

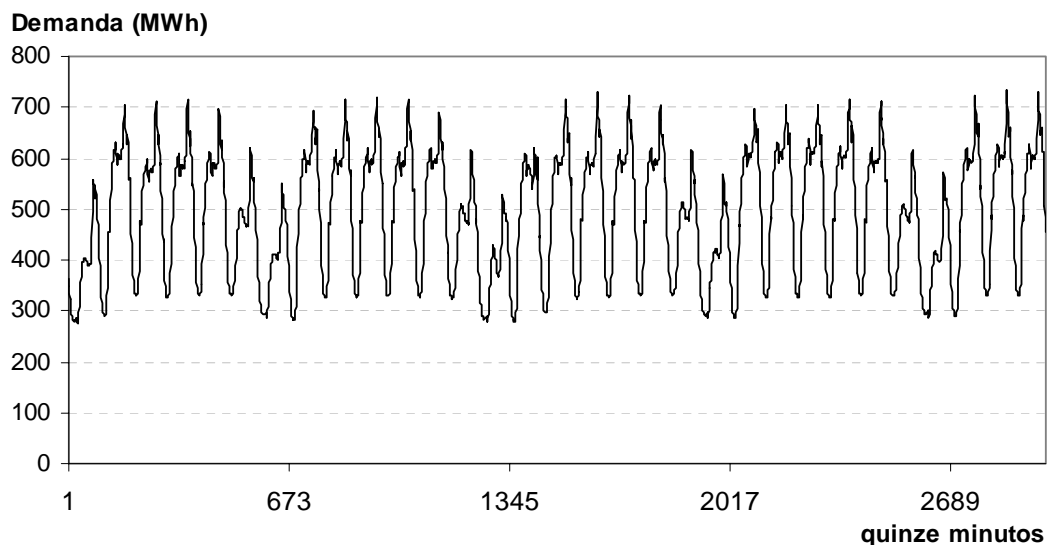
3 – MODELO PROPOSTO

3.1 – A Extensão do Modelo de Holt-Winters Padrão

O modelo Holt-Winters padrão é muito utilizado para a previsão de séries temporais que contém sazonalidade. No entanto, esta modelagem é capaz de comportar apenas um padrão sazonal.

Taylor (2003a) desenvolveu um trabalho no qual previa dados de demanda elétrica no intervalo de trinta minutos. Neste artigo, ele mostrou que havia um ciclo sazonal dentro do dia com duração de 48 períodos (de trinta minutos cada) e também havia um ciclo sazonal dentro da semana, isto é, o perfil da demanda de um dia tinha um comportamento bastante parecido com o perfil da demanda do próximo dia, principalmente ao analisar os dias de semana. Este ciclo dentro da semana, com duração de 336 períodos (esta é a quantidade de trinta minutos em uma semana), fica bastante evidente ao se analisar simultaneamente duas semanas. Estes padrões sazonais evidenciados por Taylor podem também ser verificados em dados do Brasil, como mostra a figura 3.01 abaixo.

Figura 3.01 – Curva de demanda elétrica a cada quinze minutos de 1 maio/2006 (segunda-feira) a 31 de maio/2006 (quarta-feira)



Tendo estes dois padrões sazonais em vista e sabendo que o método de Holt-Winters é conhecido por sua robustez e precisão e por isto vem sendo muito utilizado, Taylor adaptou o método de Holt-Winters padrão de tal forma que ele passasse a acomodar mais do que um padrão sazonal, isto é, múltiplos ciclos.

3.1.1 – Holt-Winters com Dois Ciclos

Este método é adequado para séries com dois ciclos, como aquela mostrada na figura anterior. Ele será apresentado para acomodar dois ciclos, e pode ser facilmente estendido para mais ciclos, caso a série a ser modelada apresente tais características.

O modelo Holt-Winters com dois ciclos, um diário e o outro semanal, é expresso matematicamente abaixo como:

$$Z_t = (S_t + T_t)D_{t-l_1}W_{t-l_2}$$

Onde: S_t o nível

T_t a tendência

D_t o primeiro ciclo

W_t o segundo ciclo

l_1 o comprimento do primeiro ciclo

l_2 o comprimento do segundo ciclo

Sendo que suas equações de atualizações são escritas como:

$$S_t = \alpha \left(\frac{Z_t}{D_{t-l_1}W_{t-l_2}} \right) + (1-\alpha)(S_{t-1} + T_{t-1})$$

$$T_t = \gamma(S_t - S_{t-1}) + (1-\gamma)T_{t-1}$$

$$D_t = \delta \left(\frac{Z_t}{S_t W_{t-l_2}} \right) + (1-\delta)D_{t-l_1}$$

$$W_t = \varpi \left(\frac{Z_t}{S_t D_{t-l_1}} \right) + (1-\varpi)W_{t-l_2}$$

Sendo $\alpha, \gamma, \delta, \varpi$ as constantes de amortecimento

No caso desenvolvido por Taylor para prever demanda de carga elétrica com intervalo de trinta minutos, a primeira sazonalidade representaria o ciclo dentro de dia e a segunda sazonalidade o ciclo dentro da semana.

Vale lembrar que, para a utilização dessas equações de atualização, já são conhecidos os valores dos parâmetros no instante anterior, isto é, em $t-1$. Devido a isto, é necessário que haja um modo para inicialização desses parâmetros em t_0 . Também, é necessário conhecer os valores das constantes de amortecimento. Essas questões serão tratadas ainda neste capítulo.

Com relação aos fatores sazonais, cabe ressaltar que somente os fatores dos ciclos 1 e 2 correspondentes ao instante T são atualizados em T (de acordo com as equações anteriores). Para os demais fatores temos que:

$$D_i(t) = D_i(t-1); \quad \text{se } i \neq (t-l_1); \text{ sendo } i = 1,2,\dots,l_1$$

$$W_j(t) = W_j(t-l_2); \quad \text{se } j \neq (t-l_2); \text{ sendo } j = 1,2,\dots,l_2$$

Além disso, é recomendável manter a restrição existente no modelo Holt-Winters multiplicativo padrão, que diz que a soma dos fatores sazonais deve ser igual ao comprimento do ciclo. Por isto, é necessário normalizar os fatores a cada atualização, de forma que:

$$\sum_{i=1}^{l_1} D_i(i) = l_1$$

$$\sum_{i=1}^{l_2} W_i(i) = l_2$$

Por fim, a equação de previsão τ passos à frente feita na origem é:

$$\hat{Z}_t(\tau) = (S_t + \tau T_t) D_{t-l_1+\tau} W_{t-l_2+\tau}$$

A formulação expressa acima representa o modelo de Holt-Winters multiplicativo com dois ciclos sazonais. Contudo, um Holt-Winters aditivo com dois ciclos pode ser desenvolvido de forma similar a partir do método de Holt-Winters aditivo padrão. Além do mais, outros padrões sazonais podem ser

acrescentados, bastando para tal incluir o índice sazonal e a equação de amortecimento para cada ciclo adicional.

3.2 – Inicialização dos parâmetros

Para a inicialização dos parâmetros, Taylor (2003a) utilizou o mesmo procedimento utilizado em Williams & Miller (1999). A mesma idéia será utilizada nesta dissertação, com pequenas alterações, como explicado abaixo.

A fim de facilitar a explicação, a série temporal será trabalhada na forma matricial, sendo que nas linhas estarão os dias a serem utilizados na definição dos parâmetros iniciais e nas colunas estará o número de períodos dentro do dia para o qual a previsão está sendo feita, isto é, se for previsão horária, a matriz terá vinte e quatro colunas, cada uma representando uma hora do dia; se for previsão a cada trinta minutos, a matriz terá 48 colunas, que irão representar os quarenta e oito períodos de trinta minutos existentes em um dia; se for previsão para cada quinze minutos, a matriz terá 96 colunas. Assim, considerando Z_T a série temporal, a notação $Z_{1,L1}$ significa que este é o dado histórico do primeiro dia, referente ao último período do dia (no intervalo em que a previsão está sendo feita). Então a série temporal pode ser representada por:

$$\begin{array}{cccc} Z_{1,1} & Z_{1,2} & \dots & Z_{1,L1} \\ Z_{2,1} & Z_{2,2} & \dots & Z_{2,L1} \\ Z_{3,1} & Z_{3,2} & \dots & Z_{3,L1} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ Z_{N,1} & Z_{N,2} & \dots & Z_{N,L1} \end{array}$$

Sendo: L1 o intervalo no qual a previsão está sendo feita (isto é, se horária será 24);

N o número de dias que está sendo utilizado para calcular os parâmetros iniciais.

O nível inicial, S_0 , é calculado como sendo a média da primeira semana de dados, isto é:

$$S_0 = \frac{\sum_{i=1}^7 \sum_{j=1}^{L_1} Z_{i,j}}{7 * L_1}$$

A tendência inicial, T_0 , é estimada subtraindo da média da segunda semana de dados a média da primeira semana de dados e dividindo o resultado pelo número de observações que há em uma semana, como abaixo:

$$T_0 = \frac{\left(\frac{\sum_{i=8}^{14} \sum_{j=1}^{L_1} Z_{i,j}}{7 * L_1} \right) - \left(\frac{\sum_{i=1}^7 \sum_{j=1}^{L_1} Z_{i,j}}{7 * L_1} \right)}{7 * L_1}$$

A inicialização do ciclo diário é calculada pelos passos a seguir:

- 1º passo – cálculo da média de cada dia, o que irá resultar em um vetor com N posições:

$$\bar{Z}_i = \frac{\sum_{j=1}^{L_1} Z_{i,j}}{L_1}, \text{ para } i=1, 2, \dots, N$$

- 2º passo – Divisão de cada elemento da linha por sua correspondente média, o que irá gerar a seguinte matriz:

$$\begin{array}{cccc} X_{1,1} & X_{1,2} & \dots & X_{1,L_1} \\ X_{2,1} & X_{2,2} & \dots & X_{2,L_1} \\ X_{3,1} & X_{3,2} & \dots & X_{3,L_1} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ X_{N,1} & X_{N,2} & \dots & X_{N,L_1} \end{array}$$

Sendo: $X_{i,j} = \frac{Z_{i,j}}{\bar{Z}_i}$ para $i=1, 2, \dots, N$ e $j=1, 2, \dots, L_1$

- 3º passo – calcular a média de cada coluna da matriz advinda do passo anterior, resultando em um vetor com L_1 posições. Já se têm os fatores sazonais, porém ainda é necessário normalizá-los.

$$Y_j = \frac{\sum_{i=1}^N X_{i,j}}{N} \text{ para } j=1, 2, \dots, L_1$$

- 4º passo – normalizar o vetor encontrado no passo anterior, o que ocorre dividindo-se cada elemento pela soma do vetor e multiplicando-se pelo comprimento desse primeiro ciclo, isto é:

$$D_j = \left(\frac{Y_j}{\sum_{j=1}^{L_1} Y_j} \right) \times L_1 \text{ para } j=1, 2, \dots, L_1$$

Este último passo irá resultar nos fatores diários, que formam um vetor com L_1 posições.

A seguir serão apresentados os passos utilizados para a estimação dos fatores correspondentes ao ciclo semanal.

- 1º passo – organizar a matriz de dados. Para o cálculo dos fatores semanais é necessário que a matriz contenha em suas linhas os dados pertencentes a uma semana e em suas colunas o número de semanas a ser utilizado nesta estimação. No caso estudado neste trabalho, a matriz terá 672 colunas, correspondendo a todos os períodos de 15 minutos em cada semana.

$$\begin{array}{cccccccccccc}
 Z_{1,1} & Z_{1,2} & \dots & Z_{1,L_1} & Z_{2,1} & Z_{2,2} & \dots & Z_{2,L_1} & \dots & Z_{7,1} & Z_{7,2} & \dots & Z_{7,L_1} \\
 Z_{8,1} & Z_{8,2} & \dots & Z_{8,L_1} & Z_{9,1} & Z_{9,2} & \dots & Z_{9,L_1} & \dots & Z_{14,1} & Z_{14,2} & \dots & Z_{14,L_1} \\
 Z_{15,1} & Z_{15,2} & \dots & Z_{15,L_1} & Z_{16,1} & Z_{16,2} & \dots & Z_{16,L_1} & \dots & Z_{21,1} & Z_{21,2} & \dots & Z_{21,L_1} \\
 \cdot & \cdot & & \cdot & & & & & & & & & \\
 \cdot & \cdot & & \cdot & & & & & & & & & \\
 \cdot & \cdot & & \cdot & & & & & & & & & \\
 Z_{N-6,1} & Z_{N-6,2} & \dots & Z_{N-6,L_1} & Z_{N-5,1} & Z_{N-5,2} & \dots & Z_{N-5,L_1} & \dots & Z_{N,1} & Z_{N,2} & \dots & Z_{N,L_1}
 \end{array}$$

Esta matriz pode ser simplificada, gerando uma matriz da seguinte forma:

$$\begin{array}{cccc}
 Q_{1,1} & Q_{1,2} & \dots & Q_{1,L2} \\
 Q_{2,1} & Q_{2,2} & \dots & Q_{2,L2} \\
 Q_{3,1} & Q_{3,2} & \dots & Q_{3,L2} \\
 \cdot & \cdot & & \cdot \\
 \cdot & \cdot & & \cdot \\
 \cdot & \cdot & & \cdot \\
 Q_{M,1} & Q_{M,2} & \dots & Q_{M,L2}
 \end{array}$$

Sendo: M o número de semanas existentes nos dados que estão sendo utilizados na estimação dos parâmetros;

L2 é o número de períodos (observações) dentro de uma semana (isto equivale ao comprimento do ciclo semanal);

Q engloba as informações contidas em uma semana.

- 2º passo – cálculo da média de cada semana, o que irá resultar em um vetor com M posições:

$$\bar{Q}_i = \frac{\sum_{j=1}^{L_2} Q_{i,j}}{L_2}, \text{ para } i=1, 2, \dots, M$$

- 3º passo – Divisão de cada elemento da linha por sua correspondente média, o que irá gerar a seguinte matriz:

$$\begin{array}{cccc}
 X^*_{1,1} & X^*_{1,2} & \dots & X^*_{1,L2} \\
 X^*_{2,1} & X^*_{2,2} & \dots & X^*_{2,L2} \\
 X^*_{3,1} & X^*_{3,2} & \dots & X^*_{3,L2} \\
 \cdot & \cdot & & \cdot \\
 \cdot & \cdot & & \cdot \\
 \cdot & \cdot & & \cdot \\
 X^*_{M,1} & X^*_{M,2} & \dots & X^*_{M,L2}
 \end{array}$$

Sendo: $X^*_{i,j} = \frac{Q_{i,j}}{Q_i}$ para $i=1, 2, \dots, M$ e $j=1, 2, \dots, L_2$

- 4º passo – calcular a média de cada coluna da matriz encontrada no passo anterior, isto é, a média de cada período, resultando em um vetor com L_2 posições.

$$Y^*_j = \frac{\sum_{i=1}^M X^*_{i,j}}{M} \text{ para } j=1, 2, \dots, L_2$$

- 5º passo – os valores encontrados no passo anterior serão divididos pelos respectivos valores iniciais de seus fatores diários:

$$U_{L_1k+j} = \frac{Y^*_{L_1k+j}}{D_j}, \text{ para } j=1, \dots, L_1 \text{ e } k=0, \dots, 6$$

- 6º passo – normalizar o vetor encontrado no passo anterior, o que ocorre dividindo-se cada elemento pela soma do vetor e multiplicando-se pelo comprimento desse primeiro ciclo, isto é:

$$W_j = \left(\frac{U_j}{\sum_{j=1}^{L_2} U_j} \right) \times L_2, \text{ para } j=1, 2, \dots, L_2$$

3.3 – Estimação das Constantes de Amortecimento

As constantes de amortecimento são definidas através da minimização da média dos erros ao quadrado (*mean square error* – MSE) na previsão um passo à frente, realizado no conjunto de dados destinado para esta definição e para a inicialização dos parâmetros.

O MSE é definido por:

$$MSE = \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{N}$$

Sendo: N o número de previsões realizadas;

y_i a saída desejada para a previsão no instante i , isto é, o valor real em i ;

\hat{y}_i a previsão no instante i .

Há vários tipos de otimização, tanto linear quanto não linear, os quais poderiam ser usados neste problema. Para este trabalho utiliza-se uma otimização através de um algoritmo genético, que foi implementado para definir as constantes de amortecimento através da minimização do MSE. No anexo A há uma apresentação deste método.

3.4 – Tratamento de Feriado

Assim como em outros países, o Brasil possui uma gama de feriados, sejam eles nacionais ou estaduais, os quais geram um perfil de demanda de carga elétrica totalmente diferente de todos os outros dias do ano. Mesmo assim, não é comum encontrar na literatura metodologia para tratamentos específicos para esses casos. Como citado por Taylor (2003a) e Taylor, Menezes & McSharry (2006), o mais comum é utilizar os métodos de previsão para períodos em que não há feriados.

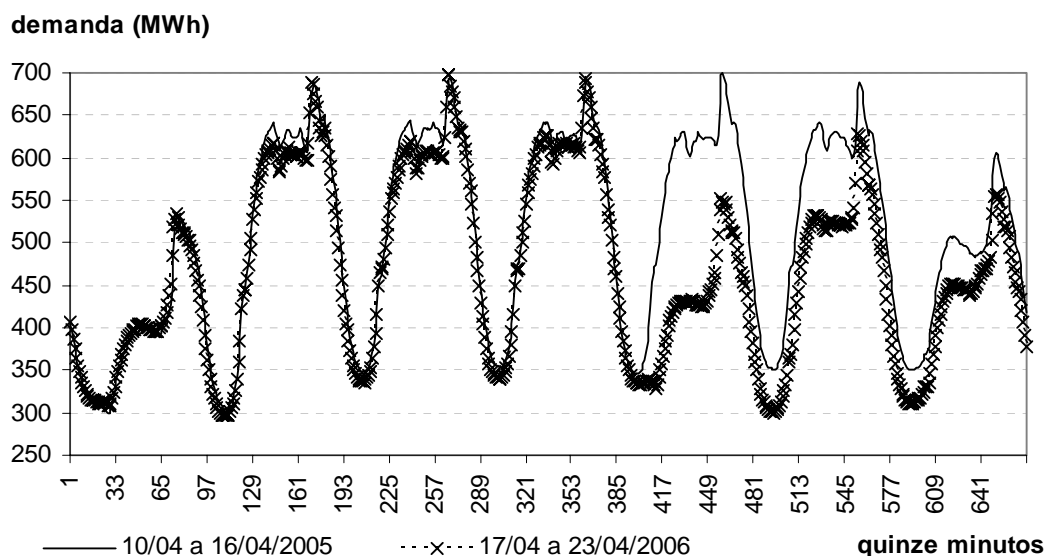
Além de causarem impacto na previsão, os feriados também distorcem a estimação dos hiperparâmetros e dos valores iniciais. Neste caso, o que é proposto por Taylor, Menezes & McSharry (2006) é o tratamento desses dias como *missing values*. Além disso, Laing & Smith (1987) sugerem que os valores históricos desses dias sejam substituídos por valores amortecidos, o que pode ser feito através da média do dia equivalente aos dias das semanas anteriores e posteriores ao dia em questão.

Contudo, tratando apenas do meio acadêmico ainda é possível ignorar os feriados ao testar o poder de previsão de um modelo, o que não ocorre no mercado, uma vez que as concessionárias de energia elétrica necessitam de previsões acuradas nestes dias a fim de evitar uma sobrecarga no sistema ou mesmo uma sub-utilização da carga gerada.

Na literatura, há algumas publicações que trazem o resultado do tratamento de feriados para aprimorar a acurácia do modelo de previsão (Cancelo, Espasa & Grafe (2007), Cancelo & Espasa (1996), Papalexopoulos & Hesterberg (1990)), tendo esses a idéia básica de detectar padrões no comportamento da demanda nos dias em que houve feriado.

Além disso, observou-se que a ocorrência de um feriado distorce não somente a carga desse dia. Dependendo do dia da semana em que o feriado ocorre, ele poderá acarretar modificações no perfil das últimas horas do dia anterior ao feriado e das primeiras horas do dia posterior, como também poderá modificar as características de um dia inteiro ou vários dias antes e após o dia do feriado, como mostrado na figura 3.02.

Figura 3.02 – Demanda de carga elétrica em uma semana que ocorreu um feriado e em uma semana onde não houve feriado



Na figura 3.02, tem-se a demanda de carga elétrica para uma área do Brasil para cada quinze minutos do dia, considerando uma semana em que houve feriado e a semana anterior a esta (em que não ocorreu feriado). O feriado em questão é o

dia 21 de abril. No ano exibido, esse feriado ocorreu em uma quinta-feira e como se pode observar, ele afetou o perfil de demanda de energia elétrica não somente do dia do feriado, mas os dois dias subsequentes também.

Cabe ressaltar aqui que três feriados devido as suas características bem específicas (vide figuras 3.03, 3.04 e 3.05), foram analisados separadamente, são eles: Natal, Ano Novo e Carnaval. Para estes feriados foram criados fatores de redução específicos, sendo eles analisados da mesma forma que a descrita acima. Como o feriado de Carnaval ocorre sempre no mesmo dia da semana, o agrupamento ocorreu somente neste dia. Já o feriado de Natal e Ano Novo foi analisado de acordo com o dia da semana em que ocorreu.

Figura 3.03 – Feriado de Carnaval (ocorreu dia 08/02)

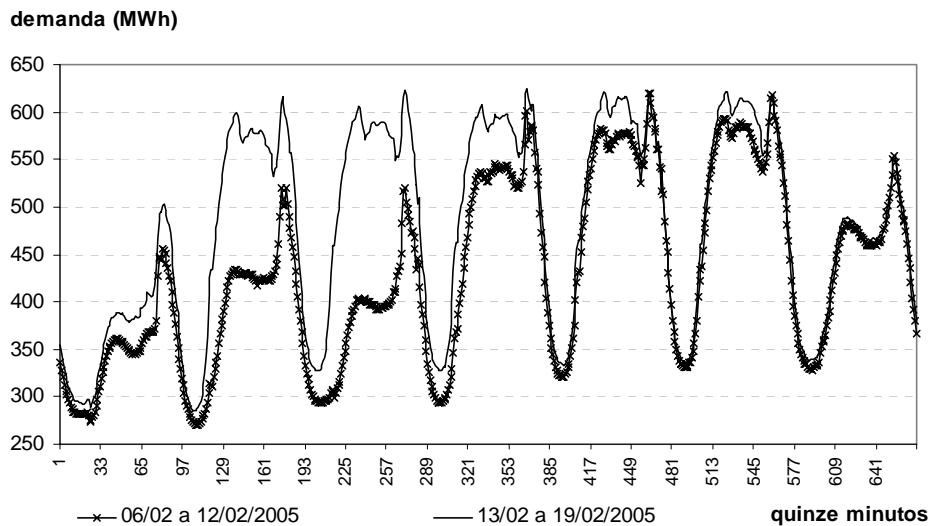


Figura 3.04 – Feriado do Natal

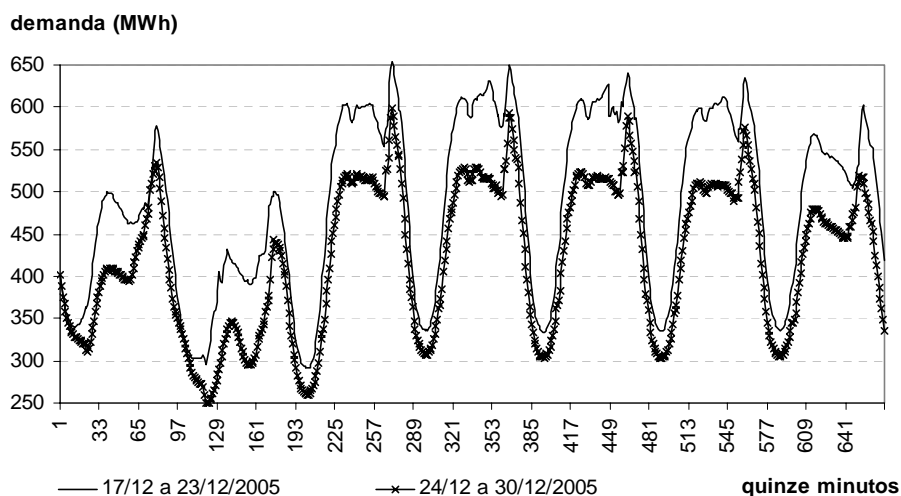
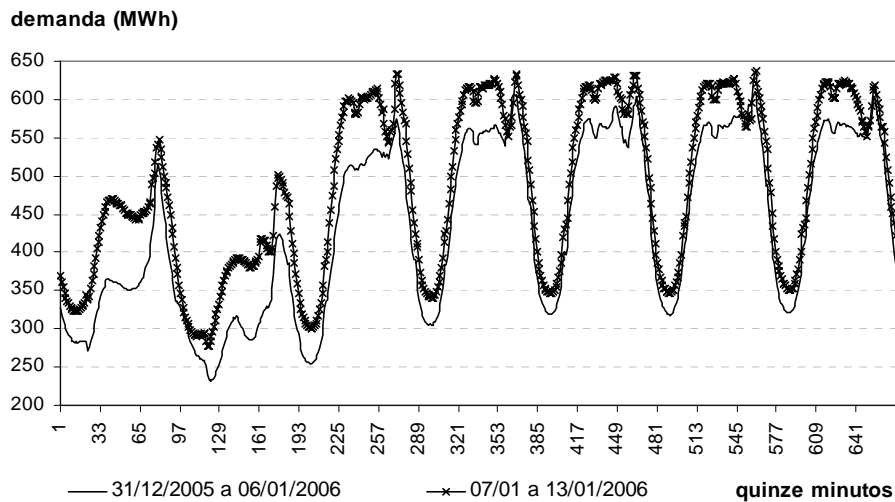


Figura 3.05 – Feriado do Ano Novo



Com isto em vista, desenvolveu-se um banco de regras para ser incorporado ao modelo de previsão de forma exógena, que contém fatores de redução/aumento que serão incorporados à previsão da energia, sendo assim capaz de ajustar melhor o modelo, reduzindo o erro de previsão.

Como o feriado gera comportamentos distintos em dias da semana diferentes, foram criadas regras para cada dia da semana, analisando também os dias (ou partes de dias) que são influenciados por esse feriado. Essas regras ainda foram ajustadas no mesmo período para o qual a previsão está sendo realizada, a fim de proporcionar uma melhor correção, isto é, se a previsão for horária, as regras para o feriado também serão, sendo então necessário analisar o que ocorre em cada hora de um dia de feriado e em cada hora dos dias próximos ao feriado; se a previsão for realizada para intervalos de 15 minutos, as análises sobre o feriado também deverão ser feitas para cada 15 minutos do dia, criando assim uma regra para um desses minutos analisados.

No caso estudado nessa dissertação, estavam disponíveis quatro anos de dados, o que para análise de feriados não é muito (dada uma média de quinze feriados por ano, sem contar os feriados-ponte¹, tem-se algo em torno de 60 casos de feriados no período estudado), mas já é suficiente para proporcionar algumas

¹ Dia em que não é feriado, mas tem o comportamento similar ao de um feriado devido à ocorrência de feriado no dia posterior ou anterior. Por exemplo, se ocorre um feriado em uma terça-feira, pode-se considerar a segunda-feira como feriado-ponte, o mesmo ocorrendo quando o feriado é em uma quinta-feira, considera-se a sexta-feira como feriado ponte.

análises capazes de reduzir os erros do modelo de previsão gerados para esses dias.

O primeiro passo desta análise foi identificar os dias que foram feriados (daqui por diante, neste inclusive, todas as vezes que for citada a palavra “feriado” estará implícita a ocorrência dos feriados-ponte também²) e os dias que não foram. Com isso, encontrou-se a diferença percentual do dia do feriado em relação ao mesmo dia da semana anterior e posterior (pois se entende que na semana antecedente e subsequente, ao observar o mesmo dia da semana, a carga terá um montante parecido, caso nenhum fato extraordinário como o fato de feriado ocorra). Isto foi feito não somente no dia do feriado, mas em aproximadamente uma semana em torno do feriado, o que visa entender que dias são afetados pelo feriado em questão.

Cabe destacar que o que se busca com isso é analisar a variação de carga elétrica de um dia normal para um dia com feriado. Além disso, busca-se também verificar se houve mudanças significativas nos dias próximos ao feriado devido à ocorrência deste. Por isto, ao fazer esta análise deve-se tomar o cuidado de sempre comparar o dia do feriado e seus dias próximos com dias em que não ocorreu feriado. Se na semana anterior/posterior a um feriado que está sendo estudado, também ocorreu um outro feriado, este dia do feriado e os dias próximos a ele não poderão ser utilizados na análise. Neste caso, o que se recomenda é retroceder/avançar uma semana a mais para fazer a análise. Apesar desse fato não ocorrer com tanta frequência, não é possível descartá-lo.

Abaixo segue uma exemplificação deste primeiro passo. Nesta, será mostrado a ocorrência de um feriado, utilizando a notação matricial anterior acrescentada de duas colunas, sendo que a primeira coluna da matriz é referente ao dia da semana (1 para domingo até 7 para sábado), a segunda coluna mostra se é feriado ou não (1 para feriado, 0 para não feriado), a terceira coluna em diante refere-se à carga do dia, no período estudado (aqui generalizado por L1).

² Salvo os casos que tiver indicação contrária.

4	0	Z _{1,1}	Z _{1,2}	Z _{1,3}	...	Z _{1,L1}
5	0	Z _{2,1}	Z _{2,2}	Z _{2,3}	...	Z _{2,L1}
6	0	Z _{3,1}	Z _{3,2}	Z _{3,3}	...	Z _{3,L1}
7	0	Z _{4,1}	Z _{4,2}	Z _{4,3}	...	Z _{4,L1}
1	0	Z _{5,1}	Z _{5,2}	Z _{5,3}	...	Z _{5,L1}
2	0	Z _{6,1}	Z _{6,2}	Z _{6,3}	...	Z _{6,L1}
N	N	N	N	N	N	N
7	0	Z _{11,1}	Z _{11,2}	Z _{11,3}	...	Z _{11,L1}
1	0	Z _{12,1}	Z _{12,2}	Z _{12,3}	...	Z _{12,L1}
2	0	Z _{13,1}	Z _{13,2}	Z _{13,3}	...	Z _{13,L1}
3	1	Z _{14,1}	Z _{14,2}	Z _{14,3}	...	Z _{14,L1}
4	0	Z _{15,1}	Z _{15,2}	Z _{15,3}	...	Z _{15,L1}
N	N	N	N	N	N	N
1	0	Z _{26,1}	Z _{26,2}	Z _{26,3}	...	Z _{26,L1}
2	0	Z _{27,1}	Z _{27,2}	Z _{27,3}	...	Z _{27,L1}

Feito isto, como já explicado, é necessário calcular a variação percentual do dia do feriado e dos dias próximos com os outros dias, o que é feito da seguinte forma:

- para a semana anterior ao feriado:

$$\left(\left(\frac{Z_{i-7-k,j}}{Z_{k,j}} * 100 \right) - 100 \right)$$

- para o dia do feriado:

$$\left\{ \left(\left(\frac{Z_{i,j}}{Z_{i-7,j}} * 100 \right) - 100 \right) - \left(\left(\frac{Z_{i,j}}{Z_{i+7,j}} * 100 \right) - 100 \right) \right\} / 2$$

- para a semana posterior ao feriado:

$$\left(\left(\frac{Z_{i+k,j}}{Z_{i+7+k,j}} * 100 \right) - 100 \right)$$

Onde: i = dia do feriado;

k = 1, 2, ..., 6;

j = periodicidade diária, $j = 1, 2, \dots, L1$.

Para facilitar o entendimento dessa metodologia mais a frente, a carga acima (onde todas as cargas são analisadas, tendo ocorrido feriado ou não) será escrita da seguinte forma: $Z_{H,j}$.

Sendo: H o dia a ser analisado

j o período do dia a ser analisado; $j=1, 2, \dots, L1$

Em seguida, são analisados somente os dias em que não houve feriado, e nestes calculou-se a variação percentual entre cada dia da semana com o mesmo dia da semana anterior, sendo esta análise feita na mesma periodicidade para a qual a previsão está sendo realizada. Com isto, foi possível ter-se uma idéia da variação normal que ocorre entre os mesmos dias da semana, em semanas próximas, em dias de não feriado. A descrição analítica desta etapa será apresentada abaixo, para duas semanas de dados (este procedimento deverá ser feito para todo o banco de dados), sendo que a primeira coluna dessa representa os dias da semana, a segunda mostra a ocorrência de feriados (que neste caso será composta somente de zeros, pois será analisado somente dias de não feriado) e da terceira coluna em diante tem-se a carga elétrica em cada período do dia que está sendo estudado.

1	0	$Z_{1,1}$	$Z_{1,2}$	$Z_{1,3}$...	$Z_{1,L1}$
2	0	$Z_{2,1}$	$Z_{2,2}$	$Z_{2,3}$...	$Z_{2,L1}$
N	N	N	N	N	N	N
7	0	$Z_{7,1}$	$Z_{7,2}$	$Z_{7,3}$...	$Z_{7,L1}$
1	0	$Z_{8,1}$	$Z_{8,2}$	$Z_{8,3}$...	$Z_{8,L1}$
2	0	$Z_{9,1}$	$Z_{9,2}$	$Z_{9,3}$...	$Z_{9,L1}$
N	N	N	N	N	N	N
7	0	$Z_{14,1}$	$Z_{14,2}$	$Z_{14,3}$...	$Z_{14,L1}$

A variação normal existente para os dias de não feriado é calculada por:

$$\left(\left(\frac{Z_{i,j}}{Z_{i-7,j}} * 100 \right) - 100 \right)$$

Onde: i = dia a ser analisado, $i= 8, \dots, r$;

r = número de dias existentes, de não feriado, no banco de dados;

j = periodicidade diária, $j = 1, 2, \dots, L1$.

Esta matriz irá gerar uma outra, sendo esta variação separada pelo dia da semana em que ocorreu, da seguinte forma:

$$\begin{array}{cccc} P_{D,1,1} & P_{D,1,2} & \dots & P_{D,1,L1} \\ P_{D,2,1} & P_{D,2,2} & \dots & P_{D,2,L1} \\ \mathbb{N} & \mathbb{N} & \mathbb{N} & \mathbb{N} \\ P_{D,N_d,1} & P_{D,N_d,2} & \dots & P_{D,N_d,L1} \end{array}$$

$$\text{Sendo } P_{D,q,j} = \left(\left(\frac{Z_{D,i_d,j}}{Z_{D,i_d-7,j}} * 100 \right) - 100 \right)$$

D = o dia da semana que está sendo avaliado, $D = 1, 2, \dots, 7$.

j = periodicidade diária, $j = 1, 2, \dots, L1$.

q = o dia que está sendo analisado, $q = 1, \dots, N_d$;

N_d = o número dados existentes por dia da semana, isto é, a quantidade de um mesmo dia da semana, sem a ocorrência de feriado;

Z = a carga que está sendo analisada;

i_d = o dia a ser analisado, por dia da semana.

Para essa variação, entendida como a variação aceitável (que geralmente ocorre) entre os dias de não feriado, foi encontrada a média e a variância, para cada dia da semana, e para cada período. Isto é, para dados de 15 minutos, dadas as variações dentro do dia para cada um desses 15 minutos, os dias da semana foram agrupados; por exemplo, agruparam-se todas as segundas-feiras e encontrou-se a média e a variância de cada uma das cargas do período do dia, nesse caso, cada um dos 15 minutos, como exemplificado abaixo:

$$\mu_{D,j} = \frac{1}{N_d} \sum_{q=1}^{N_d} P_{D,q,j} \quad \text{para } j=1,2,\dots,L1$$

$$\sigma^2_{D,j} = \frac{1}{N_d} \left(\sum_{q=1}^{N_d} (P_{D,q,j} - \mu_{D,j})^2 \right) \quad \text{para } j=1,2,\dots,L1$$

Sendo: $\mu_{D,j}$ a média da periodicidade j para o dia da semana D ;

$\sigma^2_{D,j}$ a variância da periodicidade J para o dia da semana D .

Feito isto, era necessário definir qual a distribuição na qual a variação acima se encaixava, para cada dia da semana, em cada período analisado. Ao visualizar o histograma e o *qq-plot* da média dessas variações, concluiu-se que as mesmas tinham o formato de uma distribuição normal. Sendo assim, foi utilizado como limiar, para saber qual variação nos dias de feriado poderiam ser consideradas normais e quais não poderiam, a seguinte relação: $\text{media} \pm 1,96$ desvio, pois contém cerca de 95% dos dados.

A partir disso, analisou-se a primeira variação encontrada, isto é, a variação dos dias de feriado em relação aos dias em que não eram feriados (primeiro passo), verificando em quais períodos desses dias a variação era superior ao limite escolhido ($\text{media} \pm 1,96$ desvio padrão), de acordo com cada dia da semana e em cada período do dia, isto é:

- para a semana do feriado:

$$\text{SE } \mu_{D,j} + 1,96\sigma_{D,j} > \left(\left(\frac{Z_{i-7-k,j}}{Z_{k,j}} * 100 \right) - 100 \right),$$

$$\text{ENTÃO } \Delta_{D,j,t_D} = \left(\left(\frac{Z_{i-7-k,j}}{Z_{k,j}} * 100 \right) - 100 \right)$$

- para o dia do feriado:

SE

$$\mu_{D,j} + 1,96\sigma_{D,j} > \left\{ \left[\left(\left(\frac{Z_{i,j}}{Z_{i-7,j}} * 100 \right) - 100 \right) - \left(\left(\frac{Z_{i,j}}{Z_{i+7,j}} * 100 \right) - 100 \right) \right] / 2 \right\},$$

$$\text{ENTÃO } \Delta_{D,j,t_D} = \left\{ \left[\left(\left(\frac{Z_{i,j}}{Z_{i-7,j}} * 100 \right) - 100 \right) - \left(\left(\frac{Z_{i,j}}{Z_{i+7,j}} * 100 \right) - 100 \right) \right] / 2 \right\}$$

- para a semana posterior ao feriado:

$$SE \mu_{D,j} + 1,96\sigma_{D,j} > \left(\left(\frac{Z_{i+k,j}}{Z_{i+7+k,j}} * 100 \right) - 100 \right),$$

$$\text{ENTÃO } \Delta_{D,j,t_D} = \left(\left(\frac{Z_{i+k,j}}{Z_{i+7+k,j}} * 100 \right) - 100 \right)$$

Onde: t_D = número de variações significativas por dia da semana, $t_D = 1, 2, \dots, M_D$

i = dia do feriado;

D = dia da semana;

Δ_{D,j,t_D} = variação significativa, pó dia da semana, para o período j .

É necessário verificar em qual D o dia $i-7-k$ está contido.

Com isso, a variação dos feriados foi separada de acordo com os dias da semana em que eles ocorreram, lembrando-se de agrupar junto a esses os dias próximos que também foram impactados por este feriado. Em seguida, tirou-se a média dos períodos analisados dentro do dia, mas somente daqueles que apresentaram variação significativa. Assim, foi montado o banco de regras para ocorrência de feriado, sendo esse feito para cada dia da semana e para os dias adjacentes afetados por este feriado, como mostrado a seguir:

$$B_{D,j} = \frac{1}{M_D} \sum_{t_D=1}^{M_D} \Delta_{D,j,t_D}$$

Sendo $B_{D,j}$ os fatores de redução para o dia da semana D e a periodicidade j .

3.5 – Influência da Temperatura

Vários setores da economia estão expostos às alterações decorrentes de fatores meteorológicos. Dentre esses, vários estudos já mostraram que o mercado de energia elétrica é elástico com relação às variáveis climáticas, principalmente à temperatura (Cancelo & Espasa (1996), Cancelo, Espasa & Grafe (2007), Valor, Meneu & Caselles (2001), Engle, Mustafa, Rice (1992), Yan (1998), Taylor & Buizza (2003)). Assim, o consumo de energia elétrica sofre variações de acordo com as flutuações meteorológicas. Esses fatores, assim como os efeitos causados

por um feriado, não são contemplados em um modelo univariado, que é o modelo base proposto nesta dissertação. Contudo, isso não é um fator determinante para que eles não sejam considerados. A proposta aqui segue a mesma idéia do tratamento de feriados, que consiste em encontrar fatores de desconto, neste caso, serão fatores de acréscimo, que serão aplicados nos dias em que a temperatura tiver influência sob a demanda de energia elétrica.

Vale destacar que a temperatura, apesar de ser a mais importante, não é a única variável climática que influencia o consumo de energia elétrica. O vento e a chuva também são importantes, assim como a luminosidade e a umidade existente durante cada dia. No entanto, é mais difícil encontrar um histórico disponível dessas variáveis (o que no caso dessa dissertação não estava disponível) e, caso fossem utilizadas, seria necessário modelos de previsão específicos para elas, pois qualquer variável causal incorporada ao modelo de previsão também precisará ser prevista.

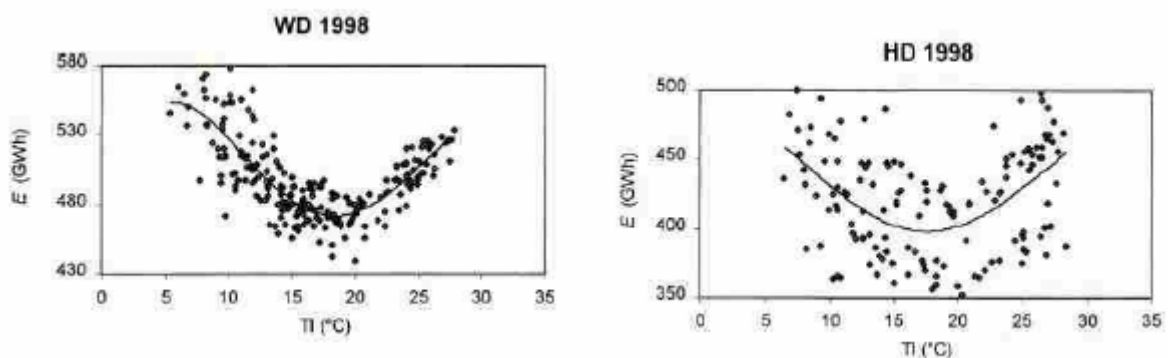
Ao contrário dos países desenvolvidos, nos quais é mais fácil encontrar dados de variáveis climáticas, como descrito por Valor, Meneu & Caselles (2001) que tinha à sua disposição dados de temperatura a cada 30 minutos durante 8 anos, no Brasil isto não ocorre. Para este estudo, estavam disponíveis apenas dados de temperatura máxima e mínima diária, o que gera o problema de não se saber exatamente o comportamento do clima dentro do dia, uma vez que é comum ter-se grandes oscilações climáticas no decorrer do dia. Dado que esta influência será aplicada a um modelo de previsão de demanda de alta frequência, seria recomendável que a temperatura estivesse disponível em intervalos menores. No entanto, isto não impossibilita sua análise, com o objetivo de aumentar a precisão do modelo.

Ao analisar a relação entre o consumo de energia e a temperatura, observa-se que ela é não linear. Assim, a variação no consumo devido à temperatura não é constante, pois depende do nível da temperatura, isto é, uma mudança de 27°C para 28°C não tem o mesmo efeito de uma mudança de 34°C para 35°C. Há também uma diferença que depende da estação do ano, assim a influência da temperatura em estações frias é diferente da influência causada pelas estações quentes. Também, existe um efeito de saturação através do qual variações na temperatura não produzem nenhum efeito no consumo. Além disso, o efeito da

temperatura é diferente de acordo com o dia da semana, ou seja, se é dia útil, fim de semana ou feriado, e também é diferente dentro do dia, em cada período.

Cabe ressaltar ainda uma diferença existente entre o comportamento da demanda de energia em relação à temperatura dos países Europeus e no Brasil. Nos países da Europa, as estações do ano são bem definidas, o que provoca invernos e verões marcantes. Por isso, nesses países, tanto a temperatura muito baixa quanto a muito alta provoca aumento no consumo de energia. Por exemplo, quando há temperatura muito baixa, existe um aumento do uso de aquecedores, já nos dias muito quentes há um aumento do uso de ar condicionado. No Brasil, devido à diferença climática (não temos um inverno com temperaturas tão baixas como as existentes lá), não há o hábito do uso de aquecedores (principalmente na região sudeste, na qual a concessionária em questão está localizada). Por isso, na Europa, o gráfico que mostra a influência da temperatura tem o formato de um “U”, isto é, as temperaturas extremas (tanto muito frias quanto muito quentes) causam um aumento no consumo de energia, como mostrado na figura 3.06, ao contrário dos dados que estão sendo analisados nesta dissertação, nos quais somente o aumento de temperatura eleva a demanda de energia, como mostrado na figura 3.07.

Figura 3.06 – Exemplo da Influência da Temperatura sobre o Consumo na Espanha³

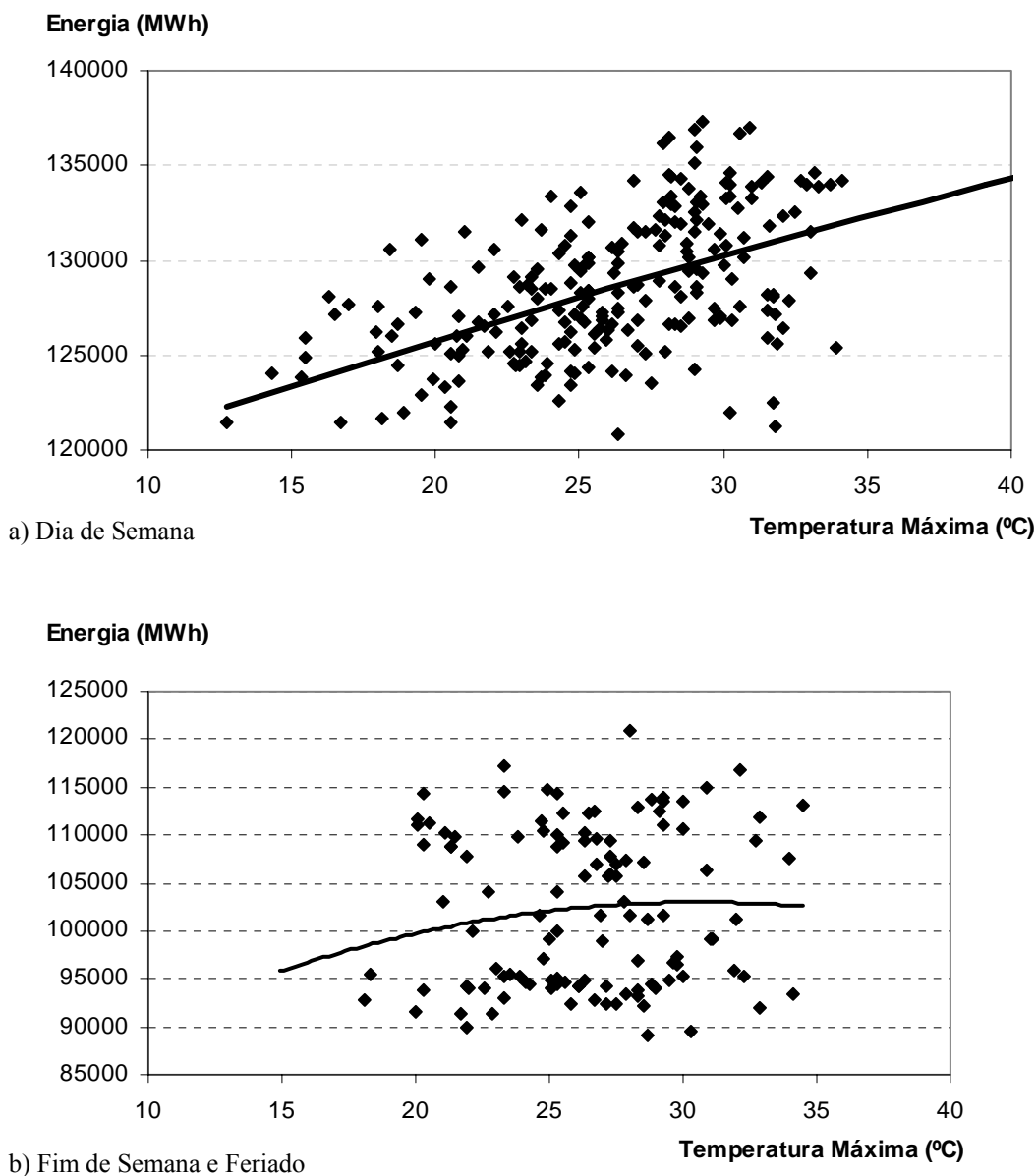


Então, existe um piso a partir do qual a temperatura influencia na demanda e também um teto a partir do qual a temperatura pode aumentar e não causar mais nenhum impacto na demanda. É necessário, ainda, analisar os meses separadamente devido ao efeito das estações do ano, e também seria interessante

³ Valor, Meneu & Caselles (2001).

analisar o efeito nas horas, mas para tanto é necessário ter dados de temperatura em intervalos menores do que o diário ou, pelo menos, ter a temperatura média que ocorreu no dia, o que ajudaria a verificar o efeito da ocorrência de temperaturas extremas. Como só está disponível para este trabalho o valor da temperatura máxima e mínima, decidiu-se verificar o seu impacto em dois momentos distintos: na demanda dos períodos de patamar de carga pesada (que ocorre no período das 8h às 18h) e de carga média (no período das 18h às 22h).

Figura 3.07 – Exemplo da Influência da Temperatura sob o Consumo no Brasil⁴



⁴ Dados do ano de 2006 utilizados nessa dissertação, pertencentes a uma concessionária da região sudeste do Brasil.

Tendo isso em vista, o primeiro passo consiste em agrupar os mesmos meses considerando os anos disponíveis e depois, dentro de cada mês, separar em dias úteis, sábados, domingos/feriados. Em seguida, encontra-se qual foi a demanda máxima diária que ocorreu em cada um dos períodos analisados (carga pesada e média).

Como está sendo analisado um histórico de quatro anos de dados de temperatura, é necessário retirar o efeito de aumento na carga (tendência) durante este período. Por isto será analisada a média da carga de cada dia do mês, para cada ano. Assim, o modelo terá variação local em cada mês, o que significa que será analisada a variação na carga, sendo esta variação calculada com base na média da carga mensal.

Para este procedimento, fazia-se necessário definir o limite a partir do qual a temperatura começa a influenciar a demanda de energia. Após alguns estudos e simulações, decidiu-se adotar como limite a média da temperatura daquele mês (considerando todos os anos). Assim, tendo, por exemplo, um histórico dos meses de janeiro, o limite é a média da temperatura (podendo neste caso ser a máxima ou a mínima) dos dias de janeiro dos anos considerados, como descrito abaixo:

$$\bar{T}_m = \frac{\left(\sum_{y=1}^a \sum_{d=1}^{n_{m,y}} T_{d,m,y} \right)}{\sum_{y=1}^a n_{m,y}}$$

Onde: \bar{T}_m é a média da temperatura do mês m para todos os anos;

$T_{d,m,y}$ é a temperatura (máxima ou mínima) do dia d do mês m do ano y .

Feito isto, será analisado o comportamento da carga dos dias cuja temperatura ultrapassou o limite. Por isso, será calculada a média da carga dos dias que tiveram temperatura abaixo do limite estabelecido, considerando cada ano existente e para cada mês, assim tem-se:

$$\bar{C}_{m,y} = \frac{\sum_{d \in \tau_y} C_{d,m,y}}{[\tau_y]}$$

Onde: C_d é a carga máxima de cada dia no intervalo do patamar (média ou pesada) considerado;

τ_y conjunto de dias do ano y com temperatura abaixo do limiar de temperatura do mês m ;

d é o dia pertencente a τ_y ;

$[\tau_y]$ é o tamanho do conjunto τ_y , isto é, o número de elementos do conjunto τ_y ;

$C_{d,m,y}$ é a carga máxima, no intervalo de patamar considerado, do dia d do mês m e ano y ;

$\bar{C}_{m,y}$ é a média das cargas dos dias $d(\in \tau_y)$, para o mês m e o ano y ;

$m = 1, 2, \dots, 12$, é o mês considerado;

$y = 1, 2, \dots, a$, é o ano considerado.

Em seguida, será analisada a variação na carga entre os dias que tiveram temperatura acima do limite e a carga média dos dias que apresentaram temperatura abaixo do limite:

$$d_{m,y,u} = \begin{cases} \frac{C_{m,y,u}}{\bar{C}_{m,y}}, \forall u \in \mathcal{G}_y \\ 0, \forall u \notin \mathcal{G}_y \end{cases}$$

Onde: \mathcal{G}_y é o conjunto de dias do ano y com temperatura (T_u) maior do que o

limite de temperatura do mês m (\bar{T}_m), $\mathcal{G}_y = \{u | T_u > \bar{T}_m\}$;

u é o dia pertencente a \mathcal{G}_y ;

$C_{m,y,u}$ é a carga máxima, no intervalo de patamar considerado, do dia u do mês m e ano y ;

$d_{m,y,u}$ é a variação na carga do dia u do mês m e ano y ;

$m = 1, 2, \Lambda, 12$ é o mês considerado;

$y = 1, 2, \Lambda, a$ é o ano considerado.

A seguir deve-se calcular a variação da temperatura (dos dias que tiveram temperatura acima do limite, o que significa que serão os mesmos dias da análise anterior) em relação ao limite, como mostra a equação a seguir:

$$t_{m,y,u} = \begin{cases} \frac{T_{m,y,u}}{\bar{T}_m}, \forall u \in \mathcal{G}_y \\ 0, \forall u \notin \mathcal{G}_y \end{cases}$$

Onde: $T_{m,y,u}$ é a temperatura (máxima ou mínima) do dia u do mês m e ano y ;

\bar{T}_m é o limite de temperatura do mês m ;

\mathcal{G}_y é o conjunto de dias do ano y com temperatura (T_u) maior do que o limite de temperatura do mês m (\bar{T}_m), $\mathcal{G}_y = \{u | T_u > \bar{T}_m\}$;

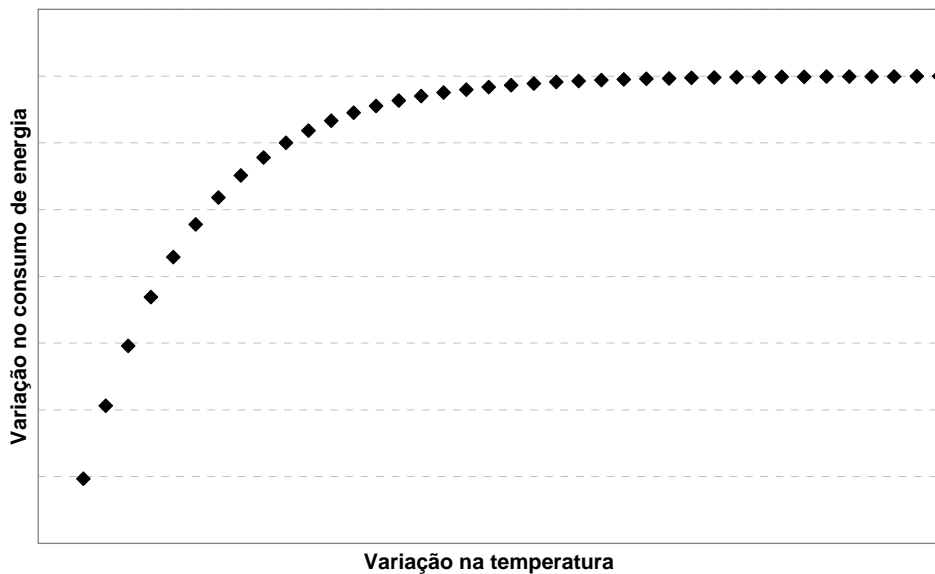
u é o dia pertencente a \mathcal{G}_y ;

$m = 1, 2, \Lambda, 12$ é o mês considerado;

$y = 1, 2, \Lambda, a$ é o ano considerado.

Feito isto, resta definir que tipo de curva deve ser utilizada para representar esse modelo. Vale ressaltar as características necessárias: a relação entre temperatura e demanda de energia é não linear, havendo um limite a partir do qual a variação da temperatura começa a influenciar a demanda e há, ainda, um ponto de saturação a partir do qual a variação na temperatura passa a não influenciar mais no consumo, como é exibido na figura 3.08.

Figura 3.08 – Modelo da curva de impacto da temperatura no consumo de energia elétrica



Uma das possíveis equações a partir da qual é possível chegar ao modelo exemplificado na figura 3.08 é a seguinte:

$$d_{m,y,u} = K_0 + (K_1 - K_0) \left[1 - e^{-\lambda t_{m,y,u}} \right] + erro$$

Onde: K_0 representa o ponto a partir do qual variações na temperatura influenciam a demanda;

K_1 representa a variação máxima que pode ocorrer na demanda devido à temperatura;

λ é a velocidade com que a curva atinge o ponto de saturação.

Como K_0 representa o ponto de início da variação no consumo (até este ponto variações na temperatura não causam nenhum efeito na demanda, assim, o consumo é a própria previsão feita pelo modelo citado no item 1 deste capítulo), ele deve ser igual a 1.

Isso simplifica o modelo para:

$$d_{m,y,u} = 1 + (K_1 - 1) \left[1 - e^{-\lambda t_{m,y,u}} \right] + erro$$

Vale lembrar que K_1 é o efeito máximo da temperatura sob o consumo e este, assim como λ , devem ser estabelecidos. Após uma análise nos dados e também de acordo com a composição da carga da concessionária em questão, na qual o consumo residencial é responsável por uma grande parcela, e ainda analisando o hábito de consumo populacional dessa região, chegou-se à conclusão que o K_1 máximo será igual a 1.2, isto significa que pode haver um aumento em até 20% no consumo devido a variações na temperatura. Feito isto tem-se que minimizar o erro dessa equação a fim de encontrar o K_1 e o λ ideais, o que será feito minimizando a soma dos erros ao quadrado (*sum square of error*), que é definida por:

$$SSE = \sum_u (erro_u)^2$$

$$SSE = \sum_u \left(d_{m,y,u} - 1 - (K_1 - 1) \left[1 - e^{-\lambda t_{m,y,u}} \right] \right)^2$$

Dessa forma, a previsão da demanda de energia irá sofrer mais uma intervenção exógena, a fim de tratar o efeito da temperatura sob o consumo de energia elétrica, proporcionando mais acurácia à previsão. Assim, dado os dias previstos, com suas respectivas temperaturas previstas, analiso a média da carga máxima (carga em questão, pode ser do patamar média ou pesada) dos dias que apresentam temperatura prevista abaixo do limite para a temperatura desse mês, conforme a seguir:

$$\bar{\hat{C}} = \frac{\sum_{q \in \psi} \hat{C}_q}{[\psi]}$$

Onde: ψ é o conjunto de dias previstos com temperatura abaixo do limite de temperatura do mês no qual os dias previstos pertencem;

$[\psi]$ é o tamanho do conjunto ψ ;

\hat{C}_q é a carga prevista dos dias que tiveram temperatura prevista abaixo do limite da temperatura do mês em que os dias pertencem;

$\bar{\hat{C}}$ é a média da carga prevista dos dias que apresentaram temperatura abaixo do limite de temperatura do mês no qual os dias previstos pertencem.

Calcula-se então a média da carga prevista devido a sua temperatura prevista estar acima do limite estabelecido para o mês em que os dias são previstos:

$$\bar{F}_p = \hat{C} \times \left[1 + (K_1 - 1) \left[1 - e^{-\lambda \left(\frac{T_p}{T_m} \right)} \right] \right]$$

Onde: T_p é a temperatura prevista do dia p ;

$p \in \varphi$;

φ é o conjunto dos dias previstos com temperatura acima do limite estabelecido para a temperatura do mês em que os dias foram previstos;

\bar{F}_p é a carga média prevista para o dia p .

Assim, o fator devido o dia previsto apresentar uma temperatura prevista maior que o limite estabelecido é:

$$\eta_p = \frac{\bar{F}_p}{\hat{C}}$$

Com isso, a correção na carga prevista será:

$$\hat{C}_p = C_p \times \eta_p$$

Onde: C_p é a carga prevista para o dia p , com $p \in \varphi$;

\hat{C}_p é a carga prevista corrigida pelo fator devido à temperatura.

Vale lembrar que esta análise é feita para toda a periodicidade no qual a previsão ocorre dentro do período do patamar de carga média e carga pesada. Assim, o mesmo fator η_p será usado durante toda a periodicidade de um dia/patamar, pois ele foi calculado com relação à média da carga prevista.