

## Previsão de Cargas Elétricas por Redes Neurais: Uma Investigação Empírica do Problema de *Overfitting*

Henrique S. Hippert<sup>1</sup>, Derek W. Bunn<sup>2</sup>, Reinaldo C. Souza<sup>3</sup>

<sup>1</sup>Depto. de Estatística, Univ. Federal de Juiz de Fora

<sup>2</sup>London Business School, U.K.

<sup>3</sup>Depto. Engenharia Elétrica, PUC-Rio

E-mails: hhippert@terra.com.br, reinaldo@ele.puc-rio.br

### Abstract

*In the last few years, models based on neural networks have been proposed with increasing frequency for load profile forecasting. Many forecasters are still sceptical about the performance of such models, arguing that most of them are based on very large neural networks that seem to be overfitted, and also, that most of them have not been properly validated, as their performance has not been compared to that of standard linear models. However, despite the scepticism, some of those NN-based models seem to have been tried out in practice, with satisfactory results. Could these models be really overfitted, and still be able to produce good forecasts? In this paper we investigate the effects of overfitting in the accuracy of NN-based forecasting systems, by comparing the performance of one such system to that of standard linear methods, over a series of actual load data.*

### 1. Introdução

Até a algum tempo atrás (início dos anos 1990), a problema de previsão de cargas elétricas parecia mais ou menos resolvido. Previsões eram necessárias para permitir a tomada de decisões técnicas (ligadas principalmente à programação de partida/parada dos geradores e ao despacho de carga), e os modelos de previsão usados na época já tinham atingido um nível de precisão “confortável” [1], sendo capazes de obter erros da ordem de 1 a 2%. Não havia muito interesse em diminuir este erro, uma vez que as companhias podiam lidar com ele simplesmente aumentando sua capacidade de geração, de modo a se precaver no caso de *underforecasting*.

A privatização e a desregularização dos sistemas elétricos que começaram a ocorrer em vários países a partir do início dos anos 1990 mudaram este panorama, tornando-o mais complexo. A energia passou a ser uma *commodity* a ser negociada no mercado, o que fez com que as companhias de eletricidade passassem a ter que tomar decisões comerciais ou financeiras, como as de compra e venda, e as de fixação do preço da energia. Isto mudou o foco dos métodos tradicionais de previsão de carga, porque os erros de previsão, além de terem consequências técnicas, passaram também a afetar os lucros da companhia, seu *market share* e o valor de suas ações [1].

Outro fator importante a afetar as pesquisas em previsão de carga foi o surgimento de técnicas novas, particularmente as de redes neurais artificiais (RNs). Os primeiros relatos da aplicação de RNs ao problema de previsão de carga surgiram no fim dos anos 80 e início dos anos 90 [2]. Desde então, o número de publicações tem crescido rapidamente; dos nove artigos sobre previsão de carga publicados em 1999 pelo *IEEE Transactions on Power Systems*, por exemplo, nada menos de seis propunham modelos baseados em RNs. A julgar pelo número de artigos, os sistemas de previsão baseados em RNs definitivamente não foram apenas “uma moda passageira”, como se temia que pudesse acontecer [3].

Estes artigos, contudo, não têm sido suficientes para convencer inteiramente os especialistas da área, e este ceticismo talvez seja justificado. Revisões recentes, e alguns livros-texto sobre previsão, afirmam que ainda há pouca evidência sistemática de que as RNs possam obter melhor resultado do que os métodos usuais, e que ainda há muito trabalho a ser feito antes que elas sejam aceitas entre as técnicas de previsão estabelecidas [4,5,6]. Os autores de [7] concluem que as RNs são “*promising*”, mas que “*a significant portion of the NN research in forecasting and prediction lacks validity. Over half of the studies suffered from validation and/or implementation problems which rendered their results suspect.*” Em [8] se observa que “*there are many inconsistent reports in the literature on the performance of ANNs for forecasting tasks*”, e por isso “*the findings are inconclusive as to whether and when ANNs are better than classical models*”.

Para investigar as razões deste ceticismo, fizemos uma revisão abrangendo cerca de 40 artigos que reportam a aplicação de RNs à previsão de carga a curto prazo [9]. As conclusões foram as de que, no todo, duas falhas importantes reduzem a credibilidade dos resultados mostrados nestes artigos. A primeira é que a maioria dos modelos propostos usavam RNs que pareciam complexas demais para o conjunto de dados que pretendiam modelar. Estas RNs tinham em geral um número muito grande de variáveis de entrada e de nós de saída, e alguns críticos afirmam que elas estariam *overfitted*. A segunda falha é que na maioria dos artigos o desempenho dos modelos propostos não foi comparado com o dos modelos lineares usados correntemente pelas

companhias de eletricidade, ou com o de algum outro padrão de referência aceitável. Os testes feitos nos modelos propostos eram em geral muito sumários e seus resultados tendiam a ser apresentados de maneira incompleta.

A revisão [2], também sobre aplicação de redes à previsão de carga, chegou a conclusões semelhantes: concluiu que os testes dos modelos propostos foram insuficientes (“*half of the reviewed articles have a test database which consists of less than half a year of data. ... In short-term load forecasting, it is not reasonable to test on less than one year ...*”), e que os artigos não revelaram detalhes que permitissem ao leitor julgar a qualidade da implementação das redes (“*Papers on STLF are generally poor of details about implementation issues*”).

Na prática, contudo, apesar do ceticismo das revisões da literatura, os sistemas de previsão baseados em RNs parecem ter produzido bons resultados. Há evidências de que estes sistemas têm sido usados rotineiramente por várias companhias de eletricidade, e obtido grande sucesso comercial [2,10,11].

Ao que parece, as redes propostas, apesar de muito complexas, são bastante robustas em relação ao *overfitting*. A fim de confirmar isto, fizemos simulações a partir de dados reais de carga e temperatura: construímos um modelo bastante complexo de RN, aparentemente superparametrizado em relação aos dados disponíveis, e comparamos seu desempenho ao de vários modelos lineares baseados nos que têm sido propostos na literatura. Os resultados, sumarizados abaixo, levam à conclusão de que estas RNs complexas, apesar de aparentemente superparametrizadas, têm realmente um desempenho melhor do que o de qualquer dos métodos tradicionais neste tipo de problema [12].

## 2. Os dados disponíveis

Os dados de carga usados foram coletados na cidade do Rio de Janeiro, e consistem numa série de medições horárias, nos anos de 1996 e 1997, da carga suprida pela concessionária local (Light). Os únicos dados meteorológicos disponíveis consistem de uma série de medições horárias da temperaturas de bulbo seco, na mesma cidade, no período de abril de 1996 a dezembro de 1997. Portanto, para estes dois anos, havia dados de carga para 104 semanas, mas dados de temperatura para apenas 91 (semanas 14 a 104), já que as primeiras 13 estavam faltantes.

Esta base de dados foi dividida em três partes. Dados de quarenta semanas (semanas 15 a 54) foram usados para estimar parâmetros e constantes dos métodos. Estes métodos foram testados num período de 30 semanas (semanas 55 a 84), e os melhores deles foram selecionado. Por fim, os métodos selecionados foram usados numa competição, prevendo para as semanas restantes da base de dados (semanas 85 a 104).

Tanto a série de carga quanto a de temperatura não tinham nenhum valor faltante, e aparentemente nenhum valor discrepante. Contudo, a série de cargas tinha que

ser tratada para que fosse retirado o efeito de eventos como feriados, jogos de futebol, greves, etc. Neste artigo, estaremos lidando apenas com dias “normais”, e os dados destes dias especiais não devem ser usados para estimação dos parâmetros dos modelos. Adotamos a opção mais simples para lidar com estes dados: substituímo-los pelos dados dos dias de mesma denominação nas semanas anteriores.

## 3. Os modelos utilizados

Os modelos usados foram agrupados em seis grupos, e se inspiram em modelos que tem sido propostos na literatura.

### Modelo I - Previsores naives (ingênuos)

Dois previsores naives serão usados como referência. O primeiro deles é definido como:  $L(d,h) = L(d-1,h)$ . O segundo é definido como:  $L(d,h) = L(d-7,h)$ , onde  $L$ : carga,  $d$ : dia,  $h$ : hora.

### Modelo II – 168 filtros univariados em paralelo

Este método usa um conjunto de 168 previsores univariados (filtros de Kalman) em paralelo para prever simultaneamente todas as 168 horas da semana. A previsão é uma-semana-à-frente: supondo que se esteja no final de uma semana, a previsão é feita para todas as 168 horas da semana seguinte. Um método similar foi proposto em [13]; a limitação mais óbvia deste método reside no fato de que cada carga é prevista com base em dados que têm pelo menos uma semana de idade

### Modelo III - Combinação de dois tipos de filtros univariados

Este método é um aperfeiçoamento do anterior, e decompõe a carga como a soma de uma média diária com uma série de 24 desvios horários em relação a esta média. Para prever a carga para um dia determinado, será preciso prever a carga média deste dia (usamos um filtro de amortecimento de Winters, uma vez que esta média é sazonal, com um período de sete dias), depois as desvios horários da média (usamos 168 filtros de Kalman em paralelo), e somar estas quantidades. Este é portanto um método de previsão *um-dia-a-frente*; ao fim de um dia, serão previstas as 24 cargas do dia seguinte. Um método similar foi proposto em [14].

### Modelos IV – Combinação de dois filtros univariados com regressão linear

Usa o tipo de modelo linear mais comum na literatura de previsão de carga (e.g. [15,16]), os modelos aditivos que decompõem a carga em três componentes:

$$L_t = B_t + W_t + e_t$$

onde:

**L**: carga; **B**: componente básico; **W**: componente devido ao efeito do clima; **e**: erro aleatório.

O “componente básico”  $B_t$  deve refletir o comportamento normal da série de cargas, supondo que não haja circunstâncias inesperadas no clima. Uma vez que este comportamento pode ser modelado pelo valor esperado da carga, dado o passado da série, usaremos as previsões do Modelo III para estimar  $B_t$ . As diferenças entre  $B_t$  e os valores observados  $L_t$  refletem os “desvios da carga” em relação aos seus valores esperados; destes “desvios de carga” ( $dL$ ), será extraída a série do “componente devido ao efeito do clima”  $W_t$ , por meio de regressão linear (modelo IV) ou de redes neurais (modelo V).

O componente  $W_t$  deve modelar a influência na carga de circunstâncias meteorológicas não usuais. Particularmente, se a temperatura a uma certa hora for mais alta do que a esperada para aquela época do ano, a carga naquela hora provavelmente também será alta do que o normal. Construímos um modelo para a temperatura “básica”, similar ao usado para a carga básica: as temperaturas médias diárias foram previstas por meio de amortecimento de Brown, e os desvios horários desta média foram previstos por 24 filtros de Kalman. As diferenças entre as temperaturas observadas e os valores básicos previstos são os “desvios de temperatura” ( $dT$ ).

O modelo de regressão para o componente  $W_t$  é dado por:

$$W_t = \beta_1 + \beta_2 dL_{t-24} + \beta_3 L_{t-24} + \beta_4 \hat{B}_t + \beta_5 d\hat{T}_t + \beta_6 t + e_t$$

onde as variáveis explicativas são:

- os valores defasados do desvio da carga  $dL_{t-24}$
- os valores defasados da carga observada  $L_{t-24}$
- os valores previstos do desvio da temperatura  $d\hat{T}_t$
- os valores previstos do componente básico  $B_t$
- a hora  $t$

Os coeficientes do modelo de regressão são reestimados a cada semana, sobre dados das três semanas anteriores. Também experimentamos usar conjuntos de 24 modelos de regressão em paralelo, um para cada hora do dia; os resultados contudo foram equivalentes aos obtidos pelo modelo mais simples acima.

A previsão final das cargas horárias será dada pela soma do componente básico  $B_t$  com o componente devido ao clima  $W_t$ .

### Modelo V: Combinação de dois filtros univariados com RNs

Este modelo usa uma RN no lugar da regressão linear do modelo acima, produzindo previsões do componente  $W_t$  a cada hora do dia seguinte. As variáveis de entrada da RN são as mesmas usadas na regressão. A RN é re-estimada a cada semana, com base nos dados das seis semanas antes da semana a ser prevista; a última destas seis semanas foi reservada para testes de parada por validação (*cross-validation*). Para escolher o número de neurônios na camada oculta, fizemos simulações sobre os dados das semanas 55 a 84; os melhores

resultados (MAPEs) foram obtidos por RNs com oito neurônios. Fizemos experimentos com diferentes defasamentos da variável  $dT$  (desvios de temperatura); valores defasados de  $dT$  produziram resultados melhores do que os não defasados, mas as diferenças entre diferentes defasamentos foi pequena.

### Modelo VI: Somente RNs

Usamos um modelo complexo de RN, com 24 neurônios de saída e 50 nós de entrada. Estes nós recebem as 24 cargas do dia antes da previsão, as 24 previsões de temperatura, e duas variáveis *dummy* que especificam o tipo de dia (dia de semana ou final de semana) do dia da previsão e o do dia antes deste. O treinamento foi feito sobre dados de uma amostra de 36 semanas. A técnica de *cross-validation* foi usada para determinar o ponto de parada das iterações, e uma amostra contendo os dados de quatro semanas foi usada para este fim. A RN não era adaptativa; foi treinada somente uma vez em cada rodada, e seus pesos não foram modificados durante o período de teste.

### 3. Resultados: comparação dos desempenhos *out-of-sample* dos métodos

A Tabela 1 compara os desempenhos dos modelos citados, com base nas previsões feitas para as 20 últimas semanas da base de dados (semanas 85 a 104). Uma vez que os modelos (v) e (vi) são baseados em procedimentos iterativos de estimação que dão resultados levemente diferentes a cada vez que são rodados, simulamos cada um deles 15 vezes; a tabela mostra o melhor e o pior resultados obtidos nestas 15 simulações.

Tabela 1 – Erros de previsão dos diferentes modelos, numa amostra de 20 semanas

		mape	mse $\times 10^4$
<b>I</b>	Naïve I	7.37	12.1
	Naïve II	6.12	6.11
<b>II</b>	168 Filtros de Kalman	3.93	2.96
<b>III</b>	168 F. Kalman + Winters	3.93	2.90
<b>IV</b>	Modelo III + regressão	3.16	1.82
<b>V</b>	Modelo III + RN	pior	3.33 2.11
		melhor	3.18 1.90
<b>VI</b>	RN única	pior	2.65 1.30
		melhor	2.35 1.16

A Figura 1 mostra diagramas de Tukey (*box-plots*) dos erros absolutos percentuais (APEs) de alguns destes modelos, durante o mesmo período. Nota-se que o modelo VI obteve erros que não apenas foram melhores do que os dos outros modelos, em termos médios, mas também tiveram menor dispersão e menos valores discrepantes (*outliers*).

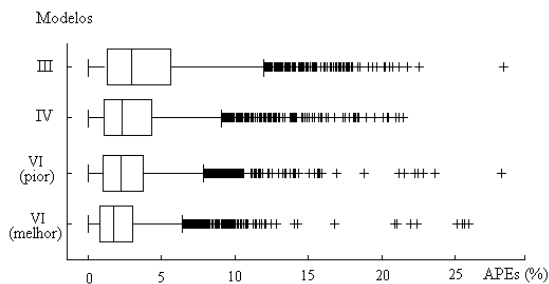


Figura 1. Distribuição dos erros em 20 semanas

#### 4. Conclusões

Os resultados acima mostram que RNs grandes podem funcionar pelo menos tão bem quanto os métodos lineares usuais na previsão de perfis de carga. Esta conclusão é confirmada pelas dos vários artigos que vem propondo métodos de previsão baseados nestas RNs, e pela uso já bastante difundido destes métodos na prática, pelas companhias de eletricidade.

Estas RNs, contudo, são tão complexas que poderíamos suspeitar que elas estejam *overfitted*, e poderíamos esperar delas maus resultados *out-of-sample*, de acordo com o raciocínio estatístico usual. O modelo de RN que fez as melhores previsões nas simulações acima (50 nós de entrada, 15 neurônios ocultos, 24 saídas) tinha 1149 pesos e *biases*, e estes foram estimados a partir de uma amostra com apenas 280 pontos amostrais (dados de 280 dias); havia quatro vezes mais parâmetros do que pontos amostrais.

É difícil afirmar, no entanto, se esta rede estava ou não super-parametrizada. Alguns autores acreditam que o número de “parâmetros efetivos” de uma RN não é o mesmo que o número de seus pesos e *biases* [17], já que alguns destes podem ser irrelevantes. Ainda não é claro o que acontece com estes parâmetros “irrelevantes”, e de que forma eles influenciam o desempenho da rede.

Os resultados mostrados nestas simulações, e em outros artigos sobre previsão de carga, sugerem que as RNs são robustas em relação ao *overfitting* neste tipo de problema, e que mesmo redes que têm grande número de pesos (em relação ao número ótimo para um dado problema) ainda podem produzir boas previsões. Vários estudos têm sido publicados sobre os efeitos da super-parametrização e do *overfitting* das RNs, mas a maioria deles [18,19,20] tratam apenas de RNs pequenas, com um único neurônio de saída. Neste caso, é mais fácil examinar o comportamento dos sistemas em estudo, e encontrar analogias com outros modelos. O problema evidentemente se torna mais complexo quando se lida com RNs de alta dimensionalidade.

Porque teriam estes modelos de RNs obtido resultados melhores do que os modelos lineares neste problema de previsão de carga? Existe alguma característica especial na série de cargas que a torna adequada à previsão por RNs? As séries de carga têm realmente a particularidade de serem facilmente previsíveis, já que reverterem à média, tem baixo ruído, não sofrem mudanças

estruturais, e mostram alta variabilidade diária (do pico de carga ao vale), em relação à variabilidade entre diferentes dias, o que torna os perfis muito repetitivos (compare-se por exemplo este tipo de problema ao problema muito mais difícil de previsão de séries financeiras). Ou estará a explicação do bom desempenho das RNs na natureza mesma do problema, “previsão de perfis”, ao invés de “previsão de valores pontuais”? Quando as NNs fazem as estimativas para cada ponto do perfil, elas levam em conta informação sobre todos os valores dos perfis dos dias anteriores, enquanto que os métodos lineares não são capazes de considerar mais do que algumas horas adjacentes de cada vez (uma vez que as cargas nas horas adjacentes são praticamente colineares). As RNs, por isso, trabalham com muito mais informação do que os modelos lineares, o que pode explicar em parte seu melhor desempenho.

Se são estas características o que explica o sucesso dos modelos com grandes RNs na previsão de carga, seria de se esperar que este sucesso se repetisse na previsão de outras séries que tenham características similares. De fato, RNs têm sido aplicadas com sucesso também na previsão de outras séries altamente previsíveis, com a de demanda diária de gás [21,22], e em outros problemas de previsão de perfis [23].

Será necessário investigar teoricamente estas características de modo a definir áreas de previsão nas quais as RNs podem ser superiores aos métodos convencionais. Estudos teóricos que expliquem o bom desempenho das RNs irão lançar nova luz sobre os relatos de aplicações práticas bem sucedidas como em [11], e ajudar a delimitar de que modo as características únicas das RNs devem ser exploradas na construção de modelos de previsão de carga.

#### Referências

- [1] Bunn DW. (2000) "Forecasting Loads and Prices in Competitive Power Markets". *Proc IEEE*, **88**: (2) 163-169.
- [2] Czernichow T, Piras A, Imhof K, Caire P, Jaccard Y, Dorizzi B, Germond A (1996) "Short Term Electrical Load Forecasting with Artificial Neural Networks". *Engineering Intelligent Syst.* **2**: 85-99.
- [3] Chatfield C. (1993) "Neural Networks: Forecasting Breakthrough or Passing Fad?" *Int J Forecast.* **9**:1-3.
- [4] Makridakis S, Wheelwright SC, Hyndman RJ. (1998) *Forecasting – Methods and Applications*. 3<sup>rd</sup> ed. John Wiley & Sons.
- [5] Gorr WL. (1994) "Research Prospective on Neural Network Forecasting". *Int J Forecast.* **10**: 1-4.
- [6] Hill T, Marquez L, O'Connor M, Remus W. (1994) "Artificial Neural Network Models for Forecasting and Decision Making". *Int J Forecast.* **10**: 5-15.
- [7] Adya M, Collopy F. (1998) "How Effective are Neural Networks at Forecasting and Prediction? A Review and Evaluation". *J Forecast.* **17**: 481-495.
- [8] Zhang G, Patuwo BE, Hu MY. (1998) "Forecasting with Artificial Neural Networks: The State of the Art". *Int J Forecast.* **14**: 35-62.

- [9] Hippert HS, Pedreira CE, Souza RC. (2000) "Neural Networks for Short-Term Load Forecasting: A Review and Evaluation". To appear in the *IEEE T Power Syst.*
- [10] Khotanzad A, Afkhami-Rohani R, Maratukulam D. (1998) "ANNSTLF - Artificial Neural Network Short-Term Load Forecaster - Generation Three". *IEEE T Power Syst.* **13**: (4) 1413-1422.
- [11] Hobbs BF, Helman U, Jitprapaikulsarn S, Konda S, Maratukulam D. (1998) "Artificial NN for short-term energy forecasting: accuracy and economic value". *Neurocomputing*, 71-84
- [12] Hippert HS, Bunn DW, Souza RC. (2000) "Forecasting Electrical Loads with Linear Methods Combined with Neural Networks". Working paper, London Business School.
- [13] Sargunraj S, Sen Gupta DP, Devi S. (1997) "Short-term Load Forecasting for Demand Side Management". *IEE Proc.-Gener. Transm. Distrib.* **144**: (1) 68-74
- [14] Infield DG, Hill DC. (1998) "Optimal Smoothing for Trend Removal in Short Term Electricity Demand Forecasting". *IEEE T Power Syst.* **13**: (3) 1115-1120.
- [15] Hyde O, Hodnett PF. (1997) "An Adaptable Automated Procedure for Short-Term Electricity Load Forecasting." *IEEE T Power Syst.* **12**: (1) 84-93.
- [16] Fan JY, McDonald JD. (1994) "A Real-Time Implementation of Short-Term Load Forecasting for Distribution Power Syst". *IEEE T Power Syst.* **9**: (2) 988-994.
- [17] Weigend AS, Rumelhart DE. (1991). "The effective dimension of the space of hidden units". *Proc. IJCNN'91.* **3**: 2069-2074
- [18] Saxen H. (1996) "Nonlinear time series analysis by neural networks: A case study" *Int J Neural Syst.* **7**: (2) 195-201.
- [19] Tetko IV, Livingstone DJ, Luik AI. (1995) "Neural network studies- comparison of overfitting and overtraining". *J Chemical Information & Computer Sc.* **35**: (5) 826-833
- [20] Markham IS, Rakes TR. (1998) "The effect of sample size and variability of data on the comparative performance of artificial neural networks and regression". *Computers & Operations Research.* **25**: (4) 251-263.
- [21] Brown RH (1996) "Development of artificial neural networking models to predict daily gas consumption" *Am Gas Assoc Forecasting Rev.* **5**: 1-22.
- [22] Khotanzad A, Elragal H, Lu TL. (2000) "Combination of artificial neural-network forecasters for prediction of natural gas consumption" *IEEE T Neural Networks.* **11**: (2) 464-473.
- [23] Khotanzad A, Davis MH, Abaye A, Maratukulam DJ. (1996) "An artificial neural network hourly temperature forecaster with applications in load forecasting" *IEEE T Power Syst.* **11**: (2) 870-876.