

## Efeito de Factores Climáticos no Consumo de Energia Eléctrica

Armando B. Mendes<sup>1</sup> José M.S. Ferreira<sup>2</sup>

<sup>1</sup> Departamento de Matemática da Universidade dos Açores e CEEAplA, amendes@uac.pt

<sup>2</sup> Electricidade dos Açores, josefer@eda.pt

### Sumário

Neste trabalho identificam-se padrões e relações entre os dados de consumo de energia eléctrica e várias variáveis climatéricas. Para tal utilizam-se técnicas de *data mining*, nomeadamente a metodologia CRISP-DM e software de *data warehouse* MS SQL Server. Os resultados confirmam que as variáveis climatéricas têm influência muito significativa no consumo de energia eléctrica, tendo sido possível prever os consumos de 2007 com um erro absoluto médio de 1,4 MW. Identificam-se ainda vários padrões no consumo, nem todos espectáveis face ao conhecimento actual.

**Palavras-chave:** árvores de modelos, clima, *data mining*, energia eléctrica.

### 1. Contexto e definição do problema

Não é claro o efeito dos factores climáticos na variação do consumo de energia eléctrica em determinado contexto. Vários estudos se têm debruçado sobre o difícil problema de prever consumos de energia eléctrica. Como se sabe, não existe nenhuma forma eficiente de armazenar energia eléctrica, pelo que há uma necessidade constante de ajustar a produção ao consumo. Se for possível estimar o consumo com alguma antecedência a produção poderá ser facilmente ajustada, tornando todo o processo mais simples e eficiente. A grande maioria dos trabalhos publicados centra-se na obtenção de previsões o mais precisas possível e com o máximo de antecedência (ver por exemplo: ; Troutt *et al.*, 1991, Engle *et al.*, 1992 e Liu *et al.*, 1993). Todas estas publicações reconhecem a importância da relação entre os factores climáticos e o consumo de energia. Segundo Smith (1989) a previsão dos factores climáticos é mesmo considerada a principal fonte de incerteza na previsão do consumo de energia.

Este problema é mais complexo em ambientes insulares especialmente sujeitos a alterações climáticas. As empresas que produzem e/ou distribuem a energia eléctrica têm de tomar decisões em tempo real de quanto produzir nas centrais termoeléctricas, de modo a satisfazer a procura a cada segundo.

Neste contexto, qualquer conhecimento adicional sobre a forma como se comporta o consumo de energia eléctrica numa dada região é muito valorizado.

### 2. Exploração de Dados e Pré-Processamento

Utiliza-se neste projecto a metodologia CRISP-DM *Cross Industry Standart Process for Data Mining*, já anteriormente descrita pelos autores em Mendes *et al.* (2009).

A utilização desta metodologia em problemas abordados por técnicas de *data mining* tem-se revelado muito útil por, no essencial, permitirem estruturar e disciplinar o processo, evitando a aplicação indiscriminada de algoritmos como reacção natural à disponibilização dos mesmos em *softwares* de simples utilização. Os retornos entre as fases descritas no modelo processual estão previstos na metodologia e constituem a espiral de modelação e extracção de conhecimento (Lavrač *et al.*, 2004). Para uma descrição completa desta metodologia ver o documento original de Chapman *et al.* (2000) e o site criado pelo projecto: [www.crisp-dm.org](http://www.crisp-dm.org).

As primeiras dificuldades prenderam-se com a obtenção dos dados e o seu indispensável tratamento para que seja possível a sua utilização com os algoritmos pretendidos.

A base de dados sobre a qual se trabalhou, refere-se aos dados meteorológicos recolhidos no Aeroporto de Ponta Delgada, no período compreendido entre os anos de 1998 e 2007. Fez-se o download do site *wunderground.com*, recorrendo a um programa em Java, construído propositadamente para este fim. Quanto aos dados dos registos instantâneos das potências de cada um dos sistemas electroprodutores da Ilha de São Miguel, para os anos 2005 e 2006, existem em meio digital de fácil acesso, tornando fácil a importação dos mesmos. No entanto, devido a limitações no sistema de informação e a restrições de segurança, o download dos dados para 2007 só foi possível em períodos de dez dias.

Além destas dificuldades de acesso aos dados, verificaram-se igualmente alguns problemas durante a exploração dos dados que tiveram por consequência a necessidade de implementar delicadas tarefas de limpeza. Este tipo de tarefas foram efectuadas recorrendo ao *SQL Server 2005*, usando ferramentas OLAP para construir um cubo como ferramenta para resumo e exploração dos dados. O processo de construção do cubo e de implementação de fluxos de processamento (*process flows*) foi semelhante ao descrito em trabalho anterior (Mendes *et al.*, 2009).

Finalmente, após ter todos os dados em formato relacional, fundiram-se numa só tabela todos os atributos referentes ao clima e à potência, sendo assim possível executar um novo cubo com a integração das variáveis de clima. Uma dificuldade relevante nesta fase prende-se com o facto de as chaves nem sempre terem correspondência nos dois conjuntos de dados. Este tipo de problema de fusão de dados (*data fusion*) surge da compatibilização de dados provenientes de fontes diversificadas, ver por exemplo Chen (2001) e Saporta, (2002). Desta forma, foi necessário confrontar as duas fontes de dados e extrair somente as instâncias cujas horas coincidiam, tal foi efectuado igualmente implementando fluxos processuais adequados.

No final da fase de exploração e pré-processamento, dispunha-se de uma base de dados bem estruturada e com dados no nível de qualidade adequado.

### **3. Construção e Validação de Modelos**

Dado o objectivo de identificar relações entre o consumo de energia eléctrica e factores climatéricos, dos algoritmos disponíveis no *MS SQL Server*, os mais

adequados para análise explicativa de um *target* quantitativo são o *Microsoft Decision Trees*, *Microsoft Linear Regression* e *Microsoft Neural Network*. Uma breve descrição destes algoritmos pode ser encontrada em Larson (2006).

Na tentativa de encontrar relações entre o clima e o consumo de energia, considerou-se a potência como a variável dependente em todos os modelos. Em cronogramas construídos foi possível observar a existência de sazonalidades tanto semanais, apresentando valores menores durante o fim-de-semana, como durante o ano observando-se um aumento do consumo durante os meses de verão e no final do ano. Foi igualmente possível observar um ligeiro aumento anual das potências utilizadas.

Na tentativa de modelar este comportamento, foram elaborados vários modelos de regressão linear. Neste sentido, foram introduzidas variáveis binárias para estimar valores médios de picos de consumo. Adicionou-se ainda uma variável binária que indicasse a existência de fins-de-semana e feriados, uma vez que o comportamento nestes dias é diferente dos restantes dias da semana. Por fim, criou-se uma nova variável “Horas” retirada do campo “Data”, que identifica a hora de cada registo.

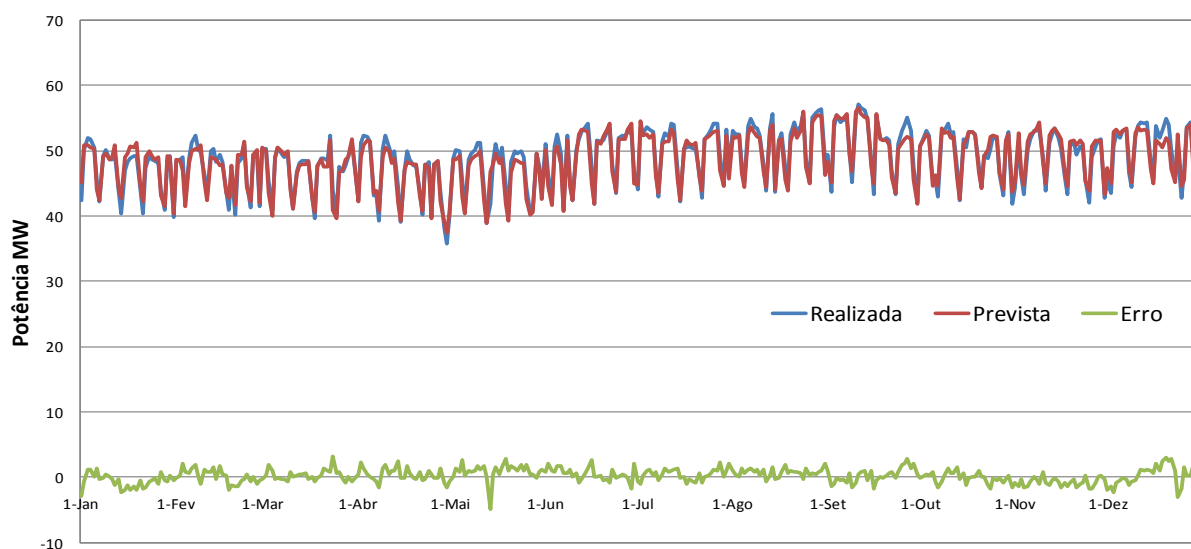
Todo o processo anterior foi repetido, para um espaço temporal mais alargado, utilizando-se registos dos anos 2005 e 2006 num total de 15.666. Esta tabela serviu como dados de treino. Usando todas as variáveis disponíveis construíram-se vários modelos e verificou-se que aquele que apresenta melhor ajuste aos dados é o que utiliza o algoritmo *MS Decision Trees*. Note-se que este algoritmo, quando são usadas variáveis dependentes numéricas, induz árvores de modelos. Ou seja, é construída uma árvore de decisão de pequena dimensão, de tal forma que em cada nó folha restam um elevado número de registos os quais são utilizados para ajustar um modelo de regressão linear. Obtêm-se assim, modelos de regressão para cada segmento de instâncias ou registos. A validação do melhor modelo foi efectuada com a base de dados de teste correspondente a valores de consumo horário para o ano de 2007. As estatísticas de ajuste apresentam-se muito boas, como pode ser observado visualmente no Figura 1. O desvio absoluto médio é de 1,4 MW o que significa que esse seria o erro médio da previsão se fosse possível obter valores de previsão das variáveis climatéricas sem erro.

#### **4. Resultados e Conclusões**

Das restantes ramificações da árvore de modelos, verifica-se que a temperatura, o ponto de orvalho e a humidade contribuem para a variação do consumo de energia, isto é, quando aumentam os valores destas variáveis há um aumento da potência consumida. Como se sabe de conhecimento de domínio ou simples senso comum, a existência de elevadas temperaturas obriga-nos a recorrer a sistemas de refrigeração. Por outro lado, quando a humidade relativa do ar é elevada, sabemos que nos sistemas de transporte de energia eléctrica existem perdas devido à corrente de fuga através dos isoladores, verificando-se ainda um aumento da utilização dos desumidificadores.

Note-se que se considera o modelo de árvores de decisão adequado, não apenas pelos resultados obtidos pelas estatísticas de qualidade de ajuste, mas porque as relações com algumas variáveis climatéricas não se mantêm ao longo de todo o ano. Por

exemplo, a temperatura pode explicar o comportamento no verão, mas não durante o Natal. Tentar ajustar a mesma regressão a todo o ano resultaria em ajustes de pouca qualidade. Assim, o ajuste por segmentos ou períodos sazonais faz, certamente, mais sentido. Aliás este tipo de comportamento das séries de consumo eléctrico é reconhecido em publicações anteriores sendo recomendada a modelação de períodos específicos como as horas de pico (ver, por exemplo: Engle *et al.*, 1992 e Liu *et al.*, 1993).



**Figura 1:** Valores médios diários de potência prevista e realizada para o ano de 2007.

Ao considerar os resultados apresentados pelos modelos é inevitável concluir que o consumo de energia é maior nos períodos de maior temperatura. No Verão a média das potências diárias é mais elevada, embora no Inverno também se verifique um consumo elevado por altura da época natalícia. Este estudo veio demonstrar que é possível explicar com elevada precisão a influência do clima no consumo de energia eléctrica.

## Referências:

- ARNOTT, DAVID & PERVAN, GRAHAM (2005). A critical analysis of decision support systems research. *Journal of Information Technology*, 20, 67-87.
- CHAPMAN, P.; CLINTON, J.; KERBER, RANDY; KHABAZA, THOMAS; REINARTZ, THOMAS; SHEARER, COLIN & WIRTH, RÜDIGER (2000). *CRISP-DM 1.0 - Step-by-step data mining guide*. USA, SPSS Inc..
- CHEN, ZHENGXIN (2001). *Intelligent Data Warehousing: From data preparation to data mining*. Boca Raton, USA, CRC Press.
- ENGLE, R.F.; MUSTAFA, C. & RICE, J. (1992). Modelling peak electricity demand. *Journal of Forecasting*, 11, 241-251.
- LARSON, B. (2006). *Delivering Business Intelligence with Microsoft SQL Server 2005*. Emeryville, McGraw-Hill.
- LAVRAČ, N.; MOTODA, H.; FAWCETT, T.; HOLTE, R.; LANGLEY, P. & ADRIAANS, P. (2004). Introduction: Lessons learned from data mining applications and collaborative problem solving. *Machine Learning*, 57, 13-34.
- LIU, LON-MU & HARRIS, JOHN L. (1993). Dynamic structural analysis and forecasting of residential electricity consumption. *International Journal of Forecasting*, 9, 437-455.
- MENDES, ARMANDO B.; FERREIRA, AIRES & ALFARO, PAULO JORGE (2009). Supporting a Telecommunications Decision with OLAP and Data Mining: A case study in Azores. *CRUZ-CUNHA, VARAJÃO E AMARAL Proceedings of the CENTERIS 2009*. Ofir, Portugal, 537-549.
- SAPORTA, GILBERT (2002). Data fusion and data grafting. *Computational Statistics & Data Analysis*, 38, 465-473.
- SMITH, D. (1989). Combination of forecasts in electricity demand prediction. *Journal of Forecasting*, 8, 349-356.
- TROUTT, M.D.; MUMFORD, L.G. & SCHULTZ, DAVID E. (1991). Using spreadsheet simulation to generate a distribution of forecasts for electric power demand. *Journal of the Operational Research Society*, 42, 931-939.