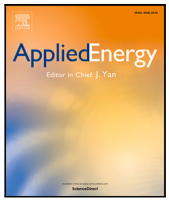
Applied Energy 321 (2022) 119343

Contents lists available at ScienceDirect

Applied Energy

journal homepage: www.elsevier.com/locate/apenergy



Site demonstration and performance evaluation of MPC for a large chiller plant with TES for renewable energy integration and grid decarbonization

Donghun Kim [a,*], Zhe Wang [a,b], James Brugger [c], David Blum [a], Michael Wetter [a], Tianzhen Hong [a], Mary Ann Piette [a]

[a] Building Technology & Urban Systems Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA
[b] The Hong Kong University of Science and Technology, Hong Kong
[c] University of California, Merced, CA, USA

ARTICLE INFO

Keywords:
MPC demonstration
Building optimal control
Model predictive control
District energy system
Carbon reduction
Renewable energy

ABSTRACT

Thermal energy storage (TES) for a cooling plant is a crucial resource for load flexibility. Traditionally, simple, heuristic control approaches, such as the storage priority control which charges TES during the nighttime and discharges during the daytime, have been widely used in practice, and shown reasonable performance in the past benefiting both the grid and the end-users such as buildings and district energy systems. However, the increasing penetration of renewables changes the situation, exposing the grid to a growing duck curve, which encourages the consumption of more energy in the daytime, and volatile renewable generation which requires dynamic planning. The growing pressure of diminishing greenhouse gas emissions also increases the complexity of cooling TES plant operations as different control strategies may apply to optimize operations for energy cost or carbon emissions. This paper presents a model predictive control (MPC), site demonstration and evaluation results of optimal operation of a chiller plant, TES and behind-meter photovoltaics for a campus-level district cooling system. The MPC was formulated as a mixed-integer linear program for better numerical and control properties. Compared with baseline rule-based controls, the MPC results show reductions of the excess PV power by around 25%, of the greenhouse gas emission by 10%, and of peak electricity demand by 10%.

1. Introduction

Energy storage technologies, including water-based thermal energy storage (TES), electrochemical, batteries, pumped storage hydropower, flywheels, and compressed air, provide operational flexibility, load balancing, renewable energy integration and frequency regulation. Of these technologies, the water-based, or chilled water tank, TES is one of the oldest (cool storage was first used commercially in the 1940s, Mitchell and Braun [1]). TES systems can decouple the HVAC energy use from the building's thermal loads and offer many benefits, including energy cost savings, ancillary services benefit, reduction in equipment size and the corresponding capital cost savings (especially when designed as partial storage systems), potentially higher energy efficiency, improved HVAC operation, and resiliency [2, Chapter 51. Thermal Storage]. Many facilities that rely on district energy systems (e.g. large-sized commercial buildings, university campuses and government institutions) already adopted this technology in the era of fossil fuel power plants and provide a key resource for achieving State and Government-mandated clean energy and carbon neutrality goals.

Traditional control strategies for TES cooling plants, i.e., the storage priority or chiller priority controls, fully charge TES using chillers during OFF peak price periods (typically a night time period) while discharging TES to meet cooling load partially or fully during ON peak price periods (typically a day time period). Although the rule-based controls have shown near-optimal behavior and provided benefits to facilities, buildings, and the grid in the past by lowering the total electrical load during peak price periods and hence requiring less spinning/non-spinning reserves and other back-up power plants in the grid, the rapid penetration of renewable energy resources is changing the situation. The intermittent nature of behind-meter renewable energy sources and the grid's movement toward more dynamic price signals challenge building or facility operators to rely on the rule-based operations. To understand impacts on the grid, refer to the CAISO (California Independent System Operator) duck curve depicted in Fig. 1. The graph shows the change of net load profile for a spring day (March 31) as more renewables are being installed. The belly of a duck appears during the mid-afternoon, and the neck of a duck follows during the evening. The belly gets deeper and might touch the baseline

* Corresponding author.
E-mail address: donghunkim@lbl.gov (D. Kim).

https://doi.org/10.1016/j.apenergy.2022.119343
Received 9 November 2021; Received in revised form 11 May 2022; Accepted 23 May 2022
Available online 6 June 2022

Acronyms

ABC	Active chilled beams
ASHP	Air source heat pump
CAISO	California independent system operator
CHP	Combined heat and power
EMS	Energy management system
GDP	Generalized disjunctive programming
GHG	Greenhouse gas
GSHP	Ground source heat pump
HVAC	Heating, ventilating, and air conditioning
ISO	Independent system operator
LEED	Leadership in energy and environmental design
MILP	Mixed integer linear programming
MOER	Marginal operating emission rate
MPC	Model predictive control
NOAA	National oceanic and atmospheric administration
nZEB	Net zero energy building
PV	Photovoltaic
RTU	Rooftop units
SOAP	Simple object access protocol
SOC	State of charge
TABS	Thermally activated building structures
TRL	Technology readiness level
TES	Thermal energy storage

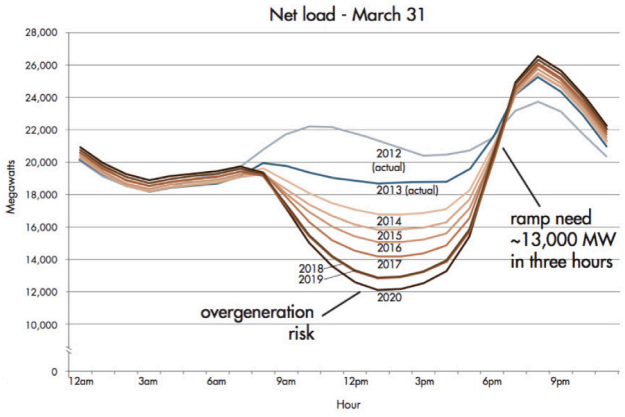


Fig. 1. The official duck chart first published by CAISO in 2013 (What the Duck Curve Tells us about Managing a Green Grid).

power supply from, e.g., the nuclear power plants. It is essential to curtail renewable energy to maintain the base load generators. Fig. 2 shows the monthly wind and solar curtailment for the CAISO from 2019 to 2021 (different colors represent different years). A significant portion of renewable energy generation is currently wasted, and the renewable energy is expected to be curtailed more as more renewable resources are being installed in CAISO territory. The other issue occurs on the neck of a duck: Because of the rapid drop of solar energy, the net load increases quickly. To meet the high ramping rate of the load, dispatchable generators with short response time (order of minutes) such as gas turbine generators have to run and emit significant CO2 (an order of thousands of metric ton CO2 equivalent per hour) during the neck period.

A large number of papers are available in the literature that present advanced control approaches, including model predictive control

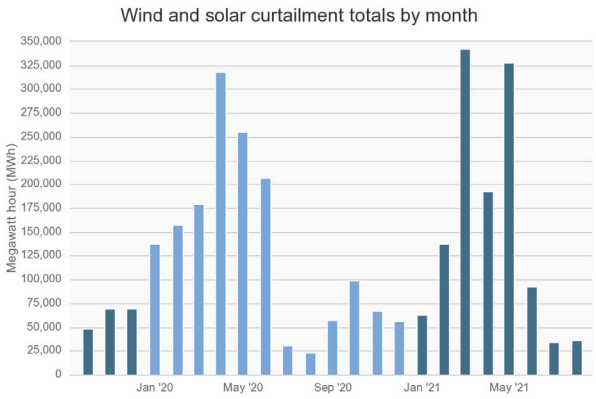


Fig. 2. Wind and solar curtailment totals by month in California ISO. The image is from CASIO (http://www.caiso.com/).

(MPC), for optimal operation of building and district energy systems with the behind-meter renewable generation or/and active energy storage (TES or battery) for renewable energy integration. Reviewed papers are summarized in Table 1, where ASHP stands for air source heat pump, GSHP for ground source heat pump, PV for Photovoltaic, TABS for thermally activated building structures, nZEB for net zero energy building, RTU for Roof Top Unit, ACB for Active Chilled Beams, CHP for combined heat and power, TES for thermal energy storage. Advanced controls, especially MPC, can help to manage energy storage and renewable generation. Kircher and Zhang [3] applied MPC to co-optimize the operation of a chiller plant and ice storage in an office building in New York, reducing the peak demand by 25%. Oldewurtel et al. [4] developed a Sequential Linear Programming based MPC for Residential and office buildings in Zurich, which reduced the peak demand by 3.5% (without battery) and by 17.5% (with battery). Ceusters et al. [5] developed a Mixed-integer Linear Programming based MPC to optimize the operation of a campus-level multi-energy system, including CHP, renewable generation (wind, PV), and battery. Zhang et al. [6] proposed to use Mixed-integer Linear Programming based MPC to operate a residential micro-grid, saving 40% costs compared with conventional rule-based strategy. In addition to MPC, LeBreux et al. [7] presented a fuzzy logic and feedforward controller to resolve the supply and demand mismatch and to minimize the grid dependency using thermal storage for a nZEB. Li et al. [8] proposed reinforcement learning and MPC to mitigate the intermittency of behind-meter renewable generation by optimizing the charging/discharging of a battery based on wind power generation prediction. It was found that MPC can help to smooth wind power scheduling and lower wind curtailment. To cover the wind intermittency, MPC demands 25% less battery capacity compared with a heuristic control algorithm.

Despite many studies of optimal operation of TES with behind-meter renewable and energy storage, a majority of them are using simulations as shown in the table, and therefore real performance of MPC including self-consumption ratio are not clear. Although there are some papers with experimental assessment of MPCs coordinating TES and HVAC systems, they are limited to small-scale systems, e.g., laboratory or small-sized building (e.g., residential) levels. For large-scale systems, there are very few papers that demonstrate an MPC for renewable energy integration and present real site performance. In addition, there is a lack of knowledge about applying MPC to mitigate CO2 emissions [17].

This paper fills the gaps by implementing an MPC for a campus-level cooling TES plant and presenting on-site performance compared with carefully selected, historical rule-based operation data for the plant. The MPC aims at promoting the self-consumption of the on-site renewable and minimizing CO2 in the grid. Because utility cost reduction is one of the key motivations for the facility, the MPC

Table 1
Summary of studies on applying advanced control to optimize the operation of energy storage, renewable and HVAC in buildings.

Study	Building	System	Goal	Control variable	Control approach	Experiment	Results
LeBreux et al. [7]	Building in Canada	TABS, direct solar radiation	Load shifting	On/off of electrical heating	Controller is based on fuzzy logic and feedforward controller	Simulation for the winter season, not-optimized	95% of heating electricity is consumed during off-peak hours
Reynders et al. [9]	Residential in Belgium	TABS, ASHP with PV	Minimize the grid dependency of nZEB	Temperature setpoint	Temperature setpoint is determined by predicted heat loss and peak hour, not-optimized	Simulation in Modelica	Peak use is reduced by 89.2% for the 4 °C comfort band, or 67.3% for the 2 °C comfort band
Kircher and Zhang [3]	Office building in New York	HVAC with ice tank, no renewable	Minimize energy bill	Operation of chiller and ice tank	MPC policy is computed implicitly at each stage through online convex optimization	Monte Carlo Simulation under different weather and utility price structure	25% peak reduction and 50% demand charge saving
Oldewurtel et al. [4]	Residential and office in Zurich	HVAC with battery, no renewable	Minimize energy bill	Building pre-cooling and battery charging/discharging	MPC using state space model for building thermal dynamics, optimization problem formed as Sequential Linear Programming	Simulation	3.5% peak reduction without battery and 17.5% with large battery
Hajiah and Krarti [10] and Hajiah and Krarti [11]	Commercial building in Boulder, CO	HVAC with ice storage (665 kWh)	Minimize energy bill	Building pre-cooling and ice storage charging/discharging	The optimization problem is solved using direct search complex method	Simulation	Combined use of building thermal mass and ice storage system can save up to 40% total energy costs
Ahmad et al. [12]	Not mentioned	ASHP, solar heating, storage tank	Minimize energy bill and comfort violations	ASHP On/Off	MPC formed as a linear programming and quadratic programming	Simulation	No quantified results are provided
Ceusters et al. [5]	Campus with multi-energy system	CHP, wind, PV, and battery	Minimize operational costs	Operation of CHP and battery	MPC is formulated as mixed-integer linear program solved by CPLEX	Simulation	RL can achieve adequately optimal performance for multi-energy management systems as MPC
Li et al. [8]		Wind farm with battery	Maximize the total revenue of the wind farm	Charge/discharge of battery	MPC with four hours wind prediction	Simulation	To cover the wind intermittency, MPC demands 25% less battery capacity compared with heuristic algorithm
Zhang et al. [6]	Residential microgrid	Wind, PV, CHP, battery, TES, EV, HVAC	Minimize operation costs	Scheduling of equipment	MPC formulated as mixed integer linear programming	Simulation	MPC saves 40% costs compared with traditional day-ahead programming strategy under perfect forecasting condition
Zhao et al. [13]	Zero Carbon Building in Hong Kong	CHP, PV, and chilled water storage	Minimize operational costs	Scheduling of CHP and thermal storage	MPC formed as nonlinear programming	Simulation	MPC helps to achieve up to 48% carbon emission reduction, 22% energy saving, and 29% cost saving
Lee et al. [14]	Office building in Tokyo	Chiller with TES	Minimize operating cost	Charging discharging of TES	MPC solved by ϵDE-RJ [15]	Simulation in EnergyPlus	MPC reduced the building operating costs by 3.4% compared to the RBC
Tarragona et al. [16]	Detached house in Spain	ASHP, PV, TES	Minimize operating costs	Operation of HP	MPC formed as mixed integer non-linear programming	Simulation with heating load from EnergyPlus	MPC saves 58% energy cost compared to RBC

is also formulated to reduce peak demand. The technical details of mathematical programming and modeling approaches are not focused on in this paper and we refer to our companion paper for those perspectives. However, in this paper, we present the complete MPC formulation, focus on the site performance evaluation, provide barriers and challenges for a large scale MPC deployment that were identified during this demonstration work, and share lessons-learned to guide and facilitate future development of MPCs for large central plants.

Section 2 describes components, overall configuration and the corresponding baseline control of the district cooling system. In Section 3, the MPC formulation for better renewable energy integration and grid decarbonization is presented, and the control infrastructure and software to supplement and implement the MPC are also included. In Section 4, a selection process of historical baseline data and performance assessment results compared with the baseline data in terms of self-consumption ratio, CO2 emission and peak demand are described. Finally, Section 5 summarizes and shares the lessons-learned to guide and facilitate future development of MPCs for central plants.

1.1. Innovations

(1) Bringing the MPC technology to real world applications and (2) demonstrating the use of district energy TES plants for renewable energy integration and grid decarbonization are innovations because there are few previous similar studies and this work advances the maturity of the MPC technology applied to district energy TES systems. MPC applications for central cooling/heating TES plants have been investigated since the early 1990's [18]. An enormous number of papers have been published and a variety of MPC approaches were introduced accordingly, but the majority of them are still simulation-based studies, despite the long history of MPC in building energy systems (as discussed in the Introduction and in Table 1). The significant, associated problem is that the lack of field demonstrations and unclear demonstrated

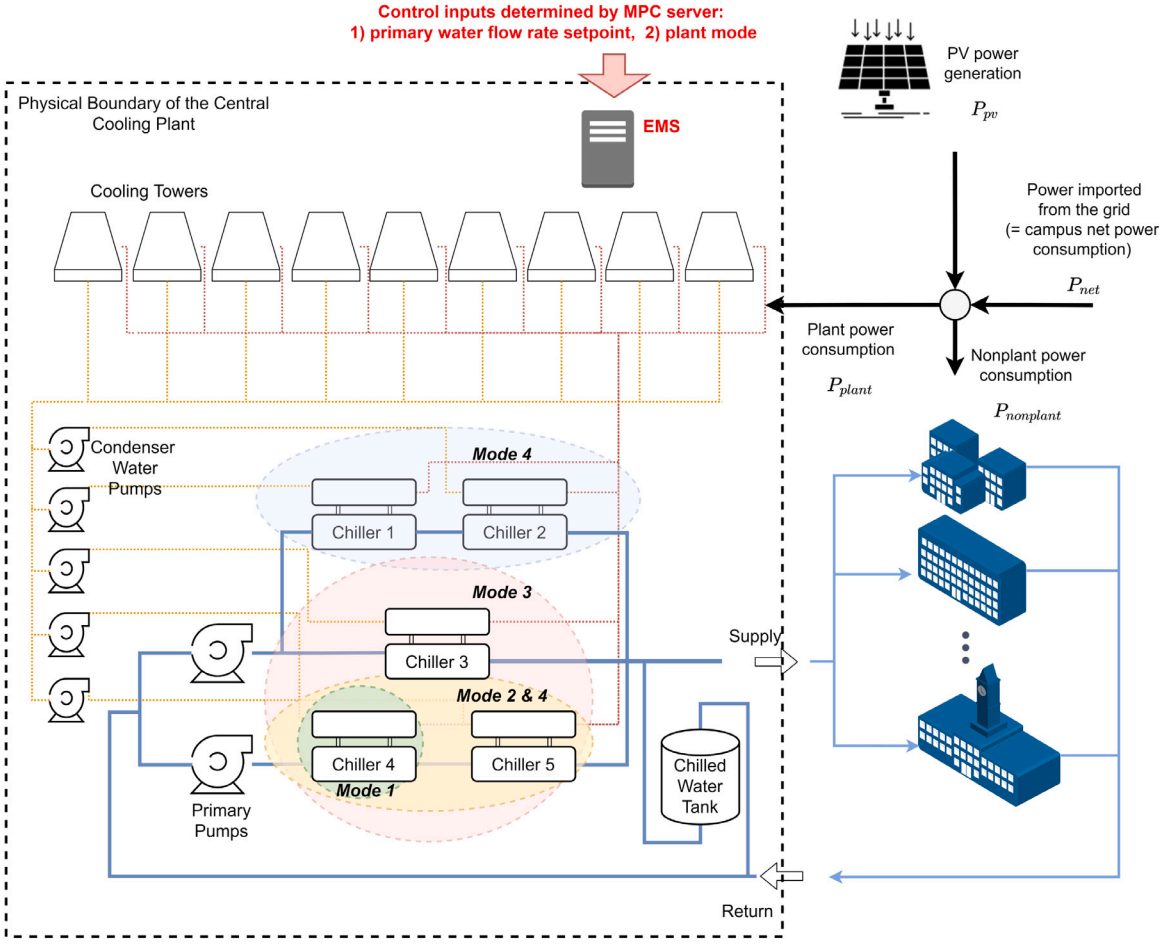


Fig. 3. The equipment configuration and schematic diagram of the central cooling plant at University of California, Merced: the black dashed box indicates the physical boundary of the cooling plant including EMS that manages device level controls. The primary water flow rate setpoint and plant mode marked in red are the variables to be determined by the MPC in this study. Power flow directions at the campus grid connection point (the circle) are shown in the upper right corner: The sum of the imported power from the grid and PV power generation should be the same as the sum of plant and non-plant power consumption.

savings at sites prevent industry from bringing MPC technologies to market. This field study should fill this gap at least partially and significantly advance the Technology Readiness Level (TRL) of MPC technology applied to district energy systems from a low level (TRL 3: proof of concept) to a higher level (≥TRL 7: system prototype demonstration in an operational environment).

2. Site description and baseline control strategy

2.1. Site description

The central plant of the University of California, Merced is the demonstration site located in California, U.S. UC Merced has been on the cutting edge of sustainability in higher education and has set stringent sustainability goals. The campus is green from the ground up, with every campus building LEED (Leadership in Energy and Environmental Design) certified by the U.S. Green Building Council. In addition, the central plant, built in 2005, was the first energy plant in the United States to be certified LEED Gold.

The cooling TES plant, shown in Fig. 3, has five chillers with a total cooling capacity of 17.6 MW, 5000 ton of refrigeration, serving around 40 buildings. It has the primary–secondary (decoupled) chilled water loop configuration, where separate variable speed pumps are dedicated to the primary and secondary water loops. The chillers in the primary water loop, where each chiller has rated cooling capacity from around 1000 to 1500 ton, are high-efficiency units with centrifugal compressors and internal capacity controls using variable speed compressor, inlet guide vane and diffuser modulation and hot-gas bypass for part-load operations. The chiller configuration is a combination of series and parallel connections. Depending on the selection of plant mode, the plant could operate a single chiller, two serially connected chillers, three chillers (two series chillers and one parallel chiller), and four chillers (two parallel chiller groups of chillers where each group has two series chillers). The variable frequency drives for the secondary water pumps adjust the pump speed to regulate the differential pressure between the supply and return pipes to transport the chilled water to multiple building load terminals. A tertiary pump "bridge" distributes the chilled water within each building.

Nine cooling towers and five condenser water pumps shown in Fig. 3 deliver the heat from the condensers to the ambient air. The operation of condenser water pumps is interlocked with the plant mode, i.e., the energy management system (EMS) switches on/off some of them according to the plant mode. The cooling tower fan speeds are modulated to meet the condenser water temperature setpoint determined by the EMS. We refer to Ashrae [2] Chapter 3. Central Cooling and Heating Plants and Chapter 12. District Heating and Cooling for more detailed descriptions of components and their functionalities, and Ashrae [19] Chapter 43. Supervisory Control Strategies and Optimization for the condenser water temperature setpoint control and cooling tower fan sequencing.

A two million gallon chilled water tank (sensible energy storage) is located at the bypass line (or decoupler) between the primary and secondary water loops. The tank is a stratified cool storage where warmer and less dense water floats on the top of colder and denser

water. The two layers are separated by the thermocline where the water temperature changes abruptly, and water temperature is uniformly distributed in each layer. When the tank is discharged, return water from building terminals fills the tank from the top pushing the colder water out of the tank. On the other hand, when the tank is charged, the chilled water fills from the bottom of the tank pushing the warm water out of the tank. There are no isolation valves on the bypass line, and therefore the mismatch between primary water flow rate and secondary flow rate determines the charging or discharging flow rate. When the secondary flow rate is higher than the primary flow rate, i.e., the deficit flow condition, the return water goes into the top of the tank making discharge. The opposite causes charging. In order not to disturb the thermocline in the water tank, the chilled water supply temperature setpoint should be maintained at around 39 °F (around 4 °C). This constraint on the chiller leaving temperature together with no-isolation valves on the bypass line results in the primary water flow rate setpoint being the only decision variable that defines the charging and discharging heat rates when chillers are running.

In addition, the UC-Merced campus has a solar farm of more than 4 MW which covers 8.5 acres southeast of the campus and an approximately 960-space parking lot. The PV has a single-axis sun tracking system, which captures up to 30% more sunlight than conventional fixed-tilt systems by following the sun throughout the day.

It is important to mention the utility tariff structure which is one of the key driving forces for MPC. Because of the scale of energy consumption, the energy price rate is much more complex than a typical TOU (time of use) structure. It purchases energy from selected energy service providers within the energy wholesale market through a negotiation process. This program makes (1) the time variation of the energy price rate small, and (2) the demand charge is much more dominant than the energy charge at the site.

2.2. Energy management system and control sequences

The primary water flow rate setpoint and plant mode are the variables determined by the MPC in this study. As shown in Fig. 3, those are transferred to the plant's EMS which consists of a number of device-level controls and manages operations of system components. The local controller for the primary water pumps in the EMS, which is a mixture of proportional–integral and ON/OFF controls, selects the right pump(s) according to the plant mode and adjusts pump speed(s) through the variable frequency drive(s) to meet the setpoint. In addition, once the plant mode signal is received by the EMS, internal logic of the EMS eventually triggers switches to turn on and off the chillers that correspond to the plant mode (See Fig. 3 for chiller(s) for each mode). It has to be mentioned that the EMS also has complex logic (not shown in Fig. 3 due to the complexity) that decides operations of isolation valves, cooling towers and condenser water pumps which are interlocked with each other for each plant mode, and that adjusts speeds of condenser pump and cooling tower fans.

2.3. Description of baseline control

The basic operation strategy for the cooling TES plant is the storage priority control that runs chiller(s) during the nighttime to store cooling energy in the chilled water tank and uses the cooling energy during the day time to fully meet the daytime campus cooling load. Fig. 4 shows a representative plant electrical power profile of the chiller plant operation. To generate this plot, the daily profiles of the total plant electrical power consumption, which aggregates chiller compressor powers, cooling tower fan powers, pump powers for the primary and condenser water pumps, over the month of August were collected and averaged. Note that the plant power is zero from 12:00 to 17:00 h. This means the chilled water tank completely serves the campus load, which clearly characterizes the control and system as the storage priority control and full storage system. It is important to remember that the storage priority control in general does not account for behind-meter, on-site renewable energy.

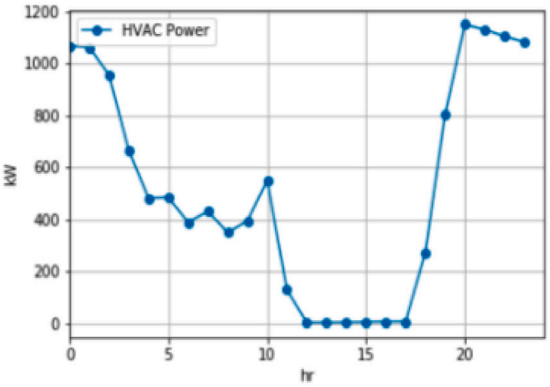


Fig. 4. Daily-averaged plant power at the UC-Merced.

3. MPC design and interface to EMS

3.1. Strategy to consider grid's carbon condition for a site-specific MPC

It is important to describe how to incorporate a grid's GHG (greenhouse gases) condition to a site-specific MPC. The grid system-wide emission rate in a specific grid region depends on the total power production rate from grid power generators, and other factors that affect system operating conditions, such as weather. The marginal operating emissions rate (MOER) is the partial derivative of system-wide emission rate with respect to the total production rate [20]. It means the change of the emission rate in the grid region with respect to the last megawatt produced by dispatchable generators having the unit of metricTon CO2-equivalent per MWh [mTonCO2e/MWh]. Intuitively, this indicates how much carbon emission rate increases/decreases in a grid region when one consumes one megawatt more/less. Therefore, MOER allows for associating the power usage at a specific site with the carbon emission rate in the grid region by simply multiplying the on-site power consumption with the MOER signal.

In this paper, we used the MOER signal calculated by WattTime, based on a proprietary model that extends the basic methodology used by both Siler-Evans et al. [21] and Callaway et al. [20] but adapted for real-time use. WattTime calculates these marginal operating emission rates in real-time, every 5 min using a combination of grid data from the respective ISO and 5 years of historical Continuous Emissions Monitoring System data [22].

3.2. MPC formulation

The MPC tested at the field site has a form of the Mixed Integer Linear Programming (MILP) problem and is presented in this Section. Let N_p and k ($\in \{0, \ldots, N_p - 1\}$) be the prediction horizon and the kth prediction step. The essential control variables are the chiller plant mode and primary water flow rate setpoint as discussed in Section 2.1. Let j and J represent a plant mode (see Fig. 3 for the mode enumeration) and the total number of plant modes. Define $s_j[k]$ ($\in \{0, 1\}$) as the indicator for ON/OFF status for the jth plant mode at time k. Denote also actions of the start-up and shut-down of the jth mode at k as $\delta_{ON,j}[k]$ ($\in \{0, 1\}$) and $\delta_{OFF,j}[k]$ ($\in \{0, 1\}$), respectively. That is, $\delta_{ON,j}[k] = 1$ implies that the jth mode was OFF at least the previous time step $(k-1)$ and is about to turn ON at k.

The MPC seeks to find the optimal sequence of plant mode and chiller load which minimizes both the peak net power consumption and carbon emission (with some weights) over the prediction horizon while meeting campus cooling load and maintaining the state of charge (SOC) within a certain level at any point of time. The net power

consumption means the total power consumption ($P_{plant} + P_{nonplant}$)[1] subtracted by the renewable generation (P_{pv}) which can be measured at the grid connection point. To promote self-consumption of the on-site PV generation, the MPC is formulated to get no incentives for negative net power consumption (that is, the excess renewable energy generation). The formulation is summarized below.

$$min \quad \sum_{k=0}^{N_p-1} \left(E[k] \times P^{+}_{net}[k]\right) + \omega_d \times d + \omega_x \times \upsilon_x \tag{1}$$

$$s_j[k] - s_j[k-1] = \delta_{ON,j}[k] - \delta_{OFF,j}[k] \tag{2}$$

$$\delta_{ON,j}[k] + \delta_{OFF,j}[k] \le 1 \tag{3}$$

$$\sum_{l=0}^{OT_j-1} \delta_{ON,j}[k-l] \le s_j[k] \tag{4}$$

$$\sum_{l=0}^{OT_j-1} \delta_{OFF,j}[k-l] \le 1 - s_j[k] \tag{5}$$

$$\sum_{j=1}^{J} s_j[k] \le 1 \tag{6}$$

$$x[k+1] = Ax[k] + B \begin{bmatrix} Q_{CH}[k] - Q_{BL}[k] \\ T_{OA}[k] - T^0_h \end{bmatrix} \tag{7}$$

$$x_{min}[k] - \upsilon_x \le x[k] \le x_{max}[k] + \upsilon_x \tag{8}$$

$$Q_{CH}[k] = \sum_{j=1}^{J} \nu_j[k] \tag{9}$$

$$Q_{min,j} s_j[k] \le \nu_j[k] \le Q_{max,j} s_j[k] \tag{10}$$

$$\sum_{j=1}^{J} (a_j \nu_j[k] + b_j s_j[k]) + P_{nonplant}[k] - P_{pv}[k] \le P^{+}_{net}[k] \tag{11}$$

$$0 \le P^{+}_{net}[k] \tag{12}$$

$$P^{+}_{net}[k] \le d \tag{13}$$

$$x[0] = z_0 \tag{14}$$

$$0 \le \upsilon_x \tag{15}$$

$$0 \le d. \tag{16}$$

The binary decision variables are

$$s_j[k], \delta_{ON,j}[k], \delta_{OFF,j}[k], \tag{17}$$

and real-number decision variables are

$$Q_{CH}[k], \nu_j[k], P^{+}_{net}[k], x[k], \upsilon_x, d, \tag{18}$$

where $k \in \{0, 1, \ldots, N_p - 1\}$ and $j \in \{0, 1, \ldots, J\}$.

The first term, $\sum_{k=0}^{N_p-1} E[k] \times P^{+}_{net}[k]$, in Eq. (1) represents the total amounts of GHG emission and energy cost over the prediction horizon. $E[k]$ indicates a weighted sum of the energy cost rate [$/kWh] and the MOER. The last two terms of d and υ_x are auxiliary variables meaning the peak power and the constraint violation for SOC, respectively. Each equality and inequality constraint is described as follows.

- Eq. (2) specifies the relationship between the ON/OFF status and start-up/shut-down variables.
- Eq. (4) enforces the minimal ON time (OT_j) for the jth plant mode. That is, once the mode is turning ON, it has to stay ON for the time period.
- Eq. (5) describes the minimal OFF time constraint (the minimal OFF time was set as the minimal ON time).
- Eq. (6) means that at most one plant mode is feasible for each time.
- Eq. (7) describes the constraint that the sum of chiller load and TES discharging heat must meet buildings' cooling load Q_{BL} at each time step, and the dynamics of the state of charge (SOC, x) in the chilled water tank. A simple, 1-dimensional thermal network model consisting of one resistance and capacitance was developed for the tank and those parameters were estimated with measurements. Then, the continuous LTI was discretized in time. (A, B) are the system matrices of the discrete LTI model. See Appendix A.1 for more detailed descriptions.
- Eq. (8) represents SOC constraint. This also defines the auxiliary variable of υ_x which is introduced to ensure a non-empty feasible set of the optimization problem.
- Eq. (9) defines auxiliary variables, ν_j, to represent the total chiller load Q_{CH}.
- Eq. (10) is an algebraic expression of the following logic: IF the jth plant mode is OFF, THEN the total chiller load (Q_{CH}) is 0. IF the jth plant mode is ON, THEN the total chiller load is bounded by $Q_{min,j}$ and $Q_{max,j}$. Note that when the jth mode is ON, these constraints make $\upsilon_{i,i\neq j} = 0$, and $\upsilon_j = Q_{CH}$ from Eq. (9).
- Eqs. (11) and (12) defines the auxiliary variable P^{+}_{net} representing the positive net power consumption. Chiller plant power (not chiller power) is modeled as the sum of the chiller plant power at each mode as shown in the underlined term. a_j is a constant and b_j is a function of wet-bulb temperature. The pair of (a_j, b_j) was obtained from the historical data for each plant mode using the Least Squares Method. See Appendix A.2 for more details.
- Eq. (13) defines d as an upper bound of the positive net power. This is to minimize the maximum power consumption over the prediction horizon using the epigraph formulation [23].

The MILP formulation adopted the Generalized Disjunctive Programming (GDP) and the Convex-Hull reformulation approaches [24–26] developed in the field of chemical processing, mathematical programming approaches for the Unit Commitment problem [27–29] in the electrical power system, and a linear reformulation method. We refer to our companion paper for those technical details.

Note that the MPC formulation only requires modeling TES, i.e., (A, B) in Eq. (8), and the plant power for each plant mode, i.e., (a_j, b_j) in Eq. (10), not individual components of chillers, cooling towers and pumps, and control sequences. This substantially reduced the sensor requirement and effort to implement the MPC compared to one that starts from individual component models. In addition, the MILP formulation allows achieving global optimality, feasibility, computational efficiency and stability. The drawbacks of this simplified modeling approach are (1) potential inaccuracy on the plant power prediction, (2) potential loss of energy savings by not optimizing the condenser/chilled water supply temperature setpoints, and (3) unclear ways of determining the lower and upper bounds of the chiller capacity for each plant mode (because chiller capacity varies with condensing and evaporating temperature). However, (1) the proposed modeling approaches resulted in reasonable accuracy (around 20% relative error over a wide range of operating conditions), (2) for many large cooling plants, heuristic control sequences that determine supply water temperature setpoints are already near energy-optimal (e.g., set the condenser water temperature setpoint near wet-bulb temperature and set the chilled water supply temperature setpoint as high as possible up to the point where the thermocline is maintained, and (3) facility operators have prior-knowledge for reliably setting those capacity bounds.

None of building level devices such as thermostats and air handling unit (AHU) supply fans were considered as controllable loads in the MPC problem. This was because the central plant energy management system (EMS) is not fully integrated with building level EMSs. Therefore, uncontrollable loads in this study are buildings' cooling load (Q_{BL}), non-plant electrical load ($P_{nonplant}$) and PV generation (P_{pv}), which are exogenous inputs to the MPC problem as indicated in Eqs. (8) and (11). On the other hand, the controllable load is the thermal load

[1] $P_{nonplant}$ represents uncontrollable electrical load occurring outside of the cooling plant such as building-level fan powers, lighting and plug loads.

for chillers (Q_{CH}). The controllable load would not affect comfort because of the hard constraint that the sum of chiller loads and tank discharging rate must meet the total buildings' cooling load.[2]

3.3. Selection of control design parameters

The hyperparameters of the MPC are weighting factors of ω_d and ω_x in Eq. (1), prediction horizon N_p and time step Δt at which an optimal decision is updated. Like other control design practices, those were configured by a combination of some parametric studies, physical insights for the system and previous studies in the literature as follows.

A large number was assigned to ω_x. This is because ω_x is essentially the penalty coefficient of the penalty function method that replaces a hard constraint with a penalty function multiplied by the penalty coefficient. It penalizes the SOC constraint violation as described in Section 3.2. Since the solution of a penalty problem converges to the optimal solution of the original, hard constraint problem as the coefficient goes to infinity [30, Chapter 9], the solution must not be sensitive to the coefficient once it is sufficiently large enough. In addition, as described at the end of Section 2.1, the demand charge is much more dominant than the energy charge. The site characteristic requests put the peak demand reduction first and therefore we set a high value to ω_d. After carrying out a few parametric studies, the value of 200 was selected for both parameters near which optimal decisions do not vary significantly: the relative 2-norm error between the optimal trajectories for $\omega_x = \omega_d = 200$ and $\omega_x = \omega_d = 500$ was less than 0.01% for all decision variables.

For choosing an appropriate time step, we considered (1) a better response to unpredicted events (e.g., abrupt increase/decrease of cooling load) and (2) a time scale separation between the MPC loop (outer control loop that determines the plant mode and flow rate setpoint) and local control loop (inner control loop that adjusts the primary pump speed to track the flow setpoint). For the first purpose, it would be helpful to reduce the time step since the MPC problem can be recalculated with updated measurements from the plant, especially SOC, and forecast data. However, for the second purpose, it is preferred to increase the time step. This is because otherwise instability might be incurred due to dynamic interactions between the inner and outer control loops [31]. With the knowledge of the time constant for the variable speed pump control loop (ranging from 5 to 10 min), a 30 min to an hour was considered to be a proper time step, and the latter was chosen in this study.

The prediction horizon was restricted by the forecast length of the carbon emission signal which is 2 days. Instead of performing a sensitivity analysis to the horizon which is not straightforward for energy systems (because increasing the horizon introduces additional forecast errors to MPC and, for the analysis, the stochastic process of the forecast errors needs to be modeled that depends on underlying forecasting mechanisms and the nature of disturbances), we set it to the upper bound of the 2-day by referring to a previous, comprehensive simulation study [32]. The paper modeled a stochastic process of forecast errors with several different stochastic difference equations and investigated the effect of the size on the MPC performance for TES systems. The results showed that the increasing prediction horizon lowers the cost even considering forecast errors and the MPC performance becomes insensitive as the horizon increases.

[2] The constraint was implicitly defined in Eq. (7) via a variable elimination: see Eq. (21).

Table 2
MPC communication data points.

Read/write	Communication data points	Unit	Symbol
Read	Campus net power consumption	[kW]	P_{net}
Read	Solar generation	[kW]	P_{pv}
Read	Central plant power consumption	[kW]	P_{plant}
Read	State of charge	[%]	x
Read	Outdoor air temperature	[°F]	T_{OA}
Read	Outdoor air wet-bulb temperature	[°F]	T_{wb}
Read	Primary water flow rate	[gpm]	V_p
Read	Chilled water supply temperature	[°F]	T_{CHWS}
Read	Return (secondary) water flow rate	[gpm]	V_r
Read	Chilled water return temperature	[°F]	T_r
Read	Entering chiller temperature	[°F]	T_{CHE}
Write	Primary water flow rate setpoint	[gpm]	$V_{p,SP}$
Write	Plant mode request	[–]	s_j

3.4. Data communication architecture between MPC and EMS

Table 2 shows the complete list of data points to implement the MPC for the campus. The SOAP (Simple Object Access Protocol) is the communication protocol utilized for exchanging data between the MPC server and the energy management system (EMS) at UC-Merced, and the suds python library was utilized for the communication. The MPC server sends read requests for the datapoints, and receives the corresponding data back using the SOAP interface. The collected data was then used to update building load forecasts with weather forecast updates and implement/analyze MPC results.

3.5. MPC implementation and interface

The flow diagram of implementing MPC at the campus cooling TES plant is shown in Fig. 5. At each sampling time ($\Delta t = 1$ h), the MPC sever receives the read datapoints (listed in the Table 2), and two-day ($N_p = 48$) ahead forecasts of ambient temperature using NOAA (National Oceanic and Atmospheric Administration) API, solar generation ($P_{pv}[\cdot]$) using pvlib [33], and MOER ($E[\cdot]$) using the WattTime API described in Section 3.1. The MPC server then predicts campus cooling load ($Q_{BL}[\cdot]$) and non-cooling plant electric power consumption ($P_{nonplant}[\cdot]$) over N_p using a Machine Learning technique (Artificial Neural Network using tensorflow [34]). Then, it solves the MPC problem (Section 3.2) with updated information and determines the sequences of plant mode s_j and chiller load Q_{CH} over N_p. GLPK (GNU Linear Programming Kit) package which is free open-source software developed to solve large-scale linear and mixed integer linear programming problems is utilized as the optimization solver. The first time step decision $s_j[0], Q_{CH}[0]$ is chosen and converted to the primary flow setpoint ($V_{p,SP}$, see Table 2), and post the write datapoints to the EMS. This process was repeated for every sampling time (1 h).

The conversion of the MPC decision variable of $Q_{CH}[0]$ to the EMS input, i.e., the flow rate setpoint for the primary pumps, is straightforward via the energy balance equation since the chilled water supply temperature setpoint is fixed (due to the restriction for maintaining the thermocline), entering chiller temperature can be measured (see Table 2), and desired chiller load was determined. More precisely, it reads the entering chiller water temperature and the chilled water supply temperature setpoint (39 °F), and computes the flow rate setpoint as follows.

$$V_{p,SP} = Q_{CH}/(\rho_{water} \times C_{p,water} \times (T_{ECHW} - T_{CHWS,SP})) \tag{19}$$

where $V_{p,SP}$ is the water flow rate setpoint [gpm] and T_{ECHW} are the entering chilled water temperature and $T_{CHWS,SP}$ is the chilled water supply temperature setpoint for (downstream) chillers. Although the plant mode is updated for the sampling time (an hour), the flow rate setpoint is more frequently updated with a finer timestep resolution (20 min) to consider the time variant entering chilled water temperature while meeting the desired chiller load.

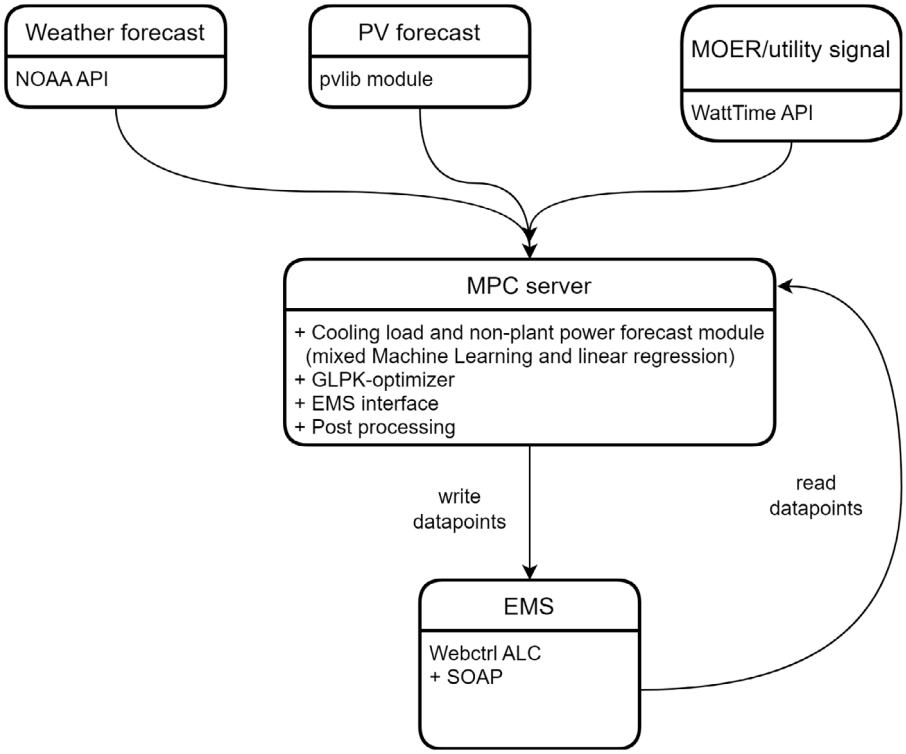


Fig. 5. Flow diagram of MPC implementation at a campus cooling plant.

4. Site performance of MPC

The presented MPC has been implemented for the campus cooling TES plant for several test periods, and sample test results for a week in May 2021 are shown and discussed in this section. We start describing our approach to evaluate performance of MPC compared with the baseline control described in Section 2.3.

4.1. Selection of baseline control period for actual performance assessment

To make fair comparisons between two controllers and evaluate actual performance for the real site, it is ideal to select a baseline period where (1) the baseline control was implemented, (2) uncontrollable disturbances including campus cooling load, MOER signal and on-site solar generation were consistent to that of the MPC period, and (3) initial and final SOC were close enough to compare. Although we have a large period historical data for the baseline control (more than a year), there was no such period where all conditions were met. This is because having a consistent solar generation limits the month of a baseline period (because a different month results in different solar trajectories of solar azimuth and zenith angles as the Earth moves through its orbit), and the grid condition is stochastic (because the MOER is a complex function of the total grid demand, weather, types of operating power generators).

To resolve this, we first selected the same month of a year for the baseline period to that of the MPC period in order to have the same solar angle trajectories. Then, we narrowed down the period such that the outdoor air temperature profile is reasonably consistent with that of the MPC-period. Here, we assumed the daily-averaged outdoor air temperature is a proxy of the campus cooling load. The correlation between daily-averaged campus cooling load and outdoor air temperature is shown in Fig. 6. The strong correlation between them validates our assumption. Then, we assumed that the grid condition during the two control periods were identical, and applied the MOER

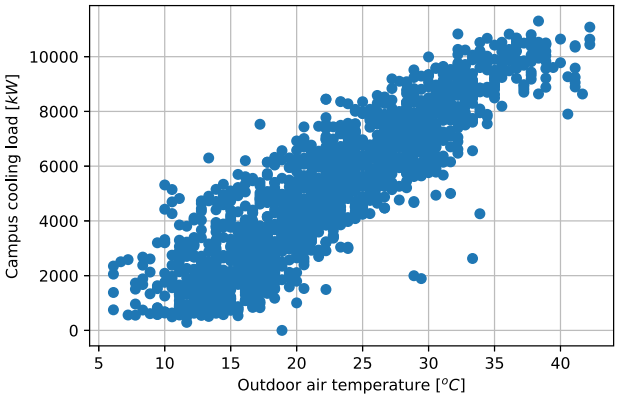


Fig. 6. Correlation between outdoor air temperature and campus cooling load (measurements).

profile for the MPC period to the baseline period to analyze carbon emission for the baseline period. This makes sense because the baseline control logic is independent of the MOER signal and therefore the change of MOER profile for the baseline control would not change the result of energy consumption and peak power.

Fig. 7 compares the profiles of the outdoor temperature, site solar generation and MOER for the selected baseline control period and the MPC period. The first subfigure shows that the MPC period is slightly warmer than the baseline period but is reasonably similar (mean values are 19.9 and 22.0 °C or 67.8 and 71.7 °F for the baseline and MPC periods, respectively). The solar generation is almost identical because the two periods are in the same month (baseline: May/2020, MPC: May/2021). The last subfigure shows the historical data for the MOER

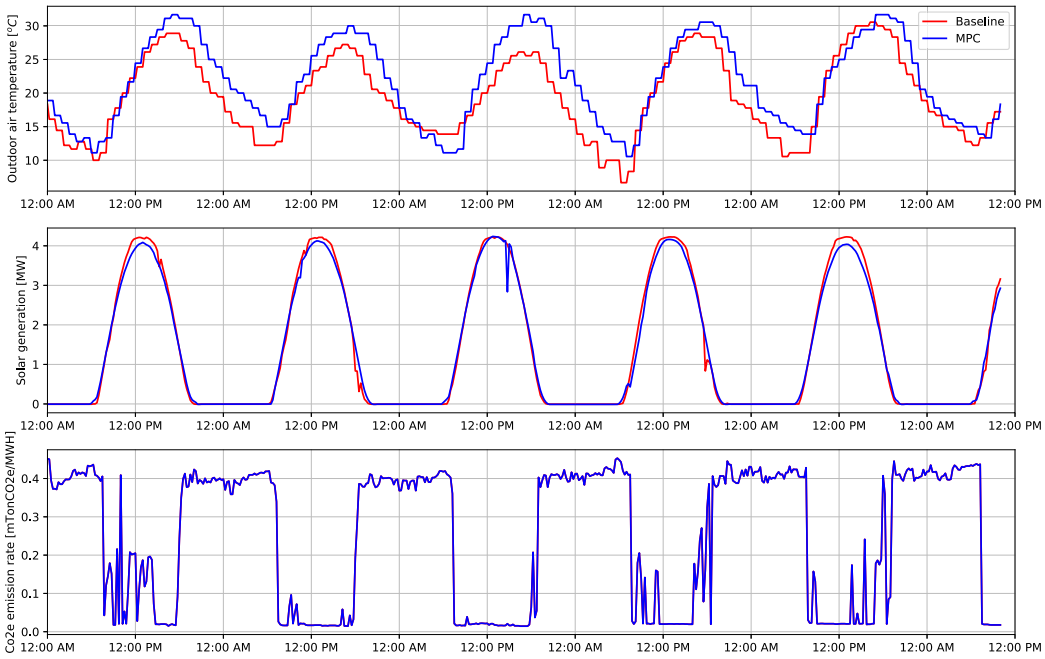


Fig. 7. Comparisons of outdoor air temperature, solar generation and CO2 emission rate between the selected baseline control period and MPC period.

signal for the MPC period and this profile was applied to estimate carbon emission for the baseline period.

4.2. MPC performance assessment

4.2.1. Numerical performance

The mixed integer linear MPC formulation, Eq. (1)–(16), consists of $3J\ N_p$ and $(J+3)N_p+2$ numbers of binary and real decision variables, and $(J+2)N_p+1$, $(4J+3)N_p$ and $2N_p+2$ numbers of equality, inequality, and bounding constraints respectively. With $N_p = 48$ and $J = 4$, the optimization problem has around 1300 constraints (289 equality, 912 inequality and 98 bounding constraints) and 1000 variables to be optimized (576 binary variables $(s_j[\cdot], \delta_{ON,j}[\cdot], \delta_{OFF,j}[\cdot])$ and 338 continuous variables $(Q_{CH}[\cdot], x[\cdot], v_j[\cdot],\ P^{+}_{net}[\cdot], \upsilon_x, d)$). For a 64-bit computer with a dual core 2.13 GHz CPU and 8 GB RAM, the computing time ith the GLPK solver was an order of seconds (the average computing time and standard deviation were 9.72 and 0.21 s) despite the size of the optimization problem.

4.2.2. Comparisons of closed-loop responses between baseline control and MPC

Fig. 8 compares the closed-loop responses of the baseline control and the MPC. The first sub-figure shows the campus net power consumption, i.e., total consumption subtracted by the on-site PV generation, and the second sub-figure shows the profiles of the state of charge (measured from the relative height of the thermocline) for the two control periods. The blue and red colors are dedicated to the baseline control and MPC, respectively. The shaded areas indicate when the MOER is higher than 0.3 [mTonCO2e/MWh], which is equal to the average level of GHG emission for a natural gas power plant in the U.S. according to the EPA's Emissions and Generation Resource Integrated Database (eGRID). In other words, the shaded areas are when the grid is dirty, and non-shaded areas are when the grid is clean. First, look at the afternoon period for the first day where the state of the charge decreases from 100% to 60% for the baseline control. This means that the control utilizes the stored thermal energy to meet the entire campus cooling load during the daytime period. For the time being, the chillers were off, causing significant excess power around 2 MW that was sent back to the grid as shown in the first subfigure. Compare the behavior of the state of charge with the MPC for the same period. The MPC actually increased the state of the charge which implies that the chillers were ON, met the entire campus cooling load and charged the storage. Since the MPC increased the chiller power during the period, the excess power export was eliminated as shown in the first subfigure. Second, when the grid condition got dirtier, the baseline control increased the state of charge and this charging process continued until the TES was fully charged (see the evening time for the first day in the second subfigure). On the other hand, the MPC consumed less energy when the grid was dirty by leveraging the stored energy from the on-site renewable. This pattern repeats for the entire test period as shown in the figure.

Table 3
Comparisons of self-consumption of on-site PV energy between a baseline control and MPC.

Control period	Solar generation [MWh/day]	Excess solar [MWh/day]	Solar self-consumed [MWh/day]	Solar self-consumption [%]
Baseline	32.6	8.9	23.7	72.8
MPC	31.7	0.2	32.5	99.2

It should be mentioned that facility operators requested to preserve 50% to 60% of the charge level during the MPC implementation period which significantly limits savings potentials. This can be confirmed in the second subfigure.

The self-consumption of on-site renewables between those two controllers are summarized in Table 3. The baseline control, which is independent of renewable generation, consumed 23.7 MWh (daily average) out of 32.6 MWh (daily average) total solar generation, resulting in 72.8% self-consumption ratio. On the other hand, the MPC control consumed almost all on-site PV energy resulting in a 99.2% self-consumption ratio. In other words, the increment of the self-consumption ratio via MPC is 26.4%.

4.2.3. Benefits of MPC to the grid

Although the time series comparison provides comprehensive information, it is difficult to capture the overall behavior of the MPC. To more clearly characterize this behavior, we averaged daily power consumption profiles for the two controls and they are shown in Fig. 9.

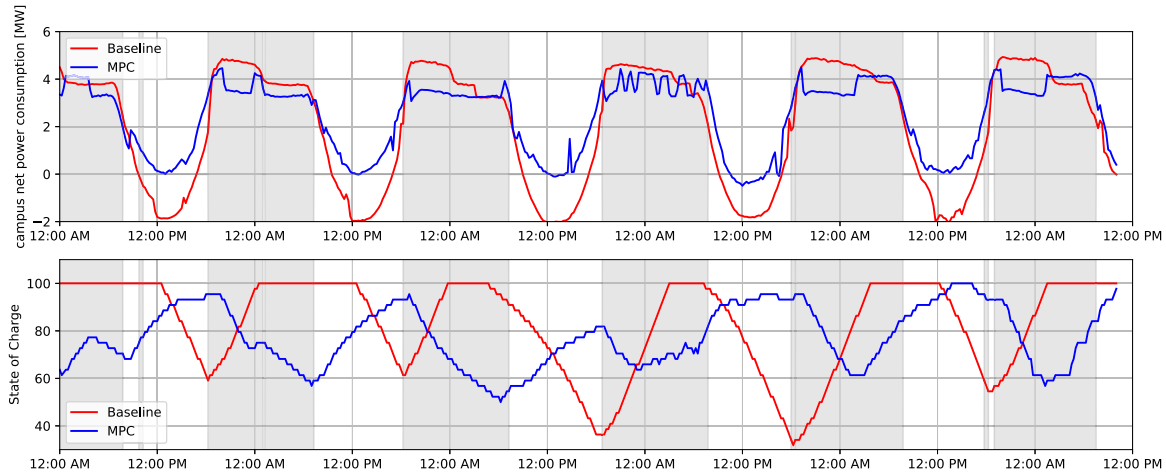


Fig. 8. Comparisons of experimental net power consumption and state of charge between the baseline and MPC at UC-Merced (red: baseline, blue: MPC).

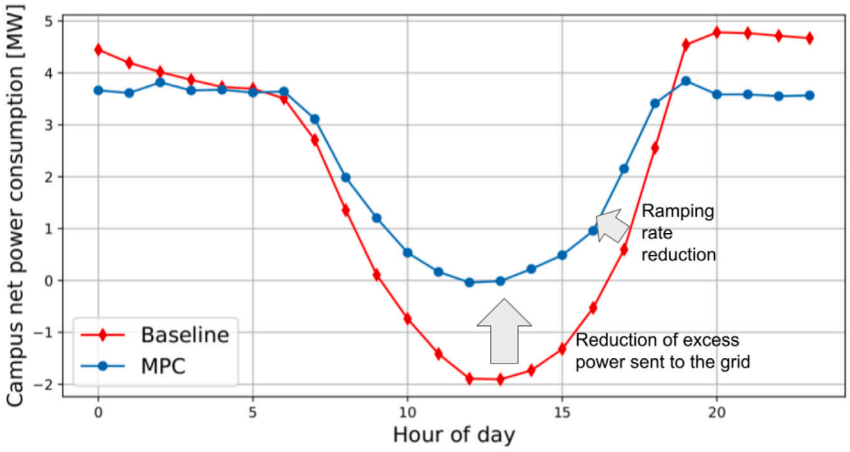


Fig. 9. Comparisons of representative (daily averaged) net power profiles between the baseline control and MPC at the UC-Merced for the weeks of evaluation in May.

Like before, the blue and red lines are associated with the baseline control and MPC. It is interesting to see that the representative (daily averaged) profiles at the site look similar to the duck-curve which the grid is experiencing. This figure clearly shows that the MPC consumes and stores more energy during the daytime using the clean on-site solar, and consumes less during evening and night time when the grid gets dirty by leveraging the stored energy.

The smooth load curve could provide significant benefits to the grid summarized below.

- The higher consumption during the daytime addresses the issue of high renewable curtailment in CAISO (see Fig. 2)
- The proper operation of the chilled water storage in response to the on-site renewable allows for tolerating the increasing PV penetration to the grid
- The reduced ramping rate during the evening time addresses the issue of running gas turbines to follow the high ramping rate of the neck of the duck-curve as discussed in Introduction
- The self-consumption of the behind-meter solar reduces the concern of voltage regulation in the distribution line and improves the grid reliability and efficiency

4.3. Site assessment of GHG reduction

The comparison of GHG emission rates are shown in Fig. 10. One can see visible reduction in the emission rate. The daily average value of the carbon reductions using the MPC is 1 [mTonCO2e] per day. The unit metric ton CO2e is equivalent to the emission rate driving a car for around 2500 miles or 4023 km (roughly San Francisco, California to Washington, DC). Again, the reduction was achieved by coordinating operations of cooling TES plant with on-site PVs in response to the grid carbon emission signal. The calculated GHG savings is around 10% but it is anticipated that more than 20% reduction can be achievable for the plant once the minimum state of charge constraint (varied between 50% to 60%) can be relaxed.

4.4. Benefit of MPC to the facility

Practically, one of the main driving forces for facility operators to adopt advanced controls is an economic benefit rather than GHG reduction. Therefore, showing a utility cost reduction by using the MPC in addition to the grid benefits and identifying the mechanism of the potential cost savings is the key for adoption. Traditionally, the peak electric load in buildings likely occurs during the daytime due to high energy consumption for the period, e.g., for running air handling unit fans and increased plug loads associated with occupants. However, the high penetration of renewables changes the situation: the high energy consumption during the daytime is compensated by the solar energy and the peak could happen during the evening or nighttime when solar is not available, especially in California. This means, if nighttime energy consumption can be shifted to daytime, it could reduce the peak power consumption. As shown above, MPC tends to operate chillers more during daytime and less during nighttime which is the mechanism of utility cost savings potential of MPC.

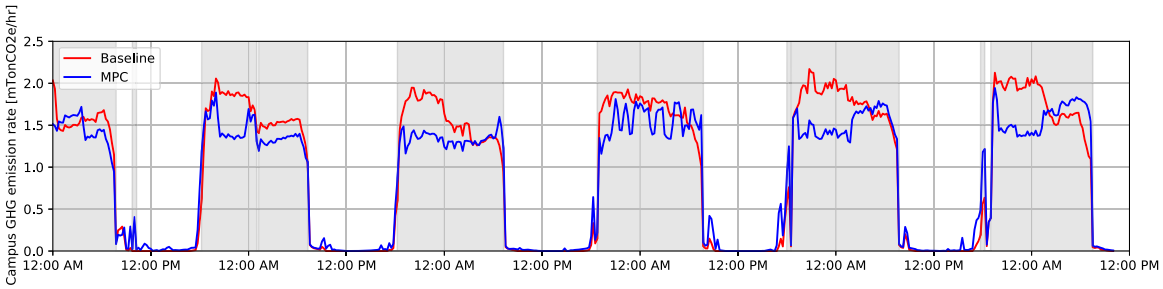


Fig. 10. Comparisons of GHG emission rate between MPC and baseline storage-priority control at the UC-Merced.

Table 4
Summary of MPC relative performance compared to baseline storage priority control for a central cooling TES plant with on-site PV when 50% of TES capacity was utilized.

Increment of PV self-consumption ratio [%]	CO2 reduction [%]	Peak demand reduction [%]
26.4	9.6	9.8

For the comparison periods discussed in the previous section (Fig. 8), the peak power reduction was around 10%. The peak demand reduction is not that significant considering the size of TES (2 million gallons of water). This is mainly because the facility operators requested to preserve 50% to 60% of the charge level during the MPC implementation period, as mentioned in Section 4.2.2. This significantly limits the capability of a peak demand reduction and CO2 reduction, because during the nighttime chillers have to run to maintain the minimum charge level. If the strong constraint has been relaxed, MPC would not need to operate chillers for the nighttime and would increase the savings. It should be mentioned that even with the conservative 10% peak reduction, it attracts the interest of facility operators because of the magnitude of power consumption for the campus-level cooling plant. For the UC-Merced case, the utility cost savings are around ten thousand dollars per month.

4.5. Summary of MPC performance

The MPC performance related to self-consumption, CO2 emission and peak power compared to the baseline storage control are summarized in Table 4. Considering the fact that only around 50% of TES capacity was utilized during this demonstration, the site demonstration results show that central cooling TES plants can be cost-effective resources for renewable energy integration and grid decarbonization by only upgrading control strategies.

5. Barriers for a large scale MPC deployment for large central plants

In this section, we describe barriers and challenges for a large scale MPC deployment that were identified during this demonstration work, and share lessons-learned to guide and facilitate future development of MPCs for large central plants. Table 5 lists the challenges, efforts required to address them, and an indication of the level of effort required. It should be mentioned that the challenges and level of effort needed to address them are from the demonstration cases, and may differ from site to site.

Haves et al. [38] carried out an experimental MPC evaluation and observed similar difficulties and concluded the following as key barriers for commercializing the MPC technology:

- The component-wise calibration process is labor intensive
- It was not possible to rely on manufacturer specifications, and performance was subject to change after maintenance and failures
- Obtaining historical data for model calibration that covers operating envelope of components is in question
- Flow-rate sensors were inaccurate for many reasons including inappropriate sensor location

In the MPC demonstration of this paper, the most significant effort was required for MPC modeling, which includes (1) identifying available and reliable sensor points from EMS, (2) understanding an existing control sequence to identify proper controllable points and to avoid potential conflicts between MPC and EMS, (3) identifying or redeveloping a suitable input–output pair and MPC model structure, according to these points and the HVAC system characteristics (including the control sequence), (4) investigating the data quality (e.g., via the correlation analysis) and operating envelope, and designing an experiment for high quality data if necessary, (5) validating the model accuracy, and (6) repeating the process of (3) to (5) until one is confident.

Since the control sequence, reliable sensor points, and data quality/comprehensiveness may differ from site to site and each of them limits a range of input–output pairs and the corresponding model structures, a customized modeling solution and MPC formulation are often required, which is clearly a barrier for MPC deployments. In addition, many of the issues described in Table 5 are time-consuming and labor-intensive to resolve and some of them could be infeasible to fix due to limited resources of time, cost and labor, and other technical and security issues. To overcome those challenges, it is necessary to investigate modeling and MPC approaches which are less interruptive and rely on less sensors. An example research topic would be combining a modeling strategy that relies on less sensors with the robust MPC technique (e.g., [39–41]) to explicitly account for the modeling and forecast errors.

The assessment of MPC formulations using a detailed simulation model before site-deployment can facilitate the control design process, mitigate potential risks and provide the value proposition. The barriers associated with simulation studies that we identified are:

- Developing a detailed, site-specific model requires significant time. This is not because of modeling the physics and hydraulics as modular models exist, but because understanding and modeling the control sequences of the EMS are time-consuming as sequences are often outdated or not well documented but can interfere with the MPC, and calibration of models is hard as proper measurement instrumentation is often lacking.
- The current state-of-the-art building energy simulation tools, e.g., EnergyPlus [42], are not detailed enough to capture some critical MPC failures and risks. For example, we found that the supply water temperature becomes unacceptably high, lasting for about 30 min to an hour, whenever the plant mode switches. This was due to the time-lag required for rearranging isolation valves and internal dynamics of chillers during mode changes which are not typically modeled. Indeed, one of our trials with an initial version of MPC that did not take this into account caused serious supply water temperature fluctuations. This could have been avoided if it was tested with a detailed simulation model.

Table 5
Issues confronted during implementation of MPC at a demonstration site.

Category	Issue	Effort and approximate level (S: small, M: medium, H: high)
Data /modeling	There are a large number of HVAC components and data points	Effort for data pre-processing and post-processing (S to M)
Data /modeling	Not all data points in the EMS are managed and up-to-date (due to the large number of data points)	May results in unrealistic models such as unreasonably high COP. Needs careful and manual review of all potential data points for MPC design (S to M)
Data /modeling	A tagging scheme for EMS data points is not standardized	Extra effort to find and confirm right data points for MPC modeling and implementation (S to M)
Data /modeling	Not all flow-rate sub-meters (including differential pressure transducers) and electric power sub-meters for individual components are available and reliable. This means the thermal load and/or power consumption data for each component is unavailable or unreliable	Challenges of using well-established HVAC modeling approaches (e.g., component-wise models). Requires identifying or developing a tailored modeling approach that relies on available measurements (H)
Data /modeling	The mass and energy balances calculated from measurements does not reasonably hold (30% to 50% mismatch), particularly due to the uncertainty in flow rate information	Challenges of using well-established HVAC modeling approaches (M-H)
Data /modeling	Historical EMS data does not cover a wide range of operating conditions, and there are strong correlations between them	Challenges for any data driven modeling approaches. Likely results in a poor model (H): refer to Ljung [35], Söderström and Stoica [36] and Kim et al. [37] for how data cross-correlations can negatively affect model estimations.
Model-ing/forecast	Models and forecasts are always uncertain	High effort to improve prediction accuracy (H). Makes an MPC to behave unexpectedly and sub-optimally
Modeling	It is a valuable step to test and compare candidate MPC solutions, before site-deployment, using a detailed simulation model. This is especially true for testing new MPC approaches or when applying to systems with unique or complex characteristics. The virtual control testbed should be detailed enough to identify potential MPC failures and risks (e.g., chiller surge, high return water temp issue during a chiller cycle due to chiller dynamics)	Developing a detailed, site-specific model requires significant time (H). Templates of testbed models would facilitate the model development and MPC design processes
Operational restriction	Potential conflicts between MPC decisions and EMS logics	Requires understanding of complex EMS control sequences, performing iterative MPC functional tests and interactions with facility operators (H)
Operational restriction	Controllability of the HVAC system is limited by system design, system faults, and/or operator requirements. For example, ON/OFF operations of components (chillers, cooling towers, pumps and isolation valves) are interlocked each other in our study and they cannot be controlled separately	May require customizing a model structure and MPC formulation: adjusting MPC model, constraints, and objective to accommodate controllability limitations (H)
Operational restriction /safety	Plant mode switching results in a large fluctuation of the chilled water supply temperature due to rearrangements of isolation valves and primary/condenser water pumps, and the internal dynamics of chillers. The undesired transient lasts from 10 min to an hour raising significant security issues	Requires adding an algorithm to an MPC to prevent frequent mode switching
Operational restriction /safety	Revising EMS logic is practically difficult and it requires identifying potential conflicts after updates, convincing facility manager and operators to accept necessary changes of the EMS, and ensuring operational safety during the MPC demonstration	Limits changes of EMS logic in favor of MPC formulation and implementation (H)
Safety	Lack of liability by MPC implementer for a potential operation failure during MPC implementation	Effort to ensure operational safety (M)
Others	There are many stakeholders for a large plant operation including programmer, IT person, multiple facility operators, facility managers	Coordination effort for MPC implementation (H)
Customer adoption	Facility operators are not familiar with the MPC concept since it is not intuitive compared with rule-based control	Iterative effort for convincing them (M-H)
Customer adoption	Unclear value proposition and/or improper incentives	Iterative effort for convincing them (M-H)

Templates of detailed simulation models for a variety of district energy system configurations, which include representative EMS control sequences and are detailed enough to identify potential MPC failures and risks would facilitate the MPC design processes.

6. Conclusions

Cooling TES plants are one of the oldest and most reliable energy storage technologies. This paper shows that cooling plants with existing TES could be one of the most cost-effective resources to achieve state and government carbon neutrality goals, since it does not require installing new energy storage but only requires changing the operation. We presented an MPC and its real site performance for a campus-level cooling TES plant that coordinates operation of multiple chillers, chilled water tank and behind-meter PVs. It aims at self-consuming on-site generation from a 4 MW solar farm, lowering carbon emission at the grid and minimizing utility bills. The performance of MPC was assessed by comparing with a carefully selected baseline period where the storage priority control has been implemented. The MPC reduced the excess PV power by around 25%, the greenhouse gas emission by 10%, and peak demand by 10%. The savings estimates were conservative, since only a partial usage of the chilled water tank was allowed during the MPC implementation period. The capability of achieving utility peak demand in addition to lowering GHG emission is the uniqueness of the MPC. Future work will include combining with a robust MPC strategy, and a longer-term implementation of the MPC (>1 year) to provide a more comprehensive savings assessment and reliability of the MPC technology.

CRediT authorship contribution statement

Donghun Kim: Conceptualization, Algorithm design, Methodology, Writing – original draft, Validation, Visualization, Investigation, Experiment. **Zhe Wang:** Literature review, Writing – original draft, Support site demonstration. **James Brugger:** Validation, Site experiment, Reviewing and editing. **David Blum:** Reviewing and editing. **Michael Wetter:** Reviewing and editing. **Tianzhen Hong:** Funding acquisition, Project administration, Supervision, Reviewing and editing, Finding demonstration site. **Mary Ann Piette:** Funding acquisition, Supervision, Reviewing and editing, Finding demonstration site.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the Assistant Secretary for Energy Efficiency and Renewable Energy, Building Technologies Office, United States and the U.S. China Clean Energy Research Center, Building Energy Efficiency (CERC-BEE) program, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. The authors would like to thank Tim Olson, Michael Pearson, Robert Stefanski, Danny Ward, Francisco Cazares-Moreno, Eric Cardoza and Victor Zaragoza from University California, Merced for their strong support for this site demonstration work. Special thanks to Michael Logsdon from Pacific West Controls, Inc. for constructive feedback and technical support, and to Erika Gupta from Building Technologies Office of U.S. Department of Energy for her support on the project.

Appendix

A.1. Chilled water tank modeling

Like the moving boundary modeling approach for chilled water tanks, it is assumed that the temperature distributions for the cold and warm water within the tank are uniform. Then, the energy balance for the chilled water tank becomes,

$$C_w \frac{d}{dt}(T_c z + T_h(1-z)) = -Q_{discharge} + \frac{T_{OA} - (T_c z + T_h(1-z))}{R}, \tag{20}$$

where C_w is the thermal capacitance of the chilled water tank [J/°C] (ideally, $C_w = \rho_w c_{p,w} V_0$), z is the relative height of the thermocline in the tank, and R is the thermal resistance between the water and outdoor air.

$Q_{discharge}(\in \mathbb{R})$ represents the discharging rate when positive and charging rate when negative. The sum of the chiller load and discharging rate must meet buildings' cooling load at every timestep in order not to interfere with comfort.

$$Q_{BL} = Q_{CH} + Q_{discharging} \tag{21}$$

It is practically useful to define a state of charge (SOC) of thermal energy storage analogous to electrochemical battery, to indicate that 100% SOC means fully charged and 0% SOC means fully depleted. To do this, we define the SOC, namely x, as scaled internal energy with two reference points as follows.

$$x := \frac{u_h^0 - u}{u_h^0 - u_c^0}, \tag{22}$$

where u is the specific internal energy of the water in the tank, u_h^0 (u_c^0) is the specific internal energy of water at a reference temperature T_h^0 (T_c^0), respectively. Since the water in the tank are sub-cooled liquid, u can be expressed as $u = c_{p,w}(zT_c) + c_{p,w}(1-z)T_h$. Then, Eq. (22) becomes

$$x = \frac{T_h^0 - (T_c z + T_h(1-z))}{T_h^0 - T_c^0}. \tag{23}$$

Plugging this to Eq. (20) results in our final model.

$$C_s \frac{dx}{dt} = -\frac{T_h^0 - T_c^0}{R} x + (Q_{CH} - Q_{BL}) - \frac{T_{OA} - T_h^0}{R}, \tag{24}$$

$$C_s = C_w(T_h^0 - T_c^0). \tag{25}$$

Parameters of C_s and R can be estimated with measurements.

After a discretization, the model can be expressed as the standard discrete-time LTI state space model,

$$x[k+1] = Ax[k] + B \begin{bmatrix} Q_{CH}[k] - Q_{BL}[k] \\ T_{OA}[k] - T_h^0 \end{bmatrix} \tag{26}$$

where A, B is defined by a selection of discretization schemes.

The SOC dynamic of a TES could be modeled analogous to the electrochemical battery with different charging and discharging efficiencies (e.g., [41]). This modeling approach naturally leads to a switching system, requiring a logical proposition to prevent simultaneous charging and discharging. However, when a TES is modeled with thermodynamics (i.e., the energy balance) and heat transfer (such as convection and conduction between the two layer waters and outdoor air temperature to model heat losses) like our paper and Henze et al. [43], the logical proposition can be avoided.

Table 6
Configuration of MPC parameters.

Symbol	Description	Value
N_p	Prediction horizon	48
J	The total number of plant modes	4
ω_d	A weight on peak power	200
ω_x	A weight on SOC violation	200
OT_j	Minimal ON and OFF time periods	2
x_{min}	The minimum charge limit	55%
x_{max}	The maximum charge limit	98%
T_h^0	A reference temperature for a warm water	58 °F (14.4 °C)
T_c^0	A reference temperature for a cold water	40 °F (4.4 °C)
R	Overall thermal resistance between water and outdoor air temperature	8.68 [°C/MW]
C_s	Thermal capacitance	391 220.52 [MJ]
$Q_{min,j}$	Cooling capacity lower bound for each plant mode	[4.33, 5.21, 8.99, 13.38] [MW]
$Q_{max,j}$	Cooling capacity upper bound for each plant mode	[4.82, 8.69, 13.08, 17.48] [MW]
a_j	A plant power coefficient for each plant mode (see Appendix A.2)	[0.126, 0.127, 0.126, 0.127] [–]
$c_{j,0}$	A plant power coefficient for each plant mode (see Appendix A.2)	[0.118, 0.191, 0.173, 0.318] [MW]
$c_{j,1}$	A plant power coefficient for each plant mode (see Appendix A.2)	[0.000, 0.006, 0.000, 0.000] [MW/°C]

A.2. Modeling total plant power as a switching system

For the total plant power modeling, we chose the following, linear-affine model structure for each mode.

$$P_{plant} = \begin{cases} 0 & \text{for the plant mode 0} \\ a_1 Q_{CH} + b_1(T_{WB}) & \text{for the plant mode 1} \\ \vdots & \\ a_J Q_{CH} + b_J(T_{WB}) & \text{for the plant mode J} \end{cases} \tag{27}$$

where Q_{CH}, T_{WB} are the total chiller load (or the sum of chiller loads), and the outdoor wet-bulb temperature, respectively. a_j is a constant and b_j is a function of T_{WB}, respectively. There is no restriction on the model structure for b_j but the linear affine form, i.e., $b_j = c_{j,0} + c_{j,1} T_{WB}$, is selected in this paper. Since T_{WB} is not an optimization variables, we denote $b_j(T_{WB})$ as b_j for the simplicity of notations. This modeling approach requires only 6 measurements (the entering chilled water temperature at the upstream of the primary pumps, supply water temperature, total primary flow rate, the outdoor wet-bulb temperature, total plant power, and mode) regardless of the number of modes J, and is simple to estimate the coefficients of $(a_1, c_{1,0}, c_{1,1}) \cdots (a_J, c_{J,0}, c_{J,1})$ via, e.g., the Least Squares Method. In this work, we let $c_{j,1}$ be zero for all modes except for the second mode, since the wet-bulb data for those modes did not cover a wide range.

A.3. MPC parameters

All MPC parameters tuned for the cooling plant are summarized in Table 6.

References

[1] Mitchell JW, Braun JE. Principles of heating, ventilation, and air conditioning in buildings. John Wiley & Sons; 2012.
[2] Ashrae, ASHRAE. 2020 ASHRAE handbook: HVAC systems and equipment. ASHRAE; 2020.
[3] Kircher KJ, Zhang KM. Model predictive control of thermal storage for demand response. In: 2015 American control conference (ACC). IEEE; 2015, p. 956–61.
[4] Oldewurtel F, Ulbig A, Morari M, Andersson G. Building control and storage management with dynamic tariffs for shaping demand response. In: 2011 2nd IEEE PES international conference and exhibition on innovative smart grid technologies. IEEE; 2011, p. 1–8.
[5] Ceusters G, Rodríguez RC, García AB, Franke R, Deconinck G, Helsen L, Nowé A, Messagie M, Camargo LR. Model-predictive control and reinforcement learning in multi-energy system case studies. Appl Energy 2021;303.
[6] Zhang Y, Zhang T, Wang R, Liu Y, Guo B. Optimal operation of a smart residential microgrid based on model predictive control by considering uncertainties and storage impacts. Sol Energy 2015;122:1052–65.
[7] LeBreux M, Lacroix M, Lachiver G. Control of a hybrid solar/electric thermal energy storage system. Int J Therm Sci 2009;48(3):645–54.
[8] Li C-T, Peng H, Sun J. MPC for reducing energy storage requirement of wind power systems. In: 2013 American control conference. IEEE; 2013, p. 6607–12.
[9] Reynders G, Nuytten T, Saelens D. Potential of structural thermal mass for demand-side management in dwellings. Build Environ 2013;64:187–99.
[10] Hajiah A, Krarti M. Optimal control of building storage systems using both ice storage and thermal mass–Part I: Simulation environment. Energy Convers Manage 2012;64:499–508.
[11] Hajiah A, Krarti M. Optimal controls of building storage systems using both ice storage and thermal mass–Part II: Parametric analysis. Energy Convers Manage 2012;64:509–15.
[12] Ahmad MW, Eftekhari M, Steffen T, Rowley P. The effect of model and objective function mismatch in model predictive control (MPC) for a solar heating system with a heat pump. In: 2014 UKACC international conference on control (CONTROL). IEEE; 2014, p. 685–90.
[13] Zhao Y, Lu Y, Yan C, Wang S. MPC-based optimal scheduling of grid-connected low energy buildings with thermal energy storages. Energy Build 2015;86:415–26.
[14] Lee D, Ooka R, Ikeda S, Choi W, Kwak Y. Model predictive control of building energy systems with thermal energy storage in response to occupancy variations and time-variant electricity prices. Energy Build 2020;225:110291.
[15] Ikeda S, Choi W, Ooka R. Optimization method for multiple heat source operation including ground source heat pump considering dynamic variation in ground temperature. Appl Energy 2017;193:466–78.
[16] Tarragona J, Fernández C, de Gracia A. Model predictive control applied to a heating system with PV panels and thermal energy storage. Energy 2020;197:117229.
[17] Tarragona J, Pisello AL, Fernández C, de Gracia A, Cabeza LF. Systematic review on model predictive control strategies applied to active thermal energy storage systems. Renew Sustain Energy Rev 2021;149:111385.
[18] Braun JE. A comparison of chiller-priority, storage-priority, and optimal control of an ice-storage system. ASHRAE Trans 1992;98:893–902.
[19] Ashrae, ASHRAE. 2019 ASHRAE handbook: HVAC applications. ASHRAE; 2019.
[20] Callaway DS, Fowlie M, McCormick G. Location, location, location: The variable value of renewable energy and demand-side efficiency resources. J Assoc Environ Resour Econ 2018;5(1):39–75.
[21] Siler-Evans K, Azevedo IL, Morgan MG, Apt J. Regional variations in the health, environmental, and climate benefits of wind and solar generation. Proc Natl Acad Sci 2013;110(29):11768–73.
[22] Environmental Protection Agency, United States. Air markets program data. 2019.
[23] Boyd S, Boyd SP, Vandenberghe L. Convex optimization. Cambridge University Press; 2004.
[24] Grossmann IE, Ruiz JP. Generalized disjunctive programming: A framework for formulation and alternative algorithms for MINLP optimization. In: Mixed integer nonlinear programming. Springer; 2012, p. 93–115.
[25] Raman R, Grossmann IE. Modelling and computational techniques for logic based integer programming. Comput Chem Eng 1994;18(7):563–78.
[26] Lee S, Grossmann IE. New algorithms for nonlinear generalized disjunctive programming. Comput Chem Eng 2000;24(9–10):2125–41.
[27] Rajan D, Takriti S, et al. Minimum up/down polytopes of the unit commitment problem with start-up costs. IBM Res Rep 2005;23628:1–14.
[28] Knueven B, Ostrowski J, Watson J-P. A novel matching formulation for startup costs in unit commitment. Math Program Comput 2020;12(2):225–48.
[29] Knueven B, Ostrowski J, Watson J-P. On mixed-integer programming formulations for the unit commitment problem. INFORMS J Comput 2020;32(4):857–76.
[30] Bazaraa MS, Sherali HD, Shetty CM. Nonlinear programming: Theory and algorithms. John Wiley & Sons; 2013.

[31] Skogestad S, Postlethwaite I. Multivariable feedback control: Analysis and design, Vol. 2. Citeseer; 2007.
[32] Henze GP, Krarti M. The impact of forecasting uncertainty on the performance of a predictive optimal controller for thermal energy storage systems. ASHRAE Trans 1999;105:553.
[33] Holmgren WF, Hansen CW, Mikofski MA. Pvlib python: A python package for modeling solar energy systems. J Open Source Softw 2018;3(29):884.
[34] Abadi M, Barham P, Chen J, Chen Z, Davis A, Dean J, Devin M, Ghemawat S, Irving G, Isard M, et al. {TensorFlow}: A System for {Large – Scale} machine learning. In: 12th USENIX symposium on operating systems design and implementation (OSDI 16). 2016, p. 265–83.
[35] Ljung L. System identification. In: Signal analysis and prediction. Springer; 1998, p. 163–73.
[36] Söderström T, Stoica P. System identification. Prentice-Hall International; 1989.
[37] Kim D, Cai J, Braun JE, Ariyur KB. System identification for building thermal systems under the presence of unmeasured disturbances in closed loop operation: Theoretical analysis and application. Energy Build 2018;167:359–69. http://dx.doi.org/10.1016/j.enbuild.2017.12.007, URL: https://www.sciencedirect.com/science/article/pii/S0378778817317322.
[38] Haves P, Hencey B, Borrell F, Elliot J, Ma Y, Coffey B, Bengea S, Wetter M. Model predictive control of HVAC systems: Implementation and testing at the University of California, Merced. Technical Report, Lawrence Berkeley National Lab.(LBNL), Berkeley, CA (United States); 2010, URL: https://www.osti.gov/servlets/purl/988177.
[39] Bemporad A, Morari M. Robust model predictive control: A survey. In: Robustness in identification and control. Springer; 1999, p. 207–26.
[40] Mayne D. Robust and stochastic model predictive control: Are we going in the right direction? Annu Rev Control 2016;41:184–92.
[41] Carli R, Cavone G, Pippia T, De Schutter B, Dotoli M. Robust optimal control for demand side management of multi-carrier microgrids. IEEE Trans Autom Sci Eng 2022.
[42] Crawley DB, Lawrie LK, Winkelmann FC, Buhl WF, Huang YJ, Pedersen CO, Strand RK, Liesen RJ, Fisher DE, Witte MJ, et al. EnergyPlus: creating a new-generation building energy simulation program. Energy Build 2001;33(4):319–31.
[43] Henze GP, Dodier RH, Krarti M. Development of a predictive optimal controller for thermal energy storage systems. HVAC&R Res 1997;3(3):233–64. http://dx.doi.org/10.1080/10789669.1997.10391376.