# Green Building Design with OR Techniques

# Abstract

This paper proposes a novel data-driven optimization framework that integrates building design decisions with actual operational data to optimize sustainability across environmental, social, and economic dimensions. The building industry faces increasing pressure to reduce emissions and environmental impacts, but existing design tools do not fully account for how buildings will be operated by occupants. The proposed optimization framework inverts the traditional design-build-operate process to a build-operate-learn-design cycle. It leverages data from sensor monitoring of constructed buildings to validate and improve computational models used for sustainable design decisions. The framework utilizes statistical learning of performance metamodels and decision optimization to explore the complex tradeoffs involved in green building certification standards like passive house, net-zero energy, and others. The focus of this paper is on developing an optimization approach performed on a gradient boost model and randomforest to investigate the causal relationship between building decisions and performance metrics. We integrate both descriptive and prescriptive analytics to inform an optimization model that provides decisions for both the design and operations of a given building. A case study is presented demonstrating the potential benefits of the optimization model compared to conventional methods. The integrative, data-driven optimization approach incorporates both categorical and continuous variables to create higher-performing, more sustainable building designs that better balance Planet, People, and Prosperity goals.

# Introduction (Dr. Jones)

Building design is the cornerstone upon which the art of form and function merge together. A building designer’s goal is to design a structure that not only operates in harmony with its surroundings but also contributes to the well-being of both present and future generations. In the United States, the building sector accounts for ~75% of total electricity consumption and ~40% of total energy use ref[]. As a result, buildings are a significant contributor to greenhouse gas emissions representing around 30% of US emissions. To combat this, the building industry is shifting from employing traditional design approaches to more eco-friendly construction ones in response to these global warming concerns and increasingly strict regulations. Additionally, the steady rise in population growth poses the pressing need for more adequate and affordable housing. Therefore, in recent years, designing cost-effective and sustainable buildings, often referred to as green buildings, have been introduced to accommodate the growing population and the demand for sustainable and affordable housing.

In this research article, we aim to: 1) determine how to design green buildings more efficiently (construct relationship between building design variables and performance metrics); 2) how to make decisions on building characteristics to optimize the three pillars of sustainability (people, planet, profit); 3) and if data driven techniques can perform better than traditional building approaches (We should rely on Dr. Jones and Dr. Chen’s conference paper to claim this, they have expert opinion results in their paper)?

We take outputs from a building simulation model and attempt to create a surrogate machine learning model to predict the outputs of the simulation for any input. Then we optimize that surrogate model to identify the best decisions. The challenge with creating a surrogate model is to gather enough data to make an accurate model and to extrapolate the outcomes of input parameters we did not run. In this paper we investigate a gradient boosted tree and a random forest.

# Background

Environmentally sustainable buildings, known as "green buildings," require simulation and performance analysis during the design phase before construction. While simulation tools allow exploring many potential designs, a comprehensive and organized exploration approach is needed instead of just trial-and-error [1]. The goal is to find a building design that minimizes energy consumption, environmental impact, and costs while maximizing occupant comfort [2], [3].

Green building design and optimization have been explored in various studies, employing techniques such as multi-objective genetic algorithms [4], [5], mixed integer linear programming for leadership in energy and environmental design (LEED) certification [6], and simplified mathematical models [7]. However, many of these approaches have limitations, including the use of simplified models that may not accurately capture the complexity of building performance, the optimization of a single objective without considering trade-offs between multiple objectives, or restricted design space exploration [8][9].

The design and analysis of computer experiments (DACE) approach has been widely used in engineering design optimization, where computer simulations or experiments are employed to model complex systems [10], [11]. Surrogate models, such as response surface models, kriging models, radial basis functions, and multivariate adaptive regression splines (MARS), have been utilized to approximate the behavior of these systems and facilitate optimization [12], [13], [14].

In the context of building design, the DACE approach has the potential to address the challenges of optimizing multiple performance objectives while accounting for both continuous and categorical decision variables. By leveraging building simulation software and constructing accurate metamodels, the DACE methodology can capture the intricate relationships between building parameters and performance metrics, enabling the exploration of a vast design space and the identification of optimal solutions [15], [16].

Recent studies have demonstrated the effectiveness of machine learning (ML) surrogate models, such as gradient boosting, neural networks, and support vector machines, in approximating complex building performance relationships [17], [18], [19]. Gradient boosting, in particular, has gained popularity due to its strong predictive performance, ability to handle heterogeneous variables, and interpretability [20], [21]. It has been successfully applied to various building optimization problems, including energy prediction [22], retrofit analysis [23], and design parameter optimization [24].

The integration of optimization formulations, such as mixed integer linear programming (MILP), further enhances the design exploration process by allowing the consideration of both continuous and categorical decision variables, as well as the incorporation of constraints and objectives specific to the building design problem [25], [26]. Researchers have investigated the integration of physics-based models with data-driven ML surrogates, such as physics-informed neural networks, to enhance prediction accuracy [27], [28].

The combination of high-fidelity simulations, advanced ML surrogate modeling, physics-based approaches, and modern optimization techniques provides a comprehensive framework for green building design optimization. However, challenges remain in improving ML generalizability, incorporating operational data, quantifying uncertainties, and assessing lifecycle sustainability impacts [29], [30].

## Surrogate model optimization methodology

The surrogate model optimization methodology leverages computationally efficient metamodels in place of expensive simulation models to drive the optimization process [31]. By constructing accurate data-driven surrogate models that can mimic the input-output behavior of the original simulations, the optimization can thoroughly explore the design space at low computational cost [32]. This surrogate-based optimization approach involves three key steps: 1) Generating an initial data set by evaluating the simulations at carefully chosen sample points using experimental design techniques [33]. 2) Fitting surrogate models, such as gradient boosting machines, to accurately approximate the simulation outputs based on the sample data [34]. 3) Formulating and solving an optimization problem that uses the surrogate models in place of the original simulations to identify optimal solutions [35]. This surrogate model management framework enables global optimization of complex systems governed by computationally expensive simulations [36].

# Data

## Decision and Response Variable

The decision variables represent the set of input parameters that can be adjusted to achieve desired performance characteristics in the simulation model [34]. For building design problems, these may include factors like insulation levels, window areas, HVAC system specifications, construction materials, building geometry, and more [37]. The decision variables can be a mix of continuous numeric, integers, categories and other data types[38]. The response variables correspond to the key performance metrics of interest that need to be optimized, such as energy consumption, utility costs, environmental impacts, occupant comfort, and others [39]. The relationship between the decision variables and response variables is typically complex and nonlinear, motivating the use of data-driven surrogate models[40].

# Gradient Boosting

Gradient boosting is a powerful ML technique that combines many simple models, typically decision trees, to create a highly accurate ensemble model [41]. It iteratively builds the model by fitting each new tree to the residuals of the previous ensemble, using a gradient descent-like procedure to minimize a loss function. This allows gradient boosting to capture complex nonlinear relationships and handle diverse data types [42]. Gradient boosting has achieved state-of-the-art performance across various domains and has been a winning algorithm in many data science competitions [43]. Its robustness, feature selection capabilities, and ability to handle mixed data types make it well-suited for building surrogate models [44]. Efficient implementations, such as XGBoost [43], LightGBM [45], and CatBoost [46], have further enhanced its scalability and performance.

The core idea behind gradient boosting is to construct each new base model (e.g. decision tree) to provide the best fit to the negative gradient of the loss function associated with the current ensemble model. Some key advantages of gradient boosting over traditional machine learning algorithms are: 1) Automatic handling of nonlinear feature interactions without explicitly engineering interaction terms. 2) Built-in feature selection and robustness to outliers due to the tree structure. 3) Natural ability to model different data types like numeric, categorical, and text features within the tree splitting process. 4) Automatic missing data imputation during tree construction. 5) Regularization mechanisms like shrinkage and tree pruning to prevent overfitting. 6) Highly parallelizable and scalable to very large datasets. These strengths make gradient boosting a flexible and powerful choice for accurate predictive modeling across many domains.

# Methodology

The proposed methodology is based on [47] and combines design of computer experiments, gradient boosting machines for metamodeling, and mathematical optimization techniques. First, relevant building design options and performance metrics are identified. Using experimental design approaches ref[Dr.Chen informs presentation], simulations are run across the design space using simulation software. Next, ensembles of gradient boosted decision trees are trained as metamodels to accurately approximate the complex relationships between design variables and performance outputs from the simulations. These metamodels serve as inexpensive surrogate models capturing the full simulation fidelity. Finally, the metamodels are integrated into an optimization formulation, such as mixed-integer linear/non-linear programming, to identify optimal building configurations balancing multiple competing objectives while satisfying constraints.

I: set of categorical variables

J set of numerical variables

:set of categories for categorical variable i

: building design numerical variables

: building design binary variables for category k of categorical design variable i

Formualtion:

Objetive

For the metamodeling step, gradient boosted trees provide a flexible and accurate approach to approximate the simulation output surfaces. Their ability to automatically handle heterogeneous variables (continuous, categorical, etc.), model high-order nonlinearities and interactions, and learn feature importance makes them well-suited for this task. Ensembling further boosts their predictive power and robustness. The resulting metamodels can closely mimic the full simulations with high accuracy across the design space while being extremely efficient to evaluate - enabling thorough design space exploration.

The metamodels are then embedded into an optimization formulation to identify optimal designs fulfilling the desired performance criteria across objectives like cost, human and environmental impact, etc. Mathematical programming approaches like mixed-integer linear/non-linear programming provide an effective framework to navigate these complex multi-objective, multi-constraint optimization problems. The flexibility of these techniques allows incorporating the metamodels along with relevant constraints on design variables, ensuring identification of feasible optimal solutions. The optimization model enables quantifying trade-offs between competing objectives and generating Pareto-optimal solution sets.

Gradient boost/optimization

1. Data Collection:
   * Identify relevant decision variables (building options) and performance metrics (e.g., energy consumption, cost, environmental impact).
   * Use building simulation software to generate data by running simulations with different combinations of building options.
   * Employ an experimental design approach to efficiently explore the decision variable space.
2. Metamodeling:
   * Construct metamodels, or surrogate models, using treed regression techniques to approximate the relationship between decision variables and performance metrics.
   * Treed regression combines decision tree algorithms with linear regression models, allowing for the handling of both continuous and categorical variables, as well as capturing nonlinear and interaction effects.
   * Develop multi-response treed regression models to simultaneously approximate multiple performance objectives.
3. Optimization:
   * Formulate a mixed integer linear programming (MILP) model that incorporates the treed regression metamodels as objective functions and constraints.
   * Include constraints to ensure feasible building options and explore trade-offs between multiple performance objectives.
   * Solve the MILP model to identify optimal building configurations that balance energy efficiency, cost, and other relevant objectives.

The proposed methodology offers several advantages over traditional building design optimization approaches. First, by leveraging building simulation software and metamodeling techniques, it can capture the complex relationships between building parameters and performance metrics, which are often nonlinear and involve interactions between variables. Second, the use of treed regression allows for the inclusion of both continuous and categorical decision variables, expanding the design space and enabling the consideration of various building materials, systems, and configurations. Third, the MILP formulation enables the optimization of multiple objectives simultaneously, providing a framework for exploring trade-offs and identifying Pareto-optimal solutions.

# Results and Discussion

The paper presents a case study involving the optimization of a single-story residential building ref[Dr. chen presentaiton]. The decision variables included various building options, such as wall and roof insulation levels, window specifications, and occupancy patterns.

Table 1: input data

| variable | description | type | values |
| --- | --- | --- | --- |
| 1 | Roof Ext Finish | Categorical | Light,Dark,Uncolored,Aluminum Paint |
| 2 | Roof Exterior Insulation | Numerical | 0,4,36,35,24,20,12,9,30,28,42,18,14,8,6 |
| 3 | Roof Add’l Insulation | Numerical | 0,19,49,38,11,3,30,13,60,21,7,15,26 |
| 4 | AGW Ext Finish | Categorical | Light,Dark,Uncolored,Medium |
| 5 | AGW Exterior Insulation | Numerical | 10.5,21,2,1.3,7,14,4,12,0,18,6,9,8 |
| 6 | AGW Interior Insulation | Numerical | 4,0,7,6 |
| 7 | AGW Add’l Insulation | Numerical | 13,19,3,11,7,15,0 |
| 8 | Ceilings Batt Insulation | Numerical | 0,21,15,13,7,60,49,45,11,38,3,26,30,19 |
| 9 | Vertical Walls BattInsulation | Numerical | 19,21,0,11,13,30 |
| 10 | GlassCategory | Categorical | Double Low E,Triple Low E,Single Low E,Quadruple Low E |
| 11 | GlassTypeThickness | Numerical | 0.125,0.25 |
| 12 | GlassTypeSpacing | Numerical | 0.25,0.5,0,0.333 |
| 13 | GlassTypeEmissivity | Categorical | High,Low |
| 14 | FrameType | Categorical | Fiberglass\_Fixed\_Mlt spacer,Wood\_Fixed Mlt Spacer,Alum w/ Brk\_ Fixed\_Mtl Spacer,Reinf'd Vinyl\_ Fixed\_Mtl Spacer |
| 15 | System1HeatingSource | Categorical | Electric Resistance,DX Coils |
| 16 | System1SystemType | Categorical | Split System Single Zone,Packaged VVT |
| 17 | SupplyFans | Categorical | Variable,Two-Speed,Forward Curved Centrifugal w/ Inlet Vanes |
| 18 | HeaterFuel | Categorical | Gas,Electricity |
| 19 | HeaterType | Categorical | Storage,instantaneous |

Three performance objectives were considered: annual utility cost, human health, and global warming potentials.

The three regression models, (CART, GB,RF) were developed using data generated from the simulation software, with 80 simulation runs covering different combinations of decision variables. The resulting trees were then incorporated into the Gurobi machine learning package [ref] to optimize the fitted trees.

|  |
| --- |
| Figure1: CART model for Cost |
|  |
| Figure2: CART model for HHP |
|  |
| Figure3: CART model for GWP |

The optimization results identified unique optimal solutions for each performance objective, highlighting the trade-offs between utility costs, human health , and global warming potential.

Table 2: Results of optimization for CART, GB and RF surrogate models

| Variable | CART | GB | RF |
| --- | --- | --- | --- |
| Roof Ext Finish | Aluminum Paint, Dark, Light, Uncolored | Aluminum Paint, Dark, Light, Uncolored |  |
| Roof Exterior Insulation | 0 | 0 |  |
| Roof Add’l Insulation | 0 | 0 |  |
| AGW Ext Finish | Dark, Light, Medium, Uncolored | Dark, Light, Medium, Uncolored |  |
| AGW Exterior Insulation | 5 | 13 |  |
| AGW Interior Insulation | 0 | 2 |  |
| AGW Add’l Insulation | 0 | 9 |  |
| Ceilings Batt Insulation | 0 | 18 |  |
| Vertical Walls BattInsulation | 0 |  |  |
| GlassCategory | Double Low E, Quadruple Low E, Single Low E, Triple Low E | Double Low E, Quadruple Low E, Single Low E |  |
| GlassTypeThickness | 0.125 | 0.125 |  |
| GlassTypeSpacing | 0 | 0.2915 |  |
| GlassTypeEmissivity | High, Low | High, Low |  |
| FrameType | AlumwBrkFixedMtlSpacer, FiberglassFixedMlt spacer, ReinforcedVinylFixedMtlSpacer, WoofFixedMltSpacer | AlumwBrkFixedMtlSpacer, FiberglassFixedMlt spacer, ReinforcedVinylFixedMtlSpacer, WoofFixedMltSpacer |  |
| System1HeatingSource | DX Coils | DX Coils, Electric Resistance |  |
| System1SystemType | Packaged VVT, Split System Single Zone | Packaged VVT, Split System Single Zone |  |
| SupplyFans | Forward Curved Centrifugal w/ Inlet Vanes,  Two-Speed, Variable | Forward Curved Centrifugal w/ Inlet Vanes,  Two-Speed, Variable |  |
| HeaterFuel | Gas | Gas |  |
| HeaterType | Storage,  instantaneous | Storage,  instantaneous |  |
|  | 3726.92 | 3274.05 | 3723.67 |
| Cost | 3726.92 | 3274.045933 |  |
|  | 166043 | 150701 | 162750 |
| GWP | 166043.47 | 150701 |  |
|  | 516.167 | 494.091 | 514.8 |
| HHP | 516.166 | 494.091 |  |

The paper also discusses the potential for future work, including the consideration of Pareto-optimal solutions for multi-objective optimization, incorporating additional performance metrics from various simulation software tools, and extending the methodology to different building types.

## Conclusion

The proposed approach, combining design and analysis of computer experiments (DACE) with mixed integer linear programming (MILP), offers a powerful framework for optimizing building options across multiple performance objectives. By leveraging building simulation software and constructing accurate metamodels using treed regression, this methodology can capture the complex relationships between building parameters and performance metrics, while allowing for the inclusion of both continuous and categorical decision variables.

The MILP formulation enables the simultaneous optimization of multiple objectives, such as cost, human and environmental impact, providing a comprehensive framework for exploring trade-offs and identifying Pareto-optimal solutions. The case study presented in the paper demonstrates the potential of this approach to identify optimal building configurations that balance various performance objectives effectively.

Overall, this methodology contributes to the field of green building design and optimization by providing a robust and flexible approach that can handle the complexity of real-world building design problems while considering multiple objectives and constraints. Future work may involve extending the methodology to different building types, incorporating additional performance metrics, and exploring alternative metamodeling and optimization techniques to further enhance the accuracy and efficiency of the approach.

# References

[1] A. T. Nguyen, S. Reiter, and P. Rigo, “A review on simulation-based optimization methods applied to building performance analysis,” *Appl Energy*, vol. 113, pp. 1043–1058, Jan. 2014, doi: 10.1016/J.APENERGY.2013.08.061.

[2] H. Gao, C. Koch, and Y. Wu, “Building information modelling based building energy modelling: A review,” *Appl Energy*, vol. 238, pp. 320–343, Mar. 2019, doi: 10.1016/J.APENERGY.2019.01.032.

[3] Z. Tian, X. Zhang, X. Jin, X. Zhou, B. Si, and X. Shi, “Towards adoption of building energy simulation and optimization for passive building design: A survey and a review,” *Energy Build*, vol. 158, pp. 1306–1316, Jan. 2018, doi: 10.1016/J.ENBUILD.2017.11.022.

[4] V. Machairas, A. Tsangrassoulis, and K. Axarli, “Algorithms for optimization of building design: A review,” *Renewable and Sustainable Energy Reviews*, vol. 31, pp. 101–112, Mar. 2014, doi: 10.1016/J.RSER.2013.11.036.

[5] R. Evins, “A review of computational optimisation methods applied to sustainable building design,” *Renewable and Sustainable Energy Reviews*, vol. 22, pp. 230–245, Jun. 2013, doi: 10.1016/J.RSER.2013.02.004.

[6] D. Castro-Lacouture, J. A. Sefair, L. Flórez, and A. L. Medaglia, “Optimization model for the selection of materials using a LEED-based green building rating system in Colombia,” *Build Environ*, vol. 44, no. 6, pp. 1162–1170, Jun. 2009, doi: 10.1016/J.BUILDENV.2008.08.009.

[7] T. R. Nielsen, “Simple tool to evaluate energy demand and indoor environment in the early stages of building design,” *Solar Energy*, vol. 78, no. 1, pp. 73–83, Jan. 2005, doi: 10.1016/J.SOLENER.2004.06.016.

[8] X. Shi, Z. Tian, W. Chen, B. Si, and X. Jin, “A review on building energy efficient design optimization rom the perspective of architects,” *Renewable and Sustainable Energy Reviews*, vol. 65, pp. 872–884, Nov. 2016, doi: 10.1016/J.RSER.2016.07.050.

[9] S. Attia, M. Hamdy, W. O’Brien, and S. Carlucci, “Assessing gaps and needs for integrating building performance optimization tools in net zero energy buildings design,” *Energy Build*, vol. 60, pp. 110–124, May 2013, doi: 10.1016/J.ENBUILD.2013.01.016.

[10] G. G. Wang and S. Shan, “Review of Metamodeling Techniques in Support of Engineering Design Optimization,” *Journal of Mechanical Design*, vol. 129, no. 4, pp. 370–380, Apr. 2007, doi: 10.1115/1.2429697.

[11] T. W. Simpson, J. D. Peplinski, P. N. Koch, and J. K. Allen, “Metamodels for computer-based engineering design: Survey and recommendations,” *Eng Comput*, vol. 17, no. 2, pp. 129–150, Feb. 2001, doi: 10.1007/PL00007198/METRICS.

[12] A. I. J. Forrester and A. J. Keane, “Recent advances in surrogate-based optimization,” *Progress in Aerospace Sciences*, vol. 45, no. 1–3, pp. 50–79, Jan. 2009, doi: 10.1016/J.PAEROSCI.2008.11.001.

[13] H. Fang, M. Rais-Rohani, Z. Liu, and M. F. Horstemeyer, “A comparative study of metamodeling methods for multiobjective crashworthiness optimization,” *Comput Struct*, vol. 83, no. 25–26, pp. 2121–2136, Sep. 2005, doi: 10.1016/J.COMPSTRUC.2005.02.025.

[14] R. Jin, W. Chen, and T. W. Simpson, “Comparative studies of metamodelling techniques under multiple modelling criteria,” *Structural and Multidisciplinary Optimization*, vol. 23, no. 1, pp. 1–13, Dec. 2001, doi: 10.1007/S00158-001-0160-4/METRICS.

[15] P. Westermann and R. Evins, “Surrogate modelling for sustainable building design – A review,” *Energy Build*, vol. 198, pp. 170–186, Sep. 2019, doi: 10.1016/J.ENBUILD.2019.05.057.

[16] A. Chakrabarti *et al.*, “Computer-based design synthesis research: An overview,” *J Comput Inf Sci Eng*, vol. 11, no. 2, Jun. 2011, doi: 10.1115/1.3593409/465824.

[17] D. Chakraborty and H. Elzarka, “Advanced machine learning techniques for building performance simulation: a comparative analysis,” *J Build Perform Simul*, vol. 12, no. 2, pp. 193–207, Mar. 2019, doi: 10.1080/19401493.2018.1498538.

[18] S. R. Vutla, S. P. Regalla, and K. Ramaswamy, “Life cycle assessment of cleanroom for micro-electro-mechanical systems fabrication with insights on sustainability,” *J Clean Prod*, vol. 282, p. 124520, Feb. 2021, doi: 10.1016/J.JCLEPRO.2020.124520.

[19] S. Seyedzadeh, F. P. Rahimian, I. Glesk, and M. Roper, “Machine learning for estimation of building energy consumption and performance: a review,” *Visualization in Engineering 2018 6:1*, vol. 6, no. 1, pp. 1–20, Oct. 2018, doi: 10.1186/S40327-018-0064-7.

[20] A. Natekin and A. Knoll, “Gradient boosting machines, a tutorial,” *Front Neurorobot*, vol. 7, no. DEC, p. 63623, Dec. 2013, doi: 10.3389/FNBOT.2013.00021/BIBTEX.

[21] J. H. Friedman, “Greedy function approximation: A gradient boosting machine.,” *https://doi.org/10.1214/aos/1013203451*, vol. 29, no. 5, pp. 1189–1232, Oct. 2001, doi: 10.1214/AOS/1013203451.

[22] S. Seyedzadeh, F. Pour Rahimian, S. Oliver, S. Rodriguez, and I. Glesk, “Machine learning modelling for predicting non-domestic buildings energy performance: A model to support deep energy retrofit decision-making,” *Appl Energy*, vol. 279, p. 115908, Dec. 2020, doi: 10.1016/J.APENERGY.2020.115908.

[23] J. Von Platten, C. Sandels, K. Jörgensson, V. Karlsson, M. Mangold, and K. Mjörnell, “Using Machine Learning to Enrich Building Databases—Methods for Tailored Energy Retrofits,” *Energies 2020, Vol. 13, Page 2574*, vol. 13, no. 10, p. 2574, May 2020, doi: 10.3390/EN13102574.

[24] Y. Zhai, Y. Wang, Y. Huang, and X. Meng, “A multi-objective optimization methodology for window design considering energy consumption, thermal environment and visual performance,” *Renew Energy*, vol. 134, pp. 1190–1199, Apr. 2019, doi: 10.1016/J.RENENE.2018.09.024.

[25] F. Ascione, N. Bianco, R. F. De Masi, G. M. Mauro, and G. P. Vanoli, “Energy retrofit of educational buildings: Transient energy simulations, model calibration and multi-objective optimization towards nearly zero-energy performance,” *Energy Build*, vol. 144, pp. 303–319, Jun. 2017, doi: 10.1016/J.ENBUILD.2017.03.056.

[26] R. Wang, S. Lu, and W. Feng, “A three-stage optimization methodology for envelope design of passive house considering energy demand, thermal comfort and cost,” *Energy*, vol. 192, p. 116723, Feb. 2020, doi: 10.1016/J.ENERGY.2019.116723.

[27] A. Warey, S. Kaushik, B. Khalighi, M. Cruse, and G. Venkatesan, “Data-driven prediction of vehicle cabin thermal comfort: using machine learning and high-fidelity simulation results,” *Int J Heat Mass Transf*, vol. 148, p. 119083, Feb. 2020, doi: 10.1016/J.IJHEATMASSTRANSFER.2019.119083.

[28] M. Raissi, P. Perdikaris, and G. E. Karniadakis, “Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations,” *J Comput Phys*, vol. 378, pp. 686–707, Feb. 2019, doi: 10.1016/J.JCP.2018.10.045.

[29] Y. Li, N. Zhu, and Y. Hou, “A novel hybrid model for building heat load forecasting based on multivariate Empirical modal decomposition,” *Build Environ*, vol. 237, p. 110317, Jun. 2023, doi: 10.1016/J.BUILDENV.2023.110317.

[30] X. Liang, T. Hong, and G. Q. Shen, “Improving the accuracy of energy baseline models for commercial buildings with occupancy data,” *Appl Energy*, vol. 179, pp. 247–260, Oct. 2016, doi: 10.1016/J.APENERGY.2016.06.141.

[31] “Recent advances in surrogate-based optimization - ScienceDirect.” Accessed: Mar. 27, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0376042108000766?via%3Dihub

[32] T. W. Simpson, J. D. Peplinski, P. N. Koch, and J. K. Allen, “Metamodels for computer-based engineering design: Survey and recommendations,” *Eng Comput*, vol. 17, no. 2, pp. 129–150, Feb. 2001, doi: 10.1007/PL00007198/METRICS.

[33] T. J. Santner, B. J. Williams, and W. I. Notz, “The Design and Analysis of Computer Experiments,” 2018, doi: 10.1007/978-1-4939-8847-1.

[34] G. G. Wang and S. Shan, “Review of Metamodeling Techniques in Support of Engineering Design Optimization,” *Journal of Mechanical Design*, vol. 129, no. 4, pp. 370–380, Apr. 2007, doi: 10.1115/1.2429697.

[35] N. V. Queipo, R. T. Haftka, W. Shyy, T. Goel, R. Vaidyanathan, and P. Kevin Tucker, “Surrogate-based analysis and optimization,” *Progress in Aerospace Sciences*, vol. 41, no. 1, pp. 1–28, Jan. 2005, doi: 10.1016/J.PAEROSCI.2005.02.001.

[36] A. I. J. Forrester, A. Sóbester, and A. J. Keane, “Engineering Design via Surrogate Modelling,” *Engineering Design via Surrogate Modelling*, Jul. 2008, doi: 10.1002/9780470770801.

[37] Z. Tian, X. Zhang, X. Jin, X. Zhou, B. Si, and X. Shi, “Towards adoption of building energy simulation and optimization for passive building design: A survey and a review,” *Energy Build*, vol. 158, pp. 1306–1316, Jan. 2018, doi: 10.1016/J.ENBUILD.2017.11.022.

[38] R. Jin, W. Chen, and T. W. Simpson, “Comparative studies of metamodelling techniques under multiple modelling criteria,” *Structural and Multidisciplinary Optimization*, vol. 23, no. 1, pp. 1–13, Dec. 2001, doi: 10.1007/S00158-001-0160-4/METRICS.

[39] A. T. Nguyen, S. Reiter, and P. Rigo, “A review on simulation-based optimization methods applied to building performance analysis,” *Appl Energy*, vol. 113, pp. 1043–1058, Jan. 2014, doi: 10.1016/J.APENERGY.2013.08.061.

[40] P. Westermann and R. Evins, “Surrogate modelling for sustainable building design – A review,” *Energy Build*, vol. 198, pp. 170–186, Sep. 2019, doi: 10.1016/J.ENBUILD.2019.05.057.

[41] J. H. Friedman, “Greedy function approximation: A gradient boosting machine.,” *https://doi.org/10.1214/aos/1013203451*, vol. 29, no. 5, pp. 1189–1232, Oct. 2001, doi: 10.1214/AOS/1013203451.

[42] T. Hastie, R. Tibshirani, and J. Friedman, “The Elements of Statistical Learning,” 2009, doi: 10.1007/978-0-387-84858-7.

[43] T. Chen and C. Guestrin, “XGBoost: A scalable tree boosting system,” *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, vol. 13-17-August-2016, pp. 785–794, Aug. 2016, doi: 10.1145/2939672.2939785.

[44] S. Seyedzadeh, F. Pour Rahimian, S. Oliver, S. Rodriguez, and I. Glesk, “Machine learning modelling for predicting non-domestic buildings energy performance: A model to support deep energy retrofit decision-making,” *Appl Energy*, vol. 279, p. 115908, Dec. 2020, doi: 10.1016/J.APENERGY.2020.115908.

[45] G. Ke *et al.*, “LightGBM: A Highly Efficient Gradient Boosting Decision Tree”, Accessed: Mar. 27, 2024. [Online]. Available: https://github.com/Microsoft/LightGBM.

[46] L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, and A. Gulin, “CatBoost: unbiased boosting with categorical features”, Accessed: Mar. 27, 2024. [Online]. Available: https://github.com/catboost/catboost

[47] V. C. P. Chen, E. C. Jones, V. Sharma, S. Gudikandula, R. Smalls, and J. Sigler, “- Training Teachers to integrate a Sustainability Assessment for Green Building and Retrofitting Project into a K-12 Classroom ,” *ASEE 2024 Annual Conference*.