A High-Granularity Approach to Modeling Energy Consumption and Savings Potential in the U.S. Residential Building Stock

Preprint

Eric Wilson, Craig Christensen, Scott Horowitz, and Henry Horsey

Presented at the ASHRAE and IBPSA-USA SimBuild 2016 Building Performance Modeling Conference
Salt Lake City, Utah
August 8-12, 2016

NREL is a national laboratory of the U.S. Department of Energy
Office of Energy Efficiency & Renewable Energy
Operated by the Alliance for Sustainable Energy, LLC

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

Contract No. DE-AC36-08GO28308

Conference Paper
NREL/CP-5500-83077
May 2022

A High-Granularity Approach to Modeling Energy Consumption and Savings Potential in the U.S. Residential Building Stock

Preprint

Eric Wilson, Craig Christensen, Scott Horowitz, and Henry Horsey

Suggested Citation
Wilson, Eric, Craig Christensen, Scott Horowitz, and Henry Horsey. 2022. *A High-Granularity Approach to Modeling Energy Consumption and Savings Potential in the U.S. Residential Building Stock: Preprint*. Golden, CO: National Renewable Energy Laboratory. NREL/CP-5500-83077. https://www.nrel.gov/docs/fy22osti/83077.pdf.

NREL is a national laboratory of the U.S. Department of Energy
Office of Energy Efficiency & Renewable Energy
Operated by the Alliance for Sustainable Energy, LLC

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

Contract No. DE-AC36-08GO28308

Conference Paper
NREL/CP-5500-83077
May 2022

National Renewable Energy Laboratory
15013 Denver West Parkway
Golden, CO 80401
303-275-3000 • www.nrel.gov

NOTICE

This work was authored by the National Renewable Energy Laboratory, operated by Alliance for Sustainable Energy, LLC, for the U.S. Department of Energy (DOE) under Contract No. DE-AC36-08GO28308. Funding provided by the U.S. Department of Energy Office of Energy Efficiency and Renewable Energy Building Technologies Office and the Bonneville Power Administration. The views expressed herein do not necessarily represent the views of the DOE or the U.S. Government. The U.S. Government retains and the publisher, by accepting the article for publication, acknowledges that the U.S. Government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this work, or allow others to do so, for U.S. Government purposes.

Cover Photos by Dennis Schroeder: (clockwise, left to right) NREL 51934, NREL 45897, NREL 42160, NREL 45891, NREL 48097, NREL 46526.

NREL prints on paper that contains recycled content.

A HIGH-GRANULARITY APPROACH TO MODELING ENERGY CONSUMPTION AND SAVINGS POTENTIAL IN THE U.S. RESIDENTIAL BUILDING STOCK

Eric Wilson[1], Craig Christensen[1], Scott Horowitz[1] and Henry Horsey, Member[1]
[1]National Renewable Energy Laboratory, Golden, CO

ABSTRACT

Building simulations are increasingly used in various applications related to energy efficient buildings. For individual buildings, applications include: design of new buildings, prediction of retrofit savings, ratings, performance path code compliance and qualification for incentives. Beyond individual building applications, larger scale applications (across the stock of buildings at various scales: national, regional and state) include: codes and standards development, utility program design, regional/state planning, and technology assessments. For these sorts of applications, representative buildings are needed for simulations to predict performance of the entire population of buildings.

Focusing on the U.S. single-family residential building stock, this paper will describe how multiple data sources for building characteristics are combined into a highly-granular database that preserves the important interdependencies of the characteristics. We will present the sampling technique used to generate a representative set of thousands (up to hundreds of thousands) of building models. We will also present results of validation against building stock consumption data.

INTRODUCTION

In what Swan (2009) refers to as bottom-up, archetype, engineering models, relatively small numbers of 'typical' or 'average' buildings have been used to represent building stocks for purposes of modeling with detailed building simulation tools (Huang et al 1999, Hopkins et al 2011, Taylor et al 2012). With today's available computing resources, it is useful to ask what is an appropriate number of representative buildings and how should they be defined. A large number of representative buildings may be appropriate, considering the possible combinations that result from the range of general building characteristics that influence energy consumption (location, vintage, size, number of stories, foundation type, heating fuel type, etc.) and specific building component characteristics (such as insulation levels, window types, etc.) that tend to vary with location and vintage. This paper describes a high-granularity approach using Latin hypercube sampling (Saltelli 2008) with building characteristic distributions to automatically generate a large number of statistically representative building archetypes, to provide model results with adequate sensitivity to address a wide range of analysis questions.

BUILDING CHARACTERISTICS

For residential building stock analysis, energy simulations of representative buildings require inputs based on characteristics of actual buildings. Table 1 shows building characteristics, dependencies, and data sources for the high-granularity approach explored in this paper.

Archetype Parameters and Variant Characteristics

As observed in building characteristics data, dependencies are such that there are certain aspects of buildings (e.g., location, vintage, heating fuel type, etc. as the headings listed under Dependencies in Table 1) upon which other building characteristics (e.g., insulation levels, window type, etc.) depend. We refer to the former as *archetype parameters* and the latter as *variant characteristics.* The hierarchy in Figure 1 illustrates such dependencies, including the fact that archetype parameters can also depend on each other.[1] Combinations of archetype parameter values and dependent variant characteristic values lead to *archetype variant buildings* to be simulated.

[1] The order of the hierarchical structure is somewhat arbitrary; for example, the same data set could be queried to develop location weighting factors as a function of vintage. Once weighting factors have been developed based on a particular hierarchical order, then that order is used for dependency-based calculations.

Characteristic Distributions and Current Values

For variant characteristics, typical or predominant values may come to mind, but the data is better represented by probability distributions. For example, many homes built in the 1950s have uninsulated walls with a few built to higher standards or since retrofitted. Attic R-values, on the other hand, are more likely to have a broader range of values, as a result of retrofitting at different times (motivated by different utility rates and/or incentives) over past decades.

For any given archetype, current buildings (as they exist today) include variant characteristics that depend on building components that are: 1) as-built, 2) retrofitted, or 3) replaced.

Envelope component characteristics are predominantly as-built – characterized with data based on new construction builder surveys, building codes, and standard construction practices. For retrofits, estimates are included for the fraction of building components that have been retrofitted and the retrofit efficiency level.

Equipment characteristics for older vintage buildings are predominantly based on replacements; data can be generated based on component lifetimes and equipment sales data. For new building vintages, as-built characteristics are still current; data can be derived from equipment energy standards. Retrofits (early replacements) are less common; data can be estimated. In all cases, efficiency levels may vary, including some consumer choice of upgrades. Diversity in probability distributions, however, primarily reflects the mix of as-built, replacement and retrofits.

Table 1. Building characteristics, dependencies, and data sources

		Dependencies								Data Sources												
	Characteristics	*Location*	*Vintage*	*Heating Fuel*	*Usage Level*	*Daytime Use*	*Floor Area*	*Number of Stories*	*Found. Type*	2009 RECS	NAHB Surveys	IECC Codes	RBSA	Ritschard et al. 1992	2012 ACS	Carmody et al. 1988	Chan et al. 2012	Wenzel et al. 1997	BECP 2009	Eng. Exp. & Calib.	Geographic Resolution	# of Options
Meta	*Location*																				TMY	216
	Vintage	✓													•						C	7
	Heating fuel	✓	✓							•											C	6
	Usage level																			•	U.S.	3
	Daytime use																				U.S.	2
Geometry	*Floor area*	✓	✓							•											R	6
	Number of stories	(✓)	✓				✓		✓	•											R	3
	Foundation type	✓									•					•					48	5
	Attached garage	(✓)	✓				✓			•											R	2
	Orientation																				U.S.	4
Envelope	Window type	✓	✓							•		•								•	R	5
	Wall insulation	✓	✓								•			•							R	8
	Ceiling insulation	✓	✓								•		•								R	7
	Foundation insulation	✓	✓								•									•	R	5
	Air leakage	✓	✓				✓	✓	(✓)								•			•	R	12
Equipment	Heating system type	✓	✓	✓						•	•										R	6
	Heating system efficiency	✓	✓								•							•		•	R	10
	Cooling system type	✓	✓							•											R	7
	Cooling system efficiency	✓	✓								•							•		•	R	7
	Duct insulation, tightness	✓	✓						✓			•							•	•	U.S.	5
	DHW system type	✓	(✓)	✓						•											R	5
	DHW system efficiency																	•		•	U.S.	3
	Cooking type	✓	(✓)	✓						•											R	10
	Clothes dryer type	✓	(✓)	✓						•											R	10
Occupancy	Heating, cooling setpoints	✓				✓				•											TMY	3
	Cooking usage				✓	✓														•	U.S.	3
	Clothes dryer usage				✓	✓														•	U.S.	3
	Lighting, Appliances, MELs				✓	✓				•										•	U.S.	3

✓= direct dependency (✓)= indirect dependency *italics = archetype parameters*
C = Census Tract R = Regional (custom) TMY = 216 TMY sub regions

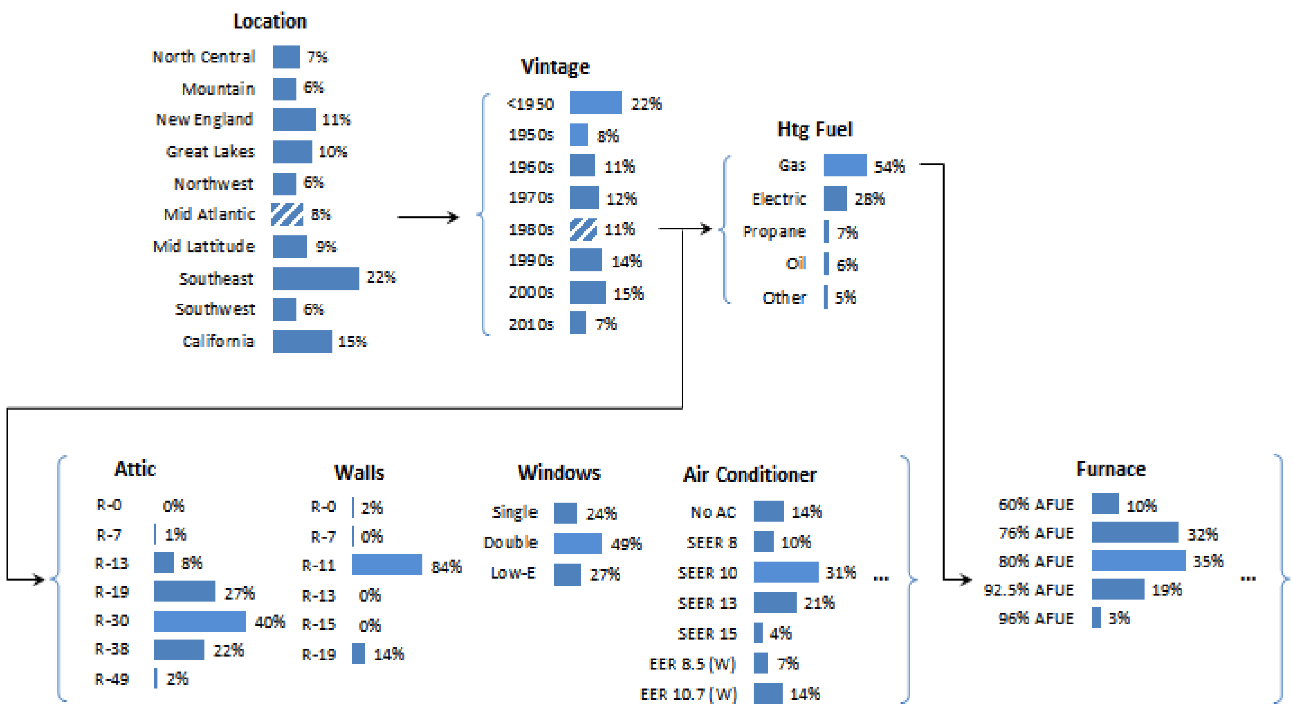

Figure 1. Variant characteristics probability distributions based on archetype parameter values. (W) indicates window air conditioners.

OVERALL PARAMETER SPACE

Theoretically, a very large number of archetype variants exist, based on all possible combinations of characteristics. However, within this parameter space, archetype variants represent differing numbers of actual homes, depending on the product of the archetype probability and the variant characteristic probabilities.

In fact, many cells in the parameter space will be essentially empty (i.e., many theoretical variants will represent zero or a statistically insignificant number of actual homes). For example, the combination of "built in the 2000s in the Southwest, with a basement and oil heat" will represent few, if any, actual homes. Obviously, modeling such variants is unnecessary. Nevertheless, the number of archetype variants associated with actual existing buildings can be very large, and approaches to limiting the number of archetype variants to be simulated will be considered (see: Modeling Approaches section).

Non-Correlated Variant Characteristics

Beyond mutual dependence on archetype parameters, we mostly lack statistical data on relationships between variant characteristics (e.g., insulation levels, window type); therefore, we assume no direct dependencies.[2] Archetype variant buildings are defined based on different combinations of these characteristics.

If further detailed data was available, additional archetype parameters could be developed where appropriate. For example, if relationships between insulation levels for different envelope building components (e.g., walls, attic/roof, foundation) beyond vintage and location dependencies were found, an archetype parameter that qualitatively describes the building envelope as well insulated, moderately insulated, or poorly insulated could be developed with the aforementioned envelope building components dependent on it.

Visualizing the Parameter Space

The parameter space can be visualized as a hierarchal tree structure covering all possible combinations of building characteristics in archetypes and variants. The tree structure branches out (based on the number of options in each probability distribution) through archetype parameters (in order of dependencies) and then through uncorrelated variant

[2] Because there are no dependencies assumed among uncorrelated characteristics, the sequential ordering in the lower part of Figure 1 is arbitrary.

characteristics. Each path from trunk to twig represents a theoretical archetype variant. Thickness at any point depends on cumulative product of probabilities to that point; at the end of the branch, the thickness represents the archetype variant direct house count.

MODELING APPROACHES

Alternative approaches to selecting representative buildings (archetype variants) to be simulated include:

EPS (Entire Parameter Space)

If granular data is used, the entire domain space includes a very large number of theoretical archetype variants (all possible combinations of characteristics). Even after eliminating combinations for which no actual homes exist (a computational challenge in itself), the domain space is too large to run detailed building energy simulations, even using high performance computing resources.

TPH (Typical Prototype Houses)

Historically, a limited number of "prototype" buildings have been used, often with each characteristic represented by a single typical, predominant or average option rather than a probability distribution. The limited sensitivity of such models may impose limitations regarding the sort of analysis questions that can be accurately addressed.

MHC (Maximum House Count)

Sampled archetype variants, selected to have the highest possible house-counts, maximize the number of actual houses directly represented. Simulations are prioritized for high house-count archetype variants, and energy results for those variants are multiplied by the associated house-count weighting factors. However, high-probability archetype parameters and characteristics end up overrepresented while low-probability archetype parameters and characteristics are underrepresented.

LHS[3] (Latin Hypercube Sampling)

Simulations are distributed across a wide variety of archetype variants according to archetype and variant characteristic probabilities. LHS is often used for computer experiments. The approach naturally includes simulations for many archetype variants with high direct house-counts resulting from high probabilities for some variant characteristics and combinations thereof, but does not focus exclusively on such variants (as the MHC approach does). Some archetype variants with lower direct house-counts are included to match the overall probability distributions. Sampled archetype variants represent fewer actual houses directly than in the MHC approach, but are designed to statistically represent the entire housing stock as best as possible for a given number of simulations.

Summary of Approaches

The first and second approaches bound the range of possibilities: using a great many simulations to using only a limited number. The third and fourth approaches use an intermediate (but perhaps, large by historical standards) number of simulations. For a highly dimensional space, such as the U.S. residential building stock:

- EPS – number of simulations prohibitive.
- TPH – granularity often insufficient for answering analysis questions accurately.
- MHC – over-represents high probability building characteristics.
- LHS – best balance between representing housing stock characteristics and number of simulations.

The LHS approach is recommended for residential building stock modeling and is described in more detail in the next section.

---

[3] By comparison to simple random (Monte Carlo) sampling which requires large sample sizes to match each building characteristic probability distribution, the LHS approach procedurally ensures a good match (limited by sample size resolution, as illustrated in Figure 4).

The described LHS approach differs from classical LHS in two ways:

a) the standard LHS implementation specifies that a variable value appears in only one sample, but here the mapping of probability distributions purposely leads to characteristic options appearing in multiple archetype variants; however, the described approach does preserve the desired LHS property of sampling ranges with equal probability and

b) classical LHS does not include dependencies, and its elegant solution based on simply selecting each sample.

THE LHS APPROACH

The LHS approach, using conditional probability distributions, (see Table 2) includes the following steps:

1. Choose number of simulations, e.g., m = 100,000
2. Construct a matrix with m rows (samples) and n columns (archetype parameters and variant characteristics—ordered by dependencies):
   a. Populate each column with m sample #'s in random order
   b. For each row,
      i. *7* For each column (from left to right): map sample # to characteristic option, using sample # ranges (see Figure 2) proportional to the applicable probability distribution (dependent on archetype characteristics in previous columns)
      ii. Repeat for next column *[when all columns for the row have been processed, the result is an archetype variant to be simulated]*
   c. Repeat for the next row *[when all rows have been processed, the result is a complete set of archetype variants to be simulated]*
3. Simulate the LHS selected archetype variants
4. Multiply each simulation result by a LHS scaling factor (= total # of houses / # of simulations)

For each LHS archetype variant, the scaling factor does not equal the direct house count that would calculated as the cumulative product of building characteristics probabilities, because each LHS archetype variant also indirectly represents additional (similar) archetype variants (not simulated). For LHS, the same-scaling factor applies to each simulation.

Table 2. Example LHS simulations (with dependencies ⟶) developed from randomly ordered sample numbers (S#) mapped to building characteristics (Ch) based on ranges in Figure 2.

	Location		Vintage		Htg Fuel		Htg Sys		Attic		Wall		Windows	
Sim	S#	Ch	S#	Ch	S#	Ch	S#	Ch	S#	Ch	S#	Ch	S#	Ch
1	89 →	CR11	14 →	1950s	24 →	Gas	1 →	AFUE-60	37 →	R-13	25 →	R-0	92 →	Single
2	13 →	CR2	93 →	2000s	16 →	Gas	58 →	AFUE-80	55 →	R-49	31 →	R-11	11 →	Double
3	61 →	CR7	59 →	1980s	98 →	Elect	10 →	HSPF-6.0	46 →	R-30	43 →	R-0	6 →	Single
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
98	77 →	CR9	2 →	<1950	46 →	Gas	76 →	AFUE-80	1 →	R-11	57 →	R-7	93 →	Double
99	66 →	CR8	100 →	2000s	62 →	Elect	50 →	HSPF-7.7	49 →	R-30	45 →	R-15	53 →	Low-E
100	43 →	CR5	27 →	1960s	47 →	Gas	92 →	AFUE-96	62 →	R-30	51 →	R-7	86 →	Double

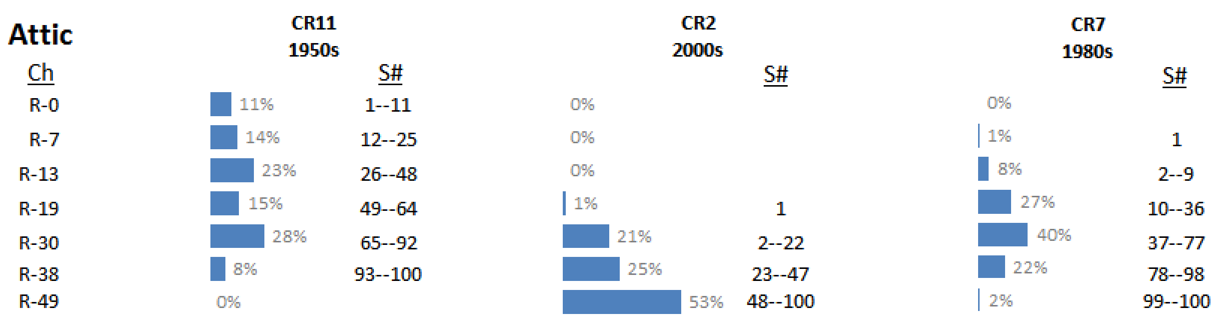

Figure 2. Sample number ranges for first three simulations in Table 2 (for a 100 simulation example) from probability distributions based on archetype dependencies.

How Many Simulations?

Beyond which building designs should be prioritized for simulation as described in the previous section, there is the question of the total number of simulations to be performed. In general, more simulations provide better coverage of the domain space and better accuracy/sensitivity of the model, but at the expense of longer runtime. The LHS approach attempts to select buildings to be simulated such that probability distributions are preserved. Therefore, the simulation distributions should match the probability distributions as closely as possible for archetype parameters.

As shown in Figure 3, simulations[4] associated with an archetype are used to create simulation distributions that best match the variant characteristic probability distributions (Figure 1). As an archetype specification becomes further defined (moving from left to right in Figure 3) the number of available simulations attributed to the archetype decreases.[5]

Using a finite number of simulations to match a probability distribution can lead to resolution problems, as seen in Figure 4.

For options with small probabilities (as can occur especially in distributions with many options), limited resolution can lead to some non-zero options with no simulations (see SEER 15 in Figure 4a) and other options with simulations that nearly double the appropriate probability (see EER 8.5 in Figure 4a).

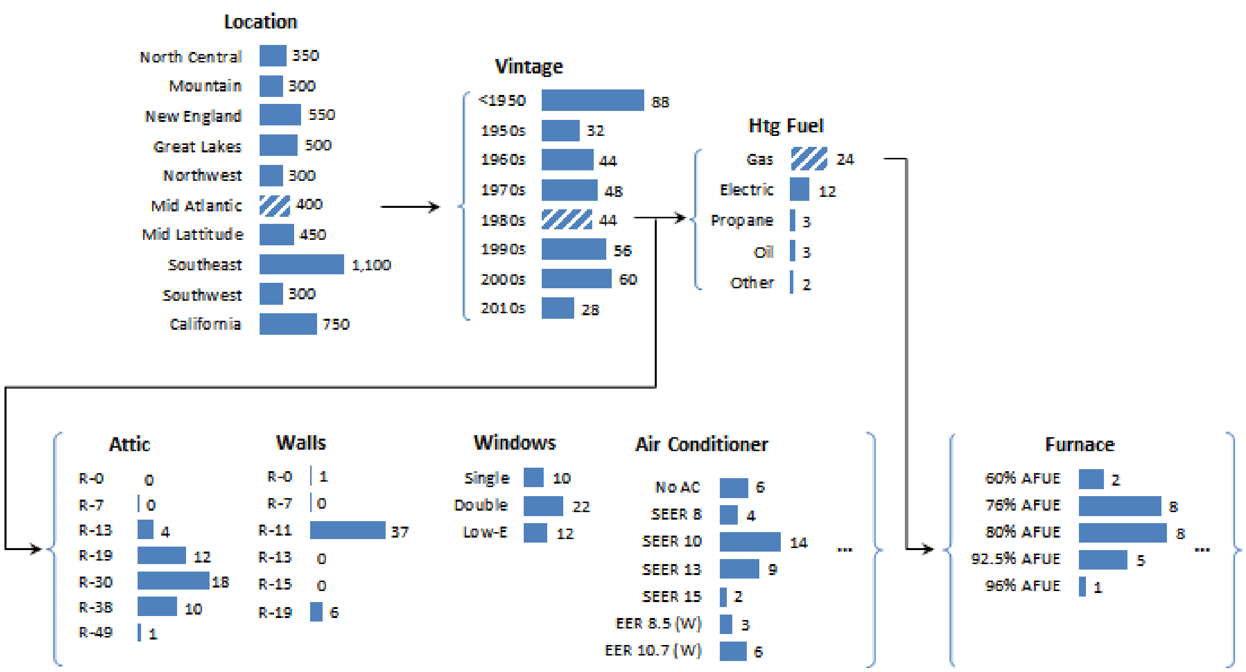

Figure 3. Number of simulations based on hierarchical conditional probability distributions.

[4] This example assumes 100,000 total simulations with ~20 TMYs per location (region). Therefore, ~5,000 simulations are available at the beginning of the hierarchical conditional probability tree.

[5] Moving left to right in Figure 3, the numbers of simulations decrease, because the hierarchical tree is dividing into a growing number of increasingly thin branches — with one path shown. Beyond the path shown, Location simulations are split off to cover multiple, not-shown Vintages, and Vintage simulations are split off to cover multiple, not-shown Heating Fuel Types.

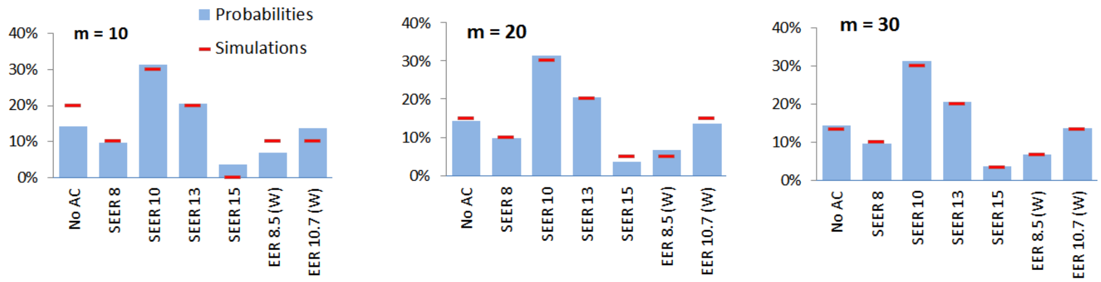

Figure 4. Air conditioner simulation distributions vs. probability distributions: a) 10 simulations, b) 20 simulations and c) 30 simulations.

The risk of inadequate resolution and significant (percent) discrepancies is highest for low-probability options within low-probability archetypes. Such cases are typically associated with relatively low house counts. However, if key to an analysis (for example, part of the target for a particular retrofit), such situations can be important. As an alternative to increasing the total number of simulations, one possibility is to focus the analysis and simulations specifically on the relevant part of the domain space.

The overall impact of the total number of simulations is difficult to accurately predict because the effects are variable across the domain space. An alternative to prediction is to monitor outputs of interest as additional simulations are run, for convergence toward a stable result (Figure 5) by tracking the minimum, maximum and average results. This approach has the advantage of finding the appropriate number of simulations depending on the specifics of different analyses.

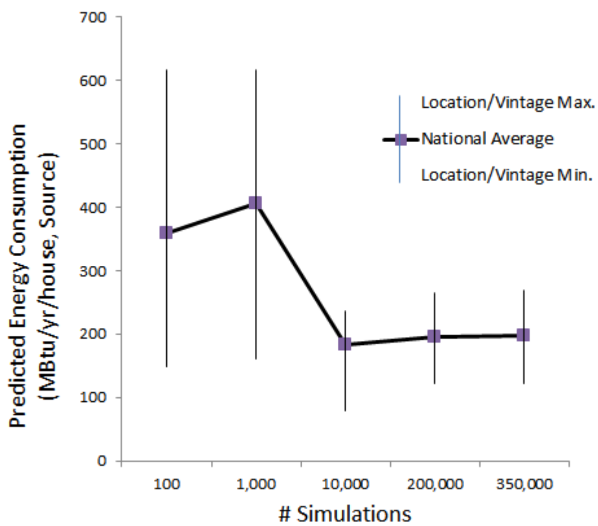

Figure 5. Convergence testing.

VALIDATION/CALIBRATION

For simulation of individual buildings, validation addresses accuracy of inputs, algorithms, and software implementation. For large-scale analysis, validation must also address archetype definitions, house-counts and dependencies. For large-scale analysis, validation involves comparing aggregated EnergyPlus model predictions to reference data (e.g., RECS consumption values).

The most basic validation is comparison of results aggregated at the highest level (e.g., national level). Comparisons of results at lower levels of aggregation (sliced by different archetype parameters, for example, as shown in Figure 6) can reveal the accuracy of the model under different circumstances and provide an indication of the model's likely usefulness for answering a range of analysis questions.

Beyond the physical characteristics captured in archetype variants, various occupant and operational factors are known to significantly affect building energy use. Simulations rely on assumed values; actual field values are not well known, nor how they vary with archetype parameters such as location and vintage. Such uncertainties may be a significant cause of lack of agreement between model results and reference data.

Calibration can be applied to 'true up' results to better match the reference data. Calibration may, in fact, be adjusting for a 'multitude of sins' in the model, and it is not clear to what degree the resulting calibrated model correctly preserves sensitivities for particular analysis questions. Therefore, calibration is a last resort. One prefers a model that validates well and requires as little calibration as possible.

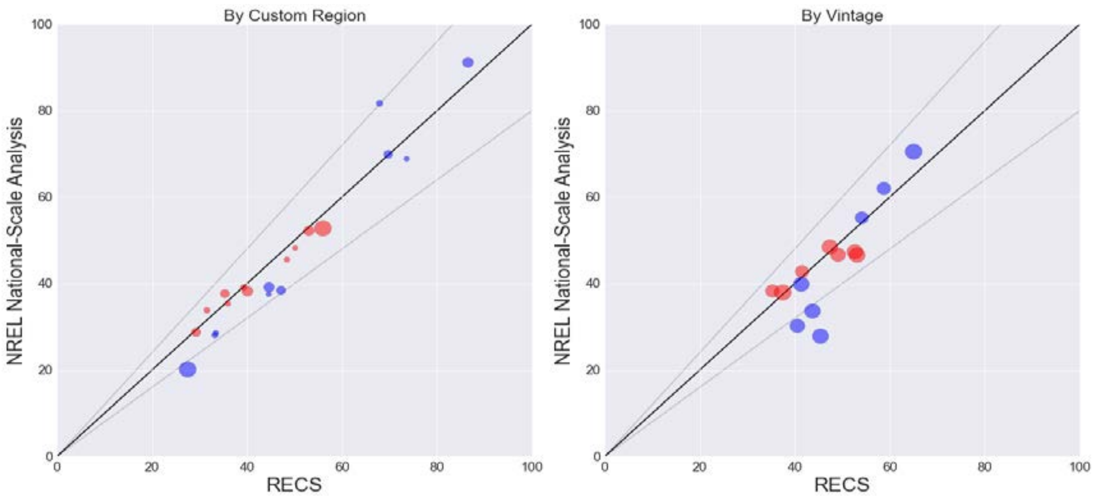

Figure 6. Consumption (source energy per house: 10^6 Btu/yr) modeled vs. RECS consumption data: by custom region and vintage; natural gas (blue) and electricity (red); bubble size indicates house count.

IMPLEMENTATION

This residential building stock modeling approach is planned for implementation on the Department of Energy's OpenStudio/EnergyPlus (www.openstudio.net and www.energyplus.net) whole-building simulation platform. The platform provides existing capabilities that include: 1) measures (scripts) that can quickly create/manipulate building models, 2) large-scale analysis using cloud computing, and 3) a framework for visualizing outputs. By leveraging these capabilities, implementation can be developed more quickly and be made available for other users/entities (e.g., utilities, local/state governments, manufacturers).

CONCLUSIONS

No single data source exists for the range of characteristics needed for residential building stock energy modeling. Data from multiple sources was used to develop a hierarchical structure of conditional probability tables that define components of a building depending on archetype parameters.

Significant diversity exists in the U.S. residential building stock, and building component characteristics are best represented by probability distributions depending on archetype parameters such as location and vintage.

Among alternative approaches for selecting representative buildings to be simulated, a Latin hypercube sampling (LHS) approach, using conditional probability distributions to capture building stock dependencies, is recommended based on a good balance between representativeness of the housing characteristics and number of simulations required. Depending on the granularity of the analysis, the total number of simulations can be scaled to achieve adequate sensitivity and avoid excessive uncertainty in the results.

FUTURE WORK

For our data set and modeling platform, we will investigate applicability of other statistical sampling approaches beyond LHS such as Monte Carlo sampling with low-discrepancy sequences including Sobol' sequences (Burhenne 2011).

ACKNOWLEDGMENT

This work was supported by the U.S. Department of Energy, Building Technologies Office and the Bonneville Power Administration.

REFERENCES

Burhenne, S., Jacob, D., and Henze, G. 2011. "Sampling Based on Sobol' Sequences for Monte Carlo Techniques Applied to Building Simulations," Proceedings of Building Simulation: 12th Conference of International Building Performance Simulation Association, Sydney.

Hopkins, A., Lekov, A., Lutz, J., Rosenquist, G., Gu, L. 2011. "Simulating a Nationally Representative Housing Sample Using EnergyPlus," LBNL-4420E, Energy Analysis Department, Lawrence Berkeley National Laboratory.

Huang, J., Hanford, J., Yang, F. 1999. "Residential Heating And Cooling Loads Component Analysis," LBNL-44636, Energy Analysis Department, Lawrence Berkeley National Laboratory.

Long, N., Ball, B., Fleming, K., Macumber, D. 2014. "Scaling Building Energy Modeling in the Cloud with OpenStudio," 2014 ASHRAE/IBPSA-USA Building Simulation Conference.

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., and Tarantola, S. 2008. Global Sensitivity Analysis. The Primer, John Wiley & Sons, Ltd.

Swan, L., Ugursal, I. 2009, "Modeling of End-Use Energy Consumption in the Residential Sector: A Review of Modeling Techniques," Renewable and Sustainable Energy Reviews.

Taylor, T., Ferdandez, N., Lucas, R. 2012. "Methodology for Evaluating Cost-Effectiveness of Residential Energy Code Changes," PNNL-21294, Pacific Northwest National Laboratory.