NorthWrite

Energy Expert: A Technical Basis

A Bin-Based Method for Baseline Performance

We have found the modeling methodology presented below (and used in the Energy Expert) to be useful in establishing performance baselines for detecting anomalies in energy consumption by buildings (Katipamula et al. 2003). It has the advantage that it can capture both linear and non-linear behavior. The method is based on the concept of data bins borrowed from the field of building energy data analysis. A bin is an interval (bin) of values of an independent variable with which a value of another (dependent) variable is associated. For example, the weather at a location can be characterized by the number of hours per year on average that the outdoor-air temperature falls into 5°F bins between some minimum temperature and some maximum temperature, as shown in Figure 1. Similarly, bins can be defined for energy uses that are correlated with outdoor-air temperature (e.g., energy use for cooling a building; see Figure 2).

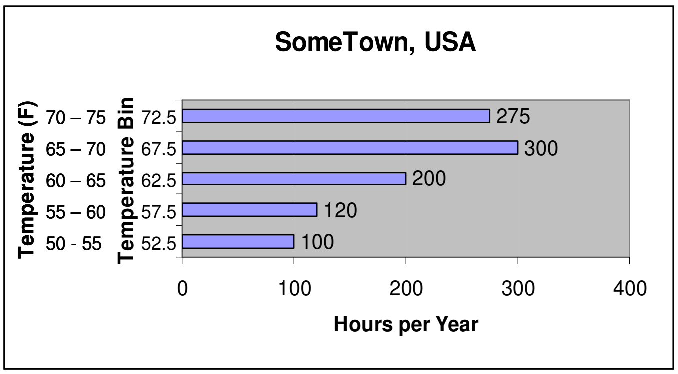
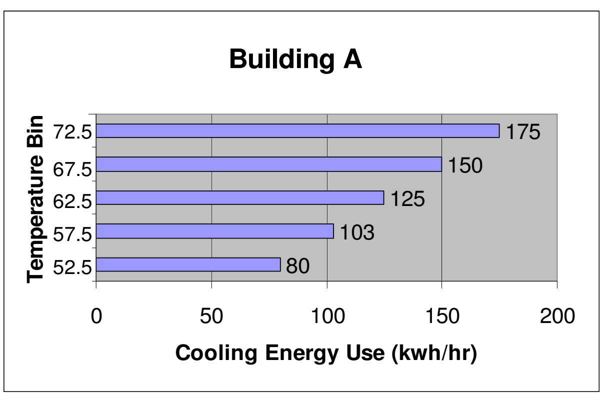


Table of Temperature Bins for SomeTown, USA

Bin (°F)	Hours per Year
72.5	275
67.5	300
62.5	200
57.5	120
52.5	100

Figure 1 – Temperature bins are shown for a fictitious location in the U.S.



Cooling Energy Bins for Building A

Bin (°F)	kWh/hr
72.5	175
67.5	150
62.5	125
57.5	103
52.5	80

Figure 2 - Example of bins for cooling energy use by a building.

When multiple variables are used to explain the variations in energy use multi-dimensional bins can be used, where a multi-dimensional bin is defined as the intersection of one-dimensional bins based on each of the variables. This is shown in Figure 3 for three-dimensional bins that characterize a variable such as energy use in terms of three explanatory variables. A representative value of the dependent variable is assigned to each bin defined by the ranges of values of the independent variables. For an energy use model, the dependent variable is energy consumption. The Energy Expert currently accepts up to 5 independent variables.

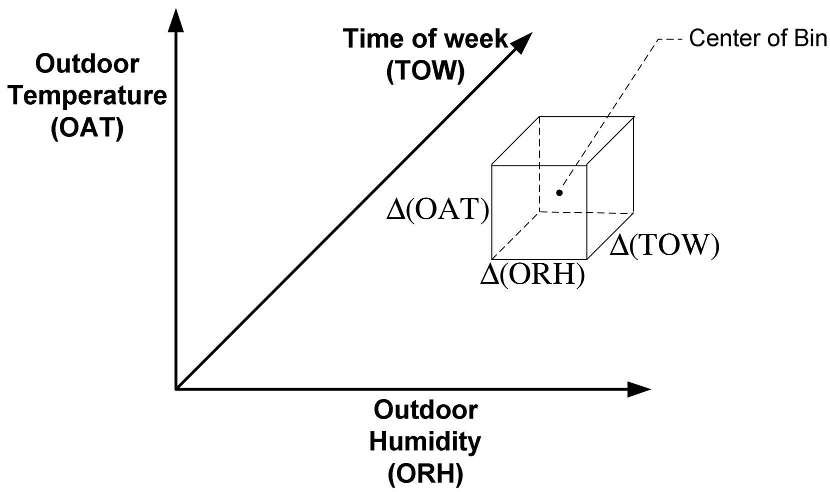


Figure 3. An example three-dimensional binning scheme with bins defined by three explanatory variables: outdoor-air temperature, outdoor-air humidity, and time of week.

The model is "trained" by collecting data empirically and assigning it to bins. Given a sample of empirical data with each set of the sample consisting of a values for a complete set of N independent explanatory variables (x_1, x_2, x_3, …, x_N) and the corresponding measured value of the dependent variable, an N-dimensional model is created by assigning each set of data in the sample to the bin in which the point defined by the values of its independent variables lies. An example bin is shown in
Figure 4. When a sufficient number of points have been assigned to each bin, the model is considered fully "trained." A representative value of the dependent variable is then assigned to each bin, completing the model. The median of the values of the dependent variable in the bin makes a good representative value for both large and small numbers of points per bin.

Once the model is trained, it is used to estimate baseline values of the dependent variable [e.g., whole building electric, refrigeration system load, etc.] given a set of measured values for the independent variables. The bin model represents the baseline behavior of the system or component during the training period. To capture continuous changes to the system over time (and refine comparisons and cost calculations), it is often advantageous to create different models to represent multiple baseline periods.

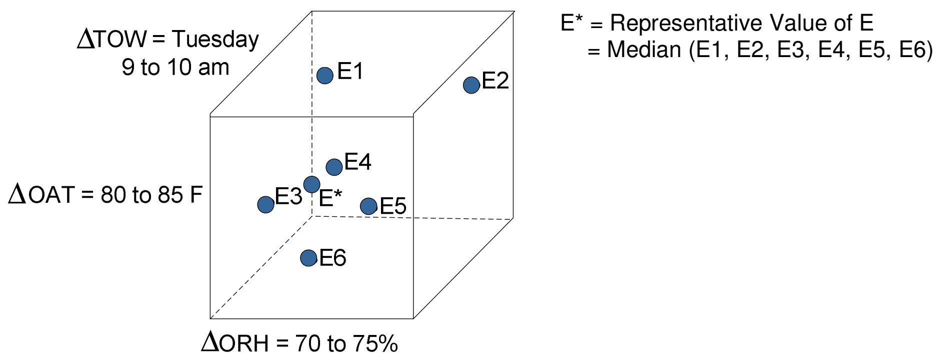


Figure 4 – An example three-dimensional energy bin is shown for outdoor-air temperature (OAT) , time of week (TOW) and outdoor relative humidity (ORH) as the independent variables. Points corresponding to sets of independent variable values and their corresponding energy values, E_i, that fall in the ranges defined by this bin are shown as points inside the bin.

To maximize use of training data and potentially minimize the length of the training period required to obtain adequate data, we utilize the concept of dynamic bins to this approach to modeling. In this approach, the bins are not defined *a priori* with data assigned to them. Instead, bins are defined as needed around a center point defined by the current values of the independent variables (thus the term "dynamic bins"). Only one bin is defined at a time as needed. For example, for the independent variables used in Figure 3 and Figure 4, the point might be defined, for example, as 9:30 am on Tuesday (TOW = 57.5), outdoor-air temperature (OAT) of 82.5°F and outdoor-air relative humidity (ORH) of 72.5%. The coordinates of this bin would then cover the independent variable intervals TOW = 57.5 ± ΔTOW/2, OAT = 82.5 ± ΔOAT/2 and ORH = 72.5% ± ΔORH/2. For ΔTOW=1 hour, ΔOAT=5°F and ΔORH=5%, the bin is defined as shown in Figure 4. All values for the independent energy variable for points in the training data set within the limits of this bin are then assigned to the bin.

An example application of this model to energy use by a chiller is shown in Figure 5 and Figure 6. In Figure 5, plots of *actual* measured energy consumption and corresponding values of the *expected* energy consumption of a chiller are shown for a 3-month period in 2002. Values of expected energy consumption were generated using a bin-based model and corresponding values for the independent variables during this time period. The top plot in Figure 6 shows the same data with expected energy consumption plotted on top of actual energy consumption, clearly revealing the differences. The bottom plot shows an energy consumption index for the same data defined as the ratio of actual to expected energy consumptions. This plot shows that the chiller is consuming more energy than it would have if maintained in its baseline (training-period) state.

Small circles have been added in the bottom plot of Figure 6 to highlight points at which a diagnostic algorithm assigned alarms to these deviations. These alarms would indicate to system operators that the deviation represents sufficient performance degradation to deserve further assessment.

The bin-based model possesses several characteristics that contribute to its strength for use in establishing performance monitoring baselines. It is conceptually simple and as a result

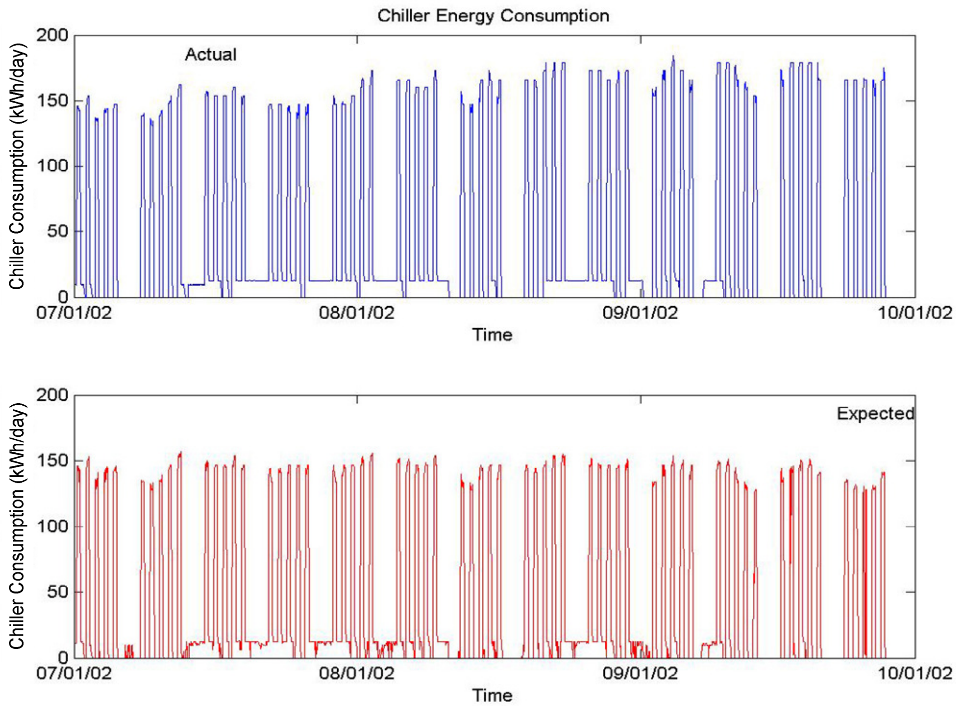


Figure 5 – The actual measured energy consumption (top) and expected energy consumption (bottom) from a bin-based model for a chiller are shown.

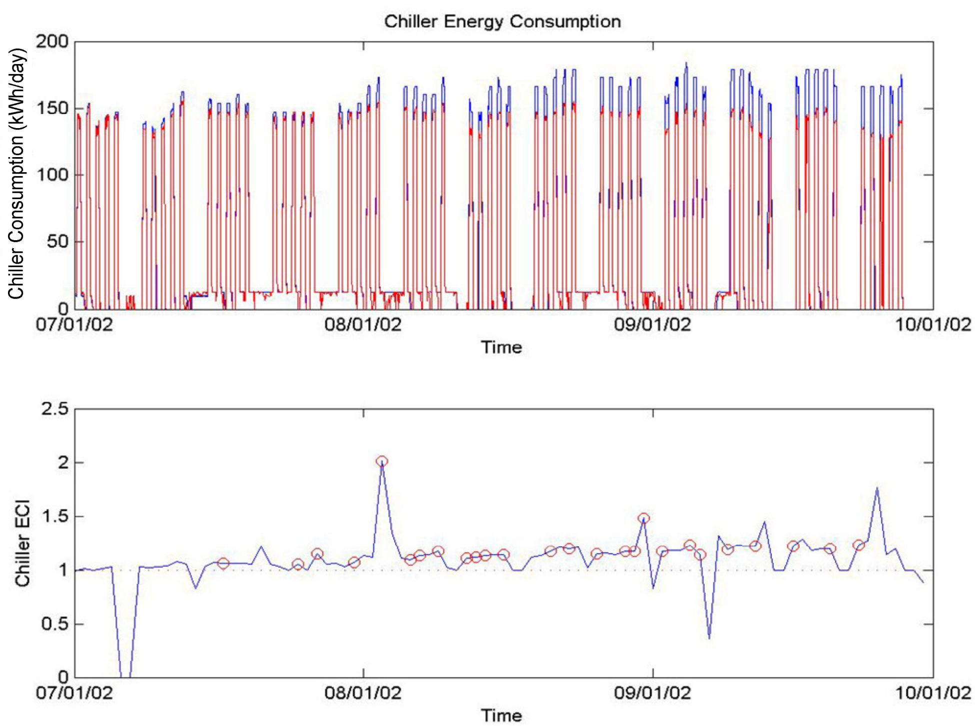


Figure 6 – Actual measured and expected energy consumption (top) and an Energy Consumption Index (bottom) are shown for the Chiller in Figure 5.

potentially appealing to users in the field. Operations staff abhor black boxes that they simply do not understand. A simple model facilitates understanding and provides an initial basis for establishing user trust in the method. Furthermore, this model has proven effective in establishing baselines for other diagnostic problems, i.e., tracking the performance of energy using systems and equipment in buildings. The method is flexible, accommodating whatever independent explanatory variables are appropriate to the system or component, and can be customized to an application's unique characteristics. Bin widths can be adjusted to tune the model to capture features of most importance. Application to building energy tracking has shown that for applications with slowly changing driving conditions (values of independent variables) the model can even be applied usefully while it is still undergoing training. Moreover, the model can capture both linear and non-linear relationships between the dependent and independent variables and transitions from regions of linear behavior to regions of non-linear behavior smoothly.

The model also possesses a few weaknesses that must be noted and assessed during application. When used to establish a performance baseline with which to compare future performance as a means to detect performance degradation, the model will absorb any degradation occurring during the training period. If the degradation is only apparent (i.e., associated with spurious measurements), it will not affect the resulting model, but if the degradation is real, persistent, or part of a trend, it will affect the model, and the model will represent behavior with some degradation present. Therefore, the best training data are measurements made during system/component operation for which performance is known to be good or proper. Accordingly, we recommend that training data be acquired over a time period immediately following verification of system commissioning or re-tuning of system operations, when performance is known (or more likely) to be at its peak.

The bin-based modeling approach is only practical for a small number of variables because the amount of data required for training grows rapidly with the number of independent explanatory variables. We generally use a rough guide of no more than three independent variables and even this depends on the range of each of the variables, the bin dimension for each variable, the frequency of data collection, and the range of operating conditions.

Finally, no physics are captured in the structure of the model. The model has essentially no structure, which makes it flexible, but as a result it has no underlying functional form from which physical relationships are easily derived. Therefore, it has little value to providing underlying knowledge of how and why a system or device behaves the way it does, but this is not the intent of the proposed application, which is merely to establish a baseline for comparison of values of performance indicators in the future to those captured by the baseline.

References

This paper is based on the work of Pacific Northwest National Laboratory which is operated by Battelle Memorial Institute for the U.S. Department of Energy.

Katipamula, S., M.R. Brambley and J. Schein. 2003. *Results of Testing WBD Features Under Controlled Conditions*. Task Report for the Energy Efficient and Affordable Small Commercial and Residential Buildings Research Program. Project 2.7 – Enabling Tools. Task 2.7.5. Included as part of *Final Report Compilation for Enabling Tools*, pp. Technical Report P-500-03-096-A7. California Energy Commission, Sacramento, California.