ORNL/TM-2021/170499

Design Space Data: Informing Common Design Decisions with Pre-Simulated Data

Brett Bass
Leland Curtis
Joshua New

November 2021

ORNL/TM-2021/170499

Energy and Transportation Science Division

DESIGN-SPACE DATA: INFORMING COMMON DESIGN DECISIONS WITH PRE-SIMULATED DATA

Brett Bass
Leland Curtis
Joshua New

November 2021

Prepared by
OAK RIDGE NATIONAL LABORATORY
Oak Ridge, TN 37831
managed by
UT-BATTELLE LLC
for the
US DEPARTMENT OF ENERGY
under contract DE-AC05-00OR22725

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

AEC	architecture, engineering, and construction
AI	artificial intelligence
AutoBEM	Automatic Building Energy Modeling
DSE	design space exploration
EUI	energy use intensity
ML	machine learning
UDSE	universal design space exploration

ABSTRACT

Design-space exploration analysis techniques represent a data-centric approach to integrating performance analysis during a project's early design phases—when the greatest potential exists to efficiently and cost effectively improve the energy efficiency of a building. This work describes a novel extension of design-space exploration, called universal design space exploration (UDSE), which leverages massive databases of pre-simulated analysis that represent all possible outcomes of common analysis workflows. The resulting databases, called *design spaces*, become universal when a single pre-simulated design space can be reapplied to future unknown projects. Unlike current simulation methods, which require a design to exist before it can be analyzed and often require minutes or hours to simulate, UDSE uses pre-simulation to deliver rapid and relevant insight as new designs are conceptualized. The data underpinning UDSE enables advanced statistical and artificial intelligence methods, which allow UDSE to deliver a greater understanding of the larger problem being explored rather than simply delivering analysis of several preconceived design options. UDSE has the potential to provide instantaneous, relevant analysis for all building design projects at a negligible cost. This paper has two main goals: to develop a relevant universal design space that showcases the potential of UDSE and to release these data freely to industry and academia, thereby removing obstacles to digital literacy in statistics, machine learning, and artificial intelligence within the architecture, engineering, and construction industries.

1. BACKGROUND

1.1 EARLY CONCEPTUAL PERFORMANCE ANALYSIS ISSUES

Architects are under increasing pressure to design energy-efficient buildings, and they need high-quality performance analysis to inform their design decisions. Traditional performance analysis workflows are well suited for the later stages of the design process when the detailed inputs required for accurate results have been determined, and when the analysis delivered a week or two after it was requested is still relevant. Traditional performance analysis workflows are not effective in early conceptual design phases when major design changes occur by the hour, and when key inputs are undetermined. However, building designers can have the greatest contribution to performance at the least cost during the early design stages because the design is still flexible. Unfortunately, performance analysis is not often performed for the design phases in which it would be most valuable.

To analyze a design, it must already be available. Analyzing the contribution of design decisions can only happen after those decisions have been made. Most designers follow an *optioneering* design approach in which several design variations or options are developed and analyzed to determine which approach is best. Performance analysis is delivered only after the options have been generated; decisions made during the development of these options are supported by experience or rules of thumb. Energy-assessment tools can provide insight into an architect's performance after the fact by heavily leveraging industry standards or best practices. The difference between typical design and utilization of design space are shown in Figure 1.

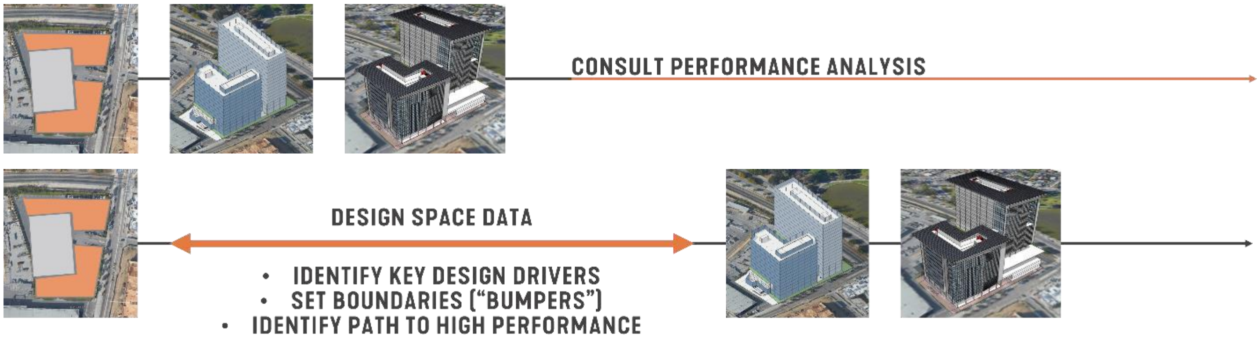

Figure 1. Design challenge flowchart illustrating how Universal Design Space Exploration (UDSE) can improve the designprocess by providing energy performance feedback on design decisions during the design process whereas traditional design cannot.

1.2 UNIVERSAL DESIGN SPACE EXPLORATION

Universal design space exploration (UDSE) has the potential to deliver instantaneous, high-quality insight and analysis for building design projects at a minimal cost. The key is in the term *universal*. Design-space exploration (DSE) techniques leverage iterative simulation of parametric models to generate databases of simulation analysis that represent a comprehensive sampling of all possible design scenarios defined by the parametric model (the design space). These data become a map of the design space, which can then be explored and mined for insight. This approach addresses some of the limitations of traditional performance analysis workflows: rather than respond to several predetermined design options, DSE provides analysis of likely alternatives and variations of these predetermined options, thereby allowing the next steps to be guided by analysis rather than experience or rules of thumb alone. Analysis of alternatives is instantaneous because the simulations and resulting analysis already exist. This approach supports real-time discussion of priorities and alternatives within integrated teams, which enhances the significance of the analysis. However, DSE is time-consuming. Simulating the hundreds or thousands of iterations required to make a useful design space can take days or weeks, thereby undermining the speed advantage of instantaneous feedback. DSE is also inflexible. Its analysis is only applicable to the design possibilities contained within the design space. By the time the design space is generated, the current design has likely evolved and left the previous design space. Expanding the design space to account for more possibilities further increases the simulation time required. These limitations prevent DSE from being widely applied in a conceptual design environment characterized by rapid change. Instead, DSE is more often applied toward specific, well-defined challenges in design development in which multiparameter exploration and optimization are required.

UDSE addresses the limitations of traditional DSE and leverages the benefits of DSE during the early conceptual design phases in which instantaneous, free, and forward-looking analysis is most needed. UDSE accomplishes this task by modeling common problems. Common problems are, by definition, well-defined and repeatable across multiple projects. If a well-understood problem is expected on a future project, then a design space can be generated long before the project requires it, which avoids the speed and interactivity issues when a project requires new, massive building simulations. If speed is no longer an issue, then the design space can become gigantic to cover more possible scenarios, thus becoming more flexible and applicable to more projects (i.e., universal). If a problem is common, then defining the parameters is more predictable, the resulting design space is more applicable, and the investment required to produce the simulation is recouped through multiple uses. In this way, a gigantic design space that encompasses a common problem becomes universal, extending across all designs, schedules, and project fees.

UDSE introduces new opportunities for statistical analysis, machine learning (ML), and app development. Unlike traditional DSE, which is typically limited to 3–7 dimensions that can be manually explored and understood, UDSE generates massive multidimensional design spaces that are difficult to comprehend using typical analysis. Common visualization methods, such as a scatter plot, can show at most 4 dimensions (*x*, *y*, color, dot shape) before becoming unintelligible. Universal design spaces can easily support more than 15 dimensions. Exploring and analyzing such large spaces requires engineers to learn tools typically employed by big data scientists. The design space exploration process is shown in Figure 2.

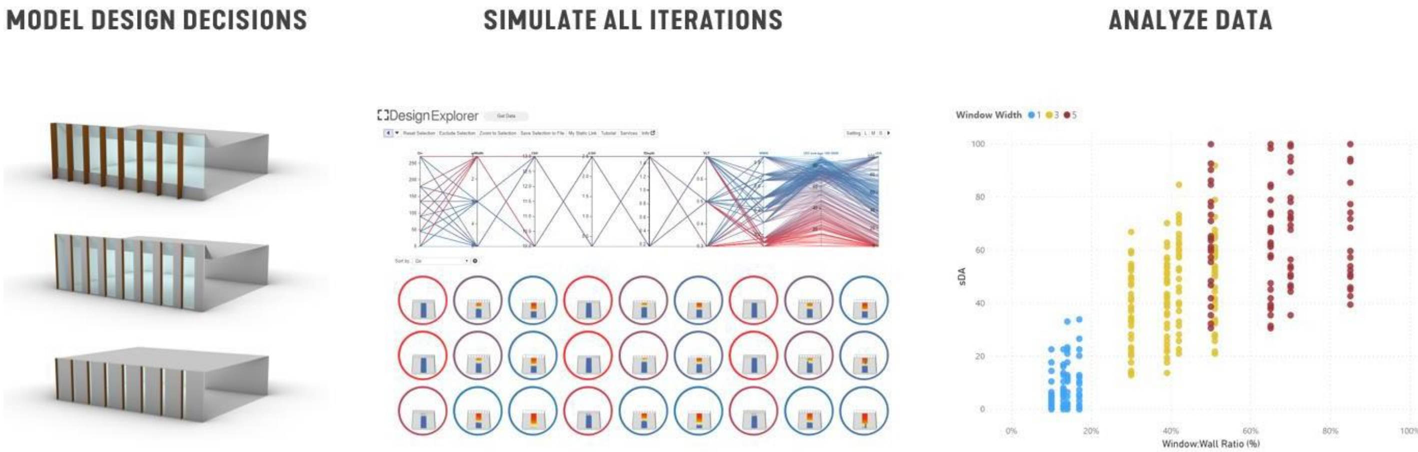

Figure 2. The DSE process.(left) Selection of design parameters, (middle) sample-based simulation ofthose parameters, and (right) summary analysis of results using metrics that inform better user decisions.

1.3 PUBLIC DATA TO SUPPORT DIGITAL LITERACY IN ARCHITECTURE, ENGINEERING, AND CONSTRUCTION

ML and artificial intelligence (AI) are transforming industries around the world, but the architecture, engineering, and construction (AEC) industry typically views these as buzzwords with few useful applications within design. The industry may be better served if these words were seen as algorithmic tools that leverage data to solve problems. The AEC industry has many problems to solve but few people who know how to deploy these novel, powerful, algorithmic tools.

Adopting these tools presents challenges. Few designers invest in learning the skills required to deploy novel tools, especially for market segments in which few examples showcase the value such skills might bring. This report describes one data-centric approach to leveraging ML within the design process. Most existing ML examples and sample data sets are designed for other industries, so AEC professionals and students have difficulty recognizing or testing the advantages that new digital methods might offer. Generating AEC-specific data requires a significant investment of time and expertise, especially in the field of performance analysis for which simulations can take hours per design iteration. This computational bottleneck is one of the problems UDSE promises to alleviate, yet it remains an obstacle. Freely providing universal design spaces that frame common design questions to industry and academia will reduce the obstacles to the development of next-generation digital tools and applications.

The methodology described in this report will enable other AEC firms, students, and start-ups to use large and complex building simulation data by developing innovative, data-centric, AI-enabled, and interactive design apps that have the potential to advance the market in unexpectedly useful ways. Expanding the reach of performance modeling to more firms and project types should result in reduced energy usage and improved building quality across the industry.

1.4 AUTOMATIC BUILDING ENERGY MODELING

The US Department of Energy's Oak Ridge National Laboratory has developed a collection of software and algorithms called Automatic Building Energy Modeling (AutoBEM), which enables users to model building energy for each structure at large geographic scales (New 2021). Within AutoBEM, building properties are detected, inferred, or predicted as inputs to generate building energy models using OpenStudio. The primary building properties necessary to generate building energy models using AutoBEM include physical characteristics, such as building footprint and height, and performance characteristics, such as building type and vintage (building code). These models are then simulated using EnergyPlus.OpenStudio is a collection of open-source software tools to support energy modeling in EnergyPlus, which is a physical building energy simulation engine (NREL 2021a; NREL 2021b). The simulation results can be customized to provide any combination of thousands of simulation outputs, which typically include building energy end uses. These results can be aggregated, analyzed, and visualized in a variety of different ways to gain insights about a building or group of buildings. AutoBEM has been used to model a utility service area (180,000 buildings) in Chattanooga, Tennessee and to create a coarse model of every building in the United States (123 million buildings).

2. REFINE FEATURES, METRICS, AND SAMPLING METHODOLOGY

The features, metrics, and sampling methodology of the design space exploration were developed by surveying designers and engineers. These experts defined which areas of the design space were most important and should be selected for simulation. The final parameter and sampling space included design parameters known to be building properties of high importance and less commonly used features that may affect building energy performance under different design conditions. The variables considered for the parametric sampling are shown in Table 1.

Table 1. Parametric sampling of design space variables, resulting in a total of about 3.35 million individual buildings

Sampling parameter	Inputs	Sampling parameter	Inputs
Program type	Higher education	Plate depth	Low
	Lab - high intensity		Typical
	Office		High
	Hospital	Floor-to-floor height	Low
	Healthcare - outpatient		Typical
	Residential		High
Climate zone	1A	Solar design	Bad
	1B		Typical
	2A		Good
	2B	Average window-to-wall ratio	0.25
	3A		0.4
	3B		0.7
	3C	Envelope quality	Baseline
	4A		High
	4B		Ultra
	4C	Construction type	Common
	5A		Less Common
	5B	Lighting power density	Baseline
	5C		Better
	6A		Best
	6B	HVAC system	Baseline
	7A		Good
	7B		Great
Total square footage	Low		Ultra
	Typical	Set points	Baseline
	High		Expanded
Target floor area	Low		
	Typical		
	High		

3. PARAMETRIC SAMPLING, BUILDING MODEL GENERATION, AND SIMULATION

3.1 PARAMETRIC SAMPLING AND BUILDING MODEL GENERATION

To generate the parametric sampling for building-energy models, the AutoBEM 1.0 workflow was modified to generate custom buildings rather than existing buildings. A previously developed data set of 122.9 million US buildings (https://doi.ccs.ornl.gov/ui/doi/339) was used to select building geometries that fit a parametric sampling of simulation inputs commonly considered during building design. Buildings were simplified to the relevant footprint area and used only four vertices for a rectangular footprint. Because footprint geometry can be extruded to any user-defined height, only the footprint area, plate length, and plate width were considered for geometry selection. If a building geometry was selected, then the orientation of the building was stored so it could be rotated properly after generation.

Each building's data was saved as a row of a table; additional design parameters could be appended to the table as columns to complete the design space. This building data table is one of several inputs required for AutoBEM's building-energy model generation. Several physical and functional building properties could be added to the table, with a new row for each variation of the design parameters. The following parameters (i.e., columns of the building data table) are required parameters in this project:

- Building type
- Building energy code (standard 90.1-2013 for this study)
- Climate zone
- Window-to-wall ratio (typical)
- Floor-to-floor height
- Number of floors (total area ÷ floor area)
- Height (number of floors × floor-to-floor height)
- HVAC type

The following optional design parameters are part of the design space and can be added to a building's row in the data table for model generation:

- Wall R-value
- Roof R-value
- Window U-value
- Window solar heat gain coefficient
- Lighting adjustment
- Setpoint adjustment
- Solar design adjustment
- Unique ID based on all distinct parameters

The typical AutoBEM workflow implements a default HVAC type for each building type, vintage, and climate zone combination. This default was adjusted to include 16 new HVAC types, each with multiple settings, for AutoBEM's generation process.

The building input table initially consisted of 3.3 million rows of buildings. This number was randomly sampled down to 256,000 for computational feasibility and initial analysis. This initial analysis indicated that a high-quality AI, with a very small difference between actual simulation results and those from the AI-generated surrogate model, could be trained on a much smaller sample of the data. The models were then generated using AutoBEM and executed as a fully parallelized workload on a 72-core server. The model generation times are shown in Table 2.

Table 2. Sample of building model generation timings; larger buildings typically take longer to generate

Building area (ft^2)	Number of buildings	Time (s)	Buildings generated per second
<50	124	24	14
<50	508	125	18
<50	128	31	17
<50	951	189	14
50–150	2,004	535	19
50–150	449	214	34
50–150	3,156	687	16
>150	3,536	1,509	31
>150	285	270	68
All	11,141	3,584	231

Two model post-generation steps were completed using OpenStudio measures and an application programming interface to make changes to the model of a building. OpenStudio measures are typically used for more complex building changes (e.g., HVAC changeouts), whereas smaller changes can be made via EnergyPlus measures or direct text-based replacement in the EnergyPlus model. The first OpenStudio measure invoked was building rotation for solar design. Building models were rotated a variable amount based on their original orientation to 0° or 90° north. Another measure was then used to adjust the window-to-wall ratio to either a good or bad solar design by adding all of the windows to two faces and removing windows from the other two (the typical prototype model solar design kept even window spacing across all four faces).

The post-generation adjustments were implemented in the models using EPPY, a python library used for editing EnergyPlus models (Santosh 2021). These changes included adjustments to wall and roof insulation (R-value), window insulation (U-value), solar heat gain coefficient, lighting power density, and heating and cooling coefficients of performance.

A final python script defined the simulation output files, variables, and frequency to be generated. A tabular hypertext markup file (with HVAC sizing data), comma-separated values file (time series end-use energy data), and log files were output for each simulation in this analysis. The settings for the output included 14 reporting variables at a monthly resolution.

3.2 BUILDING MODEL SIMULATION

The building models are simulated using Python and EnergyPlus to run in parallel on the 72-core server. Another parallelized python script is then used to aggregate building simulation results, including annual simulation energy data. This output data can be joined with the input data to obtain a single table, including all relevant inputs and outputs for the design space.

A representative sample of 4,506 building models took 1,023 s to simulate on a 72-core server. The average simulation time of each of these buildings was 16 s. A summary flowchart illustrating each step of the DSE process is shown in Figure 3.

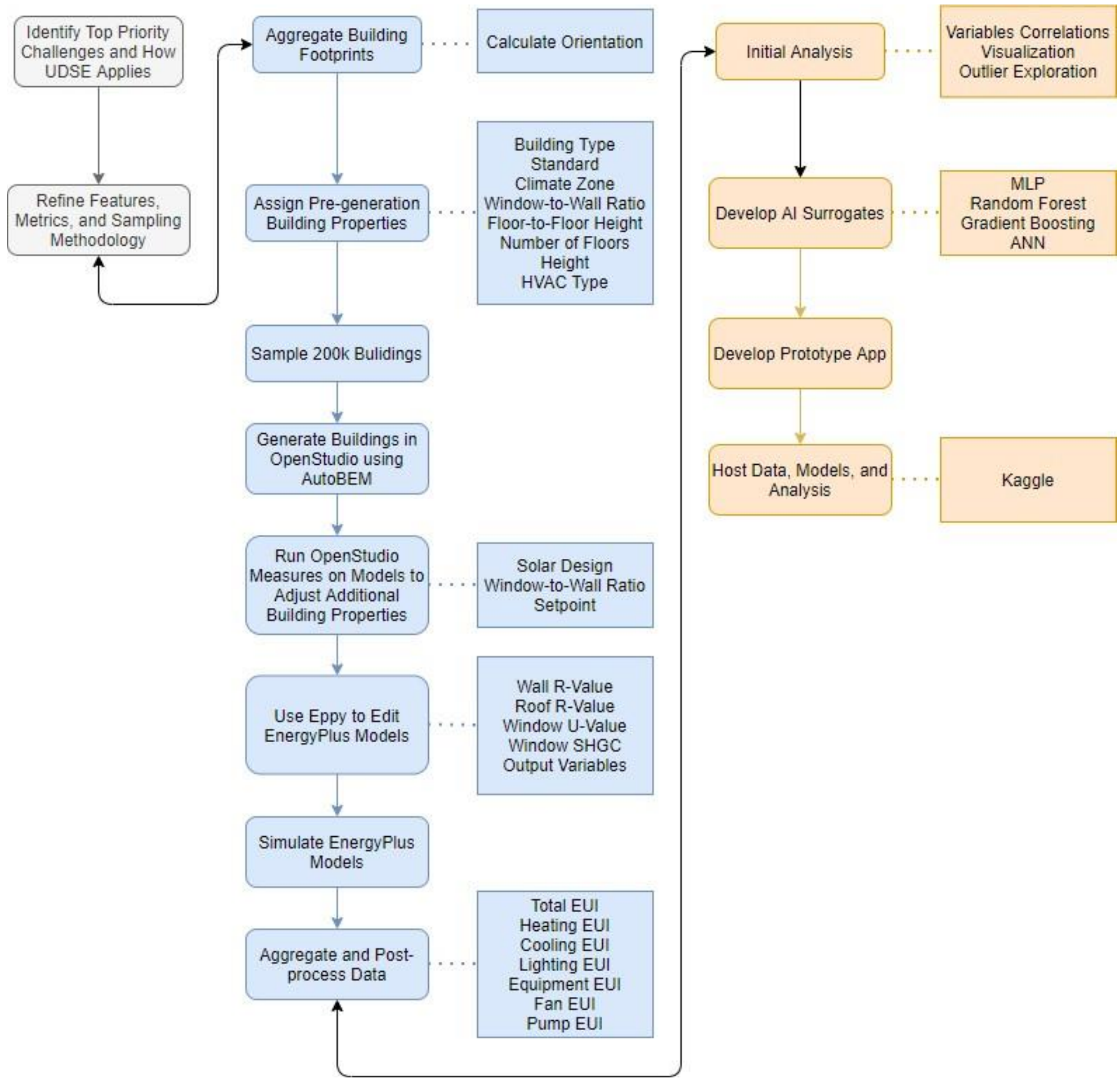

Figure 3. Flow chart of the process to generate the UDSE data set. The generated UDSE data set involved (left) development of the design space, (middle) generation and simulation of the models, and (right) analysis-based development of derivative technology to improve design decisions.

4. INITIAL RESULTS ANALYSIS

4.1 DATA VISUALIZATION

Initial data visualization, including box plots and violin charts, illustrate statistical trends and distributions of relationships among input and output variables within the design space. Figures 4 and 5 display energy use intensity (EUI), which is a well-recognized metric used by architects to normalize energy use by the floor area of the building. A violin plot comparing EUI per building type for each of the HVAC settings is shown in Figure 4, and a box plot showing the end use of each of these energy building types is shown in Figure 5. The end uses of each building include the energy use of total, electricity, cooling, natural gas, lighting, equipment, fans, pumps, heat rejection, and heat recovery.

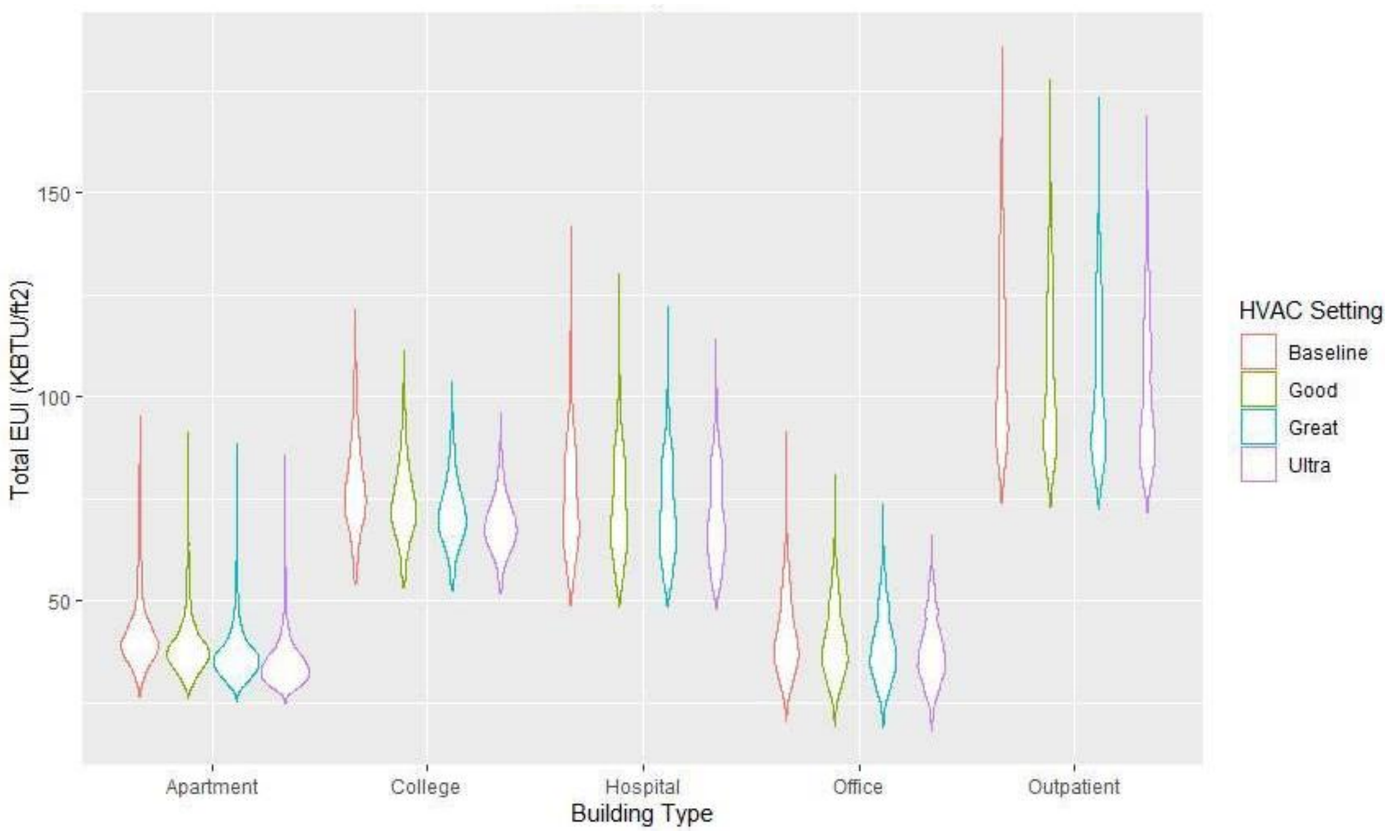

Figure 4. Total building EUI shown for five building classes. Laboratory not included for scale.

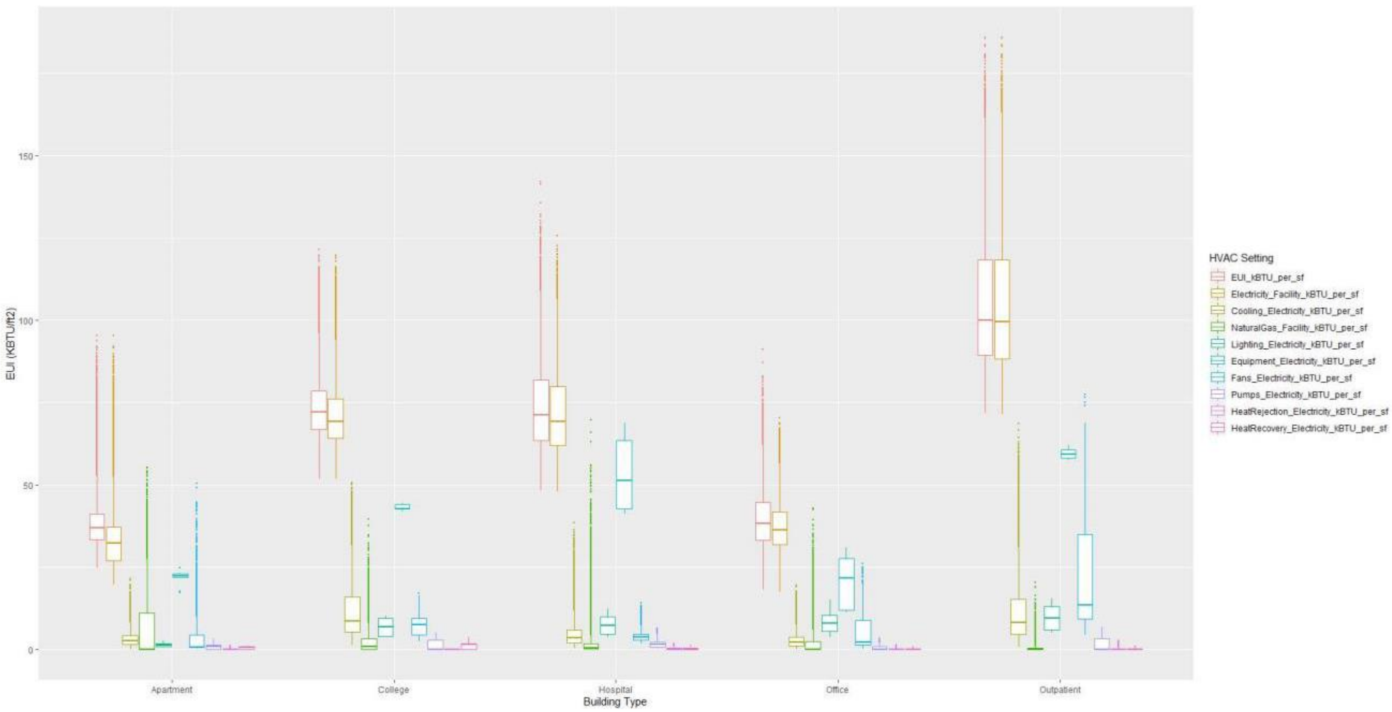

Figure 5. Building EUI split by end use for five building classes. Laboratory not included for scale.

4.2 CORRELATIONS

Relationships between building properties and building energy use is a critical component of UDSE. The data were filtered by building type and climate zone. Pearson, Spearman, and Kendall correlations between each input variable and building EUI were calculated. The Spearman and Kendall correlation statistics were employed explicitly because they can process ordinal variables, of which there were several in the feature set. These correlation values are shown in Table 3 for each building type and climate zone.

Table 3. Pearson input variable correlation coefficients to building EUI for building types and climate zones (Hospital omitted for similarity to outpatient)

	College _1A	College _2A	College _2B	College _3A	College _3B	College _3C	College _4A	College _4B	College _4C	College _5A	College _5B	College _6A	College _6B	College _7A
Height	0.25	0.20	0.26	0.35	0.34	0.17	0.26	0.30	0.25	0.30	0.30	0.28	0.27	0.40
NumberFloors	0.24	0.20	0.27	0.33	0.30	0.21	0.24	0.28	0.27	0.32	0.28	0.27	0.29	0.39
TotalArea	-0.06	-0.06	-0.04	-0.07	-0.02	-0.23	-0.08	-0.06	-0.09	0.10	-0.02	-0.03	-0.09	-0.12
WindowWallRatio	0.09	0.04	0.13	0.05	0.12	0.01	0.02	0.12	0.08	0.06	0.07	0.12	0.20	0.12
FloorHeight	0.03	0.03	0.01	0.11	0.19	-0.09	0.08	0.06	-0.08	-0.03	0.09	0.05	-0.08	0.08
PlateDepth	0.06	0.05	0.07	0.18	0.08	0.14	0.11	-0.02	0.20	0.06	0.09	0.00	0.04	0.02
PlateLength	0.14	0.07	0.12	0.12	0.08	0.12	0.15	0.22	0.26	0.21	0.11	0.23	0.33	0.30
SkinArea	0.26	0.20	0.27	0.35	0.30	0.21	0.34	0.34	0.42	0.37	0.32	0.33	0.40	0.43
SkinFloorRatio	0.26	0.21	0.27	0.33	0.28	0.25	0.34	0.35	0.43	0.35	0.32	0.33	0.40	0.44
GlassArea	0.28	0.19	0.29	0.33	0.32	0.21	0.34	0.34	0.43	0.37	0.34	0.35	0.45	0.46
EnvelopeFloorAreaRatio	0.26	0.21	0.27	0.33	0.28	0.25	0.34	0.34	0.43	0.35	0.32	0.33	0.40	0.43
EnvelopeQuality	-0.04	0.01	0.01	-0.02	-0.03	0.15	-0.04	0.14	0.01	-0.18	-0.02	-0.31	-0.21	-0.28
LightingPowerDensity	-0.37	-0.38	-0.40	-0.52	-0.49	-0.80	-0.55	-0.58	-0.59	-0.50	-0.48	-0.37	-0.50	-0.39
SetpointSetting	-0.24	-0.17	-0.38	-0.12	-0.41	-0.35	-0.24	-0.37	-0.37	-0.30	-0.33	-0.31	-0.25	-0.27
HVACSetting	0.67	0.68	0.58	0.61	0.52	0.23	0.47	0.45	0.28	0.39	0.49	0.41	0.34	0.35
SolarDesign	-0.04	-0.04	-0.05	-0.05	-0.06	-0.01	-0.09	-0.06	-0.07	-0.12	-0.06	-0.10	-0.11	-0.13
TotalBuildingRValue	-0.05	0.00	0.02	-0.01	-0.02	0.16	-0.08	0.13	-0.04	-0.18	-0.02	-0.32	-0.21	-0.27
AvgSkinRValue	0.00	-0.02	-0.11	-0.02	-0.07	-0.14	0.02	-0.17	-0.06	0.04	-0.06	0.05	-0.10	-0.05
WallRValue	-0.05	0.00	0.02	-0.01	-0.02	0.16	-0.06	0.14	-0.02	-0.17	-0.02	-0.32	-0.21	-0.20
GlassFrameUValue	0.03	-0.02	-0.01	0.02	0.03	-0.14	0.04	-0.14	0.00	0.18	0.02	0.31	0.20	0.29
RoofRValue	-0.05	-0.01	0.03	-0.01	-0.02	0.16	-0.08	0.13	-0.05	-0.18	-0.02	-0.31	-0.20	-0.28
TotalBuildingOpaqueRValue	-0.05	0.00	0.03	-0.01	-0.02	0.16	-0.08	0.13	-0.04	-0.18	-0.02	-0.32	-0.21	-0.26

	Laboratory_1A	Laboratory_2A	Laboratory_2B	Laboratory_3A	Laboratory_3B	Laboratory_3C	Laboratory_4A	Laboratory_4B	Laboratory_4C	Laboratory_5A	Laboratory_5B	Laboratory_6A	Laboratory_6B	Laboratory_7A
Height	0.08	0.14	0.13	0.19	0.14	0.19	0.20	0.17	0.23	0.26	0.16	0.25	0.27	0.27
NumberFloors	-0.05	0.02	0.00	0.06	0.00	0.06	0.05	0.06	0.10	0.15	0.00	0.11	0.14	0.12
TotalArea	-0.19	-0.09	-0.11	-0.06	-0.13	-0.01	-0.03	-0.06	-0.02	0.06	-0.04	0.02	-0.06	0.02
WindowWallRatio	-0.03	-0.02	-0.02	-0.02	-0.01	-0.01	-0.01	0.01	-0.03	0.03	0.00	-0.04	-0.03	-0.01
FloorHeight	0.42	0.41	0.43	0.45	0.46	0.45	0.47	0.48	0.47	0.46	0.49	0.49	0.51	0.49
PlateDepth	0.04	0.00	0.03	-0.02	-0.07	0.07	0.01	0.03	0.06	0.01	-0.01	0.00	-0.05	-0.05
PlateLength	0.02	0.02	-0.01	-0.04	0.10	0.02	0.01	0.04	0.08	0.01	0.00	0.00	0.03	0.02
SkinArea	0.04	0.07	0.05	0.07	0.12	0.12	0.09	0.11	0.17	0.15	0.09	0.13	0.15	0.16
SkinFloorRatio	0.11	0.12	0.09	0.10	0.17	0.13	0.12	0.14	0.19	0.14	0.11	0.14	0.16	0.16
GlassArea	0.02	0.05	0.03	0.04	0.09	0.10	0.06	0.09	0.14	0.12	0.07	0.09	0.10	0.12
EnvelopeFloorAreaRatio	0.11	0.12	0.09	0.10	0.17	0.13	0.12	0.14	0.19	0.14	0.11	0.13	0.16	0.16
EnvelopeQuality	0.11	0.00	0.01	0.00	0.03	-0.11	-0.07	-0.11	-0.10	-0.01	-0.06	-0.07	-0.16	-0.05
LightingPowerDensity	-0.01	-0.03	0.01	-0.03	-0.01	0.02	-0.01	-0.03	-0.02	-0.04	0.01	-0.01	-0.01	-0.01
SetpointSetting	-0.26	-0.25	-0.39	-0.24	-0.40	-0.33	-0.39	-0.43	-0.39	-0.38	-0.41	-0.35	-0.39	-0.32
HVACSetting	0.68	0.69	0.63	0.69	0.62	0.66	0.63	0.60	0.63	0.64	0.60	0.63	0.61	0.65
SolarDesign	0.00	0.00	0.01	0.00	0.01	0.01	0.00	0.01	0.00	0.00	0.00	0.00	0.01	0.00
TotalBuildingRValue	0.11	0.01	0.03	0.03	0.03	-0.11	-0.03	-0.10	-0.08	-0.01	-0.06	-0.07	-0.16	-0.05
AvgSkinRValue	-0.08	0.01	-0.01	-0.02	-0.01	0.09	0.04	0.07	0.09	-0.02	0.03	0.06	0.09	0.02

Note: Solid red indicates positive correlation (close to 1), solid blue indicates negative correlation (closeto −1), and white indicates no correlation (close to 0).

Table 3. Pearson input variable correlation coefficients to building EUI for building types and climate zones (Hospital omitted for similarity to outpatient) (continued)

	Laboratory_1A	**Laboratory_2A**	**Laboratory_2B**	**Laboratory_3A**	**Laboratory_3B**	**Laboratory_3C**	**Laboratory_4A**	**Laboratory_4B**	**Laboratory_4C**	**Laboratory_5A**	**Laboratory_5B**	**Laboratory_6A**	**Laboratory_6B**	**Laboratory_7A**
WallRValue	0.11	0.01	0.03	0.03	0.03	-0.11	-0.05	-0.10	-0.09	0.00	-0.06	-0.05	-0.15	-0.04
GlassFrameUValue	-0.10	0.00	0.00	0.01	-0.02	0.11	0.06	0.11	0.10	0.01	0.06	0.07	0.16	0.05
RoofRValue	0.11	0.01	0.03	0.03	0.03	-0.11	-0.02	-0.10	-0.07	-0.01	-0.06	-0.07	-0.16	-0.05
TotalBuildingOpaqueRValue	0.11	0.01	0.03	0.03	0.03	-0.11	-0.03	-0.10	-0.08	-0.01	-0.06	-0.07	-0.16	-0.05
	Outpatient_1A	**Outpatient_2A**	**Outpatient_2B**	**Outpatient_3A**	**Outpatient_3B**	**Outpatient_3C**	**Outpatient_4A**	**Outpatient_4B**	**Outpatient_4C**	**Outpatient_5A**	**Outpatient_5B**	**Outpatient_6A**	**Outpatient_6B**	**Outpatient_7A**
Height	-0.50	-0.57	-0.55	-0.55	-0.48	-0.47	-0.42	-0.50	-0.44	-0.50	-0.43	-0.44	-0.44	-0.44
NumberFloors	-0.50	-0.56	-0.56	-0.57	-0.50	-0.49	-0.45	-0.51	-0.45	-0.51	-0.45	-0.45	-0.47	-0.46
TotalArea	-0.77	-0.79	-0.77	-0.78	-0.75	-0.70	-0.67	-0.68	-0.66	-0.68	-0.66	-0.65	-0.68	-0.62
WindowWallRatio	-0.02	-0.06	0.07	0.11	0.02	0.00	0.00	-0.02	-0.01	0.07	-0.01	0.05	0.11	0.19
FloorHeight	0.06	0.00	0.01	-0.04	0.13	0.00	0.19	0.11	0.12	0.02	0.08	-0.01	0.10	0.15
PlateDepth	0.05	0.14	-0.05	-0.13	0.03	-0.05	0.08	0.03	-0.11	-0.12	-0.08	-0.02	-0.17	-0.24
PlateLength	-0.13	-0.22	-0.14	-0.09	-0.15	-0.18	-0.16	-0.23	-0.16	-0.15	-0.10	-0.09	0.02	-0.14
SkinArea	-0.47	-0.52	-0.54	-0.51	-0.46	-0.48	-0.39	-0.46	-0.44	-0.44	-0.40	-0.40	-0.43	-0.45
SkinFloorRatio	0.28	0.25	0.06	0.17	0.21	0.10	0.26	0.17	0.19	0.11	0.20	0.22	0.32	0.02
GlassArea	-0.43	-0.50	-0.47	-0.46	-0.41	-0.45	-0.36	-0.41	-0.40	-0.39	-0.37	-0.34	-0.37	-0.38
EnvelopeFloorAreaRatio	0.30	0.27	0.08	0.19	0.24	0.13	0.28	0.19	0.21	0.14	0.22	0.25	0.34	0.04
EnvelopeQuality	0.04	-0.04	0.01	0.01	0.09	0.10	0.13	-0.06	0.13	-0.01	-0.07	-0.04	-0.05	0.06
LightingPowerDensity	-0.18	-0.47	-0.25	-0.37	-0.32	-0.26	-0.17	-0.19	-0.24	-0.27	-0.12	-0.24	-0.25	-0.09
SetpointSetting	-0.14	-0.14	-0.22	-0.12	-0.21	-0.28	-0.18	-0.20	-0.15	-0.17	-0.19	-0.17	-0.14	-0.14
HVACSetting	0.36	0.27	0.27	0.19	0.21	0.06	0.15	0.11	0.05	0.13	0.09	0.13	0.07	0.10
SolarDesign	-0.02	-0.03	-0.03	-0.02	-0.03	-0.01	-0.02	-0.04	-0.01	-0.03	-0.03	-0.03	-0.03	-0.04
TotalBuildingRValue	0.00	-0.03	-0.01	-0.01	0.05	0.08	0.12	-0.05	0.11	0.00	-0.07	-0.04	-0.03	0.06
AvgSkinRValue	0.01	0.06	-0.04	-0.05	-0.06	-0.06	-0.10	0.06	-0.07	-0.04	0.05	-0.01	-0.12	-0.17
WallRValue	0.01	-0.03	-0.01	-0.01	0.05	0.08	0.13	-0.06	0.12	-0.01	-0.06	-0.05	0.03	0.04
GlassFrameUValue	-0.06	0.05	-0.02	-0.01	-0.09	-0.11	-0.13	0.06	-0.13	0.00	0.07	0.04	0.05	-0.07
RoofRValue	-0.01	-0.03	-0.01	-0.01	0.05	0.08	0.12	-0.04	0.10	0.00	-0.07	-0.04	-0.05	0.06
TotalBuildingOpaqueRValue	0.00	-0.03	-0.01	-0.01	0.05	0.08	0.12	-0.05	0.11	-0.01	-0.07	-0.04	-0.03	0.06
	Office_1A	**Office_2A**	**Office_2B**	**Office_3A**	**Office_3B**	**Office_3C**	**Office_4A**	**Office_4B**	**Office_4C**	**Office_5A**	**Office_5B**	**Office_6A**	**Office_6B**	**Office_7A**
Height	0.17	0.25	0.16	0.35	0.23	0.15	0.29	0.18	0.28	0.25	0.19	0.36	0.36	0.46
NumberFloors	0.17	0.24	0.14	0.35	0.23	0.15	0.30	0.18	0.29	0.26	0.18	0.36	0.33	0.43
TotalArea	0.20	0.19	0.14	0.35	0.18	0.16	0.27	0.10	0.21	0.12	0.12	0.25	0.24	0.32
WindowWallRatio	0.14	0.11	0.09	0.04	0.05	0.07	0.24	0.18	0.24	0.16	0.08	0.14	0.04	0.18
FloorHeight	-0.01	0.07	0.15	0.03	0.03	0.13	-0.03	0.06	0.08	-0.08	0.03	0.12	0.17	0.14
PlateDepth	-0.12	-0.18	-0.16	-0.08	-0.08	-0.13	-0.27	-0.12	-0.29	-0.16	-0.08	-0.15	-0.08	-0.07
PlateLength	0.18	0.19	0.17	0.07	0.16	0.11	0.35	0.20	0.35	0.14	0.05	0.19	0.08	0.20
SkinArea	0.18	0.25	0.15	0.32	0.23	0.13	0.32	0.20	0.29	0.22	0.17	0.36	0.33	0.48
SkinFloorRatio	-0.05	0.15	0.10	0.13	0.10	0.03	0.09	0.18	0.29	0.13	0.16	0.29	0.21	0.35
GlassArea	0.19	0.26	0.14	0.30	0.21	0.12	0.34	0.23	0.31	0.22	0.16	0.35	0.27	0.46
EnvelopeFloorAreaRatio	-0.06	0.14	0.09	0.10	0.08	0.02	0.08	0.16	0.27	0.12	0.15	0.27	0.20	0.33
EnvelopeQuality	-0.16	-0.05	0.03	-0.23	-0.11	0.00	0.10	0.01	-0.25	-0.28	-0.12	-0.20	-0.25	-0.11

Note: Solid red indicates positive correlation (close to 1), solid blue indicates negative correlation (close to −1), and white indicates no correlation (close to 0).

Table 3. Pearson input variable correlation coefficients to building EUI for building types and climate zones (Hospital omitted for similarity to outpatient) (continued)

	Office_ 1A	**Office_ 2A**	**Office_ 2B**	**Office_ 3A**	**Office_ 3B**	**Office_ 3C**	**Office_ 4A**	**Office_ 4B**	**Office_ 4C**	**Office_ 5A**	**Office_ 5B**	**Office_ 6A**	**Office_ 6B**	**Office_ 7A**
LightingPowerDensity	-0.45	-0.31	-0.32	-0.40	-0.38	-0.43	-0.43	-0.47	-0.40	-0.34	-0.31	-0.24	-0.37	-0.29
SetpointSetting	-0.07	-0.06	-0.11	-0.07	-0.09	-0.05	-0.10	-0.13	-0.09	-0.13	-0.13	-0.12	-0.10	-0.10
HVACSetting	0.16	0.12	0.14	0.11	0.11	0.03	0.14	0.11	0.09	0.17	0.15	0.19	0.14	0.18
SolarDesign	-0.03	-0.02	-0.03	-0.03	-0.03	-0.02	-0.05	-0.09	-0.02	-0.05	-0.05	-0.06	-0.06	-0.06
TotalBuildingRValue	-0.16	-0.03	0.02	-0.24	-0.07	0.00	0.04	-0.01	-0.24	-0.28	-0.12	-0.19	-0.25	-0.09
AvgSkinRValue	0.07	-0.05	-0.06	0.16	0.02	-0.02	-0.23	-0.14	-0.03	0.02	0.01	-0.10	0.08	-0.16
WallRValue	-0.16	-0.03	0.02	-0.24	-0.07	0.00	0.07	0.00	-0.25	-0.30	-0.13	-0.12	-0.24	-0.04
GlassFrameUValue	0.15	0.05	-0.03	0.22	0.12	0.00	-0.09	-0.01	0.25	0.27	0.12	0.21	0.25	0.13
RoofRValue	-0.16	-0.03	0.02	-0.24	-0.07	0.00	0.02	-0.01	-0.24	-0.28	-0.12	-0.21	-0.25	-0.10
TotalBuildingOpaqueRValue	-0.16	-0.03	0.02	-0.24	-0.07	0.00	0.03	-0.01	-0.24	-0.28	-0.12	-0.19	-0.25	-0.09

Note: Solid red indicates positive correlation (close to 1), solid blue indicates negative correlation (close to −1), and white indicates no correlation (close to 0).

4.3 AI SURROGATE MODEL DEVELOPMENT

An AI surrogate model was developed for prediction across the design space. Total building EUI was selected as the target modeling performance metric and involved all input features from the input table consisting of 256,000 rows. Numerical (non-categorical) values were used if multiple variables described the same building property. For example, the actual floor area was used instead of the floor area description (low, medium, high). Any variables that were categorical were converted to a type suitable for ML using *one-hot encoding*. One-hot encoding is the process of extracting each level of a multilevel categorical variable into a separate binary variable to prevent the model from assuming an ordinal relationship between the levels of the variable.

The data were split into 80% training data and 20% testing data. Four different ML models were evaluated in predicting building EUI: multiple linear regression, random forest (Ho 1995), gradient boosting (Friedman 2001), and multilayer perceptron (Haykin 1994). Multiple linear regression is the simplest of these models: it assumes linear relationships between the independent variables and the dependent variables. It is also easy to develop, which makes it a good model to use as a baseline to evaluate the other three, more complex models. Fivefold cross validation with a grid search was used to tune each model's hyperparameters and evaluate each model's performance. Table 4 lists the metrics for each model. Because the random forest model had the best metrics, it was chosen as the AI implementation for this analysis. Then, the best hyperparameter values were selected, and a final model was trained to use for UDSE. These hyperparameters are listed in Table 5.

Table 4. Cross-validation metrics for fivefold, three-repeat cross validation

Model	Mean absolute error in EUI (Btu/ft^2)	R^2
Random forest	2.37	0.9997
Gradient boosting	48.56	0.9349
Multilayer perceptron	7.85	0.9987
Multiple linear regression	54.57	0.9346

Table 5. Random forest hyperparameter values were determined from a grid search over the same fivefold, three-repeat cross validation

Hyperparameter	Value
Max features	Auto
Minimum samples per leaf	1
Minimum samples per split	2
Number of estimators	300

4.4 PROTOTYPE APP

A prototype app was developed that leverages this simulation-trained AI agent for interactive building design. This app can be used by designers and engineers to evaluate how their decisions will affect building energy use. It enables them to conduct detailed, flexible analysis to make critical design decisions without requiring simulation of their project. This capability encourages the design team to consider building performance early in the design process when lack of expertise, time, or money would typically prevent such considerations. This prototype app features single-design visualization and is a fully parametric model with 3D design visualization that allows the user to quickly generate a shoebox model while the surrogate AI predicts energy performance as the user interacts through the design space in real time. A view of the app's 3D model visualization is shown in Figure 6.

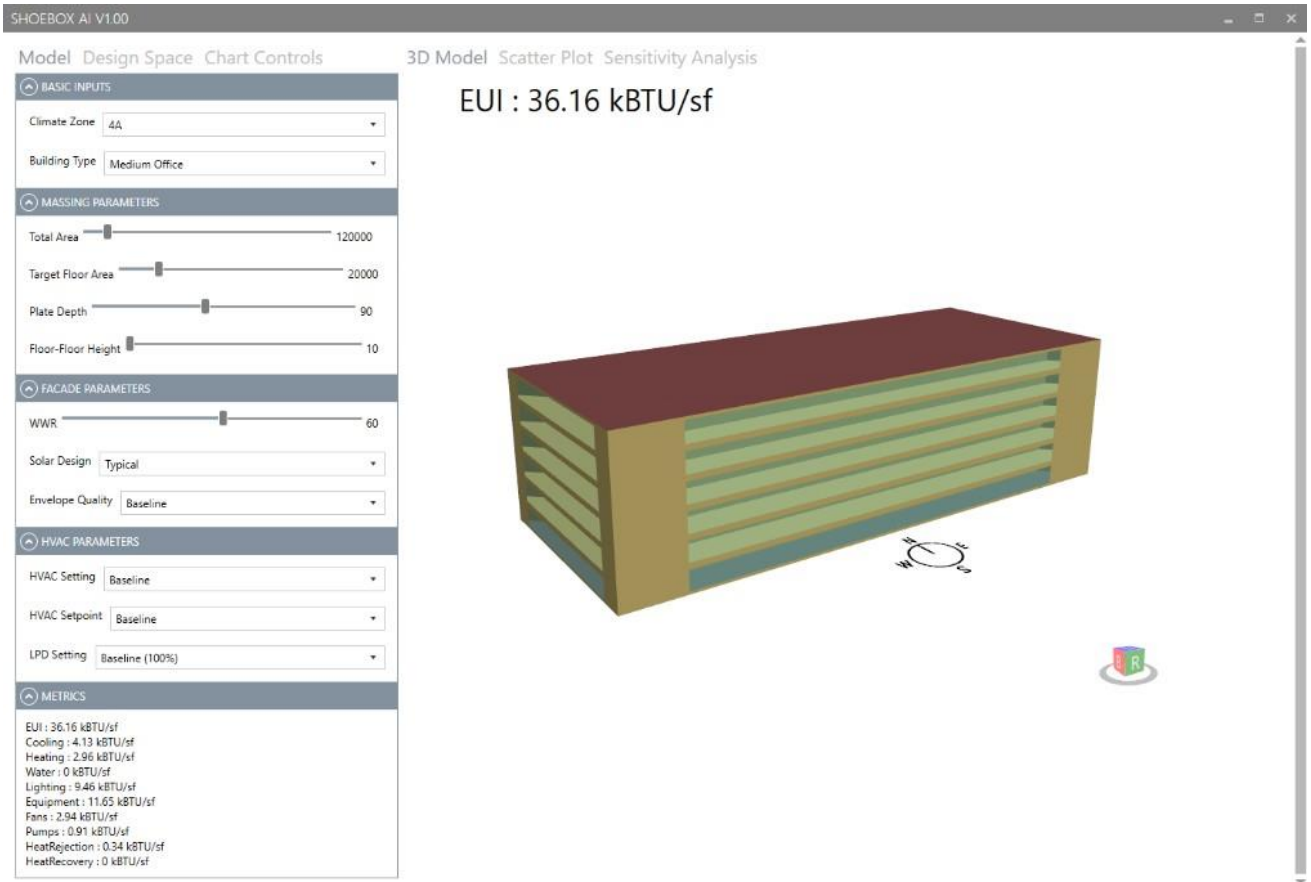

Figure 6. The 3D model visualization within the prototype app.

The prototype app also provides a scatterplot visualization to help the user understand how a single design fits within the broader design space and related energy performance. It also provides the ability to filter each of the design parameters to slice the design space. A sample scatterplot is shown in Figure 7. In addition to scatterplot visualization, the app provides real-time sensitivity analysis. This analysis helps the user clearly determine the sensitivity of EUI to the design features within the user-defined design space. A sample sensitivity analysis is shown in Figure 8.

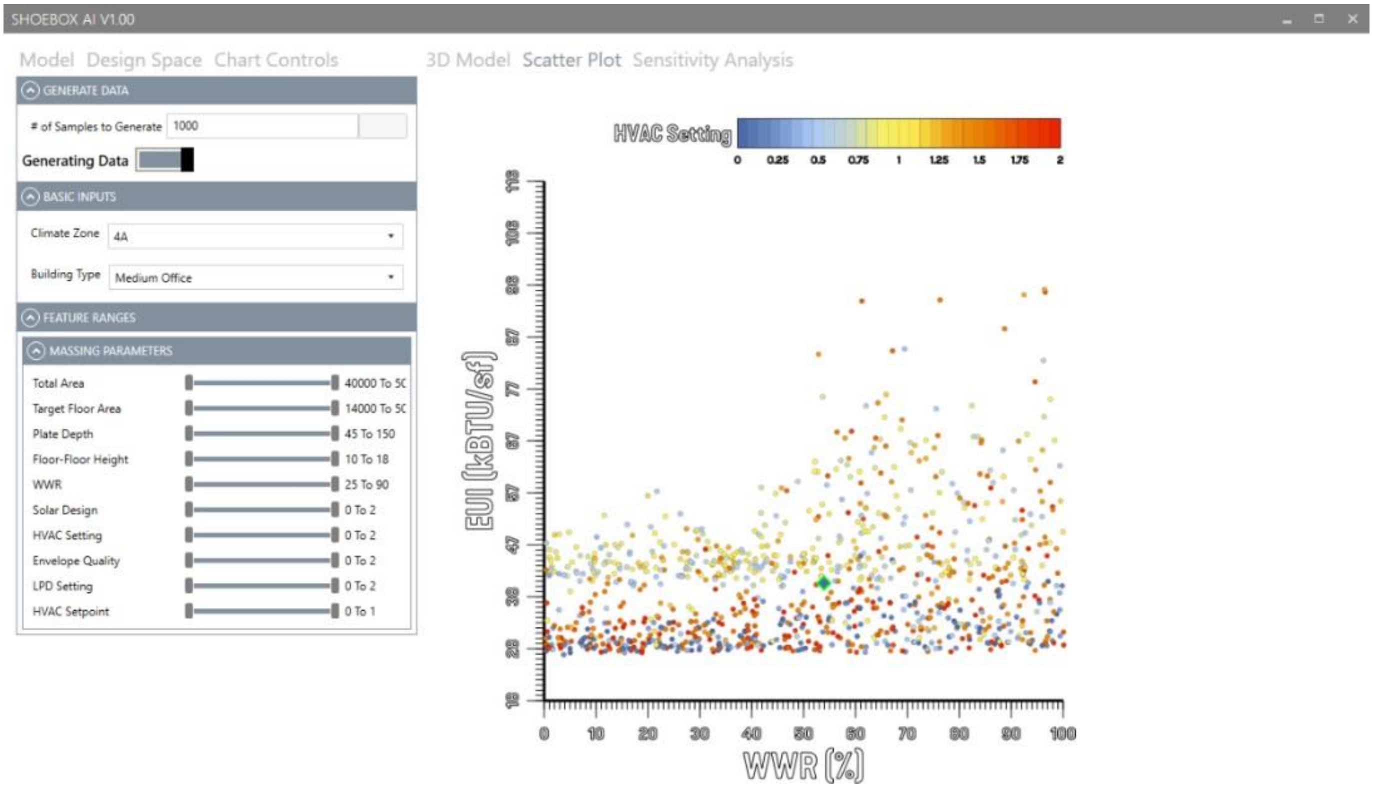

Figure 7. The scatterplot visualization within the prototype app.

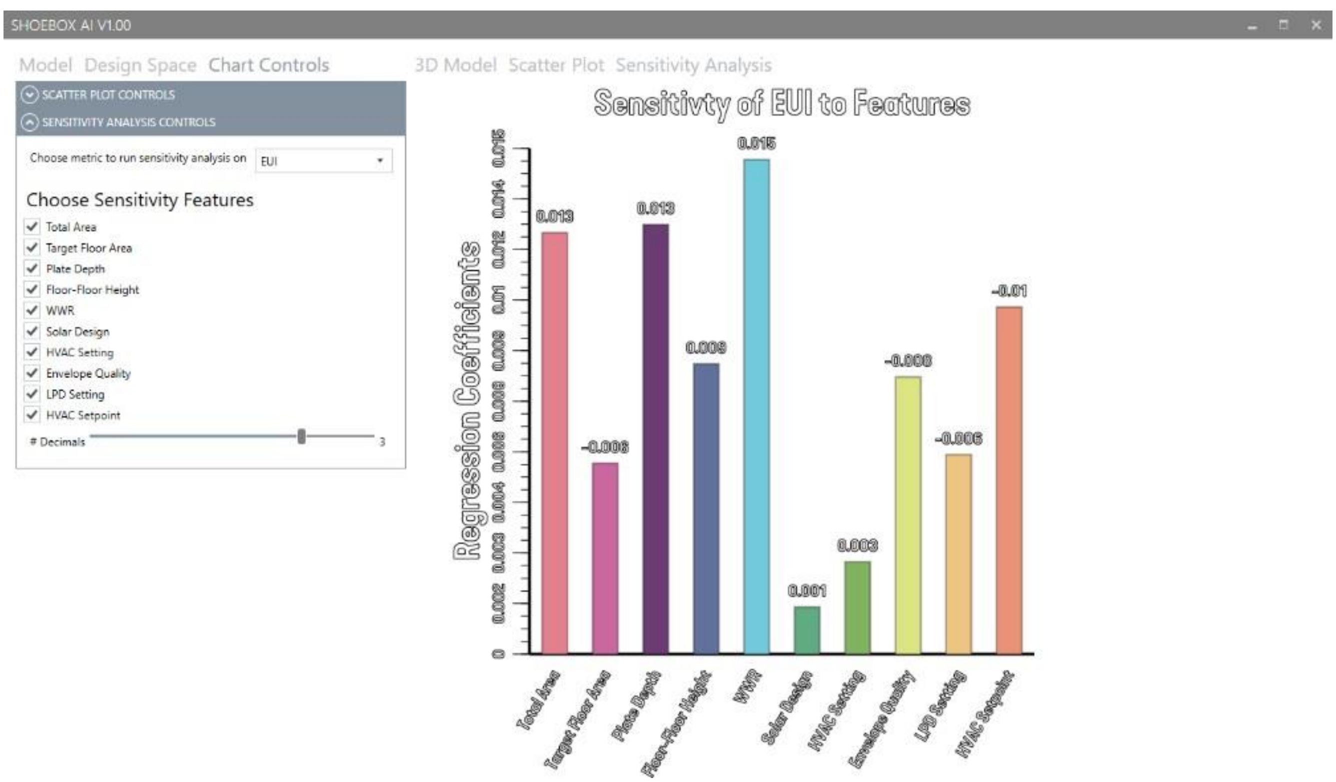

Figure 8. The real-time sensitivity analysis within the prototype app.

4.5 PUBLIC DATA HOSTING

An important part of this project was to make the full parametric design and related simulation data publicly available so they could be used and analyzed by the broader buildings' community. The data were shared to Kaggle, an online community developed by data scientists to find, publish, analyze, and model data (McNally, Bass, and Curtis 2021).

5. REFERENCES

Friedman, J. H. 2001. "Greedy function approximation: a gradient boosting machine." *Annals of Statistics* 29(5): 1189–1232.

Haykin, S. 1994. *Neural Networks: A Comprehensive Foundation*. New York: Macmillan.

Ho, T. 1995. "Random decision forests." *Proceedings of 3rd International Conference on Document Analysis and Recognition* 1: 278–282.

McNally, P., B. Bass, and L. Curtis. 2021 "Universal Design Space Building Energy Simulation." Retrieved from https://www.kaggle.com/petermcnallysg/universal-design-space-building-energy-simulation

New, J. 2021. "Automatic Building Energy Modeling (AutoBEM)." Retrieved from bit.ly/AutoBEM

NREL (National Renewable Energy Laboratory). 2021a. "EnergyPlus." Retrieved from https://energyplus.net/

NREL. 2021b. "OpenStudio." Retrieved from https://openstudio.net/

Santosh, P. 2021. "eppy 0.5.44 documentation." Retrieved from https://pythonhosted.org/eppy/index.htm