

8-28-2012

Visual Analytics for Roof Savings Calculator Ensembles

Chad Jones

University of California, Davis, cejones@ucdavis.edu

Joshua New

Oak Ridge National Laboratory, newjr@ornl.gov

Jibonananda Sanyal

Oak Ridge National Laboratory, sanyalj@ornl.gov

Kwan-Liu Ma

University of California, Davis, ma@cs.ucdavis.edu

Follow this and additional works at: http://aisel.aisnet.org/sprouts_all

Recommended Citation

Jones, Chad; New, Joshua; Sanyal, Jibonananda; and Ma, Kwan-Liu, "Visual Analytics for Roof Savings Calculator Ensembles" (2012).
All Sprouts Content. 492.

http://aisel.aisnet.org/sprouts_all/492

This material is brought to you by the Sprouts at AIS Electronic Library (AISeL). It has been accepted for inclusion in All Sprouts Content by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

Visual Analytics for Roof Savings Calculator Ensembles

Chad Jones

University of California, Davis, USA

Joshua New

Oak Ridge National Laboratory, USA

Jibonananda Sanyal

Oak Ridge National Laboratory, USA

Kwan-Liu Ma

University of California, Davis, USA

Abstract

The Roof Savings Calculator (RSC) has been deployed for DOE as an industry-consensus, web-based tool for easily running complex building energy simulations. These simulations allow both homeowners and experts to determine building-specific cost and energy savings for modern roof and attic technologies. Using a database of over 3 million RSC simulations for different combinations of parameters, we have built a visual analytics tool to assist in the exploration and identification of features in the data. Since the database contains multiple variables, both categorical and continuous, we employ a coordinated multi-view approach that allows coordinated feature exploration through multiple visualizations at once. The main component of our system, a parallel coordinates view, has been adapted to handle large-scale, mixed data types as are found in RSC simulations. Other visualizations include map coordinated plots, high dynamic range (HDR) line plot rendering, and an intuitive user interface. We demonstrate these techniques with several use cases that have helped identify software and parametric simulation issues.

Keywords: Visualization, Visual Analytics, Parallel Coordinate Plots, Linked-View Visualization, Computer Graphics Applications, Simulation and Modeling Validation and Analysis, Physical Sciences and Engineering

Permanent URL: <http://sprouts.aisnet.org/12-5>

Copyright: [Creative Commons Attribution-Noncommercial-No Derivative Works License](#)

Reference: Jones, C., New, J., Sanyal, J., Ma, K. (2012). "Visual Analytics for Roof Savings Calculator Ensembles," Proceedings > Proceedings of Energy Informatics . *Sprouts: Working Papers on Information Systems*, 12(5). <http://sprouts.aisnet.org/12-5>

INTRODUCTION

The U.S. Department of Energy (DOE) aims to reduce U.S. energy consumption and greenhouse gas emissions by enabling energy efficient technologies to enter the market at speed and scale (DOE, 2012). Building envelope and equipment experts use parametric simulation sweeps of building variables in an effort to identify optimal design of devices, select cost-effective construction, and determine the best measures for buildings. Identifying promising technologies can lead to tax incentive and loan programs designed to encourage the growth of energy saving devices, materials, and practices. However, the impact of this large volume of data is limited unless analysis tools are made available to allow researchers and administrators to dynamically formulate and fine-tune hypotheses relevant to a question under study. Using data generated from the Roof Savings Calculator (RSC), published as a free web-based tool, the visualization system described here is used to explore the multivariate database.

RSC simulation, like any other simulation model, is a numerical approximation of reality. Various uncertainties are inherent in the modeling process. These may stem from model parameters, numerical solvers, timestep of integration, and a variety of other factors. To capture a subset of the modeling uncertainties, ensembles of simulations are defined by systematic perturbations to the set of simulation input variables. The large multitude of variables and the multivariate nature of the resulting data makes analysis through thorough sampling and visualization a computationally intensive task.

The use of visual analytics helps in identifying potential flaws in the simulation models, verify existing hypothesis and uncover interesting features in the data. From a visualization standpoint, the RSC data presents challenges for producing a rich visualization system. First, the size and scale of the data is a major issue. The RSC database houses over 3 million simulation runs, each with over 50 input and output variables. Second, the variables themselves include categorical, numerical, and geospatial components. The visualization system must be able to handle this wide array of types such that filtering and feature specification can be done in any view. Thus, we present RSC Explorer, which combines various views into an interactive tool and is the result of a collaborative, cross-disciplinary partnership between building energy experts and computer scientists. In addition to providing a tool for building energy specialists to explore their data, we present modifications to parallel coordinate plots (PCP) that combines the advantages of categorical and numerical visualization techniques and applies them to a large scale problem. Using RSC Explorer, we examine groups of outliers that represent problems with the simulation code, verify existing data patterns, and discover potentially novel patterns in the simulation data.

The rest of the paper is organized as follows: we first discuss related work, followed by a section on the RSC engine and the data being used here. We then provide a detailed description of the RSC Explorer tool followed by an evaluation of the tool by experts and a discussion of the results. We finally summarize the contributions in a concluding section.

RELATED WORK

The RSC combines *DOE-2.1E* for whole-building energy simulation with *AtticSim* for simulation of modern roof and attic technologies. Both codes have been validated through field experiments and *AtticSim* is part of ASTM standard C 1340. This combined source code passes thermal behavior and other physical properties through the boundary of the attic floor each timestep of the simulation and acts as the engine of RSC. The web interface provides a simplified interface with 26 possible inputs but defaults everything to national averages with the exception of nearest U.S. location of 239 possible weather sources. This allows comparison of a base case and cool roof comparison to determine monthly changes in heating and cooling energy aggregated to yearly cost. The RSC is available at www.roofcalc.com; for more information, the interested reader is referred to (New, Miller, Desjarlais, Huang, & Erdem, 2011).

The building energy simulations provide a large multivariate dataset to explore. For the purpose of visualizing multiple variables, the system takes advantage of parallel coordinates. Parallel coordinates were introduced by Inselberg (Inselberg, 1985; Inselberg & Dimsdale, 1990) and have been widely used for visualizing relationships among multiple variables. The data is represented by drawing a vertical or horizontal axis for each data dimension, and then for each data element, a polyline is drawn to traverse each axis according to its value for that dimension. Many of the advantages, such as conveying patterns, outliers, and features, are lost when the parallel coordinates are applied to large scientific data.

In order to relieve the limitation on scalability, Fua et. al (Fua, Ward, & Rundensteiner, 1999) conveyed aggregation information through the use of hierarchical clustering with tools for drilling to individual items within the hierarchy. Novotny and Hauser (Novotny & Hauser, 2006) addressed the issue of outlier preservation in context visualization of parallel coordinates by handling outliers and trends separately. The outliers and context are calculated by using a binned data representation and then mapped to a 2D histogram. Using this technique, they have successfully visualized data with up to 3 million elements. However, the visualizations described above are made with support exclusively for numerical data. Special care must be given when integrating and visualizing RSC data which include both numerical and categorical data. Johansson and Johansson used a multiple view approach to visualizing numerical and categorical data (Johansson, 2009; Johansson & Johansson, 2010). However, their work uses quantification on categorical data to allow for numerical approaches to be used.

Drawing from the previous success of linking multiple views to explore data, the system merges new techniques and existing information visualizations into an interactive system. The interface focuses on the exploration of large building ensembles based on multi-dimensional inquiry. The drawbacks of individual visualizations are overcome by combining the features of several techniques together.

SIMULATION MODEL AND DATA

The scientists at Oak Ridge National Laboratory (ORNL) are responsible for RSC and are interested in further verifying, validating, and understanding the underlying simulation models. The RSC system diagram shows workflow with simplified input/output pages can be seen in Figure 1

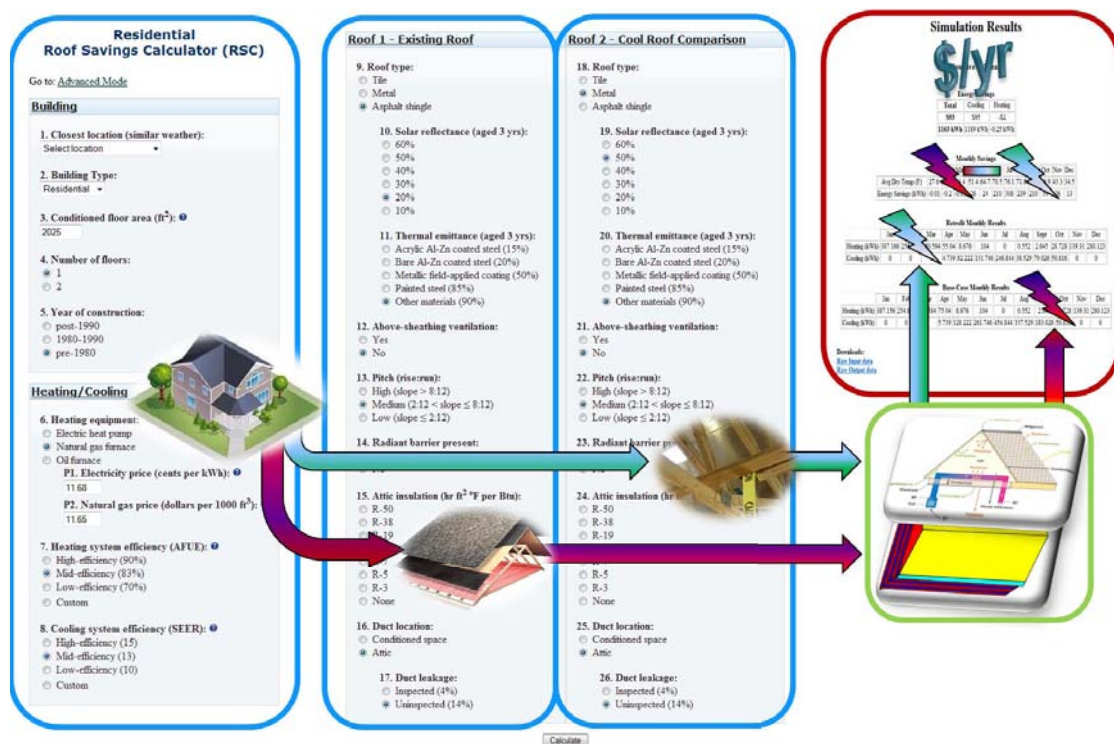


Figure 1: Roof Savings Calculator (RSC) website system diagram for www.roofcalc.com showing workflow, simulation inputs, RSC engine, and output page.

with each job request constituting 2 simulations, a baseline and a comparison. As can be seen, the simulation data used in this study derive from inputs 1-17 which can be broken up into three categories: building details with location, HVAC efficiency with utility pricing, and roof/attic properties. For more information, see the pop-up tool tips at roofcalc.com or (New et al., 2011) for more details. An advanced-mode variable not shown here is window-to-wall ratio for a total of 18 simulation inputs. Outputs are monthly electric cooling energy (kWh), electric heating energy (kWh), and natural gas heating energy (BTUs), aggregated from hourly simulation timesteps, for a total of 36 outputs. These 54 fields of data per simulation are stored in a MySQL database containing over 3 million simulations. A future version of the RSC website will use this database as a cache to provide immediate results for simulations that have already been run by anyone in the world while dynamically running and adding new simulations to the database as they are requested (or contributed through parametric sweeps).

RSC EXPLORER

RSC ensemble simulations produce data that is large and multivariate across both numerical and quantitative data. A number of visualization techniques are employed in the multiple linked view visual analytics system which we call RSC Explorer. It was implemented using the C++ programming language and uses the Qt, OpenGL, and Boost libraries. By using different views, the

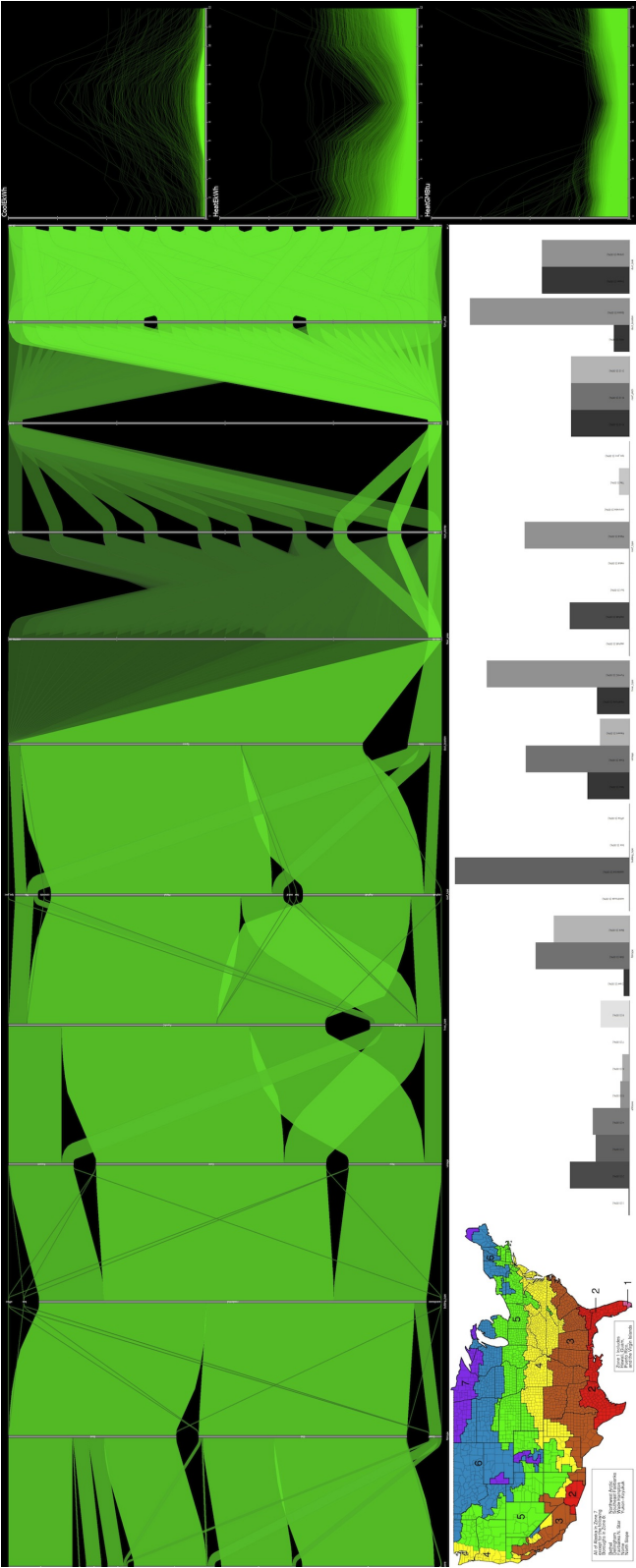


Figure 2: The image above provides an overview of the RSC Explorer. The main parallel coordinates plot (PCP) window contains both categorical and numerical data dimensions, and uses rounded polygons to provide accurate sizing and continuity of data flow through multiple dimensions. Each dimensional category can be seen as a bar chart along the bottom, next to a map of clickable ASHRAE climate zones on the left. On the right, a series of time plots using high dynamic range rendering show energy usage for heating and cooling.

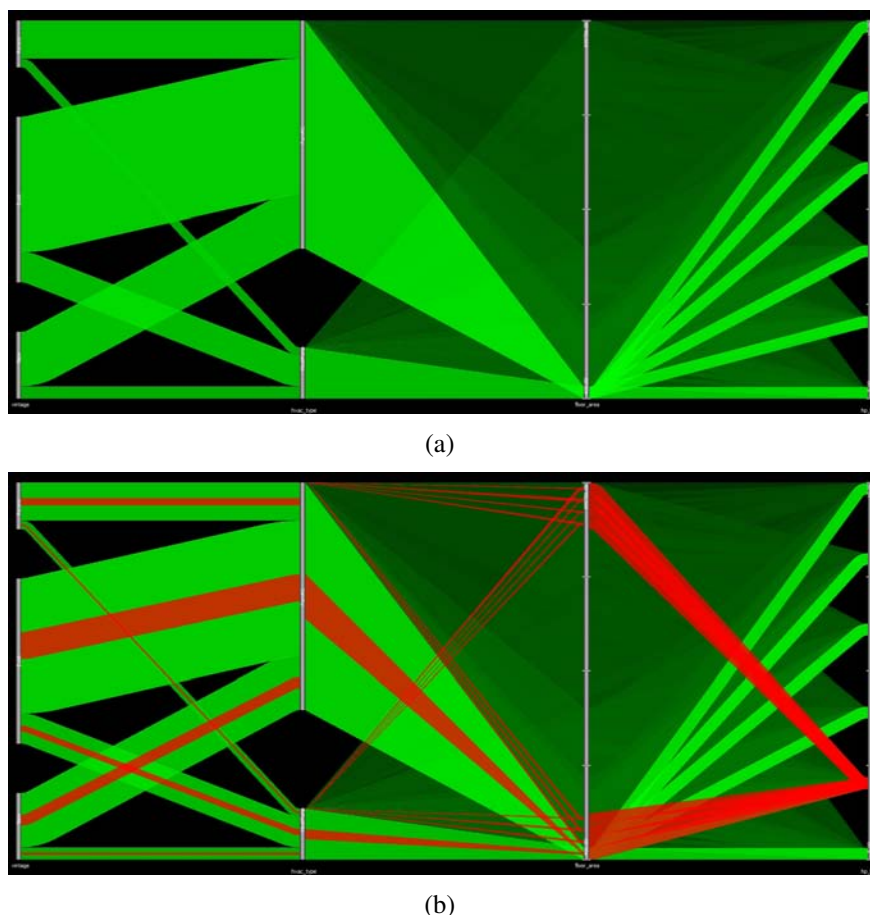


Figure 3: By integrating rendering techniques together mixed data can be visualized effectively in one plot (a). When items are selected, the red focus bins (red) are rendered on top of the green context bins (b).

visualization can be tailored to best represent the underlying simulation parameters. The main focus of the visualization system is the ability to explore input parameters using parallel coordinates. The time-varying energy usage outputs are rendered using HDR curve plots that preserve outliers, which is often crucial to the visual analytics process. The categorical and geospatial data is handled using interactive bar chart and map views, respectively. All of the views are linked together with interactive selection so that data selected in one view is also highlighted in all other views. A screenshot overview of the system is presented in Figure 2. The following sections describe each of the features in greater detail.

Parallel Coordinates

Parallel coordinate plots have been widely used as one of the most intuitive mechanisms for visualizing multivariate data. There are a few possible mechanisms for handling RSC's numerical and categorical data. One possibility would be to completely separate categorical and numerical

variables into two different plots, each with its own rendering style. However, this would limit the ability of mixing and reordering axes to gain possible correlation information between data types. Another possibility is the mapping of categorical data to a numerical value. However, this would force the implication of continuity on discrete elements that would cause visualization artifacts that can mislead users. Instead, the technique described here allows both data types to be rendered in a single plot by combining the strengths of existing techniques with a few key enhancements.

Data Aggregation

The first issue to address is how to render such a large amount of information interactively. Standard line rendering approaches cannot handle millions of data items well due to graphical processing requirements which cause very slow rendering that can break the possibility of interactivity. Another major, yet typical, issue is overplotting where line-rendering approaches fill a single pixel with the last line color even though that may have 1 line or millions of lines (of different colors and types) passing through it without any detail to convey that information. Thus, we turn to frequency-based rendering which aggregates the data to provide an overview of the data with less rendering time.

Along the works of Bendix and Novotny (Bendix, Kosara, & Hauser, 2006; Novotny & Hauser, 2006), we first generate a 2D histogram of pairwise frequency data for each pair of adjacent axes and dynamically recompute this whenever an axis is moved to a new location. This stores the image of the data flow between 2 given axes. Unlike the previous work, however, our application supports categorical data by allowing non-uniform histogram dimensions. The bin sizes are determined as either the number of categories for categorical axes or a user specified size (typically min and max) for numerical data. Figure 3(a) shows an example of the end result from the unified technique, using color intensity to denote occupancy levels.

Context Rendering

In addition to rendering a line for every simulation, previous frequency-based parallel coordinates have also used quadrilaterals (quad) to render ribbons of data that were binned together. Also, when a line or quad for data converges to an axis, there is no guarantee of continuity (i.e. that it will leave the axis at the same point). This makes tracking flow of data difficult due to hidden discontinuities at every axis. In an effort to ameliorate these concerns, we use extensions to the rounded edge rendering presented by Mansfield (Mansfield. & Ambachtsheer, 2003) which allows for smoother angle changes for incoming and outgoing data at each axis.

The way bins are rendered depends on whether the pair of axes is both numerical, both categorical, or a mix. For pairs of numerical axes, the bins are of equal size since they represent given numerical ranges from one axis to another. To emphasize the largest data concentrations, bins are ordered based on occupancy and their respective visual elements are rendered from low to high. Colormaps based on occupancy are also used to help distinguish data bins from one another. In Figure 3(a), the two axes on the furthest right are both numerical.

For pairs of categorical axes, since categorical data does not have the strict positional requirements inherent in continuous data, the size and axis positioning of bins can be modified to improve

readability. The axis is first split into its respective categories, providing a visual separate of each. Proportional sizing is allowed to enhance the ability to see the percentage of incoming and outgoing items for a given bin (much like a bar chart) beyond the standard coloring technique. Categorical also provides two types of end connections: overlapping and category stacking. Overlapping bin connections works exactly as described with numerical data, where each bin connects to the top and bottom of its respective category sub-axes. For category stacking, bin start and end vertices are stacked so one category may have multiple incoming or outgoing channels of data (that do not overlap). This allows bins to be traced more easily besides creating a much cleaner visualization. The two left-axes in Figure 3(a) are categorical.

When the pair of axes is mixed, the difference in bin sizes creates a continuity problem since there may be a small number of categories but a large quantity of numerical values. To allow for a continuous visual flow, the incoming and outgoing bin sizes for each axis must match. This causes a gradual ballooning effect as the geometry increases from the smaller to larger side. This can be seen by looking at the middle axes in Figure 3(a).

Focus Rendering

Since the visualization is intended for exploration, focus sets must also be rendered in conjunction with the context. Between any pair of axes, an aggregate focus object is rendered atop the context bin and sized according to the percentage of occupancy of focus items, as can be seen in Figure 3(b). Thus, the size of the red connection indicates the percentage of items from the relevant context bin, regardless of whether the bins are rendering using uniform or proportional sizing. This allows the user to quickly determine not only when values are connected, but also estimate how many items are present in that selection for each value pair.

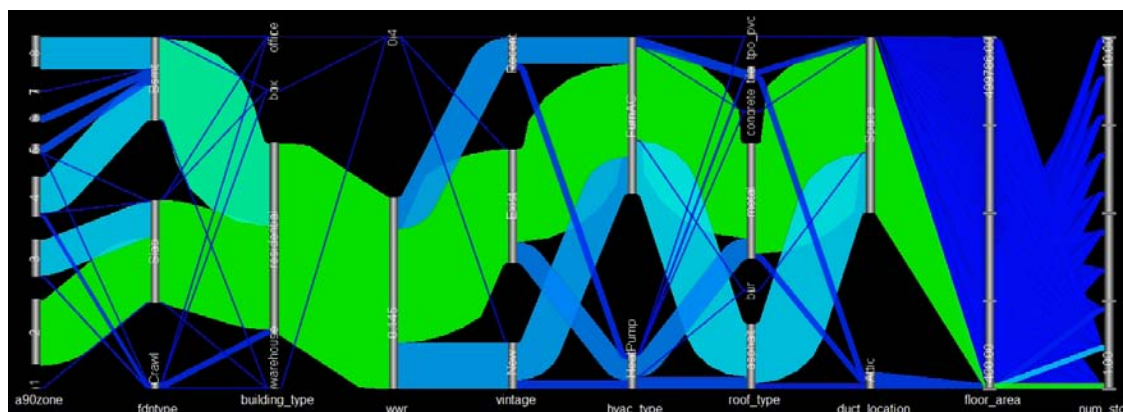
When using category bin stacking with focus items on top, the disconnect of information flow that occurs on axes was distracting for some users due to enhanced discontinuity at axis boundaries. Thus, we added focus-merging to the visualization that joins together individual focus primitives before entering and leaving an axis. Thus, the unified focus set can be seen as flowing from one category to the next.

Since the exploration process is interactive, care must be taken to ensure the combination of all these features is fast and that the software does not become unresponsive. This is accomplished using multi-threading where each thread is assigned a subset of the data items and a pair of axes to handle all required bin recalculation and rendering operations on-demand.

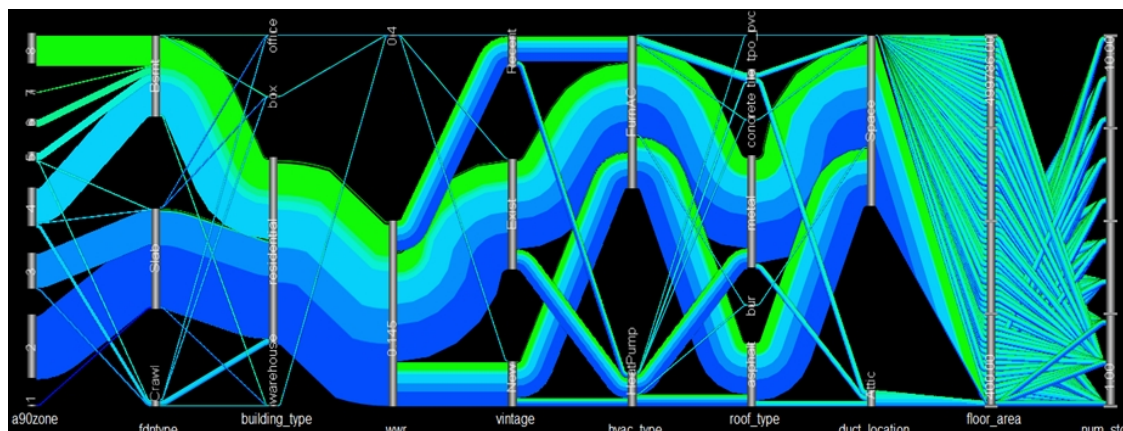
Interaction

The user is given control over the layout of the axes, the rendering styles, plot coloring and data selection. All of these operations help in generating an effective visualization and finding features of interest during the exploration process. Axis reordering is the most basic operation a user can perform. By simply clicking and dragging an axis with the mouse, it can be moved to a new location within the PCP. In addition, the controls window allow for size, font, and primitive styles, such as curves, quad, and line rendering, to be changed on the fly.

The selection of data in the PCP is done using the mouse to create selection boxes by clicking



(a)



(b)

Figure 4: The color and transparency of data can be controlled via a transfer function editor (which determines the mapping from data space to visual space). The default is coloring based on bin occupancy, shown in (a). The bins may also be colored by a chosen axis to provide flow of a single variable across the entire visualization, shown in (b).

and dragging. Each rectangle specifies a minimum and maximum data selection for one or more axes at a time to select all data items within those ranges. The same operation can be done while holding shift to deselect items in the highlighted ranges. Once a range is specified, the focus set is updated by discovering which items have been touched by the selection operation. The focus set is updated in a separate multi-threaded process to prevent lag in the system interaction.

The color and transparency for context and focus can be changed independently. The user is able to change the color through a pair of pop-up transfer function editors which use color wheels. While bin occupancy is used for coloring, any level of occupancy (from low to high) can be emphasized by the user. In addition, the user can also map a major axis to color, which will show the distribution of values for that dimension as it flows through all others as shown in Figure 4.

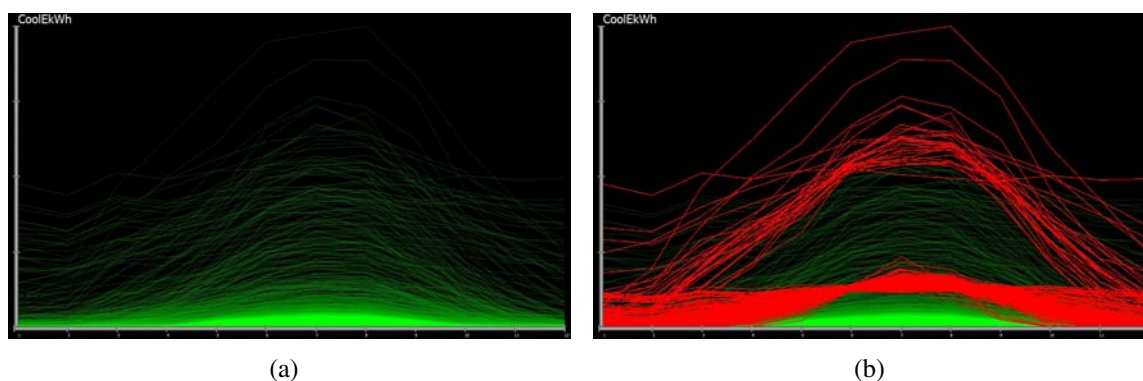


Figure 5: Time plots are rendered using high-precision textures and outlier-preserving sampling.

Linked Visual Analytics

Though the PCP represents the main analysis tool for researchers to explore their data, specialized visualizations handling time-varying energy data, categorical distributions, and geographic climate zones have been created to handle specific types of data. These visualizations are all linked into a single system to allow focus selection and rendering across all views in real-time, allowing more in depth feature exploration and support for visual analysis.

High-Density Time Plots

Time-series plots are created for line rendering of monthly data for the three output variables (electric cooling, electric heating, and natural gas heating). Due to the density of the lines, the plot can quickly become oversaturated, causing patterns to become lost. Thus, high-density rendering is performed using a high-precision element recognized by modern video cards as a framebuffer object (FBO). The FBO records the number of hits at each pixel location and a logarithmic function is then applied to color for revealing detail in saturated regions. However, a strict logarithmic function would hide any outlier details, thus we also emphasize regions below a specific threshold so that uncommon values are also shown. This is done by restricting the logarithmic opacity modification to high occupancy pixels in the fragment shader (i.e. shaders are programs computed on the graphics card) to maximize interactivity. Focus lines are rendered on top of this high-density view with a constant opacity value. An example of the time plots is illustrated in Figure 5.

With over 3 million lines to render for each plot, rendering can be quite slow even with GPU accelerated FBOs, fragment shaders, and vertex shaders. To overcome this problem, we apply an occupancy-based outlier sampling method to pick a subset of lines to render for each plot. Between each pair of months along the horizontal axis, lines are binned based on what region of the vertical axis they pass through. All curves that fall into low occupancy bins are automatically chosen to be rendered. By default, the threshold for low occupancy is 10% of the total number of lines. The remaining lines then randomly sampled to provide a distribution of the high traffic areas.

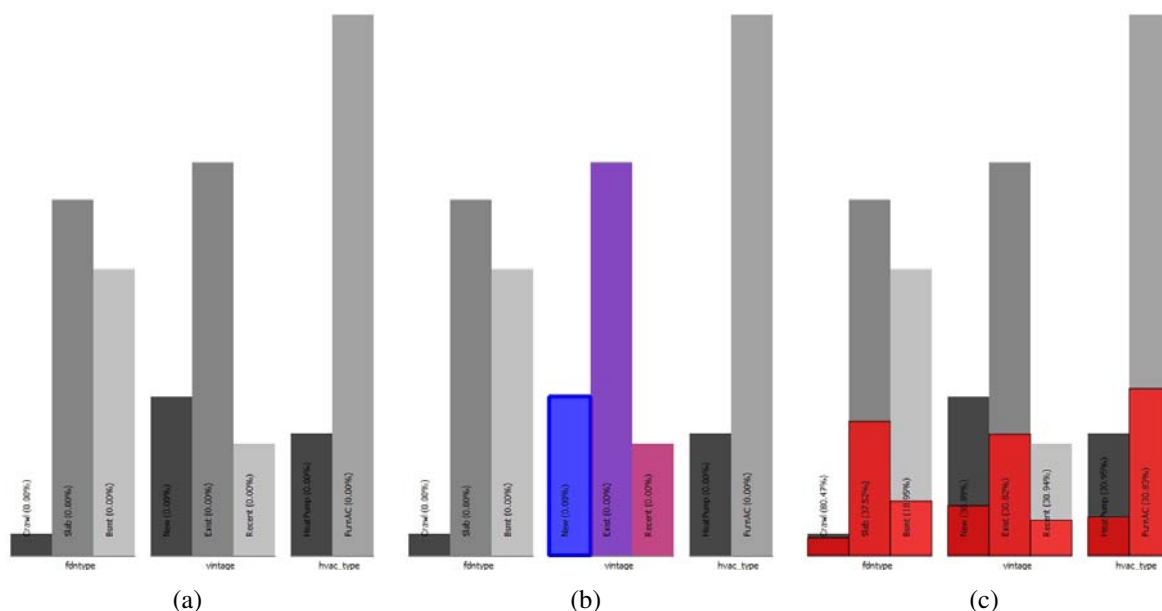


Figure 6: The category view presents a series of bar charts for each categorical dimension (a). When hovering over a particular dimension, coloring is applied to visually separate it from the others (b). When focus items are selected, a smaller red bar is drawn to show how much of each category is selected (c).

Categorical View

The category view, shown in Figure 6, is a series of simple, interactive bar charts that allow the users to quickly determine the distribution of categorical building parameters. Each categorical dimension is given a bar for every category it contains. The bars are sized base proportionally to the number of data items contained within that category. The user can simply select and deselect entire categories with the click of a mouse. On mouse over, individual categories are highlighted to provide visual emphasis. When an item from a category is selected (perhaps through interaction with other views), a red bar is drawn on top that shows how much of the category is selected.

Map View

A crucial component for building energy simulations is the local climate for each housing structure (e.g., a house in Alaska has very different energy needs when compared to one in Texas). The user can select and deselect climate zones directly on a map of the U.S. as shown in Figure 7. A standard web clickmap is used to carve the map into polygonal regions and the intersection of a mouse click with one of these regions is used to select the corresponding simulations from that ASHRAE zone. By adjusting the polygonal fill pattern to denser cross patterns, the user can approximate how much of that climate zone's data items are currently selected as a focus. Thus, polygonal fill represents focus occupancy for that region.

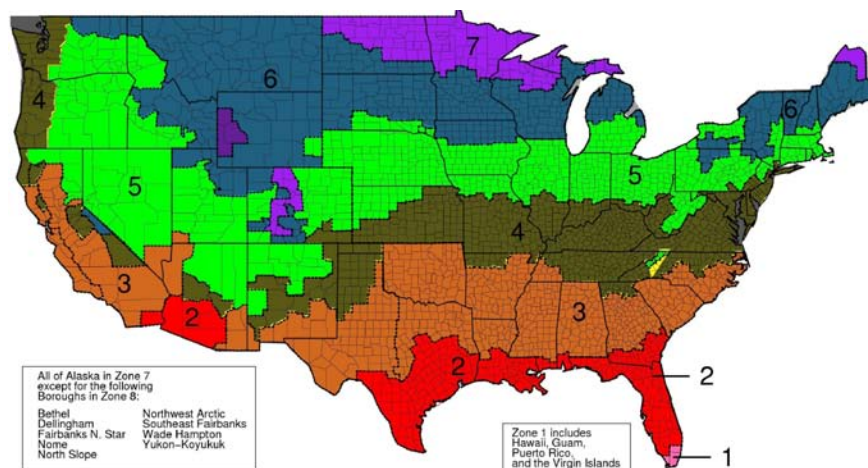


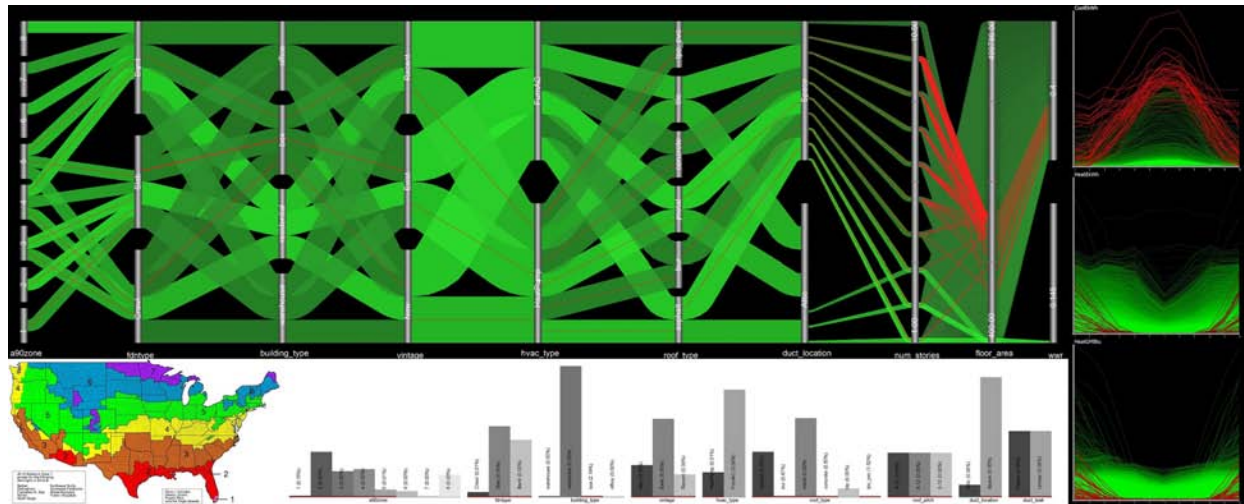
Figure 7: The map view allows users to select and deselect climate zones.

RESULTS AND DISCUSSION

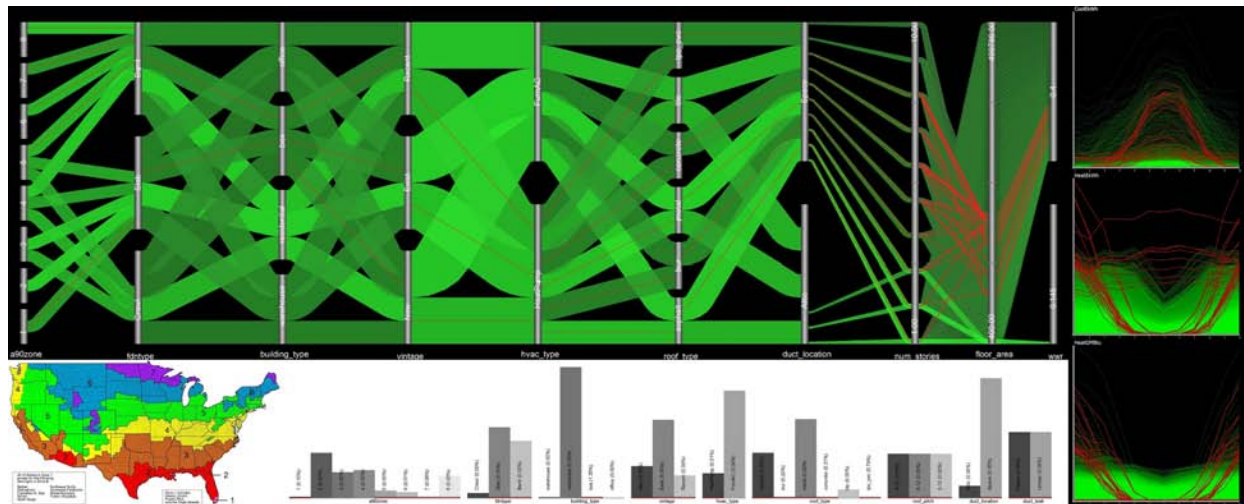
This effort is the result of a partnership of building energy simulations researchers and computer scientists. The building energy simulation experts provided the domain knowledge and expert guidance, and helped to evaluate the tool. In one session with 2 domain experts, 30 minutes was able to establish several known/expected patterns (hypothesis testing), 11 potential simulation bugs (through database mining), and 11 new patterns in the data which the domain experts found interesting and could not explain (knowledge discovery). Only a few of these are selected for discussion below.

Even before any data filtering was used, the scientists noted some interesting high-level anomalies immediately when viewing the entire data. By placing the 'Roof Type' next to 'Duct Location' it was noted that three of the commercial roof types (concrete, PVC, and built up roof) did not connect to the attic category on the parallel coordinates plot. It was found that this indicated a lack of coverage in the parametric simulations in the database since these commercial types were simulated using only a plenum and no attic space. In addition, by examining the bar charts in the category view, there was apparent bias toward modeling certain climate zone regions and building types. Namely, a vast majority of the building simulations are for residential homes, with only a few for offices and warehouses (indicating in part that most people who use the website are homeowners).

The first point of interest was discovering the causes of outliers in the energy signatures that are present in the function plots. By brushing these outlying lines directly, the focus switches in all of the coordinated views. As a sanity check, Figure 8(a) shows outliers selected from the cooling energy which verifies that no energy is being used for cooling in ASHRAE climate zone 8, as shown by the lack of red lines for the category at the top of the leftmost axis of the PCP. In Figure 8(b), interesting outliers were selected in the HDR line plot which show more use of energy for heating in the summer months than in the winter months (which doesn't make sense). As expected, the extreme heating values occur in the coldest three climate zones, but the outliers also

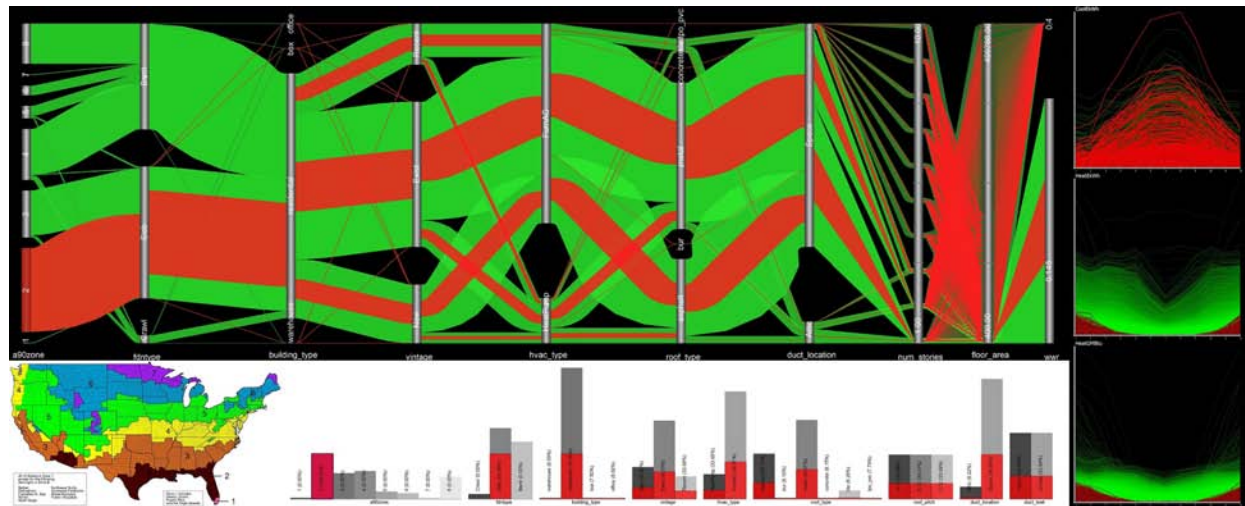


(a)

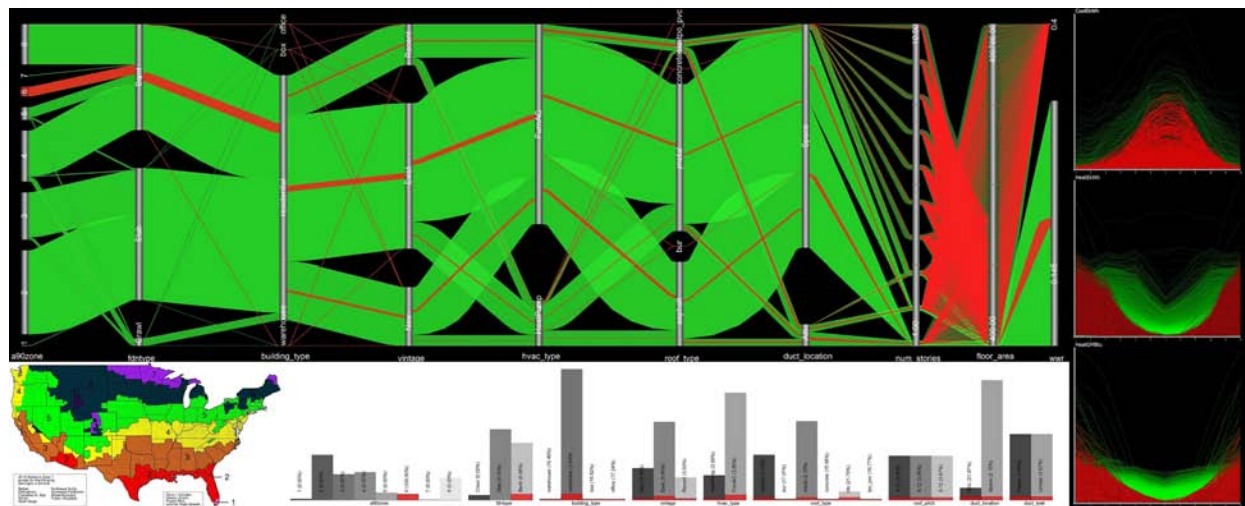


(b)

Figure 8: Outliers are selected on the electrical cooling (a) and electrical heating (b) time plot. In both cases, the focus items share the same foundation type (box) as seen in the PCP. These simulation errors were thus easily tracked to a bug in a building template file, even though these errors existed in only 8 of over 3 million simulations.



(a)



(b)

Figure 9: The images shown here have focused on zone 2 (a) and zone 6 (b). Due to their location differences, the housing parameters and heating/cooling needs are different between the two feature sets.

occur in climate zone 1. Furthermore, the 3rd axis from the left shows all outliers pass through the “box” building type. This corresponds to a DOE-2 building template file and indeed an error in the macro of the building description language (BDL) for this file was detected and corrected. In only a few minutes, domain experts that had never used this visualization technique before happened to identify an error in the complex RSC simulation system in 8 simulations out of over 3 million.

From the parallel coordinates plot, the researchers noticed that when the floor area was paired with the number of stories, that a bin was present which indicated some of the simulated buildings were 10 stories tall but only 400 square feet in area (a 6' \times 6' building). By selecting this bin, it shows that this only occurs for the office building type. Since their building simulation model is not intended to model the heating and cooling of elevator shafts, it was discovered that a minimum building size needs to be enforced to eliminate invalid building types.

Another aspect of the RSC data set that is of interest to building scientists is verifying model parameters over different climate zones. In Figure 9, two different climate zones, 2 and 6, were made the focus of RSC Explorer. The focus rendering technique developed for the parallel coordinates plot allows researchers to quickly assess the proportion of data items traveling between axes. Zone 2, shown in Figure 9(a), contains many more simulation runs than are present in zone 6. One interesting aspect separates the two zones on the second dimension, foundation type. Namely, zone 2 buildings exclusively have slab foundations while zone 6 buildings exclusively have basement foundations. This is related to the building standards and codes for these two zones. Furthermore, the difference in climate between these two zones causes different energy signatures to be seen in the energy function plots. As is seen in Figure 9(b), Zone 6, which is located further north, uses much more electricity and gas for heating during the winter season. On the other hand, zone 2 has higher energy usage for cooling during all of the months of the year. Since zone 2 contains some of the warmest parts of the country this behavior confirms what is expected from the simulation model.

The researchers also gave valuable feedback for some shortcomings that could be addressed to help improve RSC Explorer. First, while manual ordering of axes is available, it would be beneficial to the user to provide an initial ordering that avoided bad combinations. This comment was in response to some numerical axis pairs, such as HVAC efficiencies, being so evenly connected that no visible trend was present. Second, focus rendering is still difficult to follow at times for categorical axes due to the disconnect that can occur as a result of bin stacking. Third, the inability to perform PCP axis-specific zooming for ranges of interest was seen as limiting. Finally, researchers appreciated the category view with bar charts as it provided a clean indication of categories selected; however, they noted that the linear scale used caused small categories to be too small, thus limiting their usefulness.

The tool has been designed in a way that allows it to run on desktop computers as well as large, wall-sized displays. We have been able to conduct some of the analysis in the 35 million pixel EVEREST powerwall as well as the 16 million pixel TRAVL display wall at the Oak Ridge National Laboratory (Figure 10). The large size and much higher resolution allowed experts to immediately pick up features that were difficult to perceive on conventional desktops.



Figure 10: ORNL researchers exploring the result of RSC ensemble simulations on the 16 million pixel TRAVL display wall.

CONCLUSION

This paper has presented a multiview visual analytics system called RSC Explorer, which combines parallel coordinates, function plots, bar charts and map views together for visualizing numerical and categorical data. In order to allow interactive exploration of over 3 million simulations, many technical contributions were made to visualization techniques of which only a few have been discussed in this paper. Most importantly, we created a frequency-based parallel coordinates plot that is able to handle large scale data and provides a rendering style for pairs of axes. In addition, rounded bin rendering was used to alleviate aliasing and shrinking problems that are inherent in the standard quad-based rendering techniques. Coupling this versatile parallel coordinates plot with other plots provides a well rounded system for integrated exploration. Although applicable to any large scale numerical or categorical data, building scientists have been able to quickly visualize and analyze millions of simulations in an interactive fashion, verify proper visualization of known trends, query the data intuitively to test working hypotheses, and to quickly identify complex errors that would normally be buried in such large data.

ACKNOWLEDGEMENTS

This work was funded by field work proposal CEBT105 under the Department of Energy Building Technology Activity Number BT0305000. The Roof Savings Calculator at roofcalc.com was

made possible through California Energy Commission's Public Interest Energy Research (PIER) program through the U. S. Department of Energy under contract DE-AC03-76SF00098 and involved collaboration with California Energy Commission, Lawrence Berkeley National Laboratory, the Environmental Protection Agency, Oak Ridge National Laboratory, White Box Technologies, and Jacksonville State University. The authors would also like to thank the Oak Ridge Leadership Computing Facility (OLCF) for use of the Exploratory Visualization Environment for Research in Science and Technology (EVEREST) powerwall as well as the National Transportation Research Center (NTRC) for use of their Transportation Research And Visualization Laboratory (TRAVL) display wall.

Oak Ridge National Laboratory is managed by UT-Battelle, LLC, for the U.S. Dept. of Energy under contract DE-AC05-00OR22725. This manuscript has been authored by UT-Battelle, LLC, under Contract Number DE-AC05-00OR22725 with the U.S. Department of Energy. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes.

REFERENCES

- Bendix, F., Kosara, R., & Hauser, H. (2006, July). Parallel sets: Interactive exploration and visual analysis of categorical data. *IEEE Transactions on Visualization and Computer Graphics*, 12, 558–568. Retrieved from <http://portal.acm.org/citation.cfm?id=1137246.1137510> doi: 10.1109/TVCG.2006.76
- DOE. (2012). *Department of Energy FY 2012 Congressional Budget Request* (Vol. 3). Retrieved from <http://www.cfo.doe.gov/budget/12budget/Content/volume3.pdf>
- Fua, Y.-H., Ward, M. O., & Rundensteiner, E. A. (1999). Hierarchical parallel coordinates for exploration of large datasets. In *In proceedings of the ieee visualization 1999 conference* (pp. 43–50). Los Alamitos, CA, USA: IEEE Computer Society Press.
- Inselberg, A. (1985, December). The plane with parallel coordinates. *The Visual Computer*, VI(4), 69–91. Retrieved from <http://dx.doi.org/10.1007/BF01898350> doi: 10.1007/BF01898350
- Inselberg, A., & Dimsdale, B. (1990). Parallel coordinates: a tool for visualizing multi-dimensional geometry. In *In proceedings of the ieee visualization 1990 conference* (pp. 361–378). Los Alamitos, CA, USA: IEEE Computer Society Press.
- Johansson, S. (2009). Visual exploration of categorical and mixed data sets. In *Proceedings of the acm sigkdd workshop on visual analytics and knowledge discovery: Integrating automated analysis with interactive exploration* (pp. 21–29). New York, NY, USA: ACM. Retrieved from <http://doi.acm.org/10.1145/1562849.1562852> doi: <http://doi.acm.org/10.1145/1562849.1562852>

- Johansson, S., & Johansson, J. (2010, May). Visual analysis of mixed data sets using interactive quantification. *SIGKDD Explor. Newsl.*, 11, 29–38. Retrieved from <http://doi.acm.org/10.1145/1809400.1809406> doi: <http://doi.acm.org/10.1145/1809400.1809406>
- Mansfield., P. A., & Ambachtsheer, M. (2003). Programmatic rendering of directed, weighted graphs. In *Proceedings of the SVG Open 2003*.
- New, J. R., Miller, W. B., Desjarlais, A., Huang, Y. J., & Erdem, E. (2011). Development of a roof savings calculator. In *Proceedings of the rci 26th international convention and trade show*. Reno, NV.
- Novotny, M., & Hauser, H. (2006). Outlier-preserving focus+context visualization in parallel coordinates. *IEEE Transactions on Visualization and Computer Graphics*, 12(5), 893–900. (Member-Helwig Hauser) doi: <http://dx.doi.org/10.1109/TVCG.2006.170>

Editors:

Michel Avital, University of Amsterdam
Kevin Crowston, Syracuse University

Advisory Board:

Kalle Lyytinen, Case Western Reserve University
Roger Clarke, Australian National University
Sue Conger, University of Dallas
Marco De Marco, Università Cattolica di Milano
Guy Fitzgerald, Brunel University
Rudy Hirschheim, Louisiana State University
Blake Ives, University of Houston
Sirkka Jarvenpää, University of Texas at Austin
John King, University of Michigan
Rik Maes, University of Amsterdam
Dan Robey, Georgia State University
Frantz Rowe, University of Nantes
Detmar Straub, Georgia State University
Richard T. Watson, University of Georgia
Ron Weber, Monash University
Kwok Kee Wei, City University of Hong Kong

Sponsors:

Association for Information Systems (AIS)
AIM
itAIS
Addis Ababa University, Ethiopia
American University, USA
Case Western Reserve University, USA
City University of Hong Kong, China
Copenhagen Business School, Denmark
Hanken School of Economics, Finland
Helsinki School of Economics, Finland
Indiana University, USA
Katholieke Universiteit Leuven, Belgium
Lancaster University, UK
Leeds Metropolitan University, UK
National University of Ireland Galway, Ireland
New York University, USA
Pennsylvania State University, USA
Pepperdine University, USA
Syracuse University, USA
University of Amsterdam, Netherlands
University of Dallas, USA
University of Georgia, USA
University of Groningen, Netherlands
University of Limerick, Ireland
University of Oslo, Norway
University of San Francisco, USA
University of Washington, USA
Victoria University of Wellington, New Zealand
Viktoria Institute, Sweden

Editorial Board:

Margunn Aanestad, University of Oslo
Steven Alter, University of San Francisco
Egon Berghout, University of Groningen
Bo-Christer Bjork, Hanken School of Economics
Tony Bryant, Leeds Metropolitan University
Erran Carmel, American University
Kieran Conboy, National U. of Ireland Galway
Jan Damsgaard, Copenhagen Business School
Robert Davison, City University of Hong Kong
Guido Dedene, Katholieke Universiteit Leuven
Alan Dennis, Indiana University
Brian Fitzgerald, University of Limerick
Ole Hanseth, University of Oslo
Ola Henfridsson, Viktoria Institute
Sid Huff, Victoria University of Wellington
Ard Huizing, University of Amsterdam
Lucas Introna, Lancaster University
Panos Ipeirotis, New York University
Robert Mason, University of Washington
John Mooney, Pepperdine University
Steve Sawyer, Pennsylvania State University
Virpi Tuunainen, Helsinki School of Economics
Francesco Virili, Università degli Studi di Cassino

Managing Editor:

Bas Smit, University of Amsterdam

Office:

Sprouts
University of Amsterdam
Roetersstraat 11, Room E 2.74
1018 WB Amsterdam, Netherlands
Email: admin@sprouts.aisnet.org