

# A Baseline Model for Utility Bill Analysis Using Both Weather and Non-Weather Related Variables<sup>1</sup>

Robert C. Sonderegger, Ph.D.  
SRC Systems, Inc.  
2855 Telegraph Ave. Suite 410  
Berkeley, CA 94705, USA

## Abstract

Many utility bill analyses in the literature rely only on weather-based correlations. While often the dominant cause of seasonal variations in utility consumption, weather variables are far from the only determinant factors. Vacation shutdowns, "Plug Creep," changes in building operation and square footage, and plain poor correlation are all too familiar to the practicing performance contractor.

This paper presents a generalized Baseline Equation consistent with prior results by others, but extended to include other, non-weather related independent variables. Its compatibility with extensive prior research by others is shown, as well as its application to several types of facilities. The Baseline Equation as presented, can accommodate up to five simultaneous independent variables, for a maximum of eight free parameters. The use of two additional, empirical Degree-Day threshold parameters is also discussed.

The Baseline Equation presented herein is at the base of a commercial Utility Accounting software program. All case studies presented to illustrate the development of the Baseline Equation for each facility are drawn from real life studies performed by users of this program.

## Objective and Terminology

Utility performance contracting is about savings estimation and about risk management. The performance contractor promises to the client "I will reduce your annual utility bills by X in return for a payment of Y." Simple in principle, this proposition becomes ambiguous when we consider that utility bills can and do vary substantially from month to month and year to year. How will savings achieved by the contractor's work be distinguished from the naturally occurring fluctuations in utility bills?

Thus the concept of a *Baseline*, a surrogate for monthly utility bills that *would have occurred* in the absence of energy saving measures. Such surrogate utility bills can be calculated by following one of at least two paths: (1) by simulating the facility with a computer program that takes into account all relevant physical building parameters and user actions as they affect utility consumption; and (2) by correlating past utility bills to observable variables and projecting that correlation into the future. This paper is mainly about the second path.

In order to ensure mutual business satisfaction both performance contractor and client must agree on a simple, unambiguous, and technically sound manner to calculate the baseline throughout the term of the contract. This is usually achieved through a correlation of past utility bills with past weather, such as degree-days, and with past usage variables such as hours of occupancy. Once past this hurdle, contractor and client must agree on what to do if changes occur in the facility that materially affect the baseline, such as the addition of a new wing or a change in facility usage schedule. Such changes will result in *Baseline Modifications* or renormalization.

The objective of this paper is to present quantitative methodologies for calculating baselines and baseline modifications. This includes the following common questions asked by a performance contractor:

---

<sup>1</sup> To be presented at the Symposium on *Baseline Calculations For Measurement And Verification Of Energy And Demand Savings*, ASHRAE Summer Meeting, Toronto, Canada, June 18-25, 1998.

- How many past utility bills do I need to establish a baseline?
- What should I correlate utility bills with and how?
- How should I take into account the impact of vacations and holiday shutdowns?
- How do I handle plug creep?

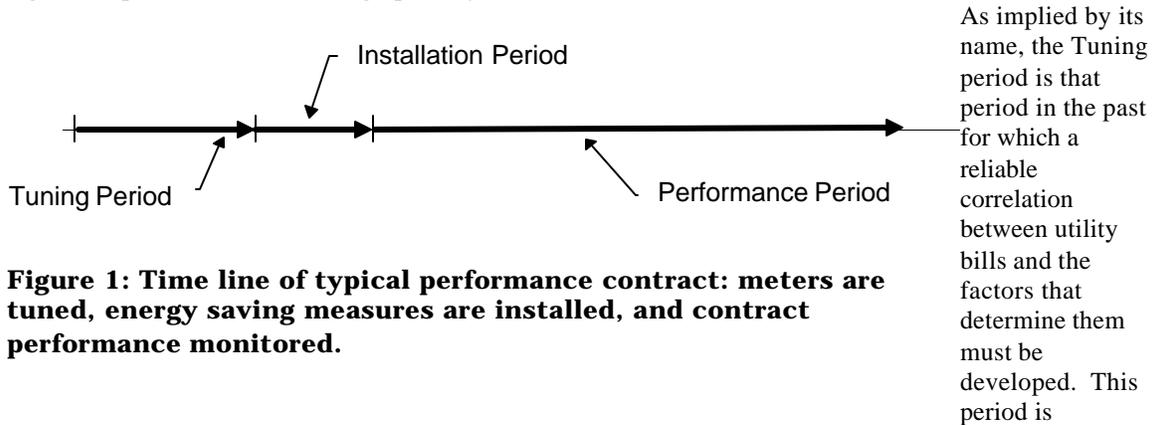
This paper addresses each question and presents quantitative methods to answer them, including a general *Baseline Equation* that is needed to discuss the quantitative aspects.

### **Typical Time Line of a Performance Contract**

For purposes of a Utility Performance Contract it is useful to distinguish three separate time periods:

- **Tuning Period** which is used to establish a reliable correlation of past utility bills;
- **Installation Period** during which the contractor implements the contracted measures;
- **Performance Period** during which utility savings resulting from the measures are monitored.

Figure 1 represents this time line graphically:



**Figure 1: Time line of typical performance contract: meters are tuned, energy saving measures are installed, and contract performance monitored.**

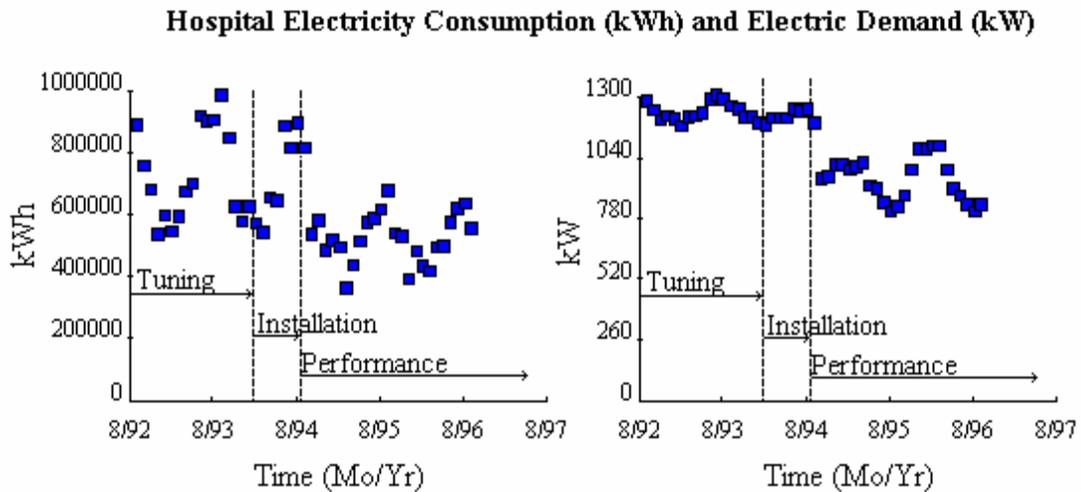
sometimes also referred to as Pre-Retrofit period.

### **How to Select an Appropriate Tuning Period**

How many bills from the Tuning period should be used to develop a Baseline correlation?. The more past utility bills are used, the more reliable the correlation. On the other hand, using utility bills spanning several years may inadvertently capture the effects of past events and trends irrelevant to current usage patterns. For example, past building construction or a change in occupancy can bias a correlation such as to render it useless as a predictor of future energy use.

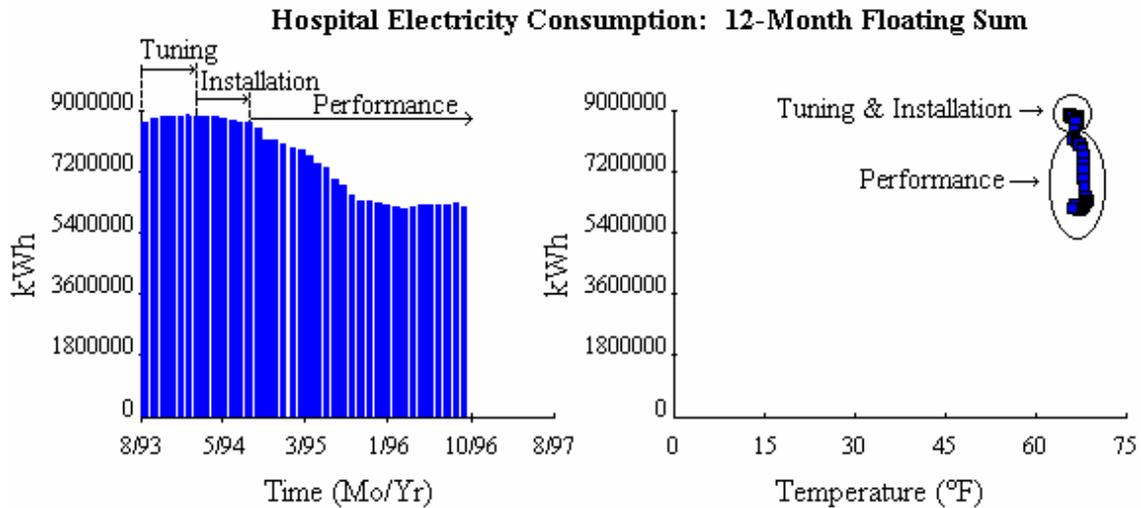
Figure 2 shows both electric consumption and demand from a hospital. Note the seasonal patterns in consumption and the visible downward trend once the energy saving measures have been installed. To determine the optimal number of bills to use from the Tuning period in order to identify a correlation the following recommendations are useful:

- Use one or two year's worth of utility bills, but no less than twelve bills.
- Include only actual, measured, bill consumptions. If a utility bill with an estimated meter reading is encountered it should be combined with one or more subsequent bills such that the result is based only on actual meter readings.
- If several years of past utility data exist, select the most appropriate tuning period by using a 12-month moving sum plot, where each point represents one year's worth of utility consumption starting with months 1 through 12 of the tuning period, then months 2 through 13, and so on. (see Fig. 3).



**Figure 2: Utility bill electric consumption and demand from a hospital before, during, and after installation of energy-saving measures. The three typical periods of a performance contract are superimposed.**

Plotting 12-month moving sums of utility bills vs. time provides the most useful visual check on what tuning period to select. Figure 3 shows electricity consumption of the same hospital plotted as 12-month moving sums. Since each plotted bar contains one year's worth of utility bills, seasonal variations caused by weather or seasonal occupancy variations are minimized.



**Figure 3: Twelve-month floating sum of utility bills. Each bar or point represents one year's worth of utility bills. Bill sums vs. time (left graph) should be flat during Tuning Period, decreasing during Performance Period. Bills sums vs. temperature (right graph) should show no variation in y-axis during Tuning Period, and no significant variation at any time on the x-axis.**

The best tuning period should be chosen so as to contain those 12 or more bars of most similar height. In the example of Fig. 3 the 12-month moving sum increases slightly during the Tuning period, flattening out

between October 1993 and February 1994. Thus, the 16 bills starting November 1992 and ending February 1994 would seem the most appropriate choice to develop a correlation.

The 12-month moving sum decreases slightly during the Installation period, and drops rapidly during the Performance period, as the effects of several energy-saving measures become apparent. The graph to the right is a check on the independent variables chosen in the baseline correlation below. In the right-hand graph we substitute a 12-month moving for the outdoor temperature for the time scale shown in the left-hand graph. The temperature-averaging periods are synchronized with the utility billing periods. As expected, the outdoor 12-month temperature average shows little variation, while the utility bill sums fall in value, consistent with the left graph.

## The Baseline Equation

Much previous work (see for example, Fels 1986, Reddy & Claridge, 1997, Reddy et al., 1996) has been done on the theoretical underpinnings of regression models as well as on the practical issues with analyzing utility data. Much of this research has concentrated on weather-dependent (mainly outdoor temperature) variables. In response to numerous field data from users of a leading Utility Accounting Program, we have developed a General Baseline Equation intended to encompass conventional, weather-sensitive utility bills as well as the more usage- and occupancy- sensitive type. Usage and derivation of the Baseline Equation, its extension to electrical demand, and its usage to quantify utility saving measures, is documented more extensively in Metrix, 1997.

This paper presents the Baseline Equation, shows its equivalence (if limited to weather-sensitive variables) to prior work, and illustrates its usage with a number of case studies drawn from the users of the same Utility Accounting Program. Because of the sensitivity of the data presented, we have omitted any information from which the identity of the facilities could be derived.

Returning to utility bill analysis, once the appropriate choice of utility bills has been made and all independent variables have been identified we are ready to establish a quantitative correlation, if one exists. All independent variables must be synchronized to the utility bill reading dates. If daily data are available, as for outdoor temperatures for example, then synchronization consists simply in summing or averaging over the days contained in each utility billing period.

If periodic data are available on a coarser time scale different from the bill reading dates, it must be apportioned over the utility bill periods. For example, monthly figures for meals served in a restaurant or widgets produced in a factory must be assigned to contemporaneous utility billing periods in proportion to the number of days shared by billing periods and variable readings. This synchronization procedure is described in more detail in a later section.

Utility bill consumptions may be correlated to the synchronized values of one, some, or all of the following types of variables:

- heating degree-days calculated for an appropriate heating balance point;<sup>2</sup>

---

<sup>2</sup> In a physical sense, the balance point temperature is the outdoor temperature above which weather no longer influences heating. Conceptually the balance point results from the average indoor temperature less the effect of *free heat* from appliances, lights, and occupants:

$$T_{BH} = T_R - \frac{IG}{BLC}$$

where

$T_R$  is the average room temperature, °F (°C);

IG is the average internal heat gain, Btu/hr (W);

- cooling degree-days calculated for an appropriate cooling balance point;
- A number, L, of other independent variables.

The Baseline Equation used to fit utility bills to these variables is:

$$\hat{Q}_i = C_D * (D_i - D_{i-1}) + C_H * HDD_{BH,i} + C_C * CDD_{BC,i} + \sum_{k=1}^L C_k * V_{k,i}$$

and

$$Q_i = \hat{Q}_i + e_i$$

where:

$i$	= index for N utility bills (i=1..N);
$Q_i$	= Actual utility bill consumptions [Utility Units, e.g., kWh];
$\hat{Q}_i$	= Best Fit to utility bill consumptions [Utility Units, e.g., kWh];
$e_i$	= <i>Errors of Fit</i> (also termed <i>Residuals</i> );
$C_D$	= Baseload consumption per unit time (Utility Units/day, e.g., kWh/day);
$D_i - D_{i-1}$	= time interval between dates $D_i$ and $D_{i-1}$ (days elapsed since an arbitrary origin);
$C_H, C_C$	= Coefficients for Heating and Cooling Degree-days (Utility units/deg-day);
$HDD_{BH,i}$	= Heating degree-days for base temperature $t_{BH}$ (°F-day or °C-day);
$CDD_{BC,i}$	= Cooling degree-days for base temperature $t_{BC}$ (°F-day or °C-day);
$C_k$	= Coefficient for other independent Variable k;
$V_{k,i}$	= Values of other independent variable k.

In practice, only the baseload consumption per unit time,  $C_D$ , and one or two of the other coefficients are usually non-zero. Though theoretically possible, it is rare to find a statistically significant correlation with more than three or four independent variables.

### **Variable-Base Degree-Days Synchronized to Utility Bills**

The heating and cooling degree-days in the Baseline Equation are calculated over an arbitrary period of days from daily temperature data: For any given period from dates  $D_{i-1}$  to  $D_i$ , heating degree-days are calculated as the sum of positive differences of a balance point and the average daily dry bulb temperatures:

$$HDD_{BH,i} = \sum_{d=D_{i-1}+1}^{D_i} (T_{BH} - T_{O,d})^+$$

where

- $HDD_{BH,i}$  are heating degree-days the days from  $T_{i-1}$  to  $T_i$ , °F-day (°C-day);
- $D_i$  is the last day of the i-th billing period;
- $D_{i-1}$  is the last day of the (i-1)-th billing period, and  $D_{i-1}+1$  is the 1st day of the i-th period.
- $T_{BH}$  is the heating balance point temperature, °F (°C);
- $T_{O,j}$  is the average outdoor temperature for the day  $D_i$ , °F (°C);
- $(X)^+$  indicates that only positive differences are to be used, and negative differences set to zero.

Typical heating balance point temperatures vary between 55 and 65°F (approximately 13 to 18°C). Since its physical determinants are usually not known, the balance point temperature is usually obtained from the best fit of the regression.

---

BLC is the so-called Building Load Coefficient, Btu/hr-°F (W/°C).

Cooling degree-days over the same period are similarly calculated as:

$$CDD_{BC,i} = \sum_{D=D_{i-1}+1}^{D_i} (T_{O,d} - T_{BC})^+$$

where

$CDD_{BC,i}$  are cooling degree-days the days from  $D_{i-1}$  to  $D_i$ , °F-day (°C-day);  
 $T_{BC}$  is the cooling balance point temperature, °F (°C);  
all other symbols are the same as in the previous equation.

Typical cooling balance point temperatures vary in a similar range as the heating balance point temperature and the physical interpretation is the same. Again, the actual value of the cooling balance point temperature is usually obtained from the best fit of the regression.

### **Other Independent Variables**

If utility bills correlate poorly or incompletely with heating and cooling degree-days, then independent variables such as occupancy hours, meals served, or widgets produced, should be accumulated over the same periods of time as the utility bill intervals. If available on a daily level, these variables can be summed over the utility periods like daily degree-days.

More commonly such non-weather related variables are available on a periodic basis with reading dates different from the utility bill reading dates. In such cases the available readings must be time-shifted to result in an equivalent series of readings whose reading dates coincide with those of the utility bills. Lacking any detailed information, time-shifting can be simply done by proportional weighting of two consecutive variable values by the fraction of the time their period overlaps with that of the utility bill.

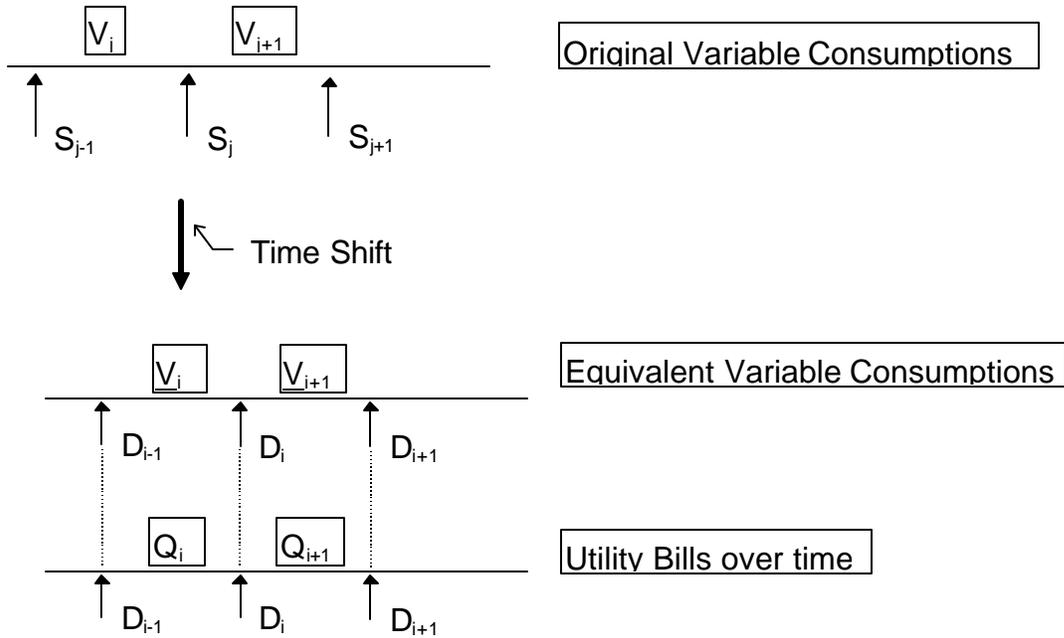
Assume, for example, that the time history of a meter, characterized by N utility bill consumptions  $Q_i$  ( $i=1..N$ ) read on N dates  $T_i$ , is to be correlated with the time history of an independent variable, characterized by M variable values  $V_j$  ( $j=1..M$ ) read on dates  $S_j$ . We want to find the equivalent variable values,  $\underline{V}_i$  after synchronizing to the utility reading dates  $T_i$ . Graphically this can be represented as shown in Fig. 4.

In the example shown in Fig. 4, varying fractions of any two original variable consumptions are combined to yield one equivalent, time-shifted consumption. A more general equation can be derived to cover cases where the original variable reading dates are more or less frequent than the utility bill reading dates.

$$\underline{V}_i = \sum_{j=1}^M V_j * \frac{[\text{Min}(D_i, S_j) - \text{Max}(S_{j-1}, D_{i-1})]^+}{S_j - S_{j-1}}$$

where

$D_i, S_j$ , represent numeric dates starting at an arbitrary origin (e.g. 1=1/1/1900);  
 $V_j$  is the value of variable V for its original time interval j (from  $S_{j-1}$  to  $S_j$ );  
 $\text{Min}(D,S)$  indicates the earlier of two dates represented by D,S;  
 $\text{Max}(D,S)$  indicates the latter of two dates D, S;  
M is the number of original variable consumptions;  
 $[X]^+$  indicates that only positive differences are to be used, negative differences ignored.



**Figure 4: Time-shift of a variable to fit utility bill date intervals. In this example the time-shifted consumption  $\underline{V}_i$  consists of about 60% of  $V_j$  and 30% of  $V_{j+1}$  (percentages need not add up to 100%).**

## Fitting Baseline Equation by Regression

To obtain the equation coefficients through regression we first normalize the equation by dividing all terms by the length (in days) of the time interval,  $D_i - D_{i-1}$ . This is equivalent to expressing all utility bills, degree-days, and other independent variables, by their daily averages. The Baseline Equation in the section of the same name can now be rewritten as:

$$\hat{q}_i = C_D + C_H * hd_{BH,i} + C_C * cd_{BC,i} + \sum_{k=1}^L C_k * v_{k,i}$$

where

$hd_{BH,i}$  are Heating degree-days per day for base temperature  $T_{BH}$  (°F or °C);

$cd_{BC,i}$  are Cooling degree-days per day for base temperature  $T_{BC}$  (°F or °C);

$v_{k,i}$  are daily average values of other independent variables, if any (variable units per day).

When the coefficient values  $C_D$ ,  $C_H$ ,  $C_C$ ,  $T_{BH}$ ,  $T_{BC}$ ,  $C_k$  are obtained through proper tuning, as discussed in more detail later, the Baseline Equation with those values represents the best fit to the meter data. The time history of the *Error* term,  $\epsilon_i$ , indicates the deviation of the actual utility bills per day from the fit. Generally speaking, the smaller the error terms,  $\epsilon_i$ , the better the fit.

Mathematically speaking, to minimize the error term is to minimize the sum of the squares of the deviation time series:

$$\sum_{i=1}^N e_i^2 = \text{Minimum}$$

Solving this basic equation yields expressions for the coefficient estimates,  $C_D, C_H, C_C, C_k$  their standard errors,  $S_D, S_H, S_C, S_k$  and the *Coefficient of Determination*,  $R^2$ , of the regression. The Coefficient of Determination expresses the goodness of fit. A value  $R^2=1$  indicate a perfect correlation between actual data and the regression equation; a value  $R^2=0$  indicates no correlation. For purposes of tuning for a performance contract, as a rule of thumb *the value of  $R^2$  should never be less than 0.75*. The Coefficient of Determination is calculated as:

$$R^2 = \frac{\sum_{i=1}^N (\hat{q}_i - \bar{q})^2}{\sum_{i=1}^N (q_i - \bar{q})^2}$$

where:

$\hat{q}_i$  = Fitted utility bill consumptions (utility units per day);

$q_i$  = Actual utility bill consumptions (utility units per day);

$\bar{q}$  = Average of daily utility bill consumptions (utility units per day).<sup>3</sup>

When more than one independent variable is included in the regression the value of  $R^2$  is no longer sufficient to determine the goodness of fit. The standard error of the estimate of the coefficients becomes the more important determinant. The smaller the standard error compared to the coefficient's magnitude, the more reliable the coefficient estimate. To facilitate the significance of individual coefficients the so-called *T-Statistics* or simply *T-values* are used, which are simply the ratio of the coefficient estimate divided by the standard error of the estimate:

$$t = \frac{\text{Coeff}}{S_{\text{Coeff}}}$$

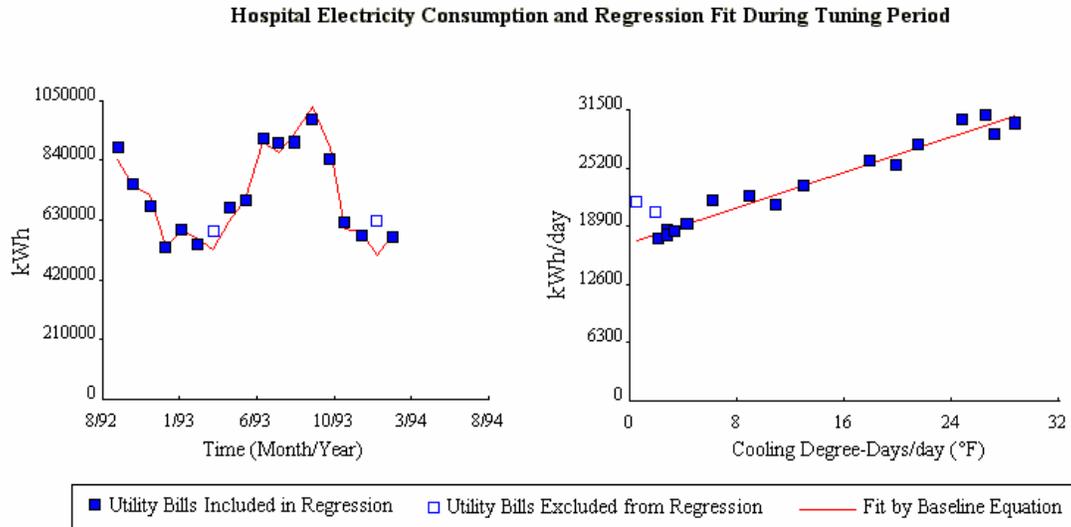
The coefficient of each variable included in the regression has a *t-statistic*. For a coefficient to be statistically meaningful, *the absolute value of its t-statistic must be at least 2.0*. Another way of stating this is that under no circumstances should a variable be included in a regression if the standard error of its coefficient estimate is greater than half the coefficient magnitude. The latter is true *even when including that variable increases the  $R^2$* . Generally speaking, including more variables in a regression results in a higher  $R^2$  but the significance of most individual coefficients will likely decrease.

---

<sup>3</sup> The general equations for the coefficients and their standard errors are more complicated and are beyond the scope of this paper. Their derivation may be found in any introductory textbook on Statistics or Data Analysis. See, for example, the chapter on "Multiple and Curvilinear Regression" in Bowen and Starr, 1982; or the chapter on "Multiple and Partial Correlation" in Croxton et al, 1967.

## Results for Weather-Related Models

When performing a regression of the Baseline Equation against the Hospital Utility Bills in Fig. 2 we find only cooling degree-days to be a significant variable, with the best fit at a base temperature of 54°F (12.2°C).



**Figure 5: Tuning Period for hospital electricity bills shown in Figure 2. Note utility bills excluded from the regression on account of degree-day threshold. Left graph shows bills vs. time, right graph shows bills vs. cooling degree-days.**

The Hospital Electricity Baseline Equation obtained from the bills in Fig. 5 is:

$$\hat{Q}_i = 17.105 * (D_i - D_{i-1}) + 481.66 * CDD_{54,i}$$

The full set of Baseline Equation Coefficients and their T-statistics and  $R^2$  of the fit is shown in Table 1 below.

**Table 1: Baseline Equation Coefficients for Hospital Electricity (N=16)**

Coeff	Value	t-stat	Comments
$C_D$	17,105 kWh/day		$R^2=0.965$
$C_H$	0		
$B_H$	-		
$C_C$	481.66 kWh/°F-day <i>866.99 kWh/°C-day</i>	19.52	2.0°F Threshold <i>(1.1°C)</i>
$B_C$	54.0°F <i>12.2°C</i>		Clg D-day Base Temp
$C_{1...3}$	0		

## Tuning the Baseline Equation

When the Baseline Equation includes only non-weather related variables, tuning the Baseline Equation is identical to performing an ordinary least-square regression as discussed earlier. There is nothing arbitrary about the process. Confronted with the same data, two analysts should always obtain the exact same results.

Unfortunately, the same categoric statement can not be made in the more prevalent case where one or more variables are weather dependent. True, the same type of regression is performed, but the (weather dependent) degree-day variables are calculated for a fixed base temperature. Each value of the base temperature will result in a slightly different degree-day series (one series member for each billing period), and slightly different coefficients, coefficient of determination  $R^2$ , and t-statistics. There is usually a base temperature that will produce a maximum value for  $R^2$ , but in our experience this maximum is quite flat and not always unambiguous. In other words, relatively large ranges of base temperatures will produce substantially the same goodness of fit. Occasionally the best fit will be achieved only with manifestly improbable base temperatures (i.e. significantly higher than indoor average room temperature).

From experience with hundreds of Baseline Equations we find that the uncertainty in base temperature for either heating or cooling degree-days is at least  $\pm 1^\circ\text{F}$  ( $\pm 0.5^\circ\text{C}$ ), often more.

### ***The Degree-Day Threshold***

An important assumption behind ordinary least square regressions is that all independent variables are normally distributed. Unfortunately, heating and cooling degree-days have an over-representation of zero or near-zero values during mild and off seasons (e.g., no heating degree-days in summer). In other words, a typical year will include a significant number of electricity bills whose weather-related portion is zero or very small.

To minimize the resulting statistical error it is convenient to exclude from the regression all utility bills corresponding to periods with degree-days below a *Degree-Day Threshold*. The Degree-Day Threshold can be expressed as average degree-days per day, corresponding to a temperature difference. A typical value is around  $2^\circ\text{F}$  ( $1.1^\circ\text{C}$ ) which signifies that any 30-day utility bill where the corresponding degree-days are below  $60^\circ\text{F-days}$  ( $33.33^\circ\text{C-days}$ ) is excluded from the regression. This was the value used for the hospital in Table 1. The value of the Degree-Day Threshold can be varied such as to exclude only bills with small, yet most uncertain cooling components.

Such exclusion is usually necessary only if cooling or heating degree-days are the only variable in the regression. For multi-variable regressions where at least one variable is not weather-related no bills need to be excluded.

### ***Excluding Bills from Regression***

There are good reasons why individual utility bills may be unsuitable to develop a baseline and should be excluded from the regression. For example, a bill may be untypically high because of a one-time equipment malfunction that was subsequently repaired. However, it is often tempting to look for reasons to exclude bills that fall far from “the line” and not question those that are close to it. For example, bills for periods containing vacations or production shutdowns may look anomalously low, but excluding them from the regression would result in a chronic over-estimate of future baseline during the same period. We will see in a later section how to handle such bills through the use of additional independent variables.

Another typical case is the estimated bill, which will be higher or lower than actual utility consumption. If the subsequent bill reflects an actual (not estimated) meter reading, the billed consumption will be lower or higher than actual consumption, respectively. While either bill will be off, together they correctly reflect the cumulative consumption over two periods. Estimated/actual bill pairs show up on time series graphs as tell-tale “double spikes” (see, for example, the two bills after 5/96 in Fig. 7). Estimated bills should be combined with the subsequent actual bill for purposes of fitting a Baseline Equation.

## **Baseline Equation vs. Mean Monthly Temperature Models**

Regression models have been used for some time to correlate utility bill consumptions with outdoor temperature (Fels 1986, Ruch and Claridge 1991, Rabl and Rihle 1992, and others). They can be divided into two broad classes: one is based on variable-base degree-days (VBDD) where the effect of outdoor temperature is embodied in variable degree-days. The other is based on monthly mean temperature

(MMT). Both types of models have similar types of non-linear parameters: base temperatures for VBDD, change points for MMT. Software programs exist that embody both approaches. Metrix (Metrix, 1997) and PRISM (Fels, Kissock, et al. 1995) fall in the VBDD category. Emodel (Kissock, Wu, et al. 1994) is based on the MMT model.

To better appreciate the basic distinction between VBDD & MMT models, consider the three-parameter MMT model appropriate for modeling monthly mean daily heating energy consumption data:

$$\hat{E} = \mathbf{b}_1 + \mathbf{b}_2 * (\bar{T}_O - \mathbf{b}_3)^-$$

where:

- $\hat{E}$  is the monthly mean daily energy use including, but not exclusive to, heating and cooling;
- $\beta_1$  is the constant daily portion of energy use;
- $\beta_2$  is the slope parameter, indicating marginal daily energy use per degree of outdoor temperature for heating;
- $\bar{T}_O$  is the monthly mean outdoor dry bulb temperature;
- $\beta_3$  is the change point, indicating outdoor temperatures where heating energy dependence on outdoor temperature begins;
- (..) indicates that only the absolute value of negative quantities in parenthesis should be considered, and should be set to zero otherwise.

The regressor variable for the MMT model is  $(\bar{T}_O - \beta_3)^-$  while the equivalent variable for the VBDD model is

$$\frac{1}{N} \sum_{i=1}^N (T_B - T_{O,i})^+ .$$

Both these quantities will be strictly equal only when  $T_B$  is higher than the

maximum outdoor temperature achievable at that location. In real cases, this would not be so and  $\beta_3$  can no longer be interpreted as the balance point temperature, but merely as a regression parameter. Reddy (1997) presented a study when both model types were applied to gas and electricity use in several Army Bases nationwide.

Only if a MMT model were expressed strictly for daily data could one show direct equivalence. To see the equivalence to the Baseline Equation, we write an MMT model showing temperature dependence for both heating and cooling energy.

$$\hat{E} = \mathbf{b}_1 + \mathbf{b}_2 * (T_O - \mathbf{b}_4)^- + \mathbf{b}_3 * (T_O - \mathbf{b}_5)^+$$

Next, we add an index  $d$  (denoting the *day*) to all daily variable quantities ( $\hat{E}_d, T_{O,d}$ ) and sum both sides over all days in a utility billing period:

$$\begin{aligned} \hat{Q} &= \sum_{d=D_{i-1}+1}^{D_i} \hat{E}_d = \sum_{d=D_{i-1}+1}^{D_i} [\mathbf{b}_1 + \mathbf{b}_2 * (T_{O,d} - \mathbf{b}_4)^- + \mathbf{b}_3 * (T_{O,d} - \mathbf{b}_5)^+] = \\ &= \sum_{d=D_{i-1}+1}^{D_i} \mathbf{b}_1 + \mathbf{b}_2 * \sum_{d=D_{i-1}+1}^{D_i} (\mathbf{b}_4 - T_{O,d})^+ + \mathbf{b}_3 * \sum_{d=D_{i-1}+1}^{D_i} (T_{O,d} - \mathbf{b}_5)^+ \end{aligned}$$

Note the identity  $(X)^- = (-X)^+$ . Now recalling the earlier definition of heating and cooling degree-days, we see that:

$$\hat{Q} = \mathbf{b}_1 * (D_i - D_{i-1}) + \mathbf{b}_2 * HDD_{BH,i} + \mathbf{b}_3 * CDD_{BH,i}$$

This simple algebraic manipulation shows the equivalence between the Baseline Equation limited to temperature sensitive variables only, and the five parameter MMT model limited to daily data:

$$\mathbf{b}_1 = C_D \quad \mathbf{b}_2 = C_H \quad \mathbf{b}_3 = C_D \quad \mathbf{b}_4 = T_{BH} \quad \mathbf{b}_5 = T_{BC}$$

The most important contribution of the Baseline Equation is in the area of non-temperature sensitive variables, though others have shown their use before (Reddy, Katipamula, Kissock, & Claridge 1995).

## Examples of Multivariate Baseline Equations

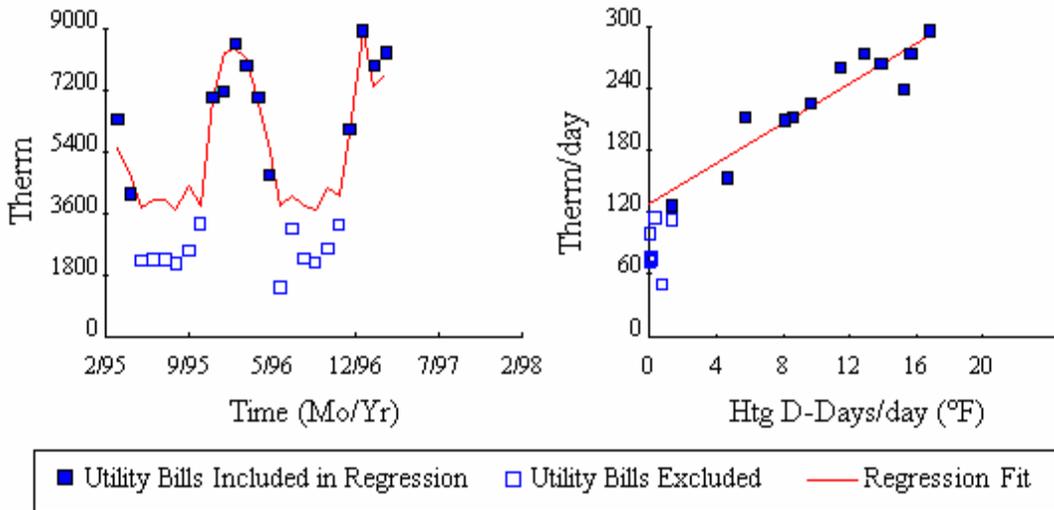
A typical example where more than one variable is needed to correlate utility bills is the gas utility consumption for a Retirement Center on the Southeastern Seaboard. A gas-fired boiler provides space heating. Separate water heating is also by natural gas. Fig. 6 shows the original bills and a regression using heating degree-days only. Note the fairly good correlation of winter bills to heating degree-days with a 61°F (16.1°C) base. The summer bills have been excluded from the regression by using the default 2°F (1.1°C) heating degree-day threshold. The results are summarized in Table 2.

**Table 2: Baseline Equation Coefficients for Retirement Center Gas (N=12)**

Coeff	Value	t-stat	Comments
$C_D$	128.08 Therm/day		$R^2=0.865$
$C_H$	9.71 Thm/°F-day <i>17.48 Thm/°C-day</i>	8.01	2.0°F Threshold <i>(1.1°C)</i>
$B_H$	61.0°F <i>16.1°C</i>		Htg D-day Base Temp
$C_C$	0		
$B_C$	-		
$C_{1..3}$	0		

No Boiler Shutoff modeled

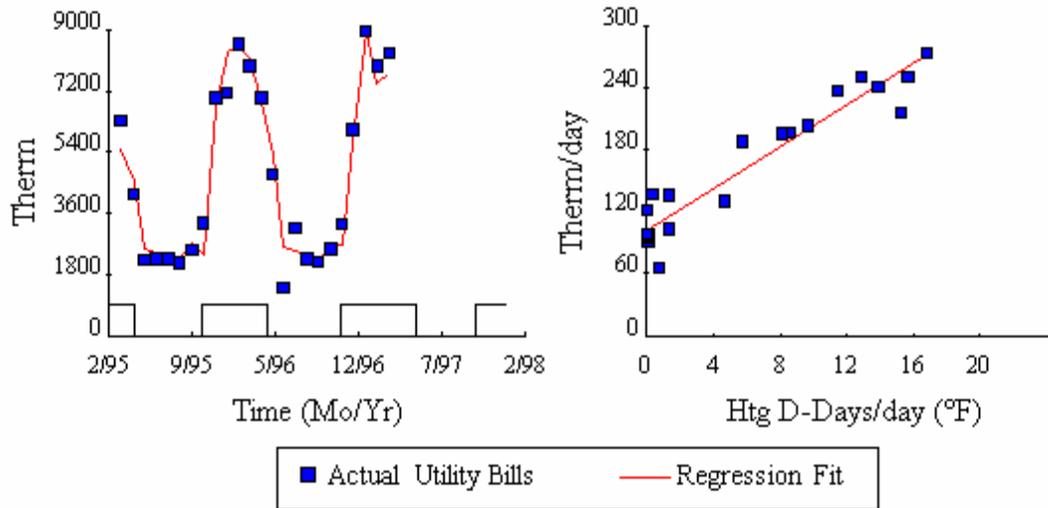
**Natural Gas vs. Time and vs. Heating Degree-Days Base 61°F (16°C)**



**Figure 6: Gas utility bills from retirement center and Baseline Equation lacking summer boiler shutoff term.**

The Baseline Equation overpredicts in summer because the boiler is shut off from approximately May through September. To model this shut off we use a simple on-off variable indicating whether the boiler is on or not. Its values and the improved regression results are shown in Fig. 7 and Table 3.

### Natural Gas vs. Time, Heating Degree-Days Base 61°F (16°C) and Boiler On-Time



**Figure 7: Same as Figure 6, but seasonal boiler shutoff is included as additional variable in regression.**

Table 3 below summarizes the Baseline Equation Coefficients for two variables. Note the increase in significance of the heating degree-day coefficient, as well as the increase in  $R^2$ .

**Table 3: Baseline Equation Coefficients for Retirement Center Gas (N=24)**

Coeff	Value	t-stat	Comments
$C_D$	78.27 Therm/day		$R^2=0.950$
$C_H$	10.13 Thm/°F-day 18.23 Thm/°C-day	8.32	2.0°F Threshold (1.1°C)
$B_H$	61.0°F    16.1°C		Htg D-day Base Temp
$C_C$	0		
$B_C$	-		
$C_1$	45.99 Thm/day	2.96	Boiler on-days
$C_{2,3}$	0		

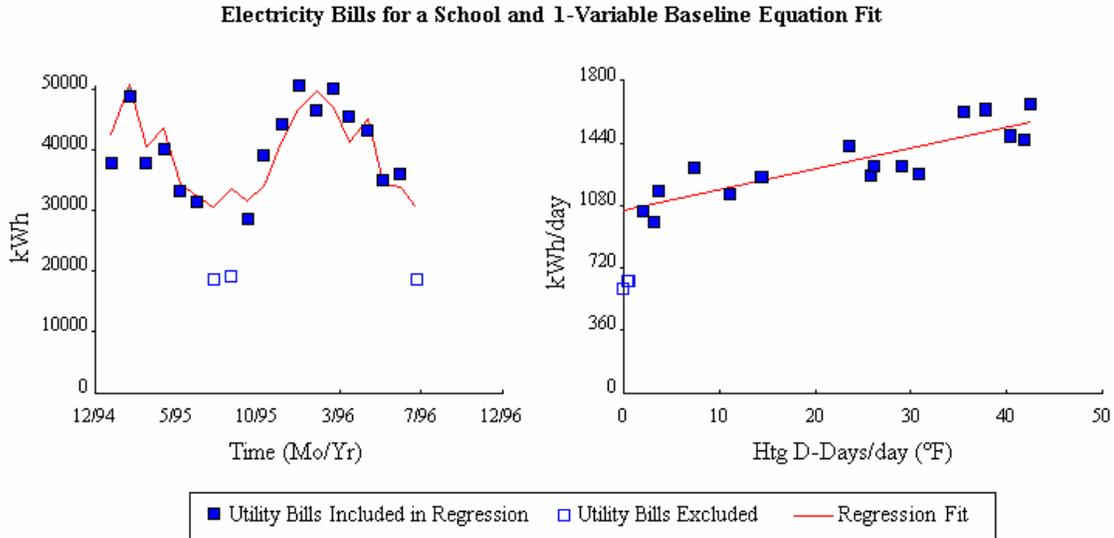
**Includes Boiler Shutoff Term**

### Occupancy Variables in Schools

Schools are notorious in the influence of occupancy on energy use. Electricity bills of a typical, electrically heated High School in a Northeastern State are shown in Fig. 8. Also shown is the fit to heating degree-days alone with a customary 2°F (1.1°C) degree-day threshold. The bills thus excluded are roughly coincident with the summer vacation periods. The Baseline Equation Coefficients are shown in Table 4. Note the mediocre fit, even with the summer bills excluded.

**Table 4: Baseline Equation Coefficients for School Electricity (N=16)**

Coeff	Value	t-stat	Comments
$C_D$	1,050.27 kWh/day		$R^2=0.735$
$C_H$	11.98 kWh/°F-day 21.56 kWh/°C-day	6.23	2°F (1.1°C) Threshold
$B_H$	65.0°F 18.3°C		Htg D-day Base Temp
$C_C$	0		
$B_C$	-		
$C_1$	0		
$C_{2...3}$	0		



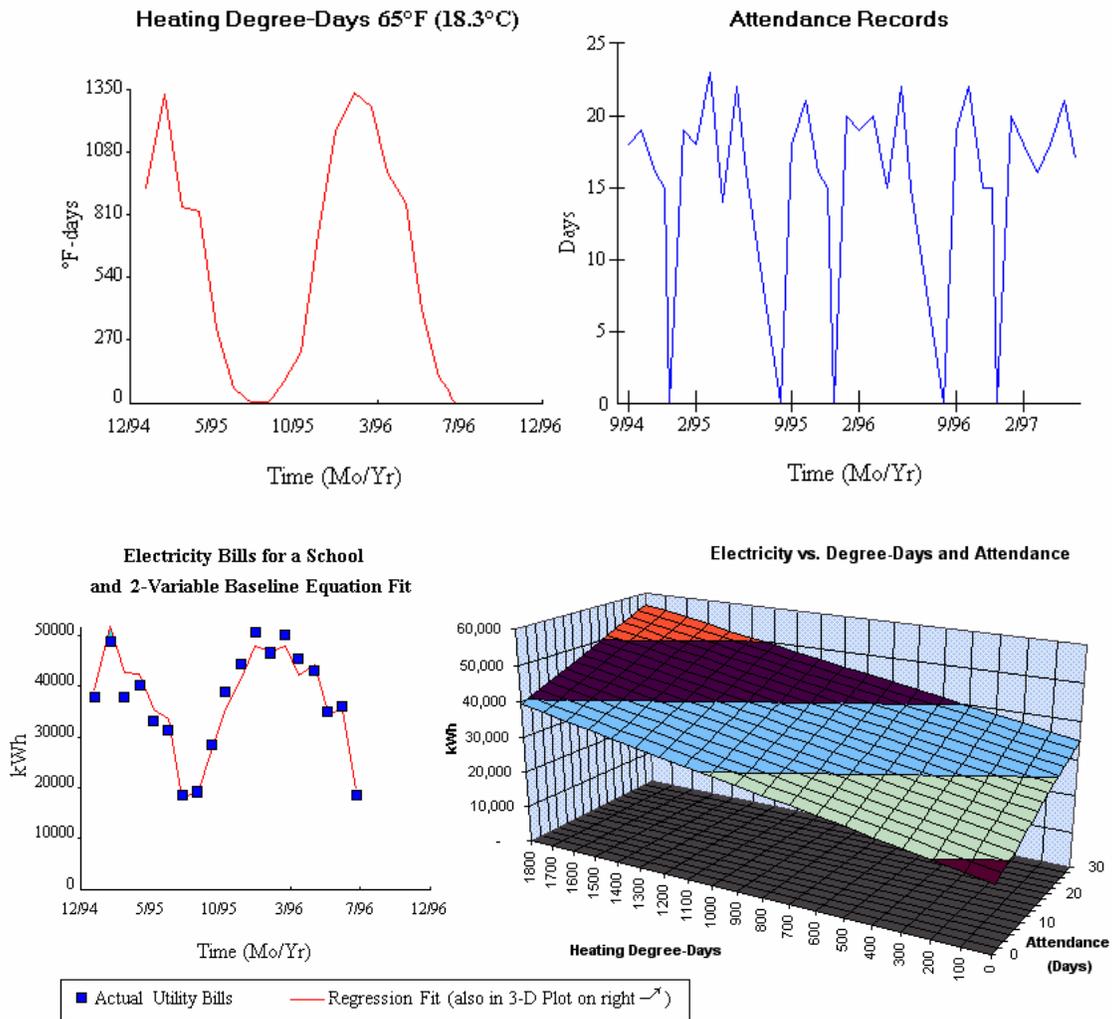
**Figure 8: Electricity consumption by a school as a function of heating degree-days only. The same bills are shown in both graphs. Bills excluded because of degree-day threshold coincide with vacation periods and fall far off fit.**

Now consider the effect of including one additional independent variable, the school’s attendance records. Both independent variables are shown side-by-side in Fig. 9a. The electric utility bills are shown again in Fig. 9b together with the 2-variable, improved fit. Also shown in the same figure is the planar fit in a 3-D graph of electricity vs. degree-days and attendance. (The numerical results are in Table 5). Note that electricity bills vary about as much over the range of degree-days as they do over the range of attendance-days. Neglecting either variable would have missed a great deal.

**Table 5: Baseline Equation Coefficients for School Electricity (N=19)**

Coeff	Value	t-stat	Comments
$C_D$	629.26 kWh/day		$R^2=0.941$
$C_H$	12.99 kWh/°F-day 23.38 kWh/°C-day	9.47	0°F (0°C) Threshold
$B_H$	65.0°F (18.3°C)		Htg D-day Base Temp
$C_C$	0		
$B_C$	-		
$C_1$	5.82 kWh/Day	7.42	Attendance, in Days
$C_{2...3}$	0		

The Baseline Equation Coefficients are shown in Table 5. Note the much better  $R^2=0.941$  and the large significance of both heating degree-days and school attendance coefficients.



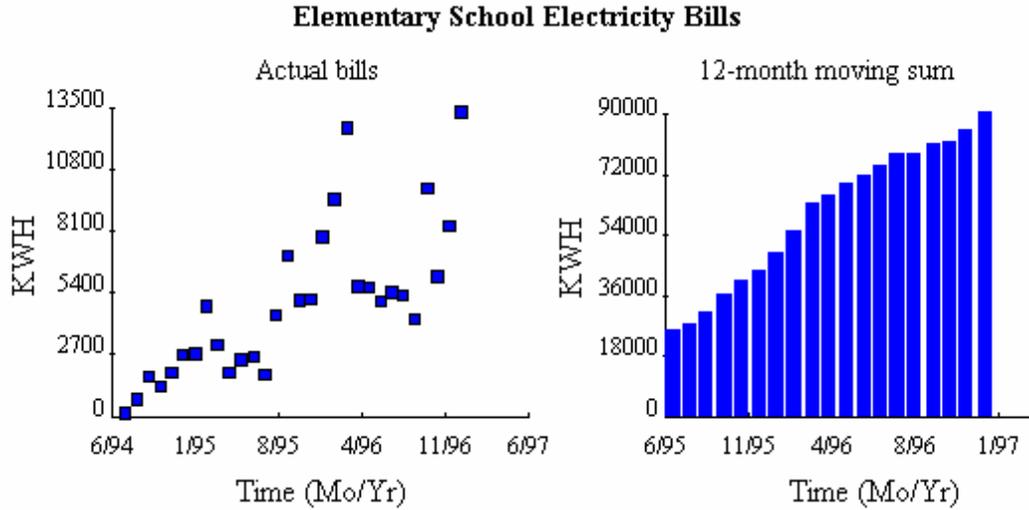
**Figure 9b: Electricity consumption by a school as a function of both heating degree-days and attendance days. Left graph shows actual electricity bills and two-variable Baseline Equation fit. Right graph shows Baseline Equation fit.**

## Plug Creep and Special Cases

Sometimes the 12-month floating sum of utility bills during the Tuning period rises appreciably with time. This phenomenon is sometimes referred to as *Plug Creep* and can be caused by the gradual increase of installed appliance load in an office, or by gradual changes in usage patterns in a hospital after a change in management or ownership. Plug creep is evident to a minor degree for the hospital in Figure 3 if the whole Tuning period is considered. Extrapolated over one year, the effect amounts to a 3% plug creep per year.

Sometimes plug creep is so significant that the only hope for establishing a meaningful baseline is to find a reliable predictor of the increase. For example, the Elementary School shown in Fig. 10 shows rapidly increasing electricity consumption in addition to the normal seasonal variations. With the seasonal

fluctuations eliminated in the graph on the right, a long-term consumption increase averaging 40,000 kWh/year is evident. How can a reliable baseline be established in this case?



**Fig. 10: Annual moving sum of utility bills vs. time. Significant *Plug Creep* is present during the Tuning Period.**

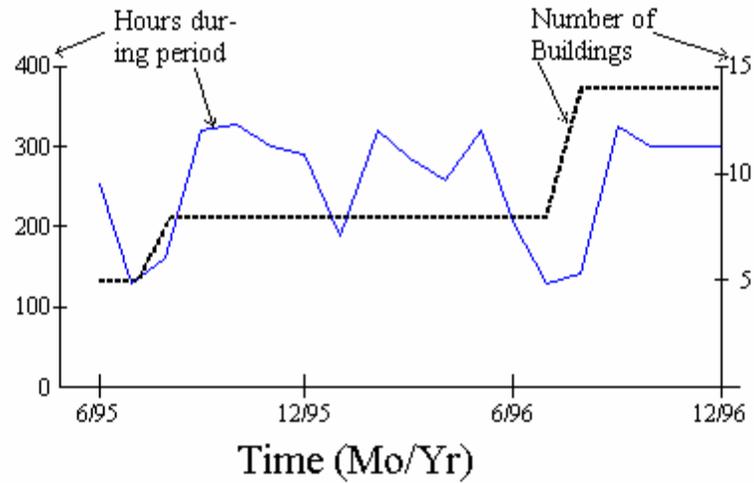
In cases of severe plug creep it is imperative to quantify the cause of the long-term increase. Usually a quick glance at a 12-month floating sum/average graph as shown earlier in Fig. 3 is warranted to ensure that there is no long-term warming (or cooling) trend that could account for the increase in energy use.

Assuming that there are no significant long-term weather trends, one or more candidate variables must be identified that could account for the long-term increase, such as the number of desktop PC's in an office, or the number of patient-days in a hospital. Failure to identify at least one such variable will doom any attempts at calculating a reliable baseline for a facility with significant plug creep, and fatally compromise the accuracy of any estimated savings.

**Table 6: Baseline Equation Coefficients for Elementary School Electricity (N=19)**

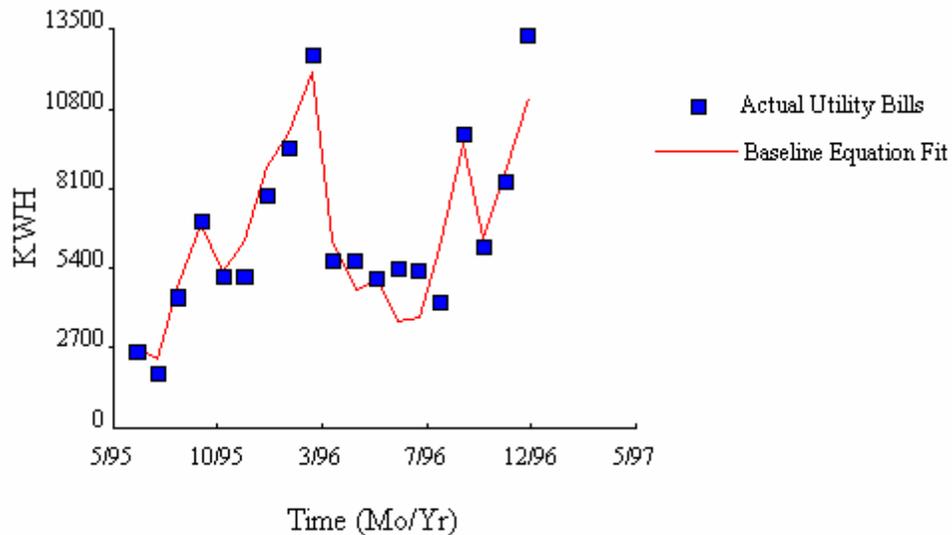
Coeff	Value	t-stat	Comments
$C_D$	-331.11 kWh/day		$R^2=0.862$
$C_H$	15.87 kWh/°F-day 28.57 kWh/°C-day	6.03	0°F (0°C) Threshold
$B_H$	65.0°F (18.3°C)		Htg D-day Base Temp
$C_C$	12.53 kWh/°F-day 22.55 kWh/°C-day	3.51	0°F (0°C) Threshold
$B_C$	64.0°F (17.8°C)		Clg D-day Base Temp
$C_1$	451.76 kWh/Bldg	4.60	Prefabricated Buildings
$C_2$	25.18 kWh/hour	3.68	Occupancy Hours
$C_3$	0		

### Hours of Occupancy and Number of Prefabricated Buildings



**Figure 11: Number of prefabricated buildings and occupancy hours for elementary school in Figure 4. Note the upward trend in both variables.**

### Elementary School Electricity Bills with 4-Variable Baseline Equation Fit



**Figure 12: Same elementary school electricity bills as in Figure 10, with superimposed four-variable Baseline Equation fit.**

For the Elementary School shown earlier in Fig. 4 two such variables accounting for long-term trends were identified: the number of prefabricated buildings used as classrooms in the expanding school, and the overall number of occupancy hours for each month, carefully accounting for vacations, holidays, and

weekends. Fig. 11 shows time histories for both variables. Table 6 shows the numerical results of the four-variable regression, and Fig. 12 shows the corresponding fit superimposed on the original utility bills.

This Elementary School is fairly unusual in that it takes fully four independent variables and seven parameters to achieve an acceptable goodness of fit. Eliminating any one variable from the Baseline Equation decreases  $R^2$  by at least 0.15 or more and decreases all t-statistics of the remaining coefficients.

## Conclusions

Based on inspection of a large number of energy saving projects set up by third parties it is this author's conclusion that a baseline must be established using more than just temperature-sensitive variables like degree-days or average outdoor temperatures. Examples of other independent variables of interest are occupancy hours in schools, meals served in a restaurant, changes in conditioned floor area, or equipment operation.

The general Baseline Equation presented serves well to establish a reliable, robust baseline for a performance contract based on utility bill analysis. It is currently estimated that the Baseline Equation as presented in this paper has been used in practice in over five-thousand performance contracts.

## Acknowledgement

I wish to thank Prof. T. Agami Reddy for invaluable review comments, editorial work, and source material provided for this paper. His contribution far exceeds what is usually expected of a reviewer.

## References

- Bowen, E.K., & Starr, M.K., *Basic Statistics for Business and Economics*, McGraw-Hill, 1982.
- Croxtan, F.E., Cowden, D.J., & Klein, S., *Applied General Statistics*, Prentice-Hall. 1967.
- Fels, M. (ed) 1986. "Special Issue Devoted to Measuring Energy Savings, The Princeton Scorekeeping Method (PRISM)", *Energy & Buildings*, Vol. 9, Nos. 1 and 2.
- Fels, M., Kissock, K., Marean, M., and Reynolds, C., *PRISM (advanced version 1.0) Users' Guide*, Princeton, N.J.: Center for Energy and Environmental Studies, Princeton University, 1995.
- Kissock, J.K., Wu, X., Sparks, R., Claridge, D., Mahoney, J., and Haberl, J., *Emodel, Version 1.4d*, College Station, Texas: Energy Systems Laboratory, Texas Engineering Experiment Station, 1994.
- Metrix Version 2.0 User's Manual*, 1997, SRC Systems Inc., 2855 Telegraph Ave., Suite 410, Berkeley, CA 94705.
- Reddy, T.A., Katipamula, S., Kissock, J.K. and Claridge, D.E., 1995. "The Functional Basis of Steady-State Thermal Energy Use in Air-Side HVAC Equipment", *ASME Journal of Solar Energy Engineering*, vol.117, pp.31-39, February.
- Reddy, T.A., & Claridge, D.E., "Uncertainty of Measured Energy Savings from Statistical Baseline Models" *HVAC & Refrigeration Journal*, Sept 1997
- Reddy, T.A., Saman, N.F., Claridge, D.E., Haberl, J.S., Turner, W.D. and Chalifoux, A. 1997. "Baselining Methodology for Facility-Level Monthly Energy Use – Part I & II", *ASHRAE Trans.*, Vol. 103, paper # BN-97-16-4 & (5).