Sustainable Cities and Society 50 (2019) 101596

Contents lists available at ScienceDirect

Sustainable Cities and Society

journal homepage: www.elsevier.com/locate/scs

Assessing environmental performance in early building design stage: An integrated parametric design and machine learning method

Kailun Feng[a,b], Weizhuo Lu[a,*], Yaowu Wang[b,c,d]

[a] Department of Civil, Environmental, and Natural Resources Engineering, Luleå University of Technology, Luleå 97187, Sweden
[b] Department of Construction Management, Harbin Institute of Technology, Harbin 150001, China
[c] Key Lab of Structures Dynamic Behavior and Control of the Ministry of Education, Harbin Institute of Technology, Harbin 150090, China
[d] Key Lab of Smart Prevention and Mitigation of Civil Engineering Disasters of the Ministry of Industry and Information Technology, Harbin Institute of Technology, Harbin 150090, China

ARTICLE INFO

Keywords:
Building early design
Parametric design
Machine learning
Environmental impact
Prediction intervals

ABSTRACT

Decisions made at early design stage have major impacts on buildings' life-cycle environmental performance. However, when only a few parameters are determined in early design stages, the detailed design decisions may still vary significantly. This may cause same early design to have quite different environmental impacts. Moreover, default settings for unknown detailed design parameters clearly cannot cover all possible variations in impact, and Monte Carlo analysis is sometimes not applicable as parameters' probability distributions are usually unknown. Thus, uncertainties about detailed design make it difficult for existing environmental assessment methods to support early design decisions.

Thus, this study developed a quantitative method using parametric design technology and machine learning algorithms for assessing buildings' environmental performance in early decision stages, considering uncertainty associated with detailed design decisions. The parametric design technology creates design scenarios dataset, then associated environmental performances are assessed using environmental assessment databases and building performance simulations. Based on the generated samples, a machine learning algorithm integrating fuzzy C-means clustering and extreme learning machine extracts the case-specific knowledge regarding designed buildings' early design associated with environmental uncertainty. Proposed method is an alternative but more generally applicable method to previous approaches to assess building's environmental uncertainty in early design stages.

1. Introduction

With the rapid urbanization in many regions, especially developing countries, large volumes of new buildings and infrastructures are under construction (Li & Yao, 2009). The new built buildings in urbanization increase the energy consumption and environmental burdens such as CO_2 emissions in cities (Alalouch, Al-Saadi, AlWaer, & Al-Khaled, 2019; Vázquez-Canteli, Ulyanin, Kämpf, & Nagy, 2019). To assist efforts to address these challenges, Ding (2008) found the life-cycle environmental performance of buildings is able to be best handled by making optimal early design decisions. Many design parameters with significant effects on environmental impact will be decided in this stage (Kohler & Moffatt, 2003). However, traditional building design is an evolutionary process, so the detailed design decisions are not fully known in early stages. This greatly hinders use of current environmental assessment methods, which require detailed design information regarding parameters such as material take-off for life-cycle assessment (LCA) (Kim, Yun, Cho, & Ha, 2017), and buildings' operational thermal and equipment properties for building performance simulation (BPS) (Gervásio, Santos, Martins, & da Silva, 2014). Thus, the information required for a full environmental assessment is not available in early design stages (Schlueter & Thesseling, 2009).

Unpredictability of the detailed design information makes it difficult to assess the environmental impact accurately during early stages (Rezaee, Brown, Augenbroe, & Kim, 2015). Kim et al. (2017) and Attia, Gratia, De Herde, and Hensen (2012) thus concluded that there is a lack of appropriate evaluation tools for designers to make environmentally friendly decisions in early stages of building design. To address this challenge, in previous studies unspecified detailed parameters have usually been set to default values (Ochoa & Capeluto, 2009) or Monte Carlo analysis has been applied (Huijbregts, Gilijamse, Ragas, & Reijnders, 2003). However, deterministic evaluations with default

* Corresponding author.
E-mail addresses: kailun.feng@ltu.se (K. Feng), weizhuo.lu@ltu.se (W. Lu), ywwang@hit.edu.cn (Y. Wang).

https://doi.org/10.1016/j.scs.2019.101596
Received 13 September 2018; Received in revised form 5 February 2019; Accepted 9 May 2019
Available online 22 May 2019
2210-6707/

settings cannot provide ranges of uncertainties and risks associated with early decisions (Rezaee et al., 2015). Possible permutations of uncertain parameters have been explored using Monte Carlo analyses to quantify performance uncertainty, in various studies, e.g. (Rezaee et al., 2015) and (Hygh, DeCarolis, Hill, & Ranjithan, 2012). However, reasonable probability distributions of unspecified parameters, which are not usually available, are needed in this sampling approach to obtain accurate performance distributions (De Wit & Augenbroe, 2002).

We thus propose an alternative approach, involving integration of parametric design technology and machine learning algorithms to quantify environmental uncertainties associated with early design decisions. In this approach, parametric design technology is used to generate reasonable design datasets that enable an algorithm combining fuzzy C-means clustering and an extreme learning machine (FCM-ELM) to learn the relationships between early design decisions and uncertainty of environmental impacts. Thus, the proposed method does not require knowledge of distributions of uncertain parameters because it makes use of big design datasets to directly infer effects of early design decisions on building performance uncertainty. The uncertainties in performance are quantified by providing prediction intervals for the planned buildings' environmental impacts. These prediction intervals describe uncertainty by specifying the lower and upper limits within which future building environmental performance is expected to fall, for some pre-defined confidence level (e.g. 95%).

The rest of the paper is structured as follows. In Section 2, we review the background of evolutionary building design and previous early design literature, and highlight the potential of the proposed method to overcome previous limitations. Section 3 describes the parametric design technology we use, presents the environmental impact assessment, then describes the uncertainty quantification using the FCM-ELM algorithm. In Section 4, we describe the evaluation of our method in a case study that considers the design of a low-rise individual-room office building. The benefits and potential applications of the proposed method are further discussed in Section 5. Finally, the theoretical and practical contributions of the study, along with its limitations and future directions, are summarized in Section 6.

2. Background

2.1. Building design stages

Conceptual, *schematic*, and *detailed* design are three building design stages that are conducted in an evolutionary manner (Mujumdar & Maheswari, 2018). In the conceptual design stage, buildings' basic geometric parameters (height, width, length, orientation, etc.) are decided. Decisions concerning the building's general floor plan, number of floors, major structural materials, and number of rooms are decided in the schematic design stage. The design parameters in the conceptual and schematic design stages are sometimes interwoven and can be regarded collectively as early design parameters. In the detailed design stage, the exact sizes of windows, walls and roofs are decided, as well as the materials from which they will be constructed. Structural engineers, for example, will design the building's structural components and electrical engineers will design its facility system. In an evolutionary design process, the detailed design parameters are not decided during the early stages, and only a rough idea or a number of possible choices for these parameters will be known (Gervásio et al., 2014). The paucity of this information greatly hinders assessment of the environmental performance of early design decisions.

However, decisions taken during early design stages significantly affect buildings' future environmental impact, as illustrated in Fig. 1. Moreover, changing early designs later is very difficult and costly, as detailed design decisions may have already been based on them. It is also extremely difficult to compensate for poor early design choices by the detailed design decisions (Chong, Chen, & Leong, 2009). Therefore, decisions in early design stages are important and have significant impacts on buildings' life cycle environmental impact.

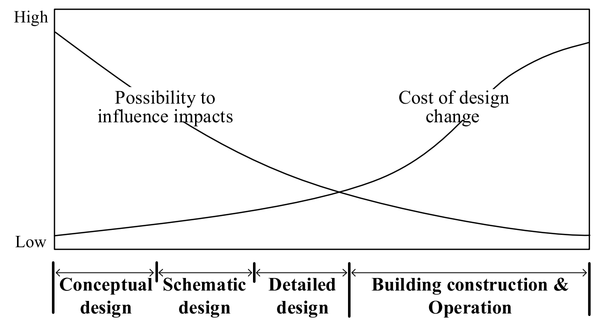

Fig. 1. Schematic illustration of the influence of building design decisions on life-cycle environmental impacts (modified from MacLeamy Curve (MacLeamy, 2004)).

2.2. Early design evaluation

Early building design has attracted considerable research attention recently due to its importance in determining buildings' life cycle performance. These researchers have identified several major sources of uncertainty (Table 1). One is the lack of knowledge about detailed building design parameters, which have not yet been decided in the early design stages. Another is occupancy behaviour that is unpredictable until a building starts operation. Other sources of uncertainty including errors in environmental modelling, software limitations, and discrepancies between buildings' theoretical and actual physical properties, throughout their life cycles. Here, we propose a novel way of handling uncertainty of undecided design, and thus improving early design decisions. Other sources of uncertainty are beyond the scope of the study.

In some recent published attempts to determine environmental impacts at an early design stage (Asadi, Amiri, & Mottahedi, 2014; Ochoa & Capeluto, 2009; Tian, Yang, Zuo, Li, & Liu, 2017), unknown variables have been set to default values. This approach has also been applied in relevant software, e.g. Insight 360 and Green Building Studio. In this way the influence of early design parameters on environmental impact will be analyzed. However, a single set of default values clearly cannot cover multiple possible design scenarios, and assigning the same values to different design possibilities implicitly assumes that choices of unknown parameters will have no effect on environmental performance. Rezaee et al. (2015) maintained that default values of unspecified parameters could impair or improve a building's environmental performance, thus influence the optimality of early design decisions. And Hiyama, Kato, Kubota, and Zhang (2014) maintained that significant differences between default and actual values are likely to reduce the validity of early decisions.

In several other studies, Monte Carlo analyses have been used to obtain values for unspecified parameters, rather than using default values (Macdonald, 2002; Macdonald & Strachan, 2001; Rezaee et al., 2015). In this approach, possible detail design scenarios and associated performance distributions are generated in Monte Carlo simulations, thus providing predictive indications of effects of unspecified detailed design parameters to inform early design decisions. However, knowledge of the probability distributions of the unknown parameters is required for robust Monte Carlo analysis (Huijbregts et al., 2003). This is highly problematic, because even if designers know possible detailed design parameters, it is challenging to determine their probability distributions (De Wit & Augenbroe, 2002; Macdonald & Strachan, 2001), especially as they will be influenced by designers' personal preferences.

A possible solution is to use machine learning algorithms, which can quantify uncertainties from appropriate datasets without knowledge of parameters' probability distributions (Shrestha & Solomatine, 2006), in conjunction with parametric design technology, which can create

Table 1
Uncertainty sources of building performance assessment.

Source	Description	Span of uncertainty	References
Undecided design	All the detailed design parameters that are unknown in early design stages	Early design stages	Gervásio et al. (2014)
Occupancy scenarios	Lack of knowledge of people's occupancy behaviour	All design stages	Rezaee et al. (2015)
Modelling	Errors in models used to simulate conditions, e.g. weather	Full building life	Østergård, Jensen, and Maagaard (2016)
Software limitations	All kinds of limitations in used software tools	Full building life	Macdonald (2002)
Physical uncertainty	Discrepancies between theoretical and actual building properties	Full building life	Rezaee et al. (2015)

alternative designs in various architectural and other design disciplines (Monizza, Bendetti, & Matt, 2018). Machine learning algorithms have been applied for building performance prediction, such as energy consumption (Manfren, Aste, & Moshksar, 2013; Vázquez-Canteli et al., 2019) and environmental impacts estimation (Zhou, Zhou, Zhu, & Li, 2015). And parametric design is frequently used to clarify the "relationship between design intent and design response" (Jabi, 2013). It is thus interesting to investigate the possibility of integrating machine learning and parametric design technology to infer relationships between early design decisions and associated environmental uncertainties.

3. Quantification of environmental uncertainty

The environmental uncertainty quantification method we propose comprises three main modules (Fig. 2). The first module explores possible design scenarios and generates full information samples (FIS) by parametric design technology. Then the design information in the FIS, including materials quantity take-off, building geometry, and thermal properties, is automatically imported into an environmental impact assessment module.

For each design sample, data from an environmental database are used to calculate embodied and recycling impacts, and building performance simulation is used to calculate operational impacts. The final assessment results and related design parameters are transferred to an environmental uncertainty modelling module.

Based on the FIS samples generated by the first module and the assessment results from the second module, the third module runs a FCM-ELM integrated algorithm to compute and model prediction intervals for environmental impact metrics, such as global warming potential. First, the samples are clustered, and prediction intervals for the clusters and individual samples are quantified. Then, the ELM model learns the connections between prediction intervals and early designs. In this way, an uncertainty feedback model is built, which then is used to support early design decisions.

3.1. Design sample generation

The parametric design technology connects design parameters such as building shape, orientation, floor plan and other building parameters into a building model and enables coordinated amendment of building design (Eltaweel & Yuehong, 2017). In early building design stages, parametric design approaches have utility for designers to explore design variations considering building performance feedback (Holzer, 2016). In our approach, the explorative ability of parametric design technology (Fig. A1 and Table A1) is used to generate full information samples (FIS) represented by building information modelling. Each design scenario is based on predefined design constraints to ensure that it is reasonable. These design samples form a design dataset that is used to infer relationships between early design decisions and associated environmental uncertainties.

Three functional sections are defined in established parametric design technology. They are input section, process section, and information extraction section, which work sequentially in parametric design technology. The *input section* selects early design parameters and the undecided parameters randomly, subject to any constraints imposed by the early designs. This information is used as input data for the *process section* that generates FIS. Then information related to the building geometry, material, and operational properties is exported to the environmental impacts assessment using the *information extraction section*.

The FIS generated by the parametric design technology are divided into three disjoint sets, in keeping with standard statistical learning

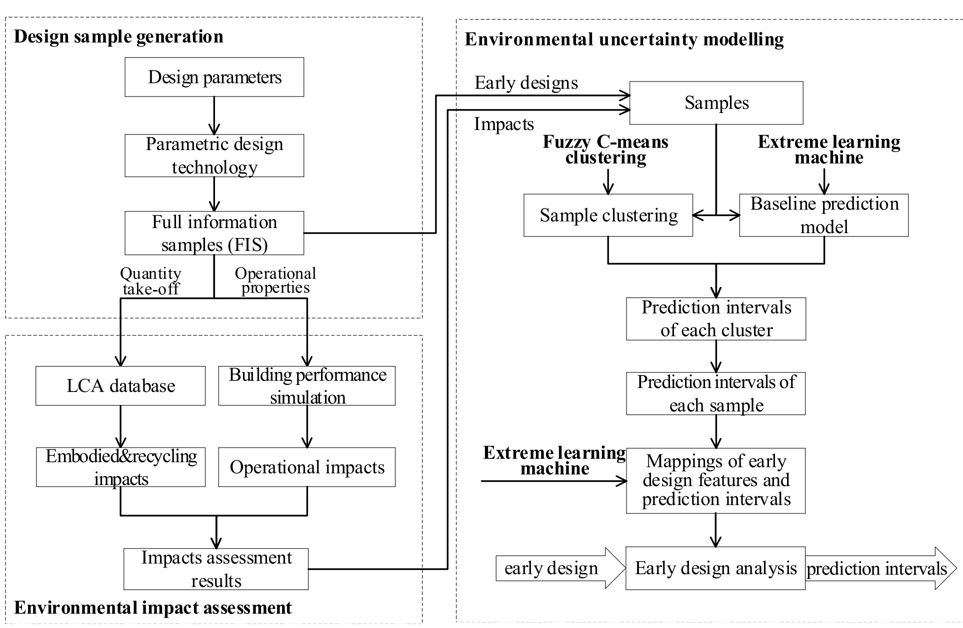

Fig. 2. Flow chart of developed quantification method for assessing environmental uncertainty in an early design stage.

Table 2
Assessment boundary for early building design.

Boundary	Embodied impact			Operational impact		End-of-life recycling
	Raw material supply	Transport	Manufacture	Operational energy use	Operational water use	Reuse/recovery/recycling
Standard code (EN 15804:2012)	A1	A2	A3	B6	B7	D
Identified impact sources (Dixit, Fernández-Solís, Lavy, & Culp, 2012; China, 2006)	Building materials			Heating-natural gas; cooling-electricity; lighting-electricity; ventilation-electricity; operating equipment-electricity	Living water; drinking water	Recycled building materials
Assessment data sources	ICE database (Hammond & Jones, 2011)			Emission factors: electricity (Yang, 2002; Zhou & Su, 2009); water (Yang, 2003); natural gas (Yang, 2002); and characterization factors from (IPCC, 2009)		ICE database (Hammond & Jones, 2011)

protocols (Friedman, Hastie, & Tibshirani, 2001), namely training, validation, and testing sets. The training samples are used to train the uncertainty quantification model. The validation samples are used to choose the optimal model settings. Finally, testing samples are used to evaluate the general performance of the established model.

3.2. Environmental impacts assessment

The ISO has standardized the environmental assessment process, starting from goal and scope definition (ISO, 2006). The scope for this study is shown in Table 2. The environmental impact in this study comprises building's embodied and operational impacts, as well as the benefits from material end-of-life recycling. Embodied and recycling impacts are governed by the building materials used for construction, whereas operational impact is governed by the building's energy and water consumption during building operation. Many previous study such as (Li, Chen, Hui, Zhang, & Li, 2013; Zhang, Shen, & Zhang, 2013) found the impacts of materials embodied, building operation, and end-of-life recycling are the major contributors to a building's life cycle impact, other impacts from construction, retrofit and demolition have been ignored due to compared negligible contributions.

The Inventory of Carbon & Energy (ICE) (Hammond & Jones, 2011) is a well-established and widely used impacts database for materials covering ‘cradle’ (A1) to factory ‘gate’ (A3), as well as some materials’ ‘recycling’ (D). This database is suitable for assessing related impacts in this study because it is an especially established and most recent assessment database for building materials based on plentiful building cases (Dixit, Fernández-Solís, Lavy, & Culp, 2010). Data on energy and water consumption data during operation are simulated by BPS and converted into emission equivalents based on local energy and water production modes. Emissions are normalized into the relevant environmental impact categories (i.e. LCIA), by converting embodied, operational and recycling environmental impacts into global warming potential (GWP, kg-CO_2-eq.) in this study.

3.3. Environmental uncertainty modelling

Based on the design samples generated by the first module and corresponding environmental impacts computed by the second module, an integrated FCM-ELM algorithm is developed to quantify and model the environmental uncertainty associated with early design choices. This method enables the evaluation of environmental uncertainty arising from early design decisions using the average and variations on environmental impacts. A fuzzy clustering and learning integrated method for uncertainty quantification was firstly investigated by Shrestha and Solomatine (Shrestha & Solomatine, 2006). And it provided better results than other statistical methods for scenarios with high levels of uncertainty.

The fuzzy clustering in this integrated method partitions the set of training samples into clusters based on early design features. Then, prediction intervals for each cluster are calculated using the samples’ residual errors between a baseline model and observed outputs. Based on the grouped clusters, the prediction intervals of each training sample are quantified using the prediction intervals of clusters and the membership grade of the sample to the cluster. Finally, a learning model in this integrated method produces mappings between prediction intervals and early design decisions. The optimal model settings, including numbers of clusters, neurons in the baseline model, and neurons in the uncertainty mapping model are determined by a multi-objective optimization algorithm. The reasonable sizes of the training, validation and testing sets are determined by a trial-and-error method. Related information is presented in Supplementary Material 1 in Appendix B .

3.3.1. Sample clustering based on fuzzy C-means clustering

Fuzzy C-means clustering (FCM) was initially proposed and developed by Dunn (Dunn, 1973) and Bezdek (Bezdek, 1981), and it has been

Table 3
Early design parameters for the case building.

No.	Parameter	Unit	Options	Reference
1	Location		Harbin; Chengdu	[a]
2	Floor plan		Floor plan 1–4 (see Fig. 3)	Wang, Li, and Zheng (1994)
3	Orientation	Angle to north	0–360	[a]
4	Building length	m	Based on floor plan and room size	[a]
5	Building width	m	≥ 4.8	(China, 2011)
6	Building height	m	≤ 24	[a]
7	Floor area	m^2	Building width × building length	
8	Office room depth	m	5.4–6	Wang et al. (1994)
9	Meeting room depth	m	5.4–6	Wang et al. (1994)
10	Office room width	m	3–3.6	Wang et al. (1994)
11	Meeting room width	m	3.6–10.8 (floor plan 1, 2 and 4); 7.2–10.8 (floor plan 3)	Wang et al. (1994)
12	Corridor width	m	1.6–2.2 (floor plan 1 and 4); 1.3–2.2 (floor plan 2); 7.2–10.8 (floor plan 3)	Wang et al. (1994)
13	Floor height	m	3.6–4.2	China (2006)
14	Floor number		1–6	
15	Office room number	/floor	4–20 (floor plan 1 and 3); 2–10 (floor plan 2); 8–20 (floor plan 4)	[a]
16	Meeting room number	/floor	0–8 (floor plan 1); 0–4 (floor plan 2); 4–6 (floor plan 3); 6–8 (floor plan 4)	[a]

[a] Choices for these parameters based on interviews.

Table 4
Relevant unspecified detailed design parameters for case building.

No.	Parameter	Unit	Range	Reference
1	Window height	m	≤ Floor height – sill height – beam lower base	[a]
2	Window width	m	1.5–2.1	[a]
3	Beam size (b × h) and steel ratio	m, t/m^3	b = (1/3–1/2) × h, h = (1/14–1/8) × beam span, steel ratio = 0.1120	Academy of Building Research China (2010), HDHUR (2010)
4	External wall thermal insulation system		Colloidal polyphenyl granule insulation paste (CPG); rock wool (RW); (see Fig. 4)	MHURD China (2014)
5	External wall thermal insulation material thickness	mm	Based on climate zone Harbin: CPG, 20–35; RW, 40 Chengdu: CPG, 20; RW, 30	IASDC (2002)
6	Roof thermal insulation system		RW; expanded pearlite thermal insulating mortar (EPT); (see Fig. 4)	MHURD China (2014)
7	Roof thermal insulation material thickness	mm	Based on climate and building sharp factor (SF)[b] Harbin: RW, 120–300 when SF ≤ 0.3; 180–300 when 0.3 < SF < 0.4; EPT: 200–520 when SF < = 0.3; 270–520 when 0.3 < SF < 0.4; Chengdu: RW, 70–300; EPT, 120–300	IASDC (2012)
8	HVAC cooling set point	°C	24–27	China (2006)
9	HVAC heating set point	°C	18–24	China (2006)
10	Living water supply	m^3/h per person	0.0033–0.0056	China (2006)
11	Drinking water supply	m^3/h per person	0.0001–0.0002	China (2006)

[a] Choices for these parameters based on interviews.
[b] The building shape factor is an indicator of a buildings' sharp and heat radiating property, obtained by dividing building's external surface by its internal volume.

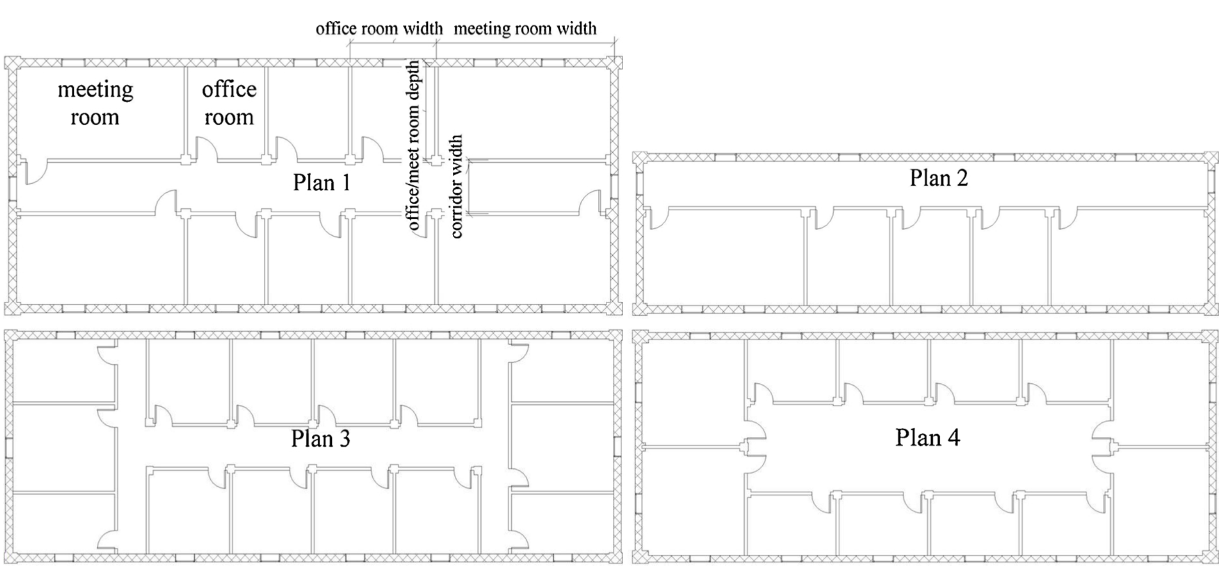

Fig. 3. Four floor plans for the case building, with rooms on two sides (Plan 1), one side (Plan 2), all four sides (Plan 3), and a large public area (Plan 4).

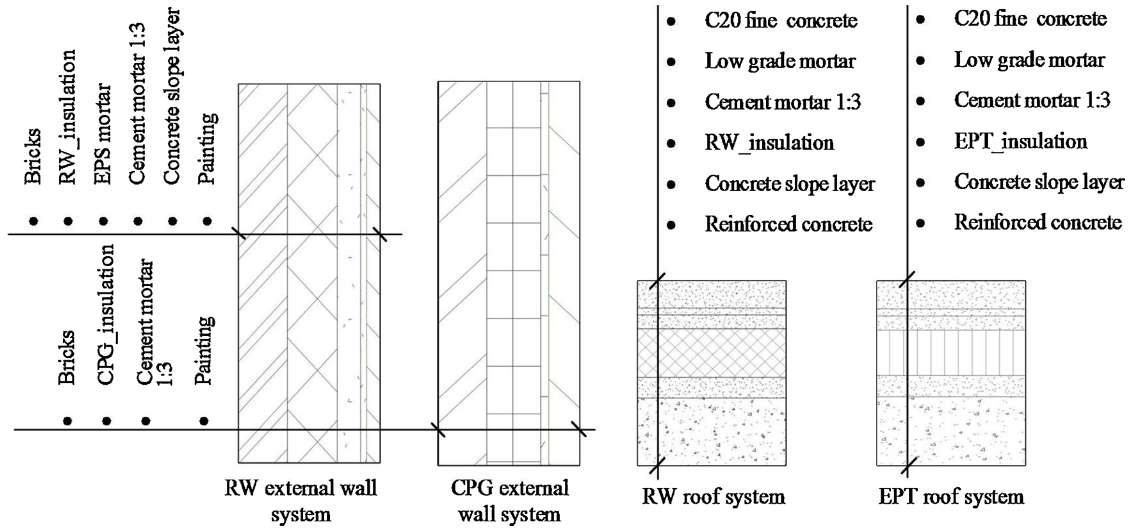

Fig. 4. External insulation system for the case building.

Table 5
Assessment data sources.

Impact sources	Unit	GWP (kg-CO_2-eq.)	Reference
Concrete-28/35 MPa	kg	0.12	Hammond and Jones (2011)
Steel rebar	kg	1.86	Hammond and Jones (2011)
Mortar (1:4)	kg	0.182	Hammond and Jones (2011)
Common brick	kg	0.24	Hammond and Jones (2011)
RW	kg	1.12	Hammond and Jones (2011)
CPG	kg	4.39	Hammond and Jones (2011)
EPT	kg	0.52 × 1.06[a]	Hammond and Jones (2011)
Glass	kg	0.91	Hammond and Jones (2011)
Sawn hardwood	kg	0.87	Hammond and Jones (2011)
Gypsum wall board	kg	0.39	Hammond and Jones (2011)
Metal stud	kg	1.46	Hammond and Jones (2011)
Paint	kg	2.91	Hammond and Jones (2011)
Electricity	kWh	1 × 1.063 = 1.063[b]	Yang (2002), Zhou and Su (2009)
Gas	m^3	1 × 0.021 + 25 × 0.00018 = 0.0255[b]	Yang (2002), IPCC (2009)
Water	m^3	1 × 0.213 = 0.213[b]	Yang (2003), IPCC (2009)

[a] Value calculated according to ICE's empirical suggestion due to lack of CO_2 equivalents.
[b] Value calculated based on emission data from Chinese energy and water production modes, and characterization factors from IPCC 2007.

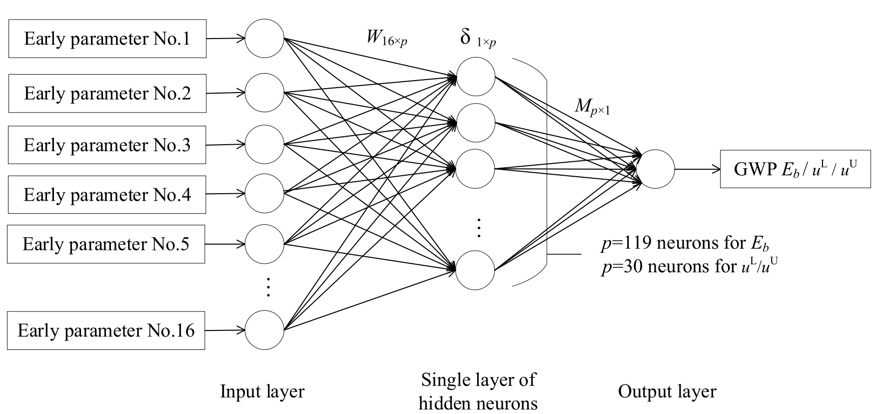

Fig. 5. The structure of ELM models for baseline regression and uncertainty mapping.

widely applied for pattern recognition. In Shrestha and Solomatine (Shrestha & Solomatine, 2006) research, FCM is used to fuzzily cluster samples and applied in conjunction with regression method to quantify uncertainty. This method set an assumption that similar samples have similar uncertainty. Here, design samples in the training set are clustered using FCM. The difference of early design parameters is set as cluster features. The FCM cost function (J_m) is defined as (1), and parameters are iteratively updated according to constraints (2) and (3) to minimize the cost function.

$$J_m = \min\left(\sum_{i=1}^{N}\sum_{c=1}^{C}(\mu_{i,c}\|x_i - m_c\|^2)\right) \tag{1}$$

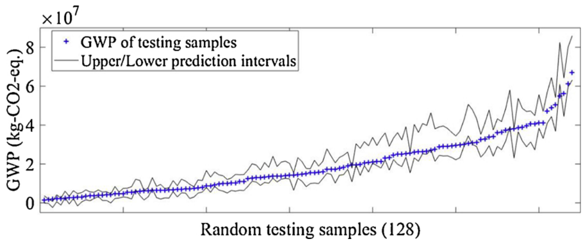

Fig. 6. Environmental impact intervals for testing samples.

$$\text{subject to } \mu_{i,c} = \frac{1}{\sum_{k=1}^{C} (||x_i - m_c||/||x_i - m_k||)^2} \tag{2}$$

$$m_c = \frac{\sum_{i=1}^{N} \mu_{i,c} x_i}{\sum_{i=1}^{N} \mu_{i,c}} \tag{3}$$

In the above equations, $||*||$ represents the Euclidean norm, x_i represents the early parameters of the ith design sample, $\mu_{i,c}$ represents the membership grade of the ith sample to the cth cluster, and is defined by (2); and m_c represents the centre of the cth cluster, defined by (3).

FCM is an iterative procedure, which terminates once a pre-determined threshold is attained or a pre-defined maximum number of iterations have been performed. Both $\mu_{i,c}$ and m_c are updated at each iteration, as determined by Eqs. (2) and (3). If the difference between the new cost function J_m of different ($\mu_{i,c}$, m_c) and the previous value is less than the defined threshold (1E-5 here), the algorithm terminates. Experiments have shown that standard FCM will generally converge to a solution in fewer than 130 iterations (Stetco, Zeng, & Keane, 2015), so we chose 130 as the maximum number of iterations.

3.3.2. Uncertainty calculation and modelling using an extreme learning machine

The extreme learning machine (ELM) (Huang, Zhu, & Siew, 2006) chooses hidden nodes at random without the need for time-consuming iterative tuning, unlike typical neural network algorithms. Its main advantage is the fast learning speed with satisfactory learning ability, as verified in many fields, such as traffic sign and human face recognition (Huang, Yu, Gu, & Liu, 2017; Mohammed, Minhas, Wu, & Sid-Ahmed, 2011). Thus, it seems highly suitable for efficient determination of relationships between design features and environmental performance in our proposed uncertainty quantification method. First a baseline environmental prediction model (E_b) is established by an ELM, which calculates prediction intervals for each cluster via Eqs. (4) and (5). The activation function in ELM is Sigmoid function.

$$U_c^L = e_i \tag{4}$$

where i is the largest value subject to $\sum_{k=1}^{i} \mu_{k,c} < \alpha/2 \sum \mu_c$.

$$U_c^U = e_j \tag{5}$$

where j is the smallest value subject to $\sum_{k=1}^{j} \mu_{k,c} > (1 - \alpha/2) \sum \mu_c$.

U_c^L and U_c^U respectively represent the lower and upper endpoints of the prediction interval of the cth cluster. These endpoints are defined using the residual error (e) of the ith and jth training samples, the maximum and minimum samples satisfying the inequalities in (4) and (5), respectively, when all samples are sorted in descending order by residual error. The error of each of the selected i and j training samples is calculated by comparing the difference between the baseline model E_b and the observed performance. As before, $\mu_{k,c}$ represents the membership grade of the kth sample to the cth cluster. $\Sigma(\mu_c)$ represents the sum, over all samples in the cth cluster, of the membership grade of the sample to the cluster. α is the confidence coefficient and is defined to be 95% in this study.

Then the environmental interval of each training sample is calculated using (6), based on the membership grade of the sample to each cluster and the cluster's prediction interval (U). We write u_k^L and u_k^U to denote the lower and upper endpoints of the uncertainty interval for the kth sample, where:

$$u_k^L = \sum_{c=1}^{C} (\mu_{k,c} U_c^L); \quad u_k^U = \sum_{c=1}^{C} (\mu_{k,c} U_c^U) \tag{6}$$

The second use of ELM, based on the uncertainty intervals of the samples, is to construct mappings from uncertainty intervals to early design scenarios x (f: $x \to u^L$ and f: $x \to u^U$). In this way, for any new early design scenarios, the mappings are used to infer environmental prediction intervals (with endpoints E^L and E^U) by summing the baseline prediction E_b with endpoints u^L and u^U as (7), where:

$$E^L = E_b + u^L; \quad E^U = E_b + u^U \tag{7}$$

4. Case study and platform development

4.1. Individual-room office building

The proposed environmental uncertainty quantification method is applied here to assess environmental uncertainty of early design for a low-rise individual-room office building, to be located in China. A series of face-to-face interviews were conducted with architects and engineers to collect design information for the case building and identify restrictions on design parameters, which should meet national design standards and the building's operational requirements. Members of the

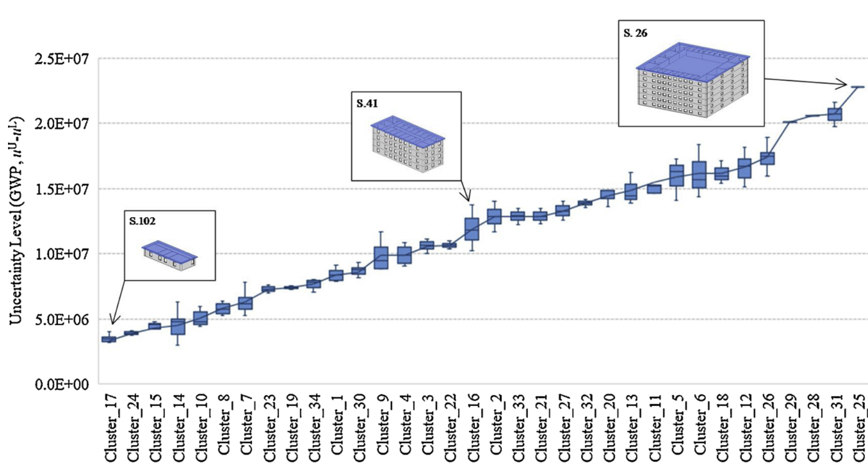

Fig. 7. FCM clustering for testing samples and exemplary designs.

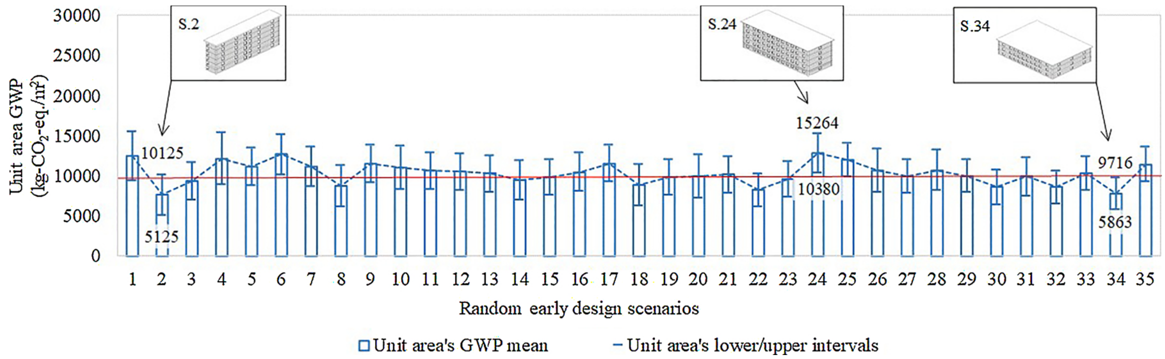

Fig. 8. Functional units' GWP means and prediction intervals.

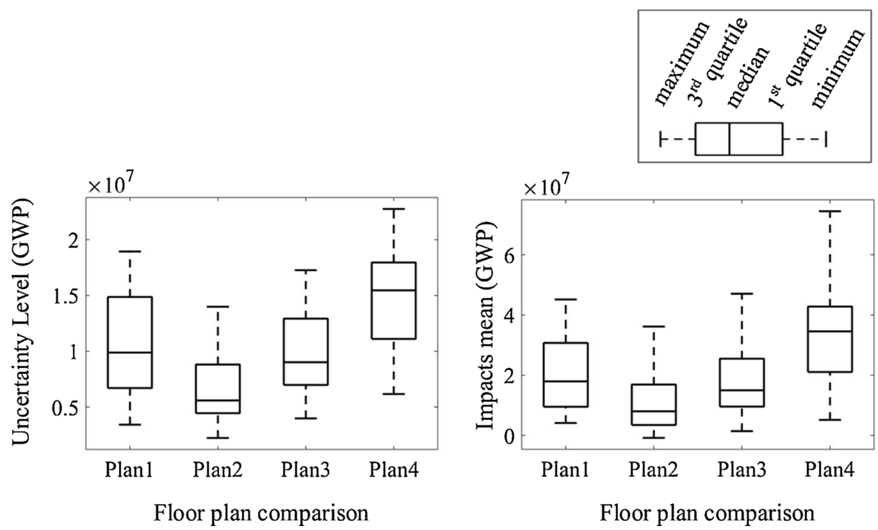

Fig. 9. GWP mean values and prediction intervals for each of the floor plans.

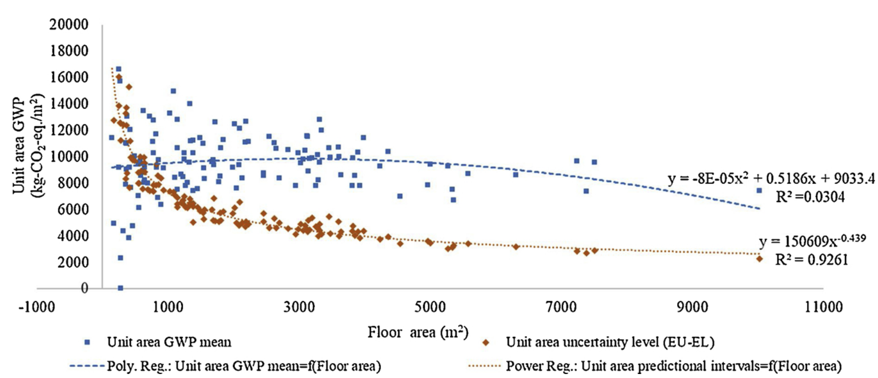

Fig. 10. Effect of floor area on mean value and uncertainty of GWP per unit area.

building design company's staff (including architects, as well as both structural and electrical engineers) participated in the interviews to provide enterprise-related knowledge regarding the case study building's design. The interviews were structured, using the questionnaire presented in Supplementary Material 2 in Appendix B.

Early design parameters are shown in Table 3. In the conceptual and schematic design stages, designers consider the early design parameters, and the detailed design parameters are unknown. However, likely ranges of values can be obtained from national building design guidelines, legal thresholds, and designers' expectations. Ranges of relevant detailed design parameters that may vary later are listed in Table 4.

Based on the interviews, four forms of floor plan were considered for the office building (Fig. 3). Generally, the preferred orientation for a new building in most Chinese cities is south-facing, to optimize effects of sunlight and prevailing winds. However, for an office building, other orientations may be considered depending on various factors at specific sites. Although for a normal design project, the building location is already determined. To assess effects of climatic differences on environmental impact assessments, we considered effects of designed buildings in Harbin (in a severely cold zone) and Chengdu (where summers are hot and winters are cold).

This section discusses possible values of unspecified parameters for the case study building. National norms in China for a public building stipulate that insulation materials must be of the 'hard-combustible' type. Colloidal polyphenyl granule insulation paste (CPG), rock wool

Table 6
Early design parameter rankings in importance for GWP uncertainty.

Ranking	Parameter	*SSD*
1	Building height	247.86
2	Floor area	231.18
3	Office room number	231.05
4	Office room width	229.55
5	Meeting room depth	228.84
6	Location	228.28
7	Building width	227.96
8	Building length	227.60
9	Corridor width	227.13
10	Elevation height	227.02
11	Floor number	226.97
12	Office room depth	226.85
13	Meeting room number	226.67
14	Orientation	226.24
15	Floor plan	199.96
16	Meeting room width	4.49

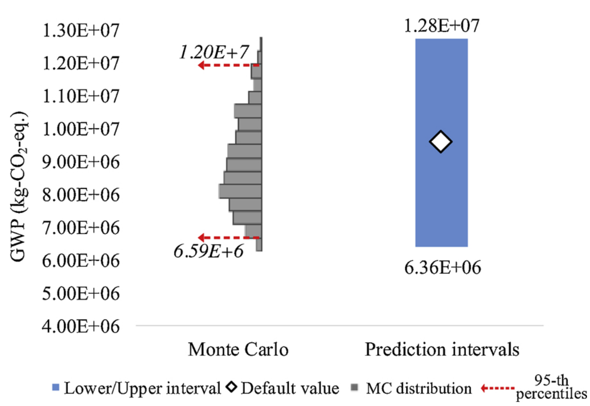

Fig. 11. Early design prediction interval, default setting and Monte Carlo comparison.

(RW), and expanded pearlite thermal insulating mortar (EPT) are recommended insulation materials for external walls and roof systems (Fig. 4). The required minimum thickness of the insulation depends on the climate associated with the building's location and its shape factor (based on superficial area and volume). Office building design standards (China, 2006) and energy-saving design standards for public buildings (Academy of Building Research China, 2015) in China provide guidelines covering the operation of office buildings. We employed empirical ranges from design guidelines (Academy of Building Research China, 2010; HDHUR, 2010) and mechanical calculations for the structure of the building from the interviewed of structural engineers. The guide values in the Construction Engineering Quota of China (2010) (HDHUR, 2010) were used to calculate ranges for the sizes and quantities of reinforcement in steel beams, columns and slabs.

4.2. Implementation platform and model establishment

The parametric design technology is developed using Autodesk® Dynamo (Fig. A1 and Table A1) and Revit. Some of the samples generated are shown in Fig. A2. The ICE database (version 2.0) (Hammond & Jones, 2011) is used to perform the embodied and recycling impacts assessment, and EnergyPlus™ to simulate the building's operation (defined as a period of 50 years). The parametric design software extracts building information from the samples automatically as MS Excel files for the ICE database, and EnergyPlus Input Files (IDF) for EnergyPlus. We programmed the FCM-ELM integrated algorithm in Matlab®. The final impacts are characterized using the GWP-100-year metric, using characterization factors in the IPCC database (IPCC, 2009). Other environmental indictors can also be included if necessary. The assessment data sources are shown in Table 5.

The sizes of sample sets for training the uncertainty quantification model and validation (for optimization of the predictive model settings) are set at 1024 and 128 samples, respectively. The optimal model settings are determined by a particle swarm optimization (PSO)-based multi-objective optimization, resulting in 34 clusters, 119 neurons for baseline regression, and 30 neurons for uncertainty mapping. Fig. 5 shows the structure of trained ELM models for baseline regression (E_b) and uncertainty mapping (u^L/u^U). Detailed information about optimal model settings is provided in Supplementary Material 1 in Appendix B.

The proposed method is assessed using a randomly generated testing set of 128 samples. The results show that the model produces prediction intervals for each sample's environmental impact are within ranges (Fig. 6), thus validating the model's ability to provide rational prediction intervals for environmental effects of early design decisions.

FCM is used in the method to generate clusters for prediction intervals calculation, as shown for the 128 testing samples (with exemplary designs) in Fig. 7. Uncertainty levels are calculated by subtracting lower from upper intervals. The results show that designs featuring a single floor with rooms along one side (as in Sample 102) are clustered into Cluster 17. These designs are calculated to have low uncertainty. Multi-floor arrangements, with rooms along two sides (as in Sample 41) are clustered into Cluster 16, which has intermediate uncertainty, while multi-floor arrangements with a large public area (as in Sample 26) are clustered into Cluster 25, which has high uncertainty. Thus, the results clearly show that different early designs will be clustered into different clusters, with substantially differing levels of uncertainty.

5. Results and discussion

5.1. Results of early design application

5.1.1. Wide scenarios comparison

The proposed method was used to assess environmental impacts and uncertainty of diverse early design scenarios for the case building. The results show that after sample generation and model training, it can efficiently yield uncertainty results for numerous design scenarios because the machine learning model is a nearly real-time feedback model. A set of random early design scenarios (Table A2) were used to validate the trained quantification model, and both means and ranges of their impacts were computed (see Fig. 8). To enable comparable, the environmental performance parameters were converted into GWP per unit area. The results are valuable when the designer do not yet have specific outlines in the early building design, thereby facilitating selection of designs that are likely to provide the best performance from wide arrays of possible scenarios.

The results shown in Fig. 8 indicate that Sample 2 has the smallest GWP/m^2 lower endpoint (and thus may provide very low impacts), while Sample 34 has the smallest GWP/m^2 upper endpoint (and thus may have relatively low impact even in extreme conditions). Thus Sample 2 could be a good early design for aggressive attempts to

Table 7
Default values for the comparative study.

Default parameter	External wall: typical mild climate insulation	Window size	Roof: typical insulation	Cooling set point	Heating set point	Living/drinking water
Unit	R = (m^2 × K)/W	m	R = (m^2 × K)/W	°C	°C	m^3/h per person
Default value	1.69 (Autodesk, 2018a, 2018b)	1.5 × 1.8	3.72 (Autodesk, 2018a, 2018b)	27.0 (China, 2006)	20.0 (China, 2006)	0.004 (China, 2006)

minimize impacts, while Sample 34 is a good choice from a conservative perspective. Clearly, if even the lower endpoint is higher than the target performance, according to the modelling, there is a 95% possibility of failure. For example, if the designers' target for GWP/m^2 is 10,000, we can immediately reject Sample 24. These results suggest that the proposed approach can provide valuable information for designers considering both specific targets and the possibilities to achieve them.

5.1.2. Analysis of specific parameters

The proposed method also supports analysis of effects of specific early design choices, which is valuable for designers involved in specific early parameter selection. In the case study, four forms of floor plan were considered. Average GWP impact and uncertainty levels (maximum, minimum, median, first and third quartile values) obtained for designs with each of the floor plans in the 128 testing samples are shown in Fig. 9. The results suggest that floor plan 2 is the optimal choice in terms of both average GWP and prediction intervals. In addition, the proposed method supports analysis of patterns of uncertainty in environmental impacts associated with variations in parameters, such as floor area. For example, as shown in Fig. 10, the magnitude of uncertainties in environmental impact decreases as floor area increases. More specifically, GWP per unit floor area starts to decrease beyond a threshold of ca. 3200 m^2, but the trend for smaller floor areas is less clear cut. The results show that for the case building, increasing the floor area leads to both reductions in impact per unit area and uncertainty in this parameter when the floor area exceeds 3200 m^2.

5.1.3. Parameter importance analysis by the partial derivatives method

The established uncertainty model can also help designers to assess the importance of early design parameters in terms of future performance uncertainty. Partial derivatives of the ELM model, as a type of neural network, expressed as (8) and (9) (Dimopoulos, Bourret, & Lek, 1995), indicate the sensitivity of the learning model's outputs to input perturbations. Thus, in this context the uncertainty of a building's performance is the sensitivity indicator, and variations in early design parameters are the input perturbations.

$$d_{ji} = s_j \sum_{h=1}^{nh} w_{ho} I_{hj} (1 - I_{hj}) w_{ih} \tag{8}$$

$$SSD_i = \sum_{j=1}^{N} (d_{ji})^2 \tag{9}$$

Here, N is 128 as the total number of samples used for sensitivity analysis, s_j is the derivative of the response of the output node to the jth sample input, w_{ho} is the weight of the hth hidden neuron to the output neuron; w_{ih} is the weight of the ith input neuron (parameter) to hidden neurons; I_{hj} is the response of the hth hidden neuron to the jth sample. SSD_i is the sum of square derivatives due to the ith parameter. Both the input and output are normalized to $(-1, +1)$. Results of using the 128 testing samples to calculate SSD show that building height, floor area and office room number have the strongest influences on the case building's overall GWP uncertainty, and meeting room width has negligible influence (Table 6). Therefore, the model can help designers identify the most influential decisions in early design stages.

5.2. Results of method comparison

The benefits of the proposed method are also substantiated by comparison with previously published methods, using a randomly generated early design (details in Table A3) as an example. Our uncertainty quantification model indicates that the GWP for this design will lie between 6.36×10^6 and 1.28×10^7 kg-CO_2-eq. (Fig. 11). As already mentioned, in various prior studies default values have been used for undecided design parameters, making the probabilistic problem deterministic. Thus, for comparison, we have chosen some default values (shown in Table 7) that are suitable for the case building for the same early design, based on ASHRAE 90.1-2010 and the CBECS database, from (Autodesk, 2018a, 2018b). The resulting GWP is 9.62×10^6 kg-CO_2-eq., as shown in Fig. 11. Clearly, this single value, arising from a deterministic calculation, cannot fully describe the potential environmental impact of an early design.

In addition, for the same early design, we compared results of our method with results based on Monte Carlo analysis of 500 samples, assuming that the unspecified parameters follow a uniform distribution pattern. The percentiles of performance distribution are observational data, while the prediction intervals are upper/lower limits for future samples in percentages (Schwarz, 2011). Therefore, reliable prediction intervals are theoretically wider than percentiles. The distribution of results (GWP) obtained using the approach based on Monte Carlo analysis is shown in Fig. 11, as well as the values within which 95% of the results lie (95th percentiles). The upper and lower intervals are 6.67% and 3.49% wider than the Monte Carlo-based 95th percentiles, respectively.

The results indicate that our method is at least as reliable and accurate as Monte Carlo methodology, although distributions obtained from the latter theoretically provide more information (if reliable) than prediction intervals. The proposed method overcomes two limitations associated with approaches based on Monte Carlo analysis. First, the learning model we proposed calculates prediction intervals based on historical residual error between the output of the model and observed samples, and then infers relationships between prediction intervals and early design decisions. Thus, the directly established uncertainty relationships are independent of the probability distributions of the unspecified parameters. Second, our method only requires satisfactory numbers of samples for model training, validation, and testing. It does not require large numbers of samples for evaluating every scenario. In contrast, Monte Carlo analysis typically requires numerous samples to obtain outcome distributions for every scenario, and thus results in high computing loads (Muthén & Muthén, 2002). This may not be efficient for early design decisions, for which many design scenarios are possible and need to be compared. Therefore, without probability distributions and requiring low sample size, our proposed method for computing prediction intervals is a more generally applicable method for evaluating early design decisions.

6. Conclusions and future work

Early design decisions have significant impacts on buildings' life-cycle environmental performance (Kohler & Moffatt, 2003). Existing methods based on default values for assessing the environmental impact of early designs may lead to unreliable results and decisions, while approaches based on Monte Carlo analysis may be unreliable because of the difficulty of determining the probability distributions of unspecified parameters. We propose a sampling and data-learning method for assessing environmental performance in early design stages, without using default values or requiring knowledge of design parameters' probability distributions. The method provides an alternative solution by enabling prediction of environmental performance intervals in early design stages. The application of the developed method in the case study demonstrated that it can provide equally reliable results and is more applicable than methods involving use of default settings or Monte Carlo analysis.

In addition to the method's utility in early design applications, the overall framework introduces an innovative way to combine and synergistically exploit advantages of both parametric design technology and machine learning algorithms. Parametric design technology can widely explore design constraints and thus create uncertainty-related knowledge by bootstrapping, and machine learning can extract associations between early design features and environmental uncertainty, then build feedback models to support further design decisions.

To apply the proposed method in early building design, relevant early design parameters and detailed design constraints need to be predefined in the parametric design software, and a local environmental impact database need to be collected for environmental assessment. Then, the uncertainty modelling using FCM-ELM is run, and the environmental feedback model is built. The proposed method has three identified advantages. First, it enables designers to explore and compare numerous potential early designs in early design stages. Second, the results not only indicate effects of early design decisions on environmental performance, but also variations in performance associated with early design parameters (Figs. 9 and 10). Third, sensitivity analysis of the developed uncertainty model enables analysis of the relative importance of early parameters for performance uncertainty (Table 6). Knowledge of this uncertainty does not make building design decisions straightforward, but it is valuable for designers to know variations in performance associated with specific design options, especially in early design stages when uncertainty levels are inherently high. Thus, as the proposed method can provide designers with uncertainty information associated with early parameters, it can help them make informed early design decisions.

Applications of the proposed method could be extended, for example, to efficiently evaluate numerous buildings in blocks or even towns, for which only basic building information is easily available but obtaining full information for every building is challenging. The method only needs the possibility of unspecified parameters, then the parametric design can create samples and machine learning can model the uncertainty. Another promising line of future work is to develop sophisticated approaches to make design decisions and optimization based on quantified uncertainty by the proposed method. In addition, as available information increases, the quantified intervals and average values derived using our method should approach real values. We intend to explore this possibility in future work.

The proposed method is promising, but some improvements could still be made. It is especially useful for early building design and complements Monte Carlo for undecided parameters uncertainty. Thus it narrows sources of uncertainty. Other uncertainty sources such as from scenario uncertainty (e.g. assessment data source) and from model uncertainty (e.g. building simulation programme) can be comprehensively considered if the proposed method is integrated with other methods, e.g. proposed bootstrapping procedures (Huijbregts et al., 2003) for the two mentioned additional sources of uncertainties.

Acknowledgments

This research was supported by the National Key Research and Development Programme of China (grant no. 2016YFC0701904) and the Swedish Research Council for Environment, Agricultural Sciences and Spatial Planning (Formas).

Appendix A

Figs. A1 and A2
Tables A1–A3

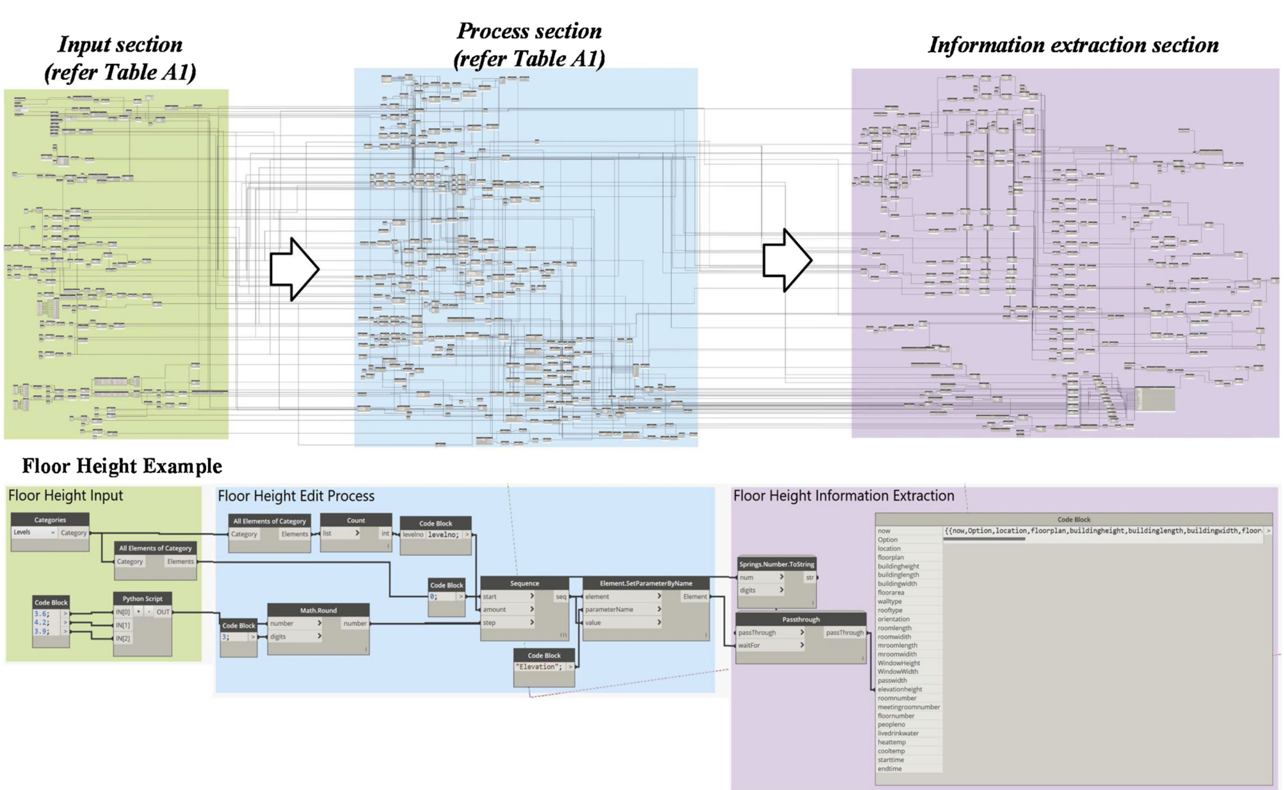

Fig. A1. Functional sections in the parametric design technology.

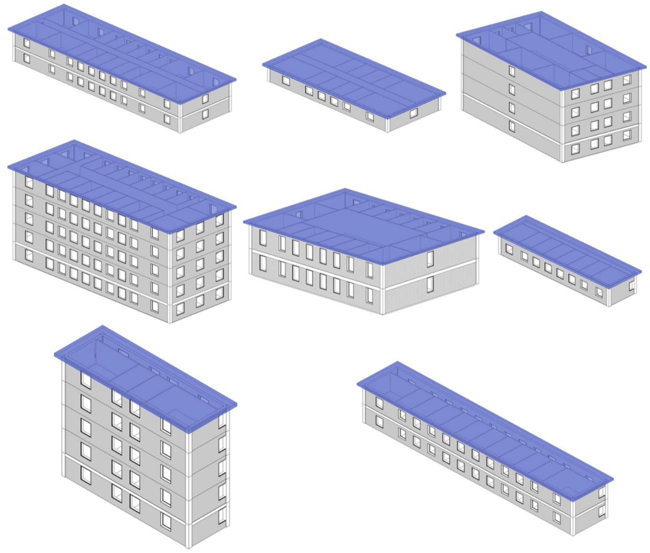

Fig. A2. Some full information samples (FIS) generated by parametric design technology.

Table A1
Design parameters in parametric design with ranges and dependencies.

Parameters in parametric design (Input section)	Range and dependency (achieved by process section)
Location	Harbin or Chengdu
Floor plan	Floor plan 1–4
Orientation	0–360°
Building length	Based on floor plan and room size
Building width	Based on floor plan and room size && ≥4.8 m
Building height	Based on floor height and floor number && ≤24 m
Floor area	Building width × building length
Office room depth	5.4–6 m
Meeting room depth	5.4–6 m
Office room width	3–3.6 m
Meeting room width	3.6–10.8 m (floor plan 1, 2 and 4); 7.2–10.8 m (floor plan 3)
Corridor width	1.6–2.2 m (floor plan 1 and 4); 1.3–2.2 m (floor plan 2); 7.2–10.8 m (floor plan 3)
Floor height	3.6–4.2 m
Floor number	1–6
Office room number	4–20 (floor plan 1 and 3); 2–10 (floor plan 2); 8–20 (floor plan 4)
Meeting room number	0–8 (floor plan 1); 0–4 (floor plan 2); 4–6 (floor plan 3); 6–8 (floor plan 4)
Window height	≤Floor height – sill height (1 m for public building) – beam lower base
Window width	1.5–2.1 m
Beam size and steel ratio	b = (1/3–1/2) × h, h = (1/14–1/8) × beam span, steel ratio = 0.1120
External wall thermal insulation system	CPG or RW
External wall thermal insulation material thickness	Harbin, CPG, 20–35 mm/RW, 40 mm; Chengdu, CPG, 20 mm/RW, 30 mm
Roof thermal insulation system	RW or EPT
Roof thermal insulation material thickness	Harbin, RW, 120–300 mm when SF ≤ 0.3, 180–300 mm when 0.3 < SF < 0.4; EPT, 200–520 mm when SF ≤ 0.3, 270–520 mm when 0.3 < SF < 0.4; Chengdu, RW, 70–300 mm/EPT, 120–300 mm
HVAC cooling set point	24–27 °C
HVAC heating set point	18–24 °C
Living water supply	0.0033–0.0056 m^3/h per person
Drinking water supply	0.0001–0.0002 m^3/h per person

Table A2
Set of random early design scenarios using parametric design technology for early design comparison.

No.	Early parameters															
	Loca.	Floor plan	Orientation	Building length (m)	Building width (m)	Building height (m)	Floor area (m^2)	Office room width (m)	Meeting room width (m)	Office room depth (m)	Meeting room depth (m)	Pass_width (m)	Floor_height (m)	Floor no.	Room no.	Meeting room no.
1	CD	2	214	45.256	7.441	23.790	2020.6497	3.379	4.556	5.761	5.761	0.84	3.965	6	48	24
2	CD	2	185	47.474	7.187	23.112	2047.1758	3.484	10.127	5.787	5.787	0.7	3.852	6	12	24
3	CD	4	105	37.137	18.706	11.226	2084.024	3.536	6.233	5.613	11.226	1.04	3.742	3	36	18
4	HRB	4	212	36.568	14.361	16.228	2100.6378	3.585	4.789	5.981	16.228	1.05	4.057	4	48	24
5	CD	3	267	36.233	20.133	11.661	2188.4069	3.031	5.993	5.896	10.065	4.17	3.887	3	48	12
6	HRB	2	138	50.650	7.222	23.460	2194.7027	3.456	8.047	5.602	5.602	0.81	3.91	6	60	12
7	HRB	2	246	73.588	7.584	15.652	2232.4952	3.541	9.545	5.444	5.444	1.07	3.913	4	40	16
8	CD	1	66	44.337	13.882	15.588	2461.942	3.147	5.931	5.931	7.937	1.01	3.897	4	32	32
9	HRB	1	197	38.281	13.324	20.740	2550.392	3.234	5.592	5.592	6.206	1.07	4.148	5	60	20
10	HRB	2	269	60.302	7.259	23.844	2626.267	3.587	6.108	5.919	5.919	0.67	3.974	6	60	24
11	HRB	1	50	48.426	13.717	15.924	2657.015	3.041	5.778	5.778	4.505	1.08	3.981	4	56	32
12	HRB	3	56	43.848	21.556	12.177	2835.532	3.219	5.831	5.758	10.773	5.02	4.059	3	60	12
13	CD	3	107	45.048	22.110	11.319	2988.035	3.361	5.718	5.465	7.371	5.59	3.773	3	60	18
14	CD	1	256	57.712	13.116	15.588	3027.760	3.551	5.728	5.728	7.326	0.83	3.897	4	48	32
15	HRB	3	253	42.334	24.165	11.337	3069.016	3.041	5.961	5.562	8.053	6.52	3.779	3	60	18
16	CD	1	250	38.558	13.372	23.652	3093.596	3.232	5.856	5.856	9.584	0.83	3.942	6	60	24
17	CD	3	165	45.199	17.310	14.476	3129.502	3.323	5.985	5.845	8.653	2.81	3.619	4	80	16
18	CD	1	275	40.739	12.855	23.508	3142.213	3.522	5.528	5.528	6.663	0.9	3.918	6	48	48
19	CD	3	254	29.905	21.185	19.600	3167.599	3.157	5.481	5.502	10.591	5.09	3.92	5	60	20
20	CD	1	207	42.005	12.717	23.694	3205.001	3.599	5.538	5.538	5.103	0.82	3.949	6	60	48
21	CD	1	159	50.899	12.937	19.055	3292.316	3.265	5.428	5.428	9.123	1.04	3.811	5	70	20
22	CD	4	199	34.257	32.058	11.334	3294.594	3.123	8.013	5.889	11.334	0.93	3.778	3	36	24
23	CD	1	130	48.177	13.755	19.945	3313.440	3.597	5.788	5.788	9.702	1.09	3.989	5	60	20
24	CD	3	228	42.931	15.456	19.045	3317.658	3.148	5.727	5.558	7.732	2.17	3.809	5	100	20
25	CD	4	280	47.572	17.575	16.276	3344.376	3.323	5.858	5.708	16.276	0.86	4.069	4	80	24
26	CD	1	129	44.804	12.859	23.532	3456.769	3.366	5.469	5.469	4.468	0.96	3.922	6	72	48
27	HRB	1	205	49.824	13.968	19.955	3479.769	3.484	5.924	5.924	7.494	1.06	3.991	5	70	20
28	HRB	4	173	34.996	17.173	23.334	3605.971	3.343	4.294	5.717	23.334	0.8	3.889	6	72	48
29	CD	4	248	35.611	20.273	19.850	3609.674	3.356	6.758	5.576	19.85	0.87	3.97	5	60	30
30	HRB	4	262	35.811	25.432	16.444	3642.909	3.469	6.357	5.506	16.444	1.03	4.111	4	48	32
31	CD	4	283	37.747	16.894	21.948	3826.100	3.568	4.225	5.737	21.948	1.09	3.658	6	72	48
32	HRB	4	241	32.356	29.769	15.972	3852.841	3.049	9.923	5.544	15.972	0.810	3.993	4	48	24
33	CD	1	100	51.781	12.556	23.904	3901.093	3.211	5.448	5.448	9.835	0.83	3.984	6	84	24
34	HRB	4	183	43.056	30.400	12.156	3926.675	3.354	10.132	5.88	12.156	1.08	4.052	3	48	18
35	HRB	4	226	39.954	16.624	22.782	3985.059	3.074	5.539	5.852	22.782	1.06	3.797	6	96	36

Table A3
Ten of 128 testing samples and a randomly generated early design for comparative study generated using parametric design technology.

No.	Early parameters															
	Loca.	Floor plan	Orientation	Building length (m)	Building width (m)	Building height (m)	Floor area (m^2)	Office room width (m)	Meeting room width (m)	Office room depth (m)	Meeting room depth (m)	Pass_width (m)	Floor_height (m)	Floor no.	Room no.	Meeting room no.
1	CD	4	152	34.408	42.927	19.355	7385.050	3.191	10.73	5.643	19.355	1.010	3.871	5	60	40
2	CD	2	206	24.097	6.872	19.350	828.002	3.386	4.331	5.512	5.512	0.680	3.870	5	10	20
3	HRB	3	200	17.894	24.373	3.958	436.131	3.159	5.788	5.836	8.126	6.350	3.958	1	4	6
4	HRB	2	157	55.596	7.735	15.528	1720.071	3.587	9.861	5.535	5.535	1.100	3.882	4	40	8
5	HRB	4	62	36.015	25.825	24.438	5580.402	3.558	6.456	5.412	24.438	0.920	4.073	6	72	48
6	CD	3	193	38.926	19.915	7.692	1550.461	3.497	5.474	5.688	9.959	4.270	3.846	2	32	8
7	HRB	4	241	32.356	29.769	15.972	3852.841	3.049	9.923	5.544	15.972	0.810	3.993	4	48	24
8	HRB	1	228	54.295	13.386	23.094	4360.627	3.379	5.673	5.673	10.252	1.020	3.849	6	84	24
9	HRB	4	113	28.388	14.952	7.722	848.936	3.235	4.984	5.526	7.722	0.910	3.861	2	16	12
10	CD	2	144	37.699	7.606	15.540	1146.958	3.502	7.674	5.466	5.466	1.070	3.885	4	8	16
…																
[a]	HRB	4	273	45.876	17.830	3.952	817.980	3.418	5.849	5.465	8.917	3.450	3.952	1	20	4

No.	Unspecified detailed parameters and results										
	Wind. h.	Wind. w.	Beam b × h	External wall thermal insulation	External wall insulation thickness (mm)	Roof thermal system	Roof insulation thickness (mm)	HVAC cooling set point (°C)	HVAC heating set point (°C)	Living and drinking water supply (m^3/h × person)	GWP (kg-CO_2-eq.)
1	1.660	1.883	0.52 × 1.08	RW	20	EPT	520	25	19	0.0054	56130013
2	1.922	1.709	0.18 × 0.43	CPG	30	RW	240	27	24	0.0050	5993429
3	1.982	1.843	0.28 × 0.60	RW	30	RW	240	24	20	0.0055	3882687
4	2.306	1.553	0.39 × 1.01	CPG	40	RW	300	26	19	0.0042	19054880
5	2.294	1.637	0.24 × 0.49	CPG	25	EPT	390	24	18	0.0047	37948713
6	2.164	1.672	0.53 × 1.1	CPG	20	EPT	390	25	21	0.0040	13549901
7	2.464	1.533	0.59 × 1.22	CPG	35	RW	280	26	24	0.0049	26323535
8	2.219	2.047	0.45 × 1.05	CPG	25	EPT	520	24	21	0.0050	38581055
9	2.197	2.08	0.23 × 0.47	CPG	25	EPT	455	27	23	0.0052	7233996
10	2.199	1.679	0.30 × 0.78	CPG	20	EPT	270	25	18	0.0040	7054385
…											

[a] Randomly generated early design for comparative study.

Appendix B. Supplementary data

Supplementary material associated with this article can be found, in the online version, at https://doi.org/10.1016/j.scs.2019.101596.

References

Academy of Building Research China (2010). *Earthquake resistant design code-GB 50011-2010.* (in Chinese).
Academy of Building Research China (2015). *Energy-saving design standards for public buildings of China.* (in Chinese).
Alalouch, C., Al-Saadi, S., AlWaer, H., & Al-Khaled, K. (2019). Energy saving potential for residential buildings in hot climates: The case of Oman. *Sustainable Cities and Society,* 101442.
Asadi, S., Amiri, S. S., & Mottahedi, M. (2014). On the development of multi-linear regression analysis to assess energy consumption in the early stages of building design. *Energy and Buildings, 85*, 246–255.
Attia, S., Gratia, E., De Herde, A., & Hensen, J. L. (2012). Simulation-based decision support tool for early stages of zero-energy building design. *Energy and Buildings, 49*, 2–15.
Autodesk (2018a). *Building type data*.
Autodesk (2018b). *Material thermal properties – Conceptual types*.
Bezdek, J. C. (1981). *Objective function clustering. Pattern recognition with fuzzy objective function algorithms*. Springer43–93.
China, M. H. U. R. D. (2006). *Office building design standards of China-JGJ67.* (in Chinese).
China, M. H. U. R. D. (2011). *Technical specifications for concrete structures in high-rise buildings-JGJ 3-2010.* China building industry press (in Chinese).
Chong, Y. T., Chen, C.-H., & Leong, K. F. (2009). A heuristic-based approach to conceptual design. *Research in Engineering Design, 20*(2), 97–116.
De Wit, S., & Augenbroe, G. (2002). Analysis of uncertainty in building design evaluations and its implications. *Energy and Buildings, 34*(9), 951–958.
Dimopoulos, Y., Bourret, P., & Lek, S. (1995). Use of some sensitivity criteria for choosing networks with good generalization ability. *Neural Processing Letters, 2*(6), 1–4.
Ding, G. K. (2008). Sustainable construction—The role of environmental assessment tools. *Journal of Environmental Management, 86*(3), 451–464.
Dixit, M. K., Fernández-Solís, J. L., Lavy, S., & Culp, C. H. (2010). Identification of parameters for embodied energy measurement: A literature review. *Energy and Buildings, 42*(8), 1238–1247.
Dixit, M. K., Fernández-Solís, J. L., Lavy, S., & Culp, C. H. (2012). Need for an embodied energy measurement protocol for buildings: A review paper. *Renewable and Sustainable Energy Reviews, 16*(6), 3730–3743.
Dunn, J. C. (1973). *A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters*.
Eltaweel, A., & Yuehong, S. (2017). Parametric design and daylighting: A literature review. *Renewable and Sustainable Energy Reviews, 73*, 1086–1103.
Friedman, J., Hastie, T., & Tibshirani, R. (2001). *The elements of statistical learning.* Springer Series in Statistics New York.
Gervásio, H., Santos, P., Martins, R., & da Silva, L. S. (2014). A macro-component approach for the assessment of building sustainability in early stages of design. *Building and Environment, 73*, 256–270.
Hammond, G., & Jones, C. (2011). *Inventory of Carbon and Energy (ICE). Version 2.0.* Sustainable Energy Research Team, Dept, Of Mechanical Engineering, University of Bathhttp://www.bath. ac. uk/mech-eng/sert/embodied.
HDHUR (2010). *Heilongjiang province construction project engineering quota.* (in Chinese).
Hiyama, K., Kato, S., Kubota, M., & Zhang, J. (2014). A new method for reusing building information models of past projects to optimize the default configuration for performance simulations. *Energy and Buildings, 73*, 83–91.
Holzer, D. (2016). Design exploration supported by digital tool ecologies. *Automation in Construction, 72*, 3–8.
Huang, G.-B., Zhu, Q.-Y., & Siew, C.-K. (2006). Extreme learning machine: Theory and applications. *Neurocomputing, 70*(1–3), 489–501.
Huang, Z., Yu, Y., Gu, J., & Liu, H. (2017). An efficient method for traffic sign recognition based on extreme learning machine. *IEEE Transactions on Cybernetics, 47*(4), 920–933.
Huijbregts, M. A., Gilijamse, W., Ragas, A. M., & Reijnders, L. (2003). Evaluating uncertainty in environmental life-cycle assessment. A case study comparing two insulation options for a Dutch one-family dwelling. *Environmental Science & Technology, 37*(11), 2600–2608.
Hygh, J. S., DeCarolis, J. F., Hill, D. B., & Ranjithan, S. R. (2012). Multivariate regression as an energy assessment tool in early building design. *Building and Environment, 57*, 165–175.
Institute of architectural standards and design China (2002). *Exterior insulation building structure design guide-10J121.* (in Chinese).
Institute of architectural standards and design China (2012). *Flat roof structure guide-12J201.* (in Chinese).
IPCC (2009). *IPCC fifth assessment report, 2*.
ISO (2006). *14040: Environmental management – Life cycle assessment – Principles and framework*.
Jabi, W. (2013). *Parametric design for architecture.* Laurence King Publ.
Kim, K. J., Yun, W. G., Cho, N., & Ha, J. (2017). Life cycle assessment based environmental impact estimation model for pre-stressed concrete beam bridge in the early design phase. *Environmental Impact Assessment Review, 64*, 47–56.
Kohler, N., & Moffatt, S. (2003). Life-cycle analysis of the built environment. *Industry and Environment, 26*(2–3), 17–21.
Li, B., & Yao, R. (2009). Urbanisation and its impact on building energy consumption and efficiency in China. *Renewable Energy, 34*(9), 1994–1998.
Li, D., Chen, H., Hui, E. C., Zhang, J., & Li, Q. (2013). A methodology for estimating the life-cycle carbon efficiency of a residential building. *Building and Environment, 59*, 448–455.
Macdonald, I. A. (2002). *Quantifying the effects of uncertainty in building simulation.* Scotland: University of Strathclyde Glasgow.
Macdonald, I., & Strachan, P. (2001). Practical application of uncertainty analysis. *Energy and Buildings, 33*(3), 219–227.
MacLeamy, P. (2004). *Collaboration, integrated information and the project lifecycle in building design, construction and operation, WP-1202.* The Construction Users Roundtable.
Manfren, M., Aste, N., & Moshksar, R. (2013). Calibration and uncertainty analysis for computer models – A meta-model based approach for integrated building energy simulation. *Applied Energy, 103*, 627–641.
MHURD China (2014). *National norm of building fire system design of China-GB50016-2014.* (in Chinese).
Mohammed, A. A., Minhas, R., Wu, Q. J., & Sid-Ahmed, M. A. (2011). Human face recognition based on multidimensional PCA and extreme learning machine. *Pattern Recognition, 44*(10–11), 2588–2597.
Monizza, G. P., Bendetti, C., & Matt, D. T. (2018). Parametric and Generative Design techniques in mass-production environments as effective enablers of Industry 4.0 approaches in the Building Industry. *Automation in Construction, 92*, 270–285.
Mujumdar, P., & Maheswari, J. U. (2018). Design iteration in construction projects – Review and directions. *Alexandria Engineering Journal, 57*(1), 321–329.
Muthén, L. K., & Muthén, B. O. (2002). How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling, 9*(4), 599–620.
Ochoa, C. E., & Capeluto, I. G. (2009). Advice tool for early design stages of intelligent facades based on energy and visual comfort approach. *Energy and Buildings, 41*(5), 480–488.
Østergård, T., Jensen, R. L., & Maagaard, S. E. (2016). Building simulations supporting decision making in early design – A review. *Renewable and Sustainable Energy Reviews, 61*, 187–201.
Rezaee, R., Brown, J., Augenbroe, G., & Kim, J. (2015). Assessment of uncertainty and confidence in building design exploration. *AI EDAM, 29*(4), 429–441.
Schlueter, A., & Thesseling, F. (2009). Building information model based energy/exergy performance assessment in early design stages. *Automation in Construction, 18*(2), 153–163.
Schwarz, C. J. (2011). *Sampling, regression, experimental design and analysis for environmental scientists, biologists, and resource managers.* Department of Statistics and Actuarial Science, Simon Fraser University57.
Shrestha, D. L., & Solomatine, D. P. (2006). Machine learning approaches for estimation of prediction interval for the model output. *Neural Networks, 19*(2), 225–235.
Stetco, A., Zeng, X.-J., & Keane, J. (2015). Fuzzy C-means++: Fuzzy C-means with effective seeding initialization. *Expert Systems with Applications, 42*(21), 7541–7548.
Tian, W., Yang, S., Zuo, J., Li, Z., & Liu, Y. (2017). *Relationship between built form and energy performance of office buildings in a severe cold Chinese region. Building simulation.* Springer11–24.
Vázquez-Canteli, J. R., Ulyanin, S., Kämpf, J., & Nagy, Z. (2019). Fusing TensorFlow with building energy simulation for intelligent energy management in smart cities. *Sustainable Cities and Society, 45*, 243–257.
Wang, S., Li, X., & Zheng, X. (1994). *Architectural design data set version 2.* China building industry press (in Chinese).
Yang, J. (2002). *Life cycle assessment and application.* Beijing, China: China Meteorological Press (in Chinese).
Yang, X. (2003). *Quantitative assessment of environmental impact on construction during planning and designing phases.* Tsinghua University.
Zhang, X., Shen, L., & Zhang, L. (2013). Life cycle assessment of the air emissions during building construction process: A case study in Hong Kong. *Renewable and Sustainable Energy Reviews, 17*, 160–169.
Zhou, Y., & Su, Y. (2009). Assessment of building energy consumption and environmental impact based on life cycle theory. *Construction Management Modernization,* (23), 109–112 (in Chinese).
Zhou, Q., Zhou, H., Zhu, Y., & Li, T. (2015). Data-driven solutions for building environmental impact assessment. *Proceedings of the 2015 IEEE 9th international conference on semantic computing (IEEE ICSC 2015)* (pp. 316–319).