ERNEST ORLANDO LAWRENCE BERKELEY NATIONAL LABORATORY

A SOFTWARE TOOL TO COMPARE MEASURED AND SIMULATED BUILDING ENERGY PERFORMANCE DATA

Author(s), Tobias Maile, Vladimir Bazjanac, James O'Donnell and Matthew Garr

Environmental Energy and Technologies Division

DISCLAIMER

This document was prepared as an account of work sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor The Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or The Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof or The Regents of the University of California.

A SOFTWARE TOOL TO COMPARE MEASURED AND SIMULATED BUILDING ENERGY PERFORMANCE DATA

Tobias Maile[1&2], Vladimir Bazjanac[1], James O'Donnell[1], Matthew Garr[2]
[1]Lawrence Berkeley National Laboratory, Univ. of California, Berkeley, USA
[2]Center for Integrated Facility Engineering (CIFE), Stanford University, Stanford, USA

ABSTRACT

Building energy performance is often inadequate when compared to design goals. To link design goals to actual operation one can compare measured with simulated energy performance data. Our previously developed comparison approach is the Energy Performance Comparison Methodology (EPCM), which enables the identification of performance problems based on a comparison of measured and simulated performance data. In context of this method, we developed a software tool that provides graphing and data processing capabilities of the two performance data sets. The software tool called SEE IT (Stanford Energy Efficiency Information Tool) eliminates the need for manual generation of data plots and data reformatting. SEE IT makes the generation of time series, scatter and carpet plots independent of the source of data (measured or simulated) and provides a valuable tool for comparing measurements with simulation results. SEE IT also allows assigning data points on a predefined building object hierarchy and supports different versions of simulated performance data. This paper briefly introduces the EPCM, describes the SEE IT tool and illustrates its use in the context of a building case study.

INTRODUCTION

The recent increase in available energy performance data from buildings generated using measurement and simulation has led to large HVAC data sets (Maile 2010). Commissioning agents, simulation experts or more generally performance analysts who are trying to analyze these data struggle with the amount of data as well as the lack of easy to use data visualization tools.

Existing software tools that plot HVAC data have various shortcomings. Control vendor products lack high quality data visualization (Haves et al. 2008) and analysis tools are burdensome to use (Isakson & Eriksson 2004). Most require on data in a specific format (e.g. Friedman & Piette 2001; Mazzarella et al. 2009), and are typically residing in some comma separated text files. However, databases are a more useful storage mechanism, given the increased magnitude of data. While simulated performance data are typically very regular and available for time periods up to one year, measured performance data are often irregular with false readings, gaps and even duplicate values. These different data set characteristics cause processing problems for a number of existing tools. A particular challenge is to analyze both measured and simulated data due to their different characteristics. Kircher (2010) recommends the integration of measured and simulated performance data as a future research area.

Due to the increased complexity of HVAC systems and their controls, the performance analyst needs data analysis tools that provide flexibility. The analyst needs to explore HVAC data relatively quickly without the need to massage data in a time-consuming manner, as is typical when using traditional tools. Existing tools often have a steep learning curve (Friedman & Piette 2001) and do not support the needs of the performance analyst who is faced with large amounts of data.

Besides the data processing, the increase in available data also leads to the need for a more defined structure to manage all available data points. Most tools do not provide a sophisticated manner with which to logically organize the different data points.

To address these shortcomings and challenges the authors developed the Stanford Energy Efficiency Information Tool (SEE IT). SEE IT provides structure through a building object hierarchy, allowing analysts to assign data points and plot configurations to specific building objects. It accesses the original data in SQL (Structured Query Language) databases and provides on-the-fly averaging though corresponding SQL statements. Three different data plot types are available: time series, carpet, and scatter plots. SEE IT supports multiple data sources, in particular multiple simulated data sources, to allow iterative simulation process. As this iterative process is a key mechanism of the Energy Performance Comparison Method, (EPCM), SEE IT is a critical enabler of the method's usability in practice.

In this paper, we summarize the EPCM, describe SEE IT's development, functionality, underlying databases, limitations and future directions, as well as illustrate its use with one case study.
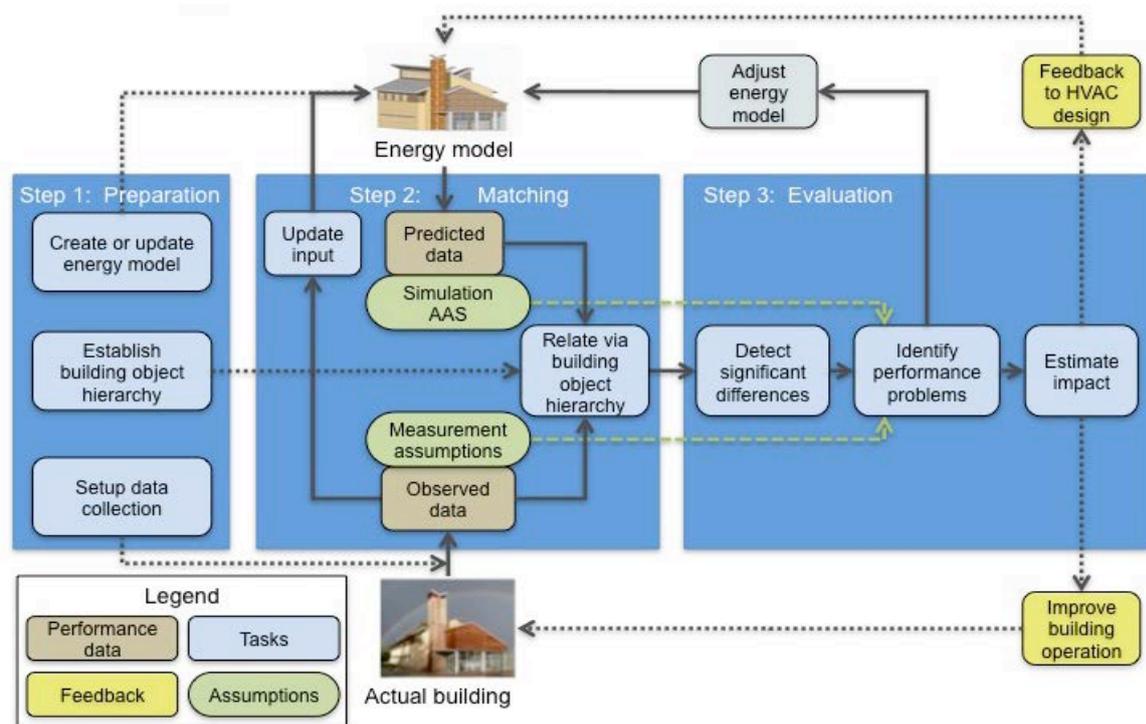
Figure 1: Energy Performance Comparison Methodology

THE ENERGY PERFORMANCE COMPARISON METHODOLOGY

This section briefly describes the Energy Performance Comparison Methodology. For a detailed description refer to Maile (2010).

EPCM aims to identify performance problems of a given building by comparing measured and simulated performance data. The concept of the tool is based on the assumption that the given building has an advanced set of measured performance data available, with at least several months of data, and one or more Building Energy Performance Simulation (BEPS) models.

EPCM guides the analyst to iteratively detect differences between measured and simulated performance data. Previous concepts to calibrate simulation models (e.g. Reddy 2006) use specific error margins (e.g. 5%) as criteria to adjust simulation models. However, these concepts do not consider the limitations of measurement systems and simulation models that influence the performance data. Without considering what measurement assumptions and simulation assumptions, approximations and simplifications (AAS)[1] are contained in performance data, a calibrated model may lead to wrong conclusions. The EPCM includes measurement assumptions and simulation AAS to describe and share the knowledge of limitations. With this increased knowledge about the measurements or the simulation model, the analyst can more reliably assess discrepancies among performance data.

The EPCM consists of three major steps.

Step 1, the preparation step, includes the setup of data collection, the update of the simulated model and the creation of the building object hierarchy. This hierarchy logically organized building objects in a tree like structure.

Step 2 is called the matching step. The analyst creates data pairs of a measured data point and the corresponding simulated data point via the building object hierarchy. Given actual weather data (such as outside air temperature), input data for the simulation are updated within step 2. The analyst also assigns measurement assumptions and simulation AAS to the data points in the building object hierarchy for the later assessment of differences.

In step 3, the evaluation step, the analyst detects differences of the just-created data pairs. The analyst can evaluate these differences with the help of the previously defined assumptions and AAS, and thereby identify performance problems. The analyst then adjusts the simulation model to mimic the behaviour of the building. This iterative process eliminates some of the differences and provides a better understanding of the reasons for found discrepancies. Estimating the impact on energy consumption and thermal comfort can help prioritize the feedback to the HVAC design or the actual building operation. Once performance problems have been fixed in

---

[1] We refer to Maile (2010) for a definition of these terms.

the real building and new performance data accumulated, one can again use the EPCM to re-evaluate the building performance.

DEVELOPMENT OF SEE IT

We used a rapid prototype development process to develop SEE IT. As apparent from the summary of EPCM, this method requires a software tool that is capable of quickly generating visual data plots to compare measured and simulated data. This need for a visualization tool was the primary motivation for developing SEE IT and initially caused the following three requirements:

- Quick generation of plots (time series plots)
- Support for measured and simulated performance data
- Definition of a way to deal with a lot of data

Its development produced three prototypical tools.
The predecessor of SEE IT is a spreadsheet-based tool that fetches performance data from databases and can generate time series plots. Version 1 of SEE IT instantaneously generates plots without any capabilities to store data point or plot assignment. Version 2 includes the building object hierarchy as well as storing capabilities.

Predecessor of SEE IT

The predecessor of SEE IT in the form of a macro-based spreadsheet tool is able to fetch data from an SQL database and cache it with the spreadsheet. It includes time series plots only and some simple unit conversion routines. This tool deals with the magnitude of data in a very pragmatic way; each instance of the spreadsheet only contains a small subset of data. Different versions of simulation results can also easily be documented with new instances of the spreadsheet. While it is able to capture measured and simulated data, it has a significant shortcoming: it has no structure, only a limited list of data points within one worksheet. The analyst must preselect data points and then add them to a specific cell in the sheet. The processing time to load the data and create a graph is considerably larger than that of the latest SEE IT version, mainly due to the averaging of data and the missing possibility to select different intervals. Based on this first prototype, we determined a set of additional requirements for inclusion in the next version:

- Additional plot types (e.g. scatter or carpet plot)
- Better performance in generating plots
- Improved unit conversions (IP and SI units)
- Improved selection of data points
- Selection of different averaging time intervals
- Improved versioning capabilities

SEE IT version 1

SEE IT version 1 addresses most of the additional requirements. It enables the use of scatter and carpet plots as well as the time series plots. For simplicity, the development started with a focus on measured data only. Version 1 is essentially a tool that allows generation of instantaneous plots based on measured data. In a later update we added plots for simulated data. SEE IT Version 1 improves the generation time of plots, since the data are fetched and formatted directly with SQL statements from the database. Two issues became apparent during the use of SEE IT version 1. One is that a plot should contain all metadata that are needed to generate the plot: (include time frame, data points, interval and units). This feature eliminates uncertainties about data contained in a particular plot. The second issue is that it would be helpful to organize data points and be able to store plot configurations, so one can reuse them later. These issues triggered the development of SEE IT version 2 and are summarized in the following requirements:

- Include all metadata within a plot
- Provide the ability to organize data points
- Enable the saving of plot configurations and data point structure

SEE IT version 2

Version 2 includes the building object hierarchy, measured and simulated data, three plot types, a storing mechanism of the current setup within SEE IT and the assignment of data points as well as plot configurations to building objects. The requirement to include all metadata into the plot is also satisfied in this version. To enable different versions of simulated data, it was first necessary to support multiple different simulated data sources. With multiple data source, two issues are important. The first one is the affiliation of each instance of a data point to a data source. This was addressed by using an identifier within the data point string. This identifier does not only show the version of the simulated data source but also differentiates between measured and simulated data sources. The second important issue is the affiliation of data points to one or the other data source, once the analyst adds a new data source. The analyst can either override the existing data source and thus assign data

points with a new one, or just add the new one, leaving the assigned data points connected to the original data source.

The authors discuss possible improvements of SEE IT version 2 in the section about the limitations and future development below.

FUNCIONALITY OF SEE IT VERSION 2

As seen in Figure 2, SEE IT provides the majority of its functionality within one window. This includes the building object hierarchy, a building object specific point list, a measured and a simulated data source point list, available plots, plot configurations, interval, units, and start and end time.
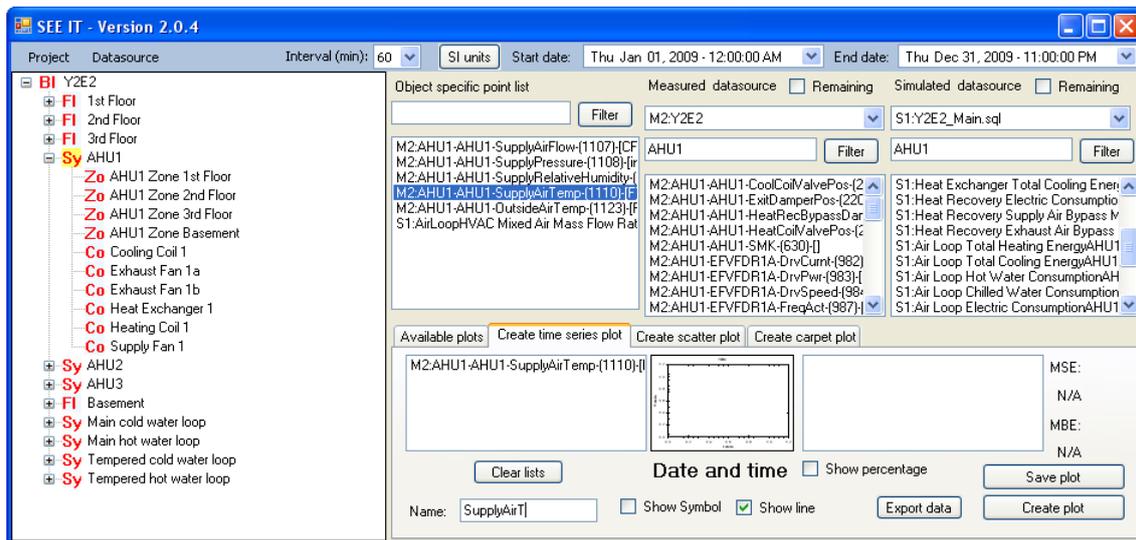
Figure 2: SEE IT version 2 screen shot

Building Object Hierarchy

The building object hierarchy in form of a tree view helps the analyst organize the different data points by assigning them to the corresponding building objects.

We used the building object hierarchy as defined by Maile (2010) as the basis for this structure. The analyst currently needs to manually create this hierarchy within the tree.

Data Sources

SEE IT allows the analyst to specify measured and simulated data sources. Both data sources are currently based on a SQL interface. The measured data source adapter enables a connection to a MySQL (Sun Microsystems 2011) database, whereas the simulated data source adapter enables a connection to a SQLite (SQLite 2011) database. The latter is the database format that is optionally generated by EnergyPlus. Multiple data sources can be generated and used within one building or project context. Only two data sources are displayed in the main window at a time. Once the analyst selects a data source, the tool displays the list of available data points, which he or she can assign to building objects in the tree. This allows the analyst to generate object-specific point lists from a number of different data sources.

Plot Generation

The lower right area of the main window enables the definition of the three different plot types: time series, carpet, and scatter plots. Any data point can be used within the plot definitions. Once a plot configuration is defined, the analyst can create a plot directly or save it and thus assign it to the currently selected building object within the tree. This functionality allows the analyst to define specific plots for building objects and reuse these at a later time. Specifically, this allows reuse of the plot configuration once a new data source is added and the data points updated. This process enables the analysis of different versions of the simulated data set.

Miscellaneous features

SEE IT provides the functionality to switch between SI and IP units. Intervals between 1 minute (original intervals) and 1 hour can be selected and the source data is averaged on the fly. Start and end time are user adjustable, but generated plots can be easily and interactively panned and zoomed in. In addition, the time series and scatter plots are able to illustrate multiple Y-axes on the left and right site of the graph, depending on the specific units of the data points.
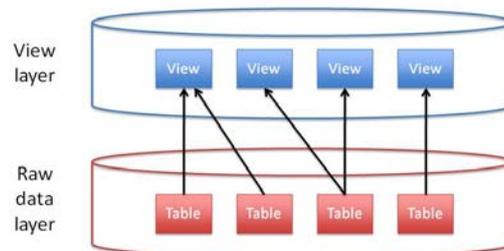
DATABASE STRUCTURE

We specifically developed the database architecture for the measured performance data, whereas EnergyPlus archives simulated data in its predefined database architecture. We therefore describe the measured data structure below and then briefly summarize the simulated data structure.

Database design for measured performance data

While the specifics of a data acquisition system greatly depend on the type of control system in a building, we defined a consistent data-archiving format. Considering the problems and time effort of data acquisition systems that store their data in files, we developed a solution based on a database. The database design is flexible enough to interface with different data acquisition systems but provides an additional generic layer that is consistent for different projects. This generic (or view) layer stores SQL queries that convert the raw data into a consistent format.

This relational MySQL database is designed with a two-layer structure consisting of raw data and views (Figure 3). The raw data layer contains summary tables and data tables. The summary tables list the available points and provide additional information for each point such as unit or type of point. The data tables contain the actual raw measured data values with attached time stamp. Summary tables use point identifiers (or ID's) to link to these data tables.

The view layer provides a more "user-friendly" and consistent perspective of the data. It contains a view that combines a complete list of available data points and single views for each point that apply correction factors, if necessary, to the raw data points.

Figure 3: Structure of the MySQL database

This two-layer structure has several advantages. First, the raw data layer can be somewhat tailored to the corresponding data acquisition system to facilitate easy and fast archiving of data. Second, the view layer enables a generic view of the data independent of specifics of the structure of the underlying raw data layer. For example, on the view layer each data point has its own view, whereas in the raw data layer, there may be multiple points per table, one point per table or multiple tables per point. The view layer uses only views that link to the raw data tables, which are created using SQL statements that query the raw data tables when the analyst requests data from a view. With this structure, the raw data layer stays intact while the views, which include all necessary data conversions, provide the correct data values to the analyst.

**Database design for simulated performance data**

EnergyPlus generates a SQLite database when providing the necessary output request. This database has only one layer and contains four tables that are of interest in this context. Two tables define the available data points (for meters and variables), one table defines the time index and one contains all the performance data. The database contains more information about the simulation model, but is not relevant for the purposes of this paper. The EnergyPlus documentation contains a detailed description of all data found in the database (DOE 2010).

ILLUSTRATIONS FROM A CASE STUDY

The case study used to illustrate the use of SEE IT is the Y2E2 building on the Stanford University campus, a building notable in this context for its comprehensive set of measurement data. Maile et al. (2010) provide a detailed description of the available data points in Y2E2. We specifically highlight the time series, scatter, and carpet plots in general and show some example of plots that compare measured and simulated performance data. Note that all plots contain the proper units on the axis, a legend referring to the specific data points, and the timeframe and interval of the plotted data. These metadata as well as the identifier (e.g., S1, S2, or M1, M2) for the data source, allow the analyst to re-create plots just based on these metadata.

Time series plot

With this case study, we illustrate the different plot types available in SEE IT. Figure 4 shows a time series plot illustrating chilled water supply and return temperatures as well as space temperatures over a two-week period.
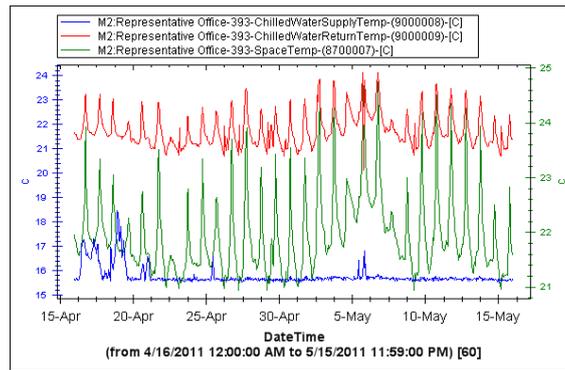
Figure 4: Example time series plot
chilled water supply and return temperature of an active chilled beam in space 393 and space temperature for a one-month period

Scatter plot
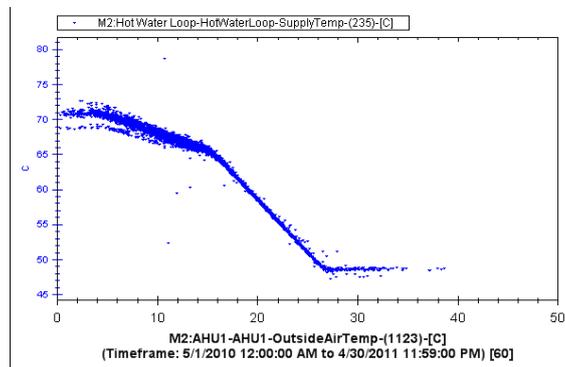
Figure 5: Example scatter plot
Hot water supply temperature of the main hot water loop versus the outside air temperature over a 12-month period

Scatter plots (Figure 5) can highlight certain relationships between different data points. In particular, in relation to the outside air temperature this plot type can reveal valuable insights. The hot water supply temperature in Figure 5, plotted against the outside air temperature, shows four different modes of operation: one in the summer (> 26 °C), two in the winter (< 6 °C and < 16 °C) and one in between. Mazzarella et al. (2009) provide a detailed description of expected patterns or signatures of scatter plots of measured data.

Carpet plot

The last plot type available in SEE IT is the carpet plot. Carpet plots can reveal daily and weekly patterns by plotting data over days (X-Axis) and time of day (Y-Axis) while varying the color by the actual data point value. As shown in Figure 6, these carpet plots can reveal specific patterns over the day, occupied and unoccupied hours, as well as weekends and holidays. We also found this type of plot useful to quickly get an overview of missing data, since data gaps can be clearly seen as white gaps in the plot.
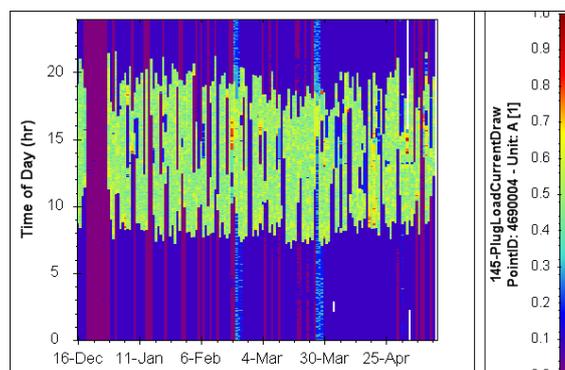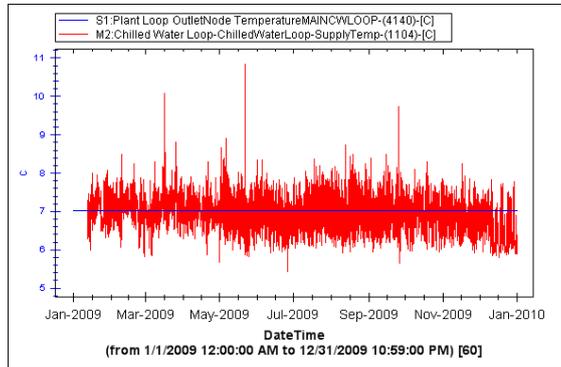
Comparison plots

In addition to allowing analysis of measured performance data in isolation, SEE IT also allows analysts to compare measured and simulated data to identify differences between measured and simulated performance. The previously introduced EPCM illustrates a method for such a comparison.
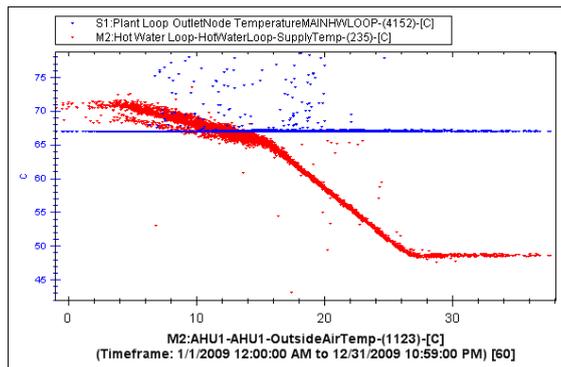
Time series comparison plot

Figure 7: Example time series plot comparing measured and simulated data
Chilled water loop supply temperature (measured and simulated) over a 12-month period

Figure 7 shows an example of a plot that compares both measured and simulated performance data. Here one can clearly see that the simulated chilled water supply temperature is constant for the whole year, whereas the measured supply temperature varies approximately plus/minus one degree Celsius. This example also clearly shows the different characteristics of simulated versus measured data. Simulated data often assume ideal undisturbed conditions, hence the constant line in the graph, while measured data often have smaller variations due to all kinds of external and internal influences.

Scatter comparison plot

Figure 8 is an example of a scatter plot comparing simulated and actual performance data for hot water loop supply temperature. In this plot, the simulated data, represented by the blue line, show only one mode of operation while the measured data show three distinct modes of operation, represented by the three slopes in the red line.

Figure 8: Example carpet plot comparing measured and simulated data
Hot water loop supply temperature (measured and simulated) versus the outside air temperature over a 12-month period

These plots illustrate the usage of the SEE IT tool to analyze measured performance data and to compare measured data with performance expectations resulting from energy simulations. A detailed analysis of this case study is currently ongoing as part of a class at Stanford University.

LIMITATIONS AND FUTURE DEVELOPMENTS

SEE IT's key feature is the combination of measured and simulated performance data. However, its current version contains a number of limitations. The following is a list of possible enhancements that could be implemented in future versions.

Automate import of building object hierarchy

Currently the analyst has to manually generate the building object hierarchy; a future version of the tool could import this structure directly from an IFC (Industry Foundation Classes) (buildingSMART 2011) file and populate the tree automatically.

Add notes and assumptions to building objects

While SEE IT enables saving of plots and data points, it currently does not provide a way to add notes or assumptions to plots or building objects.

Enable sharing of data online

The current client-side implementation does not allow collaborative sharing of plots and hierarchies between different analysts. In the class at Stanford University, the plots including descriptions and annotations are shared via a wiki page. It would be a powerful addition to implement the tool in a browser environment and save the building object hierarchy and the related data on the server so different analysts could take advantage of the work of other analysts and share their insights.

Add more data source adapters

Furthermore, the two available data sources (measured and simulated) could be extended to include more options to link measured or simulated data to SEE IT. For example, DOE-2 (DOE2 2011) is still being widely used, thus the development of a data adapter for DOE-2 could become very useful. Other measured data databases could be added easily to SEE IT given SQL functionality. Furthermore, a generic data source could be added where the advanced analyst could personally define the required SQL statements and use SEE IT to invoke those statements as necessary.

Data filters

Filtering of data is another area that needs further development in the future. Indicating weekday and weekend days or just looking at data between specific times of each day are two examples of possible filtering.

CONCLUSION

This paper describes the SEE IT tool that allows performance analysts to thoroughly investigate performance data from measured and/or simulated data sources. This combination of data sources is not found in other tools and enables a more advanced analysis of performance data. The current version supports three types of plots (carpet, scatter, and time series) as well as a building object hierarchy in the form of a tree that allows the analyst to organize the available data points and generated plots. In particular, the update of data plots within new simulation results is a very useful feature. This update is one of the key mechanisms necessary within the EPCM used to support the iteratively adjusting of the simulation model. Thus, the current version of the SEE IT tool significantly eases the application of the EPCM. A possible future version of SEE IT will implement a number of possible future developments which the authors discussed above and further enhance the use of the EPCM in practice.

ACKNOWLEDGEMENT

REFERENCES

buildingSMART, 2011. Model - Industry Foundation Classes (IFC) — buildingSMART. Available at: http://buildingsmart.com/standards/buildingsmart-standards/ifc [Accessed April 20, 2011].

DOE, 2010. EnergyPlus - Output Details and Examples. Available at: http://apps1.eere.energy.gov/buildings/energyplus/pdfs/outputdetailsandexamples.pdf [Accessed May 16, 2011].

DOE2, 2011. DOE2.com Home Page. Available at: http://www.doe2.com/ [Accessed January 31, 2011].

Friedman, H. & Piette, M.A., 2001. *Comparative Guide to Emerging Diagnostic Tools for Large Commercial HVAC Systems*, Berkeley, CA: Lawrence Berkeley National Laboratory.

Haves, P. et al., 2008. *Development of a Model Specification for Performance Monitoring Systems for Commercial Buildings*, Lawrence Berkeley National Laboratory: Lawrence Berkeley National Laboratory. LBNL Paper LBNL-63248.

Isakson, P. & Eriksson, J., 2004. Vision of a visualization tool for commissioning. In *Proceedings of the Fourth International Conference for Enhanced Building Operations*. Paris, France.

Kircher, K., 2010. Toward the Holy Grail of Perfect Information: Lessons Learned Implementing an Energy Information System in a Commercial Building. In *2010 ACEEE Summer Study on Energy Efficiency in Buildings*. Pacific Grove, CA.

Maile, T., 2010. *Comparing measured and simulated building energy performance data*. Ph.D. Thesis. Stanford, CA: Stanford University. Available at: http://purl.stanford.edu/mk432mk7379 [Accessed August 31, 2010].

Mazzarella, L. et al., 2009. *Description of European Prototype Tool for Evaluation of Building Performance and the national tools*, Available at: http://www.buildingeq-online.net/fileadmin/user_upload/Results/BEQ_report_WP5_090823.pdf [Accessed May 17, 2011].

Reddy, T.A., 2006. Literature review on calibration of building energy simulation programs: Uses, problems, procedures, uncertainty, and tools. *ASHRAE Transactions*, 112(1), pp.226-240.

SQLite, 2011. SQLite Home Page. Available at: http://www.sqlite.org/ [Accessed May 13, 2011].

Sun Microsystems, 2011. MySQL :: The world's most popular open source database. Available at: http://www.mysql.com/ [Accessed May 13, 2011].