AD-BASED SURROGATE MODELS FOR SIMULATION AND OPTIMIZATION OF LARGE URBAN AREAS

GONÇALO ARAÚJO[1], LUÍS SANTOS[2], ANTÓNIO LEITÃO[3] and RICARDO GOMES[4]
[1,4] *IN+, Técnico Lisboa*
[2] *Dept. of Architecture, Design and Media Technology, Aalborg Un.*
[3] *INESC-ID, Técnico Lisboa*
[1]*goncalo.r.araujo@tecnico.ulisboa.pt, 0000-0002-0643-8113*
[2]*lfds@create.aau.dk, 0000-0002-2664-4301*
[3]*antonio.menezes.leitao@tecnico.ulisboa.pt, 0000-0001-7216-4934*
[4]*ricardo.a.gomes@tecnico.ulisboa.pt, 0000-0003-2180-2339*

Abstract. Urban Building Energy Model (UBEM) approaches help analyze the energy performance of urban areas and predict the impact of different retrofit strategies. However, UBEM approaches require a high level of expertise and entail time-consuming simulations. These limitations hinder their successful application in designing and planning urban areas and supporting the city policy-making sector. Hence, it is necessary to investigate alternatives that are easy-to-use, automated, and fast. Surrogate models have been recently used to address UBEM limitations; however, they are case-specific and only work properly within specific parameter boundaries. We propose a new surrogate modeling approach to predict the energy performance of urban areas by integrating Algorithmic Design, UBEM, and Machine Learning. Our approach can automatically model and simulate thousands of building archetypes and create a broad surrogate model capable of quickly predicting annual energy profiles of large urban areas. We evaluated our approach by applying it to a case study located in Lisbon, Portugal, where we compare its use in model-based optimization routines against conventional UBEM approaches. Results show that our approach delivers predictions with acceptable accuracy at a much faster rate.

Keywords. Urban Building Energy Modelling; Algorithmic Design; Machine Learning in Architecture; Optimization of Urban Areas; SDG 7; SDG 12; SDG 13.

1. Introduction

To face the challenges and threats posed by climate change, large efforts and funds are being deployed to reduce carbon emissions and energy consumption worldwide (United Nations, 2020). The operation of a large part of the building stock is still

POST-CARBON, Proceedings of the 27th International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2022, Volume 2, 689-698. © 2022 and published by the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA), Hong Kong.

energy-intensive and a cause of anthropogenic carbon emissions that actively contribute to climate change (United Nations, 2018). This motivates building retrofits in consolidated urban territories to meet the sustainable development goals set by the current political agenda (European Commission, 2020). Particularly, to reduce the energy consumption of our building stock, we need to assess its performance through building energy simulation. However, these typically require expertise and become slow when analyzing numerous buildings. These limitations hinder the analysis of large urban areas that is critical to support the design and planning of cities, and the city policy-making sector (Reinhart & Cerezo Davila, 2016).

Current practices use Urban Building Energy Modeling (UBEM) approaches to address this pressing issue (Ferrando et al., 2020). However, these approaches still require a high level of expertise and the process is still time-consuming, error-prone, and tiresome (Chen et al., 2017). Thus, it is crucial to make these approaches simple-to-use, more automated, and faster.

Previous attempts to reduce the computational cost of large numbers of simulations involved surrogate models developed with machine learning techniques. A surrogate model is constructed using data-driven approaches and reproduces the behavior of a simulation model while being computationally cheaper to evaluate. Such models have been used to predict building simulation outputs such as building carbon emissions (Thrampoulidis et al., 2021), energy consumption (Bamdad et al., 2020), and daylighting (Wortmann et al., 2015). They are trained with a simulated case study dataset and substantially improve simulation run-time, deliver faster and accurate results, and promote a smoother integration with current digital design workflows. However, such models are usually case-specific and can present errors when applied to cases that are outside the boundaries of the training data. Such limitation can cause problems while using the surrogate to analyze large urban areas with a diverse building stock (Thrampoulidis et al., 2021).

To address these limitations, we propose the integration of Algorithmic Design and Analysis (ADA) (Aguiar et al., 2017) in the creation of more versatile surrogate models for urban analyses. We use algorithmically modeled building archetypes based on city-specific building properties rather than context-specific data commonly used in UBEM surrogates. This approach allows a broader use of surrogate models for larger urban areas and mitigates the case-specific limitations in the usage of such approaches.

Algorithmic Design (AD) allows us to generate buildings through algorithms (Caetano et al., 2020) and when combined with simulation analysis - ADA - we can automatically model and simulate thousands of design variations. Thus, our approach used ADA to generate and simulate many instances of parametric building archetypes. With the results, we compile a training set and test multiple regression models to build a surrogate that promptly predicts the performance of different urban settings depending on simpler inputs such as building geometry and constructions, and its accuracy and speed can be compared against conventional UBEM approaches.

2. Methodology

The core tool used in our approach was Khepri (Aguiar et al., 2017), a multi-platform AD tool that allows us to seamlessly integrate five different steps involved in the development and validation of our approach (Figure 1). The five steps are as follows: (1) import of the GIS dataset into a CAD platform; (2) model the defined building archetypes for the database and simulate them using EnergyPlus; (3) using the generated dataset to test multiple regression models available in the machine-learning library Sci-Kit Learn; (4) select the best-performant surrogate model to predict energy consumption; (5) apply it in a Multi-Objective Optimization (MOO) routine. To test the surrogate model approach in the optimization of large-scale urban retrofit scenarios, we compare its accuracy and speed with a model-based optimization supported by conventional UBEM simulations.

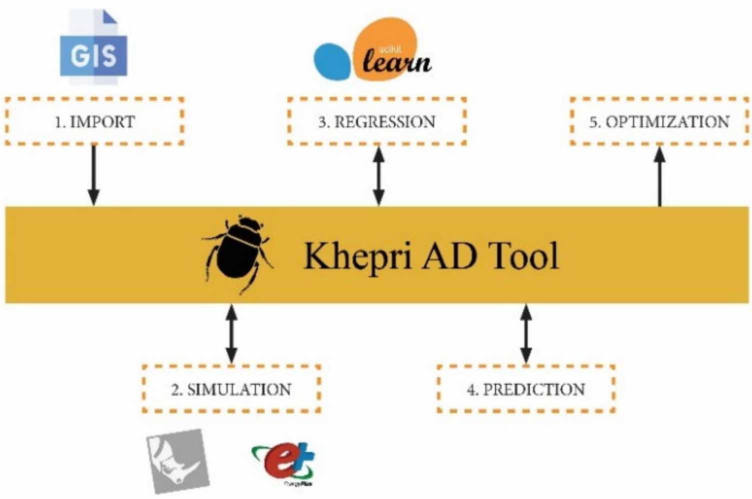

Figure 1. Workflow diagram illustrating the sequential stages of the research's methodology

2.1. CASE STUDY - IMPORT GIS DATASET

Through the analysis of a case study, we can define the domain of our surrogate model parameters and validate our approach. Figure 2 shows our building database, which contains data relative to 2193 residential buildings in Lisbon, Portugal. The city's residential buildings are characterized in Table 1 by 10 periods of construction, each with corresponding material (Santos & Matias, 2006) and retrofit solutions. Available information regarding the urban area, such as building implantation polygon, construction periods, floors, area, glazing ratio, and typology are all imported into our AD tool, and then automatically simulated in the simulation platform. This seamless integration allows us to automate all the subsequent steps of our approach.

Figure 2. Lisbon residential buildings from the dataset.

Table 1. U-values for each construction period in Lisbon.

Construction period	Wall U_Value (kWh/m^2)	Roof U_Value (W/m^2·ºC)	Floor U_Value (W/m^2·ºC)	Window U_Value (W/m^2·ºC)	Wall retrofit U-value (W/m^2·ºC)	Roof retrofit U-value (W/m^2·ºC)
<1919	2.78	1.99	1.80	2.69	0.61	0.63
1919-1945	2.78	1.99	1.80	2.69	0.61	0.63
1946-1960	1.49	1.99	1.80	2.69	0.57	0.63
1961-1970	1.08	1.99	3.03	2.69	0.49	0.63
1971-1980	1.26	1.99	3.03	2.69	0.53	0.63
1981-1990	0.50	1.99	3.03	2.69	0.32	0.63
1991-1995	0.49	1.99	3.03	2.69	0.32	0.63
1996-2000	0.46	1.99	2.31	2.69	0.29	0.63
2001-2005	0.25	1.99	2.31	2.69	0.19	0.63
>2006	0.25	1.99	2.31	2.69	0.19	0.63

2.2. MODEL AND SIMULATION

In this step, we generate and simulate both the surrogate model and our UBEM test dataset. With ADA, we set the inputs for the case study's building archetypes, which include multiple construction solutions and their possible retrofits. These archetypes represent our surrogate model's discrete parameters. In this case, we include all the construction periods and their possible retrofits (Table 1).

All the building archetypes are then modeled according to uniformly divided parameter domains such as the number of floors (from 2 to 11, step size = 3), rectangular proportion (from 1 to 5, step size = 2), orientation (from 0 to 180 degrees, where 0 is East, step size = pi/4), glazing ratio (from 0 to 0.7, step size = 0.35), and floor area (from 50 to 800 m2, step size = 75 m2). Figure 3 exemplifies some values of different parameters of one building archetype.

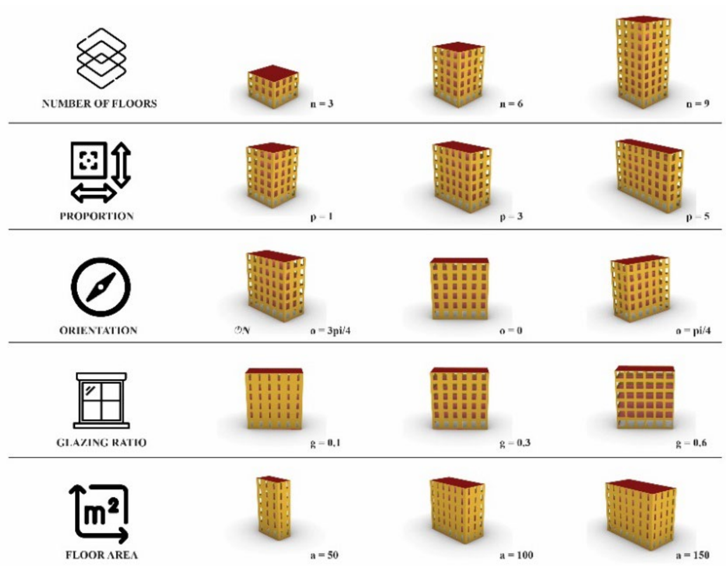

Figure 3. Parameters of a building archetype.

With all the parameters defined, we simulated our surrogate model, which due to the combination of values from different parameter domains, comprises more than 55000 different buildings. Table 2 describes the simulation settings used. The selected output was the required heating and cooling annual load (kWh/m^2).

Table 2. Simulation settings.

Timestep	ShadowCal	SolarDist/Reflex	Output timestep	Output	Context geo.
1	Polygon Clipping	Full exterior	Monthly	Annual loads	None

2.3. REGRESSION MODELS

In this stage, we extract simulation results from the AD tool to generate our training dataset. The dataset used to train our surrogate model comprises the simulation inputs and outputs of all the above-mentioned parameters.

The prediction of energy consumption in a building is a regression problem. Thus, we used different supervised learning models from the Sci-Kit learn package for Julia. From those, to find the most suitable model, we must (1) understand how the parameters affect energy consumption in a building (Araújo et al., 2021), and (2) test multiple models (Wolpert & Macready, 1997). Figure 4 shows interpolations of the energy needs (z-axis) according to our discrete (layers) and continuous (x- and y-axis) parameters step sizes. Since the interpolation graphs in Figure 4 show a polynomial behavior, we selected two linear models with polynomial features (Fan et al., 1995) of degree 3: Linear Regression, and Ridge. Additionally, we tested two ensemble regressors: RandomForest and ExtraTrees (Mendes-Moreira et al., 2012), which are also adequate for such regression problems.

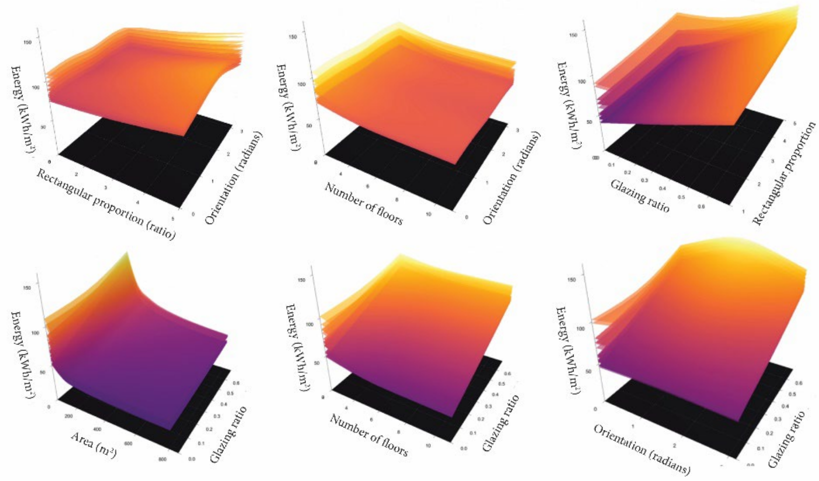

Figure 4. Heating and cooling needs (z-axis) of all the discrete parameters (represented as layers) for different continuous parameters in our dataset (x- and y-axis).

2.4. PREDICTION

After selecting our regression models, we fit them with the simulated dataset and compare our case study's UBEM with the predictions of the surrogate model. Figure 5 shows a histogram of the error, which measures the relative deviation of the surrogate's prediction from the simulated result for each building. From the figure, the best performing model appears to be the *Extra Trees regressor*, which presented the smaller error distribution of ±10 kWh/m^2. However, the *Linear Regression* and *Ridge* models also show errors within the same interval. Thus to complement the error analysis, we used the statistical indexes presented in Table 3. They show that the Extra Trees regressor model is the most accurate, showing the best error results.

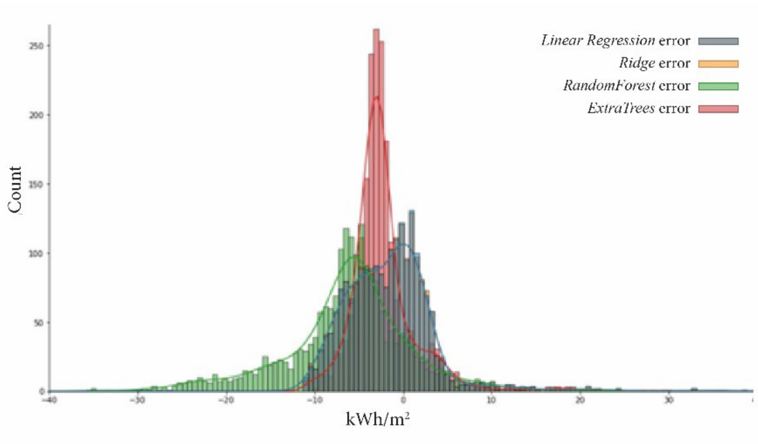

Figure 5. Histogram of the predictions error for each model.

Table 3. Regression model evaluation metrics

	Linear Regression	Ridge	Random Forest	Extra Trees
Mean Error (kWh/m^2)	15.85	15.79	-6.06	-2.21
Root Mean Squared Error (kWh/m^2)	15.40	15.40	9.88	5.44
R^2 score	0.64	0.64	0.85	0.95

2.5. OPTIMIZATION

After selecting the *Extra Trees* regression model with the best score, we tested our surrogate approach in a MOO process on a subset of the case study; a block of 21 buildings illustrated in Figure 6. For the MOO we use the Pareto Genetic Algorithm NSGA-II, which analyzed 800 solutions spread out through 40 generations.

The goals were to find (1) the cheapest and (2) fairest retrofit solution for all buildings that (3) minimized annual heating and cooling loads, i.e.,

$$\mathbf{(1)}\ \boldsymbol{min}\sum_{\boldsymbol{i=1}}^{\boldsymbol{n}}\boldsymbol{Cost_i},\ \mathbf{(2)\ min}\ \ \boldsymbol{\sigma(annual\ loads_i)},\ \mathbf{(3)}\ \boldsymbol{min}\sum_{\boldsymbol{i=1}}^{\boldsymbol{n}}\boldsymbol{annual\ loads_i},$$

Each building is represented as a variable with 4 options for its corresponding construction period: (1) no retrofit, (2) wall retrofit, (3) roof retrofit, and (4) wall and roof retrofit (Table 1). The defined objective functions for the optimization are intended to minimize annual loads (kWh/m^2), cost (€), and standard deviation (σ) to ensure some homogeneity of performance among all buildings (Araújo et al., 2021).

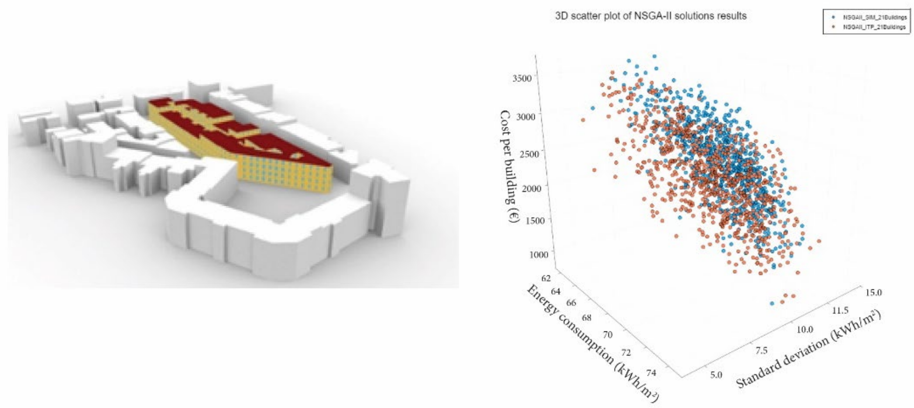

Figure 6. Building subset for optimization (left). Scatter plots of the solutions found in the optimization process (right), blue - conventional, orange - surrogate model optimization.

This MOO experiment used two methods: (1) a simulation-based approach that fully simulates each building block in each MOO evaluation, and (2) using our surrogate model. The goal was to test the effectiveness of using our surrogate approach. The scatter plot in Figure 6 shows the existing trade-offs between the different objectives. One easily seen is that solutions that are more homogenous and have smaller energy consumptions entail more expensive renovations. Additionally, to validate the surrogate model in a MOO process, we computed the cosine similarity between each objective's results for both approaches. This is illustrated in table 4, showing high values of similarity, with the least similar being the cost objective with 0.96.

Table 4: Cosine similarity for each approach's objective.

	Energy Consumption	Standard deviation	Cost per building
Cosine similarity (-1 to 1)	0.99	0.98	0.96

3. Discussion

Table 3 and Figure 5 show high levels of accuracy for the simulation of the original case study in Lisbon of 2193 residential buildings (Figure 2). The model has shown a root mean squared error (RMSE) of 5.44 kWh/m2, and a coefficient of determination (R2) score of 0.95, which explains the target results variance. Additionally, the larger error values come from extrapolations that the prediction model made outside the parameters' domain (e.g., areas of 2000 m2 caused most of the outliers since the maximum simulated area for all archetypes was 800 m2). We can assert that the prediction is very accurate for buildings within the domain of the continuous parameters of our trained surrogate model.

Besides evaluating our model's accuracy, we performed a MOO process with both the surrogate and model-based approaches. Table 4 shows high levels of spatial similarity among the objectives of both approaches MOO process, while Table 5 compares both processes by showing the elapsed time for the simulation of our original case study in Lisbon, and for the MOO process with the 21 building subset from the case study.

Table 5: Elapsed time and obtained results for the full case study simulation and 21 building optimization processes. Simulations were performed using a CPU Intel I7-10700K.

	elapsed time (seconds)	
	Dataset simulation	Optimization
Surrogate model	0.08	791.99
Simulation	5820.00	67516.70

Table 5 shows that the surrogate model largely outperforms the simulation approach and the benefits obtained from model deployment with increased speed and low error rates among simulations. However, the time comparisons made in table 5, comprise only the end-user time to perform the simulation. Thus, it does not consider the time it took to prepare the dataset and train the model, nor the time to set up both approaches. Admittedly, it takes more time to build the surrogate model than performing a simulation, but this time is quickly recovered when the surrogate is used to support numerous analyses.

Nevertheless, our approach bears some limitations. The simulation of the building archetypes does not account for shading and context geometry, which can hinder simulated and predicted results. Additionally, a low number of retrofits were added as discrete parameters, with only one retrofit for each wall and roof. In a situation where one must test multiple construction solutions, the surrogate model would take many discrete parameters, resulting in hundreds of thousands of buildings that need to be modeled and simulated, which would be highly time-consuming.

4. Conclusion

This paper presents a new surrogate model approach to quickly predict the annual heating and cooling loads of large numbers of buildings. The development and validation of our approach imparted five stages. First, we created parametric archetypal building energy models for a case study in Lisbon, Portugal. Then, we automatically modeled and simulated thousands of building archetypes. Subsequently, we utilized the simulation results to create a dataset used to fit regression models to predict simulation outputs according to the discrete and continuous parameters that defined the parametric archetypes. We select the best-performant regression model by testing its accuracy against the simulation results of our original case study in Lisbon with 2193 residential buildings. Finally, we test the effectiveness of our surrogate in a MOO that aimed to find the best retrofit solutions for a residential building block within the dataset. To that end, we compared optimization run-time of the same MOO process using our surrogate and simulation approaches. Results showed a significant decrease in optimization time for the developed surrogate. Thus, we conclude that developing surrogate models based on parametric archetypal buildings avoids the need of running expensive energy simulations and allows the surrogate model to easily adapt to other urban scenarios, opposed to those developed for specific urban areas.

Future improvements to the proposed approach will focus on outputs, accuracy, and user experience. Regarding outputs, we plan to test new metrics such as daylighting and thermal comfort. In terms of accuracy, we aim to extend the capabilities of our parametric building archetypes by including context shading, retrofit solution suites, building typologies (including commercial and office buildings), and different window types. Finally, to improve user experience, we plan to implement an easy-to-use Graphical User Interfaces (GUI) for practitioners, policy-makers, and homeowners to facilitate different urban energy analyzes.

Acknowledgments

This work was supported by international funds through Fundo Europeu de Desenvolvimento Regional (FEDER) with references POCI-01-0247-FEDER-045919, national funds through Fundacão para a Ciência e a Tecnologia (FCT) with references UIDB/50021/2020, and PTDC/ART-DAQ/31061/2017, and PhD grant under contract of FCT 2021.04849.BD.

References

Aguiar, R., Cardoso, C., & Leitão, A. (2017). Algorithmic design and analysis fusing disciplines. *Disciplines and Disruption - Proceedings Catalog of the 37th Annual Conference of the Association for Computer-Aided Design in Architecture, ACADIA* 2017, pp. 28–37.

Araújo, G., Pereira, I., Leitão, A., & Correia Guedes, M. (2021). Conflicts in passive building performance: Retrofit and regulation of informal neighbourhoods. *Frontiers of Architectural Research*. https://doi.org/10.1016/j.foar.2021.02.007

Bamdad, K., Cholette, M. E., & Bell, J. (2020). Building energy optimization using surrogate model and active sampling. *Journal of Building Performance Simulation*, 13(6), 760–776. https://doi.org/10.1080/19401493.2020.1821094

Caetano, I., Santos, L., & Leitão, A. (2020). Computational design in architecture: Defining parametric, generative, and algorithmic design. *Frontiers of Architectural Research*, 9(2), 287–300. https://doi.org/10.1016/j.foar.2019.12.008

Chen, Y., Hong, T., & Piette, M. A. (2017). Automatic generation and simulation of urban building energy models based on city datasets for city-scale building retrofit analysis. *Applied Energy*, 205(April), 323–335. https://doi.org/10.1016/j.apenergy.2017.07.128

European Commission. (2020). *Proposed Mission: 100 Climate-neutral Cities by 2030 – by and for the Citizens*. https://doi.org/10.2777/46063

Fan, J., Heckman, N. E., & Wand, M. P. (1995). Local polynomial kernel regression for generalized linear models and quasi-likelihood functions. *Journal of the American Statistical Association*, 90(429), 141–150. https://doi.org/10.1080/01621459.1995.10476496

Ferrando, M., Causone, F., Hong, T., & Chen, Y. (2020). Urban building energy modeling (UBEM) tools: A state-of-the-art review of bottom-up physics-based approaches. *Sustainable Cities and Society*, 62, 102408. https://doi.org/10.1016/J.SCS.2020.102408

Mendes-Moreira, J., Soares, C., Jorge, A. M., & De Sousa, J. F. (2012). Ensemble approaches for regression: A survey. *ACM Computing Surveys*, 45(1). https://doi.org/10.1145/2379776.2379786

Nations, U. (2020). *The Sustainable Development Goals*. In Design for Global Challenges and Goals. https://doi.org/10.4324/9781003099680-3

Reinhart, C. F., & Cerezo Davila, C. (2016). Urban building energy modeling - A review of a nascent field. *Building and Environment*, 97, 196–202. https://doi.org/10.1016/j.buildenv.2015.12.001

Santos, P. dos, & Matias, L. (2006). Coeficientes de Transmissão Térmica de Elementos da Envolvente dos Edifícios.

Thrampoulidis, E., Mavromatidis, G., Lucchi, A., & Orehounig, K. (2021). A machine learning-based surrogate model to approximate optimal building retrofit solutions. *Applied Energy*, 281, 116024. https://doi.org/10.1016/j.apenergy.2020.116024

United Nations. (2018). *The World's Cities in 2018. In The World's Cities in 2018 - Data Booklet (ST/ESA/ SER.A/417).* https://www.un.org/en/development/desa/population/publications/pdf/urbanization/the_worlds_cities_in_2018_data_booklet.pdf

Wolpert, D. H., & Macready, W. G. (1997). No free lunch theorems for optimization. *IEEE Transactions on Evolutionary Computation*, 1(1), 67–82.

Wortmann, T., Costa, A., Nannicini, G., & Schroepfer, T. (2015). Advantages of surrogate models for architectural design optimization. *Artificial Intelligence for Engineering Design, Analysis, and Manufacturing*, 29(4), 471–481. https://doi.org/10.1017/S0890060415000451