

Statistical decision assistance for determining energy-efficient options in building design under uncertainty

Singh, M.M., Geyer P.

KU Leuven, Belgium

manavmahan.singh@kuleuven.be

Abstract. Designers need to compare numerous design options in the process of designing an energy-efficient building. There are two impediments in this process, first probabilistic prediction of energy requirement at an early stage of design with uncertain design parameters and, second, selecting a design option based on the probabilistic energy prediction. The paper presents an integration of machine learning energy prediction model with building information modelling (BIM) tool to make probabilistic energy prediction, *i.e.* ranges of values. Wilcoxon rank-sum test is useful in this situation, which is capable of comparing alternatives based on probabilistic energy predictions. The tool has been developed to extract information from the BIM model, make probabilistic energy prediction using the Monte Carlo method, and perform statistical analysis. It has been found that BIM integrated machine learning model can make energy prediction of six design alternatives in 30-35 seconds with no additional modelling efforts. Higher uncertainty in the design parameters will result in larger uncertainty in the energy prediction, and the test may not be able to suggest the better option even using statistical comparison. This will require the more precise value of design parameters, *i.e.* reduced uncertainty. Different uncertainty levels in the design parameters have been tested to which extent they are sufficient to make a selection of the energy-efficient option. It is observed that uncertainty levels that are suitable for decision-making depend on the combination of design options to be compared. It is possible to differentiate among alternatives with high uncertainty in the design parameters if they are entirely different else more precise definition of the design parameters is required. This research provides a method to select a better option among the developed options based on energy performance at the early stage of design.

1 Introduction

It has been a challenging endeavour for designers to develop an energy-efficient design to meet the demand for low energy use buildings. The first step in this endeavour to design a building envelope which results in lower energy loads, *i.e.* heating and cooling load (Jin and Jeong, 2014; Méndez Echenagucia *et al.*, 2015). A designer develops several design options for building envelope which needs to be evaluated for the energy performance. The design of the building envelope takes place at the early stage of design, which offers several challenges for energy performance evaluation. But, the decision-making at the early stage enhances the performance of building with a lesser cost of change (MacLeamy, 2004). Thus, it is imperative for a designer to make informed decisions at an early stage of design from an energy-efficiency perspective.

There are two primary challenges for informed decision-making at the early stage for energy-efficient design. First, make probabilistic energy performance prediction of the design options with uncertain design parameters and, second, use the probabilistic prediction to identify better solutions. Since, a number of design parameters influencing the energy loads are uncertain at the early stage of design (Tian *et al.*, 2018), the probabilistic energy is estimated by simulating a large number of energy models (Van Gelder, Janssen and Roels, 2014). This utilises a Monte Carlo method to generate random samples of uncertain design parameters. The random samples are used with the design information (certain design parameters) to create energy models. Since

the simulation of a large number of energy models using conventional energy simulation tools such as EnergyPlus or TrnSys is time-consuming, the use of machine learning models is suggested in previous research works (Van Gelder, Janssen and Roels, 2014; Schlueter and Geyer, 2018; Singaravel *et al.*, 2018).

The energy performance of a design option is predicted using this approach, as probabilistic heating and cooling loads. The probabilistic heating and cooling loads will be a range of values; thus, it is not possible to compare the design options using quantitative comparison. The paper proposes the use of statistical method, Wilcoxon rank sum test, to assess whether two design options are significantly different from each other (Mann and Whitney, 1947). Based on the level of uncertainty in the design parameters, it may or may not be possible to draw the difference between two options. If options are indifferent against each other, reduction of the uncertainty in the design parameters is required. In this paper, this approach has been tested by reducing the uncertainty in the design parameters and its effect on the option selection. The objective of this research is to:

1. Perform quick probabilistic energy prediction to consider design uncertainty in decision making using building information modelling (BIM) with an integrated machine learning model.
2. Assist decision-making based on probabilistic energy prediction results using a statistical test for comparison with different levels of uncertainties in the design parameters. Assist means in this context to provide the designer with information if a decision between two design options is possible or if more information is required.

2 Literature review

The energy performance analysis tools offer limited integration with the design process due to the need for extensive modelling efforts and high computational time (Schlueter and Geyer, 2018). There have been few attempts to integrate energy prediction tools with the design process by the integration of BIM model with energy models (Ahn *et al.*, 2014; Negendahl, 2015). However, the challenges of the early stage of design are not well addressed. To perform quick energy predictions, integrating energy prediction model with the design process will be of paramount importance, which is possible with an integrated BIM tool. The component-based machine learning (CBML) model are developed to allow better integration with multi-level-of-detail (multi-LOD) BIM approach and extensibility to complex design cases (Geyer and Singaravel, 2018). The concept of CBML is based on the decomposition of design artefact and engineering knowledge and predicting the intermediate parameters such as heat flows before predicting the zone heating and cooling load. The use of deep learning technologies for the second generation of models improves this capability (Singaravel *et al.*, 2018). However, energy prediction using CBML at the early stage of design is performed after generating the building elements using rules (Geyer, Singh and Singaravel, 2018). Thus, it becomes a simplified representation of the building energy model, which will have a similar limitation as to the developed CBML.

One challenge of energy prediction at an early stage of design is uncertain design parameters. This issue makes the energy performance prediction process more complicated. Gelder *et al.*, (2014) has proposed the use of probabilistic prediction of energy performance with uncertain information using the Monte Carlo method and engineering surrogate models. The probabilistic estimation of energy performance with uncertain inputs is possible with the use of time-efficient meta-models and machine learning models (Van Gelder, Janssen and Roels, 2014; Singaravel

et al., 2018). However, the energy prediction model needs to be integrated with BIM tool to streamline the energy prediction process with design.

The decision-making exercise for building design is very complex, which involve assessment of building based on several contradicting performance measures such as energy consumption and thermal comfort etc. The simulation tools are used in assessing the building on these performance measures and assist in selecting a better option (de Wilde, Augenbroe and van der Voorden, 2002; Wilde and Voorden, 2004). The problem gets compounded with the inherent uncertainty in the design parameters. Thus, it is required to perform multi-criteria decision-making based on the probabilistic distribution of the performance measure (Hopfe, Augenbroe and Hensen, 2013; Rezaee *et al.*, 2015). However, the use of simulation or computation tool for decision-making requires much more efforts for collaboration among the designers and engineers (Alsaadani and Bleil De Souza, 2016). The researchers suggested using the simulation results to provide meaningful insights which assist the designers in the design process (Bleil de Souza and Tucker, 2015).

Wilcoxon rank sum test is used to test whether the randomly selected value from one sample is statistically different from a randomly selected value from another sample (Mann and Whitney, 1947; Corder and Foreman, 2014). This test is a non-parametric alternative to the two-sample t-test. Thus it is applied when the samples are not normally distributed. The method is used in quite widely used in other domain such as medicine, bio-informatics and environmental engineering (Gauthier and Hawley, 2007) but rarely used for building design application (Hu, 2019).

3 Research methodology

The research methodology is described in two parts: First, integrated tool for quick energy prediction and, second, statistical analysis to differentiate among options.

3.1 Probabilistic prediction of energy performance using BIM-integrated ML methods

CBML model is developed based on the decomposition of an energy prediction model to a building element level (Geyer and Singaravel, 2018). The element level model first predicts the heat flow for each building component - walls, floors, roof, window and ground floor. The heat flow is supplemented with the other information such as operating hours, internal heat gains and infiltration to predict the heating or cooling load at zone level, which is summed up for building level. So, the energy load prediction requires the information to be present in the BIM model at building component level. At the early stage of design, the designer develops a mass model representing the external envelope of the building. But there is no information present about building element at the early stage of the design and need to be generated using assumptions. The building elements are created using the one-zone-per-floor method described in Geyer *et al.*, (2018) (Geyer, Singh and Singaravel, 2018) as shown in Figure 1.

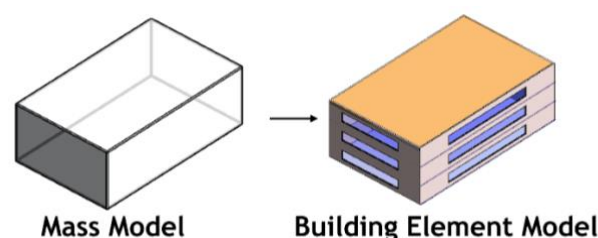


Figure 1. Generation of building elements in the mass model using the rules (Geyer *et al.*, 2018)

After creating the building elements, namely walls, floors and roof, the geometrical parameters for these elements are certain and extracted from the BIM model (data extraction). The uncertain design parameters such as technical specifications, window construction and operational design parameters are provided using user inputs, as mentioned in Table 1. For the uncertain design parameters, the Monte Carlo method is used to generate a number of random combinations. It will result in the same number of energy models to be evaluated using a machine learning model. After this step, the data is passed to CBML for energy load predictions and results are integrated back to the BIM model.

Table 1. Data collection for the CBML model

Information	Type	Source
Geometrical parameters for wall, roof, floors and roof	Certain	BIM model (Data extraction)
Technical specifications parameters ¹	Uncertain	User input form (Monte Carlo method)
Window construction parameters ²		
Operational design parameters ³		

¹Technical specifications parameters – u-value of wall, floor, roof and infiltration

²Window construction parameters – the window to wall ratios in each direction, u-value and g-value of windows

³Operational design parameters – internal heat gains and operating hours

In this research, newly developed CBML is used. It follows a similar component structure as in Geyer and Singaravel, 2018. The training data is enriched with more architectural design cases to get better performance of ML models. The newly developed models are tested on the design case outside of training data, and the performance of the models is described in Figure 2. The model shows the *goodness-of-fit* (R^2) 0.998 and 0.986 for heating and cooling load, respectively. Thus, the model can be used for energy load predictions.

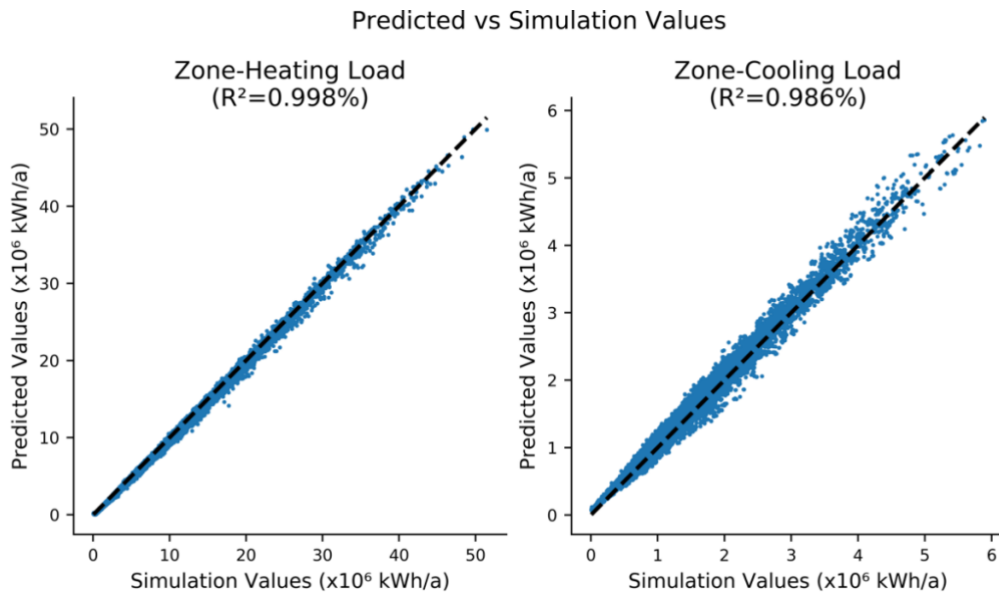


Figure 2. Performance of component-based machine learning model

The authors have developed a method integrated with a BIM authoring application to automate the process of making probabilistic energy predictions using CBML. The complete process of integration is described in Figure 3. The geometrical parameters are extracted from the BIM model, which is available as design information in the model. The remaining parameters are unknown at an early stage of the design and collected using a user input form as uncertain information. This information is passed to the CBML model to make energy predictions and integrate the results back to the BIM model.

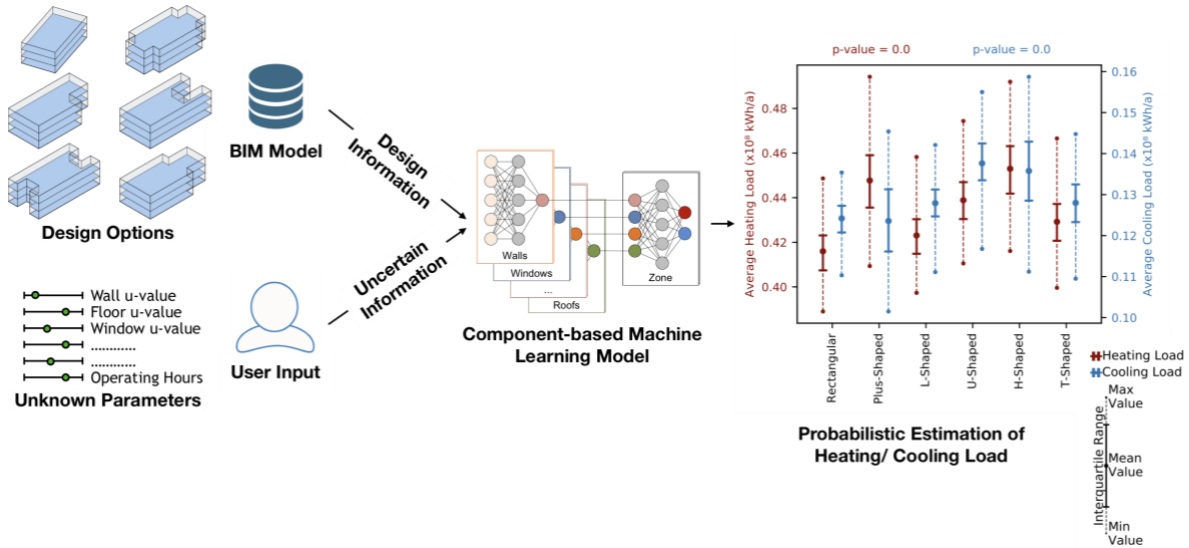


Figure 3. Probabilistic prediction of energy performance using CBML

3.2 Using statistical analysis to differentiate among options

The results are statistically analysed using non-parametric test - Wilcoxon rank sum test. Wilcoxon rank sum test is used to determine whether two independent samples selected from the population are significantly different from each other (Corder and Foreman, 2014). In this case, it is used to determine whether two design options developed by the designer differ significantly from each other concerning energy performance. If this is the case, selection and decision making is the next step of design development. Otherwise, the uncertain parameters need to be defined more precisely to differentiate among the design options. The parametric tests are not applicable as the data is not normally distributed. The probability value (p-value) is calculated for each combination of design options. The p-value is considered significant if it is less than 0.05. This p-value means we can reject the null hypothesis and state that the design options under consideration are significantly different from each other.

4 Results

The method is implemented a testcase building Tausenpfund – an average size office building in Munich. It has a rectangular floor plan with a floor area of 1200 sq. m. (3×14.8×27), distributed equally on three floors. The method is implemented on five design options developed as alternatives to rectangular floor plan building.

4.1 Integrated tool to make energy prediction using CBML model

The method based on the CBML model is integrated with BIM application Autodesk Revit to demonstrate the presented approach. The six design options are stored as conceptual masses in the application and accessed through the plugin. The successful completion of the tool allows the energy prediction and integration of energy prediction results back to the BIM model.

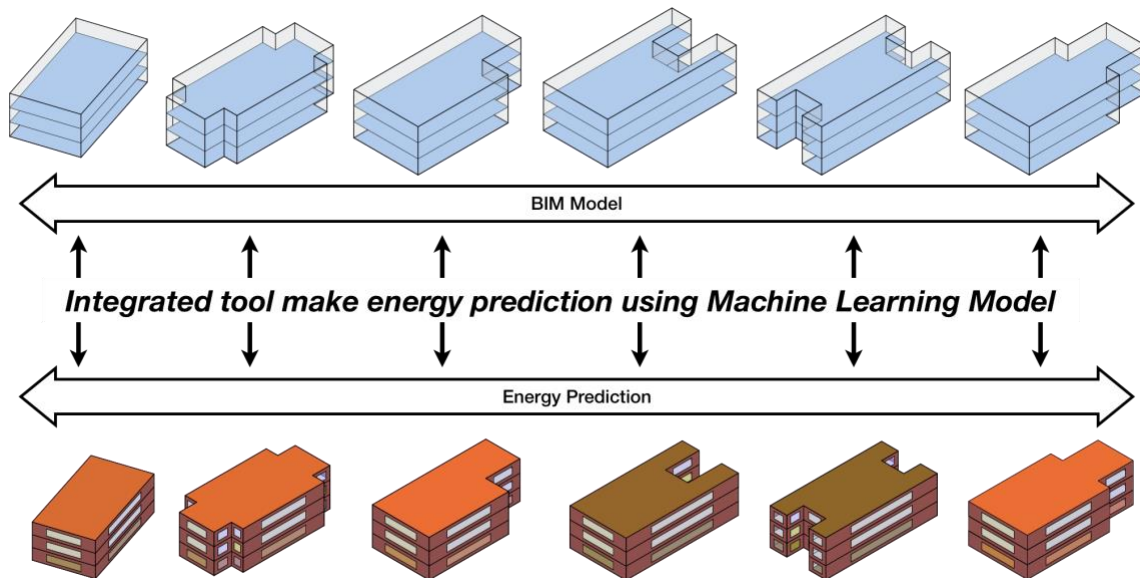


Figure 4. Integrated tool to make energy prediction using CBML model

The accuracy of the prediction is the same as the accuracy of CBML. In terms of time-efficiency, it takes 30-35 seconds to make energy prediction of six design options using 500 random samples. It means a total of 3000 energy models are evaluated in approximately 30 seconds. The process may take up to 24 hours using traditional energy simulation tools. The energy prediction is made based on the mass model and user input, which requires no additional modelling for energy prediction.

4.2 Statistical analysis to differentiate among options

Four scenarios with different level of uncertainty for input parameters are analysed to ascertain whether it is possible to make a distinction between the design options. There is a total of 15 combinations for six design options. The level of uncertainty and corresponding significant p-values for heating and cooling load is presented in Table 2 and Figure 5. The number of significant p-values is higher as the level of uncertainty decreases. With the uncertainty of $\pm 25\%$ in technical specification and operational design parameters and $\pm 50\%$ in window construction parameters, the number of significant p-values is 6 and 8 for heating and cooling load respectively for a total of 15 combinations. The number of significant p-values is 15 and 14 for heating and cooling load, respectively when the uncertainty in all the uncertain parameters is $\pm 5\%$.

Table 2. Level of uncertainty in uncertain parameters

	Technical specifications parameters	Operational design parameters	Window construction parameters	No. of significant p-values (Heating Load, Cooling Load)
Scenario 1	25	25	50	6/15, 8/15
Scenario 2	5	25	50	7/15, 9/15
Scenario 3	5	5	50	13/15, 9/15
Scenario 4	5	5	5	15/15, 14/15

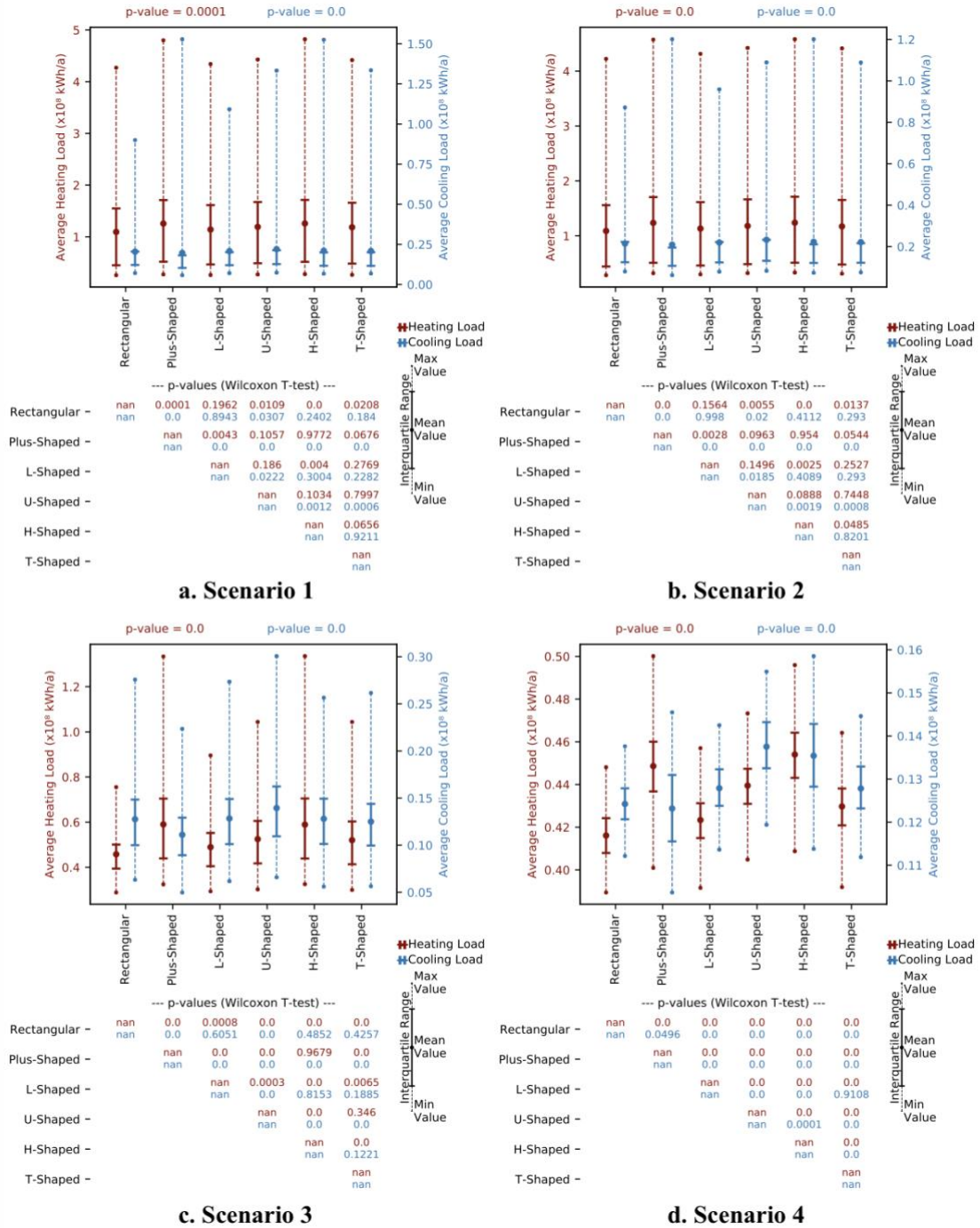


Figure 5. Energy prediction results with statistical analysis in four scenarios

5 Discussion

This paper has presented an approach of BIM integrated CBML model for probabilistic energy prediction at an early stage of design in a time-efficient way. The subsequent application of statistical analysis indicated for decision-making if two options show a significant difference in their probabilistic energy prediction of designed alternatives. The presented approach of making probabilistic estimation and statistical analysis has several limitations which are discussed in this section. The presented approach utilises CBML model, which has a limitation as all ML models, which is the applicability of models to new design cases, i.e. the generalisation. In the presented case, CBML model covers the range of design cases mentioned in the approach and energy prediction results are trustworthy. Hence, the applicability of CBML model needs to be tested extensively before using it on new design case beyond the range. A comparison of the final design configuration is recommended in this situation.

The method is only applicable to uniform floor plan buildings. The generation of building elements based on the one-zone-per-floor rule is another simplification of the energy model. On the one hand, adaptations of the method for floor plans that differ on each level are possible; however, this requires a trade-off between modelling efforts and model accuracy as well as training effort. On the other hand, such simplifications are commonly used for energy prediction at the early stage and comply with the approach of architectural design. The essential advantage of the CBML method is that it allows a higher degree of flexibility than monolithic parametric models and provides reusability of the component model and easy integration with BIM model. Flexible detail levels and creation of missing building elements are possible by generative methods.

The approach presents four scenarios with varying levels of uncertainty in the design parameters. It can be noted that the number of significant p-values increases with a lesser level of uncertainty. This fact shows the possibility of differentiating among all the options with a decreasing level of uncertainty. Also, the p-values dependent on the combination of options to be compared, so some combinations are different from each other even with a higher level of uncertainties and some are not different from each other even with precise value of design parameters. Thus, it may be possible that a designer needs to detail out one or more options before the uncertain parameters are defined more precisely.

The level of uncertainty in the design parameters is assumed to be uniformly distributed between the range of mean \pm variation, however other distributions such as triangular or normal distributions are also possible which may result in lesser uncertainty in the energy prediction results. Which means the decision can be made with higher uncertainty in the design parameters. Also, if the energy prediction results are normally distributed because of another type of distribution in the design parameters, it will be possible to use a powerful parametric statistical test.

6 Conclusion

The architectural design is assisted by making a probabilistic prediction of energy performance and statistically analysing the results for decision making. The approach offers design assistance in requiring information and prioritising decisions at the early stage of design by making quick energy prediction with BIM integrated machine learning-based energy prediction model and analysing the response statistically.

The process of making energy prediction is streamlined with the use of BIM integrated machine learning model at building element level. The inherent uncertainty at the early stage of the

design reflects in the energy prediction process. It requires an evaluation of several energy models for making probabilistic energy predictions. The whole process is made faster by extraction of data from the BIM model and energy predictions using CBML method. This method provides quick energy prediction results for the comparison of alternatives. In contrast to tradition energy simulation tool, the machine learning model is at least 3000 times faster. The accuracy of energy prediction is the same as the accuracy of the developed CBML mode, which is 0.998 and 0.986 (R^2 values) for heating and cooling load, respectively.

The uncertainty in the design parameters gets translated into the uncertainty in the energy predictions, which requires statistical analysis for comparison. The statistical test allows comparing the architectural design option considering the uncertainty that is present at the early stage of design. The possibility to compare the design alternatives depends on the uncertainty in the design parameters and the combination of design alternatives, as shown in the case presented in this paper. It will require a more precise definition of the design parameters to differentiate between two similar design options if the statistical comparison is not possible.

The integrated tool with BIM and CBML model offers an opportunity to address the design of energy-efficient buildings. There are some limitations of the current implementation of the approach such as simplifications of the energy model, applicability of machine learning model to new design case, uniform distribution of the design parameters and data extraction possible only for uniform floor plan buildings. It will require future research to address these limitations and to extend the applicability of the approach to more design case.

Acknowledgement

The research was funded by the Deutsche Forschungsgemeinschaft (DFG) in Researcher Unit 2363 “Evaluation of building design variants in early phases using adaptive levels of development”, in Subproject 4 “System-based Simulation of Energy Flows”. The computational resources and services used in this work were provided by the VSC (Flemish Supercomputer Center), funded by the Research Foundation - Flanders (FWO) and the Flemish Government – department EWI.

References

- Ahn, K. U. *et al.* (2014) ‘BIM interface for full vs. semi-automated building energy simulation’, *Energy and Buildings*. Elsevier B.V., 68(PART B), pp. 671–678. doi: 10.1016/j.enbuild.2013.08.063.
- Alsaadani, S. and Bleil De Souza, C. (2016) ‘Of collaboration or condemnation? Exploring the promise and pitfalls of architect-consultant collaborations for building performance simulation’, *Energy Research & Social Science*, 19, pp. 21–36. doi: 10.1016/j.erss.2016.04.016.
- Bleil de Souza, C. and Tucker, S. (2015) ‘Thermal simulation software outputs: a framework to produce meaningful information for design decision-making’, *Journal of Building Performance Simulation*, 8(2), pp. 57–78. doi: 10.1080/19401493.2013.872191.
- Corder, G. W. and Foreman, D. I. (2014) *Nonparametric statistics for non-statisticians: a step-by-step approach*. Second. Wiley.
- Gauthier, T. D. and Hawley, M. E. (2007) ‘STATISTICAL METHODS’, in *Introduction to Environmental Forensics*. Elsevier, pp. 129–183. doi: 10.1016/B978-012369522-2/50006-3.
- Van Gelder, L., Janssen, H. and Roels, S. (2014) ‘Probabilistic design and analysis of building performances: Methodology and application example’, *Energy and Buildings*, 79, pp. 202–211. doi: 10.1016/j.enbuild.2014.04.042.
- Geyer, P. and Singaravel, S. (2018) ‘Component-based machine learning for performance prediction in building design’, *Applied Energy*, 228, pp. 1439–1453. doi: 10.1016/j.apenergy.2018.07.011.
- Geyer, P., Singh, M. M. and Singaravel, S. (2018) ‘Component-Based Machine Learning for Energy Performance

- Prediction by MultiLOD Models in the Early Phases of Building Design', in *Advanced Computing Strategies for Engineering*, pp. 516–534. doi: 10.1007/978-3-319-91635-4_27.
- Hopfe, C. J., Augenbroe, G. L. M. and Hensen, J. L. M. (2013) 'Multi-criteria decision making under uncertainty in building performance assessment', *Building and Environment*, 69, pp. 81–90. doi: 10.1016/j.buildenv.2013.07.019.
- Hu, M. (2019) 'Does zero energy building cost more? – An empirical comparison of the construction costs for zero energy education building in United States', *Sustainable Cities and Society*, 45, pp. 324–334. doi: 10.1016/j.scs.2018.11.026.
- Jin, J.-T. and Jeong, J.-W. (2014) 'Optimization of a free-form building shape to minimize external thermal load using genetic algorithm', *Energy and Buildings*, 85, pp. 473–482. doi: 10.1016/j.enbuild.2014.09.080.
- MacLeamy, P. (2004) *MacLeamy Curve*. Available at: <http://www.msa-ipd.com/MacleamyCurve.pdf> (Accessed: 29 September 2015).
- Mann, H. B. and Whitney, D. R. (1947) 'On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other', *The Annals of Mathematical Statistics*, 18(1), pp. 50–60. doi: 10.1214/aoms/1177730491.
- Méndez Echenagucia, T. *et al.* (2015) 'The early design stage of a building envelope: Multi-objective search through heating, cooling and lighting energy performance analysis', *Applied Energy*, 154, pp. 577–591. doi: 10.1016/j.apenergy.2015.04.090.
- Negendahl, K. (2015) 'Building performance simulation in the early design stage: An introduction to integrated dynamic models', *Automation in Construction*, 54, pp. 39–53. doi: 10.1016/j.autcon.2015.03.002.
- Rezaee, R. *et al.* (2015) 'Assessment of uncertainty and confidence in building design exploration', *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, 29(4), pp. 429–441. doi: 10.1017/S0890060415000426.
- Schlueter, A. and Geyer, P. (2018) 'Linking BIM and Design of Experiments to balance architectural and technical design factors for energy performance', *Automation in Construction*, 86, pp. 33–43. doi: 10.1016/j.autcon.2017.10.021.
- Singaravel, S. *et al.* (2018) 'Deep-learning neural-network architectures and methods: Using component-based models in building-design energy prediction', *Advanced Engineering Informatics*, 38, pp. 81–90. doi: 10.1016/j.aei.2018.06.004.
- Tian, W. *et al.* (2018) 'A review of uncertainty analysis in building energy assessment', *Renewable and Sustainable Energy Reviews*, 93, pp. 285–301. doi: 10.1016/j.rser.2018.05.029.
- de Wilde, P., Augenbroe, G. and van der Voorden, M. (2002) 'Design analysis integration: supporting the selection of energy saving building components', *Building and Environment*, 37(8–9), pp. 807–816. doi: 10.1016/S0360-1323(02)00053-7.
- Wilde, P. de and Voorden, M. van der (2004) 'Providing computational support for the selection of energy saving building components', *Energy and Buildings*, 36(8), pp. 749–758. doi: 10.1016/j.enbuild.2004.01.003.