Analysing Populations of Design Variants Using Clustering and Archetypal Analysis

Kian Wee Chen[1], Patrick Janssen[2], Arno Schlueter[3]
[1]Future Cities Laboratory, Department of Architecture, ETH Zurich, Switzerland [2]Department of Architecture, National University of Singapore [3]Institute of Technology in Architecture, Department of Architecture Zurich
[1]chenkianwee@gmail.com [2]patrick@janssen.name
[3]schlueter@arch.ethz.ch

In order to support exploration in the early stages of the design process, researchers have proposed the use of population-based multi-objective optimisation algorithms. This paper focuses on analysing the resulting population of design variants in order to gain insights into the relationship between architectural features and design performance. The proposed analysis method uses a combination of k-means clustering and Archetypal Analysis in order to partition the population of design variants into clusters and then to extract exemplars for each cluster. The results of the analysis are then visualised as a set of charts and as design models. A demonstration of the method is presented that explores how self-shading geometry, envelope materials, and window area affect the overall performance of a simplified building type. The demonstration shows that although it is possible to derive general knowledge linking architectural features to design performance, the process is still not straightforward. The paper ends with a discussion on how the method can be further improved.

Keywords: *K-means clustering, Archetypal analysis, Design optimisation, Performance-based design, Computational design*

INTRODUCTION

The early architectural design stages tend to be ill-defined and explorative in nature. Architects will typically explore the project brief and design proposals simultaneously, with the problems and solutions feeding into each other to define the boundaries of what is possible (Harfield 2007, Lawson 2004).

In order to support this exploratory process, researchers have proposed the use of population-based optimisation algorithms to search for a set of well-performing design variants (Caldas 2008; Flager et al. 2009; Janssen et al. 2011; Lin and Gerber, 2014; Turrin et al. 2011). Such algorithms optimise a population of design variants based on a set of performance objectives. In an optimisation process, the design variants are generated from a parametric model based on different input parameters (Woodbury, 2010). Objective functions are then used to calculate the performance scores for each design variant. Once a population of optimised design variants has been created, architects need to be able to analyse this population. Ideally, the analysis results should give architects a better understanding of the relationship between architectural features and design performance.

Common techniques used for analysing the optimised design variants include sorting, filtering, and Pareto ranking. Essentially, these techniques filter the population, in order for architects to select a small number of design variants for further development. However, even after the filtering process, there will typically still be a large number of design variants that remain. The selection of design variants is not a straightforward process. As a result, more advanced techniques such as Multiple Criteria Decision Analysis (MCDA) (Mela et al. 2012; Pohekar and Ramachandran 2004) and Knowledge-Based Design Support System (KBDSS) (Singhaputtangkul et al. 2013) are used to support the architects in narrowing down the selection to a manageable number of design variants for further design development.

There are also techniques that extract design principles through analysing the design variants (Chichakly and Eppstein, 2013; Deb et al., 2014; Deb and Srinivasan, 2006). With these techniques, design principles are derived from the relationship between the input parameters and performance scores. These techniques are proposed for engineering design where the input parameters have a direct relationship with the performances. In early architectural design stages, architects are exploring design not only in terms of performances, but also qualitative aspects such as aesthetics. Some input parameters are modelled for the qualitative aspects of the design and have an indirect relationship to the performances. Thus, these techniques are not appropriate for the early architectural design stages.

This paper proposes a method for analysing populations of design variants through the use of Cluster Analysis and Archetypal Analysis. Cluster Analysis (Everitt and Hothorn 2011; Han et al. 2012) partitions a set of data into subsets of data or clusters such that the individual units of data in each cluster are similar to each other, while different from those in the other clusters. It is used to gain insights into the distribution of a set of data, observes characteristics unique to each cluster, and helps identify clusters of interest for further analysis. Archetypal Analysis identifies extreme values on the boundary of a data set or archetypes to represent a set of data (Cutler and Breiman 1994). An overview of the data can be approximated based on studying the archetypes.

The proposed method aims to enable architects to discover relationship between architectural features and design performance. The next section will describe the proposed method, and the demonstration section will present an example in which the method is applied to a case study. Finally, the conclusions section briefly discusses future research.

PROPOSED METHOD

The proposed method consists of two stages: clustering design variants and extracting exemplars. In the first stage, the population of design variants is hierarchically clustered into groups of design variants with distinct characteristics. Once the clusters have been created, exemplars are then extracted for each cluster using both Cluster Analysis and Archetypal Analysis. The design clusters and exemplars are then visualised in order to give architects insights into the relationship between architectural features and design performance.

Clustering Design Variants

The aim is to partition the population of design variants into clusters with distinct characteristics. A basic Euclidian distance-based clustering algorithm is sufficient. For this research, k-means analysis (Hartigan 1975) is used. It is one of the most common Euclidean distance-based algorithm used in data mining. k-means analysis starts with a random initial clustering using random selected centroids and then iterates through the data set searching for the best clusters. At each iteration, the quality of the cluster is measured using the within-cluster-variance measure. The smaller this variance, the more compact is a cluster. The analysis stops when there is no change in the within-cluster-variance for a number of iterations.

Clusters are created in two stages. In the first stage, the population of design variants is clustered according to performance scores. In the second stage, these clusters are then sub-clustered according to a set of selected architectural features derived from the design variants. These features can be any type of metrics that can be calculated to describe general characteristics of design variants. This two-stage clustering approach allows architects to understand different combinations of architectural features that can result in similar performances.

For the proposed method, the architect needs to specify the attributes to be used for clustering and the total number of resultant clusters. For the latter, a heuristic called the "elbow method" can be used (Everitt and Hothorn 2011). This method is based on the observation that an increase in the number of clusters is associated with a diminishing improvement in the quality of those clusters. This is because by splitting clusters that are already high-quality into finer clusters will have marginal reduction in the within-cluster sum of square measure. The "elbow method" can be used to find the turning point when additional clusters no longer result in any significant improvements in cluster quality.

Extracting Exemplars

Once the design variants are partitioned into clusters, a manageable number of representative design variants are extracted. This facilitates architects in qualitatively assessing the relationship between architectural features and performance scores. This is done by analysing the input parameters of the design variants of each cluster to find a set of parameters that best represent the cluster. Archetypal Analysis extracts archetypes of the clusters, which are extreme values located at the boundary of the cluster. k-means analysis extracts the centroids that are located in the centre of each cluster. Together, the archetypes and centroids give a good sampling of the design variants in the cluster. They form the exemplars of the design cluster.

Note that the exemplars are not created by selecting design variants in the cluster. Instead, they are new design variants that are reconstructed by analysing the input parameters for all design variants in the cluster. Thus, they need to be validated by ensuring that their performances and architectural features are within the range of the design cluster. For example, for a design cluster that has a daylight performance of 500-1000 lux and a shape factor of 0.2-0.5, the exemplars need to fall within these ranges to be valid. The proposed analysis method is illustrated in Figure 1.

DEMONSTRATION

The method is demonstrated on an abstract building type. The demonstration explores how self-shading geometry, envelope materials and window area will affect the overall performance of a simplified building located in the Singapore climate. The design schema is illustrated in Figure 2. The schema is based on two 4x2 grids stacked on top of each other (Figure 2a). There are four options for the location of the vertical core of the building: columns 1 and 5, columns 2 and 6, columns 3 and 7, or columns 4 and 8 (Figure 2b). One grid is chosen from each remaining column to create the building form (Figure 2c). By staggering volumes on top of each other it is possible to create self-shading geometries. There are a total of 256 possible building forms that can arise. All the external walls have windows, the heights of which range from 1.2 m to 3.6 m (Figure 2d). The walls and windows are assigned a material (Figure 2e). The building can be rotated 360º (Figure 2f). The design schema can generate 752,640 possible design variants. The example

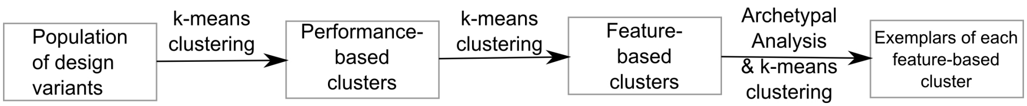


Figure 1
Procedure of the proposed analysis method

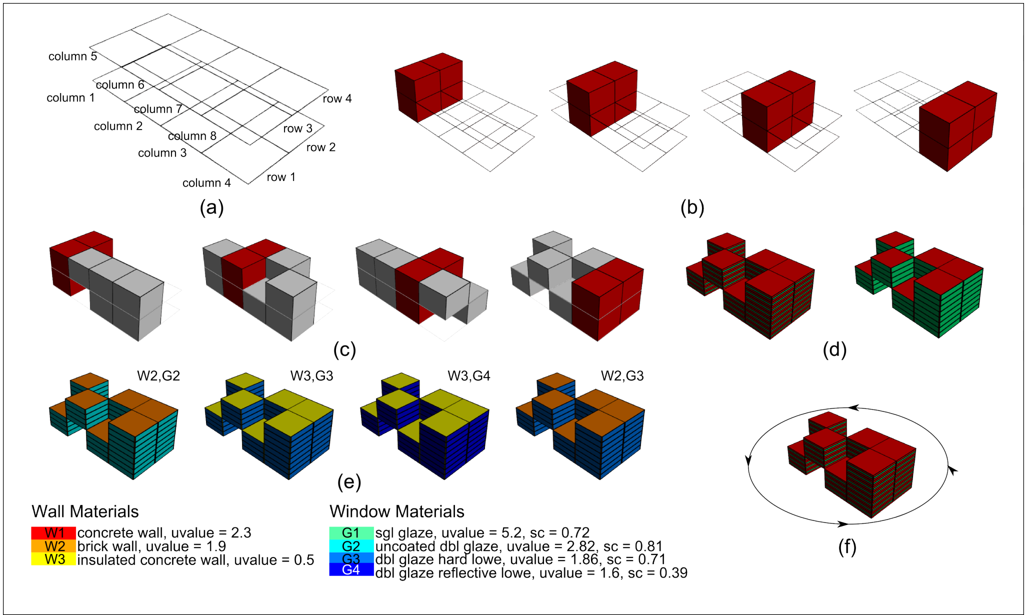


Figure 2
Design schema of proposed building

explores various forms, materials and window areas that can lead to a better performance.

The design schema is evaluated in terms of thermal transfer through the envelope, the envelope cost, and the daylight level. The thermal transfer measures the solar heat gain through the envelope in the tropical climate, and is a good performance indicator of the cooling performance of a building. It is calculated using a simplified method, as the sum of Envelope Thermal Transfer Value (ETTV) (Chua and Chou, 2010) and Roof Thermal Transfer Value (RTTV) (BCA, 2013). The envelope cost is calculated by multiplying the envelope area with the cost per square metre of the material. The better-insulated materials are more costly. The daylight level is calculated as the ratio of the floor area receiving at least 300 lux to the gross floor area. The overall thermal transfer and the envelope cost are to be minimised, while the daylight level is to be maximised.

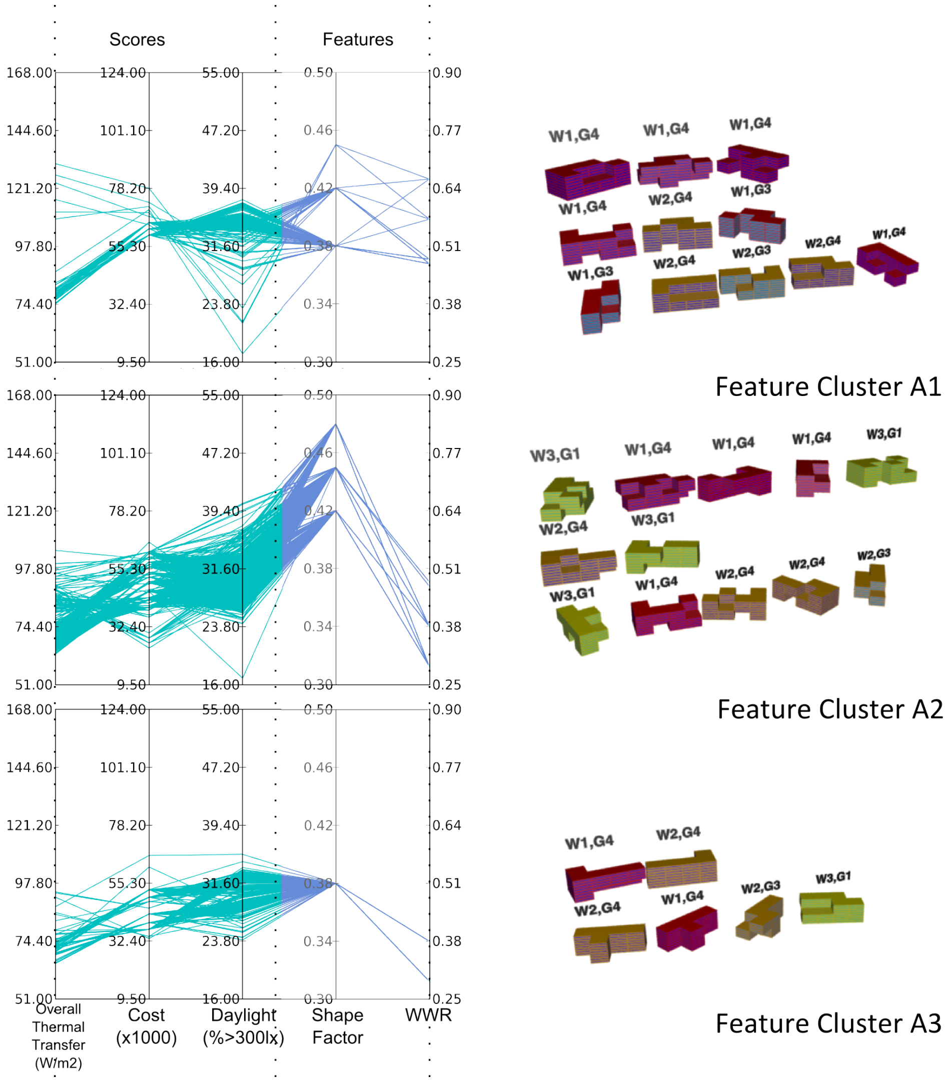
A Non-Dominated Sorting Genetic Algorithm 2 (NSGA2) (Deb et al. 2000) is used for the optimisation process. The settings of the algorithm are an initial population of 100, crossover rate of 0.9, and mutation rate of 0.01. A total of 5000 design variants are generated from 50 generations.

RESULTS

The population of 5000 design variants are analysed using the proposed method. In the clustering stage, two levels of clustering are performed. First, k-means clustering is used to cluster the population according to their performance scores. Three performance-based design clusters are produced, labelled Performance Clusters A, B and C. Cluster A achieves a balance between all three performance objectives, Cluster B focuses on low overall thermal transfer, and Cluster C focuses on low cost.

Each performance cluster is then sub-clustered based on two architectural features: the shape factor (CEN 2007) and Window Wall Ratio (WWR). The shape

Figure 3
Feature clusters A1, A2 and A3 with its exemplars



factor and WWR describes the building form and envelope of the design variants, by using these attributes for the second stage of clustering, one will be able to identify the relationship between the building form, envelope design and performance scores in the feature-based clusters. This results in a total of 9 feature clusters.

In the exemplar extraction stage, the exemplars are extracted by running an Archetypal Analysis and k-means analysis on the input parameters. The design clusters are then visualised in two forms. The performances and architectural features will be visualised as Parallel Coordinate Plots (PCP) and the exemplars as 3D models, as shown in in Figures 3 to 5. The exemplars are arranged in three rows, the centroid is located in the middle row while the archetypes are at the first and last row. The description on the top of each exemplar indicates its wall and window material. Figure 2e shows the legend of the wall and window materials.

Performance Cluster A

Performance cluster A achieves a balance between all three performance objectives. Cluster A2 and A3 (Figure 3) are the most balanced design clusters. They achieve an acceptable overall thermal transfer and daylight level while maintaining a low cost, relative to the other design clusters. The design variants in cluster A3 have a lower shape factor, which is illustrated by its exemplars with their compact forms. These compact forms have lesser over-hangs and shadings. Cluster A3 has a smaller daylight performance range than cluster A2 because of the lesser shadings: the WWR needs to be low to maintain the overall thermal transfer, and a lesser window area leads to lower daylight performance. Lastly, the long façades of the exemplars mainly face north-south to avoid the east-west sun. Cluster A2 has a high shape factor, and its exemplars have building forms that self-shade themselves with over-hangs. Due to the shading, the exemplars are able to afford higher WWR and thus achieve a higher maximum daylight level of 40.43%, compared with the 35.48% of cluster A3.

The design variants in cluster A1 (Figure 3) have a shape factor higher than that of cluster A3 but lower than that of cluster A2, and higher WWR than both design clusters. Cluster A1 is able to achieve similar overall thermal transfer and daylight performance to cluster A2, with a higher envelope cost. This higher cost is due to the high WWR, as better glazing material is required to maintain the overall thermal transfer performance with the increased window surface area.

Performance Cluster B

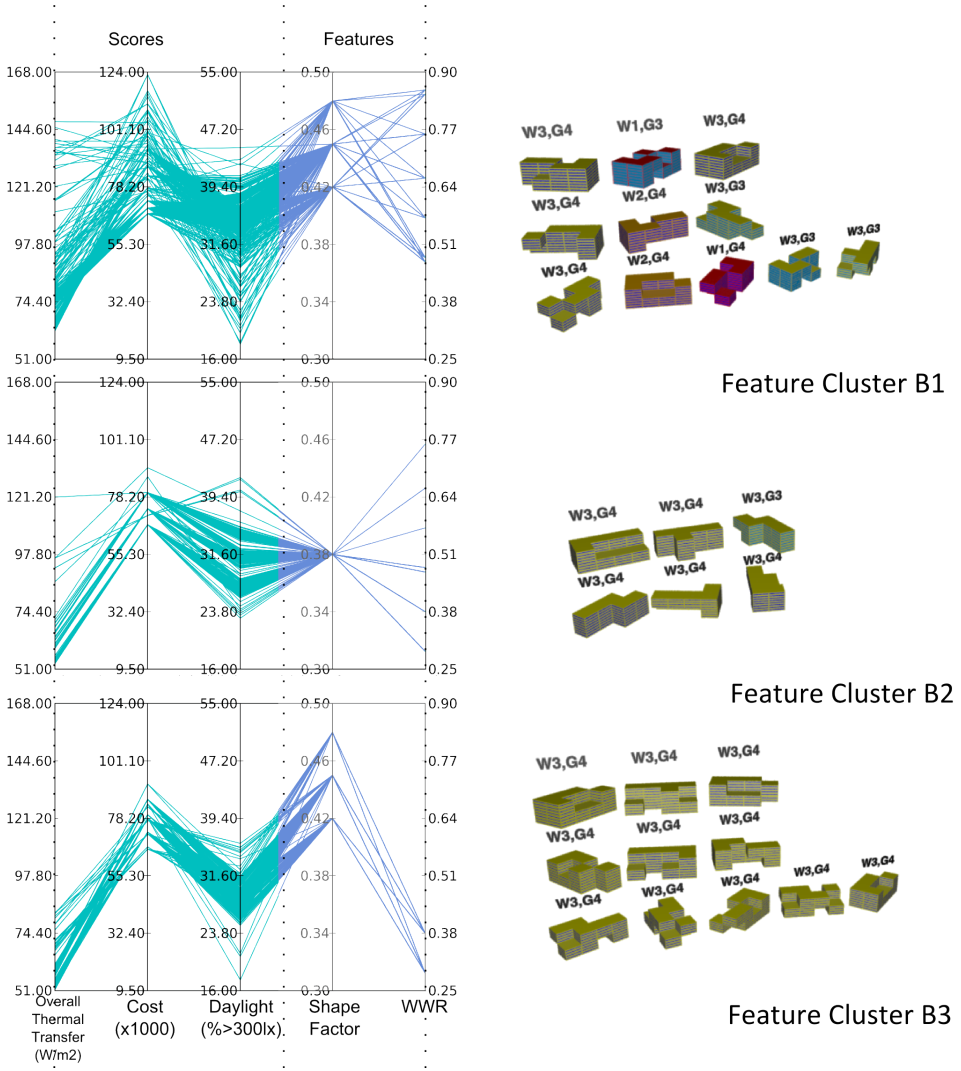
Performance cluster B consists of design variants with low overall thermal transfer but high envelope material cost. Cluster B3 (Figure 4) has the best overall thermal transfer performance and the worst daylight performance. The envelope cost is higher than cluster A2. Most of the exemplars have their long façade facing north-south, and a low WWR, as with cluster A3. Most of the exemplars use highly insulated materials, which is reflected in the higher envelope cost. The combination of good orientation, low WWR, and good envelope materials contributed to the best overall thermal transfer performance. The trade-offs are a higher envelope cost and the worst daylight performance among the nine clusters.

Overall thermal transfer and cost performance similar to cluster B3 can be achieved with an architectural design of lower shape factor and bigger range of WWR, as shown in cluster B2 (Figure 4). Design variants in cluster B1 (Figure 4) have a similar shape factor to that of cluster B3, but a higher WWR range, and the high WWR requires good insulated window materials to maintain the overall thermal transfer; as a result, the design cluster has the worst cost performance. The advantage of increasing the WWR is better daylight performance, but the daylight improvement is only 0.59-8.71%, compared with cluster B3.

Performance Cluster C

Performance cluster C consists of design variants with low envelope cost, high daylight level, but high overall thermal transfer performance. Cluster C2 (Figure 5) has the best envelope cost performance of the

Figure 4
Feature clusters B1, B2 and B3 with its exemplars



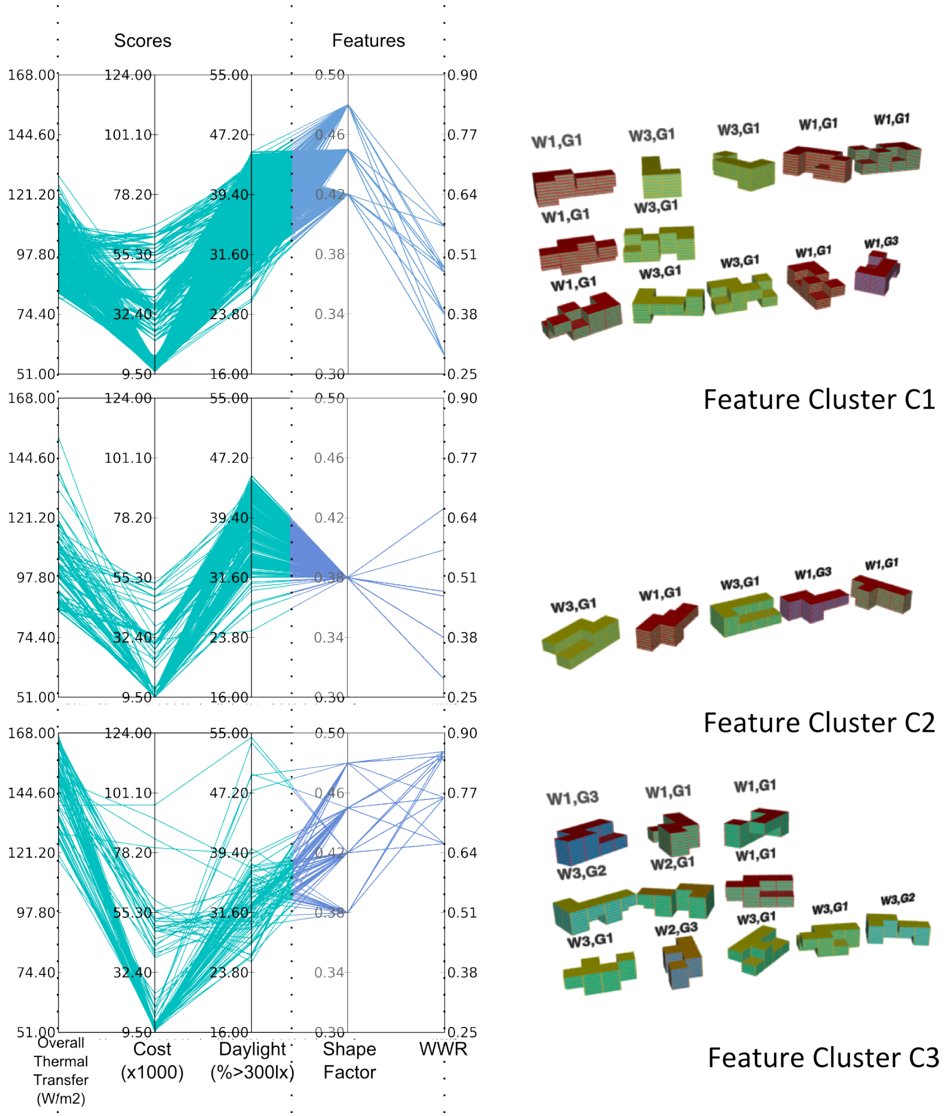


Figure 5
Feature clusters C1, C2 and C3 with its exemplars

nine design clusters. Its shape factor and WWR range are similar to those of cluster B2. The main difference between the two design clusters is the envelope material cost, as the design variants use envelope constructions of lower thermal qualities, as shown by the cluster's exemplars. It achieves similar daylight performance to that of cluster B2, but due to the low-quality envelope materials the overall thermal transfer performance is much worse than that of cluster B2.

Overall thermal transfer and daylight performances similar to those of cluster C2 can be achieved with an architectural design of higher shape factor, as shown in cluster C1 (Figure 5). The increase in shape factor increases the surface area of the envelope, and as a result the cost is 13.3k higher than that of cluster C2. Cluster C3 (Figure 5) has the best-performing daylight, while having the worst-performing overall thermal transfer performance. The exemplars are characterised as having high shape factor and high WWR with low-quality glazing materials.

CONCLUSION AND DISCUSSION

The demonstration shows how k-means clustering and Archetypal Analysis can be used to partition design variants into clusters and to extract exemplars. The PCP of the design clusters and 3D geometry of the exemplars facilitate the analysis of a large number of design variants generated from the optimisation process. The clusters are able to provide a visual summary of the 5000 design variants.

The demonstration shows that although it is possible to derive general knowledge linking architectural features to design performance, the process is still not straightforward. It is not easy with which such knowledge can be derived depends on the specific clusters being compared. For example, comparing feature cluster A2 and A3 reveals how different architectural designs can achieve similar performances. One can either have a low shape factor with low WWR, or a high shape factor with a bigger range of WWR. Other comparisons are much less revealing. For example, when comparing cluster A1 and A2, it is difficult to identify any clear relationship between architectural features and performance scores. In this case, the two sets of exemplars do not seem to have any distinct architectural features, which in turn makes it difficult to conclude anything with regards to performance.

Future research aims to improve on the current method by supporting a more interactive approach. Rather than automating the whole analysis procedure as discussed in this paper, this approach will allow architects to analyse populations of design variants by interactively applying various techniques such as clustering, archetypal analysis, and filtering. This interactive approach will allow architects to engage in an iterative process in which the analysis techniques are repeatedly tweaked in order to home in on specific relationships between architectural features and design performance.

REFERENCES

BCA, S 2013, *BCA Green Mark for New Non-Residential Buildings Version NRB/4.1*, Building and Construction Authority Singapore

Caldas, L 2008, 'Generation of energy-efficient architecture solutions applying GENE_ARCH: An evolution-based generative design system', *Advanced Engineering Informatics*, 22, pp. 59-70

CEN, E 2007, *EN 15217, Energy performance of buildings - Methods for expressing energy performance and for energy certification of buildings*, BSI European Committee for Standardization

Chichakly, K and Eppstein, M 2013, 'Discovering Design Principles from Dominated Solutions', *Access, IEEE*, 1, pp. 275-289

Chua, KJ and Chou, SK 2010, 'An ETTV-based approach to improving the energy performance of commercial buildings', *Energy and Buildings*, 42, pp. 491-499

Cutler, A and Breiman, L 1994, 'Archetypal Analysis', *Technometrics*, 36(4), pp. 338-347

Deb, K, Agrawal, S, Pratap, A and Meyarivan, T 2000, 'A Fast Elitist Non-dominated Sorting Genetic Algorithm for Multi-objective Optimization: NSGA-II', in Schoenauer, M, Deb, K, Rudolph, G, Yao, X, Lutton, E, Merelo, J and Schwefel, HP (eds) 2000, *Parallel Problem Solving from Nature PPSN VI*, Springer Berlin Heidelberg, pp. 849-858

Deb, K, Bandaru, S, Greiner, D, Gaspar-Cunha, A and

Tutum, CC 2014, 'An integrated approach to automated innovization for discovering useful design principles: Case studies from engineering', *Applied Soft Computing*, 15, pp. 42-56
Deb, K and Srinivasan, A 2006 'Innovization: Innovating Design Principles Through Optimization', *Proceedings of the 8th Annual Conference on Genetic and Evolutionary Computation*, New York, NY, USA, pp. 1629-1636
Everitt, B and Hothorn, T 2011, 'Cluster Analysis', in Everitt, B and Hothorn, T (eds) 2011, *An Introduction to Applied Multivariate Analysis with R*, Springer New York, New York, pp. 163-200
Flager, F, Welle, B, Bansal, P, Soremekun, G and Haymaker, J 2009, 'Multidisciplinary Process Integration and Design Optimisation of a Classroom Building', *Journal of Information Technology in Construction*, 14, pp. 595-612
Han, J, Kamber, M and Pei, J 2012, '10 - Cluster Analysis: Basic Concepts and Methods', in Han, J, Kamber, M and Pei, J (eds) 2012, *Data Mining (Third Edition)*, Morgan Kaufmann, Boston, pp. 443-495
Harfield, S 2007, 'On design ‘problematization’: Theorising differences in designed outcomes', *Design Studies*, 28(2), pp. 159-173
Hartigan, JA 1975, *Clustering Algorithms*, John Wiley \& Sons, New York
Janssen, P, Basol, C and Chen, KW 2011 'Evolutionary Developmental Design for Non-Programmers', *29th eCAADe Conference Proceedings*, University of Ljubljana, Faculty of Architecture (Slovenia), pp. 245-252
Lawson, B 2004, *What designers know*, Architectural Press, Oxford
Lin, SHE and Gerber, DJ 2014, 'Designing-in performance: A framework for evolutionary energy performance feedback in early stage design', *Automation in Construction*, 38, pp. 59-73
Mela, K, Tiainen, T and Heinisuo, M 2012, 'Comparative study of multiple criteria decision making methods for building design', *EG-ICE 2011 + SI: Modern Concurrent Engineering*, 26(4), pp. 716-726
Pohekar, S and Ramachandran, M 2004, 'Application of multi-criteria decision making to sustainable energy planning—A review', *Renewable and Sustainable Energy Reviews*, 8(4), pp. 365-381
Singhaputtangkul, N, Low, SP, Teo, AL and Hwang, BG 2013, 'Knowledge-based Decision Support System Quality Function Deployment (KBDSS-QFD) tool for assessment of building envelopes', *Automation in Construction*, 35, pp. 314-328
Turrin, M, Buelow, Pv and Stouffs, R 2011, 'Design explorations of performance driven geometry in architectural design using parametric modeling and genetic algorithms', *Advanced Engineering Informatics*, 25, pp. 656-675
Woodbury, R 2010, *Elements of Parametric Design*, Routledge, Oxon