2018 Building Performance Modeling Conference and
SimBuild co-organized by ASHRAE and IBPSA-USA
Chicago, IL
September 26-28, 2018

AN INTERACTIVE VISUALIZATION TOOL FOR LARGE-SCALE BUILDING STOCK MODELING

Eric Wilson and Noel Merket
National Renewable Energy Laboratory, Golden, CO, USA

ABSTRACT

Recent advancements in data science and high-performance computing are making it easier to run millions of building simulations, but meaningful visualization of such large datasets remains a challenge. This paper presents a new tool developed to view the results of large-scale OpenStudio® simulations of national, regional, or local building stocks. The tool processes millions of simulations to calculate measure savings, utility bills, carbon emissions, primary energy, and cost-effectiveness metrics at a high geographic resolution. Interactive visualizations of the building characteristics, consumption, and measure savings data include proportional symbol maps and histogram plots and can be filtered by any building characteristic.

INTRODUCTION

Background

ResStock is a bottom-up, engineering-based, residential building stock model that uses thousands to hundreds-of-thousands of representative building energy models to evaluate the energy savings potential of various energy efficiency upgrades across national, regional, and local building stocks (Wilson et al. 2017). These representative archetypes are based on statistical analysis of housing stock characteristics to capture the wide variability in construction types, equipment and appliance configurations, climate conditions, and other factors that influence building energy use. The resulting model represents the stock with much higher granularity and specificity than other approaches in the literature (Wilson et al. 2016).

Figure 1 shows a technical diagram of the free and open-source ResStock workflow. The workflow leverages the U.S. Department of Energy's OpenStudio software development kit (Roth, Goldwasser, and Parker 2016) and the EnergyPlus™ whole-building energy modeling engine (Crawley et al. 2001). OpenStudio Measures (scripts for creating and modifying individual building models or output) are orchestrated via the OpenStudio Parametric Analysis Tool (PAT), which can be used to automate large-scale simulation analysis of building stocks or portfolios of building designs. ResStock also leverages PAT and OpenStudio-Server to deploy the thousands of simulations on Amazon EC2 cloud computing (Macumber, Ball, and Long 2014).

Problem statement

ResStock analysis of the approximately 80 million single-family detached homes in the United States has typically used a set of 350,000 building archetypes, or approximately one for every 230 homes in the real world. This set of representative models, or an applicable subset, is simulated for every energy efficiency measure, package of measures, or reference scenario analyzed, leading to more than 20 million individual simulations for a typical analysis. Each of these simulations has approximately 100 building characteristic parameters associated with it (often abstracted from EnergyPlus/OpenStudio input parameters such as vintage, location, foundation type, refrigerator type, and occupant use level). Simulation results include around 30 different metrics, including annual energy use disaggregated by fuel type and end-use category, as well as calculated metrics such as total primary (source) energy use, utility bills, and carbon emissions. Upgrade scenarios have double the number of metrics after savings deltas are calculated. This means that a typical analysis of the U.S. single-family detached housing stock will have 4 billion datapoints. If hourly time series data is included for each simulation, the number of datapoints increases to more than 5 trillion.

This amount of data quickly becomes unmanageable to work with and visualize for an analyst using a standard computer. To make the analysis results accessible to decision makers, a tool is needed that makes it possible to visualize the data in meaningful ways. Most research on visualizing large-scale simulation results has focused on urban-scale modeling commonly using 3D maps of buildings (Reinhart and Cerezo Davila 2016; Giovannini et al. 2014; Fonseca and Schlueter 2015). There are geospatial platforms for visualizing national and regional energy information (NREL 2018), but these are not configured to work directly with OpenStudio output to facilitate visualization of large-scale building energy simulation results. This paper presents a new data viewer tool designed to meet these needs[1].

[1]The Data Viewer tool, with an example dataset, can be found on the

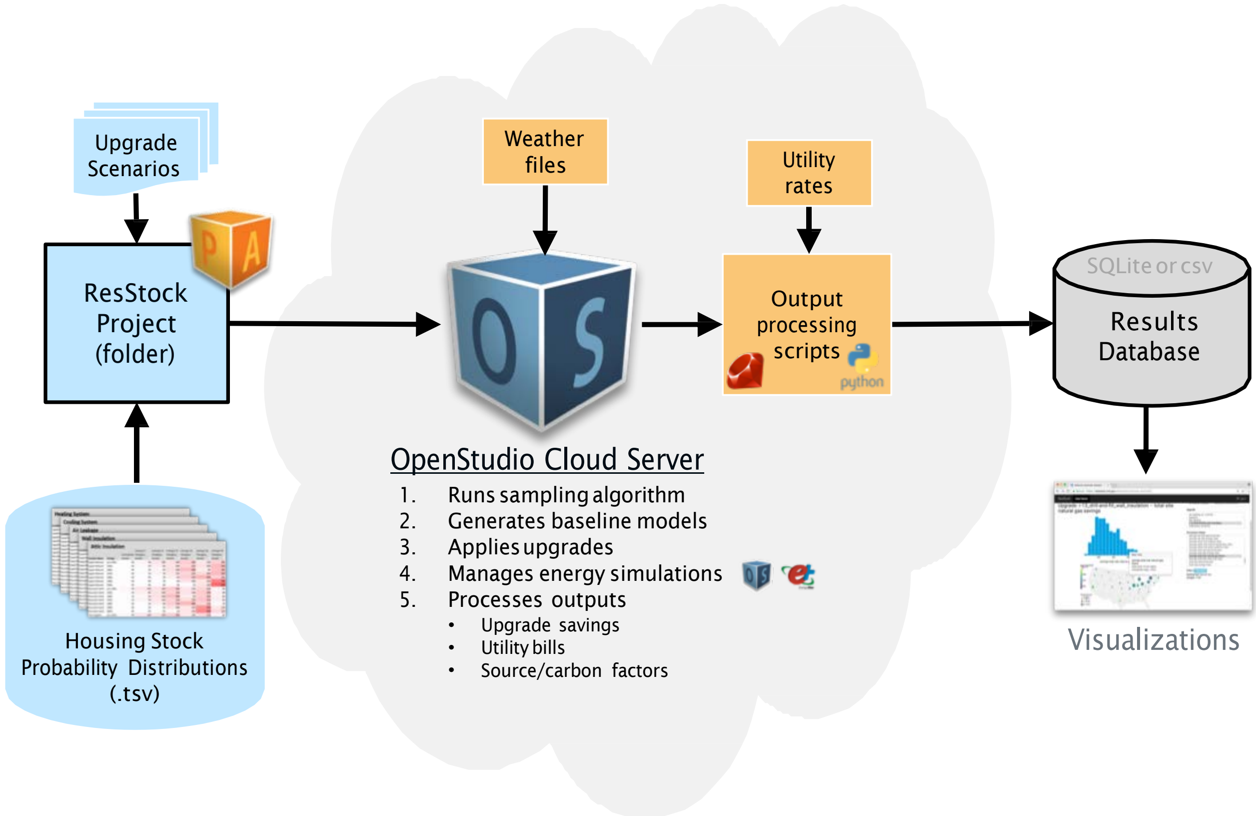


Figure 1: The ResStock workflow involves a PAT project with OpenStudio measures that 1) sample parameters from a set of probability distribution resource files, 2) generate OpenStudio/EnergyPlus input files with the sampled parameters, 3) apply upgrade scenarios to the representative models, and 4) post-process outputs. The analysis results are downloaded to a data file that can be uploaded to the ResStock website for visualization.

METHODOLOGY

System architecture

The goal of the data viewer is to provide meaningful and flexible visualization of large-scale simulation results generated by the ResStock tool and turn the data into actionable insights into residential energy consumption. Furthermore, the visualizations should be interactive and easily share-able with decision makers. Therefore, the data viewer is designed as a web application where simulation results are uploaded through a web interface, processed on the server, and then visualized using a web browser. Making this a web application, rather than a user-installed desktop application, allows result visualizations to be shared easily and removes the burden of installation and updating for end users.

The web application is built using the Flask microframework for Python in conjunction with an NGINX reverse proxy server to serve static resources. To load data for visualization, a user uploads the simulation output file from ResStock into a web interface. Data processing is performed asynchronously with the Pandas data analysis library (McKinney 2010), Dask parallel data processing library (Rocklin 2015), and NumPy (van der Walt, Colbert, and Varoquaux 2011) before it is stored in a PostgreSQL database for later retrieval. These services are coordinated and deployed on Amazon Web Services using Docker. Load balancers monitor server load and launch new server instances of the application when traffic is heavy, shutting down instances when traffic is light. See Figure 2 for a diagram of the previously mentioned web application components; the arrows indicate data flow.

The size of ResStock results can be prohibitive for efficient in-memory data processing. A novel approach used in this system is to process, aggregate, and store results in batches using a cluster of worker nodes implemented with the Celery library. This makes the system scalable to very large datasets that could not be handled on a single desktop computer and to multiple concurrent users with data-intensive tasks. The most common queries are pre-computed and stored in the database for a better user ex-

ResStock website: https://resstock.nrel.gov/.

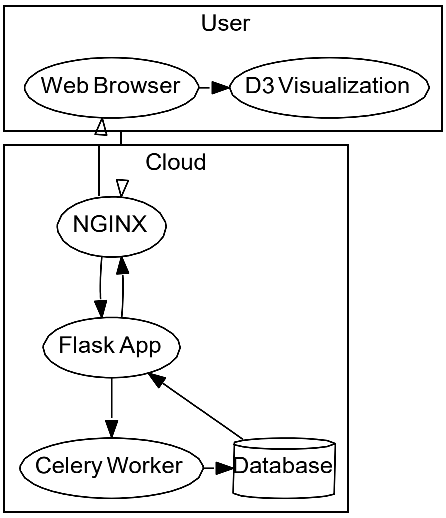


Figure 2: System architecture diagram

perience.

When a user requests the visualization through their web browser, a call is made to the server where the precomputed results are retrieved or an appropriate dataset is filtered and aggregated before returning results. Those results are then dynamically rendered into interactive scalable vector graphics (SVG) in the browser using the Data-Driven Documents (D3) library (Bostock, Ogievetsky, and Heer 2011).

Visualization types

Different visualization types are suited for answering different analysis questions. For audiences interested in national and regional results (e.g., federal policymakers, regional energy efficiency organizations, manufacturers), state-by-state maps of aggregated absolute savings and relative percentage savings illustrate the overall potential resulting from efficiency improvements, as well as where in the country the potential exists.

For the initial development of this visualization tool, we chose state-by-state *proportional symbol* maps, where the size of the circles indicate total energy use or savings and the color indicates average household energy use or savings. While state choropleth maps serve well as visual "look-up tables" for individual state information, they can lead to misinterpretation because large areas with low population or energy usage can dominate visually. Compare Figures 3a and 3b—the colors assigned to each state are identical; however, Figure 3b adds the absolute savings in each state as a second dimension, which scales the visual importance of each state based on its contribution to the total national savings potential.

One disadvantage of proportional symbol maps is that population differences between states can overwhelm other trends that affect state-to-state differences in energy use and savings potential. This is referred to as the modifiable areal unit problem, which affects choropleth maps as well (Holt et al. 1996).

If it is necessary to compare a metric between geographic units such as states, an approach that equalizes visual importance is a *tilemap*, where each state is represented as an equally-sized tile and color is used to encode the population normalized quantity, allowing a comparison similar to that of a choropleth map. The advantage is that all states (or other geographic units) carry the same visual weight. Figure 5 shows a state-by-state view of energy savings potential as a percentage, but fails to show the total savings potential.

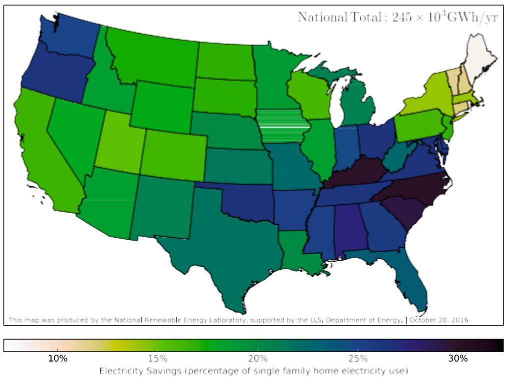
One approach to avoid sensitivity to geographic aggregation is dasymetric maps that use equal-area gridcells instead of geopolitical boundaries (Petrov 2012). Dasymetric maps can be prone to the same misinterpretation as choropleth maps because low-population areas can still dominate the visual hierarchy. Figure 4 illustrates this effect using data on primary heating fuel types from the 2008–2012 American Community Survey (Manson et al. 2017), mapped from census tracts to 10-km square gridcells covering the entire United States using the process described in Appendix F of Wilson et al. 2017. While the dasymetric maps in Figures 4a and 4b are useful for understanding frequency of fuel types at a high geographic granularity, they could be misinterpreted as showing that natural gas and propane have similar frequency at a national or regional level, because the visual prominence or coverage of the country's geographic area is relatively comparable. The state-aggregated proportional symbol maps in Figures 4c and 4d tell the real story: Propane, while common in low-population rural areas, is a minority fuel type in all states and is outnumbered nationally 7.5-to-1 by homes heated by natural gas.

Ultimately, the state-aggregated proportional symbol maps were chosen as the initial map type for presenting national and regional results in the data viewer as they limit opportunities for misinterpretation. A similar approach of proportional doughnut maps was selected for displaying the state-by-state breakdown of various building characteristics (see Figure 10).

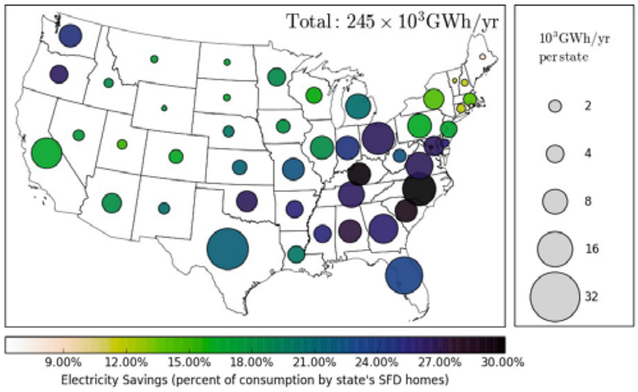
Histogram plots accompany the geospatial maps to inform viewers about the distribution of per-household energy use and savings values nationally or regionally (see Figures 7–8). Similarly, bar charts accompany the proportional doughnut maps to show the breakdown of building characteristics nationally or regionally (see Figure 10).

DISCUSSION AND RESULT ANALYSIS

Figure 6 shows annual source energy use by state as a proportional symbol map. As previously described, the size

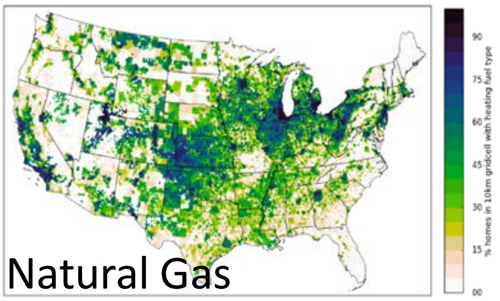


(a) Choropleth map displaying percentage electricity savings by state

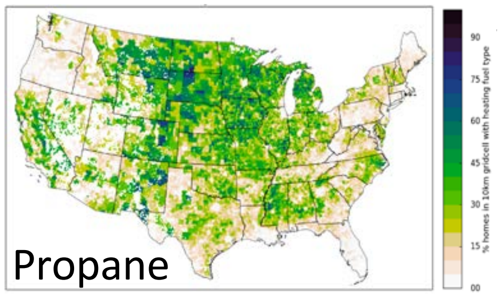


(b) Proportional symbol map displaying absolute (circle area) and percentage (circle color) electricity savings by state

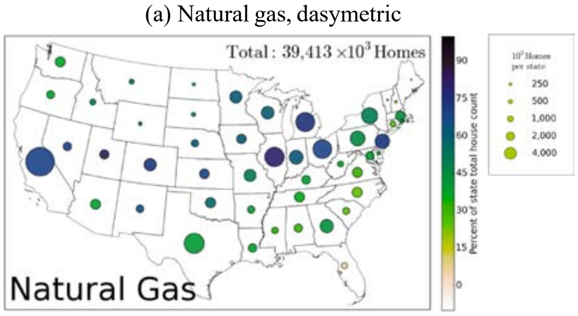
Figure 3: Comparison between choropleth and proportional symbol map for visualizing states' savings potential



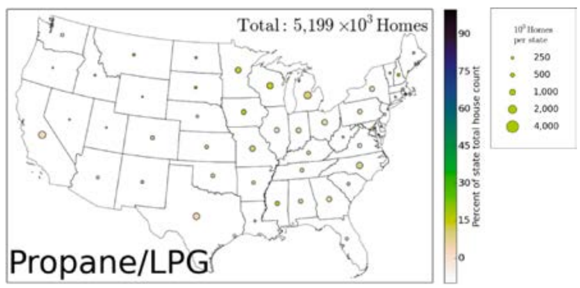
(a) Natural gas, dasymetric



(b) Propane, dasymetric



(c) Natural gas, proportional symbol



(d) Propane, proportional symbol

Figure 4: Percentage of single-family detached homes using natural gas (a, c) and propane (b, d) as their primary heating fuel; comparison of dasymetric 10-km gridcell maps (a, b) to proportional symbol maps aggregated by state (c, d)

of the circles indicates total energy use while the color of the circles indicates average household energy use. As expected, the states with the greatest population have the greatest overall energy use while states with high heating and/or cooling loads have higher per household energy use. In Figure 7 we see the same distribution of nationwide source energy use as a histogram. This illustrates the long tail of energy use among the high energy users as well as the mean. Any combination of end use and fuel type can be selected and viewed similarly.

In addition to modeling the existing housing stock, a key feature of ResStock is the ability to apply configurable upgrade scenarios to the underlying building models to ascertain energy savings potential. Figure 8a shows the natural gas savings potential of performing a drill-and-fill wall insulation upgrade to R-13 for homes with empty wall cavities. There are about 14 million homes with zero natural gas fuel savings, possibly because those homes have another primary heating fuel. To verify this, the results can be filtered to only show houses that have natural gas as their primary heating fuel, as shown in Figure 8b. This indeed shows that most of the homes with zero savings for that measure were homes not heated by natural gas.

Energy use plots can be filtered by any building characteristic, including more than one at a time, and Figure 9 shows an example of this in a screen capture of the data viewer application. In this case, we show the natural gas savings potential for the same R-13 wall drill and fill measure and filter it to homes that have natural gas heating and were built before 1960. Also, hovering over a given state displays the specific savings for that state in total and as a household average. This capability allows further investigation of which home features will yield the greatest potential energy savings, helping target programs' efforts.

Besides visualizing the energy and savings outputs from the model, a user can view the distributions of archetype characteristics that represent the building stock. Additionally, by clicking on a state, the view zooms into the results for just that state. Figure 10 shows a screen capture of a zoomed view of New York's heating fuel distribution. The bar graph serves as a legend for the data in the map and updates to represent the currently zoomed view. Mouseovers reveal even more detailed data, including the number of homes and the percent that are estimated to have each characteristic, such as that 27% of homes in New York have a primary heating fuel of fuel oil.

CONCLUSION

This paper presents a new interactive visualization tool for large-scale building stock modeling. The tool is designed to handle billions of datapoints in a flexible manner, striking a balance between precomputing common queries and executing less common queries on the fly.

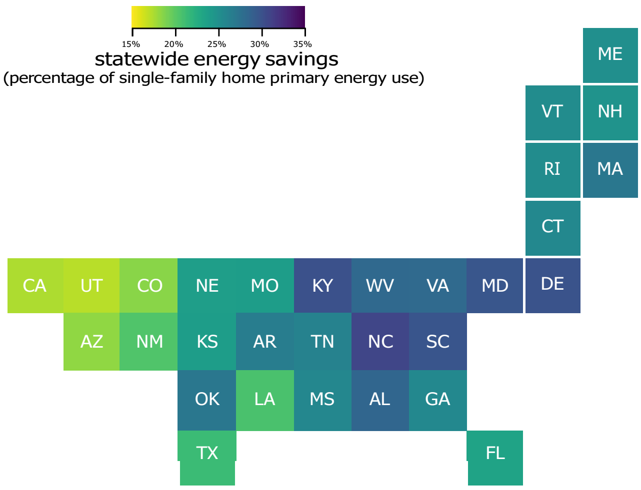


Figure 5: Tile map of energy savings potential percentage by state

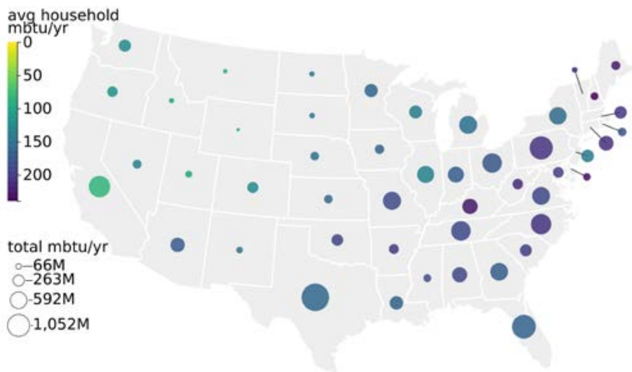


Figure 6: Household average and total annual source energy use by state

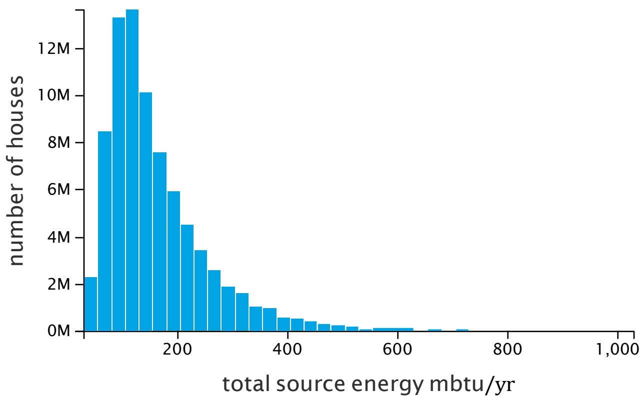
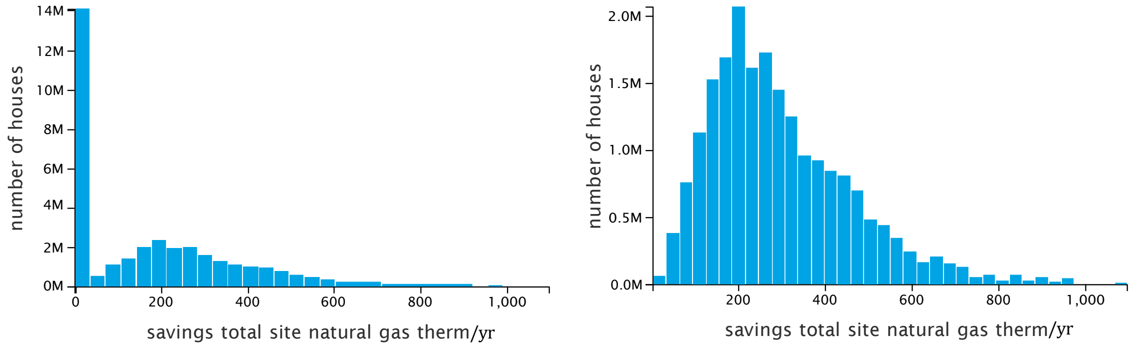


Figure 7: National annual source energy use distribution



(a) all homes (b) homes with a primary heating fuel of natural gas

Figure 8: Nationwide annual natural gas savings potential with R-13 drill-and-fill wall insulation (note different scales)

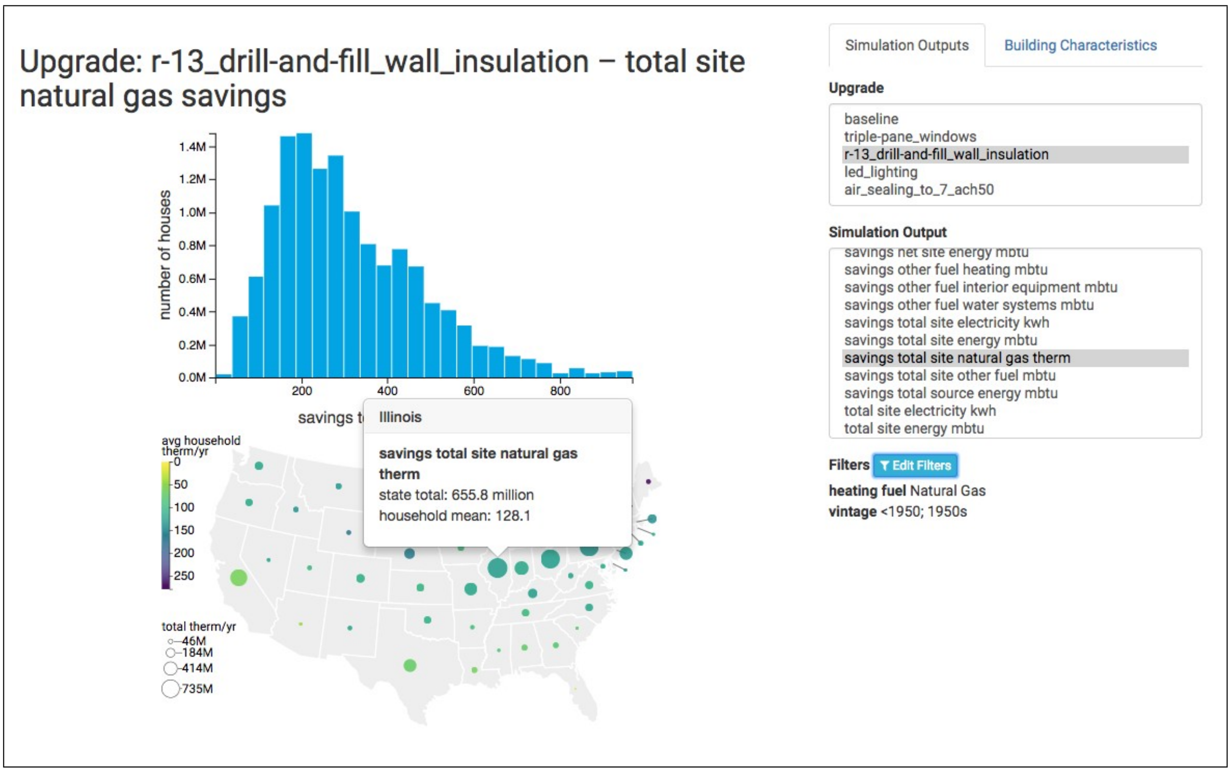


Figure 9: ResStock Data Viewer screen capture of interface showing the total annual natural gas savings for a R-13 wall insulation retrofit filtered to houses with natural gas heating that were built before 1960

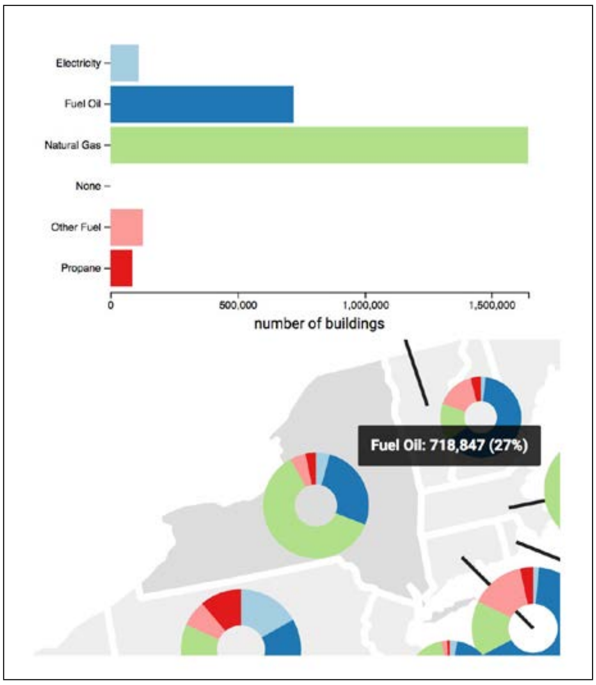


Figure 10: Heating fuel distribution in the state of New York

The initial implementation of the tool uses proportional symbol maps to visualize national and state totals and averages of energy consumption and savings data for various fuels and end uses. Histogram plots are used to visualize how those consumption and savings values are distributed. The breakdown of building stock characteristics is visualized using proportional doughnut maps and bar charts. A unique, infinitely flexible filtering system allows users to filter the maps and plots of characteristics, consumption, or savings by any building parameter (e.g., building type, vintage, heating fuel type) and the visualizations update to reflect the query in real time. This new visualization capability is free and publicly available; uploaded datasets can be shared publicly or privately.

Future Work

Future enhancements that could be made to the data viewer tool include:

- **Numeric value filtering** – Savings potential and building parameter distributions could be filtered by numeric values in addition to building parameters. For example, only the savings that meet a cost-effectiveness criteria (e.g., net present value > 0) could be displayed, and the distribution of building parameters could be displayed for the subset of buildings where an upgrade is cost-effective.
- **Additional visualizations** – There are many possibilities beyond the proportional symbol map and histogram to visualize this kind of data. One that is particularly promising is the violin plot, which allows for visualization of distributions of real valued numbers broken out by category. It serves a similar purpose to a box plot; however, by applying kernel density estimation to the distribution, it allows a more complete picture of each category's distribution. This would allow visualization of energy use or savings distributions separated by building characteristics. See Figure 11 for an example showing the total site energy as a function of bins of conditioned floor area.
- **Greater geographic granularity** – ResStock is currently being enhanced with the ability to disaggregate results by county, as well as various ranges of household income. County-level maps would provide many additional insights into how energy savings potential varies within a state, accounting for differences in building stock between urban, suburban, and rural counties.
- **Time series visualization** – For grid reliability and higher penetrations of renewables, the question of *when* energy is being used is often more important than *how much*. ResStock produces hourly time series of energy use for each building simulated. Using all the time series data would allow visualizations of load profiles and the effects of certain efficiency measures on them. Currently, the data viewer only uses the aggregated annual results for each building. The scale of the time series data is orders of magnitude larger than the annual data currently in use.

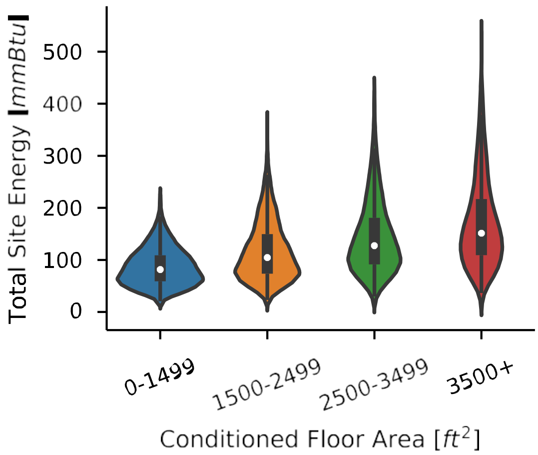


Figure 11: A violin plot of annual site energy use by house size

ACKNOWLEDGMENT

This work was authored by Alliance for Sustainable Energy, LLC, the manager and operator of the National Renewable Energy Laboratory for the U.S. Department of Energy (DOE) under Contract No. DE-AC36-08GO28308. Funding provided by U.S. Department of Energy Office of Energy Efficiency and Renewable Energy Building Technologies Office. The views expressed in the article 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.

REFERENCES

Bostock, Michael, Vadim Ogievetsky, and Jeffrey Heer. 2011. "D3: Data-Driven Documents." *IEEE Trans. Visualization & Comp. Graphics (Proc. InfoVis)*.

Crawley, Drury B., Linda K. Lawrie, Frederick C. Winkelmann, W. F. Buhl, Y. Joe Huang, Curtis O. Pedersen, Richard K. Strand, Richard J. Liesen, Daniel E. Fisher, Michael J. Witte, and Jason Glazer. 2001. "EnergyPlus: creating a new-generation building energy simulation program." *Energy and Buildings*.

Fonseca, J.A., and A. Schlueter. 2015. "Integrated model for characterization of spatiotemporal building energy consumption patterns in neighborhoods and city districts." *Applied Energy* 142:247–265.

Giovannini, Luca, Stefano Pezzi, Umberto di Staso, Federico Prandi, and Raffaele de Amicis. 2014. "Large-Scale Assessment and Visualization of the Energy Performance of Buildings with Ecomaps." *Proceedings of 3rd International Conference on Data Management Technologies and Applications*, DATA 2014. Portugal: SCITEPRESS - Science and Technology Publications, Lda, 170–177.

Holt, D., D. G. Steel, M. Tranmer, and N. Wrigley. 1996. "Aggregation and Ecological Effects in Geographically Based Data." *Geographical Analysis* 28 (3): 244–261.

Macumber, Daniel L., Brian L. Ball, and Nicholas L. Long. 2014, September. "Graphical Tool for Cloud-Based Building Energy Simulation." Technical Report NREL/CP-5500-61501, Atlanta, GA: American Society of Heating, Refrigerating, and Air-Conditioning Engineers (ASHRAE).

Manson, Steven, Jonathan Schroeder, David Van Riper, and Steven Ruggles. 2017. IPUMS National Historical Geographic Information System: Version 12.0 [Database].

McKinney, Wes. 2010. "Data Structures for Statistical Computing in Python." Edited by Stéfan van der Walt and Jarrod Millman, *Proceedings of the 9th Python in Science Conference*. 51 – 56.

National Renewable Energy Laboratory (NREL). 2018 (accessed February 9, 2018). Data Visualization and Geospatial Tools. `https://www.nrel.gov/gis/tools.html`.

Petrov, Andrey. 2012. "One Hundred Years of Dasymetric Mapping: Back to the Origin." *The Cartographic Journal* 49 (3): 256–264.

Reinhart, Christoph F., and Carlos Cerezo Davila. 2016. "Urban building energy modeling A review of a nascent field." *Building and Environment* 97 (February): 196–202.

Rocklin, Matthew. 2015. "Dask: Parallel Computation with Blocked algorithms and Task Scheduling." Edited by Kathryn Huff and James Bergstra, *Proceedings of the 14th Python in Science Conference*. 130 – 136.

Roth, Amir, David Goldwasser, and Andrew Parker. 2016. "There's a measure for that!" *Energy and Buildings* 117 (April): 321–331.

van der Walt, S., S. C. Colbert, and G. Varoquaux. 2011. "The NumPy Array: A Structure for Efficient Numerical Computation." *Computing in Science Engineering* 13 (2): 22–30 (March).

Wilson, Eric, Craig Christensen, Scott Horowitz, and Henry Horsey. 2016. "A High-Granularity Approach to Modeling Energy Consumption and Savings Potential in the U.S. Residential Building Stock." *Proceedings of SimBuild* 6, no. 1.

Wilson, Eric J., Craig B. Christensen, Scott G. Horowitz, Joseph J. Robertson, and Jeffrey B. Maguire. 2017, December. "Energy Efficiency Potential in the U.S. Single-Family Housing Stock." Technical Report NREL/TP-5500-68670, National Renewable Energy Lab. (NREL), Golden, CO (United States). DOI: 10.2172/1414819.