Capturing building data for establishing digital twins of buildings for quick energy performance assessment

Deepti Desai, (desai@contecht.eu)
Contecht GmbH

Timo Hartmann, (hartmann@contecht.eu)
Contecht GmbH

Keywords: digital twin, building performance, simulation, modeling, case study

Abstract

Operating buildings in an optimized manner can be supported by digital twins of a building with the possibility to digitally model and predict the energy use of the building. Generating such digital twins, however, requires a large effort in terms of capturing data from the existing building. Moreover, even if owners or operators are willing to undertake such an investment, capturing the required building data is often not possible as capturing technologies might interfere with the occupants of the building. In this paper, we propose a questionnaire based process to quickly and non-intrusively collect information about a building to allow for an initial energy assessment. We introduce the questionnaire to collect data and show how the questionnaire data can be converted to input for a building performance simulation model. To validate the questionnaire, we benchmark simulated results based on models of eight different buildings across Europe. For each of the buildings we compared the results of building performance simulations of simple models generated from questionnaire data with detailed models generated from BIM models. The results show that simulated deviations are between 0.75% to 40.57% between the two types of models. We conclude that the questionnaire based approach can be a reasonable starting point for first quick energy assessments for most types of buildings.

1. Introduction

Simulating building energy performance is inherently complex. Simulation models need to accurately represent dynamic systems combining complex thermal interactions between building components, such as walls, windows, and roofs. To this end, simulation models detail transfer processes of conduction, convection, and radiation. At the same time, heating, ventilation, and air conditioning (HVAC) systems and occupant behavior need to be considered. To instantiate these complex physical process models for a specific building, engineers need to gather and validate detailed data integrating interdisciplinary knowledge about a specific building (Clarke, 2007). With this complexity in building physics models, comes a complexity in input information for the simulation models.

Because of this complexity in the input that is required, it is not surprising that often there is a large difference between the predicted energy performance and the actual energy performance of a building - something referred to as the building energy gap (for a well written recent review refer to, for example, (Zou et al., 2018)). The building energy gap can be significant, according to some, the actual energy consumed in building can be up to 2.5 times more then initially simulated (Zou et al., 2018).

With the advent of complex building operation systems and digital twin solutions, the problem of the building energy performance gap also become prevalent for building operation managers (Liu et al., 2024). The core functionality of these systems and solutions is it to predict possible future states of a building so as to adjust the building operations to optimize energy use for an upcoming period. Accurate simulations of future states are key to such an optimization. In current practice, operation managers of buildings lack detailed information and understanding about the true energy performance of their buildings. Setting up simulation models that form the basis of digital twins is a cumbersome, labor-intensive, and often a disturbing process towards the operations the building is supported (Lu et al., 2020). Therefore, it is currently not possible for most building operators to generate digital twin models and to obtain accurate predictions of future building states so as to optimize building operations.

As data collection for existing buildings is cumbersome, we propose a simplified data capturing and energy performance modeling process in this paper. We suggest that instead of collecting data through a detailed inspection and engineering survey of a building, data could be collected through a short questionnaire only, at least as a starting point for an initial building performance model. This would reduce the effort that is required to capture data and the disturbance of occupants during capturing. In this paper, we present the results of a study show that the accuracy of energy performance models generated with an exemplary questionnaire tool that was developed in a recent European innovation project. Our study predicting the energy performance of seven buildings in four different European countries show that the simplified process captures energy performance of buildings within an average of 20% of what a detailed energy simulation model could predict.

This paper is structured as follows, in the next section we briefly describe current efforts to design digital twin technologies for twinning building energy performance. The section also briefly summarizes the main input information that is required for energy simulation models that form the basis of such digital twin applications. Afterwards, we explain our proposed questionnaire based method and describe how we tested it on a number of demonstration buildings in Europe. After presenting the results, we close the paper with a discussion of implications and limitations of our study.

2. Building Performance Simulations and Digital Twins

Digital twins are virtual representations of physical entities to simulate their real-world counterpart drawing upon continuously collected data of the behavior of the entity (Grieves, 2014). To this end, the digital twin integrates data from various sources to create a comprehensive, dynamic model that reflects the current state and behavior of the physical entity. The representation of the current state then allows to predict future states of a building accounting for different weather conditions, occupant behavior, and building system operations. These predictions can provide important information about how to best operate a building most effectively (Arowoiya et al., 2024) (Tan et al., 2022). While much of the focus in current digital twinning for buildings is on the collection and representation of the buildings status using advanced sensor systems, a suitable simulation model is also required to make use of the data with respect to predicting future states of the entity (Delgado & Oyedele, 2021). For buildings this requires a first digital representation of the building, in terms of its geometry and materials that can feed into a building performance simulation that uses simulation software to perform a detailed analysis of a building's energy use and energy-using systems. These simulation software

works by enacting a mathematical model that provides an approximate representation of the building operational behavior in terms of energy use and indoor climate.
Setting up such simulation models require a varied set of information input that needs to be provided. Figure 1) summarizes the different types of information that is required in terms of a building's geometrical properties, a building's materials, a building's heating, ventilation, air-conditioning, lighting, and power systems, but also about the environment of the building and the operations of the building.

3. Information required for building performance simulation

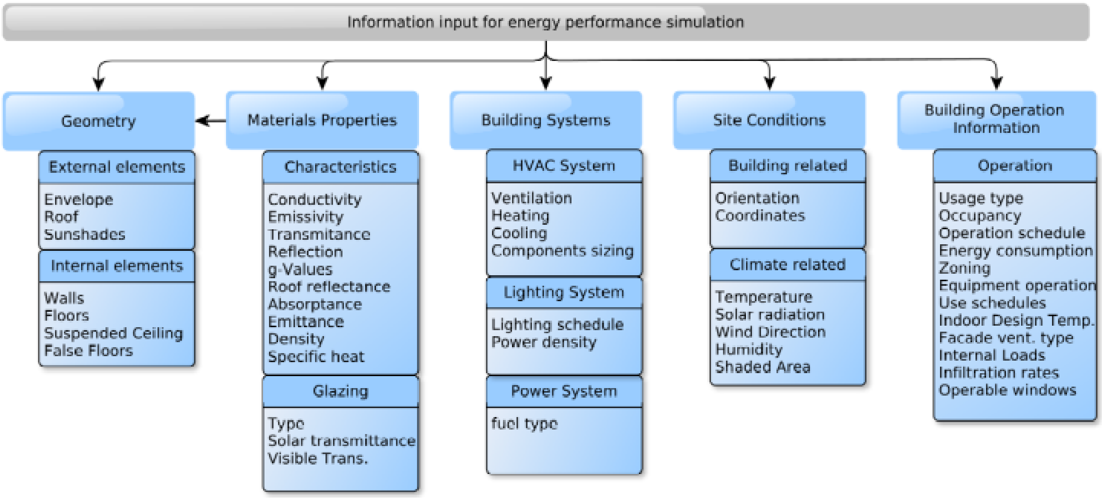


Figure 1: Information required for a building performance simulation (Gutsche & Hartmann, 2017)

Collecting all the information depicted in Figure 1 that is required is a cumbersome task. While geometrical information can usually be obtained with the help of state-of-the art surveying methods, such as laser scans (Chen et al., 2023), obtaining accurate material data would require more complex measurements (Ham & Golparvar-Fard, 2013). Additionally, information about the building's mechanical, electrical, and plumbing systems is required (Utkucu et al., 2024). Collecting all this information is expensive and often disrupts the operations of the building. The question arises whether it is required to undertake such an extensive data collection effort to establish the basis for a meaningful digital twin or whether simpler models that rely on much less accurate and fewer data, and much more on general assumptions, could be an equally suitable starting point. To shed light on this question, we tested a minimalistic data collection approach that relies on a short questionnaire developed in an European innovation project, inquiring about some main characteristics of a specific building (Desai & Hartmann, under review). We then tested the method by comparing simulation models based on the questionnaire with detailed simulation models generated from a careful assessment of seven buildings in Europe. The next section described the overall structure of the minimalistic questionnaire method.

4. A Proposed Survey Instrument for Quick Data Collection

The developed approach proposes to collect initial data about a building using a questionnaire that collects main data required by the building performance simulation (Desai & Hartmann, under review). This data is enriched with information from a standards library on materials, schedules, and equipment. This allows to conduct a first simulation

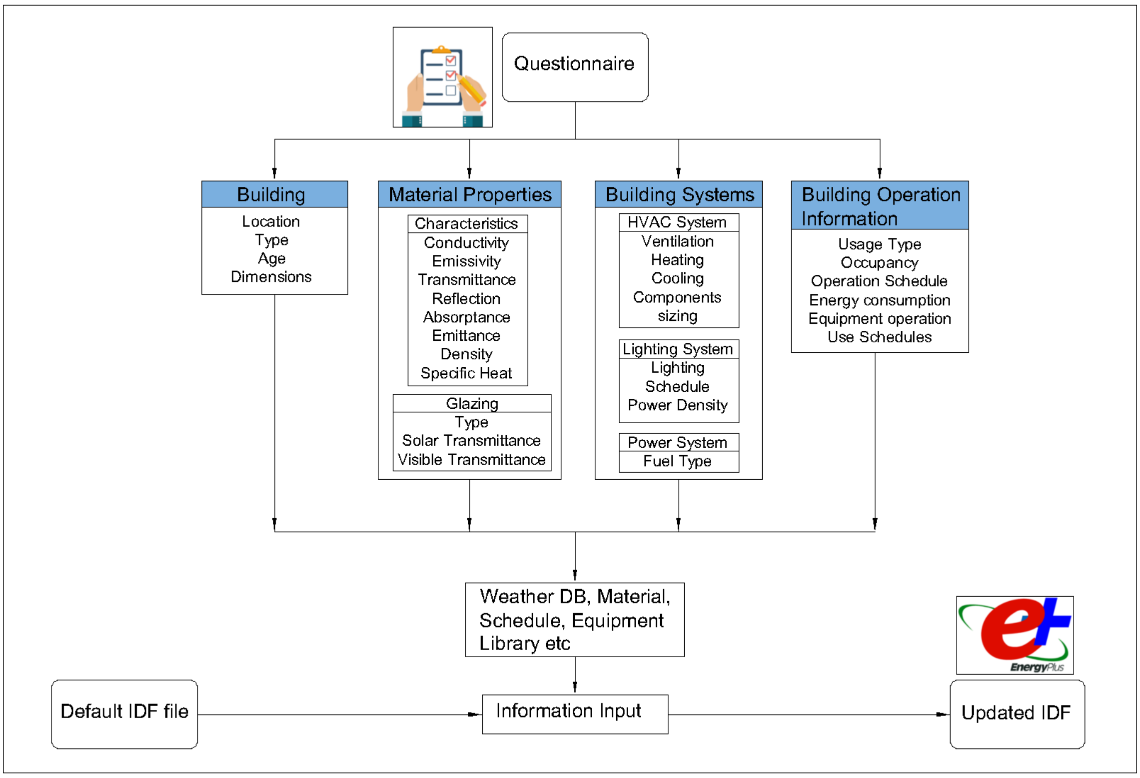


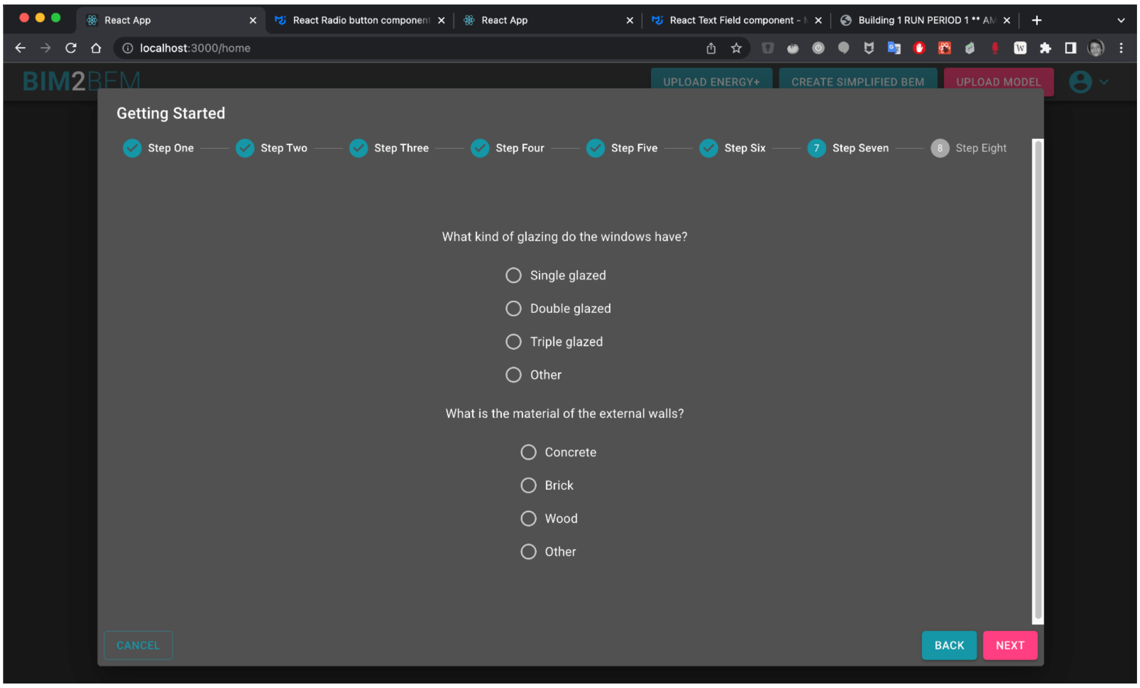
Figure 2: Overview of our proposed questionnaire approach.

of the building energy performance accounting for local weather data. Figure 2 summarized this approach. The rest of this section describes the questionnaire and how the input of the questionnaire is converted to an energy simulation model. Detailed information and conversion algorithms can be found in (Desai & Hartmann, under review). Figure ?? shows the implementation of the questionnaire in a web-supported software that was realized during the Eu innovation project that funded the effort to develop the questionnaire.

Designed to obtain all required data for running a simple building performance simulation, the questionnaire asks for a variety of information about the building such as (the HVAC system, schedule information, building materials, etc.) using a number of different sections.

The location section obtains data on the building's geographic location and geometrical shape to ensure the simulation of the precise local weather conditions that affect the building. The purpose of the building's section is to obtain the type, function, and age of the building, as well as various information about its location and position within its environment. This section addresses inputs that are relevant to the function, program, and condition of a building.

The well-being and users experience sections are about the type, behaviour, and experience of the occupancy as this part is essential to design a schedule and occupancy rate adapted to the original building. This part also provides an opportunity to assess the smartness of the building and the robustness of the simulation model output data related to comfort according to occupants' answers. The energy performance section retrieves the building's energy effectiveness information and provides a link between the efficiency of smart device deployment and the energy efficiency of the building. The structural section provides information required for the building's materials as this data



Getting Started

What kind of glazing do the windows have?

○ Single glazed
○ Double glazed
○ Triple glazed
○ Other

What is the material of the external walls?

○ Concrete
○ Brick
○ Wood
○ Other

Figure 3: Implementation of the questionnaire in a web platform.

is the core of the building's interaction with the outside environment. Finally, the facility part includes detailed information on the building's equipment, such as HVAC, hot domestic water systems, appliances, and power generating systems.

Using the information from the questionnaire our approach then suggests to convert the information into input for an energy simulation software, in our case into the IDF file input format of the widely used building performance simulation kernel EnergyPlus (of Energy, 2024). In the following we sketch this conversion. The component of an IDF file as divided into groups for instance site location is an object of the group "Group – Location – Climate – Weather File Access". The groups and components of an IDF file are:

- Site: Location - as stated before, this object belongs to an upper group, the Location class describes the parameters for the building's location which is essential to determine the weather data. The location data is directly obtained from the questionnaire.

- Building - the Building object belongs to the group "simulation parameters" it describes parameters that are used during the simulation of the building.

- Zone list, Zone, Zone control, and Building surface – this belongs to the group "Thermal Zone Description/Geometry" Without thermal zones and surfaces, the building can't be simulated. This group of objects (Zone, Building Surface) describes the thermal zone characteristics as well as the details of each surface to be modelled. Included here are shading surfaces. Moreover, the Zone List object defines a list of Zone objects, and Zone Control objects are used to control zone conditions to a specific set point. The questionnaire collects data about the overall shape and the number of storeys of the buildings. This data is then concerted to one thermal zone per building level.

- Schedule - this group of objects allows the user to influence the scheduling of many items (such as occupancy density, lighting, thermostatic controls, and occupancy activity). In addition, schedules are used to control shading element density on the building. The questionnaire collects information about the general purpose of the building from which assumptions about occupancy behavior are derived for the simulation input.

- People, Lights, ElectricEquipment and GasEquipment, and HotWaterEquipment are under the group “ Internal gains” this group describes the internal gains that influence the energy consumption other than an envelope or ambient conditions. Again coarse assumptions are made from the initial questionnaire input to model general building systems for different types of buildings as indicated in the questionnaire.
- Construction and Materials are associated with the group - Surface construction elements, this group of objects describes the physical properties and configuration of the building envelope and interior elements. That is the walls, roofs, floors, windows, and doors of the building. The questionnaire also collects initial information about the building’s materials that can be used in relation to a material database to assign the required details to the surfaces in terms of thermal properties, such as heat transfer coefficient or convection characteristics.
- Zone Equipment is a group that describes several blocks of zone equipment such as ZoneHVAC: AirDistributionUnit, EquipmentConnections, and ZoneHVAC:EquipmentList.

As already mentioned in the brief summary above, as the questionnaire can only provide some initial and simple information, a number of IDF classes are kept as default – Version, Simulation control, Building (only name would be used as an input), Shadow calculation, Heat Balance Algorithm, Timestep, Run period, Daylight saving time, and Global geometry rules.

The script uses the questionnaire data for the building’s location and downloads the climate file (EPW) and the design day file(DDY) for the building’s site from the EnergyPlus website to replicate the appropriate climate conditions that influence the building. The Latitude and Longitude of the Site are taken from the Questionnaire inputs and are entered into the IDF.

While the questionnaire has been implemented in a prototypical software that directly converts the input to an IDF input file, triggers a building performance simulation, and feeds the results of the simulation back to users, it is not clear how accurate such questionnaire based simulation can be. Therefore, we set-up an experiment to test the accuracy. This experiment is described in the next section.

5. Validation: Case Study on Multiple Buildings in Europe

We validated the simplified energy simulations using four buildings in different European states: the Netherlands, Germany, Greece, and Spain. The Dutch demonstration are two buildings in a recently constructed residential building complex that includes 18 NOM (Zero Energy) 60 m2 apartments (three rooms) for social renting. Each unit is energy-efficient, with high-quality insulation, ventilation equipment, and a floor heating system linked to a geothermal heat pump.

The German demonstration site is a three-story residential/multifamily building in Velten. The building was constructed in 1907 and comprises six flats totalling 335 m2. The building is situated in an Oceanic Climatic, which is the major climate type throughout much of Western Europe.

Table 1: Results of the validation comparing energy simulation results of models generated from the questionnaire with detailed energy models developed on the basis of careful building inspection and survey activities (Yearly energy consumption per square meter of building surface).

Pilot	Buildings	simplified model [kWh/m2]	detailed model [kWh/m2]	Difference [%]
Netherlands	Building A	108.51	77.19	40.57
	Building B	106.33	94.63	12.36
Germany		141.95	125.93	12.72
Greece	Building 1	134.59	107.86	24.78
	Building 2	114.96	115.83	0.75
Spain	Building 1	96.90	70.57	35.31
	Building 2	95.88	81.25	18

The Greece demonstration sites consist of six apartments in two different building complexes. These apartments vary in their construction years between 1950 and 2000 with an area totalling 120 m2 to 200 m2. These buildings are situated in a Hot summer Mediterranean climate typical to that part of Greece and were selected for their diverse consumption and occupancy profiles.

The Spain pilot comprises several independent buildings in a typical Spanish neighborhood. Each of the buildings is eight stories high and hosts 32 apartments, a reception, and an underground parking lot. These Residential units were built in 2006 and they cover a total area of 6,500 m2. Each building has a total area of 4,500 m2 with each apartment of an approximate area of 90 m2. These buildings are situated in a Semi-arid(Steppe) climate typical for that part of Spain.

For each of the buildings, we first ask the building operators to complete the questionnaire and generated energy simulation models using the developed software platform. We then also embarked on a detailed inspection of the buildings, including laser scans and detailed material surveys to capture accurate data describing the buildings geometry, materials, and HVAC systems. We used this information to generate detailed energy simulation models.

In a final step, we simulated the building energy performance using the questionnaire based models and the detailed models. We extracted the simulated energy consumption per year and square meter to compare the simulation results. The next section describes the results in detail.

6. Results of the Validation

Table 1 summarizes the results of the simulations in terms of yearly energy use per square meter. The results show that there is quite some variance in the difference between the simplified simulations and the detailed simulations for the various buildings ranging from 40.57% deviation to as little as 0.75% deviation. The average deviation between the simple and the detailed simulations is about 20%. The results also show that the simplified models tend to overestimate the yearly energy use compared to the simple model. In the next section, we will discuss these results in detail.

7. Discussion

If nothing more the results of our test shows that a minimized questionnaire based approach can provide simulation results that are within the difference that would be expected by the experiences of the building performance gap. The largest difference in simulation results is 40.57 %, while many of the other buildings show a much closer difference between the two simulation methods. For one building the difference is negligible with only 0.75 %.

These results are surprising in that the common sense reaction on the building performance gap was to further increase the accuracy of the simulation models with respect to representing a building's geometry and material parameters. In that sense our results are counter-intuitive and might point towards the utility of more simpler models in terms of providing indications of the energy use of buildings.

Saying this our study can only serve as an initial indication. After all simulating energy use of buildings should only be used in a context of carefully conducted simulation experiments. These experiments should aim at comparing different solutions to improve the energetic behavior of a building. Other than providing accurate indications about the "real" energy use of a building, often it is similarly important to be able to identify trends to compare different possible interventions. Equally important is that carefully conducted simulation experiments can provide information about driving factors and possible bottlenecks and so point towards meaningful possible improvements. In our simple comparative study we only compared the annual energy consumption. Future research needs to report and compare more detailed energy simulation result outputs, such as performance on peak days, performance of different thermal zones, and key operational characteristics of building systems - all aspects that are also provided as simulation output, but were not yet examined.

Future research should conduct studies to more closely compare the results of this study with other existing approaches, such as simplified assessment methods common within current norms, regulations, and energy certificates. Sensitivity studies should also be executed that can look at the impact of different input categories on the simulation results. Future research should also explore which KPIs can be meaningfully generated from the energy simulations to support the work of building operations professionals. The single value of energy usage used for teh comparison of the simple with more complex models is not suitable yet for supporting practitioners, different possibilities to provide value ranges and meaningful key performance indicators need to be explored.

Nevertheless, we believe that simple modeling methods as the one we propose here can serve as a meaningful starting point for professionals responsible for managing building operationssetting up digital twin systems of buildings cost efficient and with little disruption of operations. Such an approach also allows for the quick set-up of digital twin systems for larger building portfolios of property owners. Because the system operates already using a whole building simulation application and not one of the simplified assessment methods, these first starting points can then be iteratively improved over time with information from more detailed building inspections. Moreover, the initial models can be iteratively calibrated with building performance data collected from the ongoing operations of the building. Future research needs to explore how well simplified models would lend themselves for ongoing calibration activities, such as the approaches discussed in (Brouns et al., 2016), (Koo & Yoon, 2024), or (Yang & Becerik-Gerber, 2015). If these first initial models can then be matched with ongoing collected data about processes, but also energy usage, the models can be iteratively refined and improved already. At later stages, carefully conducted building inspections can then further enrich the initial models in an iterative step.

8. Conclusion

This paper reports on a study to compare building performance simulation results generated by simple models with the results generated by more detailed models. The simple models of the study used a questionnaire based approach to collect input data and an automated conversion script that converts the questionnaire data to the input for the simulation software EnergyPlus. The study shows that the differences are in the range between 0% and 40% difference. A difference that is very much in the range of what is to be expected in simulation accuracy concerning the building performance gap.

While our first tests show possible potential for using simple data collection approaches, important questions still remain unanswered. Most importantly is to understand better whether simulation results from simplified models provide good indications for how to improve the energy performance of a building. Saying this, considering the results, the questionnaire based approach considered here seems to be potentially valuable. At least as a quick, first starting point for generating the required simulation models for first versions of digital twin applications that can then be iteratively improved by model calibration exercises and more detailed building inspections. .

Acknowledgements

The results presented here were developed as part of projects that have received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 958284. This study is a part of the dissemination activities of the research projects "Boosting Research for a Smart and Carbon Neutral Built Environment with Digital Twins (SmartWins)" (Grant ID Number 101078997), funded under the Horizon Europe call HORIZON-WIDERA-2021-ACCESS-03-01.

References

Arowoiya, V. A., Moehler, R. C., & Fang, Y. (2024). Digital twin technology for thermal comfort and energy efficiency in buildings: A state-of-the-art and future directions. *Energy and Built Environment*, *5*(5), 641–656.

Brouns, J., Nassiopoulos, A., Bourquin, F., & Limam, K. (2016). Dynamic building performance assessment using calibrated simulation. *Energy and Buildings*, *122*, 160–174.

Chen, S., Fan, G., & Li, J. (2023). Improving completeness and accuracy of 3d point clouds by using deep learning for applications of digital twins to civil structures. *Advanced Engineering Informatics*, *58*, 102196.

Clarke, J. (2007). *Energy simulation in building design*. Routledge.

Delgado, J. M. D., & Oyedele, L. (2021). Digital twins for the built environment: Learning from conceptual and process models in manufacturing. *Advanced Engineering Informatics*, *49*, 101332.

Desai, D., & Hartmann, T. (under review). Precept building capturing & energy modelling. Retrieved June 2, 2024, from https://www.precept-project.eu/deliverables/#wp2

Grieves, M. (2014). Digital twin: Manufacturing excellence through virtual factory replication. *White paper*, *1*(2014), 1–7.

Gutsche, C., & Hartmann, T. (2017). 4m process roadmap and implementation guidelines. Retrieved April 26, 2024, from https://www.p2endure-project.eu/en/results/PublishingImages/d1-3/D2.1_4M%20Process%20Roadmap%20and%20Implementation%20Guidelines.pdf

Ham, Y., & Golparvar-Fard, M. (2013). An automated vision-based method for rapid 3d energy performance modeling of existing buildings using thermal and digital imagery. *Advanced Engineering Informatics*, *27*(3), 395–409.

Koo, J., & Yoon, S. (2024). Simultaneous in-situ calibration for physical and virtual sensors towards digital twin-enabled building operations. *Advanced Engineering Informatics*, *59*, 102239.

Liu, Y., Feng, J., Lu, J., & Zhou, S. (2024). A review of digital twin capabilities, technologies, and applications based on the maturity model. *Advanced Engineering Informatics*, *62*, 102592.

Lu, Q., Chen, L., Li, S., & Pitt, M. (2020). Semi-automatic geometric digital twinning for existing buildings based on images and cad drawings. *Automation in Construction*, *115*, 103183.

of Energy, U. D. (2024). Energyplus whole building energy simulation software. Retrieved June 2, 2024, from https://energyplus.net/

Tan, Y., Chen, P., Shou, W., & Sadick, A.-M. (2022). Digital twin-driven approach to improving energy efficiency of indoor lighting based on computer vision and dynamic bim. *Energy and Buildings*, *270*, 112271.

Utkucu, D., Ying, H., Wang, Z., & Sacks, R. (2024). Classification of architectural and mep bim objects for building performance evaluation. *Advanced Engineering Informatics*, *61*, 102503.

Yang, Z., & Becerik-Gerber, B. (2015). A model calibration framework for simultaneous multi-level building energy simulation. *Applied Energy*, *149*, 415–431.

Zou, P. X., Xu, X., Sanjayan, J., & Wang, J. (2018). Review of 10 years research on building energy performance gap: Life-cycle and stakeholder perspectives. *Energy and Buildings*, *178*, 165–181.