Determining an Accurate Counterfactual Baseline for Demand Response Events

William A. Lindsey, Tetra Tech, Madison, WI

Introduction

This poster presents a metric and analysis to aid in comparing counterfactual baselines used during a demand response event. During a demand response (DR) event, load is shed from the energy grid for a limited time in response to exceptionally high demand. For example, when the external temperature is extremely high and many customers are expected to use their air conditioning, a utility could call an event to limit participants' air conditioning use to help ease the load on the grid.

With these types of events, determining energy savings requires using a *counterfactual*, that is, hypothetically, how much energy would have been used if the event had not taken place. There are many ways to construct these counterfactuals, e.g., constructing baselines using data on non-event days and combining existing baselines with temperature data. We present a metric to compare how two counterfactual baselines perform and provide examples.

Data and Baselines

The DR events analyzed are from a southern state that runs multiple events during the summer months. We examined three events where the temperature on smart thermostats was raised by a few degrees in response to high external temperatures. Additionally, we consider six examples of counterfactual baselines: 1) Average of ten non-event weekdays preceding the event; 2) baseline #1 with a multiplicative adjustment based on actual usage the day of the event, prior to the event taking place, capped to an adjustment of ± 20 percent; 3) baseline #1 with an uncapped multiplicative adjustment; 4) baseline #1 adjusted for the increased temperature during the event; 5) the average of the four out of five non-event weekdays prior to the event with the highest usage; and 6) a cubic polynomial fit based on hours both prior and after the event.

Metric and Analysis

To compare how well the baselines perform, they need to be assessed on a day when actual energy usage is known. Each of the six baselines is calculated for the first non-event weekday prior to the event. Because this day has usage data, the baseline can be compared to the actual usage. We compute the mean squared error using the squared differences between the baseline and the usage for each hour in the event, and then average those across the entire event. For each pair of baselines, the baseline with the smaller error more accurately predicted the (known) usage on this test day.

Of the six baselines tested, the polynomial fit (6) performed the best on the test days, followed by the uncapped adjusted baseline (3). The ten-day average (1) was the worst performer. These results are logical; the further the weather measurements occur from the time of the event, the higher the likelihood that the weather will be dissimilar from the event day. Put another way, the closer the weather data are to the event hours/days, the more similar we should expect the temperature and humidity to be compared to the weather experienced during the event, and thus, the better a baseline constructed from these data will be.