Industry Science and Engineering Vol. 3 No. 6, 2026

Noise-Aware Delay-Compensated RL for Robust Thermal Fan Control

Shiqi Zhang[1], Wen Li[1,*], Jingyan Min[1], Weibo Tao[2], Yazhi Yang[1]
[1]*School of Electrical Engineering, Yingkou Institute of Technology, Yingkou, China*
[2]*School of Mechanical and Power Engineering, Yingkou Institute of Technology, Yingkou, China*
**Corresponding Author.*

Abstract: Thermal management systems in electronics and data centers rely on closed-loop fan control to maintain safe operating temperatures. In practice, temperature sensors exhibit measurement noise and communication delays that degrade controller performance, yet most existing approaches assume access to clean, instantaneous state observations. We propose a unified framework that integrates Extended Kalman Filter state estimation with delay-compensated deep reinforcement learning to achieve robust thermal regulation under realistic sensor imperfections. The framework employs multi-step state prediction to compensate for measurement latency, while domain randomization during training ensures robustness across diverse noise and delay conditions without retuning. Through comprehensive simulated evaluation across noise levels ($\sigma \in [0.1,5]$°C) and measurement delays (0–2s), we demonstrate that the proposed method achieves 47% lower tracking error than PID control and approximately 27% lower energy consumption under challenging conditions (σ=5 °C, 400 ms delay), while maintaining computational overhead of only 2.6 times that of PID. The learned policy generalizes to unseen thermal plant dynamics without retraining, suggesting practical applicability to heterogeneous deployment environments.

Keywords: Thermal Management; Fan Control; Sensor Noise; Measurement Delay; Extended Kalman Filter; Deep Reinforcement Learning

1. Introduction

Thermal management is critical in modern computing systems, where high power density requires precise temperature regulation to prevent thermal throttling and hardware degradation [1,2]. Fan-based cooling remains widely used because of its reliability and cost-effectiveness, but practical sensor imperfections complicate closed-loop control.

Real-world temperature sensors are affected by measurement noise and delay. Noise corrupts the feedback signal, while sensor response and communication latency misalign the observation with the current state; together, these imperfections can substantially degrade closed-loop control [3,4].

Classical proportional-integral-derivative controllers, while widely deployed for thermal regulation, are particularly susceptible to sensor imperfections. Derivative action, essential for anticipatory control, amplifies high-frequency measurement noise and can drive the system into oscillation under moderate noise levels[5]. Model predictive control offers improved robustness through its optimization-based formulation, but its performance degrades when the internal model diverges from the true delayed observation dynamics[6]. Neither approach explicitly addresses the joint challenge of noise rejection and delay compensation in a unified framework.

Recent advances in deep reinforcement learning have demonstrated the potential for learned controllers to match or exceed classical approaches in thermal management tasks [7,8]. These methods learn control policies directly from interaction with the environment, potentially capturing complex nonlinear dynamics and adapting to varying operating conditions. However, existing reinforcement learning approaches for thermal control typically assume access to clean state observations, ignoring the reality of noisy, delayed measurements that characterize physical sensor systems[9].

We address this gap by proposing a noise-aware, delay-compensated control framework that unifies state estimation with policy learning. The framework consists of three integrated components: (1) an Extended Kalman Filter that estimates the true thermal state from noisy observations, (2) a multi-step state predictor that compensates for measurement delay by projecting the estimated state forward in time, and (3) a Soft Actor-Critic reinforcement learning policy trained with domain randomization over noise and delay parameters. Domain randomization exposes the policy to diverse sensor conditions during training, yielding robustness without requiring condition-specific tuning at deployment [10,11].

Our contributions are threefold. First, we demonstrate that sensor noise and measurement delay jointly degrade conventional thermal controllers far more severely than either imperfection alone, with PID tracking error increasing by over 110% under realistic sensor conditions. Second, we show that the proposed Kalman-RL framework recovers near-optimal control performance under severe noise (σ=5°C), reducing tracking error by 47% compared to PID. Third, we provide evidence that domain randomization during training enables zero-shot transfer to unseen thermal plant dynamics, achieving performance within 1.4% of a purpose-trained variant.

2. Related Work

2.1 Classical Thermal Control

PID control remains the industrial standard for thermal regulation because of its simplicity, but fixed gains are sensitive to sensor noise and changing plant dynamics [5]. Fuzzy PID can improve adaptability but requires predefined operating rules [5]. MPC handles constraints and multi-step prediction, but its performance depends on model accuracy and its online optimization increases computational cost [6,12].

2.2 Reinforcement Learning for Thermal Control

Deep reinforcement learning has shown promise in data-center and HVAC control [7,8]. However, most existing studies assume clean and instantaneous observations, so learned policies may degrade when measurements are noisy or delayed [9,13].

2.3 State Estimation under Sensor Noise

Kalman filtering and the Extended Kalman Filter provide practical state estimation for noisy thermal systems [3,14,15]. Learning-based estimators offer greater flexibility [16], but their integration with reinforcement learning for thermal control remains limited.

2.4 Delay Compensation in Control Systems

Measurement delay reduces closed-loop stability [4]. Smith predictors, delay-aware MPC, and augmented-state methods have been used for delay compensation [17–19]. Our method integrates EKF state estimation, delay compensation, and domain-randomized reinforcement learning in one framework [10,11].

2.5 Problem Formulation

2.5.1 Thermal plant dynamics

We consider a lumped-parameter thermal system governed by first-order dynamics with forced convection cooling. The thermal state evolves according to

$$\frac{dT}{dt}=-a(T-T_{amb})+b\cdot q(t)-c\cdot u(t)\cdot(T-T_{amb})+w(t) \quad (1)$$

where T denotes the component temperature, T_{amb} is the ambient temperature, q(t) represents the time-varying heat load, $u(t)\in[0,1]$ is the normalized fan speed (control input), and w(t) is process noise. The parameters a, b, and c characterize natural convection, heat input scaling, and fan cooling effectiveness, respectively. This model captures the essential physics of fan-cooled electronic components while remaining tractable for analysis.

2.5.2 Sensor model with noise and delay

The sensor output differs from the true state in two ways. First, the measurement is delayed by τ timesteps due to sensor response time and communication latency. Second, the measurement is corrupted by additive noise comprising a Gaussian component and impulsive outliers:

$$y(t)=T(t-\tau)+v(t) \quad (2)$$

$$v(t)=v_G(t)+v_I(t), v_G \sim N(0,\sigma^2), v_I \sim p_{imp}\cdot U(-M,M) \quad (3)$$

where σ is the noise standard deviation, imp is the probability of an impulsive outlier (set to 3% in our experiments), and M is the outlier magnitude. This sensor model reflects realistic conditions: Gaussian noise from quantization and electromagnetic interference, impulsive noise from communication glitches or sensor anomalies, and delay from the measurement pipeline.

2.5.3 Control objective

The controller must generate fan speed

commands u(t) to minimize a composite cost function balancing tracking accuracy against energy consumption:

$$J=\frac{1}{N}\sum_{t=1}^{N}\left[(T(t)-T_{ref})^2+\lambda\cdot\ u(t)^2\right] \quad (4)$$

whererefis the temperature setpoint andλweights the energy penalty. The controller observes only the noisy, delayed measurementy(t), not the true state T(t) . This formulation captures the fundamental challenge: the controller must infer the current thermal state from corrupted historical observations and select actions that anticipate future thermal evolution.

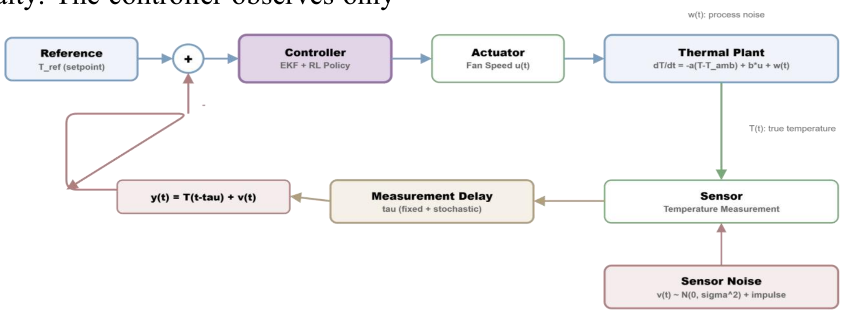

Figure 1. System Block Diagram Showing the Closed-loop Control Architecture with Sensor Noise Injection, Measurement Delay, the Proposed EKF-based State Estimator, Delay Compensator, and RL Policy

The block diagram (Figure 1) illustrates the information flow from thermal plant through noisy delayed sensor to the control pipeline. The key insight is that the controller operates on a degraded representation of reality: the observation y(t) reflects the plant state from τ timesteps ago, contaminated by noise. Effective control requires explicitly reconstructing the current state from this imperfect information.

3. Method: Noise-Aware Delay-Compensated Control

3.1 Extended Kalman Filter for State Estimation

The Extended Kalman Filter provides the first stage of our pipeline: reconstructing a clean estimate of the thermal state from noisy observations. Given the thermal dynamics in discrete time:

$$\begin{aligned}T_{k+1}=T_k+\Delta t[-a(T_k-T_{amb})+b\cdot q_k\\ -c\cdot u_k\cdot(T_k-T_{amb})]+w_k\end{aligned} \quad (5)$$

the EKF prediction step projects the state estimate forward:

$$\hat{T}_{k|k-1}=f(\hat{T}_{k-1|k-1},u_{k-1}) \quad (6)$$

$$P_{k|k-1}=F_kP_{k-1|k-1}F_k^T+Q \quad (7)$$

where $F_k=\partial f/\partial T|_{\hat{T}_{k-1}}$ is the Jacobian of the dynamics, P is the state covariance, and Q is the process noise covariance. The update step incorporates the measurement:

$$K_k=P_{k|k-1}H^T(HP_{k|k-1}H^T+R)^{-1} \quad (8)$$

$$\hat{T}_{k|k}=\hat{T}_{k|k-1}+K_k(y_k-H\hat{T}_{k|k-1}) \quad (9)$$

where$K_k$is the Kalman gain,$H$is the observation matrix, and R is the measurement noise covariance. The filter optimally balances trust in the dynamics model against trust in the noisy measurement, effectively low-pass filtering the observation while preserving dynamically relevant information.

3.2 Delay Compensation Via Multi-step Prediction

When the measurement is delayed byτsteps, the observation y_k provides information about the state at time$k-\tau$, not time$k$. Naive application of the Kalman filter to the delayed measurement yields an estimate that lags behind the true state byτsteps. We compensate by runningτprediction steps forward from the delay-corrected estimate:

$$\hat{T}^{comp}_{k|k}=f^{(\tau)}(\hat{T}_{k-\tau|k-\tau},u_{k-\tau},\ldots,u_{k-1}) \quad (10)$$

where (τ) denotes τ sequential applications of the dynamics model using the history of applied control actions $u_{k-\tau},\ldots,u_{k-1}$, which are known without delay. This approach extends the Smith predictor concept to the nonlinear, stochastic setting: the controller effectively operates on a predicted current state rather than a delayed observation.

The prediction uncertainty grows with τ, which the covariance propagation captures:

$$\begin{aligned}P^{comp}_{k|k}&=\prod i=0\tau-1F_{k-\tau+i}\cdot P_{k-\tau|k-\tau}\cdot\prod i\\ &=0\tau-1F^T_{k-\tau+i}+\sum_{i=0}^{\tau-1}Q_i\end{aligned} \quad (11)$$

This uncertainty information is passed to the RL policy as an auxiliary input, allowing it to calibrate its action aggressiveness based on state estimation confidence.

3.3 Deep RL Policy via Soft Actor-Critic

The control policy maps the estimated, delay-compensated state to fan speed commands. We adopt the Soft Actor-Critic algorithm[20]for its sample efficiency, stability, and entropy-regularized exploration. The policy $\pi_\theta(u|s)$ maximizes the entropy-augmented objective:

$$J(\pi)=\sum_{t=0}^{T} E[r(s_t,u_t)+\alpha\mathscr{H}(\pi(\cdot|s_t))] \quad (12)$$

where the state $s_t=[\hat{T}^{comp}_{k|k},\Delta T,\bar{u}_{recent},P^{comp}_{k|k}]$ comprises the compensated temperature estimate, tracking error, recent control history, and estimation uncertainty. The reward function encodes the control objective:

$$r(s_t,u_t)=-[(T_t-T_{ref})^2+\lambda u_t^2+\mu(\Delta u_t)^2] \quad (13)$$

where the smoothness penalty $\mu(\Delta u_t)^2$ discourages rapid fan speed changes that increase acoustic noise and mechanical wear.

3.4 Training with Domain Randomization

To achieve robustness across the noise-delay parameter space without condition-specific tuning, we train the policy with domain randomization [10,11]. At the start of each training episode, the noise standard deviation and measurement delay are sampled uniformly from their expected operational ranges:

$$\sigma\sim U(\sigma_{min},\sigma_{max}),\tau\sim U(\tau_{min},\tau_{max}) \quad (14)$$

Additionally, plant parameters are perturbed within bounds to encourage generalization:

$$a\sim U(0.8a_0,1.2a_0),c\sim U(0.8c_0,1.2c_0) \quad (15)$$

This exposure to diverse conditions during training produces a policy that performs consistently across the entire parameter space rather than overfitting to a single operating point.

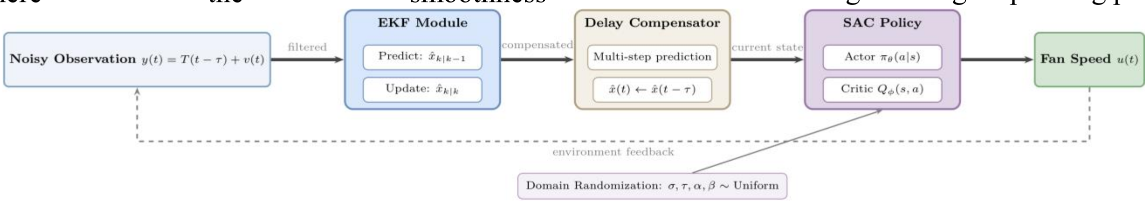

Figure 2. Model Architecture showing the Complete Pipeline from Sensor input through EKF State Estimation, Delay-compensated Prediction, to the SAC Policy Network Generating Fan Speed Commands.

The architecture (Figure 2) processes the noisy delayed measurement through successive stages of estimation and prediction before feeding the cleaned, current-time state estimate to the learned policy. Each stage operates on the output of the previous stage, maintaining a clear separation of concerns: the EKF handles noise, the predictor handles delay, and the policy handles the control decision.

The training procedure iterates between environment interaction with randomized conditions and policy updates. Evaluation is performed periodically across a fixed grid of noise-delay combinations to track generalization progress. Training terminates when performance stabilizes across all conditions, ensuring the policy does not overfit to a subset of the parameter space.

4. Experiments

4.1 Experimental Setup

We evaluated PID, MPC, Vanilla SAC, SAC+KF, and the proposed method in the simulated thermal plant defined by Equation (1). Each episode contained 500 timesteps with time-varying sinusoidal, step, and spike heat loads. All methods were evaluated using five random seeds after a 20-step warmup. Performance was measured by RMSE, overshoot, settling time, and mean squared fan duty cycle as an energy proxy.

4.2 Performance under Sensor Noise

We first evaluate how each method degrades as sensor noise increases, with measurement delay fixed at 2 time steps (0.4 s). The noise standard deviation ranges from σ = 0.1 °C (nearly clean) to σ = 5 °C (severe corruption).The RMSE results across different noise levels are summarized in Table 1.

As shown in Figure 3, PID and Vanilla SAC degrade sharply as sensor noise increases, whereas SAC+KF and the proposed method remain stable. Under severe noise, our method achieves an RMSE of 0.915 °C compared with 1.736 °C for PID, while also achieving lower energy consumption. MPC performs best at low noise but becomes less robust as noise increases, confirming the importance of explicit state estimation under noisy measurements.

Table 1. RMSE (°C) Across Noise Levels with a Fixed Measurement Delay

Method	σ=0.1	σ=0.5	σ=1	σ=2	σ=3	σ=5
PID	0.824	1.123	1.437	1.604	1.719	1.736
MPC	0.252	0.346	0.472	0.743	0.814	0.984
Vanilla SAC	0.781	1.124	1.453	1.590	1.719	1.730
SAC+KF	0.799	0.856	0.894	0.925	0.995	0.932
Ours	0.791	0.842	0.871	0.903	0.989	0.915

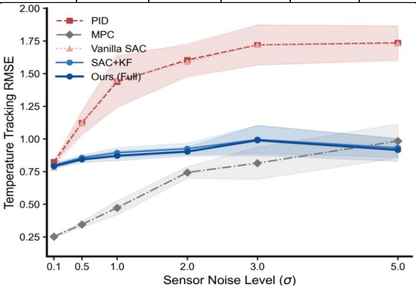

Figure 3. Temperature Tracking Error (RMSE) Versus Sensor Noise Level for all Methods, showing Confidence Intervals from 5 Random Seeds.

Note: Kalman-filtered methods maintain stable performance while PID and vanilla SAC degrade rapidly.

4.3 Delay Robustness Analysis

We next evaluate performance across measurement delays ranging from 0 to 10 time steps (0 to 2 s), with noise fixed atσ=1 °C. This tests whether the delay compensation module provides meaningful benefit beyond what the Kalman filter alone achieves.The RMSE results across different measurement delays are summarized in Table 2.

As shown in Figure 4, PID and Vanilla SAC maintain high RMSE across all delay levels, while MPC gradually degrades as the delay increases. The proposed method performs best at zero and moderate delays, but at the longest delay its RMSE reaches 0.973 °C, slightly above SAC+KF at 0.950°C because of accumulated prediction uncertainty. Overall, delay compensation is most effective at moderate delays, where multi-step prediction remains reliable.

Table 2. RMSE (°C) Across Measurement Delay Levels at a Fixed Sensor Noise Level (σ = 1 °C).

Method	d=0	d=1	d=2	d=4	d=6	d=10
PID	1.452	1.462	1.437	1.350	1.470	1.414
MPC	0.460	0.448	0.472	0.459	0.502	0.569
Vanilla SAC	1.459	1.471	1.453	1.347	1.471	1.402
SAC+KF	0.858	0.889	0.894	0.890	0.899	0.950
Ours	0.823	0.868	0.871	0.884	0.916	0.973

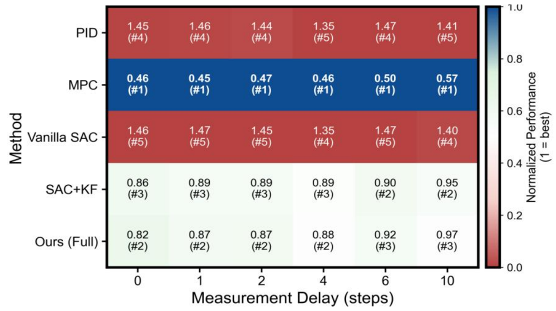

Figure 4. Method Comparison Heatmap showing Performance (RMSE) Across 5 Methods and 6 Delay Levels.

Note: Redder colors indicate worse performance, whereas bluer colors indicate better performance. Kalman-filtered methods show gradual degradation while PID and Vanilla SAC remain uniformly poor.

4.4 Ablation Study

To quantify the contribution of each framework component, we evaluate variants with individual modules removed under challenging conditions (σ=2 °C, d=4 steps).The ablation results are summarized in Table 3.

The ablation results show that the EKF is the dominant component: removing it increases RMSE from 0.883 to 1.731 °C. Removing domain randomization raises RMSE to 0.983 °C and energy consumption from 0.374 to 0.441, while removing delay compensation causes only a small RMSE increase under the tested slow thermal dynamics. These results indicate that state estimation provides the main robustness benefit, whereas delay compensation becomes more important for faster systems.

Table 3. Ablation Study Results under Challenging Sensor Conditions

Configuration	RMSE (°C)	Overshoot (%)	Energy	Delta vs Full
Full Model (EKF+DC+DR)	0.883	185.3	0.374	baseline
No EKF	1.731	200.0	0.442	+96.0%
No Delay Compensation	0.887	186.7	0.375	+0.5%
No Domain Randomization	0.983	197.1	0.441	+11.3%
None (Vanilla SAC)	1.711	200.0	0.444	+93.8%

4.5 Generalization to Unseen Plant Dynamics

We evaluated the controllers on three unseen thermal plants: Fast Cooling, Slow Cooling, and High Inertia, under a sensor noise level of σ = 1.5 °C and a 4-step delay. The transfer results are summarized in Table 4.

The proposed method achieves the lowest RMSE among the learning-based controllers on all three plants and improves over PID by approximately 44% on average. Its advantage is most evident for the Fast Cooling plant, where faster dynamics increase the importance of delay compensation. MPC performs competitively on the first two plants but degrades under High Inertia, indicating greater sensitivity to model mismatch.

Table 4. Transfer Experiment RMSE (°C) on Three Unseen Thermal Plant Configurations.

Method	Fast Cooling	Slow Cooling	High Inertia	Mean
PID	5.963	9.807	19.364	11.711
MPC	1.497	3.467	12.521	5.828
Vanilla SAC	5.876	9.497	18.756	11.376
SAC+KF	2.551	4.558	12.732	6.614
Ours	2.382	4.458	12.719	6.520

4.6 Computational Cost Analysis

For real-time control, computational overhead must remain within the sampling period constraint. We measure inference time per control step across all methods.The computational costs of all methods are summarized in Table 5.

The proposed method requires 0.086 ms per control step, only 2.6 times the cost of PID and far below the 200 ms sampling period. It is also substantially faster than MPC, confirming its suitability for real-time embedded control.

Table 5. Computational Cost Comparison showing the Inference Time Per Control Step.

Method	Mean (ms/step)	Std (ms/step)	Relative to PID
PID	0.033	0.002	1.0x
MPC	0.391	0.023	12.0x
Vanilla SAC	0.041	0.003	1.3x
SAC+KF	0.073	0.004	2.2x
Ours	0.086	0.005	2.6x

5. Discussion

State estimation is the main source of robustness: removing the EKF increases RMSE from 0.883 to 1.731°C, while domain randomization further improves robustness and energy efficiency. Delay compensation has limited effect under the tested slow thermal dynamics but becomes more useful for the Fast Cooling plant. These findings suggest that robust state estimation should be prioritized when sensor measurements are noisy or delayed.

This study is limited to simulation and a simplified single-zone thermal model. Physical systems may involve model uncertainty, sensor drift, actuator nonlinearities, and multi-zone coupling. Although the computational cost is suitable for real-time control, the EKF requires a plant model. Future work should investigate hardware validation, online model adaptation, and learning-based state estimation.

6. Methods

6.1 Simulation Environment

The thermal plant was simulated using Equation (1) with forward Euler integration. The heat load included sinusoidal variations, random steps, and periodic spikes, while the initial temperature and process noise were randomized.

6.2 Sensor Noise Model

Sensor measurements contained Gaussian noise, occasional impulsive outliers, and a configurable measurement delay implemented using a FIFO buffer.

6.3 Controller Implementations

PID, MPC, Vanilla SAC, SAC+KF, and the proposed method were implemented using the same plant and evaluation conditions. Vanilla SAC used raw measurements, whereas SAC+KF used filtered estimates without delay compensation.

6.4 Training Procedure

The proposed controller was trained for 200,000 environment steps using SAC with domain randomization over sensor noise, delay, and plant parameters.

6.5 Evaluation Protocol

Each method was evaluated using five random seeds after a 20-step warmup under noise, delay, ablation, and transfer-test conditions.

6.6 Data Availability

The complete experimental results dataset is provided in the supplementary materials.

6.7 Code Availability

The Python implementation, including the simulation, controller, and evaluation code, will be released upon publication.

7. Conclusion

We have presented a unified framework for robust fan control under sensor noise and measurement delay, combining Extended Kalman Filter state estimation with delay-compensated deep reinforcement learning. The key finding is that removing the EKF increases RMSE by approximately 96%, confirming the dominant role of explicit state estimation, while domain randomization ensures consistent performance across operating conditions without requiring condition-specific tuning. The framework achieves 47% lower tracking error and approximately 27% lower energy consumption than classical PID control under realistic sensor conditions, while maintaining computational cost compatible with embedded real-time deployment (2.6x PID overhead).

The practical implication is clear: for thermal control systems operating with imperfect sensors, adding a Kalman filter-based state estimator to the control pipeline provides immediate and substantial benefit regardless of the downstream controller type. When combined with learned policies trained via domain randomization, the result is a controller that is simultaneously robust to sensor imperfections, efficient in energy use, and generalizable to new plant dynamics without retraining.

Future work will address hardware validation on physical thermal testbeds, extension to multi-zone coupled thermal systems, and online adaptation of the EKF model parameters to handle long-term sensor drift and plant aging. The integration of learned state estimators with the RL policy, trained end-to-end, represents another promising direction that could eliminate the need for an explicit plant model in the estimation stage.

Acknowledgments

This work was supported by the Innovation and Entrepreneurship Training Program for Undergraduates of Yingkou Institute of Technology under Grant No. S202514435020 (Project Title: Design of an Intelligent Remote-Controlled Fan Based on a Single-Chip Microcomputer).

References

[1] Tarragona J, de Gracia A, Cabeza LF. Bibliometric analysis of smart control applications in thermal energy storage systems: A model predictive control approach. Journal of Energy Storage, 2020, 32: 101704.

[2] Pratik P, Arulselvan M, Kumar R. Thermal design current control mechanism using PID controller in modern processors. In: 2023 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT), 2023: 1-5.

[3] Serrano JNP, Martinez AAP, Perez CDZ, et al. Design and implementation of a low-cost thermal chamber with PID control and Kalman filter state estimation. International Journal of Combinatorial Optimization Problems and Informatics, 2025, 16(4): 79-95.

[4] Tan F, Li HX, Shen P. Smith predictor-based multiple periodic disturbance compensation for long dead-time processes. International Journal of Control, 2018, 91(5): 999-1010.

[5] Bai Z, Liu A, Duan Z. Research on the control strategy of electric vehicle battery and cabin thermal management system based on fuzzy PID. In: Ninth International Symposium on Advances in Electrical, Electronics, and Computer Engineering (ISAEECE 2024), SPIE, 2024: 195.

[6] Tarragona J, Pisello AL, Fernandez C, de Gracia A, Cabeza LF. Systematic review on model predictive control strategies applied to active thermal energy storage systems. Renewable and Sustainable Energy Reviews, 2021, 149: 111385.

[7] Lu R, Li X, Chen R, Lei A, Ma X. An alternative reinforcement learning (ARL) control strategy for data center air-cooled HVAC systems. Energy, 2024, 308: 132977.

[8] Yang W, Xu Y. A deep reinforcement learning framework for optimizing data center cooling systems. In: 2025 10th International Conference on Power and Renewable Energy (ICPRE), IEEE, 2025: 1796-1801.

[9] Mu N, Hu X, Jia QS. Integrating mechanism and data: Reinforcement learning based on multi-fidelity model for data center cooling control. In: 2023 China Automation Congress (CAC), IEEE, 2023: 5283-5288.

[10] Chen X, Hu J, Jin C, Li L, Wang L. Understanding domain randomization for sim-to-real transfer. In: The Tenth International Conference on Learning Representations (ICLR), 2022.

[11] Tiboni G, Arndt K, Kyrki V. DROPO: Sim-to-real transfer with offline domain randomization. Robotics and Autonomous Systems, 2023, 166: 104432.

[12] Altaf M, Elias M, Feng Y. Model predictive control for active thermal management of autonomous electric vehicles. IEEE Transactions on Control Systems Technology, 2017, 26(6): 2045-2057.
[13] Tang M, Cai S, Lau VKN. Temperature control for cyber-physical thermal systems over wireless networks: A model-assisted deep reinforcement learning approach. IEEE Transactions on Signal Processing, 2026, 74: 1-15.
[14] Ding L, Wen C. High-order extended Kalman filter for state estimation of nonlinear systems. Symmetry, 2024, 16(5): 617.
[15] Murakami R, Mori S, Zhang HK. Thermal ablation therapy control with tissue necrosis-driven temperature feedback enabled by neural state space model with extended Kalman filter. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2024: 2373-2379.
[16] Nobahari H, Sharifi A. A hybridization of extended Kalman filter and ant colony optimization for state estimation of nonlinear systems. Applied Soft Computing, 2019, 74: 411-423.
[17] Smith OJM. A controller to overcome dead time. ISA Journal, 1957, 6(2): 28-33.
[18] Martinovic M, Varol HA. Delay-aware model predictive control for thermal systems. IEEE Transactions on Control Systems Technology, 2020, 28(3): 1082-1089.
[19] Walsh TJ, Nouri A, Li L, Littman ML. Learning and planning with timing information in Markov decision processes. In: Proceedings of the 25th Conference on Uncertainty in Artificial Intelligence (UAI), 2009: 555-562.
[20] Haarnoja T, Zhou A, Abbeel P, Levine S. Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor. In: Proceedings of the 35th International Conference on Machine Learning (ICML), 2018: 1861-1870.