**APPENDIX O. MODELING TYPICAL METEOROLOGICAL YEAR (TMY) PRICES**

## TABLE OF CONTENTS



## LIST OF TABLES



## LIST OF FIGURES

# APPENDIX O.  MODELING TYPICAL METEOROLOGICAL YEAR (TMY) PRICES

In this appendix, the steps the Department used to construct the load-temperature and price-load models are discussed in detail.

## O.1 LOAD-TEMPERATURE MODELING

### O.1.1 Overview

There are five steps to modeling the load-temperature relationship:

- *Step 1*: The temperature and load data were binned, according to the values of the three variables HE (hour ending), DN (day/night index), and DT (day type, weekday, or weekend/holiday). The subsequent steps were carried out for each bin.

- *Step 2*: The historical data were used to calculate, by a least-squares fit, a polynomial expressing load as a function of temperature, L = F(T).

- *Step 3*: The difference between the historical load and the load predicted by the polynomial was computed and a frequency distribution for these differences was calculated.

- *Step 4*: The polynomial calculated in Step 2 was used to compute the typical meteorological year (TMY) loads as a function of the TMY temperatures in the current bin.

- *Step 5*: The frequency distribution constructed in Step 4 was used to generate a set of random corrections to the TMY loads computed from the polynomial.

### O.1.2 Illustration of the Five Steps

This section illustrates each of the steps listed above, using data for the Midwest (subdivision 3).

#### O.1.2.1 Step 1

The basic relationship between load and temperature is visible in the scatter plots presented in Appendix O. The load rises at both low and high temperatures, as space-conditioning energy use increases. Because air conditioning relies almost exclusively on electricity, while other fuels in addition to electricity are used for heating, electric loads typically rise more as temperatures increase, so most areas of the country are summer-peaking. The exception is the Northwest which, because of its mild summers and significant use of electricity for heating, typically peaks in the winter. For any given temperature, there is a very broad range of possible load values. This is due in large part to the variations in time of day and day-type. For example, a temperature of 60° may occur at 4 a.m. in the summer, or at 1 p.m. in the spring, on a weekday or on a weekend. Loads will be near their minimum at 3 a.m. on a weekend, and near their maximum at 3 p.m. on a weekday. Lighting loads represent a significant fraction of the total load, and will introduce further variation, since early morning or evening hours will correspond to day or night at different times of the year. A much more accurate relationship between load and weather can be developed if the data is first sorted on the hour, day-type. and day-night index. To sort on the day-type, the data must be associated with a particular calendar year. To construct a day-night index, a record of the daily sunrise and sunset times for each subdivision is needed. Given this information, the load-temperature data values for each hour of the year can be assigned to a bin labeled by the variables (Hour,DT,DN). As an example, Figure O.1.1 shows the data for the hour from 12-1 p.m. on a weekday, for the Midwest. The figure shows both the historical data and the TMY data. The polynomial fits are also shown. It is clear that having sorted the data in this fashion leads to a much more regular relationship between load and temperature.

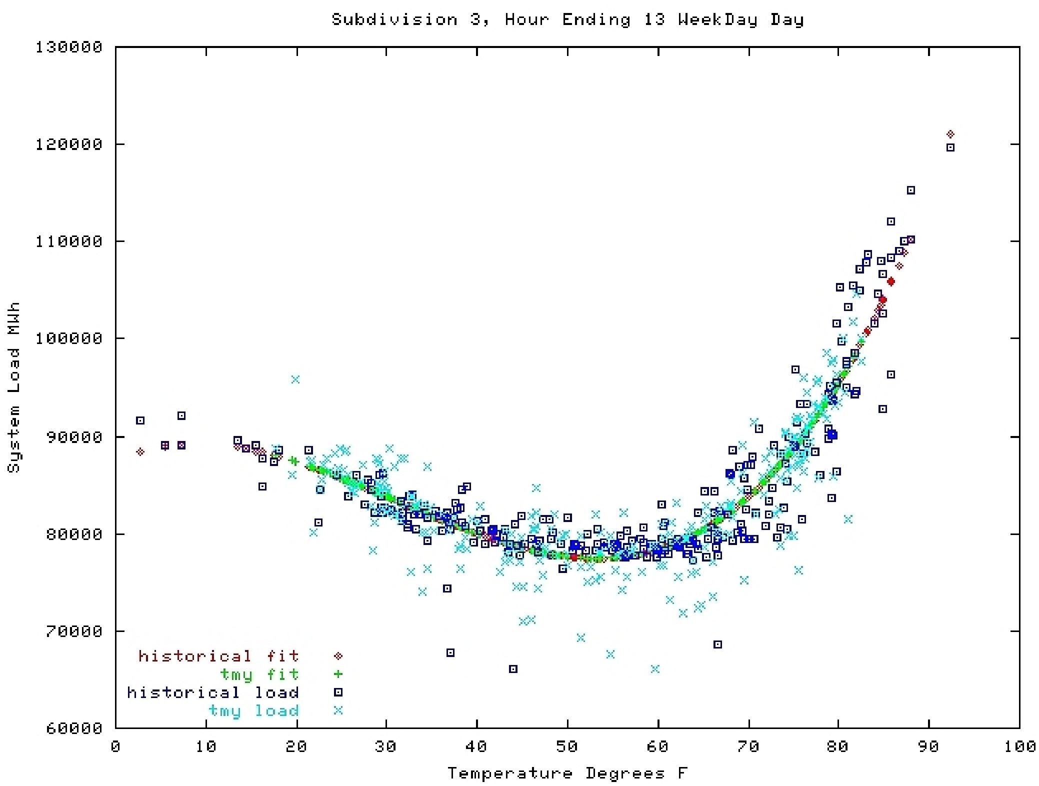


**Figure O.1.1 Load-Temperature Data for the Midwest (Subdivision 3), 12-1 p.m. on a Weekday.**

### O.1.2.2 Step 2

The data shown in Figure O.1.1 was fit to a polynomial using the standard least-squares algorithm. Some experimentation with different models showed that best results were obtained with a fit to a third-order polynomial in the temperature:

$$L = F(T) = a(0) + a(1)*T + a(2)*T^2 + a(3)*T^3$$

In the figure, the red crosses show the loads as computed by this polynomial relationship, while the dark blue squares show the original data. The coefficients $a(k)$ are computed separately for each bin and for each subdivision. It may happen that the range of temperatures seen in the historical data is less extensive than in the TMY data. For example, 1999 saw an unusually warm winter in the Northwest (subdivisions 8.1 and 9.1), so the TMY temperatures are lower than the historical temperatures. In these cases, the load-temperature model needs to extrapolate to low temperatures and to minimize the error the order of the polynomial may be lowered.

### O.1.2.3 Step 3

The polynomial derived in Step 2 gives, for each temperature $T(j)$ in the bin, a "fitted load"

$$LF(j) = a(0) + a(1)*T(j) + a(2)*T^2(j) + a(3)*T^3(j).$$

With $L(j)$ = original load value, the percentage decrement

$$DL(j) = 100*(LF(j) - L(j))/LF(j).$$

The values of $DL(j)$ typically range from -15% to +20%. These values are a measure of the degree to which the actual loads fluctuate around the fitted value predicted by the model. Figure O.1.2 shows a scatter plot of the load decrements $DL$ versus temperature for the historical data. Because there is no significant correlation between the size of the load decrement and temperature, the random fluctuations can be modeled by constructing a frequency distribution of the values $DL$ and using this distribution to generate random decrements for the TMY loads.

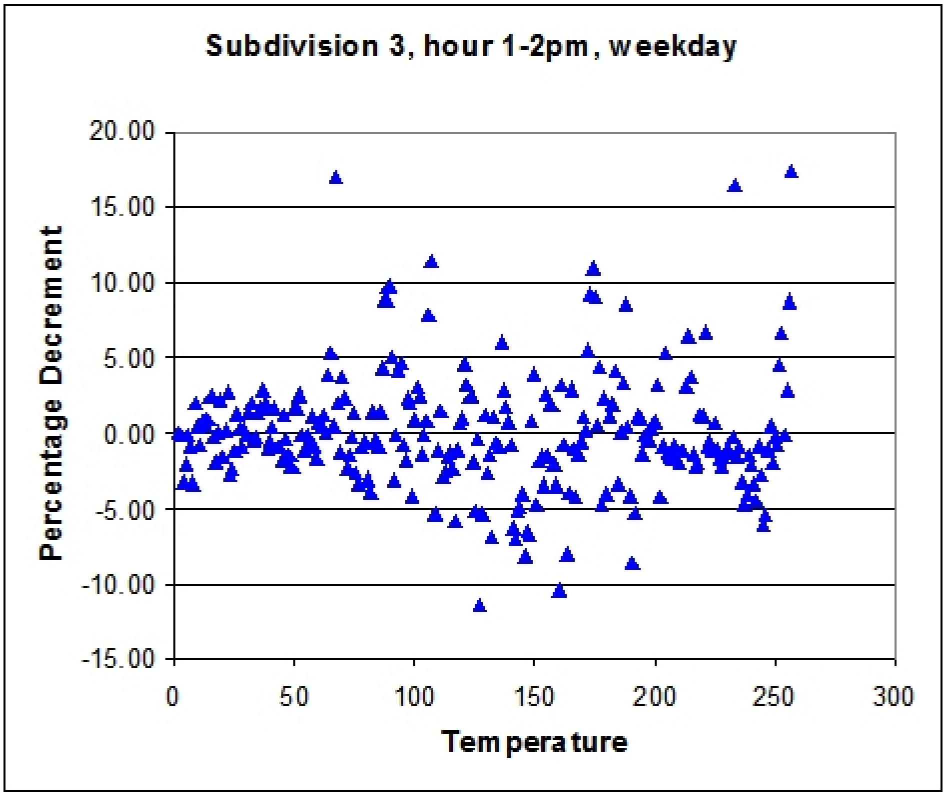


**Figure O.1.2 Scatter Plot of the Percentage Load Decrements DL versus Temperature**

### O.1.2.4 Step 4

The polynomial derived in Step 2 was then used to compute, for each TMY temperature $T_{TMY}(j)$ in the bin, a "fitted TMY load $LF_{TMY}$"

$$LF_{TMY}(j) = a(0) + a(1)*T_{TMY}(j) + a(2)*T_{TMY}{}^{2}(j) + a(3)*T_{TMY}{}^{3}(j).$$

The pairs of values of $(T_{TMY}, LF_{TMY})$ are plotted in Figure O.1.1 as green crosses. By construction, they fall on the same curve as the historical fitted values.

### O.1.2.5 Step 5

The last step was to add random fluctuations to the $LF_{TMY}(j)$ by defining a percentage decrement for the TMY data, $DL_{TMY}(j)$, and using the relationship

$$L_{TMY}(j) = LF_{TMY}(j) - 0.01*DL_{TMY}(j)*LF_{TMY}(j).$$

The values of $DL_{TMY}(j)$ were generated as random numbers distributed according to the frequency distribution constructed in Step 4. The result of adding these decrements to the fitted data is the set of TMY loads to be used in the load-price analysis. These are plotted in Figure O.1.1 in light blue. Figure O.1.3 plots the frequency distributions for the original decrements *DL* and two realizations of the random model for the decrements $DL_{TMY}$, illustrating that the model correctly represents the behavior of the random load fluctuations.

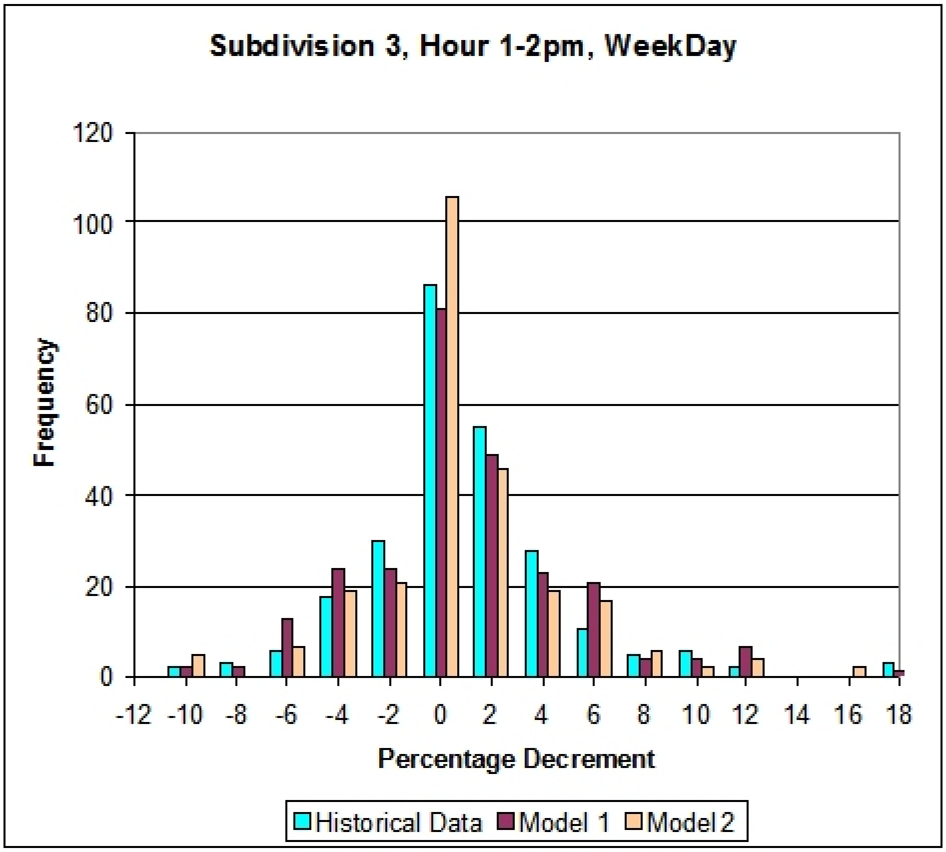


**Figure O.1.3 Frequency Distributions for the Percentage Decrements for the Historical Load Data and Two Realizations of the Random Model**

## O.2 PRICE-LOAD MODELING

### O.2.1 Overview

There are seven steps in the modeling the price-load relationship:

- *Step 1*: The price and load data were binned according to the values of the two variables: HT, which defines the hour type as peak or off-peak and SW, which defines the season as summer or winter. In this analysis, peak hours were from 7 a.m. to 7 p.m. weekdays, with all other hours defined as off-peak. Summer was defined as the months of May through

October, and winter as November through April. The subsequent steps were carried out for each bin.

- *Step 2*: Price spikes in the historical price data were separated out by setting any prices above a cap equal to the cap. The price values and hour of occurrence of the price spikes were saved to be used in Step 7. The cap was set equal to $150/MWh.

- *Step 3*: The price-load data were sorted on load, and further separated into a set of sub-bins, each containing approximately the same number of points. The average load and price in each sub-bin was computed, leading to an approximate relationship PF = G(L). Here PF is the average price in a given sub-bin and L is the average load in that sub-bin.

- *Step 4*: The function calculated in Step 3 was used to compute the fitted TMY price for each value of the TMY load.

- *Step 5*: The difference between the historical price and the fitted price predicted by G(L) was computed and a frequency distribution for these differences was calculated.

- *Step 6*: The frequency distribution constructed in Step 5 was used to generate a set of random corrections to the average TMY prices.

- *Step 7*: The price spikes were added back to the series produced in Step 6.

### O.2.2 Illustration of the Seven Steps

This section illustrates each of the steps listed above, using data for the Midwest (subdivision 3).

#### O.2.2.1 Step 1

The relationship between price and load for each subdivision is illustrated in the scatter plots in Appendix O. While there is an overall trend of higher prices for higher loads, further analysis indicates that the relationship between the two variables is not nearly as well-defined as for load-temperature. Binning the price-load data according to hour of day or day of week does not significantly reduce the spread in prices seen for a given load. The volatility in the price-load relationship generally tends to be highest during summer months and on-peak hours. As these are also the times of maximum air-conditioning energy use, to ensure the accuracy of the model, the data were sorted according to season and hour-type before beginning the quantitative analysis. This gave four data sets, each of which was modeled separately. Subdivision of the data into the four season/hour-type sets is illustrated for the Midwest in Figure O.2.1. This binning procedure was also applied to the TMY load data computed from the load-temperature model.

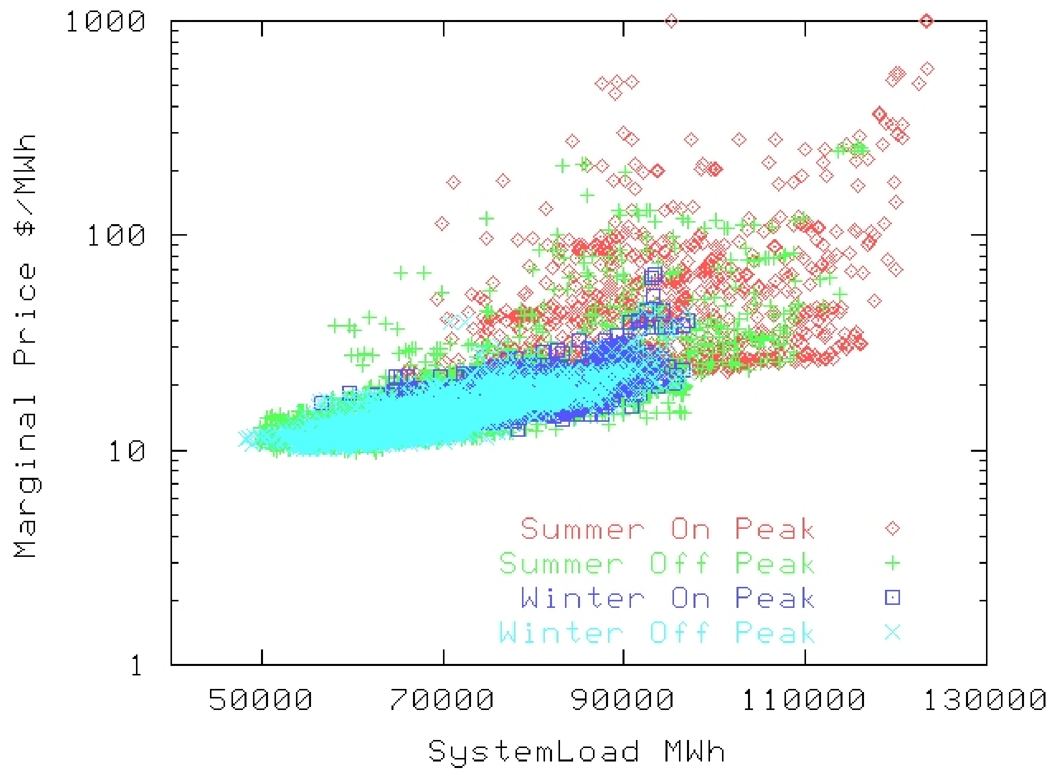


**Figure O.2.1 Price-load Data for the Midwest (Subdivision 3).**

### O.2.2.2 Step 2

Prices also show infrequent but large jumps or spikes which are difficult to model using a continuous distribution (accurate representation of rare events using a continuous distribution requires very large data sets, which were not available for this analysis). However, price spikes may contribute significantly to hourly energy costs, so it is important to represent their effects. The method developed for this analysis removed the price spikes before attempting to model the price-load relationship, then restored them to the modeled data. The advantages of this method are that the spiking behavior is included as observed, with no additional assumptions, and it is also possible to separate out the contribution of the price spikes to the overall energy costs. For *i* equal to the hour index of the data point, *P(i)* the price, and *Pcap* the price cap, for every i, if *P(i)* > *Pcap, P(I)* was set equal to Pcap. The set of values (*i,P(i)*) that were adjusted were recorded for use in Step 7. Steps 3-6 below refer to modeling the capped price data. For this analysis, the cap was set equal to $150/MWh, corresponding to the approximate upper boundary of the cloud of points in the price-load. With this cap, several subdivisions had no price spikes.

#### O.2.2.3 Step 3

Given the considerable irregularity in the price-load relationship, including the fact that prices occasionally decrease slightly as loads go up, it is not sufficiently accurate to fit the data points to a polynomial. Instead, an approximate price-load fit was computed through averaging as follows: the data in a given season/hour-type bin were sorted according to the load and grouped into separate load sub-bins. Each sub-bin was labeled by an index *m*. The minimum and maximum values in the bin were defined as *Lmin(m)* and *Lmax(m)*, respectively. Figure O.2.2 shows the result of this binning process for nine sub-bins; each colored slice represents the data in one bin. The red line represents the $150/Mwh cap. The fitted price *PF(m)* is defined to be the average price in bin *m*. This gives an approximate fit:

$PF(L) = PF(m) \text{ for } Lmin(m) \leq L < Lmax(m).$

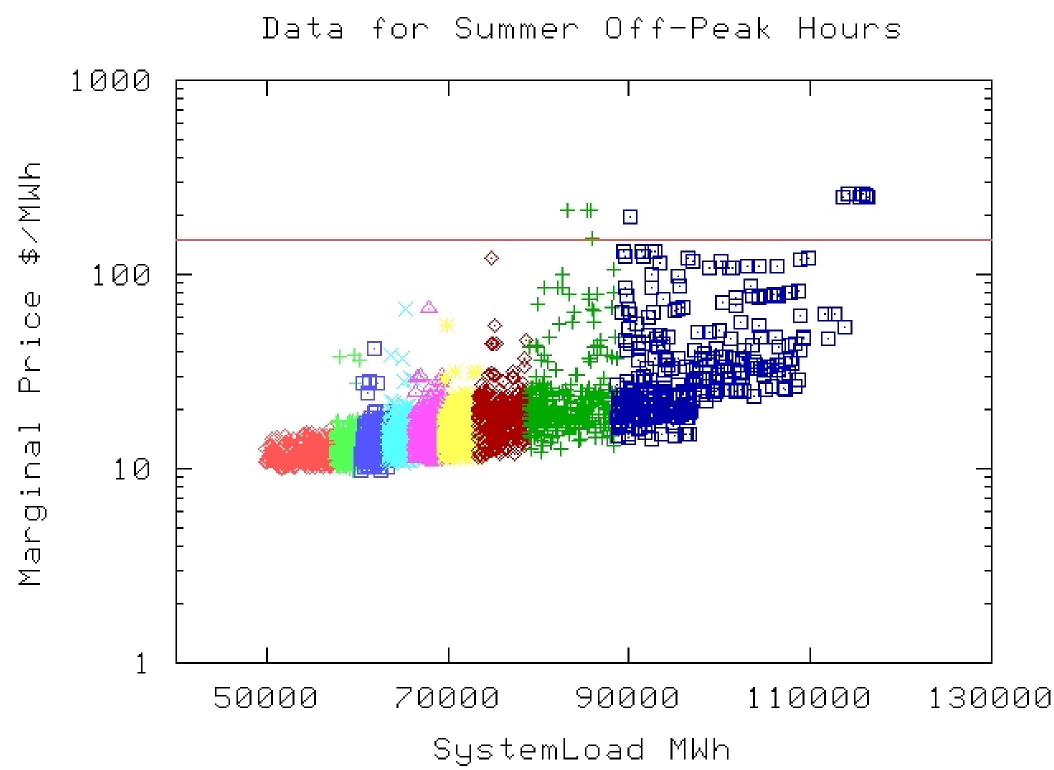


**Figure O.2.2 Price-load Data for the Summer, Off-peak Bin**

#### O.2.2.4 Step 4

The next step was to model the randomness in the price-load relationship. Here a technique similar to that used for the load-temperature data was developed. For each data pair. (*P(j),L(j)*), in load sub-bin *m*, deviations from the fitted price *PF(m)* were computed as

$$DP(j) = PF(m) - P(j), \text{ for } Lmin(m) \leq L(j) < Lmax(m),$$

and a frequency distribution for the deviations *DP(j)* was constructed. This means that the shape of the frequency distribution of *DP* depends on the load sub-bin *m*.

#### O.2.2.5 Step 5

To construct the TMY prices $P_{TMY}$, the relationship developed in Step 3 was used. For every load value $L_{TMY}(j)$, the appropriate fitted value was found:

$$PF_{TMY}(L) = PF(m) \text{ for } Lmin(m) \leq L_{TMY} < Lmax(m).$$

#### O.2.2.6 Step 6

In this step, random decrements to the fitted TMY prices were computed. For every load $L_{TMY}(j)$, the load sub-bin *m* it fell into was located, then a random decrement $DP_{TMY}(j)$ was computed using the appropriate frequency distribution. The TMY price was then defined as

$$P_{TMY}(j) = PF(m) - DP_{TMY}(j), \text{ for } Lmin(m) \leq L_{TMY}(j) < Lmax(m).$$

#### O.2.2.7 Step 7

The last step was to add back the price spikes. This was done separately for each season/hour-type bin. The two important properties of the price spikes that needed to be preserved were their values and the fact that they typically occur in clusters rather than as isolated values. This is because the external conditions which lead to price spikes in the first place will generally persist for several hours. The clustering property was preserved by grouping the price spikes as follows: Given a set of NS price spikes indexed by *k*, each price spike had associated with it an hour index *i(k)* indicating when it occurred in the original time series, and its value *P(i(k))*. The set of price spikes was ordered such that $i(k) < i(k+1)$ for all k. If price spikes occurred in consecutive hours, then $i(k+1) - i(k) = 1$. This property defined a group of price spikes. As an example, a list of price spikes, along with the indices *i(k)* and *k*, are presented in Table N.2.1. The number of spikes NS=10 and there are three groups with seven, two, and one member. The peak load for this bin is 21591 MWh.

**Table O.2.1 List of Price Spikes Occurring for the Summer, Peak Hours for New England (Subdivision 1)**

| k | i(k) | Load(I) | Price(I) |
|---|---|---|---|
| 1 | 3083 | 18018 | 185.78 |
| 2 | 3084 | 18372 | 519.94 |
| 3 | 3085 | 18502 | 1000.00 |
| 4 | 3086 | 18711 | 1000.00 |
| 5 | 3087 | 18656 | 1000.00 |
| 6 | 3088 | 18558 | 1000.00 |
| 7 | 3089 | 18348 | 1000.00 |
| 8 | 3110 | 18928 | 151.36 |
| 9 | 3111 | 18907 | 150.56 |
| 10 | 4285 | 21919 | 343.80 |

The initial hour of a group of price spikes is a random variable, more likely to occur at higher loads. Price spikes were restored to the TMY data as follows: for each group of price spikes, first an hour was chosen at random which belonged to the same season/hour-type bin. If the load in this hour was above 90 percent of the maximum load for this bin, the sequence of prices beginning in that hour were reset to the sequence of spike values in the group. In Table O.2.1, the first group of spikes contains seven consecutive values, so seven consecutive summer peak hours will have their prices reset to these values, with the initial hour chosen at random, subject to the condition that the load in that hour is above 90 percent of 21591 MWh.