Integrated data-driven and knowledge-based performance evaluation for machine assistance in building design decision support

Chen X. [1], Saluz U. [1], Staudt J. [2], Margesin M. [2], Lang W. [2], Geyer P. [1, 3]
[1] Technische Universität Berlin, Germany, [2] Technische Universität München, Germany, [3] Leibniz University Hannover, Germany
xia.chen@tu-berlin.de

Abstract. The building design process requires architects to consider interdisciplinary knowledge and data support based on the vision of sustainable development. In this context, we develop a process-integrated, dynamic machine assistance to support the decision-making process for building designers in the early design phases. In this paper, we present 1. a framework for integrating data-driven models with knowledge-based methods that provide multi-objective assistance considering energy performance and embodied environmental impact; 2. the alignment of the methods in the design process and respective decision situations to disclose the potential situated design space including its uncertainty ranges as well as detailed strategy suggestions. A case of real-world building data serves to illustrate and validate the approach. The research presented in this paper is part of research aiming at assistance by augmented intelligence for sustainable building design decision support.

1. Introduction

The vision of sustainable development already demands consideration in the early building design phases. To address this challenge, various design support tools have been developed to provide information in different aspects: building performance simulation (BPS) tools such as EnergyPlus, Sefaira (Østergård, Jensen & Maagaard, 2016), and data-driven machine learning methods (Seyedzadeh, Rahimian, Glesk & Roper, 2018) for energy performance evaluation, etc. Embodied emissions are accounted for by BIM-based life cycle assessment (LCA) tools, such as (CAALA, 2019) and (One Click LCA® software, 2022). While in fact, these tools are implemented based on different principles, requiring building designers to access interdisciplinary knowledge and data to estimate building performance. This creates the design space in the early design phases (Østergård, Jensen & Maagaard, 2017) and involves intertangled interdependencies and complexity, which surpasses the capacity of conventional design methods. Although the digitalization trend supports the integration of different methodologies, both data-driven and domain knowledge-based, up to now, the synergy of both methodologies is not integrated into design processes at different development levels.

For building energy performance, energy consumption information is relatively easily accessible via smart meter measurements with a particular set of building characteristics representation, such as building geometric features, activity behavior, material properties, etc. (Hensen & Lamberts, 2019). A diverse dataset is available via large-scale records in the real world from existing buildings or generated by validated simulation under first-principles methods. Such a dataset is suitable for data-driven approaches, or more specifically, machine learning approaches for supervised learning to capture the implicit relationship between inputs and outputs. The trained model (after learning from the dataset) is fit for a certain range of interpolation or extrapolation for new cases, which brings the model the advantage of flexibility. The effectiveness and accuracy are well-proved in many domains (Bélisle, Huang, Le Digabel & Aimen E. Gheribi, 2015; Thessen, 2016; Jia & Ma, 2017). Comprehensive reviews report the wide acceptance of data-driven approaches in our domain in the aspect of

https://doi.org/10.7146/aul.455.c202

energy performance (heating, cooling, lighting, etc.) and consumption prediction (Westermann & Evins, 2019; Amasyali & El-Gohary, 2018).

Knowledge-based methods required in this study for LCA, are involved in the sustainable building design process (Schneider-Marin & Lang, 2020; Hollberg, Tschetwertak, Schneider & Habert, 2018). They widely exist in the general engineering process for solving specific problems. Related first-principles tools require detailed information input that is typically not accessible at early design phases or suggests a level of precision that might obscure the potential outcomes of various construction types. The shared characteristics of these problems are: The data acquisition is relatively implicit; The calculation or assessment process requires induction, reasoning, and referring to other background knowledge with limited information. In this context, knowledge representations from experts are inevitably more effective and interpretable, which conducts the gap between knowledge-based and data-driven approaches in the general engineering domain and raises the research need for method integration. In most current design processes embodied emissions are evaluated at a later phase when most construction and material decisions have already been made. The prediction of embodied emissions in early design phases is subject to significant uncertainties due to a lack of detailed information (Schneider-Marin, Harter, Tkachuk & Lang, 2020; Harter, Singh, Schneider-Marin, Lang & Geyer, 2020). Bridging the gap between the lack of information in early design phases and the LCA methodology to predict embodied emissions has been identified as a significant research gap (Theißen, Höper, Wimmer, Zibell, Meins-Becker, Rössig, Goitowski & Lambertz, 2020).

In this study, we propose a general framework to integrate both data-driven and knowledge-based approaches. We intend to investigate the path of integration toward multi-objective support applied to the sustainable building design domain. In this case, both energy performance and embodied environmental impact are selected as objectives. The novelty of this framework is as follows:

- We propose an integrative modelling approach: By introducing decomposition knowledge from design, LCA and BPS simultaneously into the data generation process for data-driven model training, while both approaches share the same representation of building modelling.
- The approach combines the advantages of the flexibility and interpretability from both approaches with the shared information for supporting well-informed decision-making.
- The combined information of building environmental impact and energy performance evaluation makes trade-off analysis in the design space as an assistance for the early design phase accessible.

The remaining sections of the paper are organized as follows: Section 2 introduces example methodologies implemented in the framework in this study; Section 3 sets up a real-world case study in the early design phase scenario; Section 4 discusses the results, and Section 5 concludes the paper.

2. Methodologies for LCA and energy performance evaluation integration

The general illustration of the integrative evaluation process is presented schematically in Figure 1. In this process, buildings are represented by a digital model exhibiting parametric features based on indicators. We exclusively focus on two indicators: energy performance (heating load / total energy consumption) and embodied environmental impact (global warming potentials, GWP). By integrating approaches including data-driven (Chen & Geyer, 2022) and knowledge-based methods (Schneider-Marin, Tanja Stocker, Oliver Abele, Johannes Staudt, Manuel Margesin & Werner Lang) for prediction, uncertainty evaluation, and model interpretation, we create a shared feature list for implementing both approaches simultaneously to align with the design scenario. Eventually, indicators, uncertainties information, and model explanations are integrated to provide valuable insights for designers to conduct answers for "what-if" questions. This process of providing users with possible "assumptions" with regard to defined features as a potential design space with intervention consequences is called machine assistance (Chen & Geyer, 2022).

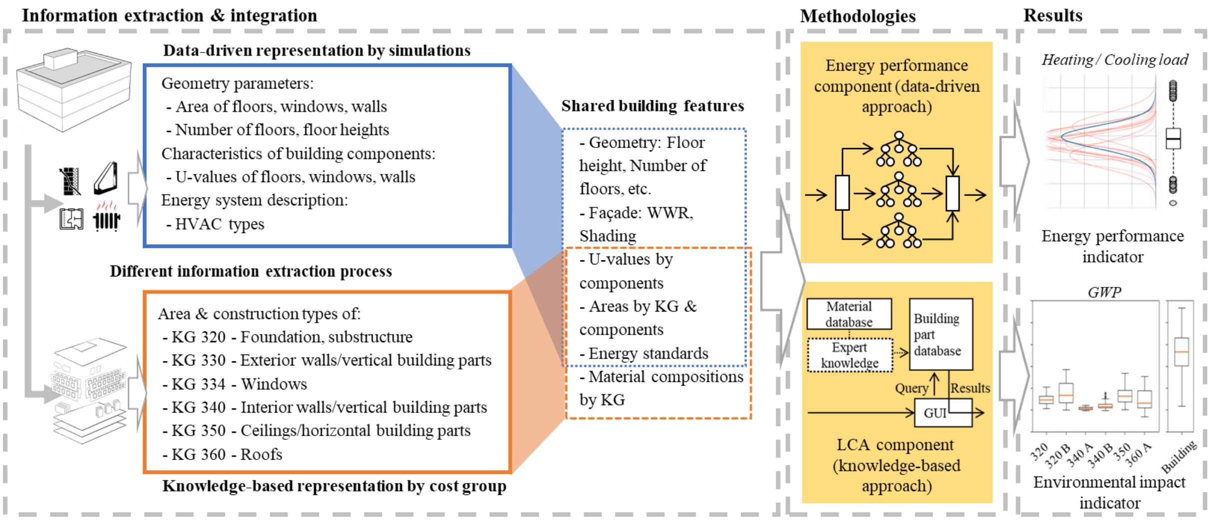

Figure 1: Performance evaluation: different approaches under the machine assistance framework

2.1 Data-driven modelling for operational energy

For the energy performance, we used a synthetic dataset generated by parametric BPS to train machine learning models to capture relationships between building features and heating load/ total energy consumption. Since we applied the framework to the building early design phase scenario, one characteristic should be necessarily included: the evaluation of information uncertainties. To represent this process, a probabilistic, tree-based surrogate model – NGBoost (Duan, Avati, Ding, Thai, Basu, Ng & Schuler, 2020) is chosen to fulfill the requirements mentioned above: Instead of generating output as a point prediction, the design of the algorithm involves the uncertainties quantification process, which provides insights into the output range within the set of feature input descriptions.

Furthermore, we consider the model explainability of the data-driven model for informed decision support. An interpretation method: SHAP (Lundberg & Lee, 2017), is integrated to analyze the feature importance and assumption consequences. Eventually, the combination improves the reusability of data-driven models by generalization and trustworthiness by explainability (Geyer, Singh & Chen, 2021). The result with the training process explanation is shown in section 3.3. For further parametric fine-tune detail and extension material, we refer to (Duan et al., 2020; Chen & Geyer, 2022).

2.2 Knowledge-based LCA evaluation methods integration

For the embodied emissions (GWP), we rely on a knowledge-based method where a database of material properties is enriched via expert input and knowledge regarding construction types and physical properties, such as thermal properties, is added to calculate results based on simple geometric models (Schneider-Marin, Tanja Stocker, Oliver Abele, Johannes Staudt, Manuel Margesin & Werner Lang). The database is fed with a multitude of options possible at an early design stage resulting in a range of outcomes. Additional modules can be integrated for cost, maintenance, repair, replacement, end-of-life, and environmental costs. This method enables users without LCA expertise to make decisions in early design phases based on reliable primary data collected in a knowledge-based decomposed LCA database ("Knowledge Database" on the basis of the publicly available database Ökobaudat (Bundesministerium des Innern, für Bau und Heimat, 2022)). This material database is enriched with expert knowledge regarding different construction types, including material composition and quantities of typical building components. Their quantities are determined by external requirements, such as energy standards or structural properties.

In this paper, we investigate different isolation standard properties as one part of material functionality in the case study. Materials are then classified according to their applicability based on the location of material and the functionality it fulfills. To organize the building elements according to their respective locations, we refer to the cost groups (Kostengruppen, KG) based on the German standard DIN 276 (Siemon, Speckhals & Siemon, 2021). In this study, we only consider KG 300 (building structure and finishes). The resulting building part properties are combined with geometric data extracted from digital models available as IFC files to predict the embodied emissions for a complete building.

3. Case study

3.1. Design scenarios

To test the described method, we used a case study of a real-world building design of a mixed-use building. The described project is called the Building.Lab on a tech campus in Regensburg, Germany (see Figure 2). The function of this 2308 sqm building is office and seminar use as well as housing. It consists of 4 above-ground stories and one underground level with a concrete skeleton structure. As we are studying an early design phase, the precise façade composition has not been determined yet. The housing areas are located in the south, and their balconies also function as passive solar protection. The larger seminar rooms are oriented to the north and can therefore be well illuminated by a high percentage of window area while avoiding overheating. The building is arranged in a U-shape around an atrium that extends across all stories.

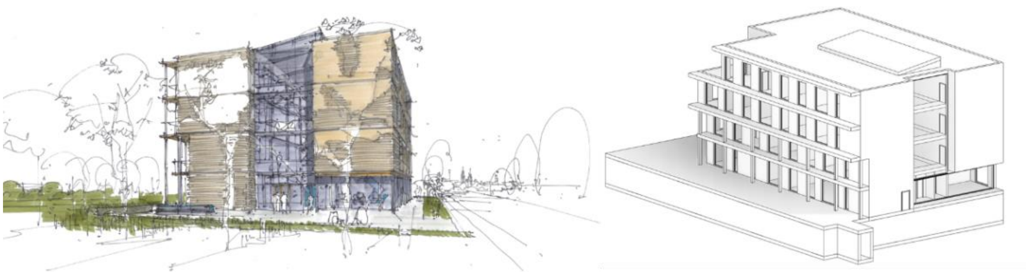

Figure 2: Building.Lab conceptual & BIM model illustration; Source: Lang Hugger Rampp GmbH / Bayerischer Bauindustrieverband e.V.

The case study examines the building design for three different window-to-wall ratios (WWR = 0.2, 0.4, 0.6), three different isolation standards (base, medium, high), and two shading options (External-Shading, Low-HGC-Value). Table 1 shows areas of the various building components for the real-world variant with a WWR of 0.4.

Table 1: Areas of building components (window-to-wall ratio = 0.4)

Cost group	Subdivision	Description	Area
KG_320	-	Foundation	966 m²
KG_330	-		
	KG_330 A	Exterior wall underground	416 m²
	KG_330 B/C	Exterior wall above ground (load bearing & non-load bearing)	901 m²
KG_334	-	Windows	600 m²
KG_340	-		
	KG_340 A	Internal wall (load bearing)	1088 m²
	KG_340 B	Internal wall (non-load bearing)	1478 m²
KG_350	-	Ceilings/Floors	1945 m²
KG_360	-	Roof (Building)	583 m²
		Roof (Garage)	369 m²

3.2. Data description

To generate data that allow well-generalizing models, we created a dataset based on the target scenario for the data-driven and knowledge-based performance evaluation. A parametric model for a generic H-shape office building has been developed that covers a wide configuration variety of building components and zones. The number of floors was set to four with an additional basement. Besides the basement, all floors have the same floor plan scheme. The modelling platform was Grasshopper (Robert McNeel & Associates, 2022). A high-level simulation interface for EnergyPlus, Honeybee (*Ladybug Tools | Honeybee)* was chosen. Figure 3 presents an illustrative sample.

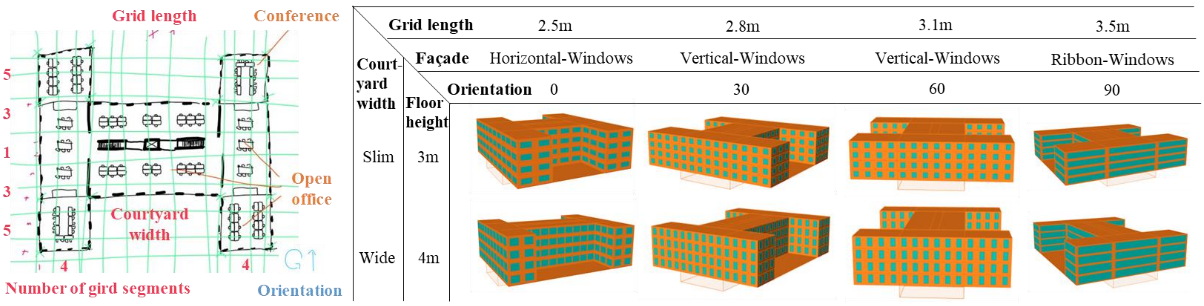

Figure 3: Floor plan scheme of generic H-shape office building and samples

For the dataset generation, the varied parameter list is shown in Table 2.

Table 2: List of input features

Parameters	Description
Standard of U-Values	Base, Medium, High
Façade	Horizontal-Windows (WWR: 0.19-0.30), Vertical-Windows (WWR: 0.20-0.46), Ribbon-Windows (WWR: 0.55-0.68)[1]
Shading	External-Shading, Low-HGC-Value
Courtyard width	Narrow – 7, Wide – 13 [number of raster segments] (17.5, 20, 22, 24.5, 30, 34, 38, 42) [m]
Floor height	3.4 [m]
Orientation	0, 30, 60, 90 [°]
Grid length	2.5, 2.8, 3.1, 3.5 [m] (Floor areas: 1150, 1400, 1800, 2250, 2700) [m²]

[1] *window-to-wall ratio*

Besides these parameters, every combination, i.e., the full factorial, has been simulated, which resulted in 1152 simulated samples. Construction and HVAC system were kept constant. For the construction, an external thermal insulation composite system (ETICS) was modelled with a massive generic material with a heat capacity of 1100 J/kg·K and 900 J/kg·K for the slabs. Furthermore, we defined three sets of u-values which would scale the thickness of the thermal layer accordingly. In this context, we selected three well-accepted energy standards in Germany and referred to u-values requirements as three isolation categories (base, medium, and high), as shown in Table 3.

Table 3: U-value requirements under different isolation standards [W/m²K.]

Standard of U-Values	Base: GEG (2020 German Energy Act for Buildings)	Medium: NZEB (Net Zero Energy Building)	High: Passive House
- Base plate	0.2625	0.206	0.15
- Roof	0.15	0.135	0.12
- Exterior wall, bearing, above ground	0.21	0.18	0.15
- Exterior wall, bearing, under ground	0.2625	0.206	0.15
- Window	0.975	0.888	0.8

The HVAC system is modelled with an ideal load air system template from EnergyPlus as it provides good comparability of loads at an early stage of the design. For zone programs, the open office area was modelled with 0.057 people/m² and conference areas with 0.053 people/m². As for shading mechanism, either windows with a solar-heat-gain-coefficient (SHGC) of 0.5 and an additional shading layer with the reflectance of 0.5, and transmission of 0.4 will be activated as soon as zone temperatures rise above 25°C or windows with a low SHGC of 0.3 have been modelled.

3.3. Model training

The surrogate model (NGBoost) consists of a set of Classification and Regression Trees (CART), which requires the input features in the form of integer or float (Loh, 2011). Only semantic features in the dataset require label-encoding by transferring descriptions into numerical categories. In this context, three features require feature engineering: *IsolationStandard*, *Façade*, and *Shading*.

The training process randomly splits the dataset into training (80%) and test (20%) sets. We set input features as in Table 2 with the prediction target of building heating load and total energy consumption. Since the dataset is relatively simple, models are trained by the default setting of hyperparameters (Schuler, 2020).

For the performance evaluation, we selected the most typical metrics in the BPS domain (Vogt, Remmen, Lauster, Fuchs & Müller, 2018) as well as in machine learning regression prediction tasks: Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), Normalized Root Mean Square Error (NRMSE) and R-squared (R2). Table 4 presents the model performance based on the test set.

Table 4: Accuracy metrics of model result

	RMSE	MAPE	NRMSE	R2
Heating Load	0.3246	0.2529%	0.8137	0.9986
Total Energy Consumption	0.4450	0.2091%	1.0868	0.9971

The result shows a decent performance for the data-driven approach, the average errors of both models are lower than 1%. More specifically, we see that the performance of the heating load model is better than the total energy consumption prediction. The reason behind it is intuitive:

total energy consumption depends on more implicit factors, e.g., cooling, equipment, and lighting load. To sum up, the surrogate model based on feature representation in Table 2 is capable to capture the building energy performance accurately.

Additionally, NGBoost as the surrogate model provides the output as a set of Gaussian distribution parameters: *loc* and *scale,* which stands for mean (point prediction) and standard deviation (uncertainty range), respectively. Figure 4 and Figure 5 illustrate how different input features impact point output and uncertainty range in the task of heating load and total energy consumption prediction.

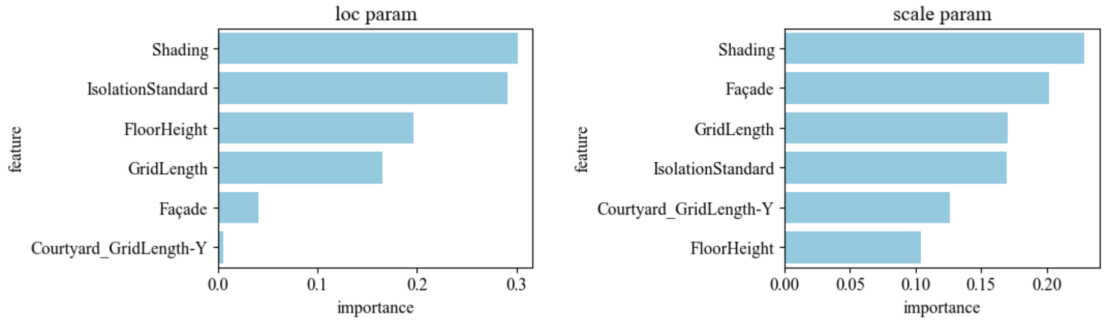

Figure 4: Feature importance for heating load prediction

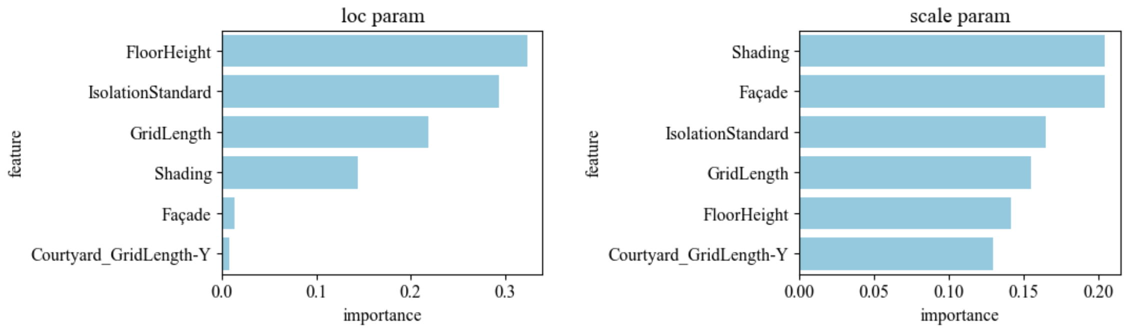

Figure 5: Feature importance for total energy consumption prediction

For heating load prediction, Figure 4 shows that the design of shading, isolation standard choices, and floor height affects the point prediction (mean value of the output) the most, followed by building geometry, especially the internal wall area. For output uncertainties, options of shading, façade, and grid length have the most effects in the uncertainty range (standard deviation of the output). Similar feature importance is also observed for the task of total energy consumption prediction (Figure 5) with regard to the question, of which shading option and building façade features should gain the most attention in consideration of building energy performance.

3.4. Results

For the building operational energy, the prediction presents regular patterns aligned with isolation standard and WWR iterations, as shown in Figure 6: The heating load varies between 80 and 115 kWh/m^2 per year, which increases with higher WWRs and decreases with higher isolation standards. The same patterns are observed in total energy consumption as well. The consumption increases with higher WWR with the option of low-HGC-value strengthening this trend. Compared with heating load, it contains more factors: Intuitively, we observed that with the higher isolation standard, external shading, and lower WWR, the total energy consumption drops accordingly.

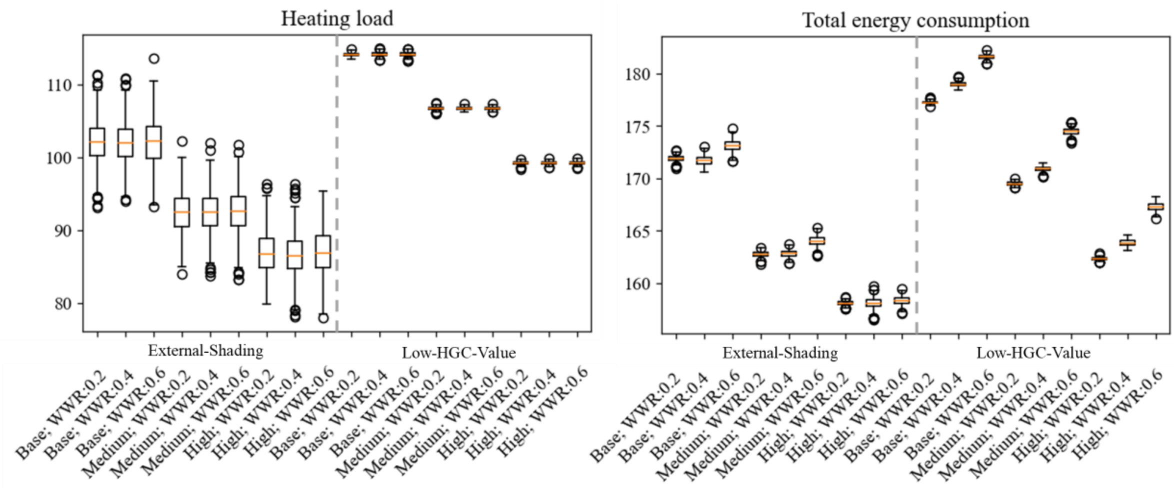

Figure 6: Heating load and total energy consumption prediction under different isolation standard, window-to-wall ratio (WWR), and shading combinations

Some interesting quantitative insights are worth mentioning from Figure 6: In heating load prediction, the range of energy consumption is partially overlapped between adjacent insulation standards. Compared to the isolation standard difference, the impact of different WWRs is relatively small. The reason is the assumption of state-of-the-art glazing and shading that allows windows to level out higher heat transfer losses by solar gains without overheating in summer. Such information provides designers with valuable benchmarks and alternative scenarios, enabling them to involve other factors (cost, CO2 emissions, etc.) for further decision support.

For the embodied emissions, the predictions show that the GWP increases for higher isolation standards as shown in Figure 7. The increase in GWP related to improved isolation standards is relatively small compared to the overall embodied emissions. For the given concrete construction type, the predictions for three WWRs show that the GWP decreases with the increasing ratio of windows with the option of low-HGC-value.

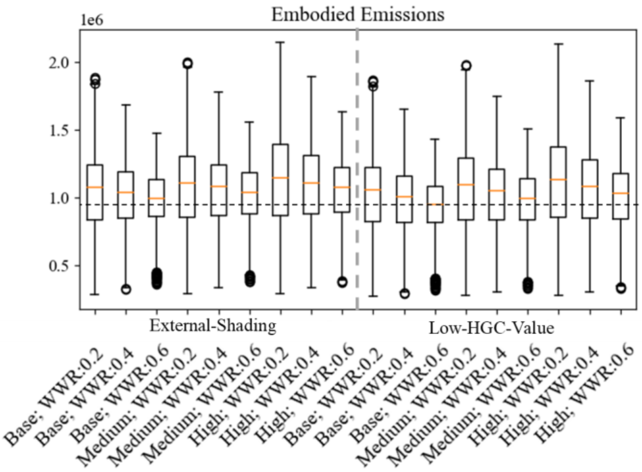

Figure 7: Embodied emissions (GWP) prediction under different isolation standard, window-to-wall ratio (WWR), and shading combinations

The simultaneous presentation of operational energy and embodied emissions predictions allows designers to evaluate parameters such as isolation standards, WWR, and shading options at early stages and adjust their design accordingly for further development and trade-off analysis. Based on these predictions, we recognized that the increase of embodied emissions for a higher isolation standard is comparatively small compared to the impact of emissions resulting from operational energy.

3. Discussion

In this paper, we explored the integration path of data-driven methods and knowledge-based methods for building design machine assistance. The key pillar of the integration depends on the connection between the available data description (features) and the knowledge-based methods' representation.

From the result of the case study, we generated alternative design scenarios in different isolation standards & window-to-wall ratio (WWR) combinations with shading options and evaluated energy performance and embodied impact. This combined information and required data are accessible in the early design phase. Both approaches contain the potential to further enhance the ability of general assistance scenarios in future research: For data-driven methods, involving, e.g., a component-based modelling process (Geyer et al. 2018) or hybrid-model approach (Chen, Guo & Geyer, 2021) would provide further insight into the energy performance at the building component level with better modelling flexibility and interpretability. For knowledge-based methods, based on the trade-off analysis, indicators to represent building cost factors would enhance the machine assistance practicability.

To explore the path of data-driven method integration into domain knowledge-based approach in general scenarios, it is vital to invest efforts in aligning representations to bridge exposed features from digital models and knowledge. For example, to provide designers with useable feedback regarding the overall life cycle impact of the design decisions taken, the operational energy results would have to be converted to GWP based on realistic primary energy scenarios. In this context, the integration between data-driven methods and knowledge-based methods is not only necessary in the feature representation process, but also vital for output interpretations.

Finally, although we particularly checked the availability of data during the early design phase to ensure practicality, the evaluation of information utility has not been addressed, especially when generated suggestions are inter-disciplinary. From this perspective, the investigation of the designers' feedback based on the integrated decision support is, in our opinion, worth further research. A related study on user effect has been carried out for the component-based prediction of operational energy planned to be repeated, including the new methods (Singh, Deb & Geyer, 2022).

4. Conclusion

This study presents a step toward a machine assistance framework for the building design process that achieves multi-objective decision support. This framework serves as augmented intelligence to accelerate the digitalization process of user decision-making assistance in domains that require not only data-driven support but also knowledge-based analysis, such as early building design. The structure organized by components and modules allows for the generalization of data-driven models aligned to domain knowledge. Moreover, this approach forms a basis for data and knowledge integration from which the design process will immensely benefit. The integration allows for involving operational energy and embodied environmental impact at an early design phase and raises the reconsideration of the design process toward the objective of sustainable development in our domain.

Acknowledgement

We gratefully acknowledge the support of the German Research Foundation (DFG) for funding the project under grant GE 1652/3-2 in the Researcher Unit FOR 2363. We would like to thank Lang Hugger Rampp GmbH / Bayerischer Bauindustrieverband e.V. for data resources support.

References

Amasyali, Kadir, & El-Gohary, Nora M. (2018). A review of data-driven building energy consumption prediction studies, *Renewable and Sustainable Energy Reviews*, vol. 81, pp. 1192–1205

Bélisle, Eve, Huang, Zi, Le Digabel, Sébastien, & Aimen E. Gheribi. (2015). Evaluation of machine learning interpolation techniques for prediction of physical properties, *Computational Materials Science*, vol. 98, pp. 170–177

Bundesministerium des Innern, für Bau und Heimat. (2022). ÖKOBAUDAT, Available online: https://www.oekobaudat.de/en.html [Accessed 2 February 2022]

CAALA. (2019). Die CAALA Software ist ein umfassendes Werkzeug zur energetischen Vordimensionierung und Ökobilanzierung - CAALA, Available online: https://caala.de/features [Accessed 15 November 2021]

Chen, Xia, & Geyer, Philipp. (2022). Machine assistance in energy-efficient building design: A predictive framework toward dynamic interaction with human decision-making under uncertainty, *Applied Energy,* vol. 307, p. 118240

Chen, Xia, Guo, Tong, & Geyer, Philipp. (2021). A hybrid-model time-series forecasting approach for reducing the building energy performance gap, EG-ICE 2021 Workshop on Intelligent Computing in Engineering

Duan, Tony, Avati, Anand, Ding, Daisy Y., Thai, Khanh K., Basu, Sanjay, Ng, Andrew Y., & Schuler, Alejandro. (2020). NGBoost: Natural Gradient Boosting for Probabilistic Prediction, Available online: http://arxiv.org/pdf/1910.03225v4

Geyer, Philipp, Singh, Manav M., & Chen, Xia. (2021). Explainable AI for engineering design: A unified approach of systems engineering and component-based deep learning, *ArXiv Preprint ArXiv:2108.13836*

Harter, Hannes, Singh, Manav M., Schneider-Marin, Patricia, Lang, Werner, & Geyer, Philipp. (2020). Uncertainty analysis of life cycle energy assessment in early stages of design, *Energy and Buildings,* vol. 208, p. 109635

Hensen, Jan L., & Lamberts, Roberto. (2019). Building Performance Simulation for Design and Operation, Available online: https://research.tue.nl/en/publications/building-performance-simulation-for-design-and-operation-expanded

Hollberg, Alexander, Tschetwertak, Julia, Schneider, Sven, & Habert, Guillaume. (2018). Design-integrated LCA using early BIM, in *Designing Sustainable Technologies, Products and Policies*, Springer, Cham, pp. 269–279

Jia, Yongna, & Ma, Jianwei. (2017). What can machine learning do for seismic data processing? An interpolation application, *Geophysics*, vol. 82, no. 3, V163-V177

Ladybug Tools | Honeybee, Available online: https://www.ladybug.tools/honeybee.html [Accessed 3 February 2022]

Loh, Wei-Yin. (2011). Classification and regression trees, *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery,* vol. 1, no. 1, pp. 14–23

Lundberg, Scott M., & Lee, Su-In. (2017). A Unified Approach to Interpreting Model Predictions, Advances in Neural Information Processing Systems, Available online: https://proceedings.neurips.cc/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf

One Click LCA® software. (2022). Automate Building Life Cycle Assessment with One Click LCA, Available online: https://www.oneclicklca.com/construction/life-cycle-assessment-software/ [Accessed 2 February 2022]

Østergård, Torben, Jensen, Rasmus L., & Maagaard, Steffen E. (2016). Building simulations supporting decision making in early design – A review, *Renewable and Sustainable Energy Reviews*, vol. 61, pp. 187–201

Østergård, Torben, Jensen, Rasmus L., & Maagaard, Steffen E. (2017). Early Building Design: Informed decision-making by exploring multidimensional design space using sensitivity analysis, *0378-7788*, vol. 142, pp. 8–22, Available online: https://www.sciencedirect.com/science/article/pii/S0378778817306916 [Accessed 23 February 2022]

Robert McNeel, & Associates. (2022). Grasshopper - New in Rhino 6, Available online: https://www.rhino3d.com/en/6/new/grasshopper/ [Accessed 3 February 2022]

Schneider-Marin, Tanja Stocker, Oliver Abele, Johannes Staudt, Manuel Margesin, & Werner Lang. Re-structuring of eco-data to include functional criteria: using data enhanced with knowledge in early phases of design, (unpubl.)

Schneider-Marin, Patricia, Harter, Hannes, Tkachuk, Konstantin, & Lang, Werner. (2020). Uncertainty analysis of embedded energy and greenhouse gas emissions using BIM in early design stages, *Sustainability*, vol. 12, no. 7, p. 2633

Schneider-Marin, Patricia, & Lang, Werner. (2020). Environmental costs of buildings: monetary valuation of ecological indicators for the building industry, *The International Journal of Life Cycle Assessment*, vol. 25, no. 9, pp. 1637–1659

Schuler, Alejandro. (2020). NGBoost User Guide, Data 100 at UC Berkeley, 26 October, Available online: https://stanfordmlgroup.github.io/ngboost/1-useage.html [Accessed 28 February 2022]

Seyedzadeh, Saleh, Rahimian, Farzad P., Glesk, Ivan, & Roper, Marc. (2018). Machine learning for estimation of building energy consumption and performance: a review, *Visualization in Engineering*, vol. 6, no. 1, pp. 1–20

Siemon, Klaus D., Speckhals, Raphael, & Siemon, Anna. (2021). Kostengruppen nach DIN 276/2018, in Klaus D. Siemon, Raphael Speckhals & Anna Siemon (eds), *Baukostenplanung und -steuerung*, Wiesbaden: Springer Fachmedien Wiesbaden, pp. 219–240

Singh, Manav M., Deb, Chirag, & Geyer, Philipp. (2022). Early-stage design support combining machine learning and building information modelling, *Automation in Construction*, vol. 136, p. 104147

Theißen, S., Höper, J., Wimmer, R., Zibell, M., Meins-Becker, A., Rössig, S., Goitowski, S., & Lambertz, M. (2020). BIM integrated automation of whole building life cycle assessment using German LCA data base ÖKOBAUDAT and Industry Foundation Classes, IOP Conference Series: Earth and Environmental Science

Thessen, Anne. (2016). Adoption of machine learning techniques in ecology and earth science, *One Ecosystem*, vol. 1, e8621

Vogt, Marcus, Remmen, Peter, Lauster, Moritz, Fuchs, Marcus, & Müller, Dirk. (2018). Selecting statistical indices for calibrating building energy models, *Building and Environment*, vol. 144, pp. 94–107

Westermann, Paul, & Evins, Ralph. (2019). Surrogate modelling for sustainable building design – A review, *Energy and Buildings*, vol. 198, pp. 170–186