2016

Closed-Loop Scheduling for Cost Minimization in HVAC Central Plants

Michael J. Risbeck
University of Wisconsin--Madison, United States of America, risbeck@wisc.edu

Christos T. Maravelias
University of Wisconsin--Madison, United States of America, christos.maravelias@wisc.edu

James B. Rawlings
University of Wisconsin--Madison, United States of America, james.rawlings@wisc.edu

Robert D. Turney
Johnson Controls, Inc., United States of America, Robert.D.Turney@jci.com

Closed-Loop Scheduling for Cost Minimization in HVAC Central Plants

Michael J. RISBECK[1*], Christos T. MARAVELIAS[1], James B. RAWLINGS[1], Robert D. TURNEY[2]

[1] University of Wisconsin–Madison, Department of Chemical and Biological Engineering,
Madison, Wisconsin, USA
Email: risbeck@wisc.edu, christos.maravelias@wisc.edu, james.rawlings@wisc.edu

[2] Johnson Controls, Inc., Building Efficiency,
Milwaukee, Wisconsin, USA
Email: robert.d.turney@jci.com

* Corresponding Author

ABSTRACT

In this paper, we examine closed-loop operation of an HVAC central plant to demonstrate that closed-loop moving-horizon scheduling provides robustness to inaccurate forecasts, and that economic performance is not seriously impaired by shortened prediction horizons or inaccurate forecasts when feedback is employed. Using a general mixed-integer linear programming formulation for the scheduling problem, we show that optimization can be performed in real time. Furthermore, we demonstrate that closed-loop operation with a moderate prediction horizon is not significantly worse than a long-horizon implementation in the nominal case, and that closed-loop operation can correct for inaccurate long-term forecasts without significant cost increase. In addition, we show that terminal constraints can be employed to ensure recursive feasibility. The end result is that forecasts of demand need not be extremely accurate over long times, indicating that closed-loop scheduling can be implemented in new or existing central plants.

1. INTRODUCTION

Economically optimal operation of HVAC central plants is complicated by time-varying electricity cost and peak demand charges that cause nominal operation to be expensive. To achieve minimum utility cost, temporal load shifting via thermal energy storage (TES) is necessary (Henze, Felsmann, & Knabe, 2004). Efficient use of thermal storage requires advance scheduling for the central plant, which is challenging due to the presence of both discrete (i.e., on/off) and continuous (i.e., how much) decisions. Because buildings are subject to disturbances and because electricity prices may be volatile, the true economic cost of a given schedule may change significantly as time passes. In addition, model error for buildings or storage tanks can accumulate toward the end of the optimization horizon, invalidating the predictions of the schedule. Finally, units may break down unexpectedly, making a given schedule infeasible at future times. Thus, central plant schedules must be periodically updated to account for altered forecasts, disturbances, and other new information.

Although it is common to include a building model within the optimization problem (Mendoza-Serrano & Chmielewski, 2012; Y. Ma et al., 2012; J. Ma, Qin, & Salsbury, 2014) to compute the demand profile based on, e.g., predicted building temperature trajectory, in this work, we assume that demand is provided as a forecast that is treated as a fixed parameter within the optimization problem. We make this choice so as to avoid the issues of building model accuracy and to focus more completely on the central plant and the scheduling decisions that must be made. As a result of the central plant focus, each optimization more closely resembles a static scheduling problem than a dynamic optimal control problem. Thus, we refer to the architecture as "closed-loop scheduling," although many of the considerations in control problems (e.g., stability, state estimation, disturbance forecasting) are still relevant. In the following two subsections, we discuss some key considerations for production scheduling in general and for HVAC central plant scheduling.

1.1 Closed-Loop Scheduling

Traditionally, production scheduling has been considered as a static process; that is, a schedule is developed (via optimization or other method) and then followed until its end and possibly repeated periodically (Pantelides, 1994). The tacit assumptions are that both the production environment and production goals do not change throughout the scheduling horizon, and that decisions after the scheduling horizon are of no consequence. Unfortunately, many production environments are actually dynamic, and time does not split nicely into independent intervals. Thus, rescheduling often becomes necessary. While event-based rescheduling (whereby the schedule is recomputed only under a specific set of circumstances) can handle disturbances (Li & Ierapetritou, 2008), a closed-loop implementation can better address the effects of the moving production window (Subramanian, Maravelias, & Rawlings, 2012). In this scheme, a new schedule with a receding (but not shrinking) horizon is determined at each time point. From this schedule, the initial decisions (e.g., unit commitments) are implemented, and the plant is allowed to operate at the current levels for the remainder of the timestep. After this period, measurements are taken to determine the current plant state (e.g., storage levels), and demand forecasts are shifted and updated to include the new point at the end of the horizon. These updated forecasts are then used to generate a new schedule, and the process is repeated.

The two main questions are how to handle the uncertainty of forecasts or other parameters and how to perform the optimization. For parametric uncertainty, three possibilities are (i) to solve for a robust schedule that can meet all of the possible forecasts (Y. Ma et al., 2012), (ii) to solve a stochastic optimization that includes recourse actions (Y. Ma, Matuško, & Borrelli, 2015), or (iii) to simply solve the deterministic optimization using the best available information. However, the robust schedule may be overly conservative when quality forecasts are available, while the stochastic optimization can greatly increase the computational difficulty of the optimization problem due to the number of scenarios. For optimization, parametric optimization approaches have been proposed to shift the computational burden off line by determining the optimal schedule as a function of problem parameters such as equipment availability (Ryu, Dua, & Pistikopoulos, 2007). Assuming the parametric program can be solved, online computation is reduced to a simple lookup; however, due to exponential scaling in the number of parameters (and implicitly the prediction horizon), these methods are limited to small systems and short horizons. Instead, we propose to solve the deterministic formulation online at each timestep using a state-of-the-art solver. As will be demonstrated in the examples, advances in computing power and solution methods allow these problems to be solved quickly and in real time, eliminating what has historically been a major hurdle for online optimization.

1.2 Central Plant Optimization

There are a number of issues specific to central plants that complicate the scheduling process. First, utility consumption can vary nonlinearly with other decision variables. Traditional scheduling problems often assume that the resources consumed by a given task are either fixed or linearly related to batch size (Shah, Pantelides, & Sargent, 1993; Pantelides, 1994), which in the central plant problem essentially becomes an assumption of constant coefficient of performance (COP). While a constant COP assumption is common (Braun, 2007; Henze, Biffar, Kohn, & Becker, 2008), most equipment models are nonlinear (Lee, Liao, & Lu, 2012), so accurate cost prediction must consider variable COP as a function of load. In addition, TES is not 100% efficient, and thus requires a (possibly nonlinear) model to be embedded within the optimization problem (Touretzky & Baldea, 2014). Finally, demand charges are particularly challenging, as they penalize (typically monthly) peak electricity usage and can thus affect system cost over long time periods. They can be dealt with by retaining the past peak as the horizon advances (Cole, Edgar, & Novoselac, 2012) and by increasing their weighting near the end of the billing period (Risbeck, Maravelias, Rawlings, & Turney, 2015), although challenges still remain.

While many techniques can address a single chiller or a simple approximation (e.g., constant COP) of the central plant, we wish to explicitly address the combinatorial problem posed by making the optimal selection from among the available units. To reduce the number of discrete decisions, various approaches to central plant scheduling have been proposed, including near-optimal heuristics (Braun, 2007) and a predefined order for discrete operating modes (Y. Ma et al., 2012; Touretzky & Baldea, 2016). These methods often come at the cost of operational flexibility. As an alternative, the optimal configuration can be determined offline as a function of total central plant load (Henze et al., 2008). This relationship then serves as a surrogate for the central plant and eliminates the discrete on/off decisions. However, such a formulation makes it difficult to appropriately constrain equipment switching. Thus, we propose to simply include all on/off decisions within the optimization problem as discrete decision variables. As we will demonstrate, carefully constructed formulations can lead to quick solution times despite the combinatorial complexity.

2. SCHEDULING MODEL

To schedule equipment usage in the central plant, we formulate a mixed-integer linear programming (MILP) model to determine the cost-optimal loading for each piece of equipment over a finite prediction horizon. Full formulation details can be found in Risbeck et al. (2015). The optimization makes the following decisions:

- Which central plant units are active?
- For active units, what is their loading?
- How much water is being sent to/from active storage tanks?

Decisions are indexed in time by $t \in \mathbf{T}$ running from 1 to the (fixed) prediction horizon T. At each timestep, using the current forecast for utility costs and resource demand, the optimization is solved, and the optimal $t = 1$ decisions are implemented. In the following sections, we discuss the mathematical form of the model as well as some relevant forecasting considerations.

2.1 Mathematical Formulation

For the scheduling model, we employ a modified version of the formulation originally presented in Risbeck et al. (2015). For brevity, we describe only the key features. To facilitate generality, the formulation uses an abstract representation of the central plant by considering "generators" (units in the central plant) and "resources" (material flows in or out of the plant). Generators are indexed by $j \in \mathbf{J}$, and the category includes, e.g., conventional chillers and pumps, while resources are indexed by $k \in \mathbf{K}$ containing, e.g., chilled water and electricity. In the following discussion, Roman letters indicate decision variables, while Greek letters denote parameters.

The objective function is to minimize operating costs, calculated as

$$\min \quad \sum_{k \in \mathbf{K}} \sum_{t \in \mathbf{T}} \rho_{kt} p_{kt} + \sum_{k \in \mathbf{K}} \rho_k^{\max} p_k^{\max}, \tag{1}$$

in which $p_{kt} \geq 0$ is the purchase of each resource, ρ_{kt} is a forecast of (time-varying) resource prices, $\rho_k^{\max}$ is the demand charge, and $p_k^{\max}$ is the maximum resource purchase, calculated by

$$p_k^{\max} \geq p_{kt} + \psi_{kt}, \quad k \in \mathbf{K}, t \in \mathbf{T}. \tag{2}$$

Here, $\psi_{kt} \leq 0$ is exogenous resource use margin (i.e., the difference between current exogenous use and peak exogenous use). Inclusion of this term is necessary because resources such as electricity may be used elsewhere in the overall system, and thus the optimizer should attribute cost only to resource use in excess of the nominal peak.

At each time t, demand ϕ_{kt} for each resource k must be met by the sum of production q_{jkt} in generators, charge or discharge y_{kt} of storage, and direct purchase p_{kt} as expressed in the following constraint:

$$\sum_{j \in \mathbf{J}} q_{jkt} + y_{kt} + p_{kt} \geq \phi_{kt}, \qquad k \in \mathbf{K}, t \in \mathbf{T}. \tag{3}$$

Depending on the particular resource, bounds may enforce that certain terms are zero, e.g., for chilled water, $p_{kt} = 0$ since chilled water cannot be purchased, and for electricity, $y_{kt} = 0$ unless there is a battery or other electricity storage available.

To schedule generators, binary variables u_{jt} are used to determine whether each generator is on or off. When a given generator is running ($u_{jt} = 1$), it often produces some resources ($q_{jkt} > 0$) while consuming others ($q_{jkt} < 0$). For example, a conventional chiller will produce chilled water while consuming electricity. Denoting these sets $\mathbf{K}_j^+$ and $\mathbf{K}_j^-$ respectively, these relationships are governed by equipment models that require constraints of the form

$$u_{jt} \xi_{jk}^{\min} \leq q_{jkt} \leq u_{jt} \xi_{jk}^{\max}, \qquad j \in \mathbf{J}, k \in \mathbf{K}_j^+, t \in \mathbf{T}, \tag{4}$$

$$q_{jkt} = u_{jt} \Xi_j(q_{j\mathbf{K}_j^+ t}), \qquad j \in \mathbf{J}, k \in \mathbf{K}_j^-, t \in \mathbf{T}, \tag{5}$$

in which the ξ_{jk} give equipment capacities and $\Xi_j(\cdot)$ is an equipment model. Note that (4) accounts for turndown when the generator is running, while (5) calculates resource consumption according to the generator model. In order to keep the optimization as an MILP, the generator models $\Xi_j(\cdot)$ are approximated as piecewise-linear and can be made arbitrarily accurate at the cost of additional discrete decision variables.

Other constraints in the formulation include a first-order linear model for storage tanks, dwell time enforcement to prevent rapid on/off switching, symmetry removal to improve solution times, and auxiliary equations for piecewise-linear models. Additional details can be found in Risbeck et al. (2015).

2.2 Forecasts

From the above constraints, we require four main types of parameters: resource price ρ_{kt}, exogenous margin ψ_{kt}, demand ϕ_{kt}, and equipment models $(\xi_{jk}^{\min}, \xi_{jk}^{\max}, \Xi_j(\cdot))$. We assume that the equipment models are readily available, e.g., empirical models based on manufacturer's or historical data (Lee et al., 2012), leaving three parameters yet to be determined. These remaining parameters are all forecasts that are subject to some error.

For electricity, time-varying prices are often published one day in advance (Albadi & El-Saadany, 2007), but if the prediction horizon is longer than 24 h, a longer forecast is required. For example, the accurate 24 h schedule could just be repeated, or historical averages can be used. Note that demand charge values are published in advance, and simpler rate structures (e.g., peak/off-peak pricing) may not be subject to any uncertainty. For central plants with heating equipment, natural gas is often sold at a constant price with no demand charge, and thus there is no forecasting uncertainty.

The more challenging forecasts to obtain are based on system demand and resource use. Both the demand and exogenous use forecasts require estimates of how much of a given resource will be consumed by the system. For heating and cooling loads, estimates can be made using linear models (Mendoza-Serrano & Chmielewski, 2012; Y. Ma et al., 2012; J. Ma et al., 2014) or detailed nonlinear models (US Department of Energy, 2016), but historical data can be used as well. For such models to make accurate predictions, weather data must be incorporated (Zavala, Constantinescu, Krause, & Anitescu, 2009). Exogenous electricity use typically requires some baseline data, although such measurements are often readily available.

3. COMPUTATIONAL EXAMPLES

We now present three simulations of different central plants being scheduled in closed loop. These simulations seek to answer the following questions:

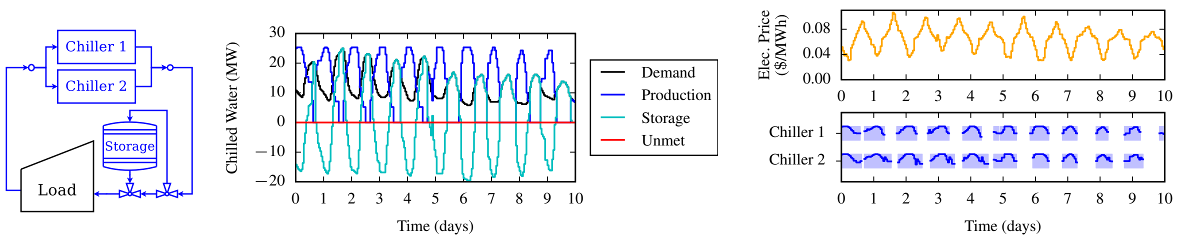
- What effect does prediction horizon have on closed-loop performance?
- What is the effect of inaccuracy in demand forecasts?
- Can the optimization problems be solved in real time?
- Can terminal constraints improve robustness or reduce the necessary horizon?

For a simulation of length N with nominal prediction horizon T, the horizon of the subproblem at time t is taken as $\min(T, N-t)$ so that no time points after $t = N$ are considered by any optimization. This choice ensures a fair cost calculation so that the long horizons are not penalized for "planning ahead" at the end of the simulation. All timesteps are 1 h, and all optimizations are performed using Gurobi 6.5 on an Intel Core 2 2.66 GHz $\times$ 4 CPU with 8 GB of RAM.

3.1 Horizon Length

For simplicity, we start with a small system consisting of two conventional chillers and a storage tank as shown in Figure 1a. The system must meet time-varying demand subject to time-varying electricity prices and a demand charge. Using this system, we examine the effect of prediction horizon T on total electricity cost. We expect that, past a certain value, horizon length does not significantly affect electricity cost. In order to favor longer horizons, we assume that the storage tanks are completely efficient, which allows chilled water to be stored for long times without penalty. We employ a simulation horizon of $N = 240$, which gives 10 days of simulation.

The optimal $T = 240$ solution is shown in Figure 1. Because it includes all decisions considered throughout the simulation horizon, this simulation is the best possible schedule in terms of electricity cost. The main features we see



(a) Central plant equipment.

(b) Chilled water production. Negative values for "Storage" indicate charging, while positive values indicate discharging.

(c) Electricity price and equipment usage. Colored blocks show active units, while dark lines show loading between min and max capacity.

Figure 1: Optimal schedule for horizon length example.

Table 1: Closed-loop costs for varying horizons. Simulation length was $N = 240$. Costs relative to horizon $T = 240$.

Horizon	Use Charges	Demand Charges	Total Charges
168	−0.28%	+8.47%	+0.02%
96	−0.66%	+22.18%	+0.13%
48	−0.47%	+15.96%	+0.10%
24	−0.27%	+16.87%	+0.32%
12	+3.01%	+106.11%	+6.58%
6	+11.70%	+246.32%	+19.83%
2	+26.80%	+502.40%	+43.27%

are that the chillers are operated at full capacity during the night (when electricity is cheap), and they are largely shut off during the day (when electricity is expensive).

Using various smaller horizons, we repeat the closed-loop simulation. Total costs for each horizon are given in Table 1. From this table, we see that a long horizon is not necessary for low cost. Once the horizon is 24 h or longer, there is only a negligible effect on electricity cost, with the 24 h horizon less than 1% higher than the full 240 h horizon. We show the chilled water production for some of the horizons in Figure 2. From this figure, it is clear that the production profile for $T = 2$ is highly suboptimal and that there are large changes in load roughly every 2 h (which also corresponds to the minimum *on* dwell time for equipment). By contrast, the $T = 24$ solution is close to the $T = 240$ solution, indicating that the much shorter horizon would be acceptable for closed-loop operation. Note that the curve for $T = 240$ corresponds to the "Production" curve from Figure 1b.

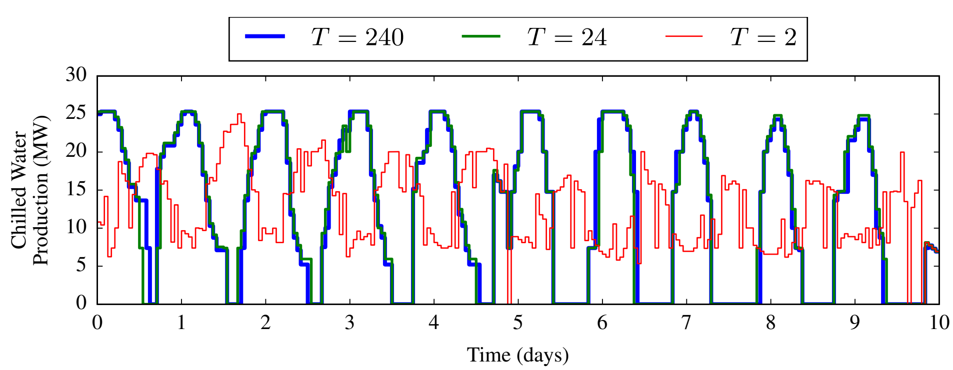
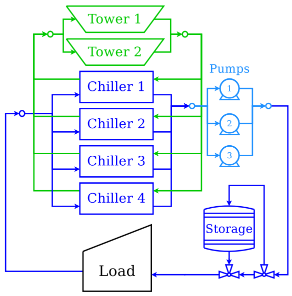
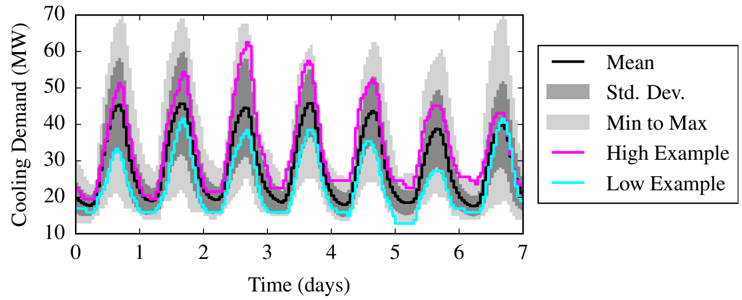


Figure 2: Chilled water production for various horizons. Horizon lengths are given in hours.



(a) Central plant equipment.



(b) Historical average demand and actual demand profile for second example. Data is averaged from May to September, 2010 in Stanford, CA.

Figure 3: System and demand for uncertain demand forecast example.

3.2 Uncertain Demand Forecasts

We turn next to examining the effect of uncertainty in the forecasts used by the model. Whereas the previous example assumed perfect forecasts were available (so as to not disadvantage a longer prediction horizon), in real systems, only an estimate of resource demand is available, and we wish to examine whether the prediction horizon becomes more or less significant in this case. In addition, we wish to determine whether closed-loop scheduling is a viable option without extremely accurate forecasts.

For this example, we use the more complex central plant shown in Figure 3a, which includes pumps and cooling towers. We simulate an approximate schedule as follows. First, we average historical data from May to September to determine the average cooling demand at each hour of the week. This average, along with two sample profiles, is shown in Figure 3b. Then, as a forecast of demand, we use a profile with 2 h of actual values followed by exponential decay from the actual to the average values. Thus, the forecast is accurate only at the beginning of the prediction horizon and quickly falls back to nominal average values. Certainly, this scheme would not be possible unless historical data is already available, but it is intended to represent the accuracy of a reasonable load-prediction methodology.

In the closed-loop simulation, each optimization is stopped when it has found the optimal solution or after 30 s, whichever comes first. If the time limit is reached, then the current incumbent solution is used. After the first timestep, the previous solution is shifted and provided as an initial guess to the optimizer. While this solution is not always feasible due to the updated forecast, the solver is often able to find a nearby feasible solution, which greatly helps solution times.

We first simulate the system using the “Low Example” from Figure 3b as the true demand profile and then then repeat the simulation using the “High Example.” Costs for the system are shown in Table 2. The clairvoyant solution is the optimal schedule using true demand profile, which is the best the system could possibly do but requires perfect forecasts. For each of the three horizons and both demand profiles, we see that cost is higher than the clairvoyant solution (as expected) and further that the differences between the horizons are largely eliminated. For the low demand profile, each solution is less expensive than the average solution (which simply meets the Average demand from Figure 3b) even though it was the basis of the forecasts. Meanwhile, the conservative policy (which meets demand one standard deviation higher than the mean) is significantly more expensive and thus is not a good choice. For the high demand profile, the average profile is highly infeasible, but no infeasibility was observed in the closed-loop solution. Thus, as long as the demand forecast is accurate in the near term and reasonably representative for the remainder of the horizon, the closed-loop implementation can correct for forecasting errors without significant cost penalty.

Another important consideration is solution time. In a practical implementation, each optimization must be completed at least within the timestep, although ideally it should be much quicker. Indeed, in this example, solution times are fast as shown in Figure 4. Note that we omit the final $T-1$ times from each set, as these optimizations use smaller horizons and are thus easier. From this figure, we see that all of the $N = 12$ and $N = 24$ optimizations were solved to optimality within 30 s, which is certainly fast enough for practical implementation. The $N = 48$ optimizations require more computational effort, although all instances were solved to within 0.5% of optimality within the time limit. We also see little difference between the high and low forecasts, indicating some insensitivity toward parameters.

Table 2: Costs for system with inaccurate demand forecast. Entries with a horizon indicate closed-loop results, while others are open-loop optimal. Costs are relative to optimal clairvoyant schedule.

(a) Costs for low demand example.

Horizon	Use Chg.	Dem. Chg.	Tot. Chg.
$N = 12$	$+0.74\%$	$+31.43\%$	$+3.40\%$
$N = 24$	-0.56%	$+42.39\%$	$+3.16\%$
$N = 48$	-0.85%	$+45.73\%$	$+3.19\%$
Average	$+27.79\%$	$+27.51\%$	$+27.76\%$
Conservative	$+67.14\%$	$+73.04\%$	$+67.65\%$

(b) Costs for high demand example.

Horizon	Use Chg.	Dem. Chg.	Tot. Chg.
$N = 12$	$+0.31\%$	$+37.24\%$	$+3.52\%$
$N = 24$	-0.17%	$+32.68\%$	$+2.68\%$
$N = 48$	-0.17%	$+33.29\%$	$+2.74\%$
Average	-18.66%	-19.00%	-18.69%
Conservative	$+6.39\%$	$+9.92\%$	$+6.70\%$

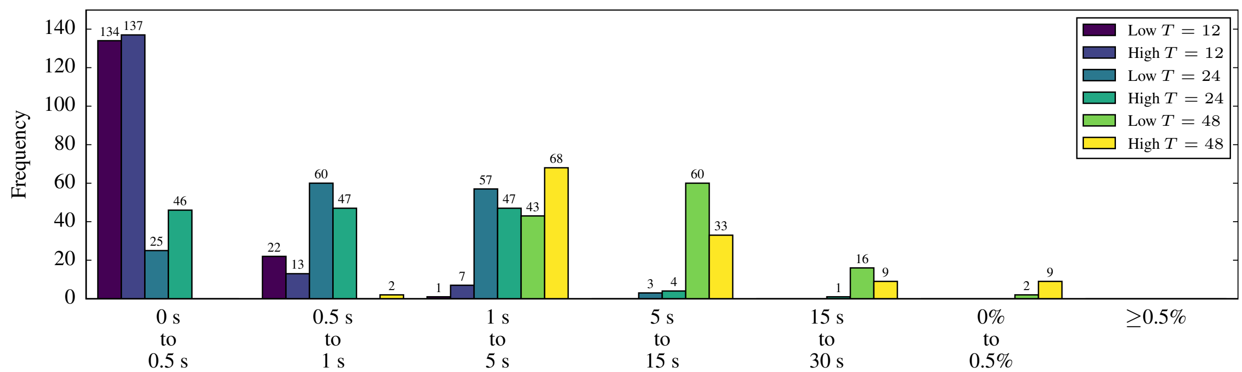


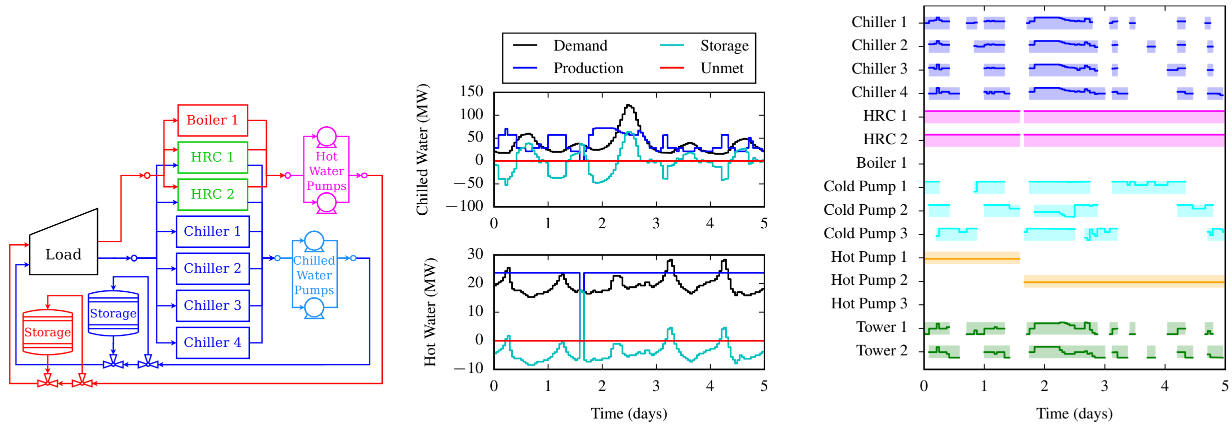
Figure 4: Solution times for uncertain demand example. Times indicate time needed to prove an optimal solution, while percentages are optimality gap after termination at 30 s. "High" and "Low" refer to demand forecast.

3.3 Terminal Constraints

While the previous examples have shown that a long prediction horizon is not necessary for low cost, it may still be desirable to obtain longer-term predictions, e.g., for diagnostics or higher-level planning. On such timescales, the discrete on/off decisions made by the full scheduling model are not as relevant, and instead the key choice is storage utilization. Thus, at the cost of less accurate predictions for resource consumption, computation can be made considerably easier by removing the discrete decision variables and piecewise-linear equipment models, replacing them with fully linear models (or other convex approximation). These changes yield a linear programming (LP) surrogate model that can be solved as a much more easily than the full scheduling model and thus can quickly provide a forecast for optimal storage use. Maximum equipment capacities are retained so that any surrogate schedule can be made feasible in the true by model relaxing dwell times and rounding up to minimum equipment capacities.

Although the surrogate model does not provide the discrete decisions needed to actually implement an equipment schedule, it can be used to enhance the solution quality of the full model by providing a reference trajectory for the storage tank. With the addition of a terminal constraint to the full model that requires the amount of storage to terminate on the reference trajectory, recursive feasibility (with respect to demand satisfaction) can be ensured. Because the surrogate model does not accurately calculate resource consumption rates, the storage trajectories it produces are likely to be suboptimal, but they are feasible. Thus, by using this trajectory as a terminal constraint, the full model is able to reoptimize the beginning of the trajectory for lower cost while still keeping enough storage to meet the future demand considered by the surrogate model.

As an example, we consider a larger system producing both hot and chilled water with the equipment shown in Figure 5a. We also use a demand profile within an extreme spike above the nominal capacity of the plant, thus requiring storage to be charged in advance. The optimal schedule is shown in Figures 5b and 5c. Notice that in the afternoon of the second day, the chillers are running at near-full capacity with significant withdrawal of storage. We then simulate the system in closed loop using both the "standard" approach used in the previous examples and also with the "surrogate" approach of this section that includes terminal constraints. Costs for these methods are given in Table 3. From this table, we see that while the longer-horizon standard approach does give the lowest (feasible) cost, the



(a) Central plant equipment. Cooling towers not shown.

(b) Chilled and hot water production.

(c) Equipment usage.

Figure 5: System and optimal schedule for terminal constraint example.

Table 3: Costs for full system. Simulation length was $N = 120$. Costs are relative to horizon Standard $T = 120$. Surrogate entries use terminal constraints from $T = 120$ surrogate models, while standard entries do not.

Horizon	Use Charges	Demand Charges	Total Charges
Standard $T = 72$	+0.27%	+0.00%	+0.24%
Standard $T = 24$	+0.89%	+0.00%	+0.78%
Standard $T = 12$*	+0.17%	+26.00%	+3.45%
Surrogate $T = 24$	+0.69%	+0.00%	+0.60%
Surrogate $T = 12$	+1.19%	+0.31%	+1.08%

*Infeasibility during simulation.

shorter horizons with the surrogate model are competitive. In particular, while the standard $T = 12$ has the lowest economic cost, it fails to meet the requisite demand due to insufficiently charging the storage tank; by contrast, the surrogate method with $T = 12$ avoids infeasibility due to the added terminal constraint.

For this larger system, we would also like to examine solution times, which are shown in Figure 6. Here we notice that the surrogate method subproblems require slightly more time due to the added terminal constraint, but are essentially just as easy to solve as the standard subproblems without the terminal constraint. The longer $T = 72$ standard models are often not solved to optimality within the time limit, but they come close. Note that the $T = 120$ surrogate models used to derive the terminal constraints can be solved in roughly 1 s (not included in Figure 6), and so they do not appreciably affect overall computational burden. Indeed, all methods are sufficiently quick for practical use.

4. CONCLUSIONS

In this paper, we have presented an MILP-based scheduling model for HVAC central plants and demonstrated that closed-loop scheduling is viable for practical use. While long-term demand forecasts may be inaccurate, only immediate decisions of a given schedule are actually implemented, and the schedule is reoptimized at the next timestep using updated forecasts before taking further action. As a result, plant operation is most directly influenced by accurate information, while the more speculative choices made at the end of the prediction horizon are not utilized. This feature reduces the burden for demand forecasts, which need only be accurate in the near term. In addition, we have shown that extremely long prediction horizons are unnecessary, with little improvement in cost for horizons longer than 48 h. Central plants with large, well-insulated storage tanks can benefit from a longer prediction horizon, but in such cases, a simplified LP surrogate model can be used to determine the storage trajectory, which can then be used as a terminal constraint in the shorter-horizon detailed MILP optimization. For two realistically-sized central plants discussed in the

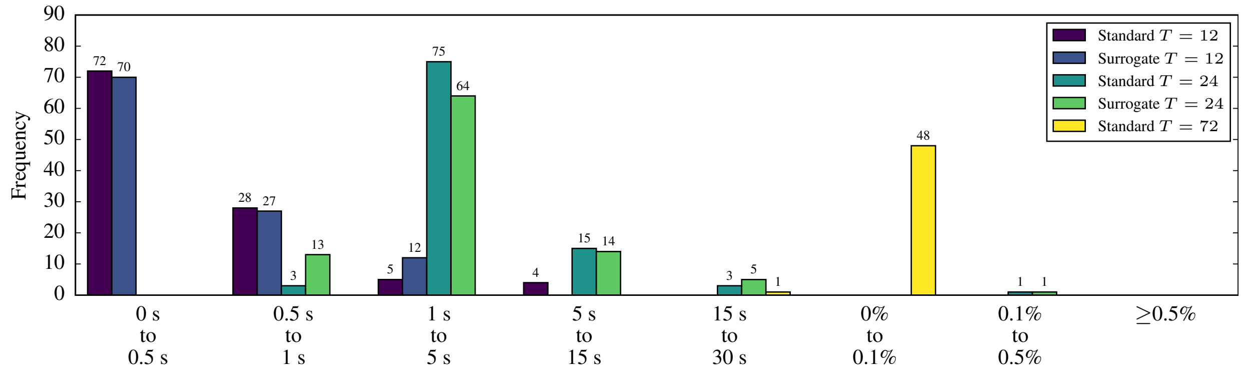


Figure 6: Solution times for terminal constraint example. Times are for solving the full model and do not include time spend solving the surrogate model to obtain terminal constraints. Bins are as in Figure 4.

examples, most subproblems could be solved to optimality within 30 s, while all could produce solutions no more than 0.5% away from optimality within that same time limit. Thus, with careful attention to problem formulation, real-time optimization is possible with standard commercial MILP solvers, leading to improved economic performance without relying on heuristic methods.

NOMENCLATURE

In the text, we use T for the length of the prediction horizon and N for the length of a simulation. All symbols used in the scheduling formulation are given below.

Sets and Subscripts

$j \in \mathbf{J}$	generators
$k \in \mathbf{K}$	resources
$\mathbf{K}_j^+ \subset \mathbf{K}$	resources produced by generator
$\mathbf{K}_j^- \subset \mathbf{K}$	resources consumed by generator
$t \in \mathbf{T}$	time points

Variables

p_{kt}	resource purchase
$p_k^{\max}$	peak resource purchase
q_{jkt}	resource production in generator
u_{jt}	generator binary on/off
y_{kt}	use of storage

Parameters

$\Xi_j(\cdot)$	model for generator consumption
$\xi_{jk}^{\min}, \xi_{jk}^{\max}$	min, max capacity for generator
ρ_{kt}	resource price
$\rho_k^{\max}$	demand charge
ϕ_{kt}	demand forecast
ψ_{kt}	exogenous margin

REFERENCES

Albadi, M. H., & El-Saadany, E. F. (2007). Demand response in electricity markets: An overview. In *2007 IEEE power engineering society general meeting* (pp. 1–5). IEEE.

Braun, J. E. (2007). A near-optimal control strategy for cool storage systems with dynamic electric rates. *HVAC&R Res.*, *13*(4), 557–580.

Cole, W. J., Edgar, T. F., & Novoselac, A. (2012). Use of model predictive control to enhance the flexibility of thermal energy storage cooling systems. In *American control conference (ACC), 2012* (pp. 2788–2793).

Henze, G. P., Biffar, B., Kohn, D., & Becker, M. P. (2008). Optimal design and operation of a thermal storage system for a chilled water plant serving pharmaceutical buildings. *Energ. Buildings*, *40*(6), 1004–1019.

Henze, G. P., Felsmann, C., & Knabe, G. (2004). Evaluation of optimal control for active and passive building thermal storage. *Int. J. Thermal Sci.*, *43*(2), 173–183.

Lee, T.-S., Liao, K.-Y., & Lu, W.-C. (2012). Evaluation of the suitability of empirically-based models for predicting energy performance of centrifugal water chillers with variable chilled water flow. *Appl. Energ.*, *93*, 583–595.

Li, Z., & Ierapetritou, M. (2008). Process scheduling under uncertainty: Review and challenges. *Comput. Chem. Eng.*, *32*(4-5), 715–727.

Ma, J., Qin, S. J., & Salsbury, T. (2014). Application of economic MPC to the energy and demand minimization of a commercial building. *J. Proc. Cont.*, *24*(8), 1282–1291.

Ma, Y., Borrelli, F., Hencey, B., Coffey, B., Bengea, S., & Haves, P. (2012). Model predictive control for the operation of building cooling systems. *IEEE Ctl. Sys. Tech.*, *20*(3), 796–803.

Ma, Y., Matuško, J., & Borrelli, F. (2015). Stochastic model predictive control for building HVAC systems: Complexity and conservatism. *IEEE Trans. Cont. Sys. Tech.*, *23*(1), 101–116.

Mendoza-Serrano, D. I., & Chmielewski, D. J. (2012). HVAC control using infinite-horizon economic MPC. In *Decision and control (CDC), 2012 IEEE 51st annual conference on* (pp. 6963–6968).

Pantelides, C. C. (1994). Unified frameworks for optimal process planning and scheduling. In *Proceedings on the second conference on foundations of computer aided operations* (pp. 253–274).

Risbeck, M. J., Maravelias, C. T., Rawlings, J. B., & Turney, R. D. (2015). Cost optimization of combined building heating/cooling equipment via mixed-integer linear programming. In *American control conference* (pp. 1689–1694). Chicago, IL.

Ryu, J.-h., Dua, V., & Pistikopoulos, E. N. (2007). Proactive scheduling under uncertainty: A parametric optimization approach. *Ind. Eng. Chem. Res.*, *46*(24), 8044–8049.

Shah, N., Pantelides, C., & Sargent, R. (1993). A general algorithm for short-term scheduling of batch operationsII. computational issues. *Comput. Chem. Eng.*, *17*(2), 229–244.

Subramanian, K., Maravelias, C. T., & Rawlings, J. B. (2012). A state-space model for chemical production scheduling. *Computers & Chemical Engineering*, *47*, 97–110.

Touretzky, C. R., & Baldea, M. (2014). Integrating scheduling and control for economic MPC of buildings with energy storage. *J. Proc. Cont.*, *24*(8), 1292–1300.

Touretzky, C. R., & Baldea, M. (2016). A hierarchical scheduling and control strategy for thermal energy storage systems. *Energ. Buildings*, *110*, 94–107.

US Department of Energy. (2016). *EnergyPlus 8.5.0.* Retrieved from `https://energyplus.net/`

Zavala, V. M., Constantinescu, E. M., Krause, T., & Anitescu, M. (2009). On-line economic optimization of energy systems using weather forecast information. *J. Proc. Cont.*, *19*(10), 1725–1736.

ACKNOWLEDGMENT

Sample data, equipment models, and funding provided by Johnson Controls, Inc.