Title

Implementation Lessons and Future Pathways for Scalable Model Predictive Control in Large Commercial Buildings

Authors

Zanetti, Ettore
Blum, David
Prakash, Anand
et al.

Publication Date

2026-07-17

Buildings & Industrial Energy Systems
Energy Technologies Area
Lawrence Berkeley National Laboratory

Implementation Lessons and Future Pathways for Scalable Model Predictive Control in Large Commercial Buildings

Ettore Zanetti, David Blum, Anand Krishnan Prakash, Lazlo Paul, Marco Pritoni

Lawrence Berkeley National Laboratory

Energy Technologies Area
July 2026

[Insert citation or DOI # here]

This work was supported by the Assistant Secretary Office of Critical Minerals and Energy Innovation (CMEI) of the US Department of Energy under Contract No. DE-AC02-05CH11231.

Implementation Lessons and Future Pathways for Scalable Model Predictive Control in Large Commercial Buildings

Ettore ZANETTI[1*], David BLUM[1], Anand Krishnan PRAKASH[1], Lazlo PAUL[1], Marco PRITONI[1],

[1] Building & Industrial Energy Systems Division, Lawrence Berkeley National Laboratory,
Berkeley, CA, USA
ezanetti@lbl.gov

* Corresponding Author

ABSTRACT

Model Predictive Control (MPC) has demonstrated potential for reducing building energy costs and integrating buildings into the electric grid, yet adoption in large commercial buildings remains limited. We developed, deployed, and demonstrated an MPC in a large office building from 2020 to 2025, an uncommonly long duration for MPC field research that has enabled us to gain valuable insights. The MPC achieved 45% energy savings in efficiency mode and an estimated 61% annual cost reduction under experimental dynamic prices compared to baseline rule-based control, leveraging thermal mass for load shifting across all four seasons. However, the demonstration revealed critical barriers to broader adoption: uncertain cost-to-benefit ratios, including dependence on specialized expertise for deployment, integration hurdles due to proprietary and many different data streams (APIs/protocols), gaining facility staff trust, and updating the Building Management System (BMS). There were also operational challenges from unexpected events such as wildfire protocols, equipment failures, and inconsistent data. Initial use of semantic data models for data acquisition is promising, but encountered performance issues with large queries. This manuscript summarizes the long-term deployment evolution and performance over time, details operational challenges and solutions, and proposes scalability pathways, including the opportunities and challenges with leveraging the use of Generative AI.

1. INTRODUCTION

Commercial buildings account for 35% of electricity use in the U.S. (EIA, 2025), and buildings greater than 50,000 sqft account for 54% of that electricity use, despite accounting for just 6% of the total number of commercial buildings (EIA, 2018). This means the potential for tapping into significant load shift potential with fewer demand flexibility (DF) customers to recruit and coordinate. Given HVAC systems account for 52% (EIA, 2018) of commercial building energy use and access inherent passive thermal energy storage, they represent a significant DF source (Tang, Wang, & Li, 2021). Model Predictive Control (MPC) and Reinforcement Learning (RL) are advanced supervisory control strategies that leverage building thermal mass, weather forecasts, and grid signals to optimize energy management and costs, and can enable significant DF. Despite extensive research demonstrating their effectiveness, widespread commercial adoption has been hindered by uncertain value propositions, reliance on specialized expertise, and poor data interoperability (Khabbazi et al., 2025). A critical review of field demonstrations reveals that long-term (higher than 30 days) deployments in occupied, large commercial buildings with more than 50,000 sqft, are rare; with only 17 out of 2892 studies (Khabbazi et al., 2025) analyzed. Furthermore, only a fraction of papers report the costs or labor associated with deployment.

To address this gap, this paper presents an analysis of an uncommonly long research deployment of MPC in a large commercial building, Building 59 (B59) at Lawrence Berkeley National Laboratory (LBNL). We summarize how the MPC and baseline building control evolved over time, summarize the results from two previous studies and a third yet unreported testing period, present operational challenges and how they were addressed, and suggest future needs and pathways for scaling MPC. This includes opportunities and challenges for the use of generative AI and agentic workflows, which is growing in advanced building control applications (Wan, Zhang, Chen, Xu, & Feng, 2025). Between 2020 and 2025, the B59 MPC was iteratively upgraded and tested across multiple projects to maintain service and test several objectives. These included initial implementation and energy minimization tests over 31 days in fall and winter (Blum et al., 2022); cost minimization tests under dynamic electricity pricing over 46 days spanning all four seasons, using prices that change each hour and repeat daily, with a different daily profile each season (Zanetti

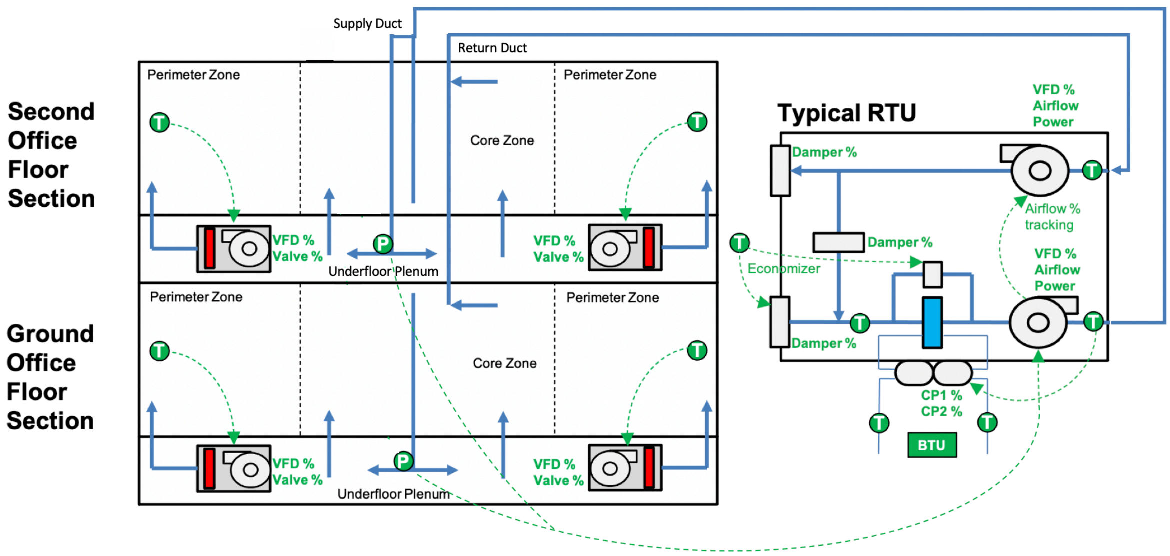

Figure 1: Design and control schematic in section view for one RTU of the UFAD HVAC system. Important sensor and control points are labeled, including temperatures (T) and static pressures (P). Original figure from (Blum et al., 2022).

et al., 2025); and new cost minimization tests first reported in this manuscript under even more dynamic electricity pricing with data acquisition through semantic queries over 27 days in spring and summer, using prices that change each hour, with a different daily profile each day for a week, where the weekly profile repeats for a given season. In total, the MPC operated the building for 104 days. By documenting the operational challenges of practical MPC deployment, including hardware and data failures, software sustainability hurdles, extreme weather conditions, and labor time, this manuscript highlights the effort required to maintain MPC operation and scale it in practice.

2. CASE STUDY

This section describes a summary of characteristics of the Building and its HVAC system for convenience to the reader. A detailed description of these systems has been reported in (Blum et al., 2022), and details about additional updates that were made to the baseline control system are reported in (Zanetti et al., 2025).

2.1 Site Description and HVAC System

The field demonstration site is Building 59 (B59) at LBNL, a four-story large commercial building built in 2015, located in Berkeley, California, ASHRAE Climate Zone 3C. The MPC was deployed to control the HVAC system serving the third and fourth floors, totaling 65,000 sqft, which predominantly consist of enclosed and open office spaces. The building structure is steel-framed with an exterior metal curtain wall system with integrated windows and foamed insulation core. In office areas, finished floors with carpeting are raised above structural concrete slabs, creating the plenum for an Underfloor Air Distribution (UFAD) system.

Heating, cooling, and ventilation are provided to the offices by the UFAD. The system uses four large Rooftop Units (RTUs) located on the roof with direct expansion (DX) coils to supply cool air to the common underfloor plenum on each level, as shown in Figure 1. Air diffuses directly to the core and perimeter zones through floor diffusers and can be pulled and heated to perimeter zones with fan-powered underfloor terminal units (UFT). Hot water for the UFT heating coils is produced by a water-source heat pump (WSHP) located on the mechanical level, in the first floor of the building. The RTUs do not provide centralized heating to the spaces.

2.2 Baseline Control

For HVAC operation the building is considered occupied only during weekdays, from 5 a.m. to 10 p.m., which can differ from actual occupancy. To meet the minimum outside air requirement, during occupied times each RTU needs to provide a minimum airflow, 25% of nominal fan capacity. The outside airflow setpoint is set to zero otherwise. The perimeter zone thermostat heating and cooling setpoints are also based on this occupancy schedule, with the average heating setpoint being 21^oC during occupied hours, 19^oC otherwise, and cooling setpoint 23^oC during occupied hours, 26^oC otherwise. The RTUs are turned off during unoccupied hours unless there is a call for heating or cooling as determined by the unoccupied setpoints.

3. MPC IMPLEMENTATION

This section briefly summarizes the MPC implementation with a focus on updates from the study from (Blum et al., 2022) to (Zanetti et al., 2025), and a follow up experiment, firstly reported in this manuscript, integrating semantic modeling with the MPC controller. For more details on the overall MPC implementation please refer to (Blum et al., 2022) and (Zanetti et al., 2025).

3.1 MPC Summary

The MPC uses grey-box R2C2 to predict zone temperatures, airflow, and power consumption across four thermal zones. RTU components are modeled with polynomial power functions. Envelope and fan parameters are updated daily through parameter estimation, while DX parameters were calibrated once.

The MPC requires weather, electricity price, internal load, and temperature setpoint forecasts. Weather forcasts come from NOAA (NOAA, 2026) while weather measurements come from a weather station at LBNL. Global Horizontal Irradiation (GHI) forecasts are estimated via a data-driven model. The dynamic prices change hourly and repeat daily in (Zanetti et al., 2025), while they change hourly and daily and repeat weekly in the latest round of tests (Liu et al., 2024). Internal loads are predicted from historical plug load averages, and setpoint forecasts coincide with the historical schedules defined in the baseline.

The controller minimizes total HVAC electricity cost over a 24-hour horizon using the dynamic prices. Zone temperature bounds are soft constraints with regularization to prevent oscillatory behavior, while hard constraints limit supply air temperature and enforce minimum outside airflow. A moving horizon estimator provides state initialization.

3.2 Software Architecture: Design and Evolution

The MPC infrastructure relies on a service-oriented architecture leveraging Docker containers and an InfluxDB time-series database as shown in Figure 2. The core optimization, parameter estimation, and state estimation services are built upon the open-source software MPCPy (Blum & Wetter, 2017) commit ecb9833 with models implemented in Modelica. Figure 2 also shows the different systems, methods where data is collected and interface with the BMS. Separate Python modules were developed to collect data from each building system, since each building system requires use of a separate API or interface. The Historic Data Collection service runs on a single Python module that centralizes the collection of data into an influxdb making it available to each of the system specific modules.

In Figure 2, services highlighted in red indicate updates from (Blum et al., 2022) to (Zanetti et al., 2025), while services highlighted in blue indicate the latest updates to integrate the semantic model. In order to update the MPC from (Blum et al., 2022) to (Zanetti et al., 2025), various data collection services needed to be updated. Plug load data acquisition was migrated from Elastic Search database query to a campus Skyspark (SkyFoundry, 2026) API, weather forecasts from DarkSky (DarkSky, 2026) to the NOAA API, and a new service was added to retrieve day-ahead dynamic electricity prices from CalFlexHub (CalFlexHub, 2024) and MIDAS (MIDAS, 2024) servers for use in the MPC cost objective. Also, after (Blum et al., 2022), the facility managers updated the baseline control to include occupancy-based schedules for zone temperature setpoints, minimum outside air flow setpoints, and HVAC system shutdown and startup. Therefore, for (Zanetti et al., 2025), the MPC was updated to add a service to retrieve zone temperature and outside air flow setpoint schedules and use them to determine constraints in the MPC formulation, and to expand the set of writeable setpoints by the MPC in the ALC BMS (ALC, 2026) to determine RTU on/off for pre-heating and pre-cooling during unoccupied hours via a new ALC setpoint *unocc_run*.

Since (Zanetti et al., 2025), a semantic model of B59 was developed using the ASHRAE Standard 223 (s223) ontology (ASHRAE, 2018). This latest modification to the software architecture included changing the data collection and parameter and state estimation services to use the semantic model for querying the necessary data streams and actuation end points, through dedicated SPARQL (W3C, 2026) queries, instead of relying on hard-coded variable names. This removed hundreds of lines of code. This architecture means that any changes in the future to the sensing and actuation infrastructure only need to be updated in the semantic model and the changes would be captured in the corresponding services, without any additional software development.

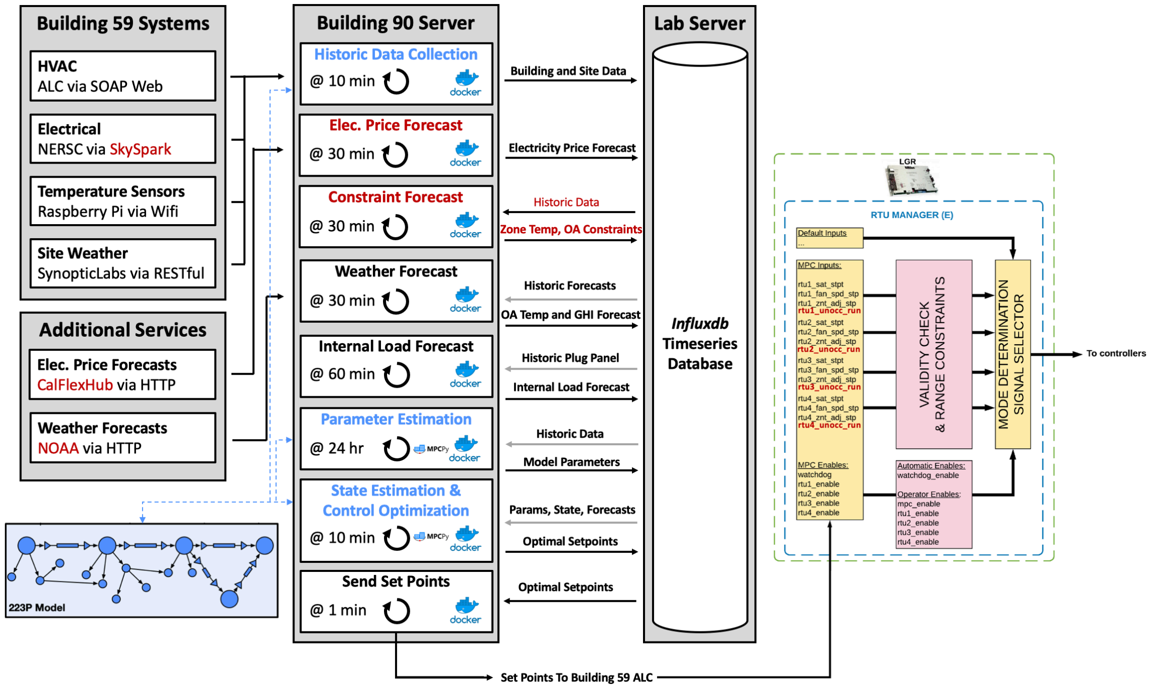

Figure 2: Schematic of the MPC software architecture, data integration, and building BMS communication, with updates from (Blum et al., 2022) to (Zanetti et al., 2025) highlighted in red. Services updated to use the 223P semantic model are highlighted in blue.

4. RESULTS

The initial energy minimization tests (Blum et al., 2022) were conducted between 10/18/2020 and 12/15/2020, alternating a week of MPC operation with a week of baseline, for a total of 31 days under MPC control and 27 days under baseline. The initial cost minimization tests using dynamic electricity hourly price profiles that repeated daily and changed seasonally (Zanetti et al., 2025) were conducted between 6/1/2023 and 5/31/2024, operating MPC control for two weeks at a time, subject to operational challenges, and otherwise under baseline. Overall, MPC control was operated 8/21/23–8/26/23, 9/27/23–10/6/23, 2/9/24–2/24/24, and 4/9/24–4/19/24. The latest cost minimization tests using highly dynamic electricity hourly price profiles that changed daily, repeated weekly, and changed seasonally and using the semantic-integrated data acquisition service were conducted between 3/1/2025 and 8/31/2025, with MPC control 3/12/2025–3/26/2025 and 7/1/2025-7/11/2025, and baseline control otherwise.

4.1 Energy, Cost, and Flexibility Performance

In (Blum et al., 2022), the energy-minimizing MPC achieved approximately 40% HVAC energy savings compared to the baseline, shown in the left scatter plot in Figure 3. During the dynamic pricing tests in (Zanetti et al., 2025), the MPC achieved an annualized estimated cost savings of approximately 61% and an energy savings of 45%, as shown in the right scatter plot in Figure 3. Comparing the two plots in Figure 3, the baseline (MPC OFF) period exhibits relatively constant consumption in the 2020 study, whereas the 2023–24 study shows more variance. This is partly attributable to a broader range of seasonal periods captured, but also reflects the change in baseline control logic mentioned in Section 3.2. During the 2020 study, RTU fans operated at nearly constant speed year-round with no occupancy scheduling. Building on insights gained from the MPC tests in (Blum et al., 2022), the building facilities team implemented new baseline logic within the BMS that introduced occupancy schedules and pressure resets, enabling better modulation of HVAC power in response to actual building demand without compromising comfort. This illustrates an additional way the MPC can add value in complex systems, where even a short deployment can provide facility managers with the confidence and data needed to optimize baseline control logic.

Figure 4 illustrates the MPC's load-shifting capability under dynamic pricing, showing hourly energy use distributions for the Spring and Summer of the initial cost-minimization tests (Zanetti et al., 2025) and the latest cost-minimization

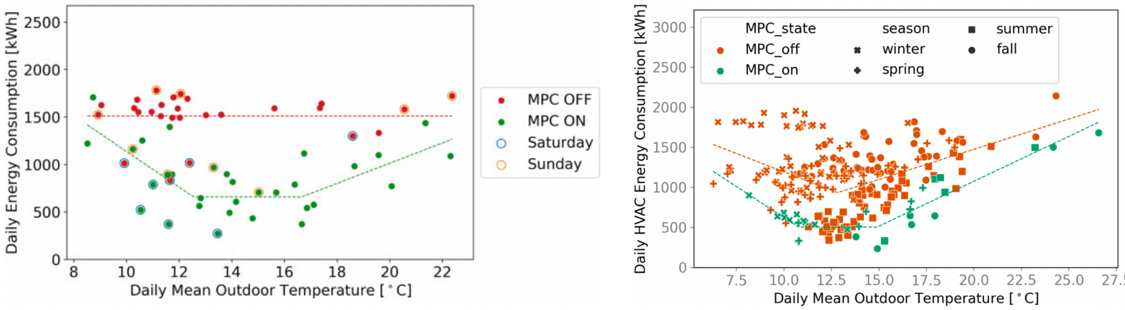

Figure 3: Daily energy consumption in (Blum et al., 2022) on the left and (Zanetti et al., 2025) on the right. Saturday and Sunday are excluded from the plot on the right since the building was unoccupied and HVAC consumption was close to zero.

tests with semantic model, under both MPC (MPC ON) and baseline control (MPC OFF). The electricity price is overlaid in blue, with fixed hourly values across each day for the (Zanetti et al., 2025) study and an hourly distribution for the latest study due to hourly profile that changed daily and repeated weekly. In both tests, the MPC successfully shifted load from high to low price periods, quantified by a Demand Decrease Percentage (DDP) during the 6-hour peak price window, ranging from 49-69% in the (Zanetti et al., 2025) Summer and Spring test to 30-80% in the latest Spring and Summer tests. Two factors explain this difference in the Spring test. First, the Spring period in the latest test experienced more extreme weather with winter- and summer-like days occurring in close succession, reducing the ability of MPC to shift load. Second, three RTU compressors failed prior to the latest tests, forcing the MPC to compensate through increased fan energy and outside air recirculation as described in the next section, while the baseline control failed to maintain comfort during portions of the day as zones overheated. In the Summer season, the latest tests had a lower energy consumption profile compared to the baseline and looking at (Zanetti et al., 2025) as a reference. The first reason is milder weather conditions during the latest Summer tests leading to an overall lower cooling consumption compared to the baseline. The second reason is that the MPC was left unattended for most of the test and, while it maintained comfort with no occupant complaints, the objective function had slightly less weighting on temperature constraints compared to previous experiments without tuning against the new prices, leading to lower consumption.

4.2 Fault Tolerance and Mitigation

One significant benefit observed during the long-term deployment was the MPC's automated robustness and capacity for proactive fault mitigation. Before the Spring tests we noticed that RTU1 completely lacked cooling capacity with no compressors running due to compressor alarms of unknown origin, and RTU3 was at half capacity with just one compressor running, similarly due to a compressor alarm. Therefore, the MPC max cooling power constraints were updated to 0% for RTU1 and 50% for RTU3 during these tests. To show the impact of changing those constraints, Figure 5 shows a comparison between the MPC before updating the constraints, in the left plot, and the MPC after updating the constraints, in the right plot, during two consecutive warm days with broken compressors. Despite the zones being warm at the end of the first day and unavailable compressors, the MPC with updated constraints was able to keep the building as comfortable as possible throughout the second day by circulating cold outside air for the whole night before. This is shown in the supply fan flow plot (b) and the outside air damper position plot (c), to the point of overcooling some of the zones. This allowed reduced discomfort during the central hours of the day as shown in plot d). The compressor control signals in plot c) are stuck at 100% but the actual compressor speeds are 0% since they were not operating.

A second fault event occurred during the (Zanetti et al., 2025) tests, when a wildfire in the area caused facility operators to close the dampers completely, leading the MPC to malfunction as it continued to assume outside air availability. In response, a dedicated wildfire mode was developed for the MPC, by reducing minimum outdoor air constraint to zero and adding a maximum outside air constraint also equal to 0, allowing the MPC to correctly account for the requirement to keep outside air dampers closed.

Together, these two fault events demonstrate how targeted adjustments to a small set of key MPC parameters, potentially triggered by automated alarms, can effectively mitigate extreme scenarios such as critical equipment failures and severe

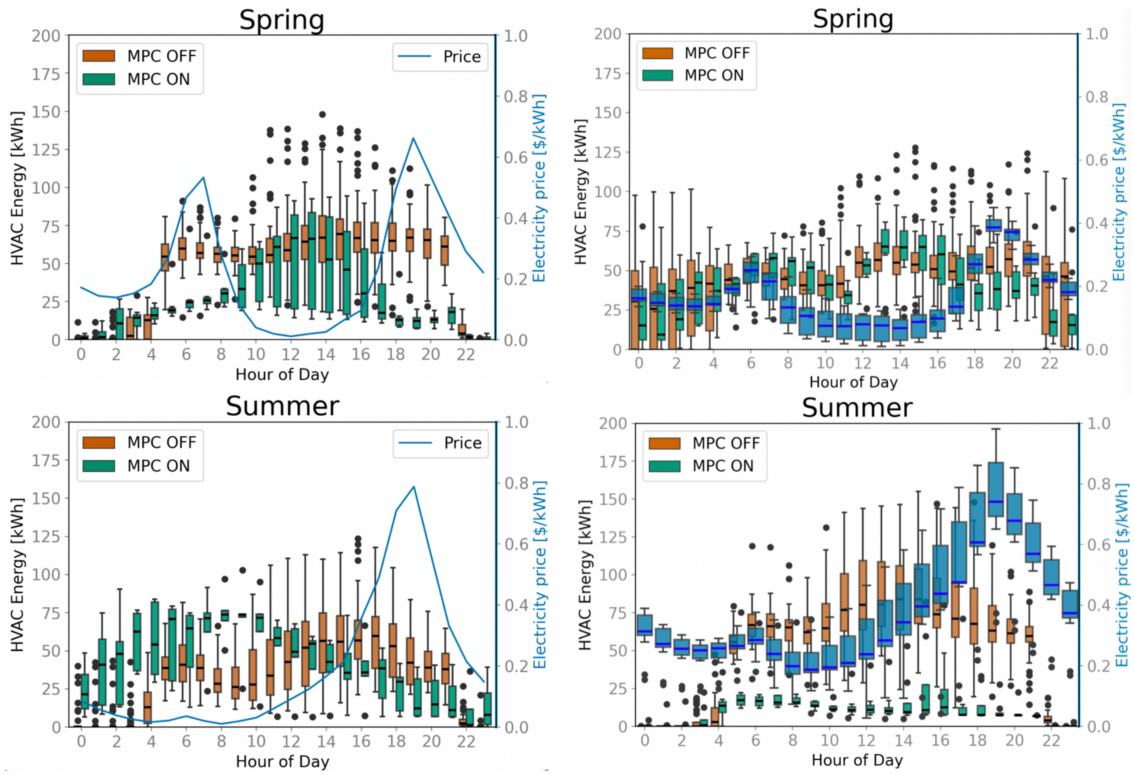

Figure 4: Hourly distribution of HVAC energy use for each day of the analysis period, on the left for Spring, top, and Summer, bottom, in (Zanetti et al., 2025) (3/1/24–5/30/24, 6/1/23–8/27/23), and on the right Spring integrating semantic modeling (3/1/25–5/30/25, 6/1/25–8/31/25)

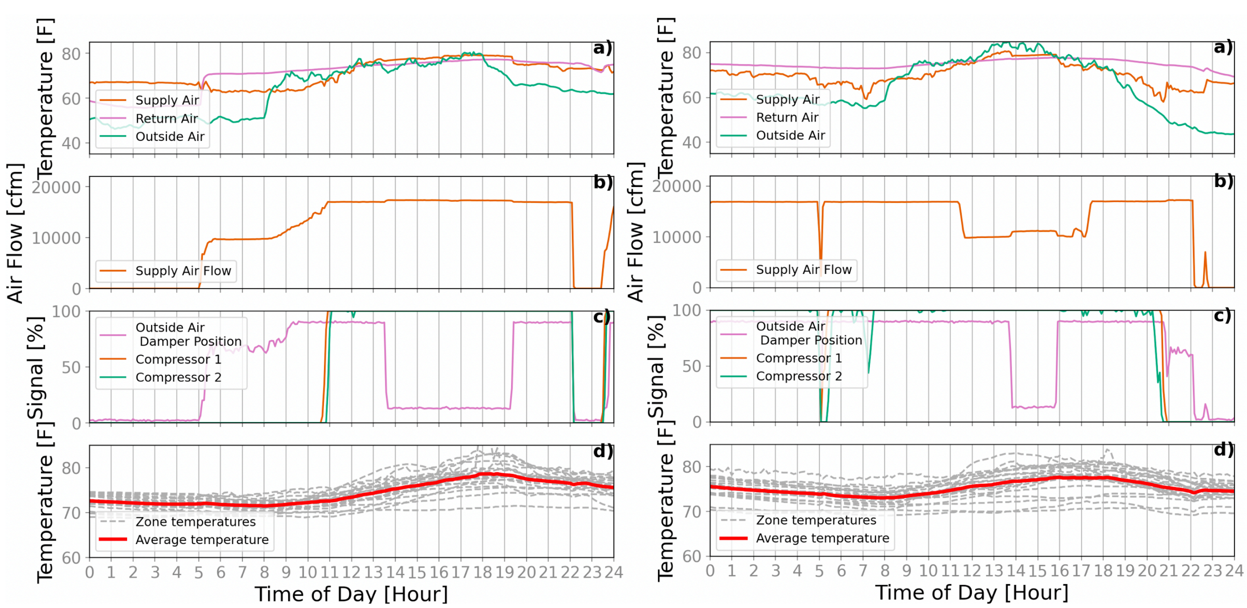

Figure 5: RTU1 warm day comparison with broken compressor MPC ON without updated constraints on the left (3/24/25) and MPC ON on with updated constraints the right (3/25/25). Plot a) shows Outside (green), Supply (orange) and Return (violet) air temperatures. Plot b) shows Supply (orange) air flow rate. Plot c) shows Compressor 1 (orange), Compressor 2 (green), and outside air damper (0 closed, 100 open) control signals. Plot d) shows Zone Temperatures (dashed grey) with Average (red).

weather events, improving overall building resilience.

4.3 Implementation Effort

Despite operational successes, we think the engineering labor required for MPC remains a primary barrier to commercialization. Table 1 details the estimated labor breakdown across projects, quantifying the person-days required for the original MPC development and subsequent follow-up efforts. Notably, a considerable portion of time was spent on data collection, mapping, and software hardening against failures, rather than on the MPC algorithm itself. The use of MPCPy partially mitigated the modeling burden, suggesting that a robust library of reusable MPC models and frameworks could significantly accelerate development. Live monitoring while MPC was in control also consumed considerable effort, though this decreased in follow-up tests as confidence in system reliability grew and error-handling infrastructure matured. These considerations align with complexities recently characterized in recent publication on MPC challenges (woo Ham et al., 2024; Henze, Kircher, & Braun, 2025).

To address the labor-intensive nature of data mapping, we evaluated the use of a semantic model using ASHRAE s223 in the latest tests. While semantic frameworks successfully unified control codes and eliminated hard coded zone logic, replacing hundreds of lines of code with semantic queries, practical limitations emerged during field deployment. Functional testing in March 2025 revealed latency when executing SPARQL queries across the semantic graph, with runtime ranging from minutes to hours due to outdated library versions and model size. This highlights that expensive graph queries over large semantic models are not suitable for real-time closed-loop control, and instead queries should be carried out once to create dedicated lookup tables, during model development, or when there are any changes in the building data pipeline.

Finally, MPC is not a standalone technology, its scalability depends heavily on the building technological readiness. Successful deployment relies on initial commissioning, sensor validation, and data mapping, tasks increasingly performed by Energy Management and Information Systems (EMIS) providers and retro-commissioning efforts (Lin, Kramer, Nibler, Crowe, & Granderson, 2022). EMIS platforms with advanced two-way communication capabilities offer a natural pathway for MPC deployment by establishing well-defined data streams and fully commissioned HVAC networks. However, a barrier persists around control responsibility and liability: while EMIS providers traditionally offer monitoring and fault detection without directly actuating equipment, MPC inherently assumes active supervisory control, transferring operational risk to the provider. The absence of a clear business and liability model for this shift in responsibility remains a source of risk aversion and an obstacle to wide-scale deployment.

5. FUTURE OUTLOOK WITH GENERATIVE AI

The B59 deployment demonstrates that while MPC is highly effective, the associated software engineering, data integration, and labor costs are uncertain, hindering scalability. The traditional paradigm relies on a “PhD-in-the-loop”—specialized engineers manually mapping metadata, tuning models, resolving API errors, and writing custom control logic. The emergence of agentic AI and Large Language Models (LLMs) offers a pathway to partially replace this bottleneck with a collaborative team of specialized, autonomous agents, as summarized in Figure 6. However, this shift introduces its own challenges that must be carefully addressed before autonomous agents can be trusted with live building control.

The proposed multi-agent architecture decentralizes MPC deployment into discrete, manageable co-current tasks that live with the building. The **data plumber** agent parses raw BMS tag lists and mechanical drawings to automatically generate standardized knowledge graphs (e.g., ASHRAE s223, Brick), connects all the other agents and generates journals for permanent memory, while the **stress test pilot** agent leverages frameworks like BOPTEST (Blum et al., 2021) to autonomously configure and execute simulations to test and validate the MPC output. The **control alchemist** agent addresses the bottleneck of developing control models starting from a validated library of equipment templates to assemble the correct optimal control problem for the building based on user requirements, like energy minimization or load shifting. Finally the **guardian of the loop** agent creates dashboards, alarms, and wikis to make human monitoring more effective, it identifies current issues with the deployment and decides whether an autonomous solution is possible or it should be raised to human supervision based on a predefined decision tree to ensure critical errors are always stopped and addressed by human supervision.

Despite its promise, this architecture faces open challenges for research. LLM hallucination risk in control-critical contexts remains a serious concern, a confidently generated but subtly incorrect setpoint or control sequence could cause equipment damage, occupant discomfort, or safety incidents before the guardian of the loop intervenes. This

Table 1: MPC implementation efforts for the different tests

Project and Category	Subtasks	Person days
Energy Minimization MPC (Blum et al., 2022)		239
Data and preparation	System design analysis, Data collection and analysis, New sensor installation, Occupant survey	70
Model development and integration	Thermal envelope, HVAC, Internal load forecast, Weather forecast	79
Controller software development	Architecture design, Service implementation, Log and error handling processes	30
Controller deployment	Facilities coordination, BMS logic editing, Functional testing, Commissioning and maintenance, Performance evaluation	60
Dynamic Price Load Shifting MPC (Zanetti et al., 2025)		88
Software and API updates	Learning code base, integrating CalFlexHub/MIDAS electricity price signals, replacing weather APIs, updating schedule and setpoint forecasting, adding wildfire mode	40
Coordination and logic edits	Monitoring data quality, editing BMS logic for unoccupied periods	23
Testing and deployment	Functional testing and commissioning, Performance monitoring and evaluation	25
Semantic-Integrated and Weekly Prices MPC		27
Semantic model development	Data acquisition, becoming familiar with the building and continuously updating the model	10
Semantic model integration	Replacing hard-coded data querying with ASHRAE 223P semantic queries, Resolving query latency	7
Testing and deployment	Functional testing and commissioning, Performance monitoring and evaluation	10

agent also relies on LLM reasoning chains, meaning the self-correction loop inherits the same failure modes it is designed to catch, particularly in novel or edge-case scenarios underrepresented in training data. Finally, replacing the "PhD-in-the-loop" with an "AI-in-the-loop" does not resolve questions of liability and trust, if anything, explainability and auditability become more critical, as facility managers and building owners must be able to interpret and validate autonomous control decisions. Addressing these challenges will require robust human oversight mechanisms, rigorous simulation-based validation through digital twins, and the development of clear accountability frameworks before agentic MPC can achieve safe, widespread deployment.

6. CONCLUSIONS

The MPC deployment at Building 59 that has run from 2020–2025 demonstrated considerable effectiveness in energy savings, cost reduction under dynamic pricing, load shifting, and fault tolerance that mitigated mechanical failures with minimal human intervention. Commercial scalability, however, remains a challenge. The estimated 327 person-days of effort were dominated by data mapping, API maintenance, and software hardening, though it should be noted that this reflects the particular context of a single-building research deployment; MPC control development for modeling and optimization would likely represent a comparably significant burden if MPC were extended to additional buildings. Liability shifts from building owners to third-party providers also remain an unresolved concern. More broadly, the cost-benefit case for MPC depends heavily on economic incentives such as price variability and demand charges, and additional field demonstrations across diverse building types, climates, and pricing contexts are needed to generate reliable implementation cost targets and payback estimates.

Several complementary paths forward can improve MPC's commercial viability. Businesses, unlike research institutions, can streamline deployment processes across successive projects, progressively reducing per-building implemen-

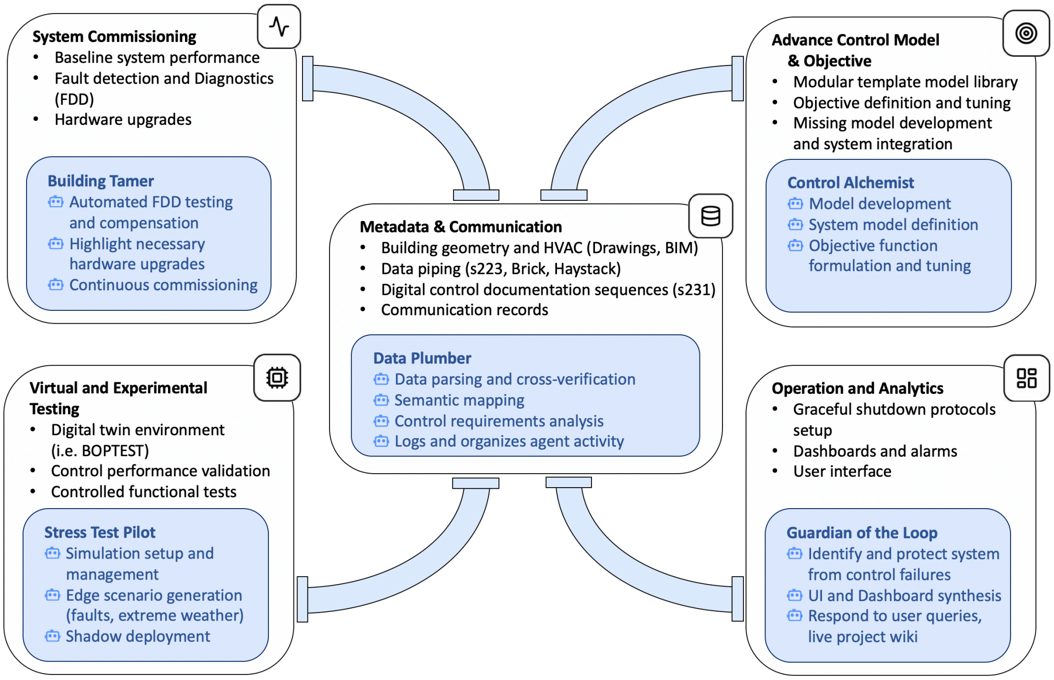

Figure 6: This diagram show the typical development pathway for advanced control in buildings and how it can be partially automated through agentic AI. The bullet points in white show the information and actions required to complete the given task, while the bullet points in blue highlight the actions that could be partially or completely be carried out by AI agents.

tation effort. Workforce development that improves MPC literacy among engineers, operators, and building owners would lower adoption barriers, while robust data pipelines with alarm systems, watchdogs, and redundant streams would reduce long-term maintenance burden. Semantic data models and reusable MPC templates for common HVAC configurations can further reduce deployment complexity. Generative AI and agentic workflows offer additional promise for partially automating deployment, though hallucination risks and accountability frameworks must be addressed before such tools can be safely trusted with live building control. Finally, expanding MPC value propositions to include grid interaction, resilience during extreme events, and indoor air quality management would strengthen the case for adoption and open new avenues for future research.

NOMENCLATURE

API	Application Protocol Interface
AI	Artificial Intelligence
BMS	Building Management System
DDP	Demand Decrease Percentage
DF	Demand Flexibility
DX	Direct Expansion
EMIS	Energy Mangement and Information System
GenAI	Generative AI
HVAC	Heating Ventilation and Air Conditioning
LBNL	Lawrence Berkeley National Laboratory
MPC	Model Predictive Control
RL	Reinforment Learning
RTU	Rooftop Unit
UFAD	Underfloor Air Distribution System
UFT	underfloor thermal units
WSHP	Water source heat pump

REFERENCES

ALC. (2026). *Alc.* Retrieved from https://www.automatedlogic.com/en/products/webctrl-building-automation-system/

ASHRAE. (2018). *Ashrae's bacnet committee, project haystack and brick schema collaborating to provide unified data semantic modeling solution.* Retrieved from Availableonlinehttps://www.ashrae.org/news/esociety/ashrae-bacnet-committee-works-with-other-organizations-on-new-standard

Blum, D., Arroyo, J., Huang, S., Drgoňa, J., Jorissen, F., Walnum, H. T., … others (2021). Building optimization testing framework (boptest) for simulation-based benchmarking of control strategies in buildings. *Journal of Building Performance Simulation*, *14*(5), 586–610.

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.

Blum, D., & Wetter, M. (2017). MPCPy: An open-source software platform for model predictive control in buildings. *Proceedings of the 15th IBPSA Conference*, 1381-1390.

CalFlexHub. (2024). *Calflexhub.* Retrieved from https://calflexhub.lbl.gov/

DarkSky. (2026). *Darksky.* Retrieved from https://darksky.net

EIA. (2018). *Cbecs table c1. total energy consumption by major fuel and table e1. major fuels consumption by end use.* Retrieved from Availableonlinehttps://www.eia.gov/consumption/commercial/data/2018/

EIA. (2025). *Table 1.2. summary statistics for the united states, 2013 - 2023.* Retrieved from Availableonlinehttps://www.eia.gov/electricity/annual/html/epa_01_02.html

Henze, G. P., Kircher, K. J., & Braun, J. E. (2025). Why has advanced commercial hvac control not yet achieved its promise? *Journal of Building Performance Simulation*, *18*(2), 217–228.

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.

Lin, G., Kramer, H., Nibler, V., Crowe, E., & Granderson, J. (2022). Building analytics tool deployment at scale: Benefits, costs, and deployment practices. *Energies*, *15*(13), 4858.

Liu, J., Piette, M. A., Pritoni, M., Nordman, B., Grant, P., Smith, S., … others (2024). Assessing customer experience and business models around price-to-device communication and smart control pathways in calflexhub. In *Aceee summer study on energy efficiency in buildings 2024.*

MIDAS. (2024). *Midas.* Retrieved from https://midasapi.energy.ca.gov/

NOAA. (2026). *National oceanic and atmospheric administration.* Retrieved from https://www.noaa.gov/

SkyFoundry. (2026). *Skyspark.* Retrieved from https://skyfoundry.com/product

Tang, H., Wang, S., & Li, H. (2021). Flexibility categorization, sources, capabilities and technologies for energy-flexible and grid-responsive buildings: State-of-the-art and future perspective. *Energy*, *219*, 119598.

W3C. (2026). *Sparql 1.1 query language.* Retrieved from https://www.w3.org/TR/sparql11-query/

Wan, H., Zhang, J., Chen, Y., Xu, W., & Feng, F. (2025). Exploring gen-ai applications in building research and industry: A review. In *Building simulation* (Vol. 18, pp. 1251–1273).

woo Ham, S., Paul, L., Casillas, A., Prakash, A., Kim, D., Brown, R., … Grant, P. (2024). Practical challenges of model predictive control (mpc) for grid interactive small and medium commercial buildings. In *Proceedings of the 2024 summer study on energy efficiency in buildings* (pp. 1–15). ACEEE.

Zanetti, E., Blum, D., Fu, H., Weyandt, C., Pritoni, M., & Piette, M. A. (2025). Commercial building hvac demand flexibility with model predictive control: Field demonstration and literature insights. *Energy and Buildings*, *345*, 116097.

ACKNOWLEDGMENT

This work was supported by the Office of Critical Minerals and Energy Innovation, Building Technologies Office, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231, and by the California Energy Commission and the California Load Flexibility Research and Deployment Hub. The authors would like to thank Christopher Weyandt, Norm Bourassa, John Elliott, Benjamin Maxwell and Raphael Vitti from Lawrence Berkeley National Laboratory, USA very much for their support of this study, access to facilities and data, and help in the overall implementation and testing of the MPC controller in Building 59.