Generation of a Large Synthetic Database of Office Tower's Energy Demand Using Simulation and Machine Learning

Ammar Alammar and Wassim Jabi

1 Introduction

A building's skin or façade is a key factor in determining the comfort and energy consumption of the building. This is because buildings are exposed to dynamic environmental factors such as solar radiation, temperature precipitation and wind and these outdoor conditions change continuously throughout the day and the year. Regardless of the outdoor climate which changes constantly, a building's skin has been typically designed as a static envelope. Fixed or static shading devices are limited in terms of their responsiveness to indoor or outdoor environmental conditions and this leads to unacceptable performance once these systems have been installed, especially if changes are required over time (Tabadkani et al., 2021). In addition, studies have shown that static facades are no-longer favourable and have limitations in terms of attaining the desired energy efficiency, adequate daylighting and control flexibility (Al-Masrani & Al-Obaidi, 2019). On the other hand, adaptive façades are capable of and effective in responding to variable climatic conditions. To that end, numerous studies have been conducted regarding substituting the static envelope with an adaptive one.

Different types of adaptive façades have been developed in terms of materials, components and systems and further developments are expected in the future (López et al., 2017). These adaptive façades have unique features or behaviours that repeat- edly and reversibly change over time according to variable boundary conditions and respond to changing performance requirements with the aim of improving the overall building performance (Loonen et al., 2014). To achieve a high-performance building with an adaptive façade is challenging due to the complexity of the system. The building envelope is influenced by a variety of physical domains (thermal, luminous, air quality, etc.) which make it difficult to accurately predict the performance of a building with an adaptive façade (Loonen et al., 2017); most building performance simulation (BPS) tools were not developed to predict building performance with an adaptive façade. Adaptive façade systems are difficult to predict due to their numerous functionalities and complexities; thus, the current paper presents a methodology for predicting adaptive façade energy performance early in the design process using machine learning (ML) approaches to overcome the limitations of BPS tools. ML-based approaches to evaluate building performance appear to be more efficient than conventional simulation-based approaches (Chakraborty & Elzarka, 2019). In various studies, decision trees (DTs) have been effectively used to assess a building's energy consumption. As a result, decision trees could be used to predict the performance of a building equipped with an adaptive façade if it is sufficiently trained with big data.

2 Predicting the Performance of Adaptive façade s with Current Tools

Innovative materials and technologies have been developed in the adaptive façade field and these can be enhanced with the use of building performance predictions. Additionally, in the design of adaptive façades it is crucial to predict building performance accurately to produce high performance buildings. According to Geyer and Singaravel (2018), evaluating the performance of adaptive façade systems during the early stage of the design is critical for determining its applicability. Loonen et al. (2017) claimed that it is a difficult task to predict the performance of buildings with adaptive façades because the system is mainly affected by the building's local boundary conditions, interactions with its occupants and other building systems. Additionally, the authors investigated simulation techniques for both static and adaptive building envelopes. The simulation process is easier in traditional static envelopes and requires certain input parameters such as the U-value and G-value in order to make predictions. The lack of available tools is a significant factor limiting studies into adaptive façade performance prediction. Loonen et al. (2017) mentioned that most software packages are described as complicated digital modelling and simulation processes and they are not user-friendly. In addition, most of the existing tools lack the ability to simulate adaptive façades within their built-in objects, apart from a few software packages that target specific types of technologies, such as thermochromic (TC) glazing technology. However, these systems experience only physiological changes, making it extremely difficult to integrate numerous variables that change over time (Sheikh & Asghar, 2019). According to Loonen et al. (2017), there are two main factors that determine the applicability of adaptive façades, as follows:

Modelling time-varying facade properties: Facade specifications (i.e., material properties or position of components) need to be changeable during simulation run-time to properly account for transient heat transfer and energy storage effects in building constructions (Loonen et al., 2014). Many state-of-the-art BPS tools have restricted functionalities for accomplishing this feature.

Modelling the dynamic operation of facade adaptation: During the operation of adaptive systems, the performance of the system is entirely dependent on the scheduling strategy (i.e., control logic) that is utilised to change the façade. Moloney (2011) states: "The design outcome in a project with kinetic facades is a process, rather than a static object or artifact."

3 Decision Tree (DT)

The decision tree is a technique that is frequently used in a wide variety of applications for classification and prediction purposes (Tung et al. 2005). A decision tree divides a set of data into multiple specified classes by employing a flowchart-like tree structure, hence providing a description, categorisation and generalisation of the given datasets (Yu et al., 2010a, 2010b). The decision tree model has a number of advantages over other models, including its simplicity of use and ability to predict with high accuracy without the need for extensive calculations. Several applications have incorporated decision tree approaches into building analysis studies (Ahmad et al., 2017). Tso and Yau (2006) compared three modelling techniques to estimate average weekly electricity energy consumption in Hong Kong (Tso & Yau, 2007). They discovered that both decision tree and ANN are more appropriate models than regression models due to their ability to analyse and predict energy consumption patterns. Haghighat et al. (2010) published another study in which they developed a prediction model to optimise building energy performance with the use of decision trees (). They used a decision tree to estimate energy use intensity (EUI) in a residential building, revealing that by utilising the decision tree approach, it is possible to precisely categorise and anticipate the energy consumption of a structure, resulting in a high-energy-performance construction.

A random forest is a collection of decision trees that can be used to make predictions. Breiman (2001) states that the random forest was initially developed as a technique for optimising the conventional decision tree method. Additionally, the author advised using a random forest as a predictive regression method. Ahmad et al. (2017) used a random forest algorithm to predict the hourly electricity usage of HVAC systems at a hotel in Madrid. Ma and Cheng (2016) employed random forests to determine the relevant levels of 171 characteristics associated with residential building regional energy consumption intensity.

4 Methods

The current study's methodology comprised three main stages (see Fig. 1). The initial phase involved creating a generative parametric simulation of office spaces equipped with an adaptive façade shading system using EnergyPlus which also functioned as the training data. In order to conduct a thorough energy analysis, the simulation settings, input parameters, material properties, occupant loads, zone programme, occupancy schedule, and thermal settings were determined. The second phase included the implementation of an automatic control system to activate the system on an hourly basis in response to two specified environmental sensors. The aim of this process was to develop a synthetic database of hourly cooling energy consumption (Wh/m^2) for use as training data. The final phase involved developing and validating a decision tree surrogate model to estimate the hourly cooling demand of an adaptive façade system in a closed office environment. For training and testing, the synthetic datasets from the simulation were imported into the decision tree model. Then a hybrid parameter tweaking approach was used to achieve the most accurate model.

4.1 Simulation Setup

As a case study, a typical mid-rise office building was developed which is located in central Riyadh, Saudi Arabia. The office building is 30 storeys and has a height of 120 m which is representative of the common height scenario found in the centre of King Abdullah Financial District (see Fig. 2a). All floors of the building have the same layout and core area measurements: 35 m * 35 m, giving a total floor space of 1225 m^2. Specifically, only shared side-lit office zones with an adaptive shading system were investigated on each floor of the proposed office building which faces the primary orientations (north, south, east and west) in order to evaluate the influence of an adaptive façade on energy performance (see Fig. 2b). Table 1 presents the spatial dimensions and characteristics of the office room.

The building context varied in each simulation (low, medium and high) to test energy loads at each level and in all the main orientations (north, south, east and west). The variation in heights of the surrounding contexts acted as one of the main features of geometric variation in the study. In addition, the average height of the surrounding buildings was used parametrically to control the vertical location of the office room in each orientation, in accordance with a lower-than-average, average, and higher-than-average height setting. This was intended to simulate the varying amounts of sunlight and daylight that the offices in a building receive. The generative parametric office tower rules and its urban context is illustrated in Fig. 3, which shows how the model were set up parametrically (generative parametric tower). The vertical location of the office is calculated using the following formula:

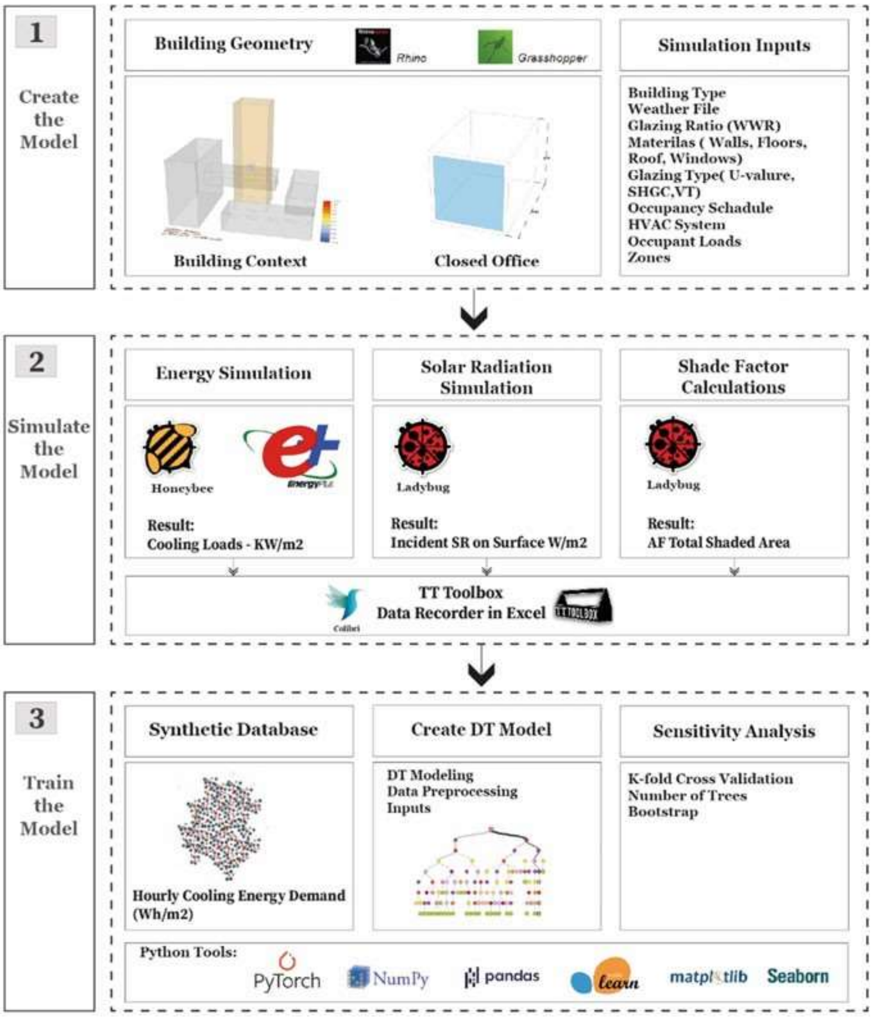

Fig. 1 Framework of the study

$$a = \sum B00 + B01 + B02 + B03/n$$

$$l = (a) * 0.50$$

$$h = (a) * 1.50$$

where a = *average,* l = *lower than average,* h = *higher than average*
B = *building context,* n = *number of variables.*

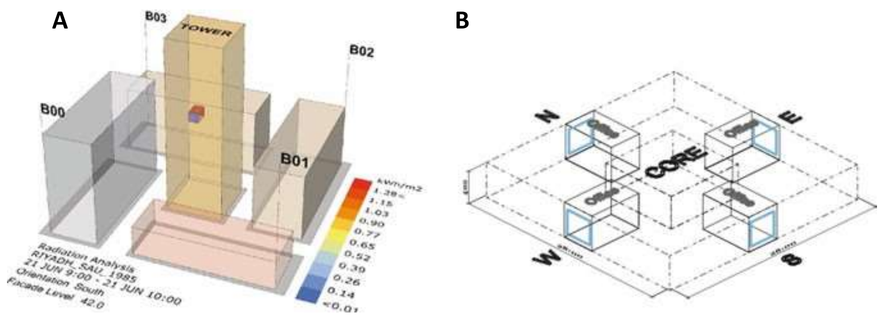

Fig. 2 **a** The 3D parametric urban context, that change parametrically in each simulation, **b** A single closed office room facing main orientations

Table 1 Spatial dimensions and characteristics of simulation inputs

Parameter	Assigned value(s)
Location	Riyadh, Saudi Arabia
Space type	Shared Office Room
Zone program	Closed Office Zone
Glazing ratio	80%
Room width	4.00 m
Room floor height	4.00 m
Room length	6.00 m
Shading reflectance	70%
Interior wall, ceiling, floor	Adiabatic
Cooling set points	24 C
Heating set points	22 C
HVAC system	ideal air load system
Number of people	2 people
Zone loads Lighting density	3 W/m^2
Number of occupants	0.5 ppl/m^2
Equipment load (W/m^2)	2 W/m^2
Infiltration ratio	0.04 cfm/sf (~0.000203 m^3/s m^2 façade)
Schedule	Sun.-Thur. 08:00 – 18:00
Prototype unit	0.80 * 0.80 c
Shadow calculation method	Time step frequency
Solar radiation sensor point 1 (P1)	3.00 m height
Operative temperature point 2 (P2)	1.5 m height

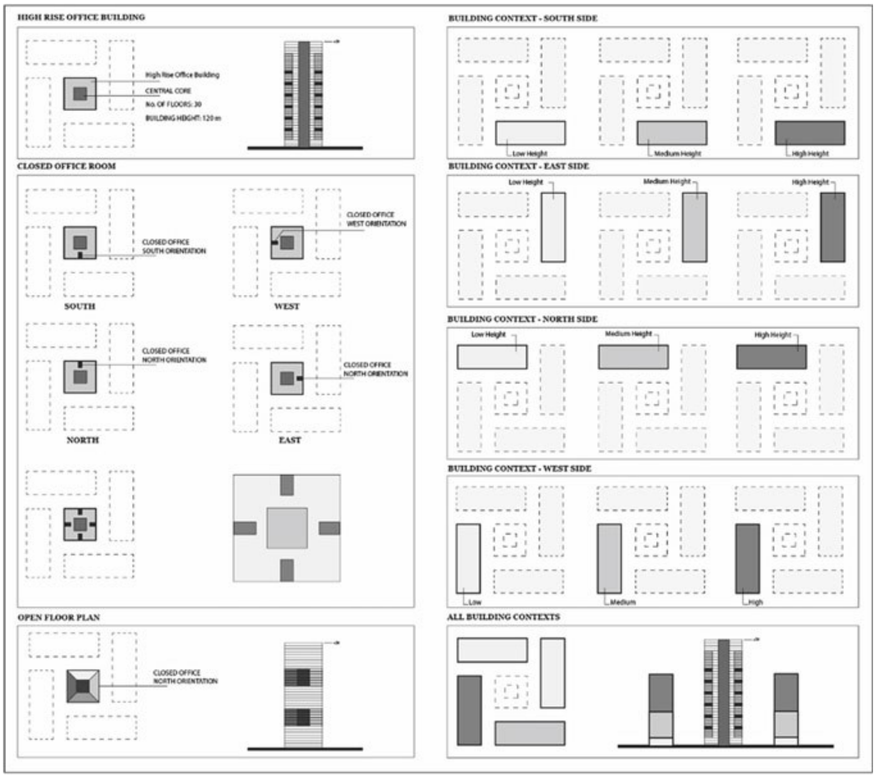

Fig. 3 Parametric model generation rules

4.2 *Validation of the Base Model*

Validation is an essential step in the simulation process to ensure that the hypothetical model is correct and represents the real environment. One of the most definitive validation methods is to compare model results with existing cases. In this study, examples of case studies are used to support the validity of the base model. The base model was established based on existing case studies and design guidelines for office buildings which exemplify the typical characteristics of office buildings. In addition, the simulation settings and parameters were determined using ASHRAE benchmark. The annual energy consumption for the base model of the study which does not integrate an adaptive façade system shows coherence with the five case studies examined. Fasiuddin and Budaiwi (2011) conducted a study on five commercial buildings in Saudi Arabia and provided detailed statistics for various factors, among which was annual energy consumption data. The annual energy consumption data for these case studies were obtained from the utility bills provided by the management or by the Saudi Electric Company (SEC) (Fasiuddin & Budaiwi, 2011).

	Case A	Case B	Case C	Case D	Case E
Gross annual consumption (kWh)	31,452,70	10,056,92	3857,49	2799,05	2107,93
Annual consumption (kWh/m^2)	273.5	267.8	275.5	249.9	263.5

Based on the above cases, the average annual consumption per unit area (m^2) for all five buildings is approximately 266 kWh/m^2/year. The developed base model was validated by comparing the simulation energy results with the annual energy consumption of the above reviewed case studies. The EUI of 266 kWh/m^2/year serves as a typical annual consumption for commercial buildings in the studied region and provides the basis for comparison. Thus, an average annual consumption per square meter less than or equal to the above value is considered to be performing within the normal range of energy consumption. However, this value could be optimised when implementing the adaptive façade system. For the studied base case, the annual energy consumption result generated from the simulation averages 232.7 kWh/m^2/year (base case-south orientation = 263.2 kWh/m^2/year, base case-west orientation = 270.9 kWh/m^2/year, base case-north orientation = 184.6 kWh/m^2/year, and base case-east orientation = 212 kWh/m^2/year). The values of 232.7 KWh/m^2 and 266 KWh/m^2 are not significantly different and the difference between the energy consumption values in the case studies and the base model is due to the use of different settings. For example, the lighting density was higher in the case studies compared to the base case model. The model of the study considers 3W/m^2 to be the average value of the considered benchmarks, whereas the case studies consider a value of 13W/m^2. In addition, the case studies use a range of 8 to 11 W/m^2 for equipment power density, whereas the base model considers 2W/m^2 to be the average benchmark value (ASHRE, 2010). This comparison verifies the capability of the study's base model and the validity of the simulation outcomes.

4.3 Inputs Parameters

The dynamic input parameters were based on a variety of factors such as the building's exterior wall U-values, the U-values of different types of glazing and the adaptive façade shading system's dynamic behaviour changes, including the hour, date, month, orientation, building context, solar radiation (SR), operative temperature, shade factor (SF), and opening ratio. A dynamic shading system integrated into the building exterior has the potential to significantly reduce energy consumption. As a result, certain variables were examined to determine whether an adaptive façade system could improve a building's energy efficiency. The dynamic input parameters used in the current study to conduct the energy analysis are listed in Table 2.

Table 2 Dynamic simulation inputs

Dynamic input parameter	Assigned value(s)	No. of iterations
Orientation	South, west, north, east	4
Building context 00	Low, medium, high	3
Building context 01	Low, medium, high	3
Façade level height	Lower than average, average, and higher than average	3
Exterior wall – U-value	0, 1, 3 W/m^2 K	3
Glazing type – U-value	0, 1, 2, 3 W/m^2 K	4
Total no. of iterations		1,296
Month	March, June, September, December	4
Day	01– 31	31
Hour	1:00–24:00	10
Shading states	A, B, C, D, E, F	6
Total no. of hourly cooling data		1,581,120

4.4 Construction Materials

The EnergyPlus recommended database (ASHRAE materials) was used to define the office's material characteristics selecting ASHRAE 90.1–2010 climate region number 1 which was assigned to a hot, dry region. Various types of external walls with different U-values were investigated for the office room. The interior walls were made of gypsum board with a U-value of 2.58 which indicates that no heat is transferred across these partition walls. The model's energy simulation also considered the effect of the glazing. Thus, different glazing systems (single, double and triple glazing) were explored for the studied model which has a variety of solar heat coefficients and thermal transmittance U-values (Gadelhak & Lang, 2016). The specifications for the building material parameters used in this investigation are listed in Table 3.

A closed office programme zone in accordance with the EnergyPlus (US Department of Energy's (DOE) office building zones (DOE 2016) was selected in all offices. These closed offices are conditioned with the default set at an ideal air load system in EnergyPlus for HVAC. In addition, the model was set for an hourly time step to calculate the energy demand of the office room. The operating time for cooling and heating was assumed to be five days per week (Monday to Friday from 7:00 to 18:00). The temperature setpoints of the HVAC system were considered to be 24 °C for cooling and 22 °C for heating. The HVAC system was set to work automatically to maintain the desired internal temperature.

Glazing type (Glaz)	U-value (W/m^2 K)	Solar heat gain coefficient (SHGC)	Visual transmittance (τvis)
Single glazing (SG)	5.82	0.82	0.88

(continued)

(continued)

Glazing type (Glaz)	U-value (W/m^2 K)	Solar heat gain coefficient (SHGC)	Visual transmittance (τvis)
Double glazing—clear (DG)	2.71	0.72	0.80
Double glazing—low-e coating (DG)	1.63	0.28	0.65
Triple glazing—Krypton filled (TG)	0.57	0.23	0.47

4.5 Automatic Control Logic

Existing BPS tools do not fully enable the adaptive behaviour of a façade to be simulated and there is a lack of a widely accepted approaches for designers to utilise when developing a control logic to test the system early in the design process. As a result of these constraints, the study developed a control scheme using the Energy Management System (EMS), an embedded feature of EnergyPlus that allows for the definition of sensors, controllers and actuators on hourly time steps (Hong and Lin 2013) (see Fig. 4). The external adaptive façade shading system is controlled by two outdoor and indoor sensors. Different shading states were designed that vary hourly and the shade factor of each shading state was calculated annually to be translated as a transmittance schedule, as well as calculating the annual incident solar radiation on the outside surface.

Sensors such as solar radiation and operative temperature (OT) were used in a closed (feedback) loop control system to adjust the opening ratio of the adaptive façade system. The first sensor point (P1) was placed at the corner of the outside wall to collect incident solar radiation on the surface, while the second sensor point (P2) was placed in the centre of the room at a height of 1.5 m to record the room air temperature. Regardless of how complicated the system is, the shade factor strategy was examined. As a result, a simple parametric unit shaped as a kinetic prismatic modular element was designed for the purpose of the current study with scaling and translating movements. Six different shading states were developed based on solar radiation and operative temperature thresholds: State-A 100%, State-B 80%, State-C 60%, State-D 40%, State-E 20%, and State-F 0% (see Fig. 5). When the external total solar radiation on the exterior surface and operative temperature surpassed the predetermined threshold, the shade system closed. The solar radiation range was 0-450W with a 50W step, whereas the operative temperature range was 21–24 °C. These criteria were established in accordance with several previous research studies that recommended an activation threshold that was appropriate for each climate zone (Touma and Ouahrani 2017; Yun et al. 2017; Tabadkani et al. 2020b). To accomplish this, an EMS conditional statement was coded to alter the required opening ratio based on the given program logic (see Fig. 6). When the solar radiation is equal to or

Table 3 Characteristics of materials used in the simulation

Name of material	Thickness (m)	Layers	U-value (W/m^2 K)	R-value (K m^2/W)
ASHRAE 90.1–2010 EXTWALLMASS CLIMATEZONE 1	0.2412	1IN Stucco 8IN CONCRETE HW RefBldg 1/2IN Gypsum	3.690821	0.270942
ASHRAE 90.1–2010 EXTWALL MASS CLIMATEZONE ALT-RES 1	0.277737	1IN Stucco 8IN CONCRETE HW RefBldg Mass Wall Insulation R-4.23 IP 1/2IN Gypsum	0.983672	1.016599
ASHRAE 90.1–2010 EXTWALL METAL CLIMATEZONE 1–2	0.154367	Metal Siding Metal Building Wall Insulation R-9.45 IP 1/2IN Gypsum	0.573406	1.743964
ASHRAE 90.1–2010 INTERIOR WALL	0.188	G01a 19 mm gypsum board F04 Wall air space resistance G01a 19 mm gypsum board	2.580645	0.3875
ASHRAE 90.1–2010 INTERIOR FLOOR	0.7291	F16 Acoustic tile F05 Ceiling air space resistance M11 100 mm lightweight concrete	1.449209	0.690031
ASHRAE 90.1–2010 INTERIOR CEILING	0.3007	M11 100 mm lightweight concrete F05 Ceiling air space resistance F16 Acoustic tile	1.449209	0.690031

Fig. 4 Energy management system (EMS) principles

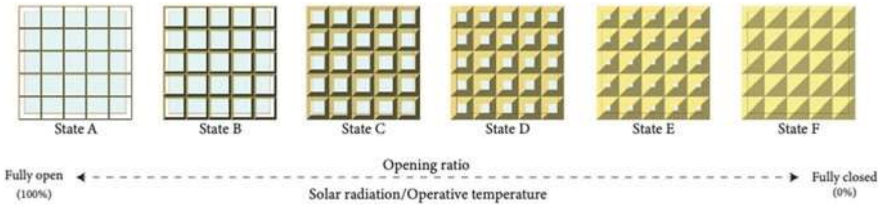

Fig. 5 Variation of the adaptive façade based on solar radiation and operative temperature

less than 50W and the operative temperature is equal to or less than 21 °C, the shade (State A) is completely open. Shade (State F) is completely closed when the solar radiation is equal to or greater than 450W and the operative temperature is greater than 24 °C. The additional shading states that fall between these two criteria were also studied (see Fig. 7).

The shade factor is calculated in response to the hourly changes in the opening ratio of each shading state. Because the grid method is similar to the ray-tracing method found within the Ladybug plug-in which is linked to the Radiance software, this method was adopted in the current research to calculate the shaded area of all of the six proposed shading states for a total of 8,760 h of the year. The shade factor ranges from 0 to 1 based on the percentage openness of the shading system, sun angle and sun position. As an example of this, Fig. 8 shows samples of different shading states when the shading system is 80 and 40% open with the grid method for shade factor calculations. To this end, the calculated shaded area of the distinct shading states was then translated into the transmittance schedule and called out within the EMS interface to select the state based on the defined threshold.

5 Modelling with Random Forest (RF)

In the current study, 1,296 simulation iterations of an office room with adaptive façade were generated using a variety of inputs to train the model. The total data collected included 3,794,688 (1296 * 4 months (122 days) *24 h) hourly cooling energy data. The cooling loads in KW/m^2 are the output by the decision tree model and thirteen variables were used as inputs: month, hour, day, orientation, building 00, building 01, façade level height, glazing type U-value W/m^{2}K, exterior wall U-value W/m^{2}K, adaptive façade opening ratio, adaptive façade—shade factor, solar radiation W/m^2, and operative temperature. Table 4 illustrates the characteristics of the input data and the ranges of each input. The generated energy results database was uploaded to the Design Explorer webpage which is a web-based tool allowing comparison analysis between the studied input parameters (see Fig. 9). To construct the decision tree model, three main steps need to be considered: (1) data pre-processing; (2) model training and hyper-parameter optimisation; and (3) model validation (Westermann & Evins, 2019).

```
EnergyManagementSystem:Sensor,
   S1,                    !- Name
   StateA,      !- Output:Variable or Output:Meter Index Key Name
   Schedule Value;  !- Output:Variable or Output:Meter Name

EnergyManagementSystem:Sensor,
   S2,                    !- Name
   StateB,      !- Output:Variable or Output:Meter Index Key Name
   Schedule Value;  !- Output:Variable or Output:Meter Name

EnergyManagementSystem:Sensor,
   S3,                    !- Name
   StateC,      !- Output:Variable or Output:Meter Index Key Name
   Schedule Value;  !- Output:Variable or Output:Meter Name

EnergyManagementSystem:Sensor,
   S4,                    !- Name
   StateD,      !- Output:Variable or Output:Meter Index Key Name
   Schedule Value;  !- Output:Variable or Output:Meter Name

EnergyManagementSystem:Sensor,
   S5,                    !- Name
   StateE,      !- Output:Variable or Output:Meter Index Key Name
   Schedule Value;  !- Output:Variable or Output:Meter Name

EnergyManagementSystem:Sensor,
   S6,                    !- Name
   StateF,      !- Output:Variable or Output:Meter Index Key Name
   Schedule Value;  !- Output:Variable or Output:Meter Name

EnergyManagementSystem:Sensor,
   S7,                    !- Name
   ZZZ,      !- Output:Variable or Output:Meter Index Key Name
   Surface Outside Face Incident Solar Radiation Rate per Area;  !- Ou


EnergyManagementSystem:Sensor,
   S8,                    !- Name
   YYY,      !- Output:Variable or Output:Meter Index Key Name
   Zone Operative Temperature ;  !- Output:Variable or Output:Meter

EnergyManagementSystem:Actuator,
   myA1,                  !- Name
   TrNS-SHD,                   !- Actuated Component Unique Name
   Schedule:Year,          !- Actuated Component Type
   Schedule Value;          !- Actuated Component Control Type

EnergyManagementSystem:ProgramCallingManager,
   MyComputedTransProg,      !- Name
   BeginTimestepBeforePredictor,  !- EnergyPlus Model Calling Point
   MyComputedTransSch;      !- Program Name 1

EnergyManagementSystem:Program,
   MyComputedTransSch,      !- Name
   Set StateA = S1,!- Program Line 1
Set StateB = S2,!- Program Line 1
Set StateC = S3,!- Program Line 1
Set StateD = S4,!- Program Line 1
Set StateE = S5,!- Program Line 1
Set StateF = S6,!- Program Line 1
SET SOL = S7,!- Program Line 2,
SET OT = S8,
   IF (SOL <= 50) && (OT < 21),
SET myA1 = StateA,
   ELSEIF (SOL <= 50) && (OT > 24),
SET myA1 = StateB,
   ELSEIF (SOL <= 50) && (OT < 24) && (OT > 21),
SET myA1 = StateA,
   ELSEIF (SOL > 50) && (SOL <=100) && (OT < 21),
SET myA1 = StateB,
   ELSEIF (SOL > 50) && (SOL <=100) && (OT > 24),
SET myA1 = StateC,
   ELSEIF (SOL > 50) && (SOL <=100) && (OT < 24) && (OT > 21),
SET myA1 = StateB,
   ELSEIF (SOL > 100) && (SOL <= 200) && (OT < 21),
SET myA1 = StateC,
   ELSEIF (SOL > 100) && (SOL <= 200) && (OT > 24),
SET myA1 = StateD,
   ELSEIF (SOL > 100) && (SOL <= 200) && (OT < 24) && (OT > 21),
SET myA1 = StateC,
   ELSEIF (SOL > 200) && (SOL <= 300) && (OT < 21),
SET myA1 = StateD,
   ELSEIF (SOL > 200) && (SOL <= 300) && (OT > 24),
SET myA1 = StateE,
   ELSEIF (SOL > 200) && (SOL <= 300) && (OT < 24) && (OT > 21),
SET myA1 = StateD,
   ELSEIF (SOL > 300) && (SOL <= 350) && (OT < 21),
SET myA1 = StateE,
   ELSEIF (SOL > 300) && (SOL <= 350) && (OT > 24),
SET myA1 = StateF,
   ELSEIF (SOL > 300) && (SOL <= 350) && (OT < 24) && (OT > 21),
SET myA1 = StateE,
   ELSEIF (SOL > 350) && (SOL <= 400) && (OT < 21),
SET myA1 = StateF,
   ELSEIF (SOL > 350) && (SOL <= 400) && (OT > 24),
SET myA1 = StateF,
   ELSEIF (SOL > 350) && (SOL <= 400) && (OT < 24) && (OT > 21),
SET myA1 = StateF,
   ELSEIF (SOL > 450),
SET myA1 = StateF,
   ENDIF;

EnergyManagementSystem:GlobalVariable,
   myglobeA1;               !- Erl Variable 1 Name

EnergyManagementSystem:OutputVariable,
   Weighted Shade Fraction Schedule,  !- Name
   myglobeA1,               !- EMS Variable Name
   Averaged,              !- Type of Data in Variable
   SystemTimestep;          !- Update Frequency
```

Fig. 6 The conditional statement coded within EMS

A random forest regressor is an algorithm based on decision tree s. decision trees are tree like structures where at each node an 'if–then-else' decision is taken about the value of an input. Based on the outcome of this decision, the tree can split into different branches. At each node, a similar decision as explained above is taken with respect to an input. The process is continued until the leaf nodes where the output is established out. A random forest regressor is a collection of such decision trees where each of the trees take a random subset of inputs to learn the 'if–then else' rules. In our experiments, random forest modelling is employed because most of the inputs are categorical and a few are continuous. Because decision trees are highly

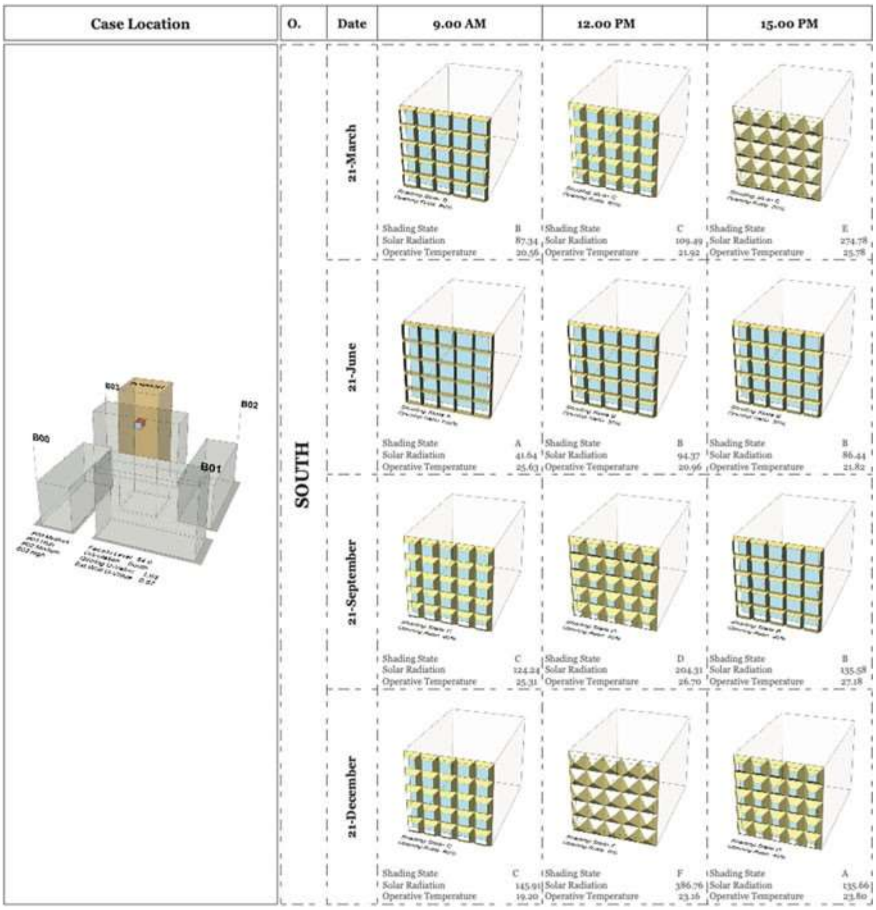

Fig. 7 Sample of hourly shading state variations chosen for three hours (9:00AM, 12:00PM, 15:00PM)

suitable for categorical as well as continuous inputs in their learning setting, random forest modelling offers a significant advantage.

There are a variety of performance indicators to quantify the model's accuracy and determine the model's precision. Based on literature, Some of the common used evaluation metrics are the root mean square error (RMSE), R-squared (R2), and mean absolute error (MAE), which used to measure the network's performance (Amasyali and El-Gohary 2018). The following formulae are used to calculate the performance measures:

$$\mathrm{RMSE} = \sqrt{\frac{1}{|y|}\sum_{i=1}^{n}\left(y_i - \hat{y}(i)\right)^2}$$

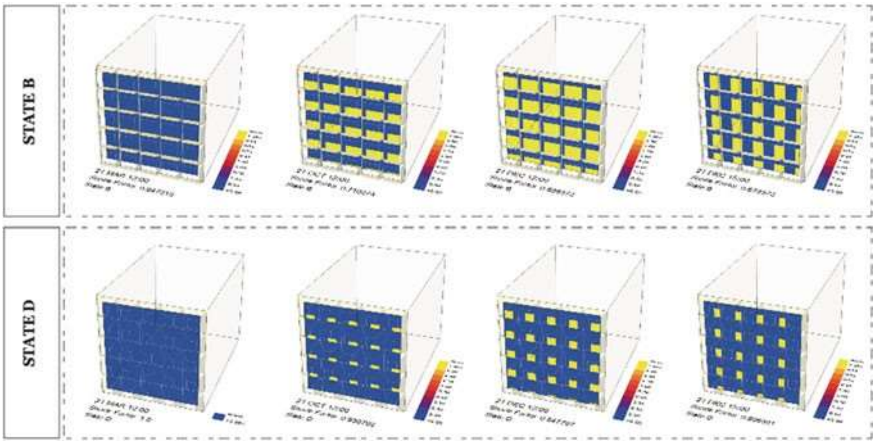

Fig. 8 Shade factor calculation for some selected cases

Table 4 The input data used for the ANN modelling

Input parameter	Input neuron type	Data range
Hours	Discrete	(1, 2,…, 24)
Month	Discrete	(March, June, September, December)
Day	Discrete	(1,2,…0.31)
Orientation	Discrete	(0,1,2,3) – (South, West, North, East)
Building 00	Discrete	(0,1,2) – (Low, Medium, High)
Building 01	Discrete	(0,1,2) – (Low, Medium, High)
Facade Level Height	Continuous	8–60
Glazing Type—U-value W/m^{2}K	Discrete	(0,1,2,3) – (SingleG0, DBL-Glz001-Clear, DBL-Glz002-low-e coating, TripleGlz-Krypton Filled)
Exterior Wall—U-value W/m^{2}K	Discrete	(0,1,2)
Adaptive Façade—Opening Ratio	Continuous	0–1
Adaptive Façade -Shade Factor	Continuous	0–1
Solar Radiation—W/m^2	Continuous	0 – 400
Operative Temperature—C	Continuous	14.00 – 30.00

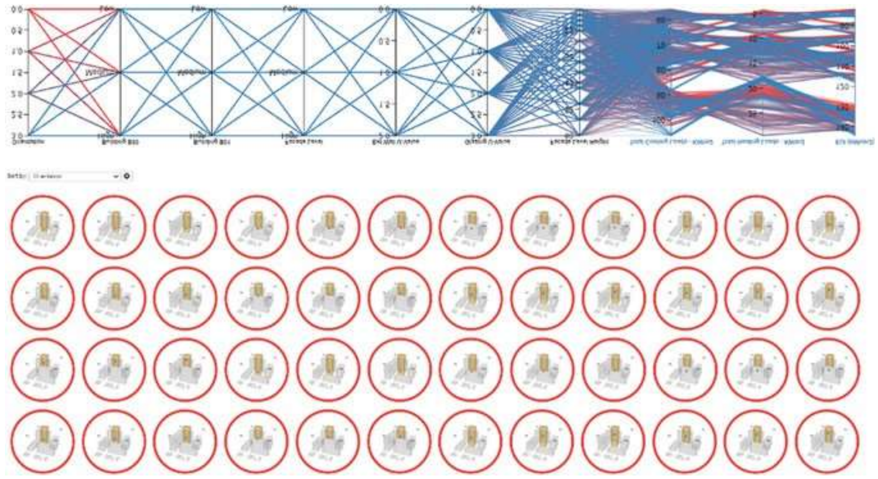

Fig. 9 Web-based of all generated cases of cooling loads (KWh/m^2) (http://ttacm.github.io/DesignExplorer/?ID=BL_3fQFpeE)

$$\mathrm{MAE} = (y, \hat{y}) = \frac{1}{|y|} \sum_{i=1}^{|y|} |y(i) - \hat{y}(i)|$$

$$R^2(y, \hat{y}) = 1 - \frac{\sum_{i=1}^{|y|} (y(i) - \hat{y}(i))^2}{\sum_{i=1}^{|y|} (y(i) - \overline{y})^2}$$

According to the above equation, y_i represents the output of the i-th data point, y(i) is the predicted value of the i-th data point, and y represents the mean of y. Here, yˆ(.) is the function approximated by the decision tree, and |y| represents the number of prediction cases. The RMSE value is the square of the difference between the actual and predicted cooling load levels. The MAE, on the other hand, defines the absolute value of the difference between the two. The distinction is often referred to as the residual. Both values can be any positive integer greater than zero and a model is said to function well if both values are as small as possible. The R^2 value is used to determine the scatter of predicted values around the regression line. It is often referred to as the coefficient of determination in statistics. IIt is defined as the ratio of variance explained by the model to the total variance or as the following equation:

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}}$$

where SS_{res} is the sum of the squares of the residuals and SS_{res} is the total sum of the squares which is a measure of variance in the data. For an optimal model, SS_{res} will be exactly equal to SS_{res} and, hence, the R^2-value is 1. Additionally, the R^2-value varies between 0 and 1 and the greater its value, the more accurately the model is able to predict.

5.1 Data Pre-Processing

The operations required to process categorical data inputs are referred to as 'data pre-processing.' Rather than using categorical inputs directly, they must be given an appropriate mathematical representation to improve the network's performance. The following data are used as discrete inputs in the current study: hours, month, date, direction, building 00, building 01, glazing type U-value, and external wall U-value. One example of hot encoding (Seger, 2018) is the pre-processing of categorical inputs. One-hot encoding is a mathematical technique for numerically describing categorical variables as a vector of zeros and ones. The vector's dimension will be equal to the number of possible values. One is assigned to the coordinate that corresponds to the value of the variable, while the remaining coordinates are set to zero. For instance, if a variable accepts values of high, medium or low, high is represented by [1,0,0], medium by [0,1,0], and low by [0,0,1].

5.2 K-fold Cross Validation

The purpose of k-fold cross validation is to choose a suitable combination of hyper-parameters such as the number of trees and bootstrap. To tune these hyper-parameters, a grid search with a k-fold cross-validation experiment is undertaken. In k-fold cross validation, the data are split into k folds. Among these k number of folds, (k-1) folds belong to the training data and the remainder to the testing data. The experiments are performed k times and each time the testing fold varies without repetition. The benefit is that each fold in one or other experiment becomes part of the training as well as the testing. Hence, the bias that can result from the binary splits is avoided. Initially the whole dataset is split into training, validation and testing sets. 80% of the data are assigned to the training set, 6.67% to the validation set and the remaining 13.37% are assigned to the testing set. The k-fold cross validation is then undertaken on the training fold. The value of k chosen is 5. The data split procedure is graphically represented in Fig. 10. Note that when undertaking k-fold cross validation, one among the fold becomes the testing set and remainder becomes the training set. In this case, one-third of the testing case will be reserved as a validation set for that particular instance of the validation procedure.

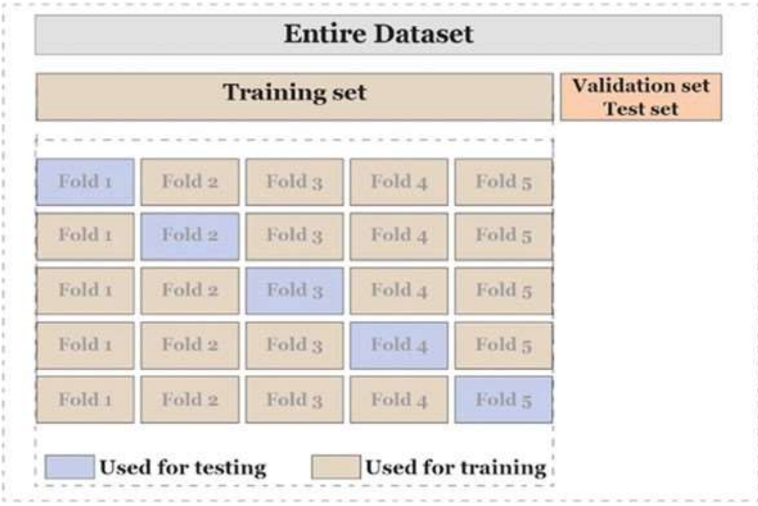

Fig. 10 The data split procedure

5.3 Optimization of Random Forest

Numerous hyper-parameters are required for random forest modelling, as follows: (1) Number of trees: The number of trees represents the total number of decision trees utilised in the random forest algorithm. The number of trees is adjusted from the set of $\{10, 20, 30, 40, 50, 60, 70, 80, 90, 100\}$. (2) Bootstrap: Bootstrapping is a statistical technique that employs random sampling from the training data. The bootstrap selects training data randomly with replacement. That is, each time a datapoint is chosen, the probability of it being chosen again and being added to the training set is equal. This strategy helps reduce the large variance of random forest models and protects against overfitting. Without the bootstrap option, the random forest algorithm is trained on the full dataset.

6 Result of Cooling Loads Prediction

This section discusses the prediction of the cooling load from the input data. The result of k-fold cross validation is given in Fig. 11. The figure contains the average RMSE, MAE and R^2-score for each of the parameter combinations. On the x-axis, bootstrap combinations are given separately along with their performance with different options for the number of trees (visualised as bars). From the experiments, the following observations can be made.

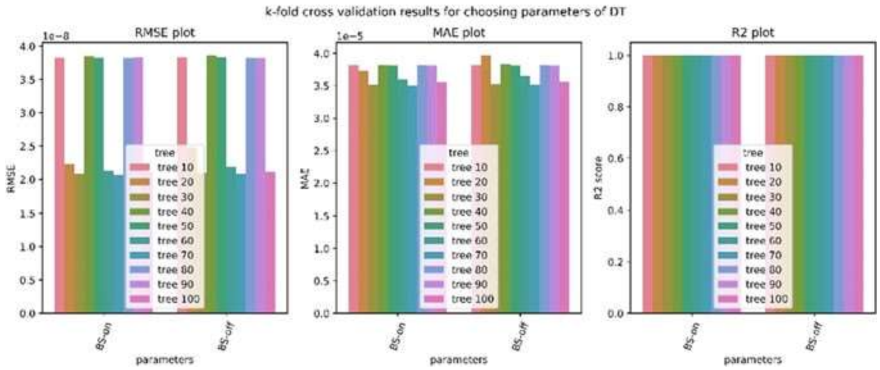

Fig. 11 The result of cooling load prediction. The figure corresponds to the results visualization of hyper-parameter tuning in random forest (BS stands for Bootstrap option)

1. The performance of the whole parameter combinations is excellent because the RMSE varies between the range of 2 × 10^-8 to 4 × 10^-8, MAE varies in the range 3.5 × 10^-5 to 4 × 10^-5, and the R^2 score is close to one.
2. There is no significant difference when the bootstrap option is enabled or not.
3. Increasing the number of trees causes the performance to exhibit an oscillatory behaviour, although the difference in performance is negligible.
4. From the experiments, the optimal model has the following hyper-parameter combination: Bootstrap option enabled and a total of 30 trees.

Figure 12 illustrates the performance metrics corresponding to the best performing models with respect to the number of trees. The results are for the bootstrap option enabled. In the k-fold cross validation, the optimal result is observed when the number of trees is 30, the ccp-alpha value is 0 and the bootstrap option is enabled. The final test results are as follows: the RMSE is 1.986 × 10^-8, MAE is 3.168 × 10^-5, and the R^2-score is 0.99985. These results correspond to the model performance selected after cross-validation with the entire training data and tested with the 20% test data. A result visualisation of random forest prediction for a randomly chosen 100 points is given in Fig. 13. It can be seen that for most of the data points, the actual and predicted value is almost the same or the prediction is very accurate.

7 Conclusion and Recommendations

The current research examined an alternative approach for evaluating the performance of adaptive façade systems, thereby resolving the issues associated with making predictions with the available BPS tools. To accomplish this, a decision tree surrogate model was utilised to estimate the hourly cooling loads of adaptive façade in hot areas. Rhino/Honeybee Grasshopper and Ladybug plugins were used to generate sufficient synthetic datasets of cooling demands for the adaptive façade

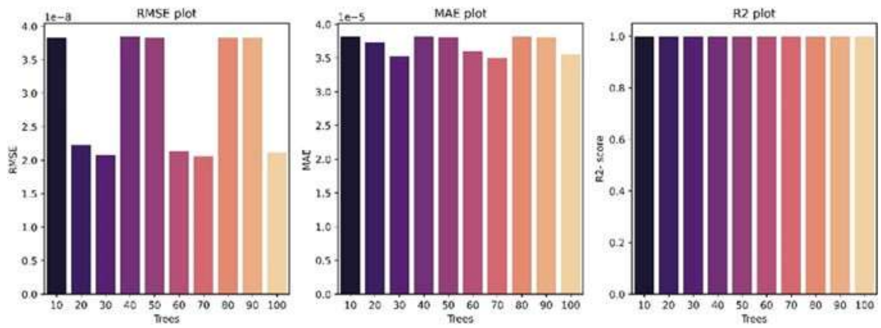

Fig. 12 Performance metrics corresponding to the best performing models

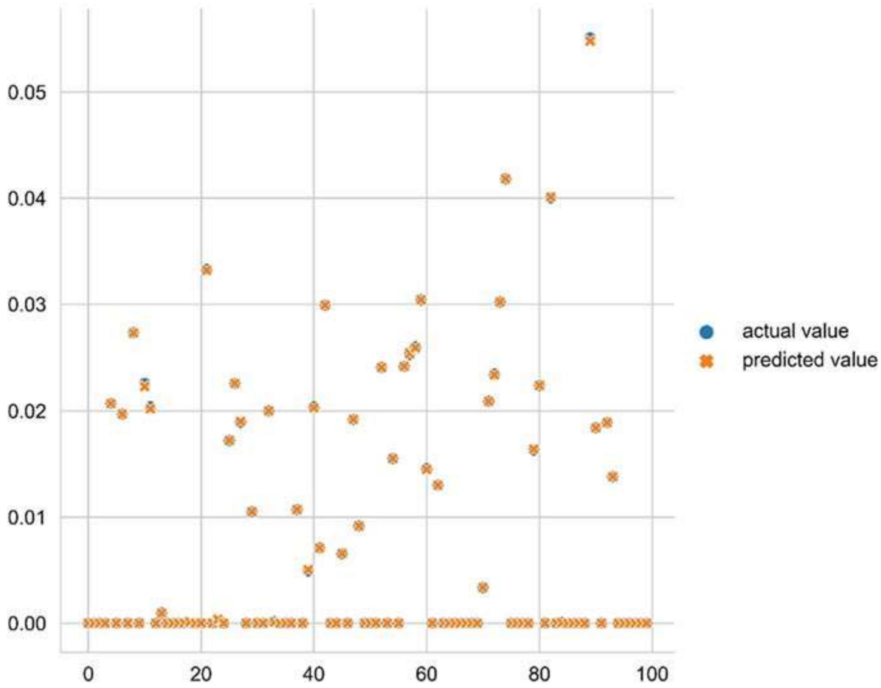

Fig. 13 The plot of the actual and predicted values for a set of 100 randomly chosen test points in the case of cooling load prediction

models. Several fixed and dynamic inputs were chosen that reflect the dynamic fluctuation of the adaptive façade system, design conditions and building envelope parameters with cooling loads as the targeted outcome. Subsequently, the data obtained were used to train, validate and test the suggested decision tree model. In addition, a hyper-parameter tuning study was performed to determine the optimal prediction model. It was found that when developing a ML surrogate model, it is essential to employ a substantial dataset to ensure highly accurate prediction outcomes. The random forest approach was utilised because most of the inputs were of a categorical nature and only a few were continuous. Random forest modelling benefits greatly from the fact that decision trees are well-suited to both categorical and continuous inputs in their learning environment. When using k-fold cross validation, the best results are obtained when using 30 trees, a ccp-alpha of 0, and the bootstrap option. The R^2-score was 0.99985, the RMSE was 1.986–0.10–8, and the MAE was 3.168 × 10–5. Using a range of random examples, the final model was tested, revealing that the predicted and actual values were similar, thereby indicating a high-prediction model. One of the main limitations of this study is the unavailability of real data; thus, simulation is used in this study to collect data because it is the most cost-effective way when data isn't available, or time and money are the main constraints. Furthermore, with simulations it is possible to analysis a wide range of design scenarios and complex modelling that cannot be accomplished easily on a real-world scale.

Because the model is fixed to a particular climate, its applicability to other climates must be determined in future research. Moreover, there is a need to test the model with other types of adaptive façade and determine if the shade factor is sufficient without the requirement of a specific geometry. For future work, the model will be imported into the grasshopper interface to test a set of new scenarios that is not part of the data used to build the developed decision tree model and compare the prediction consumption time and resources between the developed model and the existing BPS tools using large-scale generation of data. By applying this, it is possible to determine how many cases are required by DT prediction to beat the efficiency of BPS tools. Future work may include a comparison of various ML models such as artificial neural networks and recurrent neural networks. Additional future planned work regarding the adaptive façade system includes the use of other environmental control scenarios to automate the behaviour changes of the adaptive façade system.

References

Ahmad, M. W. et al. (2017). Random forests and artificial neural network for predicting daylight illuminance and energy consumption bre centre for sustainable engineering, School of Engineering, AhmadM3@Cardiff.ac.uk ; 2 HippolyteJ@Cardiff.ac.uk ; Related work. *Building Simulation* (May), pp. 1–7.

Al-Masrani, S. M., & Al-Obaidi, K. M. (2019). Dynamic shading systems: A review of design parameters, platforms and evaluation strategies. *Automation in Construction, 102*, 195–216. https://doi.org/10.1016/j.autcon.2019.01.014

Breiman, L. 2019. Random forests, pp. 1–122. doi: https://doi.org/10.1201/9780367816377-11.

Chakraborty, D., & Elzarka, H. (2019). Advanced machine learning techniques for building performance simulation: A comparative analysis. *Journal of Building Performance Simulation, 12*(2), 193–207. https://doi.org/10.1080/19401493.2018.1498538

Gadelhak, M., & Lang, W. (2016). Optimization of office building façade to enhance daylighting , thermal comfort and energy use intensity. *BSO 2016 Building Simulation & Optimization* (January 2017), pp. 1116–1124.

Geyer, P., & Singaravel, S. (2018). Component-based machine learning for performance prediction in building design. *Applied Energy, 228*(October 2017), 1439–1453. Retrieved from https://doi.org/10.1016/j.apenergy.2018.07.011.

Hong, T. et al. (2013). Building Performance Simulation. (2010). Retrieved from http://eetd.lbl.gov/publications/building-performance-simulation.

Loonen, R. C. G. M.; et al. (2014). Performance prediction of buildings with responsive building elements challenges and solutions PERFORMANCE PREDICTION OF BUILDINGS WITH RESPONSIVE BUILDING ELEMENTS: CHALLENGES AND SOLUTIONS. (June), pp. 1–8. Retrieved from https://pure.tue.nl/ws/portalfiles/portal/3992582/392669136570421.pdf.

Loonen, R. C. G. M. et al. (2017). Review of current status, requirements and opportunities for building performance simulation of adaptive facades†. *Journal of Building Performance Simulation, 10*(2), 205–223. Retrieved from https://doi.org/10.1080/19401493.2016.1152303.

López, M. et al. (2017). How plants inspire façades. From plants to architecture: Biomimetic principles for the development of adaptive architectural envelopes. *Renewable and Sustainable Energy Reviews, 67*, 692–703. Retrieved from https://doi.org/10.1016/j.rser.2016.09.018.

Magnier, L., & Haghighat, F. (2010). Multiobjective optimization of building design using TRNSYS simulations, genetic algorithm, and Artificial Neural Network. *Building and Environment, 45*(3), 739–746. Retrieved from https://doi.org/10.1016/j.buildenv.2009.08.016.

Moloney, J. (no date). A MORPHOLOGY OF PATTERN FOR KINETIC FACADES.

Seger, C. (2018). An investigation of categorical variable encoding techniques in machine learning: binary versus one-hot and feature hashing. *Degree Project Technology* , 41. Retrieved from http://www.diva-portal.org/smash/get/diva2:1259073/FULLTEXT01.pdf.

Sheikh, W. T., & Asghar, Q. (2019). Adaptive biomimetic facades: Enhancing energy efficiency of highly glazed buildings. *Frontiers of Architectural Research*. https://doi.org/10.1016/j.foar.2019.06.001

Tabadkani, A. et al. (2020). Innovative control approaches to assess energy implications of adaptive facades based on simulation using EnergyPlus. *Solar Energy, 206*(February), 256–268. Retrieved from https://doi.org/10.1016/j.solener.2020.05.087.

Tabadkani, A. et al. (2021). Design approaches and typologies of adaptive facades: A review. *Automation in Construction, 121*(October 2020), 103450. Retrieved from https://doi.org/10.1016/j.autcon.2020.103450.

Touma, A. Al, & Ouahrani, D. (2017). Shading and day-lighting controls energy savings in offices with fully-Glazed façades in hot climates. *Energy and Buildings, 151,* 263–274. Retrieved from https://doi.org/10.1016/j.enbuild.2017.06.058.

Tso, G. K. F., & Yau, K. K. W. (2007). Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks. *Energy, 32*(9), 1761–1768. https://doi.org/10.1016/j.energy.2006.11.010

Westermann, P., & Evins, R. (2019). Surrogate modelling for sustainable building design – A review. *Energy and Buildings, 198,* 170–186. Retrieved from https://doi.org/10.1016/j.enbuild.2019.05.057.

Yu, Z. et al. (2010a). A decision tree method for building energy demand modeling. *Energy and Buildings, 42*(10), 1637–1646. Retrieved from https://doi.org/10.1016/j.enbuild.2010a.04.006.

Yu, Z. et al. (2010b). A decision tree method for building energy demand modeling. *Energy and Buildings, 42*(10), 1637–1646. Retrieved from https://doi.org/10.1016/j.enbuild.2010b.04.006.