Clustering metamodel for predictive performance for dynamic shading facade

Cassio Houang Daher[1], Regina Ruschel[1]

[1] Universidade Estadual de Campinas, Brasil
c209606@dac.unicamp.br; ruschel@unicamp.br

Abstract. Dynamic shading facades present a challenge for the computational simulation of illuminance performance in the project's design phase. Such architectural elements have conflicting functions: shading without blocking natural light. The evaluation of these dynamic elements depends on multiple parameters and combinations, which results in higher complexity in compositions and difficulty in understanding. The objective is to identify the optimized positioning of dynamic elements of the facade; still in the early stages of the project, to have good illuminance performance for all busy hours of the year. The research conducted here is experimental, using computer simulation. Our model considers as the dependent variable the average annual illuminance. The independent variables analyzed are shoebox orientation, positioning of the fins, date, and time. The contribution of this research is to test the set of results of the independent variables by training an algorithm capable of replacing the simulation.

Keywords: Machine Learning, Metamodel, Dynamic Facade, Performance Evaluation, Clustering

1 Introduction

Building envelopes were progressively transformed into lightweight, transparent multilayer envelopes. These modifications led to a successive decrease in thermal mass; and, consequently, a possible increase in thermal loads (Konstantoglou & Tsangrassoulis, 2016). The materiality of the construction, mainly that which constitutes the constructive components of the envelope, can considerably affect the energy performance and comfort of the building (Kheiri, 2018). A Passive facade shading system is an effective bioclimatic means to maintain balance between visual and thermal demands. However, studies have shown that static shading elements are remarkably unable to fully respond to changing weather conditions (Al-Masrani & Al-Obaidi, 2019).

In this context, the opportunity arises to build intelligent facade systems linked to technological achievements, forming adaptive or responsive systems. Their performance aims to counteract antagonistic phenomena, such as shading and lighting, and to act as environmental moderators (Konstantoglou and Tsangrassoulis, 2016). Dynamic shading systems can be associated with smart facade applications, which have been proven to increase the comfort of users due to their interaction between the external and internal environments (Al-Masrani and Al-Obaidi, 2019). Therefore, the projects which use these systems must be designed aiming to optimize the quality of luminosity present in the internal environments (Sheikh, El and Gerber, 2011; Touloupaki and Theodosiou, 2017). In order to collect daylight more efficiently, the design of a dynamic facade system should incorporate the use of parametric design software, algorithmic design, and dynamic simulation. Altogether, these tools validate the importance of computation and automation to deal with complex problems and conflicting objectives (Sheikh, El and Gerber, 2011).

Computational technological advances have prompted designers to merge environmental needs and computational simulations, developing dynamic building envelopes that respond to different parameters (Al-Masrani and Al-Obaidi, 2019). Optimization algorithms developed in parametric architecture can be coupled with building energy simulation tools, helping designers seek almost ideal design alternatives to achieve high-performance construction projects (Kheiri, 2018). According to Chwif and Medina (2010), only with the advancement of technology; could computer simulation and mathematical models allow the performance of controlled experiments in areas of science that did not use such tools. Daylighting simulations can be applied for different purposes and vary in scale and complexity, covering solutions for internal spaces, building facades, or urban contexts (Ayoub, 2020).

Despite these advances, daytime lighting simulations are still time-consuming and expensive, requiring large computational power (Ayoub, 2020). The initial design stages are characterized by a large number of parameters and, consequently, design adjustment possibilities (Østergård et al., 2017). Knowing the performance of the solutions from the beginning can contribute to better decision-making. Interfering in the initial phases, in which most of the solutions are still presented as concepts, it effectively transforms the development process of a project, both in terms of efficiency and flexibility (Singh et al., 2021). In the conventional design, if the designer wants to change any parameter, the whole process has to be repeated, inversely, in a parametric design, which employs the use of visual programming software such as Grasshopper (Davidson, 2022), the programmer has control of the overall form throughout the design process. Eltaweel and Su (2017) define parametric architecture as the study of architectural systems aimed at defining the relationships between dimensions dependent on the various parameters. An optimization algorithm is a mathematical method whose function is to find the best suitable solution for a given problem, be it minimization or maximization (Belém et al., 2019).

Although parametric modeling allows you to easily test variables and compare solutions, this is still not enough to evaluate too much data. Therefore, optimization algorithms such as Artificial Intelligence (AI) and Machine Learning (ML) are used. AI is the field of computer science that develops computer programs or applications that have capabilities somewhat comparable to human cognitive abilities (Kumar, 2017). AM is an AI technique that allows systems to learn from data, eliminating the need to be explicitly programmed to do so. AM uses algorithms that analyze data looking for patterns, learning from these patterns and making predictions or decisions based on predefined criteria depending on the results found (Russell & Norvig, 2021). In the case of dynamic facade projects, one can take advantage of parametric architecture tools with the integration of performance simulations for data production and analysis of their results using AM. It becomes possible to implement pattern recognition strategies to optimize the performance of these models (Østergård et al., 2017).

Metamodeling is a method of building accelerated and concentrated models that correlate inputs to results obtained by complex mathematical models, replacing the simulation processes (Østergård, Jensen and Maagaard, 2017; Bracht, Melo and Lamberts, 2021). Metamodels can be developed through different methods, using AI techniques. One of its advantages is the ability to make a prediction from a reduced number of parameters with a shorter execution time. Such models have the ability to predict the actions and influences of the variables, making them capable of making decisions that favor the optimization of performance (Østergård, Jensen and Maagaard, 2018).

The search for balance in large data sets can be seen as an optimization task because it is necessary to find a set of solutions that meet multiple criteria simultaneously. In addition, the models need to be capable of generalizations in order to reduce the universe of data without losing precision (Wagdy et al., 2020). When dealing with large datasets, the search for balance needs to be considered in a different way. In this case, the challenge is finding equality between different objectives relevant to your data analysis (Russell & Norvig, 2021). Optimization functions and their consequences are not able to qualify the aesthetics of the building. Figure 1 presents a flowchart with the conceptual structure of a metamodel.

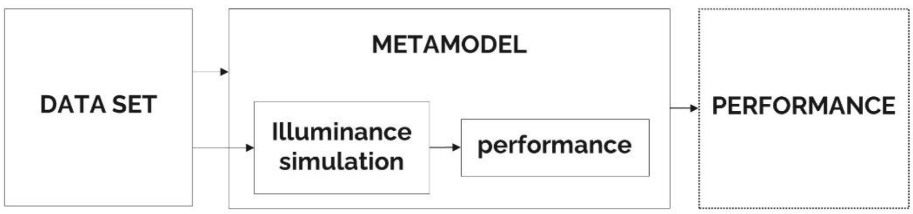

Figure 1. Flowchart that expresses the concept of how a metamodel works. Source: the authors, 2023

In complex models, several samples are essential to improve the generalization of a model since data collection approaches are specific to each of the problems and require predictions of building performance even in the early stages of design (Gao & Malkawi, 2014). In addition, exploring new design solutions during the performance simulation process is a resource that requires intensive exercise. It is imperative to identify a sample of solutions that have representation in a universe to be explored without losing their identity and the ability to solve the problem efficiently. This means that the object is to identify the optimized positioning of the fins using a minimal universe of samples.

2 Methodology

In this research, the problem addressed is the multiplicity of parameters and combinations associated with the positioning of the independent fins, which is represented by the total number of possibilities of times and orientations, generating a set of illuminances. The objective is to identify the optimized positioning of dynamic elements of the facade to achieve better indoor illuminance already in the early stages of the project.

Thus, the following research interests are presented: Shoebox shape algorithm integrated into the simulation to generate the basis for the metamodel, identification of the metamodel, evaluating and testing the metamodel, and shape algorithm integrated into the metamodel to obtain the positioning of the fins. The method of this research is experimental.

The object of study defined for this research is the Shoebox, to show the impact of the main design parameters. The dimensions of the rectangular Shoebox adopted were 3.60 m wide, 8.20 m deep, and a ceiling height of 2.80 m (Reinhart et al., 2013). Only on one side will there be an opening to the outside, 1.00 m high from the floor, 1.50 m high, and 3.00 m wide. Shoebox will be located in Campinas, SP, Brazil. Table 1 presents the set of independent variables adopted.

Table 1. Independent variables and possible combinations

Variable	Orientation	Fins vertical positioning	Fins axial positioning	Time/ Date Busy	TOTAL
Qty.	4	3	5	4745 (6:00-18:00)	1.067.625

Source: the authors, 2023

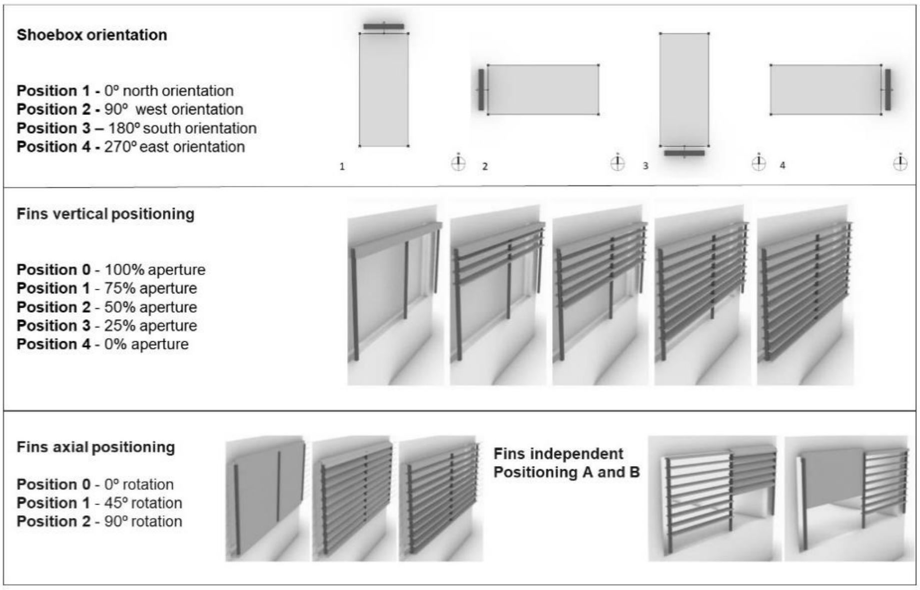

Figure 2. Operationalization of independent variables. Source: the authors, 2023

Fins were spawned in the opening created on one of the sides. The opening is 3.00 meters wide, 1.50 meters high, and 1.00 meters from the floor. The shading elements are formed by flat, rectangular fins. The fin design is not the focus of this research. One of the independent variables is the orientation of the opening, the four possible positions are represented in **Erro! Fonte de referência não encontrada.**. The fins are dynamic elements and present two forms of movement: the axial rotation of their fins and the vertical positioning of the fins.

The entire Shoebox and fins project was developed in Rhinoceros (Mcneel, 2022) and Grasshopper (Davidson, 2022). All independent variables were created using the same algorithm, allowing the exploration of all combinations. The simulations were carried out with Climate Studio (Solemma, 2021). These metrics evaluate daylight availability, which measures indoor illuminance distributions due to the incidence of daylight throughout the year.

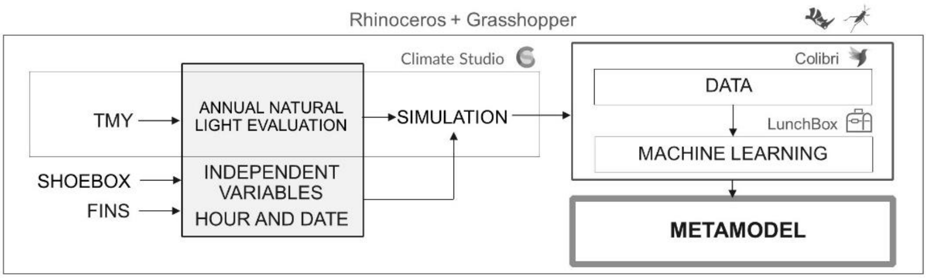

Figure 3. Environment definition flowchart. Source: the authors, 2023

The environment definition flowchart follows the logic of the software and plugins used, as shown in Figure 3. Data collection in experimental research is done by manipulating certain conditions and observing the effects produced. Climate Studio works by evaluating a designated surface, called a grid. The surface to be observed needs to be defined by the user. An algorithm capable of manipulating the shoebox geometry, with its independent variables, was developed and performed for the annual natural light assessment integrated into the Climate Studio. These metrics evaluate indoor illuminance distributions due to the incidence of daylight throughout the year.

The placement of shading elements is explored by generating the annual illuminance dataset. Once the simulation is performed, the data are used in the metamodeling process, which applies two categories of machine learning (ML) algorithms, k-Nearest Neighbor (kNN) optimization and k-Means clustering. kNN identifies the optimal values for each hour of the day, comparing all values in Lux, and finding one closest to the defined ideal value. Lux is a unit of measurement that describes the intensity of illumination perceived by a human observer, commonly used to quantify the brightness or luminosity of a light source or environment.

Clustering shows desirable dataset relationships, produces initial categories during classification, and finds cohesion in a collection of unlabeled data (Gao et al., 2014). This unlabeled data is the optimized placement of the shading fins. Optimized fin placements will be clustered according to the busy time of day using the k-means algorithm within Grasshopper. After defining the similarity between these positions, they are placed in clusters. The results of clustering aim to enable the synthesis of positions. Figure 4 shows the concept employed in the metamodel, the optimized results found are clustered, with the aim of simplifying the range of results. **Erro! Fonte de referência não encontrada.** describes the steps of the algorithm to arrive at the result.

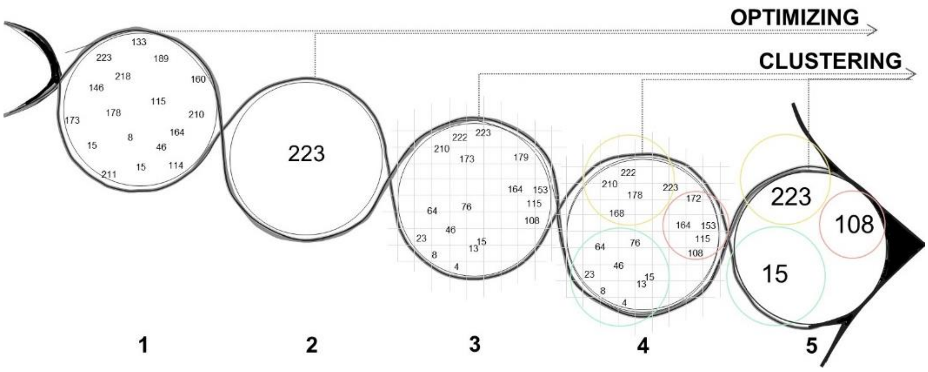

Figure 4. Conceptual diagram of the relationship between optimized results and clustering. The figure is abstract and the values presented are illustrative. Source: the authors, 2023

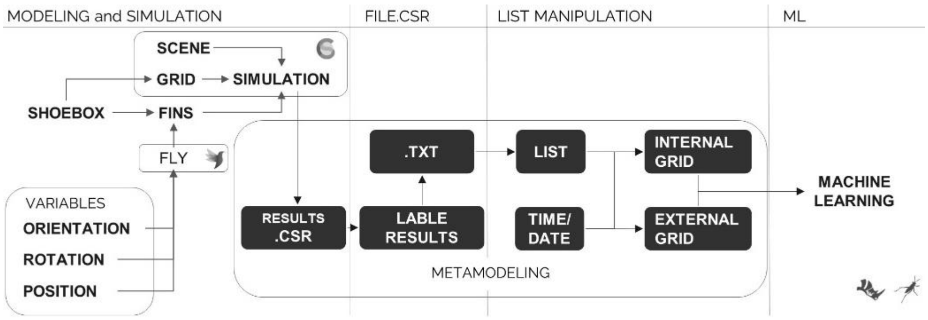

Figure 5. Metamodeling process. Source: the authors, 2023

3 Results

The results of the annual natural light evaluation were presented in average illuminance values of the annual busy hours of a Shoebox, considering only the busy hours, from 6:00 am to 6:00 pm, for office use. The Grid was defined with 40 sensors (divided into 20 external grids and 20 internal grids) and the simulation results for the Climate Studio timestep sensors were separated by date and time [lux] [wh/m²].

Metamodeling was performed in Grasshopper. The logic of the algorithm is one of a continuous line of stimuli applied first to the shape, then the simulation is performed, and then ML is applied to the simulation results.

The process of organizing the lists allows the application of an ML algorithm, classifying the performance of the solution at conception, and allowing different arrangements and combinations of the variables. The optimization algorithm called k-Nearest Neighbor (k-NN) from ML was applied to classify the optimized positions of illuminance of the fins. The .TXT files are imported and then, they are separated into three files; internal zones, external zones and labels. The example represented in **Erro! Fonte de referência não encontrada.** is of the operation of a k-NN algorithm.

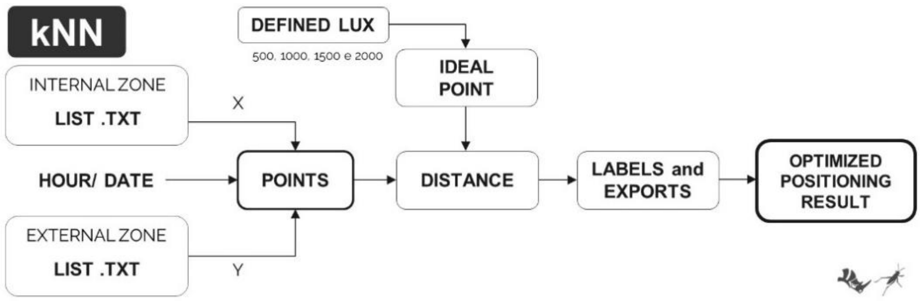

Figure 6. kNN algorithm. Source: the authors, 2023

k-NN calculates the most suitable neighbor, using distance to classify the positions. Among the 225 values for the placements and combinations of the fins, it finds the closest one and defines it as the ideal candidate. The rest of the algorithm evaluates all occupied hours in the year, labels them and exports them to a spreadsheet. The results of the metamodeling process generated the set of placements for the Shoebox fins. They were separated in spreadsheets according to comfort ranges determined by the amount of Lux. In this project, the reference values 500, 1,000, 1,500 and 2,000 lux were adopted, these illuminance ranges were defined following the premise of Nabil and Mardaljevic (2006).

3.1 Clustering

In clustering, entries are divided into groups called clusters. The entries are marked as points, and clusters are then defined based on these list objects according to proximity (Kumar, 2017). The K-means clustering algorithm is a classification analysis method and one of the most popular unsupervised learning algorithms used to identify similarities and patterns. Its function is to turn n observed values into k clusters in which each observation belongs to an exclusive cluster with the closest centroid (Gao and Malkawi, 2014). The procedure follows a simple way of sorting a given data set through a given number of clusters.

The clustering algorithm used in this research is K-means clustering for Grasshopper, developed by the company Proving Ground. The applied algorithm tries to find homogeneity in the data, creating the clusters. This algorithm works with points and is represented in Figure 7. For the data to be clustered, it needs to be transformed into two-dimensional points, with the positioning of the fins in X and the hours occupied in Y. The points are then clustered with Lunch Box K-means algorithm. **Erro! Fonte de referência não encontrada.** shows the result of the clustering.

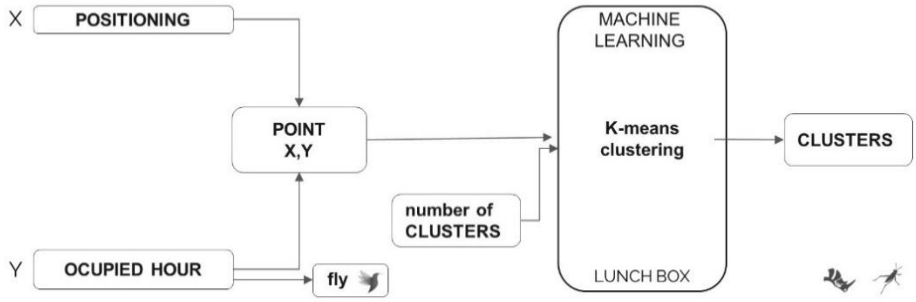

Figure 7. Clustering application algorithm. Source: the authors, 2023

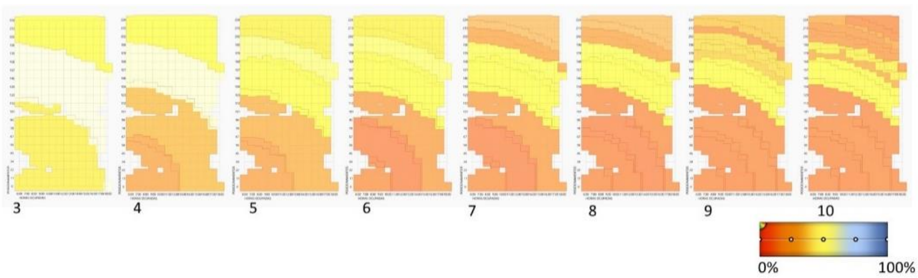

Figure 8. Percentage of placements in each cluster. Source: the authors, 2023

There are 225 different positions of the shading fins for the 5,840 results to be clustered. It was observed that some placements appear repeatedly, that's why the colors represent the percentage of placements in each of the clusters. The clusters were used to replace the optimized mates found with kNN.

Each grouping presents a positioning to test the ability to simplify the variation of positioning of the fins without loss of performance. In the graphs presented in Figure 9, the results are values in Lux for the average monthly illuminance in the Shoebox of each of the performance scenarios: 500 Lux, 1000 Lux, 1500 Lux, and 2000 Lux. Another applied variation is the number of clusters for each of the scenarios. For comparison of the performances in Lux, a graph of performance in Lux was also presented for the monthly average using the results of the optimized placements.

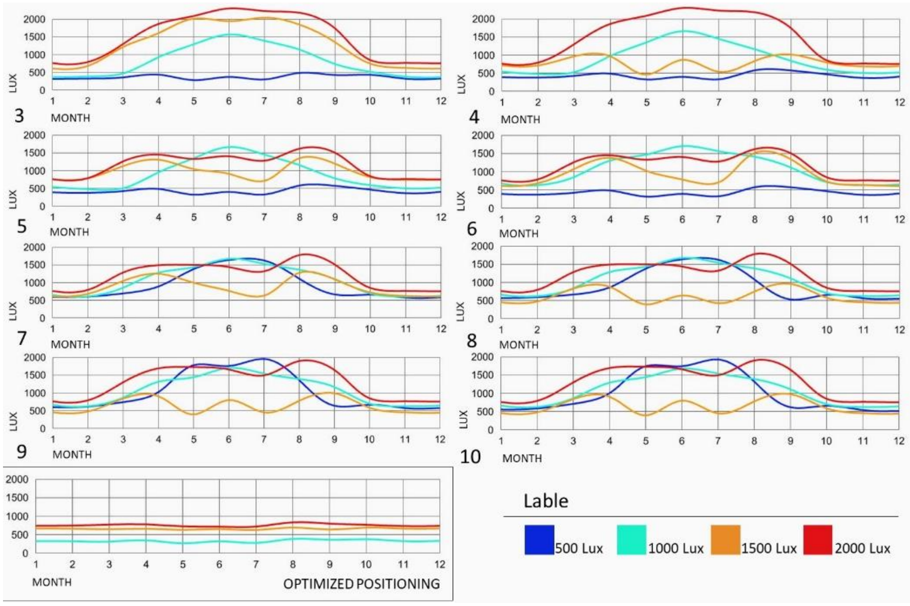

Figure 9. Graphs that compare the performance in Lux of the clustered placements, for the four different evaluated scenarios in Lux. Source: the authors, 2023

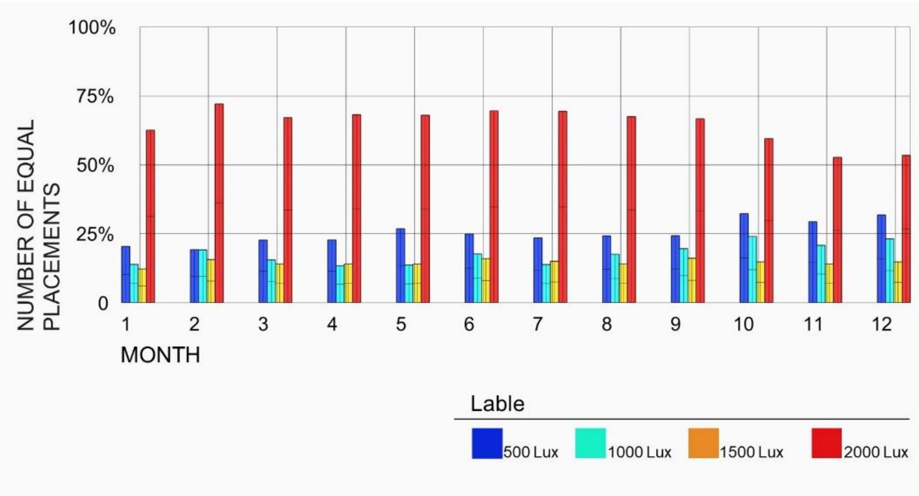

Figure 10. Number of equal placements, monthly average. Source: the authors, 2023

The results of the graphs that relate positioning equalities with the number of clusters indicate that the greater the number of clusters, the greater the number of positions that coincide with those of the optimization results. This is an expected result. Figure 10 represents the graph that compares the number of placements that are repeated between the optimization of placements and the clustering of placements per busy hour, separating the results by month. Figure 11 shows similar graph but divided by busy hours.

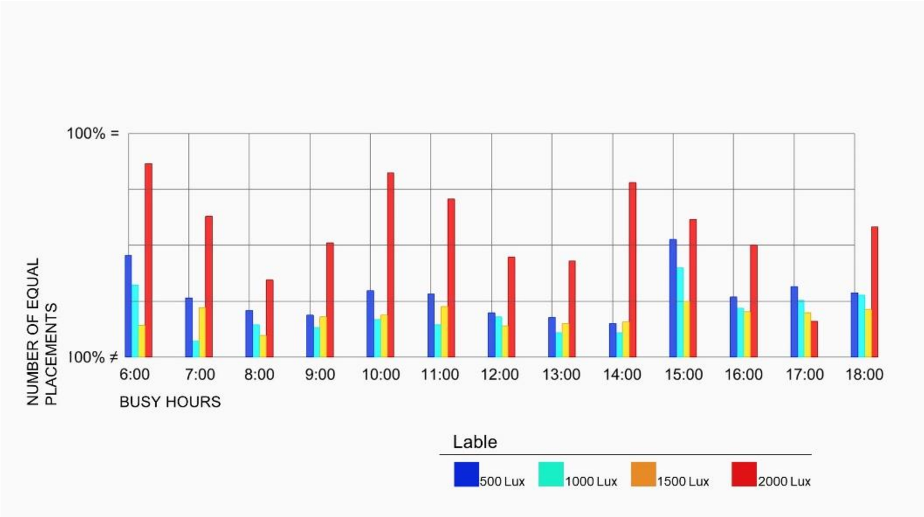

Figure 11. Number of equal placements, busy hours average. Source: the authors, 2023

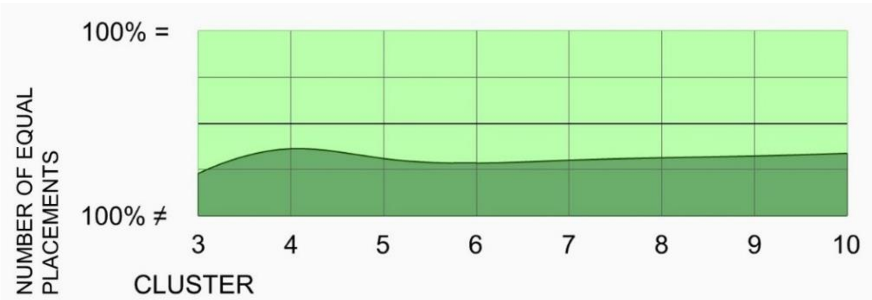

Figure 12. Percentage of equality of placements. Source: the authors, 2023

The total value of placements for each of the results found is 18,980. This means stating that if all clustered placements were equal to the optimized placements, we would have 18,980 equalities. In Figure 12, the values represent the total equality of positioning for the evaluated scenarios.

4 Discussion

Clustering aims to create an identity for each cluster, and it is expected that with this identity created, each cluster will be able to respond to specific stimuli for each busy hour of the day and year in Campinas in terms of Lux performance for the placements of the fins. In the four scenarios evaluated, for 500, 1000, 1500, and 2000 Lux, there is a difference in performance for the clustered results, demonstrating a greater ability to resolve extreme illuminance scenarios. When we compare the results presented to the research question, the metamodel appears capable of solving the optimized positioning of dynamic shading fins. Furthermore, clustering shows that it is possible to reduce the universe of positioning and combinations of fins without losing the ability to balance the function of these dynamic elements, which is to shade the facade, while at the same time bringing natural lighting to the internal environment.

5 Conclusion

The contribution of this research is to generate the set of results of the independent variables, enabling the proposition of a metamodel to replace the simulation. The purpose of applying clustering to the results is to identify the optimal positioning of set of fins, in defined Lux performance scenarios, to guide a dynamic shading system. This research helps to develop tools that would support designers in decision-making in the early stages of a project, taking into account the interactions between multiple conflicting variables.

Acknowledgements. My sincere thanks to Prof. Dr. Ana Lucia Harris (in memory), without whom this work would have no beginning, much less an end.

References

Ayoub, M. (2020). A review on machine learning algorithms to predict daylighting inside buildings. *Solar Energy*, *202*(March), 249–275. https://doi.org/10.1016/j.solener.2020.03.104

Bracht, M. K., Melo, A. P., & Lamberts, R. (2021). A metamodel for building information modeling-building energy modeling integration in early design stage. *Automation in Construction*, *121*(September 2020), 1–19. https://doi.org/10.1016/j.autcon.2020.103422

Chwif, L., & Medina, A. C. (2010). *Modelagem e Simulação de Eventos Discretos: teoria e aplicações.* (3rd ed.).

El Sheikh, M., & Gerber, D. J. (2011). Building skin intelligence: A parametric and algorithmic tool for daylighting performance design integration. *ACADIA 2011*, 170–177.

Gao, X., & Malkawi, A. (2014). A new methodology for building energy performance benchmarking: An approach based on intelligent clustering algorithm. *Energy and Buildings*, *84*, 607–616. https://doi.org/10.1016/j.enbuild.2014.08.030

Kheiri, F. (2018). A review on optimization methods applied in energy-efficient building geometry and envelope design. *Renewable and Sustainable Energy Reviews*, *92*(March), 897–920. https://doi.org/10.1016/j.rser.2018.04.080

Konstantoglou, M., & Tsangrassoulis, A. (2016). Dynamic operation of daylighting and shading systems: A literature review. *Renewable and Sustainable Energy Reviews*, *60*, 268–283. https://doi.org/10.1016/j.rser.2015.12.246

Kumar, R. (2017). Machine Learning and Cognition in Enterprises. In *Machine Learning and Cognition in Enterprises*. Apress. https://doi.org/10.1007/978-1-4842-3069-5

Østergård, T., Jensen, R. L., & Maagaard, S. E. (2017). Early Building Design: Informed decision-making by exploring multidimensional design space using sensitivity analysis. *Energy and Buildings*, *142*, 8–22. https://doi.org/10.1016/j.enbuild.2017.02.059

Østergård, T., Jensen, R. L., & Maagaard, S. E. (2018). A comparison of six metamodeling techniques applied to building performance simulations. *Applied Energy*, *211*(November 2017), 89–103. https://doi.org/10.1016/j.apenergy.2017.10.102

Touloupaki, E., & Theodosiou, T. (2017). Performance simulation integrated in parametric 3D modeling as a method for early stage design optimization - A review. In *Energies* (Vol. 10, Issue 5). MDPI AG. https://doi.org/10.3390/en10050637