scientific reports

Check for updates

OPEN

Explainable and generalizable AI-driven multiscale informatics for dynamic system modelling

Chen Luo, Ao-Jin Li, Jiang Xiao, Ming Li & Yun Li✉

Ultra-precision machining requires system modelling that both satisfies explainability and conforms to data fidelity. Existing modelling approaches, whether based on data-driven methods in present artificial intelligence (AI) or on first-principle knowledge, fall short of these qualities in high-demanding industrial applications. Therefore, this paper develops an explainable and generalizable 'grey-box' AI informatics method for real-world dynamic system modelling. Such a grey-box model serves as a multiscale 'world model' by integrating the first principles of the system in a white-box architecture with data-fitting black boxes for varying hyperparameters of the white box. The physical principles serve as an explainable global meta-structure of the real-world system driven by physical knowledge, while the black boxes enhance local fitting accuracy driven by training data. The grey-box model thus encapsulates implicit variables and relationships that a standalone white-box model or black-box model fails to capture. Case study on an industrial cleanroom high-precision temperature regulation system verifies that the grey-box method outperforms existing modelling methods and is suitable for varying operating conditions.

Keywords Grey-box model, System modelling, High-precision control system, Explainable artificial intelligence

Recently, large language models (LLMs) in artificial intelligence (AI) have shown certain potential for dynamic system modelling and control, but their industrial applications face challenges, as they are fundamentally data-driven black boxes based on artificial neural networks (ANNs)[1,2], which lack underlying knowledge that the systems were built upon. For example, high-precision temperature control for cleanroom manufacturing relies on a plant model that is derived from thermodynamic principles of a heating, ventilation, and air conditioning (HVAC) system[3–6]. In such a practical case, black-box approaches based on classical transfer functions are also inadequate since the HVAC system is intrinsically nonlinear and may be time varying[7–12]. Furthermore, this time-varying behaviour represents not only a single variation in the system parameters over the time but is also closely related to the current state of the system. For similar reasons, other black-box models such as fuzzy logic, Statistical Output Error, Box-Jenkins, and Autoregressive Moving Average (ARMA) models are also inadequate and lack explainability[11,13–21]. Nonetheless, a black-box model often yields the most straightforward data fitting solution, and can be updated regularly as the operating conditions change, although it can hardly extrapolate for operations outside the range of the training data[13,22–24].

On the contrary, a white-box model, which is based on the working mechanism of the system, offers explainability of the system being controlled[25–31]. However, a white box may not capture the detailed dynamics reflected from the data or accommodate all complexity of the system as certain dynamics are impossible to model using first principles. Factors contribute to this failure include multiple levels of system complexity, minor dynamics that are not understood, and distributed (as opposed to a fully lumped) nature of the problem[26,27,32]. Thus, white-box models often have limited data-fitting accuracy[33–37].

Conversely, the black-box approach[38–41] fits input–output (I/O) data more accurately[42–48]. At the same time, demand for explainable artificial intelligence (XAI) has emerged in recently, so as to incorporate scientific knowledge or practical experience into black-box models and thus to enhances the transparency and interpretability of machine learning and 'world models'[49,50]. Examples include Theory-Guided Data Science (TGDS) approach with Physically-Guided Neural Network (PGNN) models[51,52], Physically-Informed Neural Networks (PINN)[53], and Theory-Guided Neural Network (TgNN) frameworks to incorporate physical principles with partial differential equations (PDEs) and boundary conditions[54]. By embedding domain knowledge in AI models, their

Shenzhen Institute for Advanced Study, University of Electronic Science and Technology of China, Shenzhen 518110, China. ✉email: Yun.Li@ieee.org

trustworthiness is improved by mapping relationships between variables and by incorporating a priori knowledge existing in industrial practice. Hence, the grey-box model in this paper places an emphasis on structuring modelling on a white box for the first principles, and then organically augmenting it with local black boxes for fitting accuracy, whereas existing XAI is dominated by data-driven modelling augmented with some knowledge or experience. This paper is thus to make the best out of both black-box and white-box methods, as illustrated in Fig. 1.

For an HVAC system, white-box modelling techniques are adopted so far[25,55–57]. For process control in general, the plant model is usually constructed based on physical and/or chemical principles such as composition, mass, momentum and energy balances, forming underlying relationships between the inputs and outputs of the system. The underlying principles are usually expressed in terms of ordinary differential equations (ODEs), which are approximate to dynamics where partial differential equations (PDEs) are more adequate. The high-precision temperature control system for cleanroom manufacturing exhibits characteristics such as high thermo-inertia, time-lags, and uncertain perturbation factors[17]. These factors contribute to the system's high-order dynamics and nonlinear characteristics, necessitating accurate system identification and modelling for the design of high-performance control algorithms. However, due to assumptions, a large number of parameters or neglect of parameter variations[19,58,59], white-box models often fail to fully capture the system's complex attributes and dynamics, leading to inadequately data fitting[13,22].

Conversely, models that account for all necessary physical parameters and conditions often become overly complex, rendering it impractical to design control systems[60,61]. Therefore, hybrid models have been developed to combine the strengths of both white-box and black-box approaches[61–64]. However, HVAC system models labeled as "grey-box models" are merely a series or parallel connection of the two types[22,65,66], thus clattering the two structures and failing to offer advantages of both approaches. Table 1 lists a comparation of the three modelling techniques in areas such as semiconductor and aerospace industries.

In this paper, we define the "grey-box model" in the AI age first, and then illustrate it with an application to high-precision temperature control system modelling for industrial cleanroom ultra-precision manufacturing. Building a more accurate model for an ultra-precision system enables its control algorithm better to handle perturbations, constraints, uncertainties, time-varying dynamics, time delays, and slow-moving processes across a wide range of operating conditions.

Because the model structure is not replaced by artificial structures in a generic black-box approximator, the grey-box representation helps meet the requirements of interpretability, adaptability, and generalizability. As such a grey box approach is stemmed from a white box, it can accommodate other HVAC and complex dynamic systems for fault detection, diagnostic, or building energy efficiency strategy development through the introduction of knowledge and causality.

AI-driven grey-box modelling

To tackle the challenges of modelling complexities and control specifications, we introduced the evolutionary grey-box model first in 2002, which was validated on hydraulic system and process modelling[61]. It has since undergone further development and research[25,29,35,41,42,64,75,76,97]. In the AI age, we now define the "grey-box

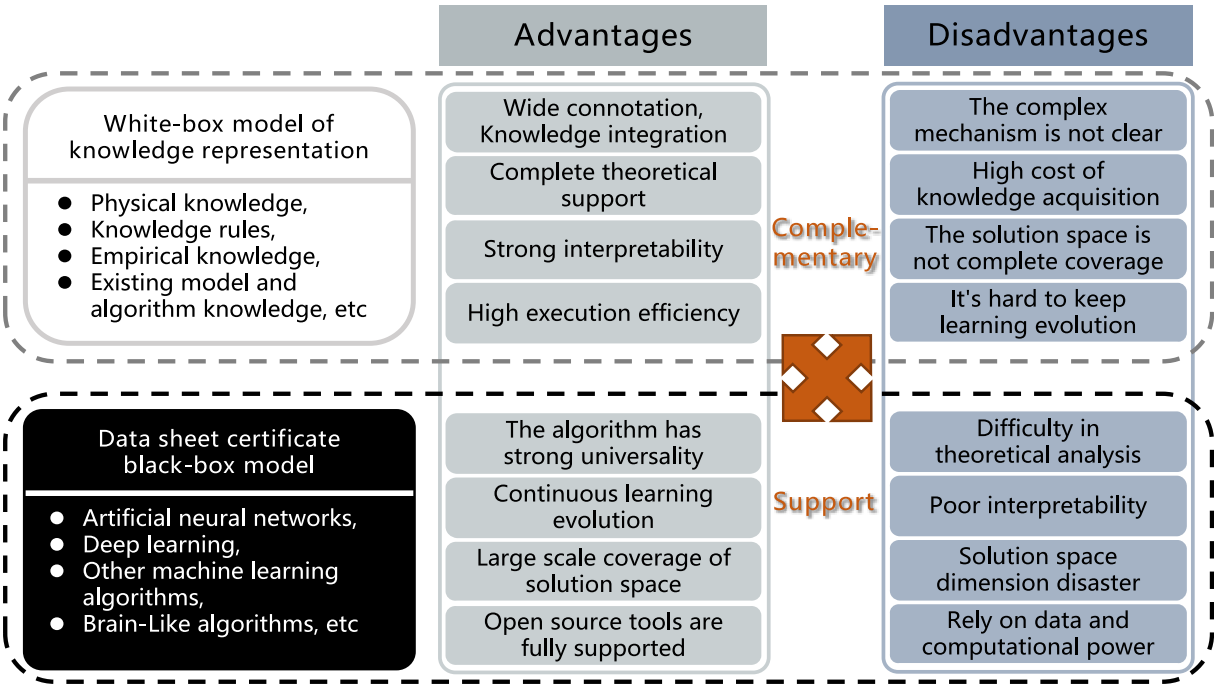

Figure 1. Advantages and disadvantages of white-box and black-box modelling approaches.

Modelling techniques	Weaknesses	Strength	Field of applications
Physics-based white box	Poor accuracy[67–69] Relies entirely on knowledge of the process and the laws of physics that control it[70,71] Complex[72,73]	Easy to analyse and represent the system more closely[74] Requires less training data[75,76] Better generalization ability[77]	Semiconductor device design[72], aerodynamic modelling[74], aerospace RTM process modelling[73], silicon-based transistor modelling[71], aerospace alloy notch fatigue assessment[75,76], bipolar transistor compact modelling[68], Gummel-Poon compact model[69], predictive modelling in aircraft design[67], computational modelling of aircraft performance[77]
Data-based black box	Requires a large amount of training data and relies entirely on measured data for input and output variables[71,78] Do not reflect actual physical behaviour[79] Poor generalization[79–81]	Do not require understanding of system physics[82,83] Provides good prediction accuracy[84] Simplicity and reduce computational cost[85,86]	Semiclassical device simulation with ML[72], semiconductor device simulation[86], subsonic blade row modelling[85]; novel transistor circuit-level analysis[79], MOS transistor modelling[83], modelling of fatigue damage processes[82], atomic modelling of devices, prediction of atomic forces[84], circuit-level analysis of novel transistors[79], electrothermal interaction modelling[80],Fatigue life prediction[81]
Gray box based on both data and physics	Requires more effort to develop[72] Accuracy is highly dependent on the richness of the data used to train the model[87]	Higher accuracy[88–91] Relatively less complex[92,93] Easy to generalize[90,94,95] Relatively low computational cost[89,96]	Semiconductor device design[72], multi-axis fatigue predictive modelling[90], helicopter fatigue predictive modelling[88], aerospace fluid dynamics modelling[93], nonlinear model simplification for computational fluid dynamics[92], semiconductor device modelling for emerging devices[91], corrosion fatigue modelling[94], wave load prediction[95], electronic conductivity prediction for large systems[89], aircraft structural health monitoring[96]

Table 1. Comparative study of the three modelling techniques.

model" as an explainable white box driven by real-world knowledge, with coefficients generalized through data-driven black boxes.

Definition—generic grey-box model:

The *Grey-Box Model* G($\boldsymbol{u}$, $\boldsymbol{y}$, t) is defined as a generic nonlinear *white-box* state-space model (SSM) derived from the fist-principle knowledge, whose hyperparameter set $\boldsymbol{\theta}(\boldsymbol{x}, \boldsymbol{u}, t)$ is dependent on input–output training data and is time varying:

$$\begin{cases} \dot{\boldsymbol{x}}(t) = \boldsymbol{f}(\boldsymbol{x}(t), \boldsymbol{u}(t), \boldsymbol{\theta}(\boldsymbol{x}, \boldsymbol{u}, t), t) \\ \boldsymbol{y}(t) = \boldsymbol{h}(\boldsymbol{x}(t), \boldsymbol{u}(t), \boldsymbol{\theta}(\boldsymbol{x}, \boldsymbol{u}, t), t) \\ \boldsymbol{\theta} = \text{argmin}_{\boldsymbol{\theta} \in \mathrm{R}^*} L[\boldsymbol{\theta}(\boldsymbol{x}, \boldsymbol{u}, t)] \end{cases} \tag{1}$$

where

- $\boldsymbol{u}(t)$ and $\boldsymbol{y}(t)$ are the plant input and output training data sets, respectively;
- $\boldsymbol{x}(t)$ are the state variables, which are dynamic intermarries between $\boldsymbol{u}(t)$ and $\boldsymbol{y}(t)$;
- $\boldsymbol{f}$ and $\boldsymbol{h}$ are white-box nonlinear functions whose hyperparameter set θ is comprised of *black-box* functional approximators;
- $R^* = R \cup R^1 \cup R^2 \cdots \cup R^\infty$;
- $L(\cdot) : R^* \rightarrow R$ is a loss function of the Grey-Box Model.

Grey-box modelling hence respects the original physical laws of the system by using a valid SSM without altering its "white" structure, thus retaining its interpretability. By filling its parameters with "black" boxes, it acknowledges the system's empirical data while reduces modelling errors. A data-driven black-box approach can hence be employed to automatically adjust the model parameters by monitoring the operating condition variables, thus enhancing the performance of the mechanistic model. Therefore, the hyperparameters accommodate diverse working conditions and operation scenario. Further, they can also be written in the form of generalized time-varying matrices, thus extending the generic *Grey-Box Model* to a generalized *Linear Time-Varying Grey-Box Model.*

Corollary—linear time-varying grey-box model:

A generalized *Linear Time-Varying Grey-Box Model* is a time-varying linearized generic *Grey-Box Model* with p inputs, q outputs and n state variables:

$$\begin{cases} \dot{\boldsymbol{x}}(t) = \mathbf{A}(\boldsymbol{x}, \boldsymbol{u})\boldsymbol{x}(t) + \mathbf{B}(\boldsymbol{x}, \boldsymbol{u})\boldsymbol{u}(t) \\ \boldsymbol{y}(t) = \mathbf{C}(\boldsymbol{x}, \boldsymbol{u})\boldsymbol{x}(t) + \mathbf{D}(\boldsymbol{x}, \boldsymbol{u})\boldsymbol{u}(t), \\ \boldsymbol{\theta} = \text{argmin}_{\boldsymbol{\theta} \in \mathrm{R}^*} L[\mathbf{A}(\boldsymbol{x}, \boldsymbol{u}, \mathrm{t}), \mathbf{B}(\boldsymbol{x}, \boldsymbol{u}, \mathrm{t}), \mathbf{C}(\boldsymbol{x}, \boldsymbol{u}, \mathrm{t}), \mathbf{D}(\boldsymbol{x}, \boldsymbol{u}, \mathrm{t})] \end{cases} \tag{2}$$

where $\mathbf{A}(\cdot)$ is termed the "state matrix", with certain elements varying dynamically with $\boldsymbol{x}(t)$, $\boldsymbol{u}(t)$ and implicitly t, and $dim[\mathbf{A}(\cdot)] = n \times n$; $\mathbf{B}(\cdot)$ is termed the "input matrix", with certain elements varying dynamically with $\boldsymbol{x}(t)$, $\boldsymbol{u}(t)$ and implicitly t, and $\dim[\mathbf{B}(\cdot)] = n \times p$; $\mathbf{C}(\cdot)$ is termed the "output matrix", with certain elements varying

dynamically with $\boldsymbol{x}(t)$, $\boldsymbol{u}(t)$ and implicitly t, and $\dim[\mathbf{C}(\cdot)] = q \times n$; $\mathbf{D}(\cdot)$ is termed the "feedthrough matrix", with certain elements varying dynamically with $\boldsymbol{x}(t)$, $\boldsymbol{u}(t)$ and implicitly t, $\dim[\mathbf{D}(\cdot)] = q \times p$. Eq. (2) form the structure of the SSM white box with **A**, **B**, **C** and **D** matrices, which are fit with black boxes determined by $\boldsymbol{\theta}$.

Such a 'world model' is constructed from physical knowledge and is hence dominantly interpretable, while it can complementarily fit unknown or complex minor dynamics accurately from training data. By decomposing it into the white-box and the black-box levels, a real-world or nonlinear dynamic system model is thus approximated by the structurally modular linear time-varying grey-box model, whose parameters can also be determined structurally modularly with multiple scales. This provides ease of informatics implementations for development when using a graphical processing unit (GPU). More detailed description is illustrated in Fig. 2.

Figure 3 presents an explainable and generalizable AI architecture for grey-box modelling, with a high-precision temperature control system as a case study. This is detailed in "Ultra-precision machining temperature conditioning system" section to enhance the robustness, generalization, and interpretability of the plant model.

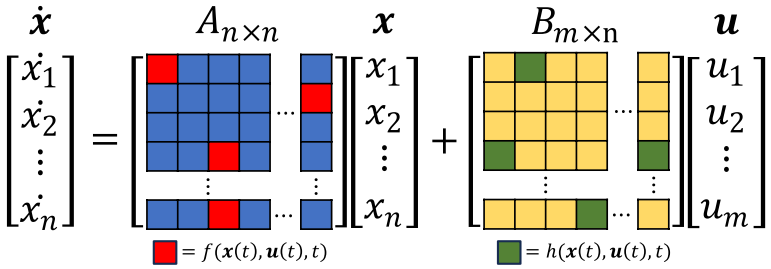

Figure 2. A generalized Linear Time-Varying Grey-Box Model for data and mechanism dual-driven modelling.

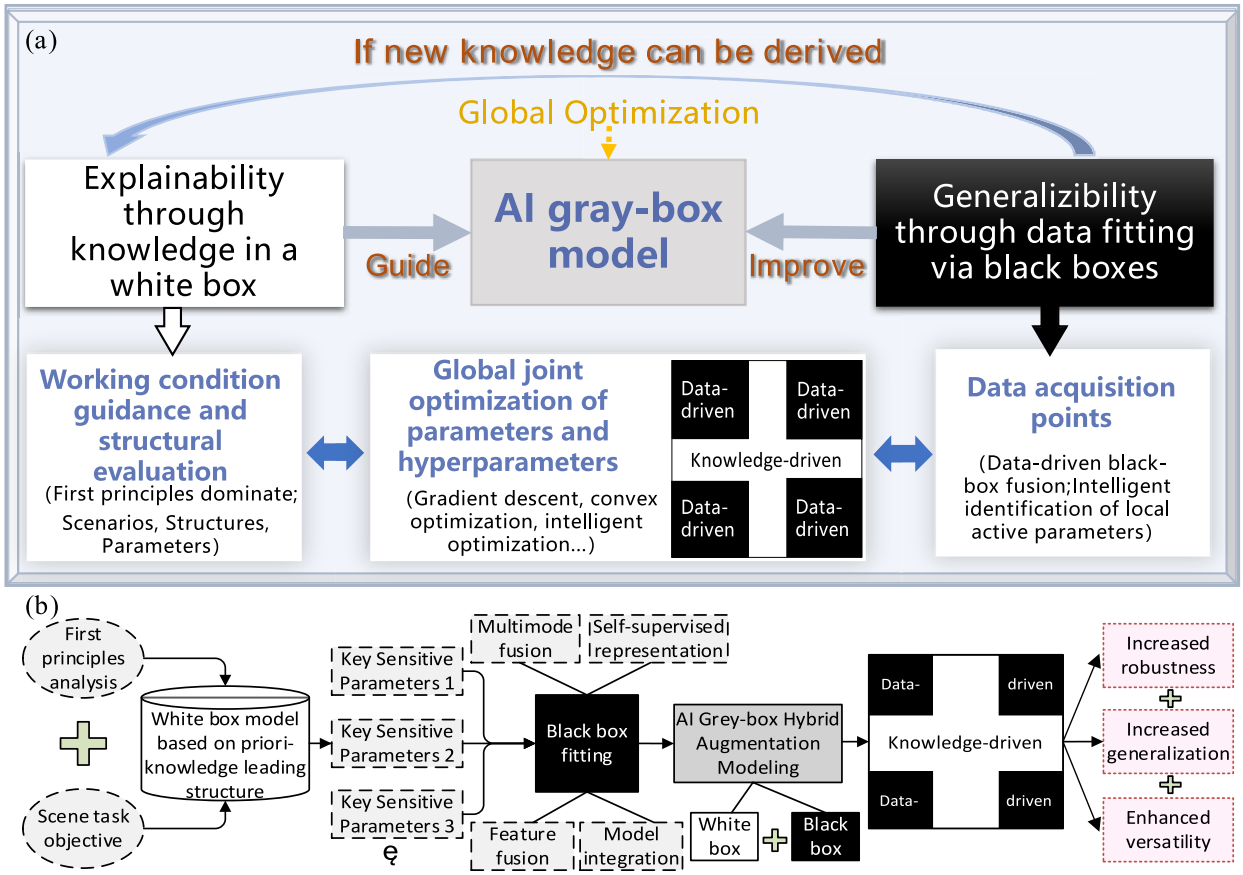

Figure 3. (**a**) Framework of explainable and generalizable AI-driven grey-box modelling. (**b**) Modelling procedures.

Ultra-precision machining temperature conditioning system

System description

System components and requirements for precision control

The high-precision temperature conditioning system ensure efficient heat transfer from stable ambient air temperature (approximately 24–26 °C) to a consistently maintained chilled water temperature (around 12–16 °C, supplied by the chillers). The system is divided into two subsystems based on the difference in outputs, as depicted in Fig. 4. The subsystem for controlling the outlet water temperature (the pink-outlined block) includes a water valve and adjustable water heating. The subsystem for controlling the outlet air temperature (the blue-outlined block) comprises a heat exchange coil and two-stage adjustable electric heating (#1 and #2). Notably, the output of the water temperature system serves as the input for the air temperature system. The system's temperature control objectives (T) are twofold: (1) achieving temperature accuracy with an error of $\Delta T \leq 0.01$ °C for the outlet air compared to the set value, and (2) maintaining the inlet air pressure and temperature, as well as the inlet water temperature, within a specified range.

The temperature conditioning system is designed to pump air through a fan and coil for heat exchange with cold water, ensuring that the ambient temperature does not exceed the reference level. This approach is used because heating air to reference level is more controllable than cooling it. The system then makes coarse and fine adjustments through two stages of air heaters, and the mixer outputs air at the required temperature. Cold water is supplied by external equipment, where a water temperature pre-adjustment system controls the flow rate and temperature of the water entering the coil. Adjustable heaters control the water and air temperatures using pulse-width modulation (PWM) with controlled duty cycles. For high-precision operation, two temperature sensors with an accuracy of 0.001 ℃ are deployed at the air inlet and outlet. Temperature sensors with an accuracy of 0.1 ℃ are installed at the water inlet and outlet, and the water valve uses the average value from these two sensors. The temperature control algorithm utilizes a trajectory controller network (TCN)[98] based on proportional-integral-derivative (PID) scheme[99]. This method is used to control both water and air temperatures in real time.

Challenges of modelling

The challenges in modelling of such a system lie in:

1. Nonlinearities. Real-world temperature control systems often exhibit nonlinear behavior, such as nonlinear relationships in heat transfer and refrigeration. Nonlinearities can lead to unpredictable system behavior, rendering standard control strategies ineffective.
2. Time-variance. Ultra-precision control is influenced by various time-varying factors including changes in internal and environmental conditions, as well as equipment aging. Time-varying characteristics can also arise when operation levels change due to system nonlinearity.
3. Distributed nature of air. The dynamics of air require modelling through partial differential equations with complex boundary conditions. However, this approach is often overly costly in engineering practice, leading to the use of ordinary differential equations (ODEs) as a lumped approximation.
4. Incomplete or noisy data representation. In practice, measured data may be incomplete for various reasons, such as sensor imperfection, disturbances or communication noise.

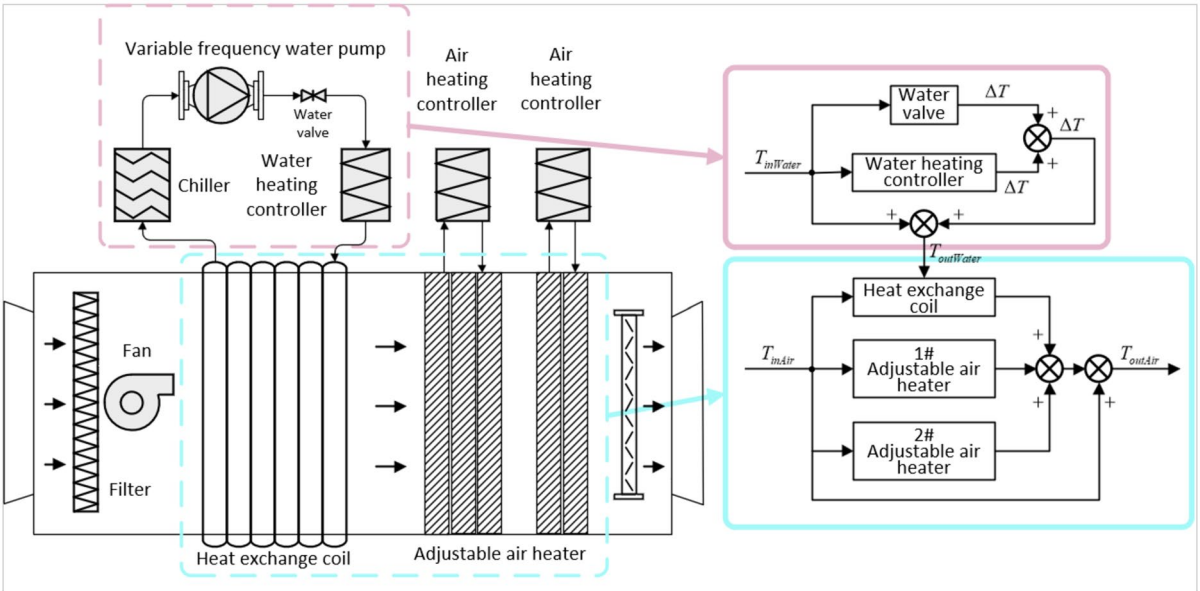

Figure 4. Schematic of a cleanroom high-precision temperature conditioning system.

To ensure optimal control for an ultra-precision temperature systems, model development must account for these factors, handle large amounts of complex data, and provide flexibility to adapt to changes. These challenges have in common that they increase the complexity of modelling and control system design.

System modelling on a white-box global structure

The global structure of the white-box model consists of three subsystems: the heat exchange coil (HEC), the adjustable air heater (AAH), and the adjustable water heater (AWH). All parameters, units and physical explanations are shown in Appendix Table 6.

1. Heat exchange coil

   The HEC serves to reduce the incoming air temperature below the target value. It functions as a secondary air temperature regulation component, with the inlet air supplied by the primary central air conditioner. Consequently, the input air temperature remains relatively stable (24–26 °C), and there is no return air as input. The inlet water for the HEC sourced from the output of the AWH, is stable at 18.5 °C and undergoes minor fluctuations based on current operational conditions and the set value of the outlet air temperature. Therefore, it is reasonable to assume that the inlet water temperature of the HEC remains relatively stable.

   Two primary heat exchange processes occur within the HEC: (1) heat exchange between the cold water and the coil metal, and (2) heat exchange between the air and the coil metal. The energy balance equation for the cooling coil is represented by Eq. (3), while Fig. 5 illustrates the energy balance diagram.

$$\begin{cases} \begin{bmatrix} \dot{T_{oW}} \\ \dot{T_{oA}} \end{bmatrix} = \begin{bmatrix} \frac{m_w C_{pw}}{C_{wm}} & -\frac{m_w C_{pw}+UA_{cc}}{C_{wm}} & 0 & \frac{UA_{cc}}{C_{wm}} \\ 0 & \frac{UA_{cc}}{C_{am}} & \frac{m_a C_{pa}}{C_{am}} & -\frac{m_a C_{pa}+UA_{cc}}{C_{am}} \end{bmatrix} \begin{bmatrix} T_{iW} \\ T_{oW} \\ T_{iA} \\ T_{oA} \end{bmatrix} + \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} \\ \theta = argmin_{\theta \in R^*} L[C_{am}, C_{wm}] \end{cases} \tag{3}$$

   Parameters to be identified are C_{wm}, C_{pw}, C_{pa}, UA_{cc}, c_1, and c_2. Here, UA_{cc} reprensents the total thermal conductivity; T is the temperature; C is the specific heat capacity; and m is the flow rate. The subscripts W, A, *i* and *o* stand for water, air, input, and output, respectively, while c_1 and c_2 are the compensation parameters. Note that the above equations and the subsequent sub-model equations appliy when the temperature of the input medium is constant and the thermal equilibrium has been reached in the flow state[100,101]:

$$T_{iW}(0) = T_{oW}(0), T_{iA}(0) = T_{iW}(0) \tag{4}$$

2. Adjustable air heater

   The AAH consists of two stages: a primary stage (2000W) and a secondary stage (500W), each with a range of 0–1000 for the open degrees. The primary electric heating stage coarsely adjusts the air temperature to approach the set temperature, while the secondary electric heating stage finely adjusts the air temperature to match the set value. Given that the equipment operates continuously and is well-sealed and insulated, heat exchange with environment and within the equipment can be neglected. The simplified energy balance equation for the AAH is provided in Eq. (5), and its corresponding energy balance diagram is shown in Fig. 6.

$$\begin{cases} \dot{T_{oA}} = \begin{bmatrix} -\frac{C_{pa} m_a}{C_{AAH}} \\ \frac{C_{pa} m_a}{C_{AAH}} \end{bmatrix} \begin{bmatrix} T_{iA} \\ T_{oA} \end{bmatrix} + \frac{K_a}{C_{AAH}} \cdot P_a \\ \theta = argmin_{\theta \in R^*} L[K_a, C_{AAH}] \end{cases} \tag{5}$$

   Parameters to be identified include K_a, C_{pa}, are C_{AAH}, where P_a denotes adjustable air heater power, K_a denotes effective power factor, and C_{AAH} denotes thermal conductivity.

3. Adjustable water heater

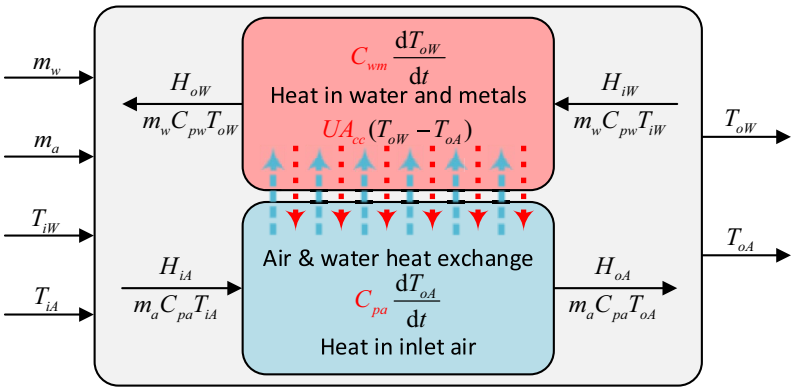

Figure 5. First-principle modelling based on energy balance of the HEC.

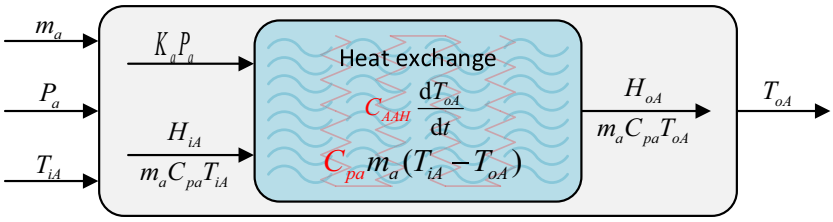

Figure 6. First-principle modelling based on energy balance of the AAH.

The AWH is similar to the AAH, with the primary difference being the heat exchange medium used. The AWH pre-heat the cold water entering the HEC to ensure a stable and controllable input to the HEC. To regulate the amount of cold water entering the HEC more effectively, the AWH incorporates a water valve to control the water flow through the heating element. Since the system is connected to a central chiller, the water valve operates under constant pressure on both sides and follows an equal percentage flow characteristic. Consequently, an exponential function approximates the flow characteristics of the water valve.

The simplified energy balance equation for the AWH is given by Eq. (6), and its corresponding energy balance diagram is illustrated in Fig. 7.

$$\begin{cases} \dot{T_{oW}} = \begin{bmatrix} -\frac{C_{pw} m_w}{C_{AWH}} \\ \frac{C_{pw} m_w}{C_{AWH}} \end{bmatrix} \begin{bmatrix} T_{iW} \\ T_{oW} \end{bmatrix} + \frac{K_w}{C_{AWH}} \cdot P_w \\ \theta = argmin_{\theta \in R^*} L[K_w, C_{AWH}] \end{cases} \tag{6}$$

$$m_w = v^{O_v} - 1 \tag{7}$$

Parameters to be identified are K_w, C_{pw}, and C_{AWH}, where P_w denotes adjustable water heater power, K_w denotes thermal conductivity, C_{AWH} denotes thermal conductivity, O_v denotes the opening of the water valve, and v denotes the flow characteristic coefficient. Equation (7) is obtained from the instruction manual for the water valve.

System model critical factor identification

The white-box mathematical models of the HEC, AAH, and AWH require prior knowledge, such as characteristic coefficients based on equipment-specific lookup tables. However, these a priori knowledge factors possess certain limitations. Firstly, the true values of these factors, including the effects of equipment wear and tear, surface fouling, and uncertainties in heat transfer rates, often differ significantly from their rated values. Secondly, during model identification and training, nonlinear factors are often simplified or neglected, leading to inaccuracies due to production tolerances of other components and measurement errors. These critical factors cannot be ignored.

Additionally, not all parameters within the entire system model hold equal importance. It is necessary to compare experimental data from multiple sets of different operating conditions to identify "critical parameters" and establish their relative importance. This ranking of parameters serves as the foundation for subsequent global optimization of the model.

Critical factor test

The experimental protocol is illustrated in Fig. 8, where the inlet air temperature and inlet water temperature were maintained as stable as possible throughout the entire test. The dashed boxes on the left side of Fig. 8 represent water heating positions, ranging from 0 to 1000 for the water valve and 70 to 100 for the water valve opening. The dashed box on the right side represents electric heating positions, ranging from 0 to 1000. The red coil indicates the heating test, while the blue coil indicates the cooling test. The purple circle marks the beginning of a group test, with the number within the circle indicating the experimental sequence. The symbol "x" denotes that a test is not set for that condition point.

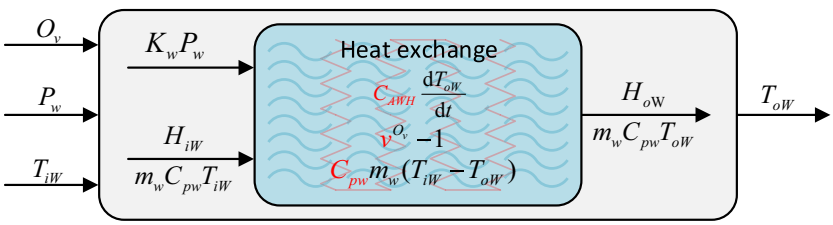

Figure 7. First-principle modelling based on energy balance of the AWH.

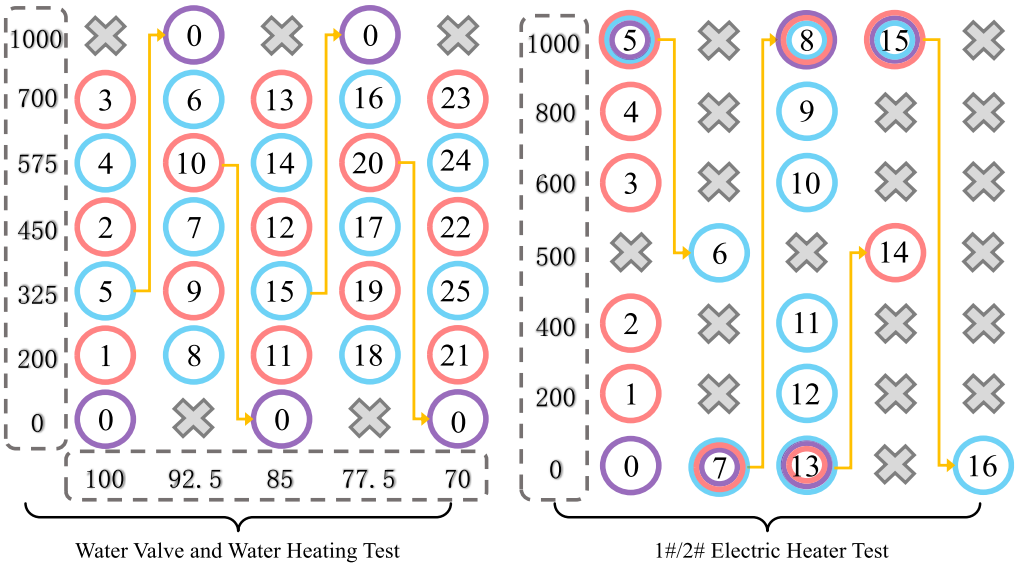

Figure 8. Critical-factor test program.

Analysis of critical factors

Experimental datasets obtained at different operating points are utilized for parameter estimation of the white-box models at each operating point using a Genetic Algorithm (GA), as shown in Fig. 9. The GA is an appropriate method for multi-modal, multi-dimensional problems where optimization objectives are known but detailed knowledge is lacking[102]. Figure 9 illustrates that the parameters of the white-box models vary within the operating envelope. Additionally, due to the different scales or large difference in the mean value, it is not feasible to compare multiple datasets by standard deviation directly. Therefore, the coefficient of variation (CV) method is employed, which is a standardized measure of data dispersion relative to its mean value. The CV is defined as the ratio of the standard deviation to the mean: $c_v = \sigma/\mu$. The estimated results are presented in Table 2.

The results reveal parameter inconsistencies across various operating conditions. Specifically, the thermal conductivity C_{am} shows higher sensitivity in the HEC system, while K_A exhibits greater sensitivity in the first AAH1. Similarly, C_{AAH} demonstrates increased sensitivity in the sceond AAH2, and C_{AWH} exhibits enhanced

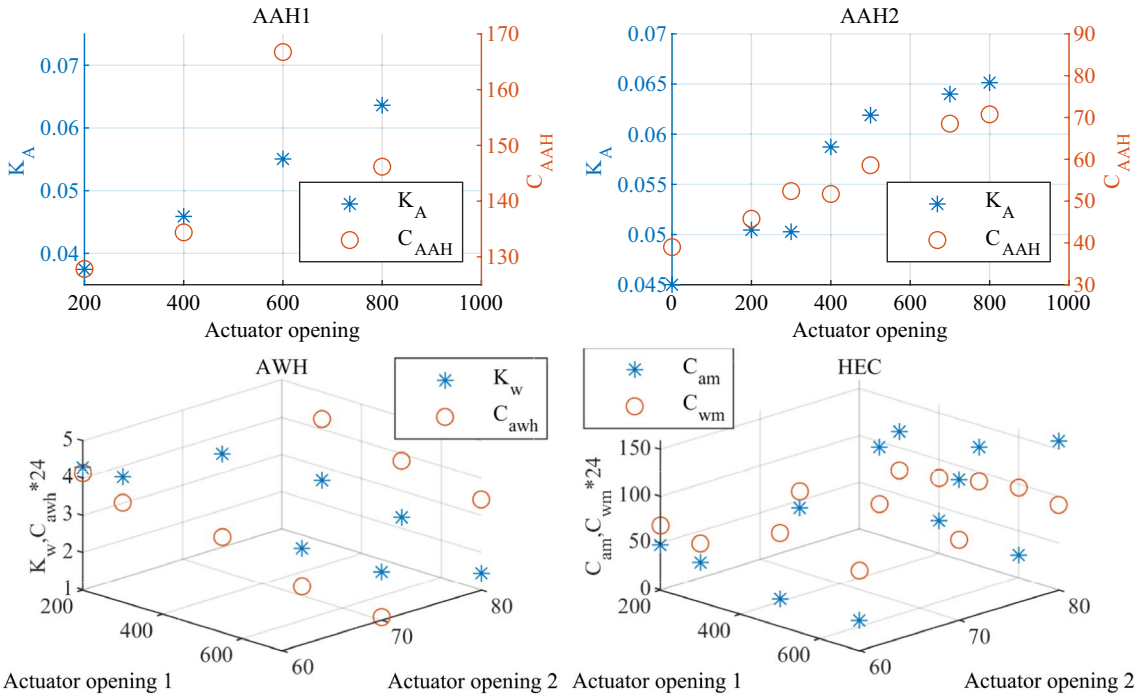

Figure 9. Identification results of subsystem parameters.

	C_{AAH1}	K_{A1}	C_{AAH2}	K_{A2}	C_{wm}	C_{am}	C_{AWH}	K_w
CV	0.1069	0.2529	0.2603	0.1434	0.1361	0.6685	− 0.4451	0.3792

Table 2. Results of the CV assessment.

sensitivity in AWH. Accurately determining these coefficients is crucial for maximizing the precision of the model's output response.

Grey-box modelling

The proposed grey-box model integrates a global physical structure with a local black-box, enabling the modelling of sensitivity coefficients associated with operating points that the white-box cannot accurately captured due to neglected nonlinearities. These sensitivity coefficients vary for each operating point and are distributed in a nonlinear manner. A black-box model for the parameters can be constructed using current operating conditions as inputs and the sensitivity parameters as outputs.

The sensitive coefficients of the white-box model are represented by a black-box component and can be modeled using power series polynomials, rational functions, fuzzy logic, neural networks, or other types of general function approximators based on basis functions.

Given the precise parameter data points obtained under known operating conditions and the strong continuity between them, this study adopts the cubic spline interpolation method for modelling (alternative methods can be considered for different operating conditions). Cubic spline interpolation is preferred for its smoothness and accurate approximation capabilities. The method involves dividing the data interval into sub-intervals and constructing continuous cubic polynomial functions within each sub-interval, ensuring the continuity of both function values and their derivative. This approach provides seamless interpolation curve between data points, effectively avoiding abrupt fluctuations.

Data set description and pre-processing

Description of the dataset

The experimental data were collected from a test center laboratory of a domestic listed company, following the open-loop test program presented in Fig. 8. To achieve decoupling of each part of the system, the openness of each actuating device was reasonably controlled. The specific parameters are described in Table 3.

Handling of data anomalies

Using the test data from the first-stage adjustable air heater (AAH) as an example, the data interval depicted in Fig. 10a corresponds to the AAH opening range of $1000 \rightarrow 500 \rightarrow 0$. The air temperature increment is calculated as the difference between the outgoing air temperature and the incoming air temperature. Figure 10a reveals abnormal fluctuations in the temperature increment curve, even when the actuator opening remains unchanged, and the system reaches a steady state. Sensor failure can be excluded as the cause, indicating the presence of other uncontrollable variables responsible for the anomalies in the experimental results. Upon observing the dataset (Fig. 10b), it is evident that the time and magnitude of the water discharge temperature change coincide with the performance of the abnormal data (Fig. 10a), excluding normal changes caused by measurable actuator opening regulation. Principle analysis further reveals that the system's heat exchange link encompasses the exchange of heat between air and water in the heat exchange coil and the water temperature from the external chiller, which falls outside the experimental control range. Consequently, this portion of the heat exchange is disregarded during data processing, confirming that the abnormal data stem from the water discharge temperature. The processed results are presented in Fig. 10c.

Parameter	Measurement	Details
Wind temperature	Sensor accuracy: 0.0001 ℃	
Water temperature	Sensor accuracy: 0.1 ℃	Significant noise in data, filtered before analysis
Inlet air temperature	Single sensor measurement	Laboratory's inlet air temperature: 23℃ with a fluctuation range of 0.1
Outlet air temperature	Average of two sensors' values	
Sampling	1 s/2 s non-equal intervals	Total sampling time: approximately 25 h
Resampling	Sampling time: 1 s	
Samples obtained	Total: about 32,900	
Selected samples for parameter estimation	Total: about 28,000	Considered sufficient for capturing system's process dynamics under different operating conditions
Inlet water temperature	Stable around 20 °C	

Table 3. Experimental data situation.

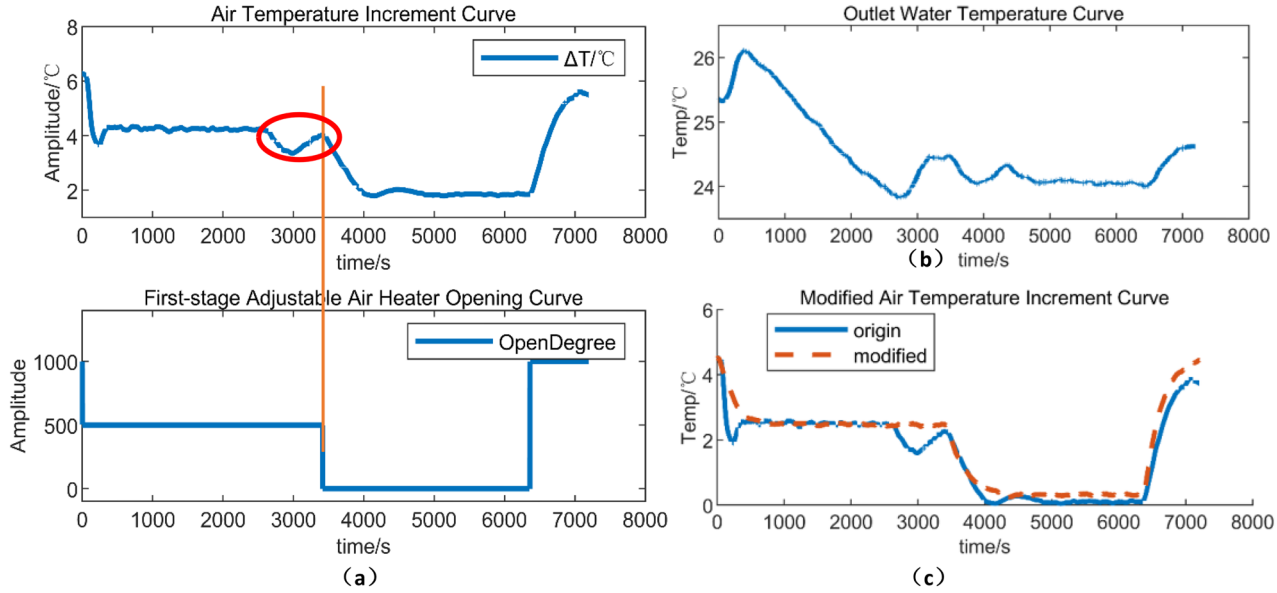

Figure 10. (**a**) First-stage AAH opening curve and air temperature increment curve; (**b**) temperature curve of water discharge; (**c**) revised incremental air temperature curve.

Parameter estimation and results

The parameters of the white-box model were optimized using a GA, and the results are presented in Table 4. All model development and validation procedures were conducted using MATLAB (R2019b). To assess and compare the performance of the white-box model and the grey-box model, the following evaluation metrics were selected. The estimated results are provided in Fig. 11 for reference.

Mean squared error (MSE):

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{y}_i - y_i\right)^2 \tag{8}$$

Coefficient of determination (R2):

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}{\sum_{i=1}^{n}\left(y_i - \overline{y_i}\right)^2} \tag{9}$$

Goodness of fit (G):

$$G = \left(1 - \frac{\sqrt{\sum_{i=1}^{n}\left(\hat{y}_i - y_i\right)^2}}{\sqrt{\sum_{i=1}^{n}\left(y_i - \frac{1}{n}\sum_{i=1}^{n} y_i\right)^2}}\right) \times 100\% \tag{10}$$

Figure 12 illustrates a visual comparison of the outlet air temperature curves for the subsystems and the overall system using both the white-box and grey-box models.

The performance metrics results for the white-box and grey-box models of the subsystem and the overall system are presented in Table 5. The overall model was further validated using an independent dataset of irregular working conditions beyond the scope of the training dataset. This additional dataset assessed the generalization capability of the model as shown in the last column of Table 4. The models generally performed better under the grey-box framework than the pure white-box framework, exhibiting a lower mean square error and higher coefficient of determination. It is worth nothing that validation performance is typically lower than fitting performance, which is expected in any real-world modelling scenario.

Subsystem	White-box parameters estimation results
HEC	$C_{wm} = 4.1251, C_{pa} = 3.1175, C_{pw} = 11.1099,$ $c_1 = 13.4870, c_2 = 4.4172$
AAH1	$K_a = 0.0988, C_{pa} = 19.2102$
AAH2	$K_a = 0.0271, C_{pa} = 19.2476$
AWH	$v = 1.0007, C_{pw} = 71.4786, K_w = 0.1721$

Table 4. White-box parameter estimation results.

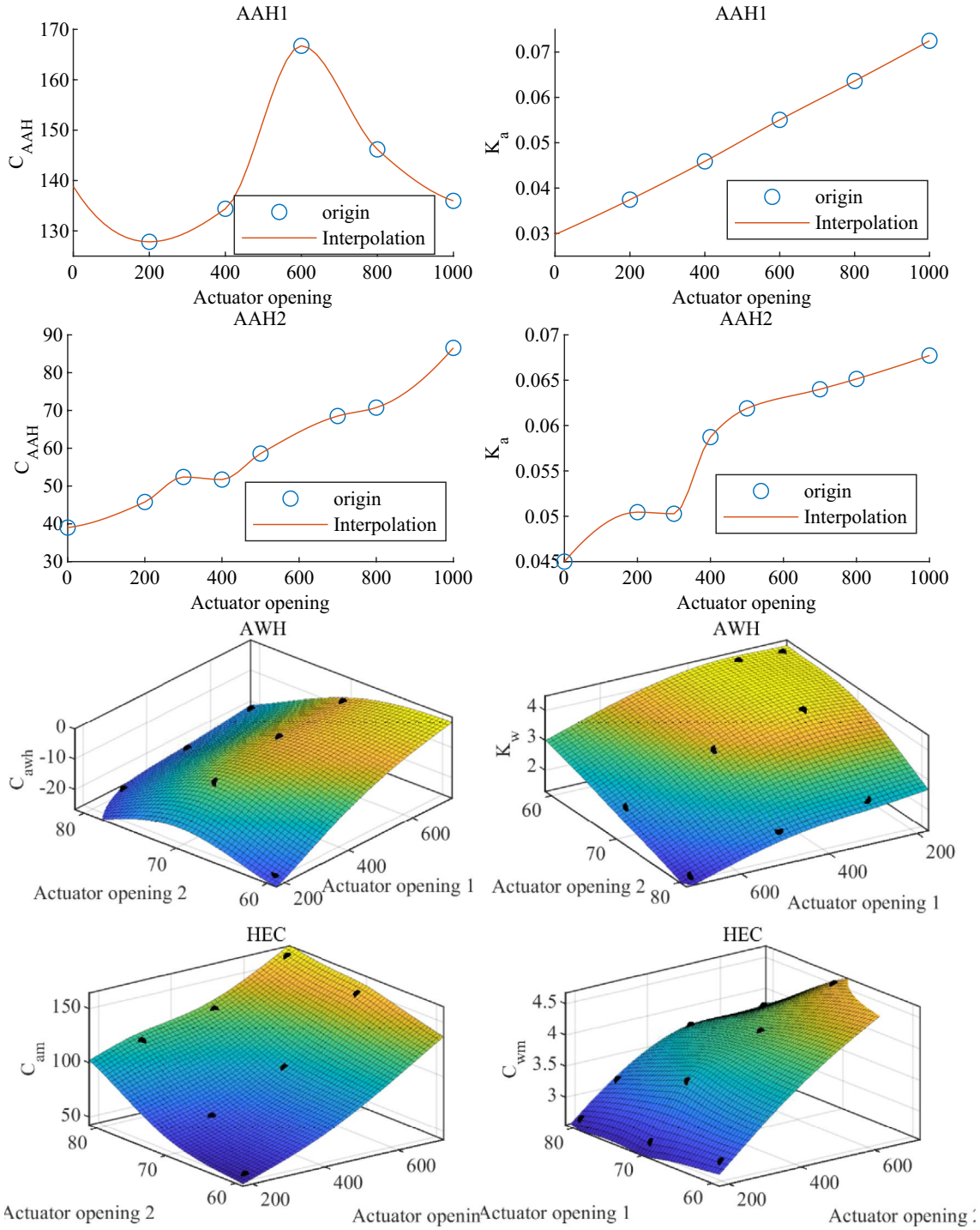

Figure 11. Parameter identification with operating conditions.

Model development and simulation

The model was developed and integrated in Simulink (R2019b), with individual subsystems constructed and interconnected according to the actual unit's configuration as illustrated in Fig. 13. Real experimental input data from the actual unit were passed into the Simulink model through MATLAB's (R2019b) workspace, enabling simulations and analyses that closely resemble real-world conditions. The subsystems were interconnected in series, requiring only the input of the unit's inlet water temperature, inlet air temperature, and the opening of each actuator to obtain the outlet air temperature response. Additionally, temperature data for intermediate stages that are not directly measurable in the experiment, such as the temperature after the coil and the temperature after the first-stage air heating, can also be obtained. To facilitate model parameter optimization by the GA, it is necessary to import the Simulink model's simulation data into Matlab's workspace and obtain the response data by calling the Simulink model within the code space for population iteration of the GA.

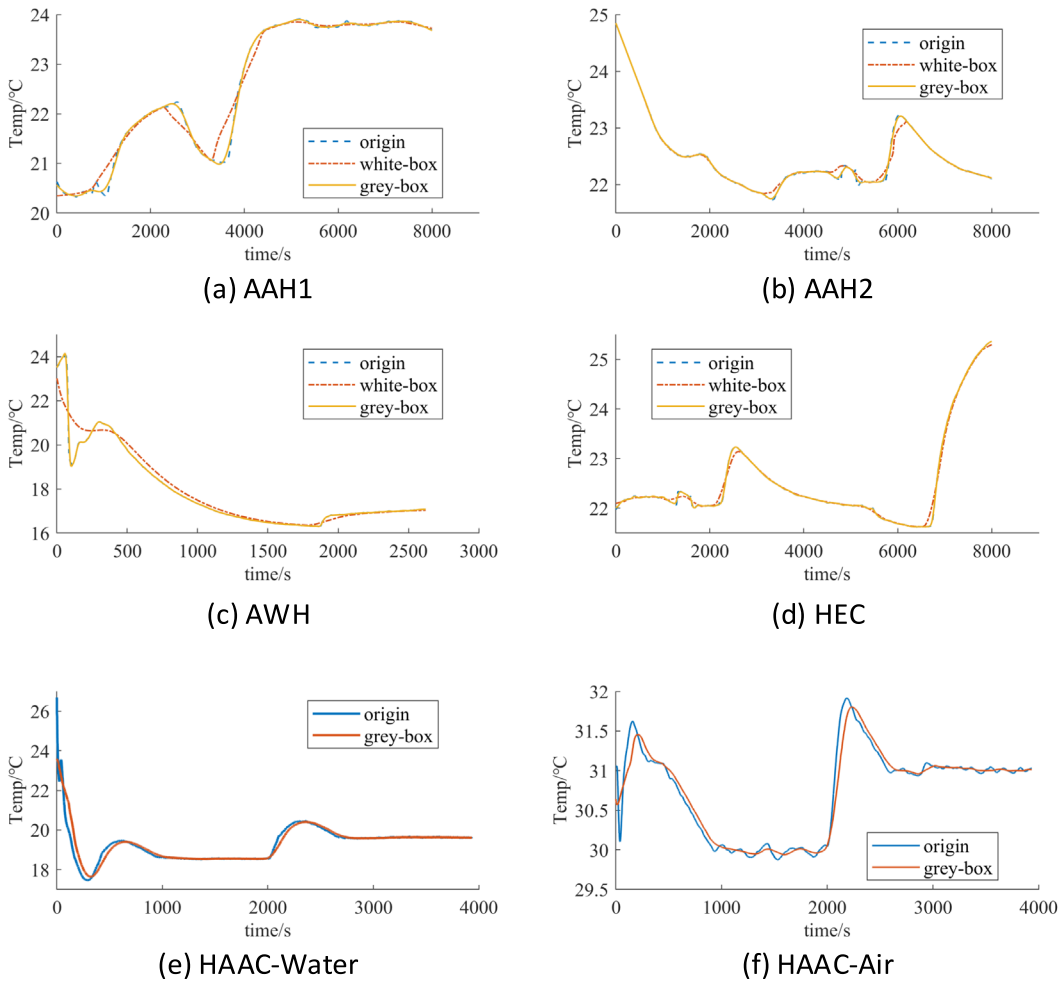

Figure 12. Temperature curves of actual outlet temperature, white-box model outlet temperature, and grey-box model outlet temperature under the same input.

	AAH1		AAH2		AWH		HEC		HAAC-Water		HAAC-Air	
	White	Grey	White	Grey	White	Grey	White	Grey	White	Grey	White	Grey
MSE	0.0181	7.0131e−4	4.8759e−4	6.0046e−5	0.1196	0.0259	0.0223	6.0959e-4	0.5715	0.1646	0.3067	0.1640
R^2	0.9774	0.9991	0.9839	0.9980	0.9893	0.9977	0.9919	0.9998	0.9161	0.9755	0.9350	0.9652
G	84.9632	97.0436	87.3081	95.5461	89.6476	95.1834	90.9824	98.5083	71.0278	84.3559	74.5033	81.3546
p	0.0389		0.0374		0.0461		0.0439		0.0473		0.0396	
Validation set	White: $MSE = 0.33$; $R^2 = 0.90$; $G = 76.8$						Grey: $MSE = 0.0161$; $R^2 = 0.856$; $G = 80.33$					

Table 5. Model performance metrics.

The model is constructed in a cascading manner, providing strong scalability that allows for simulation and research of more complex systems by adding components. By incorporating feedback modules, the development and design of a high-precision control system can be undertaken.

Conclusions

This paper develops an explainable and generalizable AI-driven grey-box modelling method that utilizes the first-principle knowledge and global structure of the real-world system. It integrates accurate black-box models to locally capture nonlinear information of unmeasurable parameters, seamlessly combining it with the white-box model to form an explainable multiscale 'world model'. Through two levels of regularisation, this model transforms a general nonlinear-structure model into a modular one that is generalizable with ease of computational implementations such as when using a GPU.

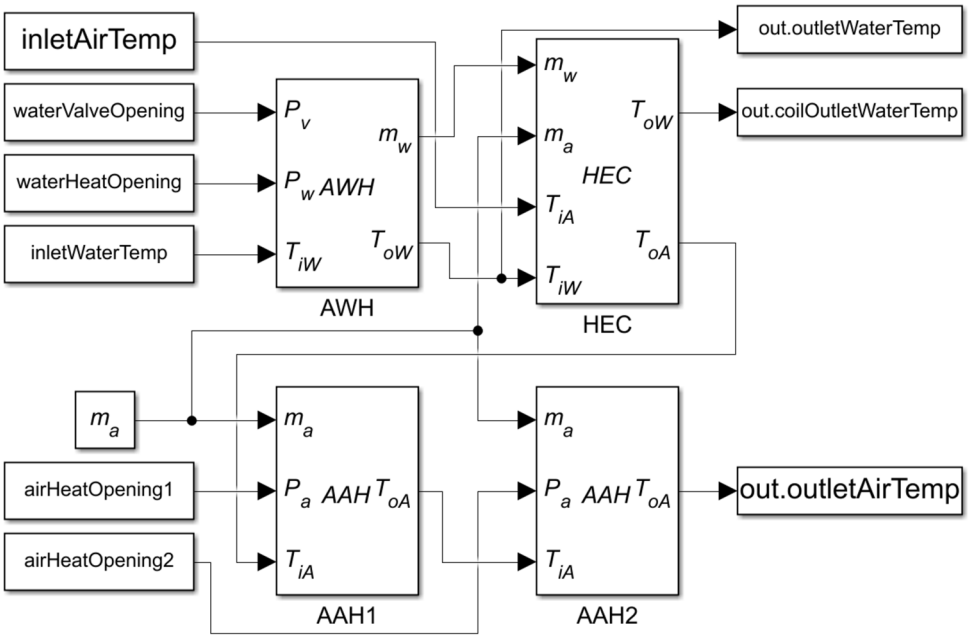

Figure 13. Simulink model.

Then, the approach is applied to and tested on a high-precision thermoregulation system. Its white-box component is effective for modelling relatively known principles, and the comparison in Table 4 reveals that the grey-box model more closely approximates the actual system while retaining explainability with varying operating conditions. A substantial improvement in the MSE metrics is observed in the overall model, where the subsystems are connected in series, causing errors to propagate. The results indicate that the grey-box model enhances accuracy, robustness, and generalization capability. This approach can not only adapt to the changing conditions of the actual engineering system to achieve better modelling accuracy and precise control, but can also provide an effective method for establishing mathematical models of thermophysical systems for ultra-high precision environmental control cabinets in different scenarios. When conditions exceed the training data range, the grey-box model still maintains a certain degree of trustworthy extrapolation owing to its white-box structure, though its accuracy may be compromised. However, the black-box component can be extended easily to other conditions as needed.

This modelling method is suitable for systems with a certain level of physical understanding and cannot be applied to systems whose internal operations are not understood, such as lottery systems, stock trading systems, and pedestrian flow systems. For a larger system, it can be decomposed into subsystems for grey-box modelling. For multi-condition modelling in a complex environment, the black-box component can be modelled using a multi-input multi-output model to enhance the model scalability. Note that, however, a grey-box model may not necessarily outperform a black-box model in terms of local accuracy due to the generalizability of the grey-box model.

The findings offer guidance for modelling similar systems in advancing engineering informatics in the AI age. In the future, robustness and reliability will be studied for enhancing the extrapolation performance of the model. Future research will also further explore the impact of additional factors on the model and will hence consider feasibility of online learning for automatic fine tuning, as well as applying this technique to develop more explainable and generalizable AI-driven control systems for various real-world problems (Supplementary Information).

Data availability

To facilitate the replication of the methods detailed in this paper, we provide the essential modelling code. Regrettably, the raw data utilized in our experiments cannot be publicly disclosed due to confidentiality agreements and were used under a license exclusively for this study. However, the data are available with permission of Shenzhen Envicool Information Technology Co., Ltd, the manufacturer of the equipment used in this study.

Appendix

See Table 6.

Parameter	Unit	Physical interpretation	Parameter	Unit	Physical interpretation
T_{iW}	°C	Temperature of inlet water	UA_{cc}	–	Total thermal conductivity
T_{oW}	°C	Temperature of outlet water	C_{AAH}	–	Thermal conductivity of adjustable air heaters
T_{iA}	°C	Temperature of inlet air	C_{AWH}	–	Thermal conductivity of adjustable water heaters
T_{oA}	°C	Temperature of outlet air	K_a	–	Effective power factor of adjustable air heaters
m_w	kg/s	Mass flow rate of water	K_w	–	Effective power factor of adjustable water heaters
m_a	kg/s	Mass flow rate of air	P_a	kW	Power of adjustable air heater
C_{pw}	J/kg °C	Specific heat capacity of water at constant pressure	P_w	kW	Power of adjustable water heater
C_{pa}	J/kg °C	Specific heat capacity of air at constant pressure	v	–	Flow characteristic coefficient
C_{wm}	J/°C	Thermal capacitance of water and metal	O_v	–	Opening of the water valve
C_{am}	J/°C	Thermal capacitance of air and metal	c_1, c_2	–	Compensation parameters

Table 6. List of symbols.

Received: 5 January 2024; Accepted: 9 July 2024
Published online: 06 August 2024

References

1. Ferreira, P. M., Ruano, A. E., Silva, S. & Conceição, E. Z. E. Neural networks based predictive control for thermal comfort and energy savings in public buildings. *Energy Build. Cool Roofs Cool Pav. Cool Cities Cool World* **55**, 238–251. https://doi.org/10.1016/j.enbuild.2012.08.002 (2012).
2. Li, Y. & Häußler, A. Artificial evolution of neural networks and its application to feedback control. *Artif. Intell. Eng.* **10**, 143–152. https://doi.org/10.1016/0954-1810(95)00024-0 (1996).
3. Di Francesco, M., Veldenz, L., Dell'Anno, G. & Potter, K. Heater power control for multi-material, variable speed automated fibre placement. *Compos. Part Appl. Sci. Manuf.* **101**, 408–421. https://doi.org/10.1016/j.compositesa.2017.06.015 (2017).
4. Guo, D. *et al.* Application of thermoelectric cooler in temperature control system of space science experiment. *Appl. Therm. Eng.* **168**, 114888. https://doi.org/10.1016/j.applthermaleng.2019.114888 (2020).
5. Liu, X., Li, Z., Jiang, Y. & Lin, B. Annual performance of liquid desiccant based independent humidity control HVAC system. *Appl. Therm. Eng.* **26**, 1198–1207. https://doi.org/10.1016/j.applthermaleng.2005.10.043 (2006).
6. Wu, J. *et al.* Intelligent diagnosis method of data center precision air conditioning fault based on knowledge graph. *Electronics* **12**, 498. https://doi.org/10.3390/electronics12030498 (2023).
7. Bi, Q. *et al.* Advanced controller auto-tuning and its application in HVAC systems. *Control Eng. Pract.* **8**, 633–644. https://doi.org/10.1016/S0967-0661(99)00198-7 (2000).
8. Huang, G., Wang, S. & Xu, X. Robust model predictive control of VAV air-handling units concerning uncertainties and constraints. *HVAC R Res.* **16**, 15–33. https://doi.org/10.1080/10789669.2010.10390890 (2010).
9. Huang, G., Wang, S. & Xu, X. A robust model predictive control strategy for improving the control performance of air-conditioning systems. *Energy Convers. Manag.* **50**, 2650–2658. https://doi.org/10.1016/j.enconman.2009.06.014 (2009).
10. Kulkarni, M. R. & Hong, F. Energy optimal control of a residential space-conditioning system based on sensible heat transfer modeling. *Build. Environ.* **39**, 31–38. https://doi.org/10.1016/j.buildenv.2003.07.003 (2004).
11. Prívara, S., Široký, J., Ferkl, L. & Cigler, J. Model predictive control of a building heating system: The first experience. *Energy Build.* **43**, 564–572. https://doi.org/10.1016/j.enbuild.2010.10.022 (2011).
12. Rehrl, J. & Horn, M. Temperature control for HVAC systems based on exact linearization and model predictive control. In *2011 IEEE International Conference on Control Applications (CCA). Presented at the 2011 IEEE International Conference on Control Applications (CCA)*, 1119–1124 https://doi.org/10.1109/CCA.2011.6044437 (2011).
13. Afram, A. & Janabi-Sharifi, F. Black-box modeling of residential HVAC system and comparison of gray-box and black-box modeling methods. *Energy Build.* **94**, 121–149. https://doi.org/10.1016/j.enbuild.2015.02.045 (2015).
14. Aydinalp-Koksal, M. & Ugursal, V. I. Comparison of neural network, conditional demand analysis, and engineering approaches for modeling end-use energy consumption in the residential sector. *Appl. Energy* **85**, 271–296. https://doi.org/10.1016/j.apenergy.2006.09.012 (2008).
15. Frausto, H. U., Pieters, J. G. & Deltour, J. M. Modelling greenhouse temperature by means of auto regressive models. *Biosyst. Eng.* **84**, 147–157. https://doi.org/10.1016/S1537-5110(02)00239-8 (2003).
16. Zhao, H. X. & Magoulès, F. Parallel support vector machines applied to the prediction of multiple buildings energy consumption. *J. Algorithms Comput. Technol.* **4**, 231–249 (2010).
17. Homod, R. Z. Review on the HVAC system modeling types and the shortcomings of their application. *J. Energy* **2013**, e768632. https://doi.org/10.1155/2013/768632 (2013).
18. Huang, H., Chen, L. & Hu, E. A neural network-based multi-zone modelling approach for predictive control system design in commercial buildings. *Energy Build.* **97**, 86–97. https://doi.org/10.1016/j.enbuild.2015.03.045 (2015).
19. Mustafaraj, G., Chen, J. & Lowry, G. Development of room temperature and relative humidity linear parametric models for an open office using BMS data. *Energy Build.* **42**, 348–356. https://doi.org/10.1016/j.enbuild.2009.10.001 (2010).
20. Okochi, G. S. & Yao, Y. A review of recent developments and technological advancements of variable-air-volume (VAV) air-conditioning systems. *Renew. Sustain. Energy Rev.* **59**, 784–817. https://doi.org/10.1016/j.rser.2015.12.328 (2016).
21. Ríos-Moreno, G. J., Trejo-Perea, M., Castañeda-Miranda, R., Hernández-Guzmán, V. M. & Herrera-Ruiz, G. Modelling temperature in intelligent buildings by means of autoregressive models. *Autom. Constr.* **16**, 713–722. https://doi.org/10.1016/j.autcon.2006.11.003 (2007).
22. Afram, A. & Janabi-Sharifi, F. Gray-box modeling and validation of residential HVAC system for control system design. *Appl. Energy* **137**, 134–150. https://doi.org/10.1016/j.apenergy.2014.10.026 (2015).
23. Liang, Y., Li, S., Yan, C., Li, M. & Jiang, C. Explaining the black-box model: A survey of local interpretation methods for deep neural networks. *Neurocomputing* **419**, 168–182. https://doi.org/10.1016/j.neucom.2020.08.011 (2021).
24. Romero, J. A., Navarro-Esbrí, J. & Belman-Flores, J. M. A simplified black-box model oriented to chilled water temperature control in a variable speed vapour compression system. *Appl. Therm. Eng.* **31**, 329–335. https://doi.org/10.1016/j.applthermaleng.2010.09.013 (2011).

25. Azarpour, A., Borhani, T. N., Alwi, S. R., Manan, Z. A. & Mutalib, M. I. A generic hybrid model development for process analysis of industrial fixed-bed catalytic reactors. *Chem. Eng. Res. Des.* **117**, 149–167. https://doi.org/10.1016/j.cherd.2016.10.024 (2017).
26. Dong, Y. & Qin, S. J. Dynamic latent variable analytics for process operations and control. *Comput. Chem. Eng.* **114**, 69–80. https://doi.org/10.1016/j.compchemeng.2017.10.029 (2018).
27. Dong, Y. & Qin, S. J. Regression on dynamic PLS structures for supervised learning of dynamic data. *J. Process Control* **68**, 64–72. https://doi.org/10.1016/j.jprocont.2018.04.006 (2018).
28. Gurden, S. P., Westerhuis, J. A., Bijlsma, S. & Smilde, A. K. Modelling of spectroscopic batch process data using grey models to incorporate external information. *J. Chemom.* **15**, 101–121. https://doi.org/10.1002/1099-128X(200102)15:2%3c101::AID-CEM602%3e3.0.CO;2-V (2001).
29. van Lith, P. F. Hybrid fuzzy-first principles modelling (2002).
30. Pasquier, R. & Smith, I. F. C. Robust system identification and model predictions in the presence of systematic uncertainty. *Adv. Eng. Inform. Collect. Intell. Model. Anal. Synth. Innov. Eng. Decision Mak.* **29**, 1096–1109. https://doi.org/10.1016/j.aei.2015.07.007 (2015).
31. Solle, D. *et al.* Between the poles of data-driven and mechanistic modeling for process operation. *Chem.-Ing.-Tech.* **89**, 542–561. https://doi.org/10.1002/cite.201600175 (2017).
32. Chai, Z. & Zhao, C. Enhanced random forest with concurrent analysis of static and dynamic nodes for industrial fault classification. *IEEE Trans. Ind. Inform.* **16**, 54–66. https://doi.org/10.1109/TII.2019.2915559 (2020).
33. Adadi, A. & Berrada, M. Peeking inside the black-box: A survey on explainable artificial intelligence (XAI). *IEEE Access* **6**, 52138–52160. https://doi.org/10.1109/ACCESS.2018.2870052 (2018).
34. Barredo Arrieta, A. *et al.* Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Inf. Fusion* **58**, 82–115. https://doi.org/10.1016/j.inffus.2019.12.012 (2020).
35. Gernaey, K. V., van Loosdrecht, M. C. M., Henze, M., Lind, M. & Jørgensen, S. B. Activated sludge wastewater treatment plant modelling and simulation: State of the art. *Environ. Model. Softw. Environ. Sci. Artif. Intell.* **19**, 763–783. https://doi.org/10.1016/j.envsoft.2003.03.005 (2004).
36. Wernick, P. & Lehman, M. M. Software process white box modelling for FEAST/1. *J. Syst. Softw.* **46**, 193–201. https://doi.org/10.1016/S0164-1212(99)00012-6 (1999).
37. Zendehboudi, S., Rezaei, N. & Lohi, A. Applications of hybrid models in chemical, petroleum, and energy systems: A systematic review. *Appl. Energy* **228**, 2539–2566. https://doi.org/10.1016/j.apenergy.2018.06.051 (2018).
38. Kourti, T. & MacGregor, J. F. Process analysis, monitoring and diagnosis, using multivariate projection methods. *Chemom. Intell. Lab. Syst.* **28**, 3–21. https://doi.org/10.1016/0169-7439(95)80036-9 (1995).
39. Ljung, L. Perspectives on system identification. *Annu. Rev. Control* **34**, 1–12. https://doi.org/10.1016/j.arcontrol.2009.12.001 (2010).
40. MacGregor, J. F. & Kourti, T. Statistical process control of multivariate processes. *Control Eng. Pract.* **3**, 403–414. https://doi.org/10.1016/0967-0661(95)00014-L (1995).
41. Rogers, T. J., Holmes, G. R., Cross, E. J., Worden, K. On a grey box modelling framework for nonlinear system identification. In *Presented at the Conference Proceedings of the Society for Experimental Mechanics Series*, 167–178. https://doi.org/10.1007/978-3-319-53841-9_15 (2017).
42. Chen, Y. & Ierapetritou, M. A framework of hybrid model development with identification of plant-model mismatch. *AIChE J.* **66**, e16996. https://doi.org/10.1002/aic.16996 (2020).
43. Du, M., Liu, N. & Hu, X. Techniques for interpretable machine learning. *Commun. ACM* **63**, 68–77. https://doi.org/10.1145/3359786 (2020).
44. Ghosh, D., Hermonat, E., Mhaskar, P., Snowling, S. & Goel, R. Hybrid modeling approach integrating first-principles models with subspace identification. *Ind. Eng. Chem. Res.* **58**, 13533–13543. https://doi.org/10.1021/acs.iecr.9b00900 (2019).
45. Hao, C. *et al.* A novel deep learning method with partly explainable: Intelligent milling tool wear prediction model based on transformer informed physics. *Adv. Eng. Inform.* **57**, 102106. https://doi.org/10.1016/j.aei.2023.102106 (2023).
46. Oliveira, R. Combining first principles modelling and artificial neural networks: A general framework. *Comput. Chem. Eng.* **28**, 755–766. https://doi.org/10.1016/j.compchemeng.2004.02.014 (2004).
47. Reis, M. S., Gins, G. & Rato, T. J. Incorporation of process-specific structure in statistical process monitoring: A review. *J. Qual. Technol.* **51**, 407–421. https://doi.org/10.1080/00224065.2019.1569954 (2019).
48. Rendall, R. & Reis, M. S. Which regression method to use? Making informed decisions in "data-rich/knowledge poor" scenarios—The Predictive Analytics Comparison framework (PAC). *Chemom. Intell. Lab. Syst.* **181**, 52–63. https://doi.org/10.1016/j.chemolab.2018.08.004 (2018).
49. Friston, K. *et al.* World model learning and inference. *Neural Netw.* **144**, 573–590. https://doi.org/10.1016/j.neunet.2021.09.011 (2021).
50. Hosain, Md. T., Jim, J. R., Mridha, M. F. & Kabir, M. M. Explainable AI approaches in deep learning: Advancements, applications and challenges. *Comput. Electr. Eng.* **117**, 109246. https://doi.org/10.1016/j.compeleceng.2024.109246 (2024).
51. Daw, A., Karpatne, A., Watkins, W. D., Read, J. S. & Kumar, V. Physics-guided neural networks (PGNN): An application in lake temperature modeling, In *Knowledge Guided Machine Learning* (Chapman and Hall/CRC, 2022).
52. Karpatne, A. *et al.* Theory-guided data science: A new paradigm for scientific discovery from data. *IEEE Trans. Knowl. Data Eng.* **29**, 2318–2331. https://doi.org/10.1109/TKDE.2017.2720168 (2017).
53. Raissi, M., Perdikaris, P. & Karniadakis, G. E. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J. Comput. Phys.* **378**, 686–707. https://doi.org/10.1016/j.jcp.2018.10.045 (2019).
54. Wang, N., Zhang, D., Chang, H. & Li, H. Deep learning of subsurface flow via theory-guided neural network. *J. Hydrol.* **584**, 124700. https://doi.org/10.1016/j.jhydrol.2020.124700 (2020).
55. Chen, Y. & Treado, S. Development of a simulation platform based on dynamic models for HVAC control analysis. *Energy Build.* **68**, 376–386. https://doi.org/10.1016/j.enbuild.2013.09.016 (2014).
56. Ghiaus, C. & Hazyuk, I. Calculation of optimal thermal load of intermittently heated buildings. *Energy Build.* **42**, 1248–1258. https://doi.org/10.1016/j.enbuild.2010.02.017 (2010).
57. Yang, W.-T., Blue, J., Roussy, A., Pinaton, J. & Reis, M. S. A physics-informed Run-to-Run control framework for semiconductor manufacturing. *Expert Syst. Appl.* **155**, 113424. https://doi.org/10.1016/j.eswa.2020.113424 (2020).
58. Attaran, S. M., Yusof, R. & Selamat, H. A novel optimization algorithm based on epsilon constraint-RBF neural network for tuning PID controller in decoupled HVAC system. *Appl. Therm. Eng.* **99**, 613–624. https://doi.org/10.1016/j.applthermaleng.2016.01.025 (2016).
59. Wang, S. & Ma, Z. Supervisory and optimal control of building HVAC systems: A review. *HVACR Res.* **14**, 3–32. https://doi.org/10.1080/10789669.2008.10390991 (2008).
60. Afroz, Z., Shafiullah, G., Urmee, T. & Higgins, G. Modeling techniques used in building HVAC control systems: A review. *Renew. Sustain. Energy Rev.* **83**, 64–84. https://doi.org/10.1016/j.rser.2017.10.044 (2018).
61. Tan, K. C. & Li, Y. Grey-box model identification via evolutionary computing. *Control Eng. Pract. Dev. High Precis. Servo Syst.* **10**, 673–684. https://doi.org/10.1016/S0967-0661(02)00031-X (2002).

62. Ghiaus, C., Chicinas, A. & Inard, C. Grey-box identification of air-handling unit elements. *Control Eng. Pract.* **15**, 421–433. https://doi.org/10.1016/j.conengprac.2006.08.005 (2007).
63. Hassanpour, H. A hybrid modeling approach integrating first-principles knowledge with statistical methods for fault detection in HVAC systems. *Comput. Chem. Eng.* **142**, 107022. https://doi.org/10.1016/j.compchemeng.2020.107022 (2020).
64. Safari, S., Shabani, F. & Simon, D. Multirate multisensor data fusion for linear systems using Kalman filters and a neural network. *Aerosp. Sci. Technol.* **39**, 465–471. https://doi.org/10.1016/j.ast.2014.06.005 (2014).
65. Sansana, J. *et al.* Recent trends on hybrid modeling for Industry 40. *Comput. Chem. Eng.* **151**, 107365. https://doi.org/10.1016/j.compchemeng.2021.107365 (2021).
66. Braun, J. E. & Chaturvedi, N. An inverse gray-box model for transient building load prediction. *HVAC R Res.* **8**, 73–99. https://doi.org/10.1080/10789669.2002.10391290 (2002).
67. Arendt, P. D., Apley, D. W. & Chen, W. Quantification of model uncertainty: Calibration, model discrepancy, and identifiability. *J. Mech. Des. Doi* **10**(1115/1), 4007390 (2012).
68. Ebers, J. J. & Moll, J. L. Large-signal behavior of junction transistors. *Proc. IRE* **42**, 1761–1772. https://doi.org/10.1109/JRPROC.1954.274797 (1954).
69. Gummel, H. K. & Poon, H. C. An integral charge control model of bipolar transistors. *Bell Syst. Tech. J.* **49**, 827–852. https://doi.org/10.1002/j.1538-7305.1970.tb01803.x (1970).
70. Huang, S. *et al.* Geometric variability aware quantum potential based quasi-ballistic compact model for stacked 6 nm-thick silicon nanosheet GAA-FETs. In *2021 IEEE International Electron Devices Meeting (IEDM). Presented at the 2021 IEEE International Electron Devices Meeting (IEDM)*, 18.5.1–18.5.4. https://doi.org/10.1109/IEDM19574.2021.9720550 (2021).
71. Zhao, Y. *et al.* A unified physical BTI compact model in variability-aware DTCO flow: Device characterization and circuit evaluation on reliability of scaling technology nodes. In *2021 Symposium on VLSI Technology. Presented at the 2021 Symposium on VLSI Technology*, 1–2 (2021).
72. Li, X. *et al.* Overview of emerging semiconductor device model methodologies: From device physics to machine learning engines. *Fundam. Res.* https://doi.org/10.1016/j.fmre.2024.01.010 (2024).
73. Wang, Y. *et al.* Application of machine learning for composite moulding process modelling. *Compos. Commun.* **48**, 101960. https://doi.org/10.1016/j.coco.2024.101960 (2024).
74. Kou, J. & Zhang, W. Data-driven modeling for unsteady aerodynamics and aeroelasticity. *Prog. Aerosp. Sci.* **125**, 100725. https://doi.org/10.1016/j.paerosci.2021.100725 (2021).
75. Yang, A., Martin, E. & Morris, J. Identification of semi-parametric hybrid process models. *Comput. Chem. Eng.* **35**, 63–70. https://doi.org/10.1016/j.compchemeng.2010.05.002 (2011).
76. Yang, X., Wang, J. & Liu, J. High temperature LCF life prediction of notched DS Ni-based superalloy using critical distance concept. *Int. J. Fatigue* **33**, 1470–1476. https://doi.org/10.1016/j.ijfatigue.2011.05.018 (2011).
77. Oswald, W. B. General formulas and charts for the calculation of airplane performance (1933).
78. Brunton, S. L. *et al.* Data-driven aerospace engineering: Reframing the industry with machine learning. *AIAA J.* https://doi.org/10.2514/1.J060131 (2021).
79. Wang, J., Xu, N., Choi, W., Lee, K.-H. & Park, Y. A generic approach for capturing process variations in lookup-table-based FET models. In *2015 International Conference on Simulation of Semiconductor Processes and Devices (SISPAD). Presented at the 2015 International Conference on Simulation of Semiconductor Processes and Devices (SISPAD)*, 309–312 https://doi.org/10.1109/SISPAD.2015.7292321 (2015).
80. Chiozzi, D., Bernardoni, M., Delmonte, N. & Cova, P. A neural network based approach to simulate electrothermal device interaction in SPICE environment. *IEEE Trans. Power Electron.* **34**, 4703–4710. https://doi.org/10.1109/TPEL.2018.2863186 (2019).
81. Zhou, T., Jiang, S., Han, T., Zhu, S.-P. & Cai, Y. A physically consistent framework for fatigue life prediction using probabilistic physics-informed neural network. *Int. J. Fatigue* **166**, 107234. https://doi.org/10.1016/j.ijfatigue.2022.107234 (2023).
82. An, D., Kim, N. H. & Choi, J.-H. Practical options for selecting data-driven or physics-based prognostics algorithms with reviews. *Reliab. Eng. Syst. Saf.* **133**, 223–236. https://doi.org/10.1016/j.ress.2014.09.014 (2015).
83. Litovski, V. B., Radjenovic, J. I., Mrcarica, Z. M. & Milenkovic, S. L. MOS transistor modelling using neural network. *Electron. Lett.* **28**, 1766–1768. https://doi.org/10.1049/el:19921124 (1992).
84. Shi, M., Mo, P. & Liu, J. Deep neural network for accurate and efficient atomistic modeling of phase change memory. *IEEE Electron Device Lett.* **41**, 365–368. https://doi.org/10.1109/LED.2020.2964779 (2020).
85. Bui-Thanh, T., Willcox, K., Ghattas, O. & van Bloemen Waanders, B. Goal-oriented, model-constrained optimization for reduction of large-scale systems. *J. Comput. Phys.* **224**, 880–896. https://doi.org/10.1016/j.jcp.2006.10.026 (2007).
86. Han, S.-C., Choi, J. & Hong, S.-M. Acceleration of semiconductor device simulation with approximate solutions predicted by trained neural networks. *IEEE Trans. Electron Devices* **68**, 5483–5489. https://doi.org/10.1109/TED.2021.3075192 (2021).
87. Le Clainche, S. *et al.* Improving aircraft performance using machine learning: A review. *Aerosp. Sci. Technol.* **138**, 108354. https://doi.org/10.1016/j.ast.2023.108354 (2023).
88. Azzam, H. A practical approach for the indirect prediction of structural fatigue from measured flight parameters. *Proc. Inst. Mech. Eng. Part G J. Aerosp. Eng.* **211**, 29–38. https://doi.org/10.1243/0954410971532479 (1997).
89. Bürkle, M. *et al.* Deep-learning approach to first-principles transport simulations. *Phys. Rev. Lett.* **126**, 177701. https://doi.org/10.1103/PhysRevLett.126.177701 (2021).
90. Gao, J., Wang, J., Xu, Z., Wang, C. & Yan, S. Multiaxial fatigue prediction and uncertainty quantification based on back propagation neural network and Gaussian process regression. *Int. J. Fatigue* **168**, 107361. https://doi.org/10.1016/j.ijfatigue.2022.107361 (2023).
91. Li, M., İrsoy, O., Cardie, C. & Xing, H. G. Physics-inspired neural networks for efficient device compact modeling. *IEEE J. Explor. Solid-State Comput. Devices Circuits* **2**, 44–49. https://doi.org/10.1109/JXCDC.2016.2636161 (2016).
92. Carlberg, K., Farhat, C., Cortial, J. & Amsallem, D. The GNAT method for nonlinear model reduction: Effective implementation and application to computational fluid dynamics and turbulent flows. *J. Comput. Phys.* **242**, 623–647. https://doi.org/10.1016/j.jcp.2013.02.028 (2013).
93. Taira, K. *et al.* Modal analysis of fluid flows: Applications and outlook. *AIAA J.* **58**, 998–1022. https://doi.org/10.2514/1.J058462 (2020).
94. Dourado, A. & Viana, F. A. C. Physics-informed neural networks for missing physics estimation in cumulative damage models: A case study in corrosion fatigue. *J. Comput. Inf. Sci. Eng.* https://doi.org/10.1115/1.4047173 (2020).
95. Pitchforth, D. J., Rogers, T. J., Tygesen, U. T. & Cross, E. J. Grey-box models for wave loading prediction. *Mech. Syst. Signal Process.* **159**, 107741. https://doi.org/10.1016/j.ymssp.2021.107741 (2021).
96. Cross, E. J. *et al.* Physics-informed machine learning for structural health monitoring. *Struct. Health Monit. Based Data Sci. Tech.* https://doi.org/10.1007/978-3-030-81716-9_17 (2021).
97. Sohlberg, B. Grey box modelling for model predictive control of a heating process. *J. Process Control* **13**, 225–238. https://doi.org/10.1016/S0959-1524(02)00030-6 (2003).
98. Li, Y. & Chong, G. Evolving trajectory controller networks from linear approximation model networks. In *Proceedings of the 2000 Congress on Evolutionary Computation. CEC00 (Cat. No.00TH8512). Presented at the Proceedings of the 2000 Congress on Evolutionary Computation. CEC00 (Cat. No.00TH8512)*, Vol. 1, 251–255 https://doi.org/10.1109/CEC.2000.870303 (2000) .

99. Ang, K. H., Chong, G. & Li, Y. PID control system analysis, design, and technology. *IEEE Trans. Control Syst. Technol.* **13**, 559–576. https://doi.org/10.1109/TCST.2005.847331 (2005).
100. Turkyilmazoglu, M. Thermal management of parabolic pin fin subjected to a uniform oncoming airflow: Optimum fin dimensions. *J. Therm. Anal. Calorim.* **143**, 3731–3739. https://doi.org/10.1007/s10973-020-10382-x (2021).
101. Turkyilmazoglu, M. Efficiency of the longitudinal fins of trapezoidal profile in motion. *J. Heat Transf.* https://doi.org/10.1115/1.4036328 (2017).
102. Li, L. *et al.* Benchmarks for evaluating optimization algorithms and benchmarking MATLAB derivative-free optimizers for practitioners' rapid access. *IEEE Access* **7**, 79657–79670. https://doi.org/10.1109/ACCESS.2019.2923092 (2019).

Acknowledgements
Project supported by the Major Research Plan of the National Natural Science Foundation of China (Grant No. 92270105).

Author contributions
C.L and A.L completed the text as well as the tables and pictures drawn throughout the article. C.L was the main idea and designer of the article. A.L was the experimental designer of the text and acquired and analysed the data. Y.L critically revised and captured the quality of the article for important intellectual content and gave final approval for publication. J.X and M.L participated in the field trials.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-024-67259-4.

Correspondence and requests for materials should be addressed to Y.L.