Energy & Buildings 361 (2026) 117501

Contents lists available at ScienceDirect

Energy & Buildings

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

An integrated framework for optimizing daylighting and thermal performance of subtropical secondary school buildings via measurement, simulation and machine learning

Gaoliang Yan [a,c], Xue Zhong [b,c,*], Lihua Zhao [c], Jianhe Luo [c], Shady Attia [d]

[a] Hunan Polytechnic University, Yueyang, Hunan 414006, China
[b] Southwest Jiaotong University, Chengdu, Sichuan 611756, China
[c] South China University of Technology, Guangzhou, Guangdong 510640, China
[d] Sustainable Building Design Lab, Department UEE, Applied Sciences, University of Liège, Belgium

ARTICLE INFO

Keywords:
Daylighting and thermal performance
Secondary school buildings
Parametric form generation
Multi-objective optimization
Explainable machine learning

ABSTRACT

In hot and humid climates, secondary school buildings face the challenge of balancing daylighting, thermal performance, and energy efficiency. Given the inherent trade-offs among such three aspects, multi-objective optimization is essential to systematically balance and optimize these competing performance criteria. Thus, this study proposes a performance-oriented framework combining field measurement, parametric form generation, multi-objective optimization (MOO), and explainable machine learning to investigate the coupled effects of building form on daylighting and thermal performance. A typical secondary school in a hot and humid region was selected as the case study. Indoor conditions were quantified through on-site environmental monitoring and questionnaires based on 7-point thermal sensation scale. Based on paired Thermal Sensation Votes (TSV) of students and measured meteorological parameters, a TSV regression model ($R^2 = 0.921$) was developed and integrated into the subsequent multi-objective evaluation. A rule-based parametric form generation tool was established to generate diverse morphological cases. Subsequently, MOO was conducted on the generated cases to evaluate and balance the daylighting and thermal performance; the resulting dataset was used to train a Gradient Boosting Regression (GBR) model for rapid performance assessment.

MOO results reveal the trade-offs among daylighting, thermal comfort, and energy performance. High-performing solutions consistently converge toward a stable performance control zone, typically characterized by 5–6 stories, 4–6 rows, and 3–5 classrooms per row, and a height-to-spacing ratio of 1.0–1.2. SHapley Additive exPlanations-based interpretation indicates that building massing parameters (rows, stories, and classrooms per row) are the most influential contributors to predictions. These findings highlight the effectiveness of the framework as a decision-support tool for performance-oriented school design in hot and humid climates.

1. Introduction

Secondary school buildings constitute a fundamental component of educational infrastructure, in which indoor daylighting and thermal conditions critically influence learning quality, visual health, cognitive performance, and students' psychological well-being [1–3]. Extensive empirical evidence demonstrates that adequate daylight enhances attention and visual comfort, while thermally stable environments support improved learning efficiency and long-term academic outcomes [2,4]. At the same time, secondary school buildings account for a substantial proportion of campus energy use, particularly in China's hot and humid regions, where cooling and lighting demands remain persistently high due to climatic conditions [5,6]. Under the national dual-carbon goals and the increasing emphasis on low-energy and healthy school environments, achieving integrated optimization of daylighting, thermal comfort, and energy use has become an urgent design challenge.

In recent years, research on the daylighting and thermal performance of educational buildings has primarily focused on environmental parameters affecting students' comfort and learning outcomes. Thermal studies typically consider indoor air temperature (T_a), relative humidity (*RH*), and thermal comfort and cooling energy consumption, while daylighting studies commonly examine illuminance, daylight factor

* Corresponding author at: Southwest Jiaotong University, Chengdu, Sichuan 611756, China.
E-mail address: Zhongxue96@swjtu.edu.cn (X. Zhong).

https://doi.org/10.1016/j.enbuild.2026.117501
Received 22 December 2025; Received in revised form 26 March 2026; Accepted 16 April 2026
Available online 19 April 2026

(*DF*), glare, spatial daylight autonomy (*sDA*) and lighting energy consumption [7,8]. These indicators are strongly influenced by building form, highlighting the critical role of architectural form in shaping indoor environmental quality. Previous studies have shown that classroom geometry and orientation significantly influence energy use and thermal comfort [9,10], while room depth and layout affect daylight penetration and overall energy performance [11,12]; additionally, window-related parameters, including size, placement, and shading strategies, play a critical role in balancing daylighting, thermal comfort, and energy demand [13,14]. In particular, envelope design, window configuration, shading strategies, and light shelves play a key role in enhancing daylight availability and uniformity while balancing thermal performance [15–17]. In current Chinese design practice, standards such as GB 50033 and GB 50099 are predominantly prescriptive, and many schools still fail to achieve high performance in operation [18], largely because static indicators overlook the coupled effects of building morphology on performance [19]. To bridge this gap, recent academic efforts have advanced the understanding of school design; however, most studies focus on the single morphological variable or the performance indicator; few treat overall building form as a set of interrelated parameters within a unified framework to examine their coupled impacts on daylighting and thermal performance [20,21].

Current research on the daylighting and thermal performance of school buildings primarily adopts three approaches. The first approach involves field measurements, which provide *in-situ* data on indoor environmental conditions and enable assessment of actual occupant comfort and building performance [22–24]. However, field measurements have limited spatial and temporal coverage and require substantial resources. They also cannot isolate individual design parameters or simulate hypothetical scenarios, which restricts their applicability in predictive and optimization analyses [25,26]. These limitations can be partially addressed by numerical simulations, which allow designers to explore a wider range of scenarios, isolate design parameters, and predict indoor conditions. To systematically search for optimal solutions within such simulation-based design spaces, evolutionary algorithms—particularly genetic algorithms—are widely adopted for multi-objective optimization [27]. Recent studies have demonstrated that genetic algorithm-based multi-objective optimization can effectively balance daylighting, thermal comfort, and energy use [28]. At the component level, Non-dominated Sorting Genetic Algorithm II (NSGA-II) has been applied to adaptive façades such as building-integrated photovoltaic (BIPV) shading and Venetian blinds, achieving simultaneous improvements in electricity generation, daylight performance, and thermal comfort while reducing overall energy consumption [29–31]. Among them, the NSGA-II is especially suitable for handling conflicting objectives and complex, non-linear building performance problems, as it enables efficient Pareto-based search without requiring gradient information [32,33]. This allows evaluation of multiple scenarios, optimization of building form and envelope parameters, and prediction of energy use and indoor comfort at early design stages [34]. Although simulation-based studies have shown the potential on multi-objective optimization (MOO) [13,35,36], building performance simulation (BPS) often relies on idealized assumptions, simplified boundary conditions, and uncertain inputs. These limitations can reduce predictive accuracy, thereby necessitating validation with field measurements [37]. With the development of artificial intelligence (AI), the third approach uses machine learning (ML) to predict indoor environmental conditions, identify key influencing factors, and support rapid design decisions for complex, multi-objective issues [38,39]. ML-based studies have demonstrated the potential of AI for predicting and optimizing indoor environmental conditions in educational buildings. ML approaches, such as Q-XGBoost reinforcement learning [40], class-weighted random forests [41], and ML models accounting for spatial variability [42], have demonstrated superior performance in predicting building-related outcomes such as thermal comfort and indoor air quality. Building on these advances, ML methods offer advantages over traditional approaches by capturing complex multi-factor interactions and extending spatial and temporal coverage. However, model performance remains highly dependent on the quality and representative of the training data, and purely ML-based models are limited in their ability to explain underlying mechanism between building form and performance.

Against this background, this study aims to develop a performance-driven design framework integrating MOO and ML to support building form optimization for daylighting, thermal comfort, and energy performance. The scope of this study encompasses a parametric exploration of typical secondary school building forms in hot–humid climates. First, field measurements of indoor daylighting, thermal conditions and occupant-related parameters were conducted to provide reliable inputs and to validate the accuracy of building performance simulations. Next, BPS were used to generate multi-scenario datasets across diverse building-form configurations and find the optimized ones to support ML analysis. In this process, as the optimization scope is inherently constrained by the number of generated design solutions and the associated computational costs, meticulous sampling is required to ensure the relative completeness of the Pareto front. Finally, ML models were trained to identify dominant form-related drivers and support early-stage design decisions by revealing how building form jointly shapes daylight and thermal conditions.Fig. 1

2. Methodology

2.1. Overall Framework

An integrated workflow (Fig. 2) was developed to support performance-oriented building form design by coupling field measurement, parametric modeling, simulation, MOO and ML-based evaluation. Unlike conventional approaches, this framework is structured around a unified Key Performance Indicators (KPIs) system that consistently links architectural morphology, simulation, optimization, and machine learning.

Firstly, indoor environmental parameters (T_a, *RH*, air velocity (V_a), globe Temperature (T_g), illuminance (E) and luminance (L)) were recorded through on-site monitoring, and *TSV* data were collected using questionnaires. A *TSV* regression model was established from the measurement as well as questionnaire, and embedded as a core indicator in subsequent performance simulations.

Secondly, a rule-based parametric plugin was developed as a computational bridge. It translates morphological rules into generative logic within Grasshopper, enabling the automated creation of diverse building forms. By encoding variables as parametric inputs, the plugin systematically defines the design space for subsequent performance evaluation and MOO.

Thirdly, parametric models were input into simulation using Ladybug Tools, and MOO was conducted with NSGA-II to explore trade-offs among daylighting, thermal comfort, and energy performance. Pareto-front analysis was applied to identify the Performance-Convergent Zones (PCZ), which determines the optimization dataset for ML training.

Finally, Gradient boosting regression (GBR) models were trained on a dataset comprising geometric parameters and performance metrics generated throughout the GA-based optimization process. Model performance was assessed using several indicators, and SHapley Additive exPlanations (SHAP)-based analysis was applied to identify dominant morphological drivers.

2.2. Field Measurement and TSV Modeling

2.2.1. Case Description

To support the proposed optimization framework, field measurements and occupant surveys were conducted in a typical secondary school building in Guangzhou (112°57′–114°03′E, 22°26′–23°56′N).

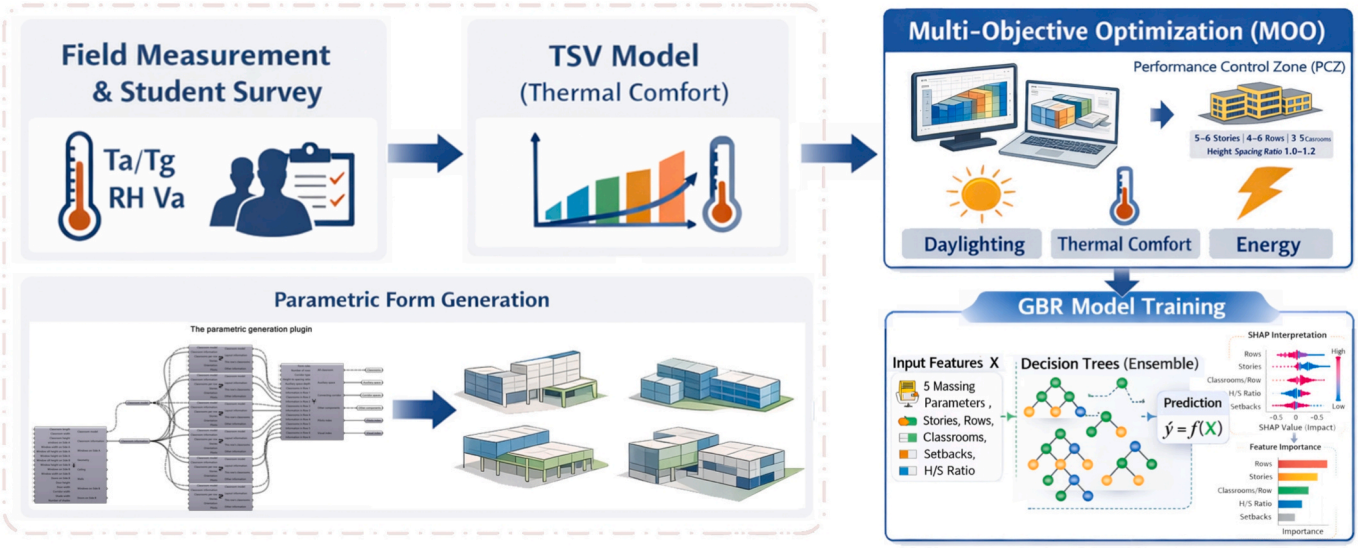


Fig. 1. Workflow from field measurement to multi-objective optimization.

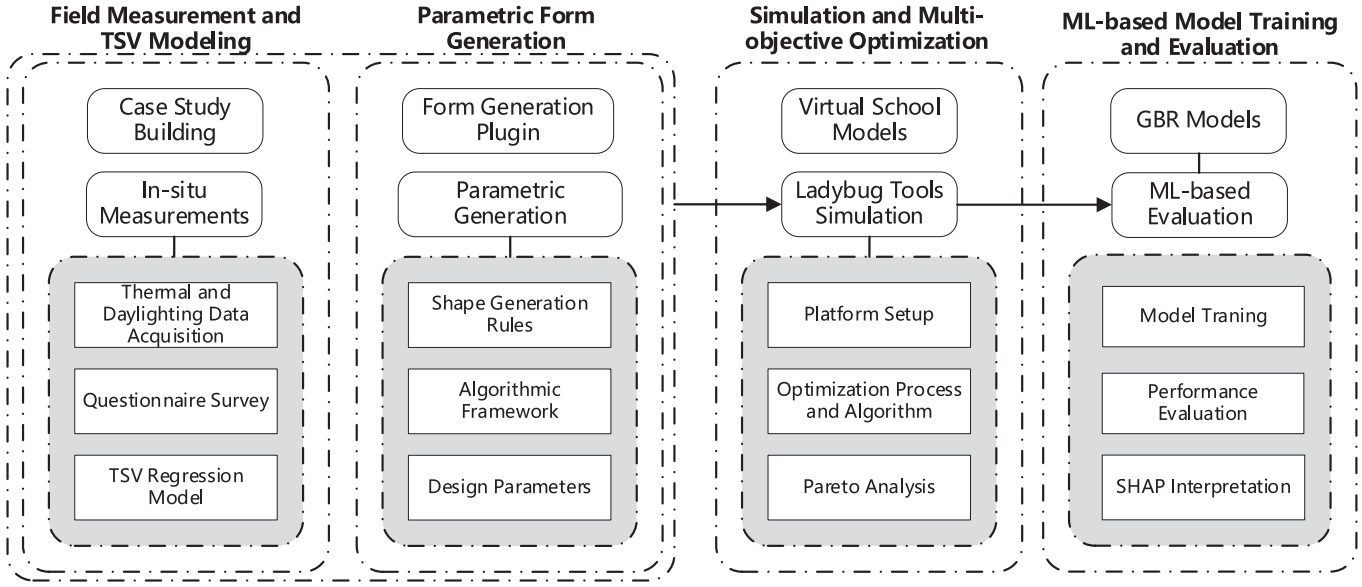


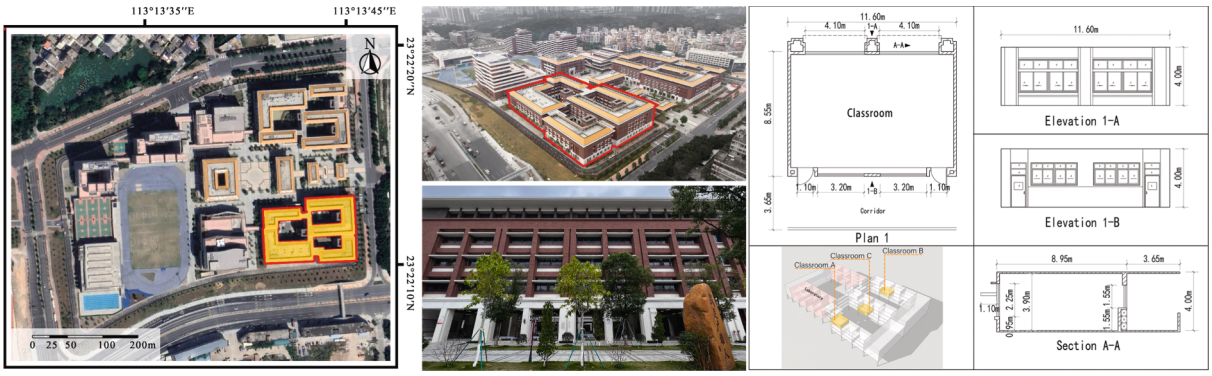
Fig. 2. Workflow of the proposed framework.

Guangzhou, located in southern China, is characterized by a humid subtropical monsoon climate (Köppen–Geiger: Cfa), featuring long, hot summers and abundant solar radiation, with annual sunshine of 1700–1900 hours and typical maximum and minimum temperatures of 39 ℃ and 5℃, respectively. Prevailing winds are northeastern in winter and southeastern to southern in summer. The selected building reflects standard spatial layouts and construction characteristics of secondary school buildings in hot–humid regions, thereby providing a representative case within the defined study scope. Experiments were conducted during regular academic semesters (March–July and September–January), when classrooms are more susceptible to excessive heat gain and daylight glare. Secondary school classrooms were selected because of their relatively stable occupancy patterns and the high thermal sensitivity of students, which enables reliable collection of thermal comfort data.

The case-study building is a five-story reinforced concrete structure with a total floor area of 27,336 m2, featuring an enclosed courtyard layout with a south-by-east orientation of 3°18′. Ordinary classrooms are primarily located in the east wing, with laboratories in the west wing. Double-glazed windows and external shading devices (such as canopies and horizontal fins) are applied (Fig. 3ab).

Three representative classrooms were selected based on orientation, floor level, and shading configuration (Fig. 3c): Classroom A (south-facing, 4th floor, canopy depth 2.0 m, 12.3 × 8.25 × 4.0 m); Classroom B (north-facing, 3rd floor, fin depth 1.1 m, 11.6 × 8.55 × 4.0 m); and Classroom C (courtyard-facing to the south, 2nd floor, fin depth 1.1 m, 11.6 × 8.55 × 4.0 m). These classrooms cover typical combinations of physical factors affecting indoor daylight and thermal performance, with their baseline conditions complying with Chinese standards GB 50099 (e.g., 1/5 window-to-floor ratio and 3% daylight factor) [43].

In addition, surrounding ground and façade materials were characterized to account for solar reflection and radiative exchange. The campus primarily consists of concrete pavements and brick façades, which influence reflected radiation, indoor heat gain and daylight distribution.



(a) School layout (b) Façade display (c) Geometric dimensions of the three classrooms

Fig. 3. Layout and façade of the sampled Guangzhou secondary school (Google Earth; authors' photographed).

2.2.2. Measurement Arrangement and Environmental Monitoring

(1) In-situ Measurement Schedule

To represent the typical operational conditions of secondary schools in a hot-summer and warm-winter climate, field measurements were conducted in Guangzhou across three seasons—transitional (May), summer (June), and winter (December). The measurement periods based on the principle of capturing the most unfavorable lighting and thermal conditions were chosen. These correspond to, during the school term, the hottest day in summer, the coldest day in winter, and a typical day in the transitional season. From the perspective of lighting conditions, luminance was measured under clear-sky conditions in summer and winter, whereas illuminance was measured under overcast conditions in the transitional season. All measurements were conducted around midday, when the solar altitude varies little to minimize temporal changes in daylight caused by variations in solar angle. The day length is approximately 13 hours and 15 minutes in the transitional season, 13 hours 30 minutes in summer, and 10 hours 42 minutes in winter.

Each measurement campaign was carried out during typical class hours (08:00–18:00), during which classrooms remain occupied for over 10 hours per day under both air-conditioned and naturally ventilated modes. Thermal measurements captured representative seasonal variations. Daylighting measurements followed GB/T 5699 protocols under controlled sky conditions: overcast for stable illuminance and clear sky for high luminance associated with potential glare. All measurements were conducted between 10:00 and 14:00 to ensure relatively stable solar altitudes, providing consistent illuminance conditions while capturing representative high-luminance scenarios. The period reflects typical occupancy and Brager and de Dear environmental framework [44]. The detailed conditions are summarized in Table 1.

Field measurements were conducted over multiple days in each season, with consecutive monitoring during each campaign. To ensure meteorological representativeness, the measured outdoor air temperature was compared with Typical Meteorological Year data for Guangzhou, confirming that the recorded conditions fall within the typical seasonal ranges rather than isolated weather events.

Table 1
Condition of field measurements.

Season	Date(2023)	Weather Condition	Measured Indicators
Transition	May 16	23–28℃, Sunny	T_a, RH, T_g, V_a, and L
	May 18	24–30℃, Overcast	E
Summer (AC)	June 13	26–34℃, Sunny	T_a, RH, T_g, V_a, and L
	June 14	25–32℃, Heavy Rain	E
Winter	Dec 19	7–11℃, Light Rain	T_a, RH, T_g, V_a, and E
	Dec 20	7–12℃, Sunny	L

*T_a= air temperature (°C); RH = relative humidity (%); T_g = globe temperature (°C); V_a = air velocity (m/s); E = illuminance (lx); L = luminance(cd/m^2).

(2) Measured Parameters, Instruments, and Sensor Layout

The monitored parameters included: T_a, RH, T_g, and V_a, which represent the core thermal environment parameters influencing human heat balance and thermal comfort. In addition, indoor E and L were measured to assess visual performance and glare risks. These indicators constitute the fundamental environmental inputs for constructing the *TSV* model and analyzing daylighting and thermal conditions in classroom spaces. The testing instruments and measured parameters are listed in Table 2.

Following the Hygienic Standards for Public Places [45], two indoor measurement points were arranged in each classroom to monitor thermal-related variables, which were evenly distributed along the longitudinal axis at a height of 1.1m to represent the occupied zone. An outdoor reference point was set in an open area near the teaching building at a height of 1.1 m to provide boundary meteorological data for simulation (Fig.4).

Daylight illuminance was measured under overcast conditions between 10:00 and 14:00, when levels were relatively stable. Measurements were conducted with curtains open and lights off. The workplane was the student desk surface, divided into a 1 × 1 m grid, with the lux meter placed horizontally at 0.75 m above the desk [13] (Fig.5).

Additionally, luminance was measured under clear-sky conditions at noon (12:00). Three measurement points were evenly arranged along the window midline toward the interior wall, and three points along the midline between windows (Fig.6). Measurement heights were 1.5 m for standing and 1.2 m for seated positions [43].

2.2.3. Thermal and Daylighting Comfort Questionnaire

A daylighting and thermal comfort questionnaire was administered concurrently with field measurement to link measured environmental conditions with occupants' subjective responses. Developed with

Table 2
Measurement parameters and instruments.

Measured Parameter	Instrument	Model	Range	Accuracy
T_a, RH	Temperature & Humidity Logger	HOBO MX2302A	-40~70℃ 0~100%	±0.2℃±2.5%
T_g, V_a	Thermal Comfort Analyzer	Delta Ohm HD32.3	-10~100℃0-5m/s	Class 1/3 DIN ±0.05m/s
E	Illuminance meter	TES1330A	0.01-20000lx	±3%rdg ±0.5%f.s
L	Luminance meter	Konica Minolta LS-110	Fast: 0.01-999,900cd/ m2Slow: 0.01-499,900cd/ m^2	0.01-9.99 cd/ m^2±2%±2 value, > 10.00 cd/ m2 ±2%±1 value

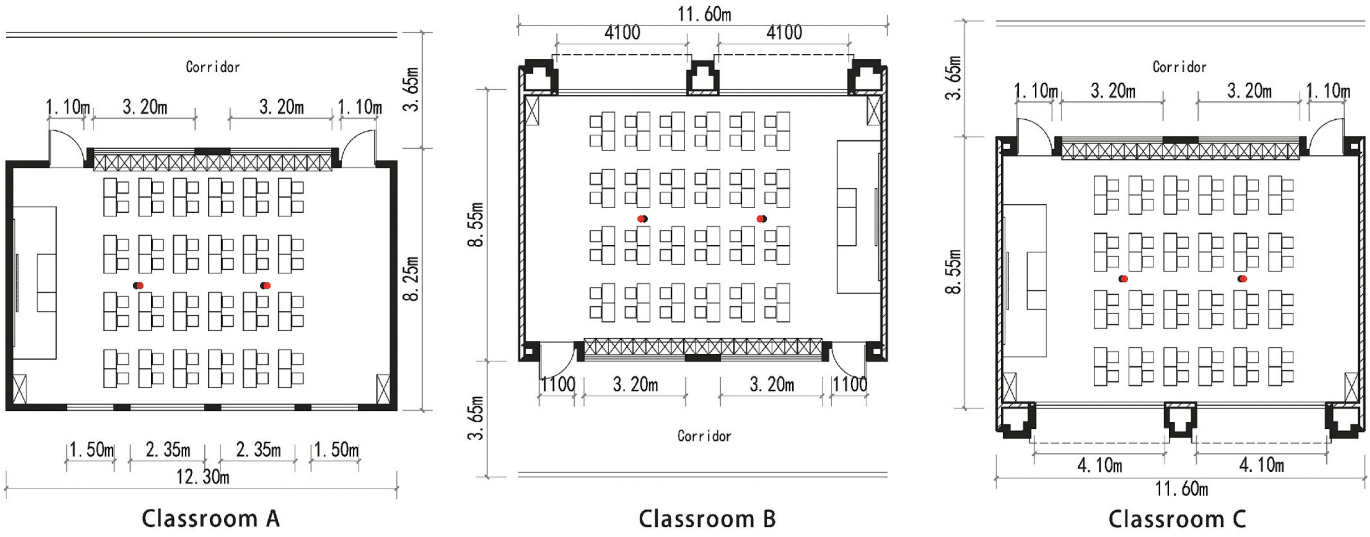


Fig. 4. Classroom layouts (a) and sensor arrangement (b).

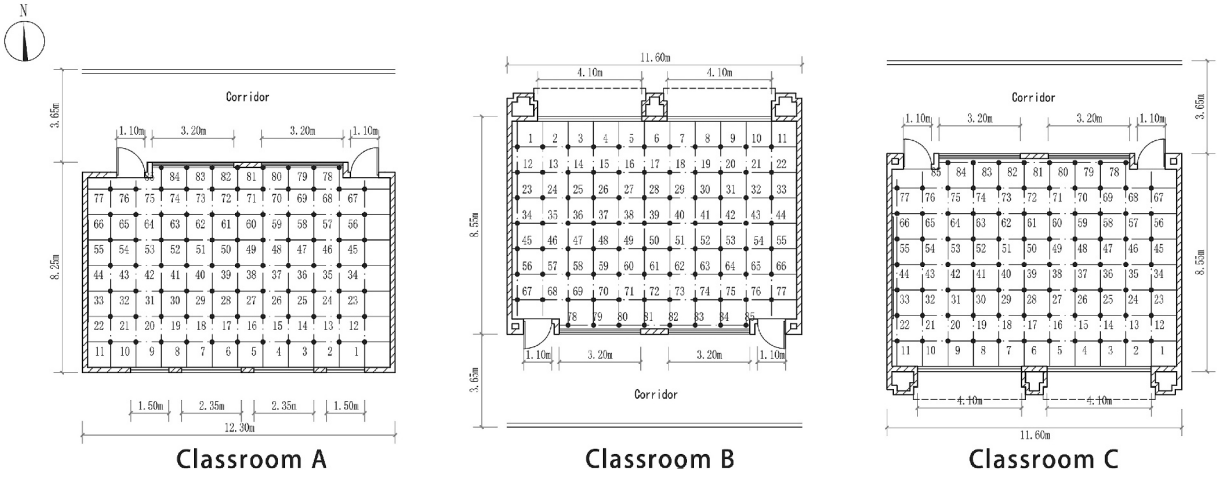
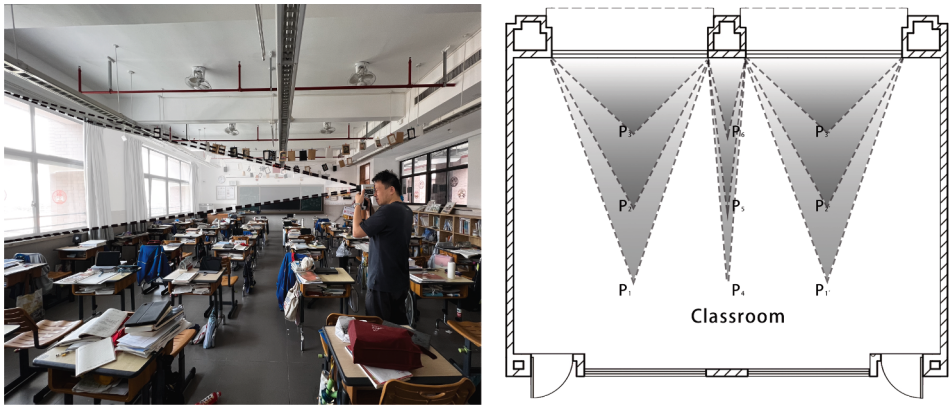


Fig. 5. Sensor arrangement for classroom illuminance.

reference to ASHRAE 55 [46] and ISO 7730 [47] and adapted to middle-school classrooms, it covered *TSV* (seven-point scale, −3 to +3) and thermal preference vote (*TPV*), humidity, airflow sensation as well as preference, visual comfort and daylight preference. Subjective and other information, including clothing insulation (clo), activity level (met), and recent window or AC operation, was also collected. Surveys were conducted during regular class hours to ensure temporal alignment with sensor measurements. A total of 513 questionnaires were distributed, of



(a) On-site luminance measurement (b) Layout of measurement points

Fig. 6. On-site luminance measurement (a) and layout of measurement points (b).

which 477 valid responses were retained after data screening. Each *TSV* record was matched to synchronous T_a, *RH*, T_g, and V_a measurements by timestamp, and standard preprocessing (outlier removal, missing-value handling, and normalization) was applied prior to analysis. The 477 valid TSV samples were distributed across seasons, classrooms, and ventilation modes, including summer (34.6%), transition (27.5%), and winter (37.9%), with both natural ventilation (65.4%) and air-conditioned conditions (34.6%) represented. This balanced distribution prevents dominance by any single subgroup and enhances the robustness of the TSV model within the defined study scope.

The questionnaire followed internationally recognized thermal comfort scales. TSV responses were paired with measured environmental parameters to develop a regression-based thermal comfort indicator. Rather than directly applying PMV/PPD indices, which are primarily developed for steady-state indoor environments, the TSV-based approach enables calibration using locally observed thermal perceptions under dynamic and adaptive conditions.

2.2.4. *TSV regression and validation*

Based on the paired dataset of questionnaire-based TSV responses and field-measured environmental parameters, a multiple linear regression (MLR) model was developed to quantify the linear relationship between thermal sensation and environmental variables in both naturally ventilated and air-conditioned classrooms. The regression coefficients were determined through least-squares estimation, and multicollinearity was checked using variance inflation factors (VIF). Model robustness was evaluated by partitioning the dataset into training and validation subsets and applying standard performance metrics (e.g., R^2) together with k-fold cross-validation. The resulting TSV regression model was implemented as a custom Python component within the Ladybug Tools workflow, enabling *TSV* to be directly incorporated as a quantitative objective in subsequent daylighting and thermal simulation and MOO.

2.3. *Parametric form generation*

A rule-based parametric modeling framework was developed in Grasshopper to automatically generate secondary school building forms. The framework follows a rule-driven logic, in which predefined morphological rules control form generation and inputs of parameters determine geometric variation.

2.3.1. *Morphological survey and rule extraction*

A parametric form-generation framework was developed based on a systematic survey of 58 secondary school buildings constructed in Guangdong Province over the past decade, ensuring the temporal representativeness of recent educational infrastructure development. The geographical distribution of these surveyed samples, primarily concentrated in the Pearl River Delta, is illustrated in Fig. 7.

The survey comprised analyses of spatial composition, topological configuration, and fractal dimensions to identify underlying patterns in school building forms, together with a statistical examination of key geometric parameters influencing building morphology (Fig. 8). To ensure reproducibility, the morphological categories were obtained using k-means clustering (k=3) on quantitative features including layout topology, building area, number of stories, classroom depth-to-width ratio, and inter-row spacing. Cluster centroids were used to define category thresholds, allowing consistent assignment of new buildings to the three categories. Based on the statistical clustering of spatial configurations, three fundamental morphological categories were identified: single-line, multi-line, and modular combination. Each category was abstracted into specific parametric rules: single-line (row count, corridor width, orientation); multi-line (spacing, alignment, offsets); and modular (dimensions, connectivity). These constraints ensure a reproducible, rule-driven basis for automated model generation. Parametric Generation Workflow (Steps 1–6):

a) Normalization: Standardize morphological features from surveyed buildings.
b) Clustering: Apply K-means (k=3) to identify morphological categories.

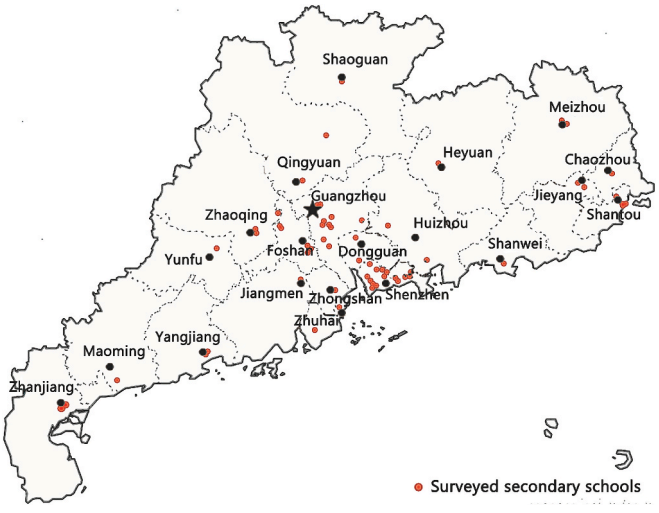


Fig. 7. Distribution of surveyed secondary schools in Guangdong.

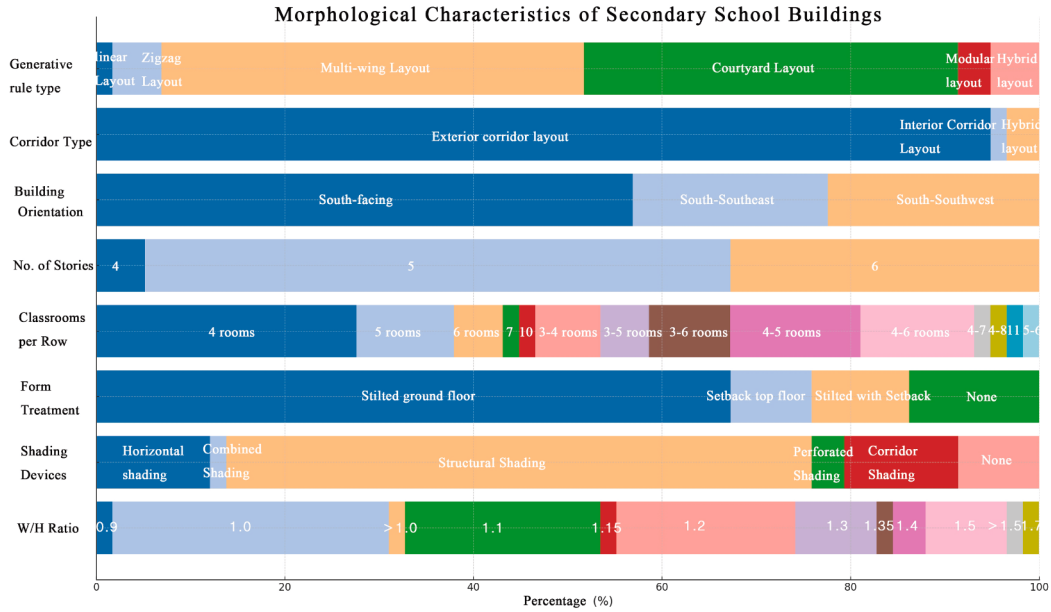


Fig. 8. Morphological Characteristics of Secondary School Buildings.

c) Thresholding: Define cluster centroids and category-specific thresholds.
d) Mapping: Link clusters to parametric rule-sets and constraints.
e) Generation: Execute rule-driven form generation in Grasshopper.
f) Integration: Output 3D models compatible with MOO workflows.

2.3.2. Rule-based parametric modeling

Following the extracted morphological rules, a rule-driven parametric framework was implemented in Grasshopper. The framework integrates two levels of generative logic: (1) unit-level assembly, in which basic functional units (e.g., classroom modules) are parameterized in terms of dimensions, orientation, and story height; and (2) building-level composition, in which unit assemblies are aggregated into complete building forms according to selected generative rules. Users can specify key morphological parameters including category type, number of stories, height-to-spacing ratio, corridor configuration, classroom dimensions, and inter-row alignment, ensuring that generated forms remain consistent with observed school patterns while allowing parametric variation. Based on these inputs, the algorithm automatically generates building forms consistent with the rule-based classification while preserving geometric consistency, spatial hierarchy, and typical school design constraints.

The framework was developed as a custom Grasshopper plugin that outputs three-dimensional models fully compatible with Ladybug Tools (the code was attached in the Supplementary material). These models can be directly used for daylighting and thermal simulations as well as MOO, thereby ensuring reproducibility and enabling seamless integration within the measurement–simulation–optimization workflow.

2.4. Simulation and MOO

An integrated simulation–optimization workflow was established based on models derived from parametric modeling.

2.4.1. Simulation, Validation, and Optimization

A simulation model and workflow for the case-study school were developed using the form-generation plugin in combination with the Ladybug Tools–Radiance–EnergyPlus workflow. Field-measured geometry, material properties, and boundary conditions were incorporated to ensure consistency with actual classroom conditions. Simulated daylighting and thermal performance were validated against on-site measurements, and model accuracy was quantified using relative deviation metrics, thereby confirming the reliability of both the simulation model and the workflow for subsequent optimization.

The validated simulation model and workflow were then integrated into a MOO framework implemented in Grasshopper, coupling the form-generation plugin, Ladybug Tools, and the NSGA-II algorithm. Rule-based parametric building forms were used as geometric inputs, while thermal, daylighting, and energy-related metrics were defined as objective performance indicators. Consistent simulation assumptions (e.g., schedules, internal gains, HVAC setpoints, ventilation, and shading) were applied across all cases, based on typical operation of the case-study school, to ensure comparability among design alternatives.

(1) Design variables

To capture the essential morphological variations of typical middle-school teaching buildings, six categories of design variables were selected based on field surveys and typological analysis. These morphological parameters are translated into geometric inputs for simulation models, where layout type and spacing dictate solar access and shading; building height and density modulate airflow and heat accumulation; and corridor and façade configurations regulate daylight distribution and envelope heat transfer. These physics-based simulations collectively determine daylighting and thermal performance, with simulation outputs subsequently quantified as performance-based KPIs to establish a direct mapping between building form and optimization objectives. The range of design variables (e.g., number of floors and height-to-spacing ratio) were constrained by local regulatory standards. Design variables and their significance for the daylighting and thermal performance of school buildings are summarized in Table 3.

(2) Multiple objectives

The objectives were defined to evaluate the daylighting and thermal performance of the parametric school-building models. To ensure a representative evaluation of the entire building, all performance indicators (*DA*, *UDI*, *SET*, *TSV*, and *Energy*) were calculated as spatio-temporal averages across all classrooms on all floors within each generated model. These KPIs were selected based on established building performance standards and previous studies on indoor environmental quality, while also addressing the climatic characteristics of Guangzhou's hot–humid context. The MOO was formulated to simultaneously improve daylighting quality, thermal comfort, and building

Table 3
Design variables and performance objectives

Category	Item	Range / Metrics	Description
Design Variables	Form type	Linear layout/multi-wing layout/ staggered layout/ courtyard layout	Defines overall layout, affecting daylighting penetration and airflow.
	Number of floors	4–6	Determines vertical stacking of classrooms, influencing floor-level daylight and thermal conditions.
	Number of building rows	2-6	Controls building density and inter-row spacing, impacting solar access, shading, and natural ventilation.
	Classrooms per row	3-6	Sets horizontal extent of building block, affecting daylight availability and internal heat load.
	Height-to-spacing ratio	1.0-1.5	Governs building height vs. spacing, influencing shading, wind flow, and thermal comfort.
	Setback levels	0-2	Alters roof exposure and shading, affecting rooftop heat gain and daylight reflection.
Objectives (annual mean)	Daylighting performance	*DA* (daylight autonomy) */UDI* (useful daylight illuminance) */$E_{\_lighting}$*	Evaluates indoor daylight availability, visual comfort, and lighting energy for daytime. *DA*/*UDI*: maximize; $E_{\_lighting}$: minimize.
	Thermal performance	*TSV*/*SET* (standard effective temperature)/$E_{\_cooling}$	Represents indoor thermal conditions, occupant comfort, and cooling energy. *SET* / | *TSV* |/ $E_{\_cooling}$: minimize.

energy performance. Specifically, *DA* and *UDI* were selected to provide a comprehensive assessment of daylighting performance: *DA* reflects the sufficiency of daylight availability, while *UDI* captures potential over-illumination and glare risks, thereby overcoming the limitations of static prescriptive thresholds. For thermal performance, *TSV* and *SET* were adopted to integrate objective environmental conditions with subjective thermal perception, as informed by the field survey. Thermal sensation is further decomposed into TSV(+) and TSV(−), representing the mean intensity of warm-side votes (>0) in the transition and cooling seasons and cool-side votes (<0) in winter, respectively, thereby avoiding mutual offsetting during seasonal aggregation. Furthermore, the inclusion of $E_{\_lighting}$ and $E_{\_cooling}$ accounts for the inherent trade-offs between daylight utilization and cooling demand. Collectively, these KPIs ensure that the optimization process is driven by environmental performance and occupant comfort rather than geometric characteristics alone. The objective functions are summarized in Table 3.

(3) Optimization engine

The optimization process followed a standard NSGA-II evolutionary procedure, including population initialization, fitness evaluation via simulation, genetic operations, and Pareto-front extraction. The simulation environment was integrated within the Grasshopper platform, utilizing Ladybug Tools (v1.6.32) as the core interface, with Radiance (v5.4a) and EnergyPlus (v22.2.0) employed for daylighting and thermal engines, respectively. Surface albedo for surrounding concrete and façade materials was set at 0.2 to account for solar reflection and its impact on indoor performance. Key parameters were set to a crossover rate of 0.9, a mutation rate of 0.05, and a population size of 50 across 30 generations, resulting in 1,500 evaluated design solutions. The total number of evaluations Surface reflectance was determined by the trade-off between the complexity of the multidimensional design space and the computational intensity of the coupled daylighting and thermal simulations. Preliminary tests indicated that 1,500 evaluations provided a robust convergence of the Pareto front where further iterations yielded diminishing returns in hypervolume improvement. This sample size aligns with established practices in building-scale optimization, balancing computational feasibility with the need for representative exploration of the design space.

2.4.2. Pareto-front Analysis

Pareto-optimal solutions were clustered to identify groups exhibiting similar performance characteristics. Within each cluster, PCZs were defined as regions in which daylighting, thermal, and energy performance metrics showed low variability, indicating robust and high-performing design solutions. Subsequently, sensitivity analysis was performed to quantify the influence of key design variables on performance outcomes, thereby revealing the underlying relationships between building form and environmental quality. Based on the combined results of clustering and sensitivity analyses, a Design Intervention Sequence (DIS) was developed to outline stepwise design strategies that prioritize robust, high-performance morphological adjustments.

While the obtained Pareto front provides a representative approximation of the trade-off space, it may not fully capture all possible optimal solutions due to computational constraints. The robustness of the Pareto analysis is inherently influenced by the number of generated solutions and the diversity of the design space explored. Increasing the number of simulations could further improve the completeness and stability of the Pareto front.

2.5. ML-based evaluation and sensitivity analysis

MOO produced a large simulation dataset linking building form to daylighting and thermal performance. To enable rapid performance evaluation of building form's impacts on performance across the design space, ML-based models were developed.

2.5.1. Dataset construction

The ML dataset was derived from the full NSGA-II-based optimization process. After removing duplicate solutions, 1,125 unique samples were retained from 1,500 candidates. Input features included morphological parameters (form type, number of stories, number of rows, classrooms per row, segment variation, height-to-spacing ratio, and setback configuration). Output variables comprised simulated performance metrics: *TSV* (split into *TSV(+)* and *TSV(−)*), *SET*, *DA*, *UDI*, and cooling and lighting energy use. These outputs variables correspond directly to the KPIs defined in the MOO process and were consistently used as target variables in the GBR model, thereby ensuring alignment between optimization objectives and machine learning predictions and forming a closed-loop mapping from morphological parameters to performance evaluation. Before model training, preprocessing steps, including missing-value handling, outlier filtering, and Min–Max normalization, were applied. The dataset was split into training (80%) and testing (20%) sets, with 5-fold cross-validation used to ensure model robustness.

2.5.2. Model selection and training

Five representative ML algorithms, including MLR, support vector regression (SVR), random forest (RF), artificial neural network (ANN), and GBR, were employed to capture the complex relationships between morphological design variables and daylighting and thermal performance. These models collectively cover linear, kernel-based, ensemble, and neural network learning paradigms, enabling a comprehensive assessment of alternative modeling strategies. Each algorithm was implemented using a representative configuration: SVR with a radial

basis function (RBF) kernel; RF and GBR with tuned tree depth and numbers of estimators; and ANN with one to two hidden layers using ReLU activation. All models were trained on the preprocessed dataset using the same training and validation protocol to ensure comparability. Hyperparameters were optimized using grid search, with mean squared error (MSE) as the loss function and R^2, root mean squared error (RMSE), and mean absolute error (MAE) as evaluation metrics. This standardized training framework ensured a fair comparison across algorithms.

2.5.3. Model Interpretability and Feature Importance

After determining the optimized model, SHAP analysis was conducted to interpret how different morphological variables influence the thermal and daylighting performance. SHAP quantifies the marginal contribution of each feature across all possible feature combinations, providing consistent and additive measures of feature importance. Given the importance of morphological indicators, representative variables were selected for SHAP-based interpretability analysis.

3. Results

3.1. Field Measurements and TSV Analysis

This section presents the results of field measurements and student *TSV* surveys conducted on representative school days, providing empirical data for comparative assessment of both objective as well as subjective thermal conditions and the established *TSV* model.

3.1.1. Field Measurement of Indoor Daylighting and Thermal Conditions

In thermal environment assessment, T_a, T_g, RH and V_a are the four key variables influencing human heat perception and determining the overall thermal load experienced by occupants. The seasonal measurement results for these variables are shown below.

(1) Thermal performance of transitional season

The indoor and outdoor thermal conditions of three classrooms during the transitional season (May 16) are shown in Fig. 9, a typical clear day with elevated outdoor temperatures. Outdoor T_a ranged 25.3 to 29.8 °C with high RH (mean 70%) and V_a (peak 2.65 m/s). Indoors, T_a averaged 24.4 to 25.6 °C (peaks 26.0 to 27.2 °C). T_g increased with solar exposure, highest in Classroom A (mean/peak: 26.4/27.4 °C), followed by B and C, indicating orientation, floor level, and shading effects. RH ranged 72.5 to 75%, inversely related to temperature, with Classroom B slightly higher. Indoor V_a was generally low (0.10 to 0.20 m/s), with Classroom C slightly higher, showing limited natural ventilation despite strong outdoor winds. Overall, the transitional-season environment was moderate, with radiative heat gain and ventilation effectiveness driving inter-classroom differences. T_g was positively correlated with solar radiation and T_a, negatively correlated with RH, and higher V_a was associated with improved thermal perception.

(2) Thermal performance of summer season

The indoor and outdoor thermal conditions of three classrooms during the summer season (June 13) are shown in Fig. 10, a typical hot day with peak horizontal irradiance of 1014 W/m^2. Outdoor T_a ranged 31.1 to 37.6 °C, and T_g peaked at 50.1 °C (mean 38.8 °C), reflecting intense solar radiation. With air conditioning set at 26.0 °C, indoor T_a averaged 27.0, 25.6, and 26.8 °C in Classrooms A, B, and C, while T_g

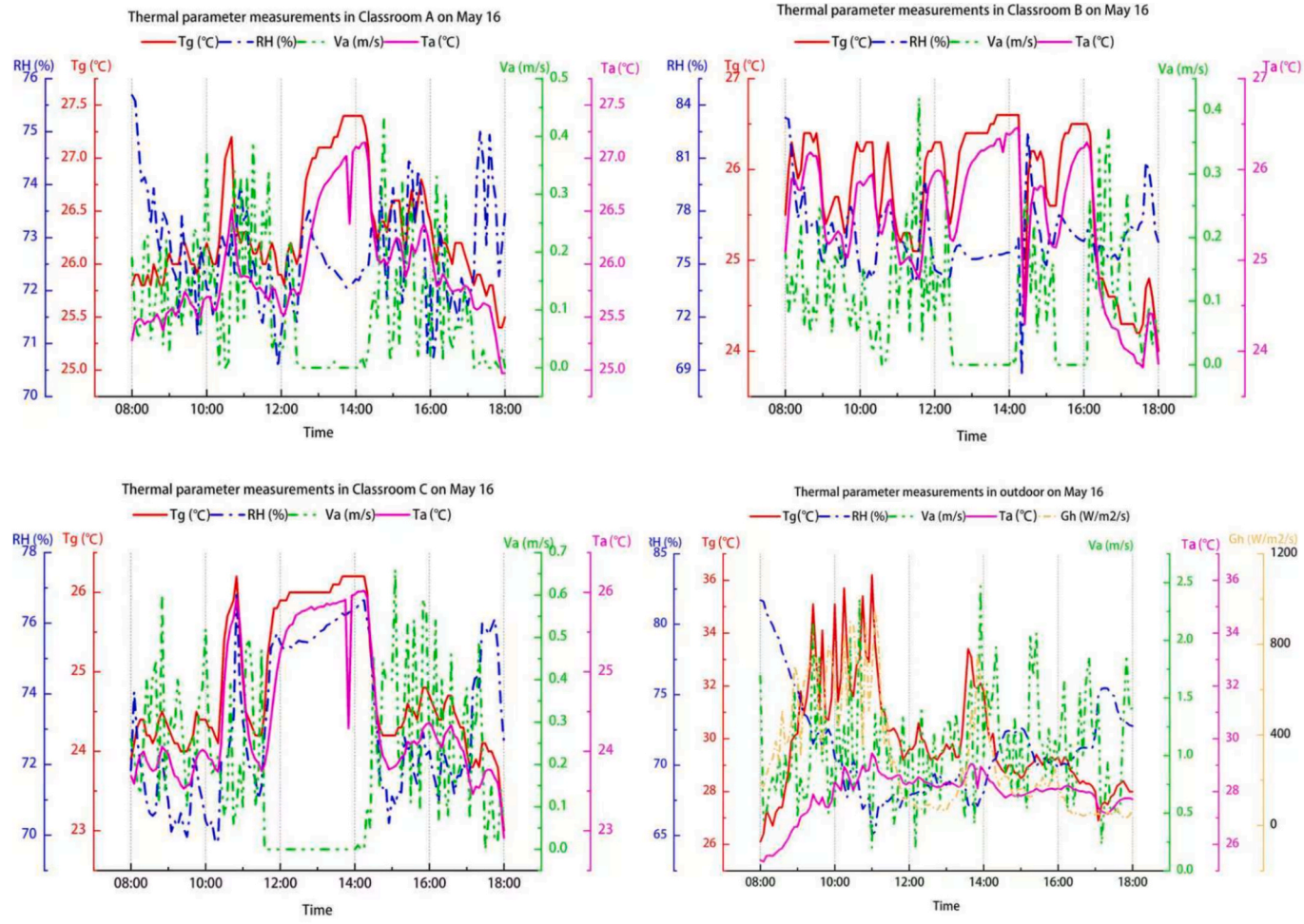


Fig. 9. Measured Results of Thermal Environment in Transitional Season (May 16).

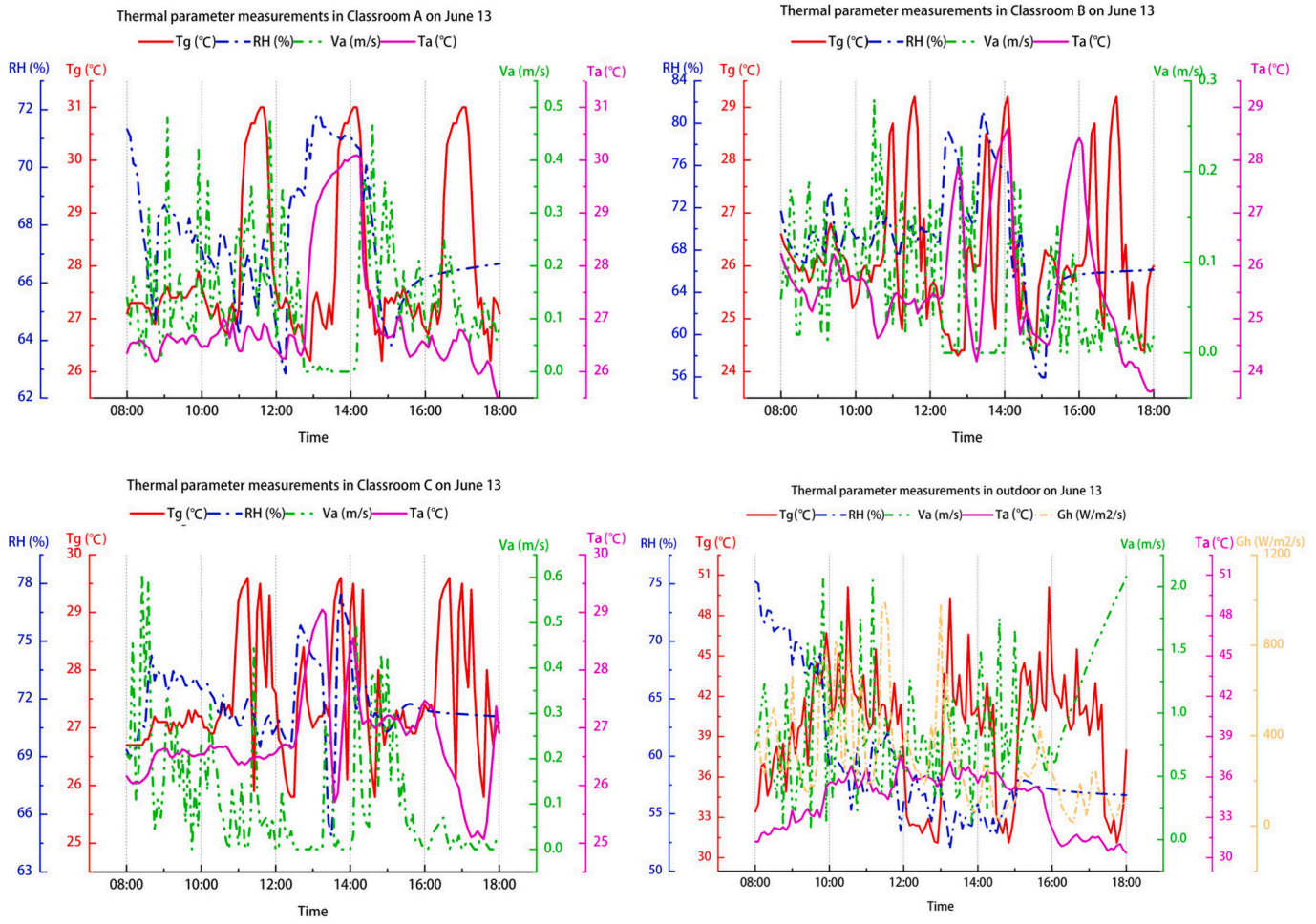


Fig. 10. Measured Results of Thermal Environment in Summer Season (June 13).

followed a similar pattern. The pronounced T_g fluctuations are primarily driven by dynamic solar radiation intensity. Indoor *RH* averaged 67.5 to 70.0%, slightly higher than the outdoor mean (60%), and V_a remained low (0.10 to 0.15 m/s) compared with outdoor airflow (0.80 m/s), indicating limited natural ventilation. Overall, mechanical cooling effectively reduced thermal loads and suppressed radiative and convective fluctuations, though differences in solar exposure, shading, and geometry still produced measurable inter-classroom variations.

(3) Thermal performance of winter season

The indoor and outdoor thermal conditions of three classrooms during the winter season (Dec. 19) are shown in Fig. 11, a cold overcast day with light rainfall. Outdoor T_a ranged 8.55–9.93 °C with high *RH*, and horizontal solar radiation peaked at 399.9 W/m^2. Indoors, with all rooms fully enclosed and no heating or ventilation, T_a averaged 19.4, 17.0, and 14.6 °C in Classrooms A, B, and C, respectively, and T_g followed the same pattern, highest in A (20.6 °C) and lowest in C (15.0 °C), reflecting differences in solar gain, shading, and orientation. Notably, short-term oscillations in indoor T_a and T_g were associated with unstable weather conditions under a cold front. Indoor RH fluctuated in a broadly synchronized pattern with these thermal variations, indicating coupled changes in both temperature and moisture conditions. Indoor *RH* (52.5 to 62.5%) remained below the outdoor mean (75%), and indoor V_a was extremely low (0.05 m/s), far lower than outdoors (1.20 m/s), indicating minimal air exchange. Overall, winter conditions produced cold and humid outdoor environments, while indoor temperature variations were mainly driven by solar exposure and envelope performance under enclosed, low-ventilation conditions.

(4) Illuminance Measurements

The measured illuminance results for the three classrooms across the three test dates are summarized in Fig. 12. The visualized distributions clearly show strong spatial and seasonal variability.

The three classrooms exhibited clear daylighting disparities across seasons, with illuminance levels peaking in summer and declining in winter. Classroom A consistently achieved the highest average illuminance (173 to 214 lx), Classroom B showed intermediate values (147 to 161 lx), while Classroom C remained the lowest (50 to 82 lx). The proportion of $DF \geq 3\%$ followed the same trend: highest in Classroom A (4.71 to 12.94%), limited occurrence in Classroom B (up to 7.06%), and consistently 0 in Classroom C. Despite meeting the prescriptive window-to-floor ratio requirements, Classroom C failed to reach the 3% daylight factor threshold, indicating that static standards cannot adequately account for the self-shading effects of complex building forms. Illuminance uniformity was generally low across all classrooms (0.10–0.42), reflecting limitations of side-window daylighting.

(5) Luminance Measurement and Glare Evaluation

Table 4 summarizes the luminance and Daylight Glare Probability (*DGP*) values of the three classrooms across three seasons. Measurement points included window centers and inter-fenestration piers to assess luminance contrast ratios. At all window-facing positions (1/4, 1/2, 3/4 depth), *DGP* remained below 0.35 in all seasons, indicating no perceptible glare [48]. *DGP* was highest in summer due to stronger solar radiation (e.g., Classroom A *DGP* up to 0.348), and lowest in winter.

Among the classrooms, A consistently exhibited the highest luminance and *DGP* because of its south-facing, unobstructed façade, while C showed the lowest values, reflecting strong shading and insufficient daylight. Across all rooms, *DGP* decreased with distance from the

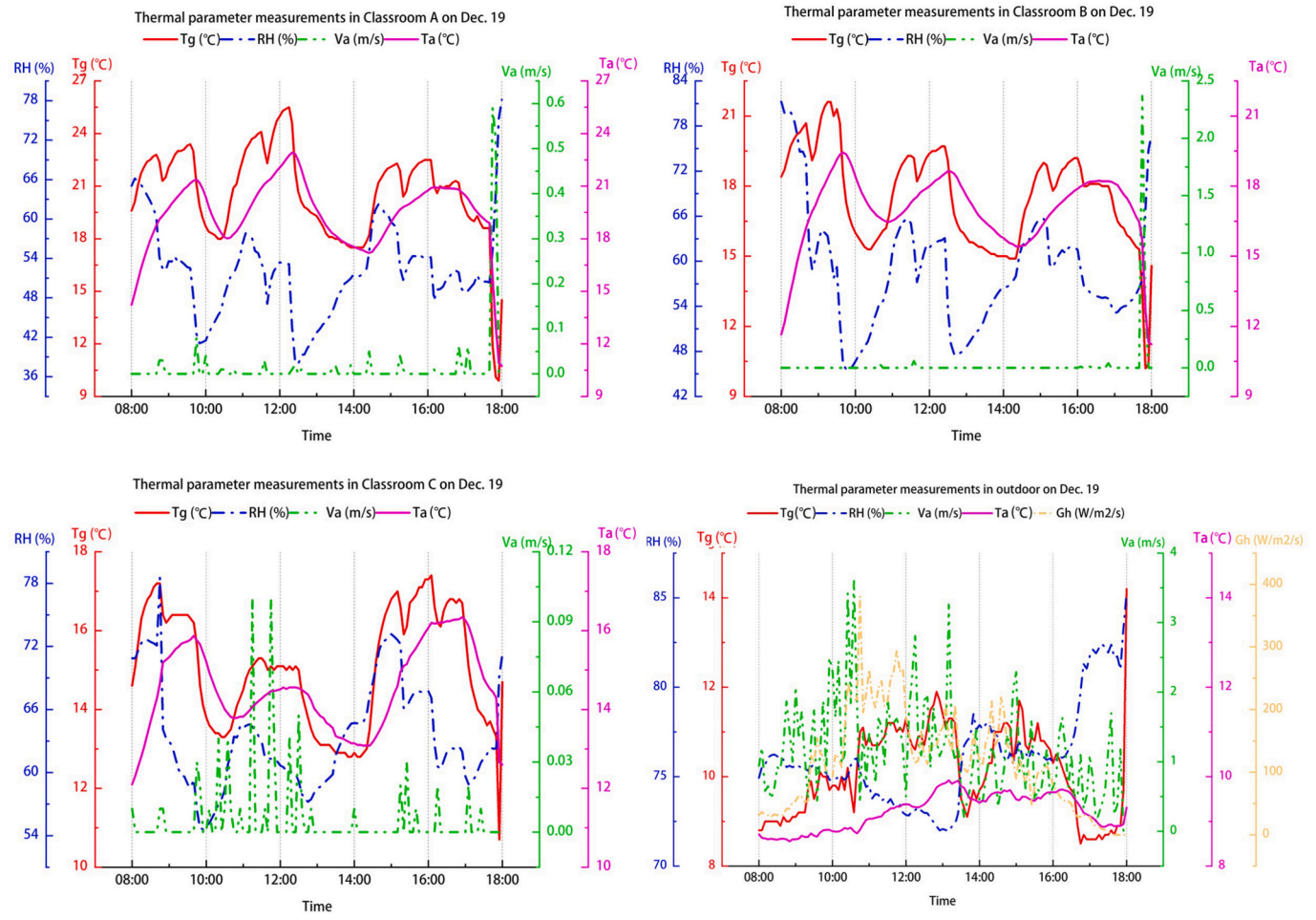


Fig. 11. Measured Results of Thermal Environment in Winter Season (Dec. 19).

window as the visible sky component diminished. The wall-between-windows areas and blackboards had low *DGP* (0.17 to 0.20) indicating minimal glare risk.

3.1.2. *Results of the Thermal Comfort Questionnaire*

The surveyed students showed a balanced gender distribution (48.85% male, 51.15% female) and were aged 12 to 17 years, with most between 13 and 15 (57.23%). Clothing insulation remained similar in the transitional season and summer (0.61 and 0.60 clo) and increased to 0.92 clo in winter. Students predominantly engaged in seated learning activities (1.2 met), with occasional light communication and brief standing (1.0 to 1.4 met).

TSV and *TPV* distributions are shown in Fig. 13. In the transitional season, "neutral (0)" dominated (61.1%), followed by "slightly cool (–1)" (26.0%). In summer under air-conditioning, "neutral (0)" was 47.3%, with "slightly cool (–1)" at 28.5%. Winter responses were cooler, with "slightly cool (–1)" (38.1%) and "cool (–2)" (23.7%) predominating. Overall, "neutral (0)" was most frequent (40.9%), indicating generally acceptable conditions, though winter discomfort was notable. *TPV* indicated most students preferred "no change" in the transitional season and summer (68.5% and 62.4%), while 59.9% preferred warmer conditions in winter.

Beyond thermal comfort, students' perceptions of indoor humidity, airflow, and daylighting were surveyed. Statistical analysis of students' humidity, airflow, and daylighting perception is presented in Fig. 14.

Subjective sensory perceptions were evaluated using a 7-point Likert scale to quantify students' environmental experiences. Humidity perception indicated that most students reported comfortable conditions during the transitional season (77.86%) and summer (72.73%), whereas comfort levels decreased markedly in winter (51.91%). In winter, 30.6% of respondents perceived the indoor environment as "slightly dry" and expressed a preference for increased humidity. These results suggest that students generally adapted well to higher humidity in warm seasons, while lower humidity in winter was more likely to cause discomfort. Perceived airflow comfort was moderate across seasons, with higher comfort levels in the transitional and summer seasons (approximately 65%) and lower levels in winter (about 44%). The perception of "slightly low" airflow was reported in all seasons. Airflow preferences were largely consistent with these perceptions, with most students indicating "no change," and increased airflow preferred primarily when airflow was perceived as insufficient. Daylighting conditions were generally rated as adequate, with reported adequacy increasing from the transitional season (68.46%) to summer (77.44%) and winter (83.43%). Daylighting preferences aligned closely with these perceptions, as most respondents preferred no change in illuminance (68.7%, 81.1%, and 86.26%, respectively). Glare evaluations further indicated favorable visual conditions, with "no glare" reported by the majority of students in all seasons (74.81%, 71.52%, and 77.3%). Summer exhibited a slightly higher incidence of glare (28.49%) compared with the transitional (25.19%) and winter seasons (22.7%), which is likely associated with stronger solar radiation.

3.1.3. *TSV modeling*

This subsection examines students' *TSV*, which reflect subjective thermal comfort under measured indoor conditions. *TSV* data were used to establish regression-based models that quantitatively link

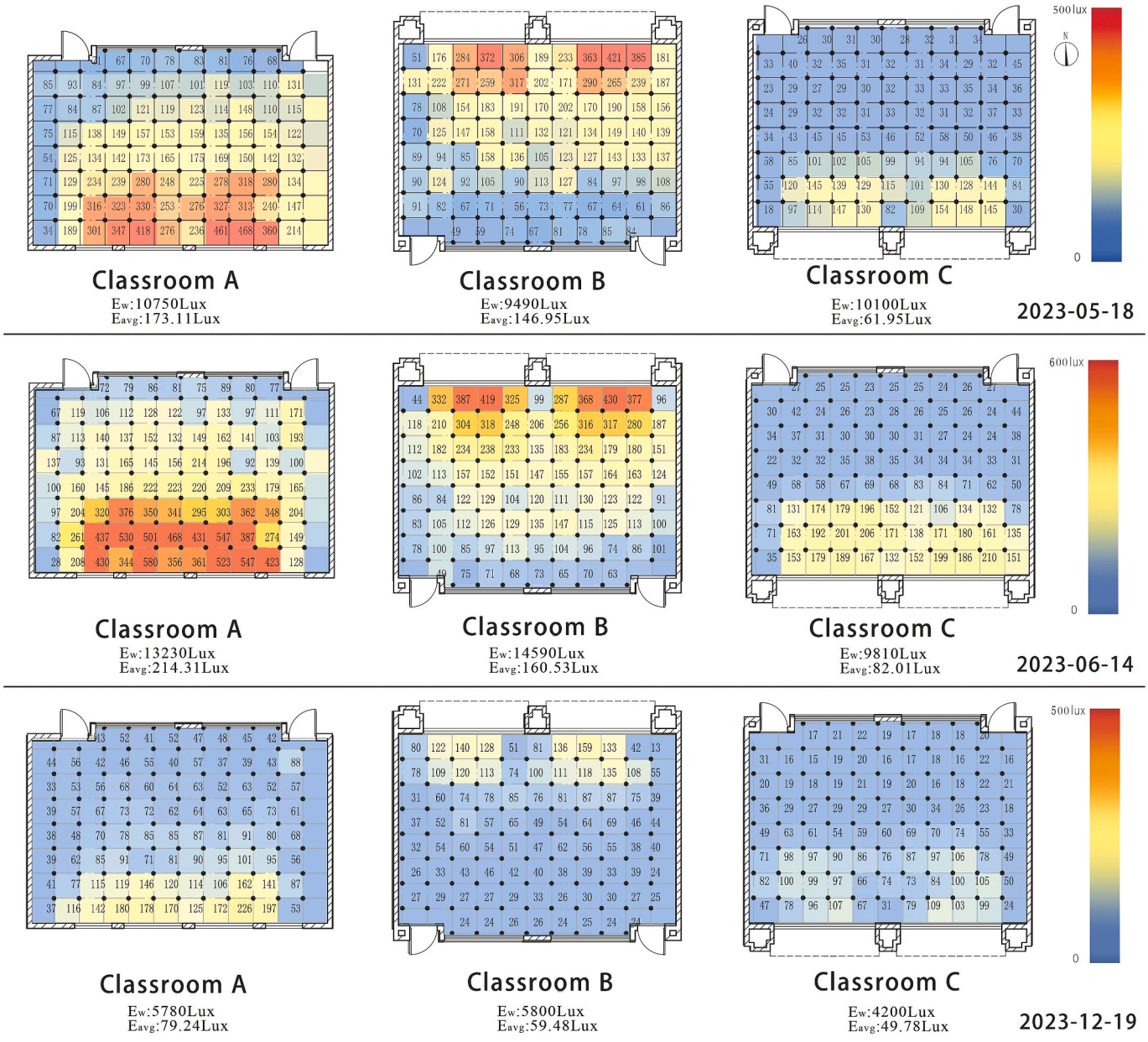
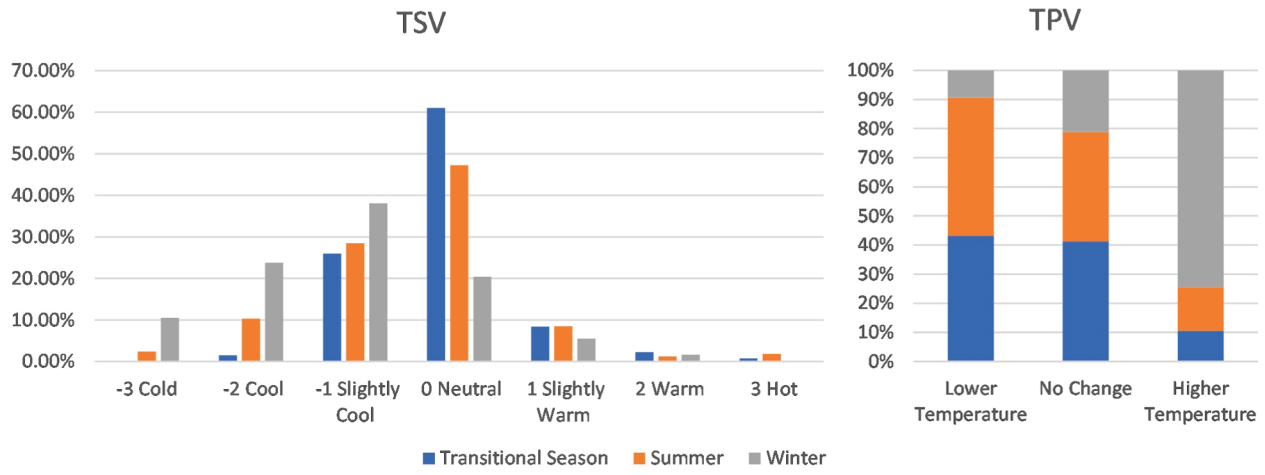


Fig. 12. Illuminance measurements for three classrooms and the outdoor environment.

Table 4
Luminance and *DGP* measurements of three classrooms across seasons.

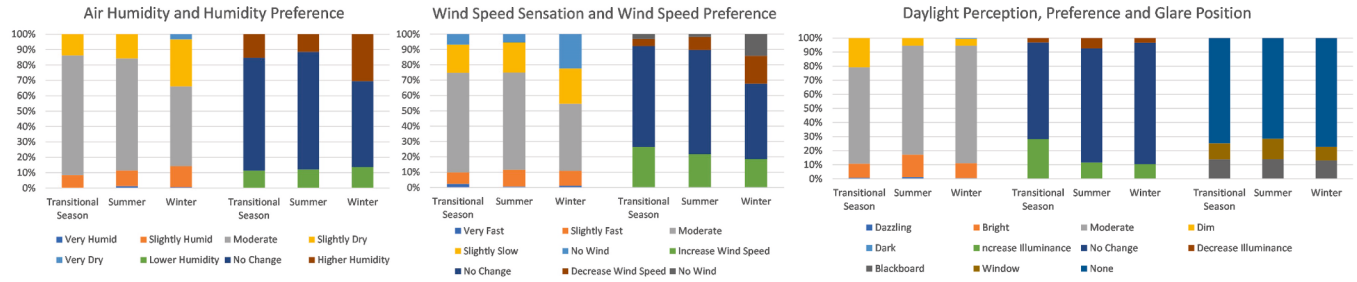
Time		May 16			June 13			June 13		
Classroom		A	B	C	A	B	C	A	B	C
Window	$L_s1/4$(cd/m^2)	4377	1922	656	5920	4023	712	2198	644	382
	$L_s1/2$(cd/m^2)	4272	1925	640	5784	4011	694	1994	648	362.4
	$L_s3/4$(cd/m^2)	4151	1831	588	5726	3842	616	2174	595.3	388.4
	$E_v1/4$(lx)	635	515	270	1062	617	390	283	262	194
	$E_v1/2$(lx)	588	519	157	720	529	233	296	240	138
	$E_v3/4$(lx)	581	495	148	826	643	153	306	168	92
	*DGP*1/4	0.337	0.303	0.255	0.348	0.353	0.245	0.321	0.255	0.235
	*DGP*1/2	0.313	0.280	0.260	0.330	0.334	0.244	0.287	0.239	0.229
	*DGP*3/4	0.288	0.258	0.237	0.302	0.302	0.239	0.269	0.232	0.236
Inter-fenestration Pier	L_s(cd/m^2)	560	415	161	731	497	247	286	171	214
	E_v(lx)	58.36	37.4	23.43	65.33	53.64	29.67	27.85	16.38	14.86
	DGP	0.193	0.184	0.170	0.203	0.189	0.175	0.177	0.170	0.173
Blackboard	L_s (cd/m^2)	43	31	30	49	36	24	18	16	16
	E_v(lx)	5.77	4.8	4.66	6.58	5.5	3.6	2.5	2.4	2.46
	DGP	0.163	0.162	0.162	0.163	0.162	0.161	0.161	0.161	0.161

*Subscripts 1/4, 1/2, and 3/4 denote measurement positions located at 25%, 50%, and 75% of the classroom depth from the window-side wall, respectively.



(a) *TSV* (b) *TPV*

Fig. 13. Distribution of *TSV* (a) and *TPV* (b).



(a) Humidity and Preference (b) Wind Speed and Preference (c) Daylighting, Glare, and Preference

Fig. 14. Seasonal distributions of students' perceived and preferred indoor humidity (a), airflow (b), and daylighting/glare (c). Note: Sensory categories (e.g., "Very Fast," "Slightly Humid") are defined by a 7-point Likert scale used to quantify the intensity of environmental stimuli. Preference categories (e.g., "Increase," "No Change") indicate the desired directional adjustment relative to the current indoor state.

environmental variables to occupants' thermal perception.

(1) Subjective–Objective Coupling Analysis

TSV matched with thermal performance indicators shows seasonal *TSV* distributions aligned closely with measured indoor conditions. *TSV* clustered near −1 to 0 in the transitional season, remained near neutral in summer under air-conditioning, and shifted toward cold sensations (−2 to −1) in winter. At the classroom scale, Classroom A showed the highest neutrality in May (70.27%), decreasing to 47.22% in winter; Classroom C exhibited the strongest cold perception in winter (approximately 40% slightly cold and 34.3% cold) consistent with the lowest T_a (14.0 °C). These patterns indicate that group-level *TSV* reliably captures seasonal shifts in indoor comfort. Pearson correlation analysis quantified subjective–objective coupling. *TSV* correlated strongly with T_g (r = 0.95) and T_a (r = 0.91), moderately with *RH* (r = 0.58), and weakly (negative) with V_a (r = −0.23). Thus, temperature variables dominate thermal sensation, while humidity and air movement act as secondary modifiers. These relationships support the validity of the coupled dataset for regression modeling.

(2) TSV Regression Modeling

A MLR model was constructed with *TSV* as the dependent variable and T_a, T_g, *RH* and V_a as predictors. The regression analysis conducted using SPSS produced the following *TSV* equation:

$$TSV = -5.0603 + 0.0293 \times T_a + 0.0959 \times RH + 0.2092 \times T_g - 1.3132 \times V_a \quad (1)$$

where T_a is air temperature (°C), *RH* is relative humidity (%), T_g is globe temperature (°C), and V_a is air speed (m/s).

Multicollinearity was assessed using VIF. All predictors showed VIF were below 5, specifically T_a (VIF = 4.84), T_g (VIF = 4.78), *RH* (VIF = 1.01), and V_a (VIF = 1.07), indicating no severe multicollinearity and stable coefficient estimation. To assess coefficient stability under potential multicollinearity, 95% confidence intervals were computed for all predictors. The intervals were narrow and excluded zero: T_a [0.005, 0.054], *RH* [0.081, 0.111], T_g [0.180, 0.238], and V_a [-1.512, -1.114]. indicating statistically significant and stable estimates ($p < 0.05$).

The MLR equation rapidly identifies dominant thermal-sensation drivers, with its coefficients providing an interpretable diagnostic of *TSV* sensitivity to T_a, T_g, *RH*, and V_a. T_g exhibits the strongest influence, typically accounting for more than 60% of the total contribution, confirming radiant temperature as the primary driver of thermal sensation. *RH* contributes roughly 20% or more, positioning humidity as the secondary factor affecting *TSV*. Model performance was evaluated using R^2,

Table 5
TSV Model Performance

Dataset	R^2	MAE	MSE	RMSE
Overall	0.921	0.254	0.094	0.307
Transition season	0.629	0.234	0.079	0.282
Summer	0.862	0.242	0.069	0.263
Winter	0.834	0.267	0.104	0.322

RMSE, and cross-validated MAE (Table 5). The MLR model achieved an overall R^2=0.9205, demonstrating strong explanatory power and reliable predictive capability for *TSV* estimation.

TSV modeling provides a quantitative representation of classroom thermal sensation and can be encapsulated in Ladybug Tools as a performance target for rapid simulation and parametric analysis. It should be noted that TSV is inherently an ordinal and bounded variable. In this study, MLR is adopted as a design-oriented approximation to enable continuous prediction and seamless integration with simulation and optimization workflows. Although R^2 is lower for the transition season (0.629) due to narrower temperature ranges and stochastic occupant behavior, a unified MLR equation is prioritized to ensure a consistent mechanism for MOO and prevent fragmented Pareto fronts. The robust overall R^2 (0.921) and stable MAE across all subsets (0.234–0.267) confirm the model's reliability in capturing core thermal sensation physics across diverse operational modes.

3.2. Parametric Generation of Building Forms

This study developed a rule-based parametric generator to create large sets of classroom-building forms representative of Guangdong's secondary schools. The system formalizes key geometric and functional constraints into parameter rules, ensuring that every generated form meets realistic design and regulatory conditions while remaining compatible with subsequent simulation workflows.

3.2.1. Generation of Building Forms

The parametric-generation plugin enables automated generation of classroom building forms under predefined geometric and functional constraints. The plugin allows users to define key parameters and generates a wide range of valid classroom building configurations automatically. Based on these parameters, the plugin automatically produces 3D building models that accurately reflect the required spatial configuration and massing conditions, demonstrating its capability for efficient exploration of design alternatives. This implementation provides a practical tool for rapid performance-oriented design exploration.

Using the parametric plugin, a diverse set of classroom building forms was generated, each conforming to the predefined geometric and functional constraints. The resulting forms cover a wide range of configurations in terms of classroom layout, building massing, and façade arrangement, demonstrating the system's capability to explore alternative design solutions systematically. Each generated form is encoded with a complete set of design parameters, enabling traceability and reproducibility for subsequent analyses. The generated morphologies are illustrated in Fig. 15, showing the diversity of spatial layouts and overall building forms produced by the plugin. This indicates that the parametric system can effectively generate valid building models that are ready for subsequent daylighting and thermal performance simulations.

3.2.2. Integration with Ladybug

The generator was embedded in Grasshopper and directly coupled with NSGA-II and Ladybug workflows, enabling seamless transfer of generated building forms into daylighting and thermal simulations. This integration supports automated batch simulation and optimization, enabling a clear mapping between geometric variations and predicted indoor environmental conditions. As a result, a continuous and reproducible workflow was established, linking parametric form generation, performance evaluation, and MOO, and providing a robust foundation for performance-driven early-stage design analysis.

3.3. Simulation and MOO

3.3.1. Model accuracy verification

Using the developed form-generation plugin and the integrated daylighting and thermal simulation workflow, the measured classroom was modeled and simulated, and the results were validated against field measurements, confirming sufficient accuracy for large-scale parametric generation and performance evaluation.

(1) T_a Validation

The simulated indoor T_a was compared with field-measured values across the three classrooms and three representative seasons (Fig. 16).

The simulated and measured indoor T_a values exhibit consistent overall trends, with simulations slightly higher than measurements throughout most periods. During the morning hours (08:00–12:00), deviations remained minimal, indicating good calibration accuracy under relatively stable occupancy and ventilation conditions. Larger discrepancies occurred in the afternoon (14:00–18:00), primarily due to unmodeled factors such as intermittent window opening, variable internal heat gains, and occupant movement. Around midday (12:00–14:00), when classrooms were unoccupied and air conditioning was turned off, the measured T_a gradually approached outdoor conditions, whereas the model did not fully capture this rapid temperature decline. Across the three observation dates, deviations between simulated and measured T_a remained within $\pm 2^{\circ}C$ for all classrooms. Classroom C showed slightly larger deviations in the transitional and winter seasons (1.42% and 1.13%), while summer deviations were smaller, as the simulation better captured the stable air-conditioned environment. The results confirm that the model and simulation workflow can reliably reproduce indoor thermal conditions.

(2) *RH* Validation

RH was simulated for the three monitored classrooms and compared against measured data (Fig. 17).

The results show that the simulated *RH* in Classrooms A and B is generally lower than the observations, suggesting that the assumed ventilation and airflow conditions in the model did not fully capture the actual air exchange rate. Both classrooms are located on higher floors, where increased interaction with outdoor air led to consistently higher measured *RH* levels. In contrast, Classroom C demonstrates good

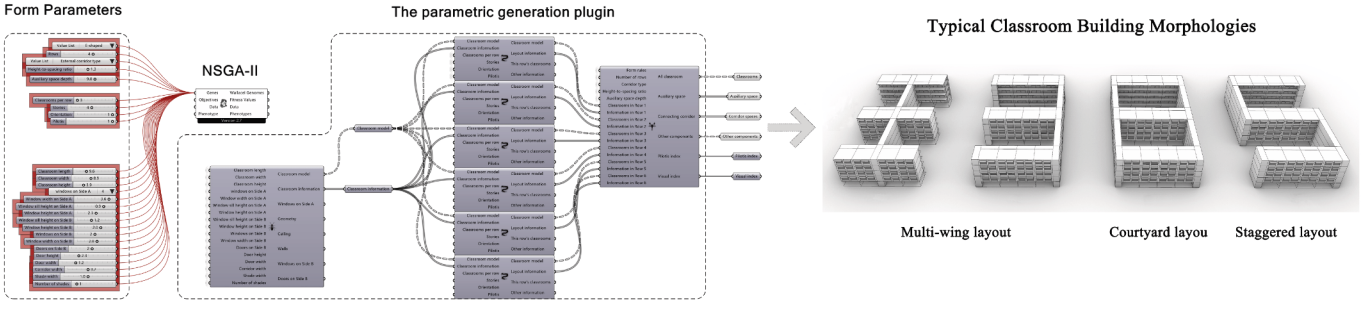
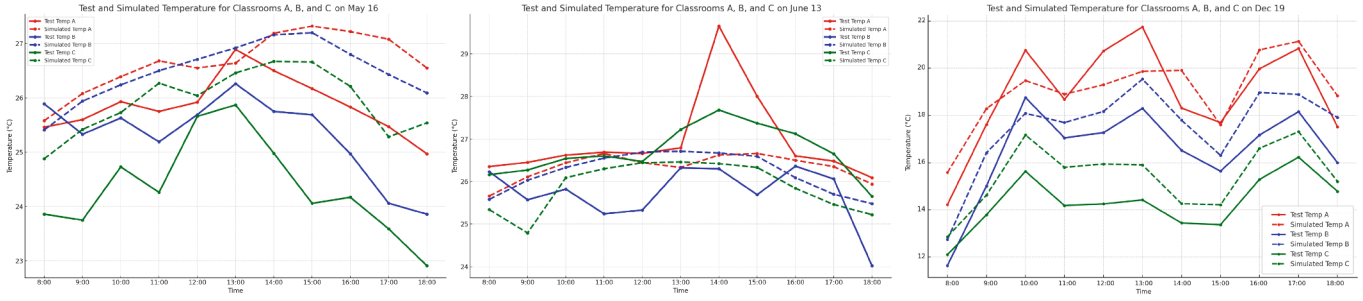
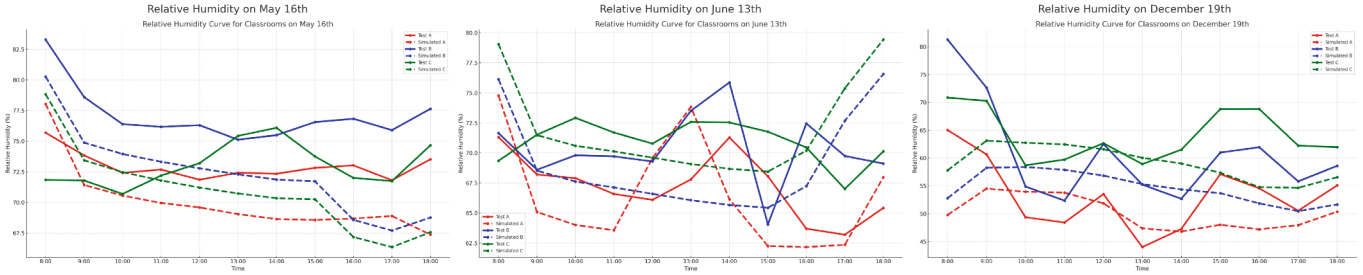


Fig. 15. Plugin-based generation of parametric building form models.



(a) Transitional T_a Validation (b) Summer T_a Validation (c) Winter T_a Validation

Fig. 16. Simulated vs. Measured T_a for Three Classrooms.



(a) Transitional *RH* Validation (b) Summer *RH* Validation (c) Winter *RH* Validation

Fig. 17. Simulated vs. Measured *RH* for Three Classrooms.

agreement between simulation and measurement. Its lower-floor location and adjacency to semi-enclosed atrium spaces reduce exposure to fluctuating outdoor moisture conditions, resulting in more stable and accurate simulation performance. Deviations between simulated and measured *RH* for all classrooms were mostly within 5%, indicating that the simulation can accurately reproduce actual indoor humidity conditions. Deviations were slightly higher in winter, ranging from 5% to 10%, showing some discrepancy with measured data, but the results remain within a reliable range suitable for most practical applications.

(3) Point-in-time illuminance (*PIT*) Validation

The onsite measurement time was used to simulate *PIT* for the three classrooms. The measurements were conducted under controlled conditions: only daylight (lights off), windows with typical curtain positions, between 10:00–14:00 on predominantly overcast days. The simulated results distribution is shown in Fig. 18.

The *PIT* simulation results show that the overall illuminance distribution trends in the three classrooms are consistent with field measurements, although simulated values are generally higher. This discrepancy is mainly due to real-world factors such as variations in interior reflectance, furniture layout, and curtain shading. To evaluate model accuracy, 792 data samples (88 grid points × 3 classrooms × 3 seasonal time points) were analyzed using Mean Bias Error (MBE) and CV(RMSE) in accordance. The results (MBE = 4.2%, CV(RMSE) = 14.7%, $R^2 = 0.91$) fall within acceptable thresholds, indicating reliable simulation performance. The relatively low measured illuminance levels (e.g., 173–214 lx in Classroom A) are attributed to overcast sky conditions during 10:00–14:00 with no electric lighting and open curtains, representing a conservative baseline. These point-in-time measurements were used to calibrate key Radiance parameters, including material reflectance, sky models, and sensor grid configuration, enabling subsequent climate-based simulations for DA and UDI. In addition, luminance and DGP measured under clear-sky noon conditions capture peak glare scenarios, supporting the extension from point-in-time validation to annual visual comfort assessment.

Overall, the comparison between simulated and measured values of representative daylighting and thermal indicators (T_a, *RH*, and *PIT*) confirms the accuracy and reliability of the model and its simulation workflow.

3.3.2. MOO

The optimization procedure produced a diverse set of Pareto-optimal solutions balancing daylighting and thermal objectives. The resulting Pareto front reveals clear trade-offs among daylighting quality (*DA*, *UDI*), thermal comfort performance (*TSV*, *SET*), and energy use for cooling and lighting, highlighting the inherent trade-offs among these multidimensional objectives. Compared with the full design space, Pareto solutions exhibit a more concentrated distribution, indicating that only a subset of geometric configurations can simultaneously achieve acceptable daylighting and thermal performance.

(1) Cluster Analysis of Pareto Solutions

Cluster analysis of the Pareto set reveals three representative solution groups, each characterized by distinct geometric tendencies (Fig.19). Cluster 1 achieves strong daylighting performance but shows relatively higher cooling energy use. Its forms are generally open in plan, often incorporating stepped massing with the largest height–depth ratios. Cluster 2 demonstrates lower cooling energy demand but comparatively weaker daylighting performance. These solutions are characterized by compact courtyard-type configurations with multiple parallel building rows. Cluster 3 constitutes the majority of solutions and presents a balanced performance across daylighting and thermal metrics. Its forms are characterized by medium-rise, multi-row configurations with minimal stepping. The dominant form types are E-shaped linear blocks and courtyard layouts, which provide a combined advantage in daylight access, ventilation, and cooling-load control.

Gene-pattern analysis (i.e., morphological parameter combinations) of Cluster 3 reveals a set of dominant morphological tendencies shared by a large proportion of high-performing Pareto solutions. In this cluster, mid-rise configurations are most commonly observed (5-6 stories),

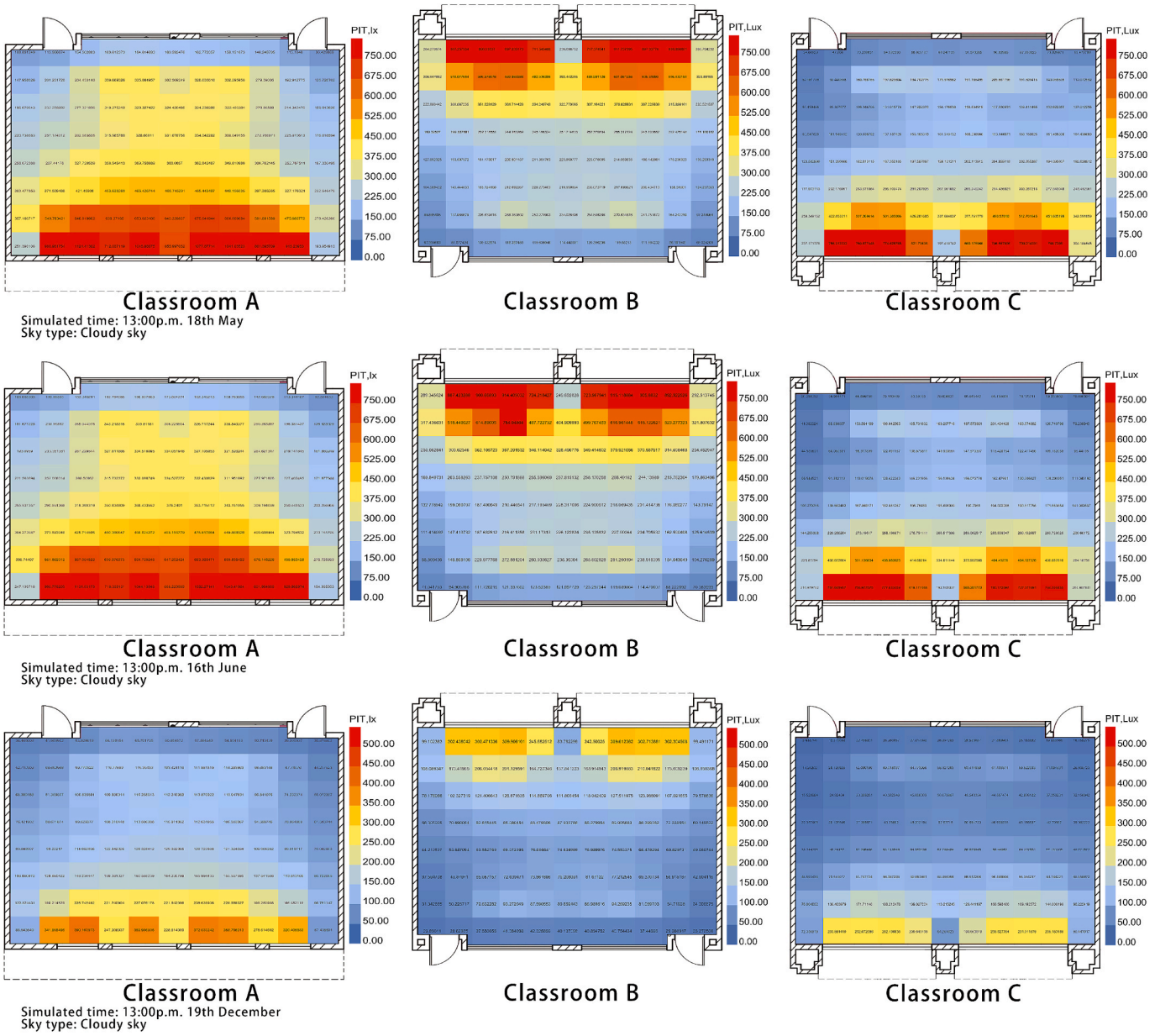


Fig. 18. Simulated *PIT* for Three Classroom.

accompanied by multi-row layouts (4-6 rows) with a moderate number of classrooms per row (3-5 classrooms per row). Building arrangements are generally compact (height-to-spacing ratio of 1.2-1.4), with limited use of setback operations (0-1 setback levels), and massing configurations frequently exhibit articulated forms such as E-shaped and courtyard typologies. Together, these features characterize a prevailing spatial organization pattern within this cluster, reflecting recurring geometric traits among Pareto-optimal solutions.

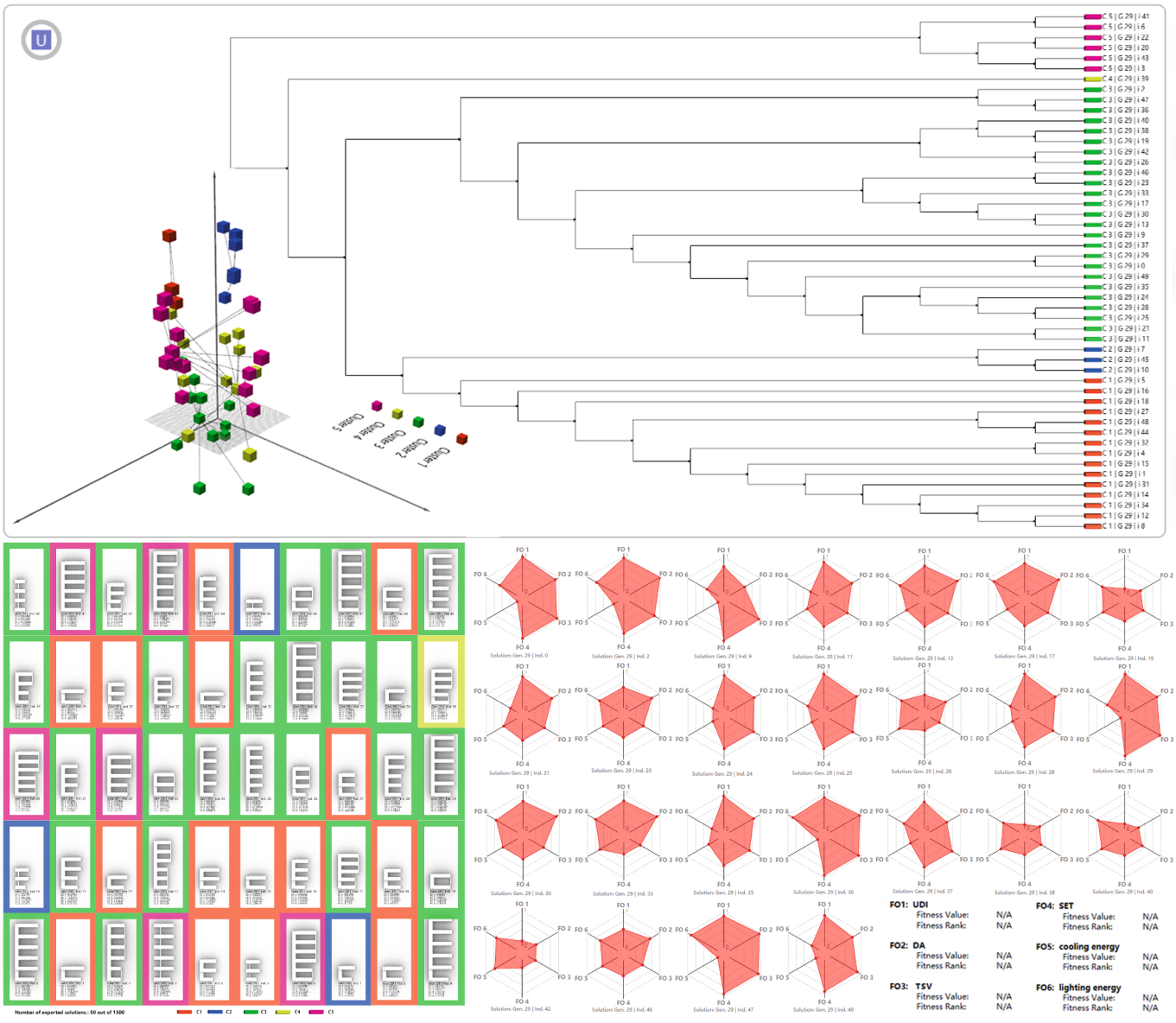
(2) Sensitivity Analysis of morphological variables

A one-at-a-time sensitivity analysis was conducted to quantify how each morphological parameter influences performance outcomes. Importance weights were used to represent sensitivity (Table 6).

The sensitivity results reveal a clear hierarchy of morphological parameters in governing both daylighting and thermal performance. The number of stories, number of rows, and classrooms per row exhibit the strongest influence on cooling energy, with absolute variation rates of 15.34%, 12.81%, and 13.83%, respectively. Height-to-spacing ratio shows the highest sensitivity for *DA*, with a variation rate of 10.88%. Setback levels moderately improve daylighting but have minimal effects on thermal indicators, while the generative rule type mainly affects daylight-related metrics and contributes little to thermal performance. These results quantify the influence of different design variables on performance indicators, providing a basis for identifying the PCZ and guiding the prioritization of parameters in the DIS.

(3) PCZ and DIS

Building upon the Pareto clustering results, a PCZ was further identified through cross-cluster comparison and robustness-oriented sensitivity analysis. Unlike the cluster-specific patterns, the PCZ extracts parameter ranges that consistently support balanced daylighting, thermal comfort, and energy performance across multiple clusters and remain stable under variations in layout and massing. The PCZ is characterized by buildings with 5–6 stories, 4–6 rows, and 3–5 classrooms per row, combined with 0–1 setback levels and a height-to-spacing ratio of 1.2–1.4. These ranges represent a compromise region in which sufficient façade exposure for daylighting is maintained without excessive envelope area that would penalize thermal and energy performance. The limited use of setbacks and the moderate height-to-spacing ratio further contribute to performance robustness by



(a) Cluster Analysis of Pareto set (b) Cluster 3 Radar Chart

Fig. 19. Pareto Set Clustering (a) and Cluster 3 Radar Chart(b).

Table 6
Importance weights of morphological parameters(%)

Design Parameter	*DA*	*UDI*	*TSV*	*SET*	Cooling Energy	Lighting Energy
Number of stories	-7.08	-1.20	-2.65	-0.52	-15.34	1.71
Number of rows	-7.72	-1.21	-2.83	-0.51	-12.81	1.98
Classrooms per row	-0.06	-0.21	-2.38	-0.44	-13.83	3.66
Setback levels	7.63	1.63	0.35	0.07	2.66	-0.48
Generative rule type	7.32	1.13	0.07	0.02	0.21	2.30
Height-to-spacing ratio	10.88	1.71	0.05	0.01	0.10	-4.82

reducing sensitivity to geometric perturbations. Accordingly, the PCZ defines a transferable design envelope rather than a cluster-dependent optimal configuration.

Based on the sensitivity ranking of parameters, the DIS was established to guide performance-driven adjustments in a defined sequence. Parameters governing thermal comfort and cooling energy are adjusted first, including the number of rows, number of stories, and classrooms per row. Daylighting performance is addressed subsequently by adjusting the height-to-spacing ratio, followed by setback levels and then the generative rule type. This ordered adjustment sequence prioritizes the most influential parameters and supports efficient iterative performance optimization.

3.4. ML-based evaluation and interpretability

3.4.1. Prediction performance

Comparative results across the ML models indicate that GBR provides the most accurate and stable performance among all candidates (Table 7). Accordingly, GBR was selected as the primary model for subsequent evaluation and interpretability analysis.

The GBR model trained on the multi-objective simulation dataset

Table 7
Overall Prediction Performance of ML Models (means)

Model	R^2	MAE	MSE	RMSE	Notes
MLR	0.52	0.182	0.091	0.301	Linear assumption limits accuracy.
SVR	0.68	0.132	0.059	0.243	Moderate; limited nonlinear fitting capability.
RF	0.77	0.105	0.043	0.207	Captures nonlinear relationships; stable.
ANN	0.80	0.097	0.036	0.190	Good overall performance and generalization.
GBR	0.81	0.080	0.029	0.171	Best overall; chosen as final model.

was evaluated on the test set, with accuracy metrics for each indicator summarized in Table 8.

The model achieved high predictive accuracy for *TSV*(+), cooling energy, and Lighting Energy, with R^2 values exceeding 0.90. This result demonstrates a strong capability to capture the nonlinear relationships between morphological features and thermal–energy performance. In contrast, prediction accuracy for daylighting indicators (*DA* and *UDI*) was relatively low. This suggests that daylighting performance is more sensitive to geometric configuration and may require additional spatial or climatic descriptors to improve model robustness. The strong performance observed for thermal indicators, particularly cooling energy and *TSV*, can be attributed to their dependence on high-sensitivity parameters identified in the Parameter Influence Hierarchy. As these parameters also define PCZ boundaries, the ML model benefits from inherent structural regularity in the dataset. By contrast, daylighting indicators exhibit weaker predictability because they are more strongly influenced by localized spatial configurations, which play a less dominant role in PCZ formation.

3.4.2. Model Interpretation

SHAP-based global interpretation was performed to elucidate the internal decision logic of the GBR model. The SHAP summary plots, feature rankings, and summary plot (Fig. 20) consistently indicate that model predictions are governed by a clear hierarchy of morphological parameters.

Beeswarm plots clarify both the directionality and dispersion patterns of key morphological parameters. Number of rows shows the widest SHAP range across all indicators, confirming its dominant and strongly nonlinear influence. For *DA*, higher row numbers are mainly associated with negative SHAP values (approximately −2.0 to −0.5), while lower values contribute positively (up to +3.0), indicating a clear trade-off between row density and daylighting availability. Conversely, for cooling energy, higher row numbers consistently yield positive SHAP values (up to +0.6), whereas lower values reduce cooling demand. Number of stories and Classrooms per row exhibit more compact but directional SHAP distributions, generally contributing positively to *SET* and *TSV*(+) while increasing cooling energy, with magnitudes typically within ±0.15. Other parameters, including setback levels, generative rule type, and height-to-spacing ratio, show tightly clustered SHAP values around zero, indicating marginal influence. Feature-importance rankings further confirm this hierarchy. Number of rows has the strongest and most consistent impact across all metrics, with mean SHAP values of 1.16 (*DA*), 0.16 (*SET*), 0.05 (*TSV*(+)), and 0.23 (cooling energy), confirming its dominant role in simultaneously shaping daylighting availability and thermal performance. Number of stories and Classrooms per row form the second tier, with notable contributions to thermal indicators—cooling energy (0.13 and 0.10) and SET (0.08 and 0.07)—while their daylighting effects remain moderate (<0.15). In contrast, Setback levels, Generative rule type, and Height-to-spacing ratio consistently show low mean SHAP values (≤0.02), indicating limited explanatory power within the explored design space. Summary plots further confirm these findings. Number of rows exhibits stable and large contributions across samples, reaching approximately ±2.1 for *DA* and ±0.48 for cooling energy. Number of stories and Classrooms per row show moderate, sample-dependent effects (typically within ±0.10–0.15 for *SET* and *TSV*(+)), while the remaining parameters remain clustered near zero (generally ≤±0.03).

The SHAP ranking aligns closely with the previously established Parameter Influence Hierarchy: parameters governing PCZ formation (rows, stories, and classrooms per row) also dominate SHAP importance, indicating that the GBR model captures the same structural relationships observed in optimization and sensitivity analyses. Overall, the SHAP analyses establish a clear hierarchy of morphological drivers, confirming that rows, stories, and classrooms per row govern most of the variance in daylighting and thermal performance.

4. Discussion

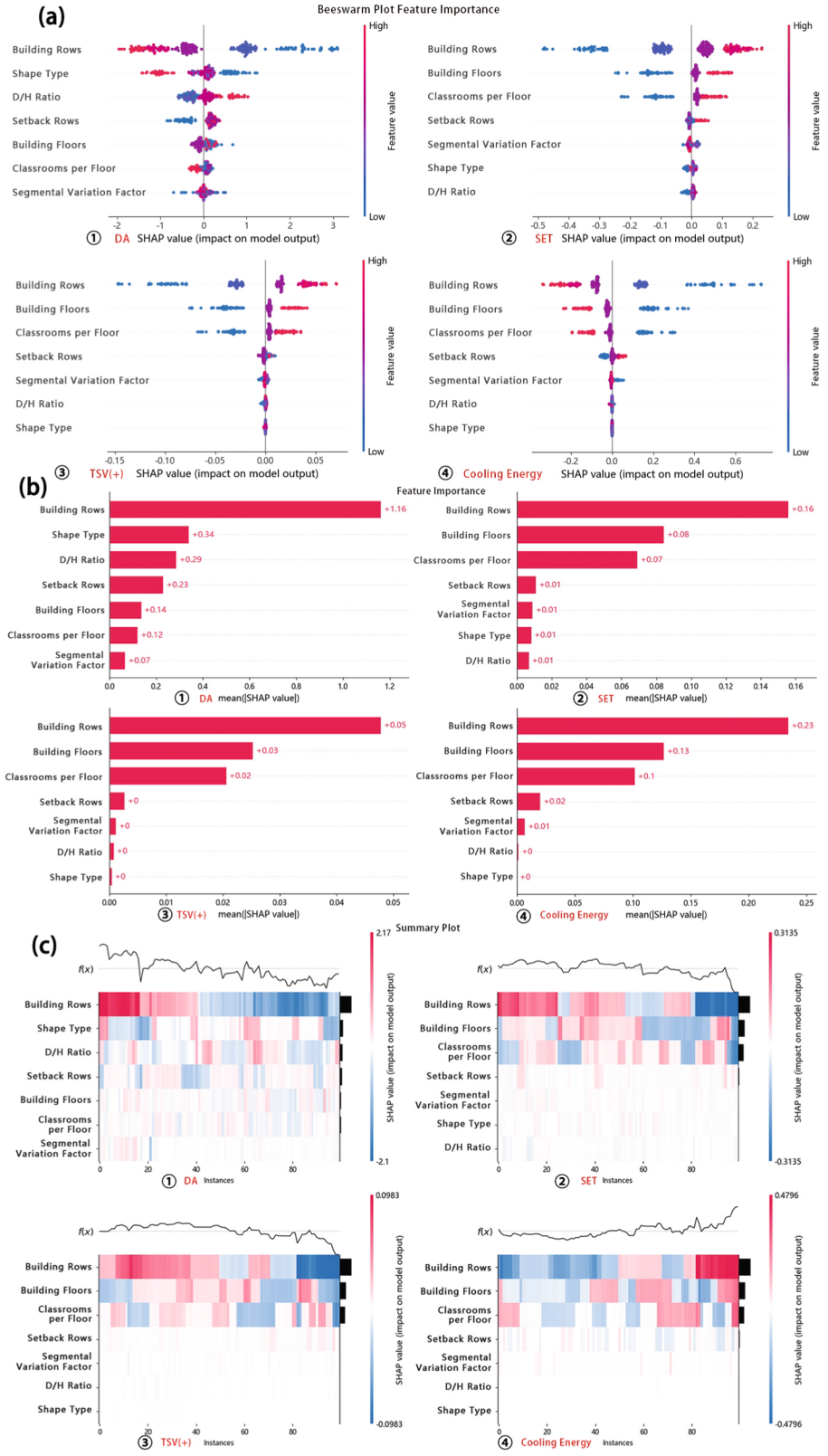
4.1. Interpretation of SHAP-Based Morphological Influence Patterns

The results reveal that building form influences performance through a clear hierarchical structure, with horizontal spatial configuration acting as the primary driver across all evaluated indicators. The SHAP-based global interpretation reveals a clear hierarchy of morphological drivers governing daylighting and thermal performance (representative mean |SHAP| = 1.16 for *DA*, 0.16 for *SET*, 0.05 for *TSV*(+), and 0.23 for cooling energy). Among all variables, number of rows consistently shows the highest SHAP importance, indicating that horizontal spatial structuring is the dominant mechanism shaping daylighting penetration, thermal exposure, and cooling demand. This dominance stems from its simultaneous control over inter-row spacing, mutual shading depth, sky-view access, and façade availability, which jointly determine solar access and radiative gains. This finding is consistent with previous studies highlighting the role of building compactness and spatial configuration in regulating solar exposure and energy demand, while extending them by quantitatively demonstrating a hierarchical dominance of horizontal layout variables.

The number of stories and classrooms per row constitute a secondary tier of influence. Their predominantly negative SHAP contributions to SET, TSV(+), and cooling energy indicate that increased vertical stacking and larger classroom groupings mitigate thermal stress by enhancing self-shading, reducing exposed envelope area, and increasing volumetric compactness. However, their effects on daylighting metrics differ: additional stories generally improve sky exposure, whereas deeper row configurations reduce daylighting autonomy, highlighting the differing sensitivities of daylighting and thermal indicators to the same morphological changes. In contrast, local geometric modifiers—including setback levels, height-to-spacing ratio, and generative rule type—exhibit consistently low SHAP contributions. These parameters primarily serve to fine-tune geometric details and do not override the dominant spatial organization defined by the number of rows, number of stories, and classroom arrangement.

Table 8
GBR prediction performance.

	DA	*UDI*	*TSV*(+)	*TSV*(-)	*SET*	Cooling Energy	Lighting Energy
R^2	0.352	0.615	0.945	0.534	0.861	0.980	0.987
MAE	1.861	0.525	0.0134	0.008	0.081	0.033	6.0×10^{-4}
MSE	5.713	0.487	3.0×10^{-4}	9.9×10^{-5}	0.009	0.002	5.6×10^{-7}
RMSE	2.390	0.698	0.017	0.01	0.098	0.042	7.0×10^{-4}



(a) Beeswarm plot (b) Feature importance (c) Summary plot

Fig. 20. SHAP-based interpretation of the GBR model.

Notably, the SHAP-derived influence hierarchy closely aligns with the dominant drivers identified through PCZ analysis and MOO. Since the SHAP analysis is based on the same simulation-derived KPIs used in the MOO process, this consistency confirms that the ML model accurately captures the underlying form–performance relationships and preserves the optimization logic within the data-driven framework. Parameters with the highest SHAP importance also govern convergence behavior, optimal form clusters, and sensitivity patterns observed during the optimization stage, confirming that the ML model captures the same underlying form–performance relationships revealed by the simulation-based workflow.

4.2. Dual Impacts of TSV on Cooling Energy and SET

Existing *TSV*-related studies primarily focus on modeling its relationship with indoor thermal environment parameters [49] or examining differences across user groups and adaptive behaviors [50]. Within these frameworks, *TSV* is generally treated as an operational evaluation or post-hoc validation indicator, rather than as a performance objective informing early-stage design. Moreover, most research remains confined to indoor or operational scales, with limited investigation into how building form and massing strategies systematically influence *TSV*, constraining its integration into performance-driven form generation. In cooling-dominated hot and humid climates, treating *TSV* as a single symmetric variable further weakens its design relevance, as warm-side and cool-side thermal sensations exhibit pronounced asymmetry in comfort perception and energy response. To address this limitation, *TSV* was explicitly decomposed into *TSV*(+) and *TSV*(−) in both the MOO and ML stages. This directional treatment improves statistical robustness and enables differentiated thermal-sensation effects to be captured. The GBR model achieves high predictive accuracy for *TSV*(+) ($R^2 = 0.918$), demonstrating that warm-side thermal sensation can be reliably learned from morphological parameters and is particularly sensitive to building-scale form characteristics.

Analysis of Pareto-optimal solutions reveals a consistent negative coupling between *TSV*(+) and cooling energy across representative clusters. In Cluster 3, a distinct low-energy band emerges at higher *TSV* (+) levels (> 0.80), with cooling energy mainly concentrated between 0.5 and 0.8 kWh/m^2, while Cluster 1 exhibits the same tendency with greater dispersion. Although this pattern appears counterintuitive, sensitivity analysis indicates that increases in key massing parameters, such as number of stories, rows, and classrooms per row, do not proportionally increase cooling demand. Instead, larger and more compact building forms distribute internal and envelope loads over a greater floor area, leading to lower cooling energy intensity per unit area. At the same time, enhanced thermal inertia and self-shading effects result in slightly higher but more spatially uniform indoor thermal conditions, reflected by increased *SET* and, consequently, higher *TSV*(+). As a result, higher *TSV*(+) values can coexist with lower cooling energy demand (Fig. 21a). This apparent contradiction challenges the conventional assumption that improved thermal comfort necessarily requires increased energy input, suggesting instead that form-induced heat redistribution can decouple comfort perception from energy intensity.

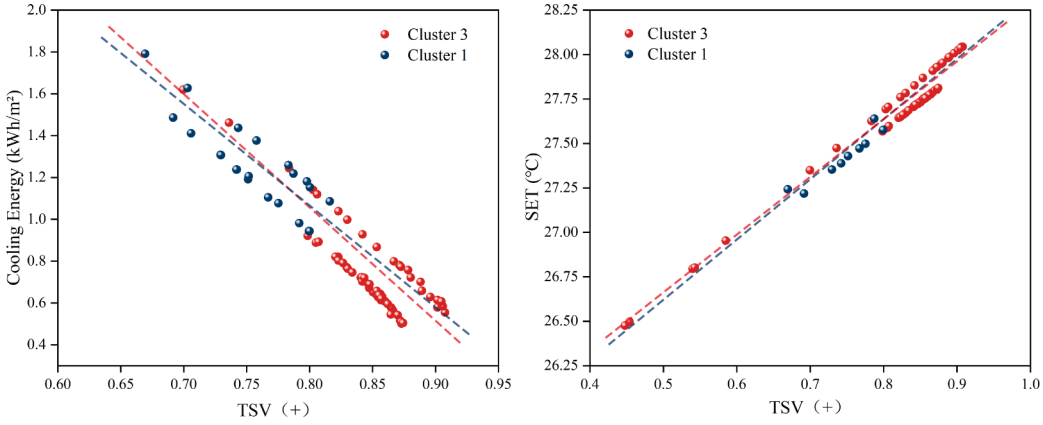
TSV(+) exhibits a strong positive correlation with SET(Fig.21b), with *SET* increasing steadily from approximately 26.5 °C to 28.0 °C as *TSV*(+) rises, following a near-linear distribution across clusters. This consistency confirms that variations in *TSV*(+) are grounded in integrated physical thermal states rather than being purely subjective responses. From a design perspective, these findings support positioning *TSV*(+) as a climate-responsive performance objective. In hot and humid school buildings, *TSV*(+) can function as a design–operational proxy that simultaneously reflects occupant thermal perception and cooling energy implications, thereby linking comfort evaluation with form-driven energy performance optimization. This finding highlights that performance trade-offs in hot–humid climates are governed by massing-induced heat redistribution rather than simple load accumulation, suggesting that form optimization should consider energy intensity and spatial thermal uniformity simultaneously.

4.3. Design implications and applicability of the integrated framework

In architectural practice, building form is developed through sequential decisions constrained by programmatic, site, and performance requirements. The DIS proposed in this study formalizes this process by mapping the relative importance of morphological parameters to corresponding design stages, while ML-based prediction provides rapid performance feedback throughout. This staged framework demonstrates how the integration of MOO-derived KPIs and ML-based prediction establishes a continuous form–performance feedback loop, enabling consistency between design generation, evaluation, and optimization.

At the early massing stage, programmatic requirements determine the total number of classrooms, thereby constraining the Number of rows, Number of stories, and Classrooms per row. These parameters form the primary layer of the DIS, as they define global spatial structure, inter-row shading, and surface-to-volume ratio. ML-based prediction enables rapid screening of massing alternatives by evaluating daylighting autonomy, thermal comfort (*SET* and *TSV(+)*), and cooling energy, allowing early exclusion of structurally inferior configurations without repeated full simulations.

Once the primary spatial framework is fixed, the site-responsive configuration stage adjusts secondary parameters, including setback



(a) Relationship between *TSV*(+) and Cooling Energy (b) Relationship between *TSV*(+) and *SET*

Fig. 21. Scatter relationships between *TSV*(+) and thermal performance indicators.

levels, height-to-spacing ratio, and generative rule type, to address site boundaries, orientation, and spacing regulations. According to the DIS hierarchy, these parameters exert limited influence on overall performance but support localized modulation of shading and daylighting distribution. ML prediction facilitates rapid sensitivity checks to ensure that site-driven adjustments remain within PCZs and preserve the established daylighting–thermal balance.

At the fine-scale design stage, façade-level parameters such as window-to-wall ratio, glazing configuration, and shading devices are introduced to refine daylighting uniformity and mitigate localized thermal stress. Although these parameters fall outside the current parametric scope, the DIS logic remains applicable: global massing remains fixed, while local modifications are evaluated against ML-predicted performance trends to maintain continuity with earlier-stage objectives. This integration transforms performance evaluation from a simulation-intensive, post hoc process into a predictive and decision-oriented workflow, significantly improving the efficiency and applicability of performance-driven design in practice.

4.4. Methodological implications and limitations

Despite the increasing adoption of integrated workflows that combine parametric modeling, simulation-based MOO, and machine learning, most existing studies primarily focus on computational efficiency or predictive accuracy, while lacking a clear methodological integration of human-centered performance metrics and design-oriented decision logic. In particular, thermal comfort indicators are often treated as post-evaluation metrics rather than as active drivers in early-stage form optimization, and the translation of optimization results into actionable design strategies remains limited. This study aims to develop an integrated, performance-oriented framework to systematically investigate and optimize the coupled effects of building form on daylighting, thermal comfort, and energy use in middle-school teaching buildings in hot–humid climates. Building on the complementary strengths of field measurements, numerical simulations, and ML, this study integrates these approaches into a unified, performance-oriented framework. Unlike conventional integrated pipelines, the proposed framework is structured around a consistent KPIs system that links morphological parameters, simulation outputs, optimization objectives, and machine learning targets. Validating simulations with measured data and using simulation-generated multi-scenario datasets to train ML models enables systematic, quantitative analysis of how building form influences daylighting, thermal comfort and energy use, while addressing limitations related to limited coverage, unstable accuracy and incomplete physical representation.

Specifically, this methodology makes three main contributions. First, thermal sensation vote (TSV) is explicitly incorporated as a performance-driven optimization objective, enabling a direct linkage between building morphology and occupant thermal perception. Second, a rule-based parametric form-generation framework is developed based on a systematic survey of 58 real-world school buildings, ensuring that the explored design space reflects realistic morphological patterns rather than abstract parameter combinations. Third, the optimization results are translated into design-oriented knowledge through the introduction of PCZ and DIS, which provide structured guidance for staged decision-making in early design.

While the proposed DIS–ML integrated framework enhances the efficiency and responsiveness of performance-driven design, its application involves inherent boundary conditions and potential "proxy-on-proxy" uncertainties. The integration of a field-derived TSV regression into a simulation-based MOO workflow means that any simulation bias in T_a, RH, T_g, and V_a could propagate into the comfort objectives. However, this risk was systematically mitigated through several robust measures. First, the simulation engine was strictly validated against field measurements, ensuring minimized input bias. Second, as evidenced by the 95% confidence intervals (T_a [0.005, 0.054]; T_g [0.180, 0.238]) and VIF values (< 5), the regression coefficients remain highly stable, ensuring that typical simulation deviations result in a TSV fluctuation of less than 0.1 units—a margin that is statistically marginal for architectural decision-support. Furthermore, although the transition season R^2 (0.629) is lower than the annual average due to narrower temperature ranges and stochastic occupant behavior, the consistent MAE (0.234–0.267) across all subsets confirms that the absolute predictive accuracy remains robust across diverse operational modes.

From a broader perspective, the current dataset is derived from school buildings in a hot–humid climate (Guangzhou), and the identified form–performance relationships should be interpreted as a decision-support tool within defined climatic and modeling assumptions. While the experimental data were collected from a single representative secondary school, the building's morphological characteristics are consistent with the 58 schools surveyed, and the strong agreement between the field-derived TSV results and existing adaptive comfort studies in Southern China further supports the broader applicability of the findings to high-density educational buildings in hot–humid climate zones [51]. While occupant adaptive behaviors and dynamic controls were not explicitly modeled in the current MOO stage, the TSV model incorporates students' subjective thermal sensation. Future work should extend this applicability by incorporating diverse climate zones, dynamic occupancy patterns, and expanded design variables to further refine the predictive fidelity of the integrated workflow.

5. Conclusion

Based on the orientation of building performance, this study proposes an integrated framework that combines field measurement, parametric form generation, simulation as well as multiple optimization, and ML-based evaluation to systematically examine how building form influences daylighting and thermal performance of secondary school buildings in hot and humid climates. Field measurements and questionnaire surveys were first conducted to collect indoor environmental data and *TSV*, enabling regression-based quantification of students' thermal perception. A parametric form-generation plugin based on typical secondary school forms was then developed to automatically generate diverse cases of building forms. These geometric models were subjected to daylighting as well as thermal simulation and MOO to assess the effects of key morphological parameters. Finally, a GBR model was trained on the optimization dataset, and its predictive performance and mechanism between crucial morphological parameters and building performance was evaluated.

The main conclusions are as follows. First, the *TSV* regression model derived from measured thermal environment indicators and questionnaire achieves high predictive accuracy, with the value of R^2 being 0.921. Second, MOO results indicate that high-performance design solutions consistently converge toward a structurally stable PCZ. This zone is characterized by mid-rise, multi-row building configurations with 5–6 stories, 4–6 rows, 3–5 classrooms per row, 0–1 setback levels, and a height-to-spacing ratio ranging from 1.2 to 1.4, reflecting balanced trade-offs among daylighting, thermal comfort, and energy performance. Third, the trained GBR model demonstrates strong predictive capability across multiple performance metrics, achieving R^2 values of 0.918 for TSV(+), 0.980 for cooling energy, and 0.987 for lighting energy. SHAP analysis indicates that building massing parameters (rows, stories, and classrooms per row) are the most influential contributors to GBR model predictions, confirms the feasibility of rapid performance evaluation based on morphological parameters.

The proposed framework and findings provide architects and planners with a practical decision-support approach to balancing daylighting and thermal performance through building form design in secondary school buildings located in hot and humid climates.

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

The work was supported by the Guangdong Natural Science Foundation (Grant No. 2024A1515011428), Postdoctoral Fellowship Program of CPSF under Grant Number GZC20251115, Chengdu Technological Innovation and R&D Program (Grant No. 2025-YF05-00029-SN).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.enbuild.2026.117501.

Data availability

Data will be made available on request.

References

[1] X. Meng, M. Zhang, M. Wang, Effects of school indoor visual environment on children's health outcomes: A systematic review[J], Health Place 83 (2023) 103021.
[2] G.C. Marchand, N.M. Nardi, D. Reynolds, et al., The impact of the classroom built environment on student perceptions and learning[J], J. Environ. Psychol. 40 (2014) 187–197.
[3] M.C. Lee, K.W. Mui, L.T. Wong, et al., Student learning performance and indoor environmental quality (IEQ) in air-conditioned university teaching rooms[J], Build. Environ. 49 (2012) 238–244.
[4] C. Cen, E. Tan, S. Valliappan, et al., Students' thermal comfort and cognitive performance in tropical climates: A comparative study[J], Energ. Buildings 115817 (2025).
[5] L. Yi, Y. Xie, C. Lin, Thermal Environment and Energy Performance of a Typical Classroom Building in a Hot-Humid Region: A Case Study in Guangzhou, China[J], Geofluids 2022 (1) (2022) 3226001.
[6] X. Gong, Y. Li, J. Cai, et al., A statistical analysis of energy consumption survey of public buildings in a hot summer and cold winter coastal zone of China[J], Buildings 13 (11) (2023) 2685.
[7] Z.S. Zomorodian, M. Tahsildoost, M. Hafezi, Thermal comfort in educational buildings: A review article[J], Renew. Sustain. Energy Rev. 59 (2016) 895–906.
[8] V. Costanzo, G. Evola, L. Marletta, A Review of Daylighting Strategies in Schools: state of the Art and Expected Future Trends [J], Buildings 7 (2) (2017) 41.
[9] A. Zhang, R. Bokel, A. van den Dobbelsteen, et al., Optimization of thermal and daylight performance of school buildings based on a multi-objective genetic algorithm in the cold climate of China[J], Energ. Buildings 139 (2017) 371–384.
[10] A. Zhang, R. Bokel, A. Van den Dobbelsteen, et al., The effect of geometry parameters on energy and thermal performance of school buildings in cold climates of China[J], Sustainability 9 (10) (2017) 1708.
[11] K. Lakhdari, L. Sriti, B. Painter, Parametric optimization of daylight, thermal and energy performance of middle school classrooms, case of hot and dry regions[J], Build. Environ. 204 (2021) 108173.
[12] S. Dervishi, N. Baçi, Early design evaluation of low-rise school building morphology on energy performance: Climatic contexts of Southeast Europe[J], Energy 269 (2023) 126790.
[13] B.J. Alkhatatbeh, Y. Kurdi, S. Asadi, Multi-objective optimization of classrooms' daylight performance and energy use in US Climate Zones[J], Energ. Buildings 297 (2023) 113468.
[14] P. Bakmohammadi, E. Noorzai, Optimization of the design of the primary school classrooms in terms of energy and daylight performance considering occupants' thermal and visual comfort[J], Energy Rep. 6 (2020) 1590–1607.
[15] Y. Liu, W. Wang, Z. Li, et al., Daylighting performance and thermal comfort performance analysis of west-facing external shading for school office buildings in cold and severe cold regions of China[J], Sustainability 15 (19) (2023) 14458.
[16] Z.S. Zomorodian, S.S. Korsavi, M. Tahsildoost, The effect of window configuration on daylight performance in classrooms: A field and simulation study[J], International journal of architectural engineering and urban planning 26 (1) (2016) 15–24.
[17] Y. Jia, Z. Liu, Y. Fang, et al., Effect of interior space and window geometry on daylighting performance for terrace classrooms of universities in severe cold regions: A case study of Shenyang, China[J], Buildings 13 (3) (2023) 603.
[18] B. Liu, Y. Liu, Q. Deng, et al., A study on daylighting metrics related to the subjective evaluation of daylight and visual comfort of students in China[J], Energ. Buildings 287 (2023) 113001.
[19] L. Gábrová, Comparison between dynamic and static metrics for daylight evaluation in the case of obstructed buildings[J], Applied Mechanics and Materials 861 (2017) 477–484.
[20] I. Kistelegdi, K.R. Horváth, T. Storcz, et al., Building geometry as a variable in energy, comfort, and environmental design optimization—A review from the perspective of architects[J], Buildings 12 (1) (2022) 69.
[21] J. Li, C. Liang, W. Zhou, A review of building physical shapes on heating and cooling energy consumption[J], Energies 17 (22) (2024) 5766.
[22] S. Di Loreto, M. Falone, M. Pierantozzi, et al., Field Measurements of Indoor Environmental Quality in School Buildings Post-COVID-19: Systematic Review[J], Appl. Sci. 15 (10) (2025) 5692.
[23] M.O. Fadeyi, K. Alkhaja, M.B. Sulayem, et al., Evaluation of indoor environmental quality conditions in elementary schools′ classrooms in the United Arab Emirates [J], Front. Archit. Res. 3 (2) (2014) 166–177.
[24] I. Ryan, X. Deng, G. Thurston, et al., Measuring students' exposure to temperature and relative humidity in various indoor environments and across seasons using personal air monitors[J], Hygiene and Environmental Health Advances 4 (2022) 100029.
[25] P. De Wilde, The gap between predicted and measured energy performance of buildings: A framework for investigation[J], Autom. Constr. 41 (2014) 40–49.
[26] Bauwens G, Ritoša K, Roels S, et al. IEA EBC Annex71: Building energy performance assessment based on in-situ measurements[J]. IEA EBC Annex71: Building energy performance assessment based on in-situ measurements-Physical parameter identification, 2021.
[27] K. Deb, A. Pratap, S. Agarwal, et al., A fast and elitist multiobjective genetic algorithm: NSGA-II[J], IEEE Trans. Evol. Comput. 6 (2) (2002) 182–197.
[28] M. Hakimazari, M. Baghoolizadeh, S.M. Sajadi, et al., Multi-objective optimization of daylight illuminance indicators and energy usage intensity for office space in Tehran by genetic algorithm[J], Energy Rep. 11 (2024) 3283–3306.
[29] S. Miraba, A. Salehi, M. Rostamzadeh-Renani, et al., Adaptive BIPV shading optimization for electricity generation and building electricity management, sDA, and DGP using multi-objective algorithms[J], Appl. Energy 402 (2025) 126860.
[30] M. Baghoolizadeh, M.B. Hamooleh, A. Alizadeh, et al., Occupant's thermal comfort augmentation and thermal load reduction in a typical residential building using genetic algorithm[J], Case Stud. Therm. Eng. 59 (2024) 104491.
[31] M. Baghoolizadeh, M. Rostamzadeh-Renani, R. Rostamzadeh-Renani, et al., Multi-objective optimization of Venetian blinds in office buildings to reduce electricity consumption and improve visual and thermal comfort by NSGA-II[J], Energ. Buildings 278 (2023) 112639.
[32] R. Li, Z. Shari, M.Z.A. Ab Kadir, A review on multi-objective optimization of building performance-Insights from bibliometric analysis[J], Heliyon 11 (4) (2025).
[33] M. Hamdy, A. Hasan, K. Siren, A multi-stage optimization method for cost-optimal and nearly-zero-energy building solutions in line with the EPBD-recast 2010[J], Energ. Buildings 56 (2013) 189–203.
[34] Y. Pan, M. Zhu, Y. Lv, et al., Building energy simulation and its application for building performance optimization: A review of methods, tools, and case studies [J], Adv. Appl. Energy 10 (2023) 100135.
[35] N. Ziaee, R. Vakilinezhad, Multi-objective optimization of daylight performance and thermal comfort in classrooms with light-shelves: Case studies in Tehran and Sari, Iran[J], Energ. Buildings 254 (2022) 111590.
[36] J. Rucińska, A. Trząski, Measurements and simulation study of daylight availability and its impact on the heating, cooling and lighting energy demand in an educational building[J], Energies 13 (10) (2020) 2555.
[37] Hong T, Langevin J, Sun K. Building simulation: Ten challenges[C]//Building simulation. Berlin/Heidelberg: Springer Berlin Heidelberg, 2018, 11(5): 871-898.
[38] P.W. Tien, S. Wei, J. Darkwa, et al., Machine learning and deep learning methods for enhancing building energy efficiency and indoor environmental quality–a review[J], Energy AI 10 (2022) 100198.
[39] J. Ogundiran, E. Asadi, M. Gameiro da Silva, A systematic review on the use of AI for energy efficiency and indoor environmental quality in buildings[J], Sustainability 16 (9) (2024) 3627.
[40] L. Haifeng, H.C. Hou, Z. Gou, User-centric approach to optimizing thermal comfort in university classrooms: Utilizing computer vision and Q-XGBoost reinforcement learning[J], Energ. Buildings 323 (2024) 114808.
[41] S. Miao, M. Gangolells, B. Tejedor, Data-driven model for predicting indoor air quality and thermal comfort levels in naturally ventilated educational buildings using easily accessible data for schools[J], Journal of Building Engineering 80 (2023) 108001.
[42] Lala B, Kala S M, Rastogi A, et al. Building matters: Spatial variability in machine learning based thermal comfort prediction in winters[C]//2022 IEEE International Conference on Smart Computing (SMARTCOMP). IEEE, 2022: 342-348.
[43] GB/T 5699–2025. Method of daylighting measurements. Beijing: China Standards Press; 2025.
[44] G.-S. Brager, R.-J. de Dear, Thermal adaptation in the built environment: a literature review[J], Energ. Buildings 27 (1) (1998) 83–96.
[45] Public Places Sanitary Inspection Methods – Part 1: Physical Factors [S]. China Standards Press, 2013.
[46] A.S.H.R.A.E.A.S.H.R.A.E. Standard, 55-2021: Thermal Environmental Conditions for Human Occupancy, ASHRAE, Atlanta, 2021.
[47] International Organization for Standardization. ISO 7730:2005. Ergonomics of the thermal environment—Analytical determination and interpretation of thermal comfort using calculation of the PMV and PPD indices and local thermal comfort criteria. Geneva: ISO; 2005.

[48] Wienold, J. *Dynamic daylight glare evaluation* [C] 11th International Building Performance Simulation Association (IBPSA) Conference, Glasgow, Scotland, 944–951.
[49] R. Elnaklah, Y. Ayyad, S. Alnusairat, et al., A comparison of students' thermal comfort and perceived learning performance between two types of university halls: architecture design studios and ordinary lecture rooms during the heating season [J], Sustainability 15 (2) (2023) 1142.
[50] G. Torriani, G. Lamberti, G. Salvadori, et al., Thermal comfort and adaptive capacities: Differences among students at various school stages[J], Build. Environ. 237 (2023) 110340.
[51] L. Yang, S. Zhao, Y. Zhai, et al., The Chinese thermal comfort dataset[J], Sci. Data 10 (1) (2023) 662.