Implementing Dynamic Subset Sensitivity Analysis for Early Design Datasets

Laura E. Hinkle*[a], Gregory Pavlak[a], Leland Curtis[b], and Nathan C. Brown[a]

[a] Department of Architectural Engineering, Penn State University, 104 Engineering Unit A, University Park, PA 16802, USA

[b] CRB Group, 1251 NW Briarcliff Pkwy #500, Kansas City, MO 64116, USA

* Corresponding author. Present/Permanent Address: Department of Architectural Engineering, The Pennsylvania State University, 104 Engineering Unit A, University Park, PA 16802, USA.

Email address: luh183@psu.edu

Abstract

Engaging with performance feedback in early building design often involves building a custom parametric model and generating large datasets, which is not always feasible. Alternatively, large parametric datasets of general design problems and filtering methods could be used together to explore specific design decisions. This paper investigates the generalizability of a method that dynamically assesses variable importance and likely influence on performance objectives as a precomputed design space is filtered down. The method first trains linear model trees to predict building performance objectives across a generic design space. Leaf node models are then aggregated to provide feedback on variable importance in different design space regions. This approach is tested on three design problems that vary in number of variables, samples, and design space structure to reveal advantages and potential limitations of the method. Algorithm improvements are proposed, and general recommendations are developed to apply it on future datasets.

Keywords

Parametric design, conceptual design, sensitivity analysis, design variable importance, surrogate model, decision tree

1. Introduction

With the integration of simulation engines into visual programming environments, parametric modeling techniques can be easily paired with simulation data to provide performance feedback during design. This approach allows designers to quickly evaluate many potential design configurations. In practice, it is not feasible to consider every design in the parametric design space, but several methods have been developed to navigate the design space efficiently. While some methods directly point the designer towards optimal performance, including automated optimization [1]–[3] and interactive optimization [4]–

[7] workflows, others intend to more gently guide the designer towards better performing designs, offering increased flexibility and opportunities for designer preference expression. Such methods include design catalogs [8]–[10], surrogate-model-based workflows that enable live manipulation [11], [12], and performance maps [13]. The latter methods can be most useful in the earliest stages when many aspects of the design are flexible [14], there are competing objectives that need to be synthesized [15], or designers have mixed quantitative and qualitative criteria [16]. In particular, surrogate modeling can be used to facilitate discussions as changes are made [17] and is accessible with modern statistical tools and libraries. However, building custom parametric models and running simulations to generate data is time-consuming, and further adjustments may be required throughout early design, requiring more effort to update the surrogate model. Design practice moves quickly, and tools get left behind if they do not provide salient information at crucial points when designers really need them. Even with newly available tools, there remains a need for responsive and accessible performance feedback from parametric design spaces.

In this vein, designers might prefer to use a general parametric model to determine which design aspects or variables tend to influence the performance before modifying the design outside a restrictive parametric framework. The general parametric model must contain many variables and configurations but have the ability to be filtered down to provide useful feedback on a specific design problem. As the design space is filtered to reflect project-specific criteria, designers can quickly discover which variables are more likely to improve performance metrics and where "good" settings tend to be for their problem. The process of determining which variables matter is a type of sensitivity analysis. Sensitivity analysis has been used for a range of building design problems, from model calibration [18] to setting up a design optimization problem [19]. While there are many existing sensitivity analysis methods appropriate for building design problems, few are suited for real-time analysis. As the general parametric model is filtered, existing sensitivity analysis methods require re-running the analysis each time, which is disruptive to the design process.

One approach to allow for real-time sensitivity analysis is to split the general parametric model design space into many regional models to be accessed during filtering. Existing regional sensitivity analysis methods have been used to develop useful qualitative feedback but encountered low accuracy in certain regions and lacked intuitive visualizations for designers [20]. Depending on the sampling technique, many regions or subsets may lack data necessary to describe the behavior [21]. For the general parametric model to be truly flexible, it must have the ability to be filtered on any design criteria and provide sensitivity analysis of sufficient accuracy for early design. With regional models, the designer can gain intuition on how variable behavior changes in each region prior to filtering to inform the initial design. However, a new method is required to provide this information along with real-time subset sensitivity analysis.

In response, this paper extends and rigorously investigates a new method called dynamic subset sensitivity analysis [22]. The method divides a general design space into many models using a decision-tree-like training process and provides real-time variable sensitivity through interpolation techniques. This paper considers the generalizability of the method by applying it to three building design problems of different domains and scales. A comparison of the three datasets shows when the method has enough data to be successful, along which what issues may arise when trying to apply the method to future parametric datasets. By presenting the analyses side-by-side, it also demonstrates how a designer might engage with multiple objectives simultaneously or iteratively as they move between decision variables and scales. Through this work, modifications to algorithm are proposed to communicate variable behavior more accurately in certain regions of the design space, particularly when the response is nonlinear. The value of the method is evaluated for each building design problem. Finally, a set of recommendations are

developed to implement the method on future datasets. The goal is to promote adoption of performance-driven parametric tools in early design, leading to more sustainable buildings.

2. Literature review

2.1 Rapid feedback in early design

Parametric modeling and design space exploration are increasingly used in early design. Researchers have been attempting to improve such design approaches through design catalogs [8], interactive and automated explorations [5], [23], and visualization techniques [24]. One of the main considerations in the development of these methods is computational time, specifically during active design exploration. General research into computation tasks shows that an interruption of more than 400ms seconds reduces productivity [25]. Building upon this finding, [26] established the roll theory, which states that "when an individual has access to the data necessary to perform the creative task at hand, when concentration is not broken by distractions, and when the individual has developed a consistent method of organizing the data, then ideas and solutions will suggest more ideas and solutions to successive steps of the creative process, in a rapid and orderly flow." Roll theory is related to the concept of creative flow [27], which has been considered while creating tools for rapid design assessment [28]. To achieve this flow, researchers have identified and tested surrogate models that approximate performance during design exploration and reduce lag [29]. Designers can explore the design space and receive rapid feedback, facilitating team discussions [30] and guiding sustainable design decisions.

While non-parametric, black-box surrogate models often achieve the highest accuracy, many researchers have implemented interpretable surrogate models with sufficient accuracy [31], [32]. Localized models such as decision-trees and piecewise models can provide granular variable sensitivity in addition to performance feedback, making them doubly advantageous if they can reach acceptable accuracy. The linear model tree utilized in this paper is an extension of the decision-tree and has been implemented in other domains such as computational fluid mechanics [33], data mining [34], and human computer interaction research [35]. The proposed method leverages the local models yielded from the linear model tree to provide real-time sensitivity analysis in early building design scenarios.

2.2 Reusable design spaces

Despite their potential benefits, many recent interactive design methods have not been widely implemented in practice due to practical considerations [36]. Building a model from scratch and running simulations is time-consuming depending on the response variable. Many researchers have shifted focus to understanding *when* and *how* building data and prediction models can be transferred from decision to decision and project to project. The idea of reusable surrogate models for engineering design is introduced in [37]. It proposes graph-based surrogate models for trusses and demonstrates its effectiveness in new design spaces via transfer learning. Several transfer learning approaches have also been proposed for building energy prediction and control [38], [39]. However, these approaches are in the early stages of development and are not yet widely used in industry. Rather than transferring data or models, another approach that is appropriate for early building design is to create a general design space that can be customized or adapted for many design problems [11], [40]. While it takes domain expertise to define a design space that balances specificity with generalizability to many projects, many design firms work repeatedly in certain geographic areas or building sectors, making this possibility feasible [41]. There are also domain-specific ways to reuse machine learning (ML) data for predicting the performance of new designs. For example, by hybridizing data modeling with physics-based modeling and/or using ML to predict the behavior of a single unit that can be aggregated to rapidly predict the performance of a full structure [42]. However, this paper focuses on the use of parametric datasets in early design.

2.3 Sensitivity analysis for building design problems

Sensitivity analysis has been widely implemented in building design problems to inform the decision-making process. It has been incorporated into model calibration procedures [18], formulating an optimization problem [19], and decision-making in design or operation [43], [44]. However, it has not yet been applied to generalizable parametric design datasets. Sensitivity analysis allocates the uncertainty in the response among the predictor variables and can be used to gauge variable importance, as well as understand variable interactions [45]. It is particularly useful in the early design stages when the designer is trying to discover which variables tend to influence the response and by how much, whether the question is related to daylight, structures, energy, acoustics, or another response variable. This process can help identify critical decisions, as well as more flexible decisions, from the onset.

There are many established methods available to perform sensitivity analysis, both with and without an accompanying regression model. Most of the widely used standalone methods are one-at-a-time (OAT), which have local and global variations that quantify the effect of each variable individually. OAT sensitivity analysis has been used to address a wide range of building design problems, ranging from improving building life cycle assessment [46] to thermal comfort [47]. Many researchers have also leveraged regression models (or surrogate models) to produce variable importance. Specifically, standardized linear regression model coefficients [48] and variable selection procedures such as stepwise regression [49] have been implemented. The main drawback of linear regression is the linearity condition, which may not be satisfied depending on the data. However, some machine learning models have their own importance metrics, such as decision trees. For example, [50] utilized the decision tree importance metric to identify which variables are most important in predicting building energy consumption patterns. Yet, the output of many machine learning models is not directly interpretable or useful to designers [51]. Finally, variance-based approaches have also been used to quantify variable importance for building systems [52]. These methods tend to achieve higher accuracy but require a large number of samples.

The methods described above compute variable importance over the entire variable domain. As the design space is refined or filtered during early design, the initial sensitivity analysis may no longer be accurate, so the calculations must be re-run from scratch. One researcher approached this issue by retraining the underlying regression model on the restricted variable domain [53]. However, depending on how the domain was restricted, predictions were not consistently accurate. Another study leveraged Monte Carlo filtering and Regional Sensitivity Analysis (RSA) [20], but also encountered low accuracy in certain regions, and did not use detailed building performance simulation software to generate data, leading to further potential inaccuracies. Nevertheless, filtering is a valuable design space exploration technique as reusable parametric models emerge as a new research area.

2.4 Data visualization for design space exploration

Making sensitivity analysis valuable for early design also requires careful consideration of how a user might engage with the data. Building design problems are often high dimensional and thus difficult to visualize. One of the most common methods in building design is parallel and radial coordinate plots [54]. Some researchers have proposed performance maps [13] or self-organizing maps [55], [56] to preserve multivariate information and convey it to designers. Others have argued that reducing the number of variables through principal component analysis or latent space [57] can guide designers towards high-performing designs more quickly. Regardless, the manner in which the results are communicated is equally important as the underlying model [58].

2.5 Research gaps and contributions

In summary, to make use of general models in early design, a new method is required that quickly and accurately updates variable importance as the design space is refined and yields results that are easily interpretable. Although dynamic subset sensitivity analysis was initially proposed in [22] on a single dataset, the method has not yet been rigorously tested. There are many data model issues that may arise when feeding in certain datasets, such as discontinuous spaces, collinearity, a lack of significance for certain regions, or even just not having enough data to make a quality assessment of importance. In this paper, we investigate the generalizability of dynamic subset sensitivity analysis by testing it on three datasets from different domains and scales. The three datasets are based on spatial daylight autonomy of a sidelit room, energy use intensity of a residential retrofit, and embodied carbon of a tall timber structure. These design problems were selected because their datasets differ in domain and scale, but also data type, linearity, number of variables, and number of samples. They are also similar in structure to common datasets being implemented in ML-based design tools by leading firms in AEC [41], to the extent that these structures are commonly known. Based on the implementation for these three datasets, we are able to derive a set of recommendations for the method to be implemented on future datasets and propose improvements to the algorithm.

3. Methodology

The overall procedure is described in Figure 1. First, three general design problems were identified, and corresponding datasets were generated or obtained, and then processed in preparation for training. The linear model trees were then trained, in addition to a simple linear regression model and traditional decision tree model for comparison. Next, the average variable sensitivity was calculated in small bins to understand where in the variable domain certain variables tended to have a large influence on the response while accounting for other variables in the model. Finally, the dynamic subset sensitivity analysis was demonstrated through a few design scenarios. The quality of the leaf node models was evaluated through coefficient p-values, and modifications to the dynamic subset sensitivity analysis algorithm were implemented. Lastly, a set of recommendations was proposed for applying this method to future datasets.

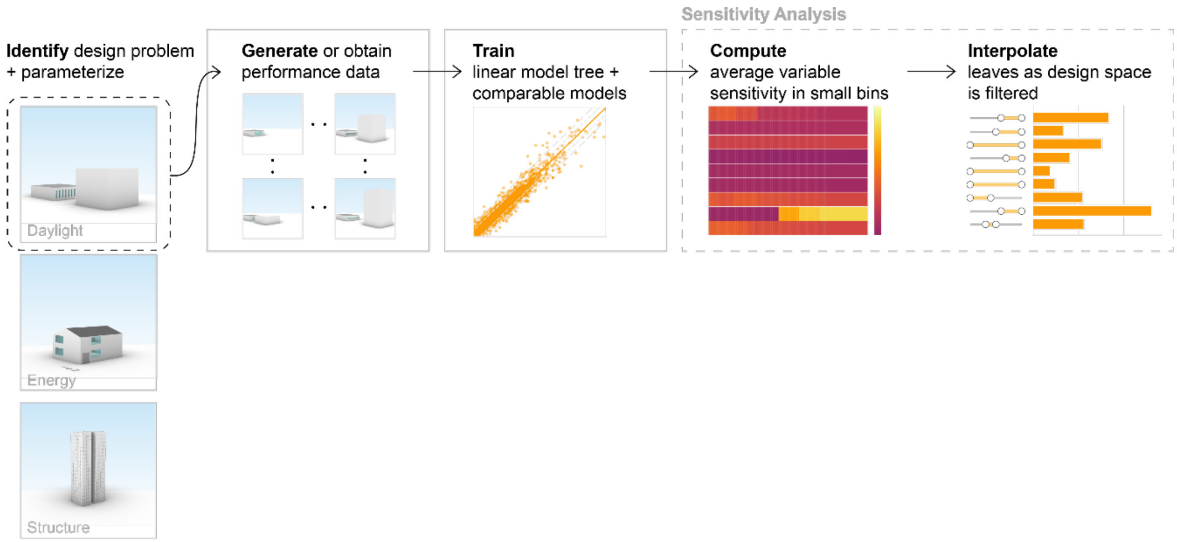

Figure 1: Overall methodology with three datasets

3.1 Problem selection

One of the goals of the proposed method was to customize a large, general dataset throughout the early design stage and across many building projects. To this purpose, three datasets were generated or selected to represent general design problems from the domains of daylighting, energy, and structure (Fig. 2).

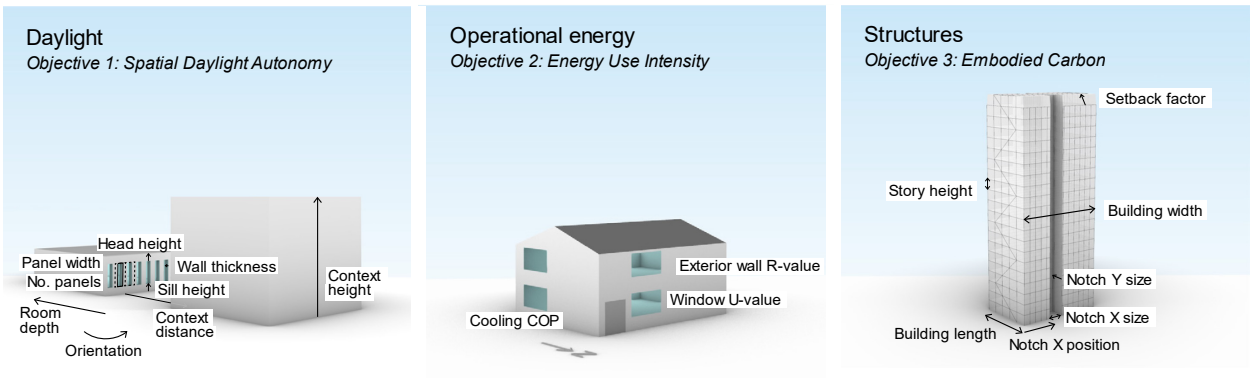

Figure 2: A visualization of the geometry for the daylight, energy, and structure design spaces

3.2 Data generation and processing

Three datasets were generated or obtained from the three design spaces described in Section 3.1. The following subsections provide details on data generation and processing for each dataset, and a summary of the variables and responses are provided in Table 1.

Table 1: Datasets summary

Dataset	Variables	Response
Daylight	Room depth, sill height, head height, orientation, context distance, context height, number of panels, panel width, wall thickness	Spatial daylight autonomy
Operational energy	Cooling COP, R-value, U-value	Energy use intensity
Structures	Building width, building length, story height, setback, notch X position, notch X size, notch Y size	Embodied carbon

3.2.1 Daylighting Model and Dataset

A sidelit room model was developed to represent the domain of daylighting. In building practice in the United States, daylight simulations are often required to obtain LEED v4 Daylight credits [59]. Therefore, this model could be useful across many spaces and projects. It is assumed that a designer would consult the model repeatedly for a single project as they establish the layout of rooms and the façade. First, the daylit room was modeled parametrically in Grasshopper to include nine variables: room depth, sill height, head height, orientation, context distance, context height, number of panels, panel width, and wall thickness (Figure 2). All room surfaces accord with LM-83 guidelines [60]. The windows were typical double-pane low-e with 61% visible transmittance and incorporated an automated shade. The shade fabric had 7.2% visible transmittance and 6.6% permeability in accordance with LM-83. Room width and room height were 9m and 3m, respectively, although they could be incorporated as variables in the future. The

variable bounds are provided in Table 2. They were set to provide enough flexibility for repeated use, but still abide by modern construction standards.

Spatial daylight autonomy (sDA) at 300 lux was the response variable, or "objective" in design space terms, generated using ClimateStudio in Grasshopper. To ensure enough samples for the regression tree, 12,500 points were sampled using Latin Hypercube sampling. The simulations were conducted in Pittsburgh, PA, USA, which is often overcast and at a 40.44° N latitude. For future datasets, sky condition and latitude could be included to make the design space more flexible, but these parameters were set to demonstrate the method. While designers might in different cases design to the typical, worst-case, or average annual behavior, these assumptions would be applicable when making a reusable dataset for buildings across a given city. The sensors were spaced at 1m and the workplane was positioned 0.762m above floor finish. Within the path-tracing settings, the number of rays emitted for each sensor at each pass was 500. The Radiance parameters considered up to 6 ambient bounces before discarding a ray. The dataset was split 80/20 for training and testing, and all predictor variables were scaled from 0-1 to ensure importance was not influenced by the variables' scale.

Table 2: Variables in spatial daylight autonomy dataset

Variable	Minimum	Maximum
Room depth (m)	6.00	15.00
Sill height (m)	0.10	1.10
Head height (m)	0.10	1.10
Orientation (deg from south)	0.00	360.00
Context distance (m)	3.00	15.00
Context height (m)	0.00	15.00
Number of panels	1	20
Panel width (relative)	0.10	0.90
Wall thickness (m)	0.20	1.00

3.2.2 Energy Model and Dataset

The second dataset was based on a residential energy retrofit scenario. This dataset represents a reusable model for within a city when testing upgrades on similar residential stock. However, the model would have to be customized based on the feasible ranges of variables to consider in each individual case. An EnergyPlus model was constructed to represent a residential home considering upgrades on the cooling COP, exterior wall insulation, and window construction. Specifically, cooling COP, R-value, and U-value were included as variables (Figure 2). The generic home was 331.23 m^2 and assumed to contain a DX cooling coil and an electric heating coil. The settings for each variable are provided in Table 3. U-value was not controlled directly, as it typically varies with other window properties. Instead, 19 window constructions were selected and used to generate data. The U-value and solar heat gain coefficient (SHGC) were extracted during data processing to represent the window constructions in the dataset. However, because U-value and solar heat gain were highly correlated, only U-value was incorporated into the linear model tree to prevent collinearity issues (Figure 7). Previous studies have also shown a correlation between U-value and SHGC among existing window constructions [61], [62]. The R-values were converted to conductivity in the exterior wall material in EnergyPlus, and the cooling COP was accessed directly in EnergyPlus. All 6,859 permutations were simulated in Altoona, Pennsylvania, USA. The total site energy per conditioned building area was the response. Although grid sampling is not recommended for the proposed method (see limitations section), simulating 19 settings for each variable

yielded high-resolution data sufficient for sensitivity analysis. The dataset was split 80/20 for training and testing, and all predictor variables were scaled from 0-1.

Table 3: Variable options for energy dataset

Variable	Options
Cooling COP	1.2, 1.4, 1.6, 1.8, 2.0, 2.2, 2.4, 2.6, 2.8, 3.0, 3.2, 3.4, 3.6, 3.8 4.0, 4.2, 4.4, 4.6, 4.8
R-value (ft^2-F-h/BTU)	12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48
U-value (W/m^2-K)	0.785, 0.992, 1.062, 1.265, 1.525, 1.624, 1.704, 1.71, 1.765, 1.772, 2.143, 2.255, 2.556, 2.72, 2.765, 3.122, 3.835, 4.513, 5.894

3.2.3 Structural Model and Dataset

The third dataset used to demonstrate the proposed method was an embodied carbon dataset initially generated by Hens et al. [63] and used to explore performance prediction for interactive parametric design in Zargar & Brown [64]. The dataset includes a wide variety of geometric configurations for a mass timber building with a post-beam-panel gravity system and a lateral system incorporating linear elements. For each geometry, a custom sizer based on timber design codes sizes each element based on applicable structural loads and fire protection criteria. Embodied carbon coefficients are then used to convert the building elements into carbon emissions equivalent values, assuming no carbon storage. The embodied carbon contributions of the elements are then summed to predict the overall embodied carbon of the entire structural system. Hens et al. [63] and Hens et al. [65] describe the methodology used to generate the dataset in more detail. In this paper, we incorporated the independent and several partially dependent variables, including building width, building length, story height, setback, notch x position, notch x size, and notch y size into the linear model tree (Figure 2). The response was embodied carbon. Because notch x position, notch x size, notch y size, and setback depend on the more fundamental variables of width and length, the linear correlations were calculated to diagnose collinearity issues before training the linear model tree (Fig. 7). However, all Pearson correlation coefficients were within the acceptable range and thus incorporated into the model. Outliers were eliminated by the interquartile range (IQR) method, which resulted in 940 data points. The variable bounds are provided in Table 4. The dataset was split 80/20 for training and testing, and all predictor variables were scaled from 0-1.

Table 4: Variables in embodied carbon dataset

Variable	Minimum	Maximum
Building width (normalized)	0.0005	0.9995
Building length (normalized)	0.0005	0.9995
Story height (m)	3.048	4.876
Setback (relative)	0.005	9.995
Notch X position (relative)	0.0005	0.9995
Notch X size (relative)	0.0005	0.9995
Notch Y size (relative)	0.00045	0.89955

3.3 Training the linear model trees

After preparing the datasets, the first step is to create regression trees that can eventually be used for sensitivity analysis and filtering. Figure 3 is a representation of a one-dimensional linear model tree, but a similar procedure follows for high dimensional spaces. The trees are built through recursive binary

splitting, where predictor X_j is split at cutpoint s such that splitting the predictor space into the regions $\{X \mid X_j < s\}$ and $\{X \mid X_j \geq s\}$ leads to the greatest reduction in the residual sum of squares (RSS). Splitting stops based on some threshold and each terminal node, or leaf (Figure 3), contains a model that applies in the j-th region only. For traditional regression trees, the estimated response $\hat{y}_{R_j}$ is the mean response for the training observations in the j-th region. However, this is often an over-simplification of the true relationships. To address this issue, linear model trees use a linear model to estimate the response. By the end of the training process, each leaf node contains its own linear model.

$$RSS = \sum_{j=1}^{J} \sum_{i \in R_j} (y_i - \hat{y}_{R_j})^2 , \quad \text{(Equation 1)}$$

In Equation 1, the outer summation accounts for each variable and the inner summation accounts for all points in the specified region. While previous studies have achieved high accuracy with nonparametric models, it is often not possible to make inferences and inform the building design process. It was hypothesized that linear model trees could achieve sufficient accuracy for early design while allowing for dynamic interpretations about variable sensitivity because of how they are constructed. The correctness of this hypothesis is tested by comparing the results across the varying datasets.

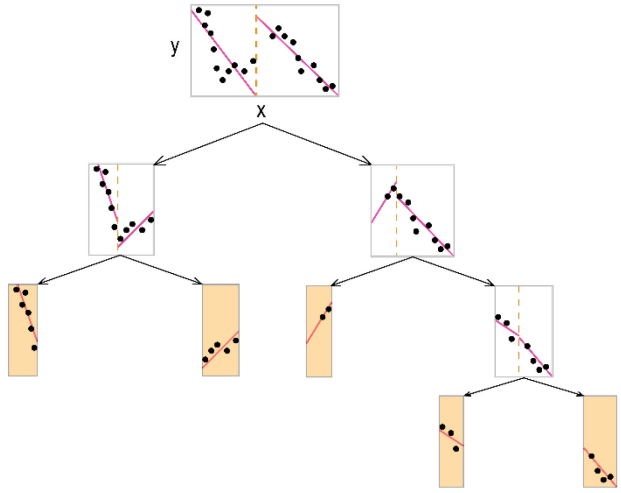

Figure 3: Linear model tree with leaf nodes in orange, after [66]

The termination criteria for a linear model tree are the maximum depth and minimum number of samples per leaf, which have to be tuned for a given dataset. For all models, the maximum depth was set to 8 and minimum number of samples per leaf was set to 30. If there are 30 samples, the distribution is considered normal based on the Central Limit Theorem from statistics. The model achieved sufficient accuracy at this depth and enforcing at least 30 points per leaf ensured the model was valid. The maximum depth of 8 was selected to control training time while ensuring enough leaf nodes for interpolation. Once the linear model tree was built, the leaves were used to compute average sensitivity in small bins.

3.4 Calculating average sensitivity over the variable domain in a multi-dimensional design space

The next step is to determine how coefficients of individual leaves should be combined to indicate local variable importance. To get a sense of sensitivity over the entire variables' domain, the average linear model coefficient was computed in small bins. The domain of each variable X_j is partitioned into 100 bins of equal length. The m-th bin is denoted by $b_m := \left[\frac{m-1}{100}, \frac{m}{100}\right)$, for $1 \leq m \leq 100$. The k-th leaf is denoted

by ℓ_k and the number of samples in ℓ_k is n_k. Then, the domain of each variable X_j is constrained by $c_{j,k} \leq X_{j,k} \leq d_{j,k}$ in leaf ℓ_k. Let $\theta_{j,k}$ be the original coefficient of $X_{j,k}$ in ℓ_k. Then the weighted coefficient restricted to bin $b_{j,m}$ is shown by $\hat{\theta}_{j,k,m}$ and is given by the following formula:

$$\hat{\theta}_{j,k,m} = \theta_{j,k} * \frac{n_k}{n(b_{j,m})} * \mathbb{I}(p - value_{\ j,k} \leq 0.05), \text{ (Equation 2)}$$

where $n(b_{j,m})$ is the number of samples in the leaves that overlap $b_{j,m}$ for X_j and $\mathbb{I}(q) = \begin{cases} 1 \; if \; q \equiv True \\ 0 \; if \; q \equiv False \end{cases}$ which is normally denoted as an indicator function. This dictates that if the hypothesis test that determines if the variable linearly affects the response fails, the coefficient is forced to zero to prevent inaccuracies in the averaging equations. Additionally, there must be at least one sample per bin. Figure 4 is a simple example to show the parts of the weighted coefficient equation.

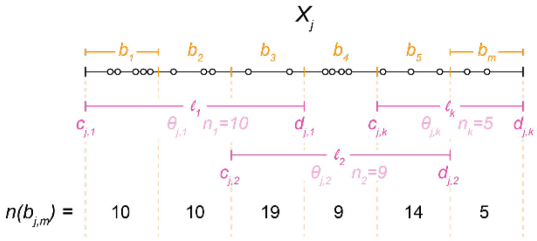

Figure 4: Weighting process in the averaging scheme

Finally, the weighted coefficient for variable X_j in b_m is given by:

$$\hat{\theta}_{j,m} = \sum_k \hat{\theta}_{j,k,m} \text{ (Equation 3)}$$

The result is a local sensitivity analysis over the entire domain that can be used to understand changes in the response. Next, the model leaves are used to update variable importance for user-defined intervals.

3.5 Real-time variable sensitivity via leaf model interpretation

While many machine learning methods can return importance metrics, they are often established through training, requiring retraining if the variables and their corresponding bounds are modified. By precomputing linear models in regions determined by the regression tree, the model coefficients can be interpolated to quickly return variable information without full model retraining. If the user-defined intervals correspond exactly to a pre-defined region, variable sensitivity is provided by that model. Otherwise, the model coefficients must be interpolated based on the "agreement" between the user-defined intervals and the variable domains in the leaves. The agreement of the user restricted intervals with the constraints of ℓ_k is given by:

$$\widetilde{w}_k = \left(\sum_{j=1}^{J} w_{k,j}^{\frac{1}{p}}\right)^p, \text{ (Equation 4)}$$

where $w_{k,j}$ is the amount of "agreement" of X_j in ℓ_k and $p > 1$ is a hyperparameter. Let $[a_j, b_j]$ be the user-defined interval on X_j. Then, the amount of agreement $w_{k,j}$ is defined as:

$$w_{k,j} = \frac{\min\{d_{j,k}, b_j\} - \max\{c_{j,k}, a_j\}}{b_j - a_j} \text{ (Equation 5)}$$

where a, b, c, and d are non-negative values. Without loss of generality, assume $\widetilde{w}_1, \widetilde{w}_2, \ldots, \widetilde{w}_t$ are the top t agreements. The total weight w_k is a function of top t agreements normalized by their sum:

$$w_k = \frac{\widetilde{w}_k}{\sum_{k=1}^{t} \widetilde{w}_k} \text{ (Equation 6)}$$

Finally, variable importance was computed using the following formula:

$$\widehat{\boldsymbol{\theta}} = \sum_{k=1}^{t} w_k \cdot abs(\boldsymbol{\theta}_k \odot \mathfrak{I}(\boldsymbol{\theta}_k)), \text{ (Equation 7)}$$

where $\boldsymbol{\theta}_k$ is the linear model coefficients at ℓ_k, $abs(\cdot)$ is element-wise absolute value of a vector, $\mathfrak{I}(\cdot)$ is element-wise $\mathbb{I}(\cdot)$ of a vector, and $\odot$ is element-wise multiplication of vectors. The procedure is presented in Algorithm 1. Note that p and t are hyperparameters that can be tuned based on the dataset. For all datasets, p and t were set to 3 and 10, respectively. For higher values of p, the contrast between the top t agreements becomes sharper. As t approaches the total number of leaves, the impact of individual leaves gets lost due to normalization. On the other hand, if $t = 1$, only one leaf is used, which might not be an accurate model of the user-defined region. Once the intervals are specified, individual predictions are made with the linear model tree itself. Single designs only fall into one leaf since the regions do not overlap. The prediction is made by the linear model in the appropriate leaf. Once this model has been established, a metric for overall variable importance and visualizations of how performance changes with variable setting modifications can both be returned to a designer without the added time of model retraining. The results section first presents the dataset itself before showing these potential visualizations for the designer.

3.6 Ensuring model significance

The algorithm mentioned above proposed an improvement to eliminate the possibility of poor linear models in the leaf nodes affecting the interpolation calculations. While this issue did not necessarily arise for the daylight dataset in [67], it is an important consideration, as some building datasets contain highly nonlinear variables that cannot be handled during the training process due to a lack of data. The improvement consists of checking the coefficient p-values in each leaf node linear model, and if the p-value is greater than the desired level of significance (in this paper, 5%), the coefficient is forced to zero in the interpolation calculations (Step 10 in Algorithm 1). If the p-value is low, we can reject the null hypothesis, which is that the coefficient is equal to zero, therefore there is evidence that the coefficient is statistically different than zero. However, if the p-value is high, there is no evidence that the coefficient is different from zero and we cannot reject the null hypothesis. In this case, the coefficient is forced to zero instead of ignored because ignoring it would eliminate information from the region and bias the interpolation towards the other models that may or may not fully cover the region. The pseudocode for the updated interpolation algorithm is provided below in Algorithm 1.

Algorithm 1: Leaf node interpolation

```
Input: Linear model tree, user-defined intervals, and hyperparameters p and t
For every leaf ℓ_k:
    For every variable j:
        Compute amount of agreement w_{k,j} according to Eqn 5
    Compute agreement w̃_k per Eqn 4
Pick top t leaves with the highest agreement w̃_k. Let these leaves be ℓ_1', ..., ℓ_t'.
Compute the normalized total weight w_k according to Eqn 6
Initialize updated coefficients θ̂ by a vector of zeros // dimension is the number of variables
Iterate through all top t leaves (Chosen in Step 5) and do the following:
```

9	Let current leaf have index $k' \in \{1', \cdots, t'\}$
10	Update the coefficient in $\widehat{\boldsymbol{\theta}}_{k'}$ by setting all the coefficients that have a *p-value* > 0.05 to zero // this describes $\boldsymbol{\theta}_k \odot \mathfrak{I}(\boldsymbol{\theta}_k)$ in Eqn 7
11	Take the absolute value of the updated coefficients and multiply by the total weight w_k
12	Replace $\widehat{\boldsymbol{\theta}}$ by $\widehat{\boldsymbol{\theta}} + \widehat{\boldsymbol{\theta}}_{k'}$ // output of Step 11
13	Return $\widehat{\boldsymbol{\theta}}$

4. Results

This section first presents linear model tree characteristics for each dataset before the results of the linear model tree interpolation procedures (Table 5). The daylight dataset produced the highest number of leaf nodes, followed by energy and structures. The training criteria enforced 30 samples in each leaf node and maximum depth of 8, but the number of samples per leaf dictated the number of leaves for the energy and structures datasets. For the daylight dataset, the number of panels and wall thickness were split the most, followed by orientation and panel width. Although orientation was split frequently, the results in the following sections show that the slopes were small; therefore, orientation was not important in most regions. Similarly, the cooling COP and R-value were split a comparable number of times, but the cooling COP has large slopes in some regions, and the R-value does not. Finally, building width was split the most for the structures dataset, followed by building length and notch Y size, which largely corresponds with the importance results in the following sections.

Table 5: Linear model tree characteristics

	Daylight	Energy	Structures
Number of leaf nodes	144	58	18
Number of splits	Number of panels: 32 Wall thickness: 29 Orientation: 26 Panel width: 18 Context height: 13 Room depth: 12 Context distance: 5 Head height: 5 Sill height: 3	U-value: 30 Cooling COP: 14 R-value: 13	Building width: 7 Building length: 3 Notch Y size: 3 Story height: 1 Setback: 1 Notch X position: 1 Notch X size: 1

Figure 5 shows a set of designs across the design space to present the range of possible designs for each domain. The daylight design options face south and assume no context building. Notably, the objectives for the daylight and structural design spaces have a visual component, while the energy objective, EUI, does not.

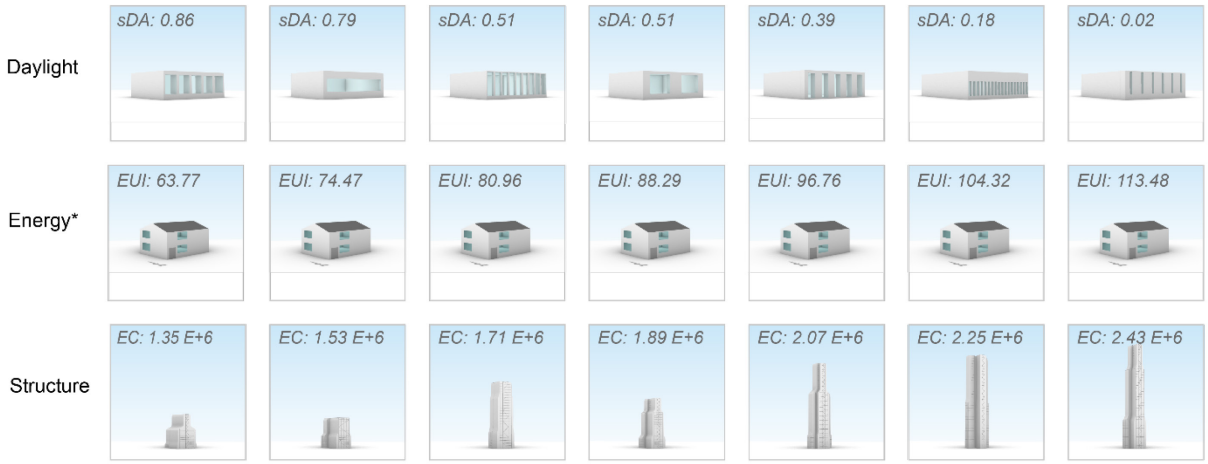

Figure 5: Range of possible design for each dataset

* The energy objective does not have a visual component, as the variables are on the material-level.

4.1 Assessing model fit

The linear model tree fit was then assessed prior to performing calculations with the leaf node model coefficients to ensure the base model was reliable. For each data point in the testing dataset, the appropriate linear model makes the prediction as determined by the linear model tree. Two parametric models were trained to provide a baseline for model performance: a multiple linear regression model and a decision tree model. Figure 6 shows the actual (simulated) response versus the predicted response for each model for the test data. For the spatial daylight autonomy dataset, the multiple linear regression model and decision tree make accurate predictions for low sDA values. However, Figure 6 shows that the linear model tree captures some nonlinear behavior in the model and makes accurate predictions, even for higher values of sDA.

The linear regression model for EUI predictions mostly falls within +/- 5 kWh/m^2 absolute error, which is sufficient for early building design. However, given the nature of the grid-sampled data, the decision tree predicts the response with even higher accuracy. The linear model tree improves upon the decision tree by fitting a linear model in each region instead of simply averaging the data. This results in a very accurate model with high interpretability. However, the linear regression model does not fit the embodied carbon data as well due to non-linear behaviors in the model and a smaller amount of data overall [63]. While the decision tree model is able to make predictions with about equal accuracy throughout the design space, it is still not accurate enough for early building design. The linear model tree is the most accurate of the three models. It is important to acknowledge that other non-parametric machine learning models such as neural networks could achieve higher accuracy, as in [64], [68] but such models would pose difficulty for interpretation. The information extracted from interpretable models is valuable to the design process and central to this paper.

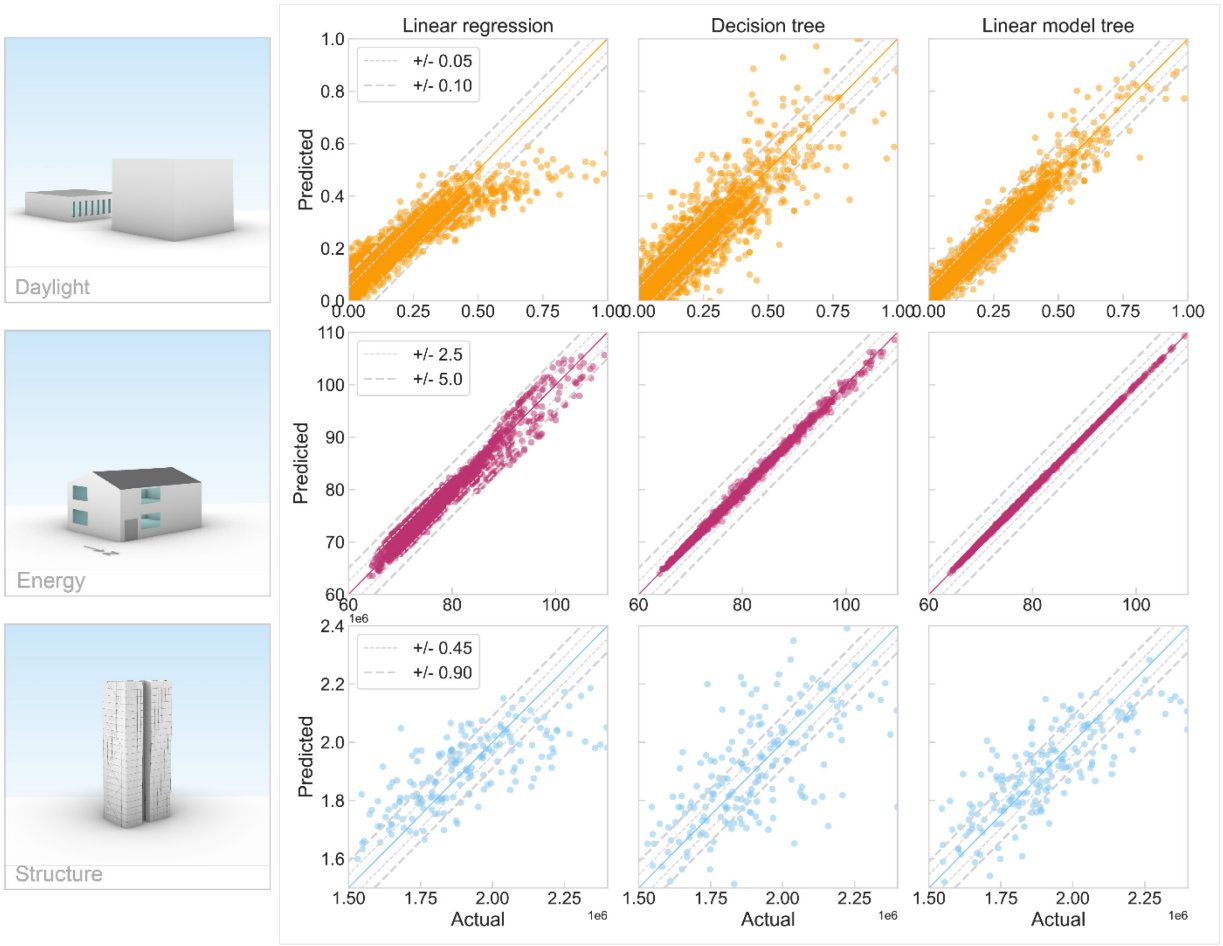

Figure 6: Model fit comparison for spatial daylight autonomy (top), energy use intensity (middle), and embodied carbon (bottom)

In addition to assessing the linear model tree fit, the linear correlations among the variables were checked to ensure collinearity issues are avoided. Figure 7 shows correlation coefficients for each dataset, including in at least one instance where a variable was eliminated due to collinearity. While the variables in the spatial daylight autonomy are not highly correlated, the window SHGC and window U-value are highly correlated. As previously mentioned, the U-value was kept in the model over the SHGC because it had a stronger linear relationship to the EUI. Finally, although the embodied carbon variables have minor correlations, the absolute value of the Pearson correlation coefficients all fall below 0.065, which is reasonable for similar building design problems in the literature [69]. Therefore, the linear regression assumption that all variables are independent is valid.

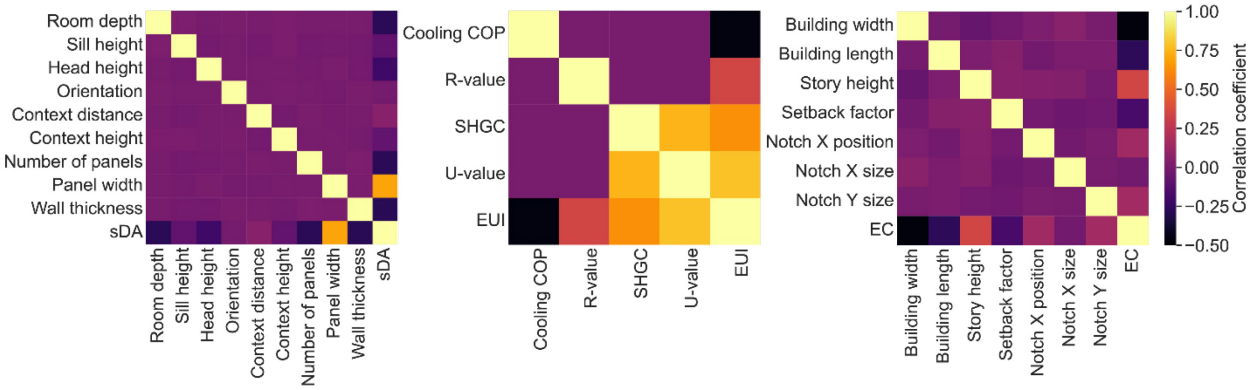

Figure 7: Pearson correlation coefficients for spatial daylight autonomy (left), energy use intensity (middle) and embodied carbon (right)

4.2 Sensitivity over the variable domain in a multi-dimensional design space

Once the linear model trees were trained, the average coefficients for each variable were plotted over their domains (Figure 8). This figure shows where the relationship to the response changes, considering all the variables in the model and all possible design directions. Although many of the variables in the spatial daylight autonomy dataset have the same slope throughout, room depth and panel width show noteworthy changes. On average, panel width does not significantly affect sDA until it reaches ~0.5 relative width of the panel. Designers can freely choose within 0.10-0.50 without affecting sDA. Similarly, room depth greatly influences sDA until it reaches about 8.7m; at this point, increasing the room depth does not change sDA. This is potentially useful information while designing floorplans. In order to achieve a high sDA, other variables must be adjusted if the room depth is beyond 8.7m.

For the energy retrofit model, only low values of cooling COP have a strong effect on the EUI. The simulations were conducted in ASHRAE climate Zone 5, which is heating dominated, so increasing the cooling COP beyond ~2.2 does not result in a significantly different EUI given other variables in the model. Adding insulation to the exterior walls (R-value variable) has a consistent though relatively smaller effect on the EUI throughout its domain. Similar to cooling COP, low U-values strongly affect EUI until about 3 W/m^2-K. The EUI includes HVAC, lighting, plug, and miscellaneous loads, and at some point, the HVAC portion is minimized. This explains the diminishing returns of the incremental insulation and COP. The diminishing returns of the incremental insulation and COP. The results in Figure 8 only consider the coefficient magnitude, but they follow domain knowledge—installing new windows with a low U-value would improve the EUI in a heating-dominated climate. Furthermore, the results in this section specify at what point increasing the variable has a negligible effect. In future sections, the coefficient sign is considered in order to better describe the relationships. Nevertheless, Figure 8 provides a high-level overview of changes in importance to EUI over the variable domain, assuming the other variables are present in the model.

In the embodied carbon dataset, building width is the strongest predictor, especially for very narrow building widths. For very small widths, the lateral system requires extremely large sections to carry the lateral forces from the broad building side, so building width significantly affects overall performance response in this region. Building length is the second-most important predictor; however, the slope is relatively consistent throughout. Among the independent and partially dependent variables considered in [63], building width and building length had the strongest linear relationships (Fig. 13 in [63]), which supports the results in this paper. The embodied carbon design space contains more non-linearities than spatial daylight autonomy and EUI, and although the linear model tree can capture non-

linear behavior through its piece-wise nature, it is restricted based on the training requirements for the number of data points per leaf node. Nevertheless, this result provides designers with a set of ranges to design within without significantly affecting the embodied carbon.

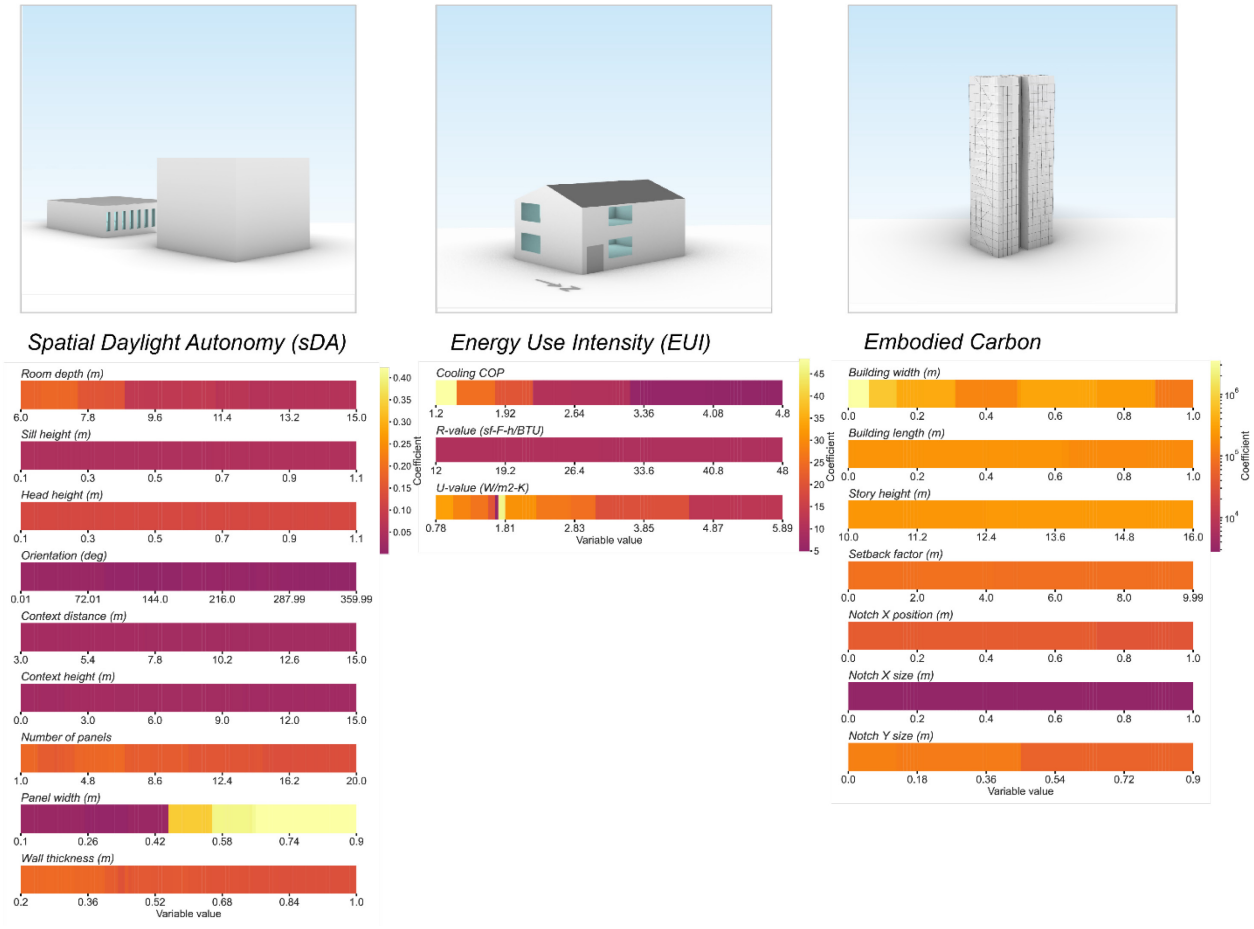

Figure 8: Average sensitivity in small bins for spatial daylight autonomy (left), EUI (middle), and embodied carbon (right)

To understand the relationships on a more granular level, Figure 9 shows the raw output of the procedure described in Section 3.1.1. The gray line represents the linear model coefficient from the overall linear regression model (shown in Fig. 6) for comparison. While Figure 8 shows the absolute value or "importance," Figure 9 shows the sign of the coefficient, which indicates the variables' tendency to increase or decrease the response in each bin or region of the domain. Comparing the two models shows similar but more detailed trends for important variables such as panel width and room depth for daylight and building width for structure. These results can also be interpreted in light of the overall model characteristics. For example, the R-value variable in the energy dataset was split the fewest number of times, so the coefficient was relatively consistent throughout the design space and very similar to the overall linear regression model. The U-value variable shows discontinuous behavior near 2 W/m^2-K because many of window constructions in the dataset had a U-value around this value but differing SHGC and other properties. While the behavior in this region is unstable, it indicates to the designer that there are many potential solutions in this region. This is a result of the real-world, discretely sampled energy variables, as well as the elimination of SHGC due to high correlation.

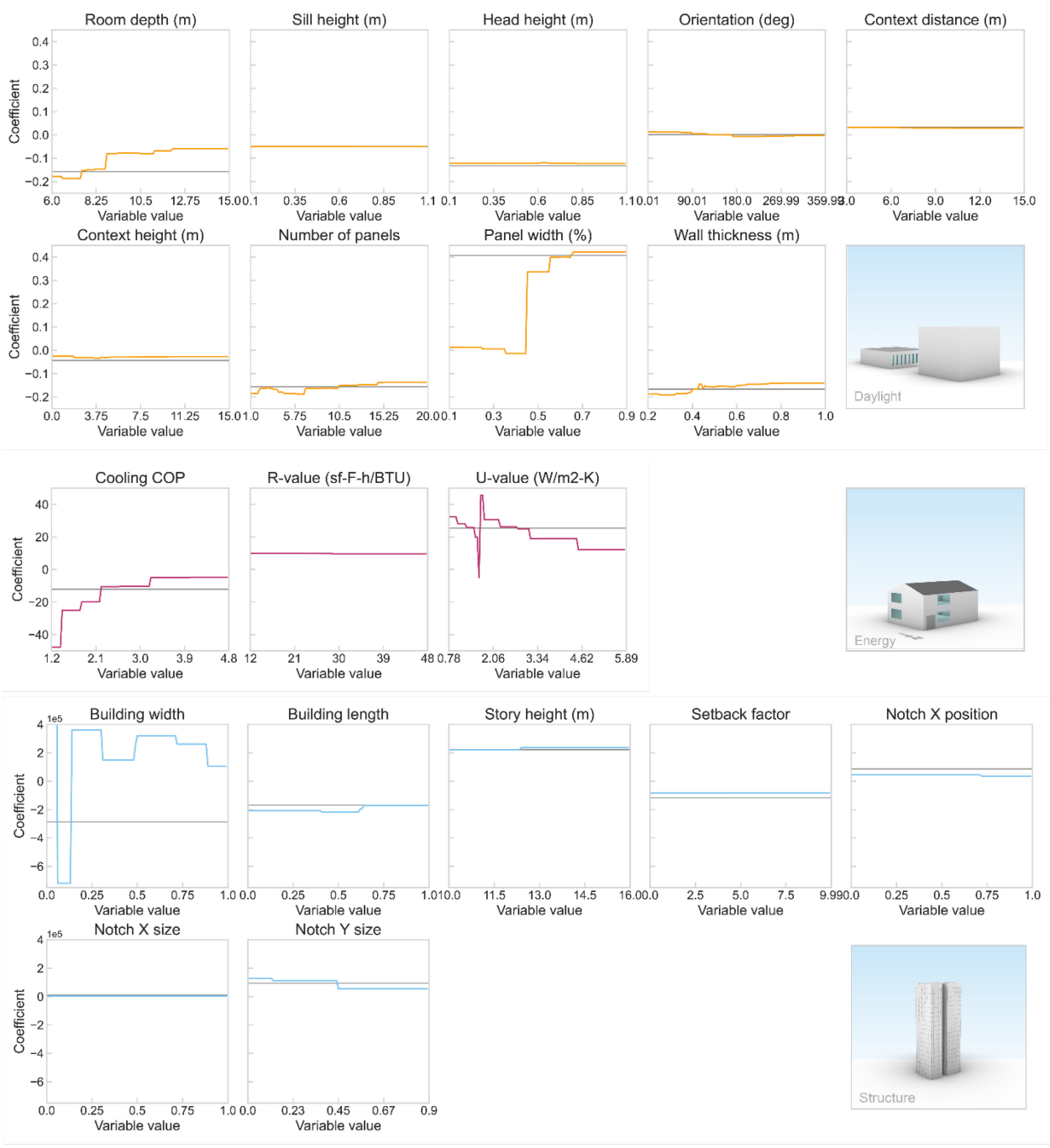

Figure 9: Average coefficient in small bins for spatial daylight autonomy (top), EUI (middle), and embodied carbon (bottom)

These results so far explain how the models were trained, how accurate they are for prediction, and how the linear model coefficients can guide designers on an expected performance response in a certain region of the design space. The following results demonstrate how these models can be aggregated to provide variable importance as designers change the possible ranges of decisions without full model retraining, since relative importance can change significantly in different regions of the design space.

4.3 Dynamic subset sensitivity analysis

4.3.1 Real-time variable importance

Data-driven parametric design often involves setting variable domains, generating data, and fitting a prediction model. As the design is refined, variable domains are narrowed until one value is ultimately selected. Previously, the prediction model needed to be re-trained on the subset of data to provide accurate variable importance and support decisions. We instead achieve subset sensitivity analysis by precomputing linear regression models in regions determined by the tree and then interpolating between regions to estimate the variable importance in the subset. Two examples per design problem are shown in Figure 10, which includes a slider for each variable, the user-defined intervals, and variable importance, presenting a potential visualization for a design tool. It is important to note that a series of visualizations presented to the designer should show both (1) *which* variables deserve attention (by virtue of producing a large effect on performance, regardless of direction) and (2) *how* such variables tend to affect performance along their domains (where the variable makes the performance trend up or down). There is some loss of precision due to the averaging in the simpler graphics, but they are intended for rapid feedback for designers that can be explored in more detail if desired. To give an indication of speed, updating the variable importance from design scenario 1 to design scenario 2 for the daylight design space takes 0.003 seconds on a desktop computer with 32 GB RAM and an Intel Core i7 2.6 GHz processor. The speed also depends on the size of the tree, but this example uses the largest tree among the three datasets. If the method were fully incorporated into an interactive tool, possibly as a plug-in to parametric design software, the rendering speed would depend on the software and would likely be more substantial than the importance calculation.

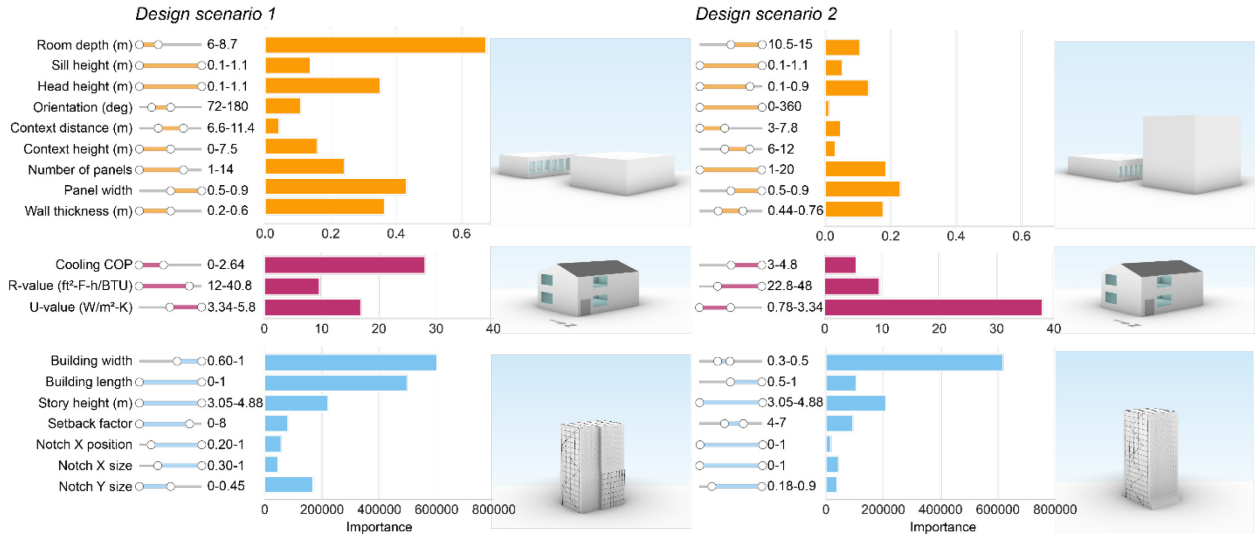

Figure 10: Dynamic subset sensitivity analysis for 2 design scenarios per dataset

Figure 10 shows two sets of design criteria imposed on each design space. Design scenario 1 for spatial daylight autonomy restricts room depth, and thus it is very sensitive in this region. With different restrictions on panel width and number of panels in design scenario 2, room depth is the most important variable. In the second design scenario, with different ranges for room depth, panel width becomes the most important variable. Similar changes are seen in the different design scenarios for energy, as Cooling COP or U-value can become the most important in different regions. In the structure dataset, building width is almost always the most important variable, but in certain scenarios other variables can approach its magnitude of importance to influencing embodied carbon.

4.3.2 Significance in leaf nodes

Although all variables are assigned a coefficient during the linear model fitting step of the linear model tree training procedure, it is possible that some of the variables do not significantly affect the response in certain regions of the design space. To determine if a variable affects the response, a hypothesis test is conducted where the null hypothesis is that the coefficient is equal to zero, which implies that there is no effect. If the p-value is less than 0.05 (5% level of significance), the null hypothesis is rejected and the relationship between the variable and the response is deemed statistically significant. Once the linear model tree was fitted, the coefficients with p-values higher than 0.05 were reset to zero from the calculations described in Sections 3.5 and 3.6. This avoids biasing the results towards coefficients that are not statistically significant.

Figure 11 illustrates how consideration of significance affects each model in this paper, as the blurred heatmap cells contain coefficients that were not statistically significant. The blurred heatmap cells have a translucent mask to represent that the coefficient p-value was higher than 0.05. The y-axis is leaf node model index and the x-axis is variables; the color represents the linear model coefficient. It was important to take coefficient p-values into account to eliminate the possibility of a high magnitude coefficient that is not statistically significant greatly influencing the calculations in Sections 3.5 and 3.6. For example, in the structure dataset leaf node model 30 has a high magnitude coefficient for the width variable, but it is not statistically significant, so it must be excluded to avoid inaccurately representing the behavior in this region of the domain. The coefficients of notch X position, notch X size, and notch Y size were not statistically significant for many leaf node models and were thus ignored. This is consistent with the initial variable assessment in [63], which does not show a clear relationship to embodied carbon throughout the domain. In contrast, the energy dataset variables have a statistically significant relationship to the response in all regions of the design space. Piece-wise linear relationships were observed in the initial data exploration, and all three variables are well-known retrofit strategies, leading to this expected result. Finally, the daylight dataset shows a mix of significant and non-significant leaf nodes, which seems to be most present for orientation, context distance, and context height.

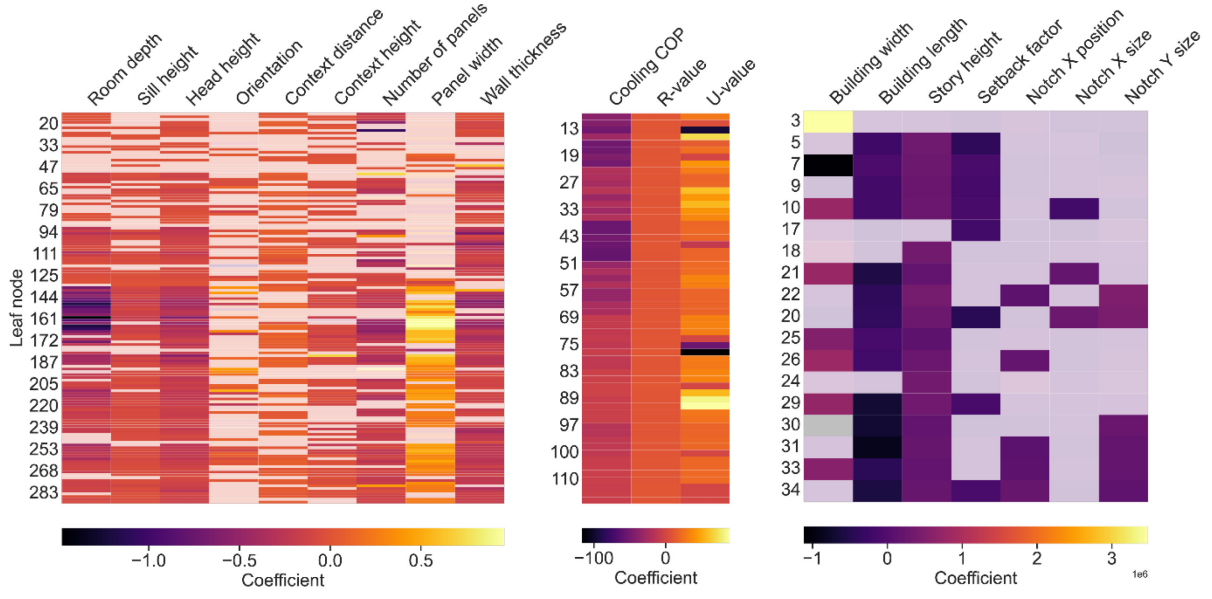

Figure 11: Linear model coefficients for each variable in each leaf node model, with a translucent mask on coefficients that do not have a statistically significant p-value

5. Discussion: recommendations for future datasets

Comparing the application of dynamic subset sensitivity analysis to several general datasets in the architectural engineering domain reveals several benefits and potential pitfalls. Resulting discussion points are included as recommendations for what could be changed or customized for use on future datasets:

- **Sampling technique**: Choose as continuous of a sampling technique as possible to ensure sufficient coverage of the design space for interpolation. If a grid-sampling technique was used to generate the data, it is possible that the variables are split at each option during the training process. At this point, the variables would no longer be treated as variables in the leaf node models. Therefore, if grid-sampling is used to generate the data, it is important to make sure the grid is fine enough. It is recommended to use Latin Hypercube sampling or similar to avoid this problem.
- **P-values in leaf nodes**: It is important to check the variable p-values in the leaf nodes, and if the p-values fall below the desired level of significance, the corresponding coefficient should be forced to zero in order to accurately represent variable importance.
- **Hyperparameters**: The hyperparameters determine the sensitivity of the interpolation calculations. Increasing the power p hyperparameter puts more emphasis on the leaf nodes with a higher agreement. For a design setting, it is recommended to keep the power low to proportionally account for the behavior in the leaf nodes, even those with a lower agreement. When choosing the appropriate number of leaf nodes in the calculations, hyperparameter t, it is important to consider the size of the dataset. The maximum t value is the total number of leaf nodes in the linear model tree, which depends on the size of the dataset and the training requirements.
- **Leaf node model fit**: It is recommended to calculate the R^2 values for the leaf node models and to assign the leaf node models with a low R^2 value a lower weight in the interpolation calculations. These models could also be useful information to the designer, as these regions are highly nonlinear and could not be handled by the linear model tree. The trends or tradeoffs in these regions may differ from the surrounding regions.
- **Traditional decision tree importance metric**: The typical decision tree has an importance metric based on how much the error metric was reduced by each split. However, this only indicates which variables are highly nonlinear, not which variables have the steepest slopes or highest importance. The metrics in this paper were developed to capture this.
- **Normalization among leaf node models:** The coefficients from all the leaf node models could be normalized, but then the method would not provide "how much" the variables matter, just a relative ranking of variable importance.
- **Number of samples**: In order to produce reliable linear models in each node, the algorithm enforces a specified number of data points per leaf node. In this paper, it was assumed the number of data points required per leaf node was 30 data points. The structural dataset contained 7 variables and 940 data points, which resulted in only eighteen leaf node models. During the interpolation process described in Section 3.1.2, there were only 18 models to consider, versus the spatial daylight autonomy dataset which had 144 leaf node models to consider.

As demonstrated in Section 4.1, it is also necessary to reduce collinearity among variables. Collinearity can be assessed by calculating the Pearson correlation coefficients. For example, the energy dataset in this paper had two variables, SHGC and U-Value, that were highly correlated, and it was necessary to eliminate one to prevent model instability issues. Because U-value showed a stronger linear relationship to the response EUI, SHGC was eliminated. Variable selection can be conducted in many other ways

including stepwise selection, forward selection, and backward elimination. It is ultimately up to the designer to determine which variables to include in the model.

6. Conclusion

6.1 Summary of contributions

This work presents a method for dynamic subset sensitivity analysis that includes a new procedure for ensuring coefficient significance. It then demonstrates the method's generalizability on three building design problems. This method updates variable importance in real-time as design criteria emerge, aiding discussion for new design directions. It also determines where in the variables' domain it tends to influence the response, which provides ranges to design within and supports design freedom.

6.2 Limitations and future work

Some aspects of this specific approach depend on having linear model coefficients. The model tree could include quadratic or cubic terms in the linear regression models to produce local polynomial models. Additionally, it is possible to implement the model tree with other node model types such as neural networks or SVM. However, linear models were selected in this method to utilize the coefficients to develop importance metrics, as well as to reduce training time. To implement the model tree with other model types, additional importance metrics must be developed, especially for nonparametric models. It is likely the training time would also increase. Another limitation for the daylight and energy datasets is using a single location. In future iterations, latitude and cloud condition could be included as variables to make it more flexible. Finally, it could be argued that the size of the embodied carbon dataset was not large enough for a model tree given the nonlinear nature of many of the variables, compared to energy [70]. However, this example was chosen to demonstrate the method on an existing dataset that was not developed directly for this method. Future general datasets in the domain of structures should be based on a larger dataset.

6.3 Concluding remarks

In this work, we investigate a new method called dynamic subset sensitivity analysis across three domains. Many factors on the dataset affect the effectiveness of the method, specifically the sampling technique and the number of samples. Considering the quality of the leaf node models through the coefficient p-value and R^2 improve the reliability of the interpolated variable importance. In the future, this work could be combined with recent work on training design agents to learn generalizable design behavior [71]. If implemented more widely, methods such as dynamic subset sensitivity analysis could track with design practice to make the greatest impact without requiring computation specialists to generate a custom parametric model and simulation data for each project.

Acknowledgments

This research is supported in part by the National Science Foundation under Grant #2033332. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF.

References

[1] L. Caldas, "Generation of energy-efficient architecture solutions applying GENE_ARCH: An evolution-based generative design system," *Advanced Engineering Informatics*, vol. 22, no. 1, pp. 59–70, Jan. 2008, doi: 10.1016/J.AEI.2007.08.012.

[2] Y. Wang and C. Wei, “Design optimization of office building envelope based on quantum genetic algorithm for energy conservation,” *Journal of Building Engineering*, vol. 35, Mar. 2021, doi: 10.1016/J.JOBE.2020.102048.

[3] Y. Fang and S. Cho, “Design optimization of building geometry and fenestration for daylighting and energy performance,” *Solar Energy*, vol. 191, pp. 7–18, Oct. 2019, doi: 10.1016/j.solener.2019.08.039.

[4] C. Mueller, “Combining parametric modeling and interactive optimization for high- performance and creative structural design,” in *Proceedings of IASS Annual Symposia, IASS 2015 Amsterdam Symposium: Future Visions – Computational Design*, 2015, pp. 1–11.

[5] M. Turrin, P. Von Buelow, and R. Stouffs, “Design explorations of performance driven geometry in architectural design using parametric modeling and genetic algorithms,” *Advanced Engineering Informatics*, vol. 25, no. 4, pp. 656–675, Oct. 2011, doi: 10.1016/j.aei.2011.07.009.

[6] N. C. Brown and C. T. Mueller, “Gradient-based guidance for controlling performance in early design exploration,” *Proceedings of the IASS Annual Symposium 2018*, 2018.

[7] C. T. Mueller and J. A. Ochsendorf, “Combining structural performance and designer preferences in evolutionary design space exploration,” *Automation in Construction*, vol. 52, pp. 70–82, Apr. 2015, doi: 10.1016/j.autcon.2015.02.011.

[8] N. Brown and C. Mueller, “Designing with data: Moving beyond the design space catalog,” in *Disciplines and Disruption - Proceedings Catalog of the 37th Annual Conference of the Association for Computer Aided Design in Architecture, ACADIA 2017*, 2017, pp. 154–163.

[9] R. Balling, “Design by shopping: A new paradigm,” in *Proceedings of the Third World Congress of structural and multidisciplinary optimization (WCSMO-3)*, 1999, pp. 295–297.

[10] G. M. Stump, M. Yukish, T. W. Simpson, and E. N. Harris, “Design Space Visualization and Its Application to a Design by Shopping Paradigm.” pp. 795–804, Sep. 02, 2003, doi: 10.1115/DETC2003/DAC-48785.

[11] P. Westermann, M. Welzel, and R. Evins, “Using a deep temporal convolutional network as a building energy surrogate model that spans multiple climate zones,” *Applied Energy*, vol. 278, Nov. 2020, doi: 10.1016/j.apenergy.2020.115563.

[12] T. Wortmann, A. Costa, G. Nannicini, and T. Schroepfer, “Advantages of surrogate models for architectural design optimization,” *Artificial Intelligence for Engineering Design, Analysis and Manufacturing: AIEDAM*, vol. 29, no. 4, pp. 471–481, Oct. 2015, doi: 10.1017/S0890060415000451.

[13] T. Wortmann, “Surveying design spaces with performance maps: A multivariate visualization method for parametric design and architectural design optimization,” *International Journal of Architectural Computing*, vol. 15, no. 1, pp. 38–53, 2017, doi: 10.1177/1478077117691600.

[14] S. Attia, E. Gratia, A. De Herde, and J. L. M. Hensen, “Simulation-based decision support tool for early stages of zero-energy building design,” *Energy and Buildings*, vol. 49. Elsevier, pp. 2–15, Jun. 01, 2012, doi: 10.1016/j.enbuild.2012.01.028.

[15] P. Pilechiha, M. Mahdavinejad, F. Pour Rahimian, P. Carnemolla, and S. Seyedzadeh, “Multi-objective optimisation framework for designing office windows: quality of view, daylight and energy efficiency,” *Applied Energy*, vol. 261, Mar. 2020, doi: 10.1016/J.APENERGY.2019.114356.

[16] N. C. Brown, “Design performance and designer preference in an interactive, data-driven conceptual building design scenario,” *Design Studies*, vol. 68, pp. 1–33, May 2020, doi: 10.1016/j.destud.2020.01.001.

[17] P. Westermann and R. Evins, “Surrogate modelling for sustainable building design – A review,” *Energy and Buildings*, vol. 198. Elsevier Ltd, pp. 170–186, Sep. 01, 2019, doi: 10.1016/j.enbuild.2019.05.057.

[18] T. Yang, Y. Pan, J. Mao, Y. Wang, and Z. Huang, “An automated optimization method for calibrating building energy simulation models with measured data: Orientation and a case study,” *Applied Energy*, vol. 179, pp. 1220–1231, Oct. 2016, doi: 10.1016/j.apenergy.2016.07.084.

[19] S. Gou, V. M. Nik, J.-L. Scartezzini, Q. Zhao, and Z. Li, “Passive design optimization of newly-built residential buildings in Shanghai for improving indoor thermal comfort while reducing building energy demand,” *Energy & Buildings*, vol. 169, pp. 484–506, 2017, doi: 10.1016/j.enbuild.2017.09.095.

[20] T. Østergård, R. L. Jensen, and S. E. Maagaard, “Interactive building design space exploration using regionalized sensitivity analysis,” in *Building Simulation Conference Proceedings*, 2017, vol. 4, pp. 1997–2006, doi: 10.26868/25222708.2017.185.

[21] J. Yang, “Convergence and uncertainty analyses in Monte-Carlo based sensitivity analysis,” *Environmental Modelling & Software*, vol. 26, no. 4, pp. 444–457, 2011, doi: 10.1016/j.envsoft.2010.10.007.

[22] L. Hinkle, G. Pavlak, N. Brown, and L. Curtis, “Dynamic Subset Sensitivity Analysis For Design Exploration,” *2022 Annual Modeling and Simulation Conference (ANNSIM)*, pp. 581–592, Jul. 2022, doi: 10.23919/ANNSIM55834.2022.9859293.

[23] N. C. Brown, “Design performance and designer preference in an interactive, data-driven conceptual building design scenario,” *Design Studies*, vol. 68, pp. 1–33, 2020, doi: 10.1016/j.destud.2020.01.001.

[24] S. P. Thole and P. Ramu, “Design space exploration and optimization using self-organizing maps,” *Structural and Multidisciplinary Optimization*, vol. 62, no. 3, pp. 1071–1088, Jul. 2020, doi: 10.1007/s00158-020-02665-6.

[25] W. J. Doherty and R. P. Kelisky, “Managing VM/CMS systems for user effectiveness,” *IBM Systems Journal*, vol. 18, no. 1, pp. 143–163, Mar. 1979, doi: 10.1147/SJ.181.0143.

[26] J. T. Brady, “A Theory of Productivity in the Creative Process,” *IEEE Computer Graphics and Applications*, vol. 6, no. 5, pp. 25–34, 1986, doi: 10.1109/MCG.1986.276789.

[27] M. Csikszentmihalyi, “Flow and the psychology of discovery and invention,” in *Flow and the psychology of discovery and invention*, New York: HarperPerennial, 1997, pp. 1–16.

[28] N. L. Jones, “Validated interactive daylighting analysis for architectural design,” Massachusetts Institute of Technology, 2017.

[29] S. Tseranidis, N. C. Brown, and C. T. Mueller, “Data-driven approximation algorithms for rapid performance evaluation and optimization of civil structures,” *Automation in Construction*, 2016, doi: 10.1016/j.autcon.2016.02.002.

[30] A. I. J. Forrester, A. Sóbester, and A. J. Keane, *Engineering Design via Surrogate Modelling*. Wiley, 2008.

[31] Z. Romani, A. Draoui, and F. Allard, “Metamodeling the heating and cooling energy needs and simultaneous building envelope optimization for low energy building design in Morocco,” *Energy and Buildings*, vol. 102, pp. 139–148, 2015, doi: 10.1016/j.enbuild.2015.04.014.

[32] P. Geyer and A. Schlüter, “Automated metamodel generation for Design Space Exploration and decision-making - A novel method supporting performance-oriented building design and retrofitting,” *Applied Energy*, vol. 119, pp. 537–556, 2014, doi: 10.1016/j.apenergy.2013.12.064.

[33] S. Shamshirband *et al.*, “Predicting Standardized Streamflow index for hydrological drought using machine learning models,” *Engineering Applications of Computational Fluid Mechanics*, vol. 14, no. 1, pp. 339–350, Jan. 2020, doi: 10.1080/19942060.2020.1715844.

[34] A. Noulas, S. Scellato, N. Lathia, and C. Mascolo, “Mining user mobility features for next place prediction in location-based services,” *Proceedings - IEEE International Conference on Data Mining, ICDM*, pp. 1038–1043, 2012, doi: 10.1109/ICDM.2012.113.

[35] J. Golbeck, C. Robles, and K. Turner, “Predicting personality with social media,” in *Conference on Human Factors in Computing Systems - Proceedings*, 2011, pp. 253–262, doi: 10.1145/1979742.1979614.

[36] T. Wortmann, J. Cichocka, and C. Waibel, “Simulation-based optimization in architecture and building engineering—Results from an international user survey in practice and research,” *Energy and Buildings*, vol. 259, Mar. 2022, doi: 10.1016/j.enbuild.2022.111863.

[37] E. Whalen and C. Mueller, “Toward Reusable Surrogate Models: Graph-Based Transfer Learning on Trusses,” *Journal of Mechanical Design, Transactions of the ASME*, vol. 144, no. 2, Feb. 2022, doi: 10.1115/1.4052298/1119281.

[38] S. Xu, Y. Wang, Y. Wang, Z. O’neill, Q. Zhu, and Q. 2020 Zhu, “One for Many: Transfer Learning for Building HVAC Control,” in *BuildSys ’20: Proceedings of the 7th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation*, 2020, pp. 230–239, doi: 10.1145/3408308.3427617.

[39] A. Li, F. Xiao, C. Fan, and M. Hu, “Development of an ANN-based building energy model for information-poor buildings using transfer learning,” *Building Simulation*, vol. 14, no. 1, pp. 89–101, 2021, doi: 10.1007/s12273-020-0711-5.

[40] B. Bass, L. Curtis, J. New, S. Sanborn, and P. McNally, “Universal design space exploration for building energy design,” *Journal of Building Engineering*, vol. 68, Jun. 2023, doi: 10.1016/j.jobe.2023.105977.

[41] “Asterisk.” Thornton Tomasetti, Accessed: May 24, 2023. [Online]. Available: https://asterisk.thorntontomasetti.com/.

[42] K.-M. Mark TAM, T. Van Mele, and P. Block, “Trans-topological learning and optimisation of reticulated equilibrium shell structures with Automatic Differentiation and CW Complexes Message Passing,” in *Proceedings of the IASS 2022 Symposium affiliated with APCS 2022 conference*, 2022, pp. 1–13.

[43] H. Samuelson, S. Claussnitzer, A. Goyal, Y. Chen, and A. Romo-Castillo, “Parametric energy simulation in early design: High-rise residential buildings in urban contexts,” *Building and Environment*, vol. 101, pp. 19–31, May 2016, doi: 10.1016/J.BUILDENV.2016.02.018.

[44] E. Brembilla, C. J. Hopfe, and J. Mardaljevic, “Influence of input reflectance values on climate-based daylight metrics using sensitivity analysis,” *Journal of Building Performance Simulation*, vol. 11, no. 3, pp. 333–349, May 2018, doi: 10.1080/19401493.2017.1364786.

[45] A. T. Nguyen and S. Reiter, “A performance comparison of sensitivity analysis methods for building energy models,” *Building Simulation*, vol. 8, no. 6, pp. 651–664, Dec. 2015, doi: 10.1007/s12273-015-0245-4.

[46] M. L. Pannier, P. Schalbart, and B. Peuportier, “Comprehensive assessment of sensitivity analysis methods for the identification of influential factors in building life cycle assessment,” *Journal of Cleaner Production*, vol. 199, pp. 466–480, Oct. 2018, doi: 10.1016/j.jclepro.2018.07.070.

[47] S. De Wit and G. Augenbroe, “Analysis of uncertainty in building design evaluations and its implications,” *Energy and Buildings*, vol. 34, no. 9, pp. 951–958, Oct. 2002, doi: 10.1016/S0378-7788(02)00070-1.

[48] S. Qiu, Z. Li, Z. Pang, W. Zhang, and Z. Li, “A quick auto-calibration approach based on normative energy models,” *Energy and Buildings*, vol. 172, pp. 35–46, Aug. 2018, doi: 10.1016/J.ENBUILD.2018.04.053.

[49] X. Chen, H. Yang, and K. Sun, “Developing a meta-model for sensitivity analyses and prediction of building performance for passively designed high-rise residential buildings,” *Applied Energy*, vol. 194, pp. 422–439, May 2017, doi: 10.1016/J.APENERGY.2016.08.180.

[50] Z. Yu, F. Haghighat, B. C. M. Fung, and H. Yoshino, “A decision tree method for building energy demand modeling,” *Energy and Buildings*, vol. 42, no. 10, pp. 1637–1646, Oct. 2010, doi: 10.1016/J.ENBUILD.2010.04.006.

[51] Z. X. Conti and S. Kaijima, “Enabling inference in performance-driven design exploration,” in *Humanizing Digital Reality: Design Modelling Symposium*, Springer, Singapore, 2017, pp. 177–188.

[52] N. Delgarm, B. Sajadi, K. Azarbad, and S. Delgarm, “Sensitivity analysis of building energy performance: A simulation-based approach using OFAT and variance-based sensitivity analysis methods,” *Journal of Building Engineering*, vol. 15, pp. 181–193, 2018, doi: 10.1016/j.jobe.2017.11.020.

[53] Y. Gao, M. Mae, and K. Taniguchi, “A new design exploring framework based on sensitivity analysis and Gaussian process regression in the early design stage,” *Journal of Asian Architecture and Building Engineering*, pp. 1–14, 2020, doi: 10.1080/13467581.2020.1783271.

[54] A. Mohiuddin and R. Woodbury, “Interactive parallel coordinates for parametric design space exploration,” in *Conference on Human Factors in Computing Systems - Proceedings*, 2020, pp. 1–9, doi: 10.1145/3334480.3383101.

[55] S. P. Thole and P. Ramu, “Design space exploration and optimization using self-organizing maps,” *Structural and Multidisciplinary Optimization*, vol. 62, no. 3, pp. 1071–1088, Jul. 2020, doi: 10.1007/s00158-020-02665-6.

[56] J. Harding, “Dimensionality Reduction for Parametric Design Exploration,” in *Advances in Architectural Geometry 2016 - Dimensionality Reduction for Parametric Design Exploration*, 2016, pp. 274–287, doi: 10.3218/3778-4_19.

[57] R. Danhaive and C. T. Mueller, “Design subspace learning: Structural design space exploration using performance-conditioned generative modeling,” *Automation in Construction*, vol. 127, Jul. 2021, doi: 10.1016/j.autcon.2021.103664.

[58] S. Wehrend and C. Lewis, “A problem-oriented classification of visualization techniques,” in *VIS ’90: Proceedings of the 1st conference on Visualization*, 1990, pp. 139–143, 46, doi: 10.1109/visual.1990.146375.

[59] "LEED v4 for Building Design and Construction - current version | U.S. Green Building Council." https://www.usgbc.org/resources/leed-v4-building-design-and-construction-current-version (accessed May 21, 2023).

[60] "IES LM-83-12 Approved Method: IES Spatial Daylight Autonomy (sDA) and Annual Sunlight Exposure (ASE)," 2012.

[61] L. Gosselin and J. M. Dussault, "Correlations for glazing properties and representation of glazing types with continuous variables for daylight and energy simulations," *Solar Energy*, vol. 141, pp. 159–165, 2017, doi: 10.1016/j.solener.2016.11.031.

[62] J. Wang, L. Caldas, L. Huo, and Y. Song, "Simulation-based optimization on window properties based on existing products," *Building Simulation and Optimization 2016*. 2016.

[63] I. Hens, R. Solnosky, and N. C. Brown, "Design space exploration for comparing embodied carbon in tall timber structural systems," *Energy and Buildings*, vol. 244, Aug. 2021, doi: 10.1016/j.enbuild.2021.110983.

[64] S. Zargar and N. C. Brown, "Deep learning in early-stage structural performance prediction: assessing morphological parameters for buildings," in *Proceedings of the IASS Annual Symposium*, 2021, pp. 1–13.

[65] I. Hens, R. Solnosky, and N. Brown, "Parametric Framework for Early Evaluation of Prescriptive Fire Design and Structural Feasibility in Tall Timber," *Journal of Architectural Engineering*, vol. 29, no. 1, Mar. 2023, doi: 10.1061/jaeied.aeeng-1455.

[66] A. Wong, "Introduction to Model Trees from scratch," *Towards Data Science*, 2018.

[67] L. Hinkle, G. Pavlak, N. Brown, and L. Curtis, "Dynamic Subset Sensitivity Analysis For Design Exploration," *Proceedings of the 2022 Annual Modeling and Simulation Conference, ANNSIM 2022*, pp. 581–592, 2022, doi: 10.23919/ANNSIM55834.2022.9859293.

[68] S. Tseranidis, N. C. Brown, and C. T. Mueller, "Data-driven approximation algorithms for rapid performance evaluation and optimization of civil structures," *Automation in Construction*, vol. 72, pp. 279–293, Dec. 2016, doi: 10.1016/J.AUTCON.2016.02.002.

[69] Z. O'Neill and F. Niu, "Uncertainty and sensitivity analysis of spatio-temporal occupant behaviors on residential building energy usage utilizing Karhunen-Loève expansion," *Building and Environment*, vol. 115, pp. 157–172, 2017, doi: 10.1016/j.buildenv.2017.01.025.

[70] N. C. Brown, V. Jusiega, and C. T. Mueller, "Implementing data-driven parametric building design with a flexible toolbox," *Automation in Construction*, vol. 118, Oct. 2020, doi: 10.1016/j.autcon.2020.103252.

[71] A. Raina, J. Cagan, and C. McComb, "Learning to Design Without Prior Data: Discovering Generalizable Design Strategies Using Deep Learning and Tree Search," *Journal of Mechanical Design*, vol. 145, no. 3, Mar. 2023, doi: 10.1115/1.4056221.