From Model-Centric to Data-Centric: A Practical MPC Implementation Framework for Buildings

Sicheng Zhan
szhan@nus.edu.sg
National University of Singapore

Matias Quintana
maqr@nus.edu.sg
National University of Singapore

Clayton Miller
clayton@nus.edu.sg
National University of Singapore

Adrian Chong
adrian.chong@nus.edu.sg
National University of Singapore

ABSTRACT

The potential of using Model Predictive Control (MPC) to improve building operation has been shown in many studies. Unfortunately, real-world applications are still restricted by the high implementation cost and the unguaranteed profitability. In the traditional paradigm of "model-centric" MPC, most effort is devoted to constructing the control-oriented model given specific building properties and data availability. Due to the significant heterogeneity among buildings, the results are hardly reproducible, and a high level of customization is required for each new building. To address this issue, we propose a new "data-centric" approach for MPC, which starts with control-oriented data curation that acquires the necessary and cost-effective data concerning the intended control purpose and the building characteristics. The foundation of data-centric MPC is a standardized framework to quantify the data requirements and the established relationships between data usage and control performance. Such an end-to-end framework promotes actual MPC applications with controllable costs and reliable outcomes. We use tropical office buildings as an example to consolidate the data-centric MPC framework. Two use cases are provided to demonstrate its benefits. Over 10% of energy saving was achieved without excessive occupant-related data, and occupant-centric control significantly improved the thermal comfort only with proper data acquisition.

CCS CONCEPTS

• **Computing methodologies → Modeling and simulation**; • **Information systems → Data model extensions**.

KEYWORDS

model predictive control, data management, smart buildings, energy efficiency

ACM Reference Format:
Sicheng Zhan, Matias Quintana, Clayton Miller, and Adrian Chong. 2022. From Model-Centric to Data-Centric: A Practical MPC Implementation Framework for Buildings. In *The 9th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation (BuildSys '22), November 9–10, 2022, Boston, MA, USA.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3563357.3564077

BuildSys 2022, November 9-10, 2022, Boston, MA

ACM ISBN 978-1-4503-9890-9/22/11…$15.00
https://doi.org/10.1145/3563357.3564077

1 INTRODUCTION

1.1 Model Predictive Control for Buildings

Model Predictive Control (MPC) is a well-established optimal control approach. Many studies have demonstrated its potential in improving building energy performance. The application scenarios include but are not limited to system efficiency optimization, occupant-centric control, and renewable energy integration. However, there were a small number of actual applications during the past decades. Benndorf et al. [2] attributed the paucity of real-world applications to the high requirements of modeling, expertise, data, hardware, usability, and costs. Considering the significant role of the building section in carbon reduction, it is essential to improve the reproducibility and scalability of MPC in buildings [4].

Unlike the traditional industries of MPC applications, buildings possess heterogeneous properties and are exposed to diverse disturbances that are impractical to fully account for. For example, internal heat gains are typically immeasurable, and it is costly to measure the ambient conditions for each building. The model and control performance could be greatly affected by these factors. Hence, the traditional one-size-fits-all MPC paradigm is not applicable. Significant configuration effort is needed for every new building, and the control performance remains uncertain until field implementations [6]. A reproducible end-to-end implementation framework could address these issues and facilitate real-world applications [3]. Yet, most existing studies focused on proposing algorithms or modeling techniques to fulfil the control objectives for the specific showcase, lacking systematic investigations on the generalizability [2, 11].

1.2 The Role of Operational Data

Building operational data and/or related metadata is an indispensable resource to establish predictive models and enable optimizations for MPC. Data availability and quality affect the downstream configuration strategy and performance. On the other hand, data acquisition and computation costs increase with a larger amount of data [1]. Currently, the data acquisition is usually conducted once when commissioning the buildings without specific purposes [7]. While many data points are not useful for control, gathering the necessary data is labor-intensive, expert-driven, and involves a high level of customization. Blum et al. [3] reported that 30% labor effort of an MPC project was devoted to completing the data preparation. Hence, a standard and fit-for-purpose data collection procedure is desired to apply MPC in practice. It is essential to understand the minimum data requirements and the marginal improvement brought by the additional desirable data.

Alongside a good understanding of the data requirements, a standardized framework is needed to describe the data availability for new buildings so that MPC can be applied in a plug-and-play manner. The increasing deployment of building information modeling (BIM) and building management systems generates miscellaneous data over building life cycles, exerting a big challenge on data management and utilization [9]. Level-of-Detail (LoD) has been applied in practice to define inputs or information requirements of the building elements in BIM[1]. This clear articulation allows model authors to justify what their models can be relied on for, and allows downstream users to clearly understand the usability and limitations of the models. Meanwhile, considerable variations in operational data usage have been overlooked. An LoD counterpart for time-series data is missing to enhance the interpretability and interoperability of data-driven applications in the operation phase.

To summarize, a more generalizable and cost-effective solution is needed to promote MPC applications in actual buildings. We address the challenges with a data-centric framework. Instead of proposing another algorithm, the framework standardizes the general MPC implementation procedure from data acquisition to subsequent configurations. In the rest of this paper, we illustrate the structure and properties of the proposed framework, followed by two case studies to demonstrate its benefits in the context of tropical offices.

2 DATA-CENTRIC MPC FRAMEWORK

Figure 1 represents the traditional workflow of "model-centric" MPC. Operational data acquisition is performed at the beginning, either arbitrarily or for monitoring. The data is used for MPC configurations after processing, where most of the development effort is devoted to constructing control-oriented models given specific building characteristics and data availability [5]. Consequently, while achieving desirable control performance in many proof-of-concept studies, considerable configuration effort is needed to reproduce the results in new buildings.

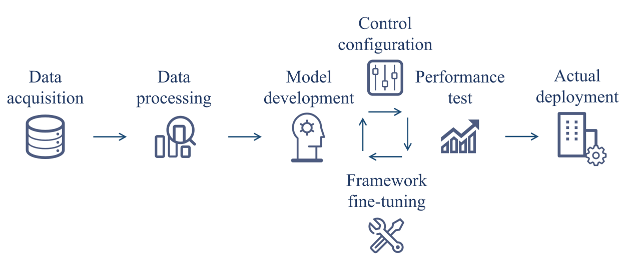

Figure 1: Schematic of the traditional "model-centric" workflow.

In contrast, the proposed data-centric workflow (Figure 2) starts with a control-oriented data curation process that selects the data points to collect regarding the intended control purpose and the building characteristics. Thereafter, the following configurations can be carried on smoothly in a better-defined scenario. Table 1 compares the data-centric workflow with the model-centric MPC. Such an end-to-end workflow avoids the trial-and-error configuration procedure in the traditional paradigm. With better reproducibility, the investment and expected outcome of a control project can be accurately evaluated, which enables larger-scale implementations. The acquisition of unnecessary data points can also be prevented.

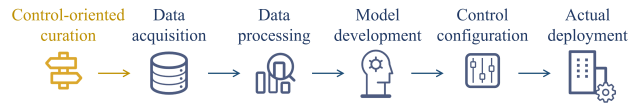

Figure 2: Schematic of the proposed data-centric workflow.

Control-oriented data curation is the first step and the core of the data-centric framework. Two pillars of the curation are a unified framework that describes the data requirements and a systematic understanding of the impact of data on downstream performance. The rest of this section provides more details about the two pillars, as well as how the curation informs model configurations.

2.1 Extended Level-of-Detail

The theme of LoD aligns with the idea of clarifying the required data and further implying the application performance. However, the original LoD definition only covers the static characteristics of building elements, overlooking the considerable variations in the usage of building operational data. Zhan and Chong [11] extended the LoD definition and quantified the availability of operational data. The original LoD denotes different levels with three-digit numbers (such as 300, 350, 500, etc.), which was inherited to respectively represent the increasing levels of time validity, measurement granularity, and temporal resolution. Thereby, the increase of extended LoD generally reflects higher data acquisition costs. The extended LoD fits the need to describe the MPC data requirements and therefore is adopted in the data-centric framework.

While time validity and temporal resolution are usually consistent for all data points in a dataset, the measurement granularity could vary across different objects in a building. Accordingly, the measurement granularity levels were respectively defined for six data categories: energy consumption, indoor condition, internal disturbance, external disturbance, system condition, and envelope condition. The design of control-oriented data curation is also driven by this categorization concerning the diverse data requirements in different categories. More details about the extended LoD and the categorization can be found in [11].

2.2 Understanding the Impact of Data

The impact of a data point on downstream applications is subject to many factors, such as where (which building) and how (control purpose) they are used. Meanwhile, the model and control performance are also affected by other factors, such as climate conditions and model complexities. Therefore, data-centric evaluation is needed to eliminate other affecting factors and investigate the impact of alternative data sources. The experiments are to be conducted under predefined scenarios, where the conclusions will be generalizable.

[1]https://bimforum.org/lod/

Table 1: Summary of the difference between model-centric and data-centric MPC.

	Model-centric	Data-centric
Aim	Apply to a specific building and control task	Generalizable to a type of building and control
Data acquisition	Performed once without specific purpose	Control-oriented data curation
Model selection	Expert-driven trial and error	Designed to match the collected data
Optimization setup	Configured according to the control purpose	
Reproduce effort	Customized from the beginning	Step-by-step guided by the framework
Performance expectation	Unpredictable before field implementation	End-to-end estimated based on data availability

Past studies allude to the two dominant factors affecting the data requirements: `building characteristics` and `data usage`. For example, fan coil unit systems required higher temporal resolution than radiant systems because of the faster system dynamics [11], and granular internal disturbance data was desirable for classrooms given the diverse and irregular daily patterns, but not for offices [12]. These two factors are used to guide the design of experiments for data-centric evaluation and control-oriented data curation, which is further elaborated on in the next section.

2.3 Related Model and Control Configurations

As `data usage` is one important factor that guides the data curation, the goal of applying MPC needs to be specified beforehand. After acquiring the data, control-oriented models can be constructed following the idea of parsimonious modeling, with the downstream performance expectable. Adequate model fidelity and informative data are both needed to form an identifiable problem that yields physically meaningful and extrapolatable models [10, 12]. Take the popular Resistor-Capacitor model, for example, too simple model structures obviously cannot fulfill the requirements. On the other hand, the desirable RC model for identification is much simpler than the standard form as per ISO standard 52016. A 3R2C model structure is typically adequate for a single room. From the practical point of view, any redundancy, either of the degree of freedom in model structure or of the dataset size and dimension, would increase the configuration cost as well as the risk of overfitting.

3 CASE STUDIES IN A TROPICAL OFFICE

The control-oriented data curation can be formalized for a certain type of building as decision-making flow charts for each data category. Figure 3 serves as an example for internal disturbance and indoor condition data in tropical office buildings. A series of factorial experiments were conducted beforehand to support the decision-making, not included here due to the space limit. There are three levels of requirements regarding the `data usage`: basic, modeling and optimization. The required measurement granularity is decided based on specific `building characteristics` or `data usage` conditions. Next, we show the benefit of data-centric MPC by integrating real-world data and simulations in two case studies.

3.1 Case I: Unneeded Internal Disturbance Data

3.1.1 Design of virtual experiments. We constructed and calibrated an emulator using Modelica Buildings library[2] to validate the decision workflow. It functions as an actual office to generate training

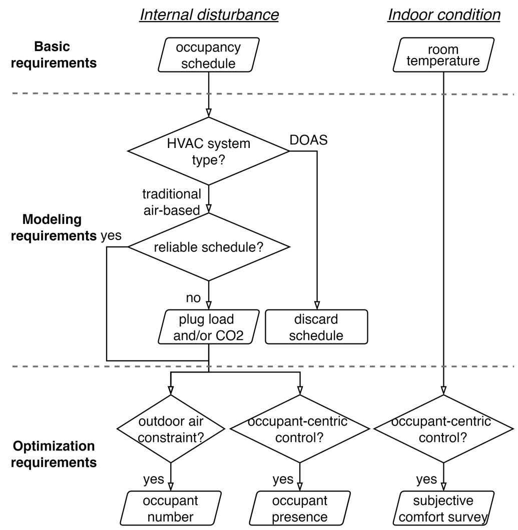

Figure 3: Part of the formalized control-oriented data curation workflow for office buildings in the tropics.

data for controller models and to apply the control actions for performance evaluation. The Modelica-based emulator models the control logic, captures transient thermal response, and has higher fidelity than other modeling tools such as EnergyPlus. To better account for the uncertainties in actual building operations, the emulator incorporates real-world boundary conditions, including weather conditions and internal heat gain sources.

According to the workflow, standard occupancy schedules are sufficient for a typical MPC task in offices with VAV systems. In contrast, plug load and/or CO_2 are usually used for real-time internal disturbance measurements in the model-centric paradigm [11]. For comparison, `schedule`, `plug load`, and `CO2 ppm` were used to identify RC models with the same structure. The models are then applied for optimization as specified in Equation 1, where cooling power P_{cool} and predicted mean vote (PMV), weighted by q_u and q_t, are minimized. The weighting factors were consistent to eliminate their impact on control performance, evaluated by prediction accuracy (RMSE) and control performance (energy and comfort).

$$J = \int_{t_0}^{t_0+30min} \left(q_u (P_{cool})^2 + q_t (PMV)^2 \right) dt \tag{1}$$

3.1.2 Comparable performance with lower cost. Different MPC configurations were compared on ten testing days, and Figure 4

[2]https://simulationresearch.lbl.gov/modelica/

presents the distributions of evaluation results on each day. While adopting the data-centric approach did not involve any real-time measurement for internal disturbance, the prediction and control results of the three alternatives were almost identical. Over 10% of energy saving, compared with the baseline control of constant 24°C setpoint, was achieved by approaching the upper bound of the thermal comfort zone. Hence, the control-oriented data curation reduces the cost of data acquisition and the potential reproducing effort. Besides, avoiding excessive data usage also eliminates the risk of drifted or faulty sensors. Meanwhile, it is worth noting the difficulty of objectively quantifying the exact amount of cost saving, which requires comparative experiments in practice.

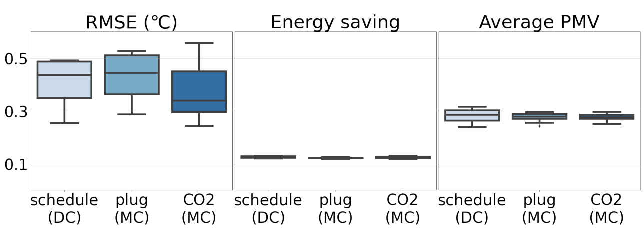

Figure 4: Model accuracy and control performance of data-centric (DC) and model-centric (MC) approaches.

3.2 Case II: Occupant-Centric Control

3.2.1 Time-varying thermal preference. Individual comfort models are typically involved in occupant-centric control. Meanwhile, as different occupants enter and leave a room, the aggregated thermal preference of the cohort changes over time. Based on the data-centric decision workflow, occupant presence and subjective comfort survey are needed to capture the time-varying preference for dynamic optimization. We demonstrate the significance of this process by incrementally integrating the required data points for comparison. To this end, actual personal comfort models and occupant presence data were introduced to the **Case I** virtual testbed.

Six people were assumed to occupy the office based on their presence data. The individual thermal preferences were generated for each occupant based on longitudinal datasets with subjective thermal preference votes [8]. Empirical density distributions of the votes were made based on the indoor temperature at which they were cast, yielding profiles of the desired temperature. The six profiles can be aggregated based on who is in the office and obtain the time-varying desired temperature $T^\star$ for the cohort. The occupant-centric objective function was formed by replacing the PMV in Equation 1 with the deviation from $T^\star$.

3.2.2 Improved thermal comfort. Figure 5 shows the resulting performance of intermediate and ultimate MPC configurations, benchmarked against the baseline control. Simply adding occupant presence data made no difference, and including the static thermal preference lowered the average deviation from the desired temperature with slightly higher energy consumption. The time-varying preference was only accounted for with the occupant presence data, which reduced the deviation by more than 50% with a similar amount of energy consumption. These indicate the necessity of data-centric MPC to fully realize the value of occupant-centric control. Arbitrary data acquisition could lead to a waste of investment.

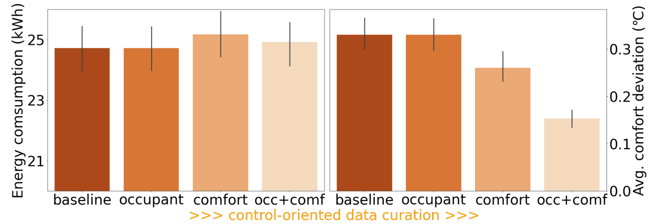

Figure 5: Energy and thermal comfort performance of occupant-centric control compared with baselines.

4 DIRECTIONS FOR FUTURE RESEARCH

This paper proposes a new data-centric MPC implementation framework and demonstrates its usefulness in tropical offices. Future research is needed to establish the cost-effectiveness analysis and consolidate the framework for more control scenarios. Besides, the framework in its current form still requires human interference and expert knowledge. Integrating quantitative metrics could strengthen the framework and automate its ancillary components, such as the data curation and following control configurations.

REFERENCES

[1] Yeonjin Bae, Saptarshi Bhattacharya, Borui Cui, Seungjae Lee, Yanfei Li, Liang Zhang, Piljae Im, Veronica Adetola, Draguna Vrabie, Matt Leach, et al. 2021. Sensor impacts on building and HVAC controls: A critical review for building energy performance. *Advances in Applied Energy* 4 (2021), 100068.

[2] Gesa A Benndorf, Dominik Wystrcil, and Nicolas Réhault. 2018. Energy performance optimization in buildings: A review on semantic interoperability, fault detection, and predictive control. *Applied Physics Reviews* 5, 4 (2018), 041501.

[3] David Blum, Zhe Wang, Chris Weyandt, Donghun Kim, Michael Wetter, Tianzhen Hong, and Mary Ann Piette. 2022. Field demonstration and implementation analysis of model predictive control in an office HVAC system. *Applied Energy* 318 (2022), 119104. https://doi.org/10.1016/j.apenergy.2022.119104

[4] Ján Drgoňa, Javier Arroyo, Iago Cupeiro Figueroa, David Blum, Krzysztof Arendt, Donghun Kim, Enric Perarnau Ollé, Juraj Oravec, Michael Wetter, Draguna L Vrabie, et al. 2020. All you need to know about model predictive control for buildings. *Annual Reviews in Control* 50 (2020), 190 – 232.

[5] Gregor P Henze. 2013. Model predictive control for buildings: a quantum leap? *Journal of Building Performance Simulation* 6 (2013), 157–158.

[6] Donghun Kim, Zhe Wang, James Brugger, David Blum, Michael Wetter, Tianzhen Hong, and Mary Ann Piette. 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 (2022), 119343.

[7] Birgit Painter, Neil Brown, and Malcolm J Cook. 2012. Practical application of a sensor overlay system for building monitoring and commissioning. *Energy and Buildings* 48 (2012), 29–39. https://doi.org/10.1016/j.enbuild.2012.01.003

[8] Matias Quintana, Mahmoud Abdelrahman, Mario Frei, Federico Tartarini, and Clayton Miller. 2021. Longitudinal personal thermal comfort preference data in the wild. In *Proceedings of the 19th ACM Conference on Embedded Networked Sensor Systems*. Association for Computing Machinery, New York, NY, USA, 556–559. https://doi.org/10.1145/3485730.3493693

[9] Shu Tang, Dennis R Shelden, Charles M Eastman, Pardis Pishdad-Bozorgi, and Xinghua Gao. 2019. A review of building information modeling (BIM) and the internet of things (IoT) devices integration: Present status and future trends. *Automation in Construction* 101 (2019), 127–139.

[10] Marija Trčka and Jan LM Hensen. 2010. Overview of HVAC system simulation. *Automation in Construction* 19, 2 (2010), 93–99.

[11] Sicheng Zhan and Adrian Chong. 2021. Data requirements and performance evaluation of model predictive control in buildings: A modeling perspective. *Renewable and Sustainable Energy Reviews* 142 (2021), 110835.

[12] Sicheng Zhan, Yue Lei, Yuan Jin, Da Yan, and Adrian Chong. 2022. Impact of occupant related data on identification and model predictive control for buildings. *Applied Energy* 323 (2022), 119580. https://doi.org/10.1016/j.apenergy.2022.119580