OPEN ACCESS

ARTICLE

Journal homepage: www.the-innovation.org/energy-use

Energy Use

INNOVATION PRESS

Energy Use 1 (2026):100039

Contents lists available at INNOVATION PRESS

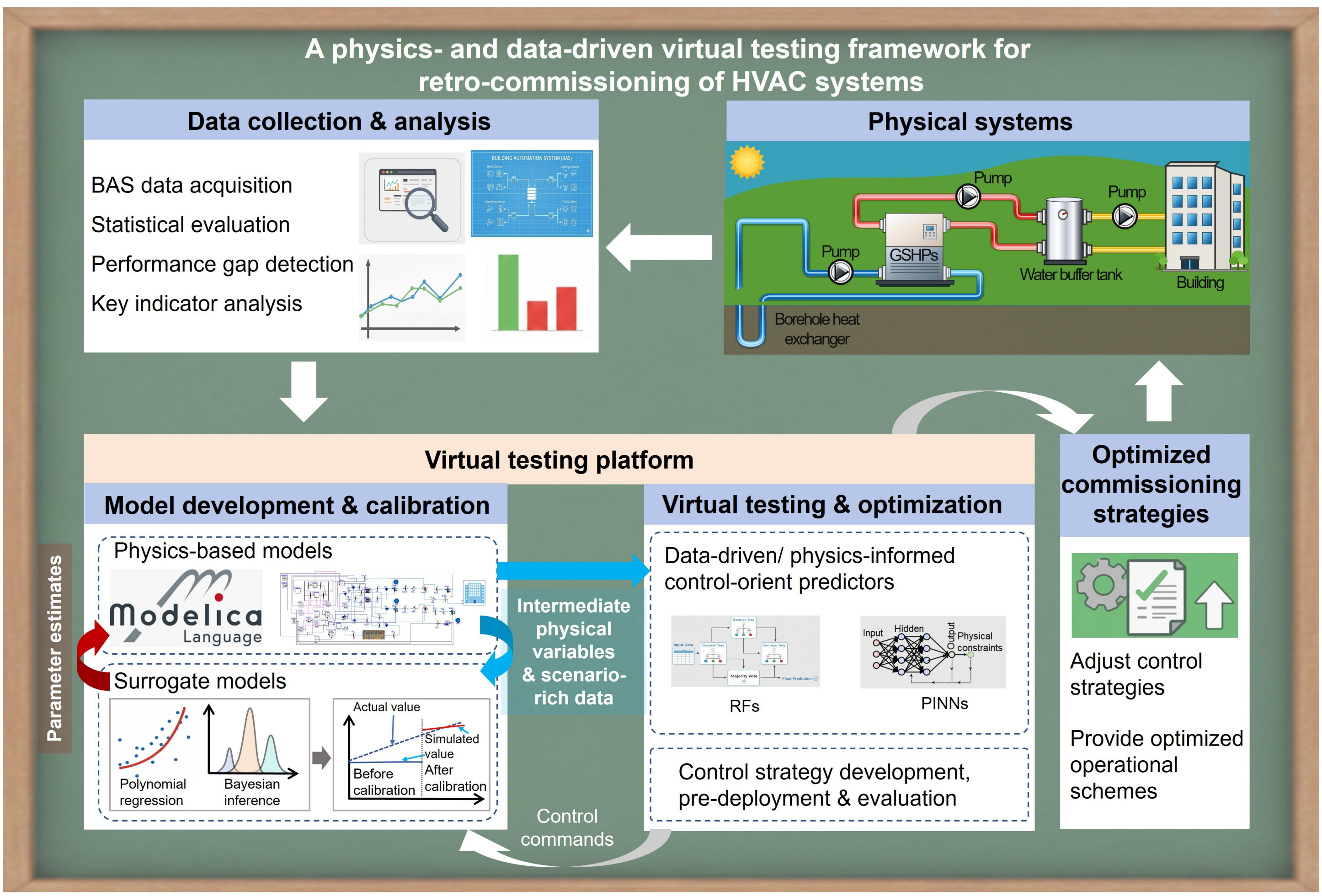
From Static to Adaptive: A Physics and Data Driven Virtual Testing Framework for Retro-Commissioning of HVAC Systems

Junjian Fang,[1] Ruchi Choudhary,[2] Anurag Goyal,[3] and Chaoqun Zhuang[1,2,*]

*Correspondence: zhuangcq@njtech.edu.cn

Received: February 5, 2026; Accepted: March 20, 2026; Published Online: March 25, 2026; https://doi.org/10.59717/ipj.energy-use.2026.100039

GRAPHICAL ABSTRACT



PUBLIC SUMMARY

- Calibrated physics-based model provides physical variables for control development.
- Bayesian calibration updates the parameters of the physics-based model within the virtual testbed.
- Optimized GSHP control cuts energy use by 16.1% and increases COP from 3.0 to 4.4.
- Virtual evaluation indicates a 14.6% reduction in electricity costs without hardware changes.

ARTICLE

Journal homepage: www.the-innovation.org/energy-use

Energy Use

INNOVATION PRESS

Energy Use 1 (2026):100039

Contents lists available at INNOVATION PRESS

From Static to Adaptive: A Physics and Data Driven Virtual Testing Framework for Retro-Commissioning of HVAC Systems

Junjian Fang,[1] Ruchi Choudhary,[2] Anurag Goyal,[3] and Chaoqun Zhuang[1,2,*]

[1]College of Urban Construction, Nanjing Tech University, No.200, North Zhongshan Road, Nanjing 210009, China

[2]Energy Efficient Cities Initiative, Department of Engineering, University of Cambridge, Cambridge CB2 1PZ, UK

[3]Department of Mechanical Engineering, Indian Institute of Technology Delhi, Hauz Khas, New Delhi 110016, India

*Correspondence: zhuangcq@njtech.edu.cn

Received: February 5, 2026; Accepted: March 20, 2026; Published Online: March 25, 2026; https://doi.org/10.59717/ipj.energy-use.2026.100039

Citation: Fang J., Choudhary R., Goyal A., et al. (2026). From Static to Adaptive: A Physics and Data Driven Virtual Testing Framework for Retro-Commissioning of HVAC Systems. *Energy Use* **2**:100039.

Traditional one-time handover commissioning often fails to identify operational inefficiencies that emerge during long-term operation of heating, ventilation, and air-conditioning (HVAC) systems, while on-site commissioning remains labor-intensive and less adaptable to changing operating conditions. This study proposes a physics- and data-driven virtual testing framework to enable virtual retro-commissioning and adaptive control optimization. A calibrated Modelica-based system model is developed as a virtual testbed, where data-driven/physics-informed predictors are used to support unit operation optimization and preheating control. Physics-based representation provides scenario-rich simulations and intermediate physical variables, while the data-driven components translate these insights into fast, control-oriented predictors. The framework is demonstrated on an underperforming ground source heat pump (GSHP) system. Results show that the optimized control strategy reduces energy consumption by 16.1% while maintaining indoor thermal conditions comparable to those under the existing strategy, and increases the operational coefficient of performance (COP) from 3.0 to 4.4. Virtual evaluation over a representative 14-day winter period indicates a 14.6% reduction in electricity costs without hardware modifications. The proposed framework enables low-risk, model-based retro-commissioning and provides a practical pathway for adaptive performance optimization of existing HVAC systems.

INTRODUCTION

Building operations account for approximately 30% of global energy demand and a substantial share of carbon emissions,[1] of which heating, ventilation, and air-conditioning (HVAC) systems typically represent 20-40% of total building energy use.[2] With population growth, increasing comfort requirements, and accelerating electrification, HVAC energy use is expected to continue rising.[3] Improving HVAC operational efficiency and reducing operational emissions are therefore critical for lowering building energy use and mitigating climate change. However, in practice, HVAC systems often deviate from design expectations due to suboptimal control, setpoint drift, and equipment or sensor biases, which can prevent the expected energy savings from being achieved.[4] This underscores the need for systematic building commissioning to improve operational performance.

Building commissioning is a key quality-assurance process that ensures HVAC systems operate as intended and meet performance requirements. Compared with hardware retrofits, commissioning relies mainly on time and labor, yet it can deliver substantial energy savings. Existing studies suggest that commissioning of operational buildings typically yields average energy savings of approximately 14%, while savings above 50% have been reported in specific cases.[5] Commissioning has also been adopted by green building rating systems (LEED[6]) and codified in standards and guidelines (ASHRAE 202[7] and ISO/CEN standards[8-9]). However, current commissioning practices are primarily concentrated at the building handover stage, while performance tends to drift over time during operation, leading to a gradual loss of the intended energy efficiency.[5] In large-sample studies, measured operational energy use is often 10%-30% higher than commissioning-based baseline estimates,[10] indicating the limitations of one-time commissioning in ensuring sustained operational performance.

To address this performance drift after handover, commissioning concepts have been extended toward retro-commissioning for existing buildings. Retro-commissioning uses operational data measured in situ over time, together with periodic assessment, to detect and correct operational deviations, thereby maintaining near-optimal system performance under changing conditions.[11] It is often implemented through continuous fault detection and diagnosis to improve energy efficiency and reduce operating costs.[12] However, this framework cannot be applied to evaluate the impact of control adjustments on performance and efficiency without costly and disruptive on-site experiments.

Physics-based simulation models can therefore support retro-commissioning and virtual evaluation of control adjustments. For instance, EnergyPlus, TRNSYS, and Modelica are widely used for building energy analysis, system performance assessment, and commissioning studies.[13] In practice, however, installation deviations, equipment aging, and evolving control strategies can induce gradual performance drift, leading to an increasing mismatch between models calibrated against design or nominal specifications and actual system behavior.[14] When model parameters are properly calibrated, these models can reproduce key thermal behaviors of real systems and can be used to evaluate alternative control strategies or equipment configurations.[15] Digital twin frameworks for HVAC and energy systems are increasingly employing continuous calibration and model updating to optimize performance and detect faults.[16-17] However, maintaining the validity of physics-based simulation models under performance drift remains a key challenge even in digital twin settings.[18]

Furthermore, retro-commissioning requires not only the capability to develop, test, and evaluate control strategies but also to support their deployment in real-world operation. In practice, this capability is rarely realized because of control modelling limitations of physics-based models.[19-20] Physics-based models are therefore often reduced to linear or reduced-order models for model predictive control,[21] or used as offline testbeds to screen candidate control changes under prescribed scenarios.[22] They offer the advantage of generating data with diverse scenarios across a wide operating

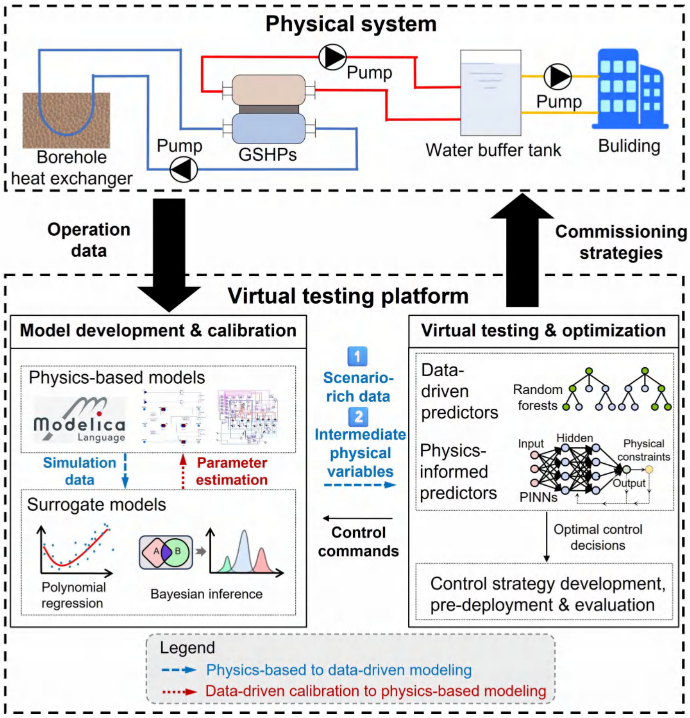


Figure 1. Proposed virtual testing framework for retro-commissioning.

range, including conditions that are rarely seen during regular operation.[23]

Most buildings, though, still rely on manual schedules and rule-based sequences of operation that often perform poorly under varying loads and operating conditions, leading to suboptimal energy performance.[24] Recent advances in sensing and data acquisition have enabled a range of purely data-driven control models, including random forests,[25] artificial neural networks,[26] support vector machines,[27] and deep neural networks.[28] As aforementioned, the deployment of these approaches for commissioning and control evaluation remains constrained by the need for large, representative datasets and by limited generalization.[29]

Variants of physics-informed machine learning[30] have recently emerged to improve robustness by incorporating physical knowledge into the machine learning process, by incorporating physical constraints, improving consistency and generalization across operating scenarios. Moreover, the simulation data can enhance limited field measurements and offer reliable training datasets for data-driven or physics-informed control decisions.[31] Additionally, these models can reveal internal system states (such as equipment-level operating conditions or intermediate thermal states), which are not directly measurable on site but are essential for analysis focused on control.[32-33] However, these approaches are typically developed for controller design, with less emphasis on commissioning-oriented applications in existing systems.

Several virtual testing platforms have been developed for control evaluation and model development in HVAC systems. Representative examples include BOPTEST, which provides a simulation-based benchmarking framework with common test cases and Modelica-based emulators for HVAC control evaluation.[22] Sinergym offers an open-source Python-based virtual testbed for large-scale simulation and continuous control.[34] AlphaDataCenterCooling is a high-fidelity virtual testbed for benchmarking operational strategies in data center cooling plants.[35] Related studies have also used existing testbeds for controller comparison,[36] developed IoT-oriented simulation-to-deployment environments,[37] built co-simulation testbeds for automated calibration,[38] and proposed digital-twin platforms for monitoring and decision support39. Collectively, these studies demonstrate the value of virtual environments for control evaluation, strategy development, and risk-reduced testing.

However, limited attention has been paid to how such platforms can better support the development and evaluation of control strategies with stronger physical consistency, robustness, and generalization capability. Moreover, in real operational systems with limited field measurements and incomplete design information, particularly when actual performance has drifted from original design conditions, how to calibrate the physics-based model of a virtual testing platform to achieve sufficient accuracy for reliable strategy evaluation and retro-commissioning has not yet been systematically addressed.

This study therefore proposes a virtual testing framework for retro-commissioning that integrates physics-based system modeling with calibration and control-oriented evaluation. The framework is demonstrated using a ground source heat pump (GSHP) system as a representative case. Within this framework, the virtual testbed features a two-way interaction between physics-based and data-driven modeling: i) The physics-based model provides intermediate physical variables for data-driven and physics-informed modeling. These variables are used as inputs to optimization-oriented control algorithms and for physical-constraint embedding, thereby improving the physical consistency, robustness, and generalizability of the virtual testing platform. ii) Meanwhile, for real operational systems with limited measurements and incomplete information, particularly when actual performance has drifted from original design conditions, Bayesian calibration based on field data is used to update the physics-based model parameters. This ensures sufficient accuracy for reliable control development, strategy evaluation, and retro-commissioning.

MATERIALS AND METHODS

The proposed virtual testing framework for retro-commissioning extends conventional one-time commissioning into an iterative process that includes monitoring, model calibration, performance analysis, and control strategy assessment. The calibrated physics-based model serves as a virtual representation of the real system, providing scenario-rich simulation data and intermediate physical variables that are not directly measurable on site. These outputs are then used to support the development of data-driven and physics-informed predictors for control strategy assessment and optimization. Figure 1 illustrates the proposed virtual testing framework, which is demonstrated in this study using a GSHP system as a representative case. Data collected from the building management system (BMS) are used for statistical analysis and performance evaluation to characterize existing control modes, identify key factors influencing energy efficiency, and diagnose potential sources of operational inefficiency.

The framework consists of two main stages. The first stage focuses on model development and calibration. A physics-based dynamic model is constructed to represent the whole system and calibrated to identify equipment performance characteristics and key thermal parameters from operational data, aligning the physics-based model with observed system behavior.

The second stage performs virtual testing and optimization of control strategies. The calibrated physics-based model serves as a virtual testbed that generates physically consistent and scenario-rich simulation data, extending beyond operating conditions typically observed during on-site commissioning and field operation. This enables systematic coverage of the operating conditions for strategy assessment under both typical and extreme scenarios. In addition, the calibrated physics-based model provides access to intermediate physical variables that are not directly measured by on-site sensor configurations, including internal operational states such as refrigerant-side thermodynamic quantities. Together, these enriched datasets support the development of data-driven/physics-informed predictors for control-oriented prediction and strategy development.

Depending on the prediction task, different control-oriented predictors can be adopted. Purely data-driven predictors[40-41] are more appropriate for system-level variables dominated by external disturbances, where predictions rely primarily on scenario-rich simulation data. Outputs from the calibrated physics-based model can also be used to construct training labels when these labels are difficult to obtain from field measurements. In contrast, physics-informed predictors[42] are used for equipment-level performance prediction, where scenario-rich simulation data and intermediate physical variables from the calibrated physics-based model are incorporated to impose physics-based constraints and enforce physical consistency. Based on these predictive models, candidate control strategies are formulated and evaluated in terms of energy efficiency, operational feasibility, and operational risks prior to field implementation. The results of the virtual testing are

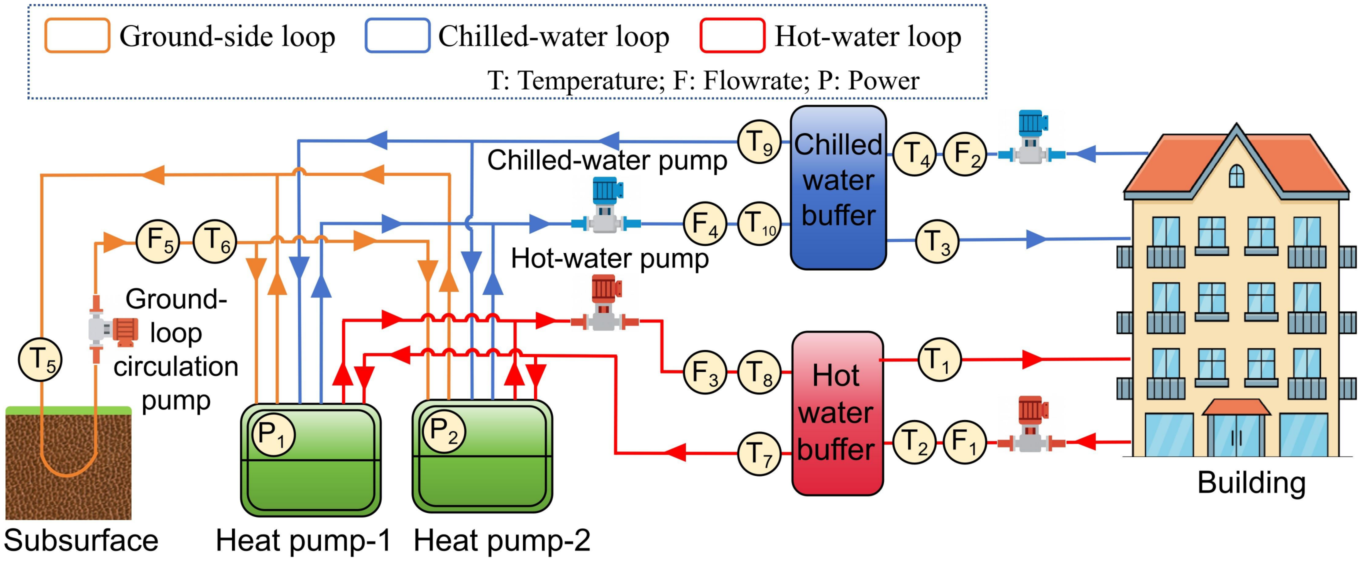


Figure 2. Schematic diagram of the GSHP system.

then used to derive commissioning recommendations for control adjustments.

The detailed procedures for physics-based modeling and calibration, as well as virtual testing and optimization, are presented in next section.

ILLUSTRATIVE STUDY

System description

The proposed physics- and data-driven virtual testing framework is demonstrated through a retro-commissioning case study of a GSHP system in an office building in the United Kingdom. The schematic configuration of the GSHP system is shown in Figure 2. The system consists of two identical four-pipe GSHP units connected in parallel, which can provide simultaneous heating and cooling. Each GSHP has a nominal cooling capacity of 148 kW and a nominal heating capacity of 160 kW. The GSHPs are connected through a common supply and return manifold to 24 vertical borehole heat exchangers, each with a diameter of 10 cm and a depth of 150 m. A 20% glycol-water solution circulates through the ground-loop as the heat transfer fluid. The system is also equipped with two buffer tanks, each with a capacity of 1500 L, to provide hydraulic separation and ensure stable system operation.

The configuration of the sensor monitoring system is generally consistent with that of a typical GSHP system. The monitored variables include the supply and return water temperatures of the building's heating and chilled-water circuits (T_1, T_2, T_3, T_4), ground-loop inlet and outlet temperatures (T_5, T_6), the condenser water inlet and outlet temperatures (T_7, T_8), the evaporator water inlet and outlet temperatures (T_9, T_{10}), the main circuit flow rates (F_1, F_2, F_3, F_4, F_5), and the operating power of each heat pump (P_1, P_2). All measurements are recorded by the BMS at 15-minute intervals. Variables such as T_1–T_6 and F_1–F_5 are measured mainly at the system headers. Although the monitoring coverage meets industry standards, several key variables are not recorded at the individual unit level. These include the inlet and outlet temperatures and flow rates on the chilled-water, hot-water, and ground-loop sides of the individual GSHP, as well as indoor air temperature and individual pump power consumption. The absence of these measurements further underscores the importance of calibrated physics-based modeling for accurate system performance assessment and optimization.

Diagnosis using on-site BMS data

On-site BMS data were first analyzed to diagnose the existing operating mode, identify the main operational deficiencies, and motivate the subsequent retro-commissioning tasks. In this study, unless otherwise noted, COP in the BMS-based performance analysis refers to the building-side operational COP, defined as the ratio of the heating delivered to the building to the recorded GSHP power input. Figures 3 A & B present the heating and cooling loads for representative winter and summer days, respectively. The loads are calculated from measured supply and return water temperatures and flow rates in the heating and chilled-water circuits. The heating system operates only during daytime occupancy and shuts down at night, with significant load fluctuations that introduce frequent transient conditions during commissioning. In contrast, the cooling system mainly serves a small data center, resulting in relatively stable and low cooling demand.

Figures 3 C & D present the operating power (P_1, P_2) of the two GSHPs during cooling and heating operation, respectively. Both units operate simultaneously under a parallel load-sharing strategy. During heating, the high and continuous demand leads to nearly equal power output, indicating stable load sharing. During cooling, the demand is low and intermittent, and short-term differences between P_1 and P_2 arise from load fluctuations, start–stop transients, and minimum part-load or operating time constraints, even though both units remain in operation. This leads to prolonged operation at part-load ratios well below the typical efficient operating range of the heat pumps. Given the relatively low cooling demand, the subsequent analysis focuses on heating operation.

As indicated in Figure 3E, the GSHPs operate at a PLR below 40% for 96.1% of the operating hours under heating conditions. Even during the coldest observed periods, the combined load of both units rarely exceeds the rated capacity of a single unit. This operating regime degrades system efficiency. Consequently, as shown in Figure 3F, the building-side operational COP falls below 1.0 during 47.5% of operating hours, because it is calculated from the heat actually delivered to the building, which can be lower than the GSHP power input due to buffer-tank dynamics and inherent system losses. It remains below 2.0 for 56% of hours and below 3.0 for more than 70% of hours. The annual average COP is 2.37, well below the typical range of 4.0-5.0 for properly commissioned GSHP systems. These conditions result in increased energy consumption and operating costs, highlighting the need for improved commissioning and control strategies.

Additionally, the building operates under a fixed occupancy schedule from 8:30 to 21:00, with a room temperature setpoint of 20 °C. To ensure the indoor temperature reaches the comfort threshold (≥ 18 °C) by 8:30, the system applies a fixed 75-min preheat starting at 7:15, irrespective of outdoor conditions. However, during mild weather, this fixed and unnecessarily early preheating often results in substantial energy waste. The combination of static controls and fixed setpoints fails to adapt to real-time heating load variations, further increasing energy consumption.

These deficiencies are addressed in the following sections through virtual testing of preheating scheduling and GSHP unit-number optimization based on the calibrated physics-based model. Table S1 summarizes the main BMS-measured input variables, optimized decision variables in the virtual testbed, and system control variables for the studied GSHP case. Although the specific variables are case-dependent, the same virtual-testing workflow can be extended to other HVAC systems through application-specific selection of

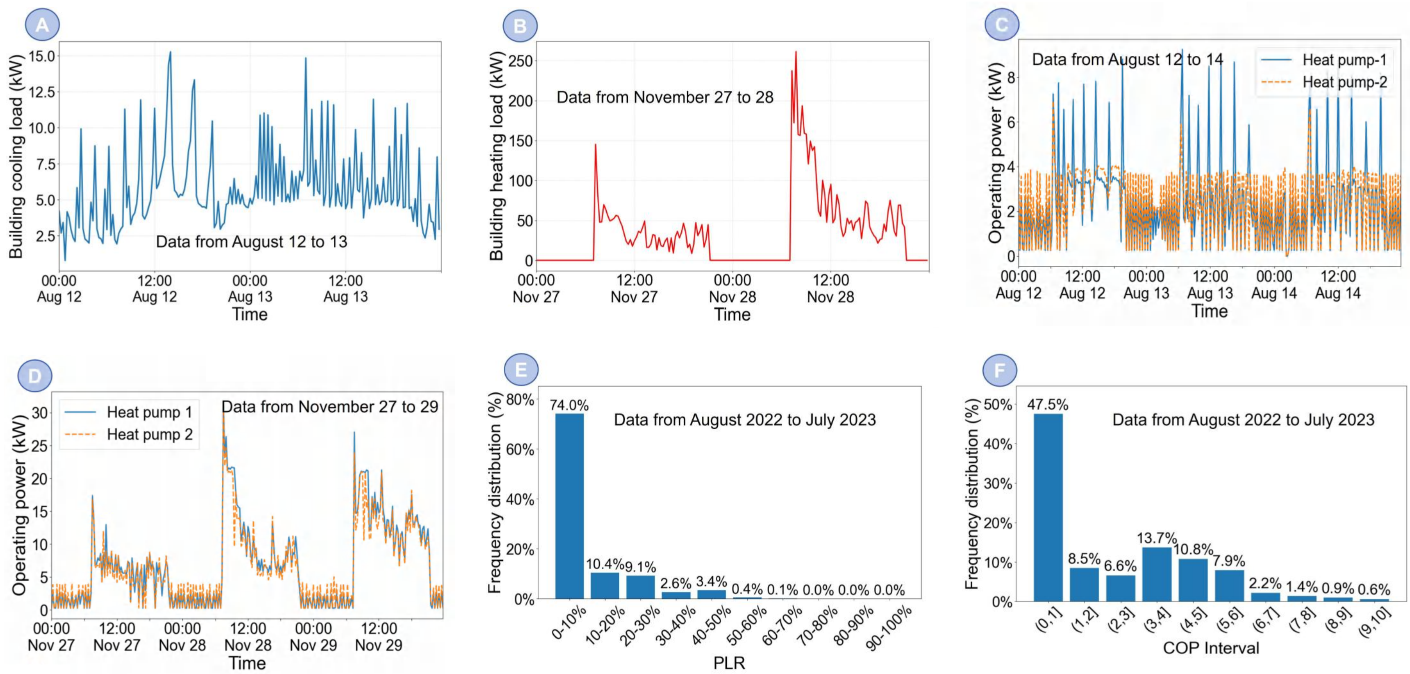


Figure 3. Operating characteristics and performance distribution of the GSHP system.

monitored variables and decision variables.

Model development

A detailed physics-based model is developed using Modelica to represent the GSHP system. The model architecture reproduces the physical configuration of the real system and includes detailed representations of the GSHPs, water buffer tank, ground heat exchangers, and room thermal models.[46] Representative components from the *Buildings* Library including *Storage.Stratified*, *Geothermal.Boreholes.UTube*, and *Movers.Flow Controlled_m_flow*, are directly employed to represent the water buffer tank, borehole heat exchanger, and circulation pumps, respectively. The GSHP unit and room thermal representations are customized based on existing components from the *Buildings* Library, including *HeatPumps.Carnot_y*, *HeatTransfer.Conduction.SingleLayer*, and *MixingVolumes.MixingVolume*, to capture system-specific operational characteristics and control logic. All adopted library components have been previously validated. The resulting subsystems are then integrated to form a system-level GSHP model.

GSHP model. The system follows a commonly used four-pipe GSHP configuration, as illustrated in Figure 4A. It features hydraulically separated hot-water and chilled-water loops, allowing both independent and simultaneous heating and cooling supply in response to building demands. Since the heat pump models directly available in the Buildings Library are mainly based on two-pipe configurations, they cannot directly represent the coupled four-pipe operating mode of the studied system. Therefore, the present four-pipe GSHP model (Figure 4B) is developed by extending the validated two-pipe heat pump model in the *Buildings* Library.[43] Specifically, the GSHP unit is modeled as two thermodynamically coupled two-pipe heat pump sub-models to represent simultaneous heating and cooling operation. One sub-model serves the hot-water loop, while the other serves the chilled-water loop. Each sub-model includes its own evaporator and condenser, and thermal coupling between heating and cooling is achieved through the shared ground-side loop.

Each two-pipe heat pump sub-model embeds the fundamental thermodynamic relations, in which the coefficient of performance (COP) of the GSHP is expressed as a function of the evaporating temperature, condensing temperature, and part-load ratio (PLR). A constant Carnot effectiveness ($\eta_{Carnot,0}$) is introduced for each component and calibrated at the nominal operating point to account for irreversibility associated with finite heat-exchange temperature differences, pressure drops, and compressor entropy generation (Equation 1). The outlet water temperature dynamics are described by Equation (2), from which the evaporating and condensing temperatures are derived (Equations 3–4). Part-load behavior is represented using empirical correction factors (Equation 5), and the GSHP COP of each two-pipe heat pump sub-model is then calculated by Equation (6) on the condenser side, with cooling performance evaluated analogously on the evaporator side.

$$\eta_{Carnot,0} = \mathrm{COP}_0 \cdot \frac{(T_{cn,0} - T_{ev,0})}{T_{cn,0}} \tag{1}$$

$$\tau \frac{Q_{nom}}{\Delta T_{nom}} \cdot \frac{dT_{out}}{dt} = mc_p(T_{in} - T_{out}) + Q \tag{2}$$

$$T_{cn} = T_{out,cn} + \frac{Q_{cn}}{Q_{cn,nom}} T_{App,cn,nom} \tag{3}$$

$$T_{ev} = T_{out,ev} - \frac{Q_{ev}}{Q_{ev,nom}} T_{App,ev,nom} \tag{4}$$

$$\eta_{PLR} = a_1 PLR^2 + a_2 PLR + a_3 \tag{5}$$

$$\mathrm{COP} = \eta_{Carnot,0} \cdot \frac{T_{cn,0}}{(T_{cn,0} - T_{ev,0})} \cdot \eta_{PLR} \tag{6}$$

Where COP_0 denotes the GSHP COP at the nominal operating point; $T_{cn,0}$ and $T_{ev,0}$ are the nominal condensing and evaporating temperatures (K), respectively; τ is the nominal time constant and is set to 60 s in this study; Q_{nom} is the nominal heat transfer rate for the loop (W); ΔT_{nom} is the nominal inlet-outlet temperature difference for the loop (K); T_{out} and T_{in} are the outlet and inlet fluid temperatures (K), respectively; m is the mass flow rate (kg/s); c_p is the constant-pressure specific heat of the fluid (J /(kg·K)); Q is the heat transfer rate between the GSHP and the water loop (W); Q_{cn} and $Q_{cn,nom}$ are the current and nominal condenser heat transfer rates (W), respectively, and $T_{App,cn,nom}$ is the nominal approach temperature on the condenser side (K); Q_{ev} and $Q_{ev,nom}$ are the current and nominal evaporator heat transfer rates (W), respectively, and $T_{App,ev,nom}$ is the nominal approach temperature on the evaporator side (K); η_{PLR} is defined as the ratio of the GSHP COP at part load to the nominal COP; a_1, a_2, a_3 are empirical, dimensionless constants.

Room thermal model. To balance model complexity and calibration difficulty while capturing the overall thermal response of the building, a single-zone room thermal model is adopted in this study. The model is constructed in Modelica using validated components from the *Buildings* Library.[43] Specifically, the indoor air and additional internal thermal mass are represented

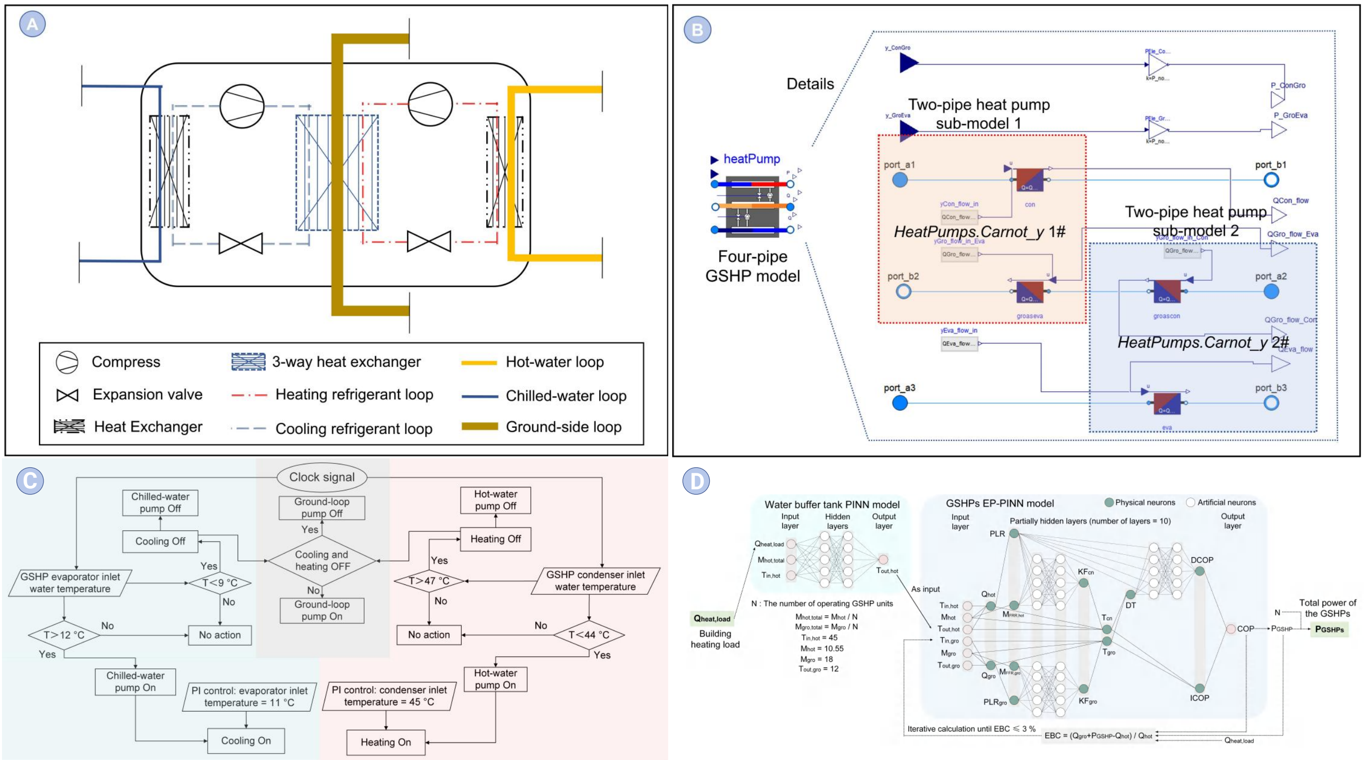


Figure 4. (A) Schematic of the four-pipe GSHP and (B) its equivalent. (C) Modelica-based model Flowchart of the existing control logic for the GSHP system. (D) Coupled physics-informed predictors for buffer-tank and GSHP.

using lumped-parameter modeling, the building envelope is represented by one-dimensional multi-node heat conduction, and the terminal radiator is modeled using a multi-segment hydronic-thermal coupled model. *MixingVolumes.MixingVolume*, *Conduction.SingleLayer,* and *Radiators.RadiatorEN442_2* are used to represent the zone air, envelope heat transfer, and terminal-side heat delivery, respectively. A *Components.HeatCapacitor* is introduced to account for interior thermal inertia not explicitly resolved elsewhere in the model, while valve, hydraulic resistance, and PID control components are included to reproduce terminal-side hydraulic flow regulation and room-temperature control. The zone air is treated as a fully mixed control volume, and its dominant thermal dynamics are described by the zone energy balance in Equation (7).

$$c_{p,air}\rho_{air}V_{room}\frac{dT_z}{dt} = Q_{wall}(t) + Q_{int}(t) + Q_{hot}(t) \tag{7}$$

Where T_z is the room temperature (K); V_{room} is the heating-zone volume (m^3); Q_{wall} is the heat transfer from the envelope to the zone (W); Q_{in} is the internal heat gain (W); Q_{hot} is the heat supplied by the terminal unit (W);$c_{p,air}$ is the specific heat capacity of the air inside the room, taken as 1005 J/(kg·K); ρ_{air} is the air density, taken as 1.2 kg/m^3.

The parameters of the GSHP system-level model are obtained from manufacturer design data, system design documents, on-site measurements, and calibration results. Nominal equipment specifications such as GSHP capacities, design flow rates, buffer tank characteristics, pump flow rates, and borehole configurations are directly adopted from manufacturer data and system design documents. System-level operational measurements, including supply and return water temperatures, flow rates, and GSHP power consumption, are used to derive thermal loads and performance indicators. The design specifications and nominal parameter settings of the system-level model are summarized in Supplementary Table S2 , while the calibrated room-model parameters are presented in next section.

Existing Control Logic. To reproduce the operating behavior of the actual system and calibrate the Modelica-based model, the existing control logic is incorporated into the system-level model. Based on the analysis of operational data, the current control strategy can be summarized as follows: i) All GSHP units start and stop simultaneously, with partial loads distributed equally. ii) A PID controller regulates GSHP power to maintain the chilled-water and hot-water buffer tank temperatures at approximately 11 °C and 45 °C, respectively. Cooling starts when the chilled-water buffer tank temperature exceeds 12 °C and stops when it drops below 9 °C. Heating starts when the hot-water buffer tank temperature falls below 44 °C and stops when it exceeds 47 °C. iii) On the GSHP side, the chilled-water, hot-water, and ground-side circulation pumps operate at fixed speeds, while the building-side pumps use variable-speed control based on differential pressure. The implemented control logic is illustrated in Figure 5.

Model calibration

Within the Modelica physical modeling environment, a calibration framework is developed that combines polynomial regression and Bayesian inference. Specifically, equipment performance parameters with well-defined performance curves are identified using regression-based methods. In contrast, parameters without explicit functional relationships, such as envelope heat transfer coefficients, surface areas, and thermal capacitances, are calibrated using sensitivity analysis and Bayesian inference due to their indirect and highly coupled influence on system outputs.

Polynomial regression. The performance characteristics of equipment can generally be described by polynomial functions, as exemplified by Equation (5). Accordingly, polynomial regression is employed to infer equipment performance coefficients from operational data, thereby effectively characterizing nonlinear efficiency variations of the equipment under different operating conditions.

Bayesian inference. To mitigate excessive model dimensionality, a Sobol sensitivity analysis is first conducted to identify key parameters, followed by Bayesian calibration using operational data. The physics-based model is evaluated under multiple parameter combinations to generate prediction data, which are used to construct a random forest (RF) surrogate.[44] Based on this surrogate, a Sobol global sensitivity analysis is performed to quantify the main and interaction effects of input parameters on output variance (Equation 8), and parameters with high first-order or total-effect indices are selected for calibration. Bayesian inference using Markov Chain Monte Carlo[45] is then applied to update these parameters (Equation 9). The mean absolute error (MAE) between measured outputs and surrogate model predictions is used as the objective function (Equation 10), and the resulting maximum a posteriori estimates are incorporated into the Modelica-based

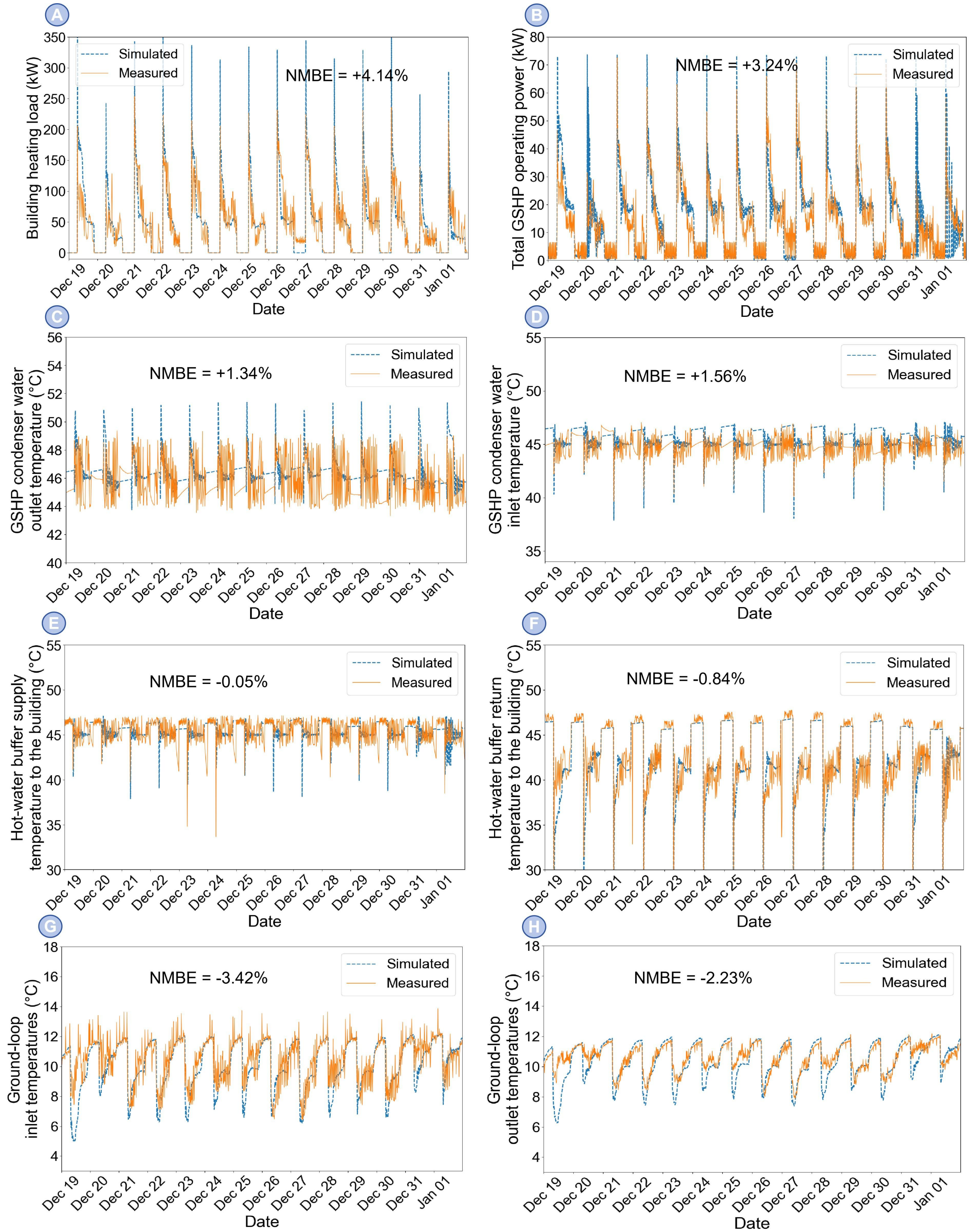


Figure 5. Validation of key system variables.

model for subsequent simulation and performance evaluation.

$$V(Y)=\sum_{i=1}^{k}V_i+\sum_{i<j}V_{ij}+\sum_{i<j<l}V_{ijl}+\ldots+V_{12\ldots k} \quad (8)$$

$$P(\theta|D)\propto P(D|\theta)P(\theta) \quad (9)$$

$$MAE=\frac{1}{n}\sum_{i=1}^{n}|Y_{m,i}-Y_{s,i}| \quad (10)$$

Where V_i represents the independent contribution of input variable X_i to the output variance, and V_{ij} represents the interaction effect between input variables X_i and X_j on the output variance; The first-order sensitivity index $S_i = V_i / V(Y)$ and the total-effect index S_{Ti} are defined to quantify the individual and overall interactive effects of each input parameter on the output variance, respectively. Larger values of these indices indicate a stronger influence of the corresponding parameter on the uncertainty of the model output; $P(\theta|D)$ denotes the posterior distribution of parameters, $P(D|\theta)$ represents the likelihood function, and $P(\theta)$ is the prior distribution of parameters; $Y_{m,i}$ and $Y_{s,i}$ denote the measured value and the corresponding predicted output of the

Table 1. Calibration parameters and prior sampling ranges

Calibration parameters	Sampling ranges for the prior distributions
Heat transfer coefficient	0.2-0.7 W/(m^2·K)
Envelope thermal capacitance	300-1000 J/(kg·K)
Envelope thickness	0.2-0.4 m
Envelope surface area	1000-4000 m^2
Equivalent envelope density	1.6-3 g/cm^3
Heating-zone volume	10000-30000 m^3

surrogate model for the i-th sample, respectively, and n is the total number of samples used for calibration.

Stable operational data from representative heating (December 1–18, 2022) and cooling (August 12–25, 2022) periods are selected for GSHP model calibration. The PLR and GSHP COP are inferred from measured hot-water inlet and outlet temperatures, flow rates, and operating power. Polynomial regression is applied to identify the GSHP part-load performance coefficients a_1, a_2, a_3 in the part-load efficiency correction function defined in Equation (5).

For room thermal model calibration, measured heating load data from December 1 to 18, 2022 are used, as this period provides high-quality measurements with minimal noise. The calibrated parameters include the heating-zone air volume, envelope surface area, heat transfer coefficient, thermal capacitance, material density, and thickness, all of which affect the building heating load and indoor thermal dynamics. An RF surrogate model is developed to predict the heating load generated by the room thermal model. The six calibration parameters listed in Table 1 are used as inputs, while the corresponding space heating load simulated by the physics-based model serves as the output. Sobol sensitivity analysis and Bayesian inference are then applied for model calibration by minimizing the mean absolute error between RF-predicted and measured heating loads.

Within the prescribed parameter space, 10,000 samples are generated using a uniform sampling strategy to support the sensitivity analysis and Bayesian calibration. The sample size is selected to ensure adequate coverage of the six-dimensional parameter space defined in Table S1 while keeping the computational cost manageable. During occupied periods, internal heat gains are represented by a constant value of 25 W/m^2, adopted from[46] and scaled according to the conditioned floor area.

Virtual testing and optimization

As mentioned in the previous section, two different predictor types are developed for virtual testing and optimization. For control tasks that are strongly influenced by external disturbances, such as weather and building thermal inertia, and where explicit physical relationships are difficult to formulate, purely data-driven predictors are employed. This predictor is used to optimize the preheating scheduling problem. In contrast, for equipment-level decisions where clear thermodynamic relationships exist and key internal states can be derived from the calibrated physics-based model, physics-informed predictors are adopted to enforce physical consistency and improve extrapolation reliability.

Preheating scheduling. Under the existing control strategy, preheating follows fixed time rules without explicit consideration of actual load conditions. To address this limitation, an adaptive preheating control strategy is proposed.

As indoor air temperature was not available in the 2022 BMS dataset used for this study, the preheating analysis is based on the calibrated single-zone room thermal model rather than direct room-by-room temperature measurements. To evaluate the suitability of this model for preheating analysis, supplementary indoor temperature measurements collected in 2025 from perimeter and interior rooms on the first floor were compared with the corresponding 2022 simulation results under similar outdoor-temperature conditions, as shown in Supplementary Figure S1-S5. The comparison covers indoor temperature evolution, initial temperature at the start of preheating, and the time required to reach 18 °C. The results show that measured temperatures vary across individual rooms, and that the simulated temperatures from the room thermal model also differ from these measurements. However, the model generally does not underestimate the preheating duration required to reach 18 °C. In cases where longer preheating was observed, the discrepancy did not exceed 15 min. These findings support the use of simulated indoor temperature from the single-zone room thermal model as an intermediate physical variable for determining the required preheating duration in this case study.

However, using the measured temperature of the worst-case room or a weighted average of all room temperatures as the control reference also has limitations. The worst-case room is not fixed and may vary with date, weather conditions, and operating scenario, while its occupancy pattern may differ from the main building demand. In addition, a weighted average may mask inter-zone thermal differences and underestimate the preheating requirement of the worst-case room. Therefore, the simulated room temperature is adopted to define the required preheating duration by tracking the time at which it first reaches 18 °C, thereby ensuring comparable heating effects prior to control strategy evaluation. This provides a practical training target for the RF predictor and demonstrates that, under incomplete field measurements, the calibrated physics-based model integrated into the virtual testing platform can support reliable control development and strategy evaluation.

An RF predictor is trained using data generated from the calibrated Modelica-based models under single-unit operation. The time required for the simulated room temperature to reach 18 °C after start-up is extracted as the prediction target, enabling adaptive preheating control. Inputs of the predictor include hourly buffer tank and outdoor air temperatures from 00:00 to 05:00 on the prediction day, representing the initial thermal state of the system before start-up, together with outdoor air temperatures over the same period on the previous day to capture cross-day thermal inertia effects.

To mitigate the risk of insufficient morning warm-up caused by prediction errors, a safety-margin adjustment is applied. Residuals between predicted values and the reference preheating durations generated by the calibrated simulation model, together with their standard deviation, are calculated during training (Equations 11–12). During prediction, a safety margin equal to six times the residual standard deviation is added to the predicted value (Equation 13). This margin is intended as a conservative engineering buffer rather than a strict statistical confidence interval. Six times the residual standard deviation is adopted to provide a high safety margin under the current data conditions, because the available training data may not fully cover all relevant operating conditions, especially rare or extreme cases. This also reflects the different practical consequences of prediction errors, since underestimation may lead to insufficient preheating before occupancy, whereas moderate overestimation mainly results in additional energy use. The safety margin can be adjusted in practice according to the similarity between the current inputs and the training data. The resulting preheating duration still varies with the predicted thermal state and weather conditions, rather than following a fixed schedule.

$$e_i = y_{true,i} - y_{pre,i} \tag{11}$$

$$\sigma = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(e_i - \bar{e})^2} \tag{12}$$

$$T_{preheat,i} = y_{pre,i} + 6\sigma \tag{13}$$

Where $y_{true,i}$ the reference preheating duration required for the simulated room temperature to reach the predefined comfort threshold, $y_{pre,i}$ denotes the preheating duration predicted by the RF model, $T_{preheat,i}$ denotes the adopted preheating duration after applying the safety margin.

In practice, users can input the above feature variables before daily GSHP start-up to determine the required preheating duration. This enables preheating schedules to be adjusted according to actual operating conditions, reducing the mismatch between fixed schedules and heating demand.

GSHP unit-number optimization. Under the current control strategy, two identical GSHP units operate simultaneously with equal PLR allocation. Although equal PLR distribution is generally considered efficient for identical units,[47] inappropriate determination of the number of operating units can still lead to prolonged low-load operation and degraded system performance. In systems equipped with buffer tanks, thermal storage introduces thermal buffering effects that create temporal and magnitude mismatches between building heating demand and GSHP output. Therefore, unit-number selection cannot be determined solely from instantaneous building load, and should also account for the thermal state of the buffer tank.

To support unit-number selection, two coupled variants of physics-informed neural network (PINN) predictors are developed to estimate GSHP operating power under different heating demands and unit-number configurations. The predictors comprise a buffer tank predictor and a GSHP performance predictor, both trained using a diverse set of data from simulations covering various operating scenarios generated from the calibrated Modelica model across a wide operating range. In the adopted hydraulic configuration, the buffer tank is hydraulically in series with the condenser-water loop. Therefore, the tank outlet temperature is identical to the condenser inlet temperature, and the tank inlet temperature is identical to the condenser outlet temperature. The PINN takes the building heating load ($Q_{heat,load}$), the hot-water mass flow rate in the GSHP-buffer tank loop ($M_{hot,total}$), and the GSHP condenser water inlet temperature ($T_{in,hot}$) as inputs and predicts the GSHP condenser water outlet temperature ($T_{out,hot}$). Physical constraints are imposed to ensure that the predicted outlet temperature increases monotonically with heating load and remains within physically feasible operating ranges, as expressed in Equations (14)-(15).

$$\frac{\partial T_{out,hot}}{\partial Q_{heat,load}} \geq 0 \tag{14}$$

$$T_{in,hot} \leq T_{out,hot} \leq T_{out,hotmax} \tag{15}$$

In parallel, GSHP performance is represented using an embedded-physics PINN (EP-PINN).[48] The EP-PINN takes condenser water inlet and outlet temperatures ($T_{in,hot}$, $T_{out,hot}$), the hot-water flow rate through each GSHP (M_{hot}), the ground-side fluid mass flow rate through each GSHP (M_{gro}), and the ground-side fluid inlet and outlet temperatures ($T_{in,gro}$, $T_{out,gro}$) as inputs, and outputs the GSHP COP. Key internal variables derived from the calibrated physics-based model, including load-related states, flow-related variables, and refrigerant-side thermodynamic quantities, are embedded as physical neurons representing intermediate quantities of the thermodynamic calculations. These neurons are structured according to the computational sequence of the thermodynamic equations. Connections are structured to reflect the physical equations, with variables propagating according to the mechanistic calculation flow, preserving the dependency relationships defined by the equations. The EP-PINN architecture and training settings are detailed in Supplementary Figure S6 and Tables S3-S4.

The predicted tank outlet temperature is then passed to the GSHP performance predictor, forming a coupled predictor for evaluating different unit-number configurations, as illustrated in Figure 4D Under the operating assumptions used for unit-number screening (a fixed tank temperature setpoint of 45 °C, constant flow in the heat pump–tank loop, and a ground-side inlet temperature approximated by the annual mean soil temperature) the operating power corresponding to different unit numbers can be evaluated as a function of building heating load. This provides a control-oriented basis for selecting the number of operating GSHP units.

MODEL EVALUATION AND VALIDATION

Different evaluation metrics are adopted according to model characteristics. For the physics-based Modelica system model, independent followed ASHRAE guidelines,[49] using the normalized mean bias error (NMBE) to quantify the agreement between simulated and measured performance (Equation 16). For the data-driven models, including polynomial regression, the RF surrogate, PINNs and RF predictors, model fitting or validation performance is evaluated using the root mean square error (RMSE) and mean absolute percentage error (MAPE) (Equations 17–18).

$$\text{NMBE} = \frac{\sum_{i=1}^{n}(M_i - S_i)}{n\bar{S}} \times 100\% \tag{16}$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(M_i - F_i)^2} \tag{17}$$

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{M_i - F_i}{M_i}\right| \times 100\% \tag{18}$$

Where M_i denotes the measured value; S_i denotes the simulated value of the physics-based Modelica system model; $\bar{S}$ is the mean of the simulated values, and n is the total number of samples; F_i denotes the fitted or predicted value of the data-driven models.

Calibration results and validation of the simulation model

For GSHP parameter calibration, polynomial regression yielded coefficients of a_1 = -0.5484, a_2 = 1.1365, a_3 = 0.4101 for the heating mode, and a_1 = -0.0455, a_2 = 0.5941, a_3 = 0.01 for the cooling mode. The heating-mode regression achieved an RMSE of 0.09 and an MAPE of 9.9%, while the cooling-mode regression yielded an RMSE of 0.02 and an MAPE of 14.6%. As shown in Figure S7, the regression results provide a satisfactory fit to the measured data and capture the efficiency degradation at low part-load ratios.

The RMSE and MAPE values indicate the fitting accuracy of the regression model, being 0.09 and 9.9% for heating and 0.02 and 14.6% for cooling, respectively.For the room thermal model, the RF surrogate achieved an RMSE of 3.1 kW and an MAPE of 5.8% in fitting the physics-based model outputs over the sampled parameter space (Figure S8). Based on this surrogate, a Sobol global sensitivity analysis was conducted. The results (Figure S9) indicate that the envelope heat transfer coefficient is the most influential parameter, followed by the envelope thermal capacitance, thickness, and surface area, while the equivalent envelope density and heating-zone volume have negligible influence. These four parameters were therefore selected for calibration using Bayesian inference, and the corresponding posterior distributions are shown in Figure S10. The final calibrated values of the envelope heat transfer coefficient, thermal capacitance, thickness, and surface area are 0.5932 W/(m^2·K), 890.6 J/(kg·K), 0.2982 m, and 3416.8 m^2, respectively.

An independent validation dataset covering 14 days from December 19, 2022, to January 1, 2023, and not used during model calibration, was used to assess model performance. As shown in Figure 5, the calibrated Modelica model achieves NMBE values of +4.14% for building heating load and +3.24% for total GSHP power consumption. For key monitoring variables, including condenser inlet and outlet temperatures, buffer tank supply and return temperatures, and ground-loop supply and return temperatures, the corresponding NMBE values range from −3.42% to +1.56%. All deviations are within ±4%, satisfying the commonly adopted validation criterion for building energy models (NMBE < ±10%).[49]

Validation of the Predictors

For the RF-based preheating predictor, validation on data not used in training yielded an RMSE of 12.7 min (Figure 6A). After applying the safety-margin adjustment, the adopted preheating durations consistently exceed the reference preheating durations from the calibrated single-zone room thermal model, particularly for samples where the RF model underestimates the required duration. As discussed in previous sections, these reference durations can represent the heating durations for the building. Adding a safety margin to the RF-predicted durations further reduces the risk of insufficient preheating across varying loads and weather conditions.

On an independent validation dataset (Figure S11), the buffer tank and GSHP predictors achieved RMSE values of 0.22 °C and 0.18, and MAPE values of 0.15% and 5.5%, respectively. The buffer tank and GSHP predictors achieve RMSE values of 0.22 °C and 0.18, and MAPE values of 0.15% and 5.5%, respectively. The results of the physics-informed buffer tank and GSHP performance predictors indicate that single-unit operation is preferable when the building heating load is below 161.3 kW, while two-unit operation becomes more efficient above this threshold (Figure 6B). As high-load conditions occur only during short morning start-up periods, single-unit operation is adopted to reduce frequent start–stop cycling.

VIRTUAL TESTING AND OPTIMIZATION

Baseline performance

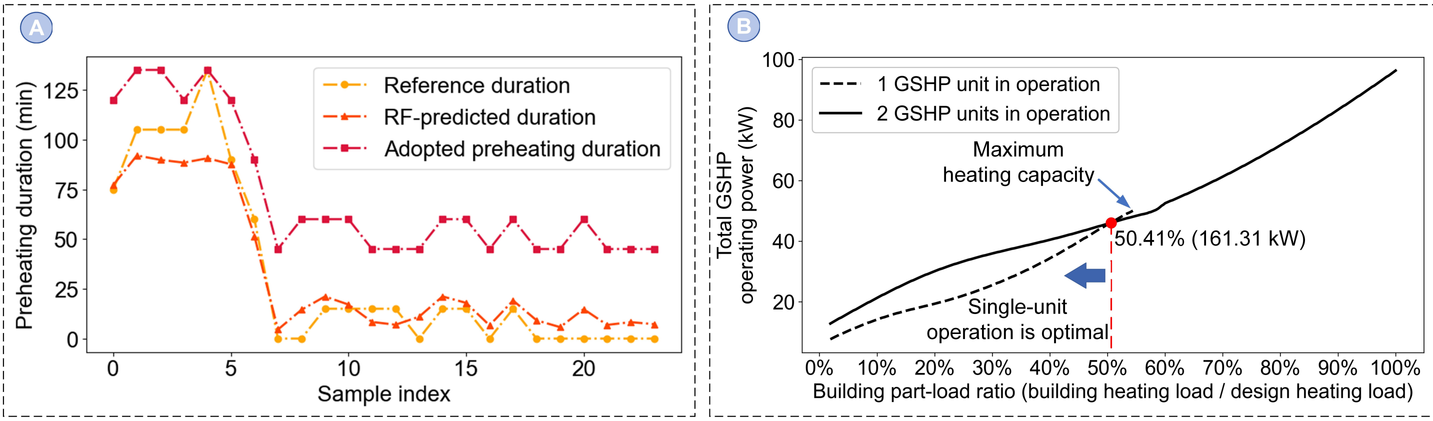


Figure 6. Coupled physics-informed predictors for buffer-tank and GSHP

The current control strategy is used as the baseline, whose performance is assessed using the calibrated simulation platform. Virtual tests across a wide range of operating conditions show that operating two GSHP units simultaneously degrades efficiency regardless of load level or ambient conditions. Energy-flow analysis identifies three major loss mechanisms: standby losses from keeping both units continuously available, cycling losses caused by frequent start–stop operation under low-load conditions, and distribution losses when operating temperatures exceed actual demand. Buffer tank simulations further indicate limited effective thermal storage, which promotes frequent cycling rather than stable continuous operation. Simulations under extreme weather conditions show that, except during the morning start-up period, the system generally retains sufficient capacity margin. A single unit can satisfy the heating demand during most operating periods while operating at a higher PLR. The short capacity deficit during morning start-up can be addressed through an appropriate preheating strategy.

Based on this analysis, the baseline operation is compared with three single-unit strategies to evaluate the feasibility and energy-saving potential of the proposed retro-commissioning approach. The baseline corresponds to the existing dual-unit operation with a fixed preheating schedule. Strategy 1 adopts single-unit operation with the current fixed preheating start time (07:15). Strategy 2 uses single-unit operation with a fixed preheating duration determined from the longest historical requirement. Strategy 3 applies single-unit operation with the proposed predictive preheating strategy.

Strategy feasibility and risk assessment

Virtual testing of alternative control strategies on the calibrated simulation platform enables systematic optimization while avoiding the risks of on-site experiments. The primary change in the tested strategies is the transition from dual-unit to single-unit operation, which increases the part-load ratio (PLR) of the active unit.

Figure 7A shows the simulation results for the coldest two weeks in the recorded period. Under Strategy 1, single-unit operation with the original fixed preheating schedule fails to raise the indoor temperature to 18 °C before occupancy, indicating insufficient morning warm-up under cold conditions. This occurs because low nighttime temperatures and accumulated heat deficits cannot be recovered within the limited preheating period when only one unit is operating. These results suggest that single-unit operation without adjusting the preheating duration may lead to thermal comfort risks. In contrast, both Strategy 2, which uses a fixed preheating duration based on the longest historical requirement, and Strategy 3, which applies the proposed predictive preheating strategy, can achieve the 18 °C setpoint before occupancy. This indicates that appropriate preheating scheduling can mitigate the capacity limitations of single-unit operation during extreme conditions.

Figure 7B presents the distribution of preheating start time under Strategy 3. The preheating duration adjusts dynamically with the operating condition. Early start-up is avoided under mild conditions, while longer preheating is applied during colder periods to meet the indoor temperature target. The longest preheating duration does not necessarily occur on the coldest day, as it depends on the overall thermal state of the building and buffer tank prior to start-up. Annual simulations show that the optimized strategy maintains the indoor temperature response at a level comparable to the baseline strategy while adapting to varying conditions. During preheating, the heat pump typically operates at a PLR of 60–80%, increasing indoor temperature while charging the buffer tank.

Energy performance improvement

The transition from dual-unit to single-unit operation alters the load profile. During normal operation, the GSHP PLR increases from approximately 20–50% to 50–80%. Figure 7C presents the building-side operational COP from December 5 to December 18, 2022. Under the baseline strategy, the COP typically ranges from 2 to 4 and frequently falls below 3. With the optimized strategies, it is maintained between 3 and 5 for most operating periods, and the mean COP increases from 3.0 to 4.4 (46.7 %). By taking account of pumping power, the system-level efficiency ratio increases from 2.3 to 3.2.

Figure 7D compares the total energy consumption of the GSHP system under different operating strategies during the coldest 14 days of the winter period. Under the baseline dual-unit operation, total energy use reaches 10,123.52 kWh. In contrast, single-unit operation with the intelligent preheating strategy reduces total energy consumption to 8,677.08 kWh, corresponding to an energy saving of 16.1%. The mean COP shows a larger increase than the reduction in total electricity consumption during the coldest two weeks of the recorded period. This is because the optimized single-unit strategy involves longer preheating periods than the baseline, which increases the GSHP runtime. From an electricity consumption perspective, these additional preheating hours contribute extra energy that is not present in the baseline strategy. From a COP perspective, the heating delivered during preheating periods is also included in the calculation. As a result, the increase in mean COP appears larger than the corresponding reduction in total electricity consumption. Over the two weeks, preheating contributes 192 kWh, accounting for only 2.3% of total energy use, while avoiding insufficient morning warm-up and improving GSHP efficiency. Compared with the fixed preheating strategy, the intelligent preheating approach further reduces energy use by 176 kWh (2%) by avoiding unnecessary early start-up during mild weather.

Single-unit operation also reduces cycling under low-load conditions, resulting in more stable operation and lower equipment wear. Although the existing buffer tank has limited thermal storage capacity, its utilization is improved under the optimized strategy. By charging the tank during preheating, stored thermal energy supports morning warm-up and can be partially carried over to the following day, maximizing the value of the existing tank without additional investment, thereby reducing retrofit and long-term maintenance costs.

Economic and environmental benefits

Strategy 3 integrates GSHP unit-number optimization with intelligent morning preheating, improving efficiency while avoiding insufficient morning warm-up and reducing energy waste. The shift from dual-unit to single-unit operation altered the system's heating load profile, increasing the GSHP PLR from 20–50% to 50–80%, keeping the unit in a higher efficiency range. Based on the calibrated virtual testbed, electricity costs over the representative 14-

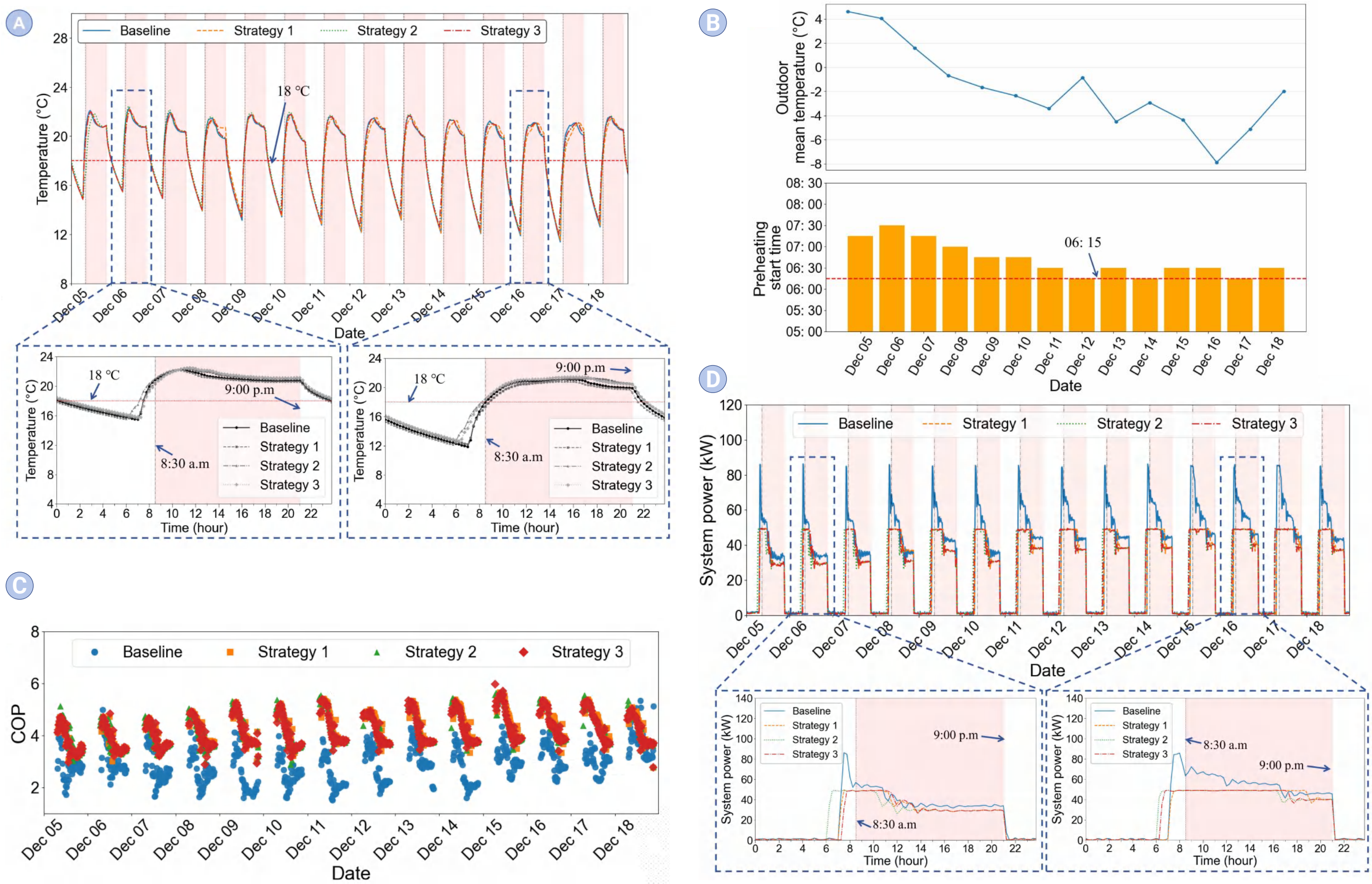


Figure 7. Comparison of (A) indoor temperature and (B) preheating start time distributions under different strategies. Comparison of (A) COP and (B) system power under different strategies.

day winter period decreased from £6,596.34 under dual-unit operation to £5,630.37 under single-unit operation with preheating, yielding a simulated saving of £965.9 (14.6%). These savings were mainly driven by better matching between heat pump operation and building heating demand.

The proposed strategy is cost-effective to deploy, requiring only control logic reprogramming and operator training, without hardware modifications, additional sensors, or equipment replacement. This approach supports the feasibility of model-based commissioning, especially in existing buildings where major retrofits are constrained by approval processes and construction limitations.

The resulting energy savings also provide environmental benefits. With a 15% energy reduction, the strategy leads to an annual emissions reduction of about 3.3 t CO_2 based on the current UK electricity grid carbon intensity. As grid decarbonization continues, absolute emissions will decrease further. The improved efficiency also reduces peak electricity demand, alleviating grid stress and supporting building electrification.

Implications for field retro-commissioning

Based on the calibrated model and virtual-testing results, a feasible pathway is proposed to translate virtual-testing results into actionable field retro-commissioning measures. The virtual testing framework for retro-commissioning could be deployed on a local supervisory computer. By reading BMS operating data stored in a local database, the testbed could update its input variables and generate optimized setpoints and control signals, which could then be written to the existing direct digital controller under the current control logic.

For the studied GSHP system, the main field adjustments would involve two aspects: revising the original dual-unit simultaneous operation to a single-unit scheme, and replacing the fixed morning preheating start time with a prediction-based dynamic schedule. Verification could be carried out on test days with similar weather and building occupancy, comparing the performance of the baseline strategy, Strategy 1, Strategy 2, and Strategy 3. Strategies 1 and 2 could be implemented through manual adjustment of unit operation and preheating settings, whereas Strategy 3 would be executed using virtual testbed–generated control variables. Additional indoor temperature measurements could be introduced during the verification stage to assess whether all rooms reach the comfort threshold before occupancy.

The four strategies can then be compared using indicators such as room temperature, GSHP operating power, total system energy consumption, and building-side operational COP, thereby further evaluating the virtual testbed's practical potential for retro-commissioning. The corresponding field deployment and experimental validation remain for future work.

CONCLUSIONS

This study presents a virtual testing–based retro-commissioning framework that integrates physics-based system modeling with data-driven methods for heating, ventilation, and air-conditioning (HVAC) systems. The framework enables control strategy development and evaluation within a calibrated virtual testbed, reducing reliance on on-site experimentation and supporting adaptive operational optimization. A ground source heat pump (GSHP) system is used to demonstrate the proposed framework. The main findings are summarized as follows:

- Based on the calibrated physics-based system model, machine learning predictors were developed to optimize GSHP unit operation (PINN predictors) and preheating control (RF predictor). Compared with the baseline strategy, the optimized control reduced energy consumption by 16.1% and increased the building-side operational COP from 3.0 to 4.4 while maintaining comparable indoor thermal conditions.
- Switching from baseline dual-unit to single-unit GSHP operation improves efficiency but introduces a morning warm-up risk without preheating. The intelligent preheating strategy increases energy use by about 2% while ensuring reliable operation under cold conditions.

• The proposed framework extends conventional one-time commissioning into an iterative, model-based optimization process. Virtual commissioning enables low-risk evaluation of control strategies, and the calibrated virtual testbed indicates a 14.6% reduction in energy cost over the representative 14-day period without hardware modifications.

Future work will therefore focus on improving the modularity and generality of the virtual testbed and framework, validating it through additional practical case studies, and further investigating adaptive supply temperature control, variable-flow operation, advanced thermal storage, and integration with demand-side response strategies. With more room-temperature measurements available, the current single-zone room thermal model, which serves as a practical approximation for preheating analysis in this study, could be further refined and extended toward a multi-zone model to better capture spatial heterogeneity across rooms.

REFERENCES

1. International Energy Agency. (2025). Energy efficiency 2025: Buildings. *IEA*. https://www.iea.org/reports/energy-efficiency-2025/buildings
2. Pérez-Lombard L., Ortiz J. and Pout C. (2008). A review on buildings energy consumption information. *Energy Build.* **40**:394–398. DOI:10.1016/j.enbuild.2007.03.007
3. Wan K. K., Li D. H., Pan W., et al. (2012). Impact of climate change on building energy use in different climate zones and mitigation and adaptation implications. *Appl. Energy* **97**:274–282. DOI:10.1016/j.apenergy.2011.11.048
4. Li G., Wang C., Liu L., et al. (2024). Study on Sensor Fault-Tolerant Control for Central Air-Conditioning Systems Using Bayesian Inference with Data Increments. *Sensors (Basel)* **24**:1150. DOI:10.3390/s24041150
5. Mills E. (2011). Building commissioning: a golden opportunity for reducing energy costs and greenhouse gas emissions in the United States. *Energy Effic.* **4**:145–173. DOI:10.1007/s12053-011-9116-8
6. Crowe E., Mills E., Poeling T., et al. (2020). Building commissioning costs and savings across three decades and 1500 North American buildings. *Energy and Buildings* **227**:110408. DOI:10.1016/j.enbuild.2020.110408
7. Borgstein E. H., Lamberts R. and Hensen J. L. M. (2018). Mapping failures in energy and environmental performance of buildings. *Energy Build.* **158**:476–485. DOI:10.1016/j.enbuild.2017.10.038
8. Verhelst J., Van Ham G., Saelens D., et al. (2017). Model selection for continuous commissioning of HVAC-systems in office buildings: A review. *Renew. Sustain. Energy Rev.* **76**:673–686. DOI:10.1002/er.4159
9. Verhelst J., Van Ham G., Saelens D., et al. (2017). Model selection for continuous commissioning of HVAC-systems in office buildings: A review. *Renew. Sustain. Energy Rev.* **76**:673–686. DOI:10.1016/j.rser.2017.01.119
10. Taskgroup I. B. E. E. (2019). Building Energy Performance Gap Issues: An International Review. *IPEEC Build. Energy Effic. Taskgroup: Paris, France.*
11. Bockelmann F. and Fisch M. N. (2019). It works—Long-term performance measurement and optimization of six ground source heat pump systems in Germany. *Energies* **12**:4691. DOI:10.3390/en12244691
12. Ouabira M. M. (2025). Future of Commissioning. *European J. Archit. Urban Plan.* **4**:13–21. DOI:10.24018/ejarch.2025.4.2.48
13. Magni M., Ochs F., de Vries S., et al. (2021). Detailed cross comparison of building energy simulation tools results using a reference office building as a case study. *Energy Build.* **250**. DOI:10.1016/j.enbuild.2021.111260
14. Zare N., Saryazdi S. M. E., Bahman A. M., et al. (2023). Investigation of heating energy performance gap (EPG) in design and operation stages of residential buildings. *Energy Build.* **301**:113747. DOI:10.1016/j.enbuild.2023.113747
15. Martinez-Viol V., Urbano E. M., Delgado-Prieto M., et al. (2022). Automatic model calibration for coupled HVAC and building dynamics using Modelica and Bayesian optimization. *Build. Environ.* **226**:109693. DOI:10.1016/j.buildenv.2022.109693
16. Chen K., Zhu X., Anduv B., et al. (2022). Digital twins model and its updating method for heating, ventilation and air conditioning system using broad learning system algorithm. *Energy* **251**:124040. DOI:10.1016/j.energy.2022.124040
17. Ward R., Choudhary R., Gregory A., et al. (2021). Continuous calibration of a digital twin: Comparison of particle filter and Bayesian calibration approaches. *Data-Centric Engineering* **2**:e15. DOI:10.1017/dce.2021.12
18. Thelen A., Zhang X., Fink O., et al. (2022). A comprehensive review of digital twin — part 1: modeling and twinning enabling technologies. *Struct. Multidis. Optim.* **65**. DOI:10.1007/s00158-022-03425-4
19. Drgoňa J., Picard D. and Helsen L. (2020). Cloud-based implementation of white-box model predictive control for a GEOTABS office building: A field test demonstration. *J. Process Control* **88**:63–77. DOI:10.1016/j.jprocont.2020.02.007
20. Henze G. P., Kircher K. J. and Braun J. E. (2025). Why has advanced commercial HVAC control not yet achieved its promise. *J. Build. Perform. Simul.* **18**:217–228. DOI:10.1080/19401493.2024.2429728
21. Picard D., Drgoňa J., Kvasnica M., et al. (2017). Impact of the controller model complexity on model predictive control performance for buildings. *Energy Build.* **152**:739–751. DOI:10.1016/j.enbuild.2017.07.027
22. Blum D., Arroyo J., Huang S., et al. (2021). Building optimization testing framework (BOPTEST) for simulation-based benchmarking of control strategies in buildings. *J. Build. Perform. Simul.* **14**:586–610. DOI:10.1080/19401493.2021.1986574
23. Wang J., Li Y., Gao R. X., et al. (2022). Hybrid physics-based and data-driven models for smart manufacturing: Modelling, simulation, and explainability. *J. Manufact. Sys.* **63**:381–391. DOI:10.1016/j.jmsy.2022.04.004
24. Lu X., Fu Y. and O'Neill Z. (2023). Benchmarking high performance HVAC Rule-Based controls with advanced intelligent Controllers: A case study in a Multi-Zone system in Modelica. *Energy Build.* **284**. DOI:10.1016/j.enbuild.2023.112854
25. Hussien A., Khan W., Hussain A., et al. (2023). Predicting energy performances of buildings' envelope wall materials via the random forest algorithm. *J. Build. Eng.* **69**:106263. DOI:10.1016/j.jobe.2023.106263
26. Afram A., Janabi-Sharifi F., Fung A. S., et al. (2017). Artificial neural network (ANN) based model predictive control (MPC) and optimization of HVAC systems: A state of the art review and case study of a residential HVAC system. *Energy Build.* **141**:96–113. DOI:10.1016/j.enbuild.2017.02.012
27. Van Every P. M., Rodriguez M., Jones C. B., et al. (2017). Advanced detection of HVAC faults using unsupervised SVM novelty detection and Gaussian process models. *Energy Build.* **149**:216–224. DOI:10.1016/j.enbuild.2017.05.053
28. Villano F., Mauro G. M. and Pedace A. (2024). A review on machine/deep learning techniques applied to building energy simulation, optimization and management. *Thermo* **4**:100–139. DOI:10.3390/thermo4010008
29. Hong T., Wang Z., Luo X., et al. (2020). State-of-the-art on research and applications of machine learning in the building life cycle. *Energy Build.* **212**:109831. DOI:10.1016/j.enbuild.2020.109831
30. Wang X. and Dong B. (2023). Physics-informed hierarchical data-driven predictive control for building HVAC systems to achieve energy and health nexus. *Energy Build.* **291**:113088. DOI:10.1016/j.enbuild.2023.113088
31. Von Krannichfeldt L., Orehounig K. and Fink O. (2025). Combining physics-based and data-driven modeling for building energy systems. *Appl. Energy* **391**:125853. DOI:10.1016/j.apenergy.2025.125853
32. Nghiem T. X., Drgoňa J., Jones C., et al. (2023). Physics-informed machine learning for modeling and control of dynamical systems. *arXiv preprint arXiv:2306.13867.* DOI:https://doi.org/10.48550/arXiv.2306.13867
33. Farkane A., Boutayeb M., Oudani M., et al. (2025). An Adaptive Physics-Informed Neural Network Observer for state estimation in nonlinear dynamical systems. *Eng. Appl. Arti. Intel.* **162**. DOI:10.1016/j.engappai.2025.112655
34. Campoy-Nieves A., Manjavacas A., Jiménez-Raboso J., et al. (2025). Sinergym – A virtual testbed for building energy optimization with Reinforcement Learning. *Energy Build.* **327**:11. DOI:10.1016/j.enbuild.2024.115075
35. Wu S., Zheng W., Wang Z., et al. (2025). AlphaDataCenterCooling: A virtual testbed for evaluating operational strategies in data center cooling plants. *Appl. Energy* **380**. DOI:10.1016/j.apenergy.2024.125100
36. Mohebi P., Liu Y., Xing S., et al. (2026). Stochastic model predictive control for building thermal management: A comparative analysis of scenario-based and tube-based approaches using the BOPTEST framework. *Energy* **351**:140808. DOI:10.1016/j.energy.2026.140808
37. Hekimoğlu M. B., Baghaee S., Büyüker A. İ., et al. (2025). IoT-Based Multi-Speed HVAC Control: Simulation Results and Testbed Development. 33rd Signal Proces. Commun. Appl. Conf.
38. Wongphong C., Wongta P. and Marurngsith W. (2026). A Co-Simulation Framework for Building Energy Management as a Testbed for Energy-Aware Data Movement Analysis. Proceed. Supercomput. Asia Int. Conf. High Perform. Comput. Asia Pacific Region Workshops. DOI:10.1016/j.enbuild.2020.109831
39. Li H. and Hong T. (2025). A digital twin platform for building performance monitoring and optimization: Performance simulation and case studies. *Build. Simul.* **18**:1561–1579. DOI:10.1007/s12273-025-1290-2
40. Pop C. B., Chifu V. R., Cordea C., et al. (2021). Forecasting the Short-Term Energy Consumption Using Random Forests and Gradient Boosting. RoEduNet.
41. Ahmad M. W., Mourshed M. and Rezgui Y. (2017). Trees vs Neurons: Comparison between random forest and ANN for high-resolution prediction of building energy consumption. *Energy Build.* **147**:77–89. DOI:10.1016/j.enbuild.2017.04.038
42. Liang X., Liu Y., Chen S., et al. (2025). Physics-informed neural network for chiller plant optimal control with structure-type and trend-type prior knowledge. *Appl. Energy* **390**. DOI:10.1016/j.apenergy.2025.125857
43. Wetter M., Zuo W., Nouidui T. S., et al. (2013). Modelica Buildings library. *J. Build. Perform. Simul.* **7**:253–270. DOI:10.1080/19401493.2013.765506
44. Fouodo C. J., Kronziel L. L., König I. R., et al. (2023). Effect of hyperparameters on variable selection in random forests. *arXiv preprint arXiv:2309.06943.* DOI:10.48550/arXiv.2309.06943
45. Hug S., Raue A., Hasenauer J., et al. (2013). High-dimensional Bayesian parameter estimation: case study for a model of JAK2/STAT5 signaling. *Math Biosci* **246**:293–304. DOI:10.1016/j.mbs.2013.04.002
46. Chow D. H. C. and Levermore G. J. (2010). The effects of future climate change on heating and cooling demands in office buildings in the UK. *Build. Serv. Eng. Res. Tech.* **31**:307–323. DOI:10.1177/0143624410371284
47. Dai Z., Zhang X., Liu J., et al. (2025). Energy-saving control strategy for the joint operation of multiple ground source heat pumps system based on TRNSYS: A research study. *Case Stud. Therm. Eng.* **67**:105834. DOI:10.1016/j.csite.2025.105834

48. Fang J., Yan C., Lu W., et al. (2025). Embedding physical neurons in physics-informed neural networks (EP-PINNs) for enhancing chiller performance prediction. *Build. Simul.* **18**:1877–1901. DOI:10.1007/s12273-025-1293-z
49. Guideline A. (2014). Measurement of energy, demand, and water savings. *ASHRAE guidel* **4**:1–150. DOI:10.1177/0143624410371284

FUNDING AND ACKNOWLEDGMENTS

This research work was supported by the National Natural Science Foundation of China (No. 52570228 & No. 22441020).

AUTHOR CONTRIBUTIONS

Conceptualization and study design were developed by all authors. Material preparation, data collection, model development, simulation, and formal analysis were carried out by Junjian Fang and Chaoqun Zhuang. Anurag Goyal contributed to the development of the modelling framework. Ruchi Choudhary and Chaoqun Zhuang supervised the research. The first draft of the manuscript was written by Junjian Fang and Chaoqun Zhuang. All authors contributed to reviewing and editing the manuscript. All authors read and approved the final manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

DATA AND CODE AVAILABILITY

Readers may contact the corresponding authors for data upon reasonable request.

SUPPLEMENTAL INFORMATION

It can be found online at https://doi.org/10.59717/ipj.energy-use.2026.100039