



## XIX Seminário Nacional de Distribuição de Energia Elétrica

SENDI 2010 – 22 a 26 de novembro

São Paulo - SP - Brasil

### Previsão do Consumo Regional da Elektro em Horizonte Anual com Modelos de Rede Neurais

|                                |                               |                            |
|--------------------------------|-------------------------------|----------------------------|
| <b>Ivette R. Luna</b>          | <b>Renata Massaro</b>         | <b>Marcos Leone</b>        |
| <b>Unicamp</b>                 | <b>Elektro</b>                | <b>Unicamp</b>             |
| ilunah@gmail.com               | Renata.massaro@elektro.com.br | marcosleonefilho@gmail.com |
| <b>Paulo Sergio F. Barbosa</b> | <b>Hermas Teixeira Franco</b> |                            |
| <b>Unicamp</b>                 | <b>Unicamp</b>                |                            |
| franco@fec.unicamp.br          | hermas.unicamp@gmail.com      |                            |

#### Palavras-chave

Inferência Fuzzy

Previsão consumo energia

Redes Neurais Artificiais

Regras Nebulosas

#### Resumo

Este trabalho apresenta o desenvolvimento de uma metodologia de modelagem com redes neurais artificiais para fins de previsão do consumo mensal de energia elétrica com base em sistema adaptativo FIS – *Fuzzy Inference System*. Tal sistema utiliza regras nebulosas e está baseado, principalmente, em modelos lineares de primeira ordem, compondo um sistema nebuloso do tipo Takagi-Sugeno (TS). O modelo proposto foi testado em diversas classes de consumo e regiões da área de concessão da ELEKTRO, sendo apresentados os resultados neste trabalho para as regiões de Limeira, Rio Claro e Guarujá, todas para a classe comercial. Estas três regionais são as de maior consumo comercial da ELEKTRO. Foi adotado o período de 2003 a 2006 para a fase de calibração e o período de 2007 e 2008 para a fase de testes. Os resultados da previsão foram de boa qualidade, seja com base na representação da estrutura sazonal de consumo, como com base na análise dos resíduos dos erros, que revelou não assimetria, e a grande maioria dos pontos situada nos limites de confiança de 95%. Tais indicadores sugerem a adequação do modelo para as regiões estudadas.

#### 1. Introdução

Como uma alternativa para o desenvolvimento dos modelos de previsão de consumo de energia elétrica mensal, surge um modelo baseado em regras nebulosas (*fuzzy*). Este modelo, proposto em Luna (2007), é construído utilizando o princípio de dividir e conquistar, pois o problema original é dividido em sub-problemas, de tal forma que modelos locais são alocados para resolver esses sub-problemas, diminuindo assim, a complexidade do problema global a ser resolvido, e facilitando a obtenção de soluções locais de melhor qualidade. Trabalhos de pesquisa prévios como os

apresentados em Murray-Smith & Johansen (1997) e em Solomatine & Siek (2006) são exemplos de aplicações em controle, identificação de sistemas e previsão de fenômenos naturais, respectivamente, sendo todos baseados em sistemas modulares, onde os especialistas são redes neurais ou apenas, modelos lineares de primeira ordem.

Assim, o modelo proposto é composto por um conjunto de especialistas (modelos locais) representados pelas regras nebulosas codificadas na estrutura do modelo. O antecedente de cada regra nebulosa define a área de ação de cada especialista. Já o especialista em si, é representado pelo conseqüente da regra associada.

Os modelos locais utilizados podem ser lineares ou não lineares, dependendo da complexidade do problema e das exigências da modelagem, tanto em termos de erro, como em facilidade de adaptação e/ou complexidade da própria estrutura do sistema.

## 2. Desenvolvimento

Em problemas de alta complexidade é necessário adotar estratégias simples, mas eficientes, que facilitem a modelagem e que sejam adaptáveis a mudanças na dinâmica do problema ao longo do tempo. Nesse contexto, abordagens baseadas em modelos locais, ou no princípio de dividir e conquistar, mas com o ajuste do modelo realizado de forma simples e adequada são de grande interesse. O modelo proposto para a geração dos previsores mensais se encaixa neste tipo de modelos, pois a sua estrutura geral surge da combinação não linear de modelos lineares. Ou seja, um modelo complexo é obtido a partir de um conjunto de modelos mais simples.

Embora exista uma terminologia diferente para a descrição do modelo e o ajuste dos seus parâmetros, a Tabela 1 mostra uma equivalência ou associação entre os termos formalmente utilizados para redes neurais e sistemas nebulosos, e os termos já conhecidos da teoria de séries temporais.

Tabela 1 - Associação entre os termos utilizados para modelos nebulosos e modelos de séries temporais.

| <b>Modelos nebulosos</b>    | <b>Modelos de séries temporais</b> |
|-----------------------------|------------------------------------|
| Entrada da rede             | Variáveis Independentes            |
| Saída da rede               | Variáveis Dependentes              |
| Erros                       | Resíduos                           |
| Treinamento ou aprendizagem | Estimação                          |
| Função erro                 | Critério de Estimação              |
| Padrões                     | Observações                        |
| Pesos, parâmetros           | Estimativas dos parâmetros         |

A estrutura do modelo baseado em regras nebulosas (FIS – *Fuzzy Inference System*) é descrita a seguir.

### 2.1 Estrutura do modelo

O sistema adaptativo proposto (FIS – *Fuzzy Inference System*) é baseado em regras nebulosas e principalmente, em modelos lineares de primeira ordem, compondo um sistema nebuloso do tipo Takagi-Sugeno (TS), descrito em Takagi & Sugeno (1985). Portanto, o sistema representa um mecanismo de inferência nebulosa, codificando na sua estrutura uma base de regras do tipo Se-Então. A parte “Se” é representada pelos antecedentes das regras nebulosas, enquanto a parte “Então” é codificada nos conseqüentes das mesmas regras.

A parte do antecedente representa uma sub-região do espaço de entrada. Esta sub-região define a região ativa associada à regra nebulosa. A parte do conseqüente é definida por um modelo local que contribuirá com uma parcela na estimativa da saída final.

A estrutura geral de um modelo composto por M regras fuzzy é ilustrada na Figura 1, onde  $x^k$  é o vetor de entrada no instante k,  $\hat{y}^k$  é a saída do modelo para uma determinada entrada  $x^k$ . Desta forma, cada regra nebulosa pode ser re-escrita como:

$R_i^k$  :Se  $x^k$  pertence à região  $i$  com um grau de pertinência  $g_i^k$  então  $\langle \hat{y}^k = y_i^k \rangle$  para  $i=1, \dots, M$  e  $y_i^k$  a saída estimada do  $i$ -ésimo modelo local.

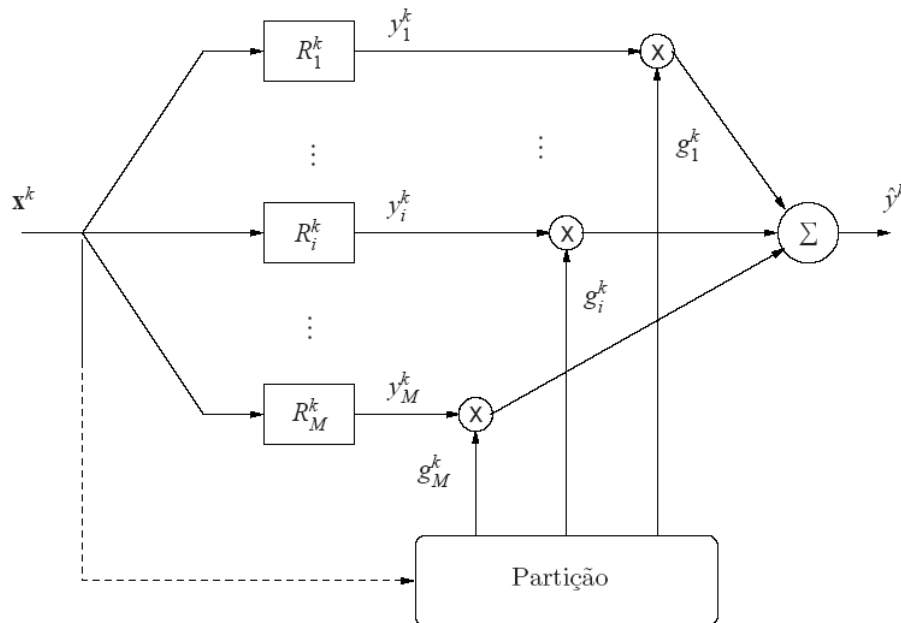


Figura 1 – Modelo dinâmico baseado em regras fuzzy.

O grau de pertinência de cada padrão de entrada  $x^k$  é representado por  $g_i(x^k) \in [0,1]$ , o qual é calculado da seguinte forma:

$$g_i(x^k) = g_i^k = \frac{\alpha_i P[i | x^k]}{\sum_{q=1}^M \alpha_q P[q | x^k]} \quad (01)$$

sendo  $\alpha_i$  coeficientes positivos que satisfazem a condição:

$$\sum_{i=1}^M \alpha_i = 1 \quad (02)$$

Estes coeficientes também podem ser considerados como índices indiretos de relevância para cada regra gerada durante o treinamento, de tal forma que, quanto mais alto seja o valor de  $\alpha_i$ , para  $i=1, \dots, M$ , maior será a importância da  $i$ -ésima regra nebulosa.

Em (01),  $P[i | x^k]$  é a probabilidade condicional de  $i$  dado  $x^k$ :

$$P[i | \mathbf{x}^k] = \frac{1}{(2\pi)^{p/2} \det(V_i)^{1/2}} \exp(-(\mathbf{x}^k - c_i)V_i^{-1}(\mathbf{x}^k - c_i)^t) \quad (03)$$

na qual  $V_i$  é uma matriz de ordem  $(p \times p)$  e representa a covariância associada à probabilidade  $P[i | \mathbf{x}^k]$ . Observa-se que  $P[i | \mathbf{x}^k]$  é uma função de densidade de probabilidade condicional, com centros em  $c_i$  e matriz de covariância  $V_i$ , a qual representa a probabilidade de escolher a regra  $i$ , dado que apenas a entrada  $\mathbf{x}^k$  é conhecida, assim como os parâmetros  $V_i$  e  $c_i$ .

A equação (03) também pode ser interpretada como a probabilidade do vetor de entrada  $\mathbf{x}^k$  pertencer à região ativa da  $i$ -ésima regra, já que, a medida que o dado de entrada encontra-se mais próximo do  $i$ -ésimo centro, maior será o valor do  $i$ -ésimo grau de pertinência. Em outras palavras, cada sub-região nebulosa é representada pelo respectivo centro  $c_i$ , assim como os parâmetros de dispersão codificados na diagonal principal da matriz de covariância associada  $V_i$ .

A saída do sistema proposto,  $\hat{y}^k$ , é calculada como uma combinação não linear de modelos locais com saídas  $y_i^k$  e os seus respectivos graus de pertinência  $g_i^k$ , ou seja:

$$\hat{y}^k = \sum_{i=1}^M g_i^k y_i^k \quad (04)$$

onde, neste caso,  $y_i^k$ , para  $i=1, \dots, M$ , é definido por:

$$y_i^k = \phi^k \theta_i^T \quad (05)$$

sendo  $\phi^k = [1 \ x_1 \ x_2 \ \dots \ x_p]$  o vetor composto por um termo constante mais as variáveis de entrada;

$\theta_i = [\theta_{i0} \ \theta_{i1} \ \theta_{i2} \ \dots \ \theta_{ip}]$  o vetor de coeficientes dos modelos locais lineares, de tamanho  $1 \times (p+1)$ .

A dinâmica do modelo pode ser interpretada como um mecanismo de inferência nebulosa, sendo a base de regras composta por regras nebulosas. Por outro lado, o modelo proposto pode também ser interpretado como uma arquitetura modular, sendo que o mecanismo de inferência representa o sistema de controle para cada módulo ou especialista (Guo e Xu, 1999), e o módulo em si é representado por cada um dos modelos lineares definidos.

Ou seja, existe uma analogia entre o modelo de mistura de especialistas e o sistema computacional proposto, já que a saída do sistema é calculada como o valor ponderado das saídas de cada regra nebulosa. Em ambas as abordagens, os coeficientes que ponderam as saídas locais dependem das entradas a cada instante e, dado que as funções de pertinência são definidas como funções gaussianas generalizadas, estes podem simultaneamente, representar funções de probabilidade e fazer uso do algoritmo EM (Expectation Maximization) para ajustar o modelo proposto.

Nesta seção, a estrutura descrita é apenas uma arquitetura geral, a qual será otimizada de acordo com a técnica de treinamento do tipo *offline*, de forma que não seja necessário definir previamente o número de regras  $M$  necessárias para compor a estrutura do modelo.

A vantagem desta estrutura é a sua relativa simplicidade quando comparada a outras abordagens não lineares, assim como a sua interpretabilidade e adaptação. Além disso, esta permite uma série de variações no treinamento do modelo, obtendo ganhos tais como, a seleção automática da estrutura, um processo de aprendizado mais rápido, capacidade de adaptação, entre outras.

## 2.2. Método de Ajuste dos Parâmetros do Modelo

O algoritmo de aprendizado utilizado para a otimização dos modelos FIS é conhecido como algoritmo da Maximização da Verossimilhança ou algoritmo EM (*Expectation Maximization*), e foi formulado como um problema de mistura de densidades de probabilidade (JACOBS, 1991) a partir da proposta inicial descrita em DEMPSTER(1997). Posteriormente, foi mostrado que o algoritmo EM poderia ser empregado como uma técnica de aprendizado supervisionado, sendo a sua aplicação estendida a outros campos, tais como o aprendizado de redes neurais multicamadas (MA et al1997), de redes neurais do

tipo RBF (LAZARO et al, 2003), assim como diversas aplicações em séries temporais (CARVALHO, 2005).

O algoritmo EM, como o próprio nome indica, consiste em maximizar o valor esperado da função objetivo, definida como a função de verossimilhança do modelo. A função de verossimilhança é baseada na interpretação do modelo um modelo de mistura para estimação da distribuição de probabilidade condicional dos dados, com funções de densidade condicional ponderando a mistura dos componentes (Mc CIACHLAN, 1988).

O algoritmo EM oferece algumas vantagens quando comparado ao algoritmo tradicional de retropropagação do gradiente descendente, dentre as quais, destacam-se a rapidez do aprendizado, já que a convergência é atingida com menor número de iterações, assim como a obtenção de ótimos locais em geral melhores. Além disso, mediante a utilização do algoritmo EM para o ajuste de sistemas modulares, a interpretação do modelo como um modelo estatístico é mais perceptível. Ou seja, existe uma dualidade inerente ao modelo (WARNER, 1996), tanto para as redes neurais como para os modelos estatísticos. Maiores detalhes referentes ao algoritmo de otimização da verossimilhança podem ser encontrados em (JACOBS, 1991).

Como foi mencionado anteriormente, a inicialização do modelo é sujeita ao número de especialistas a utilizar para construir o modelo de previsão. Embora o modelo de mistura de especialistas tenha se mostrado eficiente na previsão de séries temporais, um problema sempre existente neste tipo de abordagens é a definição da quantidade de especialistas necessários para cada problema em questão (como definir  $M$ ). Neste trabalho, foi feita a utilização de um algoritmo de agrupamento não-supervisionado, conhecido na literatura como algoritmo *Subtractive Clustering*, proposto em (CHIU, 1994). O objetivo ao utilizar este algoritmo é definir de forma automática o número de especialistas ( $M$ ) a utilizar na construção dos modelos de séries temporais, já que pela natureza do próprio algoritmo (não supervisionado), é este quem define esse valor.

Assim, o procedimento para a obtenção dos modelos de séries temporais necessários neste projeto e baseados regras nebulosas é resumido como segue:

- a) Analisar os dados disponíveis para o ajuste dos modelos e determinar as entradas a utilizar para a modelagem das séries;
- b) Pré-processar os dados e construir os padrões entradas-saída para o ajuste dos modelos;
- c) Inicializar os modelos de mistura de especialistas com a ajuda do algoritmo *Subtractive Clustering*;
- d) Ajustar os modelos utilizando o algoritmo EM;
- e) Verificar os erros de previsão até  $H$  passos à frente, de acordo com as necessidades do projeto.
- f) Se necessário, repetir o processo de otimização.

### 2.3. Aplicações a Regiões Seleccionadas da ELEKTRO

São apresentados a seguir os resultados da modelagem pelo sistema adaptativo proposto (FIS) para a classe comercial de algumas regiões da área de concessão da ELEKTRO. As séries históricas originais de consumo mensal foram submetidas ao procedimento da padronização, obtendo-se séries estacionárias. Foram testadas diversas variáveis explicativas, sendo selecionadas apenas: a) o CAGED (Cadastro Geral de Empregados e Desempregados) do mês imediatamente anterior ao mês que se deseja prever; b) o consumo padronizado do mês imediatamente anterior; c) o consumo padronizado de dois meses anteriores.

O período de calibração foi de 2003 a 2006, e o período de testes foi de 2007 a 2008.

Os resultados são apresentados adiante, apenas para as três regiões de maior consumo da ELEKTRO, por limitações de espaço.

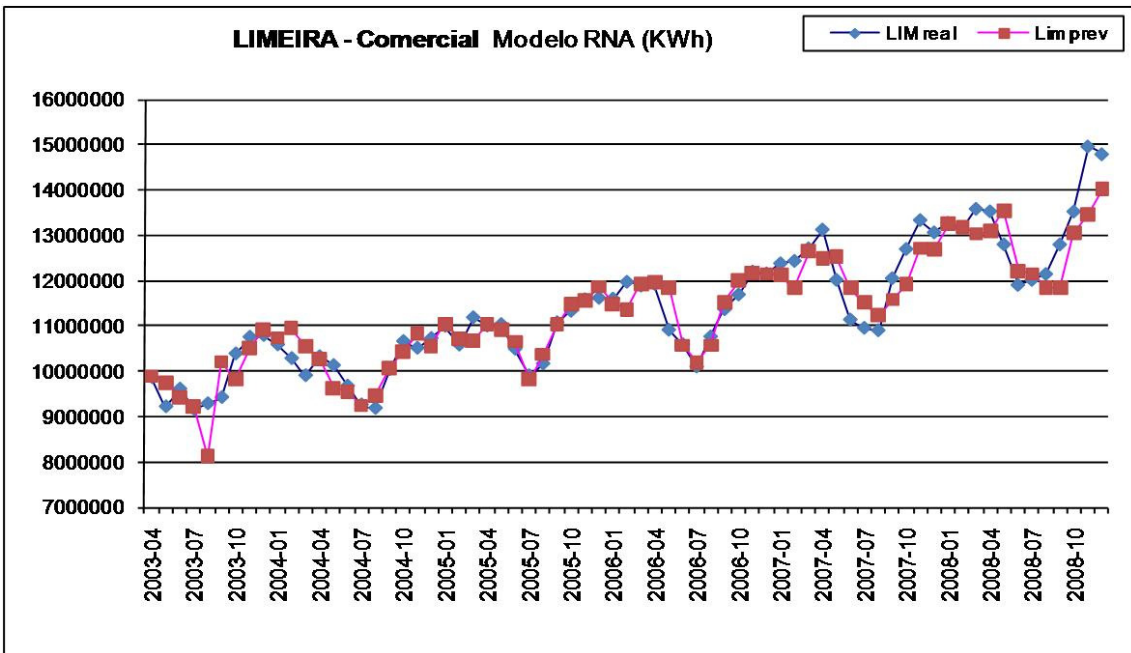


Figura 2 – Calibração e Ajuste da Região de Limeira – Classe Comercial

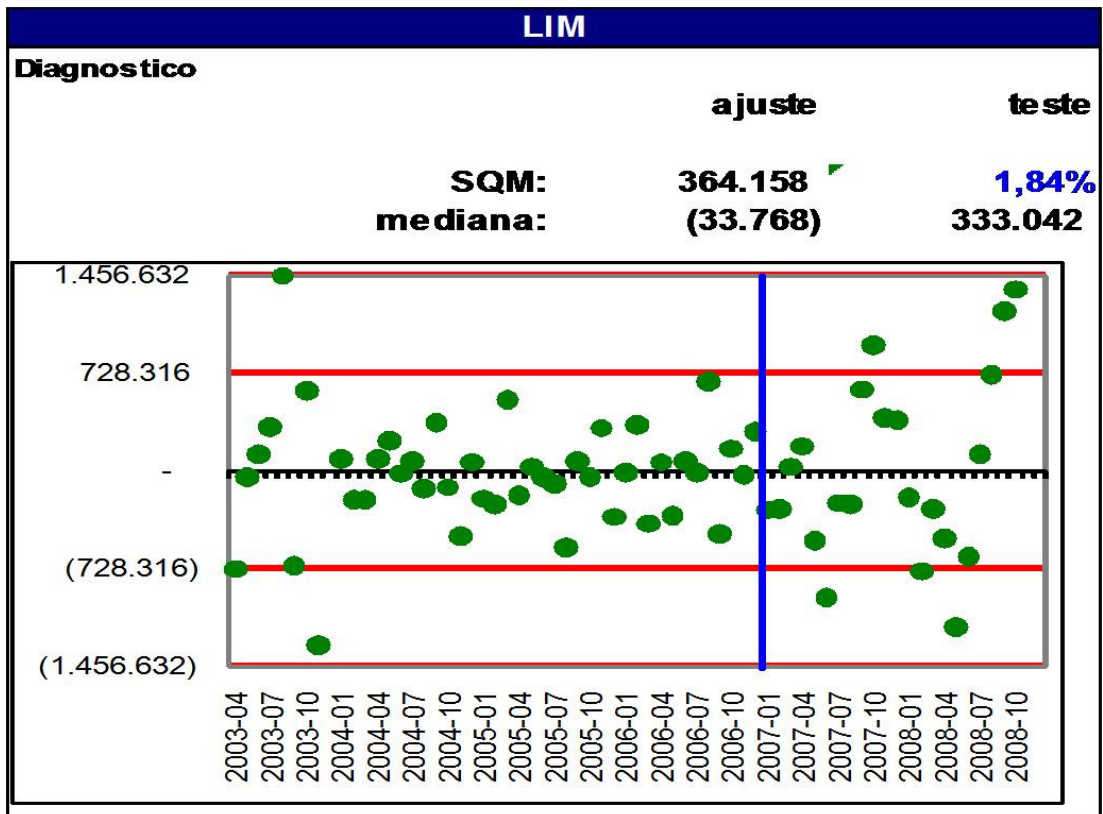


Figura 3 – Resíduos para as Fases de Calibração e Ajuste da Região de Limeira – Classe Comercial

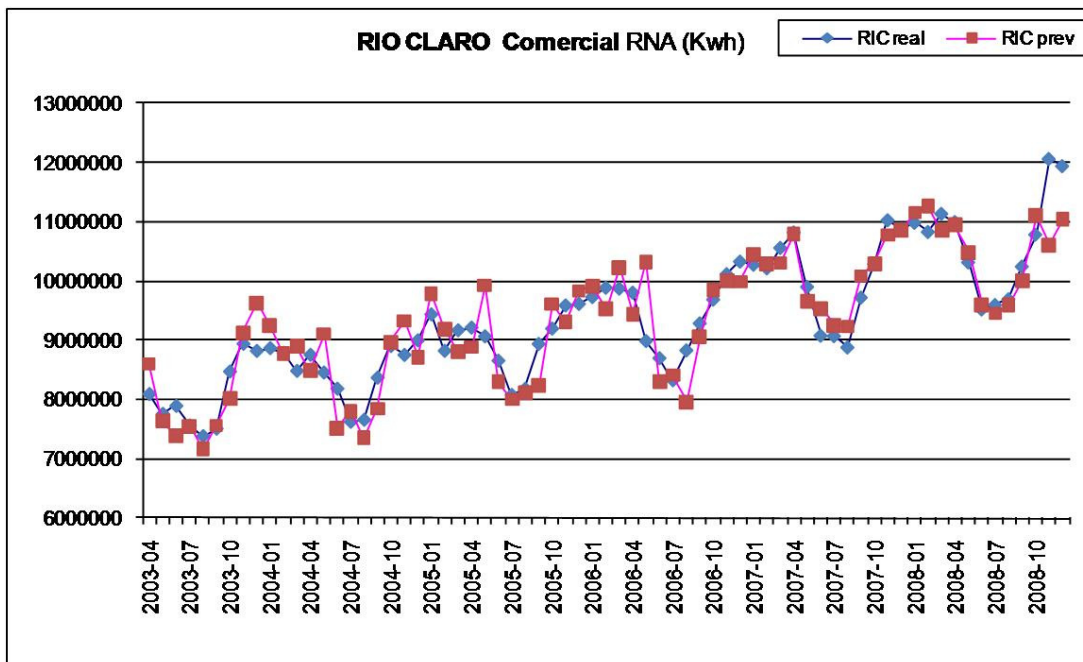


Figura 4 – Resíduos para as Fases de Calibração e Ajuste da Região de Rio Claro – Classe Comercial

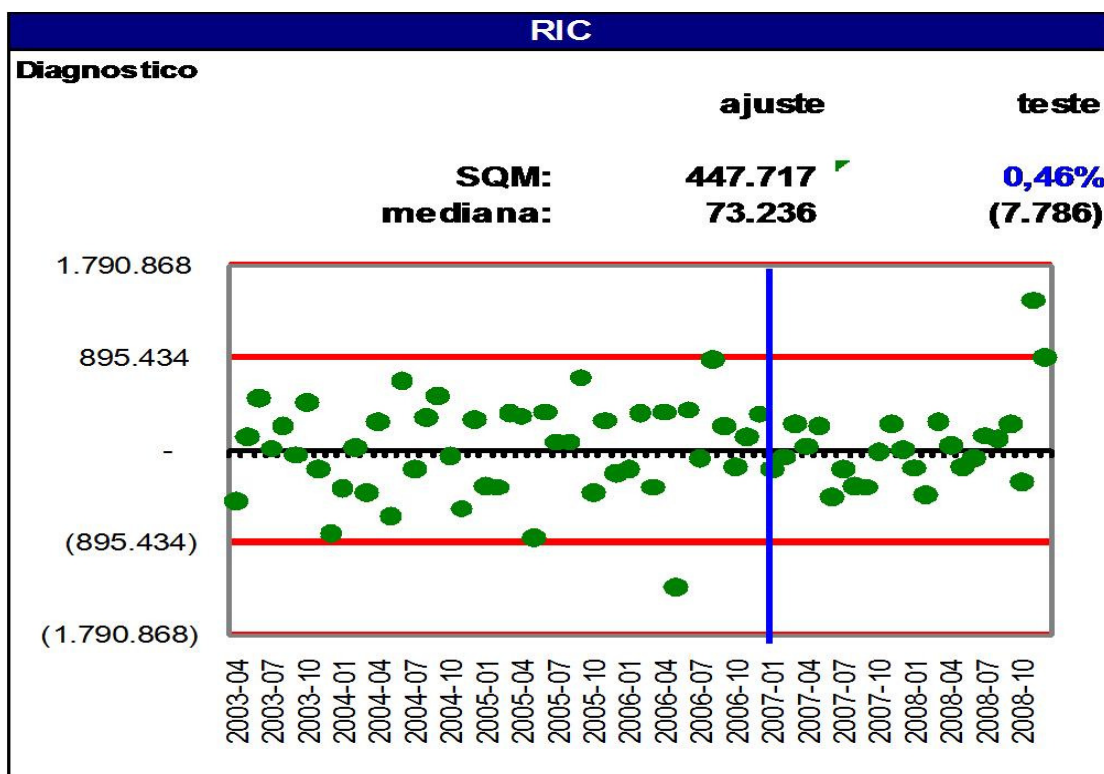


Figura 5 – Resíduos para as Fases de Calibração e Ajuste da Região de Rio Claro – Classe Comercial

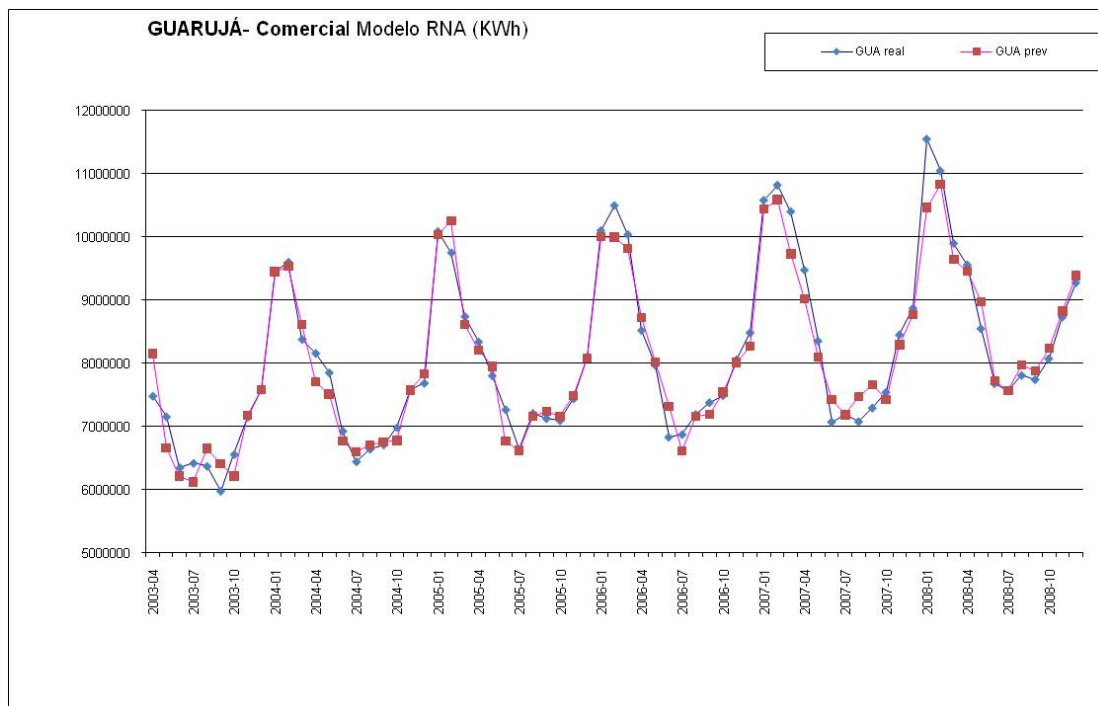


Figura 6 – Resíduos para as Fases de Calibração e Ajuste da Região de Guarujá – Classe Comercial

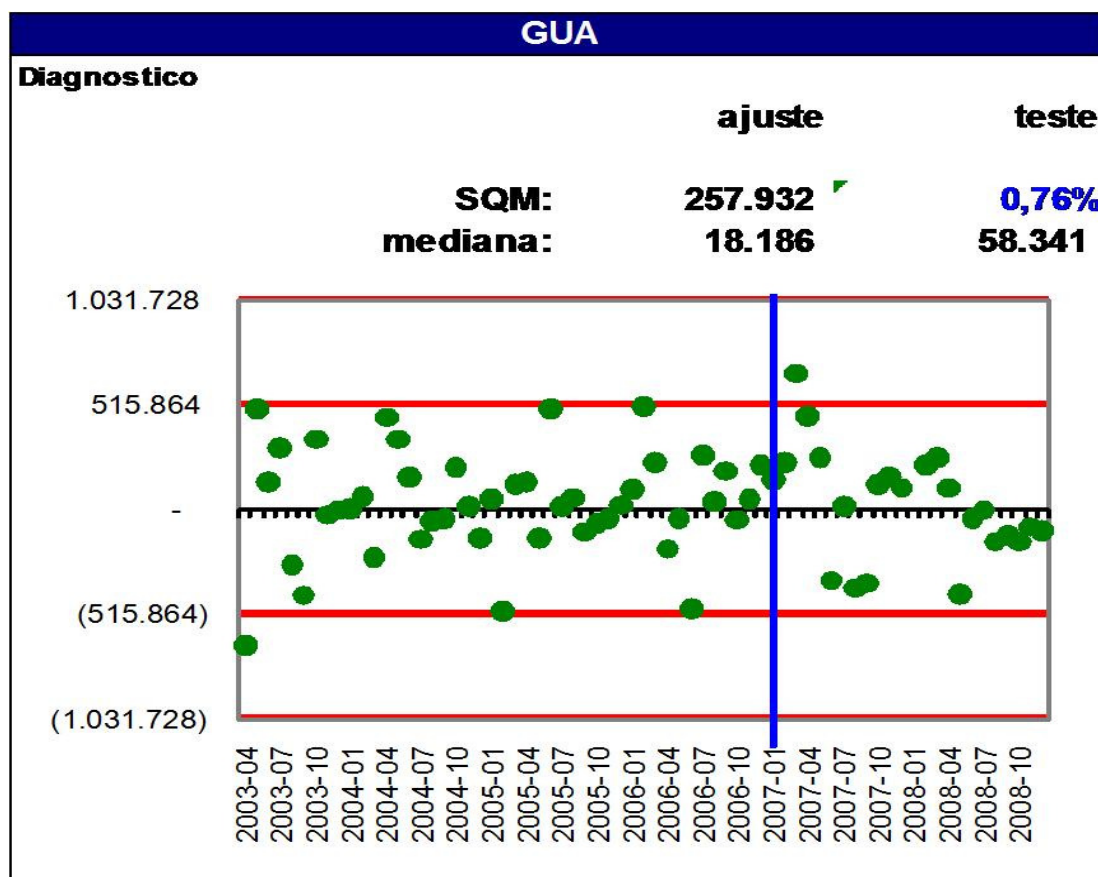


Figura 7 – Resíduos para as Fases de Calibração e Ajuste da Região de Guarujá – Classe Comercial

Para cada modelo ajustado foi feita a análise de resíduos, definindo-se, além da representação gráfica, o Erro Quadrático Médio (EQM) e a mediana dos resíduos, seja para a fase de ajuste(calibração do modelo) como para a fase de testes.



Os resíduos do modelo de sistema adaptativo FIS com base em regras nebulosas para Limeira comercial (Fig. 3) indicam não assimetria, e a grande maioria dos pontos situada nos limites de confiança de 95%, seja para a fase de calibração (à esquerda da barra vertical de cor azul) como na fase de testes (à direita da barra vertical de cor azul). Tais indicadores revelam boa qualidade preditiva do modelo.

O modelo FIS para a classe comercial de Rio Claro revela nível de ajuste razoável, com resíduos majoritariamente no interior dos limites do intervalo de confiança de 95%, conforme Fig. 5.

A adequação do modelo FIS – classe comercial para a região de Guarujá, mostrada na Fig. 6 e análise de resíduos, mostrados na Fig. 7, também indicam bom ajuste deste modelo, com boa representação dos vales e picos de consumo (Fig.6).

### 3. Conclusões

Foi proposto e testado um modelo FIS baseado em regras nebulosas (*fuzzy*) para previsão de consumo de energia elétrica em base mensal, com horizonte de um ano adiante. O sistema adaptativo proposto (FIS – *Fuzzy Inference System*) é baseado em regras nebulosas e principalmente, em modelos lineares de primeira ordem, compondo um sistema nebuloso do tipo Takagi-Sugeno (TS). Portanto, o sistema representa um mecanismo de inferência nebulosa, codificando na sua estrutura uma base de regras do tipo Se-Então.

O modelo proposto foi testado em diversas classes de consumo e regiões da área de concessão da ELEKTRO, sendo apresentados os resultados neste trabalho para as regiões de Limeira, Rio Claro e Guarujá, todas para a classe comercial. Estas três regionais são as de maior consumo comercial da ELEKTRO.

Os resultados da previsão para o ano de 2008 foram de boa qualidade, seja com base na representação da estrutura sazonal de consumo, como com base na análise dos resíduos dos erros, que revelou não assimetria, e a grande maioria dos pontos situada nos limites de confiança de 95%, seja para a fase de calibração. Tais indicadores sugerem a adequação do modelo para as regiões estudadas.

### 4. Referências bibliográficas

CARVALHO, A. X. e TANNER, M. A. (2005). Mixture-of-Experts of Autoregressive Time Series: Asymptotic Normality and Model Specification, *IEEE Transactions on Neural Networks* 16(1): 39–56.

CHIU, S. (1994). A cluster estimation method with extension to fuzzy model identification, *Proceedings of the Third IEEE Conference on Fuzzy Systems*, Vol. 2, Orlando - Florida, USA, pp. 1240–1245.

DEMPSTER, A., LAIRD, N. e RUBIN, D. (1977). Maximum likelihood for incomplete data via the EM algorithm, *Journal of the Royal Statistical Society B* 39: 1–38.

P. GUO; L. XU: “Relationship between Mixture of Experts and Ensemble Neural Networks”, *Proceedings of the 6<sup>th</sup> International Conference on Neural Information Processing*, vol. 1, 1999, pp. 246-250.

JACOBS, R., JORDAN, M., NOWLAN, S. e HINTON, G. (1991). Adaptive Mixture of Local Experts, *Neural Computation* 3(1): 79–87.

LAZARO, M.; SANTAMARIA, I.; PANTALEON, C.; A new EM-based training algorithm for RBF networks, *Neural Networks*, v.16, n.1, p.69-77, January 2003.

LUNA, I. (2007). Análise de Séries Temporais e Modelagem baseada em Regras Nebulosas, Tese de Doutorado, FEEC-UNICAMP, Brasil.

MA, S., JI, C. e FARMER, J. (1997). An efficient EM-based training algorithm for feedforward neural networks, *Neural Networks* 10(2): 243–256.

MCLACHLAN, G. e BASFORD, K. (1988). *Mixture models: inference and applications to clustering*, Marcel Dekker, Inc. SERIES: statistics: textbooks and monographs vol. 84.

MURRAY-SMITH, R.; JOHANSEN, T. A. (1997). *Multiple Model Approaches to Modelling and Control*, Taylor & Francis.

TAKAGI, T.; SUGENO, M. (1985). Fuzzy Identification of Systems and Its Applications to Modeling and Control, *IEEE Transactions on Systems, Man and Cybernetics* 15(1): 116–132.

SOLOMATINE, D. e SIEK, M. (2006). Modular learning models in forecasting natural phenomena, *Neural Networks* 10(2): 215–224.

WARNER, B. e MISRA, M. (1996). Understanding Neural Networks as Statistical Tools, *The American Statistician* 50(4): 284–293.