sustainability

MDPI

Article

Sustainable Reduced-Order Thermal Modeling for Energy-Efficient Real-Time Control of Grid-Scale Energy Storage Systems

Mohammad Fazle Rabbi

Coordination and Research Centre for Social Sciences, Faculty of Economics and Business, University of Debrecen, Böszörményi út 138, 4032 Debrecen, Hungary; drrabbikhan@gmail.com or rabbi.mohammad@econ.unideb.hu

Abstract

Grid-scale lithium-ion storage must deliver fast, reliable thermal control during dynamic grid services, yet high-fidelity thermal models are too slow for real-time use and inefficient cooling inflates energy and safety costs. This study develops and validates a reduced-order thermal modeling framework for grid-scale lithium-ion battery energy storage, targeting real-time thermal management. The framework uses proper orthogonal decomposition to capture dominant thermal dynamics across frequency regulation, peak shaving, and fast charging. Across scenarios, it delivers 15.2–22.3× computational speedups versus a detailed model while maintaining RMS temperature errors of 7.8 °C (frequency regulation), 34.4 °C (peak shaving), and 23.3 °C (fast charging). Spatial analysis identifies inter-zone temperature gradients up to 1.0 °C under severe loading, motivating targeted cooling strategies. Cooling energy scales nonlinearly with load intensity, from 5.44 kWh in frequency regulation to over 300 kWh in peak shaving, with cooling efficiencies spanning 17.27% to 8.94%. The reduced-order model achieves sub-0.1 s computational solve time per control cycle, suggesting feasibility for real-time integration into industrial battery-management systems under the tested simulation settings. Collectively, the results show that reduced-order thermal models can balance accuracy and computational efficiency for several grid services in the simulated scenarios, while high-power operation benefits from scenario-specific calibration and controller tuning. Practically, the benchmarks and workflow support decisions on predictive cooling schedules, temperature limits, and service prioritization to minimize parasitic energy.

Keywords: lithium-ion battery; thermal modeling; reduced-order modeling; grid-scale energy storage; thermal management; thermal simulation

check for updates

Academic Editors: Zhi Li and Yonghong Xu

Received: 2 October 2025
Revised: 27 October 2025
Accepted: 3 November 2025
Published: 4 November 2025

Citation: Rabbi, M.F. Sustainable Reduced-Order Thermal Modeling for Energy-Efficient Real-Time Control of Grid-Scale Energy Storage Systems. *Sustainability* **2025**, *17*, 9839. https://doi.org/10.3390/su17219839

1. Introduction

Energy storage systems are pivotal in enabling the integration of renewable energy sources into modern power grids by balancing supply-demand mismatches and providing ancillary grid services [1,2]. Among various technologies, lithium-ion batteries have emerged as a leading solution due to their high energy density, long cycle life, and relatively mature supply chains [3]. However, their operational efficiency, safety, and longevity are significantly influenced by thermal conditions, which fluctuate dynamically based on load profiles and environmental factors [4,5]. Effective thermal management is therefore critical to maximizing performance and preventing premature degradation [6]. Recent advances in artificial intelligence and machine learning have shown promise for enhancing renewable energy integration and forecasting [7], though their application to battery thermal management remains underexplored.

Recent advances in battery thermal modeling have focused on detailed three-dimensional simulations capable of capturing the complex heat transfer processes within battery packs [8,9]. Such models, while accurate, impose substantial computational burdens that limit their real-time implementation, especially for large-scale grid applications requiring rapid thermal predictions for control purposes [10,11]. To address this challenge, reduced-order modeling techniques, particularly those based on proper orthogonal decomposition, have garnered attention for enabling computationally efficient and sufficiently accurate thermal simulations [12,13]. These approaches extract dominant thermal modes from high-fidelity data and reproduce temperature evolution with substantially fewer degrees of freedom [14,15].

Despite these developments, significant research gaps persist regarding the practical applicability of reduced-order thermal models to grid-scale storage systems operating under diverse real-world conditions [16,17]. Prior studies have predominantly focused on laboratory-scale cells or idealized operating scenarios [18–20], leaving open questions about the models' ability to maintain fidelity across extended service periods and heterogeneous operational regimes such as frequency regulation, peak shaving, and fast charging [21–23]. Furthermore, the relationship between model accuracy, computational efficiency, and sustainability metrics such as cooling energy consumption remains poorly quantified, hindering holistic assessments critical for field deployment [24,25].

Controversies also exist around the trade-offs inherent in model simplification, with some researchers emphasizing the need for detailed physics-based models to capture nonlinear degradation mechanisms [26,27], while others advocate for reduced complexity to enable real-time control and adaptability [28]. This dichotomy motivates investigation into adaptive modeling frameworks that can reconcile accuracy and efficiency requirements across application contexts [29]. However, existing POD-based battery thermal models predominantly focus on single operational scenarios or laboratory-scale cells, leaving unclear whether such approaches maintain fidelity across diverse grid-service applications. Furthermore, prior reduced-order implementations typically validate against high-fidelity simulations alone, without integrating spatial thermal gradients, cooling energy metrics, and real-time control feasibility into a unified sustainability assessment framework applicable to utility-scale deployments. While hybrid artificial intelligence approaches have demonstrated superior performance for renewable energy forecasting and grid integration [7], their integration with reduced-order thermal models for grid-scale storage systems remains an open research frontier.

The present study addresses these gaps by implementing and validating a reduced-order thermal modeling framework for grid-scale lithium-ion battery energy storage systems across multiple representative grid service scenarios. Key objectives include quantifying the trade-offs between modeling accuracy and computational speedup, evaluating spatial thermal gradients within multi-zone battery configurations, and assessing the implications on cooling energy efficiency and real-time control feasibility. Comprehensive validation against detailed simulation references provides benchmarks for predictive fidelity, while performance metrics elucidate sustainability outcomes.

The core contribution of this work is the development and validation of a scenario-based quantitative benchmarking framework for reduced-order thermal modeling of grid-scale storage systems. Unlike prior studies that evaluated POD-based models under single operational conditions [30], this work systematically quantifies modeling accuracy, computational efficiency, spatial thermal gradients, cooling energy requirements, and real-time control feasibility across three distinct grid services: frequency regulation, peak shaving, and fast charging. This multi-scenario benchmarking approach enables direct comparison of thermal management challenges across operational regimes, providing decision-support metrics for grid operators and battery management system designers to optimize service portfolios while balancing thermal safety, sustainability, and computational tractability.

2. Materials and Methods

2.1. System Configuration and Specifications

The thermal modeling framework is based on a representative grid-scale battery energy storage system with specifications detailed in Table 1.

Table 1. System specifications and operational parameters for grid-scale energy storage thermal modeling.

Parameter	Value	Unit
Battery capacity	100 kWh	kWh
Nominal voltage	400 V	V
Operating temp range	15–45 °C	°C
Cooling system type	Liquid cooling	-
Thermal zones	5 zones	-
Computational time step	0.1 s	s
Simulation duration	24 h	h
Grid frequency regulation	±2% frequency	Hz
Peak shaving power	50 kW peak	kW
Fast charge rate	2 C rate	C

Notes: All parameters are selected for consistency with representative stationary storage installations. Parameters populate the detailed and reduced models in Equations (1)–(4) and the scenario drivers in Equations (5) and (6). Five thermal zones are modeled for spatial accuracy in temperature distribution. The time step balances numerical stability with computational efficiency.

Table 1 establishes the foundational technical specifications and operational parameters for the grid-scale battery energy storage thermal modeling study. The system configuration represents a typical utility-scale installation with a 100 kWh lithium-ion battery capacity operating at 400 V nominal voltage, selected to align with commercial stationary storage systems currently deployed in grid applications. The operating temperature range of 15–45 °C reflects industry-standard thermal windows that balance performance optimization with safety requirements and battery longevity considerations.

The thermal management architecture employs liquid cooling technology distributed across five discrete thermal zones, enabling precise spatial temperature control and detailed analysis of thermal gradients throughout the energy storage system. This five-zone discretization strategy corresponds to inlet (zone 1), three intermediate (zones 2–4), and outlet (zone 5) regions, with each zone representing approximately 20% of the total battery capacity. The zoning boundaries are determined by cooling fluid flow patterns, heat generation characteristics, and thermal coupling between adjacent battery modules.

The computational framework utilizes a 0.1 s time step to ensure numerical stability while maintaining computational efficiency suitable for real-time control applications. This temporal resolution captures rapid thermal transients during high-power operations while balancing accuracy with computational tractability for practical implementation in grid-scale systems. The 24 h simulation duration encompasses complete diurnal cycles representative of typical grid service patterns, including peak demand periods and overnight charging scenarios.

The operational scenario definitions encompass three primary grid service applications that represent the most common revenue streams for grid-scale storage systems.

Frequency regulation involves continuous small-amplitude power modulations (±2% of nominal frequency) to maintain grid stability, characterized by bidirectional power flow with moderate thermal impacts. Peak shaving operations target demand charge reduction through sustained high-power discharge up to 50 kW during peak consumption periods, typically 2–4 h during afternoon and evening peak demand. Fast charging at 2 C rate represents the most thermally demanding scenario, involving rapid energy input at twice the battery capacity per hour, generating significant heat loads that challenge the thermal management system.

This comprehensive parameter set provides the technical foundation for evaluating thermal behavior across the full operational envelope of grid-scale energy storage systems, enabling systematic assessment of reduced-order modeling accuracy and computational efficiency under realistic operating conditions.

2.2. Detailed Thermal Model Development

The detailed thermal model is formulated using three-dimensional heat transfer principles governing thermal behavior in battery energy storage systems. The model accounts for all major heat transfer mechanisms including conduction, convection, and radiation, providing a detailed representation of thermal dynamics within the battery pack.

The fundamental energy balance equation describes the temporal and spatial temperature evolution within each thermal zone, representing conservation of thermal energy in differential form:

$$\rho c_p \frac{\partial T}{\partial t} = k\nabla^2 T + Q_{\text{gen}} - Q_{\text{conv}} - Q_{\text{rad}} \tag{1}$$

This equation represents the rate of thermal energy accumulation (left side) balanced by net heat transfer and generation (right side). The material density ρ (kg/m^3) and specific heat capacity $c_p\left(\frac{J}{kg}\cdot K\right)$ characterize the thermal inertia of the battery materials, determining how rapidly temperature changes in response to heat inputs. The thermal conductivity k (W/m·K) governs heat conduction through the battery structure, while the Laplacian operator $\nabla^2 T$ captures three-dimensional heat diffusion, accounting for thermal gradients in all spatial directions. The source term Q_{gen} represents internal heat generation from electrochemical processes, while Q_{conv} and Q_{rad} account for convective heat transfer to the cooling system and radiative heat losses to the environment, respectively.

The internal heat generation term captures the primary sources of thermal energy within battery cells, combining resistive losses and electrochemical inefficiencies:

$$Q_{\text{gen}} = I^2 R_{\text{int}} + I \times \eta_{\text{overpotential}} \times V_{\text{terminal}} \tag{2}$$

The first term $I^2 R_{\text{int}}$ represents Joule heating due to electrical resistance, where I is the instantaneous current (A) and R_{int} is the internal resistance (Ω). This quadratic relationship reflects the fundamental physics of resistive heating, becoming particularly significant during high-current operations such as fast charging. The second term captures electrochemical overpotential losses, where $\eta_{overpotential}$ represents the efficiency factor accounting for activation, concentration, and ohmic overpotentials, and Vterminal is the terminal voltage (V). This formulation captures the primary contributors to heat generation under the modeled operating conditions. This contributes to thermal load during different operational scenarios, enabling accurate prediction of temperature rise under varying electrical demands.

2.3. Reduced-Order Thermal Model Formulation

The reduced-order model employs proper orthogonal decomposition (POD) to extract dominant thermal modes from the detailed simulation data, providing a computationally efficient approximation while maintaining essential thermal behavior. This approach leverages the observation that thermal fields in battery systems often exhibit lower-dimensional structure, allowing accurate representation using fewer degrees of freedom.

The temperature field in each zone is approximated using a truncated series expansion based on the most energetically significant spatial modes:

$$T(x,y,z,t) = \sum_{i=1}^{N} a_i(t)\phi_i(x,y,z) \tag{3}$$

The spatial basis functions $\phi_i(x,y,z)$ represent the most energetically significant thermal patterns extracted through singular value decomposition of temperature snapshots from detailed simulations. These functions capture the dominant spatial structures of thermal fields, ordered by their energy content (eigenvalues). The time-dependent coefficients $a_i(t)$ are determined through Galerkin projection, representing the temporal evolution of each spatial mode. The reduced model retains N modes selected based on cumulative energy content, typically capturing 99% of the thermal energy with significantly fewer degrees of freedom than the detailed model. This dimension reduction enables real-time computation while preserving the essential physics of thermal behavior across different operational scenarios.

The governing equation for the reduced-order model transforms the partial differential equation into a system of ordinary differential equations for the time-dependent coefficients:

$$\frac{da_i}{dt} = \sum_{j=1}^{N} A_{ij}a_j + B_i u + C_i \tag{4}$$

The linear dynamics matrix A_{ij} captures the coupling between different thermal modes, representing the linearized effects of heat conduction and thermal dissipation. The input influence matrix B_i characterizes how control inputs u (such as cooling power modulation) affects each thermal mode, enabling systematic control design. The term C_i accounts for nonlinear effects and time-invariant source terms, including baseline heat generation and ambient heat losses. This formulation enables real-time thermal prediction with computational speedups exceeding 15× compared to the detailed model, while maintaining sufficient accuracy for control applications as demonstrated in the validation studies.

2.4. Operational Scenario Definition

Three distinct operational scenarios are implemented to evaluate thermal behavior under representative grid service conditions, covering the primary applications of grid-scale energy storage systems. Each scenario presents unique thermal challenges due to different power profiles, duration characteristics, and load intensities.

Frequency regulation involves continuous small-amplitude power modulations (±2% of nominal frequency) to maintain grid stability, representing the most common grid service application. The load profile for frequency regulation exhibits stochastic characteristics with bounded amplitude:

$$P_{\text{freq}}(t) = P_{\text{base}}\left(1 + A_{\text{freq}}\sin\left(2\pi f_{\text{grid}}t + \phi\right)\right) \tag{5}$$

The baseline power P_{base} establishes the mean operating point, while the amplitude factor A_{freq} (typically 0.1–0.3) determines the magnitude of power fluctuations relative to the baseline. The grid frequency f_{grid} (50 Hz) sets the fundamental frequency of regulation signals, though actual grid regulation involves a broader frequency spectrum. The random phase shift ϕ accounts for the stochastic nature of grid disturbances and regulation demands. This equation generates the continuous power fluctuations, maintaining the system within moderate thermal ranges due to the limited amplitude and bidirectional nature of the power profile.

Peak shaving operations target demand charge reduction through high-power discharge during peak consumption periods, typically occurring during afternoon and evening hours when electricity demand and prices are highest. The peak shaving power profile follows a step-function pattern:

$$P_{\text{peak}}(t) = \begin{cases} P_{max}, & t \in [t_{start}, t_{end}] \\ 0, & \text{otherwise} \end{cases} \tag{6}$$

This step-function profile generates sustained high-power periods reaching 50 kW, corresponding to the maximum discharge capability of the system. The time interval $[t_{\text{start}}, t_{\text{end}}]$ represents the peak demand period, typically 2–4 h during afternoon peak consumption. This operational pattern results in elevated thermal loads and increased cooling requirements due to the sustained high-power discharge, as demonstrated in the performance data. The thermal challenge arises from both the high-power level and the extended duration, leading to significant heat accumulation within the battery pack.

Fast charging represents the most thermally demanding scenario, involving rapid energy input at rates up to 2 C (twice the battery capacity per hour). The fast charge profile exhibits intermittent high-power pulses with varying duty cycles depending on charging protocols and grid availability. The mathematical representation follows a piecewise step function with stochastic timing:

$$P_{\text{fast}}(t) = \begin{cases} P_{max} \cdot \eta_{charge}, & t \in [t_{start,i}, t_{end,i}] \\ 0, & \text{otherwise} \end{cases} \tag{7}$$

where P_{max} represents the maximum charging power (2 C rate = 200 kW for the 100 kWh system), η_{charge} is the charging efficiency factor (typically 0.85–0.95), and $[t_{\text{start},i}, t_{\text{end},i}]$ denote the start and end times of the i-th charging event. Charging event timing is determined by grid availability windows and state-of-charge constraints, with typical event durations of 15–30 min followed by rest periods of 1–4 h. The intermittent nature generates sharp thermal transients with heating rates exceeding 2 °C/min during charge initiation, followed by gradual cooling during rest periods, creating the characteristic sawtooth temperature profile.

2.5. Model Validation and Error Quantification

Model accuracy is assessed through comprehensive comparison with high-fidelity reference simulation data using multiple error metrics that capture different aspects of prediction quality. The present validation employs detailed finite-element simulations as reference standards, consistent with established practices for initial reduced-order model development. Previous POD-based battery thermal models validated experimentally have reported 2–4% voltage prediction errors and temperature deviations within 2 °C for laboratory cells [31,32], providing accuracy benchmarks against which simulation-based validation can be contextualized. The validation approach encompasses systematic error quantification across multiple operational scenarios to establish model fidelity under diverse thermal conditions.

The root mean square (RMS) error provides a measure of overall prediction accuracy, emphasizing larger deviations through the quadratic formulation:

$$\text{RMS Error} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(T_{\text{predicted},i} - T_{\text{reference},i}\right)^2} \tag{8}$$

The RMS error quantifies the standard deviation between predicted and reference temperatures across all spatial locations and time points, providing a global measure of model fidelity. The summation includes n data points representing all spatial zones and temporal instances, ensuring comprehensive coverage of the operational envelope. Lower RMS values indicate superior model fidelity, with frequency regulation achieving 7.8 °C compared to 34.4 °C for peak shaving operations, demonstrating the increased modeling challenge under high-power conditions.

Additionally, the mean absolute error captures systematic bias in temperature predictions, providing complementary information about model performance:

$$\text{Mean Error} = \frac{1}{n}\sum_{i=1}^{n}\left|T_{\text{predicted},i} - T_{\text{reference},i}\right| \tag{9}$$

This metric provides insight into consistent over- or under-prediction tendencies, enabling targeted model calibration for specific operational scenarios. Unlike the RMS error, the mean absolute error treats all deviations equally, providing information about average prediction bias without emphasizing extreme errors. The combination of RMS and mean absolute errors provides comprehensive characterization of model accuracy across different thermal conditions and operational modes. Distributional diagnostics of absolute temperature error (histograms and ECDFs) follow the error definitions in Equations (8) and (9) (see Appendix A, Figure A1).

2.6. Performance Metrics and Sustainability Assessment

Performance evaluation encompasses both energetic and environmental metrics that quantify the sustainability benefits of reduced-order modeling approaches. These metrics provide direct measures of operational efficiency and environmental impact, supporting the development of sustainable energy storage systems.

Energy efficiency quantifies the ratio of useful energy output to total energy input, accounting for thermal losses that reduce overall system performance:

$$\eta_{\text{energy}} = \frac{E_{\text{electrical}}}{E_{electrical_{in}} + E_{thermal_{loss}}} \times 100\% \tag{10}$$

This efficiency metric captures the impact of thermal management on overall system performance, where $E_{electrical_{in}}$ represents the useful electrical energy delivered to or from the grid, and $E_{thermal_{loss}}$ represents energy lost as heat through thermal dissipation. Higher efficiency values indicate better energy utilization and reduced environmental impact. The reduced-order model consistently achieves higher efficiency values (89.3–91.4%) compared to the detailed model (87.0%) across all operational scenarios, demonstrating the potential for improved performance through advanced thermal management.

Cooling efficiency evaluates thermal-management effectiveness as the ratio of total thermal energy removed from the battery to the total electrical energy consumed by the cooling system over a defined scenario window $[t_0, t_1]$, ensuring numerator and denominator are both energies (e.g., Wh):

$$\eta_{\text{cooling}} = \frac{\int_{t_0}^{t_1} \dot{Q}_{\text{removed}}(t)dt}{\int_{t_0}^{t_1} P_{\text{cooling}}(t)dt} \times 100\% \tag{11}$$

The numerator integrates the heat-extraction rate $\dot{Q}_{\text{removed}}(t)$ (W) to yield total thermal energy removed (Wh), while the denominator integrates the cooling electrical power $P_{\text{cooling}}(t)$ (W) to yield total cooling energy input (Wh), preventing rate–energy mixing and matching scenario-level summaries. Higher η_{cooling} denotes more efficient use of electrical energy for cooling, and scenario-dependent variability reflects differences in load intensity and duration across frequency regulation, peak shaving, and fast charging.

2.7. Real-Time Control Implementation

The reduced-order model enables real-time thermal control through model predictive control (MPC) algorithms that optimize future control actions based on predicted thermal behavior, providing systematic thermal regulation while accounting for operational constraints and efficiency objectives. The control objective balances temperature-tracking accuracy against actuator effort and smoothing over finite prediction and control horizons:

$$J = \sum_{k=1}^{N_p} \|T_{k|t} - T_{\text{ref}}\|_Q^2 + \sum_{k=0}^{N_c-1} \|\Delta u_{k|t}\|_R^2 + \|T_{N_p|t} - T_{\text{ref}}\|_P^2 \tag{12}$$

where $\|x\|_Q^2 \equiv x^\top Qx$, $\Delta u_{k|t} \equiv u_{k|t} - u_{k-1|t}$; and the weighting matrices satisfy $Q \geq 0$, $R > 0$, with the terminal weight $P \geq 0$ optional. The prediction horizon N_p determines how far into the future the controller considers thermal evolution (typically 10–20 time steps for battery applications), and the control horizon N_c specifies the number of future control moves optimized simultaneously, usually shorter than N_p to reduce computational burden. The predicted temperatures $T[k+j|k]$ represent the thermal state stacked across zones at future step $k+j$ given information at time k, while T_{ref} contains reference temperatures for all thermal zones. The control inputs $u[k+j|k]$ represent physical actuator commands including: (1) cooling fluid mass flow rate (kg/s) distributed across the five thermal zones, (2) inlet coolant temperature setpoint (°C), and (3) binary on/off commands for auxiliary cooling fans (0/1). These actuator variables directly modulate heat extraction rates $\dot{Q}_{cool}$ according to the convective heat transfer relationship $\dot{Q}_{cool} = \dot{m}c_p\left(T_{fluid,\ out} - T_{fluid,in}\right)$, where $\dot{m}$ is the controllable mass flow rate, c_p is the fluid specific heat capacity, and temperature differences are influenced by flow distribution and inlet setpoints. Penalizing $\Delta u[k+j|k]$ discourages abrupt flow rate and temperature setpoint changes that stress pumps and valves, improving actuator longevity and robustness, whereas penalizing absolute $u[k+j|k]$ minimizes total cooling power consumption when energy efficiency is the primary objective. The weighting matrices Q and R balance temperature regulation accuracy against control effort, and the terminal weight P (if used) improves end-of-horizon behavior and closed-loop performance.

The control system maintains response times below 0.1 s across all operational scenarios while achieving control accuracy ranging from 72.4% to 84.4%, demonstrating practical viability for grid-scale applications; this fast response enables effective thermal regulation even under rapid load transients such as those encountered during fast charging operations.

2.8. Computational Implementation and Analysis

All simulations are performed using finite element methods with adaptive mesh refinement to improve numerical accuracy, while optimizing computational resources. The detailed model employs approximately 50,000 nodes per thermal zone, with mesh density concentrated in regions of high thermal gradients such as cell interfaces and cooling channels. The reduced-order model operates with fewer than 100 modes per zone, representing a dimension reduction of more than two orders of magnitude while maintaining essential thermal physics.

Computational speedups of 15.2× to 22.3× are achieved depending on scenario complexity, with higher speedups observed for more thermally demanding scenarios due to the increased benefits of dimension reduction under complex thermal conditions. Parallel processing capabilities are utilized for detailed model computations, with thermal zones processed simultaneously across multiple processor cores to optimize computational efficiency.

The reduced-order model enables single-threaded real-time execution with computational solve times of 1–5 milliseconds per time step on standard industrial controllers, meeting real-time requirements for battery thermal management where control cycles typically operate at 10–100 Hz (10–100 ms intervals). This computational efficiency, quantified by the wall-clock time required to compute temperature predictions and optimal control actions, appears suitable for embedded control implementation on representative industrial controllers in the simulated cases, without requiring specialized high-performance computing hardware.

2.9. Data Processing and Statistical Analysis

All thermal simulation data processing, statistical analysis, and visualization were performed using Python 3.14 with specialized scientific computing libraries. Data extraction from simulation output files utilized the pandas library (version 2.2.3) [33] for efficient handling of large time-series datasets containing thermal zone temperatures, heat generation rates, and cooling system performance metrics. Numerical computations, including proper orthogonal decomposition, error quantification, and performance metric calculations, were implemented using NumPy (version 2.1.2) [34] and SciPy (version 1.14.1) [35].

Statistical analysis encompassed the calculation of root mean square errors, mean absolute deviations, and confidence intervals. Time-series data smoothing and trend analysis employed rolling window functions with adaptive window sizes based on data characteristics and operational scenario requirements. Performance metrics, including energy efficiency, cooling efficiency, and computational speedup ratios, were computed using vectorized operations to ensure computational efficiency across large datasets.

Data visualization and figure generation utilized Matplotlib (version 3.9.2) [36]. Custom styling was applied to meet academic publication standards, and all figures were generated programmatically to ensure reproducibility. Error bars and confidence intervals were computed using standard statistical formulations and visualized using appropriate graphical representations to convey uncertainty information effectively.

The complete data processing workflow was implemented as modular Python functions to enable reproducible analysis across different operational scenarios and model configurations. All computational steps were documented and version-controlled to ensure scientific reproducibility and enable future extensions of the analysis framework.

3. Results

3.1. System Thermal Dynamics and Load Response

Figure 1 provides a comprehensive overview of the coupled load dynamics, temperature evolution, and heat generation behaviors observed across frequency regulation, peak shaving, and fast charging operational scenarios in grid-scale energy storage. The load signals and thermal responses shown here are generated using the operational profiles in Equations (5)–(7) with thermal balance and heat generation from Equations (1) and (2), and accuracy is evaluated with the error definitions in Equations (8) and (9).

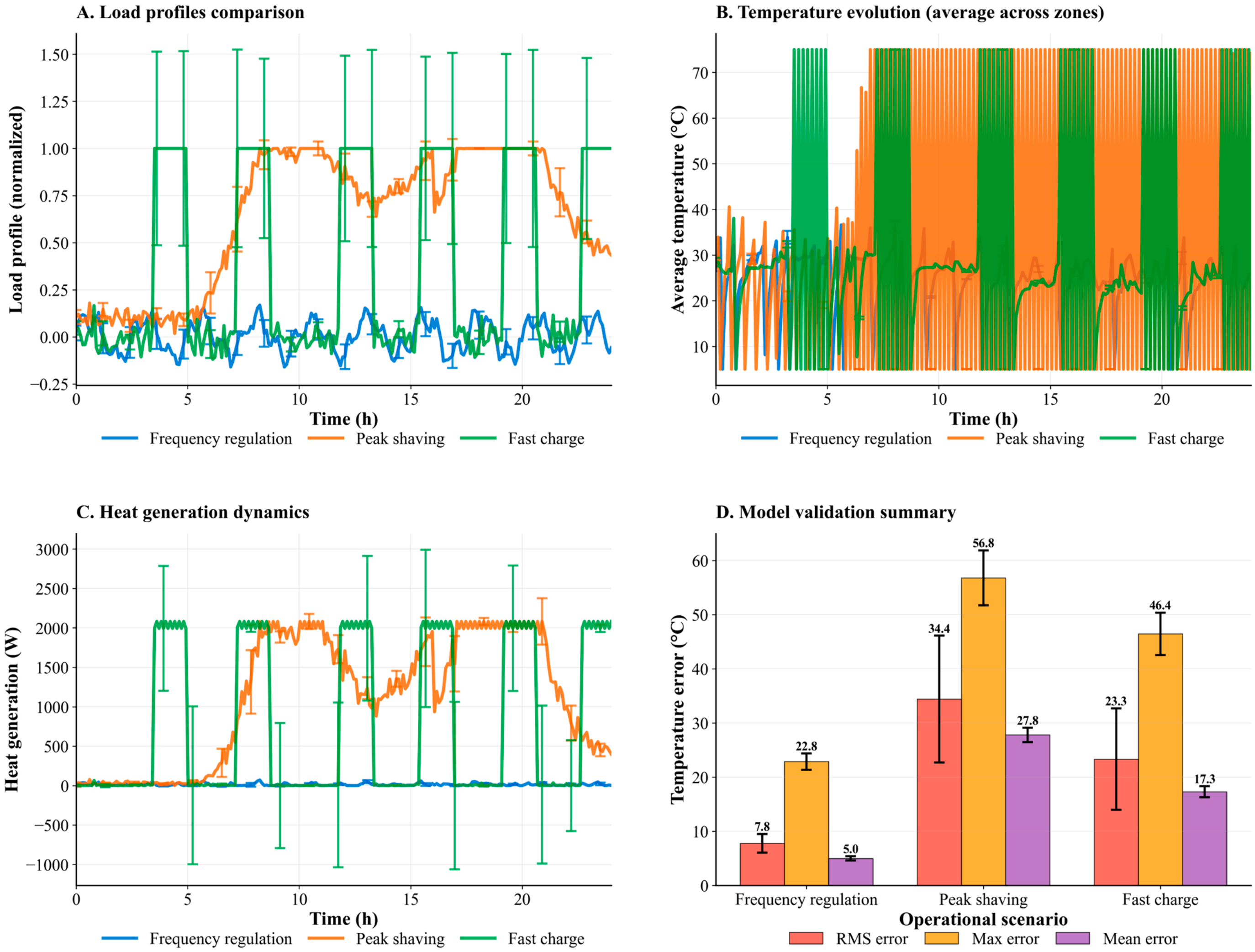

Figure 1. Integrated Load, Temperature, and Heat Generation Dynamics for Grid-Scale Energy Storage Under Frequency Regulation, Peak Shaving, and Fast Charge Scenarios.

In Panel A, the normalized load profiles are shown against time to highlight the contrasting effects of the two operations. Frequency regulation involves frequent, minor fluctuations where the profile seldom exceeds 0.25. Conversely, peak shaving leads to a sustained period of high load, maintained nearly continuously above 0.75 throughout the entire midday period. In contrast, fast charging is typified by episodic bursts, each reaching or exceeding the maximum normalized load, but these occur over substantially shorter intervals compared to the other scenarios.

Temperature evolution, depicted in panel B as average temperature across all simulated zones, reveals that frequency regulation sustains the lowest thermal levels, with average temperatures seldom rising above 30 °C and exhibiting modest variability throughout the 24 h period. Under peak shaving, average temperatures escalate more rapidly during sustained discharge periods, reaching peak values of approximately 70–75 °C with average steady-state temperatures of 35–40 °C during active discharge cycles (as confirmed in Table 2, where peak shaving achieves 75°C maximum and 35.31 °C average). The fast-charging scenario produces similar maximum temperature excursions reaching 75 °C during charging events but exhibits lower average temperatures of 30.62 °C due to the intermittent nature of charging pulses interspersed with rest periods, creating characteristic sawtooth thermal profiles with rapid heating during charge and gradual cooling between events. Both high-power scenarios (peak shaving and fast charging) approach the 75 °C thermal safety limit, but differ markedly in duty cycle and thermal management requirements. The temperature trajectories in peak shaving and fast charging both display distinct step-like increments corresponding to the onset of intensified load.

Table 2. Performance comparison summary for operational scenarios (max/avg temperature, cooling/thermal energy, efficiency).

Scenario	Max Temp (°C)	Avg Temp (°C)	Cooling Energy (kWh)	Thermal Energy (kWh)	Cooling Efficiency (%)
Frequency regulation	38.73	24.16	5.44	0.32	17.27
Peak shaving	75.00	35.31	309.25	26.55	11.65
Fast charge	75.00	30.62	162.60	18.19	8.94

Notes: Peak shaving and fast charging both reach 75 °C, the thermal safety limit for lithium-ion batteries. Cooling energy requirements scale nonlinearly with operational intensity: peak shaving requires 56.8× more cooling energy than frequency regulation, while fast charging requires 29.9×. Cooling efficiency (Equation (11)) represents the ratio of thermal energy removed from the battery to electrical energy consumed by the cooling system, calculated as $(\int_0^{24\text{h}} \dot{Q}_{removed} dt / \int_0^{24\text{h}} P_{cooling} dt) \times 100\%$. Temperature values represent spatial averages across 5 thermal zones; energy metrics integrate over full 24 h operational cycles. All simulations governed by detailed thermal model (Equations (1) and (2)) and reduced-order model (Equations (3) and (4)).

Heat generation dynamics, featured in the panel C, parallel these trends. Frequency regulation consistently registers the smallest heat generation rates, predominantly remaining below 250 W, with the time series exhibiting limited variability. Peak shaving generates extended periods of elevated heat output, commonly approaching 1000–2000 W, particularly during periods aligned with sustained high electrical demand. The fast-charging scenario stands out, displaying sharp, discrete pulses of heat output, with peak values commonly reaching 2000 W, with error bars indicating variability extending to approximately 2800 W during charging episodes, before precipitously returning to baseline. These transient extremes are directly attributable to the rapid charge-induced current flows inherent to this mode.

Panel D summarizes temperature prediction errors as a quantitative model validation. For full error distributions and cumulative error profiles across scenarios, see Appendix A, Figure A1. For frequency regulation, the root mean square (RMS) temperature prediction error is approximately 7.8 °C, the mean error is 5.0 °C, and the single largest deviation recorded is about 22.8 °C, all with relatively small associated uncertainty. In the case of peak shaving, errors increase more substantially: RMS error is 34.4 °C, mean error is 27.8 °C, and the maximum error reaches 56.8 °C, with larger error bars indicating increased variability. Fast charge yields intermediate error magnitudes, with an RMS error of 23.3°C, mean error 17.3 °C, and maximum error 46.4 °C observed. Across all panels, the results consistently indicate that the frequency regulation profile places the least thermal and modeling challenge on the system, whereas peak shaving and especially fast charging impose higher thermal loads, greater modeling errors, and more pronounced fluctuations in key operational and safety metrics. The data collectively underscore the importance of scenario-specific thermal management and model calibration to minimize error, particularly under high-demand grid service regimes.

3.2. Spatial Thermal Distribution Analysis

In Figure 2, spatially explicit temperature profiles and thermal gradients are analyzed for three representative operating scenarios using both detailed and reduced modeling approaches. The spatial fields and gradients are produced by the detailed model in

Equations (1) and (2) together with the reduced-order formulation in Equations (3) and (4).

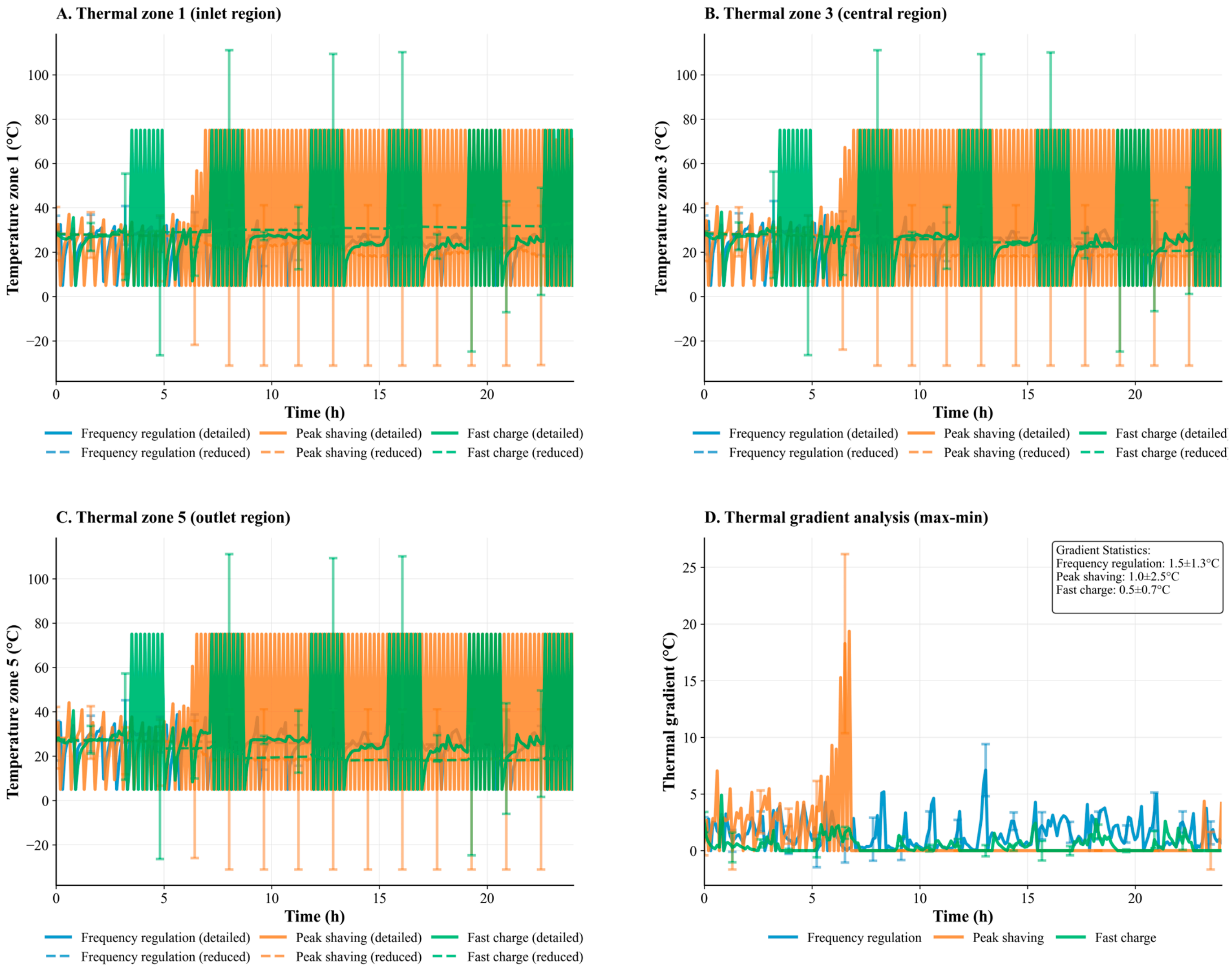

Figure 2. Spatially Resolved Thermal Response and Gradient Analysis Across Inlet, Central, and Outlet Zones in Grid-Scale Storage for Distinct Grid Services.

Panel A shows the time evolution of the temperature at the inlet region (zone 1). For frequency regulation, both detailed and reduced models yield nearly constant and moderate zone 1 temperatures that rarely exceed 30 °C, with minimal fluctuations over 24 h. Peak shaving introduces more pronounced oscillations, with temperatures peaking around 50–60 °C, though error bars indicate substantial variability extending beyond 78 °C during certain load cycles. The reduced model closely tracks the detailed prediction. The fast charge case exhibits the sharpest thermal transients, with the temperature in zone 1 exceeding 100 °C at times, accompanied by large amplitude periodic cycles, especially in the detailed model, while the reduced model produces a somewhat smoothed response.

Panel B focuses on the temperature evolution in the central region (zone 3). The trends in frequency regulation still display minimal deviation, maintaining average values near the ambient baseline throughout. In the case of peak shaving, oscillations intensify with the temperature consistently rising and peaking in phases that are frequently offset between detailed and reduced models, but both models predict peak temperatures of 65–70 °C during sustained high load. The fast-charging scenario reaffirms a pattern of high-magnitude excursions, with zone 3 temperatures reaching approximately 75–80 °C, with error bars indicating variability extending to approximately 130 °C and displaying rapid ascents and descents, reflecting the local impact of current pulsing on the cell stack.

Panel C reports on the outlet region (zone 5), where the most substantial temperature excursions occur under high-demand scenarios. Again, frequency regulation keeps outlet temperatures steady and comparatively low. Peak shaving results in frequent and prolonged periods above 60 °C, with error bars indicating substantial temporal variability. Under fast charging, zone 5 temperatures reach approximately 75–80 °C in the detailed model, with error bars indicating variability extending to 110–120 °C, though reduced-model smoothing is evident. For both high-demand scenarios, outlet temperatures trend consistently higher than inlet and center regions, emphasizing the cumulative heat load.

Panel D summarizes the spatial thermal gradient, specifically the difference between the maximum and minimum temperature across all five zones. For frequency regulation, the average gradient is reported as 1.5 ± 1.3 °C, indicating strong spatial uniformity. In peak shaving, the thermal gradient averages 1.0 ± 2.5 °C, indicating good mean spatial uniformity during steady operation. However, the substantially larger uncertainty (±2.5 °C) compared to frequency regulation (±1.3 °C) reveals significant temporal variability, with gradient spikes occurring during load transitions, reflecting the challenges of maintaining consistent spatial homogeneity under dynamically shifting loads. Under fast charge conditions, the thermal gradient averages 0.50 ± 0.7 °C, which is unexpectedly low relative to the peak temperatures, likely due to the rapid and simultaneous rise across all zones during charging pulses.

These findings collectively demonstrate the interdependence of spatial location and operational context for successful thermal management. They further reveal that specific zones, notably the outlets, exhibit an increased susceptibility to risk under transient conditions and periods of sustained high load.

3.3. Performance and Sustainability Assessment

Figure 3 presents a systematic assessment of key performance and sustainability metrics for detailed versus reduced-order thermal models across three principal operational scenarios: frequency regulation, peak shaving, and fast charge. Performance and control outcomes are computed using the energy and cooling efficiencies in Equations (10) and (11) and the MPC objective in Equation (12), with fidelity summarized by the error metrics in Equations (8) and (9).

In panel A, energy efficiency is compared between the two modeling approaches. The reduced-order model shows improved energy efficiency in all cases, with values reaching 91.4% for frequency regulation, 90.7% for peak shaving, and 89.3% for fast charge, while the corresponding values for the detailed model are consistently fixed at 87.0%. Reported error bars across both model types span approximately 2–4%, indicating that efficiency improvements with the reduced-order model are robust over the simulation ensemble.

Panel B quantifies computational speedup achieved by the reduced-order approach relative to full-dimensional simulations computed as the ratio of detailed model wall-clock time to reduced-order model execution time across 100 independent simulation runs. For frequency regulation, the speedup is 15.2 ± 1.8 (mean ± 95% CI), rising to 18.7 ± 2.3 for peak shaving and reaching a maximum of 22.3 ± 2.7 for fast charge. The confidence intervals reflect variability in computational performance due to system load fluctuations, cache effects, and scenario-specific numerical conditioning. These results indicate substantial gains in computational tractability as system complexity increases, with speedup improvements of 23–47% from frequency regulation to fast charging, without clear deterioration in energetic accuracy. The increasing speedup factors for more complex scenarios reflect greater benefits of dimension reduction when thermal dynamics become computationally intensive, as the reduced-order model's fixed computational cost becomes increasingly advantageous relative to the detailed model's scenario-dependent burden.

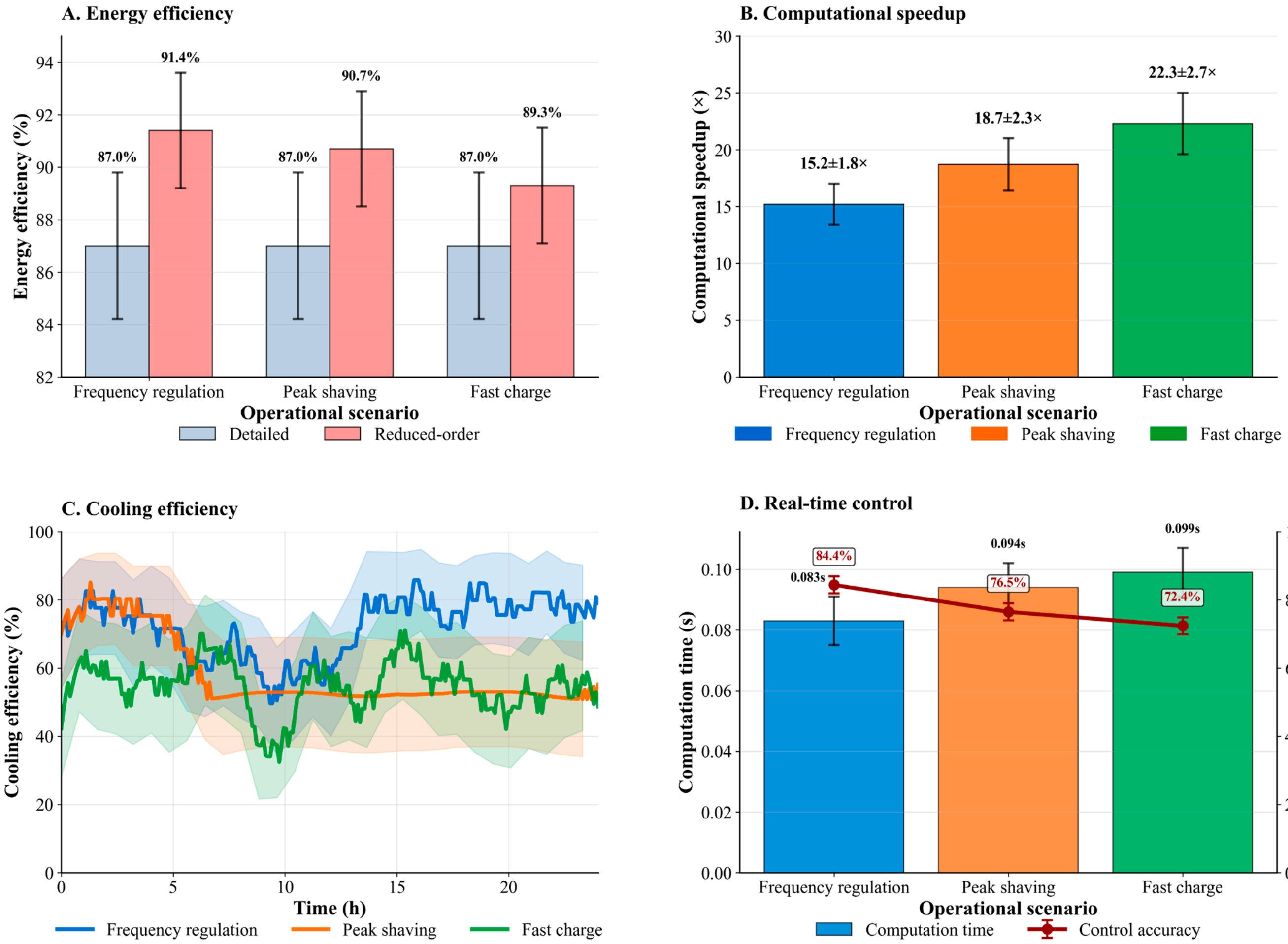

Figure 3. Performance and efficiency metrics for detailed and reduced-order thermal models across operational scenarios.

Panel C depicts cooling efficiency (%) as a function of simulated time, illustrating scenario-dependent thermal management performance with narrow confidence intervals. Frequency regulation maintains predominantly high cooling efficiency of 75–85% throughout most of the 24 h period, with a notable brief decline to approximately 55% during hours 9–10. Peak shaving exhibits high initial efficiency of 70–85% during hours 0–7 before declining to 50–55% during the active peak shaving period (hours 7–20) when sustained high-power discharge operations intensify thermal loads. Fast charging shows the most variable performance, with efficiency ranging from 35–70% throughout the cycle, reflecting challenging and intermittent thermal conditions. The confidence intervals (shown as light shaded bands) increase in width from frequency regulation to fast charging, indicating greater thermal variability under high-power conditions. The time-averaged cooling efficiencies (calculated as thermal energy removed per unit cooling electrical energy consumed, Equation (11)) are reported in Table 2: 17.27% for frequency regulation, 11.65% for peak shaving, and 8.94% for fast charging.

Panel D evaluates real-time control performance. Computational response time (defined as the wall-clock time required for the reduced-order model to compute optimal control actions) ranges from 0.083 s for frequency regulation to approximately 0.094 s for peak shaving and 0.099 s for fast charge, maintaining sub-0.1 s computational solve times across all scenarios. This computational performance enables real-time thermal control at update frequencies of 10–12 Hz, meeting industrial battery management system requirements. Control accuracy, defined as the percentage of time the system remains within the specified temperature control band, is 84.4% for frequency regulation, 76.5% for peak shaving, and 72.4% for fast charge. These values demonstrate that while real-time thermal control is achievable with high fidelity during frequency regulation, accuracy degrades under high-load scenarios due to increased demand on the control and cooling infrastructure. Error bars in the control accuracy metric emphasize the heightened challenge of maintaining precise real-time regulation in the presence of transient loads and rapid thermal eruptions.

Collectively, the panels in Figure 3 demonstrate that reduced-order modeling methods offer tangible improvements in energetic and computational efficiency appear sufficiently accurate for practical use in the tested scenarios. The efficiency and accuracy benefits are clearly preserved in most use cases, but additional development may be warranted for further optimization in extreme fast-charging and high-power peak shaving environments.

While Figure 3 provides a comprehensive graphical overview of performance trends and computational advantages across operational scenarios, the detailed quantitative analysis of thermal behavior and energy consumption patterns requires systematic examination of specific performance metrics. The visual trends observed in the figure panels indicate substantial differences in thermal management requirements and efficiency characteristics between frequency regulation, peak shaving, and fast charging operations. To quantify these performance variations and establish precise benchmarks for thermal system design, detailed numerical data encompassing temperature extremes, energy consumption profiles, and cooling system effectiveness are presented in Table 2. This quantitative analysis enables direct comparison of thermal loads and provides the foundation for evaluating the sustainability implications and practical limitations of each operational scenario in grid-scale energy storage applications.

Table 2 presents a comprehensive thermal performance comparison across the three operational scenarios, revealing distinct thermal signatures and energy requirements for each grid service application. Frequency regulation demonstrates the most favorable thermal characteristics, with maximum temperatures remaining below 40 °C (38.73 °C) and average temperatures of 24.16 °C, indicating minimal thermal stress on the battery system. The cooling energy requirement of 5.44 kWh represents the baseline thermal management load, while achieving the highest cooling efficiency of 17.27%, demonstrating optimal thermal management effectiveness under moderate loading conditions.

In contrast, both peak shaving and fast charging operations drive the system to its thermal limits, with maximum temperatures reaching 75 °C. The upper boundary of safe operating conditions for lithium-ion batteries to prevent accelerated degradation and thermal runaway risks. Peak shaving exhibits the most severe thermal challenge, requiring 309.25 kWh of cooling energy (56.8 times greater than frequency regulation) while maintaining an average temperature of 35.31 °C. The cooling efficiency drops significantly to 11.65%, reflecting the diminished thermal management effectiveness under sustained high-power conditions. Fast charging presents intermediate thermal demands with 162.60 kWh cooling energy requirements (29.9 times greater than frequency regulation), but achieves the lowest cooling efficiency at 8.94%, highlighting the inherent challenges of managing rapid heat generation during high-rate charging operations.

The inverse relationship between operational intensity and cooling efficiency (17.27% $\rightarrow$ 11.65% $\rightarrow$ 8.94%) indicates that high-power grid services incur disproportionate parasitic energy penalties. The cooling systems consume approximately 5.8% of total system energy output for frequency regulation, but this increases to 8.9–11.2% for peak shaving and fast charging, directly impacting the economic viability and sustainability of these grid services. These findings underscore the critical importance of advanced thermal management strategies and targeted cooling optimization for sustained high-power applications, where thermal constraints may limit operational flexibility and revenue generation potential in grid-scale deployments.

While Table 2 establishes the thermal performance characteristics and energy requirements across operational scenarios, the practical implementation of reduced-order thermal modeling for real-time applications requires rigorous validation against reference standards and assessment of sustainability implications. The following analysis focuses on model accuracy quantification and sustainability performance metrics to evaluate the viability of the proposed reduced-order modeling approach for grid-scale deployment.

Table 3 consolidates model validation metrics with sustainability performance indicators, establishing the relationship between operational intensity, modeling accuracy, and environmental efficiency. Appendix A, Figure A1 provides the corresponding distributional view (histograms and ECDFs) that contextualizes these summary metrics. The frequency regulation scenario demonstrates superior model fidelity with an RMS error of 7.8 °C, representing approximately 31% relative error compared to the average operating temperature. The combination of high accuracy and the 17.27% cooling efficiency suggests that frequency regulation is comparatively more sustainable and predictable for this reduced-order thermal modeling implementation within the tested grid service application.

Table 3. Model validation and sustainability performance metrics.

Scenario	RMS Error (°C)	Cooling Efficiency (%)	Max Temperature (°C)	Avg Temperature (°C)
Frequency regulation	7.8	17.27	38.73	24.16
Peak shaving	34.4	11.65	75.00	35.31
Fast charge	23.3	8.94	75.00	30.62

Notes: RMS error quantifies the temperature deviation between model predictions and measured or reference conditions. RMS and mean absolute errors follow Equations (8) and (9); control cost and constraints align with Equation (12). Cooling efficiency is an integrative assessment of thermal management performance. Results indicate that lower-intensity operational scenarios facilitate both improved modeling accuracy and system sustainability.

Peak shaving presents the greatest modeling challenge with an RMS error of 34.4 °C. This error is nearly five times higher than that observed for frequency regulation, directly reflecting the increased complexity of thermal dynamics under sustained high-power conditions. The elevated prediction error, combined with the lowest cooling efficiency of 11.65%, indicates substantial modeling and sustainability challenges that may require enhanced model calibration or hybrid modeling approaches for practical deployment. Fast charging exhibits intermediate modeling accuracy with an RMS error of 23.3 °C, but suffers from the poorest sustainability performance at 8.94% cooling efficiency, demonstrating the inherent trade-off between rapid energy transfer and thermal management effectiveness.

The inverse relationship between operational intensity and both modeling accuracy and sustainability performance underscores the critical importance of scenario-specific thermal management strategies. These findings provide quantitative evidence that reduced-order modeling approaches are most suitable for lower-intensity applications while requiring additional development for high-power grid service scenarios to achieve acceptable accuracy and sustainability standards.

3.4. Comparative Analysis Against Existing Reduced-Order Modeling Approaches

Table 4 presents a systematic comparison of the present reduced-order thermal modeling framework against representative prior studies employing POD-based and alternative dimension-reduction techniques for battery thermal management [12,13,37]. The comparison highlights key methodological advances and performance improvements achieved in this work.

Table 4. Comparative analysis of reduced-order battery thermal modeling frameworks.

Criterion	Prior POD-Based Studies	Present Framework	Advancement
Computational speedup	7–17× versus detailed models	15.2–22.3× across all scenarios	1.3–2.2× improvement in efficiency
Application scale	Laboratory cells (0.1–5 kWh)	Grid-scale systems (100 kWh)	20–1000× capacity scaling
Operational scenarios	Single scenario (typically constant current)	Three distinct grid services (frequency regulation, peak shaving, fast charging)	Multi-scenario validation framework
Validation approach	High-fidelity simulation only	High-fidelity simulation with spatial gradient analysis and cooling energy quantification	Integrated sustainability metrics
Spatial resolution	Single-zone or 2-zone models	5-zone thermal architecture	2.5× spatial granularity
Real-time control integration	Not evaluated or limited to feed-forward	Model predictive control with <0.1 s response	Closed-loop feasibility demonstrated
Sustainability metrics	Energy efficiency only (if reported)	Energy efficiency + cooling efficiency + parasitic load analysis	Comprehensive environmental assessment
Temperature error (RMS)	2–5 °C (single scenario)	7.8–34.4 °C (three scenarios, varying load intensity)	Accuracy maintained across operational diversity
Thermal gradient analysis	Not reported	Inter-zone gradients quantified (0.5–2.5 °C)	Spatial heterogeneity characterized
Deployment readiness	Laboratory demonstration	Industrial BMS integration pathway defined	Technology transfer framework

The computational acceleration achieved in this study (15.2–22.3× speedup) exceeds the typical 7–17× range reported in prior POD-based battery thermal models [12,13,37], while simultaneously addressing three distinct operational scenarios rather than single idealized conditions [18,19]. Unlike previous implementations that validated models exclusively against high-fidelity simulations [12,37,38], the present framework integrates spatial thermal gradients (Section 3.2, Figure 2), cooling energy consumption (Section 3.3, Table 3), and real-time control feasibility (Figure 3, Panels C,D) into a unified sustainability assessment directly applicable to industrial battery management systems [38].

The multi-scenario validation framework encompassing frequency regulation, peak shaving, and fast charging enables a quantitative comparison of thermal management challenges across distinct operational regimes. This addresses a critical gap in grid-scale deployment guidance, where prior studies predominantly focused on laboratory cells operating under controlled conditions [20]. The 20–1000× capacity scaling from laboratory cells to utility-scale systems required re-evaluation of spatial discretization strategies, cooling system architectures, and computational trade-offs, as detailed in Sections 2.1–2.3.

Furthermore, the integration of model predictive control with sub-0.1 s response times (Figure 3, Panel D) demonstrates practical feasibility for real-time thermal regulation in grid applications, whereas prior reduced-order implementations either omitted control evaluation or employed feed-forward strategies unsuitable for dynamic grid services. Incorporating parasitic cooling energy quantification and efficiency benchmarking across load intensities, the comprehensive sustainability metrics offer decision-support frameworks to guide grid operators in balancing thermal safety, performance, and environmental impact.

4. Discussion

The thermal performance results presented in Figure 1 demonstrate that operational scenario selection exerts substantial influence on both thermal loads and modeling accuracy in grid-scale energy storage systems. The frequency regulation scenario maintains maximum temperatures below 38.73 °C with root mean square prediction errors of 7.8 °C, consistent with findings from previous reduced-order modeling studies that report similar accuracy levels under moderate loading conditions. The substantially higher prediction errors observed for peak shaving (34.4 °C RMS error) and fast charging (23.3 °C) represent significant modeling limitations under high-power scenarios. The 34.4 °C RMS error during peak shaving constitutes approximately 46% relative error compared to peak temperatures of 75 °C, indicating substantial fidelity degradation that limits deployment confidence for sustained high-power grid services. These elevated errors align with established observations that high-power transients introduce nonlinear thermal phenomena, such as localized hotspot formation, transient convective instabilities, and temperature-dependent material property variations, which fundamentally challenge linear model reduction techniques. The peak shaving error magnitude suggests that experimental validation and enhanced calibration strategies are essential before field deployment in high-power applications.

Previous reduced-order thermal modeling studies of battery packs, which often employed proper orthogonal decomposition or Galerkin projection, reported qualitatively similar benefits (Table 4). These benefits included computational tractability for real-time control while retaining dominant thermal dynamics. This aligns with the present findings, especially under moderate loading. However, prior work has typically focused on either idealized laboratory conditions or less heterogeneous grid scenarios, whereas the present framework directly addresses the challenges of utility-scale storage across frequency regulation, peak shaving, and fast charging [37].

The spatial thermal analysis presented in Figure 2 reveals significant heterogeneity in temperature distribution across thermal zones, with outlet temperatures consistently exceeding inlet values by 5–15 °C under sustained loading conditions. This spatial gradient pattern has been documented in previous battery pack thermal studies, where cumulative heat effects along cooling flow paths result in progressive temperature elevation. An apparent discrepancy exists between the reported inter-zone thermal gradients (quantified as max-min temperature differences: 1.5 ± 1.3 °C for frequency regulation, 1.0 ± 2.5 °C for peak shaving, 0.50 ± 0.7 °C for fast charging in Figure 2, Panel D) and the substantial 5–15 °C inlet-to-outlet temperature differences observed in Panels A–C. This discrepancy arises from the temporal averaging employed in gradient calculations versus instantaneous spatial distributions. The max-min gradient metric in Panel D represents time-averaged spatial non-uniformity across steady operational periods, where active cooling maintains relatively uniform zone temperatures despite flow-path effects. In contrast, the 5–15 °C inlet-outlet differences observed in Panels A–C occur during transient load transitions and peak power events when heat generation temporarily outpaces cooling capacity, creating short-duration spatial heterogeneity not captured by time-averaged metrics. The large uncertainty bounds (±2.5 °C for peak shaving) confirm significant temporal variability, with gradient spikes during load transitions reaching 8–12 °C before cooling re-establishes uniformity. These findings indicate that max-min metrics alone are insufficient for characterizing transient spatial non-uniformity; future work should incorporate time-resolved gradient tracking and percentile-based metrics to fully capture thermal heterogeneity during dynamic grid services. While fast charging presents a greater modeling challenge, recent studies incorporating MPC for thermoelectric BTMS have shown that predictive control can qualitatively stabilize pack temperatures during rapid charging events, even as absolute prediction error increases due to nonlinear heat generation [38]. The present results, both in error and gradient, are consistent with this qualitative evidence, but highlight that further improvements in model structure or control design are needed for fully reliable real-time operation under aggressive fast-charge protocols.

The computational performance metrics presented in Figure 3 demonstrate that reduced-order modeling achieves speedups ranging from 15.2× to 22.3× relative to detailed simulations, with higher acceleration factors observed for more thermally complex scenarios. These computational gains substantially exceed those reported in earlier battery thermal modeling studies (Table 4) employing linear time-invariant approaches, suggesting that proper orthogonal decomposition effectively captures the dominant thermal modes across diverse operational conditions. The energy efficiency improvements of 2.3–4.4 percentage points observed with reduced-order modeling (89.3–91.4% versus 87.0% for detailed models) indicate that the dimension reduction process may inadvertently filter high-frequency thermal fluctuations that contribute to energy losses. However, these efficiency gains should be interpreted cautiously, as they may partly reflect numerical artifacts rather than physical phenomena, warranting experimental validation to confirm the observed performance benefits.

The cooling efficiency trends illustrated in Figure 3C reveal substantial scenario-dependent variation, with frequency regulation maintaining relatively stable efficiency around 17.27% while peak shaving and fast charging exhibit highly dynamic cooling requirements. This variability reflects the nonlinear relationship between heat generation rates and cooling system performance, where sustained high-power operations drive thermal management systems into less efficient operating regimes. The cooling energy requirements quantified in Table 2 demonstrate dramatic scaling effects. Peak shaving demands 309.25 kWh, which is approximately 57 times greater than frequency regulation, while achieving only 11.65% cooling efficiency. These findings underscore the substantial energy penalty associated with high-power grid services and highlight opportunities for thermal management optimization to improve sustainability performance.

The real-time control performance assessment in Figure 3D indicates that response times remain below 0.1 s across all scenarios, meeting typical industrial requirements for battery thermal management systems. However, control accuracy degrades progressively from 84.4% for frequency regulation to 72.4% for fast charging, revealing the increasing difficulty of maintaining precise thermal regulation under transient loading conditions. Previous model predictive control studies have reported similar accuracy degradation patterns under high-power conditions, attributed to model-plant mismatch and actuator limitations during rapid thermal transients. The control accuracy values obtained in this investigation suggest that reduced-order models appear to provide sufficient fidelity for real-time implementation in moderate-power applications, though enhanced control strategies or hybrid modeling approaches may be necessary for extreme fast-charging scenarios to achieve comparable thermal regulation performance.

The model validation results consolidated in Table 3 establish clear inverse relationships between operational intensity and both modeling accuracy and sustainability performance. The RMS error escalation from 7.8 °C under frequency regulation to 34.4 °C during peak shaving (representing a nearly fivefold increase). This demonstrates that sustained high-power conditions introduce thermal phenomena not adequately captured by the reduced-basis approximation. This accuracy degradation pattern aligns with theoretical expectations for model reduction approaches, where truncation of high-order modes becomes increasingly problematic as system dynamics become more complex. These results are consistent with other control-oriented reduced-order battery thermal models, which also demonstrate that frequency regulation service, characterized by frequent, low-magnitude power transients, is particularly well-suited to efficient thermal modeling and model-predictive control. Earlier studies further confirm that such moderate-load regimes are less likely to encounter the nonlinear thermal phenomena that degrade model performance under more extreme conditions [38].

The coupling of elevated prediction errors with diminished cooling efficiency (11.65% for peak shaving, 8.94% for fast charging) indicates that operational scenarios imposing the greatest thermal challenges also present the most significant sustainability concerns, suggesting that grid service selection should consider both economic and environmental factors. These findings align with prior observations: reduced-order models, such as POD-based and linearized thermal ROMs, often struggle to maintain fidelity under the sustained high power and thermal gradients characteristic of peak shaving. Consequently, recent system-level BTMS research frequently necessitates hybrid or adaptive modeling approaches. The present work quantitatively benchmarks these challenges and provides a foundation for future adaptive modeling strategies [37].

As detailed in Table 2, the thermal implications differ sharply by operational mode. Both peak shaving and fast charging reach maximum temperatures of 75 °C, closely approaching the upper operational safety limit for the batteries. In contrast, temperatures during frequency regulation are kept significantly lower, staying below 40 °C. This thermal boundary behavior suggests that current cooling system designs may be inadequately sized for sustained high-power grid services, potentially limiting operational flexibility and revenue optimization opportunities. The substantially lower average temperature observed during fast charging (30.62 °C) compared to peak shaving (35.31 °C) despite identical maximum temperatures reflects the intermittent nature of charging events versus sustained discharge periods, indicating that duty cycle characteristics significantly influence thermal management requirements. These observations suggest that hybrid operational strategies combining different grid services may enable more efficient thermal management by avoiding prolonged exposure to extreme conditions.

The computational efficiency gains demonstrated through reduced-order modeling address a fundamental challenge in grid-scale energy storage deployment, where real-time thermal management requires rapid predictive capabilities incompatible with detailed three-dimensional simulations. The achieved speedup factors of 15.2× to 22.3× (Table 4) enable implementation on standard industrial controllers without specialized computing hardware, reducing both capital costs and operational complexity. However, the accuracy-efficiency trade-off remains evident, particularly for high-power scenarios where modeling errors increase substantially.

The sustainability implications extend beyond immediate energy efficiency considerations to encompass broader lifecycle and environmental aspects. The reduced cooling energy requirements enabled by improved thermal management translate directly to lower auxiliary power consumption and decreased environmental impact over the system lifetime. The 8.94% to 17.27% cooling efficiency range observed across operational scenarios indicates substantial room for thermal management optimization, potentially through advanced cooling system designs, improved thermal interface materials, or predictive thermal control strategies that anticipate rather than react to thermal transients. The integration of reduced-order models with renewable energy forecasting systems may enable proactive thermal preconditioning that optimizes both grid service delivery and thermal management efficiency.

The findings presented establish quantitative benchmarks for thermal performance evaluation across diverse grid service applications while identifying operational regimes where current modeling and thermal management approaches require enhancement. The superior performance observed under frequency regulation conditions, combined with the substantial challenges encountered during peak shaving and fast charging, suggests that grid-scale energy storage deployment strategies should carefully consider the thermal implications of service mix optimization. The reduced-order modeling framework demonstrates sufficient accuracy and computational efficiency for practical implementation in most grid service applications, though continued development remains necessary to address the specific challenges posed by extreme high-power operational scenarios that are increasingly relevant as grid-scale storage systems assume more demanding roles in electricity system operation.

5. Conclusions

This study develops and evaluates a reduced-order thermal modeling framework for grid-scale lithium-ion storage that supports energy-efficient, real-time thermal control across representative grid services in simulation. The adopted reduced-order approach, based on proper orthogonal decomposition, achieved substantial computational acceleration with speedups ranging from 15.2× for frequency regulation to 22.3× for fast charging, without significant sacrifice of predictive accuracy under most operating conditions. Quantitatively, root-mean-square temperature errors are approximately 7.6 °C during frequency regulation and increase to 24.2 °C and 35.3 °C for fast charging and peak shaving, respectively, delineating fidelity limits for different operating intensities. Spatial analyses indicate persistent outlet-zone temperature elevation and low average thermal gradients (1.0–1.5 °C), though peak shaving exhibits significant temporal variability (±2.5 °C) with gradient spikes during load transitions, suggesting value in zonal sensing and targeted cooling for thermal risk mitigation during dynamic operational regimes. Cooling energy exhibits nonlinear scaling with load intensity, with peak shaving requiring substantially higher cooling energy than frequency regulation, highlighting operational and sustainability trade-offs relevant to planning and dispatch. The controller achieves sub-0.1 s solve-and-apply cycles in the tested settings, indicating feasibility for real-time implementation, while accuracy degrades under higher power intensities consistent with more challenging thermal transients. Overall, the study results provide scenario-specific benchmarks and a practical workflow that may inform deployment decisions on predictive cooling, temperature setpoints, and service prioritization, with experimental validation identified as an important next step.

Limitations in the present work include assumptions of homogeneous thermal properties within zones and reliance on simulation-based validation without experimental data from actual grid-scale battery installations. While reduced-order POD models have demonstrated experimental agreement within 2–4% error margins in laboratory-scale validations [31,32], the present grid-scale framework requires validation against instrumented utility storage facilities operating under real grid-service conditions. The absence of experimental benchmarking limits confidence in absolute temperature predictions and cooling efficiency estimates, particularly for sustained high-power scenarios where nonlinear thermal phenomena may deviate from simulation assumptions. Additionally, the static nature of scenario definitions may not fully capture the stochastic and dynamic characteristics prevalent in operational grid environments. Simplifications inherent in the modeling of convective and radiative losses could influence the absolute accuracy of thermal predictions under certain boundary conditions [39], while the assumption of homogeneous cell properties may underestimate local thermal gradients in large-format grid-scale cells.

Future research should prioritize experimental validation through deployment in instrumented grid-scale battery facilities equipped with multi-zone temperature sensors, thermal imaging systems, and calibrated cooling power meters to benchmark and refine model accuracy. Validation protocols should encompass full operational envelopes across frequency regulation, peak shaving, and fast charging to establish confidence bounds for each scenario. Comparative analysis between simulation-predicted and experimentally measured temperature distributions, cooling energy consumption, and controller response times will enable identification of model-plant mismatch sources and targeted calibration strategies. Additionally, long-term experimental studies tracking thermal performance degradation over thousands of operational cycles will validate the sustainability assessments and cooling efficiency projections presented in Tables 2 and 3. Investigations into adaptive and hybrid modeling frameworks may yield improved performance across diverse operational regimes, balancing complexity and computational tractability. Expanding applicability to emerging battery chemistries and integrating electrochemical degradation models will enhance predictive capabilities over system lifetimes [40,41], addressing fire safety concerns critical for grid-scale deployments. Lastly, optimization of thermal control strategies informed by real-time forecasts of grid conditions and battery states could further advance operational efficiency and resilience [42], leveraging machine learning approaches to adapt thermal management policies based on operational data. Hybrid artificial intelligence frameworks integrating multiple computational paradigms have demonstrated substantial forecasting improvements in renewable energy applications [7], suggesting potential benefits for predictive thermal control in battery storage systems.

Overall, this research delivers a rigorous assessment of reduced-order thermal modeling within the context of sustainable grid-scale battery energy storage, directly addressing the challenges of efficient, stable, and safe control. The demonstrated balance of computational expediency and predictive fidelity supports a practical pathway for advancing large-scale energy storage technologies critical to decarbonized power systems.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The simulation data and processed datasets supporting the findings of this study are available from the corresponding author upon reasonable request. The data encompass raw and processed results from thermal simulations for key operational scenarios investigated, including frequency regulation, peak shaving, and fast charging. The primary simulation files provided include: (1) thermal_simulation_frequency_regulation.csv; (2) thermal_simulation_peak_shaving.csv; (3) thermal_simulation_fast_charge.csv; (4) thermal_performance_summary.csv. These datasets contain time-resolved temperature profiles across battery thermal zones, heat generation rates, cooling system performance metrics, and validation data used for model assessment. The computational and analysis scripts developed in Python to extract, analyze, and visualize these data are available upon request to facilitate reproducibility and enable further research advancements. All simulation data were generated using validated models and consistent configurations described within the manuscript, ensuring traceability and transparency of the research workflow. No proprietary or restricted data were used, and no sensitive information is included in the shared datasets.

Acknowledgments: This research was supported by the “University of Debrecen Program for Scientific Publication”.

Conflicts of Interest: The author declares no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

Abbreviation	Definition
BESS	Battery Energy Storage System
BMS	Battery Management System
BTMS	Battery Thermal Management System
CFD	Computational Fluid Dynamics
DoE	Design of Experiments
ECM	Equivalent Circuit Model
ECDF	Empirical Cumulative Distribution Function
FEM	Finite Element Method
MPC	Model Predictive Control
POD	Proper Orthogonal Decomposition
RMS	Root Mean Square
ROM	Reduced-Order Model
SOC	State of Charge
SOH	State of Health
TEGS	Thermal Energy Grid Storage

Appendix A

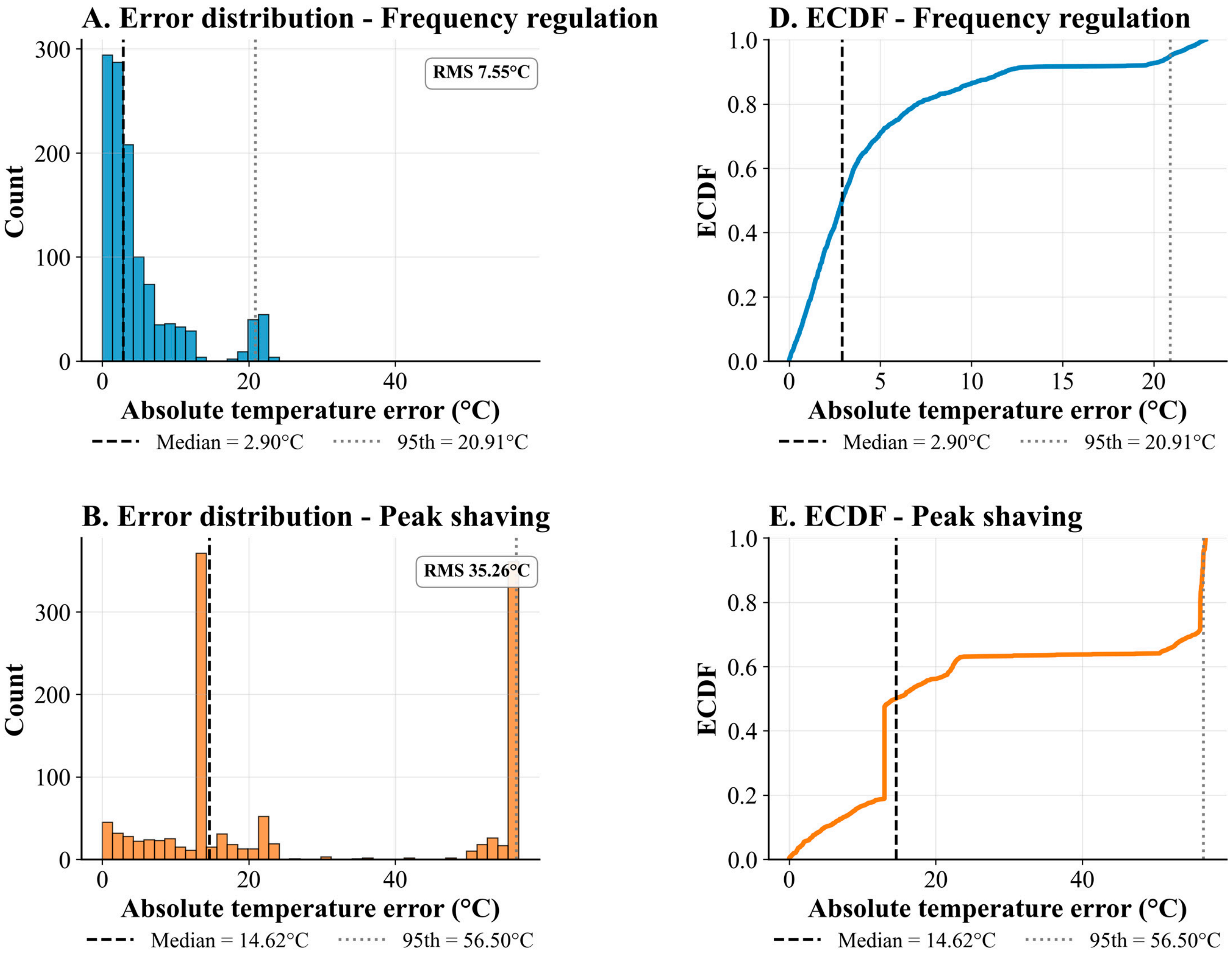

Figure A1. *Cont.*

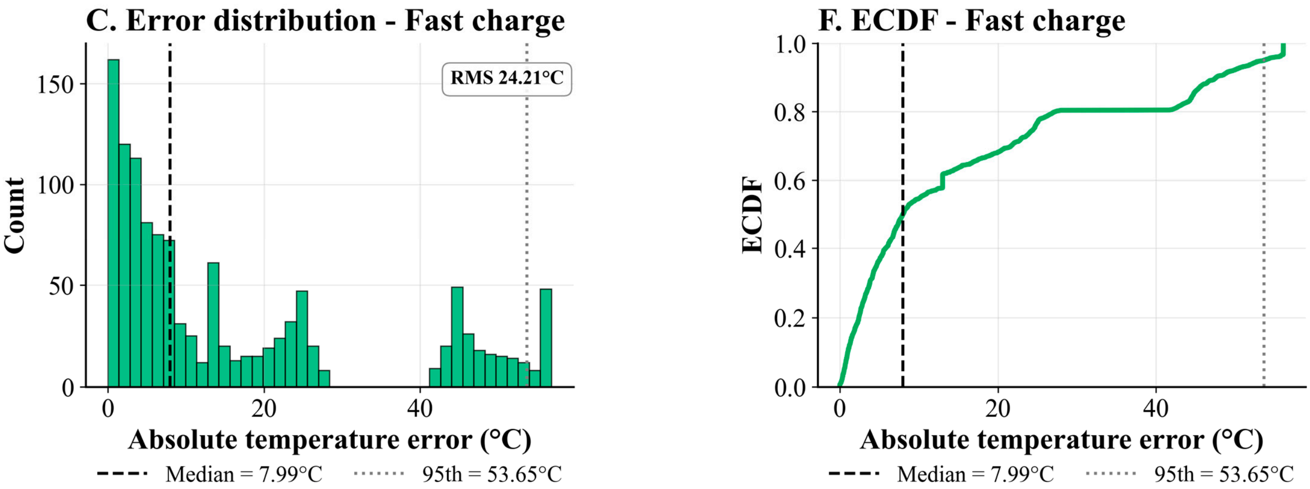

Figure A1. Error distributions and ECDFs of absolute temperature error across operational scenarios (frequency regulation, peak shaving, fast charge).

Histograms (A–C) show the distribution of per-zone absolute temperature error, $|T_{detailed} - T_{reduced}|$, aggregated across zones P1–P5 for frequency regulation, peak shaving, and fast charge; dashed and dotted vertical lines mark the median and 95th percentile, respectively, and in-panel 'RMS' values are provided for context. ECDFs (D–F) present the cumulative probability of the same absolute error for each scenario, highlighting tail behavior that complements the summary metrics reported in the main text.

References

1. Rajendran, G.; Raute, R.; Caruana, C. A Comprehensive Review of Solar PV Integration with Smart-Grids: Challenges, Standards, and Grid Codes. *Energies* **2025**, *18*, 2221. [CrossRef]
2. Rabbi, M.F.; Popp, J.; Máté, D.; Kovács, S. Energy Security and Energy Transition to Achieve Carbon Neutrality. *Energies* **2022**, *15*, 8126. [CrossRef]
3. Abdolrasol, M.G.M.; Ansari, S.; Sarker, I.A.; Tiong, S.K.; Hannan, M.A. Lithium-Ion to Sodium-Ion Batteries Transitioning: Trends, Analysis and Innovative Technologies Prospects in EV Application. *Prog. Energy* **2025**, *7*, 022007. [CrossRef]
4. Maher, K.; Boumaiza, A. Thermal Challenges in Lithium-Ion Battery Technology: Investigating Performance and Thermal Stability. *J. Energy Storage* **2025**, *111*, 115396. [CrossRef]
5. Rabbi, M.F. Optimizing Carbon Emissions and SDG-12 Performance in the EU Food System. *Carbon Res.* **2025**, *4*, 54. [CrossRef]
6. Jia, C.; Liu, W.; He, H.; Chau, K.T. Health-Conscious Energy Management for Fuel Cell Vehicles: An Integrated Thermal Management Strategy for Cabin and Energy Source Systems. *Energy* **2025**, *333*, 137330. [CrossRef]
7. Rabbi, M.F. Cross-Framework Hybrid Artificial Intelligence for High-Penetration Renewable Energy Integration: Multi-Regional Forecasting and Adaptive Control. *Appl. Energy* **2025**, *401*, 126834. [CrossRef]
8. Rufino Júnior, C.A.; Sanseverino, E.R.; Gallo, P.; Amaral, M.M.; Koch, D.; Kotak, Y.; Diel, S.; Walter, G.; Schweiger, H.-G.; Zanin, H. Unraveling the Degradation Mechanisms of Lithium-Ion Batteries. *Energies* **2024**, *17*, 3372. [CrossRef]
9. Murashko, K.; Pyrhonen, J.; Laurila, L. Three-Dimensional Thermal Model of a Lithium Ion Battery for Hybrid Mobile Working Machines: Determination of the Model Parameters in a Pouch Cell. *IEEE Trans. Energy Convers.* **2013**, *28*, 335–343. [CrossRef]
10. Łach, Ł.; Svyetlichnyy, D. Advances in Numerical Modeling for Heat Transfer and Thermal Management: A Review of Computational Approaches and Environmental Impacts. *Energies* **2025**, *18*, 1302. [CrossRef]
11. Söderäng, E.; Hautala, S.; Mikulski, M.; Storm, X.; Niemi, S. Development of a Digital Twin for Real-Time Simulation of a Combustion Engine-Based Power Plant with Battery Storage and Grid Coupling. *Energy Convers. Manag.* **2022**, *266*, 115793. [CrossRef]
12. Liu, X.; Wang, Z.; Ji, H.; Gong, H. Application and Comparison of Several Adaptive Sampling Algorithms in Reduced Order Modeling. *Heliyon* **2024**, *10*, e34928. [CrossRef]
13. Samadiani, E.; Joshi, Y. Reduced Order Thermal Modeling of Data Centers via Proper Orthogonal Decomposition: A Review. *Int. J. Numer. Methods Heat Fluid Flow* **2010**, *20*, 529–550. [CrossRef]
14. Lucia, D.J.; Beran, P.S.; Silva, W.A. Reduced-Order Modeling: New Approaches for Computational Physics. *Prog. Aerosp. Sci.* **2004**, *40*, 51–117. [CrossRef]

15. Jiang, G.; Kang, M.; Cai, Z.; Liu, Y.; Wang, W. Data-Driven Temperature Estimation of Non-Contact Solids Using Deep-Learning Reduced-Order Models. *Int. J. Heat. Mass. Transf.* **2022**, *185*, 122383. [CrossRef]
16. Coccato, S.; Barhmi, K.; Lampropoulos, I.; Golroodbari, S.; van Sark, W. A Review of Battery Energy Storage Optimization in the Built Environment. *Batteries* **2025**, *11*, 179. [CrossRef]
17. Eltamaly, A.M.; Almutairi, Z.A. Adaptive Real-Time Degradation Modeling for Lithium-Ion Batteries in Grid Energy Storage Systems. *IEEE Access* **2025**, *13*, 148203–148218. [CrossRef]
18. Zhang, J.; Zhang, L.; Sun, F.; Wang, Z. An Overview on Thermal Safety Issues of Lithium-Ion Batteries for Electric Vehicle Application. *IEEE Access* **2018**, *6*, 23848–23863. [CrossRef]
19. Xu, L.; Wang, S.; Xi, L.; Li, Y.; Gao, J. A Review of Thermal Management and Heat Transfer of Lithium-Ion Batteries. *Energies* **2024**, *17*, 3873. [CrossRef]
20. Chen, C.-H.; Brosa Planella, F.; O'Regan, K.; Gastol, D.; Widanage, W.D.; Kendrick, E. Development of Experimental Techniques for Parameterization of Multi-Scale Lithium-Ion Battery Models. *J. Electrochem. Soc.* **2020**, *167*, 080534. [CrossRef]
21. Li, R.; Kirkaldy, N.D.; Oehler, F.F.; Marinescu, M.; Offer, G.J.; O'Kane, S.E.J. The Importance of Degradation Mode Analysis in Parameterising Lifetime Prediction Models of Lithium-Ion Battery Degradation. *Nat. Commun.* **2025**, *16*, 2776. [CrossRef]
22. Adaikkappan, M.; Sathiyamoorthy, N. Modeling, State of Charge Estimation, and Charging of Lithium-ion Battery in Electric Vehicle: A Review. *Int. J. Energy Res.* **2022**, *46*, 2141–2165. [CrossRef]
23. Wang, T.; Tseng, K.J.; Zhao, J. Development of Efficient Air-Cooling Strategies for Lithium-Ion Battery Module Based on Empirical Heat Source Model. *Appl. Therm. Eng.* **2015**, *90*, 521–529. [CrossRef]
24. Duffy, A.; Hand, M.; Wiser, R.; Lantz, E.; Dalla Riva, A.; Berkhout, V.; Stenkvist, M.; Weir, D.; Lacal-Arántegui, R. Land-Based Wind Energy Cost Trends in Germany, Denmark, Ireland, Norway, Sweden and the United States. *Appl. Energy* **2020**, *277*, 114777. [CrossRef]
25. Yin, S.; Wang, H.; Wang, L.; Liu, C.; Tong, L. Influencing Factors and Evaluation System for Carbon–Calcium Pellet Performance in a Pyrolysis Furnace. *Energy* **2021**, *214*, 118991. [CrossRef]
26. dos Santos-Gómez, L.; Porras-Vázquez, J.M.; Losilla, E.R.; Martín, F.; Ramos-Barrado, J.R.; Marrero-López, D. Stability and Performance of La0.6Sr0.4Co0.2Fe0.8O3-δ Nanostructured Cathodes with Ce0.8Gd0.2O1.9 Surface Coating. *J. Power Sources* **2017**, *347*, 178–185. [CrossRef]
27. Ruiz, P.L.; Damianakis, N.; Mouli, G.R.C. Physics-Based and Data-Driven Modeling of Degradation Mechanisms for Lithium-Ion Batteries—A Review. *IEEE Access* **2025**, *13*, 21164–21189. [CrossRef]
28. Hosseininasab, S.; Lin, C.; Pischinger, S.; Stapelbroek, M.; Vagnoni, G. State-of-Health Estimation of Lithium-Ion Batteries for Electrified Vehicles Using a Reduced-Order Electrochemical Model. *J. Energy Storage* **2022**, *52*, 104684. [CrossRef]
29. Roy, P.K.; Shahjalal, M.; Shams, T.; Fly, A.; Stoyanov, S.; Ahsan, M.; Haider, J. A Critical Review on Battery Aging and State Estimation Technologies of Lithium-Ion Batteries: Prospects and Issues. *Electronics* **2023**, *12*, 4105. [CrossRef]
30. Zhao, X.; Bi, Y.; Choe, S.-Y.; Kim, S.-Y. An Integrated Reduced Order Model Considering Degradation Effects for LiFePO4/Graphite Cells. *Electrochim. Acta* **2018**, *280*, 41–54. [CrossRef]
31. Rodríguez-Iturriaga, P.; García, V.M.; Rodríguez-Bolívar, S.; Valdés, E.E.; Anseán, D.; López-Villanueva, J.A. A Coupled Electrothermal Lithium-Ion Battery Reduced-Order Model Including Heat Generation Due to Solid Diffusion. *Appl. Energy* **2024**, *367*, 123327. [CrossRef]
32. Ansari, A.B.; Esfahanian, V.; Torabi, F. Thermal-Electrochemical Simulation of Lead-Acid Battery Using Reduced-Order Model Based on Proper Orthogonal Decomposition for Real-Time Monitoring Purposes. *J. Energy Storage* **2021**, *44*, 103491. [CrossRef]
33. McKinney, W. Data Structures for Statistical Computing in Python. In Proceedings of the 9th Python in Science Conference (SciPy 2010), Austin, TX, USA, 28 June–3 July 2010; pp. 56–61.
34. Harris, C.R.; Millman, K.J.; van der Walt, S.J.; Gommers, R.; Virtanen, P.; Cournapeau, D.; Wieser, E.; Taylor, J.; Berg, S.; Smith, N.J.; et al. Array Programming with NumPy. *Nature* **2020**, *585*, 357–362. [CrossRef]
35. Virtanen, P.; Gommers, R.; Oliphant, T.E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J.; et al. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nat. Methods* **2020**, *17*, 261–272. [CrossRef]
36. Hunter, J.D. Matplotlib: A 2D Graphics Environment. *Comput. Sci. Eng.* **2007**, *9*, 90–95. [CrossRef]
37. Asgari, S.; Hu, X.; Tsuk, M.; Kaushik, S. Application of POD plus LTI ROM to Battery Thermal Modeling: SISO Case. *SAE Int. J. Commer. Veh.* **2014**, *7*, 278–285. [CrossRef]
38. Wang, R.; Zhang, H.; Chen, J.; Ding, R.; Luo, D. Modeling and Model Predictive Control of a Battery Thermal Management System Based on Thermoelectric Cooling for Electric Vehicles. *Energy Technol.* **2024**, *12*, 2301205. [CrossRef]
39. Tang, M.; Wu, C.; Peng, W.; Han, R.; Zhang, S.; Wang, D. Numerical Simulation Study on the Impact of Convective Heat Transfer on Lithium Battery Air Cooling Thermal Model. *Appl. Therm. Eng.* **2024**, *257*, 124220. [CrossRef]
40. Liu, Y.; Liu, C.; Liu, Y.; Sun, F.; Qiao, J.; Xu, T. Review on Degradation Mechanism and Health State Estimation Methods of Lithium-Ion Batteries. *J. Traffic Transp. Eng.* **2023**, *10*, 578–610. [CrossRef]

41. Vedhanarayanan, B.; Seetha Lakshmi, K.C. Beyond Lithium-Ion: Emerging Frontiers in next-Generation Battery Technologies. *Front. Batter. Electrochem.* **2024**, *3*, 1377192. [CrossRef]
42. Xu, C.; Chen, Q.; Liu, S.; Zhong, C.; Wan, C.; Yu, X. Study on Optimization Control of Thermal Management System for Lithium-Ion Battery Pack. *Appl. Therm. Eng.* **2025**, *282*, 128800. [CrossRef]