Prime Time for Model-Predictive Control? Assessing the Technical and Market Readiness of Advanced Controls in Buildings

Marco Pritoni, Donghun Kim, David Blum, Ettore Zanetti, Sang Woo Ham, Armando Casillas, Anand Krishnan Prakash, Lazlo Paul, Weiping Huang, Tao Yang, Afshan Qamar, Daniel Gerber, Jingjing Liu, Mary Ann Piette.
Lawrence Berkeley National Laboratory

ABSTRACT

Despite three decades of extensive research and field testing that have consistently validated the benefits of Model Predictive Control (MPC) in building applications, the technology has seen limited market adoption. This paper evaluates the readiness of MPC for widespread deployment, showcases recent demonstrations and field tests across diverse building types, including residential, small commercial, large commercial, and campus settings. Our results demonstrate that MPC can optimize system operations to achieve load shifting, minimize curtailment of on-site generation, and reduce energy costs by up to 80 %, while maintaining or improving occupant comfort. We also show that MPC can effectively control large assets, such as MW-sized thermal storage systems, and respond to dynamic pricing signals. However, achieving scale remains difficult due to labor-intensive workflows, reliance on a “PhD-in-the-loop” for MPC design and maintenance, susceptibility to fragile data infrastructure, and persistent workforce education and acceptance barriers. To bridge this gap, we outline a transition from bespoke, labor-intensive prototypes toward streamlined, segment-targeted deployment strategies that leverage model templates, semantic tools, and generative AI. By automating control configuration and reducing engineering effort, these recommendations provide a pathway for transforming successful research demonstrations into scalable, market-ready solutions for MPC-based controls.

Introduction and background

To maintain grid reliability and keep electricity affordable, California has set an ambitious target of achieving 7 gigawatts of load shift capacity by 2030 (CEC 2023), of which about 2 gigawatts is expected to come from dynamic-pricing-based load shift. Achieving this target requires a fundamental change in the building stock, complementing static, permanent energy efficiency with dynamic, responsive demand-side energy management. This approach needs a new generation of building control systems. These systems must automate responses to grid signals without compromising essential operations, ensuring that thermal comfort and other occupant services are maintained (Killiccote et al. 2014). As grid signals transition from infrequent demand response events to continuously varying prices and complementary emergency alerts, legacy controls are no longer adequate (Neukomm et al. 2019). While traditional controls typically rely on simple rules and fixed schedules to manage building systems, these static methods become ineffective at maintaining service levels and respond to dynamic grid conditions at the same time, as they are not effective at balancing conflicting goals (Drgoňa et al 2020).

Advanced control systems can utilize the structural thermal mass of a building, including concrete slabs, walls, and internal furnishings, to function as a thermal battery. They can *charge* the building by pre-cooling or pre-heating the interior during periods when electricity prices are low. Once the building structure reaches the target temperature, the control system can *ride* through expensive periods by significantly reducing or entirely disabling HVAC operations. This strategy allows the building to maintain internal thermal comfort solely through the slow release of stored energy from its thermal mass, partially decoupling the timing of electricity consumption from the delivery of cooling or heating services (Braun 2003; Chen et al. 2020).

Beyond thermal inertia of building structures, modern building controls provide the ability to orchestrate diverse customer-side resources, including active thermal storage, stationary batteries, electric vehicle (EV) charging, and other load-flexible end uses. These systems can strategically discharge stationary or EV battery storage to either supplement loads or export power to the grid during grid stress or high-price periods (Zhou et al 2019). Furthermore, advanced controls can automate the scheduling of non-essential appliances, such as dishwashers, laundry units, or pool pumps, to run during periods with low-price or abundant on-site generation (Nagpal et al. 2020).

What is MPC?

For three decades, academic research has explored several advanced control methodologies to allow buildings to provide load flexibility, ranging from fuzzy logic to reinforcement learning (Drgoňa et al., 2020). Among these, the most popular predictive algorithm in the literature is Model Predictive Control (MPC). MPC is a family of control methodologies characterized by the explicit use of a mathematical model to predict future behavior of a system and determine optimal control actions. This class of advanced controls is fundamentally distinct from traditional rule-based reactive strategies, such as the proportional-integral-derivative controllers, because it utilizes a forward-looking optimization paradigm rather than a backward-looking error correction mechanism (Shan et al. 2016). MPC works like a GPS for energy. Just as a navigator looks at traffic and weather ahead to find the best route, MPC looks at future weather and energy costs to find the best way to run a building. At each time step, it solves an optimization problem to balance objectives, such as minimizing energy cost or maximization of load shifted, against the constraint of maintaining occupant comfort (Figure 1). Because these forecasts are updated in real time, the system continuously recalculates its strategy to correct for unexpected changes, ensuring the building stays on the most optimal path (Drgoňa et al., 2020). A recent review found over 2,100 peer-reviewed papers published since 1997, investigating MPC's application to building HVAC systems. However, a much smaller subset of this body of research, representing only about 4% of the papers, has moved beyond simulation to validate the performance in real buildings (Khabbazi et al. 2025). Further, only a subset of these explore the use of MPC for load flexibility (Zanetti et al. 2025).

By presenting and discussing four field studies of MPC in response to dynamic pricing, this paper aims to answer the following questions:

- Is MPC effective in providing load flexibility across various building sectors?
- What are the challenges for MPC adoption and scale-up, and how can they be addressed?
- Are there alternative advanced control approaches to MPC ?

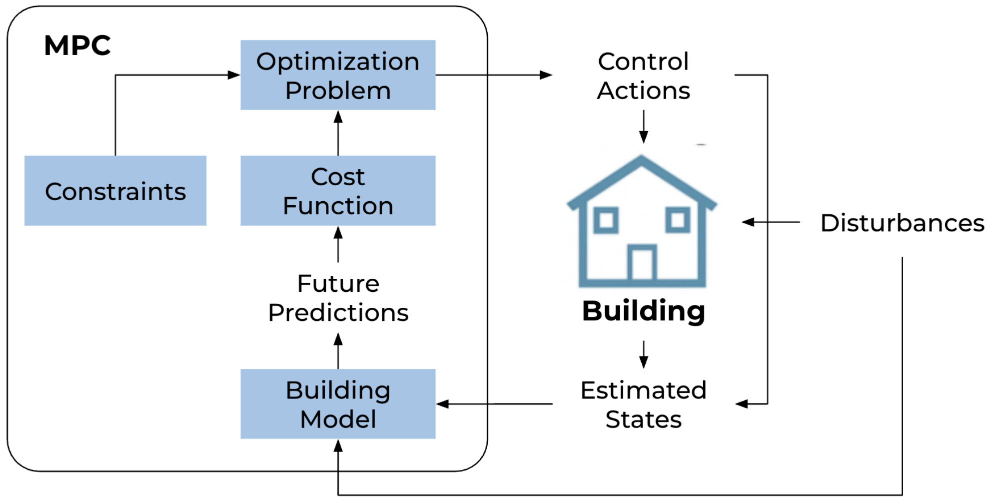

Figure 1: Conceptual operation of model predictive control (Modified from Drgoňa et al. 2020)

MPC field demonstrations characteristics and results

To evaluate the real-world effectiveness and maturity of various price-responsive technologies, the California Energy Commission-funded and Lawrence Berkeley National Laboratory-led California Load Flexibility Research and Deployment Hub (CalFlexHub) conducted field testing and demonstration of 18 automated dynamic price-responsive load-flexible technologies, including MPC (Piette et al 2022). Four of these demonstration projects representing MPC used in different building segments are summarized in Table 1. They encompass a wide range of applications, from whole-house control in single-family homes to HVAC optimization in small and large commercial buildings and even an entire district system on a university campus. These field projects were carried out between 2022 and 2025 across several locations in California. All tests were performed using fictitious price signals informed by CAISO net load data, with distinct price profiles tailored to represent Summer, Winter, Spring, and Fall conditions. The prices were hosted on a research price server, available to all the CalFlexHub project teams (Liu et al. 2024). Each technology aimed to shift electricity consumption from high- to low-price periods on a daily basis. We evaluated the performance of these MPC technologies using a standardized measurement and verification framework. Further details regarding these projects are available in the papers cited in the next section and in an upcoming final project report (Piette et al. 2026).

Table 1: MPC demonstration sites characteristics

Project #	Building Sector	# Sites	End uses controlled	Underlying control system [1]	Optimization objective and constraints
P1	Single family	1	EV charging, mini-split air conditioner,	Home-energy management systems	Obj: Minimize cost of energy

[1] In these projects MPC functioned as a supervisory layer. By controlling temperature and other setpoints, it enables the underlying control infrastructure to manage direct end-use operations, e.g. like cycling HVAC systems on or off.

Project #	Building Sector	# Sites	End uses controlled	Underlying control system [1]	Optimization objective and constraints
	(Qamar et al. 2026)		smart dryer, smart dishwasher	(HEMS), smart breakers, and residential appliance control apps	Constraints: avoid violations of max panel capacity and maintain services and comfort
P2	Small commercial (Paul et al. 2025a)	6	Rooftop heat pumps, split heat pump systems	Smart thermostats. BACnet connected thermostats, protocol translation gateways	Obj: Minimize cost of energy; Maximize shifted energy, reduce demand charges; Constraints: maintain comfort
P3	Large commercial (Zanetti et al. 2025, 2026)	1	Air-handlers with direct-expansion coil (underfloor air distribution system), water-to-water heat pump, terminal boxes	Building automation system	Obj: Minimize cost of energy Constraints: maintain comfort
P4	Campus (Kim et al. 2022)	1	Chillers, 2-million gallons chilled water storage tank	Building automation system	Obj: Minimize cost of energy; minimize peak demand charges; maximize self-consumption of on-site generation. Constraints: no impact on downstream HVAC systems

As summarized in Table 2, initial evaluations of these demonstrations highlight a substantial load impact of the technology. The MPC controllers successfully curtailed between 30% and 80% of the controlled electrical load during 6-hour high-price windows, yielding overall energy cost reductions ranging from 10% to 70%. Beyond delivering load shifting and cost reduction, the field demonstrations highlighted additional capabilities of the control systems, which addressed specific site-level constraints. In the residential application (P1), the MPC provided coordination of multiple loads to mitigate capacity limits of the electrical panel (Qamar et al. 2026). In the small commercial demonstration (P2), the controller was able to synchronize the operation of multiple rooftop units to reduce total building peak demand (Paul et al. 2025). In the large commercial application (P3), MPC demonstrated reliability and adaptability, automating responses to HVAC compressor failures and wildfire-related damper closures (Zanetti et al. 2025). Finally, in the campus-scale implementation (P4), the control system demonstrated the ability to fully utilize the available on-site generation output (Kim et al. 2022).

Table 2: Preliminary results from the field tests

Project #	Load shed in high price windows (6-hours)	Energy cost savings (% of controlled loads)	Additional benefits demonstrated
P1	50-80%	50-70%	Load coordination to avoid exceeding home electric panel capacity
P2	~30%	~10%	Coordination of multiple units to reduce peak. Improved comfort through pre-occupancy heating in winter season
P3	40-65%	50-60%	Automatic response to failed HVAC compressors and outdoor damper closing due to wildfire. Reduction of simultaneous heating and cooling
P4	~80%	~10%	Full utilization of on-site generation output

Single family house (P1)

California's consumers who want to upgrade their appliances or vehicles often face a severe physical bottleneck: the strict capacity limits of existing home electrical panels and utility service feeders. Upgrading this legacy equipment is often a prohibitively expensive and time-consuming barrier for homeowners, which can drastically slow the pace of this transition (Less et al. 2024). Project P1 addresses this challenge, by demonstrating how software-driven flexible loads can bypass the need for physical hardware upgrades while simultaneously turning homes into price-responsive grid assets. Rather than relying on a traditional smart breaker or panel, which operates blindly and simply trips to cut all power when a panel capacity is exceeded, the project developed an advanced Home Energy Management System (HEMS) that controls an EV, a mini-split air conditioner, a smart dryer and a smart dishwasher. The HEMS utilizes a two-tiered control architecture (Qamar et al. 2026). The system relies on a rule-based lower-level controller that operates on a sub-second timescale to continuously monitor real-time current and instantly curtails or modulates non-essential flexible loads before the panel's capacity is reached. Concurrently, on a longer timescale, the system's MPC layer looks ahead 24 hours to anticipate dynamic prices, strategically shifting heavy energy consumption into periods of inexpensive, abundant generation. This dual-layered approach safely absorbs sudden household electrical shocks to protect the infrastructure, while continuously optimizing the home's energy profile for grid efficiency and cost savings (Qamar et al. 2026). Figure 2 illustrates the single-family residence utilized for the field study alongside representative results from Summer 2025. The electricity price profile is shown in red (right axis), highlighting a peak window between 5:00 PM and 9:00 PM. Compared to the baseline power profile (dotted black line), the MPC-controlled profile (solid green line) demonstrates a clear shift in demand. The red-shaded areas indicate periods of increased power consumption, while the green-shaded areas denote load reductions achieved during high-price periods. In addition, this was achieved without exceeding the panel capacity. Preliminary analysis shows 50-80% of load reduction during six-hour high price periods (Table 2).

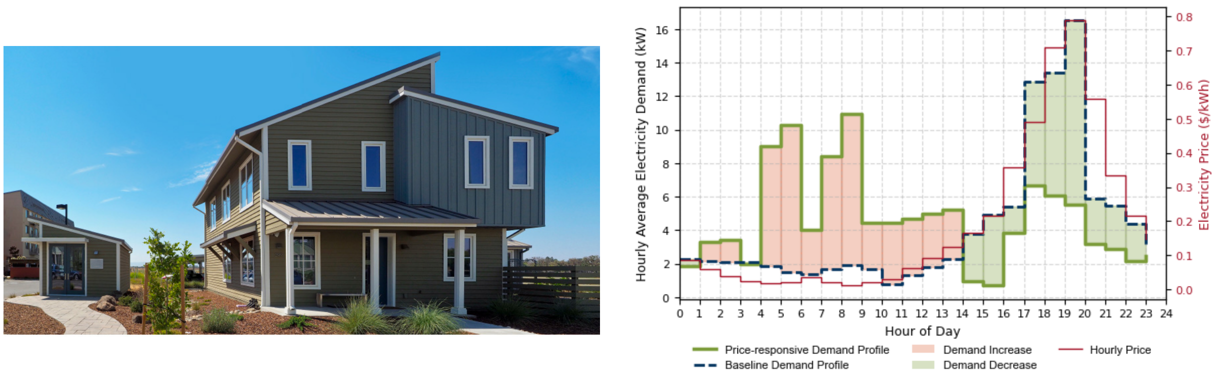

Figure 2: Test results in the single family home. The MPC orchestrates various electrical loads to optimize for seasonal pricing signals while maintaining demand below panel capacity limits (Summer 2025)

Small commercial building (P2)

Small and medium commercial buildings represent a massive but historically underserved segment of the building stock, primarily because they lack the sophisticated automation systems found in larger facilities. To unlock the load flexibility of this sector without requiring prohibitive hardware retrofits, P2 developed an open-source control platform. The software architecture relies on two components: a versatile middleware layer capable of communicating with a variety of local thermostats or control systems, and an intelligence layer where the MPC algorithm resides (Paul et al 2025a). Figure 3 illustrates this performance at one of the project's six field sites: a public library conditioned by two standard HVAC rooftop units. Facing a Spring pricing profile characterized by distinct dual peaks, one in the morning and another in the evening, the MPC anticipated the morning price surge. The results show the system actively shifting the heating load into the more inexpensive, pre-dawn hours to reduce energy costs during the morning peak. Preliminary results across the entire diverse portfolio of field deployments, spanning from libraries to small offices, the MPC platform demonstrated a robust and consistent average load reduction of 30% during high price periods (Table 2), while maintaining or improving thermal comfort. The controller was also able to synchronize the operation of multiple rooftop units to reduce total building peak demand.

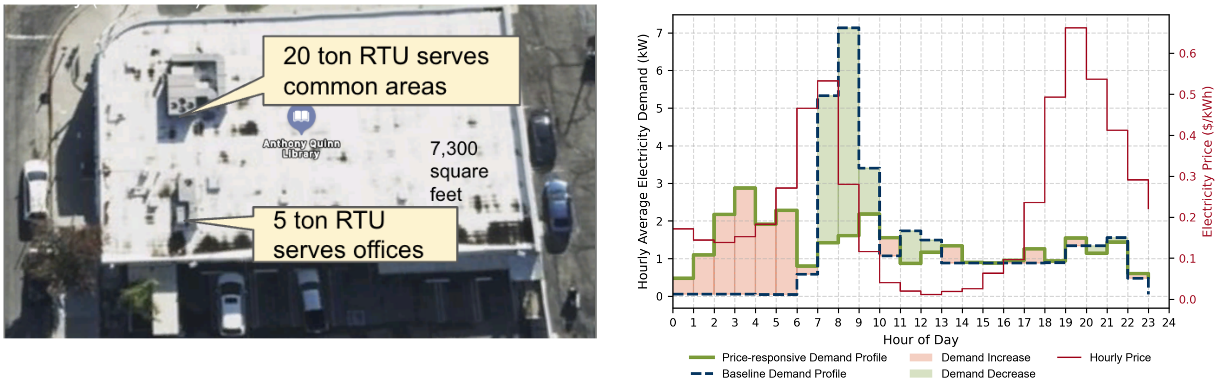

Figure 3: Test results in one of the small commercial buildings. The MPC controls two rooftop heat pumps and optimizes operation under seasonal dynamic prices (Spring 2024)

Large commercial building (P3)

Scaling up to large commercial environments, project P3 demonstrates how MPC can be effective in optimizing the operation of more complex buildings. At the site depicted in Figure 4, a 60,000-square-foot office building, the MPC was integrated into the building's existing legacy automation system to optimize a bespoke underfloor air distribution HVAC system. Rather than operating on rigid schedules, the MPC algorithm continuously ingested dynamic prices and weather information to change supply fan speeds and air temperature setpoints. The results of this optimization were substantial: the MPC achieved a reduction of up to 65% of the load during high-price winter periods, and roughly 50-60% in overall energy and cost savings during the summer (Figure 4). The large energy savings were achieved by reducing simultaneous heating and cooling caused by excessive use of reheat used to keep a minority of the offices comfortable (Zanetti at al. 2025). Crucially, these results did not come at the expense of the people inside, as the algorithm maintained more consistent temperature boundaries than the building’s default controller. Figure 4 illustrates the combination of dynamic load shifting and energy efficiency during the Summer 2024 testing period.

Beyond typical operations, the MPC demonstrated the ability to react to critical equipment failures and adverse environmental conditions. During a period of compressor failure affecting two of the four rooftop units, the controller autonomously adjusted its strategy, leveraging nighttime ventilation by adjusting fan flow and adjusting damper positions, to preserve occupant comfort despite reduced cooling capacity. Similarly, the system’s adaptability was confirmed during a wildfire event, where a dedicated "wildfire mode" was developed and deployed to account for outside air dampers closed by the facility staff. These instances highlight that the MPC can maintain operational stability through targeted parameter adjustments, successfully mitigating the impact of unexpected mechanical and external challenges (Zanetti et al. 2026).

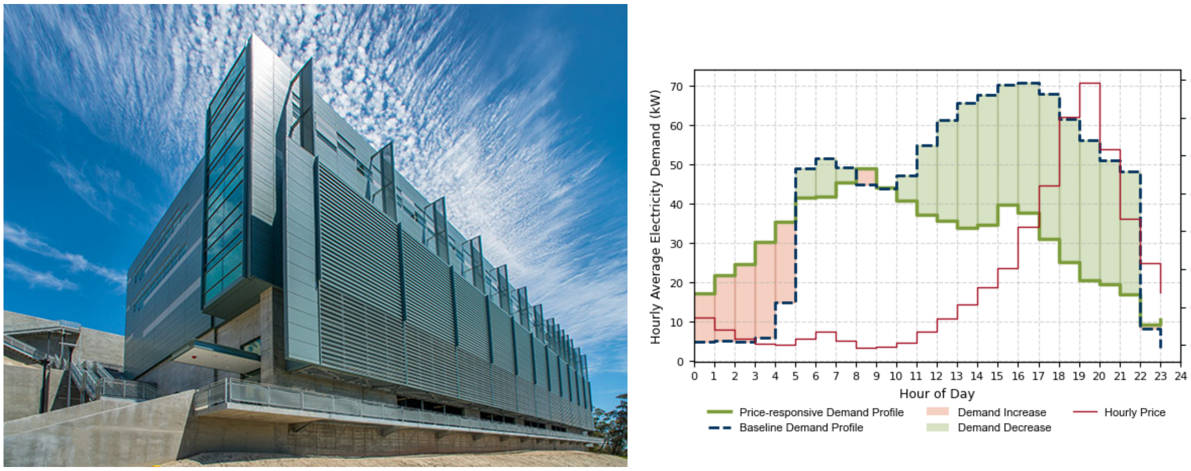

Figure 4: Test results in the large commercial. MPC controls an underfloor air distribution system and additional rooftop heat pumps to respond to seasonal dynamic prices (Summer 2024)

Campus (P4)

To illustrate the scalability of MPC to large scale campus energy systems and the potential for optimal operation of thermal energy storage (TES), project P4 deployed a supervisory MPC controller to manage the cooling system of a university campus. This system consisted of a 5,000-ton chiller plant, a 5-megawatt solar array, and a 2-million-gallon chilled water TES tank (Figure 5). Consistent with the implementation at the other field studies, the MPC continuously ingested real-time data, including dynamic electricity prices, weather forecasts, and solar generation. This data was used to optimize operation of the system to reduce energy costs while meeting cooling demands. Figure 5 shows an illustrative example of the system operation with a large reduction of the chilled water load (~80%) between 5 and 9 PM, the period with the highest electricity prices. The MPC looked ahead and aggressively utilized TES to minimize load during this high price period, whereas the manually operated baseline continuously ran the chillers with conservative TES usage.

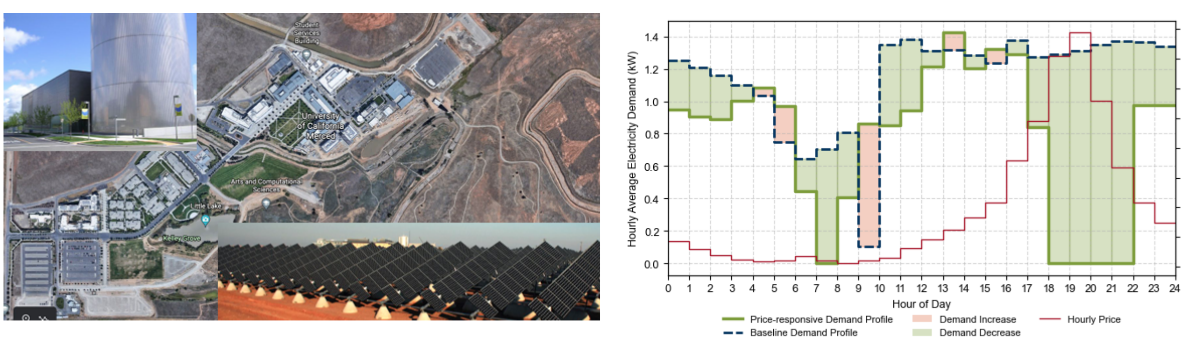

Figure 5: Test results at the university campus. MPC controls a district cooling system, central chilled-water thermal energy storage and PV array.

Discussion

Effectiveness of MPC

The four field demonstrations discussed above clearly illustrate the great potential of MPC. Across residential homes, small and large commercial buildings, and campus central plants, MPC consistently proved its ability to dynamically shift electrical loads, reduce energy cost, preserve occupant comfort, and provide additional benefits. Furthermore, the field tests demonstrated that once the underlying data pipeline and mathematical models of an MPC system are successfully implemented, the algorithms are remarkably flexible. Because MPC relies on an objective function rather than a hard-coded set of "if/then" rules, the same predictive engine can be easily changed to target entirely different operational goals. A good example of this adaptability occurred during the 2024 California wildfires: when facility operators in P3 manually closed outside air dampers to protect indoor air quality, researchers rapidly implemented a "smoke mode" within the MPC, seamlessly reprogramming the algorithm to optimize cooling without relying on fresh air intake (Zanetti et al. 2025, 2026). These real-world results are highly consistent with the broader, evolving landscape of MPC research. Recent meta-analysis of multiple MPC field studies corroborates our observations, demonstrating a strong positive correlation between dynamic price variations and energy cost savings (Zanetti et al 2025). What was predominantly a theoretical exercise in the academic literature a decade ago is now being successfully demonstrated in real buildings by researchers (Khabbazi et al. 2025; Zanetti et al. 2025) and industry[2] [3] [4] [5].

Scaling challenges in real-world deployments

Although the research and development projects highlighted above did not prioritize generalizable economic assessments of site-specific costs and savings, our portfolio of field demonstrations uncovered persistent hurdles. These challenges must be addressed to facilitate the transition from research prototypes to scalable market solutions.

Time-intensive workflows to implement and deploy the control system. From the demonstration projects, we identified steps common to all the implementations and estimated the time required for each one of them (Table 3). The process typically spans from initial customer acquisition and site audits to final system hand-off, often requiring weeks or even months of dedicated effort. The most significant time investments, MPC design, development, and tuning, require extensive site-specific customization, effectively making every deployment a unique prototype. The labor-intensive nature of manual deployment currently acts as one of the primary barriers to achieving the scalability necessary for broad market adoption (Blum et al. 2022; Ham et al. 2024; Zanetti et al. 2026;).

---

[2] Community Energy Lab'building management solution: https://communityenergylabs.com/solution/

[3] Elexity's predictive energy control platform https://www.elexity.io/

[4] Brainbox AI's AI Control https://brainboxai.com/en/solutions/ai-control

[5] QCoefficient EMeister MPC https://www.buildingsasbatteries.com/emeistermpc

Table 3: Estimated time and expertise required for various phases of MPC design and deployment during field demonstrations

Phase #	MPC Implementation Phase Description	Time Required [6]	Expertise Required
1	Customer/site acquisition and building audit	Days	Sales, Stakeholder Engagement, Energy Efficiency Audits
2	(Optional) Building retro-commissioning	Weeks	Control Systems, Mechanical Systems, Building Science
3	(Optional): Control software/hardware upgrade	Weeks	Mechanical Systems, Electrical Systems, Control Systems, IT Networking
4	Data assessment and integration	Days to weeks	Building Science, Mechanical Systems, Data Science, Computer Science, Cybersecurity
5	MPC design and development	Weeks to months	Building Science, Mechanical Systems, Control Systems, Optimization, Data Science
6	MPC deployment and tuning	Weeks	Building Science, Mechanical Systems, Control Systems, Optimization, Data Science
7	System hand-off and workforce education	Days to weeks	Stakeholder Engagement

Expertise required to design and maintain MPC. The projects required a multidisciplinary approach that integrated expertise from a variety of disciplines: mechanical and electrical engineering, building science, data management, computer science, mathematical optimization, and stakeholder engagement (Table 3). Among the workflow phases, the iterative design and operational tuning (Phases 5 and 6) of MPC implementation represent the most significant barrier to broad commercialization, demanding significant time and highly specialized expertise. The industry currently grapples with the *PhD-in-the-loop* dilemma, a reality where these advanced predictive systems often require continuous, manual oversight by highly trained researchers to prevent failure (Figure 6). During design, the MPC expert must navigate a delicate trade-off between model accuracy and computational speed. A thermodynamic model must be precise enough to maintain occupant comfort and yield control actions that achieve performance goals, yet simple enough to be mathematically solved by an optimization engine in real-time every few minutes (Drgoňa et al 2021). Further, certain systems require special design considerations

[6] Estimate ranges provided are inherently variable, as they reflect the heterogeneity in systems studied, ranging from single-family homes to large heating/cooling energy plants. Furthermore, these figures include time allocated to research-specific validation and investigative tasks that could be streamlined or eliminated in a commercial-ready deployments.

due to timescales of their systems. For example, at the campus scale (P4), researchers discovered that rapidly moving cloud cover could abruptly drop on-site solar photovoltaic generation by two megawatts in less than 15 minutes. This sudden fluctuation forces the MPC to rapidly pivot its strategy, risking unintended peak power draws from the grid. In project P1, the research team had to protect from panel overloads at second timescale, requiring a specialized hierarchical control architecture that is specific to this use case. Because large commercial buildings (P3) and campus central plant systems (P4) are highly customized, developing these initial models requires significant, site-specific engineering. Further, availability of different sensors in otherwise identical systems may also require different models or optimization formulations (Blum et al. 2022). During operation, periodic changes may be necessary to adapt models, constraints or cost functions (Figure 1) to changes or reconfigurations of mechanical systems, changes in underlying control software and logic, priority in objectives, or emergence of system faults.

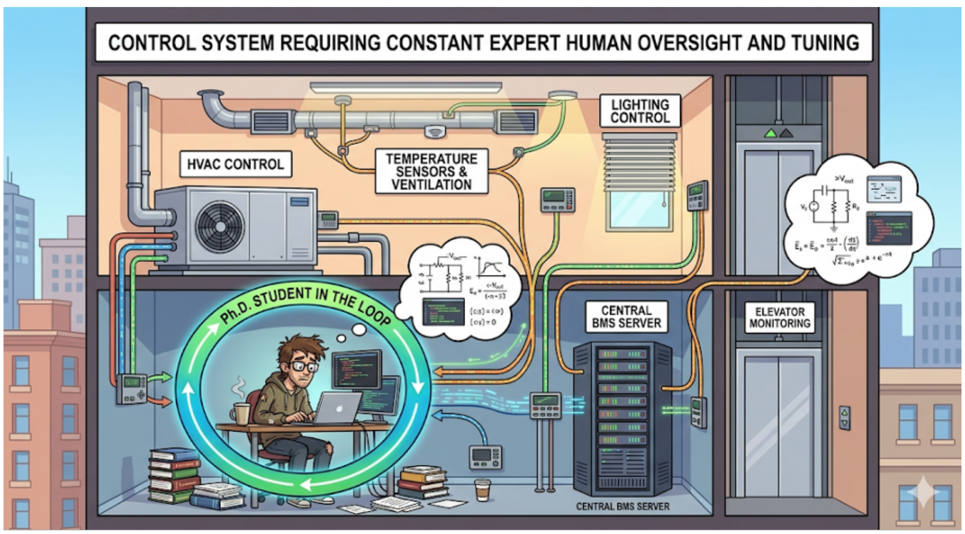

Figure 6: AI-generated (Gemini 3.5 Flash) satire illustrating the "PhD-in-the-loop" required to maintain MPC operations

Vulnerability to unreliable data infrastructure. Integration of MPC with the underlying control systems is often complex (Kim et al. 2022; Paul et al 2025b; Qamar et al 2026; Zanetti et al 2025). These field tests demonstrated that MPC systems can be brittle when exposed to real-world communication networks. Because these controllers depend on a continuous stream of data to calculate optimal setpoints, any digital interruption, such as a dropped internet connection, a deprecated Application Programming Interface (API), or a server restart, can instantly blind the optimization engine. The sheer complexity of integrating multiple distinct data sources means that failures and data interruptions are a recurring problem, currently necessitating constant human monitoring and watchdog algorithms. Moreover, technical hurdles were further compounded by a significant heterogeneity in hardware that made the data integration process excessively labor-intensive. Implementation efforts were frequently obstructed by poor documentation of existing control points and disparate cybersecurity requirements, creating barriers to achieving scalable, seamless integration.

Workforce education and acceptance. Ultimately, the broad deployment of MPC transcends software development, representing a significant human-centric challenge. Building operators and technicians frequently encounter these advanced controls for the first time, while facility professionals remain notably risk-averse, concerned that autonomous logic adjustments might trigger widespread mechanical failures. Furthermore, the legal and operational accountability for failures or safety compromises within automated control loops remains poorly defined. To eliminate the reliance on a *"PhD-in-the-loop"* for ongoing maintenance and oversight, robust training programs are critical to empowering staff to manage these complex systems with confidence and independence following the final hand-off.

Alternative approaches

To address the complexity and time required to construct MPC models, the academic literature has been exploring a hybrid approach to building MPC models with machine learning techniques, such as deep learning. These include automatically and periodically updating the model (adaptive MPC) or utilizing fully data-driven methods (data-driven MPC). Notable variants of these data-driven approaches include differentiable MPC, which allows for end-to-end policy optimization via gradients, and physics-informed MPC, which integrates domain knowledge to strictly enforce physically realistic constraints (Drgoňa et al. 2021). While these emerging methodologies offer significant potential, they currently lack the rigorous, real-world validation provided by field demonstrations.

In parallel with the evolution of MPC, the past decade has witnessed an increasing focus on reinforcement learning (RL) within academic literature. RL is a model-free approach that learns control policies through iterative trial-and-error to optimize specific reward functions. However, the need for high-fidelity simulation environments or substantial historical datasets for training RL before deployment often require comparable effort to the construction of explicit MPC models. This requirement significantly adds complexity to the initial setup phase, which, due to factors like unique configurations, must currently be performed for each building individually. Furthermore, real-world validation of RL remains scarce, with few field deployments successfully demonstrating its performance (Touzani et al. 2021; Guo et al. 2025).

Generative AI (GenAI) and foundation models are also being integrated into control loops to bridge high-level intent and system actions. Frameworks like "Office-in-the-loop" use multimodal models for HVAC adjustments (Sawada et al. 2024), while DARLIN employs retrieval-augmented generation to ground decisions in domain knowledge (Ko and Jain 2025). These approaches however still face significant problems, including high latency in inference, the risk of hallucinations in control outputs, potential cost and the lack of robust verification frameworks.

The majority of these approaches remain in early R&D stages and require additional validation before they can be successfully transitioned from academic literature to market-ready products.

Recommendations for scalable MPC

As described above, MPC remains the most mature advanced control methodology for building applications. To accelerate its transition from research into mainstream commercial products, the industry must pivot from bespoke, labor-intensive, one-of-a-kind deployment towards streamlined, scalable workflows. This transition requires a multi-pronged strategy that addresses the high costs and expertise requirements associated with various implementation phases such as customer interaction, system integration, engineering, commissioning, modeling and software update and maintenance (Table 3). These can be achieved through:

- **Segment-specific strategies for MPC development.** To mitigate the labor-intensive engineering required for constructing MPC optimization formulations, models, and constraints, deployment frameworks must be customized for specific building segments, reflecting their unique technical needs and economic viability:
  - **Residential and small commercial buildings**: Since absolute cost savings per building are lower, the priority must be pre-configured or easily configurable systems, both from the communication integration perspective and modeling approach. Developing templates for common configurations (e.g., single-speed vs. variable-capacity HVAC) (Gautier et al. 2023) and developing control approaches for families of products.
  - **Large commercial buildings** (The "Middle Ground"): These facilities are often heterogeneous and controllable assets like HVAC are custom-built, precluding simple templates. Research should focus on innovative tools that automate or semi-automate the model creation process, specifically using semantic models such as Brick Schema or ASHRAE s223[7] (Prakash et al. 2024; Paul et al.. 2025a, 2025b). This approach minimizes the configuration and calibration time required for MPC.
  - **District energy systems:** For campus-scale central cooling/heating systems, the potential for energy savings justifies bespoke, highly customized MPC implementations. In this segment, research should prioritize methods for periodic, automated model and optimization function updates to maintain accuracy and reliability over long operational lifecycles.
  - **All buildings:** GenAI should be used to support configuration of models by automating labor-intensive tasks such as metadata mapping, knowledge graph generation, and the assembly of control problems from validated equipment templates, thereby addressing the *"PhD-in-the-loop"* bottleneck and streamlining deployment workflows across diverse building types (Zanetti et al. 2026).
- **Robust, fault-tolerant communication infrastructure.** Industry must develop reliable, fault-tolerant data pipelines. These systems must be resilient to sensor failures, connectivity interruptions, and API deprecations to ensure control operations persist despite digital infrastructure challenges. Enhancing cross-system interoperability will be vital in streamlining these complex data integration efforts.
- **Workforce skills and trust.** Deployment strategies must prioritize transparent, user-friendly interfaces for facility managers. Reducing risk-aversion in the workforce requires demonstrating reliability and creating comprehensive education pathways, ensuring that stakeholders understand, trust, and can safely interact with these automated systems.
- **Broader economic and performance assessments.** While these demonstrations are promising, future investigations must prioritize the collection of rigorous longitudinal data across a broader range of building typologies and climate zones. Establishing these generalizable benchmarks is essential for quantifying costs and benefits of the technology,

[7] Semantic models, such as the Brick Schema or ASHRAE s223, can accelerate the adoption of MPC applications by providing consistent and structured naming conventions which streamline data normalization, and application configuration

ultimately enabling the industry to identify and prioritize those facilities with the most significant potential for energy cost reductions.

Conclusion

The study demonstrates that MPC is an effective technology for load flexibility across diverse building types, capable of achieving significant load shifting (30% - 80%) and cost reductions (10% to 70%) while maintaining occupant comfort. However, its widespread adoption is currently hindered by three primary barriers: 1) a time-intensive implementation process, made of several phases, 2) the 'PhD-in-the-loop' dilemma, which requires specialized, labor-intensive expertise to design and maintain MPC, and 3) fragile data infrastructures that compromise control reliability, 4) the lack of workforce skills and trust in automated control systems. To overcome these barriers and scale MPC, the industry must transition from bespoke, prototype-heavy deployments to streamlined, segment-specific strategies. This involves adopting model templates for residential and small commercial buildings, utilizing semantic tools to automate model creation for large commercial facilities, and developing automated, periodic updates for the bespoke models needed to operate massive district energy systems. Furthermore, integrating generative AI to automate labor-intensive configuration tasks and metadata mapping will be crucial in reducing engineering efforts and accelerating the path to market-ready load-flexible buildings. Future work should also establish generalizable performance benchmarks by collecting rigorous longitudinal data across diverse building typologies and climate zones.

Acknowledgements

This research was supported by the Assistant Secretary for Critical Minerals and Energy Innovation, Office of Building Technologies of the U.S. Department of Energy, under Contract No. DE-AC02-05CH11231, and by the California Energy Commission EPIC program through the EPC-20-025 (CalFlexHub) and EPC-19-013 grants.

References

Blum, D., Wang, Z., Weyandt, C., Kim, D., Wetter, M., Hong, T., & Piette, M. A. (2022). Field demonstration and implementation analysis of model predictive control in an office HVAC system. Applied energy, 318, 119104. https://doi.org/10.1016/j.apenergy.2022.119104

Braun, J. E. (2003). Load control using building thermal mass. J. Sol. Energy Eng., 125(3), 292-301. https://doi.org/10.1115/1.1592184

California Energy Commission (CEC). (2023). Senate Bill 846 load shifting goal report. (CEC-200-2023-005). https://www.energy.ca.gov/publications/2023/senate-bill-846-load-shift-goal-report

Chen, Y., et al. (2020). "Experimental investigation of demand response potential of buildings: Combined passive thermal mass and active storage." Applied Energy, 258, 114104. https://doi.org/10.1016/j.apenergy.2020.115956

Drgoňa, J., Arroyo, J., Cupeiro Figueroa, I., Blum, D., Arendt, K., Kim, D., Ollé, E. P., Tang, J., & Helsen, L. (2020). All you need to know about model predictive control for buildings: A survey and food for thought. Renewable and Sustainable Energy Reviews, 121, 109638. https://doi.org/10.1016/j.arcontrol.2020.09.001

Drgoňa, J., Tuor, A. R., Chandan, V., & Vrabie, D. L. (2021). Physics-constrained deep learning of multi-zone building thermal dynamics. Energy and Buildings, 243, 110992. 10.1016/j.enbuild.2021.110992

Gautier, A., Wetter, M., Hu, J., & Tummescheit, H. (2023). HVAC and control templates for the Modelica Buildings Library. https://eta.lbl.gov/publications/hvac-and-control-templates-modelica

Guo, F., woo Ham, S., Kim, D., & Moon, H. J. (2025). Deep reinforcement learning control for co-optimizing energy consumption, thermal comfort, and indoor air quality in an office building. Applied Energy, 377, 124467. https://doi.org/10.1016/j.apenergy.2024.124467

Khabbazi, A. J., Pergantis, E. N., Premer, L. D. R., Papageorgiou, P., Lee, A. H., Braun, J. E., ... & Kircher, K. J. (2025). Lessons learned from field demonstrations of model predictive control and reinforcement learning for residential and commercial HVAC: A review. Applied Energy, 399, 126459. https://doi.org/10.1016/j.apenergy.2025.126459

Kiliccote, S., & Piette, M. A. (2014). Buildings-to-grid technical opportunities: From the buildings perspective (No. DOE/EE--1052). Lawrence Berkeley National Lab.(LBNL), Berkeley, CA (United States). https://www.energy.gov/sites/prod/files/2014/03/f14/B2G_Tech_Opps--Building_Perspective.pdf

Kim, D., Wang, Z., Brugger, J., Blum, D., Wetter, M., Hong, T., & Piette, M. A. (2022). Site demonstration and performance evaluation of MPC for a large chiller plant with TES for renewable energy integration and grid decarbonization. Applied energy, 321, 119343 https://doi.org/10.1016/j.apenergy.2022.119343

Ko, Y. D., & Jain, R. (2025). DARLIN: Domain-guided Augmented Retrieval for LLM-based INterpretable HVAC Control. In Proceedings BuildSys 2025. https://doi.org/10.1145/3736425.3772322

Ham, S., Paul, L., Casillas, A., Prakash, A., Kim, D., Brown, R., Pritoni, M., & Grant, P. (2024). Practical challenges and impacts of low-cost model predictive control (MPC) for grid-interactive small and medium commercial buildings. ACEEE Summer Study on Energy Efficiency in Buildings 2024. https://doi.org/10.20357/B7FC7F

Less, B., Walker, I. S., Murphy, S., & Fournier, E. (2024). "Electrical Service Panel Capacity in California Households with Insights for Equitable Building Electrification." Proceedings of the ACEEE Summer Study on Energy Efficiency in Buildings.

Liu, J., Piette, M. A., Pritoni, M., Nordman, B., Grant, P., Smith, S., Brown, R., Bourg, J., Godinez, F., Fung, M., & Martinez, M. (2024). Assessing Customer Experience and Business Models around Price-to-Device Communication and Smart Control Pathways in CalFlexHub. ACEEE Summer Study on Energy Efficiency in Buildings 2024. https://doi.org/10.20357/B7K31N

Neukomm, Monica, Valerie Nubbe, and Robert Fares. "Grid-interactive buildings (GEB) technical report series: overview of research challenges and gaps." (2019). https://www1.eere.energy.gov/buildings/pdfs/75470.pdf

Nagpal, H., Staino, A., & Basu, B. (2020). Application of predictive control in scheduling of domestic appliances. Applied Sciences, 10(5), 1627. https://doi.org/10.3390/app10051627

Paul. L, Ham S W., Anand P.,Casillas A., Yang T.,, Pritoni M., (2025a) HP-FLEX: Field Demonstration of the Semantics-Driven Configuration of a Model Predictive Control System to Make Heat Pumps Flexible. ASHRAE conference. https://doi.org/10.20357/B7QP58

Paul L., de Andrade Pereira F., Prakash A., Ham S.W, Feng J.D., Brown R., Pritoni M. (2025b) Open Building Operating System: a Grid-Responsive Semantics-Driven Control Platform for Buildings. Science and Technology for the Built Environment. https://doi.org/10.1080/23744731.2024.2444819

Piette M.A., Liu J., Nordman B., Smith S., Brown R., Pritoni M., Gerke B., Yee A., Hart M., Fung M. (2022). Accelerating Decarbonization with the California Load Flexibility Research and Deployment Hub. ACEEE Summer Study on Energy Efficiency in Buildings 2022. https://escholarship.org/uc/item/40z931rt

Piette M.A. et al. (2026). California Load Flexibility Research and Deployment Hub (CalFlexHub) - Final Report. In development.

Prakash A. P., de Andrade Pereira F., Berges M., Pritoni M, Akinci B. (2024). Ontologies at Work: Analyzing Information Requirements for Model Predictive Control in Buildings in Buildings. Proceedings of the BuildSys 2024. https://doi.org/10.1145/3671127.3698189

Pritoni M., Wetter M., Paul L., Prakash A., Huang W, Bushby S., Delgoshaei P., Poplawski M., Saha A., Fierro G., Steen M, Bender J., Ehrlich P. (2024). Digital and Interoperable: the future of building automation is on the horizon. What's in it for me? ACEEE Summer Study on Energy Efficiency in Buildings 2024. https://doi.org/10.20357/B76W3X

Sawada, T., Hasegawa, T., Yokoyama, K., & Mizuno, M. (2024). Office-in-the-loop for building hvac control with multimodal foundation models. In Proceedings BuildSys 2024. https://doi.org/10.1145/3671127.3698182

Shan, K., Wang, S., Yan, C., & Xiao, F. (2016). Building demand response and control methods for smart grids: A review. Science and Technology for the Built Environment, 22(6), 692-704.. https://doi.org/10.1080/23744731.2016.1192878

Qamar, A.; Zanetti, E.; Yang, T.; Gerber, D.; Kissock, K.; Pritoni, M., (2026), Field Demonstration of Coordinated Residential Load Control for Dynamic Pricing and Panel-Capacity Protection. (Under review.)

Zanetti E., Blum D., Weyandt C., Pritoni M., Piette M. A. (2025). Commercial Building Flexibility Through Model Predictive Control: Field Demonstration and Literature Insights. Energy and Buildings. https://doi.org/10.1016/j.enbuild.2025.116097

Zanetti E, Blum D. Prakash A., Paul L., Pritoni M. (2026). Implementation Lessons and Future Pathways for Scalable Model Predictive Control in Large Commercial Buildings. 2026 Herrick Conference on High Performance Buildings. (In publication)

Zhou, Y., Cao, S., Hensen, J. L., & Lund, P. D. (2019). Energy integration and interaction between buildings and vehicles: A state-of-the-art review. Renewable and Sustainable Energy Reviews, 114, 109337. https://doi.org/10.1016/j.rser.2019.109337