VISUALIZING PATTERNS IN BUILDING PERFORMANCE DATA

Paul Raftery[1] and Marcus Keane[1]
[1]Informatics Research Unit for Sustainable Engineering, NUI Galway, Ireland
Corresponding author: praftery@nuigalway.ie, research@paulraftery.com

ABSTRACT

This paper discusses issues related to the visualization of building performance data and describes a novel visualization technique that supports effective analysis of this data. The technique illustrates a dataset against two binned variables by representing a statistical property of each bin (typically the mean) using a carpet-contour plot. Contour lines and values are also overlaid on the plot so that the analyst does not need to refer to a separate legend to identify the value at a given location. The paper presents examples from three case studies using data from: a) a whole building energy model; b) a project focusing on calibration of an energy simulation model of a large building to hourly measured data; c) a Monitoring and Targeting project for an automated lighting system.

INTRODUCTION

Whole building energy models and well-monitored buildings output significant amounts of data, typically at hourly or in most cases at 15 minute intervals. This yields a minimum of 8,760 data-points per data-stream per annum. Patterns occur at multiple frequencies within these large data-sets. Furthermore, there are interdependencies between multiple discrete (hour of day, day of week, etc.) and continuous (dry-bulb temperature, dew-point temperature, etc.) variables. There are inherent difficulties in transforming and presenting this data in a coherent and comprehensible format without excessive simplification.

Common approaches to visualizing these data-sets include simple bar charts, time-series plots, and scatter plots (Reddy 2006). Other useful techniques such as density scatter plots (van Schijndel et al. 2008), box whisker mean plots (BWM) (Haberl & Abbas 1998a, Haberl & Abbas 1998b; Haberl & Bou-Sada 1998; Saelens et al. 2011), and scatter plot matrices (Baumann 2004) are less widely used. Three-dimensional plots are also very useful for visualizing building performance data, as they efficiently present a large amount of data to the viewer. Examples include Bronson et al., 1992; Haberl et al., 1993; McCray et al., 1995; Haberl and Bou-Sada, 1998; Masoero et al., 2010; Chantrelle et al., 2011.

However, three-dimensional plots can present an issue where peaks and troughs in the data-set obscure other parts of the image. This is especially problematic in hardcopy, where the viewing perspective cannot be modified. Also, it is often difficult to infer the exact position of a data-point on a three-dimensional plot. These issues can be overcome by flattening the 3-dimensional plot and using color to represent the third axis. Such plots are known as carpet plots. Examples related to building performance include Visier et al. 2004; BuildingEQ 2009; Costa et al. 2010.

VISUALISATION

Even when carpet plots are used, the sheer number of data points in an annual simulation makes it difficult to quickly visualise the data in a meaningful manner. Consider that the output from an annual simulation (8,760 data points or more) plotted hourly would require a carpet plot 365 data-points long and only 24 wide. Such ‘tapestry’ plots take significant time to review and comprehend.

This research presents a new visualisation technique that combines binning with carpet plots. Each dataset is plotted on a carpet plot(s). Each plot uses a discrete or continuous binned variable on which the data-set is dependent, for each of the x and y axes. The color typically represents the mean of the data-points at any given position in the plot. However, median, maximum, percentile, and other statistical properties of the dataset at each point in the plot can also be presented. Subsets of the data can also be plotted. A simple example of a useful subset could be plotting only the data-points that occur during a weekday, and excluding those that occur at the weekend.

However, although color is a useful method of identifying patterns in data, it can be difficult to identify the exact value at a particular location. At best, it requires the analyst to view and compare a color to a separate legend. This issue can be overcome by overlaying contour lines and values on the carpet plot. Figure 1 shows an example of such a carpet-contour plot,

Also, the carpet-contour plots only present one calculation at a time: e.g., mean, median, upper quartile, etc. To overcome this issue, the data-set can also be viewed using BWM plots if necessary. A

BWM plot can be generated for each variable (i.e., either axis of the carpet plot) and also for smaller subsets of the data (i.e., any horizontal or vertical line in the carpet-contour plot).

This technique is useful in performance analyses that require the analyst to identify and quantify patterns in large data-sets, or compare two large datasets. In addition, this technique can also be used to compare recent performance data (e.g. last week) to a database of historical performance, highlighting where recent variable values are above or below the long term average of the data under similar operating conditions (or are in the 10th or the 90th percentile of the historical data, etc.).

Thus, this technique is applicable to three main tasks related to building performance:

1. Visualising patterns in annual building performance data: Data can be either from a simulation or from an instrumented building undergoing performance monitoring. Examples of areas where this technique is useful include whole building energy simulation and performance assessment of buildings.
2. Comparing two large data-sets: There are several display options, such as plotting the percentage difference between the means of the two datasets, or plotting the positive/negative differences only. Examples of areas where this technique is useful include calibrated building simulation (comparing simulated and measured data over one or more years) and pre/post-retrofit analysis.
3. Comparing a recent value, or sequence of values, to a large baseline data-set: The recent values are compared to the baseline set at each position and are highlighted when predefined conditions are met. Conditions could be a new maximum or minimum, or new high or low outlier, etc. Examples of areas where this technique is useful include Fault Detection and Diagnosis (FDD), ongoing performance assessment of buildings, and monitoring and targeting (M&T).

The following sections demonstrate each of the above applications using benchmark models and data from case study buildings. All of the plots were created, including all the underlying data processing, using a powerful open-source tool: the R package for statistical computing (R Core Team 2011).

DEMONSTRATION 1: VISUALISING A SINGLE ANNUAL DATASET

This section gives an example of the visualisation technique, and compares it to existing techniques such as box plots. The simulated data is from a benchmark model of a large office building in Seattle, USA (US DOE 2010a) simulated using EnergyPlus v5.0 (US DOE 2010b). The simulation runs for a year and outputs data at hourly intervals.

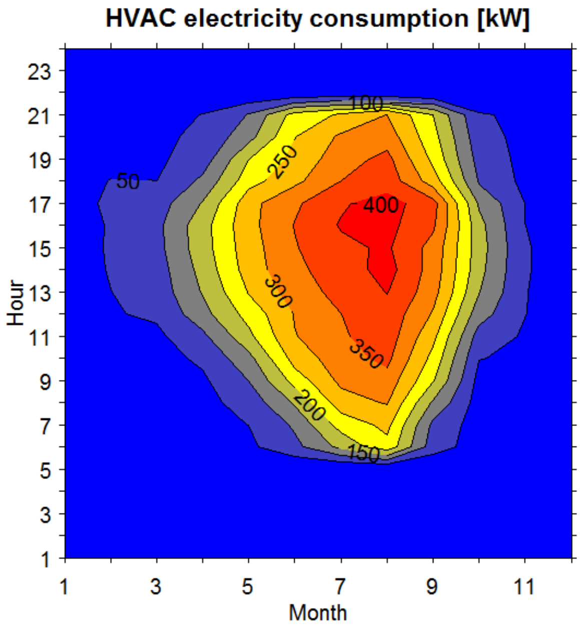

Figure 1: carpet-contour plot showing mean value of HVAC consumption against hour of day and month of year.

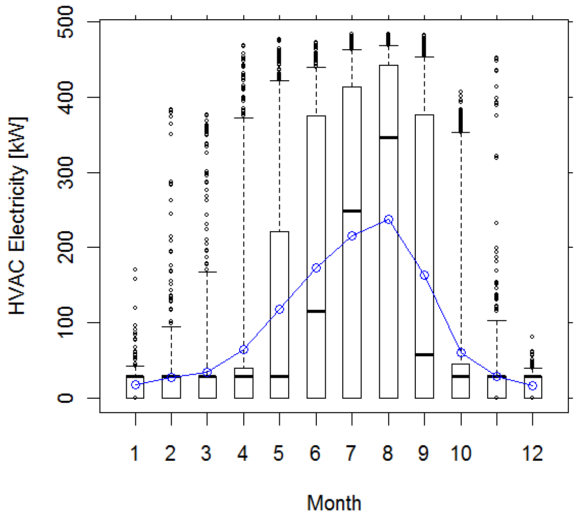

Fig 2: Monthly BWM plot of HVAC consumption

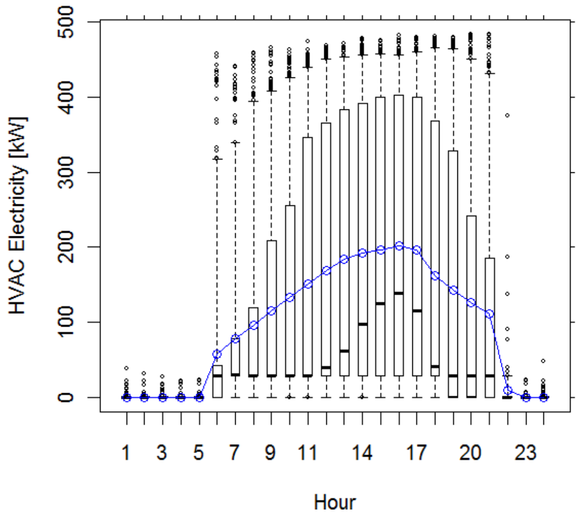

Fig 3: Hourly BWM plot of HVAC consumption

The modeled building has a total floor area of 46,320m^2, 12 floors, and an aspect ratio of 1.5:1. The HVAC system is single duct VAV with terminal reheat and ceiling return plenums. Two water-cooled chillers and a gas boiler provide heating and cooling to the building.

Figure 1 presents the mean electricity consumption of all the HVAC equipment in the model on an hour of day and month of year basis using the visualisation method described previously. The pattern of HVAC electricity consumption is clearly visible on this single plot. However, statistical properties of the data other than the means are not presented. Figures 2 and 3 present the same dataset on two BWM plots for the same variables. The thick black horizontal lines indicate the median. The outer limits of the boxes and whiskers show the 25th/75th and 5th/95th percentiles respectively. The small black circles indicate outliers, which have been defined as any points outside the whiskers. Mean values are overlaid in blue.

Figures 4, 5 and 6 show HVAC electricity consumption against outside dry-bulb temperature and solar irradiance in a similar fashion. Figures 5 and 6 would lead the viewer to conclude that the building is dependent on both solar irradiance and outdoor dry-bulb. This is true for dry-bulb temperature, however, the dependency on solar irradiance is more related to fact that the dry-bulb temperature is dependent on solar irradiance than it is related to the building's response to solar loads. This is readily apparent when viewing BWM plots of HVAC consumption with respect to solar irradiance for each bin of outdoor temperature (i.e. a BWM plot of each vertical line in Figure 4). However, the carpet-contour plot (Figure 4) clearly shows that the building is primarily dependent on outdoor dry-bulb temperature, and not solar irradiance. This is to be expected for a deep floor plan building due to the low exterior wall area to floor area ratio.

DEMONSTRATION 2: COMPARING TWO LARGE DATASETS

This section demonstrates the new visualisation technique as used to compare two large datasets (e.g. annual). Examples of where this is useful include comparing:

1. Simulated and measured data;
2. A previous year of measured data against an older year, or a multi year baseline dataset.

This demonstration is taken from a project to calibrate a detailed EnergyPlus model to hourly sub-utilities measured data using an evidence-based approach (methodology described in Raftery et al. 2009). Both the simulated and measured data are available at hourly intervals (8760 data-points per annum).

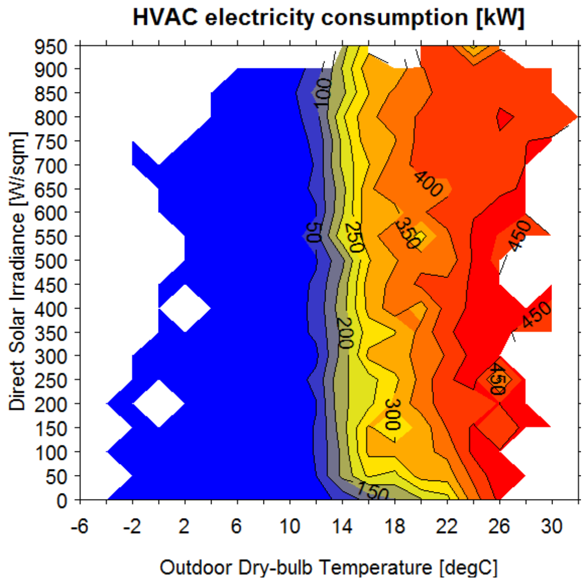

Fig 4: Carpet-contour plot showing mean value of HVAC electricity consumption against solar irradiance and outdoor dry-bulb temperature.

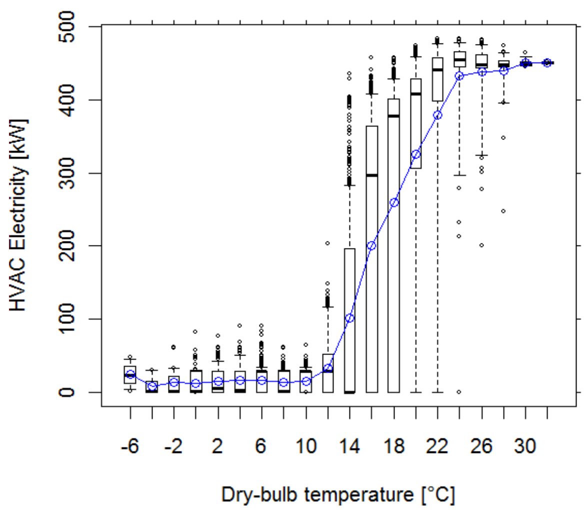

Figure 5: BWM plot of HVAC electricity against outdoor dry-bulb temperature.

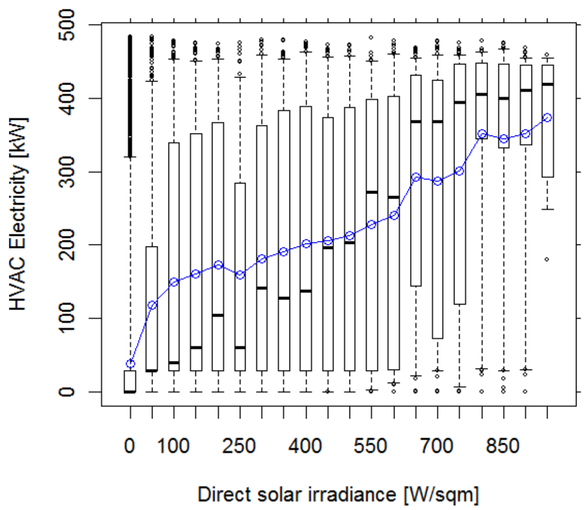

Figure 6: BWM plot of HVAC electricity against solar irradiance.

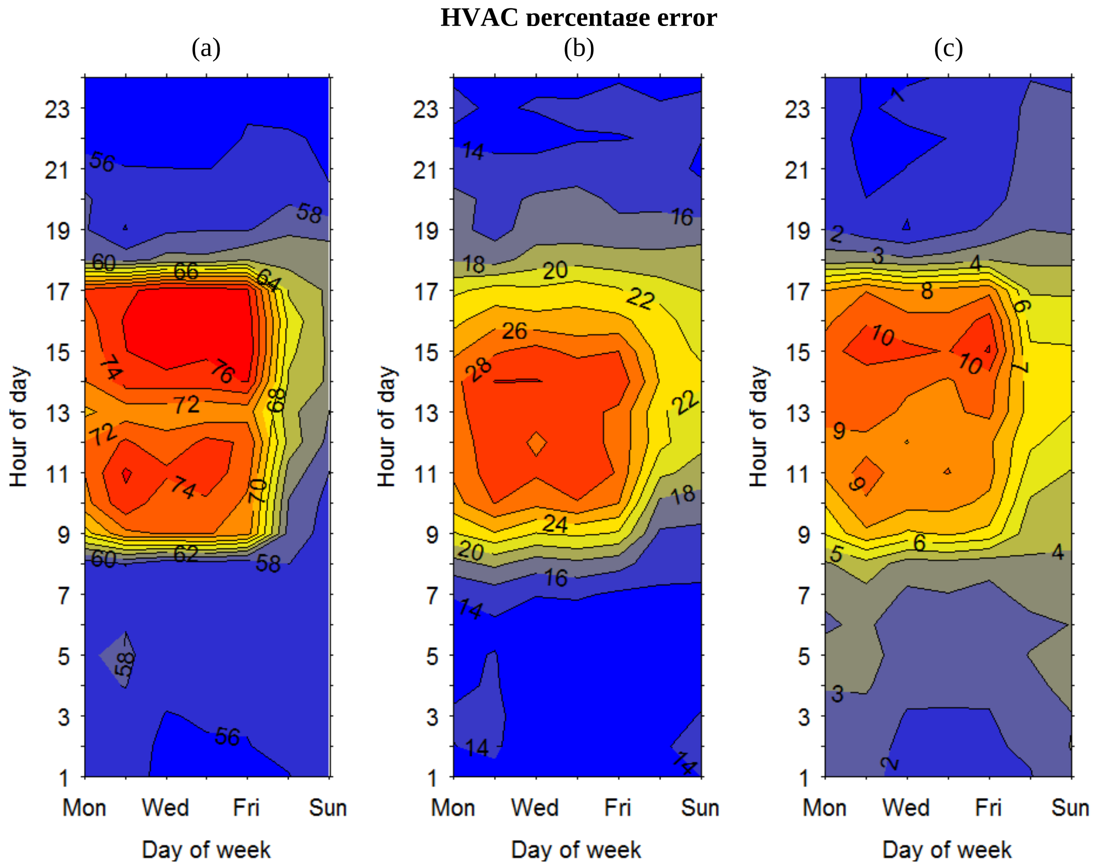

Figure 7: Mean percentage error between measured and simulated HVAC electricity consumption against hour of day and day of week for 2007. From left to right: a) model uses lighting and plug load schedules for a typical office, b) model updated to use measured lighting and plug load data in the model, c) the final model. Note that different color scales are used due to the range of variation across the images.

The 30,000m^2 case study building is located near Dublin, Ireland, has 4 floors and has an aspect ratio of 2.1:1. This industrial office building is continuously occupied on a 24 hour basis. The HVAC system is single duct VAV with terminal reheat and ceiling return plenums. District heating and cooling systems provide hot and chilled water to the building. Detailed energy monitoring system data is available for multiple years at hourly intervals, and is subdivided by end-use.

The following figures 7(a), (b) and (c) each display the results from comparing two complete annual datasets. Thus, there are two data-sets at each point on a plot - simulated data and measured data for the same building. The value displayed at each point indicates the percentage difference between the means of the two datasets. Positive and negative percentage values indicate that the model over-predicts or under-predicts energy consumption respectively. Thus, to give an example, for the year of 2007, the model depicted in figure 7(a) over-predicts HVAC energy consumption by an average of 76% on Wednesdays at 4pm, when compared to measured performance.

Figure 7(a) shows the mean percentage difference (error) between the measured electricity consumption for the HVAC equipment in the building and the results of the simulation at an early stage of the calibration process. As well as a significant constant error (the model over-predicts consumption by more than 56% at all times), the image clearly shows that there is an error related to the schedules used in the model. This scheduling error could be related to one or more of the HVAC systems, or to the scheduling of the lighting, equipment and occupancy loads on which HVAC consumption is dependent.

Later in the calibration process, it was found that the simulated internal loads were based on a typical office schedule, and were not appropriate for this building. While lighting and plug loads were reasonable when examined at a monthly resolution, when compared at higher resolution (e.g. hourly or weekly), there were large discrepancies between the measured loads and those taken from the schedules.

Once the error was identified, the model was then updated to use measured lighting and plug load data at an hourly interval in the simulation. First, constant lighting loads (e.g. corridors, emergency lighting, etc.) were identified by multiple building audits and quantified using spot measurements. These constant loads were added to the model and removed from the measured load data. The remainder of the measured load data was then applied on a floor-area weighted basis. These values were also checked to ensure that the maximum power density in the model did not exceed the installed lighting power density in each zone. Figure 7(b) shows the same comparison as Figure 7(a) after the model update.

Significant discrepancies clearly remain between the model and the real building, as would be expected in the early stages of any calibration process (Figure 7(b)). Later stages identified improved occupancy profiles based on surveys and Human Resources data, corrected HVAC equipment parameters based on commissioning documents and physical surveys, etc. These combined changes reduce the overall error significantly, as can be seen in Figure 7(c). However, a detailed discussion of the calibration project is outside the scope of this paper (a journal paper describing it in detail is under review), and the above figures adequately serve the purpose of demonstrating the comparison of two large datasets using this visualization method.

It is worth noting an issue which arose in using these plots for comparison across the wide ranges of variation between the three plots in Figure 7. If a single scale, starting at zero, was used for each plot, there would either be a large interval between color changes (e.g. 5%) which would yield a limited number of levels in each plot (particularly for Figure 7(b)), or a large number of colors which would make it difficult to distinguish changes between intervals. Thus, the color scale is different on each plot, in order to highlight the pattern in the data, and the values on the contours give the magnitude at each point across the plots.

DEMONSTRATION 3: COMPARING RECENT VALUES AGAINST AN EXISTING DATASET

This section demonstrates the novel visualisation technique as used to compare recent performance data to a large historical dataset (a baseline). The case study example illustrates how the visualisation technique was used in Monitoring and Targeting (M&T) in a large building relating to lighting. The first step is to bin the baseline data according to the two most influential variables and generate a carpet-contour plot (similar to figures 1, 4 & 7). The second stage is to compare the recent data to the baseline dataset, and to highlight where the recent data is significantly outside the expected range of values.

Lighting systems

Figure 8 shows the electricity consumption of the lighting system taking a 6 month period before the

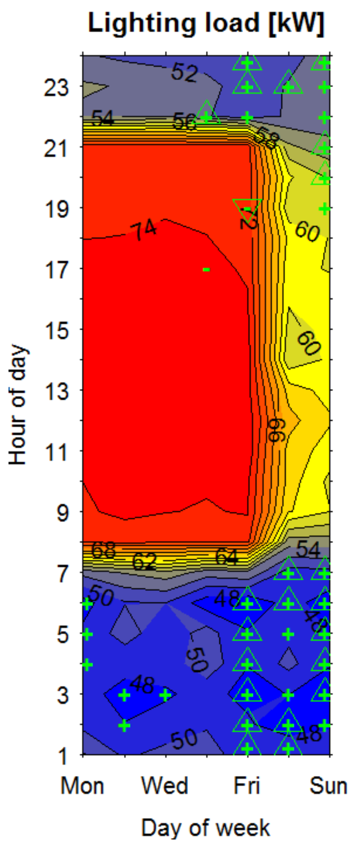

Figure 8: Carpet-contour plot showing mean lighting electricity consumption against hour of day and day of week. The green symbols (overlaid) show how the most recent week compares to the existing baseline.

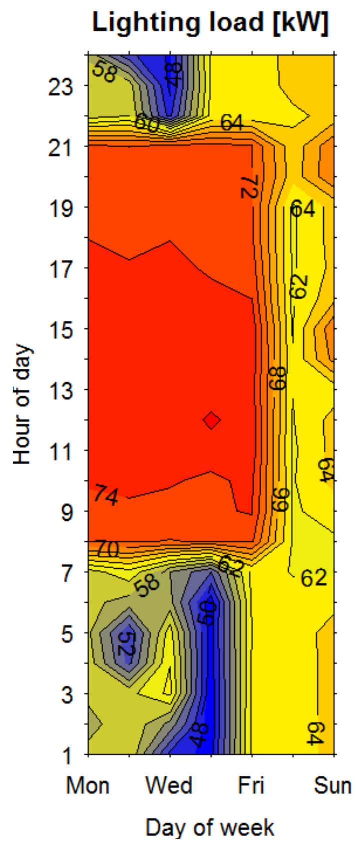

Figure 9: Carpet-contour plot showing lighting electricity consumption against hour of day and day of week (for the most the week in which the fault occurs). Color scale is identical to that in Figure 8.

fault as a baseline. The green markers are the results of a comparison of the data obtained in the week immediately after the end of the baseline. This week includes the day on which the fault occurred: 29th June 2007.

In Figure 8, the '-' or '+' symbols indicate that the most recent data-point is in the 5th or 95th percentile of the data-set at that location. The '▽' or '△' symbols indicate a new minimum or maximum. Thus, a number of these high or low symbols in a sequence indicate a significant change between the recent and baseline data.

Thus, Figure 8 shows that, compared to the baseline, there was an large increase in power consumption in the second half of the week. In this case, the fault was a change to the lighting schedules over part of the floor area. The schedule change was intended to be a temporary one, to accommodate temporary overtime during the night and weekend. However, this new schedule remained in place for more than a year, at an unnecessary electricity cost of approximately €6,500 per annum. This excludes the additional cost of the additional cooling load on the HVAC system.

Displaying a carpet-contour plot of the baseline dataset mean values and comparing that to a separate plot of the recent data is also useful in order to see the exact values for a particular week. However, without comparing the recent data to the statistical properties of the baseline dataset, an analyst has no way of knowing whether the behaviour is unusual or within the normal range of variation. Figure 9 presents such a plot, showing the electricity consumption of the lighting system during the week in which the fault occurs. As the range of variation is similar between it and Figure 8 (unlike in the three images in demonstration 2), the same color scale as the baseline image is used for clarity.

CONCLUSION

In conclusion, this paper presents a novel visualisation technique that is useful in transforming and analysing building performance data. The technique plots a dataset against two binned variables by representing a statistical property of each bin (typically the mean) using a carpet-contour plot. Contour lines and values are also overlaid on the plot so that the analyst does not need to refer to a separate legend to identify the value at a given location.

This approach is most applicable in cases where the dataset is highly dependent on two variables which are independent (e.g. day of week and hour of day), or almost independent (e.g. day of week and outdoor temperature), of each other. A useful aspect is that the plots can sometimes highlight erroneous conclusions when the two variables are slightly correlated (e.g. the discussion of Figures 4 to 6).

The technique works best for variables with regular, repeated intervals (such as hour of day, day of week, etc.). As two variables are presented simultaneously, this approach is very useful for identifying schedule related issues, whether in models or in real building systems. The technique is also useful for irregular variables (e.g. weather related), and combinations of these and schedule related variables.

The technique outlined in the 3rd demonstration is useful in ongoing performance assessment as it reliably highlights when a system is performing unexpectedly with respect to the two variables selected for the x and y axes of the plot. However, in order for this method to correctly identify faults, this approach requires a period of 'correct' operation with which to compare the current performance, which can often be a significant issue.

FUTURE WORK

Future work will focus on further developing the R scripts used in this paper to:

1. Automatically generate BWM lines upon mouse-over of any horizontal or vertical line (or axis) of a carpet-contour plot;
2. Output a table of the data-points (and statistical summaries of the data-set) upon mouse-click at any position on the carpet-contour plot.

Later work will also focus on reviewing the use of color in data presentation. The figures in this paper use a simple color palette which interpolates between blue, yellow, and red. However, a color palette where equal perceived color changes correspond to equal changes in the underlying variable would further improve these figures.

Also, in addition to the identification of new maxima/minima, the techniques in the third demonstration section could be expanded to identify a gradual decay in performance of a system. This would be applicable in cases such as ongoing performance monitoring of air handling units (due to clogged filters, passing valves, etc.).

ACKNOWLEDGEMENT

The authors would like to thank Sebastian Burhenne at the Fraunhofer ISE, Freiburg, Germany for giving us our initial introduction to the R package. This research was funded by Science Foundation Ireland through the ITOBO project (07/SRC/I1170) and by Enterprise Ireland as part of the I2E2 HVAC Appropriate Working Environments project.

REFERENCES

Baumann, O., 2004. Operation diagnostics – use of operation patterns to verify and optimize building and system operation. In Proceedings of International Conference for Enhanced Building Operations 2004 Conference. Paris, France.

Bronson, D.J. et al., 1992. Procedure for Calibrating the DOE-2 Simulation Program to Non-

Weather-Dependent Measured Loads. ASHRAE Transactions, 98(1), pp.636-652.

BuildingEQ, 2009. WP 6: Report on energy savings, Technical report, CO2 reduction and the practicability and cost benefit of developed tools and continuous commissioning, Available at: http://www.buildingeq-online.net/.

Chantrelle, F.P. et al., 2011. Development of a multicriteria tool for optimizing the renovation of buildings. Applied Energy, 88(4), pp.1386-1394.

Costa, A. et al., 2010. Building Operation & Energy Performance: Monitoring, Analysis andOptimisation Toolkit. In Proceedings of the 2nd International Conference on Applied Energy Conference. International Conference on Applied Energy. Singapore.

Haberl, J.S. & Abbas, M., 1998a. Development of Graphical Indices for Viewing Building Energy Data: Part 1. Journal of Solar Energy Engineering, 120(3), pp.156-161.

Haberl, J.S. & Abbas, M., 1998b. Development of Graphical Indices for Viewing Building Energy Data: Part 2. Journal of Solar Energy Engineering, 120(3), pp.162-167.

Haberl, J.S. & Bou-Sada, T.E., 1998. Procedures for Calibrating Hourly Simulation Models to Measured Building Energy and Environmental Data. ASME Journal of Solar Energy Engineering, 120(3), pp.193-234.

Haberl, J.S. et al., 1993. Graphical tools to help calibrate the DOE-2 simulation program. ASHRAE Journal, 35(1), pp.27-32.

Masoero, M., Silvi, C. & Toniolo, J., 2010. Assessing the energy performance of HVAC systems in the tertiary building sector by on-site monitoring. In Proceedings of 6th International Conference on Improving Energy Efficiency in Commercial Buildings. Frankfurt, Germany.

McCray, J.A. et al., 1995. Using data visualization tools for the calibration of hourly DOE-2.1 simulations. In Proceedings of the 5th International Building Performance Simulation Association Conference. Madison, Wisconsin, USA.

R Core Team, 2011. The R Project for Statistical Computing. Available at: http://www.r-project.org/.

Raftery, P., Keane, M. & Costa, A., 2009. Calibration Of A Detailed Simulation Model To Energy Monitoring System Data: A Methodology And Case Study. In Proceedings of the 11th International Building Performance Simulation Association Conference. University of Strathclyde, Glasgow, Scotland.

Reddy, T., 2006. Literature Review on Calibration of Building Energy Simulation Programs: Uses, Problems, Procedures, Uncertainty and Tools. ASHRAE Transactions, 112(1), pp.226-240.

Saelens, D., Parys, W. & Baetens, R., 2011. Energy and comfort performance of thermally activated building systems including occupant behavior. Building and Environment, 46(4), pp.835-848.

van Schijndel, A.W.M. et al., 2008. Application of an integrated indoor climate, HVAC and showcase model for the indoor climate performance of a museum. Energy and Buildings, 40(4), pp.647-653.

US DOE, 2010a. Commercial Building Reference Models. Available at: http://www1.eere.energy.gov/buildings/commercial_initiative/new_construction.html.

US DOE, 2010b. EnergyPlus v5.0. Available at: http://apps1.eere.energy.gov/buildings/energyplus/.

Visier, J.C. et al., 2004. Commissioning of buildings and HVAC systems for improved energy performance: Results of IEA ECBCS ANNEX 40, Technical Report, Available at: http://www.commissioning-hvac.org/