Cage Match or Happy Couple? Engineering Simulation Models and Billing Analysis

Lauren Gage, Bonneville Power Administration, Spokane, WA
Dave Baylon, Ecotope, Seattle, WA
Josh Rushton, Rushton Analytics, Portland, OR
Michael Baker, SBW Consulting, Seattle, WA
Justin Spencer, Navigant Consulting, Boulder, CO

ABSTRACT

Billing analysis and engineering models are often seen as competing methods and are used in relative isolation. This paper describes the strengths, weaknesses and best uses of engineering simulation models and billing analysis – both separately and together.

In the Northwest, the Regional Technical Forum (RTF) estimates savings for measures with a heavy emphasis on calculation based on engineering principles using engineering simulation models. The simulation models allow the RTF to vary savings values by climate zone, building construction, and measure specification accounting for much of the variability and savings across program participants. However, the estimated savings must be calibrated to reality. Therefore, the RTF has used billing analysis to calibrate the simulation models.

This paper describes Northwest's approach to leveraging these two methods to accurately estimate savings. Recent examples include ductless heat pumps and weatherization. For these measures, the paper will discuss the methodologies that were used to estimate savings and the ways in which billing analysis and engineering simulation models were used together to improve the reliability of the results. We have examples of our successes, but we've also found that leveraging the two can be complex, frustrating and sometimes impossible. Billing analysis and simulation models aren't a happy couple, but we need them both.

Introduction

The Pacific Northwest is a unique environment for energy efficiency (EE) evaluation. The region includes four states (Idaho, Montana, Oregon, and Washington) and four separate regulatory environments. Additionally, nearly half of the load in the region is served by over 130 public utilities, whose loads are primarily served by Federal Columbia River Hydro and sold by Bonneville Power Administration.

In 1980, Congress passed the Pacific Northwest Electric Power Planning and Conservation Act[1] that created the Northwest Power and Conservation Council (Council), which has the responsibility of assessing the resource needs of the Northwest and developing a regional power plan every five years. Each plan assesses supply and demand-side resources and recommends the least-cost, least-risk approach to meet the region's power needs over the next 20 years. In these plans, EE is a large contributor and regional targets are set for EE achievements. In this planning paradigm, it is very important to accurately quantify the current and future impact of EE, as it is being used to displace or delay new generating resources. Yet, the need to determine attribution (e.g., net to gross estimation) is

[1] 16 United States Code Chapter 12H (1994 & Supp. I 1995). Act of Dec. 5, 1980, 94 Stat. 2697. Public Law No. 96-501, S. 885.

less important as all EE savings must be accounted for in determining the resources required to satisfy regional power demands.

The Regional Technical Forum (RTF) is an advisory committee to the Council, established in 1999 to develop standardized protocols for verifying and evaluating EE savings and to ensure the region continues to meet the Council's EE targets. The RTF comprises approximately thirty voting members who are selected for their technical knowledge. The committee is guided by the RTF Operative Guidelines (RTF Guidelines, 2014) in developing estimates of savings, costs and lifetimes for Unit Energy Savings (UES) measures (i.e., simple measures with ex-ante estimated per-unit savings) as well as standard protocols for calculating savings for more complicated measures using site-specific information.

The focus of the RTF has been on developing measure-specific energy savings estimates that the Council can use in power planning and that the region's energy efficiency programs can use for program planning and regulatory filings. To be useful in power planning, savings estimates must be reliable and forward-looking (estimates published today describe savings that may take place later). Where possible, the RTF tries to increase reliability and reduce the need for *ex post* true-ups by grounding savings estimates in relevant real-world data such as *in situ* metering data.

However, metering data is expensive to collect so the sample sizes that are obtainable do not always provide detailed insight into the savings of measures whose performance depends on multiple parameters that vary from site to site. For example, the amount of energy saved by an inch of ceiling insulation is highly dependent on climate, the amount of insulation already in place, and HVAC system efficiency. In cases like this, the RTF seeks to estimate different savings values for different scenarios. This helps program administrators prioritize high-savings opportunities, and it helps power planners estimate savings potential as the building stock evolves over time. The RTF uses engineering simulation models to obtain this kind of granular insight. The simulation models allow the RTF to vary savings values by climate zone[2], building construction, and measure specification accounting for much of the variability and savings across program participants.

Utilities in the Northwest routinely conduct traditional billing analyses for overall program impact evaluations, and the RTF routinely uses simulation models to estimate measure-specific savings. Recently, the RTF has worked to leverage billing analyses and engineering simulation models together to obtain savings estimates that are grounded in real-world data and provide the granularity desired by programs and planners. We have examples of our successes, but we've also found that leveraging the two can be complex, frustrating and sometimes impossible. Billing analysis and simulation models aren't a happy couple, but we need them both.

Methodological Considerations

Typically, the evaluation of a utility-sponsored energy efficiency program has three main goals: estimate overall impacts, estimate technology-specific savings, and attribute the savings. The different goals require different data collection and analyses, so a savings estimation approach that is appropriate to one goal may not be appropriate for others. This section describes performance characteristics and limitations of three basic approaches—billing analysis, engineering simulation, and the combination of the two. This section first describes the use of billing analysis and engineering simulations when they are used independently, including the strengths and weaknesses of each approach. Then, the integration of the engineering and billing analysis methods is described.

[2] The Northwest uses nine separate climates; 3 heating zones and 3 cooling zones.

Billing Analysis

Generally, billing analysis is used to estimate the overall energy impact of the measures delivered by a program and it is often used to determine the portion of that impact that should be attributed to the program. In a billing analysis, one or more regression models are used to analyze utility billing data. The models themselves vary depending on research needs and available data, but the basic objective is always the same: empirically estimate changes in energy consumption that are associated with program interventions.

Billing records provide observations of whole-building energy consumption at (typically) one-month intervals. Because billing data is generated automatically in the regular course of business, it is relatively easy to obtain multiple years of billing records for a large sample of sites. Thus, it can provide empirical estimates of program achievements that are both reliable and inexpensive. Yet, the reliability and granularity of billing analysis results are often limited because of the coarseness of the measurement (i.e., monthly whole building data), and the many unobservable variables that affect energy consumption. For example, a particular billing analysis might yield sound estimates of a weatherization program's total savings, but it may not be able to resolve measure-specific savings (e.g., window replacements versus floor insulation versus attic insulation). More broadly, billing analyses are of limited precision when the size of the savings is small relative to the unexplained variability in total consumption. Commonly-used methods reflect these challenges and opportunities.

Two basic approaches to billing analyses are described by Agnew and Goldberg (2013). These approaches take very different forms, but both seek to estimate average overall changes in energy consumption that coincide with program interventions. Both are capable of producing reliable estimates of average, overall savings when applied correctly, and neither typically produces measure-specific estimates.

- **Pooled fixed effects analysis.** This model fits all of the billing data – all sites and all billing periods – to a single large regression model. In fitting this regression, each billing record is associated with a single "observation" of all model variables. At minimum, the explanatory variables usually include heating degree days, cooling degree days, site-specific "fixed effect" indicator variables, and either a pre-/post- indicator (whose coefficient represents average savings) or an ex ante savings variable (whose coefficient represents a realization rate).
- **Two-stage analysis.** In this approach, the analyst first fits site-specific regressions to estimate energy consumption as a function of heating- and cooling-degree days and uses the fitted models to "normalize" each home's energy consumption in the pre- and post-periods. Next, the normalized consumption estimates are used to estimate the portion of consumption changes that are due to the program (in the simplest case, this is just the average change in participants' normalized consumption). The ductless heat pump and weatherization examples below each involve a form of two-stage analysis. In both cases, the first stage fits a variable-base-degree-day (VBDD) regression to each site. For each site, the VBDD algorithm calculates heating-degree days[3] under multiple different bases, fits a regression to each, and chooses the degree-base that explains the largest fraction of the variability in the site's energy consumption.

  In most applications of the two-stage analysis, there are some sites whose energy consumption does not correlate well with cooling-degree-days or heating-degree-days. Since these sites' regression fits do not provide a reliable way to normalize their consumption, it is common

[3] These particular examples focused on heating energy, so the VBDD algorithm only considered heating degree days. The basic algorithm works just as well with cooling degree days in warmer climates, or with both cooling and heating where applicable.

practice to remove them from the analysis.[4] This practice can introduce bias since homes whose consumption is poorly correlated with weather are likely to have different energy savings than those whose consumption is well-correlated. As shown below, the RTF uses another adjustment factor to mitigate this bias.

Like all savings estimates, billing analysis results should be interpreted carefully with respect to baseline. A simple participants-only billing analysis can directly estimate the average change in energy consumption that coincided with program interventions, after accounting for knowable external variables (e.g., weather, home characteristics). These changes can be interpreted as energy savings relative to the participants' pre-existing conditions, for the period of time covered by post-case billing data. However, some important variables may not be knowable. For example, if participants adopt additional efficiency measures beyond what was reported to the program, then those measures' savings will be incorrectly attributed to the known program interventions. A comparison group can mitigate the validity threat associated with unknown variables.

A billing analysis that uses a comparison group will yield an estimate of the change in participant consumption, minus the comparison group's change in consumption. With a carefully designed comparison group (for example, a sample of participants from previous program years), this kind of difference-of-differences can estimate a version of net program-attributable savings.

In all cases, savings estimates from a billing analysis are ultimately based on physical conditions that prevailed in some set of actual sites. Thus a billing analysis cannot directly estimate savings relative to a counterfactual baseline such as current practice[5] or codes or standards. For example, an organization may define the savings of an Energy Star heat pump to be the difference between Energy Star equipment and existing federal standards. A billing analysis cannot directly estimate this version of savings because the equipment used in "pre-program" participant homes is not likely to be as efficient as current federal standards.

Therefore, billing analysis tends to work best when the goal is overall program savings which can be detected through whole building measurement (such as heating or envelope measures) or when net attribution of savings is desired. But billing analysis cannot generally provide granular insight into individual measures (especially small measures), and since the method's savings estimates relate to observable baselines, billing analysis is usually not helpful when estimating savings from a current practice baseline.

Although this paper focuses on traditional billing analysis methods, it is important to note that the increasing penetration of AMI is already leading to new methods which may provide more granular and more reliable insight into some measures. The opportunities afforded by AMI data are not yet fully understood, but it is very likely that technology will become an increasingly important evaluation tool (Eckman and Sylvia, 2014), (Armel, *et al*, 2012).

Engineering Simulation

In contrast to the billing analysis models described above, engineering simulation models are focused on estimating measure or technology-specific savings. They attempt to estimate the energy use of buildings or end uses (e.g. estimation of heating load) by applying the laws of thermodynamics. Many tools for such modelling exist including DOE 2, Trace, EnergyPlus, and the Simplified Energy Enthalpy

[4] The actual filter usually an R^2 cut-off (sites whose regression fits have R^2 values below the cut-off are removed from the analysis). To be more literal, this means that a site is eliminated if a small portion of its month-to-month variability in energy consumption can explained by HDDs an CDDs.

[5] RTF Guidelines define current practice as the conditions that would prevail in the absence of the program, as dictated by codes and standards or the current practices of the market. For these measures, the baseline is defined by the typical choices of end users in purchasing new equipment and services at the time of RTF approval.

Model[6] (SEEM). The latter tool, SEEM, is used extensively by the RTF and was developed to meet Northwest's need for a simplified, adjustable engineering simulation model for residential buildings.

SEEM and other similar models require inputs that characterize the heat loss properties of a building (e.g., level of insulation in walls and ceilings), along with the characteristics of the mechanical systems that provide heating and cooling (e.g., air leakage rate for ducts that transport conditioned air). In addition, the models require inputs that characterize occupant behavior (e.g., thermostat settings), internal loads (e.g., lighting, appliances), as well as weather data for outside air temperature. The models estimate how much energy is required to balance heat losses and gains and maintain the temperature setting on the thermostat, including any mechanical energy (pumps or fans) required to move hot and cold air throughout the building or to provide ventilation. Some models, including SEEM, have add-ons that are used to estimate other end uses (e.g. water heating). In general however these thermal simulations do not estimate the other end uses in the buildings as these are inputs to the model.

The primary advantage of an engineering simulation model is its ability to estimate energy savings for individual measures and for variations in the application of specific measures, including ones that can't be directly observed in large enough numbers to attain statistically significant results. For instance, different levels of attic insulation can be modelled, covering both pre-conditions (existing) baselines and current practice (counterfactual) baselines. These models are also useful in estimating the interaction between savings of different measures (e.g., how much does attic insulation save if installed after a heat pump). For end uses that are inputs to the model, such as lighting, the simulations can calculate the interactions with other building characteristics. For example, engineering simulation can be used to estimate heating or cooling energy changes based on lighting measure installations.

Engineering simulations are attractive because the model directly represents measureable physical processes. If you give the model the right inputs, the model can produce accurate results. In principle, SEEM and EnergyPlus should accurately predict energy consumption and savings if all of the input parameters are accurate. However, this is not possible in practice because some important parameters can never be known with certainty (for example, occupants may open windows or change thermostat settings, and a building may include spaces whose temperatures deviate from the thermostat set point). Because of this, simulation models must be calibrated to real-world data.

Furthermore, realistic engineering simulations require a large number of inputs – for example, the 'simple' model, SEEM, has approximately 60 inputs. So in addition to the problem of limited accuracy, the sheer number of required inputs means that developing the models can be complex and resource-intensive.

Combined Billing Analysis and Engineering Simulation

As described above, billing analysis and engineering simulation each have strengths and weaknesses and therefore combining the two can be advantageous. The premise for using a combined method rests on two critical assumptions:

- Engineering models are grounded in an understanding of the physical drivers of energy savings that holds true across the range of extrapolation, but generally lack sufficient available data to accurately estimate savings.
- Billing analysis provides reliable estimates of savings for an aggregation of the population, but generally lacks sufficient precision to disaggregate savings at the measure level.

[6]Developed by Ecotope, SEEM is an hourly simulation model based on the premise that simplified algorithms can provide savings estimates for duct sealing, heat pump commissioning, and other equipment measures that are: (1) Reliable enough to meet the present needs of integrated resource planners, and (2) Flexible enough to estimate heating energy consumption (and savings) across a variety of insulation and residential equipment conditions (and measures). With SEEM, users can access relatively complex simulations and modify them to measures and conditions that change over time.

Research that combines these billing and engineering models typically use two primary methods:

Statistically adjusted engineering models (SAE) have been used in evaluation for many years. With SAE models, engineering results are first developed, and then they are used as an independent variable in the econometric model that estimates savings (as in the pooled fixed effects model described in the billing analysis section above). The idea is that the engineering models explain some of the variation in the data and improve the accuracy of the econometric model and allow it to be extrapolated to a wider set of cases. SAE models can use any sort of engineering estimates of savings, ranging from a single set of deemed values to detailed, site-specific building energy simulations. SAE models can work well when engineering models explain much of the variability and when they are applied to a very large sample of participants. However, the statistical adjustments themselves take the form of a small number of very coarse adjustments that apply to broad measure categories.

Calibrated building energy simulation can be characterized as billing analysis first, then building energy simulation modeling to improve the accuracy of the savings. This method allows for a better understanding of the "why" savings results and also allows modelers to divvy up the overall billing results into measure-specific results. Calibrated building energy simulation uses data derived at the whole premise level (typical billing data) or, much more infrequently, uses end-use metering results.

Calibrated building energy simulation is currently the method of choice for developing UES estimates of savings by the RTF. For the RTF, understanding both the overall savings and the measure-specific savings are important as they provide the basis for developing reliable savings estimates. Additionally, for the RTF Guidelines, an RTF-approved Proven UES[7] measure only requires delivery verification for impact evaluation in the future (RTF Guidelines, 2014), as the savings are judged to be reliable for future participants.

Calibrated building energy simulation requires a description of the engineering and physical parameters of energy savings as well as an understanding of the overall change in energy use by the participants. Functionally, this amounts to a process that identifies and modifies inputs to simulation models as part of calibration. In the RTF case, the models have been calibrated to estimates of consumption developed from premise-level disaggregated billing consumption from detailed metering (e.g., ductless heat pumps as described in sections below) or regional surveys of building characteristics (e.g., weatherization, as described in sections below). This sort of calibration rests on two fundamental assumptions:

- Billing analysis is sufficiently unbiased that it can be used as a "truth set" for the comparison;
- A calibration based on an observational cross-section of sites, rather than pre-/post- program data, can be used to reliably estimate savings.

The first assumption is difficult to verify since we often do not have access to any higher truth set than billing data. The second implies that, even after calibration, savings results should be checked against pre/post- program data. As of this writing, the RTF has had two opportunities to compare calibrated-SEEM-based savings estimates to billing data collected by programs that promote weatherization measures. Both comparisons were of limited precision but both found reasonable agreement.

The calibration process forces a careful look at uncertainty and can be fraught with conflict between econometricians and engineers, as it reveals unresolved issues in one approach or the other. There are many more possible sources of error or bias to consider, requiring separate areas of deep technical expertise to assess, which necessitates the involvement of a diverse group of technical

[7] Measures in the UES category generally have savings based on a per-unit basis (e.g., savings per light bulb) and are estimated for a typical or average site. This category is appropriate for measures with relatively small variation in the savings that can be reliably forecast. These measures have savings determined in the prospective period.

reviewers and a longer process. However, savings subjected to this kind of rigorous calibration process are likely to be more robust and ultimately viewed as more reliable by stakeholders.

Methods in Practice – Northwest Examples

In the Pacific Northwest, there have been several large scale efforts to evaluate technologies using an integrated approach that includes both engineering and billing analysis techniques. In the residential sector these measures include high performance heat pumps, heat pump water heaters, ductless heat pumps (DHPs) and weatherization. This section will describe two of these efforts, DHP and weatherization, as examples of the integrated evaluation approach as a key part of implementing emerging technology into large scale utility programs.

Ductless Heat Pumps

The Northwest ductless heat pump research study, led by the Northwest Energy Efficiency Alliance (NEEA, 2014) was a large scale research project that combined billing analysis and engineering simulation to develop highly reliable, granular savings estimates for DHPs across the region. The project included the installation of nearly 4,000 ductless heat pumps in homes with zonal electric heat; the existing heating equipment was not removed or disabled. The DHP is particularly valuable to the Northwest due to the region's relatively large share of homes with zone-based electric baseboards[8]. The program trained contractors on quality installation practices and on careful collection of customer information that was later used in the savings assessment of the DHP technology.

The evaluation of the pilot program began in early 2009 and proceeded in four distinct steps: performance testing, detailed metering, billing analysis and combination of results. These pieces leveraged building simulation modeling and billing analysis into a final, reliable result. Yet, described below, it was not a completely happy couple.

Step 1: Performance testing: The principle goal of this step was to verify and expand the manufacturers' ratings for the performance of the equipment. Because the DHP technology employs inverter driven (variable speed drives) equipment, the performance assessed in standard ratings was not considered reliable and a testing protocol was developed to assess a wide variety of conditions that would be useful in understanding the equipment in applications in the climates of the Pacific Northwest.

The testing protocol was developed in consultation with the region and the tests were implemented on two pieces of equipment that were dominant in the regional market. Tests were conducted by Harrick Labs at Purdue University and included efficiency and heating/cooling output at a variety of temperatures ranging from -10ºF to 110ºF (i.e., beyond standard ratings but representing the temperatures typical in the region).

The results of these tests were summarized in a form that could be used to modify the regional residential simulation tool, SEEM. These performance results (NEEA, 2014), combined with the results of detailed field monitoring (Step 2), provided valuable input for the subsequent engineering evaluation of the technology (Step 4).

Step 2: In-Field Metering: The goal of detailed metering was to assess the performance of DHP equipment in actual homes and compare the results to the laboratory testing. The metering protocol included sensors on the main circuits (DHP, electric heating, hot water, and total service), outdoor and

[8] These homes have not switched to air-source heat pumps over the past 20 years of program encouragement and also are likely to use supplemental fuels (especially wood heat) which has resulted in consistently poor evaluation of savings from weatherization measures.

indoor temperature, and cooling status of the DHP (using vapor line temperature).[9] In addition to the verification of the laboratory test results, the metering protocol allowed an assessment of the interactions between the zonal electric resistance heating system and the DHP equipment. This information was critical to the simulation assessment of the DHP savings potential.

Ninety-five homes were selected at random across the region's climate zones, including 52 located in the mild climates west of the Cascade Mountains and 43 located in the more severe climates of the eastern parts of the region. The recruiting tried to screen out homes that had supplemental fuels used for heating (especially wood heat) in order to provide an analysis dataset where all energy inputs were known (just electricity in this case); supplemental fuel use is very difficult to accurately quantify.

The analysis of these homes focused on producing the information necessary to simulate the performance of the DHP in the context of the existing electric resistance heating and the actual occupant choice between these systems throughout the heating season. The simulation required that this relationship be explicitly modeled. Thus the key outputs were the amount of heat supplied by the DHP, the amount of heat supplied by the pre-existing zonal electric heating system and the ratio between these two systems. The analysis also included a variable-base degree-day (VBDD) regression applied to the metered homes. Then the detailed metering data was compared with the VBDD results in the post-installation period, which provided an estimate of the error from the VBDD heating/baseload disaggregation. The translation of the energy use from the VBDD results to the metering results in the post period was then applied to the pre period VBDD results as an attempt to correct for temperature bias (especially hot water usage) in the pre period.

Because cooling use was not expected to be assessed using the billing data (due to the short NW cooling season), the cooling performance of the DHP was assessed using Step 1's lab results and this step's field data collection (cooling energy use was directly metered).

Step 3: Billing analysis of Participants: A large scale billing analysis was performed as a way to understand the savings of DHPs in homes across the entire program and region. The real-world behaviors of occupants with the temperature settings and existing zonal systems were expected to be major contributors to energy savings. A large billing analysis would help to scale and identify aggregate behavior that affect savings independent of the performance of the equipment.

VBDD analysis was conducted on electric bills pre- and post- DHP installation. A savings estimate was calculated for each participant by taking the difference between normalized heating consumption in the pre- and post-installation periods (the analysis did not include a separate control group). From nearly 4,000 participants, approximately 3,600 participants remained for analysis after screening for incomplete or anomalous billing records. An additional 7% of cases were screened after the VBDD analysis due to very poor fits (R^2<.45) of the heating signature in the pre installation case.

In addition to information on the DHP measure installed, the program collected detailed information on the occupants, including their heating patterns and the use of supplemental fuels of all types. This detailed information allowed a direct comparison to the metering sample (Step 2 above) since it was possible to screen the entire sample to remove homes with reported supplemental fuel use and derive savings estimates comparable to the metered sample. In fact, the savings estimates from the detailed metering and the billing screened for supplemental fuel use were within 10% of each other. The program intake questionnaire also provided a basis for segmenting the billing analysis results to test the influence of different variables (e.g., climate and various occupancy variables). Ultimately, the most reliable explanatory variables were found to be the climate zone of the participants and reported use of supplemental fuels.

[9]Pre-metering of the existing heating system was not conducted; most homes had no pre-existing cooling systems, except for some window air conditioning.

The ability to estimate the impact of supplemental fuel use on DHP savings was one of the most important findings of the DHP study. The problem of supplemental fuels has plagued evaluations of weatherization space heating measures in the Pacific Northwest since these programs were introduced in the mid-1980s. The size of this sample and the reliability of the occupant information gave a direct insight into the DHP savings where supplemental fuel heating was present. To estimate the amount of supplemental fuel used, the billing analysis sample was separated into homes that used supplemental fuels and homes that do not, and the two groups' electric heating energy was compared pre and post. This comparison allowed an inference of the amount of supplemental fuel that was offset by the DHP usage.

Step 4: Combining methods: Simulation calibration and cost-effectiveness. The wealth of information derived from Steps 1-3 above provided an unprecedented amount of data and granularity in assessing the savings from the DHP installations. At this point, SEEM had been updated with a DHP and zonal-electric heat model based on the lab and field studies, The billing data had been analyzed to estimate kWh savings, to verify the supplemental fuel screens, and to estimate the saturation of supplemental fuel use in each climate zone.

The savings results were developed into a set of UES measures for use by all utilities across the region[10]. In this final step, SEEM was used to apply the DHP/electric zonal model to standard prototypes developed by the RTF. These prototypes were representative of common homes found in the Northwest for surface areas, foundation types, glazing areas, internal gains, etc. In this analysis, the RTF employed the recently available the Residential Building Stock Assessment (RBSA) (Baylon, 2013), a regional survey of existing homes that included a physical audit data and billing data. The building parameters from the RBSA and the standard prototype inputs were used as the inputs in SEEM to generalize the DHP savings estimates across the region by climate zone. The resulting UES measures account for several important findings:

- The underlying analysis linked the performance of DHPs with the manufacturer's ratings and provided reliable specifications for future utility programs.
- The measures show how supplemental fuels affect savings and cost effectiveness so utilities can determine whether to include supplemental heat screens in their potential program designs.
- DHP savings was developed for each climate zone taking into account the impacts of winter temperatures on the performance of the DHP.

Although the Northwest achieved results, the extensive and diverse data sets that resulted from these various efforts became a complication for the RTF. The billing analysis results of first year savings competed with detailed metering analysis. Adding in the impact of supplemental fuel heating was more difficult for calibration, so the RTF ultimately recommended a separate savings estimate for programs that did not screen for supplemental fuel heating use and for programs that did. The RTF spent significant time debating and reviewing the results over many months. It was difficult and complex and the detailed analysis resulted in significant reductions in savings estimates from the planning estimates for measure. The RTF, however, ultimately agreed upon UES savings for DHP measures for use in utility programs.

Weatherization

In addition to the DHP assessment described above, the RTF has conducted a large calibrated-SEEM exercise to update the weatherization measure savings (i.e., wall, ceiling and floor insulation, windows and heating equipment) estimates in the region. In this case, RBSA site-audit and billing data

[10] Most recent RTF measure workbook available at: www.rtf.council.org/measures

were used to estimate back-end adjustment factors needed to align SEEM output with observed consumption data (on average). The calibration itself, which the RTF expects to use on relevant measures on an ongoing basis, was organized in two distinct steps, or phases. Rushton and Hadley (2014) provides a detailed description; the basic steps are as follows:

Phase I: Calibration for Homes without Supplemental Heating: This part of the calibration aligned SEEM heating energy estimates with billing data estimates for RBSA homes with no off-grid heating fuels (such as wood, oil, or propane) and whose billing data exhibited clear heating energy signatures.

First, the RBSA homes whose audit data showed evidence of supplemental (i.e., non-utility heating) fuels, or whose cold-weather billing data did not correlate well with heating degree days were filtered from the Phase I analysis.[11] Two separate heating energy estimates were then calculated for each site in the Phase I sample: 1) a billing data estimate calculated through a VBDD analysis and normalized to Typical Meteorological Year (TMY) weather conditions and 2) SEEM results (generated from the audit data in the RBSA and TMY weather files).

The SEEM models used site-specific audit data for many input parameters, such as shell component U-factors, square footage, equipment, and location. Some parameters, such as internal gains, solar heat gains, and infiltration (when not measured) were based on a combination of data and judgment. The SEEM input parameters were all subject to some degree of error, but importantly, they were all determined by a clearly-defined set of standardized input conventions. Thermostat settings are particularly important to any simulation estimate, and in this case a single fixed thermostat setting (69^{0}F day / 64^{0}F night) was used for all runs in all homes regardless of the thermostat setting reported by the home's occupant. The standardized thermostat setting is likely the source of much of the systematic variation between the SEEM-based and VBDD-based estimates.

The heart of the Phase I calibration was a regression that estimated the average difference between SEEM output (with the standardized inputs just described) and the VBDD heating energy estimates. Since actual consumption was highly variable and SEEM inputs had limited resolution, the regression had to identify any trend(s) through a very noisy cloud of observations. A highly detailed model cannot be reliably fit to a few hundred data points with this degree of noise, so a central challenge of the calibration was to capture the main trend in a way that was fine-grained enough to be useful but coarse enough to be reliable.

A basic Phase I finding was that SEEM (with the standardized inputs) tends to over-estimate heating energy in inefficient homes, and it tends to under-estimate heating energy in efficient homes. This finding was consistent with the logic of comfort take-back (when an occupant increases the thermostat setting when heating becomes more affordable or, conversely, decreases the setting when heating is expensive). However, the RTF's analysis cannot, and did not attempt to, establish the root cause of the trend.

The final Phase I calibration result was a function that calculates a SEEM adjustment factor (i.e., a factor applied to SEEM's heating energy use output for a home) as a function of the home's surface-weighted average U-factor, its heating equipment type (electric resistance, heat pump, or natural gas furnace), and its heating zone.

Phase II: Calibration to the Population: The Phase II calibration accounted for differences between homes that were included in the Phase I analysis and homes that would be typical among utility program participants. That was necessary because program participation often includes homes with

[11] To eliminate an obvious bias risk, homes whose audit data suggested substantial cold-weather loads unrelated to weatherization, such as hot tubs or heated workshops, were also removed.

supplemental fuels, anomalous heating energy signatures, or intermittent occupancy. This part of the calibration estimated the fraction of heating energy that was accounted for by different fuels (electricity, gas, and off-grid fuels) and developed correction factors to adjust for the bias introduced when homes were filtered out of the Phase I sample. As with the Phase I analysis, a major challenge with Phase II was to determine the level of granularity that the available data could deliver.

After much discussion, the RTF determined that a small set of relatively coarse adjustment factors would sufficiently address the Phase-II calibration needs that were important to weatherization and heating equipment measures. For example, in heating zone 1, homes with electric heat met about 9% of their heating load, on average, with off-grid supplemental fuels.

Applying the calibrations: To estimate savings for weatherization and equipment measures, the RTF first developed SEEM inputs that describe prototypical base-case homes (based on the RBSA audit information) and efficient-case homes (based on measure specifications). In both cases, care was taken to ensure consistency with the standardized input conventions used in the Phase I analysis. Phase I adjustment factors were then calculated for the different prototypes, and these were applied to the SEEM output to obtain Phase-I-calibrated energy estimates for the two cases. The difference between the two estimates was the Phase-I-calibrated savings. This savings figure accounted for whatever combination of input-parameter errors, thermostat take-back, and modeling shortcuts were captured in the Phase I regression model.

Phase II adjustment factors were then applied. Since the Phase II adjustment was a single fixed factor (depending on heating zone), the effect was to reduce the savings of the weatherization measures to account for the effects of supplemental fuels, intermittent occupancy, and other features captured in the adjustments. (Note that these adjustments would not be appropriate for a utility program that screens participants for wood heat, consistent occupancy, or heating signature.)

During this process, the RTF has had two opportunities to compare calibrated-SEEM-based savings estimates to billing data collected by programs that offer weatherization measures. Both comparisons were of limited precision but both showed reasonable agreement. The RTF approved these measures[12] and their combined methodology, but it was not without significant effort by RTF staff and RTF members. As shown above, these steps of calibration and comparison are complex and they were difficult to follow for many RTF members. Many of the decisions embedded in the calibration (which seem reasonable in retrospect) required significant discussion and disagreement before a result was approved. Some difficulties encountered were overcome by carefully reconsidering the level of granularity required by RTF measures and others by recognizing the limitations of available data and simply deciding to press ahead.

Conclusions

In the Northwest, we are finding that the combination of billing analysis and engineering simulation allows us to continue to focus on the efficiency of equipment and the engineering of the end uses, while still understanding the overall savings that can be expected from those measures. That is, rather than using just engineering analysis (and paying less attention to the user behavior) or just a billing analysis (and paying less attention to the performance of the measures), we can bring the value of both approaches inform the final results. For the RTF, the overall effect is that both methods are included (sometimes) but that the nature of this compromise is not really predictable: It could be a Happy Couple or it could be a Cage Match, or both.

[12] Most recent RTF measure workbook available at: www.rtf.council.org/measures

REFERENCES

Agnew, Ken, and M. Goldberg. 2013. "Whole-Building Retrofit with Consumption Data Analysis Evaluation Protocol." *NREL Uniform Methods Project* (2013): Chapter 8. U.S. Department of Energy.

Armel, K.C., A. Gupta, G. Shrimali, amd A. Albert. 2012 "Is disaggregation the holy grail of energy efficiency? The case of electricity." *Energy Policy*, Volume 52, Issue C, 2012.

Baylon, D., P. Storm, K. Geraghty, R. Davis, 2012. *Residential Building Stock Assessment: Single-Family Characteristics and Energy Use*. Portland, OR: Northwest Energy Efficiency Alliance.

Crossman, K., L.Tabor, M Perussi, and D. Basak, L. 2013. "Dynamic Duo: How Combining Billing Analysis and Engineering Simulation Methods Improves Evaluation Quality and Understanding". In Proceedings of 2013 of International Energy Program Evaluation Conference, 221.

Eckman, T., and M. Sylvia. 2014. "EM&V 2.0: New Tools for Measuring Energy Efficiency Program Savings." *Electric Light and Power*, 2014.

NEEA. 2014. "Final Summary Report for the Ductless Heat Pump Impact and Process Evaluation." February 2014. Portland, OR: Northwest Energy Efficiency Alliance.

Regional Technical Forum, 2014. "Complete Operative Guidelines". Portland, OR. Northwest Power and Conservation Council.

Rushton, J., and A. Hadley, 2014. "Where Did it Go? Lost Savings Found in Real-World Data". In Proceedings 2014 of ACEEE Summer Study on Energy Efficiency in Buildings." 1-1192. Washington, D.C.: American Council for an Energy-Efficient Economy.