

Luiz Albino Teixeira Júnior

**Combinação Linear de Previsões
com Ajuste MINIMAX**

Monografia de Pós-graduação *Lato sensu*

Instituto de Ciências Exatas
Programa de Pós-graduação *lato sensu* em
Métodos Estatísticos Computacionais

Juiz de Fora
Agosto de 2011

Luiz Albino Teixeira Júnior

**Combinação Linear de Previsões
com Ajuste MINIMAX**

Monografia de Pós-graduação *Lato sensu*

Monografia apresentada como requisito parcial para obtenção do grau de pós-graduado *lato sensu* pelo Programa de Pós-graduação *Lato Sensu* em Métodos Estatísticos Computacionais da Universidade Federal de Juiz de Fora.

Orientador: Prof. Reinaldo Castro Souza

Juiz de Fora
Agosto de 2011

Agradecimentos

Agradeço a Deus primeiramente, pois é o maior responsável por chegar até aqui; à minha mãe; à minha tia Irani, à minha esposa; bem como toda a minha família.

Quero expressar também minha gratidão àqueles que colaboraram diretamente no meu curso: meu professor orientador e amigo Reinaldo Castro e a todos os professores do Departamento de Estatística da UFJF.

Resumo

Teixeira Júnior, Luiz Albino; Souza, Reinaldo Castro (Orientador). **Combinação Linear de Previsões com Ajuste MINIMAX**. Juiz de Fora, 2011. 42p. Monografia de Pós-graduação *Lato Sensu* - Instituto de Ciências Exatas, Universidade Federal de Juiz de Fora.

Esta monografia propõe a combinação linear das previsões obtidas por três métodos preditivos de séries temporais (ARIMA, Amortecimento Exponencial e Redes Neurais Artificiais) com pesos adaptativos determinados por meio de um problema de programação multiobjetivo em que se busca minimizar, simultaneamente, as estatísticas: MAE, MAPE e MSE. Os resultados obtidos são comparados com os métodos individuais.

Palavras-Chave

Previsão, Combinação de Previsões, Programação Multiobjetivo.

Abstract

Teixeira Júnior, Luiz Albino; Souza, Reinaldo Castro (Orientador). **Combinação Linear de Previsões com Ajuste MINIMAX**. Juiz de Fora, 2011. 42p. Monograph of Postgraduate *Lato Sensu* - Instituto de Ciências Exatas, Universidade Federal de Juiz de Fora.

This monograph proposes a linear combination of three methods of predicting time series (ARIMA, Exponential Smoothing and Artificial Neural Networks) with adaptive weights determined via a multi objective programming problem in which one seek to minimize, simultaneously, the statistics: MAE, MAPE and MSE. The results are compared with individual methods.

Keywords

Forecasting, Forecasting Combination, Multi Objective Programming.

Sumário

1. Introdução	7
1.1 Contextualização e Definição da Pesquisa	8
2. Modelos Individuais	
2.1. Conceitos de Processos Estocásticos, Séries Temporais e Previsão	8
2.2. Métodos de Amortecimento Exponencial	10
2.2.1. Método Multiplicativo de Amortecimento de Holt-Winters	10
2.3. Modelos de BOX & JENKINS	11
2.3.5. Modelo SARIMA	12
2.4. Método de Redes Neurais Artificiais	13
3. Combinação de Métodos	15
3.1. Combinação Linear de Previsões	16
3.2. Intervalo de Confiança do Método de Combinação Linear de Previsões	18
4. Programação Matemática	21
4.1 Programação MINIMAX	23
4.2 Programação por Metas	23
4.3 Programação Multiobjetivo	24
4.4 Programação de Metas via Desvio Mínimo	26
5 Aplicação à Série de Consumo Residencial de Energia Elétrica	26
5.1 Aplicação do Modelo ARIMA	27
5.1.2 Modelagem	27
5.2 Aplicação do Modelo Holt-Winters	28
5.2.1 Modelagem	29
5.2.2 Previsões	29
5.3 Aplicação do Método de Redes Neurais Artificiais	29
5.4 Otimização do Método de Combinação Multicritério	30
5.5 Comparação dos Métodos	33
5.5.1 Comparação dos Intervalos de Confiança dos Modelos Combinados	37
6 Conclusões	40

1

Introdução

Em países em desenvolvimento, como é o caso do Brasil, o crescimento do consumo de energia elétrica pode ser considerado uma variável relevante para a economia de um modo geral. Assim, a crescente demanda de energia requer uma correta projeção de seus níveis de consumo, sendo necessário, com isso, o estudo das condições os determinam.

A série temporal do consumo agregado de energia elétrica é geralmente segmentada em classes de consumo: comercial, industrial, residencial - sendo a última, a série temporal utilizada nessa pesquisa. Para a projeção dos níveis de consumo, são utilizadas séries históricas e informações (objetivas e subjetivas) sobre os micro e macro ambientes, bem como as de outras variáveis que podem impactar o consumo.

Assim, tem-se que a projeção do consumo residencial é relevante ao setor elétrico, desde o planejamento e controle, até a execução das demais ações, tais como: nível de investimento em infra-estrutura, adequação dos graus de necessidades de capital, gestão dos níveis do reservatório, estoques, capacidade. Por isso, a necessidade de métodos preditivos cada vez mais eficientes, minimizando a incerteza.

Quanto aos métodos estatísticos preditivos, existem três classes básicas: univariados (Amortecimento Exponencial e Box & Jenkins); causais (Função de Transferência e Regressão Dinâmica) e os multivariados (Vetores Auto-Regressivos). Outra opção é a utilização dos métodos inteligentes, como, por exemplo, Redes Neurais Artificiais (RNA), Lógica *Fuzzy* e Algoritmo Genético.

Existem inúmeras abordagens para projeção das séries de tempo, desde a utilização de modelos individuais até a combinação destes. A combinação linear de previsões, proposta por GRANGER e BATES (1969), é uma metodologia alternativa à modelagem de séries temporais. Atualmente, no setor elétrico, geralmente não é utilizada.

Os métodos de combinação linear possuem pesos (lineares) adaptativos associados às previsões dos métodos individuais que são estimados, em sua grande maioria, otimizando apenas uma função objetivo. O método proposto utiliza múltiplas funções objetivo, buscando maior precisão da previsão combinada fora da amostra.

Em suma, o método proposto consiste em calcular o valor esperado das previsões univariadas utilizando pesos adaptativos estimados via Programação Matemática não Linear Multicritério MINIMAX, com programação por Metas, tendo como funções objetivo (individuais): MAPE (*mean absolute percentual error*), MSE (*mean square error*) e MAE (*mean absolute error*). As estimativas de pesos adaptativos são utilizadas para a combinação dos cenários futuros dos métodos individuais, de modo a obter os intervalos de confiança do modelo proposto.

1.2

Objetivo

O objetivo da pesquisa é propor uma abordagem alternativa quanto à combinação linear de métodos preditivos, utilizando o arcabouço conceitual dos modelos de séries de tempo integrados a conceitos de Programação Matemática, com o uso do software AIMMS. A utilização de métodos numéricos na otimização de vários objetivos facilita a estimação dos pesos, não necessitando da abordagem algébrica para obtenção de estimadores dos pesos.

Além disso, buscam-se aprimorar os métodos vigentes na literatura de séries temporais contribuindo de forma a tornar mais eficientes suas projeções. Pequenos ganhos de em termos de precisão podem contribuir de maneira significativa em inúmeras aplicações, como nas do setor elétrico. Em particular, num horizonte de curto prazo, a proposição de um algoritmo alternativo que gere projeções mais acuradas para a série de consumo residencial mensal de energia elétrica, alternativo aos métodos individuais.

2.

Métodos Individuais

2.1

Conceito de Processos Estocásticos, Séries Temporais e Previsão

A definição de série de tempo está diretamente relacionado ao de processo estocástico (PE), onde este pode ser interpretado como um mecanismo probabilístico

gerador de dados cujo comportamento é descrito por uma equação estocástica. Em outras palavras, uma série temporal é formalmente definida como uma realização de um processo estocástico. Seu comportamento futuro, portanto, é estudado somente em termos probabilísticos.

Em termos formais, um processo estocástico é definido como uma função aleatória, y_t , indexada ao tempo e / ou a eventos, onde seu valor y_t é uma variável aleatória. Assim, a coleção $\{y_t, t \in T\}$ é denotada de espaço de estados; t , de parâmetro; T , de espaço paramétrico ou conjunto de índices e os valores de y_t , de estados do processo estocástico no instante t . Assim o sendo, um processo estocástico pode ser classificado quanto ao espaço paramétrico e ao de estados.

Em relação ao espaço paramétrico, um processo estocástico é classificado como contínuo, se o conjunto T assume valores não contáveis (contínuos). Por outro lado, se os mesmos assumem valores contáveis (discretos), tem-se um PE discreto (MORETIN, 2006). O espaço de estados, por sua vez, pode ser contínuo ou discreto, caso as variáveis aleatórias sejam, respectivamente, contínuas ou discretas.

A estrutura probabilística de um processo estocástico pode ser definida através da especificação da distribuição de probabilidade conjunta. No entanto, sua especificação é complexa. Na prática, um PE é caracterizado por um modelo, a partir do qual se torna possível obter a evolução de seus momentos principais (média, variância, covariância), possibilitando a realização de projeções no tempo.

Segundo SOUZA & CAMARGO (1996), as séries de tempo podem classificadas como discretas, contínuas, determinísticas, estocásticas, multivariadas e / ou multidimensionais.

Um processo estocástico pode ser interpretado como uma família de trajetórias (ou realizações) de uma sequência de variáveis aleatórias. O conjunto com todas as trajetórias é denominado *Ensemble*. Em sua modelagem prática, observa-se uma única série temporal e, por seguinte, estima-se um possível processo estocástico (geralmente um modelo) que a gerou, caracterizando-a.

Tendo em vista a construção dos modelos, previsões probabilísticas podem ser realizadas, com base em informações passadas e atuais. A notação de previsão normalmente é dada por $\hat{y}_T(h)$, sendo definida como a esperança condicional da variável aleatória y_{T+h} , para h passos à frente.

Dessa maneira, tem-se que $E[y_{T+h} / y_0, \dots, y_T] = \hat{y}_T(h)$, onde (y_0, \dots, y_T) é a amostra observada, e $h \in \mathbb{N}^*$. Por outro lado, quando se utilizam sistemas inteligentes, representa-se apenas um valor como previsão, não abordando, geralmente, contextos probabilísticos. Uma previsão quantitativa, portanto, pode ser caracterizada através de um número pontual projetado k passos à frente por um modelo (ou um método, como no caso das RNA), associado uma medida de incerteza (por exemplo, o MSE).

2.2

Métodos de Amortecimento Exponencial

As heurísticas de Amortecimento Exponencial foram inicialmente desenvolvidas por Robert G. Brown, no período em que trabalhava para a marinha norte-americana, durante a II guerra Mundial. Porém ganhou destaque realmente em 1970.

Os métodos de amortecimento exponencial são classificados como sendo automáticos e de validade local. Baseiam-se na premissa de que as observações mais recentes são mais informativas que as mais antigas; com efeito, seu peso decresce exponencialmente, à medida que a observação torna-se mais antiga.

2.2.1

Método Multiplicativo de Amortecimento de Holt-Winters

O método multiplicativo de amortecimento exponencial apresenta tendência aditiva em sua formulação. Captura informações das componentes simples e sazonais de alguma série de interesse, conforme em (2.1).

$$y_t = (a_1 + a_2 t) \times \rho_t + \varepsilon_t. \quad (2.1)$$

A restrição L representa o tamanho do ciclo sazonal. De acordo com MORETIM (2006), os fatores sazonais têm de obedecer ao somatório, descrito em (2.2).

$$\sum_{j=1}^L \rho_j(t) = L \quad (2.2)$$

De acordo com MORETIM (2006), o procedimento de atualização é dado pelas equações em (2.3).

$$\text{Nível: } \hat{a}_1(T) = \alpha \left(\frac{y_T}{\hat{\rho}_{m(T)}(T-1)} \right) + (1-\alpha) [a_1(T-1) + a_2(T-1)]; \quad (2.3)$$

$$\text{Tendência: } \hat{a}_2(T) = \beta (a_1(T) - a_1(T-1)) + (1-\beta) [a_2(T-1)]; \text{ e}$$

$$\text{Sazonalidade: } \hat{\rho}_{m(T)}(T) = \gamma \left[\frac{y_T}{\hat{a}_1(T)} \right] + (1-\gamma) \hat{\rho}_{m(T)}(T-1).$$

Sendo α , β e λ constantes de amortecimento estimadas (denominadas hiperparâmetros) que visam à minimização da função objetivo MSE. A componente sazonal local $\hat{\rho}_{m(T)}(T)$ é estimada pela razão do valor observado e o nível local $a_1(T)$.

Assim sendo, a equação de previsão é dada pelo valor esperado condicional, isto é: $\hat{y}_T(h) = [\hat{a}_1(T) + a_2(T)(T+h)] \hat{\rho}_{m(T+h)}(T) + \varepsilon_t$. Em MONTGOMERY (1990) *apud in* JUNIOR (2009), tem-se o procedimento de estimação dos valores iniciais $\hat{a}_1(0)$, $\hat{a}_2(0)$ e $\rho_i(0)$ - para $i=1, \dots, L$ - os quais são obtidos por meio da amostra.

2.3

Modelos de BOX & JENKINS

A modelagem estatística ARIMA foi inicialmente formulada por BOX & JENKINS, na década de 1970. Assumindo os pressupostos de estacionariedade de 2º ordem e ergodicidade de um processo estocástico subjacente, procura-se detectar o sistema probabilístico gerador da série temporal, através das informações nela contidas. Ainda, baseia-se na premissa de que uma série temporal não estacionária na média pode se tornar estacionária a partir de diferenciações, através da inclusão de um componente auto-regressivo integrado de médias móveis. O mesmo raciocínio vale à sua componente sazonal. De acordo com SOUZA & CAMARGO (1996), a metodologia BOX & JENKINS tem como base a Teoria Geral de Sistemas Lineares, na qual se mostra que a passagem de um ruído branco por um filtro de memória infinita gera um processo estacionário de segunda ordem.

A previsão h passos à frente é calculada através do valor esperado condicional à série temporal de interesse.

Isto é:

$$E[y_{T+h} / y_T, \dots, y_1] = \hat{y}_T(h), \quad \forall h \in \mathbb{N}^*. \quad (2.4)$$

Em síntese, de acordo com JUNIOR (2009), são necessárias seis etapas à modelagem BOX & JENKINS:

1. Identificação dos valores sugeridos para p, d, q, P, D, Q , a partir das análises dos correlogramas;
2. Estimação dos parâmetros do modelo (OLS, máxima verossimilhança, OLS iterativo);
3. Estatísticas de aderência (significância das estimativas, análise residual e análise das estatísticas de desempenho);
4. Se satisfatório, realiza-se o procedimento 5. Caso contrário, indica-se que outros valores para p, d, q, P, D, Q (isso é, retorna-se ao procedimento 2);
5. Teste de sobrefixação; e
6. Gerações das Previsões.

2.3.1

Modelo SARIMA

O modelo SARIMA $(p,d,q) \times (P,D,Q)_s$, também conhecido como ARIMA multiplicativo, possui duas partes: a simples e a sazonal, de modo que a sazonalidade da série temporal passa a ser considerada. O modelo SARIMA $(p,d,q) * (P,D,Q)$ é denotado por $\Phi(B^S) \phi(B) \nabla^d \nabla_s^D \tilde{y}_t = \Theta(B^S) \theta(B) a_t$, onde:

$\nabla^d = (1-B)^d$: operador diferença não sazonal de ordem d ;

$\nabla_s^D = (1-B^S)^D$: operador diferença sazonal de ordem D ;

$\theta(B) = (1-\theta_1 B^1 - \dots - \theta_q B^q)$: operador não sazonal de médias móveis;

$\phi(B) = (1-\phi_1 B^1 - \dots - \phi_q B^q)$: operador não sazonal auto-regressivo;

$\Phi(B^S) = (1-\Phi_1 B^{1S} - \dots - \Phi_p B^{pS})$: operador sazonal auto-regressivo; e

$\Theta(B^S) = (1-\Theta_1 B^{1S} - \dots - \Theta_p B^{pS})$: operador sazonal de médias móveis.

Uma vez definido o tamanho do período sazonal, analisam-se os correlogramas da ACF e PACF, a fim se de estimar os índices p, d e q , como também P, D e Q (relativos à parte sazonal). O raciocínio para estimação do modelo é similar à parte

simples, mas realizada sob os períodos $S, 2S, 3S, \dots$ - para mais detalhes, veja SOUZA & CAMARGO, 2004.

2.4

Método de Redes Neurais Artificiais

As redes neurais artificiais (RNA) são sistemas paralelos compostos por unidades de processamentos simples, conhecidas como neurônios ou processadores, que representam funções lineares e não-lineares. Os referidos são dispostos em uma ou mais camadas, sendo interligados por um grande número de conexões (sinapses) que comumente estão associadas a pesos, responsáveis por ponderar os sinais (dados) de entrada recebidos por respectivo neurônio.

De acordo com HAYKIN (2001), o funcionamento de uma RNA é inspirado nos neurônios biológicos e em sua estrutura paralela de processamento, de forma que possui a capacidade de adquirir, armazenar e utilizar conhecimento experimental, podendo ser utilizada em problemas de reconhecimento de padrões, agrupamento e previsão.

Na figura 2.1, tem-se a estruturação básica de um neurônio artificial, composto por dois módulos de processamento:

- Regra de propagação: executa uma soma ponderada das entradas multiplicadas pelos pesos sinápticos associados a cada entrada do neurônio; e
- Função de ativação: é uma função que é aplicada ao resultado da regra de propagação. O resultado da função ativação é a saída do neurônio artificial.

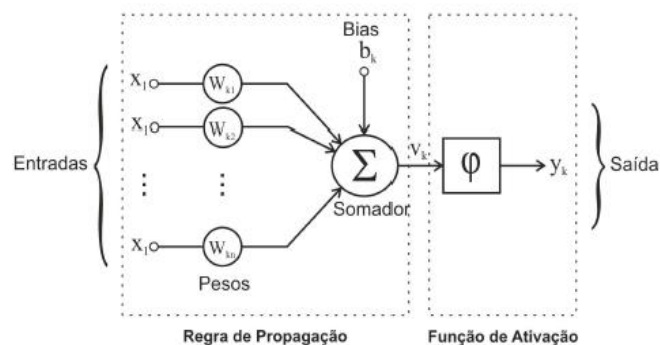


Figura 2.1 - Arquitetura de um Neurônio Artificial.

Não existe nenhum procedimento determinístico para se estruturar uma rede neural artificial, sendo sua estruturação realizada de forma heurística. Seus parâmetros básicos principais são: representação dos dados (*I/O*); tamanho das amostras de treino, validação e teste; número de camadas (*layers*); número de neurônios por camada; funções de ativação e algoritmo de aprendizado.

Na figura 2.2, tem-se a ilustração da topologia de uma rede neural artificial *feed-forward* hipotética com três camadas (de entrada, intermediária e de saída).

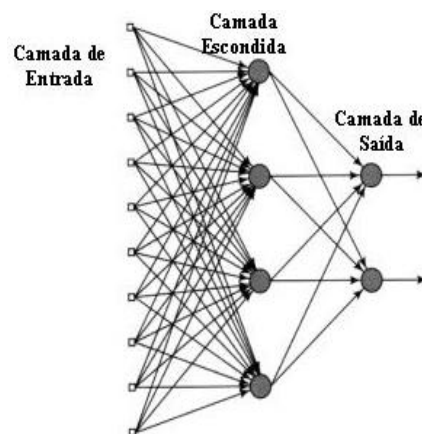


Figura 2.2 - Arquitetura Neural *Feedforward* com Três Camadas.

O algoritmo de treinamento de Levenberg-Marquardt (LM) é considerado o método mais rápido para treinamento de redes *feed-forward backpropagation*, desde que a rede possua uma quantidade “moderada” de pesos sinápticos.

O algoritmo LM utiliza uma aproximação do método de Newton, a qual é obtida a partir da modificação do método de Gauss-Newton, introduzindo-se o parâmetro μ , conforme em (2.5).

$$\Delta w = (J^T J + \mu I)^{-1} J^T \varepsilon \quad (2.5)$$

Onde:

- Δw : diferença entre os pesos inicial e final;
- μ : escalar que controla a derivação dos erros, permitindo que o termo $(J^T J)$ possa ser invertido;
- J : jacobiano dos erros da camada de saída. Cada elemento da matriz J representa uma derivada parcial de um elemento da matriz de erros com o seu correspondente peso;

- I : matriz identidade multiplicada pela constante μ ; e
- ε : vetor de erros da rede neural calculados.

Existem atualmente inúmeros algoritmos de treinamento para as redes neurais MLP, sendo geralmente do tipo supervisionado.

3

Combinação de Métodos

Considere uma situação onde existe um número p ($p > 2$) de métodos preditivos plausíveis para modelagem de um processo de séries temporais $\{y_t\}$. Note que existe incerteza associada quanto à escolha de qual utilizar. No entanto, é de interesse determinar um método, de acordo com critérios, para realização de previsões. Existem três abordagens para determinar um método preditivo:

- Escolher um modelo individual de um conjunto de métodos plausíveis $M = \{M_1, \dots, M_k\}$, baseado em algum critério de seleção, e utilizá-lo para previsão;
- Combinar as previsões oriundas dos métodos individuais de um conjunto M e utilizar a previsão combinada; e
- Combinar as densidades preditivas oriundas dos métodos do conjunto M e utilizar a preditiva combinada para obter-se a previsão.

Sobre o item *i*, existem diferentes estatísticas de aderência (como, por exemplo, MAPE, MAE, MSE) que, em consonância com a abordagem, podem ser adotadas na escolha do melhor método, que são escolhidas de modo a se obter o método que melhor se ajusta à dinâmica temporal da série. Por sua vez, os itens *ii* e *iii* tratam de abordagens distintas, embora tratem de combinação de métodos.

As medidas de aderência utilizadas para os métodos individuais também são utilizados para escolher o de combinação. Em se tratando de combinações de previsões, FLORES & WHITE (1988), destacam duas dimensões a serem definidas:

- Seleção dos métodos individuais a serem combinados; e
- Seleção do método de combinação.

De acordo com FLORES & WHITE (1988), as componentes dos métodos de combinação são denotadas como previsões base e podem ser classificadas em três categorias: objetivas, subjetivas e mistas (isto é, obtidas através da utilização da combinação de previsões objetivas e subjetivas). A categoria objetiva engloba os métodos de Amortecimento Exponencial, ARIMA, de Redes Neurais Artificiais, bem como outros procedimentos com base matemática. A subjetiva inclui todas as abordagens que envolvem o julgamento humano, tais como grupo focado ou opinião de especialistas.

A segunda dimensão concerne à maneira com a qual as técnicas devem ser combinadas. Esta é alvo de estudo há muito tempo e, portanto, uma lacuna a ser pesquisada. De acordo com CLEMEN (1989), *apud in* JUNIOR (2009), alguns métodos têm sido desenvolvidos com o objetivo de se encontrar a melhor forma de se combinar métodos individuais. Apesar da incerteza quanto à melhor combinação, a opinião dos pesquisadores no assunto é unânime: combinar previsões, em regra, conduz ao aumento de acurácia da previsão combinada, em relação às oriundas de métodos individuais.

A dimensão dos métodos de combinação envolve duas abordagens: objetiva ou subjetiva. A objetiva se refere aos métodos que fazem uso de ferramentas matemáticas, onde os resultados são passíveis de serem repetidos. A subjetiva, por sua vez, inclui esforços intuitivos no processo de combinação dos métodos individuais, empregando conhecimento humano ou opinião individual ou de grupo.

3.1

Combinação Linear de Métodos

O artigo de GRANGER & BATES (1969) é considerado o artigo seminal em combinação de métodos preditivos. Neste, os autores propuseram a combinação linear de previsões não-viesadas oriundas de dois métodos estatísticos clássicos, conforme em (3.1).

$$\hat{y}_{CL,T+h} = \hat{\omega} \hat{y}_{1,T+h} + (1 - \hat{\omega}) \hat{y}_{2,T+h} \quad (3.1)$$

Onde: $\hat{y}_{1,T+h}$, previsão do método 1; $\hat{y}_{2,T+h}$, previsão do método 2; $\hat{y}_{CL,T+h}$, previsão do método de combinação linear e $\hat{\omega}$, estimativa do peso adaptativo do método de combinação linear.

NEWBOLD & GRANGER (1974) ampliaram o número de previsões individuais, mantendo, porém, todas as suposições de GRANGER & BATES (1969). Através da combinação de três previsões, conforme em (3.2), obtida dos métodos (ARIMA, Holt-Winters e de auto-regressão *Stepwise*), concluíram que a combinação de métodos acarretou ganhos no processo preditivo.

$$\hat{y}_{CL,T+h} = \sum_{j=1}^3 \hat{\omega}_j \hat{y}_{j,T+h} \quad (3.2)$$

WINKLER & MAKRIDAKIS (1983), *apud in* JUNIOR (2009), analisaram combinações de previsões advindas de dez métodos individuais e os resultados obtidos confirmaram as conclusões de NEWBOLD & GRANGER (1974). Estes resultados consistiram na comparação da estatística MAPE de mil séries temporais, o que permitiu concluir ser melhor ignorar os efeitos da correlação no cálculo de combinações ponderadas.

GRANGER & RAMANATHAN (1984), *apud in* JUNIOR (2009), chamaram a atenção para o fato de que os métodos convencionais de combinação linear de previsões individuais podem ser interpretados como uma forma estruturada de regressão. Argumentaram ainda que este método de combinação é equivalente ao método de mínimos quadrados ordinários (MQO), considerando a previsão combinada como variável endógena e as individuais, como explicativas.

Os pesos lineares adaptativos podem ser fixos ou variáveis (não necessariamente positivos) ou somar ou não uma unidade. Na abordagem clássica de GRANGER e NEWBOLD (1986), destaca-se que as previsões pontuais são de forma linear, sendo os pesos variantes no tempo.

De forma genérica, o método de combinação linear, h passos à frente, pode ser descrito conforme em (3.3).

$$\hat{y}_{CL, T+h} = \sum_{j=1}^k \hat{\omega}_{jt} \hat{y}_{j, T+h} \quad (3.3)$$

Onde $\hat{y}_{CL,T+h}$, a previsão combinada para o instante $T+h$; $\hat{\omega}_{jt}$, o peso estimado (não necessariamente positivo ou normalizado) e $\hat{y}_{j,T+h}$, a previsão modelo j , para o instante $T+h$ - considerando uma amostra (y_1, \dots, y_T) .

Outra abordagem é a combinação linear de densidades preditivas, na qual algumas propriedades merecem destaque. De acordo com MUBWANDARIKWA & FARIA (2008), a formulação geral desta abordagem pode ser descrita por:

$$\rho_{CL(\rho_1, \dots, \rho_k)}(y_{CL,T+h} | D_T) = \sum_{j=1}^k \omega_j \rho_j(y_{j,T+h} | D_T) \quad (3.4)$$

Onde: $\rho_{CL(\rho_1, \dots, \rho_k)}(y_{CL,T+h} | D_T)$, densidade preditiva resultante da combinação linear de densidades, dadas as informações D_T ; $\rho_j(y_{j,T+h} | D_T)$, densidade preditiva do modelo j ($j = 1, \dots, k$) para o instante $T+h$, dada a informação D_T e ω_j , peso adaptativo associado à densidade do método individual j . Especificamente à projeção de consumo residencial, a combinação linear de previsões é descrita conforme a equação (3.5).

$$\hat{y}_{CL,T+h} = (BJ_{T+h}) * \hat{\omega}_1 + (HW_{T+h}) * \hat{\omega}_2 + (RNA_{T+h}) * \hat{\omega}_3 \quad (3.5)$$

Onde:

- BJ_{T+h} , previsão do modelo ARIMA para o instante $T+h$;
- HW_{T+h} , previsão do MAE para o instante $T+h$;
- RNA_{T+h} , previsão do método de RNA para o instante $T+h$; e
- $\hat{\omega}_1, \hat{\omega}_2, \hat{\omega}_3$, pesos associados linearmente às respectivas previsões.

3.2

Intervalo de Confiança do Método de Combinação Linear de Previsões

Uma vez escolhidos e estimados os modelos individuais, é possível gerar densidades preditivas através da utilização do método de Quase-Monte Carlo. O procedimento de simulação utilizado para os modelos estatísticos, nesta dissertação, inicia-se com a geração de uma sequência de números quase-aleatórios independentes

pertencentes à distribuição $U [0,1]$. Posteriormente, estes são inseridos em um algoritmo de inversão (*Inversão de Moro*) que os interpreta como probabilidades acumuladas, de forma a fornecer amostras independentes pertencente à distribuição normal-padrão. Por seguinte, as amostras normais padrão são filtradas por *Cholesky*, gerando resíduos na escala da série temporal considerada (no caso, a de consumo residencial). Assim, para cada instante, realiza-se este procedimento n vezes. Mais detalhes sobre o algoritmo de Moro, assim como as sequências de Quase-Monte-Carlo. O procedimento utilizado para as redes neurais são explanados mais adiante.

$$\Sigma_{\varepsilon_t} = ZDZ^T \quad (\text{caso multivariado})$$

$$\hat{\sigma}_{\varepsilon_t}^2 = \hat{\sigma}_{\varepsilon_t} \cdot 1 \cdot \hat{\sigma}_{\varepsilon_t} \quad (\text{caso univariado})$$

A matriz Z (desvio-padrão) multiplicada pelo vetor de erros constrói-se um vetor normal-padrão u_t ($n \times 1$) para o caso univariado, a cada instante t . (JUNIOR, 2009)

$$u_t \equiv Z^{-1} \varepsilon_t \quad (3.6)$$

A média de ε_t ainda é zero, pois os elementos de u_t foram sorteados de uma distribuição de normal-padrão e Z é uma constante.

Desse modo, com a decomposição da variância fora da amostra ($\hat{\sigma}_{\varepsilon_{T+i}}^2$), foi possível transformar um vetor de choques normais padrão independentes u_t em um de choques ε_t na escala da série temporal supracitada no horizonte de previsão considerado. A equação vetorial (3.7) explicita, em termos matemáticos, o salientado.

$$\begin{bmatrix} y_{T+i, \text{cenário } 1} \\ \cdot \\ \cdot \\ \cdot \\ y_{T+i, \text{cenário } n} \end{bmatrix} = \begin{bmatrix} \hat{y}_{T+i} \\ \cdot \\ \cdot \\ \cdot \\ \hat{y}_{T+i} \end{bmatrix} + \hat{\sigma}_{\varepsilon_{T+i}} \begin{bmatrix} u_1 \\ \cdot \\ \cdot \\ \cdot \\ u_n \end{bmatrix}, \quad i = 1, 2, \dots, h \quad (3.7)$$

A cada sorteio, os resíduos na escala da série são substituídos na equação (3.7), obtendo, ao fim de n sorteios, a respectiva densidade preditiva, para $T+h$. Alguns testes foram realizados, na presente pesquisa, a fim de verificar se houve convergência: histograma, QQ-plot, PP-plot, teste de normalidade (software @Risk). Todos a

confirmaram. Especificamente para os dois modelos estatísticos adotados, para cada instante fora da amostra, o software estatístico forneceu o desvio-padrão estimado o qual foi multiplicado pela sequência de normal padrão, gerando os respectivos resíduos (no caso, geraram-se 1.000). Em seguida, estes foram somados a cada previsão, gerando os cenários (*out of sample*).

Para as redes neurais artificiais, adotou-se outro procedimento, quanto à geração dos cenários. O problema deveu-se ao fato de as RNA não possuírem um modelo explícito que possibilite a estimação do desvio padrão amostral fora da amostra, em função dos parâmetros, conforme os modelos estatísticos. Para o último elemento da validação, o procedimento foi similar aos modelos estatísticos, ou seja, calculou-se o desvio-padrão amostral dos resíduos até o instante relativo à última observação da validação e, então, aplicou-se a equação (3.7).

Cada cenário foi inserido na janela da rede neural, gerando 1.000 cenários para o período posterior (no caso, para o primeiro elemento da amostra de teste). Como a janela possui tamanho 5, para o segundo elemento da amostra de teste, utilizou-se os quatro últimos pontos da validação e o primeiro do teste. Isso possibilitou gerar o cenário seguinte. Tanto o primeiro ponto de teste e quanto o último da validação são variáveis, ou seja, cada respectivo cenário de ambos os pontos foi inserido na janela da rede neural de forma a gerar o cenário para o instante seguinte, onde os outros três são fixos. E assim sucessivamente até o décimo segundo passo. Como foi observado um crescimento da variância dos cenários (o que naturalmente ocorre), à medida que o horizonte ficava maior, adotou-se este procedimento para os fins da pesquisa. Além disso, foram feitos teste estatísticos para verificação de normalidade e todos (a 5% de significância) não rejeitaram a hipótese de normalidade, para todos os passos à frente projetados fora da amostra.

Tendo em vista a geração de cenários para os três métodos individuais, foi possível a combinação linear dos cenários, o que possibilitou a geração de intervalos de confiança dos métodos combinados, com 95% de credibilidade. Cada cenário individual foi linearmente combinado, gerando o combinado.

Ao final, calculou-se a variância da densidade preditiva combinada e, então, procedeu-se à construção do intervalo de confiança, dada a previsão no respectivo instante.

4.

Programação Matemática

A Programação Matemática consiste em encontrar uma ou mais soluções ótimas que correspondem a valores extremos de um ou mais objetivos (representadas por funções nominadas funções-objetivo).

Os problemas de Programação Matemática podem ser subdivididos em duas classes básicas: lineares e não lineares. Assim sendo, têm-se as:

- Lineares: se todas as restrições e a função objetivo são equações ou inequações lineares (ou do 1º grau), o problema é dito *linear*; e
- Não Lineares: se pelo menos uma equação ou inequação, no PPM, for não linear, o problema é dito *não linear*.

O termo “Matemática”, neste contexto, é usado de em sentido amplo, abrangendo não apenas seus os elementos familiares, mas também os tópicos relacionados à Lógica Matemática.

O referido método é amplamente utilizado em inúmeras aplicações, particularmente em problemas de tomada de decisão. Três elementos básicos têm de ser considerados: decisões, restrições e objetivo.

As decisões de um problema matemático geralmente são representadas por símbolos X_1, \dots, X_n e são denotadas de variáveis de decisão, que, por sua vez, podem representar quantidades de forma a maximizar ou minimizar determinada função objetivo. (RAGSDALE, 2004)

De acordo com RAGSDALE (2004), as restrições podem ser expressas de várias maneiras em um Problema de Programação Matemática (PMM). Em linhas gerais, há três maneiras de expressá-las:

- Restrição menor ou igual a: $f(X_1, \dots, X_n) \leq b$;
- Restrição maior ou igual a: $f(X_1, \dots, X_n) \geq b$;
- Restrição igual a: $f(X_1, \dots, X_n) = b$.

Em cada caso, a restrição é uma função das variáveis de decisão que deve ser menor ou igual, maior ou igual ou igual, a determinado valor específico (representado pela letra b).

Por último, o objetivo de um PPM é representado por uma função no seguinte formato: $MAX(ou MIN): f(X_1, \dots, X_n)$. A formulação genérica de um PPM pode ser descrita como:

$$MAX(ou MIN): f(X_1, \dots, X_n) \quad \begin{array}{l} f(X_1, \dots, X_n) \leq b_1 \\ f(X_1, \dots, X_n) \geq b_2 \end{array}$$

É importante salientar que não existem métodos universais capazes de resolver com eficácia todos os problemas de otimização. Alguns podem ser indicados para uma classe de problemas, no entanto podem não ter desempenho satisfatório em outras aplicações.

Em problemas envolvendo mais de uma função-objetivo, a busca soluções ótimas é conhecida como otimização multiobjetivo. Assim, tendo em vista que esta abordagem envolve objetivos múltiplos, pode-se particularizar a otimização de objetivo único como um caso particular da multiobjetivo.

Teorias e algoritmos aplicáveis em PPM com objetivo único também são usados em casos multiobjetivos, a partir de transformações das funções em uma função de objetivo único. Em muitos problemas multiobjetivo, surgem diferentes aplicações científicas, onde várias pesquisas visam a desenvolver métodos para solucioná-los.

Existem vários critérios que podem ser considerados para resolver este problema complexo. Alguns são baseados em escalonar as funções, como sendo o vetor função objetivo, transformando-o em uma função escalar. Outros, porém, as tratam como novas restrições, tendo como base a abordagem MINIMAX.

Em particular, destaca-se o Método da Programação de Metas, cuja ideia principal é encontrar uma solução que atinja metas predefinidas para cada uma das soluções objetivo. Caso estas soluções não existam simultaneamente, é possível obter soluções que minimizem os desvios percentuais das funções em relação a estas metas. Outros métodos podem ser citados, como, por exemplo:

- Otimização Hierárquica, proposto por WALZ (1967);
- Método da Negociação, proposto por HAIMES e HALL (1974).

4.1

Programação MINIMAX

O conceito do método do ótimo MINIMAX é derivado da Teoria de Jogos, onde é necessário resolver situações de conflito. A Metodologia MINIMAX compara os desvios relativos obtidos para cada função objetivo em relação à sua solução ideal e desejada (isto é, a meta).

$$\begin{array}{l} \min_x Q \\ \left\{ \begin{array}{l} \omega_1 \left(\frac{\text{objetivo}_1 - f_1(\cdot)}{\text{meta}_1} \right) \leq Q \\ \cdot \\ \cdot \\ \omega_n \left(\frac{\text{objetivo}_n - f_n(\cdot)}{\text{meta}_n} \right) \leq Q \end{array} \right. \end{array}$$

O objetivo MINIMAX pode ser usado para explorar os pontos nos limites da região viável - além dos pontos extremos. Assim, quando se minimiza Q , ocorre o mesmo com os desvios percentuais absolutos ponderados com relação às respectivas metas. Desse modo, o desvio máximo ponderado de qualquer meta assume um valor igual ou inferior a Q .

4.2

Programação por Metas

Inicialmente a Programação de Metas foi utilizado em um problema de programação linear (PPL) com uma única função objetivo. De acordo com RAGSDALE (2004), a programação de metas ganhou popularidade depois do trabalho de IGNIZIO (1976), LEE (1972), dentre outros. Ainda segundo (RAGSDALE, 2004), ROMERO (1991) fez um levantamento do *estado da arte* acerca desta metodologia, listando várias aplicações, por exemplo, em Engenharia.

Comumente as técnicas de otimização apresentam restrições invioláveis, porém estas podem ser, em algumas situações, restritivas demais, ao passo de tornar o PMM

insolúvel. Por isso, torna-se factível relaxá-las para que seja possível resolver algum problema de interesse.

A programação de metas é aplicada em problemas de múltiplas funções-objetivos. Esta técnica requer um procedimento de solução interativa no qual o tomador de decisão investiga uma série de soluções viáveis a fim de encontrar a mais satisfatória.

Na otimização MINIMAX, tem-se a integração das abordagens multiobjetiva e a de metas. Os objetivos no problema MINIMAX podem ser obtidos otimizando individualmente cada função objetivo. Uma vez definidos, realiza-se a minimização da restrição Q.

Em síntese, esta metodologia requer que metas sejam especificadas para que cada objetivo a sua respectiva meta. O objetivo principal da programação de metas é encontrar uma solução que atinja metas pré-definidas para cada uma das funções objetivo. Caso não exista uma solução que atinja as metas pré-especificadas para todas as funções objetivo, objetiva-se encontrar soluções que minimizam desvios percentuais absolutos dos valores encontrados em relação às respectivas metas.

As metas são interpretadas como restrições adicionais. Com efeito, novas variáveis são acrescentadas para representação dos desvios em relação às metas pré-determinadas.

A função objetivo especifica os desvios em relação às metas e prioriza a realização (isto é, o sucesso) de cada meta, em termos quantitativos. Logo, o objetivo, neste caso, da programação de metas é minimizar os desvios percentuais absolutos.

4.3

Programação Multiobjetivo

A abordagem de otimização multiobjetivo, ou *vetorial*, ou *multicritério*, parte do suposto de existir, ao menos, duas funções objetivo em determinado PPM que devem ser minimizadas (ou maximizadas), consoantes a restrições e parâmetros. Desse modo, suas soluções podem ser de dois tipos:

- Soluções que, sob todos os objetivos simultaneamente considerados, são suplantadas por outras. Isso significa que há soluções que fornecem valores melhores de função objetivo para todos os k objetivos $f_i(.)$. Assim, estas devem ser descartadas; e

- Soluções que, comparadas com outras, são melhores em determinados objetivos, porém piores em outros. Isso significa que a escolha deve ser realizada com algum grau de subjetividade.

No último grupo, são denominadas soluções *eficientes* ou *pareto-ótimo*. Este conjunto é um objeto bem determinado, posto um problema de otimização multicritério. Sua determinação é um dos problemas centrais da abordagem multiobjetivo. Em linhas gerais, um PPM com mais de um objetivo tende a possuir um conjunto, geralmente limitado, de possíveis soluções. Dessa forma, tem-se que o objetivo de resolver um determinado PPM deve-se convergir para uma solução única.

O conjunto pareto-ótimo constitui um elenco de alternativas, candidatas a se tornarem a solução final. Para viabilizar isso, a otimização multiobjetivo emprega ainda outra etapa, em que o conjunto de soluções candidatas é reduzido até a determinação de uma única solução, através de uma sistemática; que, por sua vez, pressupõe a chamada *função utilidade*, implicando um padrão de preferências coerente e ordenado.

O problema multiobjetivo (PMO) pode, então, ser formulado como uma combinação de tais etapas de determinação de soluções eficientes e de escolha da final (aplicação da função utilidade).

Assim, formalizando, tem-se que:

Seja $x \in \mathbb{R}^n$ o vetor dos parâmetros de um problema multiobjetivo (PMO) e $\Psi_x \in \mathbb{R}^n$, um subconjunto do espaço ao qual o vetor x se encontra restrito. Seja ainda, $f(\cdot): \mathbb{R}^n \rightarrow \mathbb{R}^m$, o vetor de funções-objetivo. Dados $f(\cdot)$ e Ψ_x , o conjunto χ^ das soluções eficientes do PMO descrito por:*

$$\min_x f(x)$$

Sujeito a: $\{x \in \Psi_x\}$.

Onde $f(x)$, função objetivo; x , variável de decisão e Ψ_x , região viável. Algoritmos de determinação das soluções eficientes de um PMO são formulados, a partir das propriedades do conjunto das soluções eficientes χ^* .

4.4

Programação por Metas via Desvio Mínimo

O objetivo MINIMAX, descrito antes, pode ser usado para explorar os pontos limites região viável - além dos pontos extremos. Esta metodologia é análoga à programação de meta ponderada, porém, em vez de minimizar a soma ponderada dos desvios em relação às metas, esta soma deve obedecer a um valor máximo de desvio Q , escrito na forma de restrição e este desvio máximo é minimizado. Assim, o parâmetro Q é o desvio máximo. Em outras palavras, o objetivo é minimizar o desvio médio máximo pelo parâmetro, ou seja, Q .

Particularmente, neste trabalho, assumiu-se valor igual a um para os valores de w e, para os dos desvios absolutos das metas com relação às funções objetivo, valor máximo igual a Q .

$$FO_{MC} = MIN \{Q\}$$
$$\left\{ \begin{array}{l} d_1 = \frac{meta_1 - FO_1}{FO_1} \leq Q \\ \cdot \\ \cdot \\ \cdot \\ d_k = \frac{meta_k - FO_k}{FO_k} \leq Q \end{array} \right.$$

5.

Aplicação à Série de Consumo Residencial de Energia Elétrica

Neste capítulo, são aplicados os métodos individuais e os combinados linearmente para projeção de curto prazo da série de consumo residencial mensal de energia elétrica, tendo como objetivo uma função multicritério.

A amostra contém cento e cinquenta e nove observações, sendo doze usadas para análise fora da amostra. Os softwares utilizados foram: AIMMS, MATLAB, E-Views, R, SPSS, Excel (solver).

Após a exposição dos gráficos e tabelas, seus resultados foram comentados. Seguem-se, nesta ordem, as metodologias abordadas: ARIMA, MAE, RNA e a combinação linear multiobjetivo destes métodos.

5.1

Aplicação do Modelo ARIMA

A metodologia BOX & JENKINS impõe fortes restrições à série subjacente: estacionariedade de 2º ordem, normalidade e a série de resíduos tem de ser uma realização de um processo estocástico ruído branco.

É fundamental que as mesmas sejam observadas para que propriedades estatísticas importantes e desejáveis do modelo não sejam perdidas. Os resultados desses testes foram omitidos, porém confirmam os pressupostos.

5.1.2

Modelagem

O modelo ajustado foi o SARIMA $(1,0,0) * (1,0,3)$. Inúmeras ordens de modelos ARIMA foram testadas.

Tabela 5.1 – Estimativas dos Parâmetros e Estatísticas do Modelo ARIMA.

Termo	Coefficiente	Erro padrão	Estatística <i>t</i>	P-valor
ar [1]	0,8184	0,0498	16,414	0,0000
sar [12]	0,3681	0,0882	4,1716	0,0001
sar [24]	0,6320	0,0882	7,1600	0,0000
sma [12]	-0,2720	0,1116	-2,4374	0,0163
sma [24]	-0,8959	0,0243	-36,8597	0,0000
sma [36]	0,2581	0,1032	2,5010	0,0138

Note, na tabela 5.1, que todas as estimativas apresentam significância estatísticas, ou seja, a hipótese de que são iguais a zero foi rejeitada a 5% de significância.

Tabela 5.2 - Principais Estatísticas de Aderência.

R ²	81,03%	R ² ajustado	80,02%
DW	2,1395	Ljung-Box (p-valor)	76,81%
MAPE	5,82%	BIC	4,43E+004

Na tabela 5.2, tem-se que a estatística Durbin-Watson apresenta valor próximo a dois (valor teórico), evidenciando descorrelação de lag 1. A Ljung-Box mostra a mesma inferência, porém até o lag 18.

5.2

Aplicação do Método de Holt-Winters

Como não há restrições quanto à sua utilização, consideraram-se as estatísticas R² e MAPE para a escolha do melhor ajuste.

5.2.1

Modelagem

Testou-se, inicialmente, o modelo com sazonalidade aditiva, uma vez que a série é homocedástica, porém não obtiveram o melhor ajuste.

O modelo com melhor ajustamento foi o Holt-Winters com sazonalidade multiplicativa. A seguir, encontram-se as principais características estatísticas do método estimado.

Tabela 5.3 - Valores dos Hiperparâmetros das Componentes

Componente	Hiperparâmetro
Nível	0.51347
Sazonalidade	0.49758

5.2.2

Previsões

Na tabela 5.4, tem-se os valores de MAPE dentro e fora da amostra obtidos do Método de Holt-Winters multiplicativo.

Tabela 5.4 - MAPE's do Método Holt-Winters

Dentro da Amostra	Fora da Amostra
5,82%	5,40%

5.3

Aplicação do Método de Redes Neurais Artificiais

Os critérios de escolha da RNA foram: MAPE, MAE e MSE. A melhor RNA possui as seguintes características:

- Tamanho da janela: 5;
- Padrão de entrada: *premmx*;
- Número de camadas escondidas: 1;
- Algoritmo de treinamento: Levenberg-Marquardt (*trainlm*); e
- Número de neurônios na camada escondida: 5.

Tabela 5.5 - MAPE's da RNA

Treino	Validação	Teste
4,36%	5,39%	7,01%

5.4

Otimização do Método de Combinação Multicritério

Para a otimização do método proposto, foram necessários nove passos, para que, finalmente, fosse possível a estimação de seus pesos adaptativos.

1. Identificar as variáveis de decisão no problema;
2. Identificar os objetivos e formulá-los;
3. Identificar as restrições e formulá-las;
4. Resolver o problema uma vez para cada um dos objetivos adotados, conforme o passo 2, a fim de identificar o ótimo para cada função objetivo;
5. Declarar as metas de cada objetivo como os valores ótimos encontrados no passo 4;
6. Para cada meta, criar uma função desvio que determine a quantidade na qual qualquer solução deixe de atender a meta. Adotou-se a função desvio percentual absoluto;
7. Para cada função desvio do passo 6, atribuir um peso e criar uma restrição que demande o valor da função desvio absoluta ponderada fosse inferior à variável Q do MINIMAX;
8. Resolver o PMM resultante a fim de minimizar Q ; e
9. Verificar a solução do problema. Se factível, terminar; caso não; retornar ao passo 7.

A estimação se deu considerando a amostra de treino, combinando linearmente as previsões pontuais advindas dos três métodos base. Como anteriormente citado, as funções objetivo utilizadas foram: MAPE, MSE e MAE.

Assim, definiram-se as metas utilizando o otimizador *solver* (aplicativo Excel). A formulação simbólica da programação de cada metas é dada a seguir.

✓ Meta 1

$$FO: MIN \{MAPE\}$$

$$MAPE = \left(\sum_{t=1}^T \frac{|Z_{CL,t} - \text{observa\~{c}o}_t|}{\text{observa\~{c}o}_t} \right) \times \frac{1}{T}$$

$$\rho_{BJ} \geq 0$$

$$\rho_{HW} \geq 0 \quad \rho_{BJ} + \rho_{HW} + \rho_{RNA} = 1$$

$$\rho_{RNA} \geq 0$$

$$Z_{CL,t} = \rho_{BJ} * Z_{BJ} + \rho_{HW} ** Z_{HW} + \rho_{RNA} * Z_{RNA}$$

✓ **Meta 2:**

$$FO: MIN \{MAE\}$$

$$MAE = \frac{\sum_{t=1}^T |Z_{CL,t} - \text{observa\~{c}o}_t|}{T}$$

$$\rho_{BJ} \geq 0$$

$$\rho_{HW} \geq 0 \quad \rho_{BJ} + \rho_{HW} + \rho_{RNA} = 1$$

$$\rho_{RNA} \geq 0$$

$$Z_{CL,t} = \rho_{BJ} * Z_{BJ} + \rho_{HW} ** Z_{HW} + \rho_{RNA} * Z_{RNA}$$

✓ **Meta 3:**

$$FO: MIN \{MSE\}$$

$$MSE = \frac{\sum_{t=1}^T (Z_{CL,t} - \text{observa\~{c}o}_t)^2}{T}$$

$$\rho_{BJ} \geq 0$$

$$\rho_{HW} \geq 0 \quad \rho_{BJ} + \rho_{HW} + \rho_{RNA} = 1$$

$$\rho_{RNA} \geq 0$$

$$Z_{CL,t} = \rho_{BJ} * Z_{BJ} + \rho_{HW} ** Z_{HW} + \rho_{RNA} * Z_{RNA}$$

Uma vez definidas as metas, a formulação simbólica do PPM MINIMAX é dada a seguir.

$$\text{MIN} \{Q\}$$

$$\frac{\text{META}_1 - \text{MAPE}}{\text{MAPE}} \leq Q$$

$$\frac{\text{META}_2 - \text{MAE}}{\text{MAE}} \leq Q$$

$$\frac{\text{META}_1 - \text{MSE}}{\text{MSE}} \leq$$

$$Z_{CL,t} = \rho_{BJ} * Z_{BJ,t} + \rho_{HW} * Z_{HW,t} + \rho_{RNA} * Z_{RNA,t}$$

$$\text{MAPE} = \left(\sum_{t=1}^T \frac{|Z_{CL,t} - \text{observação}_t|}{\text{observação}_t} \right) \times \frac{1}{T}$$

$$\text{MSE} = \frac{\sum_{t=1}^T (Z_{CL,t} - \text{observação}_t)^2}{T}$$

$$\text{MAE} = \frac{\sum_{t=1}^T |Z_{CL,t} - \text{observação}_t|}{T}$$

$$\rho_{BJ} \geq 0$$

$$\rho_{HW} \geq 0$$

$$\rho_{RNA} \geq 0$$

$$\rho_{BJ} + \rho_{HW} + \rho_{RNA} = 1$$

A formulação no formato AIMMS, para estimação dos pesos adaptativos, bem como sua identificação na implementação é dada a seguir.

Restrições:

MAPE(e)

MSE(e)

MAE(e)

Conjuntos:

Períodos(p)

Métodos(m)

Métricas(e)

Parâmetros:

Metas(e)

Consumo(p)

previsão(m, p)

<i>Banco de Dados :</i>	<i>Variáveis :</i>
<i>dados</i>	<i>peso(m)</i>
<i>_ período</i>	<i>preComb(p)</i>
<i>_ consumo</i>	<i>MAPE</i>
<i>_ previsão</i>	<i>MSE</i>
<i>_ metas</i>	<i>MAE</i>
	<i>desvioMAX</i>

Tabela 5.6 – Estimativas dos Pesos Adaptativos do Modelo Linear Multicritério (fonte: AIMMS)

Métodos	Pesos
ARIMA	0,363874889
HOLT-WINTERS	0,090287294
REDES NEURAIIS	0,545837817

Uma vez estimados os pesos, foram geradas as previsões combinadas advindas do modelo proposto.

Por conseguinte, de posse das estimativas dos pesos, os cenários de Quase-Monte-Carlo dos métodos base foram combinados, gerando uma densidade preditiva do modelo linear multicritério (supostamente gaussiana, visto que foi assumido normalidade para as densidades geradas pelos modelos base) e, por consequência, um intervalo de confiança com 95% de credibilidade (predição intervalar) para o horizonte considerado fora da amostra.

5.5

Comparação dos Modelos

Os modelos base e o linear multicritério foram comparados em relação às observações fora da amostra. As células em destaque (em amarelo) destacam a previsão que mais próxima ficou do valor verdadeiro.

Tabela 5.7 - Previsões do Modelo Combinado Multicritério, Modelos Base e Valores Históricos

Data	ARIMA	Holt-Winters	Redes Neurais	Combinação Linear	Histórico
2007-12	731.885,13	699.882,19	681.354,08	701.413,91	708.970,08
2008-01	665.182,63	631.601,56	609.663,26	631.846,11	621.971,26
2008-02	619.608,13	587.424,94	506.204,20	554.802,24	531.089,20
2008-03	605.697,81	554.449,69	452.994,08	517.719,29	514.755,08
2008-04	598.781,13	542.600,25	538.311,12	560.701,89	527.584,12
2008-05	608.940,13	555.763,69	581.931,51	589.396,64	570.372,51
2008-06	622.141,81	587.499,38	567.888,81	589.400,70	585.272,81
2008-07	631.551,69	601.305,00	625.303,37	625.410,23	626.854,77
2008-08	666.884,25	651.478,75	540.935,50	596.745,74	596.979,50
2008-09	659.990,38	667.880,69	813.151,02	744.303,64	717.862,02
2008-10	649.974,50	685.670,44	601.256,68	626.605,36	649.547,68
2008-11	678.997,25	740.296,50	737.147,17	716.272,22	659.867,17

Contrapondo os métodos, percebe-se que o modelo linear multicritério obteve maior acurácia em cinco instantes. A tabela 5.8 mostra em termos percentuais a evolução temporal de cada método. Em 5.9 e 5.10, expõem-se as métricas de aderência MAE e R^2 , respectivamente.

Tabela 5.8 – Erro Percentual Absoluto (APE) dos Modelos Combinados

Data	ARIMA	MAE	RNA	Combinação Linear
2007-12	3,23%	1,28%	3,90%	1,07%
2008-01	6,95%	1,55%	1,98%	1,59%
2008-02	16,67%	10,61%	4,69%	4,46%
2008-03	17,67%	7,71%	12,00%	0,58%
2008-04	13,49%	2,85%	2,03%	6,28%

2008-05	6,76%	2,56%	2,03%	3,34%
2008-06	6,30%	0,38%	2,97%	0,71%
2008-07	0,75%	4,08%	0,25%	0,23%
2008-08	11,71%	9,13%	9,39%	0,04%
2008-09	8,06%	6,96%	13,27%	3,68%
2008-10	0,07%	5,56%	7,43%	3,53%
2008-11	2,90%	12,19%	11,71%	8,78%

Na tabela 5.9, o valor de MAPE do modelo linear multicritério foi inferior em seis instantes em relação às outras metodologias (fora da amostra). Além disso, quatro previsões obtiveram valores de APE inferior a 1%, como também menor APE máximo. Comparando-o com o Holt-Winters, quanto à métrica MAPE, o resultado fora da amostra foi 2,54% mais preciso em média. Destaca-se que teve pouca perda de desempenho fora da amostra (0,51%).

Tabela 5.9 - MAPE's dos Modelos Estimados

Métodos	Dentro da Amostra	Fora da Amostra
ARIMA	5,66%	7,88%
Holt-Winters	5,82%	5,40%
RNA	4,36%	7,02%
Combinação Linear	2,35%	2,86%

Tabela 5.10 - MAE's dos Modelos Estimados

Métodos	Dentro da Amostra	Fora da Amostra
ARIMA	32.684,26	45.354,33
Holt-Winters	32.760,43	32.765,21
RNA	26.801,18	37.057,95
Combinação Linear	13.502,32	17.320,45

O modelo linear possui menor valor de MAE, fora da amostra, mostrando maior capacidade de generalização, consoante tal métrica. Verifica-se que, em relação ao melhor modelo, o ganho de ajuste foi superior a 100%. O mesmo pode ser inferido fora da amostra.

Tabela 5.11 – Coeficiente de Explicação (R^2) dos Modelos Estimados

Métodos	Dentro da Amostra	Fora da Amostra
ARIMA	81,03%	38,40%
Holt-Winters	79,11%	66,13%
RNA	85,13%	53,54%
Combinação Linear	95,47%	86,87%

Considerando o coeficiente de explicação, tem-se que o modelo linear foi superior. Assim, foi capaz de explicar 86,87% da variabilidade que não foi utilizada na estimação do modelo multicritério. Analisando-se os gráficos 5.1 e 5.2, verifica-se que o modelo proposto obteve melhor desempenho que os modelos base, quanto à estatística MAPE, dentro e fora da amostra. Visualmente percebe-se que o ganho em termos de previsão foi considerável.

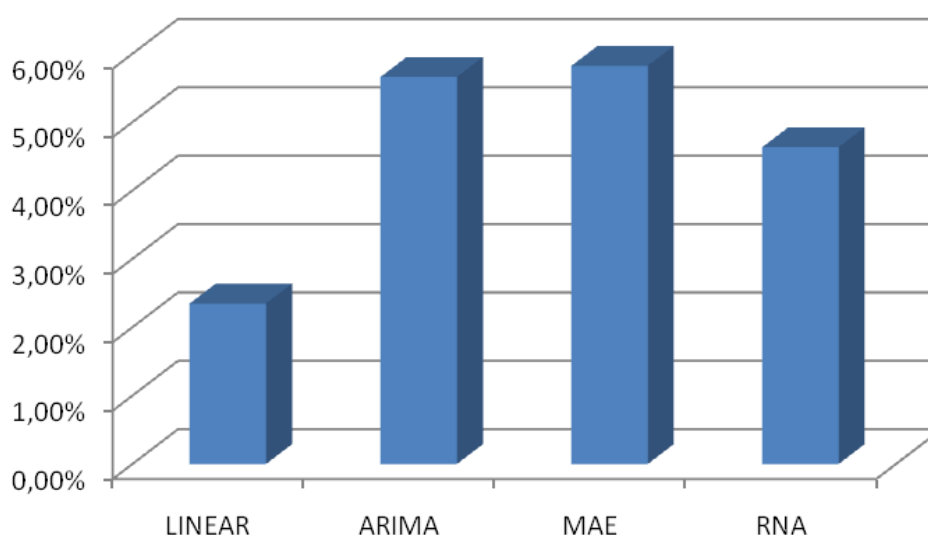


Gráfico 5.1 - Valores dos MAPE's dos Métodos (Dentro da Amostra).

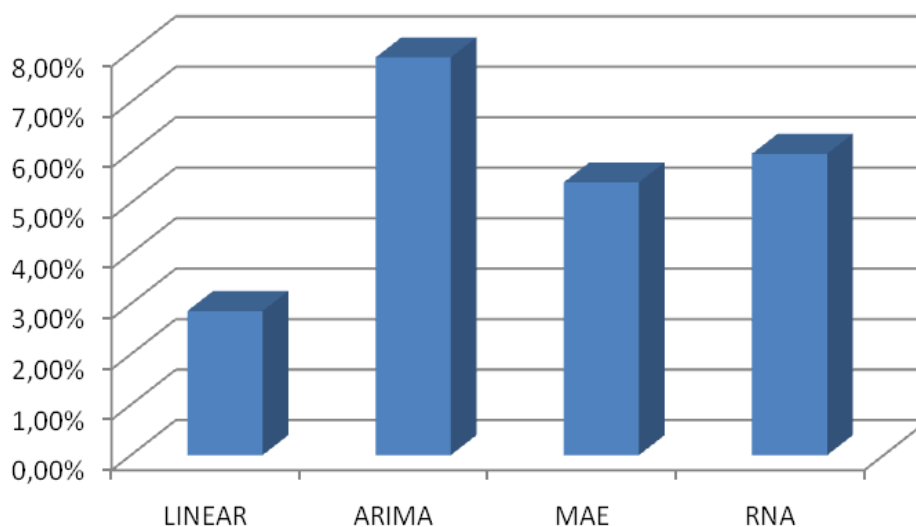


Gráfico 5.2 - Valores dos MAPE's dos Métodos (Fora da Amostra).

5.5.1

Comparação dos Intervalos de Confiança dos Modelos Combinados

Os modelos base utilizados tiveram, em cada instante, cenários gerados a partir do método de Quase-Monte-Carlo.

Dado que, na maioria dos períodos, as variâncias do modelo proposto foram menores que os individuais, foi realizado a comparação somente com o melhor neste aspecto, no caso, o Holt-Winters. A tabela 5.20 explicita os resultados auferidos.

Destaca-se que todos os valores reais se encontram dentro do intervalo de confiança, tendo em vista o horizonte considerado fora da amostra.

Além do mais, o intervalo de confiança da combinação linear, na maioria dos lag's (fora da amostra), obteve amplitudes menores que os modelos individuais.

Tabela 5.12 - Amplitudes dos Limites do Intervalo de Confiança dos Modelos Combinados em Valores Absolutos.

Instante	Holt-Winters	Linear
1	198.175,25	103.739,45
2	205.764,00	138.098,86
3	215.531,75	161.559,40
4	225.231,88	181.610,13
5	237.035,03	199.705,52
6	252.127,91	219.223,04
7	270.197,66	235.528,21
8	285.690,25	251.694,19
9	307.086,00	268.104,71
10	324.050,59	282.638,12
11	341.128,66	313.702,47
12	364.077,63	346.614,51

O modelo linear multicritério gera cenários com menor variância (medida de incerteza), em todos os instantes considerados, visto as amplitudes absolutas dos limites inferior e superior do intervalo de confiança.

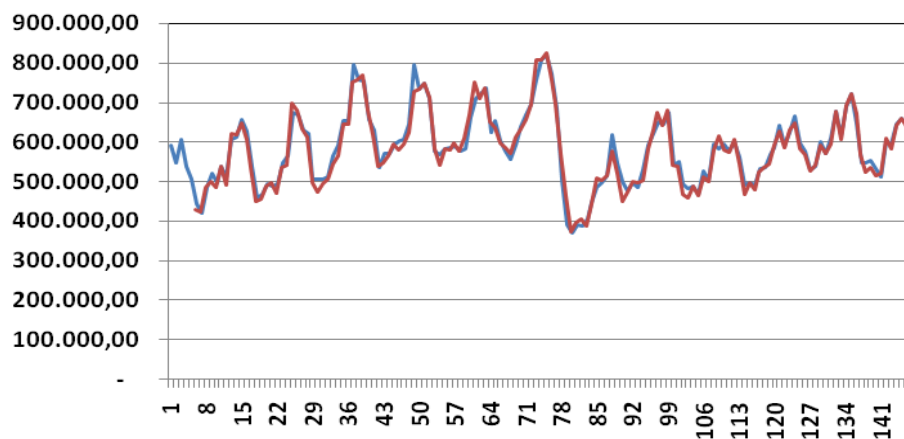


Gráfico 5.3 – Valores Reais e as Previsões Pontuais (*in sample*).

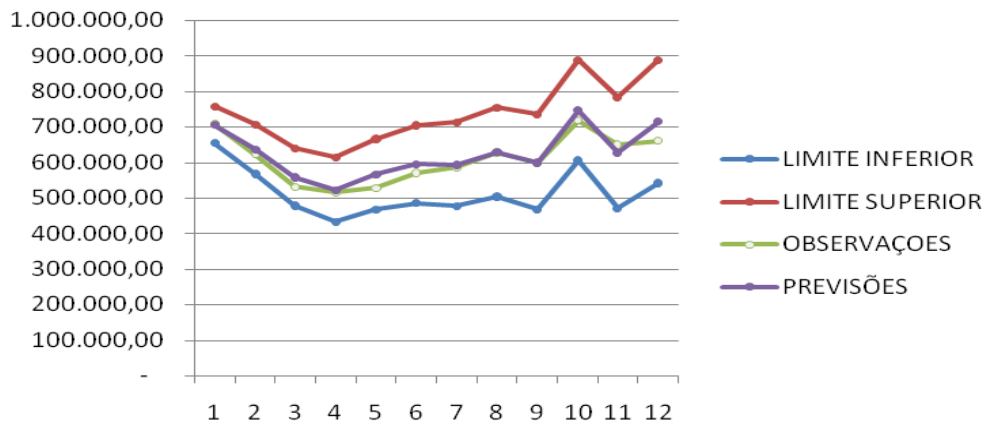


Gráfico 5.4 – Valores Reais e as Previsões Pontuais e Intervalares do Modelo de Combinação Geométrica.

Note, no gráfico 5.3, que o método proposto possui bom desempenho quanto à dinâmica passada da série considerada, pois as curvas encontram-se, a maior parte do tempo sobrepostas à das observações.

Analisando visualmente o gráfico 5.4, tem-se que:

- i. Nota-se que os pontos dos valores reais se encontram sobrepostos ao das previsões pontuais durante boa parte do horizonte considerado, mostrando novamente acurácia das previsões pontuais geométricas;
- ii. Quanto aos intervalos de confiança, verifica-se que a amplitude dos mesmos não cresce de maneira exacerbada, ou seja, a variância (incerteza) não possui um crescimento exagerado, à medida que se aumentam os lag's; e
- iii. Ainda segundo os intervalos de confiança, os mesmos mostram certa simetria entre os valores reais e os limites, superior e inferior - que é sempre desejável. Isso indica que a densidade preditiva gaussiana do modelo linear contém valores reais muito próximos seu centro de massa, ou seja, o evento que possui probabilidade máxima de acontecer, segundo o modelo, é muito próximo ao valor real, confirmando i. Pode-se dizer também que o valor real encontra-se entre os valores mais prováveis desta densidade preditiva, mostrando que os cenários produzidos são eficientes.

Conclusões

No presente trabalho, combinaram-se linearmente previsões oriundas dos modelos univariados ARIMA, Holt-Winters e Redes Neurais Artificiais, com aplicação à série de consumo residencial de energia elétrica. Para a estimação dos pesos adaptativos fixos, utilizou-se a abordagem multicritério MINIMAX, tendo como funções objetivo: MAPE, MAE e MSE. As mesmas foram individualmente otimizadas e utilizadas na análise multicritério sob forma de metas. Em síntese, calculou-se o valor esperado das previsões univariadas, utilizando uma abordagem alternativa para estimação de pesos, gerando a previsão combinada multicritério.

Os resultados do modelo de combinação linear multicritério foram superiores aos individuais, em todas as métricas de aderência consideradas.

A contribuição de cada método individual no método proposto fornece informações distintas acerca da dinâmica temporal da série analisada, de maneira que cada método individual funciona como uma componente do modelo combinado, logo a série é analisada de várias formas distintas.

A abordagem multicritério se mostrou eficiente, podendo ser utilizada em outras aplicações. Uma das características marcantes no estudo de caso proposto, foi o fato de os valores de resíduos do modelo proposto, em grande maioria, se encontrarem abaixo de um determinado nível. Em outras palavras, além de os valores de tendência central dos erros terem sido baixos, a evolução temporal destes apresentou comportamento estável. Neste aspecto, em relação ao segundo melhor método (Holt Winters), foi consideravelmente melhor.

Devido às características dos modelos, as estruturas lineares e não lineares foram bem capturadas pelo modelo multicritério. Portanto, uma ponderação adequada pode fornecer melhores previsões. Como as estimativas dos pesos se mostraram plausíveis, em termos probabilísticos, as previsões foram mais acuradas em relação aos métodos base.

Por fim, o modelo proposto obteve bom desempenho fora da amostra, o que caracteriza seu poder de generalização, no estudo proposto. Com relação à função

objetivo MAPE, por exemplo, a diferença do desempenho dentro e fora da amostra foi de 0,51%, apenas. Ou seja, o modelo de combinação linear multicritério atenua as deficiências dos métodos individuais, impactando diretamente na minimização da incerteza das projeções.

REFERÊNCIAS BIBLIOGRÁFICAS

- [1] Bates, J.M., Granger, C.W.J. Combination Forecasts. *Operations Research Quarterly*, 1969.
- [2] Box, G.E.P., Jenkins, G.M. *Time Series Analysis Forecasting and Control*. Ed. San Francisco: Holden-Day, 1976.
- [3] Faria, A.E., Mubwandarikwa, E. Multimodality on the Geometric Combination of Bayesian Forecasting Models. *International Journal of Statistics and Management System*, 3, 1-25, 2008.
- [4] Flores, B.E., White, E.M. A Framework for the Combination of Forecasts, *Journal Academic Marketing Science*, v.16, (3-4), p-95-103, 1988.
- [5] Haykin, S. *Redes Neurais Princípios e Aplicações, Segunda Edição*. Porto Alegre, 2001.
- [6] Júnior, L. A. T. *Combinação Geométrica de Métodos Preditivos; Aplicação à Série de Consumo Residencial de Energia Elétrica*. Dissertação de Mestrado. PUC-RIO, 2009.
- [7] Morettin, P.A., Toloí, L.M.C. *Análise Séries Temporais, 2ª Ed.* ABE Projeto Fisher. Ed. Edgard Blucher, 2006.
- [8] Newbold, P., Granger, C.W.J. Experience with Forecasting Univariate Times Series and the Combination of Forecasts. *Journal of the Royal Statistical Society*. v.137, p.131-165, 1974.
- [9] Ragsdale, C. *Spreadsheet Modeling & Decision Analysis: A Practical Introduction to Management Science*, fourth edition, South-Western, 2004.
- [10] Souza, R.C., Camargo, M.E. *Análise e Previsão de Séries Temporais: Os Modelos ARIMA*. Ijuí: SEDIGRAF, 1996.
- [11] Winkler, R.L., Makridakis, S. The combination of Forecasting, *Journal of the Royal Statistical Society, series A*, v.146, 1983, p.150-157, 1983.