

From Sensor Time Series to Building Performance Indicators: A BIM-Informed Framework for Contextual Data Integration

Nikol Krausz^{1*}, Árpád Barsi¹

¹ Department of Photogrammetry and Geoinformatics, Faculty of Civil Engineering, Budapest University of Technology and Economics, H-1111 Budapest, Műegyetem rkp. 3, Hungary

* Corresponding author, e-mail: krausz.nikol@emk.bme.hu

Received: 26 July 2026, Accepted: 21 September 2026, Published online: 27 September 2026

Abstract

The integration of Building Information Modeling (BIM), sensor data, and data-driven analysis can support a more contextual interpretation of building performance in existing buildings. This paper presents a pilot study on sensor-driven, BIM-informed building performance analysis, using the sensor network installed in the "F" building of the Budapest University of Technology and Economics as a case study. The aim is not to develop a complete digital twin platform, but to demonstrate how indoor environmental sensor measurements can be transformed from isolated time series into contextualized building performance information. The proposed approach combines sensor data preparation, quality assessment, sensor-to-BIM mapping, and exploratory analysis of indoor environmental patterns. Temperature and relative humidity measurements were processed to identify data quality issues, invalid measurements, outliers, temporal patterns, correlations, and clustering-based response groups. The sensor-to-BIM mapping linked measurement points to rooms, floors, functions, zones, spatial contexts, and BIM identifiers, enabling interpretation within spatial and semantic building context. The pilot results show that data quality has a direct influence on analytical outcomes, including clustering results. Contextualized sensor analysis supported the identification of stable rooms, strongly fluctuating spaces, and rooms with unusual environmental responses. A preliminary temperature prediction using a Long Short-Term Memory neural network for the F29 lecture hall also demonstrated the potential of short-term forecasting as a future extension. The study highlights that sensor-informed BIM can provide an intermediate layer between raw monitoring data and future digital twin-based operation.

Keywords

BIM, digital twin, sensor data analytics

1 Introduction

The digital transformation of the built environment has changed the role of building documentation and monitoring systems. Buildings are no longer considered only static physical assets but increasingly dynamic information systems whose operational state can be observed, analyzed, and used for decision-making. This shift is particularly important in the operation of existing buildings, where energy performance, comfort conditions, maintenance needs, and technical limitations must be assessed under real operating conditions.

Building Information Modeling (BIM) has become a central method for representing the geometric, semantic, and functional information of buildings throughout their life cycle. In new construction projects, BIM is commonly used for design coordination, quantity take-off, scheduling, and documentation. In existing buildings, however,

BIM also has significant potential as an information management platform for facility management and building operation. Volk et al. [1] emphasize that BIM for existing buildings remains a challenging field due to incomplete documentation, changing building conditions, and the need to integrate heterogeneous information sources.

In parallel, Internet of Things (IoT) sensor networks provide continuous information about the actual state of buildings. Indoor environmental measurements, such as temperature, relative humidity, air pressure, CO₂ concentration, or occupancy-related indicators, can support comfort assessment, fault detection, and maintenance planning. However, sensor data are often handled as isolated time series. Without spatial and semantic context, their interpretation remains limited. For example, similar temperature patterns may have different operational meanings

depending on whether they occur in an internal office, a lecture room, a laboratory, or a facade-adjacent space.

The integration of BIM and IoT data is therefore a key step toward sensor-informed digital twins. A digital twin combines static building information with dynamic data streams and supports monitoring, interpretation, prediction, and decision-making. Recent studies have highlighted the role of BIM, IoT, and digital twin technologies in improving construction and facility management processes [2]. Despite the increasing attention given to BIM–IoT integration and digital twins, practical methods are still needed for linking indoor environmental sensor data with spatial and semantic building context in a reproducible and interpretable way.

This paper presents a pilot study toward sensor-driven BIM-informed building performance analysis. The aim is to propose and demonstrate an initial workflow that combines indoor environmental sensor data, BIM-informed contextual information, and data-driven analysis. The workflow includes sensor data preparation, sensor-to-BIM mapping, and exploratory analysis of environmental response patterns. Rather than developing a complete automated digital twin platform, the proposed approach focuses on an intermediate stage in which raw sensor time series are quality-checked, linked to building-related spatial and semantic information, and prepared for subsequent analysis. Compared with more comprehensive BIM–IoT and digital twin solutions, the proposed workflow is a relatively simple and practical approach that can be applied in existing buildings where documentation and monitoring systems may be heterogeneous. Its main practical benefit is the structured contextualization of sensor data without requiring a fully integrated digital twin environment. A current limitation is that sensor observations are associated with building spaces at room level and are not yet fully automated or continuously synchronized with the BIM model. The study contributes a reproducible workflow that combines data quality assessment, BIM-informed contextualization, and exploratory sensor analytics, thereby bridging raw monitoring data and future digital twin applications.

2 Sensor-informed BIM for existing and heritage buildings

Heritage Building Information Modeling (HBIM) is a specialized application of BIM that focuses on the documentation and management of historic buildings and cultural heritage assets. The objective of HBIM is to create a detailed digital representation of historic and heritage buildings that

incorporates their structural, material, and historical characteristics [3]. Modern three-dimensional data acquisition technologies, such as laser scanning and photogrammetry, play a key role in developing HBIM models by providing highly accurate and realistic geometric information [4]. The key advantage of HBIM is that it serves not only as a digital documentation tool but also as a management platform supporting decision-making in heritage conservation, restoration works, condition assessment, sustainable operation and maintenance, and risk management [5].

Act C of 2023 on Hungarian Architecture [6], which integrates the former Act LXXVIII of 1997 on the Formation and Protection of the Built Environment [7] and substantial parts of Act LXIV of 2001 on the Protection of Cultural Heritage [8], consolidates building and heritage protection regulations within a unified legal framework. It places particular emphasis on the primacy of preserving the built heritage. The legislation stipulates that construction activities carried out within a protected historic environment may be authorized only if they do not cause irreversible damage. This requirement directly reinforces the importance of HBIM-based documentation and decision-support processes.

The literature increasingly emphasizes the need for dedicated heritage-specific standards for HBIM [1, 9, 10]. These standards should take into account:

- irregular geometries and historical layers;
- conservation ethics, including the principle of minimum intervention;
- long-term monitoring and sustainability considerations; and
- interoperability among the various disciplines involved, including conservators-restorers, structural engineers, building services engineers, heritage conservation specialists, and other professionals.

Recent heritage-oriented BIM studies further indicate that HBIM frameworks are increasingly expected to support multi-scale analysis, data repositories, and the integration of sensor systems for monitoring and management purposes [11].

While HBIM provides a structured representation of the physical, historical, and semantic characteristics of heritage assets, it is primarily based on static or periodically updated information. In operational building management, this information should be complemented by dynamic data that describes the building's actual state. IoT-based sensor networks can provide such data by continuously

measuring indoor environmental conditions, including temperature, relative humidity, air pressure, CO₂ concentration, or occupancy-related indicators [2, 12, 13]. In cultural heritage contexts, IoT- and digital twin-based monitoring has also been discussed as a promising direction for predictive maintenance and condition assessment [14].

Several studies demonstrate different approaches to combining BIM with sensor data and operational information. BIM–IoT integration has been investigated in relation to facility management and the renovation of existing buildings [12]. A BIM- and IoT-based monitoring system has also been developed for thermal comfort monitoring and automated alert generation [13]. An open-source IoT architecture has been proposed to support building facility maintenance [15], while a digital twin-enabled BIM–IoT framework has been investigated for monitoring, predicting, and optimizing indoor thermal comfort using machine learning [16]. In heritage applications, an open HBIM–IoT monitoring platform has been proposed for the management of historical sites and museums [17]. Together, these studies illustrate the potential of integrating building information with continuously collected operational data for monitoring, analysis, maintenance, and decision support.

The combination of BIM or HBIM with sensor data can therefore support a more contextual interpretation of building performance. By associating measurements with rooms, floors, zones, functions, and other spatial or semantic attributes, sensor time series can be interpreted in relation to the building spaces in which the measurements were collected. This is particularly relevant in existing and heritage buildings, where documentation may be fragmented, building functions may be heterogeneous, and operational interventions may be constrained by architectural or heritage considerations. At the same time, interoperability between BIM and facility management systems remains a challenge because of fragmented workflows, incompatible schemas, semantic inconsistencies, and limited real-time data exchange [18].

In the studies considered here, the emphasis is generally placed on integrated monitoring architectures, application-specific services, or higher-level operational decision support [12, 13, 15–17]. In comparison with these application-oriented approaches, the intermediate processing stage in which raw indoor environmental sensor time series are systematically assessed for data quality, preprocessed, and associated with BIM-derived spatial and semantic information is less explicitly described as a distinct and reproducible workflow. This stage is

important because missing data, erroneous measurements, inconsistent identifiers, and the absence of spatial context can directly influence the reliability and interpretability of subsequent analytical results.

The present study focuses specifically on this intermediate processing stage. It combines time-series pre-processing and data quality assessment with room-level sensor-to-BIM mapping and exploratory analysis within a reproducible workflow. Rather than implementing a fully integrated BIM–IoT or digital twin platform, the approach uses BIM-derived spatial and semantic information to contextualize sensor measurements before higher-level analysis is performed. In this way, the proposed workflow provides a practical intermediate layer between raw indoor environmental monitoring data and more advanced sensor-informed digital twin applications.

3 Case study: The BME "F" building and sensor monitoring system

The case study is based on the sensor network installed in the "F" building of the Budapest University of Technology and Economics (BME), located on the Lágymányos Campus in Budapest, Hungary. The building is a historic university building with long-standing educational, research, and laboratory functions. Lecture rooms, offices, laboratories, classrooms, and other academic spaces are accommodated within the building, which makes it suitable for investigating indoor environmental conditions under real operating conditions. From a facility management perspective, several challenges are associated with the building. The available documentation may be fragmented, the technical systems are partly aging, and modernization options may be limited by architectural and heritage-related constraints [19]. At the same time, the building remains in active daily use; therefore, comfort conditions and operational performance need to be assessed under real occupancy and usage conditions. These characteristics make the building a suitable environment for examining how indoor environmental sensor data can support the interpretation of building performance.

A sensor monitoring system is being used to collect indoor environmental data from selected rooms distributed across the building. The monitored spaces include different room types, such as lecture rooms, offices, laboratories, and classrooms. This spatial and functional diversity enables environmental patterns to be compared not only between individual rooms, but also between room groups, zones, and functions.

The monitoring system focuses primarily on indoor environmental variables that are relevant to comfort assessment and building operation. The main measured parameters include temperature and relative humidity, while CO₂ concentration was also available for selected measurement points. The sensor network was designed with practical operational constraints in mind. Since the monitored rooms are frequently used, the system is required to operate with low maintenance needs and without disturbing everyday building use. For this reason, wireless communication and battery-powered operation are important aspects of the monitoring setup. In the monitoring concept, LoRaWAN technology [20] was selected because it provides a suitable balance between long-range communication, low power consumption, and applicability in a dense built environment. The technical development and implementation of the long-range wireless sensor network used in this case study are described in detail by Szatmári et al. [21].

A BIM model of the building was previously developed. In the present pilot study, the BIM model itself was not directly integrated into the analytical workflow. Consequently, no BIM authoring platform or native BIM model format was processed as part of the sensor-data analysis. Instead, the relationship between the sensor measurements and the building context was represented through room-level identifiers used to associate each sensor with its corresponding monitored space. These identifiers provide the connection between the sensor time series and contextual attributes such as room designation, floor level, function, zone, orientation, and spatial position within the building. This simplified identifier-based approach was intentionally selected for the pilot phase in order to evaluate the usefulness of spatial and semantic building context before implementing a more tightly integrated BIM–sensor workflow. The current analysis, therefore, does not rely on direct BIM model queries, embedded sensor objects, live synchronization, or automated data exchange with a BIM environment.

4 Data preparation and quality assessment

The collected sensor measurements form timestamped time series datasets. Each record is associated with a measurement time, a sensor identifier, and one or more measured indoor environmental variables, such as temperature and relative humidity. Since the monitoring is being carried out under real building operation conditions, the raw datasets require systematic preprocessing before statistical analysis, clustering, or predictive modeling can be performed.

The overall workflow of the proposed sensor-driven BIM-informed analysis is shown in Fig. 1. The process starts with raw indoor environmental sensor measurements, followed by time-series preprocessing and data quality assessment. The cleaned and checked measurements are then linked to BIM-informed spatial and semantic information through sensor-to-BIM mapping and contextualization. Finally, analytics are applied to derive sensor-based building performance indicators, including data quality indicators, temporal response indicators, correlation-based indicators, clustering-based response groups,

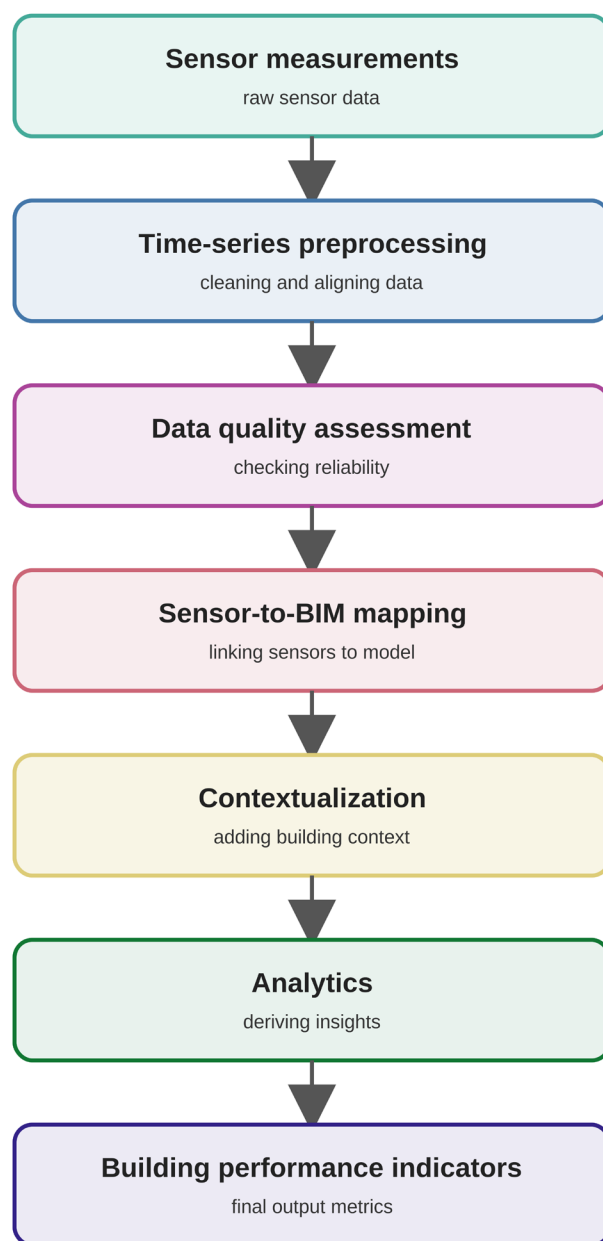


Fig. 1 Workflow from raw sensor measurements to contextualized building performance indicators

and preliminary prediction errors, which can support operational interpretation and further decision-making.

The first step of data preparation was the inspection and standardization of the timestamp information. The measurement time was converted into a consistent datetime format, and the records were ordered chronologically. This step was necessary to ensure that temporal patterns, daily variations, and aggregated indicators could be interpreted correctly. Duplicate or inconsistent timestamps, if present, were treated as data quality issues and checked before further processing.

Missing values and communication gaps were also examined. In wireless sensor networks, missing records may occur due to temporary communication problems, battery-related issues, installation phases, or data transmission interruptions. Therefore, the ratio of missing values was considered an important quality indicator for each sensor. Short gaps may have limited influence on aggregated indicators, while longer gaps can reduce the reliability of room-level or zone-level comparisons. For this reason, missing data were not treated only as technical errors, but also as information about the reliability of the monitoring process.

Invalid and physically unrealistic values were identified during the quality assessment. For indoor temperature monitoring, extreme values that cannot represent realistic building conditions indicate sensor error, communication error, or data logging problems. For example, multiple values of $-122\text{ }^{\circ}\text{C}$ were classified as invalid measurements. Instead of removing the affected records, these values were replaced by the mean of the five preceding valid measurements in order to preserve temporal continuity. This step was necessary because such values can strongly distort descriptive statistics, correlation analysis, and clustering results (Fig. 2).

Clustering was used to examine whether sensors or rooms with similar environmental behavior could be grouped based on their temperature response. This was not intended as an automatic classification of building performance, but as an exploratory method to identify stable groups, fluctuating spaces, and possible room-level deviations. The silhouette score was used to evaluate the consistency of the clustering results and to compare the effect of data cleaning on the optimal number of clusters.

After timestamp handling and data cleaning, the measurements were prepared for analysis through temporal aggregation. Depending on the analysis objective, hourly or daily aggregation can be used. Hourly aggregation is suitable for identifying daily operational patterns and short-term environmental response, while daily indicators are more appropriate for longer-term comparison between rooms, zones, or functions. In the present workflow, aggregation was used to reduce measurement noise and to create comparable indicators across sensors.

The data preparation process included the calculation of basic quality and descriptive indicators. These included the number of valid measurements, the ratio of missing data, mean values, minimum and maximum values, standard deviation, and temporal variation. These indicators provided an initial overview of both sensor reliability and indoor environmental behavior. They also supported the selection of sensors and time periods suitable for further analysis.

Outlier detection was applied after temporal ordering and data quality assessment in order to identify unusual values in the indoor environmental time series. For time-dependent outlier detection, a rolling Z-score was used, which compares each measurement with the local mean and standard deviation calculated within a moving time window, as given in Eq. (1):

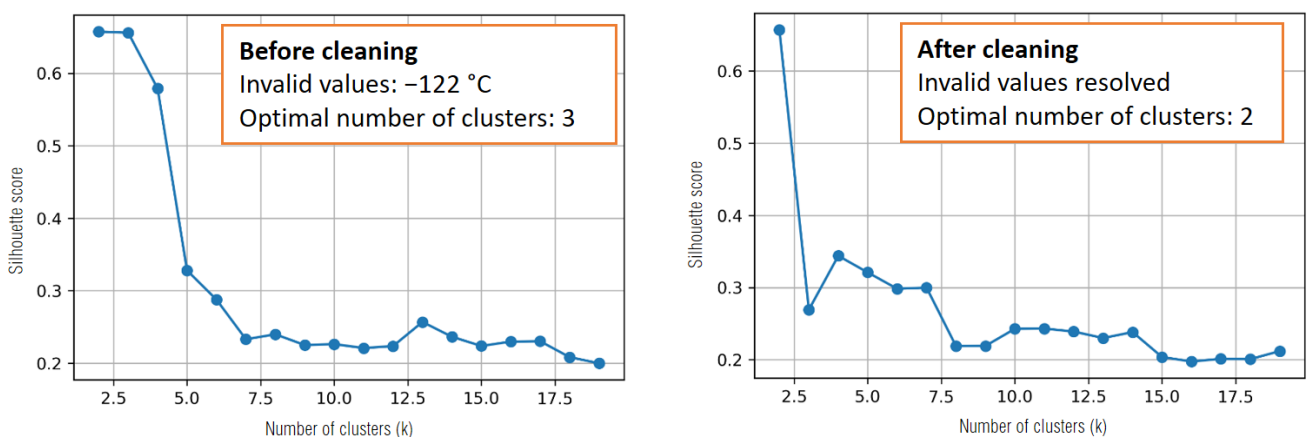


Fig. 2 Comparison of silhouette scores before and after data cleaning for temperature sensor clustering

$$Z_i = (x_i - \mu_{w,i-1}) / (\sigma_{w,i-1}^*), \quad (1)$$

where x_i is the measured value at time step i , $\mu_{w,i-1}$ and $\sigma_{w,i-1}$ are the rolling mean and rolling standard deviation calculated from the preceding moving window, and $\sigma_{w,i-1}^*$ is the adjusted rolling standard deviation. To avoid unrealistically large Z -scores caused by very small local variance, the minimum value of $\sigma_{w,i-1}^*$ was set to 0.2. Values with $|Z_i| > 3$ were classified as potential outliers.

As a more robust alternative, a median absolute deviation (MAD)-based modified Z -score was also applied. The modified Z -score was calculated according to Eq. (2):

$$M_i = (0.6745(x_i - \text{median}_{w,i-1})) / (MAD_{w,i-1}), \quad (2)$$

where $\text{median}_{w,i-1}$ is the rolling median and $MAD_{w,i-1}$ is the rolling median absolute deviation calculated from the preceding moving window. Values with $|M_i| > 3.5$ were classified as potential outliers. The resulting outlier detection for the F29 lecture hall is shown in Fig. 3.

Data cleaning was found to have a direct influence on the subsequent analytical results. Outliers and invalid measurements can create artificial differences between sensors and may lead to misleading grouping or clustering outcomes. Therefore, data preparation was considered not only as a technical preprocessing step, but as an essential part of sensor-informed building performance analysis. The reliability of the final interpretation depends strongly on the quality, consistency, and completeness of the input data.

5 Sensor-to-BIM mapping and contextualization

Following data preparation and quality assessment, the sensor measurements were linked to spatial and semantic building information to support their contextual interpretation. The purpose of this step was to ensure that the collected sensor data were not interpreted as isolated time series, but as building-related information associated with specific rooms, functions, zones, and spatial conditions.

The sensor-to-BIM mapping process was based on a structured mapping table. Each sensor identifier was assigned to a corresponding room and to the related BIM space identifier (BIM_Space_ID). In addition to the direct spatial relationship, further contextual attributes were included, such as floor level, room function, functional zone, orientation, facade adjacency or internal location, and the measured variables. These attributes enabled the sensor measurements to be grouped and compared according to building-related categories.

At the pilot stage, the mapping table was created and checked manually. This approach was considered appropriate because the objective was not to develop a fully automated BIM-IoT integration platform, but to establish a reliable and reproducible contextualization workflow. Manual mapping also enabled the identification and correction of inconsistencies in room and sensor naming, as well as in documentation, during the preparation process. The consistency of the mapping table is essential, since incorrect sensor-room assignments may lead to a misleading interpretation of environmental patterns.

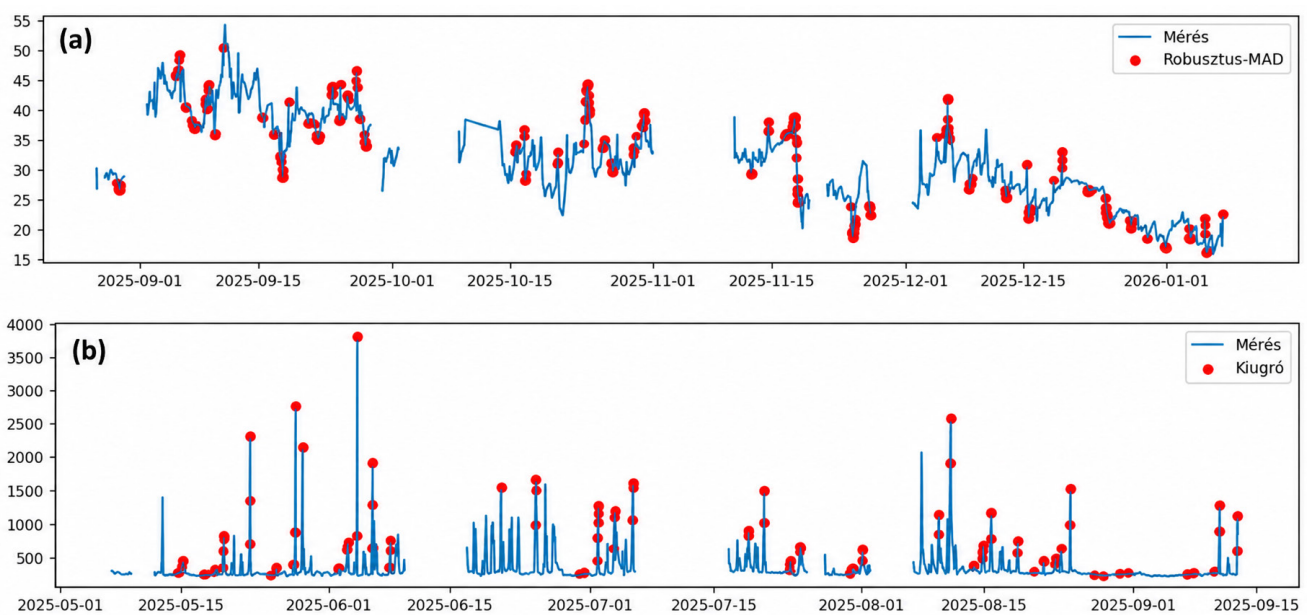


Fig. 3 Outlier detection in indoor environmental sensor data for the F29 lecture hall: (a) Robust-MAD based outliers in humidity data for 'F29-Lecture Hall'; (b) Rolling Z-score based outliers in CO₂ data for 'F29-Lecture Hall'

The main attributes used in the sensor-to-BIM mapping are summarized in Table 1. The table structure was designed to remain simple enough for practical use, while still containing the most important spatial and semantic information required for analysis.

The contextualized dataset created through this mapping process supports several types of analysis. First, measurements can be compared between rooms with similar functions, such as lecture rooms or offices. Second, rooms can be grouped according to floor level, zone, orientation, or facade adjacency. Third, unusual environmental behavior can be interpreted in relation to the spatial and functional characteristics of the room in which it occurs. This contextualization is particularly important in existing and historical buildings, where measured environmental values may be influenced by several interacting factors, including room use, heating system configuration, facade exposure, ventilation conditions, and previous architectural modifications. For example, a temperature fluctuation in a lecture room may be related to occupancy and scheduled use, while a similar fluctuation in an office or laboratory may have a different operational interpretation.

The sensor-to-BIM mapping, therefore, acts as a bridge between raw sensor measurements and building performance interpretation. It does not replace a full digital twin platform, but it provides a practical intermediate layer that enables sensor data to be analyzed within its spatial and semantic context. The resulting contextualized dataset forms the basis for the data-driven analysis of indoor environmental patterns.

The current room-level approach does not represent the exact three-dimensional position of individual sensors. This is a limitation because proximity to windows, doors, heating elements, or other building components may influence the measured conditions. Explicit 3D sensor positioning can therefore be considered in future extensions of the workflow.

6 Data-driven analysis of indoor environmental patterns

After data preparation and sensor-to-BIM contextualization, the cleaned sensor measurements were analyzed in

order to identify characteristic indoor environmental patterns. The objective of the analysis was to describe the behavior of monitored spaces, compare rooms and room groups, and identify locations with unusual environmental responses. The analysis was exploratory in nature and was intended to provide interpretable indicators for building performance assessment rather than automated decision-making.

Basic descriptive indicators and group-based comparisons were used to provide an initial overview of the monitored spaces. These indicators supported the interpretation of general indoor environmental conditions and enabled rooms to be compared according to function, zone, floor level, or spatial position.

Time series analysis was used to examine temporal changes in indoor environmental variables. Temperature and relative humidity measurements were analyzed to identify daily and longer-term patterns, sudden changes, delayed responses, and periods of increased variation. Such temporal patterns can reflect different operational effects, including room use, heating system behavior, ventilation conditions, or external environmental influence. Since the measurements were linked to room function and spatial context, the observed temporal behavior could be interpreted not only as a sensor signal, but also as a room-level environmental response.

Correlation analysis was applied to examine relationships between monitored spaces and measured variables. Correlations between sensors can indicate common building-level effects, such as heating operation or weather-related trends, while weaker or inconsistent relationships may suggest local room-specific behavior. Correlation analysis between temperature and relative humidity can also support the interpretation of indoor environmental dynamics. The contextual information provided by the sensor-to-BIM mapping enables these relationships to be examined according to room function, floor level, zone, or spatial position.

Clustering was used as an exploratory method to identify groups of sensors or rooms with similar environmental response patterns. Depending on the selected input variables, clustering can support the recognition of stable

Table 1 Example structure of the sensor-to-BIM mapping table

Sensor ID	Room	Floor	Function	Zone	Orientation/position	BIM_Space_ID	Variables
S-012	F-29	2F	lecture hall	south wing	south-facing/facade-adjacent	ifcSpace_29	T/RH/ CO ₂
S-018	I-A-08	GF	laboratory	internal	internal	ifcSpace_A08	T/RH
S-027	C-14	1F	office	north wing	north-facing/window-adjacent	ifcSpace_14	T/RH

rooms, strongly fluctuating spaces, or rooms with unusual behavior compared with similar spaces. Since clustering is sensitive to data quality, the cleaned dataset and quality indicators were used to reduce the influence of invalid measurements and outliers.

Preliminary predictive modeling can be considered in later stages of the workflow. Time series prediction methods, such as Long Short-Term Memory (LSTM) neural networks, can be used to estimate short-term indoor temperature values based on previous measurements. However, in the present study, predictive modeling is treated as a possible extension rather than the main analytical objective. Longer monitoring periods, seasonal coverage, and careful validation are required before prediction-based results can be used reliably for operational decision support.

The combined use of descriptive statistics, time series analysis, correlation analysis, group comparison, and clustering provides a structured basis for interpreting indoor environmental behavior. The goal is to identify stable rooms, strongly fluctuating spaces, and rooms with unusual environmental responses. In this way, the combined analysis can support the early recognition of operational issues that may not be obvious from individual measurements alone.

7 Pilot results

The monitoring system was installed in several stages. Initially, the first instruments were placed in three laboratories and three office rooms. By 20 September 2025, all sensors had been installed in their final positions. Consequently, the temporal differences in sensor installation had to be considered during data interpretation, since the sensors did not provide measurements from the same starting time in all cases. The available dataset contained more than 5000 records for temperature and relative humidity measurements and more than 3000 records from the CO₂ sensors. For each sensor, 24 measurements per day were provided at hourly resolution. The entire sensor network covers more than 40 rooms, supplemented by 1 outdoor and 2 attic measurement points. During data processing, the time of the first valid measurement value was determined for each sensor using Python, and the ratio of missing data was determined starting from that time (see Table 2).

The pilot analysis provided initial evidence that indoor environmental sensor data can support the interpretation of building performance when data quality and spatial context are considered together. Since the monitoring campaign is still ongoing, the results presented in this section are interpreted as preliminary indicators rather than

Table 2 Excerpt of missing-value ratios calculated from the first valid measurement of each sensor

Sensor	Room	Variables	Unit	Missing data ratio
S-012	F29	T	°C	22.09%
		RH	%	22.08%
		CO ₂	ppm	20.79%
S-029	I-1-04-Oktatoi	T	°C	0.05%
		RH	%	0.055%
S-018	I-A-08-Labor	T	°C	0.036%
		RH	%	0.018%

final operational diagnoses. The aim is to demonstrate the analytical potential of the sensor-informed approach and to identify patterns that may justify further investigation.

7.1 Data quality and preprocessing effects

The quality assessment confirmed that preprocessing has a direct influence on the reliability of subsequent analysis. Missing values, communication gaps, and occasional invalid measurements were identified in the raw datasets. These issues are typical in real building monitoring systems and can affect statistical indicators, correlations, and clustering results if they are not treated appropriately.

A clear example was observed in the temperature dataset, where multiple invalid values of −122 °C were detected. These values were classified as physically unrealistic for indoor building conditions. Instead of removing the affected records, the invalid values were replaced by the mean of the five preceding valid measurements. This correction was applied in order to preserve the temporal continuity of the time series while reducing the distorting effect of sensor or data logging errors.

The effect of this correction was also visible in the clustering analysis. Before data cleaning, the optimal number of clusters was three, while after replacing the invalid values, the optimal number of clusters changed to two. This indicates that sensor errors and extreme outliers can create artificial separation between data groups and may lead to misleading interpretations if they are not treated appropriately.

Outlier detection was also applied to other indoor environmental variables. In the relative humidity and CO₂ time series, values classified as outliers indicated periods of unusual sensor response or possible operational events. These values were not interpreted automatically as faults. Instead, they were treated as indicators requiring contextual interpretation, since high variation may result from room use, ventilation conditions, temporary occupancy effects, or measurement uncertainty.

7.2 Indoor environmental response patterns

The cleaned time series data revealed differences in the environmental response of monitored rooms. Some spaces showed relatively stable indoor conditions, while others exhibited stronger daily variation or more irregular temporal behavior. These differences suggest that room-level environmental patterns are influenced by several interacting factors, including room function, occupancy, heating system behavior, ventilation conditions, and spatial location within the building.

The distinction between stable and fluctuating rooms is relevant for building operation. Stable rooms may indicate balanced thermal conditions or limited disturbance, while strongly fluctuating spaces may require further investigation if the variation is not consistent with their expected use (Fig. 4). However, the interpretation of these patterns depends on building context. For example, higher daily variation may be acceptable in lecture rooms with scheduled occupancy, while similar behavior in offices or laboratories may require a different interpretation.

Correlation analysis supported the distinction between shared building-level effects and local room-specific behavior. Sensors with similar temporal trends may reflect common operational influences, such as heating operation or external weather-related effects. A representative example is provided by the two attic-level measurement points. The correlation between PADLAS-I and PADLAS-II was very strong, with $R = 0.971$, indicating that the two attic sensors followed highly similar environmental dynamics. In addition, the difference between their correlation profiles was small with respect to the other measurement points. The largest absolute difference between the two correlation profiles was 0.092 for I-159_136-Labor, while the smallest was 0.003 for III-2-03-Muhely/Oktatoi. This

suggests that the two attic sensors captured a consistent semi-external environmental response. In contrast, weaker or inconsistent relationships between comparable rooms may indicate local influences, including room use, facade exposure, ventilation differences, or sensor placement.

The sensor-to-BIM mapping made it possible to examine these relationships according to room function, floor level, zone, and spatial position. This contextual grouping helped avoid treating all sensor time series as equivalent measurement points.

Clustering provided an exploratory way to group rooms or sensors with similar environmental response patterns. The results showed that clustering can help identify groups of spaces with comparable indoor behavior, as well as rooms that deviate from otherwise similar spaces. The analysis also confirmed that clustering outcomes are sensitive to data quality, since invalid measurements and outliers can strongly influence the apparent structure of the data.

7.3 LSTM-based temperature prediction for the F29 Lecture Hall

As a possible extension of the exploratory analysis, a preliminary short-term temperature prediction was also tested for the F29 lecture hall. An LSTM neural network was applied to estimate indoor temperature values based on previous measurements. The aim of this step was not to develop a fully validated predictive model, but to examine whether the available sensor time series could support short-term forecasting as a future component of a sensor-informed digital twin workflow. Fig. 5 shows the measured and LSTM-predicted temperature time series for the F29 lecture hall. The predicted values follow the general trend of the measured indoor temperature, indicating that the model was able to capture the main temporal pattern

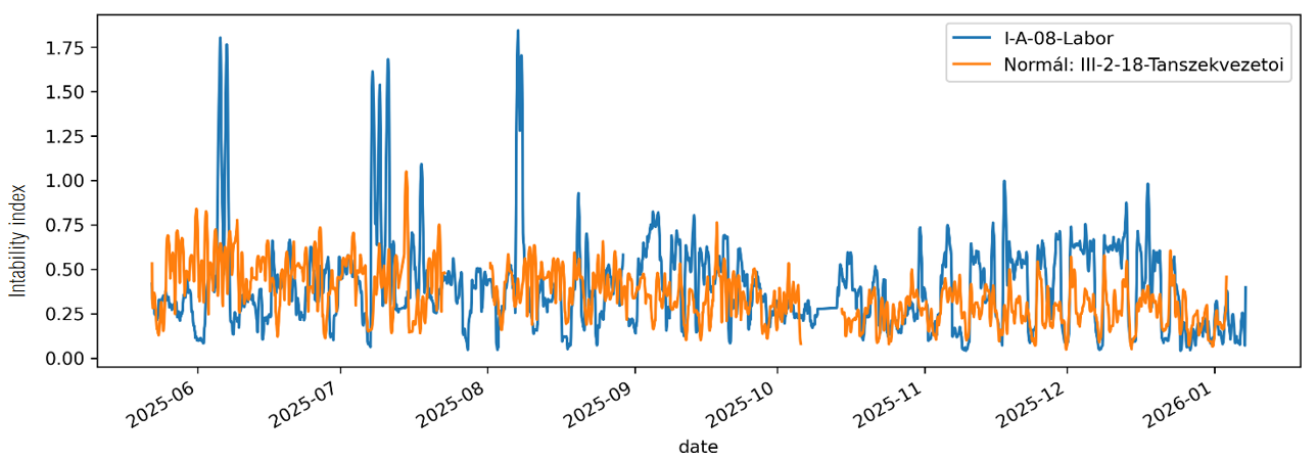


Fig. 4 Temperature instability analysis of the I-A-08 laboratory

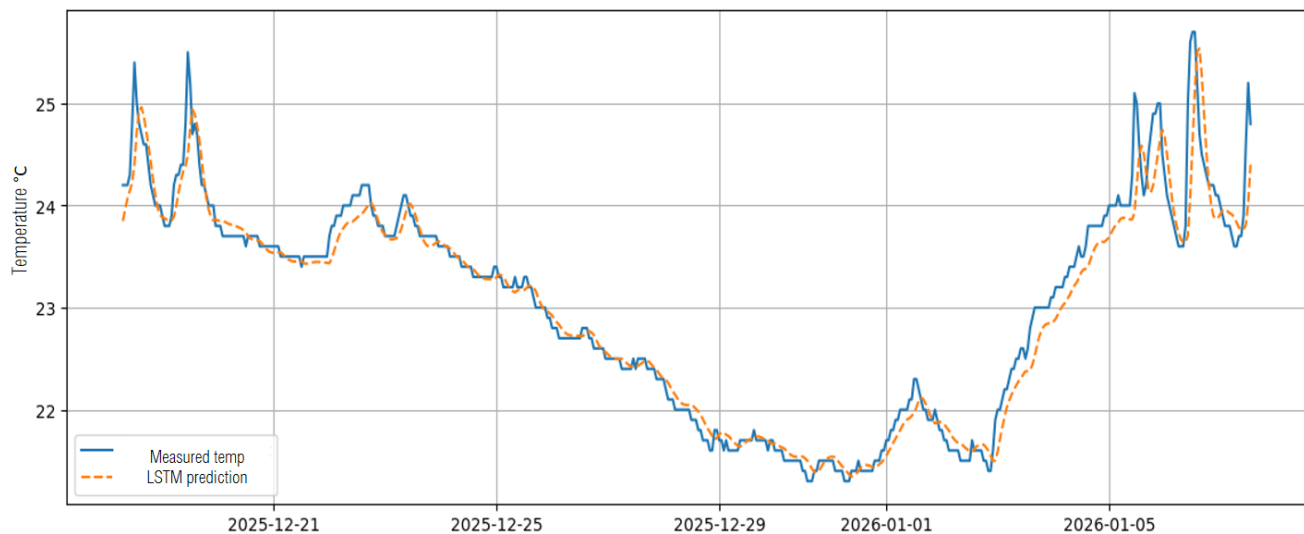


Fig. 5 Measured and LSTM-predicted temperature time series for the F29 lecture hall

of the selected period. However, rapid changes and short-term peaks were reproduced with some smoothing, which is a common limitation in short-term prediction based on limited input information.

The preliminary result suggests that sensor-based temperature prediction may be a promising extension of the workflow. Longer monitoring periods, seasonal data coverage, and additional explanatory variables, such as outdoor temperature, occupancy-related indicators, or HVAC operation data, would be required for more robust validation. Therefore, predictive modeling is considered here as an initial demonstration and a future research direction.

7.4 Sensor-informed operational interpretation

The results indicate that indoor environmental sensor data can provide useful information for building operation when they are interpreted together with spatial and semantic building context. By linking measurements to BIM-informed context, similar sensor values can be interpreted differently depending on the room type and location. This is particularly important in existing buildings, where building use, technical systems, and previous modifications may create heterogeneous indoor environmental behavior.

The main operational relevance of the proposed approach is that it supports the identification of room-level environmental patterns that may otherwise remain hidden in isolated sensor time series. Stable rooms, strongly fluctuating spaces, and rooms with unusual environmental responses can be identified as operational indicators and used to prioritize further inspection.

From a facility management perspective, the approach can support comfort-oriented operation and preliminary

fault screening. Rooms with unusually high variation, delayed thermal response, inconsistent behavior compared with similar spaces, or deviations from predicted temperature trends can be selected for further investigation.

However, the results should be interpreted as indicators rather than direct diagnoses. A detected anomaly or deviation does not automatically identify the cause of the problem. It may result from occupancy, ventilation, facade exposure, heating system behavior, sensor placement, or data quality issues. Therefore, sensor-informed analysis should be combined with expert evaluation and, where necessary, physical inspection or simulation-based assessment.

8 Operational relevance, limitations

From an operational perspective, the main value of the proposed approach lies in its ability to transform monitoring data into interpretable building-related indicators. In conventional sensor-only monitoring, measurements may reveal temporal changes, correlations, outliers, or predicted trends, but their operational meaning often remains uncertain. By combining sensor data with BIM-informed spatial and semantic context, the proposed approach supports a more structured interpretation of indoor environmental behavior. This can help facility managers identify spaces that require further attention, compare rooms with similar functions, and distinguish general building-level effects from local deviations. The approach is particularly relevant for existing and historical buildings, where operational decisions are often constrained by fragmented documentation, aging technical systems, heterogeneous room functions, and limited modernization possibilities.

Sensor-to-BIM contextualization can therefore support a more meaningful interpretation than sensor-only analysis, because measurements are evaluated together with room-level and zone-level building information.

Several limitations must be considered. First, the monitoring campaign is still ongoing, and the available dataset may not yet cover all seasonal conditions. As a result, the current findings may reflect only part of the building's annual environmental behavior. Second, sensor placement strongly influences interpretation. Measurements taken near windows, heating elements, doors, or occupied areas may represent local conditions rather than the general state of the room. Third, data quality issues such as missing values, communication gaps, invalid measurements, and outliers can affect statistical indicators, correlation analysis, clustering, and predictive modeling.

Applying the approach to another building would be an important next step for validation. Such a comparison would make it possible to distinguish between patterns that are specific to the BME "F" building and patterns that may be more generally applicable to existing university or heritage buildings. It would also help evaluate whether the proposed mapping attributes and analysis indicators are sufficient or whether additional contextual variables are required.

Overall, the pilot results indicate that sensor-informed analysis can provide a useful intermediate layer between raw monitoring data and facility management decision-making. The approach does not replace detailed building diagnostics, but it can help identify where to focus such diagnostics.

9 Conclusions

This paper presented a pilot study toward sensor-driven BIM-informed building performance analysis in an existing historical university building. The study was based on the sensor network installed in the BME "F" building

and focused on the integration of indoor environmental sensor data, BIM-informed contextual information, and data-driven analysis.

The aim was not to develop a complete digital twin platform, but to demonstrate how sensor measurements can be transformed from isolated time series into contextualized building performance information.

The results indicate that data preparation and quality assessment are essential steps in the interpretation of real building monitoring data. Missing values, communication gaps, invalid measurements, and outliers can significantly influence statistical indicators, correlations, clustering results, and preliminary predictive models. The sensor-to-BIM mapping approach provided a practical method for linking sensor measurements to rooms, floors, functions, zones, spatial context, and BIM space identifiers. This contextualization enabled the monitored spaces to be compared according to meaningful building-related categories. As a result, indoor environmental patterns could be interpreted not only as sensor signals, but also as room-level or zone-level building performance indicators. The pilot findings showed that sensor-informed analysis can support the identification of stable rooms, strongly fluctuating spaces, and rooms with unusual environmental responses. These results should be interpreted as operational indicators rather than final diagnoses.

The proposed workflow demonstrates that contextualized sensor time series enable building performance indicators that cannot be obtained from isolated measurements alone.

Overall, the presented pilot study represents an initial step toward sensor-driven, BIM-informed building operation. By combining BIM context, IoT sensor data, and data-driven analysis, the approach provides a practical basis for improving the interpretability of indoor environmental monitoring and for supporting future digital twin development in existing buildings.

References

- [1] Volk, R., Stengel, J., Schultmann, F. "Building Information Modeling (BIM) for existing buildings — Literature review and future needs", *Automation in Construction*, 38, pp. 109–127, 2014. <https://doi.org/10.1016/j.autcon.2013.10.023>
- [2] Baghalzadeh Shishegharkhaneh, M., Keivani, A., Moehler, R. C., Jelodari, N., Roshdi Laleh, S. "Internet of Things (IoT), Building Information Modeling (BIM), and Digital Twin (DT) in Construction Industry: A Review, Bibliometric, and Network Analysis", *Buildings*, 12(10), 1503, 2022. <https://doi.org/10.3390/buildings12101503>
- [3] Murphy, M., McGovern, E., Pavia, S. "Historic Building Information Modelling - Adding intelligence to laser and image based surveys of European classical architecture", *ISPRS Journal of Photogrammetry and Remote Sensing*, 76, pp. 89–102, 2013. <https://doi.org/10.1016/j.isprsjprs.2012.11.006>
- [4] Salah, R., Károlyfi, K. A., Szép, J., Géczy, N. "A structured framework for HBIM standardization: Integrating scan-to-BIM methodologies and heritage conservation standards", *Digital Applications in Archaeology and Cultural Heritage*, e00420, 2025. <https://doi.org/10.1016/j.daach.2025.e00420>

- [5] Puerto, A., Castañeda, K., Sánchez, O., Peña, C. A., Gutiérrez, L., Sáenz, P. "Building information modeling and complementary technologies in heritage buildings: A bibliometric analysis", *Results in Engineering*, 22, 102192, 2024.
<https://doi.org/10.1016/j.rineng.2024.102192>
- [6] Nemzeti Jogszabálytár "2023. évi C. törvény a magyar építészetéről" (Act C of 2023 on Hungarian Architecture), Budapest, Hungary, 2023. [online] Available at: <https://njt.hu/jogszabaly/2023-100-00-00> [Accessed: 14 January 2026]
- [7] Nemzeti Jogszabálytár "1997. évi LXXVIII. törvény az épített környezet alakításáról és védelméről" (Act LXXVIII of 1997 on the Shaping and Protection of the Built Environment), Budapest, Hungary, 1997. [online] Available at: <https://njt.hu/jogszabaly/1997-78-00-00> [Accessed: 14 January 2026].
- [8] Nemzeti Jogszabálytár "2001. évi LXIV. törvény a kulturális örökség védelméről" (Act LXIV of 2001 on the Protection of Cultural Heritage), Budapest, Hungary, 2001. [online] Available at: <https://njt.hu/jogszabaly/2001-64-00-00> [Accessed: 14 January 2026].
- [9] Murphy, M., McGovern, E., Pavia, S. "Historic building information modelling (HBIM)", *Structural Survey*, 27(4), pp. 311–327, 2009.
<https://doi.org/10.1108/02630800910985108>
- [10] Zhang, J., Zhang, L., Zhao, W., Lu, Y., Wang, X., Yao, S., Lyu, H. "HBIM for the protection and management: a case study of Chinese timber architectural heritage", *npj Heritage Science*, 13(1), 399, 2025.
<https://doi.org/10.1038/s40494-025-01926-1>
- [11] Fattore, C., Buldo, M., Priore, A., Porcari, S., Porcari, V. D., De Fino, M. "A Comprehensive Overview of Heritage BIM Frameworks: Platforms and Technologies Integrating Multi-Scale Analyses, Data Repositories, and Sensor Systems", *Heritage*, 8(7), 247, 2025.
<https://doi.org/10.3390/heritage8070247>
- [12] Altohami, A. B. A., Haron, N. A., Ales@Alias, A. H., Law, T. H. "Investigating approaches of integrating BIM, IoT, and facility management for renovating existing buildings: A review", *Sustainability*, 13(7), 3930, 2021.
<https://doi.org/10.3390/su13073930>
- [13] Valinejadshoubi, M., Moselhi, O., Bagchi, A., Salem, A. "Development of an IoT and BIM-based automated alert system for thermal comfort monitoring in buildings", *Sustainable Cities and Society*, 66, 102602, 2021.
<https://doi.org/10.1016/j.scs.2020.102602>
- [14] Cecere, L., Colace, F., Lorusso, A., Messina, B., Tucker, A., Santaniello, D. "IoT and Digital Twin: A new perspective for Cultural Heritage predictive maintenance", *Procedia Structural Integrity*, 64, pp. 2181–2188, 2024.
<https://doi.org/10.1016/j.prostr.2024.09.334>
- [15] Villa, V., Naticchia, B., Bruno, G., Aliev, K., Piantanida, P., Antonelli, D. "IoT Open-Source Architecture for the Maintenance of Building Facilities", *Applied Sciences*, 11(12), 5374, 2021.
<https://doi.org/10.3390/app11125374>
- [16] Iqbal, F., Mirzabeigi, S. "Digital Twin-Enabled Building Information Modeling–Internet of Things (BIM–IoT) Framework for Optimizing Indoor Thermal Comfort Using Machine Learning", *Buildings*, 15(10), 1584, 2025.
<https://doi.org/10.3390/buildings15101584>
- [17] Martinelli, L., Calcerano, F., Adinolfi, F., Chianetta, D., Gigliarelli, E. "Open HBIM–IoT Monitoring Platform for the Management of Historical Sites and Museums. An Application to the Bourbon Royal Site of Carditello", *International Journal of Architectural Heritage*, 19(2), pp. 153–170, 2025.
<https://doi.org/10.1080/15583058.2023.2272130>
- [18] Chatsuwan, M., Moriwaki, A., Ichinose, M., Alkhalaf, H. "BIM–FM Interoperability Through Open Standards: A Critical Literature Review", *Architecture*, 5(3), 74, 2025.
<https://doi.org/10.3390/architecture5030074>
- [19] Lechner Nonprofit Kft. "OENY", [online] Available at: <https://www.oeny.hu/oeny/veding/?telepules=Budapest%2011&hrs=5534&keppel=true> [Accessed: 14 January 2026]
- [20] LoRa Alliance "LoRaWAN Specification v1.1", [online] Available at: <https://resources.lora-alliance.org/technical-specifications/lorawan-specification-v1-1> [Accessed: 18 January 2026]
- [21] Szatmári, L., Aliyev, N., Németh, K. M., Nagy, B. "Development of Long-Range Wireless Sensor Network for Monitoring Historical Buildings", In: Nagy, B., Szalay, Z. (eds.) *Proceedings of CESBP 2025 - 6th Central European Symposium on Building Physics, Lecture Notes in Civil Engineering*, 796, Springer, Cham, pp. 1036–1053, 2026.
https://doi.org/10.1007/978-3-032-14019-7_75