

Surrogate Modeling Strategy for Urban Building Energy Simulation in Early-Stage Urban Design: A Case Study of Energy-Efficient Neighborhood Design in Seoul

Doyun Lee¹, Steven Jige Quan^{1,2,*}

¹City Energy Lab, Graduate School of Environmental Studies, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul, 08826, Republic of Korea

²Environmental Planning Institute, Graduate School of Environmental Studies, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul, 08826, Republic of Korea

Abstract. This study focuses on developing an efficient surrogate modelling strategy for early-stage urban design, aiding in energy-efficient residential neighborhood design in Seoul. The methodology involves devising a design schema with important design parameters, generating 1,000 random designs via Latin Hypercube Sampling, and simulating each design's energy use considering microclimate and shadow effects. Polynomial regression is used to develop surrogate models based on those energy simulation results, which is further validated with energy use measurements of real neighborhoods in Seoul. The developed surrogate model can provide quick evaluation of energy use with a moderate level of accuracy for early-stage energy efficient neighborhood design.

1 Introduction

The issue of energy consumption in buildings is a matter of considerable importance, as buildings are responsible for about 40% of the overall energy consumption globally [1]. The relevance of this fact becomes even more significant when considering the multifaceted factors influencing energy performance in buildings, including mass size, spaces between buildings, height, local climate and shadow effect. Due to this complexity, it is challenging to predict building energy consumption accurately, creating a need for sophisticated strategies and tools.

The prediction of energy usage in buildings holds a significant role in their performance enhancement, primarily aimed at achieving energy conservation and reducing environmental impact [2].

The early stage of urban design is critical in developing energy efficient neighborhoods, when building energy simulation models can greatly aid in design decision-making. Petersen & Svendsen [3] elucidate that the early stages of building design encapsulate numerous decisions, which leave a profound impact on the building's performance throughout its life cycle. However, due to their time-consuming simulation and the detailed design inputs required, these models are not widely used.

The role of early-stage simulations in this scenario is notably crucial. Decisions made during this phase can substantially influence life-cycle impacts and costs, thereby shaping the energy efficiency and environmental footprint of the building.

To address this issue and better utilize simulation in early-stage urban design, this study aims to develop a surrogate modelling strategy which provides simplified yet effective simulation for energy-efficient urban design support. The study applies this strategy to a neighborhood design case in Seoul and compares different types of surrogate models on their effectiveness in early-stage urban design.

The case study focusing on energy-efficient neighborhood design in Seoul is particularly relevant, given the city's urban density and the potential benefits of energy-saving strategies at a neighborhood scale. The overarching goal of this study is to develop a surrogate modelling strategy that simplifies and enhances the simulation of energy-efficient urban design. This simplified yet effective model, aims to apply urban building energy simulation for early-stage urban design support. Through validation with a case study in Seoul, the study will assess the applicability and effectiveness of this surrogate model in real-world urban design scenarios.

2 Literature Review

In the realm of urban design and building simulation, several studies have scoped on the building energy simulation and the importance of early-stage decision making in urban design.

Wu et al. [4] address the interplay between building design and the surrounding environment, focusing particularly on the impacts on pedestrians' wind comfort. The authors put forth a metamodel-based optimization

* Corresponding author: sjquan@snu.ac.kr

method aimed at developing near-optimal designs that mitigate the adverse effects of buildings on wind comfort for pedestrians in both hot and cold seasons. This study highlights the importance of considering environmental factors and their interactions with building design, suggesting that optimization methods can be beneficial in achieving more sustainable and comfortable urban environments.

Bragança et al. [5] delve into the process of design project development, with particular emphasis on the early stages. The paper underscores the critical role the initial stages of design play in influencing sustainability, performance, and life-cycle costs. The study postulates that early decision-making and planning can have far-reaching consequences on a building's overall performance, thereby emphasizing the necessity for careful and informed choices during these initial phases.

Østergård et al. [6] put forward a simulation framework that bolsters proactive, intelligent, and experience-based building simulation to support decision-making in the early stages of design. This framework is tailored to facilitate a more informed design process, thereby ensuring that decisions made are both efficient and effective in terms of building performance.

Furthermore, Quan et al. [7] explore the intricate relationship between density and energy performance in urban neighborhoods. Through simulation experiments, the authors analyze nine Shanghai neighborhoods and find that the relationship between density and energy performance. While geometry-related density appears to have a negative impact on building energy use intensity, other factors tied to neighborhood typology may alter this relationship.

The identified research gaps from previous studies include a scarcity of studies that explicitly defined variables and design schema while directly conducting simulations, despite the emphasis on the significance of early-stage building energy simulation. Furthermore, there was a limited studies focused on apartment buildings in Seoul.

3 Methodology

The spatial focus of this study is a representative residential neighborhood in Seoul, illustrating a prevalent housing format in the city: the apartment building. The primary metric of interest is the annual building energy use intensity (EUI). The unit of analysis for this research emphasizes one building within a 3x3 neighborhood grid to take into account microclimate and shadow effects.

The methodology of this study is designed to develop a surrogate modelling strategy that simplifies and effectively simulates energy-efficient urban design. It will be carried out in four main steps.

Step 1. Development of Design Schema: The initial stage of this study involves the development of a design schema. A select set of six design variables, namely building width, building depth, building height, orientation, and intervals between buildings (axis-x, axis-y) are used to define a design solution. 1000 set of designs are derived using Latin Hypercube Method. The ranges of

the design variables are adjusted between the regulations of Seoul's apartment building architecture.

Step 2. Construction of Building Energy Simulation Model: A detailed building energy simulation model is framed to reflect the characteristics of the design schema. During this phase, microclimate effects and shadow effects are considered. The microclimate conditions are derived from the Urban Weather Generator tailored for residential-heavy sectors in Seoul, subsequently leading to the formulation of a bespoke EPW (EnergyPlus Weather) file.

Step 3. Building Energy Simulation: All 1000 design samples undergo rigorous energy simulation to ascertain their EUI values. Post simulation, a validation exercise is initiated where the energy model is juxtaposed against 10 real-world apartment buildings in Seoul to measure its accuracy and reliability.

Step 4. Building Surrogate Model: With the simulation results for the 1000 designs at hand, polynomial regression is employed to sculpt a surrogate model. To facilitate better comprehension of the results, a response surface is formulated. As a final validation step, the efficacy of the surrogate model is tested against the energy performance data of 100 actual apartment structures in Seoul.

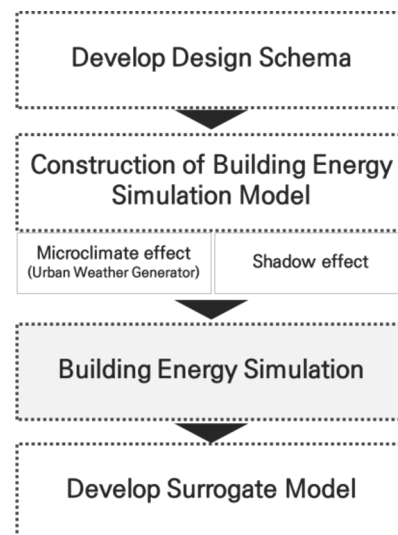


Fig. 1. Flowchart of the Study

3.1 Development of Design Schema

The initial phase in this study involved the development of a design schema, primarily centered around six variables - building height (H), building width (W), building depth (D), Building orientation (θ), and intervals between buildings (I_x, I_y). Following the regulations for apartment buildings in Seoul, the height was varied between 5 to 35 floors, considering each floor to be 3m in height. Building Width (W) and Depth (D) were set based on the architectural regulations for apartments in Seoul. Moreover, a grid spacing of 3m was adhered to for these dimensions. Building Orientation (θ) ranged from -90 degrees to 90 degrees. Intervals Between Buildings (I_x, I_y) were also set following the building codes specific to

Seoul apartments. These variables were selected based on the typical characteristics and legislation regulations of Korean apartments. Latin hypercube method is used to derive four variables. Latin hypercube sampling (LHS) is a systematic sampling technique employed to minimize the required number of simulations when evaluating the uncertainty of responses [8]. Fig 2 represents a neighborhood of the derived design samples.

Table 1. Design Variables

Design Vector Variable	Range
Building height (H)	15m – 105m (3m interval)
Building width (W)	30 – 120m (3m interval)
Building depth (D)	12 – 45m (3m interval)
Building orientation (θ)	-90°–90°
Interval (I_x)	0.8H-120m
Interval (I_y)	0.8H-120m

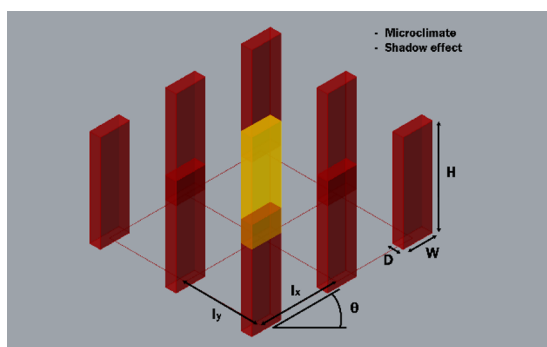


Fig. 2. Description of a Neighborhood design

3.2 Construction of Building Energy Simulation Model

3.2.1 Energy Simulation Setting

The annual building energy consumption of each of the 1000 designs was simulated considering a combination of microclimate and shadow effect. For this purpose, the Honeybee and OpenStudio Software was utilized, a platform designed to support whole-building energy modeling through EnergyPlus. The simulation was run at the level of a single neighborhood, which is composed of a 3x3 grid of buildings. The simulation's location is Seoul, with a building program of midrise apartments and a construction type of Mass. The climate zone is ASHRAE 4 - Mixed, and the building vintages are ASHRAE 90.1 2019 and IECC 2021. The simulation considered the shadow effect from the surrounding buildings in the neighborhood and also factored in the microclimate. The Energy Use Intensity (EUI) recorded from this simulation is from the building located in the very center of the 3x3 grid. EUI is calculated as the sum of all energy use (including electricity, fuel, district heating/cooling, etc.)

divided by the gross floor area, which includes both conditioned and unconditioned spaces.

Table 2. Energy Simulation Setting

Location	Seoul
Building Programs	Midrise Apartment
Climate Zones	ASHRAE 4 - Mixed
Climate Data	Seoul EPW data
Construction Type	Mass
Building Vintages	ASHRAE 90.1 2019 IECC 2021

3.2.2 Urban Weather Generator

In the pursuit of achieving a comprehensive and accurate simulation, we extracted the weather file for a typical apartment-concentrated area in Seoul by developing an Urban Weather Generator (UWG). Traditional weather files, which are predominantly based on readings from rural or less built-up locations, might not capture the intricacies of densely constructed urban environments [9]. The phenomenon of rising air temperatures due to urbanization, commonly referred to as the Urban Heat Island (UHI) effect can notably influence the energy demands of buildings in particular cities, including Seoul. The specific region selected was Nowon-gu, renowned for hosting the highest density of large-scale apartment complexes within the city. The use of the UWG is crucial for ensuring the fidelity and reliability of our simulations in the context of Seoul's apartment complexes.

3.2.3 Validation of Building Energy Simulation Model

After building the building energy simulation model, the model was tested with 10 typical Seoul neighborhoods. Table 3 presents the typical apartments of Seoul with the actual energy use intensity and the simulated result. Table 4 presents the ASHRAE Guideline 14-2002 Values in comparison to values achieved in UBEM practice at neighborhood level at hourly temporal resolutions. The model demonstrates a promising ability to predict the Energy Use Intensity (EUI) of the selected apartments, though it showcases some variations when benchmarked against the ASHRAE Guideline 14-2002. The Coefficient of Variation of the Root Mean Square Error (CVRMSE) registers at 58.7%, a deviation from the ASHRAE benchmark of 15%. This can be viewed as an indication of the model's sensitivity to capturing complex behaviors in the data that might not be encapsulated within more conventional models. Additionally, the Normalized Mean

Table 3. Comparison of Building Energy Simulation result and Actual EUI in 10 Apartments in Seoul

	H(m)	W(m)	D(m)	$\theta(^{\circ})$	I_x (m)	I_y (m)	EUI (kWh/m ²)	Simulated EUI (kWh/m ²)
1	45	105	13	23	115	66	136.8	124.5
2	27	57	10	30	68	25	105.0	165.5
3	63	61	12	48	46	60	205.3	168.6
4	105	50	10	45	70	63	377.4	160.4
5	45	70	10	-10	101	32	162.7	138
6	51	46	12	-22	63	50	173.0	152.9
7	39	107	15	1	120	62	111.4	147.9
8	105	36	36	60	93	54	236.5	134.3
9	15	63	10	9	82	32	62.8	179.3
10	36	12	105	62	60	110	116.7	149.1

Bias Error (NMBE) is approximately 11%, slightly above the ASHRAE standard of 5%. This suggests that our model has a consistent prediction trend and might lean towards slightly conservative energy efficiency estimates, potentially providing an added safety margin in real-world scenarios. This study primarily focused on constructing surrogate models; hence, validation against real-world data is for reference purposes only.

Table 4. ASHRAE Guideline 14-2002 Values in comparison to values achieved in UBEM practice

Metric	Neighborhood	
	Value (%)	ASHRAE Guideline 14-2002 (%)
CVRMSE	58.698914	15
NMBE	10.993752	5

4 Result

4.1 Development of Surrogate Model

Using the simulated building energy data, this study derived a surrogate model to predict building energy use. This model aimed to predict building energy usage based on the key building dimensions identified in the design schema - building height, width, depth, orientation, and the interval in the x and y directions. The equation of the model, based on the coefficients and the order of the polynomial features, can be written as follows:

$$\begin{aligned}
 y = & 91.8254 + 0.0033H - 0.0004W - 0.0043D + 0.0242O - 0.001I_x - 0.00006I_y + 2.44e-10H^2 \\
 & - 4.31e-9HW - 3.22e-8HD + 2.80e-7HO + 9.65e-11HI_x - 1.38e-10HI_y + 3.95e-9W^2 \\
 & - 6.75e-10WD + 1.05e-7WO - 4.56e-10WI_x - 9.45e-10WI_y \\
 & + 7.33e-8D^2 - 8.79e-7DO + 1.99e-10DI_x \\
 & + 1.17e-9DI_y + 0.0002O^2 - 8.87e-9OI_x \\
 & - 1.85e-7OI_y + 5.69e-10I_x^2 + 4.97e-10I_xI_y + 1.84e-10I_y^2
 \end{aligned}
 \tag{1}$$

where,

y: result of the building simulation (EUI),

H: height of the building,

W: width of the building,

D: depth of the building,

O: orientation of the building,

I_x : interval between buildings in the x direction,

I_y : interval between buildings in the y direction.

This model was constructed as a quadratic polynomial regression, and includes the squares of the features as well as cross-terms between the features. By plugging in new values of the building attributes into this equation, it is possible to predict the EUI.

Upon conducting validation of the derived surrogate model, the results were a Mean Squared Error (MSE) of 18.90 and an R² score of 0.9942. Overall, these metrics suggest that the surrogate model performs exceptionally well in predicting the building energy usage based on the key building dimensions identified in the design schema - building height, width, depth, orientation, and intervals in both x and y directions.

4.2 Validation

Next, this study tested the validation of the derived model using data from 100 real apartments in Seoul. The building energy data for Seoul apartments provided by 'K-apt.go' were diversely composed of individual apartments with a floor area of 60m² or less, between 60 m² and 85 m², between 85 m² and 135 m², and exceeding 135 m².

Upon validation, the Mean Squared Error (MSE) stood at 1130.626, and the R² score was calculated to be 0.723. The obtained MSE suggests some level of difference between the predicted and the actual values. This might hint that our surrogate model's performance may not align as seamlessly with real-world data as it did with the simulated dataset. An R² score of 0.723 is commendable, suggesting that our model can account for approximately 72.3% of the variability in building energy consumption. Although this score isn't as optimal as those observed with simulation data, it confidently underscores the utility of our model in predicting building energy

usage across genuine Seoul apartments. A plausible reason for these metrics could be the encompassing design schema of the simulation, covering a broad spectrum of potential apartment configurations in Seoul. In contrast, the validation dataset extracted from actual Seoul data might not showcase such extensive variability. Hence, the intricate diversity of the simulated data might not be as deeply echoed in real-world data, which could contribute to the surrogate model's slightly moderated performance with the latter.

4.3 Examining Effects of Design Variables: Response Surface Model

For the six-factor design experiment, this study utilized response surface plots illustrating the relationships between pairs of variables. Figure 3 illustrates various relationships between the design variables.

Panel A shows the relationship between building height and building width. In this case, the energy use intensity (EUI) tends to decrease as the building height decreases. The building width does not significantly impact the EUI. Panel B depicts the relationship between

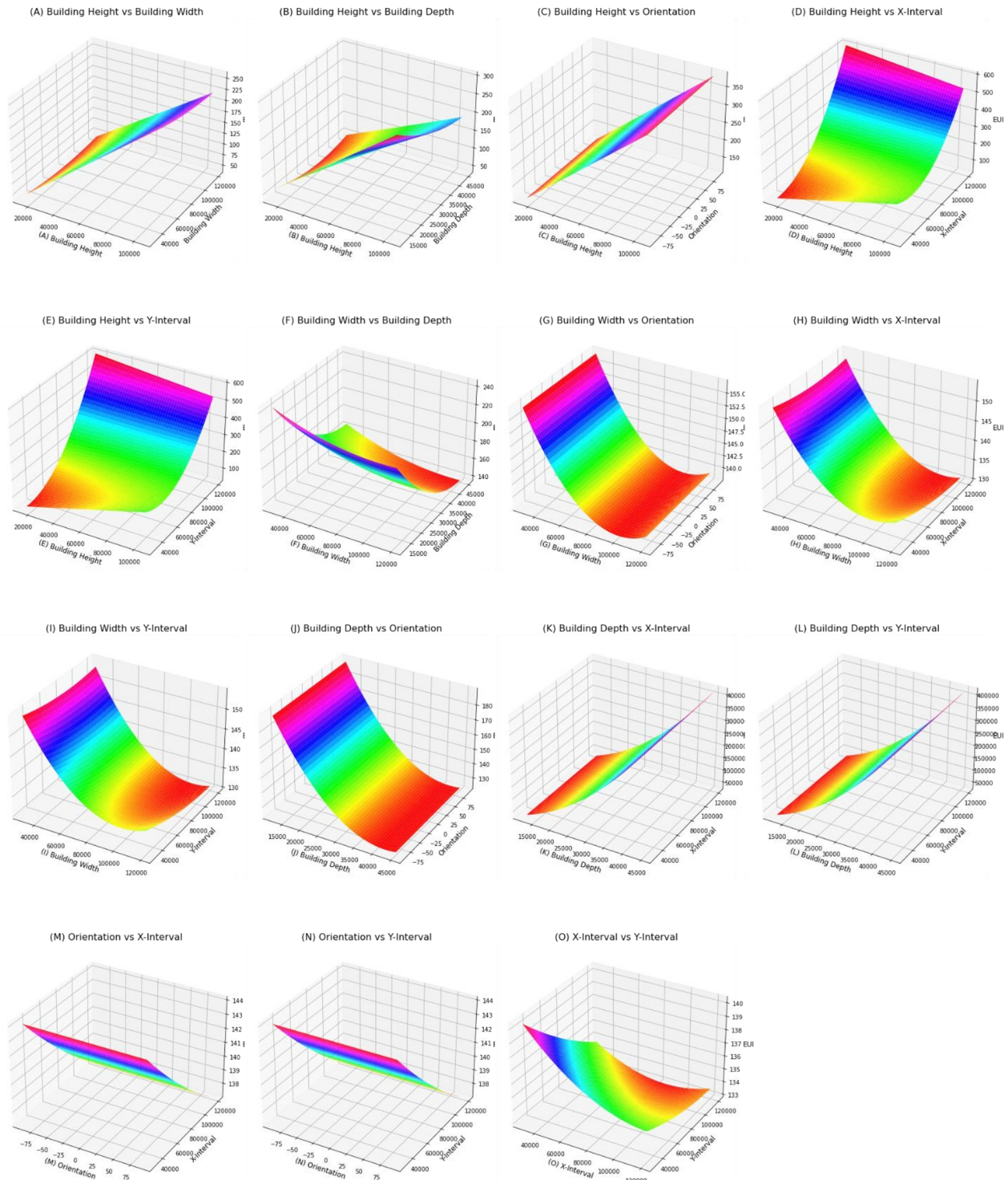


Fig. 3. Response Surface Model

building height and building depth. Here, the energy use intensity (EUI) decreases as the building height decreases, with building depth having no significant influence. Panel C showcases the interplay between building height and orientation. Again, the EUI decreases as the building height gets lower, but the orientation doesn't appear to impact this trend.

Panel D and E demonstrate the relationship between building height and X, Y-Interval. In this visualization, the energy use intensity (EUI) diminishes as both the building height and the X, Y-Interval decrease.

Panel F focuses on the relationship between building width and building depth. For these parameters, the energy use intensity (EUI) tends to elevate as both factors increase.

Panel G displays the interaction between building width and orientation. As the building width surpasses the median value, the EUI begins to rise. The orientation does not seem to sway the EUI in this scenario.

Panel H and I offer insight into the dynamic between building width and X, Y-Interval. As both variables increase, so does the EUI.

Panel J represents the relationship between building depth and orientation. The EUI notably grows as building depth increases, while orientation remains a non-influential factor.

Panel K and L visualize the connection between building depth and X, Y -Interval. The trend suggests that as the building depth decreases, the EUI also goes down. The X, Y-Interval doesn't have a marked influence in this relation.

Panel M and N delineate the relationship between orientation and X, Y-Interval. It's evident that as the X, Y-Interval goes up, the EUI follows suit.

Panel O culminates with the comparison between X-Interval and Y-Interval. The panel underscores that a surge in either or both intervals correlates with a pronounced increase in EUI.

In summary, Figure 3 provides valuable insights into how various building design parameters influence the energy use intensity (EUI) of buildings:

The building height emerges as the most dominant factor affecting EUI. A notable trend across panels is that buildings with a reduced height tend to be more energy efficient. This suggests that designs should potentially lean towards shorter structures if energy conservation is a priority. As the width of the building increases, so does its energy efficiency. Thus, broader building designs could be more beneficial in terms of energy usage. The influence of building depth on energy efficiency appears to be context-dependent. While in some scenarios it has a discernible effect on EUI, in others its impact is less straightforward. This indicates that the optimal building depth for energy efficiency may need to be determined on a case-by-case basis. Across all panels, orientation doesn't seem to have a substantial impact on the energy use intensity. It suggests that while orientation might be important for other design considerations, its influence on the overall energy efficiency of a building appears to be minimal. Both X-Interval and Y-Interval generally demonstrate that larger intervals correspond to buildings being more energy efficient. This insight might pave the

way for designs that incorporate spacious intervals to optimize for energy conservation.

Considering the above findings, architects and designers should be cognizant of the significant roles that building height and width play in energy consumption. Additionally, while depth and orientation are important for other aspects of design, their direct influence on energy efficiency might be lesser compared to the other parameters. As for intervals, it's evident that more expansive spaces could lead to more energy-efficient structures.

5 Discussion

In this study, surrogate model is developed for building energy simulations of neighborhood designs in Seoul. The results of the evaluations indicate that these models have great potential to estimate building energy use with a high level of accuracy and at a faster speed than traditional methods.

Table 5 suggest the most efficient and inefficient design cases in the given sample. This shows key design parameters influential for building energy use, which is building height, identified through the models, providing valuable insights for urban designers to consider during their design processes.

However, this study is not without limitations. The surrogate model in this study is based on the 'flat' shape of large-scale apartments commonly found in Seoul. Consequently, its applicability may be limited when it comes to diverse urban forms. The model might not be well-suited for other architectural configurations such as courtyard apartments, 'tower'-shaped apartments, houses, and so on.

Future research endeavors could consider exploring alternative modeling approaches, like Gaussian processes or gradient boosting, to further enhance the accuracy of the surrogate model. Additionally, expanding the model to cover different housing types beyond the typical apartments in Seoul would provide a more comprehensive tool for energy efficiency assessment. Incorporating specific considerations in the energy simulation model, such as wall materials, window specifications, and detailed HVAC systems, would contribute to building a more precise and reliable energy simulation model.

Table 5. Comparison between energy use intensity of the simulation model and the Surrogate model

	H(m)	W(m)	D(m)	O (°)	I _x (m)	I _y (m)	Observed EUI (kWh/m ²)	Predicted EUI (kWh/m ²)
Most Energy Efficient	15	64	42	-50	107	68	40.509	44.461
Least Energy Efficient	105	82	12	66	87	99	325.407	302.086

6 Conclusion

This research has successfully developed and validated a surrogate model for predicting building energy usage, based on the architectural parameters of building height, width, depth, orientation, and intervals. With a focus on apartment buildings in Seoul, this study marks a significant step in the understanding and consideration of energy consumption in the early stages of building design. By validating the model using data from 100 actual neighborhoods, this study strengthens the credibility and applicability of the model, illustrating its potential for real-world application. This provides urban designers with a valuable tool for making informed design decisions that can lead to more energy-efficient urban development. In conclusion, the surrogate model developed in this study stands to make a substantial contribution to the fields of architecture and urban design, providing a practical tool for incorporating energy efficiency considerations into the early stages of the design process. Future work will further enhance the model's accuracy and applicability, ultimately guiding us towards more sustainable and energy-conscious urban environments.

This work was supported by the Creative-Pioneering Researchers Program through Seoul National University (SNU) the National Research Foundation of Korea (NRF) grant funded by the Korea government (Ministry of Science and ICT) (No. 2018R1C1B5043758; No. 2022R1C1C1004953) and the National Research Foundation of Korea (NRF) grant funded by the Korea government (Ministry of Education) (No. 5120200113713).

References

1. V. S. K. V. Harish, A. Kumar, *Renew. Sust. Energ. Rev.* **56**, 1272-1292 (2016).
2. H. X. Zhao, F. Magoulès, F. *Renew. Sust. Energ. Rev.* **16(6)**, 3586-3592 (2012).
3. S. Petersen, S. Svendsen, *Energy Build.*, **42(7)**, 1113-1119, (2010).
4. Y. Wu, Q. Zhan, S.J. Quan, Y. Fan, Y. Yang, *Build. Environ.*, **195**, 107661 (2021).
5. L. Bragança, S.M. Vieira, J.B. Andrade, J. B., *Sci. World J.*, (2014).
6. T. Østergård, R.L. Jensen, S.E. Maagaard, *Renew. Sust. Energ. Rev.*, **61**, 187-201 (2016).
7. S. J. Quan, J. Wu, Y. Wang, Z. Shi, T. Yang, T., P.P.J. Yang, *Energy procedia*, **88**, 126-132, (2016).
8. M. Ali, R.C. Deo, N.J. Downs, T. Maraseni, *Monthly rainfall forecasting with Markov Chain Monte Carlo simulations integrated with statistical bivariate copulas*. In *Handbook of Probabilistic Models* (pp. 89-105). Butterworth-Heinemann (2020).
9. B. Bueno, L. Norford, J. Hidalgo, G. Pigeon, J. Build. Perform. Simul. **6(4)**, 269-281, (2013).