ORIGINAL RESEARCH **Open access**

Weather-Integrated Adaptive Reinforcement Learning for Stable Energy Management in Extreme Climate Conditions

Daniel Kim[1*], Jihoon Park[1]

Abstract

Strengthening the stability of energy management systems is vital, above all when dealing with severe weather occurrences. A standard technique to stop abrupt oscillations relies on predictive models; yet this typically calls for more elaborate models and increased computing demands. In the present study, weather predictions are embedded within a framework using Adaptive Reinforcement Learning for Energy Management (ARLEM). Because ARLEM functions as an online, model-free, reward-driven RL technique, it eliminates the need for any (predictive) modelling, with weather forecasts providing supplementary details on the agent's surroundings. The consequences of adding weather forecasts to ARLEM are analyzed for a standard urban neighbourhood in Stockholm across a winter containing two cold waves, while accounting for 17 future climate scenarios spanning 2040–2069. Several adaptive policy frameworks are examined using four forecast horizons of 3, 6, 12, and 24 h. Findings reveal that embedding weather forecasts in the decision process can markedly cut down variations in the control system, yielding greater stability in energy management. Furthermore, it raises energy efficiency through reductions in both average and peak loads, most notably amid extreme weather conditions. In summary, this advancement supports greater climate resilience in energy systems.

Keywords Energy management, Weather forecast, Adaptive reinforcement learning, Control, Stability, Climate resilience

*Correspondence:
Daniel Kim
daniel.kim@gmail.com
[1] Department of Intelligent Energy Systems, College of Engineering, Korea University, Seoul, South Korea

Introduction

Optimally managing energy in buildings and urban energy systems is difficult because of the constantly shifting interactions among multiple factors [1], such as weather patterns, occupant activities, and renewable energy integration [2]. Weather plays a decisive part in shaping building energy use [3], yet the intricate nature of weather patterns complicates precise evaluations [4]. Differences in building technologies and setups [5], along with the incorporation of renewable energy [6], add further layers of complexity to energy control processes. Rapid changes in urban settings call for flexible energy management approaches [7] and sophisticated predictive control techniques [8], along with adequate robustness to handle extreme situations and maintain strong reliability.

Predictive control techniques, such as Model Predictive Control (MPC), are frequently used to optimize energy management by enabling forward-looking adjustments based on expected conditions [9]. Recent investigations have emphasized the value of machine learning (ML) methods [10] in refining predictive models for energy use and enhancing

This is an open access article distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License (CC BY-NC-SA 4.0), which permits non-commercial use, distribution, reproduction, and adaptation in any medium or format, provided that appropriate credit is given to the original author(s) and source, a link to the license is provided, and any changes made are indicated. Adapted material must be distributed under the same license.

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

energy distribution networks [11, 12]. Progress in blockchain systems is also being considered to enhance the security and transparency of energy exchanges in smart grids [13]. A major concern in control systems is oscillations, which can lead to energy waste, operational instability, and discomfort for users [14]. Inadequately adjusted MPC algorithms, along with other control methods, often generate oscillatory responses, leading to unnecessary energy use as the system repeatedly corrects deviations from targets [15]. In HVAC systems, such oscillations can cause temperature swings that compromise thermal comfort and accelerate equipment deterioration [16]. Inaccuracies in models [14] and predictions [17], especially regarding external temperatures, often generate comfort problems. Incorporating disturbance modelling into MPC supports offset-free regulation and strengthens system stability by accounting for unforeseen environmental shifts [15]. Robust MPC approaches, such as adaptive robust MPC, were created to address uncertainties while guaranteeing constraints are met [14], for instance by factoring in flexibilities from source-network-load [18].

Various methods have been proposed to minimize oscillations in control systems. One technique involves thoughtfully adjusting weighting factors in MPC goal functions. In addition, applying limits on how quickly control inputs can vary prevents sudden actions [14]. Employing convex quadratic constraints enables capping overall energy consumption across multiple actuators, thereby restricting the system's dynamic responses within predefined bounds [15]. Robust MPC approaches strengthen closed-loop stability via bounded disturbance representations, delivering dependable performance even in the presence of uncertainties in weather and energy demand [14, 15]. Moreover, state estimation tools such as Kalman filters are essential for improving control accuracy by removing noise and addressing data gaps [19]. Multi-level control structures with primary, secondary, and tertiary layers provide an additional useful approach for managing energy distribution and mitigating oscillations in microgrids [19, 20]. Such tiered designs align system parameters, improve energy distribution, and lessen variations via an organized control setup [19]. Proportional Integral Derivative (PID), fuzzy logic, and robust control approaches have similarly been studied to raise stability and performance in energy management systems [19].

Including weather forecasts in energy management systems delivers major advantages, such as better-optimized consumption, more accurate load predictions, higher reliability, and lower operating costs [21]. This inclusion is also key to reinforcing energy system resilience toward weather-related uncertainties. In addition, weather-guided energy operations support the joint optimization of production, storage, and use — essential for meeting sustainability targets in today's power networks [22]. In microgrid applications, weather forecasts support smarter use of energy storage, load shifting, and demand response measures, thereby increasing efficiency and reducing carbon emissions [23].

By modeling different weather scenarios, energy systems become better prepared for changes in renewable supply, helping secure steady, dependable electricity delivery. Cutting-edge AI-powered forecasting tools can enhance weather prediction precision and enable more accurate energy management [24]. Research shows that brief load forecasts incorporating temperature and other weather factors greatly refine demand estimates and support oversight of the low-voltage grid [25]. Real-time online weather prediction approaches using current meteorological inputs enhance the dependability of predictive control models, particularly when paired with localized ML models for near-term estimates [15]. Still, forecast uncertainties remain a notable obstacle, necessitating robust scheduling methods that leverage source, network and load flexibilities [18].

Predictive control methods are widely used for energy management in buildings and microgrids, especially for refining HVAC operations and integrating renewable sources [26]. When weather forecasts are incorporated, these models improve energy efficiency and reduce costs without sacrificing indoor comfort [27]. A range of prediction methods — statistical, artificial intelligence, and combined — are used to anticipate energy demand and renewable output [28]. MPC has shown encouraging outcomes in managing runtime for air-source heat pumps and comparably high-energy equipment [9, 27]. Techniques for multi-step predictions based on meteorological information can also improve the precision of predictive control [29].

Numerous investigations assume ideal forecast accuracy, which is rarely realistic in practice [30]. In addition, most existing methods rely on model-driven techniques that use weather forecasts for load estimation. This presents difficulties

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

in intricate urban settings, particularly during extreme weather, when uncertainties intensify across multiple scales. Given that climate change brings previously unseen extremes, achieving optimal control is becoming increasingly difficult. Bridging this shortfall with conventional modeling requires more advanced simulation tools and adaptive transfer learning methods to increase robustness and real-world usefulness.

Reinforcement learning (RL) has exhibited significant promise for optimizing energy management in buildings connected to smart grids. This allows structures to independently adjust their energy use in response to varying surrounding conditions [31]. Approaches grounded in RL continuously refine control policies based on current information, thereby strengthening decision-making [32]. RL also enables predictive scheduling [33] and maintenance and fault detection within building systems, thereby supporting greater energy conservation and operational reliability [34]. Still, standard RL models commonly struggle to handle environments that shift over time, leading to instability and lower efficiency — especially amid major weather fluctuations [35]. In previous efforts, the author formulated RL-driven energy management strategies [36, 37]. The latest among them applies Adaptive Reinforcement Learning (ARL) specifically for Energy Management (EM) and is termed ARLEM [37]. ARLEM functions as a decentralized method for energy oversight. Here, each agent's energy output is governed by a shared flexibility signal and an individual online model-free ARL mechanism, ensuring no external data exchange with the agent. Research indicates that ARLEM substantially improves both energy efficiency and system resilience across urban settings, particularly in severe climate conditions, while safeguarding peak user privacy and minimizing processing requirements [37].

Systems based on RL that shape policies through accumulated experience can struggle to adapt to unfamiliar surroundings. ARL contributes by increasing adaptability and responsiveness in these dynamic cases. A frequent challenge across such systems — and across wider AI-supported control techniques — concerns instability or swings in operational effectiveness during rare events such as intense weather episodes. The bulk of current solutions for curbing these swings involve running predictive models and projecting near-term states with supplementary simulations. Though this proves useful in many cases, it increases the overall complexity of the control setup and can become unfeasible, particularly as both the size and number of agents grow.

This study incorporates weather forecasts directly into ARLEM to assess their influence on control outcomes, with a special focus on reducing oscillations. Beyond that, it examines whether these forecasts can also improve building energy performance and overall climate resilience within the energy grid. To achieve this, ARLEM underwent enhancements that embed two distinct signals into the decision framework: the network's flexibility signal and the weather forecast signal. This updated iteration is designated pARLEM, with the 'p' highlighting its predictive features. Testing covers four forecast horizons — 3, 6, 12 and 24 h — to determine the effect of varying time spans on the predictive model's results. On top of this, pARLEM periodically revises its policy at fixed intervals, swapping out underperforming actions for superior alternatives according to three defined selection rules. Evaluation of pARLEM targets a standard urban neighbourhood in Stockholm containing 24 buildings of four different kinds, set in a future winter span (2040–2069) that includes two cold waves and draws on detailed future climate projections [38].

Theory

In Reinforcement Learning (RL), an agent interacts with its environment to learn optimal actions through repeated cycles of actions and rewards [39]. At every step or episode, it receives feedback in the form of rewards and sharpens its policy to maximize long-term expected gains [40]. RL methods generally fall into model-based and model-free groups [35]. Model-free variants rely solely on immediate engagements and avoid creating a comprehensive environmental representation [41, 42].

Adaptive reinforcement learning (ARL) is a branch of RL in which agents continually adjust their behaviors and policies to adapt to environmental shifts [32]. The persistent learning aspect suits it to settings that are unstable and ever-changing [35], including multifaceted energy grids featuring mixed consumer and generator types, particularly during extreme climate episodes [43, 44]. ARL achieves adaptation through diverse channels, such as modified reward structures [45,

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

46], action planning modules [47, 48], alterations to network designs [49], or evolving policy mechanisms [50, 51]. On-the-spot responses to surprise developments heighten durability and versatility [52]. That said, this process normally draws on historical data while projecting ahead, which heightens demands on both simulation and choice mechanisms.

Earlier research introduced ARLEM [37] as an online, model-free ARL technique tailored for decentralized energy handling that operates without any prior knowledge of system dynamics. Online learning features continuous policy refinement that evolves as conditions change [53]. Agents refrain from delaying until large data collections build up [54]. Within ARLEM, policies receive partial refreshes at set intervals so that more suitable actions align with shifting conditions. This adaptation works by revising the policy and replacing the weakest actions with stronger options after a defined day count has passed. Updating methods vary depending on the interval length and the number of actions that need to be replaced. Here, updates occur after 1, 2, 3, 7, 15, or 30 days, with either 5 (5P) or 10 (10P) actions being swapped. Three action-selection rules are also applied: nearest actions (OneStep), two nearest actions (Two Step), or unrestricted choice (Jump). The initial pair aims to block sudden controller shifts. This online model-free ARL setup lets every agent steadily develop an ideal policy through experimentation. Rewards are computed per time step for each agent, balancing energy reduction and indoor comfort. Satisfaction on both yields a reward of 1 at time t; energy savings alone give 0.5; and comfort alone or total failure results in 0. Priority goes to actions scoring 1 during policy construction.

The state appears as the flexibility signal, expressed as an integer from 0 to 5. Agents in ARLEM rely on this signal to decode the environment and pinpoint the existing state. The signal is calculated by contrasting the current energy demand at time t with a benchmark demand for that same moment. Users select and customize the benchmark to fit specific control aims. In the present study, benchmarks reflect energy demands from matching daily periods under normal and severe weather using standard controls (absent ARLEM). A reward of 1 occurs when ARLEM-managed demand equals or stays under the benchmark for that weather type. A signal value of 0 indicates demand meets or falls below the benchmark, while values 1 through 5 indicate increasing demand intensity. The grid (such as a Distribution System Operator or DSO) broadcasts the flexibility signal to every agent at 15-minute intervals. At the highest level of 5 — denoting pressing demand reduction needs — actions scoring 0.5 also enter consideration. Over time, agents compile libraries of top actions to build their policies. This work tests two library dimensions: 10 or 24 action sets, labeled L10 and L24. Initial actions are random, but learning progressively favors stronger policy choices. A 10% exploration (randomness) rate continues to encourage discovery within the algorithm after policy establishment. Additional ARLEM details can be found in Nik and Javanroodi [37].

The current study augments ARLEM by adding a forward-looking dimension by integrating weather forecasts into decision-making, yielding pARLEM. Accordingly, the weather forecast signal is delivered to each agent concurrently with the flexibility signal, maintaining a 15-minute cadence **(Figure 1A)**. Only outdoor temperature forecasts serve as the weather signal here, given their dominant role in driving heating needs. Forecast inaccuracies chiefly hinder proactive policy tuning aimed at easing peak loads and upholding comfort. Real-world conditions show that forecasts become more reliable at shorter lead times (e.g., 3–6 h), supporting rapid policy adjustments and offsetting the drawbacks of errors. Though synthesized future weather inputs are used and forecast errors are omitted from the review (yet future climate variability is factored in), the four horizons of 3, 6, 12 and 24 h are tested. This probes interactions between forecast range (level of readiness) and refresh rate in overall system outcomes.

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

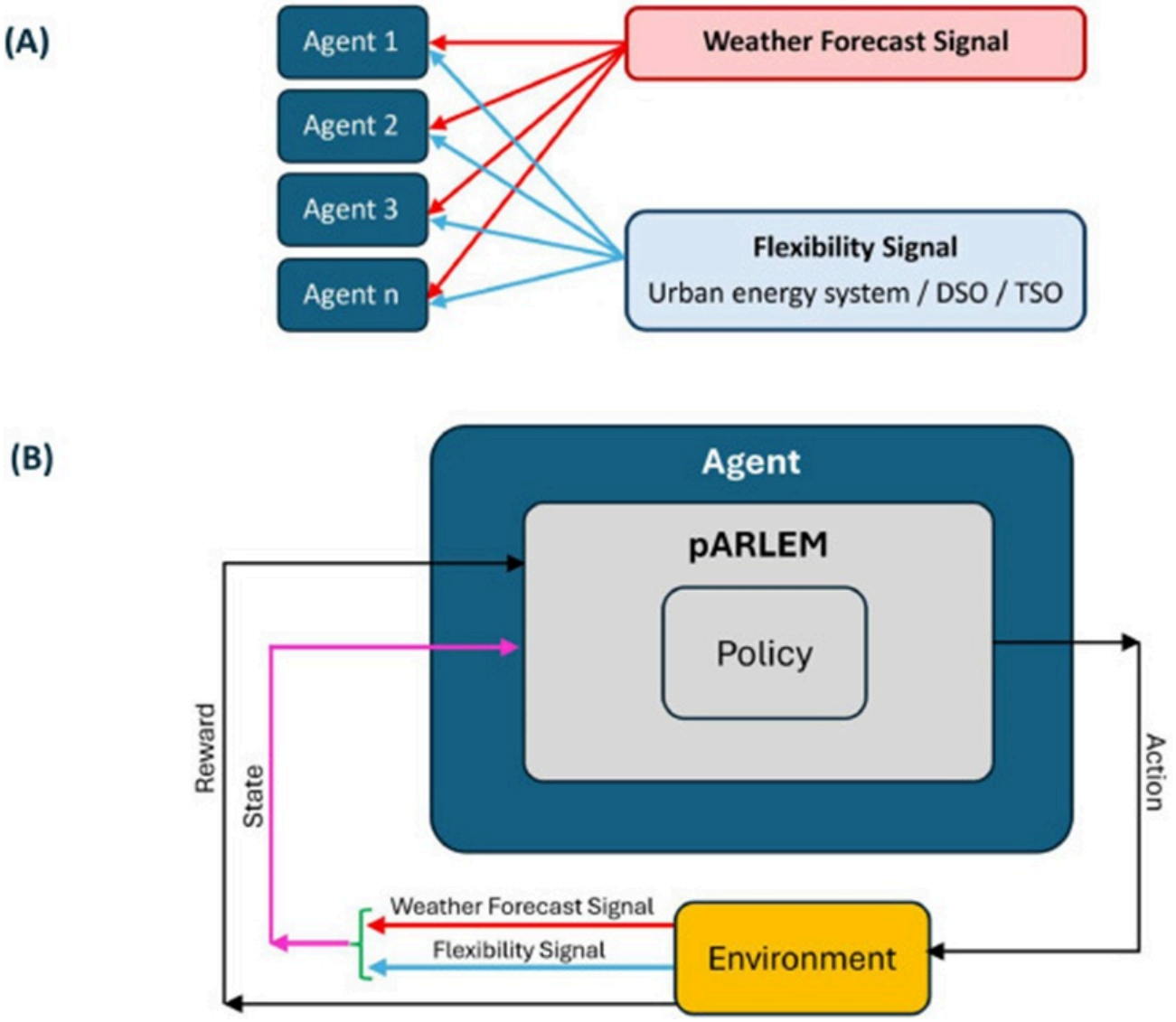

Figure 1. A) With the addition of weather forecasts to ARLEM, beyond the flexibility signal, the weather forecast signal gets delivered to every agent using the exact same timing as the flexibility signal (such as every 15 min). B) pARLEM employs online reward-based model-free ARL, so no predictive modelling occurs upon inclusion of the weather forecast signal. Instead, it serves as a state element that, paired with the flexibility signal, collectively establishes the state condition passed to the agent for use in choices.

It is worth emphasising that pARLEM builds on an online model-free ARL engine, so integrating the weather forecast signal requires no predictive modelling (for example, in contrast to MPC). The agent gets weather forecasts covering the chosen horizon, skipping any modelling or estimation of upcoming states. As supplementary environmental data, the weather forecast signal, along with the flexibility signal, fully characterizes the system state, informing the agent of current grid conditions and expected weather patterns. According to the intensity of impending weather, individual agents reinterpret the flexibility signal and activate the decision sequence carried out by the ARL engine within the agent **(Figure 1B).** High-level pseudocode for the pARLEM algorithm appears in **Figure 2**. Additionally, a summary of the primary adjustable parameters in the algorithm is supplied in **Table 1.**

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

```
Initialize control parameters and policy settings
Initialize ARLEM agent with default exploration-exploitation balance

For each control interval (e.g., 15 minutes):
    Observe current building state (e.g., indoor temperature, energy demand)
    Retrieve flexibility signal
    Retrieve weather forecast for selected horizon (e.g., 3h, 6h, 12h, 24h)
    Adjust flexibility signal

    If policy update interval is reached:
        Select policy update criterion (Jump, OneStep, TwoStep)
        Update policy set based on recent environmental observations

    Choose an action from the current policy set:
        - Balance exploration (10%) and exploitation (90%)

    Apply selected control action to the building systems
    Observe immediate reward based on:
        - Energy saving
        - Indoor comfort
    Update agent's internal experience (reinforcement signal)
End For
```

Figure 2. Pseudocode for the pARLEM algorithm.

Table 1. Overview of the major varying parameters in the pARLEM algorithm, including environmental signals, control strategies, and policy updating configurations.

Environmental factors		
Factor	**Details**	**Update frequency**
Flexibility signal	Scaled value ranging from 0–5 [-]	Every 15 min
Weather forecasts	Outdoor air temperature [°C] predicted for the following horizons: -3 h, -6 h, -12 h, -24 h	Every 15 min
Control factors		
Action selection criteria		

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

Method	**Description**	
Jump	The controller may choose any available action without restriction.	
OneStep	Selection is limited to the closest available actions (previous, current, next), with a maximum of three options.	
TwoStep	Selection is restricted to the two nearest actions, allowing up to five possible choices.	
Action library size		
Values	**Description**	
10 and 24	Total number of stored action sets in the library.	
Policy updating strategy		
Parameter	**Values**	**Description**
Updating period	1, 2, 3, 7, 15, and 30 days	Time interval after which actions are removed from the library.
Action removal	5 and 10	Number of actions eliminated from the library at each policy update step.

When energy systems contain many agents, those agents must harmonize their behaviors to fulfill the larger system's aims, a process termed "coordination" [55]. Key requirements include maintaining grid stability, complying with DSO signals, avoiding clashes or waste caused by unsynchronized behavior, and achieving superior combined outcomes rather than isolated local improvements. Often, this harmonization calls for advanced access or centralized oversight of network agents to devise plans for group excellence [56]. Nevertheless, pARLEM indirectly manages agent harmonization, given its decentralized nature and the lack of a specific harmonization procedure. Harmonization arises solely from the transmission of the flexibility signal, which conveys the full network status issued by, for instance, the DSO. Such a decentralized setup proves especially valuable for wide-ranging applications, where centralized harmonization becomes problematic due to privacy concerns and computational resource constraints. As the number of participating buildings increases, pARLEM's decentralized architecture scales effectively without the need for elaborate agent-to-agent discussions. This reduces communication demands and enhances system durability by eliminating reliance on single points of failure. Earlier research [36] revealed that this implicit harmonization style in RL-based energy management produces nearly ideal network-wide results, matching those from centrally harmonized techniques.

Within pARLEM, the weather signal functions as useful input (not a directive) and fosters natural harmonization among agents by offering a shared understanding of environmental shifts. Despite the absence of direct harmonization procedures, natural harmonization emerges as agents uniformly react to common external signals, enabling broad system adjustment without centralized direction. Essentially, even though no overt harmonization procedure applies, coordinated conduct emerges as agents independently respond to identical environmental stimuli. This decentralized approach enables pARLEM to operate effectively across diverse systems with distinct control priorities and operational patterns, thereby expanding its growth potential and practicality for real-world use, where thorough centralized harmonization is often impossible.

This work simulates and oversees a typical urban block in Stockholm, Sweden, using pARLEM. The block features 24 multi-zone residential buildings, with each one acting as a standalone agent. Four typical archetypes were formed via a multi-zone thermal modeling framework [57]. The archetypes belong to two building periods: before 1990 and after 2000. Six buildings represent each archetype inside the simulated urban block. Building details originate from the TABULA database [58]. A district heating system (DHS) delivers space heating. High-resolution building energy models (BEMs)

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

were built in EnergyPlus, and simulations were run in OpenStudio to enable detailed HVAC control. More information on the building modeling method is available in Nik and Javanroodi [37].

Here, pARLEM functions as a self-governing Building Energy Management System (BEMS) per building. A BEMS primarily manages the optimal regulation of adjustable components in a building over time. These often encompass various regulatory options, such as lighting, heating and cooling setpoints, and ventilation airflow volumes. Since pARLEM follows a model-free method, none of these possibilities or effects are directly modelled; they appear abstractly as control actions open to the agent. Through experimentation, the ARL agent builds understanding over time by testing different action mixes and spotting successful tactics against shifting surroundings. For example, the agent realizes that lowering the heating setpoint is more effective at reducing heating needs than reducing lighting levels. In this study, each agent uses control actions divided into four main categories based on existing literature [59]: (1) changing heating setpoints from 17–23 °C in 1 °C steps (7 actions), (2) managing ventilation rates per unit area from 0.4 to 1.2 l/s/m² in 0.4 l/s/m² steps (3 actions), (3) varying equipment and plug load demands from 3 to 9 W/m² in 3 W/m² steps (3 actions), and (4) changing lighting demand from 6 to 12 W/m² in 3 W/m² steps (3 actions). Overall, 189 unique action combinations exist for each building, called control or adaptation actions, that illustrate how agents modify operations amid evolving conditions. The expression ‘set(s) of actions’ denotes one particular mix from these four control choices.

Climate change causes more frequent and severe extreme weather events [6] that span multiple spatiotemporal scales [59] and impose additional burdens on energy infrastructure [60]. This work evaluates pARLEM performance under near-future climate conditions using 17 future climate scenarios from three Representative Concentration Pathways (RCPs) and five global climate models (GCMs) for 2040–2069, drawn from the ‘Coupled Model Intercomparison Project 5’ (CMIP5). In total, 510 (17 × 30) years of data created representative future weather datasets, covering the Typical Downscaled Year (TDY) and Extreme Cold Year (ECY) utilized here [38, 61]. TDY captures the most standard weather patterns from all examined years, while ECY records the coldest months. To simulate a winter with two cold waves, the TDY dataset incorporates ECY data across two time periods. The first cold wave starts on the 45th winter day and lasts about 6 days, whereas the second begins on the 72nd day and lasts about 10 days. The distributions of outdoor temperatures for standard winter conditions and extreme cold events are shown in **Figure 3.**

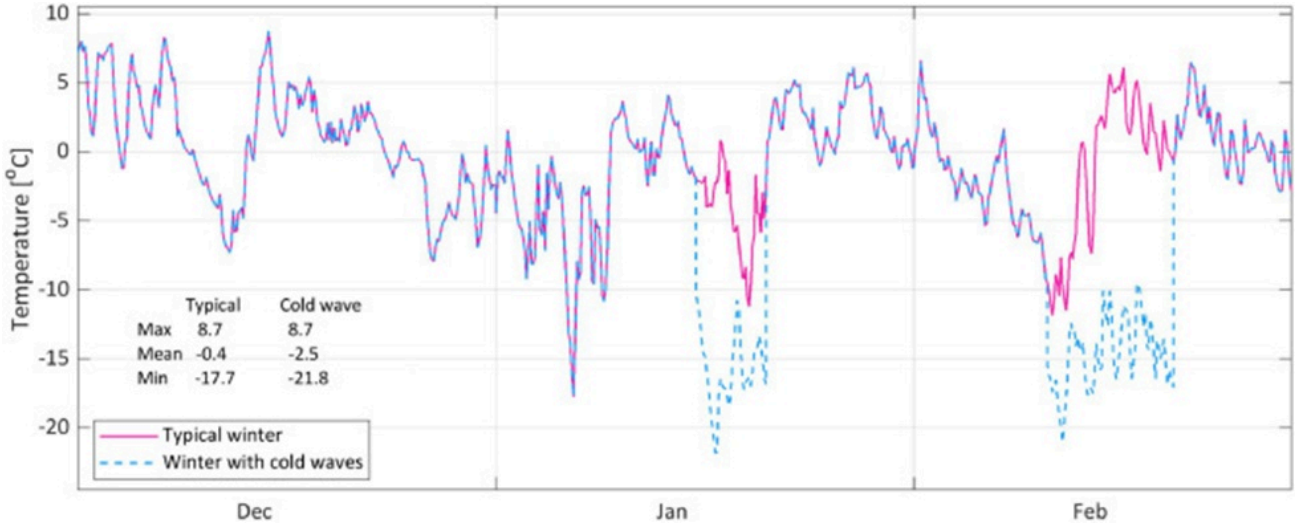

Figure 3. Outdoor temperature in Stockholm during 2040–2069 for a typical winter and a winter with two cold waves. The typical winter data are the typical winter conditions from 17 future climate scenarios. For winter with cold waves, two sets of extreme cold days are replaced using similar climate scenarios.

Results and Discussion

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

As stated before, several control approaches are being tested in this study. **Figure 4** presents heating setpoints for all buildings (grey lines) along with their mean (black line) under the two control rules, OneStep and TwoStep, when ARLEM governs building operations, both with and without weather forecasts for the next 3, 6, 12 and 24 h. In this instance, the ARLEM library includes 10 sets of actions (L10) with policy revision of 10 (P10) spanning 30 days (D30). As apparent in **Figure 4**, adding weather forecasts produces fewer controller swings, above all during cold waves highlighted by vertical dashed lines.

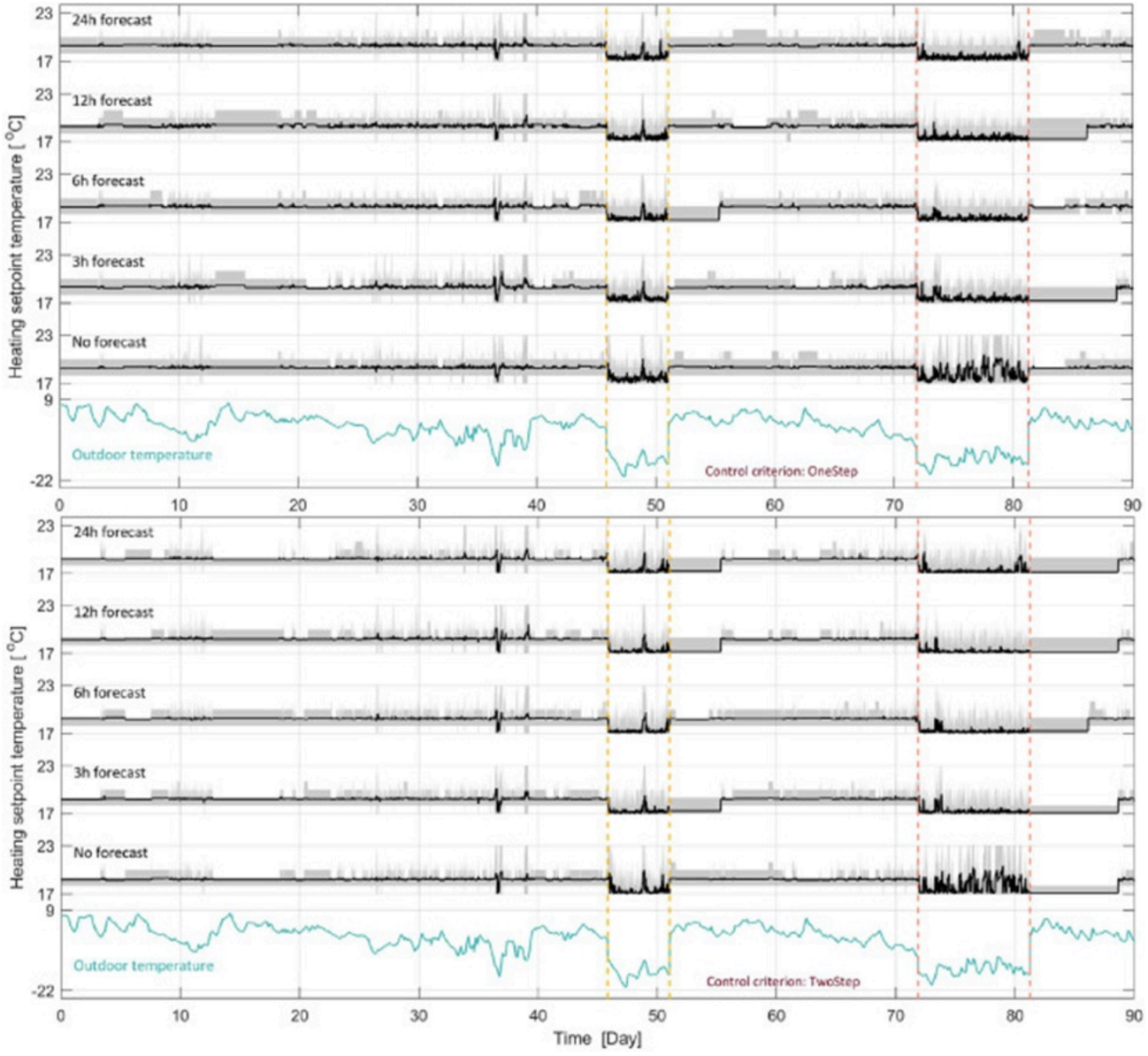

Figure 4. Heating setpoints in the buildings in Stockholm during winter with two cold waves (marked with dashed lines) for pARLEM-L10-P10-D30 control algorithm with OneStep (top) and TwoStep (bottom) control criteria and different forecast schemes. The black line shows the average in all the buildings and the grey shadow represents all the single values. The cyan line shows the outdoor temperature in Stockholm during the period considered.

Effects from embedding weather forecasts in choices were studied for every pARLEM version here, and **Figure 5** supplies a broad view of heating setpoints on ordinary winter days plus cold waves (ARLEM missing weather forecasts marked as ‘None’ on the x-axis). pARLEM outcomes are measured using the L10 and L24 libraries, with policies refreshed by discarding 5 or 10 action sets at intervals of 1, 2, 3, 7, 15, or 30 days. Two horizontal dashed lines in the

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

graphs (L10/L24 – No Policy Update) show the figures when neither policy refreshes nor weather forecasts apply. Starting from the top, the initial graph group shows mean heating setpoints on typical winter days in Stockholm. In contrast, the following group contrasts the variability of those setpoints (shifting every 15 min) across the entire span. The third and fourth graph groups display matching contrasts amid cold waves (or extreme cold days).

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

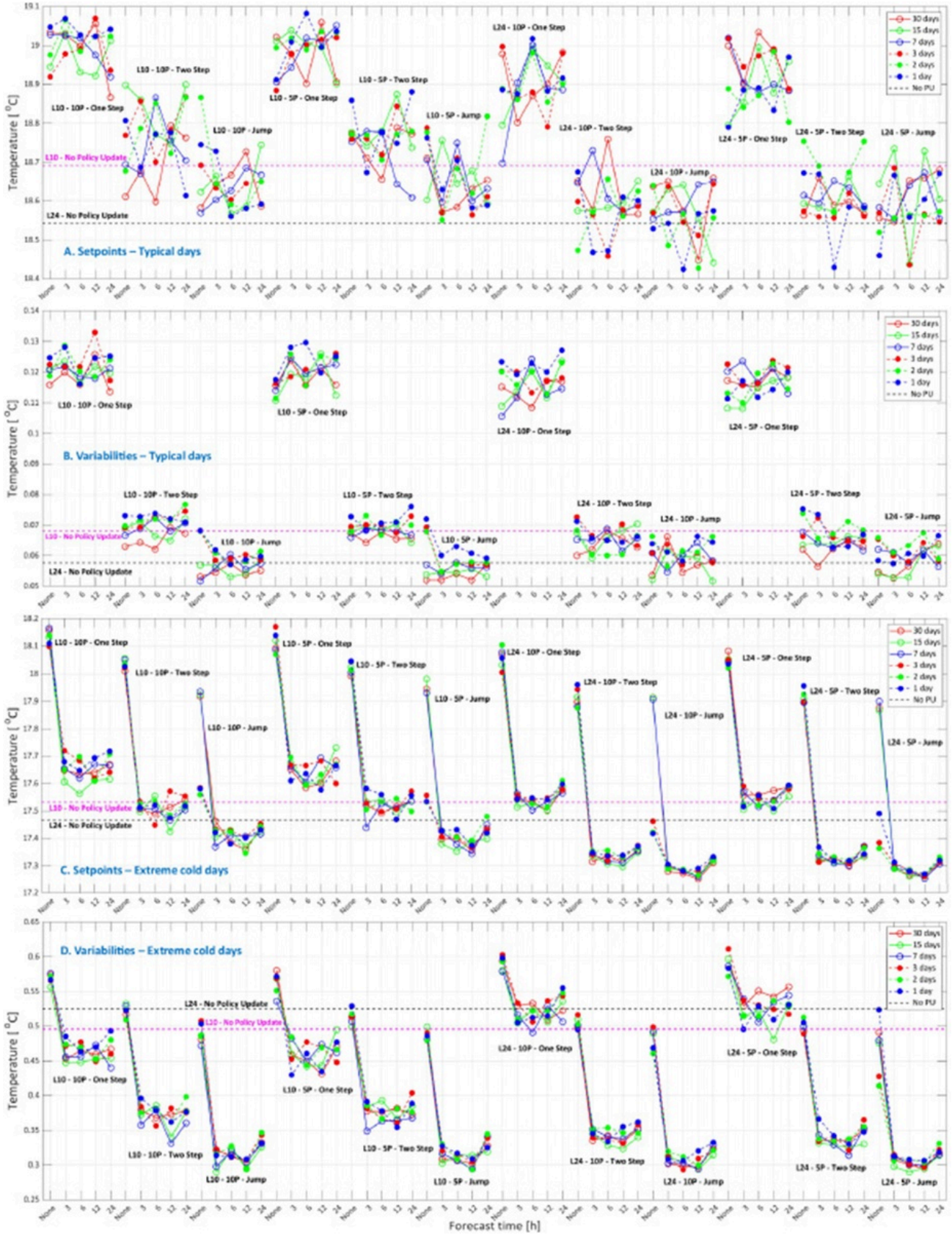

Figure 5. Average heating setpoints and their variabilities in the buildings in Stockholm during typical and extreme cold winter days for different control strategies of pARLEM.

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

Broadly, constraining the controller to choose only the closest action (OneStep) yields warmer setpoint temperatures, aligning with the prior investigation [37]. However, this creates wider swings in both controller and heating setpoint. Constraining to the two closest actions (TwoStep) or removing limits entirely (Jump) brings swings down to levels equal to or beneath those without policy refresh (the horizontal dashed lines). Further stability gains for the controller come from incorporating weather forecasts into the decision-making process. This approach works especially well during cold waves; all pARLEM versions lower heating setpoints by about 0.4 °C on average compared with no-forecast situations (marked 'None'), while reducing control swings by nearly 40 %. Among the forecast horizons, the 12-hour (12 h) forecast yields the best average performance. Overall, including weather forecasts greatly improves controller stability during cold waves, regardless of the chosen range.

Consistent management of building ventilation systems and continued reduction of their swings remain vital. **Figure 6** depicts changes in ventilation rates for the same situations shown in **Figure 4**. The black solid lines (mean ventilation rates across all buildings) show smaller swings under the TwoStep control rule. Across both control rules, adding weather forecasts to decisions markedly lowers ventilation rate swings. This is shown and verified for all control approaches in **Figure 7**. Whether on ordinary or extreme cold winter days, moving from OneStep to TwoStep and then to Jump control rules results in lower ventilation rates and reduced variations. Under OneStep, expanding the library (L24) increases variation by about 50% compared to no policy refresh. As noted previously, policy refresh tends to heighten swings inside the control system. This gets strongly mitigated under TwoStep and Jump once weather forecasts join, particularly on extreme cold days (**Figure 7**-bottom, Variabilities – Extreme cold days), achieving levels as much as 40 % below those without policy refresh.

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

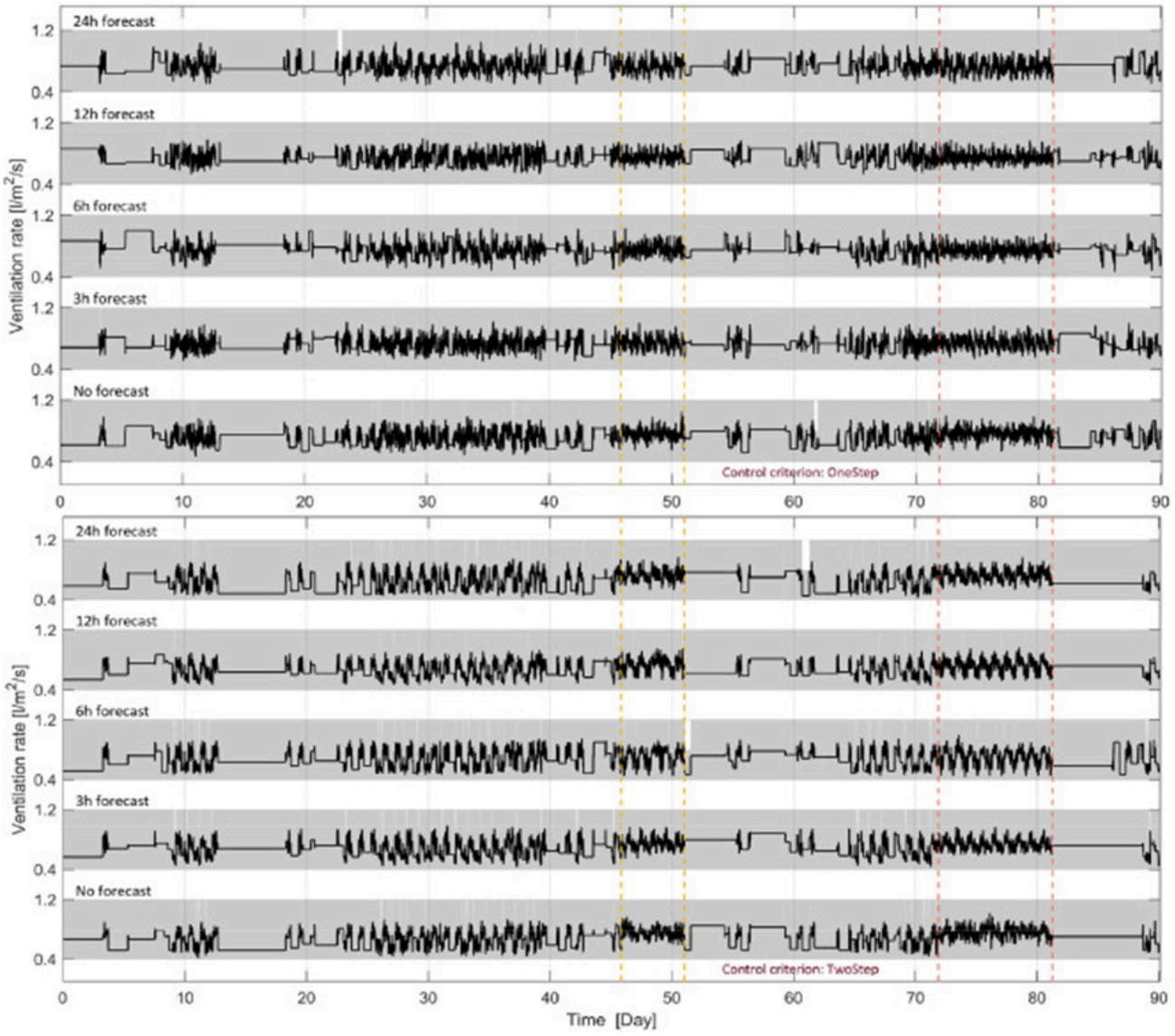

Figure 6. Ventilation rates in the buildings in Stockholm during winter with two cold waves (marked with dashed lines) for pARLEM-L10-P10-D30 control algorithm with OneStep (top) and TwoStep (bottom) control criteria and different forecast schemes. The black line shows the average in all the buildings and the grey shadow represents all the single values.

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

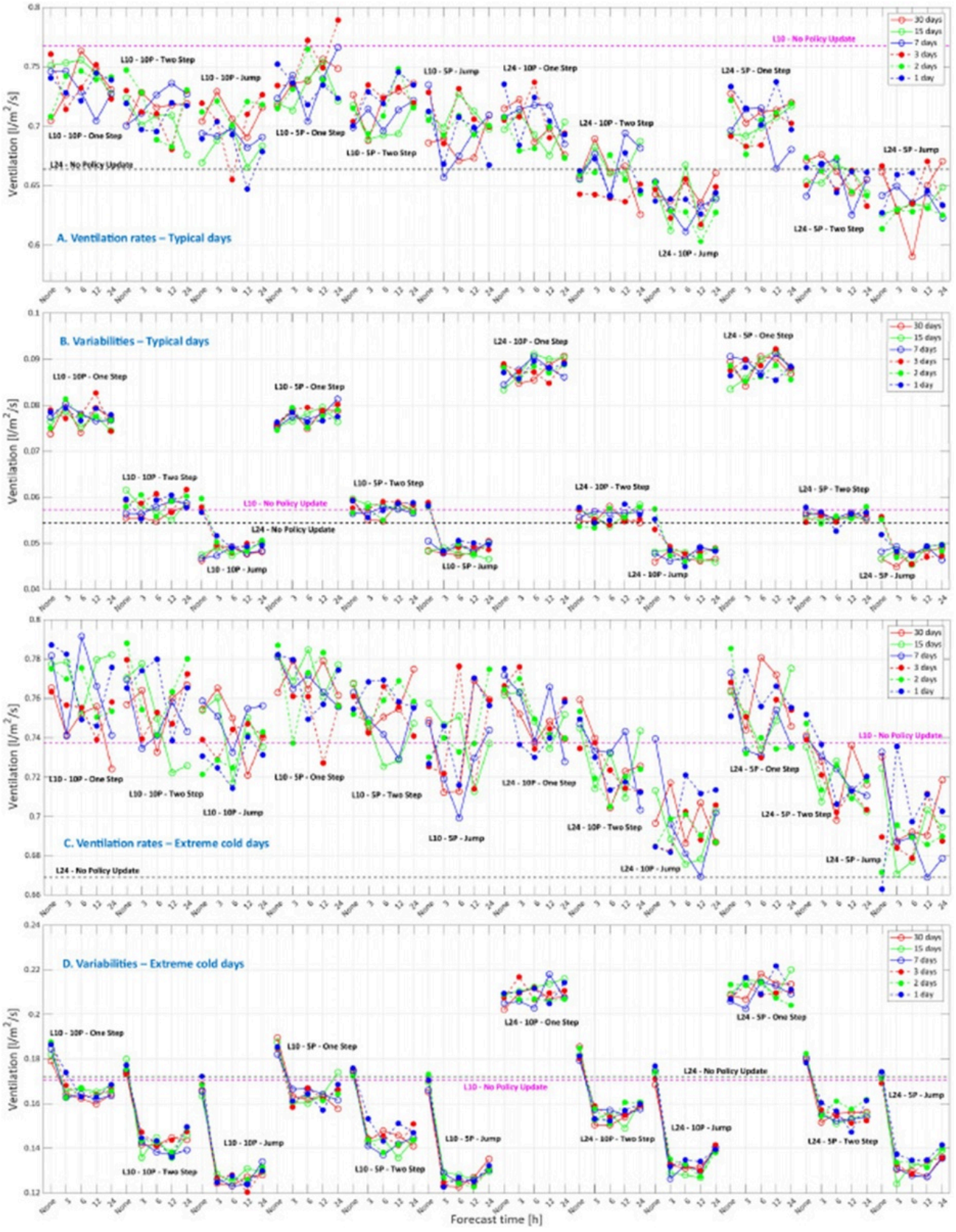

Figure 7. Average ventilation rates along with their variability levels in Stockholm buildings during standard and severe cold winter periods across various pARLEM control approaches.

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

Heating demand curves are displayed in **Figure 8** for configurations comparable to those shown in **Figure 4** and **Figure 6.** Predictably, heating demand grows during cold waves. Overall heating demands for each control option are shown in **Figure 9** for normal and intense cold winter periods. Adding weather forecasts to the choice mechanism has only a minor effect on heating demand during standard winter periods. Nevertheless, a substantial influence is evident during cold waves, as shown in **Figure 9** (bottom). The inclusion of weather forecasts can decrease hourly heating demand by nearly 5%, at times falling below levels observed in situations lacking policy refresh (illustrated by the black and magenta dashed lines). Discrepancies among forecast horizons in the ARLEM versions stay comparatively small amid cold waves. Stated another way, the presence of weather forecasts, irrespective of horizon duration, aids total energy conservation.

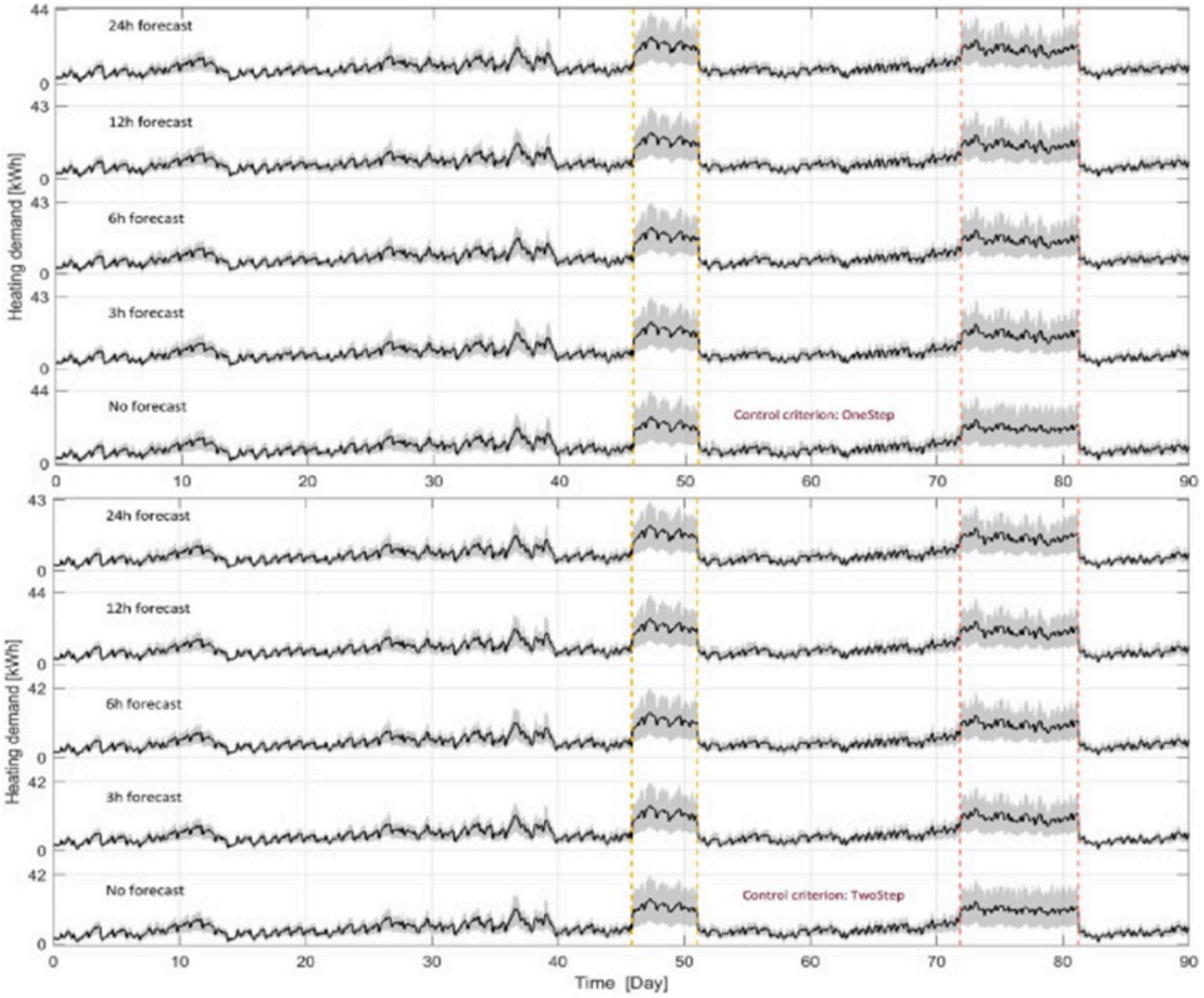

Figure 8. Heating demand in the buildings in Stockholm during winter with two cold waves (marked with dashed lines) for pARLEM-L10-P10-D30 control algorithm with OneStep (top) and TwoStep (bottom) control criteria and different forecast schemes. The black line shows the average in all the buildings and the grey shadow represents all the single values.

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

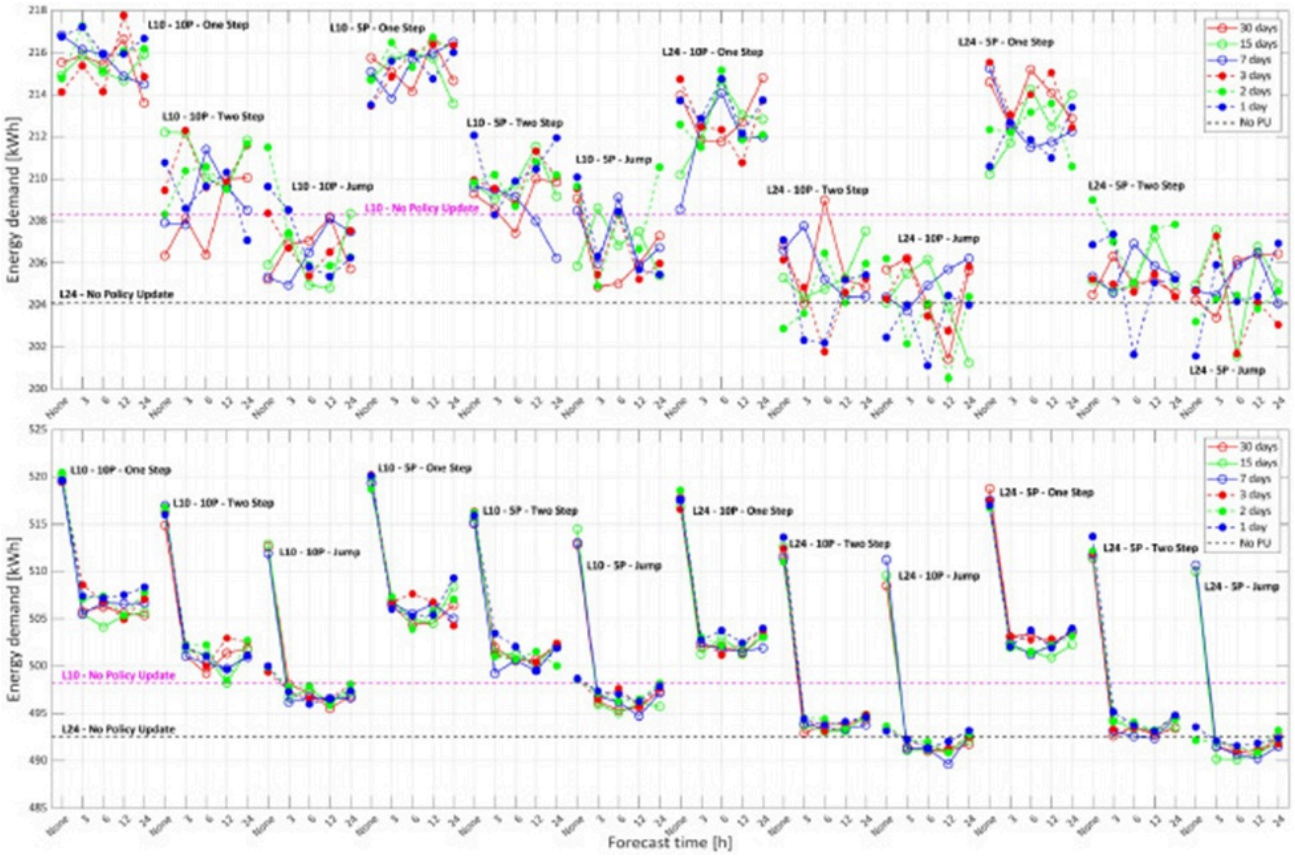

Figure 9. Average heating demand for different control strategies of pARLEM during typical (top) and extreme cold (bottom) winter days in Stockholm.

Additional analysis of the embedding of weather forecasts within choice mechanisms is provided in **Figure 10,** which presents the 98th, 90th and 80th percentiles of heating demand during cold waves across all configurations. The most striking observation concerns the control rules: enabling agents to choose actions past the closest options heightens their adaptability for cutting heating demand. Nonetheless, this could affect indoor comfort, as previously reported [37]. Adding weather forecasts helps reduce peak heating demand during cold waves. Yet, the benefit remains subtler than observed in **Figure 9**. In some instances, notably those using smaller action libraries (L10), adding weather forecasts could raise peak demand. Utilizing an expanded action library addresses this challenge, as shown for L24 instances in **Figure 10.**

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

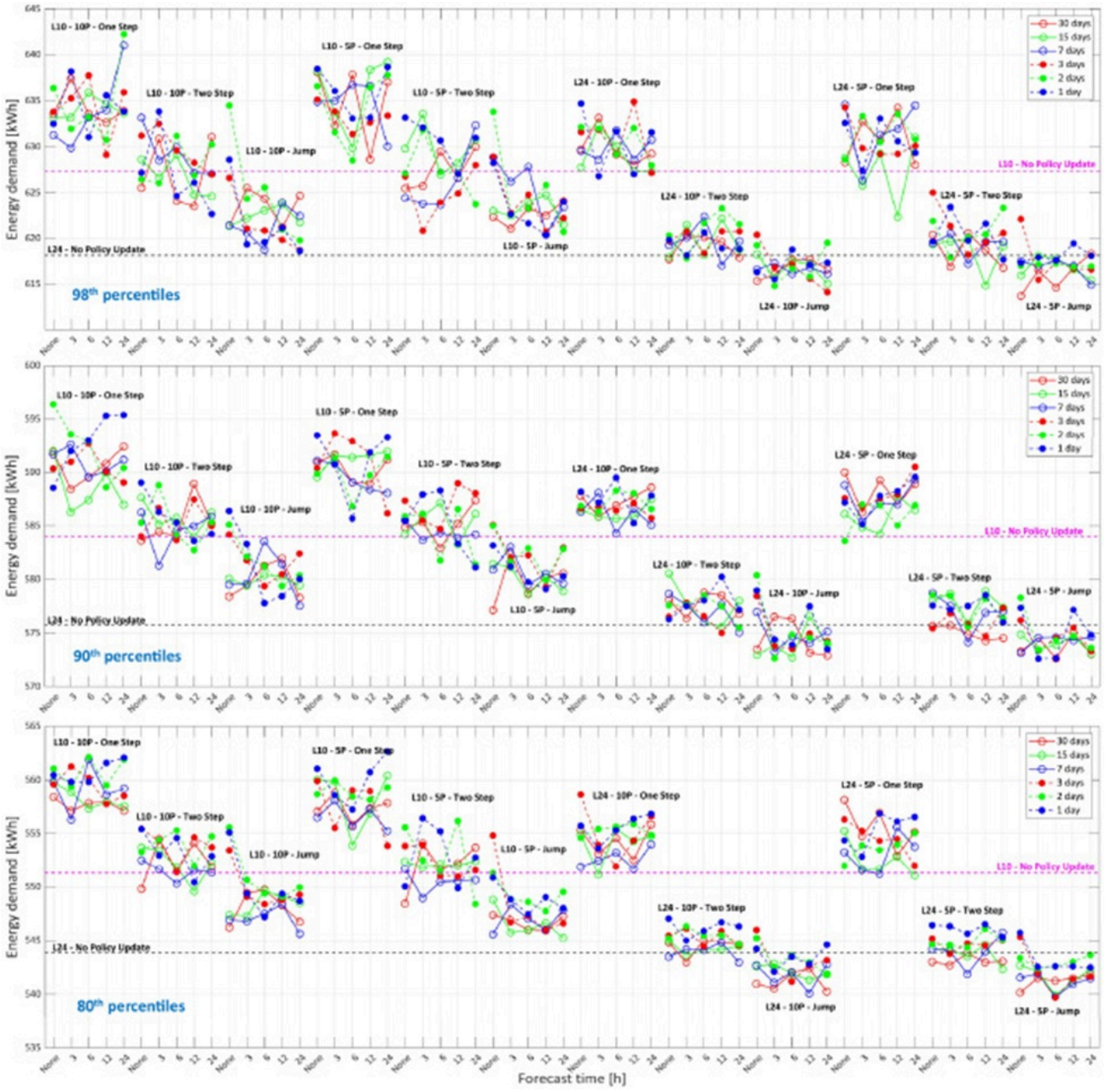

Figure 10. The 98th (top), 90th (middle) and 80th (bottom) percentiles of heating demand for different control strategies of pARLEM during the cold waves in Stockholm.

Although establishing a universal forecast error threshold remains challenging due to its connection to regional climate fluctuations and system operations, practical energy management techniques like pARLEM are typically expected to perform reliably within typical short-term forecast errors of ±1–2 °C. pARLEM's framework, marked by frequent policy refreshes and adaptable selection processes, automatically mitigates the harmful effects of these variations, as evidenced by consistent system performance across the evaluated forecast horizons.

Notably, the non-refreshing ARLEM version that omits weather forecast inclusion, shown by straight dashed lines in **Figures 9 and 10** effectively oversee building energy demand during both standard and extreme cold periods. It attains significant cuts in both mean and peak heating demands, particularly in the L24 case (the black dashed line). Combining policy refreshes with weather forecast inclusion can generate elevated energy consumption above the two fundamental

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

cases while providing the counterbalancing gain of superior indoor comfort. Importantly, energy reduction is not the exclusive target in this study; achieving more stable controller behavior is another central objective, accomplished by embedding weather forecasts into the decision algorithm, as evidenced in previous illustrations.

Conclusion

This investigation delved into embedding weather forecasts into Adaptive Reinforcement Learning for Energy Management (ARLEM) by creating a forward-looking variant, pARLEM. The purpose was to enhance the stability and effectiveness of energy control systems, particularly in extreme weather conditions. pARLEM effectiveness underwent evaluation using four forecast horizons – 3, 6, 12, and 24 h – together with assorted adaptive policy methods. These handled energy operations for a typical urban neighbourhood in Stockholm, featuring 24 buildings across four categories, under anticipated winter conditions for 2040–2069, including two cold waves.

Outcomes reveal that embedding weather forecasts into choice mechanisms, regardless of horizon duration, substantially improves system stability and energy performance. It further reduces sudden fluctuations in heating and ventilation regulation, particularly during cold waves. On average, weather forecast embedding reduces heating setpoints by roughly 0.4 °C and trims controller variations by approximately 40%, resulting in more fluid functioning. According to the outcomes, refreshing the policy alongside weather-forecast embedding could modestly increase energy demand relative to the non-forward-looking, non-adaptive reference point, while enhancing control system responsiveness and fostering steadier functioning. Concerning heating demand, distinctions between pARLEM and ARLEM remain limited on standard winter days. Still, pARLEM successfully reduces peak demand during cold waves, especially in versions with larger action libraries (i.e., those containing 24 action sets). This emphasizes the importance of pairing action-space versatility with forward-looking information. Additionally, pARLEM lowers mean heating demand by around 5% during cold waves, indicating that weather-forecast embedding across all horizons promotes additional energy conservation. Eliminating restrictions so agents can choose beyond the nearest actions amplifies their capacity to address abrupt environmental shifts and support peak-demand cuts.

Embedding weather forecasts into an ARL-powered control structure enables pARLEM to improve anticipatory precision and operational stability markedly. It delivers a durable yet exceptionally versatile structure for self-directed, decentralized oversight of energy across buildings and urban zones. By adopting a model-free ARL method, pARLEM can perform the complex task of synchronized energy oversight in urban zones with minimal computational requirements for predictive regulation. This improves energy performance, protects user privacy, and delivers strong outcomes across diverse climates. Applying live weather forecasts for continuous policy adjustment facilitates optimal energy application, lessens operational instability, and advances climate resilience, effectiveness, and sustained viability.

Acknowledgements

In loving memory of my mother, whose enduring love shapes my journey and stays ever close.

Conflict of interest

None

Financial support

This work was supported by the Swedish Research council for Sustainable Development, Formas under CARFTCIRLEM project (ID: 2022–01120), and the Joint ERA-Net Call 2020 (MICall20) on digital transformation for green energy

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

transition under DigiCiti project (ID: 108807). The efforts of Dr. Kavan Javanroodi in conducting building simulations are highly appreciated.

Ethics statement

None

Received: 08 Nov 2021 Revised: 27 Dec 2021 Accepted: 14 Jan 2022
Published online: 15 June 2022

Rights and permissions

Open Access The author(s) retain copyright. This article is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. It may be shared and adapted for non-commercial purposes with appropriate attribution, an indication of changes, and distribution of adaptations under the same license. Third-party material may be subject to separate terms identified in its credit line. View the license at https://creativecommons.org/licenses/by-nc-sa/4.0/.

References

1. Javanroodi K, Nik VM. Interactions between extreme climate and urban morphology: investigating the evolution of extreme wind speeds from mesoscale to microscale. Urban Clim. 2020;31:100544.

2. Ma Y, Borrelli F, Hencey B, Coffey B, Bengea S, Haves P. Model predictive control for the operation of building cooling systems. IEEE Trans Control Syst Technol. 2012;20(3):796-803.

3. Kamel E, Sheikh S, Huang X. Data-driven predictive models for residential building energy use based on the segregation of heating and cooling days. Energy. 2020;206:118045.

4. Nengroo SH, Har D, Jeong H, Heo T, Lee S. Continuous variable quantum reinforcement learning for HVAC control and power management in residential building. Energy AI. 2025;21:100541.

5. Lokeshgupta B, Sivasubramani S. Multi-objective home energy management with battery energy storage systems. Sustain Cities Soc. 2019;47:101458.

6. Perera ATD, Nik VM, Chen D, Scartezzini JL, Hong T. Quantifying the impacts of climate change and extreme climate events on energy systems. Nat Energy. 2020;5(2):150-9.

7. Aste N, Manfren M, Marenzi G. Building Automation and Control Systems and performance optimization: a framework for analysis. Renew Sustain Energy Rev. 2017;75:313-30.

8. Nik VM, Perera ATD, Chen D. Towards climate resilient urban energy systems: a review. Natl Sci Rev. 2021;8(6):nwaa134.

9. Cox SJ, Kim D, Cho H, Mago P. Real time optimal control of district cooling system with thermal energy storage using neural networks. Appl Energy. 2019;238:466-80.

10. Perera ATD, Wickramasinghe PU, Nik VM, Scartezzini JL. Machine learning methods to assist energy system optimization. Appl Energy. 2019;243:191-205.

11. Hasan MK, Abdulkadir RA, Islam S, Gadekallu TR, Safie N. A review on machine learning techniques for secured cyber-physical systems in smart grid networks. Energy Rep. 2024;11:1268-90.

12. Berghout T, Benbouzid M, Muyeen SM. Machine learning for cybersecurity in smart grids: a comprehensive review-based study on methods, solutions, and prospects. Int J Crit Infrastruct Prot. 2022;38:100547.

13. Ahl A, Goto M, Yarime M, Tanaka K, Sagawa D. Challenges and opportunities of blockchain energy applications: interrelatedness among technological, economic, social, environmental, and institutional dimensions. Renew Sustain Energy Rev. 2022;166:112623.

14. Drgoňa J, Arroyo J, Cupeiro Figueroa I, Blum D, Arendt K, Kim D, et al. All you need to know about model predictive control for buildings. Annu Rev Control. 2020;50:190-232.

15. Yao Y, Shekhar DK. State of the art review on model predictive control (MPC) in heating ventilation and air-conditioning (HVAC) field. Build Environ. 2021;200:107952.

16. Taheri S, Hosseini P, Razban A. Model predictive control of heating, ventilation, and air conditioning (HVAC) systems: a state-of-the-art review. J Build Eng. 2022;60:105067.

17. Lv C, Yu H, Li P, Wang C, Xu X, Li S, et al. Model predictive control based robust scheduling of community integrated energy system with operational flexibility. Appl Energy. 2019;243:250-65.

18. Wang C, Jiao B, Guo L, Tian Z, Niu J, Li S. Robust scheduling of building energy system under uncertainty. Appl Energy. 2016;167:366-76.

19. Mbungu NT, Ismail AA, AlShabi M, Bansal RC, Elnady A, Hamid AK. Control and estimation techniques applied to smart microgrids: a review. Renew Sustain Energy Rev. 2023;179:113251.

20. Mendes PRC, Isorna LV, Bordons C, Normey-Rico JE. Energy management of an experimental microgrid coupled to a V2G system. J Power Sources. 2016;327:702-13.

21. Amin A, Mourshed M. Weather and climate data for energy applications. Renew Sustain Energy Rev. 2024;192:114247.

22. Agüera-Pérez A, Palomares-Salas JC, González de la Rosa JJ, Florencias-Oliveros O. Weather forecasts for microgrid energy management: review, discussion and recommendations. Appl Energy. 2018;228:265-78.

23. Genov E, Ruddick J, Bergmeir C, Vafaeipour M, Coosemans T, Garcia S, et al. Predict. Optimize. Revise. on forecast and policy stability in energy management systems. 2025. arXiv:2407.03368.

24. Troccoli A, Fuchs T, Boscolo R, Matt E, Bastani H. Coproduction of integrated weather and climate services to support net-zero energy transitions. 2024.

25. Haben S, Giasemidis G, Ziel F, Arora S. Short term load forecasting and the effect of temperature at the low voltage level. Int J Forecast. 2019;35(4):1469-84.

26. Kuboth S, Heberle F, König-Haagen A, Brüggemann D. Economic model predictive control of combined thermal and electric residential building energy systems. Appl Energy. 2019;240:372-85.

27. Killian M, Kozek M. Ten questions concerning model predictive control for energy efficient buildings. Build Environ. 2016;105:403-12.

28. Fontenot H, Dong B. Modeling and control of building-integrated microgrids for optimal energy management – A review. Appl Energy. 2019;254:113689.

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

29. Huang H, Chen L, Hu E. A neural network-based multi-zone modelling approach for predictive control system design in commercial buildings. Energy Build. 2015;97:86-97.

30. Langer L, Volling T. An optimal home energy management system for modulating heat pumps and photovoltaic systems. Appl Energy. 2020;278:115661.

31. Wei T, Wang Y, Zhu Q. Deep Reinforcement Learning for Building HVAC Control. In: Proceedings of the 54th Annual Design Automation Conference. New York: ACM; 2017. p. 1-6.

32. Eller L, Siafara LC, Sauter T. Adaptive control for building energy management using reinforcement learning. In: 2018 IEEE International Conference on Industrial Technology (ICIT). 2018. p. 1562-7.

33. Heidarykiany R, Ababei C. Advanced day-ahead scheduling of HVAC demand response control using novel strategy of Q-learning, model predictive control, and input convex neural networks. Energy AI. 2025;20:100509.

34. Guo G, Gong Y. Multi-microgrid energy management strategy based on multi-agent deep reinforcement learning with prioritized experience replay. Appl Sci. 2023;13(5):2865.

35. Sutton RS, Barto AG. Reinforcement learning: an introduction. 2nd ed. Cambridge (MA): MIT Press; 2018.

36. Nik VM, Hosseini M. CIRLEM: a synergic integration of collective intelligence and reinforcement learning in energy management for enhanced climate resilience and lightweight computation. Appl Energy. 2023;350:121785.

37. Nik VM, Javanroodi K. Adaptive reinforcement learning for energy management – a progressive approach to boost climate resilience and energy flexibility. Adv Appl Energy. 2025;17:100213.

38. Nik VM. Making energy simulation easier for future climate – Synthesizing typical and extreme weather data sets out of regional climate models (RCMs). Appl Energy. 2016;177:204-26.

39. Wang Z, Hong T. Reinforcement learning for building controls: the opportunities and challenges. Appl Energy. 2020;269:115036.

40. Levin E, Pieraccini R, Eckert W. Using Markov decision process for learning dialogue strategies. In: Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing. Vol.1. 1998. p. 201-4.

41. Nagy Z, Henze G, Dey S, Arroyo J, Helsen L, Zhang X, et al. Ten questions concerning reinforcement learning for building energy management. Build Environ. 2023;241:110435.

42. Gao C, Wang D. Comparative study of model-based and model-free reinforcement learning control performance in HVAC systems. J Build Eng. 2023;74:106852.

43. Fan C, Jiang X, Mostafavi A. Evaluating crisis perturbations on urban mobility using adaptive reinforcement learning. Sustain Cities Soc. 2021;75:103367.

44. Taboga V, Bellahsen A, Dagdougui H. An enhanced adaptivity of reinforcement learning-based temperature control in buildings using generalized training. IEEE Trans Emerg Top Comput Intell. 2022;6(2):255-66.

45. Place A. Adaptive Reinforcement Learning with LLM-augmented Reward Functions. [Date unknown].

46. Zheng S, Yin H, Chen T, Nguyen QVH, Chen W, Zhao L. DREAM: adaptive Reinforcement Learning based on Attention Mechanism for Temporal Knowledge Graph Reasoning. In: Proceedings of the 46th International ACM SIGIR Conference. New York: ACM; 2023. p. 1578-88.

47. Wang S, Zhang X, Wang C, Wu K, Li C, Dong D. DRLAR: a deep reinforcement learning-based adaptive routing framework for network-on-chips. Comput Netw. 2024;246:110419.

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press

48. Silva C, Andrade P, Ribeiro B, Santos B F. Adaptive reinforcement learning for task scheduling in aircraft maintenance. Sci Rep. 2023;13:16605.

49. Schmidgall S. Adaptive reinforcement learning through evolving self-modifying neural networks. In: Proceedings of the 2020 Genetic and Evolutionary Computation Conference Companion. New York: ACM; 2020. p. 89-90.

50. Liu YT, Singh N, Ahmad A. Adaptive Reinforcement Learning for Robot Control. 2024. arXiv:2404.18713.

51. Li X, Lee CG. ΔV-learning: an adaptive reinforcement learning algorithm for the optimal stopping problem. Expert Syst Appl. 2023;231:120702.

52. Bifet A, Gavaldà R. Adaptive Learning from Evolving Data Streams. In: Adams NM, Robardet C, Siebes A, Boulicaut J-F, editors. Advances in Intelligent Data Analysis VIII. Berlin: Springer; 2009. p. 249-60.

53. Mnih V, Kavukcuoglu K, Silver D, Rusu AA, Veness J, Bellemare MG, et al. Human-level control through deep reinforcement learning. Nature. 2015;518(7540):529-33.

54. Kober J, Bagnell JA, Peters J. Reinforcement learning in robotics: a survey. Int J Robot Res. 2013;32(11):1238-74.

55. Nawaz A, Zhou M, Wu J, Long C. A comprehensive review on energy management, demand response, and coordination schemes utilization in multi-microgrids network. Appl Energy. 2022;323:119596.

56. Qiu D, Ye Y, Papadaskalopoulos D, Strbac G. Scalable coordinated management of peer-to-peer energy trading: a multi-cluster deep reinforcement learning approach. Appl Energy. 2021;292:116940.

57. Javanroodi K, Mahdavinejad M, Nik VM. Impacts of urban morphology on reducing cooling load and increasing ventilation potential in hot-arid climate. Appl Energy. 2018;231:714-46.

58. Loga T, Stein B, Diefenbach N. TABULA building typologies in 20 European countries—making energy-related features of residential building stocks comparable. Energy Build. 2016;132:4-12.

59. Perera ATD, Javanroodi K, Mauree D, Nik VM, Florio P, Hong T, et al. Challenges resulting from urban density and climate change for the EU energy transition. Nat Energy. 2023;8(4):397-412.

60. Deng Z, Javanroodi K, Nik VM, Chen Y. Using urban building energy modeling to quantify the energy performance of residential buildings under climate change. Build Simul. 2023;16(9):1629-43.

61. Nik VM. [Licentiate thesis]. Gothenburg: Chalmers University of Technology; 2010.

© 2022 The Author(s). | Published by Institute of Smart Energy Systems Press