

Combinação linear de previsões com ajuste numérico com programação não linear MINIMAX

Linear combination of forecasts with numerical adjustment via MINIMAX non-linear programming

Jairo Marlon Corrêa¹ - Universidade Federal do Paraná - Dep. de Matemática
Anselmo Chaves Neto² - Universidade Federal do Paraná - Dep. de Estatística
Luiz Albino Teixeira Júnior³ - Universidade Federal da Integração Latino-Americana - Dep. de Engenharia Elétrica
Edgar Manuel Carreño⁴ - Universidade Estadual do Oeste do Paraná - Dep. de Engenharia Elétrica
Álvaro Eduardo Faria⁵ - The Open University - Mathematics Department

RESUMO Este artigo propõe a combinação linear das previsões obtidas por três métodos de previsão (a saber, ARIMA, Amortecimento Exponencial e Redes Neurais Artificiais) cujos pesos adaptativos determinados por meio de um problema de programação não linear multiobjetivo, em que se busca minimizar, simultaneamente, as estatísticas: MAE, MAPE e MSE. Os resultados alcançados pela combinação proposta são comparados com a abordagem tradicional de combinação linear de previsões, onde os pesos adaptativos ótimos são determinados somente pela minimização do MSE; com o método de combinação por média aritmética; e com os métodos individuais.

Palavras-chave Séries Temporais. Combinação de Previsões. Programação Multiobjetivo.

ABSTRACT *This paper proposes a linear combination of forecasts obtained from three forecasting methods (namely, ARIMA, Exponential Smoothing and Artificial Neural Networks) whose adaptive weights are determined via a multi-objective non-linear programming problem, which seeks to minimize, simultaneously, the statistics: MAE, MAPE and MSE. The results achieved by the proposed combination are compared with the traditional approach of linear combinations of forecasts, where the optimum adaptive weights are determined only by minimizing the MSE; with the combination method by arithmetic mean; and with individual methods.*

Keywords *Time Series. Combination of Forecasts. Multi-objective Programming.*

1. Av. Cel. Francisco H dos Santos, s/n., Jardim das Américas, Curitiba, PR, 81530-900, jairocorrea@utfpr.edu.br
2. anselmo@ufpr.br
3. luiz.a.t.junior@gmail.com
4. pgesde@unioeste.br
5. alvaro.faria@open.ac.uk

CORRÊA, J. M.; NETO, A. C.; JÚNIOR, L. A. T.; CARREÑO, E. M.; FARIA, A. E. Combinação linear de previsões com ajuste numérico com programação não linear MINIMAX. **GEPROS. Gestão da Produção, Operações e Sistemas**, Bauru, Ano 11, nº 1, jan-mar/2016, p. 79-95.

DOI: 10.15675/gepros.v11i1.1322

1. INTRODUÇÃO

Uma informação fundamental no planejamento da expansão, na operação e na comercialização de energia é a previsão de valores futuros da demanda de eletricidade (Greer, 2011). Usualmente, ao invés de se prever o consumo total de energia elétrica de determinada região, faz-se uma segmentação de seu mercado, dividindo-o em distintas classes de consumo: residencial, industrial, comercial etc. Assim, são realizadas previsões para cada classe. Tal segmentação é justificada pelo fato de a dinâmica do consumo de cada classe apresentar características específicas

No que concerne aos métodos preditivos, a literatura técnica (ALFARES; NAZEERUDDIN, 2002; ALMESHAI EI; SOLTAN, 2011; GREER, 2011; HAHN *et al.*, 2009; LEE; HONG, 2015; SANSTAD *et al.*, 2014; WANG; MCDONALD, 1994) apresenta uma variedade de metodologias que podem ser empregadas na previsão das diferentes classes de consumo de energia elétrica. Entre os métodos estatísticos (CHATFIELD, 2004; HAMILTON, 1994), têm-se as abordagens: univariada (Amortecimento Exponencial e Box & Jenkins), multivariados (Vetores Auto-Regressivos) e causais (Função de Transferência e Regressão Dinâmica). Quanto aos métodos de Inteligência Artificial (PALIT; POPOVIC, 2005), destacam-se as Redes Neurais Artificiais (RNA) e a Lógica *Fuzzy*. A variedade de técnicas aplicadas na previsão da demanda de energia elétrica reflete a busca por métodos capazes de gerar previsões mais acuradas, de forma a minimizar a incerteza e, com isso, permitir um melhor uso da capacidade instalada do sistema.

Alternativamente, a projeção de séries de tempo pode ser realizada através da combinação de previsões obtidas por diferentes formas. Neste contexto, Bates e Granger (1969) propõem a combinação linear das previsões obtidas por dois modelos estatísticos preditivos, ao invés de adotar-se unicamente algum modelo individual. Na combinação linear de métodos preditivos, as previsões de cada método individual são ponderadas por pesos determinados, na maioria das vezes, de forma a minimizar o erro médio quadrático (MSE – *mean square error*) da previsão combinada.

Não obstante, o MSE é apenas uma das formas de avaliar a qualidade do ajuste de um método de previsão. Outros critérios, usualmente empregados com esta finalidade, são o MAPE (*mean absolute percentual error*) e o MAE (*mean absolute error*). Assim, neste artigo, propõe-se a combinação linear de previsões com pesos determinados mediante a otimização multiobjetivo MINIMAX (RAGSDALE, 2004).

A implementação do método proposto é ilustrada com uma série temporal mensal do consumo de energia elétrica da classe residencial. A classe residencial concentra uma parcela expressiva do consumo e a maior parte das unidades consumidoras. Salienta-se ainda que nesta classe de consumo são encontradas diversas políticas públicas do setor de energia elétrica, como a universalização do atendimento, a política de subsídios aos consumidores de baixa renda e os programas de aumento da eficiência dos eletrodomésticos (LEON; PESSANHA, 2005).

Inicialmente, as previsões do consumo mensal foram obtidas, separadamente, por três métodos distintos: ARIMA, Amortecimento Exponencial e Redes Neurais Artificiais. Na sequência, as três previsões foram combinadas linearmente com pesos determinados através de um modelo de programação matemática multiobjetivo, tendo como funções objetivo a minimização das estatísticas: MAE, MAPE e MSE. Adicionalmente, os resultados obtidos pelo método de combinação proposto foram comparados com os encontrados pela abordagem tradicionalmente empregada na combinação linear de previsões, em que os pesos atribuídos às previsões individuais são determinados através da minimização do MSE.

O artigo está organizado em seis seções. A seguir, na seção 2 tem-se uma breve descrição dos três métodos individuais para previsão de séries temporais considerados neste trabalho. A abordagem usual empregada na combinação linear de previsões é apresentada na seção 3; enquanto a abordagem multiobjetivo proposta é descrita na seção 4. Os resultados obtidos por cada método de previsão, aplicados individualmente, e pelas combinações de previsões são apresentados na seção 5. Por fim, na seção 6, são resumidas as principais conclusões do trabalho.

2. MÉTODOS DE PREVISÃO

2.1. Amortecimento Exponencial de Holt-Winters

Os métodos de amortecimento exponencial de Morettin e Toloí (2006) baseiam-se na premissa de que as observações recentes contêm informações mais relevantes que as antigas, de forma que a ponderação dos dados (série temporal) decresce exponencialmente à medida que a observação torna-se mais antiga.

Um caso particular dos métodos de amortecimento exponencial é o Holt-Winters multiplicativo, o qual realiza o mapeamento, de forma dinâmica no tempo, das componentes de nível ($a_{1,t}$), tendência ($a_{2,t}$) e sazonalidade ($\rho_{m(t)}$), conforme descrito a seguir:

$$\begin{cases} y_t = \mu_t + \varepsilon_t \\ y_t = [a_{1,t} + a_{2,t} \times t] \times \rho_{m(t)} + \varepsilon_t \end{cases} \quad (1)$$

onde os fatores sazonais têm de obedecer à restrição $\sum_{j=1}^L \rho_j(t) = L$, sendo L o tamanho do ciclo sazonal. No processo de estimação dos parâmetros da equação (1), as observações da série temporal são ponderadas por três hiperparâmetros (quantidades fixas no tempo): α , β e λ , que estão associados, respectivamente, às estimativas de nível, tendência e sazonalidade (quantidades flutuantes no tempo). Os hiperparâmetros assumem valores entre 0 e 1, com o objetivo de minimizar a estatística MSE.

2.2. Metodologia de Box & Jenkins

Um processo estocástico gaussiano estacionário de segunda ordem pode ser definido como uma família de variáveis aleatórias normalmente distribuídas cujos momentos (média, variância e covariância) são invariantes no tempo, isto é: $Y_t \sim N(\mu, \sigma^2), \forall t \in Y$. Um processo estocástico estacionário é classificado como ergódico, uma vez que uma realização do mesmo é suficiente para se obter todas as estatísticas que o descrevem (CHATFIELD, 2004; MORETTIN; TOLOI, 2006).

Assumindo-se alguma série temporal arbitrária, gerada por um processo estocástico estacionário de segunda ordem, sendo a observação no instante t dada por y_t , Box & Jenkins (2008) propõem o modelo em para o mapeamento de sua dinâmica temporal.

$$y_t = \varphi y_{t-1} + \dots + \varphi_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q} \quad (2)$$

O modelo (2) pode ser representado em termos de dois polinômios obtidos através do operador defasagem dado por $B^d Y_t = Y_{t-d}$. Então:

$$(1 - \varphi_1 B - \dots - \theta_p \varphi^p) y_t = (1 - \theta_1 B - \dots - \theta_q B^q) \varepsilon_t \quad (3)$$

onde $\phi_k, \forall k \geq 1$, e $\theta_j, \forall j \geq 1$, denotam os parâmetros do modelo ARIMA e ε_t , o erro em t.

Em (3), as defasagens de y_t representam a parte autoregressiva de ordem p – AR (p); enquanto as defasagens dos erros representam a parte média móvel de ordem q – MA (q). A equação (3) representa uma ampla classe de modelos denominada de ARMA (p,q). O exame de estacionariedade pode ser realizado através da análise da função de autocorrelação simples (ρ_k):

$$\hat{\rho}_k = \frac{\sum_{t=k+1}^T (y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=1}^T (y_t - \bar{y})^2} \quad (4)$$

onde T é o número de observações da série temporal de interesse, \bar{y} é a média da série temporal e k é a defasagem (lag) da autocorrelação.

A determinação das ordens p e q em é realizada por meio de uma análise do perfil das funções de autocorrelação (ρ_k) – ACF (*autocorrelation function*) – e de autocorrelação parcial (ϕ_{kk}) – PACF (*partial autocorrelation function*) –, cujos padrões teóricos podem ser verificados em Souza e Camargo (1996) e Hanke e Wichern (2006). Assim, uma vez identificada a ordem do modelo, são estimados os coeficientes das partes AR (p) e MA (p) e avaliadas as respectivas significâncias estatísticas (Box; Jenkins, 2008). Caso a série temporal em análise apresente tendência, a mesma deve ser diferenciada d vezes a fim de torná-la estacionária. Por exemplo, a diferenciação de primeira ordem, definida por $\Delta y_t = y_t - y_{t-1}$. Assim, a série temporal resultante das diferenciações é passível de ser modelada pela abordagem de Box & Jenkins, mais precisamente pelo modelo ARIMA (p,d,q):

$$(1 - \varphi_1 B - \dots - \theta_p \varphi^p)(1 - B)^d y_t = (1 - \theta_1 B - \dots - \theta_q B^q) \varepsilon_t \quad (5)$$

A metodologia Box & Jenkins também tem modelos que podem ser aplicados em séries temporais com sazonalidade. Analogamente aos polinômios AR (p) e MA (q), a componente sazonal pode ser modelada por polinômios AR (P) e MA (Q), cujas ordens P e Q são identificadas por meio da análise do perfil das funções de autocorrelação simples e parcial. A análise das funções de autocorrelação também permite identificar a necessidade de fazer a diferenciação sazonal de ordem D de uma série temporal. Em síntese, para as séries temporais sazonais, tem-se o modelo Box & Jenkins conhecido como ARIMA multiplicativo SARIMA (p,d,q)x(P,D,Q) S, cuja equação é descrita em (6).

$$(1 - \varphi_1 B - \dots - \varphi_p B^p)(1 - \Phi_1 B^S - \dots - \Phi_P B^{PS})(1 - B^S)^D (1 - B^d) y_t = (1 - \theta_1 B - \dots - \theta_q B^q)(1 - \Theta_1 B^S - \dots - \Theta_Q B^{QS}) \varepsilon_t \quad (6)$$

onde:

- p e q são, respectivamente, os graus dos polinômios AR e MA da componente não sazonal;
- P e Q são, respectivamente, os graus dos polinômios AR e MA da componente sazonal;
- d é a ordem das diferenças simples para remover a tendência da série;
- D é a ordem das diferenças sazonais para remover a sazonalidade;
- S é o período sazonal, por exemplo, para séries mensais $S=12$;
- $\phi_k, \forall k \geq 1$, e $\theta_j, \forall j \geq 1$, são, respectivamente, os coeficientes dos polinômios AR e MA da parte não sazonal; e
- $\Phi_n, \forall n \geq 1$, e $\Theta_m, \forall m \geq 1$, são, respectivamente, os coeficientes dos polinômios AR e MA da componente sazonal.

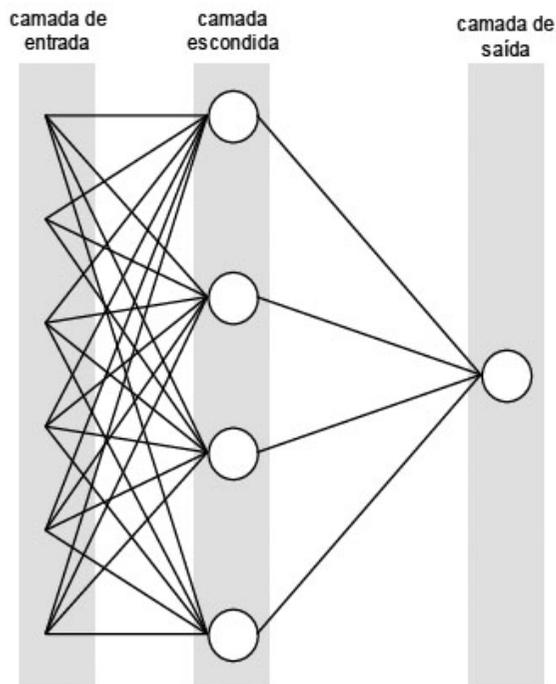
2.3 Redes Neurais Artificiais

As redes neurais artificiais (RNA) são sistemas paralelos distribuídos, compostos por unidades de processamento simples denominados neurônios e dispostas em uma ou mais camadas interligadas por um grande número de conexões (sinapses) geralmente unidirecionais e com pesos para ponderação das entradas recebidas por cada neurônio (HAYKIN, 2001).

A arquitetura mais usual é a rede neural *Multi Layer Perceptron* (MLP) com três camadas, conforme ilustra a Figura 1. De forma análoga ao processamento no cérebro humano (onde as sinapses são reforçadas ou enfraquecidas), os pesos nas conexões entre as camadas são ajustados durante o processo de aprendizagem da rede neural. A primeira camada da rede é a camada de entrada, sendo a única exposta às variáveis de entrada. Esta camada transmite os valores das variáveis de entrada para os neurônios da camada escondida para que estes extraiam as características relevantes ou padrões dos sinais de entrada e transmitam os resultados para a camada de saída, a última camada da rede. A definição do número de neurônios em cada camada é feita de forma empírica. O treinamento de uma rede neural consiste em um processo iterativo de ajuste dos pesos das conexões entre as unidades de processamento.

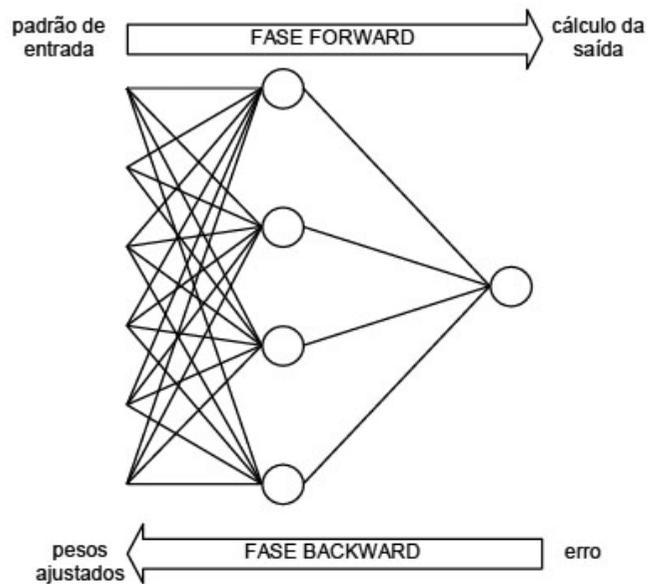
O principal algoritmo de treinamento é o *backpropagation* (PALIT; POPOVIC, 2005) cujo ajuste dos pesos ocorre mediante um processo de otimização realizado em duas fases: *forward* e *backward*, conforme ilustrado na Figura 2. Na fase *forward*, é calculada a resposta fornecida pela rede para um dado padrão de entrada. Na fase *backward*, o desvio (erro) entre a resposta desejada (alvo) e a resposta fornecida pela RNA é utilizado para o ajuste de pesos das conexões. Ao longo do treinamento da rede neural, os vários padrões de entrada e as respectivas saídas desejadas são apresentados à RNA, de forma que os pesos das sinapses sejam corrigidos com os erros, de forma iterativa, pelo algoritmo do gradiente descendente, com a finalidade de minimizar a soma dos quadrados dos erros (HAYKIN, 2001).

Figura 1 – Rede Neural Artificial *Multi-layer Perceptron* (Arquitetura da Rede Neural).



Fonte: Haykin (2001).

Figura 2 – Rede Neural Artificial *Multi-layer Perceptron* (Fases do Algoritmo *Backpropagation*).



Fonte: Haykin (2001).

3. COMBINAÇÃO LINEAR DAS PREVISÕES VIA ABORDAGEM TRADICIONAL

Flores & White (1988), *apud in* Albino (2009), destacam dois aspectos básicos na combinação de previsões: (I) seleção das previsões a serem combinadas; e (II) escolha do método de combinação. As previsões a serem combinadas (obtidas por métodos individuais) podem ser classificadas em duas categorias: objetivas e subjetivas. As previsões objetivas incluem aquelas advindas dos métodos estatísticos e de inteligência artificial; enquanto, as subjetivas englobam aquelas oriundas de julgamento humano (opinião de especialistas, por exemplo). Quanto aos métodos de combinação, destaca-se a combinação linear das previsões.

Bates e Granger (1969) propuseram a combinação linear de duas previsões objetivas não tendenciosas oriundas de dois modelos estatísticos clássicos:

$$\hat{y}_{CL,t} = [\hat{y}_{1,t} \times \hat{p}_1] + [\hat{y}_{2,t} \times \hat{p}_2] \quad (7)$$

onde a previsão combinada $\hat{y}_{CL,t}$ é dada pela média ponderada das previsões do modelo 1 ($\hat{y}_{1,t}$) e do modelo 2 ($\hat{y}_{2,t}$), em que os pesos \hat{p}_1 e \hat{p}_2 somam um e são determinados de forma a minimizar a estatística MSE (ou seja, a média dos desvios quadráticos entre os valores observados e as respectivas previsões, definidas por $\hat{y}_{CL,t}$).

Newbold e Granger (1974) combinaram três métodos estatísticos clássicos – mantendo, porém, as suposições de Bates e Granger (1969) – e concluíram que, em regra, as previsões combinadas são mais acuradas. Winkler e Makridakis (1983) analisaram combinações de modelos três a três entre dez modelos estatísticos de previsão e os resultados obtidos confirmaram as conclusões de Newbold e Granger (1974). De acordo com Aiolfi e Timmermann (2006); De Luca e Carfora (2014); Faria e Mubwandarikwa (2008) e Stock e Watson (2004); e as combinações de previsões podem promover ganho preditivo fora da amostra (*out of sample*).

Assim sendo, neste artigo, as previsões a serem combinadas de forma linear são obtidas por meio dos métodos de Box-Jenkins (BJ), Holt-Winters (HW) e Redes Neurais Artificiais (RNA). A previsão combinada, para cada instante t , pode ser interpretada como a média ponderada das previsões destes métodos individuais, onde a soma dos pesos \hat{p}_{BJ} , \hat{p}_{HW} , \hat{p}_{RNA} é igual a um, e é definida como:

$$\hat{y}_{CL,t} = [BJ_t \times \hat{p}_{BJ}] + [HW_t \times \hat{p}_{HW}] + [RNA_t \times \hat{p}_{RNA}] \quad (8)$$

Na abordagem tradicional os pesos são definidos de forma a minimizar a estatística MSE. Contudo, a avaliação da qualidade do ajuste de um modelo não se resume ao MSE, pois há outras estatísticas como, por exemplo, o MAE e o MAPE. Assim, propõe-se determinar os pesos \hat{p}_{BJ} , \hat{p}_{HW} , \hat{p}_{RNA} de maneira a minimizar, simultaneamente, as estatísticas MSE, MAE e MAPE, mediante otimização multiobjetivo com abordagem de programação por metas (descrita na seção 4).

4. COMBINAÇÃO LINEAR DAS PREVISÕES VIA MINIMAX

O problema de programação matemática (PPM) consiste em encontrar os valores para as variáveis de decisão (X_1, \dots, X_n) que correspondam aos extremos (ou aos ótimos) de um ou mais objetivos. Os objetivos são representados por funções, denominadas funções objetivo e denotadas por $f(X_1, \dots, X_n)$.

Em um PPM envolvendo mais de uma função objetivo, o procedimento de busca de soluções ótimas é denominado otimização multiobjetivo. Em particular, na combinação linear das previsões, os pesos p_{BJ} , p_{HW} , p_{RNA} são as variáveis de decisão e as estatísticas MSE, MAE e MAPE são as funções objetivo a serem minimizadas simultaneamente. Ou seja, os pesos ótimos são tais que assumem valores que minimizam as estatísticas MSE, MAPE e MAE.

Uma das possíveis abordagens para um PPM multiobjetivo denomina-se programação matemática por metas, que consiste em identificar a solução ótima que minimiza os desvios entre os valores das funções objetivos e suas respectivas metas pré-definidas (RAGSDALE, 2004). O objetivo da programação por metas pode ser definido, por exemplo, como a soma ponderada dos desvios percentuais absolutos, conforme em (9).

$$\text{Min } \omega_{MAPE} \frac{MAPE - Meta_{MAPE}}{Meta_{MAPE}} + \omega_{MAE} \frac{MAE - Meta_{MAE}}{Meta_{MAE}} + \omega_{MSE} \frac{MSE - Meta_{MSE}}{Meta_{MSE}} \quad (9)$$

Os coeficientes (pesos) ω_{MAPE} , ω_{MAE} e ω_{MSE} são constantes que podem ser fixadas pelo decisor (atribuindo preferências distintas em relação às estatísticas de ajuste MAPE, MAE e MSE) ou interpretadas como variáveis de decisão (devendo-se, assim, encontrar os seus valores ótimos). As metas indicadas na função objetivo são valores de referência obtidos, comumente, por otimização individual de cada função objetivo.

As metas são definidas com o auxílio da abordagem tradicional da combinação linear de previsões. Assim, o MSE resultante da combinação linear ótima (descrita na seção 3) é a meta a ser alcançada pelo objetivo MSE no PPM multiobjetivo. Da mesma forma, a meta do MAE corresponde ao MAE resultante da abordagem tradicional de combinação linear, porém tendo como critério a minimização do MAE e não mais do MSE. Por sua vez, a meta para o MAPE é obtida por meio da combinação linear de previsões que minimiza o MAPE. Em outras palavras, a abordagem tradicional foi aplicada três vezes, sendo que em cada vez se considerou um critério (estatística) diferente.

Um PPM multiobjetivo pode ser resolvido também através da otimização MINIMAX (RAGSDALE, 2004) com programação por metas, de forma que uma função objetivo Q impõe restrições aos desvios relativos entre as estatísticas adotadas e as suas respectivas metas, definidas como: $[\text{meta}_j - FO_{\text{meta}_j} / \text{meta}_j] < Q$, sendo $j = 1, 2, 3$. Ou seja, o valor de Q corresponde ao mínimo valor para o máximo desvio relativo entre as estatísticas e seus respectivos valores de referências (metas).

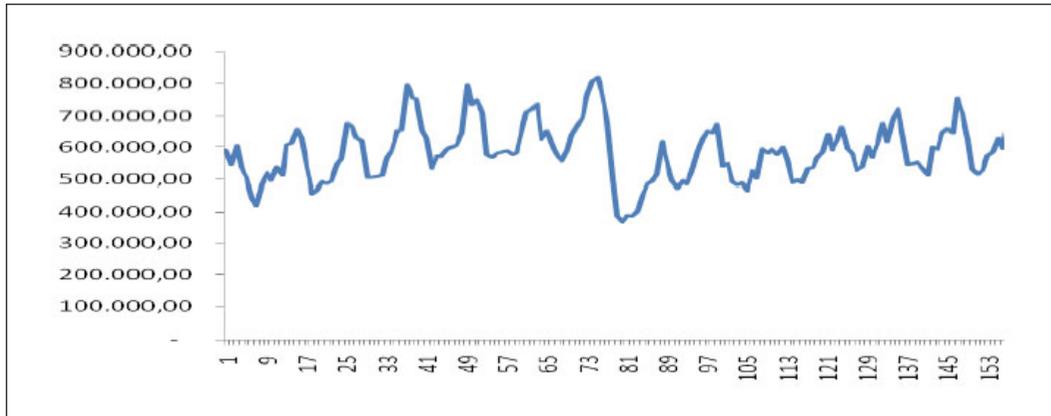
Tem-se, a seguir, o PPM proposto para a combinação linear das previsões com ajuste MINIMAX, com abordagem de programação por metas, implementado no *software* AIMMS (<http://www.aimms.com/>):

$$\begin{aligned}
 & \underset{P_{BJ}, P_{HW}, P_{RNA}, Q}{\text{Min}} \quad \{Q\} \\
 & \text{s.a.} \\
 & \left\{ \begin{aligned}
 & \omega_{MAPE} \frac{MAPE - Meta_{MAPE}}{Meta_{MAPE}} \leq Q \\
 & \omega_{MAE} \frac{MAE - Meta_{MAE}}{Meta_{MAE}} \leq Q \\
 & \omega_{MSE} \frac{MSE - Meta_{MSE}}{Meta_{MSE}} \leq Q
 \end{aligned} \right. \\
 & \left\{ \begin{aligned}
 & \hat{y}_{CL,t} = p_{BJ} \times \hat{y}_{BJ,t} + p_{HW} \times \hat{y}_{HW,t} + p_{RNA} \times \hat{y}_{RNA,t} \\
 & p_{BJ} + p_{HW} + p_{RNA} = 1
 \end{aligned} \right. \tag{10} \\
 & \left\{ \begin{aligned}
 & MAPE = \frac{1}{T} \sum_{t=1}^T \frac{|\hat{y}_{CL,t} - y_t|}{y_t} \\
 & MAE = \frac{1}{T} \sum_{t=1}^T |\hat{y}_{CL,t} - y_t| \\
 & MSE = \frac{1}{T} \sum_{t=1}^T (\hat{y}_{CL,t} - y_t)^2
 \end{aligned} \right. \\
 & \left\{ \begin{aligned}
 & 0 \leq p_{BJ} \leq 1 \\
 & 0 \leq p_{HW} \leq 1 \\
 & 0 \leq p_{RNA} \leq 1
 \end{aligned} \right.
 \end{aligned}$$

5. EXPERIMENTO COMPUTACIONAL

No experimento computacional, considerou-se a série temporal mensal de consumo residencial de energia elétrica de uma distribuidora de eletricidade que atende parte do estado do Rio de Janeiro. As observações da Figura 3 compreendem o período de janeiro de 1995 até março de 2008, totalizando 159 pontos.

Figura 3 – Srie mensal do consumo residencial de energia eltrica.



Fonte: Elaborado pelo autores (2015).

A amostra considerada na estimaco dos modelos previso (*in sample*)  composta pelas 147 primeiras observaces e a amostra de teste (*out of sample*), pelas 12 observaces subsequentes. Em funo da diversidade de mtodos utilizados, a implementaco computacional contou com o apoio de diferentes *softwares*: MATLAB 8, *E-Views*, FPW, AIMMS e Excel.

Inicialmente, fez-se a modelagem da srie temporal mensal do consumo residencial de energia eltrica, com a finalidade de produzir previses dentro e fora da amostra (12 passos  frente). Nesta modelagem, os mtodos Box-Jenkins, Holt-Winters e Redes Neurais Artificiais foram ajustados individualmente, obtendo-se as previses de cada um, a cada instante *t*. Em seguida, fez-se a combinaco linear das previses, conforme a abordagem tradicional, mas segundo diferentes critrios. Inicialmente, os pesos foram determinados mediante a minimizaco da estatstica MSE. Na sequncia, foi realizada a combinaco linear de previses segundo o critrio de minimizaco do MAE e, por ltimo, a combinaco linear resultante da minimizaco do MAPE. Ainda fez-se a combinaco de previses por mdia simples (que , de fato, um caso particular da combinaco linear de previses). Os valores das estatsticas MSE, MAPE e MAE resultantes de cada combinaco linear so interpretadas como as metas para o PPM multiobjetivo MINIMAX, descrito na seo 4. Por fim, com o auxlio do *software* AIMMS, as previses individuais so novamente combinadas de forma linear e os valores timos dos pesos adaptativos so determinados pela soluo do PPM multiobjetivo descrito na seo 4.

5.1 Modelo Holt-Winters

A seguir, na Tabela 1, têm-se os valores dos hiperparâmetros e dos doze fatores sazonais referentes ao modelo de Holt-Winters multiplicativo ajustado.

Tabela 1 – Valores dos hiperparâmetros das componentes e dos fatores sazonais.

Hiperparâmetros			
Nível	0,51347	Sazonalidade	0,49758
Fatores Sazonais			
Mês	Valor	Mês	Valor
Janeiro	1.07286	Julho	0.89065
Fevereiro	1.10144	Agosto	0.87161
Março	1.18919	Setembro	0.89276
Abril	1.12426	Outubro	0.94374
Maiο	1.01458	Novembro	0.96591
Junho	0.94362	Dezembro	1.04651

Fonte: Elaborado pelo autores (2015).

5.2 Modelo Box-Jenkins

Na família Box-Jenkins, o modelo SARIMA (1,0,0)x(1,0,3)₁₂ foi o mais aderente à série temporal do consumo da classe residencial mensal. Na Tabela 2, dado o p-valor da estatística de *Jarque-Bera* é superior a 5% (nível de significância), não se rejeita a hipótese nula de normalidade dos resíduos; além disso, os coeficientes de curtose e assimetria observados também evidenciam normalidade dos erros de previsão, visto que são próximos aos valores teóricos (a saber, 4 e 0, respectivamente).

Tabela 2 – Estatísticas para avaliação da normalidade dos resíduos.

Jarque-Bera	P-valor	Curtose	Assimetria
4,713351	9,4735%	3,763440	-0,271137

Fonte: Elaborado pelo autores (2015).

Na Tabela 3, são exibidas as estimativas de mínimos quadrados dos coeficientes do modelo SARIMA(1,0,0)x(1,0,3)₁₂, bem como as estatísticas que fornecem informações para a avaliação das respectivas significâncias estatísticas. Note que todos os coeficientes são estatisticamente significantes, a 5% de significância (pois, p-valor < 0,025).

Tabela 3 – Estimativas e estatísticas de significância.

Termo	Coefficiente	Erro padrão	Estatística t	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
Termo	Coefficiente	Erro padrão	Estatística t	P-valor
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

Fonte: Elaborado pelo autores (2015).

Na Tabela 4, o valor da estatística *Durbin-Watson* evidencia que os resíduos não apresentam autocorrelação de lag 1, pois obteve valor observado próximo ao teórico, a saber, 2. O p-valor da estatística de *Ljung-Box*, por sua vez, evidencia que os resíduos não são autocorrelacionados até o lag 20.

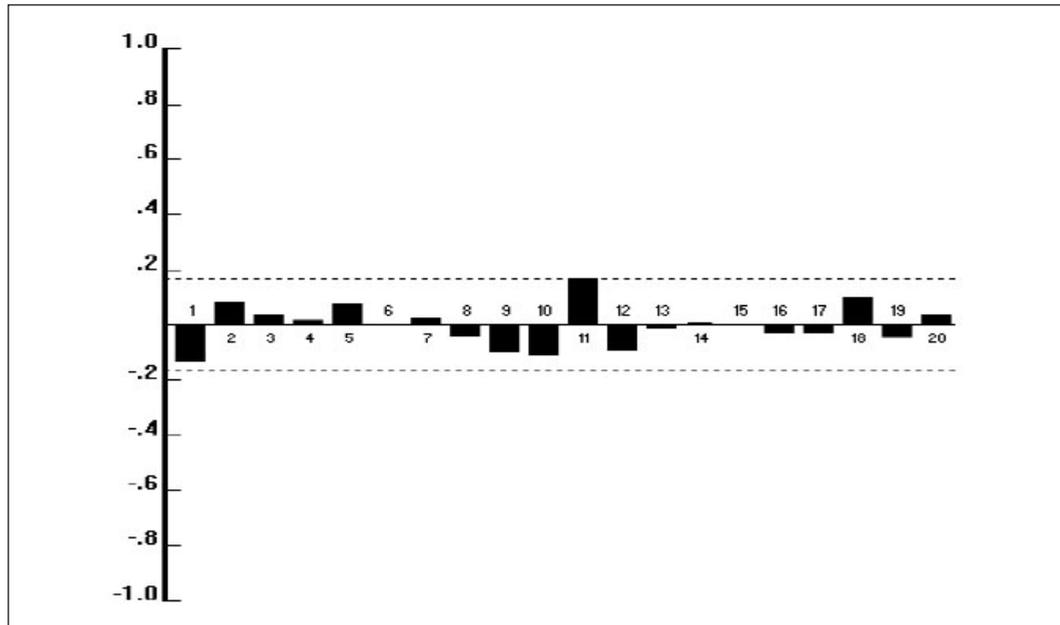
Tabela 4 – Principais estatísticas de aderência.

<i>Durbin-Watson</i>	<i>Ljung-Box</i> (p-valor)	BIC
2,1395	76,81%	4,43E+004

Fonte: Elaborado pelo autores (2015).

Os resultados da Tabela 4 são confirmados pelo gráfico da função de autocorrelação (FAC) dos erros de previsão, na Figura 4. Note que os valores das autocorrelações são significativamente, a 5%, não diferentes de zero. O critério de informação BIC (*bayesian information criterion*), na Tabela 4, foi o menor entre todas as diferentes ordens de modelo avaliadas.

Figura 4 – Correlograma da FAC dos resíduos.



Fonte: Elaborado pelo autores (2015).

5.3. Redes neurais artificiais

A RNA foi ajustada com a finalidade de fornecer previsões um passo à frente, tendo como configuração três observações de entrada (isto é, o consumo residencial de energia elétrica nos instantes $t-1$, $t-2$ e $t-3$); uma camada escondida com quatro neurônios com tangente hiperbólica como função de ativação; e um neurônio na camada de saída com função de ativação linear. O treinamento da RNA foi realizado utilizando-se o algoritmo *Levenberg Marquardt*, disponível no MATLAB 8, e aplicado aos dados normalizados no intervalo $[-1,1]$, via transformação *premnmx* (HAYKIN, 2001).

5.4. Análise de desempenho

Na Tabela 5, tem-se que os métodos de combinação de previsões apresentaram desempenhos superiores aos métodos ARIMA, Holt-Winters e Redes Neurais Artificiais, quando tomados separadamente em todas as estatísticas, dentro e fora da amostra. Dentre os métodos de combinação, o de combinação linear tradicional foi ligeiramente superior na amostra *in sample*. No entanto, na amostra *out of sample*, o de combinação linear com ajuste via otimização MINIMAX apresentou um desempenho superior em todas as estatísticas, em relação a qualquer método de previsão (individual ou de combinação), mostrando maior poder de generalização para projeção do consumo residencial mensal de energia elétrica. Salienta-se ainda que, na amostra *out of sample*, a combinação linear tradicional apresentou um desempenho ligeiramente inferior ao alcançado pela média aritmética.

Tabela 5 – Comparação de desempenho.

MÉTODO	<i>In Sample</i>			<i>Out of Sample</i>		
	MAPE	MAE	R ²	MAPE	MAE	R ²
ARIMA	5,66%	32.684,26	80,12%	7,90%	45.354,33	30,40%
Holt-Winters	5,82%	32.760,43	77,96%	5,90%	32.865,21	61,74%
Redes Neurais Artificiais	4,63%	26.801,18	85,31%	7,01 %	37.057,95	47,51%
Combinação Linear (Tradicional)	2,30%	13.231,76	95,53%	3,30%	19.720,73	84,53%
Média Aritmética	3,09%	17.586,29	93,12%	3,30%	19.145,75	84,81%
Combinação Linear (MINIMAX)	2,35%	13.515,21	95,47%	2,82%	17.145,99	86,93%

Fonte: Elaborado pelo autores (2015).

Na Tabela 6, verifique que o método MINIMAX aplicado à combinação linear favoreceu no processo preditivo fora da amostra, sendo o método proposto superior aos dois métodos de combinação (isto é, obteve os menores valores para o APE – *Absolute Percent Error*), em seis meses; além disso, o método proposto, em relação à média simples, obteve valores de APE mais próximos de zero em 7 meses. Em relação à combinação linear tradicional, o método proposto apresentou um desempenho superior em 10 meses. Destaca-se ainda que a combinação linear com ajuste MINIMAX forneceu valores de APE foram inferiores a 2% durante 6 meses.

Tabela 6 – Evolução do APE fora da amostra dos métodos de combinação.

Data	ARIMA	Holt-Winters	RNA	Média Simples	Combinação Linear (Clássico)	Combinação Linear (Multiobjetivo)
2007-12	3,23%	1,28%	3,90%	0,65%	0,43%	1,13%
2008-01	6,95%	1,55%	1,98%	2,17%	2,29%	1,52%
2008-02	16,67%	10,61%	4,69%	7,53%	5,24%	4,39%
2008-03	17,67%	7,71%	12,00%	4,46%	1,59%	0,45%
2008-04	13,49%	2,85%	2,03%	6,12%	7,33%	6,13%
2008-05	6,76%	2,56%	2,03%	2,08%	4,28%	3,21%
2008-06	6,30%	0,38%	2,97%	1,24%	1,44%	0,63%
2008-07	0,75%	4,08%	0,25%	1,19%	0,42%	0,30%
2008-08	11,71%	9,13%	9,39%	3,82%	0,50%	0,06%
2008-09	8,06%	6,96%	13,27%	0,58%	3,95%	3,69%
2008-10	0,07%	5,56%	7,43%	0,60%	3,53%	3,45%
2008-11	2,90%	12,19%	11,71%	9,16%	8,56%	8,90%

Fonte: Elaborado pelo autores (2015).

Na Tabela 7, são apresentados os pesos ótimos atribuídos aos métodos de previsão pela combinação linear tradicional e pela combinação linear proposta (MINIMAX).

Tabela 7 – Pesos atribuídos aos métodos de previsão nas distintas combinações.

Método de Combinação	Métodos Individuais		
	ARIMA	HW	RNA
Tradicional	0,428921453	0,028227626	0,54633457
Multiobjetivo (MINIMAX)	0,350393793	0,104334257	0,54527195

Fonte: Elaborado pelo autor (2015).

Estes pesos adaptativos ótimos minimizam, simultaneamente (via abordagem MINIMAX), as estatísticas de aderência MSE, MAPE e MAE. A abordagem de combinação MINIMAX atribuiu maior relevância ao método de Holt-Winters (HW) que o tradicional. Com efeito, menor ao ARIMA. Quanto à RNA, houve sensível perda de relevância.

6. CONCLUSES

No presente artigo, prope-se a combinaco linear de previses oriundas dos mtodos individuais: ARIMA, Holt-Winters e Redes Neurais Artificiais, para projecco mensal da srie temporal de consumo residencial de energia eltrica, na qual a estimaco dos pesos ocorre atravs de um problema de programaco multiobjetivo, cujo intuito  a minimizaco simultnea das funcces objetivo (estatsticas) MAPE, MAE e MSE.

O mtodo proposto para combinaco linear de previses apresentou um desempenho superior aos mtodos individuais, dentro e fora da amostra. A contribuico de cada mtodo individual no mtodo proposto forneceu informaces distintas acerca da dinmica temporal da srie em anlise, onde cada mtodo individual funciona como uma componente do modelo combinado. Adicionalmente, no perodo fora da amostra, a combinaco linear de previses com pesos determinados pela otimizaco multiobjetivo MINIMAX (com programaco por metas) apresentou o melhor desempenho em todas as estatsticas de qualidade de ajuste, ou seja, a proposta de combinaco de previses descrita no artigo apresentou maior capacidade de generalizaco.

REFERNCIAS

- AIOLFI, M.; TIMMERMANN, A. Persistence in forecasting performance and conditional combination strategies. **Journal of Econometrics**, v. 135, p. 31-53, 2006.
- ALFARES, H. K.; NAZEERUDDIN, M. Electric load forecasting: literature survey and classification of methods. **International Journal of Systems Science**, v. 33, n. 1, p. 23-34, 2002.
- ALMESHAI EI, E.; SOLTAN, H. A methodology for Electric Power Load Forecasting. **Alexandria Engineering Journal**, v. 50, p. 137-144, 2011.
- BATES, J. M.; GRANGER, C. W. J. Combination Forecasts. **Operations Research Quarterly**, v. 20, n. 4, p. 451-468, 1969.
- BOX, G. E. P.; JENKINS, G. M.; REINSEL, G. C. **Time Series Analysis: Forecasting and Control**, 4. ed. Hoboken: Wiley, 2008, p. 784.
- CHATFIELD, C. **The Analysis of Time Series: An introduction**. 6 ed. Florida: Chapman & Hall, 2004.
- DE LUCA, G.; CARFORA, A. Predicting U.S. recessions through a combination of probability forecasts. **Empirical Economics**, v. 46, p. 127-144, 2014.
- FARIA, A. E.; MUBWANDARIKWA, E. Multimodality on the Geometric Combination of Bayesian Forecasting Models. **International Journal of Statistics and Management System**, v. 3, p. 1-25, 2008.
- FLORES, B. E.; WHITE, E. M. A Framework for the Combination of Forecasts. **Journal Academic Marketing Science**, v. 16, n. 3, p. 95-103, 1988.

- GREER, M. **Electricity Cost Modeling Calculations**. Academic Press: Oxford, 2011.
- HAHN, H.; MEYER-NIEBERG, S.; PICKL, S. Electric load forecasting methods: Tools for decision making. **European Journal of Operational Research**, v. 199, p. 902-907, 2009.
- HAMILTON, J. **Time Series Analysis**. Princeton University Press, 1994.
- HANKE, J. F.; WICHERN, D. W. **Pronósticos en los negocios**. 8 ed. México: Pearson Educación, 2006.
- HAYKIN, S. **Redes Neurais Princípios e Aplicações**. 2 ed. Porto Alegre, 2001.
- LEE, W. J.; HONG, J. A hybrid dynamic and fuzzy time series model for mid-term power load forecasting. **International Journal of Electrical Power & Energy Systems**, v. 64, p. 1057-1062, 2015.
- LEON, N.; PESSANHA, J. F. M. Dinâmica da evolução do consumo de energia elétrica no setor residencial. *In: XVIII SNPTEE - Seminário Nacional de Produção e Transmissão de Energia Elétrica*, 2005. **Anais...** Curitiba, 2005.
- MORETTIN, P. A.; TOLOI, L. M. C. **Análise Séries Temporais**, 2ª Ed. ABE Projeto Fisher: Edgard Blucher, 2006.
- 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.
- PALIT, A. K.; POPOVIC, D. **Computational Intelligence in Time Series Forecasting**. London: Springer-Verlag, 2005.
- RAGSDALE, C. **Spreadsheet Modeling & Decision Analysis: A Practical Introduction to Management Science**. 4 ed. South-Western, 2004.
- SANSTAD, A. H.; MCMENAMIN, S.; SUKENIK, A.; BARBOSE, G. L.; GOLDMAN, C. A. Modeling an aggressive energy-efficiency scenario in long-range load forecasting for electric power transmission planning. **Applied Energy**, v. 128, p. 265-276, 2014.
- SOUZA, R. C.; CAMARGO, M. E. **Análise e Previsão de Séries Temporais: Os Modelos ARIMA**. Ijuí: Sedigraf, 1996.
- STOCK, J. H.; WATSON, M. (2004). Combination forecasts of output growth in a seven-country data set. **Journal of Forecasting**, v. 23, p. 105-430, 2004.
- TEIXEIRA JR, L. A. **Combinação Geométrica de Métodos Preditivos**. Aplicação à Série de Consumo Residencial de Energia Elétrica. 147 f. Dissertação (Mestrado em Engenharia Elétrica) – Pontifícia Universidade Católica do Rio de Janeiro, 2009.
- WANG, X.; MCDONALD, J. R. **Modern Power System Planning**. McGraw-Hill, 1994.
- WINKLER, R. L.; MAKRIDAKIS, S. The combination of Forecasting. **Journal of the Royal Statistical Society**, v. 146, p. 150-157, 1983.

