLP para modelagem e previsão de séries temporais. Nesta área, existem diversas formas possíveis, exemplos podem ser vistos em Zhang (2003), Aburto e Weber (2007), Zou et al. (2007), Flores (2009), Khashei e Bijari (2010), Valipour et al. (2013), entre outros. A metodologia consiste na definição: dos conjuntos de treinamento e de teste, do número de neurônios na camada de entrada, do número de camadas escondidas, do número de neurônios na(s) camada(s) escondida(s), da função de transferência e, após, na realização do treinamento do modelo por meio de um algoritmo de aprendizagem.

O conjunto de valores para treinamento, foi aquele utilizado para a obtenção dos pesos sinápticos entre as conexões de cada neurônio de cada camada, de forma a minimizar o *Mean Square Error* (MSE), dado na equação 54. O conjunto de valores de teste foi aquele utilizado para verificar se a rede apresentou a capacidade de modelar o processo gerador da série temporal, que foram os valores dos conjuntos de previsão para 6 e para 12 meses.

$$MSE = \frac{1}{N} \sum_{t=1}^N (e_t^2) \quad (54)$$

onde N é o número de elementos considerados para o cálculo do MSE e e_t^2 é o resíduo quadrático no período t , obtido por meio do modelo considerado.

O número de neurônios na camada de entrada, foi definido com base na característica sazonal observada na série temporal estudada. Buscou-se a introdução da sazonalidade aos modelos MLP, por meio da defasagem de 6 e de 12 valores nos neurônios da camada de entrada, considerando então um tipo de modelo MLP com 6 e outro com 12 neurônios. Também, foi utilizado o modelo DMLP, onde neste, além de se considerar as defasagens dos valores da série conforme nos modelos MLP, também se considerou os valores defasados de uma decomposição clássica da série, conforme realizado em Zhang e Qi (2005).

Na equação 55 se apresenta o modelo matemático genérico das entradas que foram apresentadas à rede para os conjuntos de modelos de 6 e 12 neurônios, segundo Zhang et al. (2001) e, a equação 56 mostra o modelo matemático genérico das entradas considerando ainda os valores das componentes obtidas (ZHANG; QI, 2005).

$$y_t = f(y_{t-1}, y_{t-2}, \dots, y_{t-p}) \quad (55)$$

$$y_t = f(y_{t-1}, y_{t-2}, \dots, y_{t-p}, T_{t-1}, S_{t-1}, C_{t-1}) \quad (56)$$

onde y_t é a observação dada à RNA como sendo o valor alvo no tempo t , p é a dimensão do vetor de entrada referente ao número de observações passadas da série utilizadas para prever os valores futuros, f é uma função não linear determinada pela estrutura do modelo MLP e do

conjunto de dados de treinamento apresentados, T_t é o valor da tendência no tempo t , S_t é o valor da sazonalidade no tempo t e C_t é o valor da componente irregular.

A metodologia para a decomposição da série por meio do método clássico, pode ser vista com maiores detalhes na seção 2.2.2.

Em todos os modelos MLP e DMLP, foi utilizada apenas uma camada escondida. Para cada modelo MLP e DMLP, testaram-se os seguintes números de neurônios na camada escondida, respectivamente: 5, 10, 15, 20, 25, 30, 35, 40, 45 e 50. Em todos estes modelos, foi utilizado apenas um neurônio na camada de saída, onde se obteve o valor de modelagem e de previsão, dados os valores defasados apresentados à RNA nos neurônios da camada de entrada.

Para simplificar, adotou-se a nomenclatura exemplificada a seguir para descrever os modelos. O modelo MLP será citado conforme o seguinte exemplo – MLP(6,5,1), onde 6 é o número de neurônios na camada de entrada que contém os valores defasados referentes a 6 períodos, 5 é o número de neurônios da camada escondida e 1 é o número de neurônios na camada de saída. Já o modelo DMLP, conforme o exemplo a seguir – DMLP(6,5,1), sendo a interpretação semelhante àquela do modelo MLP anterior, diferenciando-se por haver três neurônios a mais na camada de entrada, inclusos para considerar as componentes da série na modelagem e previsão.

Então, foi realizado o treinamento das RNA por meio do algoritmo *backpropagation*, que ajustou os pesos até que se tenha atingido o MSE mínimo. Os modelos foram treinados por meio do *software Matlab R2010a* e, após, foram previstos os valores futuros para cada modelo nos respectivos horizontes de tempo. Os resultados de previsão das etapas supracitadas foram contabilizados por meio dos critérios RMSE e MAPE, dados nas equações 58 e 59, respectivamente. Os valores iniciais atribuídos aos pesos dos modelos foram valores aleatórios, donde foram sendo ajustados a cada iteração.

A seleção do melhor modelo MLP e DMLP foi realizada com base no RMSE e no MAPE para as previsões de 6 e de 12 períodos adiante e na verificação da ocorrência de super ajustamento aos valores do conjunto de treinamento. A ocorrência de super ajustamento foi verificada por meio de uma regressão entre os valores de entrada da RNA e os respectivos valores gerados na saída. Caso o modelo tenha apresentado alto grau de ajustamento aos valores do conjunto de treinamento e baixo grau de ajustamento aos valores do conjunto de teste, ou seja, uma diferença significativa entre estes, concluiu-se pela sua ocorrência.

4.3 Etapa 3 – Combinação das previsões dos modelos de Box-Jenkins e MLP/RNA

Na etapa 3, foram realizadas as combinações das previsões dos modelos ARIMA, MLP e DMLP selecionados. Na Figura 21, são mostrados os passos para a obtenção das respectivas combinações.

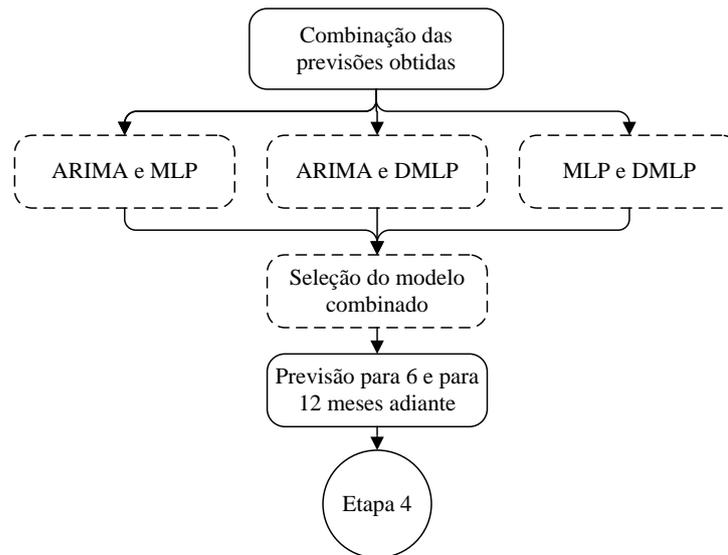


Figura 21 – Metodologia para a combinação das previsões obtidas

A combinação das previsões dos modelos foi realizada de três formas: (i) combinação das previsões dos modelos ARIMA e MLP, (ii) combinação das previsões dos modelos ARIMA e DMLP e (iii) combinação das previsões dos modelos MLP e DMLP. Para cada combinação de previsões, foram utilizados e comparados os resultados dos métodos de combinação descritos na seção 3.3.

Para simplificar a compreensão das combinações realizadas, define-se primeiramente a nomenclatura a ser usada, que é dada através de um exemplo – a combinação AR_MLP_VAR, foi obtida pela combinação da previsão do modelo ARIMA com a previsão do modelo MLP por meio do método VAR.

Para a combinação de previsões utilizando o método da variância mínima, primeiramente verificou-se a existência de correlação significativa entre os resíduos de previsão gerados pelos modelos individuais. A equação para o cálculo do coeficiente de correlação – $r_{(res1,res2)}$, que mede o grau de correlação entre os resíduos gerados pelos modelos, foi adaptada de Levine et al. (2005), conforme é mostrado na equação 57.

$$r_{(res1,res2)} = \frac{Cov(res1,res2)}{\sqrt{V(res1)}\sqrt{V(res2)}} \quad (57)$$

onde $Cov(res1,res2)$ é a covariância entre os resíduos de previsão do modelo 1 e os resíduos de previsão do modelo 2, $V(res1)$ é a variância dos resíduos de previsão do primeiro modelo e $V(res2)$ é a variância dos resíduos de previsão do segundo modelo.

Para a análise da correlação entre os resíduos de previsão obtidos pelos modelos, adotou-se o critério descrito em Levine et al. (2005). Assim, sendo o valor do coeficiente de correlação menor ou igual a 0,5 ou -0,5 (no caso de correlação negativa), a correlação entre os resíduos de previsão dos modelos pode ser considerada inexistente.

Na combinação das previsões por meio da média geométrica ponderada e da média harmônica ponderada, os pesos foram obtidos através da busca que gerou o menor MSE, considerando as restrições de cada técnica mostradas nas equações 49 e 50. Para isso, foi utilizado o método *evolutionary* disponível no complemento Solver do *software* Microsoft Excel. O *evolutionary* realiza a otimização dos valores utilizando algoritmos genéticos evolutivos, dado o critério que se deseja otimizar.

Na combinação das previsões através do método de mínimos quadrados ou de regressão, os pesos w_1 e w_2 foram obtidos pela busca que minimizou o MSE da combinação, considerando as restrições dadas na equação 51, utilizando o método *evolutionary*, assim como realizado no caso das médias ponderadas citadas no parágrafo anterior.

Para a aplicação da combinação proposta por Stock e Watson (1999), os pesos w_1 e w_2 das duas previsões foram calculados por meio de uma média móvel considerando $k=1$, assim como realizado em Andrawis et al. (2011).

Então, foi selecionada a melhor combinação das previsões dos modelos considerados. A seleção foi realizada com base no RMSE e no MAPE de previsão para 6 e para 12 meses.

4.4 Etapa 4 – Comparação dos resultados de previsão

Na última etapa deste estudo, foi realizada uma comparação entre os modelos utilizados para a previsão *versus* suas respectivas combinações e entre as combinações das previsões, com o objetivo de verificar qual das combinações resultou em um melhoramento significativo na capacidade preditiva. Após, analisou-se o desempenho obtido com as combinações das previsões em relação aos modelos utilizados. Para isso, foi necessário estabelecer o seguinte: (i) as medidas de acurácia dos modelos, (ii) a definição do método de

comparação entre os resultados gerados e (iii) a forma de mensuração do desempenho dos modelos. A seguir são descritos os passos citados.

As medidas de acurácia adotadas para mensurar os resultados obtidos nas etapas de modelagem, previsão para 6 e para 12 períodos adiante, para cada modelo, foram o *Root Mean Square Error* (RMSE) e o *Mean Absolute Percentage Error* (MAPE). Ambos os critérios mencionados são calculados por meio das equações 58 e 59, respectivamente (MAKRIDAKIS et al., 1998).

$$RMSE = \sqrt{\frac{1}{N} \sum_{t=1}^N (e_t^2)} \quad (58)$$

$$MAPE = \frac{1}{N} \sum_{t=1}^N \left| \frac{e_t}{y_t} \right| 100 \quad (59)$$

onde N é o número de elementos considerados para o cálculo do respectivo critério analisado, e_t é o resíduo no período t obtido por meio do modelo ou da combinação de previsão considerada e y_t é o valor da série temporal no período t .

A comparação dos resultados de previsão dos modelos e das combinações foi realizada conforme descrito nos passos a seguir: (i) aplicação de um teste estatístico de comparação entre os resultados de cada grupo de previsão considerando o critério RMSE; (ii) aplicação de um teste estatístico de comparação entre os resultados de cada grupo de previsão considerando o critério MAPE; (iii) análise do desempenho de cada combinação de previsão em relação aos modelos individuais considerando o critério RMSE; (iv) análise do desempenho de cada combinação de previsão em relação aos modelos individuais considerando o critério MAPE.

Para a realização das comparações utilizando um teste estatístico, conforme descrito nos itens (i) e (ii) do parágrafo anterior, foi necessário formar grupos dos resultados das previsões. As medidas de acurácia RMSE e MAPE de previsão, os resultados, foram agrupadas segundo a sua forma de obtenção. Os resultados das previsões obtidas por meio dos modelos individuais compuseram o grupo 1 (G1). Os resultados das combinações das previsões dos modelos ARIMA e MLP por meio dos oito métodos utilizados, compuseram o grupo 2 (G2). O resultado das combinações dos modelos ARIMA e DMLP compuseram o grupo 3 (G3). E, por fim, o resultado das combinações dos modelos MLP e DMLP compuseram o grupo 4 (G4). O Quadro 5 mostra a composição dos grupos e o respectivo número de resultados de previsões que foram alocados.

Grupo	Resultados das previsões alocados a cada grupo	Número total de resultados de previsão
G1	Modelos utilizados para as combinações das previsões: ARIMA, MLP e DMLP.	3
G2	Combinações do modelo ARIMA com o modelo MLP: AR_MLP_VAR, AR_MLP_MD, AR_MLP_MG, AR_MLP_MGEP, AR_MLP_MH, AR_MLP_MHP, AR_MLP_MQ e AR_MLP_IM.	8
G3	Combinações do modelo ARIMA com o modelo DMLP: AR_DMLP_VAR, AR_DMLP_MD, AR_DMLP_MG, AR_DMLP_MGP, AR_DMLP_MH, AR_DMLP_MHP, AR_DMLP_MQ e AR_DMLP_IM.	8
G4	Combinações do modelo MLP com o modelo DMLP: MLP_DMLP_VAR, MLP_DMLP_MD, MLP_DMLP_MG, MLP_DMLP_MGP, MLP_DMLP_MH, MLP_DMLP_MHP, MLP_DMLP_MQ e MLP_DMLP_IM.	8

Quadro 5 – Grupos de resultados de previsões para a comparação dos modelos e combinações e suas composições

Para a comparação dos grupos de resultados de previsões por meio do teste estatístico, primeiramente testou-se a normalidade dos conjuntos de dados de onde se obtiveram os RMSE e MAPE de previsão, pois este é um pressuposto para a sua aplicação. O teste utilizado para a verificação da normalidade das distribuições foi o de Shapiro-Wilk, pois o número de elementos amostrais para a sua realização foi menor do que 30 (LEVINE et al., 2005).

O próximo passo consistiu em realizar a comparação dos grupos por meio do teste estatístico, para identificar qual(is) o(s) tipo(s) de combinação(ões) apresenta(m) diferença significativa nos resultados de previsão. Em caso de não normalidade em uma ou mais amostras (ver Tabela 11), o teste de comparação consiste na análise de variância de um fator de Kruskal-Wallis (KW). As hipóteses nula e alternativa estabelecidas para o teste são dadas nas equações 60 e 61, respectivamente.

$$H_0 : \tau_{G1} = \tau_{G2} = \tau_{G3} = \tau_{G4} \quad (60)$$

$$H_1 : \tau_{G1}, \tau_{G2}, \tau_{G3}, \tau_{G4}, \text{ pelo menos um grupo é diferente.} \quad (61)$$

O nível de significância α estabelecido para o teste foi de 5%. Então, foram calculadas a estatística H de Kruskal-Wallis e comparada ao valor da estatística χ^2 obtida por meio da distribuição qui-quadrado, para o respectivo nível de significância e para ν graus de liberdade. O valor da estatística H de KW foi obtido por meio da equação 62.

$$H = \left(\frac{12}{n(n+1)} \sum_{j=1}^c \frac{T_j^2}{n_j} \right) - 3(n+1) \quad (62)$$

onde T_j é a soma dos postos da amostra j , n_j é o número de elementos contidos em cada amostra j e c é o número de grupos considerados no teste.

A estatística H do teste de KW segue uma distribuição qui-quadrado com $v=c-1$ graus de liberdade. A decisão pela hipótese nula (H_0) se dá caso o valor calculado de H para os valores amostrais for menor que o valor crítico da distribuição qui-quadrado e, do contrário, a decisão pela hipótese alternativa (H_1), caso o valor calculado de H para os valores amostrais, for maior que o valor crítico da distribuição qui-quadrado (DOANE, SEWARD, 2013).

Em caso de conclusão do teste favorável a H_1 , foi aplicado o teste estatístico de comparação múltipla descrito em Siegel e Castellan (1988). O objetivo consistiu na verificação de qual(is) grupo(s) diferiram significativamente. O teste de comparação múltipla é realizado conforme dado na equação 63.

$$|R_i - R_j| \geq Z_{\left(\frac{\alpha}{k(k-1)}\right)} \sqrt{\frac{N(N+1)}{12} \left(\frac{1}{n_i} + \frac{1}{n_j} \right)} \quad (63)$$

onde R_i e R_j são os postos dos grupos i e j , respectivamente; n_i e n_j são os tamanhos da amostra dos grupos i e j , respectivamente; N é o número total de elementos considerados em todas as amostras; $Z_{\left(\frac{\alpha}{k(k-1)}\right)}$ é o valor correspondente da normal padronizada dado o nível de significância α e o número de grupos k .

Em caso de ocorrência do resultado dado na equação 63, é aceita a hipótese alternativa que conclui pela existência de diferença significativa entre os grupos i e j , dado o nível de significância estabelecido. Caso contrário, a diferença não é significativa. A expressão matemática dada ao lado esquerdo da desigualdade é a diferença observada entre os postos dos respectivos grupos e, a dada no lado direito, a diferença crítica calculada (ver Tabela 12).

O último passo foi a obtenção dos desempenhos de cada combinação em relação aos modelos de previsão utilizados. O desempenho de cada combinação foi obtido através do cálculo da diferença percentual entre o resultado obtido e o resultado dos modelos, dada a medida de acurácia e o horizonte de previsão.

5 COMBINAÇÃO DAS PREVISÕES DOS MODELOS ARIMA E MLP/RNAS NO PLANEJAMENTO DA PRODUÇÃO

Neste capítulo está descrito um estudo de caso da aplicação dos modelos ARIMA, MLP, DMLP e os métodos utilizados para a combinação das previsões geradas por estes, conforme a metodologia descrita no Capítulo 4.

O estudo foi realizado em uma empresa de grande porte do ramo alimentício da região do Vale do Taquari, Rio Grande do Sul, que possui diversas unidades de negócio, incluindo a de laticínios. Esta unidade da empresa possui 58 produtos distribuídos em 7 famílias, que são: leite UHT, leite em pó, iogurtes, queijos, doces de leite, requeijão e nata.

Neste estudo, foi utilizada a série temporal da demanda agregada de leite do tipo UHT, pois é a principal família de produtos da empresa. A demanda agregada foi obtida pela demandas individuais dos cinco produtos que compõem esta família, que são: integral, semidesnatado, desnatado, de baixa lactose e rico em cálcio.

O Capítulo 5 encontra-se subdividido em duas partes: (i) onde se aplica a metodologia estabelecida e (ii) na discussão dos resultados obtidos na combinação das previsões.

5.1 Modelagem, previsão, combinação das previsões e comparação dos resultados

Na seção 5.1 é apresentada a modelagem e previsão dos valores futuros da série por meio dos modelos ARIMA e MLP, e as combinações das previsões por meio dos métodos considerados. Para tanto, primeiramente é apresentada a série temporal a ser estudada, conforme é dado na seção 5.1.1.

5.1.1 Série temporal da demanda agregada de leite tipo UHT

A série temporal da demanda agregada de leite UHT contém 90 observações espaçadas mensalmente. Para realizar as previsões, esta foi dividida em três conjuntos: (i) dos valores da demanda ocorrida entre o 1º e 78º período, que compuseram o conjunto de modelagem; (ii) dos valores da demanda ocorrida entre o 79º e 85º período, que compuseram o conjunto de previsão para 6 meses; e, (iii) dos valores da demanda ocorrida entre o 79º e 90º período, que compuseram o conjunto de previsão para 12 meses.

O conjunto de modelagem foi utilizado para a obtenção dos modelos ARIMA e MLP/RNA, onde em cada etapa foi selecionado o modelo que melhor descreveu o processo gerador da série. Os conjuntos de previsão de 6 e de 12 meses, que não foram utilizados na obtenção do modelo, foram utilizados para verificar a capacidade preditiva dos modelos e das combinações. Foram utilizados os horizontes de previsão de 6 e de 12 meses, pois para a empresa onde o estudo de caso foi aplicado, esses são os períodos considerados no planejamento da produção de médio prazo.

Na Figura 22, mostra-se a série temporal considerada. A partir desta, pode-se observar preliminarmente algumas características presentes e que serão analisadas na sequência, que são: (i) a série se desenvolve em torno de um determinado nível e (ii) há evidências que a série possua determinado grau de sazonalidade.

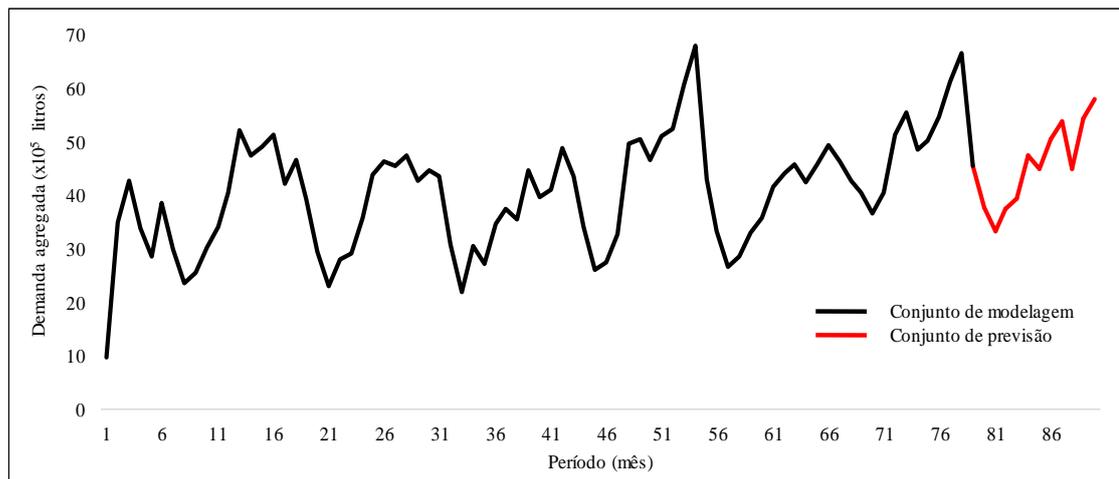


Figura 22 – Série temporal da demanda agregada de leite do tipo UHT. Fonte: dados fornecidos pela gerência da empresa estudada (2014)

5.1.2 Etapa 1 – Modelagem e previsão por meio da metodologia de Box-Jenkins

A primeira etapa para a utilização da metodologia de Box-Jenkins foi a análise das funções de autocorrelação (ACF) e de autocorrelação parcial (PACF), onde foi verificada a condição de estacionariedade e identificados os possíveis modelos para a série temporal. A modelagem e previsão por meio desta metodologia foi realizada utilizando o *software* R, versão 3.0.2 (R DEVELOPMENT CORE TEAM, 2014).

A Figura 23(a), em que se mostra a ACF da série temporal da demanda agregada do produto, demonstra que a série em análise apresenta sazonalidade do tipo autorregressiva e que possivelmente seja estacionária em tendência. A Figura 23(b), que mostra a PACF da

série, evidencia a existência de um parâmetro significativo de médias móveis a ser incluso no modelo.

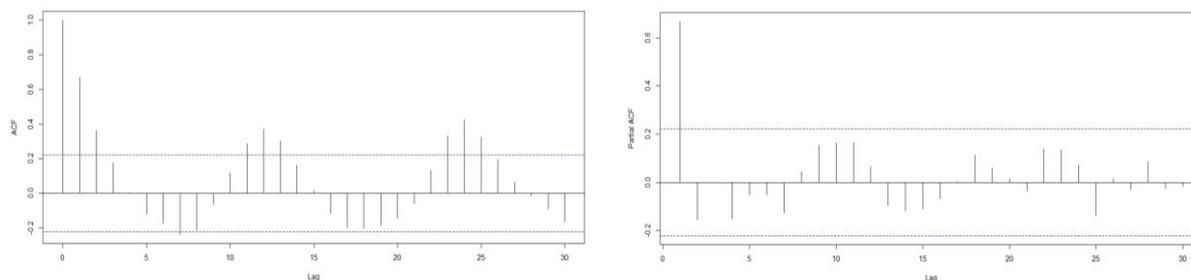


Figura 23 – Correlograma da ACF (a) e correlograma da PACF da série temporal estudada (b)

O correlograma da ACF da série, conforme Figura 23(a), demonstrou um comportamento aparentemente estacionário em tendência. Para uma adequada conclusão a respeito de tal condição, optou-se por realizar o teste de raiz unitária ADF para os níveis de confiança de 1%, 5% e 10%. Na Tabela 2 são mostrados os resultados do teste ADF para a verificação da estacionariedade nos respectivos níveis de significância.

Tabela 2 – Teste ADF para a verificação de estacionariedade da série temporal da demanda agregada de leite tipo UHT

Ordem testada	Valor do teste	Valor crítico do teste		Decisão do teste	
1	$\tau_2 = -3,504$	1%	$-3,510 / \tau_2 > \tau_c$	1%	Aceita-se H0 → série não estacionária
		5%	$-2,890 / \tau_2 < \tau_c$	5%	Rejeita-se H0 → série estacionária
		10%	$-2,580 / \tau_2 < \tau_c$	10%	Rejeita-se H0 → série estacionária

Para os níveis de significância de 5% e 10%, o teste ADF indica rejeição da hipótese nula, concluindo pela estacionariedade da série em tendência.

Observando-se o comportamento da ACF da série, verifica-se que esta apresenta dificuldades para a identificação de potenciais modelos a serem utilizados para a modelagem. Essa situação ocorre em função da persistência de padrões autorregressivos na componente sazonal da série ($s=12$), situação que pode acontecer devido à existência de raiz unitária na componente sazonal. Para tanto, foi realizada uma diferença sazonal na série, conforme a equação 64, e então analisada novamente a ACF e a PACF, com o objetivo de melhor identificar potenciais modelos.

$$\Delta^{12}y_t = y_t - y_{t-12} \quad (64)$$

Nas Figuras 24(a) e 24(b) são mostrados, respectivamente, o correlograma da ACF e o correlograma da PACF após realizada a diferença de ordem 12 ($s=12$) na série.

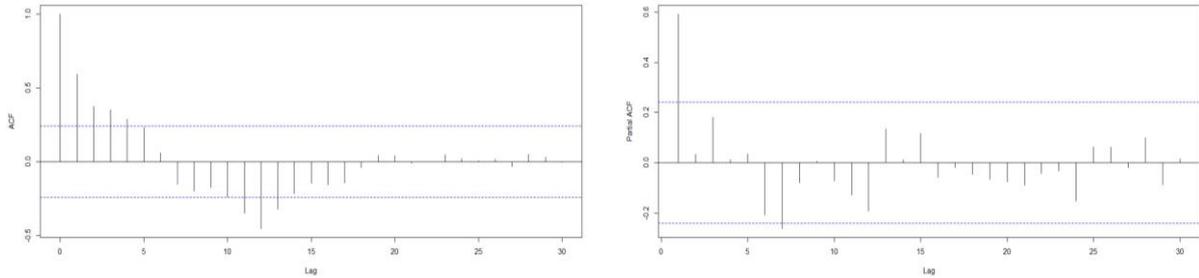


Figura 24 – Correlograma da ACF (a) e correlograma da PACF da série temporal estudada diferenciada uma vez na componente sazonal (b)

Os correlogramas da ACF e da PACF mostrados nas Figuras 24(a) e 24(b), melhoraram a capacidade de identificação de potenciais modelos em relação aos mostrados nas Figuras 23(a) e 23(b), possibilitando a identificação de cinco potenciais modelos para a modelagem do processo gerador da série, que são dados na Tabela 3.

Além dos modelos identificados, na Tabela 3 são apresentados os valores dos critérios AIC e BIC e os modelos selecionados para a estimação dos parâmetros.

Tabela 3 – Modelos SARIMA selecionados para a estimação dos parâmetros

Modelo	AIC	BIC	Observações
SARIMA(4,0,1)(1,1,0) ₁₂	441,34	456,67	
SARIMA(3,0,1)(1,1,0) ₁₂	439,92	453,06	Modelo selecionado para a estimação
SARIMA(1,0,1)(1,1,0) ₁₂	444,17	452,92	Modelo selecionado para a estimação
SARIMA(1,0,1)(2,0,1) ₆	519,85	533,99	Modelo estacionário em sazonalidade
SARIMA(1,0,1)(0,1,0) ₁₂	453,77	460,33	

Pode-se observar que o modelo SARIMA(3,0,1)(1,1,0)₁₂ e o modelo SARIMA(1,0,1)(1,1,0)₁₂, são aqueles candidatos à estimação e posterior modelagem, pois dentre os demais, minimizaram os critérios AIC e BIC. Observa-se que o AIC do primeiro modelo é menor do que o AIC do segundo, porém, o BIC do segundo é menor do que o BIC do primeiro modelo. Essa situação possivelmente tenha ocorrido devido ao BIC penalizar mais os modelos mais parametrizados, como é o caso do primeiro.

O próximo passo foi estimar os parâmetros dos modelos selecionados e observar a variância residual obtida por cada um, sendo este o critério usado para definir qual modelo SARIMA será utilizado na combinação. Na Tabela 4 são mostrados os resultados da estimação dos parâmetros, o respectivo *p-value* e a variância do erro associada a cada modelo.

Tabela 4 – Estimação dos parâmetros dos modelos selecionados

Modelos escolhidos	Valor dos parâmetros	<i>p-value</i> dos coeficientes	Variância do erro
SARIMA(3,0,1)(1,1,0) ₁₂	ϕ_1 : 0,1333	0,018	34,42
	ϕ_2 : 0,3086	0,045	
	ϕ_3 : 0,3103	0,052	
	θ_1 : 0,7361	< 0,01	
	Φ_1 : -0,6102	0,041	
SARIMA(1,0,1)(1,1,0) ₁₂	ϕ_1 : 0,7050	< 0,01	40,46
	θ_1 : 0,1971	< 0,01	
	Φ_1 : -0,5138	0,037	

Conforme se observa na Tabela 4, o modelo que obteve a menor variância residual foi o SARIMA(3,0,1)(1,1,0)₁₂. Todos os parâmetros deste modelo, conforme os respectivos *p-value*, são significativos ao nível de 5%. Portanto, este foi o modelo selecionado.

O penúltimo passo consistiu em verificar os pressupostos de ruído branco e ausência de autocorrelação dos resíduos. Na Figura 25 observa-se, os resíduos padronizados produzidos pelo modelo SARIMA(3,0,1)(1,1,0)₁₂ em função do período da observação.

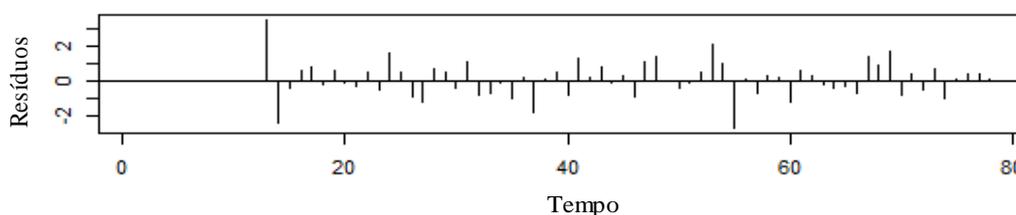


Figura 25 – Resíduos padronizados do modelo SARIMA(3,0,1)(1,1,0)₁₂ em função do período da observação

A seguir, analisou-se o comportamento da ACF dos resíduos do modelo identificado, conforme é mostrado na Figura 26.

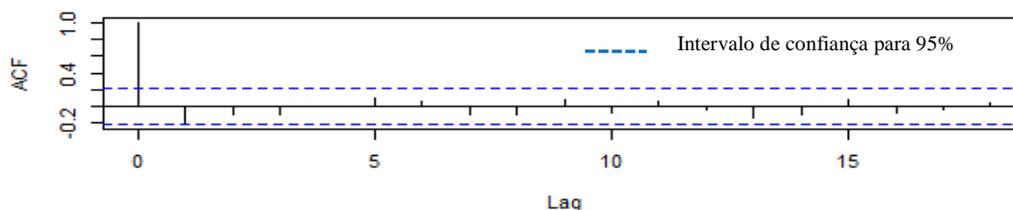


Figura 26 – Correlograma da ACF dos resíduos do modelo SARIMA(3,0,1)(1,1,0)₁₂

Observa-se que nenhum dos valores de autocorrelação dos resíduos do modelo selecionado ultrapassou o intervalo de confiança de 95%. Conclui-se então, pela ausência de autocorrelação nos resíduos, podendo o modelo SARIMA(3,0,1)(1,1,0)₁₂ ser utilizado para realizar a previsão dos valores futuros, resultado esse que é dado na Tabela 10.

5.1.3 Etapa 2 – Modelagem e previsão por meio dos modelos de MLP/RNA

A segunda etapa do estudo foi realizar a modelagem e a previsão utilizando as RNAs da topologia MLP. Tendo em vista que na etapa 1 observou-se que a série é sazonal, foram usados dois tipos de modelos: o primeiro, que utiliza de 6 e de 12 valores defasados da série para o treinamento das RNAs; e, o segundo, que utiliza de, além dos valores supracitados, os valores obtidos por meio uma decomposição clássica da série para a obtenção das componentes: tendência, sazonalidade e a componente irregular da série, conforme segue.

5.1.3.1 Modelagem e previsão por meio do modelo MLP/RNA

O primeiro passo dado na etapa 2 consistiu na modelagem da série temporal por meio dos modelos de RNAs da topologia MLP. Inicialmente foi definida a configuração, onde foram estabelecidos, entre outros – o número de neurônios na camada de entrada, na camada escondida e na camada de saída. Em seguida, realizou-se o treinamento dos modelos encontrados a partir das configurações consideradas, onde foram estabelecidos os pesos das conexões entre cada camada, para cada modelo. Por fim, realizou-se a modelagem e previsão da série por meio dos modelos MLP treinados e, então, foram contabilizados os resultados.

Considerando-se a existência da componente sazonal, as configurações testadas para a camada de entrada dos modelos MLP foram: um tipo de modelo que considera 6 valores defasados e, um tipo de modelo que considera 12 valores defasados da série. Nas Figuras 27(a) e 27(b) são mostradas, respectivamente, as representações dos modelos MLP utilizados para a modelagem e previsão da série.

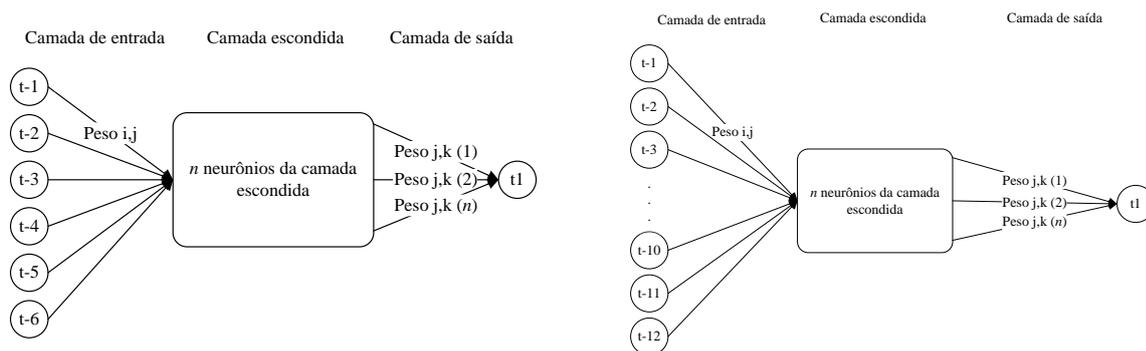


Figura 27 – Representação do modelo MLP com 6 (a) e com 12 neurônios na camada de entrada (b)

Nas Figuras 27(a) e 27(b), pode-se observar a forma pela qual os modelos MLP com 6 e com 12 neurônios na camada de entrada foram treinados, respectivamente. A cada iteração do modelo foram apresentados valores alvo, dados pelo valor da série no momento t e, na camada de entrada, a matriz dos valores defasados em n períodos, sendo n o número de neurônios inclusos no modelo.

Então foi realizado o treinamento dos modelos MLP e a partir dos valores da série, obtidos os resultados do ajustamento e previsão por meio do RMSE e do MAPE, que são mostrados na Tabela 5.

Nesta, por questões de organização na descrição dos modelos, utilizou-se a notação do modelo MLP conforme aquela utilizada por Flores (2009), por exemplo – MLP(6,5,1) é o modelo MLP com 6 neurônios na camada de entrada, 5 neurônios na camada intermediária e 1 neurônio na camada de saída.

Na Tabela 5 foi grifado em negrito o modelo MLP(12,30,1), por ter apresentado os menores valores dos critérios utilizados para medir o grau de ajustamento e previsão. Observa-se que, em termos de modelagem, o modelo selecionado encontra-se na décima posição como melhor modelo.

A não ocorrência de super ajustamento foi confirmada por meio de uma regressão linear simples entre os valores obtidos pelo modelo e os valores alvo apresentados ao modelo. Na Figura 28(a) e 28(b), são mostrados os gráficos de dispersão dos processos de treinamento e teste do modelo MLP(12,30,1), respectivamente, que demonstra tal situação. O eixo das ordenadas representa os valores obtidos pelo modelo MLP(12,30,1) e, o eixo das abscissas, os valores alvo apresentados à rede na etapa de modelagem. A linha escura (denominada *Fit*) mostrada nos gráficos representa o modelo de regressão ajustado.

Tabela 5 – Resultados dos modelos MLP/RNA utilizados para a modelagem e previsão da série temporal da demanda agregada de leite UHT

Modelos MLP	RMSE			MAPE		
	Model.	h = 6	h = 12	Model.	h = 6	h = 12
MLP(6,5,1)	5,20	10,53	9,20	9,27	18,99	15,44
MLP(6,10,1)	5,18	9,56	9,25	9,99	16,39	16,16
MLP(6,15,1)	6,00	13,71	12,30	8,44	28,80	23,08
MLP(6,20,1)	6,03	6,86	8,23	11,04	16,27	15,82
MLP(6,25,1)	7,11	12,47	10,52	15,33	30,34	21,9
MLP(6,30,1)	6,09	11,22	9,45	11,87	18,92	15,8
MLP(6,35,1)	6,72	13,57	11,42	12,69	27,67	21,9
MLP(6,40,1)	6,54	16,53	13,10	13,04	37,98	25,8
MLP(6,45,1)	6,05	14,71	11,58	11,04	30,35	20,6
MLP(6,50,1)	4,60	10,58	10,02	6,33	23,34	20,2
MLP(12,5,1)	4,97	6,08	8,10	9,00	13,08	14,81
MLP(12,10,1)	6,21	5,60	6,95	11,98	10,00	11,83
MLP(12,15,1)	4,52	4,60	6,65	9,16	8,66	11,29
MLP(12,20,1)	4,22	7,57	7,22	7,70	15,95	13,40
MLP(12,25,1)	4,63	10,06	9,97	9,88	21,60	19,07
MLP(12,30,1)*	6,44	4,31	6,00	11,44	6,66	9,93
MLP(12,35,1)	4,76	6,59	7,09	9,5	14,3	13,1
MLP(12,40,1)	4,39	8,27	7,57	6,3	17,7	14,0
MLP(12,45,1)	3,70	4,31	6,67	7,0	10,3	11,5
MLP(12,50,1)	6,63	7,09	8,46	8,3	16,1	15,7

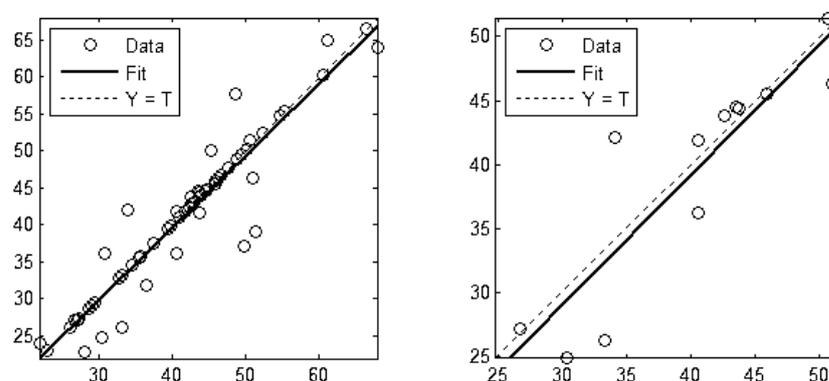


Figura 28 – Gráfico de dispersão dos resultados do processo de treinamento do modelo MLP (a) e gráfico de dispersão dos resultados do processo de teste do modelo (b)

O grau de ajuste do modelo de regressão linear simples mediu o quanto o modelo MLP(12,30,1) foi capaz de descrever os valores alvo apresentados e foi analisado por meio do

coeficiente de determinação R^2 . No treinamento, foram apresentados à RNA os valores alvo constantes no conjunto de treinamento dos modelos e, no processo de teste, foram apresentados os valores alvo não apresentados no treinamento, possibilitando verificar o potencial do modelo em prever os valores *out-of-sample* e da ocorrência de superajustamento. O valor de R^2 obtido no treinamento do modelo foi igual a 0,872 e, para o processo de validação, foi obtido o valor R^2 igual a 0,868. Assim é possível concluir que, no treinamento do modelo MLP(12,30,1), não ocorreu o superajustamento.

5.1.3.2 Modelagem e previsão por meio do modelo DMLP/RNA

Esta seção aborda o ajuste dos modelos DMLP, modelos de RNAs MLP idênticos ao da Seção 5.1.3.1, acrescido com valores dessazonalizados na camada de entrada. A configuração dos modelos DMLP foi realizada conforme descrito a seguir: (i) um tipo de modelo que considera 6 valores defasados mais os valores defasados da componente tendência, sazonalidade e irregular, obtidos por meio da dessazonalização e (ii) um tipo de modelo que considera 12 valores defasados da série mais os valores defasados da componente tendência, sazonalidade e irregular, obtidos por meio da dessazonalização.

A notação do modelo DMLP é semelhante à do modelo MLP, por exemplo, o modelo DMLP(6,5,1) possui 6 neurônios na camada de entrada, mais 3 neurônios na camada de entrada referentes aos valores defasados da decomposição, 5 neurônios na camada intermediária e 1 neurônio na camada de saída. Nas Figuras 29(a) e 29(b) são mostradas as representações dos modelos utilizados na previsão, respectivamente.

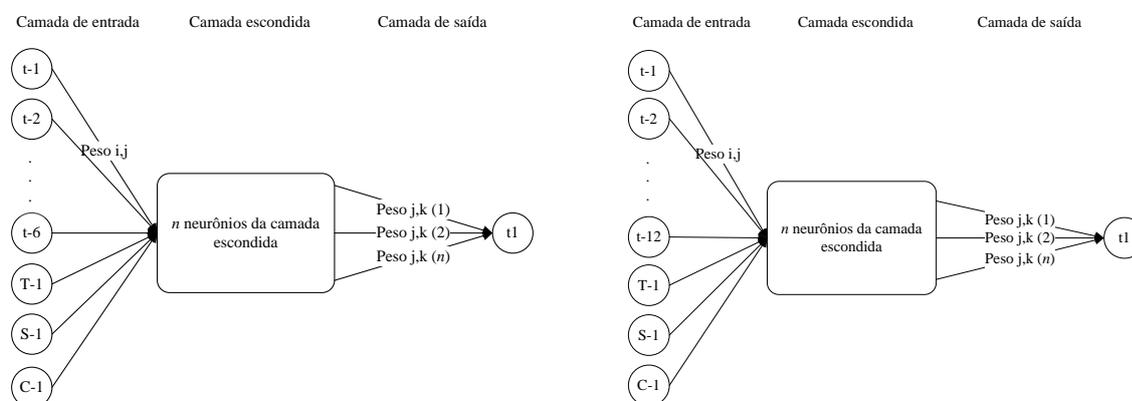


Figura 29 – Representação do modelo DMLP com 6 (a) e com 12 neurônios na camada de entrada (b)

Nas figuras, os neurônios das RNAs representados pelas variáveis $T-I$, $S-I$ e $C-I$ são, respectivamente: os valores defasados das séries de tendência, de sazonalidade e da componente irregular, em um período de tempo. A dessazonalização clássica da série temporal é realizada conforme descrita na seção 2.2.2 e os resultados obtidos por meio desta são mostrados nas Figuras 30(a) a 30(c).

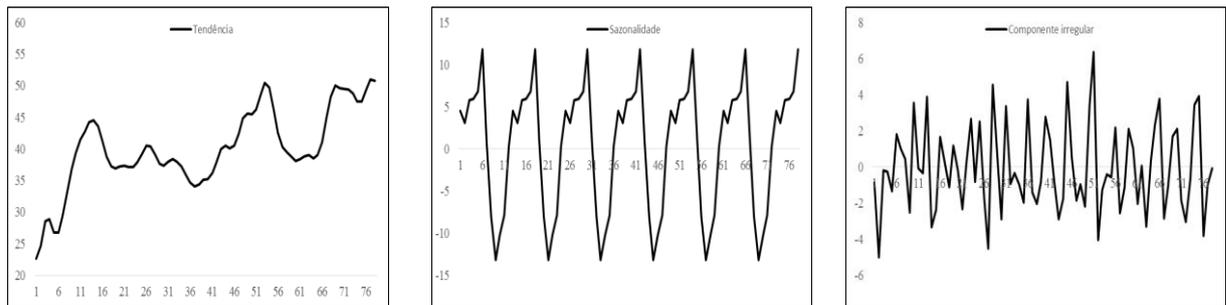


Figura 30 – Gráficos das componentes dessazonalizadas: tendência (a), sazonalidade (b) e componente irregular (c) da série temporal utilizadas como entrada nos modelos DMLP

Com os valores defasados da série e os valores obtidos por meio da dessazonalização, realizou-se o treinamento dos modelos por meio do algoritmo *backpropagation*. Após o treinamento, foram simulados os valores previstos para cada modelo DMLP nos respectivos horizontes de tempo e então contabilizados os respectivos resultados gerados. Na Tabela 6 são mostrados os resultados obtidos.

Na Tabela 6, foi grifado em negrito o modelo DMLP(6,5,1), pois ele foi selecionado para ser combinado com o modelo SARIMA(3,0,1)(1,1,0)₁₂, por apresentar os menores valores dos critérios utilizados para medir o grau de ajustamento e previsão do modelo. Observa-se que em termos de modelagem, o modelo selecionado não foi aquele que obteve o melhor resultado para os critérios.

Da mesma forma que para o MLP(12,30,1), a evidência da não ocorrência de super ajustamento foi confirmada por meio de uma regressão linear simples entre os valores obtidos pelo modelo e os valores alvo apresentados. Nas Figuras 31(a) e 31(b), são mostrados os gráficos de dispersão dos processos de treinamento e teste do modelo DMLP(6,5,1), respectivamente, que demonstra tal situação.

O valor de R^2 obtido no treinamento do modelo foi igual a 0,942 e, para o processo de teste, foi obtido o valor R^2 igual a 0,895. Assim, conclui-se que no treinamento do modelo DMLP(6,5,1) não ocorreu o super ajustamento, indicando que o modelo obtido mostra capacidade de generalização para valores *out-of-sample*.

Tabela 6 – Resultados dos modelos DMLP utilizados para a modelagem e previsão da série temporal da demanda agregada de leite UHT

Modelos DMLP	RMSE			MAPE		
	Model.	h = 6	h = 12	Model.	h = 6	h = 12
DMLP(6,5,1)*	4,06	4,01	4,48	8,41	8,55	8,50
DMLP(6,10,1)	4,22	12,44	10,66	8,32	27,35	21,06
DMLP(6,15,1)	4,45	16,45	12,96	5,39	36,72	25,80
DMLP(6,20,1)	4,18	11,35	9,74	9,61	24,76	19,00
DMLP(6,25,1)	4,58	4,49	5,28	7,24	9,89	9,9
DMLP(6,30,1)	3,99	5,09	4,31	6,43	9,97	8,58
DMLP(6,35,1)	5,85	14,11	11,09	8,16	30,11	21,4
DMLP(6,40,1)	4,60	14,94	11,60	6,96	33,56	21,7
DMLP(6,45,1)	4,17	5,58	8,67	5,32	12,74	15,4
DMLP(6,50,1)	5,11	4,32	6,25	9,51	8,61	11,4
DMLP(12,5,1)	3,38	5,73	5,15	5,43	14,89	9,92
DMLP(12,10,1)	3,88	5,50	5,50	6,75	12,44	11,44
DMLP(12,15,1)	3,83	11,65	10,21	5,80	23,70	19,20
DMLP(12,20,1)	3,31	8,75	8,75	5,87	17,98	16,31
DMLP(12,25,1)	2,43	9,54	12,92	3,36	27,05	24,54
DMLP(12,30,1)	6,48	5,24	5,72	12,76	10,19	10,23
DMLP(12,35,1)	3,24	9,54	8,40	5,0	20,8	16,2
DMLP(12,40,1)	4,44	6,33	8,00	5,0	14,9	15,1
DMLP(12,45,1)	3,17	5,86	6,36	4,2	12,2	10,4
DMLP(12,50,1)	4,79	9,35	10,47	5,6	18,5	19,2

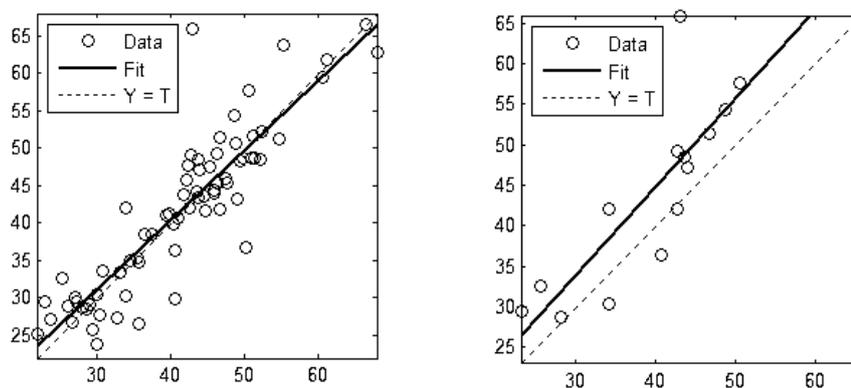


Figura 31 – Gráfico de dispersão dos resultados do processo de treinamento do modelo DMLP (a) e gráfico de dispersão dos resultados do processo de validação do modelo (b)

5.1.4 Etapa 3 – Combinação das previsões dos modelos de Box-Jenkins e MLP/RNA

Com a obtenção dos modelos, pode-se combinar as previsões para a série utilizando os métodos de combinação descritos na seção 3.3. Esta seção apresenta a etapa de combinação das previsões dos modelos SARIMA(3,0,1)(1,1,0)₁₂, MLP(12,30,1) e DMLP(6,5,1).

O primeiro método de combinação foi o da variância mínima e este foi então utilizado para combinar as previsões dos modelos anteriormente citados, obtendo os modelos AR_MLP_VAR, AR_DMLP_VAR e MLP_DMLP_VAR, respectivamente. O primeiro passo para a combinação foi a obtenção dos pesos atribuídos a cada previsão. Na Tabela 7 são mostradas as medidas estimadas para cada combinação, para a obtenção dos pesos.

Tabela 7 – Estimação das medidas e definição do método para o cálculo da combinação da variância mínima

Medidas para a combinação pelo método VAR	AR_MLP_VAR	AR_DMLP_VAR	MLP_DMLP_VAR
Covariância entre res1 e res2	3,87	-2,20	3,52
Variância do res1	10,07	10,07	39,07
Variância do res2	39,07	20,75	20,75
Correlação entre res1 e res2	0,170 (<i>p-value</i> =0,60)	-0,152 (<i>p-value</i> =0,64)	0,135 (<i>p-value</i> =0,89)
Técnica de combinação	Equação 44	Equação 44	Equação 44
Pesos <i>w</i>	0,80	0,67	0,35

Na Tabela 7, observa-se que não houve a ocorrência de correlação significativa entre os resíduos de previsão de cada modelo considerado nas combinações. Portanto, nas três combinações, utilizou-se do método dado pela equação 44 para o cálculo do peso a ser atribuído. Nas equações 65, 66 e 67 são dadas, respectivamente, a combinação das previsões – AR_MLP_VAR, AR_DMLP_VAR e MLP_DMLP_VAR, para um período adiante.

$$\hat{z}_1 = w\hat{x}_1^{(1)} + (1-w)\hat{x}_1^{(2)} = 0,80(49,45) + 0,20(36,32) = 46,76 \quad (65)$$

$$\hat{z}_1 = w\hat{x}_1^{(1)} + (1-w)\hat{x}_1^{(2)} = 0,67(49,45) + 0,33(40,04) = 46,38 \quad (66)$$

$$\hat{z}_1 = w\hat{x}_1^{(1)} + (1-w)\hat{x}_1^{(2)} = 0,35(36,32) + 0,65(40,04) = 38,74 \quad (67)$$

A segunda combinação foi realizada utilizando-se a média aritmética, onde obtiveram-se as combinações AR_MLP_MD, AR_DMLP_MD e MLP_DMLP_MD. Nas equações 68, 69 e 70, é exemplificada a combinação da previsão de cada modelo para um período à frente.

$$\hat{z}_1 = \frac{\hat{x}_1^{(1)} + \hat{x}_1^{(2)}}{2} = \frac{49,45 + 36,32}{2} = 42,88 \quad (68)$$

$$\hat{z}_1 = \frac{\hat{x}_1^{(1)} + \hat{x}_1^{(2)}}{2} = \frac{49,45 + 40,04}{2} = 44,74 \quad (69)$$

$$\hat{z}_1 = \frac{\hat{x}_1^{(1)} + \hat{x}_1^{(2)}}{2} = \frac{36,32 + 40,04}{2} = 38,18 \quad (70)$$

Por meio do cálculo da média geométrica foram obtidas as combinações AR_MLP_MG, AR_DMLP_MG e MLP_DMLP_MG. Nas equações 71, 72 e 73, é exemplificada a combinação da previsão de cada modelo para um período à frente.

$$\hat{z}_1 = \sqrt{\hat{x}_1^{(1)} * \hat{x}_1^{(2)}} = \sqrt{49,45 * 36,32} = 42,38 \quad (71)$$

$$\hat{z}_1 = \sqrt{\hat{x}_1^{(1)} * \hat{x}_1^{(2)}} = \sqrt{49,45 * 40,04} = 44,50 \quad (72)$$

$$\hat{z}_1 = \sqrt{\hat{x}_1^{(1)} * \hat{x}_1^{(2)}} = \sqrt{36,32 * 40,04} = 38,13 \quad (73)$$

Calculando-se a média geométrica ponderada de cada previsão, foram obtidas as combinações AR_MLP_MGP, AR_DMLP_MGP e MLP_DMLP_MGP. Os pesos obtidos pela busca e que satisfazem às restrições foram, respectivamente: 0,7352, 0,3394 e 0,2575; obtendo, respectivamente, os seguintes valores para o MSE: 20,57, 11,54 e 11,62. Nas equações 74, 75 e 76 são dadas as combinações da previsão de um período adiante.

$$\hat{z}_1 = [\hat{x}_1^{(1)}]^w * [\hat{x}_1^{(2)}]^{1-w} = [49,45]^{0,7352} * [36,32]^{0,2648} = 45,57 \quad (74)$$

$$\hat{z}_1 = [\hat{x}_1^{(1)}]^w * [\hat{x}_1^{(2)}]^{1-w} = [49,45]^{0,3394} * [40,04]^{0,6606} = 43,01 \quad (75)$$

$$\hat{z}_1 = [\hat{x}_1^{(1)}]^w * [\hat{x}_1^{(2)}]^{1-w} = [36,32]^{0,2575} * [40,04]^{0,7425} = 39,05 \quad (76)$$

Por meio do cálculo da média harmônica, obtiveram-se as combinações AR_MLP_MH, AR_DMLP_MH e MLP_DMLP_MH. Nas equações 77, 78 e 79 são dadas as combinações das previsões dos respectivos modelos, exemplificando a combinação para um período adiante.

$$\hat{z}_1 = \frac{2\hat{x}_1^{(1)}\hat{x}_1^{(2)}}{\hat{x}_1^{(1)} + \hat{x}_1^{(2)}} = \frac{2(49,45 * 36,32)}{(49,45 + 36,32)} = 41,88 \quad (77)$$

$$\hat{z}_1 = \frac{2\hat{x}_1^{(1)}\hat{x}_1^{(2)}}{\hat{x}_1^{(1)} + \hat{x}_1^{(2)}} = \frac{2(49,45 * 40,04)}{(49,45 + 40,04)} = 44,25 \quad (78)$$

$$\hat{z}_1 = \frac{2\hat{x}_1^{(1)}\hat{x}_1^{(2)}}{\hat{x}_1^{(1)} + \hat{x}_1^{(2)}} = \frac{2(36,32 * 40,04)}{(36,32 + 40,04)} = 38,09 \quad (79)$$

O método seguinte foi o da média harmônica ponderada, onde foram obtidas as combinações AR_MLP_MHP, AR_DMLP_MHP e MLP_DMLP_MHP. Os pesos obtidos pela busca e que satisfazem às restrições foram, respectivamente: 0,7589, 0,3230 e 0,2431; obtendo, respectivamente, os seguintes valores para o MSE: 20,66, 11,84 e 11,55. Nas equações 80, 81 e 82 são dadas as combinações para um período adiante.

$$\hat{z}_1 = \frac{\hat{x}_1^{(1)} \hat{x}_1^{(2)}}{(1-w)\hat{x}_1^{(1)} + w\hat{x}_1^{(2)}} = \frac{(49,45 * 36,32)}{(1-0,7589)49,45 + (0,7589)36,32} = 45,49 \quad (80)$$

$$\hat{z}_1 = \frac{\hat{x}_1^{(1)} \hat{x}_1^{(2)}}{(1-w)\hat{x}_1^{(1)} + w\hat{x}_1^{(2)}} = \frac{(49,45 * 40,04)}{(1-0,3230)49,45 + (0,3230)40,04} = 42,67 \quad (81)$$

$$\hat{z}_1 = \frac{\hat{x}_1^{(1)} \hat{x}_1^{(2)}}{(1-w)\hat{x}_1^{(1)} + w\hat{x}_1^{(2)}} = \frac{(36,32 * 40,04)}{(1-0,2431)36,32 + (0,2431)40,04} = 39,07 \quad (82)$$

A penúltima combinação foi a dos mínimos quadrados, onde obtiveram-se as combinações: AR_MLP_MQ, AR_DMLP_MQ e MLP_DMLP_MQ. Na primeira combinação, os pesos obtidos foram 0,7085 e 0,2935, respectivamente, e o MSE igual a 20,33. Os pesos obtidos para a segunda foram 0,3492 e 0,6488, respectivamente, e o MSE igual a 11,25. Os pesos w_1 e w_2 obtidos para a terceira combinação foram 0,2736 e 0,7283, respectivamente, e o MSE igual a 11,73. Nas equações 83, 84 e 85 é exemplificada a combinação para a previsão de um período adiante.

$$\hat{z}_{t+1} = w_1 \hat{x}_{t+h}^{(1)} + w_2 \hat{x}_{t+h}^{(2)} = 0,7085(49,45) + 0,2935(36,32) = 45,69 \quad (83)$$

$$\hat{z}_{t+1} = w_1 \hat{x}_{t+h}^{(1)} + w_2 \hat{x}_{t+h}^{(2)} = 0,3492(49,45) + 0,6488(40,04) = 43,25 \quad (84)$$

$$\hat{z}_{t+1} = w_1 \hat{x}_{t+h}^{(1)} + w_2 \hat{x}_{t+h}^{(2)} = 0,2736(36,32) + 0,7283(40,04) = 39,10 \quad (85)$$

Para o último método utilizado, o IM, fez-se necessário o cálculo dos pesos de cada uma das previsões e em cada período. O cálculo de ambos os pesos foi realizado por meio do MSE obtido por cada modelo de previsão, em cada período de tempo. O MSE, conforme Stock e Watson (1999), foi calculado por meio de uma média móvel de três período a partir dos resíduos quadráticos de cada previsão.

Na Tabela 8, são mostrados os valores obtidos para os respectivos pesos para exemplificar a combinação das previsões de um período adiante. Obtiveram-se os pesos w_1 e w_2 , por meio das equações 52 e 53 e os parâmetros de entrada destes foram o MSE do modelo ARIMA e o MSE do modelo combinado ao ARIMA – MSE₁ e MSE₂, respectivamente.

Tabela 8 – Estimação das medidas para a combinação das previsões por meio do método IM, exemplificação para a previsão de um período adiante

Medidas para a combinação pelo método IM	AR_MLP_IM	AR_DMLP_IM	MLP_DMLP_IM
MSE ₁	9,28	9,28	40,41
MSE ₂	40,41	20,81	20,81
Peso w_1	0,81	0,69	0,34
Peso w_2	0,19	0,31	0,66

Por meio dos pesos obtidos pelo método de combinação IM, para cada combinação mostrada na Tabela 8, estabeleceram-se as respectivas combinações que, nas equações 86, 87 e 88, são exemplificadas para a previsão de um período adiante.

$$\hat{z}_1 = w_1 \hat{x}_1^{(1)} + w_2 \hat{x}_1^{(2)} = 0,81(49,45) + 0,19(36,32) = 47,00 \quad (86)$$

$$\hat{z}_1 = w_1 \hat{x}_1^{(1)} + w_2 \hat{x}_1^{(2)} = 0,69(49,45) + 0,31(40,04) = 46,55 \quad (87)$$

$$\hat{z}_1 = w_1 \hat{x}_1^{(1)} + w_2 \hat{x}_1^{(2)} = 0,34(36,32) + 0,66(40,04) = 38,77 \quad (88)$$

Obtidas as combinações das previsões dos modelos, pode-se calcular os resultados de cada combinação para o conjunto de modelagem e para o conjunto de previsão e então selecionar a melhor para a previsão da série. Na Tabela 9 são mostrados os resultados.

A combinação de previsões que obteve o melhor resultado foi a AR_DMLP_IM, grifada em negrito na Tabela 9. Embora segundo o RMSE para o horizonte de 6 meses ($h=6$) tal combinação não foi aquela que gerou o melhor resultado, esta foi a escolhida considerando que a distância relativa entre ambos os modelos é maior quando se considera o MAPE para 6 meses, favorável a combinação escolhida. Nas Figuras 32(a) e 32(b) são mostrados os gráficos dos valores previstos pela combinação, além dos valores reais da série.

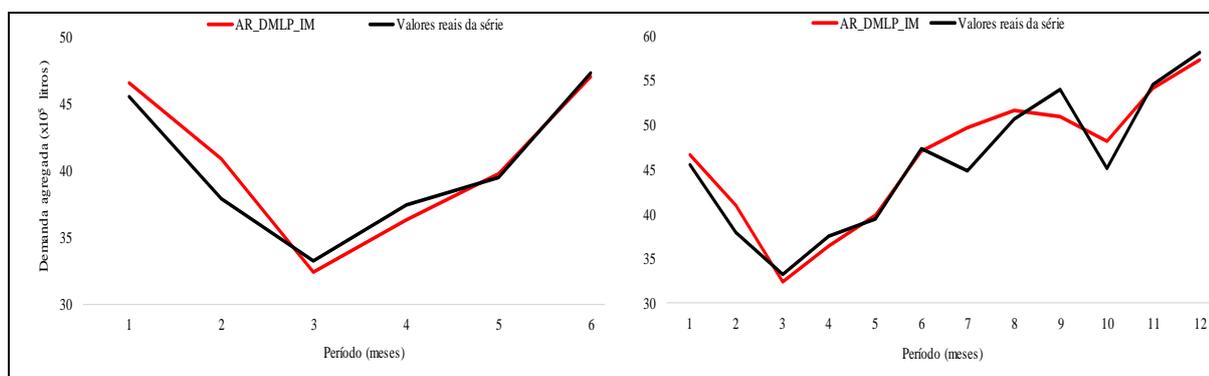


Figura 32 – Gráfico dos valores previstos pela combinação AR_DMLP_IM para 6 meses adiante (a) e dos valores previstos pela combinação para 12 meses adiante (b)

Tabela 9 – Resultados obtidos nas combinações das previsões dos modelos ARIMA, MLP e DMLP

Combinação das previsões dos modelos	RMSE		MAPE	
	h = 6	h = 12	h = 6	h = 12
AR_MLP_VAR	1,82	2,91	4,23	5,39
AR_MLP_MD	1,65	3,60	3,53	6,27
AR_MLP_MG	1,81	2,94	3,81	6,35
AR_MLP_MGP	1,58	2,94	3,33	5,14
AR_MLP_MH	1,98	3,59	4,09	6,42
AR_MLP_MHP	1,61	2,88	3,41	5,04
AR_MLP_MQ	1,53	3,04	3,26	5,32
AR_MLP_IM	1,33*	2,60	3,14	4,55
AR_DMLP_VAR	1,49	2,38	3,01	4,05
AR_DMLP_MD	1,56	2,56	3,27	4,61
AR_DMLP_MG	1,56	3,02	3,37	4,64
AR_DMLP_MGP	2,18	3,02	4,88	5,80
AR_DMLP_MH	1,56	2,56	3,47	4,66
AR_DMLP_MHP	2,28	3,07	5,11	5,93
AR_DMLP_MQ	2,10	2,96	4,69	5,65
AR_DMLP_IM	1,43	2,16*	2,86*	3,70*
MLP_DMLP_VAR	3,40	3,83	5,84	7,02
MLP_DMLP_MD	3,38	4,00	5,80	6,95
MLP_DMLP_MG	3,56	3,86	5,79	7,38
MLP_DMLP_MGP	3,48	3,86	6,35	7,29
MLP_DMLP_MH	3,41	3,95	5,96	6,99
MLP_DMLP_MHP	3,48	3,86	6,35	7,29
MLP_DMLP_MQ	3,47	3,89	6,37	7,35
MLP_DMLP_IM	2,91	2,96	4,20	5,01

5.1.5 Etapa 4 – Comparação dos resultados de previsão

Nesta seção é descrita a comparação realizada entre as combinações das previsões e uma análise do desempenho das combinações em relação aos modelos individuais utilizados. Para tanto, a seção subdivide-se em duas outras: na Seção 5.1.5.1, onde são descritos os passos para a comparação dos resultados das combinações e dos modelos e, na Seção 5.1.5.2, onde é mostrado o desempenho obtido por cada combinação em relação aos modelos de previsão de Box-Jenkins e MLP/RNA.

5.1.5.1 Comparação dos resultados das combinações e dos modelos de previsão

As comparações entre as combinações e os modelos de previsão foram realizadas por meio do teste estatístico não paramétrico de Kruskal-Wallis seguido do teste de comparação múltipla apresentado em Siegel e Castellan (1988). Inicialmente, as combinações e os modelos foram agrupados conforme delineado no Quadro 5, onde foram realizados quatro testes de comparação do RMSE e do MAPE nos horizontes de previsão, denominados: KW1, KW2, KW3 e KW4. A Tabela 10 mostra os quatro testes KW realizados.

Tabela 10 – Testes de Kruskal-Wallis para a comparação do RMSE e do MAPE dos grupos nos horizontes de previsão

Grupo/horizonte	AR	MLP	DMLP	VAR	MD	MG	MGP	MH	MHP	MQ	IM	Teste KW
G1/6	2,84	4,31	4,01	-	-	-	-	-	-	-	-	KW1
G2/6	-	-	-	1,82	1,65	1,81	1,58	1,98	1,61	1,53	1,33	
G3/6	-	-	-	1,49	1,56	1,56	2,18	1,56	2,28	2,10	1,43	
G4/6	-	-	-	3,40	3,38	3,56	3,48	3,41	3,48	3,47	2,91	
G1/12	3,06	6,00	4,48	-	-	-	-	-	-	-	-	KW2
G2/12	-	-	-	2,91	3,60	2,94	2,94	3,59	2,88	3,04	2,60	
G3/12	-	-	-	2,38	2,56	3,02	3,02	2,56	3,07	2,96	2,16	
G4/12	-	-	-	3,83	4,00	3,86	3,86	3,95	3,86	3,89	2,96	
G1/6	6,50	6,66	8,55	-	-	-	-	-	-	-	-	KW3
G2/6	-	-	-	4,23	3,53	3,81	3,33	4,09	3,41	3,26	3,14	
G3/6	-	-	-	3,01	3,27	3,37	4,88	3,47	5,11	4,69	2,86	
G4/6	-	-	-	5,84	5,80	5,79	6,35	5,96	6,35	6,37	4,20	
G1/12	6,19	9,93	8,50	-	-	-	-	-	-	-	-	KW4
G2/12	-	-	-	5,39	6,27	6,35	5,14	6,42	5,04	5,32	4,55	
G3/12	-	-	-	4,05	4,61	4,64	5,80	4,66	5,93	5,65	3,70	
G4/12	-	-	-	7,02	6,95	7,38	7,29	6,99	7,29	7,35	5,01	

O teste KW1 e o KW2 comparou os resultados dos 4 grupos no horizonte de previsão de 6 e de 12 meses, respectivamente, com base no critério RMSE. O teste KW3 e o KW4, comparou os resultados nos horizontes de previsão, com base no critério MAPE.

Foi utilizado o teste de KW para as comparações, tendo em vista que alguns dos valores mostrados na Tabela 10 não provieram de uma distribuição normal. O teste de

aderência utilizado para a verificação de normalidade nos valores foi o de Shapiro-Wilk (SW). Os resultados da estatística SW e o respectivo *p-valor* são mostrados na Tabela 11.

Tabela 11 – Resultados do teste de Shapiro-Wilk para os conjuntos de valores em que se obtiveram o RMSE e o MAPE, nos respectivos horizontes de previsão

Modelos e combinações	RMSE				MAPE			
	SW-6	<i>p-value</i>	SW-12	<i>p-value</i>	SW-6	<i>p-value</i>	SW-12	<i>p-value</i>
SARIMA	-	-	-	-	-	-	0,877**	0,080**
MLP	0,656*	0,002*	0,831*	0,021*	0,789*	0,046*	-	-
DMLP	-	-	0,855*	0,043*	-	-	-	-
AR_MLP_VAR	-	-	1,172*	0,003*	-	-	-	-
AR_MLP_MD	-	-	0,724*	0,001*	-	-	-	-
AR_MLP_MG	0,798**	0,056**	0,735*	0,002*	-	-	-	-
AR_MLP_MGP	-	-	0,759*	0,003*	-	-	-	-
AR_MLP_MH	0,721*	0,010*	0,751*	0,003*	-	-	-	-
AR_MLP_MHP	-	-	0,764*	0,004*	-	-	-	-
AR_MLP_MQ	-	-	0,743*	0,002*	-	-	-	-
AR_MLP_IM	-	-	0,734*	0,009*	-	-	-	-
AR_DMLP_VAR	0,619*	0,001*	0,775*	0,005*	0,813**	0,077**	-	-
AR_DMLP_MD	0,761*	0,025*	0,821*	0,016*	-	-	-	-
AR_DMLP_MG	0,715*	0,009*	0,809*	0,012*	0,757*	0,02*	-	-
AR_DMLP_MGP	-	-	0,839*	0,027*	-	-	-	-
AR_DMLP_MH	0,662*	0,002*	0,796*	0,008*	0,666*	0,003*	-	-
AR_DMLP_MHP	-	-	0,817*	0,015*	-	-	-	-
AR_DMLP_MQ	-	-	0,835*	0,024*	-	-	-	-
AR_DMLP_IM	0,620*	0,001*	0,688*	0,001*	0,798**	0,057**	0,849*	0,036*
MLP_DMLP_VAR	0,726*	0,011*	-	-	-	-	-	-
MLP_DMLP_MD	0,596*	0,001*	0,782*	0,006*	0,790*	0,047*	-	-
MLP_DMLP_MG	0,568*	0,001*	0,768*	0,004*	0,713*	0,008*	0,870**	0,066**
MLP_DMLP_MGP	0,802**	0,061**	-	-	-	-	-	-
MLP_DMLP_MH	0,580*	0,001*	0,787*	0,007*	0,765*	0,028*	-	-
MLP_DMLP_MHP	0,802**	0,061**	-	-	-	-	-	-
MLP_DMLP_MQ	0,780**	0,059**	-	-	-	-	-	-
MLP_DMLP_IM	0,552*	0,001*	0,683*	0,001*	0,764*	0,027*	-	-

* O conjunto de valores pelo qual se calculou o critério não segue normalidade, ao nível de 5%

** O conjunto de valores pelo qual se calculou o critério não segue normalidade, ao nível de 10%

Pode-se verificar na Tabela 11, que para várias combinações e também para os modelos, o conjunto de valores utilizados para o cálculo dos critérios de acurácia RMSE e

MAPE, nos horizontes de previsão de 6 e de 12 meses, ao nível de 5%, concluiu pela não normalidade da distribuição dos valores.

Devido a esta constatação, aplicou-se o teste de comparação não paramétrico de KW. Neste teste, a aceitação da hipótese nula implica na conclusão de que não há diferença significativa entre os resultados dos grupos, obtidos segundo o critério utilizado. A rejeição da hipótese nula, do contrário, implica que pelo menos um dos grupos difere estatisticamente, ao nível de significância estabelecido, dos demais.

Na Tabela 12 estão os resultados obtidos para os quatro testes de KW realizados para os valores mostrados na Tabela 10, e os resultados da comparação múltipla entre os grupos.

Tabela 12 – Resultados do teste de Kruskal-Wallis e do teste de comparação múltipla entre os grupos de modelos e de combinações das previsões

Teste	Valor KW	<i>p-value</i>	Grupos comparados	Diferença calculada	Diferença crítica
KW1	18,98	p<0,01	G1 e G2*	14,83	14,18
			G1 e G3*	14,83	14,18
			G1 e G4	1,83	14,18
			G2 e G3	0,00	10,47
			G2 e G4*	13,00	10,47
			G3 e G4*	13,00	10,47
KW2	15,89	p<0,01	G1 e G2	12,17	14,18
			G1 e G3*	15,04	14,18
			G1 e G4	2,04	14,18
			G2 e G3	2,87	10,47
			G2 e G4	10,12	10,47
			G3 e G4*	13,00	10,47
KW3	18,45	p<0,01	G1 e G2*	17,50	14,18
			G1 e G3*	17,00	14,18
			G1 e G4	6,00	14,18
			G2 e G3	0,50	10,47
			G2 e G4*	11,50	10,47
			G3 e G4*	11,00	10,47
KW4	15,14	p<0,01	G1 e G2	11,17	14,18
			G1 e G3*	15,54	14,18
			G1 e G4	2,54	14,18
			G2 e G3	4,37	10,47
			G2 e G4	8,62	10,47
			G3 e G4*	13,00	10,47

* Diferença significativa, ao nível de significância de 5%, entre os grupos de combinações

Os quatro testes de comparação de KW realizados e mostrados na Tabela 12 indicam que, pelo menos um dos grupos comparados e mensurados pelos critérios RMSE e MAPE, difere de forma estatisticamente significativa em relação aos demais.

Então, foi realizado o teste de comparação múltipla dos grupos para que se conheça qual tipo de combinação de previsões obteve um desempenho estatisticamente superior. Os

grupos que, ao nível de significância de 5% diferiram, foram grifados na Tabela 12. As diferenças observada e a calculada, utilizadas para a decisão se houve ou não diferença significativa entre os grupos, foram obtidas pela equação 63, dada na seção 4.4.

Nas Figuras 33(a) a 33(d) são mostrados os gráficos da resposta média de cada grupo e seus limites inferior e superior para o RMSE e para o MAPE de previsão de 6 meses e de 12 meses, respectivamente, onde se visualiza graficamente as diferenças entre os grupos.

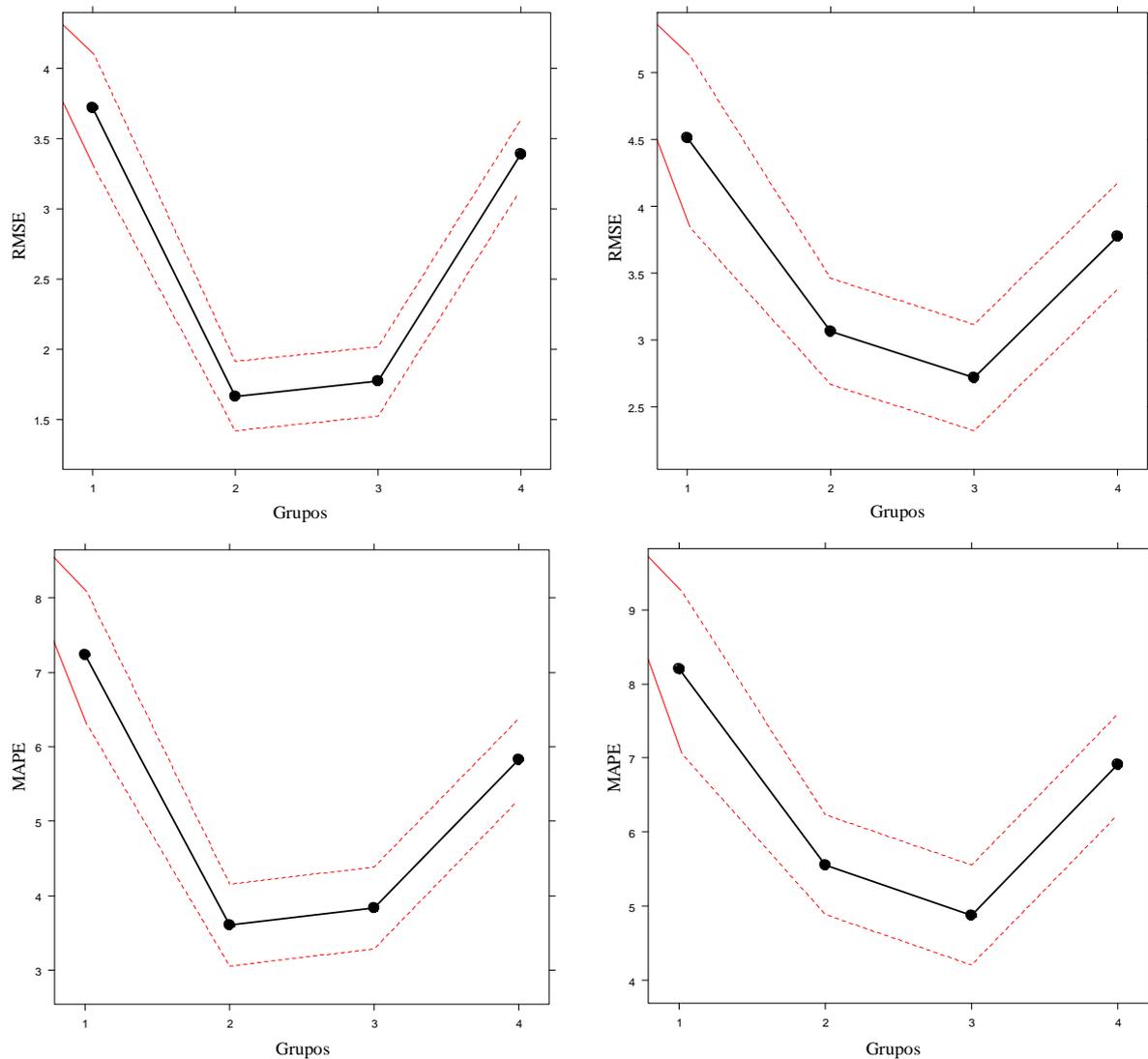


Figura 33 – Gráfico da resposta média de cada grupo, para o RMSE de previsão de 6 e de 12 meses (a,b) e da resposta média de cada grupo para o MAPE de previsão de 6 e 12 meses (c,d)

A partir da Tabela 12 e das Figuras 33(a) a 33(d), pode-se concluir que as combinações das previsões dos modelos MLP e DMLP não melhoraram significativamente o desempenho em relação aos modelos individuais, inclusive o SARIMA. Para o horizonte de previsão de 6 meses, em geral, a combinação dos modelos SARIMA e MLP apresentou o

melhor desempenho, mas a diferença não foi significativa quando comparada ao desempenho da combinação dos modelos SARIMA e DMLP. Para o horizonte de previsão de 12 meses, a combinação dos modelos SARIMA e DMLP apresentou o melhor desempenho, mas que também não é significativo quando comparado a combinação dos modelos SARIMA e MLP.

5.1.5.2 Desempenho das previsões combinadas

Nesta seção, são mostrados os desempenhos obtidos nas combinações em relação aos modelos de previsão utilizados.

Na Tabela 13 são mostrados os desempenhos obtidos com as combinações dos modelos, na comparação aos modelos do grupo 1. Os valores mostrados são expressos em unidades percentuais da variação do resultado da combinação das previsões i (onde $i = 1, 2, 3, \dots, 24$), dado na linha i , em relação ao modelo de previsão j , dado na coluna j (onde, para cada horizonte de previsão, $j = 1, 2$ e 3).

Nas Figuras 34(a) e 34(b), são mostrados os gráficos *box-plot* para o desempenho obtido nos grupos de combinações, em relação às previsões dos modelos individuais, segundo os critérios utilizados. O G2(M), por exemplo, corresponde ao *box-plot* do grupo 2 calculado por meio do critério MAPE, enquanto o G2(R), o do grupo 2 calculado pelo critério RMSE.

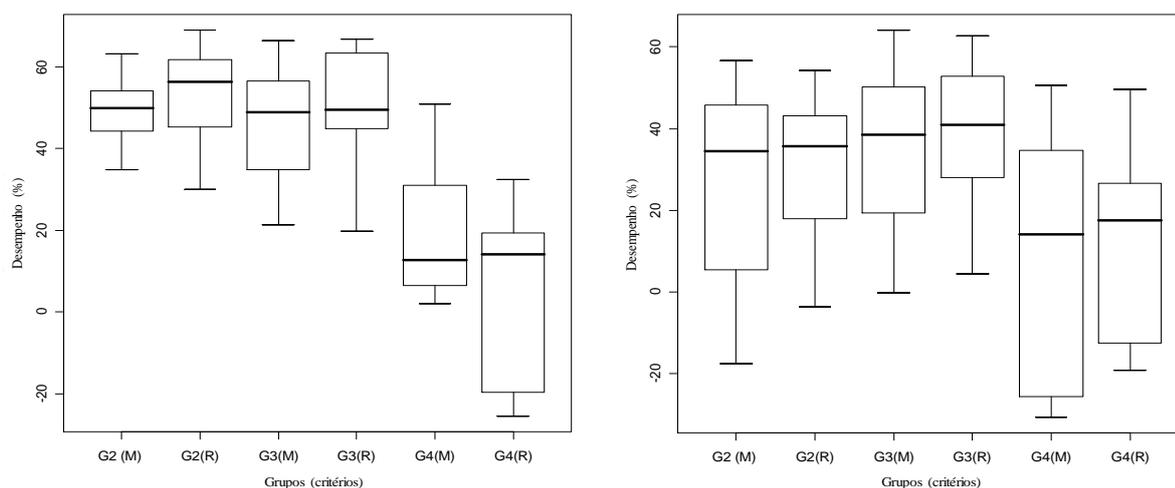


Figura 34 – *Box-plot* do desempenho de cada grupo de combinação, conforme MAPE e RMSE de previsão de 6 meses (a) e para a previsão de 12 meses (b)

Pelo RMSE, o melhor desempenho na previsão 6 meses adiante foi a AR_MLP_IM. No horizonte de previsão de 12 meses, o melhor desempenho foi obtido pela combinação AR_DMLP_IM. A combinação escolhida para realizar a previsão da série, para ambos os

horizontes, foi a AR_DMLP_IM. A diferença média do desempenho desta na previsão para 6 meses, comparado a AR_MLP_IM, conforme o MAPE, é de 4% superior em relação a segunda e, pelo RMSE, a segunda combinação é 2,6% superior em relação a AR_DMLP_IM.

Tabela 13 – Desempenho das combinações em relação aos modelos SARIMA (A), MLP (M) e DMLP (D), considerando o RMSE e o MAPE, para os respectivos horizontes de previsão

Combinação / modelos	RMSE (h=6)			RMSE (h=12)			MAPE (h=6)			MAPE (h=12)		
	A	M	D	A	M	D	A	M	D	A	M	D
AR_MLP_VAR	35,9	57,8	54,7	4,8	51,4	35,0	34,9	36,6	50,5	12,9	45,7	36,5
AR_MLP_MD	41,7	61,7	58,8	-17,6	40,0	19,7	45,7	47,1	58,7	-1,3	36,8	26,2
AR_MLP_MG	36,1	58,0	54,8	4,0	51,0	34,4	41,4	42,8	55,4	-2,5	36,1	25,4
AR_MLP_MGP	44,4	63,4	60,7	4,0	51,0	34,4	48,8	50,0	61,0	17,0	48,2	39,5
AR_MLP_MH	30,1	54,0	50,6	-17,3	40,2	19,9	37,0	38,6	52,1	-3,6	35,4	24,5
AR_MLP_MHP	43,3	62,7	59,9	5,9	52,0	35,8	47,5	48,8	60,1	18,7	49,3	40,7
AR_MLP_MQ	46,1	64,6	61,9	0,8	49,4	32,2	49,9	51,1	61,9	14,1	46,4	37,4
AR_MLP_IM*	52,9	69,0	66,7	15,1	56,7	42,0	51,6	52,8	63,2	26,5	54,2	46,5
AR_DMLP_VAR	47,6	65,5	62,9	22,2	60,3	46,8	53,6	54,8	64,7	34,6	59,2	52,4
AR_DMLP_MD	44,8	63,7	61,0	16,4	57,3	42,9	49,7	50,9	61,7	25,6	53,6	45,8
AR_DMLP_MG	45,1	63,9	61,2	1,5	49,7	32,7	48,2	49,5	60,6	25,2	53,3	45,5
AR_DMLP_MGP	22,9	49,3	45,5	1,5	49,7	32,7	24,9	26,7	42,9	6,4	41,6	31,8
AR_DMLP_MH	44,8	63,7	61,0	16,4	57,3	42,9	46,7	48,0	59,4	24,8	53,1	45,2
AR_DMLP_MHP	19,8	47,2	43,3	-0,3	48,8	31,5	21,4	23,4	40,2	4,3	40,3	30,3
AR_DMLP_MQ	26,0	51,3	47,7	3,4	50,7	34,0	27,9	29,7	45,2	8,7	43,0	33,5
AR_DMLP_IM**	49,6	66,8	64,4	29,6	64,1	51,9	56,0	57,1	66,5	40,3	62,8	56,5
MLP_DMLP_VAR	-20,0	21,1	15,2	-25,2	36,1	14,5	10,1	12,4	31,7	-13,4	29,3	17,4
MLP_DMLP_MD	-19,1	21,7	15,8	-30,7	33,3	10,8	10,8	13,0	32,1	-12,2	30,0	18,3
MLP_DMLP_MG	-25,4	17,5	11,3	-26,0	35,7	14,0	10,9	13,1	32,2	-19,2	25,7	13,2
MLP_MLP_MGP	-22,9	19,2	13,1	-26,0	35,7	14,0	2,3	4,8	25,7	-17,8	26,5	14,2
MLP_MLP_MH	-20,3	20,9	14,9	-29,0	34,2	11,9	8,2	10,5	30,2	-12,9	29,6	17,7
MLP_MLP_MHP	-22,9	19,2	13,1	-26,0	35,7	14,0	2,3	4,8	25,7	-17,8	26,5	14,2
MLP_MLP_MQ	-22,5	19,4	13,4	-26,9	35,2	13,3	2,1	4,5	25,5	-18,7	25,9	13,5
MLP_MLP_IM	-2,6	32,5	27,5	3,3	50,7	34,0	35,4	37,0	50,9	19,1	49,5	41,1

* Combinação que obteve o melhor desempenho na previsão para 6 meses pelo critério RMSE

** Combinação que obteve o melhor desempenho na previsão para 6 meses pelo critério MAPE e na previsão para 12 meses pelo critério RMSE e MAPE, respectivamente

5.2 Discussão dos resultados obtidos

Esta seção tem como objetivo descrever uma discussão acerca dos resultados obtidos com as combinações dos modelos utilizados, por meio de uma análise de seus desempenhos e das vantagens e dificuldades na utilização das combinações. As seções 5.2.1 e 5.2.2, respectivamente, apresentam os assuntos citados.

5.2.1 Desempenho das combinações das previsões

A combinação selecionada para realizar a previsão da série temporal da demanda agregada foi a AR_DMLP_IM, que combinou o modelo SARIMA(3,0,1)(1,1,0)₁₂ e o modelo DMLP(6,5,1) por meio do método *inverse mean square error* (IM). Nas Figuras 35 e 36 são mostradas as previsões por meio desta combinação e dos modelos de previsão do G1, para os horizontes de 6 e de 12 meses adiante, respectivamente.

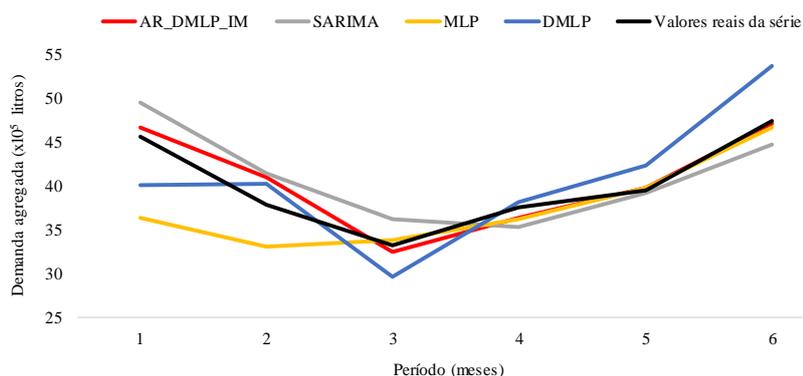


Figura 35 – Gráfico comparativo dos valores previstos pelos modelos do grupo 1 e da combinação AR_DMLP_IM para o horizonte de 6 meses adiante

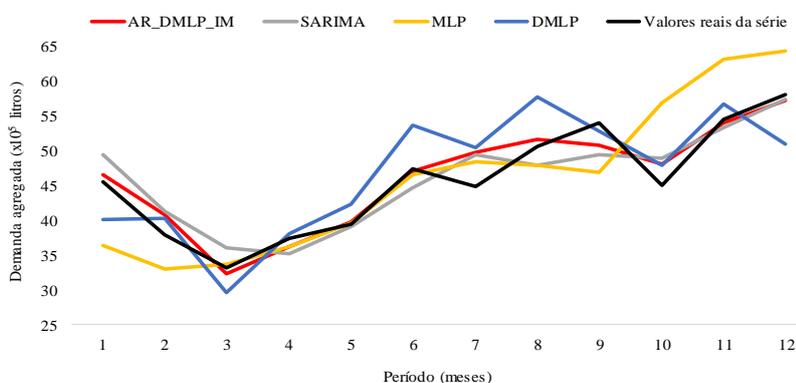


Figura 36 – Gráfico comparativo dos valores previstos pelos modelos do grupo 1 e da combinação AR_DMLP_IM para o horizonte de 12 meses adiante

Pode-se observar nas Figuras 35 e 36, a melhoria significativa na capacidade preditiva em relação aos modelos de previsão utilizados. No horizonte de previsão para 6 meses, a melhoria obtida pela combinação em relação ao modelo SARIMA(3,0,1)(1,1,0)₁₂ foi 49,6% e 56,0%, conforme o RMSE e o MAPE, respectivamente. Em relação ao modelo DMLP(6,5,1), a melhoria obtida foi 64,3% e 66,5%, pelo RMSE e MAPE, respectivamente.

No horizonte de 12 meses, a combinação melhorou a capacidade de previsão em 29,4% e em 40,2%, conforme o RMSE e o MAPE, em comparação ao modelo SARIMA(3,0,1)(1,1,0)₁₂, respectivamente. Em relação ao modelo DMLP(6,5,1), a combinação melhorou a previsão em 51,8% e em 56,5%, conforme o RMSE e o MAPE, respectivamente.

Portanto, com a combinação dos modelos mostrada nas Figuras 35 e 36, verificou-se um aumento na capacidade preditiva em relação a previsão individual de cada modelo, em ambos os horizontes de previsão e para ambas as medidas de acurácia. Pode-se notar que o resultado quando mensurado pelo RMSE, foi inferior na comparação dos resultados em relação à quando mensurado pelo MAPE. Observa-se que a principal melhoria da combinação deu-se em relação ao modelo de RNAs utilizado, e que, portanto, o modelo que mais contribuiu na previsão combinada foi o SARIMA.

Corroborando com a evidência de que foi o modelo SARIMA que mais contribuiu no resultado da previsão combinada, mostra-se na Tabela 14 os pesos w_1 , referentes ao modelo SARIMA, e w_2 , referente ao modelo DMLP, para cada horizonte h de previsão.

Tabela 14 – Pesos da combinação AR_DMLP_IM para os horizontes de previsão de 6 e de 12 meses adiante

Horizonte / Peso	1	2	3	4	5	6	Média (h=6)	7	8	9	10	11	12	Média (h=12)
w_1	0,69	0,58	0,43	0,62	0,79	0,73	0,64	0,77	0,63	0,59	0,28	0,80	0,97	0,66
w_2	0,31	0,42	0,57	0,38	0,21	0,27	0,36	0,23	0,37	0,41	0,72	0,20	0,03	0,34

Pela Tabela 14, verifica-se que o peso médio do modelo SARIMA no horizonte de previsão para 6 meses foi 0,64, indicando que a influência da previsão pelo modelo na combinação foi de aproximadamente 64%. No horizonte de previsão para 12 meses, verificou-se que a influência da previsão pelo modelo SARIMA foi similar ao ocorrido no horizonte de 6 meses, correspondendo a aproximadamente 66% no resultado da combinação. Ainda, pode-se verificar que em apenas dois casos o modelo DMLP apresentou maior

influência do que o SARIMA na previsão combinada, que foram nas previsões dos horizontes 3 e 10. O motivo deste ocorrido pode ser visualizado nas Figuras 35 e 36, onde em ambas as previsões, observa-se que o modelo SARIMA obteve um erro maior.

Quanto ao desempenho das combinações dos modelos no horizonte de previsão para 6 meses, pode-se observar na Tabela 13 que as combinações dos modelos SARIMA e MLP e dos modelos SARIMA e DMLP melhoram significativamente a capacidade preditiva em relação aos modelos individuais que as compõem, por ambas as medidas de acurácia. Também, observa-se que a combinação dos modelos supracitados gera resultados significativamente superior à combinação dos modelos MLP e DMLP. Essa situação evidencia a importância da utilização do modelo SARIMA nas combinações.

As combinações dos modelos SARIMA e MLP pelos métodos utilizados, geraram o resultado médio de 1,66 e de 3,60, segundo o RMSE e o MAPE, respectivamente, que foi o melhor entre os grupos de combinação. O aumento médio na capacidade preditiva em relação aos modelos do G1 foi de 55,3% e 50,2% conforme o RMSE e o MAPE, respectivamente. As combinações destes modelos também obtiveram um resultado médio superior de 6% em comparação às combinações dos modelos SARIMA e DMLP, porém esta melhoria não foi significativa.

Conforme pode-se visualizar na Figura 34(a), as combinações dos modelos MLP e DMLP apresentaram maior variabilidade na previsão para 6 meses, especialmente quando se analisa o RMSE. Além disso, pode ser observado na Tabela 13 que para todas as combinações destes modelos, pelo RMSE, o resultado obtido pelo modelo SARIMA foi superior. Pelo MAPE, observa-se uma situação diferente, onde as combinações obtiveram um desempenho médio 10,3% superior ao modelo SARIMA.

Analisando-se o desempenho das combinações no horizonte de 12 meses, verifica-se na Tabela 13 que apenas as combinações SARIMA e DMLP (G2) melhoram significativamente a capacidade preditiva, por ambas as medidas de acurácia. As combinações dos modelos supracitados geraram resultados significativamente superior à combinação dos modelos MLP e DMLP, diferentemente do horizonte de 6 meses, onde o desempenho das combinações dos modelos SARIMA e MLP também foram significativos.

As combinações do G2 geraram o resultado médio de 2,71 e de 4,88, pelo RMSE e MAPE, respectivamente, o melhor entre os grupos. O aumento médio na capacidade preditiva das combinações em relação aos modelos do G1 foi de 39,9% e 40,5% conforme o RMSE e o MAPE, respectivamente. Em relação às combinações dos modelos SARIMA e MLP, o

resultado médio obtido foi 11,4% e 12,2% superior, conforme RMSE e MAPE, respectivamente. Porém, não foi constatado que esse desempenho tenha sido significativo.

Pode-se observar que nas previsões para o horizonte de 12 meses, conforme mostrado na Figura 34(b), a variabilidade dos desempenhos obtidos nas combinações aumenta. Pelo RMSE, todos os grupos de combinação apresentam, pelo menos, uma combinação que não obteve desempenho superior ao modelo SARIMA, evidenciando que a probabilidade da combinação não melhorar a capacidade preditiva aumenta com o aumento do horizonte de previsão.

5.2.2 Vantagens e dificuldades na utilização das combinações dos modelos

A combinação dos modelos ARIMA com os modelos MLP/RNA apresenta vantagens que a tornam interessante para realizar a previsão de demanda. Porém, a sua utilização também apresenta algumas dificuldades que merecem ser comentadas. Esta seção tem como objetivo apresentar tais vantagens e dificuldades, com vistas à aplicação desta metodologia no contexto das organizações contemporâneas.

Os modelos ARIMA e RNA utilizados nesta dissertação apresentam a vantagem da aplicação em uma ampla gama de séries temporais. Os modelos ARIMA aplicam-se à previsão de séries com sazonalidade, séries com tendência, séries autocorrelacionadas, séries autocorrelacionadas nos resíduos, além da previsão de séries com mais de uma ou todas as situações descritas. Os modelos MLP/RNA, além destas, apresentam a capacidade de modelagem não linear (ZHANG, 2003; KHASHEI; BIJARI, 2010; LEE; TONG, 2011). A utilização conjunta de tais modelos, por meio da combinação, possibilita a captura dos diversos possíveis padrões existentes na série, os tornando interessantes para a previsão de demanda frente a outros modelos no que tange à sua acurácia.

A dificuldade na utilização prática em organizações, destes modelos e de suas combinações para a previsão de demanda ou de séries temporais, é devido, principalmente, ao conhecimento estatístico, matemático e computacional requerido. Portanto, torna-se necessário a presença de especialistas contratados, o que pode gerar um custo maior ao processo de planejamento.

Também, as empresas precisam despender esforços para coletar e armazenar dados de demanda de forma confiável. Muitas empresas atualmente ainda não adotam essa prática, o que torna o processo de previsão para o planejamento das organizações mais complicado de se realizar, tornando a empresa menos competitiva frente aos concorrentes.

6 CONSIDERAÇÕES FINAIS

A previsão de demanda é uma ferramenta importante para o planejamento das organizações, pois obtendo-se previsões e planos mais acurados, estas tornam-se mais competitivas no atual mercado globalizado. A previsão pode antecipar cenários futuros e contribuir na análise das decisões a serem tomadas, de forma a se obterem melhores resultados. Os resultados obtidos com previsões mais acuradas são devidos ao adequado ajustamento nos níveis de recursos necessários para o atendimento da demanda e ao pleno atendimento à demanda. No primeiro caso, o resultado é obtido por meio do incremento em eficiência, pois previsões mais acuradas possibilita ao tomador de decisão, utilizar, de forma mais econômica, os recursos da empresa. No segundo caso, o resultado é obtido por meio de maior eficácia da empresa, pois permite aos gestores antecipar as decisões acerca da capacidade requerida, otimizando o atendimento às oportunidades de mercado.

Para realizar a previsão de demanda, as empresas podem optar por diversas técnicas disponíveis. O que principalmente leva ao direcionamento da técnica de previsão mais adequada é o nível do planejamento e o horizonte de previsão que se deseja obter. Para o planejamento de nível estratégico, modelos subjetivos são mais adequados, tendo em vista o horizonte de previsão ser mais amplo. Para os planejamentos de nível tático e operacional, recomenda-se a utilização de modelos quantitativos, pois grande parte destes modelos são capazes de gerar previsões mais acuradas para curtos/médios horizontes de tempo.

Esta dissertação apresentou um estudo da previsão de uma série temporal da demanda agregada de leite UHT em uma empresa de laticínios, por meio da combinação de modelos. Foram escolhidos os modelos de Box-Jenkins e MLP/RNA como previsores na combinação, considerando a flexibilidade na captura dos diversos possíveis padrões existentes nas séries temporais. Embora seja possível a utilização de diversos modelos, recomenda-se estes como previsores, tendo em vista a sua utilização já consolidada na literatura da área.

A melhor combinação para a previsão da série, e que apresentou significativa melhoria, foi a dos modelos SARIMA(3,0,1)(1,1,0)₁₂ e DMLP(6,5,1). No horizonte de 6 meses, a combinação obteve um desempenho de até 56,0% melhor que o SARIMA e até 66,5% melhor que o DMLP, que foram os modelos utilizados na combinação. No horizonte de 12 meses, o desempenho da combinação foi de até 40,2% e até 56,5% superior, respectivamente.

Com a realização do estudo, pode-se verificar a importância de se utilizar modelos com distintas capacidades de captura de padrões nas séries. Nas combinações dos modelos MLP e DMLP, verificou-se que em aproximadamente 23% dos resultados obtidos, estas não melhoraram a capacidade de previsão em relação ao modelo SARIMA, comprovando a importância de sua inclusão nas combinações.

A realização deste estudo permitiu verificar o ganho obtido com as combinações de modelos para a previsão de demanda no planejamento da produção. As técnicas de combinação são simples de serem utilizadas e, em geral, melhoram a capacidade preditiva da série. Porém, devem-se utilizar bons modelos como previsores, onde para a obtenção de maior acurácia na previsão, tendo em vista os diversos padrões existentes nas séries, recomenda-se a utilização dos modelos de Box-Jenkins e os modelos MLP/RNA.

Como sugestão de pesquisa futura, outra forma de se realizar a combinação de modelos é por meio da hibridização. Neste tipo de combinação, utiliza-se um dos modelos na modelagem da série e então, a partir dos parâmetros e resultados obtidos, realiza-se a modelagem por meio do segundo. Exemplos podem ser vistos em Zhang (2003), Khashei e Bijari (2010) e em Lee e Tong (2011). Em Zhang (2003), o autor realiza a modelagem da série utilizando um modelo ARIMA e, a partir dos parâmetros p e q , utiliza um modelo MLP/RNA com p valores defasados da série e com q valores defasados do erro obtido pelo modelo ARIMA, para prever os valores futuros de uma série temporal.

Considerando o crescente número de pesquisas na área de previsão de séries temporais por meio de modelos híbridos, sugere-se a realização de um estudo com o objetivo de comparar o desempenho dessa classe de modelos na previsão de demanda, em relação aos tradicionais métodos de combinação estudados na literatura, a exemplo dos métodos utilizados nesta dissertação.

Sugere-se também, a utilização de outros modelos para a realização da previsão de demanda. Seria importante verificar, se modelos mais simples que os previsores utilizados na presente dissertação, por meio dos métodos de combinação abordados, não gerariam resultados próximos daqueles obtidos, analisando-se conjuntamente o custo *versus* o benefício obtido ao se utilizarem técnicas preditivas mais robustas, porém mais complexas, como a modelagem ARIMA e a MLP/RNA, na combinação de previsões para a previsão da demanda no planejamento da produção.

REFERÊNCIAS

- ABURTO, L.; WEBER, R. Improved supply chain management based on hybrid. **Applied Soft Computing**, Chile, n. 7, p. 136-144, 2007.
- ALLEMÃO, M. A. F. **Redes Neurais Aplicadas à Previsão de Demanda de Numerário em Agências Bancárias**. 2004. 247 p. Dissertação (Mestrado em Ciência da Computação) - Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2004.
- ANDRAWIS, R. R.; ATIYA, A. F.; EL-SHISHINY, H. Combination of long term and short term forecasts, with application to tourism demand forecasting. **International Journal of Forecasting**, n. 27, p. 870-886, 2011.
- BALLOU, R. H. **Gerenciamento da Cadeia de Suprimentos/Logística Empresarial**. 3 ed. Porto Alegre: Ed: Bookman, 2001, 465 p.
- BATES, J. M.; GRANGER, C. W. J. The combination of forecasts. **Operational Research Quarterly**, v. 20, n. 4, 1969.
- BOX, G. E. P.; JENKINS, G. M. **Time series analysis**. San Francisco: Ed: Holden-Day, 1976, 575 p.
- BOX, G. E. P.; JENKINS, G. M.; REINSEL, G. C. **Time series analysis: forecasting and control**. 3rd ed. New Jersey: Ed: Wiley, 2008, 784 p.
- BRAGA, A. P.; CARVALHO, A. P. L. F.; LUDERMIR, T. B. **Redes Neurais Artificiais: Teoria e Aplicações**. 2ª ed. Rio de Janeiro: Ed. LTC – Livros Técnicos e Científicos, 2011, 226 p.
- BURBIDGE, J. L.; CRUZ, L. H. S. **Planejamento e controle da produção**. 4ª ed. São Paulo: Ed. Atlas, 1981, 556 p.
- CHAN, W. S.; TONG, H. On tests for non-linearity in time series analysis. **Journal of Forecasting**, n.5, p.217-228, 1986.
- CHEN, H.; DING, K.; ZHANG, J. Properties of weighted geometric means combination forecasting model based on degree of logarithm grey incidence. **Proceedings of the IEEE international conference on grey systems and intelligent services**, p.673-677, 2007.
- CHOI, T. M.; YU, Y.; AU, K. F. A hybrid SARIMA wavelet transform method for sales forecasting. **Decision Support Systems**, Hong Kong, n. 51, p.130-140, 2011.
- CLEMEN, R. T. Combining forecasts: a review and annotated bibliography. **International Journal of Forecasting**, n. 5, p. 559-583, 1989.
- CORRÊA, H. L.; CORREA, C. A. **Administração de produção e operações: manufatura e serviços: uma abordagem estratégica**. São Paulo: Ed: Atlas, 2004, 690 p.

CORRÊA, H. L.; GIANESI, I. G. N.; CAON, M. **Planejamento, Programação e Controle da Produção**. 5ª ed. São Paulo: Ed. Atlas, 2011, 434 p.

COSTANTINI, M.; PAPPALARDO, C. A hierarchical procedure for the combination of forecasts. **International Journal of Forecasting**, n. 26, p. 725-743, 2010.

DAVIS, M. M.; AQUILANO, N. J.; CHASE, R. B. **Fundamentos da Administração da Produção**. Porto Alegre: Ed: Bookman, 2001, 598 p.

DOANE, D. P.; SEWARD, L. E. **Applied Statistics in Business and Economics**. 4ª ed. Ed. McGraw-Hill Irwin, 2013, 834 p.

DONALDSON, R. G.; KAMSTRA, M. Forecast combining with neural networks. **Journal of Forecasting**, n. 15, p.49-61, 1996.

ENDERS, W. **Applied Econometric Time Series**. 2nd ed. Ed. Wiley Series In Probability And Statistics, 2004, 466 p.

ENGLE, R. F. Autoregressive conditional heteroscedasticity with estimates of the variance of U.K. inflation. **Econometrica**, n. 50, p. 987-1008, 1982.

ERDOGDU, E. Electricity demand analysis using cointegration and ARIMA modelling: a case study of Turkey. **Energy Policy**, n. 35, p. 1129-1146, 2007.

FARUK, D. O. A hybrid neural network and ARIMA model for water quality time series prediction. **Engineering Applications of Artificial Intelligence**, Turquia, n. 23, p.586-594, 2010.

FAVA, V. L.; **Análise de Séries de Tempo**. In: VASCONCELLOS, M. A. S.; ALVES, D. **Manual de econometria: nível intermediário**. São Paulo: Ed. Atlas, 2000.

FERNANDES, F. C. F.; GODINHO FILHO, M. **Planejamento e controle da produção: dos fundamentos ao essencial**. 2ª ed. São Paulo: Ed. Atlas, 2010, 275 p.

FLORES, J. H. F. **Comparação de modelos MLP/RNA e modelos Box-Jenkins em séries temporais não lineares**. 2009. 133 p. Dissertação (Mestrado em Engenharia de Produção) – Universidade Federal do Rio Grande do Sul, Porto Alegre, 2009.

GAITHER, N.; FRAZIER, G. **Administração da Produção e Operações**. 8ª ed. São Paulo: Ed. Cengage Learning, 2002, 598 p.

GHEYAS, I. A.; SMITH, L. S. A novel neural network ensemble architecture for time series forecasting. **Neurocomputing**, n. 74, p. 3855-3864, 2011.

GRANGER, C. W. J.; ANDERSON, A. P. **An Introduction to Bilinear Time Series Models**. Vandenhoeck and Ruprecht, Göttingen, 1978.

GRANGER, C. W. J.; RAMANATHAN, R. Improved methods of combining forecasts. **Journal of Forecasting**. n.3, p. 197-204, 1984.

GUJARATI, D. M. **Econometria básica**. 3ª Ed. São Paulo/SP: Ed. Makron books, 2000, 846 p.

GURESEN, E.; KAYAKUTLU, G.; DAIM, T. U. Using artificial neural network models in stock market index prediction. **Expert Systems with Applications**, n. 38, p. 10389-10397, 2011.

HAYKIN, S. **Redes Neurais: Princípios e práticas**. 2 ed. Porto Alegre: Ed. Artmed S.A., 1999, 903 p.

HO, S. L.; XIE, M.; GOH, T. N. A comparative study of neural network and Box-Jenkins ARIMA modeling in time series prediction. **Computers & Industrial Engineering**, Singapore, n. 42, p. 371-375, 2002.

HO, S. L.; XIE, M. The use of ARIMA models for reliability forecasting and analysis. **Computers & Industrial Engineering**, Singapore, v. 35, n. 1-2, p. 213-216, 1998.

JACOBS, F. R.; CHASE, R. B.; SOUZA, T. C. F. **Administração da Produção e Operações: o essencial**. 3ª ed. Porto Alegre: Ed. Bookman, 2009, 424 p.

JUNTILLA, J. Structural breaks, ARIMA model and Finnish inflation forecasts. **International Journal of Forecasting**, n. 17, p. 203-230, 2001.

KHASHEI, M; BIJARI, M. An artificial neural network (p,d,q) model for timeseries forecasting. **Expert Systems with applications**, Iran, n. 37, p. 479-489, 2010.

KOTLER, P.; KELLER, K. L. **Administração de marketing**. 14ª ed. São Paulo: Ed. Pearson Education do Brasil, 2012, 765 p.

KOVÁCS, Z. L. **Redes Neurais Artificiais: Fundamentos e Aplicações**. 4ª ed. São Paulo: Ed. Livraria da Física, 2006, 177 p.

KUMAR, K.; JAIN, V. K. Autoregressive integrated moving averages (ARIMA) modelling of a traffic noise time series. **Applied Acoustics**, n. 58, p. 283-294, 1999.

LEE, Y. S.; TONG, L. I. Forecasting time series using a methodology based on autoregressive integrated moving average and genetic programming. **Knowledge-Based Systems**, Taiwan, n. 24, p. 66-72, 2011.

LEVINE, D. M. et al. **Estatística: Teoria e aplicações usando o Microsoft Excel**. 3ª ed. Rio de Janeiro/RJ: Ed. LTC, 2005, 819 p.

LUSTOSA, L.; NANJI, L.C. **Planejamento agregado e planejamento mestre da produção**. In: LUSTOSA, L.; MESQUITA, M. A.; QUELHAS, O.; OLIVEIRA, R. **Planejamento e Controle da Produção**. Rio de Janeiro: Elsevier, 2008.

- MADDALA, G. S. **Introdução à Econometria**. 3 ed. Rio de Janeiro: Ed: LTC – Livros Técnicos e Científicos, 2003, 345 p.
- MAKRIDAKIS, S.; ANDERSEN, A.; CARBONE, R.; FILDES, R.; HIBON, M.; LEVANDOWSKI, R.; NEWTON, J.; PARZEN, E.; WINKLER, R. The accuracy of extrapolation (time series) methods: results of a forecasting competition. **Journal of Forecasting**, n. 1, p.111-153, 1982.
- MAKRIDAKIS, S.; HIBON, M. The M3-competition: results, conclusions and implications. **International Journal of Forecasting**, n. 16, p.451-476, 2000.
- MAKRIDAKIS, S.; WHEELWRIGHT, S. C.; HYNDMAN, R. J. **Forecasting: Methods and Applications**. 3rd ed. Ed. Wiley & Sons, 1998, 642 p.
- MARTINS, V. L. M.; WERNER, L. Forecast combination in industrial series: A comparison between individual forecasts and its combinations with and without correlated erros. **Expert Systems with Applications**, n. 39, p. 11479-11486, 2012.
- MATOS, O. C. M. **Econometria básica: teoria e aplicações**. 3ª ed. São Paulo/SP: Ed. Atlas, 2000, 300 p.
- MENEZES, L. M.; BUNN, D. W.; TAYLOR, J. M. Review of guidelines for the use of combined forecasts. **European Journal of Operational Research**, n. 120, p. 190-204, 2000.
- MESQUITA, M. A. **Previsão de demanda**. In: LUSTOSA, L.; MESQUITA, M. A.; QUELHAS, O.; OLIVEIRA, R. **Planejamento e Controle da Produção**. Rio de Janeiro: Elsevier, 2008.
- MOREIRA, D. A. **Administração da Produção e Operações**. São Paulo: Ed. Pioneira, 2008, 619 p.
- MORETTIN, P. A. **Econometria Financeira: Um curso em séries temporais financeiras**. 1ª ed. São Paulo/SP: Ed: Blucher, 2008, 319 p.
- MORETTIN, P. A.; TOLOI, C. M. C. **Análise de séries temporais**. 2ª ed. São Paulo/SP: Ed. Edgard Blüncher, 2006, 531 p.
- OPERADOR NACIONAL DO SISTEMA ELÉTRICO. **Séries Históricas de Vazões**. Brasília, 2010. Disponível em: <http://www.ons.org.br/operacao/vazoes_naturais.aspx>. Acesso em: 20/04/2014.
- PATTON, A. J.; SHEPPARD, K. Optimal combinations of realised volatility estimators. **International Journal of Forecasting**, vol. 25, p. 218-238, 2009.
- PELLEGRINI, F. R.; FOGLIATTO, F.S. Passos para implantação de sistemas de previsão de demanda: técnicas e estudo de caso. **Revista Produção**, vol. 11, n. 1, p. 43-64, 2001.
- PELLEGRINI, F. R. **Metodologia Para Implementação de Sistemas de Previsão de Demanda**. 2001. 146 p. Dissertação (Mestrado em Engenharia) – Universidade Federal do Rio Grande do Sul, Porto Alegre, 2000.

PINDYCK, R. S.; RUBINFELD, D. L. **Econometric models & economic forecasts**. 2nd ed. Ed. McGraw-Hill Book Company, 1981, 630 p.

PINDYCK, R. S.; RUBINFELD, D. L. **Econometria: Modelos e Previsões**. 4ª ed. Rio de Janeiro/RJ: Ed. Elsevier, 2004, 726 p.

QI, M.; ZHANG, G. P. An investigation of model selection criteria for neural network time series forecasting. **European Journal of Operational Reserach**, n. 132, p. 666-680, 2001.

R CORE TEAM. **R: A language and environment for statistical computing**. Vienna, 2014. R Foundation for Statistical Computing.

RIGGS, J. L.; QUADROS, E.; KURILO, U. **Administração da Produção: planejamento, análise e controle**. 1ª ed. São Paulo: Ed. Atlas, 1981, 457 p.

RITZMAN, L. P.; KRAJEWSKI, L. J. **Administração da Produção e Operações**. 3ª ed. São Paulo: Ed. Pearson Prentice Hall, 2004, 431 p.

RUSSOMANO, V. H. **PCP: planejamento e controle da produção**. 6ª ed. São Paulo: Ed. Pioneira, 2000, 320 p.

SANDERS, N. R.; MANRODT, K. Forecasting Practices in US Corporations: Survey Results. **Interfaces**, v. 24, n. 2, p. 92-100, 1994.

SIEGEL, S.; CASTELLAN, J. **Nonparametric statistics for the behavioral sciences**. 3rd ed. Ed. McGraw-Hill, 1988, 312 p.

SILVA, R. B. **Previsão de demanda no setor de suplementação animal usando combinação e ajuste de previsões**. 2014. 64 p. Dissertação (Mestrado em Engenharia de Produção) – Universidade Federal do Rio Grande do Sul, Porto Alegre, 2014.

SLACK, N.; CHAMBERS, S.; JOHNSTON, R. **Administração da Produção**. 3ª ed. São Paulo: Ed. Atlas, 2009, 703 p.

STOCK, J. H.; WATSON, M. Combination forecasts of output growth in a seven-country data set. **Journal of Forecasting**, n. 23, p.405-430, 2004.

SU, C. T.; TONG, C. M.; LEOU, C. M. Combination of time series and neural network for reliability forecasting modeling. **Journal of the Chinese Institute of Industrial Engineers**, v. 4, n. 14, p. 419-429, 1997.

TSENG, F. M.; YU, H. C.; TZENG, G. H. Combining neural network model with seasonal time series ARIMA model. **Technological Forecasting and Social Change**, Taiwan, n. 69, p.71-87, 2002.

TUBINO, D. F. **Manual de Planejamento e Controle da Produção**. 2ª ed. São Paulo/SP: Ed. Atlas, 2000, 220 p.

TONG, H.; LIM, K. S. Threshold autoregressive, limit cycles and cyclical data. **Journal of the Royal Statistical Society**, v. 42, n. 3, p. 245-292, 1980.

VALIPOUR, M.; BANIHABID, M. E.; BEHBAHANI, S. M. R. Comparison of the ARMA, ARIMA, and the autoregressive artificial neural network models in forecasting the monthly inflow of Dez dam reservoir. **Journal of Hydrology**, n. 476, p. 433-441, 2013.

VOLLMANN, T. E. et al. **Sistemas de Planejamento & Controle da Produção para o Gerenciamento da Cadeia de Suprimentos**. 5ª ed. Porto Alegre/RS: Ed. Bookman, 2006, 647 p.

WANG, C. C. A comparison study between fuzzy time series model and ARIMA model for forecasting Taiwan export. **Expert Systems with Applications**, n. 38, p. 9.296-9.304, 2011.

WERNER, L.; RIBEIRO, J.L.D. Previsão de demanda: uma aplicação dos modelos Box-Jenkins na área de assistência técnica de computadores pessoais. **Revista Gestão & Produção**. v. 10, n. 1, p.47-67, 2003.

WINKLER, R. L; MAKRIDAKIS, S. The combination of forecasts. **Journal of the Royal Statistical Society Series A**, n. 146, p. 150-157, 1983.

WONG, W. K.; GUO, Z. X. A hybrid intelligent model for medium-term sales forecasting in fashion retail supply chains using extreme learning machine and harmony search algorithm. **International Journal of Production Economics**, n. 128, p. 614-624, 2010.

YADAV, V.; SRINIVASAN, D. A SOM-based hybrid linear-neural model for short-term load forecasting. **Neurocomputing**, n.74, p. 2874-2885, 2011.

ZHANG, G. P; PATUWO, B. E.; HU, M. Y. A simulation study of artificial neural networks for nonlinear time-series forecasting. **Computers & Operations Research** , n. 28, p. 381-396, 2001.

ZHANG, G. P.; QI, M. Neural network forecasting for seasonal and trend time series. **European Journal of Operational Research**, n. 160, p. 501-514, 2005.

ZHANG, G. P. Time series forecasting using a hybrid ARIMA and neural network model. **Neurocomputing** , [s.l], n. 50, p. 159-175, 2003.

ZOU, H. F.; XIA, G. P.; YANG, F. T.; WANG, H. Y. An investigation and comparison of artificial neural network and time series models for Chinese food grain price forecasting. **Neurocomputing** , China, n.70, p. 2913-2926, 2007.