Multi-zone optimisation of high-rise buildings using artificial intelligence for sustainable metropolises. Part 1
Background, methodology, setup, and machine learning results

Ekici, B.; Kazanasmaz, Tuğçe; Turrin, M.; Tasgetiren, Fatih; Sariyildiz, I.S.

Important note
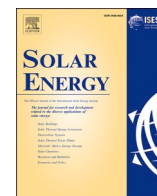To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Multi-zone optimisation of high-rise buildings using artificial intelligence for sustainable metropolises. Part 1: Background, methodology, setup, and machine learning results

Berk Ekici [a],[*], Z. Tuğçe Kazanasmaz [b], Michela Turrin [a], M. Fatih Taşgetiren [c], I. Sevil Sariyildiz [a]

[a] Delft University of Technology, Faculty of Architecture and the Built Environment, Chair of Design Informatics, Julianalaan 134, 2628 BL Delft, the Netherlands
[b] Izmir Institute of Technology, Department of Architecture, Gulbahce Kampus, Urla 35430, Izmir, Turkey
[c] Yasar University, Department of International Logistics Management, Universite Caddesi No:37-39, Agacli Yol, Bornova 35100, Izmir, Turkey

ABSTRACT

Designing high-rise buildings is one of the complex tasks of architecture because it involves interdisciplinary performance aspects in the conceptual phase. The necessity for sustainable high-rise buildings has increased owing to the demand for metropolises based on population growth and urbanisation trends. Although artificial intelligence (AI) techniques support swift decision-making when addressing multiple performance aspects related to sustainable buildings, previous studies only examined single floors because modelling and optimising the entire building requires extensive computational time. However, different floor levels require various design decisions because of the performance variances between the ground and sky levels of high-rises in dense urban districts. This paper presents a multi-zone optimisation (MUZO) methodology to support decision-making for an entire high-rise building considering multiple floor levels and performance aspects. The proposed methodology includes parametric modelling and simulations of high-rise buildings, as well as machine learning and optimisation as AI methods. The specific setup focuses on the quad-grid and diagrid shading devices using two daylight metrics of LEED: spatial daylight autonomy and annual sunlight exposure. The parametric model generated samples to develop surrogate models using an artificial neural network. The results of 40 surrogate models indicated that the machine learning part of the MUZO methodology can report very high prediction accuracies for 31 models and high accuracies for six quad-grid and three diagrid models. The findings indicate that the MUZO can be an important part of designing high-rises in metropolises while predicting multiple performance aspects related to sustainable buildings during the conceptual design phase.

1. Introduction

High-rise buildings began to emerge at the end of the 19th century to provide extra floor space in limited urban plots (Al-Kodmany and Ali, 2013). In the 20th century, population growth and urbanisation trends increased in the world (Cohen, 2006). According to a United Nations report (UN, 2019), 30% of the world's population lived in urban areas in 1950. This percentage increased to 55% in 2018, and the projection by 2050 was 68%. An increase in population and the percentage of those living in urban areas will add 2.5 billion people to the world's urban population by 2050. Moreover, there were 33 megacities with more than 10 million inhabitants in 2018. The projection indicates that this number will increase to 43 by 2030. Because of population growth and

urbanisation trends, the number and height of completed high-rise buildings have also increased over time (CTBUH, 2020).

Owing to a rapid and global increase in floor areas, the final energy use of buildings reached approximately 128 exajoules (EJ) in 2019, while it was 118 EJ in 2010 (IEA, 2020). An increasing number of high-rise buildings contribute significantly to energy use as they consume more energy with an additional effect of CO_2 emissions compared with low-rise buildings (Godoy-Shimizu et al., 2018). Another consequence of constructing more and taller high-rise buildings is the increase in building density in urban areas (Lee et al., 2017). To achieve the targets of the International Energy Agency for sustainable development scenarios, architects and engineers should consider the following challenges while designing high-rise buildings for metropolises:

Nomenclature

Daylight metrics and material properties

ASE	Annual sunlight exposure [%]
DA	Daylight autonomy [%]
DGP	Daylight glare probability
g-val.	G value of the glazing material
IL	Illumination level [lux]
sDA	Spatial daylight autonomy [%]
Tvis	Visible transmittance of the glazing material
UDI	Useful daylight illuminance [%]
U-val.	U value of the glazing material [W/m^2 K]

Machine learning

ANN	Artificial neural network
DNN	Deep neural network
FNN	Feedforward neural network
GBM	Gradient boosting model
GLM	Generalised linear model
kNN	k-nearest neighbour
MARS	Multivariate adaptive regression splines
ML	Machine learning
MLR	Multiple linear regression
MRA	Multiple regression analysis
RBF	Radial basis function

RBFNN	Radial basis function neural network
ReLU	Rectified linear units
RF	Random forest
SVM	Support vector machine

Statistics

CV	Cross-validation
MAE	Mean absolute error
MSE	Mean squared error
R^2	R-squared
Std	Standard deviation

Optimisation

GD	Gradient descent
NFL	No free lunch
SGD	Stochastic gradient descent

Others

CTBUH	Council on Tall Buildings and Urban Habitat
GH	Grasshopper 3D algorithmic modelling environment
IEA	International Energy Agency
IES	Illuminating Engineering Society
LEED	Leadership in Energy and Environmental Design
MUZO	Multi-zone optimisation
USGBC	United States Green Building Council

- Dense urban areas cause performance variations between ground and sky levels in high-rise buildings (Samuelson et al., 2016).
- Sustainable buildings require the integration of multiple performance aspects, such as natural daylight, energy consumption, and comfort (Evins, 2013).

During the design process, the conceptual phase requires a high awareness of decisions because it affects the overall performance of the buildings (Sariyildiz, 2012). Owing to the complexity of design problems, optimisation algorithms are widely used to investigate sustainable design alternatives during the conceptual design phase (Asadi and Geem, 2015). Because the performance aspects of sustainable buildings require simulations, the optimisation process entails a significant amount of time. The common approach is to integrate machine learning (ML) techniques to predict performance aspects to support swift decision-making with optimisation algorithms in computationally expensive design problems (Westermann and Evins, 2019). Optimising high-rise buildings in dense urban districts is more challenging because various floor levels require different design decisions owing to performance variations in ground and sky levels. In addition, these decisions are based on simulations, which require expensive computational time, and optimisation processes that need to cope with an enormous number of design parameters. Therefore, new methods are required to optimise the multiple floor levels of high-rise buildings when proposing sustainable alternatives within a limited time.

This paper introduces a novel multi-zone optimisation (MUZO) methodology of optimising high-rises by considering multiple floor levels as different optimisation problems to investigate sustainable alternatives during the conceptual phase. The proposed methodology includes parametric modelling and simulations of high-rise buildings, an artificial neural network (ANN) (an ML technique based on a network of neurones) for performance prediction, and computational optimisation with a decision framework. Part 1 of the MUZO study focuses on solving computationally expensive simulations while presenting the background, methodology, and setup for case studies that contain two types of shading devices: quad-grid and diagrid. The building performance model focuses on the daylight metrics of Leadership in Energy and

Environmental Design (LEED) v4.1, namely, spatial daylight autonomy (sDA) and annual sunlight exposure (ASE), for each scenario. The results present the learning scores of 40 surrogate models developed for each performance aspect using advanced ANN techniques. Part 2 of the MUZO study deals with the optimisation challenge while explaining the problem formulations to optimise the sDA and ASE using the 40 predictive models presented in this paper. Considering the near feasibility threshold adaptive penalty function, the optimisation process employs three algorithms, namely, self-adaptive differential evolution with ensemble of mutation strategies using the Optimus plug-in (Cubukcuoglu et al., 2019), covariance matrix adaptation with evolution strategy, and radial basis function optimisation using the Opossum plug-in (Wortmann, 2017). After validating the method by comparing the MUZO results with the regular high-rise scenarios, the paper discusses the advantages and disadvantages and underlines the potential and future research directions. In this paper, Section 2 presents the state of the art, Section 3 introduces the MUZO methodology, Section 4 explains the setup, Section 5 reports the ANN results, and Section 6 concludes the paper.

2. State of the art for AI in the design of sustainable high-rises

This section presents previous studies focusing on performance aspects related to sustainable high-rise buildings in three subsections: ML, computational optimisation, and ML with optimisation applications. Subsequently, the original contribution of the MUZO methodology is summarised.

2.1. Machine learning applications

Over the last two decades, ML techniques have been used to address the computational burden of simulations of high-rise buildings. An early study discussed regression models to predict energy performance (Lam et al., 1997). After a decade, Ko et al. (2008) focused on the daylight factor as part of the LEED v2.2 for building shape, layout, and façade parameters. In the following years, Li and Li (2015) examined the annual ventilation rate, in addition to energy performance. Tian et al.

(2020) developed models for energy-efficient heating design in office buildings considering conventional modelling processes and an innovative two-step method. Recently, researchers began to use sensitivity analyses with ML techniques to decrease the design complexity (Chen et al., 2017a; Chen et al., 2017b). Since the early years, various aspects have been used to predict the performance of high-rises. However, none of these studies focused on predicting the performance of an entire building. The general approach focused on a single-floor level (or part) of the high-rise model.

2.2. Computational optimisation applications

High-rise buildings are one of the complex design tasks of architecture because various decisions are required for the shape, layout, and façade parameters considering multiple performance aspects. Therefore, different methods have been examined to address the complexity of these buildings. Considering resource production systems, Imam and Kolarevic (2016) proposed a concept to optimise energy, food, water,

and land in high-rises. In addition to producing energy, two studies focused on energy performance (Chen et al., 2019a; Chen et al., 2019c), one study examined the energy demand with adaptive thermal comfort (Giouri et al., 2020), and another considered techno-economic aspects (Liu et al., 2020). In addition to façade parameters, Gan et al. (2019) investigated the geometric, position, and functional attributes to optimise energy efficiency. Despite the promising results and design alternatives, two studies (Chen et al., 2019c; Li and Li, 2015) considered the surroundings of the plots being studied, one study compared various optimisation algorithms (Chen et al., 2019a), and none of them replicated the heuristic optimisation process. Consequently, the general approach, as in ML applications, focuses on a single floor level (or part) of the high-rise model.

2.3. Machine learning and computational optimisation applications

In the high-rise domain, early examples of predictive models focused on evaluating the design performances. Some of the recent studies
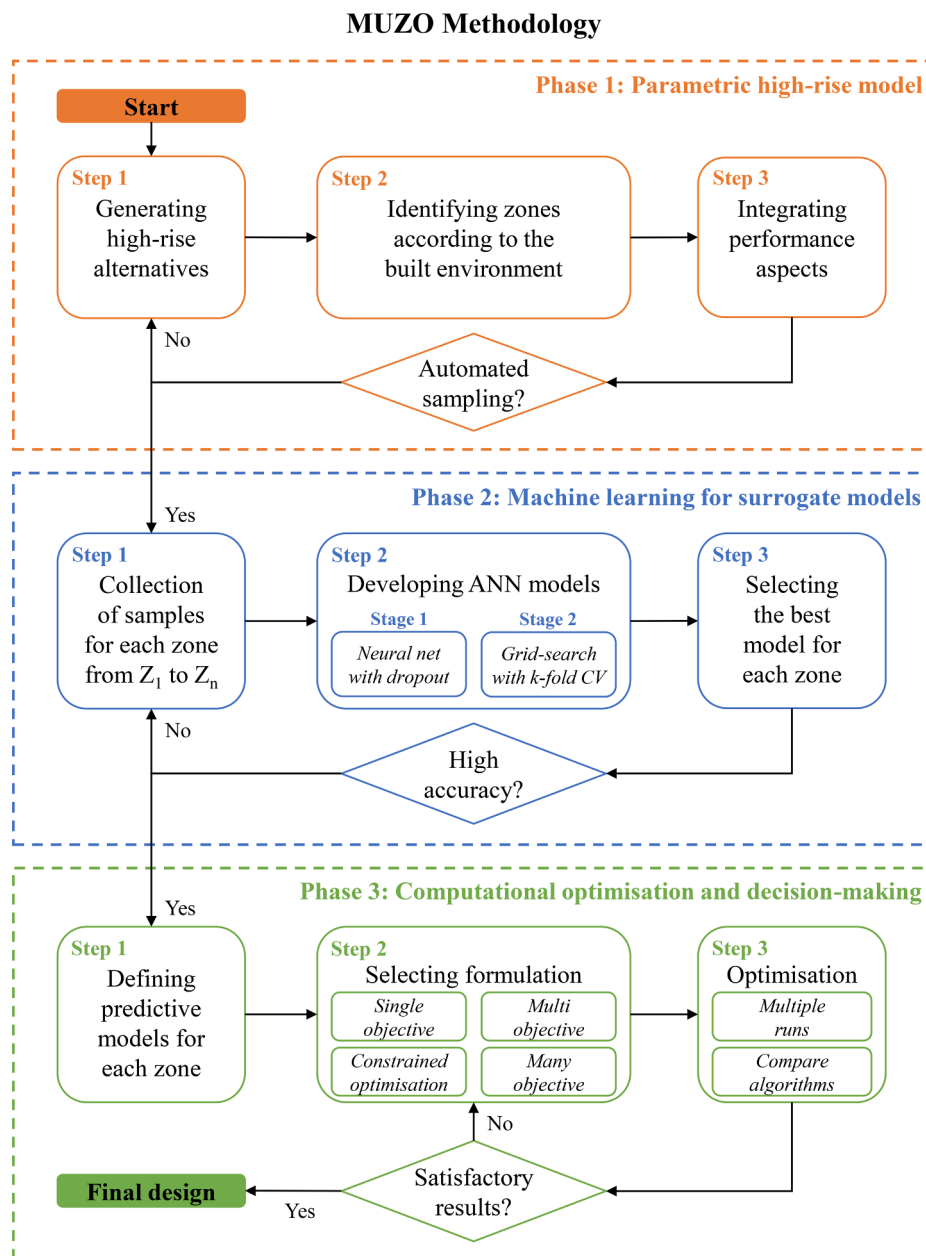

Fig. 1. MUZO methodology.

considered predictive models with optimisation algorithms because of the potential to determine optimal solutions in a short time. Early examples used regression models, support vector machines (SVMs), and multi-objective optimisation (Chen and Yang, 2017, 2018). In addition, Chen et al. (2019b) conducted a sensitivity analysis to decrease the design complexity. Despite the fast evaluation potential of using ML with optimisation, the aforementioned studies considered specific floor levels of high-rise models.

2.4. Original contribution of the research

The MUZO methodology is proposed to optimise the entire shape of a high-rise building to investigate sustainable design alternatives while addressing the computational burden. Because dense urban areas result in performance variations between the ground and sky levels, a unique optimisation strategy is required. Therefore, the MUZO methodology suggests dividing the high-rise building into equal subdivisions (or zones), which can be considered as different design problems. In addition, this paper suggests an advanced model selection to provide high prediction accuracies, as well as a decision framework by comparing the algorithms and replicating the optimisation process. Thus, the MUZO methodology aims to determine the optimal design solution by achieving sustainable high-rise alternatives for dense urban districts.

3. Multi-zone optimisation methodology

Previous studies demonstrated that the optimisation of high-rise buildings can focus on multiple performance aspects that may require various digital platforms. Considering the flexibility of integrating different software, Fig. 1 shows the phases of the MUZO methodology. The parametric high-rise model focuses on generating design alternatives with performance evaluations in phase 1. ML for surrogate models addresses the computational burden of multiple performance aspects related to sustainable buildings in phase 2. Finally, the computational optimisation and decision-making phase investigates the desirable performance for the entire high-rise building.

3.1. Parametric high-rise model

The first phase of the methodology considers the parametric model, which involves generating configurations of the high-rise building using design variables. Preparing the model requires three steps: developing the parametric high-rise model to generate design alternatives, identifying zones according to the surroundings of the plot being studied, and integrating performance aspects.

Step 1 (Generating high-rise alternatives): Initially, creating the context around the plot area is the first step during the development of the parametric high-rise model. This is because surroundings with different densities may require various design strategies and parameters in the conceptual phase. When the built environment is modelled, a parametric high-rise model that involves decision variables related to the building shape, façade design, layout, and operation is generated. Few tools are available for use in this step, i.e. Generative Components (Aish, 2003; Bentley, 2003), Dynamo (Dynamo, 2011; Keough, 2011), and Grasshopper 3D (GH) (Rutten, 2015). The MUZO methodology can include all of these parameter types and available tools during form generation.

Step 2 (Identifying zones): As mentioned previously, dense urban districts result in performance variances between the ground and sky levels. Therefore, the second step of parametric modelling identifies the zones, which involves subdividing the entire building into smaller pieces to focus on various floor levels as different optimisation problems. The number of zones is a predefined variable that depends on the density of the plot under study. For instance, in urban areas with low-density, high-rise buildings can be divided into three zones, whereas this amount may increase to five in the mid-density scenarios. For high-density scenarios, more than five zones can be used for an extensive investigation of the

effects of the surrounding at various levels. After determining the number of zones, the next step is to identify floor levels in each zone, because performance aspects, such as daylight and solar radiation, require floor surfaces for the simulation to be conducted. While a large number of selected floor levels requires an extensive simulation time, fewer selected floor levels may result in decision-making with limited awareness of the entire building's performance. Fig. 2 shows zoning scenarios for low-density, mid-density, high-density urban areas and different selections of floor levels.

Step 3 (Integrating performance aspects): The final step of the first phase in MUZO methodology involves evaluating the high-rise model using the performance aspects of sustainable buildings. The state of the art considers a limited number of performance criteria because of two reasons. First, considering multiple aspects requires extensive computational time for simulation-based evaluation. Second, the complexity of the design task increases owing to multiple performance aspects. In addition, conflicting performances introduce an additional challenge during the conceptual phase (Kirimtat et al., 2016). The proposed MUZO methodology can integrate any performance criteria to determine sustainable high-rise alternatives. Challenges on computational burden and complexity are addressed in the subsequent phases.

3.2. Machine learning for surrogate models

When the parametric model and simulations are set, various design alternatives can present the simulation results to gain awareness of the performance for different design scenarios. ANN models, which can swiftly evaluate the building performance, are used in the second phase of MUZO, which requires three steps:

Step 1 (Collection of samples): Sampling, which is the first step of ML in MUZO, is an essential process of surrogate modelling. With a specific distribution in the data, ML algorithms can learn and predict data with high accuracy. Recently, Westermann and Evins (2019) presented two types of sampling: static sampling (e.g. Latin hypercube sampling (Loh, 1996)) and adaptive sampling (i.e. sequential space-filling (Crombecq and Dhaene, 2010)). The selection of the sampling method depends on the category of surrogate models that can be either global or local models. All sampling methods using a global modelling approach can be used in the MUZO methodology. On the effect of the sample size, Chatzikonstantinou and Sariyildiz (2016) discussed that the extension of the dataset is frequently beneficial. In addition, Roman et al. (2020) presented the most commonly used sampling methods in building-performance simulations. Among these sampling methods, one of the most common approaches is

$$n_s = 22.5n_i \tag{1}$$

where n_s is the sample size and n_i is the number of independent variables. Because each subdivided part corresponds to a different optimisation model, the MUZO methodology proposes a unique sample collection framework (Fig. 3). After subdividing the high-rise building into zones, each subdivision is used to generate its own samples. When the process is complete, each generated sampling file, which belongs to one zone, can be used in different surrogate models.

Step 2 (Developing ANN models): ANNs, which correspond to the second step of the ML phase, are widely used methods in ML domains to predict various aspects of building performance. This because ANNs can manage large sample sizes for many variables and predict the performance with high accuracies (Westermann and Evins, 2019). Various ANN types, such as feedforward neural networks (FNNs) and radial basis function neural networks (RBFNNs), have been used to estimate building performance (Roman et al., 2020). In this paper, the development of ANN models consists of two stages:

Stage 1 (Neural net with dropout): The development of ANNs begins with reading and scaling the data, which frequently contain different parameters with units and metrics. After the reading process, scaling is
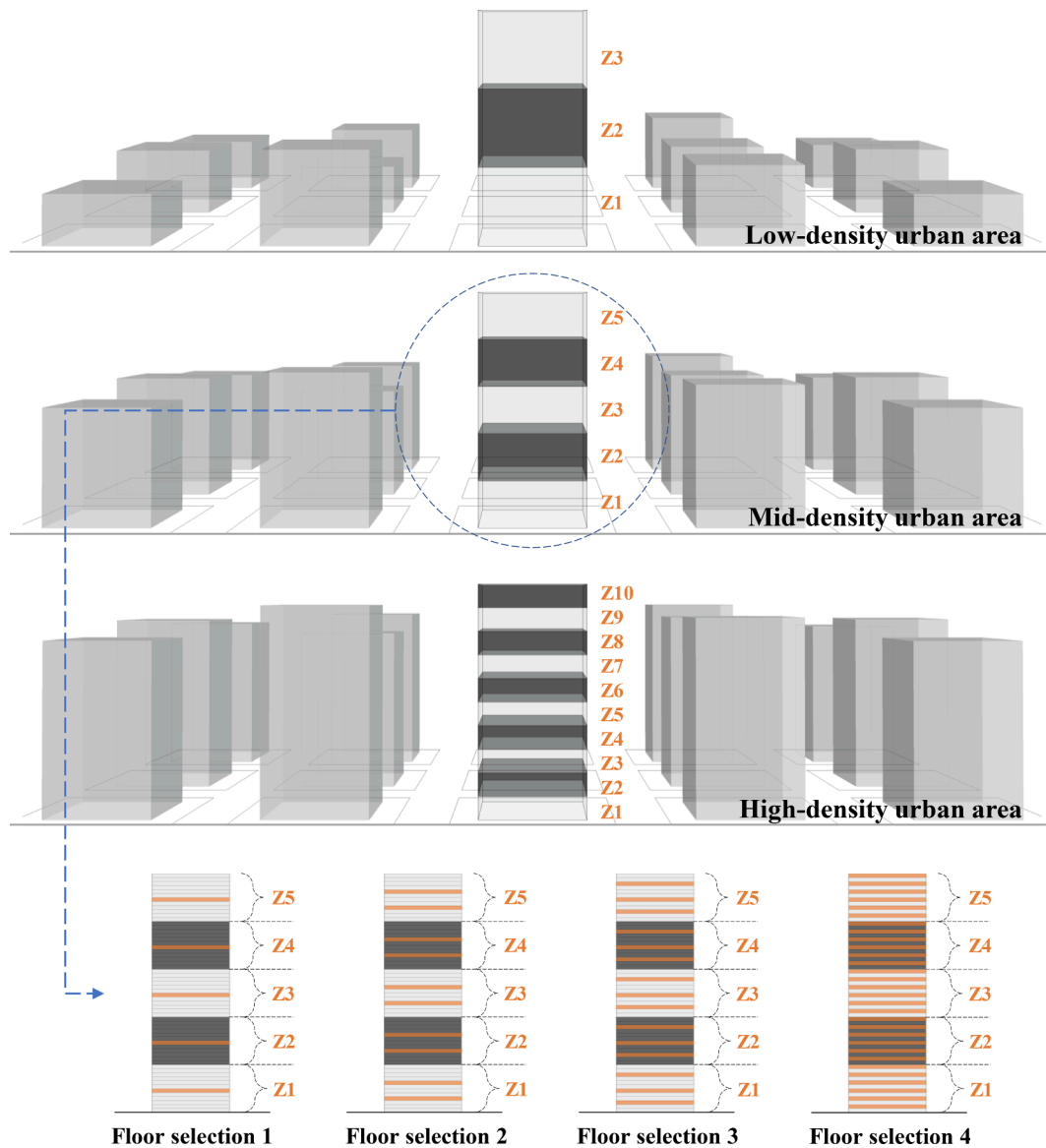
Fig. 2. Zoning examples for various urban densities and floor selections.
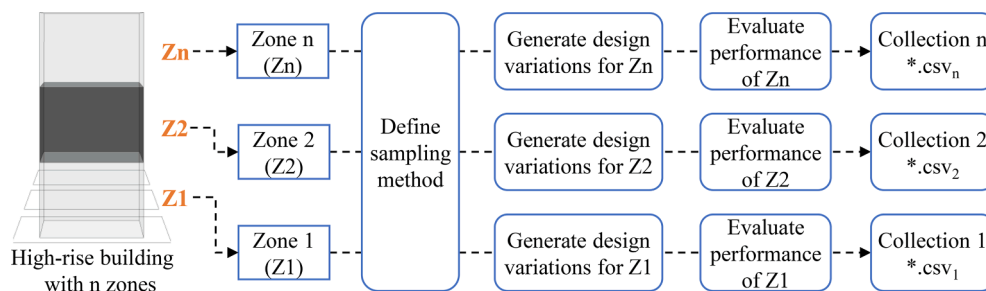

Fig. 3. Sample collection framework.

performed to obtain all inputs and outputs within the same boundaries using several scaling methods (Grus, 2019). For min–max scaling, the data is normalised as

$$x^{'} = \sigma(max(x) - min(x)) + min(x) \qquad (2)$$

where $x^{'}$ is the scaled value, x is the original value and σ is its standard deviation. Before selecting the scaling method, the problem type must be

identified, which can be either classification or regression. While classification problems focus on predicting a class label, regression problems consider predicting a quantity. In addition, splitting data is crucial for identifying training and test sets using a rate, e.g. 0.2. or 0.25, according to Westermann and Evins (2019). When the ANN model is finalised, the architecture contains various layers (Fig. 4).

Each neurone in the hidden layers receives the weighted sum of inputs to pass the result through an activation function. In the output
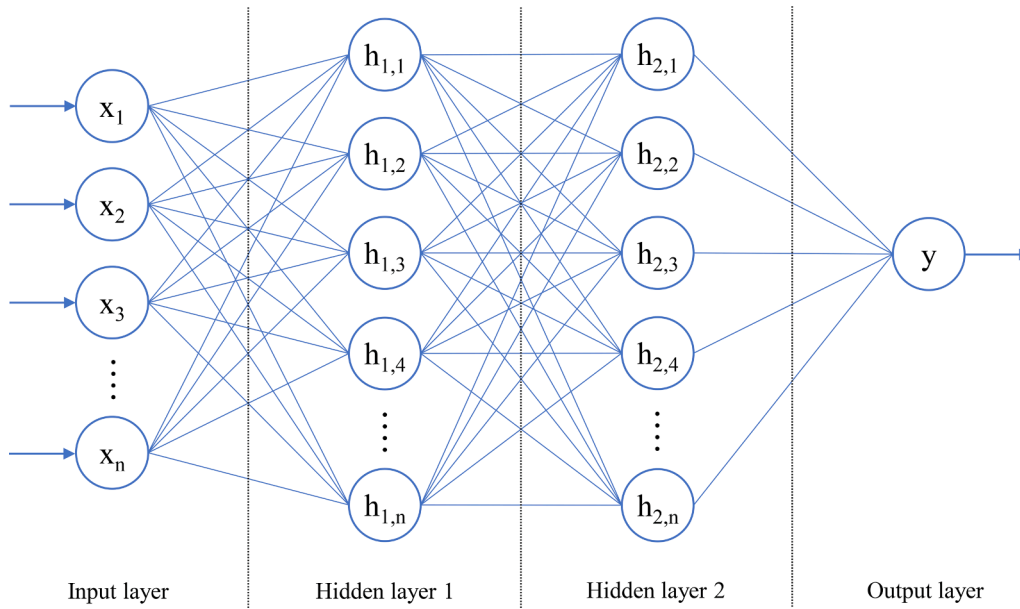
Fig. 4. ANN architecture.

layer, the ANN predicts the final solution considering this procedure for all neurones. The network also involves bias layers that can shift the result of each layer. The activation (a) of each ith layer is

$$a_i = f\left(b_i + \sum_{j=1}^{m} w_{ij}x_i\right) \tag{3}$$

where f is the activation function, b is the bias, w_{ij} is the ith layer of the jth weight, and x_i is the input vector of the ith layer. When using rectified linear units (ReLU), each neurone is activated as follows:

$$f(x) = \begin{cases} 0 & for \quad x \leqslant 0 \\ x & for \quad x > 0 \end{cases} \tag{4}$$

Different functions, e.g. sigmoid, softplus, and tanh, activate the neurones with various equations that may affect learning performance. The forward process of ANNs can predict the solution using Eq. (3). To achieve high accuracy, a backward process is necessary to determine the best values for the weights and biases. Hence, backpropagation (Hecht-Nielsen, 1992) involves a loss function and an optimisation algorithm. Researchers have widely used gradient descent (GD) (Ruder, 2016), stochastic gradient descent (SGD) (Bottou, 2010), Adam (Kingma and Ba, 2014), and RMSProp (Mukkamala and Hein, 2017) algorithms for optimisation. For the loss functions of classification problems, cross-entropy (De Boer et al., 2005) and Kullback–Leibler divergence (Kullback and Leibler, 1951) can be considered. In regression problems, researchers use the mean squared error (MSE) in Eq. (5) (Mood, 1950), mean absolute error (MAE) in Eq. (6) (Willmott and Matsuura, 2005), and R-squared (R^2) value in Eq. (7) (Draper and Smith, 1998):

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (x_i - y_i)^2 \tag{5}$$

$$MAE = \frac{\sum_{i=1}^{n} |y_i - x_i|}{n} \tag{6}$$

$$R^2 = 1 - \frac{\sum_i (x_i - y_i)^2}{\sum_i (x_i - \overline{x})^2} \tag{7}$$

where y_i is the predicted data, x_i is the observed data, $\overline{x}$ is the mean of

the observed data, and n is the sample size. Various loss functions can be used to validate the accuracy of the trained model (Chatzikonstantinou and Sariyildiz, 2016). In addition, the dropout technique (Srivastava et al., 2014), which randomly drops units from the neural network with their connections, avoids overfitting. The MUZO methodology can involve multiple loss functions and dropouts at a rate between 0 and 1.

Stage 2 (Grid search with k-fold cross-validation (CV)): Developing a surrogate model is a black-box process. One of the reasons is that multiple hyperparameters, which are the parameters of the ANN (e.g. neurone size and batch size), are involved in the learning process. Various combinations of these factors affect learning and prediction accuracies. Therefore, a model validation technique is required to evaluate the accuracy of predictions. K-fold CV (Stone, 1974) is a well-known method for accurate estimations that randomly divides the original sample into k equal-sized subsamples. While one subsample is maintained as the test set, the remaining $k-1$ subsamples are the training sets. The standard deviation (Std) indicates the difference for each error for the k-fold CV:

$$Std = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N} (x_i - \overline{x})^2} \tag{8}$$

where N is the number of observations, $\{x_1, ..., x_N\}$ are the observed values, and $\overline{x}$ is the mean value of these observations. The aim is to determine satisfactory results for the MAE, MSE, and R^2, and achieve small Std values for each accuracy metric.

Step 3 (Selecting the best model): The final step of the ML phase involves the selection of the best model using the results of the grid search. In each zone, the criteria are the highest R^2 with low MAE, MSE, and Std for this process. Using the weights and biases of the final ANN, the predictive models are ready for use in the optimisation process.

3.3. Computational optimisation and decision-making

The final phase of the MUZO methodology, which consists of three steps, involves determining the design parameters for sustainable high-rise alternatives. The first step considers the development of predictive models using the ML outputs. The second step is selecting the problem formulation. Finally, the proposed decision-making framework reveals the optimised design solution by completing the MUZO methodology.

Step 1 (Defining predictive models): The development of predictive models requires weight and bias results collected from each ANN model.

Subsequently, the collected results are transformed into matrices considering the input vector and neurone sizes for each layer to initiate the first step of the optimisation phase. The definition of activation with n layers is as follows.

$$y = f_n(f_{n-1}(f_{n-2}(f_{n-3}(x \cdot w_{n-3} + b_{n-3}) \cdot w_{n-2} + b_{n-2}) \cdot w_{n-1} + b_{n-1}) \cdot w_n + b_n) \quad (9)$$

where y is the performance criterion to be predicted, x is the input vector, w_n is the nth weight, b_n is the nth bias, and f_n is the nth activation function. For any given x, the model estimates the performance results. Having weights and biases as recorded data suggests the possibility of using predictive models in various platforms, such as C#, C++, Python, and GH, during the optimisation process.

Step 2 (Selecting formulation): When the predictive models are ready, the next step is selecting the problem formulation for the optimisation process. Previous studies on building optimisation used single objective, weighted summation, multi-objective, many objectives, and constrained optimisation problems (Ekici et al., 2019a). For n parameters, the definition of the generalised problem formulation is

$$min : f_1(X), ..., f_k(X), X = (x_1, x_2, ..., x_n) and \ X \in S$$
$$subject \ to : g_i(x) \leqslant 0, i = 1, ..., p$$
$$h_j(X) = 0, j = p + 1, ..., m \quad (10)$$

where an integer $k > 0$ is the number of objective functions, S is the entire search space, p is the number of inequality constraints and $m - p$ is the number of equality constraints. For the maximisation problem, the transformation of the function can be

$$max f(x) = -f(x) \quad (11)$$

The trade-off between building performance affects the selection of the formulation. For one aspect, e.g. maximising daylight (Mangkuto et al., 2018), the single-objective formulation is convenient for the optimisation process. Another scenario may have two conflicting objectives, such as maximising the sDA and minimising the ASE. If one of these aspects requires a threshold according to the building standards, the formulation can be a single-objective constrained optimisation (Vera et al., 2017). Otherwise, the multi-objective (Yi, 2019) or weighted summation (Wagdy et al., 2015) approaches can be alternatives. For more than three objectives, the options are multi-objective constrained or many-objective formulations (Pilechiha et al., 2020).

Step 3 (Optimisation): The final step of phase 3 involves exploring the optimal alternative for each zone. In the optimisation domain, heuristic algorithms are employed to solve complex problems by mimicking behavioural patterns and social phenomena observed in nature (Del Ser et al., 2019). Additionally, in the domain of sustainable building design,

heuristics are widely used because promising alternatives are discovered in a reasonable time frame (Evins, 2013). Despite their advantages, these algorithms do not guarantee an optimal solution. According to the no free lunch (NFL) theorem (Wolpert and Macready, 1997), a global algorithm that can determine the optimal result for all problems does not exist. In architectural design, the subject is more dynamic than the benchmark problems. Each design scenario is a specific problem owing to the variances in the surroundings. In addition, the surroundings of the different cities have diverse climate types (e.g. Mediterranean climate in Izmir, Oceanic climate in Amsterdam). Therefore, architects can propose various alternatives for the same design problem (i.e. high-rise buildings) because concerns and the required strategies are different. Thus, we may conclude that "*the global optimal of each design problem is unexplored*". Therefore, the optimisation process of the MUZO methodology involves comparing various algorithms with replications for decision-making (Fig. 5). Single-objective optimisation algorithms report the best solution that can be used as the final design alternative. In multi-objective or many-objective optimisation problems, various post-optimisation analysis methods can be used to evaluate the quality of the single best solution during the decision-making process (Si et al., 2019), e.g. weighted summation approach (Cevizci et al., 2019), TOPSIS (Kim et al., 2013), analytic hierarchy process (Goussous and Al-Refaie, 2014), minimum distance to the utopic point (Riquelme et al., 2015), auto-associative models (Chatzikonstantinou and Sariyildiz, 2017).

4. Setup of the case study

This section explains the setup for evaluating the MUZO methodology considering a hypothetical dense urban district. The first subsection describes a parametric high-rise building with variables for the two façade types. The subsequent subsection presents the selected performance aspects of the simulation setup. Finally, surrogate modelling introduces the details of the sample collection and the development of ANN models.

4.1. Parametric high-rise model and the built environment

The hypothetical district had 25 plots in GH, each with a 2500 m^2 footprint with building heights between 50 and 150 m, which were generated randomly. The focus of the study was the central plot with 60 floors, 2100 m^2 net one-floor area, 150,000 m^2 gross floor area, and 50 $\times$ 50 m façade length. Fig. 6 shows the subdivisions (zones) of the building beginning from ground-level zone 1 (Z1) to sky-level zone 10 (Z10), as well as selected floor levels (second and fifth) of every zone for simulations. Table 1 presents the façade, shape, and glazing parameters
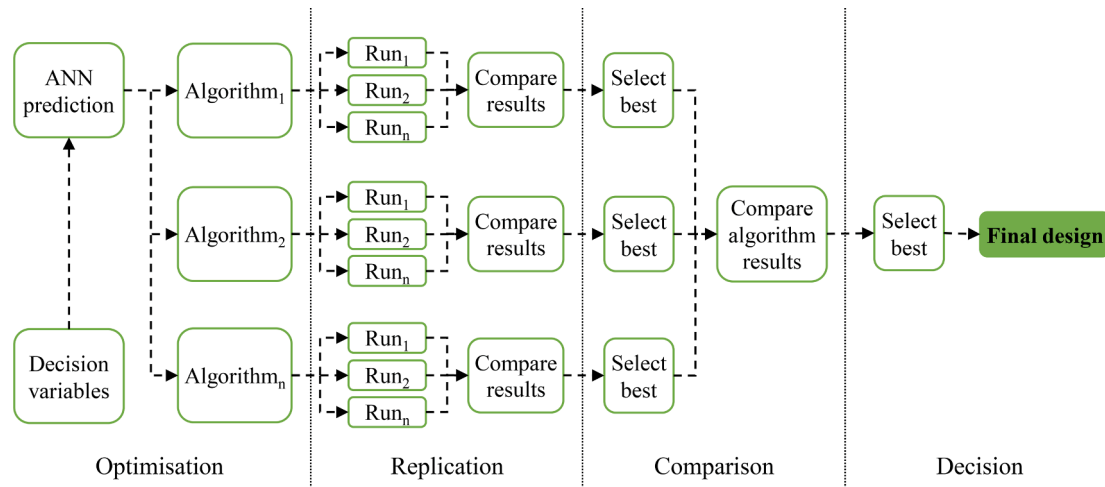

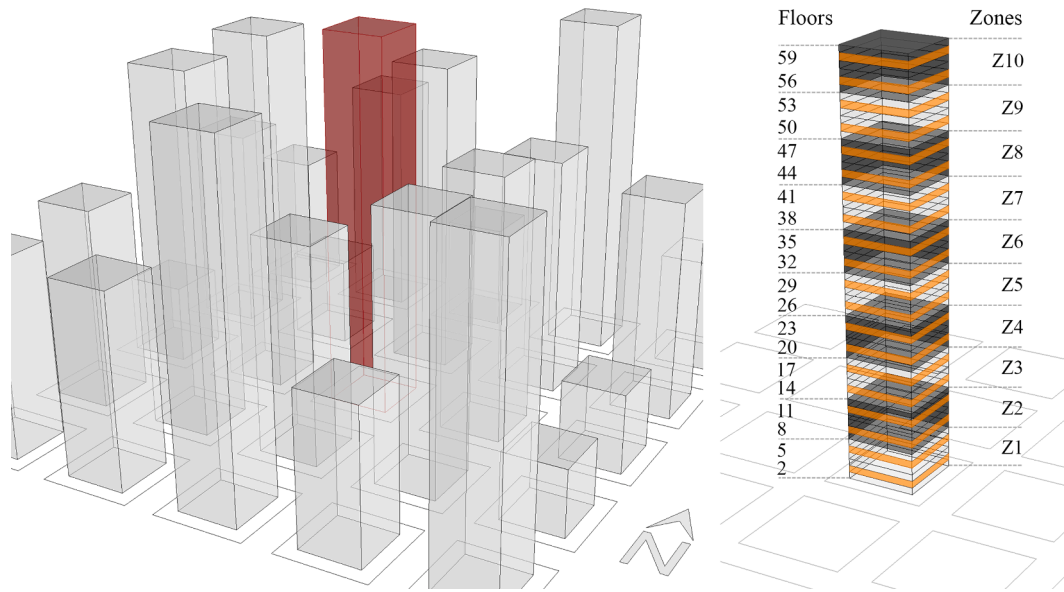**Fig. 5.** MUZO optimisation process and decision framework.

Fig. 6. Plot under study, zones (subdivisions) and selected floors of the high-rise for simulation.

Table 1
Parameters of the quad-grid and diagrid scenarios.

		Parameters	Explanation	Location	Type	Unit	Boundary
Quad-grid façade		$x_{Q1}, x_{Q6}, x_{Q11}, x_{Q16}$	Number of vertical devices	N-S-E-W	Discrete	–	[0, 8]
		$x_{Q2}, x_{Q7}, x_{Q12}, x_{Q17}$	Length of vertical devices		Continues	m	[0.0, 1.5]
		$x_{Q3}, x_{Q8}, x_{Q13}, x_{Q18}$	Rotation of vertical devices		Discrete	°	[−60, 60]
		$x_{Q4}, x_{Q9}, x_{Q14}, x_{Q19}$	Number of horizontal devices		Discrete	–	[0, 2]
		$x_{Q5}, x_{Q10}, x_{Q15}, x_{Q20}$	Length of horizontal devices		Continues	m	[0.0, 1.5]
		$x_{Q21}, x_{Q22}, x_{Q23}, x_{Q24}$	Glazing type		Discrete	–	[1, 4]
Diagrid façade		$x_{D1}, x_{D5}, x_{D9}, x_{D13}$	Length of 1st order diagonal	N-S-E-W	Continues	m	[0.0, 1.5]
		$x_{D2}, x_{D6}, x_{D10}, x_{D14}$	Length of 2nd order diagonal		Continues	m	[0.0, 1.5]
		$x_{D3}, x_{D7}, x_{D11}, x_{D15}$	Rotation of diagonal devices		Discrete	°	[−60, 60]
		$x_{D4}, x_{D8}, x_{D12}, x_{D16}$	Number of diagonal devices		Discrete	–	[0, 5]
		$x_{D17}, x_{D18}, x_{D19}, x_{D20}$	Glazing type		Discrete	–	[1, 4]
Building shape		$x_1, ..., x_{10}$	Floor-to-floor height of zones	–	Continues	m	[4.0, 5.0]
		$x_{11}, ..., x_{20}$	Rotation of zones		Discrete	°	[−10, 10]

	Type	Explanation				Tvis	U-val.	g-val.
Glazing types	1	Tinted float 8 mm blue − 12 mm air − Temperable Low-E 8 mm blue				0.22	1.6	0.28
	2	Temperable Low-E 8 mm neutral − 12 mm air − Clear float 8 mm − 12 mm air − Temperable Low-E 8 mm green				0.45	0.9	0.40
	3	Tinted float 8 mm green				0.68	5.6	0.51
	4	Ultra-clear float 8 mm − 12 mm air − Ultra clear float 8 mm				0.82	2.8	0.81

used in the parametric models for both scenarios.

The first façade design focused on horizontal and vertical shading devices using the number, length, and rotation of the devices with four glazing types. The second design considered diagonal shading devices involving the number, length, and rotation of first and second-order diagonals with the same glazing types. The design setups in Figs. 7 and 8 were used for each orientation, i.e. north (N), south (S), east (E), and west (W). Including floor-to-floor height and rotation parameters, the search space for the quad-grid scenario in one zone had 2.893399115e+28 design alternatives with 26 parameters, whereas this number was 3.054543465e+23 for the diagrid scenario with 22 parameters. Floor-to-floor height and rotation parameters of the lower zones affected the height and rotation of the higher zones. Therefore, the total amount of the design parameters in one zone increased from Z1 to Z10 (Fig. 9). Consequently, the quad-grid design had 26 parameters in Z1 and 44 variables in Z10, while the diagrid design had 22 parameters

in Z1 and 40 variables in Z10.

4.2. Performance metrics and simulation setup

We investigated two of the LEED v4.1 metrics for the case buildings, namely, the sDA and ASE, introduced for the green building certification program (USGBC, 2014). Both metrics are commonly used for various building functions to achieve sustainable design solutions (Bauer et al., 2017; Korsavi et al., 2016; Nezamdoost et al., 2018; Nezamdoost and Van Den Wymelenberg, 2017; Sherif et al., 2016; Wagdy et al., 2017). Recently, Illuminating Engineering Society (IES) presented definitions for sDA and ASE metrics (IES, 2013). The sDA evaluates the annual efficiency of ambient daylight levels in interior spaces. The calculation method results in the percentage of an analysis area with a minimum daylight illuminance level for specific hours. In contrast, the ASE indicates the potential visual discomfort in interior work environments.
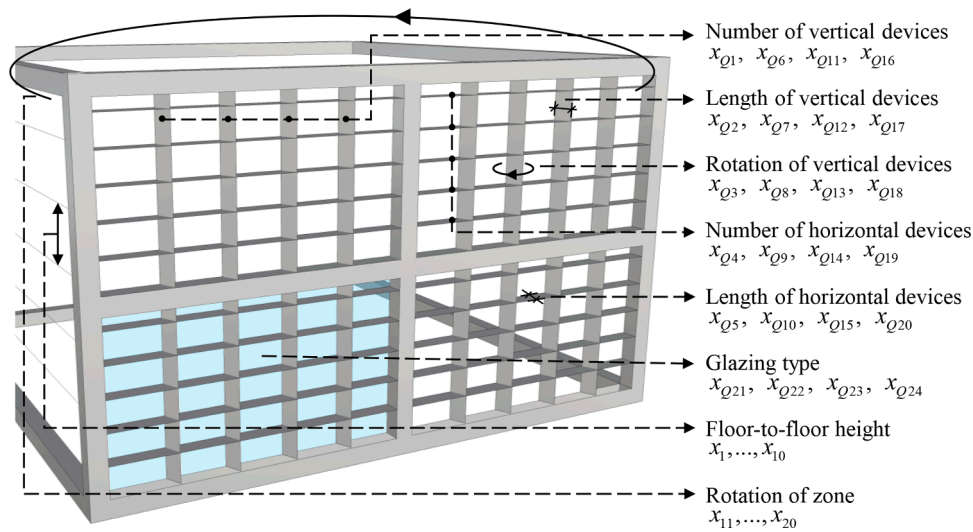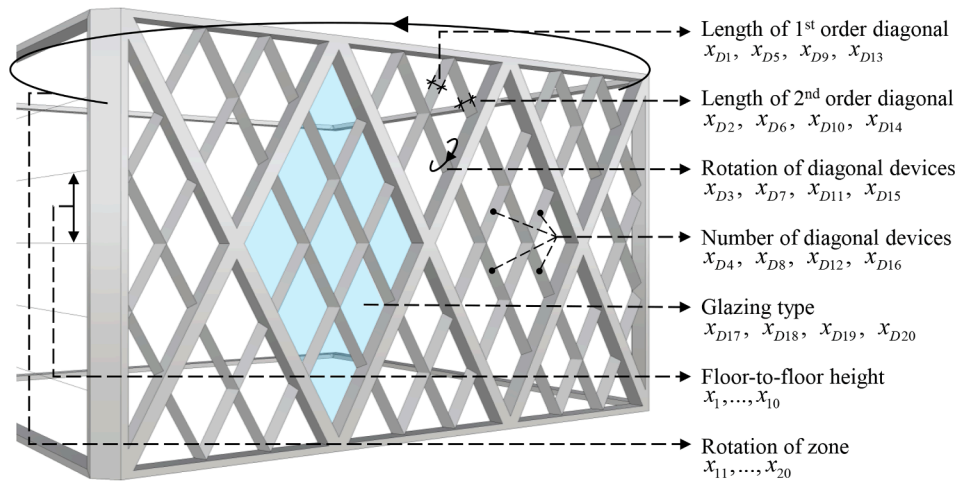
Fig. 7. Quad-grid façade design.

Fig. 8. Diagrid façade design.

The method results in the percentage of direct sunlight that exceeds a defined illuminance for the specified number of hours for the analysis area.

The simulation setup focused on the second and fifth floors of each zone (Fig. 10). The parametric model used the Diva plug-in v4.0.3.1 (Jakubiec and Reinhart, 2011) developed for GH to simulate sDA and ASE metrics with an EnergyPlus weather data file for Izmir City with a dry summer Mediterranean climate, latitude: 38.423733 and longitude 27.142826. Each zone was simulated using two analyses of planes with 180 sensor points each and was 0.8 m above the finished floor. Four glazing types, listed in Table 1, were separately used as decision variables for each orientation. As suggested by (IES, 2013), the setup simulated $sDA_{300/50\%}$ and $ASE_{1000,250h}$ for 10 h of occupation between 8 am and 6 pm. A single simulation task of one zone involved 360 sensors. For the radiance parameters listed in Table 2, the simulation process used values similar to those in previous studies because of the high computational cost. The setup was used to simulate the daylight performance of 7200 sensor points for the overall building evaluation of two scenarios.

4.3. Surrogate modelling

The surrogate modelling began with sampling collection, which considered 1000 samples for each zone using Latin hypercube sampling (Loh, 1996) and Eq. (1). One simulation required 4 min for two floors with the radiance parameters provided in Table 2. A computer with an Intel I7 4 core processor at 2.7 GHz and 16 GB DDR3 memory was used to calculate the computational burden as more than 55 days were required to collect 20,000 samples. In the next step, Python 3 (Van Rossum, 2009) was used with the additional libraries listed in Table 3 to develop ANN models with FNNs. After scaling the data for each zone with min–max scaling in Eq. (2), the SGD algorithm optimised weights and biases using Eq. (3) for all models considering 10-fold CV, three hidden layers, dropout rate with 0.1, and the ReLU activation function in Eq. (4). The automated Python program fit the model 324 times for all hyperparameter combinations for every zone. In total, the program ran 6480 different ANN models with various complexities.

5. Results

This section presents the sampling, grid search with CV, and tuned ANN results. The supplementary material provides statistics of collected samples, selected ANN models with learning scores, weights, biases, and computation time spent on the model selection for each zone.

Zone-Case	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}	x_{17}	x_{18}	x_{19}	x_{20}	$x_{Q1,...},x_{Q24}$	$x_{D1,...},x_{D20}$	(Q)	(D)
	Floor-to-floor height parameters										Rotation parameters										Façade parameters		Number of parameters	
Z1-Q	•										•										•		26	
Z1-D	•										•											•		22
Z2-Q	•	•									•	•									•		28	
Z2-D	•	•									•	•										•		24
Z3-Q	•	•	•								•	•	•								•		30	
Z3-D	•	•	•								•	•	•									•		26
Z4-Q	•	•	•	•							•	•	•	•							•		32	
Z4-D	•	•	•	•							•	•	•	•								•		28
Z5-Q	•	•	•	•	•						•	•	•	•	•						•		34	
Z5-D	•	•	•	•	•						•	•	•	•	•							•		30
Z6-Q	•	•	•	•	•	•					•	•	•	•	•	•					•		36	
Z6-D	•	•	•	•	•	•					•	•	•	•	•	•						•		32
Z7-Q	•	•	•	•	•	•	•				•	•	•	•	•	•	•				•		38	
Z7-D	•	•	•	•	•	•	•				•	•	•	•	•	•	•					•		34
Z8-Q	•	•	•	•	•	•	•	•			•	•	•	•	•	•	•	•			•		40	
Z8-D	•	•	•	•	•	•	•	•			•	•	•	•	•	•	•	•				•		36
Z9-Q	•	•	•	•	•	•	•	•	•		•	•	•	•	•	•	•	•	•		•		42	
Z9-D	•	•	•	•	•	•	•	•	•		•	•	•	•	•	•	•	•	•			•		38
Z10-Q	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•		44	
Z10-D	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•		•		40

• (Q) Quad-grid parameters   • (D) Diagrid parameters

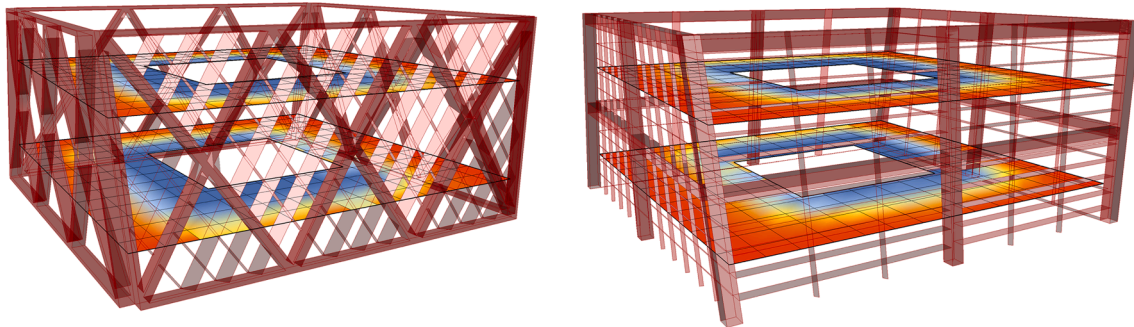**Fig. 9.** Complexity matrix for quad-grid and diagrid scenarios.

Fig. 10. Simulation setup for both scenarios.

Table 2
Radiance parameters.

	-aa	-ab	-ad	-ar	-as
This paper	0.15	2	512	256	128
(Ekici et al., 2019b)	0.15	2	512	256	128
(Kirimtat et al., 2019)	0.15	2	1000	300	20
(Yi, 2019)	0.15	2	512	256	128
(Lee et al., 2019)	0.15	2	512	256	128

Table 3
Python libraries.

Library	Explanation	Reference
Scikit-learn	ML library	(Pedregosa et al., 2011)
Keras	Deep learning library	(Chollet, 2015)
TensorFlow	Open source ML platform	(Abadi et al., 2016)
Pandas	Data analysis library	(McKinney, 2010)
Joypy	Plot library	(Taccari, 2017)
Plotly	The interactive graphing library	(Plotly, 2015)
Matplotlib	Static, animated, and interactive visualisation library	(Hunter, 2007)

5.1. Sampling results

The collected samples, which were published as an open-access dataset in (Ekici et al., 2020), contain the simulation results for the quad-grid and diagrid. Each zone had ASE and sDA results indicated as ASE_1 and sDA_1 on the second floor and ASE_2 and sDA_2 on the fifth floor. ASE_avg and sDA_avg, which represent the average values of these floors (Fig. 11), were used to develop the surrogate models.

The sDA results of the quad-grid application were between 41.9% and 100%, whereas ASE results were in the range of 9.4% and 50.7%. In the diagrid scenario, these results were 33.1% and 93.75% for sDA and 16.75% and 46.2% for ASE. For the mean values, natural daylight availability increased from Z1 to Z10 for both scenarios. However, this caused an increment in ASE results. In addition, the means of the sDA results for the quad-grid were higher than those for the diagrid.

5.2. Grid search with cross-validation results

ANN models were trained using the developed Python program, considering grid search and 10-fold CV using the collected dataset. The average results and deviations of MSE, MAE, and R^2 in Eqs. (5), (6), (7) and (8) were recorded for each parameter combination during the
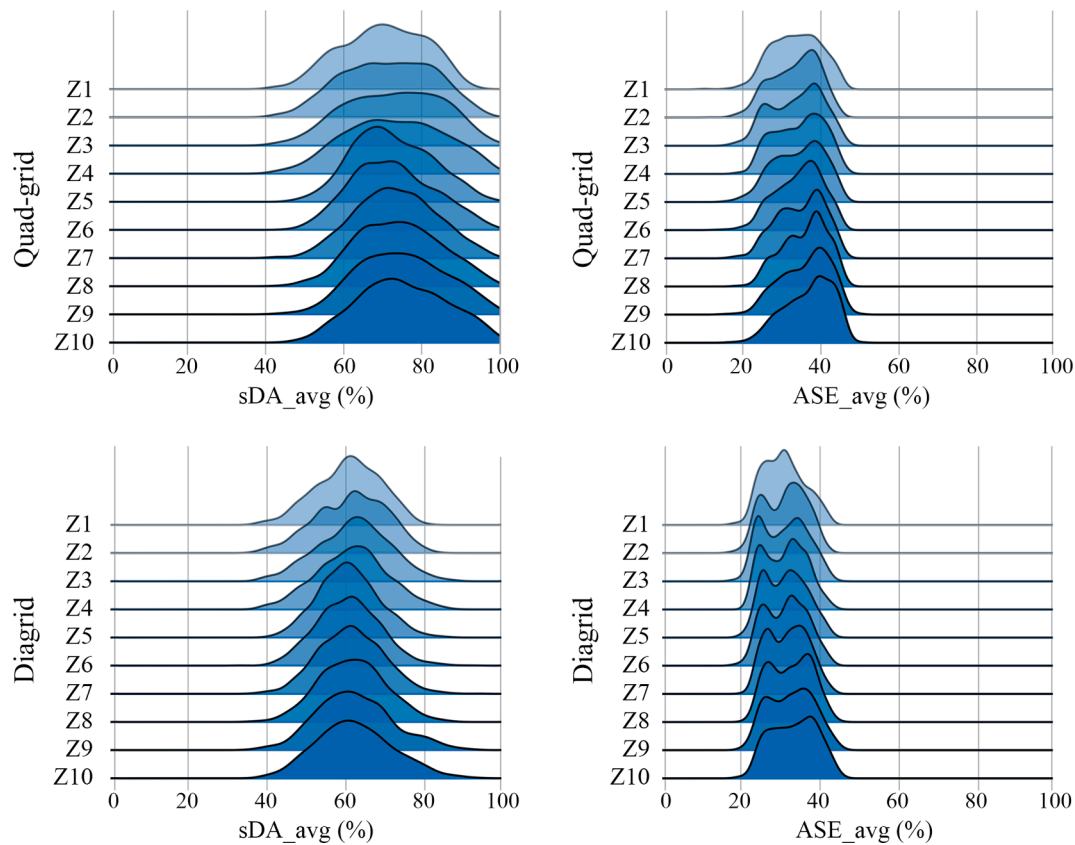
Fig. 11. Distributions of collected samples.

search process. The best hyperparameters with their statistical results are shown in Fig. 12.

The results indicated that 37 out of 40 ANN models had the best accuracy using 200 neurones. In the three models, the number of neurones was 100. For the momentum parameter, 33 models had the best score with 0.9, five models had 0.6, and two models had 0.3. Additionally, 25 models had the highest score using 0.1 for the learning rate, while twelve models using 0.05, and three models using 0.01. For epochs, eighteen models had the highest accuracy with 500, fourteen models had 750, six models using 1000, and two models had 250. Finally, the best selection for the batch size was 50 in twenty-one models, whereas it was 100 in nine models and 10 in ten models. The deviations of MAE, MSE, and R^2 indicated that all CV folds had similar results for all metrics with high accuracies. The R^2 values of 33 models were higher than 0.9, whereas in seven models they were higher than 0.8. All MAE, MSE, and Std results were less than 0.05. Consequently, the grid search results indicated promising accuracies to develop predictive models with the selected hyperparameters in the next step.

5.3. Tuned ANN results

Using the best hyperparameter sets shown in Fig. 12, ANN models were fit considering 0.2 for splitting data to demonstrate the learning behaviour and convergence by separating the dataset as training and test sets. Fig. 13 shows the R^2 results of these models, and appendices A1 and A2 provide the convergence of MSE and MAE while fitting the ANN models.

The R^2 values for all training sets were higher than 0.9. For the test sets, R^2 of fourteen quad-grid models out of twenty were higher than 0.9 and higher than 0.8 for five models. R^2 was slightly lower than 0.8 for only one model. In the diagrid application, R^2 for seventeen models was higher than 0.9, while it was higher than 0.8 for three models. The accuracy of the predictive models through the MSE and MAE results are

also provided in Appendix A. All reported MSE results were lower than 0.05. For the MAE, results of the ASE and sDA were lower than 0.05 in Z1, Z2, Z3, Z4, Z6, Z9, and Z10 for the quad-grid scenario. However, in other zones, the MAE of the ASE was slightly larger than 0.05, and it was lower than 0.05 for the sDA. In the diagrid scenario, the ASE and sDA models had MAE results lower than 0.05 in Z1, Z2, Z3, Z4, Z5, Z6, Z7, and Z9. In other zones, the MAE of the ASE was slightly higher than 0.05, whereas it was smaller than 0.05 for the sDA.

To compare the accuracy results reported for different design complexities, Fig. 14 shows the R^2 of similar studies focusing on ML applications in daylight (Chatzikonstantinou and Sariyildiz, 2016; Chen et al., 2017a,b; Chen and Yang, 2017; Kirimtat et al., 2019; Luo et al., 2021; Ngarambe et al., 2020; Sun et al., 2020). Papers in this domain have promising results for DA, sDA, illumination level (IL), and useful daylight illuminance (UDI). However, visual comfort metrics, such as daylight glare probability (DGP), have moderate accuracies for various ML applications. In this study, a similar result was achieved for the ASE metric because of the challenges in predicting comfort metrics. In addition, most of the previous studies considered design variables between 5 and 15, which provided less design complexity compared with this study. Only Kirimtat et al. (2019) considered 25 variables with R^2 values between 0.9 and 0.3. Consequently, the ML part of the MUZO methodology could address more complex designs while presenting high accuracies for all 40 models.

6. Conclusion

This paper presents the first part of the MUZO study, focusing on the background, methodology, setup, and ML results. The proposed methodology managed sampling and ANN development for 40 different models using a parametric high-rise model in a dense urban district. In addition, the developed Python program was used to investigate the best models for all zones of the two scenarios in 403 h. Based on the reported
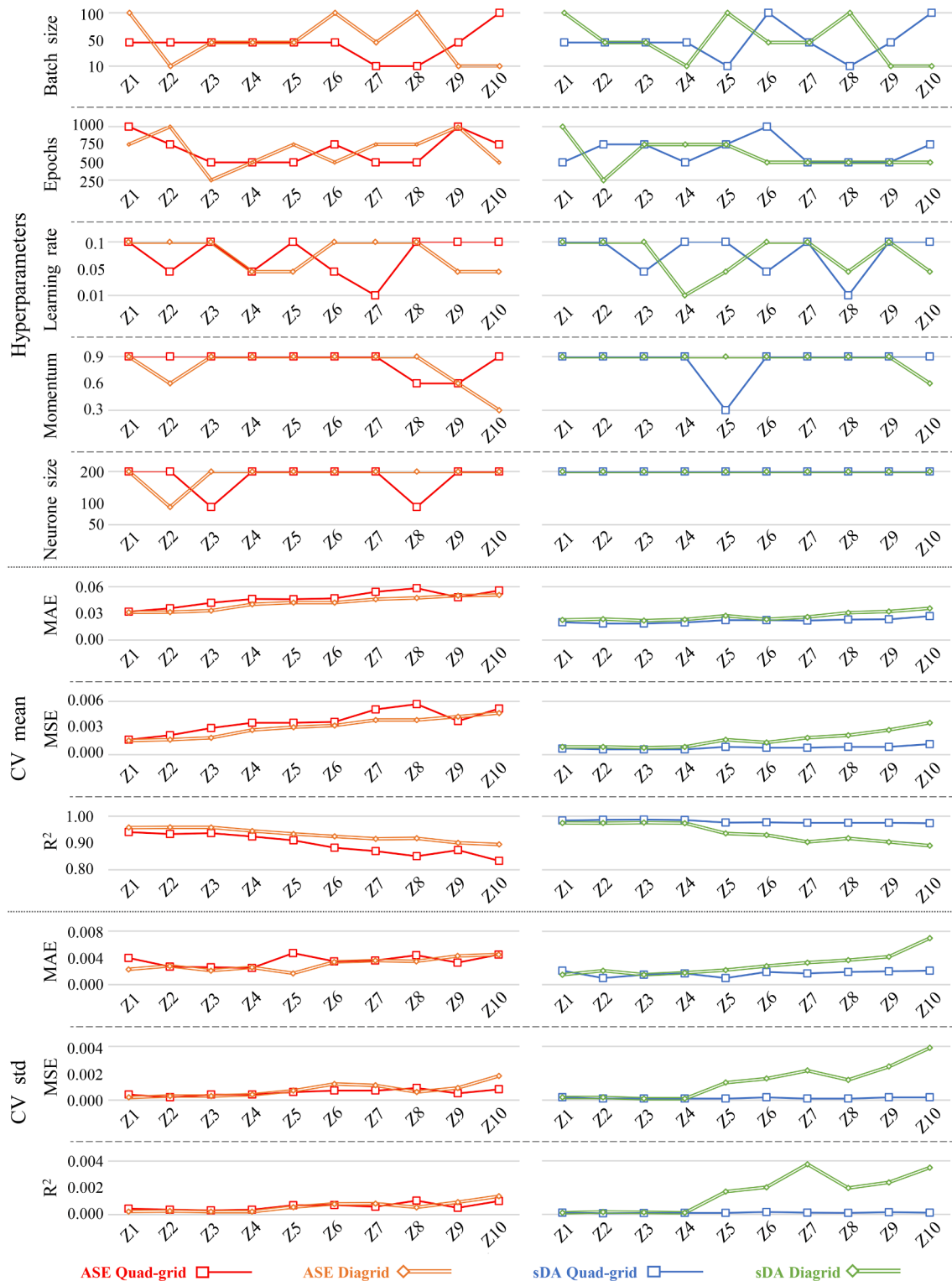
Fig. 12. Grid search results of the best ANN models.

accuracies, building zones close to the sky levels were more challenging than the ground levels because of the increasing number of design variables. The study also proved that dense urban surroundings affect the performance of high-rise buildings at various floor levels by determining different simulation results during the sampling process. Therefore, architects and engineers should consider various zones as different problems while designing sustainable high-rises in metropolises.

The ML part of the MUZO methodology indicated prediction scores with high accuracies using different hyperparameters for batch size, epoch, neurone size, momentum, and learning rate in each model despite various design complexities considering multiple performance aspects. Future research can integrate more hyperparameters, such as activation function, dropout rate, various optimisation algorithms, different numbers of hidden layers, and sample sizes. Thus, the ANN models can provide higher accuracies with an exponential increment in
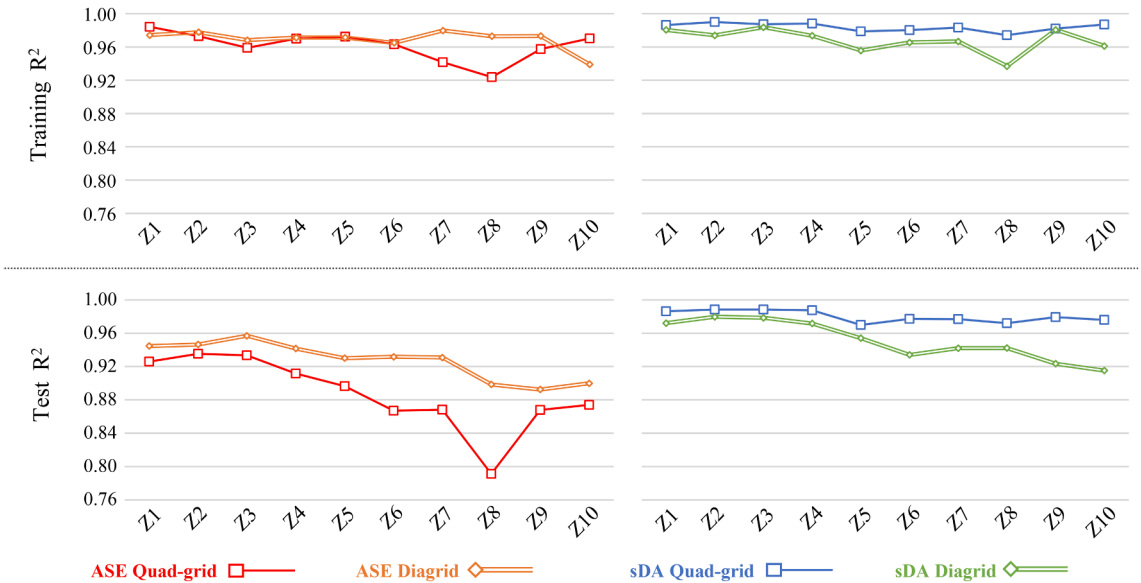
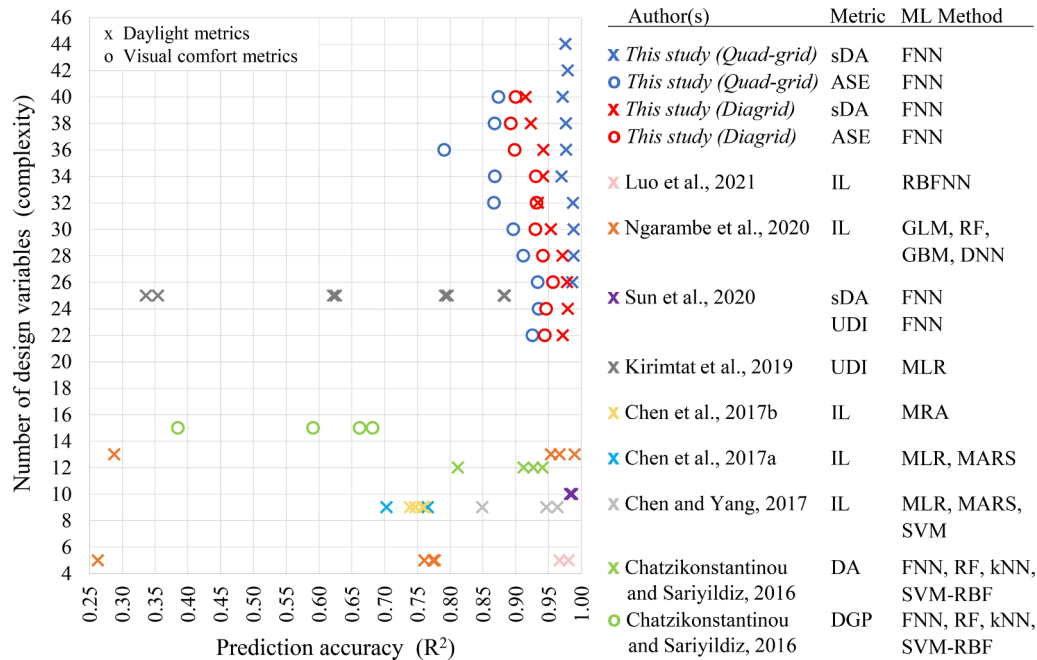**Fig. 13.** R^2 results of training and test sets.

Fig. 14. Summary of the previous studies for ML applications in daylight.

computational time. Hence, having all these parameters can be more applicable to real-world high-rise scenarios.

In conclusion, the parametric high-rise model and ML for surrogate model phases of the MUZO methodology could automate form generation, performance evaluation, sampling, data processing, ANN development, and reporting the predictive models for all zones in both high-rise scenarios. Using the ML part of the MUZO methodology, architects and engineers can address the computational burden while optimising the entirety of a high-rise building to propose sustainable alternatives in metropolises. Nevertheless, optimisation of the high-rise building, which is addressed in part 2 of this study, remains challenging owing to the high number of parameters involved in the design process.

7. Data availability

Datasets related to this article can be found at https://doi.org/10.4121/uuid:8538ac2f-3a78-4923-8fca-5beb50017299, an open-source online data repository hosted at 4TU Research Data (Ekici et al., 2020).
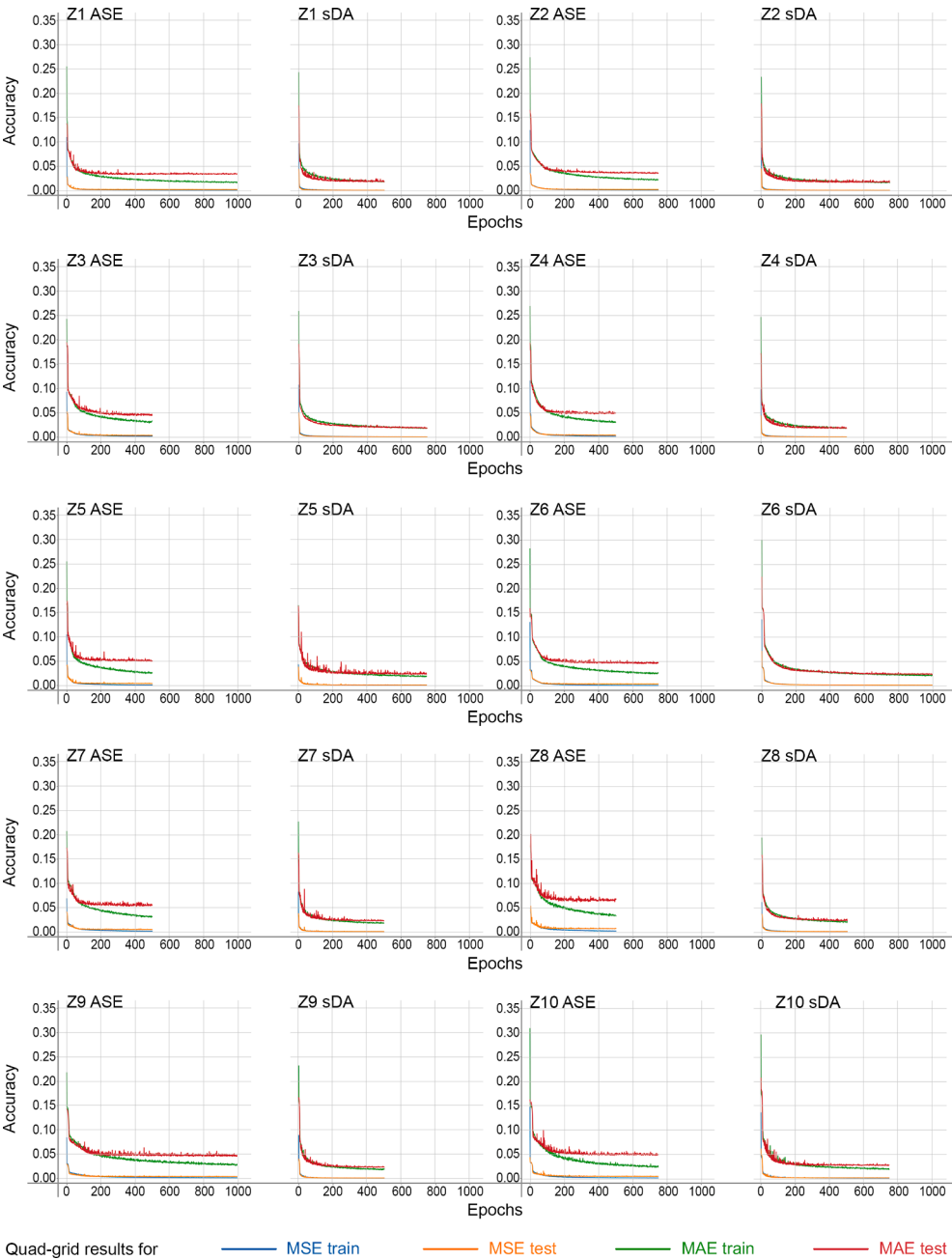
Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.
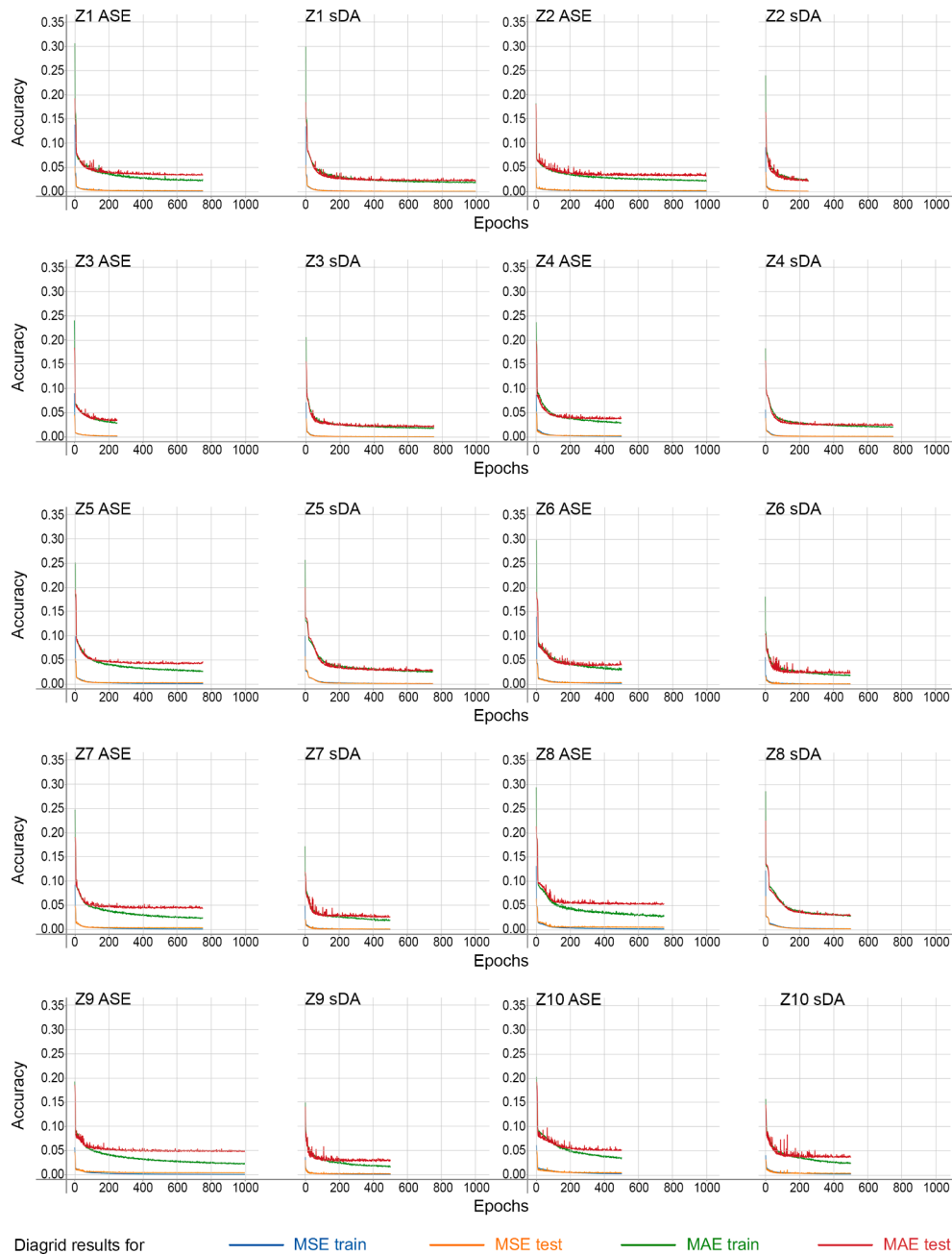
Acknowledgments

Appendix A. Quad-grid MAE and MSE results for training and test sets

Quad-grid results for — MSE train — MSE test — MAE train — MAE test

Appendix B. Diagrid MAE and MSE results for training and test sets

Diagrid results for ——— MSE train ——— MSE test ——— MAE train ——— MAE test

Appendix C. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.solener.2021.05.083.

References

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., 2016. Tensorflow: A system for large-scale machine learning. In: 12th {USENIX} symposium on operating systems design and implementation ({OSDI} 16). pp. 265–283.

Aish, R., 2003. Bentley's Generative Components: A Design Tool for Exploratory Architecture. Bentley Systems Inc.

Al-Kodmany, K., Ali, M.M., 2013. The Future of the City: Tall Buildings and Urban Design. WIT Press.

Asadi, S., Geem, Z.W., 2015. Sustainable building design: a review on recent metaheuristic methods. Recent Adv. Swarm Intell. Evolut. Comput. 203–223.

Bauer, C., Kazanasmaz, Z.T., Wittkopf, S., 2017. Climate based daylight simulations with EvalDRC–analysis of Daylight Redirecting Components. J. Facade Des. Eng. 5 (2), 45–57.

Bentley, 2003. Generative Components, https://www.bentley.com/en/products/product-line/modeling-and-visualization-software/generativecomponents, Last seen: 25/2/2021.

Bottou, L., 2010. Large-scale machine learning with stochastic gradient descent. In: Proceedings of COMPSTAT'2010. Springer, pp. 177–186.

Cevizci, E., Kutucu, S., Morales-Beltran, M., Ekici, B., Tasgetiren, M.F., 2019. Structural optimization for masonry shell design using multi-objective evolutionary algorithms. In: Optimization in Industry. Springer, pp. 85–119.

Chatzikonstantinou, I., Sariyildiz, I.S., 2017. Addressing design preferences via auto-associative connectionist models: application in sustainable architectural Façade design. Autom. Constr. 83, 108–120.

Chatzikonstantinou, I., Sariyildiz, S., 2016. Approximation of simulation-derived visual comfort indicators in office spaces: a comparative study in machine learning. Archit. Sci. Rev. 59 (4), 307–322.

Chen, X., Huang, J., Yang, H., Peng, J., 2019a. Approaching low-energy high-rise building by integrating passive architectural design with photovoltaic application. J. Cleaner Prod. 220, 313–330.

Chen, X., Huang, J., Zhang, W., Yang, H., 2019b. Exploring the optimization potential of thermal and power performance for a low-energy high-rise building. Energy Procedia 158, 2469–2474.

Chen, X., Yang, H., 2017. A multi-stage optimization of passively designed high-rise residential buildings in multiple building operation scenarios. Appl. Energy 206, 541–557.

Chen, X., Yang, H., 2018. Integrated energy performance optimization of a passively designed high-rise residential building in different climatic zones of China. Appl. Energy 215, 145–158.

Chen, X., Yang, H., Peng, J., 2019c. Energy optimization of high-rise commercial buildings integrated with photovoltaic facades in urban context. Energy 172, 1–17.

Chen, X., Yang, H., Sun, K., 2017a. Developing a meta-model for sensitivity analyses and prediction of building performance for passively designed high-rise residential buildings. Appl. Energy 194, 422–439.

Chen, X., Yang, H., Wang, Y., 2017b. Parametric study of passive design strategies for high-rise residential buildings in hot and humid climates: miscellaneous impact factors. Renew. Sustain. Energy Rev. 69, 442–460.

Chollet, F., 2015. keras, https://github.com/fchollet/keras. GitHub repository.

Cohen, B., 2006. Urbanization in developing countries: Current trends, future projections, and key challenges for sustainability. Technol. Soc. 28 (1–2), 63–80.

Crombecq, K., Dhaene, T., 2010. Generating sequential space-filling designs using genetic algorithms and monte carlo methods. In: Asia-Pacific Conference on Simulated Evolution and Learning. Springer, pp. 80–84.

CTBUH, 2020. The Skyscraper Center, Council on Tall Buildings and Urban Habitat, http://www.skyscrapercenter.com (Last Seen: 03.07.2020).

Cubukcuoglu, C., Ekici, B., Tasgetiren, M.F., Sariyildiz, S., 2019. OPTIMUS: self-adaptive differential evolution with ensemble of mutation strategies for grasshopper algorithmic modeling. Algorithms 12 (7), 141.

De Boer, P.-T., Kroese, D.P., Mannor, S., Rubinstein, R.Y., 2005. A tutorial on the cross-entropy method. Ann. Oper. Res. 134 (1), 19–67.

Del Ser, J., Osaba, E., Molina, D., Yang, X.-S., Salcedo-Sanz, S., Camacho, D., Das, S., Suganthan, P.N., Coello, C.A.C., Herrera, F., 2019. Bio-inspired computation: where we stand and what's next. Swarm Evol. Comput. 48, 220–250.

Draper, N.R., Smith, H., 1998. Applied Regression Analysis. John Wiley & Sons.

Dynamo, 2011. Dynamo: Open source graphical programming for design, https://dynamobim.org/, Last seen: 25/2/2021.

Ekici, B., Cubukcuoglu, C., Turrin, M., Sariyildiz, I.S., 2019a. Performative computational architecture using swarm and evolutionary optimisation: a review. Build. Environ. 147, 356–371.

Ekici, B., Kazanasmaz, T., Turrin, M., Tasgetiren, M.F., Sariyildiz, I.S., 2019b. A Methodology for daylight optimisation of high-rise buildings in the dense urban district using overhang length and glazing type variables with surrogate modelling. J. Phys.: Conf. Ser. IOP Publishing, 012133.

Ekici, B., Kazanasmaz, Z.T., Turrin, M., Tasgetiren, M.F., Sariyildiz, I.S., 2020. Multi-zone simulation results on ASE and sDA daylight metrics for parametric high-rise model with quad grid and diagrid facade in a highly dense hypothetical urban district using dry summer climate weather data https://doi.org/10.4121/uuid:8538ac2f-3a78-4923-8fca-5beb50017299, 4TU.ResearchData, 2020.

Evins, R., 2013. A review of computational optimisation methods applied to sustainable building design. Renew. Sustain. Energy Rev. 22, 230–245.

Gan, V.J., Wong, H., Tse, K.T., Cheng, J.C., Lo, I.M., Chan, C.M., 2019. Simulation-based evolutionary optimization for energy-efficient layout plan design of high-rise residential buildings. J. Cleaner Prod. 231, 1375–1388.

Giouri, E.D., Tenpierik, M., Turrin, M., 2020. Zero energy potential of a high-rise office building in a Mediterranean climate: using multi-objective optimization to understand the impact of design decisions towards zero-energy high-rise buildings. Energy Build. 209, 109666.

Godoy-Shimizu, D., Steadman, P., Hamilton, I., Donn, M., Evans, S., Moreno, G., Shayesteh, H., 2018. Energy use and height in office buildings. Build. Res. Inf. 46 (8), 845–863.

Goussous, J., Al-Refaie, A., 2014. Evaluation of a green building design using LCC and AHP techniques. Life Sci. J. 11 (8s), 29–40.

Grus, J., 2019. Data science from scratch: first principles with python. O'Reilly Media.

Hecht-Nielsen, R., 1992. Theory of the backpropagation neural network. In: Neural Networks for Perception. Elsevier, pp. 65–93.

Hunter, J.D., 2007. Matplotlib: A 2D graphics environment. Comput. Sci. Eng. 9 (3), 90–95.

IEA, 2020. Tracking Buildings 2020, IEA, Paris https://www.iea.org/reports/tracking-buildings-2020.

IES, 2013. Approved method: IES spatial Daylight autonomy (sDA) and annual sunlight exposure (ASE).

Imam, M., Kolarevic, B., 2016. Towards Resource-Generative Skyscrapers. Council on Tall Buildings and Urban Habitat research paper.

Jakubiec, J.A., Reinhart, C.F., 2011. DIVA 2.0: Integrating daylight and thermal simulations using Rhinoceros 3D, Daysim and EnergyPlus. In: Proceedings of building simulation. Citeseer, pp. 2202–2209.

Keough, I., 2011. Dynamo: designing a visual scripting interface for the Revit API (notes). https://github.com/ikeough/Dynamo/wiki.

Kim, Y., Chung, E.-S., Jun, S.-M., Kim, S.U., 2013. Prioritizing the best sites for treated wastewater instream use in an urban watershed using fuzzy TOPSIS. Resour. Conserv. Recycl. 73, 23–32.

Kingma, D.P., Ba, J., 2014. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

Kirimtat, A., Koyunbaba, B.K., Chatzikonstantinou, I., Sariyildiz, S., 2016. Review of simulation modeling for shading devices in buildings. Renew. Sustain. Energy Rev. 53, 23–49.

Kirimtat, A., Krejcar, O., Ekici, B., Tasgetiren, M.F., 2019. Multi-objective energy and daylight optimization of amorphous shading devices in buildings. Sol. Energy 185, 100–111.

Ko, D.H., Elnimeiri, M., Clark, R.J., 2008. Assessment and prediction of daylight performance in high-rise office buildings. Struct. Des. Tall Special Build. 17 (5), 953–976.

Korsavi, S.S., Zomorodian, Z.S., Tahsildoost, M., 2016. Visual comfort assessment of daylit and sunlit areas: a longitudinal field survey in classrooms in Kashan, Iran. Energy Build. 128, 305–318.

Kullback, S., Leibler, R.A., 1951. On information and sufficiency. Ann. Math. Stat. 22 (1), 79–86.

Lam, J.C., Hui, S.C., Chan, A.L., 1997. Regression analysis of high-rise fully air-conditioned office buildings. Energy Build. 26 (2), 189–197.

Lee, J., Boubekri, M., Liang, F., 2019. Impact of building design parameters on daylighting metrics using an analysis, prediction, and optimization approach based on statistical learning technique. Sustainability 11 (5), 1474.

Lee, M., Brauer, M., Wong, P., Tang, R., Tsui, T.H., Choi, C., Cheng, W., Lai, P.-C., Tian, L., Thach, T.-Q., 2017. Land use regression modelling of air pollution in high density high rise cities: a case study in Hong Kong. Sci. Total Environ. 592, 306–315.

Li, Y., Li, X., 2015. Natural ventilation potential of high-rise residential buildings in northern China using coupling thermal and airflow simulations. In: Building Simulation. Springer, pp. 51–64.

Liu, J., Wang, M., Peng, J., Chen, X., Cao, S., Yang, H., 2020. Techno-economic design optimization of hybrid renewable energy applications for high-rise residential buildings. Energy Convers. Manage. 213, 112868.

Loh, W.-L., 1996. On Latin hypercube sampling. Ann. Stat. 24 (5), 2058–2080.

Luo, Z., Sun, C., Dong, Q., Yu, J., 2021. An innovative shading controller for blinds in an open-plan office using machine learning. Build. Environ. 189, 107529.

Mangkuto, R.A., Feradi, F., Putra, R.E., Atmodipoero, R.T., Favero, F., 2018. Optimisation of daylight admission based on modifications of light shelf design parameters. J. Build. Eng. 18, 195–209.

McKinney, W., 2010. Data structures for statistical computing in python. In: Proceedings of the 9th Python in Science Conference. Austin, TX, pp. 51–56.

Mood, A.M., 1950. Introduction to the Theory of Statistics.

Mukkamala, M.C., Hein, M., 2017. Variants of rmsprop and adagrad with logarithmic regret bounds. arXiv preprint arXiv:1706.05507.

Nezamdoost, A., Van Den Wymelenberg, K., Mahic, A., 2018. Assessing the energy and daylighting impacts of human behavior with window shades, a life-cycle comparison of manual and automated blinds. Autom. Constr. 92, 133–150.

Nezamdoost, A., Van Den Wymelenberg, K.G., 2017. Revisiting the daylit area: Examining daylighting performance using subjective human evaluations and simulated compliance with the LEED version 4 daylight credit. Leukos 13 (2), 107–123.

Ngarambe, J., Irakoze, A., Yun, G.Y., Kim, G., 2020. Comparative performance of machine learning algorithms in the prediction of indoor daylight illuminances. Sustainability 12 (11), 4471.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., 2011. Scikit-learn: Machine learning in Python. J. Mach. Learn. Res. 12, 2825–2830.

Pilechiha, P., Mahdavinejad, M., Rahimian, F.P., Carnemolla, P., Seyedzadeh, S., 2020. Multi-objective optimisation framework for designing office windows: quality of view, daylight and energy efficiency. Appl. Energy 261, 114356.

Plotly, 2015. Plotly Technologies Inc., Collaborative data science, Montreal, QC, https://plotly.com/.

Riquelme, N., Von Lücken, C., Baran, B., 2015. Performance metrics in multi-objective optimization. In: 2015 Latin American Computing Conference (CLEI). IEEE, pp. 1–11.

Roman, N.D., Bre, F., Fachinotti, V.D., Lamberts, R., 2020. Application and characterization of metamodels based on artificial neural networks for building performance simulation: a systematic review. Energy Build. 109972.

Ruder, S., 2016. An overview of gradient descent optimization algorithms. arXiv preprint arXiv:1609.04747.

Rutten, D., 2015. Grasshopper3D, https://www.grasshopper3d.com/.

Samuelson, H., Claussnitzer, S., Goyal, A., Chen, Y., Romo-Castillo, A., 2016. Parametric energy simulation in early design: high-rise residential buildings in urban contexts. Build. Environ. 101, 19–31.

Sariyildiz, I., 2012. Performative computational design, Keynote speech. In: Proceedings of ICONARCH-I: International congress of architecture-I, Konya, Turkey, 15–17 November 2012. Selcuk University.

Sherif, A., Sabry, H., Wagdy, A., Mashaly, I., Arafa, R., 2016. Shaping the slats of hospital patient room window blinds for daylighting and external view under desert clear skies. Sol. Energy 133, 1–13.

Si, B., Wang, J., Yao, X., Shi, X., Jin, X., Zhou, X., 2019. Multi-objective optimization design of a complex building based on an artificial neural network and performance evaluation of algorithms. Adv. Eng. Inf. 40, 93–109.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., 2014. Dropout: a simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. 15 (1), 1929–1958.

Stone, M., 1974. Cross-validatory choice and assessment of statistical predictions. J. Roy. Stat. Soc.: Ser. B (Methodol.) 36 (2), 111–133.

Sun, C., Liu, Q., Han, Y., 2020. Many-objective optimization design of a public building for energy, daylighting and cost performance improvement. Appl. Sci. 10 (7), 2435.

Taccari, L., 2017. Joyplots, https://github.com/sbebo/joypy/blob/master/Joyplot.ipynb. Github repository.

Tian, Z., Wei, S., Shi, X., 2020. Developing data-driven models for energy-efficient heating design in office buildings. J. Build. Eng. 32, 101778.

UN, 2019. Department of Economic and Social Affairs, Population Division. World Urbanization Prospects: The 2018 Revision (ST/ESA/SER.A/420). United Nations, New York.

USGBC, 2014. LEED v4 for building design and construction. USGBC Inc.

Van Rossum, G.D., Fred L., 2009. Python 3 Reference Manual. CreateSpace, Scotts Valley, CA.

Vera, S., Uribe, D., Bustamante, W., Molina, G., 2017. Optimization of a fixed exterior complex fenestration system considering visual comfort and energy performance criteria. Build. Environ. 113, 163–174.

Wagdy, A., Elghazi, Y., Abdalwahab, S., Hassan, A., 2015. The balance between daylighting and thermal performance based on exploiting the kaleidocycle typology in hot arid climate of Aswan, Egypt. AEI 2015, 300–315.

Wagdy, A., Sherif, A., Sabry, H., Arafa, R., Mashaly, I., 2017. Daylighting simulation for the configuration of external sun-breakers on south oriented windows of hospital patient rooms under a clear desert sky. Sol. Energy 149, 164–175.

Westermann, P., Evins, R., 2019. Surrogate modelling for sustainable building design–A review. Energy Build. 198, 170–186.

Willmott, C.J., Matsuura, K., 2005. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. Climate Res. 30 (1), 79–82.

Wolpert, D.H., Macready, W.G., 1997. No free lunch theorems for optimization. IEEE Trans. Evol. Comput. 1 (1), 67–82.

Wortmann, T., 2017. Opossum-introducing and evaluating a model-based optimization tool for grasshopper.

Yi, Y.K., 2019. Building facade multi-objective optimization for daylight and aesthetical perception. Build. Environ. 156, 178–190.