www.itcon.org - Journal of Information Technology in Construction - ISSN 1874-4753

USING A DATA-DRIVEN APPROACH TO SUPPORT THE DESIGN OF ENERGY-EFFICIENT BUILDINGS

SUBMITTED: October 2014
REVISED: November 2014
PUBLISHED: January 2015 at http://www.itcon.org/2015/6
GUEST EDITORS: Mahdavi A. & Martens B.

Yuezhong Liu,
National University of Singapore;
liuyuezhong@nus.edu.sg

YiChun Huang,
Singapore

Rudi Stouffs, Dr.
National University of Singapore;
stouffs@nus.edu.sg

SUMMARY: *With increasing interest in sustainable design, the issue of energy-efficiency in the building design process is receiving ever more attention from designers and researchers. Greater access to building performance analysis results has led designers and researchers to increasingly address energy-efficiency concerns. However, the huge amount of performance analysis data that may be generated during the design process cannot easily be handled by traditional data analysis methods. The goal of this research is to develop a data-driven approach for the integrated design process, in order to help to improve the accuracy of performance analyses and also reduce the time required to complete such design iterations. We propose our method to include five step: 1) Requirement identification; 2) Building modelling; 3) Workflow implementation; 4) Simulation and data mining; 5) Evaluation and refinement. A case study demonstrates our data-driven workflow's ability to guide the design process with high precision. Our approach can also be extended and applied to discover useful patterns in the building design process.*

KEYWORDS: *data-driven workflow, data mining, density-based clustering, EnergyPlus, Radiance, integrated simulation.*

REFERENCE: *Yuezhong Liu, YiChun Huang, Rudi Stouffs (2015). Using a data-driven approach to support the design of energy-efficient buildings. Journal of Information Technology in Construction (ITcon), Special Issue: ECPPM 2014, Vol. 20, pg. 80-96, http://www.itcon.org/2015/6*

1. INTRODUCTION

The development of energy-efficient buildings is a sustainable vision that entails huge challenges for environmental and technical innovation. It has consequences for all professions, not the least for architectural design and building engineering, since it is here that the full complexity of building performance analysis has to be addressed and managed throughout the design process. While the huge amount of performance analysis data that is generated during the design process cannot easily be processed by traditional data analysis methods, there is a crucial need to apply more advance data analytical methods into the design process. Therefore, we propose to use a data-driven approach to support the design process of energy-efficient buildings.

1.1 Energy-Efficient Design

Energy-efficient buildings can be described as buildings that are designed to provide a significant reduction of the energy need for the building systems, including heating and cooling, lighting, etc. Design determines the building sector's energy consumption for far longer than other end-use sectors' components determine their sector's efficiency (Lysen 1996).The improvement of a building's energy efficiency at the planning stage is relatively simple while improvements after their initial construction are much more difficult: decisions made during a building's project phase will hence determine consumption over much, if not all, of a building's lifetime (Lysen 1996). In order to achieve energy efficiency, most buildings' architects will consider from the following perspectives:

1) Bioclimatic architecture: the shape and orientation of the building and its solar protections.
2) High performing building envelope: thorough insulation, high performing glazing windows, and air-sealed construction.
3) High performance controlled ventilation: mechanical insulation, and heat recovery.

Early energy-efficiency requirements for buildings responded to poor insulation levels, which could lead to health problems because of moisture or air infiltration. These requirements contained U-values, R-values and specific insulation materials or multi-glazing, and were intended to improve energy efficiency and comfort in buildings.

Lysen (1996) introduces the design concept Trias Energetica, which suggests an organized approach to reduce the dependence on fossil fuels. This concept mainly focuses on ways that deal with energy to achieve savings, reduction of dependence and environmental benefits, while maintaining the building's comfort and construction progress.

Trias Energetica proposes three main strategic actions:

1) Reduce the overall energy demand of the building
2) Supply the energy demand with renewable energy
3) Supply the remaining part with efficient use of fossil fuel

The first action refers to the insulation of the building envelope; the second action relates to natural energy gains, such as PV and windmill; the last action considers the equipment system including HVAC, lighting, control system, etc. (Fig. 1).

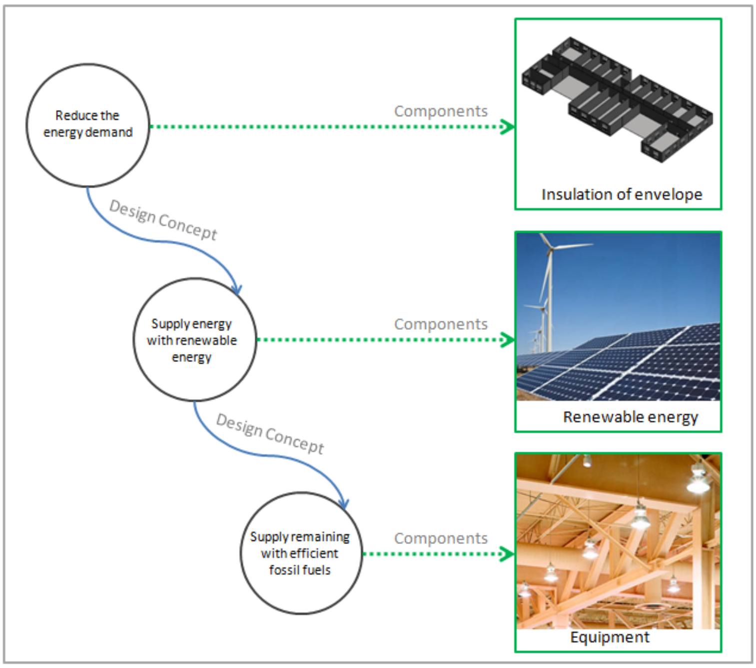

FIG. 1: The Trias Energetica concept and its relation to design components.

1.2 Integrated Design

Integrated design is a term used for the process where all the elements described above are used in an integrated way to reduce the energy consumption in a building. In this process actions are taken to reduce the energy consumption both through insulation or efficiency as through the design of the buildings and the HVAC systems. Passive use of renewable energy and other natural sources is an integrated part of the design and development process and there is an interactive process between the design of building and systems (Jens 2008). Integrated design of energy-efficient buildings requires more emphasis on energy efficiency and systems in the early planning phase than traditional design and it is difficult to regulate through building codes and energy efficiency standards, although the most advanced standards or energy performance calculation procedures include options for integrated design.

For this research, we will integrate a logical workflow informed by data mining results into the integrated design process. From this research using data mining, it is also our intention to ensure the logical correctness of the data mining analysis and its reliability in the proposed workflows. Using the data mining analysis will allow further discovery of the best correlation between different energy systems within building models, and inevitably will shorten the adaptive-iterative prototyping cycle (Fig. 2).

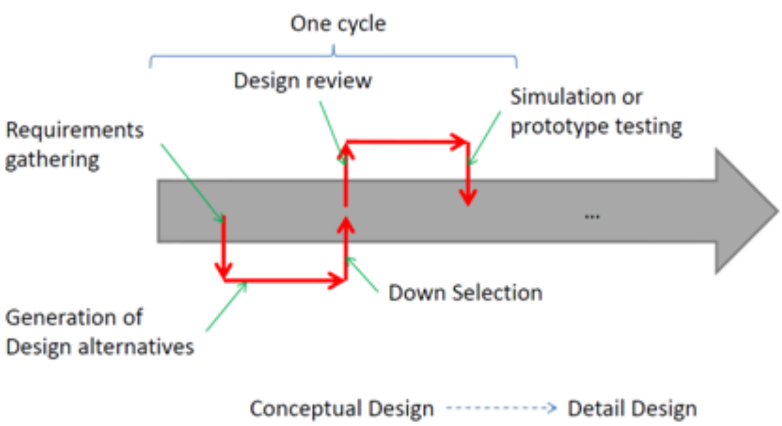

FIG. 2: Overview of the data-driven workflow cycle within the integrated design process.

As shown in the workflow in Fig. 2, data mining will be applied between the design review and simulation steps. As is well known, the data mining process allows the discovery of useful knowledge from a collection of data and is applicable in any multidisciplinary activities. This process is an iterative process and the patterns generated from the data should be validated on any relevant data records, and possess some degrees of certainty. In addition to that, these knowledge patterns can easily be updated with respect to future updates of the data source. During the data mining part, we will compare and evaluate different approaches before jumping into the next step. To understand the effectiveness of the proposed data-driven workflow for the design process, this research will focus on insulation of the envelope to improve energy efficiency by applying mainly two simulation tools: EnergyPlus and Radiance. Building physical attributes such as roof, walls, floor, and windows will be the main focus of this research.

1.3 DATA-DRIVEN WORKFLOW

We propose our design process, based on the current integrated design process, to include the following steps:

1) Identify the critical design requirements and parameters for energy and daylighting;
2) Model the building (in Revit Architecture);
3) Implement the data-driven workflow and its essential technologies, including transferring the modelling information between the different software;
4) Apply data mining techniques, including clustering, classification and associated rule, to process the integrated discipline analysis results for energy and daylighting simulation;
5) Evaluate and refine the effectiveness of the proposed workflow.

Fig. 3 presents an overview of our integrated design process; for the building modelling we adopt Revit Architecture 2013 and for the simulation tools we choose EnergyPlus and Radiance.

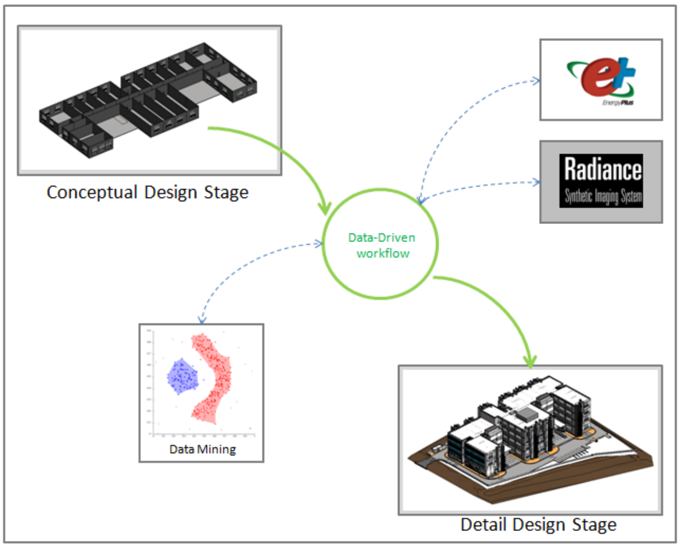

FIG. 3: Overview of the role of the data-driven workflow within the integrated design process.

1.4 CHALLENGES AND CONTRIBUTIONS

The research addresses the following two questions:

1) How to apply the data mining methods into the data-driven workflow?
2) How to control the data-driven workflow to support the design of energy-efficient buildings?

This research uses a design project as a case study; the challenges of the data-driven workflow include:

1) Data interoperability between the different file formats of the modelling software and simulation tools. For example, Revit does not support the IDF file format which is the input format for EnergyPlus.

2) Data analysis and detection. As the building information analysis is processed by data mining methods, how will the new knowledge discovered from the simulation results be applied into the next data analysis cycle?

3) Workflow control, using the data mining analysis to determine the next task. It is difficult to estimate the direction of the design process when different scenarios use different analysis methods during the data process.

After we investigate the case study, the main contributions of the research will be:

1) To overcome the data exchange problems between the modelling software and the simulation tools within the scope of this research.

2) To apply data mining methods in the design process to support the decision-making workflow.

3) To develop a data-driven workflow to support the design of energy-efficient buildings.

The rest of the paper is organized as follows. The related work will be reviewed in Section 2. Section 3 presents the design ideas and detail of our approach. Algorithm details and implementation issues are discussed in Section 4. A case study analysis is showed in Section 5. Section 6 gives the conclusions of this paper.

2. LITERATURE REVIEW

The literature review concerns three parts: I) Integrated design process; II) Data mining; II) Data-driven workflow.

2.1 INTEGRATED DESIGN PROCESS

The application of advanced control methods for building installations during building design, using building energy simulation tools, has received attention in the building automation community. Bernal et al. (2012) proposed the MLE+ tool, which can be used for performance evaluation of the proposed control methods with EnergyPlus. Liao et al. (2012) presented an occupancy driven approach for energy efficient building control system. Narayanan et al. (2010) analyzed and optimized the performance of a real building using EnergyPlus models. Auslander et al. (2013) used a Siemens Apogee controller for demand response control and predictive modelling of a campus building using EnergyPlus. The Alleyne Research Group (2012) developed THERMOSYS for analyzing the behaviour of air-conditioning and refrigeration systems. It contains dynamic models of the basic components used in compression cycles but it cannot be used for whole building simulation. Wetter (2011) described a popular tool for building energy co-simulation. None of these research projects apply an accurate and efficient workflow to direct the integrated design process. We propose to apply a data-driven workflow for this purpose.

2.2 DATA MINING

Recognizing the complexity of the search algorithms and the size of the data being analyzed when identifying useful patterns in energy modelling data, Kim et al. (2011) utilize data mining technology, which can be considered an interdisciplinary field involving concepts from machine learning, statistics, mathematics, high-performance computing, and visualization. Fayyad et al. (1996) define data mining as the nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data. John (1994) defines it as the process of discovering advantageous patterns in data. Most of these research projects do not apply data mining within a workflow and do not consider a data-driven conception.

Our research utilizes data mining methods such as filtering, clustering and classification to indentify the workflow direction and useful knowledge from the simulation result.

2.3 DATA-DRIVEN WORKFLOW

An et al. (2013) review data-driven approaches that use information from collected data to identify the characteristics of the damaged state and predict the future state without using any particular physical model. Instead, mathematical models with weight parameters are employed. The weight parameters are determined based on the training data that are obtained under the various usage conditions. Since the data-driven approaches depend on the trend of data, which often has a distinct characteristic near the end of life, it is powerful in predicting near-future behaviours, especially toward the end of life.

Data-driven approaches are divided into two categories: artificial intelligence approaches and statistical approaches. Artificial intelligence approaches include artificial neural networks (Krogh 2008, Yao 1999) and fuzzy logic (Zio and Di Maio 2010). Statistical approaches include a regression-based model such as Gaussian process (GP) regression (Mackay 1997, Seeger 2004) and a probabilistic Bayesian learning framework (Tipping 2001). However, the data-driven workflow has rarely been applied in the building design process to support energy efficiency design. In this research project, we will propose an efficient and effective data-driven workflow to guide the energy-efficiency design process.

3. RESEARCH METHODOLOGY

3.1 DENSITY-BASED CLUSTERING

In density-based clustering, clusters are defined as areas of higher density than the remainder of the data set (Fig. 4). Objects in the remaining, sparse areas that separate the clusters are usually considered to be noise and border points.

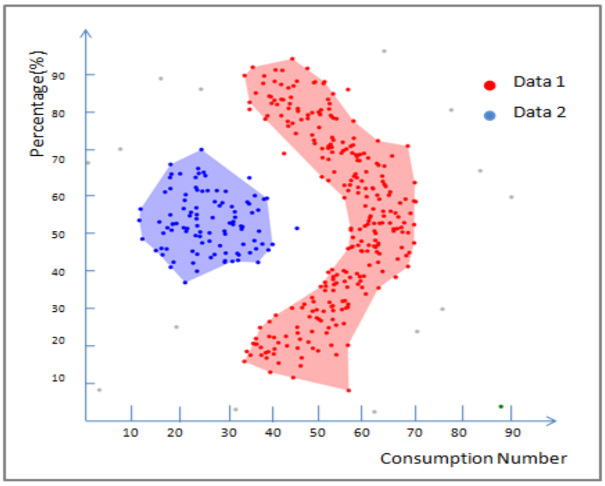

FIG. 4: Density-based spatial clustering of applications with noise (DBSCAN).

Density-based clustering is based on the notion of density reachability. Basically, a point q is directly density-reachable from a point p if q is part of the ε-neighbourhood of p, that is, the distance between p and q is at most ε, and if p is surrounded by sufficiently many points such that one may consider p and q to be part of a cluster. Also, a point q is density-reachable from p, instead of directly density-reachable, if there is a sequence (p_1, p_2, …, p_n) of points with $p_1 = p$ and $p_n = q$ where each p_{i+1} is directly density-reachable from p_i (Fig. 5, left). The density-reachable relation is not symmetric. While p may be surrounded by sufficiently many points, q may not. Two points q and r that are both directly density-reachable from p but where neither q nor r are surrounded by sufficiently many points, are said to be density-connected (Fig. 5, right).

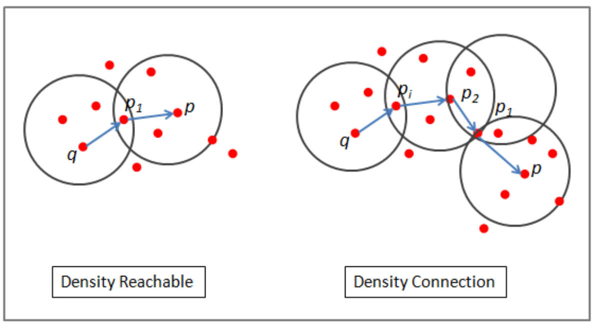

FIG. 5: Density reachable and density connection.

Consequently, the data resulting from the simulation can be simply described as follows:

Let δ_e be the energy consumption threshold, δ_l be the daylighting distribution threshold and δ_s be the density threshold, then, a group of objects q is called a simulation result analysis area, if:

1) The members of q are density-connected among themselves for the energy consumption e where $e \le \delta_e$ and the lighting distribution l where $l \le \delta_l$
2) Size $(q) \ge \delta_s$

According to this conception of the simulation result analysis area, we developed the clustering method in C# (Fig. 6). The algorithm first performs density-based clustering for all the simulation results (lines 1 – 3). Then the system refines the results retaining only the clusters with sufficient size (Lines 4 – 5). Clusters with sufficiently low energy consumption (Lines 6 – 8), or sufficiently high daylighting distribution (Lines 9 – 11) are retained as output. At last, the qualified simulation area set A is updated to the data-driven workflow (Line 12).

Density-based clustering and data analysis algorithm
Input: energy consumption threshold δ_e, daylighting distribution threshold δ_l, density threshold δ_s distance threshold δ_d, candidate set R and the simulation data stream S
Output: every qualified simulation result q

1. **for** each simulation result s of S **do**
2.   initialize new simulation analysis area set A;
3.   cluster the objects in s according to δ_d;
4.   **for** each cluster $s_i \in s$ **do**
5.   **if** size(s_i) $\ge \delta_s$ **then**
6.   **if** *energy consumption* (s_i) $\le \delta_e$ **then**
7.   calculate the density area of s_i;
8.   output s_i as a qualified energy consumption cluster q_e;
9.   **if** *daylighting distribution* (s_i) $\ge \delta_l$ **then**
10.   calculate the density area of s_i;
11.   output s_i as a qualified daylighting distribution cluster q_l;
12.   add all the qualified clusters to A;

FIG. 6: Density-based clustering and data analysis algorithm

3.2 ASSOCIATION RULE LEARNING

Association rule learning is a method for discovering potential relations between different variables. Association rules are created by analyzing data for frequent if/then patterns that satisfy the minimum support (minSup) and the minimum confidence (minConf) thresholds. The association rules problem can be described as follows: Let I = $\{i_1, i_2, \ldots, i_n\}$ be a set of items. Let D be a set of all transactions where each transaction T is a set of items such that $T \subseteq I$. Let X, Y be sets of items such that X, $Y \subseteq I$. An association rule is an implication in the form $X \Rightarrow Y$, where $X \subset I$, $Y \subset I$, $X \cap Y = \emptyset$ (Agrawal et al., 1993). Given the research requirement, we separate the energy consumption (kWh) into three levels: Low, Medium and High, where a High level indicates a reduction in energy use less than 40% and where all obtained reduction values greater or equal to 40% evenly distribute over the Low and Medium level. The possibility of each combination of material items for different energy levels will be calculated and stored in the database.

Fig. 7 shows all possible combinations of generic floor (1), wall (2), roof (3) and window (4) materials. The representation of each set is simplified by only specifying the numbers (digits) in the set. The initial set is empty, represented as *Ø*. Lines connecting item sets indicate that two or more sets can be combined to form a larger set. Fig. 8 describes the corresponding association rule algorithm.

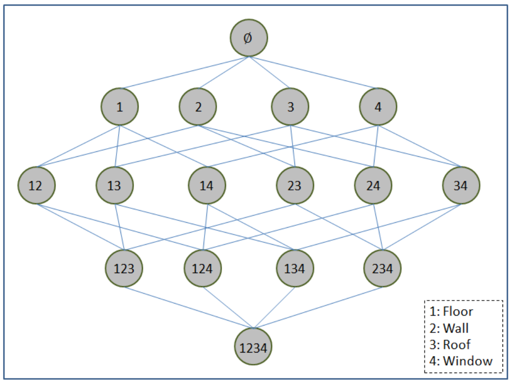

FIG. 7: All possible combinations of materials.

Association rule algorithm for all material combinations
Input: the simulation data stream S
Output: every combination result C_i

1. initialize three levels of energy consumption areas: E_h, E_m, E_l;
2. initialize a frequent item set F_1 = {large 1-itemset};
3. **for** ($k = 2$, $F_{k-1} \neq Ø$; k ++) **do**
4. set C_i = *new combination of materials*;
5. **if** *energy consumption* (S) $\leq E_h$ **then**
6. **for** each k-subset s of S **do**
7. **if** $s \in C_i$ **then**
8. *s.count++*;
9. **if** *energy consumption* (S) $\leq E_m$ **then**
10. **for** each k-subset s of S **do**
11. **if** $s \in C_i$ **then**
12. *s.count++*;
13. **if** *energy consumption* (S) $\leq E_l$ **then**
14. **for** each k-subset s of S **do**
15. **if** $s \in C_i$ **then**
16. *s.count++*;
17. save all frequent item sets = { $C_i\, E_h F_k$; $C_i\, E_h F_k$; $C_i\, E_h F_k$ };

FIG. 8: Association rule algorithm for all material combinations

3.3 DATA-DRIVEN WORKFLOW

The proposed data-driven workflow (Fig. 9) contains the following actions: 1) read the building information from the BIM model (Revit); 2) identify the direction of simulation from the data mining result (logical controller); 3) transfer the building information into an EnergyPlus file; (4) transfer the building information into a Radiance file; (5) read the simulation result from EnergyPlus; (6) read the simulation result from Radiance; (7) repeat the cycle (the iterative workflow controller controls the data-driven workflow with input from the designer).

The development of the data-driven workflow requires three main steps. The first step is to identify the project requirements. This step is a challenging one since projects have constrained budgets, schedules, and resources. It is essential that all building stakeholders – including owners, designers, engineers and contractors – have a clear understanding of the problem definition and participate in identifying a set of design alternatives early in the project planning process. The second step, corresponding to actions 5 and 6 (Fig. 7), is the generation of a large amount of data, specifically, the energy and daylighting simulation results. Exemplar data include estimated energy costs or savings in terms of building orientation, HVAC systems, lighting efficiency and control, roof and wall construction, glazing type, etc. The last step is the data mining process where we develop an overall data analysis mechanism that can also be applied to find patterns that explain or predict any behaviours resulting from the different simulations. Application of the density-based clustering method will guide the design process into an effective direction.

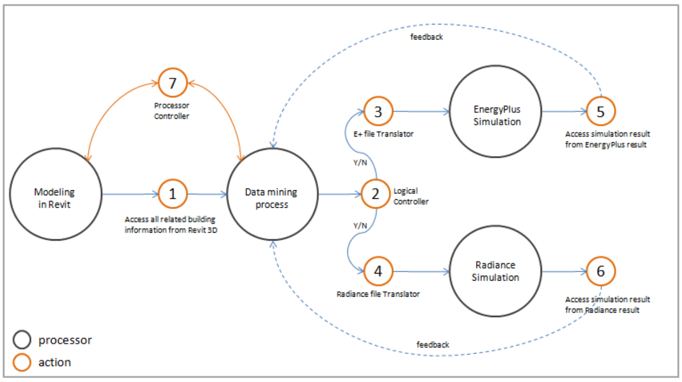

FIG. 9: Overview of proposed data-driven workflow

3.4 THE LIMITATION OF THE METHOD

1) Handling noisy data: parts of the problem situation may be irrelevant to the actual problem posed. An unsuccessful assessment of such noise as present in the problem situation may result in the same problem being unnecessarily stored numerous times in the case base because of the difference in noise assessment. In turn this implies inefficient storage and retrieval of cases.

2) Sustained learning: most of the machine-learning algorithms require a special training phase when information is extracted (knowledge generalization). This makes an on-line adaptation difficult.

3) Data oriented: machine-learning algorithms model the relationships contained in the training data set. Consequently, if the employed training data set is not a representative selection from the problem domain, the resulting model may differ from the actual problem domain. This limitation of machine learning methods is aided and abetted by the fact that most of them do not allow the use of a priori knowledge.

4. CASE STUDY

The proposed data-driven workflow outlined above is applied within a design project.

4.1 REQUIREMENT DEFINITION

The energy-efficiency design process begins when the occupants' needs are assessed and a project budget is established. Then, the proposed building is located on the site, and programmed spaces are carefully arranged to reduce energy use for heating, cooling, and lighting. Building heating and cooling loads are minimized by optimizing the building form and designing energy-efficient building elements – floors, windows, walls, and roofs. Taken together, they form the basis of the whole, integrated building design.

The case study project concerns a new concept factory design. The objective of this design is to reduce more than 40% of the usual energy consumption and maximize the daylighting amenity of this factory building. In the design of the factory, strong emphasis is placed on the functionality and the efficiency of the movement of materials within the factory. The factory incorporates many structural provisions, such as bolting points and removable concrete slabs, to enable different material handling systems of up to two tons to be easily installed by the user without needing to modify the building structure. The factory provides fire fighting, medical service, recycle service, import and export service. The factory consists of a physical working space, day room, dormitory area, office room, decontamination room, storage area/rooms, latrines, communication and electrical closets, and a mechanical room. The size of the proposed factory is approximately 21068.8 m2 (by GFA) (Fig. 10). The apparatus room is sized for a fire truck, military police car, and ambulance. The building is occupied seven days a week for 24 hours a day. A summary of the building model parameters and the thermal loading are presented in Fig. 11.

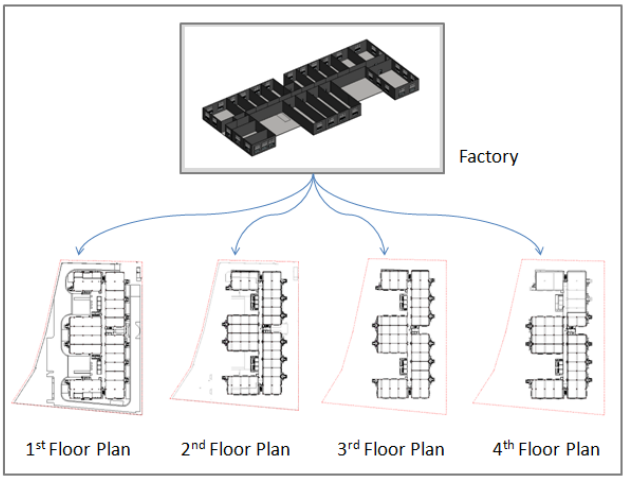

FIG. 10: Proposed factory building.

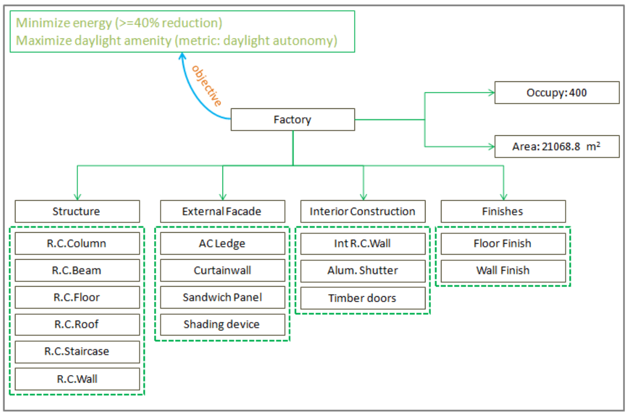

FIG. 11: Baseline building model parameters.

4.2 BUILDING MODELLING

The modelling and simulations are achieved using Autodesk Revit Architecture, EnergyPlus and Radiance. Individual energy analysis runs are often performed to identify the annual energy cost for each building feature or system under consideration, or to determine the effectiveness of applying multiple features or systems to a design. In order to perform the energy analysis, the building information is read from the Revit BIM model (Fig. 12). This completes the first action of the data-driven workflow.

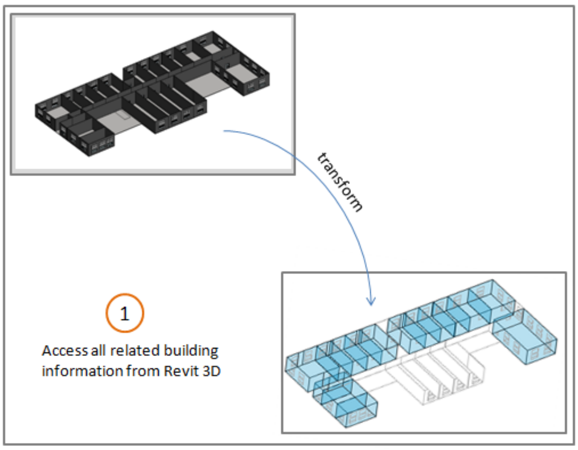

FIG. 12: Information exchange.

4.3 DATA ANALYSIS AND WORKFLOW CONTROL

When the workflow is run for the first time, there are no pre-defined data yet to follow. Thus, the design starts from the first floor plan, and the logical controller node in the workflow will choose to run the simulations in sequence, EnergyPlus then Radiance. Fig. 13 shows the clustering area of different building components from many EnergyPlus simulation results. It shows the window's clustering area to be much larger than the others. In the second action of the data-driven workflow, based on this knowledge, the logical controller guides the designer to define the window materials and sizes first. The windows material can be chosen from the following ten materials: "CLEAR 12MM", "BRONZE 6MM", "GREY 12MM", "LOW IRON 2.5MM", "BLUE 6MM", "REF A CLEAR LO 6MM", "PYR A CLEAR 3MM", "LOE CLEAR 3MM", "COATED POLY-88" and "ECABS-1 BLEACHED 6MM".

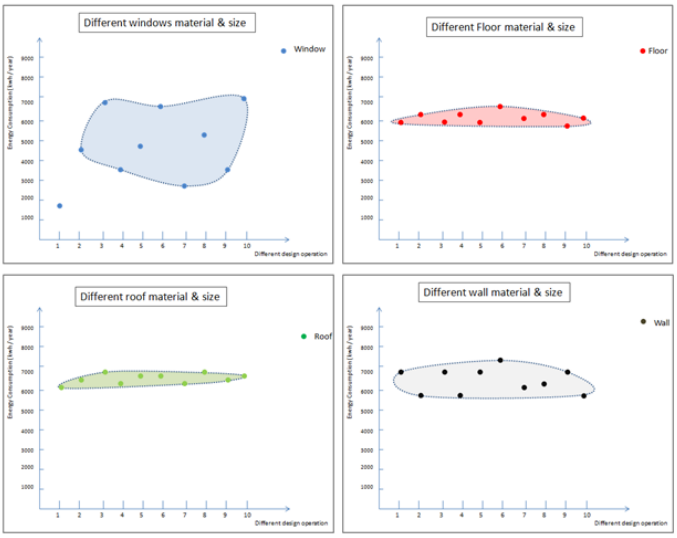

FIG. 13: Density-based clustering from simulation results.

When performing the Radiance simulations, the design alternatives initially only consider the window-to-wall-ratio (WWR). Relating to the size of windows and walls, the Radiance simulation results are compared to the daylight autonomy benchmarks. From the simulation results, the data-driven workflow finds out that when the WWR increases, the daylight autonomy will increase too (Fig. 14).

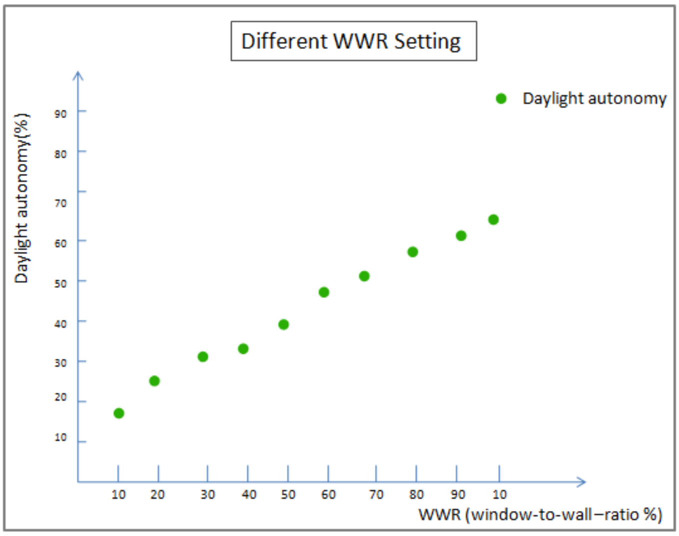

FIG. 14: Radiance simulation results.

Integrating the data from Fig. 13 and Fig.14 into an analysis with respect to daylight autonomy and energy consumption, the data-driven workflow software only includes the WWR 20%, 40% and 60% in the output charts (Fig. 15).

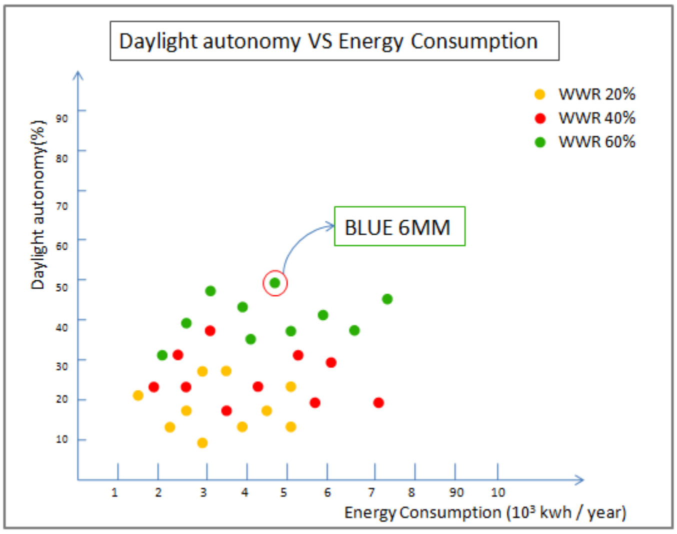

FIG. 15: Daylight autonomy and energy consumption.

From the analysis of daylight autonomy and energy consumption, the data-driven workflow can easily identify the best solution with respect to window sizes and materials. This feedback is given to the designer in action 7 of the data-driven workflow. The designer then updates the windows and walls in Revit. According to the density-based clustering results shown in Fig. 13, in the next step we will use the association rule learning to analyze the different combinations of materials. The possible materials for floor, roof and wall are shown in Fig .16.

Floor	Roof	Wall
F06 EIFS finish	M01 100mm brick	F06 EIFS finish
F07 25mm stucco	M02 150mm lightweight concrete block	F07 25mm stucco
F08 Metal surface	M03 200mm lightweight concrete block	F08 Metal surface
F09 Opaque spandrel glass	M04 300mm lightweight concrete block	F09 Opaque spandrel glass
F10 25mm stone	M05 200mm concrete block	F10 25mm stone
F11 Wood siding	M06 300mm concrete block	F11 Wood siding
F12 Asphalt shingles	M07 150mm lightweight concrete block (filled)	F12 Asphalt shingles
F13 Built-up roofing	M08 200mm lightweight concrete block (filled)	F13 Built-up roofing
F14 Slate or tile	M09 300mm lightweight concrete block (filled)	F14 Slate or tile
F15 Wood shingles	M10 200mm concrete block (filled)	F15 Wood shingles
F16 Acoustic tile	M11 100mm lightweight concrete	F16 Acoustic tile
F17 Carpet	M12 150mm lightweight concrete	F17 Carpet
F18 Terrazzo	M13 200mm lightweight concrete	F18 Terrazzo
	M14a 100mm heavyweight concrete	
	M14 150mm heavyweight concrete	
	M15 200mm heavyweight concrete	
	M16 300mm heavyweight concrete	
	M17 50mm lightweight concrete roof ballast	

FIG. 16: Materials for floor, roof and wall.

The results from the association rule mining shown in Fig. 17 indicate that the wall material "F14 SLATE" has the highest possibility (as a percentage of all material combinations) of medium and low energy consumption level. Hence, the material of the walls will be selected as "F14 SLATE". Because the design requirements only specify an energy reduction of 40%, the medium level already achieves this objective. The results from the association rule mining also show that the selection of materials for floor and roof cannot be considered independently of the selection of the wall material. Fig. 18, Fig. 19 and Fig. 20, demonstrate that while the possibility of the wall material "F14 SLATE" for the medium level is 3.52%, it is 4.3% for the Wall-Floor combination, 2.62% for the Wall-Roof combination and 6.67% for the Wall-Window combination. Multiplying the latter three percentages for floor, roof and window does not provide the same result as the percentage for wall only.

FIG. 17: *The possibility of each wall material for the three levels of energy consumption.*

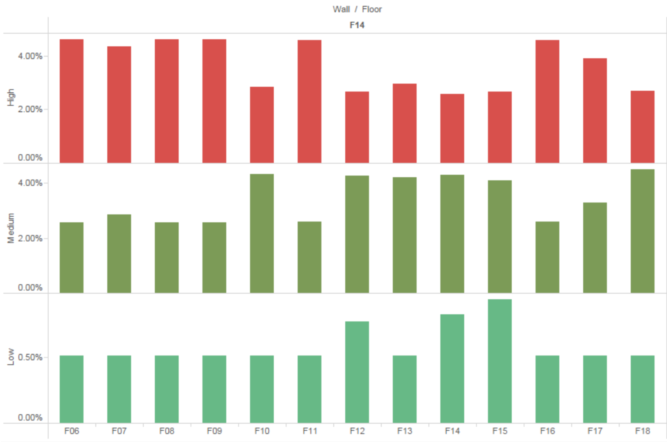

FIG. 18: *The possibility of each floor material with the selected wall material (F14) for the three levels of energy consumption.*

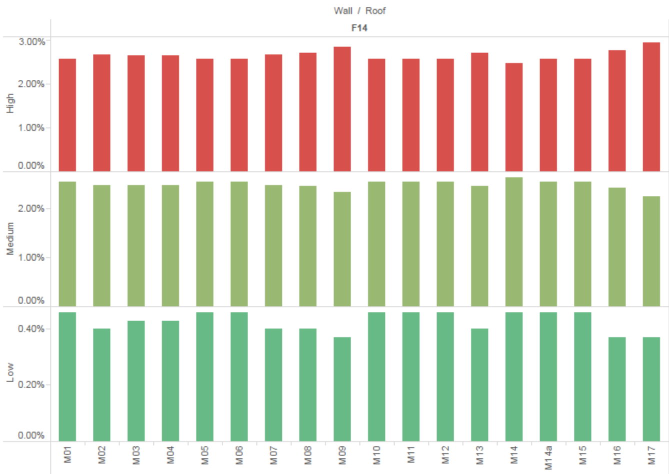

FIG. 19: *The possibility of each roof material with the selected wall material (F14) for the three levels of energy consumption.*

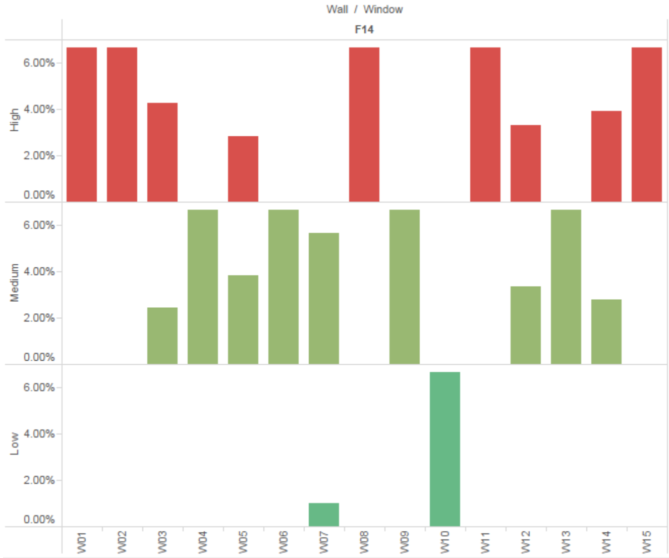

FIG. 20: The possibility of each window material with the selected wall material (F14) for the three levels of energy consumption.

From the same results, if we consider the various combinations of wall, roof and floor materials with the selected window material ("Blue 6MM"), we find that whatever material is selected for wall, roof and floor, the possibility will be limited for the medium level (Fig. 21, Fig. 22 and Fig. 23). Thus, the window material is the most influential factor for the energy consumption during the early design stage.

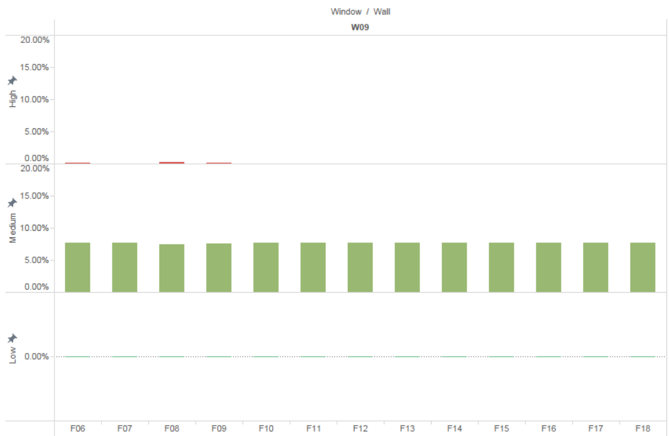

FIG. 21: The possibility of each wall material with the selected window material (Blue 6MM) for the three levels of energy consumption.

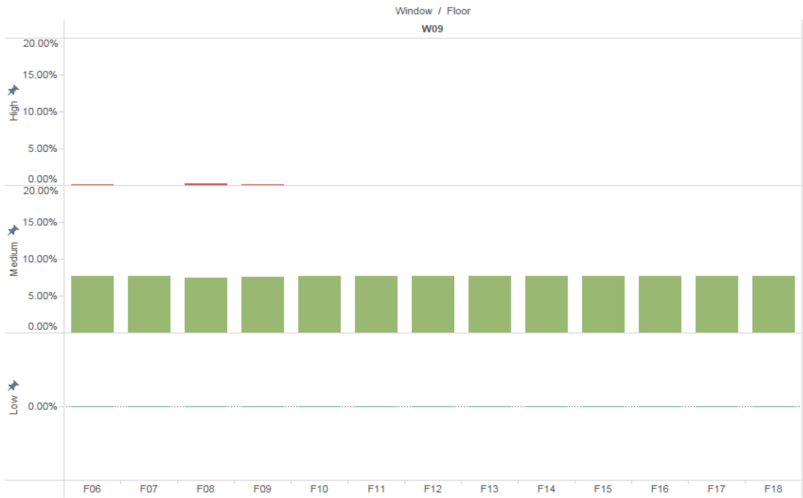

FIG. 22: The possibility of each floor material with the selected window material (Blue 6MM) for the three levels of energy consumption.

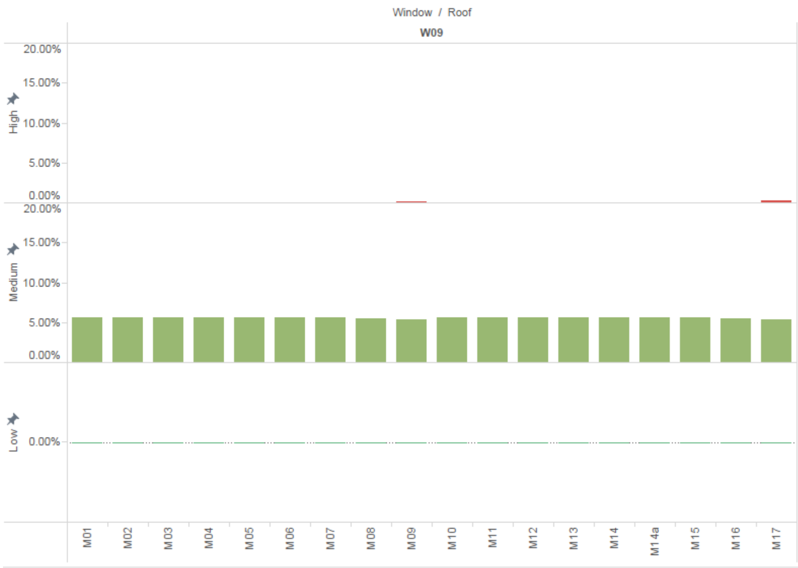

FIG. 23: The possibility of each roof material with the selected window material (Blue 6MM) for the three levels of energy consumption.

Because the lesser impact of the floor and roof material on the energy consumption, the material "F14 SLATE" is selected for the floor and the roof will be applied with 150 mm heavyweight concrete. After updating the building model, the designer triggers the data-driven workflow again to repeat the cycle and perform similar actions as already explained. The data mining process automatically updates the knowledge which has been found into its algorithms. After many cycles, the factory design is finalized as shown in Fig. 3. The data-driven workflow has realized the adaptive-iterative design process within this case study and has applied the data mining methods to draw knowledge from the simulation results. For this research, we also conclude two rules during the early design stage of energy-efficient buildings:

1) The materials of different components are non-independent factors for energy performance analysis.
2) The window materials are the most influential factors, the sequence of decision-making for energy performance will be window > wall > roof and floor

4.4 VALIDATION

Based on the "BCA Green Mark Certification Standard for New Buildings" as proposed by the Building and Construction Authority (2012) in Singapore, certification points are allocated to lighting and energy-efficient practices and features used. A minimum of 30 points must be obtained in the Energy Efficiency category to be eligible for certification. The number of points achievable for lighting and energy-efficient practices and features is capped at 50 points (including 20 bonus points that are obtainable under renewable energy).

The final design achieved the original goal of the design: 1) to reduce more than 40% of the energy consumption and 2) to maximize the daylighting amenity of this factory building. A comparison of annual energy costs for each design alternative was conducted in the different categories of building elements. The material used for both walls and floors is "F14 SLATE", the roof consists of 150 mm heavyweight concrete and windows use "BLUE 6MM". Using this combination, the design collects 42 points in the green mark certification process. From this point of view, this approach realizes the purpose of using the data-driven workflow to guide the energy-efficiency design process.

5. CONCLUSION

Utilizing data-driven workflow, this research conducted a data-oriented modelling process. The process mainly contains five steps: 1) identify the critical design requirements; 2) model the building; 3) implement the data-driven workflow and essential technologies, including transferring the modelling information between different software; 4) apply data-mining, including clustering, classification and associated rule, to perform an integrated discipline analysis for energy and daylighting simulation results; 5) evaluate and refine the effectiveness of the workflow. We anticipate that this workflow will help design teams to formally investigate the performance of many more alternatives during the different design phases, leading to improved built environments. The results of the case study demonstrate our data-driven approach's ability to guide the design process with high precision.

But the quality of the context-based analysis and assumptions depends upon 1) the breadth and quality of memory and experiences that is being drawn upon; 2) the correct identification of metrics or indicators that would accurately categorize and predict the missing attributes or information. Data-mining is performed in order to make assumptions of a high quality and is based on numerous case studies. Thus, there is an essential need to conduct further research to apply this software into more design projects. Additionally, the workflow needs to be reinforced by implementing it into a scientific system workflow software (e.g., Triana, Tavaxy, Kepler). In order to automate the data-driven workflow, the process also needs to adopt parametrical design techniques.

6. ACKNOWLEDGEMENT

This paper is an upgraded version from the ECPPM 2014 conference paper: "Using data-driven approach to support the energy efficiency building design". This research inside this paper is part of the project "Automatic Carbon Calculator Phase 2 – Data Analytics and Carbon Benchmarking" (DRA Ref: 2013-11-01), funded by JTC Corporation, Singapore.

7. REFERENCES

Alleyne Research Group. (2012). Thermosys4 user guide, University of Illinois at Urbana-Champaign. http://web.mechse.illinois.edu/media/uploads/web_sites/104/files/thermosys_2012_user_guide.20130125.5102bdb2147e45.38308161.pdf

Agrawal R., Imielinski T. and Swami A. (1993). Mining association rules between set of items in large databases, in *Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data* (SIGMOD '93), ACM, New York, NY, pp. 207-216.

An D., Kim N.H. and Choi J. (2013). Options for prognostics methods: a review of data-driven and physics-based prognostics, in S. Sankararaman and I. Roychoudhury (eds.), *Proceedings of the Annual Conference of the Prognostics and Health Management Society2013* (PHMC 2013), PHM Society.

Auslander D., Culler D., Wright P.K., Lu Y. and Piette M.A. (2013). A distributed intelligent automated demand response building management system, Final report DIADR 2103,UC Berkeley. http://i4energy.org/downloads/projects/sutardja-dai/DIADRFinalReport.pdf

Bernal W., Behl M., Nghiem T.X. and Mangharam R. (2012). MLE+: a tool for integrated design and deployment of energy efficient building controls, in *Proceedings of the Fourth ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings* (BuildSys '12), ACM, New York, NY.

Building and Construction Authority. (2012). BCA Green Mark for new non-residential buildings, Version NRB/4.1, BCA. http://www.bca.gov.sg/greenmark/others/gm_nonresi_v4.1_rev.pdf

Fayyad U.M., Piatetsky-Shapiro G. and Smyth P. (1996). From data mining to knowledge discovery: an overview, in U.M. Fayyad, G.Piatetsky-Shapiro, P. Smyth and R. Uthurusamy (eds.), *Advances in Knowledge Discovery and Data Mining*, AAAI, Menlo Park, CA.

Jens L. (2008). Energy Efficiency Requirements in Building Codes: Energy Efficiency Policies for New Buildings. http://www.iea.org/publications/freepublications/publication/Building_Codes.pdf

John G.H. (1994). Cross-validated C4.5: using error estimation for automatic parameter selection, Technical report STAN-CS-TN-94-12, Computer Science Department, Stanford University.

Kim H., Stumpf A. and Kim W. (2011). Analysis of an energy efficient building design through data mining approach, *Automation in Construction* 20(1), 37-43.

Krogh A. (2008). What are artificial neural networks? *Nature Biotechnology* 26(2), 195-197.

Liao C., Lin Y. and Barooah P. (2012). Agent-based and graphical modelling of building occupancy, *Journal of Building Performance Simulation* 5(1), 5-25.

Lysen E.H. (1996). The TriasEnergetica:solar energy strategies for developing countries, in A. Goetzberger and J Luther (eds.), *Eurosun Conference 1996*, DGS Sonnenenergie, Freiburg, Switzerland, pp. 16–19.

Mackay D.J.C. (1997). Introduction to Gaussian processes, Technical report, Cambridge University, UK. http://www.cs.utoronto.ca/~mackay/gpB.pdf.

Narayanan S, Apte M.G., Haves P, Piette M.A. and Elliott J. (2010). Systems approach to energy efficient building operation: case studies and lessons learned in a university campus, in *Proceedings of the 2010 ACEEE Summer Study on Energy Efficiency in Buildings*, Omnipress, Madison, WI.

Seeger M. (2004). Gaussian processes for machine learning, *International Journal of Neural Systems* 14(2), 69-106.

Tipping M.E. (2001). Sparse Bayesian learning and the relevance vector machine, *Journal of Machine Learning Research* 1, 211-244.

Wetter M. (2011). Co-simulation of building energy and control systems with the building controls virtual test bed, *Journal of Building Performance Simulation* 4(3), 185–203.

Yao X. (1999). Evolving artificial neural networks, *Proceedings of the IEEE* 87(9), 1423-1447.

Zio E. and Di Maio F. (2010). A data-driven fuzzy approach for predicting the remaining useful life in dynamic failure scenarios of a nuclear system, *Reliability Engineering and System Safety* 95, 49-57.