Data-Driven Digital Twining of Ventilation Systems for Performance Optimization: A University Building Case Study

Andres Sebastian Cespedes Cubides – University of Southern Denmark, Denmark – acub@mmmi.sdu.dk
Jakob Bjørnskov – University of Southern Denmark, Denmark – jabj@mmmi.sdu.dk
Muhyiddine Jradi – University of Southern Denmark, Denmark – mjr@mmmi.sdu.dk

Abstract

This study introduces the creation and application of a data-driven digital twin for building ventilation systems, focusing on a university building as a case study. It employs a grey-box energy modelling framework to accurately forecast, simulate, and monitor the ventilation system's efficiency under diverse conditions. The study collects a substantial dataset to reflect various usage patterns and environmental influences, which serves to test and validate the component models of the ventilation system. These models are integrated into a digital twin platform, providing a comprehensive overview of the system's performance and critical indicators in real time. The digital twin facilitates informed decision-making for facility managers regarding energy consumption, inefficiency identification, and the recommendation of custom retrofitting actions specific to the building's characteristics and use. The findings confirm that digital twins are effective as a tool to continuously commission and detect anomalies in buildings. The study offers a ventilation modelling and monitoring method capable of recognizing rule-based control behaviours and changes in systems that occur in cycles, like system shifts from winter to summer, and can estimate total air mass flow rate with a correlation exceeding 80%.

1. Introduction

The European Union (EU) is addressing its energy and environmental objectives for 2030 and 2050 by focusing on the building sector, which accounts for nearly 40% of its energy consumption (European commission, 2018). According to the International Energy Agency (IEA), this sector is also responsible for about 36% of total emissions, broken down into residential buildings (22%), non-residential buildings (8%), and construction projects (6%) (IEA, 2019). Digital twin technology is identified as a key innovation for improving building operations, particularly through optimizing ventilation systems which play a crucial role in maintaining indoor air quality and constitute a substantial part of building energy usage.

Recent advancements in computer science have led to the integration of digital technologies like Building Information Models (BIM) and their advanced form, Digital Twins, into building management (Lu et al., 2020). Digital Twins, particularly when combined with sensor data, enhance building energy efficiency through various methods, including model predictive control (MPC) (Smarra et al., 2018) and facilitating real-time energy-saving decisions (Agostinelli et al., 2021). These technologies enable precise and efficient modelling of a building's energy systems. A significant challenge in this field is modelling ventilation systems that must account for the variability in control actions due to different conditions such as seasonality and manual adjustments (Zhang et al., 2022). This research focuses on developing a Digital Twin for the OU44 building at the University of Southern Denmark. The building serves as a live laboratory and is equipped with CO_2 sensors and air diffuser damper position sensors. The study combines data-driven and physics-based methods to precisely estimate energy consumption and conditions in specific rooms. This Digital Twin platform offers real-time data on ventilation performance and energy utilization.

Part of
Pernigotto, G., Ballarini, I., Patuzzi, F., Prada, A., Corrado, V., & Gasparella, A. (Eds.). 2025. Building simulation applications BSA 2024. bu,press.
https://doi.org/10.13124/9788860462022

The development of the digital twin uses the *Twin4Build* framework (Bjørnskov et al., 2023), which supports grey-box energy modelling of various system components. Additionally, this research investigates continuous monitoring and anomaly detection, addressing the challenges of data collection and system integration in older systems. The findings highlight the profound impact of Digital Twin technology in building management, showcasing through the OU44 case study how such technologies can improve decision-making, enhance monitoring capacity, and adapt to other buildings equipped with basic sensors, demonstrating their broad applicability in the field.

2. Grey Box, Ontology-Based Modelling

Grey-box modelling merges theoretical knowledge and data-driven methods to produce models that balance comprehensibility with accurate system dynamics representation, especially useful for partially understood systems. It incorporates both partial theoretical knowledge and empirical data. Ontology-based modelling enhances this approach by using structured frameworks called ontologies to organize information and define relationships specific to the domain, thereby improving interoperability and facilitating knowledge reuse across different applications. The digital twin concept organizes these models and tools to serve physical asset designers and operators, defining a digital twin as a collection of digital models that predict specific outcomes, supported by a data-acquisition system for real-time physical-digital interaction. A diagram of this model-asset interaction is depicted in Figure 1. The Twin4build framework leverages the *SAREF* core ontology and its extensions *SAREF4BLDG* and *SAREF4SYST* (SAREF, 2020).

It is composed of 5 main classes which provide a streamlined modelling flow:

- Model: Represents the simulation model, composed of 1 or more component models.
- Simulator: Simulates a model instance for a given period.
- Monitor: This class offers methods to analyse estimated and measured data to assess performance and detect anomalies.
- Evaluator: Can evaluate and compare different Model instances based on user-defined quantities of interest.
- Estimator: Provides methods for performing parameter estimation and sensitivity analysis using a model instance and user-defined parameters.

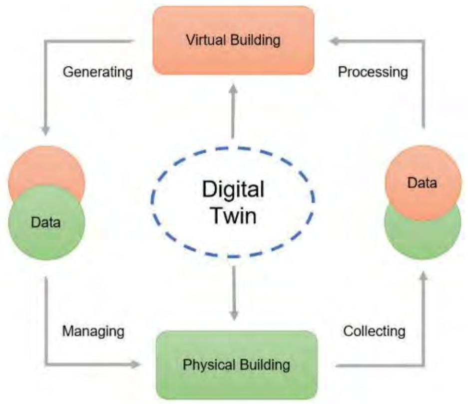

Fig. 1 – Digital twin diagram. It uses sensor data in tandem with digital models to provide a variety of services for building operators and designers (Jradi and Bjørnskov, 2023)

3. Methodology

Figure 2 shows the abstract process carried out to develop the ventilation model. The methodology encompasses an initial phase of characterization and data acquisition, followed by the application of historical data to train a machine learning algorithm. This algorithm is designed to replicate the rule-based control actions for each ventilated room. Subsequently, the approach employs historical data in conjunction with model estimates to conduct parameter estimation and refine the model's accuracy through fine-tuning.

The next sections show a general diagram of the model, with inputs and outputs. Followed by a description of a room model which is replicated for each room inside the ventilation model, the room model contains a controller and a damper model. The controller model attempts to mimic the heterogenous behaviour of the rule-based controller that controls de damper position of each room.

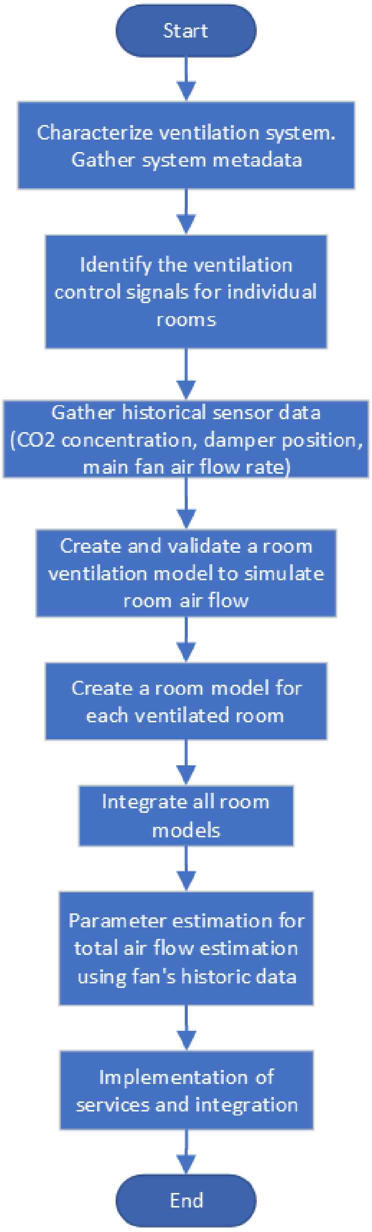

Fig. 2 – General modelling methodology

3.1 Ventilation System Model

This study considers one of the 4 nearly identical ventilation systems in the OU44 building. This subsystem provides air to a quarter of the areas on the building, distributed in 20 rooms with demand-controlled ventilation (DCV) and other 10 auxiliary, small rooms like bathrooms and copy rooms with a constant air volume (CAV). These rooms span all 4 levels of the building.

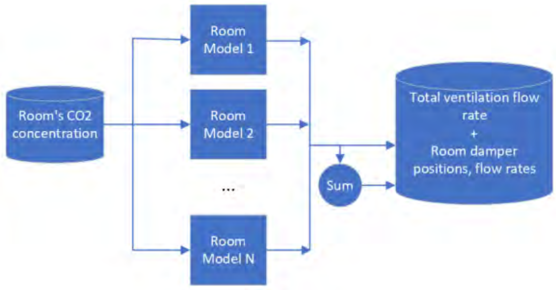

Fig. 3 – Ventilation system model

Figure 3 illustrates a model of the ventilation system. The facility's Building Management System (BMS) monitors CO_2 levels continuously, capturing real-time data every minute from rooms with DCV. This data informs the adjustment of damper positions in each room, which in turn controls the airflow according to a model specific to each room. These airflow calculations are based on the predetermined nominal flow rates designated for each room.

3.1.1 Room model

The individual room model consists of a CO_2-based controller which uses both CO_2 and time information to provide a damper opening position signal.

Fig. 4 – Room ventilation model

The damper model estimates the air flow through the damper as a function of the damper position. Using the model by Huang (Huang, 2011) in its constrained form, the model has two 2 main parameters: a corresponding to a unitless air damping coefficient and $\dot{m}_{a,max}$ which corresponds to the maximum flow rate of the room damper. The damper's air mass flow rate is described by the equation 1, where the constrains for the coefficients c and b are given by Equations 2 and 3.

$$\dot{m}_a = ae^{bu_d} + c \quad (1)$$

$$c = -a \quad (2)$$

$$b = ln\left(\frac{\dot{m}_{a,max}-c}{a}\right) \quad (3)$$

Where:
$\dot{m}_{a,max}$ is the damper's nominal air mass flow rate (kg/s)
a is a damper coefficient (dimensionless).
u_d is the damper opening position [0-1] (percentage)
The maximum air flow ($\dot{m}_{a,max}$) rate for each room is taken from the design values given in the buildings blueprints. A fine-tuned value is used for the second parameter (a) of the model.

3.1.2 Controller model

One prevalent challenge encountered in modelling ventilation systems of buildings, years after they commence operations, stems from the inconsistency in the control strategy across all rooms within the building, which can fluctuate throughout the year. This variability, compounded by the lack of comprehensive historical data on adjustments and configurations applied to the control system over time, often results in the inadequacy of simple rule-based controllers' models to accurately predict the ventilation system's actual behaviour across the entire building. To address this issue, a data-driven methodology was employed, leveraging artificial neural networks, as shown in Figure 5.

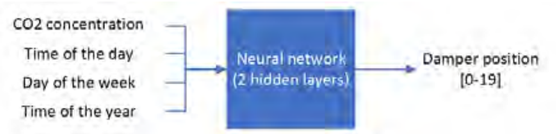

Fig. 5 – Room ventilation controller model

This approach uses both CO_2 concentrations and the rooms' damper positions to train an ANN model which attempts to mimic the behaviour of the rule-based controller assigned to each room.

3.1.3 Data embeddings

To capture the configuration of the ventilation modes for each one of the rooms and to validate the controlling signals each one of the 3 main quantities were adapted for its use in the neural network model. Equation 4 presents the discretization used for the damper position value. A continuous value from 0 to 100 is converted to 20 discrete classes with an integer from 0 to 19 for each. This is the output of the neural network.

$$d = round(\frac{D}{100} \cdot 19) \quad (4)$$

Where d is the discrete value for the damper position used as output of the neural network and D is the original data point [0,100].

Equation 5 describes the normalized CO_2 concentration, with the original values in ppm, a gaussian normalization (Z-score normalization) is made:

$$z = \frac{CO_2 - \mu_{co2,i}}{4\sigma_{co2,i}} \quad (5)$$

Where the mean and standard deviation are calculated with all the data points of each room (i) corresponding to the year 2023. The denominator includes four times the sigma value to make it less sensible to outliers. Additionally, making the values smaller and with a 0-mean.

The time variables were divided in three categories: Day of the year, day of the week and time of the day. This division allows the model to have a notion of the effects of seasons, weekdays and day-night cycles which are typically used to define ventilation rule-based controls. To provide insight into the cyclic nature of the first two time-variables, the cyclic embedding presented in Equations 6 and 7 was used.

$$x_{sin} = \sin\left(\frac{2\pi x}{P}\right) \quad (6)$$
$$x_{cos} = cos\left(\frac{2\pi x}{P}\right) \quad (7)$$

Where x are the original time of day and day of year cyclic features, P is the period of the feature (24 for time of day and 365 for day of year). This representation produces 2 variables per cyclic feature that are fed to the neural network as inputs.
The day of the week feature is encoded using one-hot-encoding.

Table 1 – Hyperparameters used for the neural networks representing the ventilation controller for each room.

Hyperparameter	Value
Input size	12
Output size	20
Hidden layer 1 size	50
Hidden layer 2 size	100
Learning rate	0.001
Number of epochs	15
Batch size	64

4. Case Study

A case study using sensor data from a living lab building from the University of Southern Denmark is considered for the implementation of the developed model. This study focuses on the ventilation system of the building. The ventilation system's digital twin aims to provide continuous commissioning services, anomaly detection and insights into the power consumption of the ventilation fan through the estimation of the total air flow rate and room ventilation conditions.

4.1 OU44 Building

Presented in Figure 6, the OU44 building of the University of Southern Denmark is a multi-purpose building equipped with different sensors throughout the building. Specifically, the main rooms and offices measure indoor CO_2 concentrations, a signal that is used in the demand-controlled ventilation system of the building.
The building has 4 nearly identical ventilation subsystems which operate independently. Each one of these is comprised of an Air Handling Unit, ventilation ducts and room air dampers controlled by a centralized controller.

Fig. 6 – Façade of the UO44 building, SDU Odense

The ventilation system, depicted in figure 7, includes a central air handling unit (AHU) that provides demand-controlled ventilation (DCV) to 20 study and multi-purpose rooms, and constant ventilation to 10 auxiliary rooms. The system is managed by a Building Management System (BMS), which controls the ventilation and aggregates data from CO_2 and temperature sensors in each of the 20 main rooms.

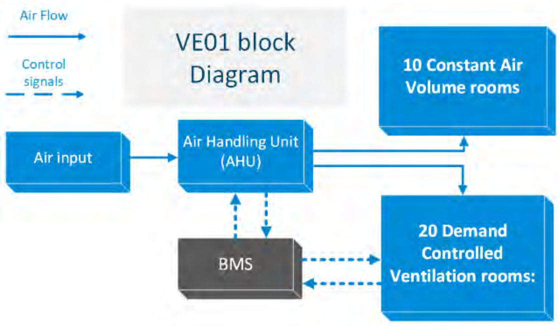

Fig. 7 – Block diagram of one of the ventilation subsystems in UO44, VE01

5. Results

5.1 Controller Model

The data-driven approach attempts to capture the rule-based controller behaviour by mimicking the measured controlling actions without having to manually craft a set of controlling rules for each room in the ventilation system. The accuracy of the Artificial Neural Network (ANN) models is evaluated by comparing them to the actual measured positions of dampers in each room. This accuracy is calculated as the percentage of instances where the ANN model's output precisely matches the discretized position of the damper at every timestamp in the test data. The results are presented in Figure 8.

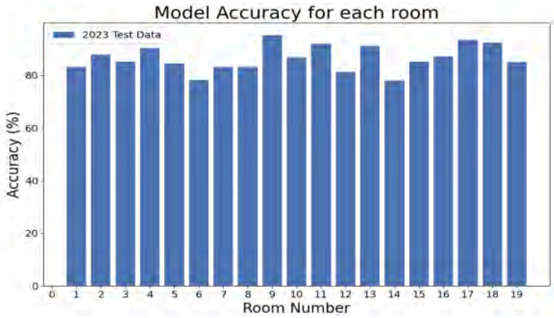

Fig. 8 – The accuracy of the Artificial Neural Network (ANN) models

The model training process starts with the data embedding described in section 3.1.3 where the CO2 concentration and time data are pre-processed. Then the data is split into training and test datasets in the following process: Out of the data for the whole year of 2023, the first three weeks of each month are used for training and the remaining days of each month for testing. Additionally, the months of January and February 2024 are used for validation.

The sum of all estimated air mass flows for each one of the rooms is shown in the Figure 9, where the continuous blue line represents sensor data, and the black dotted line are the simulation results for the total air mass flow of the AHU.

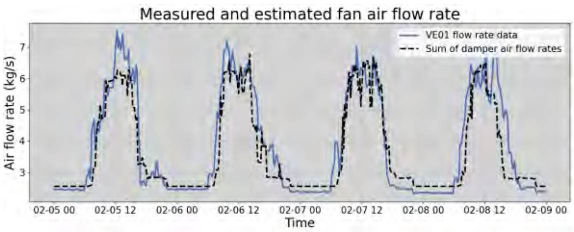

Fig. 9 – ANN-based controller estimation for a week during the month of February 2024, RMSE: 0.6581 kg/s Correlation: 0.9335

The accuracy of the ANN-based controllers is evaluated by comparing two key measurements: the Root Mean Square Error (RMSE) calculated using actual measured positions of the dampers and the RMSE calculated using the damper position signals provided by the ANN controllers. This comparison helps assess the error introduced by the ANN controllers. Table 2 presents the comparison of the RMSE obtained when estimating the total system airflow when the control signal is the measured damper position and the estimated air flow using the position signal from the ANN controller model. RMSE calculated with all validation data from 2024.

Table 2 – RMSE and correlation coefficients of mass air flow rate with and without ANN controllers.

Control signal	RMSE [kg/s]	Correlation
Measured	0.6910	0.9684
ANN control	0.6581	0.9335
Difference	0.0329	0.0349

5.1.1 Share of air flow from rooms without DCV

The proportion of total airflow to rooms without demand-controlled ventilation was determined by analysing estimated airflow values against measurements from the air handling unit (AHU). The Root Mean Square Error (RMSE) for the initial dataset, covering January 2023, was determined to be 1.566 kg/s. This RMSE is considered indicative of the constant airflow volume in rooms equipped with continuous ventilation. The accuracy of this estimation was further assessed by comparing it against data collected in subsequent periods, as shown in Figure 10.

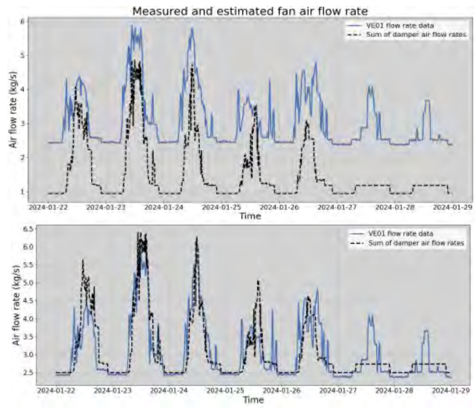

Fig. 10 – (Top) Original air flow rate estimation. RMSE: 1.6468 Correlation: 0.8241. (Bottom) Adjusted air flow rate estimation: RMSE: 0.5045 Correlation: 0.8241

5.2 Continuous Monitoring

First, the digital twin is implemented for continuous commissioning service of the ventilation system case study. In this regard, the model enables the tracking and identification of irregularities in both the overall ventilation system and within each individual room. This is done by continuously comparing an estimated signal with its measured counterpart, calculating a moving average of the error between the two signals and defining a threshold that would trigger an anomaly signal if surpassed. Figure 11 shows the continuous monitoring of the ventilation system's main air flow rate.

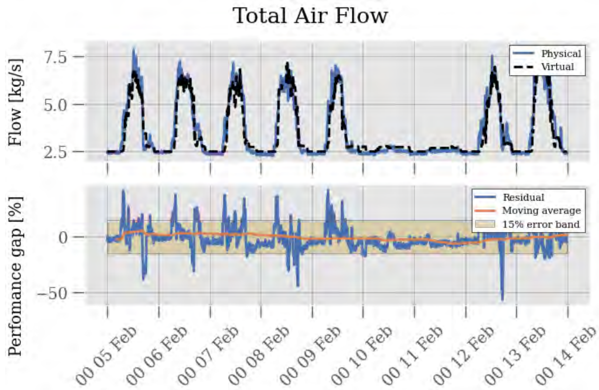

Fig. 11 – Continuous monitoring of total air handling unit inlet air flow rate from the 5th to the 14th of February 2024

The bottom plot shows the anomaly signal, which is calculated by identifying deviations between the expected and actual damper positions or airflow rates, an anomaly is considered when the error average exceeds 15% for total airflow and 20% for damper positions. In Figure 12, a detailed examination of the monitoring models for all these rooms revealed that the expected control signal deviated in behaviour for 4 rooms. Analysing further, for one of the offices with anomalies, it could be seen that the control signal was manually set to maintain a fully open damper position during work hours, disregarding CO_2 levels. This setup was modified in the following year, as depicted in Figure 13, to allow for adjustments in response to measured CO_2 levels, thereby optimizing ventilation performance and indoor air quality.

6. Conclusion

This study utilizes grey-box modeling and ontology-based methods to blend empirical data with theoretical insights for predicting and simulating ventilation system performance under various scenarios.

Fig. 12 – Anomaly signals for the rooms of the ventilation system

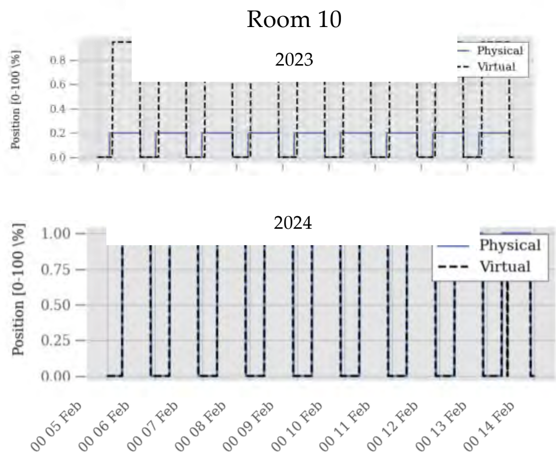

Fig. 13 – Anomalous change in the damper controller for Room 10. The behaviour of the controller is different compared to the same period in the previous year

This approach achieves a balance between interpretability and accuracy, optimizing operations without fully depending on complex theoretical processes. It replicates rule-based controllers in larger buildings by training a neural network with control signals, adjusting to cyclic patterns and periodic control strategy changes efficiently. Although effective, it can inadvertently include isolated configuration changes, which could be mitigated with larger datasets.

This research also underscores the challenges in data collection and integration within live buildings, pointing out the difficulties of merging various subsystems into a unified digital twin platform. Future efforts involve incorporating variables such as temperature and humidity to provide a more comprehensive understanding of system performance. Additionally, a promising field of study revolves around exploring forecasting ventilation power consumption and investigating optimization strategies.

Acknowledgements

This research was carried out under the "*DanRETwin: A Digital Twin solution for optimal energy retrofit decision-making and decarbonization of the Danish building stock*" project, funded by the Danish Energy Agency (*Energistyrelsen*) under the Energy Technology Development and Demonstration Program (EUDP), ID number: 640222-496754.

References

Agostinelli, S. Cumo, F. Guidi, G. Tomazzoli, C. 2021. "Cyber-Physical Systems Improving Building Energy Management: Digital Twin and Artificial Intelligence" *Energies 2021, Vol. 14, Page 2338*. https://doi.org/10.3390/EN14082338

Bjørnskov, J. Jradi, M. 2023. "An ontology-based innovative energy modeling framework for scalable and adaptable building digital twins." *Energy and Buildings*. https://doi.org/10.1016/j.enbuild.2023.113146

Cespedes-Cubides, A.S., Jradi, M. "A review of building digital twins to improve energy efficiency in the building operational stage". *Energy Inform* 7, 11 (2024). https://doi.org/10.1186/s42162-024-00313-7

European commission. 2018. "Energy performance of buildings directive." European commission. Accessed February 2024. https://energy.ec.europa.eu/topics/energy-efficiency/energy-efficient-buildings/energy-performance-buildings-directive_en

Huang, Gongsheng 2011. "Model predictive control of VAV zone thermal systems concerning bi-linearity and gain nonlinearity." *Control Engineering Practice*. DOI: https://doi.org/10.1016/J.CONENGPRAC.2011.03.005

Jradi, M., Bjørnskov, J. 2023. "A Digital Twin Platform for Energy Efficient and Smart Buildings Applications." *2023 Fifth International Conference on Advances in Computational Tools for Engineering Applications (ACTEA)*. https://doi.org/10.1109/ACTEA58025.2023.10194071

Lu, Q. Xie, X. Parlikad, A. Schooling, J., M. Konstantinou, E. 2020. "Moving from building information models to digital twins for operation and maintenance." *Proceedings of the Institution of Civil Engineers: Smart Infrastructure and Construction*. https://doi.org/10.1680/JSMIC.19.00011

SAREF. 2024. "The Smart Applications REFerence Ontology (SAREF)" SAREF. Accessed February 2024. https://saref.etsi.org/

Smarra, F. Jain, A. de Rubeis, T. Ambrosini, D. D'Innocenzo, A. Mangharam, R. 2018. "Data-driven model predictive control using random forests for building energy optimization and climate control." *Applied Energy*. https://doi.org/10.1016/J.APENERGY.2018.02.126

Zhang, Chenchen Sun, Hejiang 2022. "Analysis and Modeling of Mechanical Ventilation Operation Behaviors of Occupants in Cold Regions of North China" *Applied Sciences 2022, Vol. 12, Page 5143*. https://doi.org/10.3390/APP12105143