Navigating multi-dimensional results from large parametric building simulation studies

Graeme Stuart, Ivan Korolija
Institute of Energy and Sustainable Development, De Montfort University, Leicester, UK
gstuart@dmu.ac.uk

Ljiljana Marjanovic-Halburd
Bartlett School of Graduate Studies, University College London, London, UK

Abstract

Advances in computing in recent years allow for many thousands of building energy simulations to be computed in the time previously required for a single simulation run. Software tools exist that allow for a single input file to be modified in a number of different ways to generate thousands of self-similar input files which can then be automatically simulated. The problem with this approach is not the simulation time but the time and effort required for the analysis of the vast set of results generated.

Large, multi-dimensional result sets cannot be easily visualised as a whole. One approach is to view the results as a non-linear, interactive document in which only a small part of the results is viewed at any one time. With the addition of simple navigation to select the next sample to view, this approach allows the analyst to easily browse the large result set. More concretely, a one-dimensional sample (a selection of simulations which vary in only one aspect) can be selected from the dataset and visualised as a simple bar chart. Simple rules can then be applied to identify a collection of similar, one-dimensional samples for navigation.

To examine this approach, a prototype tool was developed as a web-based application. The basis for this tool was a multi-parameter simulation study of office building energy consumption including 1,440 individual simulations varying across six dimensions including four building types, five building fabrics, three percentages of glazing, the inclusion of daylight control, two glazing types and six HVAC system types (including building load calculations). The tool included a basic report comparing a one-dimensional sample of results and a detailed report showing time series results for an individual case. Navigation panels allowed for simple traversal of the results set and to move between the two reports. The tool was found to be very useful for navigating the multi-dimensional data and the method is generic enough to be transferable to similar datasets.

Keywords Graphical User Interface, Multi-dimensional data, Simulation, office buildings, energy consumption

1.0 Introduction

Optimising building energy consumption at early design stages is becoming increasingly important. However, in the early stages of the design process many design parameters may be unknown and there is flexibility to steer the design to a more energy efficient configuration by adjusting parameters such as glazing ratio, orientation, construction materials, etc.

A single simulation can demonstrate that a given design is fit for purpose but is of limited use in terms of influencing building design. For example, it is possible to calculate the energy consumption or hours of overheating for a given design but a single simulation cannot reveal whether the given design is optimal. What is typically required is to simulate a range of alternative designs and compare the outputs in order to inform design decisions.

The influence on energy performance of a single parameter such as glazing ratio or the thermal properties of a construction element depends on the specific building design. One approach is to focus on a single aspect of the design and conduct several simulations where the chosen aspect is varied and the remainder of the model description is kept constant. Such a study will generate a collection of outputs which reveal the impact of making the specified changes. For example, Bansal and Bhattacharya [***i***] investigated maximum heating and cooling loads and the impact of insulation thickness by simulating each of ten thicknesses. The study also investigated the effect of changes to the surface to volume ratio by varying building height, depth and width. Each parameter was investigated individually and presented as a series of scatter charts.

Building parameters such as shape, construction, percentage and type of glazing, HVAC system, orientation, control etc. can be varied from a number of base case scenarios to generate large numbers of similar simulations. For example Korolija et al [***ii***] conducted a study in three dimensions whereby the suitability of different HVAC systems was considered for open plan and cellular office buildings with and without daylight control of lighting.

Buildings and the associated HVAC systems are complex and there are a large number of parameters which can affect the energy performance of a building design. Lam et al [***iii***, ***iv***] identified 62 parameters related to building envelope and HVAC systems (both primary and secondary). 28 of these parameters were found to be correlated well with annual energy consumption and after a sensitivity analysis was conducted, 12 of these were selected as being the most significant design variables. Having a large number of significant parameters creates a significant problem if we want to explore all possible variations of a given design. For example, to simulate only three variants of each of twelve design parameters and all combinations of these variants would require over half a million individual simulation runs. This kind of parametric study can generate large amounts of data; these data and their analysis are the subjects of this paper.

A wider view of the range of potential building designs can be gained by varying a number of design parameters over realistic ranges and conducting large numbers of simulations where each unique combination of the varied parameters is simulated. These combinations define a kind of 'design-space' in which all possible alternatives are considered. Advances in computing in recent years [***v***] allow for many thousands of building energy simulations to be computed in the time previously required for a single simulation run. Software tools exist [***vi***] that allow for a single input file to be modified in a number of different ways to generate large numbers of self-similar input files which can then be automatically simulated. These enabling factors have made multi-dimensional, parametric, building energy simulation studies a reality.

In general the procedure for conducting a building energy simulation is composed of three main steps: preparation of simulation input; the simulation runs themselves;

and analysis of results [***vii***]. Each of these steps requires both time and resources to complete. The extra effort required to upscale a study from one scenario to several thousand is relatively small, the only problem is the large volume of data that needs to be analysed.

The time taken to design the thousands of input files is slightly more than that required for designing a single base case. Each input file is similar to the original and only the changes to the original will require extra effort. In many cases this may be as simple as changing the value of a single variable or designing a small number of alternative components that can be swapped in.

The time required for the actual simulation runs themselves is dependent on the number of simulations, the model complexity and the computing power available. With modern high powered parallel computing, thousands of moderately complex simulations can be conducted in a matter of days.

In contrast, the time and effort required for the analysis of these vast sets of results increases with every simulation. Multiplying the number of simulations increases the volume of data generated and the sheer volume of data can be overwhelming. Very large numbers of simulations present the danger of data overload and require significant effort to analyse. Building design professionals wanting to review the results of such a study will be presented with a significant challenge in simply viewing the results let alone conducting any meaningful analysis.

2.0 Representation of multi-dimensional results

In general, a multi-dimensional parametric study consists of an n-dimensional array (an n-hyperrectangle or n-orthotope [***viii***]) of scenarios. The simplest case, a one-dimensional study consists of a simple list of scenarios. A two-dimensional study can be visualised as a simple table of scenarios with the values of the two dimensions as column and row headings. Similarly, a three dimensional study can be visualised as a cube. Higher dimensionality becomes difficult to visualise but is easy to represent programmatically and mathematically.

Each dimension refers to an individual parameter and each cell represents a scenario and associated energy simulation run. Each simulation produces a user-defined set of outputs. These may be building loads, temperatures, thermal comfort parameters or HVAC system parameters (flow rate, temperatures etc.) reported at a given time step.

Even when only a handful of parameters are being varied this can amount to a large volume of data. For example, if eight building orientations are considered, three glazing ratios, six building fabrics and six HVAC systems then 864 simulations are generated. If each simulation generates hourly time series data over a full year then the result will include 7,568,640 hourly records and each record may contain several output fields. This adds up to a lot of data to analyse for just a small study over four dimensions.

Large volume data analysis may be conducted using a range of methods. The selected method will largely depend on the aim of the study. Smith et al [***ix***]included two parameters in a study of the influence of future climate projections on building energy simulations. Two building models were simulated with 1,500 weather files each. Due to the large number of weather files the results were presented as a mean and standard deviation for each building design.

As part of a study into the effects of future climate on air-conditioned office buildings, Chow and Levermore [**x**] varied both building design parameters and weather files. The results were presented in an aggregated form by selecting two parameters (building construction age and future climate year) to include on the x-axis and showing the range of values (for example heating/cooling degree days) within each category on the y-axis.

Descriptive statistics such as mean and standard deviation can provide summary information about the whole dataset and selected subsets. Another possibility is to employ regression analysis (for example Lam [iii,iv]) to develop predictive models from the chosen parameter values. Sanders et al [***xi***] developed a simplified regression model from a four-dimensional parametric simulation including three building design parameters (six values for each dimension) and 25 weather locations totalling 5,400 simulation runs.

However, such analytical methods will always obfuscate some detail. In some cases it is desirable to visualise the outputs of individual scenarios and make a direct comparison between scenarios.

With 1-dimensional result sets (i.e. a comparison between designs varying in only one aspect) a simple bar chart can present the annual data very effectively. Figure 1 shows two charts. The first is a simple comparison of annual consumption between scenarios with different building fabrics. A direct comparison can be made between each scenario and it is relatively easy to identify the individual scenario characteristics. Even with the simple one-dimensional visualisation it is not entirely obvious which scenario has e.g. the largest cooling requirement.

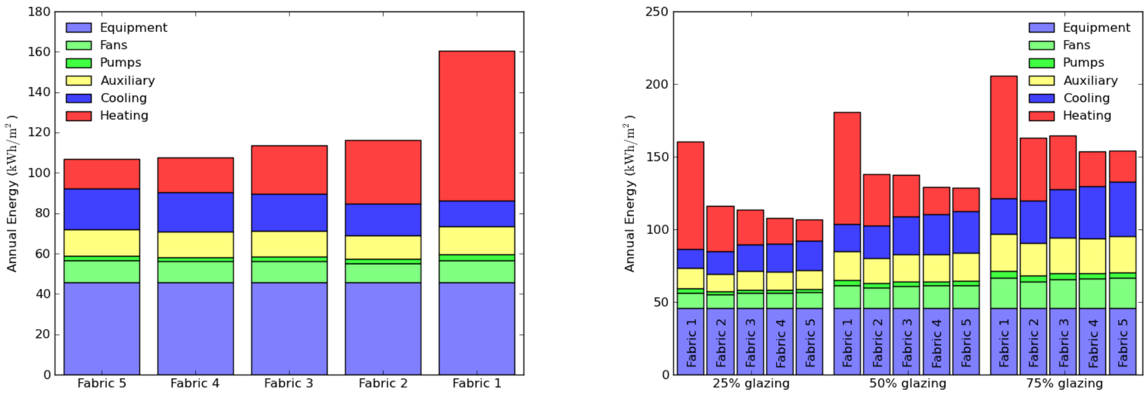


Figure 1: Bar charts showing the affect of varying building fabric alone and in combination with glazing ratio

The second chart shown in Figure 1 is a clustered bar chart showing the affect of varying both building fabric and glazing ratio. The clustering can be done in either of two ways, three groups of five fabric types or five groups of three glazing ratios. In the example the glazing ratios are grouped making it easy to discern the effect of changing the fabric with a given glazing ratio but more difficult to discern the effect of changing the glazing ratio with a given fabric.

Figure 2 shows two additional dimensions (the effect of daylighting control and the addition of a reflective coating on the glazing) by essentially repeating the clustered chart within a two dimensional grid (in this case with two rows and two columns).

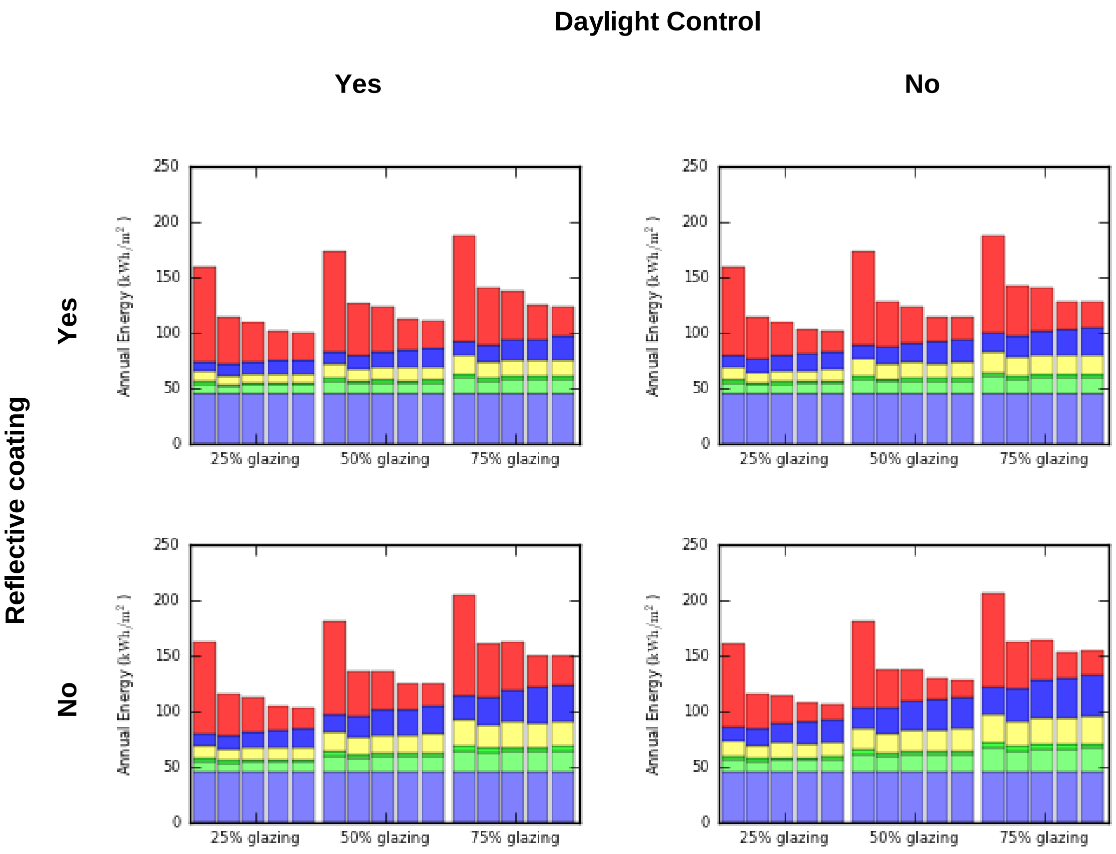
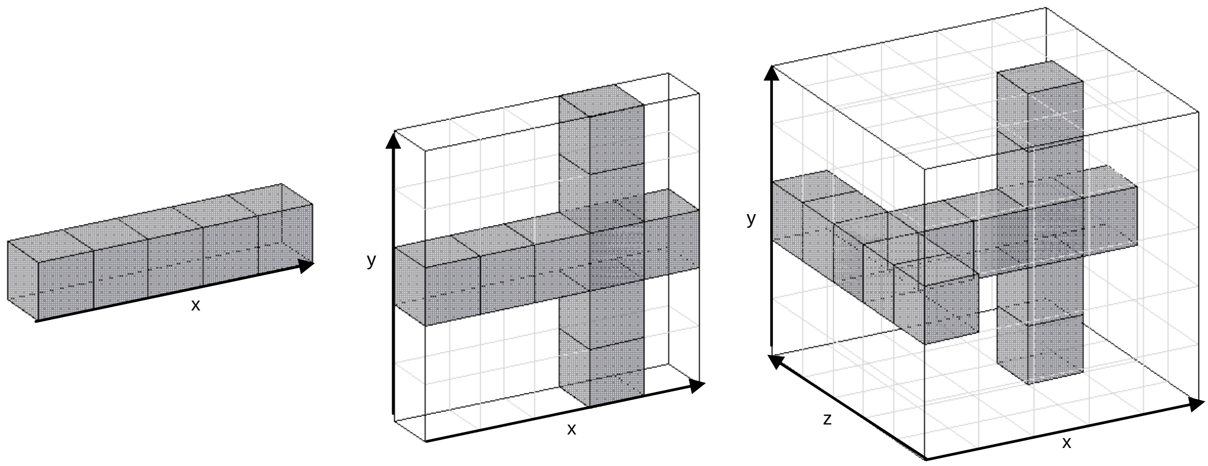


Figure 2: Example table of clustered bar charts (4-Dimensional)

Thus, four dimensions (four parameters) are now visualised. However, the ability to make direct comparison is now compromised further as there are now two identical but physically separate vertical scales. To compare scenarios with different glazing types it is necessary to move from one scale to the next.

In general, increasing the number of dimensions makes visualisation more difficult. As the examples demonstrate, a direct comparison becomes difficult when the data are not presented on the same physical scale and when they are not presented together. The ability to compare the differences between design scenarios, even when the differences are subtle, maybe be very important, especially if the interdependence of design parameters is not clear.

One solution to the problem of large numbers of dimensions is to visualise simulation results for a specifically selected one-dimensional sample of scenarios. For example, picking a one-dimensional sample from a two-dimensional dataset entails choosing either a single row or a single column. Similarly, a one-dimensional subset can be extracted from a three-dimensional dataset. Examples of one-dimensional samples taken from one, two and three dimensional datasets are illustrated in Figure 3 as shaded areas. The concept extends to n-dimensions; it is always possible to extract a one-dimensional sample.



a) 1-dimensional data b) 2-dimensional data c) 3-dimensional data

Figure 3: One-dimensional subsets taken from 1, 2, and 3-dimensional data

By selecting a one-dimensional subset from a larger dataset the results can be displayed in a simple bar chart without the complexity of the extra dimensions. However, there may be very large numbers of such subsets and to produce a chart for each of them would, in many cases generate enough charts to fill the pages of a very large static document which would be extremely difficult to navigate.

Navigation through all possible one-dimensional subsets within an n-dimensional result set in a static, linear document would be very difficult and reliant upon a complex indexing system. An interactive, non-linear document (e.g. a dynamically generated hypertext document) allows for a few simple navigation rules which can provide full coverage of a higher dimensional dataset by viewing the data one dimension at a time.

In the simplest case a one-dimensional subset can be defined by choosing a dimension to vary and specifying values for all other dimensions. This provides all the information required to extract data for a simple bar chart as in Figure 1. Figure 3b illustrates this in two dimensions. Two example subsets are highlighted, one in each dimension. The subset which varies in the x-axis has a value of y=3, the subset which varies in the y-axis has a value of x=4.

In Figure 3c a third dimension is added and three subsets are highlighted. Two are the same as in Figure 3b except they also have a value of z=3. A third subset is also shown that varies in the new third dimension and has a value of y=3 and x=1.

It is possible to navigate across the dataset in an intuitive manner by changing the value of any one dimension. If a value is specified for the dimension over which the sample is varied then an alternative dimension must be selected to vary.

Applying these rules in the two-dimensional example leads to two ways to move from one sample to another. Either the selected row or column is moved, or the selection is flipped from a row to a column. In each case the newly selected sample will be closely related to the original.

Figure 3c shows three subsets in three-dimensions. It is possible to navigate between these subsets using the rules described above. For example, starting at sample (x=4, y varies, z=3) we can select y=3 and rotate the variation to the x-axis (x varies, y=3, z=3) then, set x=1 and change the variation to the z-axis (x=1, y=3, z varies). These navigational rules define and limit the collection of subsets that can be reached from any given subset in one step. The rules generalise to the n-dimensional case.

3.0 Results from the case study

A subset of 1,440 building simulations from a large parametric study of office buildings was used to illustrate the basic principles described above. The data were generated over six dimensions including four building types, five building fabrics, three percentages of glazing, the inclusion of daylight control, two glazing types and six HVAC system types (including building load calculations).

A web-based tool was developed to navigate through this dataset and visualise one-dimensional subsets as simple stacked bar charts including heating, cooling, pumps, fans and lighting/equipment energy requirements. Figure 4 shows the main page of the tool showing the results of one selected subset of data.

Figure 4: Main scenario comparison page

The page is split into three main areas. At the top of the page is a title which shows the dimension being inspected (in this case six HVAC systems are compared) and the associated parameters which define the one-dimensional subset (in this case building type 1, fabric type 1, 25% normal glazing with no daylight control).

The central area includes six bar charts, one showing the total annual energy requirements and one each for the various sub-elements. The bar charts include data for all six HVAC systems in the chosen dimension.

Navigation panels appear on each side. There are six sub-panels, one for each dimension. Each sub-panel includes a bold link named 'by [dimension]' which is used to select the dimension which is compared. The panel associated with the currently selected dimension is disabled and greyed out (in this case the 'by system' panel). Each sub-panel also contains a link for each value that the dimension can take. For example, the glazing ratio can be set to 25%, 50% or 75%. The link for the currently selected value for each dimension is also disabled and greyed out (25% in this case).

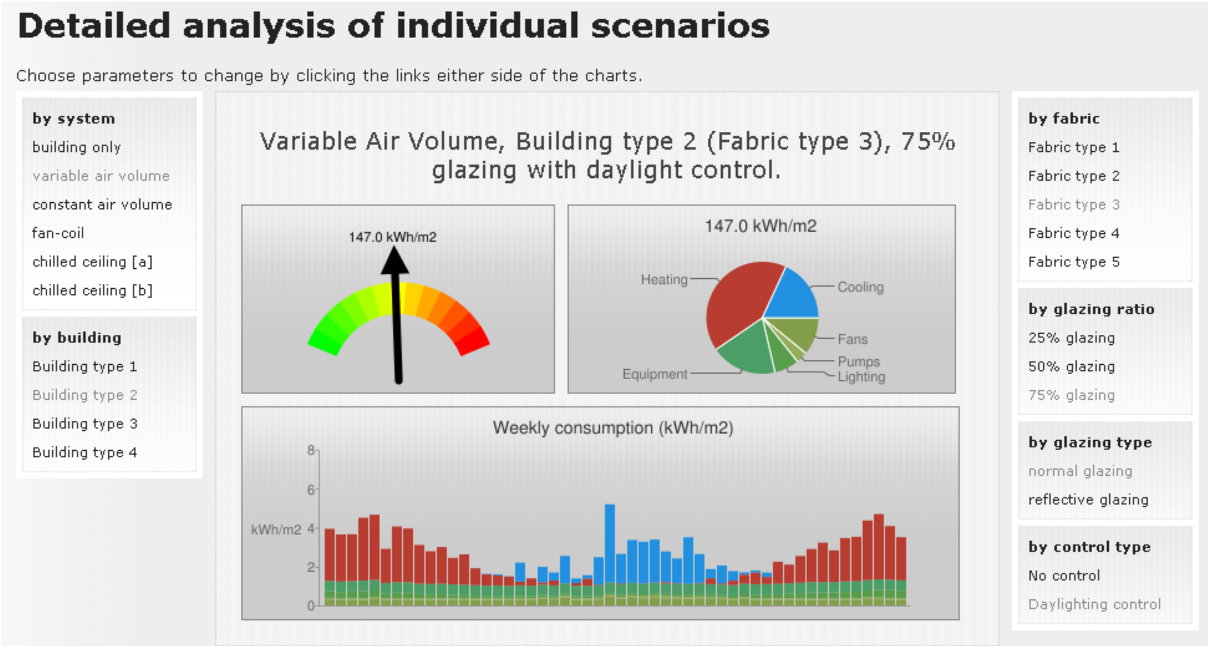


Figure 5: Individual scenario page

In addition, the tool allowed for individual scenarios to be selected and viewed in detail. Figure 5 shows the detailed analysis page which has the same format as the main page. The figure is dominated by a panel in the centre surrounded by two navigation panels. It also includes a title which describes the specific scenario.

The navigation panels include identical links to that on the main page. However, since no dimension is selected none of the sub-panels are disabled. Clicking on a dimension value will reload the page with the corresponding scenario. Selecting a 'by [dimension]' link will redirect back to the corresponding comparison on the main page.

The central panel includes three charts. In the top left is a meter which indicates the level of energy requirements for the selected scenario relative to the range covered by the whole dataset. The top right pie chart shows the break down of annual energy requirements by category. At the bottom is a stacked bar chart of weekly time series data broken down by category.

Conclusion

The ability to conduct very large numbers of simulations has introduced a significant problem in terms of visualising and analysing the very large datasets produced as presenting large sets of data in static forms can be impractical and unhelpful.

Interactive documents such as web-based systems can be designed to provide specific views of these large datasets that facilitate their analysis but also make them available to a wider audience. Very large datasets can be navigated in an intuitive manner allowing the impact of any of the varied parameters to be inspected.

The case study presented which included 365×24×1440 data points is a simple and specific example of what can be done. There is great scope for this kind of approach to be generalised in terms of building generic tools for absorbing and visualising the results of large parametric building simulation studies. There is also scope for very specific tools to be designed for individual studies whereby a particular view (or suite of views) is required to identify the complex interactions of adjusting values on large numbers of dimensions.

This study has many limitations, the small number of dimensions and the small size of each dimension in the dataset made it relatively easy to develop a navigation menu. With an order of magnitude more dimensions the space available for hyperlinks would become an issue.

Also, the study only developed a one dimensional view of the data shown in Figure 4. It is certainly possible to display results in clustered bar charts and provide a two-dimensional or even three-dimensional window into the results.

Acknowledgements

The authors would like to acknowledge financial support of this work which forms part of the CITYNET project funded via the Marie Curie Research Training Network.

References

i Bansal, N. & Bhattacharya, A. (2009), 'Parametric equations for energy and load estimations for buildings in India', *Applied Thermal Engineering* **29**(17-18), 3710 - 3715.

ii Korolija, I.; Marjanovic-Halburd, L.; Zhang, Y. & Hanby, V. I. (2011), 'Influence of building parameters and HVAC systems coupling on building energy performance', *Energy and Buildings* **43**(6), 1247 - 1253.

iii Lam, J. C.; Hui, S. C. & Chan, A. L. (1997), 'Regression analysis of high-rise fully air-conditioned office buildings', *Energy and Buildings* **26**(2), 189 - 197.

iv Lam, J.C., W. K. L. D. T. C. (2010), 'Multiple regression models for energy use in air-conditioned office buildings in different climates', *Energy Conversion and Management* **51**(12), 2692-2697.

v Moore, G. E. (1965), 'Cramming more components onto integrated circuits', *Electronics* **38**(8), 114 – 117

vi Zhang, Y., Korolija, I (2010) "Performing complex parametric simulations with jEPlus", SET2010 - 9th International Conference on Sustainable Energy Technologies, 24-27 August 2010, Shanghai, China

vii Bazjanac, V. (2001), Aquisition of building geometry in the simulation of energy performance, Seventh international IBPSA conference, Rio de Janeiro

viii Coxeter, H. S. M. (1973), Regular Polytopes, Dover.

ix Smith et al (2009), Influence of probabilistic climate projections on building energy simulation, CISBAT 2009, Lausanne

x Chow, D.H., L. G. (2010), 'The effects of future climate change on heating and cooling demands in office buildings in the UK', *Building Services Engineering Research and Technology* **31**(4), 307-323.

xi Sander et al (1993), Development of a simple model to relate heating and cooling energy to building envelope thermal characteristics